End-to-End Modeling and Simulation of MPEG-2 Transport Streams over ATM Networks with Jitter

Wenwu Zhu, Yiwei Thomas Hou, Yao Wang, and Ya-Qin Zhang

***Abstract*— In this paper, the operation of MPEG-2 systems is modeled and simulated when an MPEG-2 transport stream is delivered through an ATM network with jitter. End-to-end packet-based analysis is performed for delivery of MPEG-2 transport streams over ATM networks. A novel approach to analyzing the decoder buffer behavior in the presence of network jitter is presented. The probability density function of the interarrival time of the ATM adaptation layer 5 (AAL5) protocol data unit (PDU) is derived from an MPEG-2 video source model and an ATM network jitter model. Based on a real-time decoding requirement of the MPEG-2 transport stream (TS) system target decoder (T-STD), the decoder buffer behavior is simulated. In this simulation, the packets' arrivals follow the derived probability density function of the AAL5 PDU interarrival time. The modeling and simulation results show the interactions among packet loss ratio, decoder buffer size, and network jitter level. We found that jitter affects decoder buffer size and packet loss ratio in a significant way.**

***Index Terms*—ATM networks, decoder buffer, jitter, MPEG-2.**

I. Introduction

THERE has been great interest in video transport over ATM networks in recent years. Due to the statistical multiplexing capability of ATM and increased transmission bandwidth capacity, ATM can support multimedia applications (i.e., audio, video, and data) simultaneously. Video can be transported over ATM networks either at a constant bit rate (CBR) or variable bit rate (VBR). Recent research interest has been focusing on VBR video since it utilizes network resources much more efficiently than CBR video. However, the use of statistical multiplexing in ATM networks introduces cell delay variation (CDV) or jitter and cell/packet loss, which may affect the video quality at the receiving end. This paper models and simulates the end-to-end transport of MPEG-2 video and investigates the impact of ATM network jitter on the operation of the MPEG-2 decoder buffer.

The MPEG-2 standard (ISO/IEC 13 818) specifies the operation and interaction of video and audio coding, as well as related system functions [1]. The standard assumes that the end-to-end delay from an MPEG-2 system's encoder to decoder is constant. This is necessary to ensure that the encoder and decoder clocks operate at the same frequency and the decoder buffer does not overflow nor underflow. On the other hand, an ATM network introduces jitter when transporting MPEG-2 packets from a source to a destination. Due to this jitter, the behavior of the decoder buffer differs from that of the encoder buffer, which may cause problems in decoder operations.

To the best of our knowledge, until now, most strategies for preventing a decoder buffer from underflowing or overflowing use rate control schemes without considering the ATM network jitter [2]–[4], except in [5]. In this paper, we employ a novel approach to investigate the decoder buffer behavior in the presence of ATM network jitter. Specifically, we first generate MPEG-2 transport stream (TS) packets based on a statistical model at the source. We then segment these TS packets into ATM cells and simulate their transport over ATM networks by introducing jitter into these cells. Finally, these cells are reassembled into TS packets, which are delivered into the upper layer TS packet decoder. We simulate the behavior of the decoder buffer under a real-time decoding requirement and show the interactions among packet loss ratio, decoder buffer size, and network jitter level.

The rest of this paper is arranged as follows: in Section II, source-side modeling of MPEG-2 transport stream is presented. In Section III, we derive the statistical distribution of the MPEG-2 protocol data unit (PDU) traffic arrival process at the decoder after being transported over ATM networks. Section IV defines decoder buffer operation. In Section V, simulation results of the decoder buffer behavior are given. Section VI concludes this paper.

II. Source-Side Modeling of MPEG-2 Transport Stream

The MPEG-2 system first packetizes elementary streams (compressed video sequences) to produce packetized elementary streams (PES) packets. PES packets are further combined with system information to form TS's or program streams (PS's) by multiplexing. The PS is designed for use in error-free environments and is suitable for applications which involve software processing of system information such as interactive multimedia applications on CD-ROM. PS packets have variable and relatively greater length. The TS is designed for use in environments where the errors are likely, such as lossy storage or noisy transmission media (e.g., video distribution over long distance networks and in broadcasting systems), in which packet losses may occur. The TS packet has a fixed length of 188 bytes. In this paper, we only consider the transport of a single TS. It is known that the MPEG-2 video is VBR in nature. In this section we present how to model encoded MPEG-2 video that is composed of 188-byte TS packets.

Manuscript received May 10, 1997; revised October 1, 1997. This paper was recommended by Associate Editor J. Brailean.

W. Zhu was with the Department of Electrical Engineering, Polytechnic University, Brooklyn, NY 11201 USA. He is now with Bell Labs, Lucent Technologies, Holmdel, NJ 07733 USA.

Y. T. Hou and Y. Wang are with the Department of Electrical Engineering, Polytechnic University, Brooklyn, NY 11201 USA.

Y.-Q. Zhang is with Sarnoff Corporation, Princeton, NJ 08543-5300 USA.

Publisher Item Identifier S 1051-8215(98)01516-X.

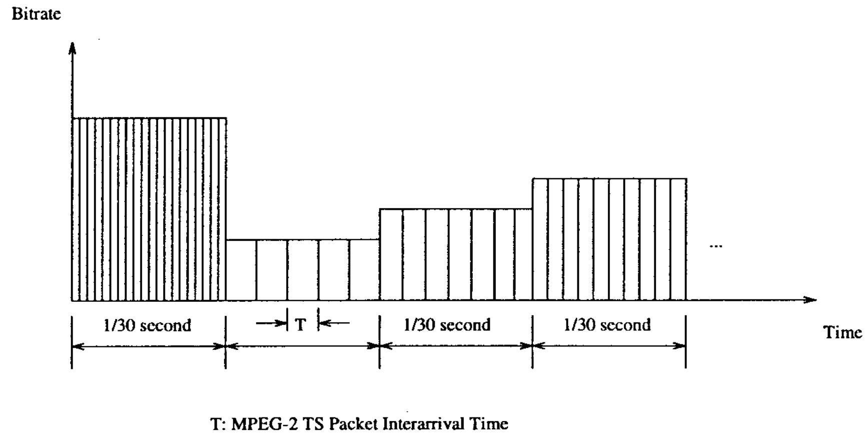

Fig. 1. MPEG-2 transport packet generation process.

To model the bit rate of a VBR video source, two classes of traffic models have been investigated (i.e., single source model [6], [8]–[10] and multiplexed source model [6], [7]). In this work, we only consider the single-source model. For low-motion videotelephony and medium-motion videoconferencing, there exist the autoregressive (AR) model [6], [8], the discrete autoregressive (DAR) model [9], and the discrete state continuous time Markov model [6], [8]. For the full motion MPEG video source, a three-class AR model with time-varying coefficients has been developed in [10].

An MPEG-2 encoded video source contains I, P, and B frames. In an I frame, each block (8×8) is coded using discrete cosine transform (DCT). In a P frame, each block is coded using motion-compensated prediction based on a previous frame. In a B frame, each block is predicted from both a previous and a future frame. The frame size (in bytes) of an MPEG-2 compressed video has a more complex statistical distribution than that of an H.261 compressed video and an MPEG-1 compressed video. This is in part caused by the mechanisms for handling interlaced video, for enabling scalability, compatibility, and error resilience, and by options for dealing with very high resolution video [1]. MPEG-2 encoded video source models have been developed for one-layer [11] and two-layer coding modes [12]. In [11], Heyman *et al.* described the statistical distribution of I, B, and P frames for one-layer mode. In [12], Chandra *et al.* modeled two-layer modes for I and P frames only.

In this paper, we only consider a video source generated by a one-layer coder. Specifically, we use the MPEG-2 video source model proposed by Heyman *et al.* [11]. In this work, the test sequence is a 10-min movie with a mean rate of 1.60 Mb/s and a peak rate of 6.543 Mb/s. The raw sequence is progressive and is coded using the main profile at main level (MP@ML) with a fixed quantizer step size. The coding parameters used are $M = 3$ (M is the space between anchor frames) and $N = 15$ (N is the space between I frames), which yield the following fixed frame pattern within each group of pictures (GOP):

I B B P B B P B B P B B P B B.

The I frames are modeled with a log-normal distribution with mean $\eta = 256.6$ and standard deviation $\sigma = 75.4$, both in units of cells.[1] After discarding outliers, the B and P frames can also be described by a log-normal distribution with $\eta = 93.4$, $\sigma = 41.6$ for P-frame and $\eta = 86.0$, $\sigma = 20.5$ for B-frame.

With the above statistics, we generate an MPEG-2 bitstream at the frame rate of $R = 30$ frames/s. This bitstream is then packetized into TS packets (with 188 bytes each) in a way that within each frame interval ($1/R = 1/30$ s), successive TS packets are generated with equal time interval. That is, TS packets are transmitted at the same rate within a frame, although the interarrival times of TS packets in separate frames are different. Illustration of an MPEG-2 TS packet generation process is shown in Fig. 1. Note that the area of each rectangular for each TS packet is 188 bytes. We can see that the MPEG-2 bitstream generated has a piecewise (or frame-wise) constant bit rate.

III. Transporting MPEG-2 TS Packets over ATM Networks

MPEG-2 TS packets need to be encapsulated into ATM cells before entering ATM networks. For this adaptation, ATM defines several layers based on the class of service. ATM adaptation layer 1 (AAL1) has been defined to support the CBR traffic, and an adaptive clock method for smoothing the network jitter has been proposed. The downside of AAL1 is that too much overhead is needed. AAL2 was proposed for VBR traffic. However, it was not well defined and has not been considered so far. ATM adaptation layer 5 (AAL5), originated for available bit rate (ABR) data transportation, can support both CBR and VBR video. We consider the transport of MPEG-2 streams over ATM networks using AAL5. In this scheme, MPEG-2 TS packets are mapped into AAL5 PDU's, each containing two TS packets. Each PDU is then converted to eight AAL5 cells in a way that the first seven AAL5 cells have 48 bytes payload each and the last AAL5 cell has 40 bytes payload plus an 8-byte trailer. Then a 5-byte ATM header is added to all AAL5 cells to form eight ATM cells with 53 bytes each. At the destination, eight ATM cells are reassembled into an AAL5-PDU, which is then decomposed into two adjacent TS packets (see Fig. 2). These TS packets are then delivered

[1] Each ATM cell contains 48 bytes of payload or 384 bits.

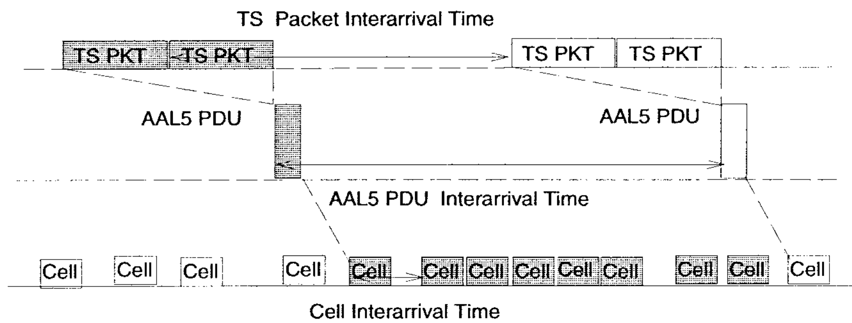

Fig. 2. The process of reassembling eight ATM cells into an AAL5-PDU.

to the upper layer TS packet decoder. Note that all the time scales in Fig. 2 are the same and the double headed arrows indicate the same interarrival time.

Let $\mathbf{T}_{\mathrm{PDU},\,S}$ denote the PDU interarrival time at the source side, $\mathbf{T}_{\mathrm{PDU},\,D}$ the PDU interarrival time at the destination, and $\mathbf{T}_{\mathrm{PDU},\,N}$ the network introduced jitter for a PDU. Then, we have

$$\mathbf{T}_{\mathrm{PDU},\,D} = \mathbf{T}_{\mathrm{PDU},\,S} + \mathbf{T}_{\mathrm{PDU},\,N}. \tag{1}$$

Let $\mathbf{T}_{\mathrm{CDV}}$ denote the jitter of a single cell (i.e., the cell delay variation) and $f_{\mathrm{CDV}}(T)$ its probability density function (PDF). Since the jitter of a PDU is determined by the jitter of the last cell in this PDU, the PDF of $\mathbf{T}_{\mathrm{PDU},\,N}$, $f_{\mathrm{PDU},\,N}(T)$, is equivalent to $f_{\mathrm{CDV}}(T)$ if we assume that the jitter of each ATM cell is independent and identically distributed. In this work, we model the cell jitter by a truncated Gaussian variable.

Assuming that $\mathbf{T}_{\mathrm{PDU},\,S}$ and $\mathbf{T}_{\mathrm{PDU},\,N}$ are independent, then

$$\begin{aligned} f_{\mathrm{PDU},\,D}(T) &= f_{\mathrm{PDU},\,S}(T) * f_{\mathrm{PDU},\,N}(T) \\ &= \int_0^{\infty} f_{\mathrm{PDU},\,N}(T-x) f_{\mathrm{PDU},\,S}(x)\, dx \end{aligned} \tag{2}$$

where $f_{\mathrm{PDU},\,S}(T)$ and $f_{\mathrm{PDU},\,D}(T)$ are the PDF's of PDU interarrival time at the source and decoder, respectively. Equation (2) gives the statistical distribution of the PDU interarrival time at the destination, from which the statistics of the PDU arrival process can be obtained (e.g., interarrival rate and squared coefficient of variation).

IV. Decoder Buffer System Definition

MPEG-2 video TS packets enter the T-STD decoder buffer after they are reassembled from ATM cells. We model the decoder buffer operation based on the MPEG-2 TS system target decoder (T-STD) [1].

T-STD is a hypothetical decoder (reference model) and provides a formalism for the timing and buffering relationship. In T-STD, data for an access unit (here an access unit means the coded representation of a picture frame) are removed at the earliest time consistent with the defined decoding time and decoding time stamp (DTS) encoded in the bitstream [1]. When the buffer does not contain the complete data for an access unit at its decoding time, the buffer is re-examined at a regular interval until the complete data is present in the buffer [1]. This is referred to as low-delay mode. Packet loss occurs when the buffer is full.

In our study, we consider a simplified version of the T-STD in which decoding is performed at a fixed time interval ($1/R$ s). We contend that even with this simplified decoder, the salient features of the MPEG-2 decoding mechanism still remain and the overall performance would not change in any significant way. Also, simulation results based on this simplified model give us the worst case scenario and provide an upper bound on the decoder buffer requirements.

In our simulations, at every frame interval of $1/R$ s, we check whether the TS packets for one video frame (an access unit) have arrived completely. If all the TS packets have arrived, they are removed from the decoder buffer instantly. Otherwise we wait and check for one more frame time interval (equal to $1/R$) to see whether the data for this frame are complete. After this extra frame time interval, the arrived data for this frame are removed from the decoder buffer even though they may be incomplete. The late PDU's are counted as lost PDU's in our packet loss ratio calculation (in addition to those PDU's lost due to decoder buffer overflow). Note that for the purpose of extracting information that may be essential for the decoding of the next frame, all the TS packets that arrive at the decoder buffer will be put into the buffer (unless the buffer is full) and decoded, even though some of them do not meet timing requirements.

V. Simulation Results

We have performed an end-to-end system simulation based on the source model and system definition described in Sections II and IV. First, MPEG-2 TS packets are generated based on the statistical distribution of the frame size at the source as described in Section II. Then these TS packets are segmented into ATM cells and transported over ATM networks. Finally, these cells are reassembled into PDU packets and delivered to the T-STD buffer. In this work, jitter is assumed to be a zero mean truncated (with 4σ truncation) Gaussian function. The variances (σ^2) of the PDU packet jitter considered include 10^{-6}, 2×10^{-6}, 5×10^{-6}, and 10×10^{-6} (in unit of s^2).

Fig. 3 shows the interactions among decoder buffer size, packet loss ratio (due to packet late arrivals and buffer overflow), and jitter levels. It can be seen that the jitter level significantly affects the packet loss ratio. For a given jitter level, as buffer size increases, the packet loss ratio decreases. In other words, if we want to keep the packet loss ratio lower, we have to increase buffer size to accommodate the jitter. It

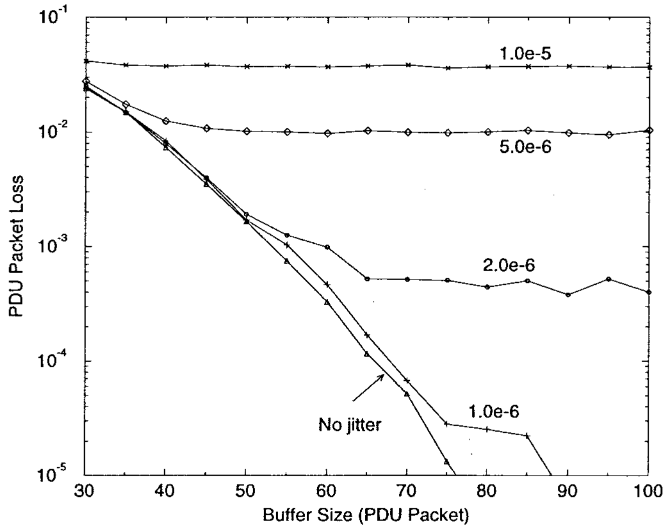

Fig. 3. Relations among packet loss ratio (due to packet late arrivals and buffer overflow), buffer size, and jitter level. The number above each curve specifies the variance of the PDU packet jitter simulated.

can been further seen that at a relatively small jitter variance, the packet loss ratio reduces when the buffer size increases. When the jitter is sufficiently large ($\sigma^2 > 5\times10^{-6}$), increasing buffer size does not affect the packet loss ratio. This is because, when the jitter is small, packet loss is mainly caused by buffer overflow. On the other hand, when the jitter is very large, late packets contribute to the majority of lost packets. In other words, the jitter level is the predominant factor that determines the packet loss ratio. It is useless to increase the physical buffer size since the late packets cannot meet the time constraint when the jitter is sufficiently large.

VI. Conclusions and Future Work

We have conducted an end-to-end modeling and simulation of MPEG-2 video transport over ATM networks with the consideration of the effect of jitter. The PDF of the PDU interarrival time at the destination is derived based on a video source model and a network jitter model. Our simulation is implemented based on the real-time decoding requirement of MPEG-2 T-STD. The modeling and simulation results show the tradeoff and interaction among the decoder buffer size, packet loss ratio, and network jitter level. When packet loss occurs, feedback information may be sent to the encoder to inform the encoder to change the bit rate by adjusting the quantization step. A study of the decoder buffer behavior when the encoder exercises this type of rate control strategy may yield more interesting and useful insights. In this work, we have derived the PDF of the PDU interarrival time at the destination. This enables us to simulate the packet arrival process at the decoder buffer. A theoretical study using queueing theory to analyze the decoder operation is another interesting topic for future work.

Acknowledgment

The authors would like to thank Dr. D. P. Heyman for the valuable discussion on MPEG-2 video source model, and Dr. A. Reibman for reviewing the draft and contributing constructive critiques.

References

[1] ISO/IEC JTC 1/SC 29/WG 11, *Generic coding of moving pictures and associate audio information, Part 1: Systems, Part 2: Video, Part 3: Audio,* CD 13818, May 1995.

[2] A. R. Reibman and B. G. Haskell, "Constraints on variable bit rate video for ATM networks," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 2, Dec. 1992.

[3] K. Joseph and D. Reininger, "Source traffic smoothing for VBR video encoders," in *Proc. 6th Int. Workshop on Packet Video,* Portland, OR, Sept. 26–27, 1994.

[4] C. Hsu and A. Ortega, "Joint encoder and VBR channel optimization with buffer and leaky bucket constraints," in *Multimedia Communications and Video Coding,* Y. Wang *et al.,* Eds. New York: Plenum, 1996, pp. 409–417.

[5] B. G. Haskell and A. R. Reibman, "Encoder/decoder buffer control for variable bit-rate channel," U.S. Patent 5 565 924, Oct. 1996.

[6] B. Maglaris, D. Anastassiou, P. Senn, and A. Tabatabai, "Performance models for statistical switching in packet video communications," *IEEE Trans. Commun.,* vol. 36, pp. 834–843, July 1988.

[7] D. A. Lucantoni, M. F. Neuts, and A. R. Reibman, "Methods for performance evaluation of VBR video traffic models," *IEEE/ACM Trans. Networking,* vol. 2, pp. 176–180, Apr. 1994.

[8] P. Senn, B. Maglaris, N. E. Rikli, and D. Anastassiou, "Models for packet switching of variable-bit-rate video sources," *IEEE J. Select. Areas Commun.,* vol. 7, pp. 865–869, June 1989.

[9] D. Heyman, A. Tabatabai, and T. V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 2, Mar. 1993.

[10] F. Yegenoglu, B. Jabbari, and Y.-Q. Zhang, "Motion-classified autoregressive modeling of variable bit rate video," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 3, pp. 42–53, Feb. 1993.

[11] D. Heyman, A. Tabatabai, and T. V. Lakshman, "Statistical analysis of MPEG2-coded VBR movie source," in *Multimedia Communications and Video Coding,* Y. Wang *et al.,* Eds. New York: Plenum, 1996, pp. 383–391.

[12] K. Chandra and A. R. Reibman, "Modeling two-layer MPEG-2 video traffic," in *Multimedia Communications and Video Coding,* Y. Wang *et al.,* Eds. New York: Plenum, pp. 419–426.