A NOVEL DETERMINISTIC TRAFFIC MODEL BASED ON A TWO-LEVEL ANALYSIS OF MPEG VIDEO SOURCES

Marcelo Dias de Amorim — amorim@gta.ufrj.br

Otto Carlos M. B. Duarte — otto@gta.ufrj.br

Networking and Automation Group – GTA
Federal University of Rio de Janeiro
COPPE – Department of Electrical Engineering
P.O. Box 68504 - 21945-970 - Rio de Janeiro - RJ - Brasil

Abstract

Integrated services networks have to support several applications with different quality of service requirements. Applications requiring strong service guarantees must be supported by bounded-delay services based on deterministic traffic models. This paper proposes a deterministic traffic model for video sources that takes advantage of specific MPEG coding characteristics. This model introduces a quality parameter that allows to balance complexity and accuracy. The results show that high accuracy can be obtained with low processing time.

1 Introduction

Computer network applications used to be much more sensible to data loss than to transport delays. The increasing number of new real-time applications introduces strict restrictions on the quality of service (QoS) guarantees to be provided by the network. As a general rule, the applications demand guarantees on the contracted QoS parameters.

In [1], two types of guaranteed services are detailed: deterministic service and statistical service. In a deterministic service, all of the packets of a connection are guaranteed to be delivered within the specified delay. For a statistical service, performance guarantees are also provided. However, these guarantees are not absolute, since they have a probability of being violated.

The quality of the service provided by the network depends basically on the traffic model, on the packet scheduler, and on the connection admission control [2, 3, 4]. Among these components, a simple and accurate traffic model is mandatory to provide an efficient service. Therefore, the traffic model must capture the traffic characteristics of an application with a minimum set of parameters that the user must send to the network manager for the effective traffic control and for the maximum utilization of the network resources.

Several deterministic traffic models have been proposed in the literature. The most important are the Fluid model [5], the Discrete model [1], the $(\vec{\sigma}, \vec{\rho})$ model [3, 4], and the D-BIND model [6, 7]. These models aim to characterize general variable bit rate (VBR) sources. However, some sorts of traffic, for example MPEG compressed video, are so important that if they are modeled by generic traffic models, specific properties will not be considered.

This work proposes a novel deterministic traffic model for MPEG compressed video traffics, called XGOP-B (eXtended GOP-Based model). This model has a simple algorithm to extract parameters from the traffics and obtains great results with a small number of parameters. These characteristics are very important, since the required policing mechanisms and connection admission control tests are reduced in terms of complexity. For the analysis and comparison among the models, several MPEG video streams are used.

2 Deterministic Traffic Characterization

A characterization is said to be deterministic when it is able to represent a traffic for all cases, i.e., when the source transmits at the required rate without buffering or discarding packets. In this paper, to develop a deterministic traffic model, the adopted approach is based on the definition of the traffic constraint function [5]. A function $\mathcal{F}(\cdot)$ is a traffic constraint function if during any interval of length δt the source does not transmit more than $\mathcal{F}(\delta t)$ bits. From another point of view, considering $\mathcal{G}[t, t+\delta t]$ as the actual number of bits transmitted by the source in the interval $[t, t+\delta t]$, the equation

$$\mathcal{G}[t, t+\delta t] \leq \mathcal{F}(\delta t), \qquad t > 0, \tag{1}$$

must be satisfied. In this way, the traffic constraint function $\mathcal{F}(\cdot)$ provides a time-invariant bound on $\mathcal{G}$.

The only restriction on $\mathcal{F}(\cdot)$ is that it must be an upper bound on the behavior of the source. Thus, there is an infinite number of possible solutions to $\mathcal{F}$. A deterministic traffic model selects, among all the possibilities, a group of solutions with specific characteristics.

For a source that transmits according to $\mathcal{G}$, the most accurate traffic constraint function that can be obtained is the empirical envelope [3, 4], defined as $\zeta(\delta t) = \texttt{sup}\ \mathcal{G}[t, t+\delta t]$, $t \geq 0$.

In practice, to represent the behavior of VBR sources by the empirical envelope may be extremely complex, requiring numerous parameters. Thus, the solution is to chose a simple upper approximation to ζ that satisfies

$$\mathcal{F}(\delta t) \geq \zeta(\delta t), \qquad \forall \delta t. \tag{2}$$

Since the traffic constraint function must be an upper approximation to ζ, and the empirical envelope is the most accurate deterministic characterization of a source, the main objective is to find a function $\mathcal{F}$ as close as possible to ζ. Analytically, the expression

$$\mathcal{F}(\delta t) - \zeta(\delta t), \qquad \forall \delta t, \tag{3}$$

must be minimized.

Several deterministic traffic models have been proposed in the literature. In this work we present four of them: the (σ, ρ) model [5], also known as Fluid model, the $(\vec{\sigma}, \vec{\rho})$ model [3], the $(x_{min}, x_{ave}, I, s_{max})$ [1], also called Discrete model, and the D-BIND model [6, 7].

In the (σ, ρ) model, for an interval of length δt, the source is limited to transmit no more than $\sigma + \rho \times \delta t$ bits. The $(\vec{\sigma}, \vec{\rho})$ model, which is an extension to the (σ, ρ) model, does not define a pair of parameters, but a two-dimensional array of them, resulting in a piece-wise concave function. Thus, this model is able to characterize a source more accurately. The Discrete model uses four parameters to model a traffic, $(x_{min}, x_{ave}, I, s_{max})$, where x_{min} is the minimum interval between two consecutive packet arrivals, x_{ave} is the maximum average interval between two packet arrivals within an interval of length I and s_{max} is the maximum packet size. The D-BIND (*Deterministic Bounding INterval Dependent*) model characterizes a source with a $R(I)$ function, where $R(I)$ is the maximum transmission rate within an interval I. The $R(I)$ function has N pairs $(R_k, I_k)|_{k=1,2,\ldots,N}$ and results in a piece-wise curve which is not necessarily concave.

Table 1 summarizes the traffic constraint functions for each one of the presented models.

3 The MPEG Standard

The MPEG (*Moving Picture Expert Group*) coding algorithm [8] was developed for storage of compressed video in digital media. However, recent applications have been using the MPEG standard for a wide variety of services, from multimedia stations to high-definition TV.

A MPEG coder produces three frame types: I frames (intra-pictures), that use intra-picture coding; P frames (predicted-pictures), that use a similar algorithm to I frames, and moving compensation based on previous I and P frames; and B frames (bidirectional-pictures), for bidirectional prediction, that use moving compensation based on previous and future reference frames.

A MPEG video is composed by a cyclic sequence of frames, called GOP (*Group Of Pictures*), which contains an initial I frame and some P and B frames, according to a pre-specified sequence, for example IBBPBBPBB. Most of MPEG video streams have I frames much larger than P and B frames. The burstiness characteristic comes from these different sizes of consecutive frames. In terms of transmissions, it means that information bursts occur at the beginning of each GOP. An example of empirical envelope for a MPEG video is shown in figure 1. In the graph, we can note the bursts produced by I frames.

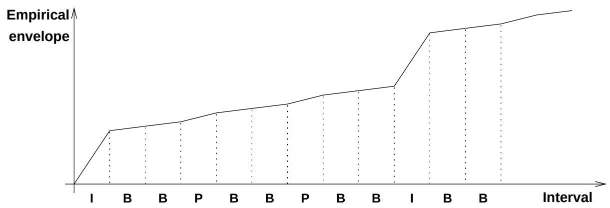


Figure 1: Empirical envelope example.

In practice, the compression rates reached by the coder vary a lot. The result may be an irregular sequence of frames. Figure 2 illustrates a real example of MPEG video trace[1].

The property described above is essential, but it is not the only one. As discussed, the compression rate for a MPEG video is not constant. The influence of this observation on the behavior of the empirical envelope is that it suffers from "sudden" deviations in the increase rate of the curve. This effect can be better observed in figure 3. The points where more accentuated variations occur are called **breakpoints** (ξ).

For the (σ, ρ) and Discrete models, the presence of breakpoints is undesired, since these models have the traffic constraint curves with constant increase rates.

[1] A video trace is an ordered sequence of frame sizes of a video sample.

Table 1: Traffic constraint function for the models.

Model	**Traffic constraint function**
(σ, ρ)	$\mathcal{F}_{(\sigma,\rho)}(\delta t) = \sigma + \rho \times \delta t$
$(\vec{\sigma}, \vec{\rho})$	$\mathcal{F}_{(\vec{\sigma},\vec{\rho})}(\delta t) = min_i\{\sigma_i + \rho_i \times \delta t\}$
Discrete	$\mathcal{F}_{Discrete}(\delta t) = \lfloor \frac{\delta t}{I} \rfloor \frac{I \times s_{max}}{x_{ave}} + min\{\lceil (\frac{\delta t}{I} - \lfloor \frac{\delta t}{I} \rfloor) \frac{I}{x_{min}} \rceil, \frac{I}{x_{ave}}\} \times s_{max}$
D-BIND	$\mathcal{F}_{D-BIND}(\delta t) = \frac{R_k I_k - R_{k-1} I_{k-1}}{I_k - I_{k-1}}(\delta t - I_k) + R_k I_k, \quad I_{k-1} \leq \delta t \leq I_k$

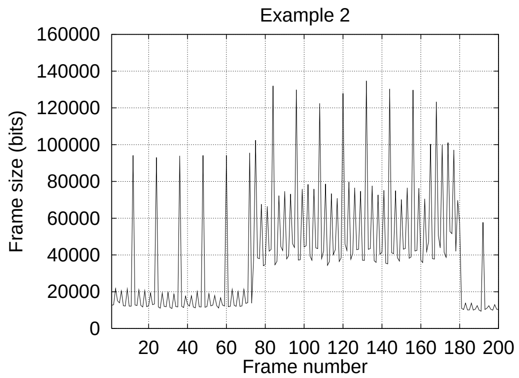


Figure 2: An example of MPEG video trace.

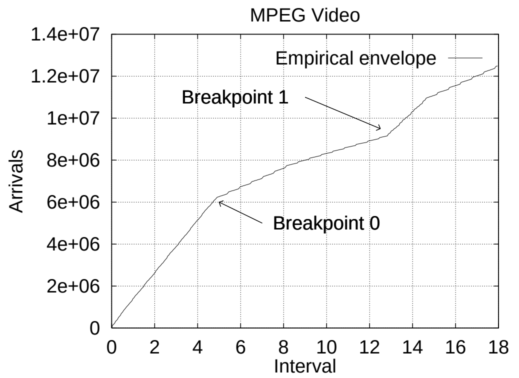


Figure 3: Influence of breakpoints on the empirical envelope.

At a negative breakpoint, i.e., when the average increase rate reduces, the traffic constraint function diverges from the empirical envelope, which may cause imprecision. In the same way, breakpoints are not suitable for the $(\vec{\sigma}, \vec{\rho})$ model when they are positive. The concave characteristic, which is intrinsic to this model, causes a divergence between the traffic constraint function and the empirical envelope at those points.

The main problem is to develop a mathematical way of identifying breakpoints. In the first step, we calculate a function that corresponds to the successive inclinations of the empirical envelope. This function stores the transmission rate of the source per time unit. These inclinations are particularly calculated with intervals proportional to the GOP size, and result in the First Derivative Function. In the next step, we obtain, from the First Derivative Function, a discrete function that represents only the variations of the First Derivative. This new function is called Second Derivative Function. The graphs in figure 4 show examples of first and second derivative functions.

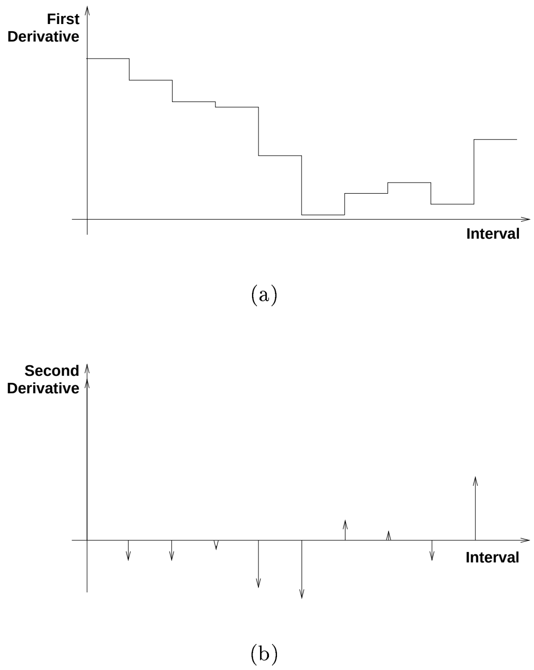


Figure 4: Examples of (a) First and (b) Second Derivatives.

The impulses plotted in figure 4(b) mark the points where the inclinations vary. However, it should be distinguished, among these points, the ones that must be chosen as breakpoints. This choice is performed according to the amount of accuracy demanded by the user. If more points are chosen, more accurate will be the model, since more parameters will be determined. On the other hand, if we select fewer points, the model will be simpler, but will be less accurate. To solve this problem, we adopt the technique of offering to the parameters extractor module a variable β called QoM-rate (Quality of Model rate). The breakpoints are chosen according to the following propositions:

Proposition 1 *Starting with an inclination, a breakpoint will be identified if this inclination varies, negatively or positively, of a factor β, with $\beta > 0$.*

Proposition 2 *Starting with an inclination, a breakpoint will be identified if the sum of an uninterrupted sequence of second derivatives varies, negatively or positively, with regard to the first inclination of this sequence, of a factor β, with $\beta > 0$.*

4 The XGOP-B Model

The XGOP-B is based on the definition of the GOP-B model (*GOP-Based model*). The GOP-B model is founded on the first analyzed property: the MPEG video sources present information bursts at the beginning of each GOP caused by the transmission of I frames. The GOP-B model defines a pair (η, ν) as traffic parameters. The η factor, called *impulse factor*, models the bursts caused by intra-frames. This factor must be added to the traffic constraint function at intervals proportional to the GOP size. The ν factor, named *average factor*, parameterizes the average increase of the traffic constraint curve.

In this way, the traffic constraint function for the GOP-B model comprises two parts. One directly proportional to all time intervals and other that is added to the traffic constraint function only for time intervals proportional to the GOP size. The GOP-B traffic constraint function is given by

$$\mathcal{F}_{GOP-B}(\delta t) = \lceil \frac{\delta t + T}{T \times D} \rceil \times \eta + \nu \times \delta t, \qquad (4)$$

where T is the inverse of the sample rate of the video and D is the GOP size[2].

The GOP-B model is based on group of pictures since the extractor parameters algorithm uses a GOP as the basic computation unit. While other models need to know *a priori* the values of the ζ function for all time intervals, the GOP-B model demands the empirical envelope only for one GOP, which significantly decreases the amount of time needed to calculate the parameters.

The GOP-B model incorporates the first property analyzed in this paper. Nevertheless, the breakpoint property is not considered. Thus, it is necessary to develop another traffic model that takes into account this property.

[2]For instance, if a GOP pattern is IBBPBB its size will be 6.

In order to solve the problem introduced by breakpoints, the XGOP-B model adopts a very simple solution. It changes the average inclination of the traffic constraint function at each breakpoint. For that, the traffic constraint function is subdivided in "zones" limited by the breakpoints, as shown in figure 5.

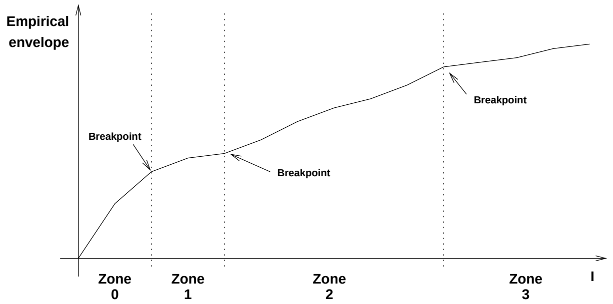


Figure 5: Empirical envelope subdivided in zones.

As for the GOP-B model, the main objective is to still obtain a pair (η, ν) to parameterize the traffic. However, we must now find several pairs (η_0, ν_0), (η_1, ν_1),..., (η_n, ν_n), one for each zone. Yet, if we provide as traffic parameters only the $(\vec{\eta}, \vec{\nu})$ array, the result will be a set of curves without any special characteristics. Thus, beyond the pairs (η_k, ν_k), we must include another array of offsets $\vec{\theta}$ for the curves and the array of breakpoints $(\vec{\xi})$. Thus, the traffic constraint function for the XGOP-B model is given by

$$\mathcal{F}_{XGOP-B}(\delta t) = \theta_k + \lceil \frac{\delta t + T}{T \times D} \rceil \times \eta_k + \nu_k \times \delta t, \quad \xi_{k-1} \le \delta t < \xi_k. \qquad (5)$$

We can note that this function is very similar to the traffic constraint function for the GOP-B model. Figure 6 graphically shows the behavior of this function.

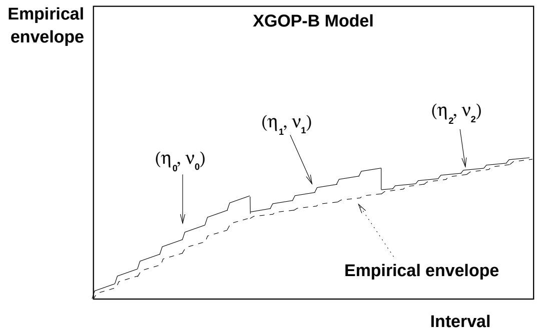


Figure 6: Traffic constraint function for the XGOP-B model.

5 Results

This section presents the comparison results among all of the models in terms of parameterization accuracy. For the measurements, 21 MPEG video samples were used [9]. These video streams were coded from films, sportive events and TV shows.

The comparison among the models is performed through a variable called Modeling Approach Loss (MAL). MAL indicates how much a result is different from the most accurate modeling, the empirical envelope. For each traffic constraint curve, its respective MAL is obtained from the difference of the areas under the traffic constraint curve and the curve given by the empirical envelope. MAL is calculated according to the following equation

$$PAM = \left(\frac{S_{\mathcal{F}}}{S_{\zeta}} - 1\right) \times 100\%, \tag{6}$$

where $S_{\mathcal{F}}$ is the area under the traffic constraint curve and S_{ζ} is the area under the empirical envelope. It should be observed that, the smaller is MAL, more accurate is the function. In a particular case, if MAL = 0 %, the two curves meet and the maximum accuracy is obtained, since the empirical envelope is the most precise function.

Figure 7 shows some of the results for the analyzed video traces. For the Discrete model, the calculus were performed with two values for the I parameter, 540ms and 1080ms, which correspond, respectively, to one and two GOP sizes. For the XGOP-B model, the QoM-rate is $\beta = 0.03$.

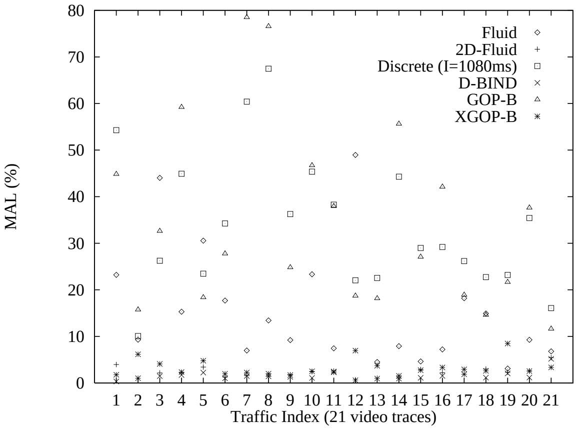


Figure 7: Modeling Approach Loss.

Figure 8 shows the traffic constraint curves for a cartoon video trace.

If the obtained results for $\beta = 0.03$ are not suitable, we can adjust β to catch the desired results. But, for smaller values of β, more breakpoints are identified and more traffic parameters are obtained, which results in greater parameterization accuracy. It is clear

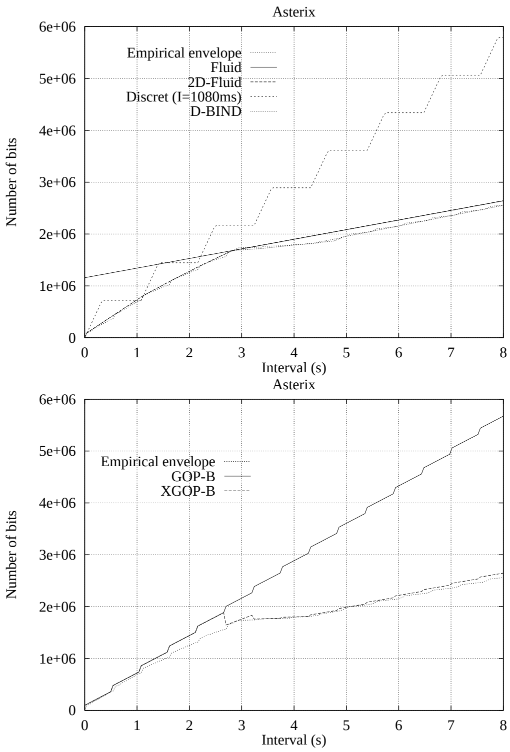


Figure 8: Traffic constraint function for the *Asterix* video trace.

that if we consider more breakpoints, the number of parameters increases and, consequently, more complex will be the policing mechanism.

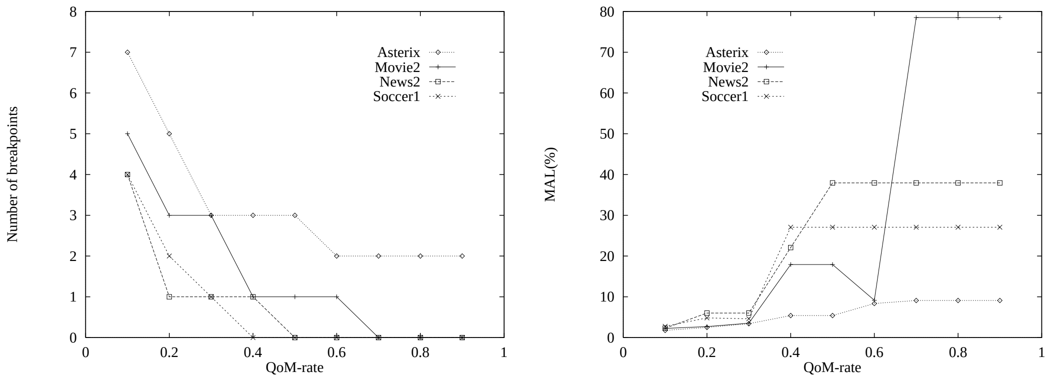
For some of the analyzed video traces, figure 9 shows the effects of β on the number of breakpoints and on the modeling accuracy.

6 Conclusions

In this work, we proposed a novel deterministic traffic model, called XGOP-B, that takes advantage of specific characteristics of MPEG-compressed video streams. This model is based on a two-level analysis of MPEG video traffics: frame-level and GOP-level analysis. The frame-level analysis considers that MPEG video sources generate information bursts at specific time intervals and the GOP-level analysis takes into account that these sources transmit at different rates for different lengths of time intervals.

In opposition to previous models, the XGOP-B parameters are obtained with a unique solution. Thus, given a video trace, the parameters extractor algorithm outputs only one set of values. It eliminates inaccuracy introduced by a bad choice.

The numerical results show that the XGOP-B model can achieve high modeling accuracy, which is quantified by calculating the achievable MAL. For the



(a) β × breakpoints

(b) β × MAL

Figure 9: β relations.

majority of the traces, we have obtained negligible values of MAL. It means that the XGOP-B model results in a traffic constraint function that almost meets the optimal characterization. Furthermore, this model introduces a quality parameter that allows to balance complexity and accuracy.

The proposed model seems to be a simple and attractive solution to obtain high accuracy with low processing time.

References

[1] D. Ferrari and D. Verma, "A scheme for real-time channel stablishment in wide-area networks," *IEEE Journal on Selected Areas in Communications*, pp. 368–379, Apr. 1990.

[2] H. Zhang and D. Ferrari, "Improving utilization for deterministic service in multimedia communication," *1994 International Conference on Multimedia Computing and Systems*, pp. 295–304, May 1994.

[3] 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 tradeoffs," *IEEE/ACM Transactions on Networking*, pp. 352–362, June 1996.

[4] E. W. Knightly, D. E. Wrege, J. Liebeherrand, and H. Zhang, "Fundamental limits and tradeoffs of providing deterministic guarantees to VBR video traffic," in *ACM Sigmetrics*, (Ottawa, Canadá), May 1995.

[5] R. L. Cruz, "A calculus for network delay, part I: Network elements in isolation," *IEEE Transactions on Information Theory*, pp. 114–121, Jan. 1991.

[6] E. Knightly and H. Zhang, "Traffic characterization and switch utilization using a deterministic bounding interval dependent traffic model," in *IEEE INFOCOM'95*, (Boston, MA), pp. 1137–1145, Apr. 1995.

[7] E. W. Knightly and H. Zhang, "D-BIND: An accurate traffic model for providing QoS guarantees to VBR traffic," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 219–231, Apr. 1997.

[8] D. L. Gall, "MPEG: A video compression standard for multimedia applications," *Communications of the ACM*, pp. 305–313, Apr. 1991.

[9] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modelling in ATM systems," *Proceedings of the 20th Annual Conference on Local Computer Networks*, 1995.