

MODELING COMPRESSED FULL-MOTION VIDEO

Benjamin Melamed

Rutgers University
Faculty of Management
Department of MSIS
New Brunswick, NJ 08903, U.S.A.

ABSTRACT

This paper presents a general approach to modeling VBR (variable bit rate) compressed full-motion video. The salient feature of such video is the existence of scenes. Scene structure presents special problems if one is to construct high-fidelity source models to drive Monte Carlo simulations. Intuitively, scene transitions are often accompanied by a substantial change of bit rate (the size of compressed pictures), as well as the level of burstiness. This paper argues that a modulated model can handle well scene modeling, and illustrates this point by an example.

1 INTRODUCTION

As high-speed (broadband) telecommunications networks proliferate, video traffic is expected to constitute a significant part of the teletraffic load. Furthermore, compressed video, also known as VBR (variable bit rate) video, is not only a high-volume teletraffic, but it also tends to be highly bursty. In a nutshell, this is due to the fact that successive video frames (pictures) differ very little (except across a scene transition), and at 24 - 30 frames a second, uncompressed video consumes large amounts of storage and bandwidth (transmission capacity). Compression is used to reduce the transmission and storage requirements of video, but even then the network must carry large volumes of bursty traffic.

The size and complexity of broadband networks, combined with the diversity of their projected traffic mix (voice, video, file transfers etc.), render the task of network design and management a difficult one. Traditional queueing models which do not address autocorrelations in traffic are often unsuitable for modeling components of telecommunications networks. Simulation studies are frequently the only viable alternative. In particular, accurate source models of traffic workload are a critical factor in the efficacy of network models.

The importance of modeling burstiness cannot be

overstated. Intuitively, burstiness is present in a traffic process, if arrival points appear to form visual clusters on the time line. The mathematical underpinnings of burstiness are complex; two main factors contributing to traffic burstiness are the marginal distribution and autocorrelation function of traffic interarrival times or bit rates, particularly strong positive short-term autocorrelations (Frost and Melamed 1994). The impact of autocorrelation in traffic processes on queueing measures such as mean queue length, mean waiting times and loss probabilities in finite buffers, can be very dramatic, even in light traffic (Hefes and Lucantoni 1998, Latouche 1985, Livny et al. 1993). Worse still, when the actual traffic has appreciable short-term positive autocorrelations, then traffic models that ignore autocorrelation err badly on the optimistic side.

Most work on video modeling addresses short sequences of empirical data consisting of one or a few scenes; furthermore, video source models are frequently designed to capture either the empirical distribution (first-order statistics) or the empirical autocorrelation function (second-order statistics), but not both; see Frost and Melamed (1994), Garrett (1993) and references therein. In contrast, we advocate models that accurately capture both. Thus, candidate models should approximate well the marginal distribution and the autocorrelation function (equivalently, spectral density) of the empirical bit rate process, *simultaneously*, and sample path realizations of the model should bear a strong “*resemblance*” to the empirical record. While time series “resemblance” is a subjective judgment which cannot be precisely quantified (its exact meaning is best left at the intuitive level), we point out that such judgments are routinely made by analysts in an informal manner. Several classes of such models are now available with practical modeling tools, including TES (Melamed 1991; Jagerman and Melamed 1992a, 1992b, 1994) and ARTA (Cario and Nelson 1996, 1997).

We now proceed to outline the proposed modeling methodology, following which we illustrate it briefly

with an example.

2 OUTLINE OF MODELING METHODOLOGY

The methodology presented here gives rise to a modulated model, with the following characteristics:

- The modulating model has few states (often a simple finite Markov chain). The states represent scene classes (types) and their transition structure aims to capture scene transitions
- The modulated models are associated to modulating states and aim to capture detailed behavior of individual scene classes. Consequently, these models are elaborate in that they are designed to capture scene characteristics of first-order and second-order statistics.

The modeling procedure consists of the following main steps:

1. The entire movie (sequence of bit rates) is first segmented into scenes.
2. The scenes are partitioned into classes by a clustering algorithm.
3. Each scene in the empirical data is assigned a class tag, yielding a random sequence of class tags.
4. The tag sequence is modeled as a finite-state process, capturing scene-transition probabilities. A typical choice is a simple Markov chain with a few states (perhaps 5 states at most).
5. The bit rate process of each scene class is separately modeled as an elaborate process, capturing individual scene statistics.
6. For each scene class, scene durations are modeled as a class-dependent process (often a renewal process).

The resulting model can be readily simulated as follows:

1. A scene class tag is generated according to the modulating process.
2. The associate duration process is sampled for a scene duration.
3. The associated bit rate process is generated for the aforementioned duration.
4. Go to Step 1.

We now proceed to describe the steps of the methodology in greater detail. As a working example, we shall use throughout the movie “Star Wars”, compressed by a variant of the JPEG compression scheme (JPEG 1990). This movie is a good representative of full-motion video, because it consist of scenes with diverse visual complexity and motion content. The data record consists of some 171,000 frame sizes (in bytes per frame) — the equivalent of about two hours of video elapsed time at the rate of 24 frames a second. A detailed description of the raw empirical bit rate data may be found in Garrett (1993), and the details of its modeling in Melamed and Pendarakis (1994).

3 SCENE IDENTIFICATION

To illustrate how scenes might be identified, consider Figure 1 which displays a portion of the empirical bit rate sequence.

The VBR sequence of bit rates, $\{X_n\}$, exhibits considerable burstiness as evidenced by Figure 7. In particular, the upper-right canvas in Figure 7 shows that the VBR sequence of bit rates, $\{X_n\}$ has high short-term empirical autocorrelations – an indication of high burstiness. Scene transitions are clearly visible in Figure 1 and the upper-left canvas of Figure 7 as abrupt large-magnitude changes in the local average bit rate.

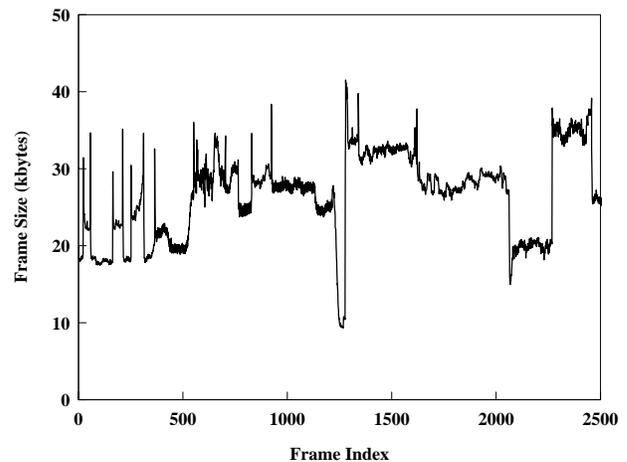


Figure 1: A Segment of the Empirical Bit Rate Record

In order to detect scene boundaries in the empirical bit rate record, one may follow a methodology similar to that introduced in Lazar et al. (1993). A scene change is declared in the bit rate stream when the absolute value of the difference between successive bit rates exceeds a certain threshold in a sus-

tained manner. For more details, refer to Melamed and Pendarakis (1994).

4 SCENE CLASSIFICATION

The next step is to construct a parsimonious mathematical description of the observed scenes, extracted from the empirical data, which is nonetheless sufficiently accurate to capture first-order and second-order statistics. Obviously, the construction of a separate model for each scene is neither parsimonious nor practical (there might be hundreds of scenes in a movie). Instead, we assume that observed scenes are drawn as realizations from a small number of stationary probability laws (classes), with class-dependent durations. For simplicity, we also assume that transitions among scene classes are governed (modulated) by a simple Markov chain.

To carry out scene classification, a classification criterion (or criteria) must be postulated first. Reasonable criteria include scene bit rate mean or scene bit rate variance (or both). Experience shows (Melamed and Pendarakis 1994), however, that scene bit rate mean alone is a good classification criterion (this point will be revisited below).

To partition the observed scenes into classes, one may use for instance *partitional clustering* (Jain 1988). The general abstract partitioning problem is formulated as follows. Given n elements in a d -dimensional metric space, find a partition of the elements into K equivalence classes (clusters), such that elements in each cluster are “closer” to each other than to elements in all other clusters. The value of K may or may not be prescribed. A clustering metric formalizing the notion of closeness has to be selected. The most common metric used in practice is the Euclidean distance, $\eta(v, w) = \sqrt{\sum_{i=1}^d (v_i - w_i)^2}$, for vectors v and w in d -dimensional Euclidean space. The clustering heuristic proceeds iteratively in a trial-and-error fashion. The algorithm requires the specification of two auxiliary parameters: the *minimum distance between cluster centroids*, η_{min} , and the *maximum variation within a cluster*, E_{max}^2 ; see Jain (1988) for more details.

Figure 2 depicts the results of applying the selected clustering procedure to the working example. Four visually distinct clusters are clearly discernible in Figure 2, each marked by a different character. The figure demonstrates the efficacy of the mean scene bit rate as a clustering criterion, and the comparable insignificance of the corresponding variance criterion. To see that, note that vertical lines have been drawn in Figure 2 to demarcate approximate cluster bound-

aries by bit rate means; on the other hand, a similar segregation by bit rate variances is neither evident nor logically compelling.

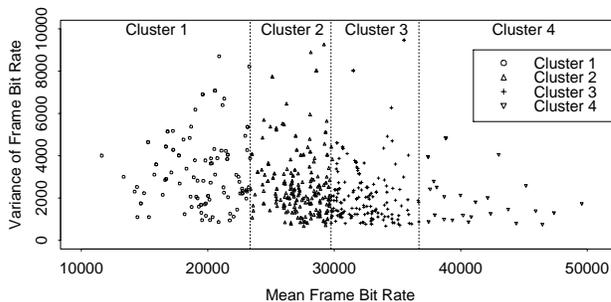


Figure 2: Clustering of a Subset of “Star Wars” Scenes by Mean and Variance Parameters

Hereafter, scenes will be clustered by bit rate mean only, and the resultant clusters will be identified with *scene classes*. The set of scene classes are characterized by non-overlapping intervals corresponding to ranges of mean scene bit rates.

5 MODELING SCENE CLASS MODULATION

The easiest way to model well the random sequence of scene classes is to assume them to be governed (modulated) by a simple Markov chain $\{M_n\}$. The estimation of the 1-step transition matrix, \mathbf{P} of $\{M_n\}$, is straightforward. A class tag is assigned to each scene in the empirical record, resulting in a realization, $\{c_n\}$, of class tags. This sequence is traversed, left to right, and class transitions are counted for each pair of classes. Normalization of these counts (division by the total number of class transitions) then yields the desired transition matrix, \mathbf{P} .

6 MODELING SCENE CLASSES

Individual scene classes may be modeled by analysts, using their favorite model. Recall, however, that accurate models capturing bit rate distribution and autocorrelation function per scene class are highly recommended. Two such modeling methodologies are TES (Transform-Expand-Sample) (Melamed 1991; Jagerman and Melamed 1992a, 1992b, 1994) and ARTA (Autoregressive To Anything) (Cario and Nelson 1996, 1997).

For each scene class, the empirical histogram of bit rates is calculated, by pulling all observed scenes from that class (cluster). For the working example, Figure

3 shows that these histograms are reasonably separated, and each is bell shaped.

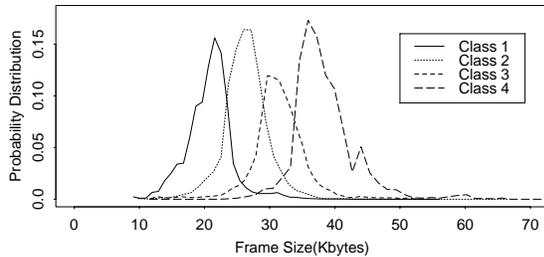


Figure 3: Empirical Histograms of Bit Rates for Individual Classes

To estimate the empirical autocorrelation of the bit rate process for each scene class, one may use the estimator in Cox and Lewis (1968), which for an empirical sequence, $\{Y_n\}_{n=0}^N$, is given by

$$\hat{\rho}(\tau) = \frac{1}{\hat{\sigma}_H \hat{\sigma}_T(N-\tau)} \sum_{j=0}^{N-1-\tau} (Y_j - \hat{m}_H)(Y_{j+\tau} - \hat{m}_T), \quad (1)$$

where

$$\hat{m}_H = \frac{1}{N-\tau} \sum_{j=0}^{N-1-\tau} Y_j, \quad \hat{\sigma}_H^2 = \frac{1}{N-\tau} \sum_{j=0}^{N-1-\tau} (Y_j - \hat{m}_H)^2,$$

$$\hat{m}_T = \frac{1}{N-\tau} \sum_{j=\tau}^{N-1} Y_j, \quad \hat{\sigma}_T^2 = \frac{1}{N-\tau} \sum_{j=\tau}^{N-1} (Y_j - \hat{m}_T)^2,$$

all depend on τ . Here the product terms in (1) include only samples from the *same* scene, as successive scenes belonging to the same class are not adjacent. The empirical autocorrelation functions per scene class (cluster) are depicted in Figure 4.

In the working example, each scene class was modeled as a TES process to accurately capture first-order and second-order statistics within scene classes (clusters).

7 MODELING SCENE CLASS DURATIONS

Scene durations for each scene class may be modeled by the analyst's favorite model, so as to capture the corresponding distributions and autocorrelation functions. Experience shows, however, that durations often form a renewal process (unlike bit rates).

For the working example, Figure 5 displays the empirical distribution of scene durations (in number of

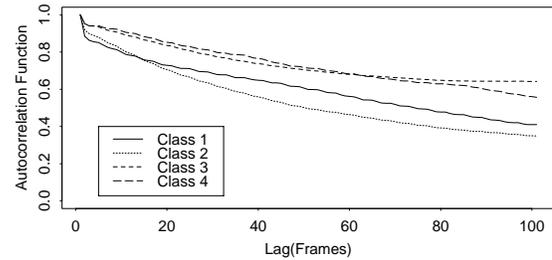


Figure 4: Empirical Autocorrelation Functions of Bit Rates for Individual Classes

frames) for each scene class. Though sharing the same functional form, a class-dependent duration distribution turns out to be a reasonable choice.

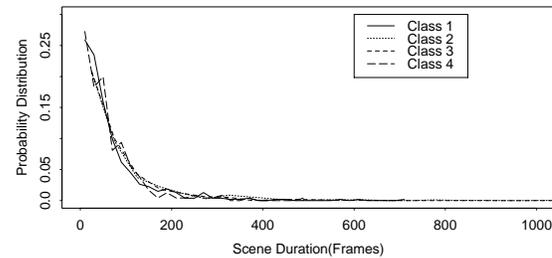


Figure 5: Empirical Distribution of Durations for Individual Classes

Figure 6 displays the autocorrelation functions of scene durations for each scene class. A visual inspection verifies the low magnitude of positive-lag autocorrelations (for lags greater than 5, the autocorrelations practically vanish).

Consequently, modeling scene durations as class-dependent iid random variables (renewal sequences) is a reasonable choice which also simplifies the model in this case.

8 MODEL VALIDATION

To validate the overall model, one runs a Monte Carlo simulation of the model to generate one or more long realizations of the VBR process, each roughly matching the size of the empirical data.

Figure 7 displays a screen image of TESTool (a software package supporting TES modeling (Hill and Melamed 1995), depicting the modeling results for

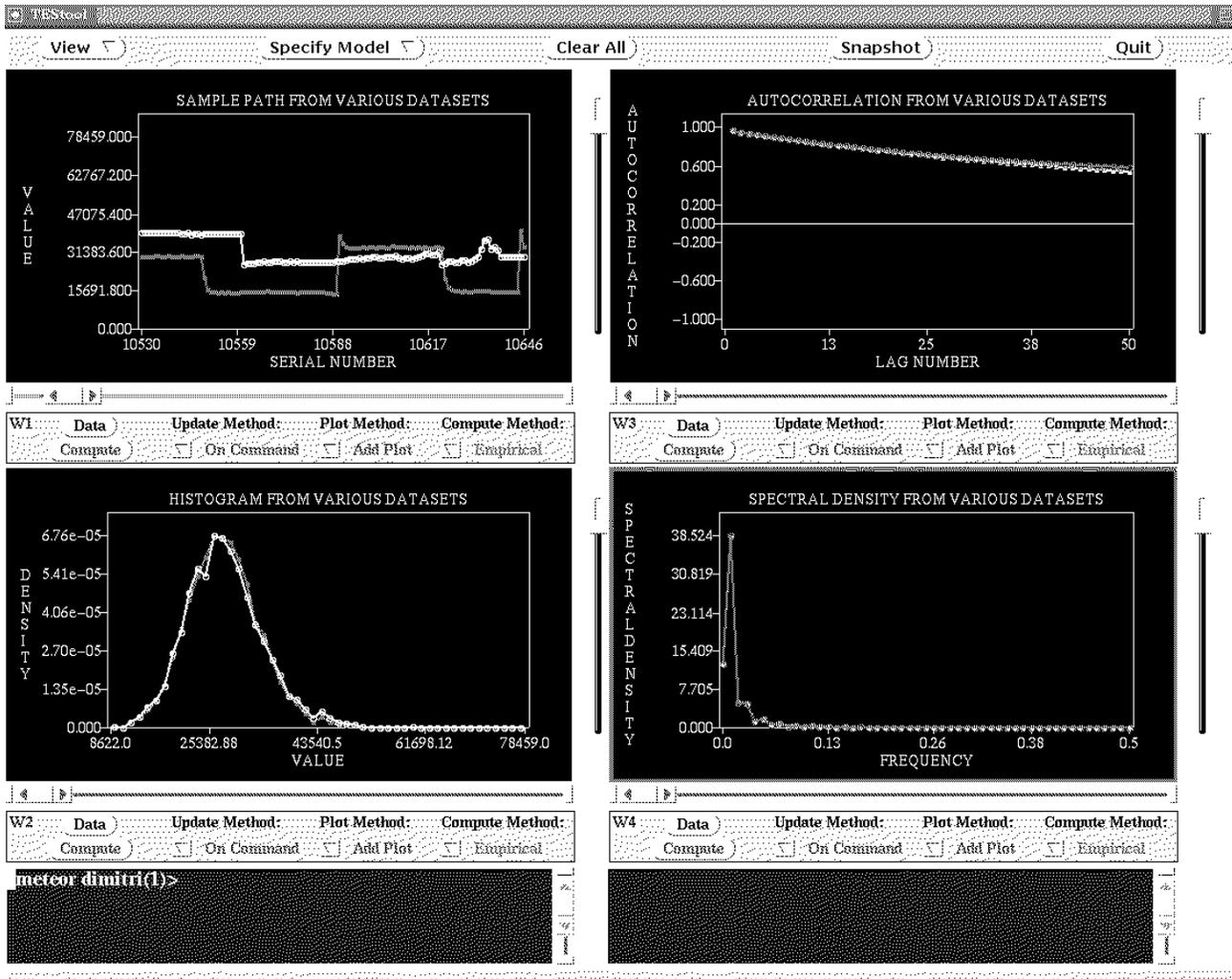


Figure 7: Comparison of Statistics of the Empirical and Simulated “Star-Wars” Bit Rate Process

the working example. The screen consists of four canvases, each superimposing simulation-based statistics (bold curves) on their empirical counterparts (fainter curves). The upper-left canvas displays a subset of the original record superimposed on its simulation-based counterpart. The lower-left canvas depicts the empirical histogram and its simulation-based counterpart, computed from the data in the upper-left canvas. Similarly, the upper-right and lower-right canvases contain the corresponding autocorrelation functions and spectral densities, respectively. Observe that all statistics exhibit excellent agreement between the empirical data and their simulation-based counterparts. In fact, the histogram, autocorrelation function and spectral density curves are barely distinguishable. In addition, the simulated time series bears a strong qualitative resemblance to the empirical record. Even though such a subjective

judgment cannot be precisely quantified, the results shown in Figure 7 support the claim that the model of the working example fits the empirical record well, both quantitatively and qualitatively. Informally, the TES processes which model individual scenes capture short-term behavior, while the modulating Markov chain governing scene changes captures longer-term scene behavior.

We conclude this section with a number of remarks and observations. First, it is somewhat surprising that four scene classes suffice here to yield a good overall model. Experimentation with fewer classes revealed deficiencies in the resultant fit, whereas more classes brought no perceptible improvement (at the cost of higher complexity). Here, the clustering analysis was extremely valuable in identifying a parsimonious model. Second, an attempt to model scene durations as class-independent resulted in a poor fit. It

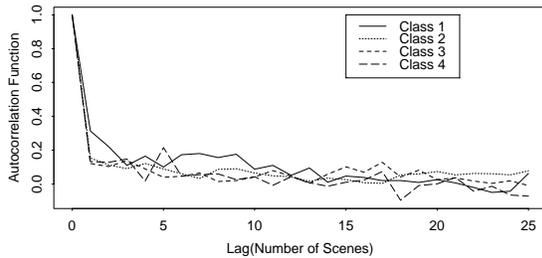


Figure 6: Empirical Autocorrelation Functions of Durations for Individual Classes

appears necessary to model these as class-dependent.

9 CONCLUSION

The results suggest that the modeling methodology proposed here can be used to create parsimonious high-fidelity models of compressed video under a variety of compression schemes. For example, TES-based models have been constructed for MPEG compression (Reininger et al. 1994) and H.261 compression (Lee et al. 1992, Melamed et al. 1992); see also Melamed and Sengupta (1992) for further details. These can then be used to create sets of representative models of full-motion video traffic to drive realistic Monte Carlo simulations of broadband telecommunications networks.

ACKNOWLEDGMENTS

This work was supported in part by DARPA/ITO grant AO F316.

The modeling work reported here was done jointly with Dimitrios E. Pendarakis. The empirical data were provided to us courtesy of Mark Garrett, Bell Communications Research (Bellcore) and Martin Vetterli, University of California, Berkeley.

REFERENCES

- Cox, D.R. and Lewis, P.A.W. 1968. *The statistical analysis of series of events*, Methuen.
- Cario, M.C. and Nelson, B.L. 1996. Autoregressive to anything: time-series input processes for simulation. 1996. *OR Letters* 19:51–58.
- Cario, M.C. and Nelson, B.L. 1997. Numerical methods for fitting and simulating autoregressive-to-anything processes. *Inform Journal on Computing*, forthcoming.

- Frost, V. and Melamed, B. 1994. Traffic modeling for telecommunications networks. *IEEE Communications Magazine* 32:3, 70–81.
- Garrett, M.W. 1993. Contributions toward real-time services on packet switched networks”, Ph.D. Dissertation, Columbia University.
- Hill, J.R. and Melamed, B. 1995. TESTool: a visual interactive environment for modeling autocorrelated time series.- *Performance Evaluation* 24(1&2):3–22.
- Heffes, H. and Lucantoni, D.M. 1986. “A markov modulated characterization of packetized voice and data and related statistical multiplexer performance”, *IEEE J. on Selected Areas in Communications* SAC-4:856–868.
- Jagerman, D.L. and Melamed, B. 1992a. The transition and autocorrelation structure of TES Processes Part I: general theory. *Stochastic Models* 8(2):193-219.
- Jagerman, D.L. and Melamed, B. 1992b. The transition and autocorrelation structure of TES processes Part II: special cases. *Stochastic Models* 8(3):499-527.
- Jagerman, D.L. and Melamed, B. 1994. The spectral structure of TES processes, *Stochastic Models* 10(3):599–618.
- Jain, A.K. 1988.- *Algorithms for clustering data*, Prentice Hall, 1988.
- Latouche, G. “An exponential semi-markov process, with applications to queueing theory”.- 1985. *Stochastic Models*, 1(2):137–169.
- JPEG technical specification:- revision (Draft), Joint Photographic Experts Group, ISO/IEC JTC1/SC2/WG8, CCITT SGVIII”, 1990.
- Lazar, A.A., Pacifici G. and Pendarakis, D.E. 1993. “modeling of video sources for real time scheduling. *Technical Report 324-93-03*, CTR, Columbia University.
- Lee, D.S., Melamed, B., Reibman, A. and Sengupta, B. 1992. TES modeling for analysis of a video multiplexer. *Performance Evaluation* 16:21–34.
- Livny, M., Melamed B. and Tsiolis, A.K. 1993. The impact of autocorrelation on queueing systems. *Management Science* 39:322–339.
- Melamed, B. 1991. TES: a Class of methods for generating autocorrelated uniform variates. *ORSA J. on Computing* 3,(4):317–329.
- Melamed, B., Raychaudhuri, D., Sengupta, B. and Zdepski, J. 1992.- TES-based traffic modeling for performance evaluation of integrated networks. *Proceedings of INFOCOM '92* 1:75–84, Florence, Italy.
- Melamed, B. and Sengupta, B. 1992. TES modeling of video traffic. *IEICE Transactions on Communi-*

cations E75-B(12):1292–1300.

- Melamed, B. and Pendarakis, D. 1994.- A TES-based model for compressed “Star Wars” video”, *Proceedings of the Communications Theory Mini-Conference at GLOBECOM '94*, 120–126, San Francisco, California.
- Reininger, D., Melamed, B. and Raychaudhuri, D. 1994. Variable bit rate MPEG video: characterization, modeling and multiplexing. *Proceedings of the 14-th International Teletraffic Congress* 1a:295–306, Antibes Juan-les-Pins, France.

AUTHOR BIOGRAPHY

BENJAMIN MELAMED is a Professor at the Rutgers University Faculty of Management, Department of MSIS. His research interests include system modeling and analysis (especially telecommunications systems and traffic), simulation, stochastic processes and visual modeling environments. He authored or co-authored over 80 papers. Melamed received a B.Sc. degree in Mathematics and Statistics from Tel Aviv University in 1972, and a M.S. and Ph.D. degrees in Computer Science from the University of Michigan in 1973 and 1976, respectively. From 1977 to 1981 he taught at the department of Industrial Engineering and Management Science at Northwestern University. He joined the Performance Analysis Department at Bell Laboratories in 1981, and later became an AT&T Fellow. Melamed moved to NEC in 1989 and served there as a Deputy Director, Head of the Performance Analysis Department, and NEC Fellow. He joined Rutgers University in 1996. Melamed was awarded an AT&T Bell Laboratories Fellow in 1988 and was elected IEEE Fellow in 1994.