

Influence of Available Bandwidth on the Statistical Characterization of Compressed Video

Paramvir Bahl

February 1996

Technical Report
TR-96-CSE-7

Department of Electrical and Computer Engineering
University of Massachusetts Amherst
Amherst, MA 01003, USA
bahl@acm.org

Influence of Available Bandwidth on the Statistical Characterization of Compressed Video

Paramvir Bahl

Department of Electrical and Computer Engineering
University of Massachusetts Amherst
Amherst, MA 01003, USA
bahl@acm.org

Abstract

The design and analysis of robust networking protocols that offer useful performance guarantees requires accurate traffic source models. In this paper we study the problem of characterizing and modeling the arrival process of compressed video. We extend earlier works in this area by including a factor that has previously been ignored, *the effect of video capture rate* on traffic characterization.

Dynamic changes in available bandwidth due to the addition and/or removal of connections can trigger re-negotiation of bandwidth between the applications and the network. Such re-negotiation may result in applications changing the capture rate of video sequences, thus effecting the traffic generation process. We show that for several popular video coding schemes, the bit rate distribution at the output of the encoder changes with the capture and compression rate. Using a combination of distributions, and exploiting knowledge of the underlying compression algorithms we characterize variable bit rate (VBR) video by application type, compression algorithm, and *frame rate*. We conclude that no single distribution can describe all video traffic, and as an alternate suggest a three dimensional matrix in which each dimension represents a different video classification aspect. Each entry in this matrix is a distribution type that best fits the given combination of the aspects. We use this result to show how the problem of network capacity planning may be tackled.

1. Introduction

An essential and necessary pre-requisite for the design of effective network protocols and their subsequent performance analysis is to have a clear and thorough understanding of the type of traffic that will be traversing the network channel. The type of traffic on the network is dependent on the applications running on the systems connected to the network. Traditionally applications that “speak” to one another using the underlying network have been limited to distributed file systems, remote logins, file transfers, electronic mail, and remote procedure calls. The modeling of such traffic as a Poisson arrival process is well accepted, well understood, and has been extensively used. A complete generation of network protocols have been designed, analyzed, simulated, and built using this model (for example variations of Aloha, variations of CSMA, and variations of token passing protocols). Applications that include digital video were not considered, as the magnitude of the associated data that had to be processed was prohibitive. With the emergence of faster processors, cheaper and greater storage, and improved data compression algorithms, the task that was previously daunting is now feasible. As a result a new class of applications that are time-critical, delay sensitive, bandwidth hungry, and which are dependent on getting sustained throughput from the network are being created. The data produced by such application is generally bursty with a packet inter-arrival time that is not exponentially distributed. Thus, Poisson source models can no longer adequately describe the traffic generation process for these new class of applications. Older protocols were never designed to support video data and understandably are unable to handle the demands put by it. New protocols that provide guaranteed bandwidth reservation through intelligent admission control, appropriate bandwidth allocation, and scheduling strategies are being designed to meet the requirements set forth by these new data types. These protocols and algorithms have to be analyzed under realistic conditions before any conclusions about their effectiveness can be made. Realistic simulation requires traffic models that can truly reflect the projected load and the demands on the network. Hence we need to understand, characterize and model these new data type and make the traffic generation process comparable to the “real world” situations.

Due to the enormous amount of data involved, when discussing packet video, video compression is almost always assumed [1]. Depending on whether or not the video encoder controls its output bit rate, it may be classified as either constant bit rate (CBR) or variable bit rate (VBR) encoder. While CBR video is attractive to network designers it is sometimes unacceptable to application vendors who place a premium on image quality. VBR encoders are better suited for such cases since they attempt to maintain promised image quality even at the expense of fluctuations in the output bit rate. Since ultimately it is the demand for quality that dictates the success of many applications, network support for VBR video becomes important.

Through proper characterization, VBR video can be supported by network protocols. Sensible parameters for admission control algorithms can be determined if the video arrival process is well understood. Quality of Service (QoS), a highly desirable feature in interactive, real-time audio-video applications, can be guaranteed. Understanding

characteristics of the traffic also helps in network capacity planning, in developing appropriate bandwidth allocation strategies and in determining buffering requirements at the nodes providing the connectivity — all of which are necessary for robust and timely video communications.

A plethora of scientific papers have used distribution-based modeling techniques to characterize video traffic, proposing a number of different distributions [2-9]. The lack of general consensus on the distribution that best describes VBR video can be attributed to the inherent content sensitivity of video compression algorithms which makes the process of characterization difficult. It is generally agreed that the problem needs to be constrained if an acceptable solution is to emerge. Previous works have thus classified video by application type and by compression schemes before applying distribution modeling. In this paper we propose and study the effect of a third additional factor that has been neglected in previous discussions, that is, the effect of available bandwidth. Dynamic change in bandwidth can occur with the addition and/or removal of connections. Under such circumstances, applications may decide to re-negotiate connection bandwidth, and subsequently vary their video capture and compression rate accordingly. We show that for inter-frame coding schemes, video-frame distribution changes with the capture and compression rate. From a pool of previously proposed asymptotic distributions, we examine the five most popular distributions and determine the ones that best describe inter-frame and intra-frame variable bit rate compressed video which has been classified by application type and frame rate. We show that no single distribution is best suited to describe all video traffic, and offer a three dimensional matrix, where each dimension represents a different video classification aspect and each entry of the matrix is the distribution type that fits the best to the given combination of the aspects. Thus, we hope to provide a rule-of-thumb for network designers who want to include models for generation of VBR video in their network simulations. As evidence of the usefulness of this matrix, we provide two important examples of using these distributions to solve real-world problems.

The paper is organized as follows: In Section 2 we formalize the three main difficulties associated with characterizing and modeling the video arrival process. Section 3 provides descriptions of the representative sequences used in our experiments and presents our approach to the problem. While the literature is replete with compression methods, we focus on those that are recognized as standards, a requirement for open and inter-operable systems. Our trace data consists of compressed video obtained from five different video codecs based on the different ISO and ITU-T video compression standards. In Section 4 we fit the observed video arrival data to the different asymptotic distributions. We employ various techniques including a segmentation approach and one that exploits knowledge of the underlying compression algorithm to determine the best fit. Additionally, we present a novel goodness-of-fit measure we used for arriving at our conclusions. In Section 5 we show how different capture rates effect the distributions of the arriving traffic. In Section 6 we present our conclusions in the form of a distribution matrix for single source distribution

modeling and give an example of how this matrix can be used for capacity planning of integrated networks.

2. Partitioning the Problem

In this section, we formalize the main difficulties encountered when characterizing VBR video. We propose the inclusion of video-frame capture rate, as a new classification when considering compressed video. We corroborate our observations with examples that clearly illustrate that characterization of VBR video can only be tackled if the problem domain is partitioned.

2.1 Characterization as a Function of Video / Application Type

Perhaps the most important reason why characterization of VBR sources is a hard problem is because, in general, video compression algorithms are sensitive to video content. Since there is considerable variation possible in video sequences (in terms of movement of objects, panning and zooming of camera, changes in scene etc.) there is an equal amount of variability associated with the output bit rate. Output from video encoders can vary from few Kilobits per second to several Megabits per second in a very short time quantum. Quantifying this variability is the hard part.

One way to break the problem into more manageable parts is to categorize video sequences in terms of content or application. Figure 1 illustrates one way of dividing video according to the application type. Examples of Request Video include video-on-demand type applications such as entertainment videos, news clips, education programs, weather information programs, tele-shopping programs, catalog videos etc. The content of such videos usually has frequent scene changes, rapid movement of objects and panning and zooming of camera. These sequences are both hard to compress and hard to characterize. Conference Video, on the other hand is easier to compress as the correlation (both spatial and temporal) between pixels is generally high. Figure 2 illustrates how a typical density function for Request Video tends to be different from that for Conference Video¹.

¹ Another name for Conference Video application, given by ITU-T, is "Real-Time Conversational Services"

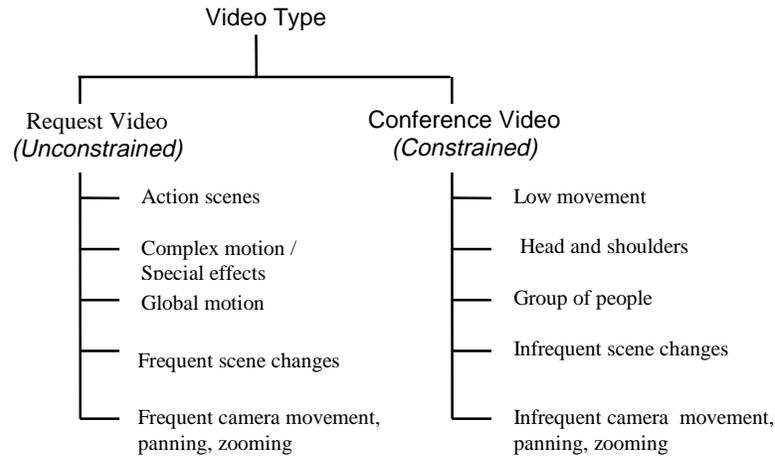


Figure 1: Properties of the classified video types

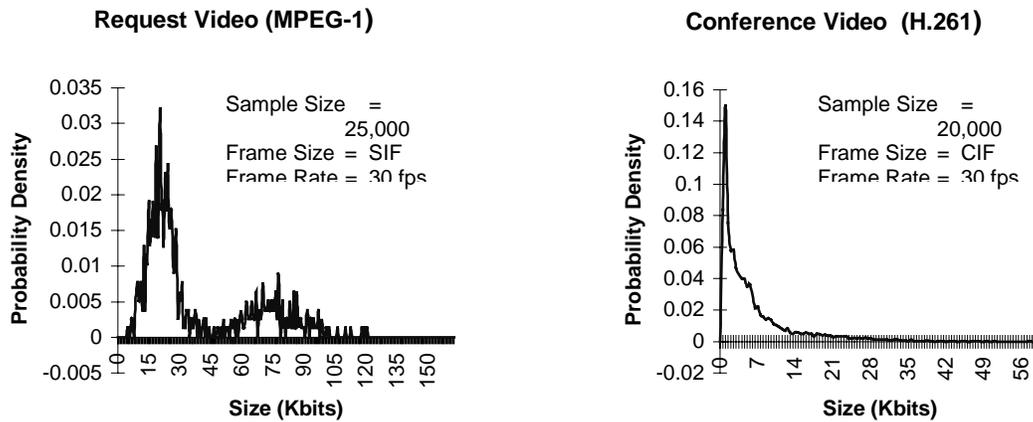


Figure 2: Characterization as a function of application type

2.2 Characterization as a Function of Compression Algorithm

Since intra-frame compression schemes only exploit spatial redundancies, the distribution of the number of bits per frame at the output of the video encoder remains relatively constant. On the other hand inter-frame compression exploits both spatial and temporal redundancies and the number of bits per frame at the output of the encoder varies considerably.

Figure 3 illustrates the differences in the density function of a video sequence that was coded using two different compression algorithms: JPEG² [10], an intra-frame scheme, and MPEG-1³ [11] an inter-frame scheme. The video sequence was part of an entertainment program broadcast over a local television channel. The first 10,000 frames were used for determining the density function. Several other sequences were also digitized, compressed and their PDFs plotted. Figure 3 was chosen as the representative sample, its shape is typical of what was seen for the other similarly coded sequences.

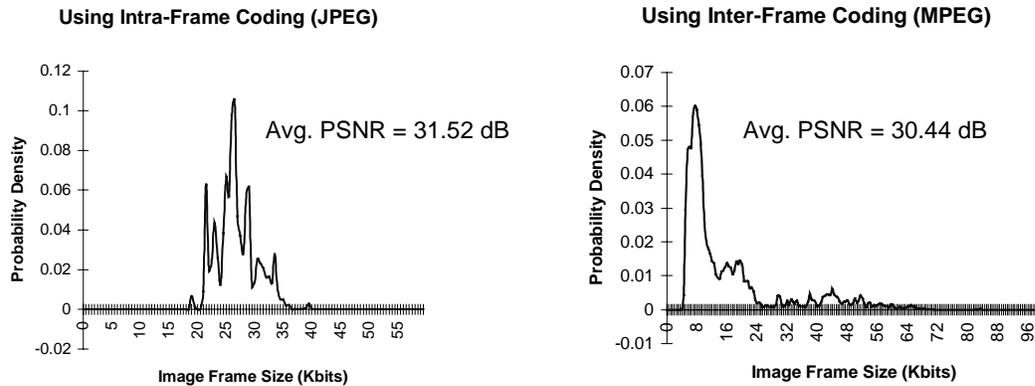


Figure 3: Characterization as a function of the compression scheme

2.3 Characterization as a Function of Available Bandwidth

A third important consideration that has so far been neglected by researchers is the rate at which the video sequence is captured and compressed before transmission. For inter-frame compression algorithms, the capture and compression rate has a direct bearing on the characterization process.

One might ask the question, why care about different capture rates? The answer is this: if the available bandwidth for transmission is not enough to accommodate transmitting a full 30 frames/seconds, the receiving application may choose to accept 15 frames/second or even fewer. So receiving less than 30 frames/second is always an option available to the application if it is willing to accept degraded real-time performance when network bandwidth is an issue.⁴ Looking at the bits coming out of the compressor the statistics and distribution are very different for different frame rates. A model that is suitable for 15 frames/second may not be suitable for 5 frames/second. Figure 4 depicts the density when video is captured and compressed at approximately 15 frames per second along with the

² Joint Photographic Picture Group, ISO/ITU standard for still compression. Sometimes used as a video compression method (M-JPEG)

³ Moving Pictures Expert Group, ISO/IEC 11172-2 Standard for coded representation of digital video

⁴ Some CBR schemes including the H-series recommendation from ITU-T (on audio-video communications) trade-off frame rate for a constant output bit rate.

density when the same video sequence is captured and compressed at 3 frames/second. Table 1 provides details of the two experiments.

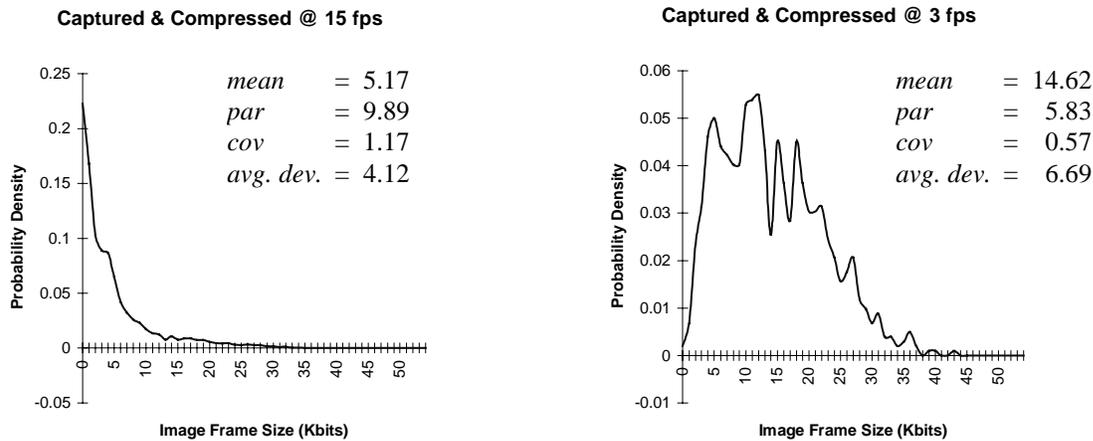


Figure 4: Characterization as a function of the capture rate

It is not difficult to conclude that the shape of the densities are different for different capture rates and hence they should be modeled differently. An interesting observation is that as the capture and compression rate is decreased, the density function for inter-frame compressed sequences approaches the density function for intra-frame compressed sequences. Intuitively this result make sense since as the capture rate decreases, the differences in sequential frames become large and frames can no longer be coded (efficiently) with respect to each other. Hence inter-frame coding degenerates to intra-frame coding and this is reflected in the resulting density functions.

Capture Rate	Compression	Resolution	No. of Frames	Length
14.8 fps	H.263	352 x 288 (CIF)	5790	6 min. 30 sec.
2.6 fps	H.263	352 x 266 (CIF)	1020	6 min. 30 sec.

Table 1: Particulars of the compressed video sequence of Figure 4

3. Data Characterization

In this section we provide details, both statistical and descriptive, of the video sequences and the video codecs we chose for characterizing VBR video traffic.

3.1 Approach: The Seven Step Process

The general process of characterization, modeling and subsequent analysis (mathematical or through simulation) is a seven step process as illustrated in Figure 5. Decisions made in each step influence the conclusions reached at the end of the process. As

is illustrated in the figure below, there are several different techniques that can be employed for modeling digital video. Details are provided in [4, 5, 7, 12-17]

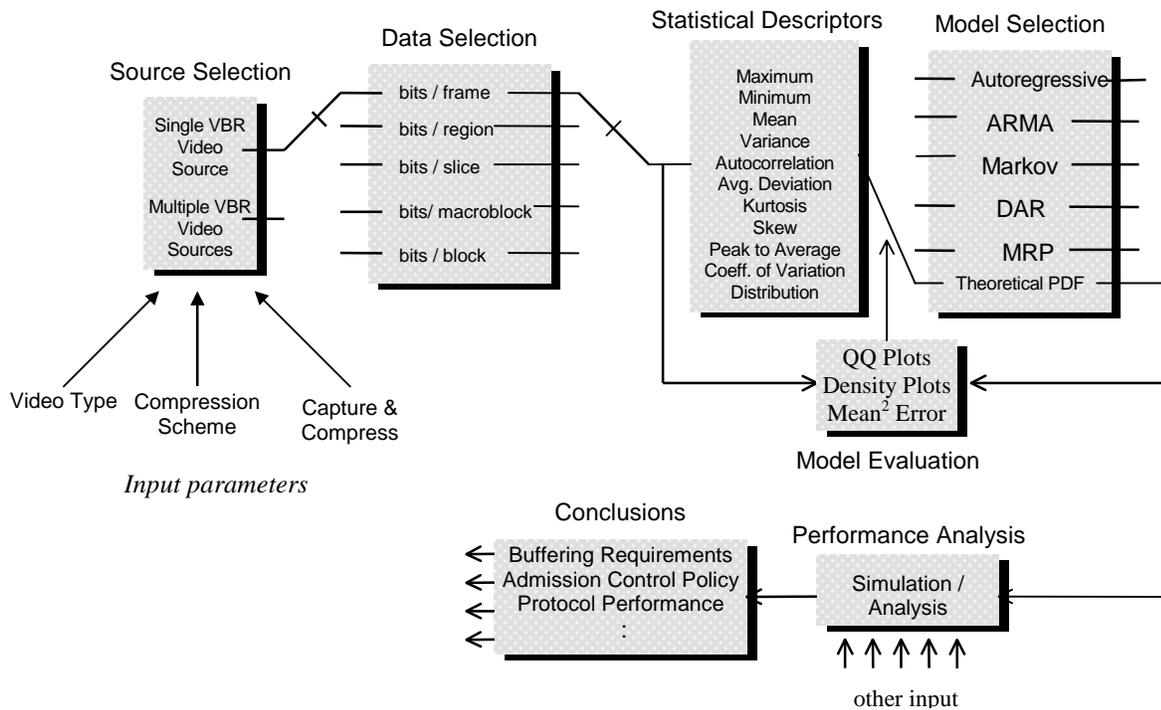


Figure 5: Steps in characterization, modeling and analysis

3.2 Representative Data Set

For this study we chose a total of twenty five video sequences which were compressed using five different ITU-T and ISO standards compliant video codecs. The representative sequences cover the two classes of applications: Conference Video and Request Video. The five compressors represent the first generation of video codecs as each employs statistical redundancy reduction techniques to achieve compression. These codecs can be classified into the two classes of coding methods: intra-frame and inter-frame. To study the effect of different capture rates, we modified the compressors to skip frames before compression. The videos ranged from thirty seconds to thirty minutes and were compressed to ratios that ranged from 196:1 (or 0.06 bits per pixel) to 23:1 (or 0.51 bits per pixel). To generate sequences for Conferencing Video we simulated a suitable conference environment and captured and compressed the video using a hardware video capture board [18] and a software video codec [19]. The software-only video codecs were based on ITU's video compression standards --- H.261⁵ [20] and H.263⁶. To generate Request Video we captured and compressed video using codecs based on ISO's MPEG-1 and MPEG-2 (or H.262) [21] video compression standards. We also included sequences that were

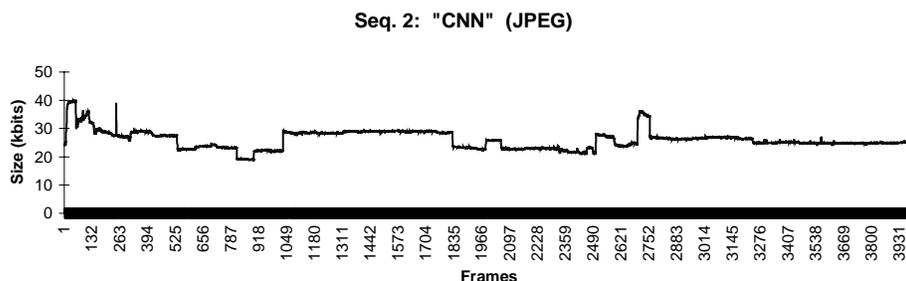
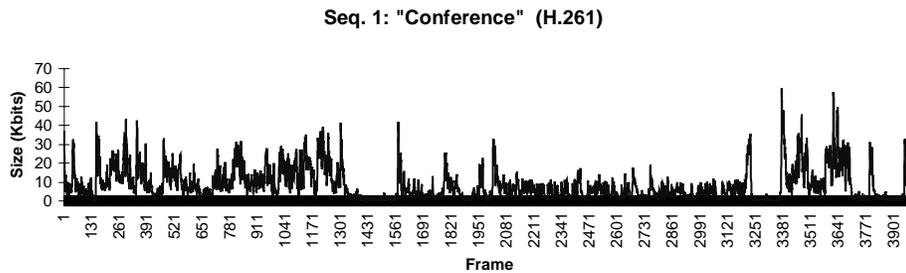
⁵ ITU-T Recommendation for Audio Visual Services over Narrowband ISDN channels, (P x 64 Kbit/sec video codec)

⁶ ITU-T Rec. for real time video over V.34 modems on the GSTN telephone networks (< 64 Kbit/sec video codec)

compressed using JPEG, a ISO standard for still image compression. (Due to the early availability and popularity of the JPEG compression standard [18], JPEG compressed images and video co-exist with MPEG and H series compressed sequences on the network). While H.261, H.263, MPEG-1 and MPEG-2 are all inter-frame compression schemes, JPEG is an intra-frame compression scheme. The five codecs were chosen as they represent the current state of art in commercially available compression technology. A description of the various building blocks behind compression algorithms and the different compression standards is provided in [19]. Table 2 provides a summary description of five representative video sequences and Figure 6 illustrates the temporal characteristics of these videos taken from the set of twenty five.

Sequence Id.	Length (min.)	No. of Frames	Frame Resolution	Capture & Compress Rate	Compression Algorithm	Encoder's Output (bits / pixel)
Seq. 1	28	12470	CIF ⁷	7.42 fps	H.261	0.06
Seq. 2	30	53946	SIF ⁸	29.97 fps	(M)JPEG	0.36
Seq. 3	30	53946	SIF	29.97 fps	MPEG-1	0.23
Seq. 4	3	5844	SIF	29.97 fps	MPEG-1	0.17
Seq. 5	0.54	781	SIF	24 fps	MPEG-1	0.51

Table 2: Video sequences used in traffic characterization



⁷ Common Interchange Format: 352 pixels x 288 lines

⁸ Source Input Format: 352 pixels x 240 lines

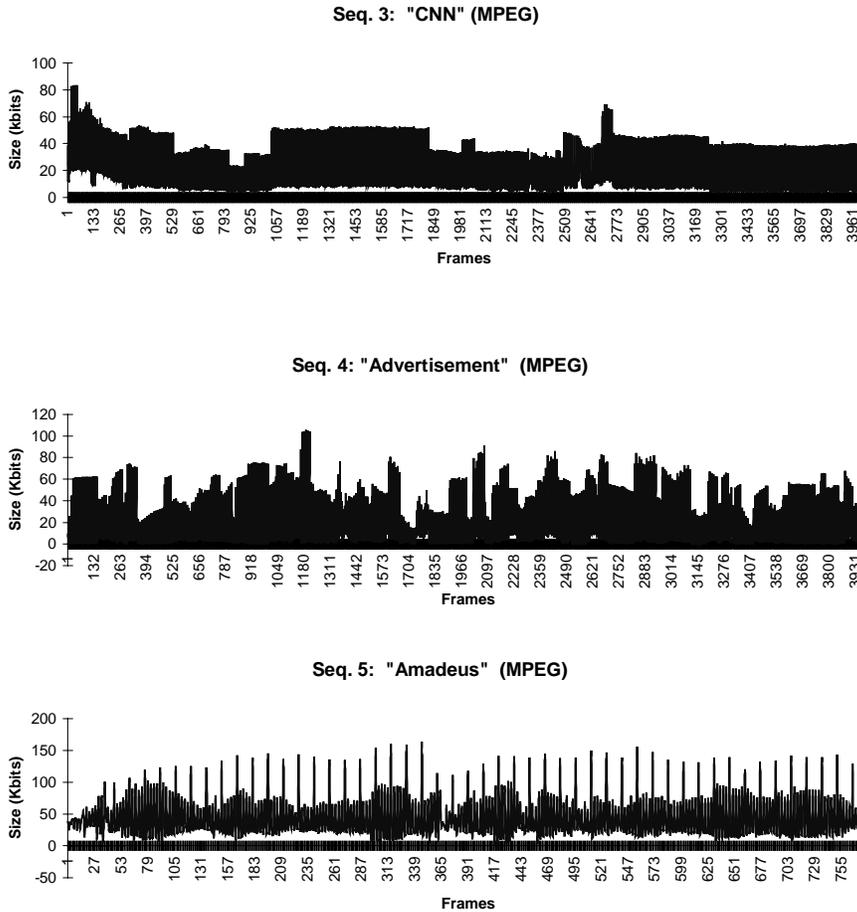


Figure 6: Temporal behavior of the video sequences

Table 3 contains the statistical descriptors for the sequences shown in Figure 6

	Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5
Sample Size	12470	53946	53946	5845	781
Mean	6.06	26.65	17.11	12.52	42.75
Variance	48.37	24.66	209.48	353.37	1209.31
Skew	2.16	1.69	1.74	2.36	1.36
Avg. Deviation	4.94	3.43	10.81	13.05	28.82
Maximum	59.44	55.20	97.84	123.02	162.44
Minimum	0.43	14.17	0.22	0.22	4.58
Median	3.58	26.23	10.46	4.27	25.09
⇓					
Std. Deviation	6.96	4.97	14.47	18.80	34.78
Coefficient of Variation	1.15	0.19	0.85	1.50	0.81
Peak-To-Avg. Ratio	9.80	2.07	5.72	9.83	3.8
Range	59.01	41.03	97.62	122.80	157.86

Table 3: Descriptive statistics of sequences considered

The probability density function for a given video sequence is derived from the numbers of bits per frame observed at the output of the VBR video encoder. (An equivalent measure is to look at the number of packets or cells generated per frame at the output of the transmitter. For fixed packet size, this is a scaled version of the bits per frame). Figure 7 illustrates the cumulative distribution function for the sequences of Figure 6 and Table 3.

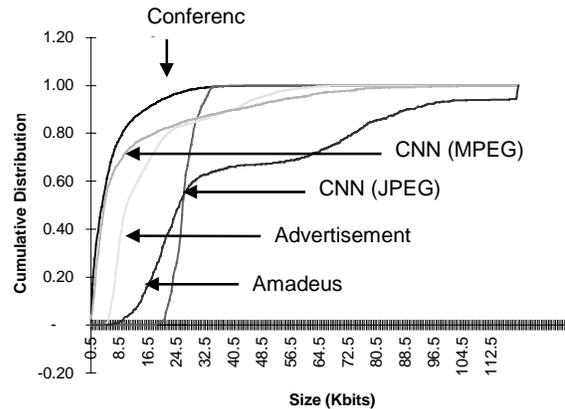


Figure 7: Cumulative distribution for the video sequences

4. Distribution Based Modeling

In this section we describe the observed video arrival data in terms of mathematical functions. We restrict our attention to video encoded at full frame rates, leaving the modeling of variable frame rate for the subsequent section. We introduce a novel goodness-of-fit measure, based on average mean-square error, that we used to arrive at our conclusions.

4.1 Parameter Estimation

After deriving the density function for the compressed video sequences, we visually examined their shapes and estimated these with known mathematical functions with similar shapes. Fitting the hypothesized probability density function $f(x)$ to the observed distribution requires the estimation of the distribution parameters. We considered both Point Estimation and Interval Estimation. In Point Estimation, the values of the parameters are derived in terms of the observed data. The values are good in terms of unbiasedness, minimum variance, etc. as defined by the estimation criterion. In Interval Estimation bounds on the parameter values are obtained which give information on the numerical value of the parameter and provide an indication on the level of confidence one can place on the estimated value derived from the observed sample [22]. The attractiveness of the Point Estimation approach comes from its simplicity. We thus present results obtained when Point Estimation was used to determine the values of the desired distribution parameters.

Within Point Estimation we used both (a) the method of Maximum Likelihood Estimation (MLE) [23], and (b) the Method of Moments [24] to estimate the values of the distribution parameters. In Maximum Likelihood Estimation, if $f_x(x; \theta_1, \theta_2, \dots, \theta_m)$ is the hypothesized density function for sample values x_1, x_2, \dots, x_n with parameters $\Theta = [\theta_1, \theta_2, \dots, \theta_m]$ that are to be estimated, then the estimation procedure for Θ consists of choosing the particular value of Θ that maximizes the Likelihood function L:

$$L(x_1, x_2, \dots, x_n; \Theta) = f(x_1; \Theta) f(x_2; \Theta) \dots f(x_n; \Theta) \quad (1)$$

In the Method of Moments, the theoretical moments (α_i 's), which are generally a function of the distribution parameters, are calculated as:

$$\alpha_i(\theta_1, \theta_2, \dots, \theta_m) = \int x^i f(x; \theta_1, \theta_2, \dots, \theta_m) dx \quad i = 1, 2, \dots \quad (2)$$

and are then equated to the sample moments obtained from the observed data.

$$M_i = \frac{1}{n} \sum_{j=1}^n x_j^i \quad i = 1, 2, \dots \quad (3)$$

By establishing and solving as many equations as there are number of parameters to be estimated, estimators for the distribution parameters are obtained.

4.2 Estimation without Prior Information

Compressed video tends to exhibit asymptotic behavior with a heavy right tail. With this observation we considered five distributions that exhibited asymptotic behavior as $x \rightarrow \infty$. The objective was to determine how well the distributions fit our trace data and whether they can be used to describe the observed video arrival process.

4.2.1 Normal Distribution

On examining the distribution of several intra-frame compressed video sequences, we concluded that, in general, the density function for high capture rate video with frequent scene changes, has a shape that is similar to a bell shaped curve typical of the familiar Normal distribution. For Normal distribution, the maximum likelihood estimators coincide with the moment estimators and the mean and variance are estimated from the sample mean and sample variance of the observed data. Figure 8 illustrates how the Normal distribution fits the observed data for two intra-frame coded sequences. Both sequences are from a Request-Video type application. The first is a 30 minute video of a Cable News Network program and the second is a 15 minute compressed video segment from a home video of a birthday party.

Besides the Normal distribution, the Lognormal, Gamma and Weibull distributions can also form bell shaped curves. We superimpose these distributions along with the Normal distribution on the observed PDF in Figure 8. The parameters for all but the Weibull distribution are obtained using the Method of Moments and are given in Table 4.

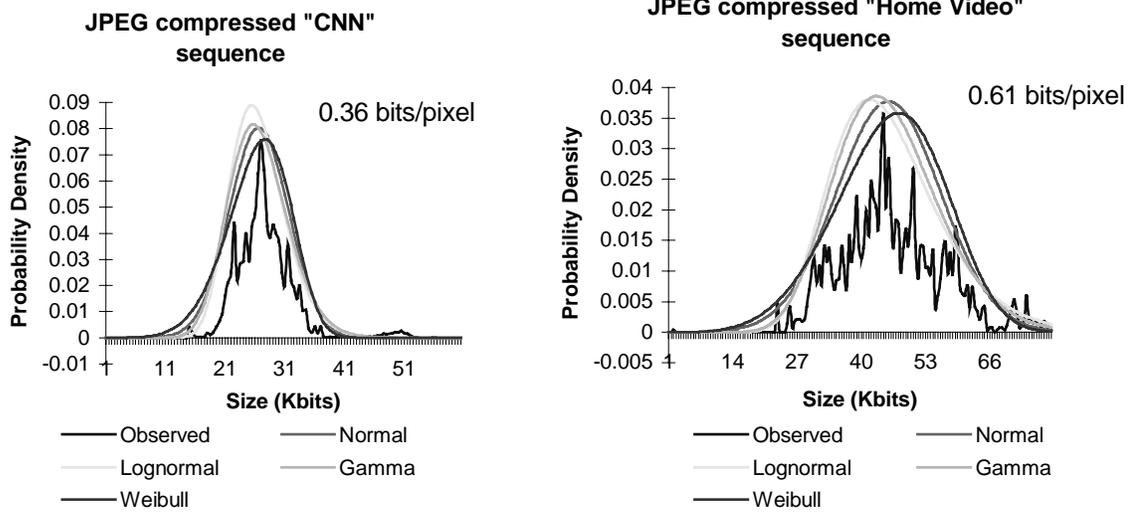


Figure 8: Estimating the distribution for a Intra-frame compressed sequence.

Looking at Figure 8 we may conclude that all four distributions seem to match the observed PDF fairly well and that each could represent the video arrival process adequately. However, in general, the Normal distribution seems to provide the best estimate for intra-frame compressed Request Video sequences. This coupled with the fact that the MLE and Moment Estimators for the distribution are identical, and the determination of these parameters is trivial, make it the distribution of choice for modeling intra-frame coded Request Video. We examine this conclusion more rigorously in a subsequent sub-section where we verify the best distribution by using our “goodness-of-fit” tests.

Sequence Id.	Gamma		Lognormal		Pareto		Weibull	
	$\alpha = \frac{\eta_x^2}{\sigma_x^2}$	$\lambda = \frac{\eta_x}{\sigma_x^2}$	$\eta_{\ln x}$	$\sigma_{\ln x}^2$	λ_{MLE}	λ_{ME}	α	λ
Seq. 1	0.759	0.125	1.233	1.195	0.811	1.197		
Seq. 2	28.795	1.081	3.267	0.030	0.306	1.038	5.802	0.035
Seq. 3	1.397	0.082	2.562	0.527	0.390	1.062	1.502	0.040
Seq. 4	0.443	0.035	1.667	1.682	0.599	1.087		
Seq. 5	1.511	0.035	3.467	0.553	0.288	1.024		

Table 4: Estimated model parameters

4.2.2 Gamma Distribution

The gamma distribution is one sided and extremely versatile in that a wide variety of shapes are possible by varying its two parameters: α and λ . α determines the shape of the distribution whereas λ is the scale parameter. The mathematical description of the Gamma distribution is given by:

$$f(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x = 1, 2, 3, \dots, \infty \quad (4)$$

where,
$$\Gamma(z) = \sum_{i=0}^{\infty} i^{z-1} e^{-i}, \quad z = 1, 2, 3, \dots \quad (5)$$

and
$$\eta_x = \frac{\alpha}{\lambda}, \quad \sigma_x^2 = \frac{\alpha}{\lambda^2} \quad (6)$$

In general $f(x)$ is unimodal with its peak at $x = 0$ for $n \leq 1$ and at $x = (n - 1)/\lambda$ for $n > 1$. The corresponding cumulative distribution function for the gamma distribution can be derived and is given by:

$$F_x(x) = \frac{\Gamma(\alpha, \lambda x)}{\Gamma(\alpha)}, \quad x \geq 0 \quad (7)$$

where,
$$\Gamma(\alpha, \lambda x) = \sum_{i=0}^{\lambda x} i^{z-1} e^{-i}, \quad z = 1, 2, 3, \dots \quad (8)$$

The versatility of the Gamma distribution coupled with the simplicity of the expressions for the mean and variance makes it an attractive candidate for fitting it to the observed PDF derived from the compressed video sequences. Table 4 contains the estimated parameters obtained with the Method of Moments for the five video sequences shown in Figure 6.

4.2.3 Lognormal Distribution

The tail in the Lognormal distribution decreases slower than any exponential function. The density function for the Lognormal distribution is described as:

$$f(x) = \frac{1}{\sigma_{\ln x} x \sqrt{2\pi}} e^{-\frac{(\ln x - \eta_{\ln x})^2}{2\sigma_{\ln x}^2}}, \quad \eta_{\ln x}, \sigma_{\ln x}^2, x > 0 \quad (9)$$

here $\eta_{\ln x}$ is the mean of $\ln(x)$ and $\sigma_{\ln x}$ is the standard deviation of $\ln(x)$. The first two moments of the distribution are $\eta_x = e^{\eta_{\ln x} + \frac{\sigma_{\ln x}^2}{2}}$, and $\sigma_x^2 = e^{2\eta_{\ln x} + \sigma_{\ln x}^2} (e^{\sigma_{\ln x}^2} - 1)$.

Figure 9 illustrates how the Gamma and Lognormal distributions fit the observed data for two inter-frame compressed sequences obtained from a Request Video application. The values for the distribution parameters was obtained using the Method of Moments. Looking at the two samples shown in Figure 9, one can conclude that the Lognormal distribution provides a better fit than the Gamma distribution.

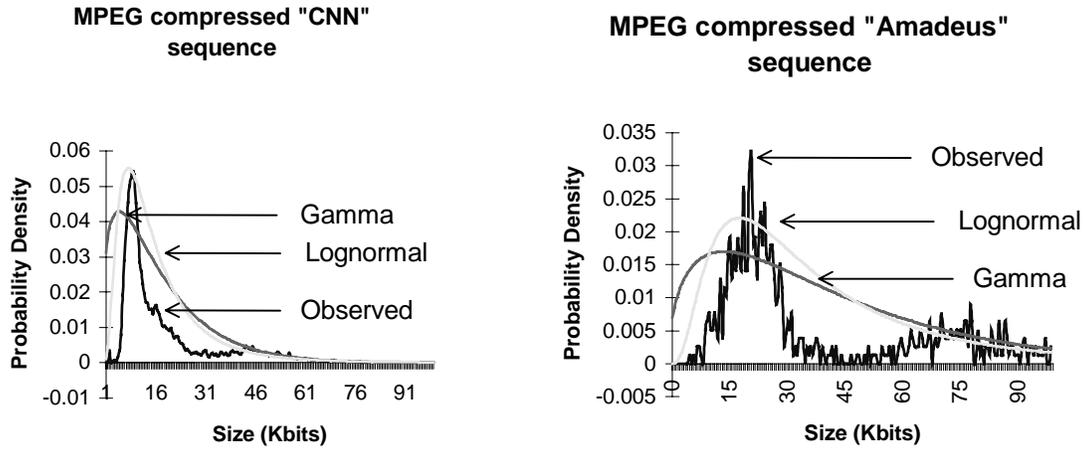


Figure 9: Estimating the distribution for Inter-frame compressed video

Seq. 5 (*Amadeus*) is not well behaved in that there is no smooth (asymptotic) decay. The second “bump” in the sequence may be attributed to the rapid scene changes in the sequence. Rapid scene changes result in increased number of bits needed for encoding the frame (differences between frames becomes large). Even though the I-B-P pattern for the MPEG encoder was fixed, P (Predicted) frames require more number of bits to encode and this is reflected by the second bump in the observed density function. This second bump has an averaging effect on both the Gamma and Lognormal distribution and to compensate for it, both under shoot and over-shoot the observed distribution at the different points. This suggests that perhaps there is a better way to estimate distributions with high action.

4.2.4 Pareto Distribution

The Pareto distribution has also been widely used for fitting observed data in a variety of different applications [25, 26]. Most recently it has been used to model segmented TELNET inter-packet arrival times and the sizes of FTP data bursts [27].

The generalized form of the Pareto distribution with shape parameters λ and k , and location parameter α has a density function given by [25]:

$$f(x) = \frac{\Gamma(\alpha + k)\lambda^\alpha x^{k-1}}{\Gamma(\alpha)\Gamma(k)(\lambda + x)^{k+\alpha}}, \quad x, \alpha, \lambda, k > 0 \quad (10)$$

A more restrictive but easier to compute form is given by letting $k = 1$, in which case the above equation reduces to the *Classical Pareto Density* given by [26]:

$$f(x) = \frac{1}{\lambda} \alpha^\lambda x^{-\frac{(\lambda+1)}{\lambda}}, \quad x > 0 \quad (11)$$

The corresponding cumulative distribution function is:

$$F(x) = P[X \leq x] = 1 - \left(\frac{\alpha}{x}\right)^{\frac{1}{\lambda}}, \quad \alpha, \lambda \geq 0, \quad x \geq \alpha \quad (12)$$

where, $\eta_x = \frac{\alpha^{\frac{1}{\lambda}}}{1-\lambda}$, $\lambda < 1$, and $\sigma_x^2 = \frac{\lambda^2 \alpha^{\frac{2}{\lambda}}}{(1-\lambda)^2(1-2\lambda)}$, $\lambda < \frac{1}{2}$

Thus when $\lambda \geq \frac{1}{2}$, the distribution has infinite variance and when $\lambda \geq 1$ the distribution has infinite mean.

On examining the density functions of inter-frame coded Conference Video sequences, we observe two features: (a) there is a minimum amount of bits for each frame and (2) the overall density function is heavily tailed towards the right. This second observation may be attributed to the fact that the number of non-key frames (bi-directional and predicted) are greater than the number of key (intra) frames. In fact it would be safe to conclude that as the compression ratio is increased (bits per pixel decreased) the distribution exhibits even heavier tails. This is specially true for cases where the camera movement, subject movement and scene changes are limited as is the case in Conference Video applications. The Pareto distribution is not suited for intra-frame coded video and for high action, inter-frame coded Request Video because of its peaked-ness.

4.2.5 Weibull Distribution

Another asymptotic distribution that can take on a wide variety of shapes and hence is a viable candidate for modeling distribution of bits per frame for compressed video is the Weibull distribution [28]. Mathematically the distribution is expressed as

$$f(x) = \frac{\alpha (\lambda x)^{\alpha}}{x} e^{-(\lambda x)^{\alpha}}, \quad \alpha, \lambda > 0, \quad x \geq 0 \quad (13)$$

where λ is the scale parameter and α is the shape parameter. The mean and variance of the Weibull distribution can be derived:

$$\eta_x = \frac{1}{\lambda} \Gamma\left(1 + \frac{1}{\alpha}\right), \text{ and } \sigma_x^2 = \frac{1}{\lambda^2} \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right]$$

It is non-trivial to obtain λ and α , the distribution parameters, in terms of the sample moments of the sequence being modeled. However if we take into account the effect these parameters have on the shape of the distribution, appropriate values for the parameters can be estimated so that the resulting distribution matches the observed distribution. For example, when $\alpha = 1$, Weibull distribution degenerates to the exponential distribution, when $\alpha < 1$ the distribution has a L shape and when $\alpha > 1$ it is bell shaped. Also from [31] we know that when $\alpha < 3.602$ the distribution has a heavy right tail, when $\alpha > 3.602$ it has heavy left tail and when $\alpha = 3.602$ the shape is close to Normal distribution. Depending on the shape of the observed distribution we can estimate α . To

determine λ , the scaling factor, we make use of the following observation: λ is inversely proportional to σ_x^2 , so when the standard deviation of the data being modeled is large (as in Conference Video) we choose a small value of λ . For bell shaped curves ($\alpha > 1$) choosing a smaller λ (< 0.04) has the effect of expanding (and slightly left shifting) the main lobe of the distribution curve while higher values of λ (> 1) make the main lobe tighter. For L shaped curved ($\alpha < 1$) changing λ has a less dramatic effect as the asymptotic behavior dominates.

An alternate method for getting the parameters of the Weibull distribution is to detect a straight line on a probability plot of $y = \ln(x)$ versus $z = \ln(-\ln(1 - F(x)))$ [29]. From a straight line fit $y = mz + c$, one obtains $\alpha = 1/m$, and $\lambda = 1/e^{-c}$.

With these observations as guiding principles we estimated the parameters for the video sequences and superimposed the resulting distribution over the distribution derived from the trace data.

4.3 Fitting a Combination of Distributions (Segmented Approach)

Modeling inter-frame video, which has rapid scene changes such as the Amadeus sequence, is possible if the density function is viewed as two separate segments. For example looking at the PDF for Amadeus a better estimate can be obtained when the shape of the density function between 0 and 50 Kbits/sec (segment 1) is modeled separately from that between 50 Kbits/sec and 120 Kbits/sec (segment 2). Modeling segment 1 and segment 2 as two concatenated Normal distributions with different means and variances (i.e. $N(\eta_{xl}, \sigma_{xl}^2) + N(\eta_{xh}, \sigma_{xh}^2)$) leads to Figure 10.

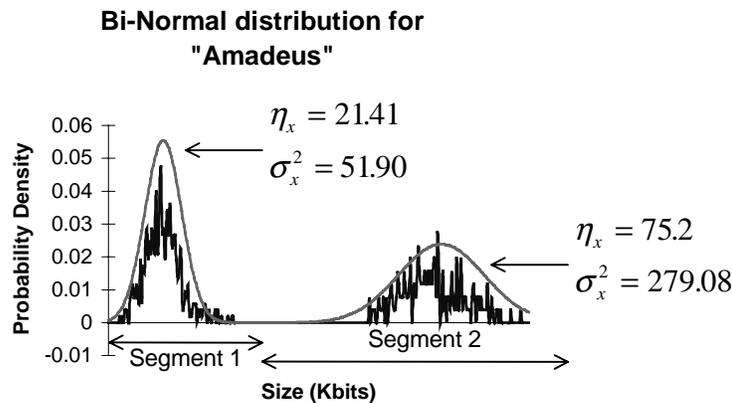


Figure 10: Rapid scene changes -- modeled as concatenated Normal Distributions

The mean and variance of the segment 1 is obtained from the sample mean and sample variance of the B frames and for Segment 2 from the sample mean and sample variance of the I + P frames.

Methods similar to the one just described have been proposed elsewhere as well. In [3] the authors propose a combination of Gamma/Exponential for estimating Conference Video sequences. In [2] the conclusion is that a combination of Gamma/Pareto fits the left and right tails of the observed distributions best, and in [30] the authors suggest a combination of three Gaussians. In our studies, where the trace data came from “real-world” applications that employed ISO and ITU-T standard video codecs, we found that a combination of two Normal distributions sufficiently estimates inter-frame compressed, high capture rate, high action, Request Video.

4.4 Estimation with knowledge of the Compression Algorithm

A different approach towards distribution modeling of VBR video is one that uses the knowledge of the underlying compression algorithm. In MPEG compression (both MPEG-1 and MPEG-2), video sequences are made up of I, P or B frames (a fourth type of frame, called D-frame or the DC-intracoded frame, is also defined in the MPEG specification, but it is hardly ever used.). The application and the encoder prior to compression generally determine the ratio and frequency of each type of frame. I frames are primarily provided to improve random access capability and when VCR type functionality (fast forward, fast reverse, step forward, set reverse etc.) are important [19]. B frames provide the bulk of the compression and P frames are used when B frames cannot be coded accurately. In terms of bits per frame, I frames produce the maximum number of bits per frame, and B frames produces the least number of bits per frame. With this knowledge we look at the distribution of bits per frame for each type of frame in MPEG compressed sequences from Request-Video applications. Figure 11 illustrates the density function for the different types of frames in the MPEG compressed “CNN” and “Advertisement” video sequences.

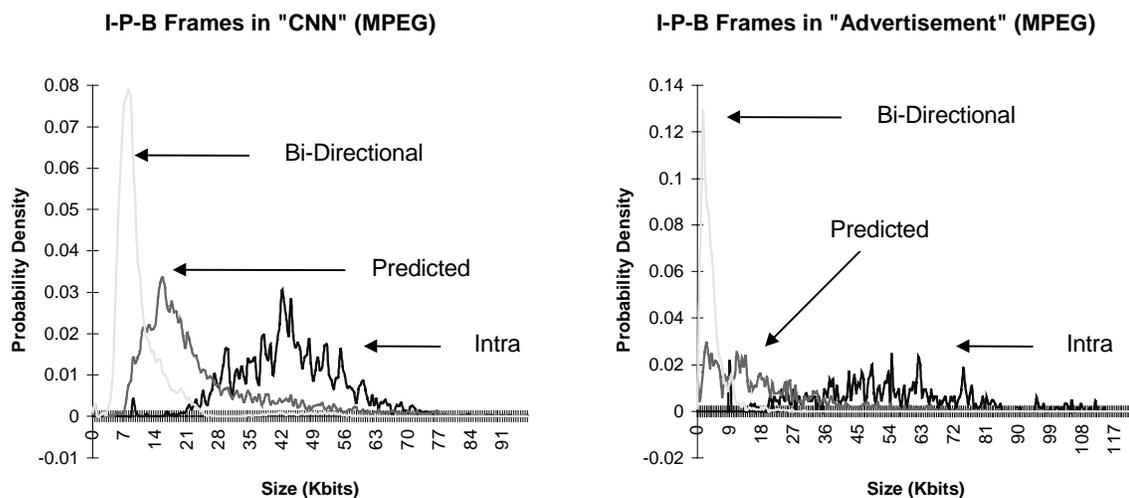


Figure 11: Dissecting MPEG compressed VBR video

Looking at each of these distributions separately we proceed to estimate the I frame distribution with the Normal distribution. Figure 12 shows a normal distribution superimposed over the observed distribution for the two sequences.

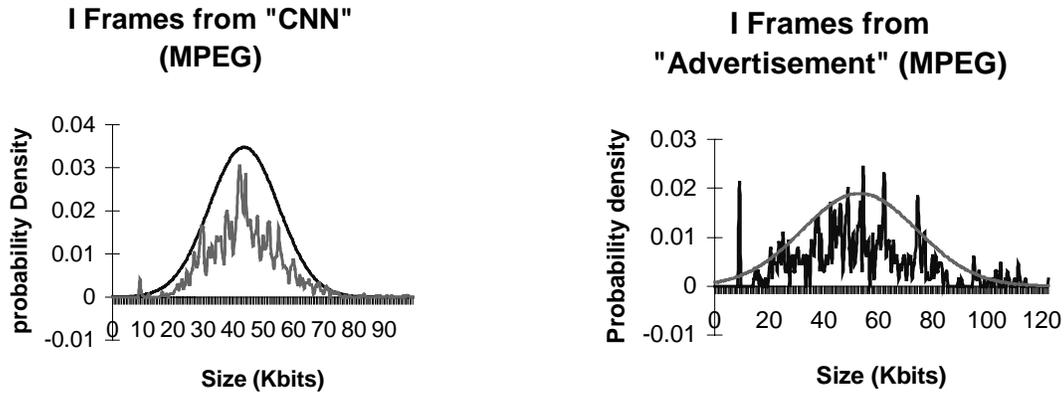


Figure 12: Normal distribution and I Frame distribution in MPEG coding

For the case of P and B frames we superimpose the Lognormal and Gamma distributions over the observed density functions. This is illustrated in Figure 13 and Figure 14.

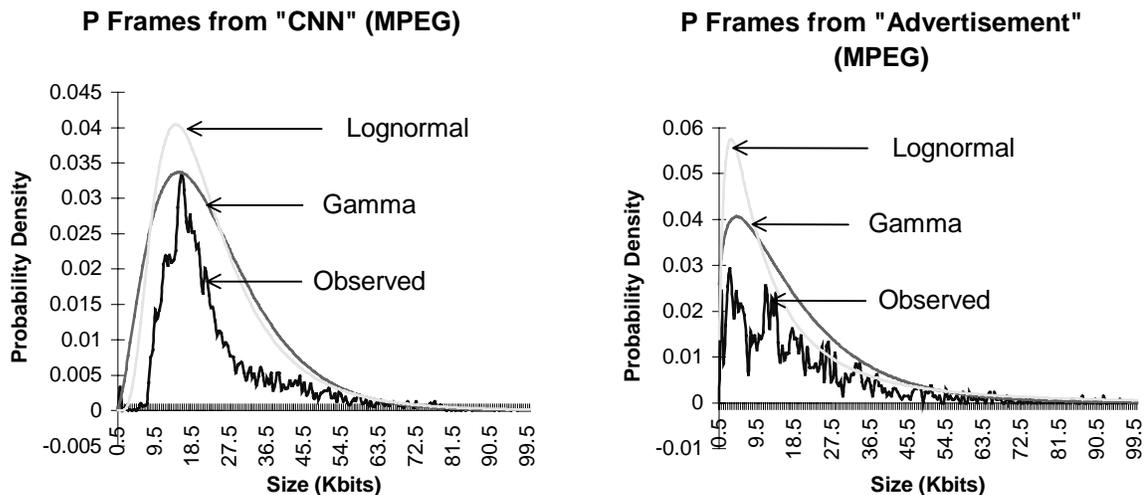


Figure 13: P Frame distribution for MPEG sequences

Looking at Figure 13 the Gamma distribution tends to do better in approximating the P frames distribution while from Figure 14 we conclude that the Lognormal distribution models B frame distribution better. Thus knowing the sequence of I-P-B frames and their frequency it is possible to generate, using a combination of Normal, Gamma and Lognormal

random variates, the representative distribution of the bits per frame for the entire video sequence (see Figure 15)

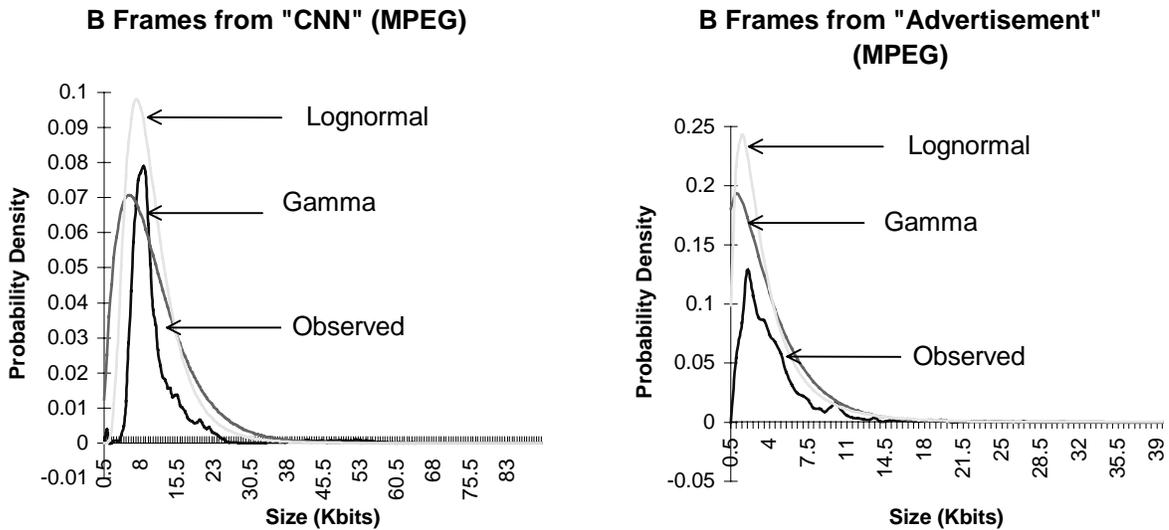


Figure 14: B Frame distribution for MPEG sequences

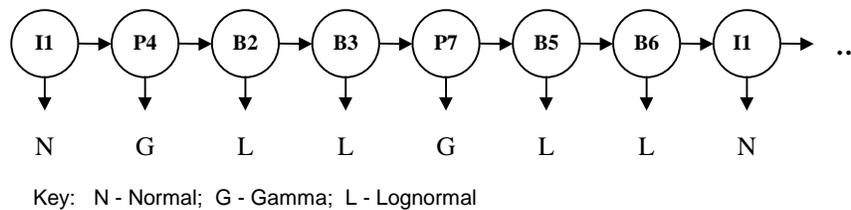


Figure 15: Random variates for a typical MPEG sequence

4.5 Metrics for Evaluating and Verifying Distribution Models

In addition to forming the histogram and frequency diagrams from the observed bits per frame and visually comparing them to the estimated distribution, we employed two other methods to reach our conclusions. The first of these is the well established method of comparing the quantiles of the observed distribution with that of the hypothesized distribution. Figure 16 is an example of such plots (the y_i -value for the q_i -th quantile is obtained as $y_i = F^{-1}(q_i)$). These particular plots are for the intra-frame compressed CNN sequence.

For Normal distribution we used

$$x_i = \eta_x + \left(4.91 \left[q_i^{0.14} - (1 - q_i)^{0.14} \right] \times \sigma_x \right) \quad (14)$$

to compute and plot the normal quantiles on the x axis. For Weibull distribution we inverted the CDF ($F_x(x) = 1 - e^{-(\lambda x)^\alpha}$) and obtained the Weibull quantiles as:

$$x_i = \frac{(-\ln(1 - q_i))^{1/\alpha}}{\lambda} \quad (15)$$

For Gamma and Lognormal distributions (not shown in this figure) we pre-generated CDF tables with a step factor of 50 and used these to determine the x_i values.

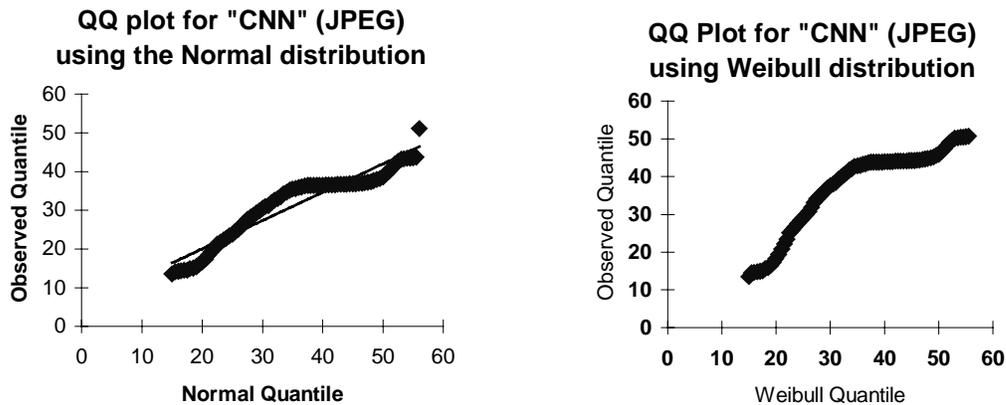


Figure 16: QQ Plots for Normal and Weibull Distributions

The second method we used for evaluating the best-fit for the estimated distribution was a more objective method of looking at the mean squared error between the observed distribution and the hypothesized distribution over a specified time quantum. The modeling error, defined as the difference between the real data and the generated data, over time was computed. Figure 17 provides an example of this method. The mean squared error for the sequence being modeled was calculated and used to determine the best fit. The estimated bits per frame were obtained from random variates. Random variates for the Pareto and Weibull distributions were generated using the inverse-transform technique. Normal variates were generated using the Box-Muller method, Lognormal variates were generated from normal variates, and Gamma variates were generated as in [31]. This test is preferred over the QQ plots since it closely reflects how the model would behave in a real simulation.

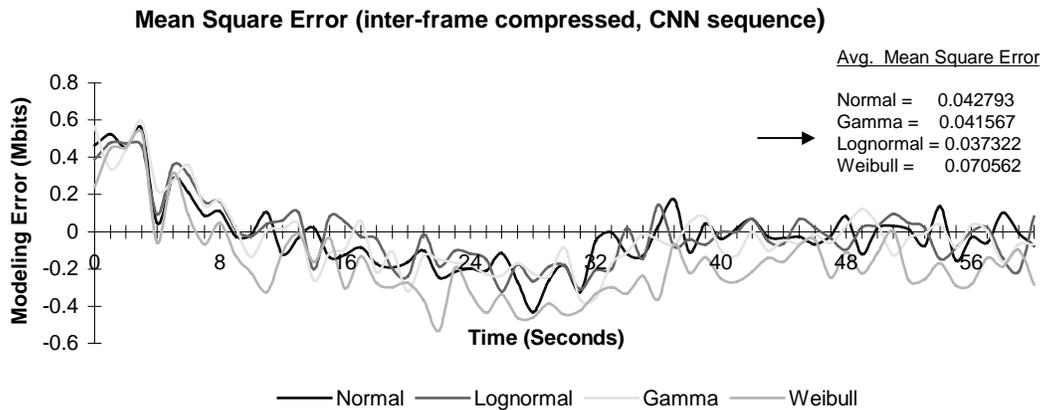


Figure 17: Modeling Error with VBR video distribution models

5 Effect of Capture Rate on Distributions

As shown in Section 2 the shape of the distribution changes with the capture and compression rate. To study this effect, we performed the following experiment: we took two typical video sequences from the two application types (Request Video and Conference Video) and created eight sequences captured at different rates (3, 6, 15, and 30 frames/second). We then compressed each sequence using the two different coding techniques. We derived the distribution of the frame size at the output of the encoder, plotted it, and then superimposed it with estimated Normal, Gamma, and Lognormal distributions. The results from these experiments are shown in Figure 18 and Figure 19.

As expected, inter-frame coded video exhibited the most change as the capture rate was decreased. At a full 30 frames/second the observed distribution exhibited a heavy right tail and was estimated well by the Gamma distribution. As the frame rate was reduced, the correlation between pixels of subsequent frames was reduced, resulting in more bits being required to maintain a constant picture quality (a constant PSNR). This is seen by the right shift in the main lobe of the distribution. The overall shape became more bell shaped, and was better estimated by the Normal distribution. For intra-frame Request Video the shape of the distribution was more consistent. Since intra-frame video does not exploit correlation between pixels of neighboring frames, the observed distribution maintained its shape even as the frame rate was decreased (see Figure 18). The Normal distribution was best suited for estimating at different frame rates.

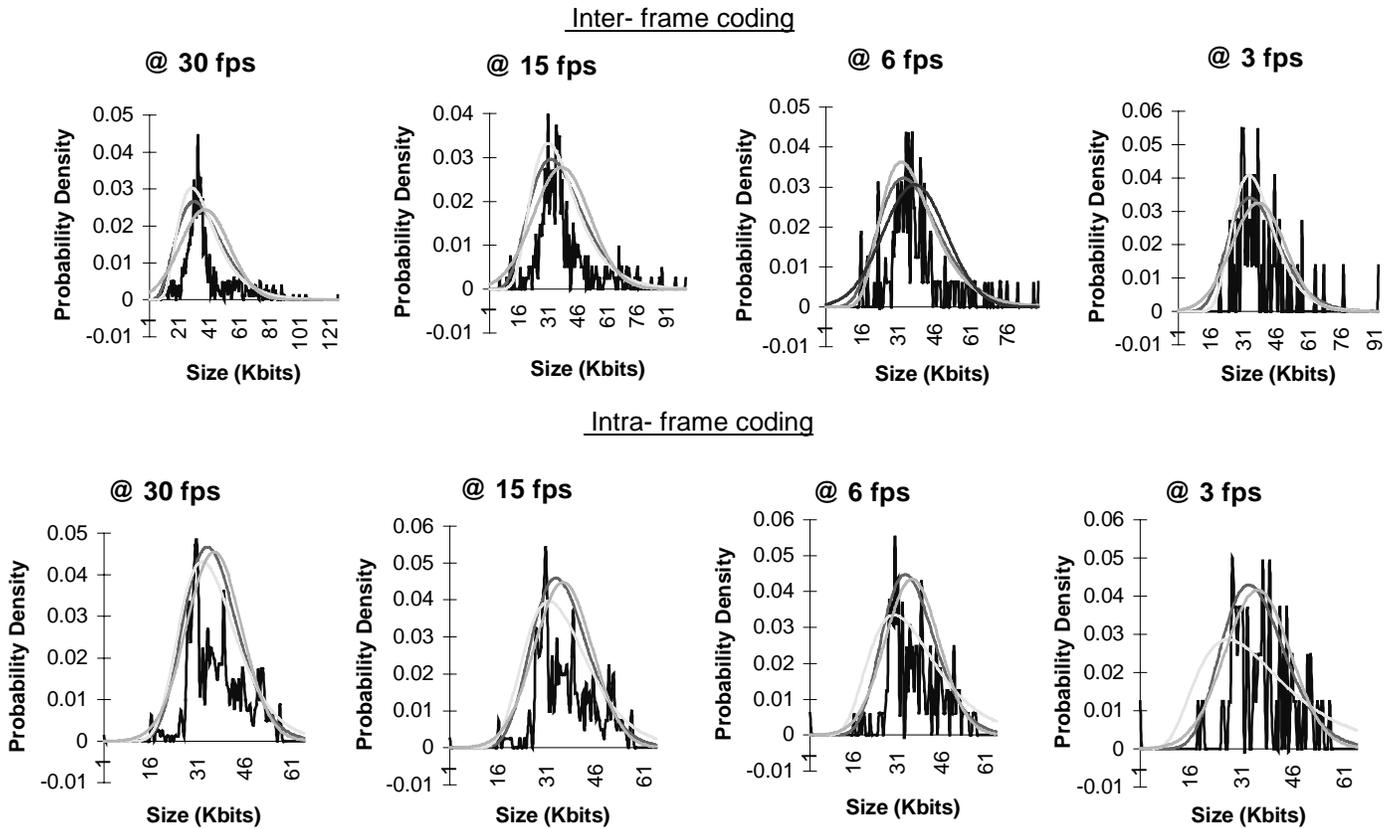


Figure 18: Distribution for Request Video using different compression schemes and different capture rates.

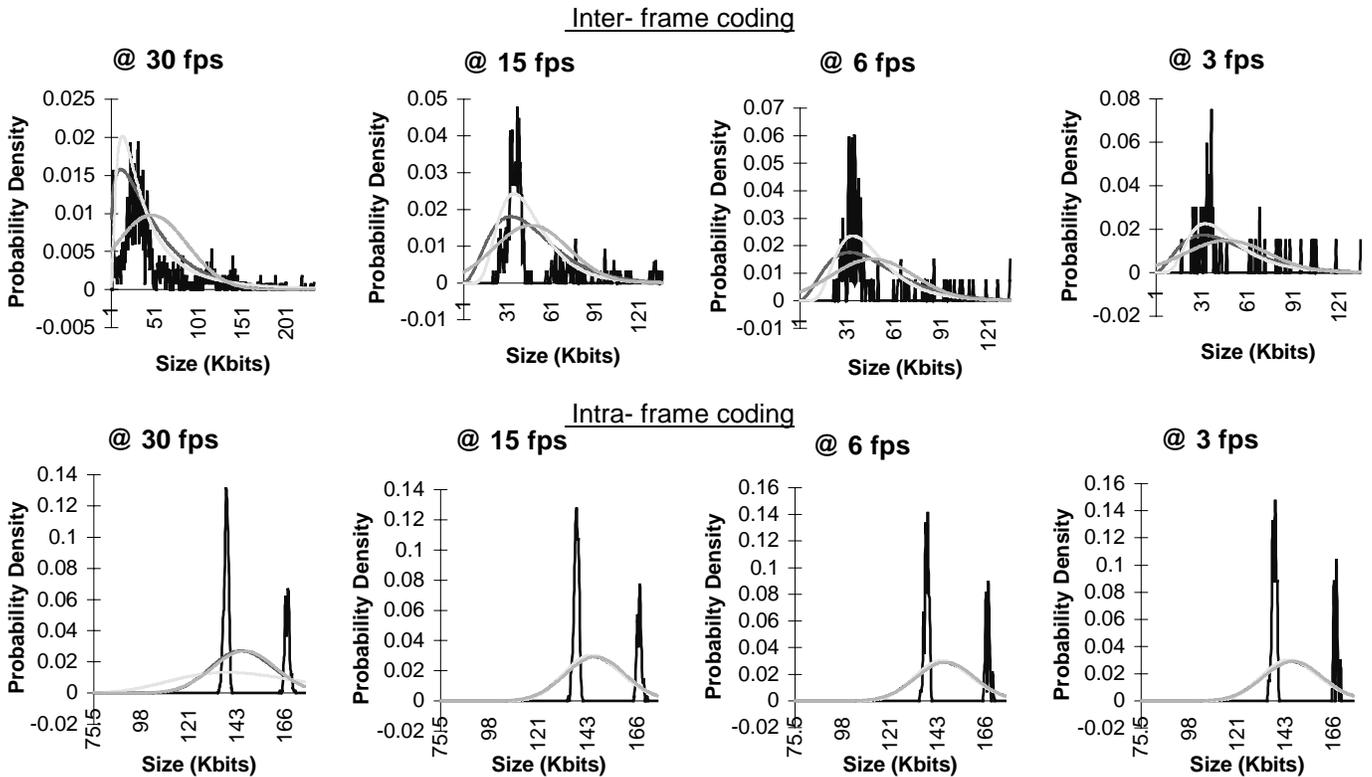


Figure 19: Distribution for Conference Video using different compression schemes and different capture rates.

Conference Video sequences compressed using a inter-frame coding scheme (Rec. H.263) exhibited a heavier right tail than the Request Video sequences. As the capture and compression rate was reduced, the main lobe of the distribution while shifting right became more peaked. At 30 frames per second, a combination of Gamma (for the main lobe) and Pareto (for the heavy tail) distributions estimated the observed distribution best, at lower frames rate all distributions performed poorly. For intra-frame coded compressed video, the shape of the distribution remained relatively unchanged for varying capture rates. The two lobes of the distribution suggest that perhaps a segmented approach leading to a concatenation of two Normal distribution, as recommended in Section 4, would provide the best estimate.

6 Discussion

The results of our experiments are summarized in Table 5. These results are applicable to characterization of single source VBR video traffic. As suggested in Section 4.4, precise modeling is possible when the details of compression algorithm are known and exploited. For Conference Video, results for only “Low Action” video are presented, since by definition Conference Video is constrained to few scene changes and low movement.

Coding Technique	Request Video				Conference Video	
	High Capture Rate		Low Capture Rate		High Capture Rate	Low Capture Rate
	<u>High Action</u>	<u>Low Action</u>	<u>High Action</u>	<u>Low Action</u>	<u>Low Action</u>	<u>Low Action</u>
Intra-Frame	Normal	Normal/Gamma	Normal/Weibull	Weibull	Bi-Normal	Bi-Normal
Inter-Frame	Bi-Normal	Gamma/Lognormal	Normal	Gamma	Gamma-Pareto	Lognormal

Table 5: Results from distribution-based modeling for single VBR video source

Note: Bi-Normal means a combination of two Normal distributions with different means.

As evidence of the usefulness of Table 5 we present an examples of how a network designer might use Table 5 to plan and design networks that are capable of supporting VBR compressed video.

6.1 Capacity Planning

A common and important task for network designers is to estimate the demands on the network for the purposes of capacity planning. In the context of video traffic, the question to be answered is: Given a multimedia (video) connection, estimate the maximum (or peak) bandwidth that the connection will require? Equivalently, when the density function for the video frame sizes is known (from Table 5), what is the estimate for the maximum video frame size?

The question can be answered by deriving the distribution for the maximum value. Let X_j , $j=1,2,\dots,n$ denote the j th frame size of n frames occurring in the video

sequence, we are then interested in the probability distribution of Y_n in terms of the random variables X_j when $n \rightarrow \infty$. The random variable Y_n is defined as:

$$Y_n = \max(X_1, X_2, \dots, X_n)$$

For simplicity, and without losing generality, we assume that X_j are independent and identically distributed. The PDF and density function for Y_n is then given as:

$$F_{Y_n}(y) = [F_X(y)]^n \quad \text{and} \quad f_{Y_n}(y) = n[F_X(y)]^{n-1} f_X(y) \quad (16)$$

While the distribution function $F_{Y_n}(y)$ becomes increasingly insensitive to the exact distribution of X_j as $n \rightarrow \infty$, no unique results can be obtained that are completely independent of the form of $F_X(x)$. Looking at Table 5, we see that the Normal, Gamma and Lognormal distributions describe the video frame size distributions best. Observing that each of the three distributions have right tails that are unbounded and are of the exponential type, that is, for each case $F_X(x)$ approaches one at least as fast as an exponential distribution, the cumulative distribution function can generally be described as:

$$F_X(x) = 1 - e^{-g(x)}, \quad \text{where } g(x) \text{ is an increasing function of } x. \quad (17)$$

Let,

$$\lim_{n \rightarrow \infty} Y_n = Y$$

$$\text{then from [32], } F_Y(y) = \exp[-e^{-\alpha(y-u)}], \quad -\infty < y < \infty \quad (18)$$

where u and $\alpha (>0)$ are the location and the scale parameters of the distribution. u is obtained from u_n as $n \rightarrow \infty$ and is the value of X_j at which $P(X_j \leq u_n) = 1 - \frac{1}{n}$. As n becomes large, $F_X(u_n)$ approaches unity or u_n is in the extreme right tail of the video frame size distribution. The scale parameter α is a limiting case of α_n and can be obtained as $\alpha_n = dg(y)/dy$ evaluated at $y = u_n$.

The mean and variance of Y are given as:

$$\eta_y = u + \left(\frac{0.577}{\alpha} \right) \quad \text{and} \quad \sigma_y^2 = \frac{\pi^2}{6\alpha^2}$$

The above distribution has a skew coefficient that is a non-negative constant, implying that the shape of the distribution is fixed with a dominant right tail.

7 Conclusion

This paper addressed the problem of characterizing and modeling compressed video traffic for efficient design of multimedia networking protocols. The existing approach of distribution-based modeling was revisited, re-evaluated, and extended to include the effect of available network bandwidth. It was shown that for inter-frame coding schemes the frame-size distribution function changes with the capture and compression rate, and for intra-frame coded video the effect of different capture rates is negligible. A matrix containing distributions that best estimate VBR video, classified by application type, compression algorithm, and frame rate was derived. An example of how a network designer might use this matrix to do network capacity planning was described.

References

- [1] N. Ohta, *Packet Video*, Artech House Inc., 1994
- [2] M. W. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self Similar VBR Video Traffic," *ACM SIGCOMM '94*, (1994): 269-280
- [3] D. P. Heyman, A. Tabatabai, and T. V. Lakshman, "Statistical Analysis and Simulation study of Video Teleconference Traffic in ATM Networks," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 1, (March 1992): 49-59
- [4] D. P. Heyman and T. V. Lakshman, "Source Models for VBR Broadcast Video Traffic," *Proceedings of the IEEE INFOCOM '94*, vol. 2, (June 1994): 664-671
- [5] M. R. Izquierdo, and D. R. Reeves, "Statistical Characterization of MPEG VBR video at the slice layer," *Multimedia Computing and Networking 1995, Proceedings of the SPIE 2417*, (February 1995): 268-279
- [6] M. Nomura, T. Fujii, and N. Ohta, "Basic Characteristics of Variable Rate Video Coding in ATM Environment," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 5, (June 1989): 752-760
- [7] P. Pancha, and M. E. Zarki, "A look at MPEG Video Coding Standard for Variable Bit Rate Video Transmission," *Proceedings of the IEEE INFOCOM '92*, vol. 1, (May 1992): 85-94
- [8] P. Sen, B. Maglaris, N. E. Rikli, and D. Anastassiou, "Models for Packet Switching of Variable-Bit-Rate Video Sources," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 5, (June 1989): 865-869
- [9] W. Verbiest, L. Pinno, and B. Voeten, "The impact of the ATM Concept on Video Coding," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 9, (December 1988): 1623-1632

- [10] ISO/IEC IS 10918-1: "Information Technology - Digital Compression and Coding of Continuous Tone Still Images, Part 1: Requirements and Guidelines," (1994)
- [11] ISO/IEC CD 11172-2: "Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to about 1.5 Mbits/s," (1993).
- [12] R. Grunenfelder, J. P. Cosmas, S. Manthorpe and A. Odinma-Okafor, "Characterization of Video Codecs as Autoregressive Moving Average Processes and Related Queuing System Performance", *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 3, (April 1991): 284-293
- [13] D. M. Lucantoni, and M. F. Neuts, "Methods for Performance Evaluation of VBR Video Traffic Models," *IEEE/ACM Transactions on Networking*, vol. 2, no. 2, (April 1994): 176-180
- [14] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Transaction on Communication*, vol. 36, no. 7, (July 1988): 834-844
- [15] B. Melamed and B. Sengupta, "TES modeling of Video Traffic," *IEICE Transactions Communications*, vol. E75-B, no. 12, (Dec. 1992): 1292-1300
- [16] P. Skelly, S. Dixit, and M. Schwartz, "A Histogram-Based Model for Video Traffic Behavior in an ATM Network Node with an Application to Congestion Control", *Proceedings of the IEEE INFOCOM '92* (1992): 95-104
- [17] F. Yegenoglu, B. Jabbari, and Y. Zhang, "Modeling of Motion Classified VBR Video Codecs," *Proceedings of the IEEE INFOCOM '92*, (1992): 105-109
- [18] P. Bahl, "The J300 family of Audio and Video Adapters - Software Architecture", *Digital Technical Journal*, vol. 7, no. 4, (1995): 34-51
- [19] P. Bahl, P. Gauthier, and R. Ulichney, "Software-Only Compression, Rendering and Playback of Digital Video", *Digital Technical Journal*, vol. 7, no. 4, (1995): 52-75
- [20] ITU-T Recommendation H.261, "Video Codec for Audiovisual Services at p x 64 kbits/s," *CDM XV-R 37-E, International Telegraph and Telephone Consultative Committee* (1990).
- [21] ISO/IEC CD 1318-2: "Generic Coding of Moving Pictures and Associated Audio, Recommendation H.262", (1994)
- [22] T. T. Soong, Probabilistic Modeling and Analysis in Science and Engineering, *John Wiley & Sons Inc.*, 1981
- [23] R. A. Fisher, "Theory of Statistical Estimation," *Proceeding of the Cambridge Phil Society*, 22 (1925): 700-725
- [24] K. Pearson, "Contributions to the Mathematical Theory of Evolution," *Phil. Transactions of the Royal Society of London, Ser. A*, 185, (1894): 71-78

- [25] B. C. Arnold, "Pareto Distribution," *Encyclopedia of Statistical Sciences*, vol. 6, John Wiley & Sons Inc., (1985): 569-574
- [26] T. Hettmansperger, and M. Keenam, "Tailweight, Statistical Interference, and Families of Distributions - A Brief Survey," *Statistical Distribution in Scientific Work*, vol. 1, (1980): 161-172
- [27] V. Paxson and S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3 (June 1995): 226-244
- [28] W. Weibull, "A Statistical Theory of the Strength of Materials," *Proceedings of the Royal Swedish Institute for Engineering Research*, no. 151 (Stockholm, 1939)
- [29] R. B. Agostino and M. A. Stephens, Goodness-of-fit Techniques, *Marcel Dekker*, 1986
- [30] R. Kishimoto, Y. Ogata, and F. Inumaru, "Generation Interval Distribution Characteristics of Packetized Variable Rate Video Coding Data Streams in an ATM Network," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 5, (March 1989): 833-841
- [31] R. Jain, The Art of Computer Systems Performance Analysis, *John Wiley & Sons, Inc.* 1991
- [32] E. J. Gumbel, Statistics of Extremes, *Columbia University Press*, New York 1988