

Video Frame Size Distribution Analysis

Denise M. Bevilacqua Masi, Ph.D.

Martin J. Fischer, Ph.D.

David A. Garbin

An analysis is performed to determine an appropriate statistical distribution for video frame sizes. As converged networks are used more often, source models of disparate types of traffic (including video) that can be used for congestion modeling are needed. Three data sets of actual video frame sizes are used, which represent two different video compression standards—H.263 and H.264. Previous analyses of video frame sizes have been performed in the literature, but research is limited; some use dated video compression standards and others do not investigate use of their hypothesized distribution in performance models. Our findings include the best fitting theoretical distributions which differ by data sets; the Erlang or gamma distributions are recommended as being suitable across the range of the three data sets. Correlations in frame sizes and the decomposition of frames into packets were also analyzed. The fitted frame size distributions were used to generate packet streams for use in packet-level congestion models. Findings are that preserving the batch nature of packet arrivals results in packet congestion estimates that approximate the congestion of the actual data fairly well, even when poorly-fitting frame size distributions are used.

Introduction

Networks are moving toward the use of Internet Protocol (IP) technology for integrated voice, data, and even video services. As converged networks are used more often, source models of disparate types of traffic (including video) are needed. Such source models can be used as inputs to performance models of the integrated networks. Accurate performance models enable quality of service (QoS) using different router configurations and queueing mechanisms to be analyzed. The purpose of this research is to determine what statistical distributions are most appropriate to represent the video traffic.

Video traffic is produced by imaging devices in frames, which contain both the audio and picture portions of the video traffic. The sizes of the video frames vary both between and within video formats. Most new videoconferencing products now include the H.264 compression standard, and the older H.263 and H.261 standards. H.263 was first developed in 1995, and H.264 is a newer, lower bit rate standard developed in 2003. H.264 supports a wide variety of applications, e.g., videoconferencing and streaming video, while H.263 is primarily used for videoconferencing. H.261 is an Integrated Services Digital Network (ISDN) standard developed in 1990.

Several researchers have analyzed the frame sizes produced by video devices. Early work in this area was conducted by Maglaris et al., [1] which developed a video frame size model. This work was done on an unnamed video standard; based on the timing of the research, it was likely similar to the H.261 predecessor, H.120. Maglaris et al. examined two models; the first model is an autoregressive discrete-time Markov process of the frame sizes, which they use as input to a dis-

crete-event simulation model. The number of compressed bits per pixel in the n^{th} frame, $\lambda(n)$, is modeled by a first-order autoregressive Markov process, $\lambda(n) = a*\lambda(n-1) + b*w(n-1)$, where a and b are constant, $|a| < 1$, and $w(n)$ is a sequence of independent Gaussian random variables with mean η and variance 1. Maglaris et al. also developed a second model, which was a continuous-time Markov process of the frame sizes, used as input to a fluid-flow queueing analysis. The Maglaris discrete-time model (including the original estimated parameters) has been used subsequently by Liu and Babiarz [2] of Nortel in simulation models reported to the Internet Engineering Task Force (IETF) Pre-Congestion Notification Working Group.

Krunz and Hughes [3] modeled the frame sizes for MPEG-2 streaming (also known as H.262). In this research, a lognormal distribution was found to fit the frame sizes best; gamma, Weibull, and lognormal distributions were considered. The fitted distributions were specific to the three types of frames present—the Intra (I), Predictive (P) and Bidirectional predictive (B) frames.

Koumaras et al. [4] found that a gamma distribution fits H.264 frame sizes well, where the distribution is again specific to each of three types of frames. They also found evidence of positive autocorrelation in the frame sizes, with a decaying rate after the first few lags, but did not incorporate autocorrelation into their model. Doggen and Van der Schueren [5] build an OPNET model using the Koumaras et al. frame size distributions. Although Doggen and Van der Schueren state that the model was tested in a local area network (LAN) environment, they do not report on any calibration results of the model versus actual laboratory behavior.

The above literature results in a range of different hypothesized models for video, depending on the video standard being modeled. Limited mention is made of utilizing the hypothesized distributions as input for performance models—only Maglaris et al. and Doggen and Van der Schueren make mention of this.

The purpose of this paper is to fit statistical distributions to the frame sizes, from several video data sets, for use in congestion models. We collected video packet trace data using the Ethereal, a popular open source network protocol analyzer which can capture packets from a variety of devices. Our data sets used two different video compression standards—H.263 and H.264. Two packet traces were collected using H.264 at high and low bit rates, and one packet trace was collected using H.263. The H.263 video data was generated by a Polycom device. The H.264 data sets were generated from an Apple iSight camera's iChat application. All three data sets were collected from point-to-point videoconferencing footage.

This paper will proceed to describe our methodology, analysis results, and conclusions.

Methodology

The recommended approach to fitting statistical distributions to a data set involves the following steps (Law and Kelton). [6]

1. Hypothesize families of distributions that appear to be appropriate on the basis of their shapes.
2. Estimate parameters for the selected families of distributions by writing code to generate one of the three types of estimators (e.g., maximum-likelihood estimators [MLEs], method of moments, or least-squares estimators).
3. Test the goodness-of-fit of the hypothesized distributions to the data set.

Three goodness-of-fit tests are often discussed in the literature on data fitting—the Kolmogorov-Smirnov (K-S) test, the Anderson-Darling test, and the Chi-Square test. Interpretation and use of the tests can be complex; some of the tests have tables of critical values available for only certain hypothesized distributions (e.g., the Anderson-Darling test and K-S test). Because critical values for the K-S test are not readily available, often approximate critical values are used for sample sizes greater than 80 (see Knuth, 1998 [7]; also reference [8]). A further complication of using goodness-of-fit tests to assess how representative a hypothesized distribution is of a data set is that goodness-of-fit tests are of limited value for large data sets ($n > 1000$), and one has to rely more on graphical comparisons; any hypothesized distribution will be rejected for virtually all real-world large data sets. [9] For more information on goodness-of-fit tests, see Law and Kelton or the *National Insti-*

tute of Standards and Technology (NIST) Engineering Statistics Handbook. [10]

Initially we began to conduct the analysis described above “manually,” using programming languages such as the R Programming Environment for Data Analysis and Graphics (“R”) or Visual Basic for Applications to write code to estimate parameters and create graphs to visualize the data and distribution fits. Our selection from among the types of statistical estimators was based on the algorithm tractability for the specific hypothesized distribution. We considered the normal, gamma, generalized exponential, and lognormal distributions, and wrote programs to estimate parameters for these distributions associated with our three data sets. For example, for the normal and gamma distributions, we used the method of moments due to simplicity. For the generalized exponential and lognormal distributions, maximum likelihood estimators were used due to availability of a straightforward algorithm (see Gupta and Kundu [11] for generalized exponential and Law and Kelton [6] for lognormal). This analysis to select an appropriate input distribution proved to be time consuming, requiring expert knowledge of parameter estimation methods for a number of distributions so that all candidate distribution families can be considered. Thus, our experience with the above distribution selection and estimation process was limited by the time required to research and program the parameter estimation routines.

We then decided to make use of a distribution-fitting software package called ExpertFit. ExpertFit automates the recommended approach to determine which probability distribution best represents a data set. The software considers about 40 statistical distributions, and can estimate parameters and generate graphical plots associated with each of them. It reports results from the three common goodness-of-fit tests described earlier, and ranks the statistical distributions to select the best fitting based on the goodness-of-fit and other results. This software can handle data sets with sample sizes of up to 100,000. ExpertFit generally uses maximum-likelihood estimators for estimating the parameters, which is recommended due to the desirable properties that MLEs possess. [6] In some cases we manually estimated parameters (using our own code programmed in R) for distributions which the ExpertFit package also recommended, enabling the results from the two packages to validate each other. ExpertFit does not consider the generalized exponential distribution, but our fits of this distribution to the video frame sizes indicated that this distribution is not suitable for these data sets. Based on the greater number of distributions that could be considered using ExpertFit versus the manual approach, and the more extensive goodness-of-fit tests and graphical comparisons that ExpertFit provides versus what we were able to perform manually, use of this software provides for a more rigorous analysis.

Analysis results

The results of the video frame size distribution fitting will be described for each of the three data sets. The H.263 video data analysis is first described, followed by the H.264 high and low bit rate data analysis. An analysis of autocorrelations in the frame sizes is then described, and, lastly, an initial analysis of the packet composition within the frames is presented.

H.263 Polycom frame size data fitting

A histogram of the H.263 Polycom frame sizes revealed that there are eight frame size values that are between 20 and 25 standard deviations from the mean frame size, while the remaining frame size values are less than seven standard deviations from the mean frame size. Further investigation revealed that these large frames were likely control or test frames to re-adjust the video rate, rather than frames carrying actual video content, so it was decided to omit these eight outliers from the

analysis data set. Figure 1 displays the revised histogram of the frame sizes with the outliers removed. All further analysis on the H.263 Polycom frame sizes uses the data set with the outliers removed. Figure 1's appearance and the non-zero skewness value indicate that the normal distribution may not be the best distribution to fit this data.

An analysis using ExpertFit's automated fitting capability resulted in the fitting of approximately 30 statistical distributions to the data set, and the ranking of the fits of the various distributions, as shown in Table 1. The distributions are ranked on the basis of a "relative score" computed by ExpertFit. The computation of the relative score is proprietary, but it does use several measures of error to rank the distributions in terms of quality-of-fit. Then the individual rankings are combined and normalized to produce the relative score. For the H.263 Polycom data set, the Weibull distribution is the best fit to the data set. Erlang, gamma, and Johnson SB distributions are also highly ranked.

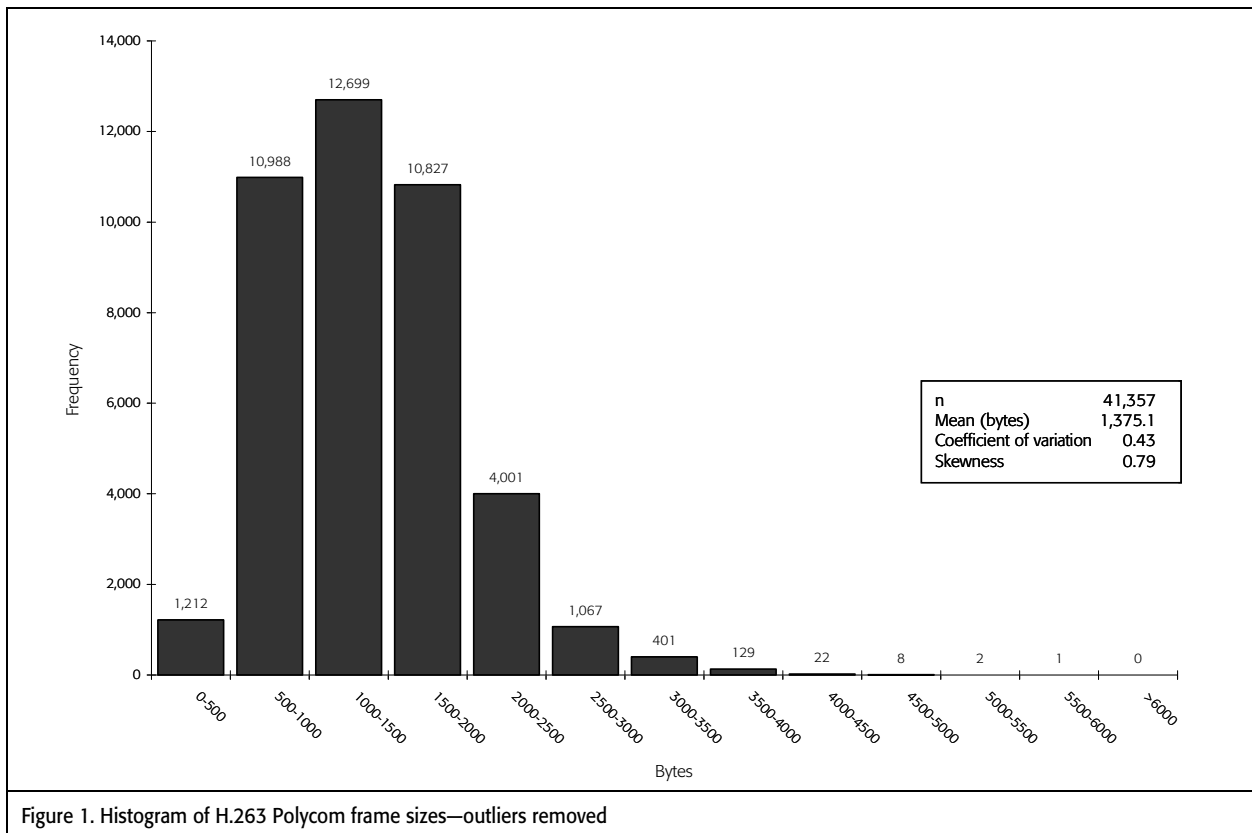


Figure 1. Histogram of H.263 Polycom frame sizes—outliers removed

Table 1. H.263 Polycom automated fitting results			
<i>Relative evaluation of candidate models</i>			
Model	Relative Score	Parameters	
1. Weibull (E)	94.38	Location	190.99720
		Scale	1,339.06913
		Shape	2.10845
2. Weibull (E)	94.38	Location	190.99035
		Scale	1,339.08734
		Shape	2.10854
3. Erlang (E)	90.63	Location	5.38666
		Scale	273.94371
		Shape	5
4. Johnson SB	90.00	Lower endpoint	0.00298
		Upper endpoint	5,928.93605
		Shape #1	2.17825
		Shape #2	1.69500
5. Erlang	89.38	Location	0.00000
		Scale	275.02104
		Shape	5
6. Gamma	88.44	Location	0.00000
		Scale	259.46888
		Shape	5.29969
7. Gamma (E)	86.56	Location	4.34856 e-4
		Scale	259.46899
		Shape	5.29969
12. Gamma (E)	73.75	Location	190.99720
		Scale	321.35096
		Shape	3.68478
14. Erlang (E)	63.13	Location	190.99720
		Scale	394.70266
		Shape	3
19. Normal	56.25	Location	1,375.10518
		Scale	593.63886

The Johnson SB is a continuous distribution that is bounded with both a minimum and maximum. The use of (E) in Table 1 indicates the location parameter is estimated from the data. Several of the distributions that are highly ranked use a location parameter close to the minimum value in the data set (e.g., Weibull), while others have location parameters equal to or close to zero (e.g., Erlang and gamma). The normal distribution has a much lower ranking of 19, and has a much lower relative score than the Weibull, Erlang, gamma, and Johnson SB distributions.

In addition to the ExpertFit automated ranking and relative scores, a number of plots and tables can be generated for evaluation of the fit of the hypothesized distributions. Figure 2 shows density histogram plots of the data overlaid by the fitted Weibull, Erlang, Johnson SB, gamma, and normal distributions. The Weibull distribution fit is perhaps slightly better than the Erlang or Johnson SB distributions, as both the Erlang and Johnson SB are slightly skewed to the left part of the empirical distribution. The Erlang and Johnson SB distributions are vir-

tually indistinguishable from each other. The gamma distribution appears to overestimate the mode slightly. The normal distribution is not a good fit—it does not fit the mode of the data well, plus it has a higher probability of low values (≤ 263 bytes) than the other distributions. With the exception of the normal distribution, any of the distributions shown in Figure 2 appear to be suitable for the H.263 Polycom data set.

Table 2 is a comparison of the test statistics and their rankings from each of the goodness-of-fit tests. Goodness-of-fit tests are of limited value for large data sets, and these tests do reject all hypothesized distributions for the H.263 Polycom data set, as predicted in discussions with Law [9] on large real world data sets. However, a comparison of the magnitude of these test statistics is still meaningful, as smaller values of the statistics indicates a smaller deviation in the hypothesized distribution from the empirical data. The rankings of the three goodness-of-fit tests indicate that although no one distribution is consistently the highest ranked for all three tests, the Weibull, Erlang, Johnson SB, and gamma are all ranked among the top 10 for the three tests.

Other graphical and tabular comparisons (not shown) that are available for evaluation of fitted distributions include distribution function absolute differences tables, probability-probability (P-P), and quantile-quantile (Q-Q) plots and tables, tables summarizing the goodness-of-fit test statistics for each of the fitted distributions and their rankings, and box plots showing various quantiles and extremes of the sample data and fitted distributions. The relative-discrepancies table (not shown) summarizes the output of the P-P and Q-Q plots; based on this output, the Weibull distribution has the best fit. Finally, the box-plot comparisons of the hypothesized distributions versus the empirical data indicate that the upper extreme of the Weibull distribution is not as high as the maximum value in the data set. The gamma and Erlang distributions are slightly better in representing the extremes of the data.

In summary, the Weibull distribution is favored based on the rankings of the relative scores, discrepancies versus hypothesized distributions, most of the goodness-of-fit tests, and graphical comparisons performed by ExpertFit. Normal is not a very good fit, based on the ExpertFit analysis (and supported by our own analysis done using the R software). The Erlang and gamma distributions also have fairly good fits. Note that the gamma (m,beta) and m-Erlang (beta) distributions are the same for integer m; in this case, the estimated gamma shape parameter is 5.3 and the Erlang shape parameter is $m = 5$, which are fairly close. The gamma distribution was also fitted separately and visualized in our own analysis using R, which corroborated the ExpertFit results for the gamma.

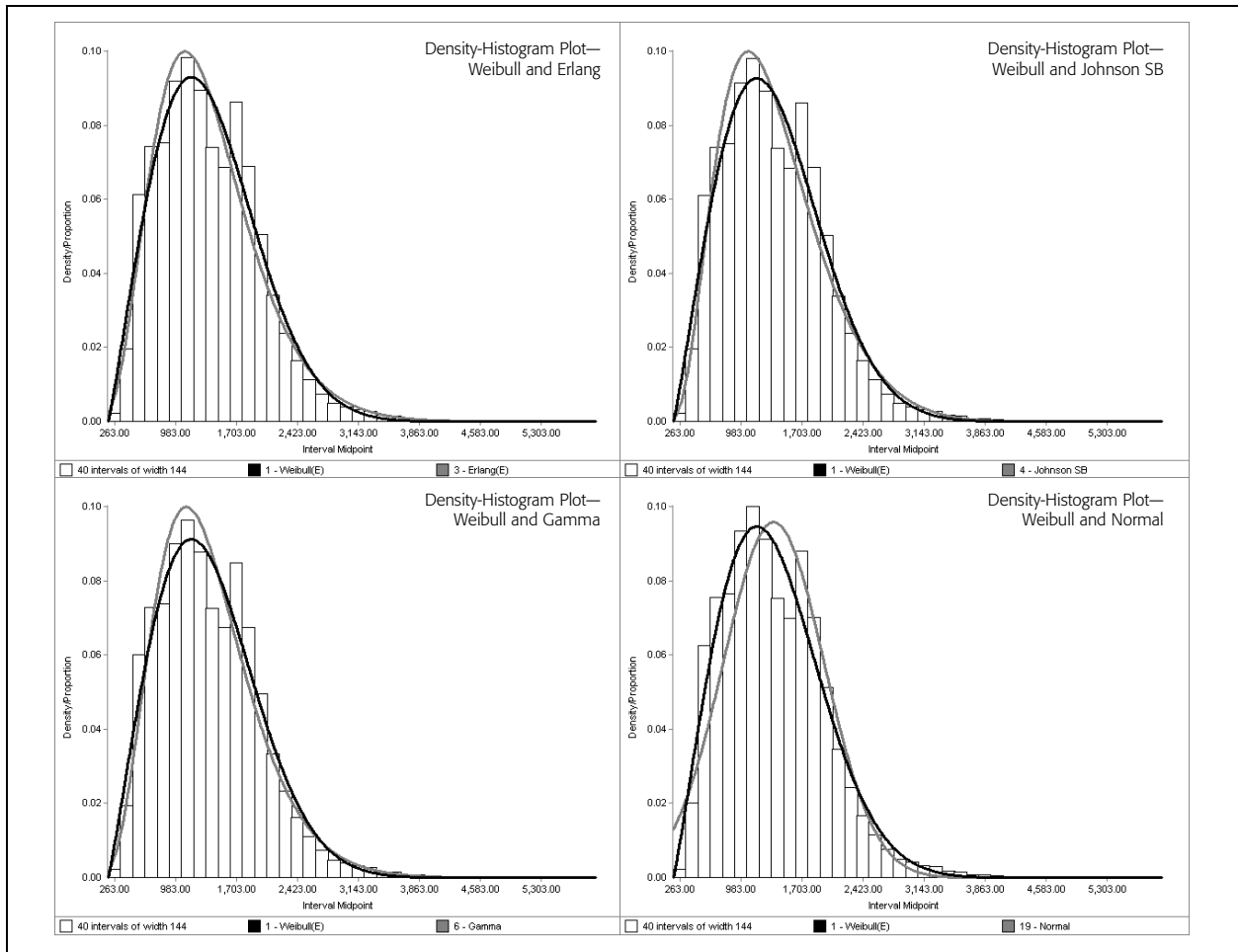


Figure 2. H.263 Polycom density histogram plots

Model	Anderson-Darling Test		Kolmogorov-Smirnov Test		Equal-Probable Chi-Square Test	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
1. Weibull (E)	39.08325	1	0.02313	2	2,203.79890	2 ½
2. Weibull (E)	39.08492	2	0.02313	1	2,203.79890	2 ½
3. Erlang (E)	59.60491	3	0.04382	8	2,321.61015	6
4. Johnson SB	58.37817	4	0.04234	6	2,111.84832	1
5. Erlang	59.87134	5	0.04349	7	2,367.19373	7
6. Gamma	64.51819	6	0.04441	9	2,233.26332	4 ½
7. Gamma (E)	64.51822	7	0.04441	10	2,233.26332	4 ½
8. Pearson Type VI	73.03936	8	0.04707	11	2,476.02561	8
9. Rayleigh (E)	78.55319	9	0.03141	4	2,488.35919	9 ½
10. Rayleigh (E)	78.56372	10	0.03141	5	2,488.35919	9 ½
11. Weibull	100.77936	12	0.03140	3	3,015.41289	13
12. Gamma (E)	94.14589	11	0.05185	15	2,574.98249	11
13. Extreme Value B	109.99816	13	0.05443	16	2,967.03801	12
14. Erlang (E)	224.83764	17	0.05080	12	3,613.46014	16

H.264 high bit rate frame size data fitting

A histogram of the H.264 high bit rate frame sizes is shown in Figure 3; outliers do not appear to be a problem for this data set. The skewness value is less than that for the H.263 Polycom data set, although it is still non-zero, so the normal distribution may not be the best distribution to fit this data.

The analysis results using ExpertFit's automated fitting capability differ from those for the H.263 data set. For the H.264 high bit rate data set, the Pearson Type VI distribution fits best, with Erlang and gamma distributions close behind. The normal distribution is ranked 21st, and has a much lower relative score than Pearson Type VI, Erlang, gamma, and log-normal. Unlike with the H.263 Polycom data set, the Weibull distribution is not a good fit for H.264 high bit rate—the rankings of various Weibull distributions are 30th, 31st, and 32nd.

Examination of the other model comparison plots is revealing. Figure 4 shows density histogram plots of the data overlaid by the fitted Pearson Type VI, Erlang, gamma, log-normal, and normal densities. The fit of the Pearson Type VI, Erlang, gamma, and lognormal distributions appears to be very similar. Normal does not fit as well—it is shifted to the right of

the mode, and has a higher probability of low values than the other distributions. With the exception of the normal distribution, any of the distributions shown in Figure 4 appear to be suitable for the H.264 high bit rate data set based on the density histogram plots.

We do not show the test-statistics comparison table for the H.264 high bit rate data set due to space limitations. The comparisons of the test-statistics from the goodness-of-fit tests show that the Erlang, Pearson Type VI, and gamma all have similar values for the test statistics. As in the case of the H.263 Polycom data set, no one distribution is consistently the highest ranked for all three tests in the H.264 high bit rate data set. The box plot comparisons are similar for the top four hypothesized distributions, and indicate that the lower extremes of the top four hypothesized distributions are not quite as low as the actual data. In the relative-discrepancies table (not shown), the Pearson Type VI has lower Q-Q discrepancies than the Erlang and gamma distributions, but the differences in the discrepancy values for all distributions are small. The Pearson Type VI, Erlang, and gamma P-P plot relative discrepancies are quite close.

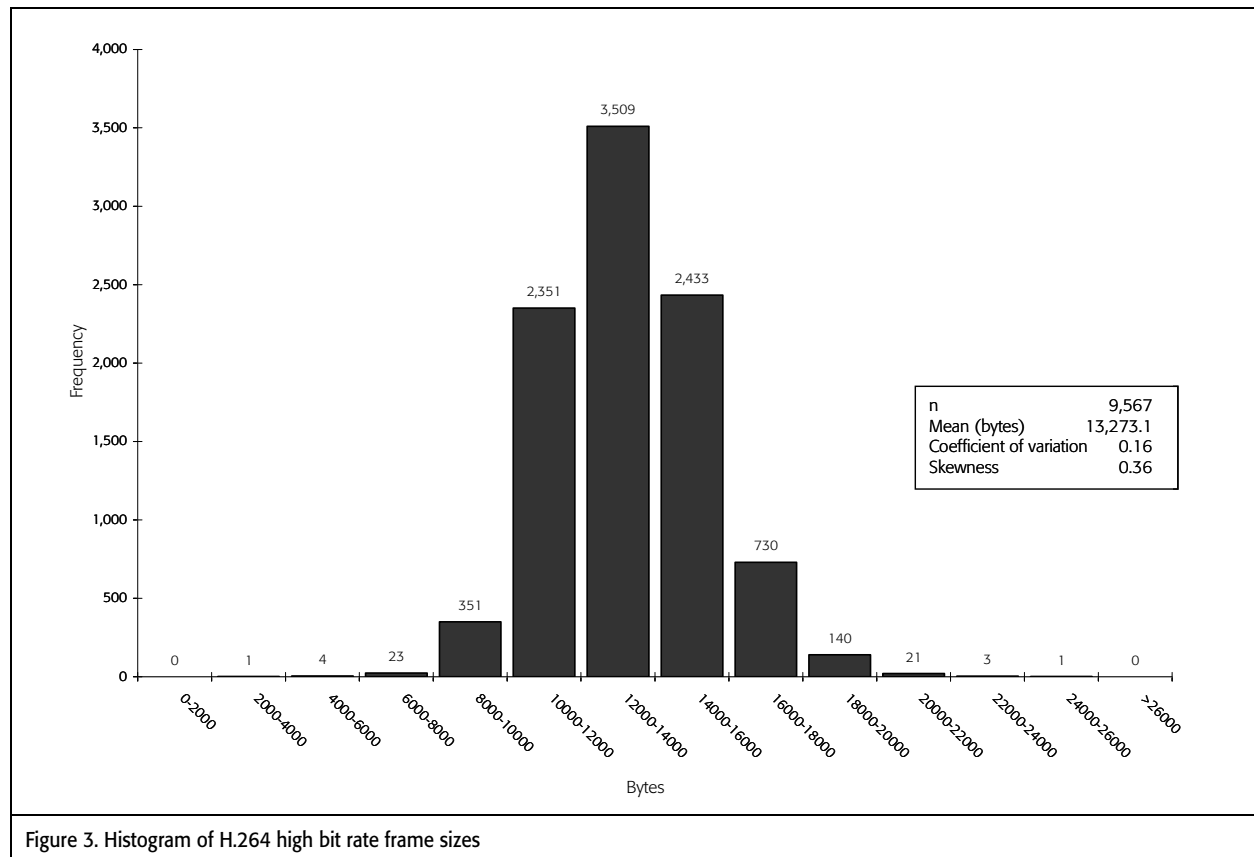


Figure 3. Histogram of H.264 high bit rate frame sizes

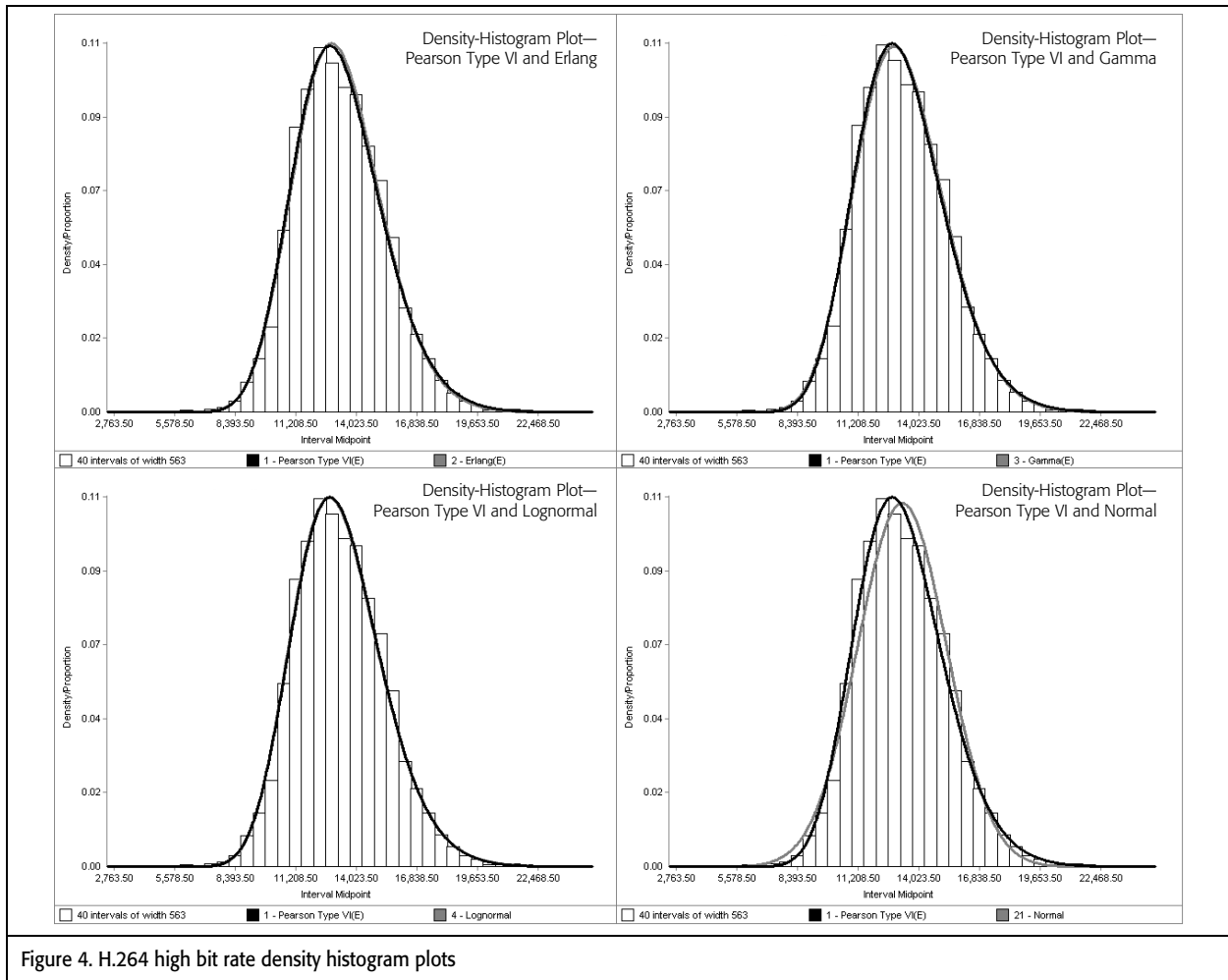


Figure 4. H.264 high bit rate density histogram plots

In summary, either the Pearson Type VI, Erlang, or gamma distributions would be a suitable fit for the H.264 high bit rate video frame data based on the rankings of the relative scores, discrepancies versus hypothesized distributions, most of the goodness-of-fit tests, and graphical comparisons performed by ExpertFit. Normal is not a very good fit to these frame sizes, based on the ExpertFit analysis and supported by our analysis using the R software.

H.264 low bit rate frame size data fitting

The analysis of H.264 low bit rate frame sizes is presented next. A histogram of the frame sizes for this data set is shown in Figure 5. Outliers do not appear to be a problem for this data set. The skewness value (.24) is lower than that for the H.263 Polycom data set and the H.264 high bit rate data set, but it is still non-zero, indicating some departure from normality.

The possible distributions were considered using ExpertFit's automated fitting capability for the H.264 low bit rate

set. The top candidate models in order of their ranking are the normal, log-logistic, and log-Laplace distributions. The Weibull, Erlang, and gamma distributions are ranked just behind the log-logistic and log-Laplace distributions.

Figure 6 shows the density histogram plots of the data overlaid by the fitted normal, log-logistic, log-Laplace, Weibull, and gamma distributions. None of the distributions appear to fit the data very well. Most of the distributions underestimate the density at the mode of the data, with the exception of the log-Laplace distribution which over-estimates the density at the mode of the data. The log-Logistic distribution is slightly shifted to the left of the mode. The normal appears to be a slightly better fit than the Weibull or gamma. The density histogram plot for the Erlang distribution is not shown, but the fit of the Erlang density is similar to the gamma. Based on these plots, the normal and log-Laplace are perhaps the best choices for the H.264 low bit rate data set, but there are deficiencies in all the distribution fits.

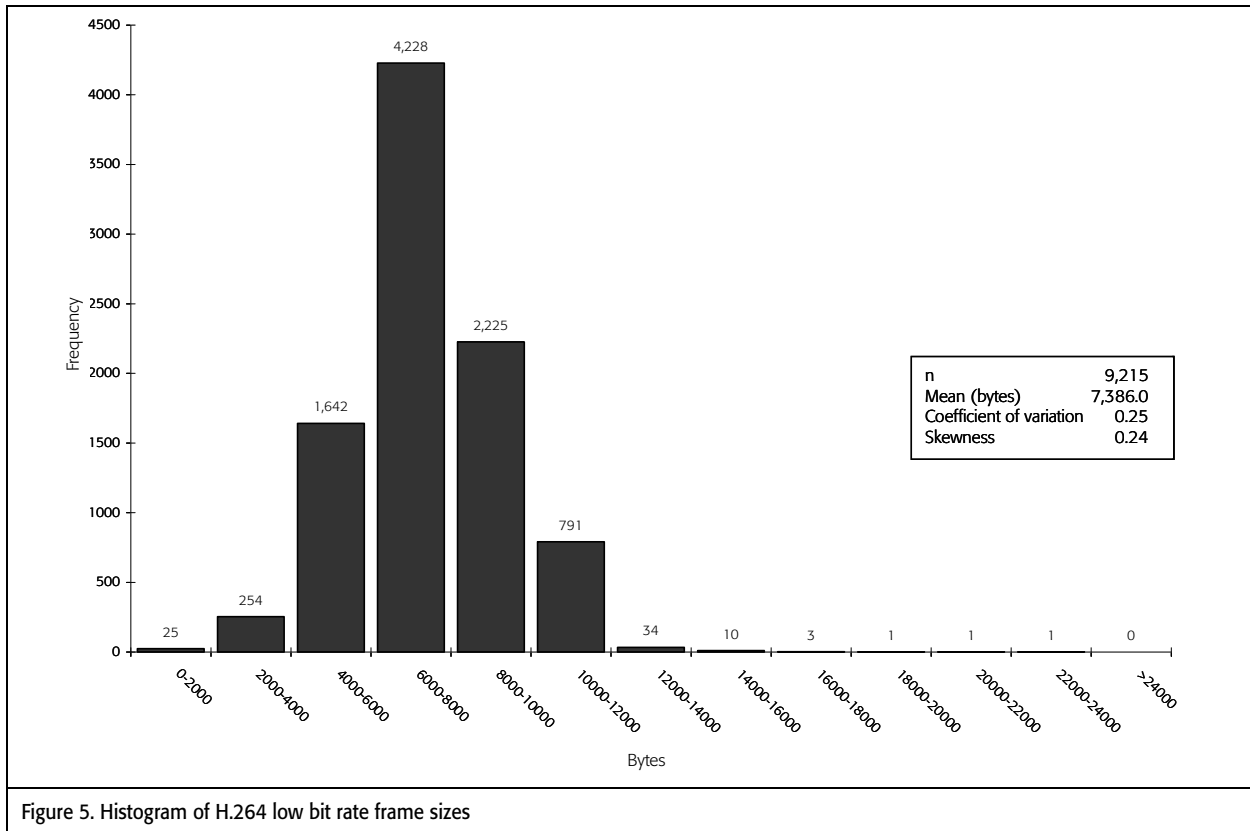


Figure 5. Histogram of H.264 low bit rate frame sizes

The comparisons of the rankings of the test statistics from the goodness-of-fit tests show that the normal is most appropriate (the test-statistics comparison table is not shown due to space limitations). As in the cases of the other data sets, no one distribution is consistently the highest ranked for all three tests in the H.264 low bit rate data set. The box-plot comparisons indicate that the log-logistic and log-Laplace distributions have upper extremes that are too high compared to the sample data. The box plots also show that the normal, Weibull, Erlang, and gamma distributions have upper extremes that are too low compared to the actual data. Based on the P-P ranking in the relative-discrepancies table (not shown), the normal distribution has the best fit. However, the best overall distributions perform poorly on the Q-Q plot rankings, and the lognormal distribution (which overall ranks 14th and 15th) does the best in terms of the Q-Q ranking, indicating the lognormal has a better fit with the order statistics of the data.

In summary, the normal distribution is most suitable for the H.264 low bit rate data set, although there is noticeable departure of this distribution from the data, and the normal is not appreciably better than some of the other distributions considered (log-logistic, log-Laplace, Weibull, Erlang, and gamma distributions).

Autocorrelation structure of frame sizes

As previously mentioned, some researchers have incorporated autocorrelations into their models for video frame sizes. We next examine the autocorrelation structure of these H.263 and H.264 frame data sets to determine if an autocorrelation component needs to be included.

One can compute and plot the autocorrelation function of the video frames, with the limits of the 95 percent confidence interval of the autocorrelations. For the H.263 frames, the autocorrelations are small, with a maximum autocorrelation value of about .11 from lags 1 through 20. Although small, these autocorrelations are statistically significant as they exceed the 95 percent confidence limits. Autocorrelations for the H.264 high and low bit rate data sets were also small, but significant—up to a lag of about 10 to 20.

We then sought to determine if our frame size model needs to account for these small, but statistically significant, amounts of autocorrelation. We tested the impact of the autocorrelation in the original data by using the trace frame sizes in a simulation model of the frame transmission on a link, compared to a simulation using frame sizes drawn independently from the hypothesized distributions. Both simulations used the same frame arrival times from the data trace. The simulations

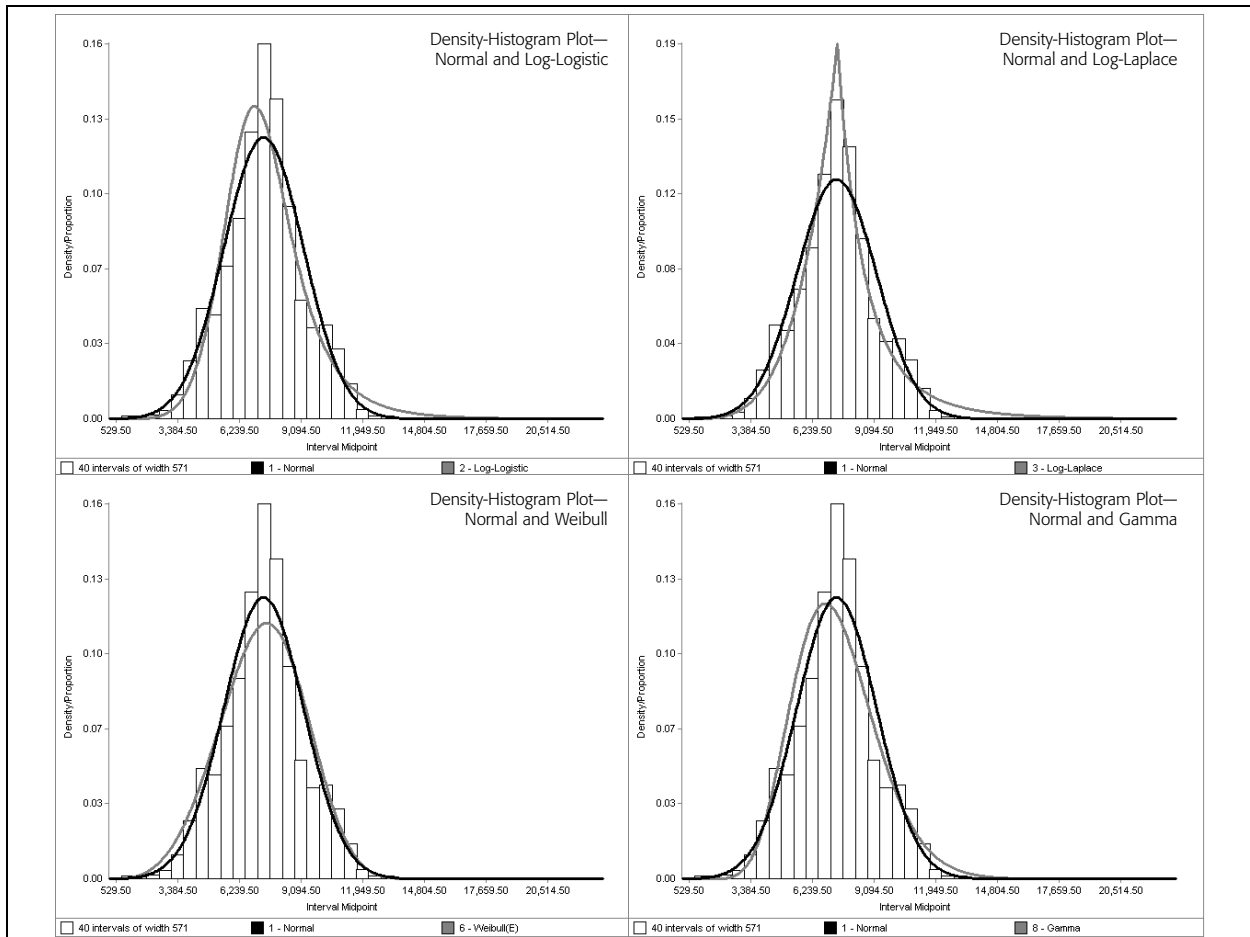


Figure 6. H.264 low bit rate density histogram plots

assumed frames are transmitted on a first come, first served basis, and used Lindley's recursion to compute the queue wait times for the frames. See Gross and Harris [12] or Fischer and Masi [13] for details on Lindley's equation, which allows one to compute queue waits given arrival and service times from a single server queue. This would serve to test whether the resulting queue waits from the simulation would be significantly impacted by using frame sizes selected independently from a hypothesized distribution versus

the actual frame sizes which do exhibit some low-level auto-correlation.

Table 3 shows the simulation results for the H.264 high bit rate data, which compare the mean and variance of the queue waits from the trace, versus the hypothesized Pearson Type VI and Erlang distributions. We found that queue waits from hypothesized distributions are lower than the trace expected queue waits and queue wait variance. Simulations were similarly conducted for the H.263 and H.264 low bit

Table 3. H.264 high bit rate simulation results

Load	Trace		1-Pearson Type VI		2-Erlang	
	EWq	Var(Wq)	EWq	Var(Wq)	EWq	Var(Wq)
0.32	0.00E+00	0.00E+00	1.61E-07	2.48E-10	0.00E+00	0.00E+00
0.40	1.77E-07	3.00E-10	6.93E-07	4.60E-09	9.93E-08	9.44E-11
0.53	7.70E-06	8.19E-08	1.58E-06	2.39E-08	7.88E-07	5.94E-09
0.80	9.46E-04	1.21E-05	7.22E-04	7.20E-06	7.65E-04	7.39E-06
0.99	5.87E-01	2.00E-01	8.16E-02	3.70E-03	1.84E-01	2.13E-02

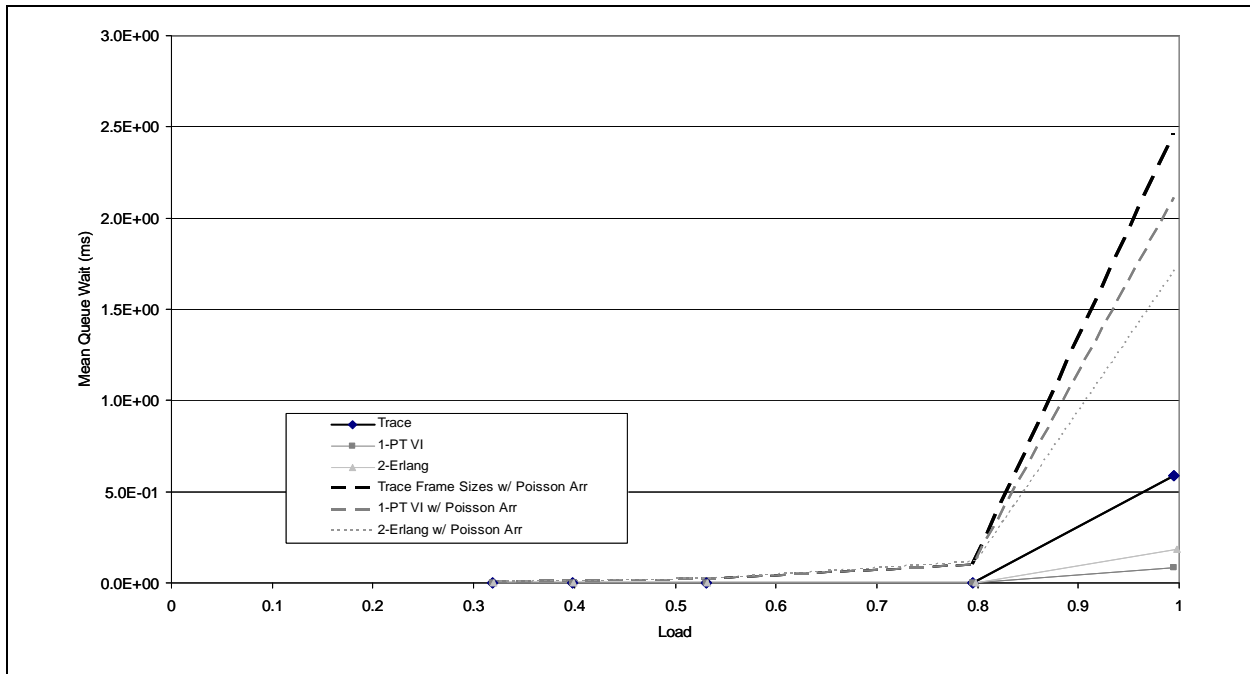


Figure 7. H.264 high bit rate mean waits for trace versus hypothesized distributions

rate trace data sets and their hypothesized frame size distributions. The magnitude of differences between the trace frame size simulation measures and hypothesized frame size simulation measures were similar to those seen for the H.264 high bit rate case, although in one case the direction of the differences was reversed.

We also investigated the impact of assuming Poisson frame arrivals, with the same mean interarrival time, rather than the actual trace arrival times. Figure 7 shows the expected queue wait results for the H.264 high bit rate simulations; very similar results were obtained for the other data sets, but are not shown. We found that in all three cases using Poisson arrivals causes the queue waits to be highly overestimated. We suspected that use of Poisson arrivals would not be appropriate, and our analysis confirms this. Poisson arrivals result in greater errors in queue waits than use of hypothesized frame distributions with actual arrival times does. Thus, the independence in the frame sizes, which is a basis for the hypothesized frame size distributions, appears to introduce small amounts of error versus the actual trace data.

In summary, differences between the simulation performance measures with the trace versus the hypothesized distributions do exist. These differences could be a factor of the fit in general of the hypothesized distributions, or the inability of the hypothesized distributions to capture the autocorrelation in the

frame sizes. However, the differences between the performance measures with the trace frame sizes and hypothesized (and independent) frame sizes are small compared to errors induced by assuming Poisson arrivals, as we had suspected would be the case.

From frames to packets: initial analysis

The composition of the frames is also of interest, as router transmission occurs for individual packets, rather than frames. For H.264 compression, each composite frame is composed of a video (picture) frame and approximately three embedded voice packets. Figure 8 shows a snapshot of the H.264 high bit rate packet arrivals and sizes; markers on the curve indicate individual packet arrival points. Each video frame is segmented into approximately equal size video packets with a maximum size of about 1000 bytes. The resulting video packets are transmitted to the router as a “packet train.” The average number of picture packets in a packet train is 13 packets, and the maximum is 23 packets, depending on the frame size. Voice packets are transmitted at about 22 ms periods, giving approximately three 40-byte voice packets after each video packet train. The average number of voice packets in a sequence after the picture packet train is 3.3, and the maximum is 6 packets. So we have a big batch of large picture packets followed by a much smaller batch of smaller size voice packets.

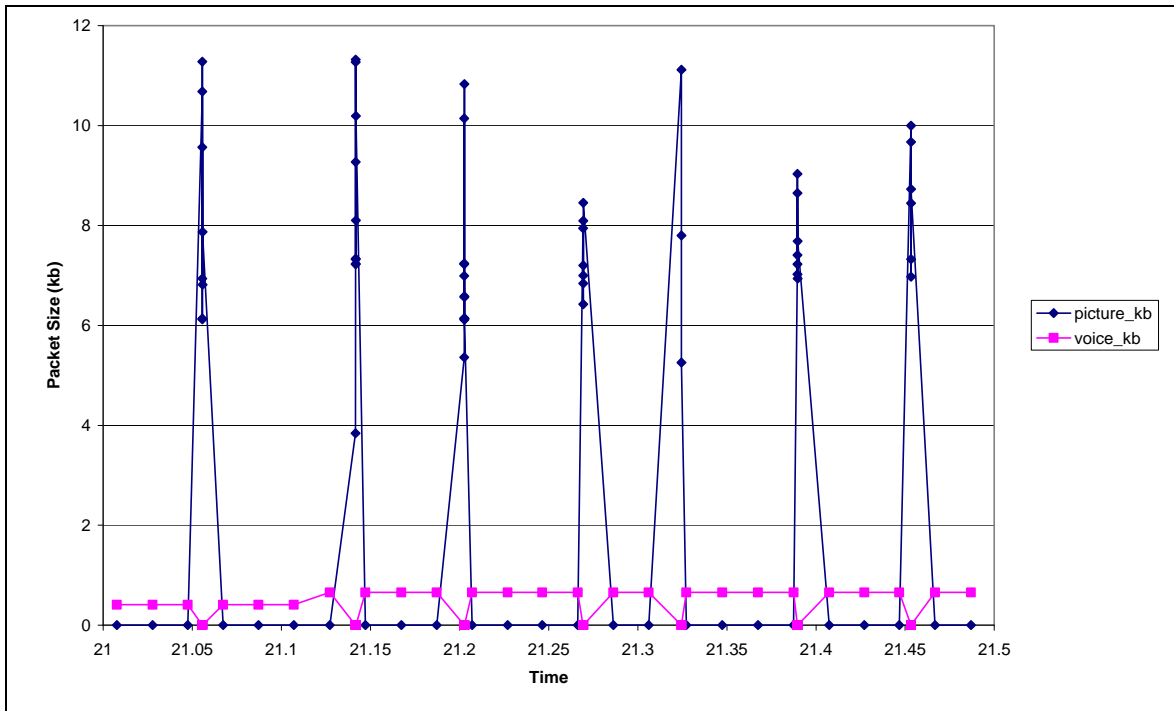


Figure 8. H.264 high bit rate packet arrivals and sizes

We can examine the sensitivity of the video performance to assumed packet characteristics. For instance, how is the packet performance impacted if we assume that packets arrive as an independent Poisson process, rather than the obviously non-Poisson arrivals depicted above? At the same time, we will also assume that packet sizes are independent. We suspected that the use of Poisson arrivals would not result in accurate estimates of packet performance, but wanted to confirm and quantify this. Another question of interest is how the packet performance was impacted if we assume that packets arrive using the batch structure shown in Figure 8, but with packet sizes based on frame sizes from the hypothesized distributions from ExpertFit?

To research this, we developed a simulation model using Lindley's recursion, which can take input from four different data sets—the original packet trace from the H.264 high bit rate data set, or three non-trace data sets. One non-trace data set again assumes independent Poisson arrivals and the voice and picture packet sizes are drawn independently from their empirical size distributions (without batches), which we suspected would not result in accurate estimates of packet performance, but we wanted to confirm and quantify this. The second non-trace data set assumes (independent) Pearson Type VI distrib-

uted frame sizes; Pearson Type VI was the best fitting frame size distribution for the H.264 high bit rate data set (see analysis in the previous sections). The third non-trace data set assumes (independent) normally-distributed frame sizes. Our goal in selecting the normally-distributed frame sizes was to determine if the use of a relatively poorly fitting frame distribution would have a significant impact on the packet congestion modeling results. Both the Pearson Type VI and normally distributed frame size data sets also assume the frame is fragmented into equally-sized picture packets, and the picture packets are sent as a packet train (e.g., with fixed interarrival times) over a Fast Ethernet connection. The number and size of picture packets is chosen for each frame so that a specified maximum packet size (e.g., 1000 bytes) is not exceeded, and all packets are of equal size. The voice packet interarrival times are also fixed (at 22 ms). Voice packet sizes are fixed at 40 bytes in both non-trace data sets. The same traffic loads are used in all scenarios.

The results of these simulations are shown in Table 4, for the case of a large buffer of 100 packets, line speed of 2000 kbps, and a load of 0.85. There is no packet loss as a result of the large buffer. We see that the use of either the Pearson Type VI-distributed or normally-distributed frames with fixed packet

Input	Wq Picture (ms)	99.9 Picture (ms)	SD(Wq) Picture (ms)	Wq Voice (ms)	99.9 Voice (ms)	SD(Wq) Voice (ms)
Trace	28.07	81.35	18.17	25.29	82.66	19.73
Pearson Type VI	26.62	76.43	16.99	24.37	76.11	18.57
Norman	26.37	71.75	16.80	24.17	69.66	18.37
Random	11.83	89.83	13.28	11.83	89.83	13.24

interarrival times by packet type approximates the trace fairly well for both the voice and picture packets. Pearson Type VI is only slightly better than normal for the 99.9 quantile of queue wait for picture and voice packets. However, the sample size from our data set is relatively small—the 9567 frames in the data set result in approximately 132,000 picture packets—so the estimates for the 99.9 quantile of queue wait may be inaccurate. The two hypothesized distributions give virtually the same results for other measures (mean queue wait and standard deviation of queue wait). Random and independent selection of packet sizes from their empirical size distribution with Poisson arrivals causes the queue waits to be underestimated versus the performance from the actual trace, as we had suspected.

A second set of models was run with the same four data sets, but now assuming a small buffer size of 10 packets, which causes packet loss to occur. The hypothesized distributions (Pearson Type VI and normal) are still fairly good for the small buffer case at approximating the trace—the mean and standard deviation of queue wait approximates trace well for these two distributions. The other non-trace data set with the random selection of packet sizes from the empirical distribution and Poisson arrivals is not a good approximation to the trace.

In summary, the use of a theoretical distribution form for frame sizes that assumes independence in the frame sizes (such as the Pearson Type VI or normal in the previous example), while preserving the batch structure and fixed interarrival times of the picture and voice packets, approximates the trace behavior fairly well, as evidenced by the close performance statistics. The use of Lindley’s equation for this simulation is of benefit because it allows one to compute congestion statistics based on an intelligently-generated data set having the desired batch structure; an analytic model or most simulations would require independent packets. The choice of the particular frame size distribution does not have a large impact on the results—use of the normal distribution, which was not ranked well by ExpertFit, gave fairly similar results as the top-ranked Pearson Type VI distribution for the H.264 hit bit rate data set. The assumption of Poisson arrival times and random/independently selected packet sizes from the empirical distribution does not approximate the packet trace well.

Summary and conclusions

Accurate input models for different traffic streams, including video, are essential in performance models of converged networks. We have seen that the best-fitting theoretical distributions to the video frame sizes differ for the three video data sets, as summarized below.

- H.263 Polycom—Weibull is most appropriate, followed by Erlang and gamma distributions; normal is much less appropriate.
- H.264 high bit rate—Pearson Type VI or Erlang are most appropriate, followed by gamma; normal is much less appropriate.
- H.264 low bit rate—normal is most suitable (although none fit very well); Erlang and gamma are in the top ten.

Ideally, due to the differences in the results, performing an analysis of the best statistical distribution for the frame sizes should be done for each data set of interest. If it is not possible to perform such an analysis on a video data set (e.g., if automated distribution fitting software like ExpertFit is not available, or due to lack of time or expertise to perform such an analysis manually), it appears that the Erlang or gamma are acceptable for all three of the data sets that we analyzed and hopefully are suitable for a wider range of video data configurations.

As our goal is to study video congestion, we want to ensure that the selected frame distribution is close enough to the actual trace data so that estimated congestion based on a hypothesized theoretical distribution is accurate. Autocorrelations in frame sizes were examined; there does appear to be low-level, although statistically significant, autocorrelations between frame sizes.

When frames are directly used in a simulation model, simulation results differ when the frame sizes from the trace are used versus frame sizes generated independently from a hypothesized distribution. The direction of the difference varies depending on the data set. Differences between simulation results for two hypothesized frame size distributions tested for

each video data set were small. However, larger differences are seen when the arrival times of the frames are assumed to be Poisson and independent. Assumptions on frame size independence have less impact on congestion estimates than do the assumptions on frame arrival time independence. Thus, it does not appear to be necessary to incorporate autocorrelation into frame size models.

The decomposition of frames into packets was also examined. The resulting packets were used in a simulation model of packet congestion. This analysis showed that use of hypothesized distributions to generate independent frame sizes can result in good approximations to actual trace data if the periodic nature of the arrivals of picture and voice packets is preserved. The choice of the hypothesized frame distribution to generate the packets had only a small impact on the final results from the congestion model.

Therefore, we conclude that if an analysis of the frame size distribution cannot be done for a data set of interest, the use of even the normally-distributed frame sizes (which had a relatively poor fit as compared to some other distributions) to generate packets can be assumed with only a small affect on accuracy of the congestion model results.

Acknowledgements

This work was partially funded by the National Communications System (NCS) Contract Number NBCH-D-02-0039 (Task Order Number D0200390095). Research funding was also provided by the Center for Network-Based Systems, a research collaboration of Noblis and George Mason University. ■

Notes and references

1. Maglaris, B. et al., "Performance Models of Statistical Multiplexing in Packet Video Communications," *IEEE Transactions on Communications*, vol. 36, pp. 834–844, July 1988; http://www.netmode.ntua.gr/courses/postgraduate/video/documents/maglari_paper.pdf.
2. Liu, X-G. and J. Babiarz, "Simulation Results for Explicit PCN Marking and Flow Termination," February 2007; http://standards.nortel.com/pcn/Simulation_EPCN.pdf.
3. Krnunz, Marwan and Herman Hughes, "A Traffic Model for MPEG-Coded VBR Streams," *Proceedings of the ACM SIGMETRICS '95 Conference*, pp. 47–55, 1995; <http://cobnitz.codeen.org:3125/citeseer.ist.psu.edu/cache/papers/cs/2253/http:zSzzSzwww.cs.umd.edu:z-richzSzcourseszScmsc710-f97zSzpaperszSzkrunz95.pdf/krunz95traffic.pdf>.
4. Koumaras, H., C. Skianis, G. Gardikis, and A. Kourtis, "Analysis of H.264 Video Encoded Traffic," *INC 2005 Fifth International Network Conference*, Samos Island, Greece, July 2005; http://aias.iit.demokritos.gr/~koumaras/Koumaras_INC2005.pdf.

5. Doggen, Jeroen and Filip Van der Schueren, MPEG4 Part 10—H.264 Modeling in OPNET, Masters Thesis, Department of Industrial Sciences and Technology, University College of Antwerpen, Belgium, July 2006; <http://users.pandora.be/jeroendoggen/OPNET/NW-Doggen.pdf>.
6. Law, A. M. and W. D. Kelton, *Simulation Modeling and Analysis*, New York: McGraw-Hill, 2nd ed., 1991.
7. Knuth, D. E., *The Art of Computer Programming, vol. 2: Seminumerical Algorithms*, 3rd ed., Reading, MA: Addison-Wesley, pp. 45–52, 1998.
8. <http://rkb.home.cern.ch/rkb/AN16pp/node143.html>.
9. Law, A. M., personal email correspondence with Denise Masi, 2007.
10. NIST Engineering Statistics Handbook; <http://www.itl.nist.gov/div898/handbook/>.
11. Gupta, R. D. and D. Kundu, "Generalized Exponential Distributions: Different Methods of Estimation," *Journal of Statistical Computation and Simulation*, vol. 69, no. 4, pp. 315–338, 2001; <http://home.iitk.ac.in/~kundu/paper62.pdf>.
12. Gross, D. and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed., John Wiley & Sons, Inc., 1998.
13. Fischer, M. J. and D. M. B. Masi, "Analyzing Internet Packet Traces Using Lindley's Recursion," *Proceedings of the 2006 Winter Simulation Conference*, L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds., 2006.

About the authors



Denise M. Bevilacqua Masi is a senior principal engineer at Noblis where her experience and research interests include queueing theory and simulation applied to telecommunications networks. She received her doctorate degree in information technology and engineering at George Mason University. Contact her at dmasi@noblis.org.



Martin J. Fischer is a senior fellow at Noblis where his experience includes network design and performance analysis in telecommunications. He has published more than 50 articles in refereed journals. He received his doctorate degree in operations research from Southern Methodist University. Contact him at mfischer@noblis.org.



David A. Garbin is a senior fellow at Noblis. His experience includes 30+ years in the telecommunications and networking field, focusing on the design and economic analysis of large networks, both for carriers and their customers. He provides quality of service in convergent Internet Protocol (IP) networks and advises government agencies in the acquisition and implementation of Voice over IP (VoIP). He holds advanced degrees from the Massachusetts Institute of Technology and is the co-author of the "New McGraw-Hill Telecom Factbook." Contact him at david.garbin@noblis.org.