

ESTIMATION OF TRAFFIC PARAMETERS IN HIGH-SPEED DATA NETWORKS

Mine Caglar, K. R. Krishnan*, and Iraj Saniee
Bellcore, New Jersey
U. S. A.

ABSTRACT

Bellcore studies have shown that Fractional Brownian Motion (FBM) is a convenient and compact mathematical representation that can account for the burstiness and long-range auto-correlation observed in the traffic of high-speed data networks. The convenience of the FBM representation for a traffic stream lies in the fact that it is described by just three parameters (m, a, H) , where m is the mean rate, a the 'peakedness' and H the Hurst parameter. Clearly, for the FBM traffic model to be applied in practice, the parameters of the model have to be estimated from the measurements that are available in the switching systems supplied by equipment vendors. On the basis of these measurements, traffic capacity management (TCM) algorithms must ensure that the provisioned capacity is sufficient to meet the anticipated traffic of broadband applications, but not so excessive as to render broadband services uneconomical.

We investigate how the estimation of the parameters of the FBM model is affected by the frequency of collection of the traffic measurements, and examine this dependency for two methods of estimation applied to two data sets: the method of variance-time plots and the recently-proposed method based on wavelets. If estimation error is judged by its impact on the engineering of link capacity, we find that both the variance-time plot method and the wavelet method offer about the same accuracy, with a slight edge for the wavelet method. Our results suggest that for high-speed traffic sources with moderate values of H , say, $0.5 \leq H \leq 0.8$, traffic measurements at 1-minute intervals might be adequate for estimation of parameters for link engineering. For traffic sources with $H > 0.85$, measurements may have to be more fine-grained, at intervals of the order of 1 second or several milliseconds. In either case, we observe that existing switch measurements at standard intervals of 15-minutes are inadequate for parameter estimation of FBM traffic models, and could lead to provisioning errors of about 10%.

**Author for correspondence*

Mine Caglar
E-mail: mcaglar@princeton.edu

K. R. Krishnan*
Bellcore
MCC 1C-353R
445 South Street
Morristown, NJ 07960-6438
Tel: 973-829-2891
Fax: 973-829-2645
E-mail: krk@bellcore.com

Iraj Saniee
Email: iis@bellcore.com

ESTIMATION OF TRAFFIC PARAMETERS IN HIGH-SPEED DATA NETWORKS

Mine Caglar, K. R. Krishnan, and Iraj Saniee
Bellcore, New Jersey
U. S. A.

1. Introduction

Bellcore studies of the traffic streams of high-speed data networks [1, 2] have shown that a considerable degree of correlation exists in the traffic bursts in such streams. Fractional Brownian Motion (FBM) has been proposed as a convenient and compact mathematical representation [3] that can account not only for the observed peakedness of such traffic, but also for the longer-than-usual auto-correlations exhibited by such traffic. The convenience of the FBM representation for a traffic stream lies in the fact that it is described by just three parameters (m, a, H) , where m stands for the mean rate, a for the ratio of variance to mean (at a pre-selected time unit) and H for the Hurst parameter. However, for the FBM traffic model to be applied in practice, the parameters of the model have to be estimated from the actual measurements that are publicly available, such as those measured and stored in the switching systems supplied by equipment vendors. On the basis of the available measurements, traffic capacity management (TCM) algorithms must ensure that the provisioned capacity is sufficient to meet anticipated demand for broadband (*e.g.*, ATM and Frame Relay) applications, but not so excessive as to render broadband services uneconomical. This task of capacity management has proved to be considerably harder in data networks than in networks for voice traffic.

Traffic measurements are collected not only to assess the current network performance and detect violations of quality of service, but also to observe the traffic trends and estimate the load for use in network engineering. One of the most critical steps in developing comprehensive TCM methods is to define the required traffic measurements (statistics). Therefore, algorithms must be developed both to estimate critical performance measures from the collected traffic measurements and to determine the parameters that describe the observed traffic. The traffic model derived from the measurements will be the basis of capacity management algorithms that determine the network capacity needed to carry current and future demands.

In this paper, we concentrate on the problem of estimating the parameters for the FBM traffic model that has been proposed for ATM and Frame Relay traffic, using available traffic measurements. Clearly, the effectiveness of traffic model estimation depends both on the intrinsic richness of the available measurements and on the accuracy of the estimation procedure that is used. Thus, an investigation of this problem has two dimensions: 1) the

sufficiency or insufficiency of the available traffic measurements for the estimation procedure and 2) a robust estimation procedure for the critical parameters. To assert, for example, that vendor-supplied equipment should provide more frequent measurements of traffic flow on a link, one has to make an economic case for the need for the finer granularity.

We investigate how the estimation of the FBM model parameters for traffic is affected by the frequency of collection of traffic measurements, and examine this dependency for two methods of FBM parameter estimation applied to two data sets: the method of variance-time plots [2] and the recently-proposed method based on wavelets [4,5]. If we judge estimation error by its impact on the engineering of link capacity, we find that both the variance-time plot method and the wavelet method offer about the same accuracy, with a slight edge for the wavelet method. Our results suggest that for high-speed traffic sources with moderate values of H , say, $0.5 \leq H \leq 0.8$, traffic measurements at 1-minute intervals might be adequate for estimation of parameters for link engineering. For traffic sources with $H > 0.85$, measurements may have to be more fine-grained, at intervals of the order 1 second or several milliseconds. In either case, we observe that existing switch measurements at standard intervals of 15-minutes are inadequate for parameter estimation of FBM traffic models, and could lead to provisioning errors of about 10%.

An outline of this paper is as follows. In Section 2, we describe the FBM model used in this study and discuss the assumptions made regarding the availability of traffic measurements. In Section 3, the variance-time plot method of estimating the FBM model parameters from the available traffic measurements is described. Also, the basis for the statistical analysis of parameter estimates derived from the traffic measurements is discussed. The method of estimation by wavelet analysis is described in Section 4. The results of the application of these methods to two data sets are analyzed in Section 5, and the engineering impact of estimation accuracy is investigated. Finally, our conclusions are stated in Section 6.

2. Traffic Model and Traffic Measurements

2.1 FBM Traffic Model

The FBM model for data traffic consists of the three parameters (m, a, H) , where

m = mean traffic rate, say, in bits/ time-unit

a = 'peakedness' factor = $\frac{\text{variance of traffic arrivals over time - unit}}{\text{mean traffic rate over time - unit}}$

H = Hurst parameter (a dimensionless measure of the persistence of correlation in data-rate, with $0.5 \leq H < 1$).

With just these three parameters, the model expresses the burstiness of data over a wide range of time-scales by its intrinsic *time-scaling* properties [1,2]. Thus, in principle, burstiness over a single time-scale determines the model

parameters, and thus the burstiness over *all* time-scales. However, in fitting an actual data set to such an FBM model, it is necessary to assure ourselves that the FBM model, is, in fact, a good fit to the data. This requires us to examine the data over different time-scales, and then, we have the problem of reconciling several sets of FBM parameters, each set obtained by considering the data at a particular granularity of time intervals.

The question for us to consider is “What granularity of traffic measurements is necessary for ATM and Frame-Relay traffic for the reliable estimation of the parameters of the FBM model corresponding to such traffic sources”? In particular, existing switches that support ATM and Frame-Relay offer only a limited set of *infrequent* traffic measurements, e.g., at intervals of 15 minutes. Are such ‘aggregate’ measurements of any value for estimating the FBM parameters m , a , and H with sufficient accuracy for purposes of link engineering?

We examine this question by analyzing two data sets of Frame-Relay and Ethernet traffic collected by Bellcore for previous studies [2]. We compare the FBM parameters estimated from the original data-sequence with those estimated from aggregated samples of the same sequence, to see what accuracy and precision are lost when working with infrequent measurements. It is known [2] that the Hurst parameter H can be reliably estimated only from long sequences, i.e., from frequent traffic measurements within a duration (say, an hour) over which the traffic may be considered stationary. It is shown that while the mean rate m may be estimated from aggregated measurements, the peakedness factor a is subject to considerable ‘sampling’ error, especially when associated with a value of H close to 1. In order to judge the *engineering* consequence of the estimation error, we determine, on the basis of the parameter estimates obtained from different levels of data-aggregation, the number of independent traffic sources that can be carried on a link, for given service criteria of delay and loss. We conclude from our analysis of these two data sets that traffic measurements at intervals of one minute may be acceptable in cases when H is no greater than about 0.8. However, measurements at intervals of the order of one second or several milliseconds may be necessary for traffic sources with $H > 0.85$. In any case, existing switch measurements at intervals of 15-minutes are quite inadequate for estimation of FBM traffic models for high-speed data networks. Whether there are other traffic models that lend themselves to reliable estimation from such aggregated traffic measurements is a question of considerable interest, and has to be investigated.

2.2 Traffic Measurements

Bellcore has issued several documents in which generic requirements for broadband traffic statistics are specified [GR-1114-CORE, GR-110-CORE, TA-NWT-001248]; however, in practice, few switch vendors are meeting such requirements. We state here the assumptions made in this paper about traffic measurements available from various switching systems.

Traffic *counts* are the basic measurements assumed to be available at regular intervals. Although other measurements, such as buffer levels and packet losses, may offer additional information for constructing traffic

models, we shall only consider traffic counts here. These counts may be available only at intervals of, say, 15 minutes, as a rule; occasionally, for ‘special studies’, one may be able to arrange for more frequent measurements, say, at 1-second intervals, but these cannot always be assumed to be *routinely* available.

The natural question regarding traffic measurements is “How often should they be collected?” The answer to this question, of course, depends on the use to which the measurements are put. Our assumption here is that the traffic measurements are used *to estimate the parameters of the mathematical model* chosen as a description of the characteristics of the traffic. Presumably, more frequent measurements will yield a higher accuracy in estimating the parameters of the traffic model, but what accuracy is it necessary to achieve? We judge the consequence of ‘inaccurate’ traffic estimation (due to infrequent measurements) by the resulting inaccuracy in the estimation of the effective bandwidth needed to carry the traffic (for a given quality of service). Thus, our plan is to begin with a sequence of fine-grained traffic measurements (obtained, let us say, by ‘special studies’), and then consider various levels of *aggregation* of this data sequence, to correspond to the data that would be obtained from less frequent measurements of the same traffic stream. We carry out the estimation of the parameters of the traffic model (the Fractional Brownian Motion model in this paper) both from the fine-grained measurements and from the more aggregated measurements. We then determine the number of independent traffic sources that can be carried on a link, (for given service criteria of delay and loss) from these various parameter-sets, and consider the magnitude of the discrepancies with respect to the results from the fine-grained measurements.

Of the three parameters (m , a , H) in the FBM traffic model, m , the stationary mean rate of the traffic, say, over an hour, can be estimated directly from the available measurements, whether fine-grained or aggregated. The parameters a and H are then *jointly* estimated by either of the two methods described below. However, H cannot be reliably estimated from very small sets of aggregated samples, e.g., on the basis of four 15-minute counts in an hour. We could assume that, in this case, it is possible to supplement the routine traffic measurements by more detailed measurements (‘special studies’ referred to above) for a limited period, such as a selected hour once a week, by means of Network Management Systems, such as SNMP (Simple Network Management Protocol). Such Network Management queries are not a suitable mechanism for the regular collection of traffic measurements, owing to the large demands placed on switch processing; however, it may not be a problem to invoke the mechanism, say, for a chosen hour once a week. From such traffic counts at a fine resolution (intervals of 1-second or smaller), we can estimate the Hurst parameter H by either of the two methods below. The rationale is that the Hurst parameter, which describes the correlation in the traffic, is a stable and intrinsic characteristic of the services (video, data, etc.) that give rise to the traffic, and not a strong function of the hourly or daily variations in the traffic rate. The estimate of H derived from the detailed measurements could then be used to obtain an estimate of the peakedness parameter a from less frequent measurements. The numerical value of a depends on the time-unit used; the following formula provides for the conversion of its value from one time-unit to another:

$$a_{(\text{for } T \text{ units of time})} = T^{2H-1} * a_{(\text{for one unit of time})}; \quad (1)$$

i.e., when the unit of time is multiplied by T , the peakedness factor is multiplied by T^{2H-1} .

3. Estimation of FBM Parameters by Variance-Time Plot

Consider the sequence of traffic counts $\{x_k, k = 1, \dots, N\}$, where x_k is the number of traffic units that arrived in the k 'th measurement period, and N is the total number of measurement samples. To construct the variance-time plot for a long data-sequence, we consider the sequence formed by the aggregation of the traffic arrivals in n consecutive measurement periods, for various values of n . Thus, the n -fold aggregated sequence $\{y_j^{(n)}\}$ is defined by

$$y_j^{(n)} = \sum_{i=(j-1)n+1}^{jn} x_i, \quad j = 1, \dots, M_n, \text{ where} \quad (2)$$

$$M_n = \left\lfloor \frac{N}{n} \right\rfloor, \quad n = 1, 2, \dots, N$$

Of course, $n = 1$ corresponds to the original sequence itself, with $M_1 = N$.

3.1 Estimation of m

The mean arrival rate, estimated from the original unaggregated sequence, is given by

$$\hat{m} = \frac{\sum_{k=1}^N x_k}{N} \quad (3)$$

where the time-unit used is the sampling interval of the original sequence. The *same* estimate is obtained even after aggregation of data, after conversion to the same time-unit as in (3), *as long as no data are ignored*, i.e., when $N = nM_n$, since

$$\hat{m} = \frac{\sum_{k=1}^N x_k}{N} = \frac{\sum_{j=1}^{M_n} \sum_{i=(j-1)n+1}^{jn} x_i}{nM_n} = \frac{1}{n} \left\{ \frac{\sum_{j=1}^{M_n} y_j^{(n)}}{M_n} \right\} \quad (4)$$

Thus, the estimate of the mean is essentially unaffected by aggregation of data, and *we can attach to it the confidence interval corresponding to the original unaggregated sequence*.

3.2 Estimation of a and H

We calculate, for several values of n , the *sample* variance $\hat{V}(n)$ of the n -fold aggregated sequence, considering, in practice, only those values of n for which there are at least 4 members in the sequence. For an FBM source with parameters (m, a, H) , the plot of $\log V(n)$ versus $\log n$, where $V(n)$ is the *true* variance of the n -fold aggregated process, is a straight line with slope $S = 2H$, with an ordinate-axis intercept of $C = \log(ma)$. Thus, we may use linear regression on the set of points $\{\log \hat{V}(n), \log n\}$ to estimate both the slope S and the ordinate-axis intercept C of the curve, and obtain an estimate of H and of the peakedness parameter a (knowing the sample mean \hat{m} of the sequence); i.e.,

$$\hat{H} = \frac{S}{2}; \quad \hat{a} = \frac{e^C}{\hat{m}} \quad (5)$$

3.3 Sources of Error in Parameter Estimation from Variance-Time Plots

The chief source of error is in the approximation of $V(n)$ by $\hat{V}(n)$; the latter is calculated from the samples in the n -fold aggregated sequence $\{y_j^{(n)}\}$. Since adjacent the members of this sequence correspond to aggregations of adjacent blocks of length n of the original sequence, it follows that the samples of $\{y_j^{(n)}\}$ can be *correlated* with each other. In this case, $\hat{V}(n)$ can be a *biased* estimate of $V(n)$. Beran [6] shows that the correction for the bias is the multiplicative factor

$$\frac{M_n - 1}{M_n - (M_n)^{2H-1}} \quad (6)$$

to be applied to $\hat{V}(n)$. However, the correction (6) itself depends on the Hurst parameter H , which is one of the parameters to be estimated by this method and unknown to begin with. Thus, a proper application of the variance-time plot would require an iterative procedure. An estimate of H obtained from (5) is used to derive the multiplicative correction factor (6) to be applied to the sample variances $\{\hat{V}(n)\}$, from which a new estimate of H is derived by the linear regression formula (5), and so on, until convergence is attained.

We also note that the correction factor (6) gets large for small N , and becomes unbounded as H tends to 1, i.e., for a highly correlated data-sequence. For this reason, one expects that estimation errors in H become magnified in the estimation of the peakedness parameter *when H is close to 1*.

3.4 Statistical Analysis of Parameter Estimates

3.4.2 The estimate of mean m

It has been noted above that the original data sequence and its various aggregated versions yield the same estimate \hat{m} for the mean m . Thus, the accuracy of \hat{m} is unaffected by data-aggregation, and we can assign to it the confidence interval corresponding to the unaggregated data-sequence. Thus, the 95% confidence interval for the mean is given by

$$\hat{m} - 1.96 \frac{\hat{s}}{\sqrt{N}} < m < \hat{m} + 1.96 \frac{\hat{s}}{\sqrt{N}} \quad (7),$$

where \hat{s} is the sample estimate of the standard deviation given by (10) below, and N is the number of samples in the original sequence.

3.4.2 The estimates of H and a

In the method of variance-time plots, the linear regression calculations use Student's t distribution, with $(M - 2)$ degrees of freedom where M is the number of points (i.e., the number of aggregation levels considered for the original sequence of traffic measurements), to determine the 95% confidence limits for the slope S , and thus for the Hurst parameter H . In the presence of long-range correlation ($H > 0.5$), the actual degrees of freedom are fewer, and, in fact, depend on the H being estimated. We ignore this complication and accept the confidence interval given by the linear regression.

Linear regression also furnishes an estimate and confidence interval for a , which, however, is based on the assumption of independence of the traffic counts. We proceed as follows to derive an estimate and approximate confidence interval for a from the data for different levels of aggregation, accounting for correlation in an approximate way.

Suppose that, for a given level of aggregation, we have the stationary data sequence $\{w_k, k = 1, \dots, M\}$, with mean m and variance σ^2 . The sample mean \hat{m} and sample variance s^2 are given by

$$\hat{m} = \frac{\sum_{k=1}^M w_k}{M} \quad (8)$$

$$s^2 = \frac{\sum_{k=1}^M (w_k - \hat{m})^2}{M - 1} \quad (9)$$

In the presence of correlation among the $\{w_k\}$, however, s^2 is a biased estimate of σ^2 . An unbiased estimate \hat{s}^2 is given by [6]

$$\hat{s}^2 = s^2 \frac{M - 1}{M - M^{2H-1}} \quad (10),$$

and the corresponding estimate of a is given by

$$\hat{a} = \frac{\hat{s}^2}{\hat{m}} \quad (11).$$

We now use the *approximation* that $\left[\frac{\sum_{k=1}^M (w_k - \hat{m})^2}{\sigma^2} \right]$ is described by the chi-square distribution with the number

of degrees of freedom given by the *integer* part of $(M - M^{2H-1})$. Thus, if χ_p^2 is the p' th percentile of the chi-square distribution, the 95% confidence interval for a is given by

$$\frac{(M - 1)\hat{a}}{\chi_{.975}^2} < a < \frac{(M - 1)\hat{a}}{\chi_{.025}^2} \quad (12)$$

The estimate \hat{a} and its confidence interval can be transformed from one time-unit to another by formula (1).

4. Wavelet Estimation Method

The wavelet estimation method is similar to the variance-time method in that the data stream is analyzed in a succession of resolution of scales. However, while the variance-time analysis is in the time domain, the wavelet method is in the frequency domain and thus has the following superior statistical properties: 1) The discrete wavelet transformation is applied to the traffic stream to get the associated *details* in the frequency domain. 2) The details are quasi-decorrelated although the original stream could be highly correlated because of long-range dependence. 3) Any linear or polynomial trends in the data are automatically factored out of the wavelet analysis. 4) The wavelet estimation method constructed on these details has low variance and high robustness.

We consider the joint estimation of H and a by the wavelet estimation method as given in [4]. Let $d(j, k)$, $k = 1, \dots, n_j$, $j = 1, \dots, J$ denote the details obtained by the discrete wavelet transform of the sequence of

traffic counts $\{x_k, k = 1, \dots, N\}$, where J is such that $2^{J+1} \leq N \leq 2^{J+2}$, and n_j is the number of coefficients available at octave j . The statistic central to the method is given by

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} d^2(j, k)$$

$j = 1, \dots, J$. Then, the Hurst parameter H and the coefficient c_f of the spectrum of the data stream are estimated through a weighted linear regression of

$$y_j = \log_2(\mu_j) - g_j \quad (13)$$

over $j = j_1, \dots, j_2$ where j_1 and j_2 are the scales relevant for long-range dependence. The constant $g_j = E(\log_2(\mu_j)) - j(2H - 1) - \log_2(c_f C)$ is introduced to ensure that the fundamental hypothesis of regression holds (with C being a constant that depends on H). Then, the slope α of the regression line is equal to $2H - 1$ and its intercept β is $\log_2 B$, where $B = c_f C$. Veitch and Abry [4] devise unbiased estimators for $2H - 1$ and $c_f C$ under mild assumptions, which effectively hold for FBM. The estimate of H is obtained by

$$\hat{H} = \frac{\alpha + 1}{2}$$

which is unbiased and consistent. The estimator of $B = c_f C$ is given by

$$\hat{B} = p \cdot 2^\beta$$

where p is a constant to ensure unbiasedness of \hat{B} . We now set

$$\hat{c}_f = \frac{\hat{B}}{\hat{C}}$$

where \hat{C} is computed using \hat{H} . The estimate \hat{c}_f is asymptotically unbiased and consistent. Both estimators have small variances close to the Cramer-Rao lower bound. Using the relationship of c_f to the variance of the traffic counts for larger scales, we obtain the estimate of the peakedness parameter a by

$$\hat{a} = \frac{\hat{c}_f (2\pi)^{2\hat{H}-1}}{2\hat{m}\Gamma(2\hat{H}-1)\cos((2\hat{H}-1)\pi/2)(2\hat{H}-1)\hat{H}} \quad (14)$$

where \hat{m} is the estimate of the mean of the traffic counts as before. As an involved function of \hat{c}_f and \hat{H} , it is not known how this estimate of a behaves. Empirical studies show that it could be slightly biased with acceptable variance [5]. We approximate a confidence interval for a through that of c_f , assuming that \hat{H} is known in Equation (14).

The wavelet estimator of H in general performs very well as expected. Other simulation studies [7,8] have also verified this. The performance of the variance-time type of analysis in these studies is found to be better than some other analysis methods. In particular, Jennane et al [9] report good performance. The wavelet method assumes that the length of the data set is a power of 2, which can lead to the exclusion of some available information. Besides, when the data set is too short, there may be too few scales for regression with this method. In these cases the variance-time method may be used.

5. Results

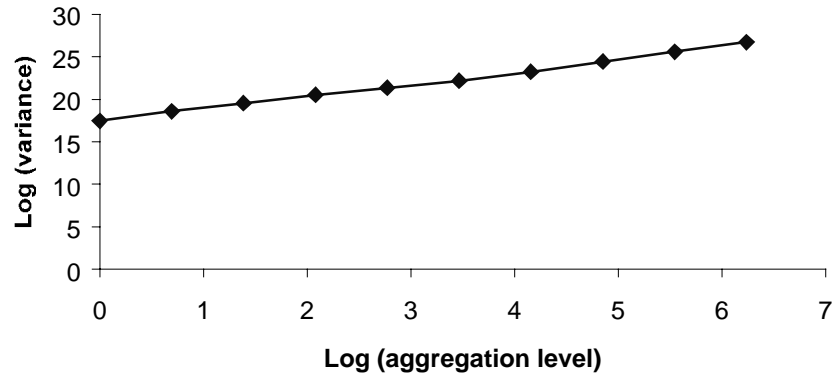
We consider two data sets:

- a) **ONESEC**: This consists of 3600 measurements (traffic counts in bytes) on a Frame Relay link, at one-second intervals, for a total duration of one hour
- b) **40MSEC**: This consists of 24000 measurements (traffic counts in bytes) on an ethernet, at 40-millisecond intervals, for a total duration of 16 minutes

5.1 Variance-Time Plot Results

ONESEC

The variance-time plot for the original data set, with variance calculated at several aggregation levels, is shown in Figure 1. This plot allows us to estimate H on the basis of the entire sequence of 3600 one-second measurements.

Figure 1: Variance-Time Plot of ONESEC

In addition to the original sequence, we then consider two levels of aggregation — up to one-minute intervals (60 samples), and 15-minute intervals (4 samples). However, H cannot be estimated from just 4 samples of 15-minute aggregation of traffic counts; hence, we *assign* to this case the value of H estimated from the original 3600-sample sequence, in order to calculate the confidence interval for a . The estimates of m, a, H and their 95% confidence intervals are indexed by the number of samples below:

$$595 < m_{3600} = m_{60} = m_4 = 800 < 1005 \text{ bytes/sec;}$$

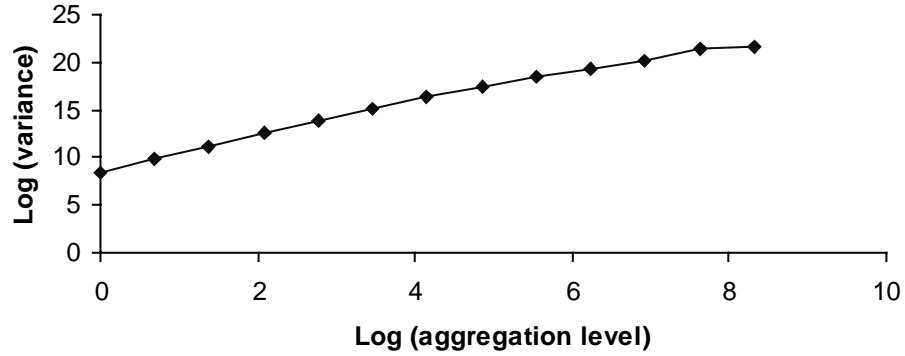
$$0.69 < H_{3600} = 0.72 < 0.76; \quad 228 < a_{3600} = 235 < 250 \text{ byte - msec}$$

$$0.77 < H_{60} = 0.84 < 0.90; \quad 32 < a_{60} = 34 < 74 \text{ byte - msec}$$

$$H_4 \text{ assigned to be } H_{3600}; \quad 214 < a_4 = 528 < 31304 \text{ byte - msec}$$

40MSEC

The variance-time plot for this data set is shown in Figure 2.

Figure 2: Variance-Time Plot of 40MSEC

In this case, along with the original sequence of 24000 measurements at 40-millisecond intervals, we also consider two levels of aggregation: one-second intervals (960 samples), and one-minute intervals (16 samples). For the 16-sample case, H is very close to 1, and the actual confidence interval obtained from the regression violates the feasible range (0.5, 1) and is replaced by the feasible range. Also, as pointed out in Section 3.3, Beran's correction (6) becomes very large for this value of H . Our approximations regarding degrees of freedom in the chi-square distribution no longer make sense, and we set the degrees of freedom to be $M - 1 = 15$ for the calculation of a for this case. The results are indexed by the number of samples below:

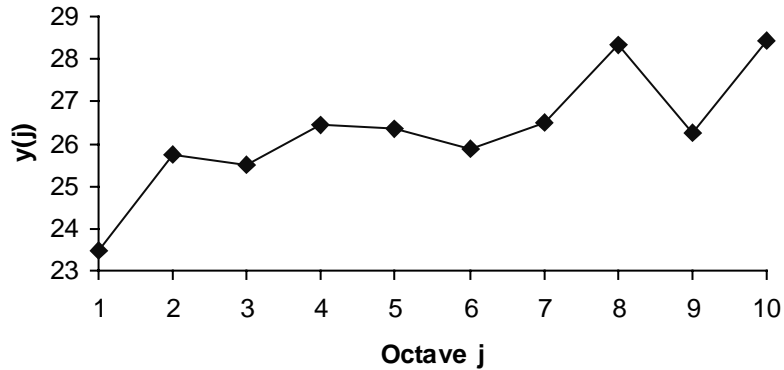
$$\begin{aligned}
3187 < m_{24000} = m_{960} = m_{16} < 3188 < 3189 \text{ bytes/sec;} \\
0.76 < H_{24000} = 0.81 < 0.87; & \quad 3.68 < a_{24000} = 3.71 < 3.83 \text{ byte - msec} \\
0.69 < H_{960} = 0.73 < 0.77; & \quad 28.79 < a_{960} = 31 < 34.48 \text{ byte - msec} \\
0.5 < H_{16} = 0.94 < 1.0 & \quad ; \quad 0.54 < a_{16} = 0.98 < 2.35 \text{ byte - msec}
\end{aligned}$$

5.2 Wavelet Results

ONESEC

We first consider the entire data set at one-second intervals. There are 3600 in total, but we make use of only the first $2048 = 2^{11}$ of these because of the nature of the wavelet method. Although this means disregarding some of the available data, our analysis of different segments of the stream has shown similar results to those obtained with the first 2048. This confirms the stationarity of the traffic within the one-hour period. After discrete wavelet transform of the data, we plot the points y_j of (13) versus scales $j = 1, \dots, 10$ in Figure 3.

Figure 3: Regression plot for one-second data



The wavelet method relies on a linear relationship of y_j versus j for higher scales j_1, \dots, j_2 . It is evident from Figure 3 that the linear relationship starts after octave 2. We take $j_1 = 3$ and $j_2 = 8$, excluding octaves 9 and 10 as n_j , the number of available details, is small for these octaves. The results are

$$0.59 < H_{2048} = 0.67 < 0.74$$

$$338 < a_{2048} = 505 < 792 \text{ byte-msec}$$

where the ranges denote the approximate 95% confidence intervals.

Second, we consider the aggregated measurements at one minute intervals. Then, there are only 60 data points of which only 32 can be used in the wavelet method. This leaves us with at most 4 octaves for regression. Using octaves 2, ..., 4, we obtain the following results

$$0 < H_{32} = 0.61 < 1$$

$$118 < a_{32} = 547 < 58940 \text{ byte-msec}$$

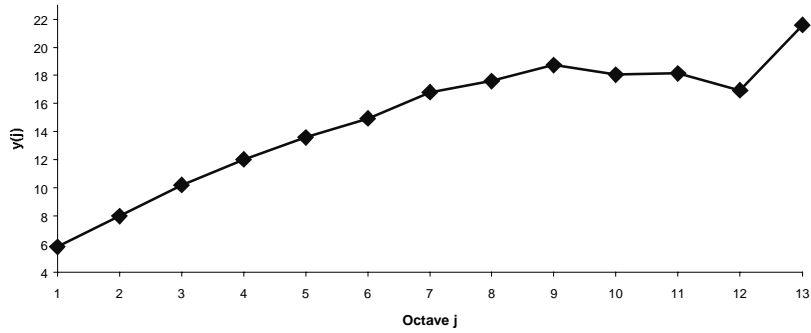
where the actual confidence interval for H is not reported as it is out of the definition range (0,1). The confidence interval for a is equally huge and not informative.

As for the highest aggregation of 15 minute intervals, there are only 4 points left. Having only 1 octave in this case, regression is not possible and we cannot analyze it with the wavelet estimation method.

40MSEC

The entire length of this data set is 24000 measurements at 40 millisecond intervals for a total duration of 16 minutes. The wavelet method makes use of the first $16384 = 2^{14}$ for estimation. The plot of the y_j of (13) versus scales $j = 1, \dots, 13$ is given in Figure 4.

Figure 4: Regression plot for 40-millisecond data



In this plot, we take $j_1 = 7$ and $j_2 = 11$. The results are

$$0.66 < H_{16384} = 0.78 < 0.90$$

$$6.1 < a_{16384} = 17.9 < 83.7 \text{ byte-msec}$$

Again, the last two octaves are disregarded because of low number of details available.

Next, we consider one-second aggregation of the data, which results in 960 samples. We use only $512 = 2^9$ of these and get

$$0.64 < H_{512} = 0.81 < 0.99$$

$$5.5 < a_{512} = 12.7 < 37.1 \text{ byte-msec}$$

which result is close to the previous one. Here, we take $j_1 = 3$ and $j_2 = 7$ disregarding only octave 8 among higher scales.

The highest aggregation level is at one-minute intervals, which results in 16 data points. Using all of these, we have 3 scales available. The use of octaves 2 and 3 only violates a technical condition of the wavelet estimation in this particular situation. So, we consider all octaves 1,...,3 in regression to get

$$0.08 < H_{16} = 0.81 < 1$$

$$0.1 < a_{16} = 0.46 < 5.4 \text{ byte-msec}$$

where the confidence interval for a does not even intersect with the previous ones. The confidence interval for H is very large although its point estimate is close to the previous estimates.

5.3 Comparison of Parameter Estimates from the Two Methods

The estimates for H and a from the variance-time plots and wavelet analysis are shown in the table below:

Table 1: Estimates of FBM Parameters from Sampled Traffic Measurements

Parameter		ONESEC			40MSEC		
		Sampling Interval			Sampling Interval		
		1-sec	1-min	15-min	40-msec	1-sec	1-min
H	Variance-Time Plot	0.72	0.81	?	0.81	0.73	0.94
	Wavelet Analysis	0.67	0.61	?	0.78	0.81	0.81
a (byte-msec)	Variance-Time Plot	235	34	?	3.7	31.0	0.98
	Wavelet Analysis	505	547	?	17.9	12.7	0.46

5.4 Link-Engineering from Sampled Measurements

To investigate the *engineering* consequence of relying on the various sampling rates, we calculate the number of (independent) traffic sources that can be carried on a T1 link (1544 kbits/sec), using the parameter estimates (m , a , H) of each sampled sequence corresponding to each data set. We used a delay criterion of 10 milliseconds, a bit loss-rate criterion of 0.0001, and a buffer of 500 bytes, and assumed a peak-rate equal to twice the mean rate for

each source. It should be noted that although the FBM model does not consider a finite ‘peak rate’ in its traffic description, the engineering algorithm that is used [10] does take account of a user-specified peak rate for the source. The results are shown in Table 2 below:

Table 2: Estimated Number of Sources Supported on a T1 Link

Estimation Method	Number of ONESEC sources			Number of 40MSEC sources		
	Sampling Interval			Sampling Interval		
	1-sec	1-min	15-min	40-msec	1-sec	1-min
Variance-Time Plot	190	193	?	57	54	53
Wavelet Analysis	190	190	?	53	53	59

It is reasonable to take the result corresponding to the smallest sampling interval as the ‘correct’ (most reliable) number of sources for each data set. Then, for ONESEC, we see that the wavelet method is *insensitive* to the aggregation of 1-second measurements to 1-minute samples, while the variance-time plot method incurs an engineering error of about 1.5%.

For 40MSEC, the wavelet method is once again *insensitive* to the aggregation of 40-millisecond measurements to 1-second samples, whereas the variance-time plot method causes an engineering error of about 5%. With 1-minute samples, both methods produce an engineering error of about 7%.

Thus, for *purposes of link engineering by the method discussed in [10]*, both the variance-time plot method and the wavelet method achieve, on the whole, about the same accuracy of parameter estimation from the traffic measurements, with a slight edge for the wavelet method.

6. Conclusions and Further Work

On the basis of the analysis of the above two data sets, we find that for high-speed data traffic sources with moderate values of H , say $0.5 \leq H \leq 0.8$, measurements at 1-minute intervals would be adequate for estimation of parameters for link engineering. For traffic sources with $0.85 \leq H < 1.0$, measurements may have to be more fine-grained, at intervals of the order 1 second or several milliseconds. In either case, we observe that existing switch measurements at standard intervals of 15-minutes are quite inadequate for parameter estimation of FBM traffic models for high-speed data networks, and could lead to significant errors in provisioning. We also conclude that both the variance-time plot method and the wavelet method for the estimation of FBM parameters offer comparable accuracies for link engineering, with a slight edge for the wavelet method

A broader question to ask is what class of traffic models is best suited for the development successful methods for traffic capacity management (which comprises both traffic estimation and link engineering). The traffic models in the literature [11] can be divided into two categories:

- (a) models which exhibit long range dependence (such as FBM, limiting aggregates of on/off models with heavy-tailed distributions for the on/off durations, and M/ Pareto / ∞ models), and
- (b) Markovian models that exhibit only short-range dependence (such as on/off models with exponential on/off distributions, Markov-modulated Poisson process, and Gaussian auto-regressive models, which typically have exponentially-decaying correlation functions).

While Markovian models are more amenable to queuing analysis than models with long-range dependence, it appears that the latter models better capture the statistical characteristics of high-speed data traffic. It may, therefore, be useful to consider Markovian models (which are short-range dependent in the mathematical sense, with asymptotic exponential decay in the auto-correlation function) with slow enough decay to capture the strong auto-correlation *over the range of time-scales that matter for the system being studied*. We expect that although Markovian traffic models lend themselves to more reliable parameter estimation than do FBM models, there would be *many* more parameters to estimate, one for each of the on and off periods. Whether the corresponding engineering methods are as robust to errors in parameter estimation as the FBM model is a question that deserves further investigation.

REFERENCES

1. W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson, On the Self-Similar Nature of Ethernet Traffic, *Statistical Science*, 10, pp. 67-85, 1995.
2. J. Beran, R. Sherman, M. Taqqu, and W. Willinger, Long-Range Dependence in Variable-Bit-Rate Video Traffic, *IEEE Transactions on Communications*, 43, pp. 1566-1579, 1995.
3. I. Norros, A Storage Model with Self-Similar Input, *Queueing Systems*, 16, pp. 387-396, 1994.
4. D. Veitch and P. Abry, A Wavelet Based Joint Estimator of the Parameters of Long-Range Dependence, submitted to special issue of IEEE Trans. Information Theory, on "Multiscale Statistical Signal Analysis and its Applications", 1997.
5. M. Caglar, Simulation of Fractional Brownian Motion with Micropulses, submitted to *IEEE/ACM Trans. Networking*, 1998.
6. J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, 1994 (p. 156)
7. R. Fischer and M. Akay, A Comparison of Analytical Methods for the Study of Fractional Brownian Motion, *Annals of Biomedical Engineering*, vol. 24, pp. 537-543, 1996.
8. A. R. Mehrabi, H. Rassamdana, and M. Sahimi, Characterization of Long-Range Correlations in Complex Distributions and Profiles, *Physical Review E*, vol. 56, pp. 712-22, 1997.
9. R. Jennane, R. Harba, and G. Jacquet, Quality of Synthesis and Analysis Methods for Fractional Brownian Motion, Proc. 1996 IEEE Digital Signal Processing Workshop.
10. K. R. Krishnan, A. L. Neidhardt, and A. Erramilli, Scaling Analysis in Traffic Management of Self-Similar Processes, Proc. 15th International Teletraffic Congress, Washington, D.C., pp. 1087-1096, 1997.
11. N. Fonesca and M. Zukerman, Tutorial on ATM dimensioning and traffic integration — management and modeling, IEEE Globecom'97, Phoenix, Arizona, November 1997.