5-th International Symposium on Digital Signal Processing for Communication Systems, DSPCS'99, Perth, 1999

Network Traffic Modeling using a Multifractal Wavelet Model

*Matthew S. Crouse, Rudolf H. Riedi, Vinay J. Ribeiro and Richard G. Baraniuk**

Department of Electrical and Computer Engineering

Rice University

6100 South Main Street

Houston, TX 77005, USA

Abstract

In this paper, we describe a new multiscale model for characterizing positive-valued and long-range dependent data. The model uses the Haar wavelet transform and puts a constraint on the wavelet coefficients to guarantee positivity, which results in a swift $O(N)$ algorithm to synthesize N-point data sets. We elucidate our model's ability to capture the covariance structure of real data, study its multifractal properties, and derive a scheme for matching it to real data observations. We demonstrate the model's utility by applying it to network traffic synthesis. The flexibility and accuracy of the model and fitting procedure result in a close match to the real data statistics (variance-time plots) and queuing behavior.

1 Introduction

Fractals models arise frequently in a variety of scientific disciplines, such as physics, chemistry, astronomy, and biology. In DSP, fractals have long proven useful for applications such as computer graphics and texture modeling [1]. More recently, fractal models have had a major impact on the analysis of data communication networks such as the Internet. In their landmark paper [2], Leland et al. demonstrated that network traffic exhibits fractal properties such as self-similarity, "burstiness," and long-range dependence (LRD) that are inadequately described by classical traffic models. Characterization of these fractal properties, particularly LRD, has provided exciting new insights into network behavior and performance.

Fractals are geometric objects that exhibit an irregular structure at all resolutions. Most fractals are *self-similar*; if we "zoom" (in or out) of the fractal, we obtain a picture similar to the original. Deterministic fractals usually have a highly specific structure that can be constructed through a few simple steps. Real-world phenomena can rarely be described using such simple models. Nevertheless, "similarity on all scales" can hold in a statistical sense, leading to the notion of random fractals.

As the pre-eminent random fractal model, fractional Brownian motion (fBm) has played a central rôle in many fields [2, 3]. FBm is the unique Gaussian process with stationary increments and the following scaling property for all $a > 0$

$$B(at) \stackrel{fd}{=} a^H B(t), \tag{1}$$

with the equality in (finite-dimensional) distribution. The parameter H, $0 < H < 1$, is known as the *Hurst parameter*. It rules the LRD of fBm, as we will see later, but it also governs its local "spikiness." In particular, for all t

$$B(t+s) - B(t) \simeq s^H, \tag{2}$$

meaning that, for $0 < H < 1$, fBm has "infinite slope" everywhere.

The statistical self-similarity (1) of fBm has proven most useful for signal modeling, since it efficiently captures signal features such as burstiness and LRD, while still allowing tractable theoretical analysis [2]. Nevertheless, models based on fBm can be too restrictive to adequately characterize many types of signals.

*This work was supported by the National Science Foundation, grant no. MIP–9457438, and by DARPA/AFOSR, grant no. F49620-97-1-0513.
Email: {mcrouse, riedi, vinay, richb}@rice.edu. URL: www.dsp.rice.edu.

First, strictly self-similar scaling behavior is not always realistic. For example, H in (1) may vary when measured over different values of a, or H in (2) may depend on t. Second, fBm models are inherently Gaussian. Many signals have positive increments and, hence, are non-Gaussian.

For applications such as network traffic [4, 5] and turbulence analysis [6], the statistics of such signals can be more accurately characterized using a *multifractal* analysis, which describes how the signal's scaling behavior varies across the signal. In this paper, we develop a multifractal signal model especially suited to positive-valued data with LRD. Modeling of such data is not only vital for networking [2, 4], but it also provides key insights into a host of other applications such as turbulence and geophysics [6].

We call our model the *multifractal wavelet model* (MWM) because of its wavelet-domain formulation. With its simple Haar-wavelet construction (Figure 1), the MWM is simple to apply, yet it can characterize LRD and a number of different multifractal properties. Since we cannot treat the MWM in full detail in this paper, we refer the interested reader to [7] for a more in-depth treatment.

2 FBm and LRD

Although we analyze fBm from a continuous-time point of view, for practical computations and simulations, we often work with sampled continuous-time fBm. The increments process of sampled fBm

$$X[n] := B(n) - B(n-1) \tag{3}$$

defines a stationary Gaussian sequence known as discrete *fractional Gaussian noise* (fGn) with covariance behavior [1]

$$r_X[k] \simeq |k|^{2H-2}, \text{ for } |k| \text{ large.} \tag{4}$$

For $1/2 < H < 1$, the covariance of fGn is strictly positive and decays so slowly that it is non-summable (i.e., $\sum_k r_X[k] = \infty$). This non-summability, corresponding to positive, slowly-decaying covariances over large time lags, defines LRD.

The LRD of fGn can be equivalently characterized in terms of how the aggregated processes

$$X^{(m)}(n) := \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i) \tag{5}$$

behave. It follows from (1) that $X(n) \stackrel{fd}{=} m^{1-H} X^{(m)}(n)$.

Hence, a log-log plot of the variance of $X^{(m)}(n)$ as a function of m — known as a *variance-time plot* — will have a slope of $2H-2$. The variance-time plot can characterize LRD in non-Gaussian, non-zero-mean data as well [2].

3 Wavelets and LRD

The discrete wavelet transform is a multi-scale signal representation of the form [8]

$$x(t) \;=\; \sum_k u_k \, 2^{-J_0/2} \, \phi\big(2^{-J_0}t - k\big) \;+\; \sum_{j=-\infty}^{J_0} \sum_k w_{j,k} \, 2^{-j/2} \, \psi\big(2^{-j}t - k\big)\,, \qquad j,k \in \mathbb{Z}$$

with J_0 the coarsest scale and u_k and $w_{j,k}$ the scaling and wavelet coefficients, respectively. The scaling coefficients may be viewed as providing a coarse approximation of the signal, with the wavelet coefficients providing higher-frequency "detail" information.

Wavelets serve as an approximate Karhunen-Loève transform for fBm [3], fGn, and more general LRD signals [9]. Thus, highly-correlated, LRD signals become nearly uncorrelated in the wavelet domain. In addition, the energy of the wavelet coefficients of continuous-time fBm decays with scale according to a power law [3]. While for *sampled* fBm the power-law decay is not exact [3], the *Haar* wavelet transform of fGn exhibits power-law scaling of the form[1] [9]

$$\text{var}(W_{j,k}) = \sigma^2 \; 2^{(2H-1)(j-1)} \; (2 - 2^{2H-1}), \tag{6}$$

where σ^2 is the variance of the fGn process.

[1] We use capital letters when we consider the underlying signal X (and, hence, its wavelet and scaling coefficients) to be random.

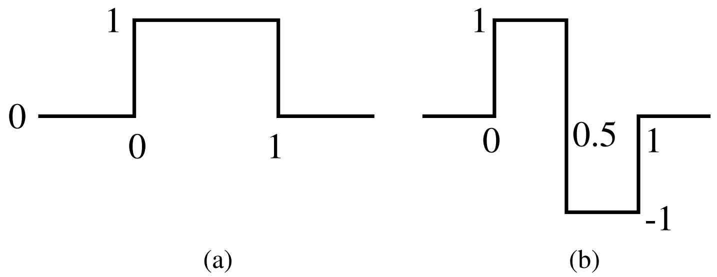


Figure 1: *(a) Haar scaling function ϕ and (b) Haar wavelet function ψ.*

4 The MWM

The basic idea behind the MWM is simple. To model non-negativity, we use the Haar wavelet transform with special wavelet-domain constraints. To capture LRD, we characterize the wavelet energy decay as a function of scale.

4.1 Haar Wavelets and Non-Negative Data

Before we can model non-negative signals using the wavelet transform, we must develop conditions on the scaling and wavelet coefficient values for $x(t)$ in (6) to be non-negative. While cumbersome for a general wavelet system, these conditions are simple for the Haar system. In a Haar transform (see Figure 1), the scaling and wavelet coefficients can be recursively computed using [8]

$$u_{j+1,k} = 2^{-1/2}(u_{j,2k} + u_{j,2k+1}) \text{ and } w_{j+1,k} = 2^{-1/2}(u_{j,2k} - u_{j,2k+1}). \tag{7}$$

Solving (7) for $u_{j,2k}$ and $u_{j,2k+1}$ we find

$$u_{j,2k} = 2^{-1/2}(u_{j+1,k} + w_{j+1,k}) \text{ and } u_{j,2k+1} = 2^{-1/2}(u_{j+1,k} - w_{j+1,k}). \tag{8}$$

For non-negative signals, $u_{j,k} \geq 0 \; \forall \; j,k$, which with (8) implies that

$$|w_{j,k}| \leq u_{j,k}, \qquad \forall \; j,k. \tag{9}$$

4.2 Multiplicative Model

The positivity constraints (9) on the Haar wavelet coefficients lead us to a very simple multi-scale, multiplicative signal model for positive processes. (See [10] for a similar model used as an intensity prior for wavelet-based image estimation.) Let $A_{j,k}$ be a random variable supported on the interval $[-1,1]$ and define the wavelet coefficients recursively by

$$W_{j,k} = A_{j,k} \, U_{j,k}. \tag{10}$$

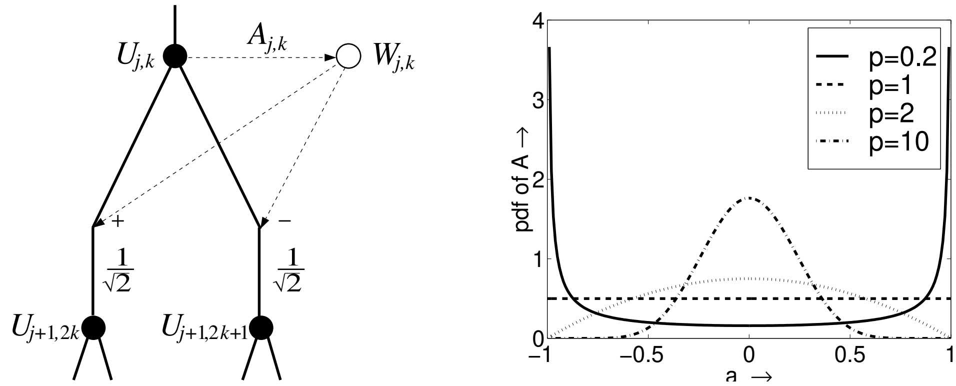
Together with (8) we obtain

$$U_{j,2k} = 2^{-1/2}(1 + A_{j+1,k})\, U_{j+1,k} \text{ and } U_{j,2k+1} = 2^{-1/2}(1 - A_{j+1,k})\, U_{j+1,k}. \tag{11}$$

The above construction can be visualized as a course-to-fine synthesis (see Figure 2(a)). Starting from the coarsest scale $j = J_0$, we can synthesize a realization of a process by iteratively applying (10) to obtain the wavelet coefficients at scale j and then applying (8) to obtain the scaling coefficients at the next finest scale $j-1$. When we reach the finest scale $j = J_1$ we obtain the desired process X, which is given by

$$X[k] = 2^{J_1/2} U_{J_1,k} \;\text{ for }\; k = 0, \ldots, 2^{J_0 - J_1} - 1. \tag{12}$$

In essence the above algorithm simultaneously synthesizes the wavelet coefficients and inverts the wavelet transform, requiring only $O(N)$ operations to create a length-N signal.



(a) (b)

Figure 2: *(a) MWM construction: At scale j, we form the wavelet coefficient as the product $W_{j,k} = A_{j,k}U_{j,k}$. Then, at scale $j-1$, we form the scaling coefficients $U_{j-1,2k}$ and $U_{j-1,2k+1}$ as sums and differences of $U_{j,k}$ and $W_{j,k}$ (normalized by $1/\sqrt{2}$). (b) Probability density function of a $\beta(p,p)$ random variable A. For $p = 0.2$, A resembles a binomial random variable, and for $p = 1$ it has a uniform density. For $p > 1$ the density appears like a truncated Gaussian density, and as p increases, the density resembles a Gaussian density more and more closely.*

4.3 β multipliers

We need to choose an appropriate distribution for the multiplier $A_{j,k}$. We will assume that $A_{j,k}$ is independent of $U_{j,k}$. Second, we will assume that $A_{j,k}$ is symmetric about 0; it is easily shown this symmetry is necessary for the resulting process to be stationary [7].

Because of its simplicity and flexibility, we will use a symmetric beta distribution, $\beta(p,p)$ (see Figure 2 (b)) for the $A_{j,k}$'s and christen the resulting model the βMWM. The variance of a random variable $A \sim \beta(p,p)$ is

$$\text{var}[A] = \frac{1}{2p+1}. \tag{13}$$

4.4 Covariance matching

Since we assume the multipliers are identically distributed within scale (i.e., $A_{j,k} \sim \beta(p_j, p_j)$), for our Haar wavelet construction we can control the wavelet energy decay across scale via

$$\frac{\text{var}(W_{j+1,k})}{\text{var}(W_{j,k})} = \frac{2\,\text{var}\,[A_{j+1,k}]}{\text{var}\,[A_{j,k}]\ (1+\text{var}\,[A_{j+1,k}])}. \tag{14}$$

Thus, to model a given process with the βMWM, we can select the parameters p_j via (13) and (14) to match the signal's theoretical wavelet-domain energy decay, such as (6). Or, given training data, we can select the parameters to match the sample variances of the wavelet coefficients as a function of scale.

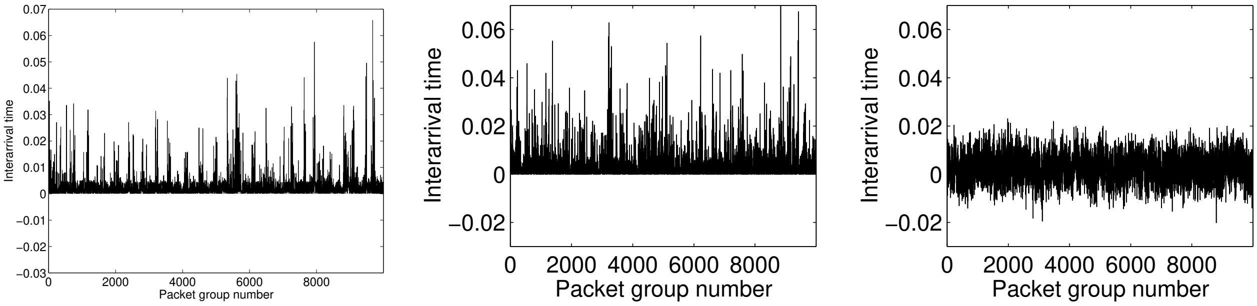
To complete the modeling, we only require to choose the coarsest scaling coefficients $U_{J_0,k}$ and the parameter $p_{(J_0)}$ of the model. From (10) and (13) we obtain

$$\text{var}(W_{J_0,k}) \cdot (2p_{(J_0)} + 1) = \mathbb{E}[U^2_{J_0,k}]. \tag{15}$$

We choose a coarsest scale J_0 and obtain $p_{(J_0)}$ from estimates of $\mathbb{E}[U^2_{J_0,k}]$ and $\text{var}(W_{J_0,k})$ using (15). We then synthesize b blocks of data, each with coarsest scale J_0 and finest scale J_1 (see Section 4.2) and concatenate them to obtain a synthetic data trace of $b \cdot 2^{J_0 - J_1}$ data points. To ensure the non-negativity of our process, we would have to model $U_{J_0,k}$ ($k = 0, .., b-1$) using multivariate densities with strictly non-negative support. However, $\mathbb{E}[U_{J_0,k}]$ will often so greatly outweigh var$[U_{J_0,k}]$ that the probability of a negative value will be negligible, even for a jointly Gaussian model.

Interestingly, the wavelet coefficients of the MWM are uncorrelated. This is easily deduced from the fact that the $A_{j,k}$'s are zero mean and independent, and from the dependency of wavelet coefficients across scale

$$W_{j,2k} = 2^{-1/2}\,\frac{A_{j,2k}}{A_{j+1,k}}\,(1 + A_{j+1,k})\,W_{j+1,k} \tag{16}$$



(a) (b) (c)

Figure 3: *Interarrival times of packets of (a) Bellcore August 1989 [2], (b) one realization of the beta multifractal wavelet model (βMWM) synthesis and (c) one realization of an fGn model. The βMWM trace resembles the Bellcore data closely, while the fGn data does not.*

and their dependency within scale

$$W_{j,2k+1} = \frac{A_{j,2k+1}}{A_{j,2k}} \, \frac{1 - A_{j+1,k}}{1 + A_{j+1,k}} \, W_{j,2k}. \tag{17}$$

[7]. Thus, by matching the variances of wavelet coefficients, the MWM would exactly capture the entire correlation structure of the wavelet coefficients, if the wavelet coefficients were truly decorrelated. Typically, the correlation between the wavelet coefficients of LRD processes are small [9], and therefore we can approximate such processes quite accurately with the MWM.

Of course, this analysis addresses only the second-order statistics of our signal. Higher-order properties of the MWM are the subject of a multifractal analysis.

5 The MWM is a Multifractal

Multifractals offer a wealth of processes that are novel in many respects. The backbone of a multifractal is typically a construction where one starts at a coarse scale and develops details of the process on finer scales iteratively in a multiplicative fashion. It follows from (11) that the MWM is a *binomial cascade*,[2] one of the simplest multifractals. The name binomial cascade is explained by applying (11) iteratively and writing $U_{j,k}$ as the product of the coarsest scale $U_{J_0,k}$ and the multipliers $2^{-1/2}(1 \pm A_{j+1,k})$.

Multiplicative structures, in particular the product representation of $U_{j,k}$, bear various consequences. First, if all multipliers $1 \pm A_{j+1,k}$ in (11) are log-normal, then the marginals $U_{j,k}$ will be log-normal as well. Similarly, if the $1 \pm A_{j,k}$ are all identically distributed, $U_{j,k}$ will be approximately log-normal by the central limit theorem.

Second, interpreting $U_{j,k}$ as the increment of a limiting process Y over the interval $[k2^{-n}, (k+1)2^{-n}]$, we find for Y a local behavior of the type (2). To see this, note that $\log |U_{j,k}| / \log 2^j$ can be written as the sum of approximately j factors of the form $\log_2 2^{-1/2}(1 \pm A_{l,m})$ normalized by $1/j$. So, we expect this number to converge to some limiting value H. It is essential to note, however, that this value H depends now on t — hence the term *multifractal* for Y.

As a further feature of interest, depending on the moments of the multipliers, the marginals $U_{j,k}$ of binomial cascades may have diverging moments of order q larger than some q_{crit}, where q_{crit} can be arbitrarily large. This broadens the realm of “heavy tailed” processes considerably.

The exact multifractal properties of the MWM are studied in detail in [7].

6 Application to Network Traffic

We demonstrate the power of the MWM for a problem of considerable practical interest — network traffic modeling. The LRD of data traffic can lead to higher packet loss than that predicted by classical queuing analysis [2]. A better understanding of the LRD data traffic can thus prove invaluable in solving network problems caused by LRD. Synthetic traffic models are well-suited for this purpose, because the effect of various model parameters on network performance

[2] For a detailed introduction to binomial cascades and multifractal theory, see [4, 7].

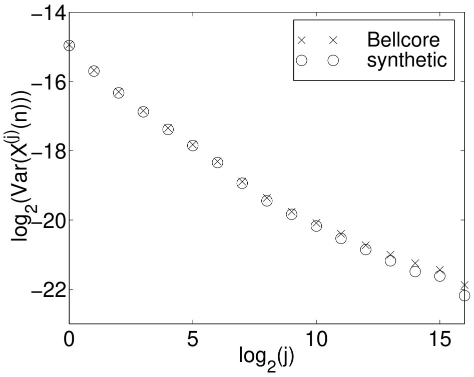


Figure 4: *Variance-time plot of the Bellcore data "×" and one realization of the βMWM synthesis "∘".*

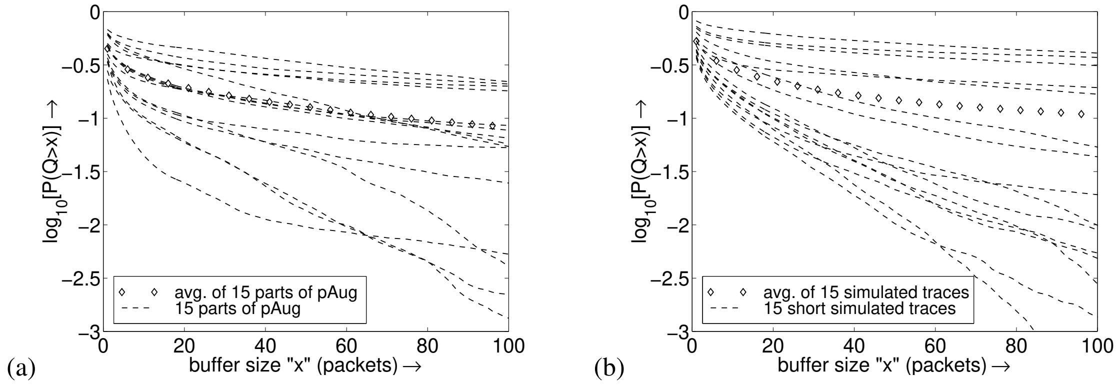


Figure 5: *Here we partition the pAug trace into 15 sub-traces of equal number of packets and compare their queueing behavior with that of 15 synthesized traces of the same length. In (a), observe that the sub-traces have a wide variation in tail queue behavior. In (b), observe that the synthesized traces display a similar variation in tail queue behavior as the sub-traces.*

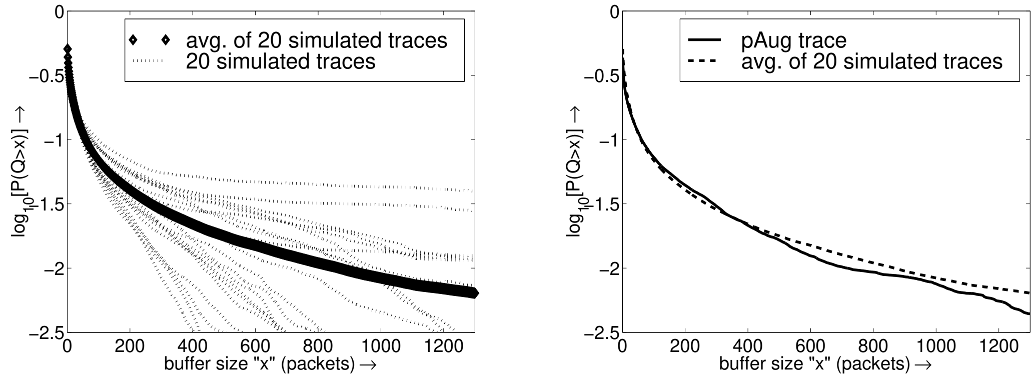
can be studied through simulation. The βMWM is one such model because it can capture vital characteristics of real traffic like LRD and queuing behavior. To show this, we focus on the August 1989 Bellcore Ethernet trace $pAug$, a record of one-million interarrival times (Figure 3(a)), as measured by Leland et al. [2]. We model this data using the βMWM. To train the βMWM, we take a Haar wavelet transform of the Bellcore data. We use (13) and (14) to choose the $\beta(p_j, p_j)$ distribution used at each scale j so that the theoretical variances of our synthesized wavelet coefficients match the measured variances of the Bellcore wavelet data. Since we require at least a few wavelet coefficients to get a good estimate of their sample variance, we cannot reliably fit our model to the coarsest scale wavelet coefficients of the Bellcore data. Thus, we model only the 16 finest wavelet scales of the Bellcore data using the βMWM, due to which the trained βMWM can generate synthetic traces of maximum length 2^{16}.

We analyze the LRD properties of the trace by estimating the variance-time plot, as shown in Figure 4. Although the data exhibits LRD (average slope corresponding to $H = 0.79$), the data does not appear to be strictly second-order self-similar, as evidenced by the "kink" in the slope. Thus, an fGn model model would be somewhat inaccurate in modeling the Bellcore data.

Figure 3 compares the Bellcore data with synthetic βMWM data and data from an fGn model. The fGn model has mean and variance equal to the sample mean and variance of Bellcore data and $H = 0.79$. We observe from Figure 3 that synthetic MWM data closely resembles the real data while fGn data does not. In fact, nearly 30% of the fGn data points go negative, which reveals that an fGn model is unrealistic for modeling positive data with a low mean and high variance like the Bellcore data considered here.

In [7], we further provide an empirical multifractal analysis of the synthesized data and show that it is closely matched by that of the Bellcore data.

To assess the accuracy and usefulness of the βMWM for traffic modeling, we compare the queuing behavior of the simulated traffic traces against that of the actual Bellcore data. In our simulation experiments, we consider the performance of an infinite length single server queue with a constant service rate of 500 packets/sec.



(a) (b)

Figure 6: *Comparison of the queuing behavior of the Bellcore pAug with that of 20 synthesized βMWM traces. Displayed are the tail probabilities of buffer occupancy against buffer size. In (a), observe the variability of the queue performance of the synthesized traces. In (b), observe that the average queue performance of simulated traces and that of the real trace match closely.*

The βMWM is a stochastic model, due to which different data sets synthesized by it will perform differently when sent through a queue. An ideal experiment to test the βMWM would be to compare the average tail queue behavior of several realizations of the real process with that of several realizations of the βMWM. Unfortunately, we have only one realization of the real process, and so try to learn about the queuing behavior of the underlying real process by partitioning the real data trace into 15 sub-traces of equal length (2^{16} packets). We assume each sub-trace to be a different realization of the underlying process and expect the queuing behavior of the sub-traces to resemble that of different sample functions of the process. We then use the βMWM to synthesize 15 traces each of 2^{16} packet interarrival times. The performance of the queue with the real and βMWM sub-traces as input are shown in Figure 5. We observe that the sub-traces of the real data have widely varying tail queue behavior, and that the simulation traces have a similar variation in queuing behavior.

We next compare the queuing performance of the entire real data trace with that of 20 traces of the approximately the same length (10^6 points), generated using the βMWM (see Figure 6). Since the trained βMWM can generate traces of 2^{16} data points, the simulation traces for this experiment were formed by concatenating 15 length-2^{16} traces generated using the βMWM. We observe that the simulated traces in Figure 6(a) exhibit a wide variation in tail queue behavior. The results of the previous experiment indicate that this is to be expected. We also observe that the average tail queue behavior of the simulated traces matches that of the real trace exceedingly well (see Figure 6(b)). However, as the previous experiment suggests, the real data can be viewed as just one possible realization of our model, and so it cannot be expected to always exhibit the same queuing behavior as the average of several simulated traces.

In summary, these queuing experiments demonstrate that our synthesized traffic traces not only match real traffic in its various statistical properties but also in its queuing behavior. This demonstrates the potential use of our model for network modeling and simulation.

7 Conclusions

The multiplicative wavelet model (MWM) combines the power of multifractals with the efficiency of the wavelet transform to form a flexible framework natural for characterizing and synthesizing positive-valued data with LRD. As our numerical experiments have shown, the MWM is particularly suited to the analysis and synthesis of network traffic data. In addition, the model could find application in areas as diverse as financial time-series characterization, geophysics (using 2-d and 3-d wavelets), and texture modeling. The parameters of the MWM are simple enough to be easily inferred from observed data or chosen a priori. Computations involving the MWM are extremely efficient — synthesis of a trace of N sample points requires only $O(N)$ computations. Finally, several extensions to the MWM are straightforward. The choice of β-distributed wavelet multipliers $A_{j,k}$ is not essential. Alternatively, we can employ mixtures of β's or even purely discrete distributions to fit higher-order multifractal moments.

References

[1] T. Lundahl, W. Ohley, S. Kay, and R. Siffert, "Fractional Brownian motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. on Medical Imaging*, vol. 5, pp. 152–161, Sep. 1986.

[2] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, pp. 1–15, 1994.

[3] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–916, Mar. 1992.

[4] R. Riedi and J. L. Véhel, "TCP traffic is multifractal: a numerical study," *IEEE Trans. Networking*, submitted Oct. 1997. Available at www.dsp.rice.edu.

[5] J. L. Véhel and R. Riedi, "Fractional Brownian motion and data traffic modeling: The other end of the spectrum," *Fractals in Engineering*, pp. 185–202, Springer 1997.

[6] B. B. Mandelbrot, "Intermittent turbulence in self similar cascades: Divergence of high moments and dimension of the carrier," *J. Fluid. Mech.*, vol. 62, p. 331, 1974.

[7] R. H. Riedi, M. S. Crouse, V. Ribeiro, and R. G.Baraniuk, "A multifractal wavelet model with application to TCP network traffic," *Submitted to IEEE Trans. on Info. Theory*, 1998. Available at www.dsp.rice.edu.

[8] I. Daubechies, *Ten Lectures on Wavelets*. New York: SIAM, 1992.

[9] L. M. Kaplan and C.-C. J. Kuo, "Extending self-similarity for fractional Brownian motion," *IEEE Trans. Signal Proc.*, vol. 42, pp. 3526–3530, Dec. 1994.

[10] K. E. Timmerman and R. D. Nowak, "Multiscale Bayesian estimation of Poisson intensities," in *Proc. 31st Asilomar Conf.*, (Pacific Grove, CA), Nov. 1997.