Simulation of nonGaussian Long-Range-Dependent Traffic using Wavelets

*Vinay J. Ribeiro, Rudolf H. Riedi, Matthew S. Crouse, and Richard G. Baraniuk**
Department of Electrical and Computer Engineering
Rice University
6100 South Main Street
Houston, TX 77005, USA

April 9, 1999

Abstract

In this paper, we develop a simple and powerful multiscale model for the synthesis of nonGaussian, long-range dependent (LRD) network traffic. Although wavelets effectively decorrelate LRD data, wavelet-based models have generally been restricted by a Gaussianity assumption that can be unrealistic for traffic. Using a multiplicative superstructure on top of the Haar wavelet transform, we exploit the decorrelating properties of wavelets while simultaneously capturing the positivity and "spikiness" of nonGaussian traffic. This leads to a swift $O(N)$ algorithm for fitting and synthesizing N-point data sets. The resulting model belongs to the class of multifractal cascades, a set of processes with rich statistical properties. We elucidate our model's ability to capture the covariance structure of real data and then fit it to real traffic traces. Queueing experiments demonstrate the accuracy of the model for matching real data. Our results indicate that the nonGaussian nature of traffic has a significant effect on queuing.

1 Introduction

Traffic models play a significant rôle in the analysis and characterization of network traffic and network performance. Accurate models enhance our understanding of these complex signals and systems by allowing us to study the effect of various model parameters on network performance through simulation.

The presence of *long-range dependence* (LRD) in modern network traffic was demonstrated convincingly in the landmark paper by Leland et. al. [1]. There, measurements of traffic load on an Ethernet were attributed to *fractal* behavior or *self-similarity*, i.e., to the fact that the data "looked statistically similar" ("bursty") on all time-scales. These features are inadequately described by classical traffic models, such as Markov or Poisson models. In particular, the LRD of data traffic can lead to higher packet losses than that predicted by classical queuing analysis [1, 2].

These findings were immediately followed by the development of new fractal traffic models [3–5]. The *fractional Brownian motion* (fBm), the most broadly applied fractal model, 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 equality in (finite-dimensional) distribution. The increment process $G(k) \equiv B(k+1) - B(k)$, called *fractional Gaussian noise* (fGn), has an autocorrelation of the form

$$r_G[k] = \frac{\sigma^2}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}). \tag{2}$$

The parameter H, $0 < H < 1$, is known as the *Hurst parameter*. It simultaneously rules the large-scale behavior and the degree of local "spikiness." In particular, for all t

$$B(t+s) - B(t) \simeq s^H \tag{3}$$

(more precisely, $B(t+s) - B(t)$ is a zero-mean Gaussian process of variance s^{2H}), meaning that fBm has "infinite slope" everywhere. Processes approximating fBm/fGn can be synthesized almost effortlessly in the wavelet domain due to the amazing decorrelating effect of the wavelet transform [6].

A strong argument for the fBm/fGn models in networks is that in many cases traffic can be viewed as the superposition of a large number of independent individual ON/OFF sources, with the ON durations heavy-tailed [7, 8]. In this case, subtracting the mean arrival rate and normalizing properly, the aggregated ON/OFF sources (cumulative arrivals) converge to Gaussian fBm by the central limit theorem (CLT) [1,3]. A "self-similar" traffic arrival model (of the increments process) is, thus, simply an "fGn+mean" model with given variance and H. The fBm/fGn models have found wide use in networking, since their Gaussianity and strong scaling (1) allows analysts to perform analytical studies of queueing behavior [9–13].

Unfortunately, the fBm/fGn models have severe limitations for network traffic applications. First, real-world traffic traces do not exhibit the strict self-similarity of (1) or (2) and are at best merely asymptotically self-similar. In other words, the single parameter H is not sufficient to capture the complicated correlation structure of real network processes. Indeed, convincing evidence has been produced establishing the importance of short-term correlations for buffering [16–18] and so-called relevant time scales have been discovered [19]. The wavelet-domain independent Gaussian (WIG) model generalizes fBm/fGn by allowing a more flexible scaling relation than (1). By matching both long and short-term correlations, the WIG model more completely matches the correlation structure of a target data set [15].

Second, the Gaussianity of fBm/fGn/WIG models can be unrealistic for certain types of traffic, for instance when the standard deviation of the data exceeds the mean. In this case, the fBm/fGn/WIG output signals take on a considerable number of negative values (see Figure 1).

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

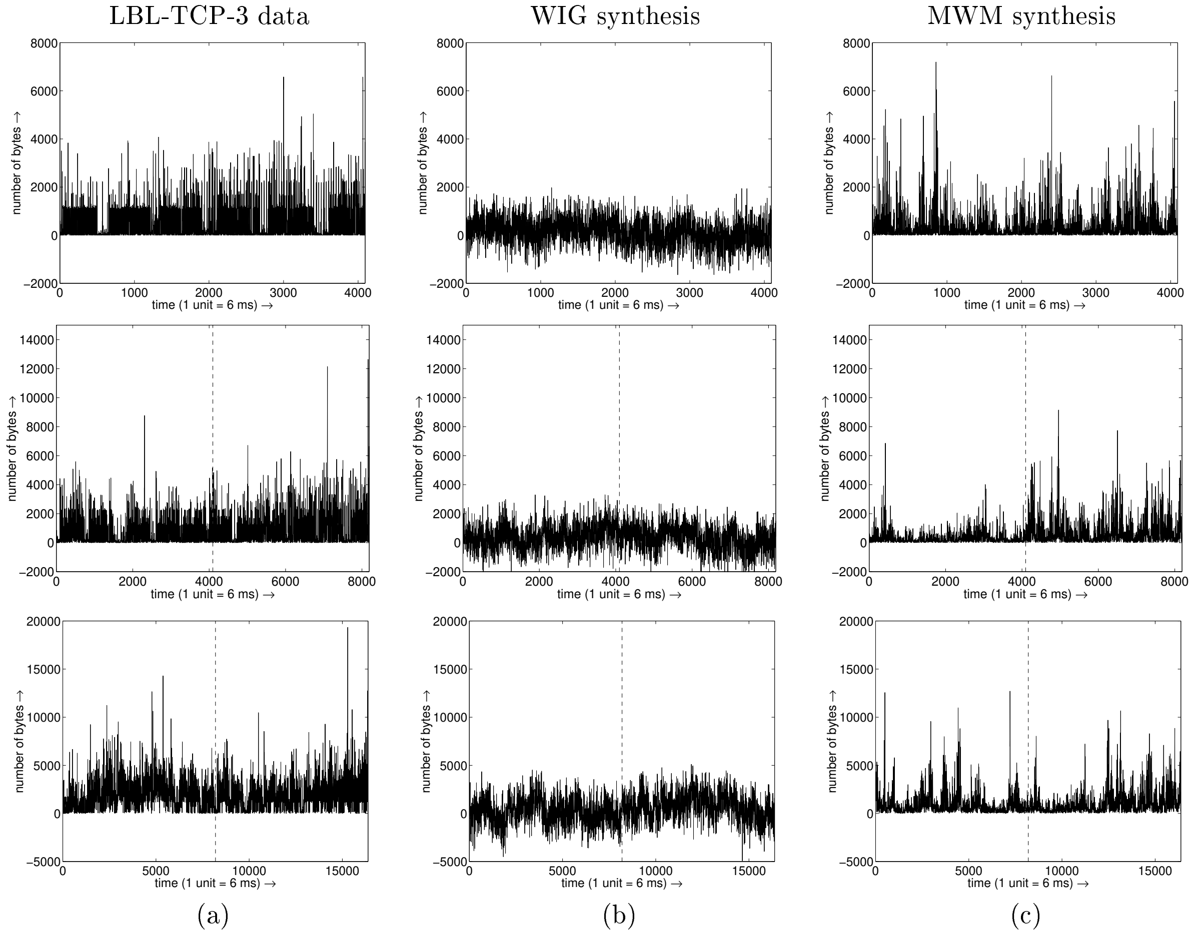


Figure 1: *Bytes-per-time arrival process at different aggregation levels for (a) wide-area TCP traffic at the Lawrence Berkeley Laboratory (trace LBL-TCP-3) [14], (b) one realization of the state-of-the-art wavelet-domain independent Gaussian (WIG) model [15], and (c) one realization of the multifractal wavelet model (MWM) synthesis. The top, middle and bottom plots correspond to bytes arriving in intervals of* 6 *ms*, 12 *ms and* 24 *ms respectively. The top and middle plots correspond to the second half of the middle and bottom plots, respectively, as indicated by the vertical dotted lines. The MWM traces closely resemble the real data closely, while the WIG traces (with their large number of negative values) do not.*

Third, in many networking applications, we are nowhere near the Gaussian limit, in particular on small time scales. Indeed, various authors have observed marginals that differ substantially from Gaussian. Usually these distributions have been observed to be heavy tailed [20, p. 364], [21]. Consequently, methods aimed at fitting marginals have been developed [22,23]. Also, more versatile models such as fractional ARIMA [24] have been applied towards better matching the short-range and long-range correlation structure present in real traces.

In this paper, we propose a new non-linear model for network traffic data. The *multifractal wavelet model* (MWM) is based on a multiplicative cascade in the wavelet domain that by design guarantees a positive output. Since each sample of the MWM process is obtained as a product of several positive independent random variables, the MWM's marginal density is approximately lognormal, a heavier-tailed distribution than the Gaussian. The MWM is thus a more natural fit for positive arrival processes, especially those with a standard deviation much larger than the mean (as observed in the traces we have studied).

In its simplest form, the MWM is closely related to the wavelet-based construction of fBm/fGn, having as few parameters (mean, variance, H). However, the MWM framework boasts the flexibility to additionally match the short-term correlations like the WIG model.

The MWM has a bursty demeanor that matches that of real traffic much more closely than fBm/fGn. The TCP traffic we have studied here exhibits local scaling similar to (3), but with an exponent H_t that depends on t. This has been termed *multifractal behavior* and was reported for the first time in [25] and subsequently in [26–29]. Amazingly, the statistical properties of H_t as a random variable in t can be described compactly through a function $T(q)$ that controls the scaling behavior of the sample moments of order q. This powerful relation, called the *multifractal formalism*, ties burstiness, higher-order dependence structure, and moments of marginals together in one unified theory.

Fitting the MWM to real traffic traces results in an excellent match, far better than the WIG model, visually (see Figure 1) and, as we will see, in the multifractal partition function $T(q)$, the burstiness as measured by the multifractal spectrum, the marginals, and the queueing behavior. Since these properties all depend on the small time-scale behavior, it appears that the multiplicative MWM approach is more appropriate than an additive Gaussian one.

In this paper, we summarize the impact of LRD on networking in Section 2. After introducing the wavelet transform and describing the WIG model in Section 3, we derive the MWM in Section 4. Section 5 reports on the results of simulation experiments with real data traces. We give an intuitive introduction to multifractal cascades in Section 6 and close with conclusions in Section 7.

2 Long-range Dependence in Network Traffic

The discovery of LRD in data traffic [1, 14] has incited a revolution in network design, control and modeling. Intuitively, the strong correlations present in a LRD process are responsible for its "bursty" nature. Thus, LRD traffic arrives in bursts that, upon entering a queue, cause excessive buffer overflows that are not predicted by traditional non-LRD traffic models such as Poisson and Markov models [2].

2.1 Long-range dependence (LRD)

Consider a discrete-time, wide-sense stationary random process $\{X_t,\ t \in \mathbb{Z}\}$ with auto-covariance function $r_X[k] = \mathrm{cov}(X_t, X_{t+k})$. A change in time scale can be represented by forming the aggregate process $X_t^{(m)}$, which is obtained by averaging X_t over non-overlapping blocks of length m replacing each block by its mean

$$X_t^{(m)} = \frac{X_{tm-m+1} + \cdots + X_{tm}}{m}. \tag{4}$$

Denote the auto-covariance of $X_t^{(m)}$ by $r_X^{(m)}[k]$. The process X is said to exhibit LRD if its auto-covariance decays slowly enough to render $\sum_{k=-\infty}^{\infty} r_X[k]$ infinite [30]. Equivalently, the power spectrum $S_X(f)$ is singular near $f = 0$ and $m\ r_X^{(m)}[0] \to \infty$ as $m \to \infty$.

An important class of LRD processes are the *asymptotically second-order self-similar processes*, which may be defined by the property $r_X[k] \sim k^{2H-2}$ for some $H \in (1/2, 1)$, or equivalently by [30]

$$\mathrm{var}(X^{(m)}) = r_X^{(m)}[0] \to m^{2H-2} \tag{5}$$

as $m \to \infty$. In words, these processes "look similar" on all scales, at least from point of view of second-order statistics. The fGn, is such a process where the *Hurst parameter* H is the same as in (1).

To estimate H by the *variance-time plot* method, we fit a straight line through the plot of an estimate of $\log \mathrm{var}(X^{(m)})$ against $\log(m)$. More reliable estimators have also been devised [24], in particular an unbiased one based on wavelets [31].

2.2 Impact of LRD on networking

The pre-eminent LRD model at present is the fGn. Its popularity stems from the fact that it is a second-order self-similar Gaussian process (2), and thus is analytically tractable. In addition, it is completely described by just two parameters — variance and H. When fGn is input to an infinite-length queue with constant service rate, the tail queue distributions decay asymptotically with a Weibullian law

$$P[Q > x] \approx \exp(-\delta x^{2-2H}), \tag{6}$$

with δ a positive constant that depends on the service rate at the queue [10, 11]. The decay of the tail queue distribution for fGn with $H > 1/2$ is much slower than the exponential decay predicted by short-range dependent (SRD) classical models [2]. This corresponds to the case $H = 1/2$.

Even though (6) shows that LRD processes have higher tail queue probabilities than SRD processes, there is still an ongoing discussion on the effect of LRD on queuing, with researchers arguing both for and against its importance [17–19, 32–34].

3 Wavelets and LRD Processes

3.1 Wavelet transform

The discrete wavelet transform provides a multiscale signal representation of a one-dimensional signal $c(t)$ in terms of shifted and dilated versions of a prototype bandpass wavelet function $\psi(t)$ and shifted versions of a lowpass scaling function $\phi(t)$ [35]. For special choices of the wavelet and scaling functions, the atoms

$$\begin{aligned} \psi_{j,k}(t) &\equiv 2^{j/2}\ \psi\big(2^j t - k\big), \\ \phi_{j,k}(t) &\equiv 2^{j/2}\ \phi\big(2^j t - k\big), \quad j,k \in \mathbb{Z} \end{aligned} \tag{7}$$

form an orthonormal basis, and we have the signal representation [35]

$$c(t) = \sum_k u_{J_0,k}\, \phi_{J_0,k}(t) + \sum_{j=J_0}^{\infty} \sum_k w_{j,k}\, \psi_{j,k}(t), \tag{8}$$

with $w_{j,k} \equiv \int c(t)\, \psi_{j,k}^*(t)\, \mathrm{d}t$, and $u_{J_0,k} \equiv \int c(t)\, \phi_{J_0,k}^*(t)\, \mathrm{d}t$. Without loss of generality, we will assume $J_0 = 0$.

In this representation, k indexes the spatial location of analysis and j indexes the *scale* or resolution of analysis — larger j corresponds to higher resolution with $j = 0$ indicating the coarsest scale or lowest resolution of analysis. In practice, we work with a sampled or finite-resolution representation of $c(t)$, replacing the semi-infinite sum in (8) with a sum over a finite number of scales $0 \le j \le n-1$, $n \in \mathbb{Z}$. Using filter bank techniques, the wavelet transform and inverse wavelet transform can be computed in $O(N)$ operations for a length-N signal. For more information on wavelet systems and their construction, see [35].

In the *Haar* wavelet transform (see Figure 2), the prototype scaling and wavelet functions are given by

$$\phi(t) = \begin{cases} 1, & 0 \le t < 1 \\ 0, & \text{else} \end{cases} \quad \text{and} \quad \psi(t) = \begin{cases} 1, & 0 \le t < 1/2 \\ -1, & 1/2 \le t < 1 \\ 0, & \text{else.} \end{cases}$$

The Haar scaling and wavelet coefficients can be recursively computed via [35]

$$\begin{aligned} u_{j-1,k} &= 2^{-1/2}(u_{j,2k} + u_{j,2k+1}), \\ w_{j-1,k} &= 2^{-1/2}(u_{j,2k} - u_{j,2k+1}). \end{aligned} \tag{9}$$

3.2 Modeling LRD data

Wavelets serve as an approximate Karhunen-Loève or decorrelating transform for fBm [6], fGn, and more general LRD signals [36]. Hence, modeling and processing of these signals in the wavelet domain is often more efficient and powerful than in the time domain.

The variance of the wavelet coefficients of continuous-time fBm decays with scale according to a power law in H [6]. For fGn, an exact power-law in H also holds for decay of the Haar wavelet coefficient variances [36]. This power-law decay, along with the decorrelation property of wavelets, has led to fast, robust algorithms for estimation [36, 37].

Gaussian LRD processes can be approximately synthesized by generating wavelet coefficients as independent zero-mean Gaussian random variables, identically distributed within scale according to $W_{j,k} \sim N(0, \sigma_j^2)$,[1] with σ_j^2 the wavelet-coefficient variance at scale j [15].

[1] We use capital letters when we consider the underlying variables to be random.

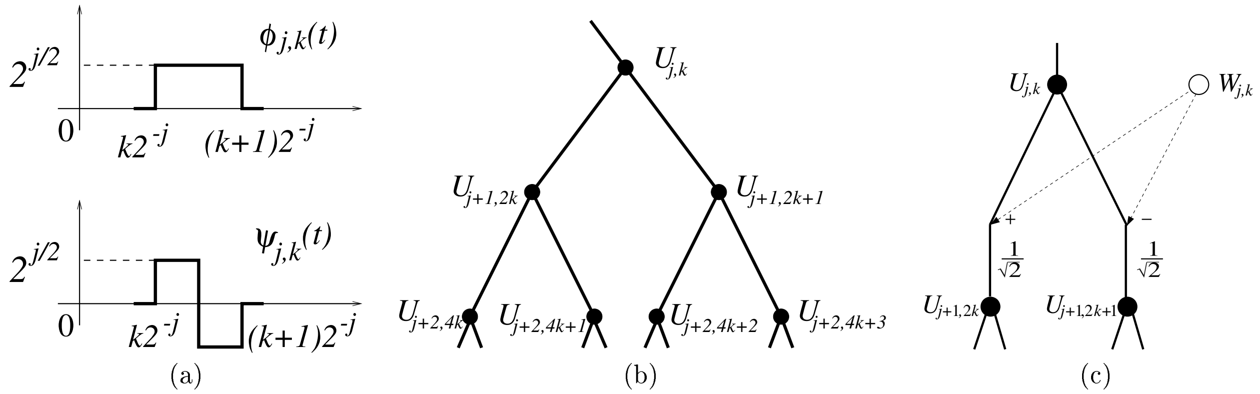


Figure 2: *(a) The Haar scaling and wavelet functions $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$. (b) Binary tree of scaling coefficients from coarse to fine scales. (c) Recursive scheme for calculating the Haar scaling coefficients $U_{j+1,2k}$ and $U_{j+1,2k+1}$ at scale $j+1$ as sums and differences of the scaling and wavelet coefficients $U_{j,k}$ and $W_{j,k}$ at scale j (normalized by $1/\sqrt{2}$). For the WIG model, the $W_{j,k}$'s are mutually independent and identically distributed within scale according to $W_{j,k} \sim N(0, \sigma_j^2)$.*

A power-law decay for the σ_j^2's leads to approximate wavelet synthesis of fBm or fGn [6]. However, while network traffic may exhibit LRD consistent with fBm or fGn, it may have short-term correlations that vary considerably from pure fBm or fGn scaling. Such LRD processes can be modeled by setting σ_j^2 to match the measured or theoretical variances of the wavelet coefficients of the desired process [15]. We call the resulting model the wavelet-domain independent Gaussian (WIG) model [15] (see Figure 2(c)). For a length-N signal, the WIG is characterized by approximately $\log_2 N$ parameters.

The WIG model assumes Gaussianity even though network traffic signals (such as loads and interarrival times) can be highly non-Gaussian. Not only are these signals strictly non-negative, but they can exhibit "spiky" behavior corresponding to a marginal distribution whose right-side tail decays much more slowly than that of a Gaussian. We seek a more accurate marginal characterization for these spiky, non-negative LRD processes, yet wish to retain the decorrelating properties of wavelets and the simplicity of the WIG model.

3.3 Modeling non-negative data with the Haar wavelet

In order to model non-negative signals using the wavelet transform, we must develop conditions on the scaling and wavelet coefficient values for $c(t)$ in (8) to be non-negative. While cumbersome for a general wavelet system,[2] these conditions are simple for the Haar system (see Figure 2), on which we focus for the balance of this paper.

Since the scaling coefficients $u_{j,k}$ represent the local mean of the signal at different scales and shifts, they are non-negative if and only if the signal itself is non-negative; that is, $c(t) \geq 0 \Leftrightarrow u_{j,k} \geq 0, \ \forall\ j, k$. This condition leads us directly to constraints on the Haar wavelet coefficients. Solving (9) for $u_{j,2k}$ and $u_{j,2k+1}$, we find

$$\begin{aligned} u_{j,2k} &= 2^{-1/2}(u_{j-1,k} + w_{j-1,k}), \\ u_{j,2k+1} &= 2^{-1/2}(u_{j-1,k} - w_{j-1,k}), \end{aligned} \tag{10}$$

which corresponds to moving down the tree in Figure 2(b) one scale level at a time.

Now, combining (10) with the constraint $u_{j,k} \geq 0$, we obtain the condition

$$c(t) \geq 0 \Leftrightarrow |w_{j,k}| \leq u_{j,k}, \qquad \forall\ j, k. \tag{11}$$

[2] The conditions are straightforward also for certain biorthogonal wavelet systems.

4 Multifractal Wavelet Model

Let us summarize our basic wavelet-based approach for modeling nonGaussian LRD network traffic. As with the WIG we will characterize the Haar wavelet variance decay as a function of scale to capture the short-range and long-range correlations. In contrast to the WIG, we will enforce the constraint (11) to ensure the non-negativity of the model output.

To keep things clear, we will introduce three different processes: the continuous-time model output $c(t)$, its integral $D(t)$, and a discrete-time approximation $C[k]$ to $c(t)$. These three signals are related by

$$C[k] \equiv \int_{k2^{-n}}^{(k+1)2^{-n}} c(t)\, \mathrm{d}t = D\left((k+1)2^n\right) - D\left(k2^n\right). \tag{12}$$

Here, $c(t)$ and $D(t)$ play rôles analogous to fGn and fBm, respectively.

For notational simplicity, we will assume that both $c(t)$ and $D(t)$ live on $[0, 1]$ and that $C[k]$ is a length-2^n discrete-time signal. Thus, there is only one scaling coefficient $U_{0,0}$ in (8). (A more general case is treated in [29].) We will primarily focus on $C[k]$.

For the Haar wavelet transform, $C[k]$ relates directly to the finest-scale scaling coefficients:

$$C[k] \;=\; 2^{n/2} U_{n,k}, \qquad k = 0, 1, \ldots 2^n - 1. \tag{13}$$

4.1 The model

The positivity constraints (11) on the Haar wavelet coefficients suggest a very simple multiscale, multiplicative signal model for positive processes. In the *multifractal wavelet model* (MWM) we compute the wavelet coefficients recursively by

$$W_{j,k} = A_{j,k}\ U_{j,k}, \tag{14}$$

with $A_{j,k}$ a random variable supported on the interval $[-1, 1]$. Together with (10), we obtain (see Figure 3)

$$\begin{aligned} U_{j,2k} &= 2^{-1/2}(1 + A_{j+1,k})\, U_{j-1,k}, \\ U_{j,2k+1} &= 2^{-1/2}(1 - A_{j+1,k})\, U_{j-1,k}. \end{aligned} \tag{15}$$

See [38] for a similar model used as an intensity prior for wavelet-based image estimation.

To generate a realization of an MWM process, we perform the following coarse-to-fine synthesis:

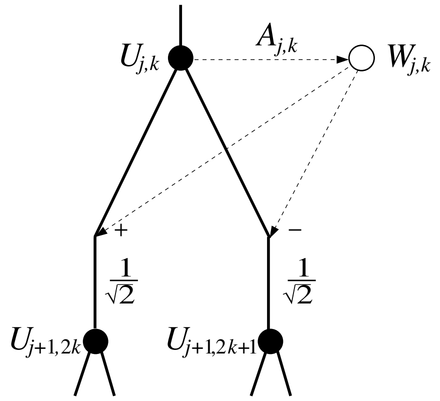


Figure 3: *MWM construction: At scale j, generate the multiplier $A_{j,k} \sim \beta(p_j, p_j)$, and then form the wavelet coefficient as the product $W_{j,k} = A_{j,k} U_{j,k}$. At scale $j+1$ of this tree, form the scaling coefficients in the same manner as the WIG model in Figure 2.*

1. Set $j = 0$. Fix or compute the coarsest (root) scaling coefficient $U_{0,0}$, establishing the global mean of the signal.
2. At scale j, generate the random multipliers $A_{j,k}$ and calculate each $W_{j,k}$ via (14) for $k = 0, \dots, 2^j - 1$.
3. At scale j, use $U_{j,k}$ and $W_{j,k}$ in (10) to calculate $U_{j+1,2k}$ and $U_{j+1,2k+1}$, the scaling coefficients at scale $j+1$, for $k = 0, \dots, 2^j - 1$.
4. Iterate steps 2 and 3, replacing j by $j+1$ until the finest scale $j = n$ is reached.

We can express the signal $C[k]$ directly as a product (or cascade) of the random multipliers $1 \pm A_{j,k}$. Decomposing each shift k into a binary expansion $k = \sum_{i=0}^{n-1} k'_i 2^{n-1-i}$, we can write

$$C[k] = 2^{-n/2} U_{n,k} = 2^{-n}\, U_{0,0} \prod_{i=0}^{n-1} \frac{(1 + (-1)^{k'_i} A_{i,k_i})}{2}, \tag{16}$$

with

$$k_0 \equiv 0, \quad \text{and} \quad k_i = \sum_{j=0}^{i-1} k'_i 2^{i-1-j}, \quad i = 1, \dots, n-1. \tag{17}$$

This result can be derived by iteratively applying (15) [29].

Since the scaling coefficients are generated simultaneously with the wavelet coefficients, there is no need to invert the wavelet transform. The finest-scale scaling coefficients $U_{n,k}$ are in fact the MWM output process (13). The total cost for computing N MWM signal samples is $O(N)$. In fact, synthesis of a trace of length 2^{18} data points takes just 8 seconds of workstation cpu time.

4.2 β multipliers

All that remains is to choose an appropriate distribution for the multipliers $A_{j,k}$. For simplicity of development, we will assume that the $A_{j,k}$'s are mutually independent and independent of $U_{j,k}$. We will also assume that the $A_{j,k}$'s are symmetric about 0 and identically distributed within scale; it is easily shown that these two conditions are necessary for the resulting process to be first-order stationary [29]. This leads us to the choice of the *symmetric beta distribution*, $\beta(p,p)$ (see Figure 4) for the $A_{j,k}$'s

$$A_{j,k} \sim \beta(p_j, p_j), \tag{18}$$

Table 1: *Comparison of the tree-based WIG and MWM models. For approximating a signal with a strict fGn covariance structure, both the WIG and MWM require only three parameters (mean, variance, and H).*

WIG	MWM
Additive	Multiplicative
Gaussian	Asymptotically Lognormal
LRD matched	LRD matched
Monofractal	Multifractal
$\log_2 N + 2$ parameters	$\log_2 N + 2$ parameters
$O(N)$ synthesis	$O(N)$ synthesis

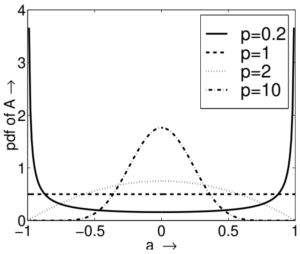


Figure 4: *Probability density function of a $\beta(p,p)$ random variable A. For $p = 0.2$, A resembles a binomial random variable, and for $p = 1$ it has a uniform density. For $p > 1$ the density appears like a truncated Gaussian density, and as p increases, the density resembles a Gaussian density more and more closely.*

with p_j the beta parameter at scale j. The beta distribution is compactly supported, easily shaped, and amenable to closed-form calculations. The variance of a random variable $A \sim \beta(p,p)$ is

$$\mathrm{var}[A] = \frac{1}{2p+1}. \tag{19}$$

In the MWM, the p_j play a rôle analogous to the σ_j^2 of the WIG model. With one beta parameter per wavelet scale, the MWM uses approximately $\log_2 N$ parameters for a trace of length N. Distributions with more parameters (e.g., discrete distributions or mixtures of betas) could be used to capture high-order data moments at a cost of increased model complexity [29]. See Table 1 for a comparison of the WIG and MWM properties.

4.3 Covariance matching

The p_j's allow us to control the wavelet energy decay across scale, since

$$\frac{\mathrm{var}(W_{j-1,k})}{\mathrm{var}(W_{j,k})} = \frac{2\,\mathrm{var}\,[A_{j-1,k}]}{\mathrm{var}\,[A_{j,k}]\,\left(1 + \mathrm{var}\,[A_{j-1,k}]\right)} = \frac{2p_j + 1}{p_{j-1} + 1}. \tag{20}$$

Thus, to model a given process with the MWM, we can select the p_j's to match the signal's theoretical wavelet-domain energy decay. 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 must choose the parameter p_0 of the model and characterize the distribution of the coarsest scaling coefficient $U_{0,0}$. From (14) and (19) we obtain

$$(2p_0 + 1)\mathrm{var}(W_{0,0}) = \mathbb{E}[U_{0,0}^2], \tag{21}$$

which allows us to obtain p_0 from estimates of $\mathbb{E}[U_{0,0}^2]$ and $\mathrm{var}(W_{0,0})$.

To precisely model $U_{0,0}$, we would have to use a strictly non-negative probability density function to ensure the non-negativity of the MWM output. However, in practice a Gaussian model at the coarsest scale (requiring $\mathbb{E}[U_{0,0}]$ and $\mathrm{var}[U_{0,0}]$) is usually sufficient if enough scales are employed (so that $\mathbb{E}[U_{0,0}] \gg$ standard deviation of $U_{0,0}$).

5 Experimental Results

In this section, we perform experiments with real data traces to demonstrate the MWM's capacity to capture important properties of real data. As expected, the MWM does an excellent job in capturing the correlation structure of real data sets. We also observe that the MWM performs well in matching the marginals and higher-order moments of real data. Recall that the Gaussian WIG model is also capable of capturing the correlation structure of training data. We thus have two models, both of which capture the correlation structure of real data but with the MWM coming closer to matching the marginals and higher-order moments. Equipped with these models, we are in an excellent position to perform queuing experiments to study if the correlation structure is by itself sufficient to capture the queuing behavior of real traffic.

5.1 Real data

We use two well-known real data traces in our experiments. The first (LBL-TCP-3) contains two hours' worth of wide-area TCP traffic between the Lawrence Berkeley Laboratory and the rest of the world [14]. This data contains the following information about each packet: the time-stamp, (renumbered) source host, (renumbered) destination host, source TCP port, destination TCP port, and number of data bytes. In our experiments we use only the time-stamp and data bytes information. We form a data trace by counting the number of bytes of packets that arrive in consecutive time intervals of 6 ms and use the first 2^{20} data points in our simulation experiments. This trace has a sample mean of 257.5 bytes/(unit time) and sample standard deviation of 562.6 bytes/(unit time).

The second real data set is one of the famous Ethernet data traces collected at Bellcore Morristown Research and Engineering facility [1]. The trace (BC-pAug89) began at 11:25 on August 29, 1989, and ran for about 3142.82 seconds (until 1,000,000 packets had been captured). As in the case of the LBL-TCP-3 data set, we obtain a data trace by summing the bytes of packets that arrived in consecutive time intervals of 2.6 ms. We use the first 2^{20} data points of this trace in our experiments. This trace has mean 345.8 bytes/(unit time) and standard deviation 703.4 bytes/(unit time). The BC-pAug89 trace is approximately a second-order self similar process with $H = 0.79$ [37].

5.2 Physical Interpretation

The MWM multipliers have a simple interpretation as recursively partitioning the arriving bytes into smaller and smaller time intervals. For instance, the value $U_{0,0}$ determines the total number of bytes in the entire trace. The value $A_{0,0}$ determines how many of these packets will be placed in the first half of the trace. The value $A_{1,0}$ then determines how many of these bytes will be placed in the first quarter of the trace, and so on.

When trained on real network data, the behavior of the multipliers $A_{j,k}$ changes with scale, with extremely low variance at coarse scales and high variance at fine scales. Amazingly, this is consistent with both the small-scale behavior of actual traffic *and* the large-scale properties resulting from the superposition of a large number of souces [7,8]. At fine scales multiplicative schemes with large variances produce bursts like those in real data (recall Figure 1). At coarse scales, the scaling coefficients (which correspond to the arrival of traffic over large time scales) involve only a handful of low-variance multipliers $A_{j,k}$. From (15) we can write, for example, at the third-coarsest scale:

$$\begin{aligned} U_{2,0} &\overset{fd}{=} \frac{U_{0,0}}{2}\,(1+A_{0,0})\,(1+A_{1,0}) \\ &\overset{fd}{\approx} \frac{U_{0,0}}{2}\,(1+A_{0,0}+A_{1,0}) \end{aligned} \qquad (22)$$

Thus, for a fixed $U_{0,0}$ at the coarsest scale, to a first-order approximation, the MWM is *additive* at the coarse scales provided the random variables $A_{j,k}$ are small in amplitude. Moreover, the $A_{j,k}$ are approximately *Gaussian* for these low-variance (high-p) symmetric β multipliers [39]. Hence, coarse-resolution MWM outputs will exhibit an additive, Gaussian-like behavior consistent with that of the previously justified ON/OFF models and notions of client behavior as a superposition of sources [7,8].

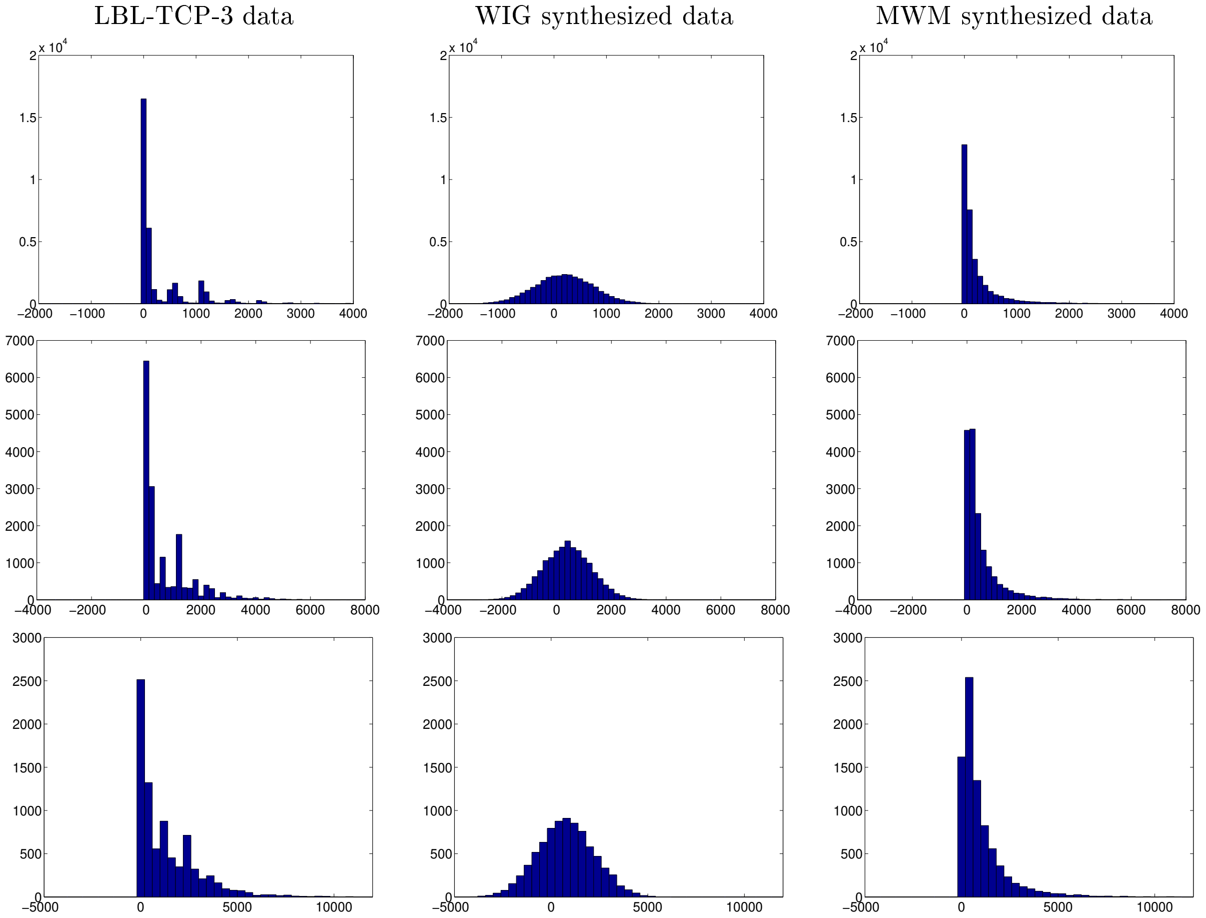
5.3 Matching of Real Data

In order to study how well the MWM and WIG models can match real data, we train them on the the real data traces. To fit the WIG and MWM models to the data, we use the procedure outlined in Sections 3.2 and 4.3, which involves taking a Haar wavelet transform of the real data and estimating the variances σ_j^2 of the wavelet coefficients at each scale. We estimate these variances only at the 15 finest scales, because at coarser scales there are not a sufficient number of coefficients to obtain good variance estimates. As a result, we synthesize data traces of maximum length 2^{15} data points. For both the MWM and WIG, we model the coarsest-scale scaling coefficient $U_{0,0}$ as a Gaussian random variable with mean and variance equal to the sample mean and variance of the scaling coefficients of the real data at this scale. With trained models in hand, we now generate synthetic data traces.

Due to space constraints, we present fitting results only for the LBL-TCP-3 trace. Recall from Figure 1 that visually the synthetic MWM looks very similar to the real trace. We compare the marginals of MWM and WIG traces to that of the LBL-TCP-3 trace at three different aggregation levels. From Figure 5 observe that the MWM marginals are similar to that of the real data trace, while the Gaussian WIG marginals differ significantly. We also observe that the WIG traces have a considerable number of negative points, a result of the low mean and high standard deviation of the real data trace.

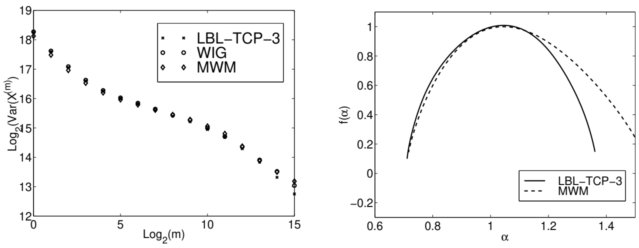
We next compare the correlation matching abilities of the two models. We compare the variance-time plots of the real data, the MWM traces, and the WIG traces in Figure 6(a). The variance-time plot estimates were obtained by averaging the empirical variance-time plots of 32 independent realizations of the models. We observe that, as expected, both the MWM and WIG models do a good job of matching the correlation structure of the real data.

We plot the multifractal spectra (see Section 6) of the LBL-TCP-3 data and the synthetic MWM trace in Figure 6(b) (calculations for the negative moments of the WIG data become numerically unstable and hence the spectra for the WIG is not included). We observe that the spectra match extremely well except for large values of α. This corresponds to a close match of the scaling of higher-order moments, but a somewhat less accurate match of the scaling of the negative moments.



(a) (b) (c)

Figure 5: *Histograms of the bytes-per-times process at different aggregation levels for (a) wide-area TCP traffic at the Lawrence Berkeley Laboratory (trace LBL-TCP-3) [14], (b) one realization of the WIG model, and (c) one realization of the MWM synthesis. The top, middle and bottom plots correspond to bytes arriving in intervals of 6 ms, 12 ms and 24 ms respectively. Note the large probability mass over negative values for the WIG model.*



(a) (b)

Figure 6: *(a) Variance-time plot of the LBL-TCP-3 data "×", the WIG-data "◇", and one realization of the MWM synthesis "○". (b) Multifractal spectra of the LBL-TCP-3 data and one realization of the MWM synthesis.*

5.4 Queuing results

Much effort has been exerted studying the effect of the correlation structure on queuing performance [2,17–19]. Gaussian models that capture the correlation structure of traffic have been proposed [1, 15] and theoretical results for the tail queue probability have been obtained [9, 11, 40]. These models are in many cases appropriate for modeling data traffic. For example, the WIG model has been shown to capture the queuing behavior of video traffic well [15]. For the WAN/LAN data traces that we consider here, however, this is not the case.

The multiplicative structure of the MWM captures both the correlations of the real data as well as the higher-order moments. With its inherent approximately lognormal marginal distribution, it also comes close to matching the marginals of the real data traces.

Intuitively, the more traffic characteristics a model matches, the better will it match the queuing behavior of real traffic. Hence, it is not surprising that a perfect fitting of second-order correlations *and* marginals as done in [22] leads to a good match of queueing behavior.

Here, we take a different approach comparing two simple and quite related models in their ability to capture the queueing behavior of the two real data sets. With this experiment we hope to shed some light on the impact of marginals and higher-order correlations on queuing behavior.

In all experiments, data traces are fed as input to an infinite length single-server queue with link capacity 800 bytes/unit time. We estimate the tail queue probabilities of the various data traces as

$$\widehat{P}_i[Q > x] = \frac{\text{number of time instants } Q > x}{\text{total time duration of trace } i}. \qquad (23)$$

We also provide confidence intervals with confidence level of 90% for the estimated queue distribution $(1/L)\Sigma_{i=1}^{L}\widehat{P}_i[Q > x]$, where L is the total number of traces, assuming that it is a Gaussian random variable [41].

With both real traces, we performed the same queuing experiment. We first trained the MWM and WIG models on the real traces as described in Section 5.3. We then synthesized 480 MWM and WIG traces of length 2^{15}, fed them as input to our theoretical queue and obtained their queuing behavior. Recall that both the WIG and MWM capture the mean, variance and correlation structure of the real data.

In Figure 7(a) we compare the average queuing behavior of the MWM and WIG traces to that of the real trace LBL-TCP-3. We observe that the MWM traces match the queuing behavior of the real data trace much better than the WIG traces. From Figure 5 we notice that the WIG data traces have a considerable number of negative data points. This is because the LBL-TCP-3 data set has a large ratio of standard deviation to mean, which when modeled by a Gaussian process leads to a large fraction of data points going negative. In order to test whether these negative values are the cause for the poor performance of the WIG model, we set negative values to zero in the WIG traces and obtained the queuing behavior of these new traces. We call the new data traces WIG+. We see from Figure 7(a) that the queuing performance of the WIG+ traces is not substantially better than that of the WIG traces.

Thus, we conclude that the Gaussian WIG traces do not give a good approximation to the queuing behavior of the real data set in spite of capturing the correlation structure of the real data trace. Furthermore, the ad hoc procedure of setting all negative values to zero does not improve matters. In fact, the ad hoc procedure used in creating the WIG+ data traces destroys the statistics of the traces. Other ad hoc procedures like excluding all negative data points or setting all negative points to their absolute value also destroy the statistics of the traces. This reveals some of the problems associated with Gaussian models for modeling traffic with marginals similar to those of the real data traces considered here.

The results for the real trace BC-pAug89 are shown in Figure 7(b) and are similar to those for the LBL-TCP-3 trace. Clearly, the MWM again performs far better than the WIG model in capturing the queuing behavior of the real data.

These queuing experiments indicate that the correlation structure of traffic is not the only factor that decides the queuing behavior of data traffic. Since the MWM outperforms the WIG model in matching queuing behavior, we conclude that the additional traffic characteristics of real data captured by the MWM, like marginals and higher order moments, have a substantial effect on the queuing behavior of traffic with statistics similar to the real data sets that we considered here.

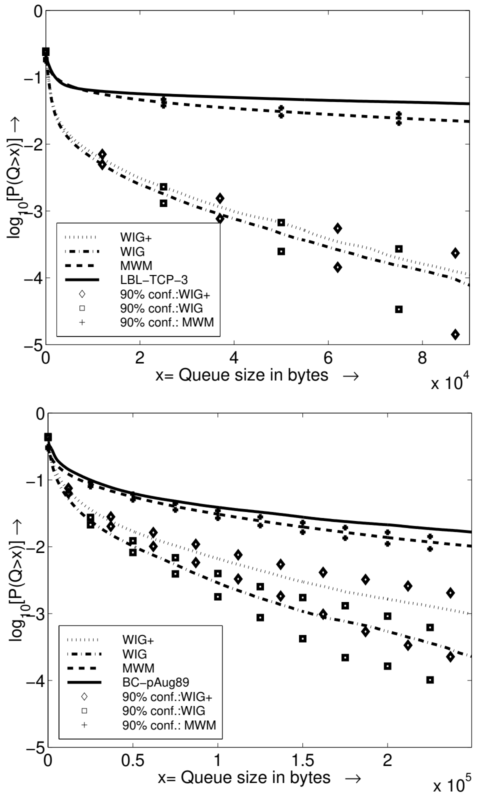


Figure 7: *Comparison of the queuing performance of real data traces with those of synthetic WIG and MWM traces. In (a), we observe that the MWM synthesis matches the queuing behavior of the LBL-TCP-3 data closely, while the WIG synthesis does not. Even when negative values of the WIG data are set to 0 (WIG+), the WIG traces do not come close to matching the correct queuing behavior. In (b), we observe a similar behavior with the BC-pAug89 data.*

6 MWM is a Cascade

We now link the MWM with the theory of cascades. The technical details in this section are not necessary for understanding or applying the MWM and can be omitted on a first reading. Multiplicative cascades generalize the self-similarity of fractal models such as fGn and fBm by offering greater flexibility and richer scaling properties, including burstiness and scaling of higher-order moments [25, 29]. Identifying the MWM algorithm with a multiplicative cascade allows us to benefit from the accumulated theoretical and practical knowledge of the field of *multifractals*, including a precise understanding of the convergence of the MWM algorithm, properties of the marginal distributions, advantages over monofractal fGn models, and a range of possible refinements and extensions. For these reasons, we find it useful to examine the MWM within the context of cascades and multifractals.

The backbone of a cascade is a construction where one starts at a coarse scale and develops details of the process on finer scales iteratively in a multiplicative fashion. The MWM is one such cascade, as (15) and (16) reveal. In accordance with the notation for cascades, setting

$$M_0^0 = U_0^0 \quad \text{and} \quad M_{k_i}^i = \frac{(1 + (-1)^{k'_{i-1}} A_{i-1,k_{i-1}})}{2}, \quad 0 < i \le n, \tag{24}$$

and substituting into (16) leads us to (see Figure 8(a))

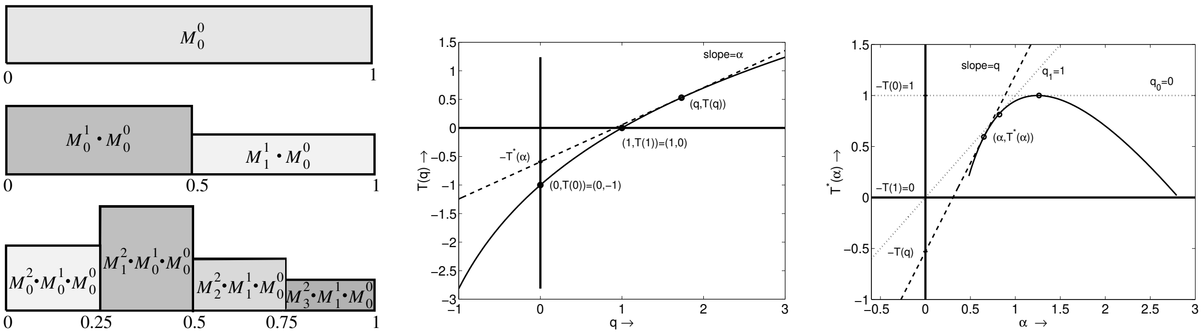
$$C[k] = 2^{-n} M_0^0 \prod_{i=1}^{n} M_{i,k_i}, \tag{25}$$

with the k_i and k'_i defined in the same way as for (16).

Our aim in this section is to both introduce and give an intuitive understanding of cascades to the reader. After studying the nature of the MWM's marginals, we compare cascades with Gaussian LRD processes such as the WIG. As already hinted in the introduction, cascades such as the MWM are ideal for modeling burstiness. We explain this here by developing the multifractal formalism (for further details, see [29]).

6.1 Lognormal marginals

Multiplicative structures, in particular the product representation (25), naturally lead to lognormal marginals. If the $M_{k_i}^i$ are all positive and identically distributed, then $C[k]$ will be approximately lognormal by the CLT. Figure 5 shows that Gaussian modeling seems unfit in this network scenario; various other authors make a case for marginal distributions, including the lognormal, with tails that are much heavier than the Gaussian [20, p 364], [21]. We do not claim that the lognormal is appropriate for all traffic



(a) (b) (c)

Figure 8: *(a): The MWM translates immediately into a multiplicative cascade in the time domain (cf. (25)). (b) $\mapsto$ (c): We demonstrate the Legendre transform $T \mapsto T^*$ in the simple case of concave, differentiable functions such as the spectra of a typical MWM ((33) with $p = 1.66$, $H = .85$). Set $\alpha = T'(q)$, then $T^*(\alpha)$ is such that the tangent at $(q, T(q))$ passes through $(0, -T^*(\alpha))$. In other words, $-T^*(\alpha) + q\alpha = T(q)$. By symmetry, the tangent at $(\alpha, T^*(\alpha))$ has slope q and passes through $(0, -T(q))$. There are two notable special values of q. Trivially, $T(0) = -1$, whence the maximum of T^* is 1. In addition, every positive increment process has $T(1) = 0$, whence $T*$ touches the bisector.*

at all scales, and for a limited number of scales a cascade signal can behave differently from a lognormal. However, this link between cascades and useful marginal models for traffic points to the viability of cascades for providing realistic traffic models.

6.2 Cascades vs. fGn

There is a fundamental difference between cascade modeling and modeling via self-similar processes such as fGn or the WIG, which treat traffic as a mean rate superimposed with fractal noise. Additive self-similar models "hover" around the mean with occasional outbursts in both positive and negative directions, while multiplicative cascades "sit" just above the zero line and emit occasional positive jumps or spikes. In mathematical terms this distinction is best captured by examining negative moments: for self-similar models, these are the negative moments of the fractal noise, hence they capture uninterestingly small variations around the mean; for cascades, on the other hand, these are the negative moments of the process itself, so they capture unnaturally small values and provide useful information.

6.3 Measuring burstiness

For the ease of notation let $k_n 2^{-n} \to t$ mean that $t \in [k_n 2^{-n}, (k_n + 1)2^{-n})$ and $n \to \infty$. The strength of growth, also called the degree of *Hölder continuity,* at time t of a process $Y(t)$ (that corresponds to $D(t)$ of the MWM) with positive increments can be characterized by

$$\begin{aligned} \alpha(t) &= \lim_{k_n 2^{-n} \to t} \alpha^n_{k_n} \qquad \text{where} \\ \alpha^n_{k_n} &\equiv -\frac{1}{n} \log_2 \left| Y((k_n+1)2^{-n}) - Y(k_n 2^{-n}) \right|. \end{aligned} \tag{26}$$

The smaller the $\alpha(t)$, the larger the increments of Y, and the "burstier" it is at time t. The frequency of occurrence of a given strength α, as visible from an analysis on coarse scales can be measured by the *multifractal spectrum*:

$$\begin{aligned} f(\alpha) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log_2 \#\{k_n &= 0, \ldots, 2^n - 1 \\ &: \alpha^n_{k_n} \in (\alpha - \varepsilon, \alpha + \varepsilon)\}. \end{aligned} \tag{27}$$

By definition, f takes values between 0 and 1 and is often shaped like a $\cap$ and concave, but not always. The smaller $f(\alpha)$ is, the "fewer" points t will show $\alpha(t) \simeq \alpha$. If $\overline{\alpha}$ denotes the value $\alpha(t)$ assumed by "most" points t then $f(\overline{\alpha}) = 1$.

Note that this analysis via increments (26) is sufficient provided $Y(t)$ has no polynomial trends. If, on the other hand, polynomial terms are present, then the increment-analysis will yield $f(\alpha) = 1$ for $\alpha = k \in \mathbb{N}$ where k is the order of the first non-vanishing derivative of Y. Then one has to eliminate the polynomial influence, a) via wavelets or, b) by subtracting the trend if known. The known trend for self-similar processes is none other than the "mean arrival rate".

It is, therefore, important to mention that our analysis of real traces in Section 5 shows no integer scaling exponent $\alpha(t)$, except for $\alpha(t) = 1$ for a small number of t, that is, $f(1) < 1$. Thus, we conclude that polynomial trends are not present in the real traffic traces studied here. Since we did not remove any trend from the real data prior to our analysis, this result suggests that the data is not well characterized by self-similar models.

6.4 Higher-order moments and the MF spectrum

Cascades such as the MWM possess rich multifractal spectra. Unlike cascades, the strong self-similarity of the fBm (1) forces it to have a trivial multifractal behavior. To be precise, for the fBm, $\alpha(t) = H$ for all t. To demonstrate this, we will use information about the *scaling* of higher-order moments of the two types of processes to obtain their multifractal spectra.

Let us define

$$\begin{aligned} T(q) &\equiv \lim_{n \to \infty} \frac{1}{-n} \log_2 \mathbb{E}\left[S_n(q)\right], \qquad \text{where} \\ S_n(q) &\equiv \sum_{k_n=0}^{2^n-1} \left| Y((k_n+1)2^{-n}) - Y(k_n 2^{-n}) \right|^q \\ &= \sum_{k_n=0}^{2^n-1} 2^{n\alpha^n_{k_n}}. \end{aligned} \tag{28}$$

Note that T is always concave, since $\log_2 \mathbb{E} S_n(q)$ is concave. For a typical plot of T and f see Figure 8 (b) and (c).

The multifractal spectrum $f(\alpha)$ and $T(q)$ are closely related as the following instructive hand-waving argument shows. Grouping in the sum $S_n(q)$ of (28) the terms according to $\alpha^n_{k_n} \simeq \alpha$, and using

(27) we get

$$\begin{aligned} S_n(q) &= \textstyle\sum_\alpha \sum_{\alpha_n \sim \alpha} (2^{-n\alpha})^q \simeq \sum_\alpha 2^{n f_G(\alpha)} 2^{-nq\alpha} \\ &\simeq 2^{-n \inf_\alpha (q\alpha - f_G(\alpha))}. \end{aligned} \tag{29}$$

We conclude that we must "expect" $T(q)$ to equal $\inf_\alpha(q\alpha - f_G(\alpha))$. For the special case of an MWM process, i.e., $Y = D$, it can be shown (see [42]) that the dual relation holds. This relation is called the *multifractal formalism* and reads

$$f(\alpha) = T^*(\alpha) := \inf_q (q\alpha - T(q)). \tag{30}$$

Simple calculus shows that $T^*(\alpha) = q\alpha - T(q)$ at $\alpha = T'(q)$ provided $T''(q) < 0$. This relation via the *Legendre transform* T^* is typical of the theory of large deviations [43]. The goal there is to establish relations such as (30) under most general assumptions. To use the correct terminology, f is the *rate function* of a so-called *large deviation principle* (LDP): it measures how frequently or how likely the observed $\alpha^n_{k_n}$ deviates from the "expected value" $\overline{\alpha}$.

In order to estimate $T(q)$ from data, it is customary to use the approximation $2^{-nT(q)} \simeq S_n(q)$. For the MWM this is equivalent to

$$2^{-jT(q)} \sim \sum_{k=0}^{2^j-1} |2^{-j/2} U_{j,k}|^q. \tag{31}$$

Any linear fit of $\log S_{(j)}(q)$ against j will give the slope $T(q)$.

Let us calculate $T(q)$ for the MWM model, i.e., $Y = D$. Using independence of the multipliers $M^i_{k_i}$ and denoting by $\sum'$ the sum over all $k_n = 0, ..., 2^n - 1$ we find

$$\begin{aligned} \mathbb{E}[S_n(q)] &= \sum{}' \mathbb{E}(M^n_{k_n})^q \cdots \mathbb{E}(M^1_{k_1})^q \cdot \mathbb{E}(M^0_0)^q \\ &= \sum{}' \mathbb{E}(M^{(n)})^q \cdots \mathbb{E}(M^{(1)})^q \cdot \mathbb{E}(M^0_0)^q \\ &= \mathbb{E}(M^0_0)^q \cdot 2^n \cdot \prod_{i=1}^n \mathbb{E}\left(M^{(i)}\right)^q. \end{aligned} \tag{32}$$

In the second step we made use of the fact that the multipliers $M^i_{k_i}$ are identically distributed to $M^{(i)}$. To this we add the fact that the moments of the $M^{(i)}$ converge to the ones of the limiting random variable M for the next equation, and end by assuming that $M = (1 + A)/2$ with A being β-distributed as in (19) to obtain:

$$\begin{aligned} \text{MWM:} \quad T(q) &= -1 - \log_2 \mathbb{E}[M^q] \qquad (33) \\ &= -1 - \log_2 \frac{\Gamma(p+q)\Gamma(2p)}{\Gamma(2p+q)\Gamma(p)} \text{ if } q > -p, \end{aligned}$$

and $T(q) = -\infty$ if $q \le -p$.

The function $T(q)$ is a simple statistical description of the process that captures marginal information, but which also governs the "burstiness" through the multifractal formalism. It must be emphasized here that the multifractal parameters $T(q)$ of the MWM process do not necessarily imply that the process cannot be modeled parsimoniously. For example, in the case of the MWM, the β-distributions for the multipliers are controlled by the parameters p_j (20). If one replaced the right side of (20) by the powerlaw for fGn then all values $T(q)$ would be determined by H [29].

Now let us compute $T(q)$ for the self-similar fBm. From (1) we find

$$\begin{aligned} &\mathbb{E} \sum_{k_n=0}^{2^n-1} |B((k_n+1)2^{-n}) - B(k_n 2^{-n})|^q \\ &= 2^n \mathbb{E}\left[|B(2^{-n})|^q\right] = 2^{n-nqH} \mathbb{E}\left[|B(1)|^q\right] \end{aligned} \tag{34}$$

which yields for fBm

$$\text{fBm:} \qquad T(q) = \begin{cases} qH - 1 & \text{for } q > -1, \\ -\infty & \text{for } q \le -1. \end{cases} \tag{35}$$

This is probably the most compact way to express the monofractal character of fBm: taking the Legendre transform of T shows that fBm possesses only one degree of "burstiness" ($\alpha(t) = H$) which is omnipresent (compare also (3)).

6.5 Multifractal scaling of moments and LRD

The multifractal scaling exponent $T(2)$ of a process Y is closely related to the LRD parameter H, since both measure the power-law behavior of some second-order statistics. More precisely, $T(2)$ measures the scaling behavior of the second sample moments (28), while H (5) can be estimated from the scaling of the sample variance.

For a process Y with zero-mean increments Z this can be made precise. Let us denote resolutions by $m = 2^{-n}$. We define $Z^{(m)}$ such that $mZ^{(m)}$ is the increment of Y at resolution m. From Section 2.1 we find first that $\mathrm{var}(Z^{(m)}) \simeq m^{2H-2}$. Next, (28) gives (28) gives $2^{(-n)T(2)} \simeq \mathbb{E} \sum_{k=0\dots 2^n-1} |mZ^{(m)}_k|^2 = 2^n 2^{2(-n)} \mathbb{E}|Z^{(m)}|^2 = \mathrm{const} \cdot 2^{-n} \mathrm{var}(Z^{(m)})$. Comparing the exponents of 2^{-n} we find $T(2) = 1 + (2H - 2)$, or,

$$H = \frac{T(2)+1}{2} \qquad \text{for zero-mean processes.} \tag{36}$$

This is in agreement with the theoretical formulas (33) and (35) for the spectra of MWM and fBm, respectively.

For multifractal measures such as the MWM a first difficulty in establishing a relation between H and $T(2)$ arises from the fact that these processes are not second-order stationary. So, LRD cannot be defined as usual through the decay of the auto-correlations. However, alternative fractal properties, such as the decay of aggregate variances (Section 2.1) or wavelet coefficients (Section 3.2), which are equivalent to LRD in the presence of second-order stationarity, can still be defined and calculated, leading to the same result (36).

7 Conclusions

The MWM provides a new multiscale tool for synthesis of nonGaussian LRD traffic. Computations involving the MWM are extremely efficient — synthesis of a trace of N sample points requires only $O(N)$ computations. In fact, synthesis of a trace of length 2^{18} data points takes just 8 seconds of workstation cpu time. The parameters of the MWM, numbering approximately $\log N$, are identical in number to the WIG model and are simple enough to be either inferred from observed data or chosen a priori. We can reduce the number of parameters further by developing a parametric characterization of the wavelet energy decay across scale.

With the MWM and WIG models, we have been able to fit actual traffic traces, and have developed preliminary queueing results that demonstrate the importance of the nonGaussian nature (including scaling of higher-order moments) of traffic in determining queueing performance.

Apart from being a useful tool for fast synthesis of realistic data traffic, the MWM is a promising analysis tool for the network researcher. Further research could make the MWM viable for data prediction. The parameters of the MWM could also be used to capture the effect of different protocols on shaping data traffic (e.g., the TCP protocol). In short, the use of the MWM in real-time network protocols and control algorithms seems very promising.

References

[1] 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.

[2] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent traffic," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 209–223, April 1996.

[3] M. Taqqu and J. Levy, *Using renewal processes to generate LRD and high variability.* In: Progress in probability and statistics, E. Eberlein and M. Taqqu eds., vol. 11. Birkhaeuser, Boston, 1986. pp 73–89.

[4] N. Likhanov, B. Tsybakov, and N. Georganas, "Analysis of an ATM buffer with self-similar input traffic," *Proc. IEEE, Info com '95 (Boston 1995)*, pp. 985–992, 1995.

[5] F. Brichet, J. Roberts, A. Simonian, and D. Veitch, "Heavy traffic analysis of a fluid queue fed by a superposition of ON/OFF sources," *COST*, vol. 242, 1994.

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

[7] M. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic. Evidence and possible causes," in *Proceedings of SIGMETRICS '96*, May 1996.

[8] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, "Self-similarity through high-variability: Statistical analysis of ethernet LAN traffic at the source level," *IEEE/ACM Trans. Networking (Extended Version)*, vol. 5, pp. 71–86, Feb. 1997.

[9] J. Choe and N. Shroff, "Supremum distribution of gaussian processes and queueing analysis including long-range dependence and self-similarity," *Stochastic Models* submitted, 1997.

[10] I. Norros, "On the use of fractional Brownian motion in the theory of connectionless networks," *COST*, vol. 242, 1994.

[11] N. Duffield and N. O'Connell, "Large deviations and overflow probabilities for the general single-server queue, with applications," *Math. Proc. Cambr. Phil. Soc.*, vol. 118, pp. 363–374, 1995.

[12] I. Norros, "Four approaches to the fractional Brownian storage," *Fractals in Engineering*, pp. 154–169, 1997.

[13] G. Gripenberg and I. Norros, "On the prediction of fractional Brownian motion," *preprint ilkka.norros@vtt.fi*, 1995.

[14] V. Paxson and S. Floyd, "Wide-area traffic: The failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226–244, 1995.

[15] S. Ma and C. Ji, "Modeling video traffic in the wavelet domain," in *Proc. of 17th Annual IEEE Conf. on Comp. Comm., INFOCOM*, pp. 201–208, Mar. 1998.

[16] N. Duffield, "Economies of scale for long-range dependent traffic in short buffers," *Telecommunication Systems,* to appear, 1998.

[17] B. K. Ryu and A. Elwalid, "The Importance of Long-range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities," *Proc. ACM SIGCOMM Conf.*, vol. 26, no. 4, pp. 3–14, 1996.

[18] D. P. Heyman and T. V. Lakshman, "What are the implications of long-range dependence for VBR-video traffic engineering?," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 301–317, June 1996.

[19] A. Neidhardt and J. Wang, "The concept of Relevant Time Scales and its application to queuing analysis of self-similar Traffic," *Proc. SIGMETRICS '98/PERFORMANCE '98*, pp. 222–232, 1998.

[20] J. Roberts, U. Mocci, and J. V. (eds.), "Broadband network teletraffic," in *Lecture Notes in Computer Science, No 1155*, Springer, 1996.

[21] S. Bates and S. McLaughlin, "The estimation of stable distribution parameters from teletraffic data," *preprint*, 1998.

[22] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. Kaye, "Modeling and simulation of self-similar VBR compressed video: a unified approach," *Computer-Communication-Review*, vol. 25, pp. 114–125, Oct. 1995.

[23] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. Kaye, "Fast simulation of self-similar traffic in ATM networks," *ICC '95 Seattle*, vol. 1, pp. 438–44, 1995.

[24] M. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for long-range dependence: An empirical study," *Fractals.*, vol. 3, pp. 785–798, 1995.

[25] R. Riedi and J. L. Véhel, "Multifractal properties of TCP traffic: A numerical study," *Technical Report No 3129, INRIA Rocquencourt, France*, Feb, 1997. Available at `www.dsp.rice.edu`.

[26] P. Mannersalo and I. Norros, "Multifractal analysis of real ATM traffic: a first look," *COST257TD*, 1997.

[27] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic," *Proc. ACM/Sigcomm 98*, vol. 28, pp. 42–55, 1998.

[28] A. C. Gilbert, W. Willinger, and A. Feldmann, "Scaling analysis of random cascades, with applications to network traffic," *IEEE Trans. on Info. Theory, (Special issue on multiscale statistical signal analysis and its applications)*, vol. 45, April 1999.

[29] R. H. Riedi, M. S. Crouse, V. Ribeiro, and R. G. Baraniuk, "A multifractal wavelet model with application to network traffic," *IEEE Trans. Info. Theory, (Special issue on multiscale statistical signal analysis and its applications)*, vol. 45, April 1999. Available at www.dsp.rice.edu.

[30] D. R. Cox, "Long-range dependence : A review," *Statistics : An Appraisal*, pp. 55–74, 1984.

[31] P. Abry, P. Gonçalvès, and P. Flandrin, "Wavelets, spectrum analysis and $1/f$ processes," *preprint*, 1996.

[32] B. V. Rao, K. R. Krishnan, and D. P. Heyman, "Performance of Finite-Buffer Queues under Traffic with Long-Range Dependence," *Proc. IEEE GLOBECOM*, vol. 1, pp. 607–611, November 1996.

[33] M. Paulekar and A. M. Makowski, "Tail probabilities for a multiplexer with self-similar traffic," *Proc. IEEE INFOCOM*, pp. 1452–1459, 1996.

[34] M. Grossglauser and J.-C. Bolot, "On the relevance of long-range dependence in network traffic," *Computer-Communication-Review*, vol. 26, pp. 15–24, October 1996.

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

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

[37] P. Abry and D. Veitch, "Wavelet analysis of long range dependent traffic," *IEEE Trans. Inform. Theory*, vol. 4, no. 1, pp. 2–15, 1998.

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

[39] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1-2. New York: John Wiley & Sons, 1994.

[40] I. Norros, "A storage model with self-similar input," *Queueing Systems*, vol. 16, pp. 387–396, 1994.

[41] R. Jain, *The Art of Computer Systems Performance:Techniques for experimental design, measurement, simulation, and modeling.* John Wiley & Sons, Inc., 1991.

[42] R. H. Riedi, "Multifractal processes," *IEEE Info. Theory*, submitted 1999.

[43] J.-D. Deuschel and D. W. Stroock, *Large Deviations.* Academic Press, 1984.