Network Traffic Modeling using a Multifractal Wavelet Model

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

Abstract. In this paper, we develop a simple and powerful multiscale model for synthesizing 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.

1. Introduction

Fractal models arise frequently in a variety of scientific disciplines, such as physics, chemistry, astronomy, and biology. More recently, fractal models have had a major impact on the analysis of data communication networks such as the Internet. In their landmark paper [1], Leland et al. demonstrated that network traffic loads exhibit 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.

As the pre-eminent random fractal model, fractional Brownian motion (fBm) has played a central rôle in many fields [1, 2]. FBm is the unique Gaussian process with stationary increments and the following scaling property for all $a > 0$

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

| LBL-TCP-3 data | WIG synthesis | MWM synthesis |

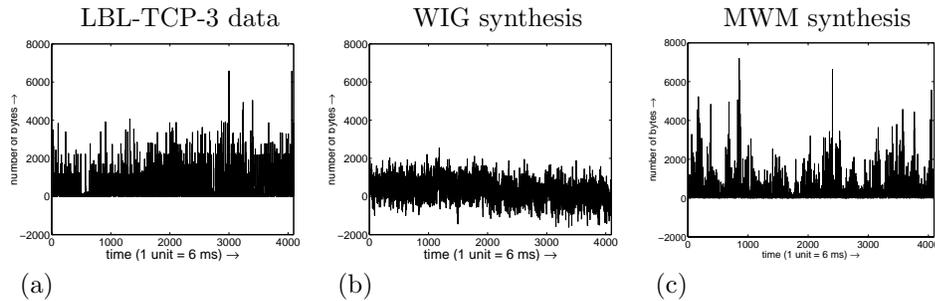(a)   (b)   (c)

FIGURE 1. Bytes-per-time arrival process for (a) wide-area TCP traffic at the Lawrence Berkeley Laboratory (trace LBL-TCP-3) [4], (b) one realization of the state-of-the-art wavelet-domain independent Gaussian (WIG) model [3], and (c) one realization of the multifractal wavelet model (MWM) synthesis. The MWM traces closely resemble the real data, while the WIG traces (with their large number of negative values) do not.

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. Gaussian processes with a more flexible scaling relation than (1) can be synthesized by the wavelet-domain independent Gaussian (WIG) model [3].

Real network traffic traces, however, do not exhibit the strict self-similarity of (1) and are positive and nonGaussian in nature thus limiting the use of fBm as a traffic model. The transmission control protocol (TCP) traffic we study in this paper exhibits local scaling similar to (2), but with an exponent H_t that depends on t. This has been termed *multifractal behavior* and was reported for the first time in [5] and subsequently in [6, 7, 8, 9]. 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.

In this paper, we propose a new non-linear model for network traffic data. The *multifractal wavelet model* (MWM) is based on a multifractal cascade in the wavelet domain that by design guarantees a positive output. Each sample of the MWM process is obtained as a product of several positive independent random variables resulting in an approximately lognormal marginal density.

Fitting the MWM to real traffic traces results in an excellent match, far better than the Gaussian 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.

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

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 [10]

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

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 property is called 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 $X(n) \overset{fd}{=} m^{1-H} X^{(m)}(n)$.

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 [1].

3. Wavelets and LRD Processes

3.1. Wavelet transform

The discrete wavelet transform is a multi-scale signal representation of the form [11]

$$c(t) = \sum_k u_k \, 2^{-J_0/2} \, \phi\big(2^{-J_0}t - k\big) + \sum_{j=-\infty}^{J_0} \sum_k w_{j,k} \, 2^{-j/2} \, \psi\big(2^{-j}t - k\big), \; j,k \in \mathbf{Z}$$
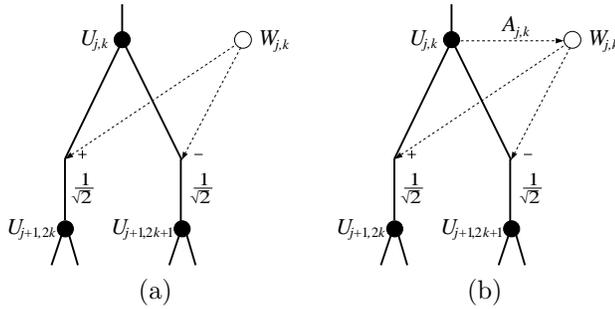
FIGURE 2. (a) WIG construction: Generate the $W_{j,k}$'s as mutually independent and identically distributed within scale according to $W_{j,k} \sim N(0, \sigma_j^2)$. Then compute the Haar scaling coefficients $U_{j+1,2k}$ and $U_{j+1,2k+1}$ at scale $j+1$ as sums and differences of the scaling and wavelet coefficients $U_{j,k}$ and $W_{j,k}$ at scale j (normalized by $1/\sqrt{2}$). (b) MWM construction: At scale j, generate the multiplier $A_{j,k} \sim \beta(p_j, p_j)$, form the wavelet coefficient as the product $W_{j,k} = A_{j,k}U_{j,k}$ and them form scaling coefficients at scale $j+1$ as in (a).

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. Using filter bank techniques, the wavelet transform and inverse wavelet transform can be computed in $O(N)$ operations for a length-N

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

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

The Haar scaling and wavelet coefficients can be recursively computed via [11]

$$u_{j-1,k} = 2^{-1/2}(u_{j,2k} + u_{j,2k+1}), \ w_{j-1,k} = 2^{-1/2}(u_{j,2k} - u_{j,2k+1}). \tag{6}$$

3.2. Modeling LRD data

Wavelets serve as an approximate Karhunen-Loève or decorrelating transform for fBm [2], fGn, and more general LRD signals [12]. Hence, modeling and processing of these signals in the wavelet domain is often more efficient and powerful than in the time domain.

Gaussian LRD processes can be approximately synthesized by generating wavelet coefficients as independent zero-mean Gaussian random variables, identi-

cally 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 [3]. We call the resulting model the wavelet-domain independent Gaussian (WIG) model [3] (see Figure 2(a)). A power-law decay for the σ_j^2's leads to approximate wavelet synthesis of fBm or fGn [2].

The WIG model assumes Gaussianity even though network traffic signals (such as loads and interarrival times) are positive and can be highly nonGaussian. We seek a more accurate marginal characterization for these spiky, non-negative LRD processes.

3.3. Modeling non-negative data with the Haar wavelet

It is easily shown that the Haar scaling coefficients $u_{j,k}$ are non-negative if and only if the signal itself is non-negative; that is, $c(t) \geq 0 \Leftrightarrow u_{j,k} \geq 0, \ \forall \ j, k$. Solving (6) for $u_{j,2k}$ and $u_{j,2k+1}$, we find

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

Combining (7) with the constraint $u_{j,k} \geq 0$, we obtain the condition

$$c(t) \geq 0 \Leftrightarrow |w_{j,k}| \leq u_{j,k}, \ \forall \ j, k. \tag{8}$$

4. Multifractal Wavelet Model

The positivity constraints (8) on the Haar wavelet coefficients suggest a very simple multiscale, multiplicative signal model for positive processes. In the *multifractal wavelet model* (MWM) – a name which we will explain in a moment – we compute the wavelet coefficients recursively by

$$W_{j,k} = A_{j,k} \ U_{j,k}, \tag{9}$$

with $A_{j,k}$ a random variable supported on the interval $[-1, 1]$. Together with (7), we obtain (see Figure 2(b))

$$U_{j,2k} = 2^{-1/2}(1 + A_{j+1,k}) \, U_{j-1,k}, \ U_{j,2k+1} = 2^{-1/2}(1 - A_{j+1,k}) \, U_{j-1,k}. \tag{10}$$

Thus, to generate a MWM realization we first synthesize $U_{0,0}$ and then compute scaling coefficients at finer scales recursively using (10) till we reach a desired finest scale. This algorithm synthesizes an N-point using $O(N)$ computations.

The synthesized signal $C[k]$ is a discrete time approximation of the continuous time signal $c(t)$. Decomposing each shift k into a binary expansion $k = \sum_{i=0}^{n-1} k_i' 2^{n-1-i}$, we can write

$$C[k] = 2^{-n/2} U_{n,k} = 2^{-n} \, U_{0,0} \prod_{i=0}^{n-1} \frac{(1 + (-1)^{k_i'} A_{i,k_i})}{2}, \tag{11}$$

with

---

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

$$k_0 \equiv 0, \text{ and } k_i = \sum_{j=0}^{i-1} k_i' 2^{i-1-j}, \ i = 1, \dots, n-1. \tag{12}$$

In our experiments we choose the *symmetric beta distribution*, $\beta(p, p)$ for the $A_{j,k}$'s

$$A_{j,k} \sim \beta(p_j, p_j), \tag{13}$$

with p_j the beta parameter at scale j. We set the p_j's to get the desired decay of the variances of $W_{j,k}$'s.

5. MWM is a Cascade

The MWM is a special case of the rich class of *multiplicative cascades*. Cascades provide a natural framework for producing positive "bursty" processes and offer greater flexibility and richer scaling properties than fractal models such as fGn and fBm. The subtle structure of cascades is best understood in terms of the powerful theory of *multifractals*, a statistical tool for measuring "burstiness" superior to LRD which merely measures "high variability".

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 [5, 9]. For these reasons, we find it useful to examine the MWM within the context of cascades and multifractals.

5.1. Cascades

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, e.g., is a multiplicative cascade: as (10) and (11) reveal we may write

$$C_{\text{MWM}}[k] = 2^{-n} M_0^0 \prod_{i=1}^{n} M_{i,k_i}, \text{ with } M_{k_i}^i = \frac{(1 + (-1)^{k_{i-1}'} A_{i-1,k_{i-1}})}{2}. \tag{14}$$

This construction procedure naturally results in a process that "sits" just above the zero line and emits occasional positive jumps or spikes. In contrast, additive self-similar models such as fGn and the WIG "hover" around the mean with occasional outbursts in both positive and negative directions.

5.2. Multifractal analysis

Intuitively, multifractal analysis measures the frequency with which bursts of different strengths occur in a signal. Consider a positive process $Y(t)$. The strength of the burst of Y at time t, also called the degree of *Hölder continuity*, can be characterized by

$$\alpha(t) = \lim_{k_n 2^{-n} \to t} \alpha_{k_n}^n \text{ where } \alpha_{k_n}^n := -\frac{1}{n} \log_2 \left| Y((k_n + 1)2^{-n}) - Y(k_n 2^{-n}) \right| \quad (15)$$

where $k_n 2^{-n} \to t$ means that $t \in [k_n 2^{-n}, (k_n + 1)2^{-n})$ and $n \to \infty$. The smaller the $\alpha(t)$, the larger the increments of Y around time t, and the "burstier" it is at time t. The frequency of occurrence of a given strength α, can be measured by the *multifractal spectrum*:

$$f(\alpha) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \tfrac{1}{n} \log_2 \#\{k_n = 0, \ldots, 2^n - 1 : \alpha_{k_n}^n \in (\alpha - \varepsilon, \alpha + \varepsilon)\}. \quad (16)$$

By definition, f takes values between 0 and 1 and is often shaped like a $\cap$ and concave. The smaller the $f(\alpha)$, the "fewer" points t will exhibit $\alpha(t) \approx \alpha$. If α_0 denotes the value $\alpha(t)$ assumed by "most" points t, then $f(\alpha_0) = 1$. See Figure 3(b) for the multifractal spectrum of the LBL-TCP-3 data set and of synthetic MWM data. We observe that the MWM captures the spectrum of the real data except for large values of α. This means that the MWM does not generate as many small values as the signal possesses.

5.3. Multifractal spectrum and higher-order moments

Though (16) gives us a simple measure of burstiness in data, in practice it is impossible to compute the right side of (16). However, $f(\alpha)$ can be obtained through the use of high and low-order moments of the signal $Y(t)$.

Define the *partition function* that captures the scaling of different moments of Y as

$$T(q) := \lim_{n \to \infty} \frac{1}{-n} \log_2 \mathbb{E}\left[S_n(q) \right], \quad (17)$$

with

$$S_n(q) := \sum_{k_n=0}^{2^n-1} \left| Y((k_n + 1)2^{-n}) - Y(k_n 2^{-n}) \right|^q = \sum_{k_n=0}^{2^n-1} 2^{-qn\alpha_{k_n}^n}. \quad (18)$$

The multifractal spectrum $f(\alpha)$ and $T(q)$ are closely related, as the following hand-waving argument shows. Grouping in the sum $S_n(q)$ of (18) the terms behaving as $\alpha_{k_n}^n \approx \alpha$, and using (16) we get

$$S_n(q) = \sum_{\alpha} \sum_{\alpha_n \sim \alpha} \left(2^{-n\alpha} \right)^q \approx \sum_{\alpha} 2^{nf(\alpha)} 2^{-nq\alpha} \approx 2^{-n \inf_\alpha (q\alpha - f(\alpha))}. \quad (19)$$

We conclude that we must "expect" $T(q)$ to equal $\inf_\alpha(q\alpha - f(\alpha))$, the so-called *Legendre transform* of $f(\alpha)$. For the special case of an MWM process, i.e., $Y(t) =$
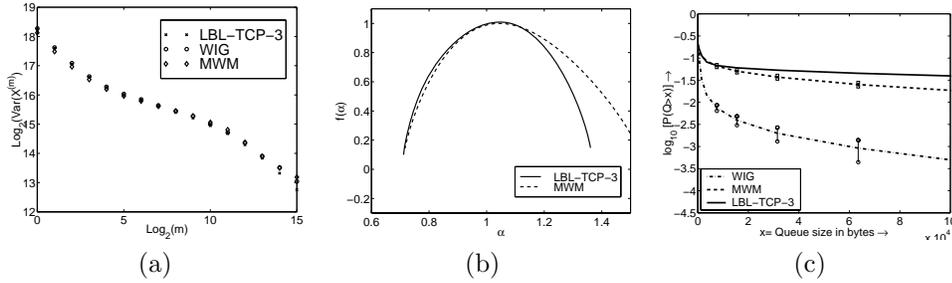
(a)   (b)   (c)

FIGURE 3. (a) Variance-time plot of the LBL-TCP-3 data "×",
the WIGdata "⋄", and one realization of the MWM synthesis
"∘". (b) Multifractal spectra of the LBL-TCP-3 data and one
realization of the MWM synthesis. (c) Comparison of the queu-
ing performance of real data traces with those of synthetic WIG
and MWM traces. Observe that the MWM synthesis matches the
queuing behavior of the LBL-TCP-3 data closely, while the WIG
synthesis does not.

$\int_0^t c(t)\, \mathrm{d}t$, it can be shown (see [13]) that the inverse relation holds, called the
multifractal formalism

$$f(\alpha) = T^*(\alpha) := \inf_q \left(q\alpha - T(q) \right). \tag{20}$$

This relation makes the multifractal spectrum more accessible for practical
purposes.

6. Experimental Results

In order to compare the MWM and WIG models we train them on a well-known
real data trace, the LBL-TCP-3 [4]. We compare the correlation matching abilities
through the variance-time plots of the real data, the MWM traces, and the WIG
traces in Figure 3(a). Observe that, as expected, both the MWM and WIG models
perform well in matching the correlation structure of the real data.

We plot the multifractal spectra (see Section 5) of the LBL-TCP-3 data and
the synthetic MWM trace in Figure 3(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.

In Figure 3(c) 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.

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

[3] 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.

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

[5] 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`.

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

[7] 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.

[8] 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.

[9] 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*, vol. 45, pp. 992–1018, April 1999. Available at `www.dsp.rice.edu`.

[10] 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.

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

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

[13] R. H. Riedi, "Multifractal processes," *Technical Report, ECE Dept. Rice Univ., TR 99-06*. To appear in "Long range dependence: Theory and applications," eds. Doukhan, Oppenheim and Taqqu (2000). Available at `www.dsp.rice.edu/publications`.

Department of Electrical and Computer Engineering
Rice University
6100 South Main Street
Houston, TX 77005, USA
E-mail address: {`mcrouse, riedi, vinay, richb`}`@rice.edu`
URL: `www.dsp.rice.edu.`