ADDITIVE AND MULTIPLICATIVE MIXTURE TREES FOR NETWORK TRAFFIC MODELING

Shriram Sarvotham, Xin Wang, Rudolf H. Riedi, and Richard G. Baraniuk

Department of Electrical and Computer Engineering, Rice University
6100 South Main Street, Houston, TX 77005, USA
Email: {shri, xinwang, riedi, richb}@rice.edu; Web: dsp.rice.edu.

ABSTRACT

Network traffic exhibits drastically different statistics, ranging from nearly Gaussian marginals and Long range dependence at very large time scales to highly non-Gaussian marginals and multifractal scaling on small scales. This behavior can be explained by forming two components of the traffic according to the speed of connections, one component absorbing most traffic and being mostly Gaussian, the other constituting virtually all the small scale bursts. Towards a better understanding of this phenomenon, we propose a novel tree-based model which is flexible enough to accommodate Gaussian as well as bursty behavior on different scales in a parsimonious way.

Keywords: Network traffic modeling, Haar wavelet, Long range dependence, multifractals, fractional Brownian motion, Lévy stable motion.

1. MOTIVATION AND SUMMARY

Network traffic analysis and modeling play a major rôle in characterizing network performance. Models that accurately capture the salient characteristics of traffic are useful for analysis and simulation, and they further our understanding of network dynamics and so aid design and control.

Numerous studies have found that aggregate traffic exhibits *long-range-dependence* (LRD) [1], and that traffic can be extremely bursty, resulting in a non-Gaussian marginal distribution and multifractal properties [2]. These findings are in sharp contrast to classical Markovian type traffic models and their predictions [1, 3, 4] and, therefore, merit close attention.

1.1. Models for Network Traffic

Fractional Gaussian noise (fGn) is a stationary Gaussian process with LRD (see Section 3) which has become very popular as a model of traffic arrivals at a node a communication network such as the internet. It owes its credibility to the fact that it can be obtained as the limit of the superposition of a large number of independent individual ON/OFF sources which transmit data at a given rate during the heavy-tailed ON-phase and no data during the OFF phase [1].

The fGn process can be synthesized using an *additive* tree-based model such as the Wavelet-domain-independent-Gaussian model (WIG) [5, 2]. Although the fGn and the additive WIG model match the LRD of network traffic and fit well at large time-scales, previous work has shown that traffic data is highly non-Gaussian, especially at small time-scales, and exhibits multifractal scaling behavior. *Multiplicative* tree-based cascades such as the Multifractal Wavelet Model (MWM), proposed in [2], can reproduce these properties with accuracy. Indeed, the additive WIG and the multiplicative MWM are both flexible enough to match the variance of network traffic at all scales; however, the multiplicative MWM provides a superior match of marginal distributions, especially at smaller scales [2]. This can be attributed to the high non-Gaussianity of traffic traces which is particularly prominent on time scales of seconds and below.

1.2. Components of Network Traffic

In search of the causes of the apparent non-Gaussianity and the well-documented burstiness of network traffic we revisited the classical ON/OFF model. Within this framework, traffic bursts arise only from a "constructive interference", i.e., large number of connections transmitting data simultaneously. A close look at measured traffic, though, does not confirm this scenario. Rather, in most cases only *one* connection[1] dominates the traffic arrivals during a burst [6].

Thus motivated, we call any connection which sends more than a threshold of bytes during any time interval of a given size T an *alpha connection*. The (large) threshold is chosen based on the mean of the aggregate traffic at time-scale T plus a few standard deviations. We call all bytes sent by alpha connections the *alpha* traffic component. The residual traffic is called the *beta* component.[2] Our procedure thus decomposes an aggregate traffic trace into

$$\text{total traffic} \;=\; \text{alpha traffic} \;+\; \text{beta traffic}. \qquad (1)$$

[1] In this analysis, a connection is identified through source and destination IP address and port number.

[2] By analogy to the dominating *alpha males* and submissive *beta males* observed in the animal kingdom.

This work was supported by NSF, DARPA/AFOSR, DOE, and ONR.

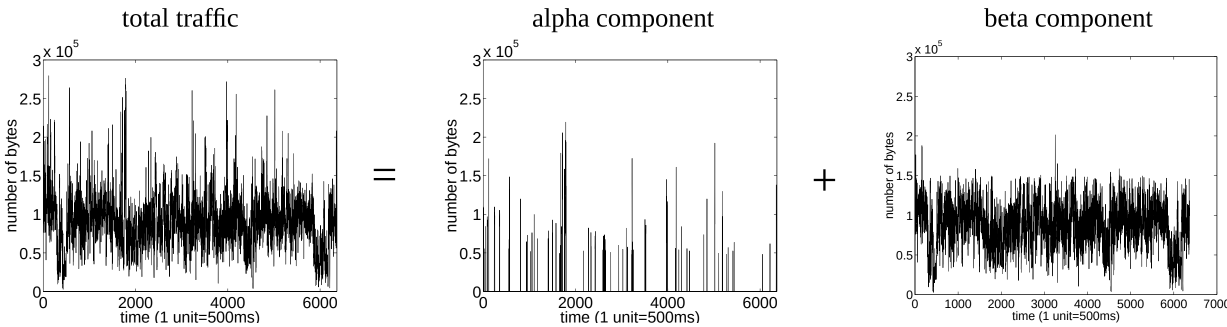

Figure 1: Decomposition of the traffic trace into the sum of a bursty alpha component and an fGn beta component.

We have applied the alpha/beta traffic decomposition to many real-world traffic traces at time scales T varying from 50 to 500 ms and found tremendous consistency in our results [6]. The statistical properties of the components can be summarized as follows.

Beta traffic: At time-scales coarser than the round-trip time, the beta component is very close to Gaussian and strongly dependent, provided a sufficiently large number of connections are present. Moreover, the beta component carries the same fractal scaling (LRD) exponent as the aggregate traffic and can be well approximated by fGn.

Alpha traffic: The alpha component constitutes a small fraction of the total workload but is entirely responsible for the bursty behavior. Alpha traffic is highly non-Gaussian.

It is notable that this decomposition in networking terms (based on connection-level information) achieves a separation in statistical terms. It leads, thus, to a better understanding and modeling of the overall network traffic.

1.3. Multiscale estimation and modeling

Additive models accurately capture Gaussian marginals found in network traffic when aggregated to large time-scales, as well as in the beta component. Multiplicative models, on the other hand, prove ideal to match non-Gaussian, bursty signals such as network traffic at small time-scales, as well as the alpha component. Wavelet based models combine the advantage of the tree structure for fast and simple synthesis with the ability to simulate scaling behavior such as long range dependence, another prominent property of network traffic.

Motivated by the presence of these drastically different statistics of measured network traffic, we develop a new tree-based model based on mixing additive and multiplicative innovations which is able to accommodate Gaussian as well as non-Gaussian bursty statistics parsimoniously. We use *kurtosis*, a fourth-order statistic, to test the closeness of fit of the model to real traffic data. With an interest in fast traffic decomposition for on-line monitoring we also propose a wavelet based separation scheme. After doing so we present a short overview of tree-based multiscale models and proceed to the novel mixture model.

2. WAVELET-BASED TRAFFIC SEPARATION

The computationally intense connection level separation of alpha and beta traffic does not lend itself to massive data processing or on-line monitoring. Approximate separation of alpha and beta traffic can be done using a novel wavelet-based thresholding scheme that does not require explicit connection information. This scheme is based on the fact that we can treat the beta component as "noise" and the alpha component as the "signal", and use well-known denoising techniques to separate the two. We use Wavelet based denoising techniques, with coefficient thresholding. For colored denoising (since beta traffic is colored noise, fGn), we use different thresholds for wavelet coefficients at different scales. Kaplan and Kuo [7] have shown that for Haar wavelet, the variance progression of the wavelet transform of fGn satisfies a power-law decay (c.f. (3)). In colored denoising scheme, the threshold at each scale is made proportional to the expected standard deviation of the wavelet coefficients at that scale. Thus, knowing the Hurst parameter, we can fix the threshold at each scale using equation (3). Johnstone *et al* [8] have shown that this thresholding scheme is optimal for colored denoising.

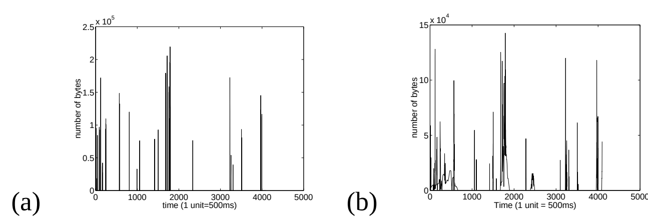

Figure 2: *(a) Isolated bursts in the real trace. (b) Bursts isolated using the wavelet separation scheme.*

3. ADDITIVE AND MULTIPLICATIVE MODELS

The models of interest here are based on a multi-resolution description of the traffic signal. At the heart of these schemes lies the fact that the wavelet transform is an approximate Karhunen-Loève transform for LRD signals. We design the wavelet coefficients to be uncorrelated in two ways, employing additive, resp. multiplicative innovations.

3.1. WIG Additive Model

The discrete wavelet transform is a multiscale signal representation of the form [9]

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

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

It has been shown [10] that highly-correlated LRD signals become nearly uncorrelated in the wavelet domain. In addition, the *Haar* wavelet transform of fGn exhibits power-law scaling of the form[3] [7]

$$\mathrm{var}(W_{j,k}) = \sigma^2 \, 2^{(2H-1)(j-1)} \, (2 - 2^{2H-1}). \tag{3}$$

Thus by generating independent wavelet coefficients $W_{j,k}$ with appropriate decay of energy with scale and inverting the wavelet transform, one can synthesize Gaussian LRD processes. Using efficient multiscale tree structures, this model provides fast $O(N)$ synthesis algorithms to synthesize N-point data sets [5, 10].

3.2. MWM Multiplicative Model

As a consequence of the Gaussian nature, the additive WIG model can produce unrealistic synthetic traffic traces, in particular since it can take negative values and cannot capture the burstiness of traffic at small time-scales. The basic idea behind the non-Gaussian MWM model 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.

To guarantee non-negativity of signals, notice that in Haar wavelet transform the scaling and wavelet coefficients can be recursively computed using

$$\begin{aligned} u_{j,2k} &= 2^{-1/2}(u_{j+1,k} + w_{j+1,k}) \\ u_{j,2k+1} &= 2^{-1/2}(u_{j+1,k} - w_{j+1,k}). \end{aligned} \tag{4}$$

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

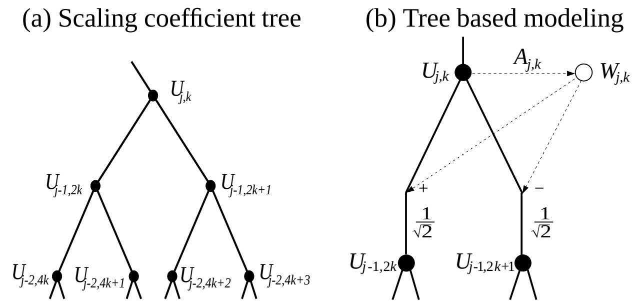

Figure 3: *(a) Binary tree of Haar scaling coefficients. (b) Tree-based modeling: At scale j, we produce innovations in form the wavelet coefficients $W_{j,k}$. For the* additive *WIG model these coefficients are chosen as independent Gaussian random variables. For the* multiplicative *MWM they are formed as the product $W_{j,k} = A_{j,k}U_{j,k}$ with independent positive multipliers $A_{j,k}$.*

For non-negative signals, $u_{j,k} \geq 0, \ \forall \ j,k$, which with (4) implies that

$$|w_{j,k}| \leq u_{j,k}, \qquad \forall \ j,k. \tag{5}$$

The positivity constraints (5) on the Haar wavelet coefficients lead us to a very simple multiscale, multiplicative signal model for positive processes. Let $A_{j,k}$ be a random variable supported on the interval $[-1,1]$ and define the wavelet coefficients recursively by

$$W_{j,k} = A_{j,k} \, U_{j,k}. \tag{6}$$

Together with (4) we obtain (see Figure 3(b))

$$\begin{aligned} U_{j,2k} &= 2^{-1/2}(1 + A_{j+1,k})\, U_{j+1,k} \\ U_{j,2k+1} &= 2^{-1/2}(1 - A_{j+1,k})\, U_{j+1,k}. \end{aligned} \tag{7}$$

We use beta distributions for $A_{j,k}$.

4. ADDITIVE AND MULTIPLICATIVE MIXTURE MODEL

With strikingly different statistical properties present in network traffic it is desirable to develop models which combine flexibility with parsimony. Using one single model capable to adapt to Gaussian as well as spiky, non-Gaussian signals is not only satisfactory from a signal processing point of view, but may also provide deeper inside into the causes of the complex traffic dynamics.

From the statistical description given in Section 1 it is apparent that WIG should provide an ideal model for the beta component, while the alpha component will find a better match with the MWM. This suggests to model the overall traffic as a superposition of WIG and MWM. However, such an approach implies an inflation of parameters which are hard to estimate as they rely on the traffic decomposition.

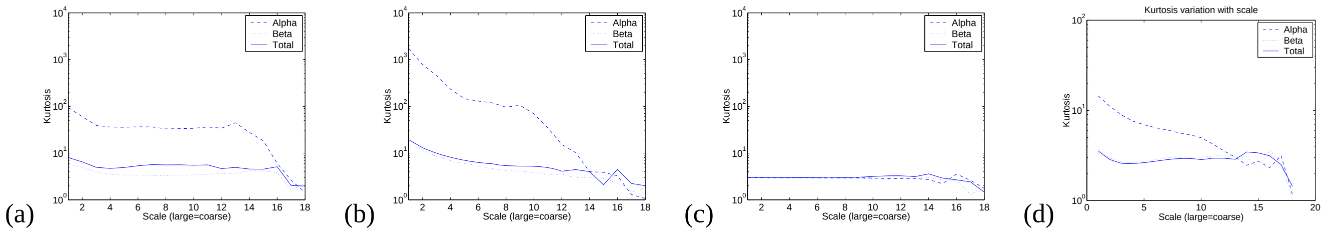

Figure 4: Kurtosis versus time-scale for (a) real traces; (b) MWM model; (c) WIG model; (d) WIG-MWM mixture model.

Rather, we propose to use both, additive as well as multiplicative innovations in one single tree. A first simple scheme works as follows: At each parent node on the tree we first try additive innovations with parameter such as to match the variance at the corresponding scale. The resulting children nodes at which the signal turns negative are replaced by recomputing the value using multiplicative innovations instead. Note that this scheme inherently preserves positivity. More importantly, the proportion of multiplicative innovations increases as the variances of the wavelet coefficients increase compared to its mean. This is desirable because non-Gaussian, spiky signals exhibit a large wavelet variance to mean ratio, leading to a largely multiplicative mixture tree. On the other hand, a nearly Gaussian, positive signal will produce a mixture tree with an overwhelming part of additive innovations.

To measure the level of Gaussianity of signal and models at different time-scales we use the well-known Gaussianity measure, kurtosis, which is defined as the ratio of its fourth central moment to the square of its variance. The kurtosis for Gaussian random variable is 3. Random variables with fatter tails have a kurtosis greater than 3, and vice versa.

Figure 4 displays the kurtosis values of signals and models against scale (coarse to the right) in log-log. It should be recalled foremost that all models use as parameters only the variance on binary scales which they match perfectly (not shown, see [2]). Notably, the WIG model is a poor approximation due to a deviation from Gaussianity even for the beta component on fine scales. The multiplicative MWM model, though more accurate, exhibits too large kurtosis on fine scales. The mixture model, while still not ideal, convinces through its uniformly superior match for both, Gaussian as well as bursty traces, and a flexibility which allows to match large and small kurtosis values.

Discussion

The main advantage of the novel mixture tree model lies in its parsimony and flexibility in matching Gaussian as well as non-Gaussian multiscale marginals. More elaborate models come to mind immediately, such as using distributions for the multiscale multipliers $A_{j,k}$ with more than one parameters in order to match the fourth moment in addition to the variance. A superposition of WIG and MWM might be superior, but significantly increases the difficulty of parameter estimation.

The good match of mixture tree models is satisfactory since they agree with the networking intuition that large scale behavior is governed by aggregation (an additive operation) while small scale behavior is strongly affected by multiplexing and queueing with their inherent non-Gaussianity.

5. REFERENCES

[1] 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*, vol. 5, no. 1, pp. 71–86, Feb. 1997.

[2] R. H. Riedi, M. S. Crouse, V. Ribiero, and R. G. Baraniuk, “A multifractal wavelet model with application to TCP network traffic,” *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 992–1018, April 1999.

[3] A. Erramilli, O. Narayan, and W. Willinger, “Experimental queueing analysis with long-range dependent traffic,” *IEEE/ACM Trans. Networking*, pp. 209–223, April 1996.

[4] 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.

[5] Sheng Ma and Chuanyi Ji, “Modeling video traffic in the wavelet domain,” in *Proc. of 17th Annual IEEE Conf. on Comp. Comm., INFOCOM*, Mar. 1998, pp. 201–208.

[6] S. Sarvotham, R. Riedi, and R. Baraniuk, “Connection-level analysis and modeling of network traffic,” in *Proc. IEEE/ACM Network Measurement Workshop*, 2001.

[7] L. M. Kaplan and C.-C. J. Kuo, “Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis,” *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3554–3562, Dec. 1993.

[8] I. M. Johnstone and B. W. Silverman, “Wavelet threshold estimators for data with correlated noise,” *J. Royal Stat. Soc. B*, , no. 59, pp. 319–351, 1997.

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

[10] L. M. Kaplan and C.-C. J. Kuo, “Extending self-similarity for fractional Brownian motion,” *IEEE Trans. Signal Proc.*, vol. 42, no. 12, pp. 3526–3530, Dec. 1994.