

# Modeling Heterogeneous Network Traffic in Wavelet Domain

Sheng Ma, *Member, IEEE*, and Chuanyi Ji

**Abstract**—Heterogeneous network traffic possesses diverse statistical properties which include complex temporal correlation and non-Gaussian distributions. A challenge to modeling heterogeneous traffic is to develop a traffic model which can accurately characterize these statistical properties, which is computationally efficient, and which is feasible for analysis. This work develops wavelet traffic models for tackling these issues. In specific, we model the wavelet coefficients rather than the original traffic. Our approach is motivated by a discovery that although heterogeneous network traffic has the complicated short- and long-range temporal dependence, the corresponding wavelet coefficients are all “short-range” dependent. Therefore, a simple wavelet model may be able to accurately characterize complex network traffic. We first investigate what short-range dependence is important among wavelet coefficients. We then develop the simplest wavelet model, i.e., the independent wavelet model for Gaussian traffic. We define and evaluate the (average) autocorrelation function and the buffer loss probability of the independent wavelet model for Fractional Gaussian Noise (FGN) traffic. This assesses the performance of the independent wavelet model, and the use of which for analysis. We also develop (low-order) Markov wavelet models to capture additional dependence among wavelet coefficients. We show that an independent wavelet model is sufficiently accurate, and a Markov wavelet model only improves the performance marginally. We further extend the wavelet models to non-Gaussian traffic through developing a novel time-scale shaping algorithm. The algorithm is tested using real network traffic and shown to outperform FARIMA in both efficiency and accuracy. Specifically, the wavelet models are parsimonious, and have the computation complexity  $O(N)$  in developing a model from a training sequence of length  $N$ , and  $O(M)$  in generating a synthetic traffic trace of length  $M$ .

**Index Terms**—Long-range dependence, network traffic modeling, self-similar traffic, wavelets.

## I. INTRODUCTION

**T**RAFFIC modeling and understanding is imperative to network design and simulation, to providing quality of service (QoS) to diverse applications, and to network management and control. Numerous models have been proposed in the past for modeling network traffic. However, it remains open how to model heterogeneous network traffic possessing two pertinent statistical properties: complex temporal correlation and marginal distributions that result from the complexity of (IP) net-

works and diverse network applications. The goal of this work is to develop a traffic model that is both accurate in capturing the aforementioned statistical properties and computationally efficient for developing a model as well as generating synthetic traffic.

The complex temporal correlation of network traffic can be characterized by the short-range (SRD) and the long-range dependence (LRD). Examples of traffic exhibiting SRD include voice-over IP (VoIP) [41] and VBR video traces [21]; and examples of traffic possessing LRD include web request traffic [12] and Ethernet data traffic [32]. The autocorrelation function of SRD traffic decays exponentially, and that of LRD traffic decays hyperbolically. For real-time applications, it has been shown that only SRD is relevant [25], [55], [7], [21], and [32]. Numerous models corresponding to short-range-dependent processes can be used to model SRD reasonably well. These models include variants of Markov processes [3], [20], [49], [57], [61], [64] and DAR [16], [25]. For nonreal time applications such as video-on-demand, most of the data communications and some network management tasks, a traffic model needs to capture the temporal dependence at large time scales, i.e., the long-range dependence. The Markov-type models, when extended to capture LRD, often result in a complicated structure with many states/parameters [3]. Models such as Fractional Gaussian Noise (FGN) processes [32] can capture the long-range dependence but not the short-range dependence. In fact, network traffic such as VBR video can exhibit a complex mixture of SRD and LRD. That is, the corresponding autocorrelation function behaves similarly to that of long-range dependent processes at large lags, and to that of short-range dependent processes at small lags [7], [21]. Models developed to characterize both SRD and LRD include FARIMA [21], a model based on the Hosking procedure [26], the scene-based model [27], the Markov Modulated Process [3], [52], the fractal point process [54], and the  $M/G/\infty$  model [31]. Among these methods, the scene-based model [27] and the Markov Modulated process [3], [52] provide a physically interpretable model to include both long-range and short-range dependence. However, due to the stochastic nature of network traffic, it is difficult to accurately define and segment network traffic into different states in the time domain. FARIMA models are not computationally efficient. They may require a large number of model parameters, and  $O(N^2)$  computational time to develop a model from a traffic trace of length  $N$  and to generate synthetic traffic of length  $N$  [21], [26]. The  $M/G/\infty$  model has been shown to have a moderate number of parameters. However, it is a point process. Its efficiency in generating a high volume of synthetic traffic needs further investigation.

Manuscript received March 3, 1999; revised March 13, 2000; recommended by IEEE/ACM TRANSACTIONS ON NETWORKING Editor T. V. Lakshman. This work was supported by the National Science Foundation under contracts NCR 9805338 and CAREER IRI-9502518.

S. Ma is with Machine Learning for Systems, IBM T. J. Watson Research Center, Hawthorne, NY 10532 USA (e-mail: shengma@us.ibm.com).

C. Ji is with the Department of Electrical and Computer Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: chuanyi@ecse.rpi.edu).

Publisher Item Identifier S 1063-6692(01)09000-8.

Non-Gaussian distribution is another important statistical property of heterogeneous traffic. It has been shown that both video and data traffic have heavy-tailed non-Gaussian marginal probability density functions (PDFs) [21], [53], [30]. Moreover, the higher order statistics of traffic can have a significant impact on accurately predicting the buffer overflow probability [26], [23]. Algorithms have been proposed to incorporate higher order statistics by matching the marginal distribution and the second-order statistics of network traffic [26], [44], [24]. However, their performance, measured by the queueing results, is still not as desirable [26]. This is because these algorithms only model the marginal distribution of the traffic at the finest time scale, whereas the marginal distributions across different time scales should be modeled for accurately predicting queueing behavior.

The goal of this work is to develop a traffic model which can capture complex temporal dependence in terms of both LRD and SRD, which can model non-Gaussian distributions to achieve accurate queueing performance, and which is computationally efficient. Why are these aspects difficult to be achieved simultaneously? The main reason is that the heterogeneous traffic is intrinsically complex in the time domain. This makes it difficult for the time-domain models to be both accurate and efficient. Models have been developed in the frequency domain [34]. Although the harmonics are not a generic representation of heterogeneous traffic, the idea of modeling in a transformation domain [34] motivates our work in this paper.

Which transformation domain may be suitable for modeling network traffic? Sherman *et al.* [56] demonstrated that the statistical properties of the aggregated traffic are self-similar across a wide range of time scales. In addition, network control and management are often performed at different time scales. All these motivate us to model heterogeneous traffic based on time scales. We show later that time scales can be naturally represented by wavelets. We also demonstrate that the wavelet representation matches the properties of the bursty network traffic, in that the wavelet coefficients are short-range dependent even though the corresponding heterogeneous traffic may be long-range dependent. Consequently, a simple yet accurate model can be derived in the wavelet domain.

Wavelet models have been developed for fractional Brownian motion (FBM) and scale-invariant processes [4], [5], [35], and [18] in signal processing. Wornell *et al.* [63] have proven that the spectrum of the independent wavelet model of an FBM process is very close to that of  $1/f$  processes. Therefore, the independent wavelet model has been proposed to rapidly generate FBM of FGN-like synthetic sample paths. However, the previous investigation on (asymptotic) correlation structure of wavelet coefficients has been focused mostly on a limited scope for FBM [63], [19], [42], [59], FGN [28], or AR(1) [14]. The correlation structure has not been well studied for short-range dependent processes nor for a mixture of long- and short-range dependent processes. To apply wavelet models to networking related applications, [2], [1], [18] proposed to estimate Hurst parameters by wavelet models. The possibility of using wavelets for modeling network traffic was mentioned in [48] and [17]. However, wavelet approaches have not been applied to modeling heterogeneous traffic when this work is developed [37], [38]. Recently,

[51], [50] have applied multiplicative wavelet models to model network traffic.

The main contributions of this work include: 1) the application of the wavelet approach for modeling heterogeneous traffic; 2) the development of a novel time-scale shaping algorithm to incorporate non-Gaussian distributions of network traffic; and 3) investigation of the performance of the wavelet models using networking-related performance measures.

Herein, we first investigate whether/why the wavelet modeling approach is indeed capable of capturing the complex temporal dependence in heterogeneous traffic. For this, we derive analytical results on the correlation of wavelet coefficients. These results show that a key advantage of using wavelets is their ability to reduce the complex temporal dependence so significantly that the wavelet coefficients only possess the short-range dependence. With these results, we first develop wavelet models for both SRD and LRD Gaussian traffic, and then develop a novel multitime-scale shaping algorithm for the non-Gaussian distributed traffic. After we assess our models through intensive experiments, we further assess our wavelet models analytically for modeling FGN traffic. In particular, we show that the average buffer overflow probability of the independent wavelet model, when used to model FGN traffic, is asymptotically close to that of the original traffic. We show that the autocorrelation function of the independent wavelet model of FGN traffic has the same (hyperbolic) decay rate as the original traffic, with an approximation error of less than 15%. From the computational efficiency perspective, we show that wavelet models are parsimonious, and have the lowest computational complexity. In specific, the computational complexity is  $O(N)$  in developing a wavelet model from a training sequence of length  $N$ , and  $O(M)$  in generating a synthetic sample path of length  $M$ .

The paper is organized as follows. Section II provides background knowledge. Section III studies why wavelet models are good candidates for long- and short-range dependent traffic. Section IV investigates Gaussian wavelet models. Section V develops a time-scale shaping algorithm to extend the independent wavelet models to non-Gaussian traffic. Section VI presents experimental results which validate the performance of the wavelet models. Section VII provides analysis on the autocorrelation function and the buffer loss probability of the independent wavelet model of FGN traffic to further evaluate the performance of the wavelet models. Section VIII discusses our results. Section IX concludes the paper.

## II. BACKGROUND

### A. Long/Short-Range Dependence and Performance Measures

Intuitively, long-range dependence (LRD) can be considered as a phenomenon that current observations are significantly correlated to the observations that are far away in time. One formal definition [21] of a long-range dependent stationary process can be described as that the sum of its correlation function  $r(k)$  over all lags is infinite.<sup>1</sup> This implies that the correlation function  $r(k)$  decays asymptotically as a hyperbolic function of  $k$ , i.e.,

<sup>1</sup>Please see [32], [21], [11], and [6] for other definitions and properties of the LRD.

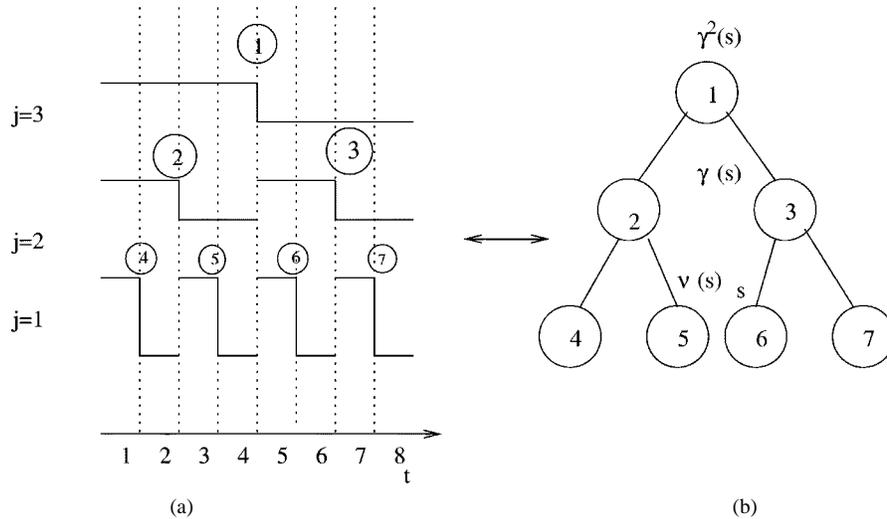


Fig. 1. (a) Haar wavelet basis functions. (b) Corresponding tree diagram and two types of operations. The number in the circle represents the one-dimensional index of the wavelet basis functions. For example, the equivalent notation of  $d_1^2$  is  $d_6$ .  $s$ ,  $\nu(s)$  and  $\gamma(s)$  represent the one-dimensional index of wavelet coefficients.  $\gamma(s)$  is defined to be the parent node of node  $s$ .  $\nu(s)$  is defined to be the left neighbor of node  $s$ .

$r(k) \sim O(k^{-(2-2H)})$  for  $k \geq 0$ .  $H(0.5 < H < 1)$  is the so-called Hurst parameter, which is an important quantity for characterizing the LRD. Examples of such long-range dependent processes include the FGN process and the fractional autoregressive integrated moving average process (FARIMA). The nature of these random processes is “self-similar,” i.e., the corresponding statistical properties are invariant at different time scales [6], [11]. In particular, FGN is a Gaussian process, and can be completely specified by three parameters: mean, variance, and the Hurst parameter. FARIMA  $(p, d, q)$  is a fractional differentiation of an auto-regressive moving average (ARMA  $(p, q)$ ) process, where  $p$  and  $q$  represent the orders of the ARMA  $(p, q)$  process and  $d(0 < d < 0.5)$  is a differentiation degree. The Hurst parameter  $H$  of FARIMA  $(p, d, q)$  equals to  $0.5 + d$ . FARIMA  $(p, d, q)$  has  $p + q + 3$  parameters, and is much more flexible than FGN in terms of simultaneously modeling of both long-range dependence and short-range dependence in network traffic [6]. Examples of short-range dependent random processes include auto-regressive (AR) and ARMA processes with exponentially decaying correlation functions, i.e.,  $r(k) \sim \rho^k (-1 < \rho < 1)$ .

The criteria that we use to measure the performance of the wavelet models are the autocorrelation function and the buffer loss rate. The autocorrelation function is an important quantity characterizing the second-order statistics of a wide-sense-stationary process. If a model is able to capture both LRD and SRD components in network traffic, it should be able to match the autocorrelation function of network traffic in a long enough range. The buffer loss rate is chosen as one other criterion, since an important goal for traffic modeling is to assist designing the buffer size of a server,<sup>2</sup> and estimating the packet loss rate.

### B. Wavelet Transformation

Wavelets are complete orthonormal bases which can be used to represent a signal as a function of time [13]. In  $L^2(R)$ , discrete wavelets can be represented as

$$\phi_j^m(t) = 2^{-j/2} \phi(2^{-j}t - m) \quad (1)$$

<sup>2</sup>This can be modeled as a single queue with capacity  $C$  and a buffer size  $B$ .

where  $j$  and  $m$  are positive integers. The dilation index  $j$  characterizes the function  $\phi(t)$  at different time scales.  $m$  represents the translation in time. Because  $\phi_j^m(t)$  are obtained by dilating and translating a mother function  $\phi(t)$ , they have the same shape as the mother wavelet and therefore are self-similar to each other.

A discrete-time process  $x(t)$  can be represented through its inverse wavelet transform

$$x(t) = \sum_{j=1}^K \sum_{m=0}^{2^{K-j}-1} d_j^m \phi_j^m(t) + \phi_0 \quad (2)$$

where  $0 \leq t < 2^K$ .  $\phi_0$  is equal to the average value of  $x(t)$  over  $t \in [0, 2^K - 1]$ . Without loss of generality,  $\phi_0$  is assumed to be zero for the rest of this paper.  $d_j^m$ 's are wavelet coefficients and can be obtained through the wavelet transform

$$d_j^m = \sum_{t=0}^{2^K-1} x(t) \phi_j^m(t). \quad (3)$$

The mother wavelet we choose in this work is the Haar wavelet, where

$$\phi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1/2, \\ -1 & \text{if } 1/2 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

To explore the relationships among wavelets, Willsky *et al.* defines a tree diagram and the corresponding one-dimensional indices of wavelet coefficients [5], [35], [10]. Fig. 1(a) shows an example of Haar wavelets for  $K = 3$ , and Fig. 1(b) shows the corresponding tree diagram. The circled numbers represent the one-dimensional indices of the wavelet basis functions, and are assigned sequentially to wavelet coefficients from the top to the bottom and the left to the right. The one-dimensional index  $s$  is thus a one-to-one mapping to the two-dimensional index  $(j(s), m(s))$ , where  $j(s)$  and  $m(s)$  represent the scale and the shift indices of the  $s$ th wavelet. The equivalent notation<sup>3</sup> of  $d_s$

<sup>3</sup>For example,  $d_6$  is  $d_1^2$  in the given example. (The shift index  $m$  starts from 0.)

is then  $d_j^{m(s)}$ . In addition, we denote the parent and the neighboring wavelets of a wavelet through the tree diagram. As shown in Fig. 1,  $\gamma(s)$  and  $\nu(s)$  are the parent and the left neighbor of node  $s$ , respectively. We use both the one-dimensional and two-dimensional indices of a wavelet coefficient in this paper.

A key advantage of using Haar wavelets is simplicity. The computational complexity of the (Haar) wavelet transform and inverse transform is  $O(N)$ , where  $N$  is the length of the time series.

When  $x(t)$  is a random process, which is of interest to this work, the corresponding wavelet coefficients  $d_j^m$ 's define a two-dimensional random processes in terms of  $j$  and  $m$  (see [22], [63], [5], and references therein for details). Due to the one-to-one correspondence between  $x(t)$  and its wavelet coefficients, the statistical properties of the wavelet coefficients are completely determined by those of  $x(t)$ . Likewise, if the statistical properties of the wavelet coefficients are well specified, they can be used to characterize the original random process. This motivates our approach of traffic modeling by characterizing statistical properties of wavelet coefficients.

### III. WHY WAVELETS: CORRELATION OF WAVELET COEFFICIENTS

One of the main motivations for using wavelets is their ability to reduce the temporal correlation so that wavelet coefficients are less correlated. In this section, we first provide (asymptotic) analysis on correlation structures of wavelet coefficients for well-known LRD and SRD processes. We then provide empirical studies to show that the correlation structures are dominated by only a few key elements. This motivates traffic modeling in the wavelet domain, and the simple wavelet models we will choose in Section IV.

#### A. Analysis on Correlation Structure of Wavelet Coefficients

1) *Correlation Structure of Wavelet Coefficients of LRD Processes:* The correlation structure of (long-range dependent) FGN process has been investigated extensively in [28], [63], [19], and can be applied to the problem we consider in this work.

*Theorem 1:* (Kaplan and Kuo [28]; Flandrin [19]): Let  $x(t)$  be a FGN process with Hurst parameter  $H(0.5 < H < 1)$ . Let  $d_j^m$ 's be the (Haar) wavelet coefficients of  $x(t)$ . Then:

- 1) For a given time scale  $j$ ,  $d_j^m$ 's are i.i.d. Gaussian random variables with zero mean and variance  $2^j(2^{2H-1})(2^{2(1-H)} - 1)\sigma^2$ , where  $\sigma$  is the variance of  $x(t)$ .
- 2) For  $(m_1 + 1)2^{j_1} - m_2 2^{j_2}$  large, where  $j_1, j_2, m_1$  and  $m_2$  are the dilation and the translation indices of two different wavelet coefficients, respectively, the correlation between two wavelet coefficients is

$$E(d_{j_1}^{m_1} d_{j_2}^{m_2}) \sim O\left(|2^{j_1} m_1 - 2^{j_2} m_2|^{-2(1-H')}\right) \quad (5)$$

where  $H' = 1 - H$ .

Here  $|2^{j_1} m_1 - 2^{j_2} m_2|$  is the shortest distance between two wavelets, and greater than 1. The exponent,  $1 - H'$ , is between 0 and 0.5 since  $0.5 < H < 1$  for FGN processes. This shows that the correlation changes from the

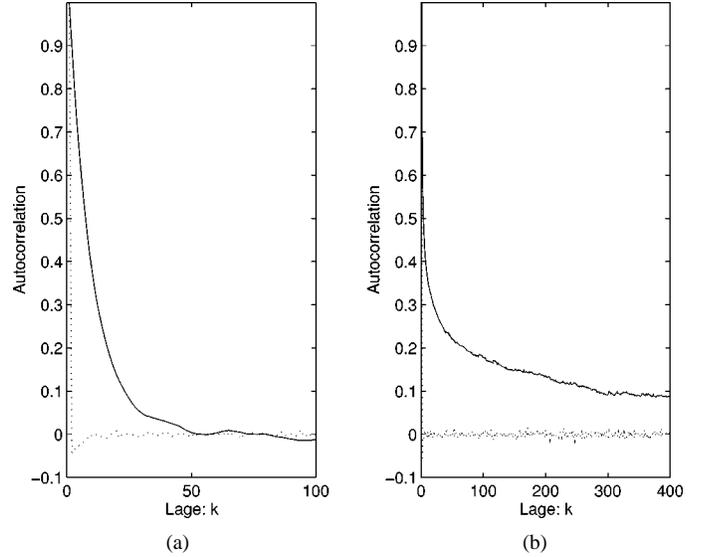


Fig. 2. Solid line: Autocorrelation coefficients of the original process. Dotted line: the normalized autocorrelation of wavelet coefficient, i.e.,  $(E(d_1^m d_1^{m+k})/\sigma_{d_1^m} \sigma_{d_1^{m+k}})$ . (a) AR(1) process. (b) FARIMA (0, 0.4, 0) process.

mean-revert ( $0.5 < H < 1$ ) in the time domain to the mean-avert ( $0 < H' < 0.5$ ) in the wavelet domain. Recall that the temporal autocorrelation of FGN decays at a rate  $O(|k|^{-2(1-H)})$  for  $k$  being the lag, and is thus non-summable. The above theorem indicates that the wavelet transformation has changed the long-range dependence in the time domain so significantly that the summation of the correlation of wavelet coefficients converges to a constant. Fig. 2 illustrates how drastic the reduction is by comparing the autocorrelation function of the original FARIMA (0, 0.4, 0) process to the corresponding autocorrelation function<sup>4</sup> of wavelet coefficients (of  $d_1^m$  and  $d_1^{m+k}$ ).

2) *Correlation Structure of Wavelet Coefficients of SRD Processes:* For short-range dependent processes, we derive the correlation of wavelet coefficients.<sup>5</sup>

*Theorem 2:* Let  $x(t)$  be a zero mean wide-sense-stationary (discrete) Gaussian process with the autocorrelation  $r(k)$ , where  $r(k) = \sigma^2 \rho^{|k|}$  with  $|\rho| < 1$ ,  $k$  is an integer and  $\sigma^2$  is the variance of  $x(t)$ . Let  $d_j^m$ 's be the (Haar) wavelet coefficients of  $x(t)$ . Then:

- 1) For a given time scale  $j$ ,  $d_j^m$ 's are Gaussian random variables with a zero mean and a variance  $\sigma^2(1 + (2\rho/1 - \rho) - (3\rho/(1 - \rho)^2 2^{j-1})) + O(\rho^{2^{j-1}})$ .
- 2) For  $m_1 2^{j_1} - (m_2 + 1)2^{j_2} > 0$

$$\begin{aligned} E(d_{j_1}^{m_1} d_{j_2}^{m_2}) &= 2^{(-j_1 - j_2/2)} \rho^{m_1 2^{j_1} - (m_2 + 1)2^{j_2}} \\ &\quad \times \left(1 - \rho^{2^{j_2 - 1}}\right)^2 \left(1 - \rho^{2^{j_1 - 1}}\right)^2 \frac{\rho}{(1 - \rho)^2} \sigma^2. \quad (6) \end{aligned}$$

<sup>4</sup>It can be easily shown that the time series  $d_j^m$  for a fixed  $j$  is stationary in terms of  $m$ . Therefore, the autocorrelation exists.

<sup>5</sup>No previous results exist on the explicit correlation structure of wavelet coefficients for discrete processes except the bounds for some of the continuous random processes [14].

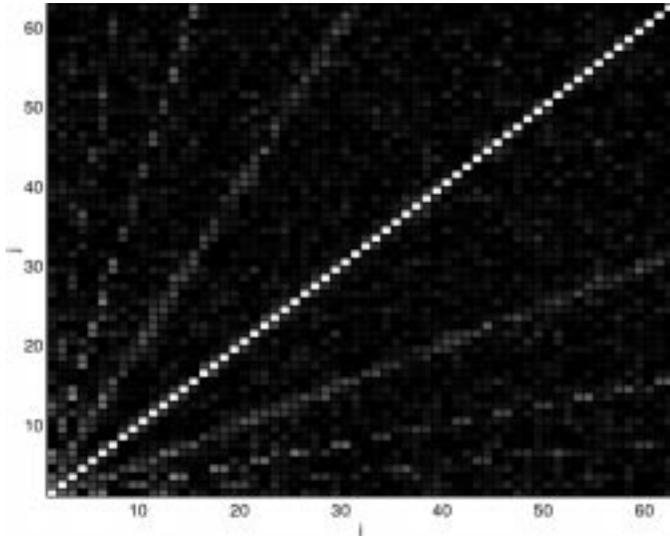


Fig. 3. Correlation matrix of FARIMA (0, 0.4, 0).

The sketch of the proof of the theorem is in Appendix A, and the details are in [38], [36].

This theorem shows that the correlation of wavelet coefficients decays exponentially as  $|m_1 2^{j_1} - m_2 2^{j_2}|$ , and therefore remains short-range dependent in the wavelet domain. In fact, the decay rate is even faster than the corresponding correlation in the time domain due to the “differencing operation” performed by the Haar wavelet transform. Fig. 2 illustrates the decay rate by comparing the (temporal) correlation for an AR(1) process with that of its wavelet coefficient for  $j_1 = j_2 = 1$ .

The above investigations suggest that a complex short- and/or long-range dependent process in the time domain may be sufficiently modeled by a short-range dependent process in the wavelet domain. That is, simple models which are insufficient for the temporal process may be accurate when used to model the wavelet coefficients.

#### B. Empirical Studies on Correlation Structure of LRD and SRD

What short-range dependence needs to be captured among wavelet coefficients? Unfortunately, an answer to this question cannot be provided by Theorems 1 and 2, since they only hold for  $|(m_1 - 1)2^{j_1} - m_2 2^{j_2}|$  large. We thus address this issue through experiments. Using sample paths of FARIMA (0, 0.4, 0) and AR(1), we obtain the corresponding correlation matrices of wavelet coefficients plotted in Figs. 3 and 4, respectively. A pixel  $(i, k)$  in an image represents the correlation between the  $i$ th and the  $k$ th wavelet coefficients, where  $i$  and  $k$  are the one-dimensional indices shown in Fig. 1. The gray level is proportional to the magnitude of the correlation. The higher the magnitude of the correlation, the whiter the pixel in the image. These figures show that in addition to the diagonal line,<sup>6</sup> there are four pairs of lines having “visible” correlation.<sup>7</sup> They correspond to the correlation between  $\gamma^k(s)$  and  $s$ , where  $\gamma(s)$  represents the parent of the node  $s$ , and  $\gamma^k(s)$  denotes the parent of

<sup>6</sup>In order to have enough gray level to see more subtle details, the diagonal pixels, which is always 1, is set to 0.5.

<sup>7</sup>We only consider  $K = 5$  which has only five levels in the tree diagram.

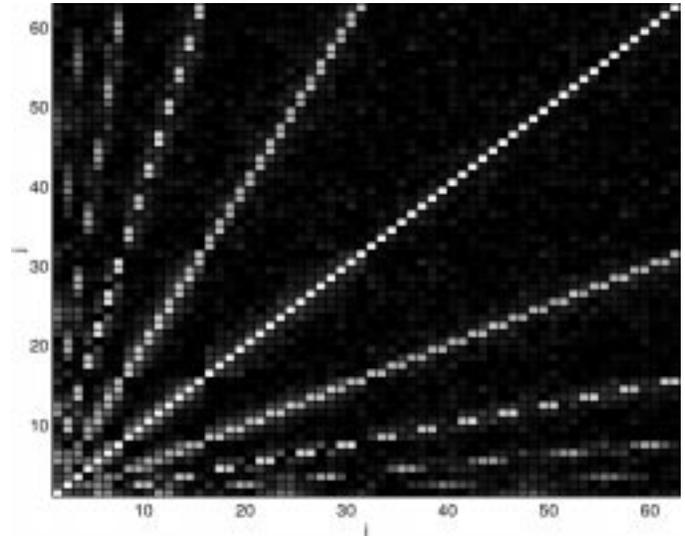


Fig. 4. Correlation matrix of AR(1).

the node  $\gamma^{k-1}(s)$  with  $k$  being 1, 2, 3, 4 from the diagonal line. We then conclude that the most significant correlation is due to the parent-child relationship. Since the complicated temporal correlation concentrates on only a few key correlation patterns in the wavelet domain, we can use a parsimonious model in the wavelet domain to represent the original traffic.

#### IV. GAUSSIAN WAVELET MODELS

We begin developing the wavelet models from the traffic with Gaussian distribution, for which we only need to characterize the autocovariance function through the wavelet models.

##### A. General Markov Models in Wavelet Domain

We capture the short-range dependence among wavelet coefficients using Markov models. Such Markov models can be implemented through a linear model on wavelet coefficients [5], where<sup>8</sup>

$$d_s = \sum_{l=1}^{s-1} a_s(l) d_l + b_s w_s. \quad (7)$$

Here,  $a_s(l)$  ( $1 \leq l \leq N$ ) and  $b_s$  are weighting factors depending on the one-dimensional index  $s$ , and  $w_s$  is i.i.d Gaussian noise with zero mean and a unit variance.

The order of the Markov model,  $s$ , can be chosen to make a tradeoff between model complexity and performance. Since our empirical study has demonstrated the regular patterns of the correlation structure for a wide range of SRD and LRD processes, we can choose  $s$  accordingly to capture several strongest correlations and ignore the insignificant ones.

##### B. Independent Wavelet Model for Gaussian Traffic

The simplest model is the independent wavelet model when  $s$  is chosen to be 1, and  $a_s(l) = 0$  for all  $l$ . This corresponds to the case that  $d_j^m$ 's are independent Gaussian random variables with zero mean and variance  $\sigma_j$ .  $\sigma_j$  can be estimated from data at

<sup>8</sup>Here we assume causal relations among wavelet coefficients.

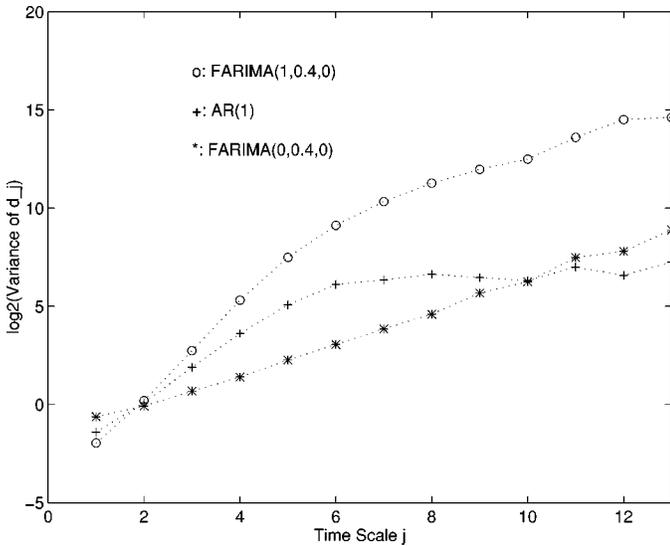


Fig. 5. Log 2 of variance of  $d_j$  versus the time scale  $j$ .

each  $j$  independently. This model only characterizes the mean and the variance of individual wavelet coefficients, and completely neglects the interdependence among them. Therefore, the resulting correlation is

$$E(d_{j_1}^{m_1} d_{j_2}^{m_2}) = \begin{cases} \sigma_j^2, & m_1 = m_2 \text{ and } j_1 = j_2 = j \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

How can the independent wavelet model represent LRD/SRD processes in the time domain? To provide answers to this question, Fig. 5 plots the variances of wavelet coefficients from an independent wavelet model for several well-known processes: a long-range dependent FARIMA (0, 0.4, 0), a short range dependent process AR(1), and a mixture of long-range and short-range dependent process FARIMA (1, 0.4, 0).

The figure shows that variances of the three processes exhibit different behavior. In particular, the variance of LRD increases with  $j$  exponentially for all  $j$ . Intuitively, this is due to the fact that the statistical variation of a LRD process persists through all time scales. The variance of SRD first increases rapidly at small time scale  $j$ , but saturates when  $j$  is large. This is because that the statistical variation of SRD processes only persists at small time scales.<sup>9</sup> For a mixture of LRD and SRD, the variance shows the mixed behaviors from both SRD and LRD. These plots suggest that the variances of independent wavelet coefficients are capable of distinguishing LRD from SRD for Gaussian processes.

### C. (Low-Order) Markov Wavelet Models

The next simplest model is the first-order Markov model, which captures the parent-child relationship, the most significant correlation among wavelet coefficients as shown in the previous section. Specifically

$$d_s = a'(j(s))d_{\gamma(s)} + b'(j(s))w_s \quad (9)$$

<sup>9</sup>As an extreme case, the variance of an i.i.d. temporal process, which includes Poisson-type processes and i.i.d. Gaussian processes, the variances of the corresponding wavelet coefficients do not vary with respect to the time scale at all [47], [62].

where  $a'(j(s))$  and  $b'(j(s))$  are the parameters to be determined from a training sequence,  $s$  is the one-dimensional index of a wavelet coefficient, and  $j(s)$  is the scale index of  $s$ .  $w_s$  is Gaussian noise with zero mean and the unit variance. This is a special case of (7) for all  $a_s(l)$ s to be zero except  $a_s(\gamma(s))$ , where  $\gamma(s)$  is the parent node of  $s$ . The resulting correlation pattern of this wavelet model consists of the diagonal line and the next brightest off-diagonal line in Figs. 3 and 4. Markov wavelet models with even higher orders can be used to capture more correlations among wavelet coefficients, and can be implemented in a similar manner.

### D. Algorithm for Developing Wavelet Models and Generating Synthetic Traffic

Once the form of a wavelet model is chosen, two issues need to be considered: 1) how to obtain parameters of a wavelet model from a training sequence; and 2) how to generate synthetic traffic from the obtained model. The algorithm given below implements these two tasks. Assume a training sequence  $\hat{x}(t)$  of length  $N$  is given from a Gaussian process.

*Algorithm:*

- 1) Estimate parameters from  $\hat{x}(t)$ .
  - Perform the wavelet transform on  $\hat{x}(t)$  to obtain the corresponding training sequence of wavelet coefficients,  $\hat{d}_j^m$ s.
  - Estimate the required parameters in the selected wavelet correlation model (Section IV) from  $\hat{d}_j^m$ s.<sup>10</sup>
- 2) Generate synthetic traffic.
  - Generate coefficients  $d_j^m$  from the wavelet correlation model using the estimated parameters for all  $m$  and  $j$ .
  - Perform the inverse wavelet transform to the generated wavelet coefficients ( $d_j^m$ s). This results in the synthetic traffic in the time domain.

Efficiency of traffic models can be measured through two quantities: 1) the computational time needed to develop a model using a training sequence and to generate synthetic traffic; and 2) the number of parameters of a model. In particular, the wavelet transform, the inverse transform, and the parameter estimation for the wavelet models are all linear [13]. Thus, the computational time is  $O(N)$  for developing a wavelet models, and  $O(M)$  for generating synthetic traffic of length  $M$ . As a comparison, FARIMA requires  $O(N^2)$  for estimating parameters<sup>11</sup> using a training sequence of length  $N$ , and  $O(M^2)$  for generating a synthetic trace. In terms of the actual computing time, it usually takes at most a few minutes on a Sparc station to develop a wavelet model and to generate a trace of length  $2^{18}$ , whereas a FARIMA model needs at least several hours to complete the same task.

As for the number of parameters of a wavelet model, an independent wavelet model has at most  $\log N$  parameters for modeling a training sequence of length  $N$ . An  $n$ th order Markov wavelet model has about  $(n + 1) \log N$  parameters (in this work,  $n = 1, 2, 3$ ).

<sup>10</sup>For example, for the independent wavelet model, the sample variances of wavelet coefficients are estimated.

<sup>11</sup>Through maximum-likelihood estimation.

## V. NON-GAUSSIAN INDEPENDENT WAVELET MODEL

As heterogeneous network traffic often possesses a non-Gaussian distribution, we extend the Gaussian wavelet models to non-Gaussian traffic. The key idea is to shape the distributions of synthetic traffic at multiple time scales. To motivate this approach, we first discuss the relationships among queueing behaviors, time scales, and wavelets. We then present our algorithm.

### A. Time Scale and Buffer Overflow Probability

What should be modeled to accurately predict buffer overflow probability? Queueing analysis shows that the marginal distribution of the cumulative process of traffic at the critical time scale is crucial for determining the buffer overflow probability [15], [8], [9].

To briefly review these results in the large deviation theory, we let  $C$  be the capacity of a single first-in-first-out (FIFO) buffer with an infinite waiting room,  $Q_t$  be the buffer size at time  $t$ , and  $B$  be the threshold for the buffer overflow.  $Q_t$  is known to satisfy  $Q_t = \sup_{s \geq 1} (X_t(s) - cs)$  [29], where  $X_t(s)$  is the cumulated process of a work load  $x(t)$

$$X_t(s) = \sum_{j=0}^{s-1} x(t+j) \quad (10)$$

and  $s$  represents the period for cumulation.

The buffer overflow probability has shown to satisfy [15], [45], [8], [9]

$$\Pr(Q_t > B) \approx \Pr(X_t(s^*) > cs + B) \quad (11)$$

where  $s^*$  is the so-called critical time scale, and  $s^* = \arg \sup_{s \geq 1} \Pr(X_t(s) > cs + B)$ . This approximation has been shown to be valid asymptotically (for large  $B$ ) for a wide range of traffic including the long-range dependent FGN process [15]. It has also been shown to be a reasonable approximation even for a moderate buffer size with various traffic loads [8], [9].

An important implication of the above approximation is that the tail distribution of the cumulated process at the critical time scale determines the buffer overflow probability. Therefore, an accurate traffic model should capture the tail distribution of the cumulated process of the original traffic at the critical time scale. However, the critical time scale depends on the buffer size  $B$ , the capacity  $C$ , and the utilization. Therefore, in order to perform well under a wide range of conditions, a traffic model should match marginal distributions of the cumulated process at a wide range of time scales.

### B. Time Scales and Wavelets

The time scale has a natural relationship with wavelets. Specifically, as given in [39], the (Haar) wavelet coefficients can be related to the so-called scale coefficients by

$$v_j^{2^m} = \frac{1}{\sqrt{2}} (d_{j+1}^m + v_{j+1}^m) \quad (12)$$

where the scale coefficient  $v_j^m$  is defined as

$$v_j^m = 2^{-j/2} \sum_{t=m2^j}^{(m+1)2^j-1} x(t) \quad (13)$$

for  $j \geq 1$  and integer of  $m$ . By comparing this equation with the definition of cumulative process  $X_t(s)$  [see (10)], we can relate the scale coefficient  $v_j^m$  to the cumulative process  $X_t(s)$  as

$$v_j^m = 2^{-j/2} X_{m2^j}(2^j). \quad (14)$$

In other words, a scale coefficient  $v_j^m$  is simply the weighted cumulative process over the interval  $[m2^j, (m+1)2^j - 1]$  with a length  $s = 2^j$  and a starting point  $t = m2^j$ .

The scale coefficient can be further related to wavelet coefficients through the recursive relation [see (12)] as

$$v_j^m = \sum_{k=j+1}^K w_k d_k^{m2^k} + w_K v_K^0 \quad (15)$$

where  $w_k = 2^{-(k-j/2)}$  is a weighting factor, and  $v_K^0$  is the scale coefficient at the coarsest time scale. Finally, combining (15) and (14), wavelet coefficients can be related to a cumulative process through scale coefficients.

### C. Time-Scale Shaping Algorithm

Using the relationships among wavelet coefficients, scale coefficients, and the cumulative process, we can now derive our shaping algorithm for non-Gaussian traffic. The key idea consists of the following: 1) generating the so-called background wavelet coefficients by Gaussian wavelet models (Section IV-D); 2) computing the empirical distributions of scale coefficients of a training sequence; and 3) shaping the background wavelet coefficients so as to match the empirical distributions of the scale coefficients.

The idea can be implemented through a top-down procedure, i.e., the background wavelet coefficients are shaped from the coarsest to the finest time scales. Specifically, let  $\tilde{d}_k^m$  and  $\check{d}_k^m$  be the unshaped and the shaped wavelet coefficients at the  $k$ th time scale, respectively. Assume that the shaping has been done from the coarsest ( $K$ th) to the  $[(K-j+1)$ th] time scale. At step  $j$ , we fix the following:  $\tilde{d}_k^m$ 's for all  $m$  and  $K \geq k > K-j+1$ , and  $\tilde{v}_{K-j}^m$  [related to  $\tilde{d}_k^m$  through (15)]. Our objective at step  $j$  is to transform the (unshaped) wavelet coefficient  $\tilde{d}_{K-j}^m$  to the shaped wavelet coefficient  $\check{d}_{K-j}^m$  so as to match the empirical distribution of the scale coefficients  $v_{K-j-1}^m$ 's. To do so, we define an intermediate variable  $\mathbf{z}_{K-j}^m = (1/\sqrt{2})(\mathbf{d}_{K-j}^m + \tilde{v}_{K-j}^m)$ . We note that if an appropriate transform can be applied to  $\mathbf{d}_{K-j}^m$ , then  $\mathbf{z}_{K-j}^m$  would become the new scale coefficient,  $\tilde{v}_{K-j-1}^{2^m}$  [see (12)],<sup>12</sup> whose desired distribution can be computed by the training sequence at scale  $K-j-1$ . Hence, we obtain

$$\check{d}_{K-j}^m = \sqrt{2} F_{\check{v}_{K-j-1}^{2^m}}^{-1} \left( F_{\mathbf{z}_{K-j}^m}(\mathbf{z}_{K-j}^m) \right) - \tilde{v}_{K-j}^m \quad (16)$$

where  $F_{\check{v}_{K-j-1}^{2^m}}(\cdot)$  represents the distribution of the scale coefficient at scale  $K-j-1$ ,<sup>13</sup> and can be estimated through the histogram [57], [26] of training traffic using (14).

<sup>12</sup>By (12),  $\tilde{v}_{K-j-1}^{2^m}$  needs to be obtained to match the marginal of a cumulative process. But from (13),  $v_{K-j-1}^m = (1/\sqrt{2})(v_{K-j}^{2^m} + v_{K-j}^{2^{m+1}})$ . Therefore, either  $v_{K-j}^{2^m}$  or  $v_{K-j}^{2^{m+1}}$  can be chosen arbitrarily for shaping in order to match a desired marginal with that of a cumulative process. We choose to shape  $v_{K-j}^{2^m}$  in this work.

<sup>13</sup>We assume that  $v_{j-1}^m$  for a fixed  $j$  has the same distribution for different  $m$ . This is true for stationary traffic. Therefore, the distribution of  $\mathbf{v}_{j-1}^{2^m}$  can be estimated through a histogram of  $v_{j-1}^m$  for a fixed  $j$ .

To understand what transformation  $F_{z_{K-j}^m}(z_{K-j}^m)$  is, we note that for an independent wavelet model,  $\mathbf{d}_j^m$  is independent of  $\mathbf{v}_j^m$ . This is because  $\mathbf{v}_j^m$  only depends on wavelet coefficients at time scales larger than  $j$ . Therefore, we have

$$F_{z_{K-j}^m}(z_{K-j}^m) = F_{\mathbf{d}_{K-j}^m}(d_{K-j}^m) \quad (17)$$

where  $d_{K-j}^m$  is a Gaussian random variable determined by the wavelet model. For Markov wavelet models with dependent wavelet coefficients, the cumulative distribution function  $F_{z_{K-j}^m}(z_{K-j}^m)$  is equal to the conditional cumulative distribution function of  $\mathbf{d}_{K-j}^m$  given  $v_{K-j}^m$ ,

$$F_{z_{K-j}^m}(z^m) = F_{\mathbf{d}_{K-j}^m | v_{K-j}^m}(d_{K-j}^m). \quad (18)$$

Such a conditional distribution is difficult to estimate empirically. Therefore, the transformation for the independent wavelet model [see (17)] can be used as an approximation.

Combining (16) and (17), we have

$$\tilde{d}_{K-j}^m = \sqrt{2} F_{\tilde{v}_{K-j-1}^{2m}}^{-1} \left( F_{\mathbf{d}_{K-j}^m}(d_{K-j}^m) \right) - \tilde{v}_{K-j}^m. \quad (19)$$

From the shaped wavelet coefficient  $\tilde{d}_j^m$ , the shaped scale coefficient can be obtained as

$$\tilde{v}_{K-j-1}^{2m} = \frac{1}{\sqrt{2}} \left( \tilde{d}_{K-j}^m + \tilde{v}_{K-j}^m \right). \quad (20)$$

It can be verified that the (shaped) scale coefficient,  $\tilde{v}_{K-j-1}^{2m}$  has indeed the (targeted) distribution  $F_{\tilde{v}_{K-j-1}^{2m}}(\cdot)$ .

The above procedures can be summarized by the following algorithm.

#### Time-Scale Shaping Algorithm

Input: a training sequence (network traffic)  $\hat{x}(t)$ . Output: synthetic traffic  $\tilde{x}(t)$  or model parameters (wavelet coefficients,  $\tilde{d}_j^m$ s)

- Traffic modeling
  1. Do wavelet transform on the training sequence  $\hat{x}(t)$  to obtain wavelet coefficients  $\hat{d}_j^m$ s and then scale coefficients  $\hat{v}_j^m$ s [see (15)].
  2. Estimate the variance  $\hat{\sigma}_j^2$  of wavelet coefficients  $\hat{d}_j^m$  at each time scale  $j$ .
  3. Estimate the cumulative probability function of scale coefficients,  $F_{\hat{v}_j}(\cdot)$ , at each time scale  $j$  using a training sequence on the cumulative process<sup>14</sup>  $\hat{X}_{m2^j}(2^j)$ .
- Synthetic traffic (or the model parameters) generation
  1. Generate background (Gaussian) wavelet coefficients  $\tilde{d}_j^m$  from a wavelet correlation model.
  2. Recursively compute the (shaped) wavelet coefficients  $\tilde{d}_j^m$  [(16) and (19)] and scale coefficients  $\tilde{v}_{j-1}^m$  [see (20)] from  $j = K$  to 1 for all  $m$ .
  3. Do wavelet inverse transformation and obtain the synthetic traffic  $\tilde{x}(t)$ .

<sup>14</sup>Which can be obtained by aggregating the original training sequence at various time scales.

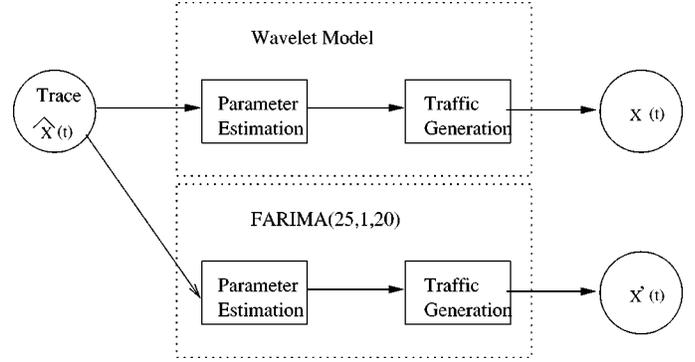


Fig. 6. Experimental setup.

## VI. PERFORMANCE OF WAVELET MODELS: NUMERICAL INVESTIGATIONS

In this section, we report numerical investigations on the performance of the wavelet models, where the performance for modeling Gaussian traffic is measured by the autocovariance function, and that for modeling non-Gaussian traffic is measured by the buffer loss probability. The results on modeling Gaussian traffic are verified using training sequences generated from known processes, and those for non-Gaussian traffic are validated using real network measurements.

### A. Simulation Setup

The experimental setup we use is shown in Fig. 6. A trace  $\hat{x}(t)$  is fed into a traffic model for estimating parameters of a wavelet model. The model obtained is then used to generate synthetic traffic  $x'(t)$ . The original and synthetic traffic traces are then used to obtain empirical autocovariance functions and the buffer losses probabilities, which are compared to measure the performance of the wavelet models. A FARIMA model is used similarly to further compare with the performance of the wavelet model. Both the sample autocorrelation functions and the buffer loss probabilities are obtained and compared with the true autocovariance functions and the buffer loss rates. The results for wavelet models are averaged over ten random sample paths.

### B. Performance on Gaussian Traffic

The performance of three types of wavelet models are investigated, which are independent wavelet models and the first-order and the third-order Markov wavelet models. A sample path  $\hat{x}(t)$  of length  $10^5$ – $10^6$  is generated from either an AR or a FARIMA model. As various model parameters are used in our simulations, we report results in this paper based on two representative cases: a short-range dependent process, AR(1) with an AR parameter 0.9, and a mixture of both SRD and LRD processes, FARIMA(1, 0.4, 0), with an AR parameter 0.9 and the Hurst parameter 0.9.

Figs. 7 and 9 show the sample autocorrelation functions for AR(1) and FARIMA(1, 0.4, 0); and Figs. 8 and 10 plot the buffer loss rates, respectively. To examine further the performance of each wavelet model, we plot the mean square error  $\text{MSE}(\tau)$  between the original autocorrelation and the one from a wavelet model summing up to a lag  $\tau$ , where

$$\text{MSE}(\tau) = \frac{1}{\tau} \sum_{k=1}^{\tau} (r(k) - \hat{r}(k))^2. \quad (21)$$

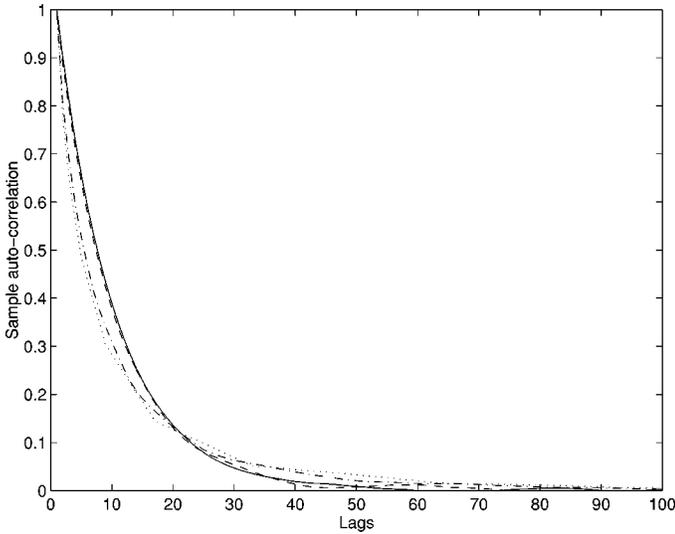


Fig. 7. Sample correlations. —: AR(1) (the true autocovariance function); - - : third-order Markov wavelet model; — · — : first-order Markov wavelet model; ... : independent wavelet model.

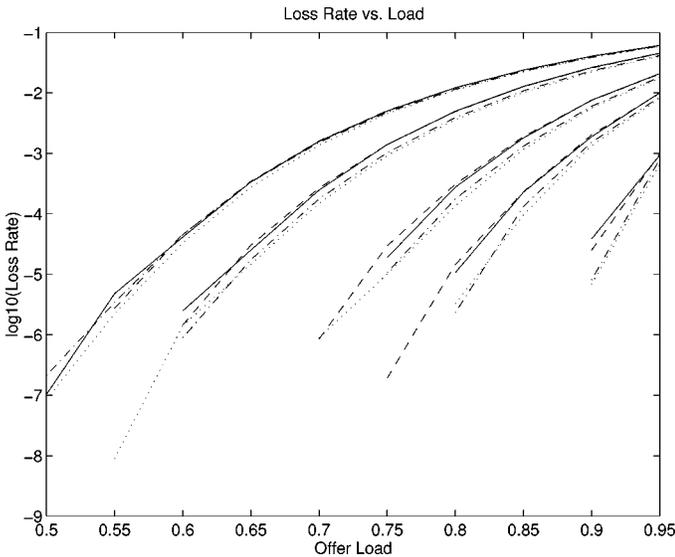


Fig. 8. Buffer response. x-axis: utilization. y-axis:  $\log_{10}$ (overflow probability). —: AR(1) (the true buffer loss probability); - - : third-order Markov wavelet model; — · — : first-order Markov wavelet model; ... : independent wavelet model. The normalized buffer size is 0.1, 0.5, 1, 10 from top down.

Figs. 11 and 12 plot  $MSE(\tau)$  for AR(1) and FARIMA(1, 0.4, 0), respectively.

As observed from the figures, the independent wavelet model, which neglects all the dependence in the wavelet domain, performs reasonably well. Markov wavelet models which capture more correlations among wavelet coefficients improve the performance only marginally.

### C. Performance on Real Network Traffic

The performance of the independent wavelet model and the time-scale algorithm is investigated in this section using real network traffic. Specifically, two widely used traffic traces are used. One is JPEG-coded *Star Wars* at the frame level [21]. The trace is obtained by applying a JPEG-like encoder to each of 171 000 frames at intervals of 1/24 second per frame of the

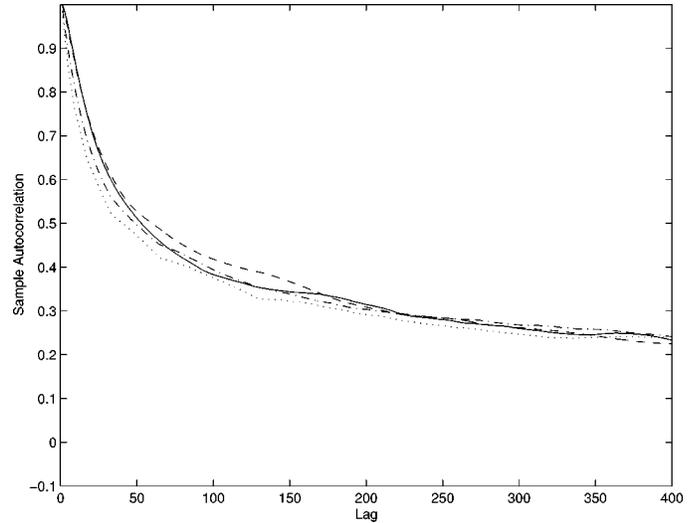


Fig. 9. Sample correlations. —: FARIMA(1, 0.4, 0) (the true autocovariance function); — · — : third-order; - - : first-order; ... : independent wavelet model.

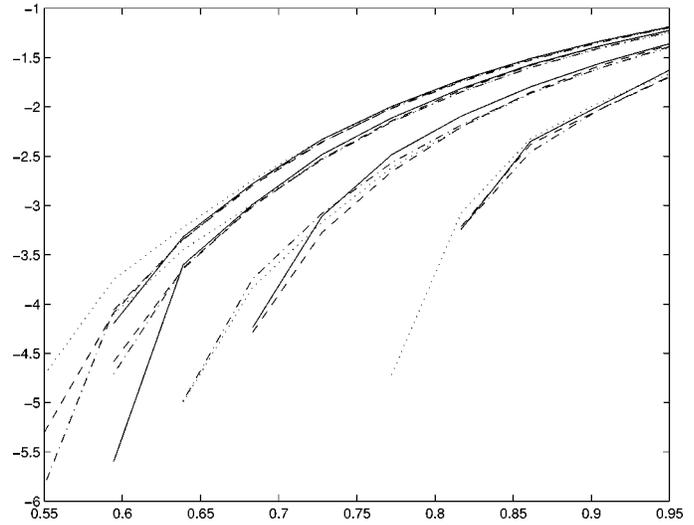


Fig. 10. Buffer response. x-axis: utilization. y-axis:  $\log_{10}$ (overflow probability). —: FARIMA(1, 0.4, 0) (the true buffer loss rate); — · — : third-order Markov wavelet model; - - : first-order Markov wavelet model; ... : independent wavelet model. The normalized buffer size is 0.1, 0.5, 1, 10 from top down.

two-hour movie *Star Wars*. The other is an Ethernet data trace used by Leland *et al.* [33].<sup>15</sup> The data set records the number of bits for every 10 ms during the half hour collection period. The length of this training sequence is 176 000.

An independent wavelet model is used and the time-scale shaping algorithm developed in Section V are applied. FARIMA(25, d, 20) model is used for comparison, where FARIMA(25, d, 20) has 25 AR parameters and 20 MA parameters.<sup>16</sup> The algorithms used for FARIMA to estimate its parameters and to generate synthetic traffic are from a commercial software package, Splus [58]. As the two training sequences have non-Gaussian marginal distributions, the generated Gaussian traffic by FARIMA is further transformed using a standard method [26] so that the resulting synthetic

<sup>15</sup>We only report results on the data set collected in August 1989.

<sup>16</sup>FARIMA(25, d, 20) is selected by compromising performance and complexity through multiple trials.

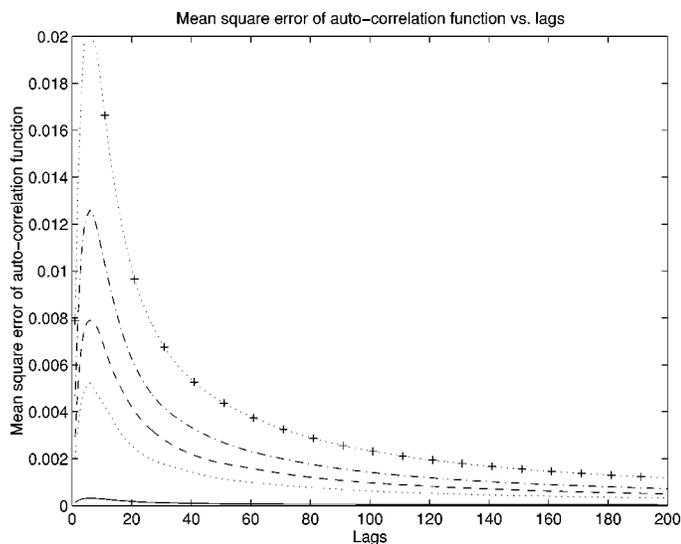


Fig. 11. MSE compared with AR(1). x-axis:  $\tau$ . y-axis:  $MSE(\tau)$ . — : fourth-order Markov wavelet model; ... : third-order; - - : second-order; - · - : first-order; · · · : independent wavelet model.

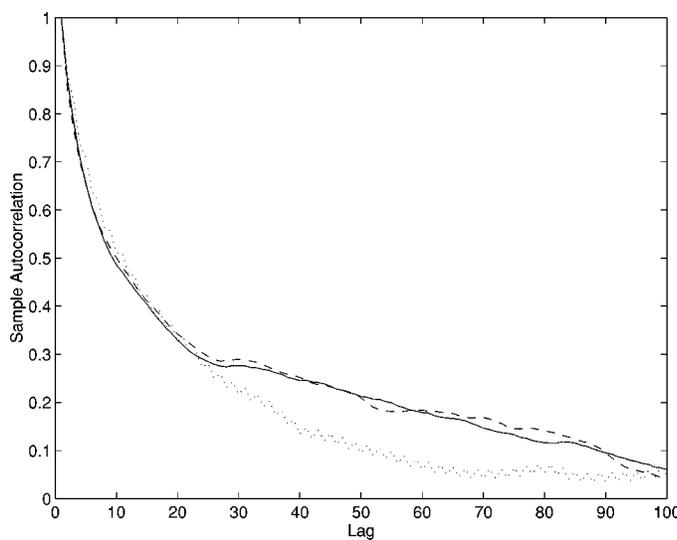


Fig. 13. — : Autocorrelation of *Star Wars*; - - : FARIMA(25, d, 20); ... : Algorithm 2.

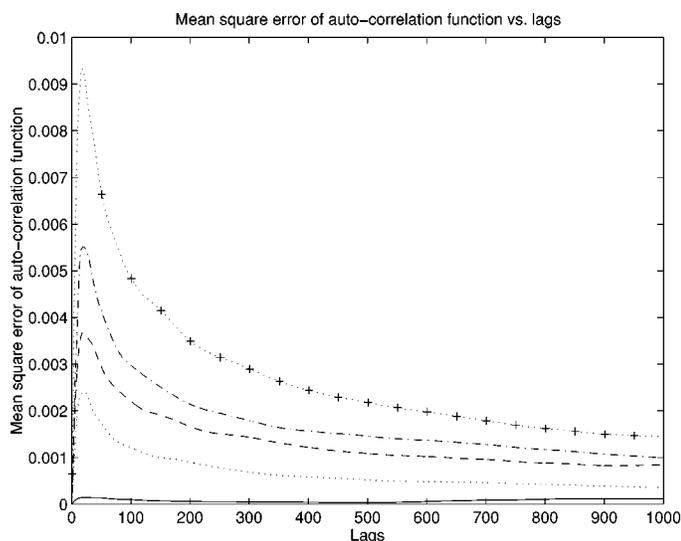


Fig. 12. MSE compared with FARIMA(1, 0.4, 0). x-axis:  $\tau$ . y-axis:  $MSE(\tau)$ . — : fourth-order Markov wavelet model; ... : third-order; - - : second-order; - · - : first-order; · · · : independent wavelet model.

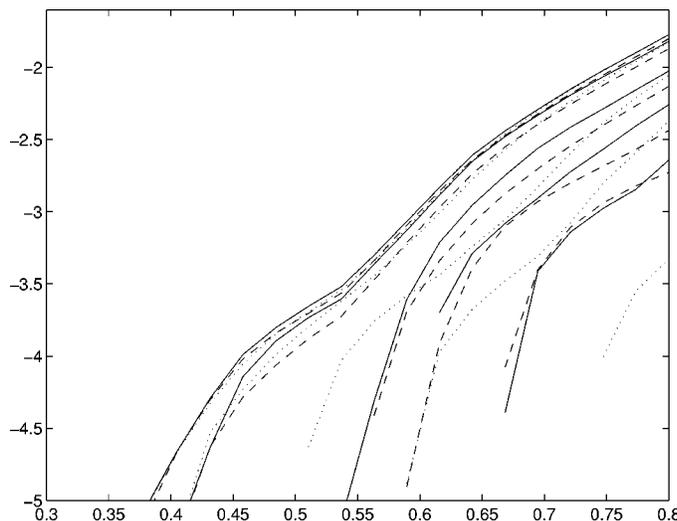


Fig. 14. Vertical axis:  $\log_{10}(\text{loss rate})$ ; horizontal axis: work load. — : single video source; ... : FARIMA(25, d, 20); - - : Algorithm 2. Normalized buffer size: 0.1, 1, 10, 30 and 100 from the top down.

traffic has a desired marginal distribution as estimated from the real trace.

Figs. 13 and 16 plot the autocorrelation functions resulting from the FARIMA(25, d, 20) and the wavelet model for both traces, respectively. These figures show that the wavelet model has a comparable performance to that of FARIMA in terms of modeling the second-order statistics. Figs. 14 and 15 give the results on the buffer loss probabilities for both traces, respectively. As observed, the performance of the wavelet model resulting from the time-scale shaping algorithm is comparable to that of FARIMA(25, d, 20) at small buffer sizes but is much improved at large buffer sizes. This shows the importance for the wavelet model and the time-scale shaping algorithm to match the non-Gaussian marginal distributions of the cumulative process at a wide range of time scales.

### VII. PERFORMANCE OF INDEPENDENT WAVELET MODELS: ANALYSIS

One observation from our empirical studies is that independent wavelet models are rather accurate measured by both the autocovariance function and the buffer overflow probability, and the Markov models which include additional dependence only improve the performance marginally. This motivates us to further access the performance of independent wavelet models through analysis. We focus our analysis in this work on a limited case when independent wavelet models are used to model an FGN process. An FGN process is of the particular interest, since it is the only long-range dependent process with an explicit autocorrelation function. In addition, the independent wavelet model of an FGN process can be expressed explicitly [28]. This makes it possible for us to analyze the corresponding buffer overflow probability and the autocorrelation explicitly,

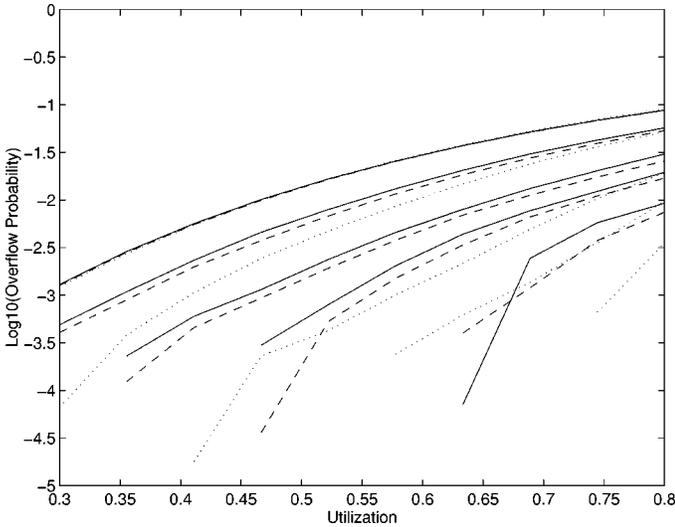


Fig. 15. Buffer response. —: data trace; - -: wavelet model; ... : FARIMA. Normalized buffer size  $B/C$ : 0.1, 1, 10, 30 and 100 from the top down.

and compare with those of the original FGN process given in the prior work [15], [46], [43].

#### A. Definitions and Notations

Let  $\hat{x}(t)$  represent an FGN process.  $x_{wK}(t)$  is a random process resulting from the (Haar) independent wavelet model, where

$$x_{wK}(t) = \sum_{j=1}^K \sum_{m=0}^{\infty} d_j^m \phi_j^m(t). \quad (22)$$

$t \geq 0$ , and  $K$  represents the limited resolution.  $\phi_j^m(t)$  is a (Haar) wavelet basis function with a scale index  $j \geq 1$ , and a shift index  $m \geq 0$ .  $d_j^m$  is an independent Gaussian random variable defined in (8). Let  $x_w(t)$  be the limit<sup>17</sup> of  $x_{wK}(t)$  with respect to  $K$ , i.e.

$$x_w(t) = \lim_{K \rightarrow \infty} x_{wK}(t). \quad (23)$$

Since  $x_{wK}(t)$  is a cyclostationary rather than a stationary process [63], we need to define the average buffer overflow probability of the independent wavelet model.

#### B. Average Buffer Overflow Probability

Consider a discrete time queue with an infinity buffer and the capacity  $c$ .  $x_w(t)$  and  $\hat{x}(t)$  are fed into two such queues at the beginning of a discrete time slot  $t$  for  $t \geq 0$ . Let  $B_{wt}$  and  $\hat{B}_t$  be the buffer sizes at the end of the  $t$ th time slot due to  $x_w(t)$  and  $\hat{x}(t)$ , respectively,<sup>18</sup> where

$$B_{wt} = \sup_{s \geq 1} (X_{wt}(s) - sc) \quad (24)$$

and

$$\hat{B}_t = \sup_{s \geq 1} (\hat{X}_t(s) - sc). \quad (25)$$

$\hat{X}_t(s)$  and  $X_{wt}(s)$  are the cumulated process,  $X_{wt}(s) = \sum_{i=0}^{s-1} x_w(t-i)$ , and  $\hat{X}_t(s) = \sum_{i=0}^{s-1} \hat{x}(t-i)$ .

<sup>17</sup>Assume the limit exists.

<sup>18</sup>As the subscript  $w$  represents for the wavelet model,  $x_{wt}(t)$  corresponds to the synthetic traffic from the independent wavelet model, and  $B_{wt}$  is the corresponding buffer size.

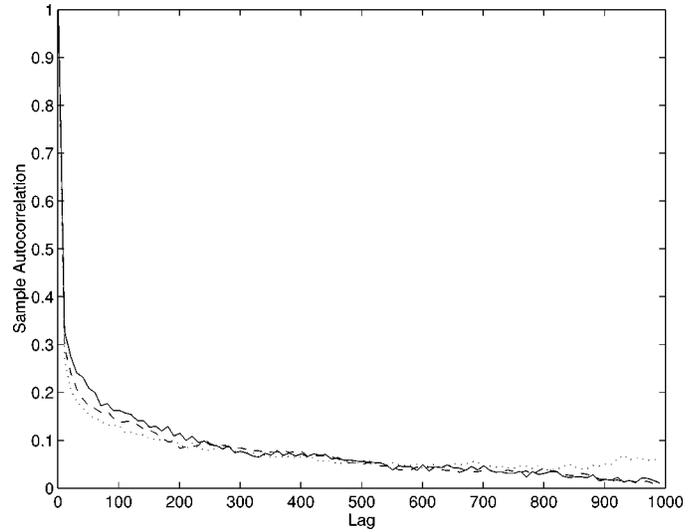


Fig. 16. Sample autocorrelation. x-axis: lag. y-axis: sample autocorrelation. —: data trace; - -: wavelet model; ... : FARIMA model.

Let  $\Pr(B_{wt} > B)$  denote the buffer overflow probability due to the independent wavelet model at time-slot  $t$ , where

$$\Pr(B_{wt} > B) = \Pr\left(\sup_{s \geq 1} (X_{wt}(s) - sc) > B\right). \quad (26)$$

Since  $x_w(t)$  is nonstationary [63] as mentioned in the previous section,  $\Pr(B_{wt} > B)$  varies with  $t$ . Therefore, to have a meaningful buffer loss rate of the independent wavelet model, we define the average buffer overflow probability as

$$\tilde{L}_{WL}(B) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \Pr(B_{wt} > B). \quad (27)$$

Since an FGN process is a stationary process, its buffer overflow probability,  $L_{FGN}(B)$ , is

$$L_{FGN}(B) = \Pr(\hat{B}_t > B) \quad (28)$$

for any positive integer  $t$ . The average buffer overflow probability of the independent wavelet model for an FGN process can then be derived and compared with the true value as shown by the following theorem.

*Theorem 3:* When the buffer size  $B$  goes to infinity, the buffer overflow probability of the independent model,  $\tilde{L}_{WL}(B)$ , and that of the original FGN process,  $L_{FGN}(B)$ , satisfies

$$\begin{aligned} & \lim_{k_0 \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log \tilde{L}_{WL}(B) \\ &= \lim_{k_0 \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log L_{FGN}(B) \\ &= \frac{(c - \mu)^2 \left(\frac{1}{c - \mu}\right)^{2(1-H)} \left(\frac{1-H}{H}\right)^{2H}}{2\sigma^2(1-H)^2} \end{aligned} \quad (29)$$

for

$$B = \frac{(c - \mu)2^{k_0}(1-H)}{H} \quad (31)$$

where  $H$  ( $0.5 < H < 1$ ) is the Hurst parameter of the FGN process, and  $k_0$  is a positive integer.

This theorem shows that when the buffer size  $B$  is assumed to be a subset of all possible values, and is large, the independent wavelet model is asymptotically close to that of the original FGN [15], [46]. In other words, the independent wavelet model can faithfully capture a long-range dependent FGN traffic. Therefore, the result shows, in this special case, the capability and performance of the independent wavelet model, as well as the feasibility of using the model for queue analysis. Meanwhile, we would like to note that even for this special case, the proofs of the theorem involve elaborate analysis. The difficulty is due to the cyclostationarity of the independent wavelet model [63] which leads to the time-varying buffer overflow probability. This is, in fact, the consequence of using an independent wavelet model by ignoring all the dependence in the wavelet domain. As the result, techniques such as the large deviation cannot be used directly, and the lower and upper bounds have to be derived for the average buffer overflow probability. The main idea of the proof is given in Appendix B, and the details can be found in [36].

### C. Average Autocorrelation Function

The performance of independent wavelet models can be further evaluated through the autocorrelation function. To deal with cyclostationarity of the independent wavelet model, we need to define the average autocorrelation function. Specifically, let the average autocorrelation function of the resolution-limited independent wavelet model be

$$\overline{R}_K(\tau) = \frac{1}{2^K} \sum_{t=1}^{2^K} \mathbf{E}(x_{wK}(t)x_{wK}(t+\tau)). \quad (32)$$

The limit defines the average autocorrelation function of  $x(t)$ ,<sup>19</sup> where

$$\overline{R}(\tau) = \lim_{K \rightarrow \infty} \overline{R}_{x_{wK}}(\tau). \quad (33)$$

Using the variance  $\sigma_j^2$  of the independent wavelet model for FGN process (see Part 1 of Theorem 1) in (33), we can derive the following theorem.

*Theorem 4:* The average autocorrelation function of the independent wavelet model for an FGN process is bounded by

$$c_1(H)|\tau|^{2H-2} \leq \overline{R}(\tau) \leq c_2(H)|\tau|^{2H-2} \quad (34)$$

where  $0.5 < H < 1$  is the Hurst parameter.  $c_1(H)$  and  $c_2(H)$  are defined by a function  $f(\alpha; H)$ , where

$$f(\alpha; H) = \alpha^{2H-2}p_1 + \alpha^{2H-3}p_2. \quad (35)$$

$p_1$  and  $p_2$  are

$$p_1 = 2 - 2^{2(1-H)} \quad (36)$$

and

$$p_2 = 2^{2(1-H)} - 1 - \frac{3(1 - 2^{2H-2})}{1 - 2^{2H-3}}. \quad (37)$$

Then

$$c_1(H) = f(1; H) \quad (38)$$

<sup>19</sup>Assuming the limit exists.

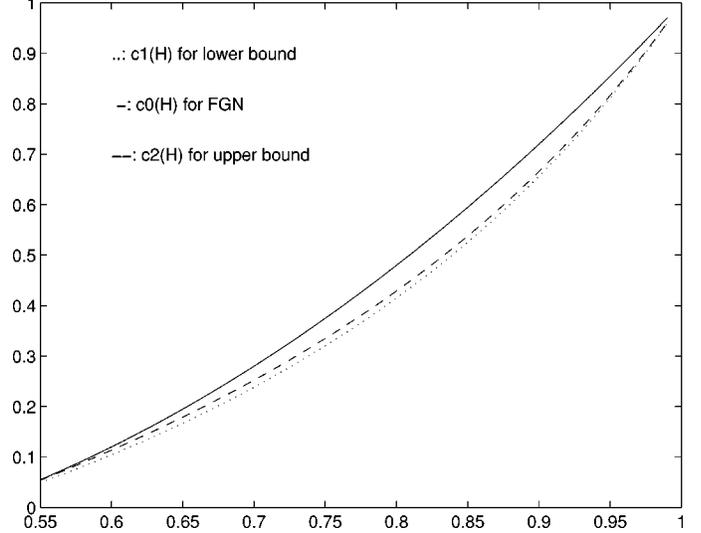


Fig. 17. x-axis:  $H$ . Solid line:  $c_0(H)$ . Dashed line:  $c_2(H)$ . Dotted line:  $c_1(H)$ .

and

$$c_2(H) = f(\alpha^*; H) \quad (39)$$

where  $\alpha^* = (-(2H - 3)p_2)/((2H - 2)p_1)$ .

The proof of Theorem 4 is in Appendix C. To understand the results given by the theorem, we recall that the autocorrelation function of an FGN process is [40]

$$R_{\text{FGN}}(\tau) = \frac{1}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}) \quad (40)$$

$$= c_0(H)|\tau|^{2H-2} + O(|\tau|^{2H-3}) \quad (41)$$

where  $c_0(H) = H(2H - 1)$ . The performance of independent wavelet models measured by autocorrelation functions can then be evaluated by comparing (34) with (41). The theorem (34) shows that the average autocorrelation function of independent wavelet model decays hyperbolically. This demonstrates that an independent wavelet model is capable of modeling long-range dependence in network traffic. In addition, the rate of decay ( $O(|\tau|^{2H-2})$ ) for large lags is the same as that of the autocorrelation function of the original FGN process. By further comparing the constants  $c_0(H)$  with  $c_1(H)$  and  $c_2(H)$  shown in Fig. 17, we can examine the difference between the autocorrelation function of the independent wavelet model and that of the original FGN process. Specifically, the relative difference between the constants ( $|(c_1(H) - c_0(H))/c_0(H)|$  and  $|(c_2(H) - c_0(H))/c_0(H)|$ ) is plotted as a function of Hurst parameter  $H$  in Fig. 18, and shown to be no more than 15% between the average autocorrelation function of the independent wavelet model and that of the FGN.

## VIII. DISCUSSION

### A. Performance and Efficiency of Wavelet Models

Why do independent wavelet models perform so well even when they neglect all the dependence in the wavelet domain? Intuitively, the (deterministic) self-similar structure of wavelets is a natural match to the statistical self-similarity of traffic. As wavelet basis functions have “absorbed” the long-range and short-range dependence by differencing the averages at all time

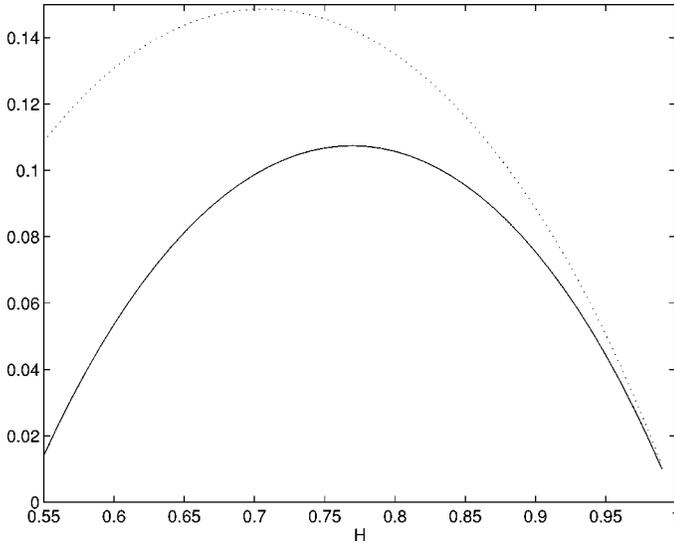


Fig. 18. x-axis:  $H$ . Solid line:  $|c_0(H) - c_1(H)|/c_0(H)$ . Dotted line:  $|c_0(H) - c_2(H)|/c_0(H)$ .

scales, the wavelet coefficients are short-range dependent. This makes it possible to model wavelet coefficients as independent (or low-order Markov dependent) random variables without losing much information. The resulting wavelet model is therefore simple and parsimonious. In addition, since there exist fast algorithms for both wavelet transforms and inverse transforms [13], our method is able to achieve the lowest computational complexity in developing the model and in generating synthetic traffic.

### B. More on the Related Work

In signal processing, [4], [5], [35] have established a general framework for multiscale representations of a random process through the dyadic tree. [63], [19], [42], [59] have shown that wavelets can provide compact representations for an FBM process. Moreover, it has proven [63] that the spectrum of the independent wavelet model of an FBM process is very close to that of  $1/f$  processes. Therefore, the independent wavelet model has been proposed to rapidly generate FBM or FGN-like synthetic sample paths. But the previous investigation on (asymptotic) correlation structure of wavelet coefficients has been focused mostly on a limited scope for FBM [63], [19], [42], [59], FGN [28], or AR(1) [14]. The correlation structure has neither been well studied for short-range dependent processes<sup>20</sup> nor for a mixture of long- and short-range dependent processes. Since network traffic has both short- and long-range dependence, we have extended the previous work to a broader class of Gaussian processes in order to study correlation structures. Wavelets were also used to estimate Hurst parameters [2], [1], [18]. The possibility of using wavelets for modeling network traffic was mentioned in [48], [17].

## IX. CONCLUSION

This work is motivated by the fact that wavelet coefficients of network traffic with complex long-range and short-range depen-

<sup>20</sup>Bounds but not the actual correlation function were derived in [14].

dence are no longer long-range dependent. Therefore, simple models can be developed in the wavelet domain. In this work, we have investigated thoroughly the independent wavelet model, the simplest wavelet model. In that, we have shown that they are capable of characterizing both long- and short-range dependent (temporal) processes through variances of wavelet coefficients at different time scales. We have derived autocorrelation functions and the queue loss rate using the independent wavelet model for the case of FGN traffic. Further, we have developed Markov wavelet models which capture the dependence among wavelet coefficients. We have compared the performance of the independent and Markov models, and show that independent wavelet models are sufficiently accurate and Markov wavelet models only improve the performance marginally. Finally, we have developed a time-scale shaping algorithm that extends the (Gaussian) wavelet models to non-Gaussian traffic. The algorithm shapes traffic at different time scales by exploiting relationships among (Haar) wavelet coefficients, scale coefficients, and the cumulative process. We also have demonstrated that the wavelet models are parsimonious, and have the lowest computational complexity achievable.

A possible future direction is to extend our initial (queue and autocorrelation function) analysis to a more general setting. Another issue of interest is to better deal with the nonstationary nature of the independent wavelet model, which we have discussed somewhat in this work. Other issues of interest include how to apply wavelet models as well as the concept of time scales to assist network design, control, and management.

## APPENDIX A PROOF OF THEOREM 2

*Proof:* Since the proof of Theorem 2(b) is similar to that of 2(a), in this appendix, we only sketch the proof of 2(a).

Since  $x(t)$  is stationary and  $d_j^m$  is obtained through the wavelet transform which is linear,  $d_j^m$  is stationary in terms of  $m$ . Without the loss of generality, we only need to consider  $d_j^0$ . From definition of Haar wavelet coefficients, we have

$$\text{Var}(d_j^0) = 2^{-j} \mathbf{E} \left( \sum_{t=0}^{2^{j-1}-1} x(t) - \sum_{t=2^{j-1}}^{2^j-1} x(t) \right)^2 \quad (42)$$

$$= 2^{-j} \mathbf{E} \left[ \left( \sum_{t=0}^{2^{j-1}-1} (x(t) - \mu) \right)^2 + \left( \sum_{t=2^{j-1}}^{2^j-1} (x(t) - \mu) \right)^2 \right. \quad (43)$$

$$\left. - \sum_{t_1=0}^{2^{j-1}-1} \sum_{t_2=2^{j-1}}^{2^j-1} 2(x(t_1) - \mu)(x(t_2) - \mu) \right]. \quad (44)$$

Through straightforward algebraic manipulations (see [38], [36] for details), we can compute the two terms of (44), and thereby prove the theorem.

APPENDIX B  
MAIN IDEA OF THE PROOF OF THEOREM 3

As the buffer loss probability of independent wavelet models is defined in the average sense to account for the nonstationarity, the proof of Theorem 3 involves comprehensive analysis to bound the average loss rate. Here, we provide the main ideas of the proof. More details can be obtained in [36], [38].

Since proving Theorem 3 is equivalent to proving that

$$\lim_{B \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log \tilde{L}_{WL}(B) \leq \lim_{B \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log L_{FGN}(B) \quad (45)$$

and

$$\lim_{B \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log \tilde{L}_{WL}(B) \geq \lim_{B \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log L_{FGN}(B) \quad (46)$$

where  $B$  is defined by (31), we need to show (45) and (46) hold respectively.

Note that the buffer overflow probability [see (26)] can be generally lower bounded [15], [60] by

$$\Pr \left( \sup_{s \geq 1} X_t(s) > cs + B \right) \geq \sup_{s \geq 1} \Pr(X_t(s) > cs + B) \quad (47)$$

and upper bounded by the union bound as

$$\Pr \left( \sup_{s \geq 1} X_t(s) > cs + B \right) \leq \sum_{s=1}^{\infty} \Pr(X_t(s) > cs + B). \quad (48)$$

For an FGN process ( $0.5 < H < 1$ ), Duffield [15] has shown that the upper and the lower bounds of the buffer overflow probability are asymptotically close to each other. That is

$$\lim_{B \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log L_{FGN}(B) \quad (49)$$

$$= \lim_{B \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log \Pr \left( \hat{X}(s^*) > cs^* + B \right) \quad (50)$$

$$= \lim_{B \rightarrow \infty} \frac{1}{B^{2(1-H)}} \log \sum_{s=1}^{\infty} \Pr \left( \hat{X}(s) > cs + B \right) \quad (51)$$

$$= \frac{(c - \mu)^2 \left( \frac{1}{c - \mu} \right)^{2(1-H)} \left( \frac{1-H}{H} \right)^{2H}}{2\sigma^2(1-H)^2} \quad (52)$$

where  $s^* = \arg \sup_{s > 0} \Pr(\hat{X}_t(s) > cs + B)$  is the so-called critical time scale [45].

Therefore, to prove (45), it is sufficient to prove that

$$\Pr(X_t(s) > cs + B) \leq \Pr \left( \hat{X}_t(s) > cs + B \right) \quad (53)$$

for all integers  $t$  and  $s$ , and using (48) and (51). On the other hand, the lower bound [see (46)] can be obtained through (47) and (50), if we can show that

$$\Pr(X_t(s^*) > cs^* + B) = \Pr \left( \hat{X}_t(s^*) > cs^* + B \right) \quad (54)$$

where  $s^*$  is the critical time scale of the FGN, holds under certain conditions on  $B$  and  $t$ .

Since  $X_t(s)$ , the cumulative process of the independent wavelet model, is a Gaussian random variable for fixed  $t$  and  $s$ , a key step for us to derive the theorem through proving (53) and (54) is to derive the variance of  $X_t(s)$ , and relate it to that of  $\hat{X}_t(s)$ .

Because  $X_t(s)$  is a function of  $t$ , our proof contains two main steps. The first step is to show that the conclusion holds for the special time slot,  $N-1$  ( $N = 2^K$  and  $K$  is a large integer). This can be done through deriving the wavelet representation of  $X_{N-1}(s)$  and calculating the variance of  $X_{N-1}(s)$ . The second step is to relate the variance of  $X_t(s)$ , for any  $t \geq 0$ , to that of  $X_{N-1}(s)$ . The proofs for these two steps are done mostly through algebraic manipulations, and are lengthy due to the nonstationary nature of independent wavelet models. (Please refer [36] for details.) Intuitively, because wavelets provide the multi-scale representation of a signal [39], which in our case is traffic, the cumulated processes resulting from the independent wavelet model  $X_t(s)$  are equal in probability to the cumulated process of an FGN process  $\hat{X}_t(s)$  at special set of  $t$  and  $s$ . This results in the fact that  $X_t(s)$  is very close to  $\hat{X}_t(s)$  for the rest of  $t$  and  $s$ . We prove that the above intuitions are true in [36], and therefore prove the theorem.

APPENDIX C  
PROOF OF THEOREM 4

To prove the theorem, we first need to derive an expression for the average autocorrelation function.

Inserting the wavelet representation of  $x_{wK}(t)$  defined in (22) into (33), we can obtain through some algebraic manipulations

$$\bar{R}_K(\tau) = \sum_{j=1}^K \sigma_j^2 h_j(\tau) \quad (55)$$

where

$$h_j(\tau) = \begin{cases} \left(1 - \frac{3\tau}{T_j}\right) & 0 \leq \tau < T_j/2 \\ \left(\frac{\tau}{T_j} - 1\right) & T_j/2 < \tau < T_j \\ 0 & \text{otherwise.} \end{cases} \quad (56)$$

$\sigma_j$  represents the variance of wavelet coefficients of an FGN process and is given in Theorem 1 as

$$\sigma_j^2 = 2^{j(2H-1)} \left(2^{2(1-H)} - 1\right). \quad (57)$$

Then the average autocorrelation function [see (55)] of an independent wavelet model for an FGN process is

$$\bar{R}(\tau) = \sum_{j=1}^{\infty} \left(2^{2(1-H)} - 1\right) 2^{j(2H-2)} h_j(\tau). \quad (58)$$

Let  $k = \lfloor \log_2(\tau) \rfloor + 1$ . Replacing  $h_j(\tau)$  by (56), we have

$$\begin{aligned} \bar{R}(\tau) &= \left(2^{2(1-H)} - 1\right) \left( 2^{k(2H-2)} \left(2^{2(1-H)} - 1\right) \left(\frac{\tau}{2^k} - 1\right) \right. \\ &\quad \left. + \sum_{j=k+1}^{\infty} 2^{j(2H-2)} \left(1 - \frac{3\tau}{2^j}\right) \right) \end{aligned} \quad (59)$$

$$= 2^{k(2H-2)} \left( 2 - 2^{2(1-H)} + \frac{\tau}{2^k} \left( 2^{2(1-H)} - 1 - \frac{3(1 - 2^{-2(1-H)})}{2(1 - 2^{-(3-2H)})} \right) \right). \quad (60)$$

The above equation is obtained based on the fact that  $h_j(\tau)$  is zero for  $j < k$ .

Equation (60) can be further written as

$$\bar{R}(\tau) = f(\alpha; H) \tau^{2(1-H)} \quad (61)$$

where

$$\alpha = \frac{2^k}{\tau} \quad (62)$$

and

$$f(\alpha; H) = \alpha^{2H-2} p_1 + \alpha^{2H-3} p_2. \quad (63)$$

$p_1$  and  $p_2$ , defined by

$$p_1 = 2 - 2^{2(1-H)} \quad (64)$$

and

$$p_2 = 2^{2(1-H)} - 1 - \frac{3(1 - 2^{2H-2})}{1 - 2^{2H-3}} \quad (65)$$

are weighting functions which only depend on the Hurst parameter  $H$ .

Using the above expressions, we can bound the function  $f(\alpha; H)$  as follows.

Because  $\tau \leq 2^k \leq 2\tau$  and (62), we have  $1 \leq \alpha \leq 2$ . Through some algebraic manipulations, it can be shown that  $f(1; H) = f(2; H)$ . Therefore, there is an extreme value between  $1 < \alpha < 2$ . Through setting the derivative of  $f(\alpha; H)$  with respect to  $\alpha$  to zero, the only root can be found to be

$$\alpha^* = \frac{-(2H-3)p_2}{(2H-2)p_1}. \quad (66)$$

Since the derivative of  $f(\alpha; H)$  with respect to  $\alpha$  is nonnegative in (1, 2), we have

$$f(1; H) \leq f(\alpha; H) \leq f(\alpha^*; H) \quad (67)$$

for  $1 \leq \alpha \leq 2$ . The conclusion follows by setting

$$c_1(H) = f(1; H) \quad (68)$$

and

$$c_2(H) = f(\alpha^*; H). \quad (69)$$

Q.E.D

#### ACKNOWLEDGMENT

The authors would like to thank O. Rose and M. Garrett for making the video traces available, W. Willinger, B. Ryu, J. Thomas, X. Tian, J. Modestino and K. Vastola for helpful discussions, and the anonymous reviewers and the Associate Editor for helpful comments.

#### REFERENCES

- [1] P. Abry, P. Goncalves, and P. Flandrin, "Wavelet, spectrum analysis and  $1/f$  processes," *Lecture Notes in Statistics*, vol. 103, pp. 15–30, 1995.
- [2] P. Abry and V. Darryl, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2–16, Jan. 1998.
- [3] A. T. Andersen and B. F. Nielsen, "An application of superpositions of two state Markovian sources to the modeling of self-similar behavior," in *Proc. IEEE INFOCOM*, Kobe, Japan, 1997, pp. 196–204.
- [4] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 766–784, Mar. 1992.
- [5] M. Basseville, A. Benveniste, and A. S. Willsky, "Multiscale autoregressive processes," *IEEE Trans. Signal Processing*, vol. 40, pp. 1915–1954, 1992.
- [6] J. Beran, *Statistics for Long-Memory Processes*. London, U.K.: Chapman & Hall, 1994.
- [7] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, pp. 1565–1579, 1995.
- [8] J. Choe and N. B. Shroff, "A new method to determine the queue length distribution at an ATM multiplexer," in *Proc. IEEE INFOCOM*, Kobe, Japan, 1997, pp. 549–556.
- [9] —, "New bounds and approximations using extreme value theory for the queue length distribution in high-speed networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, 1998, pp. 364–371.
- [10] K. C. Chou, A. S. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 464–478, 1994.
- [11] D. R. Cox, "Long-range dependence: A review," in *Statistics: An Appraisal*, H. D. David and H. T. David, Eds. Ames, IA: Iowa State Univ. Press, 1984, pp. 55–74.
- [12] M. Crovella and A. Bestavros, "Self similarity in world wide web traffic: Evidence and possible causes," *IEEE/ACM Trans. Networking*, vol. 5, pp. 835–846, 1997.
- [13] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
- [14] R. W. Dijkerman and R. R. Mazumdar, "Wavelet representations of stochastic processes and multiresolution stochastic models," *IEEE Trans. Signal Processing*, vol. 42, pp. 1640–1652, July 1994.
- [15] N. G. Duffield and N. O'Connell, "Large deviations and overflow probabilities for the general single server queue, with applications," Dublin Inst. for Advanced Studies, DIAS-STP-93-30, 1993.
- [16] A. Elwalid, D. Heyman, T. V. Lakshman, D. Mitra, and A. Weiss, "Fundamental bounds and approximations for ATM multiplexers with applications to video conferencing," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1004–1016, June 1995.
- [17] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent packet traffic," *IEEE/ACM Trans. Networking*, vol. 4, pp. 209–223, 1996.
- [18] P. W. Fieguth and A. S. Willsky, "Fractal estimation using models on multiscale trees," *IEEE Trans. Signal Processing*, vol. 44, pp. 1297–1300, 1996.
- [19] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–917, Mar. 1992.
- [20] V. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Commun. Mag.*, vol. 32, pp. 70–80, 1994.
- [21] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," in *Proc. SIGCOM*, London, U.K., 1994, pp. 269–279.
- [22] M. Goldberg, "Applications of wavelets to quantization and random process representations," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1993.
- [23] B. Hajek and L. He, "On variations of queue response for inputs with the same mean and autocorrelation function," presented at the Conf. Information Sciences and Systems, Princeton Univ., Princeton, NJ, 1996.
- [24] D. Heyman, "The GBAR source model for VBR video conferences," *IEEE/ACM Trans. Networking*, vol. 5, pp. 554–560, 1997.
- [25] D. P. Heyman and T. V. Lakshman, "What are the implications of long-range dependence for VBR-video traffic engineering?," *IEEE/ACM Trans. Networking*, vol. 4, pp. 301–317, June 1996.
- [26] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye, "Modeling and simulation of self-similar variable bit rate compressed video: A unified approach," in *Proc. SIGCOM*, Stanford, CA, 1996, pp. 114–125.
- [27] P. R. Jelenkovic, A. A. Lazar, and N. Semret, "The effect of multiple time scales and subexponentiality in MPEG video streams on queueing behavior," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1052–1071, June 1997.
- [28] L. M. Kaplan and C. J. Kuo, "Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis," *IEEE Trans. Signal Processing*, vol. 42, pp. 3554–3562, Dec. 1993.
- [29] L. Kleinrock, *Queueing Systems*. New York: Wiley, 1975.

- [30] M. Krnz and S. K. Tripathi, "Modeling bit rate variations in MPEG source," Preprint, 1995.
- [31] M. M. Krnz and A. M. Makowski, "Modeling video traffic using  $m/g/\infty$  input processes: A compromise between Markovian and lrd models," *IEEE J. Select. Areas Commun.*, pp. 733–748, 1998.
- [32] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.
- [33] W. E. Leland and D. V. Wilson, "High time resolution measurements and analysis of LAN traffic: Implications for LAN interconnection," in *Proc. IEEE INFOCOM*, Bal Harbour, FL, 1991, pp. 1360–1366.
- [34] S. Q. Li and C.-L. Hwang, "Queue response to input correlation functions: Discrete spectral analysis," *IEEE/ACM Trans. Networking*, pp. 317–329, Feb. 1993.
- [35] M. R. Luetgen, W. C. Karl, and A. S. Willsky, "Multiscale representations of Markov random fields," *IEEE Trans. Signal Processing*, vol. 41, pp. 3377–3396, 1993.
- [36] S. Ma, "Traffic modeling and analysis," Ph.D. dissertation, Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, 1998.
- [37] S. Ma and C. Ji, "Modeling video traffic using wavelets," in *35th Annu. Allerton Conf. Communication, Control, and Computing*, Oct. 1997.
- [38] —, "Modeling video traffic in the wavelet domain," in *Proc. IEEE INFOCOM*, San Francisco, CA, Apr. 1998, pp. 201–208.
- [39] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, 1989.
- [40] B. B. Mandelbrot and J. W. Van Ness, "Fractal Brownian motions, fractal noises, and applications," *SIAM Rev.*, vol. 10, pp. 422–437, 1968.
- [41] M. Mandjes, I. Saniee, S. Stolyar, and R. Schmidt, "Load characterization, overload prediction and load anomaly detection for voice-over IP traffic," in *Proc. Allerton Conf.*, 2000.
- [42] E. Masry, "The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 39, pp. 260–264, 1993.
- [43] G. Mayor and J. Silvester, "Time scale analysis of an ATM queueing system with long-range dependent traffic," in *Proc. IEEE INFOCOM*, Kobe, Japan, 1997, pp. 205–212.
- [44] B. Melamed, D. Raychaudhuri, B. Sengupta, and J. Zdepski, "TES-based video source modeling for performance evaluation of integrated networks," *IEEE Trans. Commun.*, vol. 42, pp. 2773–2783, 1994.
- [45] M. Montgomery and G. De Veciana, "On the relevance of time scales in performance oriented traffic characterizations," in *Proc. IEEE INFOCOM*, San Francisco, CA, 1996, pp. 513–520.
- [46] I. Norros, "A storage model with self-similar input," *Queueing Syst.*, vol. 16, pp. 387–396, 1994.
- [47] —, "On the use of fractional Brownian motion in the theory of connectionless networks," *IEEE J. Select. Areas Commun.*, vol. 13, Aug. 1995.
- [48] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Networking*, vol. 3, pp. 226–244, 1995.
- [49] D. Reininger, D. Raychaudhuri, B. Melamed, B. Sengupta, and J. Hill, "Statistical multiplexing of VBR MPEG compressed video on ATM networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, 1993, pp. 919–926.
- [50] V. J. Ribeiro, R. H. Riedi, M. S. Crouse, and R. G. Baraniuk, "Multiscale queueing analysis of long-range-dependent network traffic," in *Proc. IEEE INFOCOM*, vol. 2, 2000, pp. 1026–1035.
- [51] V. J. Ribeiro, R. H. Riedi, M. S. Crouse, and R. G. Baraniuk, "Simulation of non-Gaussian long-range-dependent traffic using wavelets," in *Proc. ACM SIGMETRICS*, 1999, pp. 1–12.
- [52] S. Robert and J. Y. Le Boudec, "Can self-similar traffic be modeled by Markovian processes," in Stockholm Management Committee Meeting, 1995, s.Z. COST242 Tech. Doc. TD95-26. [Online] Available [http://r-cwww.epfl.ch/PS\\_files/robert\\_stockholm.p](http://r-cwww.epfl.ch/PS_files/robert_stockholm.p).
- [53] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM traffic engineering," Univ. of Wurzburg, Wurzburg, Germany, Tech. Rep. 101, 1995.
- [54] B. K. Ryu, "Fractal network traffic: From understanding to implications," Ph.D. dissertation, Columbia Univ., New York, 1996.
- [55] B. K. Ryu and A. Elwalid, "The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities," in *Proc. ACM SIGCOM*, Stanford, CA, 1996, pp. 3–14.
- [56] R. Sherman, M. S. Taqqu, W. Willinger, and D. V. Wilson, "Statistical analysis of Ethernet LAN traffic at the source level," in *Proc. ACM SIGCOM*, Stanford, CA, 1995, [Online] Available <ftp://ftp.bell-core.com/pub/walter/sigcomm95.ps.Z>.
- [57] P. Skelly, M. Schwartz, and S. Dixit, "A histogram-based model for video traffic behavior in an ATM multiplexer," *IEEE/ACM Trans. Networking*, vol. 1, pp. 446–459, Aug. 1993.
- [58] StatSci, "S-PLUS guide to statistical and mathematical analysis, Version 3.3," Statistical Sciences, a division of MathSoft, Inc., Seattle, WA, 1995.
- [59] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 904–909, 1992.
- [60] D. Tse, "Variable-rate lossy compression and its effects on communication networks," Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, M.I.T., Cambridge, MA, 1994.
- [61] D. Tse, R. G. Gallager, and J. N. Tsitsiklis, "Statistical multiplexing of multiple time-scale Markov streams," *IEEE J. Select. Areas Commun.*, vol. 43, pp. 1566–1579, 1995.
- [62] G. Veciana and J. Walrand, "Effective bandwidths: Call admission, traffic policing and filtering for ATM networks," *Queueing Syst.*, vol. 20, pp. 37–59, 1995.
- [63] G. W. Wornell and A. V. Oppenheim, "Wavelet-based representations for a class of self-similar signals with application to fractal modulation," *IEEE Trans. Inform. Theory*, vol. 38, pp. 785–800, 1992.
- [64] F. Yegenoglu, B. Jabbari, and Y.-Q. Zhang, "Motion classified autoregressive modeling of variable bit rate video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 42–53, Mar. 1993.



**Sheng Ma** (M'01) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, R.O.C., in 1987. He received the M.S. and Ph.D. degrees in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1995 and 1998, respectively.

He joined the IBM T. J. Watson Research Center, Hawthorne, NY, as a Research Staff Member in 1998. He is now a Manager of the Machine Learning for Systems department. His current research interests are in the areas of network traffic modeling and control,

network and computer system management, machine learning, and data mining.

**Chuanyi Ji** received the B.S. degree (honors) from Tsinghua University, Beijing, R.O.C., in 1983, the M.S. degree from the University of Pennsylvania, Philadelphia, in 1986, and the Ph.D. degree from the California Institute of Technology, Pasadena, in 1992, all in electrical engineering.

She joined the faculty at Rensselaer Polytechnic Institute, Troy, NY, in November 1991, where she is now an Associate Professor of electrical and computer systems engineering. She spent her sabbatical at Bell Labs Lucent in 1999, and was a visiting faculty at LIDS, MIT, in the fall of 2000. Her research interests are in the areas of computer communication networks and adaptive systems, stochastic modeling, analysis and control heterogeneous network traffic in both wired and wireless networks, network management, adaptive algorithms, statistics and Inform. Theory, and their applications in networking and communication.

She is the recipient of the Early Career Award from Rensselaer Polytechnic Institute in 2000, the CAREER Award from the National Science Foundation in 1995, and the Ming-Li Scholarship from the California Institute of Technology in 1990.