Modeling Network Traffic in Wavelet Domain*

Sheng Ma

IBM T.J. Watson
Hawthorne, NY 10532
Tel: (914) 784 7837
shengma@us.ibm.com

Chuanyi Ji
Dept. of ECSE
Rensselaer Polytechnic Inst.
Troy, NY 12180
Tel: (518) 276 6534
chuanyi@ecse.rpi.edu

January 1999

Abstract

A significant discovery from this work is that although network traffic has the complicated short- and long-range temporal dependence, the corresponding wavelet coefficients are no longer long-range dependent. Therefore, a "short-range" dependent process can be used to model network traffic in the wavelet domain. Both independent and Markov models are investigated. Theoretical analysis shows that the independent wavelet model is sufficiently accurate in terms of the buffer overflow probability for Fractional Gaussian Noise traffic. Any model which captures additional correlations in the wavelet domain only improves the performance marginally.

The independent wavelet model is then used as a unified approach to model network traffic including VBR MPEG video and Ethernet data. The computational complexity is $O(N)$ for developing such wavelet models and generating synthesized traffic of length N, which is among the lowest attained.

*To be appeared in special issue of International Journal on Chaos Theory and Applications

1 Introduction

As high speed networks aim at providing integrated services to various applications with diverse statistical characteristics and Quality of Service (QoS) requirements, modeling different types of traffic, and generating synthesized traffic from a model will be crucial to network design and network simulation. In this work, we will focus on these important issues.

Two of the key issues [14] of accurate traffic modeling are performance and computational efficiency. The former addresses the ability of a model to characterize significant statistical properties in network traffic. The latter deals with the complexity of a model, and the computational complexity needed to develop such a model and to generate a large volume of synthesized traffic. Both issues are important for simulating high-speed networks with extremely low loss probability.

Although numerous studies have been conducted on traffic modeling and performance analysis (see for example [14][33] [38][43][41] and references herein), new traffic models are needed to deal with new properties introduced by data and video traffic. One of such significant statistical properties is the so-called long-range dependence (LRD) recently found in Ethernet data[21] and VBR video traffic[15]. It is characterized by a hyperbolically decaying autocorrelation which is different from an exponentially decaying autocorrelation of Markov-type of models. The long-range dependence questions[9][31][10] the feasibility of traditional Markov models[1], especially in such applications as video-on-demand, broadcast video, and most data communications where delay requirements are less stringent[2]. Furthermore, an even more complex statistical property, the co-existence of long-range dependence (LRD) and strong short-range dependence[5][15], has been found in VBR video traces. This means that the auto-correlation function of video traffic behaves similarly to that of long-range dependent processes at the large lags, and to that of short-range dependent processes such as DAR processes[17] at the small lags[5][15][21][37]. In addition, periodic statistics have been shown in the standardized Motion Picture Expert Group (MPEG) encoded sources due to periodic appearances of different types of frames[35]. All these suggest a complex temporal behavior of network traffic, and make accurate traffic modeling a challenging task.

Several models have been developed to model long-range dependence in network traffic, such as Fractional Gaussian Noise (FGN) processes[21], the fractal point processes[36] and methods based on chaotic maps[12][11]. Since these models can not model strong short-range dependence (SRD) in network traffic, they can hardly be used to model video traffic. Models, which can model both long-range and short-range dependence, include FARIMA models[15], a model based on Importance Sampling[18], scene-based models[19] and the Markov Modulated Processes[1][34]. A common feature of all these methods is that they model both long-range and short-range dependence in the time domain. Among these methods, the scene-based modeling[19] and the Markov Modulated models[1][34] provide a physically interpretable model to include both long-range and short-range dependence. However, due to the dynamic and stochastic nature of network traffic, it is

[1] It is possible to use a Markov-type of model[1] to model long-range dependent processes. However, it usually results in a complicated model with a complex structure and too many parameters.

[2] As shown in [17][37], a short-range dependent model may be sufficient to model real-time traffic in terms of a buffer loss rate because of a small relevant critical time-scale[37]. However, long-range dependent models may be needed for non-real-time applications. In addition, long-range dependent models may be proven to be important for estimating other quantities such as a loss pattern in addition to a buffer loss rate [10].

difficult to accurately define and segment network traffic into different states of a Markov model. The rest of the methods all suffer the computational complexity too high to be used for generating a large volume of synthesized traffic[15]. A more computationally efficient method based on Fast Fourier Transform has been proposed [32] to model Ethernet traffic in the frequency domain. Another method based on Markov models has been proposed to model the frequency components of video traffic[23]. Both methods suggest that interesting properties of either Ethernet or video traffic could be investigated in the Frequency domain. However, none of the methods are yet able to capture the long-range and the short-range dependence simultaneously.

Therefore, the question remains open on how to develop a computationally efficient model which can capture both long-range and strong short-range dependence in network traffic. In this work, we will tackle this problem by developing a new method based on wavelets. Instead of modeling network traffic directly in the time-domain, we model the statistical properties in the wavelet domain.

Our work is motivated by previous work in the signal processing community on using the independent wavelet model to represent and model a Fractional Brownian Motion (FBM) process. Particularly, [2][3][24] have established a general framework for multiscale representations of a random process through the dyadic tree. [42][13][29][40] have shown that wavelets can provide compact representations for an FBM process. Moreover, Wornell[42] has 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 or FGN-like synthesized sample path.

However, since we are interested in modeling long- and short- range dependent network traffic, we need to further address (1) whether wavelet domain modeling is suitable for modeling a mixture of short- and long-range dependence; (2) what is the quantitative networking related performance, such as auto-correlation function and buffer overflow probability; and (3) how to model real network traffic, such as data traffic and video traffic. In this work[3], we will first demonstrate that the wavelet coefficients of a long-range dependent temporal process are no longer long-range dependent. Then the wavelet coefficients can be modeled by simple statistics corresponding to the "short-range dependence" alone. We will then show that a simple wavelet model based on independence assumptions is capable of capturing both long-range dependence for more general LRD processes and short-range dependence as well, and thus provides a parsimonious and unified model to capture both long-range and short-range dependence in network traffic. In addition, since computational complexity of wavelet transforms and inverse transforms are in the order of N with N being the length of synthesized traffic, our wavelet models can rapidly generate synthesized network traffic with a computational complexity $O(N)$, and thereby provide one of the most efficient methods to synthesize high quality network traffic. Furthermore, because wavelets have a natural link to the time-scale (multi-resolution) representation[28], they enable time-scale modeling and provide a systematic method for modeling periodic MPEG video traffic.

The format of the paper is as follows. In Section 2, we will provide background knowledge. In Section 3, we will give outlines of our approach. In Section 4, we will investigate statistical properties of wavelet coefficients, and develop as well as evaluate wavelet models for the long-range and the short-range dependent Gaussian processes, respectively. In Section 5, we will provide statistical properties of wavelet coefficients for real network traffic, and derive our algorithms based on these properties. We will further discuss our approach

[3]The early results of this work has been published in [26][27]

in Section 6, and then conclude the paper.

2 Background

2.1 Long-Range versus Short-Range Dependence

Roughly speaking, Long-Range Dependence (LRD) can be considered as a phenomenon that current observations are significantly correlated to the observations that are farther away in time. This phenomenon is of particular interest to traffic modeling, since it has been discovered recently that both Ethernet traffic[21] and video sources[5][15] possess long-range dependence.

One formal definition [15] 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 [4]. This implies that the correlation $r(k)$ decays asymptotically as a hyperbolic function of k, i.e., $r(k) \sim O(k^{-(2-2H)})$ for $k \geq 0$. H $(0.5 < H < 1)$ is the so-called Hurst parameter, which is an important quantity used to characterize the LRD. Examples of such long-range dependent processes include the Fractional Gaussian Noise (FGN) process and the Fractional Autoregressive Integrated Moving Average process (FARIMA). The nature of these random processes is "self-similar", i.e., the statistical properties of these processes are invariant at different time scales [4][6]. In particular, FGN is a Gaussian process, and can be completely specified by three parameters. 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 (SRD) in network traffic. More details can be found in [4] on FGN and FARIMA processes. Examples of short-range dependent random processes include Auto-Regressive (AR) processes and Auto-Regressive-Moving-Average (ARMA) processes with exponentially decaying correlation functions. That is, $r(k) \sim \rho^k$ $(-1 < \rho < 1)$.

2.2 Criteria for Measuring The Performance of A Model

Both the auto-correlation function and the buffer loss rate will be used as performance measures for wavelet models[5] in this work. The auto-correlation 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 auto-correlation function of network traffic in a very long range. The buffer loss rate is chosen as one other criterion to measure the performance of a model, since one of important goals for traffic modeling is to help designing the buffer size of a server[6], and estimating the loss rate as a measure for the Quality of Service (QoS). Therefore, synthesized traffic generated by a good traffic model should have a similar buffer loss rate to that of the actual traffic.

[4] Please see [21][15][6][4] for other definitions and properties of the LRD.

[5] Because the marginal distribution can be shaped to match original traffic, we do not consider it as a criterion.

[6] A server can be modeled as a single queue with capacity C and a buffer size B.

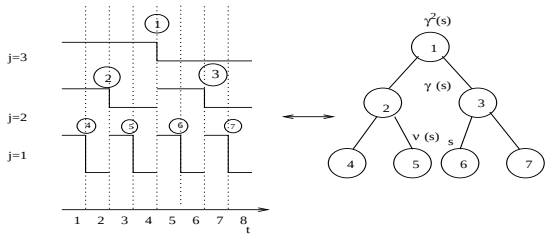

Figure 1: Left figure shows the Haar wavelet basis functions. Right figure illustrates the corresponding tree diagram and two types of operations. The number in the circle represents the one dimension index of the wavelet basis functions. For example, the equivalent notation of d_2^1 is d_3. s, $\nu(s)$ and $\gamma(s)$ represent the one dimension 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.

2.3 Wavelet Transformation

The model we will develop on network traffic is based on wavelets. Wavelets are complete orthonormal bases which can be used to represent a signal as a function of time[7]. In $L^2(R)$, discrete wavelets can be represented as

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

where j and m are positive integers. j represents the dilation, which 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 self-similar to each other.

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

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

where $0 \le t < 2^K$. ϕ_0 is equal to the average value of $x(t)$ over $t \in [0, 2^K - 1]$. d_j^m's are wavelet coefficients and can be obtained through the wavelet transform

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

The mother wavelet we choose in this work is the Haar wavelet. It is defined as

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

The Haar wavelet basis have been illustrated in the left figure of Figure 1 for $K = 3$. The right figure of Figure 1 shows the corresponding tree diagram[3]. In the tree diagram, we define the one-dimension index of wavelet coefficients. This will be used later. The motivation for using Haar wavelets is due to simplicity which results in computationally efficient methods for the wavelet transform and the inverse transform. The computational complexity of the wavelet transform and the inverse transform is in the order of $O(N)$[7], where $N = 2^K$ is the length of the time series.

When $x(t)$ is a random process, which is of the interest to this work, the corresponding wavelet coefficients d_j^m's define a two-dimensional random processes (see [16][42][3] and references herein for details) in terms of j and m. 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 using wavelets to model network traffic, i.e., to statistically model wavelet coefficients d_j^m's.

3 Outlines of Our Approach

To use wavelets to model network traffic, we will investigate the following issues: (1) Why is it good to use wavelets? (2) What statistical properties of wavelet coefficients are pertinent to model network traffic? (3) How to include these important properties into wavelet models? (4) What is the performance of the wavelet models measured both experimentally and theoretically?

As the first step, we will investigate why we should consider wavelet models, and what statistical properties are important through investigating the correlation structure of wavelet coefficients for well-known long-range and short-range dependent Gaussian processes. We will show that a key advantage of using wavelets is their ability to reduce the dependence in the originally temporal process so significantly that the wavelet coefficients only possess the "short-range" dependence in the wavelet domain. We will then develop algorithms to model wavelet coefficients using either independent or Markov models. As empirical results show that statistically independent wavelet coefficients will lead to an accurate model in terms of buffer loss rate, we will provide theoretical analysis on the buffer overflow probability for the independent wavelet model when the workload is FGN. The theoretical results will confirm that the wavelet model based on independent wavelet coefficients is sufficient. Any model which captures dependence among wavelet coefficients will only improve the buffer loss rate marginally.

To model real network traffic, we will extend the wavelet models developed for Gaussian processes to incorporate the marginal distributions of the wavelet coefficients and the periodic structure of the (MPEG) video traffic. The performance of our wavelet models will be tested experimentally on three different video traces and one Ethernet data trace, and compared with that of other models.

4 Wavelet Modeling of Long-Range and Short-Range Dependent Gaussian Processes

As the first step to investigate the feasibility and advantage of wavelets to model network traffic, we will investigate wavelet model in this section for well-known long-range and short-range dependent Gaussian processes, respectively.

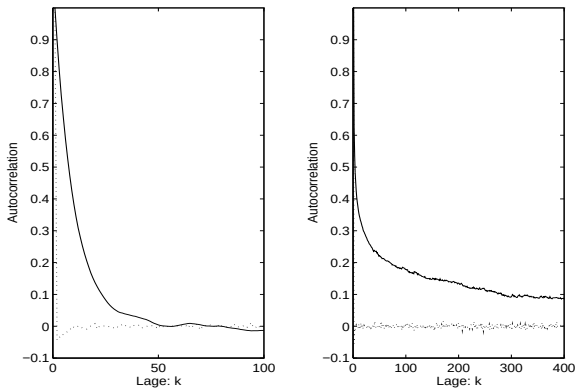

Figure 2: Solid line: Autocorrelation coefficients of the original process. Dotted line: the normalized auto-correlation of wavelet coefficient, i.e., $\frac{E(d_1^m d_1^{m+k})}{\sigma_{d_1^m}\sigma_{d_1^{m+k}}}$. The left figure is for the AR(1) process. The right figure is for the FARIMA(0,0.4,0) process.

4.1 Theoretical Studies on The Correlation Structure of Wavelet Coefficients

4.1.1 The Correlation Structure of Wavelet Coefficients of LRD Processes

The correlation structure of (long-range dependent) FGN process has been investigated extensively in [20][42][13], and can be summarized as follows.

Theorem 1 *(***Kaplan and Kuo[20], Flandrin[13]***)* *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 $2^{j(2H-1)}(2^{2(1-H)}-1)\sigma^2$, where σ 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

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

where $H' = 1 - H$.

We recall that the corresponding temporal auto-correlation of Fractional Gaussian Noise decays at a rate $O(|k|^{-2(1-H)})$, where k is the lag between two samples and $0.5 < H < 1$. This rate leads to a divergent summation of the auto-correlations. 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. This is because that the correlation changes from mean revert $(0.5 < H < 1)$ in time domain to mean avert $(0 < H' < 0.5)$ in the wavelet domain. Figure 2 illustrates how drastic the reduction is by comparing the autocorrelation function of the original $FARIMA(0, 0.4, 0)$ process to the normalized auto-correlation function [7] for d_1^m and d_1^{m+k}[8].

[7] It can be easily shown that the time series d_j^m for fixing j is stationary in terms of m. Therefore, the auto-correlation exists.
[8] $FARIMA(0, 0.4, 0)$ is an asymptotically self-similar process and is very similar to a FGN process.

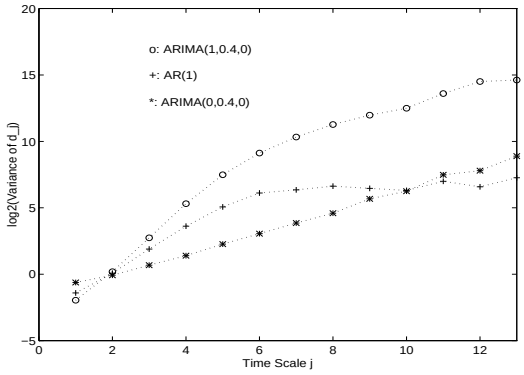

Figure 3: Log 2 of Variance of d_j versus the time-scale j

4.1.2 The Correlation Structure of Wavelet Coefficients of SRD processes

For short-range dependent processes, we derive the correlation wavelet coefficients given in a theorem below[9]

Theorem 2 *Let $x(t)$ be a zero mean wide-sense-stationary (discrete) Gaussian process with the auto-correlation $r(k)$, where $r(k) = \sigma^2\rho^{|k|}$ with $\mid \rho \mid < 1$, k is an integer and σ^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 $\sigma^2(1 + \frac{2\rho}{1-\rho} - \frac{3\rho}{(1-\rho)^2 2^{j-1}}) + O(\rho^{2^{j-1}})$.
(2) for $m_1 2^{j_1} - (m_2+1)2^{j_2} > 0$,

$$E(d_{j_1}^{m_1} d_{j_2}^{m_2}) = 2^{\frac{-j_1-j_2}{2}} \rho^{m_1 2^{j_1} - (m_2+1)2^{j_2}} (1-\rho^{2^{j_2-1}})^2 (1-\rho^{2^{j_1-1}})^2 \frac{\rho}{(1-\rho)^2}\sigma^2. \quad (6)$$

The sketch of the proof of the theorem can be found in Appendix A. Details are given in [25].

This theorem shows that the correlation of wavelet coefficients decays exponentially as $|m_1 2^{j_1} - m_2 2^{j_2}|$, the shortest distance between supports of two wavelet base functions, and therefore remains to be short-range dependent in the wavelet domain[10]. Figure 2 compares the (temporal) correlation for an $AR(1)$ process with that of its wavelet coefficient for $j_1 = j_2 = 1$ to illustrate the rate of decays.

4.1.3 Implications of The Theorems

One of the implications of the two theorems is that the variances of wavelet coefficients for a long-range dependent process are very different from those for a short-range dependent process as indicated by Theorems 1 and 2. To illustrate this, Figure 3 draws the variances of wavelet coefficients for a long-range dependent process FARIMA$(0, 0.4, 0)$, a mixture of long-range and short-range dependent process FARIMA$(1, 0.4, 0)$, and a short range dependent process AR(1). As observed from the figure, the variance of LRD increases with

[9] There exist no previous results on the explicit correlation structure of wavelet coefficients for discrete processes except the bounds for some of the continuous random processes[8].

[10] The rate of decay is even faster than the corresponding correlation in the time domain.

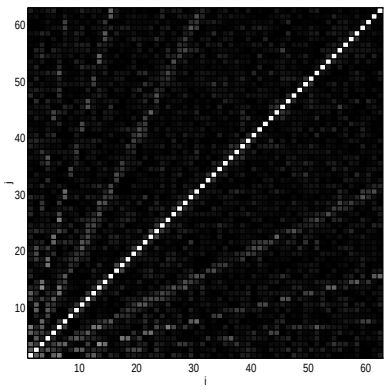

Figure 4: Correlation Matrix of "FARIMA(0,0.4,0)".

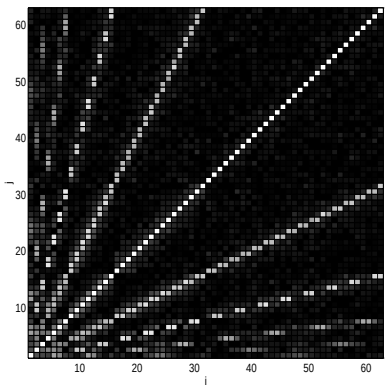

Figure 5: Correlation Matrix of "AR(1)".

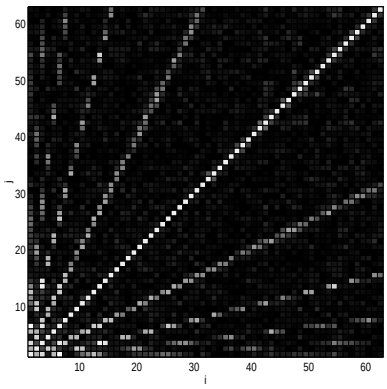

Figure 6: Correlation Matrix of "Star Wars".

j exponentially for all j. The variance of SRD increases at an even faster rate than that of LRD when j is small but saturates when j is large. For a mixture of LRD and SRD, the variance shows the mixed properties from both SRD and LRD. These results indicate that the variance of wavelets are capable of distinguishing LRD from SRD for Gaussian processes.

The second indication of the theorems is that the wavelet transform significantly reduce the temporal dependence so that a complicated mixture of short- and long-range dependence in the time domain may be sufficiently modeled by a "short-range" dependent process in the wavelet domain.

4.2 Empirical Studies on The Correlation Structure of LRD and SRD

What short-range dependence need to be captured among wavelet coefficients? Unfortunately, an answer to this question can not be provided by Theorems 1 and 2, since they only hold for $|(m_1 - 1)2^{j_1} - m_2 2^{j_2}|$ large. Therefore, we plot the correlation matrices of $FARIMA(0, 0.4, 0)$, $AR(1)$ and a JEPG encoded video source "Star Wars" (details will be given latter in this paper) in Figures 4, 5 and 6 respectively to visualize the correlation structure. A pixel (i, j) in an image represents the correlation between the i-th and the j-th wavelet coefficients, where i and j are the (node) one dimension index in the tree diagram (see Figure 1 details). 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 besides the diagonal line[11], there are 4 pairs of lines having "visible" correlations[12]. They correspond to the correlation between $\gamma^k(s)$ and s, where $\gamma(s)$ represents the parent of the node s (see Figure 1 for explanations on $\gamma(s)$), and $\gamma^k(s)$ denotes the parent of the node $\gamma^{k-1}(s)$ with k being $1, 2, 3$ and 4 counting from the diagonal line. From the figures, we can conclude that the most significant correlation is due to the parent-child relationship. Since the complicated correlation in the time domain actually concentrates on certain types of correlations in the wavelet domain, we can use a parsimonious model in the wavelet domain to represent the original traffic.

[11] In order to have enough gray level to see more subtle details, the diagonal pixels, which is always 1, is set to 0.5.

[12] We only consider $K = 5$ which has only 5 level in the tree diagram.

4.3 Modeling The Correlation Structure

We propose several models to model partial correlations among wavelet coefficients, ranging from the simplest to the most complex. The simplest model assumes that wavelet coefficients are statistically independent. The most complicated model is a third order Markov model.

Model 1: Independent model, which models the variances of wavelet coefficients[13] . d_j^m's are modeled as independent Gaussian variables with a zero mean and a variance σ_j.

Model 2: The first-order Markov model, which models the correlation between $\gamma(s)$ and s, i.e., the parent-child relationship. Such a Markov model can be implemented through an $AR(1)$ process as

$$d_{s_j} = a_j d_{\gamma(s_j)} + b_j w_j, \tag{7}$$

where a_j and b_j are the parameters to be determined from data. w_j is Gaussian noise with zero mean and unit variance. This model captures the diagonal lines as well as the two whitest off-diagonal lines in the (correlation) graph.

Model 3: The third order Markov model, which models the correlation among $\gamma(s)$, $\gamma^2(s)$, $\nu(s)$ and s. This model incorporates additional correlation among node s and its neighboring nodes. In the (correlation matrix) graph, it is equivalent to matching the first two strongest lines as well as the (barely visible) line near the diagonal, which represents the neighboring relationship. Please refer to [25] for details on the algorithm.

As for the complexity of the aforementioned models, Model 1 only needs one parameter (the variance) σ_j at each level j and thus requires $log_2(N)$ parameters in total. Models 2 and 3 have two and four parameters at each level, respectively.

4.4 An Algorithm for Generating Wavelet Models for Gaussian Processes

The models of the correlation structure can now be included in an algorithm to obtain wavelet models for a Gaussian process.

Let $\hat{x}(t)$ be a trace of length N from a Gaussian process.

Algorithm 1

1. Perform wavelet transform on $\hat{x}(t)$ to obtain $\hat{d}_j^m$'s (wavelet coefficients of $\hat{x}(t)$).
2. Estimate the required parameters in a selected wavelet correlation model (Section 4.3) from $\hat{d}_j^m$'s.
3. Generate coefficients d_j^m from the wavelet correlation model for all m and j.
4. Do inverse wavelet transform on d_j^m's to get the synthesized traffic in the time domain.

To estimate the computational complexity of the algorithm, we notice that the computational complexity of the wavelet transform (Step 1) and the inverse transform (Step 4) is $O(N)$, respectively. The computational complexity of Steps 2 and 3 is also $O(N)$[14]. Then the total computational cost of the algorithm is $O(N)$.

[13] The mean of d_j^m is zero for a stationary process.

[14] Assume that $O(1)$ time is needed to generate one Gaussian random variable.

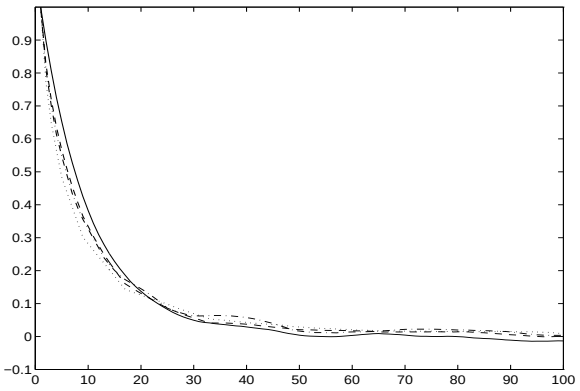

Figure 7: Sample Correlations. "-": AR(0.9);"–": Model 3; "-.": Model 2; "..": Model 1.

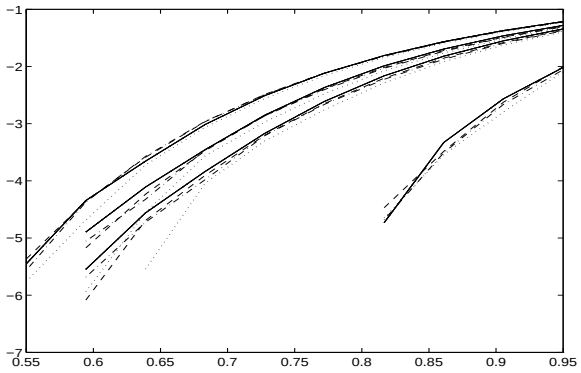

Figure 8: Buffer Response. x-axis: utilization. y-axis: log10(Loss probability). "-": AR(0.9);"–": Model 3; "-.": Model 2; "..": Model 1. The normalized buffer size ($\frac{B}{C}$ is 0.1, 0.5, 1, 10 from top down.

	R/S (FARIMA)	$V-T$ (FARIMA)	R/S (Algorithm 1)	$V-T$ (Algorithm 1)
FARIMA(0,0.2,0)	0.711	0.703	0.711 ± 0.009	0.716 ± 0.018
FARIMA(0,0.3,0)	0.801	0.776	0.795 ± 0.018	0.755 ± 0.017
FARIMA(0,0.4,0)	0.840	0.840	0.852 ± 0.014	0.830 ± 0.033

Table 1: Hurst parameters of the FARIMA models and the wavelet models.

4.5 Experimental Results

In this section, experimental results will be provided to evaluate (1) the performance of three wavelet models discussed in Section 4.3, and (2) the capability of wavelet models to capture long-range dependence. For each experiment, five sample paths will be generated from either AR(1) or FARIMA, one of them will be used to obtain a wavelet model through Algorithm 1, while the other four are used to obtain averaged loss rate and sample auto-correlation for testing the performance of wavelet models.

To compare the performance of three wavelet models proposed in Section 4.3, sample paths of length 2^{18} are generated from an AR(1) process with the parameter to be 0.9. We plot the sample correlation and buffer response in Figures 7 and 8, respectively, for all three wavelet correlation models. We observe from the figure that the simplest wavelet model which neglects the dependence in the wavelet domain performs reasonably well. The models which capture more correlations among wavelets only improve the performance slightly.

To further test the performance of the independent wavelet model (Model 1) on capturing long-range dependence, sample paths with length 2^{17} are generated from FARIMA$(0, d, 0)$ (LRD alone) for $d = 0.2, 0.3, 0.4$, respectively, using a simulator in SPLUS.

Three quantities are used to test the performance: the estimated Hurst parameter, the (sample) auto-correlation function in the time-domain and the estimated buffer loss probability estimated from the sample paths.

Table 1 gives the results on the Hurst parameter using the R/S test and the Variance-Time (V-T) test

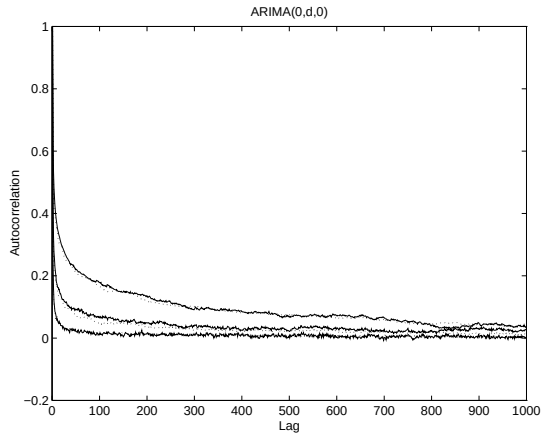

Figure 9: Autocorrelation functions for FARIMA(0,d,0) (solid lines) and Algorithm 1 (dotted lines). d=0.2,0.3 and 0.4 from the bottom up.

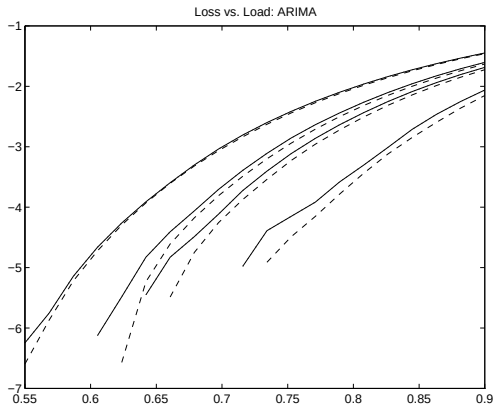

Figure 10: The vertical axis: $\log_{10}$(Loss Rate). The horizontal axis: work load. The solid lines: FARIMA(0,0.4,0). Dotted lines: Algorithm 1. The normalized buffer size ($\frac{B}{C}$): 0.1, 0.5,1, and 10 from the top down.

[15]. Each quantity obtained from Algorithm 1 is averaged over 5 different runs. The standard deviations are also given in the table. As can be seen, the Hurst parameters from the wavelet models are very close to those from the original FARIMA processes.

The correlation functions estimated from the sample paths are plotted in Figure 9. The loss rate is estimated from the sample path from the original FARIMA process and the wavelet model, we plot the log of loss rate versus the work load for the normalized buffer size $0.1, 0.5, 1$, and 5, respectively. The results show that the correlation function and the loss rate due to the wavelet model are very close to those due to the actual FARIMA process.

4.6 Analysis on The Buffer Overflow Probability of Independent Wavelet Model

Although empirical results show that the independent wavelet coefficients provide a reasonably good model for a long-range dependent process, valid buffer loss rates below 10^{-5} are hard to obtain due to the lengthy computational time needed in generating a long sequence for FARIMA by SPLUS. Therefore, it is important to further confirm our observation through analyzing the buffer overflow probability. We will analyze the buffer overflow probability of the independent wavelet model for a FGN process and compare that resulting from the original FGN process. Our result can be summarized as follows.

Theorem 3 *Let B_t and $\hat{B}_t$ be the buffer sizes at the t-th time slot due to the synthesized traffic obtained by the independent wavelet model, and by the FGN process, respectively. Let C represent the capacity of a single infinity buffer queue. The buffer overflow probability due to the synthesized (wavelet) traffic $x(t)$ satisfies*

$$\begin{aligned} Pr(B_N > B) &\sim Pr(\hat{B}_N > B), \\ &\sim \exp(-\frac{(C-\mu)^2(\frac{B}{C-\mu})^{2(1-H)}(\frac{1-H}{H})^{2H}}{2\sigma^2(1-H)^2}), \end{aligned} \tag{8}$$

[15] See [5] [21][15] for details on R/S test and V-T test.

where "$\sim$" represents asymptotically and logarithmically equal. $N = 2^K$ is the length of the traffic with K being a positive integer. B is assumed to be

$$B = \frac{(c-\mu)2^{k_0}(1-H)}{H}, \tag{9}$$

where k_0 is an positive integer for simplicity.

The proof of the theorem can be found in Appendix B. We would like to point out that this theoretical result is limited by two conditions: (a) the buffer size B_t is only considered at time $t = N$ instead of all t, and (b) the buffer full size B is assumed to be $2^{k_0}(C-\mu)$, which is only a subset of all possible values. In [25], we relax the condition (a) and prove that the conclusion of the theorem 3 holds for the average buffer overflow probability. Due to the complexity of the proof, we only give the weaker version of our theorems in [25].

This result shows that the simple wavelet model which ignores the correlations among different wavelet coefficients has the same Weibull decaying buffer overflow probability as that of FGN. This confirms that the independent wavelet model is accurate enough for modeling Fractional Gaussian Noise processes, and models which capture additional correlations only improve the buffer loss rate marginally. In what follows, we will use the independent wavelet model to model real network traffic.

5 Wavelet Modeling of Network Traffic

5.1 The Traffic Sources

Four network traffic sources are chosen to test our wavelet models: (1) JPEG coded "Star Wars" [15], (2) multiplexed "Star Wars", (3) MPEG-I coded videos, and (4) Ethernet data trace with 10 ms time interval.

The trace "Star Wars" [15] is obtained by applying JPEG-like encoder to each of 171,000 frames at an interval of $\frac{1}{24}$ second per frame of the 2-hour movie of "Star Wars" [16]. This source is used to test our model, since the source provides rich variations in terms of scene changes.

Since modeling multiplexed video traffic is of practical importance, we combine M copies of the VBR trace of "Star Wars" to obtain a multiplexed source as suggested in [15], where M is chosen to be 100. Each copy is offset by a random number of frames. Upon reaching the end of the trace, each copy is wrapped around to the beginning. Mathematically, since the multiplexed source is a summation of individual sources, and wavelet transform is a linear operation, multiplexing video sources in the time domain is equivalent to multiplexing the corresponding wavelet coefficients of each source in the wavelet domain. Therefore, the marginal distribution of wavelet coefficients of the multiplexed traffic is very close to Gaussian.

The third data set is the MPEG coded video source constructed from MPEG-I encoded video sequences created[35]. We choose a video called "Jurassic Park", which has 53332 frames at an interval of $40ms$. A standard MPEG encoder generates three types of compressed frames: I, P and B. I frames are compressed using intra-frame coding only, while P and B frame, in addition to intra-frame coding, allow using motion compensation techniques. As a result, I frames are the largest in size, followed by P frames and B frames. In addition, many MPEG encoders have fixed Group-of-Picture (GOP) pattern. The GOP pattern in "Jurassic Park" consists of 12 frames as $IBBPBBPBBPBB$.

[16] Detailed description on the "Star Wars" trace can be found in [15].

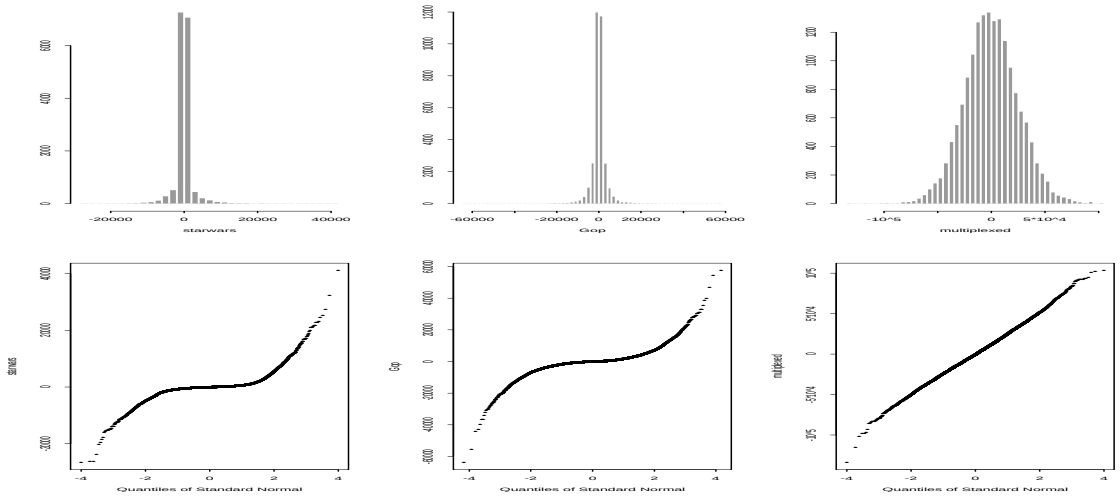

Figure 11: Marginal distribution of d_j^m for $j = 3$. Left: Star Wars. Middle: "MPEG GOP". Right: Multiplexed (M=100). Up: Histogram. Down: Quantile-Quantile plots (horizontal axis: Normal distribution.)

The fourth data set is an Ethernet traffic trace. The Ethernet trace obtained from [22][17] is one of benchmark data sets for traffic modeling. This data set is collected on an Ethernet in Bellcore from 11:25 August 29, 1989, and details can be found in [22]. We choose the time unit for the discrete time queuing model to be $10ms$, which leads to discrete traffic with length 176000[18].

5.2 Modeling Network Traffic

5.2.1 Marginal Probability Density Functions at Different Time-Scales

As Algorithm 1 generates wavelet models for Gaussian processes, a question to ask is whether wavelet coefficients from the real network traffic have Gaussian marginals.

The histograms and Q-Q plots[19] are given in Figure 11 for the wavelet coefficients of three video sources at $j = 3$[20]. The figure shows that the estimated marginal density functions of wavelet coefficients for individual video sources at the given time-scale have a much heavier tail than the normal distribution, whereas the density function for the multiplexed source fits the normal distribution well. This is due to the fact that marginal density function of wavelet coefficients for a single source is usually dominated by the nature of the source itself, and can be different from Gaussian in general. The marginal of wavelet coefficients of the multiplexed source, which are sums i.i.d. random variables, is approximately normal when the number of

[17] This Ethernet trace data can be obtained from http:
ita.ee.lbl.govhtmlcontribBC.html. In this database, we only report results on the data set collected in August 1989.

[18] Although a multiplexer of data network works in continuous time, it is convenience to use the discrete time queuing model for simulation[32][14].

[19] The $Q - Q$ plot is a standard statistical tool to measure the deviation of a marginal density function from a normal density. The $Q - Q$ plots of a process with a normal marginal is a straight line. The deviation from the line indicates the deviation from the normal density. See [15] and references herein for more details.

[20] at GOP level for MPEG-I. Similar behaviors have been observed at the other time-scales. Non-Gaussian marginals are observed for Ethernet data trace but omitted here.

multiplexed sources is sufficiently large.

5.2.2 Shaping The Marginal Density Functions

To obtain wavelet coefficients d_j^m's with the same empirical density functions as that of the network traffic, a transformation is needed on the wavelet coefficients generated by Algorithm 1 with a Gaussian marginal density function. Such a transformation can be done easily through a method described in [18] with little computation. To illustrate the idea, let X be a random variable with a cumulative distribution function $F_X(x)$. Let $F_Y(\cdot)$ be the desired cumulative distribution function. Then the transformation needs to be performed on X is simply

$$y = F_Y^{-1}(F_X(x)). \tag{10}$$

It can be easily verified that the resulting random variable Y has the desired cumulative distribution function $F_Y(y)$. In the modeling process, $F_X(x)$ is the cumulative distribution function for an d_j^m generated by a Gaussian distribution. $F_Y(y)$ is the desired cumulative distribution function, which can be estimated using the histograms from the data. Then the transformation can be easily performed to obtain the wavelet coefficients with a desired marginal density function. Interested readers may see [18] for more details.

5.2.3 An Algorithm for Generating Wavelet Models for Network Traffic

Adding an additional step to shape the marginal, we have Algorithm 2 for the independent wavelet model (Model 1 in Section 4.3) as follows. Let $\hat{x}(t)$ be a traffic trace of length N.

Algorithm 2

1. Perform wavelet transform on $\hat{x}(t)$ to obtain $\hat{d}_j^m$'s (wavelet coefficients of $\hat{x}(t)$).
2. Compute variance $\hat{\sigma}_j$ at the j-th time-scale.
3. Generate wavelet coefficients d_j^m through a Gaussian random variable with zero mean and variance $\hat{\sigma}_j$.
4. Perform the transformation on d_j^m according to Section 5.2.1.
5. Do the inverse wavelet transform on the generated wavelet coefficients to get the synthesized traffic.

It can be seen that the total parameters needed for this algorithm is $\log_2(N)$, where N is the length of synthesis traffic, and the computational time is in the order of $O(N)$.

5.3 Modeling MPEG-coded Video Traffic

To deal with the periodic structure of MPEG-coded video (see Figure 12 for illustration), Algorithm 2 can be further modified for MPEG traffic by taking the advantage of wavelet modeling applicable at different time-scales. To do so, we note that as the interframe redundancy is reduced by using P and B frames, I, P, B frames have significantly different statistical characteristics. Therefore, I, P, and B frames should be distinguished from one another at the frame level for modeling, whereas no such distinction is needed from the GOP level. Therefore, we can apply Algorithm 2 directly to model the MPEG video traffic above the GOP

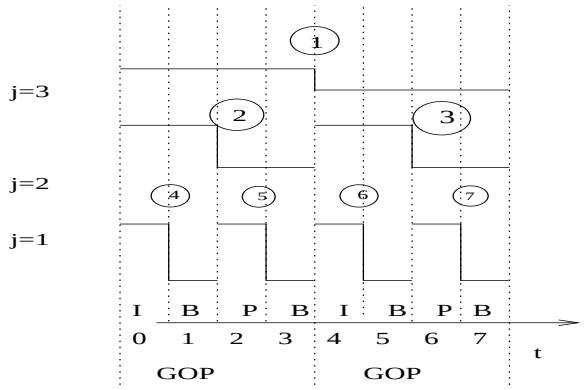

Figure 12: MPEG traffic with GOP pattern IBPB and its relationship with wavelet basis function. I: Intra-coded frame. P: predictive frame. B: bidirectional frame.

	R/S (Video trace)	$V-T$ (Video trace)	R/S (Alg. 2)	$V-T$ (Alg. 2)
Multiplexed (M=100)	0.869	0.824	0.842 ± 0.0267	0.839 ± 0.012
(Single) Star Wars	0.815	0.774	0.811 ± 0.009	0.779 ± 0.017

Table 2: Hurst parameters

level. Below the GOP level[21], we treat wavelet coefficients with the same relative position in a GOP pattern as a group with the same statistical properties. For example, the variances and the correlation (graph) of the parent-child relationship are the same for every wavelet coefficient in the same group. We can then model different groups of wavelet coefficients with different parameters.

5.3.1 Experimental Results

To evaluate both the performance and the computational efficiency of the wavelet models (Algorithm 2), we apply the algorithm to obtain wavelet models for all four sources, and to generate synthesized traffic. FARIMA models are used to model the sources as well for comparison. In particular, FARIMA models are chosen to have 25 AR terms and 20 MA terms. A maximum-likelihood algorithm provided by SPLUS [39] is used to determine all 45 parameters for the FARIMA models using the trace. Furthermore, the Gaussian marginal distribution of the FARIMA process is shaped to be the same as that of trace data by the same method described in Section 5.2.2. Synthesized traffic from both wavelet and FARIMA models are used to obtain sample auto-correlation functions and the buffer loss rate. Since FARIMA model can not deal with (periodic) MPEG source easily, the results are absent in corresponding figures.

The auto-correlation functions of FARIMA$(25, d, 20)$ and of the wavelet model are plotted in Figures 13, 15, 17 and 19 for the single "Star Wars" source, multiplexed "Star Wars" source and the "Ethernet" trace, respectively. As can been seen, the wavelet models have consistently a better match to the auto-correlation functions than the FARIMA models, especially for single sources. The reason is possibly due to that the wavelet model can match the marginal distribution at different time scales. Hurst parameters for "Star Wars"

[21] We assume the GOP pattern is periodic as in most cases. If the length of GOP pattern is not a power of 2, we will add zero size frames to the end of each GOP to make it a power of 2.

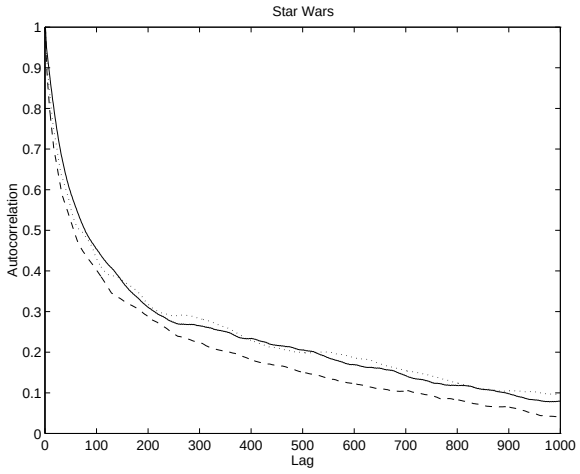

Figure 13: "-": Autocorrelation of "Star Wars"; "- -": FARIMA(25,d,20); "..": Algorithm 2

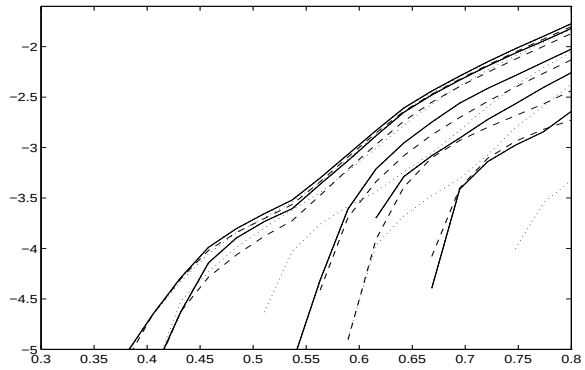

Figure 14: Vertical axis: $\log_{10}$(Loss Rate); horizontal axis: work load. "-": the single video source; "..": FARIMA(25,d,20); "- -" Algorithm 2. The normalized buffer size ($\frac{B}{C}$): 0.1, 1, 10,30 and 100 from the top down.

and "Multiplexed" traffic are also reported in Table 2 for comparison. The results of the loss rate of a single buffer due to each source are given in Figures 14, 16, 18 and 20, respectively. The wavelet model has a significantly better performance at medium and large buffer sizes[22].

To compare the complexity of the wavelet and FARIMA models, we note that FARIMA has 46 parameters but the independent wavelet model has only $15 \sim 17$ parameters. As for the computational time, it takes more than 5-hour CPU time on a SunSPARC 5 workstation for an FARIMA$(25, d, 20)$ model to estimate its parameters from the data and to generate synthesized video traffic of length $171,000$. It only takes 3 minutes on the same machine for Algorithm 2 to complete the same task. The computational complexity to generate synthesized traffic of length N is $O(N^2)$ for an FARIMA model, and only $O(N)$ for a wavelet model.

6 Discussions

6.1 Discussions on Wavelet-Domain Modeling

Why do wavelet models work so well? As conjectured in [32] that wavelets might be feasible to model self-similar data traffic, the complicated mixture of long-range and short-range dependence in traffic can be naturally embedded in the self-similar structure of wavelets. Since the self-similar structure of wavelets has "absorbed" the long-range and short-range dependence in different time-scales, the wavelet coefficients correspond to a "short-range dependent" process in the wavelet domain. This results in fast decaying correlation in the wavelet domain and makes it possible to to model wavelet coefficients as independent or Markov-dependent random variables. The resulting algorithms for generating wavelet coefficients are therefore simple. In addition, since there exist fast algorithms for both transforms and inverse transforms[7], our method for

[22] At a very small buffer size (an extreme case is a bufferless queue), the buffer loss rate is dominated by the marginal distribution. Because both algorithms are shaped to have the same marginal distribution as that of traces, both algorithms give very good buffer loss estimation.

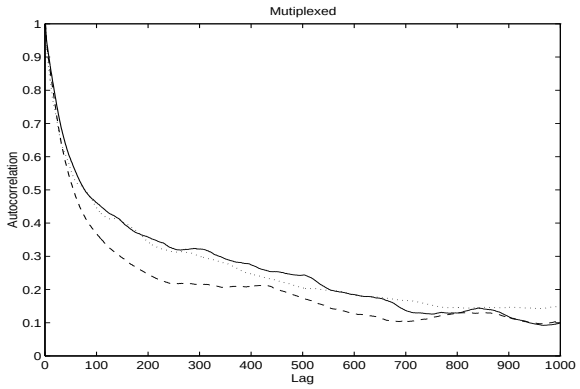

Figure 15: "-": Autocorrelation of Multiplexed Source (M=100); "- -": FARIMA(25,d,20); "..": Algorithm 2

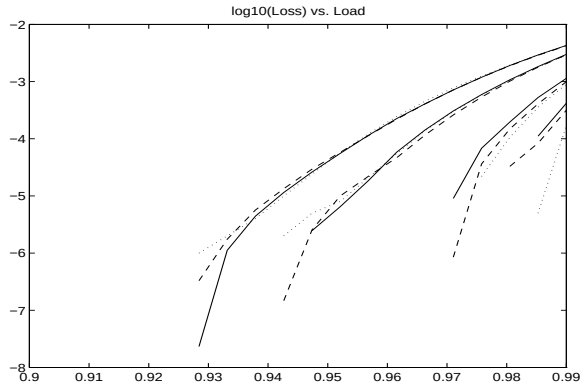

Figure 16: Vertical axis: $\log_{10}$(Loss Rate); horizontal axis: work load. "-": the multiplexed source; "..": FARIMA(25,d,20); "- -" Algorithm 2. The normalized buffer size: 0.1, 1, 10,30 from the top down.

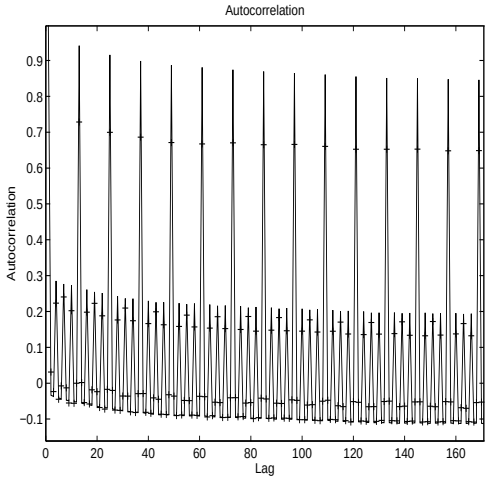

Figure 17: Autocorrelation. "-": "Jurassic Park";"+": Wavelet Model

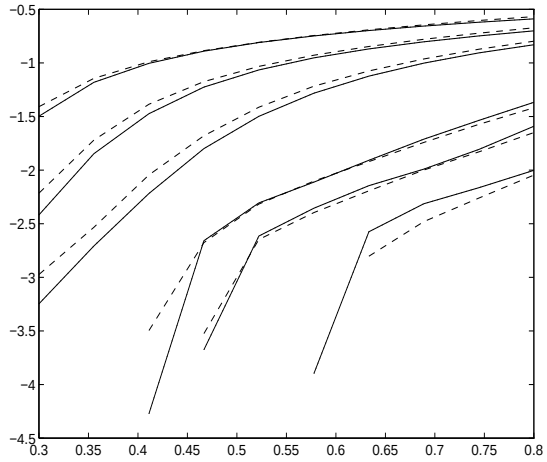

Figure 18: Buffer Response. x-axis: utilization. y-axis: log10(Loss probability). "-": "Jurassic Park";"–": Wavelet Model. The normalized buffer size: 0.1,0.5,1,10,30,100 from the top down.

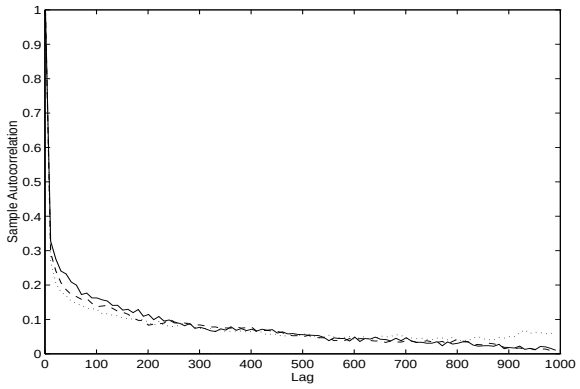

Figure 19: Sample Autocorrelation. x-axis: Lag. y-axis: Sample Autocorrelation. "-": "Data Trace";"- -": Wavelet Model; "..": FARIMA model.

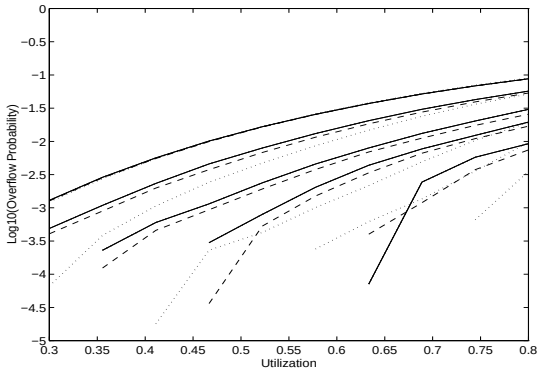

Figure 20: Buffer Response. "-": Data Trace;"- -": Wavelet Model; "..": FARIMA. Normalized buffer size B/C: 0.1, 1, 10, 30 and 100 from the top down.

developing the models and generating synthesized traffic has been able to achieve the lowest complexity, which is $O(N)$ for a time-series of length N.

Another appealing property of wavelets is that wavelet basis are defined by a scaled and shifted mother wavelet with a finite support, and thereby can be naturally linked to the time-scale analysis. This makes it easy for wavelet models to handle the periodic structure in MPEG video.

Furthermore, since (Harr) wavelets are simple, it is possible to use the model for theoretical analysis as shown in Section 4.6.

6.2 Comparison Among Different Traffic Models

To further illustrate the strength (and possible weakness) of wavelet models, we summarize in Table 3 strength/weakness of different models for network traffic. In the first column, "SRD/LRD" represents a model's capability to capture both the short-range and the long-range dependence. In the second column, we characterize the complexity of a modeling algorithm from two perspectives: the computational complexity on generating synthesized traffic of length N, and the implementation complexity, which refers to the amount of efforts involved in developing an accurate model which may involve tuning different parameters. In the table, we also include the feasibility of a model for theoretical analysis ("Analysis"), and the flexibility of the method in terms of modeling at different time scales.

7 Conclusions

An important discovery from this work is that wavelet coefficients of network traffic with complicated short- and long-range temporal dependence are no longer long-range dependent. Therefore, the "short-range" dependent process can be used to model traffic in the wavelet domain. This opens up new possibilities for modeling, analyzing and controlling long-range dependent network traffic.

In this work, we have developed wavelet models for network traffic including both video and data traffic.

	SRD/LRD	Computational Complexity /Implementation Complexity	Analysis	Time-Scale Modeling
TES[30]	Yes	O(N)/High	Difficult	No
DAR [17]	Only SRD	O(N)/Low	Yes	No
Scene-Based [19]	Yes	O(N)/High	Yes	Yes
Importance Sampling [18]	Yes	$O(N^2)$/Low	No	No
Wavelet Modeling	Yes	O(N)/Low	Yes	Yes
FARIMA [21]	Yes	$O(N^2)$/High	No	No

Table 3: Comparison on traffic models.

A good performance of the model has been obtained through extensive tests on three video sources and one Ethernet data trace using both the auto-correlation and the buffer loss rate as performance measures. The buffer loss rate of the wavelet model has shown analytically to be similar to that of the FGN process for a FGN workload, and thereby demonstrates the capability and the performance of the independent wavelet models. Since the self-similar structure of wavelets naturally matches the statistical self-similarity in network traffic, the resulting wavelet models are parsimonious and have much fewer parameters than FARIMA models. The computational complexity for developing such a wavelet model and for synthesizing a large volume of traffic has shown to be $O(N)$, which is the lowest attained.

In our future work, we will further investigate the capability of independent wavelet models, and extend the analysis to non-Gaussian traffic in order to make up for insufficient the network data at low loss rate. We will also investigate other performance measures to account for effects of long-range dependence on loss patterns.

Acknowledgment

We would like to thank Olive Rose and Mark Garrett for making the video traces available, and would like to thank Water Willinger, Bo Ryu, Marina Thottan and Xusheng Tian for helpful comments.

The support from National Science Foundation ((CAREER) IRI-9502518) and DARPA (DARPA BAA-9704) is gratefully acknowledged.

References

[1] Allan T. Andersen and Bo Friis Nielsen. An application of superpositions of two state Markovian sources to the modeling of self-similar behavior. In *Proceedings of INFOCOM*, pages 196–204, Kobe, Japan, 1997.

[2] Michele Basseville, Albert Benveniste, Kenneth C. Chou, Stuart A. Golden, Ramine Nikoukhah, and Alan S. Willsky. Modeling and estimation of multiresolution stochastic processes. *IEEE Transactions on Information Theory*, 38(2):766–784, 1992.

[3] Michele Basseville, Albert Benveniste, and Alan S. Willsky. Multiscale autoregressive processes. *IEEE Transactions on Signal Processing*, 40:1915–1954, 1992.

[4] Jan Beran. *Statistics for Long-Memory Processes*. Chapman & Hall, 1994.

[5] Jan Beran, Robert Sherman, Murad S. Taqqu, and Walter Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications*, 43:1565–1579, 1995.

[6] D.R. Cox. "Long-Range Dependence : A Review". In H.D.David and H.T.David, editors, *Statistics : An Appraisal*, pages 55–74. The Iowa State University Press, 1984.

[7] I. Daubechies. *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.

[8] R.W. Dijkerman and R.R. Mazumdar. Wavelet representations of stochastic processes and multiresolution stochastic models. *IEEE Transactions on Signal Processing*, 42(7):1640–1652, 1994.

[9] N.G. Duffield and N. O'Connell. "Large deviations and overflow probabilities for the general single server queue, with applications". DIAS-STP-93-30, Dublin Institute for Advanced Studies, 1993.

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

[11] A. Erramilli, R.P. Singh, and P. Pruthi. An application of chaotic maps to packet traffic modeling. *Queueing System*, 20:171–206, 1995.

[12] A. Erramilli, R.P. Singh, and P. Pruthi. An application of deterministic chaotic maps to model packet traffic. *Queueing System*, 20:171–206, 1995.

[13] Patrick Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, 38(2):910–917, 1992.

[14] V. Frost and B. Melamed. Traffic modeling for telecommunications networks. *IEEE Communication Magazines*, 32:70–80, 1994.

[15] Mark W. Garrett and Walter Willinger. Analysis, modeling and generation of self-similar VBR video traffic. In *Proceedings of SIGCOM*, pages 269–279, London, UK, 1994.

[16] Marc Goldburg. *Applications of Wavelets to Quantization and Random Process Representations*. PhD thesis, Standford University, 1993.

[17] D.P. Heyman and T.V. Lakshman. What are the implications of long-rang dependence for VBR-video traffic engineering? *IEEE/ACM Transactions on Networking*, 4(3):301–317, 1996.

[18] 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 *Proceedings of SIGCOM*, pages 114–125, Stanford, CA, 1996.

[19] Predrag R. Jelenkovic, Aurel A. Lazar, and Nemo Semret. The effect of multiple time scales and subexponentiality in MPEG video streams on queueing behavior. *IEEE Journal on Selected Area in Communications*, 15(6):1052–1071, 1997.

[20] L.M. Kaplan and C.J. Kuo. Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis. *IEEE Transactions on Information Theory*, 41(12):3554–3562, 1993.

[21] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of Ethernet traffic. *IEEE/ACM Transactions on Networking*, 2(1):1–15, February 1994. Available through ftp://ftp.bellcore.com/pub/wel/tome.ps.Z.

[22] W.E. Leland and D.V. Wilson. High time resolution measurements and analysis of LAN traffic: Implications for LAN interconnection. In *Proceedings of IEEE INFOCOM*, pages 1360–1366, Bal Harbour, FL, 1991.

[23] S.Qi Li and C-L Hwang. Queue response to input correlation functions: Discrete spectral analysis. *IEEE/ACM Transactions on Networking*, 1:317–329, 1993.

[24] Mark R. Luettgen, William C. Karl, and Alan S Willsky. Multiscale representations of markov random fields. *IEEE Transactions on Signal Processing*, 41:3377–3396, 1993.

[25] S. Ma. *Traffic modeling and analysis*. PhD thesis, Department of Electrical, Computer and systems Engineering, Rensselaer Polytechnic Institute, 1998.

[26] S. Ma and C. Ji. Modeling video traffic in the wavelet domain. *IEEE Communications Letters*, pages 100–103, 1998.

[27] S. Ma and C. Ji. Modeling video traffic in the wavelet domain. In *Proceedings of INFOCOM*, San Francisco, CA, April 1998.

[28] S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligent*, 11:674–693, 1989.

[29] E. Masry. The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. *IEEE Transactions on Information Theory*, 39:260–264, 1993.

[30] B. Melamed, D. Raychaudhuri, B. Sengupta, and J. Zdepski. TES-based video source modeling for performance evaluation of integrated networks. *IEEE Transactions on Communications*, 42:2773–2783, 1994.

[31] I. Norros. "A storage model with self-similar input". *Queueing System*, 16:387–396, 1994.

[32] V. Paxson. Fast approximation of self-similar network traffic. Available through ftp://ftp.ee.lbl.gov/papers/fast-approx-selfsim.ps. Z, 1995.

[33] D. Reininger, D. Raychaudhuri, B. Melamed, B. Sengupta, and J. Hill. Statistical multiplexing of VBR MPEG compressed video on ATM networks. In *Proceedings of INFOCOM*, pages 919–926, San Francisco, CA, 1993.

[34] S. Robert and J.Y. Le Boudec. Can self-similar traffic be modeled by Markovian processes. Available through http://lrcwww.epfl.ch/PS_files/robert_stockholm.p s.Z, COST242 Technical Document, TD95-26, presented at the Stockholm Management Committee Meeting, 1995.

[35] O. Rose. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM traffic engineering. Technical Report 101, University of Wurzburg, 1995.

[36] B.K. Ryu. *Fractal Network Traffic: From Understanding to Implications.* PhD thesis, Columbia University, 1996.

[37] B.K. Ryu and A. Elwalid. The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities. In *Proceedings of SIGCOM*, pages 3–14, Stanford, CA, 1996.

[38] P. Skelly, M. Schwartz, and S. Dixit. A histogram-based model for video traffic behavior in an ATM multiplexer. *IEEE/ACM Transactions on Networking*, 1(4):446–459, 1993.

[39] Statistical Sciences, Seattle: StatSci, a division of MathSoft, Inc. *S-PLUS Guide to Statistical and Mathematical Analysis, Version 3.3*, 1995.

[40] A.H. Tewfik and M. Kim. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Transactions on Information Theory*, 38:904–909, 1992.

[41] D. Tse, R.G. Gallager, and J.N. Tsitsiklis. Statistical multiplexing of multiple time-scale Markov streams. *IEEE Journal on Selected Area in Communications*, 43:1566–1579, 1995.

[42] G.W. Wornell and A.V. Oppenheim. Wavelet-based representations for a class of self-similar signals with application to fractal modulation. *IEEE Transactions on Information Theory*, 38:785–800, 1992.

[43] F. Yegenoglu, B. Jabbari, and Ya-Qin Zhang. Motion classified autoregressive modeling of variable bit rate video. *IEEE Transactions on Circuits and Systems for Video Technology*, 3(2):42–53, 1993.

Appendix A: Proof of Theorem 2

Proof: Since the proof of 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

$$Var(d_j^0) = 2^{-j}\mathbf{E}(\sum_{t=0}^{2^{j-1}-1} x(t) - \sum_{t=2^{j-1}}^{2^j-1} x(t))^2 \quad (11)$$

$$= 2^{-j}\mathbf{E}[(\sum_{t=0}^{2^{j-1}-1} (x(t)-\mu))^2 + (\sum_{t=2^{j-1}}^{2^j-1} (x(t)-\mu))^2 - \sum_{t_1=0}^{2^{j-1}-1}\sum_{t_2=2^{j-1}}^{2^j-1} 2(x(t_1)-\mu)(x(t_2)-\mu)], \quad (12)$$

Through straight forward algebraic manipulations (see [25] for details), we can compute the two term of Equation (12), and thereby prove the theorem.

Appendix B: Proof of Theorem 3.

To prove Theorem 3, we will first define the averaged aggregate-process, and establish its wavelet representation with independent wavelet coefficients generated by Algorithm 1. We will then develop three lemmas needed for proving the main theorem. Using the lemmas, we will obtain an upper and a lower bound for the buffer loss rate, and then show that bounds are of the same order.

Definitions

An Averaged Aggregate-Process

The averaged aggregate-process $X(s)$ of the synthesized process $x(t)$ generated by a wavelet model at time t_0 is defined as

$$X(s) = \frac{1}{s}\sum_{i=1}^{s} x(t_0 - i), \tag{13}$$

for $1 \leq s \leq t_0$. $X(s)$ represents the average value of $x(t)$ over $[t_0 - s, t_0 - 1]$. To facilitate the analysis, we consider a special case that $t_0 = N$.

A Wavelet Representation of The Averaged Aggregate-Process

Replacing $x(t)$ by its wavelet representation given in Equation (2), the wavelet representation of $X(s)$ can be easily obtained as

$$X(s) = \frac{1}{s}\sum_{j=1}^{K} a_j(s) w_j d_j, \tag{14}$$

for $1 \leq j \leq K$, where $w_j = 2^{-0.5j}$ is a weighting factor. d_j is a wavelet coefficient at the j-th time-scale obtained by Algorithm 1 for independent wavelet coefficients[23]. When used to model the FGN process with a zero mean, the variance of d_j is given in [42][13] as

$$Var(d_j) = 2^{-j(1-2H)}(2^{2(1-H)} - 1)\sigma^2(1 - O(2^{-2(1-H)K})). \tag{15}$$

$a_j(s)$ is a weighting factor which depends on the aggregation length s, where

$$a_j(s) = \begin{cases} r_j^s, & \text{for either } r_j^s = r_{j-1}^s \text{ or } j = 1, \\ 2^{j-1} - r_{j-1}^s, & \text{otherwise,} \end{cases} \tag{16}$$

with $r_j^s = s - \lfloor s2^{-j} \rfloor * 2^j$, and $\lfloor \cdot \rfloor$ represents the largest integer smaller than s.

Examples of such $a_j(s)$ are given in Table 3. As can be seen from the table, $a_j(s)$ is periodic with a period 2^j. Furthermore, some of the properties of $a_j(s)$ can be derived directly from the Equation (16) as given below.

Properties of $a_j(s)$

[23] Strictly speaking, $d_j = d_j^{m_j}$, where m_j is an index on m which depends on j. However, since there is only one such $d_j^{m_j}$ at a particular time scale j, we drop the superscript m_j for simplicity.

j s	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
2	1	2	1	0	1	2	1	0	1	2	1	0	1	2	1	0
3	1	2	3	4	3	2	1	0	1	2	3	4	3	2	1	0
4	1	2	3	4	5	6	7	8	7	6	5	4	3	2	1	0
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Table 4: The values of $a_j(s)$. s increases from 1 to 16 horizontally and j increases from 1 to 7 vertically.

1. For any positive integer satisfying $s = 2^{\lfloor \log_2(s) \rfloor}$, $a_j(s)$ can be represented as

$$a_j(s) = \begin{cases} s, & \text{if } j \geq log_2(s) + 1 \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

2. For any positive integer $s > 2^{\lfloor \log_2(s) \rfloor}$, we define $s' = 2^{\lfloor \log_2(s) \rfloor + 1}$ and $s'' = s' - s$. Then $s' > s > 2^{\lfloor \log_2(s) \rfloor} > s'' > 0$ and $0 < \frac{s''}{s'} < 0.5$. For example, when $s = 3$ ($s' = 4$ and $s'' = 1$), we have $4 > 3 > 2 > 1$ and $\frac{s''}{s'} = \frac{1}{4}$.

3. For s, s', s'' defined in Property 2, $a_j(s)$ can be expressed as

$$a_j(s) = |a_j(s') - a_j(s'')|, \tag{18}$$

   where $|z|$ represents absolute value of z. This can be verified using Equation (16).

4. For s, s', s'' defined in Property 2, $a_j(s'')$ and $a_j(s')$ can be related as

$$a_j(s')a_j(s'') = a_j^2(s')\frac{s''}{s'} \tag{19}$$

   This can be obtained using Property 1 or Equation (16).

Lemmas

Three lemmas need to be developed in order to prove the main theorem.

Let $\hat{X}(s)$ represent the averaged aggregate-process of $\hat{x}(t)$. Let $V(s)$ and $\hat{V}(s)$ denote the variances of $X(s)$ and $\hat{X}(s)$, respectively. $V(s)$ can be computed directly through Equation (14) and Equation (15) as

$$V(s) = \frac{1}{s^2}\sum_{j=1}^{K} a_j^2(s)2^{-j(2-2H)}(2^{2(1-H)} - 1)\sigma^2(1 - O(2^{-2K(1-H)})). \tag{20}$$

$\hat{V}(s)$ has been shown in [31] as

$$\hat{V}(s) = \frac{1}{s^{2-2H}}\sigma^2. \tag{21}$$

The following lemmas will relate $V(s)$ with $\hat{V}(s)$. In particular, Lemma 1 shows that $V(s) \leq \hat{V}(s)$ for $s = 2^k$, where $k = 0, 1, 2, 3, \ldots$. Lemma 3 shows that $V(s) < \hat{V}(s)$ for other positive integer s. Lemma 2 is used by Lemma 3.

Lemma 1 *For* $s = 2^{\lfloor log_2(s) \rfloor}$,

$$V(s) \leq \hat{V}(s). \tag{22}$$

Proof: By inserting the expression of $a_j(s)$ (Property 1) into Equation (20), we have

$$\begin{aligned} V(s) &= \frac{1}{s^2} \sum_{j=\log_2(s)+1}^{K} s^2 2^{-j(2-2H)}(2^{2(1-H)} - 1)\sigma^2(1 - O(2^{-2(1-H)K}) & (23) \\ &= s^{-2(1-H)}\sigma^2(1 - O(2^{-2(1-H)K}) & (24) \\ &\leq s^{-2(1-H)}\sigma^2. & (25) \end{aligned}$$

From the definition of $\hat{V}(s)$ (Equation 21), we have $V(s) \leq \hat{V}(s)$ for $s = 2^{\lfloor log_2(s) \rfloor}$.

Lemma 2 *Define* $g(\alpha)$ *to be*

$$g(\alpha) = 1 + \alpha^{2H} - 2\alpha - (1 - \alpha)^{2H}. \tag{26}$$

then $g(\alpha) < 0$, *for* $\alpha \in (0, 0.5)$ *and* $H \in (0.5, 1)$.

Proof: It can be verified easily that $g(0) = g(0.5) = 0$, and second derivation of $g(\alpha)$ is

$$g''(\alpha) = (2H)(2H - 1)\{\alpha^{2H-2} - (1 - \alpha)^{2H-2}\}. \tag{27}$$

For $0.5 < H < 1$ and $0 < \alpha < 0.5$, $g''(\alpha) > 0$. Therefore, $g(\alpha) < 0$ for $\alpha \in (0, 0.5)$.

Lemma 3 *For* $s > 2^{\lfloor log_2(s) \rfloor}$,

$$V(s) < \hat{V}(s), \tag{28}$$

i.e.,

$$s^{2H} > \sum_{j=1}^{K} a_j^2(s) 2^{-j(2-2H)}(2^{2(1-H)} - 1). \tag{29}$$

Proof: We will use the induction for the proof. We will start by first proving the inductive hypothesis. We will then show the conclusion holds for $s = 3$ as well [24].

Inductive hypothesis: we assume Lemma 3 holds for any positive integer n satisfying $3 < n < s$, i.e. $n^{2H} > \sum_{j=1}^{K} a_j^2(n) 2^{-j(2-2H)}(2^{2(1-H)} - 1)$. Now we will prove the lemma holds for s.

Using the expression for $a_j(s)$ given by Equation (29), we have

$$\begin{aligned} & \sum_{j=1}^{K} a_j^2(s) 2^{-j(2-2H)}(2^{2(1-H)} - 1) & (30) \\ = & \sum_{j=1}^{K} (a_j(s') - (a_j(s''))^2 2^{-j(2-2H)}(2^{2(1-H)} - 1) \\ = & (2^{2(1-H)} - 1) \sum_{j=1}^{K} a_j^2(s') 2^{-j(2-2H)} - (2^{2(1-H)} - 1) \sum_{j=1}^{K} 2a_j(s')a_j(s'') 2^{-j(2-2H)} \end{aligned}$$

[24] The cases for $s = 1, 2$ have been included in Lemma 1.

$$+ \quad (2^{2(1-H)}-1)\sum_{j=1}^{K} a_j^2(s'')2^{-j(2-2H)} \tag{31}$$

$$= \quad (2^{2(1-H)}-1)(1-2\frac{s''}{s'})\sum_{j=1}^{K} a_j^2(s')2^{-j(2-2H)} + (2^{2(1-H)}-1)\sum_{j=1}^{K} a_j^2(s'')2^{-j(2-2H)}, \tag{32}$$

where the Equation (32) is obtained by applying Property 4 and by combining the first two terms of Equation (31). To upper bound Equation (32), we notice the first term can be upper bounded by $s'^{2H}(1-2\frac{s''}{s'})$ due to Lemma 1 for $s' = 2^{\lfloor log_2(s') \rfloor + 1}$ (true by the definition of s'). The second term of Equation (32) is bounded by s''^{2H} due to either the inductive hypothesis for $s'' \neq 2^{\lfloor log_2(s'') \rfloor}$, or Lemma 1 for $s'' = 2^{\lfloor log_2(s'') \rfloor}$. Therefore, we have

$$\sum_{j=1}^{K} a_j^2(s)2^{-j(2-2H)}(2^{2(1-H)}-1) \tag{33}$$

$$\leq \quad s'^{2H} + s''^{2H} - 2s'^{2H}\frac{s''}{s'}. \tag{34}$$

Since $\frac{s''}{s'} = \alpha$ and $\alpha \in (0, 0.5)$ due to Property 2, using Lemma 2 we can further obtain

$$\sum_{j=1}^{K} a_j^2(s)2^{-j(2-2H)}(2^{2(1-H)}-1) \tag{35}$$

$$\leq \quad s'^{2H}(1+\alpha^{2H}-2\alpha) \tag{36}$$

$$< \quad s'^{2H}(1-\alpha)^{2H} \tag{37}$$

$$= \quad s^{2H}. \tag{38}$$

Therefore, the inductive hypothesis holds for all positive integer $s > 3$ and $s \neq 2^{\lfloor \log_2(s) \rfloor}$.

For $s = 3$, $s' = 4$ and $s'' = 1$. Since both s' and s'' satisfy the condition of Lemma 1, the proof will follow what given from Equations (33) through (38). Then for all $3 \leq s \leq K$, the hypothesis holds.

Q.E.D.

Proof of the Main Theorem

Proof: Two steps are needed to prove the main theorem. First, we will prove an upper bound using the lemmas. We will then obtain a lower bound with a Weibull-decay, and show that the two bounds are of the same order.

Proof of The Upper Bound

The proof of the upper bound, $P(B_N > B) \leq P(\hat{B}_N > B)$, is based on the lemmas developed in the previous section. First, we note that a sufficient condition for this to hold is

$$P(X(s) > v) \leq P(\hat{X}(s) > v) \tag{39}$$

for any integer s $(1 \leq s \leq N)$ and v bigger than the mean of $X(s)$.

Since for every t, $x(t)$ and $\hat{x}(t)$ are Gaussian random variable, $X(s)$ and $\hat{X}(s)$ are Gaussian as well with the same mean but different variances. Therefore, to show the inequality, we only need to show that the

variance of $X(s)$ is no bigger than that of $\hat{X}(s)$ for all $s \in [0, N-1]$, i.e.,

$$V(s) \leq \hat{V}(s). \tag{40}$$

This is true due to Lemmas 1 and 3. Therefore, we complete our proof for the upper bound.

Proof of the Lower Bound

To obtain a lower bound, we use a similar method to that in [31] but different quantities from the synthesized traffic due to independent wavelet coefficients. As in [31], the probability $P(B_N > B)$ can be lower bounded as

$$P(B_N > B) \geq \sup_{k=1}^{K} P(X(2^k) > (C-\mu) + 2^{-k}B). \tag{41}$$

Here we consider only $s = 2^k$ with k being an integer ($1 \leq k \leq K$). Define

$$P_L^k \stackrel{\text{def}}{=} P(X(2^k) > (C-\mu) + B2^{-k}). \tag{42}$$

Since the (independent) wavelet coefficients generated by Algorithm 1 are Gaussian random variables, $X(2^k)$ is a Gaussian random variable with a zero mean and a variance $2^{2(k-1)(1-H)}\sigma^2(1-O(\frac{1}{N^{2(1-H)K}}))$ for $1 \leq k \leq K$ as given by Equation 15. Then we can obtain

$$P_L^k \geq Q(\frac{(C-\mu)f(k)}{\sigma}), \tag{43}$$

where $f(k) = \frac{1+2^{-k}\frac{B}{C-\mu}}{2^{-k(1-H)}}$, and $Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(\frac{-t^2}{2})dt$.

Let k^* be the the critical time-scale, which indicates that the loss is more likely to happen at the 2^{k^*} time interval. We have $k^* = \arg\min_k f(k)$, where $k^* = k_0$, with B as given in Equation (9). Then $P_L^{k*} \geq Q(\frac{(C-\mu)f(k^*)}{\sigma})$. Furthermore, using the approximation $Q(x) \sim \exp(-\frac{x^2}{2})$ for x large, and $\sup_k(P_L^k) \geq P_L^{k^*}$, we have

$$P(B_N > B) \geq \exp(-\frac{(C-\mu)^2(\frac{B}{C-\mu})^{2(1-H)}(\frac{1-H}{H})^{2H}}{2\sigma^2(1-H)^2}). \tag{44}$$

The bound obtained above has a similar form to the buffer overflow probability for the continuous-time FGN process[31].

Combining The Lower and The Upper Bound

Since it has been shown in [31] that $P(\hat{B}_N > B) \sim \exp(-\frac{(C-\mu)^2(\frac{B}{C-\mu})^{2(1-H)}(\frac{1-H}{H})^{2H}}{2\sigma^2(1-H)^2})$ as B tends to infinity, the upper and the lower bounds for $P(B_N > B)$ are of the same order when $K - k^*$ is large. Hence, we have

$$P(B_N > B) \sim \exp(-\frac{(C-\mu)^2(\frac{B}{C-\mu})^{2(1-H)}(\frac{1-H}{H})^{2H}}{2\sigma^2(1-H)^2}), \tag{45}$$

or $P(B_N > B) \sim P(\hat{B}_N > B)$.

Q.E.D.