Measurement-based Real-time Traffic Model Classification

Yi Zeng
Department of Electrical Engineering
School of Engineering
Southern Methodist University
Dallas, TX 75275
Email: yizeng@engr.smu.edu

Thomas M. Chen
Department of Electrical Engineering
School of Engineering, PO Box 750338
Southern Methodist University
Dallas, TX 75275-0338, USA
Email: tchen@engr.smu.edu

Abstract— **A new method for real-time traffic model classification is proposed and evaluated. The method classifies the current measured traffic to a "best-fit" model selected from a library of candidate models using statistical estimation techniques. A two-model system has been prototyped and evaluated through simulation experiments. The experimental system consists of a short-range dependent model and long-range dependent model, and uses the estimated Hurst parameter to select between the two models to choose the model requiring an equivalent bandwidth (EB) that is cloest to the actual required EB. Results demonstrate that the two-model system can classify observed traffic to the correct model with fair accuracy, and can automatically detect a change in traffic characteristics after a delay. The design parameters effecting the classification accuracy and delay to detect traffic changes are discussed.**

I. Introduction

Traditionally, traffic modeling is performed off-line on historical traffic data, and the results may not be relevant to current network conditions. It would be more useful to analyze traffic in real time if possible, but the problem of real-time traffic modeling is complicated by the fact that traffic characteristics may vary over time and depend on the particular traffic type. Traffic studies suggest that a single model cannot adequately represent all types of traffic. To address this problem, we have developed a new method for traffic modeling that classifies observed traffic to the "best-fit" traffic model from a library of models using statistical estimation techniques. The model classification is updated continuously in real time as more traffic is observed. With accurate real-time traffic modeling, it will be possible to better adapt resource allocation to current network conditions.

In this paper, we focus on the problem of model classification using a two-model library to choose a model with the equivalent bandwidth (EB) that is closest to the actual required EB. In Section II, we describe the general model classification approach. Section III presents theoretical and experimental results for the simple two-model scheme. In Section IV, we study the system for equivalent bandwidth calculation. Section V are the conclusions and research issues for future work.

II. Model-library traffic classification

Many traffic models have been developed over the years for various types of traffic. It can be seen from these studies that a single model cannot adequately represent all types of Internet traffic. Moreover, traffic characteristics can vary randomly over time. Hence, we have investigated an adaptive method where the best-fit traffic model is dynamically selected based on current traffic measurements. An overview of the general approach is shown in Fig. 1. A model library consists of a number of candidate traffic models. It is designed intentionally to be modular, instead of an integrated expert system, so that individual models can be added or changed without effecting the entire system. As the traffic rate is observed (represented by a time series), statistics for the traffic are continually updated and used to dynamically select one of the candidate models as the best-fit model. The selected model represents the current traffic behavior. Ultimately, the best-fit traffic model that is output from this system can be used to adapt resource allocation algorithms in real time.

In the general case with N candidate models, it is difficult to identify a sufficient set of statistics for the model selection. Also, it is very complicated to fully evaluate the accuracy of a general system. Therefore, we have chosen to focus our initial study on a simple two-model system where a single statistic is sufficient to differentiate between the two candidate models. The feasibility of the two-model system is demonstrated and evaluated through simulation experiments. Our objective is to generalize the results from the two-model system to better design the general N-model system.

III. Results for Two-Model Classification

A. Experimental Two-Model System

In our initial experiments, two candidate models were implemented: a Poisson process to represent short-range dependent (SRD) traffic and a fractional Gaussian noise (fGn) process for long-range dependent (LRD) traffic. The Poisson process has been widely used for traditional SRD data sources.

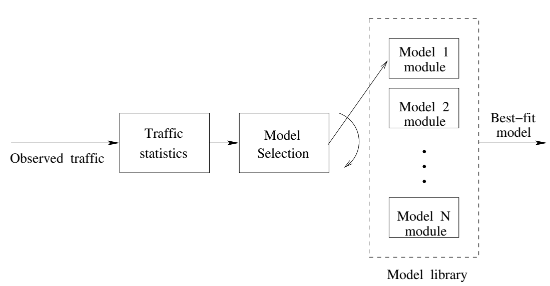

Fig. 1. General model classification system

Recently, Ethernet data traffic was found to exhibit self-similar properties [1]. Evidence for self-similarity was also reported in wide area traffic [2], variable-bit rate (VBR) video traffic [3] [4], World Wide Web traffic [5], and SS7 traffic [6].

An LRD or self-similar process $X(t)$ has an autocovariance function that decays hyperbolically,

$$\gamma_X(k) \sim |k|^{-2(1-H)} \tag{1}$$

when $|k| \rightarrow \infty$ and $0.5 < H \leq 1$. In contrast, SRD processes have exponentially decaying autocovariance. The Hurst parameter H is a property of the process indicating its degree of self-similarity (higher H indicates more self-similarity). Because of the central importance of the Hurst parameter H in characterizing LRD processes, it is natural to use the parameter as the statistic to choose between the SRD or LRD candidate models. We investigated the use of a threshold value for $\hat{H}$, denoted by T, to choose between the LRD and SRD candidate models. If $\hat{H} < T$, the Poisson model was selected; otherwise, the fGn model was chosen. However, the randomness of the estimation of H may result in selection of the wrong model, so the probability of misclassification is an issue which is discussed in section III-C.

B. Wavelet-based Hurst Parameter Estimation

The first issue examined was accurate estimation of the Hurst parameter. Considering the different known estimation methods, the Abry-Veitch (AV) wavelet-based estimator was chosen because it has good statistical characteristics [7] [8]. The AV wavelet-based estimator performs a time average of wavelet coefficients $|w_x(j,k)|^2$:

$$\Gamma_x = \frac{1}{n_j} \sum_{k=1} n_j |w_x(j,k)|^2 \tag{2}$$

where n_j is the number of wavelet coefficients at a given scale level j. An estimator $\hat{H}$ for the Hurst parameter can be obtained by making a linear regression of $\log_2(\Gamma_x)$ in the scaling range $[j_1, j_2]$ which should have a slope of $2H-1$:

$$\log_2(\Gamma_x) = (2H-1)j + \hat{e} \tag{3}$$

In this study, the whole interval of the scales worked well as the scaling range.

For the wavelet-based estimator, the number of data samples is required to be a power of 2. Previous studies have indicated that the variance of the estimator decreases quickly with the sample size [10]. Thus, for accuracy, a larger sample size will always be better. On the other hand, larger sample sizes will increase the time needed for model selection. The choice of sample size involves a trade-off between estimation accuracy and time for model selection. In the following experiments, we used a sample size of 1024 which appeared to be a reasonable compromise. In practice, the choice of sample size may be constrained by a desired level of accuracy or time for model selection.

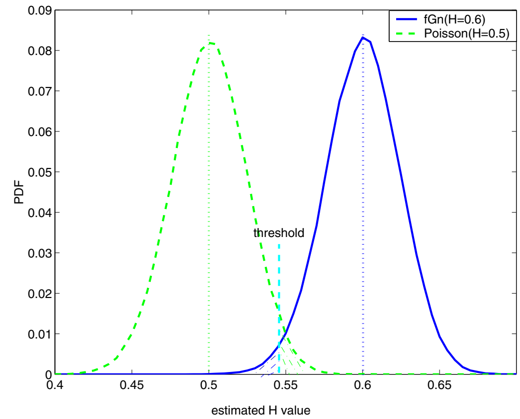

Fig. 2. Empirical PDF of $\hat{H}$ measured for $H = 0.5, 0.6$ data

C. Accuracy of Model Classification Using Thresholds

The next issue examined was the optimal choice of threshold and resulting accuracy of model classification. Theoretical studies have found that $\hat{H}$ is unbiased and approximately normally distributed, that is, $\hat{H} \sim N(H, \sigma^2_{\hat{H}})$ [8]. Moreover, the variance does not depend on H and is given by

$$\sigma^2_{\hat{H}} = \frac{2^{j_1-1}}{n \ln^2 2} \frac{1-2^{-J}}{1-2^{-(J+1)}(J^2+4)+2^{-2J}} \tag{4}$$

where $J = j_2 - j_1 + 1$, $[j_1, j_2]$ is the scaling range, and n is the sample size. Here, $j_1 = 1$, $j_2 = 10$, $n = 1024$, then the standard deviation $\sigma_{\hat{H}}$ from (4) turned out to be 0.046.

Fig. 2 shows examples of empirical PDFs (probability density functions) of $\hat{H}$ measured for 100,000 simulation runs of Poisson and fGn ($H = 0.6$) data. The empirical PDF is normally distributed around H. The only notable difference from the theoretical PDF is a smaller measured standard deviation $\hat{\sigma}_{\hat{H}} = 0.025$ (compared to 0.046).

Given $\hat{H} \sim N(H, \sigma^2)$, we can calculate the probability of misclassification of the system for a specific threshold, and find the optimal threshold to minimize the misclassification probability. It was observed in our simulation experiments that the optimal threshold could be obtained in the midpoint of two model's Hurst values, given that the traffic was one of the two models with equal likelihood. In the following we will derive the probability of misclassification for a specific threshold value T, as well as the optimal threshold T_o if model is selected with a priori probability p.

Suppose we have two candidate models with Hurst parameters H_1 and H_2, respectively, where $H_1 < H_2$. Model 1 is selected if $\hat{H} < T$, and model 2 is selected if $\hat{H} > T$. If the actual traffic is model 1, then the conditional probability of misclassification will be

$$\begin{aligned} P_{m1} &= Pr(\hat{H} > T \mid model1) \\ &= \int_T^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-H_1)^2}{2\sigma^2}} dy \end{aligned} \tag{5}$$

If the actual traffic is model 2, then the conditional probability of misclassification will be

$$P_{m2} = Pr(\hat{H} < T \mid model2) = \int_{-\infty}^{T} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-H_2)^2}{2\sigma^2}} dy \quad (6)$$

If the actual traffic is model 1 with probability p, then the total probability of misclassification will be

$$P_m = P_{m1} \cdot p + P_{m2} \cdot (1-p) \quad (7)$$

which is proportional to the shaded area in Fig. 2. The optimal threshold minimizing the total probability of misclassification can be found by taking the derivative of (7) with respect to T and setting it equal to 0:

$$0 = \frac{\partial}{\partial T} P_m = \frac{1-p}{\sqrt{2\pi}\sigma} e^{-\frac{(T-H_2)^2}{2\sigma^2}} - \frac{p}{\sqrt{2\pi}\sigma} e^{-\frac{(T-H_1)^2}{2\sigma^2}} \quad (8)$$

This simplifies to

$$lnp - \frac{(T_o - H_1)^2}{2\sigma^2} = ln(1-p) - \frac{(T_o - H_2)^2}{2\sigma^2} \quad (9)$$

The solution is:

$$T_o = \frac{H_2^2 - H_1^2 + 2\sigma^2 \ln \frac{p}{1-p}}{2(H_2 - H_1)} \quad (10)$$

When the traffic is either model 1 or 2 with equal likelihood (p = 0.5), T_o is simply the midpoint between H_1 and H_2:

$$T_o = \frac{H_1 + H_2}{2} \quad (11)$$

Fig. 3 shows the misclassification probability P_m when model 1 is Poisson ($H_1 = 0.5$) and model 2 is fGn with $H_2 = 0.6$ with equal likelihood. It was observed that theoretical and experimental results for misclassification probabilities agreed very closely when the experimental standard deviation ($\sigma_{\hat{H}} = 0.025$) was used. Fig. 3 also shows that the total misclassification probability P_m is minimized by a threshold of $T_o = 0.55$ as expected from (11).

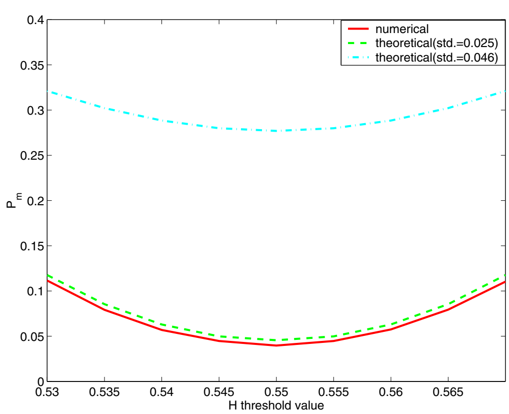

Fig. 3. Misclassification Probability P_m for $H = 0.5, 0.6$ traffic

Moreover, given that p is known *a priori*, the probability of misclassification P_m for a specific threshold T can be calculated from (7). Likewise, for a given P_m, we can find out how close the two models' Hurst values can be and choose one threshold to meet the given probability.

D. Detection of Traffic Changes

In the previous experiments, the simulated traffic was entirely SRD or LRD, and the problem was to select the correct traffic model. The estimation of the Hurst parameter was relatively straightforward. In more realistic circumstances, the traffic characteristics may be dynamic and changing over time. The Hurst parameter estimation must be updated continuously to detect the changes. In the following experiments, the simulated traffic consisted of alternating SRD ($H = 0.5$) and LRD ($H = 0.8$) intervals, and the problem is to correctly classify the intervals to model 1 (Poisson) or model 2 (fGn with $H = 0.8$). With an optimal threshold of $T = 0.65$, the misclassification probability was very small, $P_m \sim 10^{-9}$. The performance metric of interest is the time required for the system to detect a change in the traffic.

As before, the Hurst parameter estimation is computed over windows (intervals) of 1024 data samples. As more traffic is observed, the window is advanced by a step size s, and the Hurst parameter is re-estimated, as shown in Fig. 4. Thus, consecutive calculations of the Hurst parameter estimator share an interval of $1024 - s$ data samples in common. Whenever the Hurst parameter is updated, the model selection is also recalculated (using the threshold). In the case that Hurst parameter is a function of time, we assume that in the sliding window period of time, the Hurst parameter is a constant. An example of the procedure is shown in Fig. 5 with a sliding window size $s = 64$. The traffic consists of 4096 Poisson data samples followed by 4096 fGn data samples and another 4096 Poisson data samples. It took 320 data samples for $\hat{H}$ to fall below T and detect the change from fGn to Poisson traffic, and 640 data samples to detect the change from Poisson to fGn traffic.

Let $D_{s\text{-}fp}$ denote the delay (in data samples) to detect a change from fGn to Poisson traffic, and $D_{s\text{-}pf}$ the delay to detect a change from Poisson to fGn. The delay is measured as the difference between the time of the actual traffic change to the corresponding change in the output of the model classification system. There must be some delay before enough data samples cause the estimator $\hat{H}$ to cross the threshold and change the model selection.

Two factors effect the delays $D_{s\text{-}fp}$ and $D_{s\text{-}pf}$: the step

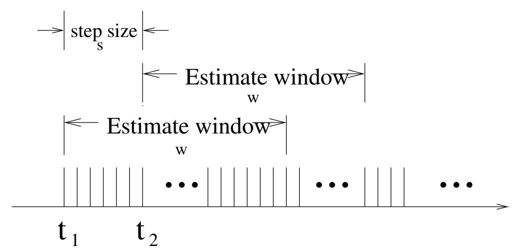

Fig. 4. Sliding windows used for estimation

size s and the window size w that is the sample size for each calculation of the Hurst parameter. These factors should be chosen to shorten the delay to detect traffic changes, but there is a trade-off between shorter delay and accurate estimation. Fig. 6 shows that D_{s_fp} and D_{s_pf} generally increase with larger step size s, as might be expected. Larger steps mean that the Hurst parameter is updated less frequently, so it takes longer to detect a traffic change. However, s cannot be updated too frequently due to computation cost. In Fig. 7, the step size is fixed to $s = 64$, and D_{s_fp} and D_{s_pf} are shown to increase with w. It appears that w should be minimized, but small w will result in larger variance in the estimator $\hat{H}$ and more likelihood of model misclassification.

IV. Equivalent Bandwidth Calculation

The concept of equivalent bandwidth (EB) has been developed recently to provide a measure of resource usage and it is a useful approach for resource allocation [11]. The equivalent bandwidth of a source is the minimum required bandwidth such that QoS requirement can be met [12]. For a stationary source, the equivalent bandwidth is defined as [11]:

$$\alpha(s,t) = \frac{1}{st} \log E[e^{sX[0,t]}] \quad 0 < s, t < \infty, \qquad (12)$$

where $X[0,t]$ is the amount of traffic that arrives from a source in $[0,t]$ interval and $X[0,t]$ has stationary increments.

Equivalent bandwidth formulae have been developed for several models, eg, fractional Brownian Motion (fBM) model and Poisson model. The equivalent bandwidth (EB) formula for fBM model is [13]:

$$EB_{fBM} = m + (H^H(1-H)^{1-H}\sqrt{-2ln\epsilon})^{\frac{1}{H}} v^{\frac{1}{2H}} B^{1-\frac{1}{H}} m^{\frac{1}{2H}}, \qquad (13)$$

where m stands for the mean rate of the incoming traffic, H is the Hurst parameter, ϵ is the required cell loss ratio (CLR), B is the buffer size of the server, and v is the variance of the traffic.

The EB formula of a Poisson process with mean rate λ is:

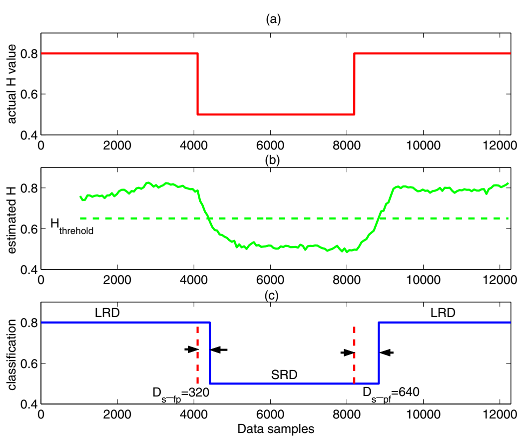

Fig. 5. Detection of traffic changes with $w = 1024$, $s = 64$: (a) True Hurst value of simulated traffic; (b) estimated Hurst value $\hat{H}$; (c) Hurst value of classified traffic.

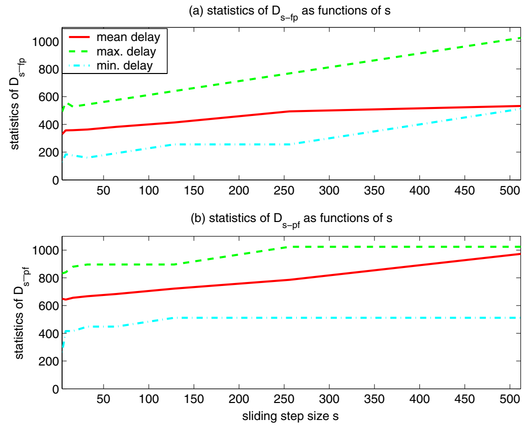

Fig. 6. D_{s_fp} and D_{s_pf} measured as functions of s: (a) statistics of D_{s_fp} as functions of s; (b) statistics of D_{s_pf} as functions of s.

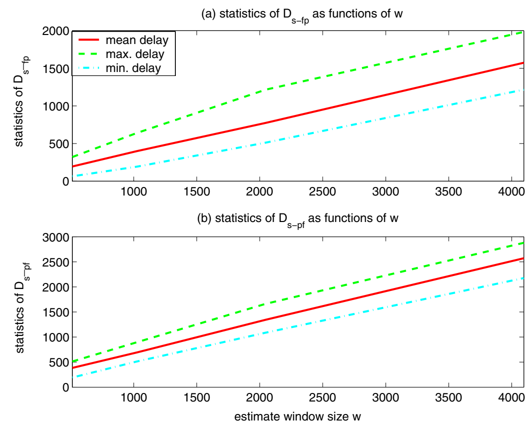

Fig. 7. D_{s_fp}, D_{s_pf} measured as functions of w: (a) statistics of D_{s_fp} as functions of w; (b) statistics of D_{s_pf} as functions of w.

$$EB_{poisson} = \frac{\lambda(e^{\theta} - 1)}{\theta}, \qquad (14)$$

Here $\theta \approx -\frac{ln\epsilon}{B}$, so we can see the EB of a Poisson process only has relation to do with mean rate, CLR and the buffer size, but has no relation to do with Hurst parameter and the variance of the traffic.

We learned from (13) that there is a relationship between Hurst parameter and equivalent bandwidth, as shown in Figure 8. In our classification system, the ultimate purpose of classification is to estimate the true EB for the incoming traffic. To do that, we should choose the traffic model that comes closest to the true EB, eg, model i is chosen if $|EB_i - EB_{true}|$ has the minimum value. However, the true EB is unknown and instead we must choose a model based on the estimated H parameter. Then the question becomes how to identify the model with the EB closest to the true EB using only the estimated H. (13) seems too complicated to find the reverse function, but when we study the relation between $log10(EB)$ and H, we find that it is nearly linear and can be fit by a regression model. Figure 9 shows the relation between $log10(EB)$ and H, and

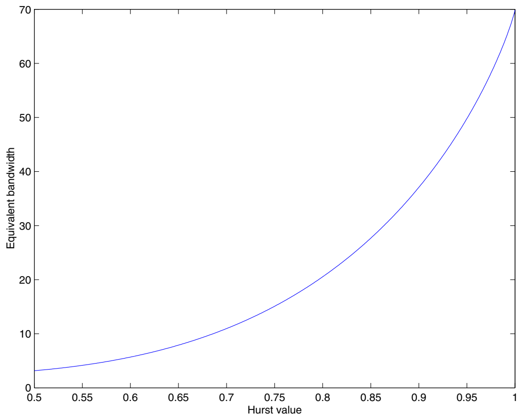

Fig. 8. Relationship between Hurst parameter and equivalent bandwidth

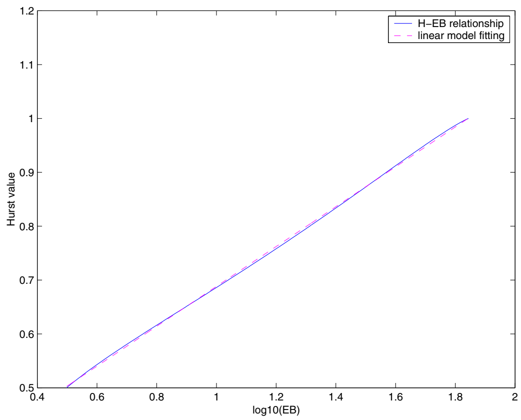

Fig. 9. Relationship between $log10(EB)$ and Hurst parameter

the fitted linear model. In this figure we used the parameters: mean rate = 2 (Mbps), buffer size = 1000, cell loss ratio = 0.00001, Hurst value $\in [0.5, 1]$, and variance = 10; then the linear model is

$$H = 0.37 \cdot \log 10(EB) + 0.3181. \quad (15)$$

This gives a way to do the threshold transformation. For example, in the classification system we have a fBM model with $H = H_1$ and a Poisson model with $H = H_2$. The EB of the coming traffic could be either of EB_{fBM} or $EB_{Poisson}$. The EB threshold is set to be the mid-point value of EB_{fBM} and $EB_{Poisson}$ since we choose model A if $|EB_A - EB_{true}| < |EB_B - EB_{true}|$. We can find from (15) that the corresponding Hurst parameter for the EB threshold as the Hurst threshold to do the classification. In one experiment, we set $H_1 = 0.8$ and $H_2 = 0.5$, using the parameters that stated above, we got $EB_{fBM} = 20.5300$, $EB_{Poisson} = 0.0232$, and the $H_{threshold} = 0.6925$, nearly the mid-point of $H = 0.8$ and $H = 0.5$.

V. Conclusions

This paper has presented a new model classification method and demonstrated its feasibility with the simple two-model case. The results for the two-model system using a simple threshold show that it can be fairly accurate when the traffic characteristics are static. When traffic characteristics are changing, the model classification problem is complicated by the need to compute the Hurst parameter over long data intervals (for accuracy) and the opposite need for short data intervals (to detect traffic changes quickly). With sliding windows, experimental results indicate that small step sizes and short windows can help to reduce the delay to detect traffic changes, but at a cost of more computation and higher likelihood of model misclassification.

We studied how to choose the model to meet the equivalent bandwidth requirement while only use the estimated Hurst parameter. But the probobility of misclassification based on the EB requirement is still needed to be studied.

Future work will concentrate on how to apply this model to call admission control algorithm and how to develop it into N model system (In this case, instead of using Hurst parameter alone, a set of parameters need to be chosen for classification, e.g., autocorrelation function and probability distribution function of the traffic). Meanwhile, more models like Markov-modulated Poisson process (MMPP) model, autoregressive moving average (ARMA) model, fractional autoregressive integrated moving average (p, d, q) (FARIMA (p, d, q)) model, etc, will be built into the library.

References

[1] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, "On the self-similar nature of ethernet traffic(extended version)," *IEEE/ACM Trans. Networking*, no. 2, pp. 1–25, 1994.

[2] V. Paxon and S. Floyd, "Wide-area traffic: the failure of poisson modeling," *IEEE/ACM Trans. Networking*, vol. 3, pp. 226–244, June 1995.

[3] M. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar vbr video traffic," in *Proc. of ACM SIGCOMM'94*, 1994.

[4] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit rate video traffic," *IEEE Trans. Commun.*, vol. 43, 1995.

[5] M. E. Crovella and A. Bestavros, "Self similarity in www traffic: evidence and possible causes," *IEEE/ACM Trans. Networking*, vol. 5, pp. 835–846, Dec. 1997.

[6] D.E. Duffy, A.A. McIntosh, M. Rosenstein, and W. Willinger, "Statistical analysis of ccsn/ss7 traffic data from working subnetworks," *IEEE J. Select. Areas Commun.*, vol. 12, no. 3, 1994.

[7] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Inform. Theory*, vol. 44, no. 1, Jan. 1998.

[8] D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 878–896, Apr. 1999.

[9] J.C. Goswami and A.K. Chan, *Fundamentals of wavelets: theory, algorithms, and applications*. John Wiley & Sons, Inc, 1999.

[10] M. Roughan, D. Veitch and P. Abry, "Real-time estimation of the parameters of long-range dependence," *IEEE/ACM Trans. Networking*, vol. 8, pp. 467–478, Aug. 2000.

[11] F. Kelly, "Notes on Effective Bandwidths," *Stochastic networks: theory and applications*, pp. 141–168, Oxford Press, 1996.

[12] N.L.S. Fonseca, G.S. Mayor and C.A.V. Neto, "On the equivalent bandwidth of self-similar sources," *ACM Trans. Modeling and Computer Simulation*, vol. 10, Issue 2, April 2000.

[13] I. Norros, "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 953–962, Aug. 1995.