to appear in IPOM 2002 , October 2002

Automatic Model Classification of Measured Internet Traffic

Yi Zeng and Thomas M. Chen
Department of Electrical Engineering
PO Box 750338
Southern Methodist University
Dallas, TX 75275
yizeng@engr.smu.edu, tchen@engr.smu.edu

***Abstract*— In this paper, a new method for automatic traffic model classification is investigated. This method tries to classify the current measured traffic to a "best-fit" model selected from a model library in real time. In our initial simulation experiments, a model library has been constructed with two models representing short-range dependent and long-range dependent traffic, respectively. We use the measured Hurst parameter statistic to choose between the two traffic models. The accuracy of the method is evaluated under different circumstances with simulated traffic.**

***Index Terms*— Model classification, long-range dependence, Hurst parameter, wavelet-based estimator**

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 naturally be more useful to analyze traffic in real time. For example, resource allocation could be better adapted given an accurate model of the current network traffic. However, the problem of real-time traffic modeling is complicated by the fact that traffic characteristics may vary over time and depend on the particular type of source. Traffic studies suggest that a single model cannot adequately represent all types of traffic. To address this problem, we have investigated an approach based on a modular library of traffic models. The "best-fit" traffic model for measured traffic is selected dynamically from the library using statistical estimation techniques. The model selection is updated continuously in real time as more traffic is observed.

In this paper, we focus on the simple problem of model classification using a two-model library. Ultimately, the lessons learned will be generalized to a system with an arbitrary number of models. In Section II, we describe the general model classification approach. Section III presents preliminary simulation results for the simple two-model scheme. In the conclusions, we discuss ideas for extending our method to the general N-model case and outline research issues for future work.

II. Model-library traffic classification

Many traffic models have been developed over the years for various types of traffic. From these studies, it appears that a single model cannot adequately represent all types of Internet traffic. Moreover, traffic characteristics can vary over time and depend on the particular type of sources. Hence, we have investigated an adaptive method where the best-fit traffic model is dynamically chosen based on current traffic measurements. An overview of the general approach is shown in Figure 1. A model library consists of a number of pre-programmed candidate traffic models. It is designed to be modular (instead of integrated) to allow different models to be added, changed, or removed, without effecting the overall system. As traffic is observed in real time (represented by a time series), traffic statistics are continually updated, and the statistics are used to dynamically select one of the candidate models as the best-fit model. Ultimately, the best-fit traffic model that is output from this system can be used to adapt resource allocation algorithms.

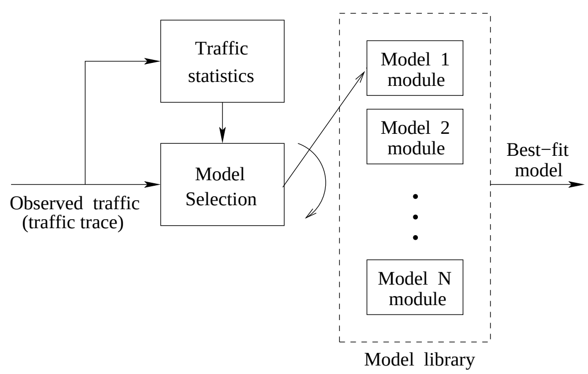


Fig. 1. General model classification system

In the general case with N candidate models, it is difficult to identify a sufficient set of statistics for the model selection. Also, it would be 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 evaluated by means of simulation experiments. Our objective is to genaralize the results from the two-model system to better design the general N-model system.

III. Results for two-model classification

A. An Experimental Two-Model System

In initial experiments, two candidate models were implemented: a Poisson process to model short-range dependent (SRD) traffic and a fractional Gaussian noise (fGn) process to model long-range dependent (LRD) traffic. The Poisson process has been widely used for traditional SRD data sources. Recently, Ethernet data traffic was found to exhibit self-similar properties [1]. Evidence for self-similarity was reported in wide area traffic [2], variable-bit rate (VBR) video traffic [3] [4], World Wide Web traffic [5], and SS7 traffic as well [6].

An LRD or self-similar process has an autocovariance function that decays hyperbolically

$$Cov(X_n, X_{n+k}) \sim |k|^{-\beta} \quad (1)$$

when $|k| \rightarrow \infty$ and $0 < \beta < 1$. In contrast, SRD processes have exponentially decaying autocovariance. The Hurst parameter H ($0.5 \leq H \leq 1$) is used to indicate the degree of self-similarity of a process (higher H indicates more self-similarity). The parameter β is related to the Hurst parameter by $\beta = 2(1 - H)$. 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.

B. Wavelet-based Hurst Parameter Estimation

Wavelets are a powerful tool to analyze the scale-based properties of data. Since a self-similar process has similar statistical characteristics over a broad range of scales, wavelet transforms are naturally well suited to examine the self-similarity of network traffic. In this study, the Abry-Veitch wavelet-based estimator for the Hurst parameter was chosen because it has been reported to be a good choice for on-line applications and has advantages over Whittle's method [7] [8].

1) Scaling range: Wavelet transforms provide a means for simultaneously examining a time series at a range of different scales a, while maintaining the time dimension information. Multiresolution analysis (MRA) theory defines a dyadic grid, $(a, t) = (2^j, 2^j k), j, k \in N$, and shows that no information is lost if the continuous wavelet coefficients are sampled at the grid. Thus leads to the discrete wavelet transform [9]. Given a time series $X(k)$ and the initial scaling function $a_x(0, k) = X(k)$, we can compute the wavelet coefficients $w_x(j, k)$ at scale 2^j $(1 \leq j \leq \log_2 N)$ by convolving the scaling functions with filters $h(n)$ and $g(n)$:

$$a_x(j, k) = h a_x(j - 1, k) \quad (2)$$

$$w_x(j, k) = g a_x(j - 1, k) \quad (3)$$

where $h(n)$ and $g(n) = (-1)^n h(1 - n)$ are discrete-time low-pass and high-pass filters, respectively.

TABLE I

Vanishing moment test for fGn data.

	H = 0.6			H = 0.7		
	N = 1	N = 2	N = 3	N =1	N =2	N =3
$\hat{H}_{max}$	0.6684	0.6702	0.7045	0.7994	0.8314	0.8589
$\hat{H}_{min}$	0.5040	0.4839	0.4908	0.6290	0.6065	0.6053
$\hat{H}_{mean}$	0.5986	0.6064	0.6115	0.7017	0.7176	0.7225
$\hat{H}_{stat}$	0.0341	0.0374	0.0410	0.0368	0.0423	0.0420
	H = 0.8			H = 0.9		
	N = 1	N = 2	N = 3	N =1	N =2	N =3
$\hat{H}_{max}$	0.9037	0.8993	0.9122	0.9959	1.0250	1.0824
$\hat{H}_{min}$	0.6816	0.7180	0.6847	0.7799	0.8080	0.8419
$\hat{H}_{mean}$	0.7959	0.8188	0.8262	0.8992	0.9190	0.9323
$\hat{H}_{std}$	0.0389	0.0384	0.0429	0.0375	0.0387	0.04

The wavelet-based Hurst parameter estimator performs a time average of wavelet coefficients $|w_x(j, k)|^2$ at a given scale j:

$$\Gamma_x = \frac{1}{n_j} \sum_{k=1} n_j |w_x(j, k)|^2 \quad (4)$$

where n_j is the number of wavelet coefficients at scale level j. The linear relationship between $\log_2(\Gamma_x)$ and scale level j for a range of scales $[j_1, j_2]$ indicates the presence of self-similarity or long-range dependence. An estimator $\hat{H}$ for the Hurst parameter can be obtained by making a linear regression of $\log_2(\Gamma_x)$ on scale level j in the scaling range $[j_1, j_2]$:

$$\log_2(\Gamma_x) = (2H - 1)j + \hat{e} \quad (5)$$

The scaling range $[j_1, j_2]$ is an important factor affecting the accuracy of the wavelet-based H estimator. In this study, the whole interval of the scales was used for the scaling range because it appeared to be a good choice.

2) Vanishing moments: In the wavelet-based estimator, the vanishing moment N of the wavelet is an important parameter. A wavelet is said to have vanishing moments of order N if

$$\int_{-\infty}^{\infty} t^p \psi_0(t) \mathrm{d}t = 0, p = 0, ..., N - 1 \quad (6)$$

where $\psi_0(t)$ is the mother wavelet [10]. A larger N results in better estimation theoretically [7]. However, larger N also means fewer data points are available for the estimation because of the border effect [7]. In this study, different values of N were applied to simulated fGn data and Poisson data, and the statistics for $\hat{H}$ were computed as shown in Table I and Table II. From the results, it was observed that the best statistics for $\hat{H}$ were obtained (in most cases) when $N = 1$. Hence, the vanishing moment $N = 1$ was used in our wavelet-based estimator.

3) Data sample size: For the wavelet-based estimator, the number of data samples is required to be a power of 2. Table III shows the results of statistical tests of $\hat{H}$ for various lengths of fGn and Poisson data with $N = 1$. It can be seen that larger sample sizes naturally result in more accurate estimation

TABLE II

VANISHING MOMENT TEST FOR POISSON DATA.

		Vanishing moment	
		N=1	N=2
$\hat{H}_{stat}$	$\hat{H}_{max}$	0.6292	0.6222
	$\hat{H}_{min}$	0.3533	0.3618
	$\hat{H}_{mean}$	0.5006	0.4980
	$\hat{H}_{std}$	0.0359	0.04

TABLE III

ESTIMATION RESULTS FOR VARIOUS LENGTHS OF FGN (H=0.6) AND POISSON (H=0.5) DATA

	Data sample of fGn data				
	256	512	1024	2048	4096
$\hat{H}_{max}$	0.7376	0.6684	0.6836	0.6419	0.6320
$\hat{H}_{min}$	0.4760	0.5040	0.5432	0.5652	0.5712
$\hat{H}_{mean}$	0.5971	0.5986	0.6021	0.6004	0.5985
$\hat{H}_{std}$	0.0587	0.0341	0.0238	0.0156	0.0127
	Data sample of Poisson data				
	256	512	1024	2048	4096
$\hat{H}_{max}$	0.6925	0.6292	0.5861	0.5591	0.5377
$\hat{H}_{min}$	-0.3587	0.3533	0.4108	0.4494	0.4621
$\hat{H}_{mean}$	0.5003	0.5006	0.5002	0.4999	0.5001
$\hat{H}_{std}$	0.0577	0.0359	0.025	0.0165	0.0112

(smaller variances and closer mean to the true value). For accuracy, a larger data sample will always be better. However, 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. For this study, we decided on a data length of 1024 somewhat arbitrarily. In practice, the choice of sample size may be constrained by a desired level of accuracy or time for model selection.

C. Accuracy of Model Selection

1) Model selection using threshold: We investigated the use of a threshold value for $\hat{H}$ to choose between the LRD and SRD candidate models. Obviously, $0.5 < H < 1$ indicates a presence of long range dependence and $H = 0.5$ implies short range dependence. However, the estimation of H is random which can result in selection of the wrong model. To find a suitable threshold for the system, we attempted various threshold values to choose between a Poisson ($H = 0.5$) model and a fGn ($H = 0.6$) model. If $\hat{H} < H_{threshold}$, the Poisson model was selected; otherwise, the fGn model was chosen. But we found from Table I that some $\hat{H}$s of fGn (H = 0.6) and those of Poisson overlapped, so there could be a challenging case for accurate model selection. In simulation experiments, the probabilities of mis-classification were examined. For fGn data, a estimation with $\hat{H} < H_{threshold}$ was considered to be a wrong one and the probability of mis-classification was defined as the ratio of number of wrong estimations to the total number of estimations. Same definition was applied to Poisson data with wrong estimator having $\hat{H} \geq H_{threshold}$. For each threshold

TABLE IV

PROBABILITIES OF MIS-CLASSIFICATION TEST

$\hat{H}_{threshold}$	Prob. for Poisson	Prob. for fGn (H=0.6)
0.5547	0.011	0.034
0.55	0.0181	0.028
0.5470	0.026	0.026
0.5450	0.0315	0.02

value, the probabilities were measured as shown in Table IV for Poisson and fGn data. It was noted that a threshold value of $H = 0.5470$ resulted in equal probabilities for fGn and Poisson data, so we would not biase to either model. And the combined ratios turned out to be only 0.052. Thus, it seems that model selection using threshold can be fairly accurate even when the candidate models are similar.

Studying Tables I and II revealed a fact that although there was a overlap between $\hat{H}$s of fGn (H = 0.6) data and Poisson data, $\hat{H}$s of fGn data with higher H values had very few overlapping probability with Poisson data. Thus, we can expect that if fGn data has higher H value, higher threshold will result in less and even zero wrong selection probability. Then using the threshold found from fGn (h=0.6) will degrade the system performance for fGn data with higher H values. However, thus will ensure the overall system performance for fGn with different H values. So we chose this value as the system threshold.

2) Model selection when traffic is changing: In the previous experiment, the simulated traffic was purely SRD or LRD which simplifies estimation of the Hurst parameter. In more realistic circumstances, the traffic characteristics may be changing over time and the Hurst parameter estimation must be updated continuously to detect the changes. In the next simulation experiment, we generated traffic consisting of alternating SRD and LRD segments. The SRD traffic was simulated by multiplexing 10 on/off sources data. The simulated LRD traffic consisted of fGn with $H = 0.8$ instead of $H = 0.6$. The reason was that since we assumed fGn (H=0.6) as a lowest self-simlarity case in our study, using it as source traffic would have the worst system performance. So we chose fGn (H=0.8) for the following model changing simulations, and tried to focus on addressing the selection delay problem instead of studying it with the wrong model selection problem together.

Three types of traffic were generated:

1) 4096-length segments of LRD traffic alternating with 4096-length segments of SRD traffic;
2) a 1536-length segment of SRD traffic followed by a 1536-length of LRD traffic and then a 1024-length of SRD traffic;
3) a 1536-length segment of SRD data followed by a 1024-length segment of LRD traffic and then a 512-length of SRD traffic.

The Hurst parameter is estimated simply over consecutive, non-overlapping 1024-length segments of the traffic. That is, the Hurst parameter is computed over the first 1024-length segment, then computed over the next 1024-length segment, and so on. Clearly, this method is well suited for traffic that is LRD or SRD with 1024-length intervals, namely the first simulated traffic type. When there is a mixture of LRD and SRD traffic within the same 1024-length segment, estimation of H may result in the selection of the wrong traffic model.

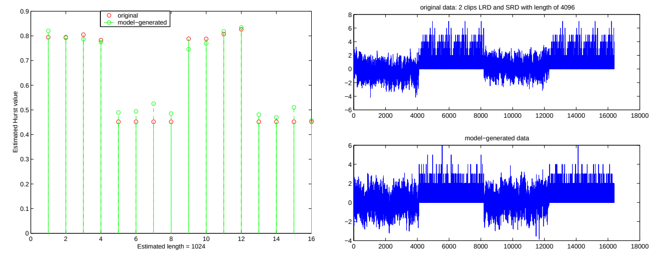
In the first simulated traffic data, the 4096-length segments of LRD or SRD traffic are multiples of 1024. Hence, estimating the Hurst parameter over 1024-length segments work well. Figure 2(a) shows that the Hurst parameter is estimated accurately, and Figure 2(b) shows that the model classification works well.

In the second simulated traffic data, the LRD and SRD segments are not exactly multiples of 1024 lengths. The first 1024-length segment consists entirely of SRD traffic, so the Hurst parameter computed over this segment is accurate, as shown in Figure 3(a). The second 1024-length segment is a combination of SRD and LRD traffic, resulting in an estimated $H = 0.67$ and classification as LRD as shown in Figure 3(b). The third 1024-length segment is entirely LRD and the fourth 1024-length segment is entirely SRD; both segments were classified correctly.

In the third simulated traffic data, the first 1024-length segment is entirely SRD which is classified correctly, as shown in Figure 4(b) and Figure 4(a). The second and third 1024-point segments consist of a mixture of LRD and SRD traffic. Both segments were classified as LRD because of the computed H values.

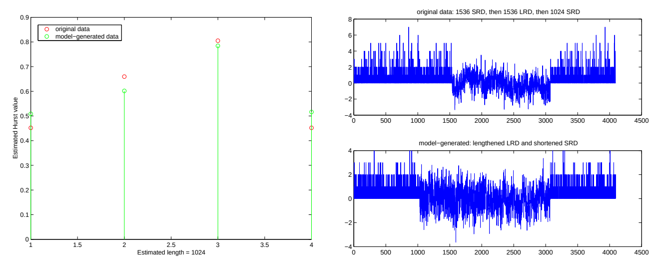
3) Improving model selection with sliding windows and exponential weights: The previous experiments showed that estimation of the Hurst parameter over non-overlapping segments could be problematic when the traffic changes do not happen to coincide with the 1024-length intervals. As an improvement, the next simulation experiments tried a sliding window; the Hurst parameter is still computed over 1024-length data segments, but this window is advanced (slide) in 64-length steps. Thus, consecutive calculations of the Hurst parameter share an interval of 960 data points in common. For another improvement, we used exponential weights to discount older data samples in the Hurst estimation.

Figures 5(a) to 7(a)) show the improvement in the Hurst parameter estimation, and Figures 5(b) to 7(b)) show the improvement in model classification. The sliding window and exponential weights can greatly improve the capability to detect changes in the traffic characteristics.



(a) Hurst value comparison (b) Data comparison

Fig. 2. Trace 1: Hurst value and data comparisons between original and model-generated traffic



(a) Hurst value comparison (b) Data comparison

Fig. 3. Trace 2: Hurst value and data comparisons between original and model-generated traffic

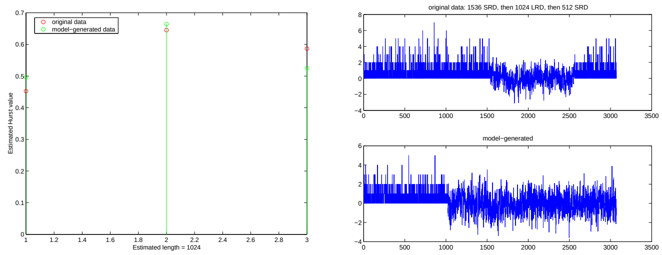
IV. Conclusions

This paper focused on a two-model system for model classification. The results show that a simple threshold can be used with good accuracy when the traffic characteristics are unchanging. When traffic characteristics are changing over time, 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). We investigated sliding windows and exponential weights to improve the model selection when traffic characteristics are dynamic. It remains for future work to more fully evaluate and improve this system.

For a library which contains more than two models, more statistics of traffic have to be measured and a more complicated method has to be developed for model selection. In the two-model experiment, we used one statistic, Hurst parameter of the traffic, to switch between candidate models. Generalizing to the N-model library, N statistics should be measured and form a N dimensional vector. This vector, denoted by $(x_1, ..., x_N)$, will be used to switch among the N models. If we divide the N-dimensional space into N disjoint regions, model j will be selected as the "best-fit" model if $(x_1, ..., x_N)$ falls into the region j. Future works include what statistics should be measured, how to construct the regions for N models, and the way to evaluate the system.

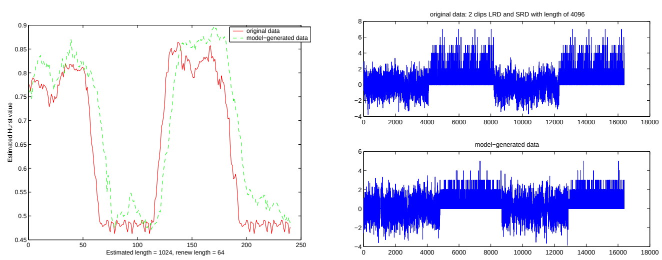
References

[1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traf-



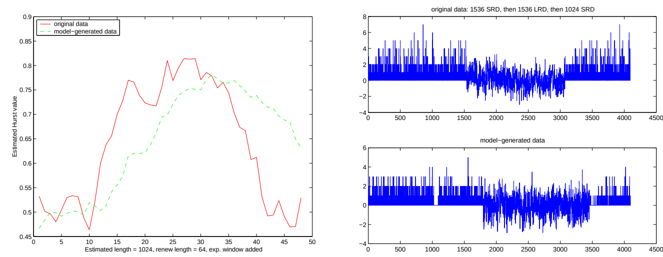
(a) Hurst value comparison (b) Data comparison

Fig. 4. Trace 3: Hurst value and data comparisons between original and model-generated traffic



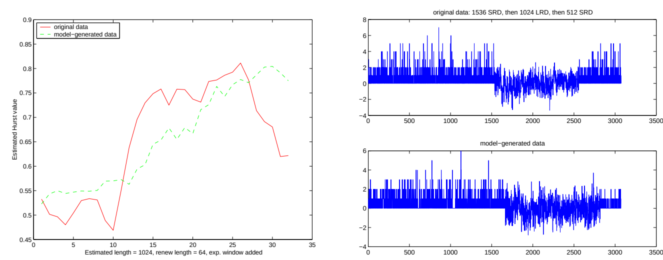
(a) Hurst value comparison (b) Data comparison

Fig. 5. Trace 1: Hurst value and data comparisons with sliding window and exponential weights



(a) Hurst value comparison (b) Data comparison

Fig. 6. Trace 2: Hurst value and data comparisons with sliding window and exponential weights



(a) Hurst value comparison (b) Data comparison

Fig. 7. Trace 3: Hurst value and data comparisons with sliding window and exponential weights

fic(extended version)," *IEEE/ACM Trans. on Networking*, no. 2, pp. 1–25, 1994.

[2] V. Paxon and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," *IEEE/ACM Trans. on 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. on Commun.*, vol. 43, 1995.

[5] M. E. Crovella and A. Bestavros, "Self similarity in WWW traffic: evidence and possible causes," *IEEE/ACM Trans. on Networking*, vol. 5, pp. 835–846, December 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. on Selected Areas in Commun.*, vol. 12, no. 3, 1994.

[7] P. Abry and D. Veitch, "Wavelet analysis of long-range-Dependent traffic," *IEEE Trans. on Inform. Theory*, vol. 44, January 1998.

[8] D. Veitch and P. Abry, " A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. on Inform. Theory*, vol. 45, pp. 878–896, April 1999.

[9] I. Daubechies, *Ten lectures on wavelets*. Philadelphia,PA: SIAM, 1992.

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