Self-similar Traffic Prediction Using Least Mean Kurtosis

Hong Zhao, Nirwan Ansari, and Yun Q. Shi
Center for Communications and Signal Processing Research
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
University Heights, Newark, NJ 07102, USA

Abstract

Recent studies of high quality, high resolution traffic measurements have revealed that network traffic appears to be statistically self similar. Contrary to the common belief, aggregating self-similar traffic streams can actually intensify rather than diminish burstiness. Thus, traffic prediction plays an important role in network management. In this paper, Least Mean Kurtosis (LMK), which uses the negated kurtosis of the error signal as the cost function, is proposed to predict the self similar traffic. Simulation results show that the prediction performance is improved greatly over the Least Mean Square (LMS) algorithm. [1]

1. Introduction

Recent studies have shown that aggregate Internet traffic does not comply to the Poisson model. It exhibits long term correlations which cannot be modeled by a Markov model. Leland *et al.* [5] analyzed the Ethernet traffic data and showed that the generally accepted argument for the "Possion-like" nature of aggregate traffic that aggregate traffic becomes smoother as the number of traffic sources increases has very little to do with reality. "Self-similar" or "fractal like" processes can describe more effectively the actual network traffic. A self similar phenomenon exhibits structural similarity across all (or at least a wide range) of time scales. For self similar traffic, there is no natural length for a burst; traffic bursts appear on a wide range of time scales. From a mathematical point of view, self similar traffic differs from other traffic models in the following way [1]. Let s be a time unit representing a time scale, such as $s = 10^m$ seconds ($m = 0, \pm1, \pm2 \ldots$). For every time scale, let $X^{(s)} = X_n^{(s)}$ denote the time series computed as the number of units (packets, bytes, cells, etc) per time unit s in the traffic stream. Traditional traffic models possess the property that as s increases, the "aggregated" process, $X^{(s)}$, tends to be a sequence of *i.i.d.* random variables (covariance stationary white noise). But for a self-similar process, they either appear visually indistinguishable from one another ("exactly self-similar") but distinctively different from pure noise, or they converge to a time series with a non-degenerate autocorrelation structure ("asymptotically self-similar"). A mathematical definition of a self-similarity process will be given in Section 2.

Two formal mathematical models that yield elegant representations of the self-similarity phenomenon are the Fractional Gaussian Noise (FGN) and the Fractional Autoregressive Integrated Moving Average (FARIMA) processes. Since the FARIMA process is much more flexible with regard to the simultaneous modeling of the short term and long term behavior of a time series than the FGN process, the FARIMA process is adopted here to simulate the network traffic.

The prediction of network traffic plays an important role in resource allocation and network management. Many types of traffic have the property of long range dependence (LRD), and aggregated Internet traffic also shows long term correlations. The higher correlation in the time domain, the longer the mean queue size, and thus the delay will be long. From the network perspective, the key point is to assign the link capacity to a given traffic in order to provide guaranteed quality of service. By prediction, we can not only achieve this but also keep the bandwidth utilization high. In this paper we propose to predict the self similar traffic by the least mean kurtosis (LMK) algorithm. This prediction is based on a higher order statistics rather than the second order statistics used in the LMS algorithm. Simulation results show that this LMK algorithm achieves better performance than LMS in predicting self similar traffic generated by the FARIMA model. The rest of the paper is organized as follows: In Section 2 self similar traffic and its widely used model are introduced; the LMK algorithm is proposed to

[1] This work has been supported in part by the NJ Commission on Higher Education via the NJI-TOWER project and New Jersey Commission on Science and Technology via the NJ Center for Wireless Telecommunications

0-7695-1916-4/03 $17.00

predict the self similar traffic. Section 3 presents the performance analysis of our proposed scheme; finally concluding remarks are included in Section 4.

2. Traffic prediction

The ability to predict traffic within a network is one of the fundamental requirements of network design and management. The prediction quality depends on the amount of uncertainty that accompanies the prediction and the nature of traffic itself. Prediction must be as accurate as possible so that bandwidth and buffer resources are not wasted and at the same time QoS can be guaranteed. The LMK algorithm is proposed to predict the self similar traffic, and is shown to outperform the LMS algorithm.

2.1. The self similar traffic model

A self-similarity time series has the property that when aggregated, the new short time series has the same autocorrelation function as the original. Each point in the short time series is the sum of multiple original points. The self-similar process is defined as follows [5]: Let $X = (X_t : t = 0, 1, 2, \cdots)$ be a covariance stationary stochastic process with mean μ, variance σ^2, and autocorrelation function $r(k), k > 0$. In particular, we assume that X has an autocorrelation function of the form

$$r(k) \sim k^{-\beta} L(t) \text{ as } k \to \infty, \tag{1}$$

where $0 < \beta < 1$ and L is slowly varying at infinity. For our discussion below, we assume for simplicity that L is asymptotically constant. For each $m = 1, 2, 3 \cdots$, let $X^{(m)} = (X_k^{(m)} : k = 1, 2, 3 \cdots)$ denote the new covariance stationary time series obtained by averaging the original series X over non-overlapping blocks of size m. That is, for each $m = 1, 2, 3, \cdots$, $X^{(m)}$ is given by

$$X_k^{(m)} = \frac{1}{m}(X_{km-m+1} + \cdots + X_{km}), k \geq 1.$$

Its corresponding autocorrelation function is $r^{(m)}$. The process is called (exactly) second-order self-similar with self-similarity parameter $H = 1 - \beta/2$, if for all $m = 1, 2, \cdots$, $var(X^{(m)}) = \sigma^2 m^{-\beta}$ and

$$r^{(m)}(k) = r(k) \text{ as } k \geq 0. \tag{2}$$

X is called (asymptotically) second-order self-similar with self-similarity parameter $H = 1 - \beta/2$ if for all k large enough,

$$r^{(m)}(k) \to r(k), \text{ as } m \to \infty, \tag{3}$$

with $r(k)$ given by (1). Intuitively, the most striking feature of (exactly or asymptotically) second-order self similar process is that their aggregated process $X^{(m)}$ possesses a non-degenerate correlation structure, as $m \to \infty$.

In practice, traffic model is often used to simulate the network traffic. An important requirement of practical traffic modeling is to generate synthetic data sequences that exhibit similar features as the measured traffic. For self-similar traffic, there are two formal mathematical models: FGN and FARIMA processes, that generate elegant representations of the self similarity phenomenon. Since the FARIMA process is more accurate in simulating network traffic than FGN, we adopt FARIMA to generate traffic used in this paper.

The FARIMA(p, d, q) process, an extension to ARIMA(p, d, q), is defined as [3]:

$$\phi(B)\nabla^d X_t = \theta(B)\epsilon_t, \tag{4}$$

where d is the indicator for the strength of long range dependence and assumes the value between 0 and $1/2$. ϵ_t is a Gaussian white noise, and

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots \theta_q B^q$$

are polynomials of degree p and q, respectively, in the backward shift operator B. The operator $\nabla^d = (1-B)^d$ can be expressed by the binomial expansion

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k, \tag{5}$$

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}, \tag{6}$$

where $\Gamma(x)$ denotes the gamma function; note that for all positive integers, only the first $d+1$ terms are non-zero in Eq. (6). The FARIMA$(0, d, 0)$ process with $0 < d < 1/2$, is stationary and long range dependence with an autocorrelation function

$$\rho_k = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \text{ as } k \to \infty. \tag{7}$$

To generate FARIMA traffic data, the FARIMA process can be approximated by the linear process in the form of [4]

$$X_k = \sum_{i=0}^{I} c_{k-i}\epsilon_i, \tag{8}$$

where ϵ_i is an $i.i.d$ random variable. ϵ_i may be Gaussian or non-Gaussian. For Gaussian FARIMA(0,d,0),

$$c_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}. \tag{9}$$

c_k can be iteratively obtained as follows

$$c_0 = 1, \tag{10}$$

$$c_{k+1} = \frac{k+d}{k+1} \cdot c_k. \tag{11}$$

$H = d + 0.5$, and thus we can generate FARIMA traffic according to the parameter H. In our simulations, the FARIMA traffic with $H = 0.8$ is used to evaluate the performance of the LMK predictor.

2.2. The LMK algorithm

Since the FARIMA model can yield elegant representations of the self similarity phenomenon, and it is more flexible than FGN with regard to the simultaneous short-term and long-term behavior of a time series, FARIMA is used to simulate the network traffic.

Let $X(n)$ be the time series traffic generated by the FARIMA model. A pth-order linear predictor has the form

$$\hat{x}(n+1) = \sum_{l=0}^{p-1} h(l)x(n-l), \tag{12}$$

$$e(n) = \hat{x}(n) - x(n) = \mathbf{h}^T\mathbf{X} - x(n), \tag{13}$$

where $\mathbf{h}$, $\mathbf{X}$ are vectors of adaptive filter coefficients and input signal, respectively. Let

$$\mathbf{h} = [h(0), h(1), \cdots h(p-1)]^T,$$

$$\mathbf{X} = [x(n-1), x(n-2), \cdots x(n-p)]^T.$$

In the LMK algorithm, the cost function is defined to be the negated kurtosis [7]:

$$\begin{aligned} J_{LMK}(\mathbf{h}) &= 3E^2[e(n)^2] - E[e^4(n)] \\ &= 3[E(\mathbf{h}^T\mathbf{X} - x(n))^2]^2 - E[\mathbf{h}^T\mathbf{X} - x(n)]^4 \end{aligned} \tag{14}$$

Taking the gradient with respect to the vector $\mathbf{H}$,

$$\begin{aligned} \nabla J_{LMK}(\mathbf{h}) &= 12E[\mathbf{h}^T\mathbf{X} - x(n)]^2 E[\mathbf{h}^T\mathbf{X} - x(n)]\mathbf{X} \\ &\quad - 4E[\mathbf{h}^T\mathbf{X} - x(n)]^3\mathbf{X} \\ &= 4\{3E[e^2(n)]E[e(n)] - 4E[e^3(n)]\}\mathbf{X}. \end{aligned} \tag{15}$$

The mean value $E[\mathbf{h}^T\mathbf{X} - x(n)]^2$ will be estimated specially by the following recursive equation:

$$G(n) = \beta G(n-1) + (1-\beta)e^2(n). \tag{16}$$

Using this estimate and the ensemble estimate of $E[\mathbf{h}^T\mathbf{X} - x(n)]$, we can get the following equation:

$$\tilde{\nabla} J_{LMK}(\mathbf{h}(n)) = 4[3G(n) - e^2(n)]e(n)\mathbf{X}. \tag{17}$$

According to the method of the steepest descent adaptive weight-update algorithm [2], LMK can be characterized by

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \frac{1}{4}\mu\{\tilde{\nabla} J_{LMK}(\mathbf{h}(n))\}, \tag{18}$$

where μ is the step size, $\tilde{\nabla} J_{LMK}(\mathbf{h}(n))$ is an approximation of the gradient vector $\nabla J_{LMK}(\mathbf{h})$, $G(n)$ is an iterative approximation of $E[\mathbf{h}^T\mathbf{X} - x(n)]$, and β is a forgetting factor that controls the memory of the error power estimator. $\beta = 0.7$ is empirically found to work well in all our simulations. Eq. (18) can further be normalized as follows:

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \frac{\mu[3G(n) - e^2(n)]e(n)\mathbf{X}}{(\mathbf{X}^T\mathbf{X})^2}. \tag{19}$$

Tanrikulu *et. al* [7] have compared the computational complexities of LMS and LMK, and found that LMS requires $O[2N + 1M, N + 1A]$ and LMK requires $O[2N + 5M, N + 3A]$ where N is the number of adaptive coefficients, M denotes multiplication, and A denotes addition. Therefore only four extra multiplications and two extra additions which are independent of N are necessary for the LMK algorithm.

3. Performance analysis and comparison

A simulation is conducted on the self-similar traffic generated by the FARIMA model. Since the self similar process has the the property of long range dependence, and the history of long range dependence process has significant impact on the present value of the process, it is natural to assume that the longer the dependence, the better the prediction. Östring and Sirisena [6] considered the prediction of long range dependent process and demonstrated that long-range dependence has only marginal value in improving prediction. It is the *short term correlation* within the structure of a self similar process rather than the long term correlation that dominates the performance of the predictors. So the Akaike information criterion (AIC) is used to choose the best order not greater than 12. The AIC associates a cost function with the order of the filter. It was found by numerous simulations that the autocorrelation of the prediction error $e(n)$ is close to that of the white noise. Thus, we use one-step, 12 order adaptive filter for both the LMK and LMS algorithm. The performance of the algorithm is quantified by the inverse Signal to Noise Ratio ($SNR^{-1} = \frac{\sum e^2(n)}{\sum x^2(n)}$). Table 1 shows the performance comparison of LMK based and LMS based predictors.

The Hurst parameter of the generated self-similar traffic is $H = 0.8$; this experiment is repeated 50 times, and the resulting ensemble averaged SNR^{-1} is plotted for LMK and LMS in Fig. 1. Note that the step sizes are chosen as

Table 1. Comparison of the SNR^{-1} performance of LMK and LMS predictors on self similar traffic

Method	SNR^{-1}										
	1	2	3	4	5	6	7	8	9	10	Average
LMK	0.0156	0.0169	0.0152	0.0161	0.0156	0.0170	0.0154	0.0202	0.0170	0.0159	0.0165
LMS	0.0398	0.0385	0.0410	0.0370	0.0424	0.0376	0.0421	0.0415	0.0349	0.0422	0.0397

$\mu_{LMS} = 0.6$, $\mu_{LMK} = 0.7$ that provide the least mismatch for each algorithm, in other words, the least SNR^{-1}. The forgetting factor in LMK is chosen to be $\beta = 0.7$ which provides less mismatch and faster convergence. In Fig. 1, it is clear that LMK not only converges as quickly as LMS but also produces significantly less prediction error; thus, LMK outperforms LMS in predicting the self similar traffic. Table 1 shows the performance comparison of SNR^{-1} for LMK and LMS. Note that LMK incurs smaller prediction error in all experiments than LMS; due to the length limit, only ten results and their average values are listed here. Thus the performance has improved greatly if we use the LMK algorithm instead of the LMS algorithm to predict the self similar traffic.

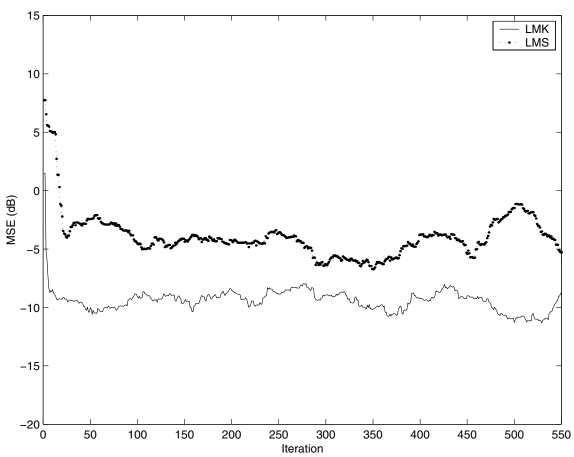

Figure 1. Averaged output SNR^{-1} versus number of iteration for self similar traffic

4. Conclusions

Aggregate Internet traffic does not comply to the Poisson model, but can be more effectively described by the self-similar process. In this paper, we propose the LMK adaptive algorithm which uses the negated kurtosis of the error signal as the cost function to predict the self-similar network traffic generated by the FARIMA model. Simulation results show that LMK incurs much smaller prediction error as compared with LMS. Since the prediction performance can be improved greatly with only a small extra computation, LMK can be used to effectively predict the real time network traffic.

References

[1] V. Frost and B. Melamed. Traffic modeling for telecommunications networks. *IEEE Communications Magazine*, 32:70–81, March 1994.

[2] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 1991.

[3] J. Hosking. Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research*, 20:1898–1908, 1984.

[4] P. Kokoszka and M.Taqqu. Parameter estimation for infinite variance fractional ARIMA. *The Annals of Statistic*, 24:1880–1993, 1996.

[5] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2:1–15, February 1994.

[6] S. Östring and H. Sirisena. The influence of long range dependence on traffic prediction. *Proc. ICC2001*, 4:1000–1005, 2001.

[7] O. Tanrikulu and A. Constantinides. Least-mean kurtosis: A novel high-order statistics based adaptive filtering algorithm. *IEE Electronics Letters*, 30:189–190, February 1994.