ITC19/ Proceedings of Poster 2307-2316
Beijing University of Posts and Telecommunications Press

A Practical Approach for Multi-Scale Performance Analysis of Internet Traffic

JI Yusheng

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, 101-8430, Japan
kei@nii.ac.jp

Abstract: Recent studies suggest that long-memory property in nowadays internet traffic shows a complicated scaling behavior: unlike the unchanged Hurst parameter of a self-similar process, scaling exponents of such processes change over time or scale ranges. In this paper, we propose a practical method, which uses the simple FBM process, to model long-range dependent traffic with changing scaling exponents over time scales. By using multiple self-similar processes to capture the statistical properties in different time scales, queueing performance of the multi-scale natured traffic can be presented by a composed performance curve produced by each self-similar process. Simulation results using traffic data measured from a backbone network is used to show the efficiency of our method.
Keywords: long-range dependence, self-similarity, multi-scale, fractional Brownian motion, traffic modeling

1. INTRODUCTION

Long-range dependence have been found in various kinds of network traffic, such as in local and wide area networks [4][7][11][12]. In order to model such kind of traffic, self-similar processes are introduced instead of traditional Poisson-based models. Recent studies also suggest that such long-memory property in nowadays internet traffic shows a more complicated scaling behavior: unlike the unchanged Hurst parameter of a exactly self-similar process, scaling exponents of such processes change over time or time scale ranges[1][5]. In order to capture these natures of network traffic, multifractal and other stochastic models are proposed [8][14][15]. As these models try to describe the real traffic feature more precisely, however, it results in an increased complexity of analysis. In [6] we introduced a method to model long-range dependent traffic by manually dividing the second-order property of the traffic into different parts, and approximating each part by a second-order self-similar process. In order to reduce the influence of possible errors produced by manual operations, a more elegant and systematic approach is required.

In this paper, we propose a practical and semi-parametric method, which uses the simple FBM processes, to model long-range dependent traffic with changing scaling exponents over time scales. By using multiple self-similar processes to capture the statistical properties in different time scales, queueing performance of the multi-scale natured traffic can be presented by a composed performance curve produced by each self-similar process. Simulation results using real traffic data measured from a backbone network show that our method is efficient for aggregated network traffic which fits Gaussian property.

2. STATISTICAL BEHAVIOR OF INTERNET TRAFFIC

We use sample traces monitored from the Science Information Network (SINET) [18], which is a nation-wide internet backbone run by the National Institute of Informatics (NII), and connects more than 700 universities and institutions in Japan. Two example traces used here are obtained in 2001, during the daily peak hours in the afternoon. Each data trace is lasted for 10,000 seconds, during which the long-term average rate is kept almost constant. At the time the data was collected, ATM switches and IP routers were used to connect the OC-3 backbone links. By monitoring the backbone traffic on ATM switches, raw data was obtained and assembled into time stamps and lengths for each IP packet including IP header. In the following statistical analysis, the basic time units in each trace are chosen to be 1 millisecond to ensure that the time stamps are accurate, and that the characteristics during short time intervals are not lost.

2.1. Variance-Time Plot

For a discrete time process $X(t)$ with stationary increments, if the average of $X(t)$ for consecutive number of m is $X^{(m)}(k) = \frac{1}{m}\sum_{t=(k-1)m+1}^{km} X(t), \quad k = 1, 2, \cdots,$ then the second-order self-similarity can be expressed as

$$\mathrm{Var}[X^{(m)}(k)] = \sigma^2 m^{-2(1-H)}, \tag{1}$$

where σ^2 is the variance of $X(t)$, and H is the Hurst parameter. Second-order self-similarity is a statistical property that the correlation structure is preserved under time aggregation.

Variance-time plots can be used to verify the second-order self-similarity or long-range dependence and estimate the Hurst parameter by observing the variance function of $X^{(m)}(k)$ versus time scale m in a log-log scaled coordinates. Let $X^{(m)}(k)$ be the number of bits arrived during each consecutive time period of length $m\Delta t$, where Δt is the basic time unit, then the variance-time plot shows the variance of $X^{(m)}(k)$ against time scale m.

Fig.1 shows the variance-time plots of two sample traces. For both traces, the decaying rates of variance are not constant along time scale m. They show a kinked variance-time property, like it is reported in [13], with two scaling regimes: When m is small, the rate of decay for variance is high, and is rather close to the exponential process; As m increases, the rate of decay for variance reduces, showing a stronger dependence. In Fig. 1, it is evident that long-range dependence exists in the traffic. Furthermore, the dependent features are more complex than exactly second-order self-similar.

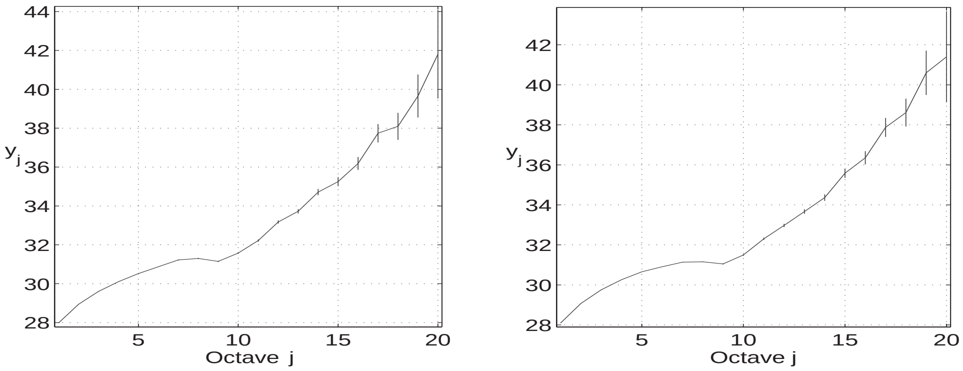

Fig. 2. The second-order log-scale diagrams of sample traces (Left: May 7, 2001; Right: May 9, 2001)

3. MULTI-SCALE PERFORMANCE ANALYSIS

As a simple and ideal model, fractional Brownian motion (FBM) or its incremental process fractional Gaussian noise (FGN) have been used to describe the long-range dependence in network traffic. An FBM process is a Gaussian centered process with stationary increments. Differing from a regular Brownian motion, the increments of an FBM are dependent, and an FBM process has exact self-similarity.

For an FBM process with average arrival rate λ and variance coefficient a, the number of arrivals up to time t can be expressed as

$$A(t) = \lambda t + \sqrt{\lambda a} Z_H(t), \quad t \in \mathbb{R}, \tag{3}$$

where $Z_H(t)$ is a normalized FBM variable, with zero-mean, a variance given by $\mathrm{Var}[Z_H(t)]=|t|^{2H}$, and a Hurst parameter $H\in[0.5, 1)$. From its Gaussian property, an FBM process is completely described by the triple $\{\lambda, a, H\}$.

The complementary distribution of queue length for a queue with service rate C fed by an FBM process $\{\lambda, a, H\}$ is given by [10]

$$\Pr(Q > x) \sim e^{-\gamma x^{2(1-H)}} \quad as \quad x \to \infty, \tag{4}$$

where $\gamma \equiv \frac{1}{2a\lambda(1-H)^2}\left\{\frac{(C-\lambda)(1-H)}{H}\right\}^{2H}$. This is obtained using the following lower bound

$$\begin{aligned} \Pr(Q > x) &\geq \max_{t\geq 0} \Pr(A(t) > Ct + x) \\ &= \max_{t\geq 0} \bar{\Phi}\left(\frac{(C-\lambda)t + x}{\sqrt{a\lambda} t^H}\right), \end{aligned} \tag{5}$$

```
input C, λ;                                    // service and average rates
input nn;                                      // square root of the time scale range
input m0;                                      // the basic time scale in units of second
For each i {
    m1 = m(i) / nn;                            // smallest time scale of the i'th FBM sub-process
    m2 = m(i) * nn;                            // largest time scale of the i'th FBM sub-process
    read v(m1), v(m2);                         // variance values at these time scales
    β = {log(v(m1)) - log(v(m2))} / {log(m2) - log(m1)};
    h = 1 – β / 2;                                                      // see (6)
    a = (power(10, log(v(m1)) + β * log(m1/m0))) / (λ * power(m0, 2h));  // see (7)
    x = m(i) * (C - λ) * (1 – h) / h;                                   // see (8)
    γ = ComputeGamma(h, a, C, λ);                                       // using (4)
    p = ComputeProb(x, β, γ);                                           // using (4)
}
```

Fig. 3. The pseudo code for performance calculation

With known triples $\{\lambda, a_i, H_i\}$ for each FBM sub-process, the last thing we need to know for obtaining the performance using (4) is the relationship between time scale m and queue length x. These two parameters can be connected by the notion which is called the "critical time scale" in [16] and the "relevant time scale" in [9]. If x is the queue length in a infinite buffer-sized queueing system with FBM input, the critical time scale for the queueing performance of an FBM process $\{\lambda, a, H\}$ can be expressed as

$$m = \frac{x}{C-\lambda}\frac{H}{1-H}. \tag{8}$$

Note that the critical time scale m expressed by (8) is exactly the length of time t when the lower bound of the RHS of the equality of (5) is achieved.

From m, H_i, and the excess bandwidth C-λ, the corresponding queue length x can be calculated by (8). Consequently, the queueing performance in whole time scale range can be obtained from (4). Fig. 3 shows the pseudo code for the performance evaluation algorithm described above.

For a kinked variance-time property, the Hurst parameter changes significantly around the kink. In (6), the interval between the two time scale values m_1^i and m_2^i, i.e., the time scale range of each FBM sub-process (corresponding to nn^2 in the pseudo code) is used to determine the Hurst parameter H_i of i-th FBM sub-process. Note that in log-scaled coordinates, the horizontal distance between two abscissa values (say m_1 and m_2) is the proportion (m_2/m_1) of these two values.

The time scale range will control largely the calculated value of Hurst parameter around the kink, while be almost unrelated in the area where second-order log-scale diagram or variance-time plot is closed to a straight line. If the difference between the Hurst parameters of two neighboring FBM sub-processes is large, two values of x calculated using (8) from two neighboring values of m will have reversed order against the order of m's. Let us denote the

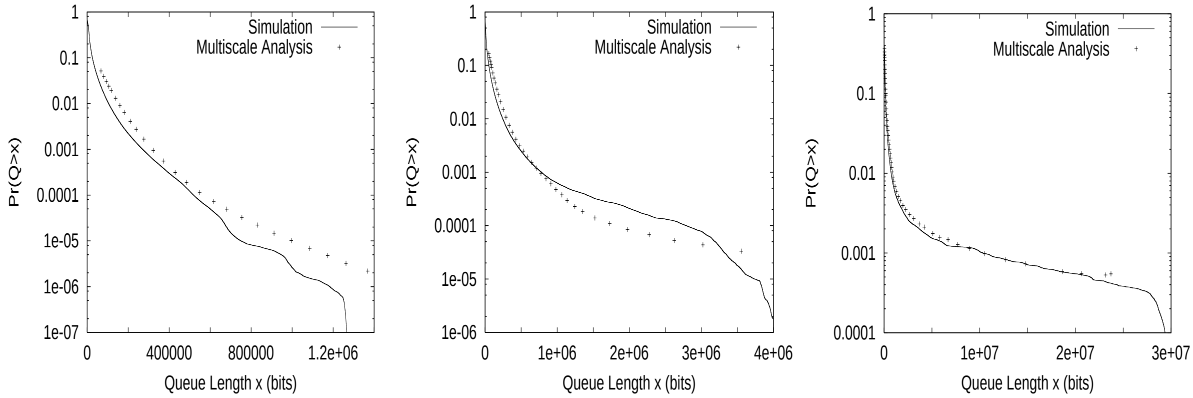

Fig. 4. Complementary distributions of queue length of trace in May 7 (From left to righte: Load = 0.7, 0.75, 0.8)

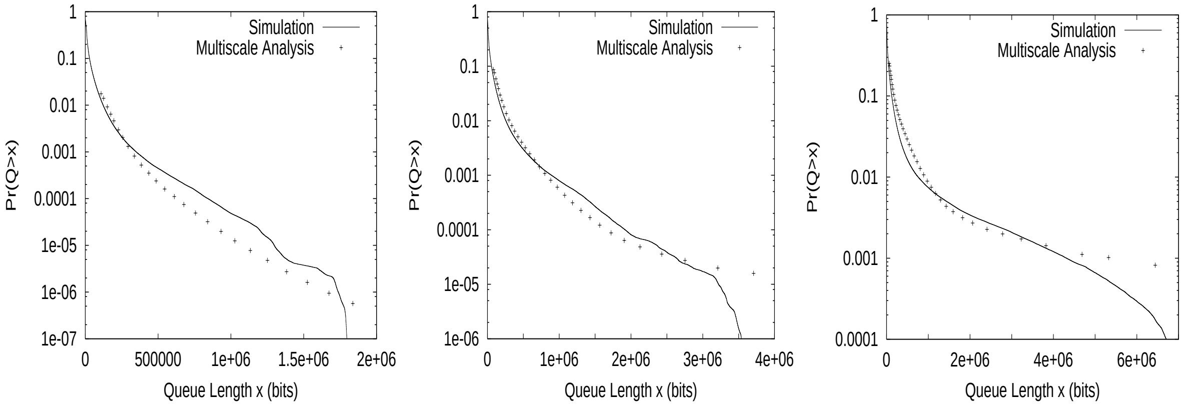

Fig. 5. Complementary distributions of queue length of trace in May 9 (From left to right: Load = 0.7, 0.75, 0.8)

4.2 The Impact of Time Scale Range on Performance

To show the impact of the value of time scale range used for calculating Hurst parameters of each FBM sub-process, Fig. 6 captures the differences of performance obtained by different time scale ranges using example of May 9 data with an offered load of 0.75. Two other performance curves obtained from two FBM processes corresponding to the two straight line areas in the variance-time plots in Fig.1 (Right) with h_1 = 0.69 and h_2 = 0.932 are also shown in the figure.

The performance of FBM process with h_1 = 0.69 can only provide the objective performance in small time scale area, and the other one with h_2 = 0.932 represents the performance only in large time scale. Since the cross point of two performance curves of these two FBM processes corresponds to the time scale of the kink point of the variance-time plot, we can notice that the differences of performance produced by different values of time scale range mainly occur at around the kink point. A sufficiently large time scale range can prevent an inversed order of x's as the case when time scale range equals to 1.2 in the figure, and produce a more smoothened performance curve along the whole range of queue length.

2. J.-M. Bardet and P. Bertrand, “Definition, properties and wavelet analysis of multiscale fractional Brownian motion,” Preprint, 2003.
3. J.-M. Bardet and P. Bertrand, “Detecting abrupt change on the Hurst parameter of a multi-scale fractional Brownian motion with applications,” Int. Conf. on Self-Similarity and Applications, Clermonts-Ferrand, France, 2002.
4. A. Erramilli, O. Narayan and W. Willinger, “Experimental queueing analysis with long-range dependent packet traffic”, IEEE/ACM Trans. Networking, vol.4, no.2, pp.209-223, 1996.
5. A. Feldmann, A. Gilbert, W. Willinger and T. Kurtz, “The Changing Nature of Network Traffic: Scaling Phenomena,” ACM Computer Communication Review, Vol.28, No.2, pp.5-29, 1998.
6. Y. Ji, T. Fujino, S. Abe, J. Matsukata and S. Asano, “On the Impact of Time Scales on Tail Behavior of Long-Range Dependent Internet Traffic,” Proc. IEEE ICON, pp.13-18, 2003.
7. W. Leland, M. Taqqu, W. Willinger and D. Wilson, “On the self-similar nature of Ethernet traffic (extended version),” IEEE/ACM Trans. Networking, vol.2, no.1, pp.1-15, 1994.
8. C. A. Melo and N. L. Fonseca, “An Envelope Process for Multifractal Traffic Modeling, Proc. IEEE ICC, 2004.
9. A. Neidhardt and J. Wang, “The concept of relevant time scales and its application to queuing analysis of self-similar traffic” Proc. ACM SIGMETRICS’98, pp.222-232, 1998.
10. I. Norros, “A storage model with self-similar input,” Queueing Systems, vol.16, pp.387-396, 1994.
11. I. Norros, “On the use of fractional Brownian motion in the theory of connectionless network,” IEEE JSAC, vol.13, no.6, pp.953-962, 1995.
12. V. Paxson and S. Floyd, “Wide area traffic: the failure of Poisson modeling,” IEEE/ACM Trans. Networking, vol.3, no.3, pp.226-244, 1995.
13. V. J. Ribeiro, “Queuing Analysis of Long-Range Dependent Traffic,” ELEC 599 Project Report, Department of Electrical and Computer Engineering, Rice University, 1999.
14. R. Riedi, M. Crouse, V. Ribeiro and R. Baraniuk, “A multifractal wavelet model with application to network traffic,” IEEE Trans. Info. Theory, vol.45, no.3, pp.992-1019, 1999.
15. S. Roux, D. Weitch, P. Abry, L. Huang, J. Micheel and P. Flandrin, “Statistical Scaling Analysis of TCP/IP Data Using Cascades,” Proc. ICASSP, 2001.
16. B. K. Ryu and A. Elwalid, “The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities,” ACM Comp. Comm. Review, 26(4), pp.3-14, 1996.
17. D. Veitch and P. Abry, “A Wavelet-Based Joint Estimator of the Parameters of Long-Range Dependence,” IEEE Trans. Info. Theory, Vol. 45, No. 3, 1999.
18. http://www.sinet.ad.jp/english/