An Empirical Autocorrelation Form for Modeling LRD Traffic Series

Ming Li[1], Jingao Liu[1], and Dongyang Long[2]

[1] School of Information Science & Technology, East China Normal University
Shanghai 200026, PR China
{mli, jgliu}@ee.ecnu.edu.cn
[2] Department of Computer Science, Zhongshan University, Guangzhou 510275, PR China
{issldy}@zsu.edu.cn

Abstract. Paxson and Floyd (*IEEE/ACM T. Netw.* 1995) remarked the limitation of fractional Gaussian noise (FGN)) in accurately modeling LRD network traffic series. Beran (1994) suggested developing a sufficient class of parametric correlation form for modeling whole correlation structure of LRD series. M. Li (*Electr. Letts*., 2000) gave an empirical correlation form. This paper[1] extends Li's previous letter by analyzing it in Hilbert space and showing its flexibility in data modeling by comparing it with FGN (a commonly used traffic model). The verifications with real traffic suggest that the discussed correlation structure can be used to flexibly model LRD traffic series.

1 Introduction

Modeling long-range dependent (LRD) series has been widely studied, see e.g., [1] ~ [6], where exactly self-similar (ess) process (i.e., fractional Gaussian noise (FGN)) is a commonly used tool, e.g., [1] [2] [5] [7]. However, in communication networks, autocorrelation function (ACF) form of ess processes is too narrow for accurately modeling actual series [8]. On the other hand, accurate models of actual series are at the heart of some applications. For instance, accurate models of actual traffic series are crucial to performance evaluation of communication networks [9]. In addition, ACF has impact on queuing systems [10]. Motivated by those, we extend Li's early work [6] for an empirically derived 3-parameter ACF form in Section 2. Verifications of this ACF form are given in Section 3 and conclusions in Section 4.

1 This paper is in part sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, PRC.

H. Jin et al. (Eds.): NPC 2004, LNCS 3222, pp. 399-402, 2004.

2 Empirical 3-Parameter Correlation Form

Let x be an LRD time series and r be its ACF. Then, $r(\tau) \sim c\tau^{-\beta}$ ($\tau \to \infty$), where $c > 0$ is a constant, $0 < \beta < 1$. Then, we aims at finding a function $R(\tau)$ to best fit $r(\tau)$.

Generally, an ACF is nonlinear. Thus, modeling a measured ACF can be regarded as an issue of nonlinear least squares fitting. If a model is characterized by several parameters, nonlinear least squares in multi-dimensions may result in a set of nonlinear equations. Since a set of nonlinear equations may have no (real) solutions [11], it is needed to prove the existence of solutions. As numerical solutions for the root finding of a set of nonlinear equations are in the sense of approximation, a criterion is needed to evaluate the quality of curve fitting.

Denote a measured traffic trace as $x(t_i)$, indicating the number of bytes in a packet at t_i, $i \in I_0$ ($= 0, 1, 2, \cdots$). Let r be the measured ACF of x, R be the modeling of r and $M^2(R) = E[(R - r)^2]$ be the mean square error. Then, $M^2(R)$ is used to evaluate the quality of curve fitting. In our scheme, $M^2(R) < 10^{-4}$ was required. Hence, our method to model r is to find R that fits r with the constrain of $M^2(R) < 10^{-4}$.

Let the error $e = R - r$. Construct the functional below

$$f(e) = \sqrt{\sum_k \left[R(k) - r(k)\right]^2}. \tag{1}$$

Based on the experiments, we present the following normalized correlation form

$$R(k) = (|k| + 1)^{-a} + Lu(|k| - m),\ a \geq 0,\ 1 \geq L \geq 0,\ m = 1, 2, \cdots, k \in I, \tag{2}$$

where u is the unit step function. Consequently, $f(e)$ stands for a 3-D cost function

$$J(a, L, m) \triangleq f(e). \tag{3}$$

Due to the evenness of ACF, we only consider $k \geq 0$ in what follows. An approximated root (a_0, L_0, m_0) of $J = 0$ can be determined by iteration based on nonlinear least squares fitting for a given r. The existence of solutions is explained below.

In fact, a measured traffic trace is of finite length. Without losing the generality, the maximum possible length of x is assumed as $p \in I_0$. Let $N \in I_0$ and $N >> p$. Then, N may be regarded as an "infinite" in the engineering sense. Denote

$\|r\| = \sqrt{<r, r>} = \sqrt{\sum_0^{N-1} |r|^2}$. Then, $l_N^2 = \{r; \|r\| < \infty\}$ is a Hilbert space [5]. Denote

$\mathcal{A}_1 = \{R; R(k) = c[(k+1)^{-a} + Lu(k - m)]\}$. Then, $\mathcal{A}_1 \subset l_N^2$.

Statement. Let $r \in l_N^2$ be a measured autocorrelation sequence. There exists a unique element $R \in \mathcal{A}_1$ such that $\|r - R\| = \inf_{s \in \mathcal{A}_1} \|r - s\|$, where $\|r - R\| = f(e)$.

Proof: l_N^2 is an obvious convex set and $f(e)$ is a convex functional defined on l_N^2. Therefore, the extremum of $f(e)$ exists. Thus, Statement follows.

According to Statement, for a given $r \in l_N^2$, if $R = R(k;\, a_0, L_0, m_0)$ is such that $\|r - R\| = \inf_{s \in \mathcal{A}_1} \|r - s\|$, (a_0, L_0, m_0) is called approximated root of $J = 0$ and $(a_0, L_0, m_0) = \arg\min J(a, L, m)$. Thus, if $M^2(R) < 10^{-4}$, $R(k;\, a_0, L_0, m_0)$ is acceptable in our scheme. In the paper, (a_0, L_0, m_0) is obtained by Levenberg-Marquardt method [11].

3 Verifications

Four well known real-traffic traces (dec-pkt-1, dec-pkt-2, dec-pkt-3 and dec-pkt-4) are analyzed. Denote $R(k)$ as $R(k;\, a, L, m)$. Then, the cost function for dec-pkt-1 is given by $J(a, L, m) = \|R(k;\, a, L, m) - r_{pkt1}(k)\|$, where $r_{pkt1}(k)$ is the measured ACF of dec-pkt-1. By Levenberg-Marquardt method, one obtains $(a_0, L_0, m_0) = (2.091, 0.377, 1)$. At this point, $M^2(R) = 1.952 \times 10^{-5}$. Therefore, $r_{pkt1}(k)$ is modeled by

$$R(k) = (k+1)^{-2.091} + 0.377u(k-1). \tag{4}$$

Fig. 1 indicates dec-pkt-1 and Fig.2 the fitting the data for modeling $r_{pkt1}(k)$.

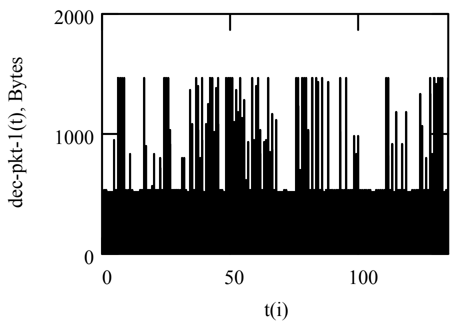

Fig. 1. TCP trace of dec-pkt-1

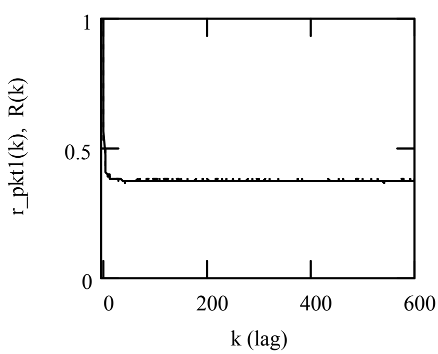

Fig. 2. The result of fitting the data: …… measured ACF; — modeled ACF

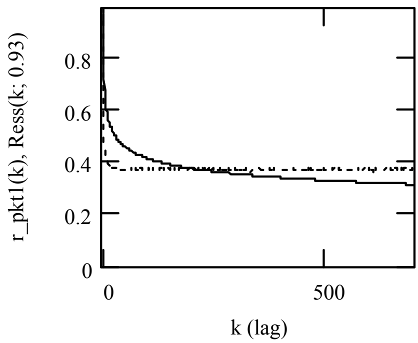

Fig. 3. Fitting $r_{pkt1}(k)$ with ess model: …… measured ACF; — modeled ACF with FGN

Similarly, we have (2.088, 0.402, 1), (3.14, 0.341, 1), and (3.14, 0.341, 1) for $r_{pkt2}(k)$, $r_{pkt3}(k)$ and $r_{pkt4}(k)$, respectively.

To evaluate the benefit of model (2), we use FGN to fit $r_{pkt1}(k)$. The normalized ACF of FGN is given by $R_{ess}(k, H) = 0.5[(k + 1)^{2H} - 2k^{2H} + (k - 1)^{2H}]$. By using least squares fitting, we have the result $R_{ess}(k; 0.93)$ with $M^2(R_{ess}) = 0.003$. As $R_{ess}(k; 0.93)$ is the best result in the ess sense, the benefit of our model is obvious, see Fig. 3.

4 Conclusions

A correlation form for modeling LRD traffic series has been given. The verifications show that it has a noteworthy flexibility to model LRD traffic and satisfactorily fits the real traffic investigated. This model has an advantage over models based on single parameter such as that of ess model. Because the modeled ACFs are non-summable, the long-range dependence of traffic has also been verified in this way.

References

1. B. B. Mandelbrot, *Gaussian Self-Affinity and Fractals*, Springer, 2001.
2. J. Beran, *Statistics for Long-Memory Processes*. Chapman & Hall, 1994.
3. M. Li and C.-H. Chi, *Journal of the Franklin Institute*, 340 (6/7), 2003, 503-514.
4. V. Paxson and S. Floyd, *IEEE/ACM T. on Networking*, 3 (3), June 1995, 226-244.
5. M. Li, W. Zhao, and et al., *Applied Mathematical Modelling*, 27 (3), 2003, 155-168.
6. M. Li, and et al, *Electronics Letters*, 36 (19), 2000, 1168-1169.
7. V. Paxson, *Computer Communication Review*, 27 (5), 1997, 5-18.
8. B. Tsybakov and N. D. Georganas, *IEEE T. on Inform. Theory*, 44 (5), 1998, 1713-1725.
9. A. Adas, *IEEE Communications Magazine*, 35 (7), 1997, 82-89.
10. S. Q. Li and C. L. Hwang, *IEEE/ACM T. on Networking*, 1 (6), Dec. 1993, 678-692.
11. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd Ed., Cambridge University Press, 1992.