INTERNATIONAL JOURNAL OF COMPUTERS
Issue 3, Volume 1, 2007

Error Order of Magnitude for Modeling Autocorrelation Function of Interarrival Times of Network Traffic Using Fractional Gaussian Noise

Ming Li

***Abstract*—** Fractional Gaussian noise (FGN) is a commonly used model of computer network traffic time series with long-range dependence (LRD). It has been realized that FGN may not be enough to accurately model real traffic. However, quantitative evidence about this is seldom reported. To this end, this paper gives quantitative descriptions, based on processing real traffic, on the error order of magnitude for modeling autocorrelation functions of interarrival times of four types of traffic, namely, TCP, UDP, IP, OTHER, using FGN. The present results exhibit that modeling accuracy, which is expressed by mean square error, by using FGN is usually in the order of magnitude of 10^{-3}. The main reason to cause error by using FGN model is that FGN might not satisfactorily fit the short-term lags of real traffic.

***Keywords*—** Time series; Fractional Gaussian noise; Long-range dependence; Interarrival time series of network traffic; Curve fitting.

I. Introduction

Time series with long-range dependence (LRD) has been widely studied in many fields of sciences and engineering, including network traffic (traffic for short), see e.g. [1, 2] and references therein. Fractional Gaussian noise (FGN) is a commonly used model of traffic, see e.g. [3-11].

Computer scientists have noticed that FGN may not be enough for accurately modeling autocorrelation function (ACF) of real traffic, see e.g. [10,11,12,13], but quantitative description with respect to the error order of magnitude of the curve fitting of ACF using FGN is rarely seen. In addition, the cause of error resulted from the curve fitting of ACF modeling using FGN is seldom reported. In this paper, we shall give quantitative descriptions of the error order of magnitude of the curve fitting of ACF modeling using FGN. Besides, we shall point out that the main cause of the error of the curve fitting is that FGN might not satisfactorily fit the short-term lags of real traffic. The traffic series investigated in this research is interarrival time series in computer communications networks.

Let $x[t(i)]$ be a traffic time series, indicating the number of bytes in the ith packet at the time $t(i)$, where $i \in I$ (= 0, 1, 2, …). We call $t(i)$ timestamp series, implying the timestamp of the ith packet. Let the increment series of $t(i)$ be

$$s(i) = t(i + 1) - t(i). \tag{1}$$

Then, $s(i)$ is called interarrival time series of $x[t(i)]$.

Note 1. $x[t(i)]$ and $s(i)$ are non-negative while $t(i)$ is non-decreasing. □

Experimental processing of actual traffic exhibits that x is of LRD. The letter [14] briefs the LRD property of $s(i)$. This paper experimentally studies four types of traffic of 28 series. They are TCP traffic, UDP traffic, IP traffic and OTHER one, where TCP means Transmission Control Protocol, UDP implies User Datagram Protocol, IP stands for Internet Protocol, and OTHER traffic represents non-TCP, non-UDP, non-encapsulated traffic.

In the rest of paper, Section 2 describes the preliminaries. Experimental investigations are given in Section 3. Section 4 concludes the paper.

II. Preliminaries

A. FBM and FGN

Let $B(t)$, $t \in (0, \infty)$, be Brownian motion. Let $B_H(t)$ be fractional Brownian motion (FBM) with $H \in (0, 1)$. Let $\Gamma(\cdot)$ be Gamma function. Then,

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H+1/2)}\left\{\begin{array}{l}\int_{-\infty}^{0}[(t-u)^{H-0.5}-(-u)^{H-0.5}]dB(u)+ \\ \int_{0}^{t}(t-u)^{H-0.5}dB(u)\end{array}\right\}. \tag{2}$$

Let the increment series of $B_H(t)$ be

$$G(t) = B_H(t+a) - B_H(t), \tag{3}$$

where a is a real number. Then, $G(t)$ is called FGN. Denote

$$\sigma^2 = (H\pi)^{-1}\Gamma(1-2H)\cos(H\pi)$$

the intensity of FGN [15]. Then, the ACF of FGN for $\tau > 0$ is given by

Manuscript received May 2, 2007; Revised version received August 10, 2007. This work was supported in part by the National Natural Science Foundation of China under the project grant number 60573125.

Ming Li is with the School of Information Science & Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, PR. China. (Tel.: +86-21-5434 5193; fax: +86-21-5434 5119; e-mails: ming_lihk@yahoo.com, mli@ee.ecnu.edu.cn,

URL: http://www.ee.ecnu.edu.cn/teachers/mli/js_lm(Eng).htm).

$$\rho(\tau) = \frac{\sigma^2}{2}[(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}]. \quad (4)$$

The normalized ACF of FGN is given by

$$R(\tau) = \frac{1}{2}[(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}]. \quad (5)$$

Below, we use $R(k)$ (k is integer) to indicate the ACF of FGN in the discrete case.

Note 2. A series is of LRD if its ACF is non-summable and it is of short-range dependence (SRD) if its ACF is summable [2]. □

Note 3. FGN for $H \in (0.5, 1)$ is of LRD while it is of SRD for $H \in (0, 0.5)$. □

B. H Estimation

The parameter H plays a vital role in time series with LRD. There are various methods for H estimation, see e.g., [1, 2]. This paper uses the method introduced in [8].

A series measured in practice is of finite length. In fact, it is of finite length when numeric computation is involved. Let R be the ACF of a traffic series $s(i)$ with LRD. Then, for $H \in (0.5, 1)$,

$$R(k) = \frac{E\{[s(i+k)-\mu][s(i)-\mu]\}}{\sigma^2} \sim ck^{2H-2} \, (k \to \infty),$$

where $c > 0$ is a constant, $\mu = E(s)$, where E is the mean operator. Without losing generality, the maximum possible length of R is assumed to be N. Define the norm of R as the inner product given by

$$\|r\| = \sqrt{<r, r>} = \sqrt{\sum_{k=0}^{N-1} |r|^2}, N \in I. \quad (6)$$

Then, the following

$$l_N^2 = \left\{ r; \sqrt{\sum_{k=0}^{N-1} |r|^2} < \infty \right\} \quad (7)$$

is a Hilbert space [8,16].

Define the set $\mathcal{E}$ as

$$\mathcal{E} = \{R;\ R(k) = 0.5[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}], H \in (0.5,\ 1)\},$$
$$(k = 0, 1, \ldots, N-1), \quad (8)$$

Then,

$$\mathcal{E} \subseteq l_N^2. \quad (9)$$

According to the theorem of existence of a unique minimizing element in Hilbert space [8,16], for an ACF of a real series $r \in l_N^2$, there exists a unique $R \in \mathcal{E}$ such that

$$\|r - R\| = \inf_{a \in \mathcal{E}} \|r - a\|. \quad (10)$$

Let

$$J(H) = \frac{1}{N}\sum_k [r(k;\ H) - R(k;\ H)]^2. \quad (11)$$

Then, minimizing $J(H)$ yields an estimate

$$H_0 = \arg\min J(H). \quad (12)$$

The value of $J(H_0)$ is the minimum mean square error, which is denoted by $M^2(R) = E[(r - R)^2]$.

The research thought is stated like this. By investigating the $M^2(R)$s of 28 real traces, we may experimentally observe the error order of magnitude by the curve fitting of ACF modeling. Moreover, we may explore the possible cause that produces errors by using FGN model.

III. Experimental Investigations

A. Real Data Used

Real data used in this paper consist of 28 series. They are 6 series of TCP traffic (Table 1), 10 series of UDP traffic (Tables 2 and 3), 6 of IP traffic (Table 4), and 6 of OTHER traffic (Table 5). The series with the prefix DEC were measured at Digital Equipment Corporation, those with Lbl were recoded at the Lawrence Berkeley Laboratory, and the series with NUS were collected at the National Univeristy of Singpaore. In Tables 1-5, the first column stands for series name, the second for record date, and the third for series length. We denote $R(k)$ by $R(k;\ H)$ for facilitating the illustrations in what follows.

B. Demonstrations

Demonstration with DEC-pkt-1.TCP: The series $x[t(i)]$ of DEC-pkt-1.TCP is indicated in Fig. 1 (a) and timestamp series $t(i)$ is in Fig. 1 (b). The interarrival series $s(i)$ is in Fig. 2. The measured ACF of $s(i)$ is shown in Fig. 3 (a). Minimizing $J(H)$ yields H_0 = 0.923 with $M^2(R)$ = 2.264×10^{-3}. Therefore, the modeled ACF $R(k)$ of $s(i)$ of DEC-pkt-1.TCP using FGN is indicated in Fig. 3 (b). Fig. 3(c) shows the fitting the data. By eye, one sees that FGN does not satisfactorily fits the ACF of $s(i)$ of DEC-pkt-1.TCP for short-term lags.

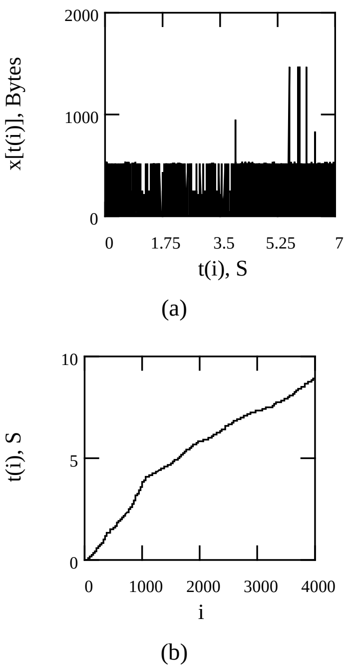


Fig. 1. Real traffic series DEC-pkt-1.TCP. (a). series in packet size $x[t(i)]$. (b) Timestamp series $t(i)$.

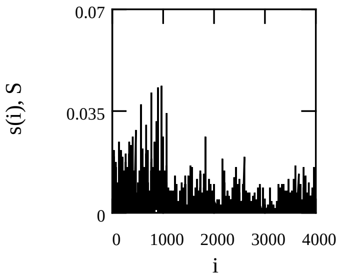
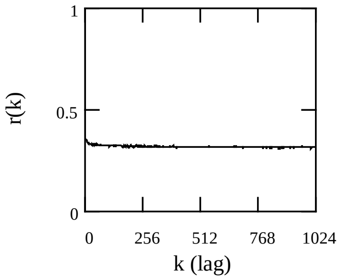
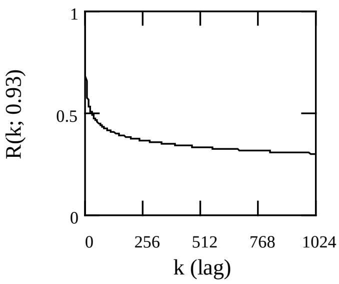


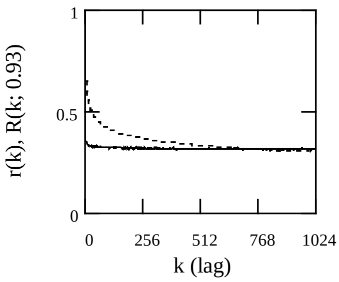
Fig. 2. Interarrival series $s(i)$ of DEC-pkt-1.TCP.



(a)



(b)



(c) —— $r(k)$, ······ $R(k)$.

Fig. 3. Modeling procedure. (a). $r(k)$: Measured ACF of $s(i)$ of DEC-pkt-1.TCP. (b). $R(k)$: Modeled ACF based on FGN. (c). Fitting the data.

Demonstration with DEC-pkt-1.UDP: Real series $t(i)$ for DEC-pkt-1.UDP is shown in Fig. 4 and $s(i)$ in Fig. 5, resepctively. The measured ACF of $s(i)$ is shown in Fig. 6 (a). Minimizing J yields $H_0 = 0.945$ with $M^2(R) = 6.09\times10^{-3}$. Fig. 6 (b) indicates the modeled ACF using FGN and Fig. 6 (c) shows the fitting the data of ACF of $s(i)$ of DEC-pkt-1.UDP based on FGN.

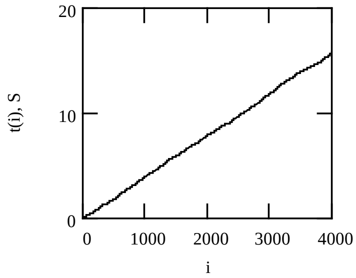


Fig. 4. Real series $t(i)$ for DEC-pkt-1.UDP.

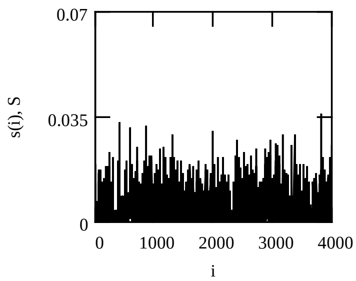
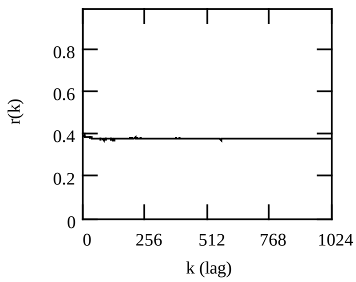
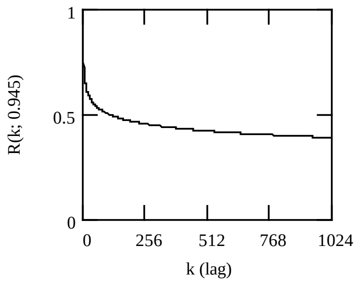


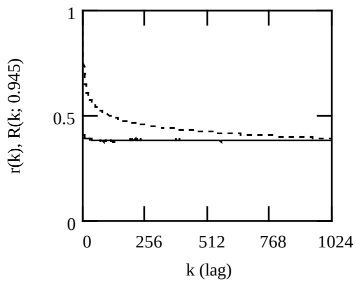
Fig. 5. Real series $s(i)$ for DEC-pkt-1.UDP.



(a)



(b)



(c) —— $r(k)$, ······ $R(k)$.

Fig. 6. Modeling procedure. (a). Measured ACF of $s(i)$ of DEC-pkt-1.UDP. (b). $R(k)$: Modeled ACF based on FGN. (c). Fitting the data.

Demonstration with DEC-pkt-1.IP: Timestamp series for DEC-pkt-1.IP is plotted in Fig. 7 and $s(i)$ in Fig. 8. The measured ACF of $s(i)$ is in Fig. 9 (a). Minimizing J yields H_0 = 0.958 with $M^2(R) = 4.133 \times 10^{-3}$. Fig. 9 (b) indicates the modeled ACF $R(k)$ of $s(i)$ of DEC-pkt-1.IP using FGN and Fig. 9 (c) fitting the data.

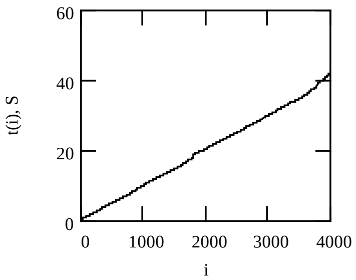


Fig. 7. Real series $t(i)$ for DEC-pkt-1.IP.

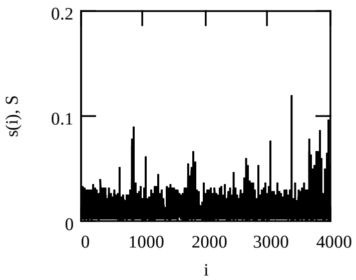
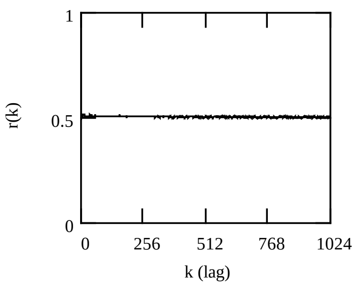
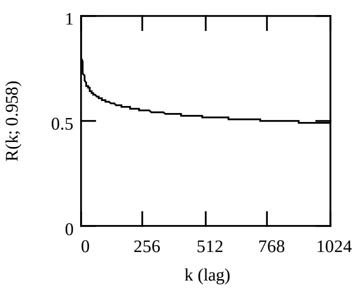


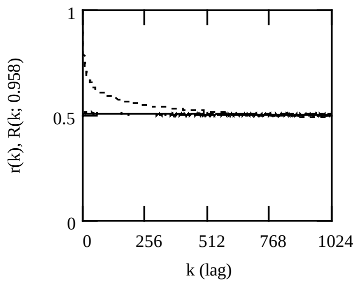
Fig. 8. Real series $s(i)$ for DEC-pkt-1.IP.



(a)



(b)



(c) —— $r(k)$, ⋯⋯ $R(k)$.

Fig. 9. Modeling procedure. (a). Measured ACF of $s(i)$ of DEC-pkt-1.IP. (b). $R(k)$: Modeled ACF based on FGN. (c). Fitting the data.

Demonstration with DEC-pkt-1.OTHER: The series $t(i)$ of DEC-pkt-1.OTHER is indicated in Fig. 10 and $s(i)$ in Fig. 11. The measured ACF of $s(i)$ is in Fig. 12 (a). Minimizing J yields H_0 = 0.937 with $M^2(R) = 5.038 \times 10^{-3}$. Fig. 12 (b) indicates the modeled ACF model of $s(i)$ of DEC-pkt-1.OTHER using FGN. Fig. 12 (c) gives the fitting the data.

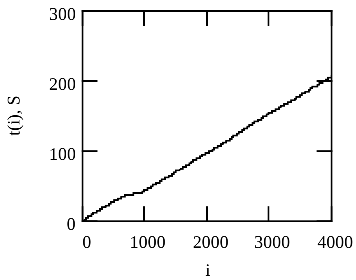


Fig. 10. Real series $t(i)$ for DEC-pkt-1.OTHER.

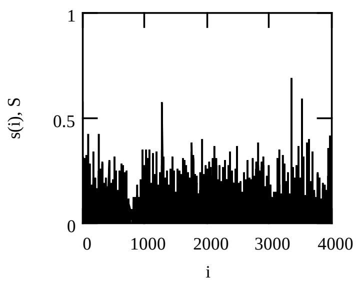
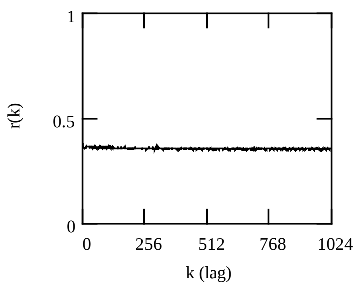
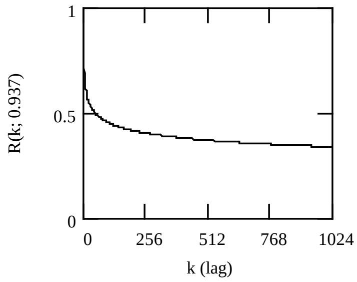


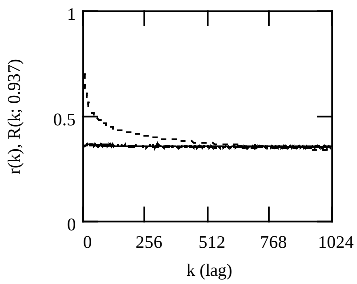
Fig. 11. Real series $s(i)$ for DEC-pkt-1.OTHER.



(a)



(b)



(c) —— $r(k)$, ⋯⋯ $R(k)$.

Fig. 12. Modeling procedure. (a). Measured ACF of $s(i)$ of DEC-pkt-1.OTHER. (b). $R(k)$: Modeled ACF based on FGN. (c). Fitting the data.

C. *Summary*

We summarize the experimental results for all 28 series in the columns 5-6 in Tables 1-4, where the fourth column stands for H estimate and the fifth for $M^2(R)$.

Table 1. Six real series of TCP traffic.

Series name	Record date	Series length	H_0	$M^2(R)$
DEC-pkt-1.TCP	08Mar95	3.3×10^6	0.923	2.264×10^{-3}
DEC-pkt-2.TCP	09Mar95	3.9×10^6	0.920	2.282×10^{-3}
DEC-pkt-3.TCP	09Mar95	4.3×10^6	0.925	2.270×10^{-3}
DEC-pkt-4.TCP	09Mar95	5.7×10^6	0.922	2.320×10^{-3}
Lbl-pkt-4.TCP	21Jan94	862946	0.930	2.208×10^{-3}
Lbl-pkt-5.TCP	28Jan94	710614	0.925	2.220×10^{-3}

Table 2. Four real series of UDP traffic.

Series name	Record date	Series length	H_0	$M^2(R)$
NUS-1.UDP	24Mar03	1×10^6	0.920	3.654×10^{-3}
NUS-2.UDP	24Mar03	1×10^6	0.915	3.469×10^{-3}
NUS-3.UDP	26Mar03	1×10^6	0.915	3.415×10^{-3}
NUS-4.UDP	26Mar03	1×10^6	0.920	3.534×10^{-3}

Table 3. Six real series of UDP traffic.

Series name	Record date	Series length	H_0	$M^2(R)$
DEC-pkt-1.UDP	08Mar95	829759	0.935	7.729×10^{-3}
DEC-pkt-2.UDP	09Mar95	805802	0.935	2.881×10^{-3}
DEC-pkt-3.UDP	09Mar95	1035457	0.935	2.883×10^{-3}
DEC-pkt-4.UDP	09Mar95	1187454	0.935	2.886×10^{-3}
Lbl-pkt-4.UDP	21Jan94	33744	0.904	2.886×10^{-3}
Lbl-pkt-5.UDP	28Jan94	69358	0.875	2.182×10^{-3}

Table 4. Six real series of IP traffic.

Series name	Record date	Series length	H_0	$M^2(R)$
DEC-pkt-1.IP	08Mar95	225237	0.955	2.416×10^{-3}
DEC-pkt-2.IP	09Mar95	335556	0.938	2.884×10^{-3}
DEC-pkt-3.IP	09Mar95	325833	0.900	2.517×10^{-3}
DEC-pkt-4.IP	09Mar95	511287	0.935	2.624×10^{-3}
Lbl-pkt-4.IP	21Jan94	303055	0.890	4.264×10^{-3}
Lbl-pkt-5.IP	28Jan94	195241	0.890	4.312×10^{-3}

Table 5. Six real series of OTHER traffic.

Series name	Record date	Series length	H_0	$M^2(R)$
DEC-pkt-1.OTHER	08Mar95	74135	0.931	2.893×10^{-3}
DEC-pkt-2.OTHER	09Mar95	78021	0.931	3.040×10^{-3}
DEC-pkt-3.OTHER	09Mar95	105410	0.931	2.874×10^{-3}
DEC-pkt-4.OTHER	09Mar95	92361	0.931	2.662×10^{-3}
Lbl-pkt-4.OTHER	21Jan94	121140	0.878	1.105×10^{-3}
Lbl-pkt-5.OTHER	28Jan94	401231	0.890	2.012×10^{-3}

IV. Conclusion

The results in Tables 1-5 suggest that $s(i)$ of traffic (TCP, UDP, IP, OTHER) is of LRD and the modeling accuracy of ACF based on FGN is in the order of magnitude of 10^{-3}. The plots in Fig. 3 (c), Fig. 6 (c), Fig. 9 (c), and Fig. 12 (c) imply that FGN may not satisfactorily fit the short-term lags of those traffic data. This might likely be the main error source with respect to the curve fitting of ACF modeling based on FGN.

References

[1] B. B. Mandelbrot, *Gaussian Self-Affinity and Fractals*, Springer, 2001.
[2] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, 1994.
[3] Ian W. C. Lee and A. O. Fapojuwo, "Stochastic processes for computer network traffic modeling," *Comput Commun*, Vol. 29, No. 1, Dec. 2005, pp. 1-23.
[4] A. Adas, "Traffic models in broadband networks," *IEEE Communications Magazine*, Vol. 35, No. 7, 1997, pp. 82-89.
[5] H. Michiel and K. Laevens, "Teletraffic engineering in a broad-band era," *Proc. the IEEE*, Vol. 85, No. 12, Dec. 1997, pp. 2007-2033.

[6] M. Li, "An approach to reliably identifying signs of ddos flood attacks based on lrd traffic pattern recognition," *Computers & Security*, Vol. 23, No. 7, 2004, 549-558.
[7] M. Li and C.-H. Chi, "A correlation based computational model for synthesizing long-range dependent data," *Journal of the Franklin Institute*, Vol. 340, No. 6-7, Sep.-Nov. 2003, pp. 503-514.
[8] M. Li, and et al., "Modeling autocorrelation functions of self-similar teletraffic in communication networks based on optimal approximation in Hilbert space," *Applied Mathematical Modelling*, Vol. 27, No. 3, 2003, pp. 155-168.
[9] J. Beran, R. Shernan, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable bit-rate video traffic," *IEEE Trans. Commun*, Vol. 43, 1995, pp. 1566-1579.
[10] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poison modeling," *IEEE Trans. Networking*, Vol. 3, No. 3, June 1995, pp. 226-244.
[11] B. Tsybakov and N. D. Georganas, "Self-similar processes in communications networks," *IEEE Trans. Information Theory*, Vol. 44, No. 5, Sep. 1998, pp. 1713-1725.
[12] M. Li, "Modeling autocorrelation functions of long-range dependent teletraffic series based on optimal approximation in Hilbert space-a further study," *Applied Mathematical Modelling*, Vol. 31, No. 3, Mar. 2007, pp. 625-631.
[13] M. Li, S. C. Lim, and W. Zhao, Long-Range Dependent Network Traffic: A View from Generalized Cauchy Process, in *Progress in Applied Mathematical Modeling*, edited by F. Yang, et al., Nova Science Publishers, USA, 321-338, 2008.
[14] M. Li, W. Jia and W. Zhao, Correlation form of timestamp increment sequences of self-similar traffic on ethernet, *Electronics Letters*, Vol. 36, No. 19, 2000, pp. 1168-1169.
[15] M. Li and S. C. Lim, "A rigorous derivation of power spectrum of fractional Gaussian noise," *Fluctuation and Noise Letter*, Vol. 6, No. 4, Dec. 2006, pp. C33-C36.
[16] J. P. Aubin, *Applied Functional Analysis*, 2nd Edition, John Wiley & Sons, 2000.

Ming Li was born in 1955 in Wuxi, China. He completed his undergraduate program in electronics engineering at Tsinghua University. He received the M.S. degree in ship structural mechanics from China Ship Scientific Research Center and Ph.D. degree in computer science from City University of Hong Kong, respectively. From 1990 to 1995, he was a researcher in CSSRC. From 1995-1999, he was with the Automation Department, Wuxi University of Light Industry. From 2002 to 2004, he was with the School of Computing, National University of Singapore. In 2004, he joined East China Normal University (ECNU) as a professor in both electronics engineering and computer science. He is currently a Division Head for Communications & Information Systems at ECNU. His research areas relate to applied statistics, computer science, measurement & control. He has published over 70 papers in international journals and international conferences in those areas. Li is a member of IEEE.