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 of the curve fitting of ACF using FGN. Besides, we shall describe with respect to the error order of magnitude of the cause of error resulted from the curve fitting of ACF modeling using FGN is rarely seen. In addition, the description with respect to the error order of magnitude of the traffic. The traffic series investigated in this research is interarrival time series in computer communications networks.

Let \( x(t) \) be a traffic time series, indicating the number of bytes in the \( i \)th packet at the time \( t(i) \), where \( i \in I = \{0, 1, 2, \ldots\} \). We call \( t(i) \) timestamp series, implying the timestamp of the \( i \)th packet. Let the increment series of \( t(i) \) be

\[
\Delta t(i) = t(i+1) - t(i).
\]

Then, \( \Delta t(i) \) is called interarrival time series of \( x(t(i)) \).

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.

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,

\[
B_H(t) - B_H(0) = \Gamma(H + 1/2) \left\{ \int_0^t [(t-u)^H - (-u)^H] dB(u) + \int_0^t (-u)^{H-0.5} dB(u) \right\}. \tag{2}
\]

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_lmh(Eng).htm).
\[
\rho(\tau) = \frac{\sigma^2}{2} [(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}].
\]

(4)

The normalized ACF of FGN is given by
\[
R(\tau) = \frac{1}{2} [(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}].
\]

(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 \(H \in (0, 0.5)\).

### 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 recorded at the Lawrence Berkeley Laboratory, and the series with NUS were collected at the National University of Singapore. 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 \(x[t(i)]\) is shown in Fig. 1 (a). Minimizing \(J(H)\) yields \(H_0 = 0.923\) with \(M^2(R) = 2.264 \times 10^{-3}\). Therefore, the modeled ACF \(R(k)\) of \(x[t(i)]\) of DEC-pkt-1.TCP using FGN is indicated in Fig. 1 (b). Fig. 3(c) shows the fitting the data. By eye, one sees that FGN does not satisfactorily fits the ACF of \(x[t(i)]\) of DEC-pkt-1.TCP for short-term lags.
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, respectively. The measured ACF of \( s(i) \) is shown in Fig. 6 (a). Minimizing \( J \) yields \( H_0 = 0.945 \) with \( M^2(R) = 6.09 \times 10^{-3} \). 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. 2. Interarrival series \( s(i) \) of DEC-pkt-1.TCP.

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

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

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

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.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.

\[ J \]

\[ H_0 = 0.958 \]

\[ M^2(R) = 4.133 \times 10^{-3} \]

\[ R(k) \]

\[ s(i), S \]

\[ t(i), S \]

\[ r(k), R(k) \]

\[ r(k), \ldots, R(k) \]

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.
estimate and the fifth for columns 5-6 in Tables 1-4, where the fourth column stands for NUS-4.UDP 26Mar03 1
NUS-2.UDP 24Mar03 1
NUS-1.UDP 24Mar03 1
NUS-4.UDP 26Mar03 1×10^6 0.920 3.534×10^{-1}

Table 3. Six real series of UDP traffic.

<table>
<thead>
<tr>
<th>Series name</th>
<th>Record date</th>
<th>Series length</th>
<th>H₀</th>
<th>M(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEC-pkt-1.UDP</td>
<td>08Mar95</td>
<td>829759</td>
<td>0.935</td>
<td>7.729×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-2.UDP</td>
<td>09Mar95</td>
<td>805802</td>
<td>0.935</td>
<td>2.881×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-3.UDP</td>
<td>09Mar95</td>
<td>1035457</td>
<td>0.935</td>
<td>2.883×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-4.UDP</td>
<td>09Mar95</td>
<td>1187454</td>
<td>0.935</td>
<td>2.886×10^{-3}</td>
</tr>
<tr>
<td>Lbl-pkt-4.UDP</td>
<td>21Jan94</td>
<td>33744</td>
<td>0.904</td>
<td>2.886×10^{-3}</td>
</tr>
<tr>
<td>Lbl-pkt-5.UDP</td>
<td>28Jan94</td>
<td>69358</td>
<td>0.875</td>
<td>2.182×10^{-3}</td>
</tr>
</tbody>
</table>

Table 4. Six real series of IP traffic.

<table>
<thead>
<tr>
<th>Series name</th>
<th>Record date</th>
<th>Series length</th>
<th>H₀</th>
<th>M(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEC-pkt-1.IP</td>
<td>08Mar95</td>
<td>225237</td>
<td>0.955</td>
<td>2.416×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-2.IP</td>
<td>09Mar95</td>
<td>335556</td>
<td>0.938</td>
<td>2.884×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-3.IP</td>
<td>09Mar95</td>
<td>325833</td>
<td>0.900</td>
<td>2.517×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-4.IP</td>
<td>09Mar95</td>
<td>511287</td>
<td>0.935</td>
<td>2.624×10^{-3}</td>
</tr>
<tr>
<td>Lbl-pkt-4.IP</td>
<td>21Jan94</td>
<td>303055</td>
<td>0.890</td>
<td>4.264×10^{-3}</td>
</tr>
<tr>
<td>Lbl-pkt-5.IP</td>
<td>28Jan94</td>
<td>195241</td>
<td>0.890</td>
<td>4.312×10^{-3}</td>
</tr>
</tbody>
</table>

Table 5. Six real series of OTHER traffic.

<table>
<thead>
<tr>
<th>Series name</th>
<th>Record date</th>
<th>Series length</th>
<th>H₀</th>
<th>M(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEC-pkt-1.OTHER</td>
<td>08Mar95</td>
<td>74135</td>
<td>0.931</td>
<td>2.893×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-2.OTHER</td>
<td>09Mar95</td>
<td>78021</td>
<td>0.931</td>
<td>3.040×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-3.OTHER</td>
<td>09Mar95</td>
<td>105410</td>
<td>0.931</td>
<td>2.874×10^{-3}</td>
</tr>
<tr>
<td>DEC-pkt-4.OTHER</td>
<td>09Mar95</td>
<td>92361</td>
<td>0.931</td>
<td>2.662×10^{-3}</td>
</tr>
<tr>
<td>Lbl-pkt-4.OTHER</td>
<td>21Jan94</td>
<td>121140</td>
<td>0.878</td>
<td>1.105×10^{-3}</td>
</tr>
<tr>
<td>Lbl-pkt-5.OTHER</td>
<td>28Jan94</td>
<td>401231</td>
<td>0.890</td>
<td>2.012×10^{-3}</td>
</tr>
</tbody>
</table>

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 plots. The curve fitting of ACF modeling based on FGN may likely be the main error source with respect to the curve fitting of ACF modeling based on FGN.

REFERENCES


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.