A Novel Description of Multifractal Phenomenon of Network Traffic Based on Generalized Cauchy Process

Ming Li[1], S.C. Lim[2], and Huamin Feng[3]

[1] School of Information Science & Technology, East China Normal University,
Shanghai 200062, P.R. China
`mli@ee.ecnu.edu.cn, ming_lihk@yahoo.com`
[2] Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Selanger, Malaysia
`sclim@mmu.edu.my`
[3] Key Laboratory of Security and Secrecy of Information, Beijing Electronic Science and
Technology Institute, Beijing 100070, P.R. China
`fenghm@besti.edu.cn`

Abstract. Let $D(n)$ and $H(n)$ be the fractal dimension and the Hurst parameter of traffic in the nth interval, respectively. Thus, this paper gives the experimental variance analysis of $D(n)$ and $H(n)$ of network traffic based on the generalized Cauchy (GC) process on an interval-by-interval basis. We experimentally infer that traffic has the phenomenon $\text{Var}[D(n)] > \text{Var}[H(n)]$. This suggests a new way to describe the multifractal phenomenon of traffic. That is, traffic has local high-variability and global robustness. Verifications of that inequality are demonstrated with real traffic.

Keywords: Generalized Cauchy process, Fractal dimension, Hurst parameter, Multifractals, Network traffic, Time series, Variance analysis.

1 Introduction

Network traffic (traffic for short) has multifractal behavior (Taqqu, Teverovsky and Willinger [1]). Multifractional Brownian motion (mBm) is a way to describe the multifractal of a time series by extending the Hurst parameter H to a time-dependent function $H(t)$ on a point-by-point basis, see e.g. Peltier and Levy-Vehel [2], Lim and Muniandy [3]. From a view of networking, however, time-varying H is usually expressed by $H(n)$, where n is the index of the nth interval (see e.g. Willinger, Paxson, Riedi and Taqqu [4], Li [5]), since servers in computer networks usually serve arriving traffic on an interval-by-interval basis. In this paper, we investigate the variances of $H(n)$ and $D(n)$ of real traffic, where $D(n)$ is the fractal dimension of traffic in the nth interval.

A single parameter model, such as the standard Fractional Gaussian noise (fGn for short), is a widely used tool in traffic modeling, see e.g. [6-9,31,35], where H is related to D by the linear expression $D = 2 - H$. However, Tsybakov and Georganas [8, Paragraph 1, Section II] noticed that "the class of exactly self-similar processes (i.e., fGn) is too narrow for modeling actual network traffic." In addition, Paxson and

Floyd [9, Last sentence, Paragraph 4, § 7.4] remarked that "it might be difficult to characterize the correlations over the entire trace with a single Hurst parameter."

Recently, Gneiting and Schlather [12] introduced stationary Gaussian processes indexed by two parameters. A simple one is called the Cauchy class, since the processes in that class can be regarded as an extension of the generalized Cauchy process used in geostatistics [11]. For simplicity, we call a process belonging to such a class as the generalized Cauchy (GC) process as stated in [14]. Li and Lim applied it to traffic modeling in the mono-fractal sense [13] and discussed its properties in [14]. A key point of the GC process is that it has the functionality to separately characterize D and H of a time series with LRD, where H is a measure of LRD while D is a measure of roughness or local irregularity ([12] and Mandelbrot [15,16]). In passing, we note that the linear relation $D = 2 - H$, resulting from fGn, implies that local properties are reflected in global ones for fGn, as can be seen from Mandelbrot [10], [12].

Experimental processing of traffic reveals that traffic is robust at large time scaling [9] but highly irregular at small time scaling (Feldmann, Gilbert, Willinger and Kurtz [17], Willinger, Govindan, Jamin, Paxson and Shenker [18]). This may be a tough issue to explain in traffic analysis from the point of view of a single parameter model such as fGn. Intuitively, such multifractal phenomena of traffic imply $\text{Var}[D(n)] > \text{Var}[H(n)]$, because a high value of $\text{Var}[D(n)]$ represents local high-variability while a small value of $\text{Var}[H(n)]$ implies global robustness. Nevertheless, variance analysis based on fGn results in, due to the relation $D = 2 - H$, $\text{Var}[D(n)] = \text{Var}[H(n)]$, which conflicts with the traffic's multifractal phenomenon experimentally observed in [1,9,17,18].

Now, suppose that H and D of traffic can be decoupled. Then, we can separately characterize $\text{Var}[D(n)]$ and $\text{Var}[H(n)]$ for the goal of investigating multifractal phenomenon of traffic. Recently, we introduced GC to model traffic [13], which provides a useful tool to achieve that goal. It is our belief that this paper is the first attempt at: 1) doing the variance analysis of $D(n)$ and $H(n)$ of traffic based on GC model, and 2) experimentally inferring that traffic has the phenomenon $\text{Var}[D(n)] > \text{Var}[H(n)]$, which may serve as a novel description of the properties that traffic has local high-variability and is globally robust

The rest of this paper is organized as follows. The GC process is briefed in Section 2. The variance analysis of $D(n)$ and $H(n)$ of traffic is discussed in Section 3, which is followed by conclusions.

2   Brief of GC Process

$X(t)$ is called the GC process if it is a stationary Gaussian centered process with the autocorrelation given by $C(\tau) = E\left[X(t+\tau)X(t)\right] = \left(1 + |\tau|^{\alpha}\right)^{-\beta/\alpha}$, where $0 < \alpha \le 2$ and $\beta > 0$ [12-14]. Since $C(\tau)$ is an even function. The following considers $\tau \ge 0$ unless otherwise stated:

$$C(\tau) = (1 + \tau^{\alpha})^{-\beta/\alpha}, \quad \tau \ge 0, \tag{1}$$

where $C(\tau)$ is positive-definite for the above ranges of α and β, and it is a completely monotone for $0 < \alpha \leq 1$, $\beta > 0$. When $\alpha = \beta = 2$, one gets the usual Cauchy process.

Recall that a self-similar process $X(t)$ with self-similarity index κ requires for $r > 0$,

$$X(rt) \triangleq r^{\kappa} X(t), \tag{2}$$

where $\triangleq$ denotes equality in joint finite distribution. The above equation describes a global property. It is known that a stationary Gaussian random process $X(t)$ that is not an exactly self-similar process may satisfy a weaker self-similar property known as local self-similarity. Taking into account the definition of local self-similarity by Kent and Wood [19], we say that a Gaussian stationary process is locally self-similar of order α if $C(\tau)$ satisfies for $\tau \to 0$,

$$C(\tau) = 1 - \beta' \mid \tau \mid^{\alpha} \{1 + O(\mid \tau \mid^{\gamma})\}, \quad \gamma > 0.$$

For the GC process, $\beta' = \dfrac{\beta}{\alpha}$ and $\gamma = \alpha$. The fractal dimension D of a locally self-similar process of order α is given by (see e.g., [19] and Adler [20])

$$D = 2 - \frac{\alpha}{2}. \tag{3}$$

Note that the local irregularities of the sample paths are measured by α, which can be regarded as the fractal index of the process (Davies and Hall [21], Hall [22], Constantine and Hall [23], Hall and Roy [24], Chan, Hall, and Poskitt [25]). Thus, the behavior of $C(\tau)$ at the origin to a great extent determines the roughness of the random process.

The large time lag behavior of the correlation is given by the following hyperbolically decaying correlation

$$C(\tau) \sim \tau^{-\beta}, \quad \tau \to \infty, \tag{4}$$

which implies LRD for $0 < \beta < 1$. The process becomes short-range dependent if $\beta > 1$. Thus, the index β characterizes LRD. Comparing the large asymptotic value of the correlation $\tau^{-\beta}$ with τ^{2H-2} yields

$$H = 1 - \beta/2. \tag{5}$$

Therefore, the LRD condition $0 < \beta < 1$ implies $0.5 < H < 1$.

One thing worth noting is that the two parameters α and β can vary independently. Hence, D of the GC process is independent of H. The separate characterization of D and H gives the GC model the flexibility that is lacking in single parameter models like fGn.

At the end of this section, we note that the GC process is non-Markovian since its correlation $C(t_1, t_2)$ does not satisfy the triangular relation given by

$$C(t_1, t_3) = C(t_1, t_2)C(t_2, t_3)/C(t_2, t_2), \quad t_1 < t_2 < t_3,$$

which is a necessary condition for a Gaussian process to be Markovian (Todorovic [26]). In fact, up to a multiplicative constant, the Ornstein-Uhlenbeck process is the only stationary Gaussian Markov process (Lim and Muniandy [27], Wolperta and Taqqu [28]).

3   Experimental Variance Analysis

On an interval-by-interval basis, we write

$$C(\tau; n) = (1 + \tau^{\alpha(n)})^{-\beta(n)/\alpha(n)}, \quad \tau \geq 0. \tag{6}$$

Hence, we have

$$D(n) = 2 - \frac{\alpha(n)}{2}, \tag{7}$$

$$H(n) = 1 - \frac{\beta(n)}{2}. \tag{8}$$

Consider three sets of test data. The first set consists of four traces measured at the Bellcore (BC) in 1989. The second includes traces recorded at Digital Equipment Corporation (DEC) in March 1995. The third contains 4 traces collected by the National Laboratory for Applied Network Research (NLANR) in November 2005. These data are available freely from [29,30]. They have been used in the research of traffic analysis, see e.g. [4,7,9,13,17,31,32]. The third is relatively recent, in comparison with the other two.

Now, we use three test series for demonstrations. They are DEC-PKT-1, BC-Oct89Ext, and AMP-1131669938-1.psize. Denote $x[t(i)]$ a traffic series, indicating the number of bytes in a packet at time $t(i)$ ($i = 0, 1, 2, \cdots$). Then, $x(i)$ is a series, representing the number of bytes in the ith packet. Figs. 1 (a) ~ (c) indicate the first 1024 points of 3 series, namely DEC-PKT-1 at DEC in 1995, BC-Oct89Ext at BC in 1989, and AMP-1131669938-1.psize at NLANR in 2005, respectively.
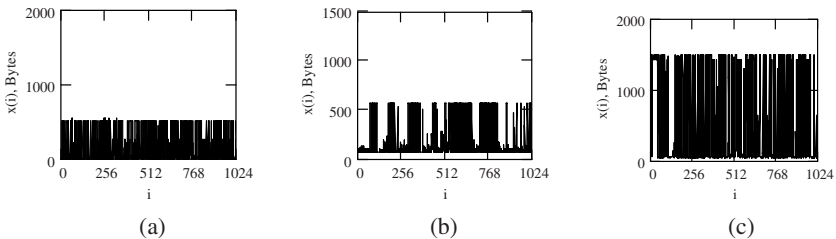

(a)   (b)   (c)

Fig. 1. Real traffic. (a). DEC-PKT-1. (b). BC-Oct89Ext. (c). AMP-1131669938-1.psize.

The computation settings are as follows. Block size L is 1024 and average times N = 10. Sectioning the series as $x(i)$, $i = (n - 1)(L \times N)$, …, $n(L \times N)$ for $n = 1$, …, 30. By computing 30 correlations for each series and using the least square fitting, we

indicate $H(n)$ of those in Fig. 1 by Fig. 2 (a), (b), and (c). Since $H(n)$ appears random, we need investigating its variation statistically. Figs. 3 (a) ~ (c) give their histograms (see [36] for the computation of the histogram of a series). By numeric computation, we have $Var[H(n)] = 0.073$ for DEC-PKT-1, $Var[H(n)] = 0.074$ for BC-Oct89Ext, and $Var[H(n)] = 0.073$ for AMP-1131669938-1.psize.
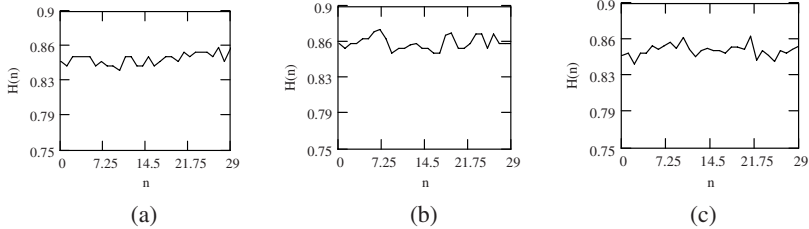

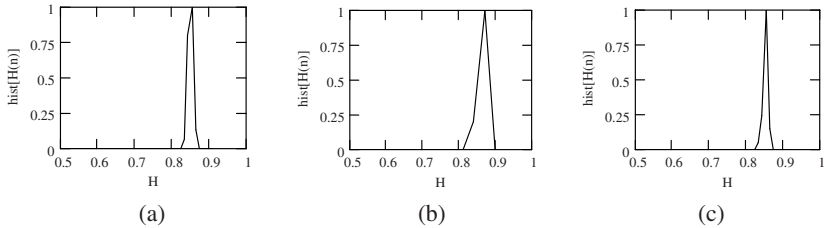**Fig. 2.** $H(n)$. (a). For DEC-PKT-1. (b). For BC-Oct89Ext. (c). For AMP-1131669938-1.psize.

Fig. 3. Histograms of $H(n)$. (a). For DEC-PKT-1: $Var[H(n)] = 0.073$. (b). For BC-Oct89Ext: $Var[H(n)] = 0.074$. (c). For AMP-1131669938-1.psize: $Var[H(n)] = 0.073$.

On the other hand, $D(n)$ of each test series is indicated in Fig. 4 (a), (b), and (c), respectively. Fig. 4 exhibits that $D(n)$ also appears random. Thus, we use the histogram to observe its fluctuation. Figs. 5 (a) ~ (c) show their histograms. According to numeric computation, therefore, we obtain $Var[D(n)] = 0.298$ for DEC-PKT-1, $Var[D(n)] = 0.297$ for BC-Oct89Ext, and $Var[D(n)] = 0.304$ for AMP-1131669938-1.psize. Taking into account $Var[H(n)]$ and $Var[D(n)]$ mentioned above, we have

$$Var[D(n)] = 0.298 > Var[H(n)] = 0.073 \text{ for DEC-PKT-1}, \tag{9}$$

$$Var[D(n)] = 0.299 > Var[H(n)] = 0.074 \text{ for BC-Oct89Ext}, \tag{10}$$

$$Var[D(n)] = 0.304 > Var[H(n)] = 0.073 \text{ for AMP-1131669938-1.psize}. \tag{11}$$

Tables 1 ~ 3 give the comparisons of $Var[H(n)]$ and $Var[D(n)]$ for 3 sets of test series, respectively. One particular thing worth noting is that $Var[H(n)]$ is in the order of magnitude of 10^{-2} while $Var[D(n)]$ is in the order of magnitude of 10^{-1}. Hence, we experimentally infer that traffic has the phenomenon expressed by

$$\text{Var}[D(n)] > \text{Var}[H(n)], \tag{12}$$

which is valid for the test data from the past (1989) to the current (2005) for either Ethernet traffic or WAN one. The mean square errors of the estimations of $D(n)$ and $H(n)$ are in the order of magnitude of 10^{-4}.
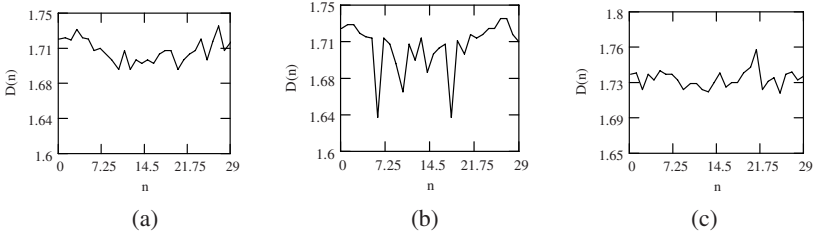

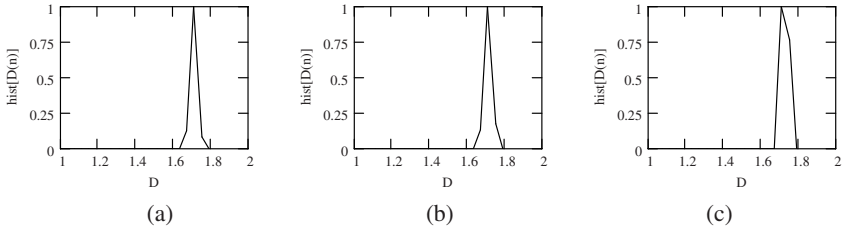**Fig. 4.** $D(n)$. (a). For DEC-PKT-1. (b). For BC-Oct89Ext. (c). For AMP-1131669938-1.psize.

Fig. 5. Histograms of $D(n)$. (a). For DEC-PKT-1: $\text{Var}[D(n)] = 0.298$. (b). For BC-Oct89Ext: $\text{Var}[D(n)] = 0.299$. (c). For AMP-1131669938-1.psize: $\text{Var}[D(n)] = 0.304$.

Table 1. Comparisons of $\text{Var}[H(n)]$ and $\text{Var}[D(n)]$ of traffic at Bellcore

Data series	$\text{Var}[H(n)]$	$\text{Var}[D(n)]$
pAug89	0.074	0.299
pOct89	0.076	0.300
OctExt	0.075	0.298
Octext41	0.077	0.301

Table 2. Comparisons of $\text{Var}[H(n)]$ and $\text{Var}[D(n)]$ of traffic at DEC

Data series	$\text{Var}[H(n)]$	$\text{Var}[D(n)]$
DECPKT1	0.073	0.298
DECPKT2	0.072	0.300
DECPKT3	0.074	0.301
DECPKT4	0.074	0.302

Table 3. Comparisons of Var[$H(n)$] and Var[$D(n)$] of traffic by NLANR

Data series	Var[$H(n)$]	Var[$D(n)$]
AMP-1131409930-1.psize	0.075	0.306
AMP-1131495398-1.psize	0.075	0.305
AMP-1131580868-1.psize	0.075	0.304
AMP-1131669938-1.psize	0.073	0.304

Finally, we note that the generalized GC process provides a new way to investigate or describe the multifractal [phenomenon of traffic. By "new way," we mean that D and H of traffic are separately characterized, which substantially differs from single parameter based models, such as fGn or local Hurst function. We highly appreciate an anonymous referee who proposed a challenging issue of how D statistically correlates H of traffic. Let it be the topic of the future work.

4 Conclusions

We have explained the variance analysis of $D(n)$ and $H(n)$ of traffic based on the GC process. The present result implies that traffic has the property Var[$D(n)$] > Var[$H(n)$], meaning that the variability of $D(n)$ of traffic is greater than that of $H(n)$ in general. This may be a novel description to the multifractal of traffic.

Acknowledgements

References

1. Taqqu, M. S., Teverovsky, V., Willinger W.: Is Network Traffic Self-Similar or Multifractal? Fractals **5** (1997) 63-73
2. Peltier, R. F., Levy-Vehel, J.: Multifractional Brownian Motion: Definition and Preliminaries Results. INRIA TR 2645, 1995
3. Lim, S. C., Muniandy, S. V.: On Some Possible Generalizations of Fractional Brownian Motion. Physics Letters A **226** (2000) 140-145
4. Willinger, W., Paxson, V., Riedi, R. H., Taqqu, M. S.: Long-Range Dependence and Data Network Traffic, in Long-range Dependence: Theory and Applications. Doukhan, P., Oppenheim, G., Taqqu, M. S., (eds.): Birkhauser (2002) 625-715

5. Li, M.: Change Trend of Averaged Hurst Parameter of Traffic under DDOS Flood Attacks. Computers & Security **25** (2006) 213-220

6. J. Beran, J., Shernan, R., Taqqu, M. S., Willinger, W.: Long-Range Dependence in Variable Bit-Rate Video Traffic. IEEE T. Communications **43** (1995) 1566-1579

7. Li, M., Zhao, W., et al.: Modeling Autocorrelation Functions of Self-Similar Teletraffic in Communication Networks based on Optimal Approximation in Hilbert Space. Applied Mathematical Modelling **27** (2003) 155-168

8. Tsybakov, B., Georganas, N. D.: Self-Similar Processes in Communications Networks. IEEE T. Information Theory **44** (1998) 1713-1725

9. Paxson, V., Floyd, S.: Wide Area Traffic: the Failure of Poison Modeling. IEEE/ACM T. Networking **3** (1995) 226-244

10. Mandelbrot, B. B., The Fractal Geometry of Nature. W. H. Freeman, New York (1982)

11. Chiles, J.-P., Delfiner, P.,: Geostatistics, Modeling Spatial Uncertainty. Wiley, New York (1999)

12. Gneiting, T., Schlather, M.: Stochastic Models that Separate Fractal Dimension and Hurst Effect. SIAM Review **46** (2004) 269-282

13. Li, M., Lim, S. C.: Modeling Network Traffic Using Cauchy Correlation Model with Long-Range Dependence. Modern Physics Letters B **19** (2005) 829-840

14. Lim, S. C., Li, M.: Generalized Cauchy Process and Its Application to Relaxation Phenomena. J. Phys. A: Math. Gen. **39** (2006) 2935-2951

15. Mandelbrot, B. B.: Multifractals and $1/f$ Noise. Springer, 1998

16. Mandelbrot, B. B.: Gaussian Self-Affinity and Fractals. Springer (2001)

17. Feldmann, A., Gilbert, A. C., Willinger, W., Kurtz, T.G.: The Changing Nature of Network Traffic: Scaling Phenomena. Computer Communication Review **28** (1998) 5-29

18. Willinger, W., Govindan, R., Jamin, S., Paxson, V., Shenker, S.: Scaling Phenomena in the Internet: Critically Examining Criticality. Proceedings of Natl. Acad. Sci. USA, **99 (Suppl. 1)** (2002) 2573-2580

19. Kent, J. T., Wood, T. A.: Estimating the Fractal Dimension of a Locally Self-Similar Gaussian Process by Using Increments. J. R. Statit. Soc. B **59** (1997) 579-599

20. Adler, R. J.: The Geometry of Random Fields. Wiley, New York (1981)

21. Davies, S., Hall, P.: Fractal Analysis of Surface Roughness by Using Spatial Data. Journal of the Royal Statistical Society Series B **61** (1999) 3-37

22. Hall, P.: On the Effect of Measuring a Self-Similar Process. SIAM J. Appl. Math. **35** (1995) 800-808

23. Constantine, A. G., Hall, P.: Characterizing Surface Smoothness via Estimation of Effective Fractal Dimension. Journal of the Royal Statistical Society Ser. B **56** (1994) 97-113

24. Hall, P., Roy, R.: On the Relationship between Fractal Dimension and Fractal Index for Stationary Stochastic Processes. The Annals of Applied Probability **4** (1994) 241-253

25. Chan, G., Hall, P., Poskitt, D. S., Periodogram-Based Estimators of Fractal Properties. The Annals of Statistics **23** (1995) 1684-1711

26. Todorovic, P.: An Introduction to Stochastic Processes and Their Applications. Springer Verlag, New York (1992) 100-

27. Lim, S. C., Muniandy, S. V.: Generalized Ornstein-Uhlenbeck Processes and Associated Self-Similar Processes. J. Phys. A: Math. Gen. **36** (2003) 3961-3982

28. Wolperta, R. L., Taqqu, M. S.: Fractional Ornstein–Uhlenbeck Lévy Processes and the Telecom Process Upstairs and Downstairs. Signal Processing **85** (2005) 1523-1545

29. On-line available: http://ita.ee.lbl.gov/html/traces.html

30. On-line available: http://pma.nlanr.net/Traces/Traces/daily/

31. Leland, W., Taqqu, M. S., Willinger, W., Wilson, D.: On the Self-Similar Nature of Ethernet Traffic (extended version). IEEE/ACM T. Networking **2** (1994) 1-15
32. Li, M., Jia, W. J., Zhao, W.: Correlation Form of Timestamp Increment Sequences of Self-Similar Traffic on Ethernet. Electronics Letters **36** (2000) 1168-1169
33. http://www.nlanr.net/
34. http://moat.nlanr.net/
35. Li, M., Lim, S. C.: A Rigorous Derivation of Power Spectrum of Fractional Gaussian Noise. Fluctuation and Noise Letters **6** (2006) C33-C36
36. Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P.: Numerical Recipes in C: the Art of Scientific Computing. 2nd Edition, Cambridge University Press (1992)