Dynamics of Temporal Correlation in Daily Internet Traffic

Kensuke Fukuda
NTT Network Innovation Laboratories
Tokyo, 180-8585, Japan
Email: fukuda@t.onlab.ntt.co.jp

Luís A. Nunes Amaral
Dept. of Chemical Engineering,
Northwestern University
Evanston, IL 60208, USA

H. Eugene Stanley
Center for Polymer Studies and
Dept. of Physics,
Boston University
Boston, MA 02215, USA

***Abstract*— In order to characterize the dynamics of self-similar behavior in daily Internet traffic, we analyze the time series of traffic volume for 24-hour period in wide-area Internet, by using detrended fluctuation analysis (DFA)—a well-known method of characterizing nonstationarity in a time series. We show that the estimated scaling exponent (which is directly related to the Hurst parameter) of traffic fluctuations has a dependency on the level of human activity for a time scale greater than 30s. Thus, the temporal correlation for traffic fluctuations is close to $1/f$-noise during the day, and becomes weaker at night. This result suggests that Internet traffic cannot be modeled using the unique value of the Hurst parameter.**

I. Introduction

Recent analyses of Internet traffic have shown that traffic volume fluctuations in wide-area and local-area networks are characterized by self-similarity, or long-range dependency [1]–[3]. Self-similarity is a scale invariant property under time-scale translation, i.e. it yields the existence of clustering and bursty characteristics in the flow over wide time scales. Thus, self-similar traffic causes larger queueing delays than the estimation by Poissonian traffic [4], [5].

Theoretically, self-similar time series is characterized by a power law type power spectrum density

$$S(f) \propto f^{-\beta}. \tag{1}$$

To quantify the self-similarity, the Hurst parameter, $H = \frac{\beta+1}{2}$, is frequently used; For positive temporal correlation of a self-similar time series, $0.5 < H < 1.0$, while in a temporally uncorrelated time series, $H = 0.5$. When $H = 1.0$, the time series produces $1/f$ noise. Many analyses of Internet traffic have reported that H is broadly distributed around $0.6 < H \leq 1.0$ for several measurement points, indicating that traffic fluctuations have a long range correlation in time. Thus, self-similarity and long range correlation would appear to be intrinsic properties in Internet traffic. However, the value of the scaling parameter H depends on the environment of the measurement points. From the perspective of statistical physics, traffic behavior can be viewed as phase transition phenomena between non-congested and congested phases [6], and phase transition analysis shows that traffic fluctuations become $1/f$-noise (i.e., $H = 1.0$) at the critical point between the two phases [6]. Moreover, away from this critical point, the temporal correlation becomes weaker, and traffic fluctuations are close to white noise. Namely, the phase transition view includes self-similar and non-self-similar states. However, we do not fully understand the temporal dynamics of the correlation during the day; In other words, how does the scaling parameter fluctuate during 24 hours?

In this paper, we focus on the temporal dynamics of this scaling parameter during a 24-hour period in order to characterize daily traffic dynamics. Using detrended fluctuation analysis (DFA) [7], we analyze two wide-area Internet traffic datasets. DFA is a method widely used to characterize a nonstationary [1] time series and it can estimate the scaling parameter more accurately than such traditional methods as power-spectrum analysis and R/S analysis [7], [8]. Our DFA results reveal that the value of the scaling exponent (which is directly related to the Hurst parameter) of traffic fluctuations depends on the level of human activity for a time scale longer than 30s, although Internet traffic fluctuations have a positive temporal correlation during a 24-hour period. Thus, the traffic fluctuations are approximately $1/f$-noise during the day, while the correlation is low at night.

II. Method and datasets

A. Detrended Fluctuation Analysis

The DFA method of characterizing a nonstationary time series is based on the root mean square analysis of a random walk [7]. DFA is better for this task than power spectrum analysis or R/S analyses because it produces results that are independent of the effect of the trend, i.e., DFA screens out spurious self-similarity [8]. For this reason, DFA is widely used in analyzing real-world time series (e.g. physiology, and econophysics) [9], [10].

To use the DFA method, we (also, see Fig. 1):

1) integrate the time series;
2) divide the time series into "boxes" of length n;
3) in each box, perform a least-squares polynomial fit of order p to the integrated signal;
4) in each box, calculate the root-mean-square deviations of the integrated signal from the polynomial fit, corresponding to the elimination of the local trend; and
5) repeat this procedure for different box sizes (i.e., time scales) n.

[1] Here, stationarity means that the mean and variance of a time series does not change under time translation.

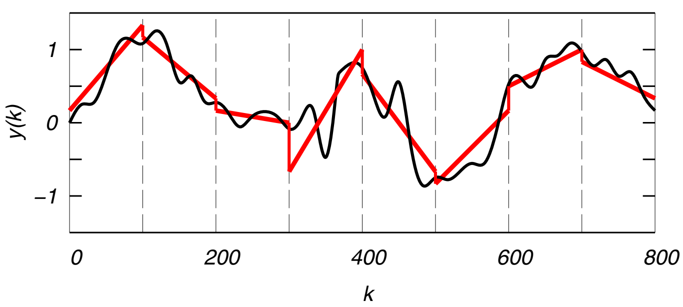

Fig. 1. Schematic representation of detrended fluctuation analysis (DFA), which is a method of characterizing a nonstationary time series based on the root-mean-square method. $y(k)$ is a value obtained by integrating the original time series for the time k (black curve). Gray lines indicate the polynomial fit ($p = 1$) in a box ($n = 100$). $F(n)$ is calculated by the root-mean-square of deviation between $y(k)$ and the polynomial fit. For a self-similar time series, we observe a power law in Eq. (2) between $F(n)$ and n. The value of exponent α of the power law in the the DFA result is directly related to the value of the power law in the power spectrum density; $\beta = 2\alpha - 1$. An advantage of the DFA is that it is independent of the number of points in the data set. Also, the power spectrum density tends to be affected by trends.

For a self-similar time series, we find a power law relation

$$F(n) \sim n^{\alpha} \tag{2}$$

between the average magnitude of deviated fluctuations $F(n)$ and the size of boxes n.

Scaling exponent α is related to the exponent of the power law in power spectrum density with

$$\beta = 2\alpha - 1. \tag{3}$$

For $0 < \alpha < 0.5$, the time series is anti-correlated, while for $\alpha > 0.5$, the time series has a temporal positive correlation [11]. In particular, $\alpha = 0.5$, $\alpha = 1.0$, and $\alpha = 1.5$ correspond to white noise, $1/f$-noise, and Brownian motion, respectively. From Eq.(3), the scaling exponent corresponds to the Hurst parameter, $H = \alpha$ for $0.5 \leq \alpha \leq 1.0$.

B. Datasets

We analyzed the two wide-area Internet traffic datasets shown in Table 1. Both measurement points are the entrance between the organization and the Internet. A raw traffic data volume consists of the header part of the packet and the time stamp when the packet is captured. From a raw data volume, we reconstructed the time series for the number of bytes of the data packet passing through the link from the Internet to the organization with 0.1s bin.

Figure 2 displays an example of traffic fluctuations, which is captured at the link between NTT laboratories and the Internet. These were from (a) 13:00–14:00 (day) and (b) 05:00–06:00 (night). It is visually apparent that both fluctuations are characterized by nonstationarity, and that daytime traffic behavior largely differs from that at night. The mean traffic flow density is 80 KB during the day, and 1.5 KB at night. Also, the mean number of the existing connections for 0.1 s is ≈ 600 during the day and ≈ 50 at night.

[2]The original data volumes are available at http://www.nlann.net/auckland

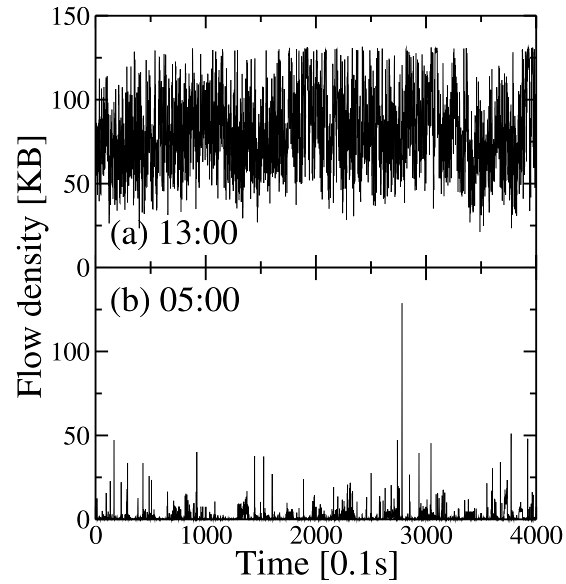

Fig. 2. An example of Internet traffic fluctuations for 400 s of 0.1 s bin at (a) 13:00 and (b) 05:00. Data is the number of bytes passing through the link between the Internet and NTT Laboratories in Jun. 2001. It is visually apparent that the fluctuations during the day are more violent than those during the night.

III. Results

Figure 3 displays the result of the DFA for two one-hour traffic time series, indicating the difference in the time of the day; 13:00–14:00 (day) and 05:00–06:00 (night). It is visually apparent that both plots follow the power law decays given in Eq. (2), but with a crossover point at $n_{\times} \approx 300$, corresponding to 30 s. For $n < n_{\times}$, the slopes of plots are closer at $\alpha \approx 0.85$ during day and night, meaning that the activity of users had no likely effect on temporal correlation for a smaller time scale. However, for $n > n_{\times}$, α is characterized by a different value; $\alpha_{day} = 1.01$ and $\alpha_{night} = 0.66$. Namely, during the day, the traffic behavior is approximately $1/f$-noise for larger time scales. On the other hand, during the night, the traffic behavior is still temporally correlated, but the level of the temporal correlation is lower than that during the day. This result suggests that in order to estimate scaling exponent α, one should not calculate from a longer time series, because the analysis of such long time series conceals the details on the dynamics of the scaling exponent.

Next, in order to verify the results of DFA, we evaluate the same traffic time series by using the power spectrum analysis as shown in Fig. 4. The results are not so obvious in estimating the value of the slope of the power law in Eq. (1). For $f < 10^{-1}$, the value of exponent β is close to 0.9 for the 13:00 dataset. Also, $\beta \approx 0.0$ for 05:00 dataset, meaning that the temporal correlation during the night is quite low, and close to white noise. The result of the power spectrum analysis also supports our finding that the temporal correlation of traffic fluctuations depends on the level of human activity. Our result also indicates that DFA is a more useful method for estimating the scaling exponent than power spectrum analysis.

TABLE I

DESCRIPTION OF MEASUREMENT POINTS.

Dataset	Measurement point	Bandwidth	Datalink	Duration
NTT	NTT laboratories (Tokyo)	11Mbps	ATM	Jul. 2001 (168 hours)
Auckland[2]	Auckland University (NZ)	5Mbps	ATM	Feb. 2001 (72 hours)

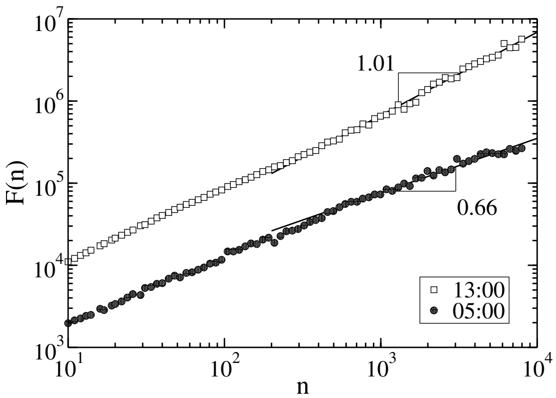

Fig. 3. Results of the DFA for two one-hour traffic fluctuations at NTT. Both plots are characterized by power laws with crossover points at $n_\times \approx 300$, which corresponds to 30 s. For $n < n_\times$, the slopes for 05:00 and 13:00 are close to 0.85. Thus, for the time scale below $n_\times$, temporal correlation might be generated by the same mechanism during the night and during the day. Also, for $n > n_\times$, scaling exponent α is above 0.5, meaning that traffic fluctuations have a positive temporal correlation. Remarkably, α has different values during the night and during the day. In particular, the exponent at 13:00 is close to 1.0, indicating that traffic fluctuations become $1/f$-noise. This result suggests that one should not analyze a long dataset at once, because the analysis of such a long time series conceals details on the behavior of the exponent.

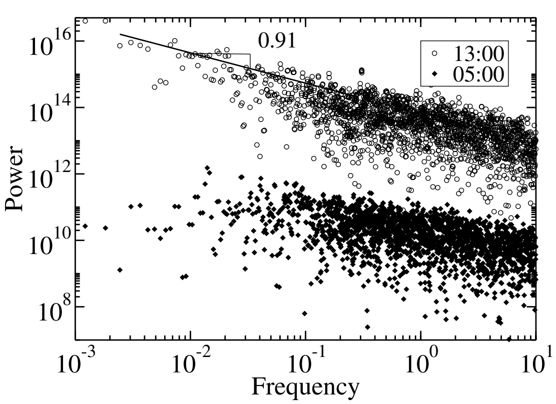

Fig. 4. Results of the power spectrum analysis. Original data are the same as those for the previous DFA results. For both data, there are crossover points at around 10^{-1} (= 10s). Also it is difficult to estimate the slope of the plots correctly for $f < 10^{-1}$. However, it is visually apparent that the statistical properties of the traffic are different during day and night for long time scales (> 10s)

Moreover, in order to investigate the dynamic changes in the scaling exponent during a 24 hour period, we used DFA to estimate the value of scaling exponent α for each 1-hour time series. Figure 5 displays (a) variations in the mean traffic flow density and (b) variations in the value of scaling exponent α. Variations of the mean flow density indicate that the amount of the traffic is small between 22:00-08:00, and remains high from 10:00–18:00, corresponding to the daily working patterns of researchers. The exponent was estimated from the result of DFA by fitting the region for $n > n_\times$. In Fig. 5(b), α remains smaller around 0.6–0.9 for 23:00–08:00, while α is stable at $\approx$ 0.95 for 10:00–19:00. Approximately, traffic fluctuations are $1/f$-noise for >3 Mbps. This result indicates that the temporal correlation of traffic fluctuations are affected by the level of the activity of the users.

For a further check, we also plotted the dynamics of the scaling exponent for Auckland traffic dataset as shown in Fig. 6. The absolute value of mean flow density is different from that of the NTT dataset, but the shape of the functional form resembles each other. Also, the scaling exponent remains around 0.7–0.8 for 23:00–07:00 and increases ($\approx$ 0.9) for 09:00–18:00. For >1Mbps, the value of the exponent stays around 0.9. Though the usage of the users in a university is different from that of the researchers, the dynamics of the scaling exponent for the Auckland traffic dataset is consistent with the dynamics for the NTT traffic dataset. Thus, this dynamics of the exponent is most likely an intrinsic property of Internet traffic.

IV. DISCUSSION

Here, we discuss a possible cause for the observed dynamics of the statistical property from the viewpoint of the file size, and the protocol components. Interestingly, cumulative distribution $P(s)$ of transfered byte size s in each connection follows a power law form $P(s) \propto s^{-1.0}$, independent of the time of day as shown in Fig. 7. This result is consistent with the distribution of file sizes in a Unix file system [12]. If the distribution for the transfered size directly dominates the temporal correlation of traffic, the unique exponents would be observed during day and night independent of the measurement points. In particular, this effect is expected to be clear for the case of small amounts of traffic. From our result for $n < n_\times$, the estimated exponent becomes a unique value independent of the difference in the time of day. For this reason, the file size effect has a possibility to generate temporal correlations for smaller time scale. However, our results for longer time scales (>30 s) shows that the scaling exponent is characterized by the different value depending on the time of day. Thus, it is unlikely that the exponent α observed in traffic

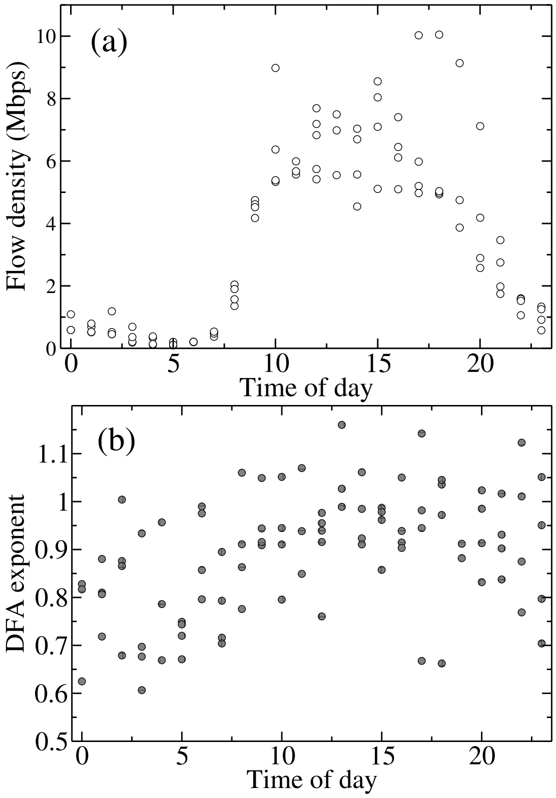

Fig. 5. (a) Dynamic behavior of traffic fluctuations at the link between NTT laboratories and the Internet. It is visually apparent that the mean flow density of traffic depends on the level of the activity by users, corresponding to the working pattern of researchers. (b) Behavior of the estimated value of scaling exponent α. We estimate α from the fitting of the curve of the DFA result for $n > n_{\times}$. All plots are above 0.5, so that traffic fluctuations have a positive correlation. However, we observed that the value of the exponent is also related to the level of the activity by the users. Namely, the exponents are close to 0.6-0.9 during 01:00-07:00, though they are close to 0.95-1.05 during 10:00-19:00. Again, the value of the exponent decreases during 19:00-24:00.

fluctuations is determined by the file size distribution.

It is plausible that the temporal correlations generated for longer time scales are an effect of TCP [14]–[17], which is based on the feedback control with a retransmission mechanism that provides reliable communication between hosts [13]. Reference [14] showed that the retransmission mechanism is an important factor in generating temporal correlations in traffic fluctuations. Considering that retransmission events occur when the network is congested (at least at a bottleneck link), this effect can explain our DFA results, i.e., $1/f$-behavior during the day, and temporal correlations weakening during night. Moreover, Refs. [16], [17] indicate that the simple feedback control generates $1/f$-behavior at the critical flow density between non-congested and congested phases; This is because each host fits its transmission rate to the current network status by using round-trip time information as the network becomes congested. This implicit cooperative behavior among hosts changes the statistical behavior of the network traffic: it becomes highly correlated. Both mechanisms may have an effect on the statistical properties of the network traffic when the time scales are longer.

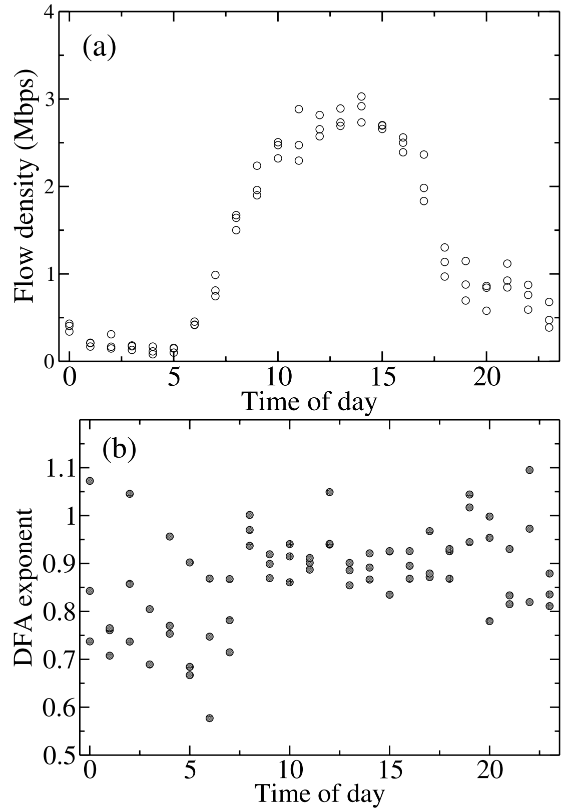

Fig. 6. (a) Behavior of mean flow density in the Auckland dataset, and (b) behavior of the estimated DFA exponent. The exponents converge around 0.9 during 09:00-20:00, though the exponents are broadly distributed around $0.7 - 0.8$ during 00:00-07:00. This result is consistent with the result for the NTT dataset.

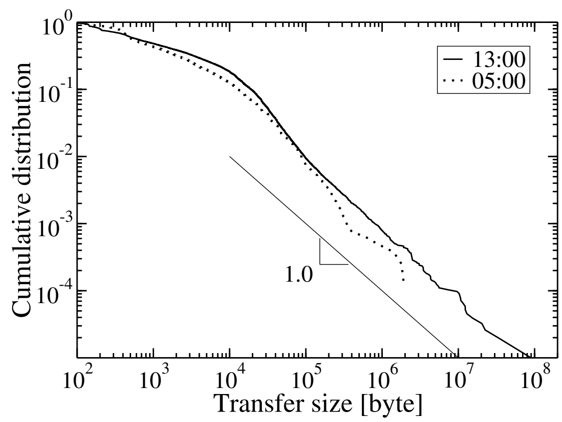

Fig. 7. Cumulative distribution of transfered byte size in a TCP connection; 05:00 (night) and 14:00 (day). By using tcptrace command [18], we count the unique transfered byte size in a successful TCP connection, in which both SYN and FIN packets appeared. The distribution follows a power law decay with an exponent of ≈ 1.0 for $> 10^4$, identical to that of the cumulative distribution of file sizes in a Unix file system. Although the distribution has a cutoff point at around 10^6 during the night, the values of the exponent are both close to 1.0. Note also that the observed exponent usually deviates from the scaling exponent α during the night. This suggests that the scaling exponent is independent of the exponent in the transferred size.

The value of the scaling exponent for longer time scales seems to be closely related to the mean flow density of the traffic as shown in Figs. 5 and 6. However, although the mean flow density of the Auckland dataset during the day is much smaller than that of the NTT dataset, the scaling exponent is close to 0.9–1.0 during day for both datasets. From this result, it is hard to directly link the value of the scaling exponent to the absolute mean flow density of traffic fluctuations. The value of the scaling exponent probably depends on the "degree of congestion," which is determined by the topological environment and the current network status (e.g. the location of the bottleneck, the ratio of the bandwidth usage, etc). In this sense, the phase transition view can be useful in explaining fluctuations of the scaling exponent [6]. Reference [6] shows that the correlation length tends to diverge at the critical traffic flow density and, at the critical point, traffic fluctuations become $1/f$-noise. Thus, we can interpret that the traffic status during the day is close to the critical point, and below the critical point during the night.

V. Conclusion

We analyzed wide-area Internet traffic to understand the dynamics of its temporal correlation. Our analysis of each one-hour traffic dataset revealed that the value of the estimated scaling exponent of the DFA, which directly corresponds to the Hurst parameter, fluctuates between 0.6 and 1.0 for a 24-hour period. In particular, we found that the temporal correlation becomes smaller at night, while the traffic behavior is characterized by $1/f$-noise during day. Thus, our results indicate that Internet traffic cannot be modeled using self-similarity with the unique Hurst parameter.

Acknowledgment

We wish to thank T. Mori for providing NTT dataset, and Auckland University team for providing the Auckland dataset.

References

[1] W. E. Leland, M.S. Taqqu, W. Willinger, and D.V. Willson. "On the Self-Similar Nature of Ethernet Traffic," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1–15, 1994.

[2] V. Paxson and S. Floyd. "Wide-Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226-244, 1995.

[3] M. Crovella and A. Bestavros. "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 835–846, 1997.

[4] I. Norros. "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 953–962, 1995.

[5] A. Erramilli, O. Narayan, and W. Willinger. "Experimental Queueing Analysis with Long-Range Dependent Traffic," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 209–223, 1996.

[6] M. Takayasu, H. Takayasu, and K. Fukuda. "Dynamic Phase Transition Observed in the Internet Traffic Flow," *Physica A*, vol. 277, pp. 248–255, 2000.

[7] C-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger. "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," *Chaos*, vol. 5, pp. 82–87, 1995.

[8] K. Hu, P. Ch. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley. "Effect of trends on detrended fluctuation analysis," *Physical Review E*, vol. 64, pp. 011114, 2001.

[9] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. Ch. Ivanov, C-K. Peng, and H. E. Stanley. "Fractal dynamics in physiology: Alterations with disease and aging," *Proceedings of the National Academy of Sciences USA*, vol. 99 Supp. 1, pp. 2466-2472, 2002.

[10] H. Takayasu (ed.), "Empirical Science of Financial Fluctuations: The Advent of Econophysics," Springer-Verlag, Tokyo, 2002.

[11] M. F. Shlesinger. "Fractal time and $1/f$ noise in complex systems," *Ann. NY Acad. Sci.*, vol. 504, pp. 214–228, 1987.

[12] G. Irlam. "Unix file size survey - 1993," Available at `http://www.base.com/gordoni/ufs93.html`, 1994.

[13] J. Postel. "Transmission Control Protocol," RFC 793, 1981.

[14] A. Feldmann, A. C. Gilbert, P. Huang, and W. Willinger. "Dynamics of IP Traffic: A Study of the Role of Variability and the Impact of Control," *Proc. ACM SIGCOMM99*, pp. 301-313, Cambridge, USA (1999).

[15] A. Veres, and M. Boda. "The Chaotic Nature of TCP Congestion Control". *Proc. IEEE INFOCOM2000*, pp. 1715–1723, TelAviv, Israel (2000).

[16] K. Fukuda, M. Takayasu, and H. Takayasu. "Analysis of Minimal Model of Internet Traffic," *Proc. Traffic and Granular Flow 2001*, Springer (2001).

[17] K. Fukuda, M. Takayasu, and H. Takayasu. "Origin of Self-Similarity in TCP Traffic," *Proc. SPECTS 2003*, pp. 18–26, Montreal, Canada (2003).

[18] S. Ostermann. "Tcptrace home page. `http://irg.cs.ohiou.edu/software/tcptrace/`."