Point Process Models of 1/f Noise and Internet Traffic

V. Gontis, B. Kaulakys, and J. Ruseckas

Institute of Theoretical Physics and Astronomy, Vilnius University, A. Goštauto 12, LT-01108 Vilnius, Lithuania

Abstract. We present a simple model reproducing the long-range autocorrelations and the power spectrum of the web traffic. The model assumes the traffic as Poisson flow of files with size distributed according to the power-law. In this model the long-range autocorrelations are independent of the network properties as well as of inter-packet time distribution.

Keywords: computer networks, 1/f noise, point processes, traffic statistics
PACS: 05.10.-a, 05.40.-a, 05.45.Pq, 05.60.Cd, 87.18.Sn

INTRODUCTION

The power spectra of large variety of complex systems exhibit $1/f$ behavior at low frequencies. It is widely accepted that $1/f$ noise and self-similarity are characteristic signatures of complexity [1, 2]. Studies of network traffic and especially of Internet traffic prove the close relation of self-similarity and complexity. Nevertheless, there is no evidence whether this complexity arises from the computer network or from the computer file statistics. We already have proposed a few stochastic point process models exhibiting self-similarity and $1/f$ noise [3, 4, 5, 6]. The signal in these models is a sequence of pulses or events. In the case of δ-type pulses (point process) the signal is defined by the stochastic process of the interevent time [5]. We have shown that the Brownian motion of interevent time of the signal pulses [3] or more general stochastic fluctuations described by multiplicative Langevin equation are responsible for the $1/f$ noise of the model signal [5]. It looks very natural to model computer network traffic exhibiting self-similarity by such stochastic point process signal. In case of success it would mean that self-similar behavior is induced by the stochastic arrival of requests from the network. Another possibility is to consider that the self-similar behavior is induced by the server statistics, rather than by the arrival process. The empirical analysis of the computer network traffic provides an evidence that the second possibility is more realistic [7]. This imposed us to model the computer network traffic by Poisson sequence of pulses with stochastic duration. We recently showed that under suitable choice of the pulse duration statistics such a signal exhibited $1/f$ noise [6].

In this contribution we provide the analytical and numerical results consistent with the empirical data and confirming that self-similar behavior of the computer network traffics is related with the power-law distribution of files transferred in the network.

CP776, *Science of Complex Networks: From Biology to the Internet and WWW; CNET 2004,*
edited by J. F. F. Mendes, S. N. Dorogovtsev, A. Povolotsky, F. V. Abreu, and J. G. Oliveira

SIGNAL AS A SEQUENCE OF PULSES

We will investigate a signal consisting of a sequence of pulses. We assume that:

1. the pulse sequences are stationary and ergodic;
2. interevent times and the shapes of different pulses are independent.

The general form of such signal can be written as

$$I(t) = \sum_k A_k(t - t_k) \tag{1}$$

where the functions $A_k(t)$ determine the shape of the individual pulses and the time moments t_k determine when the pulses occur. Time moments t_k are not correlated with the shape of the pulse A_k and the interevent times $\tau_k = t_k - t_{k-1}$ are random and uncorrelated. The occurrence times of the pulses t_k are distributed according to Poisson process.

The power spectrum is given by the equation

$$S(f) = \lim_{T\to\infty} \left\langle \frac{2}{T} \left| \int_{t_i}^{t_f} I(t) e^{-i2\pi f t} dt \right|^2 \right\rangle \tag{2}$$

where $T = t_f - t_i$ and the brackets $\langle ... \rangle$ denote the averaging over realizations of the process. The power spectral density of a random pulse train is given by Carson's theorem

$$S(f) = 2\bar{\nu} \left\langle |F_k(\omega)|^2 \right\rangle, \quad \omega = 2\pi f \tag{3}$$

where

$$F_k(\omega) = \int_{-\infty}^{+\infty} A_k(u) e^{i\omega u} du \,. \tag{4}$$

is the Fourier transform of the pulse A_k and

$$\bar{\nu} = \lim_{T\to\infty} \left\langle \frac{N+1}{T} \right\rangle \tag{5}$$

is the mean number of pulses per unit time. Here $N = k_{\max} - k_{\min}$is the number of pulses.

Pulses of variable duration

Let the only random parameter of the pulse is the duration. We take the form of the pulse as

$$A_k(t) = T_k^{\beta} A\left(\frac{t}{T_k}\right), \tag{6}$$

where T_k is the characteristic duration of the pulse. The value $\beta = 0$ corresponds to the fixed height pulses; $\beta = -1$ corresponds to constant area pulses. Differentiating the fixed area pulses we obtain $\beta = -2$. The Fourier transform of the pulse (6) is

$$F_k(\omega) = \int_{-\infty}^{+\infty} T_k^{\beta} A\left(\frac{t}{T_k}\right) e^{i\omega t} dt = T_k^{\beta+1} \int_{-\infty}^{+\infty} A(u) e^{i\omega T_k u} du \equiv T_k^{\beta+1} F(\omega T_k).$$

From Eq. (3) the power spectrum is

$$S(f) = 2\bar{\nu} \left\langle T_k^{2\beta+2} |F(\omega T_k)|^2 \right\rangle. \tag{7}$$

Introducing the probability density $P(T_k)$ of the pulses durations T_k we can write

$$S(f) = 2\bar{\nu} \int_0^{\infty} T_k^{2\beta+2} |F(\omega T_k)|^2 P(T_k) dT_k. \tag{8}$$

If $P(T_k)$ is a power-law distribution, then the expressions for the spectrum are similar for all β.

Power-law distribution

We take the power-law distribution of the pulse durations

$$P(T_k) = \begin{cases} \frac{\alpha+1}{T_{\max}^{\alpha+1} - T_{\min}^{\alpha+1}} T_k^{\alpha}, & T_{\min} \le T_k \le T_{\max}, \\ 0, & \text{othervise.} \end{cases} \tag{9}$$

From Eq. (8) we have the spectrum

$$\begin{aligned} S(f) &= 2\bar{\nu} \frac{\alpha+1}{T_{\max}^{\alpha+1} - T_{\min}^{\alpha+1}} \int_0^{\infty} T_k^{\alpha+2\beta+2} |F(\omega T_k)|^2 dT_k \\ &= \frac{2\bar{\nu}(\alpha+1)}{\omega^{\alpha+2\beta+3}(T_{\max}^{\alpha+1} - T_{\min}^{\alpha+1})} \int_{\omega T_{\min}}^{\omega T_{\max}} u^{\alpha+2\beta+2} |F(u)|^2 du. \end{aligned}$$

When $\alpha > -1$ and $\frac{1}{T_{\max}} \ll \omega \ll \frac{1}{T_{\min}}$ then the expression for the spectrum can be approximated as

$$S(f) \approx \frac{2\bar{\nu}(\alpha+1)}{\omega^{\alpha+2\beta+3}(T_{\max}^{\alpha+1} - T_{\min}^{\alpha+1})} \int_0^{\infty} u^{\alpha+2\beta+2} |F(u)|^2 du. \tag{10}$$

If $\alpha + 2\beta + 2 = 0$ then in the frequency domain $\frac{1}{T_{\max}} \ll \omega \ll \frac{1}{T_{\min}}$ the spectrum is

$$S(f) \approx \frac{2\bar{\nu}(\alpha+1)}{\omega(T_{\max}^{\alpha+1} - T_{\min}^{\alpha+1})} \int_0^{\infty} |F(u)|^2 du. \tag{11}$$

Therefore, we obtained $1/f$ spectrum. The condition $\alpha + 2\beta + 2 = 0$ is satisfied, e.g., for the fixed area pulses ($\beta = -1$) and uniform distribution of pulse durations ($\alpha = 0$) or for fixed height pulses ($\beta = 0$) and uniform distribution of inverse durations $\gamma = T_k^{-1}$, i.e. for $P(T_k) \propto T_k^{-2}$.

Rectangular pulses

We will obtain the spectrum of the rectangular fixed height pulses ($\beta = 0$). The height of the pulse is a and the duration is T_k. The Fourier transform of the pulse is

$$F(\omega T_k) = a\int_0^1 du\, e^{i\omega T_k u} = a\frac{e^{i\omega T_k}-1}{i\omega T_k} = ae^{i\frac{\omega T_k}{2}}\frac{2\sin\left(\frac{\omega T_k}{2}\right)}{\omega T_k}. \qquad (12)$$

Then the spectrum according to Eqs. (8), (9) and (12) is

$$\begin{aligned} S(f) &= \frac{4\bar{\nu}a^2}{\omega^2} + \frac{4\bar{\nu}a^2(\alpha+1)}{\omega^{\alpha+3}(T_{\max}^{\alpha+1}-T_{\min}^{\alpha+1})} \\ &\times \mathrm{Re}\left\{i^{-1-\alpha}\left(\Gamma(\alpha+1, i\omega T_{\max}) - \Gamma(\alpha+1, i\omega T_{\min})\right)\right\} \end{aligned} \qquad (13)$$

where $\Gamma(a,z)$ is the incomplete gamma function, $\Gamma(a,z) = \int_z^\infty u^{a-1}e^{-u}du$.

For $\alpha = -2$ we have the uniform distribution of inverse durations. The term with $\Gamma(\alpha+1, i\omega T_{\max})$ is small and can be neglected. We also assume that $T_{\min} \ll T_{\max}$ and neglect the term $\left(\frac{T_{\max}}{T_{\min}}\right)^{\alpha+1}$. Then we obtain $1/f$ spectrum

$$S(f) \approx \frac{\bar{\nu}a^2}{f}T_{\min}. \qquad (14)$$

Further we will investigate how the variable duration of the pulses is related with the variable size of files transferred in the computer networks.

MODELING COMPUTER NETWORK TRAFFIC BY SEQUENCE OF PULSES

In this section we provide numerical simulation results of the computer network traffic based on the description of signals as uncorrelated sequence of variable size web requests. We model empirical data of incoming web traffic publicly available on the Internet [8]. Our assumptions are closely related with the model description in the previous section, with the empirical data and analysis provided in Ref. [7]. First of all from Eq. (14) it is clear that the sequence of requests distributed as power law (9) for $\alpha = -2$ yields $1/f$ spectrum, as observed in the empirical data [8]. For the numerical calculations we use the positive Cauchy distribution instead

$$P(x) = \frac{2}{\pi}\frac{s}{s^2+x^2} \qquad (15)$$

which better approximates the empirical request size distribution [8]. Where $s = 4100$ *bytes* is empirical parameter of distribution and x is a stochastic size of the file requests in *bytes*. Requested files arrive as Poisson sequence with mean inter-arrival time $\tau_f = 0.101$ seconds. The files arrive divided by the network protocol into $n_p = x/1500$ packets.

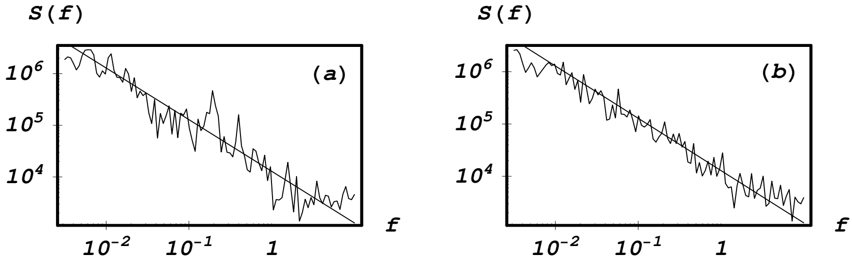

FIGURE 1. Power spectral density $S(f)$ versus frequency f calculated numerically according to Eq. (2): a) from empirical incoming web traffic presented in [8]; b) from numerically simulated traffic dividing files into Poisson sequence of packets with $s = 4100$, $\tau_f = 0.101$, $\tau_p = 11.6 \times 10^{-6} \times 10^{\varepsilon}$, where ε is a random variable equally distributed in the interval [0,3]. Straight lines represent theoretical prediction (14) with empirical parameters according to Eq. (16).

According to our assumption these packets spread into another Poisson sequence with mean inter-packet time τ_p. The total incoming web traffic is a sequence of packets resulting from all requests. This procedure reconstructs the described Poisson sequence of variable duration pulses into self-similar point process modeling traffic of packets. Our numerical results confirm that the spectral properties of the packet traffic are defined by the Poisson sequence of variable duration and are independent of file division into packets. It is natural to expect that mean inter-packet time τ_p depends on the position of computer on the network from which the file is requested. Consequently, the inter-packet time distribution measured from the empirical histogram or calculated in this model must depend on the computer network structure when the spectral properties and autocorrelation of the signal are defined by the file size statistics independent of network properties. Our numerical simulation of the web incoming traffic and its power spectrum, presented on Fig. 1, confirm that the flow of packets exhibits $1/f$ noise and long-range autocorrelation induced by the power law (positive Cauchy) distribution of transferred files. Both empirical and simulated spectrum are in good agreement with theoretical prediction (14), which we rewrite with empirical parameters of the model as:

$$S(f) \approx \frac{s \ln 10}{f \tau_f p \tau_{p,max}}. \tag{16}$$

Where $p = 1500$ is a standard packet size in *bytes* and $\tau_{p,max} = 11.6 \times 10^{-3}$ is a maximum inter-packet time.

In Fig.2. we present the empirical and numerically simulated histograms of the inter-packet time τ_p We assume a very simple model to reproduce empirical distribution of the packet arrivals. Files arrive divided into packets with inter-packet time $\tau_p = 11.6 \times 10^{\varepsilon-6}$, where ε is a random variable equally distributed in the interval [0,3]. This assumption reproduces empirical distribution of the inter-packet time pretty well, as seen in Fig.2.

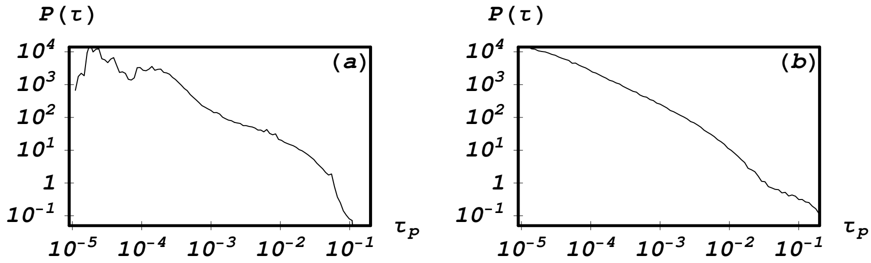

FIGURE 2. Inter-packet time histograms: a) empirical data of web incoming traffic [8]; b) numerical simulation with same parameters as in Fig.1.

CONCLUSIONS

In this contribution we present a very simple model reproducing the long-range autocorrelations and power spectrum of the web traffic. The model assumes the traffic as Poisson flow of files distributed according to the power-law. In this model the long-range autocorrelations are independent of the network properties and of the inter-packet time distribution. We reproduced the inter-packet time distribution of incoming web traffic assuming that files arrive as Poisson sequence with mean inter-packet time equally distributed in a logarithmic scale. This simple model may be applicable to the other computer networks as well.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Uli Harder in making empirical data available on the Internet. This contribution was prepared with the support of the Lithuanian State Science and Studies Foundation.

REFERENCES

1. D. L. Gilden, T. Thornton and M. W. Mallon, *Science* **267**, 1837–1839 (1995).
2. H. Wong, *Microel. Reliab.* **43**, 585–599 (2003).
3. B. Kaulakys and T. Meskauskas, *Phys. Rev. E* **58**, 7013–7019 (1998).
4. B. Kaulakys, *Microel. Reliab.* **40**, 1787–1790 (2000).
5. V. Gontis and B. Kaulakys, *Physica A* **343**, 505–514 (2004).
6. J. Ruseckas, B. Kaulakys, and M. Alaburda, *Lith. J. Phys.* **43**, 223–228 (2003).
7. A. J. Field, U. Harder, and P. G. Harrison, *IEE Proceedings - Communications* **151**, 355–363 (2004).
8. Empirical data: `http://www.doc.ic.ac.uk/~uh/QUAINT/data/`