Modeling of Flows with Power-Law Spectral Densities and Power-Law Distributions of Flow Intensities

Bronislovas Kaulakys, Miglius Alaburda, Vygintas Gontis, Tadas Meskauskas, and Julius Ruseckas

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

Summary. We present analytical and numerical results of modeling of flows represented as correlated non-Poissonian point process and as Poissonian sequence of pulses of different size. Both models may generate signals with power-law distributions of the intensity of the flow and power-law spectral density. Furthermore, different distributions of the interevent time of the point process and different statistics of the size of pulses may result in $1/f^{\beta}$ noise with $0.5 \lesssim \beta \lesssim 2$. A combination of the models is applied for modeling Internet traffic.

1 Introduction

Modeling and simulations enable one to understand and explain the observable phenomena and predict new ones. This is true, as well, for mathematical studies and modeling of traffic flow with the aim to get a better understanding of phenomena and avoid some problems of traffic congestion. Traffic phenomena are complex and nonlinear, they show cluster formation, huge fluctuations and long-range dependencies. Almost twenty years ago it was detected from empirical data that fluctuations of a traffic current on a expressway obey a $1/f$ law for low spectral frequencies [1]. Similarly, $1/f$ noise is observable in the flows of granular materials [2, 3].

$1/f$ noise, or $1/f$ fluctuations are usually related with power-law distributions of other statistics of the fluctuating signals, first of all with the power-law decay of autocorrelations and the long-memory processes (see, e.g., the comprehensive bibliography of $1/f$ noise on the website [4], review articles [5, 6] and references in the recent paper [7]). The appearance of clustering and large fluctuations in traffic and granular flows may be a result of synchronization of the ensemble of the nonlinear system subjected to common random external perturbations, which may result in nonchaotic behavior of Brownian-type motions, intermittency and $1/f$ noise [8, 9].

Traffic and granular flows usually may be considered as consisting of discrete identical objects such as vehicles, pedestrians, granules, packets and so on. They may be represented as consisting of pulses or elementary events and further simplified to a point process model [7, 10–12]. Moreover, from the modeling of traffic it was found that $1/f$ noise may be the result of clustering and jumping [10] similar to the point process model of $1/f$ noise [7, 11, 12].
On the other hand, $1/f$ noise may be conditioned by the flow consisting of uncorrelated pulses of variable size with a power-law distribution of pulse durations [13]. In Internet traffic the flow of the signals primarily is composed of power-law distributed file sizes. The files are divided by the network protocol into equal packets [14]. Therefore, the total incoming web traffic is a sequence of packets arising from a large number of requests. Such a flow exhibits $1/f$ fluctuations as well [14, 15].
Long-range correlations and power-law fluctuations of expressway traffic flow have recently been observed on a wide range of time-scales from minutes to months and investigated using the method of detrended fluctuation analysis [16]. There are no explanations why traffic flow exhibits $1/f$ noise behavior in such a large interval of time.
It is the purpose of this paper to present analytical and numerical results for the modeling of flows represented as sequences of different pulses and as a correlated non-Poissonian point process resulting in $1/f$ noise and to apply these results to the modeling of Internet traffic.

2 Signal as a Sequence of Pulses

We will investigate a signal of flow consisting of a sequence of pulses,

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

Here the function $A_k(t - t_k)$ represents the shape of the pulse k having influence on the signal $I(t)$ in the region of time t_k.

2.1 Power Spectral Density

The power spectral density of the signal (1) can be written as

$$S(f) = \lim_{T\to\infty} \left\langle \frac{2}{T} \sum_{k,k'} e^{i\omega(t_k - t_{k'})} \int\limits_{t_i - t_k}^{t_f - t_k} \int\limits_{t_i - t_{k'}}^{t_f - t_{k'}} A_k(u) A_{k'}(u') e^{i\omega(u - u')} du du' \right\rangle, \tag{2}$$

where $\omega = 2\pi f$, $T = t_f - t_i \gg \omega^{-1}$ is the observation time and the brackets $\langle \ldots \rangle$ denote the averaging over realizations of the process. We assume that the pulse shape functions $A_k(u)$ decrease sufficiently fast when $|u| \to \infty$. Since $T \to \infty$, the bounds of the integration in Eq. (2) can be changed to $\pm\infty$.

When the time moments t_k are not correlated with the shape of the pulse A_k, the power spectrum is [2]

$$S(f) = \lim_{T\to\infty} \frac{2}{T} \sum_{k,k'} \left\langle e^{i\omega(t_k - t_{k'})} \right\rangle \left\langle \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} A_k(u) A_{k'}(u') e^{i\omega(u-u')} du du' \right\rangle . \tag{3}$$

After introduction of the functions [13]

$$\Psi_{k,k'}(\omega) = \left\langle \int_{-\infty}^{+\infty} A_k(u) e^{i\omega u} du \int_{-\infty}^{+\infty} A_{k'}(u') e^{-i\omega u'} du' \right\rangle \tag{4}$$

and

$$\chi_{k,k'}(\omega) = \left\langle e^{i\omega(t_k - t_{k'})} \right\rangle \tag{5}$$

the spectrum can be written as

$$S(f) = \lim_{T\to\infty} \frac{2}{T} \sum_{k,k'} \chi_{k,k'}(\omega) \Psi_{k,k'}(\omega). \tag{6}$$

2.2 Stationary Process

Equation (6) can be further simplified for the stationary process. Then all averages can depend only on $k - k'$, i.e.,

$$\Psi_{k,k'}(\omega) \equiv \Psi_{k-k'}(\omega) \tag{7}$$

and

$$\chi_{k,k'}(\omega) \equiv \chi_{k-k'}(\omega). \tag{8}$$

Equation (6) then reads

$$S(f) = \lim_{T\to\infty} \frac{2}{T} \sum_{k,k'} \chi_{k-k'}(\omega) \Psi_{k-k'}(\omega). \tag{9}$$

Introducing a new variable $q \equiv k - k'$ and changing the order of summation yields

$$\begin{aligned} S(f) &= \lim_{T\to\infty} \frac{2}{T} \sum_{q=1}^{k_{\max}-k_{\min}} \sum_{k=k_{\min}}^{k_{\max}-q} \chi_q(\omega) \Psi_q(\omega) \\ &+ \lim_{T\to\infty} \frac{2}{T} \sum_{q=k_{\min}-k_{\max}}^{-1} \sum_{k=k_{\min}-q}^{k_{\max}} \chi_q(\omega) \Psi_q(\omega) + \lim_{T\to\infty} \frac{2}{T} \sum_{k=k_{\min}}^{k_{\max}} \Psi_0(\omega). \end{aligned} \tag{10}$$

Here $k_{\min}$ and $k_{\max}$ are minimal and maximal values of the index k in the interval of observation T. Eq. (10) may be simplified to the structure

$$S(f) = 2\bar{\nu}\Psi_0(\omega) + \lim_{T\to\infty} 4\sum_{q=1}^{N}\left(\bar{\nu} - \frac{q}{T}\right)\mathrm{Re}\,\chi_q(\omega)\Psi_q(\omega) \tag{11}$$

where $\bar{\nu}$ is the mean number of pulses per unit time and $N = k_{\max} - k_{\min}$ is the number of pulses in the time interval T.
If the sum $\frac{1}{T}\sum_{q=1}^{N} q\,\mathrm{Re}\,\chi_q(\omega)\Psi_q(\omega) \to 0$ when $T \to \infty$, then the second term in the sum vanishes and the spectrum is

$$S(f) = 2\bar{\nu}\Psi_0(\omega) + 4\bar{\nu}\sum_{q=1}^{\infty}\mathrm{Re}\,\chi_q(\omega)\Psi_q(\omega) = 2\bar{\nu}\sum_{q=-\infty}^{\infty}\chi_q(\omega)\Psi_q(\omega). \tag{12}$$

2.3 Fixed Shape Pulses

When the shape of the pulses is fixed (k-independent) then the function $\Psi_{k,k'}(\omega)$ does not depend on k and k' and, therefore, $\Psi_{k,k'}(\omega) = \Psi_{0,0}(\omega)$. Then equation (6) yields the power spectrum

$$S(f) = \Psi_{0,0}(\omega)\lim_{T\to\infty}\frac{2}{T}\sum_{k,k'}\chi_{k,k'}(\omega) \equiv \Psi_{0,0}(\omega)S_\delta(\omega). \tag{13}$$

Eq. (13) represents the spectrum of the process as a composition of the spectrum of one pulse,

$$\Psi_{0,0} = \left|\int_{-\infty}^{+\infty} A_k(t)e^{i\omega t}dt\right|^2, \tag{14}$$

and the power density spectrum $S_\delta(\omega)$ of the point process

$$I_\delta(t) = a\sum_k \delta(t - t_k) \tag{15}$$

with the area of the pulse $a = 1$.

3 Stochastic Point Processes

The shapes of the pulses mainly influence the high frequency power spectral density, i.e., at $\omega \geq 1/\Delta t_p$, with Δt_p being the characteristic pulse length. Therefore the power spectral density at low frequencies for not very long pulses is mainly conditioned by the correlations between the transit times t_k, i.e., the signal may be approximated by the point process.
The point process model of $1/f^\beta$ noise has been proposed [11, 12], generalized [7], analysed and used for financial systems [17]. It has been shown that when the average interpulse, interevent, interarrival, recurrence or waiting times $\tau_k = t_{k+1} - t_k$ of the signal diffuse in some interval, the power spectrum

of such process may exhibit the power-law dependence, $S_\delta(f) \sim 1/f^\beta$, with $0.5 \lesssim \beta \lesssim 2$. The distribution density of the signal (15) intensity defined as $I = 1/\tau_k$ may be of the power-law, $P(I) \sim I^{-\lambda}$, with $2 \leqslant \lambda \leqslant 4$, as well. The exponents β and λ are depending on the manner of diffusion-like motion of the interevent time τ_k and, e.g., for the multiplicative process are interrelated [7, 17]. For the pure multiplicative process [7]

$$\beta = 1 + \alpha, \quad \lambda = 3 + \alpha, \tag{16}$$

where α is the exponent of the power-law distribution, $P_k(\tau_k) \sim \tau_k^\alpha$, of the interevent time. In general, for relatively slow fluctuations of τ_k, the distribution density of the flow I,

$$P(I) \sim P_k(I^{-1})I^{-3}, \tag{17}$$

is mostly conditioned by the multiplier I^{-3}. Since the point process model has recently [7, 17] been analysed rather properly we will not repeat the analysis here and present only some new illustrations.

Figure 1 demonstrates that for essentially different distributions of τ_k, the power spectra and distribution densities of the point processes are similar.

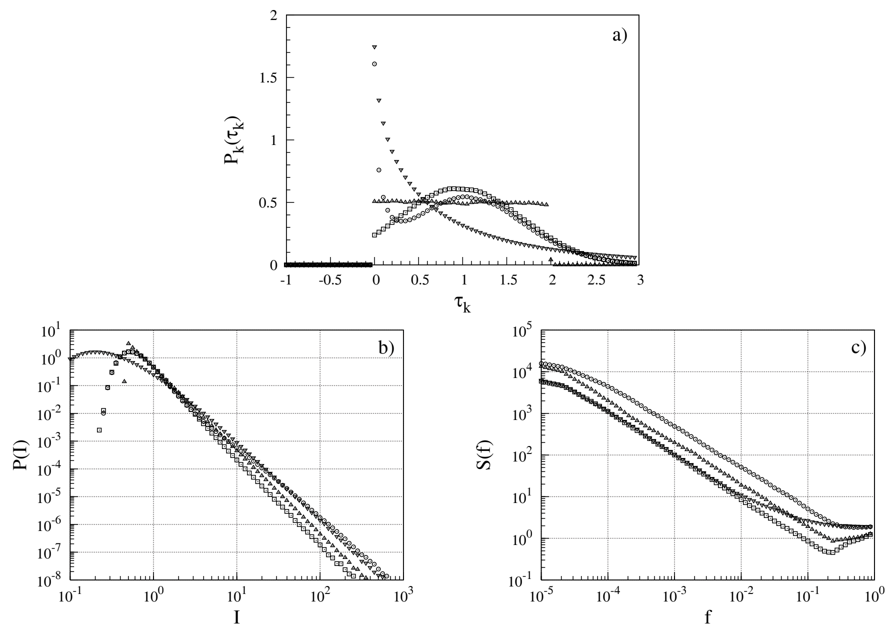

Fig. 1. Distribution densities of the interevent time τ_k, (a), of the flow $I(t)$, (b), and of the power spectra $S(f)$, (c), for different point processes with slow diffusion-like motion of the average interevent time. Different symbols correspond to different types of generation of the interevent sequences.

Further we proceed to the flow consisting of the pulses of different durations and application of this approach for modeling of the Internet traffic.

4 Flow Consisting of Pulses of Variable Duration

When the occurrence times t_k of the pulses are uncorrelated and distributed according to a Poisson process, the power spectrum of the random pulse train is given by Carlson's theorem

$$S(f) = 2\bar{\nu} \left\langle |F_k(\omega)|^2 \right\rangle , \tag{18}$$

where

$$F_k(\omega) = \int_{-\infty}^{+\infty} A_k(t) e^{i\omega t} dt \tag{19}$$

is the Fourier transform of the pulse A_k. Suppose that the random parameters of the pulses are the duration and the area (integral) of the pulse. We can take the form of the pulses as

$$A_k(t - t_k) = T_k^{\rho} A \left(\frac{t - t_k}{T_k} \right) , \tag{20}$$

where T_k is the characteristic duration of the pulse. The value of the exponent $\rho = 0$ corresponds to the fixed height but different durations, the telegraph-like pulses, whereas $\rho = -1$ corresponds to constant area pulses but of different heights and durations, and so on.
For the power-law distribution of the pulse durations,

$$P(T_k) = \begin{cases} \frac{\delta+1}{T_{\max}^{\delta+1} - T_{\min}^{\delta+1}} T_k^{\delta}, & T_{\min} \le T_k \le T_{\max}, \\ 0, & \text{otherwise}, \end{cases} \tag{21}$$

from Eqs. (18) and (19) we have the spectrum

$$S(f) = \frac{2\bar{\nu}(\delta + 1)}{(T_{\max}^{\delta+1} - T_{\min}^{\delta+1}) \omega^{\delta+2\rho+3}} \int_{\omega T_{\min}}^{\omega T_{\max}} |F(u)|^2 u^{\delta+2\rho+2} du. \tag{22}$$

For $\tau_{\max}^{-1} \ll \omega \ll \tau_{\min}^{-1}$ when $\delta > -1$ the expression (22) may be approximated as

$$S(f) \approx \frac{2\bar{\nu}(\delta + 1)}{(T_{\max}^{\delta+1} - T_{\min}^{\delta+1}) \omega^{\delta+2\rho+3}} \int_{0}^{\infty} |F(u)|^2 u^{\delta+2\rho+2} du. \tag{23}$$

Therefore, the random pulses with the appropriate distribution of the pulse duration (and area) may generate signals with the power-law distribution of

the spectrum with different slopes. So, the pure $1/f$ noise generates, e.g., the fixed area ($\rho = -1$) with the uniform distribution of the durations ($\delta = 0$) sequences of pulses, the fixed height ($\rho = 0$) with the uniform distribution of the inverse durations $\gamma = T_k^{-1}$ and all other sequences of random pulses satisfying the condition $\delta + 2\rho = -2$.
In such a case we have from Eq. (23)

$$S(f) \sim \frac{(\delta+1)\bar{\nu}}{(T_{\max}^{\delta+1} - T_{\min}^{\delta+1})f}. \tag{24}$$

5 Internet Traffic

In this Section we will apply the results of Section 4 for modeling Internet traffic. The incoming traffic consists of a sequence of packets, which are the result of the division of the requested files by the network protocol (TCP). The maximum size of a packet is 1500 bytes. Therefore, the information signal is as in the point process (15) with pulse area $a = 1500$ bytes. Further, we will analyse the flow of the packets and will measure the intensity of the flow in packets per second. In such a system of units in Eq. (15) we should put $a = 1$.
We exploit the empirical observation [14, 18] that the distribution of the file sizes x may be described by the positive Cauchy distribution

$$P(x) = \frac{2}{\pi}\frac{s}{s + x^2} \tag{25}$$

with the empirical parameter $s = 4100$ bytes. This distribution asymptotically exhibits the Pareto distribution and follows Zipf's law $P\,(X > x) \sim 1/x$. The files are divided into packets of a maximum size of 1500 bytes or less by the network protocol. In Internet traffic the packets spread into the Poissonian sequence with average inter-packet time τ_p (see Fig. 2). The total incoming flow of the packets to the server consists of packets arising from the Poissonian request of the files with average interarrival time of files τ_f.
The files are requested from different servers located at different distance. This results in the distribution of the average inter-packet time τ_p in some interval. For reproduction of the empirical distribution of the interpacket time τ_k we assume the uniform distribution of $\lg \tau_k$ in some interval $[\tau_{k,min}, \tau_{k,max}]$, similarly to the McWhorter model of $1/f$ noise [7]. As a result, the presented model reproduces sufficiently well the observable non-Poissonian distribution of the arrival interpacket times and the power spectral density, as well (see Fig. 3).

6 Conclusion

In the paper it was shown that processes exhibiting $1/f$ noise and power-law distribution of the intensity may be generated starting from the signals as

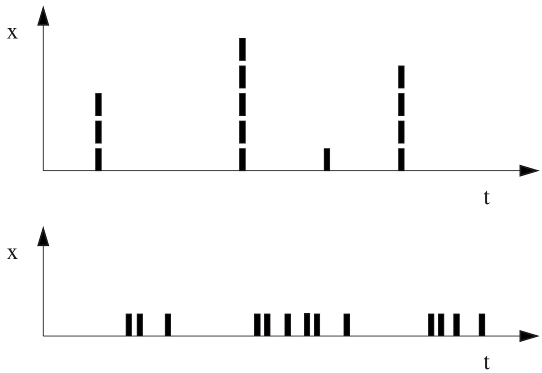

Fig. 2. Division of the requested files into equal size packets with some inter-packet time.

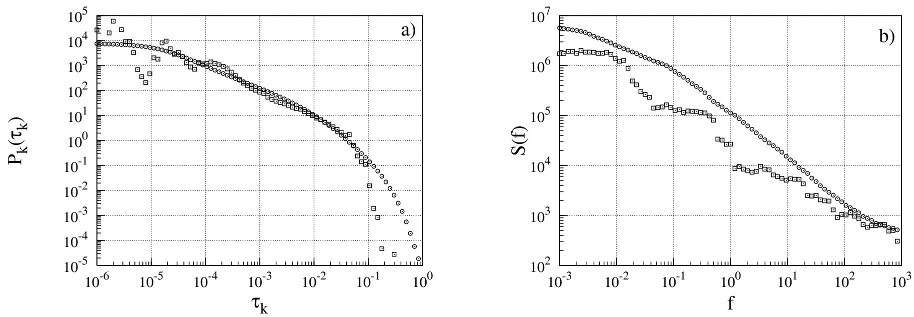

Fig. 3. Distribution densities of (a) the interpacket time τ_k, and (b) the power spectra, for the simulated point process (open circles ∘) and empirical data (open squares □). The used parameters are as in the empirical data [14, 18], $\tau_f = 0.101s, \tau_{k,min} = 11.6\mu s$ and $\tau_{k,max} = 1000\ \tau_{k,min}$.

sequences of constant area pulses with correlated appearance times as well as of different size Poissonian pulses. Combination of both approaches enables the modeling of signals in Internet traffic.

Acknowledgements

Support by the Lithuanian State Science and Studies Foundation is acknowledged.

References

1. T. Musha, H. Higuchi: Jpn. J. Appl. Phys. **15**, 1271 (1976);
2. K. L. Schick, A. A. Verveen: Nature **251**, 599 (1974).
3. G. Peng, H. J. Herrman: Phys. Rev. E **51**, 1745 (1995).
4. W. Li: www.nslij-genetics.org/wli/1fnoise.
5. M. B. Weismann: Rev. Mod. Phys. **60**, 537 (1988).
6. H. Wong: Microelectron. Reliab. **43**, 585 (2003).
7. B. Kaulakys, V. Gontis, M. Alaburda: Phys. Rev. E **71**, 051105 (2005).
8. B. Kaulakys, G. Vektaris: Phys. Rev. E **52**, 2091 (1995).
9. B. Kaulakys, F. Ivanauskas, T. Meskauskas: Int. J. Bifurcation and Chaos **9**, 533 (1999).
10. X. Zhang, G. Hu: Phys. Rev. E **52**, 4664 (1995).
11. B. Kaulakys, T. Meskauskas: Phys. Rev. E **58**, 7013 (1998).
12. B. Kaulakys: Phys. Lett. A **257**, 37 (1999).
13. J. Ruseckas, B. Kaulakys, M. Alaburda: Lith. J. Phys **43**, 223 (2003).
14. A. J. Field, U. Harder, P. G. Harrison: IEE Proc.-Commun. **151**, 355 (2004).
15. V. Gontis, B. Kaulakys, J. Ruseckas: AIP Conf. Proceed. **776**, 144 (2005).
16. S. Tadaki et al: these proceedings and private communication.
17. V. Gontis, B. Kaulakys: Physica A **343**, 505 (2004); **344**, 128 (2004).
18. A. J. Field, U. Harder, P. G. Harrison: http://www.doc.ic.ac.uk/∼uh/QUAINT/data/.