Statistica Neerlandica (2000) Vol. 54, nr. 3, pp. 293–314

Heavy tails versus long-range dependence in self-similar network traffic

A. Stegeman*

University of Groningen, Department of Mathematics, P.O. Box 800, NL-9700 AV Groningen

Empirical studies of the traffic in computer networks suggest that network traffic exhibits *self-similarity* and *long-range dependence.* The ON/OFF model considered in this paper gives a simple 'physical explanation' for these observed phenomena. The superposition of a large number of ON/OFF sources, such as workstations in a computer lab, with strictly alternating and heavy-tailed ON- and OFF-periods, can produce a cumulative workload which converges, in a certain sense, to fractional Brownian motion. Fractional Brownian motion exhibits both self-similarity and long-range dependence. However, there are two sequential limits involved in this limiting procedure, and if they are reversed, the limiting process is stable Lévy motion, which is self-similar but exhibits no long-range dependence. We study simulations limit regimes and provide conditions under which either fractional Brownian motion or stable Lévy motion appears as limiting process.

Key Words and Phrases: self-similarity, ON/OFF model, fractional Brownian motion, stable Lévy motion, teletraffic.

1 Introduction

Computers play a predominant role in the modern society. Networks such as the World Wide Web (WWW) have made it possible to access vast amounts of information at the touch of a button. However, as any surfer on the Web will have experienced, transmission times can be extremely long. In order to improve upon network performance, the characteristics of network traffic have been studied. This has been done in networks such as Ethernet LANs (Local Area Networks) or the WWW. Recent measurements and theoretical analysis of data traffic have shown the presence of three characteristic phenomena:

- *heavy tailed distributions*
- *self-similarity*
- *long-range dependence (LRD)*

(In Section 2.1 we will explain in detail what we mean by those notions.) This

* a.w.stegeman@math.rug.nl

implies that traditional traffic models, based on classical queuing theory with exponential inter-arrival times, are not appropriate for describing high-speed network traffic, a conclusion which can be found, for example in FOWLER and LELAND (1991) and in PAXTON and FLOYD (1995). Empirical evidence on the existence of heavy tails, self-similarity and LRD in traffic measurements was further provided in the studies by LELAND et al. (1993), CROVELLA and BESTAVROS (1996), CROVELLA et al. (1996) and WILLINGER et al. (1995).

To understand why heavy tails, self-similarity and LRD are present in the traffic data, WILLINGER et al. (1995) considered a simple ON/OFF model. In this model, traffic is generated by a large number of independent ON/OFF sources such as workstations in a large computer network. An ON/OFF source transmits data at a constant rate if it is ON and remains silent if it is OFF. Every individual ON/OFF source generates an ON/OFF process consisting of alternating ON- and OFF-periods. The lengths of the ON-periods are identically distributed and so are the lengths of OFF-periods. Moreover, the sequences of lengths of ON- and OFF-periods are supposed to be independent. WILLINGER et al. (1995) provide an exploratory statistical analysis of Ethernet LAN traffic of individual sources and conclude that the lengths of the ON- and OFF-periods are heavy-tailed in the sense that the distributions of those lengths are Pareto-like with tail parameters between 1 and 2; see Section 2.1. In particular, the lengths of the ON- and OFF-periods have finite means but infinite variances. This fact is further supported by empirical research in LELAND et al. (1993) and CROVELLA and BESTAVROS (1996). The latter authors studied the traffic on the World Wide Web. They found evidence of Pareto-like tails in file lengths, transfer times and idle times. See also CROVELLA et al. (1996).

HEATH et al. (1998) studied the ON/OFF model at the source level. They constructed a stationary version of the ON/OFF process of an individual source. Assuming heavy-tailed (Pareto-like) lengths of ON- and OFF-periods, they showed that the ON/OFF process of an individual source necessarily exhibits LRD, see again Section 2.1 for a precise definition of this notion.

WILLINGER et al. (1995) studied the superposition of many iid ON/OFF sources. They focused on the cumulative workload process which is the aggregate network traffic through time. Their main result is that the cumulative workload process (properly normalized) of an increasing number of iid ON/OFF sources converges to fractional Brownian motion (see Section 2.3) in the sense of convergence of the finite−dimensional distributions. Their result involves a double limit: first, the number of sources goes to infinity and then they let a time-scaling parameter converge to infinity. This order of taking limits is crucial for obtaining fractional Brownian motion as a limiting process. Indeed, when limits are taken in reversed order, TAQQU et al. (1997) showed that the limits of the finite−dimensional distributions are those of infinite variance stable Lévy motion (see again Section 2.3). The increment process of fractional Brownian motion, fractional Gaussian noise, exhibits LRD reflecting the LRD in the original workload process. This is in contrast to stable Lévy motion whose increments are independent: LRD disappears in the limit.

In both, the results of WILLINGER et al. (1995) and TAQQU et al. (1997), a double limit is involved and the limit regime is sequential. In practice, the behavior of the cumulative workload process depends on the relative sizes of the number of sources and the time-scaling parameter. We study simultaneous limit regimes, in which both parameters go to infinity at the same time. We provide conditions on their relative speeds in order to ensure that the limit process is either stable Lévy motion or fractional Brownian motion.

The paper is organised as follows. In Section 2 we give definitions of heavy tails, self-simlarity and LRD and discuss some of the methods used to observe these phenomena in network traffic data. We also define fractional Brownian motion and stable Lévy motion. In Section 3 we give a firm definition of the ON/OFF model and provide the necessary assumptions. In Section 4 we present the results of WILLINGER et al. (1995) and TAQQU et al. (1997). In Section 5 we present our main results, involving simultaneous limit regimes, which were proved by MIKOSCH and STEGEMAN (1999). A preliminary analysis was performed in STEGEMAN (1998). In Section 6 we give a sketch of the proof.

2 Traffic characteristics

2.1 *Heavy tails, self-similarity and long-range dependence*

First we introduce the concept of heavy tails. If (a_n) and (b_n) are real sequences, we use the notation $a_n \sim b_n$ to denote $\lim_{n\to\infty} a_n/b_n = 1$.

DEFINITION 1. *We say that a random variable Z has a* heavy right (or left) tail *if*

$$P(Z > z) \sim c_1 z^{-\alpha}, \quad \text{or } P(Z \leqslant -z) \sim c_2 z^{-\alpha}, \quad \text{as } Z \to \infty, \tag{1}$$

respectively, where $\alpha \in (0, 2)$ and c_1, c_2 are positive constants. A random variable Z is heavy tailed *if it has a heavy left or right tail.*

REMARK 1. It is common use to define heavy tails are regularly varying. For ease of presentation we restrict ourselves to the special case (1).

Notice that if Z is heavy tailed, then $\mathrm{Var}(Z) = \infty$. If Z is heavy tailed with tail parameter $\alpha < 1$, then also $E|Z| = \infty$. An example of a heavy tailed distribution is the Pareto distribution, which is defined by

$$P(Z > z) = \left(\frac{\kappa}{\kappa + z}\right)^{\alpha}, \quad z \geqslant 0, \quad \text{with } \kappa > 0 \text{ and } \alpha \in (0, 2).$$

Other examples of such distributions are infinite variance stable distributions (see Section 2.3) and the infinite variance Fréchet distributions (see for example EMBRECHTS et al., 1997).

Next we define self-similarity.

DEFINITION 2. *A real-valued stochastic process* $(Z(t),\ t \geqslant 0)$ *is* self-similar with parameter $H > 0$ *if for all* $a > 0$,

$$(a^{-H} Z(at),\ t \geqslant 0) \stackrel{d}{=} (Z(t),\ t \geqslant 0), \tag{2}$$

which means that the finite-dimensional distributions of the processes in (2) are identical.

Notice that the distribution of a self-similar process is invariant under rescaling both in time and space, as implied by (2). The perhaps best known example of a self-similar process is Brownian motion: recall that a Gaussian, mean-zero process $(B(t))_{t \geqslant 0}$ is called standard *Brownian motion* if it has stationary, independent increments, continuous sample paths with probability 1 and variance $\mathrm{Var}(B(t)) = t$. Since $\mathrm{Cov}(B(s),\ B(t)) = \min(s,\ t)$ it is not difficult to see that Brownian motion is self-similar with parameter $H = 1/2$.

Next we define long-range dependence (LRD).

DEFINITION 3. *Let* $(Z_n,\ n \geqslant 0)$ *be a real-valued, stationary, finite-variance stochastic process with autocorrelation function* $\rho(k)$*, i.e.*

$$\rho(k) = \frac{\mathrm{Cov}(Z_n,\ Z_{n+k})}{\mathrm{Var}(Z_n)}, \quad k = 0,\ 1,\ 2,\ \ldots.$$

Then Z exhibits long-range dependence (LRD) *if*

$$\sum_k |\rho(k)| = \infty. \tag{3}$$

Instead of LRD the term *long memory* is frequently used. In view of (3), a process has LRD if the autocorrelations are not absolutely summable. If the autocorrelations are absolutely summable the process is said to have *short-range dependence* or *short-memory.* This is true for the most frequently applied class of time series models, the ARMA processes.

In the literature, LRD is sometimes defined by describing the rate at which $\rho(k)$ decreases to zero as $k \to \infty$. Following this approach, a stationary process has LRD if

$$\rho(k) \sim c_\rho k^{-\beta} \quad \text{as } k \to \infty, \tag{4}$$

where c_ρ is a positive constant and $\beta \in (0,\ 1)$. Clearly, (4) implies (3).

An analogous way to define LRD is through the spectral density. The *spectral density* of a stationary, finite variance process $(Z_n,\ n \geqslant 0)$ is defined as (see BROCKWELL and DAVIS, 1991, Section 4.3)

$$f(\lambda) = \frac{\sigma^2}{2\pi} \sum_k \rho(k) \mathrm{e}^{-\mathrm{i}k\lambda} \quad \text{for } \lambda \in [-\pi,\ \pi],$$

where $\sigma^2 = \mathrm{Var}(Z_n)$. Since the autocorrelations are not absolutely summable under LRD, the spectral density has a singularity at $\lambda = 0$. One can show (under some technical conditions) that (4) is equivalent to

$$f(\lambda) \sim c_f \lambda^{\beta-1} \quad \text{as } \lambda \to 0, \tag{5}$$

where c_f is a positive constant depending on c_ρ.

LRD in a stochastic process has a substantial impact on the variance of the sample mean

$$\overline{Z}_N := \frac{1}{N}\sum_{j=1}^{N} Z_j.$$

If (4) or, equivalently, (5) holds, then as $N \to \infty$

$$\mathrm{Var}(\overline{Z}_N) \sim c_v\, \sigma^2 N^{-\beta}, \tag{6}$$

where β is the same as in (4) and c_v is a positive constant depending on β and c_ρ. For a short memory process, $\mathrm{Var}(\overline{Z}_N)$ would decrease proportionally to N^{-1}, whereas under LRD, the variance decreases at a slower rate.

According to Definition 3 only a *stationary* process can exhibit LRD. Since there are no reliable statistical tools for testing the stationarity of a real-life time series, the question arises as to whether slowly decreasing autocorrelations might be the result of non-stationarities. It turns out that this can indeed be the case. For example, TEVEROVSKY and TAQQU (1995) include shifting means and slowly declining trends into their models and show that the sample autocorrelation function of the resulting non-stationary model behaves like the autocorrelation function of a stationary process with LRD. Therefore, without additional information on the stationarity of time series, it is not justified to conclude LRD from the various graphical or statistical methods available.

The term LRD is used to indicate that the interplay of events that are far apart in time is not negligible. We use Definition 3 to define LRD, but, as we have tried to indicate, there are plenty of other definitions available. Several alternatives can be found in the monograph by BERAN (1994). We mention at this point that the best known examples of processes with LRD are fractional ARIMA models and fractional Gaussian noise (see Section 2.3).

2.2 *Traffic data analysis*

In this section we discuss some of the statistical methods that are used to detect the presence of heavy tails, self-similarity and LRD.

CROVELLA et al. (1996) have found evidence of heavy tails in the distributions of file lengths and transmission times in their empirical study of the WWW. They use two graphical methods for estimating the tail parameter α in (1). These methods are also used by WILLINGER et al. (1995) to find evidence of heavy tails in activity periods and idle times of individual computers.

The first method is called the log-log *complementary distribution* (LLCD) plot. Let F_n denote the empirical distribution function of n observed file lengths and let $\overline{F}_n = 1 - F_n$ denote the 'complementary' distribution, i.e. its right tail. An LLCD plot shows $\log(\overline{F}_n(x))$ versus $\log(x)$. In the heavy tailed case, $\log(\overline{F}_n)$ should approximate a straight line with slope $-\alpha$ for large values of x. In this way an estimate for α can be obtained. This method is also used in CROVELLA and BESTAVROS (1996).

The second method is based on the *Hill* estimator. Suppose $Z_1, \ldots, Z_n$ are iid with a heavy tailed distribution function F and tail parameter α. Denote the order statistics by $Z_{n,n} \leqslant \cdots \leqslant Z_{1,n}$. The Hill estimator of α uses the k largest order statistics to give the estimate

$$\hat{\alpha}_k = \left(\left(\frac{1}{k} \sum_{j=1}^{k} \log(Z_{j,n}) \right) - \log(Z_{k,n}) \right)^{-1}.$$

Since $k = k_n \to \infty$ and $k/n \to 0$ are necessary conditions for consistency of $\hat{\alpha}_k$, the estimator $\hat{\alpha}_k$ is plotted against k for a variety of values k which are small compared to n. The graphical output of this procedure is called the *Hill plot*. In the heavy tailed case the estimator stabilizes at a level $\hat{\alpha}$ for certain values of k. This level $\hat{\alpha}$ is taken as an estimate for the tail parameter α. In Figure 2 a Hill plot is shown for realisations of stable random variables. In practice there are some serious problems concerning the accuracy of the Hill estimator and related tail-estimation techniques, see e.g. RESNICK (1997) or EMBRECHTS et al., 1997, Section 6.4.

Next we discuss the detection of self-similarity in the workload data. The workload data consist of measurements of the number of bytes or *packets* that arrive on the network per time unit. A packet consists of a collection of bytes that belong to the same file. If a file is sent through the network, it is decomposed into several packets. Usually, the workload is measured for a couple of hours with a time unit of several microseconds (1 second contains 10^6 microseconds). This procedure yields very large data sets. In Figure 1 packet counts are depicted from measurements on the Ethernet LAN at Bellcore in August 1989. These traffic data were also used in LELAND and WILSON (1991) and LELAND et al. (1993). Traffic data can be obtained from `http: /www.acm.org/sigcom/ITA`.

There are no reliable statistical tests for self-similarity as such, unless one specifies a parametric models, estimates its parameters from the data and obtains the self-similarity parameter from the estimated parameters, possibly via a functional relationship. In practice one usually depends on a 'pictorial proof'. In LELAND et al. (1993) the workload data are shown on different time scales. The assumtion of self-similarity is justified on the basis of the observation that there are no dramatic changes in the relative variability of the data. This is consistent with the definition of self-similarity in Definition 2. However, we should keep in mind that self-similarity means that the *distribution* of the process is invariant under rescaling in time and

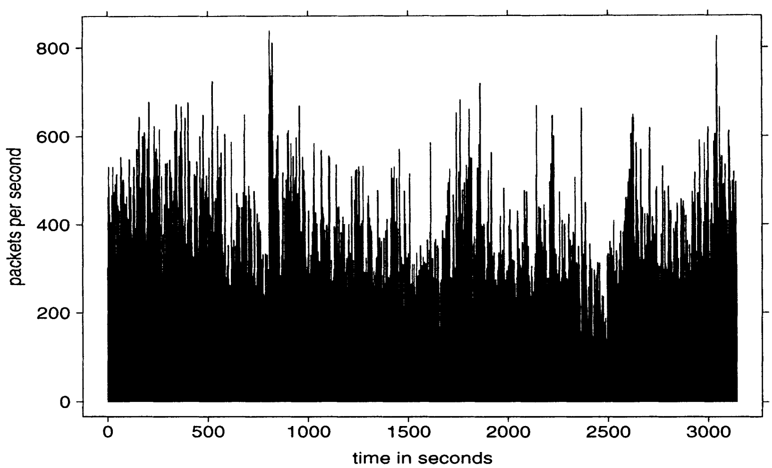

Fig. 1. Packet counts of Ethernet LAN traffic measurements at Bellcore, August 1989.

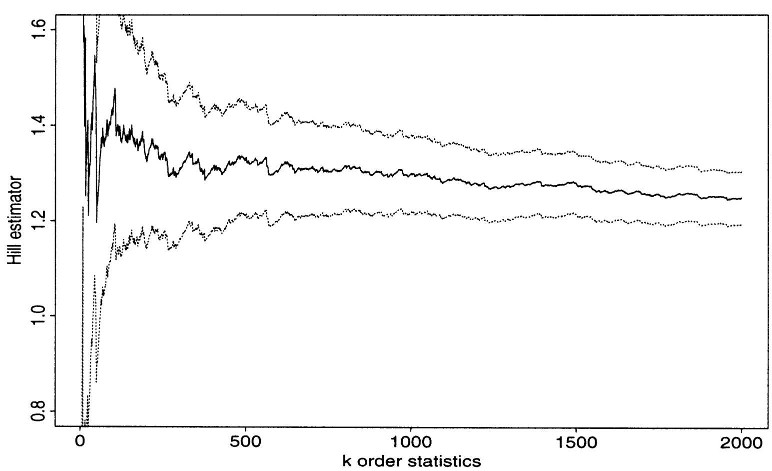

Fig. 2. The Hill plot with asymptotic 95% confidence bands for 100 000 simulated α-stable random variables with tail parameter $\alpha = 1.2$.

space and *not the sample paths.* The 'pictorial proof' can be considered as a handy graphical tool, but nothing else.

Finally, we consider the detection of LRD in the workload data. Since an infinite sum of autocorrelations is impossible to verify in practice, usually (4) is used to define LRD. There are several methods to detect the presence of LRD in a time series and to estimate the parameter β. A detailed description can be found in BERAN (1994). For a performance analysis of various estimators we refer to TAQQU and TEVEROVSKY (1995, 1996). Empirical studies of LRD in workload data are given in

LELAND et al. (1993) and CROVELLA and BESTAVROS (1996). Here we described three exploratory statistical methods which are used to detect LRD as suggested in (4)–(6).

The first method uses the sample autocorrelation function $\hat{\rho}(k)$ and plots $\log(\hat{\rho}(k))$ against $\log(k)$. This is called the log–log *correlogram* plot. If LRD in the sense of (4) is present, the points of this plot should be randomly scattered around a straight line with slope $-\beta$ for sufficiently large values of k. A disadvantage of this method is that, for large k, relatively to the sample size of n, the estimte $\hat{\rho}(k)$ is unreliable (see BROCKWELL and DAVIS, 1991, Section 7.2). In Figure 3 the sample autocorrelation function and the log–log correlogram of the Ethernet packet counts are shown. The autocorrelation function decreases to zero very slowly, which gives rise to the belief that LRD is present in the data. Linear regression in the log–log correlogram yields the estimate $\hat{\beta} \approx 0.41$ of the parameter β in (4).

The second method focuses on the spectral density. If LRD is present, the spectral density has a peak at zero. A natural estimate for the spectral density is the *periodogram* $I(\lambda)$ (see BROCKWELL and DAVIS, 1991, Section 10.3). To detect the presence of LRD in the sense of (5), $\log(I(\lambda))$ is plotted against $\log(\lambda)$ for frequencies λ close to zero. This is called the log–log *periodogram* plog. If the points are randomly scattered around a straight line with slope between zero and -1, this indicates that LRD in the sense of (5) might be present. However, if the autocorrelations are not absolutely summable, the periodogram fluctuates much more for frequencies tending to zero, than for short memory processes (see BERAN, 1994, Theorem 3.8). In Figure 4, a smoothed log periodogram and the corresponding log–log periodogram of the Ethernet packet counts are shown. Clearly, the periodogram has a peak at zero. Linear regression in the log–log periodogram yields the estimate $\hat{\beta} \approx 0.38$ of the parameter β in (5). The small amount of points for small frequencies indicates that reliable estimates of β can be expected *only* for huge sample sizes. Also, notice that the residuals of a least squares regression in the log–log correlogram or the log–log periodogram are not uncorrelated. This means that the standard assumptions on least squares regression do not apply.

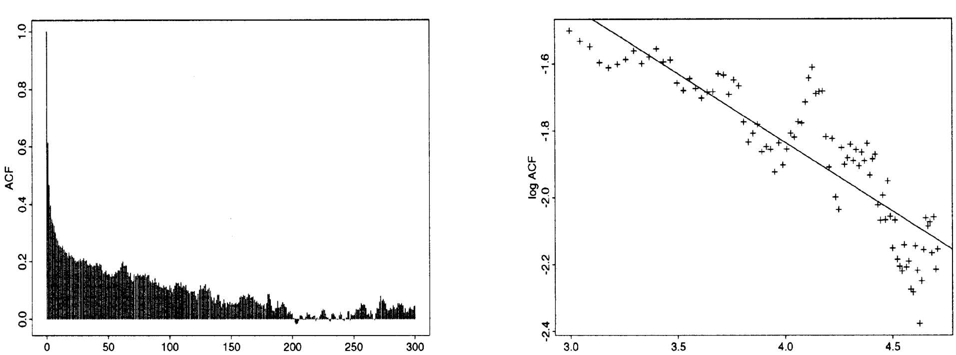

Fig. 3. The sample autocorrelation function (left) and the log–log correlogram (right) of the Ethernet packet counts.

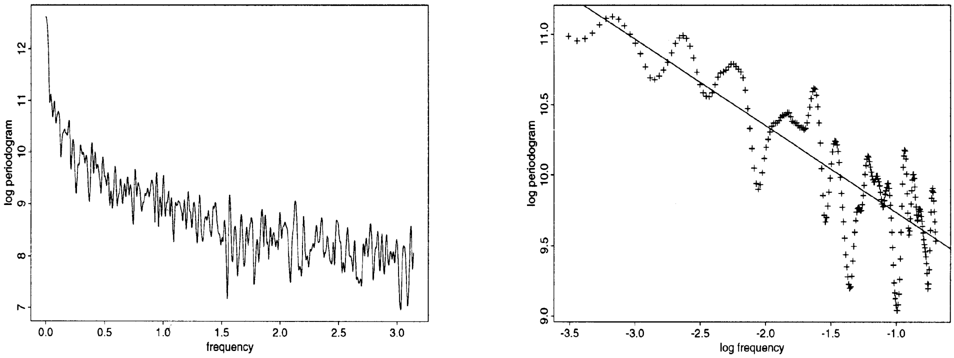

Fig. 4. A smoothed log periodogram (left) and the corresponding log–log periodogram (right) of the Ethernet packet counts.

The third method is used to detect LRD in the sense of (6). Suppose we have workload measurements $z_1, \ldots, z_N$. The procedure is to plot the variance of the sample mean $\bar{z}_n$ against $n (n \leqslant N)$ on log–log scales. This is called the *variance-time* plot. If, for large n, the points are scattered around a straight line with slope between zero and -1, an indication of the presence of LRD has been found. The variance of the sample mean $\overline{Z}_n$ is approximated as follows. The data are split up into blocks of size m (with $m \ll n$). For the kth block we compute the mean value $\bar{z}_k^{(m)}$, with

$$\bar{z}_k^{(m)} = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} z_i, \qquad k = 1, 2, \ldots, [n/m].$$

The sample variance of $\{\bar{z}_1^{(m)}, \ldots, \bar{z}_{[n/m]}^{(m)}\}$ is taken as an approximation of the variance of $\bar{z}_n$.

2.3 *Fractional Brownian motion and stable Lévy motion*

In this section we define the stochastic processes fractional Brownian motion and stable Lévy motion. They can be obtained as limit processes for the cumulative workload process in the ON/OFF model, which is described in Section 3. We start with the definition of fractional Brownian motion.

DEFINITION 4. *Standard* fractional Brownian motion $(B_H(t), t \geqslant 0)$ *is a Gaussian process with mean* $E(B_H(t)) = 0$ *and autocovariance function*

$$\mathrm{Cov}(B_H(s), B_H(t)) = \tfrac{1}{2}(s^{2H} + t^{2H} - |t - s|^{2H}),$$

where $H \in (0, 1)$.

We refer to the process $(\sigma_0 B_H(t), t \geqslant 0)$, with $\sigma_0 > 0$, as fractional Brownian motion. Fractional Brownian motion has stationary increments and is self-similar with parameter H. The parameter H is called the Hurst parameter after the hydrologist Hurst, who in 1951 found evidence for LRD in a data set of annual minimal water levels of the River Nile. For $H = 1/2$ fractional Brownian motion becomes ordinary Brownian motion and hence, has independent increments. The increment process $(B_H(n) - B_H(n-1),\ n = 1, 2, \ldots)$ is called *fractional Gaussian noise* and has LRD in the sense of (4) if $H \in (1/2, 1)$. If $H \in (0, 1/2]$ then fractional Gaussian noise has short-range dependence.

Stable Lévy motion is a process with α-stable finite-dimensional distributions. The univariate stable distribution is denoted by $S_\alpha(\sigma, \beta, \mu)$, where $\alpha \in (0, 2]$ is the index of stability, $\sigma > 0$ is the scale parameter, $\beta \in [-1, 1]$ is the skewness parameter and $\mu \in \mathbb{R}$ is the shift parameter. If $Z \sim S_\alpha(\sigma, \beta, \mu)$, then its characteristic function is given by

$$E\,\mathrm{e}^{\mathrm{i}\theta Z} = \begin{cases} \exp\left\{-\sigma^\alpha|\theta|^\alpha\left(1 - \mathrm{i}\beta\,\mathrm{sign}(\theta)\tan\left(\dfrac{\pi\alpha}{2}\right)\right) + \mathrm{i}\mu\theta\right\} & \text{if } \alpha \neq 1, \\ \exp\left\{-\sigma|\theta|\left(1 + \mathrm{i}\beta\dfrac{2}{\pi}\,\mathrm{sign}(\theta)\ln|\theta|\right) + \mathrm{i}\mu\theta\right\} & \text{if } \alpha = 1. \end{cases}$$

The case $\alpha = 2$ corresponds to the Gaussian distribution. For $\alpha < 2$, Z has infinite variance.

Stable Lévy motion is defined as follows.

DEFINITION 5. *Stable Lévy motion* $(\Lambda_{\alpha,\sigma,\beta}(t),\ t \geqslant 0)$ *is a process with stationary, independent increments, cadlag sample paths and marginal distributions* $\Lambda_{\alpha,\sigma,\beta}(t) \sim S_\alpha(\sigma t^{1/\alpha}, \beta, 0)$, *where* $\alpha \in (0, 2]$, $\sigma > 0$ *and* $\beta \in [-1, 1]$.

REMARK 2. *Cadlag* is a French acronym (continue à droite, limites à gauche). A function is cadlag if it is continuous from the right and has left limits.

Stable Lévy motion is Brownian motion for $\alpha = 2$. Stable Lévy motion is self-similar with parameter $1/\alpha$.

More properties of fractional Brownian motion and stable Lévy motion can be found in the monograph by SAMORODNITSKY and TAQQU (1994).

In Figure 5 sample paths of fractional Brownian motion and stable Lévy motion are shown. Fractional Brownian motion has continuous and nondifferentiable sample paths. As H gets closer to 1, the sample paths become smoother. If $H \in (1/2, 1)$ the LRD in the increments of the process can cause long up- or downward movements. This is in contrast to Brownian motion ($H = 1/2$), whose increments are independent. Stable Lévy motion with $\alpha < 2$ has discontinuous sample paths (although the sample path in Figure 5 looks continuous due to Matlab). Large jumps are possible since the process has infinite variance (provided $\alpha < 2$). The independent increments cause the rapid up and down fluctuations.

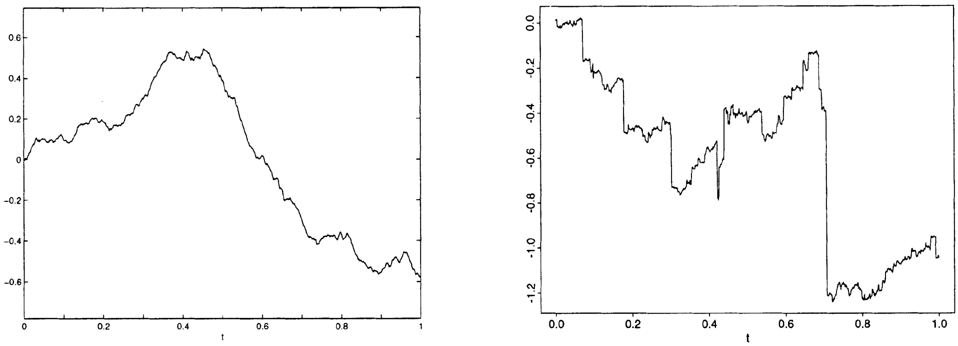

Fig. 5. Sample paths of fractional Brownian motion with $H = 0.9$ (left) and stable Lévy motion with $\alpha = 1.2$ (right).

3 The ON/OFF model

Here we present the ON/OFF model for computer network traffic, which was introduced by WILLINGER et al. (1995). We commence by considering a single ON/OFF source (representing one computer).

During an ON-period, the source generates traffic at a constant rate of 1 byte per time unit. During an OFF-period, the source remains silent. Let $X_1, X_2, \ldots$ be iid non-negative random variables representing the lengths of ON-periods and $Y_1, Y_2, \ldots$ be iid non-negative random variables representing the lengths of OFF-periods. The X- and Y-sequences are supposed to be independent. For any distribution function F we write $\overline{F} = 1 - F$ for the right tail. By F_{on} and F_{off} we denote the common distributions of ON- and OFF-periods, respectively.

In what follows, we assume that the lengths of ON- and OFF-periods are heavy tailed, i.e.

$$\overline{F}_{\text{on}}(x) \sim c_{\text{on}} x^{-\alpha_{\text{on}}} \quad \text{and} \quad \overline{F}_{\text{off}}(x) \sim c_{\text{off}} x^{-\alpha_{\text{off}}}, \quad \text{as } x \to \infty, \tag{7}$$

where $\alpha_{\text{on}}, \alpha_{\text{off}} \in (1, 2)$ and $c_{\text{on}}, c_{\text{off}}$ are positive constants. Hence, both distributions F_{on} and F_{off} have finite means μ_{on} and μ_{off} but their variances are infinite. Notice that the tail parameters α_{on} and α_{off} may be different, hence the extremes of the ON- and OFF-periods can differ significantly.

An ON/OFF source generates an *ON/OFF process* $\{W_t, t \geqslant 0\}$ of alternating ON- and OFF-periods. So $W_t = 1$ if t is in an ON-period and $W_t = 0$ if t is in an OFF-period. The ON/OFF process W is stationary if t runs from $-\infty$ to $+\infty$. But here t starts at zero, so some assumptions have to be imposed on the distribution of the 0th ON/OFF-period in order to make W stationary. HEATH et al. (1998) solved this problem by considering the renewal sequence generated by the alternating ON- and OFF-periods. Renewals happen at the beginnings of the ON-periods, the inter-arrival distribution is $F_{\text{on}} * F_{\text{off}}$ and the mean inter-arrival time $\mu_{\text{on}} + \mu_{\text{off}}$. In order to make the renewal sequence stationary (see RESNICK, 1992, p. 224, for a definition), a non-negative delay random variable D is introduced which is independent of the X_is and the Y_is and has distribution

$$P(D \leqslant x) = \frac{1}{\mu_{\text{on}} + \mu_{\text{off}}} \int_0^x (1 - F_{\text{on}} * F_{\text{off}}(u))\,\mathrm{d}u. \tag{8}$$

A stationary version of the renewal sequence is then given by

$$(S_n,\ n \geqslant 0) := \left(D,\ D + \sum_{i=1}^{n} (X_i + Y_i),\ n \geqslant 1 \right). \tag{9}$$

The first renewal S_0 is the starting time of the first ON-period. HEATH et al. (1998) constructed the delay random variable D as follows. They define four independent random variables B, $X_{\text{on}}^{(0)}$, $Y_{\text{off}}^{(0)}$ and Y_{off}, independent of the X_is and the Y_is. B is a Bernoulli random variable with

$$P(B = 1) = \frac{\mu_{\text{on}}}{\mu_{\text{on}} + \mu_{\text{off}}},$$

and

$$P(X_{\text{on}}^{(0)} \leqslant x) = \frac{1}{\mu_{\text{on}}} \int_0^x \overline{F}_{\text{on}}(u)\,\mathrm{d}u \quad \text{and} \quad P(Y_{\text{off}}^{(0)} \leqslant x) = \frac{1}{\mu_{\text{off}}} \int_0^x \overline{F}_{\text{off}}(u)\,\mathrm{d}u.$$

The random variable Y_{off} has distribution F_{off}. The delay variable D is now defined by

$$D = B(X_{\text{on}}^{(0)} + Y_{\text{off}}) + (1 - B)Y_{\text{off}}^{(0)}. \tag{10}$$

HEATH et al. (1998) showed that this D has indeed the distribution in (8). The interpretation of (10) is as follows. If $B = 1$ the renewal sequence starts during an ON-period (of which the remaining lifetime at $t = 0$ is $X_{\text{on}}^{(0)}$), which is followed by a full OFF-period Y_{off}. If $B = 0$ the renewal sequence starts during an OFF-period of which the remaining lifetime at $t = 0$ is $Y_{\text{off}}^{(0)}$.

The ON/OFF process W is now defined as

$$W_t = B 1_{[0, X_{\text{on}}^{(0)})}(t) + \sum_{n=0}^{\infty} 1_{[S_n, S_n + X_{n+1})}(t), \qquad t \geqslant 0.$$

The stationarity of the renewal sequence (9) implies strict stationarity of the process W with mean

$$EW_t = P(W_t = 1) = \frac{\mu_{\text{on}}}{\mu_{\text{on}} + \mu_{\text{off}}}.$$

The main result of HEATH et al. (1998) involves the autocovariance function $\gamma(k)$ of W. HEATH et al. (1998) give the precise rate of decay for $\gamma(k)$, under the assumptions (7) and $\alpha_{\text{on}} \neq \alpha_{\text{off}}$. Define

$$\alpha_{\min} = \min(\alpha_{\text{on}}, \alpha_{\text{off}}).$$

As $k \to \infty$

$$\gamma(k) \sim const\, k^{-(\alpha_{\min}-1)}, \tag{11}$$

so the ON/OFF process W exhibits LRD in the sense of (4).

So far we have considered the ON/OFF process W of a single source. WILLINGER et al. (1995) consider a network of M iid sources. Each source generates an ON/OFF process $W_t^{(m)}$. The total traffic in the network at time t is defined by

$$W_M(t) = \sum_{m=1}^{M} W_t^{(m)}, \qquad t \geqslant 0.$$

We call W_M the *workload process.* Since the sources are iid, (11) yields that the workload process exhibits LRD in the spirit of (4). The total traffic until time t is then given by

$$W_M^*(t) = \int_0^t \left(\sum_{m=1}^{M} W_u^{(m)} \right) \mathrm{d}u, \qquad t \geqslant 0. \tag{12}$$

We call W_M^* the *cumulative workload process.* In Figure 6 three ON/OFF processes, their workload and cumulative workload are shown.

4 Convergence results for the cumulative workload process

In this section we concentrate on limit theorems for the cumulative workload process, as both M and T go to infinity. In Section 4.1 we discuss the case in which first $M \to \infty$ and then $T \to \infty$. This yields convergence to fractional Brownian motion. In Section 4.2 we reverse the limits in M and T and obtain stable Lévy motion as the limit process.

4.1 *Convergence to fractional Brownian motion*

We start with the case when first $M \to \infty$ and then $T \to \infty$. The main result of WILLINGER et al. (1995) states that, in this case, the finite-dimensional distributions of $(W_M^*(Tt),\ t \geqslant 0)$ converge to those of fractional Brownian motion. Define the process

$$V_{(M,T)}(t) := \frac{W_M^*(Tt) - EW_M^*(Tt)}{T^H M^{1/2}},$$

where $H = (3 - \alpha_{\min})/2$.

THEOREM 1. *Let H be as above and B_H be standard fractional Brownian motion. Then for any $x_1, \ldots, x_n \in \mathbb{R}$, $t_1, \ldots, t_n \geqslant 0$ and $n \geqslant 1$*

$$\lim_{T\to\infty} \lim_{M\to\infty} P(V_{(M,T)}(t_1) \leqslant x_1, \ldots, V_{(M,T)}(t_n) \leqslant x_n)$$
$$= P(\sigma_0 B_H(t_1) \leqslant x_1, \ldots, \sigma_0 B_H(t_n) \leqslant x_n), \tag{13}$$

where σ_0 is a certain positive constant.

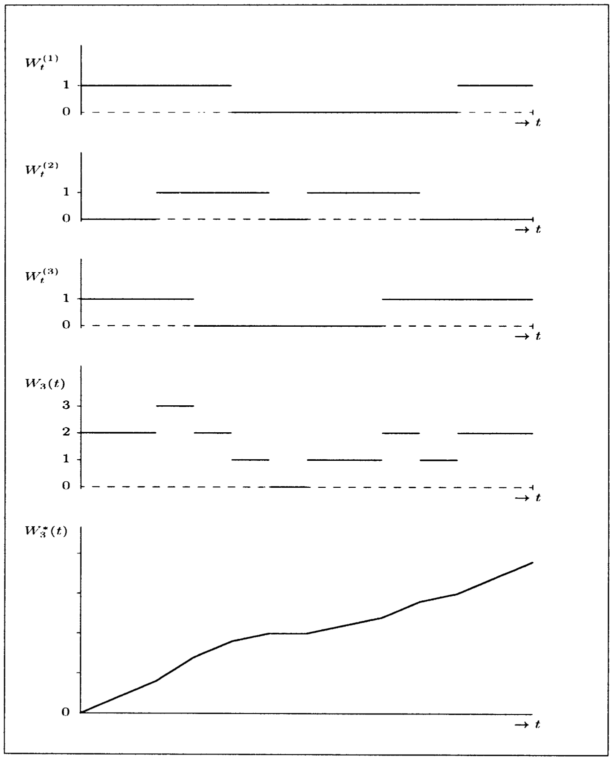

Fig. 6. ON/OFF processes, the workload process and the cumulative workload process in the case $M = 3$.

Usually, convergence in the sense of the finite-dimensional distributions involves only one limit. In this case, however, the process $V_{(M,T)}$ depends on two parameters, which both go to infinity.

The second limit $T \to \infty$ is used only to stabilize the variance of the process. WILLINGER et al. (1995) showed that

$$\mathrm{Var}(W_M^*(T)) \sim \sigma_0^2 M T^{3-\alpha_{\min}}, \quad \text{as } T \to \infty,$$

which explains the relation between H and $\alpha_{\min}$.

Fractional Brownian motion B_H is self-similar with parameter H. Since $1 < \alpha_{\min} < 2$, $H \in (1/2, 1)$ and so the corresponding fractional Gaussian noise sequence exhibits LRD. This reflects the LRD in the workload process. The theorem gives one possible 'physical explanation' of the observed self-similarity and the LRD in high-speed network traffic provided one accepts that the limits of M and T are taken in the proposed order. As we will see in the following sections, for (13) to hold, it is essential that the limits are performed in the order indicated.

4.2 *Reversed limits: convergence to stable Lévy motion*

In this section we discuss the convergence of the finite-dimensional distributions of the cumulative workload process, if first $T \to \infty$ and then $M \to \infty$. In this case, the limit process is α-stable Lévy motion. The following theorem states that, if first $T \to \infty$, the finite-dimensional distributions of the cumulative workload process (properly normalized) converge to those of infinite variance stable Lévy motion. This result can be found in TAQQU et al. (1997). Define the process

$$\tilde{V}_{(M,T)}(t) := \frac{W_m^*(Tt) - EW_M^*(Tt)}{(MT)^{1/\alpha_{\min}}}.$$

THEOREM 2. *For any* $x_1, \ldots, x_n \in \mathbb{R}$, $t_1, \ldots, t_n \geqslant 0$ *and* $n \geqslant 1$

$$\lim_{M\to\infty} \lim_{T\to\infty} P(\tilde{V}_{(M,T)}(t_1) \leqslant x_1, \ldots, \tilde{V}_{(M,T)}(t_n) \leqslant x_n)$$
$$= P(c\Lambda(t_1) \leqslant x_1, \ldots, c\Lambda(t_n) \leqslant x_n), \qquad (14)$$

where $\Lambda := \Lambda_{\alpha_{\min},\sigma,\beta}$ *is stable Lévy motion, c and* σ *are certain positive constants and* $\beta \in [-1, 1]$.

This result should be compared with Theorem 1: If the limits are reversed and a different normalization is used, another self-similar process, infinite variance stable Lévy motion, appears as limit process. In contrast to the fractional Brownian motion of Theorem 2, one loses the LRD completely: the limit process has independent increments.

The main reason for the different results of Theorems 1 and 2 is the infinite variance of the lengths of the ON- and OFF-periods. Indeed, suppose the variances of both X and Y are finite. Then WILLINGER et al. (1995) show that (with $\alpha_{\min} = 2$) (13) yields Brownian motion ($H = 1/2$) as limiting process. Equivalently, in the proof of (14) TAQQU et al. (1997) show that (with $\alpha_{\min} = 2$) the limiting process is 2-stable Lévy motion, which is Brownian motion. Notice that if $\alpha_{\min} = 2$, the pre-limit workload process has short-range dependence.

There is no particular (practical of theoretical) reason why one should prefer the limit regime in (13) to that in (14). In order to better understand the interplay of the rôles of the limits of M and T to infinity, we study the weak limit behavior of the cumulative workload process for large M and T. In the next section we focus on the simultaneous limit regimes in which M and T go to infinity at the same time.

5 Simultaneous limit regimes

In this section we present our main results under the assumption that both parmeters, M and T, go to infinity simultaneously. In particular, we assume that $M = M_T$ is some integer-valued function such that

$$M_T \text{ is non-decreasing in } T \text{ and } \lim_{T\to\infty} M_T = \infty.$$

For ease of presentation we usually suppress the dependence of M and T.

Recall the definition of the cumulative workload process W_M^* from (12). The following theorem gives conditions on M which ensure that the finite-dimensional distributions of W_M^* have an infinite variance stable Lévy motion as limiting process.

THEOREM 3. *Let c, σ and β be as in Theorem 2. If*

$$(M_T T)^{1/\alpha_{\min}} = o(T) \text{ as } T \to \infty, \tag{15}$$

then for any $x_1, \ldots, x_n \in \mathbb{R}$, $t_1, \ldots, t_n \geqslant 0$ *and* $n \geqslant 1$

$$\lim_{T\to\infty} P(\tilde{V}_{(M_T,T)}(t_1) \leqslant x_1, \ldots, \tilde{V}_{(M_T,T)}(t_n) \leqslant x_n)$$
$$= P(c\Lambda(t_1) \leqslant x_1, \ldots, c\Lambda(t_n) \leqslant x_n), \tag{16}$$

where $\Lambda := \Lambda_{\alpha_{\min},\sigma,\beta}$ *is stable Lévy motion.*

Notice that (16) is equivalent to convergence of the finite-dimensional distributions of the process $\tilde{V}_{(M_T,T)}$ to those of stable Lévy motion.

THEOREM 4. *Let* σ_0 *be as in Theorem 1,* $H = (3 - \alpha_{\min})/2$, *and* B_H *be standard fractional Brownian motion. If*

$$T = o(M_T^{1/2} T^H) \quad \text{as } T \to \infty, \tag{17}$$

then

$$[V_{(M_T,T)}(t)]_{t\geqslant 0} \xrightarrow{d} [\sigma_0 B_H(t)]_{t\geqslant 0}, \tag{18}$$

where $\xrightarrow{d}$ *denotes weak convergence in the space* $\mathbb{C}[0, \infty)$ *of continuous functions on the interval* $[0, \infty)$ *endowed with the sup-norm on compact intervals.*

The weak convergence in (18) implies convergence of the finite-dimensional distributions of the process $V_{(M_T,T)}$ to those of fractional Brownian motion. Notice that (15) is equivalent to $M_T = o(T^{\alpha_{\min}-1})$, while (17) requires $T^{\alpha_{\min}-1} = o(M_T)$. The results of Theorems 3 and 4 tell us that it depends on the relative speed of M with respect to T, which process is obtained as a limit for the cumulative workload. If M goes to infinity at a slower rate than $T^{\alpha_{\min}-1}$, then stable Lévy motion appears, which has independent increments. If M grows faster than $T^{\alpha_{\min}-1}$, then the limiting process is fractional Brownian motion, of which the increments exhibit LRD.

An intuitive explanation for the totally different dependence structures in the limiting processes is as follows. The ON/OFF process of a single source exhibits LRD. If we add up M iid sources, the workload process still has LRD. The effect of

$T \to \infty$ is that the time scale is blown up. Values of t that were close to each other are far apart as T becomes large.

This destroys the dependence structure within an ON/OFF process. This effect is slowed down if the adding up ON/OFF processes happens very rapidly, i.e. if M grows fast. The dependence structure of the workload process remains intact in Theorem 4 since M is allowed to grow fast. In Theorem 3, however, M is not allowed to grow fast and the dependence in the workload process vanishes in the limit. See also Remarks 3 and 4.

Notice that if

$$M_T \sim const\ T^{\alpha_{\min}-1} \quad \text{as } T \to \infty,$$

then both conditions (15) and (17) are not satisfied. In this case it is not clear whether the cumulative workload process converges in some sense at all.

6 Sketch of the proofs of Theorems 3 and 4

Here we give the ideas underlying the proofs of Theorems 3 and 4. We also try to explain why different limit processes appear under different conditions on M. A detailed proof can be found in MIKOSCH and STEGEMAN (1999), in which general regularly varying tails are assumed for F_{on} and F_{off} (including slowly varying functions). See also MIKOSCH et al. (1999).

6.1 *The stable Lévy motion case*

We follow the approach of TAQQU et al. (1997). First, we consider a single ON/OFF source and focus on the mean corrected cumulative workload process:

$$G_t = \int_0^t (W_u - EW_u)\,\mathrm{d}u, \qquad t \geqslant 0.$$

We will represent this process as a random sum with remainder term. This representation will show the basic structure of G.

Notice that

$$W_t - EW_t = \begin{cases} \dfrac{\mu_{\text{off}}}{\mu_{\text{on}} + \mu_{\text{off}}} =: r_{\text{on}} & \text{if } W_t = 1, \\ -\dfrac{\mu_{\text{on}}}{\mu_{\text{on}} + \mu_{\text{off}}} =: -r_{\text{off}} & \text{if } W_t = 0. \end{cases}$$

Recall the definition of the underlying renewal sequence (S_n) from (9) and define the corresponding renewal counting process

$$N_t := \sum_{n=0}^{\infty} 1_{[0,t]}(S_n).$$

Notice that $S_{N_t-1} \leqslant t < S_{N_t}$. For $k \geqslant 1$, let J_k be the mean-corrected cumulative workload in the kth renewal interval $[S_{k-1}, S_k)$:

$$J_k = \int_{S_{k-1}}^{S_k} (W_u - EW_u)\,\mathrm{d}u = r_{\mathrm{on}} X_k - r_{\mathrm{off}} Y_k$$

$$= r_{\mathrm{on}}(X_k - EX_k) - r_{\mathrm{off}}(Y_k - EY_k).$$

Notice that the J_ks are iid mean-zero random variables.

The random sum representation of G_t depends on whether $t \geqslant S_0$ or $t < S_0$.

The case $t \geqslant S_0$. Then G_t consists of three parts. The first part is the contribution of the 0th renewal interval $[0, S_0)$, which we denote by R_1. The second part is the contribution of the renewal intervals $[S_{k-1}, S_k)$, for $k = 1, \ldots, N_t$, given by

$$\sum_{k=1}^{N_t} J_k.$$

It remains to subtract the contribution of the interval $(t, S_{N_t}]$. We denote the contribution of this interval by $R_2(t)$.

The case $t < S_0$. In this case we denote the contribution of the interval $[0, t]$ by $R_3(t)$. The complete expression for G_t is as follows.

LEMMA 1. *The mean-corrected cumulative workload process has the following representation as a random sum with remainder terms:*

$$G_t = \left[R_1 + \sum_{k=1}^{N_t} J_k - R_2(t) \right] 1_{[S_0,\infty)}(t) + R_3(t) 1_{[0,S_0)}(t). \tag{19}$$

Next, consider a large number M of iid ON/OFF sources and define for source m

$$G_t^{(m)} = \int_0^t (W_u^{(m)} - EW_u^{(m)})\,\mathrm{d}u, \qquad t \geqslant 0.$$

We also adapt the notation of (19) to the mth source. Let $\alpha := \alpha_{\min}$ and $a_T = (MT)^{1/\alpha}$. We consider the process

$$\left(a_T^{-1} \sum_{m=1}^{M} G_{Tt}^{(m)}, \qquad t \geqslant 0 \right),$$

for large T. As in (19), we can write

$$a_T^{-1} \sum_{m=1}^{M} G_{Tt}^{(m)} = I_T(t) + II_T(t) - III_T(t) + IV_T(t), \tag{20}$$

where

$$I_T(t) = a_T^{-1} \sum_{m=1}^{M} [R_1^{(m)} 1_{[S_0^{(m)},\infty)}(Tt)],$$

$$II_T(t) = a_T^{-1} \sum_{m=1}^{M} \sum_{k=1}^{N_{Tt}^{(m)}} [J_k^{(m)} 1_{[S_0^{(m)},\infty)}(Tt)],$$

$$III_T(t) = a_T^{-1} \sum_{m=1}^{M} [R_2^{(m)}(Tt) 1_{[S_0^{(m)},\infty)}(Tt)],$$

$$IV_T(t) = a_T^{-1} \sum_{m=1}^{M} [R_3^{(m)}(Tt) 1_{[0,S_0^{(m)})}(Tt)].$$

The next result states that the terms I_T, III_T and IV_T vanish in the limit.

LEMMA 2. *Suppose*

$$a_T = o(T), \quad \text{as } T \to \infty.$$

Then for every $t \geqslant 0$,

$$I_T(t) \xrightarrow{P} 0 \quad \text{and} \quad III_T(t) \xrightarrow{P} 0 \quad \text{and} \quad IV_T(t) \xrightarrow{P} 0,$$

as $T \to \infty$.

Since (see BILLINGSLEY, 1968, Theorem 4.1) $\tilde{X}_n \xrightarrow{d} \tilde{X}$ and $\tilde{Y}_n \xrightarrow{P} 0$ together imply that $\tilde{X}_n + \tilde{Y}_n \xrightarrow{d} \tilde{X}$, Lemma 2 yields that we only have to consider the finite-dimensional distributions of II_T. According to the next result they converge to stable Lévy motion.

LEMMA 3. *Let* c, σ *and* β *be as in Theorem 3. Suppose*

$$a_T = o(T), \quad \text{as } T \to \infty.$$

Then for any $x_1, \ldots, x_n \in \mathbb{R}$, $t_1, \ldots, t_n \geqslant 0$ *and* $n \geqslant 1$

$$\lim_{T\to\infty} P(II_T(t_1) \leqslant x_1, \ldots, II_T(t_n) \leqslant x_n) = P(c\Lambda(t_1) \leqslant x_1, \ldots, c\Lambda(t_n) \leqslant x_n),$$

where $\Lambda := \Lambda_{\alpha,\sigma,\beta}$ *is the* α*-stable Lévy motion.*

Sketch of proof First we establish the convergence of the marginal distributions, i.e. the random variables $II_T(t)$ for fixed $t \geqslant 0$, to $S_\alpha(c\,\sigma\, t^{1/\alpha}, \beta, 0)$. For ease of presentation we assume that $t = 1$. We write

$$II_T(1) = \sum_{m=1}^{M} Z_{T,m}, \tag{21}$$

where

$$Z_{T,m} := 1_{[S_0^{(m)},\infty)}(T) a_T^{-1} \sum_{k=1}^{N_T^{(m)}} J_k^{(m)}.$$

PETROV (1975), Theorem 8 in Chapter IV, gives the following necessary and sufficient conditions for the sums (21) of rowwise iid random variables to converge weakly to an α-stable distribution $S_\alpha(c\,\sigma, \beta, 0)$: as $T \to \infty$

$$(A)\ \ M\ P(Z_{T,1} > x) \to C_\alpha \frac{1+\beta}{2} (c\,\sigma)^\alpha x^{-\alpha}, \quad \forall x > 0,$$

$$(B)\ \ M\ P(Z_{T,1} \leqslant -x) \to C_\alpha \frac{1-\beta}{2} (c\,\sigma)^\alpha x^{-\alpha}, \quad \forall x > 0,$$

$$(C)\ \ \lim_{\epsilon\downarrow 0} \limsup_{T\to\infty} M\,\mathrm{Var}(Z_{T,1} 1_{\{|Z_{T,1}|<\epsilon\}}) = 0,$$

where C_α is given by

$$C_\alpha = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}.$$

(A) and (B) require the application of a large deviations result for heavy tailed random sums. This boils down to showing that the renewal counting process N_T can be replaced by its mean μ_T, i.e. values that are far from the mean are asymptotically negligible. (C) then follows from (A) and (B), since

$$\mathrm{Var}(Z_{T,1} 1_{\{|Z_{T,1}|<\epsilon\}}) \leqslant E(Z_{T,1}^2 1_{\{|Z_{T,1}|<\epsilon\}})$$
$$= \int_0^{\epsilon^2} P(Z_{T,1} > \sqrt{x})\,\mathrm{d}x + \int_0^{\epsilon^2} P(Z_{T,1} \leqslant -\sqrt{x})\,\mathrm{d}x.$$

The convergence of the finite-dimensional distributions is established by applying the Cramér-Wold device i.e. by showing that any linear combination of the random variables $II_T(t_i)$, for $t_1 < \cdots < t_k$, converges to the corresponding linear combination of the stable random variables $\Lambda_{\alpha,\sigma,\beta}(t_i)$. □

So far we have shown the convergence of the finite-dimensional distributions to stable Lévy motion. A natural thought is to consider the possibility of functional weak convergence in the path space of cadlag functions $\mathbb{D}[0, \infty)$. Unfortunately, we do not have weak convergence under the J_1-topology on $\mathbb{D}$ (see SKOROKHOD, 1956). KONSTANTOPOULOS and LIN (1998) showed that a sequence of stochastic processes with continuous sample paths, converging in finite-dimensional distributions to a process with discontinuous sample paths, is not tight under the J_1-topology. However, it might be possible to have weak convergence under a weaker topology on $\mathbb{D}$, for instance Skorokhod's M_1-topology.

REMARK 3. Notice that the independent increments of the limit process arise in a very natural way. We used the decomposition (20) and asymptotically, we are left with II_T, which is a random sum of iid heavy tailed random variables. Basically, the counting process N_T can be replaced by its mean μ_T, since values far from the mean are asymptotically negligible. This means that for large T, II_T behaves like a sum of iid random variables and hence has independent increments.

6.2 *The fractional Brownian motion case*

In this section we consider the case when M increases faster than $T^{\alpha-1}$. Let $a_T = M^{1/2}T^H$, where H is as in Theorem 4. As before, $\alpha := \alpha_{\min}$. This time $T = o(a_T)$ as $T \to \infty$ and the result of Lemma 2 does not hold. Therefore, we have to consider the process (20) as a whole. The proof of the convergence of the finite-dimensional distributions to fractional Brownian motion is similar to the proof of Lemma 3. For fixed $t \geqslant 0$, the cumulative workload is represented as a sum of M iid random variables. For the convergence of the marginal distributions, we have to show that as $T \to \infty$

$$a_T^{-1} \sum_{m=1}^{M} G_{Tt}^{(m)} \xrightarrow{d} N(0, \sigma_0^2 t^{3-\alpha}) \stackrel{d}{=} \sigma_0 B_H(t). \tag{22}$$

Necessary and sufficient conditions for (22) to hold can be found in PETROV (1995), Theorem 4.2. The convergence of the finite-dimensional distributions is again established by considering linear combinations. The tightness is proved by applying BILLINGSLEY (1968), Theorem 12.3.

REMARK 4. Contrary to the convergence to stable Lévy motion, the LRD in the workload process is preserved in the limit. This can be explained using the decomposition (20). This time we have to consider the whole proces (20), in which the LRD is still present. In Theorem 4, M is allowed to grow fast compared to T and the dependence structure of the workload process remains intact (see also Remark 3).

Acknowledgements

I would like thank Thomas Mikosch for his advice, comments and mathematical support on this work, which is an extended version of my Masters Thesis. I also thank Statistica Neerlandica for giving me the opportunity to write this paper as a result of winning the 1999 VVS-prize for my Masters Thesis.

References

BERAN, J. (1994), *Statistics for long-memory processes*, Chapman & Hall, New York.

BILLINGSLEY, P. (1968), *Convergence of probability measures*, Wiley, New York.

BROCKWELL, P. J. and R. A. DAVIS (1991), *Time series: theory and methods*, (2nd ed.), Springer-Verlag, New York.

CROVELLA, M. E., M. S. TAQQU and A. BESTAVROS (1996), Heavy-tailed probability distributions in the world wide web, *preprint*.

CROVELLA, M. A. and A. BESTAVROS (1996), Self-similarity in world wide web traffic: evidence and possible causes, *Performance Evaluation Review* **24**, 160–169.

EMBRECHTS, P., C. KLüPPELBERG and T. MIKOSCH (1997), *Modelling entremal events*, Springer-Verlag, Berlin.

FOWLER, H. J. and W. E. LELAND (1991), Local area network traffic characteristics with implications for broadband network congestion management, *IEEE Journal on Selected Areas in Communications* **9**, 1139–1149.

HEATH, D., S. I. RESNICK and G. SAMORODNITSKY (1998), Heavy tails and long range dependence in ON/OFF processes and associated fluid models, *Mathematics of Operations Research* **23**, 145–165.

KONSTANTOPOULOS, T. and S. LIN (1998), Macroscopic models for long-range dependent network traffic, *Queueing Systems Theory and Applications* **28**, 215–243.

LELAND, W. E. and D. V. WILSON (1991), High time-resolution measurement and analysis of LAN traffic: implications for LAN interconnection, *Proceedings of the IEEE INFOCOM'91*, 1360–1366.

LELAND, W. E., W. WILLINGER, M. S. TAQQU and D. V. WILSON (1993), On the self-similar nature of Ethernet traffic, *Computer Communications Review* **23**, 183–193.

MIKOSCH, T. and A. W. STEGEMAN (1999), The interplay between heavy tails and rates in self-similar network traffic, *preprint available at* http://www.math.rug.nl/∼stegeman

MIKOSCH, T., S. I. RESNICK, H. ROOTZÉN and A. W. STEGEMAN (1999), Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *preprint available at* http:/ / www.math.rug.nl/∼stegeman

PAXTON, V. and S. FLOYD (1995), Wide-area traffic: the failure of Poisson modeling, *IEEE/ACM Transactions on Networking* **3**, 226–244.

PETROV, V. V. (1975), *Sums of independent random variables*, Springer-Verlag, Berlin.

PETROV, V. V. (1995), *Limit theorems of probability theory*, Oxford University Press, Oxford.

RESNICK, S. I. (1992), *Adventures in stochastic processes*, Birhäuser, Boston.

RESNICK, S. I. (1997), Heavy tail modeling and teletraffic data, *Annals of Statistics* **25**, 1805–1869.

SAMORODNITSKY, G. and M. S. TAQQU (1994), *Stable non-Gaussian random processes*, Chapman & Hall, New York.

SKOROKHOD, A. V. (1956), Limit theorems for stochastic processes, *Theory of Probability and its Applications* **1**, 261–290.

STEGEMAN, A. W. (1998), *Modeling traffic in high-speed networks by ON/OFF models*, Masters Thesis at the University of Groningen.

TAQQU, M. S. and V. TEVEROVSKY (1995), Estimators for long-range dependence: an empirical study, *Fractals* **3**, 785–798.

TAQQU, M. S. and V. TEVEROVSKY (1996), On estimating the intensity of long-range dependence in finite and infinite variance time series, *preprint*.

TAQQU, M. S., W. WILLINGER and R. SHERMAN (1997), Proof of a fundamental result in self-similar traffic modeling, *Computer Communications Review* **27**, 5–23.

TEVEROVSKY, V. and M. S. TAQQU (1995), Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator, *preprint*.

WILLINGER, W., M. S. TAQQU, R. SHERMAN and D. V. WILSON (1995), Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level (extended version), *Computer Communications Review* **25**, 100–113.

Received: April 1999. Revised: January 2000.