Appears in the Scandinavian Journal of Statistics, 29 (2002) 273-295

The modeling of Ethernet data and of signals that are heavy-tailed with infinite variance

Murad S.Taqqu
Boston University

April 13, 2001

Running title: The modeling of heavy-tailed signals

Keywords and phrases: computer networks, renewal reward processes, heavy tails, fractional Brownian motion, self-similar processes, stable processes.

Abstract

We provide an overview of some of the research of the last ten years involving computer network data traffic. We describe the original Ethernet data study which suggested that computer traffic is inherently different from telephone traffic and that in the context of computer networks, self-similar models such as fractional Brownian motion, should be used. We show that the on-off model can physically explain the presence of self-similarity. While the on-off model involves bounded signals, it is also possible to consider arbitrary unbounded finite-variance signals or even infinite-variance signals whose distributions have heavy tails. We show that, in the latter case, one can still obtain self-similar processes with dependent increments, but these are not the infinite-variance fractional stable Lévy motions which have been commonly considered in the literature. The adequate model, in fact, can either have dependent or independent increments, and this depends on the respective size of two parameters, namely, the number of workstations in the network and the time scale under consideration. We indicate what happens when these two parameters become jointly asymptotically large. We conclude with some comments about high frequency behavior and multifractals.

1 Introduction

The Ethernet is a local area network (LAN) technology which is still widely used today. Among the features that make Ethernets so successful are ease of maintenance and administration, ease of network reconfiguration access by a single, passive medium that is shared by all host stations and the absence of a central controller allocating access to the channel. (For a detailed description of the Ethernet technology, see Schoch and Hupp (1980)).

Starting with the extensive analyses of traffic measurements from Ethernet LANs over a 4-year period described in Leland, Taqqu, Willinger and Wilson (1994), there have been a number of recent empirical studies that provide evidence of the prevalence of *self-similar* traffic patterns in measured traffic. Self-similar processes were introduced by Kolmogorov (1941). They were brought to the attention of statisticians and probabilists by Mandelbrot and his co-workers in the late 1960s and early 1970s. Intuitively, the attractive feature of such processes is that they look the same at different scales. From another perspective, actual traffic exhibits correlations over a wide range of time scales (i.e., has *long-range dependence*), while traditional traffic models typically focus on a very limited range of time scales and are thus *short-range dependent* in nature.

While technically short-range dependent models such as Markov models can be used to approximate self-similar ones, this approximation tends to be ad hoc and can involve many parameters. One advantage of self-similar models is that they are parsimonious. A model, however, should not only fit the data but also have a physical basis. It is important, therefore, to provide an answer to the following question: *What is the physical "explanation" for the observed self-similar nature of measured traffic from today's packet networks?* We do that below.

In this paper, we summarize a number of results that were obtained in the last ten years. For more details, the reader should refer to the original papers. A detailed bibliography can be found in Willinger, Taqqu and Erramilli (1996). For a comprehensive overview of research in the networking area, the reader should also consult the recent monograph by Park and Willinger (2000). It contains contributed chapters spanning a cross-section of topics that cover traffic measurement, modeling, queues, performance analysis and traffic control for self-similar network traffic.

We then present an answer to question stated above by providing the appropriate mathematical results. In particular, we present a plausible and simple explanation for the observed self-similarity of measured aggregate packet traffic in terms of the nature of the traffic generated by the individual sources or source-destination pairs that contribute to the aggregate packet stream. Developing an approach originally suggested by Mandelbrot (1969), we show that the superposition of many (idealized) on-off sources having *heavy-tailed* distributions with infinite variance (e.g., certain Pareto and stable distributions), results in self-similar aggregate traffic. Being able to phrase the results in the well-known framework of on-off source models (also known as "packet train models"), we identify the presence of these heavy tails as the essential point of departure from traditional to self-similar traffic modeling. Intuitively, if the length of an *on* or *off* period is highly variable but can be very large with non-negligible probability, then their superposition exhibits characteristics that cover a wide range of time scales. The parameter α describing the "heaviness" of the tail of these distributions measures the intensity of the effect. We also provide a simple relation between α and the self-similarity parameter H of the aggregate traffic stream.

In sharp contrast to our findings, traditional traffic modeling, when cast in the framework of on-off source models, assumes without exception finite variance distributions for the on- and off-periods (e.g., exponential distribution, geometric distribution). These assumptions drastically limit the on-off activities of an individual source, and as a result, the superposition of many such sources behaves like white noise in the sense that the aggregate traffic stream is void of any significant correlations, except possibly some in the short range. This behavior is in clear contrast with measured network traffic (for details, see for example Leland, Taqqu, Willinger and Wilson (1994)).

The Ethernet data is presented in Section 2. We suggest in Section 4 that it can be approximated by fractional Gaussian noise (FGN) which are the increments of fractional Brownian motion (FBM). The processes FGN and FBM are defined in Section 3. In Section 5, we present two models that explain the presence of long-range dependence, the on-off model and the infinite sources Poisson model. Both of these models involve two parameters. The on-off model, for example, involves M, the number of workstations in the network and T, the time scale. Fractional Brownian motion appears in the limit as $M \to \infty$ first followed by $T \to \infty$. If $T \to \infty$ first and then $M \to \infty$, we obtain instead the Lévy-stable motion, an infinite variance process with independent increments. Summarizing results obtained by Mikosch, Resnick, Rootzén and Stegeman (1999), we also describe what happens if M and T go jointly to infinity.

The on-off model involves bounded signals (1 or 0). In Section 6, we describe what happens when the signals are unbounded with either finite or infinite variance. The results for finite variance are similar to the on-off case. When the signals have infinite variance, the limiting process will have infinite variance as well, but it is not the fractional Lévy stable motion which is the "common" extension of fractional Brownian motion to the infinite variance case. It is a new self-similar process, denoted Z_β which we shall describe. We show that one obtains in the limit either Z_β or the Lévy stable motion, depending, in particular, on which one of either M or T goes to infinity first. In Section 7, we describe what happens when the signals have infinite variance and M and T go to infinity jointly, summarizing results obtained by Pipiras, Taqqu and Levy (2000). We conclude in Section 8 by mentioning some recent developments in research.

2 The Ethernet data

We shall focus on very high quality, high time-resolution Ethernet LAN traffic data collected by Leland and Wilson (1991). It was collected at the Bellcore Morris Research and Engineering Center (MRE). The network environment in this center is probably typical of a research or software development environment where workstations are the primary machines on people's desks. This data has been widely analyzed and is now regarded as "classical". The monitoring system used to collect the data for the present study was custom-built in 1987-1988, records all packets seen on the Ethernet under investigation with accurate time stamps (to about 100 μs) and can do so for week-long runs without interruption. (A packet consists of a header followed by a variable number of bytes.) For a detailed description of the monitor, including extensive testing of its capacity and accuracy, see Leland and Wilson (1991). There is no intrinsic limitation on the amount of traffic that can be collected. For example, in the case at hand, 1 to 2 days of uninterrupted monitoring of the Ethernet cable typically resulted in about 27 million packets filling a single 2.4-Gbyte 8-mm tape.

Table 1 of Willinger, Taqqu, Leland and Wilson (1995) gives a summary description of the Ethernet traffic data. There are four sets of traffic measurements, each one representing between 20 and 40 consecutive hours of Ethernet traffic and each one consisting of tens of millions of Ethernet packets. The data were collected on different intracompany LAN networks at different periods in time over the course of approximately four years, exhibiting a number of different network utilizations and host populations. For illustration here, we shall consider the data set labeled AUG89.P from the August 1989 measurements. It is a time series which measures the number of packets in the network every 10 milliseconds. There is 27 hours worth of data. Since we would like to model the data as a stationary time series, we focused on subsets of 1 hour long. We identified what are considered "typical" low (L), medium (M) and high (H) activity hours. For example, AUG89.MP, referred to in Table 1 of Willinger, Taqqu, Leland and Wilson (1995) is a time series of length 360,000, each observation representing the number of packets per 10 milliseconds; a total of 968,631 packets were observed during this medium-traffic hour (2.25pm to 3.25pm).

The 27 consecutive hours worth of monitored Ethernet traffic from the August 1989 measure-

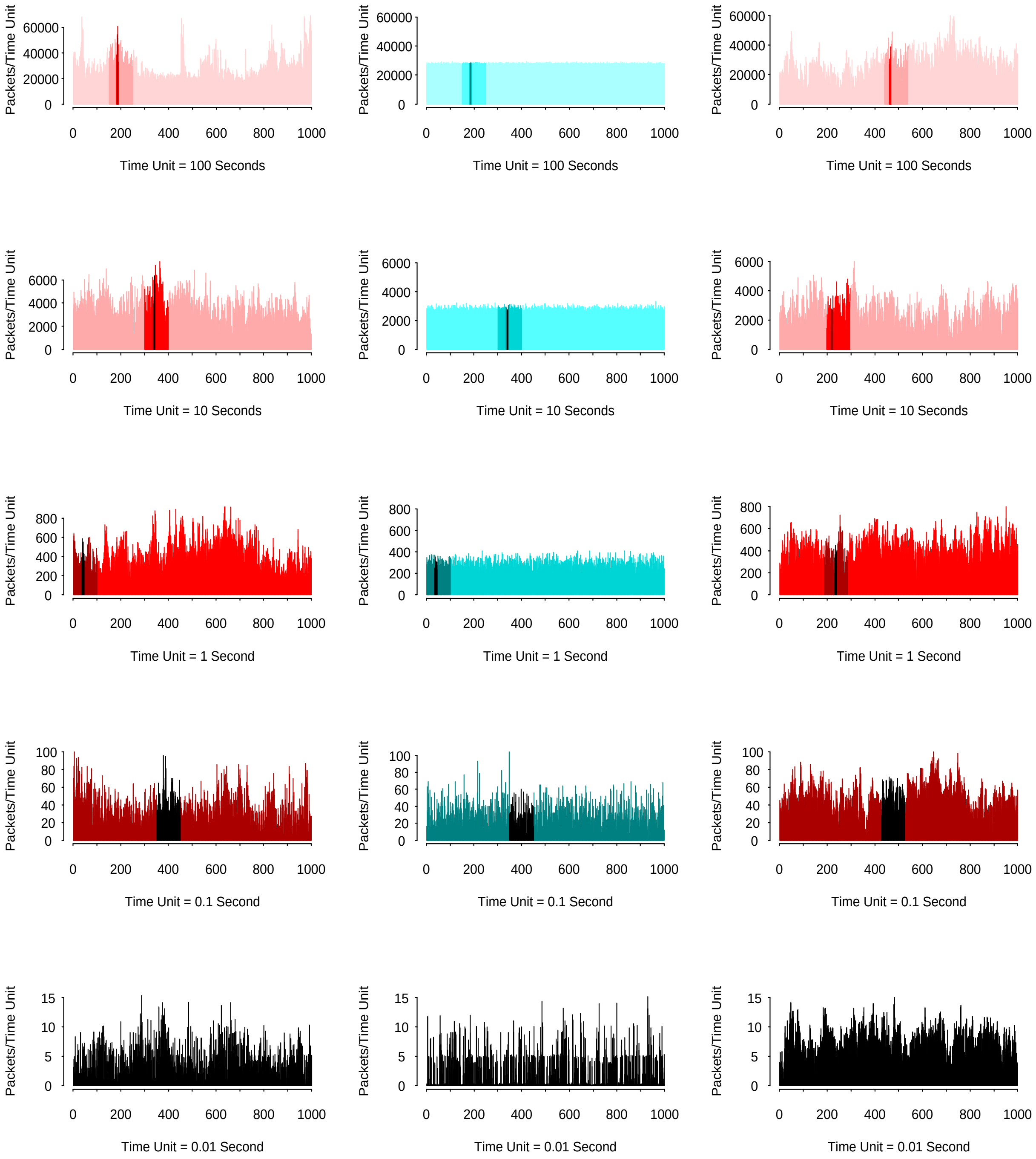


Figure 1: Actual Ethernet traffic (left column), synthetic time series generated from a traditional Poisson traffic model (middle column), and synthetic time series generated from an on-off traffic model (right column) – on five different time scales. Different gray levels indicate the same segments of traffic on the different time scales.

ments are displayed in the first column of Figure 1. (Figure 1 displays three columns of five graphs. We shall label the graphs in the first column by a-e respectively, those in the second column by a'-e' and those in the third column by a"-e", the plots in the first row being denoted a, a' and a" respectively). Figure 1a-e depicts a sequence of plots of the Ethernet packet counts (i.e., number of packets per time unit) for five different choices of time units (levels of aggregation). Starting with a time unit of 100 seconds (Figure 1a), each subsequent plot is obtained from the previous one by increasing the time resolution by a factor of 10 and by concentrating on a randomly chosen subinterval (indicated by a darker shade). Recall that the time unit corresponding to the finest time scale is 10 milliseconds (Figure 1e). In order to avoid the visually irritating quantization effect associated with the finest resolution level, plot (e) depicts a "jittered" version of the number of packets per 10 milliseconds, that is, a small amount of noise has been added to the actual arrival rate. Observe that with the possible exception of plot (a), which suggests the presence of a daily cycle, all plots are intuitively very "similar" to one another (in a distributional sense), that is, Ethernet traffic seems to look the same in the large time scales, i.e., minutes or hours, as in the small, i.e., seconds or milliseconds. In particular, notice the absence of a natural length of a "burst": at every time scale ranging from milliseconds to minutes and hours, bursts consist of bursty subperiods separated by less bursty subperiods. This feature of Ethernet traffic is drastically different from both conventional telephone traffic and from stochastic models for packet traffic currently considered in the literature. The latter typically produce plots of packet counts which are indistinguishable from white noise after aggregating over a few hundred milliseconds, as illustrated in Figure 1 with the sequence of plots (a')-(e'); this sequence was obtained in the same way as the sequence (a)-(e), except that it depicts synthetic traffic generated from a comparable compound Poisson process in terms of average packet size and arrival rate. (Note that while the choice of a compound Poisson process is admittedly not very sophisticated, even more complicated Markovian arrival processes would produce plots indistinguishable from Figure 1a'-e'.) Figure 1a-e suggests the use of self-similar stochastic processes for traffic modeling. The plots (a")-(e") are synthetic plots using the on-off models discussed below, which can explain the presence of self-similarity. Observe that the synthetic plots (a")-(e") look similar to the plots of real date (a)-(e) of Ethernet data.

3 Fractional Brownian motion and fractional Gaussian noise

The original data is nonnegative. We shall aggregate it, subtract its mean and use fractional Gaussian noise to model it at high levels of aggregation. In this section, we define fractional Gaussian noise, starting with its "parent" process fractional Brownian motion. Fractional Brownian motion $B_H(t)$, $t \in \mathbb{R}$, is a stochastic process whose finite-dimensional distributions are characterized by the following properties:

1. The process is Gaussian. It starts at zero ($B_H(0) = 0$) and has zero mean ($EB_H(t) = 0$).
2. It has stationary increments.
3. $EB_H^2(t) = \sigma^2|t|^{2H}, \quad 0 < H < 1.$

By using these properties, one can easily derive the covariance

$$EB_H(t)B_H(s) = \frac{1}{2}\{EB_H^2(t) + EB_H^2(s) - E(B_H(t) - B_H(s))^2\}$$

$$= \frac{\sigma^2}{2}\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\}. \tag{3.1}$$

Fractional Brownian motion, moreover, is *self-similar* with parameter H, namely

$$B_H(at) \stackrel{d}{=} a^H B_H(t), \quad \forall a > 0,$$

that is, the finite-dimensional distribution of $B_H(at)$ are the same as those of $a^H B_H(t)$.

When $H = 1/2$, $B_H(t)$ becomes the usual Brownian motion with independent increments since, in this case, (3.1) reduces to $\min(s,t)$ which is the covariance of Brownian motion. The value $H = 1/2$ divides the range $(0,1)$ of values of H into two distinct parts $0 < H < 1/2$ and $1/2 < H < 1$, which have quite different properties. These are best noted when looking at the (discrete-time) increment process

$$X_i = B_H(i+1) - B_H(i), \quad i \in \mathbb{Z},$$

which is called *fractional Gaussian noise* (FGN). FGN is a stationary Gaussian process with mean zero and covariance

$$r(k) = EX_iX_{i+k} = \frac{\sigma^2}{2}\{|k+1|^{2H} - 2|k|^{2H} - |k-1|^{2H}\} \tag{3.2}$$

Thus $r(0) = \mathrm{Var}X_i = \sigma^2$ and, if $H = 1/2$, $r(k) = 0$ for $|k| \geq 1$. But if $H \neq 1/2$,

$$r(k) \sim \sigma^2 H(2H-1)|k|^{2H-2} \qquad \text{as} \qquad k \to \infty. \tag{3.3}$$

(We say that the functions $a(k)$ and $b(k)$ are asymptotic to each other if $\lim_{k\to\infty} a(k)/b(k) = 1$ and we denote this by $a(k) \sim b(k)$). Thus $r(k)$ tends to zero like a power function. When $1/2 < H < 1$, $r(k)$ is non-negative and tends to zero, but

$$\sum_{k=-\infty}^{\infty} r(k) = \infty. \tag{3.4}$$

The covariances tend to zero but so slowly that the sum of the covariances diverges. Because of this, one says that the stationary process $\{X_i,\ i \in \mathbb{Z}\}$ has *long-range dependence*, or equivalently, *long memory* or *strong dependence.*

When $0 < H < 1/2$, the right-hand side of (3.3) is negative and in fact, $r(k) < 0$ for all $k \neq 0$. One says that $\{X_i,\ i \in \mathbb{Z}\}$ has *negative dependence.* In the context of the modelling of computer network traffic, we shall only be interested in the cases $H = 1/2$ and $H > 1/2$.

One can also adopt the spectral domain perspective and consider the spectral density

$$s(x) = \sum_{k=-\infty}^{\infty} r(k)e^{ikx}, \quad -\pi/2 < x < \pi/2$$

which satisfies

$$s(x) \sim c_H|x|^{1-2H} \qquad \text{as} \qquad x \to 0 \tag{3.5}$$

for some constant $c_H > 0$. Thus, $s(x)$ blows up at the origin when $H > 1/2$ and tends to zero at the origin when $H < 1/2$. The index "x" corresponds to frequencies, and thus $x \approx 0$ means very low frequencies.

The self-similarity of $B_H(t)$ induces another type of self-similarity on the *stationary* sequence $\{X_i\}$, namely

$$\frac{1}{N^H} \sum_{j=iN}^{(i+1)N-1} X_j \stackrel{d}{=} X_i \tag{3.6}$$

that is, summing the original sequence on blocks of size N and normalizing by $N^H = (\text{block size})^H$ preserves the finite-dimensional distributions.

In practice, sequences $\{X_i,\ i \in \mathbb{Z}\}$ that one encounters may not have the specific covariance structure (3.2) at all lags k. One is often interested only at what happens at large lags or low frequencies, wanting to know if long-range dependence is present or not, that is, whether the asymptotic relations (3.3) or (3.5) are satisfied. In this case, the relation (3.6), instead of holding for all block sizes N, may hold only asymptotically, as the block size $N \to \infty$, with the right-hand side of (3.6) replaced by fractional Gaussian noise. One says then, that $\{X_i,\ i \in \mathbb{Z}\}$ is *asymptotically self-similar*. Asymptotic self-similarity can also be verified by focusing on the cumulative sums and requiring that the finite-dimensional distributions of the process $N^{-H} \sum_{j=1}^{[Nt]} X_j$, $0 \le t \le 1$, converge to those of $B_H(t)$, $0 \le t \le 1$.

4 Data analysis

The presence of long-range dependence in traffic measurements taken from communication systems *other than* Ethernet LANs (e.g., certain types of video traffic) has been demonstrated in Beran, Sherman, Taqqu and Willinger (1995) and Garrett and Willinger (1994). In the case of the Ethernet data, self-similar models fit the data better than conventional traffic models, all of which ignore the presence of long-term correlations in the data. It is already possible to arrive at this conclusion through simple plots of the traffic over a range of different time scales; simple back-of-the-envelope calculations based on the scaling behavior observed in these plots even yield a rough estimate of the degree of self-similarity.

Graphical methods for checking the presence of long-range dependence are based on the fact that if aggregated data with their mean subtracted can be approximated by a self-similar model, for example, fractional Gaussian noise, then one can derive statistics which, when plotted in log/log scale, would appear, within a range, as a straight line. The slope of the line is related to the self-similarity parameter H of the model. Many of these methods do not require the data to be Gaussian. One does not even need to aggregate the data at the outset since this is typically done while implementing the method.

Figure 2 displays the histogram of the AUG89.MP data for a level of aggregation of ten seconds. The other figures display the output of various graphical methods for checking long-range dependence. Figure 3 involves the R/S method, Figure 4 the Aggregated Variance method, Figure 5 the Variance of Residuals method, Figure 6 the wavelet method and Figure 7 the modified periodogram method. These methods are described in Taqqu, Teverovsky and Willinger (1995) and Taqqu and Teverovsky (1998). The wavelet method is described in Abry and Veitch (1998) and Abry, Flandrin, Taqqu and Veitch (2001). In Figure 6, it is used with the following parameters "Daubechies wavelet with 2 zero moments, starting scale $j_1 = 3$, ending scale $j_2 = 15$, no initialization" and yields $\hat{H} = 0.90$. There are also MLE-based techniques which are described in Taqqu, Teverovsky and Willinger (1995) and Taqqu and Teverovsky (1997, 1998), such as Whittle, aggregated Whittle

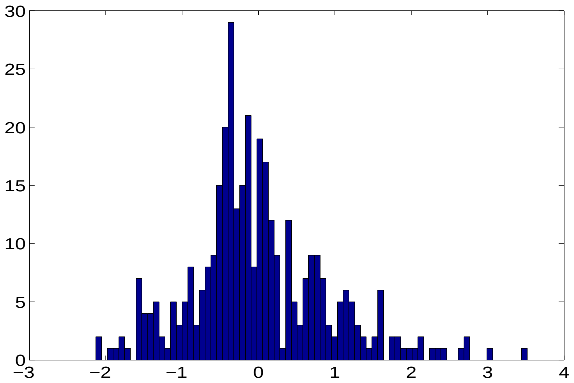


Figure 2: Histogram of aggregated data 10 sec

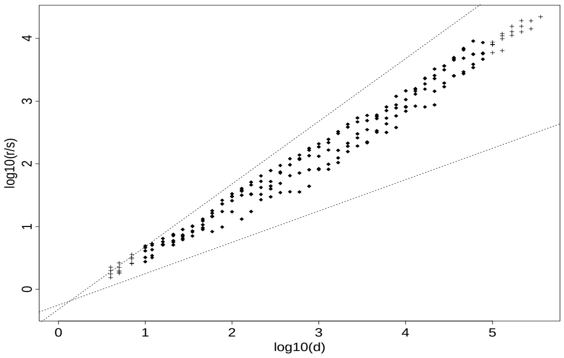


Figure 3: R/S method

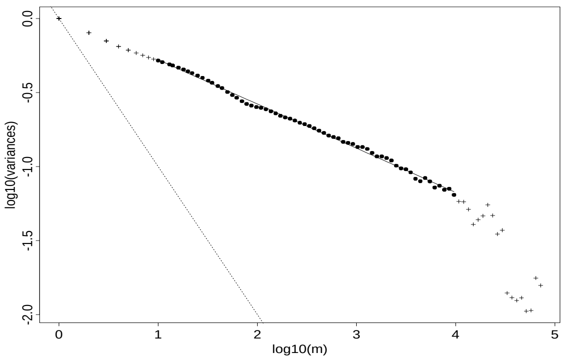


Figure 4: Aggregated Variance method

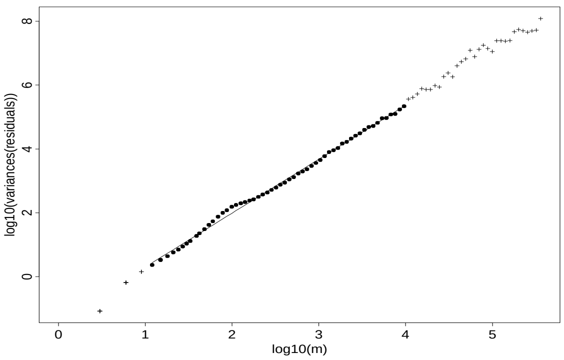


Figure 5: Variance of Residuals method

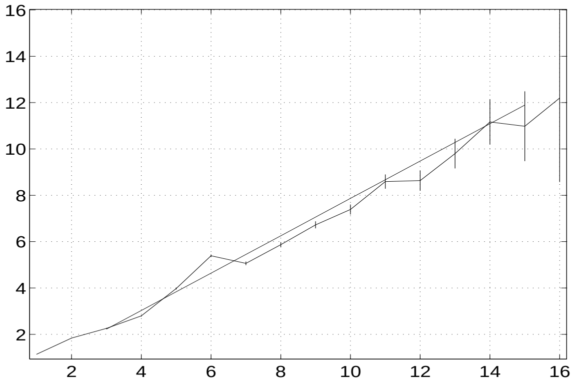


Figure 6: Wavelet method

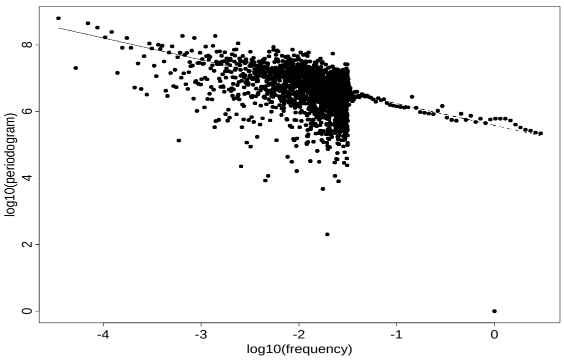


Figure 7: Modified Periodogram (log/log)

and local Whittle. Depending on the method and their tuning parameters, one gets an estimated value of H ranging from 0.85 to 0.9.

5 Explaining the presence of long-range dependence

5.1 The on-off model

The use of an on-off model to generate fractional Brownian motion can be related to ideas originally proposed by Mandelbrot (1969) in an economic context (see Taqqu and Levy (1986) for a description). In our setting, the model is supported by detailed statistical analyses of real-time traffic measurements from Ethernet LANs at the level of individual sources. These are described in Willinger, Taqqu, Sherman and Wilson (1997). Resnick and Samorodnitsky (2000) provide an overview of on-off models in a queuing context.

Consider a network with a number of workstations (sources) communicating with each other. Let us first model the behavior of a single source using an on-off alternating renewal process, as follows. The source alternates between an *on* state where it sends packets to the network and an *off* state where it is idle. Let $\{W(t),\ t \geq 0\}$ be a stationary process, where

$$W(t) = \begin{cases} 1 & \text{if time } t \text{ is an } \textit{on} \text{ interval,} \\ \\ 0 & \text{if time } t \text{ is in } \textit{off} \text{ interval.} \end{cases}$$

(We will simply say that $W(t)$ is 1 if t is *on* and 0 if t is *off*.) Viewing $W(t)$ as the reward at time t, we have a reward of 1 throughout an *on* interval, then a reward of 0 throughout the next *off* interval, then 1 again, and so on. The length of the *on* intervals are i.i.d., those of the *off* intervals are i.i.d. and the lengths of *on* and *off* intervals are independent. An *off* interval always follows an *on* interval and it is the pair of *on* and *off* intervals that defines an interrenewal period.

Let F_{on} and F_{off} denote the cumulative distribution function of the *on* and *off* interval respectively, and let $\overline{F} = 1 - F$ denote a complementary cumulative distribution function. Let also σ_{on} and σ_{off} denote the respective variances. Assume as $x \to \infty$,

$$\text{either} \;\; \overline{F}_{on}(x) \sim \ell_{on} x^{-\alpha_{on}},\ 1 < \alpha_{on} < 2 \;\; \text{or} \;\; \sigma_{on} < \infty,$$

and

$$\text{either} \;\; \overline{F}_{\mathit{off}}(x) \sim \ell_{\mathit{off}} x^{-\alpha_{\mathit{off}}},\ 1 < \alpha_{\mathit{off}} < 2 \;\; \text{or} \;\; \sigma_{\mathit{off}} < \infty.$$

Here ℓ_{on} and ℓ_{off} are constants, but they could be replaced by slowly varying functions. A slowly-varying function $L(x)$ at infinity satisfies $L(xa)/L(x) \to 1$ as $x \to \infty$ for any $a > 0$. Constants, $\log x$, $1/\log x$ are examples of slowly-varying functions. For simplification, we do not use general slowly varying functions in this section.

When $1 < \alpha_{on} < 2$, the distribution of the *on* times has infinite variance but finite expected value μ_{on}. This means that the on time can be very long with relatively high probability. We will refer to the case $\sigma_{on}^2 < \infty$ as "$\alpha_{on} = 2$." In this case, the *on* times are not likely to last very long. They can be for example exponential, as in the classical model of Poisson arrivals. Similar remarks hold for the distribution of the *off* times.

We are interested in understanding the behavior of the aggregated load $\int_0^t W(u)du$ at large times t. It has variance

$$V(t) = \mathrm{Var}\left(\int_0^t W(u)du\right) = 2\int_0^t \left(\int_0^v \gamma(u)du\right) dv \tag{5.1}$$

where

$$\gamma(u) = EW(u)W(0) - (EW(0))^2 \tag{5.2}$$

denotes the covariance function of W. One can show (see Taqqu, Willinger and Sherman (1997)) that this implies that

$$V(t) \sim \sigma^2 t^{2H} \quad \text{as} \quad t \to \infty \tag{5.3}$$

for some constant $\sigma > 0$. The constant H is given by

$$H = \frac{3 - \min(\alpha_{on}, \alpha_{off})}{2}. \tag{5.4}$$

To grasp the importance of (5.3), suppose momentarily that this relation holds. Then, with M independent and identically distributed sources, the total load at time t will be $\sum_{m=1}^{M} W^{(m)}(t)$. As $M \to \infty$, the renormalized load satisfies

$$\mathcal{L} \lim_{M\to\infty} M^{-1/2} \sum_{m=1}^{M} (W^{(m)}(t) - EW^{(m)}(t)) = G(t), \quad t \geq 0,$$

by the usual Central Limit Theorem ($\mathcal{L}$ denotes convergence of the finite dimensional distributions). The process $\{G(t),\ t \geq 0\}$ is Gaussian and stationary (since the $W_m(t)$'s are stationary) and has covariance function $\{\gamma(t),\ t \geq 0\}$. If we also aggregate and renormalize time, we get, by (5.3),

$$\mathcal{L} \lim_{T\to\infty} T^{-H} \int_0^{Tt} G(u)du = \sigma B_H(t),\ t \geq 0, \tag{5.5}$$

where $\sigma > 0$ is some constant and $\{B_H(t),\ t \geq 0\}$ is fractional Brownian motion. Indeed, the limit in (5.5) must be Gaussian with mean zero and have stationary increments since the integral of G has these properties. Moreover, by (5.3), its variance must be $\sigma^2 t^{2H}$ for fixed t. Since these properties characterize fractional Brownian motion (see Section 3), Relation (5.5) follows, and hence we have

$$\mathcal{L} \lim_{T\to\infty} \lim_{M\to\infty} \frac{1}{T^H} \frac{1}{\sqrt{M}} \int_0^{Tt} \sum_{m=1}^{M} (W^{(m)}(u) - EW^{(m)}(u))du = \sigma B_H(t),\ t \geq 0, \tag{5.6}$$

which is the result we are after. It states that the aggregated total load converges to fractional Brownian motion as the number of sources M and the time scale factor T become large (for details see Taqqu, Willinger and Sherman (1997)).

This result is tailor-made for parallel computing: letting every processor of a parallel machine generate traffic according to an alternating on-off model (same α), simply adding the outputs over all processors produces self-similar traffic. For example, producing a synthetic trace of length 100,000 on a MasPar MP-1216, a massively parallel computer with 16,384 processors, takes on the order of a few minutes. In fact, Figure 1 shows the result of a simulation where this method was to generate 27 hours worth of Ethernet-like traffic at the 10 millisecond time scale (i.e., a time series of approximately 10,000,000 observations). The value $\alpha = 1.2$ was chosen (corresponding to the estimated $H = 0.9$ that is consistent with the August 1989 data set), $M = 500$ (number of processors used to generate traffic, corresponding roughly to the number of active source-destination pairs during the observed period), and strictly alternating on-off sources with the same α-value for the distributions of the *on* and *off* periods. To check whether or not the resulting synthetic traffic trace "looks like" actual Ethernet LAN traffic as measured in August 1989, the synthetic trace is plotted on five different time scales in Figure 1 (right most column). As can be seen, this synthetic traffic passes the "visual" test easily, with the possible exception of the plot in the top row (the effect of the diurnal cycle in the 27 hour trace of Ethernet traffic on the 100s time scale becomes noticeable, especially because it is – by definition – not part of the stationary model that gave rise to the top right plot). On a more rigorous level, the trace also fits the data well in a statistical sense, i.e., the estimated H parameter matches the one from the data.

5.2 Reversing the limits

If we reverse the limits, we must change the normalization accordingly and in this case

$$\mathcal{L} \lim_{M\to\infty} \lim_{T\to\infty} \frac{1}{M^{1/\alpha}} \frac{1}{T^{1/\alpha}} \int_0^{Tt} \sum_{m=1}^{M} (W^{(m)}(u) - EW^{(m)}(u))du = \Lambda_{\alpha,\sigma,\kappa}(t), \ t \geq 0, \tag{5.7}$$

where,

$$\alpha = \min(\alpha_{on}, \alpha_{off}), \tag{5.8}$$

and where $\Lambda_{\alpha,\sigma,\kappa}(t)$ is, in general, a *skewed* Lévy stable motion of index α. The process $\Lambda_{\alpha,\sigma,\kappa}(t)$ has independent and stationary increments but infinite variance. The characteristic function of the random variable $\Lambda_{\alpha,\sigma,\kappa}(1)$ is

$$E\exp\{i\theta\Lambda_{\alpha,\sigma,\kappa}(1)\} = \exp\{-\sigma^\alpha|\theta|^\alpha - i\kappa \ \mathrm{sign}(\theta)\tan(\pi\alpha/2)\},$$

with α denoting the stability index, $\sigma > 0$ the scale coefficient and $\kappa \in [-1, 1]$ the skewness (the distribution is symmetric if $\kappa = 0$). The skewness parameter is usually denoted β but, in order to avoid confusion, we use κ instead because β appears in the sequel with a different meaning. The process $\Lambda_{\alpha,\sigma,\kappa}(1)$ is said to be *standard* if $\sigma = 1$ and *symmetric* if $\kappa = 0$. In (5.7), σ is some positive constant, $\kappa = 1$ if $\alpha_{on} < \alpha_{off}$, $\kappa = -1$ if $\alpha_{on} > \alpha_{off}$ and $\kappa \in [-1, 1]$ if $\alpha_{on} = \alpha_{off}$. To understand for example why $\kappa = 1$ if $\alpha_{on} < \alpha_{off}$, note that a renewal is made up of an *on* and *off* period. The mean reward is $EW(t) = P(W(t) > 0) = \mu_{on}/(\mu_{on} + \mu_{off}) \in (0, 1)$. Since $\alpha_{on} < \alpha_{off}$, the *on* rewards dominate. More precisely, in each renewal interval, the distribution of the cumulative reward minus the mean has a heavier tail on the right. Therefore, the limit of the normalized sum over renewals will be stable with $\kappa = 1$ (for details, see Taqqu, Willinger and Sherman (1997)).

Relations (5.6) and (5.7) can also be expressed in terms of the aggregated total load $W^*(t, M) = \int_0^t \sum_{m=1}^{M} W^{(m)}(u)du$, as follows:

$$\mathcal{L} \lim_{T\to\infty} \lim_{M\to\infty} \frac{W^*(Tt, M) - EW^*(Tt, M)}{M^{1/2}T^{(3-\alpha)/2}} = \sigma B_H(t), \ t \geq 0,$$

and

$$\mathcal{L} \lim_{M\to\infty} \lim_{T\to\infty} \frac{W^*(Tt, M) - EW^*(Tt, M)}{M^{1/\alpha}T^{1/\alpha}} = \Lambda_{\alpha,\sigma,\kappa}(t), \ t \geq 0.$$

5.3 Joint growth

What happens if T and M go together to infinity? Mikosch, Resnick, Rootzén and Stegeman (1999) assume that $M(T)$ is an integer-valued function which is non-decreasing in T and which tends to infinity as $T \to \infty$. They show that under the fast growth condition

$$\lim_{T\to\infty} \frac{M(T)}{T^{\alpha-1}} = \infty,$$

the limit is the fractional Brownian motion $B_H(t)$ and under the slow growth condition

$$\lim_{T\to\infty} \frac{M(T)}{T^{\alpha-1}} = 0,$$

the limit is the skewed Lévy stable motion $\Lambda_{\alpha,\sigma,\kappa}(t)$ which appears in (5.7). These growth conditions are motivated in Section 7.1 below.

5.4 The infinite source Poisson model

An alternative to the on-off model is the *infinite source Poisson model.* In this case sources come in at a rate of λ and each source transmits for a random time with distribution $\overline{F}_{on}(x) \sim lx^{-\alpha}$, $x \to \infty$, $1 < \alpha < 2$, and upon completing transmission, it becomes inactive. The variance of the transmission times is infinite but their mean μ_{on} is finite. The transmission times are assumed i.i.d.

Let $W(u)$ denote the number of active sources at time $u > 0$ and let $A(t) = \int_0^t W(u)du$ be the aggregated load from time 0 to time $t \geq 0$. The behavior of the system depends on the relative size of the arrival rate λ and the time horizon T. It is λ, here, which plays the role of M. Mikosch, Resnick, Rootzén and Stegeman (1999) show that under the fast growth condition

$$\lim_{T\to\infty} \frac{\lambda(T)}{T^{\alpha-1}} = \infty,$$

one has

$$\mathcal{L} \lim_{T\to\infty} \frac{A(Tt) - EA(Tt)}{\lambda^{1/2} T^{(3-\alpha)/2}} = c_1 B_H(t),\ t \geq 0,$$

and under the slow growth condition

$$\lim_{T\to\infty} \frac{\lambda(T)}{T^{\alpha-1}} = 0,$$

one has

$$\mathcal{L} \lim_{T\to\infty} \frac{A(Tt) - EA(Tt)}{\lambda^{1/\alpha} T^{1/\alpha}} = \Lambda_{\alpha,c_2,1}(t),\ t \geq 0,$$

where $\Lambda_{\alpha,c_2,1}(t)$ is a Lévy stable motion of index α with skewness parameter equal to 1. Observe that one has $EA(t) = \mu_{on}\lambda t$.

6 Heavy-tailed infinite variance signals

In the on-off case considered earlier, the signals where bounded (1 or 0) and one had an alternating renewal process. One can also consider renewal processes, where the signals have infinite variance. We can model this as a renewal reward process, the reward being the signal intensity and ask what happens to the total aggregated load when there are M workstations, and when both M and T tend to infinity.

The limit will now depend on the relative magnitudes of $1 < \alpha < 2$ (heavy-tailed exponent of the renewal interval) and $0 < \beta < 2$ (heavy-tailed exponent of the reward). Before describing what happens in more details, we briefly summarize the results. One gets the Lévy stable motion in some cases and a new process in others. Specifically, the limit is the Lévy stable motion if $\beta < \alpha$ in the case when either $T \to \infty$ first and then $M \to \infty$ or in the case when the order of M and T is reversed, and also in the case $\beta > \alpha$ when $T \to \infty$ first, followed by $M \to \infty$ (Levy and Taqqu (1987)).

The interesting case is $1 < \alpha < \beta < 2$, with $M \to \infty$ first followed by $T \to \infty$. This is the case which gives fractional Brownian motion when $\beta = 2$ (finite variance rewards). The limit turns out to be β-stable, self-similar with stationary increments (Levy and Taqqu (2000)). But it is not fractional stable motion, which is the commonly used extension of fractional Brownian motion to the stable domain. The limit is a "mixed moving average" type process which is characterized below.

In contrast with the preceding section 5, we shall allow for the presence of slowly varying functions and specify the value of the constants that appear in the limits. The results will be formulated as theorems. To help the reader, they are also summarized in Figure 8. Some aspects of the case $0 < \alpha < 1$ are discussed in Resnick and Rootzén (1998).

6.1 Assumptions on the renewals and the rewards

Assumptions on the renewal times: Let $\{U_i\}_{i \geq 1}$ be a sequence of i.i.d. random variables with range of positive integers, such that either

$$P(U \geq u) = u^{-\alpha} L_U(u),\ u = 1, 2, \ldots, \quad \alpha \in (1, 2), \tag{U1}$$

or

$$P(U = u) = \alpha u^{-\alpha - 1} l_U(u),\ u = 1, 2, \ldots, \quad \alpha \in (1, 2), \tag{U2}$$

where U denotes a generic U_i and where L_U and l_U are slowly varying functions at infinity. Let also

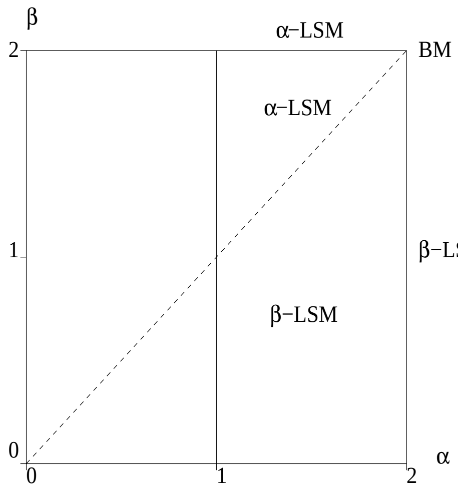
$$\mu = EU.$$

The random variable U_0 will be the first arrival time, independent of the sequence $\{U_i\}_{i \geq 1}$ and having the distribution

$$P(U_0 = u) = \frac{1}{\mu} P(U \geq u + 1),\ u = 0, 1, 2, \ldots \tag{6.1}$$

The random variables U_i will be called *interrenewal times* and they define the interrenewal intervals of a renewal process. The special choice of U_0 allows this renewal process to have stationary

$\lim_{M\to\infty} \lim_{T\to\infty}$



$\boxed{\alpha = 2}$

$\beta = \alpha = 2$: BM

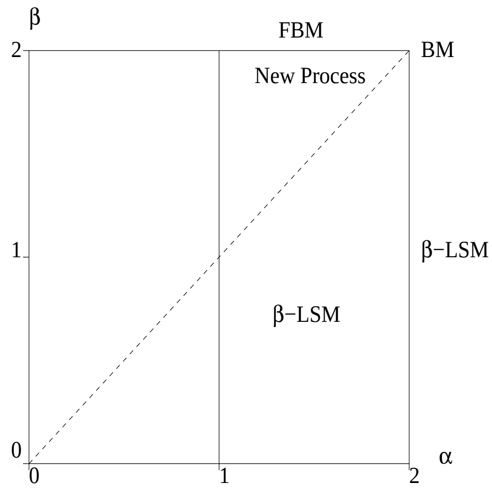
$0 < \beta < \alpha = 2$: β-LSM

$\boxed{1 < \alpha < 2}$

$1 < \alpha < \beta = 2$: α-LSM

$0 < \beta < \alpha < 2$: β-LSM

$\lim_{T\to\infty} \lim_{M\to\infty}$



$\boxed{\alpha = 2}$

$\beta = \alpha = 2$: BM

$0 < \beta < \alpha = 2$: β-LSM

$\boxed{1 < \alpha < 2}$

$1 < \alpha < \beta = 2$: FBM

$1 < \alpha < \beta < 2$: Z_β

$0 < \beta < \alpha < 2$: β-LSM

Figure 8: The figure displays the possible limits. They depend on the values of the interrenewal exponent α and of the reward exponent β. The possible limits are BM = Brownian motion, FBM = fractional Brownian motion, LSM = Lévy stable motion, Z_β = the new self-similar stable process with index β and dependent increments.

increments. It is well-known that condition (U2) implies (U1) with

$$\lim_{u\to\infty} L_U(u)/l_U(u) = 1, \tag{6.2}$$

while, for (U1) to imply (U2), the function L_U has to satisfy additional assumptions (see, for example, Bingham, Goldie and Teugels (1987)). The distribution $\overline{F}(u) = P(U \geq u)$ plays a role similar to the earlier distributions $\overline{F}_{on}(x)$ and $\overline{F}_{off}(x)$.

Assumptions on the rewards: Let $\{W_n\}_{n\geq 0}$ be a sequence of i.i.d. random variables, referred to as *signals* or *rewards*, independent of the interrenewal times sequence $\{U_n\}_{n\geq 0}$, and such that either

$$EW = 0, \qquad \sigma^2 = EW^2 < \infty \tag{FVR}$$

or

$$P(|W| \geq w) = 2P(W \geq w) = w^{-\beta}L_W(w),\ w > 0. \tag{IVR}$$

Here W denotes a generic W_n, L_W is a slowly varying function at infinity, and either

$$\alpha < \beta < 2 \tag{IVRL}$$

(the tail of the reward is lighter than the tail of the renewal) or

$$0 < \beta < \alpha, \tag{IVRH}$$

(the tail of the reward is heavier than the tail of the renewal.)

Condition (IVR) assumes that W is symmetric and implies that W has infinite variance $EW^2 = \infty$. The assumption of symmetry is not essential but it is assumed to simplify the notation.

Notation. To help the reader, we use special labels to distinguish between the various assumptions on the rewards (finite variance versus infinite variance) and on the heaviness of the tails. The labels (FVR) and (IVR) stand for "finite variance rewards" and "infinite variance rewards", respectively, and the labels (IVRL) and (IVRH) indicate that, in addition, the tails of rewards are "lighter" or "heavier" than those of the interrenewal times, respectively. Because the exponents appear with a negative sign in (IVR) and (U1), (U2), the tails of the rewards are "lighter" if their index β is greater than the index α of the interrenewal times. The labels (U1) and (U2) refer to assumptions on the interrenewal times U.

The renewal reward process is then $\{W(t),\ t \geq 0\}$, where

$$W(t) = W_n \quad \text{if} \quad t \quad \text{belongs to the } n^{th} \text{ interrenewal interval.}$$

Consider now a sequence of renewal reward processes $\{W^{(m)}(t),\ t \geq 0\}$, $m = 1, 2, \ldots$, which are i.i.d. copies of $W(t)$, and interpret each $W^{(m)}(t)$ as the the centered load of the workstation m at time t. For M workstations, the aggregated load is again

$$W^*(t, M) = \sum_{m=1}^{M} \int_0^t W^{(m)}(u)du. \tag{6.3}$$

We are interested in the behavior of $W^*(Tt, M)$, adequately normalized, as T and M go to infinity.

If $L : (0,\infty) \mapsto (0,\infty)$ is a slowly varying function at infinity and $\gamma > 0$, we also denote by L^*_γ a slowly varying function such that, for all $x > 0$,

$$L^*_\gamma(u)^{-\gamma} L(u^{1/\gamma} L^*_\gamma(u)x) \to 1, \tag{6.4}$$

as $u \to \infty$. We shall write $L^*_\gamma = L^*_U$ when $\gamma = \alpha$ and $L = L_U$, and $L^*_\gamma = L^*_W$ when $\gamma = \beta$ and $L = L_W$. It is well-known that the functions L^*_U and L^*_W appear in the normalization term for the partial sums $\sum_{k=1}^n U_k$ and $\sum_{k=1}^n W_k$ to converge to a stable random variable as $n \to \infty$.

6.2 Limiting results

The first theorem involves the limit $T \to \infty$ (second) and $M \to \infty$ (first) and deals with finite variance rewards ($\sigma^2 = EW^2 < \infty$). It is the analog to the result (5.6) in the on-off setup.

Theorem 6.1 (Taqqu and Levy (1986), Theorem 6, (ii)). *Under assumptions* (U1) *on the renewals and* (FVR) *on the rewards,*

$$\mathcal{L} \lim_{T\to\infty} \lim_{M\to\infty} \frac{W^*(Tt, M)}{T^{(3-\alpha)/2}(L_U(T))^{1/2} M^{1/2}} = \sigma_0 B_H(t), \tag{6.5}$$

where $t \in [0,1]$, B_H is a standard fractional Brownian motion with parameter $H = (3-\alpha)/2$ and $\sigma_0^2 = 2\sigma^2(\mu(\alpha-1)(2-\alpha)(3-\alpha))^{-1}$.

The next theorem characterizes the limit when the rewards in Theorem 6.1 have infinite variance instead.

Theorem 6.2 (Levy and Taqqu (2000), Theorem 2.1, and Pipiras and Taqqu (2000), Proposition 2.1). *Under assumptions* (U2) *on the renewals and* (IVRL) *on the rewards, namely, if $\alpha < \beta < 2$, then*

$$\mathcal{L} \lim_{T\to\infty} \lim_{M\to\infty} \frac{W^*(Tt, M)}{(T^{\beta-\alpha+1} l_U(T))^{1/\beta} M^{1/\beta} L^*_W(M)} = (\mu C_\beta)^{-1/\beta} Z_\beta(t), \tag{6.6}$$

where $t \in [0,1]$ and

$$C_\beta^{-1} = \Gamma(2-\beta)\cos(\pi\beta/2)/(1-\beta). \tag{6.7}$$

The limit process Z_β is a symmetric β-stable process described below.

The following integral representation for Z_β provides some insight about its structure.

Proposition 6.1 *The process Z_β can be represented in the sense of the finite-dimensional distributions as*

$$Z_\beta(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ((t\wedge v - u)_+ - (0\wedge v - u)_+) M(du, dv) \tag{6.8}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\int_0^t 1_{\{u<s<v\}} ds\right) M(du, dv), \tag{6.9}$$

where M is an independently scattered symmetric β-stable random measure on $\mathbb{R}^2$ with a control measure

$$m(du, dv) = \alpha(v-u)_+^{-\alpha-1} du dv. \tag{6.10}$$

The random measure M has a simple heuristic meaning. It is random noise, independent on disjoint rectangles (du, dv). The noise has a β-stable distribution, that is, the characteristic function of the random variable $M(\Delta)$, $\Delta \in \mathbb{R}^2$, is $E\exp\{i\theta M(\Delta)\} = \exp\{-|\theta|^\beta m(\Delta)\}$, where $m(\Delta) = \int_\Delta m(du, dv)$ is the scale parameter. (For a precise definition of the random measure M, see Section 3.3 in Samorodnitsky and Taqqu (1994)).

As shown in Levy and Taqqu (2000), the process Z_β in (6.8) has stationary *dependent* increments and is self-similar with exponent

$$H = (\beta - \alpha + 1)/\beta. \tag{6.11}$$

The process Z_β has (a version with) continuous paths. This is because, for any $p < \beta$, the p-th moment $E|Z_\beta(t)|^p$ is finite and, by stationarity of the increments and H self-similarity of the process Z_β,

$$E|Z_\beta(t_2) - Z_\beta(t_1)|^p = E|Z_\beta(t_2 - t_1)|^p = E|Z_\beta(1)|^p\, |t_2 - t_1|^{Hp} = C|t_2 - t_1|^{Hp}. \tag{6.12}$$

Since

$$H = \frac{\beta - \alpha + 1}{\beta} > \frac{1}{\beta} \quad \text{or} \quad \beta > \frac{1}{H}$$

we can choose $p < \beta$ such that $pH > 1$. The existence of a continuous modification (version) of Z_β then follows from (6.12) and Kolmogorov's criterion.

Note that, if one sets $\beta = 2$ (the case of finite variance), one recovers the self-similarity exponent H given in Theorem 6.1. In fact, supposing that all slowly varying functions asymptotically equal 1, by setting $\beta = 2$ in the normalization $T^{(\beta-\alpha+1)/\beta} M^{1/\beta}$ of Theorems 6.2, one recovers the normalization $T^{(3-\alpha)/2} M^{1/2}$ used in Theorem 6.1. One may thus view Theorem 6.1 as the boundary case $\beta = 2$ of Theorems 6.2.

The process Z_β is not one of the linear fractional stable motions (LFSM)

$$L_\beta(t) = L_{\beta,H}(a, b; t) = \int_{-\infty}^{\infty} f_{\beta,H}(a, b; t, u) M(du), \tag{6.13}$$

where

$$f_{\beta,H}(a, b; t, u) = a\left((t-u)_+^{H-1/\beta} - (-u)_+^{H-1/\beta}\right) + b\left((t-u)_-^{H-1/\beta} - (-u)_-^{H-1/\beta}\right), \tag{6.14}$$

a and b are two reals such that $|a| + |b| > 0$, the self-similarity index $H \in (0,1)$ and M is an independently scattered symmetric β-stable random measure on $\mathbb{R}$ with Lebesgue control measure (see, for example, Section 7.4 in Samorodnitsky and Taqqu (1994)). The LFSMs (6.13) are often considered the "natural" extensions of fractional Brownian motion to the infinite variance stable case. Pipiras and Taqqu (2000) prove that the process Z_β that appears in the limit in Theorem 6.2 is not LFSM.

The next result describes what happens when $\beta < \alpha$, that is, when the tail of the reward distribution is heavier than that of the renewals. The limit in this case is the independent increments Lévy stable motion.

Theorem 6.3 (Levy and Taqqu (2000), Theorem 2.1). *Under assumptions* (U2) *on the renewals and* (IVRH) *on the rewards, namely, if* $0 < \beta < \alpha$, *then*

$$\mathcal{L} \lim_{T\to\infty} \lim_{M\to\infty} \frac{W^*(Tt, M)}{T^{1/\beta} M^{1/\beta} L_W^*(M)} = (\mu C_\beta)^{-1/\beta} (EU^\beta)^{1/\beta} \Lambda_{\beta,1,0}(t), \tag{6.15}$$

where $t \in [0,1]$, and C_β is defined as in (6.7) and where $\Lambda_{\beta,1,0}(t)$ is a standard symmetric Lévy stable motion with index β.

If we reverse the order of the limits and consider $M \to \infty$ (second), $T \to \infty$ (first), then after changing the normalization factors accordingly, the limit is Lévy stable motion. But, depending on the assumption on the rewards, this limiting process can have index α or β.

Theorem 6.4 (Taqqu and Levy (1986), Theorem 6, (i), and Levy and Taqqu (1987), Theorem 1). *Under assumptions* (U1) *on the renewals,*

$$\mathcal{L}\lim_{M\to\infty}\lim_{T\to\infty}\frac{W^*(Tt,M)}{M^{1/\alpha}T^{1/\alpha}L_U^*(T)} = (\mu C_\alpha)^{-1/\alpha}(E|W|^\alpha)^{1/\alpha}\Lambda_{\alpha,1,0}(t) \tag{6.16}$$

if $EW^2 < \infty$ (assumption FVR) or $1 < \alpha < \beta < 2$ (assumption IVRL), and

$$\mathcal{L}\lim_{M\to\infty}\lim_{T\to\infty}\frac{W^*(Tt,M)}{M^{1/\beta}T^{1/\beta}L_W^*(T)} = (\mu C_\beta)^{-1/\beta}(EU^\beta)^{1/\beta}\Lambda_{\beta,1,0}(t) \tag{6.17}$$

if $0 < \beta < \alpha < 2$ (assumption IVRH). The processes $\Lambda_{\alpha,1,0}(t)$ and $\Lambda_{\beta,1,0}(t)$, $t \in [0,1]$ are standard symmetric Lévy stable motions with index α and β, respectively.

Suppose now that M is a function of T and that $M = M(T) \to \infty$, as $T \to \infty$. We want to know when the total aggregated load $W^*(T\cdot, M(T))$ converges to any of the above limits as $T \to \infty$. As previously mentioned, we have to distinguish between two cases. Since the limit process in Theorems 6.3 and 6.4 is the same under the assumption (IVRH), we may expect $W^*(T\cdot, M(T))$ to always converge to this limit as $T \to \infty$. This is indeed the case. Under the assumptions (FVR) and (IVRL), however, the limit processes in Theorems 6.1 or 6.2, and Theorem 6.4 are different. In this case, we shall find conditions $M(T)$ for the normalized total aggregated load $W^*(T\cdot, M(T))$ to converge to any of the three limits obtained in the above theorems.

7 Joint convergence when the signals have infinite variance

We continue to view the signals as rewards in a renewal reward process and ask what happens as T and M go to infinity together. The idea of taking M as a function of T comes from Mikosch, Resnick, Rootzén and Stegeman (1999), who considered the on-off case described in Subsection 5.3, where the rewards take values 1 or 0. We focus here on the convergence of the total aggregated process $W^*(Tt, M)$ under the assumptions (FVR) and (IVRL). For details see Pipiras, Taqqu and Levy (2000).

7.1 Fast and slow growth conditions under finite variance rewards (FVR)

First, a matter of notation. To indicate that a function $a(T)$ grows to infinity faster than a function $b(T)$, we use below the suggestive notation $a(T) \gg b(T)$, instead of $b(T) = o(a(T))$.

We want to get an an idea about the conditions on $M(T)$ that are needed for the convergence of $W^*(T\cdot, M(T))$. In the heuristics below, we shall suppose that all slowly varying functions asymptotically constant.

It is best to consider first the case of finite variance rewards. In view of Theorem 6.4, when T is large compared to M, we still expect $W^*(T\cdot, M(T))/(M^{1/\alpha}T^{1/\alpha})$ to converge to stable Lévy motion. The question is how fast the parameter $M(T)$ should grow with T. Observe that, for finite T and M, the process $W^*(T\cdot, M)$ has always finite variance, since it involves interrenewal times that cannot be greater than T. Moreover, by Taqqu and Levy (1986), p. 87, the standard deviation

$$\left(E|W^*(T,M)|^2\right)^{1/2} \sim \text{const } T^{(3-\alpha)/2}M^{1/2}$$

for large T and M. Therefore, if we set $M = M(T)$ and expect the limit of $W^*(T\cdot, M)/(M^{1/\alpha}T^{1/\alpha})$ to be the (infinite variance) stable Lévy motion as $T \to \infty$, then the standard deviation of $W^*(T\cdot, M)/(M^{1/\alpha}T^{1/\alpha})$ should diverge. In other words,

$$\frac{\left(E|W^*(T,M)|^2\right)^{1/2}}{M^{1/\alpha}T^{1/\alpha}} \sim \text{const } \frac{T^{(3-\alpha)/2}M^{1/2}}{M^{1/\alpha}T^{1/\alpha}} \gg 1,$$

or,

$$\begin{aligned} T^{(3-\alpha)\alpha-2} &\gg M^{2-\alpha} \\ T^{(2-\alpha)(\alpha-1)} &\gg M^{2-\alpha} \\ T^{\alpha-1} &\gg M, \end{aligned}$$

as $T \to \infty$. This is the regime where we expect stable Lévy motion in the limit. On the other hand, when $T^{\alpha-1} << M$, we expect the normalization of $W^*(T\cdot, M(T))$ to be $T^{(3-\alpha)/2}M(T)^{1/2}$ and the limiting process to be fractional Brownian motion.

7.2 Fast and slow growth conditions under infinite variance rewards, lighter than interrenewal times (IVRL)

If the rewards have heavy tails with index $\beta > \alpha$, one may argue in a similar way. Observe that, since $E|W|^\alpha < \infty$ for $\alpha < \beta$, the process $W^*(T, M)$ has always the α^{th} moment and the convergence (6.15) suggests that

$$(E|W^*(T,M)|^\alpha)^{1/\alpha} \sim \text{const } T^{(\beta-\alpha+1)/\beta}M^{1/\beta}$$

for large T and M. Now, if we expect $W^*(T\cdot, M(T))/M^{1/\alpha}T^{1/\alpha}$ to converge to stable Lévy motion, which has an infinite α^{th} moment, the normalization $M^{1/\alpha}T^{1/\alpha}$ should satisfy

$$\frac{T^{(\beta-\alpha+1)/\beta}M^{1/\beta}}{M^{1/\alpha}T^{1/\alpha}} \gg 1 \tag{7.1}$$

or,

$$\begin{aligned} T^{(\beta-\alpha+1)/\beta}M^{1/\beta} &\gg T^{1/\alpha}M^{1/\alpha} \\ T^{(\beta-\alpha+1)/\beta-1/\alpha} &\gg M^{1/\alpha-1/\beta} \\ T^{(\beta-\alpha+1)\alpha-\beta} &\gg M^{\beta-\alpha} \\ T^{(\alpha-1)(\beta-\alpha)} &\gg M^{\beta-\alpha} \\ T^{\alpha-1} &\gg M, \end{aligned}$$

as $T \to \infty$. These computations are illuminating because they show that condition (7.1) does not in fact involve β, and is therefore the same as the one obtained in the case of finite variance rewards. When $M \gg T^{\alpha-1}$, we expect the normalization of $W^*(T\cdot, M(T))$ to be $T^{(\beta-\alpha+1)/\beta}M(T)^{1/\beta}$ and the limiting process to be Z_β.

7.3 Limits under fast and slow growth

While in the preceding heuristics, we assumed that the slowly varying functions were asymptotically constant, we now drop this simplifying assumption and provide the rates of convergence in the general case. We shall consider the following two regimes:

Fast growth:

$$\lim_{T\to\infty} \frac{M}{T^{\alpha-1}}(L_U^*(MT))^\alpha = \infty, \tag{7.2}$$

Slow growth:

$$\lim_{T\to\infty} \frac{M}{T^{\alpha-1}}(L_U^*(MT))^\alpha = 0. \tag{7.3}$$

The following theorems are our main results under the assumptions (FVR) and (IVRL).

Theorem 7.1 *Under the fast growth condition (7.2) above and assumptions* (U1) *on the renewals and* (FVR) *on the rewards,*

$$\mathcal{L}\lim_{T\to\infty} \frac{W^*(Tt,M)}{T^{(3-\alpha)/2}M^{1/2}(L_U(T))^{1/2}} = \sigma_0 B_H(t), \tag{7.4}$$

where $t \in [0,1]$, σ_0 is given in Theorem 6.1 and B_H is a standard FBM.

Theorem 7.2 *Under the fast growth condition (7.2) above and assumptions* (U2) *on the renewals and* (IVRL) *on the rewards,*

$$\mathcal{L}\lim_{T\to\infty} \frac{W^*(Tt,M)}{T^{(\beta-\alpha+1)/\beta}M^{1/\beta}(l_U(T))^{1/\beta}L_W^*(T^{-\alpha+1}Ml_U(T))} = (\mu C_\beta)^{-1/\beta} Z_\beta(t), \tag{7.5}$$

where $t \in [0,1]$ and Z_β is the symmetric β-stable process of Theorem 6.2.

In some cases, one may need a stronger convergence result than convergence in the sense of the finite-dimensional distributions. It is fairly easy to show that the convergence (7.4) or (7.5) can be extended to weak convergence in the space $D[0,1]$ equipped with the usual Skorokhod J_1–topology. (Recall that $D[0,1]$ is the space of right-continuous functions on $[0,1]$ which have left-limits.) One can apply the usual argument, that the limit processes B_H and Z_β are smooth enough. Their path regularity is easy to establish. For example, since $H = (3-\alpha)/2 > 1/2$, fractional Brownian motion B_H is long-range dependent with smoother paths as H increases and the well-known Kolmogorov's criterion applies to $E|B_H(t) - B_H(s)|^2$ (there is no need to consider a power higher than 2). Therefore, (7.4) and (7.5) extend to weak convergence in the space $D[0,1]$ equipped with the J_1–topology.

We now consider what happens under the slow growth condition (7.3).

Theorem 7.3 *Under the slow growth condition (7.3) and assumptions* (U1) *on the renewals and either* (FVR) *or* (IVRL) *on the rewards,*

$$\mathcal{L}\lim_{T\to\infty} \frac{W^*(Tt,M)}{T^{1/\alpha}M^{1/\alpha}L_U^*(TM)} = (\mu C_\alpha)^{-1/\alpha}(E|W|^\alpha)^{1/\alpha}\Lambda_{\alpha,1,0}(t), \tag{7.6}$$

where $t \in [0,1]$ and $\Lambda_{\alpha,1,0}$ is the standard symmetric Lévy motion as in Theorem 6.4 under the assumptions (FVR) *or* (IVRL).

7.4 When growth conditions are not required

Finally we consider the assumption (IVRH), namely $0 < \beta < \alpha < 2$, and conclude with a result which requires no assumptions on the growth of the function $M = M(T)$.

Theorem 7.4 *Suppose that $M = M(T) \to \infty$, as $T \to \infty$. Then, under the assumptions* (U2) *on the renewals and* (IVRH) *on the rewards,*

$$\mathcal{L} \lim_{T\to\infty} \frac{W^*(Tt, M)}{T^{1/\beta} M^{1/\beta} L_W^*(TM)} = (\mu C_\beta)^{-1/\beta} (EU^\beta)^{1/\beta} \Lambda_{\beta,1,0}(t), \tag{7.7}$$

where $t \in [0, 1]$, and $\Lambda_{\beta,1,0}$ is the standard symmetric Lévy motion as in Theorem 6.4 under the assumption (IVRH).

8 Outlook

The processes obtained here provide a statistical description of the traffic as seen on a single link and when aggregated over all users/connections and over sufficiently large time periods. They model deviations of the signal from its mean and become apparent only when the time scale T is large. These models, however, are not adequate at small time scales. By their very definition, asymptotic properties (in the limit as we look over larger and larger time scales) leave the small-time scaling behavior of the measured traffic essentially unspecified and allow for different possible descriptions of the fine structure of the observed traffic. In order to develop models of measured network traffic capable of accounting for small time scales, we need a new mathematical framework, one that can handle highly irregular or regular behavior well-localized in time. This leads to the study of the differentiability, fractality and regularity of stochastic processes; that is, properties that are inherently related to the high-frequency behavior or small-lag correlations and have little to do with low-frequency components and/or high-lag correlations. One is then led to measures of local behavior such as the fractal Hausdorff dimension or Hölder exponents, and to use multifractals. For details, see for example Willinger, Paxson, Riedi and Taqqu (2001).

Taqqu, Teverovsky and Willinger (1997) perform a time-domain based multifractal analysis on the Ethernet data and conclude that the use of multifractals is not necessary to describe that data set. Riedi and Lévy Véhel (1997) consider a wider range of TCP/IP wide area network data and reach a similar conclusion, as do Mannersalo and Norros (1997) who analyzed various short time series of measured aggregate wide area network ATM traffic and basically confirmed the results reported in (1997).

A wavelet-domain based approach, however, can be more sensitive to the possibility of multifractal scaling in network traffic. It was originally proposed by Feldmann et al. (1998, 1999) in an effort to validate the findings reported in Riedi and Lévy Véhel (1997) using different data sets. The wavelet-domain based approach exploits the fact that irregular/regular patterns in a given signal can be well localized (in time and space) and studied by considering the wavelet coefficients associated with the signal's wavelet decomposition. Feldmann et al. (1998, 1999) demonstrate that one can associate known network-specific features with qualitative characteristics (e.g., scaling regions, "breakpoints", "spikes" or "dips") in the output of the wavelet scaling analysis of the data. A relatively regular pattern that is well localized in time and scale will give rise to corresponding small wavelet coefficients; in contrast, well-localized irregular patterns can be expected to yield large wavelet coefficients in the corresponding time-scale or wavelet-domain "neighborhood" of the underlying signal.

Despite progress in relating certain network-specific features to the output of the wavelet domain-based scaling analysis, nothing close to a physical explanation of the observed small-time scale behavior (including multifractal scaling) exists to date. Yet, for any type of observed scaling behavior to have a genuine impact on networking, it is essential that their applications move beyond the traditional descriptive stage and yield answers as to why network traffic exhibits a certain scaling property.

Acknowledgments

This research was partially supported by the NSF Grant ANI-9805623 at Boston University. I would like also to thank Vladas Pipiras, Stilian Stoev and the referees for their comments.

References

Abry, P. & Veitch, D. (1998), 'Wavelet analysis of long range dependent traffic', *IEEE Transactions on Information Theory* **44(1)**, 2–15.

Abry, P., Flandrin, P., Taqqu, M. S. & Veitch, D. (2001), Self-similarity and long-range dependence through the wavelet lens, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, 'Long-range Dependence: Theory and Applications', Birkhäuser. Preprint.

Beran, J., Sherman, R., Taqqu, M. S. & Willinger, W. (1995), 'Long-range dependence in variable-bit-rate video traffic', *IEEE Transactions on Communications* **43**, 1566–1579.

Bingham, N. H., Goldie, C. M. & Teugels, J. L. (1987), *Regular Variation*, Cambridge University Press.

Feldmann, A., Gilbert, A. & Willinger, W. (1998), Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic, *in* 'Proceedings of the ACM/SIGCOMM'98', Vol. 28, Vancouver, Canada, pp. 42–55.

Feldmann, A., Gilbert, A., Huang, P. & Willinger, W. (1999), 'Dynamics of IP traffic: A study of the role of variability and the impact of control', *Computer Communication Review* **29**(4), 301–313. Proceedings of the ACM/SIGCOMM'99, Cambridge, MA.

Garrett, M. W. & Willinger, W. (1994), Analysis, modeling and generation of self-similar VBR video traffic, *in* 'Proceedings of the ACM Sigcomm '94, London, UK', pp. 269–280.

Kolmogorov, A. N. (1941), 'Local structure of turbulence in an incompressible liquid for very large Reynolds numbers', *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS (N.S.)* **30**, 299–303. Reprinted in Friedlander & Topper *Turbulence: classic papers on statistical theory*, 1961 New York: Interscience.

Leland, W. E. & Wilson, D. V. (1991), High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection, *in* 'Proceedings of the Infocom '91', Bal Harbour, FL, pp. 1360–1366.

Leland, W. E., Taqqu, M. S., Willinger, W. & Wilson, D. V. (1994), 'On the self-similar nature of Ethernet traffic (Extended version)', *IEEE/ACM Transactions on Networking* **2**, 1–15.

Levy, J. & Taqqu, M. S. (1987), 'On renewal processes having stable inter-renewal intervals and stable rewards', *Les Annales des Sciences Mathématiques du Québec* **11**, 95–110.

Levy, J. B. & Taqqu, M. S. (2000), 'Renewal reward processes with heavy-tailed interrenewal times and heavy-tailed rewards', *Bernoulli* **6**(1), 23–44.

Mandelbrot, B. B. (1969), 'Long-run linearity, locally Gaussian processes, H-spectra and infinite variances', *International Economic Review* **10**, 82–113.

Mannersalo, P. & Norros, I. (1997), Multifractal analysis of real ATM traffic: a first look, Technical report, *COST257TD*, VTT Information Technology.

Mikosch, T., Resnick, S., Rootzén, H. & Stegeman, A. (1999), Is network traffic approximated by stable Lévy motion or fractional Brownian motion?, Preprint.

Park, K. & Willinger, W., eds (2000), *Self-Similar Network Traffice and Performance Evaluation*, J. Wiley & Sons, Inc., New York.

Pipiras, V. & Taqqu, M. S. (2000), 'The limit of a renewal-reward process with heavy-tailed rewards is not a linear fractional stable motion', *Bernoulli* **6**, 607–614.

Pipiras, V., Taqqu, M. S. & Levy, J. B. (2000), Slow, fast and arbitrary growth conditions for renewal reward processes when the renewals and the rewards are heavy-tailed, Preprint.

Resnick, S. I. & Rootzén, H. (1998), Self-similar communication models and very heavy tails, ORIE technical report 1228, Cornell University.

Resnick, S. I. & Samorodnitsky, G. (2000), Fluid queues, leaky buckets, on-off processes and teletraffic modeling with highly variable and correlated inputs, *in* K. Park & W. Willinger, eds, 'Self-Similar Network Traffic and Performance Evaluation', Wiley (Interscience Division).

Riedi, R. H. & Véhel, J. L. (1997), Multifractal properties of TCP traffic: a numerical study, INRIA Research Report 3129. See also: J. Lévy Véhel and R. Riedi, "Fractional Brownian motion and Data Traffic Modelling," in *Fractals in Engineering*, pp. 185-202, Springer.

Samorodnitsky, G. & Taqqu, M. S. (1994), *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York, London.

Schoch, J. F. & Hupp, J. A. (1980), 'Measured performance of an ethernet local network', *Communications of the ACM* **23**, 711–721.

Taqqu, M. S. & Levy, J. (1986), Using renewal processes to generate long-range dependence and high variability, *in* E. Eberlein & M. S. Taqqu, eds, 'Dependence in Probability and Statistics', Birkhäuser, Boston, pp. 73–89.

Taqqu, M. S. & Teverovsky, V. (1997), 'Robustness of Whittle-type estimates for time series with long-range dependence', *Stochastic Models* **13**, 723–757.

Taqqu, M. S. & Teverovsky, V. (1998), On estimating the intensity of long-range dependence in finite and infinite variance series, *in* R. Adler, R. Feldman & M. S. Taqqu, eds, 'A Practical Guide to Heavy Tails: Statistical Techniques and Applications', Birkhäuser, Boston, pp. 177–217.

Taqqu, M. S., Teverovsky, V. & Willinger, W. (1995), 'Estimators for long-range dependence: an empirical study', *Fractals* **3**(4), 785–798. Reprinted in *Fractal Geometry and Analysis*, C.J.G. Evertsz, H-O Peitgen and R.F. Voss, editors. World Scientific Publishing Co., Singapore, 1996.

Taqqu, M. S., Teverovsky, V. & Willinger, W. (1997), 'Is network traffic self-similar or multifractal?', *Fractals* **5**, 63–74.

Taqqu, M. S., Willinger, W. & Sherman, R. (1997), 'Proof of a fundamental result in self-similar traffic modeling', *Computer Communications Review* **27**(2), 5–23.

Willinger, W., Paxson, V., Riedi, R. H. & Taqqu, M. S. (2001), Long-range dependence and data network traffic, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, 'Long-range Dependence: Theory and Applications', Birkhäuser. Preprint.

Willinger, W., Taqqu, M. S. & Erramilli, A. (1996), A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks, *in* F. P. Kelly, S. Zachary & I. Ziedins, eds, 'Stochastic Networks: Theory and Applications', Clarendon Press, Oxford, pp. 339–366.

Willinger, W., Taqqu, M. S., Leland, W. E. & Wilson, V. (1995), 'Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements', *Statistical Science* **10**, 67–85.

Willinger, W., Taqqu, M. S., Sherman, R. & Wilson, D. V. (1997), 'Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level', *IEEE/ACM Transactions in Networking* **5**(1), 71–86. Extended Version of the paper with the same title that appeared in *Computer Communications Review* Vol. 25 (1995) 100-113.

Murad S. Taqqu
Department of Mathematics
Boston University
111 Cummington St.
Boston, MA 02215, USA
murad@math.bu.edu