*Statistical Science*
1995, Vol. 10, No. 1, 67–85

# Self-Similarity in High-Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements

**Walter Willinger, Murad S. Taqqu, Will E. Leland and Daniel V. Wilson**

*Abstract.* Traffic modeling of today's communication networks is a prime example of the role statistical inference methods for stochastic processes play in such classical areas of applied probability as queueing theory or performance analysis. In practice, however, statistics and applied probability have failed to interface. As a result, traffic modeling and performance analysis rely heavily on subjective arguments; hence, debates concerning the validity of a proposed model and its predicted performance abound.

In this paper, we show how a careful statistical analysis of large sets of actual traffic measurements can reveal new features of network traffic that have gone unnoticed by the literature and, yet, seem to have serious implications for predicted network performance. We use hundreds of millions of high-quality traffic measurements from an Ethernet local area network to demonstrate that Ethernet traffic is statistically self-similar and that this property clearly distinguishes between currently used models for packet traffic and our measured data. We also indicate how such a unique data set (in terms of size and quality) (i) can be used to illustrate a number of different statistical inference methods for self-similar processes, (ii) gives rise to new and challenging problems in statistics, statistical computing and probabilistic modeling and (iii) opens up new areas of mathematical research in queueing theory and performance analysis of future high-speed networks.

*Key words and phrases:* Self-similarity, long-range dependence, Hurst effect, $R/S$ analysis, variance-time analysis, periodogram-based estimation, traffic modeling for high-speed networks, queueing theory.

## 1. INTRODUCTION

### 1.1 From Megabit to Gigabit Networks: LAN, MAN and BISDN

Local area networks (LAN's) were introduced in the mid-1970's to interconnect data processing equipment (host computers, file servers, PC's, workstations, terminals, printers, plotters etc.) in office or R & D environments, or within university departments. The Ethernet was an early LAN technology (for a detailed description of the Ethernet technology, see [47]), and it remains one of the most popular in use today. Among the features that make Ethernets so successful are ease of maintenance and administration, ease of network reconfiguration (stations can be moved—disconnected from one point and reconnected at another—without the need to take down the whole network), 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. From today's perspective, two of the major disadvantages of Ethernets are their relatively slow speed of 10 Mbit/s and their limited range (their physical span is limited to a few kilometers, i.e., to a small campus, a single building or to just one floor of a building). The increasing availability of high-performance workstations with sustained I/O bus bandwidths of

*Walter Willinger, Will Leland, and Daniel V. Wilson are Members of the Technical Staff at Bellcore, 445 South Street, Morristown, New Jersey 07960. Murad S. Taqqu is Professor, Department of Mathematics, Boston University, Boston, Massachusetts 02215.*

100 Mbit/s or more, supercomputers and parallel machines is a driving force toward higher speed LAN's (so-called *gigabit LAN's*) with at least 100 times more bandwidth than today's Ethernets. The success and large number of existing LAN's (of the order of hundreds of thousands) is a major cause for the current proliferation of *metropolitan area networks* (MAN's), that is, systems capable of interconnecting different LAN's within a limited geographic area of about 100 km (e.g., university campus or a small community).

Due to the increasing demand for MAN's and gigabit LAN's, and given the technological progress in the areas of transmission and switching, the current trend in telecommunication has been moving away from the existing service-specific networks (i.e., separate networks for voice, data and video) toward a single, service-independent, flexible and efficient network, the so-called *broadband integrated services digital network* (BISDN). The transfer mode traditionally used in telephone networks is circuit communication, where a dedicated circuit is provided for the complete duration of a connection. In contrast, BISDN is based on a high-speed packet communication methodology that enables all services to be transported and switched in a common digital form, namely, as fixed-sized 53-octet packets. Clearly, packet communication is more complex than circuit communication since it generates traffic spanning vastly different time scales (from microseconds to seconds and minutes).

In the absence of practical experience with gigabit networks, the challenge for today's traffic engineers is to gain an understanding of the likely characteristics of future BISDN traffic. Due to the already existing need, LAN interconnection services are expected to become an immediate and major BISDN application. Therefore, understanding the characteristics of LAN traffic such as Ethernet traffic can provide valuable insight into the time dynamics of realistic future BISDN and other gigabit network traffic scenarios. It will help guide future choices of appropriate traffic models, which in turn will be used as inputs to queueing systems in order to assess problems related to the economic design, control and performance of future networks.

This paper uses statistical techniques to arrive at an Ethernet traffic model. In doing so, it deliberately diverges from the usual approach in telecommunication systems, where traffic models are typically judged by how well they predict the performance of the queueing model alone, and almost never by how well the model fits actual traffic data in a statistical sense. In fact, in the related literature on teletraffic models, one rarely finds a validation of a given model against actual data; at the same time, comparisons of Monte Carlo simulations of a given model with one's analytic computations abound. As discussed in [44], in teletraffic practice, the choice and validation of most queueing models depend heavily on the modeler's intuitive understanding of the application at hand and on extramathematical considerations. Most arguments concerning the validity of the underlying models and the resulting performance predictions stem from this subjective quality of the modeling. Here, instead, a detailed analysis of actual traffic data drives the conclusion—that Ethernet traffic is statistically self-similar.

### 1.2 Self-Similarity and Long-Term Correlations in Teletraffic

Self-similar processes were introduced by Kolmogorov [26]. They were brought to the attention of statisticians and probabilists by Mandelbrot and his co-workers in the late 1960's and early 1970's. Intuitively, their attractive feature is that they look the same at different scales. This property suggests (see [38]) the existence of a multilevel hierarchy of underlying mechanisms (in the case at hand, one per time scale) whose combined effect is the same as that of self-similarity. While it is tempting to invoke such a hierarchical structure to account for self-similarity in nature, it is not easy to demonstrate its physical reality (see, however, [8] for self-similar models of hierarchical variation in a textile context and [6] in a physical context). A rather different construction that also attempts to provide a phenomenological explanation for the observed self-similarity in a given data set and that is especially appealing in the teletraffic context considered in this paper can be found in [36] and will be discussed in more detail in Section 5.

In terms of stochastic modeling, self-similar processes or their increment processes are almost exclusively used in situations where the modeler tries to account for the presence of long-term correlations in a parsimonious manner (see Section 2). As is cogently discussed in [18], it seems to be the rule rather than the exception that long time series of absolute measurements violate the assumption of independence or short-range dependence. Due to their by now well-documented omnipresence in many naturally occurring empirical records, statistical methods for data with long-range dependence are slowly making their way into the mainstream (e.g., see [3]).

The presence of long-term correlation in traffic measurements taken from communication systems *other than* Ethernet LAN's (e.g., certain types of video traffic) has recently been demonstrated in [4] and [13]. In the case of the Ethernet data, our

analysis below clearly shows that self-similar models fit Ethernet LAN 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. We know of no examples in the literature of empirical records where the use of a self-similar model becomes evident simply by plotting the raw data on a number of different time scales, the obvious reason being the absence of really large and high-quality data sets (typically, data sets from other disciplines reported in the literature contain a few hundreds or a few thousands of observations). Furthermore, for the Ethernet data, computationally fast graphical methods for estimating the degree of self-similarity in a given set of data turn out to be extremely accurate due to the extraordinarily large number of available observations. Finally, both size and quality of the traffic data allow for computationally intensive but statistically rigorous estimation methods which result in a clear and coherent picture of the self-similar nature of Ethernet traffic.

The most striking result of our analysis of the Ethernet traffic measurements is that, in a statistical sense, one can clearly distinguish the measured Ethernet data from data predicted by practically all the stochastic models for packet traffic currently considered in the literature, including Markov-modulated Poisson processes [20], fluid-flow models [1], ARMA models, TES processes [24] and packet-train models [25]. While it is becoming common knowledge among statisticians that ignoring long-range dependence can have drastic consequences for many statistical inference methods (see [18] or [3]), only direct arguments, detailing the impact on network performance, will convince the teletraffic community of the value of self-similar traffic models for performance analysis. However, queueing analysis with self-similar input processes represents a new area of research and is likely to require a new set of mathematical tools. Thus, while practically no analytic results are available at this time, some simulation results (using traces of actual Ethernet traffic; see [11]) and approximate analytic results (see [43]) do already exist and indicate that the performance of queueing models with self-similar input processes is drastically different from the performance predicted by traditional models. We refer to [30] and [32] for a discussion of some of the network-related implications of these observed differences.

### 1.3 Traffic Measurements

In this paper, we concentrate exclusively on the data analysis and modeling aspects of the Ethernet traffic measurements. To this end, we use very high quality, high time-resolution Ethernet LAN traffic data collected by Leland and Wilson [29]. 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 $\mu$s) 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 [29]. 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. These traffic measurements are of unusual quality. Because of their size, they require data analytic methods that go beyond the traditional approaches (see Section 3). We also illustrate how the existence of such data sets gives rise to new and challenging problems in statistical computing, for example, real-time parameter estimation (see Section 4).

The traffic measurements analyzed in this paper were 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. Table 1 gives a summary description of the traffic data. We consider 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 each of the four sets of traffic measurements, we identified what are considered "typical" low (L), medium (M) and high (H) activity hours and whether the data measure bytes (B) or packets (P). For example, AUG89.LB is a time series of length 360,000, each observation representing the number of bytes per 10 milliseconds. The sum of all these observations equals 224,315,439 bytes per hour. The time series AUG89.LP is of the same length but represents the number of packets per 10 milliseconds; a total of 652,909 packets were observed during this low-traffic hour.

With the resulting data sets, we are able to inves-

TABLE 1
*Qualitative description of the sets of Ethernet traffic measurements used in the analysis in Section 4*

| Traces of Ethernet Traffic Measurements | | | | | |
|---|---|---|---|---|---|
| Measurement Period | | data set | total number of bytes | total number of packets | Ethernet utilization |
| AUGUST 1989 | total (27.45 hours) | | 11,448,753,134 | 27,901,984 | 9.3% |
| Start of trace: Aug. 29, 11:25am | low hour (6:25am-7:25am) | AUG89.LB<br>AUG89.LP | 224,315,439 | 652,909 | 5.0% |
| End of trace: Aug. 30, 3:10pm | normal hour (2:25pm-3:25pm) | AUG89.MB<br>AUG89.MP | 380,889,404 | 968,631 | 8.5% |
| | busy hour (4:25pm-5:25pm) | AUG89.HB<br>AUG89.HP | 677,715,381 | 1,404,444 | 15.1% |
| OCTOBER 1989 | total (20.86 hours) | | 14,774,694,236 | 27,915,376 | 15.7% |
| Start of trace: Oct. 5, 11:00am | low hour (2:00am-3:00am) | OCT89.LB<br>OCT89.LP | 468,355,006 | 978,911 | 10.4% |
| End of trace: Oct. 6, 7:51am | normal hour (5:00pm-6:00pm) | OCT89.MB<br>OCT89.MP | 827,287,174 | 1,359,656 | 18.4% |
| | busy hour (11:00am-12:00am) | OCT89.HB<br>OCT89.HP | 1,382,483,551 | 2,141,245 | 30.7% |
| JANUARY 1990 | total (40.16 hours) | | 7,122,417,589 | 27,954,961 | 3.9% |
| Start of trace: Jan. 10, 6:07am | low hour (Jan. 11, 8:32pm-9:32pm) | JAN90.LB<br>JAN90.LP | 87,299,639 | 310,038 | 1.9% |
| End of trace: Jan. 11, 10:17pm | normal hour (Jan. 10, 9:32am-10:32am) | JAN90.MB<br>JAN90.MP | 182,636,845 | 643,451 | 4.1% |
| | busy hour (Jan. 11, 10:32am-11:32am) | JAN90.HB<br>JAN90.HP | 711,529,370 | 1,391,718 | 15.8% |
| FEBRUARY 1992 | total (47.91 hours) | | 6,585,355,731 | 27,674,814 | 3.1% |
| Start of trace: Feb 18, 5:22am | low hour (Feb. 20, 1:21am-2:21am) | FEB92.LB<br>FEB92.LP | 56,811,435 | 231,823 | 1.3% |
| End of trace: Feb. 20, 5:16am | normal hour (Feb. 18, 8:21pm-9:21pm) | FEB92.MB<br>FEB92.MP | 154,626,159 | 524,458 | 3.4% |
| | busy hour (Feb. 18, 11:21am-12:21am) | FEB92.HB<br>FEB92.HP | 225,066,741 | 947,662 | 5.0% |

tigate features of the observed traffic (e.g., self-similarity) that persist across the network as well as across time, irrespective of the utilization level of the Ethernet and of the network topology. Although only one LAN could be monitored at any one time (making it impossible to study correlations in the activity on different LAN's) and all data were collected from LAN's in the same company (making it not representative of all LAN traffic), we believe that some of the characteristics uncovered by our analysis of the data in Table 1 are likely to be present in non-Ethernet LAN traffic and in many high-speed networks of the future.

The traffic was mostly from services that use the Internet protocol (IP) suite for such capabilities as remote login or electronic mail, and the network file system (NFS) protocol for file service from servers to workstations. For example, the first two data sets were collected from a typical workgroup or laboratory network which was isolated from the rest of the Bellcore network by a router (see Figure 1). At the time of collection of the first data set, the laboratory consisted of about 140 people, most of whom had diskless Sun-3 class workstations on their desks. The network in the laboratory consisted of two cable segments (see Figure 1) separated by a bridge, implying that not all traffic within the laboratory could be seen by the monitor. The hosts on this network consisted of workstations, their file servers and a pair of minicomputers. Only a small number of hosts used reduced instruction set (RISC) processors. However, by the time the second data set was collected, an extensive upgrade of the Sun-3 class machines to RISC-based machines had taken place, as well as a small increase in the number of hosts (from about 120 to about 140). This upgrade explains the large difference in traffic volume in the first two data sets. For

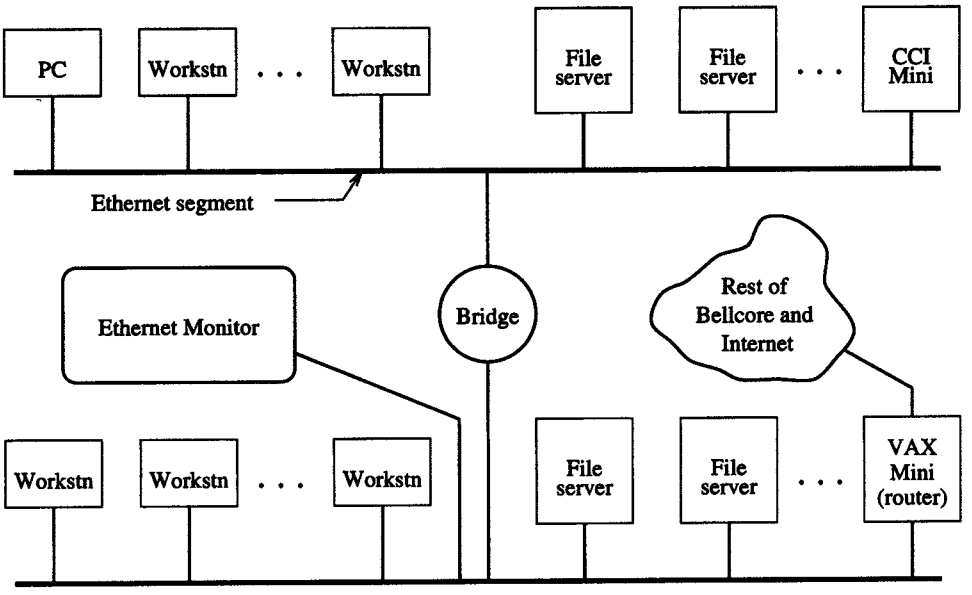


FIG. 1. *Network from which the August and October* 1989 *measurements were taken.*

a more detailed description of the four data sets in Table 1, see [32].

## 2. STOCHASTIC MODELING OF SELF-SIMILAR PHENOMENA

### 2.1 A Picture is Worth a Thousand Words

For 27 consecutive hours of monitored Ethernet traffic from the August 1989 measurements (first row in Table 1), Figure 2a–e depicts a sequence of simple plots of the packet counts (i.e., number of packets per time unit) for five different choices of time units. Starting with a time unit of 100 seconds (Figure 2a), 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 2e). 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 (minutes, hours) time scales as in the small (seconds, 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 scale-invariant or "self-similar" 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 2 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 (in terms of average packet size and arrival rate) compound Poisson process. (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 2a'–e'.) Figure 2 suggests the use of self-similar stochastic processes for traffic modeling. The presentation below of the concept of self-similar processes closely follows [7] and [4]; see also [46].

### 2.2 Definition of Self-Similar Processes

Let $X = (X_t: t = 0, 1, 2, \ldots)$ be a *covariance stationary* (sometimes called *wide-sense stationary*) stochastic process with mean $\mu$, variance $\sigma^2$ and autocorrelation function $r(k)$, $k = 0, 1, 2, \ldots$. In particular, we assume that $X$ has an autocorrelation function of the form

$$r(k) \sim k^{-\beta} L_1(k) \quad \text{as } k \to \infty, \tag{1}$$

where $0 < \beta < 1$ and $L_1$ is slowly varying at infinity, that is, $\lim_{t \to \infty} L_1(tx)/L_1(t) = 1$ for all $x > 0$

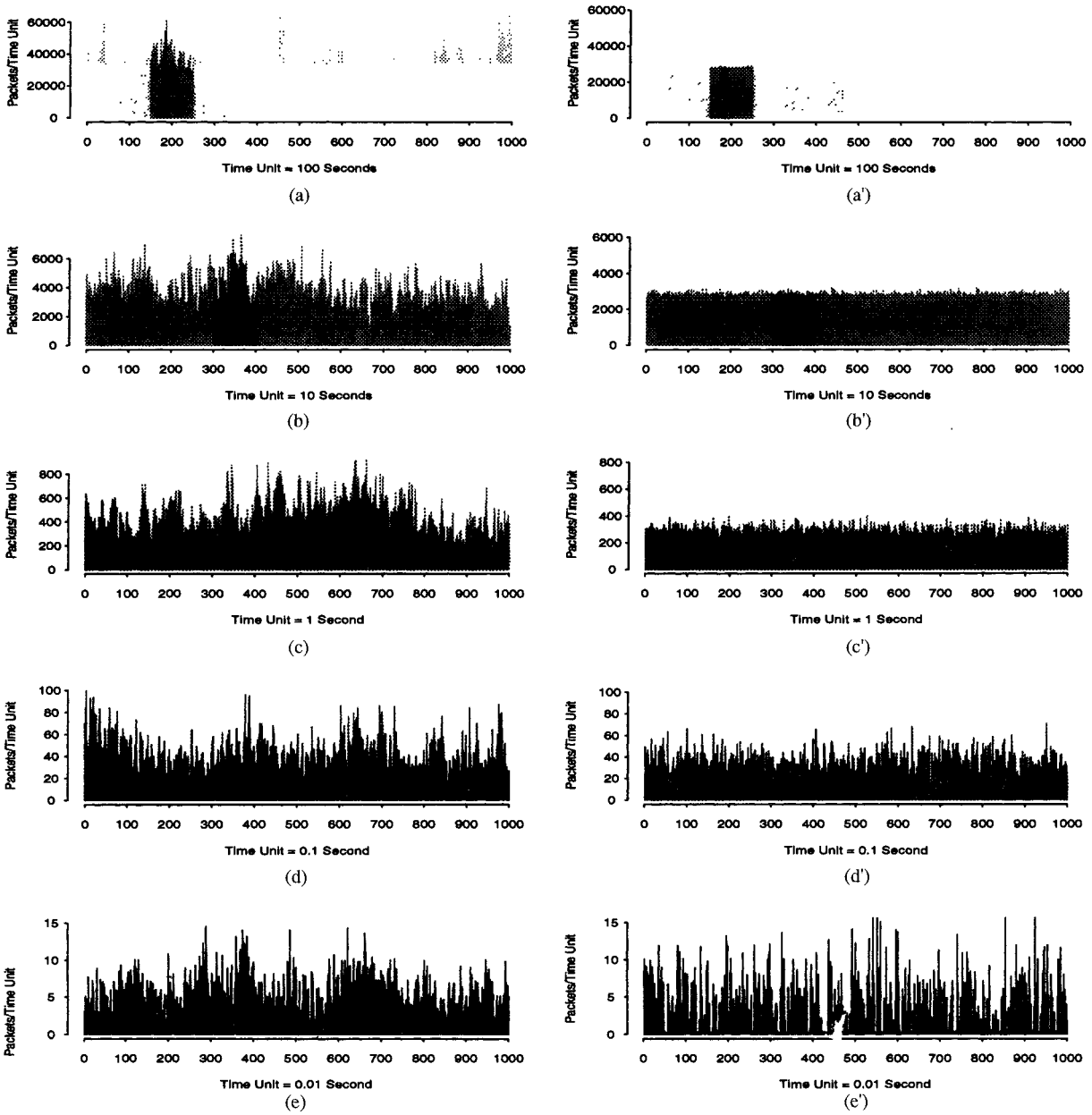


FIG. 2. *Indication of self-similarity: Ethernet traffic (packets per time unit) on five different time scales* (a)–(e); *for comparison, synthetic traffic from an appropriately chosen compound Poisson model on the same five different time scales* (a')–(e'). (*Different gray levels are used to identify the same segments of traffic on the different time scales.*)

[examples of such slowly varying functions are $L_1(t) = \text{const.}$ and $L_1(t) = \log(t)$]. For each $m = 1, 2, 3, \ldots$, let $X^{(m)} = (X_k^{(m)}: k = 1, 2, 3, \ldots)$ denote a new time series obtained by averaging the original series $X$ over nonoverlapping blocks of size $m$; that is, for each $m = 1, 2, 3, \ldots$, $X^{(m)}$ is given by

$$X_k^{(m)} = 1/m(X_{km-m+1} + \cdots + X_{km}), \quad k = 1, 2, 3, \ldots. \tag{2}$$

Note that for each $m$ the aggregated time series $X^{(m)}$ defines a covariance stationary process; let $r^{(m)}$ denote the corresponding autocorrelation function.

The process $X$ is called (*exactly*) *self-similar* with self-similarity parameter $H = 1 - \beta/2$, if, for all $m = 1, 2, 3, \ldots$, $1/m^H(X_{km-m+1} + \cdots + X_{km})$, $k = 1, 2, 3, \ldots$, has the same finite-dimensional distributions as $X$. It is (*exactly second-order*) *self-similar* with self-similarity parameter $H = 1 - \beta/2$ if for all $m = 1, 2, 3, \ldots, 1/m^H(X_{km-m+1} + \cdots + X_{km})$ has the same variance and autocorrelation as $X$. In terms of the aggregated processes $X^{(m)}$, this means

that, for all $m = 1, 2, 3, \ldots$, $\mathrm{var}(X^{(m)}) = \sigma^2 m^{-\beta}$ and

$$r^{(m)}(k) = r(k) = 1/2\,\delta^2(|k|^{2-\beta}), \quad k = 0, 1, 2, \ldots, \tag{3}$$

where $\delta^2(f)$ denotes the second central difference operator applied to a function $f$, that is, $\delta^2(f(k)) = f(k+1) - 2f(k) + f(k-1)$. An example of an exactly self-similar process with self-similarity parameter $H$ is *fractional Gaussian noise* (FGN) with $1/2 < H < 1$, that is, the increment process of *fractional Brownian motion* with parameter $H$, introduced by Mandelbrot and Van Ness [39].

The process $X$ is called (*asymptotically second-order*) *self-similar* with self-similarity parameter $H = 1 - \beta/2$ if

$$r^{(m)}(k) \to 1/2\,\delta^2(k^{2-\beta}) \quad \text{as} \quad m \to \infty,\ k = 0, 1, 2, \ldots. \tag{4}$$

Thus, an asymptotically self-similar process has the property that, for large $m$, the corresponding aggregated time series $X^{(m)}$ have a fixed correlation structure, solely determined by $\beta$; moreover, due to the asymptotic equivalence (for large $k$) of differencing and differentiating, $r^{(m)}$ agrees asymptotically with the correlation structure of $X$ given by (1). The *fractional autoregressive integrated moving-average processes* or *fractional* ARIMA$(p, d, q)$ with $0 < d < 1/2$ are examples of asymptotically second-order self-similar processes with self-similarity parameter $H = d + 1/2$. (For more details, see [17] and [21].)

Intuitively, the most striking feature of (exactly or asymptotically) self-similar processes is that their aggregated processes $X^{(m)}$ possess a nondegenerate correlation structure as $m \to \infty$. This behavior is in stark contrast to the more conventional stochastic models, all of which have the property that their aggregated processes $X^{(m)}$ tend to second-order pure noise (as $m \to \infty$), that is

$$r^{(m)}(k) \to 0, \quad \text{as } m \to \infty,\ k = 1, 2, 3, \ldots. \tag{5}$$

Note that we have chosen the above definitions of self-similarity over the mathematically more convenient definition of a *self-similar* continuous-time stochastic process $X = (X_t: t \geq 0)$ with mean zero and stationary increments, namely, for all $a > 0$,

$$X_{at} = a^H X_t, \tag{6}$$

where equality is understood in the sense of equality of the finite-dimensional distributions, and the exponent $H$ is the self-similarity parameter. Definitions (4) and (5) have the advantage that they do not obscure the connection with standard time series theory, and they reflect the fact that we are mainly interested in large $m$'s (time "scales"); here we are less concerned about deviations from self-similarity for $m \to 0$. From a modeling perspective, the crucial point is that both the discrete-time and the continuous-time definitions involve a wide range of time scales. One advantage of definition (6) in the presence of large data sets is that it allows for a quick heuristic method for estimating the self-similarity parameter $H$ from simple plots like the ones in Figure 2; if the original time series $X$ represents the number of Ethernet packets per 10 milliseconds [plot (e)], then plots (a)–(d) depict segments of the aggregated time series representing the number of packets per 100 seconds, 10 seconds, 1 second and 0.1 second, respectively. All of the plots (a)–(e) in Figure 2 look "similar," suggesting a more or less identical nondegenerate autocorrelation function for all of the aggregated processes. In fact, a naive inference from the successive plots (a)–(e) in Figure 2 (subtracting the sample mean of $X$ and using simple statistics such as range and histogram) yields $H$-values of about 0.8 for relation (6). In contrast, plots (a′)–(e′) in Figure 2 show the pure white noise behavior of the aggregated processes generated from the synthetic Poisson batch traffic model: identical but degenerate autocorrelation structures for the $X^{(m)}$'s (for $m > 100$).

### 2.3 Properties of Self-Similar Processes

2.3.1 *Long-range dependence and the Hurst effect.* A stochastic process satisfying relation (1) is said to exhibit *long-range dependence* (see, e.g., [3], [7], [27] or [49]). In Mandelbrot's terminology, long-range dependence is also called the *Joseph effect*, referring to the "seven fat years and seven lean years" in the Biblical story of Joseph. Thus, processes with long-range dependence are characterized by an autocorrelation function that decays hyperbolically as the lag increases. Moreover, it is easy to see that (1) implies $\Sigma_k r(k) = \infty$. This nonsummability of the correlations captures the intuition behind long-range dependence, namely, that while high-lag correlations are all individually small, their cumulative effect is important and gives rise to features which are drastically different from those of the more conventional (i.e., short-range) dependent processes. The latter are characterized by an exponential decay of the autocorrelations, that is, $r(k) \sim \rho^k$, as $k \to \infty$, $0 < \rho < 1$, resulting in a summable autocorrelation function $0 < \Sigma_k r(k) < \infty$. Also note that the nonsummability of the correlations is needed in order to guarantee a nondegenerate correlation structure of the aggregated processes $X^{(m)}$ as $m \to \infty$.

When working in the frequency domain, long-range dependence manifests itself in a spectral density that obeys a power law near the origin. In fact,

equivalently to (1) (under weak regularity conditions on the slowly varying function $L_1$), there is long-range dependence in $X$ if

$$f(\lambda) \sim \lambda^{-\gamma} L_2(\lambda) \quad \text{as } \lambda \to 0, \tag{7}$$

where $0 < \gamma < 1$, $L_2$ is slowly varying at 0 and $f(\lambda) = \sum_k r(k) e^{ik\lambda}$ denotes the spectral density function. Thus, from the point of view of spectral analysis, long-range dependence implies that $f(0) = \sum_k r(k) = \infty$, that is, it requires a spectral density which tends to $+\infty$ as the frequency $\lambda$ approaches 0 ("$1/f$-noise"). On the other hand, short-range dependence is characterized by a spectral density function $f(\lambda)$ which is positive and finite for $\lambda = 0$.

From our earlier discussion, it follows that both fractional Gaussian noise processes (with $1/2 < H < 1$) and fractional ARIMA$(p, d, q)$ processes (with $0 < d < 1/2$) exhibit long-range dependence. The parameters $H$ and $d$, respectively, measure the degree of long-range dependence and can be estimated from empirical records (see Section 3). Heuristically, long-range dependence manifests itself in the presence of cycles of all frequencies and orders of magnitude, displays features suggestive of nonstationarity and has been found to be relevant in economics, in hydrology and geology and in telecommunication (for references, see [3], [7], [18] and [48]).

Historically, the importance of self-similar processes as defined in Section 2.2 lies in the fact that they provide an elegant explanation and interpretation of an empirical law that is commonly referred to as *Hurst's law* or the *Hurst effect*. Briefly, for a given set of observations $(X_k: k = 1, 2, \ldots, n)$ with sample mean $\overline{X}(n)$ and sample variance $S^2(n)$, the *rescaled adjusted range* or the *R/S-statistic* is given by

$$\begin{aligned} &R(n)/S(n) \\ &= 1/S(n)[\max(0, W_1, W_2, \ldots, W_n) \\ &\qquad - \min(0, W_1, W_2, \ldots, W_n)], \end{aligned} \tag{8}$$

with $W_k = (X_1 + X_2 + \cdots + X_k) - k\overline{X}(n)$, $k = 1, 2, \ldots, n$. Hurst [22, 23] found that many naturally occurring time series appear to be well represented by the relation

$$E[R(n)/S(n)] \sim cn^H \quad \text{as } n \to \infty, \tag{9}$$

with *Hurst parameter* $H > 0.5$, and $c$ a finite positive constant that does not depend on $n$. On the other hand, if the observations $X_k$ come from a short-range dependent model, then Mandelbrot and Van Ness [39] showed that

$$E[R(n)/S(n)] \sim dn^{0.5} \quad \text{as } n \to \infty, \tag{10}$$

where $d$ is a finite positive constant, independent of $n$. The discrepancy between (9) and (10) is generally referred to as the *Hurst effect* or *Hurst phenomenon*.

2.3.2 *Slowly decaying variances.* From a statistical point of view, the most salient feature of self-similar processes as defined in Section 2.2 is that the variance of the arithmetic mean decreases more slowly than the reciprocal of the sample size; that is, it behaves like $n^{-\beta}$ for some $\beta \in (0, 1)$, instead of like $n^{-1}$ for the processes whose aggregated series converge to second-order pure noise. For our discussion below, we assume for simplicity that the slowly varying functions $L_1$ and $L_2$ in (1) and (7), respectively, are asymptotically constant. Cox [7] showed, in fact, that a specification of the autocorrelation function satisfying (1) [or, equivalently, of the spectral density function satisfying (7)] is the same as a specification of the sequence $(\text{var}(X^{(m)}): m \geq 1)$ with the property

$$\text{var}(X^{(m)}) \sim am^{-\beta} \quad \text{as } m \to \infty, \tag{11}$$

where $a$ is a finite positive constant independent of $m$, and $0 < \beta < 1$; in fact, the parameter $\beta$ is the same as in (1) and is related to the parameter $\gamma$ in (7) by $\beta = 1 - \gamma$. On the other hand, for covariance stationary processes whose aggregated series $X^{(m)}$ tend to second-order pure noise [i.e., (5) holds], it is easy to see that the sequence $(\text{var}(X^{(m)}): m \geq 1)$ satisfies

$$\text{var}(X^{(m)}) \sim bm^{-1} \quad \text{as } m \to \infty, \tag{12}$$

where $b$ is a finite positive constant independent of $m$.

The consequences of the slowly decaying variances $\text{var}(X^{(m)})$ for classical statistical tests and confidence or prediction intervals can be disastrous (e.g., see [3], [19] and [28]), since the usual standard errors (derived for conventional models) are wrong by a factor that tends to infinity as the sample size increases.

2.3.3 *Parsimonious modeling.* Since we are always dealing with finite data sets, it is in principle not possible to decide whether the asymptotic relationships (1), (3), (4) and so on hold. For processes that are not self-similar in the sense that their aggregated series converge to second-order pure noise [property (5)], the correlations will eventually decrease exponentially, continuity of the spectral density function at the origin will eventually show up, the variances of the aggregated processes will eventually decrease as $m^{-1}$ and the rescaled adjusted range will eventually increase as $n^{0.5}$. For finite sample sizes, distinguishing between these asymptotics and the ones corresponding to self-similar processes is, in general, problematic.

In the present context of Ethernet measurements, we typically deal with time series with hundreds of thousands of observations and are, therefore, able to employ statistical and data analytic techniques which are impractical for small data sets. Moreover, with such sample sizes, parsimonious modeling becomes a necessity due to the large number of parameters needed when trying to fit a conventional process to a "truly" self-similar model. Modeling, for example, long-range dependence with the help of ARMA processes is equivalent to approximating a hyperbolically decaying autocorrelation function by a sum of exponentials. Although always possible, the number of parameters needed will tend to infinity as the sample size increases, and giving physically meaningful interpretations for the parameters becomes more and more difficult. In contrast, the long-range dependence component of the process can be modeled by a self-similar process with only one parameter. Finally, from a modeling perspective, it would be very unsatisfactory to use for a single empirical time series two different models, one for a short sequence, another for a long sequence.

## 3. INFERENCE FOR SELF-SIMILAR PROCESSES

### 3.1 Statistical Methods for Testing for Self-Similarity

From a theoretical point of view, slowly decaying variances, long-range dependence and a spectral density of the form (7) are different manifestations of one and the same property of the underlying covariance stationary process $X$, namely, that $X$ is (asymptotically or exactly second-order) self-similar. Subsequently, we can approach the problem of testing for and estimating the degree of self-similarity from three different angles: (i) analysis of the variances of the aggregated processes $X^{(m)}$; (ii) time-domain analysis based on the $R/S$-statistic; (iii) periodogram-based analysis in the frequency domain. This subsection provides a brief description of the corresponding statistical and graphical methods (for more details, see [4] and the references therein); their use in analyzing the Ethernet data will be illustrated in Section 3.2. For a similar analysis that uses different data sets from Table 1, see [31].

3.1.1 *Variance–time plots.* We have seen in Section 2.3 that for self-similar processes the variances of the aggregated processes $X^{(m)}$, $m = 1, 2, 3, \ldots$, decrease linearly (for large $m$) in log–log plots against $m$ with slopes arbitrarily flatter than $-1$ [see (11)]. On the other hand, none of the short-range dependent processes commonly considered in the teletraffic literature yield a power law for the variances of the form (11); this behavior can be approximated for some transient period of time by short-range dependent models with a large number of parameters, but the variance of $X^{(m)}$ will eventually decrease linearly in log–log plots against $m$ with a slope equal to $-1$ [see (12)]. The so-called *variance–time plots* are obtained by plotting $\log(\mathrm{var}(X^{(m)}))$ against $\log(m)$ ("time") and by fitting a simple least squares line through the resulting points in the plane, ignoring the small values for $m$. Values of the estimate $\hat{\beta}$ of the asymptotic slope between $-1$ and 0 suggest self-similarity, and an estimate for the degree of self-similarity is given by $\hat{H} = 1 - \hat{\beta}/2$.

Clearly, variance–time plots are not reliable for empirical records with small sample sizes. However, as we will demonstrate below, with sample sizes of the magnitude of the Ethernet traffic data sets, "eyeball tests" such as the variance–time plots become highly useful and give a rather accurate picture about the self-similar nature of the underlying time series and about the degree of self-similarity.

3.1.2 *$R/S$ analysis.* The objective of the $R/S$ analysis of an empirical record is to infer the degree of self-similarity $H$ (Hurst parameter) in relation (9) for the self-similar process that presumably generated the record under consideration. In practice, $R/S$ analysis is based on a heuristic graphical approach (originally described in detail in [40]) that tries to exploit as fully as possible the information in a given record. The following graphical method has been used extensively in the past. Given a sample of $N$ observations $(X_k: k = 1, 2, 3, \ldots, N)$, one subdivides the whole sample into $K$ nonoverlapping blocks and computes the rescaled adjusted range $R(t_i, n)/S(t_i, n)$ for each of the new "starting points" $t_1 = 1$, $t_2 = N/K + 1$, $t_3 = 2N/K + 1, \ldots$ which satisfy $(t_i - 1) + n \leq N$. Here, the $R/S$ statistic $R(t_i, n)/S(t_i, n)$ is defined as in (8) with $W_k$ replaced by $W_{t_i+k} - W_{t_i}$, and $S^2(t_i, n)$ is the sample variance of $X_{t_i+1}, X_{t_i+2}, \ldots, X_{t_i+n}$. Thus, for a given value ("lag") of $n$, one obtains many samples of $R/S$, as many as $K$ for small $n$ and as few as 1 when $n$ is close to the total sample size $N$. Next, one takes logarithmically spaced values of $n$, starting with $n \approx 10$. Plotting $\log(R(t_i, n)/S(t_i, n))$ versus $\log(n)$ results in the *rescaled adjusted range plot* (also called the *pox diagram of $R/S$*). When the parameter $H$ in relation (9) is well defined, a typical rescaled adjusted range plot starts with a transient zone representing the nature of short-range dependence in the sample, but eventually settles down and fluctuates in a straight "street" of a certain slope. Graphical $R/S$ analysis is used to

determine whether such asymptotic behavior appears supported by the data. In the affirmative, an estimate $\hat{H}$ of the self-similarity parameter $H$ is given by the street's asymptotic slope (typically obtained by a simple least squares fit), which can take any value between 1/2 and 1.

With respect to the effectiveness of $R/S$ analysis as a function of the sample size, similar comments as in Section 3.1.1 apply. For practical purposes, the most useful and attractive feature of the $R/S$ analysis is its relative robustness against changes of the marginal distribution. This feature allows for practically separate investigations of the self-similarity property of a given empirical record and of its distributional characteristics.

3.1.3 *Periodogram-based analysis with aggregation.* While variance–time plots and pox plots of $R/S$ are very useful tools for identifying self-similarity (in a mostly heuristic manner), the absence of any results for the limit laws of the corresponding statistics make them inadequate when a more refined data analysis is required (e.g., confidence intervals for the degree of self-similarity $H$, model selection criteria and goodness-of-fit tests). In contrast, a more refined data analysis is possible for maximum likelihood type estimates (MLE) and related methods based on the periodogram. In particular, for Gaussian processes $X = (X_k\colon\ k = 0, 1, 2, \ldots)$, Whittle's approximate MLE has been studied extensively (see [52], [2], [12] and [9]) and is defined as follows. Let $f(x;\theta) = \sigma_\varepsilon^2 f(x;(1,\eta))$ be the spectral density of $X$ with $\theta = (\sigma_\varepsilon^2, \eta) = (\sigma_\varepsilon^2, H, \theta_3, \ldots, \theta_k)$, where $H = (\gamma + 1)/2$ [with $\gamma$ as in (7)] describes the degree of self-similarity and $\theta_3, \ldots, \theta_k$ model the short-range dependence structure of the process. As the scale parameter, we use the variance $\sigma_\varepsilon^2$ of the innovation $\varepsilon$ in the infinite AR-representation of the process, that is, $X_j = \sum_{i\geq 1}\alpha_i X_{j-i} + \varepsilon_j$, with $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_j)$. Note that this implies

$$\int \log(f(x;(1,\eta)))\,dx = 0. \tag{13}$$

The Whittle estimator $\hat{\eta}$ of $\eta$ minimizes

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{I(x)}{f(x;(1,\eta))}\,dx, \tag{14}$$

where $I(\cdot)$ denotes the *periodogram* of $X$ defined by

$$I(x) = (2\pi n)^{-1}\left|\sum_{j=1}^{n} X_j e^{ijx}\right|^2, \tag{15}$$

and the estimate of $\sigma_\varepsilon^2$ is given by

$$\hat{\sigma}_\varepsilon^2 = \int_{-\pi}^{\pi} \frac{I(x)}{f(x;(1,\hat{\eta}))}\,dx. \tag{16}$$

Then $n^{1/2}(\hat{\theta} - \theta)$ is asymptotically normally distributed if $(X_j)_{j\geq 1}$ can be written as an infinite moving average process. For Gaussian processes, $\hat{\theta}$ has the same asymptotic distribution as the MLE and is asymptotically efficient. For other periodogram-based methods, see, for example, [14], [15] and [45].

In the context of periodogram-based methods, problems of robustness due to (i) deviations from Gaussianity and (ii) deviations from the assumed model spectrum are commonly encountered. Transforming the data so as to obtain approximately the desired marginal (normal) distribution is generally considered a viable method to overcome (i). For problem (ii), there are several proposals in the literature, including estimating $H$ from the periodogram ordinates at low frequencies only, or bounding the influence of $I(x)$ at high frequencies. In the presence of large data sets, an alternative and more direct method for tackling (ii) uses the method of aggregation (see Section 2.2): if $(X_i)_{i\geq 1}$ is a Gaussian process satisfying (7), then

$$m^{-H}L^{-1/2}(m)\sum_{i=(j-1)m+1}^{mk}(X_i - E[X_i]), \qquad j = 1, 2, 3, \ldots, [n/m], \tag{17}$$

converges (in distribution) to fractional Gaussian noise as $m \to \infty$ [$L(\cdot)$ is a slowly varying function at infinity]. The same holds true if $X_i = \mu + G(Y_i)$, where $(Y_i)_{i\geq 1}$ is a Gaussian process satisfying (7), $E[G(Y_i)] = 0$, $E[G^2(Y_i)] < \infty$ and $E[G(Y_i)Y_i] \neq 0$. Hence, for sufficiently large $m$, fractional Gaussian noise is a good model for the aggregated $X^{(m)}$ so that we can apply a maximum likelihood type estimator for fractional Gaussian noise.

Combined, Whittle's approximate MLE approach and the aggregation method give rise to the following operational procedure for obtaining confidence intervals for the self-similarity parameter $H$. For a given time series, consider the corresponding aggregated processes $X^{(m)}$ with $m = 100, 200, 300, \ldots$, where the largest $m$-value is chosen such that the sample size of the corresponding series $X^{(m)}$ is not less than about 100. For each of the aggregated series, estimate the self-similarity parameter $H^{(m)}$ via a discretized version of (14) (replace the integral by a Riemann sum), where $f(x;(1,H^{(m)}))$ denotes the spectral density of fractional Gaussian noise. This procedure results in estimates $\hat{H}^{(m)}$ of $H^{(m)}$ and corresponding, say, 95%-confidence intervals of

the form $\hat{H}^{(m)} \pm 1.96\hat{\sigma}_{H^{(m)}}$, where $\hat{\sigma}^2_{H^{(m)}}$ is given by the central limit theorem for $\hat{\theta}$ mentioned earlier. Finally, we plot the estimates $\hat{H}^{(m)}$ of $H^{(m)}$ together with their 95%-confidence intervals versus $m$. Such plots will typically vary a lot for small aggregation levels, but will stabilize after a while and fluctuate around a constant value, our final estimate of the self-similarity parameter $H$. Once stabilization is observed, we choose for our confidence interval the one with the smallest value for $m$, because the size of the confidence intervals increases in $m$ (the more we aggregate, the fewer observations).

### 3.2 The Self-Similar Nature of Ethernet Traffic

3.2.1 *Ethernet traffic over a one-day period.* We first consider the August 1989 snapshot of Ethernet traffic (row 1 in Table 1) and analyze the three subsets AUG89.LB, AUG89.MB and AUG89.HB. Each sequence contains 360,000 observations, and each observation represents the number of bytes sent over the Ethernet every 10 milliseconds. Figure 3 depicts (a) the pox plot of $R/S$, (b) the variance–time curve and (c) the periodogram plot, corresponding to the sequence AUG89.MB. The pox plot of $R/S$ (Figure 3a) shows an asymptotic slope that is distinctly different from 0.5 (lower dotted line) and 1.0 (upper dotted line) and is easily estimated (using the "brushed" points) to be about 0.79. The variance–time curve (Figure 3b), which has been normalized by the sample variance of the whole sequence, shows an asymptotic slope that is clearly different from $-1$ (dotted line) and is easily estimated to be about $-0.40$, resulting in a practically identical estimate of the Hurst parameter $H$ of about 0.80. Finally, looking at the periodogram plot corresponding to the time series AUG89.MB, we observe that although there are some pro-

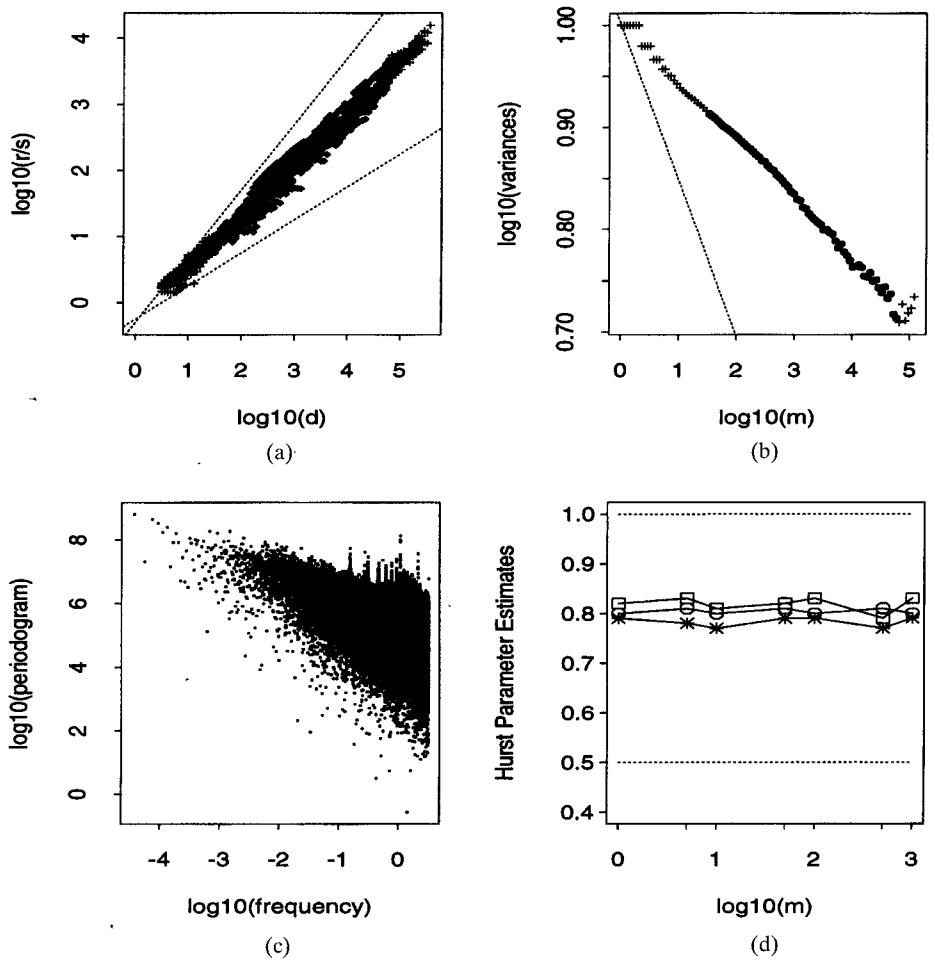


FIG. 3. (a) *Pox plot of R/S*; (b) *variance–time plot*; (c) *periodogram plot of sequence AUG89.MB. The asymptotic slopes are readily estimated (using the "brushed" points) to be about* $-0.79$ *in* (a) *and* $-0.40$ *in* (b), *and (using the lowest* 10% *of all frequencies) about* $-0.64$ *in* (c). *Plot* (d) *gives the estimates of the Hurst parameter H for the sequence AUG89.MB as a function of the aggregation level m* (○ *variance–time plot*, ∗ *pox plot of R/S estimate*, □ *periodogram plot estimate*).

nounced peaks in the high-frequency domain of the periodogram, the low-frequency part is characteristic for a power-law behavior of the spectral density around zero. In fact, by fitting a simple least squares line using only the lowest 10% of all frequencies, we obtain a slope estimate $\hat{\gamma} \approx 0.64$ which results in a Hurst parameter estimate of about 0.82. Thus, together the three graphical methods suggest that the sequence AUG89.MB is self-similar with self-similarity parameter $H \approx 0.80$. Moreover, Figure 3d indicates that the normal hour Ethernet traffic of the August 1989 data is, for practical purposes, (exactly) self-similar: it shows the estimates of the Hurst parameter $H$ for selected aggregated time series derived from the sequence AUG89.MB, as a function of the aggregation level $m$. For aggregation levels $m = 1, 5, 10, 50, 100, 500, 1000$, we plot the Hurst parameter estimate $\hat{H}^{(m)}$ [based on the pox plots of $R/S$ ($*$), the variance-time curves ($\bigcirc$), and the periodogram plots ($\square$)] for the aggregated time series $X^{(m)}$ against the logarithm of the aggregation level $m$. Notice that the estimates are extremely stable and practically constant over the depicted range of aggregation levels $1 \leq m \leq 1000$. Because the range includes small values of $m$, the sequence AUG89.MB can be regarded as (exactly) self-similar. Thus, in terms of their second-order statistical properties, the aggregated series $X^{(m)}$, $m \geq 1$, can be considered to be identical and produce, therefore, realizations that have similar overall structure and look very much alike. This observation agrees with the visual assessment of plots (a)–(e) in Figure 2 made earlier. Similar results are obtained for the sequences AUG89.LB and AUG89.HB, and for the corresponding packet-count processes AUG89.LP, AUG89.MP and AUG89.HP. Together, these observations show that Ethernet traffic over approximately a 24-hour period is self-similar, with the degree of self-similarity depending on the utilization level of the Ethernet (increasing as the utilization increases).

3.2.2 *Ethernet traffic over a four-year period.* Figure 4 shows a sample result of the MLE-based estimation method mentioned in Section 3.1.3 when combined with the method of aggregation. For each of the four sets of traffic measurements described in Table 1, we use the time series representing the packet counts during normal traffic conditions (i.e.,AUG89.MP, OCT89.MP, JAN90.MP and FEB92.MP)and consider the corresponding aggregated time series $X^{(m)}$ with $m = 100$, 200,

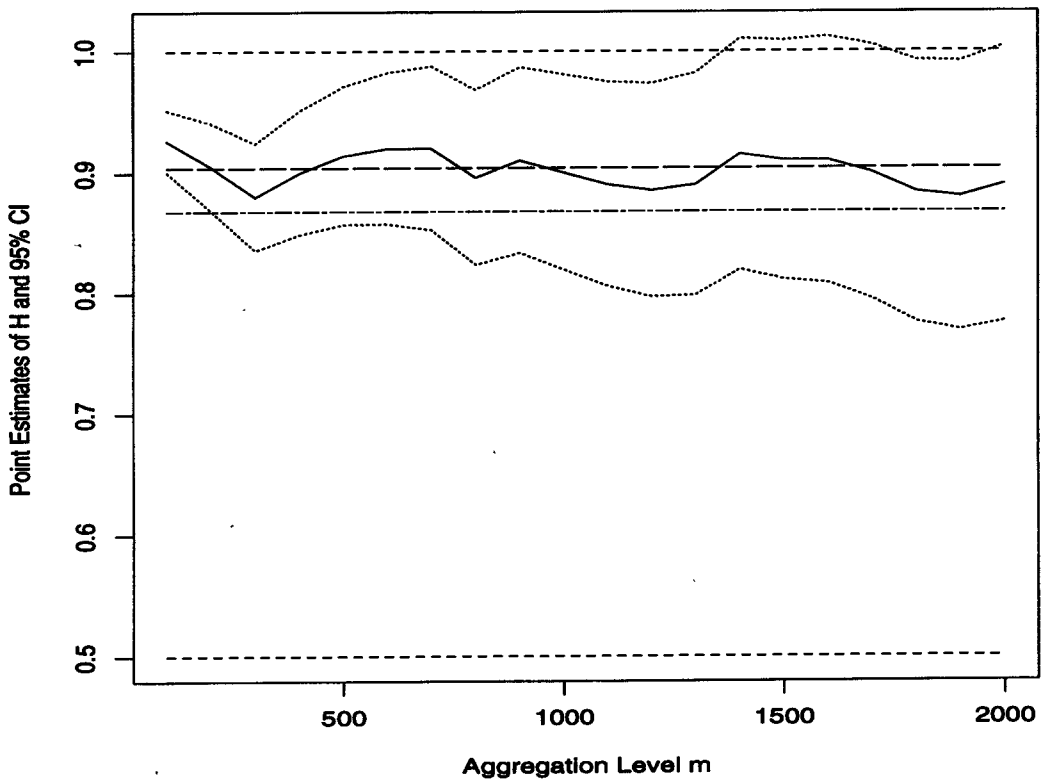


FIG. 4. *Periodogram-based MLE estimate $\hat{H}^{(m)}$ of $H$ (—) and 95%-confidence intervals ($\cdots$), as a function of the aggregation level $m$, for sequence AUG89.MP; $m \approx 300$ is an appropriate aggregation level for sequence AUG89.MP, yielding a point estimate $\hat{H} = \hat{H}^{(300)} = 0.90$ and a 95%-confidence interval [0.85, 0.95]. For comparison, we also added to the plot the estimates of $H$ based on the variance-time plot (-·-·-) and the $R/S$-based estimate of $H$ (- - -).*

300, ..., 1900, 2000 (representing the packet counts per 1, 2, ..., 20 seconds, respectively). We plot the Hurst parameter estimates $\hat{H}^{(m)}$ of $H^{(m)}$ obtained from the aggregated series $X^{(m)}$, together with their 95%-confidence intervals, against the aggregation level $m$. The resulting plot for the time series AUG89.MP is given in Figure 4 (the plots for the other time series are very similar and are not given here) and shows that the values of $\hat{H}^{(m)}$ are quite stable and fluctuate only slightly in the 0.85 to 0.95 range throughout the aggregation levels considered. The same holds for the 95%-confidence interval bands, indicating strong statistical evidence for self-similarity of the time series AUG89.MP with a degree of self-similarity of about 0.90. The relatively stable behavior of the Hurst parameter estimates $\hat{H}^{(m)}$ for the different aggregation levels $m$ also confirms our earlier finding that Ethernet traffic during normal traffic hours can be considered to be exactly self-similar rather than asymptotically self-similar. For exactly self-similar time series, determining a single point estimate for $H$ and the corresponding 95%-confidence interval is straightforward and can be done by visual inspection of plots such as the one in Figure 4 (see below). Notice that in Figure 4, we added two lines corresponding to the Hurst parameter estimates obtained from the pox diagrams of $R/S$ and the variance–time plots, respectively. These lines fall well within the 95%-confidence interval bands, which shows that for these long time series considered here, graphical estimation methods based on $R/S$ or variance–time plots can be expected to be very accurate.

In addition to the four normal hour packet data time series, we also applied the combined MLE–aggregation method to the other traffic data sets described in Table 1. Figure 5 depicts all Hurst parameter estimates (together with the 95%-confidence interval) for each of the 12 *packet* data time series. (A similar plot, not shown here, was also obtained for the 12 time series representing the number of bytes per 10 milliseconds.) We also included in this summary plot the Hurst parameter estimates obtained through the $R/S$ analysis (*) and variance–time plots (○) in order to indicate the accuracy of these "graphical" estimators when compared to the statistically more rigorous Whittle estimator (●). Figure 5 shows convincingly the self-similar nature of Ethernet traffic, irrespective of when and where the data were collected during the four years of measurements: the value of $H$

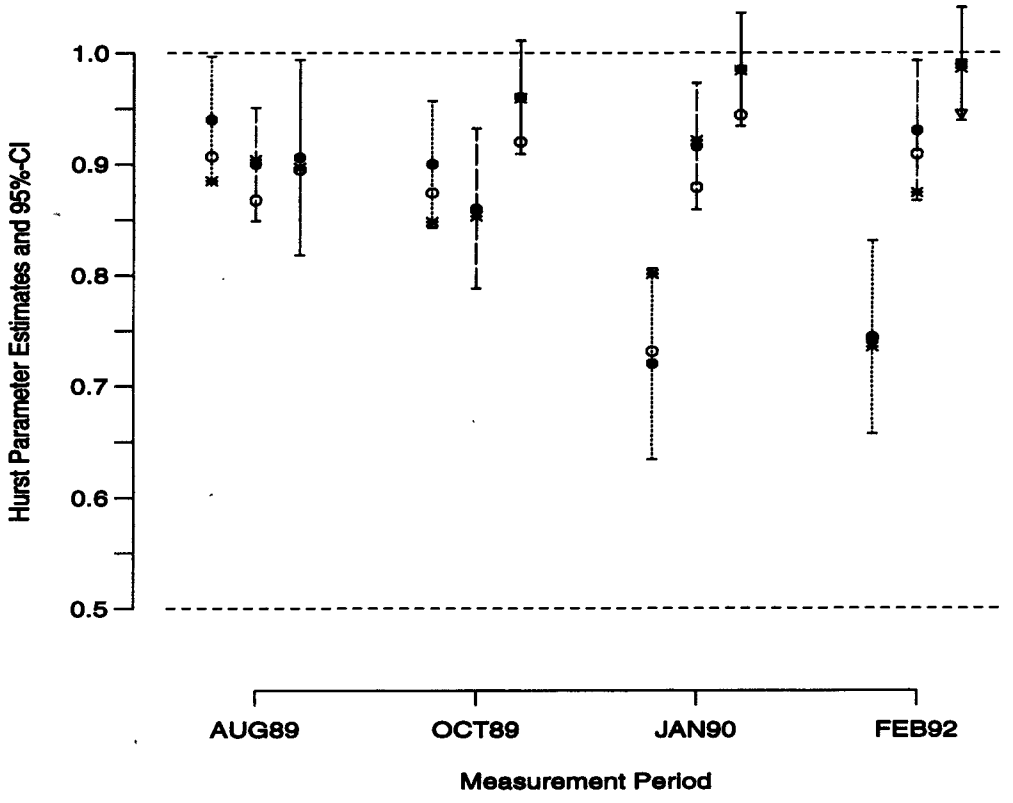


FIG. 5. *Summary plot of estimates of the Hurst parameter H for all low-, medium- and high-traffic packet data sets of the measurement periods given in Table 1 (● periodogram-based MLE estimate and the corresponding 95%-confidence interval; ○ point estimate of H based on the asymptotic slope of the variance-time plot; * point estimate of H based on the asymptotic slope of the pox plot of R/S).*

may change, but none of the 95%-confidence intervals comes even close to covering the value $H = 0.50$, the value of the Hurst parameter that characterizes stochastic models for packet traffic currently considered in the literature. These findings not only apply to *internal* traffic consisting of all packets on a LAN, but also to *remote* or *external* Ethernet traffic (all IP packets with a source or destination address that is not on any of the Bellcore networks), and *external TCP* traffic, the portion of external traffic using the transmission control protocol (TCP) and IP.

Note that all the results in this section have been obtained by treating the Ethernet packets essentially as black boxes, that is, we only counted the number of packets and/or bytes per time unit but did not look into the packet header fields or distinguished packets based on their source or destination addresses or based on particular applications. As a result, drawing conclusions about which network design issues and/or network usage features are likely to result in the different observed values of the Hurst parameter for the different data sets displayed in Table 1 is somewhat speculative at this stage; further work that requires looking into the header field of every Ethernet packet is currently in progress (see also Section 5) and is likely to shed more light on this question. However, some intuitive arguments for why the $H$-value varies from data set to data set can already be given at this stage. For example, a quick look at Figure 5 reveals an obvious difference between the low-traffic hours of the pre-1990 data (i.e., August 1989 and October 1989) and the later measurements (January 1990 and February 1992). In order to explain this difference, it is important to know that while the pre-1990 measurements consist of mostly host-to-host workgroup traffic, the later measurements are predominantly from router-to-router traffic. Although this change does not appear to alter drastically the self-similarity parameter of the data sets corresponding to the typical normal or busy hours, it clearly affects the $H$-values of the low-traffic hours. Intuitively, low period router-to-router traffic consists mostly of machine-generated packets which tend to form a smoother arrival stream than low period host-to-host traffic which is typically produced by a smaller than average number of actual Ethernet users, for example, researchers working late hours.

## 4. SELF-SIMILAR TRAFFIC MODELING AND HIGH-PERFORMANCE COMPUTING

As we have seen in Section 3, exactly self-similar models such as fractional Gaussian noise (or some nonlinear transformation of fractional Gaussian noise) or asymptotically self-similar models such as fractional ARIMA processes can be used to fit hour-long traces of Ethernet traffic very well. Their successful application to packet traffic modeling relies on (i) having readily available statistical methods for "real-time" parameter estimation for these processes, (ii) being able to generate quickly long traces of synthetic observations from these processes and (iii) demonstrating network-related implications of self-similarity which, when not accounted for, lead to mediocre or unacceptable network performance. While all three items open up new areas of research in statistics, statistical computing and applied probability–queueing theory, respectively, in this section we specifically address the problem of real-time parameter estimation for very large sets of self-similar data and the question of efficiently generating long traces of synthetic data from self-similar models.

### 4.1 Parameter Estimation and Distributed Computing

Fractional Gaussian noise is characterized by its mean $\mu$, variance $\sigma^2$ and Hurst parameter $H$. Each of these three parameters has an obvious physical interpretation. When there are indications of a particular short-range dependence structure in the traffic measurements, asymptotically self-similar models such as the fractional ARIMA($p$, $d$, $q$) can be more appropriate; that is, by adding a few parameters (typically one or two), it is not only possible to fit the low-frequency components in the data but also to capture the contributions of the high-frequency components. Parameter estimation techniques are known in both cases but the statistically rigorous methods often turn out to be computationally too intensive in order to be feasible for large data sets.

However, recent work by Beran and Terrin [5] shows how existing parameter estimation techniques for self-similar data can be adapted and result in fast ("real-time") parameter estimation methods even for very long time series. For example, instead of estimating the Hurst parameter $H$ from the whole series $(X_1, \ldots, X_n)$, divide the series into $k$ subseries $(X_1, \ldots, X_i)$, $(X_{i+1}, \ldots, X_{2i}), \ldots, (X_{(k-1)i+1}, \ldots, X_{ki})$, $i > 0$, $k = [n/i]$. Estimate each $\hat{H}_j$, $j = 1, \ldots, k$, using existing techniques (e.g., via the Whittle estimate) and define the "grand" estimate $\hat{H}$ of $H$ to be equal to $\hat{H} = 1/k \sum \hat{H}_j$. Beran and Terrin show that, as $n \to \infty$ such that $i/n \to \gamma > 0$, $n^{1/2}(\hat{H} - H)$ has the same asymptotic distribution as the Whittle estimator based on the whole series. Moreover, they show that, for Gaussian processes, $\hat{H}$ is asymptotically efficient. Obviously, the procedure suggested

by Beran and Terrin allows for real-time parameter estimation for very long time series, given an appropriate computing environment consisting of a high-speed communication network of high-performance workstations and mass storage disk arrays. This result also holds if there are additional parameters besides $H$ and if the estimation method is combined with the aggregation technique of Section 3.2. We are currently in the process of implementing and experimenting with the Beran–Terrin method in an experimental high-performance computing environment that is available at Bellcore.

Apart from making parameter estimation for very large self-similar data sets computationally feasible, the procedure proposed by Beran and Terrin also provides a method for checking whether $H$ (and possibly additional parameters) remains constant over the whole time series. The problem of deciding whether inhomogeneities in $H$ over time are real (due to actual changes in the dependence structure in the data) or are due to randomness is very delicate because even optimal estimates of $H$ turn out to vary considerably when calculated for disjoint parts of a long-range dependent time series. In order to assess quantitatively how much the estimates of $H$ can vary when estimated from different portions of the data, Beran and Terrin obtained the joint asymptotic distribution of the Whittle estimates of $H$ based on the $k$ disjoint subseries. More precisely, in order to test the hypothesis $H_0$: $H \equiv H_0$ for the whole series, against the alternative $H_a$: $H \neq H_0$, that is, among the $k$ subseries with corresponding $H$-parameters $H_1, \ldots, H_k$, there exists at least one pair $j \neq i$ such that $H_j \neq H_i$, define the test statistic $T_{1,2,\ldots,k} = \Sigma(\hat{H}_j - \hat{H})^2[\hat{\sigma}_j^2 i]^{-1}$, where $\hat{\sigma}_j^2$ is given by a generalization of the central limit theorem for Whittle's estimator (see [5]). Beran and Terrin then show that, under the null hypothesis $H_0$, $T_{1,2,\ldots,k}$ is asymptotically $\chi^2$-distributed with $k - 1$ degrees of freedom. Hence we reject $H_0$ at the level of significance $\alpha$, if $T_{1,2,\ldots,k} > \chi^2_{k-1;\alpha}$, where $\chi^2_{k-1;\alpha}$ is the upper $(1 - \alpha)$-quantile of the $\chi^2$-distribution with $k - 1$ degrees of freedom.

We illustrate the procedure for testing for a constant $H$-parameter using the series AUG89.MB. More precisely, in order to reduce the amount of computation, we consider the corresponding aggregated time series $X^{(100)}$ of length 3600 depicted in Figure 6, representing the number of bytes per second during the normal-traffic hour of the August 1989 data set. Figure 6 also shows the different ways we partitioned the data into disjoint subsets, together with the Whittle estimate of $H$ for each corresponding subset. Finally, for every partition, the corresponding $P$-values are given at the right side of the figure and show that, for all but two partitions, the null hypothesis of a constant $H$-value is never rejected, with $P$-values larger than 0.20. Also notice that while the $H$-estimates fluctuate quite a bit for the finest partition (i.e., 10 nonoverlapping 10-minute blocks), these fluctuations are well within the allowed range and decrease as the number of blocks in the partition gets smaller. We applied this procedure also to the other data sets described in Table 1, especially to those which resulted in estimated Hurst parameters close to 1. When finding an $H$-estimate close to 1, it is generally advised to analyze the time series further in order to ensure that the high degree of self-similarity is genuine and cannot be explained by elementary arguments. With the exception of the sets FEB92.MP and FEB92.MB, all sequences appear to be adequately modeled using a constant $H$-value.

### 4.2 Generating Synthetic Sequences and Parallel Computing

It is important to be able to generate synthetic data sequences that exhibit features similar to the measured traffic when doing practical work. While exact methods for generating synthetic traces from fractional Gaussian noise and fractional ARIMA models exist (see [41] and [21], respectively), they are, in general, only appropriate for short traces (about 1000 observations). For longer time series, short memory approximations have been proposed such as the *fast fractional Gaussian noise* by Mandelbrot [37]. However, such approximations also often become inappropriate when the sample size becomes exceedingly large.

Here, we briefly mention two methods for generating asymptotically self-similar observations. To our knowledge, both methods are new, although in both cases, the underlying theoretical results have been known for quite some time. The first method exploits a convergence result obtained by Granger [16], who showed that when aggregating many simple AR(1)-processes, where the AR(1) parameters are chosen from a beta distribution on $[0, 1]$ with shape parameters $p$ and $q$, then the superposition process is asymptotically self-similar with self-similarity parameter $H = (3 - q)/2$. This method is obviously tailor-made for parallel computers, and producing a synthetic trace of length 100,000 on a MasPar MP-1216, a massively parallel computer with 16384 processors, takes about 3–5 minutes. In contrast, Hosking's method to produce 100,000 observations from a fractional ARIMA$(0, d, 0)$ model requires a few hours of CPU time on a Sun SPARCstation 2.

The second method is based on a construction

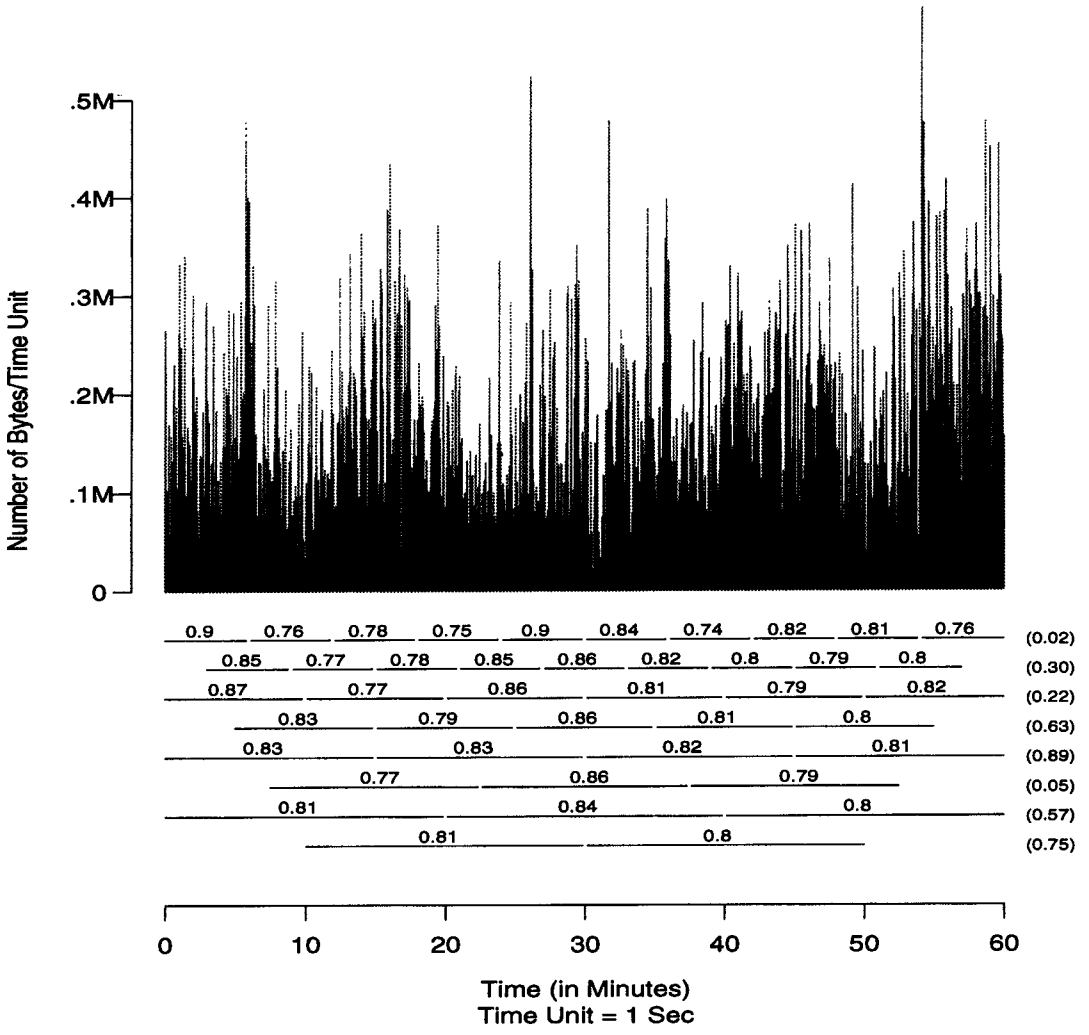


FIG. 6. *Testing the hypothesis that H is constant for the time series $X^{(100)}$ corresponding to the sequence AUG89.MB: indicated are the different ways of partitioning the data into disjoint blocks, together with the resulting H-estimate for each block and the P-value (in parentheses) for each partition.*

originally introduced in an economic context by Mandelbrot [36]. Appropriately normalized, the superposition of an increasing number of i.i.d. renewal reward processes (each one exhibiting heavy-tailed interrenewal times with index $1 < \alpha < 2$, i.e., interrenewal times with infinite variance) over an every increasing time horizon converges (in the sense of the finite-dimensional distributions) to fractional Brownian motion with self-similarity parameter $H = (3 - \alpha)/2$ (see also [50]). Again, letting each processor of a massively parallel machine generate a simple renewal reward process, we see that Mandelbrot's construction of fractional Brownian motion is also ideally suited for parallel computers. Implementations of and experimentations with this method are currently under way.

## 5. DISCUSSION

The main finding of our statistical analysis of hundreds of millions of high-quality, high time-resolution Ethernet LAN traffic measurements is that Ethernet LAN traffic is statistically self-similar. Some sequences are asymptotically self-similar and others can be regarded, for practical purposes, as exactly self-similar. The self-similarity property allows us to distinguish clearly between the packet traffic models currently considered in the literature and our measured data. These results, however, are both disturbing and stimulating. Indeed, the collection, analysis and modeling of traffic data taken from high-speed communication networks and the subsequent use of the proposed models for performance analysis involve multiple disciplines. The data analyst must work closely with teletraffic engineers and experts in computers and applied probability. The results of this paper can be disturbing to the teletraffic engineer, for they question traditional traffic models and thus cast doubt on predicted network performance that are based on queueing models with statistically questionable input processes. On the other hand,

the fact that self-similarity is ubiquitous in our measured data and cannot be captured by conventional traffic models gives rise to stimulating new research problems in statistics, statistical computing, stochastic modeling and queueing theory.

Data analysts are rarely confronted with data sets of the size and quality of the Ethernet LAN traffic measurements considered in this paper. These new types of data sets are likely to stimulate the development of statistical methods that can take advantage of high-performance computing environments. Two such examples are discussed in this paper: fast parameter estimation techniques for large sets of self-similar data and quick generation methods of long traces of synthetic data from a self-similar model. In terms of stochastic modeling of self-similar phenomena, the use of self-similar stochastic processes or their increment processes seems to be the most natural approach. However, there are also some recent promising attempts, mostly of an experimental nature, for describing the "fractal" nature of our measured data with the help of packet interarrival time distributions with infinite means (see [51]) or using deterministic nonlinear chaotic maps (e.g., see [10]). For queueing and performance analysis, all these approaches pose completely new and very challenging problems which are likely to require a new set of mathematical tools. Ultimately, in the context of teletraffic, it is the predicted performance of appropriately chosen queueing systems that will decide about the relevance or irrelevance of self-similar arrival processes.

Finally, there is the ever-present question about a physical "explanation" for the observed self-similarity property in a given data set. To this end, we recall a construction by Mandelbrot [36] as expounded in Taqqu and Levy [50] (see also [33]), originally cast in an economic framework involving commodity prices (see also Section 4.2). In the context of Ethernet LAN traffic, Mandelbrot's construction provides an intuitively appealing argument for the visually obvious (see Figure 2a–e) and statistically significant (see Figure 5) self-similarity property of the aggregate traffic in terms of the behavior of individual Ethernet users. In its simplest form, Mandelbrot's result states that if an individual traffic source goes through "active" periods [during which it generates packets (or bytes) at regular intervals] and "inactive" periods (when no packets are generated) and if the lengths of the active and inactive periods are i.i.d. (and independent from one another) and have infinite variance (or, using Mandelbrot's terminology, exhibit the *Noah effect*), then aggregating many such sources produces traffic that is self-similar in the limit (as the number of sources increases). This convergence result relies heavily on the Noah effect, which assumes that with nonnegligible probability the active and inactive periods can last a very long time. In light of the way a typical Ethernet host (workstation user, file server, router) contributes to the overall traffic on an Ethernet, this seems to be a plausible property. We plan to extract individual user traffic from the aggregate traffic data in Table 1 in order to investigate the validity of the infinite-variance assumption. However, as reported in [42], evidence in support of the Noah effect in packet traffic measurements already exists! In this context, it should also be mentioned that this is not the first study in the area of telecommunications that demonstrates the presence of the Joseph effect (long-range dependence) and Noah effect in actual traffic data. In fact, Mandelbrot himself (see [34] and [35]) introduced these concepts in his analysis of error clustering in analog transmission channels. It is safe to expect that the Joseph and Noah effects will play an increasingly important role in the traffic modeling work of tomorrow's gigabit networks.

## ACKNOWLEDGMENTS

This work could not have been done without the help of J. Beran and R. Sherman, who provided the *S*-functions that made the statistical analysis of an abundance of data possible. We also acknowledge many helpful discussions with A. Erramilli about his dynamical systems approach to packet traffic modeling. Finally, we would like to thank Associate Editor David Griffeath for his numerous comments on an earlier version of the paper that led to a much improved exposition of the material. M. S. Taqqu was supported, at Boston University, by ONR Grant N00014-90-J-1287.

## REFERENCES

[1] ANICK, D., MITRA, D. and SONDHI, M. M. (1982). Stochastic theory of a data-handling system with multiple sources. *Bell System Technical Journal* **61** 1871–1894.

[2] BERAN, J. (1986). Estimation, testing and prediction for self-similar and related processes. Ph.D. dissertation, ETH Zurich.

[3] BERAN, J. (1992). Statistical methods for data with long-range dependence. *Statist. Sci.* **7** 404–427.

[4] BERAN, J., SHERMAN, R., TAQQU, M. S. and WILLINGER, W. (1995). Long-range dependence in variable-bit rate video traffic. *IEEE Trans. Comm.* To appear.

[5] BERAN, J. and TERRIN, N. (1994). Estimation of the long-memory parameter, based on a multivariate central limit theorem. *J. Time Ser. Anal.* **15** 269–278.

[6] CASSANDRO, M. and JONA-LASINIO, G. (1978). Critical behaviour and probability theory. *Adv. in Phys.* **27** 913–941.

[7] COX, D. R. (1984). Long-range dependence: a review. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds.) 55–74. Iowa State Univ. Press.

[8] COX, D. R. and TOWNSEND, M. W. H. (1947). The use of the

correlogram in measuring yarn irregularities. *Proc. Roy. Soc. Edinburgh Sec. A* **63** 290–311.

[9] DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766.

[10] ERRAMILLI, A. and SINGH, R. P. (1992). An application of deterministic chaotic maps to characterize packet traffic. Unpublished manuscript.

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

[12] FOX, R. and TAQQU, M. S. (1986). Large-Sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

[13] GARRETT, M. W. and WILLINGER, W. (1994). Analysis, modeling and generation of self-similar VBR video traffic. Proceedings of ACM Sigcomm '94. *Computer Communication Review* **24** 269–280.

[14] GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4** 221–237.

[15] GRAF, H. P. (1983). Long-range correlations and estimation of the self-similarity parameter. Ph.D. dissertation, ETH Zurich.

[16] GRANGER, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *J. Econometrics* **14** 227–238.

[17] GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1** 15–29.

[18] HAMPEL, F. R. (1987). Data analysis and self-similar processes. In *Proceedings of the 46th Session of the ISI* **4** 235–254. Internat. Statist. Inst., Tokyo.

[19] HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

[20] HEFFES, H. and LUCANTONI, D. M. (1986). A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications* **4** 856–868.

[21] HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176.

[22] HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116** 770–799.

[23] HURST, H. E. (1956). Methods of using long-term storage in reservoirs. *Proceedings of the Institution of Civil Engineers, Part I*, **5** 519–590.

[24] JAGERMAN, D. L. and MELAMED, B. (1992). The transition and autocorrelation structure of TES processes part I: general theory. *Stochastic Models* **8** 193–219.

[25] JAIN, R. and ROUTHIER, S. A. (1986). Packet trains: measurements and a new model for computer network traffic. *IEEE Journal on Selected Areas in Communications* **4** 986–995.

[26] KOLMOGOROV, A. N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *C. R. Acad. Sci. URSS (N. S.)* **30** 301–305. [Translation in *Turbulence* (S. K. Friedlander and L. Topper, eds.) 151–155. (1961). Interscience, New York.]

[27] KUENSCH, H. R. (1986). Statistical aspects of self-similar processes. In *Proceedings of the 1st World Congress of the Bernoulli Society* (Yu. Prohorov and V. V. Sazonov, eds.) **1** 67–74. VNU Science Press, Utrecht.

[28] KUENSCH, H., BERAN, J. and HAMPEL, F. (1993). Contrasts under long-range dependence. *Ann. Statist.* **21** 943–964.

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

[30] LELAND, W. E., TAQQU, M. S., WILLINGER, W. and WILSON, D. V. (1993). On the self-similar nature of Ethernet traffic. Proceedings of ACM Sigcomm '93. *Computer Communication Review* **23** 183–193.

[31] LELAND, W. E., TAQQU, M. S., WILLINGER, W. and WILSON, D. V. (1993). Statistical analysis of high time-resolution Ethernet LAN traffic measurements. In *Proceedings of the 25th Symposium on the Interface. Computing Science and Statistics* (M. E. Tarter, and M. D. Lock, eds.) **25** 146–155. Interface Foundation of North America, Fairfax Station, VA.

[32] LELAND, W. E., TAQQU, M. S., WILLINGER, W. and WILSON, D. V. (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking* **2** 1–15.

[33] LEVY, J. B. and TAQQU, M. S. (1987). On renewal processes having stable inter-renewal intervals and stable rewards. *Ann. Sci. Math. Québec* **11** 95–110.

[34] MANDELBROT, B. B. (1965). Self-similar error clusters in communication systems and the concept of conditional stationarity. *IEEE Transactions on Communications Technology* **COM-13** 71–90.

[35] MANDELBROT, B. B. (1967). Some noises with $1/f$ spectrum, a bridge between direct current and white noise. *IEEE Trans. Inform. Theory* **13** 289–298.

[36] MANDELBROT, B. B. (1969). Long-run linearity, locally Gaussian processes, $H$-spectra and infinite variances. *Internat. Econom. Rev.* **10** 82–113.

[37] MANDELBROT, B. B. (1971). A fast fractional Gaussian noise generator. *Water Resources Research* **7** 543–553.

[38] MANDELBROT, B. B. and TAQQU, M. S. (1979). Robust $R/S$ analysis of long run serial correlation. In *Proceedings of the 42nd Session of the ISI* **48** (Book 2) 69–99. Internat. Statist. Inst., Tokyo.

[39] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.

[40] MANDELBROT, B. B. and WALLIS, J. R. (1969). Some long-run properties of geophysical records. *Water Resources Research* **5** 321–340.

[41] MCLEOD, A. I. and HIPPEL, K. W. (1978). Preservation of the rescaled adjusted range 1: a reassessment of the Hurst phenomenon. *Water Resources Research* **14** 491–508.

[42] MEIER-HELLSTERN, K., WIRTH, P. E., YAN, Y.-L. and HOEFLIN, D. A. (1991). Traffic models for ISDN data users: office automation application. In *Teletraffic and Datatraffic in a Period of Change. Proceedings of the 13th ITC* (A. Jensen and V. B. Iversen, eds.) 167–172. North-Holland. Amsterdam.

[43] NORROS, I. (1994). A storage model with self-similar input. *Queueing Systems* **16** 387–396.

[44] RAMASWAMI, V. (1988). Traffic performance modeling for packet communication: whence, where and whither. In *Proceedings of the 3rd Australian Teletraffic Seminar*.

[45] ROBINSON, P. M. (1994). Semiparametric analysis of long-memory time series. *Ann. Statist.* **22** 515–539.

[46] SAMORODNITSKY, G. and TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York.

[47] SHOCH, J. F. and HUPP, J. A. (1980). Measured performance of an Ethernet local network. *Comm. ACM* **23** 711–721.

[48] TAQQU, M. S. (1985). A bibliographical guide to self-similar processes and long-range dependence. In *Dependence in Probability and Statistics* (E. Eberlein and M. S. Taqqu, eds.) 137–165. Birkhäuser, Boston.

[49] TAQQU, M. S. (1987). Self-similar processes. In *Encyclopedia of Statistical Sciences* **8** 352–357. Wiley, New York.

[50] TAQQU, M. S. and LEVY, J. B. (1986). Using renewal processes to generate long-range dependence and high variability. In *Dependence in Probability and Statistics* (E. Eberlein and M. S. Taqqu, eds.). *Progr. Probab. Statist.* **11** 73–89. Birkhäuser, Boston.

[51] VEITCH, D. (1992). Novel models of broadband traffic. In *Proceedings of the 7th Australian Teletraffic Research Seminar*.

[52] WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434.


Time Unit
60000
40000

Time Unit
60000
40000