Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level

Walter Willinger (Bellcore), Murad S. Taqqu (Boston University), Robert Sherman (Bellcore) and Daniel V. Wilson (Bellcore)

Abstract— **A number of recent empirical studies of traffic measurements from a variety of working packet networks have convincingly demonstrated that actual network traffic is *self-similar* or *long-range dependent* in nature (i.e., bursty over a wide range of time scales) – in sharp contrast to commonly made traffic modeling assumptions. In this paper, we provide a plausible physical explanation for the occurrence of self-similarity in high-speed network traffic. Our explanation is based on convergence results for processes that exhibit *high variability* (i.e., infinite variance) and is supported by detailed statistical analyses of real-time traffic measurements from Ethernet LAN's at the level of individual sources.**

Our key mathematical result states that the superposition of many ON/OFF sources (also known as *packet trains*) whose ON-periods and OFF-periods exhibit the *Noah Effect* (i.e., have high variability or infinite variance) produces aggregate network traffic that features the *Joseph Effect* (i.e., is self-similar or long-range dependent). There is, moreover, a simple relation between the parameters describing the intensities of the Noah Effect (high variability) and the Joseph Effect (self-similarity). An extensive statistical analysis of two sets of high time-resolution traffic measurements from two Ethernet LAN's (involving a few hundred active source-destination pairs) confirms that the data at the level of individual sources or source-destination pairs are consistent with the Noah Effect. We also discuss implications of this simple physical explanation for the presence of self-similar traffic patterns in modern high-speed network traffic for (i) parsimonious traffic modeling, (ii) efficient synthetic generation of realistic traffic patterns, and (iii) relevant network performance and protocol analysis.

I. Introduction

Starting with the extensive analyses of traffic measurements from Ethernet LAN's over a 4-year period described in [16], there have been a number of recent empirical studies that provide evidence of the prevalence of *self-similar* or *fractal* traffic patterns in measured traffic from today's high-speed networks. Prominent among these studies are the in-depth statistical analysis of large amounts of wide-area traffic measurements reported in [24] and the detailed investigation of traffic data collected at the packet level from multiple NSFNET core switches presented in [13]. One of the most surprising findings from these and other studies concerns the ease with which it is possible to statistically distinguish between measured network traffic and traditional traffic models: 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. Although such findings can in general be expected to favor the use of self-similar models over traditional models from a statistical perspective, there has been considerable resistance toward self-similar traffic modeling on practical grounds. One of the major reasons for this resistance has been the absence of satisfactory answers to the following 2 questions. (1) What is a physical "explanation" for the observed self-similar nature of measured traffic from today's packet networks? and (2) What is the impact of self-similarity on network and protocol design and performance analysis?

In this paper, we present an answer to question (1) by providing the appropriate mathematical results and by validating our findings with detailed statistical analyses of two representative sets of high time-resolution traffic measurements from two different Ethernet LAN's. In particular, we provide 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 [19], we show that the superposition of many (idealized) ON/OFF sources, each of which exhibits a phenomenon called the "Noah Effect", results in self-similar aggregate traffic. By expressing the results in the well-known framework of ON/OFF source models (also known as "packet train models"), we identify the Noah Effect as the essential point of departure from traditional to self-similar traffic modeling. Intuitively, the Noah Effect for an individual ON/OFF source model results in highly variable ON- and OFF-periods, i.e., "train lengths" and "intertrain distances" that can be very large with non-negligible probability. In other words, the Noah Effect guarantees that each ON/OFF source individually exhibits characteristics that cover a wide range of time scales. The Noah Effect is synonymous with the *infinite variance syndrome* – the empirical observation that many naturally occurring phenomena can be well described using distributions with infinite variance. Mathematically, we use *heavy-tailed* distributions with infinite variance (e.g., certain Pareto and stable distributions) to account for the Noah Effect, and the parameter α describing the "heaviness" of the tail of such a distribution gives a measure of the intensity of the Noah Effect. We also provide a simple relation between α and the Hurst parameter H, where the latter has been suggested in [16] as a measure of the degree of self-similarity (or equivalently, of the "Joseph Effect") of the aggregate traffic stream.

W. Willinger and M.S.Taqqu were partially supported by the NSF grant NCR-9404931. M.S. Taqqu was also partially supported by the NSF grant DMS-9404093.

To appear in: *Proc. of the ACM/Sigcomm'95*

In sharp contrast to our findings, traditional traffic modeling, when cast in the framework of ON/OFF source models, without exception assumes 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 [17]). Note that the results of the present study suggest yet another, equally simple, statistical method for distinguishing between traditional and self-similar traffic: an analysis of network traffic that checks for the presence or absence of the Noah Effect in the traffic generated by the individual sources or source-destination pairs. To demonstrate the effectiveness of such an analysis, we used two sets of Ethernet LAN traffic measurements, generated by about 100 and 3,200 individual sources (resulting in about 700 and 10,000 active source-destination pairs), respectively. The data were collected at the Bellcore Morristown Research and Engineering Center (MRE). One of the data sets was collected in August of 1989, has been studied extensively in the past (at the aggregate packet level), and was, in fact, part of the analysis presented in [16]. The second data set represents a very recent (December 1994) collection of high time-resolution traffic measurements, was obtained from a different Ethernet LAN than the first set, and includes applications that were non-existent in the first data set (e.g., WWW and Mbone).

Although the main objective of this paper is to provide an answer to question (1) (physical "explanation"), our results concerning individual source behavior are clearly significant for answering question (2) (impact of self-similarity on network and protocol design and performance analysis). Starting with the work by Norros [22], there has been mounting evidence that clearly shows that the performance of queueing models with self-similar inputs can be radically different from the performance predicted by traditional traffic models, especially by Markovian models (e.g., see [5], [4], [7]). Here we complement this evidence by illustrating the practical relevance of our findings for (i) parsimonious traffic modeling for high-speed networks, (ii) efficient simulation of actual network traffic, and (iii) analyzing queueing models and protocols under realistic traffic scenarios.

Two previous studies of LAN traffic measurements are of particular relevance in the present setting. Jain and Routhier [12] used packet data collected at a ring network at MIT and proposed a "packet train" (or ON/OFF) source model in order to capture the observed burstiness in actual packet streams. In this context, our results show that packet train models are consistent with traffic data collected at the level of individual source-destination pairs – once the Noah Effect for the packet-train lengths and the inter-train distances has been accounted for. By doing so, some of the shortcomings of the original packet train modeling approach (e.g., lack of any physical interpretation, arbitrary choice of crucial parameter values) are remedied. Of particular importance to our work are Gusella's extensive studies [8], [9], [10] of traffic measurements from a 10 Mb/s Ethernet LAN. In view of the results discussed in the present paper, Gusella's work falls strictly within the traditional approach to traffic modeling: phenomena like the Joseph and Noah Effects are attributed to non-stationarity in the data and are ignored in subsequent data modeling. Naturally, the resulting models, based on burstiness characterizations using indices of dispersion, are adequate only over a limited range of time scales. Our approach suggests a viable alternative: by expanding the range of traditional traffic models to account for the Joseph and Noah Effects, it is possible to describe these phenomena in a strictly stationary setting. The benefits for doing so include new insights into the time dynamics of high-speed network traffic, and the applicability of simple models for the very complex traffic patterns observed in today's networks. Finally, with regard to an intuitive physical explanation at the *application level* of the empirically observed self-similar nature of wide-area network traffic, we refer to [24], as well as to [13]. Note that in these application-level studies, the Noah Effect also plays a crucial role.

The rest of the paper is organized as follows. In Section II, we introduce an idealized ON/OFF source model and present the convergence theorems that form the basis of our approach. In Section III, we discuss the available traffic measurements and present our statistical analysis of these data, concentrating on detecting the Noah Effect in traffic generated by individual source-destination pairs. Finally, in Section IV we illustrate the significance of the presence of the Noah Effect at the source level and its implications for aggregate traffic streams with a number of examples that are relevant and of practical importance for the design and performance analysis of modern communication networks and protocols.

II. Self-Similarity Through High-Variability

The models presented here were first introduced by Mandelbrot [19] and Taqqu and Levy [29] and were originally cast in an economic setting involving commodity prices. The models are based on renewal reward processes and are rephrased here in the context of packet traffic modeling. Intuitively, they take into account the presence of long packet trains ("ON-periods", i.e., periods during which packets arrive at regular intervals) and long inter-train distances ("OFF-periods", i.e., periods with no packet arrivals) in traffic generated by individual sources or individual source-destination pairs in a LAN. We will show in this section that the superposition of many such packet trains exhibits, on large time scales, the self-similar behavior that has been observed in the Ethernet LAN traffic data in [16].

We consider here *idealized* ON/OFF models where an ON-period can be followed by an ON-period and an OFF-period can succeed another OFF-period. Although the ON/OFF models commonly considered in the communications literature have strictly alternating ON- and OFF-periods and hence differ from the idealized models considered here, we chose the idealized setting because it allows for an immediate application of some known results in [19], [29]. It also allows for the distributions of the ON and OFF times to vary. The study of the more traditional (i.e., alternating) ON/OFF models, as well as a comparison between the idealized and strictly alternating setting will appear in a subsequent paper. Moreover, below we present only the simplest case of an idealized ON/OFF model; generalizations (e.g., different distributions for the ON- and OFF-periods, non-homogeneous sources) are possible, but the details are also deferred to a later paper.

Following Mandelbrot's original work, an *idealized* ON/OFF source model, or simple packet train model, is characterized by a *reward sequence* $\{W(l), l = 0, 1, \ldots\}$, i.e., a 0/1-valued discrete time stochastic process$\{W(l)\}$, with $W(l) = 1$ or 0, depending on whether or not there is a packet at time l. Thus, the reward sequence $\{W(l)\}$ consists of a sequence of 1's ("ON-periods") and 0's ("OFF-periods"). Let $p \equiv P($ a given period is an ON-period $) = 1/2$, and assume that the lengths of the ON- and OFF-periods are independent and identically distributed (i.i.d.) positive random variables, denoted $U_k,\ k = 1, 2, \ldots$ (U denotes an arbitrary U_k, with finite expectation $E(U)$). Let $S_k = S_0 + U_1 + U_2 + \ldots + U_k,\ k \geq 0$ be the corresponding renewal times. We assume that $\{S_k,\ k \geq 0\}$ is stationary. This can be achieved by choosing the distribution of S_0 in the following special way:

$$P(S_0 = u) = (E(U))^{-1} P(U \geq u + 1),\ u = 0, 1, 2, \ldots$$

To ensure the stationarity of the reward sequence $\{W(l),\ l \geq 0\}$, let $l = 0$ be in an ON-period with probability 1/2.

Next suppose that there are M i.i.d. sources, where the m^{th} source ($m = 1, \ldots, M$) has it own reward sequence $\{W^{(m)}(l),\ l \geq 0\}$. Then the superposition or cumulative reward ("packet load") at time l is $\sum_{m=1}^{M} W^{(m)}(l)$. Aggregating this load through (non-overlapping) time blocks of size b, we get

$$W^*_{M,b}(j) = \sum_{l=bj+1}^{b(j+1)} \sum_{m=1}^{M} W^{(m)}(l), \quad j = 0, 1, 2, \ldots,$$

where j denotes the aggregation block number. We are interested in the statistical behavior of the sequence $\{W^*_{M,b}\}$ for large M and b. This behavior can only depend on the distribution of U, the one element we have not yet specified. Motivated by the empirically derived fractional Gaussian noise model for aggregate packet traffic in [16], we want to choose the distribution of U in such a way that, as $M \to \infty$ and $b \to \infty$, $\{W^*_{M,b}\}$ adequately normalized is *fractional Gaussian noise* $\{G_{H,\sigma}(t),\ t \geq 0\}$, the only Gaussian sequence which is self-similar (with *Hurst parameter* $\frac{1}{2} \leq H < 1$) at all scales. By this we mean that the finite-dimensional distributions of $b^{-H} \sum_{l=bj+1}^{b(j+1)} G_{H,\sigma}(l),\quad j = 0, 1, 2, \ldots$ are the same whatever the value of the block aggregation size b. (For more information about fractional Gaussian noise and the corresponding cumulative process, called *fractional Brownian motion*, we refer for example to [27, Chapter 7].) In our setting, to obtain fractional Gaussian noise we suppose that U has a hyperbolic tail distribution, that is, it satisfies

$$P(U > u) \sim c\, u^{-\alpha} \quad \text{as} \quad u \to \infty,\ 1 < \alpha < 2, \tag{1}$$

where c is a positive finite constant, independent of u. Mandelbrot refers to property (1) as the *infinite variance syndrome* or the *Noah Effect*. Note that $\alpha < 2$ implies $E(U^2) = \infty$, while the choice $\alpha > 1$ ensures that $E(U) < \infty$ and hence that S_0 is not infinite. For example, U may have a discrete Pareto-type distribution or be some discrete version of a one-sided stable distribution (e.g.,see [27]). One can show, as in [29], that under the conditions stated above the following holds.

Theorem 1. *For large enough source number M and block aggregation size b, the cumulative load $\{W^*_{M,b}(j),\ j \geq 0\}$ behaves statistically as*

$$bM\frac{1}{2} + b^H M^{1/2} G_{H,\sigma}(j)$$

where $H = \frac{3-\alpha}{2}$ and $\sigma^2 = \frac{c}{4E(U)2(\alpha-1)(2-\alpha)(3-\alpha)}$. More precisely,

$$\mathcal{L}\lim_{b\to\infty} \mathcal{L}\lim_{M\to\infty} b^{-H} M^{-1/2} \left(W^*_{M,b}(j) - \frac{bM}{2} \right) = G_{H,\sigma}(j),$$

where $\mathcal{L}$ lim means convergence in the sense of the finite-dimensional distributions (convergence in law).

Heuristically, Theorem 1 states that the mean level $bM/2$ provides the main contribution for large M and b; fluctuations from that level are given by the fractional Gaussian noise $G_{H,\sigma}(j)$ scaled by a lower order factor $b^H M^{1/2}$. Note that it is essential that the limits be performed in the order indicated. Also note that $1 < \alpha < 2$ implies $1/2 < H < 1$. Thus, the main ingredient that is needed for the limiting result to hold is the hyperbolic tail behavior (1), which guarantees the infinite variance property (i.e., high variability) of the ON- and/or OFF-periods of a "typical" source; whether the ON/OFF periods form a strictly alternating renewal process or an i.i.d sequence is not essential.

Theorem 1 can be generalized in a number of different directions. In particular, we mention here the possibility for (i) allowing rewards in an ON-period to be given by positive i.i.d. random variable with finite variance (e.g., the rewards can equal the number of bytes in a packet), and (ii) considering heterogeneous sources (i.e., each source type satisfies the hyperbolic tail property of the form (1), where the index α is type-dependent). In the case of heterogeneous sources, the limit is a superposition of independent fractional Gaussian noises with different (type-specific) H's. As far as the fluctuations are concerned, however, the term with the highest H (or equivalently, the term corresponding to the smallest α for which the corresponding proportion of source types to total number of sources does not converge to 0) ultimately dominates as $b \to \infty$. When the distribution of the length of packet trains has finite variance, the corresponding source types will contribute to the limit an ordinary white noise component. Details about the proofs of these generalizations of Theorem 1, as well as a statement for the case of the superposition of strictly alternating ON/OFF models will appear in a later paper.

III. Ethernet LAN Traffic Measurements at the Source Level

In this section, we first describe two sets of Ethernet LAN traffic measurements that will be used in our subsequent analysis. The two sets result in about 500 and 10,000 active source-destination pairs, respectively. This wealth of data presents a considerable challenge when trying to investigate in a statistically rigorous manner the presence of the Noah Effect in the traffic streams generated by all or part of these individual active source-destination pairs. Thus, one of the main objectives of this section is to illustrate the use of exploratory data analysis tools that can assist in extracting essential information out of an abundance of traffic measurements. While some of the tools applied below are well-known, others are less familiar and will be explained in more detail as they are used. Finally, note that we are not interested in precise point estimates for the index α appearing in equation (1) nor in measuring the exact intensity of the Noah Effect for a given source or source-destination pair, but are instead concerned about the range of α-values that is consistent with the data representing individual source-destination pairs.

A. *Traffic Measurements*

The first set of traffic measurements is the busy hour of the August 1989 Ethernet LAN measurements presented and analyzed (and denoted by AUG89.HB and AUG89.HP in Table 1) in [16]. (Source-level analyses of the other data sets considered in [16] result in similar conclusions – for further details about these data sets, see also [17].) In addition to the information about time stamp and size (in bytes) of every packet seen during this hour, this first data set also contains the source and destination address of each recorded packet. During this busy hour, 105 hosts sent or received packets over the network (out of 121 hosts that were active during the whole 27 hour long monitoring session). Upon further inspection, out of 11,025 possible source-destination pairs, only 748 or about 6.8% were actually sending or receiving packets (this effect has also been observed in previous traffic studies, e.g., [3], [23]). The most active hosts were sources 1, 7, 11, 27, 32, 58 (6 Sun-3 fileservers), sources 2 and 47 (2 DEC 3100 fileservers), source 34 (a Sun-4 server), sources 6, 15, 20, 25, 30, 63 (6 diskless Sun-3 clients), source 8 (a DEC 3100 client), and sources 10 and 17 which served as routers. Only about 5% of the traffic on this network was *external*, i.e., destined for machines on other networks or outside the company.

The second data set is new and represents a "typical" hour of Ethernet LAN traffic collected during a 9 day-long measurement period in December 1994 (additional hour-long periods of this traffic trace have also been analysed and show similar results, despite differences in the traffic mix). The traffic was gathered from the stub Ethernet between the router provided by Bellcore's Internet service provider and a second Bellcore-controlled router that enforces security. The measurements are made up entirely of *remote* traffic, i.e., of packets destined for points on the Internet outside of Bellcore or for Bellcore from the outside (all via a 1.5 Mb/s T-1 link). Our motivation for including a very recent set of traffic measurements in our study was to demonstrate the robustness of traffic characteristics such as the Joseph and Noah Effects under a variety of changes that working LANs experience in practice over time. LAN environments undergo considerable changes with regard to network configuration, host population, hardware and software upgrades, user applications, etc. For this data set, the number of active hosts (based on IP addresses) turns out to be about 3,500, while the percentage of active to possible source-destination pairs is about 0.25%. The most active host in this data set was the machine outside of Bellcore that sent *Mbone* packets (see below and Section III.F for more details regarding *Mbone*). Also included in the most active machines were four machines outside of Bellcore supplying data in response to file transfer (FTP) sessions, along with one Bellcore host supplying file transfer, E-mail, and Domain name service to the outside world. Other active hosts included one Bellcore host supplying Network News to the outside and three machines supplying news to Bellcore. Of the two Bellcore machines mentioned above, one is a Sun Sparcserver 690MP and the other (the network news supplier) is a Sparc-1. Note that all remote traffic is bandwidth limited by Bellcore's 1.5 Mb/s link to the outside world.

It is also known that LAN environments can experience drastic changes at the user application level within relatively short periods of time (for a similar finding regarding WAN traffic, see [23]). A brief investigation of what services generated this second hour-long data set revealed that a new Internet service called *Mbone* (see for example, [18]) was responsible for over 50% of the recorded traffic (in bytes). Another new service, the World Wide Web (WWW) information retrieval service (e.g., see [1]), made up 9.4% of the total traffic. Neither Mbone nor WWW traffic was present in the first data set, nor in any of the earlier data sets studied in [16]. Services such as file transfer (14.5%), telnet/rlogin (2.8%), electronic mail (SMTP) (3.2%) and Network News transfer (NNTP) (12.2%) still present significant components of the total traffic but no longer dominate it.

B. *Textured Plots and the Packet Train Assumption*

We consider here the first data set that has been shown in [16] to be consistent with second-order self-similarity, with a Hurst parameter of $H \approx 0.90$ for the time series representing the packet counts per 10 milliseconds. This conclusion was reached by treating the Ethernet packets as black boxes, i.e., without using any information contained in the packet header fields. In contrast, for the present study, we extracted from the header field of each packet monitored during this hour the corresponding pair of source-destination addresses. This process resulted in 105 individual time series representing the packet arrivals on the Ethernet from the 105 hosts that were active (i.e., sent or received packets) during this hour. Furthermore, separating the packets generated by a given source depending on the packet's destination address yields a total of 748 time series corresponding to the number of active source-destination pairs. In view of the results presented in Section II, we are thus faced with the challenging task of analyzing 748 time series with sufficient statistical rigor and accuracy to conclude whether or not these data support our physical explanation for self-similarity, i.e., whether or not the data are consistent (i) with the ON/OFF traffic model assumption for individual sources or source-destination pairs and (ii) with the crucially important assumption of the Noah Effect for the corresponding ON- and OFF-periods. To this end, our goal is not to provide a single point estimate for the intensity α of the Noah Effect, but to examine if there is evidence for the Noah Effect in the data and if so, to determine the "typical" range of α-values. Note that because of the basic relation $H = (3 - \alpha)/2$ (see Theorem 1), the earlier findings in [16] of $H \approx 0.90$ for the time series of (aggregate) packet counts suggests the presence of the Noah Effect with a low α-value of about 1.20.

For the purpose of checking the appropriateness of the ON/OFF traffic modeling assumption for individual sources or source-destination pairs, we first make use of a simple exploratory data analysis tool called *textured dot strip plot* or simply *textured plot*, originally proposed in [30] (see also [28]). Intuitively, the idea of textured plots is to display one-dimensional data points in a strip in an attempt to show all

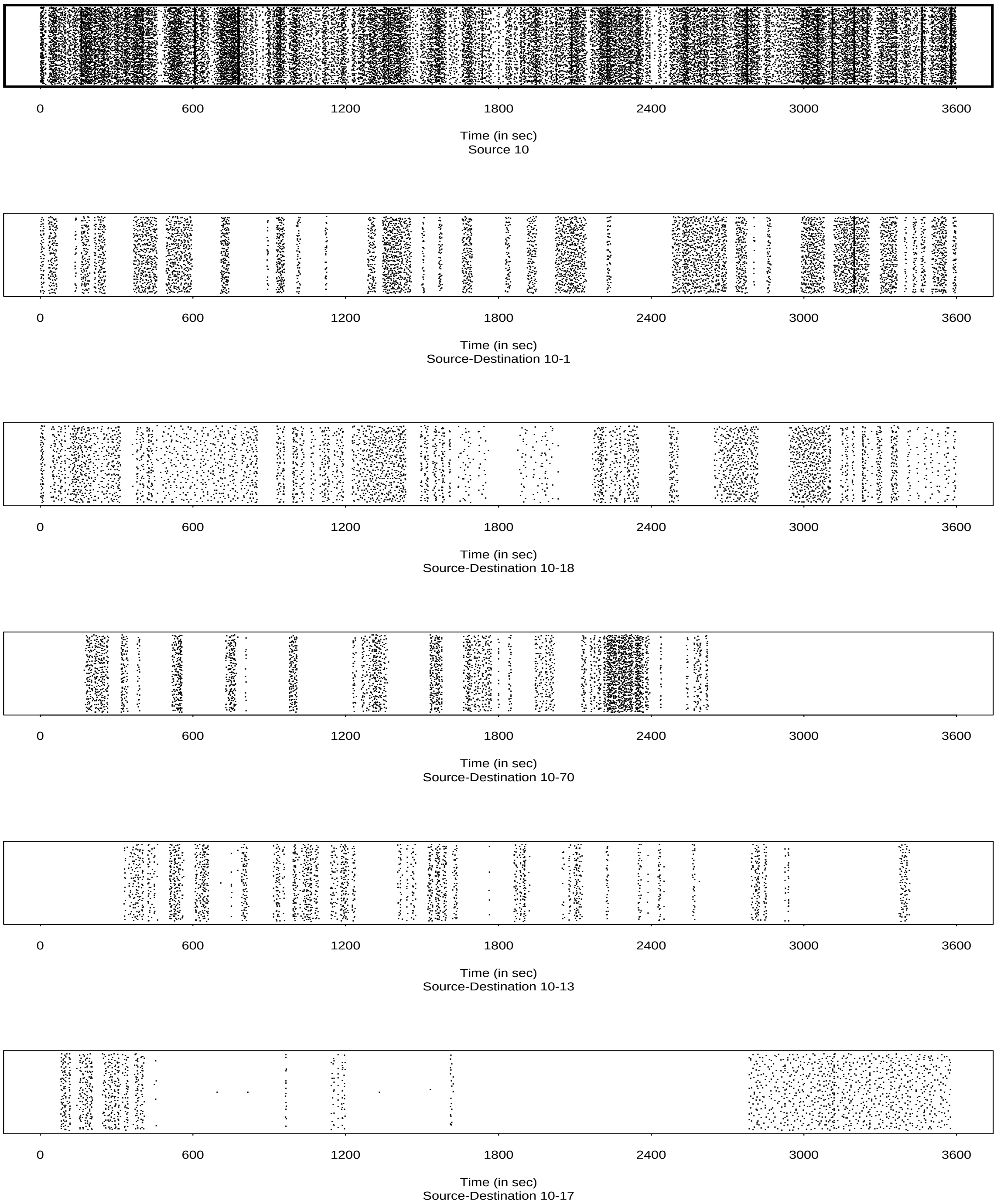

Fig. 1. Textured plots of packet arrival times for source 10 and source-destination pairs 10-1, 10-18, 10-70, 10-13 and 10-17.

data points individually. Thus, if necessary, the points are displaced vertically by small amounts that are partly random, partly constrained. The resulting textured dot strip facilitates a visual assessment of changing patterns of data intensities in a way other better-known techniques such as histogram plots, one-dimensional scatterplots, or box-plots are unable to provide, especially in the presence of extreme values. To illustrate the effectiveness of textured plots for assessing the bursty or ON/OFF nature of traffic generated by an individual source or source-destination pair, we display in Figure 1 six textured plots associated with source 10 (other sources result in similar plots). Each point in the plots represents the time of a packet arrival. Serving as a router, source 10 contributed 1.85% to the overall number of packets and sent data to 25 different destinations. The top plot in Figure 1 represents the textured dot strip corresponding to the arrival times of all packets originating from source 10 (there are 26,330 packets), and the subsequent 5 panels result from applying the textured plot technique to the arrival times of all packets originating from source 10 and destined for sources 1, 18, 70, 13 and 17, respectively. These 5 source-destination pairs were responsible for 5,901, 4,050, 3,407, 2,135, and 1,918 packets, respectively, and make up about 66% of all the packets generated by source 10.

Figure 1 supports two important observations regarding the bursty behavior of traffic generated by (i) a reasonably active individual source (e.g., source 10) and (ii) a "typical" individual source-destination pair (e.g., source-destination pair 10-18). First, a close look at the top plot of Figure 1 clearly reveals the burstiness expected from actual packet traffic, but offers little hope for supporting the ON/OFF nature of the underlying traffic that gave rise to this strip plot. The plot is even more discouraging from the point of view of hoping for some "objective" criterion for identifying ON/OFF periods (but see Section III.D for an appealing robustness property). However, a glance at the 5 source-destination plots in Figure 1 makes the ON/OFF behavior of the traffic generated by the individual source-destination pairs immediately apparent. There is little question of what is meant by an ON- or OFF-period, and subsequent extraction of the lengths of the ON- and OFF-periods from a given source-destination traffic trace is greatly facilitated by use of of the corresponding textured plot of packet arrival times.

C. Checking for the Noah Effect

Next we illustrate the techniques used for determining the presence or absence of the Noah Effect for the ON- and OFF-periods derived from the traffic data generated by individual sources or source-destination pairs. In the case where the data are found to be consistent with the Noah Effect, these techniques also allow for fast procedures (partly heuristic, partly rigorous) for estimating the intensity α of the Noah Effect. As we will demonstrate, these techniques work best when applied in combination with each other and with attention to the physical structure of the data. Specifically, we make extensive use of *complementary distribution plots* (related to the *qq-plot* method [14]) and *Hill's method* [11], [26] for estimating α.

In order to determine the presence or absence of the Noah Effect in a given data set, we take logarithms of both sides of relation (1), obtaining

$$\log(P(U > u)) \sim log(c) - \alpha\ log(u), \quad \text{as} \quad u \to \infty, \qquad (2)$$

where $1 < \alpha < 2$. Using complementary distribution plots, i.e., plotting (on a $\log - \log$ scale) the complementary empirical distribution function of a sample that was presumably drawn from a distribution that exhibits hyperbolic tails (i.e., satisfies (1)), results in an approximately straight line for large u-values, with a slope of $-\alpha, 1 < \alpha < 2$. To illustrate the effectiveness of this technique for the data at hand, we concentrate on source-destination pair 10-18 (Figure 1, panel 3); other soure-destination pairs yield similar results. Based on its textured plot, we define an OFF-period to be *any interval of length $t \geq 2$ seconds that does not contain any packet*; this, in turn, defines the ON-periods unambiguously and results in a total of 202 ON-periods and the same number of OFF-periods for this source-destination pair. (We will return to the issues of the particular choice of the threshold t and of the robustness of our results under different threshold values in Section III.C below.) Figure 2 depicts the complementary distribution plots of the ON-periods (top left) and OFF-periods (top right) and indicates a straight line behavior for large u-values, i.e., a hyperbolic tail distribution satisfying (1) for the ON- and OFF-periods. In fact, a heuristic estimate (obtained by "eyeballing" a straight line through the points to the right of the dashed vertical line) yields $\alpha \approx 1.7$ for the ON-periods and $\alpha \approx 1.2$ for the OFF-periods. To compare, Figure 2 also includes the complementary distribution plot of (i) an exponential distribution with the same mean of 7.2 s as the ON-periods (bottom left), and (ii) a Pareto distribution with $\alpha = 1.2$ and the same mean of 10.5 s as the OFF-periods (bottom right). Clearly, when compared to the tail of the ON-periods, the tail of the exponential distribution is concentrated on a very narrow range of u-values and falls off far too fast. On the other hand, the Pareto distribution covers practically the same range of time scales as the empirically observed OFF-periods and matches the straight line behavior of the data over practically the whole x-axis.

While complementary distribution plots often provide solid evidence for or against the Noah Effect in a given data set, the eyeballing method described above for producing a rough estimate for α is cumbersome and unsatisfactory. A statistically more rigorous method for estimating the intensity of the Noah Effect is known as *Hill's estimate* and is described in [11] (see also [26]). Briefly, let $U_1, U_2, \ldots, U_n$ denote, for example, the observed ON-periods, and write $U_{(1)} \leq U_{(2)} \leq \ldots \leq U_{(n)}$ for the corresponding order statistics. Hill's estimate is given by

$$\hat{\alpha_n} = \left(1/k \sum_{i=0}^{i=k-1} (\log U_{(n-i)} - \log U_{(n-k)}) \right)^{-1}, \qquad (3)$$

where the choice of $1 < k \leq n$ indicates how many of the largest observations enter into the calculation of formula (3). In practice, one plots Hill's estimate $\hat{\alpha_n}$ as a function of k, for a range of k-values. In the presence of a tail behavior in the data that is consistent with (1), a typical *Hill plot* varies considerably for small values of k (i.e., only a small fraction of the largest observations are considered), but becomes more stable as more and more data points in the tail of the distribution are included (often up to a cut-off value, to the left of which (1) no longer holds). An apparent straight line behavior for large u-values in the complementary distribution plot corresponds to a region of k-values where the Hill estimate remains stable. In the absence of such a straight line behavior, Hill's estimate will continue to decrease as k increases, a strong indication that the data are not consistent with the hyperbolic tail assumption (1). For further properties (and shortcomings) of the Hill estimate, see for example [26]. Figure 3 depicts the Hill estimate plots corresponding to the data used in Figure 2. Recalling that each data set contains 202 observations, the top left plot depicts the Hill estimate for the ON-periods and should be viewed together with the top left plot in Figure 2; note the region of stability in the Hill plot (k-values between 20 to about 70), i.e., the tail of the distribution that is consistent with the hyperbolic decay as given in (1) contains about 40% of all the observations. Moreover, the Hill estimate can be readily read off from the y-axis and yields $\hat{\alpha} \approx 1.7$. In the case of the Hill plot for the OFF-periods (top right) and the fitted Pareto model (bottom right), the situation is obvious and in agreement with the information contained in the corresponding complementary distribution plots in Figure 2. Finally, the

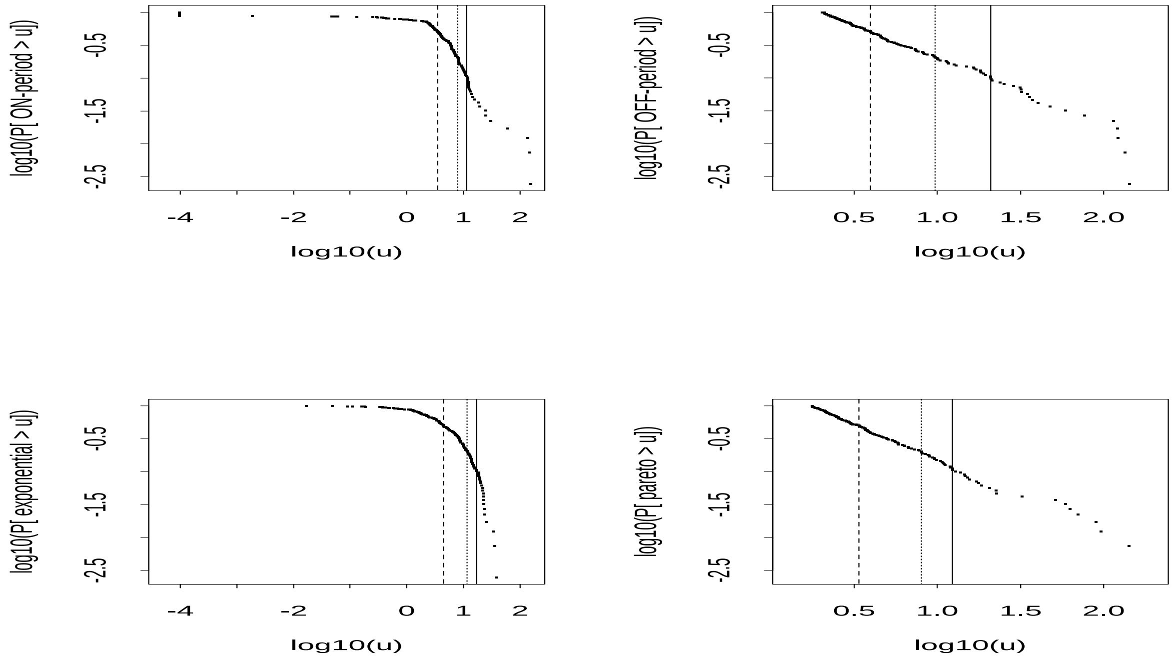

Fig. 2. Complementary distribution plots for *ON-periods* (top left) and *OFF-periods* (top right) for the source-destination pair 10-18, using a threshold value of $t = 2s$; for a sample from an *exponential distribution* that matches the ON-periods (lower left), and for a sample from a *Pareto distribution* that matches the OFF-periods (lower right). (The vertical solid, dotted and dashed lines indicate that 10%, 20% and 50% of all data points are to the right of the respective lines.)

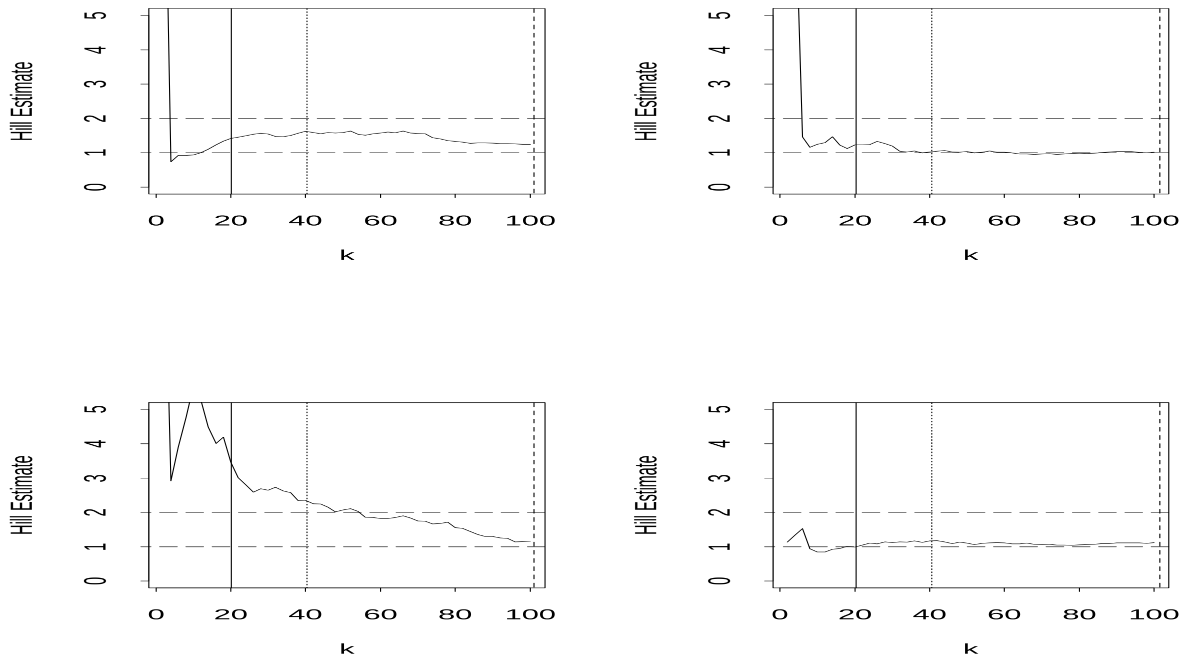

Fig. 3. Hill estimate plots for *ON-periods* (top left) and *OFF-periods* (top right) for the source-destination pair 10-18, using a threshold value of $t = 2s$; for a sample from an *exponential distribution* that matches the ON-periods (lower left), and for a sample from a *Pareto distribution* that matches the OFF-periods (lower right). (The vertical solid, dotted and dashed lines indicate that 10%, 20% and 50% of the largest order statistics have been included in the Hill estimation calculation.)

bottom left plot in Figure 3 illustrates the typical behavior of the Hill plot when the data are inconsistent with assumption (1); the plot does not settle down but continues to decrease as more and more of the smaller order statistics are included in the calculation of the estimator. Intuitively, this behavior is caused by the concave shape (throughout the whole x-axis) of the complementary distribution plot (see bottom left plot in Figure 2).

D. A Robustness Property of the Noah Effect

Before checking for the presence of the Noah Effect in the traffic traces generated by the remaining source-destination pairs, we first point out a robustness feature of the Noah Effect that should greatly diminish any reluctance toward using ON/OFF or packet train models at the source or source-destination level. In the past, such reluctance has typically been based on a lack of physical interpretation or intuition for defining objectively the notion of an OFF-period (or, using the notation of Section III.B, for selecting the "right" threshold value t). In the packet train terminology, the problem is to decide in a coherent manner on the "appropriate" intertrain distance, i.e., on deciding when the "departure" of the previous train took place and when the "arrival" of the next train occurs. Here we show, that *as far as the Noah Effect is concerned, it does not matter how the OFF-periods or intertrain distances (and subsequently, the ON-periods or packet train lengths) have been defined*. In other words, the Noah Effect is robust under a wide range of choices for the threshold value t that we used in Section III.B to explicitly define OFF-periods as any interval of length t seconds or larger that sees no packet arrival.

The reason behind this insensitivity of the Noah Effect for (non-degenerate) OFF-periods to different choices of the threshold t is the well-known scaling property of distributions that satisfy the hyperbolic tail condition (1). Here, by scaling property we mean that if the distribution of U satisfies (1) and t denotes a threshold value, then for sufficiently large u, t with $u > t$,

$$P(U > u | U > t) \sim (\frac{u}{t})^{-\alpha}, 1 < \alpha < 2, \tag{4}$$

(see [20] and also [24], where this property is discussed and used for an intuitive explanation of the Joseph Effect in measured TELNET traffic traces). Thus, the tail behavior of the (conditional) distributions of U given $U > t$, for different choices of the threshold t, differs only by a scaling factor and hence gives rise to complementary distribution plots with identical asymptotic slopes but different intercepts. This appealing robustness property of the Noah Effect for the OFF-periods with respect to the choice of t is illustrated in Figure 4 where we show the complementary distribution plots (top row) and corresponding Hill estimate plots (bottom row) for three different ways of defining the OFF-periods for the traffic associated with source 10 (see top plot of Figure 3). More specifically, we chose t-values that span 3 orders of magnitude, namely $t = 1s$ (left column, 313 observations), $t = 0.2s$ (middle column, 4,537 observations) and $t = 0.025s$ (right column, 19250 observations). Figure 4 confirms the robustness property of the Noah Effect under the different choices of t, with an estimated intensity between $1.6 - 1.9$.

Next recall that defining OFF-periods for a given source or source-destination pair by using a specific threshold value t implicitly defines a corresponding sequence of ON-periods. However, in contrast to the OFF-periods, where the choice of threshold values $s < t$ leaves the OFF-periods obtained via the t-threshold unaffected, ON-periods that were defined based on the t-threshold will typically get fragmented into smaller ON- and OFF- periods when threshold values $s < t$ are used. However, a similar investigation of the sensitivity of the ON-period distributions to the choice of threshold value reveals the same appealing robustness feature of the Noah Effect for the ON-periods as we observed for the OFF-periods. For an intuitive explanation (assuming the idealized modeling assumptions in Section II), recall that a t-ON period (i.e., an ON-period that was obtained using the threshold value t) typically consists of a number of s-ON and s-OFF periods where $s < t$. However, if the s-ON/OFF periods satisfy relation (1), then so does their sum (and vice versa; e.g., see [27]). Subsequently, when interested in the intensity of the Noah Effect for the ON-periods, fragmentation into smaller ON/OFF-periods as the threshold value t decreases should have no effect, and suggests that the Noah Effect for the ON-period is robust. Note that in the case of the ON-periods, determining a sensible range of t-values for investigating this robustness property is facilitated by the fact that as t gets small, the number of resulting ON-periods quickly approaches the number of total packets in a given traffic trace; in other words, for small enough t, practically every packet is counted as an ON-period, and every packet interarrival time as an OFF-period. Clearly, sensible threshold choices are those that stay well above such critical values. To illustrate the robustness property of the Noah Effect for the ON-periods, we consider source 6 that generated a total of 48,275 packets (of which 97% went to source 32), and depict in Figure 5 the complementary distribution plots and Hill plots corresponding to the sequences of ON-periods defined via the threshold values $t = 1.0s, 0.1s$ and $0.05s$. As can be seen, the intensity of the Noah Effect remains essentially unaffected (and equals about 1.3), even though t varies from seconds to 100 milliseconds to tens of milliseconds. Sources other than source 6 yield similar results and provide convincing evidence of the robustness property of the Noah Effect of the ON-periods: although there is no "natural" division into ON/OFF-periods at the source level, such a division becomes apparent at all (a wide range of) time scales; moreover, these divisions appear in a consistent manner.

E. Self-Similarity and the Noah Effect: 1989 Traffic Traces

To facilitate the full-fledged statistical analysis of the busy hour of the August 1989 traffic measurements at the source-destination level, we considered in detail only the 181 most active (out of a total of 748 active) source-destination pairs. Together, these 181 source-destination pairs generated more than 93% of all the packets seen on the Ethernet during this hour and represent more than 98% of the overall traffic (in bytes). As a rule, we chose to neglected all source-destination pairs that generated fewer than about 300 packets during the whole hour. Our statistical analysis of this abundance of traffic data benefited tremendously from the observed robustness property of the Noah Effects for the ON- and OFF-periods and from the availability of graphical tools that allow for effective visualization of complex data structures.

The summary plots in Figure 6 were obtained by checking, for each of the 181 source-destination pairs, for the presence or absence of the Noah Effect in their corresponding sequences of ON- and OFF-periods that were generated using a variety of different threshold values, typically ranging between 5 s and 0.01 s, depending on the observed traffic. Instead of determining exact point estimates for the intensities α of the Noah Effects

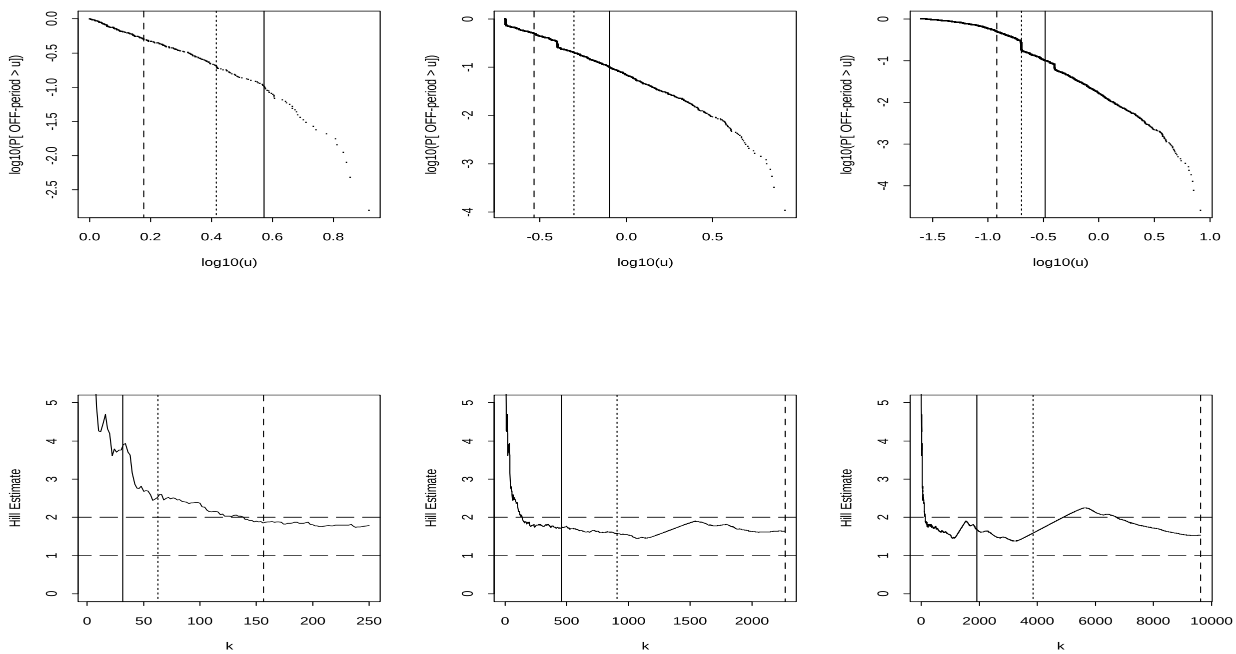

Fig. 4. An illustration of the robustness property of the Noah Effect for the OFF-periods (using source 10). For threshold values $t = 1.0s, 0.20s$ and $0.025s$, the top row gives the complementary distribution plots and the bottom row the corresponding Hill plots.

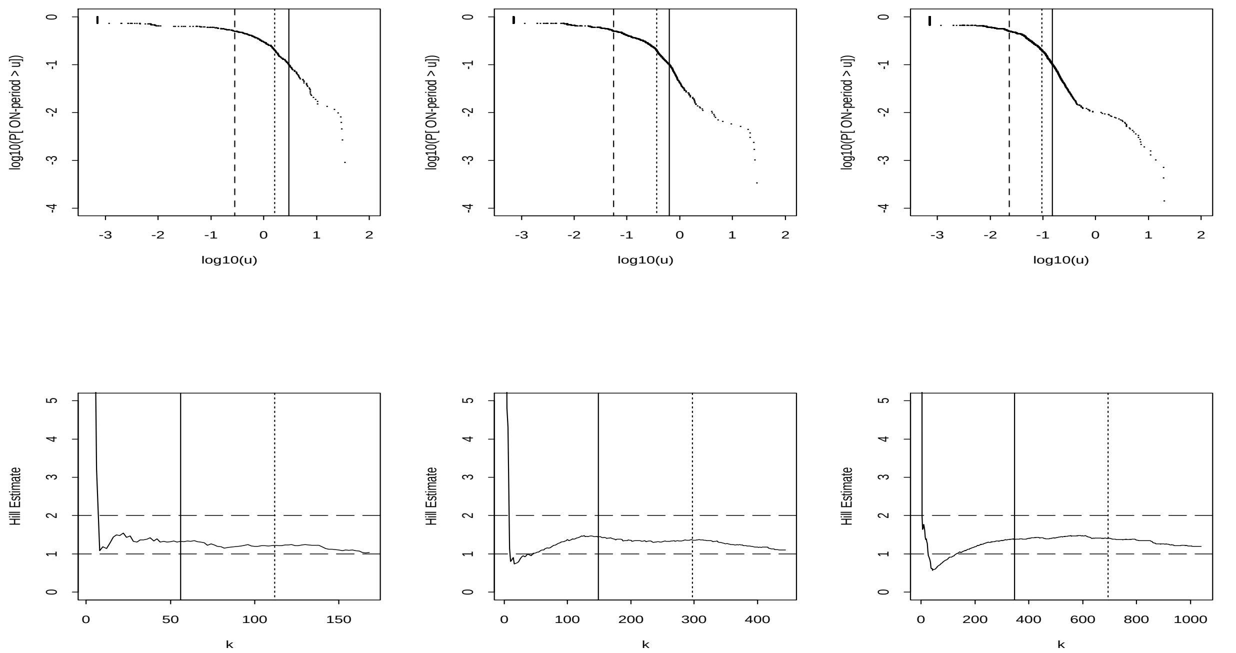

Fig. 5. An illustration of the robustness property of the Noah Effect for the ON-periods (using source 6). For threshold values $t = 1.0s, 0.10s$ and $0.05s$, the top row gives the complementary distribution plots and the bottom row the corresponding Hill plots.

for the ON- and OFF-periods, we marked each source-destination pair, depending on whether the corresponding estimates seemed to be consistent with a range of α-values in $(0, .85), (.75, 1.35), (1.25, 1.75), (1.65, 2.25)$ or $(2.25, 2.75)$, representing the intuitively easy to define cases "definitely below 1.0", "around 1.0", "somewhere in the middle of the interval (1,2)", "around 2.0", and "definitely above 2.0 or inconclusive", respectively. For each source-destination pair considered, the decision was based on a combination of (i) textured plots for visual assessment of ON/OFF nature of the traffic, (ii) complementary distribution plots as a quick heuristic method for checking the tail behavior of a distribution, and (iii) a careful interpretation of Hill plots (typically in connection with information obtained via (ii)). The results are shown in Figure 6, where we plot for each of the 181 source-destination pairs its "load" (in bytes, on log scale)

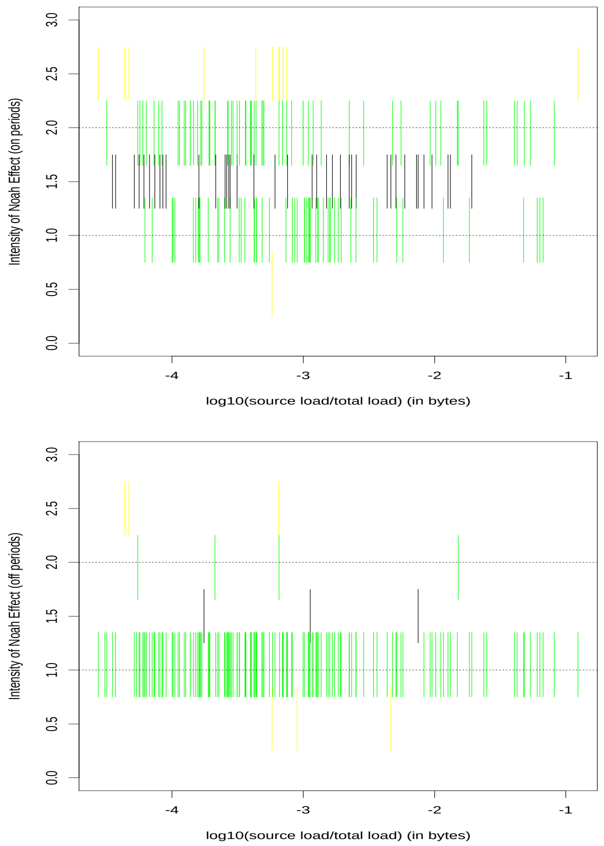

Fig. 6. Summary plot of ranges for the α-estimates for the ON-periods (top) and OFF-periods (bottom) of the 181 most active source-destination pairs, as a function of their "loads".

against the range of α-values that is consistent with its traffic trace. As can be seen, in the case of the ON-periods (top plot), the α-estimates consistent with the data cover pretty much the whole interval $(1, 2)$. In comparison, the bottom plot in Figure 6 shows that in the case of the OFF-periods, α-estimates in the lower part of the interval $(1, 2)$ clearly dominate the picture.

When combined, the two plots in Figure 6 provide strong statistical evidence in favor of our proposed physical explanation for the empirically observed self-similarity property of the aggregate packet process, in terms of the nature of traffic generated by each individual source-destination pair that sent packets over the Ethernet. In particular, our analysis shows that the data at the source-destination level are consistent with an ON/OFF modeling assumption for individual sources or source-destination pairs, and are in strong agreement with the assumption of the Noah Effect for the distributions of the corresponding ON/OFF-periods. (We have also done extensive testing of the independence assumption for the ON-periods and OFF-periods and have found the data to be in full agreement with it.) In fact, one of the most astonishing findings from our analysis has been the extremely widespread and often very obvious presence of the Noah Effect, expressed via relation (1), in measured source-level traffic data, regardless of whether the source represents a fileserver or a client machine. Possible "explanations" for this phenomenon include an empirically observed hyperbolic tail behavior for the file sizes residing in file systems such as file servers (see the discussion and references in [24]), a Pareto-like tail behavior for measured CPU time used by a typical Unix process (see [15]), and a measurements study of an ISDN office automation application reported in [21] that suggests that human-computer interactions occur over a wide range of time scales and thus, may require models based on infinite variance distributions. Finally note that the results in Figure 6 seem to suggest the possibility of different α-values for the ON- and OFF-periods (higher for ON than for OFF), thus pointing toward scenarios that are relevant in practice and for which the mathematical results in Section II should be generalized. While efforts in this direction are already under way, it is important to keep in mind that the crucial piece of information contained in Figure 6 is the strong evidence for the presence of the Noah Effect in the data (see Section II).

F. Self-Similarity and the Noah Effect: 1994 Traffic Traces

Since networks such as Bellcore's various Ethernet LANs are known to often undergo drastic changes within relatively short time periods, the question arises about the relevance of traffic characteristics (e.g., the Joseph and Noah Effects) observed in 5 year-old traffic traces for today's networks. In other words, how sensitive are traffic characteristics such as the Joseph and Noah Effects under changes in network topologies, hardware and software upgrades, changes at the user application level, etc.? Our main objective in including the second data set collected during December 1994 in the present study was to (i) provide insight into these questions and (ii) demonstrate how the information gained from a detailed statistical analysis of data sets of high time-resolution traffic measurements from networks in the past can be used to contribute to a better understanding of the traffic transported on today's networks and to make educated guesses about future network traffic dynamics. Here we briefly discuss some of the findings from our analysis of this second data set, without going into any details; the results are based on an analysis that combines techniques introduced in [16] for analyzing aggregate traffic streams and methods illustrated earlier in this section for dealing with individual source-destination traffic traces.

Given the information provided in Section III.A about the second data set, we first split the hour-long trace into 2 subsets; the first subset represents the traffic sent from the machine that furnishes *Mbone* traffic to Bellcore. It makes up about 52% of the total traffic (in bytes). The second subset consists of the remaining, i.e., all *non-Mbone* packets recorded during the given hour. Concentrating first on the *non-Mbone* traffic, an analysis along the lines of [16] of the aggregate traffic (number of packets per 10 milliseconds) reveals that this data set is consistent with second-order asymptotic self-similarity, and as an estimate of the degree of the Joseph Effect (i.e., the Hurst parameter), we obtain an H-value between 0.85 and 0.90. After separating the aggregate traffic into traffic traces generated by individual source-destination pairs, we get that the 300 most active (out of a total of about 10,000) source-destination pairs are responsible for 83% of the *non-Mbone* traffic. Analyzing these 300 traces in the same way as above shows again consistency of the data with the ON/OFF source modeling assumption. However, in contrast to the first data set, our findings strongly suggest different intensities for the Noah Effects for the ON- and OFF-periods, namely α-values between 1.0 and 1.5 for the OFF-periods, and α-values around 2.0 (i.e., on the borderline between finite and infinite variance) for the ON-periods. In light of the remarks in Section II, this observation suggests that when moving from the internal Ethernet traffic (first data set) to traffic that consists entirely of remote packets, the empirically observed self-similarity property of the aggregate packet stream is mainly due to the relatively strong presence of the Noah Effect in the OFF-periods of the individual source-destination pairs. However, with the ever increasing demand for services such as WWW, it is reasonable to expect that in the future, the intensity of the Noah Effect for the ON-periods corresponding to the individual source-destination pairs to increase (resulting in α-values that move from around 2.0 toward 1.0); as a result, future aggregate *non-Mbone* traffic can be expected to move closer toward an exactly self-similar setting.

In the case of the *Mbone* traffic data set, only an analysis of the aggregate packet stream was performed. The results indicate that *Mbone* traffic is asymptotically self-similar, with an H-value in the high 0.9 range. Its distinctive feature, however, is that only after aggregation levels beyond 100 ms does the strong intensity of the Joseph Effect become obvious, i.e., does the correlation structure of *Mbone* traffic remain unchanged as aggregation levels increase further (WAN traces analyzed in [24] exhibit similar properties). Based on our current understanding (see Section II) and on extensive simulation studies (see Section IV and, especially [25]), this property of *Mbone* traffic suggests the absence of the Noah Effect for the ON-periods (i.e., a corresponding intensity level for the ON-period α that exceeds 2.0) and at the same time, the presence of a very strong (i.e., α-values close to 1.0) intensity of the Noah Effect for the OFF- periods of the individual user applications that typically run over *Mbone*. On an intuitive basis, this observation agrees with the facts that *Mbone* uses compression on the video and audio streams it carries, and that the compression algorithms in use are relatively unsophisticated, resulting in packet bursts separated by comparatively large idle periods. As before, it is reasonable to speculate that as *Mbone* will use more sophisticated compression schemes and will provide a more heterogeneous mix of applications, the Noah Effect of the resulting ON-periods will intensify, while the Noah Effect for its OFF-periods is likely to become less pronounced. In turn, *Mbone* traffic of the future is likely to remain self-similar, possibly over a wider range of time scales than today.

IV. Implications of the Noah Effect in Practice

Recall that the empirically observed self-similarity property in measured network traffic allows for a clear distinction – on statistical grounds – between traditional traffic models and actual traffic collected from working networks. The proposed physical explanation based on the Noah Effect enables us to phrase the essential difference between self-similar and traditional traffic modeling in the setting of the well-known ON/OFF source models. To wit, traditional traffic modeling assumes finite variance distributions for the ON- and OFF-periods (in fact, exponential or geometric distributions are used almost exclusively), while self-similar modeling is based on the assumption of the Noah Effect, i.e., requires infinite variance distributions. Moreover, traditional traffic modeling becomes a special case of the self-similar approach by chosing α-values bigger than 2.0. From a more applied viewpoint, questions related to the impact of self-similarity in practice (e.g., generating realistic network traffic, performance of networks, protocols, and controls) can be reduced to the more basic question of the practical implications of the Noah Effect. In this section, we illustrate its impact with examples concerning traffic modeling, synthetic traffic generation, and network performance analysis.

A. Traffic Modeling and Generation

There is no question that today's network traffic is complex. Often, this is interpreted as saying that a mathematical model of this traffic must be complicated in nature, i.e., must be highly parameterized in order to realistically account for the observed complexity. One of the main results of this paper is that although network traffic is intrinsicly complex, *parsimonious modeling* is still possible; even more so, we demonstrate that it gives rise to a physical explanation for the self-similarity phenomenon that is simple and fully consistent with actual traffic measurements. Thus, for aggregate traffic measurements, insistence on parsimonious modeling has lead to the use of self-similar (or long-range dependent) processes for traffic modeling at the aggregate level. In this paper, the desire for a "phenomenological" explanation of self-similarity in network traffic has resulted in new insights into the nature of traffic generated by the individual sources that contributed to the aggregate stream. We identified the Noah Effect as one of the essential ingredients of these new insights, thus reducing the problem of accurately and realistically modeling traffic in today's networks to estimating a single parameter, namely the intensity α of the Noah Effect in the ON- and OFF-periods of a "typical" network host. We have also shown that while this result (i.e., Theorem 1) holds, strictly speaking, only in an idealized setting (e.g., i.i.d. ON- and OFF-periods, one source type), generalizations accommodating more realistic conditions are possible (see the discussion in Section II), maintain the simplicity of the basic result, and may require the addition of only a small number of physically meaningful parameters.

Explaining, and hence modeling self-similar phenomena in the traffic context in terms of the superposition of many ON/OFF sources with infinite variance distribution for the lengths of their ON/OFF-periods, leads to a straightforward method for generating long traces of self-similar traffic within reasonable (i.e., linear) time – assuming a parallel computing environment. Indeed, the results (e.g., Theorem 1) are tailor-made for parallel computing: letting every processor of a parallel machine generate traffic according to an ON/OFF model (same α), simply adding (i.e., aggregating) 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 7 shows the result of a simulation where we used this method (with alternating rather than i.i.d. ON/OFF periods – see Section II) 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). More precisely, our objective here was to experimentally "verify" Theorem 1 in the context of the August 1989 traffic measurements, i.e.,

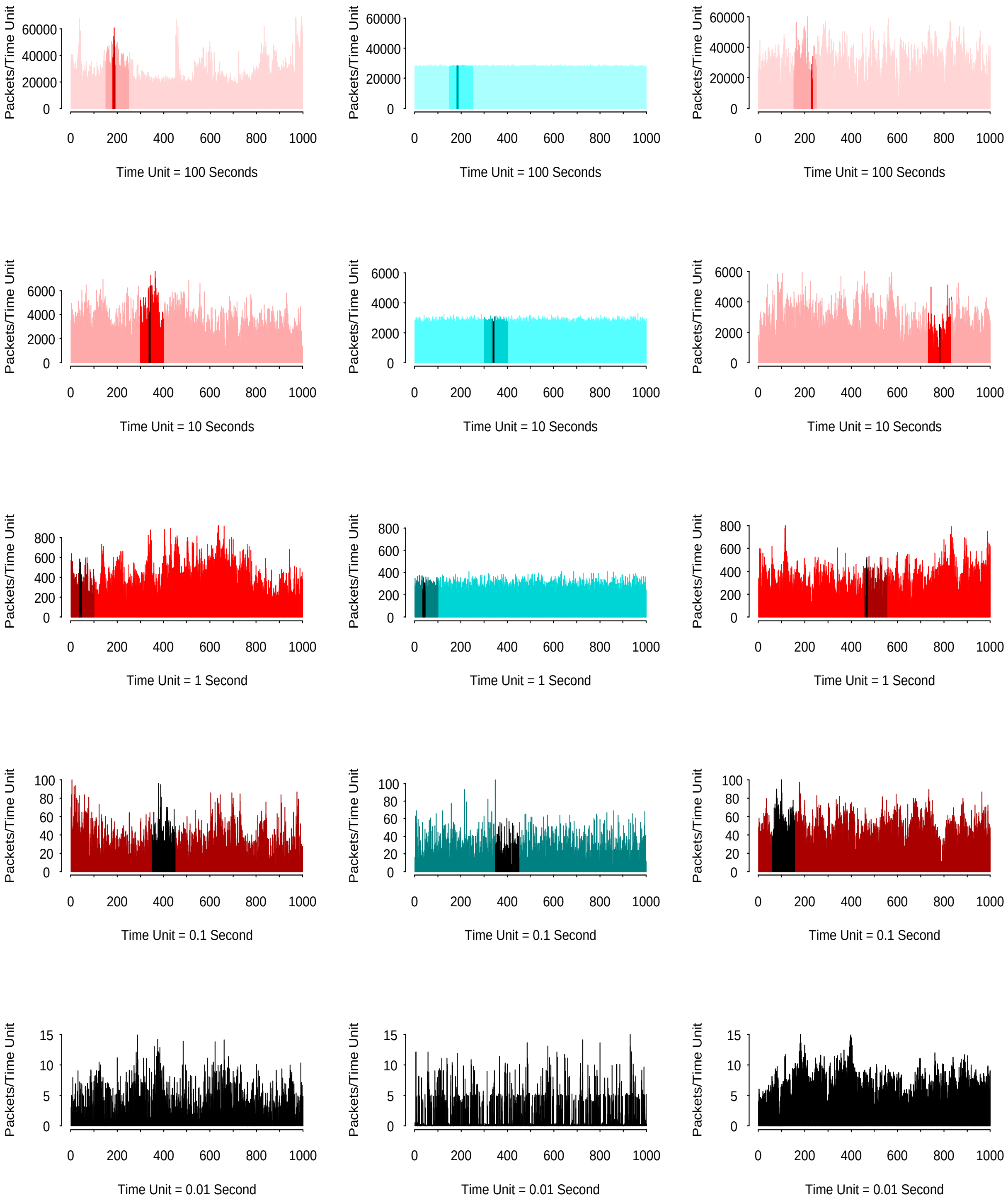

Fig. 7. Actual Ethernet traffic (left column), synthetic trace generated from an appropriately chosen traditional traffic model (middle column), and synthetic trace generated from an appropriately chosen self-similar traffic model with a single parameter (right column) – on five different time scales. Different gray levels indicate the same segments of traffic on the different time scales.

to choose $\alpha = 1.2$ (corresponding to the estimated Hurst parameter of $H = 0.9$ that is consistent with the August 1989 data set) and $M = 500$ (number of processors used should correspond roughly to the number of active source-destination pairs during the observed period), and to check whether or not the resulting synthetic traffic trace "looks like" actual Ethernet traffic as measured in August 1989. We plot in Figure 7 (right most column) the synthetic trace on 5 different time scales, the same way it was done in [17], the original traffic measurements (left most column), and a synthetic trace (middle column) generated from an appropriately matched batch Poisson process (the latter was taken as representative of traditional traffic modeling). As can be seen, our synthetic traffic passes the "visual" test easily, with the possible exception of the plot in the top row (the effect of the daily cycle in the 27 hour trace of Ethernet traffic on the 100 s 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 (not shown here), the trace also fits the data well in a statistical sense, i.e., the estimated Hurst parameter matches the one from the data. Similarly striking agreement between synthetically generated traffic and actual Ethernet traces was obtained in a number of different scenarios, e.g., chosing $M = 16,000$ (close to the total number of processors on the MasPar machine), allowing for different source types, selecting different α-values for the ON- and OFF-period distributions (including different combinations of finite/infinite variance scenarios), and generating under the i.i.d. and alternating renewal assumptions, respectively (see Section II). Recall, that the Ethernet-like behavior of the synthetically generated trace has been accomplished *with only one parameter*, namely the intensity α of the Noah Effect for the ON/OFF-periods of the traffic generated by a "typical" user – parsimonious modeling at its best, and proof that today's complex network traffic dynamics can be modeled in a simple manner without requiring highly parameterized mathematical models. For a more refined modeling of the trace, 2-3 additional parameters may be necessary.

B. Performance and Protocol analysis

The practical benefits of parsimonious modeling of measured network traffic become especially apparent when focusing on the potential impacts of traffic characteristics such as the Joseph and Noah Effects on queueing and network performance, protocol analysis, and network congestion controls. Clearly, the appeal lies in the small number of physically meaningful parameters whose practical impacts need to be investigated. Starting with the empirical finding of self-similarity in Ethernet LAN traffic data reported in [16], there has been mounting evidence for the practical importance of the Hurst parameter H for traffic engineering purposes. In particular, work in [22], [5] (see also [7], [2]) demonstrates a significant difference in queueing performance (expressed in terms of the queue length distribution) between traditional (Markovian) traffic models and those exhibiting the Joseph Effect. More specifically, while the queue length distribution of the former decreases exponentially fast, that of the latter decreases much more slowly (depending on the intensity H of the Joseph Effect), namely like the tail of a Weibull distribution. In practice, not accounting for the Joseph effect at the modeling stage can lead to overly optimistic performance predictions and thus to quality-of-service requirements that are impossible to guarantee in a realistic network scenario. This observation is of particular importance in the context of the widely used concept of *equivalent bandwidth* [6]. At the same time, the presence of the Joseph Effect in measured traffic still allows for economies of scale (i.e., statistical multiplexing gains) via multiplexing a large number of such sources (see [4]).

In view of our physical explanation that the Joseph Effect in aggregate traffic streams is caused by the Noah Effect in the individual ON/OFF sources that generate the aggregate stream, understanding the impacts of the Noah Effect in simple ON/OFF source models on queueing performance becomes essential and is likely to provide valuable new insights into questions related to the design of efficient protocols and effective controls for realistic network traffic. In fact, work is already under way that provides such new insights (e.g., see [25], [2]). For example, investigating the queue length distribution for ON/OFF traffic that exhibits the Noah Effect (either directly or via a corresponding M/G/1 model), these authors show that the Noah Effect gives rise to an infinite mean waiting time, i.e., to queue length distributions that themselves exhibit the Noah Effect and decrease even slower than a Weibull distribution (i.e., the corresponding distribution obtained when multiplexing many such ON/OFF sources). Clearly, this is bad news from the point of view of trying to keep the traffic generated by individual sources isolated from other traffic as far into the network as possible: the resulting buffer requirements at each network node and the ensuing potential delays will be overwhelming. Similarly, traffic shaping at the source may not be feasible in practice due to the naturally occurring large ON-periods for these sources, which in turn would require huge buckets and thus give rise to unreasonably large delays. On the other hand, the results strongly suggest the idea of statistically multiplexing a large number of sources that exhibit the Noah Effect at the earliest possible stage in the network. By doing so, the theory predicts smaller buffer requirements for the network elements and hence smaller packet delays. The simulation results presented in [25] also suggest a wide range of possibilities for protocols and controls for dealing with traffic scenarios that consist of sources with different combinations of infinite variance/finite variance ON- and OFF-periods. For example, protocol design should be expected to be sensitive to and take into account knowledge about network traffic such as the presence or absence of the Noah Effect in a "typical" traffic source. However, how to effectively design protocols that take such information into account remains largely an open issue. Similarly, congestion control schemes that incorporate information about the presence or absence of the Noah Effect at the source level and the Joseph Effect at the aggregate level have yet to be proposed and investigated.

V. Conclusion

Traditional ON/OFF source models typically assume exponential or geometric distributions for their ON- and OFF-periods (or more generally, finite variance distributions). These models are widely used and are especially popular with queueing and performance analysts because of their analytic tractability. However, in recent years, it has been recognized that multiplexing a large number of these sources results in aggregate traffic that is inconsistent with traffic measurements from working networks. On the other hand, Jain and Routhier's packet train models [12], which arose directly from traffic measurement studies, were criticized because of their lack of a clear definition of a "train", their lack of suggestions for chosing the crucial model parameters, and their lack of a physical interpretation. Motivated by the desire to provide a physical explanation for the empirically observed self-similarity property in actual network traffic, we propose in this paper to expand the range of traditional traffic modeling at the level of individual sources to account for the Noah Effect, i.e., for the ability of individual sources to exhibit characteristics that cover a wide range of time scales ("high-variability sources"). By doing so, the criticisms for the ON/OFF source model as well as for the packet train model are deflected. Our results in Section II show that the superposition of many ON/OFF models,, each of which exhibits the Noah Effect, results in aggregate packet streams that are consistent with measured network traffic, and exhibits the same self-similar or fractal properties as can be observed in the data. Moreover, our statistical analysis in Section III confirms the presence of the Noah Effect in measured Ethernet LAN traffic at the source level, and demonstrates an appealing robustness property that renders the stated objections against packet train source models irrelevant.

By (i) reducing the self-similarity phenomenon for aggregate traffic streams to properties of the individual traffic components that make up the aggregate stream, (ii) expressing essential difference between traditional and self-similar traffic modeling in the context of the well-known ON/OFF source models, and (iii) identifying the Noah Effect as the main point of departure from traditional to self-similar traffic modeling, we hope to facilitate the acceptance of self-similar traffic models as viable and practically relevant alternatives to traditional models. The benefits for doing so are immediate and include parsimonious and physically meaningful models for the seemingly very complex traffic dynamics in today's networks, and new insights into problems related to the performance and analysis of protocols and network controls. We have discussed in Section IV some of the mounting evidence for the practical importance of the Noah and Joseph Effects for network engineering. We expect these empirically observed traffic characteristics to play an increasingly important role in the traffic modeling and network performance work for tomorrow's high-speed networks.

Acknowledgment

We gratefully acknowledge the many helpful comments and specific suggestions of Vern Paxson.

References

[1] T. Berners-Lee, R. Cailliau, A. Loutonen, H. F. Nielsen and A. Secret. The World-Wide Web. *Communications of the ACM*, Vol. 37, pp. 76–82, 1994.

[2] F. Brichet, J. Roberts, A. Simonian, and D. Veitch. Heavy traffic analysis of a fluid queue fed by on/off sources with long-range dependence. Preprint, 1995.

[3] P. Danzig, S. Jamin, R. Cáceres, D. Mitzel and D. Estrin. An Empirical Workload Model for Driving Wide-Area TCP/IP Network Simulations. *Internetworking: Research and Experience*, Vol. 3, pp. 1–26, 1992.

[4] N. G. Duffield. Economies of scale in queues with sources having power-law large deviation scalings. Preprint, 1994.

[5] N. G. Duffield and N. O'Connell. Large deviations and overflow probabilities for the general single-server queue, with applications. *Proc. Cambridge Phil. Soc.*, 1995 (to appear).

[6] A. I. Elwalid and D. Mitra. Effective bandwidths of general Markovian sources and admission control of high speed networks. *IEEE/ACM Transactions on Networking*, Vol. 1, pp. 329–343, 1993.

[7] A. Erramilli, O. Narayan, and W. Willinger. Experimental queueing analysis with long-range dependent traffic. Preprint, 1994.

[8] R. Gusella. A characterization of the variability of packet arrival processes in workstation networks. Ph.D. dissertation, Univ. of California, Berkeley, 1990.

[9] R. Gusella. A measurement study of diskless workstation traffic on an Ethernet. *IEEE Transactions on Communications*, Vol. 38, pp. 1557–1568, 1990.

[10] R. Gusella. Characterizing the variability of arrival processes with indexes of dispersion. *IEEE Journal on Selected Areas in Communications*, Vol. 9, pp. 203–211, 1991.

[11] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, Vol. 3, pp. 1163–1174, 1975.

[12] R. Jain and S. A. Routhier. Packet trains: Measurements and a new model for Computer network traffic. *IEEE Journal on Selected Areas in Communications*, Vol. 4, pp. 986–995, 1986.

[13] S. M. Klivansky, A. Mukherjee, and C. Song. On Long-Range Dependence in NSFNET Traffic. Preprint, 1994.

[14] M. F. Kratz and S. I. Resnick. The QQ-Estimator and Heavy Tails. Preprint, 1995.

[15] W. E. Leland and T. J. Ott. Unix process behavior and load balancing among loosely-coupled computers. In O. J. Boxma, J.—W. Cohen, and H. C. Tijms, editors, *Teletraffic Analysis and Computer Performance Evaluation*, pp. 191–208, Amsterdam, 1986, Elsevier Science Publishers B. V.

[16] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic. *ACM/SIGCOMM Computer Communications Review*, Vol. 23, pp. 183–193, 1993. Proceedings of the ACM/SIGCOMM'93, San Francisco, September 1993.

[17] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic (Extended Version). *IEEE/ACM Transactions on Networking*, Vol. 2, pp. 1–15, 1994.

[18] H. Eriksson. Mbone: The Multicast Backbone. *Communications of the ACM*, Vol. 37, pp. 54–60, 1994.

[19] B. B. Mandelbrot. Long-run linearity, locally Gaussian processes, H-spectra and infinite variances. *International Economic Review*, Vol. 10, pp. 82–113, 1969.

[20] B. B. Mandelbrot. *The Fractal Geometry of Nature*. Freeman, New York, 1983.

[21] K. Meier-Hellstern, P. E. Wirth, Y.-L. Yan, and D. A. Hoeflin. Traffic models for ISDN data users: Office Automation application. In A. Jensen and V. B. Iversen, editors, *Teletraffic and Datatraffic in a Period of Change, Proc. of ITC13, Copenhagen*, pp. 167–172, Amsterdam, 1991, Elsevier Science Publishers B. V.

[22] I. Norros. A storage model with self-similar input. *Queueing Systems*, Vol. 16, pp. 387–396, 1994.

[23] V. Paxson. Growth Trends in Wide-Area TCP Connections. *IEEE Network*, Vol. 8, pp. 8–17, 1994.

[24] V. Paxson and S. Floyd. Wide-area traffic: The failure of Poisson modeling. *Proceedings of the ACM/SIGCOMM'94*, pp. 257–268, 1994.

[25] P. Pruthi and A. Erramilli. Heavy-tailed on/off source behavior and self-similar traffic. *Proc. IEEE ICC'95*, Seattle, June 1995.

[26] S. I. Resnick and C. Starcia. Smoothing the Hill Estimator. Preprint, 1995.

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

[28] D. F. Swayne, D. Cook, and A. Buja. XGobi: Interactive Dynamic Graphics in the X Window System with a Link to S. *1991 Proceedings of the Section on Statistical Graphics*, pp. 1–8, 1991.

[29] M. S. Taqqu and J. Levy. Using renewal processes to generate long-range dependence and high variability. In E. Eberlein and M. S. Taqqu, editors, *Dependence in Probability and Statistics*, pp. 73–89, Boston, 1986. Birkhäuser.

[30] J. W. Tukey and P. A. Tukey. Strips displaying empirical distributions: I. Textured Dot Strips. *Bellcore Technical Memorandum*, 1990.