Wide-Area Traffic: The Failure of Poisson Modeling

Vern Paxson and Sally Floyd
Lawrence Berkeley Laboratory and
EECS Division, University of California, Berkeley
1 Cyclotron Road, Berkeley CA 94720
vern@ee.lbl.gov, floyd@ee.lbl.gov

To appear in SIGCOMM '94

Abstract

Network arrivals are often modeled as Poisson processes for analytic simplicity, even though a number of traffic studies have shown that packet interarrivals are not exponentially distributed. We evaluate fifteen wide-area traces, investigating a number of wide-area TCP arrival processes (session and connection arrivals, FTPDATA connection arrivals within FTP sessions, and TELNET packet arrivals) to determine the error introduced by modeling them using Poisson processes. We find that user-initiated TCP session arrivals, such as remote-login and file-transfer, are well-modeled as Poisson processes with fixed hourly rates, but that other connection arrivals are less persuasively Poisson; that modeling TELNET packet interarrivals as exponential grievously underestimates the burstiness of TELNET traffic, but using the empirical Tcplib[DJCME92] interarrivals preserves burstiness over many time scales; and that FTPDATA connection arrivals within FTP sessions come bunched into "connection bursts", the largest of which are so large that they completely dominate FTPDATA traffic. Finally, we offer some speculations regarding how our findings relate to the possible *self-similarity* of wide-area traffic.

1 Introduction

When modeling network traffic, packet and connection arrivals are often assumed to be Poisson processes because such processes have attractive theoretical properties. A number of studies have shown, however, that for both local-area and wide-area network traffic, the distribution of packet interarrivals clearly differs from exponential [JR86, Gusella90, FL91, DJCME92]. Recent work argues convincingly that LAN traffic is much better modeled using statistically *self-similar* processes [LTWW93], which have much different theoretical properties than Poisson processes. For self-similar traffic, there is no natural length for a "burst"; traffic bursts appear on a wide range of time scales. In this paper we show that for wide-area traffic Poisson processes are valid only for modeling the arrival of user sessions (TELNET connections, FTP control connections); that they fail as accurate models for other WAN arrival processes; and that WAN arrival processes appear much better modeled using self-similar processes.

For our study we analyze fifteen traces of wide-area TCP traffic. We consider both previous and new models of aspects of FTP and TELNET traffic, discuss the implications of these models for burstiness at different time scales, and compare the results of the models with the trace data. We show that in some cases commonly-used Poisson models result in serious underestimations of the burstiness of TCP traffic over a wide range of time scales. (We restrict our study to time scales of 0.1 seconds and larger.)

We first show that for interactive TELNET traffic, *connection* arrivals are well-modeled as Poisson with fixed hourly rates. However, the exponentially-distributed interarrivals commonly used to model *packet* arrivals generated by the user side of a TELNET connection grievously underestimate the burstiness of those connections, and high degrees of multiplexing do not help. Using the empirical Tcplib [DJ91, DJCME92] distribution for TELNET packet interarrivals instead results in packet arrival processes significantly burstier than Poisson arrivals, and in close agreement with traces of actual traffic. From these findings we then construct a model of TELNET traffic parameterized by only the hourly connection arrival rate and show that it accurately reflects the burstiness found in actual TELNET traffic. The success with this model of using Tcplib packet interarrivals confirms the finding in [DJCME92] that the arrival pattern of user-generated TELNET packets has an invariant distribution, independent of network details.

For small machine-generated bulk transfers such as SMTP (email) and NNTP (network news), connection arrivals are not well-modeled as Poisson, which is not surprising since both types of connections are machine-initiated and can be timer-driven. Previous research has discussed how the periodicity of machine-generated IP traffic such as routing updates can result in network-wide traffic synchronization [FJ93], a phenomenon impossible with Poisson models.

For large bulk transfer, exemplified by FTP, the traffic structure is quite different than suggested using Poisson models. As with TELNET connections, user-generated FTP session arrivals are well-modeled as Poisson with fixed hourly rates. However, we find that FTPDATA connections within a single FTP session are clustered into bursts. Both FTPDATA connection and FTPDATA burst arrivals are not well-modeled as Poisson processes. Furthermore, one of our key findings is that the distribution of the number of bytes in each burst has a very heavy upper tail; a small fraction of the largest bursts carries almost all of the FTPDATA bytes. This implies that faithful modeling of FTP traffic should concentrate heavily on the characteristics of the largest bursts.

Poisson arrival processes are quite limited in their burstiness, especially when multiplexed to a high degree. Our findings, however, show that wide-area traffic is much burstier than Poisson models predict, over many time scales. This greater burstiness has implications for many aspects of congestion control and traffic performance. We conclude the paper with a discussion of how our burstiness results might mesh with self-similar models of network traffic, and then with a look at the general implications of our results.

2 Traces used

Dataset	Date	Duration	What	Drops
Bellcore (BC)	10Oct89	13 days	TCP S/F*	0
DEC (DEC-1)	26Nov91	24 hours	TCP S/F	?
LBL 1-7	See refs.	30 days	TCP S/F	$\leq 15 \cdot 10^{-6}$
UCB (UCB)	31Oct89	24 hours	TCP S/F*	0
LBL PKT-1	17Dec93	2 hours	All TCP	$5 \cdot 10^{-4}$
LBL PKT-2	19Jan94	2 hours	All TCP	$9 \cdot 10^{-4}$
LBL PKT-3	20Jan94	2 hours	All TCP	$2 \cdot 10^{-4}$
LBL PKT-4	21Jan94	1 hour	All pkts.	$7 \cdot 10^{-4}$
LBL PKT-5	28Jan94	1 hour	All pkts.	$5 \cdot 10^{-4}$

Table 1: Summary of Wide-Area Traces

Table 1 summarizes the traces of wide-area traffic used in our study. The first rows represent traces previously analyzed: the BC and UCB traces in [DJCME92][1], the DEC-1 trace in [P93], and the LBL traces in [P94, P93]. The "LBL 1-7" row represents 7 wide-area TCP SYN/FIN traces, each spanning 30 days. The "drops" column gives the fraction of packets dropped by the trace program during each of the traces, always quite low. The final five rows reflect new traces we gathered for our study. Each of these traces began at 2PM; the first three captured all TCP packets, and lasted two hours. The final two traces captured all packets and lasted one hour.

3 TCP connection interarrivals

This section examines the connection start times for several TCP protocols. The pattern of connection arrivals is dominated by a 24-hour pattern, as has been widely observed before. We show that for TELNET connection arrivals and for FTP session arrivals, within one-hour intervals the arrival process can be well-modeled by a homogeneous Poisson process; each of these arrivals reflects an individual user starting a new session. For NNTP and SMTP, a Poisson model of connection arrivals is questionable. We also show that for individual FTPDATA connections, the arrival process is definitely not Poisson, but, as is discussed later in Section 6, there is noteworthy structure in the arrival process of individual FTPDATA connections within an FTP session.

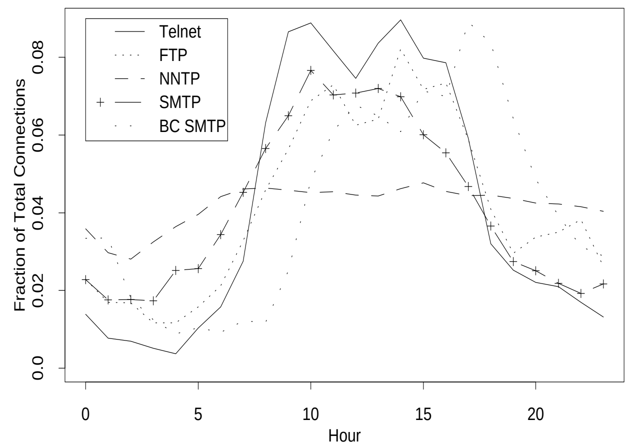

Figure 1: Mean, relative, hourly connection arrival rate for LBL-1 through LBL-4 datasets.

Figure 1 shows the mean hourly connection arrival rate for datasets LBL-1 through LBL-4. For the different protocols, we plot for each hour the fraction of an entire day's connections of that protocol occurring during that hour.[2] For example, TELNET connections occur primarily during normal office hours, with a lunch-related dip at noontime; this pattern has been widely observed before. FTP file transfers have a similar hourly profile, but they show substantial renewal in the evening hours, when presumably users take advantage of lower networking delays. The NNTP traffic maintains a fairly constant rate throughout the day, only dipping somewhat in the early morning hours (but the mean size of each connection varies over the course of the day; see [P93]). The SMTP traffic is interesting because it shows a morning bias for the LBL site (west-coast U.S.) and an afternoon bias for the Bellcore site (east-coast U.S.); perhaps the shift is due to cross-country mail arriving earlier in the Pacific time zone and later in the Atlantic time zone.

Figure 1 shows enough daily variation that we cannot reasonably hope to model connection arrivals using simple homogeneous Poisson processes, which require constant rates. The next simplest model is to postulate that during fixed-length intervals (say, one hour long) the arrival rate is constant and the arrivals within each interval might be well modeled by a homogeneous (fixed-rate) Poisson process. Telephone traffic, for example, is fairly well modeled during one-

[1] These traces captured all WAN packets, but our analysis in this paper uses only the TCP SYN/FIN packets

[2] In Figure 1, FTP refers to FTP sessions.

hour intervals using homogeneous Poisson arrival processes [FL91].

To evaluate these Poisson models, we developed a simple statistical methodology (Appendix A) for testing whether arrivals during a given one-hour or ten-minute period are Poisson with a fixed rate. If the arrivals during the period are truly Poisson, then we would expect 95% of the tested periods to pass the test. Note that we would expect testing ten-minute periods to perhaps be more successful than testing one-hour periods, because using ten-minute periods allows the arrival rate to change six times each hour rather than remaining constant throughout the hour.

We applied our methodology to the LBL-5 and LBL-6 datasets for TELNET, FTP, FTPDATA, SMTP, and NNTP connections. Here FTP refers to an FTP *session* (i.e., an FTP control connection), while FTPDATA refers to the data-transfer connections spawned by these control connections [3]. We also tested arrivals of FTPDATA *bursts* (see Section 6 below) for the LBL-6 and LBL-7 datasets.

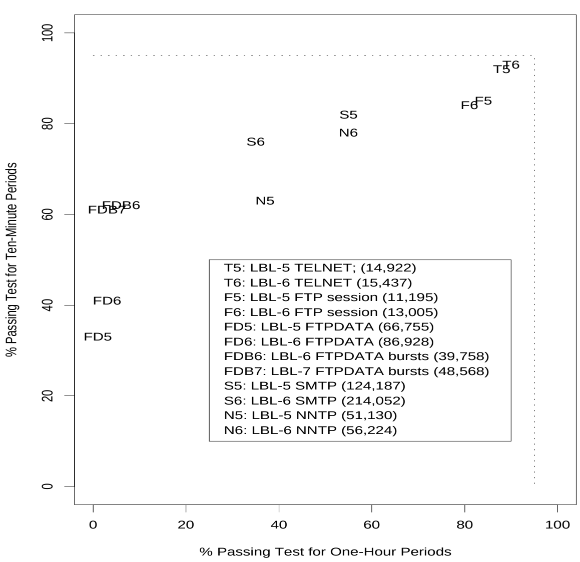

Figure 2: Results of testing for homogeneous Poisson arrivals at the 5% significance level.

Figure 2 shows the results of our tests. Along the x-axis we plot the percentage of tested one-hour intervals that passed the statistical test, and along the y-axis the same for ten-minute intervals. The dashed lines correspond to a 95% pass-rate, indicating arrivals indistinguishable from Poisson. The legend identifies the protocol and dataset corresponding to each of the points, along with the total number of such connections (or bursts) during the dataset.

We see immediately that TELNET connection arrivals are almost exactly Poisson, both for 1-hour and 10-minute fixed rates. FTP session arrivals are also well-modeled as Poisson.

On the other hand, from these tests NNTP and SMTP arrivals appear not well-modeled as Poisson. This finding is not conclusive, however, because (as explained in Appendix A) the higher number of NNTP and (especially) SMTP connections may simply be allowing the statistical test more opportunity to detect relatively insignificant deviations from Poisson arrivals.

That NNTP and SMTP arrivals appear not well-modeled as Poisson is not too surprising. Because of the flooding mechanism used to propagate network news, NNTP connections can immediately spawn secondary connections as new network news is received from one remote peer and in turn offered to another. NNTP and SMTP connections also are often timer-driven. Finally, SMTP connections are perturbed by mailing list explosions in which one connection immediately follows another, and possibly by timer effects due to using the Domain Name Service to locate MX records [Stevens94].

FTPDATA connection arrivals are clearly not Poisson over one-hour periods and only poorly modeled as such over 10-minute periods. This finding can be readily attributed to the fact that "multiple-get" file transfers often result in a rapid succession of FTPDATA connections, one immediately following another [P93]. Coalescing multiple-connection FTPDATA bursts (see Section 6) into single arrivals improves the 10-minute Poisson fit somewhat, but the arrivals still fail to fit as well as even NNTP.

The finding that TELNET connection arrivals are well-modeled as a Poisson process with fixed hourly rates is at odds with that of [MM85], who found that user interarrival times looked "roughly log-normal". We believe the discrepancy is due to characterizing all interarrivals lumped together, rather than postulating separate hourly rates; for example, the distribution of all LBL-5 TELNET interarrivals looks close to log-normal.

Given that TELNET connection arrivals appear Poisson over one-hour intervals, one might imagine that other human-initiated traffic such as RLOGIN and X11 will also fit this model. We find that RLOGIN does and X11 does not. We conjecture that the difference is that during a single X11 *session* (corresponding to running an instance of *xterm*, say) a user initiates multiple X11 connections, while TELNET and RLOGIN sessions are comprised of a single TCP connection. Thus, TELNET arrivals correspond to users deciding to *begin* using the network; X11 arrivals correspond to users deciding to do something new *during* their use of the network. The former behavior is likely to be close to uncorrelated, memoryless arrivals, since each arrival generally involves a new user; the latter does not have the memoryless property (and hence is not exponential), since a single user is involved in generating new arrivals. If we could discern between X11 session arrivals and X11 connection arrivals, then we conjecture we would find the session arrivals to be Poisson.

[3] We first removed the periodic "weather-map" FTP traffic discussed in [P94].

4 TELNET packet interarrivals

The previous section shows that start times for TELNET connections are well-modeled by a Poisson process. In this section we look at the packet arrival process within a TELNET connection. We restrict our study to packets generated by the TELNET connection originator; this in general is a user typing at a keyboard. Because these packets are initiated by a human, we might hope that the arrival process is to some degree “invariant”; that is, the process may be independent of network dynamics and instead mainly reflect the delays and bursts of activity associated with people typing commands to a computer. Indeed, our empirical results of the interarrival times between packets in a single TELNET connection are consistent with the empirical Tcplib distribution found by previous researchers. Unlike the exponential distribution, the empirical distribution of TELNET packet interarrival times is *heavy-tailed*; we show that using the exponential distribution results in seriously underestimating the burstiness both of TELNET traffic within a single connection and of multiplexed TELNET traffic. Modeling TELNET packet arrivals by a Poisson process, as is generally done, can result in simulations and analyses that significantly underestimate performance measures such as average packet delay.

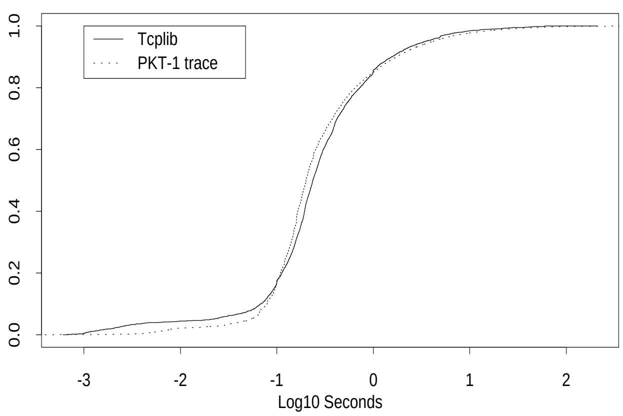

Figure 3: Empirical distributions of packet-interarrivals within TELNET connections.

Figure 3 shows two empirical distributions of the interarrival times of packets within TELNET connections. The solid line shows the distribution used by Tcplib[DJ91, DJCME92]; the dashed line shows the same for the PKT-1 trace. Above 0.1 seconds, the agreement is quite good, especially in the upper tail. That different sites produce the same distribution argues heavily that the distribution is independent of network dynamics and instead reflects human typing dynamics. Below 0.1 seconds the interarrival times probably *are* dominated by network dynamics; but, as stated earlier, in this paper we are not concerned with time scales below 0.1 seconds.

Even ignoring the lower tail, the interarrival distribution is not even close to exponential in shape (note that the x-axis is logarithmically scaled). Rather, the main body of the distribution fits very well to a Pareto distribution (doubly-exponential; see Appendix B) with parameter $\alpha = 0.85$, and the upper 3% tail to a Pareto distribution with $\alpha = 0.95$. Interestingly, a Pareto distribution with $\alpha < 1$ has infinite mean and variance; a very different beast than an exponential distribution.

It is not surprising that interactive packet arrivals do not fit a Poisson model, since earlier work looking at many different components of interactive traffic failed to find any statistically significant exponential fits to the observed distributions [FJ70]. This leaves the question: What are the consequences of using Poisson packet arrivals rather than the empirical Tcplib distribution for TELNET traffic?

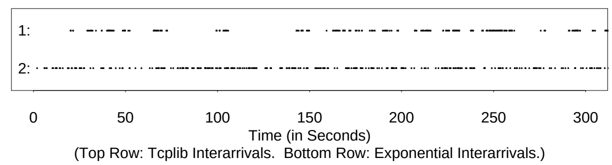

Figure 4: Comparisons of models for TELNET packet interarrival times.

Figure 4 shows packets from two simulated TELNET connections, each lasting 300 seconds. The top line shows Method 1, where packet interarrival times are independently drawn from the Tcplib distribution. The bottom line shows Method 2, where packet interarrival times are independent random variables from an exponential distribution with a mean of 1.1 seconds (the mean of the Tcplib empirical distribution). Both methods generate roughly 300 TELNET packets in the 300-second period. We have plotted a dot for each packet arrival, where the x-axis gives the time of the arrival. As expected, the packets are more clustered with Method 1 than with Method 2.

This difference in burstiness between the two methods persists to some extent for multiplexed TELNET connections. For example, we ran 10-minute simulations with 100 active TELNET connections, where all connections were active for the entire duration of the simulation. In one set of simulations, Method 1 was used to generate packet interarrival times for each connection, and for the other set of simulations Method 2 was used. We found the packet arrival process for Method 1 visibly more bursty than the packet arrival process with Method 2, and can quantify the burstiness as follows. For each method, consider the number of TELNET packets arriving during each one-second interval. For the 600-second simulation using Method 1, this aggregate number had a mean of 92 and a variance of 240; with Method 2, this aggregate number had a mean of 92 and a variance of 97. Even a quite high degree of statistical multiplexing fails to smooth away the difference between the two packet arrival processes.

One of the natural performance measures for TELNET traffic is average packet delay. It would not be hard to construct simulations, one using Tcplib and the other using exponential interarrivals, where making the mistake of using exponential interarrivals instead of Tcplib results in significantly underestimating the average queueing delay for TELNET packets.

The above shows that the Tcplib packet interarrival distribution behaves quite differently than a Poisson process, even in the presence of multiplexing. We now show that for measured network traffic, these differences extend far beyond the time scale of individual packets. To look at the difference in burstiness at different time scales, we first extracted all TELNET originator packets[4] from the two-hour PKT-2 trace. These packets belonged to 277 separate TCP connections. Of these connections, 4 were anomalously large and rapid (more than 2^{10} bytes transferred by the originator at sustained data rates exceeding 8 bytes/sec). These are unlikely to correspond to human typing, were clear outliers, and are probably better modeled as bulk transfer connections. Removing the outliers left us with 273 connections.

We then synthesized several two-hour packet traces as follows. For each of the TELNET connections, we synthesized a connection with the same starting time within the two-hour period and the same size (in packets). One of the synthesized traces used the Tcplib empirical distribution for the packet interarrivals within each connection ("TCPLIB"); one used exponential interarrivals with mean 1.1 ("EXP"); and one uniformly distributed each connection's packet arrivals over the interval between when the connection began and when during the PKT-2 trace the connection terminated ("VAR-EXP"). This last method corresponds to exponential interarrivals with the mean adjusted to reflect the connection's actual observed packet rate. Thus, for the TCPLIB and EXP schemes, we generated connections with the same starting times and packet sizes as their counterparts in the PKT-2 trace, but perhaps with different durations, while with the VAR-EXP scheme, the generated connections shared starting time, packet size, and duration.

A valuable tool for assessing burstiness over different time-scales is the "variance-time plot" [LTWW93, Garrett93], which we describe here by example rather than rigorously. Suppose we have an arrival process consisting of 72,000 observations, corresponding to a two-hour trace viewed every 0.1 seconds. If we are interested in the process's burst-structure on a time scale of 10 seconds, we could construct a "smoothed" version of the process by averaging the first 100 observations to obtain the process's mean value during the first 10 seconds, the next 100 observations for the next 10 seconds, and so on. In general we can do this sort of smoothing for any aggregation level M, where in the previous example $M = 100$.

To construct a variance-time plot, we smooth the process for different values of M and then plot the variance of the smoothed process on the y-axis vs. the aggregation level (M) on the x-axis, using logarithmic scales.

Variance-time plots are useful for gauging burstiness over many different time scales; straight lines on variance-time plots are also suggestive of self-similarity (see Section 7 for further discussion).

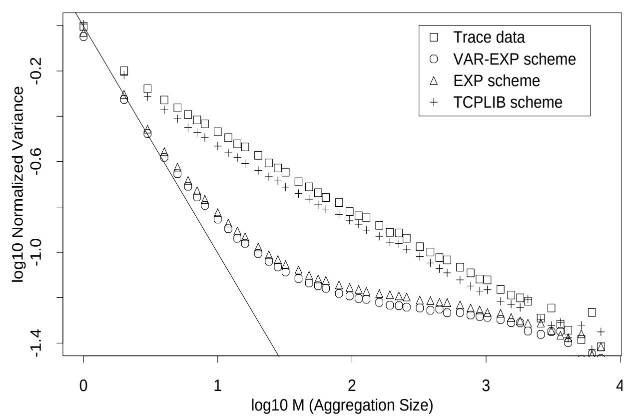

Figure 5: Variance-Time Plot for TELNET packet arrival process. The line from the upper left corner has slope -1 (see Section 7 for significance).

Figure 5 shows such a plot for the PKT-2 TELNET trace and for the three schemes discussed above. Here the unaggregated process ($M = 1$) corresponds to 72,000 observations of the number of TELNET originator packets arriving during 0.1-second intervals. The y-axis is the variance of the aggregated process normalized by dividing by the square of the average number of packets per 0.1-second. This normalization allows us to compare the variance of processes with different numbers of arrivals.[5]

From the plot it is immediately clear that the variance of the TCPLIB scheme agrees closely with the PKT-2 trace data, while both EXP and VAR-EXP exhibit far less variance, indicating they are much less bursty.

Figure 6 shows this explicitly. Here we plot the arrival process corresponding to 5-second intervals ($M = 50$) for the PKT-2 trace and for the EXP trace. The x-axis shows the time in seconds, and the y-axis shows the total number of TELNET packets in each 5-second interval. The average number of packets in the two traces are similar; the PKT-2 trace has an average of 59 packets in each 5-second interval, and the fixed-rate exponential trace has an average of 57 packets in each 5-second interval. The variances, however, are quite different. With 5-second bins, the PKT-2 trace has a variance of 672, while the exponential trace has a variance of 260.

Clearly, this difference in the packet-generation rate over 5-second intervals could have consequences for queueing delays in simulations using these two different traces. As the variance-time plot shows, the PKT-2 trace is more bursty over many time intervals, not only over the five-second intervals shown here. The conclusions are that using exponential packet interarrival times for TELNET connections results in substantial underestimations of the burstiness of multiplexed

[4]Except for "pure ack" packets, containing no user data.

[5]The traces consisted of between 82,500 and 86,000 packets.

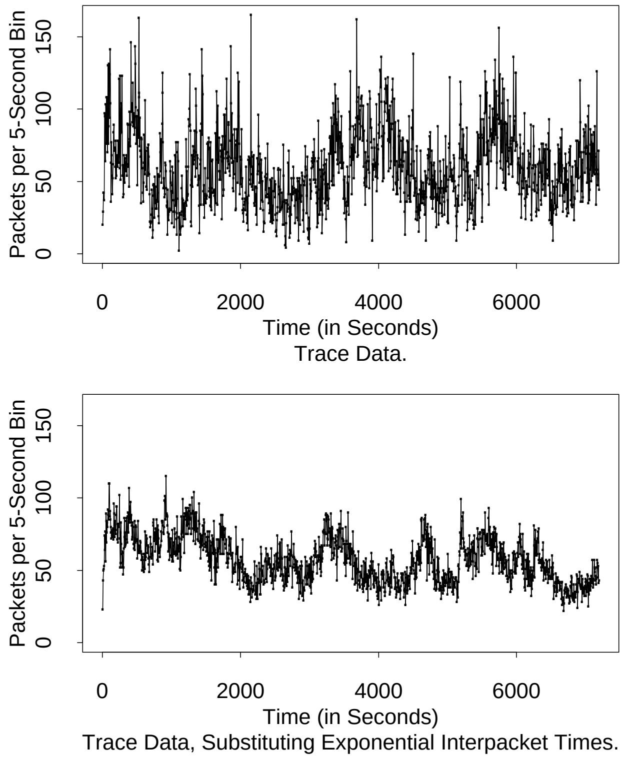

Figure 6: Comparisons of actual and exponential TELNET packet interarrival times.

TELNET traffic, but using i.i.d. interarrivals drawn from the Tcplib distribution faithfully reproduces the burst structure.

5 Fully modeling TELNET originator traffic

Section 3 has shown that over 1-hour periods, TELNET connection arrivals are well-modeled as Poisson processes, and Section 4 has shown that within a TELNET connection, packet interarrival times can be modeled using the heavy-tailed distribution in Tcplib. The connection size in *bytes* has previously been modeled by a log-extreme distribution [P93]; the distribution of the connection size in *packets* is somewhat different, and seems to be better modeled by a log-normal distribution (see below). In this section, we put these three pieces together to construct a complete model of TELNET originator traffic that is parameterized only by the connection arrival-rate. Variance-time plots show that this model corresponds well to empirical measurements, and suggest that both the model and the empirical traces could be self-similar, with burstiness persisting over a wide range of time scales.

First, we look at the difference in the distributions of originator bytes per connection vs. originator packets. Previous work reports that the number of bytes sent by the originator in a wide-area TELNET connection is well-modeled using a log-*extreme* distribution with parameters $\alpha = \log_2 100$ and $\beta = \log_2 3.5$ [P93]. We experimented with using this distribution to produce sizes for an equal number of TELNET connections as appeared in the PKT-2 trace. We found that the distribution consistently generates connection sizes (in bytes) much larger than the connection sizes (in packets) observed in the trace. We attribute this difference to two effects:

- The [P93] fit was made using month-long traces of TELNET connections, allowing for much longer and larger connections than present in our two-hour trace;
- The [P93] fit models connection size in *bytes* and not in *packets*. One generally assumes that each TELNET originator packet conveys one byte of user data, corresponding to a keystroke. Often, however, a packet carries more than one byte, either due to effects of the Nagle algorithm or because the TELNET connection is operating in "line mode" [Stevens94]. For example, the PKT-2 TELNET originator traffic comprised about 85,000 packets carrying 139,000 user data bytes.

Given these difficulties, we attempted to fit the observed TELNET connection sizes (in packets) with another simple analytic distribution. We found that a $\log_2$-normal distribution with $\log_2$-mean $\bar{x} = \log_2 100$ and $\log_2$-standard deviation $\sigma = 2.24$ fit the connection size in packets well visually[6], considerably better than a log-extreme distribution with parameters fitted to the data.[7]

Putting all of this together, we have a complete model for TELNET traffic, FULL-TEL-MODEL, parameterized only by the TELNET connection arrival rate. FULL-TEL-MODEL uses Poisson connection arrivals, log-normal connection sizes (in packets), and Tcplib packet interarrivals.

We then used FULL-TEL-MODEL to generate three synthetic traces of TELNET originator traffic, using a connection arrival rate of 273 connections in 2 hours. Because such traces start off with no traffic and build up to a steady-state corresponding to the connection arrival rate, we trimmed the traces to just their second hour. We then used variance-time plots to compare the traces with the second hour of the PKT-2 TELNET trace.

Figure 7 shows the results of the comparison. In general the agreement is quite good, though the models have slightly higher variance than the trace data for $M > 10^2$. We conclude that FULL-TEL-MODEL faithfully captures TELNET originator traffic, except to be a bit burstier on time scales above 10 seconds.[8]

[6] The exact numerical values of $\bar{x}$ and σ should not be taken too seriously, as they came from a small sample.

[7] We also found that a log-extreme distribution fit the connection size in bytes better than a log-normal distribution.

[8] We also tested the model's fit to the PKT-1 and PKT-3 TELNET traces; the results were similar.

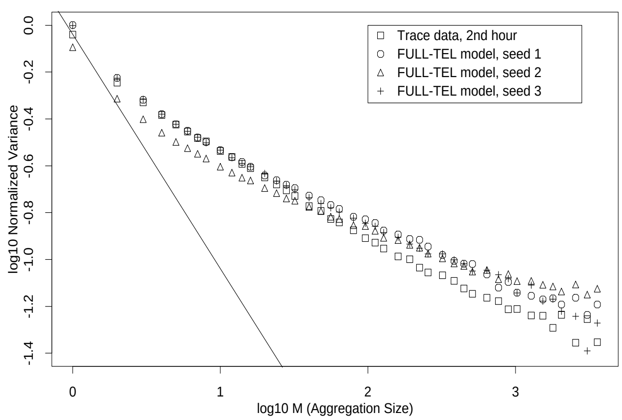

Figure 7: Variance-time plot comparing PKT-2 trace data with the complete TELNET model.

6 FTPDATA connection arrivals

This section investigates arrival processes for FTP traffic. Modeling FTP is particularly important because FTPDATA connections currently carry the bulk of the data bytes in wide area networks ([CBP93]). Section 3 showed that while FTP session arrivals can be modeled as Poisson processes, this is not the case for FTPDATA connection arrivals. This section shows that FTPDATA connections within a session are clustered in bursts, and that the distribution of burst sizes in bytes is quite heavy-tailed; roughly half of the FTP traffic volume comes from the largest 0.5% of the FTPDATA bursts. These large bursts are likely to completely dominate FTP traffic dynamics.

In this paper, we do not attempt to model FTPDATA packet arrivals within a connection. Unlike TELNET connections, where the originator packet arrival process is largely determined by packet generation pattern at the source, the packet arrival process for an FTPDATA connection is largely determined by network factors such as the available bandwidth, congestion, and details of the transport-protocol congestion control algorithms. Previous studies have found that FTPDATA packet interarrivals are far from exponential [DJCME92]; this is not surprising, since the above network factors lead to a process quite different from memoryless arrivals.

To begin, Section 3 shows that FTPDATA connection arrivals are not well-modeled as Poisson, even if we aggregate closely-spaced connections into single bursts. Each FTP session spawns a number of FTPDATA connections; one key question is how these connections are distributed within the duration of the FTP session.

We computed the distribution of spacing between FTPDATA connections spawned by the same FTP session for five datasets: LBL-1, LBL-6, LBL-7, DEC-1, and UCB. Here, "spacing" refers to the amount of time between the end of one FTPDATA connection within a session and the beginning of the next. In each case the upper tail of the distribution is much heavier than exponential, and is better approximated using a log-normal or log-logistic distribution.

When plotted using a log-scale, all of the distributions showed inflection points at spacings between 2 and 6 seconds. We conjecture that spacings shorter than these points reflect sequential FTPDATA connections due to multiple transfers (the FTP "mget" command) or a user issuing a "list directory command" very shortly followed by a "get". Thus spacings of less than these values might well be interpreted as sequential FTPDATA connections corresponding to a single "burst" of file-transfer activity. We somewhat arbitrarily chose a spacing of ≤ 4 seconds as defining connections belonging to the same *burst*, and we note that such spacings are not inordinately larger than the 1-2 second spacings that can occur internal to a single FTPDATA connection due to packet loss timeouts.

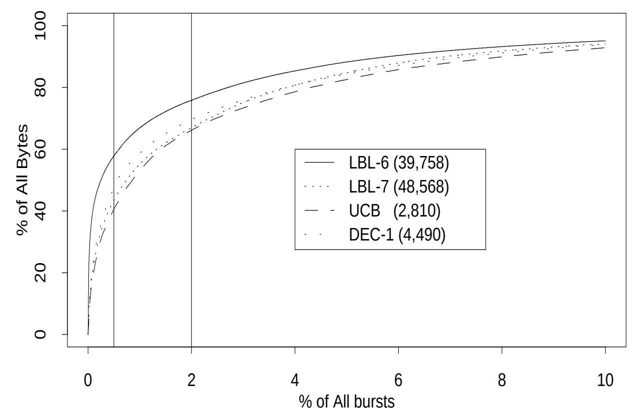

Figure 8: Percentage of all FTPDATA bytes due to largest 10% FTPDATA bursts.

With this definition of a burst of FTPDATA connections, we analyzed the same five datasets to measure the distribution of the number of bytes transferred during a single connection burst. The distribution proves to be remarkably heavy-tailed. Figure 8 shows percentage of all FTPDATA bytes (y-axis) due to the largest 10% of the FTPDATA bursts (x-axis; we have omitted LBL-1 as it is virtually the same as DEC-1). The numbers in parentheses in the legend give the total number of FTPDATA bursts occurring during each trace. The first vertical line marks the upper 0.5% of the FTPDATA bursts, and the line to its right, the upper 2%.

The key point to draw from this figure is that the upper 0.5% tail of the FTPDATA bursts holds *between 40% and 60% of all of the data bytes*. Thus, at any given time FTP traffic will most likely be *completely dominated by a single or small handful of bursts*. This finding means that for many aspects of network behavior, modeling small FTP sessions or bursts is irrelevant; all that matters is the behavior of a few huge bursts. The sizes and durations of these bursts will vary considerably from one time to another; but they *will* be present.[9]

[9] Our finding that the size of FTP burst has a heavy tail matches a survey

We did simple fitting of the upper tail of the distribution of data bytes per FTPDATA burst and found that for all five datasets, the upper 5% tail is fit well to a Pareto distribution with $0.9 \leq \alpha \leq 1.1$. As the Pareto distribution is heavy-tailed (see Appendix B), this agrees with our findings in Figure 8.

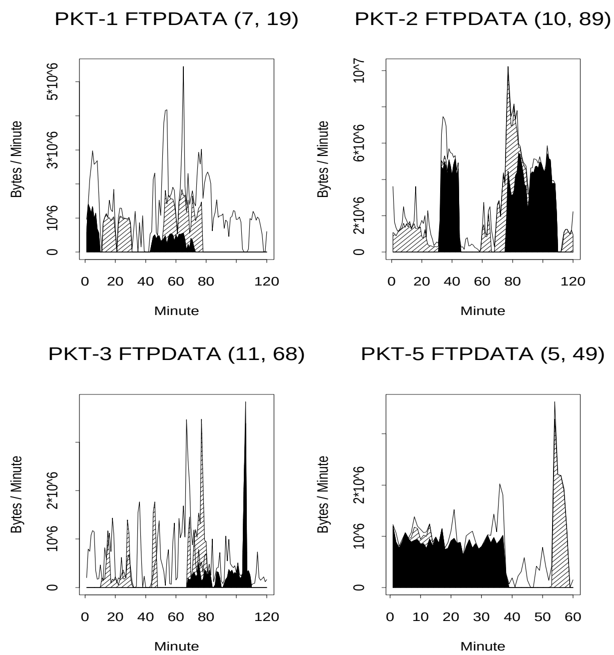

Figure 9: Proportion of FTPDATA traffic due to largest 2% (shaded) and 0.5% (black) connection bursts.

Figure 9 graphically illustrates the dominance of the upper FTPDATA-burst tail. The four plots show the FTPDATA traffic rate in bytes/minute for the PKT-1, PKT-2, PKT-3, and PKT-5 datasets. The shaded areas represent traffic contributed by the largest 2% of the bursts, and the black areas the largest 0.5%. The numbers in parentheses give the number of bursts and FTPDATA connections comprising the 2% burst upper-tail. (For example, the upper 2% tail of the PKT-1 bursts was made up of 7 bursts consisting of a total of 19 FTPDATA connections.) We see that sometimes bursts contain many separate connections and sometimes not. Indeed, the distribution of the number of connections per burst is well modeled as a Pareto distribution.[10]

For PKT-1 (364 bursts) and PKT-3 (552 bursts), the upper 2% and 0.5% tails hold around 50% and 15% of all the traffic; for PKT-2 (483 bursts) and PKT-5 (238 bursts), 85% and 60%. The large degree of difference between PKT-1/PKT-3 and PKT-2/PKT-5 illustrates how volatile the upper-tail behavior is; a trace comprising 400 bursts (and substantially more FTPDATA connections) might well be completely dominated by 2 of the bursts, or it might not, since 2 is a very small sample of the upper-tail behavior. Thus we are left in the difficult position of knowing that upper-tail behavior dominates traffic, but with such small numbers of bursts that we cannot reliably use large-number laws to predict what we are likely to see during any given short trace.

We would also like to know whether the arrivals of the upper-tail bursts can be modeled as a Poisson process, as that would provide a first step toward predicting their effect on network traffic. We analyzed the 199 upper-0.5%-tail LBL-6 bursts, first removing effects due to daily variation in traffic rates by looking at interarrivals in terms of number of intervening bursts instead of seconds. We found that the dataset failed the statistical test (Appendix A) for Poisson arrivals at all significance levels, and its autocorrelation function peaks at a lag of one burst, indicating that large bursts are themselves clustered (though the correlation is not dramatic). Thus, caution must be used if approximating large-burst arrivals using a Poisson process; further analysis is needed to model the burst-clustering.

7 Speculations regarding self-similarity

We have argued in the previous sections that on any time-scale smaller than user-session arrivals, modeling wide-area TCP traffic using compound Poisson processes fails to faithfully capture the traffic's dynamics. Recent work [LTWW93] shows that local-area Ethernet traffic (and perhaps wide-area TCP traffic) is much better modeled using *self-similar* processes, which display substantially more burstiness over a wide range of time scales than do compound Poisson processes. The pattern of burstiness has implications for many aspects of traffic performance, such as delay distributions, lengths of congestion periods, and traffic admissions control procedures.

In this section we discuss the degree to which the PKT-1 through PKT-5 traces of TELNET traffic, FTPDATA traffic, and general wide-area traffic are suggestive of self-similarity. We first look at a subtlety regarding inferring self-similarity from variance-time plots. We then give an overview of two methods for generating truly self-similar traffic (due to [LTWW93] and [W94]) and discuss how the models developed in this paper might match these methods. We finish with a preliminary look at aggregate wide-area traffic, which appears to be self-similar when fully aggregated (all network protocols), but less so when restricted to just TCP traffic, leaving us with a number of interesting but as-yet unanswered questions.

As explained in [LTWW93], straight lines on variance-time plots with slopes < -1, such as that for the PKT-2 TELNET trace in Figure 5, are suggestive of self-similarity[11]. In

conducted by Irlam [Irlam93] of the sizes of files in 1,000 file systems comprising 12 million files and 250 GB of data: 1.9% of the files accounted for 71% of the bytes, and 0.5% accounted for 54% of the bytes.

[10] For example, one of the bursts in the LBL-7 dataset was made up of 979 separate FTPDATA connections.

[11] A self-similar stochastic process is one which, when aggregated as explained in the text accompanying Figure 5, has the same autocorrelation

general, the slope of an arrival process's variance-time plot is a function of the process's autocorrelation function [Cox84]. One important point to bear in mind is that it is possible for a non-self-similar process with complicated autocorrelational structure to produce a straight line on a variance-time plot over a wide range of aggregation. For example, we have found that arrival processes using i.i.d. Pareto or heavy ($\sigma = 4$) log-normal interarrivals have this property. Thus, we do not claim that our PKT-2 TELNET trace is self-similar, as that requires a more thorough investigation[12] beyond the scope of this paper. We also note that the level of TELNET originator traffic in our traces is on the order of 100,000 packets over two hours, a packet rate much lower than that of the "low hour" link-level (all protocols) traces that the authors of [LTWW93] found exhibited only asymptotic self-similarity.

There are several methods for producing self-similar traffic that could account for self-similarity in TCP traffic. As discussed in [LTWW93], self-similar traffic can be produced by multiplexing ON/OFF sources that have exponential start times, a fixed rate in the ON periods, and ON/OFF period lengths that are *heavy-tailed* (see Appendix B).

A second method for generating self-similar traffic that could fit TCP traffic is an *immigration* model of individuals with Poisson arrival times and durations or lifetimes from a heavy-tailed distribution with infinite variance [Cox84, W94]. In this model (also called the M/G/∞ queue model [W94]), X_t is the number of individuals or connections in the system at time t. The process $\{X_t\}_{t=0,1,2,\ldots}$ is a self-similar process.[13] The immigration model implies that self-similar traffic would result from multiplexing constant-rate connections that have Poisson connection arrivals and a heavy-tailed distribution for connection lifetimes.

Our model of TELNET connections presented in Section 5 in some respects matches the immigration model described above. For example, TELNET connection sizes in packets do fit a heavy-tailed distribution, the log-normal distribution, although this is not quite as heavy-tailed as the Pareto distribution. The fact that the model for TELNET packet arrivals is not a constant-rate process does not necessarily reduce the self-similarity of the aggregated traffic. Thus, the immigration model can give some intuition to the suggested self-similarity of our TELNET traces and models.

Our model of FTP traffic also fits in some respects to the immigration model described above. For FTPDATA connections, the connection duration in seconds is roughly proportional to the connection size in packets (although the exact duration in seconds depends on such factors as the TCP window, the roundtrip time, and the available bandwidth on the bottleneck link). Similarly, over larger time scales the packet arrival process within an FTPDATA connection can be plausibly approximated as constant-rate. Thus, if we can approximate FTPDATA burst arrivals as Poisson (a bit of a stretch, as shown in Section 3 above), then multiplexed FTP bursts fit the immigration model above.

Unfortunately, although our model suggests that traces of FTP traffic could easily be self-similar, it is not clear from our variance-time plots that this is in fact the case.[14] Further work is required to understand the discrepancy.

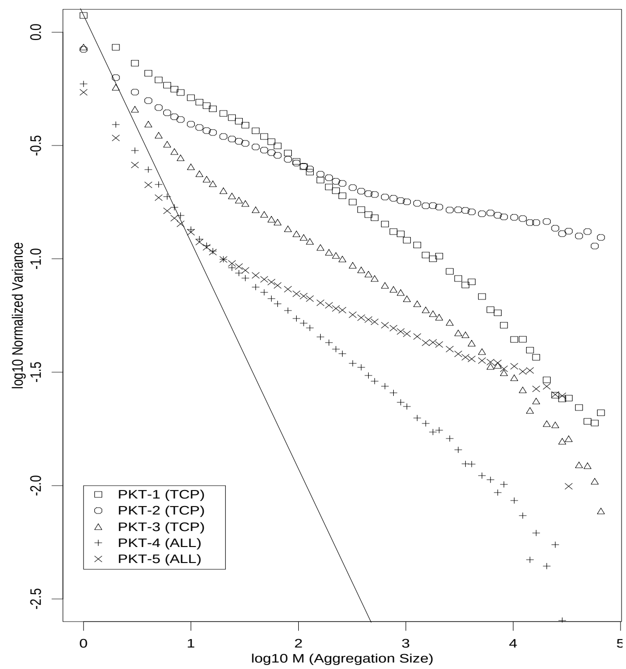

Figure 10: Variance-time plot for all TCP / all link-level packet arrivals.

We finish this section with a preliminary look at whether wide-area traffic aggregated over different protocols appears self-similar. Figure 10 shows variance-time plots for all of the "PKT" traces listed in Table 1. Here, the unaggregated process ($M = 1$) corresponds to observing the packets arriving during each 0.01 second interval.

Recall that the first three traces captured all TCP packets for two hours, and the last two captured all wide-area packets appearing on the gateway Ethernet for one hour. The first three traces consist of between 1.7 and 2.4 million packets, and the last two traces each have around 1.3 million packets. The corresponding rates of packets/hour are above those of the "low hours" in [LTWW93], so we would hope to find that the traces exhibit exact self-similarity.

function as the unaggregated process.

[12] I.e., computing a periodogram-based maximum likelihood estimate of the Hurst parameter with accompanying confidence intervals, as explained in [LTWW93, Garrett93].

[13] This method has the added attraction of giving an efficient procedure for generating self-similar traffic [W94]. We used this method to generate self-similar traces, in order both to develop intuition and to verify our statistical tests for self-similarity.

[14] Furthermore, we investigated the empirical autocorrelation functions for the PKT-1 through PKT-5 FTPDATA arrival processes and found significant differences in the functions between observing the process every 0.1 seconds and observing it every 1.0 seconds, much more so than for TELNET. If FTPDATA traffic is self-similar at this level of aggregation, the autocorrelation functions should appear similar, too.

We see that PKT-4 and PKT-5, the full link-level traces, both yield straight lines consistent with asymptotic self-similarity for $M \geq 10$ (0.1 second). For the TCP traces, PKT-2 appears consistent with asymptotic self-similarity for $M \geq 10^3$ (10 seconds); PKT-1 is concave down for small and large M; and while PKT-3 has a straight section between $M = 10$ and $M = 10^3$, it starts concave up and ends concave down.

The difference between the link-level and TCP traces is surprising. A possible implication is that TCP traffic is not well described as self-similar, but when multiplexed with Mbone (primarily audio traffic, carried in multicast UDP datagrams tunneled inside of IP datagrams) and DECnet (the other main contributor) traffic, the combination proves self-similar.[15] It may be that self-similarity results from how aggregated traffic sources affect one another, and that lulls in one type of traffic that lead to curves in variance-time plots of just that traffic, are coupled with bursts in other types of traffic. In any case, we emphasize that we are engaged here primarily in speculation; we lack sufficient data and testing to make any definitive statements regarding wide-area self-similarity.

We end with a related comment regarding the balance between link-level modeling and protocol-specific modeling. One approach to investigating self-similarity is to model aggregate link traffic as self-similar, without attempting to model individual connections. This approach could have many uses in simulations and in analysis. For example, aggregate self-similar traffic could be used instead of Poisson traffic to model cross-traffic, or aggregate self-similar traffic could be used in simulations investigating link-sharing between two different classes of traffic.

However, for many simulation experiments, the simulator needs to model individual sources. For example, in simulations that investigate the effects of different transport protocols or different gateway scheduling algorithms on network traffic, the simulator requires source models; the traffic patterns on the link will depend on the transport protocols and scheduling algorithms that are used in the simulations, as well as on the pattern of traffic generated by the source. It is to this end that further investigation of wide-area arrival processes will prove fruitful.

8 Implications

This paper's findings are summarized in the Introduction. In this section we conclude with a look at the implications of our results.

Several researchers have previously discussed the implications of long-range dependence (burstiness across different time scales) in network traffic. Modeling TCP traffic using Poisson or other models that do not accurately reflect the long-range dependence in actual traffic will result in simulations and analyses that significantly underestimate performance measures such as average packet delay or maximum queue size.

[FL91] examines the burstiness of data traffic over a wide range of time scales, and discusses the impact of this burstiness for network congestion. Their conclusions are that congested periods can be quite long, with losses that are heavily concentrated; that, in contrast to Poisson traffic models, linear increases in buffer size do not result in large decreases in packet drop rates; and that a slight increase in the number of active connections can result in a large increase in the packet loss rate. They suggest that, because the level of busy period traffic is not predictable, it would be difficult to efficiently size networks to reduce congestion adequately. They observe that, in contrast to Poisson models, in reality "traffic 'spikes' (which cause actual losses) ride on longer-term 'ripples', that in turn ride on still longer-term 'swells' ". They suggest that a filtered variable can be used to detect the low-frequency component of congestion, giving some warning before packet losses become significant. Our TELNET findings suggest that for some time scales of TELNET traffic, "swells" might correspond to (heavy-tailed) TELNET connection durations, "ripples" to lulls in TELNET user activity (the upper tail of the Tcplib interarrival distribution), and "spikes" to occasional flurries of user activity (the much smaller-scale lower tail of the Tcplib distribution).

[LTWW93] discusses some additional implications of long-range dependence of packet traffic. These include an explanation of the inadequacy of many commonly-used notions of burstiness, and the somewhat counter-intuitive observation that the modeling of individual connections can gain insight from an understanding of the fundamental characteristics of aggregate traffic. In this paper observations of the characteristics of aggregate traffic motivated our revisitation of models for individual connections; indeed, we originally set out to challenge the notion that wide-area traffic might be self-similar, and have come full circle.

[Garrett93] has examined the long-range dependence of variable-bit-rate (VBR) video traffic. His empirical measurements of VBR traffic show strong low-frequency components, and he proposes source models for video traffic that display the same long-range dependence. Given the likelihood that VBR traffic will soon comprise a large fraction of Mbone traffic, we soon will have aggregate wide-area traffic of which a substantial portion is perforce self-similar, simply due to the source characteristics of its individual connections.

There are some additional respects in which the burstiness and long-range dependence of TCP traffic can affect traffic performance. Consider a link with priority scheduling between classes of traffic, where the higher-priority class has no enforced bandwidth limitations (other than the link bandwidth itself). In such a partition, interactive traffic such as TELNET might be given priority over bulk-data traffic such as FTP. If the higher-priority class has long-range depen-

[15] In the PKT-4 and PKT-5 traces, neither the Mbone traffic nor the DECnet traffic alone appears self-similar.

dence and a high degree of variability over long time scales, then the bursts from the higher-priority traffic could starve the lower-priority traffic for long periods of time.

A second impact of the long-range dependence of packet traffic concerns classes with admissions control procedures that are based on measurements of recent traffic, rather than on enforced traffic parameters of individual connections. As has been show by numerous researchers, such admissions control procedures could lead to a much more effective use of the available bandwidth. Nevertheless, if the measured class has high burstiness and long-range dependence, then the admissions control procedure could be easily mislead following a long period of fairly low traffic rates.[16]

We believe one of the most important open questions is the degree to which network dynamics can themselves affect the burstiness of network traffic. For example, our WAN traffic traces are largely composed of TCP, UDP (mostly Mbone), and DECnet traffic. How do we interpret the fact that the aggregate link traces are more suggestive of self-similarity than are the traces of the three component parts? In what ways does the burstiness of one class of traffic affect the burstiness of another class? What would be the structure of the TCP, UDP, or DECnet traffic if the classes were not competing with each other for link bandwidth?

In summary: we should abandon Poisson-based modeling of wide-area traffic for all but user session arrivals. For TELNET traffic, we offer a faithful model of originator traffic parameterized by only the hourly connection arrival rate. Modeling the TELNET responder remains to be done. For FTP traffic, we have shown that modeling should concentrate heavily on the extreme upper tail of the largest bursts. A busy wide-area link might have only one or two such bursts an hour, but they tend to strongly dominate that hour's FTP traffic. Finally, our look at aggregate TCP and all-protocol traffic suggests that anyone interested in accurate modeling of wide-area traffic should begin by studying self-similarity.

9 Acknowledgments

We gratefully acknowledge helpful discussions with Johan Bengtsson, Kim Claffy, Mark Garrett, Van Jacobson, Allyn Romanow, and Walter Willinger.

We would like to thank Ramón Cáceres and Sugih Jamin for providing the UCB, BC, and DEC-1 datasets, which were gathered by Ramón, D. V. Wilson, and Jeff Mogul, respectively; and Craig Leres and Steve McCanne, who helped us gather the LBL traces.

This work was supported by the Director, Office of Energy Research, Scientific Computing Staff, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

[16] This is similar to a situation in California geology some decades ago. Because there hadn't been a large earthquake for a long time, people began to believe it unlikely that there would be another one.

A Methodology for testing for Poisson arrivals

To test whether a trace of connection arrivals corresponds to a nonhomogeneous Poisson process, we first pick an interval length I over which we hypothesize that the arrival rate does not change. If the trace spans a total of T time units, we divide the entire trace into T/I intervals each of length I. We then separately test the interarrivals in each interval against the null hypothesis that they come from an exponential distribution whose mean is the same as that of the interval's interarrivals, which would be the case if they come from a homogeneous Poisson process.

For our statistical goodness-of-fit determination, we use the Anderson-Darling (A^2) test, recommended by Stephens in [DS86] because it is generally much more powerful than either the better-known Kolmogorov-Smirnov or χ^2 tests, and is particularly good for detecting deviations in the tails of a distribution. A^2 is an *empirical distribution test*; it looks at the entire observed distribution, rather than reducing the distribution into bins as is required by χ^2.

We associate a *significance level* with each A^2 test. For example, a test with a significance level of 5% will correctly confirm the null hypothesis (if it is correct) 95% of the time; the other 5% of the time, the test will erroneously declare the hypothesis false. Thus, the significance level indicates the percentage of "false negatives" (in general it is difficult to assess the corresponding percentage of "false positives"). We can use significance-level testing as follows: if we test 100 one-hour intervals for exponential interarrivals and 95% of them pass the A^2 test at the 5% significance level, then we can say that the interarrivals appear statistically indistinguishable from those drawn from a true exponential distribution. If, for example, 90% of the intervals pass A^2 at the 5% level, then the exponential model is quite good, but not exact. If substantially fewer pass the test, then the exponential model is not very good.

There are two important details for correctly applying and interpreting the A^2 test. The first is that estimating the parameters of our model from the data to be tested alters the significance levels of the A^2 test (this applies to our null hypothesis above, in which we derive the mean of the exponential fit from the data rather than knowing it *a priori*). The second is that the number of data points tested also alters the significance levels. In general, the more points tested, the more likely the test will detect an incorrect null hypothesis. [DS86] gives procedures for incorporating both of these considerations into A^2 tests.

B Pareto distributions

The Pareto distribution plays a role both in TELNET packet interarrivals and in the size of FTPDATA bursts. In this appendix we summarize the Pareto distribution and discuss its interpretation and occurrence in the physical world.

A Pareto distribution with parameter α has the following cumulative distribution function:[HK80]

$$P[X \leq x] = 1 - (a/x)^{\alpha}, \quad \alpha, a \geq 0, \quad x \geq a,$$

with the corresponding probability density function:

$$f(x) = \alpha a^{\alpha} x^{-\alpha-1}.$$

Following [LTWW93], we say that a distribution is *heavy-tailed* if

$$P[X \geq x] \sim x^{-\alpha}, \quad \text{as } x \rightarrow \infty, \quad \alpha \geq 0.$$

If $\alpha \leq 2$, then the heavy-tailed distribution has infinite variance; if $\alpha \leq 1$, then it has infinite mean. Examples of heavy-tailed distributions include the Pareto, log-normal, and Weibull distributions [DMRW94] (though the log-normal distribution is less heavy-tailed than the Pareto distribution [Garrett93, p.96]).

For a light-tailed distribution of waiting times such as the uniform distribution, the longer you have waited, the sooner you are likely to be done. An example of a medium-tailed distribution is the memoryless exponential distribution; the future waiting time is independent of the waiting time so far. In contrast, the Pareto distribution is in some sense scale-invariant; the probability that the wait is at least $2n$ seconds is a fixed fraction of the probability that the wait is at least n seconds, regardless of the value of n. Thus, the longer you have waited, the more likely you will have to wait still longer.

The Pareto distribution (referred to as the *power-law* distribution in some publications) has been used to model such diverse distributions as incomes exceeding a minimum value, asteroid sizes, extinction events, and energy released by earthquakes [K93]. Bak and Chen explained the Pareto distribution of sizes of avalanches as resulting from “self-organized criticality”, where a single grain of sand added to a sandpile occasionally triggers a large avalanche [BC91]; it is not clear that this notion of criticality adds insight to the occurrence of Pareto distributions in TCP traffic modeling. In communications, heavy-tailed distributions have been used to model telephone call holding times [DMRW94]. Thus, the presence of the Pareto distribution in network traffic distributions is not overly peculiar, though its proper physical interpretation remains unclear.

References

[BC91] P. Bak and K. Chen, “Self-Organized Criticality”, Scientific American, January 1991, pp. 46-53.

[CBP93] K. Claffy, H.-W. Braun and G. Polyzos, “Long-term traffic aspects of the NSFNET”, Proceedings of INET ’93, SDSC Report GA-A21238, San Diego Supercomputer Center, 1993.

[Cox84] D. R. Cox, *Long-Range Dependence: A Review*, in “Statistics: An Appraisal,” Proceedings 50th Anniversary Conference, Iowa State Statistical Library, H. A. David and H. T. David, editors, Iowa State University Press, pp. 55-74, 1984.

[DMRW94] D. Duffy, A. McIntosh, M. Rosenstein, and W. Willinger, “Statistical Analysis of CCSN/SS7 Traffic Data from Working CCS Subnetworks”, to appear in IEEE JSAC, 1994.

[DS86] R. B. D’Agostino and M. A. Stephens, editors, “Goodness-of-Fit Techniques”, Marcel Dekker, Inc., 1986.

[DJ91] P. Danzig and S. Jamin, “tcplib: A Library of TCP Internetwork Traffic Characteristics”, *Report CS-SYS-91-01*, Computer Science Department, University of Southern California, 1991. Available via anonymous FTP to excalibur.usc.edu as pub/jamin/tcplib/tcplib.tar.Z.

[DJCME92] 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, 3 (1), pp. 1-26, 1992.

[FJ93] S. Floyd and V. Jacobson, “The Synchronization of Periodic Routing Messages,” SIGCOMM ’93, September 1993, pp. 33-44.

[FL91] H. Fowler and W. Leland, *Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management*, IEEE JSAC, 9 (7), pp. 1139-1149, September, 1991.

[FJ70] E. Fuchs and P. E. Jackson, *Estimates of Distributions of Random Variables for Certain Computer Communications Traffic Models*, Communications of the ACM, 13(12), pp. 752-757, December, 1970.

[Garrett93] M. Garrett, Contributions Toward Real-Time Services on Packet Switched Networks, CU/CTR/TR 340-93-20, Columbia University, 1993.

[Gusella90] Riccardo Gusella, *A Measurement Study of Diskless Workstation Traffic on an Ethernet*, IEEE Transactions on Communications, 38 (9), pp. 1557-1568, September, 1990.

[HK80] T. Hettmansperger, and M. Keenan, “Tailweight, Statistical Inference, and Families of Distributions - A Brief Survey”, in *Statistical Distributions in Scientific Work*, Patil, G.P., et al. (eds), Kluwer Boston, 1980, pp. 161-172. V.1

[Irlam93] G. Irlam (gordoni@netcom.com), "ufs'93 [Updated file size survey results]", USENET newsgroup comp.os.research, message 2ddp3b$jn5@darkstar.UCSC.EDU, Nov. 29, 1993.

[JR86] R. Jain and S. Routhier, *Packet Trains — Measurements and a New Model for Computer Network Traffic*, IEEE JSAC, 4 (6), pp. 986-995, September, 1986.

[K93] S. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, 1993.

[LTWW93] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic", SIGCOMM '93, September 1993, pp. 183-193.

[MM85] W. T. Marshall and S. P. Morgan, *Statistics of Mixed Data Traffic on a Local Area Network*, Computer Networks and ISDN Systems 10(3,4), pp. 185-194, 1985.

[P94] V. Paxson, "Growth Trends in Wide-Area TCP Connections", to appear in *IEEE Network*, 1994.

[P93] V. Paxson, "Empirically-Derived Analytic Models of Wide-Area TCP Connections: Extended Report", technical report LBL-34086, Lawrence Berkeley Laboratory, May, 1993. Available as *WAN-TCP-models.1.ps.Z* and *WAN-TCP-models.2.ps.Z* via anonymous FTP to ftp.ee.lbl.gov.

[Stevens94] W. Richard Stevens, "TCP/IP Illustrated, Volume 1: The Protocols", Addison-Wesley, 1994.

[W94] W. Willinger, private communication.