Contradictory Relationship between Hurst Parameter and Queueing Performance

Ronn Ritke*, Xiaoyan Hong and Mario Gerla
UCLA – Computer Science Department,
405 Hilgard Ave., Los Angeles, CA 90024
ritke@cs.ucla.edu, hxy@cs.ucla.edu, gerla@cs.ucla.edu

Abstract

Long Range Dependent (LRD) network traffic does not behave like the traffic generated by the Poisson model or other Markovian models. The main difference is that LRD traffic increases queueing delays due to the burstiness of the traffic over many time scales. LRD traffic has been measured in different types and sizes of networks for different applications (eg. WWW) and different traffic aggregations. Since LRD behaviour is not rare nor isolated, accurate characterization of LRD traffic is very important in order to predict performance and to allocate network resources. The Hurst parameter is used to describe the degree of LRD and the burstiness of the traffic. In this paper we analyze UCLA Computer Science Department network traffic traces and compute their Hurst parameters. Queueing simulation is generally used to study the impact of LRD and to determine if the Hurst parameter accurately describes such LRD. Our results show that the Hurst parameter is **not** by itself an accurate predictor of the queueing performance for a given LRD traffic trace.

Key words: Hurst parameter, Long Range Dependence, Queueing performance.

*Ronn Ritke is the corresponding author, Fax: 1 310 8257578

1 Introduction

Accurate characterization of Internet traffic is very important for precise modeling and network design decisions. Modeling of Internet traffic is based on the traffic characteristics and the resulting models often serve as input for simulations. The results of simulations are used for a number of network design decisions. For many years the Poisson model was widely used to model Internet traffic, but in the last few years new characteristics have been discovered in Internet traffic. Long Range Dependence (LRD) has been discovered in LANs [1] [2] and WANs [3]. It has also been discovered in different services and applications: Aggregate [1] [2] , World Wide Web [4] [5], Variable-Bit-Rate (VBR) video traffic [6] and different types of computer networks: Ethernet [1] [2] , ISDN [9] and CCSN/SS7 [10]. The traffic with the LRD property is more bursty than traffic generated with the Poisson model. The Poisson model is Short Range Dependent and does not accurately model LRD traffic [3]. In comparison to LRD traffic, the use of the Poisson model (or other Markovian models) results in overly optimistic queueing performance. The queue length distribution decays much more slowly for LRD traffic. The queueing delay rises dramatically with increasing LRD [7] and the Hurst parameter quantifies this long range dependence.

In this paper we look at the relationship between the Hurst parameter and queueing performance. The rest of the paper is organized as follows: Information about the computer network traffic traces used in this paper is presented in Section 2. The definition for long range dependence is given in Section 3. The computation of the Hurst parameter and a discussion on the variance-time plot and examples of its use in estimating the Hurst parameter with UCLA network traffic is covered in Section 4. The performance of LRD traffic in a queueing simulation in order to determine the effectiveness of the Hurst parameter in predicting the resulting queueing performance is covered in Section 5. Lastly, in Section 6 we summarize our findings.

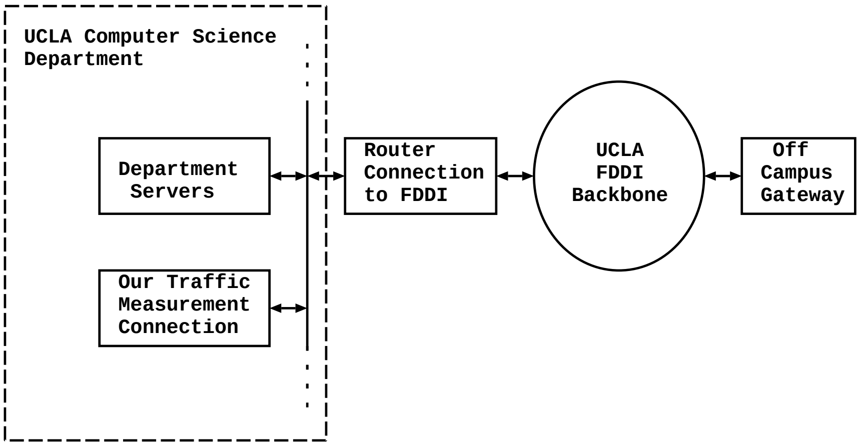


Figure 1: UCLA CSD Measurement Connection Diagram.

2 Traffic Traces

Network traffic traces were taken at UCLA CSD over a 5 week period (Feb - March 1998). Network traffic information was collected at a host running Tcpdump [15]. This host was connected (via a special link) to department servers and to the router that connects the CSD to the FDDI backbone (see Figure 1). The traces represent the network traffic in the Computer Science Department. The resulting output was processed to obtain the format needed to test for LRD (arrival time and packet length for each packet). Information for each traffic trace obtained is summarized in Table 1. Note that the 'All' in the trace name signifies that it is an aggregate traffic trace.

Trace Name	Start Date	Start Time	Duration	Arrivals
All1	3/6/98	1pm	2 Hour	510519
All4	2/26/98	10am	2 Hour	1238081
All7	2/23/98	9am	2 Hour	2100013
All8	3/23/98	12pm	3 Hour	686962

Table 1: UCLA Computer Science Dept Traffic Trace Information.

In order to determine which applications are present in the UCLA CSD network traffic we made use of the fact that (following RFC 1700 recommendations) a number of TCP and UDP applications were assigned to well known port numbers. For example, http network traffic was assigned to port 80. Each well known port has both a UDP and a TCP application assigned to it even if the application supports only the TCP implementation. Tcpdump allows for the collection of all the packet traffic (our All traces use this Tcpdump option) and it can be used to 'filter' the traffic so that only TCP or UDP packets are collected in the traffic trace. For a 7 day traffic trace taken earlier, Tcpdump was used to capture only TCP traffic. As a result, when a traffic analysis tool counted the number of packets or bytes sent to the well known ports there was no ambiguity whether it is TCP or UDP. Two different traffic summaries - packet and byte - are presented in Tables 2 and 3 for a 7 day trace. Both summaries are shown because some applications have a large number of packets but a small number of bytes per packet while other applications have a small number of packets but each packet has a large number of bytes. To get an overview of the network traffic we look at the applications that make up the bulk of the measured traffic and only the applications that have over 0.9 percent of the total bytes or total packets are reported.

Port	Packets Percent	Application
6000	24.73	X-Windows
513	6.543	login
23	5.87	Telnet
80	3.849	http (World Wide Web)
119	3.42	nnt
514	1.855	Syslog
20	1.596	FTP-data
515	1.483	printer

Table 2: Summary of the Applications that have over 0.9% of the total packets.

3 Definition of Self-similar and LRD processes

Our approach is to define LRD following the definitions given in [1] [13].

Port	Bytes Percent	Application
6000	34.4	X-Windows
514	6.528	Syslog
119	2.739	nntp
515	2.62	printer
25	1.376	SMTP
20	1.253	FTP-data
513	0.998	login
23	0.952	Telnet
80	0.907	http (World Wide Web)

Table 3: Summary of the Applications that have over 0.9% of the total bytes.

Let $X = (X_t : t = 0, 1, 2, \ldots)$ be a covariance stationary stochastic process with mean μ, variance σ^2 and autocorrelation function $r(k), k \geq 0$. Assume $r(k)$ is of the form

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

where $0 < \beta < 1$.

For each m=1,2,3,..., let $X^{(m)} = (X_t^{(m)} : t = 1, 2, 3, \ldots)$ denote the new covariance stationary time series obtained by averaging the original series X over non-overlapping blocks of size m, i.e.,

$$X_t^{(m)} = (X_{tm-m+1} + \cdots + X_{tm})/m, \quad t \geq 1 \tag{2}$$

The process X is called (exactly) second-order self-similar if for all $m = 1, 2, 3, \ldots,\ var(X^{(m)}) = \sigma^2 m^{-\beta}$ and

$$r^{(m)}(k) \sim r(k), \quad k \geq 0 \tag{3}$$

The process X is called (asymptotically) second-order self-similar if for all k large enough,

$$r^{(m)}(k) \to r(k), \quad as \quad m \to \infty \tag{4}$$

The key property of this class of self similar processes is the fact that the covariance does not change under block aggegation and time scale changes. The relationship between the Hurst parameter and β is $H = 1 - \beta/2$. Note that here $1/2 < H < 1$, since $0 < \beta < 1$.

A self similar process with $1/2 < H < 1$ (i.e., $\beta < 1$) is long range dependent (LRD). Since $\beta < 1$ the function $\sum_k r(k) = \sum r^{-\beta} = \infty$. By contrast, a short-range dependent process (eg. Poisson Process) has fast decaying autocorrelation function (i.e., $\beta > 1$), hence, $\sum_k r(k) < \infty$. The Hurst parameter is thus a key indicator of LRD behavior. One immediate consequence of LRD behavior is that the traffic exhibits the same burstiness across many time scales. A commonly held belief is that the degree of LRD and of traffic burstiness is completely characterized by the Hurst parameter. I.e, the higher the H value, the burstier the traffic.

Having introduced the definitions of Hurst parameter and LRD, in the following section we define some important time series and their relationship to the original traffic trace.

4 Time series generation and H computation

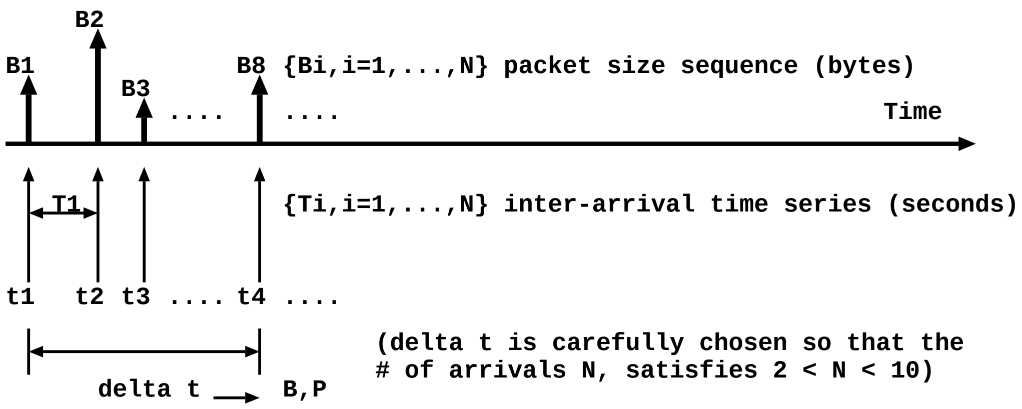


Figure 2: Time Series Diagram.

A. Time Series Generation

The original traffic trace is characterized by two variables: time (of arrival of a packet) and length (of the packet). From this trace, time series with only one variable must be generated in order to estimate the Hurst parameter for this variable. There are several methods for generating such single variable time series from data traces. Researchers from

Bellcore [1] [2] have proposed the {Traffic B} and {Traffic P} time series while researchers from Boston University [4] [5] have introduced two other time series denoted as $\{T_i\}$ and $\{B_i\}$. We will examine all 4 time series. The relationship of the four time series from the packet arrival times and packet lengths is displayed in Figure 2. Each of the four time series captures different aspects of the traffic trace. $\{T_i\}$ is the inter-arrival time series (no notion of packet length) and $\{B_i\}$ represents the packet size sequence (no notion of arrival times). To compute {Traffic B} and {Traffic P}, we choose a time interval Δt which typically contains between 2 and 10 arrivals (see Figure 2). Within non-overlapping time intervals of size Δt we sum the number of bytes $\overline{B}_i$ arriving in each interval Δt_i and get the time series {Traffic B} = $\{\overline{B}_i, i = 1, 2, 3, \ldots\}$. Next, let $\overline{P}_{ave}$ be the mean packet size computed over the entire duration of the experiment. Consider the new traffic sequence where the actual packet size is replaced by the mean packet size $\overline{P}_{ave}$. Within non-overlapping time intervals of size Δt we sum the number of bytes $\overline{P}_i$ arriving in each interval Δt_i to get a time series {Traffic P} = $\{\overline{P}_i, i = 1, 2, 3, \ldots\}$.

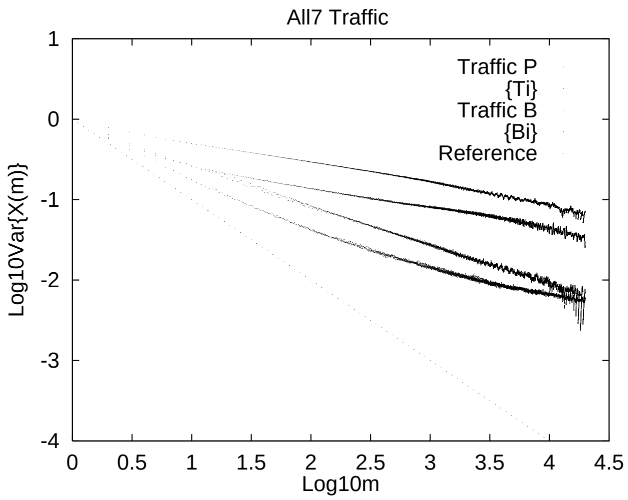


Figure 3: Variance-time plots for the time series: $\{T_i\}$, $\{B_i\}$, {Traffic P} and {Traffic B}.

B. Estimation of the Hurst Parameter

We present two different views of the Hurst parameter: one is a visual estimated view (variance-time plot), the other is a computed view (see next subsection). To visually estimate the Hurst parameter, we plot $var(X^{(m)})$ as a function of m. The variance-time plot draws the variance vs. m in a log-log scale, which shows the slowly decaying variance of a self-similar series. If the input traffic has the LRD property, the curve should be linear (for large m) with slope larger than -1. The 'Reference' line on the variance-time plot (Figures 3 and 4) represents the slope of the line of $\beta = 1$, that is $var(X^{(m)}) = m^{-1}$, then H = 1/2. Any line with a slope less than 0 and greater than this reference line exhibits LRD and has an H parameter value $1/2 < H < 1$. Figure 3 shows the variance-time plot for all four time series mentioned previously. The captions represent the curves top down on the graph. By inspection of the variance-time plot, it is apparent that all four curves have an H value greater than 1/2 and less than 1, demonstrating that all four curves show the property of long range dependence.

C. Computation of the Hurst Parameter

The Hurst parameters were computed for the same traffic trace files that served as input for the variance-time plots. The Least-Squares Curve Fitting [14] was used to get an equation for the curve. The resulting equation is in the Slope-Intercept form of the equation of a line $y = mx + b$ where m is the slope. From the definition in the previous section we know that $-\beta$ is the slope of the curve (so here $-\beta = m$). The Hurst value is then computed using the relation $H = 1 - \beta/2$.

The queueing simulation requires both the packet arrival times and the packet lengths. We can not use the time series $\{T_i\}$ (has only packet size information) and $\{B_i\}$ (has only the interarrival time information) for the queueing simulation. As a result, we focus on {Traffic P} and {Traffic B} for the Hurst parameter estimation and for the queueing simulation. Table 4 contains the computed Hurst values of {Traffic B} and {Traffic P} for each of the four traffic traces used in this paper. The variance-time plots for {Traffic B} and {Traffic P} for all four traces are shown in Figure 4. Notice that all four computed {Traffic P} Hurst values are greater than their respective computed {Traffic B} values. Simulation is used in the next section to show the relationship between the Hurst parameter and queueing performance.

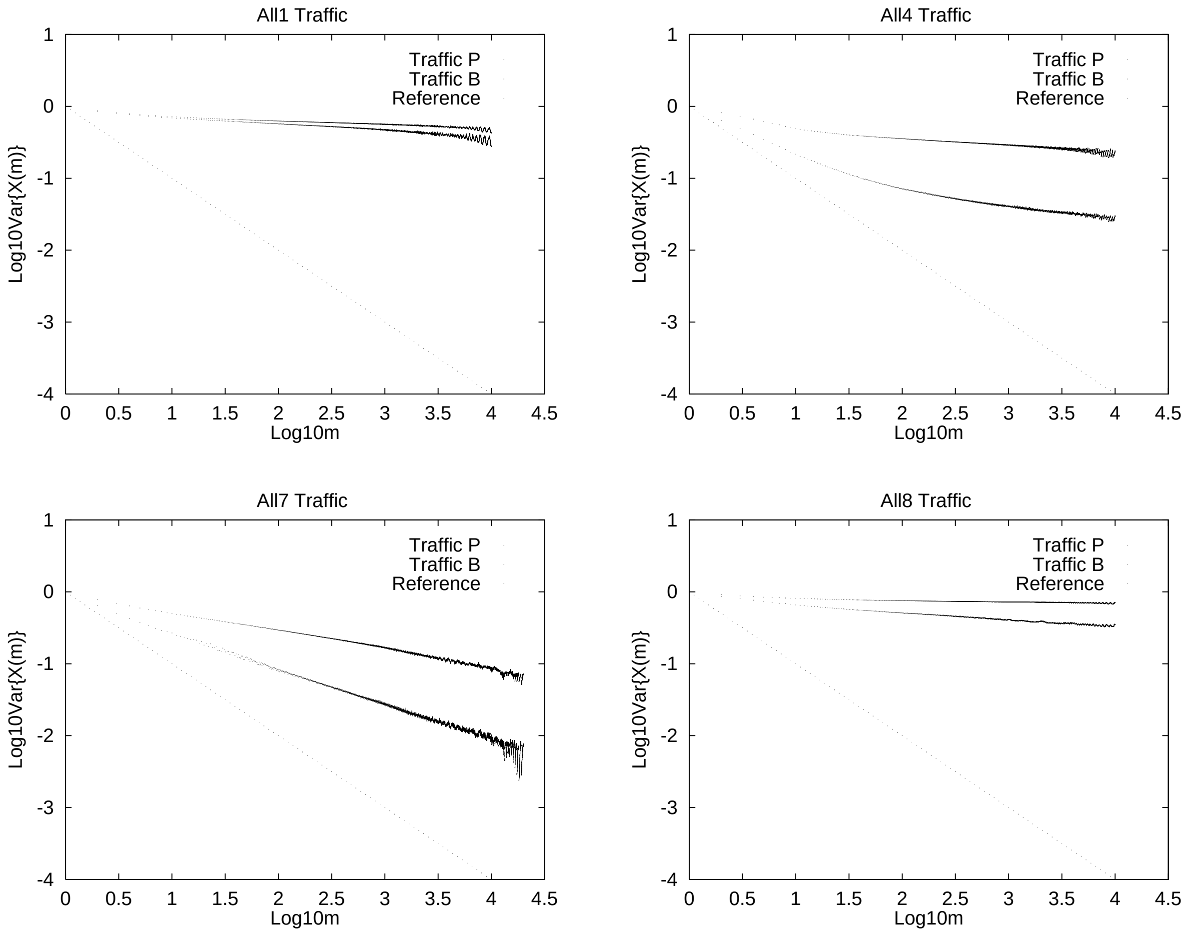


Figure 4: Variance-time plots for the 4 traces.

Traffic Trace	{Traffic B}	{Traffic P}
All1	0.9338	0.9652
All4	0.8991	0.9427
All7	0.7395	0.8475
All8	0.9542	0.9901

Table 4: Hurst values for the 4 traces.

5 The Queue Simulator

In this section, we introduce our queuing simulation which uses real traffic traces and then we discuss the simulation results. Previous studies on the queueing simulation are either driven by real traffic traces or by traffic models. As was done in many other modeling approaches, we use a single parameter (Hurst parameter) to describe the self-similar property of network traffic. Our paper differs from previous studies, however in that we discuss the relationship between Hurst parameter and queueing performance. We use UCLA traffic, which exhibits the long range dependency property, to drive a queueing simulation. It is well known that LRD traffic is burstier than the traditional Poisson model, and thus requires a much larger queue size. By observing the influence of such traffic on the queueing system, we will see that the Hurst parameter alone does not sufficiently quantify the LRD property, nor does it characterize the burstiness of real traffic.

The queueing system utilized in the simulation has a single server, infinite buffer size and FIFO discipline. Experiments were run with five different server utilizations (0.3, 0.5, 0.7, 0.9, 0.97). Only the experiments with utilizations (0.5, 0.7) are reported here. The queueing simulation is driven by {Traffic B} and {Traffic P} sequences, which in turn were derived from real traces.

6 Experimental Results

In the experiments, we measure the complementary distribution of the queue length. Let Q(t) be the number of bytes in the queue over time. In the plots we show $P(Q(t) > x)$, the probability that the queue length is greater than x, in log scale. The longer the tail of the distribution, the burstier the traffic.

In order to provide a reference, Figure 5 shows the queue length distributions for an M/M/1 queueing system with utilizations 0.3, 0.5 and 0.7. The M/M/1 queueing model uses the average interarrival time (0.0019 sec) and average packet size (472 bytes) extracted from the trace All1. In contrast, in Figure 6 we show the simulation results obtained by applying real traffic traces to the FIFO queue. In particular, Figure 6 (a) corresponds to the All1 Trace. It should be pointed out that the queue distribution yielded by {Traffic B} trace was found to be identical to that produced by the original (measured) trace. Thus, in all the following plots we only show {Traffic B} results. It is very clear by comparing Figure 5 and Figure 6(a) that there is a large difference (3 orders of magnitude) in queue length distributions between the M/M/1 queue and the real traffic queue (corresponding to {Traffic B}) and as a result the Poisson process can not be used as a substitute for the real LRD traffic. The disparity between Poisson and Self Similar queues is revealed also by the analytical models. In particular, for the Poisson process (M/M/1 queue) the $P(Q(t) > x) \sim e^{-\eta x}$ [6] but for the self similar process the $P(Q(t) > x) \sim e^{-\alpha x^{2-2H}}$ [6] [7].

Simulation results for the real traffic trace and {Traffic B} are identical so the figures are not presented here.

Researchers in the area of self-similarity widely accept the hypothesis that a large H value indicates greater burstiness [1]. The Hurst parameter is most important in traffic modeling because it concisely describes the self similarity property and burstiness of the traffic [3] [7]. As confimed by simulation, the burstier the traffic, the longer the tail of queue length distribution. However, our experiments show that the widely accepted hypothesis that a larger H value indicates a greater burstiness is not always true. Several queueing experiments and discussions of the queueing simulation results provide evidence from several different points of view to support such an argument.

First, let us look at the behavior of the {Traffic B} and {Traffic P} synthetic traces which were derived from the same traffic trace. Consider trace All1 (Figure 6 (a)) for example.

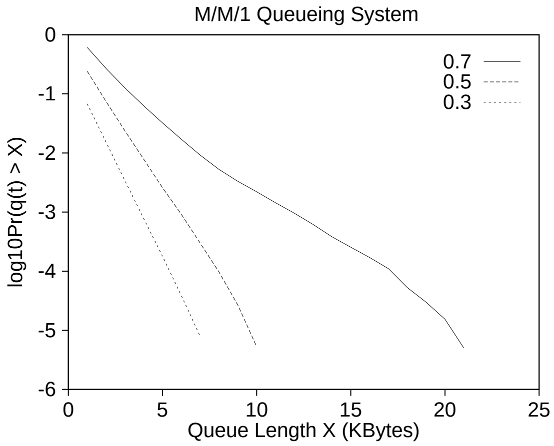


Figure 5: Queueing experiment for M/M/1 model using λ and μ from All1 trace .

The computed Hurst parameters show that the H values of Traffic B (0.9338) and P (0.9652) of trace All1 (Table 4) are very close. But the tail of {Traffic B} is much longer than that of {Traffic P}. This suggests that {Traffic B} is much burstier than {Traffic P}. Moreover, trace All7 in Figure 6 (c) and Figure 4 (c) shows that while {Traffic P} has a greater H value than {Traffic B} (i.e. 0.8475 vs. 0.7395), the tail of the queue length distribution of {Traffic P} is shorter than that of {Traffic B}. These 'inversions' support our claim that the value of the Hurst parameter does not accurately reflect the relative burstiness between {Traffic B} and {Traffic P} from the same trace.

Traffic Trace	{Traffic B}	{Traffic P}
All7 Seg1	0.9171	0.9587
All7 Seg2	0.9387	0.9710
All7 Seg3	0.9462	0.9713
All7 Seg4	0.9574	0.9748
All7	0.7395	0.8475

Table 5: Hurst values for the four segments of trace All7.

More support of our argument comes from the simulation experiments with segments from the same trace. Previous experiments on segments of the original trace were done by Abry

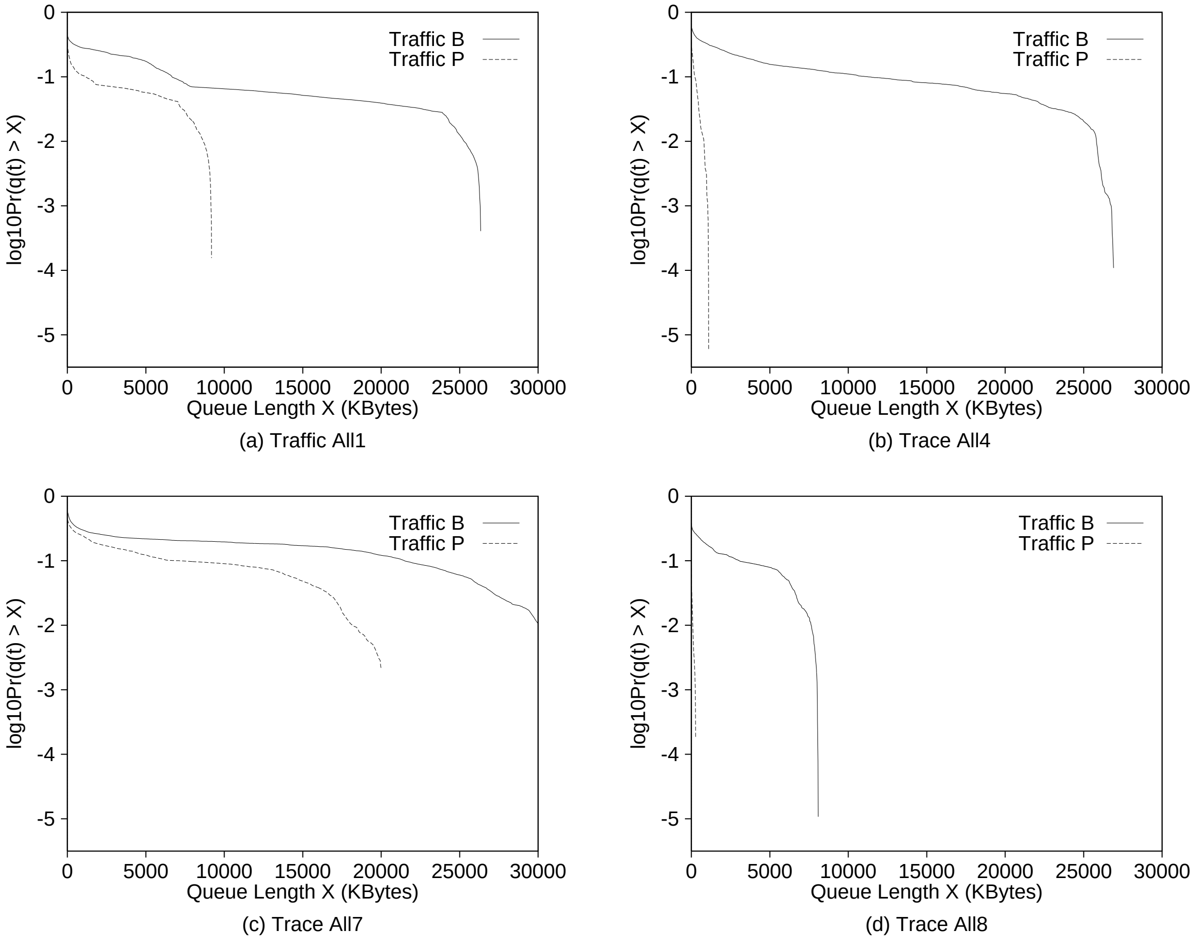
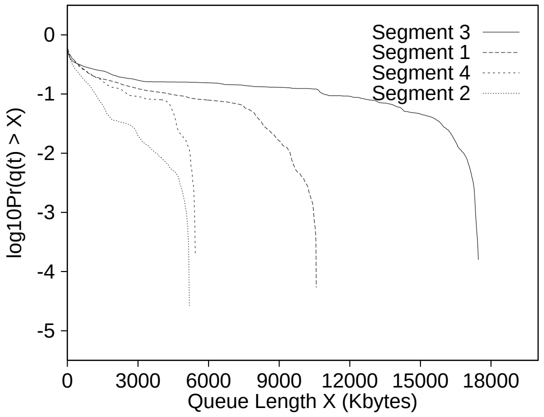
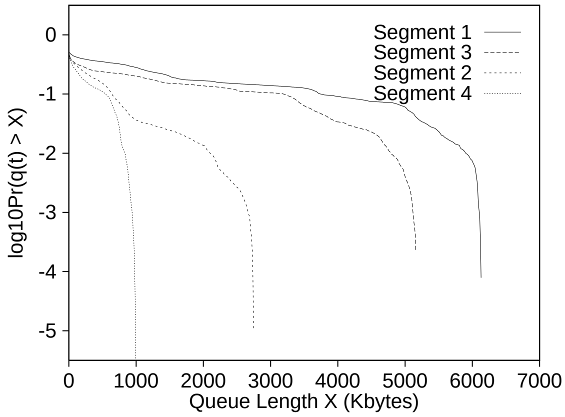


Figure 6: Queueing experiment of each trace with 0.7 utilization.

[12] to show that the Hurst parameter is constant across the segments. Our segmented experiments show that parts of the entire trace perform differently from the original whole one. The original trace All7 is divided into four sections. Each section is treated as an individual trace. We derive {Traffic P} and {Traffic B} for each trace, estimate their H values, and feed them into the queueing system. The utilization is kept at 70 percent. The results given in Table 5 shows that the {Traffic P} and {Traffic B} segments have H values similar to each other. However, when the results of queueing experiments are considered (Figure 7), strong differences between queue length distributions among the four segments are observed. For example, {Traffic B} of Segment3 has a heavier tail than the other segments,



(a) Traffic B



(b) Traffic P

Figure 7: Queueing experiment of Segments from trace All7 with 0.7 utilization.

but it doesn't have the largest H value. So the queueing performance is different for segments with similar H parameters . Furthermore, the tails of these distributions are much lighter than the tail of the queue length distribution of the entire trace All7 (see Figure 6 (c)). That is, the largest queue length of the four segments of {Traffic B} is near 18MBytes (Figure 7 (a)), but the probability of a queue length larger than 25MByte is still near 10 percent for {Traffic B} of the entire trace (Figure 6 (c)). However the value of the Hurst parameter for All7 is smaller than that of the segments. The segments can have Hurst values larger than the original because traffic can be more bursty in smaller time scales. In other words the Hurst value can vary with different time scales.

The queue length distributions of the traces when the system is at 50 percent load is shown in Figure 8. The distribution of queue buffer size decays faster than in 70 percent utilization. Just as Erramilli, et. al., showed in their paper [7], generally, a traffic load of 0.5 is near the "knee" of the delay-utilization curve. When the utilization is greater than 0.5, the queueing delay increases very fast. From the queue length distribution, this point can be clearly seen, where the tail of a heavy traffic load (Figure 6) is much longer than that of a light load (Figure 8). Therefore, our discussions have focused on the queuing performance under heavy load (0.7), which is of most interest. For the sake of completeness, however, we

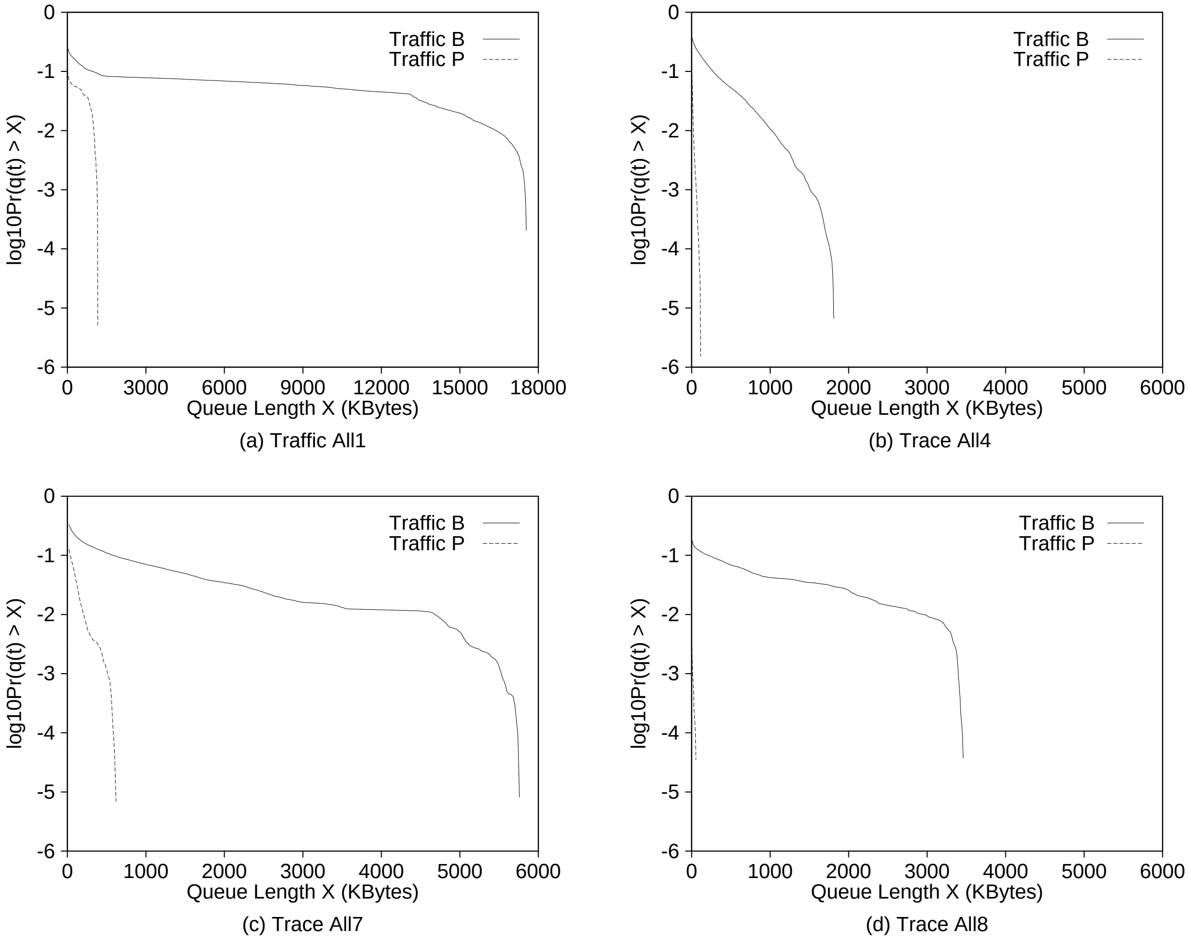


Figure 8: Queueing experiment of each trace with 0.5 utilization.

have compared the All1 simulation (both {Traffic P} and {Traffic B}) for loads 0.3, 0.9, 0.97 (see Figure 9). We note that the property that queueing lengths for {Traffic B} are longer than for {Traffic P}, for each traffic trace run at each utilization (0.3, 0.5, 0.7, 0.9, 0.97).

All these comparisons lead us to the same conclusion: that the Hurst parameter is not an accurate indicator of the traffic burstiness. Abry [12] has pointed out that the Hurst parameter is only a single aspect, albeit the most important aspect of describing the traces. Veitch [11] used a second parameter for the purpose of quantitative analysis of the LRD phenomenon, but he didn't measure queueing performance. Our queueing experiments clearly

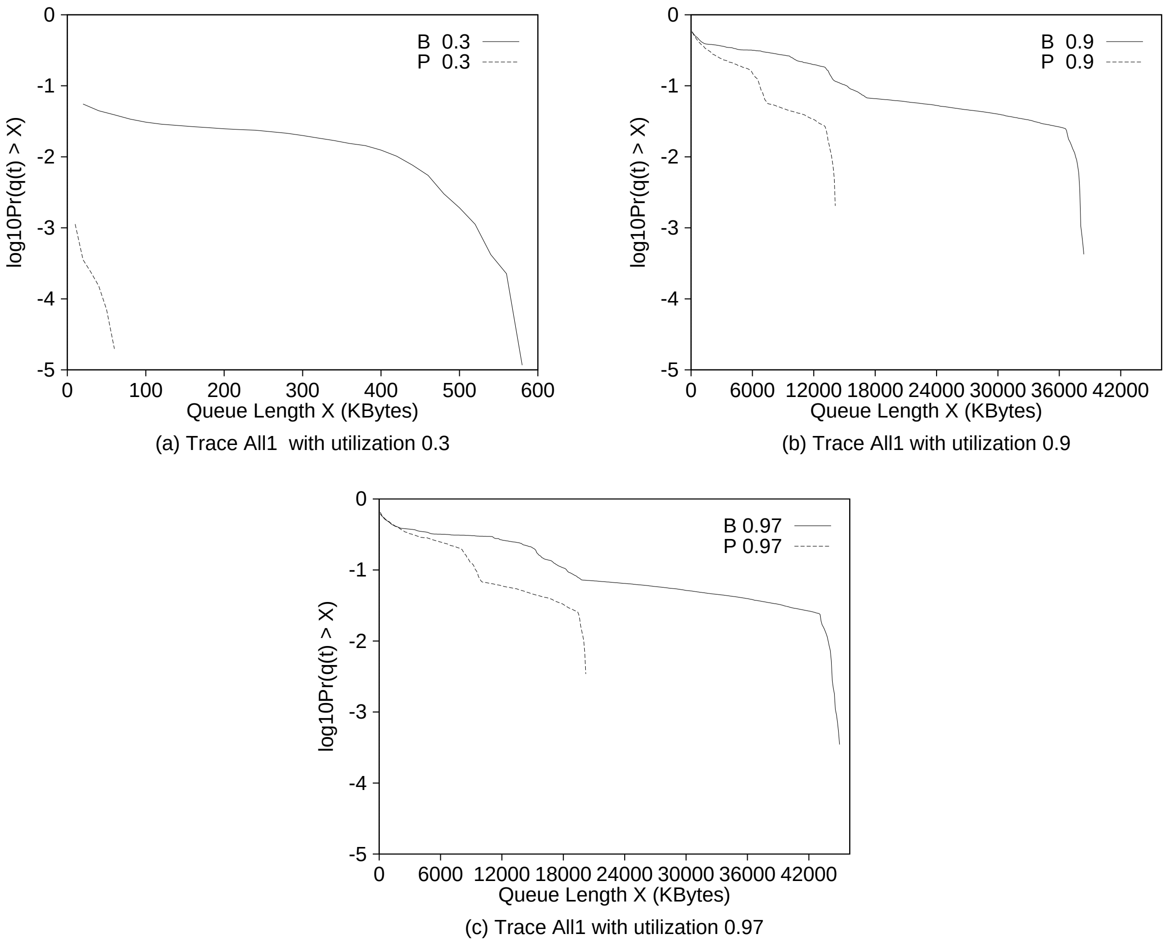


Figure 9: Queueing experiments for trace All1 with 0.3, 0.9 and 0.97 utilizations.

show that the Hurst parameter alone is not sufficient to predict the burstiness.

7 Conclusion

The commonly held point of view in this field is that the Hurst parameter can fully and completely capture the burstiness characteristic of self-similar traffic. Namely, the Hurst parameter indicates the degree of burstiness: the higher the H value, the burstier the traffic. However, the simulation results presented in this paper have shown that the H parameter is not a consistent, monotonic indicator of burstiness. For example, our results show that the

{Traffic B} and {Traffic P} time series derived from the same trace are such that {Traffic P} has a greater Hurst value than {Traffic B}, i.e. {Traffic P} should be more bursty. Yet, the queue length distributions show an 'inversion' since {Traffic B} causes longer queues than {Traffic P}. These contradictory relations hold true for all four traffic traces. Namely, the value of the Hurst parameter does not fully reflect the relative burstiness of {Traffic B} and {Traffic P}, even though they were from the same trace. Moreover, the differences in queueing performance among segments and between the segments and the entire trace confirm the inaccuracy of the H value in predicting queue burstiness. The main conclusion of this paper is that the H parameter alone is not sufficient to fully describe the LRD property of a traffic source and to predict its queueing impact.

References

[1] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, 1994: On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. on Networking*, **2**, no. 1, pp. 1-15, Feb. 1994.

[2] W. Willinger, M.S. Taqqu, M.S. Sherman, and D.V. Wilson, 1996: Self-similarity through high-variability: Statistical analysis of ethernet LAN traffic at the source level (extended version). *IEEE/ACM Trans. on Networking*, **5**, pp. 71-86, December 1996.

[3] Vern Paxson and Sally Floyd, 1995: Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Trans. on Networking*, **3**, no. 3, pp. 226-244.

[4] Mark E. Crovella and Azer Bestavros , 1995: Explaining World Wide Web Traffic Self-Similarity. *Technical Report TR-95-015, Boston University, August 1995* .

[5] Mark E. Crovella and Azer Bestavros , 1996: Self-Similarity in World Wide Web Traffic Evidence and Possible Causes . *1996 ACM SIGMETRICS Int. Conf. Measurement, Modeling of Computer System*, pp. 160-169, May 1996.

[6] Jan Beran, Robert Sherman, M.S. Taqqu and W. Willinger, 1995: Long-Range Dependence in Variable-Bit-Rate Video Traffic. *IEEE/ACM Trans. on Communications* , **2**, no. 4, pp. 1566-1579.

[7] A. Erramilli, O. Narayan, and W. Willinger, 1996: Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Trans. on Networking*, **4**, no. 2, pp. 209-223, April 1996.

[8] R. G. Addie , M. Zukerman, and T. Neame , 1995: Performance of a Single Server Queue with Self Similar Input. *IEEE INFOCOM'95 Seattle, June, 1995.*

[9] H. J. Fowler and W. E. Leland , 1991: Local area network traffic characteristics, with inplications for broadband network congestion management. *IEEE J. Select. Areas Commun.*, **9**, pp. 1139-1149.

[10] D. E. Duffy, A. A. McIntosh, M. Rosenstein, and Walter Willinger , 1994: Statistical analysis of CCSN/SS7 traffic data from working subnetworks. *IEEE J. Select. Areas Commun.*, **12**, no. 3, pp. 544-551.

[11] Darryl Veitch and Patrice Abry, 1997: A Wavelet Based Joint Estimator of the Parameters of Long-Range Dependence, Presented at GRESTI Grenoble, Sept.1997, http://universe.serc.rmit.edu.au/ darryl/.

[12] Patrice Abry, Darryl Veitch, 1998: Wavelet Analysis of Long Range Dependent Traffic, *Trans. Info. Theory*, **44**, no. 1, pp. 2-15.

[13] W. Willinger, M.S. Taqqu, et. al., 1995: Self-Similarity in High-Speed Packet Traffic: Analysis of Modeling of Ethernet Traffic Measurements. *Statistical Science*, **10**, no. 1, pp. 71-86, December 1995.

[14] Trishor S. Trivedi, 1982: Probability & Statistics with Reliability, Queueing, and Computer Science Applications. *Prentice-Hall Inc.*, Englewood Cliffs New Jersey, pp. 543-545.

[15] Van Jacobson, Craig Leres and S. McCanne, 1989: The Tcpdump Manual Page. *Lawrence Berkeley Laboratory*, http://ee.lbl.gov/, ftp://ftp.ee.lbl.gov/tcpdump.tar.Z.