In *Proceedings of the Tenth International Conference on on Parallel and Distributed Computing Systems (PDCS-97),* pages 322-327, October 1997

Consequences of Ignoring Self-Similar Data Traffic in Telecommunications Modeling

P.M. Fiorini and Lester Lipsky
Department of Computer Science and Engineering
University of Connecticut
{pierre, lester}@brc.uconn.edu

Mark Crovella
Deparment of Computer Science
Boston University
crovella@cs.bu.edu

Abstract

This paper is concerned with the performance ramifications of "self-similar" or "fractal" data traffic upon networked systems. It is expected in the near future that communication systems will experience serious performance degregations as a consequence of this phenomena. Thus, the development of tractable models to analyze behaviors exhibited by these processes is important. Interarrival processes which possess self-similar properties are discussed and analytic queueing models to ascertain performance measurements are presented. We conclude by analyzing the performance consequences of self-similar processes upon these systems.

1 Introduction

Recent measurements of data traffic have revealed interesting properties with potentially significant ramifications to the modeling, design, and control of broadband ISDN networks. For instance, Leland et.al. at Bellcore Morristown Research Center have analyzed millions of packets on several Ethernet LAN's and millions of frame data by Variable-Bit-Rate (VBR) video services [9, 10]. In [9, 10] and other studies [1, 2, 3, 7] packet traffic is characterized by variability or "burstyness" across a wide range of scales. That is, network traffic looks the same when measured over time scales ranging from milliseconds to minutes to hours, etc. Data traffic of this type is said to be *self-similar* or *fractal* in nature [9, 10]. One property of these self-similar processes is that $r(k)$, the autocorrelation function lag-k of the number of arrivals per time interval, $N(t)$ (the counting process), converges to zero so slowly such that $\sum_{k=1}^{\infty} r(k) = \infty$ [12]. When the aforementioned occurs it is said that the counting process has the property of *long-range autocorrelations* or *long-range dependencies*. Thus, it can be said, in general, important properties of these processes include: high-variability; counting processes which possess long-range dependencies; and, interarrival processes with long-range autocorrelations.

One way to emulate some of the aforementioned behaviors is by using *power-tail* distributions [8]. Note that power-tail distributions encompass *ALL* Lévi-Pareto functions. It has been shown that the counting process of power-tail distributions possess an autocorrelation function where the lags, k, decay as $\frac{1}{k^{\alpha-1}}$ [12].

In this paper, concepts of self-similarity and long-range dependencies will be discussed and its consequences upon system performance will demonstrated by comparing performance measurements of queues with self-similar interarrival patterns with conventional approaches (i.e. using the $GI/M/1$ queue where GI is, for example, exponential or hyperexponential). Studying the impact of self-similar network traffic upon communication systems is a timely and important research area because many researchers have argued that the ramifications of self-similar network traffic will (if not already) have grave performance consequences upon networked systems in the near future [2, 9, 10]. The paper goes on to discuss one interarrival process, a power-tail distribution merged with a Poisson stream ($PT \otimes \lambda$), developed by [6] which exhibits properties of self-similarity and long-range dependencies its counting and interarrival times. When this process is fed to an exponential server, the response times are computed. These performance measurements are compared to a TPT-32/M/1 and $M/M/1$ queues with comparable means. This is done to illustrate performance bounds for queues which exhibit LRD behavior in their interarrival processes. We conclude by speculating upon the impact of long-range dependencies upon these systems.

1.1 Background

This section presents the background for this paper. The notions self-similarity, power-tail distributions, and long-range autocorrelations in the counting and interarrival processes will be addressed.

1.1.1 Self-Similarity

From our perspective, when the notion of self-similar data traffic is discussed, we mean that the data traffic looks similar when measured across many time scales. It should be noted that stochastic self-similarity is not the same as topological self-similarity. Roughly speaking, topological self-similarity implies that if we examine a small portion, S' of some set S, S' resembles the original set S. In other words, there exists some *exact* affine transformation which maps S' directly onto S.

Stochastic self-similarity is a weaker version of its topological cousin since there can be no *exact* affine transformations. However, one way to quantify self-similarity is to realize that there exists some scaling factor, κ, such that random variables in a self-similar distribution can be generated in the following manner [8]:

$$Z_n = n^{\kappa}[S_n - E(X)], \tag{1}$$

where $\kappa = 1 - \frac{1}{\alpha}$ for $1 < \alpha < 2$. Note that S_n represents a sample average drawn from the distribution, $F_1(\cdot)$. In other words, $S_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ where $X = \{x_1, x_2, \cdots, x_n\}$ is a sequence of iid random variables generated from $F_1(\cdot)$. In this context it can be shown that all distributions are asymptotically self-similar. The random variable Z_n in (2) approaches what has been described as a 4-parameter family of distributions called the α-*stable distributions* [14]. Furthermore note that the α-stable distributions do *not* adhere to convergence properties inherent in distributions which obey the *Central Limit Theorem* (CLT). For instance, random variables which follow the CLT can be generated in the following way:

$$Z_n = n^{\frac{1}{2}}[S_n - E(X)], \tag{2}$$

where again, S_n represents a sample average drawn from some distribution, $F_2(\cdot)$; and, $S_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ where $X = \{x_1, x_2, \cdots, x_n\}$ is a sequence of iid random variables generated from $F_2(\cdot)$.

What equations (1) and (2) are implying is that the α-stable or "self-similar" distributions contract much more slowly when compared to distributions which adhere to the CLT. In other words, in general, it takes many more samples from an α-stable distribution for the random variable Z_n in equation (1) to converge than it does Z_n in equation (2).

1.1.2 Power-Tail Distributions

Power-tail distributions have been used to model for network traffic behavior [8]. One reason is that these distributions inherently exhibit "bursty" behavior. Power-tails have been shown to be self-similar functions, have long-range autocorrelations in the counting process, and possess infinite variance [8, 12]. Power-tail distributions encompass all Lévi-Pareto functions. The power-tail's reliability function, $R(x)$, which is defined to be $Pr(X > x)$, has the following property for large x:

$$R(x) \Rightarrow \frac{c}{x^{\alpha}}, \tag{3}$$

where $\alpha > 0, c > 0$. Its pdf, after differentiation, is $f(x) = \frac{\alpha c}{x^{\alpha+1}}$. Note that the variance does not exist for $f(x)$ if $0 < \alpha \leq 2$. In general, for all the moments, it can be shown that $E(X^{\ell}) = \infty \;\; \forall \;\; \ell \geq \alpha$.

These distributions have been known to exist for a long time. Pareto used them in describing the distribution of wealth in economics. Lèvy showed that all *stable* distributions with infinite variance have power-tails. For a more thorough discussion, the reader is referred to Feller [5]. These distributions have been ignored in the past because an extremely large number of events must occur for the tail to be felt. Only recently, with the arrival of the *information highway* can we expect to see so many events in a short time.

In general, simple power-tail distributions where of the form $f(x) = cx^{\mu-1}/(1+x)^{\alpha+\mu}$ and were difficult to utilize for Laplace transforms, and do not have a direct matrix representation. Greiner, however, developed a family of functions which emulates the power-tail distribution and can be utilized in analytic models generally [8]. The reliability function for a TPT-M distribution is given by

$$R_M(x) = \frac{1-\theta}{1-\theta^M} \sum_{n=0}^{M-1} \theta^n \exp\left(-\frac{\mu x}{\gamma^n}\right), \tag{4}$$

where $0 < \theta < 1$ and $\gamma > 1$. When the limiting function is defined, then Greiner has shown that $R(x)$ satisfies (3), and α is related to θ and γ by $\theta\gamma^{\alpha} = 1$, or $\alpha = -\log(\theta)/\log(\gamma)$ [8]. Furthermore, Greiner has shown that all moments are infinite as $M \to \infty$. We refer to these functions, $R_M(x)$, as *truncated power-tails* or TPT-M *distributions*, because, depending on the size of M, they look like their limit function, the true power-tail, $R(x)$ [8] . But for some large x, depending upon M, they drop off exponentially. These functions are easy to manipulate algebraically. Furthermore, they are m-dimensional *phase* distributions which have vector-matrix representations. See Lipsky or Neuts for a more detailed information concerning matrix representation of phase distributions [11, 13].

It should be noted that by utilizing power-tails as arrival processes to queues, it can be expected that long interarrival times will occur. The reason is that power-tail distributions have the property of infinite variance. That is, assuming some fixed mean, we have

$$\sigma^2 \sim \int_y^\infty xR(x)dx = \infty \quad \forall\, y \geq 0. \qquad (5)$$

One interpretation of Equation (5) is that no matter what y is chosen to be, there will always be a non-negligible probability that large deviations from the mean (i.e. long interarrival times) will occur.

1.1.3 Parameters of the TPT-M Distributions

Physically, the TPT-M distributions are hyperexponential-M distributions which satisfy certain criteria. Consequently, this enables TPT-M distributions to be rich in parameters and facilitates modeling of teletraffic data. The α parameter, in a sense, indicates the degree of self-similarity. The closer α is to 1, the more "bursty" the TPT-M distribution is. Alternatively, as α approaches 2, the system becomes less "bursty". The M parameter of the TPT distribution indicates the degree of truncation. Essentially, where the TPT-M is truncated more or less determines when its reliability function, $R_M(x)$, begins to decay exponentially. The entrance vector, $\mathbf{p}$, of the TPT-M distribution is geometrically distributed with parameter θ. When the power-tail is truncated, this vector has to be normalized by $(1-\theta)/(1-\theta^M)$ so that $\sum_i \mathbf{p}_i = 1$. The parameter γ^n, where n is the phase of the TPT-M distribution whose value ranges from $[0 \ldots M-1]$, is the mean service time in the sub-system for that phase. Note that the values for θ, γ, and α are determined by the relationship $\theta\gamma^\alpha = 1$. The parameter μ is equal to $(1-\theta)/(1-\theta\gamma)$ and can be used as a normalization constant so the mean can be fixed to a specific value.

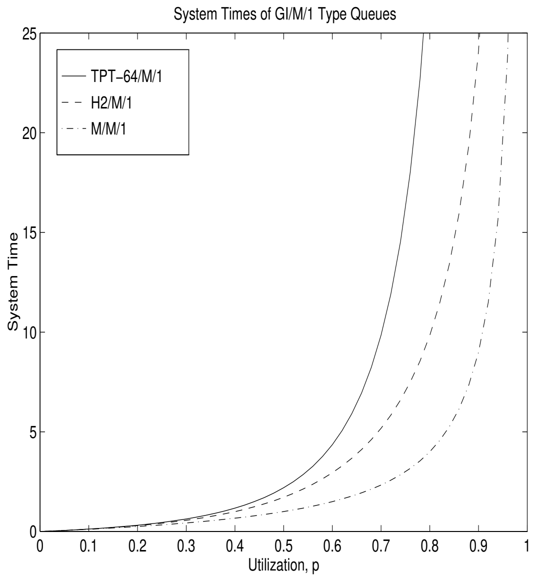

Figure 1: **The System Times of the (TPT-64)/M/1, $H_2/M/1$, and $M/M/1$ Queues**. The power-tail interarrival process (TPT-64) has clearly the worst performance compared to the hyperexponential-2 (H_2) and the exponential arrival streams.

1.2 Performance Consequences of Self-Similar Arrivals

In this section, the impact of self-similar arrivals to a $GI/M/1$ queue will be examined. These results will be compared to $H_2/M/1$ and $M/M/1$ queues with comparable mean interarrival times.

It is clear from the figure that the TPT-64/M/1 queue has the worst performance followed by the $H_2/M/1$ and the $M/M/1$ systems. What is occurring is long-interarrival times are being generated by the TPT-64 distribution. The consequence of this behavior is that this statistically necessitates "bursts" of arrivals which backlogs the queue. Note that as M gets larger for the TPT distribution, performance of the (TPT-M)/M/1 queue will only get worse. What this implies is that if network data traffic is "self-similar", then we can expect serious performance problems. Furthermore, queueing models which attempt to emulate communication systems which utilize conventional techniques (i.e. Batch Poisson, Markov Modulated Poisson Processes (MMPP), etc.) will surely be inadequate to appropriately characterize behavior's of these systems. Some of these reasons are that the aforementioned processes are not self-similar, and do not possess long-range autocorrelations in the count-

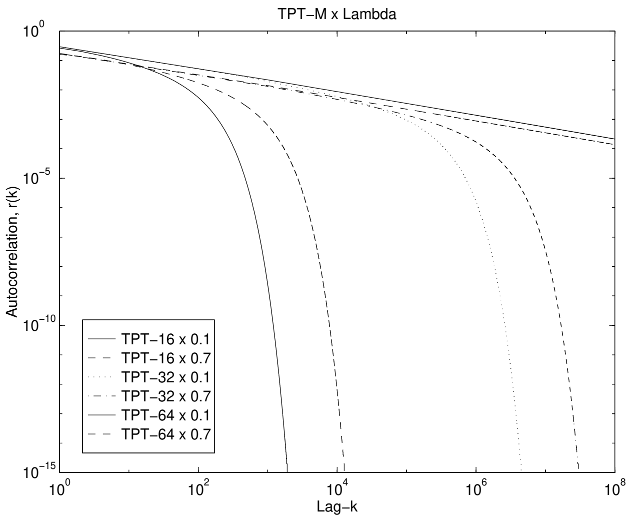

Figure 2: **TPT-M Distributions Merged with Varying amounts of a Poisson Process.** Note that for the TPT-64 $\otimes\lambda$ process, that $r(k)$ appears linear for more than 8 orders of magnitude.

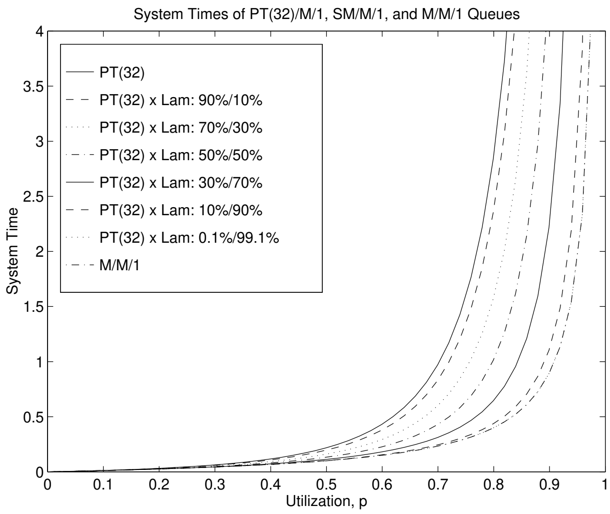

Figure 3: **The system times of the TPT-$32/M/1$, $SM/M/1$, and $M/M/1$ Queues**. The amount of Poisson added to the $PT \otimes \lambda$ process ranges from 10% to 90%.

ing or arrival time processes.

1.3 The $PT \otimes \lambda$ Stream: A Self-Similar Interarrival Process

The $PT \otimes \lambda$ process is a Poisson process superpositioned upon TPT-M distribution. In this section we compute the autocorrelation lag-k for the $PT \otimes \lambda$ interarrival process. This is done with power-tails using various truncation sizes, that is, for $M = 16, 32$, and 64 in order to study its impact upon the autocorrelations. Also, for the TPT-M distributions, parameters such as the interarrival rate, $\bar{x}$, and α are fixed to 1 and 1.4 respectively. α was fixed to 1.4 in order to fit the data in [10]. In order to compute the autocorrelations we utilized the following equation from [?]. The autocorrelation between the Xth and X_{+k}th event (lag-k) can be computed by the following

$$r(k) = \frac{\wp \left[\mathcal{V}[\mathcal{Y}^k - \varepsilon' \wp]\mathcal{V}\right] \varepsilon'}{2\wp\mathcal{V}^2\varepsilon' - [\wp\mathcal{V}\varepsilon']^2}. \tag{6}$$

When examining Figure (6) note that the autocorrelations for the TPT-16$\otimes\lambda$ and TPT-32$\otimes\lambda$ distributions decay relatively quickly as compared to the TPT-64$\otimes\lambda$ distribution. One observation regarding Figure (6) is that the autocorrelations appear to be "long-range". The reason long-range dependencies occur is a result of the "distance" between power-tail interarrival points. In this point process, as more power-tail arrivals are generated, large power-tail interarrival times will eventually occur. The "gaps" generated by large power-tail interarrival times are filled with Poisson arrivals. This behavior tends to cause long-range dependencies since there exists some deviation between the means of the Poisson and the $PT \otimes \lambda$ processes. This, in effect, contributes significantly to the autocorrelations. In essence, what is occurring is that the Poisson arrivals retain the "memory" of long power-tail interarrival times.

1.4 Queueing Performance

In this section, we compute the steady-state system times of the TPT-$32/M/1$, $SM/M/1$, and $GI(SM)/M/1$ queues, analyze queueing performance, and discuss the nexus between power-tails and LRD.

From Figure (3) it is apparent that the TPT-$32/M/1$ queue has the overall worst performance. One way to understand why this is occurring is realize with power-tail distributions, large events will eventually occur. Long interarrival times create "gaps" in the arrival stream which indicate long periods of no arrivals to the system. The consequence of these long periods of no arrivals is that it statistically necessitates intervals where large numbers of arrivals will occur in "bursts". The consequence of these bursts of traffic is that the performance of the system, as a whole, degrades.

Figure (3) also indicates that varying the amount of Poisson which is added to the arrival process affects the performance of the $SM/M/1$ queue. For the TPT-32 and $PT \otimes \lambda$ streams in Figure (3), when the amount of Poisson added to the process is 10%, the behavior of the queue is very similar to that of the TPT-32/$M/1$ queue. This implies that the $PT \otimes \lambda$ interarrival stream has the similar bursty behavior of the power-tail renewal process, however, not quite as severe.

Adding more Poisson to the stream has the effect of subduing the bursty behavior of the $PT \otimes \lambda$ process. The reason is that *fewer* bursts of traffic are needed to engender the mean since Poisson arrivals are breaking up long interarrival time gaps generated by the power-tail distribution. Consequently, as more Poisson is added to the stream, performance of this system improves.

1.4.1 Performance Bounds for the $SM/M/1$ Queue

Figure (3) indicates that the TPT-32/$M/1$ and $M/M/1$ cases represent performance boundaries for the $SM/M/1$ queue. That is, as the percentage of the power-tail distribution is augmented, the performance of the $SM/M/1$ queue approaches that of the TPT-32/$M/1$ queue. Alternatively, as the amount of Poisson added to the $PT \otimes \lambda$ stream is diminished, its performance approaches that of the $M/M/1$ queue. This is interesting since *ALL* the $SM/M/1$ queues exhibit LRD in its interarrival stream. In fact, it has been shown that an infinitesimal amount of Poisson added to a $PT \otimes \lambda$ stream results in LRD in its associated interarrival process [6]. Note that it should be obvious that $SM/M/1$ queueing behavior asymptotically approaches that of the $M/M/1$ system as the percentage of Poisson added to the $PT \otimes \lambda$ process approaches 100%. This becomes evident when the amount of Poisson added to the merged stream is 99.9% as since the performance of the $SM/M/1$ queue is comparable to the $M/M/1$ system. This is interesting since the SM interarrival process in this case *does* exhibit LRD in its interarrival process yet it's performance is nearly identical to that of the $M/M/1$ queue! Thus, the following question needs to be addressed: What is the precise impact of LRD upon queueing performance? Figure (3) indicates when the amount of Poisson added to the $PT \otimes \lambda$ stream is small, then system performance is adversely affected. Thus, if the TPT-32/$M/1$ queue is not considered, it could be inferred that LRD has a deleterious impact upon performance. On the other hand, as the amount of Poisson is augmented in the

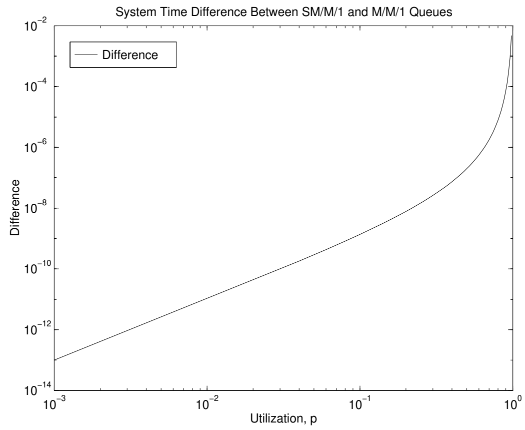

Figure 4: **Comparing the performance of an $SM/M/1$ Queue with LRD to the $M/M/1$ system.** This Figure indicates the differences between the performance the queues. The following question needs to be addressed: What is the precise impact of LRD in the interarrival process?

$PT \otimes \lambda$ process, then it could be implied that LRD does not have a significant impact upon queueing delay since its behavior is similar to that of the $M/M/1$ queue.

When considering performance bounds for this particular $SM/M/1$ queue, remember that both the TPT-32/$M/1$ and the $M/M/1$ queues interarrival processes are renewal processes. In other words, by definition, these arrival streams do *not* possess any correlated behavior. Thus, one conclusion is that an interarrival stream that exhibits LRD behavior could be viewed as one particular characteristic of an interarrival process; and, the magitude of the correlated behavior in the interarrival stream should be the more general indicator of performance degregation.

1.5 Power-Tails and LRD

It is our contention that LRD are a *consequence* of power-tail distributions. It is well known that the autocorrelations in interarrival processes described by Leland [10] and others decay via $r(k) \sim \frac{c}{k^{\alpha-1}}$ where $1 < \alpha \leq 2$. Consequently, the autocorrelations decay so slowly that they tend to infinity. The interpretation of the above is that it has long memory. In other words, dependence between events which are far apart diminishes slowly with increasing lag. We surmise the cause of LRD in interarrival processes merged or mixed with a power-tail is a result of the "distance" between power-tail interarrival points. For example, in the $PT \otimes \lambda$ process, the autocorrelations are caused by the power-tail generating long interarrival time "gaps" or "distances". These gaps are filled with Poisson arrivals and correlations result. LRD in the $PT \otimes \lambda$ process occurs because the autocorrelations retain the memory of long interarrival times generated by the power-tail distribution. As previously mentioned, one property of power-tail distributions is that it has a predisposition to generate large interarrival times because they have infinite variance. The reason long-interarrival times are generated a result of the power-tail distribution experiencing large deviations from the mean. This, in effect, is responsible for the "distance" phenomena.

1.6 Conclusion

This paper has illustrated that teletraffic of the nature described by Leland and others [9, 2] is certainly different; and, one consequence of modeling these systems using conventional methodologies is that overly optimistic performance measurements should be expected. Thus, traditional traffic processes (i.e. Batch Poisson, MMPP, etc.) behaviors must certainly be inadequate to characterize behaviors such as self-similarity and long-range dependencies because they inherently do not possess these properties. In other words, one important property of self-similar functions from a network modeling perspective is that of appropriately characterizing "burstyness"; and, while conventional models do attempt to incorporate this feature, it is unclear whether these models can ever emulate this behavior adequately. One self-similar arrival process with long-range dependencies was discussed (the $PT \otimes \lambda$ stream) and performance measurements were obtained. These results were compared to a power-tail renewal process. Furthermore, it was argued, at least in the $PT \otimes \lambda$ process, that long-range dependencies were a consequence of large power-tail interarrival times.

References

[1] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-range Dependence in Variable-Bit-Rate Video Traffic," *IEEE Trans. Commun.*, vol. 43, pp. 1566-1576, 1995.

[2] Mark E. Crovella and Azer Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes", In *Proceedings of SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, 1996.

[3] D. E. Duffy, A. A. McIntosh, M. Rosenstein, and W. Willinger, "Statistical Analysis of CCSN/SS7 Traffic Data from Working Subnetworks," *IEEE J. Select Areas Commun.,*, vol. 1, no. 3, pp. 544-551, 1994.

[4] A. Erramilli, O. Narayan, and W. Willinger, "Experimental Queueing Analysis with Long-Range Dependent Packet Traffic," *IEEE/ACM Trans. Networking*, Vol. 4, No. 2, 1996.

[5] William Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, John Wiley and Sons, New York, 1971.

[6] Pierre M. Fiorini and Lester Lipsky "A Markov Renewal Process with Long-Range Autocorrelations," Technical Report CSE-TR-97-01, Department of Computer Science and Engineering, University of Connecticut, 1997.

[7] M. W. Garrett and W. Willinger, "Analysis, Modeling, and Generation of Self-Similar VBR Video Traffic," in *Proc. ACM Sigcomm '94*, London, UK, 1994, pp. 269-280.

[8] Michael Greiner, Manfred Jobmann, and Lester Lipsky, "The Importance of Power-Tail Distributions for Telecommunications Traffic Models," Technical Report, Institute für Informatik, Technische Universität München, München, Germany, 1995. (Also submitted for publication in *Operations Research*)

[9] 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.

[10] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Trans. Networking*, Vol. 2, No. 1, pp. 1-15, 1994.

[11] Lester Lipsky, *Queueing Theory: A Linear Algebraic Approach*, MacMillan and Company, New York, 1992.

[12] Lester R. Lipsky and Pierre M. Fiorini, "Auto-Correlation of Counting Processes Associated with Renewal Processes," Technical Report, Booth Research Center, University of Connecticut, 1995.

[13] Marcel F. Neuts, *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins University Press, Baltimore and London, 1981.

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