

LONG-LASTING TRANSIENT CONDITIONS IN SIMULATIONS WITH HEAVY-TAILED WORKLOADS

Mark E. Crovella

Lester Lipsky

Department of Computer Science
Boston University
Boston, MA, 02215 U.S.A.

Department of Computer Science and Engineering
The Taylor L. Booth Center for Computer Applications and Research
University of Connecticut, U-155
Storrs, Connecticut 06269-3155 U.S.A.

ABSTRACT

Recent evidence suggests that some characteristics of computer and telecommunications systems may be well described using *heavy tailed* distributions — distributions whose tail declines like a power law, which means that the probability of extremely large observations is non-negligible. For example, such distributions have been found to describe the lengths of bursts in network traffic and the sizes of files in some systems. As a result, system designers are increasingly interested in employing heavy-tailed distributions in simulation workloads. Unfortunately, these distributions have properties considerably different from the kinds of distributions more commonly used in simulations; these properties make simulation stability hard to achieve. In this paper we explore the difficulty of achieving stability in such simulations, using tools from the theory of stable distributions. We show that such simulations exhibit two characteristics related to stability: slow convergence to steady state, and high variability at steady state. As a result, we argue that such simulations must be treated as effectively always in a transient condition. One way to address this problem is to introduce the notion of time scale as a parameter of the simulation, and we discuss methods for simulating such systems while explicitly incorporating time scale as a parameter.

1 INTRODUCTION

Recently the phenomenon of network traffic *self-similarity* has received significant attention in the networking community (Leland *et al.* 1994). Self-similarity refers to the condition in which a time-series's autocorrelation function declines like a power-law, leading to positive correlations among widely separated observations. Thus the fact that network traffic often shows self-similarity means that it shows noticeable bursts at a wide range of time scales—typically at least four or five orders of magnitude.

A related observation is that the distribution of file sizes in some systems also declines like a power-law—meaning that file sizes also often span many orders of magnitude (Crovella and Bestavros 1996). Both of these conditions have been shown to be well described using distributions that are *heavy tailed*—distributions whose tails follow a power law.

Heavy tailed distributions behave quite differently from the distributions more commonly used to describe characteristics of computing systems, such as the Normal distribution and the exponential distribution, which have tails that decline exponentially (or faster). In contrast, because their tails decline relatively slowly, the probability of very large observations occurring when sampling random variables that follow heavy tailed distributions is non-negligible. In fact, the distributions we discuss in this paper have *infinite variance*, reflecting the extremely high variability that they capture.

As a result, designers of computing and telecommunication systems are increasingly interested in employing heavy-tailed distributions to generate workloads for use in simulation. However, simulations employing such workloads may show unusual characteristics; in particular, they may be much less stable than simulations with less variable inputs. In this paper we discuss the kind of instability that may be expected in simulations with heavy-tailed inputs and show that they may exhibit two features: first, they will be very slow to converge to steady state; and second, they will show highly variable performance at steady state. To explain and quantify these observations we rely on the theory of *stable* distributions (Feller 1971, Samorodnitsky and Taqqu 1994).

To deal with the slow convergence of these simulations to steady state, we suggest that simulations explicitly incorporate the notion of time scale as a parameter. This means that researchers recognize that their simulations are not achieving steady state, but rather finite-time-scale approximations to steady state. To do so requires methods to measure the

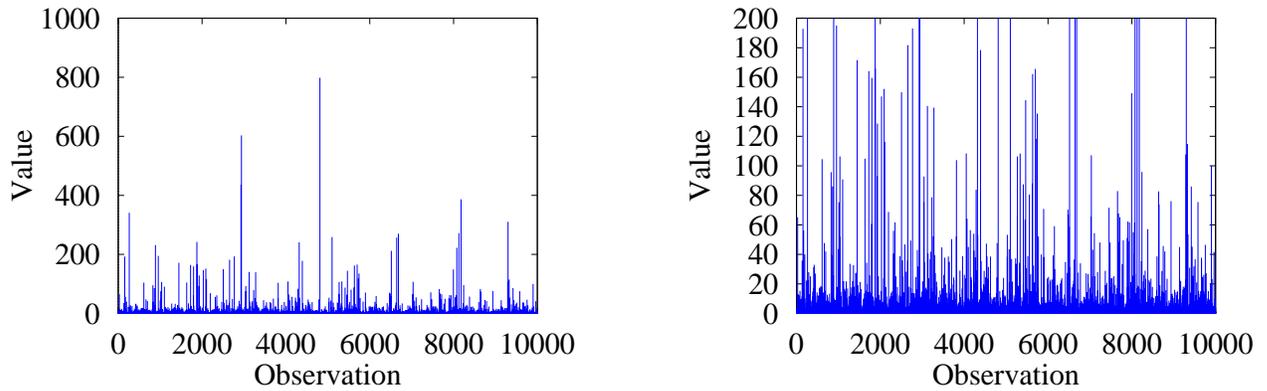


Figure 1: Sample Data from Heavy Tailed Distribution with $\alpha = 1.2$

movement toward steady state in a simulation as a function of the number of observations that have been made of the heavy tailed random variable. We suggest simple first steps in this regard by using order statistics for heavy random variables.

In general however many of the problems associated with the simulations using heavy-tailed workloads seem quite difficult to solve. This paper does not primarily suggest solutions but rather draws attention to these problems, both to yield insight for researchers using simulation and to suggest areas in which more research is needed. As a result we conclude with a summary of the issues that should be addressed when using simulations with heavy-tailed workloads.

2- HEAVY TAILED DISTRIBUTIONS

2.1- Background

Let X be a random variable with cdf $F(x) = P[X \leq x]$ and complementary cdf (ccdf) $\bar{F}(x) = 1 - F(x) = P[X > x]$. We say here that a distribution $F(x)$ is *heavy tailed* if

$$\bar{F}(x) \sim cx^{-\alpha} \quad 0 < \alpha < 2 \quad (1)$$

for some positive constant c , where $a(x) \sim b(x)$ means $\lim_{x \rightarrow \infty} a(x)/b(x) = 1$. If $F(x)$ is heavy tailed then X shows very high variability. In particular, X has infinite variance, and if $\alpha \leq 1$, X has infinite mean. Section 2.2 will explore the implications of infinite moments in practice; here we note simply that if $\{X_i, i = 1, 2, \dots\}$ is a sequence of observations of X then the sample variance of $\{X_i\}$ as a function of i will tend to grow without limit, as will the sample mean if $\alpha \leq 1$.

The simplest heavy tailed distribution is the *Pareto* distribution which is power-law over its entire range. The Pareto distribution has pmf

$$p(x) = \alpha k^\alpha x^{-\alpha-1} \quad 0 < k \leq x$$

and cdf

$$F(x) = P[X \leq x] = 1 - (k/x)^\alpha \quad (2)$$

in which the positive constant k represents the smallest possible value of the random variable.

In practice, random variables that follow heavy tailed distributions are characterized as exhibiting many small observations mixed in with a few large observations. In such datasets, most of the observations are small, but most of the contribution to the sample mean or variance comes from the few large observations.

This effect can be seen in Figure 1, which shows 10,000 synthetically generated observations drawn from a Pareto distribution with $\alpha = 1.2$ and mean $\mu = 6$. On the left hand side of the figure the scale allows all observations to be shown; on the right the y axis is expanded to show the region from 0 to 200. These figures show the characteristic, visually striking behavior of heavy tailed random variables. From the left plot it is clear that a few large observations are present, some on the order of hundreds to one thousand; while from the right plot it is clear that most observations are quite small, typically on the order of tens or less.

An example of the effect of this variability on sample statistics is shown in Figure 2. This figure shows the running sample mean of the data points from Figure 1, as well as a level line showing the mean of the underlying distribution (6). Note that the sample

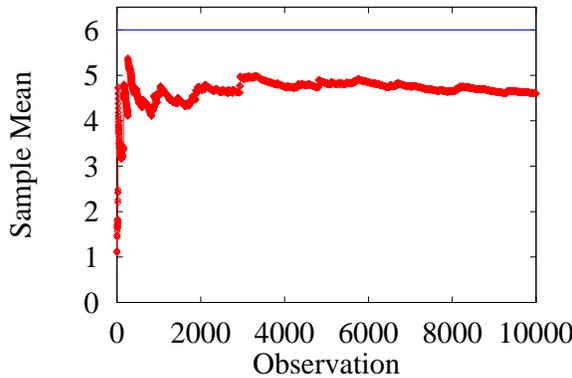


Figure 2: Running Mean of Data from Figure 1

mean starts out well below the distributional mean, and that even after 10,000 observations it is not close in relative terms to the distributional mean.

2.2- Heavy Tails in Computing Systems

A number of recent studies have shown evidence indicating that aspects of computing and telecommunication systems can show heavy tailed distributions. Measurements of computer network traffic have shown that autocorrelations often show heavy tails; this is the phenomenon of *self similarity* (Garrett and Willinger 1994, Leland *et al.* 1994). Measurements of file sizes in the Web (Arlitt and Williamson 1996, Crovella and Bestavros 1996) and in I/O patterns (Peterson 1996) have shown evidence that file sizes can show heavy tailed distributions. And finally, the CPU time demands of Unix processes have also been shown to follow heavy tailed distributions (Harchol-Balter and Downey, 1996).

The presence of heavy tailed distributions in measured data can be assessed in a number of ways. The simplest is to plot the ccdf on log-log axes, and visually inspect the resulting curve for linearity over a wide range (several orders of magnitude). This is based on Equation (1), which can be recast as:

$$\lim_{x \rightarrow \infty} \frac{d \log \bar{F}(x)}{d \log x} = -\alpha$$

so that for large x , the ccdf of a heavy tailed distribution should appear to be a straight line on log-log axes with slope $-\alpha$.

An example empirical dataset is shown in Figure 3, which is taken from Crovella and Bestavros (1996). This figure is the ccdf of file sizes transferred through the network due to the Web, plotted on log-

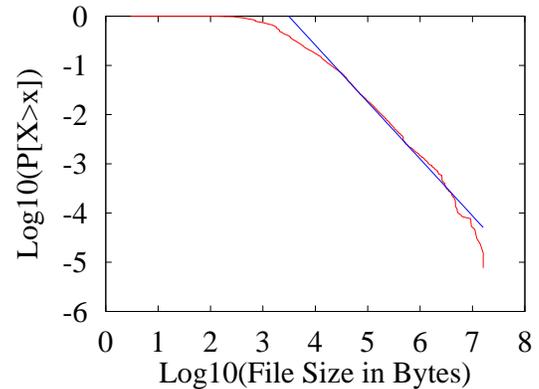


Figure 3: Log-log Complementary Distribution of Sizes of Files Transferred Through the Web

log axes. The figure shows that the file size distribution appears to show power law behavior over approximately three orders of magnitude. The slope of the line fit to the upper tail is approximately -1.2 , yielding $\hat{\alpha} \approx 1.2$.

3- STABILITY- IN- SYSTEMS- WITH HEAVY TAILED WORKLOADS

As heavy tailed distributions are increasingly used to characterize workload characteristics of computing systems, researchers interested in simulating such systems are beginning to use heavy tailed inputs to simulations. For example, Paxson (1995) describes methods for generating self-similar time series for use in simulating network traffic and Park, Kim, and Crovella (1996) use heavy-tailed file sizes as inputs to a network simulation. However, an important question arises: how stable are such simulations? This can be broken down into two questions:

1. How long until such simulations reach steady state?, and
2. How variable is system performance at steady state?

In this section we will show that if simulation outputs are dependent on all the moments of the distribution F then the answers to the above questions can be surprising. Essentially, we show that such simulations can take a very long time to reach steady state; and that such simulations can be much more variable at steady state than is typical for traditional systems.

Note that some simulation statistics may not be directly affected by all the moments of the distribution F , and our conclusions do not necessarily apply

to those cases. For example, the mean number of customers in an $M/G/\infty$ queueing system may not show unusual behavior even if the service time distribution F is heavy tailed because that statistic only depends on the mean of F . Also, some measured statistics may not be made unstable by high variance, but rather may be a measure of high variance—*e.g.*, the Hurst parameter H which measures the high variance in burst length of self-similar network traffic.

Since not all simulation statistics will be affected by heavy tailed workloads, we choose a simple statistic to show the generality of our observations: the sample mean of the heavy tailed inputs. Since our results apply to the sample mean of the input, we expect that any system property that behaves like the sample mean should show similar behavior. For example, assume we want to achieve steady state in a particular simulation. This implies that the measured system utilization λ/\bar{x} (where λ is the average interarrival time and \bar{x} is the sample mean of service times over some period) should be close to the desired system utilization ρ . For this to be the case, \bar{x} must be close to its desired mean μ .

To analyze the behavior of the sample mean, we are concerned with the convergence properties of sums of random variables. The normal starting point for such discussions would be the Central Limit Theorem (CLT). Unfortunately, the CLT applies only to sums of random variables with finite variance, and so does not apply in this case. In the place of the CLT we instead have limit theorems for heavy tailed random variables first formulated by Lévy (Feller 1971, Samorodnitsky and Taqqu 1994).

To introduce these results we need to define the notation $A \xrightarrow{d} B$ which means that the random variable A converges in distribution to B (roughly, has distribution B for large n). Then the usual CLT can be stated as: for X_i i.i.d. and drawn from some distribution F with mean μ and variance $\sigma^2 < \infty$, define

$$A_n = \frac{1}{n} \sum_{i=1}^n X_i$$

and

$$Z_n = n^{-1/2}(A_n - \mu); \quad (3)$$

then

$$Z_n \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad (4)$$

where $\mathcal{N}(0, \sigma^2)$ is a Normal distribution.

However, when X_i are i.i.d. and drawn from some distribution F that is heavy tailed with tail index $1 < \alpha < 2$, then if we define

$$Z_n = n^{1-1/\alpha}(A_n - \mu) \quad (5)$$

we find that

$$Z_n \xrightarrow{d} \mathcal{S}_\alpha \quad (6)$$

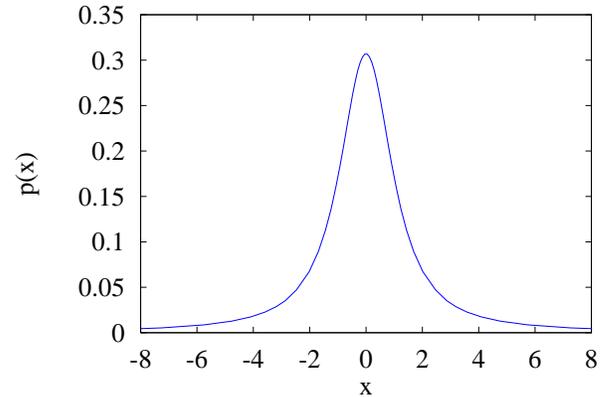


Figure 4: Pmf of an α -Stable Distribution

where \mathcal{S}_α is an α -Stable distribution. The α -Stable distribution has four parameters: α , a location parameter (analogous to the mean), a scale parameter (analogous to the standard deviation), and a skewness parameter. Based on the value of the last parameter, the distribution can be either skewed or symmetric. A plot of the symmetric α -Stable distribution with $\alpha = 1.2$ and location zero is shown in Figure 4. From the figure it can be seen that this distribution has a bell-shaped body much like the Normal distribution but that it has much heavier tails. In fact the α -Stable distribution has power-law tails that follow *the same* α as that of the distribution F from which the original observations were drawn.

From Equations (5) and (6) we can make two observations about the behavior of sums of heavy tailed random variables. First, Equation (5) states that such sums may converge much more slowly than is typical in the finite variance case. Second, Equation (6) states that even after convergence, the sample mean will show high variability—it follows a heavy tailed distribution.

These effects can be seen graphically in Figure 5. This figure shows histograms of A_n for varying values n . On the left we show the case in which the X_i s were drawn from an Exponential distribution; on the right we show the case in which the X_i s were drawn from a strictly positive heavy tailed distribution with $\alpha = 1.4$; in both cases the mean of the underlying distribution was 1. The plot on the left shows that the most likely value of the sample mean is equal to the true mean, even when summing only a small number of samples. In addition, it shows that as one sums larger numbers of samples, the sample mean converges quickly to the true mean. However, neither of these observations are true for the case of the

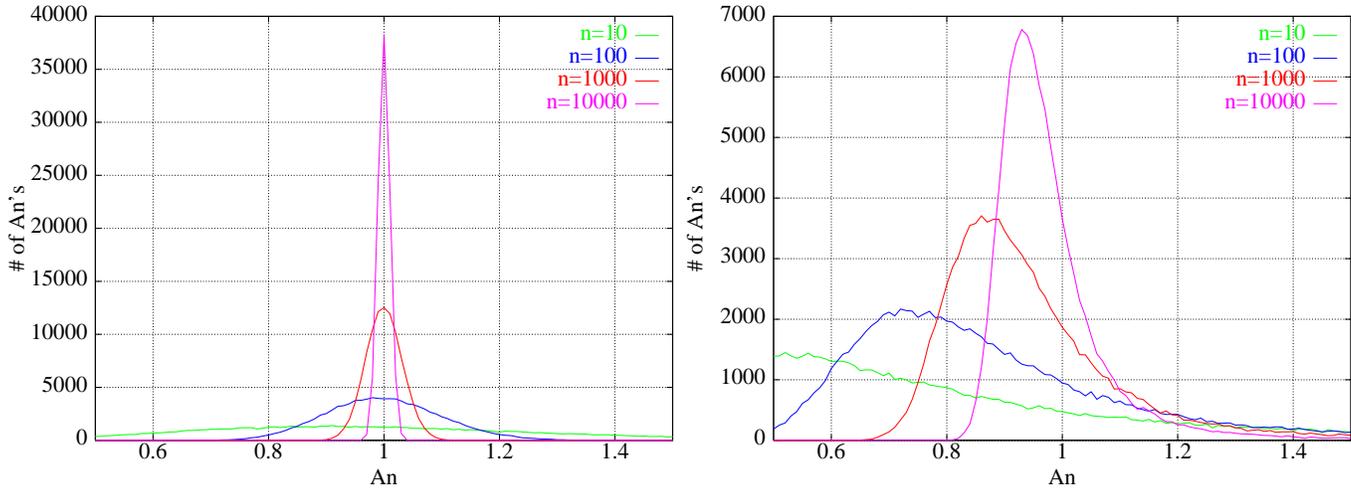


Figure 5: Histogram of A_n as n varies for Exponential (left) and Heavy Tailed (right) Random Variables.

heavy tailed distribution on the right. When summing small numbers of samples, the most likely value of the sample mean is far from the true mean, and the distribution progresses to its final shape rather slowly.

Thus we have seen that the convergence properties of sums of heavy tailed random variables are quite different from those of finite variance random variables. We relate this to steady state in simulation as follows: presumably for a simulation to reach steady state, it must at a minimum have seen enough of the input workload to observe its mean. Of course it may be necessary for much more of the input to be consumed before the simulation reaches steady state, so this condition is a relatively weak one. Still, we show in the next two subsections that this condition has surprising implications for simulations.

3.1- Slow Convergence to Steady State

In Equation (5), Z_n represents a constant (since for large n , Z_n converges in distribution). Thus another way of formulating Equation (5) is:

$$|A_n - \mu| \sim n^{1/\alpha-1}.$$

In this form it is more clear how slowly A_n converges to μ . If α is close to 1, then the rate of convergence, measured as the difference between A_n and μ , is very slow—until, for $\alpha = 1$, the average does not converge at all, reflecting the fact that the mean is infinite.

Suppose one would like to use A_n to form an estimate of the mean μ that is accurate to k digits. Alternatively, one might state that a simulation has reached steady state when the observed mean of the input A_n agrees with μ to k digits. Then we would

like

$$|A_n - \mu|/\mu \leq 10^{-k}.$$

Now, as a rough approximation:

$$|A_n - \mu| = c_1 n^{1/\alpha-1}$$

for some positive constant c_1 . Then we find that:

$$n \geq c_2 10^{\frac{k}{1-1/\alpha}}.$$

We can say that given this many samples, k digit accuracy is “highly likely.”

For example, assume we would like 2-digit accuracy in A_n , and suppose $c_2 \approx 1$. Then the number of samples n necessary to achieve this accuracy is shown in Table 1. This table shows that as $\alpha \rightarrow 1$, the number of samples necessary to obtain convergence in the sample mean explodes. Thus, it is *not feasible* in any reasonable amount of time to observe steady state in such a simulation as we have defined it. Over any reasonable time scale, such a simulation is *always in transient state*.

Table 1: Number of Samples Necessary to Achieve 2 Digit Accuracy in Mean as a Function of α

α	n
2.0	10,000
1.7	72,000
1.5	1,000,000
1.2	10^{12}
1.1	10^{22}

3.2- High Variability at Steady State

Equation (6) shows that even at steady state, the sample mean will be distributed according to a heavy

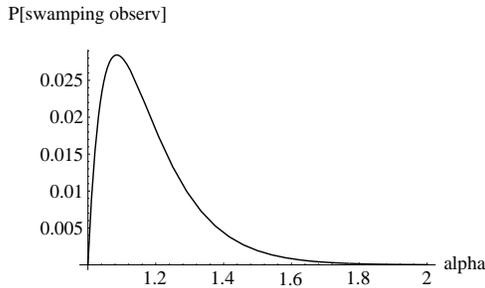


Figure 6: Probability of a Swamping Observation in 10^5 Inputs as a Function of α

tailed distribution, and hence will show high variability. Thus, the likelihood of an erroneous measurement of μ is still non-negligible. Equivalently, the simulation still behaves erratically.

To see this more clearly, let us define a *swamping* observation as one whose presence causes the estimate of μ to be at least twice as large as it should be. That is, if we happen to encounter a swamping observation in our simulation, the observed mean of the input will have a relative error of at least 100%.

In a simulation consisting of n inputs, a swamping observation must have value at least $n\mu$. Let us assume that the inputs are drawn from a Pareto distribution. Such a distribution has $\mu = k\alpha/(\alpha - 1)$. Then the probability $p_{n\mu}$ of observing a value of $n\mu$ or greater is

$$p_{n\mu} = P[X > n\mu] = \left(\frac{k}{nk\alpha/(\alpha - 1)} \right)^\alpha = \left(\frac{\alpha - 1}{n\alpha} \right)^\alpha$$

and the probability p of observing such a value at least once in n trials is

$$p = 1 - (1 - p_{n\mu})^n.$$

Figure 6 shows a plot of p as a function of α for $n = 10^5$. (The figure is not significantly different for other values of n , e.g., $10^6, 10^7$.) It shows that even in a relatively long simulation, the probability of a swamping observation is not negligible; when α is below about 1.3, such an observation could occur more often than once in a hundred simulations. The probability declines very rapidly for $\alpha < 1.1$ not because the variability of the simulation is declining, but because of the way we have defined the swamping observation: in terms of the distributional mean. When $\alpha = 1$, the mean is infinite, and so it becomes impossible to observe a value greater than the mean.

Taken together, Table 1 and Figure 6 also provide some insight into the value of α above which

it may be possible to obtain convergent, consistent simulations. The table shows that simulation convergence becomes impractical when α is somewhere in the region between 1.7 and 1.5; and the Figure shows that simulations become erratic at steady state in approximately the same region. As a result, we can conclude that the difficulties inherent in simulations with heavy tailed inputs are likely to be particularly great when α is less than about 1.7; and that when α is greater than or equal to about 1.7 it may be feasible (given sufficient computing effort) to obtain consistent steady state in simulation.

4- TIME SCALE AS A SIMULATION PARAMETER

The observations in the previous section suggest that for simulations involving heavy tailed workloads, steady state is elusive—especially when α is less than approximately 1.7. One way to address this in simulation is to explicitly incorporate the notion of time scale. Doing so recognizes that no system ever sees an infinite sized input, or runs for an infinite amount of time. Instead, real systems experience *finite-time-scale* approximations to steady state. Therefore all sample moments of the input will be finite, even when some moments of the underlying distribution are infinite.

To do this, we need some way to relate the infinite-moment underlying distribution (one with infinite support) to the proper finite-moment distribution (one with finite support) as a function of the amount of time that the simulation runs. If we express the amount of time that the simulation runs in terms of the number of observations made of the input random variable, then one way to address this problem is through *order statistics*. Order statistics provide estimates of quantities such as the largest out of n observations of some random variable (David 1981).

Let $X_i, i = 1, 2, \dots, n$ be i.i.d. samples from some distribution $F(x)$. Define $Y = \max_n X_i$; then the distribution of Y is $F^n(x)$. If F is heavy tailed, then $E[Y]$ exists if $\alpha > 1$. For example, consider the Pareto distribution defined in Equation (2). Then:

$$E[Y] = knB(n, 1 - 1/\alpha) \approx E[X]n^{1/\alpha}$$

where B is the Beta function (Lipsky, Zhang, and Kang 1996). Then, over a sequence of n observations, the X_i s may be considered to be drawn from a finite-support distribution bounded above by $E[Y]$.

In addition, this allows use to analyze systems whose workloads are heavy tailed, but with bounded support. For example, a Web server will have some largest file—this constitutes the upper bound on its

file size distribution. Call this upper bound B . Then if $E[Y] \ll B$, the system behaves as if the X_i s were drawn from an unbounded distribution. But if the simulation runs long enough for $E[Y]$ to approach or exceed B , then the system behaves as if the X_i s were drawn from the bounded distribution.

5 CONCLUSIONS

We have shown that a difficult problem arises when simulating systems with heavy tailed workloads. In such systems, steady-state behavior can be elusive, because average-case behavior depends on the presence of many small observations as well as a few large observations.

This problem has two implications:

1. Since a number of large but rare observations must occur before average case behavior is evident, convergence of a simulation to steady state may be slow. It may not be possible in any reasonable time to achieve steady state.
2. Since many small observations must occur to balance the presence of large observations, large observations can have a dominating effect on performance results even at steady state. Simulations may still behave erratically even at steady state.

Our results indicate the researchers running simulations using heavy tailed workloads with α less than about 1.7 should consider carefully the stability of their results. In those cases, a fruitful approach may be to incorporate time scale explicitly into the stability analysis using techniques from order statistics.

ACKNOWLEDGMENTS

This research was supported in part by NSF grant CCR-9501822 and by a grant from Hewlett-Packard Laboratories.

REFERENCES

- Martin F. Arlitt and Carey L. Williamson. 1996. Web server workload characterization: The search for invariants. In *Proceedings of the 1996 SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pages 126–137.
- Mark E. Crovella and Azer Bestavros. 1996. Self-similarity in World Wide Web traffic: Evidence and possible causes. In *Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 160–169.

- H. A. David. 1981. *Order Statistics*. John Wiley and Sons, Inc.
- William Feller. 1971. *An Introduction to Probability Theory and Its Applications*, volume II. Second edition. John Wiley and Sons.
- Mark Garrett and Walter Willinger. 1994. Analysis, modeling, and generation of self-similar VBR video traffic. In *Proceedings of ACM SIGCOMM '94*.
- Mor Harchol-Balter and Allen Downey. 1996. Exploiting process lifetime distributions for dynamic load balancing. In *Proceedings of SIGMETRICS '96*, pages 13–24.
- 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 Transactions on Networking*, 2:1–15.
- Lester Lipsky, Tao Zhang, and Seonglin Kang. 1996. On the performance of parallel computers: Order statistics and Amdahl's law. *International Journal of Computers and Their Applications*.
- Kihong Park, Gi Tae Kim, and Mark E. Crovella. 1996. On the relationship between file sizes, transport protocols, and self-similar network traffic. In *Proceedings of the Fourth International Conference on Network Protocols (ICNP'96)*, pages 171–180.
- Vern Paxson. April, 1995. Fast approximation of self-similar network traffic. Technical Report LBL-36750, Lawrence Berkeley National Laboratory.
- David L. Peterson. December, 1996. Data center I/O patterns and power laws. In *CMG Proceedings*.
- Gennady Samorodnitsky and Murad S. Taqqu. 1994. *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Chapman and Hall, New York.

AUTHOR BIOGRAPHIES

MARK CROVELLA is Assistant Professor of Computer Science at Boston University. He received his B.S. from Cornell University in 1982 and his Ph.D. from the University of Rochester in 1994. His research interests include performance evaluation of networks and parallel systems. Recently he has focused on performance evaluation of the Internet, the Web, and Web servers. He has given numerous invited talks on network performance evaluation and self-similarity, tutorials on performance characteristics of the Web, and has had performance models based on his work adopted by ADSL Forum.

LESTER LIPSKY is Professor of Computer Science and Engineering at the University of Connecticut. He received a BME from C.C.N.Y., and earned his Ph.D. in Theoretical Atomic Physics from the

University of Connecticut in 1965. Professor Lipsky's major research interests are in the analytical modelling of the performance of computer systems and networks, and related mathematical and numerical problems. In recent years, he has become increasingly interested in the theoretical aspects of queueing theory, particularly non-steady state and "chaotic" phenomena, using a linear-algebraic formulation. He has authored or co-authored over 100 research articles. His book, *QUEUEING THEORY: A Linear Algebraic Approach*, was published by Macmillan, NY in 1992. He is a member of the A.C.M., I.E.E.E., the American Physical Society, Sigma Xi, Upsilon Pi Epsilon and Sigma Pi Sigma.