EXAMINE THE SCHEDULING IN INTERNET QUEUEING SYSTEMS WITH THE TRANSFORM APPROXIMATION METHOD

B.Balamurugan[1]

Abstract: *The underlying arrival and service distributions of Internet traffic are drawn from a family of distributions, which are known as heavy-tailed distributions. These distributions have tails that decay much slower than exponential and do not possess a closed form Laplace transform. We have developed an analysis method, known as the Transform Approximation Method (TAM), to overcome this problem. TAM has been successfully applied to the analysis of the queue length process of systems with heavy-tailed service distributions and to systems with heavy-tailed arrivals. In this paper, we examine the use of TAM in analyzing the waiting time process in systems with heavy-tailed service times. The original version of TAM was modified and successfully applied to this problem.*

Key words: Scheduling in internet, Queueing models, Transform Approximation Method

1. INTRODUCTION

In a recent Business Week article, [1] a mathematical concept known as fractals was listed as one of the top ten technologies of the future. This article referenced many potential applications for fractals, including the Internet. Fractals are applicable when the underlying process being mathematically modeled has a similar appearance regardless of the time or observation scale.

In many modern applications of queueing theory, the classical assumption of exponentially decaying service distributions does not apply. Often, appropriate distributions for modeling in these areas are *heavy-tailed.* [8]Such distributions decay more slowly than any exponential function, giving a non-trivial probability of an extremely large event

Why do these distributions limit the use of queueing theory results? Many of the available results from queuing theory require the existence of the Laplace transform of the underlying interarrival or service time distributions. The Pareto (P), Lognornal (LN), and Weibull (W) probability distributions do not possess closed-form mathematical expressions for their Laplace transform. Thus, standard queueing results cannot be applied.

We shall use the one-parameter (shape only) version of the Pareto given by

$$F(x) = 1 - \frac{1}{\left[(1+x)^{\alpha}\right]} \qquad (x \geq 0),$$

where α is the shape parameter. The corresponding density function is

$$f(x) = \frac{\alpha}{\left[(1+x)^{\alpha+1}\right]} \qquad (x \geq 0).$$

A major consequence of heavy-tailed behavior of the Pareto is the disappearance of moments. It is easy to see that for a Pareto to have its k^{th} moment, $E[X^k]$, we need $\alpha > k$. If $\alpha > 1$, then the mean, $E[X]$, is $E[X] = 1/(\alpha - 1)$ and if $\alpha > 2$, it follows that $E[X^2] = \frac{2}{[(\alpha-2)(\alpha-1)]}$.

Thus, no matter what the value of the parameter α, a Pareto random variable cannot have all its moments and, hence, does not have an analytic Laplace transform. Therefore, since the Laplace transform plays such a vital role in the mathematical analysis of queueing systems, analysis possibilities become quite limited when the interarrival or service time distribution does not have an analytic Laplace transform.

Much of this research has focused on phase-type approximations to the heavy-tailed distributions. It requires fitting the approximation to the underlying distribution and this may require the determination of a large number of constants. But, as pointed out by Feldmann and Whitt, [2] there is a serious need to investigate other potential congestion-analysis methods that do not depend on fitting approximating distributions. In this paper, we present another possible approach that uses the actual heavy-tailed distributions empirically.

In the following sections of this paper, we present the TAM and some basic required results from the M/G/1 queue; the development of a recursive method coupled with TAM; some numerical examples comparing the two TAM sampling techniques, as well as comparisons with simulation; and a few concluding remarks.

The Transform Approximation Method

To see how the TAM works, If f (x) is heavy-tailed, then $f^*(s)$ does not have a closed form. Harris and Marchal [13] propose approximating $f^*(s)$ by,

$$f^*(s) = \int_0^\infty e^{-sx} f(x)\,dx \qquad \mathrm{Re}(s) \geq 0.$$

$$f^*(s) \approx \frac{1}{N}\sum_{i=1}^{N} e^{-sx(i)}$$

where the $x(i)$, $i = 1, 2, \ldots, N$, are chosen as an equi probable N-point discrete approximation for the original random variable X such that $F(x(i)) = i/(N+1)$. We call x(i) the TAM samples. When the underlying distribution is a Pareto, it follows from the form of the Pareto CDF that

$$x(i) = \frac{1}{\left[1 - i/(N+1)\right]^{1/\alpha}} - 1.$$

The Weibull also has a simple representation for the TAM samples given by

$$x(i) = \beta\left[-\ln\left(1-\frac{i}{N+1}\right)\right]^{1/\alpha}.$$

Unfortunately, the Lognormal does not have a simple representation for the TAM samples, so we used Microsoft Excel's Visual Basic code and its Lognormal inversion module to generate the M/LN/1 results. We call this method of generating the TAM sample points the Uniform Transformation Approximation Method (UTAM), since the sampling points are chosen uniformly.

Fischer and Harris [3] first investigated the use of UTAM on heavy-tailed queueing systems. Harris et al. [4] further examined queues with heavy-tailed interarrival distributions, while Brill et al. [5] examined the use of UTAM in analyzing the queue length process in M/G/1 queues where the service distribution had heavy-tails. In this paper, we give an improved version of TAM and examine its use in analyzing the queue waiting time process in this queue.

Following Gross and Harris, [6] let $Wq^*(s)$ be the Laplace transform of the queue wait in M/G/1. $Wq(t)$ is the cumulative distribution function of the waiting time in the queue; then, for Re(s)≥0 we have

$$Wq^*(s) = \int_0^\infty e^{-st} Wq(t)dt = \frac{(1-\rho)}{s-\lambda(1-B^*(s))}. \tag{1}$$

It represents the number of arrivals during an average service time. Also in Equation (1), the parameter s is a complex number and for $Wq^*(s)$ to be well defined mathematically we restrict the real part of s to be non-negative. Without this restriction $Wq^*(s)$ may not be defined.

For $Wq^*(s)$, we need the Laplace transform of the service time distribution, $B^*(s)$, and heavy-tailed distributions may not have one in closed form. Thus, any numerical procedure to invert $Wq^*(s)$ is very difficult. What we do is replace $B^*(s)$ with an empirical version using TAM samples. That is $B^*(s) = \sum_{i=1}^{N} e^{-sx(i)} / N$,

where $x(i)$ is the TAM sample from the appropriate service distribution.

A standard way to invert $Wq^*(s)$ is known as the Fourier-Series method. Abate and Whitt [7] have studied that method when applied to the M/G/1 queue. They require the service time have a Laplace transform, which is not available in the case of heavy-tailed service times. Of course, one could use TAM and the Fourier- Series method discussed in Abate and Whitt. [7] That approach will be discussed in a forthcoming paper. [8] Here we use a modification of the method presented in Fischer and Knepley [9] and TAM to study *Wq(t)*.

The TAM Recursion for Computing the Cumulative Distribution Function of the Waiting Time in the Queue

The method presented in Fischer and Knepley [20] can be used to invert $Wq^*(s)$. It changes the continuous Laplace transform $Wq^*(s)$ into a generating function. Let $T > 0$ be some small constant, and consider the transformation $z = e^{-sT} \sim (1 - sT)$. The procedure uses Equation (1) to develop the generating function $R(z) = \sum_{n=0}^{\infty} r_n z^n$,

where $r_n = Wq(nT)$. The quantity T is a sampling interval of Wq(t), where the smaller values of T give more accuracy. The technique developed by Fischer and Knepley [9] was used to analyze several queueing systems, but its application to the M/D/1 queue is most appropriate on this problem. It has to be extended for our use, but serves as background and helps with understanding. Let b be the deterministic service time in M/D/1, then $B^*(s) = e^{-sb}$ we then have

$$Wq^*(s) = \int_0^{\infty} e^{-st} Wq(t)dt = \frac{(1-\rho)}{s - \lambda(1 - e^{-sb})}.$$

For a given T, let k be a positive integer such that $kT=b$. Note the selection of k depends on T and b, but can be accomplished as long as b has a finite number of decimal points. We consider the transformation $z = e^{-sT}$, since $kT=b$ then $e^{-sb} = z^k$. Furthermore, we have

$$Wq^*(s) = \int_0^{\infty} e^{-st} Wq(t)dt \approx T\sum_{n=0}^{\infty} e^{-snT} Wq(nT)$$
$$\approx T\sum_{n=0}^{\infty} r_n z^n = TR(z).$$

Combining this equation and the M/D/1 equation given above for $Wq^*(s)$ with $s \approx (1-z)/T$ we get

$$\sum_{n=0}^{\infty} r_n z^n = \frac{(1-\rho)}{(1 - \lambda T - z + \lambda T z^k)}.$$

Equating coefficients of z^n yields $r_0 = (1-\rho)/(1-\lambda T)$,

for n = 1,2 , ..., k-1 $r_n = r_{n-1}/(1-\lambda T)$, and for n = k, k+1, k+2 , ... $r_n = (r_{n-1} - \lambda T r_{n-k})/(1-\lambda T)$.

In order to invert Equation (1) with $B^*(s) = \sum_{i=1}^{N} e^{-sx(i)} / N$,

we have to extend the method just presented for M/D/1 to the situation where there is a sum of exponentials in the denominator. Let *x(i)* be the TAM samples from the underlying service distribution and *k(i)* be an integer such that *T k(i) ~ x(i)* for all *i* (specifically, let *k(i)=round(x(i)/T)*). We then have

$$B^*(s) = \sum_{i=1}^{N} e^{-sx(i)} / N = \sum_{i=1}^{N} e^{-sk(i)T} / N .$$

The selection of *k(i)* is dependent on *T* and the desired level of accuracy. (Note in the M/D/1 case we had one k such that *kT=b*.) Next we set *c(n)* equal to the number of *k(i) = n*. Since $z = e^{-sT}$, we have $B^*(s) = \sum_{n=0}^{\infty} c(n) z^n / N$.

Not all the c(n) is positive; in fact, many may be zero. Also, if *N* is large and *T* is small, the highest value of n such that *c(n)>0* may be large. In addition, for a given *T* and *N*, there is an N^* such that *c(n)= 0* for all $n > N^*$. Thus, *B*(s)* is in reality a finite sum; but, for ease of notation, we will keep the upper limit as infinity.

Proceeding just as we did for M/D/1, we have

$$R(z) = \frac{1-\rho}{1 - z - \lambda T\left(1 - \sum_{n=0}^{\infty} c(n) z^n / N\right)}.$$

Equating like coefficients of z^n we get $r_0 = \dfrac{1-\rho}{1-\lambda T}$

and for n = 1,2,3, ... $(1-\lambda T) r_n = r_{n-1} - \dfrac{1}{N} \lambda T \sum_{j=0}^{n} c(n-j) r_j$.

This equation is equivalent to the M/D/1 recursion because there *N=1*, and *c(k)=1* and zero for any other *k*. As discussed in Brill et al., [16] the runtimes in computing *Wq(t)* can become excessive. The reason was because the sample size, *N*, was large (~ 10000) and the sum,

$$\sum_{j=0}^{n-1} c(n-j) r_j ,$$

has to be computed for each n. The large N was used to ensure we get enough samples in the tail of the underlying distribution. We have modified UTAM to get around this problem.

Let $F(x)$ is the cumulative distribution function of interest, and for $k = 1,2, \ldots, N$ set

$$y(k) = 1 - q^k \tag{2}$$

for some q in (0,1) and solve the equation $F(x(k)) = y(k)$

for $x(k)$ and $k = 1,2,\ldots,N$. The samples, $x(k)$, are weighted by the probabilities $p(0) = (y(1)+y(2))/2$, $P(N)=1-(y(N-1)+y(N))/2$ and
$$p(k) = [y(k+1) - y(k-1)]/2$$

for k = 2,3,..., N-1. The TAM approximation for $B^*(s)$ becomes

$$B^*(s) = \sum_{k=1}^{N} p(k)e^{[-sx(k)]},$$

we call this method the geometric TAM or GTAM. The overall recursion for $Wq(t)$

presented above and coupled with TAM we refer to as the TAM Recursion for computing $Wq(t)$.

For GTAM, there is slight modification in the TAM recursion. The quantities $c(n)$ become the sum of the $p(i)$'s such that $k(i)$ =n. In addition, $R(z)$ would be

$$R(z) = \frac{1-\rho}{1 - z - \lambda T\left(1 - \sum_{n=0}^{\infty} c(n)z^n\right)},$$

and the basic recursion for r_n becomes

$$r_0 = \frac{1-\rho}{1-\lambda T}$$

and for $n = 1,2,3, \ldots$ $(1-\lambda T)r_n = r_{n-1} - \lambda T \sum_{j=0}^{n} c(n-j)r_j$.

The final issue that needs to be addressed is how one select "q" in Equation (2) does. For a given N, we use a search procedure to find "q" that equates the mean found via GTAM and the actual mean of the underlying distribution. It is not hard to see that for a given N, GTAM generates more samples in the tail of the distribution than UTAM.

2. CONCLUSION

We have presented two versions of TAM and have shown that GTAM was significantly better than UTAM. Previously, we used UTAM in determining the queue length process in M/G/1. There we used the Fast Fourier Transform (FFT) method [10] to invert the queue length moment generating function. It performed well and had acceptable runtimes. With GTAM's success on the waiting time problem, we returned to see how much GTAM could help in generating the queue length process.The TAM Recursion requires determination of $Wq(t)$ for all t's less than or equal to t. If one is interested in the value of $Wq(t)$ for a desired t^*, then all the $Wq(t)$ need be computed for $t \leq t^*$. This can be time consuming, although GTAM is quite efficient

3. REFERENCES

1. Business Week, April 7, 2011.
2. Feldman, A. and W. Whitt, "Fitting Mixtures of Exponentials to Long-Tail Distributions to Analyze Network Performance Models," *Performance Evaluation*, 31, pp. 245-279, 2008.
3. Fischer, M. J. and C. M. Harris, "A Method For Analyzing Congestion in Pareto and Related Queues," *The Telecommunication Review*, Volume 10, Mitretek Systems, McLean, VA, pp. 15-28, 1999.
4. Harris, C. M., P. H. Brill, and M. J. Fischer, "Internet-Type Queues With Heavy-Tailed Interarrival Times and Computational Methods for Their Analysis," *INFORMS Journal on Computing*, 12, pp. 261-271, 2011.
5. Brill, P. H., M. J. Fischer, D. Gross, and D. M. Masi, "Internet-Type Queues With Heavy-Tailed Service Times and Their Computational Solution," *INFORMS Journal on Computing*, 2000.
6. Gross, D. and C. M. Harris, *Fundamentals of Queueing Theory*, Third Edition, John Wiley, New York, 1998.
7. Abate, J. and W. Whitt, "The Fourier-Series Method For Inverting Transforms of Probability Distributions," *Queueing Systems*, 10, pp. 5-88, 1992.
8. Shortle, J. F. and M. J. Fischer, "Using the Transformation Approximation Method to Analyze Queues With Pareto Service," Submitted to Annals of Operation Research, 2011.
9. Fischer, M. J. and J. E. Knepley, "A Numerical Solution for Some Computational Problems Occurring in Queueing," *Theory Algorithmic Methods in Probability, Studies in Management Science*, pp. 271-285, 1977.
10. Tijms, H., *Stochastic Models: An Algorithm Approach*, John Wiley, New York, 1994
11. Fischer, M. J., D. Gross, J. F. Shortle, D. M. Masi, and P. H. Brill, "Experiences in Using the Transform Approximation Method to Analyze Queues with Heavy-Tailed Service Distributions," *INFORMS Applied Probability Conference*, New York City, 2011.

****

HOD of Maths – Velammal Institute of Technology, Chennai
hodofmaths@Gmail.com