Proceedings of the 1984
Winter Simulation Conference
S. Sheppard, U. Pooch, D. Pegden (eds.)

APPROXIMATING TIME-DEPENDENT NON-EXPONENTIAL QUEUES

Michael R. Taaffe and Kim Ong
School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907

Moment-differential equations, moment-matching, surrogate distributions and numerical integration have been combined to form a small closed set of differential equations that can accurately approximate huge sets of Kolmogorov-forward equations for complex nonstationary queueing models.

INTRODUCTION

This paper outlines a method for approximating the time-dependent distribution of queue size for Ph(t)/M(t)/S/C and Ph(t)/Ph(t)/1/C queueing models. The general method, called the surrogate distribution approximation (SDA) method, will be outlined and graphical results for selected test cases will be discussed.

Real service systems typically exhibit time-dependent behavior. Computer communication systems, telephone systems, and air-traffic control communication systems all can reasonably be expected to have arrival-traffic patterns that vary with the time of day. Likewise, mean rates of service may be time-dependent (e.g., server maintenance, a peak traffic server may be replaced by faster servers...). Mathematical models of stochastic service systems that include nonstationary input ($\lambda(t)$ the arrival rate and $\mu(t)$ the service rate) are intractable for all but the simplest systems. Simulation experimentation for nonstationary systems can be quite costly [Fishman, 1978]. Numerical procedures for evaluating the time-dependent distribution of system size such as Runge-Kutta numerical integration of the Kolmogorov-forward equations has long been an effective method [Taaffe, 1982] [Giffin, 1978]. However, numerical integration loses its appeal as a solution method for models that have large state-spaces, because for every additional point in the state-space, an additional differential equation is needed. Multivariate queueing models can easily be constructed which require thousands or millions of differential-difference equations. Priority queues, networks of queues and queues having phase distributions of interarrival times and/or service times are examples of models that can have state-spaces which are too large to analyze via numerical integration of the Kolmogorov-forward equations [Taaffe, 1982].

The SDA method for a multivariate queueing model partitions the state-space into subspaces which have uniform flow or transition rates. The probability of being in a particular state of the system can be constructed by a set of conditional distributions (conditioned on being in a particular subspace) and a set of probabilities relating the subspaces. In addition, each of the conditional distributions are approximated by matching the first two moments with the first two moments of an approximating distribution called the surrogate distribution. The result of the SDA approach when applied to the Ph(t)/M(t)/S/C model is to reduce the number of differential equations from K(C+1) to 6K where K is the number of phases in the arrival process. The Ph(t)/Ph(t)/1/C is described by $K_1 + CK_1K_2$ equations and is approximated by $K_1 + 3K_1K_2$ where K_1 is the number of phases in the arrival process and K_2 is the number of phases in the service process. In both cases the approximation replaces C (the capacity) with a constant in the expression describing the number of differential equations. The approximation, thus, makes feasible analyzing even systems with very large state-spaces. Next, the SDA method will briefly be reviewed. For details, the reader should consult Clark [1981] or Taaffe [1982].

SDA

The SDA method will be described in this section by using the well-known M(t)/M(t)/S/C as an example. More complex models are described by more complex differential equations, but the SDA approach for the these models is similar to the M(t)/M(t)/S/C. The Kolmogorov-forward equations for the M(t)/M(t)/S/C are

$$dPn(t)/dt = -(\lambda(t)+n\mu(t))P_n(t) + \overline{\delta}_{n0}\lambda(t)P_{n-1}(t)+(n+1)\mu(t)P_{n+1}(t) \qquad 0 \le n \le S-1$$

$$dPn(t)/dt = -(\overline{\delta}_{nc}\lambda(t)+S\mu(t))P_n(t)+\lambda(t)P_{n-1}(t) + \overline{\delta}_{nc}S\mu(t)P_{n+1}(t) \qquad S \le n \le C$$

where

$$\delta_{ij} = \begin{bmatrix} 1, & i=j \\ 0, & i\neq j \end{bmatrix}$$

$$\overline{\delta}_{ij} = 1 - \delta_{ij}$$

There are C+1 states in this system. Therefore, C+1 differential equations can be numerically integrated to analyze the time-dependent behavior. The SDA method can approximate this model by calculating conditional moments via differential equations and a few approximate probabilities. Two conditional distributions can describe the M(t)/M(t)/S/C. They are the distribution of system size given not all servers are busy and the distribution of system size given all the servers are busy. The conditional moment differential equations for this system are:

$$dE_1^{(p)}(t)/dt = \lambda(t)\ (\sum_{q=0}^{p-1} \binom{p}{q} E_1^{(q)}(t) - S^p P_{S-1}(t)\ )$$

$$+ \mu(t)\ (\sum_{q=0}^{p-1} \binom{p}{q} (-1)^{p-q} E_1^{q+1}(t)$$

$$+ S(S-1)^p P_S(t)\ )$$

and

$$dE_2^{(p)}(t)/dt = \lambda(t)\ (\sum_{q=0}^{p-1} \binom{p}{q} (E_2^{(q)} - C^q P_C(t)\ )$$

$$+ S^p P_{S-1}(t))$$

$$+ S\mu(t) \sum_{q=0}^{p-1} \binom{p}{q} (-1)^{p-q} E_2^{(q)}(t)$$

$$- (S-1)^p P_S(t)\ )$$

where $E_1^{(p)} \equiv E(N(t)^p | N(t) < S)$

$E_2^{(p)} \equiv E(N(t)^p | N(t) \geq S)$

Of course $E_1^{(0)}(t) + E_2^{(0)}(t) = 1$ and $E_1^{(0)}(t) = P(N(t) < S)$ and $E_2^{(0)}(t) = P(N(t) \geq S)$.

The moments of the system size are often the performance measures of interest. One could numerically integrate these five moment differential equations ($E_1^{(0)}(t)$ and $E_2^{(0)}(t)$ are redundant, thus one can be ignored) to describe system behavior. To do so, however, requires knowledge of $P_{S-1}(t)$, $P_S(t)$, and $P_C(t)$. Thus the set of moment differential equations is not closed. The SDA approach is a moment matching method. In this case $(E_1^{(1)}(t), E_1^{(2)}(t))$ are matched by a Polya-Eggenberger (PE) distribution and $(E_2^{(1)}(t), E_2^{(2)}(t))$ are matched by a second PE distribution. The probabilities needed to numerically solve the conditional moment differential equations are then approximated by PE probabilities. The entire procedure can be summarized as follows:

1. $E_1^{(p)}(t), E_2^{(p)}(t)$ (The conditional moments) are given at time t for p=0,1,2.
2. Assume both of the two conditional distributions is approximated by a PE.
3. Solve for the parameters of both PE surrogate distributions given

$$(E_1^{(1)}(t), E_1^{(2)}(t)) \text{ and } (E_2^{(1)}(t), E_2^{(2)}(t))$$

4. Calculate $dE_i^{(p)}(t)/dt$ (p=0,1,2 and i=1,2) using, in part, probabilities emanating from the surrogate distributions.
5. Calculate $E_i^{(p)}(t+\Delta t)$ (p=0,1,2 and i=1,2) by numerical integration.

The result of using the SDA method on the M(t)/M(t)/S/C is five differential equations approximating C+1 equations with excellent accuracy (≈ 5% maximum error). See Clark [1981] for details of the SDA method for the M(t)/M(t)/S system or Taaffe [1982] for the M(t)/M(t)/S/C.

The remainder of this paper outlines use of the SDA approach for two models with time-dependent phase distributions as input.

<u>Ph(t)/M(t)/S/C</u>

The Ph(t)/M(t)/S/C model has the following state-space:

$$S = \{(n,i,t) | 0 < n < C,\ 1 < i < k,\ t > 0\}$$

n = number of customers in the system

i = arrival phase of the next arrival

k = the number of phases comprising the arrival

The size of the state-space is the cardinality of S, which is also the number of differential equations needed to describe this system.

$$|S| = K(C+1)$$

The state-space for the Ph(t)/M(t)/S/C system can be partitioned into 2K regions, where K is the number of phases needed to describe the arrival process. There are two subspaces for each phase of the arrival process: one subspace for Phase i and the number of customers < S and another for Phase i and the number of customers $\geq$ S. The result of this partitioning is 2K conditional distributions:

$P(N(t)=j | A(t)=i,\ N(t) < S)$ $j=0,\ldots S-1$, $i=1,2,\ldots K$

$P(N(t)=j | I(t)=i,\ N(t) \geq S)$ $j=S,S+1,\ldots,C$, $i=1,2,\ldots,K$

where $A(t) \equiv$ current phase of the next arrival at time t

$N(t) \equiv$ number of customers in the system at time t

The SDA approach requires matching two moments for each conditional distribution and a probability for the system being in a particular subspace; i.e.,

$$E_{i;1}^{(p)}(t) \equiv E(N(t)^p | A(t)=i, N(t) < S)$$

and

$$E_{i;2}^{(p)}(t) \equiv E(N(t)^p | A(t)=i, N(t) \geq S)$$

for i=1,2,...,K and p=0,1,2. there are 2K conditional distributions and 3 moments per distribution (0^{th}, 1^{st} and 2^{nd}). The probabilities required in evaluating the conditional moment differential equations are

1. $P_{S-1,\ell}(t)$ $\quad \ell \in A_{(K)}$
2. $P_{S,i}(t)$ $\quad i=1,2,\ldots,K$

where $A_{(K)}$ ≡ set of phase indices which may be terminal phases for the arrival process.

Implementation of the SDA method on the Ph(t)/M(t)/S/C model has resulted in an efficient and accurate approximation. Test cases examined thus far have exhibited a maximum error in the first moment of less than 2%. Similar accuracy has also been the case for other performance measures such as VAR(N(t)) and P(N(t)≠0). Figure 1 is a plot of E(N(t)) for a typical Ph(t)/M(t)/3/20 model. In this case the inter-arrival distribution is an Erlang order 3 distribution with a time-dependent rate parameter ($E_3(t)$).

Ph(t)/Ph(t)/1/C

The state-space for the Ph(t)/Ph(t)/1/C model is

$$S = \{(0,i,0) | 1 \leq i \leq K_1\} \cup \{(n,i,j) | 1 \leq n \leq C, \; 1 \leq i \leq K_1, \; 1 \leq j \leq K_2\}$$

where (n,i,j) ≡ n customers, i is the current phase of the next arrival and j is the current phase of the customer in service

K_1 ≡ number of phases in the arrival process

K_2 ≡ number of phases in the service process

The size of the state-space is

$$|S| = K_1 + CK_1K_2$$

Clearly for moderate K_i and large C, the number of equations can become too large for convenient analysis. As in the Ph(t)/M(t)/S/C model the SDA approach will be applied after partitioning the state-space by phase, both service and arrival. The conditional moments associated with the conditional distributions for the subspaces are

$$E_{ij}^{(p)}(t) = E(N(t)^p | A(t)=i, I(t)=j) \qquad 1 \leq i \leq K_1, \quad 1 \leq j \leq K_2, \quad p=0,1,2$$

A(t) ≡ current phase of the next arrival at time t

S(t) ≡ current phase of service for the customer in service at time t

The probability terms that appear on the right-hand side of the conditional moment differential equations $dE_{ij}^{(p)}(t)/dt$ are:

1. $P_{0,\ell,0}(t)$
2. $P_{1,i,m}(t)$
3. $P_{C,\ell,j}$

for $1 \leq i \leq K_1$,

$1 \leq j \leq K_2$,

$\ell \in A_{(K_2)}$ and

$M \in S_{(K_2)}$

and $A_{(K_1)}$ ≡ set of indices of all possible terminal phases of the arrival process

$S_{(K_2)}$ ≡ set of indices of all possible terminal phases of the service process

Probabilities (2) and (3) are easily approximated by the usual SDA method. Probabilities (1) could be approximated similarly, however, since there are so few of them, they can easily be computed via the Kolmogorov equations directions. Thus the number of differential equations to approximate the Ph(t)/Ph(t)/1/C is $K_1 + 3K_1K_2$. Again the accuracy has thus far been excellent across all performance measures and test cases examined.

SUMMARY

The combination of moment-differential, moment matching, surrogate distributions and numerical integration has been used to construct algorithms resulting in a small closed set of differential equations that can accurately approximate huge sets of Kolmogorov-forward equations for complex nonstationary queueing models. The method described is called the SDA method and has been used successfully on the M(t)/M(t)/S, M(t)/M(t)/S/C, and M(t)/M(t)/1/C p-priority models in the past [Rothkopf, 1979] [Clark, 1981] [Taaffe, 1982]. This paper outlined the SDA approach for models with time-dependent phase distributions, which proved to be an excellent approximation. the SDA method also is being applied to some overflow models. The method shows great promise for making complex multivariate queueing model analysis more computationally feasible.

REFERENCES

Taaffe, M.R. (1982), Approximating Nonstationary Queueing Systems, Ph.D. dissertation, The Ohio State University, Columbus, Ohio 43210.

Clark, G.M. (1981), "Use of Polya Distributions in Approximate Solutions to Nonstationary M/M/S Queues", Comm. of The ACM, Vol. 24, No. 4, April 1981.

Rothkopf, M.H. and Oren, S.S., "A Closure Approximation for the Nonstationary M/M/S Queue, Management Science, Vol. 25, No. 6, June 1979.

Fishman, G.S., Principles of Discrete Event Simulation, John Wiley & Sons, Inc., New York, 1978.

Giffin, W.C., Queueing: Basic Theory and Applications, Grid, Inc., Columbus, Ohio, 1978.

PH(t)/M(t)/S/C

Example: $H_3(t)/M(t)/S/C$

Capacity: 30

Servers: 5

Initial number in the system: 0

Arrival Phase Probabilities P_i	Phase Arrival Rates $\lambda_i(t)$
.2	$45 + 3\sin(\pi t/2)$
.3	$38 + 8\sin(\pi t/3)$
.5	$28 + 9\sin(\pi/3)$

Service Rate $\mu(t)$

$6 + 3\sin(\pi t/2)$