APPROXIMATING NONSTATIONARY QUEUEING SYSTEMS

Michael R. Taaffe

School of Industrial Engineering
Purdue University

Gordon M. Clark

The Ohio State University
Columbus, Ohio

Proceedings of the 1982
Winter Simulation Conference
Highland * Chao * Madrigal, Editors

82CH1844-0/82/0000-00 9 $00.75 © 1982 IEEE

Abstract

This paper describes approximations for nonstationary queueing models. Explicit consideration of the time-dependent vector of probabilities associated with complex queueing systems requires numerical integration of the Chapman-Kolmogorov differential difference equations. The number of differential equations increases nonlinearly as the complexity of the system increases, for example priority or network systems. The research methods described in this paper are approximation approaches to reduce the number of differential equations numerically integrated, or simulated, necessary to represent complex stochastic models to a small and possibly constant number.

INTRODUCTION

Continuous simulation is a powerful tool in analyzing time-dependent behavior of complex systems. In continuous simulations, transient and time-varying behavior of systems are described by numerically integrating sets of difference, differential or differential difference equations. Forrester popularized the use of continuous simulation with his work on industrial dynamics, urban dynamics, principles of systems, and world dynamics [Forrester, 4-7]. Representation of delays in continuous simulation languages such as DYNAMO, GASPIV or SLAM makes use of n^{th} order exponential delays. An n^{th} order delay process is equivalent to the time spent in a stochastic service system which has a time-dependent arrival rate, infinite number of servers and independent service times from the same Erlang distribution. In order to represent nonstationary systems that allow queueing, in a continuous simulation, numerical integration of the differential difference equations describing the time-dependent state probabilities is required. These differential difference equations describing time-dependent state probabilities are called the Kolmogorov-forward equations [Giffin, 8]. Kolesar, Rider, Crabill and Walker made use of linear programming and numerical integration of the Kolmogorov-forward equations for a M/M/c queueing system in a large model of scheduling police patrol cars to meet time-varying demands [Kolesar, et al., 10]. One serious limitation in the study mentioned is that no priority structure in dispatching the calls is made. The authors point out that modeling priority calls is possible but the complexity of computation would greatly increase.

Hartman, Koopman, Giffin, Walters and Bundy have all made use of Runge-Kutta integration of Kolmogorov-forward equations to study aspects of air traffic control [Hartman, 9], [Giffin, 8], [Bundy, 1]. Rothkopf and Oren developed an approximation for M/M/c nonstationary queues to study centralized photocopy centers [Rothkopf, 17]. Rothkopf and Oren noted as did Kolesar, et al., that the lack of complexity in the model such as priorities, was an important research issue.

Actual demand in real service systems such as air traffic, telephone traffic and data-communication network traffic can reasonably be expected to exhibit nonstationary patterns. Typically demand for on-line card-catalog computer equipment experiences three peak periods of use per day: mid-morning, early afternoon and early evening [Knox, Miller, 12]. Thus, steady state analytic models of a nationwide computer bibliographic utility service that accepted demands from several different time zones of the country and had dynamically distributed processing capability, would be of limited use to network planners and management. The sudden spike in the demand curve for telephone signal switching and long line use and the corresponding degradation of the overall long distance system resulting from Media Stimulated Calling (MSC) such as the demand that occurred immediately following the 1980 Presidential debate is an example of the importance transient analysis of stochastic service systems [LaPadula, 13]. Continuous simulation models of nonstationary systems, such as those mentioned above, require explicit representation of queues.

The curse of dimensionality limits the utility of numerical integration or continuous simulation of the Kolmogorov-forward equations, to systems that have a relatively small number of possible system states. The number of states in stochastic service systems increases rapidly and in a nonlinear manner for increasingly complex systems. As the complexity of the system increases or equivalently as the dimension of the vector describing the state of the system increases, the utility of numerical methods decreases.

The research methods described in this paper are approximation approaches to reduce the number of differential equations numerically integrated or simulated, necessary to represent complex stochastic models to a small and possibly constant number. Clark points out, approximation techniques are needed to reduce the number of differential equations integrated numerically to allow explicit consideration of queueing delays in continuous simulation [Clark, 3]. When complex stochastic models can be approximated by a relatively few differential equations, then numerical integration of the reduced set of equations is easily accomplished and continuous simulation models of large complex systems that include as components queueing systems, is feasible.

Two methods of reducing the number of differential equations will be investigated. The first method is a moment matching surrogate distribution approach. The second method is to approximate the departure processes from a node in a queueing system with a time-dependent Poisson process. Two important classes of complex nonstationary queueing models that are represented by multidimensional probability density functions will be investigated. The first is priority queueing and the second is a set of special cases of queueing network systems that have two nodes called tandem systems. Other complex queueing models are not considered. However, the approximation approaches presented in this research and combinations of these approaches may prove to be useful in the analysis of other complex nonstationary systems.

The standard approach for representing the transient and nonstationary behavior of M/M/1/K queueing systems is to numerically integrate the time differential difference equations representing the probabilities of being in each of the system states. The time derivatives are called the Kolmogorov-forward equations and are frequently used to analyze time-varying behavior of queues [Giffin, 8], [Hartman, 9], [Koopman, 11]. For an M/M/1/K nonstationary system the Kolmogorov-forward equations are

$$\frac{dPr(N(t)=0)}{dt} = -\lambda(t)Pr(N(t)=0) + \mu(t)Pr(N(t)=0) \qquad (1)$$

$$\frac{dPr(N(t)=i)}{dt} = -(\lambda(t)+\mu(t))Pr(N(t)=i) + \lambda(t)Pr(N(t)=i-1) + \mu(t)Pr(N(t)=i+1) \qquad (2)$$

$$i=1,2,\ldots,K-1$$

$$\frac{dPr(N(t)=K)}{dt} = -\mu(t)Pr(N(t)=K) + \lambda(t)Pr(N(t)=K-1) \qquad (3)$$

where

$N(t)$ = i indicates that the number of entities in the system at time t is i (4)

$Pr(N(t)=i)$ = Probability that $N(t)=i$ (5)

$\frac{dPr(N(t)=i)}{dt}$ = time derivative of $Pr(N(t)=i)$ (6)

$\lambda(t)$ = nonstationary mean Poisson arrival rate at time t

$\mu(t)$ = nonstationary mean service rate of exponentially distributed service times at time t

Equations (1) through (3) are very smooth differential-difference equations that can be numerically integrated quite accurately [Whitlock, 19], [Olson, 15], [Taaffe, 18]. To represent a system with capacity K, K+1 differential equations need to be numerically integrated, so a large model which has several independent M/M/1/K nonstationary queues as components can require extensive computational resources. Further, for even moderate size models which have interacting nonstationary queues as components, the state space increases nonlinearly causing great dimensional problems. The motivation for producing an approximation for nonstationary M/M/1/K queues is to reduce the computational resources requirements for evaluating the queue size distribution through time with little loss of accuracy.

In characterizing the behavior of a queueing system, the values of the first moment of the distribution of the number of entities in the system is important. To represent the time-varying behavior of the first moment, one approach is to numerically integrate the differential difference equations of the probabilities of being in each of the possible system states then compute the moment, when desired, by multiplying each probability by the appropriate index and summing, i.e.,

$$E(N(t)) = \sum_{i=0}^{K} Pr(N(t)=i)\,i \qquad \text{first moment} \qquad (7)$$

Similarly for the second moment

$$E(N(t)^2) = \sum_{i=0}^{K} Pr(N(t)=i)\,i^2 \qquad \text{second moment} \qquad (8)$$

In order to compute sums (7) and (8), of course the Kolmogorov-forward equations still need to be numerically integrated. If K is large, the number of arithmetic operations may be large enough to cause significant round-off errors. Not only do all K+1 differential equations need to be numerically evaluated but also the two sums, (7) and (8), have to be evaluated, each of which has K+1 sumands. Another approach is to derive the moment-differential equations directly and numerically integrate them rather than numerically integrating all of the Kolmogorov-forward equations and then summing.

$$\frac{dE(N(t))}{dt} = \sum_{i=0}^{K} \frac{dPr(N(t)=i)\,i}{dt} \qquad (9)$$

$$\frac{dE(N(t))}{dt} = \lambda(t)(1-Pr(N(t)=K) \qquad (10)$$

$$- \mu(t)(1-Pr(N(t)=0))$$

Thus, to numerically integrate the first moment differential equation only the probability of the system being at capacity and the probability of the system being idle need to be available. If somehow these two probabilities are available at time t then the first moment can be represented as a function of time and the K-1 probabilities need not be integrated. Similarly for the second moment differential equation:

$$\frac{dE(N(t)^2)}{dt} = \sum_{i=0}^{K} \frac{dPr(N(t)=i)}{dt} i^2$$

$$\frac{dE(N(t)^2)}{dt} = \lambda(t)(1-Pr(N(t))=K) \qquad (11)$$

$$+ 2\lambda(t)(E(N(t))-KPr(N(t)=K))$$

$$+ \mu(t)(1-Pr(N(t)=0))$$

$$- 2\mu(t)(E(N(t)))$$

Once again the second moment differential equation can be computed using only the value of the arrival rate $\lambda(t)$, the service rate $\mu(t)$, the first moment E(N(t)) and the probabilities of a full system Pr(N(t)=K) and empty system Pr(N(t)=0). Thus, if somehow the values for Pr(N(t)=0) and Pr(N(t)=K) were available for all t, then two differential equations would be all that is needed to get the time-varying values of the mean and variance of the number in the system. Numerical accuracy is also necessarily less of a concern in evaluating moment differential equations since significantly fewer arithmetic operations are required, e.g., only two differential equations would be numerically evaluated instead of K+1 differential equations and two summations with K+1 terms each.

Rider [16], Chang [2] and Clark [3] numerically integrated the first moment differential equation for an M/M/1 queue and approximated Pr(N(t)=0), the only probability needed, by various methods. Rothkopf and Oren [17] also approximated nonstationary queues by numerically integrating moment differential equations. Rothkopf and Oren's approximation was the first to use a probability distribution as a surrogate for the actual and unknown distribution. The negative binomial served as a surrogate from the Rothkopf and Oren approximation. Clark [3] extended the approximation procedure of Rothkopf and Oren and eliminated the need for correction factors for queues with multiple servers. Clark used the Polya-Eggenberger distribution as the surrogate distribution. For a review of a comparison of the negative-binomial versus the Polya-Eggenberger as a surrogate, see Clark [3].

The Surrogate Distribution Approximation Approach

If the distributional form of Pr(N(t)=i) were known and it was a two parameter distribution, then computing Pr(N=0,t) and Pr(N=K,t) could be simple given the first two moments. Of course, the problem is that the functional form of Pr(N(t)=i) is not known. Even if the functional form of Pr(N(t)=i) were known, only two probabilities are actually used to compute the moment differential equations. Because knowing the distributional form of Pr(N(t)=i) would make things much easier, assuming it is a two parameter distribution, one approximation method is to use a surrogate distribution for the unknown distribution. Even if the form of the surrogate distribution does not change with time, the surrogate probabilities do change with time since the moments change with time, thus the surrogate distribution parameters are also changing. The surrogate distribution moment-matching approach is the first method discussed in this paper and will henceforth be referred to as the surrogate distribution approximation. In outline the simplest version of the surrogate distribution approximation consists of five steps:

1) Assume E(N(t)) and $E(N^2(t))$ are known.

2) Assume Pr(N(t)=i) is approximated by Pr(X=i) where X is a random variable with an associated two parameter distribution called the surrogate.

3) Solve for the surrogate distribution parameters given $E(N(t))$ and $E(N^2(t))$.

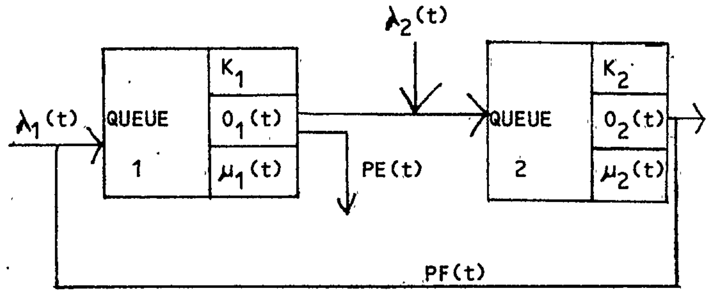
4) Calculate $\frac{dE(N(t))}{dt}$ and $\frac{dE(N^2(t))}{dt}$ using $Pr(N(t)=0)$ and $Pr(N(t)=K)$ as approximated by $Pr(X=0)$ and $Pr(X=K)$.

5) Calculate $E(N(t+\Delta t))$ and $E(N^2(t+\Delta t))$ by numerical integration.

Step one is a natural result of an iterative use of the approximation procedure. The surrogate distribution approximation is a heuristic procedure and the reason is that the surrogate distribution at any time will have the correct first two moments but nothing can be said a priori about the accuracy of any of the surrogate probabilities emanating from the surrogate distribution. As demonstrated earlier, only two probabilities from the surrogate distribution of $Pr(N(t)=i)$ need to be accurate if the surrogate approach were to be used for the M/M/1/K, i.e., $Pr(N(t)=0)$ and $Pr(N(t)=K)$. Step three implies that the choice of surrogate, in part, must be based on the computability of the distribution's parameters given the first two moments. A distribution whose parameters are difficult to compute would not serve as an efficient surrogate. Step four merely calculates the moment-differential equations based on their current value, the current mean arrival and service rates and, in the case of the M/M/1/K, two probabilities from the surrogate distribution. By numerical integration, as Step five indicates, the first two moments are computed for time $t+\Delta t$. At time $t+\Delta t$, the procedure is repeated. With each increment in time, a new set of surrogate probabilities are used to compute the derivatives for the moments which in turn are used to compute the next set of moments and thus surrogate distribution parameters. The surrogate distribution approximation has been applied to the M/M/c, M/M/c/k, M/M/1/k 2 priority nonpre-emptive models and some special cases of the M/M/1/k p-priority models. Results from all of these applications indicate excellent accuracy and robustness. Detailed description of the accuracy of the approximations are available in Clark [3] and Taaffe [18].

Tandem Queueing Systems

The second type of queueing system for which approximations will be discussed in this paper are several special cases of the nonstationary tandem queueing system. The following is a diagram of the general tandem considered in this paper.



$\lambda_i(t)$ = mean rate of the time-dependent Poisson external arrival process to node i at time t.

$\mu_i(t)$ = mean rate of the time-dependent service process at node i at time t.

$0_i(t)$ = mean rate of the departure process from node i at time t.

K_i = capacity of node i.

PF(t) = probability at time t of an entity departing two returning to node node one, i.e., the probability of feedback at time t.

PE(t) = probability at time t of an entity one and exiting the departing node one system.

The dimension of the state space is two and the number of states the system can realize is $(K_1+1)(K_2+1)$. The joint probability of finding the system in any one of the possible states is represented by

$$P_{i,J}(t) = Pr(N_1(t)=i, N_2(t)=J)$$

where $N_i(t)$ is the number of entities at node i at time t.

Clearly, if the queueing node capacities are large then the number of states and thus the number of differential equations gets large quickly. In general, in a N node network with each node having capacity k the number of states is $(k+1)^N$. The approach taken in the tandem network approximations is to analyze the marginal Chapman-Kolmogorov equations and make approximations to reduce the number of differential equations to $(K_1+1)+(K_2+1)$ rather than the product. Further, combining the surrogate approximation outlined above with the tandem approximation the number of differential equations is reduced to a small constant number.

In this class of approximation a time-dependent surrogate distribution is used not to describe the number in the queue but to describe the time between departures from each of the queueing nodes in the network.

The approximation strategy used consisted of making a heuristic assumption about the functional form of a few joint probabilities which result in departure processes that are time-dependent Poisson processes (TDPP). For example: for node 1, which has a component of its input process a fraction of the output process of node 2, consider the joint probability

$Pr(N_1(t)=i, N_2(t)=0)$

$N(t)$ = number of entities at node 1 at time t

$N_2(t)$ = number of entities at node 2 at time t

to be the product of its marginal probabilities, i.e.,

$$Pr(N_1(t)=i, N_2(t)=0) = Pr(N_1(t)=i) Pr(N_2(t)=0)$$

If this assumption is made then the output process from node two and thus a component of the input process at node one, is TDPP [Taaffe, 18]. The approximating assumption described above does not assert that node one and node two queue size random variables are stochastically independent, rather it is an assumption about the functional form of a few joint probabilities.

The performance of the approximation for several special cases as well as the general tandem queueing system are analyzed and presented in detail in Taaffe [18].

Summary

The empirical results of the approximations discussed above are encouraging. The surrogate distribution approximation for priority queues is robust and quite accurate. The strengths and weaknesses of the tandem approximation are apparent from the empirical results. Future work combining the two types of approximations should prove useful in approximating more complex nonstationary systems.

BIBLIOGRAPHY

1. Bundy, W.W., "Queues with Delayed Probabilistic Feedback as a Model of Air Traffic Control Communications," M.S. Thesis, The Ohio State University, Columbus, Ohio, 1976.

2. Chang, S.S.L., "Simulation and Time-Varying Conditions in Queueing Networks," Proceedings of the Pittsburgh Conference of Modelling and Simulation, 1976.

3. Clark, G.M., "Use of Poly Distributions in Approximate Solutions to Nonstationary M/M/s Queues," Communications of the ACM, Vol. 24, No. 4, pp. 206-218, 1981.

4. Forrester, J.W., Industrial Dynamics, John Wiley & Sons, New York, 1961.

5. Forrester, J.W., Urban Dynamics, The M.I.T. Press, 1969.

6. Forrester, J.W., Principles of Systems, Wright-Allen Press, 1971.

7. Forrester, J.W., World Dynamics, Wright-Allen Press, 1971.

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

9. Hartman, Michael G., "The Truncated, Time-Varying Queue as a Model of an Air Traffic Control System," Master's Thesis, The Ohio State University, Columbus, Ohio, 1971.

10. Kolesar, P.J., Rider, K.L., Cragill, T.B., and Walker, W.E., "A Queueing-Linear Programming Approach to Scheduling Police Patrol Cars," Operations Research, Vol. 23, No. 6, 1975.

11. Koopman, B.O., "Air Terminal Queues under Time-Dependent Conditions," Operations Research, Vol. 20, 1972.

12. Knox, A.W., and Miller, B.A., "Predicting the Number of Public Computer Terminals Needed for an On-Line Catalog: A Queueing Theory Approach," Paper presented at The American Library Association Research Round Table Research Forum, Dallas, Texas, 1979.

13. LaPadula, C.A., "On the Accuracy of a Closure Approximation for Non-Stationary Multi-Server Queueing Systems," ORSA/TIMS Meeting, Toronto, 1981.

14. Newell, G.F., Approximate Behavior of Tandem Queues, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, New York, 1979.

15. Olson, Stuart W., "The Numerical Solution of Transient Queueing Problems," AD 744802 Technical Report, Systems Analysis Division, Plans and Analysis Directorate, HQ, U.S. Army Weapons command, Rock Island, Illinois.

16. Rider, K.L., "A Simple Approximation to the Average Queue Size in the Time-Dependent M/M/1 Queue," J. ACM, Vol. 23, No. 2, 1976.

17. Rothkopf, M.H., and Oren, S.S., "A Closure Approximation for the Nonstationary M/M/s Queue," Communications of the ACM, Vol. 24 No. 4, 1981.

18. Taaffe, M.R., "Approximating Nonstationary Queueing Systems," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1982.

19. Whitlock,Jr., James Smith, "Modeling Computer Systems with Time Varying Markov Models," Ph.D. Dissertation, University of North Carolina, Chapel Hill, 1973.