TRANSIENT QUEUEING APPROXIMATIONS FOR COMPUTER NETWORKS

A Thesis

by

WILLIAM A. BAKER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 1986

Major Subject: Electrical Engineering

TRANSIENT QUEUEING APPROXIMATIONS FOR COMPUTER NETWORKS

A Thesis

by

WILLIAM A. BAKER

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Pierce E. Cantrell
(Chair of Committee)

Richard M. Feldman
(Member)

Jerry D. Gibson
(Member)

Karan Watson
(Member)

Jo W. Howze
(Head of Department)

December 1986

Major Subject: Electrical Engineering

ABSTRACT

Transient Queueing Approximations for Computer Networks. (December 1986)

William A. Baker, B.S., Rutgers University

Chair of Advisory Committee: Dr. P. E. Cantrell

The objective of this thesis was to evaluate the performance of several transient queue approximations. The approximations were tested and characterized for a single M/M/1 queue and a tandem queue (two node) network.

The five approximations tested in this thesis used a closure assumption to obtain the probability of an empty system. Then, depending on the method, equations were integrated to obtain the mean and, in some cases, the variance. Johnston's and Rider's methods solved for just the mean. Rothkopf/Oren's and Chang/Wang's methods obtained mean and variance values, and Clark's method produced several quantities which were used to find mean and variance statistics.

For the M/M/1 case, the approximations by Clark and Chang were very accurate over a wide range of input patterns and initial conditions. Rothkopf's was accurate over all conditions but never as accurate as Chang or Clark. Johnston's and Rider's approximations performed acceptably only over some of the cases. The hardest conditions to follow, based on relative error, were low utilization cases with a large number in the queue at $t = 0$.

For nonstationary arrival patterns into the M/M/1 queue, Clark's method was superior to all others; mean and variance values were always within three percent of the exact.

For the tandem queue, equations for dM/dt and dV/dt were derived to observe dependencies on joint probabilities between the queues. While the rate of change of the mean was only a function of the marginal probabilities of each queue, the rate of change for the variance included joint probability terms. An assumption of queue independence was made in order to implement the closure assumptions for the tandem queue.

The approximations by Chang and Clark were very accurate in producing the mean. For low utilization cases, the methods experienced difficulties in following the true variance values. This was due to inaccuracies in the assumption that the two queues were independent of each other.

In conclusion, the methods by Chang/Wang and Clark hold promise for use in modeling computer networks, particularly for the mean in each queue.

To my mother, Rose

ACKNOWLEDGMENTS

Special thanks to Dr. P. E. Cantrell, whose guidance and patience made this thesis possible. I would also like to thank Drs. R. M. Feldman, J. D. Gibson, and K. Watson for their time and useful comments in reviewing this document.

I am indebted to Dr. G. M. Clark of Ohio State University for help in implementing his approximation method.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The objective of this thesis is to evaluate the performance of several transient queueing approximations for a network of queues. These approximations will be tested and characterized for a single M/M/1 and a tandem queue (2 node) network [1].

The statistics of queueing systems, such as the mean number in each queue and the variance, are often computed using steady state assumptions. In many systems, however, the queue parameters change with time and steady state assumptions lead to erroneous mean and variance quantities. It is therefore desirable to solve the transient system. Unfortunately, solutions to transient queueing systems are difficult to obtain. Although an analytic solution exists for the single M/M/1 queue, a network of two such queues remains an open problem. The approximation methods are used reduce the computational complexity of existing transient solutions and to provide insight into the behavior of systems for which no analytic solution exists. It is hoped that this research will serve to improve present methods of modelling computer networks [2].

A. Queueing Theory Background

Central to interpreting results from any queueing model is the understanding of the underlying queueing process. If you now refer to Fig. 1, you will see an example of the LATEX picture environment. Of particular importance are five basic characteristics [3]: arrival pattern of customers, service pattern of servers, queue discipline, system capacity, and the number of service channels.

Journal model is *IEEE Transactions on Automatic Control.*

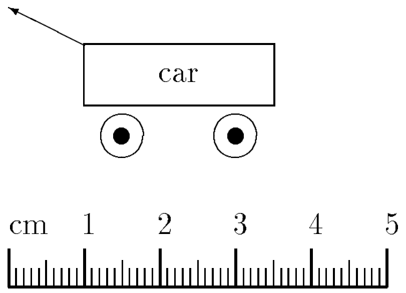

Fig. 1. A sample `picture` environment.

B. Network Applications

The study of a network of queues can be used to provide useful information for the design and maintenance of computer networks, where several computers are communicating with each other. On the design side, the modeling of a network can provide statistics such the average number of packets waiting to be transmitted at each computer [4].

C. Solution Methods

The most common numerical solution to the transient queueing model is found through the use of the Kolmogorov forward equations [5]. This method can handle non-stationary arrival and service rates and, for reasonable error bounds, provides an exact solution. One equation is integrated to find the probability of being in a particular state.

D. Thesis Structure

Chapter II starts by reviewing some of the fundamentals in queueing theory upon which the approximations are based. The closure approximations are presented for the M/M/1 queue and compared to reveal differences in structure [6].

In Chapter III the M/M/1 approximations are tested against exact methods for both stationary and nonstationary arrival patterns. The stationary cases are compared against exact results from Cantrell [7, 8], while the nonstationary cases will be compared to solutions from Kolmogorov forward equations. Each approximation will be characterized to show areas of weakness and strength.

In Chapter IV the methods proving to be most accurate will be tested in a two node feed-forward network, otherwise known as the tandem queue. The results are compared against the Kolmogorov forward equation solutions and results from the previous chapter to see the effect of the first node on the accuracy of the second node results.

In Chapter V final conclusions are drawn and suggestions for further research topics are suggested. An equation using the equation environment

$$\lim_{x \to 0} \frac{\sin x}{x} = 1, \tag{1.1}$$

and one using the displaymath environment

$$\sqrt{1 + \sqrt{1 + \sqrt{1 + x}}}.$$

are displayed here. Now refer to Fig. 2 for another example of what you can do with the LaTeX picture environment.

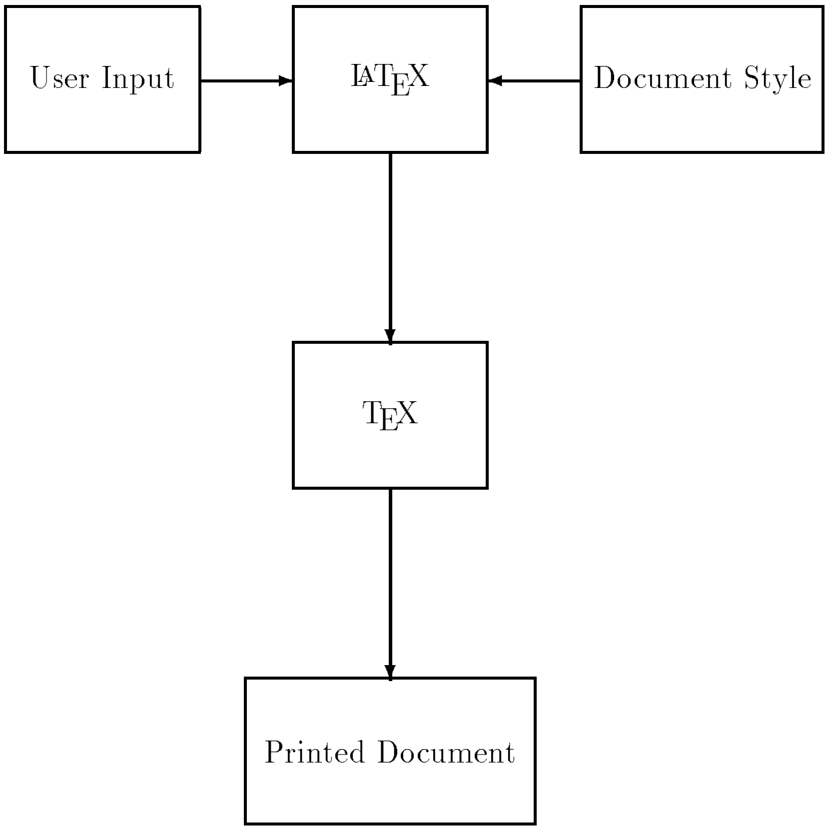

Fig. 2. Overall Structure

CHAPTER II

CLOSURE APPROXIMATIONS IN THE TANDEM QUEUE

The purpose of this chapter is to extend the results from the M/M/1 queue to a two queue system consisting of a M/M/1 queue whose output is directed to a second Markovian queue. This small network is known as a tandem queue and is depicted in Fig. 3. The size of this network makes possible a solution by near-exact methods so that the closure methods can be evaluated for the dependencies of the mean and variance of the second queue on the state of the first queue. Since the first queue of the tandem is simply M/M/1, this chapter will concentrate on the results from the second queue. The two most accurate closure assumptions, Clark and Chang/Wang, will be compared against the Kolmogorov solution [9].

A. The Kolmogorov Solution

The state space for the tandem queue is a two-dimensional lattice of states indexed by the number in each queue. For example, $P_{1,2}(t)$ is the probability that there is one in the first queue and two in the second. The size of the state space depends on the maximum number in each queue. If each queue can hold 49 items, including server, than the number of possible states is 50^2 or 2500 [10].

The Kolmogorov solution for the tandem queue was obtained using a stochastic balance between various states of the birth-death process. Fig. 4 shows the stochastic balance used to obtain (2.4). The Kolmogorov equation set for the tandem queue was found to be

$$\begin{aligned} \frac{dP_{0,0}}{dt} &= -\left(\gamma_1+\gamma_2\right)P_{0,0}+\mu_2 P_{0,1} \qquad (2.1)\\ \frac{dP_{0,i}}{dt} &= -\left(\gamma_1+\gamma_2+\mu_2\right)P_{0,i}+\mu_2 P_{0,i+1} \end{aligned}$$

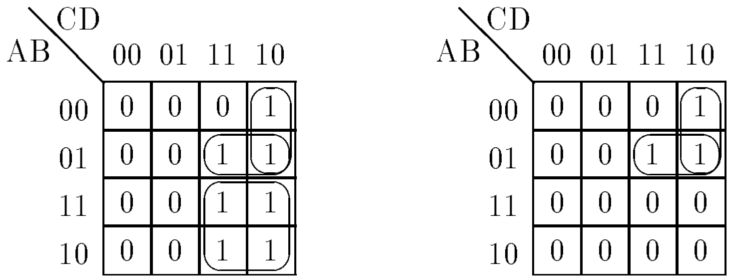

Fig. 3. The two node tandem queue.

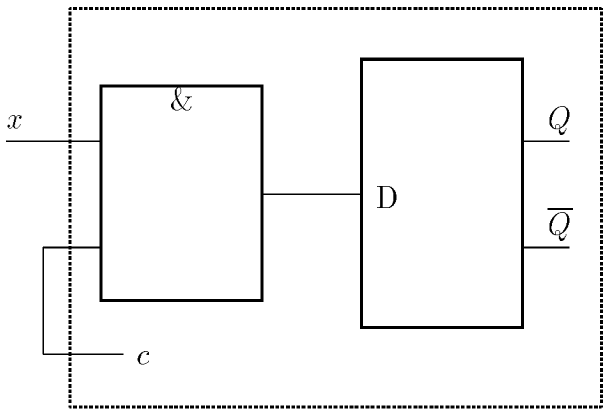

Fig. 4. Stochastic balance for tandem queue without feedback.

$$
\begin{aligned}
&+ \quad \gamma_2 P_{0,i-1} + \mu_1 P_{1,i-1} && i = 1,2,3... && (2.2)\\
\frac{dP_{j,0}}{dt} &= -(\gamma_1 + \gamma_2 + \mu_1) P_{j,0} + \gamma_1 P_{j-1,0} + \mu_2 P_{j,1} && j = 1,2,3... && (2.3)\\
\frac{dP_{j,i}}{dt} &= -(\gamma_1 + \gamma_2 + \mu_1 + \mu_2) P_{j,i} + \gamma_1 P_{j-1,i} + \mu_2 P_{j,i+1} && && \\
&+ \quad \gamma_2 P_{j,i-1} + \mu_1 P_{j+1,i-1} && j,i = 1,2,3... && (2.4)
\end{aligned}
$$

The mean and variance statistics for the second queue are obtained by the following equations:

$$M_2 = \sum_{i=1}^{\infty} i \cdot \sum_{j=0}^{\infty} P_{j,i}$$

and

$$V_2 = \sum_{i=1}^{\infty} i^2 \cdot \sum_{j=0}^{\infty} P_{j,i} - M_2^2.$$

Calculation of the mean and variance requires the truncation of the M/M/1/∞ to some maximum number of states. Stated differently, the M/M/1/∞ queue model is approximated by an M/M/1/k queue. While it is impossible to evaluate the error in this approximation, an indication of the truncation error can be obtained by summing all the probability states up to state k and subtracting this total from one. This yields the probability of being in a state greater than k. If this value is very small then its product with i and i^2 will also be small.

It is easy to see how large and complicated the Kolmogorov equation set can become for just a small network, and the usefulness of an accurate, state-reducing approximation [11].

B. Approximations for the Tandem Queue

1. Independent Queue Assumption

Jackson [12] showed that a network of queues can be analyzed as a group of independent M/M/1 queues when the network is operating under steady-state conditions.

One method to approximate the tandem queue state space is to assume that the independence holds under transient conditions as well. By assuming the two queues are independent, the joint probability $P_{j,i}$ simply becomes the product of the marginal probabilities, P_j and P_i. Thus, the number of states needed to model the tandem M/M/1/50 queue by the Kolmogorov equations decreases from 2500 to 100.

Since the primary motivation behind the approximation methods is to obtain accurate mean and variance statistics for the queues, it is of interest to investigate errors induced by assuming the queues to be independent. The mean and variance statistics for the first and second queues are defined as

$$\begin{aligned} M_1 &= \sum_{j=1}^{\infty} j \cdot P_j \\ V_1 &= \sum_{j=1}^{\infty} j^2 \cdot P_j - M_1^2, \end{aligned}$$

and

$$M_2 = \sum_{i=1}^{\infty} i \cdot P_i \tag{2.5}$$

$$V_2 = \sum_{i=1}^{\infty} i^2 \cdot P_i - M_2^2. \tag{2.6}$$

The accuracy of P_j for $j > 0$ will determine the effectiveness of the independence assumption. By definition, $P_j = \sum_{i=0}^{\infty} P_{j,i}$. By summing (2.3) and (2.4), we obtain

$$\begin{aligned} \frac{dP_j}{dt} &= -(\gamma_1 + \gamma_2 + \mu_1) \sum_{i=0}^{\infty} P_{j,i} - \mu_2 \sum_{i=1}^{\infty} P_{j,i} + \gamma_1 \sum_{i=0}^{\infty} P_{j-1,i} \\ &\quad + \mu_1 \sum_{i=1}^{\infty} P_{j+1,i-1} + \mu_2 \sum_{i=0}^{\infty} P_{j,i+1}. \end{aligned}$$

By gathering similar terms and summing, the above equation simplifies to

$$\frac{dP_j}{dt} = -(\gamma_1 + \mu_1) P_j + \gamma_1 P_{j-1} + \mu_1 P_{j+1}, \qquad j = 1, 2, 3...$$

which is identical to (2.4) developed for the single M/M/1 queue. This is true because the addition of the second queue does not effect the first in any manner. If, however, there was feedback from the second queue to the first then this result would no longer hold.

The equation for dP_i/dt for the second queue will now be derived to show how the joint probability state must be decoupled to obtain the independent queue probability equations.

2. Closure Approximations for the Tandem Queue

The approximations by Clark and Chang/Wang were shown in the previous chapter to be most accurate for the M/M/1 queue. In this section, we will investigate the extension of these approximations for the tandem queue. The resulting equation for dM_2/dt is

$$\frac{dM_2}{dt} = \gamma_2 + \mu_1 (1 - P0_1) - \mu_2 (1 - P0_2) . \tag{2.7}$$

To derive dV_2/dt, we differentiate (2.6) to obtain

$$\frac{dV_2}{dt} = \sum_{i=1}^{\infty} i^2 \cdot \frac{dP_i}{dt} - 2M_2 \cdot \frac{dM_2}{dt}. \tag{2.8}$$

C. Implementation and Results

Clearly, there are two issues concerning the accuracy of the closure approximations in a tandem queue. The first is the accuracy of the assumption of independent queues. When is the assumption that $P0_2$ is independent on the state of the first queue a good one? Also, what error results from the approximation for $V_2(t)$ via (2.8)? The second concern is how well the closure approximations model the independent tandem queue. Since the independence assumption makes the tandem queue a network of two M/M/1 queues, the second issue was largely answered in the previous chapter. Therefore this chapter will be dedicated to investigating the performance of the independent queue assumption [13].

1. Test Conditions

Three approximations were compared against the truncated Kolmogorov solution for the tandem queue: the independent Kolmogorov solution, Chang/Wang's approximation, and Clark's approximation. The test cases were the same as those discussed in Chapter I, except that cases with ρ close to or greater than one could not be included. This is because the truncated Kolmogorov equation set models the tandem queue as two dependent M/M/1/k queues, requiring the integration of k^2 equations. If ρ becomes too large then the probability of being in a state with greater than k in a queue can no longer be neglected, resulting in mean and variance inaccuracies. We used $k = 50$ which limited $\rho \leq 0.8$.

2. Results

The approximations all performed well for most of the conditions presented. The most accurate of the three was the Kolmogorov independent solution by a very small margin over Clark. Chang/Wang's method also was accurate, but it encountered difficulty with the high M_0, low utilization cases. See Table I for the full comparison. The comparable performance of the approximations is shown in Fig. 5. The Appendix also contains plots for the worst case percent error and the mean-square error.

As can be seen from Table II, Chang's method is much faster than the rest of the approximations. Clark's method also provides significant computational savings over both the dependent and the independent Kolmogorov methods. As is usually the case, increased accuracy and information accompanies increased computation.

This concludes the study of the tandem queue. To summarize, both Clark's and

Table I. Results for Nonstationary M/M/1 Queue

Test case parameters			Average Percent Error, e_{ave}, in %							
			John.	Rider	Rothkopf		Chang		Clark	
$\frac{\lambda}{\mu}$	a	T	$M(t)$	$M(t)$	$M(t)$	$V(t)$	$M(t)$	$V(t)$	$M(t)$	$V(t)$
0.5	1.0	10	28.96	9.77	1.98	11.27	3.39	5.00	0.05	0.47
0.5	1.0	20	28.35	11.75	4.28	21.06	6.21	9.37	0.17	0.65
0.5	1.0	40	25.76	14.24	7.32	32.02	10.60	16.07	0.64	1.96
0.5	1.0	60	24.25	16.48	8.65	33.88	9.97	19.25	1.03	2.47
0.5	1.0	80	22.17	17.04	8.99	32.19	15.68	17.41	1.24	2.70
0.5	1.0	100	19.60	14.92	8.18	20.33	14.77	18.63	1.17	2.75
0.5	1.0	120	17.45	13.03	4.51	11.37	6.60	21.80	0.86	2.19
0.9	0.25	10	12.26	4.82	2.52	9.26	1.27	4.50	0.19	0.67
0.9	0.25	20	7.59	3.71	2.37	11.26	1.08	4.39	0.17	0.76
0.9	0.25	40	6.44	3.81	1.72	13.10	1.50	5.65	0.47	1.20
0.9	0.25	60	7.08	4.13	2.12	14.26	2.05	7.72	0.85	1.71
0.9	0.25	80	7.88	4.37	2.74	15.13	2.64	10.26	1.23	2.20
0.9	0.25	100	8.47	4.65	3.41	15.79	3.22	13.22	1.58	2.73
0.9	0.25	120	8.89	5.09	3.96	16.25	3.89	16.52	1.88	3.27

Fig. 5. e_{ave} for stationary tandem queue, $M_0 = 0$.

Table II. CPU Times for Stationary Tandem Queue

Test case parameters		CPU time for VAX 8650 , in secs.			
		Exact	Independent		
$\frac{\lambda}{\mu}$	T_{final}	Kolmogorov	Kolmogorov	Chang	Clark
0.1	39	20.76	0.51	0.08	0.16
0.3	56	24.21	0.48	0.07	0.27
0.6	120	44.96	0.84	0.04	0.43
0.8	300	119.89	2.27	0.05	1.44

Chang/Wang's performed strongly for all tests when $\rho > 0.3$. For low utilization cases, the approximations incurred larger errors with respect to e_{ave} and e_{wor}. This however was due to numerical accuracy problems for small values of the mean coupled with large values (close to one) of $P0$ in both queues. The e_{wor} criterion in the Appendix did not show any model weakness for the low utilization cases.

REFERENCES

[1] Joe Author. Title of report. Research Report 23, Department of Publication, Name of University, City, State, October 1999.

[2] Arthur Author and Joe Author. Title of journal article. *Title of Journal*, 76:27–45, July 1776.

[3] Author Book. *Title of Book*, volume 1. Name of Publishing Company, City, State of Publication, 16th edition, 1988.

[4] Matthew Wright, James Chodzko, and Danny Luk. *Title of Book*, chapter 4–5, pages 56–99. Name of Publisher, College Station, Texas, 1988.

[5] Han Chen. Title of booklet. Texas A&M University, College Station, Texas, July 1989.

[6] Russell L. Oertel and Pierce Cantrell. LaTeX *Example Manual*. Department of Electrical Engineering, Texas A&M University, College Station, Texas, September 1988.

[7] Joseph W. Simpson. Title of article. In Editor of Book, editor, *Title of Book*, chapter 3, pages 400–402. Name of Publishing Company, City, State of Publication, 1983.

[8] Peter J. Rizzo, editor. *Title of Conference Proceedings*, Location of Symposium, June 1987. Sponsoring Organization, Name of publisher.

[9] Arthur Article. Title of conference article. In *Title of Conference Proceedings Record*, Location of Symposium, January 1989. Sponsoring Organization, Name of Publisher.

[10] P. Kenneth and G. E. Taylor. Solution of variational problems with bounded control variables by means of the generalized Newton-Raphson method. presented at the Symp. on Recent Advances in Optimization Techniques, Carnegie Inst. of Tech., Pittsburgh, Pa., April 1965.

[11] R. E. Larson. *Title of Dissertation.* PhD thesis, Stanford University, Stanford, California, 1964.

[12] Elaine G. Haschke. Title of thesis. Master's thesis, Drofnats University, West Yellowstone, Montana, May 1993.

[13] J. D. Simon. private communications, November 1963.

APPENDIX A

SUPPLEMENTAL RESULTS

Fig. 6. Rothkopf/Oren's $\widehat{P_0}$ results for stationary M/M/1 queue.

VITA

Mr. William A. Baker was born on May 28, 1958 in Camden N.J. He received his B. S. in Mechanical Engineering at Rutger's College of Engineering in 1980. His mailing address is 6 Suburban Blvd., Delran N.J.

The typist for this thesis was William A. Baker.