1981

The Response Times of Priority Classes under Preemptive Resume in M/G/m Queues

A. B. Bondi

J. P. Buzen

Report Number:
81-404

The Response Times of Priority Classes under Preemptive Resume in M/G/m Queues*

A.B. Bondi

Department of Computer Sciences,

Purdue University, West Lafayette, Indiana

J.P. Buzen

BGS Systems, Inc., Waltham, Massachusetts

CSD-TR No. 404, November 1982

Revised July 1983

ABSTRACT

Approximations are given for the mean response times of each priority level in a two-class multi-server M/G/m queue operating under preemptive resume scheduling. The results have been tested against simulations for different numbers of servers.

The Response Times of Priority Classes under Preemptive Resume in M/G/m Queues*

A.B. Bondi

Department of Computer Sciences,

Purdue University, West Lafayette, Indiana

J.P. Buzen

BGS Systems, Inc., Waltham, Massachusetts

CSD-TR No. 404, November 1982

Revised July 1983

Introduction

Many computer systems have central processing units which are allocated to jobs for short bursty periods whose durations have arbitrary distributions. Jobs at low priority levels may be preempted by those at higher ones, and the mean burst size may be different at each priority level. This note examines the queueing delays of jobs at two priority levels subject to preemptive resume scheduling.

In a recent paper, Buzen and Bondi (1983) [3] derived approximate expressions for the mean response times of each priority level in an $M/M/m$ preemptive priority queue. These approximations have been found to agree with the exact results for two class systems of Mitrani and King [8] as well as with simulation results for three classes

* This work was supported in part by NSF grant number MCS78-01729. Authors' addresses: A.B. Bondi, Dept. of Computer Sciences, Purdue University, West Lafayette, IN 47906; J.P. Buzen, BGS Systems Inc., 1 University Office Park, Waltham, MA 02254.

given here for the first time. The computational cost and complexity of the approximate method are very small.

In this note, the heuristic arguments of [3] are extended to treat the problem of approximating the response times of $M/G/m$ queues with two priority levels.

1. Definitions and Notation

Consider a preemptive resume $M/G/m$ queueing system with customers at r levels of priority such that

class i customers have priority over class j customers if $1 \le i < j \le r$,

customers within the same priority class follow the FCFS discipline,

customers of class i arrive in a Poisson stream with rate λ_i, and have i.i.d. general service times with mean $1/\mu_i$.

Denote the mean response time of class i customers by R_i. We shall also use $\lambda_{(p)}$ to denote the sum of the first p values of λ_i, and $\overline{R}_{(p)}$ to denote the overall average of the mean response times of the p highest priorities. Thus,

$$\lambda_{(p)} = \sum_{i=1}^{p} \lambda_i, \quad p = 1,2,\ldots,r. \tag{1}$$

By Little's Law,

$$\overline{R}_{(p)} = \sum_{i=1}^{p} \frac{\lambda_i R_i}{\lambda_{(p)}}, \quad p = 1,2,\ldots,r. \tag{2}$$

To ensure the existence of finite waiting times for the r priority classes [4], also assume that the total traffic intensity satisfies

$$\rho_{(r)} = \sum_{i=1}^{r} (\lambda_i / m\mu_i) < 1 \tag{3}$$

The overall mean service rate of the p highest priority levels, weighted by arrival rates, will be denoted by $\overline{\mu}_{(p)}$, i.e.,

$$\overline{\mu}_{(p)} = \frac{1}{\sum_{j=1}^{p} \lambda_j / \mu_j} \sum_{i=1}^{p} \lambda_i \tag{4}$$

The response time of a p-class m-server system with discipline d, arrival rate vector

$\underline{\lambda}_{(p)}=(\lambda_1, \ldots, \lambda_p)$, and service rate vector $\underline{\mu}_{(p)}=(\mu_1, \ldots, \mu_p)$ will be denoted by $R(d, \underline{\mu}_{(p)}, \underline{\lambda}_{(p)}, m)$. In this discussion, d will be either first come first served (FCFS) or preemptive resume (PRI). The waiting time of a system with the same parameters will be denoted analagously by $W(.,.,.,.)$.

In this section, the waiting time of a single class M/G/m queue will be denoted by $W(G)$, with G replaced by D or M as appropriate.

Our analysis makes use of previously published expressions for mean waiting times in single class queueing systems. Let S be a random variable denoting service time. By [9] the mean waiting time $W(G)$ of an m server system with generally distributed service and Poisson arrivals with rate a may be approximated by

$$W(G) \approx \frac{a^m E[S^2](E[S])^{m-1} p_0(\rho)}{2(m-1)!(m-aE[S])^2} \tag{5}$$

where

$$\rho = aE[S]/m \tag{6}$$

and

$$p_0(\rho) = [\sum_{i=0}^{m-1} \frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!(1-\rho)}]^{-1} \tag{7}$$

Notice that when the service time is exponentially distributed with $E[S]=1/u$, equation (5) yields the following exact expression:

$$W(M) = \frac{\rho P_m(\rho)}{a(1-\rho)}. \tag{8}$$

where

$$P_m(\rho) = \frac{p_0(\rho)(m\rho)^m}{m!(1-\rho)} \tag{9}$$

$P_m(\rho)$ is the probability of having m or more customers in the queue. $p_0(\rho)$ is the probability that the queue is empty and is given by

$$p_0(\rho) = [\sum_{i=0}^{m-1} \frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!(1-\rho)}]^{-1} \tag{10}$$

An alternative expression for the waiting time of an M/G/m queue with coefficient of variation other than zero and one has been proposed by Boxma et al (BCH) [2]:

$$W(G) \approx [W(M)+(2W(D)-W(M))(E[S^2]u/\gamma_1-m-1)/(m-1)]^{-1} \times (1+CS^2)W(M)W(D) \quad (11)$$

where CS^2 denotes the coefficient of variation squared and γ_1 depends on the service time distribution. $W(D)$ has been approximated by Cosmetatos [5] as

$$W(D) \approx \frac{1}{2} W(M)[1+(16m\rho)^{-1}(1-\rho)(m-1)((4+5c)^{1/2}-2)] \quad (12)$$

For deterministic service, the value of γ_1 recommended by [6] is

$$\gamma_1 \approx (1-CS^2)E[S]/(m+1)+CS^2E[S]/m \quad (13)$$

For two-stage hyperexponential service with rate parameters u_1 and u_2 and stage selection probability q,

$$\gamma_1 = \sum_{i=0}^{m} \frac{\binom{m}{i} r^i (1-r)^{m-i}}{iu_1+(m-i)u_2} \quad (14)$$

where

$$r = \frac{q/u_1}{q/u_1+(1-q)/u_2}$$

It can be shown that equation (11) reduces to $W(D)=W(D)$ or $W(M)=W(M)$ when equation (14) is used to compute γ_1. Equations (11)-(14) have been reported in [6] who have described them as being more accurate than that of Nozaki and Ross.

For the sake of completeness, numerical results based on both the Nozaki-Ross and BCH approximations to the single class waiting time will be presented.

2. Brief Description of the Simulations

Intensive simulations of two-class M/G/m preemptive resume priority queues were carried out with deterministic and two-stage hyperexponential service time distributions. Simulations of exponential service were carried out as a check. The simulation output was analyzed using the method of batched means [10]. Startup effects were eliminated by discarding data for the first 10000 seconds of simulated time and using only the remaining 40000 seconds. Simulations of three-class systems with exponential service were treated the same way.

3. Computing the Response Times of the Individual Classes

As noted by Barberis [1], equation (3) may be rearranged to yield the response time of the pth priority level, i.e.

$$R_p = \frac{\lambda_{(p)}\overline{R}_{(p)} - \lambda_{(p-1)}\overline{R}_{(p-1)}}{\lambda_p}, \; p=2,3,\ldots,r \tag{15}$$

This equation is valid for any preemptive priority system for which the p highest priority classes have finite waiting times and in which no overhead is attributed to the process of preemption. Notice that the overall response time $\overline{R}_{(p)}$ of a preemptive system may be computed as though customer classes $p+1, p+2, \ldots$ are absent.

It follows that the problem of finding approximations to the response times of the individual classes is reduced to the problem of finding approximate values for the $\overline{R}_{(p)}$'s.

When the service time distributions of all classes are exponential with the same mean $1/\mu$, the overall response time $\overline{R}_{(p)}$ of a p-priority level system with combined arrival rate $\lambda_{(p)}$ is given exactly as in equation (6), with $a=\lambda_{(p)}$ and $u=\mu$. Formally,

$$R(PRI, \mu\underline{e}, \underline{\lambda}_{(p)}, m) = R(FCFS, \mu\underline{e}, \underline{\lambda}_{(p)}, m) \tag{16}$$

where $\underline{e}$ is a vector of p 1's.

Now, consider the more general case in which the service time distributions are not exponential or the individual values of the μ_i's are not all equal. In this case, equation (16) does not hold because the mean time spent waiting in the queue is not the same under FCFS and PRI scheduling. That is, the ratio

$$\eta = \frac{R(PRI, \underline{\mu}_{(p)}, \underline{\lambda}_{(p)}, m) - 1/\overline{\mu}_{(p)}}{R(FCFS, \underline{\mu}_{(p)}, \underline{\lambda}_{(p)}, m) - 1/\overline{\mu}_{(p)}} \tag{17}$$

is not equal to unity in general unless all the service times are exponential and $\mu_1 = \mu_2 = \cdots = \mu_p = \overline{\mu}_{(p)}$.

Equation (17) defines η as the ratio of two waiting times. In the discussion which follows, $W(.,.,.,.)$ will be used to denote the waiting time of a system with response time $R(.,.,.,.)$. Thus, equation (17) may be rewritten as

$$\eta = \frac{W(PRI,\underline{\mu}_{(p)},\underline{\lambda}_{(p)},m)}{W(FCFS,\underline{\mu}_{(p)},\underline{\lambda}_{(p)},m)}$$

The next step is to develop an approximation for η. Suppose the m servers in the original system are replaced by a single server that is m times as fast.The service rate of priority level p will then be $m\mu_p$, and the arrival rate for that level will remain λ_p. Define the quantity η' for this modified system as follows.

$$\eta' = \frac{W(PRI,m\underline{\mu}_{(p)},\underline{\lambda}_{(p)},1)}{W(FCFS,m\underline{\mu}_{(p)},\underline{\lambda}_{(p)},1)}$$

Note that η and η' both reflect the effect on overall response time of converting from FCFS scheduling to priority scheduling. Even if the service time distributions of the priority levels differ significantly in mean or form, the ratios among the service times will still be the same in the single and multiple server systems because the relationship between service time and order of selection from the queue of waiting customers will be the same. This relationship is the primary factor influencing η and η'.

The principal source of difference between η and η' is the difference in the composition of the queue of waiting customers. The m server system allows m customers to be in service simultaneously. This will, in general, reduce the proportion of waiting high priority customers relative to the proportion in the single server case. Hence, if m is large and the queue of high priority customers is long, η' may be a less satisfactory approximation for η when the μ_i's differ significantly from each other.

In other cases, η' should be a reasonably good approximation for η, and the analysis will continue with the following assumption, which was also used in [3]:

$$\frac{W(PRI,\underline{\mu}_{(p)},\underline{\lambda}_{(p)},m)}{W(FCFS,\underline{\mu}_{(p)},\underline{\lambda}_{(p)},m)} \approx \frac{W(PRI,m\underline{\mu}_{(p)},\underline{\lambda}_{(p)},1)}{W(FCFS,m\underline{\mu}_{(p)},\underline{\lambda}_{(p)},1)} \tag{18}$$

Rearranging to obtain approximations to the quantities needed in equation (15),

$$W(PRI,\underline{\mu}_{(p)},\underline{\lambda}_{(p)},m) \approx W(PRI,m\underline{\mu}_{(p)},\underline{\lambda}_{(p)},1)\,\frac{W(FCFS,\underline{\mu}_{(p)},\underline{\lambda}_{(p)},m)}{W(FCFS,m\underline{\mu}_{(p)},\underline{\lambda}_{(p)},1)} \tag{19}$$

To evaluate the right hand side of (19), note first that $R(PRI,m\underline{\mu}_{(p)},\underline{\lambda}_{(p)},1)$ is derivable from the response times of the individual classes, $r_k, k=1,2,...,p$ in an M/G/1 preemptive-resume priority system [7]. The the average response time over the p

highest priority classes is given by

$$R(PRI, m\underline{\mu}_{(p)}, \underline{\lambda}_{(p)}, 1) = \frac{1}{\lambda_{(p)}} \sum_{i=1}^{p} \lambda_i r_i \tag{20}$$

Now consider the remaining terms in equation (19). Let

$$\gamma = \frac{W(FCFS, \underline{\mu}_{(p)}, \underline{\lambda}_{(p)}, m)}{W(FCFS, m\underline{\mu}_{(p)}, \underline{\lambda}_{(p)}, 1)} \tag{21}$$

The overall mean service time in the numerator is clearly $1/\overline{\mu}_{(p)}$, that in the denominator $1/m\overline{\mu}_{(p)}$.

If we let S be a random variable denoting the service time in the m-server system, the overall waiting time in the single server system in (16) will be given by

$$W(FCFS, m\underline{\mu}_{(p)}, \underline{\lambda}_{(p)}, 1) = \frac{\lambda_{(p)} E[S^2]}{2m^2(1-\rho_{(p)})} \tag{22}$$

by the Pollaczek-Khinchin formula. Using the Nozaki-Ross approximation in equation (5), the overall waiting time for the multiple server system is approximately

$$W(FCFS, \underline{\mu}_{(p)}, \underline{\lambda}_{(p)}, m) \approx \frac{\lambda_{(p)} E[S^2](m\rho)^{m-1} p_0(\rho_{(p)})}{2(m-1)!(1-\rho_{(p)})^2} \tag{23}$$

Notice that it is not necessary to obtain $E[S^2]$ in order to compute γ approximately because it appears in the numerator in equations (22)and (23). Substituting into (21) from (22) and (23), $E[S^2]$ cancels and the following expression is derived:

$$\gamma \approx \frac{P_m(\rho_{(p)})}{\rho_{(p)}} \tag{24}$$

This expression for γ in (24) is independent of the distribution of service times. It is the same expression for γ that was derived in [3]. The numerical results in Tables 3 and 5 indicate that the use of this expression for γ will lead to inaccurate answers when the coefficient of variation is larger than one.

Computing γ using the BCH-based waiting times in the numerator of equation (21) usually gives more accurate numerical results. In place of equation (21), let

$$\gamma = \frac{W(FCFS, \overline{\mu}_{(p)}, \lambda_{(p)}, m)}{W(FCFS, m\overline{\mu}_{(p)}, \lambda_{(p)}, 1)} \tag{25}$$

This is a ratio of single class waiting times with the given parameters. It is clear from

272

equation (11) that this expression for γ does depend on $E[S^2]$. For (25) to reduce to (24) in the exponential case, one must assume that the overall service time CV is the average of the class CV's weighted by arrival rates. This is only true for deterministic service. However, the assumption is implicit in the derivation given in [3], and using it gives more accurate numerical results than those based on the theoretically correct CV for all the non-zero CV values investigated here.

Using either the Nozaki-Ross or BCH expressions, we have that

$$\bar{R}_{(p)}=1/\bar{\mu}_{(p)}+W(PRI,\underline{\mu}_{(p)},\underline{\lambda}_{(p)},m)$$
$$\approx 1/\bar{\mu}_{(p)}+[\frac{1}{\lambda_{(p)}}\sum_{k=1}^{p}\lambda_k r_k-\frac{1}{m\bar{\mu}_{(p)}}]\gamma \qquad (26)$$

where the r_k's ($k=1,2,\ldots,r$) are as mentioned above. The approximate values of R_i ($i\geq 2$) may then be derived directly from equation (15). R_1 may be evaluated as though the other classes did not exist, using equation (5) or (11) with $a=\lambda_1$ and $u=\mu_1$.

The approximate method of obtaining response times works well for two and three classes when the service time distributions of all priority levels are exponential (see [3] and Tables 7-9). For general service, the effect of multiple servers and coefficient of variation of service time (CV) on response time must be accounted for. The BCH response times are usually within 10% of the simulated results (Tables 4,6) except when the system's arrival and service parameters make its response times highly variable. For example, the priority ordering in Table 4 results in frequent preemption of customers that do not arrive very often, a policy which would not be recommended if the response time of the low priority jobs were a crucial factor. Another example of instability is the family of systems in Table 9.2. Here, the most frequent arrivals require the shortest service time, but they are blocked by two classes of much longer jobs at higher priority levels. By contrast, Table 6 illustrates the response times of a balanced family, and the relative errors there are correspondingly small.

4. Conclusion

A simple method for computing the approximate mean response times of individual customer classes in an $M/G/m$ preemptive resume priority system has been given. It attempts to account for the influence of preemption on each priority level as the parameters and service time distribution of each class are considered. Comparisons with published results for the two priority case with exponential service, for which an exact solution exists, show that the approximation is accurate to within 5% in most cases involving exponential service. Comparisons with simulation results show that the approximation correctly predicts the qualitative behavior of a system with two priority levels when the service time coefficient of variation differs from unity, and at far less cost than a simulation would. Because of its logical consistency and ease of implementation, the approximation should enjoy a wide range of applications to the modeling of priority systems.

Acknowledgements

This research was supported in part by NSF grant number MCS78-01729. Computing facilities were provided by the Department of Computer Sciences at Purdue University. The first author has benefited from discussions with Michael Taaffe and Bruce Schmeiser. We would like to thank Subhash C. Agrawal for his valuable comments. We also wish to thank Dr. Daniel Heyman of Bell Laboratories for drawing our attention to the paper [9] by Nozaki and Ross, and an anonymous reviewer for his reference to

References

1. Barberis, G., "A Useful Tool in the Theory of Priority Queueing," *IEEE Trans. on Comm.* **COM-28**(9) pp. 1753-1761 (1980).

2. Boxma, O.J., Cohen, J.W., and Huffels, N., "Approximations of the Mean Waiting Time in an M/G/s Queueing System," *Ops. Res.* **27** pp. 1115-1127 (1979).

3. Buzen, J.P. and Bondi, A.B., "The Response Times of Priority Classes under Preemptive Resume in M/M/m Queues," *Opns. Res.* **31**(3) pp. 456-465 (1983).

4. Cobham, A., "Priority Assignment - A Correction," *Opns. Res.* **3** p. 547 (1955).

5. Cosmetatos, G.P., "Approximate explicit formulae for the average queueing time in the processes M/D/r and D/M/r," *Informatics*, pp. 328-331 (1975).

6. Groenevelt, H., Hoorn, M.H. van, and Tijms, H.C., "Tables for M/G/c queueing systems with phase-type service," Res. Report no. 85, Dept. of Act. Sci. and Econometrics, Vrije Universiteit Amsterdam (July 1982).

7. Kleinrock, L., *Queueing Systems, Volume 2: Computer Applications*, John Wiley, New York (1976).

8. Mitrani, I. and King, P.J.B., "Multiprocessor Systems with Preemptive Priorities," *Performance Evaluation* **1** pp. 118-125 (1981).

9. Nozaki, S.A. and Ross, S.M., "Approximations in Finite-Capacity Multiserver Queues with Poisson Arrivals," *J. Appl. Prob.* **15** pp. 826-834 (1978).

10. Schmeiser, B., "Batch size effects in the analysis of simulation output," *Opns. Res.* **30**(3) pp. 556-568 (1982).

Table 1: Comparison of Approximate and Exact Response Times for N Servers One Class with Heavy Traffic, One Class with Light Traffic						
	Class 1			Class 2		
Arrival Rates	N*1.600			N*0.050		
Service Rates	2.000			1.000		
Traffic Intensity	0.8			0.05		
N	Exact	Approx.	Rel. Err.	Exact	Approx.	Rel. Err.
2	1.389	1.389	0.000	10.465	10.691	0.022
4	0.873	0.873	0.000	5.680	5.899	0.039
6	0.716	0.716	0.000	4.073	4.252	0.044
8	0.642	0.642	0.000	3.262	3.412	0.046
10*	0.602	0.602	0.000	2.728	2.900	0.063

* The exact response times in this row differ from the corresponding ones in Mitrani and King (1981). We have been informed by Dr. King that the numbers shown here are the correct ones.

Table 2: Comparison of Approximate and Exact Response Times for N Servers Both Classes with Moderate Traffic						
	Class 1			Class 2		
Arrival Rates	0.450			0.300		
Service Rates	1.000/N			2.000/N		
Traffic Intensity	0.45			0.15		
N	Exact	Approx.	Rel. Err.	Exact	Approx.	Rel. Err.
2	2.508	2.508	0.000	3.288	3.288	0.000
4	4.234	4.234	0.000	3.608	3.574	0.010
6	6.118	6.118	0.000	4.161	4.140	0.005
8	8.063	8.063	0.000	4.850	4.841	0.002
10	10.034	10.034	0.000	5.648	5.628	0.004

374

Table 3: Predicted Response Times and Confidence Limits of High and Low Priority Levels with Mixed Traffic and N Servers; Response Times based on Nozaki-Ross Approximation

Parameters for N Servers	Priority High	Priority Low
Arrival rate	N*1.6	N*0.05
Service rate	2.0	1.0

A: Coefficient of Variation 0

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	0.944	0.965	0.030	0.944	0.987	0.948	0.983
	2	6.764	7.401	0.521	7.029	7.774	7.100	7.702
4	1	0.686	0.691	0.008	0.686	0.697	0.687	0.696
	2	3.855	4.146	0.222	3.987	4.304	4.017	4.274
6	1	0.608	0.613	0.003	0.612	0.615	0.611	0.615
	2	2.871	3.139	0.126	3.048	3.229	3.066	3.212
8	1	0.572	0.576	0.003	0.574	0.578	0.574	0.577
	2	2.375	2.600	0.109	2.522	2.678	2.537	2.663
10	1	0.551	0.555	0.002	0.555	0.556	0.554	0.556
	2	2.075	2.270	0.063	2.225	2.315	2.235	2.307

B: Coefficient of Variation 1

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	1.389	1.388	0.124	1.300	1.477	1.316	1.460
	2	10.691	9.882	1.348	8.918	10.846	9.101	10.664
4	1	0.873	0.881	0.035	0.856	0.906	0.861	0.902
	2	5.899	5.950	0.599	5.521	6.378	5.602	6.297
6	1	0.716	0.709	0.013	0.700	0.718	0.701	0.717
	2	4.253	3.919	0.194	3.780	4.058	3.807	4.032
8	1	0.643	0.645	0.011	0.637	0.653	0.638	0.651
	2	3.412	0.229	0.229	3.218	3.546	3.249	3.514
10	1	0.602	0.603	0.008	0.597	0.609	0.598	0.608
	2	2.901	2.833	0.241	2.661	3.005	2.693	2.972

Table 3: Confidence Limits of High and Low Priority Response Times with Mixed Traffic and N Servers (Nozaki-Ross Method), continued

C: Coefficient of Variation root 3

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.278	2.185	0.143	2.083	2.288	2.102	2.268
	2	18.544	16.305	2.961	14.187	18.423	14.588	18.021
4	1	1.246	1.187	0.054	1.149	1.225	1.156	1.218
	2	9.987	8.401	0.651	7.935	8.867	8.024	8.779
6	1	0.932	0.893	0.055	0.854	0.933	0.862	0.925
	2	7.016	6.086	1.067	5.323	6.849	5.467	6.704
8	1	0.786	0.739	0.032	0.716	0.762	0.721	0.758
	2	5.488	3.988	0.456	3.662	4.314	3.724	4.253
10	1	0.705	0.668	0.028	0.648	0.688	0.652	0.684
	2	4.552	3.532	0.444	3.214	3.850	3.274	3.789

D: Coefficient of Variation 3

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	4.944	5.066	1.309	4.130	6.002	4.308	5.825
	2	42.102	45.765	32.654	22.407	69.123	26.837	64.693
4	1	2.364	2.137	0.203	1.992	2.281	2.019	2.254
	2	22.250	17.525	4.719	14.149	20.900	14.789	20.260
6	1	1.588	1.287	0.131	1.193	1.380	1.211	1.363
	2	15.305	9.744	1.977	8.330	11.158	8.598	10.890
8	1	1.215	1.084	0.079	1.027	1.141	1.038	1.130
	2	11.714	7.911	1.625	6.749	9.074	6.970	8.853
10	1	1.012	0.863	0.060	0.820	0.906	0.828	0.898
	2	9.505	5.449	0.690	4.955	5.942	5.049	5.849

Table 4: Predicted Response Times and Confidence Limits of High and Low Priority Levels with Mixed Traffic and N Servers; Response Times based on BCH Approximation

Parameters for N Servers	Priority High	Priority Low
Arrival rate	N*1.6	N*0.05
Service rate	2.0	1.0

A: Coefficient of Variation 0

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	0.950	0.965	0.030	0.944	0.987	0.948	0.983
	2	6.763	7.401	0.521	7.029	7.774	7.100	7.703
4	1	0.693	0.691	0.008	0.686	0.697	0.687	0.696
	2	3.864	4.146	0.222	3.987	4.304	4.017	4.274
6	1	0.613	0.613	0.003	0.612	0.615	0.611	0.615
	2	2.886	3.139	0.126	3.048	3.229	3.066	3.212
8	1	0.576	0.576	0.003	0.574	0.578	0.574	0.577
	2	2.393	2.600	0.109	2.522	2.678	2.537	2.663
10	1	0.555	0.555	0.002	0.555	0.556	0.554	0.556
	2	2.096	2.270	0.063	2.225	2.315	2.235	2.307

B: Coefficient of Variation 1

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	1.389	1.388	0.124	1.300	1.477	1.316	1.460
	2	10.691	9.882	1.348	8.918	10.846	9.101	10.664
4	1	0.873	0.881	0.035	0.856	0.906	0.861	0.902
	2	5.899	5.950	0.599	5.521	6.378	5.602	6.297
6	1	0.716	0.709	0.013	0.700	0.718	0.701	0.717
	2	4.253	3.919	0.194	3.780	4.058	3.807	4.032
8	1	0.643	0.645	0.011	0.637	0.653	0.638	0.651
	2	3.412	0.229	0.229	3.218	3.546	3.249	3.514
10	1	0.602	0.603	0.008	0.597	0.609	0.598	0.608
	2	2.900	2.833	0.241	2.661	3.005	2.693	2.972

Table 4: Confidence Limits of High and Low Priority Response Times with Mixed Traffic and N Servers (BCH Method), continued

C: Coefficient of Variation root 3								
N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.204	2.185	0.143	2.083	2.288	2.102	2.268
	2	18.689	16.305	2.961	14.187	18.423	14.588	18.021
4	1	1.191	1.187	0.054	1.149	1.225	1.156	1.218
	2	9.982	8.401	0.651	7.935	8.867	8.024	8.779
6	1	0.890	0.893	0.055	0.854	0.933	0.862	0.925
	2	6.936	6.086	1.067	5.323	6.849	5.467	6.704
8	1	0.753	0.739	0.032	0.716	0.762	0.721	0.758
	2	5.375	3.988	0.456	3.662	4.314	3.724	4.253
10	1	0.678	0.668	0.028	0.648	0.688	0.652	0.684
	2	4.420	3.532	0.444	3.214	3.850	3.274	3.789

D: Coefficient of Variation 3								
N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	4.564	5.066	1.309	4.130	6.002	4.308	5.825
	2	42.870	45.765	32.654	22.407	69.123	26.837	64.693
4	1	1.991	2.137	0.203	1.992	2.281	2.019	2.254
	2	21.944	17.525	4.719	14.149	20.900	14.789	20.260
6	1	1.283	1.287	0.131	1.193	1.380	1.211	1.363
	2	14.374	9.744	1.977	8.330	11.158	8.598	10.890
8	1	0.983	1.084	0.079	1.027	1.141	1.038	1.130
	2	10.520	7.911	1.625	6.749	9.074	6.970	8.853
10	1	0.827	0.863	0.060	0.820	0.906	0.828	0.898
	2	8.215	5.449	0.690	4.955	5.942	5.049	5.849

Table 5: Predicted Response Times and Confidence Limits of High and Low Priority Levels with Moderate Traffic and N Servers; Response Times based on Nozaki-Ross Approximation

Parameters for N Servers	Priority High	Priority Low
Arrival rate	0.45	0.30
Service rate	1.0/N	2.0/N

A: Coefficient of Variation 0

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.254	2.266	0.021	2.251	2.281	2.254	2.278
	2	2.281	2.358	0.153	2.249	2.467	2.269	2.446
4	1	4.117	4.130	0.016	4.119	4.142	4.121	4.139
	2	2.885	3.023	0.160	2.908	3.137	2.930	3.116
6	1	6.059	6.067	0.011	6.059	6.076	6.061	6.074
	2	3.637	3.773	0.132	3.678	3.867	3.696	3.849
8	1	8.031	8.037	0.008	8.032	8.042	8.033	8.041
	2	4.468	4.580	0.126	4.490	4.670	4.507	4.653
10	1	10.017	10.019	0.007	10.014	10.024	10.015	10.023
	2	5.348	5.443	0.098	5.373	5.512	5.386	5.499

B: Coefficient of Variation 1

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.508	2.494	0.143	2.392	2.596	2.412	2.577
	2	3.255	3.230	0.411	2.937	3.524	2.992	3.469
4	1	4.233	4.170	0.148	4.064	4.276	4.084	4.256
	2	3.574	3.477	0.422	3.175	3.779	3.232	3.722
6	1	6.119	6.092	0.163	5.976	6.209	5.998	6.187
	2	4.140	4.088	0.275	3.892	4.285	3.929	4.248
8	1	8.063	7.946	0.174	7.822	8.071	7.845	8.047
	2	4.841	4.953	0.292	4.745	5.162	4.784	5.122
10	1	10.034	9.991	0.235	9.823	10.159	9.855	10.127
	2	5.628	5.637	0.235	5.469	5.805	5.501	5.773

Table 5: Confidence Limits of High and Low Priority Response Times with Moderate Traffic and N Servers (Nozaki-Ross Method), continued

C: Coefficient of Variation root 3								
N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	3.016	2.821	0.253	2.640	3.002	2.674	2.967
	2	5.024	4.587	0.755	4.047	5.127	4.149	5.025
4	1	4.674	4.270	0.344	4.024	4.516	4.071	4.469
	2	4.952	4.477	1.306	3.550	5.410	3.726	5.229
6	1	6.237	6.046	0.262	5.859	6.234	5.894	6.198
	2	5.146	4.332	0.547	3.941	4.723	4.015	4.649
8	1	8.126	8.098	0.333	7.860	8.337	7.905	8.291
	2	5.587	5.057	0.979	4.356	5.757	4.489	5.624
10	1	10.069	9.986	0.431	9.678	10.294	9.736	10.236
	2	6.186	5.806	0.366	5.545	6.068	5.594	6.019

D: Coefficient of Variation 3								
N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	4.539	4.148	1.318	3.205	5.091	3.384	4.912
	2	11.049	9.159	3.308	6.793	11.525	7.242	11.076
4	1	5.169	4.587	0.784	4.026	5.148	4.133	5.042
	2	9.087	6.535	3.211	4.238	8.832	4.674	8.396
6	1	6.592	6.121	0.697	5.623	6.620	5.717	6.525
	2	6.121	5.076	1.565	3.957	6.196	4.169	5.983
8	1	3.315	8.083	0.810	7.504	8.663	7.614	8.553
	2	7.825	5.719	1.748	4.469	6.969	4.706	6.732
10	1	10.172	9.873	0.908	9.223	10.522	9.346	10.399
	2	9.873	6.804	1.755	5.549	8.059	5.787	7.821

Table 6: Predicted Response Times and Confidence Limits of High and Low Priority Levels with Moderate Traffic and N Servers; Response Times based on BCH Approximation

Parameters for N Servers	Priority High	Priority Low
Arrival rate	0.45	0.30
Service rate	1.0/N	2.0/N

A: Coefficient of Variation 0

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.271	2.266	0.021	2.251	2.281	2.254	2.278
	2	2.316	2.358	0.153	2.249	2.467	2.269	2.446
4	1	4.136	4.130	0.016	4.119	4.142	4.121	4.139
	2	2.952	3.023	0.160	2.908	3.137	2.930	3.116
6	1	6.074	6.067	0.011	6.059	6.076	6.061	6.074
	2	3.712	3.773	0.132	3.678	3.867	3.696	3.849
8	1	8.041	8.037	0.008	8.032	8.042	8.033	8.041
	2	4.541	4.580	0.126	4.490	4.670	4.507	4.653
10	1	10.024	10.019	0.007	10.014	10.024	10.015	10.023
	2	5.414	5.443	0.098	5.373	5.512	5.386	5.499

B: Coefficient of Variation 1

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.508	2.494	0.143	2.392	2.596	2.412	2.577
	2	3.255	3.230	0.411	2.937	3.524	2.992	3.469
4	1	4.234	4.170	0.148	4.064	4.276	4.084	4.256
	2	3.574	3.477	0.422	3.175	3.779	3.232	3.722
6	1	6.119	6.092	0.163	5.976	6.209	5.998	6.187
	2	4.140	4.088	0.275	3.892	4.285	3.929	4.248
8	1	8.063	7.946	0.174	7.822	8.071	7.845	8.047
	2	4.841	4.953	0.292	4.745	5.162	4.784	5.122
10	1	10.034	9.991	0.235	9.823	10.159	9.855	10.127
	2	5.628	5.637	0.235	5.469	5.805	5.501	5.773

Table 6: Confidence Limits of High and Low Priority Response Times with Moderate Traffic and N Servers (BCH Method), continued

C: Coefficient of Variation root 3								
N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.845	2.821	0.253	2.640	3.002	2.674	2.967
	2	4.876	4.587	0.755	4.047	5.127	4.149	5.025
4	1	4.348	4.270	0.344	4.024	4.516	4.071	4.469
	2	4.520	4.477	1.306	3.550	5.410	3.726	5.229
6	1	6.143	6.046	0.262	5.859	6.234	5.894	6.198
	2	4.372	4.332	0.547	3.941	4.723	4.015	4.649
8	1	8.083	8.098	0.333	7.860	8.337	7.905	8.291
	2	5.225	5.057	0.979	4.356	5.757	4.489	5.624
10	1	10.044	9.986	0.431	9.678	10.294	9.736	10.236
	2	5.884	5.806	0.366	5.545	6.068	5.594	6.019

D: Coefficient of Variation 3								
N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	3.342	4.148	1.318	3.205	5.091	3.384	4.912
	2	9.486	9.159	3.308	6.793	11.525	7.242	11.076
4	1	4.178	4.587	0.784	4.026	5.148	4.133	5.042
	2	6.576	6.535	3.211	4.238	8.832	4.674	8.396
6	1	6.022	6.121	0.697	5.623	6.620	5.717	6.525
	2	5.782	5.076	1.565	3.957	6.196	4.169	5.983
8	1	8.003	8.083	0.810	7.504	8.663	7.614	8.553
	2	5.823	5.719	1.748	4.469	6.969	4.706	6.732
10	1	10.000	9.873	0.908	9.223	10.522	9.346	10.399
	2	6.246	6.804	1.755	5.549	8.059	5.787	7.821

Table 7.0: Predicted Response Times and Confidence Bounds for Three Priority Levels, Moderate Traffic, N Servers, and Deterministic Service

Parameters for N Servers	Priority 1	Priority 2	Priority 3
Arrival rate	0.45	0.30	1.0
Service rate	2.0/N	1.0/N	5.0/N

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.271	2.277	0.034	2.253	2.302	2.256	2.297
	2	2.316	2.400	0.160	2.285	2.514	2.307	2.493
	3	3.934	4.117	0.636	3.662	4.572	3.748	4.486
4	1	4.136	4.141	0.028	4.122	4.161	4.125	4.158
	2	2.952	3.050	0.143	2.948	3.152	2.968	3.133
	3	3.919	4.159	0.659	3.688	4.630	3.777	4.541
6	1	6.074	6.077	0.022	6.061	6.093	6.064	6.090
	2	3.712	3.806	0.143	3.704	3.909	3.723	3.889
	3	4.008	4.249	0.673	3.768	4.731	3.859	4.640
8	1	8.041	8.044	0.018	8.031	8.056	8.033	8.054
	2	4.541	4.618	0.144	4.516	4.721	4.535	4.702
	3	4.154	4.377	0.673	3.896	4.858	3.987	4.767
10	1	10.024	10.028	0.016	10.016	10.039	10.018	10.037
	2	5.414	5.465	0.135	5.368	5.561	5.387	5.543
	3	4.338	4.534	0.676	4.051	5.017	4.142	4.926

Table 7.1: Predicted Response Times and Confidence Bounds for Three Priority Levels, Moderate Traffic, N Servers, and Exponential Service

Parameters for N Servers	Priority 1	Priority 2	Priority 3
Arrival rate	0.45	0.30	1.0
Service rate	2.0/N	1.0/N	5.0/N

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.508	2.520	0.163	2.403	2.636	2.426	2.614
	2	3.255	3.393	0.584	2.975	3.812	3.055	3.732
	3	7.112	7.097	1.688	5.890	8.305	6.119	8.076
4	1	4.234	4.262	0.135	4.165	4.359	4.184	4.340
	2	3.574	3.640	0.449	3.319	3.962	3.380	3.901
	3	6.611	6.798	1.250	5.904	7.693	6.074	7.523
6	1	6.118	6.093	0.107	6.016	6.169	6.031	6.155
	2	4.140	4.102	0.262	3.915	4.289	3.950	4.253
	3	6.351	7.130	2.262	5.512	8.748	5.819	8.441
8	1	8.063	8.027	0.163	7.911	8.143	7.933	8.121
	2	4.841	4.787	0.262	4.600	4.975	4.636	4.939
	3	6.221	6.507	1.578	5.378	7.636	5.592	7.422
10	1	10.034	10.050	0.273	9.855	10.246	9.892	10.209
	2	5.628	5.687	0.384	5.413	5.962	5.465	5.910
	3	6.179	6.754	1.854	5.428	8.081	5.679	7.829

Table 7.2: Predicted Response Times and Confidence Bounds for Three Priority Levels, Moderate Traffic, N Servers, and Hyperexponential Service; Coefficient of Variation 1.732 (root 3)

Parameters for N Servers	Priority 1	Priority 2	Priority 3
Arrival rate	0.45	0.30	1.0
Service rate	2.0/N	1.0/N	5.0/N

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	2.845	2.965	0.304	2.747	3.182	2.788	3.141
	2	4.875	4.977	1.236	4.092	5.861	4.260	5.693
	3	13.095	12.719	4.585	9.439	15.999	10.061	15.377
4	1	4.348	4.600	0.391	4.321	4.880	4.374	4.827
	2	4.520	5.224	1.483	4.163	6.284	4.365	6.083
	3	11.432	14.976	6.367	10.422	19.531	11.286	18.667
6	1	6.164	6.133	0.230	5.968	6.298	6.000	6.267
	2	4.732	4.218	0.401	3.931	4.504	3.985	4.450
	3	10.378	9.043	2.936	6.943	11.143	7.341	10.744
8	1	8.083	8.113	0.212	7.962	8.264	7.990	8.236
	2	5.225	5.058	0.794	4.490	5.626	4.598	5.518
	3	9.657	9.243	3.378	6.826	11.659	7.284	11.201
10	1	10.044	10.198	0.534	9.816	10.580	9.888	10.507
	2	5.884	5.711	0.636	5.256	6.166	5.343	6.080
	3	9.150	9.558	4.316	6.471	12.646	7.056	12.060

Table 7.5: Predicted Response Times and Confidence Bounds for
Three Priority Levels, Moderate Traffic, N Servers, and Hyperexponential Service;
Coefficient of Variation 3.0

Parameters for N Servers	Priority 1	Priority 2	Priority 3
Arrival rate	0.45	0.30	1.0
Service rate	2.0/N	1.0/N	5.0/N

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	3.770	4.446	0.970	3.752	5.140	3.884	5.008
	2	9.485	11.951	5.014	8.365	15.538	9.045	14.858
	3	30.512	49.453	20.249	34.968	63.937	37.715	61.190
4	1	4.561	4.599	0.562	4.197	5.001	4.274	4.925
	2	6.578	5.828	1.294	4.902	6.753	5.078	6.578
	3	23.724	21.573	7.510	16.202	26.945	17.220	25.926
6	1	6.231	6.637	0.820	6.051	7.223	6.162	7.112
	2	5.782	6.767	2.985	4.632	8.903	5.037	8.498
	3	19.587	26.703	17.332	14.306	39.101	16.657	36.750
8	1	8.108	7.856	0.452	7.532	8.179	7.593	8.118
	2	5.822	5.376	0.755	4.836	5.916	4.939	5.814
	3	16.910	14.765	5.364	10.928	18.602	11.655	17.874
10	1	10.054	10.018	0.891	9.381	10.656	9.502	10.535
	2	6.246	6.791	2.722	4.844	8.737	5.213	8.368
	3	15.047	15.907	12.379	7.052	24.762	8.731	23.083

Table 8.0: Predicted Response Times and Confidence Bounds for Three Priority Levels, Moderate Traffic, N Servers, and Deterministic Service; Alternative Ordering

Parameters for N Servers	Priority 1	2	3
Arrival rate	1.0	0.45	0.30
Service rate	5.0/N	2.0/N	1.0/N

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	0.410	0.410	0.001	0.409	0.410	0.409	0.410
	2	2.935	3.037	0.059	2.994	3.079	3.002	3.071
	3	5.632	5.855	0.582	5.438	6.272	5.517	6.193
4	1	0.802	0.802	0.001	0.801	0.802	0.801	0.802
	2	4.687	4.811	0.054	4.772	4.850	4.779	4.842
	3	6.156	6.543	0.608	6.108	6.978	6.191	6.895
6	1	1.200	1.200	0.000	1.200	1.201	1.200	1.201
	2	6.524	6.640	0.047	6.606	6.673	6.612	6.667
	3	6.789	7.248	0.644	6.788	7.709	6.875	7.621
8	1	1.600	1.600	0.000	1.600	1.600	1.600	1.600
	2	8.409	8.507	0.053	8.469	8.545	8.476	8.538
	3	7.483	7.937	0.605	7.504	8.370	7.587	8.288
10	1	2.000	2.000	0.000	2.000	2.000	2.000	2.000
	2	10.324	10.413	0.049	10.378	10.447	10.384	10.441
	3	8.217	8.650	0.585	8.231	9.068	8.311	8.989

Table 8.1: Predicted Response Times and Confidence Bounds for Three Priority Levels, Moderate Traffic, N Servers, and Exponential Service; Alternative Ordering

Parameters for N Servers	Priority 1	2	3
Arrival rate	1.0	0.45	0.30
Service rate	5.0/N	2.0/N	1.0/N

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	Upper	90% Limits Lower	Upper
2	1	0.417	0.414	0.005	0.411	0.418	0.412	0.417
	2	3.626	3.590	0.341	3.346	3.833	3.392	3.787
	3	9.320	8.224	1.417	7.210	9.238	7.403	9.046
4	1	0.802	0.806	0.010	0.799	0.813	0.800	0.811
	2	5.145	5.080	0.222	4.921	5.239	4.952	5.209
	3	9.344	9.692	2.800	7.689	11.696	8.069	11.316
6	1	1.200	1.196	0.012	1.187	1.204	1.189	1.203
	2	6.847	6.781	0.258	6.596	6.965	6.631	6.930
	3	9.603	9.782	1.651	8.601	10.963	8.825	10.739
8	1	1.600	1.595	0.016	1.583	1.606	1.585	1.604
	2	8.644	8.513	0.293	8.304	8.722	8.343	8.683
	3	9.994	8.981	0.899	8.338	9.624	8.460	9.502
10	1	2.000	2.000	0.019	1.986	2.013	1.989	2.011
	2	10.499	10.450	0.414	10.153	10.746	10.209	10.690
	3	10.474	11.085	2.509	9.290	12.880	9.631	12.540

Table 6.2: Predicted Response Times and Confidence Bounds for Three Priority Levels, Moderate Traffic, N Servers, and Hyperexponential Service; Coefficient of Variation 1.732; Alternative Ordering

Parameters for N Servers	Priority 1	2	3
Arrival rate	1.0	0.45	0.30
Service rate	5.0/N	2.0/N	1.0/N

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	0.421	0.431	0.024	0.414	0.448	0.417	0.444
	2	4.818	4.888	1.031	4.151	5.626	4.291	5.486
	3	16.318	16.861	7.526	11.477	22.244	12.498	21.223
4	1	0.803	0.805	0.027	0.786	0.825	0.790	0.821
	2	5.853	5.809	0.625	5.362	6.256	5.447	6.171
	3	15.116	15.256	4.427	12.090	18.423	12.690	17.822
6	1	1.200	1.204	0.044	1.172	1.235	1.178	1.229
	2	7.307	6.857	0.531	6.477	7.237	6.549	7.165
	3	14.500	11.651	3.438	9.192	14.111	9.658	13.644
8	1	1.600	1.602	0.029	1.582	1.623	1.586	1.619
	2	8.960	8.927	0.846	8.322	9.532	8.437	9.418
	3	14.218	15.259	7.106	10.176	20.343	11.140	19.378
10	1	2.000	2.017	0.034	1.993	2.042	1.997	2.037
	2	10.723	10.728	0.541	10.341	11.115	10.414	11.041
	3	14.158	14.035	3.726	11.370	16.701	11.875	16.195

Table 8.3: Predicted Response Times and Confidence Bounds for Three Priority Levels, Moderate Traffic, N Servers, and Hyperexponential Service; Coefficient of Variation 3.0; Alternative Ordering

Parameters for N Servers	Priority 1	2	3
Arrival rate	1.0	0.45	0.30
Service rate	5.0/N	2.0/N	1.0/N

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	0.437	0.446	0.035	0.421	0.471	0.425	0.466
	2	8.200	.7.321	1.913	5.953	8.690	6.213	8.430
	3	36.787	28.394	11.572	20.116	36.671	21.686	35.101
4	1	0.803	0.808	0.032	0.785	0.831	0.789	0.827
	2	7.424	8.286	3.187	6.006	10.566	6.439	10.133
	3	30.030	33.831	21.051	18.773	48.889	21.629	46.033
6	1	1.200	1.168	0.045	1.136	1.200	1.142	1.194
	2	8.163	8.233	1.432	7.209	9.258	7.403	9.064
	3	25.871	22.661	14.194	12.507	32.814	14.433	30.888
8	1	1.600	1.587	0.061	1.543	1.631	1.551	1.622
	2	9.481	9.824	1.024	9.091	10.557	9.230	10.418
	3	23.273	21.901	8.703	15.676	28.127	16.856	26.946
10	1	2.000	2.005	0.094	1.937	2.072	1.950	2.060
	2	11.062	10.687	1.174	9.847	11.527	10.006	11.368
	3	21.577	17.102	9.095	10.597	23.608	11.830	22.374

Table 9.1: Predicted Response Times and Confidence Bounds for Three Priority Levels with Mixed Traffic, N Servers, and Exponential Service

Parameters for N Servers	Priority 1	Priority 2	Priority 3
Arrival rate	1.0*N	4.0*N	0.2*N
Service rate	20.0	40.0	0.2857
Traffic intensity	0.05	0.10	0.70

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	0.050	0.050	0.001	0.049	0.051	0.050	0.050
	2	0.026	0.026	0.000	0.026	0.026	0.026	0.026
	3	12.686	12.676	3.153	10.621	15.132	11.048	14.704
4	1	0.050	0.050	0.000	0.050	0.050	0.050	0.050
	2	0.025	0.025	0.000	0.025	0.025	0.025	0.025
	3	7.563	7.260	0.654	6.792	7.728	6.881	7.639
6	1	0.050	0.050	0.000	0.050	0.050	0.050	0.050
	2	0.025	0.025	0.000	0.025	0.025	0.025	0.025
	3	5.953	6.081	0.868	5.460	6.701	5.578	6.584
8	1	0.050	0.050*	0.001	0.050	0.051	0.050	0.051
	2	0.025	0.025	0.000	0.025	0.025	0.025	0.025
	3	5.168	5.187	0.909	4.537	5.837	4.660	5.714
10	1	0.050	0.050*	0.001	0.049	0.050	0.050	0.050
	2	0.025	0.025	0.000	0.025	0.025	0.025	0.025
	3	4.750	4.942	1.238	4.057	5.828	4.225	5.660

*The runs for 8 and 10 servers were considerably shorter than the others.

Table 9.2: Predicted Response Times and Confidence Bounds for Three Priority Levels with Mixed Traffic N Servers, and Exponential Service: Alternative Ordering

Parameters for N Servers	Priority 1	2	3
Arrival rate	1.0*N	0.2*N	4.0*N
Service rate	20.0	0.2857	40.0
Traffic intensity	0.05	0.70	0.10

N	Pri.	Pred. RT	Sim. Mean RT	Sim. Std. Dev.	95% Limits Lower	95% Limits Upper	90% Limits Lower	90% Limits Upper
2	1	0.050	0.050	0.001	0.049	0.050	0.049	0.050
	2	8.011	7.978	1.101	7.191	8.766	7.340	8.617
	3	30.170	28.785	6.653	24.026	33.545	24.929	32.642
4	1	0.050	0.050	0.000	0.050	0.050	0.050	0.050
	2	5.288	5.256	0.324	5.024	5.488	5.068	5.444
	3	13.337	13.276	3.060	11.087	15.466	11.503	15.050
6	1	0.050	0.050	0.000	0.050	0.050	0.050	0.050
	2	4.487	4.540	0.259	4.355	4.726	4.390	4.691
	3	8.064	9.299	3.140	7.053	11.545	7.479	11.119
8	1	0.050	0.050	0.000	0.050	0.050	0.050	0.050
	2	4.126	4.059	0.167	3.940	4.178	3.962	4.156
	3	5.560	6.017	1.637	4.846	7.187	5.068	6.965
10	1	0.050	0.050	0.000	0.050	0.050	0.050	0.050
	2	3.931	3.988	0.182	3.857	4.118	3.882	4.093
	3	4.124	5.396	2.585	3.547	7.245	3.898	6.895