



# An Approximate Numerical Solution for Multiclass Preemptive Priority Queues with General Service Time Distributions

by

Patricia M. Snyder\* William J. Stewart\*\*

\*Operations Research Program, North Carolina State University, Raleigh, N.C. 27695

\*\*Department of Computer Science, North Carolina State University, Raleigh, N.C. 27695

This work was supported in part by NSF Grant MCS-83-00438

In this paper an approximate numerical solution for a multiclass preemptive priority single server queue is developed. The arrival process of each class follows a Poisson distribution. The service time distribution must have a rational Laplace transform, but is otherwise arbitrary and may be different for different classes. The work reported here was motivated by a desire to compute the equilibrium probability distribution of networks containing preemptive priority servers. Such networks are frequently encountered when modeling computer systems, medical care delivery systems and communication networks. We wish to use an iterative technique which constructs a series of two station networks consisting of one station from the original network and one "complementary" station whose behavior with respect to the original station mimics that of the rest of the network. At each iteration, it is necessary to compute the equilibrium probability distribution of one or more preemptive priority queues.

Although such queues have been studied for some time, the resulting solutions have most often been developed utilizing transforms or probability generating functions, e.g. Jaiswal [1968]; in many cases of interest, inversion has not been attempted. Miller [1981] presented explicit solutions for two class priority queues but Miller's work, which is based on that of Neuts, is limited to exponential service times and two classes. The approach presented here is applicable to many classes and to more general service time distributions than have previously been considered.

The algorithm utilizes a bootstrap approach, a concept borrowed from dynamic programming. The solution for class 1 is trivial. Once we have solved the

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

system with  $k$  different classes, we have available all the necessary information to solve the system with  $k+1$  classes. We shall assume that each class has a distinct service time distribution  $G_k$ , with mean  $g_k$  and variance  $s_k^2$ . Let class  $k$  have preemptive priority over class  $l$  if  $k < l$ .

Successive steps of the algorithm are based upon the machine breakdown and repair model, previously used by Keilson [1962]; White and Christie [1958]; Gaver [1962]; and Avi-Itzhak and Noar [1963] to model preemptive priority queues. When we are solving for the equilibrium probability distribution of class  $k$  jobs,  $k > 1$ , we consider a model with one machine whose service time distribution is  $G_k$ . The breakdown rate of the machine is the sum of the arrival rates of all higher priority jobs. The downtime or repair time of the machine has mean  $\gamma_{k-1}$  and variance  $\sigma_{k-1}^2$ ; these parameters are the mean and variance of the busy period in a preemptive priority system with  $k-1$  classes.

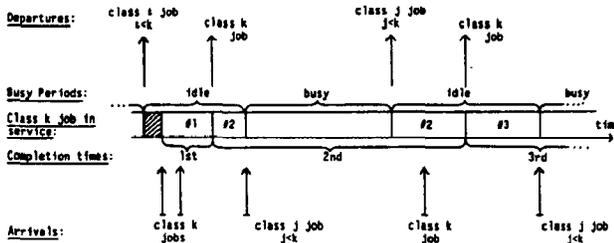
The first step in the solution of all the machine breakdown and repair models considered herein is to construct the infinitesimal generator  $Q$  which specifies the transitions into and out of each state for the given model. Given  $Q$ , it is a simple matter to write the global balance equations, i.e., the Chapman-Kolmogorov equations which relate flow into and out of each state of the model at equilibrium. These global balance equations are second order difference equations, i.e. if  $\underline{X}_k(n)$  is the probability of finding  $n$  class  $k$  jobs in the system at equilibrium,  $\underline{X}_k(n)$  is a function of  $\underline{X}_k(n-1)$  and  $\underline{X}_k(n+1)$ . The next step is to construct another set of balance equations, which we call the aggregate balance equations, and to use the results in Snyder and Stewart [1985] to reduce each of the second order difference equations (the global balance equations) to first order difference equations. As long as the generator  $Q$  is irreducible and positive recurrent, it is now possible to define a coefficient matrix  $R_k$  which relates the probability vectors  $\underline{X}_k(n)$  and  $\underline{X}_k(n-1)$ ;

specifically, we shall construct real matrices  $R_k$  whose elements are given explicitly as a function of the model parameters and for which  $\underline{X}_k(n) = \underline{X}_k(n-1) R_k$ .

## 1. Background

When solving for the marginal equilibrium probability distribution of class  $k$  jobs,  $k > 1$ , we shall consider a machine breakdown and repair model in which the downtime is distributed as the busy period of a preemptive priority queue with classes  $1, 2, \dots, (k-1)$ . Jaiswal [1968] shows that the  $n$ th moment of the busy period in such a preemptive priority queue depends on the 1st through  $n$ th moments of the busy period in a preemptive priority queue with classes  $1, 2, \dots, k-2$ , and on the 1st through  $n$ th moments of  $G_{k-1}(t)$ , the service time distribution of class  $(k-1)$  jobs. Consequently, if we know  $G_k(t)$  for  $k=1, 2, \dots$ , we can generate as many moments of the busy period distribution as we wish. This allows us to express the parameters of the busy period model as functions of the parameters of the service time distributions,  $G_k(t)$ ,  $k=1, 2, \dots$ . If, on the other hand, we are conducting a series of trials in which we measure the service time for each class  $k$  and subsequently compute the experimental moments, Jaiswal's formulae are still applicable.

To utilize Jaiswal's formulae, we introduce a quantity which reflects the time a class  $k$  job actually spends being processed. Let us define completion time as the duration of a period that begins the instant the service of a customer starts and ends the instant the server becomes free to take another customer of the same class. For a preemptive priority system with  $k$  classes we can illustrate these events with the following diagram:



### Busy periods

refer to the busy period distribution in a system with job classes  $1, 2, \dots, k-1$ .

### Completions

refer to the completion times for jobs of class  $k$ .  indicates that no class  $k$  job is present; during this period, the system may actually be idle or servicing a job of class  $i$ ,  $i > k$ .

The concept of completion time provides an understanding of the inaccuracy encountered in heuristic methods which use a reduced work rate approximation. These heuristics employ the observation that in a preemptive priority queue, the service of lower priority customers appears to be provided by a server whose average capacity for work is reduced due to the servic-

ing of higher priority customers. If  $\rho_k$  is the class  $k$  utilization at the priority center, these approximations replace the original preemptive priority queue with dedicated servers for each class, each server working with a reduced mean service rate of  $g_k^{-1}(1 - \sum_{j=1}^{k-1} \rho_j)$ .

This reduced mean service rate is identical to the reciprocal of class  $k$  mean completion time and fails to take into account those periods of time an arriving class  $k$  job finds the server occupied with a higher priority job, thus overestimating the reduced mean service rate. Moreover, these heuristics neglect the fact that the variance of the modified service time is not the same as the variance of the original service time distribution.

Following Jaiswal, let  $\Gamma_k$  denote the random variable "repair time for class  $k+1$  jobs" which of course, is also equal to the busy period in a preemptive priority queue with classes  $1, 2, \dots, k$ . Denote by:

$\bar{g}_k(s)$  the Laplace transform of the probability density function for  $G_k$ .

$\bar{\gamma}_k(s)$  the Laplace transform of the probability density function for  $\Gamma_k$ .

$\bar{B}_k(s)$  the Laplace transform of the busy period density function for a single class queue with service time distribution  $G_k$ .

$\bar{c}_k(s)$  the Laplace transform of the probability density function for  $C_k$ , the completion time for class  $k$ .

$\bar{\beta}_k(s)$  the Laplace transform of the busy period density function for a single class queue with service time distribution  $C_k$ .

The moments of the busy period of a preemptive priority queue with  $k$  classes may be obtained by means of the following computational scheme which is adapted from Jaiswal [1968]:

The Laplace transform of the pdf for  $\Gamma_1$  may be written as

$$\bar{\gamma}_1(s) = \bar{g}_1(\lambda_1\{1 - \bar{\gamma}_1(s)\} + s).$$

Although it may not be possible to obtain an analytic expression for  $\bar{\gamma}_1(s)$ , we can compute the moments of  $\Gamma_1$  by using the relation  $E[\Gamma_1^i] = (-1)^i \bar{\gamma}_1^{(i)}(0)$  where  $\bar{\gamma}_1^{(i)}(0)$  denotes the  $i$ th derivative of  $\bar{\gamma}_1(s)$  with respect to  $s$ , evaluated at  $s=0$ . The first three moments (to which reference is made in section 2) are thus given by

$$\Gamma_1 = E[\Gamma_1] = g_1/(1 - \lambda_1 g_1) \quad \text{and}$$

$$E[\Gamma_1^2] = E[G_1^2]/(1 - \lambda_1 g_1)^3$$

$$E[\Gamma_1^3] = \frac{E[G_1^3]}{(1-\lambda_1 g_1)^4} + \frac{3\lambda_1\{E[G_1^2]\}^2}{(1-\lambda_1 g_1)^5}$$

In order to compute the busy period in a preemptive priority queue with 2 classes, we must first determine the completion time for class 2 jobs. Jaiswal shows that the Laplace transform of the pdf for class 2 completion time is simply the Laplace transform of the pdf for  $G_2$  with parameter  $(\lambda_1\{1-\bar{\gamma}_1(s)\}+s)$  i.e.

$$\bar{c}_2(s) = \bar{g}_2(\lambda_1\{1-\bar{\gamma}_1(s)\}+s).$$

Here also, we may compute as many moments as desired by differentiation of  $\bar{c}_2(s)$ . The first three moments are

$$c_2 = E[C_2] = g_2 \cdot (1 + \lambda_1 \gamma_1) \quad \text{and}$$

$$E[C_2^2] = E[G_2^2] \cdot (1 + \lambda_1 \gamma_1)^2 + \lambda_1 E[\Gamma_1^2] \cdot g_2.$$

$$E[C_2^3] = E[G_2^3] (1 + \lambda_1 \gamma_1)^3 +$$

$$3E[G_2^2] (1 + \lambda_1 \gamma_1) \lambda_1 E[\Gamma_1^2] + g_2 \lambda_1 E[\Gamma_1^3]$$

Once we know  $\bar{c}_2(s)$  we can write an expression for the Laplace transform of the pdf for  $\Gamma_2$

$$\bar{\gamma}_2(s) = \frac{\lambda_2}{\lambda_1 + \lambda_2} \bar{\beta}_2(s) + \frac{\lambda_1}{\lambda_1 + \lambda_2} \bar{\gamma}_1(\lambda_2\{1-\bar{\beta}_2(s)\}+s).$$

Some insight into this formulation may be gained by considering  $\bar{\beta}_2(s)$  as the busy period given that the first arrival to an idle server is a class 2 job and  $\bar{\gamma}_1(\lambda_2\{1-\bar{\beta}_2(s)\}+s)$  as a busy period given that the idle period is terminated by a class 1 job so that the arriving class 2 job is delayed. The first three moments of  $\Gamma_2$  may be expressed as

$$\gamma_2 \equiv E[\Gamma_2] = \frac{\lambda_2}{\lambda_1 + \lambda_2} \cdot \frac{c_2}{1 - \lambda_2 c_2} + \frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot \frac{\gamma_1}{1 - \lambda_2 c_2}$$

$$E[\Gamma_2^2] = \frac{\lambda_1 \cdot E[G_1^2] + \lambda_2 E[G_2^2]}{(\lambda_1 + \lambda_2)(1 - \lambda_1 g_1 - \lambda_2 g_2)^3}$$

$$E[\Gamma_2^3] = \lambda_2 \Lambda_2^{-1} E[B_2^3] + \lambda_1 \Lambda_2^{-1} \gamma_1 \lambda_2 E[B_2^3]$$

$$+ 3\lambda_1 \lambda_2 \Lambda_2^{-1} E[\Gamma_1^2] E[B_2^2] (1 + \lambda_2 E[B_2])$$

$$+ \lambda_1 \Lambda_2^{-1} E[\Gamma_1^3] (1 + \lambda_2 E[B_2])^3$$

where

$$E[B_2] = \frac{c_2}{1 - \lambda_2 c_2}$$

$$E[B_2^2] = \frac{E[C_2^2] (1 + \lambda_2 c_2)^2}{1 - \lambda_2 c_2}$$

and

$$E[B_2^3] = \frac{E[C_2^3] (1 + \lambda_2 c_2)^2 + 3\lambda_2 E[C_2^2] E[B_2^2] (1 + \lambda_2 c_2)}{1 - \lambda_2 c_2}$$

Now suppose that we have obtained Laplace transforms for the pdf of  $\Gamma_{k-1}$  and  $C_{k-1}$ . To write the

Laplace transform of the pdf for  $\Gamma_k$ , we first determine the Laplace transform for the pdf of class k job completion times. This is

$$\bar{c}_k(s) = \bar{g}_k(\Lambda_{k-1}\{1-\bar{\gamma}_{k-1}(s)\}+s)$$

where  $\Lambda_k = \sum_{i=1}^k \lambda_i$ .

We may then write

$$\bar{\gamma}_k(s) = \frac{\lambda_k}{\Lambda_k} \bar{\beta}_k(s) + \frac{\Lambda_{k-1}}{\Lambda_k} \bar{\gamma}_{k-1}(\lambda_k\{1-\bar{\beta}_k(s)\}+s)$$

and the desired moments may be obtained by differentiation.

## 2. Modeling the Busy Period Distribution

In general, we will not be able to invert the Laplace transform of the k-class busy period distribution but we can compute as many of its moments as desired. We must decide on the number of moments to compute and on the form of the model to construct from these moments.

To explore this area, four standard models, each with three parameters or less, were fitted to the first three moments ( $\eta_1, \eta_1, \eta_3$ ) of the k-class busy period distribution and the performance of these models under varying conditions was studied. Figure 1 illustrates the models. The simplest model is that of an exponential distribution with parameter  $\mu = \eta_1$ . A two-phase Coxian distribution was used for two of the models. In model 2, the parameters are determined by the computational formulae:

$$\mu_1 = 2/\eta_1$$

$$\mu_2 = \eta_1/(\eta_2 - \eta_1^2)$$

$$\alpha = \eta_1^2/(2\eta_2 - \eta_1^2)$$

Marie [1978]

Model 3 matches three busy period moments using the formulae

$$a = 2\eta_1\eta_3 - 3\eta_2^2$$

$$b = 6\eta_1\eta_2 - 2\eta_3$$

$$c = 6\eta_2 - 12\eta_1^2$$

$$\mu_1 = \frac{-b + (b^2 - 4ac)^{1/2}}{2a}$$

$$\mu_2 = \frac{-b - (b^2 - 4ac)^{1/2}}{2a}$$

$$\alpha = \frac{\mu_2^2(\eta_2 - 2\eta_1^2)}{(2 - 2\mu_2\eta_1)}$$

for  $\eta_2/\eta_1 \neq 1$ . Model 4 is Sauer's hypergeometric model which forces equal flow through each phase. The parameters for Sauer's model are determined as follows:

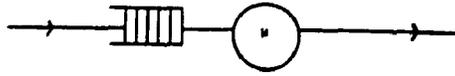
$$cv^2 = \eta_2/\eta_1^2 - 1$$

$$\alpha = \frac{cv^2 + 1 - [(cv^2)^2 - 1]^{1/2}}{2(cv^2 + 1)}$$

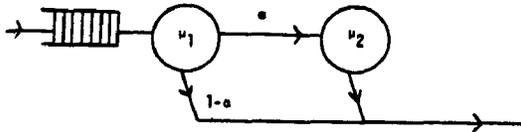
$$\mu_1 = 2/\eta_1$$

$$\mu_2 = 2(1-\alpha)/\eta_1$$

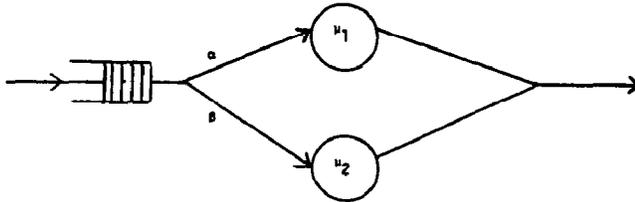
for  $cv^2 \geq 1$ .



Model 1: Exponential



Models 2 and 3: Coxian-2



Model 4: Sauer

Figure 1. "Models used to represent the busy period".

Since the machine interference model produces the exact solution for the highest priority class, the class 1 equilibrium distribution was not investigated with these models. In all cases, the mean queue length,  $L_j$ , was computed using the formula

$$L_j = \frac{\rho_j}{1 - \sum_{i=1}^{j-1} \rho_i} + \frac{\lambda_j \sum_{i=1}^j \lambda_i E[G_i^2]}{2(1 - \sum_{i=1}^{j-1} \rho_i)(1 - \sum_{i=1}^j \rho_i)}$$

Jaiswal [1968]

When the service time distribution functions of class 1 and class 2 jobs are exponential, the equilibrium probability of an idle server,  $X(0,0)$  and the equilibrium probability of finding no class 2 jobs at the

server,  $\underline{X}_2(0)$ , were also computed for each of the models since the exact values may be obtained from the formulae:

$$X(0,0) = 1 - \lambda_1 g_1 - \lambda_2 g_2$$

$$\underline{X}_2(0) = X(0,0)/(1-r_0)$$

where

$$r_0 = \frac{\lambda_1 + \lambda_2 + g_1^{-1} - [(\lambda_1 + \lambda_2 + g_1^{-1})^2 - 4\lambda_1 g_1^{-1}]^{1/2}}{2g_1^{-1}}$$

The results for a 2 class preemptive priority system in which the service time distribution functions of both classes follow an exponential distribution appear in Tables 1 and 2. All models produce exceedingly accurate values for  $X(0,0)$ . When the busy period coefficient of variation exceeds 2, Marie's Coxian 2 model provides the best approximation to  $\underline{X}_2(0)$ ; when the busy period coefficient of variation is less than 2 and greater than 1, Sauer's model and the three moment model approximate  $\underline{X}_2(0)$  somewhat more accurately, but Marie's model is good in this range also with a maximum error of less than 0.0015. The mean queue length calculations for the approximate models were set to terminate if either a class 2 equilibrium probability of less than  $10^{-5}$  was encountered or if the number of class 2 jobs reached 99. Marie's model performed best with respect to mean queue length for all these examples.

The results for a three class preemptive priority system in which the service time distributions of all three classes follow Coxian distributions appear in Tables 4, 5 and 6. While these results are less decisive than those for the 2 class system with exponentially distributed service times, Marie's Coxian 2 model produces more accurate mean queue length approximations than the other models most of the time. Marie's model is always best when the busy period coefficient of variation exceeds 10; when the other models are superior, the mean queue length computed from Marie's model differs from the best approximation by less than  $5 \times 10^{-4}$ . Under all conditions, the accuracy of the exponential model was poor compared with the models fitting at least 2 busy period moments. Although implementation of the numerical method proposed in this paper requires close to twice as many floating point operations for two phase models than for the exponential model, the extra computational effort is warranted by the increase in accuracy. A surprising fact which emerged from this study was that matching three busy period moments does not increase the accuracy of the approximation; consequently, Marie's two phase model is recommended.

A discussion of the tradeoff between accuracy and computational effort would be incomplete without mention of Miller's algorithm for the solution of two class preemptive priority queues with exponential service time distributions. In order to compare the various models, the approximate methods were set to ter-

minate if either a class 2 equilibrium probability of less than  $10^{-5}$  was encountered or if the number of class 2 jobs reached 99. The implementation of Miller's algorithm allowed 100 class 1 jobs and 100 class 2 jobs. The slowest approximate method required at most .03 seconds of cpu time while Miller's algorithm required at least .48 seconds of cpu time. In most cases the return for the substantial additional cpu time was only a slight improvement (.001) in the class 2 mean queue length accuracy (see Table 2) although  $\underline{X}_2(0)$  as determined by Miller's algorithm was exact for all the applicable examples.

### 3. Case of Coxian-2 Distributions

In this section, we shall assume that all service time and busy period distributions follow a law of Cox of order 2. In such circumstances, the infinitesimal generators are particularly easy to construct. In section 4 we shall consider the case when the service time distributions and/or the busy period distributions are modeled using more general Coxian distributions.

We introduce the following parameters:

- (i)  $\lambda_k$  is the arrival rate of class k jobs.
- (ii)  $m_{k1} = 2g_k^{-1}$  is the mean service rate of class k jobs at phase 1,

$m_{k2} = g_k/s_k^2$  is the mean service rate of class k jobs at phase 2,

$a_k = g_k^2/2s_k^2$  is the probability that a class k job which has completed service at phase 1 will require service at phase 2. Let  $b_k = 1-a_k$ .

- (iii)  $\mu_{k1} = 2\gamma_{k-1}^{-1}$  is the mean repair rate for class k jobs at phase 1,

$\mu_{k2} = \gamma_{k-1}/\sigma_{k-1}^2$  is the mean repair rate for class k jobs at phase 2,

$\alpha_k = \gamma_{k-1}^2/2\sigma_{k-1}^2$  is the probability that a machine serving class k jobs requires a phase 2 repair upon completion of a phase 1 repair. Let  $\beta_k = 1-\alpha_k$ .

If a service is interrupted when it is in phase 1,  $l=1,2$ , it resumes from that phase.

As indicated above, the solution for class 1 jobs, those of the highest priority, is trivial. Under a preemptive priority service discipline, this queueing situation is simply a  $\lambda_1/C_2/1$  system. Marie and Pel-laumail, [1983]; Neuts, [1981] and Carroll et al. [1982] all provide a numerical algorithm from which we can obtain the equilibrium probability distribution for queues of this type.

Now suppose we desire the equilibrium probability distribution of class k jobs,  $k > 1$ . We shall consider a machine breakdown and repair model with one machine. The arrival rate of class k jobs to this system is  $\lambda_k$  while the service time follows a Coxian distribution with parameters  $m_{k1}, m_{k2}, a_k, b_k$ . The breakdown rate is  $\sum_{i=1}^{k-1} \lambda_i$ . Upon occurrence of a breakdown, the down time follows a Coxian-2 distribution with parameters  $\mu_{k1}, \mu_{k2}, \alpha_k, \beta_k$ .

Let us define a state of this system as  $(n_k, j, z)$  where

$n_k$  is the number of class k jobs in the system.

$j$  is the phase of the current service for class k jobs or the phase from which an interrupted class k service will resume.

$z$  is the state of the machine and takes on the values:

0, if the machine is operating.

1,  $l = 1, 2$ , if the machine is down and repair work is in phase 1.

Let  $x_k(n_k, j, z)$  be the probability of finding the system in state  $(n_k, j, z)$  at equilibrium and let

$$\underline{X}_k(n) = \begin{bmatrix} x_k(n, 1, 0) \\ x_k(n, 1, 1) \\ x_k(n, 1, 2) \\ x_k(n, 2, 0) \\ x_k(n, 2, 1) \\ x_k(n, 2, 2) \end{bmatrix}$$

In the representation of the matrices which follow, zero-valued elements and subblocks are mostly omitted.

For class k jobs the infinitesimal generator  $\underline{Q}_k$  is given by

$$\underline{Q}_k = \begin{bmatrix} \underline{B}_{k0} & \underline{A}_{k0} & & & & \\ \underline{A}_{k2} & \underline{A}_{k1} & \underline{A}_{k0} & & & \\ 0 & \underline{A}_{k2} & \underline{A}_{k1} & \cdots & & \\ 0 & & \underline{A}_{k2} & \cdots & & \end{bmatrix}$$

where  $\underline{A}_{k0} = \lambda_k \underline{I}_{(6 \times 6)}$



- I.  $\lambda_k$  is the arrival rate of class k jobs to the queue.
- II. (a)  $l_k$  is the number of phases in the representation of the class k service time distribution.
- (b)  $d_k(j)$  is the departure rate of class k jobs from the system when the current service is in phase j.
- (c)  $t_k(i,j)$  is the transition rate of class k jobs from service phase i to service phase j. Note that there is no restriction that  $j > i$ .
- (d)  $w_k(j)$  is the probability that following the departure of a class k job, the next class k job begins service in phase j. If the system is empty,  $w_k(j)$  is the probability that a class k arrival to an empty system begins in phase j.
- III. (a)  $\xi_k$  is the number of phases in the representation of the class k repair time (k-1 class busy period) distribution.
- (b)  $\delta_k(j)$  is the departure rate from repair phase j for class k jobs (departure rate from busy period phase j for a preemptive priority queue with k-1 classes).
- (c)  $\tau_k(i,j)$  is the transition rate for machines serving class k jobs from repair phase i to repair phase k.
- (d)  $\omega_k(j)$  is the probability that when a machine servicing class k jobs breaks down the first repair phase is j.

$$\underline{X}_k(n) = \begin{bmatrix} x_k(n,1,0) \\ x_k(n,1,1) \\ \vdots \\ x_k(n,1,\xi_k) \\ x_k(n,2,0) \\ \vdots \\ x_k(n,2,\xi_k) \\ \vdots \\ x_k(n,l_k,0) \\ \vdots \\ x_k(n,l_k,\xi_k) \end{bmatrix}$$

Each block in the infinitesimal generator  $\underline{Q}_k$  is of order  $(\xi_k+1)l_k$ . We have the usual block tridiagonal form for  $\underline{Q}_k$ :

$$\underline{Q}_k = \begin{bmatrix} \underline{B}_{k0} & \underline{A}_{k0} & & & \\ \underline{B}_{k1} & \underline{A}_{k1} & \underline{A}_{k0} & & \\ & \underline{A}_{k2} & \underline{A}_{k1} & \cdots & \\ & & \vdots & & \\ & & \vdots & & \\ & & \vdots & & \end{bmatrix}$$

The subblocks are as follows:

$$\underline{B}_{k0} = \begin{bmatrix} (\underline{B}_{k0})_{11} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & & & \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{0} \end{bmatrix}$$

where

$$(\underline{B}_{k0})_{11} = \begin{bmatrix} -\Lambda_k & \Lambda_{k-1}\omega_k(1) & \cdots & \Lambda_{k-1}\omega_k(\xi_k) \\ \delta_k(1) & * & \vdots & \tau_k(1,\xi_k) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \delta_k(\xi_k) & \tau_k(\xi_k,1) & \cdots & *\xi_k \end{bmatrix}$$

and

$$*_{m} = -\{\lambda_k + \sum_{\substack{j=1 \\ j \neq m}}^{\xi_k} \tau_k(m,j) + \delta_k(m)\} ;$$

If service is interrupted when it is in phase j,  $j=1,2,\dots, l_k$ , it resumes from phase j.

We again observe that obtaining the equilibrium probability distribution for class 1 jobs is straightforward; Marie and Pellaumail [1983] provide the numerical solution for the general model proposed above.

For an arbitrary class k,  $k > 1$ , the state description will again be  $(n_k, j, z)$ ; j now takes on values  $1, 2, \dots, l_k$  while z has values  $0, 1, 2, \dots, \xi_k$ . the steady-state probability distribution vector is given by:

$$\underline{B}_{k1} = \begin{bmatrix} (\underline{B}_{k1})_{11} & \underline{0} & \cdots & \underline{0} \\ (\underline{B}_{k1})_{21} & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ (\underline{B}_{k1})_{l_k 1} & \underline{0} & \cdots & \underline{0} \end{bmatrix}$$

where the elements of  $(\underline{B}_{k1})_{ij} \in \mathbb{R}^{(\xi_k+1) \times (\xi_k+1)}$  are given by

$$((\underline{B}_{k1})_{j1})_{mn} = \begin{cases} d_k(j) & \text{if } m=n=0 \\ 0 & \text{otherwise;} \end{cases}$$

(Here we use  $(\underline{B})_{ij}$  to denote the  $i$ -th element (or block) of the matrix  $\underline{B}$ , so that  $((\underline{B}_{k1})_{j1})_{mn}$  denotes the  $mn$ -th element of the  $j$ l-th block of  $\underline{B}_{k1}$ .)

$$\underline{A}_{k0} = \begin{bmatrix} w_k(1)\lambda_k \underline{I} & w_k(2)\lambda_k \underline{I} & \cdots & w_k(l_k)\lambda_k \underline{I} \\ \underline{0} & \underline{0} & & \\ \underline{0} & \cdot & \cdots & \cdot \\ & \underline{0} & & \underline{0} \end{bmatrix}$$

$$\underline{A}_{k0} = \begin{bmatrix} \lambda_k \underline{I} & & & \\ & \lambda_k \underline{I} & & \\ & & \cdot & \\ & & & \lambda_k \underline{I} \end{bmatrix}$$

$$\underline{A}_{k1} = \begin{bmatrix} (\underline{A}_{k1})_{11} & (\underline{A}_{k1})_{12} & \cdots & (\underline{A}_{k1})_{1l_k} \\ (\underline{A}_{k1})_{21} & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ (\underline{A}_{k1})_{l_k 1} & (\underline{A}_{k1})_{l_k 2} & \cdots & (\underline{A}_{k1})_{l_k l_k} \end{bmatrix}$$

Each subblock is again of order  $(\xi_k+1)$ . The diagonal blocks have the form

$$(\underline{A}_{k1})_{rr} = \begin{bmatrix} *_{rr} & \Lambda_{k-1}\omega_k(1) & \cdots & \Lambda_{k-1}\omega_k(\xi_k) \\ \delta_k(1) & *_{1r} & \cdots & \tau_k(1, \xi_k) \\ \cdot & \cdot & & \cdot \\ \delta_k(\xi_k) & \tau_k(\xi_k, 1) & \cdots & *_{\xi_k r} \end{bmatrix}$$

$$\text{with } *_{rr} = -\{\Lambda_k + d_k(r) + \sum_{\substack{j=1 \\ j \neq m}}^{l_k} \tau_k(r, j)\}$$

$$\text{and } *_{mr} = -\{\lambda_k + \delta_k(m) + \sum_{\substack{j=1 \\ j \neq m}}^{\xi_k} \tau_k(m, j)\}$$

The off-diagonal blocks have elements

$$((\underline{A}_{k1})_{ij})_{mn} = \begin{cases} \tau_k(ij) & \text{if } m=n=0 \\ 0 & \text{otherwise;} \end{cases}$$

$$\underline{A}_{k2} = \begin{bmatrix} (\underline{A}_{k2})_{11} & (\underline{A}_{k2})_{12} & \cdots & (\underline{A}_{k2})_{1l_k} \\ (\underline{A}_{k2})_{21} & (\underline{A}_{k2})_{22} & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ (\underline{A}_{k2})_{l_k 1} & (\underline{A}_{k2})_{l_k 2} & \cdots & (\underline{A}_{k2})_{l_k l_k} \end{bmatrix}$$

Each subblock  $(\underline{A}_{k2})_{ij}$  is of order  $(\xi_k+1)$  and has elements

$$((\underline{A}_{k2})_{ij})_{mn} = \begin{cases} w_k(i)d_k(i) & \text{if } m=n=0 \\ 0 & \text{otherwise;} \end{cases}$$

By construction  $\underline{A}_{k2}$  is a rank 1 matrix. We again refer to the results in Snyder and Stewart [1985] which assures us that we may perform a reduction procedure on the Chapman-Kolmogorov equations relating flow into and out of model states at equilibrium. Implementation of the reduction procedure is straightforward. Following the steps detailed in section 1 we may write the equilibrium probability distribution of class  $k$  jobs explicitly as

$$\underline{x}_k(1)^T = \underline{x}(0)^T \underline{A}_{k0} [-\underline{A}_{k1} - \underline{\lambda}_k \underline{w}_k^T]^{-1}$$

$$\underline{x}_k(n)^T = \underline{x}(n-1)^T \underline{R}_k, \quad n > 1$$

where

$$\underline{R}_k = \underline{A}_{k0} [-\underline{A}_{k1} - \underline{\lambda}_k \underline{w}_k^T]^{-1}$$

and  $\underline{\lambda}_k$  and  $\underline{w}_k$  are column vectors each with  $(\xi_k+1)l_k$  elements

$$\underline{\lambda}_k = [\lambda_k, \lambda_k, \dots, \lambda_k]^T$$

$$\underline{w}_k = \underbrace{[w_k(1), 0, \dots, 0, w_k(2), 0, \dots, 0, \dots, w_k(l_k), 0, \dots, 0]}_{\xi_k}^T$$

We again note that the  $\underline{x}_n$  must satisfy the normalizing equation. The final step in the computation of the equilibrium probability distribution of class  $k$  jobs in a  $k$  class preemptive priority queueing system is to sum over all states  $x_k(n, j, z)$  for each  $n=0, 1, \dots, N_k$ , i.e.

$$\underline{x}_k(n) = \sum_{j=1}^{l_k} \sum_{z=0}^{\xi_k} x_k(n, j, z)$$

## 5. Conclusion and Extensions

In this paper an explicit approximate numerical solution for a multiclass preemptive priority queue with generally distributed service times was developed. The solution is based on earlier results by the authors, [1985]: specifically, if the matrix specifying transitions due to customer departures is of rank 1, then the Chapman-Kolmogorov equations, (which are normally second-order difference equations) may be reduced to a set of first order difference equations and solved recursively to obtain an explicit solution. This solution may be written immediately in terms of the model parameters.

The validity of the approximation was tested on several models constructed from the first three moments of the busy period. Over a wide range of test values, the Coxian 2 model of Marie performed best in terms of accuracy and computational efficiency. The simplest case, with service and busy period distributions modeled by such a law of Cox, was presented in detail to illustrate the solution technique. The model was then generalized to allow arbitrary service and busy period distributions.

It is our expectation that this approximation will be used in an iterative method to determine equilibrium probability distributions of otherwise product form networks which contain preemptive priority stations. We believe that this approximation may also be used with Marie's iterative technique, Marie [1979] to solve an otherwise nonproduct form network which contains a preemptive priority queue.

Table 3

cv <sup>2</sup> of busy period	$\rho_1^{[1]}$	$\rho_2^{[1]}$ solution	exact	model 1	model 2	model 3	model 4
7	.75	.10	0.200000	0.214286	0.207558	0.208290	0.208290
4	.60	.10	0.391548	0.412500	0.396920	0.399351	0.399351
2.33	.40	.10	0.625000	0.642857	0.624324	0.627753	0.627753
2.33	.40	.40	0.250000	0.257143	0.249730	0.251101	0.251101
1.858	.30	.50	0.242070	0.245154	0.241575	0.242385	0.242385
1.858	.30	.30	0.484139	0.492308	0.483150	0.487700	0.484770
1.858	.30	.1	0.725209	0.738462	0.724724	0.727155	0.727155
1.22	.10	.10	0.871563	0.872727	0.871456	0.871568	0.871568
1.22	.10	.30	0.653673	0.654545	0.653592	0.653676	0.653676
1.22	.10	.50	0.435782	0.436363	0.435728	0.435784	0.435788

$$[1]\rho_1 = \lambda_1 E[G_1]$$

Probability that there are zero class 2 jobs in system

Example Number	Service Time CV			Busy Period CV	
	Class 1	Class 2	Class 3	1 Class	2 Class
1	4.0	4.0	4.0	5.667	13.455
2	2.0	2.0	2.0	3.000	7.673
3	10.0	10.0	10.0	13.667	30.904
4	2.0	10.0	10.0	3.000	20.642
5	2.0	10.0	2.0	3.000	20.642
6	10.0	2.0	10.0	13.667	17.840
7	10.0	2.0	2.0	13.667	17.840
8	0.5	0.5	0.5	1.000	3.336
9	0.5	1.0	0.5	1.000	4.148
10	1.0	0.5	1.0	1.667	3.970
11	0.5	2.0	2.0	1.000	5.771
12	0.5	2.0	0.5	1.000	5.771
13	0.5	2.0	10.0	1.000	5.771
14	2.0	0.5	2.0	3.000	5.238
15	2.0	0.5	0.5	3.000	5.238
16	2.0	0.5	10.0	3.000	5.238
17	1.0	1.0	1.0	1.667	5.782
18	1.0	2.0	2.0	1.667	6.405
19	1.0	2.0	0.5	1.667	6.405
20	1.0	2.0	10.0	1.667	6.405
21	2.0	1.0	2.0	3.000	6.405
22	2.0	1.0	0.5	3.000	6.045
23	2.0	1.0	10.0	3.000	6.045

3 class examples for isolated preemptive priority queue

$$\rho_1 = .25, \rho_2 = .4, \rho_3 = .1875$$

Table 2

CV of Busy Period	$\rho_1$	$\rho_2$	Exact Solution	Miller	Model 1	Model 2	Model 3	Model 4
0.7	0.75	0.10	30.8667	17.8951	8.1575	17.2809	16.6384	16.6304
4.0	0.60	0.10	6.3333	6.2210	2.7318	6.2942	6.2700	6.2700
2.33	0.40	0.10	1.2667	1.2667	0.8797	1.2854	1.2851	1.2851
2.33	0.40	0.40	4.8667	4.8667	3.5981	4.6629	4.8626	4.8626
1.86	0.30	0.10	0.5952	0.5952	0.4666	0.5949	0.5948	0.5948
1.86	0.30	0.30	1.3929	1.3928	1.1998	1.3923	1.3923	1.3923
1.86	0.30	0.50	3.7857	3.7857	3.3982	3.7836	3.7836	3.7836
1.22	0.10	0.10	1.1528	0.1528	0.1500	0.1528	0.1528	0.1528
1.22	0.10	0.30	0.5370	0.5370	0.5333	0.5370	0.5370	0.5370
1.22	0.10	0.50	1.3055	1.3055	1.2998	1.3053	1.3053	1.3053

$$\rho_1 = \lambda_1 E[G_1]$$

[2] The large queue length error in this example results from the decision to terminate computation of  $nP(n)$  at  $n=99$  jobs

Mean number of class 2 jobs in system for 2 class exponential examples

Table 4

Example no.	Model 1		Model 2		Model 3		Model 4	
	P2(0)	P3(0)	P2(0)	P3(0)	P2(0)	P3(0)	P2(0)	P3(0)
1	0.4200	0.3686	0.4029	0.3018	0.4031	0.3028	0.4043	0.3043
2	0.4200	0.3686	0.4083	0.3118	0.4089	0.3136	0.4098	0.3150
3	0.4200	0.3686	0.3981	0.2930	0.3981	0.2934	0.3990	0.2943
4	0.4200	0.3686	0.4083	0.2985	0.4089	0.3024	0.4098	0.2984
5	0.4200	0.3686	0.4083	0.2985	0.4089	0.3024	0.4098	0.2984
6	0.4200	0.3686	0.3981	0.2982	0.3981	0.3115	0.3990	0.3002
7	0.4200	0.3686	0.3981	0.2982	0.3981	0.3115	0.3990	0.3002
8	0.4200	0.3686	0.4200	0.3330			0.4200	0.3385
9	0.4200	0.3686	0.4200	0.3268			0.4200	0.3304
10	0.4200	0.3686	0.4145	0.3280	0.4154	0.3332	0.4154	0.3316
11	0.4200	0.3686	0.4200	0.3183			0.4200	0.3218
12	0.4200	0.3686	0.4200	0.3183			0.4200	0.3218
13	0.4200	0.3686	0.4200	0.3183			0.4200	0.3218
14	0.4200	0.3686	0.4083	0.3207	0.4089	0.3273	0.4098	0.3242
15	0.4200	0.3686	0.4083	0.3207	0.4089	0.3273	0.4098	0.3242
16	0.4200	0.3686	0.4083	0.3207	0.4089	0.3273	0.4098	0.3242
17	0.4200	0.3686	0.4145	0.3230	0.4154	0.3268	0.4154	0.3268
18	0.4200	0.3686	0.4145	0.3158	0.4154	0.3185	0.4154	0.3192
19	0.4200	0.3686	0.4145	0.3158	0.4154	0.3185	0.4154	0.3192
20	0.4200	0.3686	0.4145	0.3158	0.4154	0.3185	0.4154	0.3192
21	0.4200	0.3686	0.4083	0.3171	0.4089	0.3215	0.4098	0.3206
22	0.4200	0.3686	0.4083	0.3171	0.4089	0.3215	0.4098	0.3206
23	0.4200	0.3686	0.4083	0.3171	0.4089	0.3215	0.4098	0.3206

Equilibrium probability of no class 2 jobs at server, P2(0)  
Equilibrium probability of no class 3 jobs at server, P3(0)

Table 5

Example no.	Exact	Model 1	Model 2	Model 3	Model 4
1	3.2478	2.4127	3.2443	3.2442	3.2441
2	2.1819	1.8042	2.1808	2.1808	2.1807
3	6.5048	4.2356	6.4817	6.4815	6.4585
4	4.8000	4.2356	4.5919	4.5919	4.5918
5	4.6000	4.2356	4.5919	4.5919	4.5918
6	4.0668	1.8042	4.0558	4.0557	4.0556
7	4.0668	1.8042	4.0558	4.0557	4.0556
8	1.3476	1.3474	1.3474		1.3474
9	1.5000	1.4997	1.4997		1.4997
10	1.4668	1.3474	1.4663	1.4663	1.4663
11	1.8048	1.8042	1.8042		1.8042
12	1.8048	1.8042	1.8042		1.8042
13	1.8048	1.8042	1.8042		1.8042
14	1.7048	1.3474	1.7039	1.7040	1.7039
15	1.7048	1.3474	1.7039	1.7040	1.7039
16	1.7048	1.3474	1.7039	1.7040	1.7039
17	1.6181	1.4997	1.6188	1.6186	1.6186
18	1.9238	1.8042	1.9231	1.9230	1.9230
19	1.9238	1.8042	1.9231	1.9230	1.9230
20	1.9238	1.8042	1.9231	1.9230	1.9230
21	1.8571	1.4997	1.8563	1.8562	1.8561
22	1.8571	1.4997	1.8563	1.8562	1.8561
23	1.8571	1.4997	1.8563	1.8562	1.8561

Class 2 mean queue length of 3 class models

Table 6

Example no.	Exact	Model 1	Model 2	Model 3	Model 4
1	6.7789	2.7291	6.7682	6.7678	6.7678
2	4.2816	3.3236	4.2776	4.2777	4.2774
3	14.2706	4.5772	13.0882	13.0789	13.0612
4	10.9739	4.5772	10.6620	10.6241	10.6512
5	8.5014	2.1121	8.4267	8.4045	8.4205
6	10.0508	4.5772	9.8630	9.8000	9.8558
7	7.5783	2.1121	7.5490	7.5180	7.5459
8	2.4087	1.6490	2.4078		2.4075
9	2.6724	1.6490	2.6712		2.6711
10	2.7692	1.8035	2.7680	2.7677	2.7679
11	3.6635	2.1121	3.6608		3.6607
12	3.1997	1.6490	3.1978		3.1976
13	6.1360	4.5773	6.1245		6.1250
14	3.4904	2.3907	3.4878	3.4876	3.4877
15	3.0268	2.3907	3.0249	3.0249	3.0251
16	5.9629	2.3907	5.9524	5.9518	5.9521
17	3.0330	1.8035	3.0313	3.0312	3.0312
18	3.8695	2.1121	3.8664	3.8664	3.8663
19	3.4059	1.6490	3.4034	3.4033	3.4035
20	6.3420	4.5772	6.3301	6.3307	6.3306
21	3.7541	2.1121	3.7511	3.7513	3.7510
22	3.2905	1.6490	3.2885	3.2881	3.2883
23	6.2267	4.5772	6.2151	6.2155	6.2156

Class 3 mean queue length for 3 class examples

- [2] Carroll, J.L., A. van de Liefvoort and L. Lipsky 1982. "Solutions of M/G/1/N-type Loops with Extensions to M/G/1 and GI/M/1 Queues." *Opns. Res.* Vol. 30, pp. 490-514.
- [3] Gaver, D.P. 1962. "A Waiting Line with Interrupted Service, Including Priorities." *J. Royal Stat. Soc., Series B*, Vol. 24, pp.73-90.
- [4] Jaiswal, N. 1968. *Priority Queues*. Academic Press. New York and London.
- [5] Keilson, J, 1962. "Queues Subject to Service Interruption." *Ann. Math. Stat.* Vol. 33, pp. 1314-1322.
- [6] Kleinrock, L. 1975. *Queueing Systems, Vol. 1: Theory*. John Wiley and Sons, New York.
- [7] Marie, R.A. 1978. Modelisation par Reseaux de files d'attente, theses: Docteur-es-Sciences Mathematiques, Universite de Rennes, France, November.
- [8] Marie, R.A. 1979. "An Approximate Analytical Method for General Queueing Networks." *IEEE Trans. on Software Engr.*, Vol. SE-5, pp. 530-538.
- [9] Marie, R.A. and J.M. Pellaumail 1983. "Steady-State Probabilities for a Queue with a General Service Distribution and State Dependent Arrivals." *IEEE Trans. on Software Engr.* Vol. SE-9, pp. 109-113.
- [10] Miller, R. 1981. "Computation of Steady-State Probabilities for M/M/1 Priority queues." *Opns. Res.* Vol. 29, pp. 945-957.
- [11] Neuts, M.F. 1981. *Matrix Geometric Solutions in Stochastic Models-An Algorithmic Approach*. Johns Hopkins Univ. Press, Baltimore.
- [12] Snyder, P.M. and W.J. Stewart 1984. "Explicit and Iterative Numerical Approaches to Solving Queueing Networks." *Opns. Res.* Vol. 33, pp. 183-202.
- [13] White, H. and L.S. Christie 1958. "Queueing with preemptive priorities or with breakdown." *Opns. Res.* Vol. 6, pp. 79-95.

## References

- [1] Avi-Itzhak, B. and P. Naor 1963. "Some queueing Problems with Service Station Subject to Breakdown." *Opns. Res.* Vol. 11, pp. 303-320.