279

REPRINTED FROM

NAVAL RESEARCH LOGISTICS QUARTERLY

SEPTEMBER 1973
VOL. 20, NO. 3

OFFICE OF NAVAL RESEARCH

NAVSO P-1278

THE SINGLE SERVER QUEUE IN DISCRETE TIME-NUMERICAL ANALYSIS III

Marcel F. Neuts* and Eugene Klimko

Purdue University

ABSTRACT

This paper deals with the *stationary analysis* of the finite, single server queue in discrete time. The following *stationary* distributions and other quantities of practical interest are investigated: (1) the joint density of the queue length and the residual service time, (2) the queue length distribution and its mean, (3) the distribution of the residual service time and its mean, (4) the distribution and the expected value of the number of customers lost per unit of time due to saturation of the waiting capacity, (5) the distribution and the mean of the waiting time, (6) the asymptotic distribution of the queue length following departures.

The latter distribution is particularly noteworthy, in view of the substantial difference which exists, *in general*, between the distributions of the queue lengths at arbitrary points of time and those immediately following departures.

1. INTRODUCTION

This paper is a direct sequel to [2], to which we refer for a detailed definition and for the assumptions of the finite, discrete time queue. For easy reference, we only give a summary of the notation here.

NOTATION

- L_1 Maximum number of customers allowed in the system at any time. All excess customers are lost and do not return.
- L_2 Maximum duration of the service time of a single customer.
- r_j Probability that a service lasts for j units of time, $j=1, \ldots, L_2$. We assume without loss of generality that $r_{L_2} > 0$. Also $r_1 + \ldots + r_{L_2} = 1$.
- K Maximum number of arrivals during a unit of time. It is assumed that $K < L_1$.
- p_j Probability that j customers arrive during a unit of time, $j=0, 1, \ldots, K$. We assume without loss of generality that $p_0 > 0$. and $p_K > 0$. Also $p_0 + \ldots + p_K = 1$.
- X_n The number of customers in the system at time $n+$.
- Y_n The number of time units until the customer in service at time $n+$ completes service. We note that $0 \leqslant Y_n \leqslant L_2$ and that $Y_n = 0$ if and only if $X_n = 0$.

In [2], it was shown that the bivariate sequence $\{(X_n, Y_n), n \geqslant 0\}$ is an irreducible, aperiodic Markov chain with state space $\{(0, 0)\} \cup \{(1, 2, \ldots, L_1) \times (1, \ldots, L_2)\}$. Its transient behavior was discussed and investigated numerically in [2]. In this paper we first discuss the *stationary* joint distribution of the queue length X_n and the residual service time Y_n.

*The research of this author was supported by the National Science Foundation, Contract No. GP 28650

2. THE EQUATIONS FOR THE STATIONARY JOINT PROBABILITIES OF X_n AND Y_n.

We denote the stationary probabilities by $P(i, j)$ for $i=1, \ldots, L_1$ and $j=1, \ldots, L_2$; $P(0, 0)$ is the stationary probability that the queue is empty. The stationary joint density of X_n and Y_n is the unique solution to the following system of linear equations.

(1) a.
$$P(0, 0) = p_0[P(1, 1) + P(0,0)],$$

b.
$$P(i, j) = \sum_{\nu=0}^{i} p_{i-\nu} P(\nu, j+1) + r_j r_{L_2}^{-1} P(i, L_2),$$

for $1 \leq i \leq K$, $1 \leq j \leq L_2 - 1$.

c.
$$P(i, j) = \sum_{\nu=i-K}^{i} p_{i-\nu} P(\nu, j+1) + r_j r_{L_2}^{-1} P(i, L_2),$$

for $K+1 \leq i \leq L_1 - 1$, $1 \leq j \leq L_2 - 1$.

d.
$$P(L_1, j) = P(L_1, j+1) + \sum_{\nu=1}^{K} \left(1 - \sum_{k=0}^{\nu-1} p_k\right) P(L_1 - \nu, j+1) + r_j r_{L_2}^{-1} P(L_1, L_2),$$

for $1 \leq j \leq L_2 - 1$.

e.
$$P(i, L_2) = r_{L_2} \left\{ \frac{p_0 p_i}{1 - p_0} P(1, 1) + \sum_{\nu=0}^{i} p_{i-\nu+1} P(\nu, 1) \right\},$$

for $1 \leq i \leq K$.

f.
$$P(i, L_2) = r_{L_2} \sum_{\nu=i-K+1}^{i+1} p_{i-\nu+1} P(\nu, 1),$$

for $K+1 \leq i \leq L_1 - 1$.

g.
$$P(L_1, L_2) = r_{L_2} \sum_{\nu=1}^{K} \left(1 - \sum_{k=0}^{\nu-1} p_k\right) P(L_1 - \nu + 1, 1),$$

h.
$$P(0, 0) + \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} P(i, j) = 1.$$

The system (1) contains $L_1 L_2 + 1$ *independent* linear equations in $L_1 L_2 + 1$ unknowns. We shall show that its solution may be conveniently expressed in terms of the solution of a homogeneous system of

L_1 equations in L_1 unknowns. Moreover, the latter system has a particular structure which greatly simplifies its numerical solution.

We denote by $\underline{P}_j$ the L_1-tuple $[P(1, j), \ldots, P(L_1, j)]$ for $j=1, \ldots, L_2$. We also introduce the $L_1 \times L_1$ matrices A and B defined as follows:

$$A=\begin{pmatrix}
p_0 & p_1 & p_2 & \cdots & p_K & 0 & \cdots & 0 & 0 & 0 \\
0 & p_0 & p_1 & \cdots & p_{K-1} & p_K & \cdots & 0 & 0 & 0 \\
0 & 0 & p_0 & \cdots & p_{K-2} & p_{K-1} & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & \cdots & p_{K-3} & p_{K-2} & \cdots & 0 & 0 & 0 \\
\cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\
 & & & & & & & p_{K-2} & p_{K-1} & p_K \\
 & & & & & & & p_{K-3} & p_{K-2} & p_{K-1}+p_K \\
 & & & & & & & \cdot & \cdot & \cdot \\
 & & & & & & & \cdot & \cdot & \cdot \\
 & & & & & & & \cdot & \cdot & \cdot \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & p_1 & p_2 & 1-p_0-p_1-p_2 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & p_0 & p_1 & 1-p_0-p_1 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & p_0 & 1-p_0 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 1
\end{pmatrix}$$

$$B=\begin{pmatrix}
\frac{p_1}{1-p_0} & \frac{p_2}{1-p_0} & \frac{p_3}{1-p_0} & \cdots & \frac{p_k}{1-p_0} & 0 & \cdots & 0 & 0 & 0 \\
p_0 & p_1 & p_2 & \cdots & p_{K-1} & p_K & \cdots & 0 & 0 & 0 \\
0 & p_0 & p_1 & \cdots & p_{K-2} & p_{K-1} & \cdots & 0 & 0 & 0 \\
0 & 0 & p_0 & \cdots & p_{K-3} & p_{K-2} & \cdots & 0 & 0 & 0 \\
\cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\
 & & & & & & & p_{K-2} & p_{K-1} & p_K \\
 & & & & & & & p_{K-3} & p_{K-2} & p_{K-1}+p_K \\
 & & & & & & & \cdot & \cdot & \cdot \\
 & & & & & & & \cdot & \cdot & \cdot \\
 & & & & & & & \cdot & \cdot & \cdot \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & p_2 & p_3 & 1-p_0-\cdots-p_3 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & p_1 & p_2 & 1-p_0-p_1-p_2 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & p_0 & p_1 & 1-p_0-p_1 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & p_0 & 1-p_0
\end{pmatrix}$$

In terms of A and B, the equations $(1b-g)$ may be written as

(2) $$\underline{P}_j = \underline{P}_{j+1}A + r_j r_{L_2}^{-1}\underline{P}_{L_2}, \qquad 1 \leq j \leq L_2 - 1,$$

$$\underline{P}_{L_2} = r_{L_2}\underline{P}_1 B.$$

The latter system is equivalent to the equations

(3) $$\underline{P}_j = r_{L_2}^{-1}\underline{P}_{L_2} \sum_{\nu=j}^{L_2} r_\nu A^{\nu-j}, \qquad 1 \leq j \leq L_2 - 1,$$

$$\underline{P}_{L_2} = \underline{P}_{L_2}\left(\sum_{\nu=1}^{L_2} r_\nu A^{\nu-1} \right) B.$$

We now observe that both A and B are stochastic matrices, that A is upper triangular and that the matrix B has only one subdiagonal. We shall say, for brevity, that B is *nearly upper triangular.* Since $r_1 + \ldots + r_{L_2} = 1$, and A is an upper triangular stochastic matrix, the matrix $\sum_{\nu=1}^{L_2} r_\nu A^{\nu-1}$ is stochastic and upper triangular. The stochastic matrix B is irreducible, so that the matrix

(4) $$Q = \sum_{\nu=1}^{L_2} r_\nu A^{\nu-1} B,$$

is irreducible and stochastic. Finally it is easy to verify that Q is nearly upper triangular.

The vector $\underline{P}_{L_2}$ is therefore proportional to the vector of the stationary probabilities of the matrix Q. The nearly upper triangular form of the matrix Q makes the numerical computation of the vector $\underline{P}_{L_2}$—up to a positive multiplicative constant—particularly simple. The vector $\underline{P}_{L_2}$ is proportional to the vector $(t_1, t_2, \ldots, t_{L_1})$, whose components may be computed recursively as follows

(5) $$t_1 = 1,$$

$$t_2 = (1 - q_{11}) q_{21}^{-1},$$

$$t_k = q_{k,k-1}^{-1} \left[t_{k-1}(1 - q_{k-1,k-1}) - \sum_{\nu=1}^{k-2} t_\nu q_{\nu,k-1} \right], \qquad 3 \leq k \leq L_1.$$

It is easy to verify that none of the entries $q_{k,k-1}$, $2 \leq k \leq L_1$ vanish, so that by using the first equation in (2), the vectors $\underline{P}_j$, $j = 1, \ldots, L_2 - 1$, may be computed up to a common, positive multiplicative constant. Equation (1a) is then used to determine $P(0, 0)$ up to the same multiplicative constant. This constant may finally be computed using Equation (1h). The stationary joint density of the queue length and the residual service time is therefore determined.

3. THE STATIONARY DENSITY OF THE WAITING TIME

The support of the stationary density $\{w_j\}$ of the waiting time consists of the integers 0, 1, . . ., L_1L_2. Clearly $w_0=P(0, 0)$ and for $1 \leq j \leq L_1L_2$, the density may be written symbolically as the convolution polynomial

$$\{w_j\}=P(1,\cdot)+P(2,\cdot)*\{r_\nu\}+P(3,\cdot)*\{r_\nu\}^{(2)}+\ldots+P(L_1,\cdot)*\{r_\nu\}^{(L_1-1)}, \tag{6}$$

where $\{r_\nu\}$ is the density of the service time.

The numerical computation of the w_j, $1 \leq j \leq L_1L_2$, by using a convolution analogue of *Horner's algorithm* for polynomials was discussed in [2].

4. THE STATIONARY DENSITY OF THE NUMBER OF LOST CUSTOMERS PER UNIT OF TIME

Since the waiting room is finite, it is possible that customers will be lost due to the waiting room being full at their arrival time. It is therefore of interest to know the stationary density $\{\varphi_j\}$ of the number of lost customers per unit of time. It has its support on the integers 0, 1, . . . K and may be determined by the explicit expressions

$$\varphi_j=\sum_{k=j}^{K} p_k \sum_{\nu=1}^{L_2} P(L_1-k+j,\nu), \qquad 1 \leq j \leq K, \tag{7}$$

$$\varphi_0=1-\sum_{j=1}^{K} \varphi_j.$$

Knowing the joint density discussed in section 2, the probabilities $\{\varphi_j\}$ are readily computed.

5. THE STATIONARY DENSITY OF THE QUEUE LENGTH AT DEPARTURES

The probabilities associated with the queue length at departure times, are primarily of interest in the analytic treatment of queues of $M|G|1$ type. Although they are frequently examined, their inherent applied interest is limited.

As we shall indicate below, the density of the queue length following departures may easily be obtained from auxiliary quantities which are computed in the process of evaluating the joint stationary density, discussed in section 2. In view of the importance ascribed to this density in the applied queueing literature, we decided to investigate its computational aspects. Note the very substantial difference which may exist between it and the stationary density of X_n.

The queue lengths following departures form an irreducible, aperiodic Markov chain with state space $\{0, 1, \ldots, L_1-1\}$. Let us denote its transition probability matrix by T. Furthermore, let $\theta_k(i,\nu)$ be the probability that in k consecutive units of time during which no departures occur, ν customers join the queue, given that the queue length at the beginning of the first unit of time was i.

The entries of T are then given by

$$T_{0j}=\sum_{k=1}^{L_2} r_k \sum_{h=1}^{K} p_h(1-p_0)^{-1}\,\theta_k(h,j-h+1), \qquad \text{for } 0 \leq j \leq L_1-1, \tag{8}$$

$$T_{ij}=\sum_{k=1}^{L_2} r_k\theta_k(i,j-i+1), \qquad \text{for } 1\leqslant i\leqslant j+1,$$

$$T_{ij}=0, \qquad \text{for } i>j+1.$$

We note that the transition probability matrix T is *nearly upper triangular.* The stationary probabilities corresponding to T may be calculated by a simple recursion such as in Formula (5). In order to evaluate the entries of the matrix T, we first show that

$$\theta_k(i,j-i+1)=(A^k)_{i,j+1}, \qquad \text{for } 1\leqslant i\leqslant L_1,\ 0\leqslant j\leqslant L_1-1, \tag{9}$$

where A is the upper triangular matrix defined in section 2.

For $k=1$, we find that

$$\begin{aligned}\theta_1(i,j-i+1)&=p_{j-i+1}, && \text{for } 0\leqslant j-i+1\leqslant K,\ j\leqslant L_1-2,\\ &=\sum_{\nu=L_1-i}^{K} p_\nu, && \text{for } L_1-K\leqslant i\leqslant L_1,\ j=L_1-1,\\ &=0, && \text{for all other pairs } (i,j),\end{aligned} \tag{10}$$

so that Equation (8) holds for $k=1$. Furthermore

$$\theta_{k+1}(i,j-i+1)=\sum_{\nu=\max(0,j-K)}^{j}\theta_k(i,\nu-i+1)p_{j-\nu}, \tag{11}$$

for $0\leqslant j\leqslant L_1-2$, and

$$\theta_{k+1}(i,L_1-i)=\sum_{\nu=L_1-K}^{L_1}\theta_k(i,\nu-i)\sum_{h=L_1-\nu}^{K}p_h,$$

for $1\leqslant i\leqslant L_1$. When expressed in terms of the matrix A, Formula (11) proves (9) inductively.

The matrix T can be compactly written as

$$T=C\sum_{k=1}^{L_2} r_kA^k, \tag{12}$$

where $C_{1j}=p_j(1-p_0)^{-1}$ for $1\leqslant j\leqslant K$; $C_{i-1,i}=1$, for $2\leqslant i\leqslant L_1$, and $C_{ij}=0$, for all other pairs (i,j).

The relation between the limiting distribution of the queue length following the nth departure and the stationary queue length distribution is noteworthy. A well-known theorem, from Reference [3], states that in a stable $M|G|1$ queue with single arrivals, the queue length at time t and the queue length following the nth departure have *the same limiting distribution* as t and n respectively tend to infinity.

An analogous result holds for the discrete time queue, discussed by Dafermos and Neuts [1], provided that the probability of two or more arrivals during a unit of time is zero. This result is proved by an exact analogue of the argument used for the $M|G|1$ queue, so we shall omit the proof.

In the case of group arrivals ($K \geq 2$) no simple relation exists between those two limiting distributions. Theorems which relate those distributions can be proved using the theory of Markov renewal processes, but the resulting formulas are not illuminating. We shall not pursue this topic here, but we offer as an illustration some numerical results for a queue which has rare arrivals of large groups of customers.

We considered a queue with $L_1 = 100$, $L_2 = 2$, $K = 20$, and with $p_0 = 0.975$, $p_{20} = 0.025$, $r_1 = r_2 = 0.5$. Although the traffic intensity ρ for the unbounded queue is 0.75, examination of the transient behavior shows that this queue converges *very slowly* to its stationary phase.

The limiting distribution of the queue length *following a departure* has a mean equal to 32.2864. In contrast, the limiting distribution of the queue length *at time n* has a mean equal to 24.1752. In addition, we list a summary of the numerical values of the two stationary distributions. π_k is the stationary probability of at most k customers at time n; π_k^* is the stationary probability of at most k customers, following a departure from the system (see Table 1).

TABLE 1

k	π_k	π_k^*
0	0.263	0.013
10	0.384	0.180
20	0.561	0.409
30	0.673	0.560
40	0.768	0.688
50	0.838	0.782
60	0.890	0.854
70	0.930	0.908
80	0.960	0.948
90	0.983	0.979

The greater limiting probability of longer queue lengths following departures may appear to be paradoxical at a casual reading. A moment's reflection shows however that, on the contrary, this is to be anticipated in stable queues with group arrivals. In our example, the queue length will typically be zero for long intervals of time because of the high value of p_0. The averaging procedure involved in the stationary distribution of the queue length at time n heavily favors the lower values of k. The limiting distribution of the queue length following the nth departure effectively ignores the long idle periods and results primarily from the behavior of the queue during the service of the large groups of customers. The high probabilities of the larger values of k in this distribution are therefore not surprising.

This example strikingly shows that the asymptotic distribution of the queue features may be of limited practical value, even in very stable queues. Most realizations of the queue length process in our example will exhibit very substantial fluctuations, which are not reflected in the asymptotic distributions. The practical questions related to queues of this type can only be answered after analyzing their *transient* behavior. The exclusive concern with asymptotic results in "practical" discussions of queueing theory is therefore regrettable.

6. COMPUTATIONAL ORGANIZATION

In order to minimize both the computation time and the required memory storage, we took advantage of the highly structured form of the matrices Q and T in Equations (4) and (11), respectively.

The basic matrix is the upper triangular polynomial matrix $Q^* = \{q^*_{ij}\}$

$$Q^* = \sum_{\nu=1}^{L_2} r_\nu A^{\nu-1}. \tag{13}$$

The rows of this matrix are similar in the sense that

$$q^*_{i,\,i+\nu} = q^*_{1,\,\nu+1} \tag{14}$$

$$\text{for } \nu = 0, 1, 2, \ldots L_1 - i - 1;\ i = 2, 3, \ldots, L_1 - 1.$$

Furthermore the matrix Q^* is stochastic, so that

$$q^*_{i,\,L_1} = 1 - \sum_{j=i}^{L_1-1} q^*_{ij}. \tag{15}$$

Therefore, the first row determines the entire matrix. This permits the storage of Q^* using only L_1 memory spaces, rather than the $(L_1^2 + L_1)/2$ spaces required for an arbitrary upper triangular matrix. The resulting saving in memory space is substantial for large queues and in fact makes the analysis of queue lengths up to 800 feasible. Computation of the matrix Q^* is performed by using Horner's method for the formation of polynomials, i.e., by recursive computation as follows

$$Q^*_1 = (r_{L_2} A + r_{L_2-1} I) \tag{16}$$

$$Q^*_n = Q^*_{n-1} A + r_{L_2-n} I \qquad n = 2, \ldots, L_2 - 1.$$

Each of the successive matrices Q^*_n is completely determined by its top row. The right-most elements are not needed and therefore are not computed. The top row entries of Q^*_n are rapidly calculated by means of the formulas

$$q^{*(n+1)}_{1j} = \sum_{i=0}^{\min(K,j)} p_i q^{*(n)}_{j-i} + r_{L_2-n} p_j, \quad \text{for } j \leq k \tag{17}$$

$$q^{*(n+1)}_{1j} = \sum_{i=0}^{\min(K,j)} p_i q^{*(n)}_{j-i}, \qquad \text{for } K < j \leq L_1.$$

The matrix Q has the form

$$\begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} & \cdots & q_{1,L_1-4} & q_{1,L_1-3} & q_{1,L_1-1} & q_{1L_1} \\ q_{21} & q_{22} & q_{23} & q_{24} & \cdots & q_{2,L_1-3} & q_{2,L_1-2} & q_{2,L_1-1} & q_{2L_1} \\ 0 & q_{21} & q_{22} & q_{23} & \cdots & q_{2,L_1-4} & q_{2,L_1-3} & q_{3,L_1-1} & q_{3,L_1} \\ 0 & 0 & q_{21} & q_{22} & \cdots & q_{2,L_1-5} & q_{2,L_1-4} & q_{4,L_1-1} & q_{4,L_1} \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdots & q_{21} & q_{22} & q_{L_1-2,L_1-1} & q_{L_1-2,L_1} \\ 0 & 0 & \cdot & \cdot & \cdots & 0 & q_{21} & q_{L_1-1,L_1-1} & q_{L_1-1,L_1} \\ 0 & 0 & \cdot & \cdot & \cdots & 0 & 0 & q_{L_1,L_1-1} & q_{L_1,L_1} \end{pmatrix}$$

where the third through the last rows, except for the last two columns, are essentially repetitions of the second row. The last column is determined by the condition that the rows sum to one. We therefore need to compute and store only the first and second rows and the (L_1-1)-st column. This requires $3L_1-4$ memory cells for the storage of the Q matrix rather than the $L_1+(L_1+2)(L_1-1)/2$ required for an arbitrary nearly upper triangular matrix. The top row elements of Q are given by

$$q_{1j}=\frac{p_j}{1-p_0}\,q_{11}^*+\sum_{i=0}^{\min(K,j-1)} p_i q_{1,j-i+1}^*, \qquad \text{for } j\le K \tag{18}$$

$$=\sum_{i=0}^{K} p_i q_{1,j-i+1}^*, \qquad \text{for } K<j<L_1-1$$

$$q_{1,L_1-1}=p_0\left(1-\sum_{j=1}^{L_1-1} q_{1j}^*\right)+\sum_{i=1}^{K} p_i q_{1,L_1-i}^*;$$

the second row elements of the Q matrix are

$$q_{2j}=\sum_{i=0}^{\min(K,j-1)} p_i q_{1,j-1}^*, \qquad \text{for } 1\le j\le L_1-2; \tag{19}$$

and the (L_1-1)-st column elements are calculated by using

$$q_{i,L_1-1}=p_0\left(1-\sum_{j=1}^{L_1-i} q_{1j}^*\right)+\sum_{k=1}^{\min(K,L_1-i)} p_k q_{1,K_1-k}^*, \tag{20}$$

for $2\le i\le L_1$.

The stationary probabilities of the Q matrix were determined using Formula (5) and its compact representation by Formulas (18)–(20). For this purpose, a subroutine called STAPROB was written. The resulting stationary probability vector was identified temporarily with the vector $\underline{P}_{L_2}$. The vectors

$\underline{P}_{L_2}, \ldots, \underline{P}_1$ were successively obtained by (2). Throughout this computation essentially only the top row of the matrix A is needed. The multiplication formula is

$$P(i,j) = \sum_{\nu=0}^{\min(K,j-1)} p_\nu P(i-\nu, j+1) + r_j r_{L_2}^{-1} P(i, L_2), \tag{21}$$

for $j = L_2 - 1, \ldots, 1$.

Finally $P(0,0)$ is computed and all $P(i,j)$ are adjusted so as to satisfy the normalization condition $(1h)$.

The waiting-time distribution was calculated according to (6) by a subroutine called WAIT. This subroutine was adapted from the program, discussed in [2]. In cases where L_1L_2 is large, one may wish to print only the percentage points of the waiting-time distribution. A routine to do this was also written.

The computational procedure for the queue length following a departure is similar to that for the stationary queue length distribution. The polynomial

$$\sum_{\nu=1}^{L_2} r_\nu A^\nu = AQ^* \tag{22}$$

is first computed and then the matrix T is determined. It is represented in a manner similar to that of the matrix Q. Only a modicum of additional computation is involved. The stationary distribution is then calculated by the subroutine STAPROB.

Testing

In addition to testing the program for its correctness, we compared the stationary probabilities with the transient probabilities after 60 time units. The latter were obtained by the methods developed in [2].

Computational Experience

Practical limits on the problem size are determined by the memory requirements. The available memory space of 150K octal required that $L_1L_2 \leq 20{,}000$, approximately. This permits, for instance, queue lengths of size 800 with service times concentrated on 25 points. For problems of this magnitude the computation time was a limiting factor only in the evaluation of the waitingtime distribution. We ran examples, both with and without the distribution of the waitingtime. The central processing times on the CDC 6500 at Purdue University for these examples are shown in Table 2. T_1 and T_2 are the actual program running times in seconds (without compilation and loading times), respectively, with and without the computation of the waitingtime distribution. For the example with $L_1 = 800$, $L_2 = 25$, $K = 4$, the time T_1 was in excess of *3,000* seconds and the computations were not completed even then. In all the examples, we used the same arrival distribution $p_0 = 0.8$, $p_1 = p_2 = p_3 = p_4 = 0.05$. The service time distribution for the first three examples was $r_1 = 0.675$, $r_2 = r_3 = r_4 = 0.05$, and $r_5 = 0.175$. In the last example, the service time distribution was a truncated geometric with $p = 0.5$ and the residual probability was added to r_{25}.

TABLE 2

L_1	L_2	K	T_1	T_2
100	5	4	5.751	0.945
200	5	4	22.539	2.221
400	5	4	69.774	6.612
800	25	4	> 3,071.032	26.290

7. CONCLUSIONS

Large discrete, single server queues in the stationary phase may be analyzed numerically. As we have shown, most queue features of interest, with the possible exception of the stationary waiting-time distribution, can be computed without the use of excessive processing times. This should be contrasted with simulation methods which are inherently ill-suited for the study of the stationary phase.

The prohibitive processing times required for the waiting-time distribution in large queues, raise the interesting question of how to evolve efficient numerical procedures for the evaluation of expressions of the general type

$$\sum_{i=1}^{n} A_i(\cdot) * F^{(i)}(\cdot)$$

which appear frequently in stochastic models of varied applied interest.

Finally, the example discussed in section 5, shows that in queues exhibiting large fluctuations, it may be hazardous to base conclusions on a single stationary distribution. In such cases one should study the transient behavior, whenever possible.

For further information on the algorithms discussed in this paper, one may contact either of the authors at the Department of Statistics, Purdue University, West Lafayette, Ind. 47907.

REFERENCES

[1] Dafermos, S. and M. F. Neuts, "A Single Server Queue in Discrete Time," Cahiers du Centre de Recherche Opérationnelle *13*, 23–40 (1971).

[2] Neuts, M. F., "The Single Server Queue in Discrete Time—Numerical Analysis I," Nav. Res. Log. Quarterly, *20*, 297–304 (1973).

[3] Takàcs, L., *Introduction to the Theory of Queues* (Oxford University Press, New York, 1962).