AD-A090 771 DELAWARE UNIV NEWARK APPLIED MATHEMATICS INST F/G 12/1
ALGORITHMIC SOLUTION OF SOME QUEUES WITH OVERFLOWS.(U)
APR 80 M F NEUTS, S KUMAR AFOSR-77-3236
UNCLASSIFIED TR-51B AFOSR-TR-80-0993 NL

END
DATE
FILMED
11-80
DTIC

AFOSR-TR- 80-0993

LEVEL II

5

SC

AD A090771

DDC FILE COPY

SIGILLVM VNIVERSITATIS DELAVARIENSIS 1833 SCIENTIA SOL MENTIS EST

GRAMM PHILOL RHETOR ETHICA METAPH LOGICA MATHEM PHYSICA

DTIC ELECTE OCT 24 1980 S D E

APPLIED MATHEMATICS INSTITUTE

University of Delaware
Newark, Delaware

80 10 6 105

LEVEL II (5)

ALGORITHMIC SOLUTION OF SOME QUEUES

WITH OVERFLOWS

by

Marcel F. Neuts

and

Seshavadhani Kumar

Department of Mathematical Sciences
University of Delaware
Newark, Delaware 19711

Applied Mathematics Institute
Technical Report # 51B
April 1980

This research was supported by the Air Force Office of Scientific Research, Air Force System Command, USAF, under Grant No. AFOSR-77-3236 and by the National Science Foundation, under Grant No. ENG-7908351.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

19

1. REPORT NUMBER: 18 AFOSR TR-80-0993
2. GOVT ACCESSION NO.: AD-A090771
3. RECIPIENT'S CATALOG NUMBER: 9

4. TITLE (and Subtitle): 6 ALGORITHMIC SOLUTION OF SOME QUEUES WITH OVERFLOWS

5. TYPE OF REPORT & PERIOD COVERED: Interim rept.,

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): 10 Marcel F./Neuts
Seshavidhani/Kumar

8. CONTRACT OR GRANT NUMBER(s): 15 AFOSR-77-3236, NSF-ENG79-08351

9. PERFORMING ORGANIZATION NAME AND ADDRESS
University of Delaware
Applied Mathematics Institute
Newark, DE 19711

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 17 A5
61102F 16 2304/A5

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Office of Scientific Research/NM
Bolling AFB, Washington, DC 20332

12. REPORT DATE: 11 Apr [redacted]80

13. NUMBER OF PAGES: 23 12 24

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 14 TR-51B

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Overflow in queues, matrix-geometric solution, phase type distributions, systems of queues, computational probability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The overflow stream from an M/PH/1 queue of finite capacity is used as the input to an unbounded queue with one or more exponential servers. It is shown that the combined system, consisting of the two queues, may be studied as a highly structured Markov process. In the stable case, this Markov process has a matrix-geometric invariant vector. Particular features of the infinitesimal generator of the process may be used to simplify the numerical computation of various steady-state features of the model. Several variants and numerical examples are discussed.

Abstract

The overflow stream from an M/PH/1 queue of finite capacity is used as the input to an unbounded queue with one or more exponential servers. It is shown that the combined system, consisting of the two queues, may be studied as a highly structured Markov process. In the stable case, this Markov process has a matrix-geometric invariant vector. Particular features of the infinitesimal generator of the process may be used to simplify the numerical computation of various steady-state features of the model. Several variants and numerical examples are discussed.

Key Words

Overflow in queues, matrix-geometric solution, phase type distributions, systems of queues, computational probability.

1. The Model

We study a service system with two units, each with a single server. Unit 1 is of finite capacity K. Unit 2 has as its input the stream of overflow customers from Unit 1, i.e. any customer who upon arrival to the system finds K customers (waiting or in service) in Unit 1 proceeds for service to Unit 2. The queue in Unit 2 is unbounded. Tractable variants of this model, involving multi-server units with exponential servers, are discussed in Sections 4 and 5.

The arrivals to the system are according to a Poisson process of rate λ. The successive service times in Unit 1 are independent and have a common distribution of phase type [6,11], given by the irreducible representation $(\underline{\beta}, S)$ of order m. Service times are assumed to be positive random variables, so that $\beta_{m+1} = 0$. Customers, who upon arrival find K customers in Unit 1, go to Unit 2. Service times in that unit are independent exponential random variables with parameter μ. The interarrival times and the successive service times in both units form independent sequences of random variables. For the present, we assume that there are no independent arrivals directly to Unit 2.

This queueing system may be viewed as an $M/M/1$ queue in a Markovian environment [8,9,11]. In order to see this, we first consider Unit 1 alone. It may be studied as a continuous-parameter Markov chain on the state space $E_1 = \{0\} \cup \{(i,j): 1 \leq i \leq K, 1 \leq j \leq m\}$. The index i denotes the number of customers in Unit 1 and the index j gives the phase of the current service. The state 0 corresponds to the case where Unit 1 is empty.

The infinitesimal generator Q of this Markov process is given by

$$
(1) \qquad Q = \begin{vmatrix}
-\lambda & \lambda\underline{\beta} & & & & \\
\underline{S}^\circ & S-\lambda I & \lambda I & & & \\
 & S^\circ B^\circ & S-\lambda I & \lambda I & & \\
 & & S^\circ B^\circ & S-\lambda I & \lambda I & \\
 & & & \cdots & & \\
 & & & S^\circ B^\circ & S-\lambda I & \lambda I \\
 & & & & S^\circ B^\circ & S
\end{vmatrix} ,
$$

where the matrix $S^\circ B^\circ$ is defined by $\underline{S}^\circ \cdot \underline{\beta}$. The states (i,j), $1 \le i \le K$, $1 \le j \le m$, are ordered in the lexicographic manner.

The process Q now clearly acts as an environmental process, which modulates the arrival rate to the second unit. Whenever the process Q is in one of its states (K,j), $1 \le j \le m$, there are arrivals to Unit 2 according to a Poisson process of rate λ. At all other times the arrival process to that unit is interrupted.

For future use, we note the following explicit formulas for the unique stationary probability vector $\underline{\pi}$ of Q, which satisfies the equations $\underline{\pi}\, Q = \underline{0}$, $\underline{\pi}\, \underline{e} = 1$.

<u>Lemma 1</u>

After partitioning into the form $[\pi_0, \underline{\pi}_1, \underline{\pi}_2, \ldots, \underline{\pi}_K]$, the vector $\underline{\pi}$ is given by

$$
\pi_0 = \left[1 + \underline{\beta} \sum_{\nu=1}^{K-1} C^{\nu} \underline{e} + \lambda\, \underline{\beta}\, C^{K-1}(-S^{-1})\underline{e} \right]^{-1} ,
$$

$$
(2) \qquad \underline{\pi}_i = \pi_0\, \underline{\beta}\, C^{i} , \qquad \text{for } 1 \le i \le K-1 ,
$$

$$
\underline{\pi}_K = \lambda\, \pi_0\, \underline{\beta}\, C^{K-1}(-S^{-1}) ,
$$

where

(3) $$C = \lambda(\lambda I - \lambda B^{\circ\circ} - S)^{-1} .$$

The matrix $B^{\circ\circ}$ is defined by $B^{\circ\circ} = \underline{e} \cdot \underline{\beta}$.

<u>Proof</u>

The vector $\underline{\pi}$ is the stationary probability vector for the (Km+1)-state Markov process which describes the bounded M/PH/1 queue. The simple proof of the lemma, which is given in [11], is repeated here for completeness.

The steady-state equations $\underline{\pi}\, Q = \underline{0}$, may be written as

$$
(4)\quad
\begin{aligned}
&-\lambda\, \pi_0 + \underline{\pi}_1\, \underline{S}^\circ = 0 , \\
&\lambda\, \pi_0\, \underline{\beta} + \underline{\pi}_1\, (S-\lambda I) + \underline{\pi}_2\, S^\circ B^\circ = \underline{0} , \\
&\lambda\, \underline{\pi}_{i-1} + \underline{\pi}_i (S-\lambda I) + \underline{\pi}_{i+1}\, S^\circ B^\circ = \underline{0} , \qquad \text{for} \quad 2 \le i \le K-1 \\
&\lambda\, \underline{\pi}_{K-1} + \underline{\pi}_K\, S = \underline{0} .
\end{aligned}
$$

Postmultiplying all equations in (4), but the first, by $\underline{e}$, we obtain $\lambda\, \underline{\pi}_i\, \underline{e} = \underline{\pi}_{i+1}\, \underline{S}^\circ$, for $1 \le i \le K-1$, or equivalently $\lambda\, \underline{\pi}_i\, B^{\circ\circ} = \underline{\pi}_{i+1}\, S^\circ B^\circ$. Replacing the rightmost term in the equations for $i = 1, \ldots, K-1$ by $\lambda\, \underline{\pi}_i\, B^{\circ\circ}$, we obtain

$$
\begin{aligned}
&\lambda\, \pi_0\, \underline{\beta} = \underline{\pi}_1\, (\lambda I - \lambda B^{\circ\circ} - S) \\
&\lambda\, \underline{\pi}_{i-1} = \underline{\pi}_i\, (\lambda I - \lambda B^{\circ\circ} - S) , \qquad \text{for} \quad 2 \le i \le K-1 .
\end{aligned}
$$

This readily leads to the formulas (2). The non-singularity of the matrix C may easily be proved by contradiction.

The entire system may now be studied as a Markov process on the state space $E = \{k \geq 0\} \quad E_1$, with generator Q^* given by

$$
(5) \qquad Q^* = \begin{vmatrix} Q-\lambda\Delta & \lambda\Delta & & & \\ \mu I & Q-\lambda\Delta-\mu I & \lambda\Delta & & \\ & \mu I & Q-\lambda\Delta-\mu I & \lambda\Delta & \\ & & \mu I & Q-\lambda\Delta-\mu I & \cdots \\ & & \vdots & \vdots & \end{vmatrix} ,
$$

where Δ denotes a diagonal matrix of order $Km+1$, whose last m diagonal elements are equal to one. All its other elements are zero. In Formula (5), I denotes an identity matrix of order $Km+1$.

The matrix Q^* is a particular case of the generators, discussed in [11]. The following theorem is therefore immediate.

<u>Theorem 1</u>

The queueing system is <u>stable</u> if and only if

$$
(6) \qquad \lambda \, \underline{\pi}_K \, \underline{e} < \mu .
$$

The stationary probability vector $\underline{x}$ of Q^*, partitioned into vectors $\underline{x}_k$, $k \geq 0$, of dimension $Km+1$, is given by

$$
(7) \qquad \underline{x}_k = \underline{\pi} \, (I-R) \, R^k , \qquad \text{for } k \geq 0 .
$$

The matrix R is the minimal nonnegative solution to the equation

$$
(8) \qquad \mu \, R^2 + R \, (Q-\lambda\Delta-\mu I) + \lambda\Delta = 0 ,
$$

and has spectral radius less than one.

As was pointed out in [11], the fact that all but the last m rows of Δ vanish implies that the same is true of the matrix R. Although the matrix R is of order $Km+1$, only its last m rows are positive. This special structural property may be exploited to compute R efficiently.

2. Algorithmic Procedure

Taking the simplified structure of the matrices R and Q into account, we may rewrite Equation (8) as follows. We partition the last m rows of R into $[\underline{R}_0, R_1, R_2, \ldots, R_K]$, where $\underline{R}_0$ is a column vector of dimension m and R_i, $1 \leq i \leq K$, are square matrices of order m. It then follows from (1) and (8) that

$$
(9)\quad
\begin{aligned}
&\mu R_K\underline{R}_0 - (\lambda+\mu)\underline{R}_0 + R_1\underline{S}^\circ = \underline{0}, \\
&\mu R_K R_1 + \lambda\underline{R}_0\underline{\beta} + R_1[S-(\lambda+\mu)I] + R_2 S^\circ B^\circ = 0, \\
&\mu R_K R_i + \lambda R_{i-1} + R_i[S-(\lambda+\mu)I] + R_{i+1} S^\circ B^\circ = 0, \\
&\qquad\qquad \text{for } 2 \leq i \leq K-1, \\
&\mu R_K^2 + \lambda R_{K-1} + R_K[S-(\lambda+\mu)I] + \lambda I = 0.
\end{aligned}
$$

Setting $M = [(\lambda+\mu)I-S]^{-1}$, the equations (9) may be written in the convenient form

$$
(10)\quad
\begin{aligned}
&\underline{R}_0 = (\lambda+\mu)^{-1}(\mu R_K\underline{R}_0 + R_1\underline{S}^\circ), \\
&R_1 = [\mu R_K R_1 + (\lambda\underline{R}_0 + R_2\underline{S}^\circ)\underline{\beta}] M, \\
&R_i = [\mu R_K R_i + \lambda R_{i-1}] M + R_{i+1}\underline{S}^\circ \cdot \underline{\beta} M, \\
&\qquad\qquad \text{for } 2 \leq i \leq K-1, \\
&R_K = [\mu R_K^2 + \lambda R_{K-1} + \lambda I] M.
\end{aligned}
$$

It may be verified by an elementary induction argument that the Gauss-Seidel type iterates, defined for $n \geq 0$, by

$$\underline{R}_0(n+1) = (\lambda+\mu)^{-1} [\mu\, R_K(n+1)\, \underline{R}_0(n) + R_1(n+1)\underline{S}^\circ] ,$$

$$R_1(n+1) = \{\mu\, R_K(n+1)\, R_1(n) + [\lambda\, \underline{R}_0(n) + R_2(n+1)\underline{S}^\circ]\, \underline{\beta}\}M ,$$

$$R_i(n+1) = [\mu\, R_K(n+1)\, R_i(n) + \lambda\, R_{i-1}(n)]\, M + R_{i+1}(n+1)\underline{S}^\circ \cdot \underline{\beta}\, M , \qquad \text{for } 2 \leq i \leq K-1 ,$$

$$R_K(n+1) = [\mu\, R_K^2(n) + \lambda\, R_{K-1}(n) + \lambda I]\, M ,$$

with $R_K(0) = \lambda M$, $R_i(0) = 0$, $1 \leq i \leq K-1$, $\underline{R}_0(0) = \underline{0}$, converges increasingly to the desired matrix R. This method of computation has clear advantages over successive substitutions.

As was proved in [8], the matrix R satisfies the relation

$$\mu\, \tilde{R}\, \underline{e} = \mu \left[\underline{R}_0 + \sum_{j=1}^{K} R_j\, \underline{e}\right] = \lambda\, \underline{e} , \tag{11}$$

which provides a useful accuracy check on the computation of R. The matrix $\tilde{R}$ is formed by the last m rows of R.

From Equation (7), by using the structure of R, we obtain

$$\begin{aligned} \underline{x}_0 &= \underline{\pi} - \underline{\pi}_K\, \tilde{R} , \\ \underline{x}_k &= \underline{\pi}_K\, (I-R_K)\, R_K^{k-1}\, \tilde{R} , \qquad \text{for } k \geq 1 . \end{aligned} \tag{12}$$

By virtue of (11), the marginal density of the queue length in Unit 2 is given by

$$\begin{aligned} y_0 &= 1 - \lambda\, \mu^{-1}\, \underline{\pi}_K\, \underline{e} , \\ y_k &= \lambda\, \mu^{-1}\, \underline{\pi}_K\, (I-R_K)\, R_K^{k-1}\underline{e} , \qquad \text{for } k \geq 1. \end{aligned} \tag{13}$$

It is of particular interest to study the conditional densities of the queue length in Unit 2, given the number of customers present in Unit 1. To this end, we partition the vector $\underline{x}_k$ as $[\underline{x}_k^{(0)}, \underline{x}_k^{(1)}, \ldots, \underline{x}_k^{(K)}]$, where the $\underline{x}_k^{(r)}$, $1 \leq r \leq K$, are m-vectors. The conditional queue length density, given that there are j customers in Unit 1 is then given by

$$
(14) \qquad \begin{aligned} q_j^{(k)} &= (\underline{\pi}_j \underline{e})^{-1} \underline{x}_k^{(j)} \underline{e}, \qquad \text{for } 1 \leq j \leq K, \\ q_0^{(k)} &= \pi_0^{-1} x_k^{(0)}, \end{aligned}
$$

for $k \geq 0$.

From the formulas (12), (13) and (14), expressions for the first and second moments of the corresponding queue lengths may be routinely calculated.

3. The Waiting Time Distribution

We now consider the stationary distribution of the waiting time of an arriving customer, under the first-come, first-served queue discipline. There are two cases. Either the customer is admitted into Unit 1 and is served immediately or he waits for a time not exceeding x units of time; or the customer finds the first unit full and goes to the second unit. If it is free, he is immediately taken into service. If not, he may have to wait there for not more than x units of time. We denote the conditional waiting time distributions, corresponding to these two alternatives by $W_1(\cdot)$ and $W_2(\cdot)$ respectively.

$W_1(x)$ may be viewed as the distribution of the time till absorption in the Markov Process with infinitesimal generator

$$Q_1 = \begin{array}{c} 0 \\ \underline{1} \\ \underline{2} \\ \underline{3} \\ \underline{4} \\ \vdots \\ \\ \underline{K-1} \end{array} \left[\begin{array}{ccccccc} 0 & \underline{0} & \underline{0} & & \cdots & & \\ \underline{S}^\circ & S & & & & & \\ \underline{0} & S^\circ B^\circ & S & & & & \\ & 0 & S^\circ B^\circ & S & & & \\ & & 0 & S^\circ B^\circ & & & \\ \vdots & \vdots & \vdots & \vdots & & & \\ & & & & & S & 0 \\ & & & & & S^\circ B^\circ & \underline{S} \end{array} \right]$$

with the initial probability vector,

(15) $$\underline{y}_0 = (1-\underline{\pi}_K\, \underline{e})^{-1} (\pi_0, \underline{\pi}_1, \underline{\pi}_2, \ldots, \underline{\pi}_{K-1}) \ .$$

In order to compute this distribution, we consider the system of differential equations,

(16) $$\begin{aligned} y_0'(x) &= \underline{y}_1(x)\, \underline{S}^\circ \ , \\ \underline{y}_i'(x) &= \underline{y}_i(x)\, S + \underline{y}_{i+1}(x) \underline{S}^\circ \cdot \underline{\beta} \ , \qquad \text{for } 1 \le i \le K-2 \ , \\ \underline{y}_{K-1}'(x) &= \underline{y}_{K-1}(x)\, S \ , \end{aligned}$$

with initial conditions, given by (15) .

Clearly $[\underline{y}_i(x)]_j$ is the probability, that the customer is in the state (i,j), $1 \le j \le m$, of the Markov process Q_1 at time x. From this it follows that $W_1(x) = y_0(x)$, for $x \ge 0$.

The conditional distribution $W_2(\cdot)$ is obtained as follows. Since the service time in Unit 2 is exponential with parameter μ, the Laplace-Stieltjes transform $W_2^*(\cdot)$ is given by,

$$W_2^*(s) = (\underline{\pi}_K \underline{e})^{-1} \sum_{r=0}^{\infty} \left(\frac{\mu}{\mu+s}\right)^r \underline{x}_r^{(K)} \underline{e}$$

$$(17) \qquad = (\underline{\pi}_K \underline{e})^{-1} \sum_{r=0}^{\infty} \left(\frac{\mu}{\mu+s}\right)^r \underline{\pi}_K (I-R_K) R_K^r \underline{e}$$

$$= (\underline{\pi}_K \underline{e})^{-1} \underline{\pi}_K (I-R_K)\underline{e} + (\underline{\pi}_K \underline{e})^{-1} \mu \underline{\pi}_K (I-R_K) R_K [sI-\mu(R_K-I)]^{-1} \underline{e} .$$

Upon inversion, we obtain

$$(18) \qquad W_2(x) = 1 - (\underline{\pi}_K \underline{e})^{-1} \underline{\pi}_K R_K \exp[\mu x(R_K-I)]\underline{e} , \qquad \text{for} \quad x \geq 0 .$$

To compute $W_2(\cdot)$, it suffices to solve the system of differential equations

$$\underline{v}'(x) = \mu \underline{v}(x) (R_K-I) ,$$

with initial conditions $\underline{v}(0) = (\underline{\pi}_K \underline{e})^{-1} \underline{\pi}_K R_K$, and to form $W_2(x) = 1 - \underline{v}(x)\underline{e}$, for $x \geq 0$.

Finally, the unconditional waiting time distribution is obtained as,

$$W(x) = (1-\underline{\pi}_K \underline{e}) W_1(x) + \underline{\pi}_K \underline{e} W_2(x) , \qquad \text{for} \quad x \geq 0 .$$

The mean waiting time of an arriving customer is computed by considering the two units separately. The Laplace-Stieltjes transform of $W_1(x)$ is given by,

$$W_1^*(s) = \pi_0 + \sum_{r=1}^{K-1} \underline{\pi}_r [(sI-S)^{-1} S^\circ B^\circ]^r \underline{e}$$

$$= \pi_0 + \sum_{r=1}^{K-1} \underline{\pi}_r (sI-S)^{-1} \underline{S}^\circ f^{r-1}(s) ,$$

where $f(s) = \underline{\beta} (sI-S)^{-1} \underline{S}^\circ$. The mean of $W_1(\cdot)$ is hence given by ,

$$- W_1^{*'}(0) = -\sum_{r=1}^{K-1} \underline{\pi}_r S^{-1}\underline{e} + \mu_1' \sum_{r=2}^{K-1} (r-1) \underline{\pi}_r \underline{e} ,$$

where μ_1' is the mean service time in Unit 1. Similarly, upon differentiating $W_2^*(s)$ we get

$$- W_2^{*'}(0) = \mu^{-1} \left[(\underline{\pi}_k \underline{e})^{-1} \underline{\pi}_k (I-R_k)^{-1} \underline{e}-1 \right] .$$

The mean waiting time W is then given by

$$W = (1-\underline{\pi}_K \underline{e}) (-W_1^{*'}(0)) + (\underline{\pi}_K \underline{e}) (-W_2^{*'}(0)) .$$

4. Several Servers in Unit 2

Here we allow the second unit to have c exponential servers of service rate μ. The queueing system is now studied as a Markov process on the state space $E = \{k \geq 0\} \times E_1$, with infinitesimal generator

$$Q_2 = \begin{vmatrix} A_{00} & A_{01} & & & & & & & \\ A_{10} & A_{11} & A_{12} & & & & & & \\ & & \cdots & & & & & & \\ & & & A_{c-2,0} & A_{c-2,1} & A_{c-2,2} & & & \\ & & & & A_{c-1,0} & A_{c-1,1} & A_0 & & \\ & & & & & A_2 & A_1 & A_0 & \\ & & & & & & A_2 & A_1 & \cdots \\ & & & & & & & \cdots & \end{vmatrix} .$$

The square blocks of dimension $Km+1$ are given by,

$$A_{00} = Q - \lambda\Delta \quad ,$$
$$A_{01} = A_{12} = \cdots = A_{c-1,2} = A_0 = \lambda\Delta \ ,$$
$$A_{i0} = i\mu I \ , \qquad \text{for} \quad 1 \le i \le c-1 \ ,$$
$$A_{i1} = Q-\lambda\Delta-i\mu I \ , \qquad \text{for} \quad 1 \le i \le c-1 \ ,$$
$$A_1 = Q-\lambda\Delta-c\mu I \ ,$$
$$A_2 = c\mu I \ ,$$

where the matrices Q and Δ are as given in Section 1.

Then, corresponding to Theorem 1, we have

Theorem 2. The queueing system is stable if and only if

$$\lambda \, \underline{\pi}_K \, \underline{e} < c \, \mu \ .$$

The stationary probability vector $\underline{x}$ is partitioned as $(\underline{x}_0, \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_c, \underline{x}_{c+1}, \ldots)$, where each of the component vectors is of dimension $Km+1$. The vectors $\underline{x}_k$, for $k \ge c$, are given by,

$$\underline{x}_k = \underline{x}_{c-1} \, R^{k-c+1} \ . \tag{19}$$

The matrix R is the unique solution in the set of nonnegative matrices of spectral radius less than one, of the matrix-quadratic equation,

$$c \, \mu \, R^2 + R \, (Q-\lambda\Delta-c\mu I) + \lambda\Delta = 0 \ . \tag{20}$$

The last m rows of R are positive; all others are zero.

The vector $(\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_{c-1})$ is the left eigenvector, corresponding to the eigenvalue zero, of the irreducible semi-stable matrix

$$(21) \quad T = \begin{vmatrix} A_{00} & A_{01} & 0 & \cdots & & & \\ A_{10} & A_{11} & A_{12} & \cdots & & & \\ & A_{20} & A_{21} & \cdots & & & \\ & & & & A_{c-2,0} & A_{c-2,1} & A_{c-2,2} \\ & & & & & A_{c-1,0} & A_{c-1,1}+RA_2 \end{vmatrix}$$

and is normalized by

$$\sum_{i=0}^{c-2} \underline{x}_i \, \underline{e} + \underline{x}_{c-1} \, (I-R)^{-1} \, \underline{e} = 1 \ .$$

Proof. The statements about the steady-state vector $\underline{x}$ were proved in [9]. That the matrix R has the particular structure is clear from the discussion in Section 1.

The matrix R is computed using the algorithmic procedure described in Section 2, as are the steady-state vector and the other quantities of interest.

The Waiting Time Distribution

In the computation of the waiting time distribution of an arriving customer, only the distribution $W_2(\cdot)$ requires some discussion. Its Laplace-Stieltjes transform $W_2^*(s)$ is given by

$$W_2^*(s) = (\underline{\pi}_K \underline{e})^{-1} \left[\sum_{i=0}^{c-1} \underline{x}_i^{(K)} \underline{e} + \sum_{i=c}^{\infty} \left(\frac{c\mu}{c\mu+s} \right)^{i-c+1} \underline{x}_i^{(K)} \, \underline{e} \right] \ .$$

Since $\underline{x}_i^{(K)} = \underline{x}_{c-1}^{(K)} R_K^{i-c+1}$, for $i \geq c$, we obtain

$$W_2^*(s) = (\underline{\pi}_K \underline{e})^{-1} \left[\sum_{i=0}^{c-1} \underline{x}_i^{(K)} \underline{e} + c\mu \ \underline{x}_{c-1}^{(K)} R_K \ (sI+c\mu I-c\mu R_K)^{-1} \underline{e} \right] \ .$$

Upon inversion and simplification, one obtains

$$W_2(x) = 1 - (\underline{\pi}_K \underline{e})^{-1} \underline{x}_{c-1}^{(K)} (I-R_K)^{-1} R_K \exp[-c\mu x(I-R_K)]\underline{e}\ ,$$

$$\text{for} \quad x \geq 0 \ .$$

We see that the distribution $W_2(\cdot)$ may again be computed by the solution of a simple system of differential equations, similar to that obtained in Section 3.

5. Overflows from a Multi-server Unit 1 to a Unit 2 with Several Servers

In this section, we consider the system with Unit 1 having r exponential servers with parameter μ_1, and the second unit with c exponential servers, with parameter μ_2. A particular case of this model, in which both units have single exponential servers with the same rate and in which the waiting room in unit 2 is also of finite capacity, has been studied in [2].

We describe the unit 1 of the present model by a Markov process on the state space $E_1 = \{0, 1, 2, \ldots, K\}$, with infinitesimal generator

$$Q_3 = \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ r \\ r+1 \\ \vdots \\ K \end{array} \left|\begin{array}{ccccccccc} -\lambda & \lambda & & & & & & & \\ \mu_1 & -\lambda-\mu_1 & \lambda & & & & & & \\ & 2\mu_1 & -\lambda-2\mu_1 & \lambda & & & & & \\ & & \cdots & & & & & & \\ & & & r\mu_1 & -\lambda-r\mu_1 & \lambda & & & \\ & & & & r\mu_1 & -\lambda-r\mu_1 & \lambda & & \\ & & & & & \cdots & & & \\ & & & & & & & r\mu_1 & -r\mu_1 \end{array}\right| .$$

The stationary probability vector $\underline{z}$ of Q_3 is given by

$$z_n = \left(\frac{\lambda}{\mu_1}\right)^n \frac{1}{n!} z_0 , \qquad \text{for} \quad 1 \le n \le r-1 ,$$
$$= \left(\frac{\lambda}{\mu_1}\right)^n \frac{1}{r^{n-r} r!} z_0 , \qquad \text{for} \quad r \le n \le K ,$$

where

$$z_0 = \left[\sum_{n=0}^{r-1} \left(\frac{\lambda}{\mu_1}\right)^n \frac{1}{n!} + \frac{1}{r!} \sum_{n=r}^{K} \left(\frac{\lambda}{\mu_1}\right)^n \frac{1}{r^{n-r}}\right]^{-1} .$$

The queueing model now readily leads to a quasi-birth-and-death process on the state space $E = \{i \ge 0\} \times E_1$, with generator

$$Q_4 = \left|\begin{array}{ccccccccc} A_{00} & A_{01} & & & & & & & \\ A_{10} & A_{11} & A_{12} & & & & & & \\ & & \cdots & & & & & & \\ & & & A_{c-2,0} & A_{c-2,1} & A_{c-2,2} & & & \\ & & & & A_{c-1,0} & A_{c-1,1} & A_0 & & \\ & & & & & A_2 & A_1 & A_0 & \\ & & & & & & A_2 & A_1 & \cdots \\ & & & & & & & \cdots & \end{array}\right| ,$$

where

$$A_{00} = Q_3 - \lambda\Delta_1 ,$$
$$A_{01} = A_{12} = \cdots = A_{c-2,2} = A_0 = \lambda\Delta_1 ,$$
$$A_{i0} = i\mu_2 I , \qquad \text{for} \quad 1 \le i \le c-1 ,$$
$$A_{i1} = Q_3 - \lambda\Delta_1 - i\mu_2 I , \qquad \text{for} \quad 1 \le i \le c-1 ,$$
$$A_2 = c\mu_2 I ,$$
$$A_1 = Q_3 - \lambda\Delta_1 - c\mu_2 I ,$$

and Δ_1 is the diagonal matrix of order $K+1$, whose last diagonal element is one and all others are zero.

The following result is then immediate.

<u>Theorem 3</u>. Under the equilibrium condition

$$\lambda\, z_K < c\,\mu_2 ,$$

the stationary probability vector $\underline{x}$ of Q_4, in partitioned form, is given by,

$$\underline{x}_i = \underline{x}_{c-1}\, R^{i-c+1} , \qquad \text{for } i \geq c . \tag{22}$$

The matrix R is the unique solution in the set of non-negative matrices of spectral radius less than one, of the equation

$$c\,\mu_2\, R^2 + R\,(Q_3 - c\mu_2 I - \lambda\Delta_1) + \lambda\Delta_1 = 0 . \tag{23}$$

The vector $\underline{y} = (\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_{c-1})$, is obtained as the left eigenvector corresponding to the eigenvalue zero of an irreducible, semi-stable matrix, similar to the matrix of Equation (21). It is normalized by the condition $\sum_{i=0}^{c-2} \underline{x}_i\, \underline{e} + \underline{x}_{c-1}(I-R)^{-1}\, \underline{e} = 1$.

Only the last row of R differs from zero and is strictly positive.

The elements $R_0, R_1, \ldots, R_K$ in the last row of R are computed by iterative solution of the system of equations

$$c\mu_2\, R_K\, R_0 - (\lambda + c\mu_2)\, R_0 + \mu_1\, R_1 = 0 ,$$
$$c\mu_2\, R_K\, R_j + \lambda\, R_{j-1} - (\lambda + j\mu_1 + c\mu_2)\, R_j + (j+1)\,\mu_1\, R_{j+1} = 0 ,$$
$$\text{for } 1 \leq j \leq r-1 ,$$

$$c\mu_2 R_K R_j + \lambda R_{j-1} - (\lambda+r\mu_1+c\mu_2) R_j + r\mu_1 R_{j+1} = 0 ,$$
$$\text{for} \quad r \le j \le K-1 ,$$
$$c\mu_2 R_K^2 + \lambda R_{K-1} - (\lambda+r\mu_1) R_K + \lambda = 0 .$$

The accuracy check

$$\sum_{j=0}^{K} R_j = \frac{\lambda}{c\mu_2} ,$$

should be satisfied.

6. Additional Variants

For each of the three preceding models, the Unit 2 may be assumed to have its own Poisson arrival process of rate λ_1. The general results on the matrix-geometric form of the stationary probability vector $\underline{x}$ remain valid for this case, but the major simplifications due to the particular structure of R are lost for $\lambda_1 > 0$. The matrix R is then strictly positive. In the computation of R, it is still worthwhile to write that matrix in partitioned form, so that the highly sparse and regular structure of the coefficient matrices may be exploited. How this may be done is obvious once the analogues of the equations (9) are written down.

The overflow streams from the bounded M/PH/1 and M/M/r queues are particular cases of the versatile Markovian point process, introduced in [10]. By implementing the general, but much more belabored algorithm in [12], it is possible to study the case where Unit 2 has a single server with a general service time distribution. If one allows the service time distribution in that unit to be of phase type, the methods of the present paper may be implemented. The large matrices, which now arise, need to be handled with care to avoid problems of storage and large computation time. A brief discussion of this case may be found in Chapter 6 of [11].

7. Numerical Examples

In some of the small number of papers, which give numerical results for overflow models, it has been noted that such results may be quite astonishing and that naive approximations to the overflow stream may lead to gross errors [3-5]. This observation may be made generally for queues in which arrival and service rates are subject to (random) fluctuations [8, 11]. The underlying cause for the highly overdispersed queue length and waiting time distributions for such models lies in the large variations, which typically occur in the interarrival times to Unit 2. In order to illustrate this adequately, it is necessary to consider several related numerical examples.

We consider the system consisting of an M/PH/1 queue of finite capacity $K = 4$, with overflow to a single exponential server. The service time distribution in Unit 1 was chosen to be $\frac{38}{39} E_2(8,x) + \frac{1}{39} E_1(0.1,x)$. Its mean is 0.5. Its particular form was chosen to reflect a case in which occasionally long service times occur. The arrival rate λ to the system is 20, so that Unit 1 is saturated 90% of the time.

The service rate μ in Unit 2 was chosen so as to obtain three different cases, respectively with $\rho = 0.92$, 0.94 and 0.96. Before giving the numerical results, that are obtained, we briefly consider an appealing, but naive "approximation". It may be argued that the overflow stream from Unit 1 could be "approximated" by a Poisson process of rate 0.9λ. This would lead, by elementary formulas for the M/M/1 queue, the following values for the mean and the standard deviation of the queue length in Unit 2.

ρ	mean	st. dev.
0.92	11.50	11.99
0.94	15.67	16.16
0.96	24.00	24.49

The correct computed values are however much larger.

ρ	mean	st. dev.
0.92	18.99	23.39
0.94	27.64	33.44
0.96	45.24	53.33

The reasons why the naive approximation fares so poorly are similar to those discussed in [8]. During periods of overflow, the second unit behaves like a mildly unstable M/M/1 queue, which may recover from its build-up during the periods when Unit 1 is not filled to capacity. During the rare long service times in Unit 1, the latter will almost certainly become saturated and remain so for a fairly long time. The build-ups in the second unit will then be substantial and cause large queue lengths for a long time thereafter.

The numerical results for Unit 2 are not significantly affected by a moderate increase in the capacity of Unit 1. A change of e.g. $K = 4$, to $K = 8$, resulted in only a minute change in the computed characteristics of Unit 2. The additional four waiting spaces are occupied most of the time and do not appreciably affect the rate of overflow.

As is to be expected, the queue in Unit 2 is much smoother when the service times in Unit 1 exhibit less random variation. This is seen by replacing the earlier service time distribution, e.g. by $E_3(3,x)$.

A variety of other aspects of overflow in queues may easily be numerically investigated by implementing the algorithm proposed in this paper. The numerical results are often astonishing at first, but are seen - after some reflection - to correspond to intuitive qualitative behavior of the queue. They suggest that great care is needed in the interpretation of numerical results for networks with capacity constraints.

REFERENCES

[1] ÇINLAR, E. and DISNEY, R. L. (1967)
Streams of Overflows from a Finite Queue
Opns. Res., 15, 131-134.

[2] DISNEY, R. L. (1972)
A Matrix Solution for the Two Server Queue with Overflow
Mgmt. Sci., 19, 254-265.

[3] McNICKLE, D. C. (1979)
Congestion in a Simple Overflow Model
New Zealand Operational Research, 7, 151-157.

[4] McNICKLE, D. C.
Congestion in Overflow Queues
Proc. Operations Res. Soc., New Zealand, August 1979, 83-87.

[5] McNICKLE, D. C.
Notes on Congestion in Overflow Queues
Tech. Rept., Economics Dept., Univ. of Canterbury, New Zealand, December 1979.

[6] NEUTS, M. F. (1975)
Probability Distributions of Phase type
Liber Amicorum Prof. Emeritus H. Florin, Depart. Math., Univ. Louvain, Belgium, 173-206.

[7] NEUTS, M. F. (1978)
Markov Chains with Applications in Queueing Theory which have a Matrix-Geometric Invariant Vector
Adv. Appl. Prob., 10, 185-212.

[8] NEUTS, M. F. (1978)
The M/M/1 Queue with Randomly Varying Arrival and Service Rates
Opsearch, 15, 139-157.

[9] NEUTS, M. F. (1978)
Further Results on the M/M/1 Queue with Randomly Varying Rates
Opsearch, 15, 158-168.

[10] NEUTS, M. F. (1979)
A Versatile Markovian Point Process
J. Appl. Prob., 16, 764-779.

[11] NEUTS, M. F. (1981)
<u>Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach</u>
The Johns Hopkins University Press, Baltimore, MD - to appear.

[12] RAMASWAMI, V. (1980)
The N/G/1 Queue and its Detailed Analysis
Adv. Appl. Prob., 12, 222-261.

[13] RATH, J. H.
An Approximation for a Queueing System with Two Queues and Overflows
Tech. Rept., Bell Labs, Holmdel, New Jersey 07733, USA

[14] RATH, J. H. and SHENG, DIANE (1979)
Approximations for Overflows from Queues with a Finite Waiting Room
Opns. Res., 27, 1208-1216.

LMED

-8