Z. Wahrscheinlichkeitstheorie verw. Gebiete
57, 441-452 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Asymptotic Behavior of the Stationary Distributions in the GI/PH/c Queue with Heterogeneous Servers*

Marcel F. Neuts[1] and Yukio Takahashi[2]

[1] Department of Mathematical Sciences, University of Delaware, Newark DE 19711, USA
[2] Department of Economics, Tohoku University, Kawauchi-Sendai 980, Japan

Summary. This paper deals with the stable c-server queue with renewal input. The service time distributions may be different for the various servers. They are however all probability distributions of phase type. It is shown that the stationary distribution of the queue length at arrivals has an exact geometric tail of rate $\eta, 0<\eta<1$. It is further shown that the stationary waiting time distribution at arrivals has an exact exponential tail of decay parameter $\xi>0$. The quantities η and ξ may be evaluated together by an elementary algorithm. For both distributions, the multiplicative constants which arise in the asymptotic forms may be fully characterized. These constants are however difficult to compute in general.

1. Introduction

Very few algorithmically tractable results are known for multi-server queues, except in the restrictive case where the service time distributions are exponential. In this paper, we obtain results on the tail behavior of the stationary distributions of the queue length and the waiting time for a c-server queue with renewal input. The various servers are allowed to be *heterogeneous*, i.e. the service time distributions may be different for different servers. The c service time distributions are however required to be *of phase type.*

If P_m and $W(x)$ denote respectively the stationary probabilities that a customer arriving to the queue finds at least m customers in the system and that he has to wait for a time at most x, then we establish the asymptotic formulas

$$P_m = K\eta^m + o(\eta^m), \qquad \text{as } m\to\infty,$$

and

$$1-W(x) = K_1 e^{-\xi x} + o(e^{-\xi x}), \qquad \text{as } x\to\infty,$$

* This research was supported by the National Science Foundation under Grant Nr. ENG-7908351 and by the Air Force Office of Scientific Research under Grant Nr. AFOSR-77-3236.

0044-3719/81/0057/0441/$02.40

where $K_1 = \eta^c K$. The constants η and ξ, which satisfy $0 < \eta < 1$ and $\xi > 0$, will be shown to satisfy a system of equations, which may easily be solved by elementary numerical procedures. The positive constants K and K_1 will be fully characterized, but their direct numerical evaluation is seen to be difficult, in general.

In the course of the discussion, a number of results of theoretical interest on the GI/PH/c queue with heterogeneous servers will be obtained. Because of the huge dimensions of the matrices, which arise in these results, their feasibility for algorithmic implementations is limited to particular cases.

The proofs of the various results in this paper rely heavily on the elementary properties of phase type distributions [4, 5, 8] and on the theory of block-partitioned stochastic matrices with a matrix-geometric invariant vector [6, 7, 7]. Specific references will be given for each prior result, which is used, but no easily available proofs will be repeated. The particular results for the GI/PH/1 queue were established in Neuts [8, 9]. Corresponding results for the PH/PH/c queue with identical servers are discussed in Takahashi [11], while iterative numerical procedures for that model are proposed in Takahashi and Takami [10].

The formal description of the model is as follows. Customers arrive to a c-server system according to a renewal process with interarrival time distribution $F(\cdot)$ of finite mean λ'. The distribution $F(\cdot)$ satisfies $F(0+)=0$. The service time distribution of a customer may depend on the server to which he gains access. Services by the j-th server, $1 \leqq j \leqq c$, have a common distribution *of phase type*, with irreducible representation $[\boldsymbol{\beta}(j), S(j)]$ where $\boldsymbol{\beta}(j)$ is a probability row-vector of dimension $\nu(j)$ and $S(j)$ is a square, stable matrix of order $\nu(j)$. The corresponding vector $\mathbf{S}^0(j)$ is defined by $\mathbf{S}^0(j) = -S(j)\mathbf{e}$. Throughout this paper, the symbol $\mathbf{e}$ will denote a column vector with all its components equal to one and of dimension appropriate to the formula in which it occurs. The mean service time $\mu'(j)$ of the j-th server is given by $\mu'(j) = -\boldsymbol{\beta}(j) S^{-1}(j)\mathbf{e}$.

It will be assumed that all service times are independent of the arrival process. For any $k \geqq 2$ successive customers, the service times are assumed to be conditionally independent, given the labels of the servers by which they are processed. Under the stated assumptions, the model has a Markov chain, embedded at the successive epochs of arrival. Its states are of three types. For $i \geqq c$, the state described by $(i, h_1, h_2, \ldots, h_c)$ signifies that, immediately prior to the arrival, there are i customers in the system and that for $1 \leqq j \leqq c$, the j-th server is in the phase h_j of his PH-distribution. The set of all such states with a fixed index i will be denoted by $\mathbf{i}$. The states $(c-1; h_1, \ldots, h_c)$ are similarly defined, but one of the phase states now corresponds to the initial service phase, selected by the arriving customer. The remaining states correspond to the case where, prior to the arrival, there are fewer than $c-1$ customers in the system. The precise labeling of these states is immaterial to our discussion. The set of all such states will be denoted by E. The states $(i, h_1, \ldots, h_c)$, for $i \geqq c-1$, $1 \leqq j \leqq c$, $1 \leqq h_j \leqq \nu(j)$, are listed in lexicographic order. The states in $E \cup (\mathbf{c-1})$ are called *the boundary states.*

In order to define the transition probability matrix $\tilde{P}$ of the embedded Markov chain, we introduce the matrices $P_j(r, t)$, for $r \geqq 0$, $t \geqq 0$, and $1 \leqq j \leqq c$.

These matrices satisfy the differential equations

$$\begin{aligned} P_j'(0,t) &= P_j(0,t)S(j), \\ P_j'(r,t) &= P_j(r,t)S(j) + P_j(r-1,t)\mathbf{S}^0(j)\boldsymbol{\beta}(j), \end{aligned} \tag{1}$$

for $r \geqq 1$, with the initial conditions $P_j(r,0)=\delta_{r0}I$, for $r \geqq 0$. Their significance is the same as in [5] or [6].

We note that

$$\sum_{r=0}^{\infty} P_j(r,t)z^r = \exp\{[S(j)+z\mathbf{S}^0(j)\boldsymbol{\beta}(j)]t\}, \tag{2}$$

for $t \geqq 0$, $0 \leqq z \leqq 1$.

The matrices A_k, $k \geqq 0$, are defined by

$$A_k = \sum \int_0^{\infty} P_1(r_1,t) \otimes P_2(r_2,t) \otimes \ldots \otimes P_c(r_c,t)\,dF(t), \tag{3}$$

for $k \geqq 0$. The summation is over all c-tuples $(r_1,\ldots,r_c)$, which satisfy $r_1 \geqq 0,\ldots,r_c \geqq 0$, $r_1+\ldots+r_c=k$. The symbol $\otimes$ stands for the Kronecker product of matrices. As shown in [6], the matrices $P_j(r,t)$ are positive for $r \geqq 1$, $t>0$, and $1 \leqq j \leqq c$. It is therefore clear that the matrices A_k, $k \geqq c$, are positive. The matrices A_k, $k \geqq 0$, are of order m, given by

$$m = \prod_{j=1}^{c} v(j). \tag{4}$$

From Formulas (2) and (3), it follows that

$$\begin{aligned} A^*(z) &= \sum_{k=0}^{\infty} A_k z^k \\ &= \int_0^{\infty} \exp\{[S(1)+z\mathbf{S}^0(1)\boldsymbol{\beta}(1)]t\} \otimes \ldots \otimes \exp\{[S(c)+z\mathbf{S}^0(c)\boldsymbol{\beta}(c)]t\}\,dF(t). \end{aligned} \tag{5}$$

The matrix $A^*(z)$ is positive for $0<z \leqq 1$.

The transition probability matrix $\tilde{P}$ is now given by

$$\tilde{P} = \begin{array}{c|cc|ccccc} & E & \mathbf{c-1} & \mathbf{c} & \mathbf{c+1} & \mathbf{c+2} & \mathbf{c+3} & \ldots \\ \hline E & & & 0 & 0 & 0 & 0 & \ldots \\ \mathbf{c-1} & & & A_0 & 0 & 0 & 0 & \ldots \\ \hline \mathbf{c} & & & A_1 & A_0 & 0 & 0 & \ldots \\ \mathbf{c+1} & & & A_2 & A_1 & A_0 & 0 & \ldots \\ \mathbf{c+2} & & & A_3 & A_2 & A_1 & A_0 & \ldots \\ \mathbf{c+3} & & & A_4 & A_3 & A_2 & A_1 & \ldots \\ \vdots & & & \vdots & \vdots & \vdots & \vdots & \end{array} \tag{6}$$

The elements in the columns, labeled E and $\mathbf{c}-\mathbf{1}$, are immaterial to our discussion. They are, in general, exceedingly complicated and depend on the rule by which arriving customers are assigned to free servers. Explicit, but highly involved expressions are given for the case $c=2$, in Chap. 4 of [8]. In all properly defined cases, the matrix $\tilde{P}$ is irreducible.

The matrix $A=\sum_{k=0}^{\infty} A_k=A^*(1)$, is a strictly positive, stochastic matrix. Let $\boldsymbol{\theta}(j)$ be the positive probability vector satisfying

$$\boldsymbol{\theta}(j)[S(j)+\mathbf{S}^0(j)\boldsymbol{\beta}(j)]=\mathbf{0} \qquad \boldsymbol{\theta}(j)\mathbf{e}=1, \tag{7}$$

for $1 \leqq j \leqq c$, then it readily follows from (5) that the m-vector $\boldsymbol{\pi}$, which satisfies $\boldsymbol{\pi}A=\boldsymbol{\pi}$, $\boldsymbol{\pi}\mathbf{e}=1$, is given by

$$\boldsymbol{\pi}=\boldsymbol{\theta}(1)\otimes\ldots\otimes\boldsymbol{\theta}(c). \tag{8}$$

By using elementary formulas, proved in [5], we may also express the vector $\boldsymbol{\beta}^*=\sum_{k=1}^{\infty} kA_k\mathbf{e}$, explicitly in terms of the data of the model. We then easily verify that

$$\boldsymbol{\pi}\boldsymbol{\beta}^*=\lambda' \sum_{j=1}^{c} \mu'^{-1}(j). \tag{9}$$

As shown in Chap. 1 of [8], the Markov chain $\tilde{P}$ is positive recurrent if and only if $\boldsymbol{\pi}\boldsymbol{\beta}^*>1$, or equivalently

$$\lambda'^{-1} < \sum_{j=1}^{c} \mu'^{-1}(j). \tag{10}$$

The arrival rate λ'^{-1} to the queue must be less than the combined service rate of the c servers. This intuitive equilibrium condition may also be proved by applying the main theorem in Lavenberg [3].

The invariant probability vector $\mathbf{x}$ of $\tilde{P}$ is now partitioned into vectors $\mathbf{x}_E$, $\mathbf{x}_{c-1}$, $\mathbf{x}_c$, $\mathbf{x}_{c+1}, \ldots$, where the vectors $\mathbf{x}_i$, $i \geqq c-1$, are m-vectors and the vector $\mathbf{x}_E$ is of dimension card (E).

It then follows from general results, proved in [7] or [8], that

$$\mathbf{x}_i=\mathbf{x}_{c-1}R^{i-c+1}, \qquad \text{for } i \geqq c-1, \tag{11}$$

where the *positive* matrix R is the minimal nonnegative solution to the nonlinear matrix equation

$$R=\sum_{k=0}^{\infty} R^k A_k. \tag{12}$$

The vectors $\mathbf{x}_E$ and $\mathbf{x}_{c-1}$ are determined, up to a multiplicative constant, by solving a homogeneous system of linear equations. That constant is determined by use of the normalizing equation

$$\mathbf{x}_E\mathbf{e}+\mathbf{x}_{c-1}(I-R)^{-1}\mathbf{e}=1. \tag{13}$$

The spectral radius $\eta = \text{sp}(R)$, is the unique solution in (0, 1) of the equation

$$z = X(z), \tag{14}$$

where $X(z)$ is the spectral radius of $A^*(z)$. The matrix R is also the unique nonnegative solution of spectral radius less than one to the Eq. (12).

2. Preliminary Results

The probability distribution of phase type with irreducible representation $[\boldsymbol{\beta}(j), S(j)]$ has a rational Laplace-Stieltjes transform, given by

$$\phi_j(s) = \boldsymbol{\beta}(j)[sI - S(j)]^{-1}\mathbf{S}^0(j), \qquad \text{for Re } s \geqq 0. \tag{15}$$

Let the abscissa of convergence of $\phi_j(s)$ be $-\tau_j < 0$. The function $\phi_j(s)$ is then defined, positive and convex decreasing on the interval $(-\tau_j, \infty)$.

Lemma 1. *The equation*

$$z\phi_j(s) = 1, \tag{16}$$

has a unique real solution $s_j = \psi_j(z)$, *for every* z *in* $(0, 1]$. *The function* $\psi_j(\cdot)$ *satisfies* $-\tau < \psi_j(z) \leqq 0$, *and is strictly increasing on* $(0, 1]$. *Moreover* $\psi_j(0+) = -\tau_j$, *and* $\psi_j'(1) = \mu'^{-1}(j)$.

The quantity $\psi_j(z)$ *is the eigenvalue of maximal real part of the matrix* $S(j) + z\mathbf{S}^0(j)\boldsymbol{\beta}(j)$. *The corresponding left eigenvector* $\mathbf{u}(j, z)$, *normalized by* $\mathbf{u}(j, z)\mathbf{e} = 1$, *is given by*

$$\begin{aligned} \mathbf{u}(j, z) &= z(z-1)^{-1}\psi_j(z)\boldsymbol{\beta}(j)[\psi_j(z)I - S(j)]^{-1}, && \text{for } 0 < z < 1, \\ &= \boldsymbol{\theta}(j), && \text{for } z = 1. \end{aligned} \tag{17}$$

Proof. Essentially the same results were proved in [11]. For easy of reference, we repeat the proof. The first set of properties of $\psi_j(z)$ follow readily from consideration of the graph of $\phi_j(s)$ and from the Eq. (16). The equation

$$\mathbf{u}(j, z)[S(j) + z\mathbf{S}^0(j)\boldsymbol{\beta}(j)] = \psi_j(z)\mathbf{u}(j, z), \tag{18}$$

leads to

$$\mathbf{u}(j, z) = z[\mathbf{u}(j, z)\mathbf{S}^0(j)]\boldsymbol{\beta}(j)[\psi_j(z)I - S(j)]^{-1}.$$

Postmultiplication by $\mathbf{S}^0(j)$ leads to $z\phi_j[\psi_j(z)] = 1$. The inner product $\mathbf{u}(j, z)\mathbf{S}^0(j)$ does not vanish, since the matrix $\psi_j(z)I - S(j)$ is nonsingular for $0 < z \leqq 1$.

The vector $\mathbf{u}(j, z) = \int_0^\infty \exp[-\psi_j(z)t] \cdot \boldsymbol{\beta}(j)\exp[S(j)t]\,dt$, is positive, since the vector $\boldsymbol{\beta}(j)\exp[S(j)t]$ is positive for $t > 0$, as was shown in [6]. This implies that the eigenvalue $\psi_j(z)$ of the irreducible stable matrix $S(j) + z\mathbf{S}^0(j)\boldsymbol{\beta}(j)$, is the eigenvalue of maximal real part. The normalization $\mathbf{u}(j, z)\mathbf{e} = 1$, readily yields (17).

Lemma 2. *The maximal eigenvalue* $\chi(z)$ *of* $A^*(z)$ *is given by*

$$\chi(z)=f\left[-\sum_{j=1}^{c}\psi_j(z)\right], \quad \text{for } 0<z\leqq 1, \tag{19}$$

where $f(\cdot)$ *is the Laplace-Stieltjes transform of the interarrival time distribution* $F(\cdot)$*. The corresponding eigenvector* $\mathbf{u}(z)$ *is given by*

$$\mathbf{u}(z)=\mathbf{u}(1,z)\otimes\ldots\otimes\mathbf{u}(c,z). \tag{20}$$

Proof. The vector $\mathbf{u}(z)$ is clearly positive and satisfies $\mathbf{u}(z)\mathbf{e}=1$. It readily follows from (18) that

$$\mathbf{u}(j,z)\exp\{[S(j)+z\mathbf{S}^0(j)\boldsymbol{\beta}(j)]\,t\}=\exp[\psi_j(z)\,t]\,\mathbf{u}(j,z),$$

and hence by (5), that

$$\mathbf{u}(z)A^*(z)=f\left[-\sum_{j=1}^{c}\psi_j(z)\right]\mathbf{u}(z).$$

This clearly implies (19) and completes the proof.

Let now η be the unique solution in (0, 1) of the Eq. (14). The vector $\mathbf{u}(\eta)=\mathbf{u}(1,\eta)\otimes\ldots\otimes\mathbf{u}(c,\eta)$, is then given by

$$\mathbf{u}(j,\eta)=\eta(\eta-1)^{-1}\psi_j(\eta)\,\boldsymbol{\beta}(j)[\psi_j(\eta)I-S(j)]^{-1}, \quad \text{for } 1\leqq j\leqq c.$$

As shown in Chap. 1 of [8], the vector $\mathbf{u}(\eta)$ is also the left eigenvector of the matrix R, corresponding to its Perron eigenvalue η.

3. Asymptotic Behavior of the Queue Length Density

Theorem 1. *The stationary density of the queue length at arrivals satisfies*

$$\sum_{i=k}^{\infty}\mathbf{x}_i\mathbf{e}=(1-\eta)^{-1}(\mathbf{x}_{c-1}\mathbf{z})\eta^{k-c+1}+o(\eta^k), \quad \text{as } k\to\infty, \tag{21}$$

where $\mathbf{z}$ *is the* right *eigenvector of* R*, corresponding to the eigenvalue* η *and satisfies* $\mathbf{u}(\eta)\mathbf{z}-1$.

Proof. Let us write $\mathbf{u}$ for $\mathbf{u}(\eta)$. A classical property of irreducible, nonnegative matrices now yields that

$$R^i=\eta^i\mathbf{z}\mathbf{u}+o(\eta^i), \quad \text{as } i\to\infty.$$

Since

$$\sum_{i=k}^{\infty}\mathbf{x}_i\mathbf{e}=\mathbf{x}_{c-1}R^{k-c+1}(I-R)^{-1}\mathbf{e},$$

Formula (21) readily follows.

Remark. We clearly also have

$$\mathbf{x}_i=\mathbf{x}_{c-1}R^{i-c+1}=\eta^{i-c+1}(\mathbf{x}_{c-1}\mathbf{z})\,\mathbf{u}+o(\eta^i), \quad \text{as } i\to\infty. \tag{22}$$

This implies that

$$\frac{\mathbf{x}_i}{\mathbf{x}_i\mathbf{e}} \to \mathbf{u}, \quad \text{as } i \to \infty. \tag{23}$$

The left hand side is the conditional probability density of the c-tuple of service phases, given that the arriving customer finds i customers in the queue. For large i, we find that the joint conditional contribution of the c residual service times has the *limit distribution* with c-fold Laplace-Stieltjes transform

$$\begin{aligned} h(s_1, \ldots, s_c) &= \mathbf{u} \cdot \{[s_1 I - S(1)]^{-1} \mathbf{S}^0(1) \otimes \ldots \otimes [s_c I - S(c)]^{-1} \mathbf{S}^0(c)\} \\ &= \prod_{j=1}^{c} \mathbf{u}(j, \eta)[s_j I - S(j)]^{-1} \mathbf{S}^0(j). \end{aligned}$$

We see that for large i, the conditional residual service times in the c servers, given that the arriving customer finds i customers in the system, are *approximately independent*. The marginal distribution of the residual service time for the j-th server is then approximately the PH-disribution with representation $[\mathbf{u}(j, \eta), S(j)]$, for $1 \leqq j \leqq c$.

4. Asymptotic Behavior of the Stationary Waiting Time Distribution at Arrivals

Once the vector $\mathbf{x}_{c-1}$ and the matrix R are known, the probability distribution $W(\cdot)$ of the waiting time at arrivals under the first-come, first-served discipline, may be computed by solving a finite, highly structured system of differential equations with constant coefficients. This will be shown by generalizing the proof, given for the case of the GI/PH/1 queue, in [8] or [9]. For multi-server queues, the practical utility of this result is severely limited by the dimension m. The proof of this result is however essential to the derivation of the desired asymptotic formula for $W(\cdot)$.

Let the matrices C and D be defined by

$$\begin{aligned} C &= S(1) \oplus S(2) \oplus \ldots \oplus S(c), \\ D &= \mathbf{S}^0(1)\,\boldsymbol{\beta}(1) \oplus \mathbf{S}^0(2)\,\boldsymbol{\beta}(2) \oplus \ldots \oplus \mathbf{S}^0(c)\,\boldsymbol{\beta}(c), \end{aligned} \tag{24}$$

where the symbol $\oplus$ is the *Kronecker sum* of matrices [1]. The distribution $W(\cdot)$ may be viewed as the distribution of the time till absorption into the state $*$ in the Markov process with generator Q, given by

$$Q = \begin{array}{c|cccccc} & * & \mathbf{c} & \mathbf{c+1} & \mathbf{c+2} & \mathbf{c+3} & \ldots \\ \hline * & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots \\ \mathbf{c} & D\mathbf{e} & C & 0 & 0 & 0 & \ldots \\ \mathbf{c+1} & \mathbf{0} & D & C & 0 & 0 & \ldots \\ \mathbf{c+2} & \mathbf{0} & 0 & D & C & 0 & \ldots \\ \mathbf{c+3} & \mathbf{0} & 0 & 0 & D & C & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{array} \tag{25}$$

in which the initial probability vector is given by

$$\mathbf{x}_E\mathbf{e}+\mathbf{x}_{c-1}\mathbf{e},\ \mathbf{x}_{c-1}R,\ \mathbf{x}_{c-1}R^2,\ \ldots.$$

It is readily seen that the Laplace-Stieltjes transform $w(s)$ of $W(\cdot)$ is given by

$$\begin{aligned} w(s)&=\mathbf{x}_E\mathbf{e}+\mathbf{x}_{c-1}\sum_{i=0}^{\infty}R^i[(sI-C)^{-1}D]^i\mathbf{e} \\ &=\mathbf{x}_E\mathbf{e}+\mathbf{x}_{c-1}\Psi^*(s)\,\mathbf{e}, \end{aligned} \tag{26}$$

where $\Psi^*(s)$ is the square matrix of order m, which satisfies

$$\Psi^*(s)=I+R\,\Psi^*(s)(sI-C)^{-1}D. \tag{27}$$

The Eq. (27) is now transformed in the same manner as discussed in [9]. If $\Psi(\cdot)$ is the matrix of mass functions with Laplace-Stieltjes transform $\Psi^*(s)$ and $\psi(x)$ is the m^2-vector obtained by forming the direct sum of rows of $\Psi(x)$, then we derive from (27) that

$$\psi(x)=\mathbf{v}-\mathbf{v}(I\otimes C+R^T\otimes D)^{-1}\{I\otimes I-\exp[(I\otimes C+R^T\otimes D)x]\}(R^T\otimes D), \tag{28}$$

for $x\geqq 0$. The vector $\mathbf{v}$ is the m^2-vector obtained by forming the direct sum of the identity matrix. R^T is the transpose of the matrix R.

We now set $\mathbf{v}^0=-\mathbf{v}(I\otimes C+R^T\otimes D)^{-1}$, and

$$\boldsymbol{\Theta}(x)=\mathbf{v}^0\exp[(I\otimes C+R^T\otimes D)\,x], \qquad \text{for } x\geqq 0.$$

The $m\times m$ matrices V^0 and $\Theta(x)$ have the vectors $\mathbf{v}^0$ and $\boldsymbol{\Theta}(x)$ as the direct sums of their respective rows. They satisfy the equations

$$V^0C+RV^0D=-I, \tag{29}$$

and

$$\Theta'(x)=\Theta(x)C+R\,\Theta(x)D, \qquad \Theta(0)=V^0, \tag{30}$$

for $x\geqq 0$.

By virtue of (28), the matrix $\Psi(x)$ is then given by

$$\Psi(x)=I+RV^0D-R\,\Theta(x)D, \qquad \text{for } x\geqq 0. \tag{31}$$

The distribution $W(\cdot)$ is given by

$$W(x)=\mathbf{x}_E\mathbf{e}+\mathbf{x}_{c-1}\mathbf{e}+\mathbf{x}_{c-1}RV^0D\mathbf{e}-\mathbf{x}_{c-1}R\Theta(x)D\mathbf{e}, \tag{32}$$

for $x\geqq 0$. This expression may be further simplified. We post-multiply in (29) by $\mathbf{e}$ and note that $C\mathbf{e}+D\mathbf{e}=\mathbf{0}$. This yields that $V^0D\mathbf{e}=(I-R)^{-1}\mathbf{e}$. Upon substitution into (32). we readily obtain

$$W(x)=1-\mathbf{x}_{c-1}R\,\Theta(x)\,D\mathbf{e}, \qquad \text{for } x\geqq 0. \tag{33}$$

We see that the probability distribution $W(\cdot)$ may, in principle, be computed by first evaluating the matrix V^0 and then solving the matrix-differential Eq.

(30). In order to obtain the asymptotic formula, we need a number of preliminary lemmas.

Lemma 3. *The matrix* $C+\eta D$, *given by*

$$C+\eta D=[S(1)+\eta \mathbf{S}^0(1)\boldsymbol{\beta}(1)]\oplus\ldots\oplus[S(c)+\eta\mathbf{S}^0(c)\boldsymbol{\beta}(c)], \tag{34}$$

is an irreducible, stable matrix. Its eigenvalue $-\xi$ *of maximal real part is given by*

$$-\xi=\sum_{j=1}^{c}\psi_j(\eta). \tag{35}$$

The corresponding left eigenvector is given by $\mathbf{u}=\mathbf{u}(\eta)$. *The corresponding right eigenvector* $\mathbf{u}^0$, *normalized by* $\mathbf{u}\mathbf{u}^0=1$, *is given by the Kronecker product* $\mathbf{u}^0=\mathbf{u}^0(1)\otimes\ldots\otimes\mathbf{u}^0(c)$, *where*

$$\mathbf{u}^0(j)=\frac{\eta-1}{\eta\psi_j(\eta)\cdot\boldsymbol{\beta}(j)[\psi_j(\eta)I-S(j)]^{-2}\mathbf{S}^0(j)}\cdot[\psi_j(\eta)I-S(j)]^{-1}\mathbf{S}^0(j), \tag{36}$$

for $1\leqq j\leqq c$.

Proof. Since each of the matrices $S(j)+\eta\mathbf{S}^0(j)\boldsymbol{\beta}(j)$, $1\leqq j\leqq c$, is an irreducible stable matrix, so is the matrix $C+\eta D$, [11]. The matrix $C+\eta D$ is the sum of c matrices of the form

$$I\otimes\ldots\otimes I\otimes[S(j)+\eta\mathbf{S}^0(j)\cdot\boldsymbol{\beta}(j)]\otimes I\otimes\ldots\otimes I.$$

This readily yields, by (18), that

$$\mathbf{u}(C+\eta D)=\sum_{j=1}^{c}\psi_j(\eta)\cdot\mathbf{u}.$$

The vector $\mathbf{u}^0(j)$ is clearly a right eigenvector of $S(j)+\eta\mathbf{S}^0(j)\cdot\boldsymbol{\beta}(j)$, corresponding to $\psi_j(\eta)$. Furthermore $\mathbf{u}(j,\eta)\mathbf{u}^0(j)=1$. Since

$$\mathbf{u}\mathbf{u}^0=\prod_{j=1}^{c}\mathbf{u}(j,\eta)\mathbf{u}^0(j)=1,$$

the proof is complete.

Lemma 4. *The matrix* $I\otimes C+R^T\otimes D$ *has nonnegative off-diagonal elements and is irreducible. Its eigenvalue of maximal real part is* $-\xi$. *The corresponding left and right eigenvectors are respectively given by* $\mathbf{z}^T\otimes\mathbf{u}$ *and* $\mathbf{u}^T\otimes\mathbf{u}^0$. *Their inner product is one.*

Proof. The off-diagonal elements of $I\otimes C+R^T\otimes D$ are clearly nonnegative. The irreducibility of the matrix $I\otimes C+R^T\otimes D$ follows from the positivity of R and the irreducibility of the representation of the service time distributions.

We have

$$\begin{aligned}&(\mathbf{z}^T\otimes\mathbf{u})(I\otimes C+R^T\otimes D)\\ &\quad=\mathbf{z}^T\otimes\mathbf{u}C+\eta\mathbf{z}^T\otimes\mathbf{u}D=\mathbf{z}^T\otimes\mathbf{u}(C+\eta D)=-\xi(\mathbf{z}^T\otimes\mathbf{u}),\end{aligned}$$

and a similar calculation for the right eigenvector. Since both eigenvectors are positive, $-\xi$ is the eigenvalue of maximal real part [2].

Lemma 5. *The vector* $\mathbf{v}^0 = -\mathbf{v}(I \otimes C + R^T \otimes D)^{-1}$, *satisfies*

$$\mathbf{v}^0(\mathbf{u}^T \otimes \mathbf{u}^0) = \xi^{-1}. \tag{37}$$

Also

$$\mathbf{u}D\mathbf{e} = \xi(1-\eta)^{-1}. \tag{38}$$

Proof. It follows from the definition of $\mathbf{v}^0$ that

$$\mathbf{v}^0(\mathbf{u}^T \otimes \mathbf{u}^0) = -\mathbf{v}(I \otimes C + R^T \otimes D)^{-1}(\mathbf{u}^T \otimes \mathbf{u}^0) = \xi^{-1}\mathbf{v}(\mathbf{u}^T \otimes \mathbf{u}^0).$$

However

$$\mathbf{v}(\mathbf{u}^T \otimes \mathbf{u}^0) = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m] \begin{bmatrix} u_1 \mathbf{u}^0 \\ u_2 \mathbf{u}^0 \\ \vdots \\ u_m \mathbf{u}^0 \end{bmatrix} = \sum_{v=1}^{m} u_v(\mathbf{e}_v \mathbf{u}^0) = \mathbf{u}\mathbf{u}^0 = 1.$$

The vectors $\mathbf{e}_v$ are the m unit-vectors of dimension m. This proves Formula (37).

A typical term of $D\mathbf{e}$ is the Kronecker product $\mathbf{e} \otimes \mathbf{e} \otimes \ldots \otimes \mathbf{e} \otimes \mathbf{S}^0(j) \otimes \mathbf{e} \otimes \ldots \otimes \mathbf{e}$. Premultiplication by $\mathbf{u} = \mathbf{u}(1) \otimes \ldots \otimes \mathbf{u}(c)$ yields

$$\begin{aligned} \mathbf{u}(j)\mathbf{S}^0(j) &= \eta(\eta-1)^{-1}\psi_j(\eta)\boldsymbol{\beta}(j)[\psi_j(\eta)I - S(j)]^{-1}\mathbf{S}^0(j) \\ &= (\eta-1)^{-1}\psi_j(\eta), \end{aligned}$$

so that

$$\mathbf{u}D\mathbf{e} = (\eta-1)^{-1} \sum_{j=1}^{c} \psi_j(\eta) = \xi(1-\eta)^{-1}.$$

Theorem 2. *The waiting time distribution* $W(\cdot)$ *satisfies*

$$1 - W(x) = \eta(1-\eta)^{-1}(\mathbf{x}_{c-1}\mathbf{z})e^{-\xi x} + o(e^{-\xi x}), \quad \text{as } x \to \infty. \tag{39}$$

Proof. It follows from the definition of the vector $\boldsymbol{\theta}(x)$ and the properties of the matrix $I \otimes C + R^T \otimes D$, that

$$\begin{aligned} \boldsymbol{\theta}(x) &= \mathbf{v}^0[(\mathbf{u}^T \otimes \mathbf{u}^0) \cdot (\mathbf{z}^T \otimes \mathbf{u})]\,\mathbf{e}^{-\xi x} + o(e^{-\xi x}) \\ &= \xi^{-1}(\mathbf{z}^T \otimes \mathbf{u})e^{-\xi x} + o(e^{-\xi x}), \quad \text{as } x \to \infty. \end{aligned}$$

The vector $\mathbf{z}^T \otimes \mathbf{u}$ is the direct sum of the rows of the matrix $\mathbf{z} \cdot \mathbf{u}$. The preceding formula may therefore be equivalently written as

$$\Theta(x) = \xi^{-1}(\mathbf{z} \cdot \mathbf{u})e^{-\xi x} + o(e^{-\xi x}), \quad \text{as } x \to \infty.$$

Substitution into (33) yields

$$1 - W(x) = \xi^{-1}(\mathbf{x}_{c-1}R\mathbf{z})(\mathbf{u}D\mathbf{e})e^{-\xi x} + o(e^{-\xi x}), \quad \text{as } x \to \infty.$$

Since $R\mathbf{z} = \eta\mathbf{z}$, and by using Formula (38), we readily obtain (39).

Remark. Results, similar to those in Theorems 1 and 2, may be proved for the stationary distributions of the queue length and the waiting time at an arbitrary time. The proofs proceed along the same lines as in the single server case, discussed in [9]. The same decay parameters η and ξ are obtained, but the multiplicative constants are different.

5. Computational Procedure and Applications

The decay parameters η and ξ may be computed together by elementary algorithms. There are various alternative methods. It is advisible to solve the equation

$$z=f\left[-\sum_{j=1}^{c} \psi_j(z)\right], \tag{40}$$

for η in $(0,1)$ by a method which does not involve derivatives. The secant or bisection methods may be implemented with equal ease.

At each stage of the computation, we have two values z_1 and z_2 satisfying

$$0<z_1<f\left[-\sum_{j=1}^{c} \psi_j(z_1)\right]<\eta<f\left[-\sum_{j=1}^{c} \psi_j(z_2)\right]<z_2<1,$$

since the right hand side of (40) is increasing. As the next trial value z' is obtained, either by bisection or the secant method, the corresponding values $\psi_j(z')$. $1 \leqq j \leqq c$, are computed by solving the equations

$$z\boldsymbol{\beta}(j)[\psi_j(z)I-S(j)]^{-1}\mathbf{S}^0(j)=1, \tag{41}$$

for their unique solutions in the intervals $(-\tau_j, 0)$, $1 \leqq j \leqq c$. One clearly only solves those equations which are actually different. The monotonicity properties of the $\psi_j(z)$, proved in Lemma 1 are useful in solving the Eq. (41). When the interval (z_1, z_2), which brackets η is sufficiently small, we evaluate a final value $\hat{\eta}$, which is the computed value of η. The computed value $\hat{\xi}$ of ξ is obtained by setting $\hat{\xi}=-\sum_{j=1}^{c} \psi_j(\hat{\eta})$.

For many PH-distributions of interest, the Laplace-Stieltjes transform is, of course, explicitly available, so that the Eq. (41) can then be written in a computationally more convenient form.

Except for queues with a very small number of servers and then only for PH-distributions with few phases, the computation of the matrix R, and hence of the vectors $\mathbf{x}_{c-1}$ and $\mathbf{z}$, is not practically feasible. Even without explicit knowledge of the constant $\mathbf{x}_{c-1}\mathbf{z}$, the asymptotic results of Theorems 1 and 2 have practical uses.

With η and ξ so easily computable, these results may be used to test the merits of simulation procedures for the queue length and waiting times in multiserver queues. The estimates of $-\log[1-W(x)]$, for example, should for large x lie approximately on a straight line of slope ξ. Assuming that the simulation procedure can correctly identify the parameter ξ, it should also be sufficiently accurate to provide a good estimate for the intercept of the linear asymptote of $-\log[1-W(x)]$. We will then have an estimate of $\mathbf{x}_{c-1}\mathbf{z}$, which may be used in the asymptotic formulas to provide estimates of tail probabilities for the queue length and waiting time.

As a point of theoretical interest, it appears likely that the asymptotic results of Theorems 1 and 2 remain valid for the GI/G/c queue with heterogeneous servers, provided that each of the c service time distributions have a Laplace-Stieltjes transform with a negative abscissa of convergence. This may probably be proved by appropriate continuity arguments and the approximation of the service time distributions by PH-distributions. This matter, as well as the applications to simulation methodology, will be taken up elsewhere.

References

1. Bellman, R.: Introduction to Matrix Analysis. New York: McGraw Hill 1960
2. Gantmacher, F.R.: The Theory of Matrices. New York: Chelsea 1959
3. Lavenberg, S.S.: Stability and Maximum Departure Rate of Certain Open Queueing Networks having Finite Capacity Constraints. R.A.I.R.O. Informatique/Computer Science **12**, 353–370 (1978)
4. Neuts, M.F.: Probability Distributions of Phase Type. Liber Amicorum Prof. Emeritus H. Florin. Dept. Math., Univ. Louvain, Belgium, 173–206 (1975)
5. Neuts, M.F.: Renewal Processes of Phase Type. Naval. Res. Logist. Quart. **25**, 445–454 (1978)
6. Neuts, M.F.: Markov Chains with Applications in Queueing Theory, which have a Matrix-geometric Invariant Probability Vector. Advances in Appl. Probability **10**, 185–212 (1978)
7. Neuts, M.F.: The Probabilistic Significance of the Rate Matrix in Matrix-geometric Invariant Vectors. J. Appl. Probability **17**, 291–296 (1980)
8. Neuts, M.F.: Matrix-geometric Solutions in Stochastic Models. An Algorithmic Approach. Baltimore, MD.: Johns Hopkins University Press 1981
9. Neuts, M.F.: Stationary Waiting Time Distributions in the GI/PH/1 Queue. J. Appl. Probability **18** (1981)
10. Takahashi, Y., Takami, Y.: A Numerical Method for the Steady-state Probabilities of a GI/G/c Queueing System in a General Class. J. Operations Res. Soc. Japan **19**, 147–157 (1976)
11. Takahashi, Y.: Asymptotic Exponentiality of the Tail of the Waiting Time Distribution in a PH/PH/c Queue. Advances in Appl. Probability **13** (1981)

Received September 5, 1980