ALGORITHMIC METHODS FOR PH/PH/1 QUEUES WITH BATCH ARRIVALS OR SERVICES

Yutaka Baba
Tokyo Institute of Technology

(Received December 23, 1981; Revised October 13, 1982)

Abstract We discuss algorithms for the computation of the steady state features of PH/PH/1 queues with bounded batch arrivals or batch services. Many characteristic quantities such as the mean and variance of queue length in continuous time, the mean waiting time, the waiting probability, etc. are obtained in computationally tractable forms. We also show various numerical values.

1. Introduction

In most of all studies on the queueing systems with batch arrivals or services, it has been assumed that the interarrival time distribution or the service time distribution was exponential. Then we shall consider here a batch arrival queue denoted by PH\(X\)/PH/1, in which the distributions of interarrival and service times are of phase type. We only discuss the case that consecutive batch sizes are independent, and have the common probability \(\{g_i\}\), \(1 \leq i \leq K\), with mean \(g\), where \(K\) is the largest index for which \(g_K > 0\).

And we also consider a batch service queue denoted by PH/PH\(Y\)/1. The batch service discipline is one introduced by Bailey [1]. A server can serve up to \(K\) customers simultaneously. At each service epoch, the first \(K\) customers in the queue may be received their services. If the queue length is shorter than \(K\), all customers in the queue are served.

Phase type distributions were introduced by Neuts [6]. A probability distribution on \((0,\infty)\) is said to be of phase type if it is obtained as the distribution of the time till absorption in a finite Markov chain with continuous parameter. The class of phase type distributions includes a number
of well-known and important distributions such as generalized Erlang and hyperexponential distributions appeared in the queueing theory, and it is dense in the class of all probability distributions concentrated on \((0,\infty)\). So the classes of models of types \(PH[X]/PH/1\) and \(PH/PH[Y]/1\) queues are also tractable for numerical analysis.

Recently Neuts [9] showed that an important class of infinite stochastic matrices had an invariant probability vector of a matrix-geometric form. The models discussed in this paper have this property. Based on Neuts's result, we shall obtain many characteristic quantities such as the mean and variance of queue length in continuous time, mean waiting time, waiting probability, etc. These quantities are obtained in computationally tractable forms. We also show various numerical examples and obtain qualitative insight into the effect of varying the parameter values.

2. Preliminaries

In this section, we prepare some notations and basic properties related to matrices and phase type distributions.

We denote by \(I\) the identity matrix, by \(O\) the zero matrix, by \(0\) the zero vector, and by \(e\) the column vector with all entries equal to 1.

In the later discussions we will frequently use Kronecker products of matrices (e.g., Bellman [2]). Let \(X = (x_{ij})\) and \(Y = (y_{ij})\) be \(m \times n\) and \(m_y \times n_y\) matrices respectively. Then we denote by \(X \otimes Y\) the Kronecker product of \(X\) and \(Y\), which is an \((m \times m_y) \times (n \times n_y)\) matrix whose entry in the \(((i-1)m_y + k)\)th row and in the \(((j-1)n_y + l)\)th column is \(x_{ij}y_{kl}\).

Now we examine some basic properties of a phase type distribution \(F(x)\). Suppose that \(F(x)\) is represented by a continuous time Markov chain on the state space \(\{0,1,2,\ldots,m\}\) with the initial probability vector

\[
\alpha^o = (0, \alpha) = (0, \alpha_1, \alpha_2, \ldots, \alpha_m)
\]

and the infinitesimal generator

\[
Q^o = \begin{pmatrix}
0 & 0 \\
\alpha & Q
\end{pmatrix} =
\begin{pmatrix}
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
t_1 & q_{11} & q_{12} & \cdots & q_{1m} \\
t_2 & q_{21} & q_{22} & \cdots & q_{2m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
t_m & q_{m1} & q_{m2} & \cdots & q_{mm}
\end{pmatrix}.
\]

State 0 is the single absorbing state and states \(\{1,2,\ldots,m\}\) are tran-
sient (or nonrecurrent) states. All the entries of $a^0$ and off-diagonal entries of $Q^0$ are nonnegative and $Q^0 + T^0 = 0$.

The distribution function $F(x)$ is written as

$$F(x) = 1 - \exp(Qx)^{a^0}, \quad \text{for } x \geq 0.$$  

The pair $(a, Q)$ is called a representation of $F(*)$. We can show that the moments $\mu_k$, $k \geq 1$, about the origin all exist and are given by the formula

$$\mu_k = (-1)^k k! \exp(Q)^{-1} a^0, \quad \text{for } k \geq 1.$$  

Examples:

(a) For the exponential distribution with parameter $\lambda$, the matrix $Q$ is given by

$$Q = -\lambda \quad \text{and} \quad a_1 = 1,$$

so that $F(*)$ has the simple representation $(1, -\lambda)$.

(b) The generalized Erlang distribution obtained by the convolution of $m$ exponential distributions with parameters $\lambda_1, \ldots, \lambda_m$ has as one of its representations the pair $(a, Q)$ given by

$$a = (1, 0, \ldots, 0)$$

$$Q = \begin{bmatrix}
-\lambda_1 & \lambda_1 & 0 & \cdots & 0 \\
0 & -\lambda_2 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & -\lambda_m
\end{bmatrix}$$

with $T^0 = (0, \ldots, 0, \lambda_m)^T$.

(c) The hyperexponential distribution

$$F(x) = \sum_{i=1}^{m} \alpha_i (1 - \exp(-\lambda_i x)), \quad x \geq 0,$$

may be represented by $a = (\alpha_1, \ldots, \alpha_m)$ and $Q = \text{diag}(-\lambda_1, -\lambda_2, \ldots, -\lambda_m)$, with $T^0 = (\lambda_1, \ldots, \lambda_m)^T$.

For any representation $(a, Q)$, we can define a matrix $Q^*$ by

$$Q^* = Q + T^0 a^0,$$

where $T^0 = (T^0, \ldots, T^0)$ and $A^0 = \text{diag}(a_1, \ldots, a_m)$.

The matrix $Q^*$ may be considered to be the infinitesimal generator of an
\( m \)-state Markov chain, which has a close relationship to the probability distribution \( F(\cdot) \).

The significance of \( Q^* \) is as follows. At any time that an absorption occurs in the Markov chain \( Q^* \), we instantaneously perform a multinomial trial with outcomes \( 1, \ldots, m \) and probabilities \( \alpha_1, \ldots, \alpha_m \) to pick a new initial state. Considering each absorption into the state \( 0 \) as a renewal, we obtain a renewal process for which the time between any two successive renewals has cdf \( F(\cdot) \), the phase type distribution given by (2.3). Such a renewal process is called a renewal process of phase type (Neuts [8]).

The above procedure also constructively defines a new Markov chain with state space \( \{1, \ldots, m\} \), initial probability vector \( a \) and infinitesimal generator \( Q^* \). This Markov chain describes the "phase" of the system. In Neuts [6], it is shown that one may, without loss of generality, assume that the representation \((\alpha, Q)\) of \( F(\cdot) \) is so chosen as to make \( Q^* \) irreducible, and we shall henceforth assume that this is indeed the case.

We let \( \pi \) denote the stationary vector of the Markov chain \( Q^* \), i.e., the unique (strictly positive) vector satisfying \( \pi Q^* = 0 \), \( \pi e = 1 \). It may be easily shown that \( \pi = -\lambda a \mu^{-1} \), where \( \mu^{-1} = -\alpha \mu^{-1} e \) is the mean of \( F(\cdot) \).

3. PH/PH/1 queue with batch arrivals

Consider the \( PH^{[X]}/PH/1 \) model, in which customers arrive in batches. We shall assume that the probability distribution \( F(\cdot) \) of the batch inter-arrival times is of phase type with a representation \((\alpha, T)\), where \( T \) is a matrix of order \( M \) and that the probability distribution \( G(\cdot) \) of the service times is of phase type with a representation \((\beta, S)\), where \( S \) is a matrix of order \( N \). Moreover we put that \( T^o = -Te \), \( T^o = (T^o, \ldots, T^o) \), \( \alpha^o = \text{diag} (\alpha_1, \ldots, \alpha_M) \), \( S^o = -Se \), \( S^o = (S^o, \ldots, S^o) \), and \( \Sigma^o = \text{diag} (\beta_1, \ldots, \beta_N) \).

We assume that consecutive batch sizes are independent and have the common probability function \( \{g_i ; 1 \leq i \leq K\} \), with mean \( g \), where \( K \) is the largest index for which \( g_K > 0 \).

In this section, we discuss the stationary distributions of the number of customers in the system at an arbitrary epoch \( t \). In order to obtain the stationary distribution of the \( PH^{[X]}/PH/1 \) queue, we shall consider the continuous parameter Markov chain which represents the transitions of the states of the states of the \( PH^{[X]}/PH/1 \) queue at arbitrary epoch. In order to define the states of this Markov chain, we shall define that the index \( i \) (\( i \geq 1 \)) denotes the number of customers in the system, the index \( j \)
(1 ≤ j ≤ N) represents the phase of service process, and the index \( h \)
(1 ≤ h ≤ M) represents the phase of arrival process.

Then the \( PH[X]/PH/1 \) queue may be studied in terms of a continuous
parameter Markov chain with state space \((\mathcal{Q}_j)\), where \( 0 = \{(0, h), 1 \leq h \leq M\} \)
and \( \mathcal{I} = \{(i, j, h), i \geq 1, 1 \leq j \leq N, 1 \leq h \leq M\} \). The infinitesimal generator
of the Markov chain is given by the matrix

\[
Q = \begin{pmatrix}
0 & 1 & 2 & \cdots & \kappa & \kappa+1 & \kappa+2 & \kappa+3 & \cdots \\
\tilde{C}_0 & g_1 \tilde{D}_0 & g_2 \tilde{D}_0 & \cdots & g_k \tilde{D}_0 & 0 & 0 & 0 & \cdots \\
\tilde{B}_0 & \tilde{C}_0 & g_1 \tilde{D}_1 & \cdots & g_k \tilde{D}_1 & g_{k+1} \tilde{D}_1 & 0 & 0 & \cdots \\
\tilde{B}_1 & \tilde{C}_1 & g_1 \tilde{D}_1 & \cdots & g_k \tilde{D}_1 & g_{k+1} \tilde{D}_1 & g_{k+2} \tilde{D}_1 & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\end{pmatrix}
\]

where \( \tilde{B}_0 = S^0 \otimes I, \tilde{C}_0 = S^0 \otimes I, \tilde{C}_0 = T, \tilde{C}_0 = T \otimes I + I \otimes S, \tilde{D}_0 = \tilde{B} \otimes T^0 A^0, \)
and \( \tilde{B} = I \otimes T^0 A^0 \). \( I \) is the identity matrix of order \( N \) when it is in the
right of the symbol \( \otimes \) and the identity matrix of order \( M \) when it appears
to the left. We see that the dimensions of the matrices \( \tilde{B}_0, \tilde{C}_0, \tilde{D}_0 \) are
\( MN \times M, M \times M, M \times MN \), respectively. All other blocks in \( Q \) are square
and of order \( MN \). It is easily verified that the \( PH[X]/PH/1 \) queue is
stable if and only if \( \tilde{B}_0 S^{-1}_0 < \mu_1 \), where \( \mu_1^{-1} = -\tilde{C}_0 T^{-1} \tilde{C}_0 \) and \( \mu_1^{-1} = -\tilde{B} S^{-1} \).

We denote the invariant probability vector of \( Q \) by \( \pi \), which we may
write in the partitioned form \( \{x_0, x_1, \ldots, x_i, \ldots\} \), where \( x_0 \) is \( M \)-vector
and \( x_i, i \geq 1 \), are \( MN \)-vectors.

The component \( x_{0,h} \) of \( x_0 \) is the stationary probability that, at an
arbitrary time epoch \( t \), the queue is empty and the arrival process is in
the phase state \( h \), 1 ≤ h ≤ M. The component \( x_{i,j,h} \) of \( x_i \),
\( i \geq 1 \), is the stationary probability that there are \( i \) customers in the
system, the service process is in the \( j \)-th phase, 1 ≤ j ≤ N, and the arrival
process is in the \( h \)-phase, 1 ≤ h ≤ M at \( t \).

Let \( \tau \) be the real number such that

\[
\tau = \max \{ \max_{1 \leq i \leq M} \{ -\tilde{C}_0 ii \}, \max_{1 \leq i \leq MN} \{ -\tilde{C}_0 ii \} \} > 0,
\]

then the discrete parameter Markov chain with transition matrix
where $C_0 = I + \tau^{-1}C\tau$, $C = I + \tau^{-1}C$, $B_0 = \tau^{-1}B\tau$, $B = \tau^{-1}B$, $D_0 = \tau^{-1}D\tau$, and $D = \tau^{-1}D$, has the same invariant probability as the continuous time Markov chain whose infinitesimal generator is $Q$. In this paper, we shall analyze about the discrete parameter Markov chain with transition matrix $P$.

Neuts [9] shows that a class of infinite, block-partitioned, stochastic matrices has a matrix-geometric invariant probability vector of the form $(u_0, u_1, \ldots)$, where $u_k = u_0^k$, for $k \geq 0$. The rate matrix $R$ is an irreducible, nonnegative matrix of spectral radius less than one. The spectral radius $sp(R)$ of $R$ is the maximum absolute value of the eigenvalues of $R$. The matrix $R$ is the minimal solution, in the set of nonnegative matrices of spectral radius at most one, of a nonlinear matrix equation. We shall show that the stochastic matrix $P$ in (3.3) has a matrix-geometric invariant vector.

Let us define some matrices and vectors given by

\begin{align*}
F &= C_0, \\
G &= (g_1^D, \ldots, g_k^D), \\
E_0 &= \begin{pmatrix} 0 & 0 & \cdots & 0 & B \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \\
E &= \begin{pmatrix} 0 & 0 & \cdots & 0 & B \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \\
F &= \begin{pmatrix} C & g_1^D & \cdots & g_k^D \\ B & C & \cdots & \vdots \\ 0 & B & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C \end{pmatrix}, \\
G &= \begin{pmatrix} g_1^D & 0 & \cdots & 0 \\ g_2^D & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_k^D & g_{k-1}^D & \cdots & g_k^D \end{pmatrix}, \\
\bar{y}_0 &= \bar{x}_0, \\
\bar{y}_i &= (\bar{x}_{(i-1)k+1}, \ldots, \bar{x}_{ik}) \quad (i \geq 1).
\end{align*}
Then the stochastic matrix $P$ is written by

$$
P = \begin{pmatrix}
F_0 & G_0 & 0 & 0 & \ldots \\
0 & F & G & 0 & \ldots \\
0 & 0 & F & G & 0 & \ldots \\
& & & & \ddots & \ddots & \ddots \\
& & & & & \ddots & \ddots & \ddots \\
& & & & & & \ddots & \ddots & \ddots \\
& & & & & & & \ddots & \ddots & \ddots \\
& & & & & & & & \ddots & \ddots & \ddots \\
\end{pmatrix}
$$

(3.5)

The stochastic matrix $P$ is modified block tri-diagonal. We can see that the dimensions of the matrices $F_0, G_0, E_0$ are $M \times M$, $M \times MN$, $MN \times M$, respectively, the matrices $E, F, G$ are square and of order $MN$, $y_0$ is a $M$-vector and $y_i$ ($i \geq 1$) are $MNK$-vectors. Then the steady state equations with respect to $P$ are given by

$$
y_0 = y_0F_0 + y_1E_0 
$$

(3.6)

$$
y_1 = y_0G_0 + y_1F + y_2E 
$$

$$
y_i = y_{i-1} + y_iF + y_{i+1}E 
$$

($i \geq 2$).

We define the sequence of matrices $\{R(n), n \geq 0\}$ by

$$
R(0) = G, \quad R(n+1) = G + R(n)F + [R(n)]^2 E. 
$$

(3.7)

Neuts [9,12] showed that if the stochastic matrix $H = E + F + G$ is irreducible, the sequence $\{R(n)\}$ converges to a matrix $R \geq 0$, which satisfies

$$
sp(R) < 1, \quad R = G + RF + R^2 E. 
$$

(3.8)

The invariant vector $y_i$ ($i \geq 1$) satisfies

$$
y_{i+1} = y_iR 
$$

($i \geq 1$).

Therefore the invariant vector $y_0$ and $y_1$ will be got if we can solve the linear system of equations given by

$$
y_0 = y_0F_0 + y_1E_0 
$$

(3.10)

$$
y_1 = y_0G_0 + y_1(F + RE) 
$$

$$
y_0 + y_1(I - R)^{-1}e = 1.
$$

Remark

Since $sp(R) < 1$, the inverse matrix $(I - R)^{-1}$ exists. Further, the matrix $(I - R)^{-1}$ is strictly positive for $R$ is irreducible.
Next we shall consider the steady state equations with respect to \( x_i \) \((i \geq 0)\). These equations are given by

\[
\begin{align*}
    x_0 &= x_0 C_0 + x_1 B_0 \\
    x_1 &= x_0 g_1 D_0 + x_1 C + x_2 B \\
    x_2 &= x_0 g_2 D_0 + x_1 g_1 D + x_2 C + x_3 B \\
    &\vdots \\
    x_K &= x_0 g_K D_0 + \sum_{i=1}^{K-1} x_i g_{K-i} D + x_K C + x_{K+1} B \\
    x_{K+j} &= \sum_{i=1}^{K} x_{i+j-1} g_{K+i-j} D + x_{K+j} C + x_{K+j+1} B \quad (i \geq 1).
\end{align*}
\]

(3.11)

In order to get the moments of system size, we introduce the vector generating function \( \bar{x}(z) = \sum_{i=1}^{\infty} x_i z^i \) \(|z| \leq 1\). Multiplying both hands in (3.11) by \( z^i \) and summing over \( i \geq 1 \), we can easily obtain that \( \bar{x}(z) \) satisfies the equation

\[
\begin{align*}
    \bar{x}(z) &= x_0 D_0 \sum_{i=1}^{K} g_i z^i + \sum_{i=2}^{\infty} z^i \sum_{j=1}^{i} x_{i-j} g_j D + x(z) C \\
    &\quad + z^{-1}[\bar{x}(z) - x_1 z]B.
\end{align*}
\]

(3.12)

Rearranging (3.12), we have

\[
\begin{align*}
    \bar{x}(z)[zI - A(z)] &= x_0 D_0 \sum_{i=1}^{K} g_i z^{i+1} - zx_1 B, \\
    \text{where } A(z) &= B + zC + \sum_{i=1}^{K} g_i z^{i+1} D.
\end{align*}
\]

(3.13)

In deriving the moments of system size, some explicit expressions of the derivatives of the Perron-Frobenius eigenvalues and the associated eigenvalues of the matrix \( A(z) \) may be necessary.

For \(|z| \leq 1\), the matrix \( A(z) \) has the Perron-Frobenius eigenvalue \( \delta(z) \) uniquely. Let \( u(z) \) and \( v(z) \) be the corresponding right and left eigenvectors such as

\[
[A(z) - \delta(z)I]u(z) = v(z)[A(z) - \delta(z)I] = 0.
\]

(3.14)

The following relations are assumed to be hold.
(3.15) \( v(z)u(z) = v(z)e = 1 \), \( v(1) = \pi \), and \( u(1) = e \),

where \( \pi \) is the steady state vector of the stochastic matrix \( A(1) \).

We denote by \( A(j)(z) \) the matrix obtained by differentiating \( j \) times each entry of \( A(z) \).

We have to prepare the following.

Theorem 3.1 \cite{5}. The derivatives \( \delta(j)(1), u(j)(1), v(j)(1), j \geq 0, \) may be computed recursively for each \( j \). These recursion formulae are follows.

\[
\delta(0)(1) = 1, \quad u(0)(1) = e, \quad v(0)(1) = \pi
\]

(3.16) \[
\begin{align*}
u(1)(1) &= [I - A(1) + \Pi]^{-1}[A(1)(1) - \delta(1)(1)I]e \\
v(1)(1) &= \pi[A(1)(1) - \delta(1)(1)I][I - A(1) + \Pi]^{-1}
\end{align*}
\]

and for \( j \geq 2 \)

\[
\begin{align*}
\delta(j)(1) &= \sum_{i=0}^{j} \binom{j}{i} \pi A(i)(1) u(j-i)(1) - \sum_{i=1}^{j-1} \binom{j}{i} \delta(i)(1) \pi u(j-i)(1) \\
\delta^2(j)(1) &= \sum_{i=0}^{j-1} \binom{j}{i} \nu(i)(1) A(j-i)(1) \nu(j-i)(1) e \\
\nu(j)(1) &= \sum_{i=0}^{j-1} \binom{j}{i} \nu(i)(1) [A(j-i)(1) - \delta(j-i)(1)I][I - A(1) + \Pi]^{-1},
\end{align*}
\]

where \( \Pi \) is the square matrix of order \( MN \), and each row of it is \( \pi \). \( \square \)

We want to derive the first two moments of the system size, i.e., we shall deduce the formulae for computing \( X(1)(1)e \) and \( X(2)(1)e \). Let us recall the equation (3.13):

\[
K(z)[zI - A(z)] = x_0 P_0 \sum_{i=1}^{K} g_i z^{i+1} - z K_1 B.
\]

If let \( z \) tend to \( 1 \), we get

(3.18) \( K(1)[I - A(1)] = x_0 P_0 - K_1 B. \)

Adding \( K(1)\Pi = (K(1)e)\pi = (1 - x_0 e)\pi \) to both sides of the equation (3.18) and recognizing that \( \pi[I - A(1) + \Pi]^{-1} = \pi \), we have

(3.19) \( K(1) = [x_0 P_0 - K_1 B][I - A(1) + \Pi]^{-1} + (1 - x_0 e)\pi. \)
Differentiating (3.13) once with respect to \( z \) yields

\[
(3.20) \quad \frac{\partial}{\partial z} \left[ x(z) \right] = x_0 D_0 \sum_{i=1}^{K} (i+1)g_i z^i - x_1 B.
\]

Hence as \( z \to 1 \), we get

\[
(3.21) \quad \frac{\partial}{\partial z} \left[ x(1) \right] = x_0 D_0 \sum_{i=1}^{K} (i+1)g_i - x_1 B.
\]

We note that \( I - A(1) \) is singular but \( I - A(1) + \Pi \) is nonsingular and \( x(1) \Pi = (x(1) \phi) \Pi \). Therefore (3.21) becomes

\[
(3.22) \quad x(1) = -x(1) \left[ I - C - \sum_{i=1}^{K} (i+1)g_i D \right] + x_0 D_0 \sum_{i=1}^{K} (i+1)g_i u(1) (1)
\]

Using (3.19) and (3.22), we have the following theorem.

Theorem 3.2. The first and second moments of the system size are given by

\[
(3.23) \quad x(1) = -x(1) u(1) + \frac{1}{2[1 - \delta(1)(1)]} \delta(2)(1) x(1) \\
+ x_0 D_0 \sum_{i=1}^{K} (i+1)g_i e + 2x_0 D_0 \sum_{i=1}^{K} (i+1)g_i u(1) (1)
\]

\[
+ x_0 D_0 u(2) (1) - 2x_1 Bu(1) (1) - x_1 Bu(2) (1)
\]

\[
(3.24) \quad x(2) = -x(1) u(2) (1) - 2x(1) u(1) (1)
\]

\[
+ \frac{1}{3[1 - \delta(1)(1)]} \left[ 3x(1) \delta(2)(1) + 3x(1) u(1) \delta(2)(1) \right]
\]

\[
+ x(1) \delta(3)(1) + \sum_{i=1}^{K} (i+1)i(i-1)g_i x_0 D_0 e
\]

\[
+ 3 \sum_{i=1}^{K} (i+1)i g_i x_0 D_0 u(1) (1) + 3 \sum_{i=1}^{K} (i+1)g_i x_0 D_0 u(2) (1)
\]

\[
+ x_0 D_0 u(3) (1) - 3x_1 Bu(2) (1) - x_1 Bu(3) (1)
\].

Proof: Multiplying (3.13) on the right by \( u(z) \) we get

\[
(3.25) \quad [z - \delta(z)]x(z) u(z) = x_0 D_0 \sum_{i=1}^{K} g_i z^{i+1} u(z) - x_1 Bu(z).
\]
Differentiating this with respect to $z$ and rearranging the results,

\begin{equation}
(3.26) \quad \frac{X^{(1)}}{u}(z) = -\frac{X(z)u^{(1)}}{u} + \frac{1}{z-\delta(z)} \left[-l - \delta^{(1)}(z)\right] X(z)u(z) \\
+ x_0 \sum_{i=1}^{K} (i+1) g_i z^i u(z) + x_0 \sum_{i=1}^{K} g_i z^i u(z) \\
- x_1 B u(z) - z x_1 B u^{(1)}(z).
\end{equation}

Letting $z \rightarrow 1$ and using L'Hopital's rule, we have (3.23) after some tedious computations.

Differentiating (3.25) twice with respect to $z$ and rearranging the terms, we get

\begin{equation}
(3.27) \quad \frac{X^{(2)}}{u}(z) = -\frac{X(z)u^{(2)}}{u} - 2 \frac{X(z)u^{(1)}}{u} (z) + \frac{1}{z-\delta(z)} \delta^{(2)}(z) X(z)u(z) \\
- 2 \delta^{(1)}(z) X(z)u(z) - 2 \delta^{(1)}(z) X(z)u^{(1)}(z) \\
+ x_0 \sum_{i=1}^{K} (i+1) g_i z^i u^{(1)}(z) + 2 x_0 \sum_{i=1}^{K} (i+1) g_i z^i u(2)(z) \\
+ x_0 \sum_{i=1}^{K} g_i z^i u^{(2)}(z) - 2 x_1 B u^{(1)}(z) - z x_1 B u^{(2)}(z).
\end{equation}

Letting $z \rightarrow 1$ and using L'Hopital's rule, we have (3.24) after tedious computations.

The higher factorial moments of the system size can be computed in the same manner but the formulae become uninspiringly complicated. Thus they will not be shown here.

We denote the steady state vector of $Q^* = T + T A^*$ by $\gamma$. Let $W_1$ and $W$ be the time until the first customer of the batch and an arbitrary customer of the batch enter the service, respectively. Then we obtain the next theorem.

**Theorem 3.3**

\begin{equation}
(3.28) \quad P(W_1 > 0) = 1 - \frac{x_0 z^0}{\gamma z^0}.
\end{equation}

\begin{equation}
(3.29) \quad P(W > 0) = 1 - \frac{x_0 z^0}{\gamma z^0 g}.
\end{equation}

**Proof:** Since $\gamma T z^0 dt$ is the probability that in the steady state an arrival of the batch occurs in $[t, t+dt]$ and $x_0 z^0$ the probability that an arrival of the batch occurs in $[t, t+dt]$ and finds no customers in the system, we have $P(W_1 > 0) = x_0 z^0 / \gamma z^0$. Therefore (3.28) holds.
$P(\text{an arbitrary customer in the batch is the first customer of the batch})$
$= 1/g$, so $P(W = 0) = P(W_1 = 0)/g = T^0 / g$. Hence we obtain (3.29).

From Theorems 3.2 and 3.3, we can get many characteristic quantities such as the mean queue size and the mean system size in continuous time, variance of queue size and variance of system size in continuous time, mean waiting time, mean system time, waiting probability of the first customer of the batch and an arbitrary customer of the batch, and the probability that the server is busy in continuous time. For example, the mean system time is obtained from (3.23) and Little's formula. We calculated mean waiting times for various interarrival time distribution, service time distribution, and arrival batch size. In Table 3.1 and Fig. 3.1, we shall show the mean waiting time for constant batch size $K$ ($K \geq 1$). This model is denoted by $PH[K]/PH/1$. In Fig. 3.2, we also illustrate the waiting probability of an arbitrary arriving customer.

Table 3.1 and Fig. 3.1 show that if we fix the interarrival and the service time distributions then the larger the arrival batch size, the longer the mean waiting time. These also show that the larger the variation of coefficient of the interarrival or the service time distribution, the longer the mean waiting time, if $K$ and the traffic intensity $\rho$ are fixed. It was also found from Fig. 3.2 that the waiting probability of an arbitrary customer depends on the interarrival time distribution, the service time distribution, or the batch size $K$, even if $\rho$ is fixed.

Table 3.1  (Mean waiting time)/(Mean service time) for $PH[K]/PH/1$

<table>
<thead>
<tr>
<th></th>
<th>$K=1$</th>
<th>$K=2$</th>
<th>$K=4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.3 0.6 0.9</td>
<td>0.3 0.6 0.9</td>
<td>0.3 0.6 0.9</td>
</tr>
<tr>
<td>$M/M/1$</td>
<td>0.429 1.500 9.000</td>
<td>1.143 2.750 14.000</td>
<td>2.571 5.250 24.000</td>
</tr>
<tr>
<td>$M/E_2/1$</td>
<td>0.321 1.125 6.750</td>
<td>1.036 2.375 11.750</td>
<td>2.464 4.875 21.750</td>
</tr>
<tr>
<td>$E_2/M/1$</td>
<td>0.217 0.984 6.588</td>
<td>0.762 1.761 9.218</td>
<td>1.866 3.336 14.499</td>
</tr>
<tr>
<td>$E_2/E_2/1$</td>
<td>0.131 0.631 4.356</td>
<td>0.683 1.418 6.999</td>
<td>1.791 3.001 12.289</td>
</tr>
<tr>
<td>$H_2/H_2/1$</td>
<td>0.643 2.250 13.500</td>
<td>1.357 3.500 18.500</td>
<td>2.786 6.000 28.500</td>
</tr>
<tr>
<td>$H_2/M/1$</td>
<td>0.602 2.202 13.484</td>
<td>1.425 4.021 22.813</td>
<td>3.063 7.621 41.425</td>
</tr>
<tr>
<td>$H_2/H_2/1$</td>
<td>0.860 3.018 18.048</td>
<td>1.698 4.880 27.427</td>
<td>3.347 8.523 46.091</td>
</tr>
</tbody>
</table>

$H_2$ represents a hyperexponential distribution with two phases and its variation coefficient is $\sqrt{2}$.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.
\( \rho \) is the traffic intensity and is given by \( \rho = \frac{K \lambda_1}{\mu_1} \).

Fig. 3.1 (Mean waiting time)/(Mean service time) for \( PH[K]/PH/1 \)

Fig. 3.2 The waiting probability of an arbitrary customer for \( PH[K]/PH/1 \)
4. PH/PH/1 queue with batch services

Consider the PH/PH[\(Y\)]/1 model, in which customers are served in batches. We assume that the interarrival time distribution \( F(\cdot) \) is of phase type with a representation \((\mathbf{0}, T)\), where \( T \) is a matrix of order \( M \), and the service time distribution \( G(\cdot) \) is of phase type with a representation \((\mathbf{B}, S)\), where \( S \) is a matrix of order \( N \). Moreover we put that \( T^0 = -T_0 \), \( T^0 = (T^0, \ldots, T^0) \), \( A^0 = \text{diag}(\alpha_1, \ldots, \alpha_M) \), \( S^0 = -S_0 \), \( S^0 = (S^0, \ldots, S^0) \) and \( S^0 = \text{diag}(\beta_1, \ldots, \beta_N) \). In our batch service discipline, a server can serve up to \( K \) customers simultaneously. Hence at each service initiation the first \( K \) customers from the queue receive their services. If the queue length is shorter than \( K \), the entire queue is served. This model is denoted by the notation \( \text{PH/PH}[K]/1 \).

The \( \text{PH/PH}[K]/1 \) queue may be studied in terms of a continuous parameter Markov chain with state space \((Q \cup i^*)\), where \( \bar{Q} = \{(0,h), 1 \leq h \leq M\} \) and \( i^* = \{(i,j,h), i \geq 0, 1 \leq j \leq N, 1 \leq h \leq M\} \). The index \( i \) stands for the queue size (excluding the customer in service), and \( 0 \) is the state that the server is idle. The infinitesimal generator of the Markov chain is given by the matrix

\[
Q = \begin{pmatrix}
0 & 0^* & 1^* & 2^* & \ldots & (K-1)^* & K^* & (K+1)^* & \ldots \\
0 & \tilde{C}_0 & \tilde{b}_0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots \\
0^* & \tilde{b}_0 & \tilde{C} & \tilde{b} & 0 & \ldots & 0 & 0 & 0 & \ldots \\
1^* & 0 & \tilde{b} & \tilde{C} & \tilde{b} & 0 & \ldots & 0 & 0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
(K-1)^* & 0 & \tilde{b} & 0 & \tilde{C} & \tilde{b} & 0 & \ldots & 0 & \ldots \\
K^* & 0 & \tilde{b} & 0 & \tilde{C} & \tilde{b} & \ldots & 0 & \ldots & \ldots \\
(K+1)^* & 0 & 0 & \tilde{b} & 0 & \tilde{C} & \tilde{b} & 0 & \ldots & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{pmatrix}
\]

(4.1)

where \( \tilde{b}_0 = S^0 \otimes I \), \( \tilde{b} = S^0 \otimes \tilde{I} \), \( \tilde{C}_0 = T \), \( \tilde{C} = T \otimes I + I \otimes S \), \( \tilde{b}_0 = \tilde{b} \otimes T^0A^0 \), and \( \tilde{b} = I \otimes T^0A^0 \). It is easily verified that the \( \text{PH/PH}[K]/1 \) queue is stable if and only if \( \lambda_1 < K\mu_1 \), where \( \lambda_1^{-1} = -\alpha T^{-1} e \) and \( \mu_1^{-1} = -\beta S^{-1} e \).

We denote the invariant probability vector of \( Q \) by \( \mathbf{x} \). We may write it in the partitioned form \( \{x_0, x_0^*, \ldots, x_i^*, \ldots\} \), where \( x_0^* \) is an \( M \)-vector and \( x_i^*, i \geq 0 \), are \( MN \)-vectors. Analogously to section 3, we construct the discrete parameter Markov chain with a transition matrix
\[
\begin{pmatrix}
0 & 0^* & 1^* & 2^* & \ldots & (K-1)^* & K^* & (K+1)^* & \ldots \\
0 & C & D & 0 & \ldots & 0 & 0 & 0 & \ldots \\
0^* & B & C & D & \ldots & 0 & 0 & 0 & \ldots \\
1^* & 0 & B & C & D & \ldots & 0 & 0 & 0 & \ldots \\
(K-1)^* & 0 & B & 0 & 0 & \ldots & C & D & 0 & \ldots \\
K^* & 0 & B & 0 & 0 & \ldots & 0 & C & D & \ldots \\
(K+1)^* & 0 & 0 & B & 0 & \ldots & 0 & 0 & C & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{pmatrix}
\]

where \( \tau = \max \{ \max_1 \{-\tilde{c}_{0,i}\}, \max_{1 \leq i \leq MN} \{-\tilde{c}_{i,i}\} \} \), \( C_0 = I + \tau^{-1} \tilde{c}_0 \), \( C = I + \tau^{-1} \tilde{c} \), \( B_0 = \tau^{-1} \tilde{b}_0 \), \( B = \tau^{-1} \tilde{b} \), \( D_0 = \tau^{-1} \tilde{d}_0 \) and \( D = \tau^{-1} \tilde{d} \). This Markov chain has the same invariant probability vector as the continuous parameter Markov chain whose infinitesimal generator is \( Q \). Therefore the steady state equations with respect to \( P \) are given by

\[
\begin{align*}
X_0 &= X_0 C_0 + X_0 B_0 \\
X_0^* &= X_0 D_0 + X_0^* C + \sum_{i=1}^{K} X_i^* B \\
X_j^* &= X_{j-1}^* D + X_{j-1}^* C + X_{j+K}^* B \\ & \quad (j \geq 1) \\
X_0^e + \sum_{i=0}^{\infty} X_i^e &= 1 .
\end{align*}
\]

We define the sequence of matrices \( \{R(n), n \geq 0\} \) by

\[
R(0) = D , \quad R(n+1) = D + R(n) C + [R(n)]^{K+1} B.
\]

Neuts [9,12] shows that if the stochastic matrix \( E = B + C + D \) is irreducible, the sequence \( \{R(n)\} \) converges to a matrix \( R \geq 0 \), which satisfies

\[
sp(R) < 1 , \quad R = D + RC + R^{K+1} B
\]

and invariant vector \( X_{-1}^* \), \( i \geq 0 \), satisfies

\[
X_{i+1}^* = X_i^* R \quad (i \geq 0) .
\]

Therefore, in order to get the invariant vector \( X_0 \) and \( X_0^* \), it is only requested to solve the linear system of equations given by

\[
X_0 = X_0 C_0 + X_0^* B_0
\]
(4.7) \[ x_0^* = x_0^0 + x_0^* (c + \sum_{i=1}^{K} R_i B) \]
\[ x_0^e + x_0^* (I - R)^{-1} e = 1. \]

If we introduce the vector generating function \( X(z) = \sum_{i=0}^{\infty} x_i^* z^i \) (\( |z| \leq 1 \)), we have

(4.8) \[ X(z) = \sum_{i=0}^{\infty} x_0^* R^i z^i = x_0^* (I - Rz)^{-1} \]
using (4.6). From (4.8), we obtain the next theorem.

**Theorem 4.1.** The first and second factorial moments of the queue size (excluding the customer in service) are given by

(4.9) \[ X^{(1)}(z) = x_0^* (I - R)^{-1} R(I - Rz)^{-1} \]
(4.10) \[ X^{(2)}(z) = 2 x_0^* [(I - R)^{-1} R]^2 (I - Rz)^{-1}. \]

**Proof:** Differentiating (4.8) with respect to \( z \) once and twice, we have

(4.11) \[ X^{(1)}(z) = x_0^* (I - Rz)^{-1} R(I - Rz)^{-1} \]
and

(4.12) \[ X^{(2)}(z) = 2 x_0^* [(I - Rz)^{-1} R]^2 (I - Rz)^{-1}. \]

Letting \( z \rightarrow 1 \) and multiplying \( e \) on the right, we obtain (4.9) and (4.10).

Similarly to Theorem 3.3, we have the waiting probability of an arriving customer as follows.

**Theorem 4.2.**

(4.13) \[ P(W > 0) = 1 - \frac{x_0^* e}{\gamma T^*}, \]
where \( W \) is the waiting time of an arbitrary customer and \( \gamma \) is the steady state vector of infinitesimal generator \( Q^* = T + T^* A^0 \).

**Proof:** The proof of this theorem is similar to Theorem 3.3, so we omit the proof.

From Theorems 4.1 and 4.2, we have many characteristic quantities such as mean queue size and variance of queue size in continuous time, mean waiting time, mean system time, waiting probability of an arbitrary arriving customer, and the probability that the server is busy. Mean waiting time is obtained from (4.9) and Little's formula. We calculated mean waiting

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.
times for various interarrival time distribution, service time distribution, and service batch size in Table 4.1. We also illustrate mean waiting times in Fig. 4.1 and waiting probability in Fig. 4.2 for $E_2/E_2[K]/1$ and $H_2/H_2[K]/1$ queueing models, where $H_2$ represents the hyperexponential distribution whose number of phases is two.

Table 4.1 and Fig. 4.1 demonstrate many interesting results. For example, in Fig. 4.1, the larger the variation coefficients of the interarrival and the service time distributions, the longer the mean waiting time. It is also found that, in the light traffic case, the larger the service batch size, the longer the mean waiting time, but, in the heavy traffic case, the larger the service batch size, the shorter the mean waiting time, if the interarrival time distribution, the service time distribution, and the traffic intensity $\rho$ are fixed. In Fig. 4.2, it is found that the waiting probability depends on the service batch size. However, in the light traffic case, the larger the variation coefficients of the interarrival time distribution and the service time distribution, the larger the waiting probability, but, in the heavy traffic case, the larger the variation coefficients of the interarrival time distribution and the service time distribution, the smaller the waiting probability, if the service batch size is fixed and only when $\kappa \geq 2$.

Table 4.1 (Mean waiting time)/(Mean service time) for $PH/PH[K]/1$

<table>
<thead>
<tr>
<th></th>
<th>$\kappa=1$</th>
<th>$\kappa=2$</th>
<th>$\kappa=4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.3</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>$M/M/1$</td>
<td>0.429</td>
<td>1.500</td>
<td>9.000</td>
</tr>
<tr>
<td></td>
<td>0.620</td>
<td>1.591</td>
<td>7.311</td>
</tr>
<tr>
<td></td>
<td>0.828</td>
<td>1.600</td>
<td>6.310</td>
</tr>
<tr>
<td>$M/E_2/1$</td>
<td>0.321</td>
<td>1.125</td>
<td>6.750</td>
</tr>
<tr>
<td></td>
<td>0.456</td>
<td>1.124</td>
<td>4.942</td>
</tr>
<tr>
<td></td>
<td>0.612</td>
<td>1.077</td>
<td>3.889</td>
</tr>
<tr>
<td>$E_2/M/1$</td>
<td>0.217</td>
<td>0.984</td>
<td>6.588</td>
</tr>
<tr>
<td></td>
<td>0.471</td>
<td>0.893</td>
<td>3.780</td>
</tr>
<tr>
<td></td>
<td>0.771</td>
<td>1.487</td>
<td>5.720</td>
</tr>
<tr>
<td>$E_2/E_2/1$</td>
<td>0.131</td>
<td>0.631</td>
<td>4.356</td>
</tr>
<tr>
<td></td>
<td>0.319</td>
<td>0.893</td>
<td>3.780</td>
</tr>
<tr>
<td></td>
<td>0.569</td>
<td>0.976</td>
<td>3.306</td>
</tr>
<tr>
<td>$M/H_2/1$</td>
<td>0.643</td>
<td>2.250</td>
<td>13.500</td>
</tr>
<tr>
<td></td>
<td>0.964</td>
<td>2.564</td>
<td>12.097</td>
</tr>
<tr>
<td></td>
<td>1.310</td>
<td>2.719</td>
<td>11.215</td>
</tr>
<tr>
<td>$H_2/M/1$</td>
<td>0.602</td>
<td>2.202</td>
<td>13.484</td>
</tr>
<tr>
<td></td>
<td>0.766</td>
<td>1.986</td>
<td>9.572</td>
</tr>
<tr>
<td></td>
<td>0.911</td>
<td>1.803</td>
<td>7.458</td>
</tr>
<tr>
<td>$H_2/H_2/1$</td>
<td>0.860</td>
<td>3.018</td>
<td>18.048</td>
</tr>
<tr>
<td></td>
<td>1.140</td>
<td>2.977</td>
<td>14.358</td>
</tr>
<tr>
<td></td>
<td>2.401</td>
<td>3.912</td>
<td>13.355</td>
</tr>
</tbody>
</table>

$H_2$ represents a hyperexponential distribution with two phases and its variation coefficient is $\sqrt{2}$.

$\rho$ is the traffic intensity and is given by $\rho = \lambda_1/K\mu_1$. Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.
Fig. 4.1 (Mean waiting time)/(Mean service time) for \( PH/PH[K]/1 \)

![Graph showing the relationship between mean waiting time and service time for \( PH/PH[K]/1 \) system for different values of \( K \).]

Fig. 4.2 The waiting probability of an arbitrary customer for \( PH/PH[K]/1 \)

![Graph depicting the waiting probability for different values of \( K \) and \( \rho \).]
Acknowledgements

I wish to express to my deep appreciation to Professor Hidenori Morimura for his helpful suggestions and criticisms. I am also grateful to the referees for their valuable comments on this paper.

References


Yutaka Baba: Department of Information Sciences, Faculty of Science, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, Japan.