International Journal of Operations Research

International Journal of Operations Research Vol. 3, No. 1, 7−15 (2006)

A General Input Queue with *N* Policy and Service Rate Depending on Bulk Size

Jau-Chuan Ke[1,*] and Kuo-Hsiung Wang[2]

[1]National Taichung Institute of Technology, Department of Statistics No. 129, Sec. 3, Sanmin Rd., Taichung 404, Taiwan, Republic of China

[2]Department of Applied Mathematics, National Chung-Hsing University, Taichung 402, Taiwan, Republic of China

Received January 2005; *Accepted October 2005*

Abstract—An embedded Markov chain is used to analyze a G/M/1 queuing system with N policy. When the system is empty, the server remains idle (deactivates) and does not start serving the waiting customers in the queue until the number of arrivals reaches N. The service is performed in batches of min(n, N) if there are n customers waiting at the completion of service. Service times of the server depend on the batch sizes. We utilize the matrix-geometric method in the solution procedure and solve the stationary probabilities of the number of customers in the system by means of simultaneous linear equations. We further obtain a number of explicit and computationally tractable results such as mean queue length and mean waiting time in the queue. A numerical example illustrates the validation of the solution procedure.

Keywords—Bulk service, Bontrol policy, Queue, Waiting time distribution

1. INTRODUCTION

The N policy applied to queueing system was originally considered by Yadin and Naor (1963). The N policy M/G/1 queueing system was first studied by Heyman (1968) and was investigated by several researchers such as Bell (1971, 1972), Kimura (1981), Tijms (1986), Teghem (1987), Gakis et al. (1995), Wang and Ke (2000), and others. Recently, Wang and Yen (2003) derived the analytical steady-state results for the N policy M/H$_k$/1 queueing system.

A general bulk-service queue was first introduced by Neuts (1967). He considered an ordinary M/G$^{[a,b]}$/1 queue with bulk-service. The bulk-service rule is applied as follows; let there be n customers waiting at the completion of a service. If $0 \leq n < a$, a group of size n gets service and if $n > b$, a group of size b is served. Neuts (1979) examined model by Neuts (1967) deriving queue length distribution. Curry and Feldman (1985) treated a bulk-service M/M$^{[a,b]}$/1 queue with service rate depending on its service bulk size. They derived the distribution of the number of customers in service as well as in the queue through the matrix geometric solution procedure by Neuts (1981). Two good references on the subject are the books of Chaudhry and Templeton (1984) and Medhi (1984). Recently, Laxmi and Gupta (1999) investigated finite-buffer bulk-service G/M$^{[1,b]}$/1 queueing system using the supplementary variable technique. Baba (1996) used matrix geometric solution procedure to study the ordinary G/M$^{[1,b]}$/1 queueing system with bulk-service rule and service rate depending on its service size.

In this paper we study a bulk-service G/M/1 queueing system, in which the server operates N policy and service rates depending on his service bulk size. The results are more general than those of Baba (1996). The arrivals occur according to a renewal process with interarrival time distribution $A(t)$ of finite mean $1/\lambda$ and Laplace-Stieltjes transform (LST) $A^*(s)$. The successive interarrival time random variables are denoted by A. The server deactivates whenever the system is empty. As soon as there are N customers queued in the system, the server reactivates to begin serving the waiting customers until the system is empty. Customers are served in batches, the sizes of which are not greater than N (i.e., the server can serve in batches of min(n, N) at a time) and they are taken to services in the same order as they arrives. Service times depend on the batch sizes. The service times of successive batches are mutually independent. The distribution function of service time H_j of a batch size j $(1 \leq j \leq N)$ follows an exponential distribution with mean $1/\mu_j$. Further assume that $\mu_i \neq \mu_k$ when $i \neq k$.

As a practical application fitting our general model is the following produce to order system for a product which based on the work of Zhang et al. (2001). Assume that the interarrival times of production orders follow a general distribution. It is desirable that the production begins whenever the number of orders reaches a critical value N

*Corresponding author's email: jauchuan@ntit.edu.tw

(the minimum set-up lot size satisfied benefit). In other words, if the number of orders is less than N the production waits until the accumulated orders are reached N. The management policy is to set up the facility and begins production when there are N orders in the queue. Arriving production orders form a single waiting line at the facility based on the order of their arrival. When n orders present in the queue after beginning production, the facility processes min(n, N) orders at a time, which its production (processing) time follows an exponential distribution with rate μ_n. Whenever the production ends and no orders arrive, the production facility is shut down (turned off). The tuned-off period of the facility may be referred to machine maintenances and other secondary works. Another interesting examples of bulk service can be found in the operation of an unscheduled car ferry or a single ground floor station of an elevator, or the control of a traffic flow (Neuts, 1967).

The purpose of this paper is threefold. The first is to show the queue size and the service batch size at points of arrivals form an embedded Markov chain. The second is to present the matrix geometric form for the steady-state probabilities of this Markov chain. The third is to develop explicit analytic expressions for the steady-state queue length distribution at points of arrivals as well as the LST of the stationary waiting time distribution of an arbitrary customer.

2. EMBEDDED MARKOV CHAIN

We consider a N policy G/M/1 queueing system with service rates depending on its service batch size. Suppose that I_r denotes the queue length immediately prior to the r-th arrival and L_r the number of customers in service immediately prior to the r-th arrivals, respectively. Further, suppose that τ_r denotes the time between $(r-1)$ st and r-th arrivals. For convenience, we choose the time origin at an epoch of arrival and set $\tau_0 = 0$. Then, it is easily seen that the sequences $\{(I_r, \tau_r): 0 \le r < N\}$ and $\{(I_r, L_r, \tau_r): 0 \le r\}$ are two Markovian renewal sequences on the state spaces $\{(i,t): 0 \le i < N,\ t \ge 0\}$ and $\{(i,\ \ell,\ t): 0 \le i,\ 1 \le \ell \le N, t \ge 0\}$, respectively.

Because the service time have memoryless property, we have

$$a_j \equiv \Pr[H_j > A] = \int_0^\infty e^{-\mu_j t} dA(t) = A^*(\mu_j), \tag{1}$$

where $A^*(\mu_j)$ besides being the LST of the interarrival time, also represents the probability that a service is longer than an interarrival.

Let $b_{i,\ell,j}$ $(i = 1,\ 2,\ \ldots,\ N; \ell = 0,\ 1,\ \ldots; j = 1,\ 2,\ \ldots,\ N)$ be the probability that when the customers of batch size j are served immediately prior to an arrival, the service of batch size j finishes, the service of batch size N finishes ℓ times and the customers of batch size i are served immediately prior to next arrival. It follows from APPENDIX A that we have

Case 1: $i = j = N$,

$$b_{N,\ell-1,N} = \int_0^\infty \frac{(\mu_N t)^\ell}{\ell!} e^{-\mu_N t} dA(t), \quad \text{for } \ell \ge 1 \tag{2}$$

Case 2: $1 \le i \le N-1$ and $j = N$,

$$b_{i,\ell,N} = \int_0^\infty dA(t) \int_0^t e^{-\mu_i (t-x)} \mu_N \frac{(\mu_N x)^\ell}{\ell!} e^{-\mu_N x} dx, \quad \text{for } \ell \ge 0 \tag{3}$$

Case 3: $1 \le j \le N-1$ and $i = N$,

$$b_{N,\ell,j} = \int_0^\infty dA(t) \int_0^t e^{-\mu_j (t-x)} \mu_j \frac{(\mu_N x)^\ell}{\ell!} e^{-\mu_N x} dx, \quad \text{for } \ell \ge 0 \tag{4}$$

Case 4: $1 \le i = j \le N-1$,

$$b_{i,0,i} = \int_0^\infty \mu_i t e^{-\mu_i t} dA(t), \text{ for } \ell = 0 \tag{5}$$

$$b_{i,\ell,i} = \int_0^\infty \mu_i \mu_N e^{-\mu_i t} dA(t) \int_0^t (t-x) \frac{(\mu_N x)^{\ell-1}}{(\ell-1)\ !} e^{(\mu_i - \mu_N)x} dx, \quad \text{for } \ell \ge 1 \tag{6}$$

Case 5: $1 \le i \ne j \le N-1$,

$$b_{i,0,j} = \int_0^\infty dA(t) \int_0^t \mu_j e^{-\mu_j x} e^{-\mu_i (t-x)} dx, \text{ for } \ell = 0 \tag{7}$$

$$b_{i,\ell,j} = \int_0^\infty dA(t) \int_0^t \mu_N \frac{(\mu_N x)^{\ell-1}}{(\ell-1)!} e^{-\mu_N x} dx \times \int_0^{t-x} \mu_j e^{-\mu_j y} e^{-\mu_i (t-x-y)} dy, \text{ for } \ell \ge 1. \tag{8}$$

Equation (2) to (8) can be calculated using a readily computable form. For example, (5) is equivalent to

$$b_{i,0,i} = -\mu_i A^{*(1)}(\mu_i).$$

The transition probability matrix $\tilde{P}$ of the embedded Markov chain $\{I_r, r \ge 0\} \cup \{(I_r, L_r), r \ge 0\}$, displayed for $N = 3$, is then of the form as

$$
\tilde{P}=\begin{pmatrix}
c_0 & 0 & 0 & \alpha_2\mathbf{1} & \alpha_1\mathbf{1} & \alpha_0\mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_1 & 0 & 0 & B_{10} & A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_2 & 0 & 0 & B_{20} & 0 & A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_3 & 0 & 0 & B_{30} & 0 & 0 & A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_4 & 0 & 0 & B_{11} & B_{30} & 0 & 0 & A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_5 & 0 & 0 & B_{21} & 0 & B_{30} & 0 & 0 & A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_6 & 0 & 0 & B_{31} & 0 & 0 & B_{30} & 0 & 0 & A & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_7 & 0 & 0 & B_{12} & B_{31} & 0 & 0 & B_{30} & 0 & 0 & A & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_8 & 0 & 0 & B_{22} & 0 & B_{31} & 0 & 0 & B_{30} & 0 & 0 & A & 0 & 0 & 0 & 0 & \cdots \\
c_9 & 0 & 0 & B_{32} & 0 & 0 & B_{31} & 0 & 0 & B_{30} & 0 & 0 & A & 0 & 0 & 0 & \cdots \\
c_{10} & 0 & 0 & B_{13} & B_{32} & 0 & 0 & B_{31} & 0 & 0 & B_{30} & 0 & 0 & A & 0 & 0 & \cdots \\
c_{11} & 0 & 0 & B_{23} & 0 & B_{32} & 0 & 0 & B_{31} & 0 & 0 & B_{30} & 0 & 0 & A & 0 & \cdots \\
c_{12} & 0 & 0 & B_{33} & 0 & 0 & B_{32} & 0 & 0 & B_{31} & 0 & 0 & B_{30} & 0 & 0 & A & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
$$

We see that, in general, for the present model, the square matrices are given

$$
A=\begin{pmatrix}
a_1 & & & & & & & \\
 & a_2 & & & & & & \\
 & & a_3 & & & & & \\
 & & & a_4 & & & & \\
 & & & & \ddots & & & \\
 & & & & & a_{N-2} & & \\
 & & & & & & a_{N-1} & \\
 & & & & & & & a_N
\end{pmatrix},
$$

$$
\text{and } B_{i\ell} == \begin{array}{c} \\ \end{array}
\begin{array}{c|c}
 & \begin{matrix} 1 & 2 & \cdots & i^{th} & \cdots \end{matrix} \\
\begin{matrix} 1 \\ 2 \\ \vdots \\ \\ \\ \\ \end{matrix} & \begin{pmatrix} & & & b_{i,\ell,1} & \\ & & & b_{i,\ell,2} & \\ & & & b_{i,\ell,3} & \\ & & & b_{i,\ell,4} & \\ & & & \vdots & \\ & & & b_{i,\ell,N} & \end{pmatrix}
\end{array},
$$

where A is a $N\times N$ square matrix with $a_i = A^*(\mu_i)$ $(1\le i\le N)$, $B_{i\ell}$ $(1\le i\le N,\ \ell\ge 0)$ are $N\times N$ square matrices with nonzero elements only in the ith column, $b_{i,\ell}=(b_{i,\ell,1},b_{i,\ell,2},b_{i,\ell,3},\cdots,b_{i,\ell,N})^t$, the appropriate dimensional column vectors c_k are determined to satisfy that each row sum of $\tilde{P}$ is equal to unity and they are assumed to zero in blank of matrices. Furthermore, let $\mathbf{1}=(1,\ 0,\ 0,\ \ldots,\ 0)$ be a N-dimensional row vector and

$$
\alpha_n=\int_0^\infty \frac{(\mu_N t)^n e^{-\mu_N t}}{n!}dA(t)=\frac{(-\mu_N)^n}{n!}A^{*(n)}(\mu_N). \tag{9}
$$

Using the results of Neuts (1981), it is trivial to obtain the equilibrium condition $\rho=\lambda/(N\mu_N)<1$. In this case, the Markov chain represented by the transition matrix $\tilde{P}$ is positive recurrent.

Let $\mathbf{q}=(p_0, p_1, \ldots, p_{N-1}, \mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{N-1}, \mathbf{q}_N, \mathbf{q}_{N+1}, \ldots)$ be the steady-state probability vector of $\tilde{P}$. That is, $\mathbf{q}$ is the solution to $\mathbf{q}^{\tilde{P}}=\mathbf{q}$, $\mathbf{qe}=1$. Note that p_i $(i = 0, 1, 2, \ldots, N-1)$ are the steady-state probabilities corresponding to state i when the server is idle, and the vectors of order N, $\mathbf{q}_i$ $(i=1,\ 2,\ \cdots)$, are the steady-state probability vectors corresponding to the states $\{(i,1),\ \cdots,\ (i,N)\}$ when the server is working. $\mathbf{q}_i$ is partitioned as $\mathbf{q}_i$ = $(q_{i,1},\ q_{i,2},\ \ldots,\ q_{i,N})$.

Applying the results of Neuts (1981) and Baba (1996), when $\lambda<N\mu_N$, we have

$$
q_n=q_N \mathrm{R}^{n-N},\ \text{for } n=N, N+1, N+2,\cdots, \tag{10}
$$

where the sub rate matrix R is now the minimal nonnegative solution to the equation

$$
\mathrm{R}=A+\sum_{n=1}^{\infty}\mathrm{R}^{nN}B_{N,n-1}, \tag{11}
$$

where

$$
\mathrm{R}=\begin{pmatrix}
a_1 & & & & & & & r_1 \\
 & a_2 & & & & & & r_2 \\
 & & a_3 & & & & & r_3 \\
 & & & a_4 & & & & r_4 \\
 & & & & \ddots & & & \vdots \\
 & & & & & a_{N-2} & & r_{N-2} \\
 & & & & & & a_{N-1} & r_{N-1} \\
 & & & & & & & r_N
\end{pmatrix}, \tag{12}
$$

with r_N being the unique real root between 0 and 1 of the equation

$$z = A^*(\mu_N(1-z^N)),$$

and

$$r_i = \mu_i A^*(\mu_i)^N[A^*(\mu_i)-r_N][A^*(\mu_N(1-A^*(\mu_i)^N)) - A^*(\mu_i)] \div [\mu_i - \mu_N + \mu_N A^*(\mu_i)^N][A^*(\mu_i)-r_N + A^*(\mu_N(1-r_N^N)) - A^*(\mu_N(1-A^*(\mu_i)^N))].$$

The derivations of r_i can be referred to Baba (1996).

3. STATIONARY QUEUE LENGTH AT ARRIVAL

In this section, we derive explicit expressions for the steady-state probability vector $\mathbf{q}$ of $\tilde{P}$. There are a number of obvious simplifications, which the particular structure of $\tilde{P}$ induces into the computation of the vector $(p_0, p_1, \cdots, p_{N-1}, \mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_N)$, which is given by the system of linear equations

$$p_1 = p_2 = \cdots = p_{N-1}, \tag{13}$$

$$\begin{aligned} q_{1,1} &= p_0\alpha_{N-1} + \sum_{n=0}^{\infty} \mathbf{q}_{nN+1} b_{1,n} \\ &= p_0\alpha_{N-1} + \mathbf{q}_1 b_{1,0} + \mathbf{q}_N \sum_{n=0}^{\infty} R^{nN+1} b_{1,n+1}, \end{aligned} \tag{14}$$

$$\begin{aligned} q_{1,i} &= \sum_{n=0}^{\infty} \mathbf{q}_{nN+i} b_{i,n} \\ &= \mathbf{q}_i b_{i,0} + \mathbf{q}_N \sum_{n=0}^{\infty} R^{nN+i} b_{i,n+1}, \text{ for } 2 \le i \le N-1 \end{aligned} \tag{15}$$

$$q_{1,N} = \sum_{n=0}^{\infty} \mathbf{q}_{nN+N} b_{N,n} = \mathbf{q}_N \sum_{n=0}^{\infty} R^{nN} b_{N,n}, \tag{16}$$

$$q_{k,1} = p_0\alpha_{N-k} + q_{k-1,1}a_1, \text{ for } 2 \le k \le N \tag{17}$$

$$q_{k,i} = q_{k-1,i}a_i, \text{ for } 2 \le k \le N \text{ and } 2 \le i \le N-1 \tag{18}$$

$$\begin{aligned} q_{k,N} &= q_{k-1,N}a_N + \sum_{n=0}^{\infty} \mathbf{q}_{(n+1)N+(k-1)} b_{N,n} \\ &= q_{k-1,N}a_N + \mathbf{q}_N \sum_{n=0}^{\infty} R^{nN+(k-1)} b_{N,n}, \text{ for } 2 \le k \le N. \end{aligned} \tag{19}$$

It is to be noted that

$$b_{i,0} = (b_{i,0,1}, b_{i,0,2}, b_{i,0,3}, \ldots, b_{i,0,N}), \text{ for } i = 1, 2, \ldots, N-1, \tag{20}$$

where

$$b_{i,0,l} = \begin{cases} -\mu_i A^{*(1)}(\mu_i) & i = l, \\ \dfrac{\mu_l}{\mu_l - \mu_i}[A^*(\mu_i) - A^*(\mu_l)] & i \ne l. \end{cases}$$

Let $\mathbf{e}$ be the N dimensional column vectors with all elements equal to one and $\mathbf{I}$ be the identity matrix of order N. Using the normalizing equation, it finally yields

$$p_0 + Np_1 + \sum_{n=1}^{N-1} \mathbf{q}_n \mathbf{e} + \mathbf{q}_N (\mathbf{I} - R)^{-1} \mathbf{e} = 1. \tag{21}$$

Since the sub rate matrix R is completely determined in (12) and (13)-(21) form a system of simultaneous linear equations with $(N \times N)+1$, we can solve these equations if we can calculate

$$\varphi_i = (\varphi_{i,1}, \varphi_{i,2}, \cdots, \varphi_{i,N})^t = \sum_{n=0}^{\infty} R^{nN+i-1} b_{N,n}, \quad i = 1, 2, \cdots, N \tag{22a}$$

$$\phi_i = (\phi_{i,1}, \phi_{i,2}, \cdots, \phi_{i,N})^t = \sum_{n=0}^{\infty} R^{nN+i} b_{i,n+1}, \quad i = 1, 2, \cdots, N. \tag{22b}$$

First, we calculate φ_i. For notational convenience, we denote $A^*(\mu_k)^m = [A^*(\mu_k)]^m$, $A^{*(1)}(s) = dA^*(s)/ds$ and $r_k^{(m)}$ the (k, N) element of R^m.

(i) For $1 \le i \le N$ and $1 \le j \le N-1$, using (2) and (4), the calculations of $\varphi_{i,j}$ are given by

$$\begin{aligned} \varphi_{i,j} &= \sum_{n=0}^{\infty} [a_j^{nN+i-1} b_{N,n,j} + r_j^{(nN+i-1)} b_{N,n,N}] \\ &= \sum_{n=0}^{\infty} [a_j^{nN+i-1} \int_0^{\infty} dA(t) \int_0^t \mu_j e^{-\mu_j(t-x)} \frac{(\mu_N x)^n}{n!} e^{-\mu_N x} dx \\ &\quad + r_j \sum_{k=1}^{nN+i-1} a_j^{nN+i-1-k} r_N^{k-1} \int_0^{\infty} \frac{(\mu_N t)^{n+1}}{(n+1)!} e^{-\mu_N t} dA(t)] \\ &= \frac{\mu_j A^*(\mu_j)^{i-1}\left[A^*(\mu_j) - A^*[\mu_N(1-A^*(\mu_j)^N]\right]}{\mu_N - \mu_j - \mu_N A^*(\mu_j)^N} \\ &\quad + \frac{r_j}{A^*(\mu_j) - r_N}\Big\{A^*(\mu_j)^{i-N-1} \\ &\quad \times \left(A^*[\mu_N(1-A^*(\mu_j)^N)] - A^*(\mu_N)\right) \\ &\quad - r_N^{i-N-1}\left[A^*(\mu_N[1-r_N^N]) - A^*(\mu_N)\right]\Big\}. \end{aligned} \tag{23}$$

(ii) For $1 \le i \le N$ and $j = N$, from (2), the calculations of $\varphi_{i,N}$ are as follows

$$\begin{aligned}\varphi_{i,N} &= \sum_{n=0}^{\infty} r_N^{nN+i-1} b_{N,n,N} \\ &= \sum_{n=0}^{\infty} r_N^{nN+i-1} \int_0^{\infty} \frac{(\mu_N t)^{n+1}}{(n+1)!} e^{-\mu_N t} dA(t) \\ &= \frac{A^*[\mu_N(1-r_N^N)] - A^*(\mu_N)}{r_N^{N+1-i}}. \end{aligned} \tag{24}$$

Finally, we calculate ϕ_i.

(i) For $i = N$ and $1 \le j \le N-1$, it follows from (2) and (4) that the calculations of $\phi_{N,j}$ are given by

$$\begin{aligned}&\phi_{N,j} \\ &= \sum_{n=0}^{\infty} [a_j^{nN+N} b_{N,n+1,j} + r_j^{(nN+N)} b_{N,n+1,N}] \\ &= \sum_{n=0}^{\infty} [a_j^{nN+N} \int_0^{\infty} dA(t) \int_0^t \mu_j e^{-\mu_j(t-x)} \frac{(\mu_N x)^{n+1}}{(n+1)!} e^{-\mu_N x} dx \\ &\quad + r_j \sum_{k=1}^{nN+N} a_j^{nN+N-k} r_N^{k-1} \int_0^{\infty} \frac{(\mu_N t)^{n+2}}{(n+2)!} e^{-\mu_N t} dA(t)] \\ &= \mu_j [\frac{A^*(\mu_j) - A^*[\mu_N(1-A^*(\mu_j)^N)]}{\mu_N - \mu_j - \mu_N A^*(\mu_j)^N} \\ &\quad + \frac{A^*(\mu_N) - A^*(\mu_j)}{\mu_N - \mu_j}] + \frac{r_j}{A^*(\mu_j)^N [A^*(\mu_j) - r_N]} \\ &\quad \times [A^*(\mu_N[1-A^*(\mu_j)^N]) - A^*(\mu_N) + A^{*(1)}(\mu_N)] \\ &\quad - \frac{r_j}{r_N^N [A^*(\mu_j) - r_N]} [A^*(\mu_N[1-r_N^N]) - A^*(\mu_N) \\ &\quad + A^{*(1)}(\mu_N)]. \end{aligned} \tag{25}$$

(ii) For $i = j = N$, using (2), the calculations of $\phi_{N,N}$ are given by

$$\begin{aligned}\phi_{N,N} &= \sum_{n=0}^{\infty} r_N^{nN+N} b_{N,n+1,N} \\ &= \sum_{n=0}^{\infty} [r_N^{nN+N} \int_0^{\infty} \frac{(\mu_N t)^{n+2}}{(n+2)!} e^{-\mu_N t} dA(t)] \\ &= \frac{A^*[\mu_N(1-r_N^N)] - A^*(\mu_N) + r_N^N \mu_N A^{*(1)}(\mu_N)}{r_N^N}. \end{aligned} \tag{26}$$

(iii) For $1 \le i \le N-1$ and $j = N$, from (3), we have the calculations of $\phi_{i,N}$ as follows

$$\begin{aligned}\phi_{i,N} &= \sum_{n=0}^{\infty} r_N^{nN+i} b_{i,n+1,N} \\ &= \sum_{n=0}^{\infty} r_N^{nN+i} \int_0^{\infty} dA(t) \int_0^t e^{-\mu_i(t-x)} \mu_N \frac{(\mu_N x)^{n+1}}{(n+1)!} e^{-\mu_N x} dx \\ &= \mu_N r_N^{i-N} [\frac{A^*(\mu_i) - A^*[\mu_N(1-r_N^N)]}{\mu_N - \mu_i - \mu_N r_N^N} \\ &\quad + \frac{A^*(\mu_N) - A^*(\mu_i)}{\mu_N - \mu_i}]. \end{aligned} \tag{27}$$

(iv) For $1 \le i = j \le N-1$, since it follows from (3) and (6) that the calculations of $\phi_{i,i}$ are given by

$$\begin{aligned}&\phi_{i,i} \\ &= \sum_{n=0}^{\infty} [a_i^{nN+i} b_{i,n+1,i} + r_i^{(nN+i)} b_{i,n+1,N}] \\ &= \sum_{n=0}^{\infty} [a_i^{nN+i} \int_0^{\infty} \mu_i \mu_N e^{-\mu_i t} dA(t) \int_0^t (t-x) \frac{(\mu_N x)^n}{n!} e^{(\mu_i - \mu_N)x} dx \\ &\quad + r_i \sum_{k=1}^{nN+i} a_i^{nN+i-k} r_N^{k-1} \int_0^{\infty} dA(t) \int_0^t \mu_N e^{-\mu_i(t-x)} \frac{(\mu_N x)^{n+1}}{(n+1)!} e^{-\mu_N x} dx] \\ &= \mu_i \mu_N A^*(\mu_i) \Big[A^*(\mu_N(1-A^*(\mu_i)^N)) \\ &\quad + (\mu_i - \mu_N + \mu_N A^*(\mu_i)^N) A^{*(1)}(\mu_i) - A^*(\mu_i) \Big] \\ &\quad / [\mu_i - \mu_N + \mu_N A^*(\mu_i)^N]^2 + \frac{r_i \mu_N A^*(\mu_i)^{i-N}}{A^*(\mu_i) - r_N} \\ &\quad \times [\frac{A^*[\mu_N(1-A^*(\mu_i)^N)] - A^*(\mu_i)}{\mu_i - \mu_N + \mu_N A^*(\mu_i)^N} + \frac{A^*(\mu_N) - A^*(\mu_i)}{\mu_N - \mu_i}] \\ &\quad - \frac{r_i \mu_N r_N^{i-N}}{A^*(\mu_i) - r_N} \times [\frac{A^*[\mu_N(1-r_N^N)] - A^*(\mu_i)}{\mu_i - \mu_N + \mu_N r_N^N} \\ &\quad + \frac{A^*(\mu_N) - A^*(\mu_i)}{\mu_N - \mu_i}]. \end{aligned} \tag{28}$$

(v) For $1 \le i \ne j \le N-1$, using (3) and (8), the calculations of $\phi_{i,j}$ are given by

$$\begin{aligned}&\phi_{i,j} \\ &= \sum_{n=0}^{\infty} [a_j^{nN+i} b_{i,n+1,j} + r_j^{(nN+i)} b_{i,n+1,N}] \\ &= \sum_{n=0}^{\infty} [a_j^{nN+i} \int_0^{\infty} dA(t) \int_0^t \mu_N \frac{(\mu_N x)^n}{n!} e^{-\mu_N x} dx \\ &\quad \times \int_0^{t-x} \mu_j e^{-\mu_j y} e^{-\mu_i(t-x-y)} dy + r_j \sum_{k=1}^{nN+i} a_j^{nN+i-k} r_N^{k-1} \int_0^{\infty} dA(t) \\ &\quad \times \int_0^t \mu_N e^{-\mu_N x} \frac{(\mu_N x)^{n+1}}{(n+1)!} e^{-\mu_i(t-x)} dx] \\ &= \frac{\mu_j A^*(\mu_j)^{i-N} [A^*(\mu_i - \mu_N A^*(\mu_j)^N) - A^*(\mu_i)]}{(\mu_j - \mu_i)} \\ &\quad - \frac{\mu_j \mu_N A^*(\mu_j)^i [A^*(\mu_i - \mu_N A^*(\mu_j)^N) - A^*(\mu_j)]}{(\mu_j - \mu_i)(\mu_j - \mu_i + \mu_N A^*(\mu_j)^N)} \\ &\quad + \frac{r_j \mu_N A^*(\mu_j)^{i-N}}{A^*(\mu_j) - r_N} [\frac{A^*(\mu_i) - A^*[\mu_N(1-A^*(\mu_j)^N)]}{\mu_N - \mu_i - \mu_N A^*(\mu_j)^N}] \\ &\quad + \frac{r_j \mu_N r_N^{i-N}}{A^*(\mu_j) - r_N} [\frac{A^*(\mu_i) - A^*[\mu_N(1-r_N^N)]}{\mu_N - \mu_i - \mu_N r_N^N}]. \end{aligned} \tag{29}$$

From (23)-(29), we can easily calculate φ_i and ϕ_i

($i = 1,\ 2,\ \cdots,\ N$) by using the LST of the interarrival distribution, $A^*(\theta)$, and its first derivative, $A^{*(1)}(\theta)$, the service rate μ_i and the sub rate matrix *R*. Calculating $b_{i,\ell}$ in (20), and α_n in (9), and $\varphi_{i,j}$ and $\phi_{i,j}$ in (23)-(29), the steady-state probability distribution at points of arrivals can be determined by means of solving a linear system of equations (14)-(19) and (21).

4. SYSTEM CHARACTERISTICS

In this section, we develop the steady-state characteristics of this system, such as mean queue length, the expected waiting time in the queue, and so on.

Mean queue length

Let L_0 be the steady-state queue length at points of arrivals while the server is turned off. From (13), we have

$$E[L_0] = \sum_{n=0}^{N-1} np_n. \tag{30}$$

Let L_w be the steady-state queue length at points of arrivals while the server is working and $G(z)$ be its generating function. Using $\mathbf{q}_n$ ($n = 1,\ 2,\ \ldots$) and *R*, $G(z)$ is expressed as

$$\begin{aligned} G(z) &= \sum_{n=1}^{\infty} \mathbf{q}_n \mathbf{e} z^n \\ &= \sum_{n=1}^{N-1} \mathbf{q}_n \mathbf{e} z^n + \sum_{n=0}^{\infty} \mathbf{q}_N R^n \mathbf{e} z^{N+n} \\ &= \sum_{n=1}^{N-1} \mathbf{q}_n \mathbf{e} z^n + \mathbf{q}_N z^N (\mathbf{I} - Rz)^{-1} \mathbf{e}. \end{aligned} \tag{31}$$

The mean of L_w is given by

$$\begin{aligned} E[L_w] &= \frac{dG(z)}{dz}\Big|_{z=1} \\ &= \sum_{n=1}^{N-1} n\mathbf{q}_n \mathbf{e} + \mathbf{q}_N R(\mathbf{I} - R)^{-2} \mathbf{e} + N\mathbf{q}_N (\mathbf{I} - R)^{-1} \mathbf{e}. \end{aligned} \tag{32}$$

Let L_S be the mean queue length for the *N* policy G/M/1 queue system with bulk-services depending its service size. Using (30) and (32), it finally yields

$$\begin{aligned} E[L_s] = \sum_{n=0}^{N-1} np_n + \sum_{n=1}^{N-1} n\mathbf{q}_n \mathbf{e} + \mathbf{q}_N R(\mathbf{I} - R)^{-2} \mathbf{e} \\ + N\mathbf{q}_N (\mathbf{I} - R)^{-1} \mathbf{e}, \end{aligned} \tag{33}$$

where p_0 and $\mathbf{q}_n$ ($n = 1,\ 2,\ \cdots,\ N$) are obtained through solving linear equations of (13)-(19) and (21).

Waiting Time Distribution

Let W and $W^*(s)$ denote the steady-state waiting time and its LST, respectively.

If upon arrival a customer finds that the server is idle, then the test customer needs to wait queue length of $N-1$. Further if upon arrival a customer finds that the number of waiting customers is $iN + j$ ($i \geq 0$, $0 \leq j \leq N-1$) and the number in service is *k* then the conditional waiting time has LST as

$$\frac{\mu_k}{s+\mu_k}\left(\frac{\mu_N}{s+\mu_N}\right)^i.$$

Hence $W^*(s)$ is given by

$$\begin{aligned} W^*(s) &= \sum_{n=0}^{N-1} p_n [A^*(s)]^{N-n-1} + \sum_{i=0}^{\infty} \sum_{j=0}^{N-1} \sum_{k=1}^{N} \mathbf{q}_{iN+j,k} \\ &\quad \times \left(\frac{\mu_k}{s+\mu_k}\right)\left(\frac{\mu_N}{s+\mu_N}\right)^i \\ &= \sum_{n=0}^{N-1} p_n [A^*(s)]^{N-n-1} + \sum_{i=0}^{\infty} \sum_{j=0}^{N-1} \mathbf{q}_{iN+j} \mathbf{u}(s) \left(\frac{\mu_N}{s+\mu_N}\right)^i \\ &= \sum_{n=0}^{N-1} p_n [A^*(s)]^{N-n-1} + \sum_{j=0}^{N-1} \mathbf{q}_j \mathbf{u}(s) \\ &\quad + \sum_{i=1}^{\infty} \sum_{j=0}^{N-1} \mathbf{q}_N R^{(i-1)N+j} \mathbf{u}(s) \left(\frac{\mu_N}{s+\mu_N}\right)^i \\ &= \sum_{n=0}^{N-1} p_n [A^*(s)]^{N-n-1} + \sum_{j=0}^{N-1} \mathbf{q}_j \mathbf{u}(s) \\ &\quad + \mathbf{q}_N \left(\frac{\mu_N}{s+\mu_N}\right)\left(\mathbf{I} - \frac{\mu_N}{s+\mu_N} R^N\right)^{-1} (\mathbf{I} - R^N) \\ &\quad \times (\mathbf{I} - R)^{-1} \mathbf{u}(s), \end{aligned} \tag{34}$$

where

$\mathbf{u}(s) = (\mu_1/(s+\mu_1), \mu_2/(s+\mu_2), \cdots, \mu_N/(s+\mu_N))^t$ and $\mathbf{q}_0 = (p_0, p_1, \ldots, p_{N-1})$.

Differentiating (34) once and inserting $s = 0$, we have

$$\begin{aligned} &E[W] \\ &= \sum_{n=0}^{N-1} \frac{(N-n-1)}{\lambda} p_n + \sum_{j=0}^{N-1} \mathbf{q}_j \mathbf{u}_1 + \mathbf{q}_N \Big[\frac{1}{\mu_N} R^N (\mathbf{I} - R^N)^{-1} \\ &\quad \times (\mathbf{I} - R)^{-1} \mathbf{e} + (\mathbf{I} - R)^{-1} \mathbf{u}_1 + \frac{1}{\mu_N} (\mathbf{I} - R)^{-1} \mathbf{e}\Big], \end{aligned} \tag{35}$$

where

$$\mathbf{u}_1 = \left(\frac{1}{\mu_1}, \frac{1}{\mu_2}, \cdots, \frac{1}{\mu_N}\right)^t.$$

5. NUMERICAL EXAMPLE

As an example, suppose $N = 2$, $\lambda = 0.8$, $\mu_1 = 1.0$,

$\mu_2 = 2.0$, and $\mu_3 = 3.0$.

From the definition of a_j in (1) and (9), we get

$$a_1 = \frac{\lambda}{\lambda+\mu_1} = 0.44444,\ a_2 = \frac{\lambda}{\lambda+\mu_2} = 0.28571,$$
$$a_3 = \frac{\lambda}{\lambda+\mu_3} = 0.21053,\ \alpha_0 = \frac{\lambda}{\lambda+\mu_3} = 0.21053,$$
$$\alpha_1 = \frac{\lambda\mu_3}{(\lambda+\mu_3)^2} = 0.1662,\ \alpha_2 = \frac{\lambda\mu_3^2}{(\lambda+\mu_3)^3} = 0.13121.$$

From (12), we have $z = 0.21212$, $r_1 = 0.011744$, $r_2 = 0.003691$ and then the sub matrix is given by

$$R = \begin{pmatrix} 0.44444 & 0 & 0.011744 \\ 0 & 0.28571 & 0.003691 \\ 0 & 0 & 0.21212 \end{pmatrix}.$$

From (23) and (29), we obtain

$$\varphi_{1,1} = 0.12623,\ \varphi_{1,2} = 0.15329,\ \varphi_{1,3} = 0.16747,$$
$$\varphi_{2,1} = 0.058069,\ \varphi_{2,2} = 0.04415,\ \varphi_{2,3} = 0.035523,$$
$$\varphi_{3,1} = 0.026226,\ \varphi_{3,2} = 0.012821,\ \varphi_{3,3} = 0.0075352,$$
$$\phi_{1,1} = 0.096782,\ \phi_{1,2} = 0.745,\ \phi_{1,3} = 0.059205,$$
$$\phi_{2,1} = 0.10747,\ \phi_{2,2} = 0.047588,\ \phi_{2,3} = 0.0080734,$$
$$\phi_{3,1} = 0.24155,\ \phi_{3,2} = 0.15487,\ \phi_{3,3} = 0.0012619.$$

From (20), we have $b_{1,0} = (0.24691,\ 0.31746,\ 0.35088)$ and $b_{2,0} = (0.15873,\ 0.20408,\ 0.22556)$. Thus, it finally follows from (14)-(19) and (21) yielding the following system of simultaneous linear equations

$$\begin{aligned}
q_{1,1} &= 0.13121 p_0 + 0.24691 q_{1,1} + 0.31746 q_{1,2} + 0.35088 q_{1,3} \\
&\quad + 0.096782 q_{3,1} + 0.745 q_{3,2} + 0.059205 q_{3,3}, \\
q_{1,2} &= 0.15873 q_{2,1} + 0.20408 q_{2,2} + 0.22556 q_{2,3} + 0.10747 q_{3,1} \\
&\quad + 0.047588 q_{3,2} + 0.0080734 q_{3,3}, \\
q_{1,3} &= 0.12623 q_{3,1} + 0.15329 q_{2,3} + 0.16747 q_{3,3}, \\
q_{2,1} &= 0.1662 p_0 + 0.44444 q_{1,1}, \\
q_{2,2} &= 0.28571 q_{1,2}, \\
q_{2,3} &= 0.21053 q_{1,2} + 0.058069 q_{3,1} + 0.044415 q_{3,2} + 0.035523 q_{3,3}, \\
q_{3,1} &= 0.21053 p_0 + 0.44444 q_{2,1}, \\
q_{3,2} &= 0.28571 q_{2,2}, \\
q_{3,3} &= 0.21053 q_{2,2} + 0.026226 q_{3,1} + 0.012821 q_{3,2} + 0.0075352 q_{3,3}, \\
&3p_0 + q_{1,1} + q_{1,2} + q_{1,3} + q_{2,1} + q_{2,2} + q_{2,3} \\
&+1.8268 q_{3,1} + 1.4066 q_{3,2} + 1.2692 q_{3,3} = 1.
\end{aligned}$$

Solving the equations listed above yields

$$p_0 = 0.224298,\ p_1 = 0.224298,\ p_2 = 0.224298,$$
$$q_{1,1} = 0.065303,\ q_{1,2} = 0.022288,\ q_{1,3} = 0.010529,$$
$$q_{2,1} = 0.066301,\ q_{2,2} = 0.006368,\ q_{2,3} = 0.009347,$$
$$q_{3,1} = 0.076689,\ q_{3,2} = 0.001819,\ \text{and } q_{3,3} = 0.003401.$$

It follows from (10) that we have the steady-state probabilities $q_n(n > N)$. Inserting (33) and (35), we obtain $E[L_s] = 1.492852$ and $E[W] = 2.651378$.

The above bulk-service M/M/1 system with N (=2) policy with dependent service rates is presented numerically to demonstrate the implement and efficiency of the proposed procedure (Section 3). Once the matrix R has been constructed, we easily calculate the probability distributions at points of arrivals by solving a linear system of equations. This example is different from the ordinary M/M^r/1 queuing system (by Kleinrock (1975)) in that when the system is empty the server deactivates and does not provide a varying bulk-service rate until the number of customers reaches *N*. Kleinrocks' results are obtained by solving the equation of the *r*-th order. Note that when interarrival time is not exponential, the matrix geometric procedure yields an efficient computational method for obtaining the steady-state probabilities (see Neuts (1981)).

6. CONCLUSIONS

For the bulk service G/M/1 (or M/G/1) queuing systems with dependent service rate, the mathematical modeling is not easily tractable. Although embedded Markovian process is frequently used to model G/M/1 (or M/G/1) systems (see Neuts (1967), Baba (1996)), it should be noted that Neuts' (1967) approach is unsuitable for computational purpose. In this paper, we study a variant bulk service G/M/1 queuing system with an *N* policy. We show that the queue size and the service batch size at points of arrivals form an embedded Markov chain and the steady-state probabilities of this Markov chain have the matrix geometric form. By the arguments of Neuts (1981), the rate matrix *R* of the matrix geometric solution procedure is derived in a readily computable form. We also obtain some explicit and computationally tractable results such as mean queue length and mean waiting time in the queue. Once *R* is obtained, we easily find the steady-state results. A variant M/M/1 example is demonstrated numerically to verify the validation of the solution procedure.

Bulk-service queuing models arise naturally in many applications. For example, transportation processes including buses, airplanes, etc., and industrial applications such as heat treatments and metal plating operations, all have a common feature of bulk service. For efficient purposes, the control of such system may have certain advantages in some practical applications. In future, the work can be generalized for other policies (such as *T* policy and *D* policy) or triadic policies (such as *NT* policy and *ND* policy).

ACKNOWLEDGEMENTS

The authors are pleased to thank three anonymous referees and Editor-in-Chief, Prof. G.-H. Tzeng, for their valuable comments and suggestions, which improved the presentation of this paper.

REFERENCES

1. Baba, Y. (1996). A bulk service GI/M/1 queue with service rates depending on service batch size. *Journal of the Operations Research Society of Japan*, 39: 25-35.
2. Bell, C.E. (1971). Characterization and computation of optimal policies for operating an M/G/1 queueing system with removable server. *Operations Research*, 19: 208-218.
3. Bell, C.E. (1972). Optimal operation of an M/G/1 priority queue with removable server. *Operations Research*, 21: 1281-1289.
4. Chaudhry, M.L. and Templeton, J.G.C. (1984). *A First Course in Bulk Queues*. John Wiley and Sons, New York.
5. Curry, G.L. and Feldman, R.M. (1985). An M/M/1 queue with a general bulk service rule. *Naval Research Logistics Quarterly*, 32: 595-603.
6. Gakis, K.G., Rhee, H.K., and Sivazlian, B.D. (1995). Distributions and first moments of the busy and idle periods in controllable M/G/1 queueing models with simple and dyadic policies. *Stochastic Analysis and Applications*, 13: 47-81.
7. Heyman, D.P. (1968). Optimal operating policies for M/G/1 queueing system. *Operations Research*, 16: 362-382.
8. Ke, J.-C. and Wang, K.-H. (2002). A recursive method for the N policy G/M/1 queueing system with finite capacity. *European Journal of Operational Research*, 142: 577-594.
9. Kimura, T. (1981). Optimal control of an M/G/1 queueing system with removable server via diffusion approximation. *European Journal of Operational Research*, 8: 390-398.
10. Kleinrock, L. (1975). *Queueing Systems*. Vol. I. Theory, Wiley, New York.
11. Laxmi, P.V. and Gupta, U.C. (1999). On the finite-buffer bulk-service queue with general independent arrivals: GI/M[x]/1/N. *Operations Research Letters*, 25: 241-245.
12. Medhi, J. (1984). *Recent Developments in Bulk Queueing Models*. Wiley Easter, Ltd. New Delhi.
13. Neuts, M.F. (1967). A general class of bulk queues with Poisson input. *Annals of Mathematical Statistics*, 38: 759-770.
14. Neuts, M.F. (1979). Queues solvable without Rouche's theorem. *Operations Research*, 27: 767-781.
15. Neuts, M.F. (1981). *Matrix Geometric Solutions in Stochastic Models: an Algorithmic Approach*. The John Hopkins University Press, Baltimore.
16. Teghem, J. Jr. (1987). Optimal control of a removable server in an M/G/1 queue with finite capacity. *European Journal of Operational Research*, 31: 358-367.
17. Tijms, H.C. (1986). *Stochastic Modelling and Analysis*. Wiley, New York.
18. Wang, K.-H. and Ke, J.-C. (2000). A recursive method to the optimal control of an M/G/1 queueing system with finite capacity and infinite capacity. *Applied Mathematical Modeling*, 24: 899-914.
19. Wang, K.-H. and Yen, K.-L. (2003). Optimal control of an M/H_k/1 queueing system with a removable server. *Mathematical Methods of Operations Research*, 57: 255-262.
20. Yadin, M. and Naor, P. (1963). Queueing systems with a removable service station. *Operational Research Quarterly*, 14: 393-405.
21. Zhang, Z.G., Vickson R., and Love, E. (2001). The optimal service policies in an M/G/1 queueing system with multiple vacation types. *INFOR*, 39: 357-366.

APPENDIX A: THE DERIVATION OF $b_{i,\ell,j}$

First, we let T_n denote the time between the $(r-1)$-st and the r-th customers. Since the interarrival times are assumed independent, the random variable T_n can be denoted by T, and we denote its CDF by $A(t)$ (see Section 1).

Case 1: For $i = j = N$, $b_{N,\ell-1,N}$ is corresponding to the probability that ℓ services with rate μ_N progress during an interval time T. Then

$$b_{N,\ell-1,N} = \int_0^\infty \frac{(\mu_N t)^\ell}{\ell!} e^{-\mu_N t} dA(t), \text{ for } \ell \ge 1 \quad \text{(A-1)}$$

Case 2: For $1 \le i \le N-1$ and $j = N$, $b_{i,\ell,N}$ can obtained as follows; (see Chaudhry and Templeton, 1984)

After the r-th customer arrives, a time X has elapsed when the service of batch size N has been served $\ell+1$ times with rate μ_N. This case when the time X has elapsed, departures from the system can refer to a Poisson process. The time X, then, is Erlang-distributed, being the sum of $\ell+1$ exponential random variables with μ_N. Thus the p.d.f. of X is given by

$$\frac{\mu_N (\mu_N x)^\ell}{\ell!} e^{-\mu_N x}. \quad \text{(A-2)}$$

When the time required for the waiting customers of batch size i to enter into service is less than the interarrival time, that is when $X < T$, then there is time $T-X$ remaining to have the customers of batch size i will start their services. The probability that exactly i size of batch will be served before time $T-X$ expires is given by

$$e^{-\mu_i (T-X)}. \quad \text{(A-3)}$$

From (A-2) and (A-3), $b_{i,\ell,N}$ is expressed as

$$b_{i,\ell,N} = \int_0^\infty dA(t)\int_0^t e^{-\mu_i(t-x)}\mu_N \frac{(\mu_N x)^\ell}{\ell!} e^{-\mu_N x} dx,$$
for $\ell \ge 0$ (A-4)

Case 3: For $1 \le j \le N-1$ and $i = N$, similar to the analysis for Case 2, $b_{N,\ell,j}$ is given by

$$b_{N,\ell,j} = \int_0^\infty dA(t)\int_0^t e^{-\mu_j(t-x)}\mu_j \frac{(\mu_N x)^\ell}{\ell!} e^{-\mu_N x} dx,$$
for $\ell \ge 0$ (A-5)

Case 4: For $1 \le i = j \le N-1$,

(1). Consider $\ell = 0$ case, $b_{i,0,i}$ = Pro[the service of batch size *i* during the period *T*] is given by

$$b_{i,0,i} = \int_0^\infty \mu_i t e^{-\mu_i t} dA(t), \text{ for } \ell = 0 \tag{A-6}$$

(2). Consider $\ell \ne 0$ case, $b_{i,\ell,i}$ is derived from the following: After the *r*-th customer arrives, a time X has elapsed when the service of batch size N has been served ℓ times with rate μ_N. Similar to the analysis Case 2, we get p.d.f. of X is given by

$$\frac{\mu_N (\mu_N x)^{\ell-1}}{(\ell-1)!} e^{-\mu_N x}. \tag{A-7}$$

The time is again elapsed Y when the service of batch size *i* has been processed. Similar arguments, we have

$$\mu_i e^{-\mu_i y}. \tag{A-8}$$

In contrast, the probability that exactly *i* size of batch will be served before time $T-X-Y$ expires is given by

$$e^{-\mu_i(T-X-Y)}. \tag{A-9}$$

From (A-7) - (A-9), $b_{i,\ell,i}$ is expressed as

$$\begin{aligned} b_{i,\ell,i} &= \int_0^\infty dA(t)\int_0^t \mu_N \frac{(\mu_N x)^{\ell-1}}{(\ell-1)!} e^{-\mu_N x} dx \\ &\quad \times \int_0^{t-x} \mu_i e^{-\mu_i y} e^{-\mu_i(t-x-y)} dy \\ &= \int_0^\infty \mu_i \mu_N e^{-\mu_i t} dA(t) \int_0^t (t-x)\frac{(\mu_N x)^{\ell-1}}{(\ell-1)!} \\ &\quad \times e^{(\mu_i-\mu_N)x} dx, \quad \text{for } \ell \ge 1 \end{aligned} \tag{A-10}$$

Case 5: For $1 \le i \ne j \le N-1$, $b_{i,\ell,i}$ is obtained by the same procedure as Case 4.

APPENDIX B: EQUATION (13) IS DERIVED BY SUPPLEMENTARY VARIABLE TECHNIQUE

The state of the system at time *t* is given by
$X(t) \equiv$ number of customers in the system, and
$V(t) \equiv$ remaining interarrival time for the customer who is arriving.

Let us define

$p_n(v,t)dv = \Pr\{X(t)=n,\ v < V(t) \le v+dv\},\ v \ge 0,$
$n = 0,\ 1,\ 2,\ ...,\ N-1$

$q_{i,n}(v,t)dv = \Pr\{X(t)=n,\ v < V(t) \le v+dv\},\ v \ge 0,$
$n = 1,\ 2,\ ...,\ N,\ i = 1,2,...\infty.$

Relating the state of the system at time *t* and *t*+*dt*, we easily set up the following partial differential equations for server idle as follows

$$\begin{aligned}(\frac{\partial}{\partial t} - \frac{\partial}{\partial v}) p_0(v,t) &= \mu_1 q_{1,1}(v,t) + \mu_2 q_{1,2}(v,t) + \cdots \\ &+ \mu_N q_{1,N}(v,t),\end{aligned} \tag{B-1}$$

$$(\frac{\partial}{\partial t} - \frac{\partial}{\partial v}) p_n(v,t) = \frac{d}{dv} A(v) p_{n-1}(0;t), \quad 1 \le n \le N-1 \tag{B-2}$$

In steady-state, (B-1) and (B-2) becomes

$$-\frac{d}{dv} p_0(v) = \mu_1 q_{1,1}(v) + \mu_2 q_{1,2}(v) + \cdots + \mu_N q_{1,N}(v), \tag{B-3}$$

$$-\frac{d}{dv} p_n(v) = \frac{d}{dv} A(v) p_{n-1}(0), \quad 1 \le n \le N-1. \tag{B-4}$$

We introduce the following Laplace-Stieltjes transforms:

$$p_n^*(\theta) = \int_0^\infty e^{-\theta v} p_n(v) dv,$$
$$p_n = p_n^*(0) = \int_0^\infty p_n(v) dv.$$

Similar to the analysis by Ke and Wang (2002), It follows from (B-4) that

$$p_1 = p_2 = \ ... \ = p_{N-1} \tag{B-5}$$

Note that the partial differential equations for server busy are constructed in this manner. In this case, their solution is not easily tractable. Thus, this paper develops the queue size distribution using embedded Markov chain.