Chapter 5

MARKOVIAN MULTISERVER VACATION MODELS

In the three previous chapters, we focused on single server vacation models of different types. In this and the next chapter, we will discuss the multiserver vacation models.

5.1 Introduction to Multiserver Vacation Models

In many practical queueing systems, multiple servers attend to the queue. Call centers, banks, and fast food restaurants are a few examples. A common feature of these systems is that the servers can perform some secondary, nonqueueing tasks when they are not busy. For example, call center agents may make outbound calls to potential customers when no inbound calls are on hold. These outbound calls are secondary or supplementary jobs that can be done by the idle agents. To model this feature, we use the multiserver model with vacations that represent the durations of secondary jobs. Compared with single server vacation models, the multiserver vacation models are more complex to analyze. Levy and Yechiali (1976) studied the M/M/c queue with exponential vacations and obtained the distribution of the number of busy servers and the expected number of customers in the system. Neuts (1981) developed the matrix analytical method, which provides a powerful tool in studying complex stochastic systems. Vinod (1986) presented the analysis of M/M/c queue with vacations by using the quasi-birth-and-death (QBD) process. By finding the explicit expression of the rate matrix, Tian and Zhang (2000) obtained the distributions of the queue length and the waiting time in various M/M/c queueing systems with vacations and established the conditional stochastic decomposition properties for the queue length and the waiting time. Like the unconditional stochastic decomposition properties for the single server vacation model, the conditional stochastic decomposition properties also indicate the relationship between the multiserver vacation system and the corresponding the classical M/M/c system.

The multiserver vacation models have more complex and different system dynamics than the single server vacation models. Below is an overview of the vacation policies used in multiserver vacation models.

(1) *Synchronous All-Server Vacation Policy.* Under such a policy in an M/M/c queue, all c servers start a random vacation V simultaneously. As in the single server model, for the multiple vacation case, if the system remains empty at a vacation completion instant, these servers take another vacation together, and they repeat this process until they find the waiting customer(s) in the system. Then the c servers resume serving the queue. This type of policy applies to the situation where the servers are controlled by the same means or are required to perform a teamwork-type job. For instance, in a mainframe computer system with multiple user terminals, the user terminals are considered to be the servers and the mainframe computer's shutdowns due to power failures or maintenance activities can be treated as synchronous vacations. In this and the next chapter, we denote the multiple and synchronous vacation system by (SY, MV). Similarly, for the single vacation case, when the system becomes empty at a service completion instant, all c servers take only one vacation together. After completing the vacation, these servers either start serving the customers, if any, or stay idle if the system remains empty. The single and synchronous vacation system is denoted by (SY, SV). The third case is that all servers are turned off when the system becomes empty at a service completion instant and are turned on with a setup or warmup period when the next customer arrives. This type of system is called a *synchronous setup model* and is denoted by (SY, SU). Note that these policies are exhaustive service type.

(2) *Asynchronous All-Server Vacation Policy.* Under such a policy in an M/M/c queue, any of c servers starts a vacation independently if this server finds no waiting customer in the system at his or her service completion instant. At this instant, other servers may be serving customers, or on vacation, or idle (for single vacation case). Since the servers take individual vacations independently, we say that the servers follow an *asynchronous vacation policy.* The condition for taking a vacation now is that there is no waiting customer. Thus there may be still some customers in service in the system when a server starts a vacation. Therefore, the policy is also said to be semiexhaustive. If the servers take individual vacations consecutively as long as the queue length is zero, the servers follow a multiple vacation policy. Therefore, the sys-

tem is denoted by (AS, MV). On the other hand, if the server takes only one vacation when no waiting customer is in line at a service completion instant and either resumes service or stays idle, the system then is under a single vacation policy. This system is denoted by (AS,SV). Similarly, if a server is turned off when there is no waiting customer at the server's service completion instant and is turned on with a setup (or a warmup) period when the next customer arrives, the system is called an *asynchronous setup model* and is denoted by (AS, SU).

(3) *Some-Server Vacation Policy.* In some situations, we want to limit the number of servers who can take vacations in the system. Under either an SY or an AS vacation policy in the M/M/c queue, all c servers are eligible for taking vacations. However, the maximum number of servers on vacation at a time is no more than a prespecified number d $(0 < d \leq c)$. This limit also implies that the number of servers attending to the queue (either serving or being idle) is at least $c - d$. This class of policies offer more flexibility in allocating the servers' time to multiple tasks or controlling the servers' utilization level. Clearly, the special case $d = c$ becomes the all-server vacation policy. The some-server vacation policies can be either an SY or an AS type. For each type, the policies can be further classified into multiple vacation, single vacation, or setup time models according to the rules of resuming queue service.

(4) *Threshold Vacation Policy.* As a generalization of the some-server vacation policy, we may introduce more control parameter(s) into the policy. The basic threshold policy is similar to the threshold policy in the single server model and is called the *all-server N-policy with or without vacations.* Under such a policy in an M/M/c queue, all servers start taking a vacation at a service completion instant when the system becomes empty. If the servers keep taking synchronous vacations until the number of customers in the system is at least N at a vacation completion instant, and then resume serving the queue, we call the servers follow an N-threshold vacation policy. If the servers are shut down at a service completion instant when the system is empty, and start serving the customers immediately when the number of customers in the system reaches N, we say the servers follow an *N-policy without vacations.* Another threshold-type policy is a generalization of the some-server vacation policy. Here is how it works. In an M/M/c queue, the servers are allowed to take vacations only when the number of idle servers reaches d at a service completion instant. When this condition is met, a subset of e $(\leq d)$ servers take a vacation together. These e servers keep taking synchronous vacations until there are waiting customers at a vacation completion instant. Then these e servers resume attending to the queue. This policy is called an (e, d) *policy.* As a further extension of the (e, d)

policy, we may introduce the threshold N for service resumption to have a three number (e, d, N) policy. Under such a policy, a group of e servers starts taking a vacation whenever d $(\leq c)$ servers become idle at a service completion instant and keep taking the vacations until the number of customers in the system reaches N at a vacation completion instant; then the e servers resume attending the queue. Note that in the (e, d, N) policy, parameter d controls when the server vacation period starts, parameter e controls the number of servers on vacation, and parameter N controls when the vacationing servers return to the queue service.

It is well known that the stochastic decomposition theorems play a central role in the theory of single server vacation models. However, we cannot establish the corresponding theorems in the multiserver vacation models due to the complexity of the system dynamics. Our research indicates that the relation between the multiserver vacation model and the corresponding classical nonvacation model in terms of stationary performance measures can be established under the condition when all servers are busy. Therefore, we present a set of conditional stochastic decomposition theorems in this and the next chapter. It can be proven that for a steady-state system, given that all servers are busy, the conditional queue length or waiting time in the multiserver vacation model can be decomposed into the sum of two independent random variables. One random variable is the conditional queue length or waiting time in the corresponding nonvacation model, and the other random variable is the additional queue length or the additional delay due to the vacation effect. In fact, the conditional stochastic decomposition properties also exist in the single vacation models (see Doshi,1989) and are the common laws for both single server and multiserver vacation models.

5.2 Quasi-Birth-and-Death Process Approach

5.2.1 QBD Process

Most studies on the multiserver vacation systems focus on the M/M/c systems. These Markovian queueing systems can be modeled as Quasi-Birth-and-Death (QBD) processes and can be analyzed by using the matrix analytical method (MAM). The MAM, mainly developed by Neuts (1981) and other mathematicians, provides a powerful tool in developing the stationary distributions for the QBD processes. A QBD process is the generalization of a birth-and-death (BD) process from a one-dimensional state space to a multidimensional state space. Like the infinitesimal generator of a BD process with the tri-diagonal structure, the infinitesimal generator of a QBD is a block-partitioned tri-diagonal matrix. For the purpose of the model development in this and the next chapter, we present only some relevant materials concerning the QBD processes. For details about the QBD processes and the MAM theory, see Neuts (1981) and Lotouche and Ramaswami (1999).

Consider a two -dimensional Markov process $\{(X(t), J(t)), t \geq 0\}$ with state space

$$\Omega = \{(k, j) : k \geq 0, 1 \leq j \leq m\}.$$

The process $\{(X(t), J(t)), t \geq 0\}$ is called a *QBD process* if the infinitesimal generator of the process is given by

$$\widetilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{C}_0 & & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C}_1 & & & \\ & \mathbf{B}_2 & \mathbf{A}_2 & \mathbf{C}_2 & & \\ & & \mathbf{B}_3 & \mathbf{A}_3 & \mathbf{C}_3 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}, \qquad (5.2.1)$$

where all submatrices are $m \times m$ matrices; $\mathbf{A}_k, k \geq 0$, have negative diagonal elements and nonnegative off-diagonal elements; and $\mathbf{C}_k, k \geq 0$, and $\mathbf{B}_k, k \geq 1$, are all nonnegative matrices satisfying

$$(\mathbf{A}_0 + \mathbf{C}_0)\mathbf{e} = (\mathbf{B}_k + \mathbf{A}_k + \mathbf{C}_k)\mathbf{e} = 0, \qquad k \geq 1.$$

State set $\{(0, 1), ..., (0, m)\}$ is said to be the boundary level; state set $\{(k, 1), ..., (k, m)\}$ is said to be level k. In many applications, we have a special case of (5.2.1) where the nonboundary submatrices of the infinitesimal generator are independent of level k. Thus $\widetilde{\mathbf{Q}}$ is written as

$$\widetilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{C}_0 & & & & \\ \mathbf{B}_1 & \mathbf{A} & \mathbf{C} & & & \\ & \mathbf{B} & \mathbf{A} & \mathbf{C} & & \\ & & \mathbf{B} & \mathbf{A} & \mathbf{C} & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}. \qquad (5.2.2)$$

Assume that the QBD process is positive recurrent, and let (X, J) be the limit of $\{(X(t), J(t))\}$ as $t \to \infty$. Denote the stationary probabilities by

$$\pi_{kj} = P\{X = k, J = j\} = \lim_{t \to \infty} P\{X(t) = k, J(t) = j\}, \ (k, j) \in \Omega.$$
$$\pi_k = (\pi_{k1}, \pi_{k2}, ..., \pi_{km}), \qquad k \geq 0.$$

We present the following theorems without the proofs. For the proofs of these results, see Neuts (1981).

Theorem 5.2.1. The irreducible QBD process is positive recurrent if and only if the matrix equation

$$\mathbf{R}^2\mathbf{B} + \mathbf{R}\mathbf{A} + \mathbf{C} = \mathbf{0} \tag{5.2.3}$$

has the minimum nonnegative solution $\mathbf{R}$, with spectral radius $sp(\mathbf{R}) < 1$, and a set of linear homogeneous equations

$$\pi_0(\mathbf{A}_0 + \mathbf{R}\mathbf{B}_1) = \mathbf{0}$$

has the positive solution. Furthermore, the stationary distribution can be expressed as the matrix geometric form

$$\pi_k = \pi_0\mathbf{R}^k, \qquad k \geq 0,$$

where π_0 is the positive solution of the set of linear homogeneous equations and satisfies the normalization condition

$$\pi_0(\mathbf{I} - \mathbf{R})^{-1}\mathbf{e} = \mathbf{1}.$$

In practical applications, we often encounter the variants of the standard or so-called *canonical form QBD process* presented above. In a noncanonical QBD process, the infinitesimal generator, denoted by $\widetilde{\mathbf{Q}}^*$, still has the same structure as in (5.2.2), where $\mathbf{A}_0$ is an $m_1 \times m_1$ matrix and $\mathbf{C}_0$ and $\mathbf{B}_1$ are $m_1 \times m$ and $m \times m_1$ matrices, respectively. In other words, the number of states for the boundary level is different from the number of states for the nonboundary levels. These noncanonical QBD processes with $\widetilde{\mathbf{Q}}^*$are called *QBD processes with complex boundary behavior* and follow the theorem below.

Theorem 5.2.2. The irreducible QBD process with $\widetilde{\mathbf{Q}}^*$ is positive recurrent if and only if the matrix equation (5.2.3) has the minimum nonnegative solution $\mathbf{R}$, with the spectral radius $sp(\mathbf{R}) < 1$, and the $m_1 + m$ linear homogeneous equations below have the positive solution

$$(\pi_0, \pi_1)B[\mathbf{R}] = \mathbf{0},$$

where $B[\mathbf{R}]$ is the $(m_1 + m) \times (m_1 + m)$ matrix

$$B[\mathbf{R}] = \begin{bmatrix} \mathbf{A}_0 & \mathbf{C}_0 \\ \mathbf{B}_1 & \mathbf{A} + \mathbf{R}\mathbf{B} \end{bmatrix}.$$

Furthermore, the stationary distribution can be expressed as the matrix geometric form

$$\pi_k = \pi_1\mathbf{R}^{k-1}, \qquad k \geq 1,$$

and (π_0, π_1) satisfies the normalization condition as

$$\pi_0\mathbf{e} + \pi_1(\mathbf{I} - \mathbf{R})^{-1}\mathbf{e} = 1.$$

The stationary distribution $\pi_{\underset{\sim}{k}}, k \geq 0$, is called the *modified matrix geometric solution*. Note that $\widetilde{\mathbf{Q}}^*$ differs from the canonical form of (5.2.1) only in the transition probabilities from the boundary states. The probabilistic significance of the matrix $\mathbf{R}$ therefore remains the same, but is applied to nonboundary states only. To analyze the M/M/c queue with vacations, we encounter a more complex variant of the QBD process: For a system with c servers, not only the $k = 0$ level but also $k = 1, 2, ..., c-1$ levels become boundary levels, and these boundary levels contain different numbers of states. Thus we denote the number of states at level k by m_k, $0 \leq k \leq c-1$. The infinitesimal generator is given by

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{C}_0 & & & & & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C}_1 & & & & & & \\ & \cdots & \cdots & & & & & & \\ & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} & \mathbf{C}_{c-1} & & & \\ & & & & \mathbf{B}_c & \mathbf{A} & \mathbf{C} & & \\ & & & & & \mathbf{B} & \mathbf{A} & \mathbf{C} & \\ & & & & & & \mathbf{B} & \mathbf{A} & \cdots \\ & & & & & & & \ddots & \ddots \end{bmatrix}, \tag{5.2.4}$$

where $\mathbf{A}_k$ are the $m_k \times m_k$ matrices, $0 \leq k \leq c-1$; $\mathbf{B}_k$ are the $m_k \times m_{k-1}$ matrices, $0 \leq k \leq c-1$; $\mathbf{C}_k$ are the $m_k \times m_{k+1}$ matrices, $0 \leq k \leq c-2$; $\mathbf{C}_{c-1}$ and $\mathbf{B}_c$ are the $m_{c-1} \times m$ and $m \times m_{c-1}$ matrices, respectively; and $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are all $m \times m$ square matrices. We assume that the QBD process is positive recurrent. The stationary distribution is denoted by

$$\pi_k = (\pi_{k1}, \pi_{k2}, ..., \pi_{km_k}), \qquad 0 \leq k \leq c-1,$$
$$\pi_k = (\pi_{k1}, \pi_{k2}, ..., \pi_{km}), \qquad k \geq c.$$

The matrix $\mathbf{Q}$ in (5.2.4) can be repartitioned into the block form that is consistent with $\widetilde{\mathbf{Q}}^*$as follows: Let

$$\mathcal{A}_0 = \begin{bmatrix} \mathbf{A}_0 & \mathbf{C}_0 & & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C}_1 & & & \\ & \ddots & \ddots & & & \\ & & & \mathbf{B}_{c-2} & \mathbf{A}_{c-2} & \mathbf{C}_{c-2} \\ & & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} \end{bmatrix},$$

$$\mathcal{B}_1 = (0, \mathbf{B}_c), \qquad \mathcal{C}_0 = \begin{pmatrix} 0 \\ \mathbf{C}_{c-1} \end{pmatrix},$$

where $\mathcal{A}_0$ is the square matrix of order $m^* = m_0 + \cdots + m_{c-1}$, and $\mathcal{B}_1$ and $\mathcal{C}_0$ are the $m \times m^*$ and $m^* \times m$ matrices, respectively. Thus (5.2.4) can be rewritten as

$$\mathbf{Q} = \begin{bmatrix} \mathcal{A}_0 & \mathcal{C}_0 & & & & \\ \mathcal{B}_1 & \mathbf{A} & \mathbf{C} & & & \\ & \mathbf{B} & \mathbf{A} & \mathbf{C} & & \\ & & \mathbf{B} & \mathbf{A} & \mathbf{C} & \\ & & & \ddots & \ddots & \ddots \end{bmatrix},$$

and this QBD process becomes a variant of the canonical form. Theorem 5.2.2 now is modified as follows:

Theorem 5.2.3. The irreducible QBD process is positive recurrent if and only if the matrix equation

$$\mathbf{R}^2\mathbf{B} + \mathbf{R}\mathbf{A} + \mathbf{C} = \mathbf{0}$$

has the minimum nonnegative solution, $\mathbf{R}$, with the spectral radius $sp(\mathbf{R}) < 1$, and the linear homogeneous equations

$$(\pi_0, \cdots, \pi_{c-1}, \pi_c)B[\mathbf{R}] = \mathbf{0} \tag{5.2.5}$$

have a positive solution where

$$B[\mathbf{R}] = \begin{bmatrix} \mathbf{A}_0 & \mathbf{C}_0 & & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C}_1 & & & \\ & & \ddots & \ddots & & \\ & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} & \mathbf{C}_{c-1} \\ & & & & \mathbf{B}_c & \mathbf{A} + \mathbf{R}\mathbf{B} \end{bmatrix}.$$

Furthermore, the stationary distribution can be expressed as the matrix geometric form

$$\pi_k = \pi_c \mathbf{R}^{k-c}, \qquad k \geq c, \tag{5.2.6}$$

where $(\pi_0, \cdots, \pi_{c-1}, \pi_c)$ is the positive solution of (5.2.5) and satisfies the normalization condition

$$\sum_{k=0}^{c-1} \pi_k \mathbf{e} + \pi_c(\mathbf{I} - \mathbf{R})^{-1}\mathbf{e} = \mathbf{1}.$$

5.2.2 Conditional Stochastic Decomposition

First, we prove an important property of the matrix geometric distribution, which is the foundation of developing the conditional stochastic decomposition results in this and the next chapter. Assume that the two -dimensional nonnegative random vector (X, J) has the joint distribution

$$\pi_{kj} = P\{X = k, J = j\}, \qquad k \geq 0, 0 \leq j \leq c,$$

and let

$$\pi_k = (\pi_{k0}, \pi_{k1}, \cdots, \pi_{kc}), \qquad k \geq 0.$$

Furthermore, we assume that (X, J) follows a matrix geometric distribution and that there exists a nonnegative square matrix $\mathbf{R}$ of order $c+1$ with $sp(\mathbf{R}) < 1$. Therefore we have

$$\pi_k = \beta \mathbf{R}^k, \qquad k \geq 0; \qquad \beta(\mathbf{I} - \mathbf{R})^{-1}\mathbf{e} = 1,$$

where $\beta = \pi_0 = (\beta_0, \beta_1, \cdots, \beta_c)$. Now we only consider the case where $\mathbf{R}$ is a triangular block-partitioned matrix,

$$\mathbf{R} = \begin{bmatrix} \mathbf{H} & \eta \\ \mathbf{0} & r \end{bmatrix}, \tag{5.2.7}$$

where $\mathbf{H}$ is a $c \times c$ matrix, η is a $c \times 1$ column vector, and r is a real number. It follows from $sp(\mathbf{R}) < 1$ that $sp(\mathbf{H}) < 1$ and $0 < r < 1$. Defining the conditional random variable

$$X^{(c)} = \{X \mid J = c\},$$

we have the stochastic decomposition theorem.

Theorem 5.2.4. If $\mathbf{R}$ has the form given in (5.2.7), $X^{(c)}$ can be decomposed into the sum of two independent random variables,

$$X^{(c)} = X_0 + X_d,$$

where X_0 follows a geometric distribution with parameter r and X_d follows a discrete PH distribution of order c, with the p.g.f.

$$X_d(z) = \frac{1}{\sigma}\left\{\beta_c + z(\beta_0, \beta_1, \cdots, \beta_{c-1})(\mathbf{I} - z\mathbf{H})^{-1}\eta\right\}, \tag{5.2.8}$$

where

$$\sigma = \beta_c + (\beta_0, \beta_1, \cdots, \beta_{c-1})(\mathbf{I} - \mathbf{H})^{-1}\eta.$$

Proof: Since $\mathbf{R}$ is a triangular block-partitioned matrix, we have

$$\mathbf{R}^k = \begin{bmatrix} \mathbf{H}^k & \sum_{i=0}^{k-1} r^i \mathbf{H}^{k-1-i}\eta \\ \mathbf{0} & r^k \end{bmatrix}, \qquad k \geq 1.$$

Substituting $\mathbf{R}^k$ into the matrix geometric expression, we get

$$\begin{aligned}
\pi_k =&(\pi_{k0}, \pi_{k1}, \cdots, \pi_{kc}) = \beta \mathbf{R}^k \\
=&(\beta_0, \beta_1, \cdots, \beta_c)\begin{bmatrix} \mathbf{H}^k & \sum_{i=0}^{k-1} r^i \mathbf{H}^{k-1-i}\eta \\ \mathbf{0} & r^k \end{bmatrix} \\
=&\left((\beta_0, \beta_1, \cdots, \beta_{c-1})\mathbf{H}^k, \beta_c r^k + (\beta_0, \beta_1, \cdots, \beta_{c-1})\sum_{i=0}^{k-1} r^i \mathbf{H}^{k-1-i}\eta\right), \\
&k \geq 0.
\end{aligned}$$

From this expression, we obtain the joint probability

$$\pi_{kc} = \beta_c r^k + (\beta_0, \beta_1, \cdots, \beta_{c-1}) \sum_{i=0}^{k-1} r^i \mathbf{H}^{k-1-i} \eta, \qquad k \geq 0. \qquad (5.2.9)$$

Using (5.2.9), it is easy to compute the probability of the condition event

$$\begin{aligned} P\{J = c\} =& \sum_{k=0}^{\infty} \pi_{kc} \\ =&\beta_c \sum_{k=0}^{\infty} r^k + (\beta_0, \beta_1, \cdots, \beta_{c-1}) \sum_{k=1}^{\infty} \sum_{i=0}^{k-1} r^i \mathbf{H}^{k-1-i} \eta \\ =&\frac{1}{1-r} \left[\beta_c + (\beta_0, \beta_1, \cdots, \beta_{c-1})(\mathbf{I} - \mathbf{H})^{-1} \eta \right] \\ =&\frac{\sigma}{1-r}. \end{aligned}$$

Now the conditional probability is given by

$$P\{X^{(c)} = k\} = \frac{1-r}{\sigma} \pi_{kc}, \qquad k \geq 0.$$

Taking the p.g.f., we have

$$\begin{aligned} X^{(c)}(z) &= \sum_{k=0}^{\infty} z^k P\{X^{(c)} = k\} \\ &= \frac{1-r}{\sigma} \left\{ \beta_c \sum_{k=0}^{\infty} (zr)^k + (\beta_0, \beta_1, \cdots, \beta_{c-1}) \sum_{k=1}^{\infty} \sum_{i=0}^{k-1} r^i \mathbf{H}^{k-1-i} \eta \right\} \\ &= \frac{1-r}{\sigma} \left\{ \frac{\beta_c}{1-zr} + z(\beta_0, \beta_1, \cdots, \beta_{c-1}) \frac{1}{1-zr} (\mathbf{I} - z\mathbf{H})^{-1} \eta \right\} \\ &= \frac{1-r}{1-zr} \frac{1}{\sigma} \left\{ \beta_c + z(\beta_0, \beta_1, \cdots, \beta_{c-1})(\mathbf{I} - z\mathbf{H})^{-1} \eta \right\} \\ &= X_0(z) X_d(z), \end{aligned}$$

where $X_0(z) = (1-r)(1-zr)^{-1}$ is the p.g.f. of the geometric distribution. Expanding $X_d(z)$ gives

$$P\{X_d = k\} = \begin{cases} \frac{\beta_c}{\sigma} & k = 0, \\ \frac{1}{\sigma}(\beta_0, \beta_1, \cdots, \beta_{c-1}) \mathbf{H}^{k-1-i} \eta & k \geq 1. \end{cases}$$

Therefore, X_d follows a matrix geometric distribution. Based on Lemma 4.1.1 in Sengupta (1991), X_d is a discrete PH distribution of order c. □

If $\mathbf{R}$ is a lower triangular block-partitioned matrix,

$$\mathbf{R} = \begin{bmatrix} r & 0 \\ \xi & \mathbf{H} \end{bmatrix}, \tag{5.2.10}$$

where $\mathbf{H}$ is the $c \times c$ square matrix, ξ is the $c \times 1$ column vector, and r is a real number in $(0, 1)$. Defining the conditional random variable

$$X^{(0)} = \{X \mid J = 0\}$$

and using the same approach, we can prove the following theorem.

Theorem 5.2.5. If $\mathbf{R}$ has the form given in (5.2.10), $X^{(c)}$ can be decomposed into the sum of two independent random variables,

$$X^{(c)} = X_0 + X_d,$$

where X_0 follows a geometric distribution with parameter r and X_d follows a discrete PH distribution of order c, with the p.g.f.

$$X_d(z) = \frac{1}{\sigma}\left\{\beta_0 + z(\beta_1, \beta_2, \cdots, \beta_c)(\mathbf{I} - z\mathbf{H})^{-1}\xi\right\}, \tag{5.2.11}$$

where

$$\sigma = \beta_0 + (\beta_1, \beta_2, \cdots, \beta_c)(\mathbf{I} - \mathbf{H})^{-1}\xi.$$

5.3 M/M/c Queue with Synchronous Vacations

5.3.1 Multiple Vacation Model

Consider an M/M/c system with arrival rate λ, service rate μ, and FCFS service order. The detailed analysis of this classical queueing system can be found in any book on queueing theory (see Kleinrock (1975), Harris and Gross (1985), etc.). For the convenience of reference, we present the main results of the M/M/c queue that are relevant to the vacation models in this chapter. If $\rho = \lambda(c\mu)^{-1} < 1$, the system is positive recurrent, and there exists the stationary distribution of the queue length. In the steady-state, the number of waiting customers given that all servers are busy, denoted by $L_0^{(c)}$, follows a geometric distribution with parameter ρ. That is

$$P\{L_0^{(c)} = k\} = (1-\rho)\rho^k, \qquad k \geq 0. \tag{5.3.1}$$

Given that a customer arrives at a state when all the servers are busy, this customer's conditional waiting time $W_0^{(c)}$ follows an exponential distribution with parameter $c\mu(1-\rho)$. Therefore, its distribution function and LST are, respectively,

$$W_0^{(c)}(x) = 1 - e^{-c\mu(1-\rho)x}, x \geq 0; \quad W_0^{*(c)}(s) = \frac{c\mu(1-\rho)}{s + c\mu(1-\rho)}. \tag{5.3.2}$$

Allowing servers to take multiple synchronous vacations in an M/M/c system, we have the multiserver vacation model denoted by M/M/c (SY, MV). In such a system, all c servers start taking a vacation together when the system becomes empty at a service completion instant. At a vacation termination instant, if the system remains empty, these servers take another vacation together; if there are $1 \le j < c$ customers in the system, then j servers start serving customers and $c - j$ servers stay idle; if there are $j \ge c$ customers in the system, all c servers start serving the customers and $j - c$ customers wait in the line. We assume that the vacations are i.i.d. random variables, denoted by V, following a PH distribution of order m with the irreducible representation $(\alpha, \mathbf{T})$ and $\alpha\mathbf{e} = 1$. This means that there is no positive probability that the vacation is zero and the LST of V is $v(s) = \alpha(s\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0$. It is also assumed that the vacation times, the service times, and the interarrival times are mutually independent.

Let $L_v(t)$ be the number of customers in the system at time t and define

$$J(t) = \begin{cases} 0 & \text{the servers are not on vacation,} \\ j & \text{the servers are on vacation at phase } j, \quad j = 1, 2, ..., m. \end{cases}$$

Since the vacations are synchronous, at least one server is busy during the nonvacation period, and some servers may be idle. Note that the servers' being idle is different from their being on vacation. With the (SY, MV) policy, $\{((L_v(t), J(t)), t \ge 0\}$ is a QBD process with state space

$$\Omega = \{(0, j) : 1 \le j \le m\} \cup \{(k, j) : k \ge 1, 0 \le j \le m\}.$$

The infinitesimal generator can be rewritten in the block-partitioned form

$$\mathbf{Q} = \begin{bmatrix} \mathcal{A}_0 & \mathcal{C}_0 & & & & \\ \mathcal{B}_1 & \mathbf{A} & \mathbf{C} & & & \\ & \mathbf{B} & \mathbf{A} & \mathbf{C} & & \\ & & \mathbf{B} & \mathbf{A} & \mathbf{C} & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{5.3.3}$$

In (5.3.3), $\mathcal{A}_0$ is a square matrix of order $m^* = (c - 1)(m + 1) + m$, representing the transitions among the boundary states, where the number of customers in the system is no more than $c - 1$. $\mathcal{B}_1$ and $\mathcal{C}_0$ are the $(m + 1) \times m^*$ and $m^* \times (m + 1)$ matrices, respectively. These matrices can be written as

$$\mathcal{A}_0 = \begin{bmatrix} \mathbf{A}_0 & \mathbf{C}_0 & & & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C} & & & & \\ & \mathbf{B}_2 & \mathbf{A}_2 & C & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \mathbf{B}_{c-2} & \mathbf{A}_{c-2} & \mathbf{C} \\ & & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} \end{bmatrix},$$

$$\mathcal{B}_1 = (\mathbf{0}, \mathbf{B}_c), \qquad \mathcal{C}_0 = \begin{pmatrix} \mathbf{0} \\ \mathbf{C} \end{pmatrix},$$

where $\mathbf{A}_0 = -\lambda\mathbf{I}+\mathbf{T}+\mathbf{T}^0\alpha$ is the square matrix of order m, $\mathbf{C}_0 = (\mathbf{0}, \lambda\mathbf{I})$ is the $m \times (m+1)$ matrix, and $\mathbf{C} = \lambda\mathbf{I}$ is the square matrix of order $(m+1)$. Moreover, we have

$$\mathbf{A}_k = \begin{bmatrix} -(\lambda + k\mu) & \mathbf{0} \\ \mathbf{T}^0 & -\lambda\mathbf{I} + \mathbf{T} \end{bmatrix}_{(m+1)\times(m+1)}, \qquad 1 \le k \le c-1,$$

$$\mathbf{B}_1 = \begin{pmatrix} \mu\alpha \\ \mathbf{0} \end{pmatrix}_{(m+1)\times m}, \quad \mathbf{B}_k = \begin{bmatrix} k\mu & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{(m+1)\times(m+1)}, \; 2 \le k \le c-1.$$

$\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ in (5.3.3) are all the square matrices of order $m+1$, as follows:

$$\mathbf{A} = \begin{bmatrix} -(\lambda + c\mu) & \mathbf{0} \\ \mathbf{T}^0 & -\lambda\mathbf{I} + \mathbf{T} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} c\mu & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{C} = \lambda\mathbf{I}.$$

Theorem 5.3.1. If $\rho = \lambda(c\mu)^{-1} < 1$, the matrix equation $\mathbf{R}^2\mathbf{B} + \mathbf{R}\mathbf{A} + \mathbf{C} = 0$ has the minimum nonnegative solution

$$\mathbf{R} = \begin{bmatrix} \rho & \mathbf{0} \\ \rho\mathbf{e} & \lambda(\lambda\mathbf{I} - \mathbf{T})^{-1} \end{bmatrix}. \tag{5.3.4}$$

Proof: Since $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are all the lower triangular block-partitioned matrices, the solution to the matrix equation must have the same form. Assume that

$$\mathbf{R} = \begin{bmatrix} r & \mathbf{0} \\ \xi & \mathbf{H} \end{bmatrix},$$

where r is a real number, $\mathbf{H}$ is a square matrix of order m, and ξ is a $m \times 1$ column vector. Substituting $\mathbf{R}$ into the matrix equation, we have

$$\begin{cases} c\mu r^2 - (\lambda + c\mu)r + \lambda = 0 \\ \mathbf{H}(-\lambda\mathbf{I} + \mathbf{T}) + \lambda\mathbf{I} = 0 \\ c\mu(r\mathbf{I} + \mathbf{H})\xi - (\lambda + c\mu)\xi + \mathbf{H}\mathbf{T}^0 = 0 \end{cases} \tag{5.3.5}$$

If $\rho < 1$, the first equation of (5.3.5) has the minimum nonnegative solution $r = \rho$ (the other solution is $r = 1$). The second equation of (5.3.5) gives $\mathbf{H} = \lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}$, which is nonnegative. Substituting ρ and $\mathbf{H}$ into the third equation of (5.3.5) and using the fact that $-\mathbf{Te} = \mathbf{T}^0$, we have

$$\begin{aligned}
\xi &= \frac{\lambda}{c\mu}\left\{I - \lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right\}^{-1}(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0 \\
&= \rho\left\{(\lambda\mathbf{I} - \mathbf{T})\left[\mathbf{I} - \lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]\right\}^{-1}\mathbf{T}^0 \\
&= \rho(-\mathbf{T})^{-1}\mathbf{T}^0 = \rho\mathbf{e}.
\end{aligned}$$

□

Note that $\mathbf{H} = \lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}$ is a substochastic matrix with $sp(\mathbf{H}) < 1$. It follows from the structure of $\mathbf{R}$ that $sp(\mathbf{R}) = \max\{\rho, sp\{\mathbf{H}\}\}$. Thus the necessary and sufficient condition for $sp(\mathbf{R}) < 1$ is $\rho < 1$. It is easy to verify that under the condition $\rho < 1$, the matrix

$$B[\mathbf{R}] = \begin{bmatrix} \mathcal{A}_0 & \mathcal{C}_0 \\ \mathcal{B}_1 & \mathbf{RB} + \mathbf{R} \end{bmatrix}$$

is a finite, aperiodic, and irreducible infinitesimal generator, and the linear homogeneous equation set (5.2.5) must have a positive solution. For instance, if $\mathbf{x}$ is the stationary probability vector of $B[\mathbf{R}]$, then any positive vector $K\mathbf{x}$ is a positive solution of (5.2.5), where K is any constant factor. It follows from Theorem 5.2.3 that the system is positive recurrent if and only if $\rho < 1$.

Assume that $\rho < 1$ and let (L_v, J) be the stationary limit of $\{L_v(t), J(t)\}$, with the stationary probability distribution denoted by

$$\begin{aligned}
x_k &= P\{L_v = k, J = 0\}, \qquad k \geq 1, \\
\pi_{kj} &= P\{L_v = k, J = j\}, \qquad k \geq 0,\ 1 \leq j \leq m, \\
\pi_k &= (\pi_{k1}, \pi_{k2}, \cdots, \pi_{km}\}, \qquad k \geq 0.
\end{aligned}$$

Theorem 5.3.2. If $\rho < 1$, the distribution of (L_v, J) in the M/M/c (SY, MV) system is given by

$$\begin{cases}
\pi_j = K\beta\left[\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^j, & j \geq 0, \\
x_j = K\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \psi_j & 1 \leq j \leq c-1, \\
x_j = x_{c-1}\rho^{j-c+1} + \rho\pi_{c-1}\sum_{i=0}^{j-c}\rho^i\left[\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^{j-c-i}\mathbf{e}, & j \geq c,
\end{cases} \tag{5.3.6}$$

where

$$\beta = (\beta_1, \cdots, \beta_m) = \frac{\lambda}{1 - v(\lambda)}\alpha(\lambda\mathbf{I} - \mathbf{T})^{-1};\ \beta\mathbf{e} = 1,$$

$$\psi_j = \beta \sum_{i=0}^{j-1} i! \left[\mu(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^i \mathbf{e}, \qquad 1 \le j \le c-1,$$

$$K = \left\{ \sum_{j=1}^{c-1} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \psi_j + \frac{\rho}{1-\rho}\frac{\left(\frac{\lambda}{\mu}\right)^{c-1}}{(c-1)!}\psi_{c-1} \right.$$
$$\left. + \beta\left[\mathbf{I} - \frac{\rho}{1-\rho}\left(\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right)^{c-1}\right](\mathbf{I} - \lambda\mathbf{T}^{-1})\mathbf{e} \right\}^{-1}.$$

Proof: The stationary distribution is rewritten in the segment partitioned vector form as

$$\mathbf{\Pi} = (\pi_0, (x_1, \pi_1), \cdots, (x_n, \pi_n), \cdots).$$

Clearly, $\mathbf{\Pi Q} = \mathbf{0}, \mathbf{\Pi e} = 1$. Since every column containing $(-\lambda\mathbf{I} + \mathbf{T})$ has only this nonzero submatrix, we have

$$\begin{aligned} &\pi_j = \pi_0\left[\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^j, \qquad j \ge 0, \\ &\pi_0(-\lambda\mathbf{I} + \mathbf{T} + \mathbf{T}^0\alpha) + x_1\mu\alpha = \mathbf{0}. \end{aligned} \tag{5.3.7}$$

Using $\lambda\mathbf{I} - \mathbf{T} - \mathbf{T}^0\alpha = (\lambda\mathbf{I} - \mathbf{T})\left[\mathbf{I} - (\lambda\mathbf{T} - \mathbf{T})^{-1}\mathbf{T}^0\alpha\right]$, for $j \ge 1$, we have

$$\left[(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0\alpha\right]^j = [v(\lambda)]^{j-1}(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0\alpha \to \mathbf{0},\ \text{as } j \to \infty.$$

It follows that $\mathbf{I} - (\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0\alpha$ is invertible, and thus $\lambda\mathbf{I} - \mathbf{T} - \mathbf{T}^0\alpha$ is also invertible. From (5.3.7), we obtain

$$\begin{aligned} \pi_0 &= x_1\mu\alpha\left(\lambda\mathbf{I} - \mathbf{T} - \mathbf{T}^0\alpha\right)^{-1} \\ &= x_1\mu\alpha\left[\mathbf{I} - (\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0\alpha\right]^{-1}(\lambda\mathbf{I} - \mathbf{T})^{-1} \\ &= x_1\mu\alpha\left\{\mathbf{I} + \sum_{j=1}^{\infty}[v(\lambda)]^{j-1}(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0\alpha\right\}(\lambda\mathbf{I} - \mathbf{T})^{-1} \\ &= x_1\frac{\mu}{\lambda}\frac{\lambda}{1 - v(\lambda)}\alpha(\lambda\mathbf{I} - \mathbf{T})^{-1} \\ &= K\beta, \end{aligned}$$

where $K = \lambda^{-1}x_1\mu$ is a constant to be determined by the normalization condition. Note that

$$1 - v(\lambda) = \alpha\left[\mathbf{I} + (\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}\right]\mathbf{e} = \lambda\alpha(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{e},$$

and it is easy to verify that $\beta\mathbf{e} = 1$. Using Theorem 5.2.3, we have

$$(x_k, \pi_k) = (x_{c-1}, \pi_{c-1})\mathbf{R}^{k-c+1}, \qquad k \geq c-1.$$

Substituting $\mathbf{R}$, given in (5.3.4), into the matrix geometric solution above, we obtain the last equation of (5.3.6). Now we need to get x_j, $j = 1, ..., c-1$, and K. It follows from the equilibrium equation $\mathbf{\Pi Q} = \mathbf{0}$ that

$$\begin{cases} 2\mu x_2 - \lambda x_1 = \mu x_1 - \pi_1\mathbf{T}^0 \\ (j+1)\mu x_{j+1} - \lambda x_j = j\mu x_j - \lambda x_{j-1} - \pi_j\mathbf{T}^0, \qquad j = 2, ..., c-1. \end{cases} \tag{5.3.8}$$

Substituting the relation $\mu x_1 = \lambda\pi_0\mathbf{e}$ into the first equation of (5.3.8), we have

$$2\mu x_2 - \lambda x_1 = \lambda\pi_0\mathbf{e} + \lambda\pi_0(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}\mathbf{e} = \lambda^2\pi_0(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{e}.$$

Taking the sum from $j = 2$ to $j = k$, $2 \leq k < c-1$, we get

$$\begin{aligned} (k+1)\mu x_{k+1} - \lambda x_k &= 2\mu x_2 - \lambda x_1 - \sum_{j=2}^{k}\pi_j\mathbf{T}^0 \\ &= \lambda^2\pi_0(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{e} + \pi_0\sum_{j=2}^{k}\left[\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^j\mathbf{T}\mathbf{e} \\ &= \lambda\pi_0\left[\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^k\mathbf{e}, \end{aligned}$$

which can be written as a recursive relation as

$$x_{k+1} = \frac{\lambda}{(k+1)\mu}x_k + \frac{\lambda}{(k+1)\mu}\pi_0\left[\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^k\mathbf{e}, \quad 1 \leq k \leq c-1.$$

Using this relation repeatedly, we obtain

$$x_j = \frac{1}{j!}\pi_0\sum_{i=1}^{j-1}i!\left(\frac{\lambda}{\mu}\right)^{j-i}\left[\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^i\mathbf{e} = K\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j\psi_j, \ 1 \leq j \leq c-1.$$

Finally, K is determined by the normalization condition.□

From Theorem 5.3.2, we can get the stationary performance measures of the M/M/c (SY, MV) system. The distribution of the queue length is given by

$$P\{L_v = 0\} = K, \qquad P\{L_v = j\} = x_j + \pi_j\mathbf{e}, \qquad j \geq 1,$$

and the distribution of the number of waiting customers is given by

$$
\begin{aligned}
P\{Q_v = 0\} =& \sum_{k=1}^{\infty} x_k = 1 - \frac{\rho}{1-\rho} x_{c-1} \\
& - \pi_0 \left[-\lambda \mathbf{T}^{-1} + \frac{\rho}{1-\rho} \left[\lambda(\lambda \mathbf{I} - \mathbf{T})^{-1}\right]^{c-1} (\rho \mathbf{I} - \lambda \mathbf{T}^{-1}) \right] \mathbf{e}, \\
P\{Q_v = j\} =& x_{c-1} \rho^{j+1} \\
& + \pi_0 \Big\{ \left(\lambda(\lambda \mathbf{I} - \mathbf{T})^{-1}\right)^j \\
& + \rho \left[\lambda(\lambda \mathbf{I} - \mathbf{T})^{-1}\right]^{c-1} \sum_{i=0}^{j} \rho^i \left[\lambda(\lambda \mathbf{I} - \mathbf{T})^{-1}\right]^{j-i} \mathbf{e} \Big\}, \; j \geq 1.
\end{aligned}
$$

For the waiting time, consider a customer arriving at state (k, h), $k = 0, 1, ..., c-1$, $1 \leq h \leq m$. This customer's waiting time is the residual life of a vacation. The probability that this waiting time is no more than x is the hth component of the vector

$$
\int_0^x \exp(\mathbf{T}t) dt \mathbf{T}^0, \qquad x > 0.
$$

If a customer arrives at state (k, h), $k \geq c, 1 \leq h \leq m$, his or her waiting time is the sum of the residual life of a vacation and $k - c$ i.i.d. exponential random variables with rate $c\mu$. Using the conditional argument, we obtain the the distribution function of the waiting time W_v:

$$
\begin{aligned}
W_v(x) =& 1 - \frac{\rho}{1-\rho} x_{c-1} e^{-c\mu(1-\rho)x} \\
& - \Big\{ \frac{\rho}{1-\rho} K\beta \left[\lambda(\lambda \mathbf{I} - \mathbf{T})^{-1}\right]^{c-1} \\
& \times (\lambda \mathbf{I} - \mathbf{T}) \left[(\lambda - c\mu)\mathbf{I} - \mathbf{T}\right]^{-1} \exp\{-c\mu(1-\rho)x\} \mathbf{e} \Big\} \\
& + \Big\{ K\beta \left(\mathbf{I} - \left[\lambda(\lambda \mathbf{I} - \mathbf{T})^{-1}\right]^{c-1} \left[(\lambda - c\mu)\mathbf{I} - \mathbf{T}\right]^{-1} \right) \\
& \times (\mathbf{I} - \lambda \mathbf{T})^{-1} \exp(\mathbf{T}x) \mathbf{e} \Big\}, \qquad x \geq 0.
\end{aligned}
$$

It can be proved that the number of waiting customers Q_v and the waiting time W_v follow the discrete and continuous PH distributions of order $m + 1$, respectively (see Tian and Li (2000)).

Obviously, the expressions for the distributions of the queue length and the waiting time are quite complex. Thus we cannot establish the stochastic decomposition relations for the queue length and the waiting time as in single server vacation systems. However, we can prove some conditional stochastic decomposition properties in the M/M/c (SY, MV) system. Define the conditional random variable

$$L_v^{(c)} = \{L_v - c | L_v \geq c, J = 0\}$$

as the number of waiting customers in the system, given that all servers are busy. Furthermore, from the PH distribution $(\alpha, \mathbf{T})$ of the vacation time, we build a PH random variable U of order m with the representation $(\gamma, \mathbf{T})$, where

$$\gamma = \beta \left[\lambda (\lambda \mathbf{I} - \mathbf{T})^{-1}\right]^{c-1} = \frac{1}{1 - v(\lambda)} \alpha \left[\lambda (\lambda \mathbf{I} - \mathbf{T})^{-1}\right]^{c}. \tag{5.3.9}$$

The mean of U is given by

$$E(U) = \beta \left[\lambda (\lambda \mathbf{I} - \mathbf{T})^{-1}\right]^{c-1} \left(-\mathbf{T}^{-1}\right) \mathbf{e}.$$

Theorem 5.3.3. If $\rho < 1$, $L_v^{(c)}$ in an M/M/c (SY, MV) system can be decomposed into the sum of two independent random variables $L_v^{(c)} = L_0^{(c)} + L_d$, where $L_0^{(c)}$ is the number of waiting customers in the system, given that all servers are busy, in a classical M/M/c queue and follows the geometric distribution with parameter ρ. L_d is the additional queue length due to the vacation effect and follows a discrete PH distribution of order m, with the irreducible representation $(\delta, \mathbf{S})$. Here,

$$\delta = \frac{\lambda}{\sigma} \rho \gamma (-\mathbf{T}^{-1}), \qquad \mathbf{S} = \lambda (\lambda \mathbf{I} - \mathbf{T})^{-1},$$

$$\delta_{m+1} = \frac{\rho}{\sigma} \left[\frac{1}{(c-1)!} \left(\frac{\lambda}{\mu}\right)^{c-1} \psi_{c-1} + \gamma \mathbf{e}\right], \qquad \mathbf{S}^0 = (\lambda \mathbf{I} - \mathbf{T})^{-1} \mathbf{T}^0,$$

and

$$\sigma = \rho \left[\frac{1}{(c-1)!} \left(\frac{\lambda}{\mu}\right)^{c-1} \psi_{c-1} + \gamma \mathbf{e} + \lambda E(U)\right]$$

is a constant. γ is the m -dimensional row vector determined by (5.3.9).

Proof: Since $\mathbf{R}$ is a lower triangular block-partitioned matrix, we can use a similar approach to that in Theorem 5.2.5. It follows from (5.3.6) that the probability that all servers are busy at an arbitrary time is

$$\begin{aligned}
P\{L_v \geq c, J = 0\} & \\
&= \sum_{j=c}^{\infty} x_j \\
&= x_{c-1} \sum_{j=c}^{\infty} \rho^{j-c+1} + \rho \pi_{c-1} \sum_{j=c}^{\infty} \sum_{i=0}^{j-c} \rho^i \left[\lambda (\lambda \mathbf{I} - \mathbf{T})^{-1} \right]^{j-c-i} \mathbf{e} \\
&= \frac{\rho}{1-\rho} x_{c-1} + \frac{\rho}{1-\rho} \pi_{c-1} \left[\mathbf{I} - \lambda (\lambda \mathbf{I} - \mathbf{T})^{-1} \right]^{-1} \mathbf{e} \\
&= \frac{K\rho}{1-\rho} \left\{ \frac{1}{(c-1)!} \left(\frac{\lambda}{\mu} \right)^{c-1} \psi_{c-1} + \gamma (\lambda \mathbf{I} - \mathbf{T}) (-\mathbf{T}^{-1}) \mathbf{e} \right\} \\
&= \frac{K\rho}{1-\rho} \left\{ \frac{1}{(c-1)!} \left(\frac{\lambda}{\mu} \right)^{c-1} \psi_{c-1} + \gamma \mathbf{e} + \lambda E(U) \right\} \\
&= \frac{K\sigma}{1-\rho}.
\end{aligned}$$

Thus the distribution of $L_v^{(c)}$ can be rewritten as

$$P\{L_v^{(c)} = j\} = P\{L_v = c + j | L_v \geq c, J = 0\} = \frac{1-\rho}{K\sigma} x_{j+c}, \qquad j \geq 0.$$

Taking the p.g.f. and using Theorem 5.3.2, we obtain

$$\begin{aligned}
L_v^{(c)}(z) &= \frac{1-\rho}{K\sigma} \sum_{j=c}^{\infty} z^{j-c} x_j \\
&= \frac{1-\rho}{K\sigma} \left\{ x_{c-1} \sum_{j=c}^{\infty} z^{j-c} \rho^{j-c+1} \right. \\
&\qquad \left. + \rho \pi_{c-1} \sum_{j=c}^{\infty} z^{j-c} \sum_{i=0}^{j-c} \rho^i \left[\lambda (\lambda \mathbf{I} - \mathbf{T})^{-1} \right]^{j-c-i} \mathbf{e} \right\} \\
&= \frac{1-\rho}{1-z\rho} \frac{1}{\sigma} \left\{ \frac{\rho}{(c-1)!} \left(\frac{\lambda}{\mu} \right)^{c-1} \psi_{c-1} \right. \\
&\qquad \left. + \rho\gamma \left[\mathbf{I} - z\lambda (\lambda \mathbf{I} - \mathbf{T})^{-1} \right]^{-1} \mathbf{e} \right\}. \qquad (5.3.10)
\end{aligned}$$

Note that, from (5.3.1), $L_0^{(c)}(z) = (1-\rho)(1-z\rho)^{-1}$ is the p.g.f. of the corresponding conditional random variable in the M/M/c queue. For the remaining factor of (5.3.10), we have

$$\begin{aligned}
&\frac{\rho}{\sigma}\left\{\frac{1}{(c-1)!}\left(\frac{\lambda}{\mu}\right)^{c-1}\psi_{c-1}+\gamma\left[\mathbf{I}-z\lambda\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\right]^{-1}\mathbf{e}\right\}\\
&\quad=\delta_{m+1}-\frac{\rho}{\sigma}\gamma\mathbf{e}+\frac{\rho}{\sigma}\gamma\left[\mathbf{I}-z\lambda\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\right]^{-1}\mathbf{e}\\
&\quad=\delta_{m+1}+\frac{\rho}{\sigma}\gamma\left\{-\mathbf{I}+\lambda z\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}+\mathbf{I}\right\}\left(\mathbf{I}-z\mathbf{S}\right)^{-1}\mathbf{e}\\
&\quad=\delta_{m+1}+z\delta(\mathbf{I}-z\mathbf{S})^{-1}\mathbf{S}^{0},
\end{aligned}$$

which is the p.g.f. of a PH distribution with $(\delta, \mathbf{S})$.□

For the conditional stochastic decomposition property, we have the following probability interpretation.

Remark 5.3.1: δ_{m+1} is the conditional probability that there is no waiting customer in the system when all the servers are busy. The additional queue length L_d is the number of customers arriving during a random interval U^* that follows the PH distribution of order m with the representation $(\gamma^*, \mathbf{T})$. Here,

$$\gamma^* = \frac{\rho}{\sigma}\beta\left[\lambda\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\right]^{c-1}(-\mathbf{T}^{-1}),$$

and δ in Theorem 5.3.3 can be written as

$$\delta = \frac{\lambda\rho E(U)}{\sigma E(U)}\beta\left[\lambda\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\right]^{c-1}(-\mathbf{T}^{-1}) = \frac{\lambda E(U)}{\sigma}\gamma^*.$$

Therefore, L_d is equal to the number of arrivals during the residual life of U with probability $p^* = \lambda E(U)\sigma^{-1}$ and is zero with probability $1-p^* = \delta_{m+1}$. The average number of waiting customers in the system, given that all the servers are busy, is given by

$$E(L_v^{(c)}) = \frac{1}{1-\rho} + \frac{\rho\left[\lambda^2 E(U^2) + 2\lambda E(U)\right]}{2\sigma}. \tag{5.3.11}$$

We can also prove the conditional stochastic decomposition property for the waiting time $W_v^{(c)}$. Define

$$W_v^{(c)} = \{W_v | L_v \geq c, J = 0\}$$

as the conditional waiting time, given that this customer arrives at a state where all the servers are busy.

Theorem 5.3.4. If $\rho < 1$, $W_v^{(c)}$ can be decomposed into the sum of two independent random variables, $W_v^{(c)} = W_0^{(c)} + W_d$, where $W_0^{(c)}$ is the corresponding conditional waiting time in a classical M/M/c queue, with the LST as in (5.3.2). W_d is the additional delay due to the vacation effect and follows a PH distribution of order m with the irreducible representation $(\delta, \mathbf{L})$, where

$$\mathbf{L} = c\mu\left(\lambda\mathbf{I} - \mathbf{T}\right)^{-1}\mathbf{T}, \qquad \mathbf{L}^0 = c\mu\left(\lambda\mathbf{I} - \mathbf{T}\right)^{-1}\mathbf{T}^0, \tag{5.3.12}$$

and δ is given in Theorem 5.3.3.

Proof: If a customer arrives at state $(j, 0)$, $j \geq c$, then his or her waiting time W_{cj} follows an Erlang distribution with parameters $j-c+1$ and $c\mu$, with the LST

$$W^*_{cj}(s) = \left(\frac{c\mu}{s + c\mu}\right)^{j-c+1}, \qquad j \geq c.$$

Thus the LST of $W_v^{(c)}$ can be written as

$$\begin{aligned}
&W_v^{*(c)}(s)\\
&\quad = \frac{1-\rho}{K\sigma}\sum_{j=c}^{\infty} x_j W^*_{cj}(s)\\
&\quad = \frac{1-\rho}{K\sigma}\left\{x_{c-1}\sum_{j=c}^{\infty}\rho^{j-c+1}\left(\frac{c\mu}{s+c\mu}\right)^{j-c+1}\right.\\
&\qquad \left. +\rho\pi_{c-1}\sum_{j=c}^{\infty}\sum_{i=0}^{j-c}\rho^i\left(\frac{c\mu}{s+c\mu}\right)^{j-c+1}\left[\lambda\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\right]^{j-c-i}\mathbf{e}\right\}\\
&\quad = \frac{c\mu(1-\rho)}{s+c\mu(1-\rho)}\frac{1}{\sigma}\left\{\frac{\rho}{(c-1)!}\left(\frac{\lambda}{\mu}\right)^{c-1}\psi_{c-1}\right.\\
&\qquad\qquad \left. +\rho\gamma\left[\mathbf{I}-\frac{c\mu}{s+c\mu}\lambda\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\right]^{-1}\mathbf{e}\right\}.
\end{aligned}$$

It follows from (5.3.2) that the first factor of the expression above is the LST of the corresponding conditional random variable $W_0^{(c)}$. For the second factor, we have

$$\frac{1}{\sigma}\left\{\frac{\rho}{(c-1)!}\left(\frac{\lambda}{\mu}\right)^{c-1}\psi_{c-1}+\rho\gamma\left[\mathbf{I}-\frac{c\mu}{s+c\mu}\lambda\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\right]^{-1}\mathbf{e}\right\}$$

$$=\frac{1}{\sigma}\left\{\frac{\rho}{(c-1)!}\left(\frac{\lambda}{\mu}\right)^{c-1}\psi_{c-1}+(s+c\mu)\rho\gamma\left[s\mathbf{I}-c\mu\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\mathbf{T}\right]^{-1}\mathbf{e}\right\}$$

$$=\delta_{m+1}-\frac{\rho}{\sigma}\gamma\mathbf{e}+\frac{1}{\sigma}(s+c\mu)\rho\gamma(s\mathbf{I}-\mathbf{L})^{-1}\mathbf{e}$$

$$=\delta_{m+1}+\frac{\lambda}{\sigma}\rho\gamma(-\mathbf{T}^{-1})(s\mathbf{I}-\mathbf{L})^{-1}c\mu\left(\lambda\mathbf{I}-\mathbf{T}\right)^{-1}\mathbf{T}^{0}$$

$$=\delta_{m+1}+\delta\left(s\mathbf{I}-\mathbf{L}\right)^{-1}\mathbf{L}^{0}.$$

□

We can interpret the conditional stochastic decomposition property and the additional delay W_d similarly to those of Theorem 5.3.3. The expected conditional waiting time of a customer given that he or she arrives at a state where all the servers are busy in the M/M/c (SY, MV) system, is

$$E(W_v^{(c)})=\frac{1}{c\mu(1-\rho)}+\frac{\rho\left[\lambda^2E(U^2)+2\lambda E(U)\right]}{2\sigma c\mu}=\frac{1}{c\mu}E(L_v^{(c)}).$$

5.3.2 Single Vacation and Setup Time Models

In a synchronous single vacation system, denoted by M/M/c (SY, SV), all servers take a single vacation together at a service completion instant when the system becomes empty. At the vacation termination instant, the servers either stay idle or serve the customers if any are present in the system. We again assume that the vacation time follows a PH distribution of order m with the representation $(\alpha, \mathbf{T})$, $\alpha\mathbf{e} = 1$. After each vacation, there are three possible cases: (i) If no customers are in the system, the c servers stay idle; (ii) if $1 \le j < c$ customers are in the system, then the j servers start serving the customers and the $c-j$ servers become idle; (iii) if $j \ge c$ customers are in the system, then all the c servers start serving the customers and $c-j$ customers are waiting in the line. As with the M/M/c (SY, MV) model developed in the previous section, $\{(L_v(t), J(t)), t \ge 0\}$ is a QBD process with the state space

$$\Omega = \{(k,j) : k \ge 0, 0 \le j \le m\},$$

where state $(0,0)$ represents case (i). The infinitesimal generator has the same structure as (5.3.3), where $\mathcal{A}_0$ is the square matrix of order $c(m+1)$, and $\mathcal{B}_1$ and $\mathcal{C}_0$ are the $(m+1)\times c(m+1)$ and $c(m+1)\times(m+1)$ matrices, respectively. The only difference from the M/M/c (SY, MV) system is in the following matrices:

$$\mathbf{A}_0 = \begin{bmatrix} -\lambda & \mathbf{0} \\ \mathbf{T}^0 & -\lambda\mathbf{I} - \mathbf{T} \end{bmatrix}, \qquad \mathbf{B}_1 = \begin{bmatrix} 0 & \mu\alpha \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{C}_0 = \lambda\mathbf{I}.$$

Other entry blocks of the infinitesimal matrix $\mathbf{Q}$ are the same as the M/M/c (SY, MV) system.

Another variation of the M/M/c type vacation model is the system with synchronous setup times, denoted by M/M/c (SY, SU). In such a system, whenever the system becomes empty at a service completion instant, all c servers are shut down or turned off. When the next customer arrives, the c servers are turned on and experience a set-up time before serving the customers. After the setup time, there are only two possible cases concerning the number of customers in the system: (i) $j \geq c$ and (ii) $1 \leq j \leq c$. In the first case, all the c servers start serving the customers, and in the second case, only the j servers start serving the customers and the $c - j$ servers become idle. The setup time, also denoted by V, follows the same PH distribution as in the (SY, SV) case. Now the QBD process $\{(L_v(t), J(t)), t \geq 0\}$ has the state space

$$\Omega = \{(0,0)\} \cup \{(k,j) : k \geq 1, 0 \leq j \leq m\},$$

where state $(0,0)$ is the state where all servers are turned off. When a customer arrives at state $(0,0)$, a PH setup time starts at phase j with probability $\alpha_j, 1 \leq j \leq m$, $\alpha = (\alpha_1, ..., \alpha_m)$. The infinitesimal generator has the same structure as (5.3.3) where $\mathcal{A}_0$ is the square matrix of order $m^* = (c-1)(m+1)+1$, and $\mathcal{B}_1$ and $\mathcal{C}_0$ are the $(m+1) \times m^*$ and $m^* \times (m+1)$ matrices, respectively. Now we have the following matrices:

$$\mathbf{A}_0 = -\lambda, \qquad \mathbf{B}_1 = \begin{pmatrix} \mu \\ \mathbf{0} \end{pmatrix}_{(m+1)\times 1}, \qquad \mathbf{C}_0 = \begin{pmatrix} 0 & \lambda\alpha \end{pmatrix}_{1\times(m+1)}.$$

Other entry blocks of the infinitesimal matrix $\mathbf{Q}$ are the same as in the M/M/c (SY, MV) system. Since both the M/M/c (SY,SV) and the M/M/c (SY, SU) have the same $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ matrices as in the M/M/c (SY, MV) treated in the previous section, they have the same rate matrix $\mathbf{R}$ of (5.3.4). However, we need to compute the boundary-state probabilities using (5.3.7) and (5.3.8). Similar to Theorem 5.3.2, we have the following theorems.

Theorem 5.3.5. If $\rho < 1$, the distribution of (L_v, J) in the M/M/c (SY,SV) system is given by

$$\begin{cases} \pi_j = K\beta\left[\lambda(\lambda\mathbf{I}-\mathbf{T})^{-1}\right]^j, & j \geq 0, \\ x_0 = \frac{1}{\lambda}K\beta T^0, & \\ x_j = K\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \varphi_j & 1 \leq j \leq c-1, \\ x_j = x_{c-1}\rho^{j-c+1} + \rho\pi_{c-1}\sum_{i=0}^{j-c}\rho^i\left[\lambda(\lambda\mathbf{I}-\mathbf{T})^{-1}\right]^{j-c-i}e, & j \geq c, \end{cases} \tag{5.3.13}$$

where

$$\beta = (\beta_1, \cdots, \beta_m) = \frac{\lambda}{1-v(\lambda)}\alpha(\lambda\mathbf{I}-\mathbf{T})^{-1}; \ \beta\mathbf{e} = 1,$$

$$\varphi_j = \beta\left\{\lambda^{-1}(\lambda\mathbf{I}-\mathbf{T}) + \sum_{i=0}^{j-1} i!\left[\mu(\lambda\mathbf{I}-\mathbf{T})^{-1}\right]^i\right\}\mathbf{e}, \qquad 1 \leq j \leq c-1,$$

$$K = \left\{\frac{\beta T^0}{\lambda} + \sum_{j=1}^{c-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j\varphi_j + \frac{\rho}{1-\rho}\frac{\left(\frac{\lambda}{\mu}\right)^{c-1}}{(c-1)!}\varphi_{c-1}\right.$$
$$\left. +\beta\left[\mathbf{I} + \frac{\rho}{1-\rho}\left(\lambda(\lambda\mathbf{I}-\mathbf{T})^{-1}\right)^{c-1}\right](\mathbf{I}-\lambda\mathbf{T}^{-1})\mathbf{e}\right\}^{-1}.$$

Proof: We solve the following equations for the boundary-state probabilities

$$\begin{cases} -\lambda x_0 + \pi_0\mathbf{T}^0 = 0 \\ \pi_0(-\lambda I + T) + \mu x_1\alpha = 0 \\ (j+1)\mu x_{j+1} - \lambda x_j = j\mu x_j - \lambda x_{j-1} - \pi_j T^0, & 1 \leq j \leq c-1. \end{cases}$$

Similarly to the proof of Theorem 5.3.2, if we use the matrix geometric solution and recursively solve these equations, we have (5.3.13). □

Theorem 5.3.6. If $\rho < 1$, the distribution of (L_v, J) in the M/M/c (SY,SU) system is given by

$$\begin{cases} \pi_j = K\alpha\left[\lambda(\lambda\mathbf{I}-\mathbf{T})^{-1}\right]^j, & j \geq 1, \\ x_j = K\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j h_j & 0 \leq j \leq c-1, \\ x_j = x_{c-1}\rho^{j-c+1} + \rho\pi_{c-1}\sum_{i=0}^{j-c}\rho^i\left[\lambda(\lambda\mathbf{I}-\mathbf{T})^{-1}\right]^{j-c-i}e, & j \geq c. \end{cases} \tag{5.3.14}$$

where

$$h_j = \alpha \sum_{i=0}^{j} i! \left[\mu(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^i \mathbf{e}. \qquad 0 \le j \le c-1,$$

$$K = \left\{ 1 + \lambda E(V) + \sum_{j=1}^{c-1} \frac{1}{j!} \left(\frac{\lambda}{\mu}\right)^j h_j \right.$$
$$\left. + \frac{\rho}{1-\rho} \left[\frac{\left(\frac{\lambda}{\mu}\right)^{c-1}}{(c-1)!} h_{c-1} + \lambda\alpha \left[\lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\right]^{c-2} (-\mathbf{T}^{-1})\mathbf{e} \right] \right\}^{-1}.$$

Proof: The equations for the boundary-state probabilities now become

$$\begin{cases} -\lambda x_0 + \mu x_1 = 0 \\ \lambda x_0 \alpha - \pi_1(\lambda\mathbf{I} - \mathbf{T}) = \mathbf{0} \\ -(\lambda+\mu)x_1 + 2\mu x_2 + \pi_1 \mathbf{T}^0 = 0 \\ (j+1)\mu x_{j+1} - \lambda x_j = j\mu x_j - \lambda x_{j-1} - \pi_j \mathbf{T}^0, \qquad 1 \le j \le c-1. \end{cases}$$

Using the same method of solving the equations as in the proof of Theorem 5.3.2 yields (5.3.14). □

From (5.3.13) and (5.3.14), we can obtain the stationary distributions for the queue length and the waiting time for both the M/M/c (SY, SV) and the M/M/c (SY, SU) systems. We can also prove the corresponding conditional stochastic decomposition properties. All these results are similar to Theorems 5.3.3 and 5.3.4.

As special cases of the PH distributed vacations, we present the examples with exponential vacations.

Example 1: M/M/c (SY, MV) with exponential vacations.

Assume that the vacation time V follows the exponential distribution with parameter θ and $V(x) = 1 - e^{-\theta x}, x \ge 0$. Then we have $v^*(s) = \theta(s+\theta)^{-1}$, $T = -\theta$, $T^0 = \theta$, $\alpha = 1$. The vector β in Theorem 5.3.2 is reduced to 1 and

$$\psi_j = \sum_{i=0}^{j-1} i! \left(\frac{\mu}{\lambda+\theta}\right)^i, \qquad j = 1, ..., c-1.$$

The distribution of (L_v, J) is given by

$$\pi_j = K\left(\frac{\lambda}{\lambda+\theta}\right)^j, \qquad j \geq 0$$

$$x_j = \begin{cases} K\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \sum_{i=0}^{j-1} i!\left(\frac{\mu}{\lambda+\theta}\right)^i, & 1 \leq j \leq c-1, \\ x_{c-1}\rho^{j-c+1} + \rho\pi_{c-1}\sum_{i=0}^{j-c}\rho^i\left(\frac{\lambda}{\lambda+\theta}\right)^{j-c-i}, & j \geq c \end{cases}$$

where

$$K = \left\{\sum_{j=1}^{c-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \psi_j + \frac{\rho}{1-\rho}\frac{\left(\frac{\lambda}{\mu}\right)^{c-1}}{(c-1)!}\psi_{c-1} \right.$$
$$\left. + \frac{\lambda+\theta}{\theta}\left[1 + \frac{\rho}{1-\rho}\left(\frac{\lambda}{\lambda+\theta}\right)^{c-1}\right]\right\}^{-1}.$$

In the conditional stochastic decomposition expression, we have

$$\sigma = \rho\left\{\frac{1}{(c-1)!}\left(\frac{\lambda}{\mu}\right)^j \psi_{c-1} + \left(\frac{\lambda}{\lambda+\theta}\right)^{c-1}\frac{\lambda+\theta}{\theta}\right\}.$$

The additional queue length L_d follows the modified geometric distribution

$$P\{L_d = k\} = \begin{cases} \frac{\rho}{\sigma}\left\{\frac{1}{(c-1)!}\left(\frac{\lambda}{\mu}\right)^{c-1}\psi_{c-1} + \left(\frac{\lambda}{\lambda+\theta}\right)^{c-1}\right\}, & k = 0, \\ \frac{\rho}{\sigma}\left(\frac{\lambda}{\lambda+\theta}\right)^c\left(\frac{\lambda}{\lambda+\theta}\right)^{k-1}, & k \geq 1. \end{cases}$$

Note that L_d is the mixture of two random variables:

$$L_d = (1-p^*)X_0 + p^*X_d,$$

where X_0 has the probability density concentrated at the origin and X_d follows the geometric distribution with parameter $\lambda(\lambda+\theta)^{-1}$. That is,

$$P\{X_d = j\} = \left(1 - \frac{\lambda}{\lambda+\theta}\right)\left(\frac{\lambda}{\lambda+\theta}\right)^j, \qquad j \geq 0,$$

and

$$p^* = \frac{\rho}{\sigma}\left(\frac{\lambda}{\lambda+\theta}\right)^{c-1}\frac{\lambda+\theta}{\theta}.$$

From Theorem 5.3.4, it follows that the additional delay W_d follows the modified exponential distribution, with the distribution function

$$W_d(x) = 1 - \frac{\rho}{\sigma}\left(\frac{\lambda}{\theta}\right)\left(\frac{\lambda}{\lambda+\theta}\right)^{c-1} e^{-\frac{\theta c\mu}{\lambda+\theta}x}, \qquad x \geq 0.$$

Finally, given that all the servers are busy, the expected values of $L_v^{(c)}$ and $W_v^{(c)}$ are given, respectively, by

$$E(L_v^{(c)}) = \frac{1}{1-\rho} + \frac{\rho}{\sigma}\left(\frac{\lambda}{\lambda+\theta}\right)^{c-1}\frac{\lambda(\lambda+\theta)}{\theta^2},$$
$$E(W_v^{(c)}) = \frac{1}{c\mu(1-\rho)} + \frac{\rho}{\sigma}\left(\frac{\lambda}{\lambda+\theta}\right)^{c-1}\frac{\lambda(\lambda+\theta)}{c\mu\theta^2}.$$

Example 2: M/M/c (SY, SV) with exponential vacations.

For the exponential vacation time with parameter θ in an M/M/c (SY, SV) queue, we have

$$\varphi_j = \frac{\lambda+\theta}{\lambda} + \sum_{i=1}^{j-1} i!\left(\frac{\mu}{\lambda+\theta}\right)^i, \qquad 1 \leq j \leq c-1.$$

Thus the distribution of (L_v, J) is given by

$$\pi_j = K\left(\frac{\lambda}{\lambda+\theta}\right)^j, \qquad j \geq 0,$$
$$x_0 = \frac{\theta}{\lambda}K$$
$$x_j = \begin{cases} K\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \varphi_j & 1 \leq j \leq c-1 \\ x_{c-1}\rho^{j-c+1} + \rho\pi_{c-1}\sum_{i=0}^{j-c}\rho^i\left(\frac{\lambda}{\lambda+\theta}\right)^{j-c-i}, & j \geq c \end{cases}$$

where

$$K = \left\{\frac{\theta}{\lambda} + \sum_{j=1}^{c-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \varphi_j + \frac{\rho}{1-\rho}\frac{\left(\frac{\lambda}{\mu}\right)^{c-1}}{(c-1)!}\varphi_{c-1} + \frac{\lambda+\theta}{\theta}\left[1 + \frac{\rho}{1-\rho}\left(\frac{\lambda}{\lambda+\theta}\right)^{c-1}\right]\right\}^{-1}.$$

Now, replacing ψ_{c-1} with φ_{c-1} in σ, we can obtain the conditional stochastic decomposition expression and the distributions of L_d and W_d, which have the same forms as in Example 1.

Example 3: M/M/c (SY, SU) with exponential setup times.

Assume that the setup time follows the exponential distribution with parameter θ. In this case, vector α is reduced to 1. From Theorem 5.2.6, we have

$$h_j = \sum_{i=0}^{j} i! \left(\frac{\mu}{\lambda+\theta}\right)^i, \qquad 0 \le j \le c-1.$$

Therefore, the distribution of (L_v, J) is given by

$$\pi_j = K\left(\frac{\lambda}{\lambda+\theta}\right)^j, \qquad j \ge 1,$$

$$x_j = \begin{cases} K\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \sum_{i=0}^{j} i!\left(\frac{\mu}{\lambda+\theta}\right)^i, & 0 \le j \le c-1, \\ x_{c-1}\rho^{j-c+1} + \rho\pi_{c-1}\sum_{i=0}^{j-c}\rho^i\left(\frac{\lambda}{\lambda+\theta}\right)^{j-c-i}, & j \ge c. \end{cases}$$

Similarly to the examples above, replacing ψ_{c-1} with h_{c-1} gives all the corresponding results as in Example 1.

5.4 M/M/c Queue with Asynchronous Vacations

5.4.1 Multiple Vacation Model

In an M/M/c system with arrival rate λ and service rate μ, any server starts a vacation as long as there is no waiting customer in the system at the service completion. At a server's vacation termination instant, if there is no waiting customer, the server takes another vacation; and if there are waiting customers, the server resumes serving the customers. Since the servers take vacations individually and independently, this system is called the *asynchronous multiple vacation model*, denoted by M/M/c (AS, MV). This type of vacation model was studied by Levy and Yechiali (1976), Vinod (1986), and Tian and Li (1999). Assume that the vacation time follows the exponential distribution with parameter θ and that the interarrival times, the service times, and the vacation times are mutually independent. Let $L_v(t)$ be the number of customers in the system at time t, and, and let $J(t)$ be the number of busy servers. According to the (AS, MV) policy, the server is either busy or on vacation. Thus $\{(L_v(t), J(t)), t \ge 0\}$ is a QBD process with the state space

$$\Omega = \{(k,j) : 0 \le k \le c-1, 0 \le j \le k\} \cup \{(k,j) : k \ge c, 0 \le j \le c\}.$$

Using the lexicographical sequence for the states, the infinitesimal generator can be written in the block-partitioned form as in (5.3.3) where $\mathcal{A}_0$ is the square matrix of order $c^* = \frac{1}{2}c(c+1)$, and $\mathcal{B}_1$ and $\mathcal{C}_0$ are the $(c+1) \times c^*$ and $c^* \times (c+1)$ matrices, respectively. These matrices can be written as

$$\mathcal{A}_0 = \begin{bmatrix} A_0 & \mathbf{C}_0 & & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C} & & & \\ & \mathbf{B}_2 & \mathbf{A}_2 & C & & \\ & & \ddots & \ddots & \ddots & \\ & & & \mathbf{B}_{c-2} & \mathbf{A}_{c-2} & \mathbf{C}_{c-2} \\ & & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} \end{bmatrix},$$

$$\mathcal{B}_1 = (\mathbf{0}, \mathbf{B}_c), \qquad \mathcal{C}_0 = \begin{pmatrix} \mathbf{0} \\ \mathbf{C}_{c-1} \end{pmatrix}, \tag{5.4.1}$$

where $A_0 = -\lambda$, $\mathbf{C}_0 = (\lambda, 0)$, and $\mathbf{B}_1 = (0, \mu)^T$. For $\mathbf{A}_k$, $1 \leq k \leq c-1$, we have

$$\mathbf{A}_k = \begin{bmatrix} -h_0 & c\theta & & & \\ & -h_1 & (c-1)\theta & & \\ & & \ddots & \ddots & \\ & & & -h_{k-1} & (c-k+1)\theta \\ & & & & -(\lambda+k\mu) \end{bmatrix}_{(k+1)\times(k+1)},$$

$$1 \leq k \leq c-1,$$

where $h_k, 0 \leq k \leq c$, is defined as

$$h_k = h_k(\lambda, \theta, \mu) = \lambda + k\mu + (c-k)\theta, \qquad 0 \leq k \leq c.$$

$\mathbf{B}_k$ and $\mathbf{C}_k$ are the $(k+1)\times k$ and $(k+1)\times(k+2)$ matrices, respectively, $1 \leq k \leq c-1$, and are written as

$$\mathbf{B}_k = \begin{bmatrix} 0 & & & \\ & \mu & & \\ & & \ddots & \\ & & & (k-1)\mu \\ 0 & 0 & \cdots & k\mu \end{bmatrix}_{(k+1)\times k},$$

$$\mathbf{C}_k = \begin{bmatrix} \lambda & & & 0 \\ & \lambda & & 0 \\ & & \ddots & \vdots \\ & & \lambda & 0 \end{bmatrix}_{(k+1)\times(k+2)}.$$

Finally, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in the infinitesimal generator (5.3.3) are the square matrices of order $c+1$ and are given by

$$\mathbf{B} = diag(0, \mu, 2\mu, \cdots, c\mu), \ \ \mathbf{C} = \lambda\mathbf{I},$$

$$\mathbf{A} = \begin{bmatrix} -h_0 & c\theta & & & \\ & -h_1 & (c-1)\theta & & \\ & & \ddots & \ddots & \\ & & & -h_{c-1} & \theta \\ & & & & -h_c \end{bmatrix}.$$

To find the minimum nonnegative solution to the matrix equation (5.2.3), we need the following lemma.

Lemma 5.4.1. If $\rho = \lambda(c\mu)^{-1} < 1$, the equation

$$k\mu z^2 - [\lambda + k\mu + (c-k)\theta]z + \lambda = 0, \qquad 1 \le k \le c,$$

has two roots, namely, $r_k < r_k^*$ and $0 < r_k < 1$, $r_k^* \ge 1$.

Proof: It is easy to verify that the equation has two real roots which are

$$r_k^*, r_k = \frac{\lambda + k\mu + (c-k)\theta \pm \sqrt{[\lambda + k\mu + (c-k)\theta]^2 - 4\lambda k\mu}}{2k\mu}.$$

Note that

$$\begin{aligned} [\lambda - k\mu + (c-k)\theta]^2 &\le [\lambda + k\mu + (c-k)\theta]^2 - 4\lambda k\mu \\ &\le [\lambda + k\mu + (c-k)\theta]^2, \quad \text{if } \lambda \ge k\mu, \\ [k\mu - \lambda + (c-k)\theta]^2 &\le [\lambda + k\mu + (c-k)\theta]^2 - 4\lambda k\mu \\ &\le [\lambda + k\mu + (c-k)\theta]^2, \quad \text{if } \lambda < k\mu. \end{aligned}$$

Substituting these estimations into the expressions r_k^*, and r_k, we obtain $0 < r_k < 1$ and $r_k^* > 1$, $1 \le k \le c-1$. Finally, if $k = c$, we have $r_c = \rho < 1$ and $r_c^* = 1$. □

Theorem 5.4.1. If $\rho < 1$, the matrix equation (5.2.3) has the minimum nonnegative solution

$$\mathbf{R} = \begin{bmatrix} r_0 & r_{01} & \cdots & r_{0c} \\ & r_1 & \cdots & r_{1c} \\ & & \cdots & \vdots \\ & & & r_c \end{bmatrix}, \tag{5.4.2}$$

where $r_0 = \lambda(\lambda + c\theta)^{-1}$, and r_k, $1 \le k \le c-1$, are given in Lemma 5.4.1, and $r_c = \rho$. The nondiagonal entries satisfy the recursive relation

$$j\mu \sum_{i=k}^{j} r_{ki} r_{ij} + (c-j+1)\theta r_{k,j-1} - [\lambda + j\mu + (c-j)\theta] r_{kj} = 0,$$
$$0 \le k \le c-1, \quad k+1 \le j \le c, \tag{5.4.3}$$

where $r_{jj} = r_j$, $0 \le j \le c$, and $sp(\mathbf{R}) < 1$.

Proof: Since $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are all upper triangular matrices, the solution to (5.2.3) should also be an upper triangular matrix with the same structure as in (5.4.2). Thus the entries of $\mathbf{R}^2$ are given by

$$\begin{aligned} (\mathbf{R}^2)_{jj} &= r_j^2, && 0 \le j \le c, \\ (\mathbf{R}^2)_{kj} &= \sum_{i=k}^{j} r_{ki} r_{ij}, && 0 \le k \le c-1, \quad k < j \le c. \end{aligned}$$

Substituting $\mathbf{R}$ and $\mathbf{R}^2$ into (5.2.3), we have

$$\begin{cases} \lambda - (\lambda + c\theta) r_0 = 0, \\ k\mu r_k^2 - [\lambda + k\mu + (c-k)\theta] r_k + \lambda = 0, \qquad 1 \le k \le c, \\ j\mu \sum_{i=k}^{j} r_{ki} r_{ij} + (c-j+1)\theta r_{k,j-1} - [\lambda + j\mu + (c-j)\theta] r_{kj} = 0, \\ \qquad 0 \le k \le c-1, \ k+1 \le j \le c. \end{cases}$$

The first equation gives $r_0 = \lambda(\lambda + c\theta)^{-1}$. From Lemma 5.4.1, to obtain the minimum nonnegative solution, we take $r_k < 1$ as the root of the quadratic equation. The last equation gives the recursive relation (5.4.3). Clearly, the spectral radius of $\mathbf{R}$ satisfies

$$sp(\mathbf{R}) = \max\left\{ \frac{\lambda}{\lambda + c\theta}, r_1, \cdots, r_{c-1}, \rho \right\} < 1.$$

□

Lemma 5.4.2. Rate matrix $\mathbf{R}$ satisfies $\mathbf{R}\mathbf{T}^0 = \lambda \mathbf{e}$, where

$$\mathbf{T}^0 = \mathbf{B}\mathbf{e} = (0, \mu, \cdots, c\mu)^T$$

is the m -dimensional column vector.

Proof: Note that $\mathbf{A}\mathbf{e} = -(\lambda \mathbf{e} + \mathbf{T}^0), \mathbf{B}\mathbf{e} = \mathbf{T}^0$, and $\mathbf{C}\mathbf{e} = \lambda \mathbf{e}$. Multiplying both sides of (5.2.3) by $\mathbf{e}$, we obtain

$$\mathbf{R}^2\mathbf{T}^0 - \mathbf{R}(\lambda \mathbf{e} + \mathbf{T}^0) + \lambda \mathbf{e} = \mathbf{0},$$
$$(\mathbf{I} - \mathbf{R})(\lambda \mathbf{e} - \mathbf{R}\mathbf{T}^0) = \mathbf{0}.$$

Since $\mathbf{I}-\mathbf{R}$ is invertible, we have $\mathbf{RT}^0=\lambda\mathbf{e}$. □

Using (5.4.3), we can recursively compute the nondiagnoal entries from the entries on the diagonal. In (5.4.3), setting $j=k+1$, we have

$$(k+1)\mu(r_k r_{k,k+1}+r_{k,k+1}r_{k+1}-[\lambda+(k+1)\mu+(c-k-1)\theta]r_{k,k+1} \\ =-(c-k)r_k, \qquad 0\le k\le c-1.$$

That is

$$\{\lambda+(k+1)\mu+(c-k-1)\theta-(k+1)\mu r_{k+1}-(k+1)\mu r_k\}\, r_{k,k+1} \\ =(c-k)\theta r_k.$$

Note that

$$\lambda+(k+1)\mu+(c-k-1)\theta-(k+1)\mu r_{k+1}=(k+1)\mu r^*_{k+1}.$$

Substituting this expression into the previous one, we obtain

$$r_{k,k+1}=\left(\frac{c-k}{k+1}\right)\left(\frac{\theta}{\mu}\right)\frac{r_k}{r^*_{k+1}-r_k}, \qquad 0\le k\le c-1.$$

In (5.4.3), letting $j=k+2,k+3,...$ and using similar recursive computation, we have

$$r_{k,k+2}=\frac{(c-k)(c-k-1)}{(k+1)(k+2)}\left(\frac{\theta}{\mu}\right)^2\frac{r_k r^*_{k+2}}{D_{k,k+2}}, \qquad 0\le k\le c-2,$$

$$r_{k,k+3}=\frac{(c-k)(c-k-1)(c-k-2)}{(k+1)(k+2)(k+3)}\left(\frac{\theta}{\mu}\right)^3\frac{r_k r^*_{k+3}(r^*_{k+2}r^*_{k+3}-r_k r_{k+1})}{D_{k,k+3}}, \\ 0\le k\le c-3,$$

where

$$D_{kn}=\prod_{n\ge j>i\ge k}(r^*_j-r_i), \qquad n>k.$$

Since (5.4.3) is a nonlinear double-subscript recursive relation, it is difficult to find a general expression for r_{kj}. However, we can follow a specific sequence to recursively compute these nondiagonal entries. This sequence starting with the diagonal entries is illustrated in Figure 5.4.1 for a $c=4$ case.

$$\begin{array}{ccccc} r_0 \rightarrow & r_{01} \rightarrow & r_{02} \rightarrow & r_{03} \rightarrow & r_{04} \\ & \uparrow & \uparrow & \uparrow & \uparrow \\ & r_1 \rightarrow & r_{12} \rightarrow & r_{13} \rightarrow & r_{14} \\ & & \uparrow & \uparrow & \uparrow \\ & & r_2 \rightarrow & r_{23} \rightarrow & r_{24} \\ & & & \uparrow & \uparrow \\ & & & r_3 \rightarrow & r_{34} \\ & & & & \uparrow \\ & & & & \rho \end{array}$$

Figure 5.4.1 The sequence of computing nondiagonal entries for the c=4 case.

If $\rho < 1$, let $\{L_v, J\}$ be the stationary limit of the stochastic process $\{L_v(t), J(t)\}$ of an M/M/c (AS, MV) system as $t \to \infty$. Denote the stationary probability by

$$\pi_{kj} = P\{L_v = k, J = j\} = \lim_{t\to\infty} P\{L_v(t) = k, J(t) = j\}, \ (k, j) \in \Omega.$$

To accommodate the block-partitioned structure of $\mathbf{Q}$, we express the distribution of $\{L_v, J\}$ as segment vectors as

$$\pi_0 = \pi_{00}, \quad \pi_1 = (\pi_{10}, \pi_{11}), \cdots, \pi_k = (\pi_{k0}, \pi_{k,1}, \cdots, \pi_{k,k}), \quad 0 \le k \le c,$$

and when $k \ge c$, all π_k's are $(c + 1)$ -dimensional row vectors

$$\pi_k = (\pi_{k0}, \pi_{k,1}, \cdots, \pi_{k,c}).$$

Now the square matrix $B[\mathbf{R}]$ of order $\frac{1}{2}(c+1)(c+2)$ is constructed as

$$B[\mathbf{R}] = \begin{bmatrix} \mathcal{A}_0 & \mathcal{C}_0 \\ \mathcal{B}_1 & \mathbf{RB} + \mathbf{A} \end{bmatrix}$$
$$= \begin{bmatrix} A_0 & \mathbf{C}_0 & & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C}_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} & \mathbf{C}_{c-1} \\ & & & & \mathbf{B}_c & \mathbf{RB} + \mathbf{A} \end{bmatrix}, \qquad (5.4.4)$$

which is an aperiodic and irreducible infinitesimal generator with finite dimension. Hence, the linear equation system

$$(\pi_0, \pi_1, \cdots, \pi_c) B[\mathbf{R}] = \mathbf{0} \qquad (5.4.5)$$

must have positive solutions. For instance, the stationary probability vector of the infinitesimal generator $B[\mathbf{R}]$ multiplied by any constant $K > 0$ is a positive solution to (5.4.5).

Theorem 5.4.2. If $\rho < 1$, the distribution of (L_v, J) is given by

$$\pi_k = K\beta_k, \qquad 0 \le k \le c,$$
$$\pi_k = K\beta_c \mathbf{R}^{k-c}, \quad k \ge c,$$

where $\beta_k, 0 \le k \le c$ is the positive solution to (5.4.5), and the constant K is

$$K = \left\{ \sum_{k=0}^{c-1} \beta_k \mathbf{e} + \beta_c (\mathbf{I} - \mathbf{R})^{-1} \mathbf{e} \right\}^{-1}.$$

Proof: Using Theorem 5.2.3 immediately gives the results.□

For the stationary probability vectors, there exists the following relation.

Theorem 5.4.3. If $\rho < 1$, the stationary probability vectors satisfy

$$\lambda \pi_k \mathbf{e} = \pi_{k+1} \mathbf{T}_{k+1}^0$$

where

$$\mathbf{T}_k^0 = (0, \mu, \cdots, k\mu)^T, \ 0 \le k \le c-1,$$
$$\mathbf{T}_k^0 = (0, \mu, \cdots, c\mu)^T, \ k \ge c.$$

Proof: Let $\Pi = (\pi_0, \pi_1, \cdots)$. The equilibrium equation $\mathbf{\Pi Q} = \mathbf{0}$ gives

$$\begin{cases} \pi_1 \mathbf{B}_1 + \lambda \mathbf{A}_0 = 0, \\ \pi_{k-1} \mathbf{C}_{k-1} + \pi_k \mathbf{A}_k + \pi_{k+1} \mathbf{B}_{k+1} = \mathbf{0}, \quad 1 \le k \le c-1 \\ \pi_{c-1} \mathbf{C}_{c-1} + \pi_c \mathbf{A} + \pi_{c+1} \mathbf{B}_{c+1} = \mathbf{0}, \\ \pi_{k-1} \mathbf{C} + \pi_k \mathbf{A} + \pi_{k+1} \mathbf{B} = \mathbf{0}, \quad k \ge c+1. \end{cases}$$

Using

$$\mathbf{C}_k \mathbf{e} = \lambda \mathbf{e}, \quad \mathbf{A}_k \mathbf{e} = -(\lambda \mathbf{e} + \mathbf{T}_k^0), \quad \mathbf{B}_k \mathbf{e} = \mathbf{T}_k^0.$$

and right-multiplying both sides of the equilibrium equations by $\mathbf{e}$, we obtain

$$\lambda \pi_k \mathbf{e} - \pi_{k+1} \mathbf{T}_{k+1}^0 = 0.$$

□

It is possible to solve (5.4.5) numerically. However, the computation is quite complex. To compare the M/M/c (AS, MV) system with the classical M/M/c system, we define the conditional random variables. Let $L_v^{(c)} = \{L_v - c | J = c\}$ be the number of waiting customers in the system given that all the servers are busy in the M/M/c (AS,MV) system. Rewrite the vector β_c and the rate matrix $\mathbf{R}$, respectively, as

$$\beta_c = (\beta_{c0}, \beta_{c1}, ..., \beta_{cc}) = (\delta, \beta_{cc}),$$
$$\mathbf{R} = \begin{pmatrix} \mathbf{H} & \eta \\ \mathbf{0} & \rho \end{pmatrix}, \tag{5.4.6}$$

where $\delta = (\beta_{c0}, \beta_{c,1}, \cdots, \beta_{c,c-1})$ is a c -dimensional row vector. Comparing with (5.4.2), we find that $\mathbf{H}$ is a $c \times c$ square matrix and η is an $c \times 1$ column vector as follows:

$$\mathbf{H} = \begin{bmatrix} r_0 & r_{01} & \cdots & r_{0,c-1} \\ & r_1 & \cdots & r_{1,c-1} \\ & & \ddots & \vdots \\ & & & r_{c-1} \end{bmatrix}, \qquad \eta = \begin{bmatrix} r_{0c} \\ r_{1c} \\ \vdots \\ r_{c-1,c} \end{bmatrix}.$$

Obviously, the spectral radius of $\mathbf{H}$, $sp(\mathbf{H})$ is less than 1.

The following theorems show the relationship between the vacation model and the classical M/M/c model in terms of the conditional queue length and the conditional waiting time.

Theorem 5.4.4. If $\rho < 1$, $L_v^{(c)}$ in an M/M/c (AS, SV) system can be decomposed into the sum of two independent random variables,

$$L_v^{(c)} = L_0^{(c)} + L_d,$$

where $L_0^{(c)}$ is the corresponding random variable in the classical M/M/c system and has the geometric distribution of (5.3.1) and L_d is the additional queue length due to the vacation effect and follows the PH distribution of order c,

$$P\{L_d = k\} = \begin{cases} \frac{1}{\sigma}\beta_{cc}, & k = 0, \\ \frac{1}{\sigma}\delta\mathbf{H}^{k-1}\eta, & k \geq 1, \end{cases} \tag{5.4.7}$$

where

$$\sigma = \beta_{cc} + \delta(\mathbf{I} - \mathbf{H})^{-1}\eta.$$

Proof: Based on the triangular structure of $\mathbf{R}$ in (5.4.6) and the matrix geometric solution, we have

$$\pi_{kc} = K\beta_{cc}\rho^{k-c} + K\delta \sum_{j=0}^{k-c-1} \rho^j \mathbf{H}^{k-c-1-j}\eta, \qquad k \geq c.$$

Using this expression, we get the probability of the conditional event:

$$\begin{aligned} P\{J=c\} =& \sum_{k=c}^{\infty} \pi_{kc} \\ =& K\beta_{cc} \sum_{k=c}^{\infty} \rho^{k-c} + K\delta \sum_{k=c+1}^{\infty} \sum_{j=0}^{k-c-1} \rho^{j} \mathbf{H}^{k-c-1-j} \eta \\ =& \frac{K}{1-\rho} \left\{ \beta_{cc} + \delta(\mathbf{I}-\mathbf{H})^{-1}\eta \right\} = \frac{K}{1-\rho}\sigma. \end{aligned}$$

Thus the distribution of $L_v^{(c)}$ is

$$\begin{aligned} P\{L_v^{(c)}=k\} =& P\{L_v = k+c | J=c\} \\ =& \frac{1-\rho}{K\sigma} \pi_{k+c,c} \\ =& \frac{1-\rho}{\sigma} \left\{ \beta_{cc}\rho^{k} + \delta \sum_{j=0}^{k-1} \rho^{j} \mathbf{H}^{k-1-j} \eta \right\}, \qquad k \geq 0. \end{aligned}$$

Taking the p.g.f. of $L_v^{(c)}$, we have

$$\begin{aligned} L_v^{(c)}(z) &= \sum_{k=0}^{\infty} z^k P\{L_v^{(c)}=k\} \\ &= \frac{1-\rho}{\sigma} \left\{ \beta_{cc} \sum_{k=0}^{\infty} z^k \rho^k + \delta \sum_{k=1}^{\infty} z^k \sum_{j=0}^{k-1} \rho^{j} \mathbf{H}^{k-1-j} \eta \right\} \\ &= \frac{1-\rho}{1-z\rho} \frac{1}{\sigma} \left\{ \beta_{cc} + z\delta(\mathbf{I}-z\mathbf{H})^{-1}\eta \right\} \\ &= L_0^{(c)}(z) L_d(z), \end{aligned}$$

where

$$L_d(z) = \frac{1}{\sigma} \left\{ \beta_{cc} + z\delta(\mathbf{I}-z\mathbf{H})^{-1}\eta \right\}.$$

Expanding $L_d(z)$ as a power series, we obtain (5.4.7). □

Note that $\mathbf{H}$ may not be a stochastic submatrix. Sengupta (1991) proved that the probability distribution of (5.4.7) must be a discrete PH distribution of order c and provided a method of constructing the PH expression for the distribution. From Theorem 5.4.4, we find that the expected value of $L_v^{(c)}$ is

$$E(L_v^{(c)}) = \frac{1}{1-\rho} + \frac{1}{\sigma}\delta(\mathbf{I}-\mathbf{H})^{-2}\eta.$$

Define the conditional waiting time $W_v^{(c)} = \{W_v | J = c\}$. We have the following theorem for the conditional stochastic decomposition property of the waiting time.

Theorem 5.4.5. If $\rho < 1$, $W_v^{(c)}$ in an M/M/c (AS, MV) system can be decomposed into the sum of two independent random variables,

$$W_v^{(c)} = W_0^{(c)} + W_d.$$

where $W_0^{(c)}$ is the corresponding conditional waiting time in a classical M/M/c system without vacations and follows an exponential distribution with parameter $c\mu(1-\rho)$. W_d is the additional delay due to the vacation effect and follows a matrix exponential distribution

$$P\{W_d \le x\} = 1 - \frac{1}{\sigma}\delta \exp\{-c\mu(\mathbf{I}-\mathbf{H})x\}(\mathbf{I}-\mathbf{H}))^{-1}\eta, \qquad x \ge 0. \quad (5.4.8)$$

Proof: Assume that a customer arrives at state (k, c) for $k \ge c$. If we condition on this event, the customer's waiting time, denoted by W_{vk}, has the LST

$$W_{vk}^*(s) = \left(\frac{c\mu}{s + c\mu}\right)^{k-c+1},$$

for $k \ge c$. The LST of $W_v^{(c)}$ is given by

$$\begin{aligned}
W_v^{*(c)}(s) &= \sum_{k=c}^{\infty} P\{L_v^{(c)} = k\} W_{vk}^*(s) \\
&= \frac{1-\rho}{\sigma}\left\{\beta_{cc}\sum_{k=c}^{\infty}\rho^{k-c}\left(\frac{c\mu}{s+c\mu}\right)^{k-c+1}\right. \\
&\qquad \left. + \delta \sum_{k=c+1}^{\infty}\left(\frac{c\mu}{s+c\mu}\right)^{k-c+1}\sum_{j=0}^{k-c-1}\rho^j \mathbf{H}^{k-c-1-j}\eta\right\} \\
&= \frac{c\mu(1-\rho)}{s + c\mu(1-\rho)}\frac{1}{\sigma}\left\{\beta_{cc} + \delta\left(\mathbf{I} - \frac{c\mu}{s+c\mu}\mathbf{H}\right)^{-1}\eta\right\} \\
&= \frac{c\mu(1-\rho)}{s + c\mu(1-\rho)}\frac{1}{\sigma}\left\{\beta_{cc} + c\mu\delta\left(s\mathbf{I} - c\mu(\mathbf{H}-\mathbf{I})\right)^{-1}\eta\right\} \\
&= W_0^*(s) W_d^*(s),
\end{aligned}$$

where

$$W_d^*(s) = \frac{1}{\sigma}\left\{\beta_{c0} + c\mu\delta\left(s\mathbf{I} - c\mu(\mathbf{H}-\mathbf{I})\right)^{-1}\eta\right\}.$$

It follows from $W_d^*(s)$ that the distribution function of W_d can be written as (5.4.8). □

From (5.4.8), the expected value of $W_v^{(c)}$ is given by

$$E(W_v^{(c)}) = \frac{1}{c\mu(1-\rho)} + \frac{1}{c\mu\sigma}\delta(\mathbf{I}-\mathbf{H})^{-2}\eta = \frac{1}{c\mu}E(L_v^{(c)}).$$

5.4.2 Single Vacation or Setup Time Model

We now consider a system with asynchronous single vacation policy, denoted by M/M/c (AS, SV). In this system, any server who finds no waiting customer at his or her service completion instant takes only one vacation and then either serves a customer, if any, or stays idle. Therefore the server can be in one of three possible states: serving a customer, taking a vacation, or staying idle. Assume that the vacation time follows an exponential distribution with parameter θ and is independent of the service time and the interarrival time.

$L_v(t)$ is defined as before, and $J(t)$ now represents the number of servers who are not on vacations (busy or idle). Then $\{(L_v(t), J(t)), t \geq 0\}$ is a QBD process with the state space

$$\Omega = \{(k,j) : k \geq 0,\ 0 \leq j \leq c\}.$$

For example, state $(0,0)$ represents the state in which there is no customer in the system and all servers are on vacations, and state $(0,j), 1 \leq j \leq c-1$, is the state in which no customers are in the system and $c-j$ servers are on vacations and j servers are idle. The structure of the infinitesimal generator $\mathbf{Q}$ is the same as in (5.3.3), and the $(c+1)\times(c+1)$ matrices $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are the same as in the M/M/c (AS, MV) system. $\mathcal{A}_0, \mathcal{B}_1$, and $\mathcal{C}_0$ are the $c(c+1)\times c(c+1), (c+1)\times c(c+1)$, and $c(c+1)\times(c+1)$ matrices, respectively, and have the same structures as in (5.4.1). However, for $1 \leq k \leq c-1$, $\mathbf{A}_k, \mathbf{B}_k$, and $\mathbf{C}_k$ are now the $(c+1)\times(c+1)$ matrices as follows:

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{M}_k & & & & \\ & (c-k-2)\theta & & & \\ & \ddots & & \ddots & \\ & & & -h_{k,c-1} & \theta \\ & & & & -(\lambda+k\mu) \end{bmatrix},$$

where

$$\mathbf{M}_k = \begin{bmatrix} -h_0 & c\theta & & & & & \\ & -h_1 & (c-1)\theta & & & & \\ & & \ddots & \ddots & & & \\ & & & -h_k & (c-k)\theta & & \\ & & & & -h_{k,k+1} & (c-k-1)\theta & \\ & & & & & -h_{k,k+2} & \end{bmatrix},$$

$$h_k = \lambda + k\mu + (c-k)\theta, \qquad 0 \le k \le c-1,$$
$$h_{kj} = \lambda + k\mu + (c-j)\theta, \qquad 0 \le k \le c-1, \quad k \le j \le c-1;$$

$$\mathbf{B}_k = \begin{bmatrix} 0 & & & & & & & \\ 0 & \mu & & & & & & \\ & 0 & 2\mu & & & & & \\ & & \ddots & \ddots & & & & \\ & & & 0 & (k-1)\mu & & & \\ & & & & k\mu & 0 & & \\ & & & & & \ddots & \ddots & \\ & & & & & & k\mu & 0 \end{bmatrix};$$

$$\mathbf{C}_k = \mathbf{C} = \lambda \mathbf{I};$$

and finally $\mathbf{B}_c = \mathbf{B}$.

Similarly, we can also discuss the M/M/c queue with asynchronous setup times, which is denoted by M/M/c (AS, SU). In such a system, a server is turned off when no customers are waiting at its service completion instant and is turned on again at the next arrival instant. The server starts serving the customer after a setup (or warmup) time. Note that an arrival may see not only busy or turned-off servers but also servers in the process of setup. If an arrival sees k servers are busy or in the setup process, $0 \le k \le c-1$, then $c-k$ servers are in the turned-off state, and this arrival causes one of these $c-k$ servers to be turned on. Note that if the arrival sees some servers in the setup process, then the first server completing the setup time starts serving waiting customers according to the FCFS order. Due to the random setup times, the server that first finishes setup may not be the server that is first turned-on. When a server is experiencing setup time, other servers may be still serving customers. Therefore, at a server's setup time completion instant, it is possible that there are no waiting customers in the system and this server is turned off again without serving any customers. We use the same symbol V as for vacation to denote the setup time that is exponentially distributed with parameter θ. It is also assumed that the setup time, the interarrival time, and the service time are mutually independent. Let $J(t)$ be the number of busy servers at time t. Thus the number of servers in the setup processes at time t is

$$\min\{L_v(t), c\} - J(t).$$

In the M/M/c (AS, SU) system, $\{L_v(t), J(t)\}$ is a QBD process with the state space

$$\Omega = \{(k,j) : 0 \le k \le c-1, 0 \le j \le k\} \cup \{(k,j) : k \ge c, 0 \le j \le c\}.$$

State (k,j), for $0 \le k \le c-1, 0 \le j \le k$, represents the state where j servers are busy, $k-j$ servers are in the set-up processes, $c-k$ servers are off, and $k-j$ customers are waiting in the line. Similarly, state (k,j), for $k \ge c, 0 \le j \le c$, represents the state where $c-j$ servers are in the setup process, j servers are busy, and there are no turned-off servers. The infinitesimal generator of the QBD process, $\mathbf{Q}$, is still given by (5.3.3), where $\mathcal{A}_0, \mathcal{B}_1$, and $\mathcal{C}_0$ are the $\frac{1}{2}c(c+1) \times \frac{1}{2}c(c+1)$, $(c+1) \times \frac{1}{2}c(c+1)$, and $\frac{1}{2}c(c+1) \times (c+1)$ matrices, respectively, and can be expressed in the block-partitioned structure of (5.4.1). For $1 \le k \le c-1$, $\mathbf{A}_k$, the $(k+1) \times (k+1)$ matrix is written as

$$\mathbf{A}_k = \begin{bmatrix} l_{k0} & k\theta & & & & \\ 0 & l_{k1} & (k-1)\theta & & & \\ & 0 & l_{k2} & (k-2)\theta & & \\ & & & \ddots & \ddots & \\ & & & & l_{k,k+1} & \theta \\ & & & & & l_{kk} \end{bmatrix},$$

where

$$l_{kj} = -(\lambda + j\mu + (k-j)\theta), \quad 0 \le k \le c-1,\ 0 \le j \le k,$$

$\mathbf{B}_k, 1 \le k \le c$, $\mathbf{C}_k, 0 \le k \le c-1$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are the same as in the M/M/c (AS, MV) system.

Based on the matrix structures of the QBD processes, we find that the analysis of the M/M/c (AS, SV) system or the M/M/c (AS, SU) system is similar to that of the M/M/c (AS, MV) system. The main difference is in solving (5.4.5) for the boundary-state probabilities. In particular, these systems have similar conditional stochastic decomposition properties to Theorems 5.4.4 and 5.4.5. The only distinguished component is the vector $\beta_c = (\beta_{c0}, \beta_{c1}, \cdots, \beta_{cc}) = (\delta, \beta_{cc})$, which is obtainable from a set of different equations. As an example, we give the results of these models for the M/M/2 queue.

Example 1: The M/M/2 (AS, MV) system.

The infinitesimal generator for $\{(L_v(t), J(t)), t \geq 0\}$ becomes

$$\mathbf{Q} = \begin{bmatrix} A_0 & \mathbf{C}_0 & & & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C}_1 & & & & \\ & \mathbf{B}_2 & \mathbf{A} & \mathbf{C} & & & \\ & & \mathbf{B} & \mathbf{A} & \mathbf{C} & & \\ & & & \mathbf{B} & \mathbf{A} & \mathbf{C} & \\ & & & & \ddots & \ddots & \ddots \end{bmatrix}, \tag{5.4.9}$$

where $A_0 = -\lambda$, $\mathbf{C}_0 = (\lambda, 0)$, $\mathbf{B}_1 = (0, \mu)^T$, and

$$\mathbf{B}_2 = \begin{bmatrix} 0 & 0 \\ 0 & \mu \\ 0 & 2\mu \end{bmatrix}, \quad \mathbf{A}_1 = \begin{bmatrix} -(\lambda + 2\theta) & 2\theta \\ 0 & -(\lambda + \mu) \end{bmatrix},$$

$$\mathbf{C}_1 = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \end{bmatrix}.$$

$\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are the 3×3 matrices, as follows:

$$\mathbf{A} = \begin{bmatrix} -(\lambda + 2\theta) & 2\theta & 0 \\ 0 & -(\lambda + \mu + \theta) & \theta \\ 0 & 0 & -(\lambda + 2\mu) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & & \\ & \mu & \\ & & 2\mu \end{bmatrix},$$

$$\mathbf{C} = \lambda \mathbf{I}.$$

Let $r_1 < r_1^*$ denote the two roots of the quadratic equation $\mu z^2 - (\lambda + \mu + \theta)z + \lambda = 0$, $\rho = \lambda(z\mu)^{-1} < 1$. The rate matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \begin{bmatrix} \frac{\lambda}{\lambda+2\theta} & \frac{2\theta}{\mu}\frac{r_0}{r_1^*-r_0} & \rho\frac{\theta}{\mu}\frac{1}{1-r_1}\frac{1}{r_1^*-r_0} \\ & r_1 & \frac{\theta}{2\mu}\frac{r_1}{1-r_1} \\ & & \rho \end{bmatrix},$$

where $r_0 = \lambda(\lambda + 2\theta)^{-1}$. Note that

$$B[\mathbf{R}] = \begin{bmatrix} A_0 & \mathbf{C}_0 & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C}_1 \\ & \mathbf{B}_2 & \mathbf{A} + \mathbf{R}\mathbf{B} \end{bmatrix}$$

becomes a 6×6 matrix. Let $r_{ij} = (\mathbf{R})_{ij}$ be the (i, j) entry of $\mathbf{R}$. It can be verified by direct computation that $\mathbf{R}\mathbf{T}^0 = \lambda \mathbf{e}$, where $\mathbf{T}^0 = (0, \mu, 2\mu)^T$.

Solving $(\pi_0, \pi_1, \pi_2)B[\mathbf{R}] = \mathbf{0}$ gives

$$\begin{aligned}
\pi_0 &= K, \\
\pi_1 &= K\beta_1 = K\left(\frac{\lambda}{\lambda + 2\theta}, \frac{\lambda}{\mu}\right), \\
\pi_2 &= K\beta_2 = K(\beta_{20}, \beta_{21}, \beta_{22}),
\end{aligned}$$

where

$$\begin{aligned}
\beta_{20} &= \left(\frac{\lambda}{\lambda + 2\theta}\right)^2, \\
\beta_{21} &= \frac{\lambda}{\mu} r_1 + \frac{2\theta}{\mu} \frac{r_0^2}{r_1^* - r_0}, \\
\beta_{22} &= \frac{\theta}{2\mu} \frac{1}{1 - r_1} \frac{\lambda}{\mu} \left(\frac{r_0}{r_1^* - r_0} + r_1\right), \\
K &= \left\{\frac{\lambda}{\mu} + \frac{2(\lambda + \theta)}{\lambda + 2\theta} + \beta_2(\mathbf{I} - \mathbf{R})^{-1}\mathbf{e}\right\}^{-1}.
\end{aligned}$$

From these results, we can easily obtain various performance measures and the conditional stochastic decompositions for the queue length and the waiting time.

Example 2: The M/M/2 (AS, SV) system.

The infinitesimal generator is still given by (5.4.9) where all elements are the 3×3 matrices as follows:

$$\mathbf{A}_0 = \begin{bmatrix} -(\lambda + 2\theta) & 2\theta & 0 \\ 0 & -(\lambda + \theta) & \theta \\ 0 & 0 & -\lambda \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 0 & 0 & 0 \\ \mu & 0 & 0 \\ 0 & \mu & 0 \end{bmatrix},$$

$$\mathbf{A}_1 = \begin{bmatrix} -(\lambda + 2\theta) & 2\theta & 0 \\ 0 & -(\lambda + \mu + \theta) & \theta \\ 0 & 0 & -(\lambda + \mu) \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 2\mu & 0 \end{bmatrix}.$$

Matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{R}$ are the same as in Example 1. $B[\mathbf{R}]$ is the 9×9 matrix. Solving $(\pi_0, \pi_1, \pi_2)B[\mathbf{R}] = \mathbf{0}$ gives

$$\begin{aligned}
\pi_0 &= K\beta_0 = K\left(1,\ \frac{2\theta}{\lambda} + \frac{\theta}{\lambda + \mu + \theta},\ \frac{\theta}{\lambda}\left(\frac{2\theta}{\lambda} + \frac{\theta}{\lambda + \mu + \theta}\right)\right), \\
\pi_1 &= K\beta_1 = K\left(\frac{\lambda}{\lambda + 2\theta},\ \frac{\lambda + 2\theta}{\mu},\ \frac{\theta}{\lambda}\left(\frac{2\theta}{\lambda} + \frac{\lambda + \theta}{\lambda + \mu + \theta}\right)\right), \\
\pi_2 &= K\beta_2 = K(\beta_{20},\ \beta_{21},\ \beta_{22}),
\end{aligned}$$

where

$$\beta_{20} = \left(\frac{\lambda}{\lambda + 2\theta}\right)^2,$$
$$\beta_{21} = \frac{\lambda}{\mu + 2\theta} r_{01} + \frac{\lambda + 2\theta}{\mu} r_1,$$
$$\beta_{22} = \frac{\lambda}{\lambda + 2\theta} r_{02} + \frac{\lambda + 2\theta}{\mu} r_{12} + \frac{\theta}{2\mu}\left(\frac{2\theta}{\mu} + \frac{\lambda}{\mu} + \frac{\lambda + \theta}{\lambda + \mu + \theta}\right),$$

and K can be determined by the normalization condition. Again, from these results, we can obtain the major performance measures and the conditional stochastic decomposition properties.

Example 3: The M/M/c (AS, SU) system.

The structure of the infinitesimal generator remains the same as in (5.4.9), where the only different entry is

$$\mathbf{A}_1 = \begin{bmatrix} -(\lambda + \theta) & \theta \\ 0 & -(\lambda + \mu) \end{bmatrix}.$$

All other entries of $\mathbf{Q}$ are the same as in Example 1. Thus the rate matrix $\mathbf{R}$ is the same as in Example 1, and $B[\mathbf{R}]$ is the 6×6 matrix. Solving $(\pi_0, \pi_1, \pi_2)B[\mathbf{R}] = \mathbf{0}$ gives

$$\pi_0 = K,$$
$$\pi_1 = K\beta_1 = K\left(\frac{\lambda}{\lambda + \theta}, \frac{\lambda}{\mu}\right),$$
$$\pi_2 = K\beta_2 = K\left(\frac{\lambda^2}{(\lambda + \theta)(\lambda + 2\theta)}, \frac{\lambda}{\lambda + \theta} r_{01} + r_1, \frac{\lambda}{\lambda + \theta} r_{02} + \frac{\lambda}{\mu} r_{12}\right),$$

where r_{01}, r_{12},and r_{02} are the entries of $\mathbf{R}$. From these results, we can develop the major performance measures and the conditional stochastic decomposition properties.

5.5 M/M/c Queue with Synchronous Vacations of Some Servers

5.5.1 (SY, MV, d)-Policy Model

In the vacation models discussed in the previous sections, we assume that all servers may be on vacation. This means that a customer may see that no servers are available at his or her arrival instant. In practical situations, we may wish to keep at least a certain number of servers always on duty (in either busy or idle status). For a system with synchronous vacation policy, this means that only a certain number of servers (not all) are allowed to take a vacation each time. For example, a border-crossing station between the U.S. and Canada operates 24 hours a day and requires at least one or two lanes to be open to traffic. Therefore, we need to study the vacation model with vacations of some but not all servers. Ikagi (1992) studied an M/M/2 system where at most one server can take vacations. We now discuss an M/M/c system where only a subset of servers is allowed to take vacations. Introducing a control parameter d ($1 \le d \le c$), we design the following policy: at a service completion instant, if the number of idle servers reaches d (or the number of customers in the system is reduced to $c-d$), these d servers start a vacation together and the remaining $c-d$ servers either serve customers or stay idle; at a vacation completion instant, if the number of customers does not exceed $c-d$, these d servers take another vacation together; otherwise, these d servers resume serving customers. Note that when d servers start a vacation, there are still customers in the system. Thus the policy is said to be *semi-exhaustive*. The system is denoted by M/M/c (SY, MV, d). It is assumed that the vacation time follows an exponential distribution with parameter θ and is independent of the interarrival time and the service time. The service sequence is FCFS. At a vacation completion instant with $j > c-d$ customers in the system, there are two possible cases of resuming the queue service: (i) if $c-d < j \le c$, then $j-c+d$ returning servers start serving customers and $c-j$ servers become idle; (ii) if $j > c$, then all returning servers start serving customers and $j-c$ customers are waiting in the line. Now, there is a distinguished feature of this type of vacation model compared with the single server vacation model or the multiserver vacation model with synchronous vacations for all servers. That is, in the M/M/c (SY, MV, d) system, the number of customers in the system during the vacation may either increase or decrease, since $c-d$ servers still attend the queue, while in the M/M/c (SY, MV) system or single server vacation system, the number of customers never decreases during the vacation.

Let $L_v(t)$ be the number of customers in the system at time t, and let

$$J(t) = \begin{cases} 0 & d \text{ servers are on vacation at time } t, \\ 1 & \text{no servers are on vacation at time } t. \end{cases}$$

$\{L_v(t), J(t)\}$ is a QBD process with the state space

$$\Omega = \{(k,0) : 0 \le k \le c-d\} \cup \{(k,j) : k > c-d, j = 0, 1\}.$$

Note that a customer departure in state $(c-d+1, 1)$ makes the process transfer to state $(c-d, 0)$, and the d servers start a vacation. If we use the lexicographical sequence for the states, the infinitesimal generator can be written in the block-partitioned form as

$$\mathbf{Q} = \begin{bmatrix} A_0 & C_0 & & & & & & & \\ B_1 & A_1 & C_1 & & & & & & \\ & B_2 & A_2 & C_2 & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & \mathbf{B}_c & \mathbf{A}_c & \mathbf{C}_c & & & \\ & & & & \mathbf{B} & \mathbf{A} & \mathbf{C} & & \\ & & & & & \mathbf{B} & \mathbf{A} & \mathbf{C} & \\ & & & & & & \ddots & \ddots & \ddots \end{bmatrix}, \tag{5.5.1}$$

where

$$\mathbf{A}_k = \begin{cases} -(\lambda + k\mu), & 0 \le k \le c-d, \\ \begin{pmatrix} -[\lambda + (c-d)\mu + \theta] & \theta \\ 0 & -(\lambda + k\mu) \end{pmatrix}, & c-d < k \le c, \end{cases}$$

$$\mathbf{C}_k = \begin{cases} \lambda, & 0 \le k < c-d, \\ (\lambda, 0), & k = c-d, \\ \lambda\mathbf{I}, & c-d < k \le c. \end{cases}$$

$$\mathbf{B}_k = \begin{cases} k\mu & 1 \le k \le c-d, \\ \begin{pmatrix} (c-d)\mu \\ (c-d+1)\mu \end{pmatrix} & k = c-d+1, \\ \begin{pmatrix} (c-d)\mu & 0 \\ 0 & k\mu \end{pmatrix} & c-d+1 < k \le c. \end{cases}$$

Furthermore,

$$\mathbf{B} = \begin{pmatrix} (c-d)\mu & 0 \\ 0 & c\mu \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} -[\lambda + (c-d)\mu + \theta] & \theta \\ 0 & -(\lambda + c\mu) \end{pmatrix},$$
$$\mathbf{C} = \lambda\mathbf{I}.$$

To obtain the rate matrix $\mathbf{R}$, we need to know that the quadratic equation

$$(c-d)\mu z^2 - [\lambda + \theta + (c-d)\mu]z + \lambda = 0 \tag{5.5.2}$$

has the two real roots $r < r^*$ and $0 < r < 1$, $r^* > 1$.

$$r^*, r = \frac{1}{2(c-d)\mu}\{\lambda + \theta + (c-d)\mu \pm \sqrt{H}\},$$

where

$$H = [\lambda - (c-d)\mu]^2 + \theta^2 + 2\theta[\lambda + (c-d)\mu] > 0.$$

We are now ready to prove the following theorem.

Theorem 5.5.1. If $\rho = \lambda(c\mu)^{-1} < 1$, the matrix equation

$$\mathbf{R}^2\mathbf{B} + \mathbf{R}\mathbf{A} + \mathbf{C} = \mathbf{0} \tag{5.5.3}$$

has the minimal nonnegative solution

$$\mathbf{R} = \begin{pmatrix} r & \frac{\theta r}{c\mu(1-r)} \\ 0 & \rho \end{pmatrix}, \tag{5.5.4}$$

and $sp(\mathbf{R}) < 1$.

Proof: The coefficient matrices of (5.5.3) are all upper-triangular. Let

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}.$$

Substituting this $\mathbf{R}$ into (5.5.3) gives the following set of equations:

$$\begin{cases} (c-d)\mu r_{11}^2 - [\lambda + \theta + (c-d)\mu]r_{11} + \lambda = 0 \\ c\mu r_{22}^2 - (\lambda + c\mu)r_{22} + \lambda = 0 \\ c\mu r_{12}(r_{11} + r_{22}) + \theta r_{11} - (\lambda + c\mu)r_{12} = 0. \end{cases}$$

To obtain the minimal nonnegative solution, let $r_{11} = r$ in the first equation and let $r_{22} = \rho$ in the second equation (the other root for the second equation is $r_{22} = 1$). Substituting r and ρ into the third equation, we obtain $r_{12} = \frac{\theta r}{c\mu(1-r)}$ and $sp(\mathbf{R}) = \max(r, \rho) < 1.\Box$

From Theorems 5.5.1 and 5.2.3, it can be easily proved that $\{L_v(t), J(t)\}$ is positive recurrent if and only if $\rho < 1$.

Lemma 5.5.1. $\mathbf{R}$ satisfies $\mathbf{RBe} = \lambda\mathbf{e}$ and there exists the relation

$$\lambda + \theta + (c-d)\mu(1-r) = \frac{\theta}{1-r} + (c-d)\mu = \frac{\lambda}{r}. \tag{5.5.5}$$

Proof: Multiplying both sides of (5.5.3) from the right by $\mathbf{e}$ gives

$$\mathbf{R}^2\mathbf{Be} - \mathbf{R}(\lambda\mathbf{e} + \mathbf{Be}) + \lambda\mathbf{e} = \mathbf{0},$$

and rearranging the terms results in

$$(\mathbf{I} - \mathbf{R})(\lambda\mathbf{e} - \mathbf{RBe}) = \mathbf{0}.$$

Because the inverse of $\mathbf{I} - \mathbf{R}$ exists, $\mathbf{RBe} = \lambda\mathbf{e}$, which gives

$$\theta + (c-d)\mu(1-r) = \frac{1-r}{r}\lambda.$$

Adding λ to both sides of the equation above yields (5.5.5).$\Box$

The infinitesimal generator $\mathbf{Q}$ can be repartitioned as follows:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_{01} & & & \\ \mathbf{H}_{10} & \mathbf{A} & \mathbf{C} & & \\ & \mathbf{B} & \mathbf{A} & \mathbf{C} & \\ & & \mathbf{B} & \mathbf{A} & \ddots \\ & & & \vdots & \vdots \end{bmatrix},$$

where

$$\mathbf{H}_0 = \begin{bmatrix} \mathbf{A}_0 & \mathbf{C}_0 & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C}_1 & & \\ & \ddots & \ddots & \ddots & \\ & & \mathbf{B}_{c-2} & \mathbf{A}_{c-2} & \mathbf{C}_{c-2} \\ & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} \end{bmatrix},$$

$$\mathbf{H}_{10} = (\mathbf{0}, \mathbf{B}_c),\ \mathbf{H}_{10} = \begin{pmatrix} \mathbf{0} \\ \mathbf{C}_{c-1} \end{pmatrix}.$$

Note that the repartitioned $\mathbf{Q}$ is not in the standard canonical form and has a more complicated structure near the lower boundary. However, the matrix analytical method can still be applied by using a modified matrix-geometric invariant vector, as shown in section 1.5 of Neuts (1981).

Let $\{L_v, J\}$ be the stationary random variables for the queue length and the status of servers. Denote the joint probability by

$$\pi_{kj} = P\{L_v = k, J = j\} = \lim_{t\to\infty} P\{L_v(t) = k, J(t) = j\},\ (k, j) \in \Omega,$$

where $\pi_k = (\pi_{k0}, \pi_{k1})$, for $k \geq c - d + 1$. We show below that $\{\pi_{kj}$ $(k, j) \in \Omega\}$ exist and can be obtained.

Define the $(c - d + 1) \times (c - d + 1)$ matrix

$$B[\mathbf{R}] = \begin{bmatrix} A_0 & C_0 & & & & \\ B_1 & A_1 & C_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} & \mathbf{C}_{c-1} \\ & & & & \mathbf{B}_c & \mathbf{A} + \mathbf{R}\mathbf{B} \end{bmatrix}$$

and the $2(c - d + 1)$ -dimensional vector

$$\Pi_{c+d+1} = (\pi_{00}, \pi_{10}, ..., \pi_{c-d,0}, (\pi_{c-d+1,0}, \pi_{c-d+1,1}), ..., (\pi_{c0}, \pi_{c1})).$$

Lemma 5.5.2. $\Pi_{c+d+1} B[\mathbf{R}] = 0$ has a positive solution:

$$\begin{aligned} \pi_{j0} &= \tfrac{K}{j!} \left(\tfrac{\lambda}{\mu}\right)^j, \qquad & 0 \leq j \leq c - d, \\ \pi_j &= K(\beta_{j0}, \beta_{j1}), \qquad & c - d < j \leq c, \end{aligned}$$

where

$$\beta_{j0} = \frac{1}{(c-d)!}\left(\frac{\lambda}{\mu}\right)^{c-d} r^{j-(c-d)}, \quad c-d+1 \le j \le c$$

$$\beta_{j1} = \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^{j} \frac{\theta r}{\lambda(1-r)}\left[1 + \frac{1}{(c-d)!}\sum_{i=1}^{j-(c-d)-1}(c-d+i)!\left(\frac{r\mu}{\lambda}\right)^{i}\right],$$
$$c-d+1 \le j \le c,$$

The empty summation $\sum_{i=1}^{0}$ is defined to be zero.

Proof: Using $\mathbf{R}$ in (5.5.4), we have

$$\mathbf{A} + \mathbf{RB} = \begin{pmatrix} -\{\lambda+\theta+(c-d)\mu(1-r)\} & \frac{\theta}{1-r} \\ 0 & -c\mu \end{pmatrix},$$

which appears in the last row of $B[\mathbf{R}]$. Then the matrix equation $\Pi_{c+d+1}B[\mathbf{R}] = 0$ can be written as a set of equations:

$$\begin{cases} -\lambda\pi_{00} + \mu\pi_{10} = 0 & \text{(Eq. 1)} \\ \lambda\pi_{j-1,0} - (\lambda + j\mu)\pi_{j0} + (j+1)\mu\pi_{j+1,0} = 0,\ 1 \le j < c-d, & \text{(Eq. 2)} \\ \lambda\pi_{c-d-1,0} - (\lambda + (c-d)\mu)\pi_{c-d,0} + (c-d)\mu\pi_{c-d+1,0} & \\ \quad + (c-d+1)\mu\pi_{c-d+1,1} = 0 & \text{(Eq. 3)} \\ \theta\pi_{c-d+1,0} - (\lambda + (c-d+1)\mu)\pi_{c-d+1,1} & \\ \quad + (c-d+2)\mu\pi_{c-d+2,1} = 0 & \text{(Eq. 4)} \\ \lambda\pi_{j-1,0} - [\lambda + (c-d)\mu + \theta]\pi_{j0} + (c-d)\mu\pi_{j+1,0} = 0, & \\ \qquad c-d < j \le c-1 & \text{(Eq. 5)} \\ \lambda\pi_{j-1,1} - \theta\pi_{j,0} - (\lambda + j\mu)\pi_{j,1} + (j+1)\mu\pi_{j+1,1} = 0, & \\ \qquad c-d+1 < j \le c-1 & \text{(Eq. 6)} \\ \lambda\pi_{c-1,0} - (\lambda + \theta + (c-d)\mu(1-r))\pi_{c0} = 0 & \text{(Eq. 7)} \\ \lambda\pi_{c-1,1} + \frac{\theta}{1-r}\pi_{c0} - c\mu\pi_{c1} = 0 & \text{(Eq. 8)} \end{cases}$$

From (5.5.5) and (Eq. 7), we obtain $\pi_{c0} = r\pi_{c-1,0}$. In (Eq. 5), letting $j = c-1$, we get

$$\begin{aligned}\lambda\pi_{c-2,0} &= (\lambda + \theta + (c-d)\mu)\pi_{c-1,0} - (c-d)\mu r\pi_{c-1,0} \\ &= (\lambda + \theta + (c-d)\mu(1-r))\pi_{c-1,0} = \frac{\lambda}{r}\pi_{c-1,0}\end{aligned}$$

so that $\pi_{c-1,0} = r\pi_{c-2,0}$. Repeating using (Eq. 5) recursively, gives

$$\pi_{c0} = r^{j}\pi_{c-j,0} \qquad 0 \le j \le d. \tag{5.5.7}$$

Let $\pi_{0,0} = K$. From (Eq. 1), we obtain $\pi_{10} = \lambda\mu^{-1}K$. Successively substituting equations in (Eq. 2), we have

$$\pi_{j,0} = \frac{K}{j!}\left(\frac{\lambda}{\mu}\right)^{j}, \qquad 0 \le j \le c-d. \tag{5.5.8}$$

In (5.5.8), letting $j = c - d$ and comparing it with (5.5.7), we get

$$\pi_{c0} = \frac{K}{(c-d)!} \left(\frac{\lambda}{\mu}\right)^{c-d} r^d.$$

Then substituting it back to (5.5.7) yields

$$\pi_{j0} = \frac{K}{(c-d)!} \left(\frac{\lambda}{\mu}\right)^{c-d} r^{j-(c-d)}, \qquad c-d < j \le c.$$

Substituting $\pi_{c-d-1,0}, \pi_{c-d,0}$ and $\pi_{c-d+1,0}$ into (Eq. 3) gives

$$\pi_{c-d+1,1} = \frac{K}{(c-d+1)!} \left(\frac{\lambda}{\mu}\right)^{c-d+1} \left[1 - \frac{\mu}{\lambda}(c-d)r\right].$$

From (5.5.5), it is easy to verify that

$$1 - (c-d)\frac{\mu}{\lambda}r = \frac{\theta r}{\lambda(1-r)}.$$

Using this relation, we have

$$\pi_{c-d+1,1} = \frac{K}{(c-d+1)!} \left(\frac{\lambda}{\mu}\right)^{c-d+1} \frac{\theta r}{\lambda(1-r)}. \tag{5.5.9}$$

Substituting (5.5.9) and (5.5.8) into (Eq. 4), we get

$$\pi_{c-d+2,1} = \frac{K}{(c-d+2)!} \left(\frac{\lambda}{\mu}\right)^{c-d+2} \frac{\theta r}{\lambda(1-r)} \left[1 + (c-d+1)\frac{\mu r}{\lambda}\right].$$

Successively substituting this expression and (5.5.9) into (Eq. 6), we obtain

$$\pi_{j1} = \frac{K}{j!} \left(\frac{\lambda}{\mu}\right)^{j} \frac{\theta r}{\lambda(1-r)} \left\{ 1 + \frac{1}{(c-d)!} \sum_{i=1}^{j-(c-d)-1} (c-d+i)! \left(\frac{\mu r}{\lambda}\right)^i \right\},$$
$$c-d < j \le c.$$

Finally, using direct substitution, we can verify (Eq. 8).□

Based on the modification method in section 1.5 of Neuts (1981) for the infinitesimal generator $\mathbf{Q}$ with complex lower boundary, it is obvious that if and only if $sp(\mathbf{R}) < 1$ and linear equation system $\Pi_{c+d+1}B[\mathbf{R}] = 0$ has a positive solution, then the QBD process $\{L_v(t), J(t)\}$ is positive recurrent. Based on Theorem 5.5.1 and Lemma 5.5.2, these conditions are satisfied if and only if $\rho < 1$.

For the joint distribution queue length and the status of servers, we have the following theorem.

Theorem 5.5.2. If $\rho < 1$, the distribution of $\{L_v, J\}$ is

$$\begin{cases} \pi_{j0} = \frac{K}{j!}\left(\frac{\lambda}{\mu}\right)^j, & 0 \le j \le c-d \\ \pi_j = K(\beta_{j0}, \beta_{j1}), & c-d < j \le c \\ \pi_{j0} = K\beta_{c0} r^{j-c}, & j > c \\ \pi_{j1} = K\beta_{c1}\rho^{j-c} + K\beta_{c0}\frac{\theta r}{c\mu(1-r)}\sum_{i=0}^{j-c-1} r^i \rho^{j-c-1-i}, & j > c, \end{cases}$$

where β_{j0} and β_{j1} are given in Lemma 5.5.2 and the constant K is as follows:

$$K = \left[\sum_{j=0}^{c-d} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \sum_{j=c-d+1}^{c-1} (\beta_{j0} + \beta_{j1}) + (\beta_{c0}, \beta_{c1})(\mathbf{I} - \mathbf{R})^{-1}\mathbf{e}\right]^{-1}.$$

Proof: Based on Theorem 5.2.3, we have

$$\pi_k = (\pi_{k0}, \pi_{k1}) = K(\beta_{c0}, \beta_{c1})\mathbf{R}^{k-c} \qquad k \ge c,$$

and $\pi_{00}, \pi_{10}, ...\pi_{c-d,0}, (\pi_{k0}, \pi_{k1}), c-d < k \le c$, are given by Lemma 5.5.2. Substituting $\mathbf{R}$ in (5.5.4) into the expression above gives (π_{j0}, π_{j1}) for $j > c$. K can be determined by the normalization condition. □

The distribution of the number of customers in the system at any time is

$$P\{L_v = j\} = \begin{cases} \pi_{j0}, & 0 \le j \le c-d, \\ \pi_{j0} + \pi_{j1}, & j > c-d \ , \end{cases}$$

Note that, based on Theorem 5.5.2, the distribution of waiting time can be obtained by conditioning on each state $(k, j) \in \Omega$. However, this distribution is very complex and is not convenient to use. It is also hard to compare this multiserver vacation system with its classical M/M/c system in terms of unconditional distributions. Therefore, we again present the conditional stochastic decomposition properties.

Let $L_v^{(c)} = \{L_v - c | L_v \ge c, J = 1\}$ and $W_v^{(c)} = \{W_v | L_v \ge c, J = 1\}$ represent the queue length and the waiting time, respectively, given that all servers are busy.

Theorem 5.5.3. If $\rho < 1$, $L_v^{(c)}$ can be decomposed into the sum of two independent random variables,

$$L_v^{(c)} = L_0^{(c)} + L_d,$$

where $L_0^{(c)}$ is the conditional queue length of the classical M/M/c system without vacation and L_d is the additional queue length due to the vacation effect. The p.g.f. of L_d is given by

$$L_d(z) = \frac{1}{\sigma}\left\{\beta_{c1} + \frac{\theta r}{c\mu(1-r)}\beta_{c0} z \frac{1}{(1-zr)}\right\}, \tag{5.5.10}$$

where

$$\sigma = \beta_{c1} + \frac{\theta r}{c\mu(1-r)^2}\beta_{c0}.$$

Proof: From Theorem 5.5.2, the probability that all servers are busy is

$$\begin{aligned}
P\{L_v \geq c, J = 1\}
&= \sum_{j=c}^{\infty} \pi_{j1} \\
&= K\beta_{c1}\sum_{j=c}^{\infty}\rho^{j-c} + K\frac{\theta r}{c\mu(1-r)}\beta_{c0}\sum_{j=c+1}^{\infty}\sum_{k=0}^{j-c-1} r^k \rho^{j-c-1-k} \\
&= \frac{K}{1-\rho}\beta_{c1} + \frac{K}{1-\rho}\beta_{c0}\frac{\theta r}{c\mu(1-r)^2} \\
&= \frac{\sigma}{1-\rho}K.
\end{aligned}$$

The conditional probability distribution of $L_v^{(c)}$ is obtained as

$$\begin{aligned}
P\{L_v^{(c)} = j\}
&= P\{L_v = j + c | L_v \geq c, J = 1\} \\
&= \frac{1-\rho}{\sigma}\left\{\beta_{c1}\rho^j + \frac{\theta r}{c\mu(1-r)}\beta_{c0}\sum_{k=0}^{j-1} r^k \rho^{j-1-k}\right\}, \ j \geq 0.
\end{aligned} \tag{5.5.11}$$

Taking the p.g.f. of (5.5.11), we have

$$\begin{aligned}
L_v^{(c)}(z) &= \sum_{j=0}^{\infty} P\{L_v^{(c)} = j\} z^j \\
&= \frac{1-\rho}{\sigma}\left\{\beta_{c1}\sum_{j=0}^{\infty}\rho^j z^j + \frac{\theta r}{c\mu(1-r)}\beta_{c0}\sum_{j=1}^{\infty} z^j \sum_{k=0}^{j-1} r^k \rho^{j-1-k}\right\} \\
&= \frac{1-\rho}{1-z\rho} \times \frac{1}{\sigma}\left\{\beta_{c1} + \frac{\theta r}{c\mu(1-r)}\beta_{c0} z \frac{1}{1-zr}\right\} \\
&= L_0^{(c)}(z) L_d(z).
\end{aligned}$$

□

Expanding $L_d(z)$ yields the distribution of L_d as

$$P\{L_d = j\} = \begin{cases} \frac{1}{\sigma}\beta_{c1}, & j = 0, \\ \frac{1}{\sigma}\frac{\theta r}{c\mu(1-r)^2}\beta_{c0}(1-r)r^{j-1}, & j \geq 1. \end{cases} \tag{5.5.12}$$

Note that (5.5.12) indicates that with probability $\beta_{c1}\sigma^{-1}$, $L_d = 0$ and with probability $1 - \beta_{c1}\sigma^{-1}$, L_d follows a geometric distribution with parameter r. The following theorem gives the conditional stochastic decomposition property of the waiting time.

Theorem 5.5.4. If $\rho < 1$, $W_v^{(c)}$ can be decomposed into the sum of two independent random variables,

$$W_v^{(c)} = W_0^{(c)} + W_d,$$

where $W_0^{(c)}$ is the conditional waiting time in a classical M/M/c system without vacations, and W_d is the additional delay due to the vacation effect. $W_0^{(c)}$ follows an exponential distribution with parameter $c\mu(1-\rho)$, and W_d has the LST

$$W_d^*(s) = \frac{1}{\sigma}\left\{\beta_{c1} + \frac{\theta r}{c\mu(1-r)^2}\beta_{c0}\frac{c\mu(1-r)}{s + c\mu(1-r)}\right\}. \tag{5.5.13}$$

Proof: Assume that a customer arrives at state $(k, 1)$ for $k \geq c$. If we condition on this state, this customer's waiting time W_{ck} has the LST

$$W_{ck}^*(s) = \left(\frac{c\mu}{s + c\mu}\right)^{k-c+1}, \qquad k \geq c.$$

The conditional waiting time when all servers are busy has the LST

$$\begin{aligned} W_v^{*(c)}(s) &= \sum_{k=c}^{\infty} P\{L_v^{(c)} = k - c\}W_{ck}^*(s) \\ &= \frac{1-\rho}{\sigma}\left\{\beta_{c1}\frac{c\mu}{s + c\mu(1-\rho)}\right. \\ &\qquad \left. + \frac{\theta r}{c\mu(1-r)}\beta_{c0}\frac{c\mu}{s + c\mu(1-\rho)}\frac{c\mu}{s + c\mu(1-r)}\right\} \\ &= \frac{c\mu(1-\rho)}{s + c\mu(1-\rho)}\frac{1}{\sigma}\left\{\beta_{c1} + \frac{\theta r}{c\mu(1-r)^2}\beta_{c0}\frac{c\mu(1-r)}{s + c\mu(1-r)}\right\} \\ &= W_0^*(s)W_d^*(s). \end{aligned}$$

□

Note that (5.5.13) indicates that W_d has the probability density at the origin with probability

$$q^* = \frac{1}{\sigma}\beta_{c1},$$

and follows an exponential distribution with parameter $c\mu(1-r)$ with probability $1-q^*$.

From the conditional stochastic decomposition properties, we can obtain the means of $L_v^{(c)}$ and $W_v^{(c)}$:

$$E(L_v^{(c)}) = \frac{\rho}{1-\rho} + \frac{1}{\sigma}\frac{\theta r}{c\mu(1-r)^3}\beta_{c0},$$

$$E(W_v^{(c)}) = \frac{1}{c\mu(1-\rho)} + \frac{1}{\sigma}\frac{\theta r^2}{c\mu(1-r)^3}\beta_{c0}\frac{1}{c\mu}.$$

Remark 5.5.1. Using a similar analysis, we can study the single vacation model, denoted by M/M/c (SY, SV, d), where the d servers take only one vacation simultaneously when the number of customers in the system is reduced to $c-d$ at a service completion instant and return to serve the queue or stay idle after the vacation. We can also analyze the M/M/c (SY, SU, d) model where the d servers are turned off when the number of customers in the system becomes $c-d$ at a service completion instant and are turned on with a setup time when the number of customers in the system is increased to $c-d+1$. Note that in both models, the number of servers on duty never falls below $c-d$. If we assume that the vacation time or the setup time follows the exponential distribution with parameter θ and is independent of the interarival time and the service time, the analysis of the M/M/c (SY, SV,d) or the M/M/c (SY, SU,d) is the same as in sections 5.5.1 and 5.5.2. The infinitesimal generator is still given by (5.5.1) and the matrices $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are the same as in the M/M/c (SY, MV,d) system. The only difference from the M/M/c (SY, MV, d) model is the transition rates among the boundary states, where the number of customers in the system is no more than c. All three models have the same rate matrix $\mathbf{R}$ of (5.5.4). The structures of the conditional stochastic decompositions in these models remain the same as illustrated in Theorems 5.5.3 and 5.5.4 except for the expressions of β_{c0} and β_{c1}, which are determined by different equations.

5.5.2 (SY, MV, e-d)-Policy Model

Now we consider an M/M/c system where only a batch of idle servers (not all) are allowed to take synchronous multiple vacations. The servers in such a system follow a so-called (e,d) policy. The (e,d) policy prescribes that when any d $(\leq c)$ servers become idle or the number of customers in the system is reduced to $c-d$ at a service completion instant, then e $(\leq d)$ idle servers start to take a vacation together. At a vacation completion instant, if the number of customers in the system is no more than $c-e$ (still no waiting customers), these e servers take another vacation together until they find that there are more than $c-e$ customers in the system at a vacation completion instant. Then these e servers return to serve the queue. The vacation time is assumed to be exponentially distributed with parameter θ. The service order is FCFS and interarrival times, service times, and vacation times are mutually independent.

At a vacation completion instant, if there are j customers in the system where $c-e<j\leq c$, then $j-c+e$ returning servers will serve the customers immediately and $c-j$ servers become idle; if $j>c$, all e returning servers serve the customers immediately and $j-c$ customers are waiting in the line.

Let $L(t)$ be the number of customers in the system at time t, and let

$$J(t)=\begin{cases}0, & e \text{ servers are on vacation at time } t,\\ 1, & \text{no servers are on vacation at time } t.\end{cases}$$

Then $\{L(t),J(t)\}$ is a QBD process with the state space

$$\Omega=\{(k,0):0\leq k\leq c-d\}\bigcup\{(k,j):k>c-d,\ j=0,1\}.$$

Using the lexicographical sequence for the states, the infinitesimal generator is given by (5.5.1) where the entries are modified as follows:

$$\mathbf{A}_k=\begin{cases}-(\lambda+k\mu), & 0\leq k<c-d,\\ \begin{pmatrix}-(\lambda+k\mu) & 0\\ 0 & -(\lambda+k\mu)\end{pmatrix}, & c-d\leq k\leq c-e,\\ \begin{pmatrix}-(\lambda+\theta+(c-e)\mu) & \theta\\ 0 & -(\lambda+k\mu)\end{pmatrix}, & c-e<k\leq c-1,\end{cases}$$

$$\mathbf{B}_k=\begin{cases}k\mu, & 1\leq k\leq c-d,\\ \begin{pmatrix}(c-d+1)\mu\\ (c-d+1)\mu\end{pmatrix}, & k=c-d+1,\\ \begin{pmatrix}k\mu & 0\\ 0 & k\mu\end{pmatrix}, & c-d+1<k\leq c-e,\\ \begin{pmatrix}(c-e)\mu & 0\\ 0 & k\mu\end{pmatrix}, & c-e<k\leq c-1,\end{cases}$$

$$\mathbf{C}_k = \begin{cases} \lambda, & 0 \le k < c-d, \\ (\lambda, 0), & k = c-d, \\ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, & c-d < k \le c-1, \end{cases}$$

$$\mathbf{B} = \begin{pmatrix} (c-e)\mu & 0 \\ 0 & c\mu \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} -[\lambda + (c-e)\mu + \theta] & \theta \\ 0 & -(\lambda + c\mu) \end{pmatrix},$$
$$\mathbf{C} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}.$$

Note that a customer departure in state $(c-d+1, 1)$ makes a transition to state $(c-d, 0)$, in which e servers are on vacation. Because the matrices $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are the same as before, the expression of $\mathbf{R}$ is still given by (5.5.4) and the expression of r in Theorem 5.5.1 is slightly changed to

$$r = \frac{1}{2(c-e)\mu}\{\lambda + \theta + (c-e)\mu - \sqrt{H}\},$$

where $H = [\lambda - (c-e)\mu]^2 + \theta^2 + 2\theta[\lambda + (c-e)\mu]$. Define

$$\Pi = (\pi_{00}, \cdots, \pi_{c-d,0}, \pi_{c-d+1}, \pi_{c-d+2}, \cdots), \tag{5.5.14}$$

where $\pi_k = (\pi_{k0}, \pi_{k1})$, for $k \ge c-d+1$. To obtain the distribution $\{\pi_{kj}$ $(k,j) \in \Omega\}$, we define

$$\psi_k = \left(\frac{\lambda}{\mu}\right)^k \left[(c-e)! \left(\frac{\mu}{\lambda}\right)^{c-e} + \frac{r}{1-r}\frac{\theta}{\lambda} \sum_{\nu=k}^{c-e-1} \nu! \left(\frac{\mu}{\lambda}\right)^{\nu}\right],$$
$$c-d \le k \le c-e.$$

For the ease of computation, the recursive relation

$$\mu\psi_k - \lambda \quad \psi_{k-1} = -(k-1)!\frac{\theta r}{1-r}, \tag{5.5.15}$$

can be used. Using the same approach of treating the M/M/c (SY, MV, d) system, we can verify that $\mathbf{\Pi} B[\mathbf{R}] = 0$ has the positive vector solution as:

$$\pi_{k0} = \begin{cases} K\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k, & 0 \le k \le c-d, \\ K\frac{1}{k!}\frac{\psi_k}{\psi_{c-d}}\left(\frac{\lambda}{\mu}\right)^{c-d}, & c-d \le k \le c-e, \\ K\frac{1}{\psi_{c-d}}\left(\frac{\lambda}{\mu}\right)^{c-d} r^{k-c+e}, & c-e \le k \le c, \end{cases}$$

$$\pi_{k1} = \begin{cases} K\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \frac{r}{1-r}\frac{\theta}{\lambda}\psi_{c-d}^{-1}\sum_{\nu=c-d}^{k-1}\nu!\left(\frac{\mu}{\lambda}\right)^{\nu-c+d}, \\ \qquad c-d+1 \le k \le c-e, \\ K\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \frac{r}{1-r}\frac{\theta}{\lambda}\frac{1}{\psi_{c-d}}\left[\sum_{\nu=c-d}^{c-e-1}\nu!\left(\frac{\mu}{\lambda}\right)^{\nu-c+d}\right. \\ \qquad \left. + \sum_{\nu=c-e}^{k-1}\nu!\left(\frac{\mu}{\lambda}\right)^{\nu-c+d} r^{\nu-c+e}\right], \\ \qquad c-e \le k \le c. \end{cases} \tag{5.5.16}$$

where the constant K can be determined by the normalization condition.

Similarly to the proof of Theorem 5.5.2, we can easily obtain the following theorem.

Theorem 5.5.5. The joint distribution of $\{L, J\}$, denoted by $\{\pi_{kj}, (k,j) \in \Omega\}$ for $0 \le k \le c$, is given by (5.5.16) and

$$\begin{cases} \pi_{k0} = K\left(\frac{\lambda}{\mu}\right)^{c-d}\psi_{c-d}^{-1} r^{k-c+e}, & k \ge c+1, \\ \pi_{k1} = \pi_{c1}\rho^{k-c} + \pi_{c0}\frac{\theta r}{c\mu(1-r)}\sum_{\nu=0}^{k-c-1} r^{\nu}\rho^{k-c-1-\nu}, & k \ge c. \end{cases} \tag{5.5.17}$$

The constant K is

$$K = \left[\begin{array}{c} 1 + \sum_{k=0}^{c-d}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \left(\frac{\lambda}{\mu}\right)^{c-d}\psi_{c-d}^{-1}\left(\frac{1}{1-r} + \sum_{k=c-d+1}^{c-e-1}\frac{\psi_k}{k!}\right) \\ + \sum_{k=c-d+1}^{c-1}\beta_{k1} + \frac{1}{1-\rho}\left(\beta_{c1} + \beta_{c0}\frac{\theta r}{c\mu(1-r)^2}\right) \end{array}\right]^{-1},$$

where $\beta_{k1} = K^{-1}\pi_{k1}, c-d+1 \le k \le c$, and π_{k1} can be determined by (5.5.16) and $\beta_{c0} = \left(\frac{\lambda}{\mu}\right)^{c-d}\psi_{c-d}^{-1} r^e$.

The distribution of the number of customers in the system at any time is

$$P\{L_v = k\} = \begin{cases} \pi_{k0}, & 0 \le k \le c-d, \\ \pi_{k0} + \pi_{k1}, & k \ge c-d+1. \end{cases}.$$

The distribution of the number of busy servers, denoted by M_B, is

$$P\{M_B = j\} = \begin{cases} \pi_{j0}, & 0 \le j \le c-d \\ \pi_{j0} + \pi_{j1}, & c-d+1 \le j < c-e \\ \sum_{\nu=c-e}^{\infty} \pi_{\nu 0} + \pi_{c-e,1} & j = c-e. \\ \pi_{j1} & c-e < j \le c-1, \\ \sum_{\nu=c}^{\infty} \pi_{\nu 1}, & j = c. \end{cases}$$

Let W and $W^*(s)$ be the stationary waiting time and its LST, respectively. To obtain the waiting time distribution, we establish the following lemmas.

Assume that $X^{(\nu)}$ follows an Erlang distribution with parameters α and ν, and V follows an exponential distribution with parameter θ. In addition, $X^{(\nu)}$ and V are independent. Now we have

Lemma 5.5.3. Given $X^{(\nu)} < V$, $\nu \ge 1$, the conditional probability distribution, $\{X^{(\nu)} | X^{(\nu)} < V\}$, follows an Erlang distribution with parameters ν and $\theta + \alpha$.

Proof: Assume that $X^{(\nu)}$ follows an Erlang distribution with parameters ν and α. The p.d.f. and LST are $f_\nu(x) = \frac{\alpha(\alpha x)^{\nu-1}}{(\nu-1)!}e^{-\alpha x}$ and $\widetilde{f}_\nu(s) = \left(\frac{\alpha}{\alpha+\nu}\right)^\nu$ for $\nu \ge 1$, respectively. Also assume that V follows an exponential distribution with parameter θ and is independent of X. It is well known that

$$P\{X^{(v)} < V\} = \left(\frac{\alpha}{\theta+\alpha}\right)^\nu, \qquad \nu \ge 1.$$

Given the event $\{X^{(\nu)} < V\}$, the conditional distribution function of $X^{(\nu)}$ is

$$\begin{aligned} F_{X^{(\nu)}}(x | X^{(\nu)} < V) =& \frac{P\{X^{(\nu)} < x, X^{(\nu)} < V\}}{P\{X^{(\nu)} < V\}} \\ =& \left(\frac{\alpha+\theta}{\alpha}\right)^\nu \int_0^x \frac{\alpha(\alpha t)^{\nu-1}}{(\nu-1)!} e^{-\theta t} e^{-\alpha t} dt \\ =& \int_0^x (\alpha+\theta) \frac{[(\alpha+\theta)t]^{\nu-1}}{(\nu-1)!} e^{-(\theta+\alpha)t} dt. \end{aligned}$$

□

Lemma 5.5.4. Given $\{X^{(\nu)} < V < X^{(\nu+1)}\}$, $\nu \ge 1$, the conditional probability distribution, $\{V | X^{(\nu)} < V < X^{(\nu+1)}\}$, follows an Erlang distribution with parameters $\nu+1$ and $\theta+\alpha$.

Proof: First, it is easy to compute the probability of the conditional event as

$$P\{X^{(\nu)} < V < X^{(\nu+1)}\} = \frac{\theta}{\theta+\alpha}\left(\frac{\alpha}{\theta+\alpha}\right)^\nu, \quad \nu \ge 1. \qquad (5.5.18)$$

From the independence property, we have

$$\begin{aligned}
&P\{V < x, X^{(\nu)} < V < X^{(\nu+1)}\} \\
&\quad = \int_0^x P\{V < x, t < V < t + x\} f_\nu(t)dt \\
&\quad = \int_0^x f_\nu(t)dt \int_t^x e^{-\alpha(u-t)} \theta e^{-\theta u} du \\
&\quad = \frac{\theta}{\theta+\alpha} \int_0^x \left(e^{-\theta t} - e^{-\theta x} e^{-\alpha(x-t)} \right) f_\nu(t)dt.
\end{aligned}$$

Using (5.5.18), given the event $\{X^{(\nu)} < V < X^{(\nu+1)}\}$, the conditional distribution function of V is

$$\begin{aligned}
&F_V(x|X^{(\nu)} < V < X^{(\nu+1)}) \\
&\quad = \frac{P\{V < x, X^{(\nu)} < V < X^{(\nu+1)}\}}{P\{X^{(\nu)} < V < X^{(\nu+1)}\}} \\
&\quad = \left(\frac{\alpha+\theta}{\alpha}\right)^\nu \int_0^x \left(e^{-\theta t} - e^{-\theta x} e^{-\alpha(x-t)} \right) f_\nu(t)dt.
\end{aligned}$$

Taking the derivative with respective to x, we obtain the p.d.f. as

$$\begin{aligned}
&f_V(x|X^{(\nu)} < V < X^{(\nu+1)}) \\
&\quad = \left(\frac{\alpha+\theta}{\alpha}\right)^\nu \int_0^x (\alpha+\theta) e^{-\theta x} e^{-\alpha(x-t)} f_v(t)dt \\
&\quad = (\alpha+\theta) \left(\frac{\alpha+\theta}{\alpha}\right)^\nu e^{-(\alpha+\theta)x} \int_0^x \frac{\alpha(\alpha t)^{\nu-1}}{(\nu-1)!} dt \\
&\quad = (\alpha+\theta) \frac{\left((\alpha+\theta)\, x\right)^\nu}{\nu!} e^{-(\alpha+\theta)x}.
\end{aligned}$$

□

In the following discussion, let $\alpha = (c-e)\mu$ and

$$\alpha_\nu = \frac{\theta}{\theta + (c-e)\mu} \left(\frac{(c-e)\mu}{\theta + (c-e)\mu} \right)^\nu, \qquad \nu \geq 0.$$

Let H_0(or H_1) be the probability that at an arrival instant, e servers are off (or on) duty and the arriving customer has to wait. Obviously we have

$$H_0 = \sum_{\nu=c-e}^{\infty} \pi_{\nu 0} = K \left(\frac{\lambda}{\mu}\right)^{c-d} \psi_{c-d}^{-1} \frac{1}{1-r},$$

$$H_1 = \sum_{\nu=c}^{\infty} \pi_{\nu 1} = \frac{1}{1-\rho} \left(\pi_{c1} + \frac{\theta r}{c\mu(1-r)^2} \pi_{c0} \right).$$

Theorem 5.5.6. The LST of W is

$$\begin{aligned} W^*(s) = 1 - H_0 - H_1 & \\ & + H_0 \frac{\theta + (c-e)\mu(1-r)}{s + \theta + (c-e)\mu(1-r)} \left[\delta + (1-\delta) \frac{c\mu(1-r)}{s + c\mu(1-r)} \right] \\ & + H_1 \frac{c\mu(1-\rho)}{s + c\mu(1-\rho)} \left[\sigma + (1-\sigma) \frac{c\mu(1-r)}{s + c\mu(1-r)} \right], \end{aligned} \tag{5.5.19}$$

where

$$\delta = \frac{\theta(1-r^e) + (c-e)\mu(1-r)}{\theta + (c-e)\mu(1-r)}, \ \sigma = \frac{\pi_{c1}}{H_1(1-\rho)}.$$

Proof: The probability of no waiting is

$$P\{W = 0\} = \sum_{\nu=0}^{c-e-1} \pi_{\nu 0} + \sum_{\nu=c-d+1}^{c-1} \pi_{\nu 1} = 1 - H_0 - H_1.$$

If a customer (called a *tagged customer*) arrives at state $(c-e+j, 0)$, $0 \le j < e$, then the number of waiting customers before the tagged customer is less than e. Therefore, as soon as the vacation is completed, the tagged customer will get immediate service. Before the vacation completion, $c-e$ servers are busy. Based on Lemma 5.5.4, given that $\{X^{(\nu)} < V < X^{(\nu+1)}\}$, $0 \le \nu \le j$, the conditional waiting time of the tagged customer follows the Erlang distribution with parameters $\nu+1$, and $\theta + (c-e)\mu$. If $V > X^{(j+1)}$, which means that the vacation is not completed until the service of the tagged customer starts, then based on Lemma 5.5.3, the conditional waiting time also follows the Erlang distribution with parameters of $j+1$ and $\theta + (c-e)\mu$. Thus, the LST of the waiting time for the tagged customer arriving at state $(c-e+j, 0)$, $0 \le j < e$, is

$$W^*_{c-e+j,0}(s) = \sum_{\nu=0}^{j} \alpha_v \widetilde{f}_{\nu+1}(s) + \left(1 - \sum_{\nu=0}^{j} \alpha_v \right) \widetilde{f}_{j+1}(s).$$

Substituting α_ν and $\widetilde{f}_\nu(s)$ into the equation above gives

$$W^*_{c-e+j,0}(s) = \frac{\theta}{s+\theta} + \frac{s}{s+\theta} \left(\frac{(c-e)\mu}{s + \theta + (c-e)\mu} \right)^{j+1}, \ 0 \le j < e.$$

Therefore, we obtain

$$\sum_{j=0}^{e-1} \pi_{c-e+j,0} W^*_{c-e+j,0}(s)$$
$$= K\left(\frac{\lambda}{\mu}\right)^{c-d} \psi_{c-d}^{-1} \left\{ \frac{\theta}{s+\theta}\frac{1-r^e}{1-r} + \frac{s}{s+\theta}\frac{(c-e)\mu}{s+\theta+(c-e)\mu(1-r)} \right.$$
$$\left. \times \left[1-\left(\frac{(c-e)\mu}{s+\theta+(c-e)\mu}\right)^e r^e\right]\right\}. \quad (5.5.20)$$

If a customer (tagged customer) arrives at state $(c+j,0)$, $j \geq 0$, then the number of waiting customers before this tagged customer is $j+e$. If during the residual vacation, ν services are completed, that is, $\{X^{(\nu)} < V < X^{(\nu+1)}\}$, $0 \leq \nu \leq j$, then, after the e returning servers start serving customers, there are $j-\nu$ customers before the tagged customer. Note that at this vacation completion instant, all c servers are busy. If $\{X^{(\nu)} < V < X^{(\nu+1)}\}$, $j+1 \leq \nu \leq j+e$, then, at this vacation completion instant, the tagged customer gets service immediately. If $\{V > X^{(j+e+1)}\}$, the tagged customer gets the service before the vacation is completed. Thus, based on Lemmas 5.5.3 and 5.5.4, the LST of the conditional waiting time for this customer is

$$W^*_{c+j,0}(s) = \sum_{\nu=0}^{j} \alpha_v \widetilde{f}_{\nu+1}(s)\left(\frac{c\mu}{s+c\mu}\right)^{j-\nu+1} + \sum_{\nu=j+1}^{j+e} \alpha_\nu \widetilde{f}_{\nu+1}(s)$$
$$+ \left(1-\sum_{\nu=0}^{j+e}\alpha_v\right)\widetilde{f}_{j+e+1}(s)$$
$$= \sum_{\nu=0}^{j} \alpha_v \widetilde{f}_{\nu+1}(s)\left(\frac{c\mu}{s+c\mu}\right)^{j-\nu+1}$$
$$+ \left(\frac{(c-e)\mu}{s+\theta+(c-e)\mu}\right)^{j+1}$$
$$\times \left\{\frac{\theta}{s+\theta} + \frac{s}{s+\theta}\left(\frac{(c-e)\mu}{s+\theta+(c-e)\mu}\right)^e\right\}.$$

Therefore, from this expression, we obtain

$$\begin{aligned}
&\sum_{j=0}^{\infty}\pi_{c+j,0}W^*_{c+j,0}(s)\\
&\quad = K\left(\frac{\lambda}{\mu}\right)^{c-d}\psi_{c-d}^{-1}r^e\\
&\qquad \times\left(\frac{c\mu}{s+c\mu(1-r)}\frac{\theta}{s+\theta+(c-e)\mu(1-r)}\right.\\
&\qquad\quad +\left\{\frac{(c-e)\mu}{s+\theta+(c-e)\mu(1-r)}\right.\\
&\qquad\qquad \left.\left.\times\left[\frac{\theta}{s+\theta}+\frac{s}{s+\theta}\left(\frac{(c-e)\mu}{s+\theta+(c-e)\mu}\right)^e\right]\right\}\right). \qquad (5.5.21)
\end{aligned}$$

Using (5.5.20) and (5.5.21) and simplifying the expression yields

$$\begin{aligned}
&\sum_{j=c-e}^{\infty}\pi_{j,0}W^*_{j,0}(s)\\
&= K\left(\frac{\lambda}{\mu}\right)^{c-d}\psi_{c-d}^{-1}\left\{\frac{\theta}{s+\theta}\left[\frac{1-r^e}{1-r}+\frac{1-r^e}{1-r}\frac{(c-e)\mu(1-r)}{s+\theta+(c-e)\mu(1-r)}\right]\right.\\
&\qquad +\frac{1}{1-r}\frac{(c-e)\mu(1-r)}{s+\theta+(c-e)\mu(1-r)}\\
&\qquad \left.+\frac{r^e}{1-r}\frac{\theta}{s+\theta+(c-e)\mu(1-r)}\frac{c\mu(1-r)}{s+c\mu(1-r)}\right\}\\
&= H_0\left\{\frac{\theta+(c-e)\mu(1-r)}{s+\theta+(c-e)\mu(1-r)}\right.\\
&\qquad \left.-\frac{\theta r^e}{s+\theta+(c-e)\mu(1-r)}\frac{s}{s+c\mu(1-r)}\right\}\\
&= H_0\frac{\theta+(c-e)\mu(1-r)}{s+\theta+(c-e)\mu(1-r)}\left\{\frac{\theta(1-r^e)+(c-e)\mu(1-r)}{\theta+(c-e)\mu(1-r)}\right.\\
&\qquad \left.+\frac{\theta r^e}{\theta+(c-e)\mu(1-r)}\frac{c\mu(1-r)}{s+c\mu(1-r)}\right\}\\
&= H_0\frac{\theta+(c-e)\mu(1-r)}{s+\theta+(c-e)\mu(1-r)}\left\{\delta+(1-\delta)\frac{c\mu(1-r)}{s+c\mu(1-r)}\right\}. \qquad (5.5.22)
\end{aligned}$$

Finally, if a customer arrives at state $(c+j,1)$, $j\geq 0$, his or her waiting time follows the Erlang distribution with parameters $j+1$ and $c\mu$. Hence, we have

$$W^*_{c+j,1}(s)=\left(\frac{c\mu}{s+c\mu}\right)^{j+1}, \quad j\geq 0.$$

Using (5.5.16) and (5.5.17), we obtain

$$\begin{aligned}
&\sum_{j=0}^{\infty} \pi_{c+j,1} W^*_{c+j,1}(s) \\
&= \sum_{j=0}^{\infty} \left(\frac{c\mu}{s + c\mu} \right)^{j+1} \left\{ \pi_{c1}\rho^j + \frac{\theta r}{c\mu(1-r)} \pi_{c0} \sum_{\nu=0}^{j-1} r^{\nu} \rho^{j-1-v} \right\} \\
&= \frac{c\mu}{s + c\mu(1-\rho)} \left(\pi_{c1} + \frac{\theta r}{c\mu(1-r)^2} \pi_{c0} \frac{c\mu(1-r)}{s + c\mu(1-r)} \right) \\
&= H_1 \frac{c\mu(1-\rho)}{s + c\mu(1-\rho)} \left\{ \frac{\pi_{c1}}{H_1(1-\rho)} + \frac{\theta r \pi_{c0}}{H_1(1-\rho)c\mu(1-r)^2} \frac{c\mu(1-r)}{s + c\mu(1-r)} \right\} \\
&= H_1 \frac{c\mu(1-\rho)}{s + c\mu(1-\rho)} \left\{ \sigma + (1-\sigma) \frac{c\mu(1-r)}{s + c\mu(1-r)} \right\}. \qquad (5.5.23)
\end{aligned}$$

Combining (5.5.22) and (5.5.23), we have (5.5.19).□

Note that (5.5.19) has an interesting probability interpretation. The stationary waiting time is zero with probability $1 - H_0 - H_1$, is the sum of an exponential random variable of parameter $\theta + (c-e)\mu(1-r)$ and a modified exponential random variable with probability H_0, and is the sum of an exponential random variable of parameter $c\mu(1-\rho)$ and a modified exponential random variable with probability H_1. Thus, the distribution function and the mean of the waiting time are obtained from (5.5.19) as

$$\begin{aligned}
F_W(x) =& 1 - H_0 - H_1 \\
&+ H_0 \left(1 - e^{-[\theta + (c-e)\mu(1-r)]x} \right) \left[\delta + (1-\delta)(1 - e^{-c\mu(1-r)x}) \right] \\
&+ H_1 \left(1 - e^{-c\mu(1-\rho)x} \right) \left[\sigma + (1-\sigma)(1 - e^{-c\mu(1-r)x}) \right].
\end{aligned}$$

and

$$\begin{aligned}
E(W) =& H_0 \left[\frac{1}{\theta + (c-e)\mu(1-r)} + (1-\delta) \frac{1}{c\mu(1-r)} \right] \\
&+ H_1 \left[\frac{1}{c\mu(1-\rho)} + (1-\sigma) \frac{1}{c\mu(1-r)} \right].
\end{aligned}$$

The probability distribution of the waiting time is very useful in computing the service level of queueing systems, such as the probability that a customer waits less than a certain amount of time. Now we present the conditional stochastic decomposition properties in this vacation model.

Let $L^{(1)} = \{L - c | L \geq c, J = 1\}$ and $W^{(1)} = \{W | L \geq c, J = 1\}$ be the conditional queue length and the conditional waiting time, respectively, given that all servers are busy, and let $L_0^{(1)} = \{L - c | L \geq c\}$ and $W_0^{(1)} = \{W | L \geq c\}$ be the corresponding conditional random variables in the classical M/M/c queue.

Theorem 5.5.7. The conditional waiting time and the conditional queue length given that all servers are busy can be decomposed into the sum of two independent random variables,

$$W^{(1)} = W_0^{(1)} + W_d^{(1)},$$

$$L^{(1)} = L_0^{(1)} + L_d^{(1)},$$

where $W_d^{(1)}$ is the additional delay due to the vacation effect and has the LST

$$W_d^{*(1)}(s) = \sigma + (1 - \sigma) \frac{c\mu(1 - r)}{s + c\mu(1 - r)}, \tag{5.5.24}$$

and $L_d^{(1)}$ is the additional queue length due to the vacation effect and has the p.g.f.

$$L_d^{(1)}(z) = \sigma + (1 - \sigma) \frac{z(1 - r)}{1 - zr}. \tag{5.5.25}$$

Proof: Note that $P\{L \geq c, J = 1\} = \sum_{\nu=c}^{\infty} \pi_{\nu 1} = H_1$. Given the condition $\{L \geq c, J = 1\}$, the probability that there are j customers in the system is

$$P\{L^{(1)} = j\} = P\{L = c + j | L \geq c, J = 1\} = H_1^{-1} \pi_{c+j,1},\ j \geq 0.$$

Hence, the LST of $W^{(1)}$ is

$$W^{*(1)}(s) = H_1^{-1} \sum_{j=0}^{\infty} \pi_{c+j,1} \left(\frac{c\mu}{s + c\mu} \right)^{j+1}.$$

Using (5.5.23) in the expression above, we get

$$W^{*(1)}(s) = \frac{c\mu(1 - \rho)}{s + c\mu(1 - \rho)} \left\{ \sigma + (1 - \sigma) \frac{c\mu(1 - r)}{s + c\mu(1 - r)} \right\}.$$

From $P\{L^{(1)} = j\} = H_1^{-1}\pi_{c+j,1}$ and (5.5.17), we have

$$
\begin{aligned}
L^{(1)}(z) &= \sum_{j=0}^{\infty} z^j P\{L^{(1)} = j\} \\
&= H_1^{-1}\left\{\sum_{j=0}^{\infty}(z\rho)^j + \frac{\theta r}{c\mu(1-r)}\pi_{c0}\sum_{j=1}^{\infty} z^j \sum_{\nu=0}^{j-1} r^\nu \rho^{j-1-\nu}\right\} \\
&= H_1^{-1}\left\{\frac{\pi_{c1}}{1-z\rho} + \frac{\theta\mu\pi_{c0}}{c\mu(1-r)}\sum_{\nu=0}^{\infty} z^{\nu+1} r^\nu \sum_{j=\nu+1}^{\infty} z^{j-\nu-1}\rho^{j-\nu-1}\right\} \\
&= H_1^{-1}\frac{1}{1-z\rho}\left\{\pi_{c1} + \frac{\theta\mu}{c\mu(1-r)}\pi_{c0}\frac{z}{1-zr}\right\} \\
&= \frac{1-\rho}{1-z\rho}\left\{\frac{\pi_{c1}}{H_1(1-\rho)} + \frac{\theta\mu}{H_1(1-\rho)c\mu(1-r)^2}\pi_{c0}\frac{z(1-r)}{1-zr}\right\} \\
&= \frac{1-\rho}{1-z\rho}\left\{\sigma + (1-\sigma)\frac{z(1-r)}{1-zr}\right\}.
\end{aligned}
$$

□

Note that (5.5.25) indicates that $L_d^{(1)}$ is zero with probability σ and is 1 plus a geometrically distributed random variable with parameter r with probability $1-\sigma$. Using (5.5.24) and (5.5.25), we have the expected values of these conditional random variables as follows:

$$
\begin{aligned}
E(L_d^{(1)}) &= \frac{1-\sigma}{1-r}, \\
E(W_d^{(1)}) &= \frac{1-\sigma}{1-r}\frac{1}{c\mu}, \\
E(L^{(1)}) &= \frac{\rho}{1-\rho} + \frac{1-\sigma}{1-r}, \\
E(W^{(1)}) &= \frac{1}{c\mu(1-\rho)} + \frac{1-\sigma}{c\mu(1-r)}.
\end{aligned}
$$

Another condition is when e servers are on vacation and the other $c-e$ servers are busy. Let $L^{(0)} = \{L - c + e | L \geq c - e, J = 0\}$ and $W^{(0)} = \{W | L \geq c - e, J = 0\}$. We also have the conditional stochastic decomposition property for the conditional waiting time.

Theorem 5.5.8. $L^{(0)}$ follows a geometric distribution with parameter r. $W^{(0)}$ can be decomposed into the sum of two independent random variables,

$$W^{(0)} = W_0^{(0)} + W_d^{(0)},$$

where $W_0^{(0)}$ follows an exponential distribution with parameter $\theta + (c - e)\mu(1-r)$, and $W_d^{(0)}$ follows a modified exponential distribution with the LST

$$W_d^{(0)}(s) = \delta + (1-\delta)\frac{c\mu(1-r)}{s + c\mu(1-r)}. \tag{5.5.26}$$

Proof: Note that $P\{L \geq c-e, J=0\} = \sum_{\nu=c-e}^{\infty} \pi_{\nu 0} = H_0$. Thus, the probability distribution of $L^{(0)}$ is

$$P\{L^{(0)} = j\} = P\{L = c-e+j | L \geq c-e, J=0\} = H_0^{-1}\pi_{c-e+j,0}, \quad j \geq 0.$$

Taking the p.g.f. of this distribution gives

$$\begin{aligned} L^{(0)}(z) &= \sum_{j=0}^{\infty} z^j P\{L^{(0)} = j\} = H_0^{-1} \sum_{j=0}^{\infty} z^j \pi_{c-e+j,0} \\ &= H_0^{-1} K \left(\frac{\lambda}{\mu}\right)^{c-d} \frac{1}{1-r} \psi_{c-d}^{-1} \frac{1-r}{1-zr} = \frac{1-r}{1-zr}. \end{aligned}$$

Therefore, $L^{(0)}$ follows a geometric distribution with parameter r. Given the condition of $\{L^{(0)} = j\}$, the waiting time is no longer the sum of $j+1$ exponential random variables with parameter $(c-e)\mu$. As indicated in the proof of Theorem 5.5.6, the waiting process also depends on the vacation completion instant. Note that (5.5.22) gives the joint distribution of W and event $\{L \geq c-e, J=0\}$, and hence, the LST of the conditional waiting time $W^{(0)}$ is

$$W^{*(0)}(s) = \frac{\theta + (c-e)\mu(1-r)}{s + \theta + (c-e)\mu(1-r)} \left\{ \delta + (1-\delta)\frac{c\mu(1-r)}{s + c\mu(1-r)} \right\}.$$

This completes the proof. □

5.6 M/M/c Queue with Asynchronous Vacations of Some Servers

In this section, we consider an M/M/c queue where servers can take vacations independently when they become idle. The service policy now prescribes the following: at a service completion instant or at a vacation completion instant, if the server finds no waiting customers and the number of servers on vacations is less than d, this server will take a vacation individually. With such a policy, the number of servers on duty (busy or idle) is at least $c-d$ at any time. Because servers take vacations individually and continue taking vacations if the vacation condition is satisfied, the vacation policy is called an *asynchronous multiple vacation policy*. The vacation time is assumed to be exponentially distributed with parameter θ. The service order is FCFS and interarrival times, service times, and vacation times are mutually independent. This system is denoted by M/M/c (AS, MV, d).

With this vacation policy, if the number of customers $k \le c-d$, there must be d servers on vacations and $c-d-k$ servers that stay idle; if $c-d < k \le c$, there are at least $c-k$ servers on vacations and no idle servers.

Let $L_v(t)$ be the number of customers in the system at time t, and let $J(t)$ be the number of servers on vacations at time t. Then $0 \le J(t) \le d$, and $\{L_v(t), J(t)\}$ is a QBD with the state space

$$\begin{aligned}\Omega =&\{(k,d) : 0 \le k \le c-d\} \cup \{(k,j) : c-d < k \le c-1, c-k \le j \le d\}\\ &\cup \{(k,j) : k \ge c, 0 \le j \le d\}.\end{aligned}$$

For a given k, the state set $\{(k,j),\ (k,j) \in \Omega\}$, called *level* k, contains the states that are sequenced in descending j starting with $j = d$. Using the lexicographical sequence for the states, the infinitesimal generator for the QBD has the same block structure as in (5.3.3), where $\mathcal{A}_0, \mathcal{B}_1$, and $\mathcal{C}_0$ can be written in the block-partitioned form as in (5.4.1). Letting $h_k = \lambda + (c-k)\mu + k\theta$, $0 \le k \le d$, the submatrices of the infinitesimal generator are given by

$$\begin{aligned} A_k &= -(\lambda + k\mu), && 0 \le k \le c-d,\\ B_k &= k\mu, && 1 \le k \le c-d,\\ C_k &= \lambda, && 0 \le k \le c-d-1, \end{aligned}$$

$$\mathbf{B}_k = \begin{bmatrix} (c-d)\mu & & & \\ & (c-d+1)\mu & & \\ & & \ddots & \\ & & & (k-1)\mu \\ 0 & 0 & \cdots & k\mu \end{bmatrix}, \quad c-d < k \le c-1;$$

$$\mathbf{C}_k = \begin{bmatrix} \lambda & & & 0 \\ & \lambda & & 0 \\ & & \ddots & \vdots \\ & & \lambda & 0 \end{bmatrix}, \qquad c-d \le k < c-1;$$

$$\mathbf{A}_k = \begin{bmatrix} -h_d & d\theta & & & & \\ & -h_{d-1} & (d-1)\theta & & & \\ & & \ddots & \ddots & & \\ & & & -h_{c-k-1} & (c-k-1)\theta \\ & & & & -(\lambda+k\mu) \end{bmatrix},$$

$$c-d < k \le c-1.$$

Other submatrices $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are the $(d+1)\times(d+1)$ matrices, as follows:

$$\mathbf{A} = \begin{bmatrix} -h_d & d\theta & & & & \\ & -h_{d-1} & (d-1)\theta & & & \\ & & \ddots & \ddots & & \\ & & & -h_1 & \theta \\ & & & & -(\lambda+c\mu) \end{bmatrix},$$

$\mathbf{B} = diag((c-d)\mu, (c-d+1)\mu, ..., c\mu)$, and $\mathbf{C} = \lambda\mathbf{I}$. Therefore, $\mathcal{A}_0$ is the square matrix with order $d^* = (c-d) + \frac{1}{2}d(d+1)$. $\mathcal{B}_1$ and $\mathcal{C}_0$ are the $(d+1)\times d^*$ and $d^* \times (d+1)$ matrices, respectively.

To obtain the explicit expression for $\mathbf{R}$, we need the following lemmas.

Lemma 5.6.1. For any $0 \le k < d$, the quadratic equation

$$(c-d+k)\mu z^2 - [\lambda + (c-d+k)\mu + (d-k)\theta]z + \lambda = 0 \qquad (5.6.1)$$

has two different real roots $r_k < r_k^*$ and $0 < r_k < 1$, $r_k^* > 1$.

Proof: Let $j = c-d+k$. Then $c-d+1 \le j \le c$, and hence, (5.6.1) can be rewritten as

$$j\mu z^2 - [\lambda + j\mu + (c-j)\theta]z + \lambda = 0, \qquad c-d+1 \le j < c.$$

Then the result follows from the same approach used in the proof of Lemma 5.4.1.□

If $k = d$, (5.6.1) becomes

$$c\mu z^2 - (\lambda + c\mu)z + \lambda = 0,$$

and its two roots are $r_d = \rho = \lambda(c\mu)^{-1}$ and $r_d^* = 1$.

Lemma 5.6.2. The rate matrix $\mathbf{R}$ satisfies $\mathbf{RBe} = \lambda\mathbf{e}$.

Proof: Note that $\mathbf{Ae} = -(\lambda\mathbf{e} + \mathbf{Be})$. Multiplying both sides of (5.5.3) from the right by $\mathbf{e}$ gives

$$\mathbf{R}^2\mathbf{Be} - \mathbf{R}(\lambda\mathbf{e} + \mathbf{Be}) + \lambda\mathbf{e} = \mathbf{0},$$

and rearranging the terms results in

$$(\mathbf{I}-\mathbf{R})(\lambda\mathbf{e}-\mathbf{RBe})=\mathbf{0}.$$

Because the inverse of $\mathbf{I}-\mathbf{R}$ exists, so $\lambda\mathbf{e}=\mathbf{RBe}$. □

Now we show the theorem for computing $\mathbf{R}$.

Theorem 5.6.1. If $\rho=\lambda(c\mu)^{-1}<1$, the matrix equation $\mathbf{R}^2\mathbf{B}+\mathbf{RA}+\mathbf{C}=0$ has the minimal nonnegative solution

$$\mathbf{R}=\begin{bmatrix} r_0 & r_{01} & r_{02} & \cdots & r_{0,d-1} & r_{0d} \\ & r_1 & r_{12} & \cdots & r_{1,d-1} & r_{1d} \\ & & r_2 & \cdots & r_{2,d-1} & r_{2d} \\ & & & \cdots & \cdots & \cdots \\ & & & & r_{d-1} & r_{d-1,d} \\ & & & & & \rho \end{bmatrix}, \tag{5.6.2}$$

where r_k, $0\le k\le d-1$, is the solution of (5.6.1) that is between 0 and 1 and the nondiagonal entries in $\mathbf{R}$ satisfy the equations

$$\begin{aligned}(c-d+k)\mu\sum_{i=j}^{k} r_{ji}r_{ik}-[\lambda+(c-d+k)\mu+(d-k)\theta]r_{jk}\\ +(d-k+1)\theta r_{j,k-1}=0,\\ 0\le j\le d-1,\quad j+1\le k\le d.\end{aligned} \tag{5.6.3}$$

In (5.6.3), if $k=j$, let $r_{kk}=r_k$, $0\le k\le d$.

Proof: Since $\mathbf{A},\mathbf{B}$, and $\mathbf{C}$ are all upper-triangular matrices, the solution to the matrix equation, $\mathbf{R}$, must be an upper-triangular matrix. Let $\mathbf{R}$ be in form of (5.6.2). Then the entries of $\mathbf{R}^2$ are

$$(\mathbf{R}^2)_{kk}=r_k^2,\qquad 0\le k\le d,$$

$$(\mathbf{R}^2)_{jk}=\sum_{i=j}^{k} r_{ji}r_{ik},\ 0\le j\le d-1, j<k\le d.$$

Substituting $\mathbf{R}^2,\mathbf{R},\mathbf{A},\mathbf{B}$,and $\mathbf{C}$ into the matrix equation gives a set of equations:

$$\begin{cases}(c-d+k)\mu r_k^2-[\lambda+(c-d+k)\mu+(d-k)\theta]r_k+\lambda=0, \\ \qquad\qquad 0\le k\le d, \\ (c-d+k)\mu\sum_{i=j}^{k} r_{ji}r_{ik}+(d-k+1)\theta r_{j,k-1} \\ \quad =[\lambda+(c-d+k)\mu+(d-k)\theta]r_{jk}, \\ \qquad\qquad 0\le j\le d-1,\ j+1\le k\le d.\end{cases} \tag{5.6.4}$$

Based on Lemma 5.6.1, we can obtain the minimal nonnegative solution by letting r_k be the root of (5.6.1) in (0,1) where $0 \le k \le d-1$, and letting $r_d = \rho$. The second equation of (5.6.4) is the recursive relation (5.6.3).□

From (5.6.2), we find that the spectral radius

$sp(\mathbf{R}) = \max(r_0, \cdots, r_{d-1}, \rho)$, so $sp(\mathbf{R}) < 1$ if and only if $\rho < 1$. Therefore, $\rho < 1$ is the necessary and sufficient condition for $\{(L_v(t), J(t)), t \ge 0\}$ to be positive recurrent.

Because (5.6.3) is a set of nonlinear recursions, it is not possible to get the explicit expression for every r_{jk} $(j < k)$. However, as in section 5.4.1, it is feasible to recursively compute every r_{jk}. In addition, $\mathbf{RBe} = \lambda\mathbf{e}$ in Lemma 5.6.2 is a set of d linear equations that the nondiagonal entries satisfy. Note that we cannot use these $d+1$ equations to determine every nondiagonal entry. However, we can use the recursive relations in (5.6.3) and Lemma 5.6.2 jointly to determine the nondiagonal entries of $\mathbf{R}$. For example, letting $k = j+1$ in (5.6.3) and using the same method of section 5.4.1, we obtain

$$r_{j,j+1} = \frac{d-j}{c-d+j+1}\frac{\theta}{\mu}\frac{r_j}{r^*_{j+1} - r_j}, \qquad j = 0, 1, ...d-1.$$

With this relation, we can compute these entries on the first off-diagonal line parallel to the diagonal of $\mathbf{R}$.

If $\rho < 1$, let $\{L_v, J\}$ be the queue length and the number of vacationing servers for the steady state system. Denote its joint probability by

$$\pi_{kj} = P\{L_v = k, J = j\} = \lim_{t\to\infty} P\{L_v(t) = k, J(t) = j\}, \ (k, j) \in \Omega.$$

To accommodate the block structure of $\mathbf{Q}$, we express the distribution of $\{L_v, J\}$ as three probability vectors

$$\pi_k = \begin{cases} \pi_{kd}, & 0 \le k \le c-d, \\ (\pi_{kd}, \pi_{k,d-1}, \cdots, \pi_{k,c-k}), & c-d < k \le c-1, \\ (\pi_{kd}, \pi_{k,d-1}, \cdots, \pi_{k,1}, \pi_{k0}), & k \ge c, \end{cases}$$

where π_k, $0 \le k \le c-d$, is a real number; π_k, $c-d+1 \le k \le c-1$, is a $(k-c+d+1)$ -dimensional row vector; and π_k, $k \ge c$, is a $(d+1)$ -dimensional row vector. The marginal probability

$$\mathbf{\Pi}_c = (\pi_0, \cdots, \pi_{c-d}, \pi_{c-d+1}, \cdots, \pi_c)$$

is a row vector of $(c-d) + \frac{1}{2}(d+1)(d+2)$ dimensions.

Note that the square matrix of order $(c-d)+\frac{1}{2}(d+1)(d+2)$,

$$B[\mathbf{R}] = \begin{bmatrix} \mathcal{A}_0 & \mathcal{C}_0 \\ \mathcal{B}_1 & \mathbf{RB}+\mathbf{A} \end{bmatrix}$$
$$= \begin{bmatrix} A_0 & C_0 & & & & \\ B_1 & A_1 & C_1 & & & \\ & \vdots & \vdots & \vdots & & \\ & & & \mathbf{B}_{c-1} & \mathbf{A}_{c-1} & \mathbf{C}_{c-1} \\ & & & & \mathbf{B}_c & \mathbf{RB}+\mathbf{A} \end{bmatrix},$$

is an aperiodic and irreducible infinitesimal generator with finite dimensions. Hence,

$$\mathbf{\Pi}_c B[\mathbf{R}] = \mathbf{0} \tag{5.6.5}$$

must have positive solutions. Using Theorem 5.2.3, we obtain the following theorem.

Theorem 5.6.2. If $\rho < 1$, the joint distribution of (L_v, J) is

$$\pi_k = \begin{cases} K\beta_k, & 0 \le k \le c-d, \\ K\beta_k = K(\beta_{kd}, \beta_{k,d-1}, \cdots, \beta_{k,k-c}), & c-d < k \le c-1, \\ K\beta_k = K\beta_c\mathbf{R}^{k-c}, & k \ge c, \end{cases} \tag{5.6.6}$$

where $(\beta_0, \cdots, \beta_{c-d,}\beta_{c-d+1} \cdots, \beta_c)$ is the positive solution to (5.6.5) and the constant K is

$$K = \left\{ \sum_{j=0}^{c-d} \beta_j + \sum_{j=c-d+1}^{c-1} \beta_j\mathbf{e} + \beta_c(\mathbf{I}-\mathbf{R})^{-1}\mathbf{e} \right\}^{-1}.$$

Furthermore, we can get the distribution of the number of customers in the system:

$$P\{L_v = k\} = \begin{cases} \pi_k, & 0 \le k \le c-d, \\ \pi_k\mathbf{e}, & c-d < k \le c-1, \\ \pi_c\mathbf{R}^{k-c}\mathbf{e}, & k \ge c. \end{cases}$$

Note that, from Theorem 5.6.2, the distribution of the waiting time can be obtained by conditioning on every state $(k, j) \in \Omega$. However, this approach is very complex and not convenient to use.

Let $L_v^{(c)} = \{L_v - c | L_v \ge c, J = 0\}$ and $W_v^{(c)} = \{W_v | L_v \ge c, J = 0\}$ be the conditional queue length and the conditional waiting time, respectively, given that all servers are busy. Rewrite β_c and $\mathbf{R}$ as

$$\beta_c = (\delta, \beta_{c0}), \qquad \mathbf{R} = \begin{pmatrix} \mathbf{H} & \eta \\ \mathbf{0} & \rho \end{pmatrix},$$

where $\delta = (\beta_{cd}, \beta_{c,d-1}, \cdots, \beta_{c1})$ is a d-dimensional vector. Comparing with (5.6.2), we find that $\mathbf{H}$ is a $d \times d$ matrix and η is a $d \times 1$ column vector as follows:

$$\mathbf{H} = \begin{bmatrix} r_0 & r_{01} & \cdots & r_{0,d-1} \\ & r_1 & \cdots & r_{1,d-1} \\ & & \ddots & \vdots \\ & & & r_{d-1} \end{bmatrix}, \qquad \eta = \begin{bmatrix} r_{0d} \\ r_{1d} \\ \vdots \\ r_{d-1,d} \end{bmatrix}.$$

Obviously, $sp(\mathbf{H}) < 1$.

Theorem 5.6.3. If $\rho < 1$, the conditional queue length $L_v^{(c)}$ can be decomposed into the sum of two independent random variables,

$$L_v^{(c)} = L_0^{(c)} + L_d,$$

where $L_0^{(c)}$ is the conditional queue length of the classical M/M/c system without vacation and follows a geometric distribution with parameter ρ. L_d is the additional queue length due to the vacation effect and follows a matrix geometric distribution of order d. L_d has the p.g.f.

$$L_d(z) = \frac{1}{\sigma}\left\{\beta_{c0} + z\delta(\mathbf{I} - z\mathbf{H})^{-1}\eta\right\}, \tag{5.6.7}$$

where

$$\sigma = \beta_{c0} + \delta(\mathbf{I} - \mathbf{H})^{-1}\eta.$$

Proof: Based on the structure of $\mathbf{R}$, we have

$$\mathbf{R}^k = \begin{pmatrix} \mathbf{H}^k & \sum_{j=0}^{k-1} \rho^j \mathbf{H}^{k-1-j}\eta \\ \mathbf{0} & \rho^k \end{pmatrix}, \qquad k \geq 1.$$

Substituting the kth power of $\mathbf{R}$ and $\beta_c = (\delta, \beta_{c0})$ into the matrix geometric expression in Theorem 5.6.2 yields

$$\pi_k = K(\delta\mathbf{H}^{k-c},\ \beta_{c0}\rho^{k-c} + \delta\sum_{j=0}^{k-c-1} \rho^j\mathbf{H}^{k-c-1-j}\eta), \qquad k \geq c.$$

If $k = c$, the empty sum of the second term is 0, so the last element of π_k is

$$\pi_{k0} = K(\beta_{c0}\rho^{k-c} + \delta\sum_{j=0}^{k-c-1} \rho^j\mathbf{H}^{k-c-1-j}\eta), \qquad k \geq c.$$

The probability that all servers are busy is

$$
\begin{aligned}
P\{L_v \geq c, J=0\} =& \sum_{k=c}^{\infty} \pi_{k0} \\
=& K\left\{\beta_{c0}\frac{1}{1-\rho} + \delta\sum_{k=c+1}^{\infty}\sum_{j=0}^{k-c-1}\rho^j \mathbf{H}^{k-c-1-j}\eta\right\} \\
=& \frac{K}{1-\rho}\left\{\beta_{c0} + \delta(\mathbf{I}-\mathbf{H})^{-1}\eta\right\} = \frac{K}{1-\rho}\sigma.
\end{aligned}
$$

The distribution of $L_v^{(c)}$ is

$$
\begin{aligned}
P\{L_v^{(c)} = k\} =& P\{L_v = k+c | L_v \geq c, J=0\} \\
=& \frac{1-\rho}{K\sigma}\pi_{k+c,0} \\
=& \frac{1-\rho}{\sigma}\left\{\beta_{c0}\rho^k + \delta\sum_{j=0}^{k-1}\rho^j\mathbf{H}^{k-1-j}\eta\right\}, \quad k \geq 0. \qquad (5.6.8)
\end{aligned}
$$

Taking the p.g.f. of (5.6.8), we have

$$
\begin{aligned}
L_v^{(c)}(z) =& \frac{1-\rho}{\sigma}\left\{\beta_{c0}\sum_{k=0}^{\infty}(z\rho)^k + \delta\sum_{k=1}^{\infty}z^k\sum_{j=0}^{k-1}\rho^j\mathbf{H}^{k-1-j}\eta\right\} \\
=& \frac{1-\rho}{\sigma}\left\{\frac{\beta_{c0}}{1-z\rho} + z\delta\sum_{j=0}^{\infty}(z\rho)^j\sum_{k=j+1}^{\infty}(z\mathbf{H})^{k-1-j}\eta\right\} \\
=& \frac{1-\rho}{1-z\rho}\frac{1}{\sigma}\left\{\beta_{c0} + z\delta(\mathbf{I}-z\mathbf{H})^{-1}\eta\right\} \\
=& L_0^{(c)}(z)L_d(z).
\end{aligned}
$$

Expanding (5.6.7), we obtain

$$
P\{L_d = k\} = \begin{cases} \frac{1}{\sigma}\beta_{c0}, & k=0, \\ \frac{1}{\sigma}\delta\mathbf{H}^{k-1}\eta, & k \geq 1. \end{cases}.
$$

Hence, L_d follows a matrix geometric distribution. □

Note that (5.6.7) implies that L_d has a PH expression of order d. However, $(\sigma^{-1}\delta, \mathbf{H})$ may not be a PH representation because $\mathbf{H}$ may not be a stochastic submatrix. Sengupta (1991) proved that the distribution of L_d must be a discrete PH distribution of order d and provided a method of constructing the PH representation for this type of distribution.

From Theorem 5.6.3, we find that the expected conditional queue length, given that all servers are busy, is

$$E(L_v^{(c)}) = \frac{1}{1-\rho} + \frac{1}{\sigma}\delta(\mathbf{I}-\mathbf{H})^{-2}\eta.$$

The following theorem gives the conditional stochastic decomposition property of the waiting time.

Theorem 5.6.4. If $\rho < 1$, $W_v^{(c)}$ can be decomposed into the sum of two independent random variables

$$W_v^{(c)} = W_0^{(c)} + W_d,$$

where $W_0^{(c)}$ is the conditional waiting time in the classical M/M/c when all servers are busy and follows an exponential distribution with parameter $c\mu(1-\rho)$. W_d is the additional delay due to the vacation effect and has the LST

$$W_d^*(s) = \frac{1}{\sigma}\left\{\beta_{c0} + c\mu\delta(s\mathbf{I} - c\mu(\mathbf{H}-\mathbf{I}))^{-1}\eta\right\}. \tag{5.6.9}$$

Proof: Assume that a customer arrives at state $(k, 0)$ for $k \geq c$, if we condition on this state, this customer's waiting time, W_{k0}, has the LST

$$W_{k0}^*(s) = \left(\frac{c\mu}{s+c\mu}\right)^{k-c+1}, \qquad k \geq c.$$

The LST of the conditional waiting time when all servers are busy is

$$\begin{aligned}
W_v^{*(c)}(s) &= \sum_{k=c}^{\infty} P\{L_v^{(c)} = k\} W_{k0}^*(s) \\
&= \frac{1-\rho}{\sigma}\left\{\beta_{c0}\sum_{k=c}^{\infty}\rho^{k-c}\left(\frac{c\mu}{s+c\mu}\right)^{k-c+1}\right. \\
&\qquad \left. + \delta\sum_{k=c+1}^{\infty}\left(\frac{c\mu}{s+c\mu}\right)^{k-c+1}\sum_{j=0}^{k-c-1}\rho^j\mathbf{H}^{k-c-1-j}\eta\right\} \\
&= \frac{c\mu(1-\rho)}{s+c\mu(1-\rho)}\frac{1}{\sigma}\left\{\beta_{c0} + \delta\left(\mathbf{I} - \frac{c\mu}{s+c\mu}\mathbf{H}\right)^{-1}\eta\right\} \\
&= \frac{c\mu(1-\rho)}{s+c\mu(1-\rho)}\frac{1}{\sigma}\left\{\beta_{c0} + c\mu\delta\left(s\mathbf{I} - c\mu(\mathbf{H}-\mathbf{I})\right)^{-1}\eta\right\} \\
&= W_0^*(s)W_d^*(s).
\end{aligned}$$

□

Based on (5.6.9), the distribution function of W_d can be written as

$$P(W_d \le x\} = 1 - \frac{1}{\sigma}\delta \exp\{-c\mu(\mathbf{I}-\mathbf{H})x\}(\mathbf{I}-\mathbf{H})^{-1}\eta, \qquad x \ge 0.$$

This expression indicates that the additional delay W_d follows a matrix exponential distribution. From Theorem 5.6.4, we can get the mean of the conditional waiting time:

$$E(W_v^{(c)}) = \frac{1}{c\mu(1-\rho)} + \frac{1}{c\mu\sigma}\delta(\mathbf{I}-\mathbf{H})^{-2}\eta = \frac{1}{c\mu}E(L_v^{(c)}). \tag{5.6.10}$$

5.7 Bibliographic Notes

Avi-Itzahak and Naor (1962) studied the M/M/c queue where servers are subject to failures. Neuts (1981) also analyzed the M/M/c queue with a repairable server by using the QBD process. The early work on the multiserver vacation model was done by Levy and Yechiali (1976). They treated the M/M/c queue where servers may take individual vacations (asynchronous vacations) and obtained the expected number of customers in the system and the stationary distribution of the number of busy servers. Due to the complexity of multiserver vacation systems, the classical birth-and-death-process approach is not appropriate. The matrix analytical method (MAM) developed in the 1980s is more suitable for analyzing the multiserver vacation systems that can be formulated as QBD processes. For stochastic models analyzed by the MAM, see Neuts (1981, 1989, 1995) and Latouche and Rammaswami (1999). Vinod (1986) first used the QBD process to study the M/M/c queue with exponential vacations and suggested using the numerical method to find the stationary distributions of queue length and waiting time. Using the MAM and the numerical method, Chao and Zhao (1998) analyzed the multiserver vacation models with *station* vacations (synchronous vacations) or *server* vacations (asynchronous vacations). Igaki (1992) studied an M/M/2 queue where only one server is allowed to take vacations and showed the conditional stochastic decomposition properties for the performance measures when two servers are busy. Madan et al. (2003) presented an analysis of a two-server vacation model with a Bernoulli schedule and a single vacation. The M/M/c vacation model presented in section 5.3 was presented in Tian and Li (2000). They used the QBD process to analyze multiserver systems with PH-type vacations or setup times. The M/M/c queue with asynchronous vacations presented in section 5.4 was obtained in Tian et al. (1999). Multiserver queues with some-server vacations of both asynchronous and synchronous types were studied by Zhang and Tian (2003a, 2003b). Multiserver vacation models with two or three threshold policies were treated by Tian and Zhang (2004) and Zhang and Tian (2004). For a variety of other M/M/c vacation models with some server vacations and threshold policies, see Zhang and Tian (2003a, 2003b) and Tian and Zhang (2004, 2006). Most past studies on multiserver vacation models were M/M/c systems; other types of multiserver vacation models are more difficult to study. GI/M/c type vacation models will be discussed in the next chapter. Like the non-vacation M/G/c queues, M/G/c type vacation models might be studied in the future by using approximation methods.