

Chapter 2

M/G/1 TYPE VACATION MODELS: EXHAUSTIVE SERVICE

This chapter focuses on single server vacation systems where the server follows an exhaustive-service policy: in other words, the server does not take any vacations until the system becomes empty. The systems considered are the M/G/1 type, where interarrival times are exponentially distributed i.i.d. random variables and service times are generally distributed i.i.d. random variables. The rules for resuming queue service at a vacation completion instant are numerous. However, they can be generally classified into two categories. The rules in the first category are mainly based on the number of vacations taken before the first customer arrives at the empty system. These rules usually require the server to serve the queue at a vacation completion instant if waiting customers exist. The rules in the second category are based on the number of waiting customers at a vacation completion instant. If the server returns to serve the queue only when the number of waiting customers reaches a critical value, the rule is called a *threshold policy*. In section 2.1, we consider the multiple adaptive vacation (MAV) policy, a general rule of the first category. In section 2.2, we demonstrate that several common vacation models are special cases of the MAV policy model. The threshold policy models are presented in section 2.3. Other variations of the M/G/1 type exhaustive-service models are also discussed in this chapter. Specifically, the discrete-time vacation models are presented in section 2.4. Vacation models with Markov arrival process (MAP) are considered in section 2.5. Vacation models with batch arrivals or batch services are discussed in section 2.6. Finally, the finite-buffer vacation models are given in section 2.7.

2.1 M/G/1 Queue with Multiple Adaptive Vacations

2.1.1 Classical M/G/1 Queue

We first present briefly some well-known results for a classical M/G/1 queue without vacations. The details of developing these results can be found in any queueing theory books (for example, see Gross and Harris (1985)). In such a system, customers arrive according to a Poisson process with rate λ and service times are i.i.d random variables with a general distribution function, denoted by $B(t)$. Let

$$\frac{1}{\mu} = \int_0^\infty t dB(t), \quad b^{(2)} = \int_0^\infty t^2 dB(t), \quad B^*(s) = \int_0^\infty e^{-st} dB(t).$$

Assume that the service order is first-come-first-served (FCFS) and that interarrival times and service times are independent.

Denote by L_n the number of customers in the system at the n th customer departure instant, $\{L_n, n \geq 1\}$ is an embedded Markov chain of the queueing process, satisfying

$$L_{n+1} = \begin{cases} L_n - 1 + A_{n+1}, & L_n \geq 1, \\ A_{n+1}, & L_n = 0, \end{cases}$$

where A_{n+1} is the number of arrivals during the $(n+1)$ service time. Obviously these numbers are i.i.d. random variables and can be denoted by A , with respective probability distribution and mean

$$a_j = P(A = j) = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dB(t), \quad j \geq 0, \quad E(A) = \frac{\lambda}{\mu} = \rho.$$

ρ is called the *traffic intensity* of the system and is the ratio of arrival rate to service rate. The probability generating function (p.g.f.) of A is $A(z) = B^*(\lambda(1-z))$, and the transition probability matrix of the embedded Markov chain is

$$\mathbf{P} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ & a_0 & a_1 & a_2 & \cdots \\ & & a_0 & a_1 & \cdots \\ & & & \vdots & \ddots \end{bmatrix}. \quad (2.1.1)$$

It can be proved that $\{L_n, n \geq 1\}$ is positive recurrent and the system reaches the steady state if and only if $\rho < 1$. Therefore, when $\rho < 1$, the p.g.f.s of the stationary number of customers in the system, L , and the

stationary number of customers waiting in line, Q , and the LST of the stationary waiting time, W , are as follows:

$$\begin{aligned} L(z) &= \frac{(1-\rho)(1-z)B^*(\lambda(1-z))}{B^*(\lambda(1-z)) - z}, \\ Q(z) &= \frac{(1-\rho)(1-z)}{B^*(\lambda(1-z)) - z}, \\ W^*(s) &= \frac{(1-\rho)s}{s - \lambda(1 - B^*(s))}. \end{aligned} \quad (2.1.2)$$

The means of these stationary random variables are, respectively,

$$\begin{aligned} E(L) &= \rho + \frac{\lambda^2 b^{(2)}}{2(1-\rho)}, \\ E(Q) &= \frac{\lambda^2 b^{(2)}}{2(1-\rho)}, \\ E(W) &= \frac{\lambda b^{(2)}}{2(1-\rho)} = \frac{1}{\lambda} E(Q). \end{aligned} \quad (2.1.3)$$

These formulas are called *Pollaczek-Khinchin formulas*. Note that (2.1.2) gives the p.g.f. of the queue length distribution at a customer departure instant, called the *departure distribution*. It can be shown that the departure distribution is the same as the distribution seen by an arriving customer, called the *arrival distribution*. Furthermore, due to the well-known Poisson Arrivals See Time Averages (PASTA) property (see Wolff (1982)), the arrival distribution is the same as the distribution of the queue length at any time t . Therefore, the departure distribution obtained in (2.1.2) is the same as the distribution at any time. This important property holds in all M/G/1 vacation models discussed in this chapter.

A busy period, denoted by D , is defined as the period from the arrival instant of the first customer at an empty system to the departure instant of a customer that leaves an empty system. It is well known that the LST of D satisfies the functional relation

$$D^*(s) = B^*(s + \lambda(1 - D^*(s))).$$

Based on this relation, the mean of the busy period is obtained as

$$E(D) = \frac{1}{\mu(1-\rho)} = \frac{1}{\lambda - \mu}. \quad (2.1.4)$$

2.1.2 Multiple Adaptive Vacation Model

In an M/G/1 queue, the server follows the following vacation policy. When the server finishes serving all customers in the system, it starts to take a vacation. The server will take vacations consecutively until either a customer has arrived at a vacation completion instant or a maximum number, denoted by H , of vacations have been taken. In the case of arrivals occurred during a vacation, the server resumes serving the queue immediately at that vacation completion instant. In the case of no arrivals occurring after the server has completed H vacations, the server stays idle and waits to serve the next arrival. H , called the *stages of vacations*, is assumed to be a discrete random variable, with respective distribution and p.g.f.

$$P\{H = j\} = h_j, \quad j \geq 1; \quad H(z) = \sum_{j=1}^{\infty} h_j z^j.$$

The consecutive vacations, denoted by V_k , $k = 1, 2, \dots, H$, are i.i.d. random variables with the distribution function of $V(x)$, the LST of $v^*(s)$, and the finite first and second moments. The queueing system of this policy is called a *vacation model with exhaustive service, multiple adaptive vacations* (MAV), or simply an *E-MAV model*, denoted by M/G/1 (E, MAV). The E-MAV policy reflects the flexibility of allowing the server to work on both the primary randoml-arrival jobs (the queue) and a random number of secondary jobs (the vacations) during the idle time. Assume that the interarrival times, the service times, the vacation times, and the stages of vacations are mutually independent and the service order is FCFS.

Define two events

$A_I = \{\text{a busy period starts with the ending of an idle period}\},$

$A_v = \{\text{a busy period starts with the ending of a vacation}\},$

we have

$$\begin{aligned} P\{A_I\} &= \sum_{j=1}^{\infty} P\{H = j\} P\{T > V_1 + \dots + V_j\} \\ &= \sum_{j=1}^{\infty} h_j \int_0^{\infty} e^{-\lambda t} dV^{(j)}(t) \\ &= \sum_{j=1}^{\infty} h_j [v^*(\lambda)]^j = H[v^*(\lambda)], \end{aligned}$$

where $V^{(j)}(t)$ is the j th convolution of $V(t)$. Obviously,

$$P\{A_v\} = 1 - H[v^*(\lambda)].$$

Letting L_n be the number of customers left behind by the n th customer, we have

$$L_{n+1} = \begin{cases} L_n - 1 + A, & \text{for } L_n \geq 1, \\ Q_b - 1 + A, & \text{for } L_n = 0, \end{cases}$$

where Q_b is the number of customers in the system when a busy period starts. Note that the case of $Q_b = 1$ is for M/G/1 queue without vacations.

Lemma 2.1.1. The p.g.f. and the mean of Q_b are, respectively,

$$\begin{aligned} Q_b(z) &= H[v^*(\lambda)]z + \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} \{v^*(\lambda(1 - z)) - v^*(\lambda)\}, \\ E(Q_b) &= H[v^*(\lambda)] + \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} \lambda E(V). \end{aligned} \quad (2.1.5)$$

Proof: The event $\{Q_b = 1\}$ occurs if either of two mutually exclusive cases happens: (1) the busy period starts with a customer arriving at an idle server; or (2) the busy period starts with the ending of a vacation during which only one customer arrives. Hence, we have

$$P\{Q_b = 1\} = H[v^*(\lambda)] + \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} v_1,$$

where $v_j = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dV(t)$ is the probability that j customers arrive during a vacation time. For $j \geq 2$, $\{Q_b = j\}$ represents the case in which the busy period starts with the ending of a vacation during which j customers have arrived. Thus,

$$P\{Q_b = j\} = \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} v_j, \quad j \geq 2.$$

Taking the p.g.f. of the distribution of Q_b yields $Q_b(z)$ and computing $Q'_b(1)$ gives $E(Q_b)$. \square

Under the E-MAV policy, the transition probability matrix of the embedded chain of $\{L_n, n \geq 1\}$ becomes

$$\mathbf{P} = \begin{bmatrix} b_0 & b_1 & b_2 & b_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ & a_0 & a_1 & a_2 & \cdots \\ & & a_0 & a_1 & \cdots \\ & & & \vdots & \vdots \end{bmatrix}, \quad (2.1.6)$$

where

$$\begin{aligned} b_j &= P\{Q_b - 1 + A = j\} \\ &= H[v^*(\lambda)]a_j + \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} \sum_{i=1}^{j+1} v_i a_{j+1-i}, \quad j \geq 0. \end{aligned} \quad (2.1.7)$$

Similar to the classical M/G/1 queue, from (2.1.6) it can be proved that the embedded chain $\{L_n, n \geq 1\}$ is positive recurrent if and only if $\rho = \lambda\mu^{-1} < 1$. When $\rho < 1$, let L_v be the limiting (or stationary) random variable of L_n as $n \rightarrow \infty$, with the stationary distribution

$$\Pi = (\pi_0, \pi_1, \dots, \pi_n, \dots),$$

where $\pi_j = P\{L_v = j\} = \lim_{n \rightarrow \infty} P\{L_n = j\}$, for $j \geq 0$. We now give the stochastic decomposition property for the stationary queue length.

Theorem 2.1.1. For $\rho < 1$, L_v can be decomposed into the sum of two independent random variables,

$$L_v = L + L_d,$$

where L is the queue length of a classical M/G/1 queue without vacations with its p.g.f. given in (2.1.2). L_d is the additional queue length due to the vacation effect, with the p.g.f.

$$L_d(z) = \frac{1 - Q_b(z)}{E(Q_b)(1 - z)}, \quad (2.1.8)$$

where $Q_b(z)$ is given in Lemma 2.1.1.

Proof: Based on the equilibrium equation of $\Pi\mathbf{P} = \Pi$ and (2.1.6), we have

$$\pi_k = \pi_0 b_k + \sum_{j=1}^{k+1} \pi_j a_{k+1-j}, \quad k \geq 0. \quad (2.1.9)$$

From (2.1.7), we obtain the p.g.f of $\{b_k, k \geq 0\}$:

$$\sum_{k=0}^{\infty} z^k b_k = \frac{1}{z} B^*(\lambda(1 - z)) Q_b(z).$$

Multiplying both sides of (2.1.9) by z^k and summing over k gives

$$\begin{aligned}
 L_v(z) &= \sum_{k=0}^{\infty} z^k \pi_k \\
 &= \pi_0 \frac{1}{z} B^*(\lambda(1-z)) Q_b(z) + \sum_{k=0}^{\infty} z^k \sum_{j=1}^{k+1} \pi_j a_{k+1-j} \\
 &= \pi_0 \frac{1}{z} B^*(\lambda(1-z)) Q_b(z) + \frac{1}{z} B^*(\lambda(1-z)) [L_v(z) - \pi_0].
 \end{aligned}$$

Solving the equation above for $L_v(z)$, we get

$$L_v(z) = \frac{\pi_0 B^*(\lambda(1-z)) [1 - Q_b(z)]}{B^*(\lambda(1-z)) - z}. \quad (2.1.10)$$

Using the normalization condition and the L'Hopital rule, we have

$$\pi_0 = \frac{1 - \rho}{E(Q_b)},$$

and substituting it into (2.1.10) gives

$$\begin{aligned}
 L_v(z) &= \frac{(1 - \rho)(1 - z) B^*(\lambda(1 - z))}{B^*(\lambda(1 - z)) - z} \frac{1 - Q_b(z)}{E(Q_b)(1 - z)} \\
 &= L(z) L_d(z).
 \end{aligned}$$

This completes the proof. \square

Note that $L_d(z)$ in (2.1.8) is a p.g.f of a probability distribution. Define a distribution as

$$q_j = \frac{1}{E(Q_b)} \sum_{n=j+1}^{\infty} P\{Q_b = n\}, \quad j = 0, 1, \dots$$

Then the p.g.f. of $\{q_j, j \geq 0\}$ is

$$\begin{aligned}
 \bar{Q}_b(z) &= \sum_{j=0}^{\infty} q_j z^j \\
 &= \frac{1}{E(Q_b)} \sum_{j=0}^{\infty} z^j \sum_{n=j+1}^{\infty} P\{Q_b = n\} \\
 &= \frac{1}{E(Q_b)(1 - z)} \sum_{n=1}^{\infty} P\{Q_b = n\} (1 - z^n) \\
 &= \frac{1 - Q_b(z)}{E(Q_b)(1 - z)}.
 \end{aligned}$$

Based on Theorem 2.1.1, the following expected value formulas are obtained:

$$\begin{aligned} E(L_d) &= \frac{E(Q_b^2)}{2E(Q_b)}, \\ E(L_v) &= \rho + \frac{\lambda^2 b^{(2)}}{2(1-\rho)} + \frac{E(Q_b^2)}{2E(Q_b)}. \end{aligned} \quad (2.1.11)$$

Using $Q_b(z)$ in (2.1.5), we have

$$E(Q_b^2) = \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} \lambda^2 E(V^2).$$

For the stationary waiting time, there exists a similar stochastic decomposition property.

Theorem 2.1.2. For $\rho < 1$, the stationary waiting time, denoted by W_v , can be decomposed into the sum of the two independent random variables,

$$W_v = W + W_d,$$

where W is the waiting time of a classical M/G/1 queue without vacations, with its LST given in (2.1.2). W_d is the additional delay due to the vacation effect, with the LST

$$W_d^*(s) = \frac{H[v^*(\lambda)]}{E(Q_b)} + \frac{\lambda E(V)}{E(Q_b)} \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} \frac{1 - v^*(s)}{E(V)s}, \quad (2.1.12)$$

where $E(Q_b)$ is given in Lemma 2.1.1.

Proof: Based on the independent increment property of Poisson arrivals and the fact that the number of customers left behind by a departing customer is the same as the number of arrivals during this customer's time (waiting and service) in the system, we have

$$\begin{aligned} L_v(z) &= \sum_{k=0}^{\infty} z^k \int_0^{\infty} \int_0^{\infty} \frac{[\lambda(x+y)]^k}{k!} e^{-\lambda(x+y)} dW_v(x) dB(y) \\ &= \int_0^{\infty} \int_0^{\infty} e^{-\lambda(x+y)(1-z)} dW_v(x) dB(y) \\ &= W_v^*(\lambda(1-z)) B^*(\lambda(1-z)). \end{aligned}$$

Substituting $L_v(z)$ into the formula above gives

$$W_v^*(\lambda(1-z)) = \frac{(1-\rho)(1-z)}{B^*(\lambda(1-z)) - z} \frac{1 - Q_b(z)}{E(Q_b)(1-z)}. \quad (2.1.13)$$

Letting $\lambda(1 - z) = s$, we have

$$\begin{aligned} W_v^*(s) &= \frac{(1 - \rho)s}{s - \lambda(1 - B^*(s))} \frac{\lambda[1 - Q_b(1 - \frac{s}{\lambda})]}{E(Q_b)s} \\ &= W^*(s)W_d^*(s). \end{aligned}$$

Using (2.1.2), we find that the additional delay W_d has an LST of

$$W_d^*(s) = \frac{\lambda[1 - Q_b(1 - \frac{s}{\lambda})]}{E(Q_b)s}. \quad (2.1.14)$$

Substituting $Q_b(z)$ from (2.1.5) into (2.1.14) and simplifying yields (2.1.12). \square

Formula (2.1.12) indicates that the additional delay W_d is zero with probability of $p = H[v^*(\lambda)][E(Q_b)]^{-1}$ and is equal to the residual vacation time with probability of $1 - p$. It is easy to verify that the number of arrivals during W_d is the additional queue length due to the vacation effect, L_d . The means of the additional delay and the waiting time can be obtained as

$$\begin{aligned} E(W_d) &= \frac{\{1 - H[v^*(\lambda)]\}\lambda E(V^2)}{2(1 - v^*(\lambda))E(Q_b)}, \\ E(W_v) &= \frac{\lambda b^{(2)}}{2(1 - \rho)} + \frac{\{1 - H[v^*(\lambda)]\}\lambda E(V^2)}{2(1 - v^*(\lambda))E(Q_b)}. \end{aligned} \quad (2.1.15)$$

Let us now provide the busy-period analysis of the M/G/1 (E,MAV) model. Denote by D_v the busy period of the vacation system and by D the busy period of the classical M/G/1 system. Note that the only difference between D_v and D is the number of customers present in the system when the busy period starts. Due to the memoryless property of the exponential interarrival times, the busy period starting with k customers in the system is equal to the sum of k independent M/G/1 queue busy periods D . It follows immediately that

$$D_v^*(s) = Q_b[D^*(s)],$$

where $D^*(s)$ is the LST of D . Thus

$$E(D_v) = \frac{1}{\mu(1 - \rho)} E(Q_b).$$

Let J be the number of consecutive vacations taken by the server. Based on the MAV policy, we have

$$J = \min\{H, k : V^{(k-1)} < T < V^{(k)}\}.$$

It is easy to verify that

$$P\{J \geq 1\} = 1,$$

$$P\{J \geq j\} = P\{H \geq j\}P\{V^{(j-1)} \geq T\} = [v^*(\lambda)]^{j-1} \sum_{k=j}^{\infty} h_k, \quad j \geq 2.$$

Therefore, we have

$$\begin{aligned} \sum_{j=1}^{\infty} P\{J \geq j\} z^j &= \frac{z(1 - J(z))}{1 - z} \\ &= \sum_{j=1}^{\infty} z^j [v^*(\lambda)]^{j-1} \sum_{k=j}^{\infty} h_k = z \frac{1 - H[v^*(\lambda)z]}{1 - v^*(\lambda)z}. \end{aligned}$$

From this relation, we obtain

$$\begin{aligned} J(z) &= 1 - \frac{1 - z}{1 - v^*(\lambda)z} \{1 - H[v^*(\lambda)z]\}, \\ E(J) &= \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)}. \end{aligned}$$

Denote the total length of J consecutive vacations by V_G . Then

$$E(V_G) = E(J)E(V) = \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} E(V). \quad (2.1.16)$$

The idle period, denoted by I_v , occurs only when event A_I happens. Hence,

$$E(I_v) = H[v^*(\lambda)] \frac{1}{\lambda}. \quad (2.1.17)$$

Define the busy cycle B_c as the time period between two consecutive busy-period ending instants. Then we have

$$\begin{aligned} E(B_c) &= E(D_v) + E(V_G) + E(I_v) \\ &= \frac{1}{\mu(1 - \rho)} E(Q_b) + \frac{1 - H[v^*(\lambda)]}{1 - v^*(\lambda)} E(V) + H[v^*(\lambda)] \frac{1}{\lambda} \\ &= \frac{1}{\lambda(1 - \rho)} E(Q_b). \end{aligned} \quad (2.1.18)$$

Let p_b, p_v , and p_i be the probabilities of the server's being busy, on vacation, and idle, respectively. We then have

$$\begin{aligned} p_b &= \frac{E(D_v)}{E(B_c)} = \rho, \\ p_v &= \frac{E(V_G)}{E(B_c)} = \frac{1 - H[v^*(\lambda)]}{(1 - v^*(\lambda))E(Q_b)} \lambda(1 - \rho)E(V), \\ p_i &= \frac{E(I_v)}{E(B_c)} = \frac{1}{E(Q_b)}(1 - \rho)H[v^*(\lambda)]. \end{aligned} \quad (2.1.19)$$

2.2 Some Classical M/G/1 Vacation Models

In this section, we show that several classical vacation models are the special cases of the E-MAV model presented in the previous section.

2.2.1 Multiple Vacation Model

Consider an M/G/1 queue where the server follows an exhaustive-service and multiple vacation (E, MV) policy. This policy requires the server to keep serving customers until the system is empty and then to take vacations for as long as the system is empty. The server returns to serve the queue when there are some customers waiting in the system at a vacation completion instant. This type of system, denoted by M/G/1 (E, MV), has been extensively studied. The multiple vacation policy allows the server to maximize the use of idle time for supplementary work. However, the server does not have any idle time in such a system (where idle time means either serving the queue or being on vacation), if taking a vacation represents doing productive work. Obviously, this situation is the $H = \infty$ case for the E-MAV model.

If $H = \infty$, $H(z) = 0$. From (2.1.5), the busy period starts with Q_b customers in the system. The p.g.f. and the mean of Q_b are, respectively,

$$\begin{aligned} Q_b(z) &= \frac{v^*(\lambda(1 - z)) - v^*(\lambda)}{1 - v^*(\lambda)}, \\ E(Q_b) &= \frac{\lambda E(V)}{1 - v^*(\lambda)}. \end{aligned} \quad (2.2.1)$$

As a special case, it follows directly from Theorem 2.1.1 that the stochastic decomposition properties exist in the M/G/1 (E, MV).

Theorem 2.2.1. For $\rho < 1$, in an M/G/1 (E, MV) system, the queue length L_v can be decomposed into the sum of two independent random variables,

$$L_v = L + L_d,$$

where L is the queue length of a classical M/G/1 queue without vacations, with its p.g.f. given in (2.1.2). L_d is the additional queue length due to the vacation effect, with the p.g.f.

$$L_d(z) = \frac{1 - v^*(\lambda(1 - z))}{\lambda E(V)(1 - z)}. \quad (2.2.2)$$

Proof: Substituting $Q_b(z)$ and $E(Q_b)$ of (2.2.1) into (2.1.8) gives (2.2.2). \square

The means of L_d and L_v are, respectively,

$$\begin{aligned} E(L_d) &= \frac{\lambda E(V^2)}{2E(V)}, \\ E(L_v) &= \rho + \frac{\lambda^2 b^{(2)}}{2(1 - \rho)} + \frac{\lambda E(V^2)}{2E(V)}. \end{aligned} \quad (2.2.3)$$

Theorem 2.2.2. For $\rho < 1$, in an M/G/1 (E, MV) system, the stationary waiting time W_v can be decomposed into the sum of two independent random variables,

$$W_v = W + W_d,$$

where W is the waiting time of a classical M/G/1 queue without vacations, with its LST given in (2.1.2). W_d is the additional delay due to the vacation effect, with the LST

$$W_d^*(s) = \frac{1 - v^*(s)}{E(V)s}. \quad (2.2.4)$$

Proof: In (2.1.12), letting $H(z) \equiv 0$ and substituting $E(Q_b)$ into (2.2.1) gives (2.2.4). \square

The means of W_d and W_v are, respectively,

$$\begin{aligned} E(W_d) &= \frac{E(V^2)}{2E(V)}, \\ E(W_v) &= \frac{\lambda b^{(2)}}{2(1 - \rho)} + \frac{E(V^2)}{2E(V)}. \end{aligned} \quad (2.2.5)$$

Remark 2.2.1. It can be proved that there exist several closure properties of phase-type (PH) distributions for the vacation effect: (1) Note that the additional delay W_d is just the residual life of a vacation V . If the vacation is a PH distributed random variable with a representation

of (α, \mathbf{T}) , where \mathbf{T} is an $m \times m$ matrix and $\alpha_{m+1} = 0$, then W_d follows a PH distribution with a representation of (π, \mathbf{T}) , where π is the stationary probability vector of the infinitesimal generator $\mathbf{T}^* = \mathbf{T} + \mathbf{T}^0\alpha$. (2) The additional queue length L_d is the number of arrivals during W_d . For the PH vacations, L_d follows a discrete PH distribution with an irreducible representation (γ, \mathbf{U}) , where

$$\begin{aligned}\gamma &= \lambda\pi(\lambda\mathbf{I} - \mathbf{T})^{-1}, & \mathbf{U} &= \lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}, \\ \gamma_{m+1} &= \pi(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0, & \mathbf{U}^0 &= \lambda(\lambda\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0.\end{aligned}$$

For details about PH distribution, see Chapter 2 of Neuts (1981).

Substituting $Q_b(z)$ and $E(Q_b)$ of (2.2.1) into the results of the busy-period analysis in the E-MAV model, we obtain the corresponding formulas for the M/G/1 (E, MV) system:

$$\begin{aligned}E(D_v) &= \frac{\rho E(V)}{(1 - \rho)(1 - v^*(\lambda))}, \\ E(V_G) &= \frac{1}{1 - v^*(\lambda)} E(V), \\ E(B_c) &= \frac{E(V)}{(1 - \rho)(1 - v^*(\lambda))}, \\ p_v &= \frac{E(V_G)}{E(V_G) + E(D_v)} = 1 - \rho, \\ p_b &= \frac{E(D_v)}{E(V_G) + E(D_v)} = \rho.\end{aligned}$$

2.2.2 Single Vacation Model

Another important vacation model is the M/G/1 queue with exhaustive service and single vacation (E, SV). In this system, the server takes exactly one vacation immediately at the end of each busy period. If it finds no customer in the system upon returning from the vacation, it becomes idle until the next arrival. A customer arriving at an idle server does not wait, while a customer arriving during a server's vacation must wait until the end of the vacation. Note that the server now can be in one of three possible states, namely, serving the queue, taking a vacation, and staying idle. In practice, the single vacation after each busy period can be considered as a maintenance activity if the server represents a machine. Obviously, this situation is the $H \equiv 1$ case for the M/G/1 (E, MAV) model.

If $H \equiv 1$, then $H(z) = z$. From (2.1.5), we have

$$\begin{aligned} Q_b(z) &= v^*(\lambda(1-z)) - v^*(\lambda)(1-z), \\ E(Q_b) &= v^*(\lambda) + \lambda E(V). \end{aligned} \quad (2.2.6)$$

With (2.2.6), Theorem 2.1.1 becomes the following:

Theorem 2.2.3. For $\rho < 1$, in an M/G/1 (E, SV) system, the queue length L_v can be decomposed into the sum of two independent random variables,

$$L_v = L + L_d,$$

where L is the queue length of a classical M/G/1 queue without vacations with its p.g.f. given in (2.1.2). L_d is the additional queue length due to the vacation effect, with the p.g.f.

$$L_d(z) = \frac{1 + (1-z)v^*(\lambda) - v^*(\lambda(1-z))}{[v^*(\lambda) + \lambda E(V)](1-z)}. \quad (2.2.7)$$

Note that (2.2.7) can be rewritten as

$$L_d(z) = \frac{v^*(\lambda)}{v^*(\lambda) + \lambda E(V)} + \frac{\lambda E(V)}{v^*(\lambda) + \lambda E(V)} \frac{1 - v^*(\lambda(1-z))}{\lambda E(V)(1-z)}.$$

This expression indicates that L_d is zero with probability of $p = v^*(\lambda) \times [v^*(\lambda) + \lambda E(V)]^{-1}$ and is the number of arrivals to the system during the residual life of the vacation with probability of $1 - p$. Now, the means of L_d and L_v are, respectively,

$$\begin{aligned} E(L_d) &= \frac{\lambda^2 E(V^2)}{2[v^*(\lambda) + \lambda E(V)]}, \\ E(L_v) &= \rho + \frac{\lambda^2 b^{(2)}}{2(1-\rho)} + \frac{\lambda^2 E(V^2)}{2[v^*(\lambda) + \lambda E(V)]}. \end{aligned} \quad (2.2.8)$$

Similarly, from Theorem 2.1.2 for the M/G/1 (E, MAV), we get the following theorem.

Theorem 2.2.4. For $\rho < 1$, in an M/G/1 (E, SV) system, the stationary waiting time W_v can be decomposed into the sum of two independent random variables,

$$W_v = W + W_d,$$

where W is the waiting time of a classical M/G/1 queue without vacations, with its LST given in (2.1.2). W_d is the additional delay due to the vacation effect, with the LST

$$W_d^*(s) = \frac{sv^*(\lambda) + \lambda(1-v^*(s))}{[v^*(\lambda) + \lambda E(V)]s}. \quad (2.2.9)$$

Now, (2.2.9) can be rewritten as

$$W_d^*(s) = \frac{v^*(s)}{v^*(\lambda) + \lambda E(V)} + \frac{\lambda E(V)}{v^*(\lambda) + \lambda E(V)} \frac{1 - v^*(s)}{E(V)s}. \quad (2.2.10)$$

From (2.2.10), we see that W_d is zero with probability $p = v^*(\lambda)[v^*(\lambda) + \lambda E(V)]^{-1}$ and is the residual life of a vacation with probability $1 - p$. The means of W_d and W_v are given by

$$\begin{aligned} E(W_d) &= \frac{\lambda E(V^2)}{2[v^*(\lambda) + \lambda E(V)]}, \\ E(W_v) &= \frac{\lambda b^{(2)}}{2(1 - \rho)} + \frac{\lambda E(V^2)}{2[v^*(\lambda) + \lambda E(V)]}. \end{aligned} \quad (2.2.11)$$

Remark 2.2.2. Equation (2.2.10) shows that W_d is a mixture of zero and the residual life of a vacation. If the vacation is a PH-distributed random variable with a representation of (α, \mathbf{T}) , where \mathbf{T} is an $m \times m$ matrix and $\alpha_{m+1} = 0$, then W_d also follows a PH distribution with a representation of (γ, \mathbf{T}) , where

$$\gamma = \frac{\lambda E(V)}{v^*(\lambda) + \lambda E(V)} \pi, \quad \gamma_{m+1} = \frac{v^*(\lambda)}{v^*(\lambda) + \lambda E(V)}.$$

π is the stationary probability vector of the infinitesimal generator $\mathbf{T}^* = \mathbf{T} + \mathbf{T}^0 \alpha$. Note that the additional queue length L_d is the number of arrivals during W_d . For the PH distributed vacations, L_d follows a discrete PH distribution with an irreducible representation (η, \mathbf{U}) , where

$$\begin{aligned} \eta &= \frac{\lambda E(V)}{v^*(\lambda) + \lambda E(V)} \lambda \pi (\lambda \mathbf{I} - \mathbf{T})^{-1}, \quad \eta_{m+1} = \frac{1}{v^*(\lambda) + \lambda E(V)}, \\ \mathbf{U} &= \lambda (\lambda \mathbf{I} - \mathbf{T})^{-1}, \quad \mathbf{U}^0 = (\lambda \mathbf{I} - \mathbf{T})^{-1} \mathbf{T}^0. \end{aligned}$$

Using the results of the busy period analysis for the M/G/1 (E, MAV), we have

$$\begin{aligned}
E(D_v) &= \frac{1}{\mu(1-\rho)}[v^*(\lambda) + \lambda E(V)], \\
E(V_G) &= E(V), \\
E(I_v) &= \frac{v^*(\lambda)}{\lambda}, \\
p_b &= \rho, \\
p_v &= \frac{\lambda(1-\rho)E(V)}{v^*(\lambda) + \lambda E(V)}, \\
p_i &= \frac{v^*(\lambda)(1-\rho)}{v^*(\lambda) + \lambda E(V)}.
\end{aligned}$$

2.2.3 Setup Time Model

Consider an M/G/1 system where the first customer in each busy period requires a random setup time U . For example, in a production system, to reduce the operating cost the machine is shut down, and when the next job arrives, the facility is turned on again and must experience a warmup or setup period before processing the job. The setup time may also represent the switchover time from working on supplementary jobs to serving the arriving customer that initiates the busy period. We denote this system by M/G/1 (E, SU).

We first illustrate the relationship between M/G/1 (E, MV) and M/G/1 (E, SU), as was established by Levy and Kleinrock (1986). In a multiple vacation model, the waiting time of the first customer, denoted by R , in each busy period is the time interval from its arrival instant to the current vacation completion instant. Note that R is equivalent to the setup time triggered by the first arrival in a setup time model. The following preliminary result is useful.

Lemma 2.2.1. In an M/G/1 (E, MV) with FCFS service sequence, the LST and the mean of R are, respectively,

$$\begin{aligned}
R^*(s) &= \frac{\lambda[v^*(s) - v^*(\lambda)]}{[1 - v^*(\lambda)](\lambda - s)}, \\
E(R) &= \frac{E(V)}{1 - v^*(\lambda)} - \frac{1}{\lambda}.
\end{aligned} \tag{2.2.13}$$

Proof: Due to the memoryless property of exponential interarrival times, if a customer arrival occurs during the server's vacation, the interarrival time T can be counted from the instant of starting the vacation and $T < V$. Therefore, the distribution function of R can be written as

$$R(t) = P\{R \leq t\} = P\{V - T \leq t | V > T\}.$$

Taking the LST of the distribution of R , we get

$$\begin{aligned} R^*(s) &= E[e^{-s(V-T)} | V > T] \\ &= \frac{\int_0^\infty dV(x) \int_0^x e^{-s(x-y)} \lambda e^{-\lambda y} dy}{\int_0^\infty (1 - e^{-\lambda x}) dV(x)}. \end{aligned} \quad (2.2.14)$$

Note that the denominator of (2.2.14) is $1 - v^*(\lambda)$ and the numerator is

$$\begin{aligned} \int_0^\infty dV(x) \int_0^x e^{-s(x-y)} \lambda e^{-\lambda y} dy &= \lambda \int_0^\infty e^{-sx} \left[\int_0^x e^{-(\lambda-s)y} dy \right] dV(x) \\ &= \frac{\lambda}{\lambda - s} [v^*(s) - v^*(\lambda)]. \end{aligned}$$

Substituting these results into (2.2.14) gives (2.2.13). From (2.2.13) we have $E(R)$. \square

Letting U and $u^*(s)$ be the setup time and its LST in the M/G/1 (E, SU) and using the relation between M/G/1 (E, MV) and M/G/1 (E, SU), we have the stochastic decomposition property for the queue length.

Theorem 2.2.5. For $\rho < 1$, in an M/G/1 (E, SU) system, the stationary queue length L_v can be decomposed into the sum of two independent random variables,

$$L_v = L + L_d,$$

where L is the queue length of a classical M/G/1 queue without vacations with its p.g.f. given in (2.1.2). L_d is the additional queue length due to the setup time effect, with the p.g.f.

$$L_d(z) = \frac{1 - zu^*(\lambda(1-z))}{[1 + \lambda E(U)](1-z)}. \quad (2.2.15)$$

Proof: Consider a fictitious M/G/1 (E, MV) in which U is the waiting time of the first customer of a busy period, and let V be the vacation time of this system. From Lemma 2.2.1, U and V satisfy the relation

$$\begin{aligned} u^*(s) &= \frac{\lambda[v^*(s) - v^*(\lambda)]}{[1 - v^*(\lambda)](\lambda - s)}, \\ \lambda E(V) &= [1 + \lambda E(U)](1 - v^*(\lambda)). \end{aligned} \quad (2.2.16)$$

In the first equation of (2.2.16), replacing s with $\lambda(1-z)$, we have

$$v^*(\lambda(1-z)) = z(1 - v^*(\lambda))u^*(\lambda(1-z)) + v^*(\lambda).$$

Now, substituting $\lambda E(V)$ and $v^*(\lambda(1-z))$ into L_d of (2.2.2) gives (2.2.15). \square

Note that (2.2.15) can be rewritten as

$$L_d(z) = \frac{1}{1 + \lambda E(U)} + \frac{\lambda E(U)}{1 + \lambda E(U)} z \frac{1 - u^*(\lambda(1 - z))}{\lambda E(U)(1 - z)}.$$

This expression indicates that L_d is zero with probability of $p = [1 + \lambda E(V)]^{-1}$ and is the number of arrivals occurring during the residual setup time plus one customer that triggers the setup time with probability of $1 - p$. The means of L_d and L_v in the M/G/1 (E, SU) are, respectively,

$$\begin{aligned} E(L_d) &= \frac{2\lambda E(U) + \lambda^2 E(U^2)}{2(1 + \lambda E(U))}, \\ E(L_v) &= \rho + \frac{\lambda^2 b^{(2)}}{2(1 - \rho)} + \frac{2\lambda E(U) + \lambda^2 E(U^2)}{2(1 + \lambda E(U))}. \end{aligned} \quad (2.2.17)$$

Theorem 2.2.6. For $\rho < 1$, in an M/G/1 (E, SU) system, the stationary waiting time W_v can be decomposed into the sum of two independent random variables,

$$W_v = W + W_d,$$

where W is the waiting time of a classical M/G/1 queue without vacations, with its LST given in (2.1.2). W_d is the additional delay due to the vacation effect, with the LST

$$W_d^*(s) = \frac{\lambda - (\lambda - s)u^*(s)}{[1 + \lambda E(U)]s}. \quad (2.2.18)$$

Proof: Consider the same M/G/1 (E, MV) system used in the proof of Theorem 2.2.5. From (2.2.16), we get

$$\begin{aligned} v^*(s) &= \frac{1}{\lambda} u^*(s)[1 - v^*(\lambda)](\lambda - s) + v^*(\lambda), \\ E(V) &= \frac{1}{\lambda} [1 - v^*(\lambda)][1 + \lambda E(U)]. \end{aligned}$$

Substituting these results into $W_d^*(s)$ of (2.2.4) yields (2.2.18). \square

Now, (2.2.18) can be rewritten as

$$W_d^*(s) = \frac{1}{1 + \lambda E(U)} u^*(s) + \frac{\lambda E(U)}{1 + \lambda E(U)} \frac{1 - u^*(s)}{E(U)s}.$$

From this expression, we see that W_d is a complete setup time U with probability $p = [1 + \lambda E(U)]^{-1}$ and is the residual life of a setup time (or

residual setup time) with probability $1 - p$. This is because the expected number of customers in the system at the beginning of the busy period is $E(Q_b) = 1 + \lambda E(U)$. For these customers, the first customer that triggers the setup time must wait U ; other customers behind the first customer, including those arriving during the busy period, have to wait, on average, the additional time of residual setup time as compared with a classical M/G/1 system. The means of W_d and W_v are obtained, respectively, as

$$\begin{aligned} E(W_d) &= \frac{2E(U) + \lambda E(U^2)}{2[1 + \lambda E(U)]}, \\ E(W_v) &= \frac{\lambda b^{(2)}}{2(1 - \rho)} + \frac{2E(U) + \lambda E(U^2)}{2[1 + \lambda E(U)]}. \end{aligned} \quad (2.2.19)$$

2.3 M/G/1 Queue with Threshold Policy

In this section, we discuss the M/G/1 systems with threshold policy. In this type of system, the server becomes unavailable at the end of a busy period and resumes serving the queue instantly either when the queue length reaches a critical number N or at a vacation termination instant when the queue length equals or exceeds N . This type of policy is called a *threshold* or *N-policy*. Compared with the MAV model, the server's returning to queue service under the *N-policy* may be further delayed. We first treat the *N-policy* model without vacations.

2.3.1 N-Threshold Policy Model

In an M/G/1 queue with *N-policy* without vacations, at the end of a busy period, the server is shut down until the N th customer arrival instant, and then the server starts another busy period with $N \geq 1$ customers. Note that we can still consider the sum of N interarrival times as a *special server vacation*. This model is motivated by some practical systems where a significant setup cost occurs for each busy period and thus there is an economic benefit in reducing the frequency of setups. In fact, finding the cost-minimization *N-policy* is a typical optimal control problem in queueing theory.

Now, the busy period starts with exactly N customers in the system. Thus, the p.g.f and the expected value of Q_b are given, respectively, by

$$Q_b(z) = z^N, \quad E(Q_b) = N.$$

The embedded Markov chain at customer departure instants $\{L_n, n \geq 1\}$ has the probability transition matrix

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & \cdots & a_0 & a_1 & \cdots \\ a_0 & a_1 & \cdots & a_{N-1} & a_N & \cdots \\ & a_0 & \cdots & a_{N-2} & a_{N-1} & \cdots \\ & & \cdots & \cdots & \cdots & \cdots \\ & & & \vdots & \vdots & \vdots \end{bmatrix}. \quad (2.3.1)$$

With the classical method, it can be proved that $\{L_n, n \geq 1\}$ is positive recurrent if and only if $\rho = \lambda\mu^{-1} < 1$.

Theorem 2.3.1. For $\rho < 1$, in an M/G/1 system with N -policy, the stationary queue length L_v can be decomposed into the sum of two independent random variables,

$$L_v = L + L_d,$$

where L is the queue length of a classical M/G/1 queue without vacations with its p.g.f. given in (2.1.2). L_d is the additional queue length due to the effect of N -policy, with the p.g.f.

$$L_d(z) = \frac{1 - z^N}{N(1 - z)}. \quad (2.3.2)$$

Proof: Using the equilibrium equation $\Pi\mathbf{P} = \Pi$ and (2.3.1), we have

$$\begin{aligned} \pi_k &= \sum_{j=1}^{k+1} \pi_j a_{k+1-j}, & 0 \leq k \leq N-2, \\ \pi_k &= \pi_0 a_{k-N+1} + \sum_{j=1}^{k+1} \pi_j a_{k+1-j}, & k \geq N-1. \end{aligned}$$

The p.g.f. of $\{\pi_k, k \geq 0\}$ is obtained as follows.

$$\begin{aligned} L_v(z) &= \sum_{k=0}^{N-2} z^k \sum_{j=1}^{k+1} \pi_j a_{k+1-j} + \sum_{k=N-1}^{\infty} z^k \left[\pi_0 a_{k+1-N} + \sum_{j=1}^{k+1} \pi_j a_{k+1-j} \right] \\ &= \sum_{k=0}^{\infty} z^k \sum_{j=1}^{k+1} \pi_j a_{k+1-j} + \pi_0 \sum_{k=N-1}^{\infty} z^k a_{k-N+1} \\ &= \sum_{j=1}^{\infty} \pi_j z^{j-1} \sum_{k=j-1}^{\infty} z^{k-j+1} a_{k-j+1} + \pi_0 z^{N-1} \sum_{k=N-1}^{\infty} z^{k-N+1} a_{k-N+1} \\ &= \frac{1}{z} [L_v(z) - \pi_0] B^*(\lambda(1-z)) + \pi_0 z^{N-1} B^*(\lambda(1-z)). \end{aligned}$$

Solving this equation for $L_v(z)$, we get

$$L_v(z) = \frac{\pi_0(1 - z^N)B^*(\lambda(1 - z))}{B^*(\lambda(1 - z)) - z}. \quad (2.3.3)$$

Using the normalization condition $L_v(1) = 1$, we find $\pi_0 = (1 - \rho)N^{-1}$. Substituting π_0 into (2.3.3) gives

$$\begin{aligned} L_v(z) &= \frac{(1 - \rho)(1 - z)B^*(\lambda(1 - z))}{B^*(\lambda(1 - z)) - z} \frac{1 - z^N}{N(1 - z)} \\ &= L(z)L_d(z). \end{aligned}$$

□

From this stochastic decomposition property, the expected values of L_d and L_v are given, respectively, by

$$\begin{aligned} E(L_d) &= \frac{N - 1}{2}, \\ E(L_v) &= \rho + \frac{\lambda^2 b^{(2)}}{2(1 - \rho)} + \frac{N - 1}{2}. \end{aligned} \quad (2.3.4)$$

The LST and the expected value of the busy period D_v can be obtained easily as follows:

$$D_v^*(s) = [D^*(s)]^N; \quad E(D_v) = \frac{N}{\mu(1 - \rho)}.$$

The idle period follows an Erlang distribution, with the respective LST and expected value given by

$$v^*(s) = \left(\frac{\lambda}{\lambda + s} \right)^N; \quad E(V) = \frac{N}{\lambda}.$$

The busy cycle B_c has the expected value

$$E(B_c) = E(V) + E(D_v) = \frac{N}{\lambda(1 - \rho)}.$$

Using $E(B_c)$, it is easy to show that the proportion of busy or idle time is $p_b = \rho$ or $p_v = 1 - \rho$. Note that the waiting time for a customer arriving during a server's idle period depends on the interarrival times of customers arriving later. Let A_v and A_b represent the arrival of a customer during an idle period and during a busy period, respectively. Due to the property of complete randomness of the exponential distribution, for any particular one of these N arrivals during the idle period,

the probability that the customer is the k th arrival is N^{-1} . The waiting time of the first of these N arrivals is the sum of $N - 1$ interarrival times. The waiting time of the second arrival is the sum of $N - 2$ interarrival times plus one service time, and so on. Conditioning on event A_v , we have the LST of the waiting time:

$$\begin{aligned} W_v^*(s|A_v) &= \frac{1}{N} \sum_{j=0}^{N-1} \left(\frac{\lambda}{\lambda + s} \right)^{N-1-j} [B^*(s)]^j \\ &= \frac{1}{N} \left(\frac{\lambda}{\lambda + s} \right)^{N-1} \frac{\lambda^N - (\lambda + s)^N [B^*(s)]^N}{\lambda - (\lambda + s)B^*(s)}. \end{aligned} \quad (2.3.5)$$

Let us now prove a conditional stochastic decomposition property for the waiting time. In fact for the multiserver vacation models to be discussed in chapters 5 and 6, we can establish only the conditional decomposition properties at the time of writing this book. We use the method of the *delayed busy period* developed by Conway (1960), Nair and Neuts (1969), and Kleinrock (1975) to give the following result.

Theorem 2.3.2 For $\rho < 1$, the conditional waiting time for customers arriving in a busy period, $(W_v|A_b)$, can be decomposed into the sum of two independent random variables,

$$(W_v|A_b) = W + (W_d|A_b),$$

where W is the waiting time of a classical M/G/1 queue without vacations, with its LST given in (2.1.2). $(W_d|A_v)$ is the conditional additional delay due to the effect of N -policy, with the LST

$$W_d^*(s|A_b) = \frac{\mu \{1 - [B^*(s)]^N\}}{N s}. \quad (2.3.6)$$

Proof: Let X_0 be the sum of the first N customer service times, called the *initial delay* or *initial phase* of a busy period. According to the FCFS sequence, let X_1 be the sum of the service times of all customers arriving during X_0 , called the *first phase* of the busy period. In general, the sum of the service times of the customers arriving during the $(m - 1)$ phase X_{m-1} is called the *mth phase* and is denoted by X_m . Thus we have the busy period

$$D_v = \sum_{m=0}^{\infty} X_m.$$

Let $D_m(t)$ and $d_m^*(s)$ be the distribution function and the LST of X_m , respectively. Then $d_0^*(s) = [B^*(s)]^N$. If there are j arrivals during X_{m-1} ,

X_m is the sum of j service times. Therefore, we get

$$\begin{aligned} d_m^*(s) &= \sum_{j=0}^{\infty} \int_0^{\infty} [B^*(s)]^j \frac{(\lambda t)^j}{j!} e^{-\lambda t} dD_{m-1}(t) \\ &= \int_0^{\infty} e^{-\lambda(1-B^*(s))t} dD_{m-1}(t) \\ &= d_{m-1}^*(\lambda(1-B^*(s))), \quad m \geq 1. \end{aligned} \quad (2.3.7)$$

If a customer arrives at an instant of y time units before the end of the m th phase of length X_m , then the waiting time of this customer is y plus the sum of the service times of all customers arriving in the time interval $X_m - y$. Thus the LST of the conditional waiting time is

$$\begin{aligned} E\{e^{-sW_m} | X_m = t, \text{ arriving at an instant of } y \text{ time units before the end of } X_m\} \\ = \sum_{n=0}^{\infty} e^{-sy} [B^*(s)]^n \frac{[\lambda(t-y)]^n}{n!} e^{-\lambda(t-y)} \\ = \exp\{-[sy + \lambda(t-y)(1-B^*(s))]\}. \end{aligned}$$

Due to Poisson arrivals, given that the customer arrives in $[0, t]$, the arrival instant is uniformly distributed over $[0, t]$ with density of $t^{-1}dy$. Conditioning on y , we have

$$\begin{aligned} E\{e^{-sW_m} | X_m = t\} &= \int_0^t \exp\{-[sy + \lambda(t-y)(1-B^*(s))]\} \frac{1}{t} dy \\ &= \frac{1}{t} e^{-\lambda(1-B^*(s))t} \int_0^t \exp\{-[s - \lambda(1-B^*(s))]y\} dy \\ &= \frac{e^{-\lambda(1-B^*(s))t} - e^{-st}}{t[s - \lambda(1-B^*(s))]} \end{aligned} \quad (2.3.8)$$

Given that a customer has arrived during X_m , the conditional probability that the arrival occurs in $(t, t + dt)$ is

$$\frac{t}{E(X_m)} dD_m(t).$$

Unconditioning (2.3.8), we have

$$\begin{aligned} W_m^*(s) &= \int_0^{\infty} E\{e^{-sW_m} | X_m = t\} \frac{t}{E(X_m)} dD_m(t) \\ &= \frac{1}{E(X_m)[s - \lambda(1-B^*(s))]} \int_0^{\infty} [e^{-\lambda(1-B^*(s))t} - e^{-st}] dD_m(t) \\ &= \frac{d_{m+1}^*(s) - d_m^*(s)}{E(X_m)[s - \lambda(1-B^*(s))]} \end{aligned} \quad (2.3.9)$$

Given that a customer arrives in the busy period D_v , the probability that this arrival occurs in the m th phase is $E(X_m)[E(D_v)]^{-1}$. Moreover, for $\rho < 1$, with probability of 1, D_v ends in a finite time interval. That is,

$$\lim_{m \rightarrow \infty} X_m = 0, \quad \text{a.s.}; \quad \lim_{m \rightarrow \infty} d_m^*(s) = 1.$$

Now from (2.3.9), we have

$$\begin{aligned} W_v^*(s|A_b) &= \sum_{m=0}^{\infty} \frac{E(X_m)}{E(D_v)} W_m^*(s) \\ &= \frac{\sum_{m=0}^{\infty} [d_{m+1}^*(s) - d_m^*(s)]}{E(D_v)[s - \lambda(1 - B^*(s))]} \\ &= \frac{1 - d_0^*(s)}{E(D_v)[s - \lambda(1 - B^*(s))]} \end{aligned}$$

Substituting $E(D_v) = N[\mu(1 - \rho)]^{-1}$ and $d_0^*(s) = [B^*(s)]^N$ into the equation above, we get

$$W_v^*(s|A_b) = \frac{(1 - \rho)s}{s - \lambda(1 - B^*(s))} \frac{\mu\{1 - [B^*(s)]^N\}}{Ns}.$$

□

(2.3.6) indicates that the additional delay for the customers arriving during a busy period is the residual life of the sum of N service times, and its expected value is given by

$$E(W_v|A_b) = \frac{\lambda b^{(2)}}{2(1 - \rho)} + \frac{N - 1}{2\mu}.$$

Furthermore, from (2.3.4), (2.3.5), and (2.3.6), we get the LST of the unconditional waiting time distribution as

$$W_v^*(s) = (1 - \rho)W^*(s|A_v) + \rho W^*(s|A_b).$$

2.3.2 Other Threshold Policy Models

Due to different practical applications, several related threshold-type policies have been studied in the past. Heyman (1977) presented a T -policy M/G/1 model. In such a model, the server is turned off for a fixed time interval T at the end of each busy period and then either resumes the queue service or stays idle depending on whether or not there are waiting customers at the end of T . Obviously, the T -policy model is equivalent to the M/G/1 (E,SV) with a constant vacation. In section 2.2.2, letting

$$E(V) = T, \quad v^*(\lambda) = e^{-\lambda T}, \quad v^*(s) = e^{-sT},$$

we obtain the results of the T -policy model.

Another variant of the threshold policy model is the D -policy M/G/1 model, which was studied by Balachandran and Tijms (1975). With the D -policy, after a busy period, the server will not start another busy period until the cumulative work (or the total service times of waiting customers) exceeds a critical number D . The detailed analysis of the D -policy model is more complex and can be found in Balachandran and Tijms (1975).

As an extension of the N -policy, Yadin and Naor (1963) investigated the M/G/1 queue with N -policy and setup and closedown times. In this system, the server needs a random closedown delay time C , with the LST $c^*(s)$. If a customer arrives during C , the customer is served immediately or a new busy period starts at the arrival instant; if no customer arrives during C , the server is shut down and will not be turned on until the number of waiting customers reaches N . When the server is turned on, it must experience a random setup time V , with the LST $v^*(s)$. Again, letting Q_b be the number of customers in the system at the beginning of a busy period, we have

$$Q_b(z) = [1 - c^*(\lambda)]z + c^*(\lambda)z^N v^*(\lambda(1 - z)),$$

$$E(Q_b) = 1 + c^*(\lambda)[N - 1 + \lambda E(V)].$$

Similarly, we can prove the stochastic decomposition property on the queue length. That is, $L_v = L + L_d$, where the p.g.f. and the expected value of L_d are given, respectively, by

$$L_d(z) = \frac{1 - (1 - c^*(\lambda))z - c^*(\lambda)z^N v^*(\lambda(1 - z))}{\{1 + c^*(\lambda)[N - 1 + \lambda E(V)](1 - z)\}},$$

$$E(L_d) = \frac{c^*(\lambda)[N(N - 1) + \lambda N E(V) + \lambda^2 E(V^2)]}{2\{1 + c^*(\lambda)[N - 1 + \lambda E(V)]\}}.$$

Because the waiting time of a customer is not independent of the inter-arrival times after its arrival, the analysis of the waiting time is fairly complex. Using a similar approach to that of the M/G/1 with N -policy model, we can establish the conditional decomposition property on the waiting time. The LST of the waiting time is

$$W_v^*(s) = \frac{(1 - c^*(\lambda))(1 - \rho)s + c^*(\lambda)(1 - \rho)[1 - v^*(s)(B^*(s))^N]}{\{1 + c^*(\lambda)[N - 1 + \lambda E(V)]\}[s - \lambda(1 - B^*(s))]}$$

$$+ \frac{c^*(\lambda)(1 - \rho)v^*(s) \left[\left(\frac{\lambda}{s + \lambda} \right)^N - [B^*(s)]^N \right]}{\{1 + c^*(\lambda)[N - 1 + \lambda E(V)]\} \left[\frac{\lambda}{s + \lambda} - B^*(s) \right]}.$$

The combination of N -policy and multiple vacations is also a well-known vacation policy. Under this policy, at the end of a busy period, the server takes i.i.d. random vacations consecutively until the number of customers in the system at a vacation completion instant is at least N , and then it resumes serving the queue. Consider a set of Markov points comprising the vacation completion and busy period ending instants. Let q_k be the joint probability that a randomly selected Markov point is the vacation completion instant and that the number of customers in the system at that instant is k . Let h_0 be the probability that a randomly selected Markov point is the busy-period ending instant. We then have

$$q_k = h_0 v_k + \sum_{j=0}^{\min(k, N-1)} q_j v_{k-j}, \quad k \geq 0,$$

$$1 = h_0 + \sum_{k=1}^{\infty} q_k,$$

where

$$v_k = \int_0^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} dV(t).$$

Defining the p.g.f. as

$$q(z) = \sum_{k=0}^{\infty} q_k z^k$$

and using the transition relation, we have

$$\begin{aligned} q(z) &= h_0 \sum_{k=0}^{\infty} v_k z^k + \sum_{k=0}^{N-1} z^k \sum_{j=0}^k q_j v_{k-j} + \sum_{k=N}^{\infty} z^k \sum_{j=0}^{N-1} q_j v_{k-j} \\ &= h_0 v^*(\lambda(1-z)) + \sum_{j=0}^{N-1} q_j z^j \sum_{k=j}^{\infty} z^{k-j} v_{k-j} \\ &= \left[h_0 + \sum_{j=0}^{N-1} q_j z^j \right] v^*(\lambda(1-z)). \end{aligned}$$

Furthermore, let

$$q_N(z) = \frac{1}{h_0} \sum_{k=0}^{N-1} q_k z^k.$$

Thus $q(z)$ can be rewritten as

$$q(z) = h_0 [1 + q_N(z)] v^*(\lambda(1-z)). \quad (2.3.10)$$

The coefficients of $q_N(z)$, q_0, q_1, \dots, q_{N-1} , can be determined by solving a set of equations

$$q_k = h_0 v_k + \sum_{j=0}^k q_j v_{k-j}, \quad k = 0, 1, \dots, N-1.$$

Note that the busy period does not start at a vacation completion instant when the number of customers in the system is less than N . Therefore, the p.g.f. of Q_b is given by

$$Q_b(z) = \frac{\sum_{k=N}^{\infty} q_k z^k}{\sum_{k=N}^{\infty} q_k} = \frac{q(z) - h_0 q_N(z)}{q(1) - h_0 q_N(1)}. \quad (2.3.11)$$

From (2.3.10), we have $q(1) = h_0(1 + q_N(1))$, and hence

$$q(1) - h_0 q_N(1) = h_0.$$

Now substituting $q(z)$ of (2.3.10) into (2.3.11) gives

$$Q_b(z) = v^*(\lambda(1-z)) - q_N(z) [1 - v^*(\lambda(1-z))].$$

Based on the method used before and $Q_b(z)$, we can obtain the stationary distribution of the queue length and the corresponding decomposition property. However, like the N -policy M/G/1 system, the residual life of the vacation may depend on the arrival process after a customer's arrival; the waiting time of this customer cannot be determined by using the classical relation between L_v and W_v . Therefore, we should use the same approach as in the N -policy M/G/1 model to obtain the stationary waiting time.

2.4 Discrete-Time Geo/G/1 Queue with Vacations

In this section, we discuss some discrete-time vacation models. In a discrete-time queueing system, the time axis is divided into fixed-length intervals called *slots*, and customer arrivals and service completions occur only at discrete time instants, which can be either the starts or the ends of the slots. In computer and telecommunication systems, the basic time unit is a fixed interval called a *packet* or *ATM cell* of transmission time. Therefore, the discrete-time models in this section are more appropriate for studying computer and telecommunication systems. The early work in this area was presented by Meisling (1958), and the discrete-time queueing models, including vacation models, have been developed as continuous counterparts (see Hunter (1983) and Takagi (1993a)).

2.4.1 Classical Geo/G/1 Queue

We first describe the classical discrete-time Geo/G/1 queueing system. In this system, we assume that customer arrivals can only occur at discrete time instants $t = n^-, n = 0, 1, 2, \dots$. The service starting and ending times can only occur at discrete time instants $t = n^+, n = 1, 2, \dots$. The model is called a *late arrival system*. The interarrival times are i.i.d. discrete random variables, denoted by T , with a geometric distribution of parameter p . That is,

$$P\{T = j\} = p\bar{p}^{j-1}, \quad j = 1, 2, \dots,$$

where $\bar{p} = 1 - p$. Thus the number of arrivals in interval $[0, n]$, C_n , follows a Binomial distribution

$$P\{C_n = j\} = \binom{n}{j} p^j \bar{p}^{n-j}, \quad j = 0, 1, \dots, n.$$

The service times are also i.i.d. discrete random variables, denoted by S , with a general distribution. We have

$$P\{S = j\} = g_j, \quad j \geq 1; \quad G(z) = \sum_{j=1}^{\infty} z^j g_j.$$

We assume that the interarrival times and the service times are independent and that the service order is FCFS.

Let A be the number of customers arriving during a service time. We have

$$\begin{aligned} k_j &= P(A = j) = \sum_{k=j}^{\infty} P\{S = k\} \binom{k}{j} p^j \bar{p}^{k-j} \\ &= \sum_{k=j}^{\infty} g_k \binom{k}{j} p^j \bar{p}^{k-j}, \quad j \geq 0. \end{aligned}$$

The p.g.f. and the expected value of A are given, respectively, by

$$\begin{aligned} A(z) &= \sum_{j=0}^{\infty} z^j k_j = \sum_{j=0}^{\infty} z^j \sum_{k=j}^{\infty} g_k \binom{k}{j} p^j \bar{p}^{k-j} \\ &= \sum_{k=0}^{\infty} g_k \sum_{j=0}^k \binom{k}{j} (pz)^j \bar{p}^{k-j} \\ &= G[1 - p(1 - z)]. \end{aligned} \tag{2.4.1}$$

$$E(A) = pE(S) = \rho.$$

Let L_n be the number of customers in the system at the n th customer departure instant. Thus $\{L_n, n \geq 1\}$ is a Markov chain, with the transition probability matrix

$$\mathbf{P} = \begin{bmatrix} k_0 & k_1 & k_2 & k_3 & \cdots \\ k_0 & k_1 & k_2 & k_3 & \cdots \\ & k_0 & k_1 & k_2 & \cdots \\ & & k_0 & k_1 & \cdots \\ & & & \vdots & \ddots \end{bmatrix}.$$

It can be proved that $\{L_n, n \geq 1\}$ is positive recurrent if and only if $\rho < 1$. For $\rho < 1$, let L be the stationary queue length or the limiting random variable of $\{L_n, n \geq 1\}$, and let W be the stationary waiting time. Now L and W are nonnegative integer random variables. Like the Pollaczek-Khinchin formulas for the continuous-time M/G/1 system, we have

$$\begin{aligned} L(z) &= \frac{(1-\rho)(1-z)G[1-p(1-z)]}{G[1-p(1-z)]-z}, \\ W(z) &= \frac{(1-\rho)(1-z)}{(1-z)-p(1-G(z))}, \\ E(L) &= \rho + \frac{p^2}{2(1-\rho)}E[S(S-1)], \\ E(W) &= \frac{p}{2(1-\rho)}E[S(S-1)]. \end{aligned} \tag{2.4.2}$$

The busy period of the Geo/G/1 queue is also a positive integer random variable, with the p.g.f. $D(z)$ satisfying the functional equation

$$D(z) = G[zD(1-p(1-z))],$$

and the expected value

$$E(D) = \frac{E(S)}{1-\rho}. \tag{2.4.3}$$

2.4.2 Geo/G/1 Queue with MAVs

Like the continuous-time M/G/1 (E, MAV) model, we introduce the multiple adaptive vacation policy into the Geo/G/1 system. For the Geo/G/1 (E, MAV) model, the server attempts to consecutively take a maximum number of H vacations. H is a random variable, with the respective distribution and p.g.f.

$$P\{H = j\} = h_j, \quad j \geq 1; \quad H(z) = \sum_{j=1}^{\infty} z^j h_j.$$

The vacations are i.i.d. discrete random variables, with the respective distribution and p.g.f.

$$P\{V = j\} = v_j, \quad j \geq 1; \quad v(z) = \sum_{j=1}^{\infty} z^j v_j.$$

If no customer arrives during H consecutive vacations, the server becomes idle and is ready to serve the next arrival. If the first customer arrives during the k th vacation, where $1 \leq k \leq H$, then the server starts serving the customer (or starts a busy period) at the k th vacation completion instant. Let J be the actual number of vacations consecutively taken by the server between the two busy periods. Obviously, J depends on H and the arrival process. Let T be the interarrival time and $V^{(k)}$ the k th convolution of vacation time V . Then we have

$$J = \min\{H, k : V^{(k-1)} < T \leq V^{(k)}\}.$$

Define the events A_I and A_v as in section 2.1.2. We get

$$\begin{aligned} P\{A_I\} &= \sum_{i=1}^{\infty} P\{H = i\} \sum_{k=i}^{\infty} P\{V^{(i)} = k\} \bar{p}^k \\ &= \sum_{i=1}^{\infty} h_i [v(\bar{p})]^i = H[v(\bar{p})], \\ P(A_v) &= 1 - H[v(\bar{p})]. \end{aligned}$$

Let L_n be the number of customers in the system at the n th departure instant. $\{L_n, n \geq 1\}$ is a Markov chain. We have

$$L_{n+1} = \begin{cases} L_n - 1 + A, & L_n \geq 1; \\ Q_b - 1 + A, & L_n = 0, \end{cases}$$

where A is the number of arrivals during a service time, and its p.g.f. and expected value are as in (2.4.1). Q_b , as defined earlier, is the number of customers in the system at the beginning of a busy period. The case $Q_b \equiv 1$ is the classical Geo/G/1 queue. Let c_j be the probability that exactly j customers arrive during a vacation V . It follows that

$$c_j = \sum_{k=j}^{\infty} v_k \binom{k}{j} p^j \bar{p}^{k-j}, \quad j = 0, 1, \dots,$$

with respective p.g.f. and expected value

$$C(z) = v(1 - p(1 - z)), \quad E(C) = \sum_{j=0}^{\infty} j c_j = pE(V).$$

To establish the stochastic decomposition theorem, we first present the following lemma.

Lemma 2.4.1. The p.g.f. and the expected value of Q_b are given, respectively, by

$$\begin{aligned} Q_b(z) &= H[v(\bar{p})]z + \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})}[v(1 - p(1 - z)) - v(\bar{p})], \\ E(Q_b) &= H[v(\bar{p})] + \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})}pE(V). \end{aligned} \quad (2.4.4)$$

Proof: For $Q_b = 1$, we must have either an arrival occurring during an idle period or only one arrival occurring during a vacation time. Then it follows that

$$P\{Q_b = 1\} = H[v(\bar{p})] + \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})}v_1.$$

For $j \geq 2$, we have

$$p\{Q_b = j\} = \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})}v_j.$$

Multiplying $P\{Q_b = j\}$ by z^j and taking the sum of these products from $j = 1$ to ∞ gives the $Q_b(z)$ in (2.4.4). Computing $Q'_b(1)$ yields $E(Q_b)$. \square

The probability transition matrix of $\{L_n, n \geq 1\}$ is

$$\mathbf{P} = \begin{bmatrix} b_0 & b_1 & b_2 & b_3 & \cdots \\ k_0 & k_1 & k_2 & k_3 & \cdots \\ & k_0 & k_1 & k_2 & \cdots \\ & & k_0 & k_1 & \cdots \\ & & & \vdots & \ddots \end{bmatrix}, \quad (2.4.5)$$

where k_j is defined as before and

$$\begin{aligned} b_j &= P\{Q_b - 1 + A = j\} \\ &= H[v(\bar{p})]k_j + \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})} \sum_{i=1}^{j+1} c_i k_{j+1-i}, \quad j \geq 0. \end{aligned} \quad (2.4.6)$$

Theorem 2.4.1. For $\rho < 1$, in a Geo/G/1 (E, MAV) system, the stationary queue length L_v can be decomposed into the sum of two independent random variables,

$$L_v = L + L_d,$$

where L is the queue length of a classical Geo/G/1 queue without vacations, with its p.g.f. as given in (2.4.2). L_d is the additional queue length due to the vacation effect, with the p.g.f.

$$L_d(z) = \frac{1 - Q_b(z)}{E(Q_b)(1 - z)}, \quad (2.4.7)$$

where $Q_b(z)$ is given in Lemma 2.4.1.

Proof: It follows from the equilibrium equation $\mathbf{\Pi P} = \mathbf{\Pi}$ and (2.4.5) that

$$\pi_j = \pi_0 b_j + \sum_{i=1}^{j+1} \pi_i k_{j+1-i}, \quad j \geq 0. \quad (2.4.8)$$

Using (2.4.6), we can compute the p.g.f. of $\{b_j, j \geq 0\}$ as

$$\sum_{j=0}^{\infty} z^j b_j = \frac{1}{z} G(1 - p(1 - z)) Q_b(z).$$

Multiplying both sides of (2.4.8) by z^j and taking the sum over j , we have

$$\begin{aligned} L_v(z) &= \pi_0 \sum_{j=0}^{\infty} z^j b_j + \sum_{j=0}^{\infty} z^j \sum_{i=1}^{j+1} \pi_i k_{j+1-i} \\ &= \frac{\pi_0}{z} G(1 - p(1 - z)) Q_b(z) + \sum_{i=1}^{\infty} \pi_i \sum_{j=i-1}^{\infty} z^j k_{j+1-i} \\ &= \frac{\pi_0}{z} G(1 - p(1 - z)) Q_b(z) + \frac{1}{z} G(1 - p(1 - z)) [L_v(z) - \pi_0]. \end{aligned}$$

Solving this equation for $L_v(z)$ gives

$$L_v(z) = \frac{\pi_0 G(1 - p(1 - z)) [1 - Q_b(z)]}{G[1 - p(1 - z)] - z}.$$

Using the normalization condition $L_v(1) = 1$, we can determine $\pi_0 = (1 - \rho)[E(Q_b)]^{-1}$. Substituting π_0 into $L_v(z)$, we get

$$\begin{aligned} L_v(z) &= \frac{(1 - \rho)(1 - z)G[1 - p(1 - z)]}{G[1 - p(1 - z)] - z} \frac{1 - Q_b(z)}{E(Q_b)(1 - z)} \\ &= L(z)L_d(z). \end{aligned}$$

□

From the stochastic decomposition theorem, we can obtain the expected values as follows:

$$\begin{aligned} E(L_d) &= \frac{p^2 E(Q_b^2)}{2E(Q_b)} = \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})} \frac{p^2 E(V(V-1))}{2E(Q_b)}, \\ E(L_v) &= \rho + \frac{p^2}{2(1-\rho)} E(S(S-1)) + \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})} \frac{p^2}{2E(Q_b)} E(V(V-1)). \end{aligned} \quad (2.4.9)$$

Let D_k be the system time of the k th customer, which extends from its arrival instant to its departure instant. We have

$$D_{k+1} = \begin{cases} D_k - T + S, & D_k - T \geq 0; \\ \Omega + S, & D_k - T < 0, \end{cases} \quad (2.4.10)$$

where T and S are the interarrival time and service time, respectively, and Ω is the waiting time of the first customer of a busy period. Similarly to Lemma 2.2.1, we have the following:

Lemma 2.4.2. The p.g.f. and the expected value of Ω are given, respectively, by

$$\begin{aligned} \Omega(z) &= H[v(\bar{p})] + \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})} \frac{p[v(z) - v(\bar{p})]}{z - \bar{p}}, \\ E(\Omega) &= \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})} \frac{p^2 E(V) - p(1 - v(\bar{p}))}{p^2}. \end{aligned} \quad (2.4.11)$$

Proof: For $j \geq 0$, we have

$$P\{\Omega = j\} = P\{A_I\}\delta_{j0} + P\{A_v\}P\{\Omega = j|A_v\},$$

where δ_{j0} is the Kronecker symbol. If A_v occurs and the vacation during which customers arrive is the first vacation, then $\Omega = (V_1 - T|V_1 \geq T)$; otherwise, due to the memoryless property of Poisson process, the conditional probability of Ω , given that event A_v occurs, can be computed from the start of the second vacation. That is,

$$\begin{aligned} P\{\Omega = j|A_v\} &= P\{V_1 \geq T\}P\{V_1 - T = j|V_1 \geq T\} \\ &\quad + P\{V_1 < T\}P\{\Omega = j|A_v\}. \end{aligned} \quad (2.4.12)$$

Note that

$$P\{V_1 - T = j|V_1 \geq T\} = \frac{1}{1 - v(\bar{p})} \sum_{k=j+1}^{\infty} v_k p \bar{p}^{k-j-1}, \quad j \geq 0.$$

Taking the p.g.f. of the conditional probability distribution above, we have

$$E\{z^{V_1-T} | V_1 \geq T\} = \frac{p[v(z) - v(\bar{p})]}{(1 - v(\bar{p}))(z - \bar{p})}.$$

Taking the p.g.f. of the probability distribution Ω and using (2.4.12) and the conditional p.g.f. above, we obtain $\Omega(z)$. \square

Theorem 2.4.2. For $\rho < 1$, in a Geo/G/1 (E,MAV), the stationary waiting time W_v can be decomposed into the sum of two independent random variables,

$$W_v = W + W_d,$$

where W is the waiting time of a classical Geo/G/1 queue without vacations, with its p.g.f. given in (2.4.2). W_d is the additional delay due to the vacation effect, with the p.g.f.

$$W_d(z) = \frac{p - (z - \bar{p})\Omega(z)}{E(Q_b)(1 - z)}, \quad (2.4.13)$$

where $\Omega(z)$ is as in (2.4.11).

Proof. Note that the D_{k+1} and D_k have the same stationary distribution. From (2.4.10), taking the p.g.f., we have

$$D(z) = P\{D - T \geq 0\}E(z^{D-T} | D \geq T)E(z^S) + P\{D < T\}E(z^\Omega)E(z^S). \quad (2.4.14)$$

Because $P\{D \geq T\} = D(\bar{p})$, we have

$$\begin{aligned} E(z^{D-T} | D \geq T) &= \frac{1}{1 - D(\bar{p})} \sum_{j=0}^{\infty} z^j \sum_{k=j+1}^{\infty} P\{D = k\} p \bar{p}^{k-j+1} \\ &= \frac{1}{1 - D(\bar{p})} \sum_{k=1}^{\infty} P\{D = k\} p \sum_{j=0}^{k-1} \bar{p}^{k-j+1} z^j \\ &= \frac{1}{1 - D(\bar{p})} \frac{p}{\bar{p} - z} (D(\bar{p}) - D(z)). \end{aligned}$$

Substituting the equation above into (2.4.14) gives

$$D(z) = \frac{D(\bar{p}) [p - (z - p)\Omega(z)] G(z)}{pG(z) - z + \bar{p}}. \quad (2.4.15)$$

Using the normalization condition and the L'Hopital rule, we have

$$D(\bar{p}) = \frac{1 - p}{1 + p\Omega'(1)}.$$

Note that $1 + p\Omega'(1) = E(Q_b)$. Substituting $D(\bar{p})$ above into (2.4.15) gives

$$\begin{aligned} D(z) &= W_v(z)G(z) \\ &= \frac{(1-\rho)(1-z)}{pG(z) - z - \bar{p}} \frac{p - (z - \bar{p})\Omega(z)}{E(Q_b)(1-z)} G(z). \end{aligned}$$

From this expression, we obtain the stochastic decomposition property.

□

Based on Theorem 2.4.2, we can get the expected values as follows:

$$\begin{aligned} E(W_d) &= \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})} \frac{pE(V(V-1))}{2E(Q_b)}, \\ E(W_v) &= \frac{p}{2(1-p)} E(S(S-1)) + \frac{1 - H[v(p)]}{1 - v(p)} \frac{p}{2E(Q_b)} E(V(V-1)). \end{aligned}$$

Note that the p.g.f. of L_d can be rewritten as

$$L_d(z) = \frac{H[v(\bar{p})]}{E(Q_b)} + \frac{1}{E(Q_b)} \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})} pE(V) \frac{1 - v[1 - p(1-z)]}{pE(V)(1-z)}.$$

This expression indicates that L_d is a mixture of two random variables. That means that L_d is zero with probability $p^* = H[v(\bar{p})][E(Q_b)]^{-1}$ and, with probability $1 - p^*$, is equal to the number of customers arriving during the residual life of a vacation. Similarly, the p.g.f. of W_d can be rewritten as

$$W_d(z) = \frac{H[v(\bar{p})]}{E(Q_b)} + \frac{1}{E(Q_b)} \frac{1 - H[v(\bar{p})]}{1 - v(\bar{p})} pE(V) \frac{1 - v(z)}{E(V)(1-z)},$$

which shows that W_d is zero with probability p^* and is equal to the residual life of a vacation with probability $1 - p^*$. In addition, we can perform the busy-period analysis for this discrete-time system in the same way as for the M/G/1 (E, MV) system.

2.4.3 Special Cases of the MAV Model

There are several classical models that can be considered as special cases of the Geo/G/1 (E, MAV) model.

Example 1. Discrete-time Geo/G/1 with multiple vacations.

To obtain the results for the Geo/G/1 with multiple vacation and exhaustive service, we can simply let $H = \infty$, $H(z) = 0$. From (2.4.4),

we have

$$Q_b(z) = \frac{1}{1 - v(\bar{p})} [v(1 - p(1 - z)) - v(\bar{p})],$$

$$E(Q_b) = \frac{pE(V)}{1 - v(\bar{p})}.$$

From (2.4.11), we get

$$\Omega(z) = \frac{p}{1 - v(\bar{p})} \frac{v(z) - v(\bar{p})}{z - \bar{p}}.$$

$$E(\Omega) = \frac{1}{1 - v(\bar{p})} \frac{p^2 E(V) - p(1 - v(\bar{p}))}{p^2}.$$

Substituting these expressions into (2.4.7) and (2.4.13) gives the stochastic decomposition properties for the queue length L_v and the waiting time W_v . The p.g.f.'s and the expected values of the additional queue length and delay are given by

$$L_d(z) = \frac{1 - v[1 - p(1 - z)]}{pE(V)(1 - z)},$$

$$W_d(z) = \frac{1 - v(z)}{E(V)(1 - z)},$$

$$E(L_d) = \frac{p}{2E(V)} E(V(V - 1)),$$

$$E(W_d) = \frac{1}{2E(V)} E(V(V - 1)).$$

Like the M/G/1 (E, MV), now W_d is the residual life of a discrete-time vacation and L_d is the number of customers arriving during the residual life. If the vacation time follows a discrete PH distribution of order m , we can use the closure property of the PH distribution to prove easily that L_d and W_d are also discrete PH distributions.

Example 2. Discrete-time Geo/G/1 with single vacation.

Let $H \equiv 1$; then $H(z) = z$. From (2.2.4) and (2.2.11), we have

$$Q_b(z) = v[1 - p(1 - z)] - v(\bar{p})(1 - z),$$

$$E(Q_b) = v(\bar{p}) + pE(V),$$

$$\Omega(z) = v(\bar{p}) + \frac{p[v(z) - v(\bar{p})]}{z - \bar{p}}.$$

Substituting these expressions into (2.4.7) and (2.4.13) gives the stochastic decomposition properties. The p.g.f.'s and the expected values

of the additional queue length and delay are given by

$$\begin{aligned} L_d(z) &= \frac{1 + v(\bar{p})(1 - z) - v[1 - p(1 - z)]}{[v(\bar{p}) + pE(V)](1 - z)}, \\ W_d(z) &= \frac{v(\bar{p})(1 - z) + p(1 - v(z))}{[v(\bar{p}) + pE(V)](1 - z)}, \\ E(L_d) &= \frac{p^2}{2[v(\bar{p}) + pE(V)]} E(V(V - 1)), \\ E(W_d) &= \frac{p}{2[v(\bar{p}) + pE(V)]} E(V(V - 1)). \end{aligned}$$

Note that L_d can be rewritten as

$$L_d(z) = \frac{v(\bar{p})}{v(\bar{p}) + pE(V)} + \frac{pE(V)}{v(\bar{p}) + pE(V)} \frac{1 - v[1 - p(1 - z)]}{pE(V)(1 - z)}.$$

This means that L_d is zero with probability $p^* = v(\bar{p})[v(\bar{p}) + pE(V)]^{-1}$ and is equal to the number of customers arriving during the residual life of a vacation with probability $1 - p^*$. Similarly, $W_d(z)$ can be rewritten as

$$W_d(z) = \frac{v(\bar{p})}{v(\bar{p}) + pE(V)} + \frac{pE(V)}{v(\bar{p}) + pE(V)} \frac{1 - v(z)}{E(V)(1 - z)}.$$

Therefore, W_d is zero with probability p^* and equals the residual life of a vacation with probability $1 - p^*$.

Example 3. Discrete-time Geo/G/1 with setup time.

A Geo/G/1 queue with setup time can be considered as an equivalent Geo/G/1 (E, MV) with the waiting time of the first customer of a busy period being equal to the setup time, U . Let V and $v(z)$ represent the vacation time and its p.g.f., respectively. Using Lemma 2.4.2 or the relation between $\Omega(z)$ and $v(z)$ in Example 1, we have

$$u(z) = \frac{p}{1 - v(\bar{p})} \frac{v(z) - v(\bar{p})}{z - \bar{p}}. \quad (2.4.16)$$

Now to express the $v(z)$ and $E(V)$ of the equivalent Geo/G/1 (E, MV) in terms of the known $u(z)$ and $E(U)$, we take the derivative of both sides of (2.4.16) at $z = 1$ and obtain

$$E(U) = \frac{p^2 E(V) - p(1 - v(\bar{p}))}{(1 - v(\bar{p}))p^2}. \quad (2.4.17)$$

From (2.4.17), we get

$$\begin{aligned} E(V) &= \frac{1}{p}(1 - v(\bar{p})) + (1 - v(\bar{p}))E(U), \\ v(z) &= \frac{1}{p}(1 - v(\bar{p}))(z - \bar{p})u(z) + v(\bar{p}). \end{aligned}$$

Substituting these two equations into $W_d(z)$ of the Geo/G/1 (E, MV) in Example 1 gives

$$W_d(z) = \frac{p - (z - \bar{p})u(z)}{[1 + pE(U)](1 - z)}. \quad (2.4.18)$$

Similarly, replacing z with $1 - p(1 - z)$ in (2.4.16) yields

$$v[1 - p(1 - z)] = zu[1 - p(1 - z)](1 - v(\bar{p})) + v(\bar{p}).$$

Substituting this relation into $L_d(z)$ of the Geo/G/1 (E, MV) in Example 1, we obtain the p.g.f. of the additional queue length due to the setup time effect

$$L_d(z) = \frac{1 - zu[1 - p(1 - z)]}{[1 + pE(U)](1 - z)}. \quad (2.4.19)$$

Note that (2.4.19) can be rewritten as

$$L_d(z) = \frac{1}{1 + pE(U)} + \frac{pE(U)}{1 + pE(U)} z \frac{1 - u[1 - p(1 - z)]}{pE(U)(1 - z)}.$$

This expression indicates that L_d is zero with probability of $p^* = [1 + pE(U)]^{-1}$ and equals the number of arrivals during the residual life of a setup time plus one with probability $1 - p^*$. Similarly, (2.4.18) can be rewritten as

$$W_d(z) = \frac{1}{1 + pE(U)} u(z) + \frac{pE(U)}{1 + pE(U)} \frac{1 - u(z)}{E(U)(1 - z)}.$$

This equation means that the additional delay W_d is equal to a complete setup time with probability p^* and is the residual life of a setup time with probability $1 - p^*$. It is easy to verify that all results for the discrete-time Geo/G/1 type vacation system are similar to those for the corresponding continuous-time M/G/1 vacation system.

2.5 MAP/G/1 Vacation Models

In this section, we discuss the vacation model with nonrenewal arrival process. The Markov arrival process (MAP) is a tractable non-renewal process that can realistically represent the bursty input process in many computer and telecommunication systems. Some popular input processes, such as the Markov-modulated Poisson process (MMPP) and the PH-renewal process, are special cases of the MAP. The complete analysis of MAP/G/1 (E, MV) has been performed by Lucantoni et al. (1990). We present here some main results concerning this type of system. The detailed derivations of these results and other MAP-arrival vacation models can be found in the references provided in the bibliographic notes for this chapter.

MAP Arrival Process. Consider a Markov process on the finite state space $\{1, 2, \dots, m+1\}$, where $\{1, 2, \dots, m\}$ are transient states and $\{m+1\}$ is an absorbing state. The arrival process is defined as follows: The Markov process evolves until the absorption occurs. The epoch of absorption corresponds to an arrival in the arrival process. The Markov process is then instantaneously restarted in a transient state, where the selection of the new state is allowed to depend on the state from which absorption occurred. The sojourn in a transient state i is exponentially distributed with parameter λ_i . When the sojourn time has elapsed, there are two possibilities. With probability p_{ij} $1 \leq j \leq m$, the Markov process enters the absorbing state and is instantaneously restarted in the transient state j . With probability q_{ij} , $1 \leq j \leq m$, $j \neq i$, the process immediately enters the transient state j . Note that

$$\sum_{\substack{j=1 \\ j \neq i}}^m q_{ij} + \sum_{j=1}^m p_{ij} = 1, \quad 1 \leq i \leq m.$$

Equivalently, if for each i , $1 \leq i \leq m$, we define $D_{ij} = \lambda_i p_{ij}$, $1 \leq j \leq m$, $C_{ij} = \lambda_i q_{ij}$, $1 \leq i, j \leq m$, and $C_{ii} = -\lambda_i$, then the probability of an arrival in an infinitesimal interval of length dt that leaves the Markov process in state j , given that the Markov process is in state i , is $D_{ij}dt$. Similarly, the probability that the process enters the transient state j (without an arrival) in an interval of length dt , given that it is in state i , is $C_{ij}dt$, $i \neq j$. In fact, the MAP can be considered as a semi-Markov process whose transition probability matrix $\mathbf{F}(\cdot)$ is of the form

$$\mathbf{F}(x) = \int_0^x e^{\mathbf{C}u} du \mathbf{D} = (\mathbf{I} - e^{\mathbf{C}x})(-\mathbf{C}^{-1})\mathbf{D},$$

where $\mathbf{C} = [C_{ij}]$ and $\mathbf{D} = [D_{ij}]$ are, respectively, a stable matrix and a nonnegative matrix whose sum is an irreducible infinitesimal generator (see Ramaswamy (1990) for properties of the matrix exponential). Let N_t be the number of arrivals in $(0, t]$ and J_t the state of the Markov process at time t . Now let

$$P_{ij}(n, t) = P\{N_t = n, J_t = j | N_0 = 0, J_0 = i\}$$

be the (i, j) entry of the matrix $\mathbf{P}(n, t)$. $\mathbf{P}(n, t)$ satisfies the forward Chapman-Komogorov equations

$$\begin{aligned} \mathbf{P}'(n, t) &= \mathbf{P}(n, t)\mathbf{C} + \mathbf{P}(n-1, t)\mathbf{D}, \quad n \geq 1, t \geq 0, \\ \mathbf{P}(0, 0) &= \mathbf{I}, \end{aligned}$$

and the matrix generating function $\mathbf{P}(z, t) = \sum_{n=0}^{\infty} \mathbf{P}(n, t)z^n$ is explicitly given by

$$\mathbf{P}(z, t) = e^{(\mathbf{C} + z\mathbf{D})t}, \quad |z| \leq 1, t \geq 0.$$

The stationary vector π of this Markov process satisfies the equations

$$\pi(\mathbf{C} + \mathbf{D}) = 0, \quad \pi \mathbf{e} = 1.$$

The fundamental mean of the transition probability matrix $\mathbf{F}(\cdot)$ is given by $\lambda'_1 = (\pi \mathbf{D} \mathbf{e})^{-1}$, so $(\lambda'_1)^{-1}$ is the fundamental arrival rate of the MAP. Note that the assumption that the absorption is certain, starting from any transient state, is equivalent to the nonsingularity of the matrix \mathbf{C} and $-\mathbf{C}^{-1} \geq 0$.

The Embedded Markov Renewal Process. For the MAP/G/1 (E, MV) system with i.i.d service and i.i.d vacation times, denoted by H (rather than B , as defined in most sections of this book) and V , respectively, we can define the embedded Markov renewal process at customer departure instants as follows. Let τ_k be the epoch of the k th departure from the queue, with $\tau_0 = 0$, and let (ξ_k, J_k) be the number of customers in the system and the phase of the arrival process at τ_k^+ . Then $(\xi_k, J_k, \tau_{k+1} - \tau_k)$ is a semi-Markov process on the state space $\{(i, j) : i \geq 0, 1 \leq j \leq m\}$. Let μ'_1 and $E(V)$ be the means of the service time and the vacation time, respectively. The semi-Markov process is positive recurrent when the traffic intensity $\rho = \mu'_1/\lambda'_1$ is less than 1 (note that the symbols μ'_1 and λ'_1 are means rather than rates, as used in other sections). The transition probability matrix is given by

$$\tilde{\mathbf{P}}(x) = \begin{bmatrix} \tilde{\mathbf{B}}_0(x) & \tilde{\mathbf{B}}_1(x) & \tilde{\mathbf{B}}_2(x) & \cdots \\ \tilde{\mathbf{A}}_0(x) & \tilde{\mathbf{A}}_1(x) & \tilde{\mathbf{A}}_2(x) & \cdots \\ 0 & \tilde{\mathbf{A}}_0(x) & \tilde{\mathbf{A}}_1(x) & \cdots \\ 0 & 0 & \tilde{\mathbf{A}}_0(x) & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \quad x \geq 0, \quad (2.5.1)$$

where for $n \geq 0$, $\tilde{\mathbf{A}}_n(x)$ and $\tilde{\mathbf{B}}_n(x)$ are the $m \times m$ matrices of mass functions defined as follows:

$[\tilde{\mathbf{A}}_n(x)]_{ij}$ is the probability that, given a departure at time 0 that left at least one customer in the system and the arrival process in phase i , the next departure occurs no later than time x with the arrival process in phase j , and during that service there were n arrivals; $[\tilde{\mathbf{B}}_n(x)]_{ij}$ is the probability that, given a departure at time 0 that left the system empty and the arrival process in phase i , the next departure occurs no later than time x with the arrival process in phase j , leaving n customers in the system. In addition, we introduce the conditional probabilities

$[\tilde{\mathbf{V}}_n(x)]_{ij}$, the probability that, given a vacation beginning at time 0 with the arrival process in phase i , the end of the vacation occurs no later than x with the arrival process in phase j , and during the vacation there were n arrivals. From the definition of $\mathbf{P}(n, t)$, we have

$$\tilde{\mathbf{A}}_n(x) = \int_0^x \mathbf{P}(n, t) dH(t), \quad \tilde{\mathbf{V}}_n(x) = \int_0^x \mathbf{P}(n, t) dV(t), \quad (2.5.2)$$

We define the transform matrices of $\tilde{\mathbf{A}}_n(x)$ as

$$\mathbf{A}_n^*(s) = \int_0^\infty e^{-sx} d\tilde{\mathbf{A}}_n(x), \quad \mathbf{A}(z, s) = \sum_{n=0}^\infty \mathbf{A}_n^*(s) z^n,$$

and the matrices $\mathbf{A}_n = \mathbf{A}_n(0) = \tilde{\mathbf{A}}_n(\infty)$ and $\mathbf{A} = \mathbf{A}(1, 0)$. Using the properties of $\mathbf{P}(n, t)$, we get

$$\mathbf{A}(z, s) = \int_0^\infty e^{-[s\mathbf{I} - \mathbf{C} - z\mathbf{D}]t} dH(t), \quad \mathbf{V}(z, s) = \int_0^\infty e^{-[s\mathbf{I} - \mathbf{C} - z\mathbf{D}]t} dV(t). \quad (2.5.3)$$

From these expressions, we see that $\mathbf{A} = \int_0^\infty e^{(\mathbf{C} + \mathbf{D})t} d\mathbf{B}(t)$ and matrix \mathbf{A} is stochastic. Note that the stationary vector π satisfies $\pi\mathbf{A} = \pi$, $\pi\mathbf{e} = 1$. The corresponding transform matrices for $\mathbf{B}_n(x)$ can be developed as follows:

$$\begin{aligned} \tilde{\mathbf{B}}_n(x) = & \sum_{i=0}^\infty \sum_{j=1}^{n+1} \int_0^x \int_0^y \int_0^{y-u} dV^{(i)}(u) e^{\mathbf{C}u} \mathbf{P}(j, v) dV(v) dH(y - u - v) \\ & \times \mathbf{P}(n - j + 1, y - u - v). \end{aligned}$$

This expression is obtained by using the decomposition based on the law of total probability. That is: there are i vacations with no arrivals, and the i th vacation ends at time u . The next vacation is of length v , and there are $j \geq 1$ arrivals during that vacation. The first service of the busy period ends at $y \leq x$, and there are $n - j + 1$ arrivals during that service. The transform matrices of $\tilde{\mathbf{B}}_n(x)$ are

$$\mathbf{B}_n^*(s) = \int_0^\infty e^{-sx} d\tilde{\mathbf{B}}_n(x), \quad \mathbf{B}(z, s) = \sum_{n=0}^\infty \mathbf{B}_n^*(s) z^n,$$

It can be shown that the transform matrix $\mathbf{B}(z, s)$ is given by

$$\mathbf{B}(z, s) = \frac{[\mathbf{V}(z, s) - \mathbf{V}(0, s)]\mathbf{A}(z, s)}{z[\mathbf{I} - \mathbf{V}(0, s)]}.$$

It is also easy to prove that the matrices $\mathbf{B}_n(s)$ satisfy

$$\mathbf{B}_n(s) = \sum_{j=0}^n \mathbf{V}_j^0(s) \mathbf{A}_{n-j}(s),$$

where $\mathbf{V}_j^0(s) = [\mathbf{I} - \mathbf{V}_0(s)]^{-1} \mathbf{V}_{j+1}(s)$, for $n \geq 0$. Note that the matrix $\mathbf{V}_j^0 = \mathbf{V}_j^0(0) = (\mathbf{I} - \mathbf{V}_0)^{-1} \mathbf{V}_{j+1}$, for $j \geq 0$, is the probability that, following a sequence of vacations without arrivals, there are $j+1$ arrivals during the first vacation in which arrivals occur.

The Stationary Queue Length at Departures. The stationary vector of Markov chain $\mathbf{P} = \tilde{\mathbf{P}}(\infty)$, embedded at departures from the queue, is the joint probability density of the stationary queue length and the phase of the arrival process. From (2.5.1), we have

$$\mathbf{P} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \mathbf{B}_2 & \cdots \\ \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \cdots \\ 0 & \mathbf{A}_0 & \mathbf{A}_1 & \cdots \\ 0 & 0 & \mathbf{A}_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \quad (2.5.4)$$

Writing the stationary probability vector \mathbf{x} of \mathbf{P} in the petitioned form $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \cdots)$, we get the set of equations

$$\mathbf{x}_i = \mathbf{x}_0 \mathbf{B}_i + \sum_{v=1}^{i+1} \mathbf{x}_v \mathbf{A}_{i+1-v}, \quad i \geq 0. \quad (2.5.5)$$

Once the vector \mathbf{x}_0 is obtained, an efficient recursion presented in Ramaswami (1988) can be used to compute the vectors \mathbf{x}_i , $i \geq 1$. It takes a few steps to compute \mathbf{x}_0 , as shown in Lucantoni et al. (1990). The first step is to study the first-passage times from level $\mathbf{i} + \mathbf{1}$ to \mathbf{i} . Define $\tilde{G}_{jj'}^{[r]}(k; x)$ as the probability that the first passage from state $(i+r, j)$ to state (i, j') , $i \geq 1$, $1 \leq j, j' \leq m$, $r \geq 1$, occurs in exactly k transitions no later than time x , and that (i, j') is the first state visited in level \mathbf{i} . $\tilde{\mathbf{G}}^{[r]}(k; x)$ is the matrix with elements $\tilde{G}_{jj'}^{[r]}(k; x)$. By the first-passage argument, it can be shown (see Neuts (1976)) that the joint transform matrix $\mathbf{G}(z, s)$, defined as

$$\mathbf{G}(z, s) = \sum_{k=1}^{\infty} \int_0^{\infty} e^{-sx} d\tilde{\mathbf{G}}^{[1]}(k, x) z^k, \quad |z| \leq 1, \quad \text{Re } s \geq 0,$$

satisfies the nonlinear matrix equation

$$\mathbf{G}(z, s) = z \sum_{v=0}^{\infty} \mathbf{A}_v(s) \mathbf{G}^v(z, s). \quad (2.5.6)$$

Let us define the matrices

$$\mathbf{G}(z) = \mathbf{G}(z, 0) = z \sum_{v=0}^{\infty} \mathbf{A}_v \mathbf{G}^v(z),$$

$$\mathbf{G} = \mathbf{G}(1) = \sum_{v=0}^{\infty} \mathbf{A}_v \mathbf{G}^v.$$

The matrix \mathbf{G} is stochastic when $\rho < 1$. It can also be shown that $\mathbf{G}(z, s)$ satisfies the functional equation

$$\mathbf{G}(z, s) = z \int_0^{\infty} e^{-sx} e^{[\mathbf{C} + \mathbf{D}\mathbf{G}(z, s)]x} dH(x),$$

which implies that

$$\mathbf{G} = \int_0^{\infty} e^{(\mathbf{C} + \mathbf{D}\mathbf{G})x} dH(x).$$

For $\rho < 1$, the stationary probability vector \mathbf{g} of the positive recurrent stochastic matrix \mathbf{G} satisfies

$$\mathbf{g}\mathbf{G} = \mathbf{g}, \quad \mathbf{g}\mathbf{e} = 1.$$

It can also be shown that \mathbf{g} is the stationary vector of the infinitesimal generator $\mathbf{C} + \mathbf{D}\mathbf{G}$. It is shown in Lucantoni and Ramaswami (1985) that the matrix \mathbf{G} may be efficiently computed by the following recursive scheme. Start with $\mathbf{G}_0 = \mathbf{0}$, and for $k = 0, 1, 2, \dots$, compute

$$\mathbf{H}_{n+1, k} = [\mathbf{I} + \theta^{-1}(\mathbf{C} + \mathbf{D}\mathbf{G}_k)]\mathbf{H}_{n, k}, \quad n = 0, 1, 2, \dots,$$

$$\mathbf{G}_{k+1} = \sum_{n=0}^{\infty} \gamma_n \mathbf{H}_{n, k},$$

where $\mathbf{H}_{0, k} = \mathbf{I}$, $\theta = \max_i(-C_{ii})$, and $\gamma_n = \int_0^{\infty} e^{-\theta x} \frac{(\theta x)^n}{n!} dH(x)$. It is shown in Lucantoni and Ramaswami (1985) that the sequence \mathbf{G}_k converges monotonically to \mathbf{G} . After computing \mathbf{G} , we can obtain \mathbf{g} . The next step is to compute \mathbf{x}_0 . The quantity $(x_{0j})^{-1}$ is the mean recurrence time of the state $(0, j)$ in the Markov chain \mathbf{P} . Considering the chain \mathbf{P} only at its visits to the level $\mathbf{0}$ and recording the indices of the states visited as well as the number of transitions in \mathbf{P} between consecutive visits to $\mathbf{0}$, we obtain an irreducible m -state Markov renewal process, with the transition matrix given by the matrix generating function $\mathbf{K}(z)$. The matrix $\mathbf{K}(z)$ is obtained as follows. Define $\tilde{K}_{jj'}(k; x)$, $k \geq 1, x \geq 0, 1 \leq j, j' \leq m$, as the conditional probability that the

Markov renewal process, starting in the state $(0, j)$, returns to set $\mathbf{0}$ for the first time in exactly k transitions and no later than time x , by hitting the state $(0, j')$. The joint transform matrix of $\tilde{\mathbf{K}}(k; x) = \{\tilde{K}_{jj'}(k; x)\}$ is defined by

$$\mathbf{K}(z, s) = \sum_{k=1}^{\infty} \int_0^{\infty} e^{-sx} d\tilde{\mathbf{K}}(k; x) z^k, \quad |z| \leq 1, \quad \operatorname{Re}(s) \geq 0.$$

A first-passage argument shows that $\mathbf{K}(z, s)$ satisfies

$$\mathbf{K}(z, s) = z \sum_{v=0}^{\infty} \mathbf{B}_v^*(s) \mathbf{G}^v(z, s).$$

We define the matrices

$$\begin{aligned} \mathbf{K}(z) &= \mathbf{K}(z, 0) = z \sum_{v=0}^{\infty} \mathbf{B}_v \mathbf{G}^v(z), \\ \mathbf{K} &= \mathbf{K}(1) = \mathbf{K}(1, 0) = \sum_{v=0}^{\infty} \mathbf{B}_v \mathbf{G}^v. \end{aligned}$$

It can be shown that

$$\mathbf{K}(z, s) = \frac{\mathbf{V}(\mathbf{G}(z, s), s) - \mathbf{V}_0(s)}{\mathbf{I} - \mathbf{V}_0(s)},$$

and, therefore,

$$\mathbf{K} = \mathbf{K}(1, 0) = \frac{\mathbf{V}(\mathbf{G}) - \mathbf{V}_0}{\mathbf{I} - \mathbf{V}_0}.$$

In Neuts (1989), it has been shown that \mathbf{x}_0 can be expressed in terms of the stationary probability vector κ of \mathbf{K} , which satisfies $\kappa \mathbf{K} = \kappa$, $\kappa \mathbf{e} = 1$, and the vector $\kappa^* = \mathbf{K}'(1) \mathbf{e}$, of the row-sum means of $\mathbf{K}(z)$. Specifically, we have

$$\mathbf{x}_0 = \frac{\kappa}{\kappa \kappa^*}.$$

Furthermore, we can show (see Lucantoni et al. (1990)) that

$$\mathbf{x}_0 = \frac{\lambda'_1(1 - \rho)}{E(V)} \mathbf{g}(\mathbf{I} - \mathbf{V}_0).$$

Once \mathbf{x}_0 has been obtained, the remaining components of \mathbf{x} are efficiently computed using a recursion developed by Ramaswami (1988). Defining $\mathbf{X}(z) = \sum_{i=0}^{\infty} \mathbf{x}_i z^i$, we get from (2.5.5)

$$\mathbf{X}(z)[z\mathbf{I} - \mathbf{A}(z)] = \mathbf{x}_0[z\mathbf{B}(z) - \mathbf{A}(z)].$$

Using the expressions for $\mathbf{B}(z)$ and \mathbf{x}_0 , it follows that

$$\begin{aligned}\mathbf{X}(z)[z\mathbf{I} - \mathbf{A}(z)] &= x_0(\mathbf{I} - \mathbf{V}_0)^{-1}[\mathbf{V}(z) - \mathbf{I}]\mathbf{A}(z) \\ &= \frac{\lambda'_1(1 - \rho)}{E(V)}\mathbf{g}(\mathbf{V}(z) - \mathbf{I})\mathbf{A}(z), \quad |z| \leq 1. \quad (2.5.7)\end{aligned}$$

Next, we present a natural matrix analogue of the stochastic decomposition of the queue length at departures in M/G/1 (E, MV).

Theorem 2.5.1. For $|z| \leq 1$, $\mathbf{X}(z) = \mathbf{X}_0(z)\mathbf{V}(z)$, where $\mathbf{X}_0(z)$ is the corresponding transform of the MAP/G/1 queue without vacation and where

$$\mathbf{V}(z) = \frac{\mathbf{V}(z) - \mathbf{I}}{E(V)(\mathbf{C} + z\mathbf{D})},$$

is the matrix generating function of the number of arrivals during a time interval with the same distribution as the residual life of a vacation time.

Proof: For $|z| < 1$, it can be shown (see Heffes and Lucantoni (1986)) that

$$\begin{aligned}\mathbf{X}_0(z) &= \lambda'_1(1 - \rho)\mathbf{g}(\mathbf{C} + z\mathbf{D})\mathbf{A}(z)[z\mathbf{I} - \mathbf{A}(z)]^{-1} \\ &= \lambda'_1(1 - \rho)\mathbf{g}\mathbf{A}(z)\mathbf{A}(z)[z\mathbf{I} - \mathbf{A}(z)]^{-1}(\mathbf{C} + z\mathbf{D}).\end{aligned}$$

The second expression follows from the commutativity of the matrices $\mathbf{C} + z\mathbf{D}$ and $\mathbf{A}(z)$. Based on this expression, we can obtain

$$\mathbf{X}_0(1) = \pi + \lambda'_1(1 - \rho)\mathbf{g}(\mathbf{C} + \mathbf{D})\mathbf{A}(\mathbf{I} - \mathbf{A} + e\pi)^{-1}.$$

Now

$$\mathbf{V}(z) = \sum_{n=0}^{\infty} \int_0^{\infty} P(n, t) \frac{[1 - V(t)]}{E(V)} dt z^n = \int_0^{\infty} e^{(\mathbf{C} + z\mathbf{D})t} \frac{[1 - V(t)]}{E(V)} dt.$$

Integrating by parts and using the commutativity of $\mathbf{C} + z\mathbf{D}$ and $\mathbf{V}(z)$, we have $\mathbf{V}(z) = E(V)^{-1}(\mathbf{C} + z\mathbf{D})^{-1}[\mathbf{V}(z) - \mathbf{I}]$, from which it follows that

$$\mathbf{V}(1) = e\pi - E(V)^{-1}(\mathbf{V} - \mathbf{I})(e\pi - \mathbf{C} - \mathbf{D})^{-1}.$$

Thus, for $|z| < 1$, we have

$$\mathbf{X}_0(z)\mathbf{V}(z) = \lambda'_1(1 - \rho)\mathbf{g}\mathbf{A}(z)[z\mathbf{I} - \mathbf{A}(z)]^{-1} \frac{[\mathbf{V}(z) - \mathbf{I}]}{E(V)}$$

by the commutativity of $\mathbf{C} + z\mathbf{D}$, $\mathbf{V}(z)$, and $\mathbf{A}(z)$. Using (2.5.7), we obtain the decomposition property. \square

In addition, we can relate the queue length distribution at an arbitrary time t to the stationary queue length distribution at departures

by using a classical argument based on the key renewal theorem (see Cinlar (1969)). Therefore, we can obtain all the corresponding results at an arbitrary time and the waiting time distribution of the MAP/G/1 (E,MV) system. Readers are referred to Lucantoni et al. (1990) for the detailed development of these results and for more references on the MAP processes.

2.6 General-Service Bulk Queue with Vacations

2.6.1 $M^X/G/1$ Queue with Vacations

The batch arrival vacation model appears in many situations such as computer communication systems. The common method of studying the batch arrival queueing system with vacations is by using supplementary variables. We present $M^X/G/1$ (E, MV) as an example of this class of models (see the work by Baba (1986)).

Consider an $M^X/G/1$ queue where customers arrive in batches according to a Poisson process with rate λ . The batch size X is a random variable, with the distribution function and p.g.f.

$$P(X = j) = g_j, \quad j = 1, 2, \dots, \quad G(z) = \sum_{j=1}^{\infty} g_j z^j, \quad (2.6.1)$$

respectively, the mean of $g = E(X)$; and the second moment of $g^{(2)} = E(X^2)$. The service times are i.i.d. random variables denoted by B , with general distribution $B(x)$ and probability density $b(x)$. The vacation times are also i.i.d. random variables, denoted by V , with general distribution $V(x)$ and probability density $v(x)$. In addition, the service time and the vacation time are independent. To study the queue length distribution, we use the residual service time or the residual vacation time as the supplementary variable. At an arbitrary time, the steady state of the system can be described by the following random variables:

$$\xi = \begin{cases} 0 & \text{if the server is on vacation,} \\ 1 & \text{if the server is busy,} \end{cases}$$

L_v = the number of customers present,

\widehat{B} = the residual service time for customer in service,

\widehat{V} = the residual vacation time for the server on vacation.

Now we define

$$\begin{aligned} \pi_n(x)dx &= P(L_v = n, x < \widehat{B} \leq x + dx, \xi = 1), \quad n = 1, 2, \dots, \\ \omega_n(x)dx &= P(L_v = n, x < \widehat{V} \leq x + dx, \xi = 0), \quad n = 0, 1, \dots, \end{aligned}$$

and the LST

$$\pi_n^*(s) = \int_0^\infty e^{-sx} \pi_n(x) dx, \quad \omega_n^*(s) = \int_0^\infty e^{-sx} \omega_n(x) dx. \quad (2.6.2)$$

By considering the steady-state transitions, we obtain the following differential difference equations:

$$\begin{aligned} -\frac{d\pi_1(x)}{dx} &= -\lambda\pi_1(x) + \pi_2(0)b(x) + \omega_1(0)b(x), \\ -\frac{d\pi_n(x)}{dx} &= -\lambda\pi_n(x) + \sum_{j=1}^{n-1} \lambda g_j \pi_{n-j}(x) + \pi_{n+1}(0)b(x) + \omega_n(0)b(x), \quad n \geq 2, \\ -\frac{d\omega_0(x)}{dx} &= -\lambda\omega_0(x) + \pi_1(0)v(x) + \omega_0(0)v(x), \\ -\frac{d\omega_n(x)}{dx} &= -\lambda\omega_n(x) + \sum_{j=1}^n \lambda g_j \omega_{n-j}(x), \quad n \geq 1. \end{aligned} \quad (2.6.3)$$

Taking the LST on both sides of the equations of (2.6.3), we have

$$\begin{aligned} -s\pi_1^*(s) + \pi_1(0) &= -\lambda\pi_1^*(s) + \pi_2(0)B^*(s) + \omega_1(0)B^*(s), \\ -s\pi_n^*(s) + \pi_n(0) &= -\lambda\pi_n^*(s) + \sum_{j=1}^{n-1} \lambda g_j \pi_{n-j}^*(s) \\ &\quad + \pi_{n+1}(0)B^*(s) + \omega_n(0)B^*(s), \\ -s\omega_0^*(s) + \omega_0(0) &= -\lambda\omega_0^*(s) + \pi_1(0)V^*(s) + \omega_0(0)V^*(s), \\ -s\omega_n^*(s) + \omega_n(0) &= -\lambda\omega_n^*(s) + \sum_{j=1}^n \lambda g_j \omega_{n-j}^*(s), \quad n \geq 1. \end{aligned} \quad (2.6.4)$$

We also define

$$\begin{aligned} \pi(z, 0) &= \sum_{n=1}^{\infty} \pi_n(0)z^n, & \omega(z, 0) &= \sum_{n=0}^{\infty} \omega_n(0)z^n, \\ \pi^*(z, s) &= \sum_{n=1}^{\infty} \pi_n^*(s)z^n, & \omega^*(z, s) &= \sum_{n=0}^{\infty} \omega_n^*(s)z^n. \end{aligned} \quad (2.6.5)$$

Multiplying the second equation by z^n , summing over n , and using the first equation of (2.6.4) and $G(z)$, we have

$$\begin{aligned} [s - \lambda - \lambda G(z)]\pi^*(z, s) &= -B^*(s)[\pi(z, 0) - \pi_1(0)z]/z \\ &\quad - [\omega(z, 0) - \omega_0(0)]B^*(s) + \pi(z, 0). \end{aligned} \quad (2.6.6)$$

Similarly, multiplying the fourth equation by z^n , summing over n , and using the third equation of (2.6.4), we have

$$[s - \lambda + \lambda G(z)]\omega^*(z, s) = \omega(z, 0) - \pi_1(0)V^*(s) - \omega_0(0)V^*(s). \quad (2.6.7)$$

Substituting $s = \lambda - \lambda G(z)$ into (2.6.6) and (2.6.7), it follows that

$$\begin{aligned} & -B^*(\lambda - \lambda G(z))[\pi(z, 0) - \pi_1(0)z]/z \\ & - [\omega(z, 0) - \omega_0(0)]B^*(\lambda - \lambda G(z)) + \pi(z, 0) = 0, \\ & \omega(z, 0) - \pi_1(0)V^*(\lambda - \lambda G(z)) - \omega_0(0)V^*(\lambda - \lambda G(z)) = 0. \end{aligned} \quad (2.6.8)$$

Next, inserting $z = 0$ in the second equation of (2.6.8) and using $\omega(0, 0) = \omega_0(0)$, we have

$$\omega_0(0) = V^*(\lambda)\pi_1(0)/[1 - V^*(\lambda)]. \quad (2.6.9)$$

Substituting (2.6.9) into the second equation of (2.6.8) gives

$$\omega(z, 0) = V^*(\lambda - \lambda G(z))\pi_1(0)/[1 - V^*(\lambda)]. \quad (2.6.10)$$

From the first equation of (2.6.8) and (2.6.10), we obtain

$$\pi(z, 0) = \frac{zB^*(\lambda - \lambda G(z))[V^*(\lambda - \lambda G(z)) - 1]\pi_1(0)}{[1 - V^*(\lambda)][z - B^*(\lambda - \lambda G(z))]}. \quad (2.6.11)$$

Substituting (2.6.9), (2.6.10), and (2.6.11) into (2.6.6), we get

$$\pi^*(z, s) = \frac{z[V^*(\lambda - \lambda G(z)) - 1][B^*(\lambda - \lambda G(z)) - B^*(s)]\pi_1(0)}{[1 - V^*(\lambda)][z - B^*(\lambda - \lambda G(z))][s - \lambda + \lambda G(z)]} \quad (2.6.12)$$

Substituting (2.6.9) and (2.6.10) into (2.6.7) yields

$$\omega^*(z, s) = \frac{[V^*(\lambda - \lambda G(z)) - V^*(s)]\pi_1(0)}{[1 - V^*(\lambda)][s - \lambda + \lambda G(z)]}. \quad (2.6.13)$$

Since $\pi^*(1, 0) + \omega^*(1, 0) = 1$, using the L'Hopital's rule on (2.6.12) and (2.6.13), we obtain

$$\pi_1(0) = (1 - \rho)[1 - V^*(\lambda)]/E(V).$$

Therefore, the expected number of customers in the system is

$$\begin{aligned} E(L) &= \left. \frac{\partial \pi^*(z, s)}{\partial z} \right|_{z=1, s=0} + \left. \frac{\partial \omega^*(z, s)}{\partial z} \right|_{z=1, s=0} \\ &= \rho + \frac{\lambda g E(V^2)}{2E(V)} + \frac{\lambda[\lambda g^2 b^{(2)} + g^{(2)} E(B)]}{2(1 - \rho)}. \end{aligned} \quad (2.6.14)$$

Now we give the waiting time and the busy-period analysis for this model. The stationary waiting time W_v of an arbitrary or test customer in an arriving batch can be decomposed into the sum of two independent random variables. We first write $W_v = W_1 + W_2$, where W_1 is the waiting time of the first customer in the test customer's batch and W_2 is the waiting time for the service of the batch-mates who are served before the test customer under consideration. The LST of W_1 can be written as

$$\begin{aligned} W_1^*(s) &= \sum_{k=1}^{\infty} \pi_k^*(s) [B^*(s)]^{k-1} + \sum_{k=1}^{\infty} \omega_k^*(s) [B^*(s)]^k \\ &= \pi^*(B^*(s), s) / B^*(s) + \omega^*(B^*(s), s) \\ &= \frac{(1 - \rho)[1 - V^*(s)]}{E(V)[s - \lambda + \lambda G(B^*(s))]} \end{aligned} \quad (2.6.15)$$

Let r_n ($n = 1, 2, \dots$) be the probability of the test customer being in the n th position of the arriving batch. Using the result in the renewal theory (Burke (1975)), we have

$$r_n = \frac{1}{g} \sum_{k=n}^{\infty} g_n.$$

Hence, we have

$$W_2^*(s) = \sum_{k=1}^{\infty} r_k [B^*(s)]^{k-1} = \frac{1 - G(B^*(s))}{g[1 - B^*(s)]}. \quad (2.6.16)$$

Using (2.6.15), (2.6.16), and the fact that W_1 and W_2 are independent, it follows that

$$\begin{aligned} W^*(s) &= W_1^*(s) W_2^*(s) \\ &= \frac{(1 - \rho)s[1 - G(B^*(s))]}{g[s - \lambda + \lambda G(B^*(s))][1 - B^*(s)]} \frac{1 - V^*(s)}{sE(V)}. \end{aligned} \quad (2.6.17)$$

This expression gives the following stochastic decomposition property of the stationary waiting time.

Theorem 2.6.1. For $\rho = \lambda gE(B) < 1$, in an $M^X/G/1$ (E, MV) system, the waiting time W_v can be decomposed into the sum of two independent random variables,

$$W_v = W + W_d,$$

where W is the waiting time of a classical $M^X/G/1$ queue without vacations, with its LST given as

$$W^*(s) = \frac{(1 - \rho)s[1 - G(B^*(s))]}{g[s - \lambda + \lambda G(B^*(s))][1 - B^*(s)]},$$

and W_d is the residual vacation time for the server on vacation, with the LST

$$W_d^*(s) = \frac{1 - V^*(s)}{sE(V)}.$$

From (2.6.17), the expected value of the waiting time is given by

$$E(W_v) = \frac{E(V^2)}{2E(V)} + \frac{\lambda gb^{(2)}}{2(1-\rho)} + \frac{g^{(2)}E(B)}{2g(1-\rho)}.$$

Let us now obtain the LST and the expected value of the busy period D_v . Define D_{vn} as the busy period starting with n customers in the system at a vacation completion instant where $n = 1, 2, \dots$. In Ramaswami (1980), it is shown that $D_{v1}^*(s)$ is the root with the smallest absolute value in z of the equation

$$z = B^*(s + \lambda - \lambda G(z)) \quad (2.6.18)$$

and satisfies

$$D_{vn}^*(s) = [D_{v1}^*(s)]^n.$$

Thus, the LST of D_v is given by

$$\begin{aligned} D_v^*(s) &= \sum_{j=0}^{\infty} D_{vj}^*(s) \int_0^{\infty} \sum_{k=0}^j \frac{(\lambda x)^k}{k!} e^{-\lambda x} g_j^{*k} dV(x) \\ &= \sum_{j=0}^{\infty} [D_{v1}^*(s)]^j \int_0^{\infty} \sum_{k=0}^j \frac{(\lambda x)^k}{k!} e^{-\lambda x} g_j^{*k} dV(x) \\ &= V^*(\lambda - \lambda G(D_{v1}^*(s))), \end{aligned} \quad (2.6.19)$$

where g_j^{*k} is the k th-fold convolution of g_j itself, with $g_j^{*0} = \delta_{j0}$. Taking the first derivative with respect to s and letting $s = 0$ in (2.6.18), we have

$$E(D_{v1}) = E(B)/(1 - \rho). \quad (2.6.20)$$

Similarly, taking the first derivative of (2.6.19) at $s = 0$ and using (2.6.20), we have

$$E(D_v) = \rho E(V)/(1 - \rho). \quad (2.6.21)$$

Using a similar approach, Choudhury (2002) provided a complete analysis on the single vacation batch arrival model ($M^X/G/1$ (E, SV)).

2.6.2 M/G^X/1 Queue with Vacations

Now we discuss the batch service vacation model by using an M/G^(a,b)/1 (E, SV) queue. In such a system, customers arrive according to a Poisson process and are served in batches of maximum size b and minimum threshold a . The server takes a single vacation when it finds less than a customers after the service completion. The results in this section are mainly based on the study of Sikdar and Gupta (2005). For other batch service vacation models including the M/G^(a,b)/1 (E, MV), we provide several references in the bibliographic notes for this chapter.

Similarly to the batch arrival vacation model, the supplementary variable method is used to develop the results below. At an arbitrary time, the steady state of the system can be described by the following random variables:

$$\xi = \begin{cases} 0 & \text{if the server is dormant and ready to serve,} \\ 1 & \text{if the server is on vacation,} \\ 2 & \text{if the server is busy,} \end{cases}$$

L_v = the number of customers present,

\hat{B} = the residual service time of the batch in service,

\hat{V} = the residual vacation time for the server on vacation.

Note that there are differences in the definitions of ξ and \hat{B} between the batch service model and the batch arrival model in the previous section. Accordingly, we define

$$\begin{aligned} \pi_n(x)dx &= P(L_v = n, x < \hat{B} \leq x + dx, \xi = 2), \quad n = 0, 1, 2, \dots, \\ \omega_n(x)dx &= P(L_v = n, x < \hat{V} \leq x + dx, \xi = 1), \quad n = 0, 1, 2, \dots, \\ R_n &= P(L_v = n, \xi = 0), \quad n = 0, 1, 2, \dots, a - 1. \end{aligned}$$

and the LSTs

$$\pi_n^*(s) = \int_0^\infty e^{-sx} \pi_n(x) dx, \quad \omega_n^*(s) = \int_0^\infty e^{-sx} \omega_n(x) dx.$$

It follows from the above that

$$\pi_n^*(0) \equiv \pi_n = \int_0^\infty \pi_n(x) dx, \quad \text{and} \quad \omega_n^*(0) \equiv \omega_n = \int_0^\infty \omega_n(x) dx.$$

It is clear that $\pi_n(\omega_n)$, $n \geq 0$, represents the probability of n customers in the queue when the server is busy (on vacation) at arbitrary time instants.

By considering the steady-state transitions, we obtain the following system of the differential difference equations:

$$\begin{aligned}
0 &= -\lambda R_0 + \omega_0(0), \\
0 &= -\lambda R_n + \lambda R_{n-1} + \omega_n(0), \quad 1 \leq n \leq a-1, \\
-\frac{d\pi_0(x)}{dx} &= -\lambda \pi_0(x) + b(x) \sum_{n=a}^b (\pi_n(0) + \omega_n(0)) + \lambda R_{a-1} b(x), \\
-\frac{d\pi_n(x)}{dx} &= -\lambda \pi_n(x) + \lambda \pi_{n-1}(x) + b(x) (\pi_{n+b}(0) + \omega_{n+b}(0)), \quad n \geq 1, \\
-\frac{d\omega_0(x)}{dx} &= -\lambda \omega_0(x) + \pi_0(0) v(x), \\
-\frac{d\omega_n(x)}{dx} &= -\lambda \omega_n(x) + \lambda \omega_{n-1}(x) + \pi_n(0) v(x), \quad 1 \leq n \leq a-1, \\
-\frac{d\omega_n(x)}{dx} &= -\lambda \omega_n(x) + \lambda \omega_{n-1}(x), \quad n \geq a.
\end{aligned} \tag{2.6.22}$$

Taking the LST on both sides of the last five equations in (2.6.22), we have

$$\begin{aligned}
(\lambda - s)\pi_0^*(s) &= B^*(s) \sum_{n=a}^b (\pi_n(0) + \omega_n(0)) + \lambda R_{a-1} B^*(s) - \pi_0(0), \\
(\lambda - s)\pi_n^*(s) &= \lambda \pi_{n-1}^*(s) + B^*(s) (\pi_{n+b}(0) + \omega_{n+b}(0)) - \pi_n(0), \quad n \geq 1, \\
(\lambda - s)\omega_0^*(s) &= V^*(s) \pi_0(0) - \omega_0(0), \\
(\lambda - s)\omega_n^*(s) &= \lambda \omega_{n-1}^*(s) + V^*(s) \pi_n(0) - \omega_n(0), \quad 1 \leq n \leq a-1, \\
(\lambda - s)\omega_n^*(s) &= \lambda \omega_{n-1}^*(s) - \omega_n(0), \quad n \geq a.
\end{aligned} \tag{2.6.23}$$

Now, using the first two equations of (2.6.22) and all equations of (2.6.23), we obtain a set of results that later lead to queue length distribution at various epochs.

Lemma 2.6.1. There exist two relations

$$\sum_{n=0}^j \omega_n(0) = \lambda R_j, \quad 0 \leq j \leq a-1, \text{ and} \tag{2.6.24}$$

$$\sum_{n=0}^{a-1} \pi_n(0) = \sum_{n=0}^{\infty} \omega_n(0). \tag{2.6.25}$$

Proof: Using the first equation and letting $n = 1$ in the second equation of (2.6.22), we obtain $\sum_{n=0}^1 \omega_n(0) = \lambda R_1$. Recursively, for $n = 2, 3, \dots, a-1$, from the second equation of (2.6.22), we get (2.6.24).

Setting $s = 0$ in the first two equations of (2.6.23), we have

$$\pi_0(0) = \sum_{n=a}^b (\pi_n(0) + \omega_n(0)) + \lambda R_{a-1} - \lambda \pi_0, \quad (2.6.26)$$

$$\pi_n(0) = \pi_{n+b}(0) + \omega_{n+b}(0) + \lambda(\pi_{n-1} - \pi_n), \quad n \geq 1. \quad (2.6.27)$$

Summing over n on (2.6.27), adding (2.6.26), and using (2.6.24), we obtain (2.6.25) after some simplification. \square

Define the nonserving period D_v^c as the sum of a vacation V and an idle time I_v . We then have the following lemma.

Lemma 2.6.2. The expected value of D_v^c is given by

$$E(D_v^c) = E(V) + \frac{1}{\sum_{n=0}^{a-1} p_n^+} \left[\sum_{i=0}^{a-1} p_i^+ \sum_{j=0}^{a-1-i} h_j \frac{(a-i-j)}{\lambda} \right], \quad (2.6.28)$$

where p_i^+ is the stationary probability that i customers are left at a departure instant of a batch, and $h_j = \int_0^\infty \frac{(\lambda x)^j}{j!} e^{-\lambda x} dV(x)$.

Proof: Let $N(t)$ (the number of customers in the system at time t) be the state of the system at time t . Thus, at the end of a busy period, $N(t)$ enters the set of vacation states $\mathbf{S} \equiv \{0, 1, 2, \dots, a-1\}$. The conditional probability that $N(t)$ enters state $i \in \mathbf{S}$, given that $N(t)$ enters \mathbf{S} , is $p_i^+ / \sum_{n=0}^{a-1} p_n^+$. For fixed $i \in \mathbf{S}$, if $j \leq (a-1-i)$ customers arrive during a vacation with probability h_j , then at the vacation completion instant, $N(t)$ enters the set of idle states $\mathbf{U} \equiv \{k : k = a-i-j\}$. Note that $N(t)$ leaves the set \mathbf{U} when $a-(i+j)$ customers arrive. Thus the expected sojourn time of $N(t)$ in \mathbf{U} is $(a-(i+j))/\lambda$. Using the conditional argument and $E(D_v^c) = E(V) + E(I_v)$, we obtain (2.6.28). \square

Lemma 2.6.3. The probability that the server is busy is given by

$$p_b = \frac{\lambda E(B)}{\lambda E(B) + \lambda E(V) \sum_{i=0}^{a-1} p_i^+ + \sum_{n=0}^{a-1} p_n^+ \sum_{j=0}^{a-1-n} A_j}, \quad (2.6.29)$$

where $A_j = \sum_{i=0}^j h_i$.

Proof: Using $p_b = E(D_v)/(E(D_v) + E(D_v^c))$, $E(D_v) = E(B) / \sum_{i=0}^{a-1} p_i^+$ (derived on page 324 in Chaudhry and Templeton (1983)), and (2.6.28), we obtain (2.6.29) after substitution and simplification. \square

In addition, we can find the probability that the server is in the idle state p_{idle} as follows:

$$\begin{aligned}
 p_{idle} &= P(\text{server is in the nonserving period}) \\
 &\times P(\text{server is idle} | \text{server is in the nonserving period}) \\
 &= (1 - p_b)[E(I_v)/E(D_v^c)] \\
 &= (1 - p_b) \frac{\left(1/\lambda \sum_{n=0}^{a-1} p_n^+\right) \left[\sum_{i=0}^{a-1} p_i^+ \sum_{j=0}^{a-1-i} A_j\right]}{E(V) + \left(1/\lambda \sum_{n=0}^{a-1} p_n^+\right) \left[\sum_{i=0}^{a-1} p_i^+ \sum_{j=0}^{a-1-i} A_j\right]}.
 \end{aligned} \tag{2.6.30}$$

Alternatively, by the definition of R_j , we have

$$p_{idle} = \sum_{j=0}^{a-1} R_j. \tag{2.6.31}$$

The probability that the server is on vacation p_v is given by

$$p_v = E(V) \sum_{n=0}^{\infty} \omega_n(0). \tag{2.6.32}$$

Using the fact that $p_b + p_{idle} + p_v = 1$, (2.6.30), (2.6.31), and (2.6.32), we get the following result after some simplification:

$$\sum_{n=0}^{\infty} \omega_n(0) = \frac{1 - p_b}{E(V) + \left(1/\lambda \sum_{n=0}^{a-1} p_n^+\right) \left[\sum_{i=0}^{a-1} p_i^+ \sum_{j=0}^{a-1-i} A_j\right]}. \tag{2.6.33}$$

Now we are ready to get the p.g.f. of the queue length distribution at various epochs.

Theorem 2.6.2. The p.g.f.'s of sequences $\{R_n\}_{n=0}^{a-1}$, $\{\pi_n(0)\}_{n=0}^{\infty}$, $\{\omega_n\}_{n=0}^{\infty}$,

$\{\pi_n^*(s)\}_{n=0}^\infty$, and $\{\omega_n^*(s)\}_{n=0}^\infty$, denoted by $R(z)$, $\pi(z, 0)$, $\omega(z, 0)$, $\pi^*(z, s)$, and $\omega^*(z, s)$, respectively, are given by

$$R(z) = \frac{1}{\lambda(z-1)} \sum_{n=0}^{a-1} (z^a - z^n) \omega_n(0), \quad (2.6.34)$$

$$\begin{aligned} \pi(z, 0) &= \frac{B^*(\lambda(1-z))}{z^b - B^*(\lambda(1-z))} \\ &\times \left[(V^*(\lambda(1-z)) - 1) \sum_{n=0}^{a-1} \pi_n(0) z^n \right. \\ &\quad \left. + \sum_{n=a}^b (z^b - z^n) (\pi_n(0) + \omega_n(0)) + \sum_{n=0}^{a-1} \omega_n(0) (z^b - z^n) \right], \end{aligned} \quad (2.6.35)$$

$$\omega(z, 0) = V^*(\lambda(1-z)) \sum_{n=0}^{a-1} \pi_n(0) z^n, \quad (2.6.36)$$

$$\begin{aligned} \pi^*(z, s) &= \frac{B^*(\lambda(1-z)) - B^*(s)}{(s - \lambda + \lambda z)(z^b - B^*(\lambda(1-z)))} \\ &\times \left[(V^*(\lambda(1-z)) - 1) \sum_{n=0}^{a-1} \pi_n(0) z^n \right. \\ &\quad \left. + \sum_{n=a}^b (z^b - z^n) (\pi_n(0) + \omega_n(0)) + \sum_{n=0}^{a-1} \omega_n(0) (z^b - z^n) \right], \end{aligned} \quad (2.6.37)$$

$$\omega^*(z, s) = \frac{V^*(\lambda(1-z)) - V^*(s)}{s - \lambda + \lambda z} \sum_{n=0}^{a-1} \pi_n(0) z^n. \quad (2.6.38)$$

Proof: From (2.6.22), multiplying the second equation by z^n , summing over n from 1 to $a-1$, and adding the first equation, we get (2.6.34). Now from (2.6.23), multiplying the second equation by z^n , summing over n ($n \geq 1$), and adding the first equation, we have

$$\begin{aligned} &(\lambda - s - \lambda z) \pi^*(z, s) \\ &= \frac{B^*(s) - z^b}{z^b} \pi(z, 0) + \frac{B^*(s)}{z^b} \sum_{n=a}^b (\pi_n(0) + \omega_n(0)) (z^b - z^n) \\ &\quad + \frac{B^*(s)}{z^b} \left(\omega(z, 0) - \sum_{n=0}^{a-1} \pi_n(0) z^n + \sum_{n=0}^{a-1} \omega_n(0) (z^b - z^n) \right). \end{aligned} \quad (2.6.39)$$

Similarly, from (2.6.23), multiplying the fourth and the fifth equations by z^n , summing over n ($n \geq 1$), and adding the third equation, we get

$$(\lambda - s - \lambda z)\omega^*(z, s) = V^*(s) \sum_{n=0}^{a-1} \pi_n(0)z^n - \omega(z, 0). \quad (2.6.40)$$

Inserting $s = \lambda(1 - z)$ in (2.6.39) and (2.6.40), we have

$$\begin{aligned} \pi(z, 0) &= \frac{B^*(\lambda(1 - z))}{B^*(\lambda(1 - z)) - z^b} \\ &\times \left[\sum_{n=a}^b (\pi_n(0) + \omega_n(0))(z^b - z^n) - \omega(z, 0) \right. \\ &\quad \left. + \sum_{n=0}^{a-1} \pi_n(0)z^n - \sum_{n=0}^{a-1} \omega_n(0)(z^b - z^n) \right], \end{aligned} \quad (2.6.41)$$

$$\omega(z, 0) = V^*(\lambda(1 - z)) \sum_{n=0}^{a-1} \pi_n(0)z^n. \quad (2.6.42)$$

Using (2.6.42) in (2.6.41) and (2.6.40), we obtain (2.6.35) and (2.6.38). Also, using (2.6.42) and (2.6.35) in (2.6.39), after simplification we get (2.6.37). \square

Note that the p.g.f.'s of sequences $\{\pi_n\}_0^\infty$ and $\{\omega_n\}_0^\infty$ are $\pi(z) = \pi^*(z, 0)$ and $\omega(z) = \omega^*(z, 0)$, respectively. Setting $s = 0$ in (2.6.37) and (2.6.38), these p.d.f.'s are given by

$$\begin{aligned} \pi(z) &= \frac{B^*(\lambda(1 - z)) - 1}{\lambda(z - 1)(z^b - B^*(\lambda(1 - z)))} \\ &\times \left[(V^*(\lambda(1 - z)) - 1) \sum_{n=0}^{a-1} \pi_n(0)z^n \right. \\ &\quad \left. + \sum_{n=a}^b (z^b - z^n)(\pi_n(0) + \omega_n(0)) + \sum_{n=0}^{a-1} \omega_n(0)(z^b - z^n) \right], \end{aligned} \quad (2.6.43)$$

$$\omega(z) = \frac{V^*(\lambda(1 - z)) - 1}{\lambda(z - 1)} \sum_{n=0}^{a-1} \pi_n(0)z^n. \quad (2.6.44)$$

Furthermore, we obtain the p.g.f. of the queue length of the system.

Theorem 2.6.3. For $\rho < 1$, in an M/G^(a,b)/1 (E, SV) system, the p.g.f. of the stationary queue length L_v at arbitrary time is given by

$$\begin{aligned}
 L_v(z) = & \frac{B^*(\lambda(1-z)) - 1}{\lambda(z-1)(z^b - B^*(\lambda(1-z)))} \\
 & \times \left[\sum_{n=a}^b (z^b - z^n)(\pi_n(0) + \omega_n(0)) + \sum_{n=0}^{a-1} \omega_n(0)(z^b - z^n) \right] \\
 & + \frac{V^*(\lambda(1-z)) - 1}{\lambda(z-1)} \frac{z^b - 1}{z^b - B^*(\lambda(1-z))} \sum_{n=0}^{a-1} \pi_n(0) z^n \\
 & + \frac{1}{\lambda(z-1)} \sum_{n=0}^{a-1} (z^a - z^n) \omega_n(0). \tag{2.6.45}
 \end{aligned}$$

Proof: For the number of customers in the system L_v , we have

$$P\{L_v = n\} = \begin{cases} R_n + \pi_n + \omega_n & 0 \leq n \leq a-1, \\ \pi_n + \omega_n & n \geq a. \end{cases} \tag{2.6.46}$$

Multiplying both sides of (2.6.46) by z^n and summing over n , we obtain

$$L_v(z) = R(z) + \pi(z) + \omega(z). \tag{2.6.47}$$

Substituting (2.6.34), (2.6.43), and (2.6.44) into (2.6.47), we get (2.6.45) after simplification. \square

If we consider the queue length at service and vacation completion instants, we get an embedded Markov chain with two state variables. One is the queue length and the other is an indicator variable φ , with $\varphi = 0$ representing a service completion instant and $\varphi = 1$ a vacation completion instant. For $n \geq 0$, let π_n^+ (ω_n^+) be the probability of n customers in the queue at a service completion (vacation completion) instant. From $\sum_{n=0}^{\infty} (\pi_n^+ + \omega_n^+) = 1$, it follows that

$$\pi_n^+ = \frac{1}{\sigma} \pi_n(0), \quad \omega_n^+ = \frac{1}{\sigma} \omega_n(0),$$

where $\sigma = \sum_{n=0}^{\infty} (\pi_n(0) + \omega_n(0))$. From (2.6.35) and (2.6.36), it is easy to find the p.g.f.'s of π_n^+ and ω_n^+ , respectively, as

$$\begin{aligned} \pi^+(z) &= \frac{B^*(\lambda(1-z))}{z^b - B^*(\lambda(1-z))} \\ &\times \left[(V^*(\lambda(1-z)) - 1) \sum_{n=0}^{a-1} \pi_n^+ z^n + \sum_{n=a}^b (z^b - z^n)(\pi_n^+ + \omega_n^+) \right. \\ &\quad \left. + \sum_{n=0}^{a-1} (z^b - z^n) \omega_n^+ \right], \end{aligned} \quad (2.6.48)$$

$$\omega^+(z) = V^*(\lambda(1-z)) \sum_{n=0}^{a-1} \pi_n^+ z^n. \quad (2.6.49)$$

As defined earlier, p_j^+ , $j \geq 0$, is the stationary probability that j customers are left in the system at a departure epoch of a batch (service completion instant). To find its p.g.f., we introduce two symbols E_1 and E_2 as follows:

$$\begin{aligned} E_1 &= p_b(\lambda E(V)) \left(\sum_{i=0}^{a-1} p_i^+ + \sum_{k=0}^{a-1} p_k^+ \sum_{m=0}^{a-1-k} A_m \right) \\ &\quad + (1 - p_b)(\lambda E(B)) \sum_{i=0}^{a-1} p_i^+, \\ E_2 &= \lambda E(V) \sum_{k=0}^{a-1} p_k^+ + \sum_{k=0}^{\infty} p_k^+ \sum_{m=0}^{a-1-k} A_m. \end{aligned}$$

It is easy to get

$$\sigma = \frac{E_1}{E(B)E_2}. \quad (2.6.50)$$

Now by differentiating the first two equations of (2.6.23) with respect to s at $s = 0$, we obtain

$$\lambda \pi_0^{*(1)}(0) - \pi_0 = -E(B) \sum_{n=a}^b (\pi_n(0) + \omega_n(0)) - \lambda E(B) R_{a-1}, \quad (2.6.51)$$

$$\lambda \pi_0^{*(1)}(0) - \pi_0 = -E(B) \sum_{n=a}^b (\pi_n(0) + \omega_n(0)) - \lambda E(B) a_{a-1}, \quad n \geq 1. \quad (2.6.52)$$

Adding (2.6.51) and (2.6.52) and using (2.6.24) and (2.6.25) gives

$$\sum_{n=0}^{\infty} \pi_n = E(B) \sum_{n=0}^{\infty} \pi_n(0), \quad (2.6.53)$$

which is also the probability that the server is busy, p_b .

Similarly, differentiating the remaining three equations of (2.6.23) with respect to s , setting $s = 0$, and using the same approach, we obtain

$$\sum_{n=0}^{\infty} \omega_n = E(V) \sum_{n=0}^{\infty} \omega_n(0). \quad (2.6.54)$$

Using (2.6.53) and (2.6.50), we get

$$\sum_{i=0}^{\infty} \pi_i^+ = \frac{p_b E_2}{E_1}. \quad (2.6.55)$$

From $p_n^+ = \pi_n^+ / \sum_{i=0}^{\infty} \pi_i^+$ and (2.6.55), it follows that

$$p_n^+ = \left(\frac{E_1}{p_b E_2} \right) \pi_n^+. \quad (2.6.56)$$

Multiplying both sides of (2.6.56) by z^n , summing over n , and substituting $\pi^+(z)$ from (2.6.48), we get the following theorem.

Theorem 2.6.4. The p.g.f. of p_n^+ is given by

$$\begin{aligned} P^+(z) = & \left[\frac{E_1}{p_b E_2} \frac{B^*(\lambda(1-z))}{z^b - B^*(\lambda(1-z))} \right] \\ & \times \left[(V^*(\lambda(1-z)) - 1) \sum_{n=0}^{a-1} \pi_n^+ z^n \right. \\ & \left. + \sum_{n=a}^b (z^b - z^n)(\pi_n^+ + \omega_n^+) + \sum_{n=0}^{a-1} (z^b - z^n) \omega_n^+ \right]. \quad (2.6.57) \end{aligned}$$

Based on the transform equations of (2.6.23), we can develop some relations among these queue length distributions at various epochs that are useful in numerically computing these distributions. Here are a few

important relations.

$$\omega_n = \frac{\sigma}{\lambda} \sum_{j=0}^n (\pi_j^+ - \omega_j^+), \quad 0 \leq n \leq a-1, \quad (2.6.58)$$

$$\omega_n = \frac{\sigma}{\lambda} \left(\sum_{j=0}^{a-1} \pi_j^+ - \sum_{j=0}^n \omega_j^+ \right), \quad n \geq a, \quad (2.6.59)$$

$$R_n = \frac{\sigma}{\lambda} \sum_{i=0}^n \omega_i^+, \quad 0 \leq n \leq a-1. \quad (2.6.60)$$

$$\pi_n = \frac{\sigma}{\lambda} \left(\sum_{i=0}^{b+n} \omega_i^+ + \sum_{i=a}^{b+n} \pi_i^+ - \sum_{i=0}^n \pi_i^+ \right), \quad n \geq 0. \quad (2.6.61)$$

Define $a_j = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dB(t) = \frac{(-\lambda)^j}{j!} B^{*(j)}(\lambda)$, $j \geq 0$, and $h_j = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dV(t) = \frac{(-\lambda)^j}{j!} V^{*(j)}(\lambda)$, $j \geq 0$. Under the , the probabilities $\{\pi_n^+\}_0^\infty$, and $\{\omega_n^+\}_0^\infty$ satisfy the following equations:

$$\pi_0^+ = a_0 \sum_{i=0}^{a-1} \omega_i^+ + a_0 \sum_{i=a}^b (\pi_i^+ + \omega_i^+), \quad (2.6.62)$$

$$\pi_n^+ = a_n \sum_{i=0}^{a-1} \omega_i^+ + \sum_{k=1}^n a_{n-k} (\pi_{b+k}^+ + \omega_{b+k}^+) + a_n \sum_{k=a}^b (\pi_k^+ + \omega_k^+), \quad n \geq 1, \quad (2.6.63)$$

$$\omega_n^+ = \sum_{j=0}^n a_j \pi_{n-j}^+, \quad 0 \leq n \leq a-1, \quad (2.6.64)$$

$$\omega_n^+ = \sum_{j=1}^a a_{n-a+j} \pi_{a-j}^+, \quad n \geq a, \quad (2.6.65)$$

and $\sum_{n=0}^\infty (\pi_n^+ + \omega_n^+) = 1$.

From (2.6.58), (2.6.59), (2.6.60), and (2.6.61), it is clear that $\{R_n\}_0^\infty$, $\{\pi_n\}_0^\infty$, and $\{\omega_n\}_0^\infty$ can be obtained by using $\{\pi_n^+\}_0^\infty$ and $\{\omega_n^+\}_0^\infty$. Note that $\{\pi_n^+\}_0^\infty$ is dependent on $\{\omega_n^+\}_0^\infty$. From (2.6.64) and (2.6.65), we find that $\{\pi_n\}_0^{a-1}$ is needed to compute $\{\omega_n^+\}_0^\infty$. In addition, from (2.6.63) we also need to get $\{\pi_n\}_a^b$. These probabilities can be obtained by using $\{p_n^+\}_0^\infty$, which are computed by solving a set of equations $\mathbf{p}^+ = \mathbf{p}^+ \mathbf{P}$, where $\mathbf{p}^+ = [p_0^+, p_1^+, \dots, p_j^+, \dots]$ and $\mathbf{P} = [p_{ij}]$ is the transition probability matrix of the Markov chain embedded at the batch departure instants,

with p_{ij} 's given by

$$p_{ij} = \begin{cases} \sum_{n=0}^{b-i} a_n g_0, & 0 \leq i \leq a-1, j=0, \\ \sum_{n=0}^{b-i} a_n g_j + \sum_{m=b-i+1}^{b+j-i} a_m g_{j-m+b-i}, & 0 \leq i \leq a-1, j \geq 1, \\ g_j, & a \leq i \leq b, j \geq 0 \\ g_{j-(i-b)}, & j \leq i-b, i \geq b+1, j \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

The system of equations can be solved by using the algorithm given in Latouche and Ramaswami (1999). The algorithm is based on the state truncation method, in which p_{ij} is truncated so that $\sum_{j=0}^l p_{ij} = 1$ for all i , i.e., $p_{il} = 1 - \sum_{j=0}^{l-1} p_{ij}$, $0 \leq i \leq l$, where l indicates a sufficiently large i and j ($i = j$) so that \mathbf{P} becomes an $l \times l$ square matrix. Here is a summary of the procedure of computation:

- Step 1: Using the algorithm called GTH in Latouche and Ramaswami (1999) to solve the equation system $\mathbf{p}^+ = \mathbf{p}^+ \mathbf{P}$ and get $\{p_i^+\}_0^l$.
- Step 2: Compute p_b using (2.6.29).
- Step 3: Compute $\sum_{n=0}^{\infty} \pi_n^+$ using (2.6.55).
- Step 4: Compute $\{\pi_n^+\}_0^l$ using the relation $\pi_n^+ = p_n^+ \sum_{n=0}^{\infty} \pi_n^+$.
- Step 5: Compute $\{\omega_i^+\}_0^l$ using (2.6.64) and (2.6.65).
- Step 6: Compute σ using (2.6.50).
- Step 7: Compute $\{\omega_i\}_0^{a-1}$ and $\{\omega_i\}_0^l$ using (2.6.58) and (2.6.59), respectively.
- Step 8: Compute $\{R_i\}_0^{a-1}$ and $\{\pi_i\}_0^{a-1}$ using (2.6.60) and (2.6.61), respectively.
- Step 9: Finally, compute $\{p_i\}_0^{a-1}$ and $\{p_i\}_a^l$ using (2.6.46).

2.7 Finite-Buffer M/G/1 Queue with Vacations

The vacation models discussed in the previous sections have infinite buffers for waiting customers. However, some practical queueing systems in computer or telecommunication networks have finite-buffers for waiting messages. The early work on the finite buffer vacation system was reported by Lee (1984) using the embedded Markov chain at both service and vacation completion epochs, the supplementary variable, and the sample-biasing technique. Frey and Takahashi (1997) studied the same vacation system using a simpler analysis. The results in this section are based mainly on the work by Frey and Takahashi (1997).

Consider an M/G/1 (E, MV) system with a finite buffer of capacity N , denoted by M/G/1/N (E, MV). We assume that the service discipline is nonpreemptive and FCFS order. With the same symbols used before, such as B for the service time, V for the vacation time, and

$$a_j = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dB(t), \quad v_k = \int_0^\infty \frac{(\lambda t)^k}{k!} e^{-\lambda t} dV(t),$$

we have the probability that j customers arrive (and are accepted) during an idle period (the server is on vacation), denoted by φ_j , as

$$\begin{aligned} \varphi_j &= \sum_{l=0}^{\infty} (v_0)^l v_j = \frac{v_j}{1 - v_0}, \quad j = 1, \dots, N-1, \\ \varphi_N &= \sum_{l=0}^{\infty} (v_0)^l v_N^c = \frac{v_N^c}{1 - v_0}, \quad \text{where } v_N^c = \sum_{j=N}^{\infty} v_j. \end{aligned}$$

Let π_j , $j = 0, \dots, N-1$, be the stationary probability that j customers are left in the system at a customer departure instant, and define

$$a_j^c = \sum_{i=j}^{\infty} a_i.$$

Clearly, the stationary distribution π_j satisfies the following equilibrium equations:

$$\begin{aligned} \pi_j &= \pi_0 \sum_{i=1}^{j+1} \varphi_i a_{j-i+1} + \sum_{i=1}^{j+1} \pi_i a_{j-i+1}, \quad j = 0, \dots, N-2, \\ \pi_{N-1} &= \pi_0 \sum_{i=1}^N \varphi_i a_{N-i}^c + \sum_{i=1}^{N-1} \pi_i a_{N-i}^c, \\ \sum_{j=0}^{N-1} \pi_j &= 1. \end{aligned} \tag{2.7.1}$$

From (2.7.1), we can numerically solve the stationary distribution $\{\pi_j\}_{j=0}^{N-1}$ recursively.

It is worth noting that π_j 's are different from the probabilities of the number of customers in the system, p_j 's, given in Lee (1984), where vacation completion epochs are also considered. The relationship is given by

$$\pi_j = \frac{p_j}{\sum_{i=0}^{N-1} p_i}, \quad j = 0, \dots, N-1.$$

Now we derive the relationship between the queue length distribution at an arbitrary time, denoted by $\{\pi_j^*\}_0^N$, and the queue length distribution at a customer departure epochs $\{\pi_j\}_0^{N-1}$. We focus on the service facility (or the server), excluding the waiting buffer. From the PASTA property, it follows that π_N^* is also the probability that N customers are in the system just before an arrival epoch. Thus the rate $\lambda(1 - \pi_N^*)$ is the arrival rate of customers accepted by the system, which is also the throughput of the service facility. Using Little's law, the expected number of customers in the service facility is equal to $\rho' = \lambda(1 - \pi_N^*)E(B)$, which is also the probability that the server is busy. The following lemma gives another expression of ρ' .

Lemma 2.7.1. ρ' is given by

$$\rho' = \frac{E(B)(1 - v_0)}{E(V)\pi_0 + E(B)(1 - v_0)}. \quad (2.7.2)$$

Proof: Consider two point processes. One is formed by the beginning epochs of busy periods, and the other is formed by the end epochs of busy periods. Denote the rates of these two point processes by λ_b and λ_e , respectively. Note that $(1 - \rho')/E(V)$ is the rate of the point process formed by the vacation termination instants, and the probability that the system is not empty is $1 - v_0$. Thus we have

$$\lambda_b = \frac{(1 - \rho')(1 - v_0)}{E(V)}. \quad (2.7.3)$$

On the other hand, the rate of the point process formed by the service completion instants is $\rho'/E(B)$, and the probability that no customer is left in the system at these instants is π_0 . Therefore, we get

$$\lambda_e = \frac{\rho'\pi_0}{E(B)}. \quad (2.7.4)$$

Using the fact that $\lambda_b = \lambda_e$, (2.7.3) and (2.7.4), we obtain (2.7.2). \square

Theorem 2.7.1. The stationary queue length distribution at an arbitrary time $\{\pi_j^*\}_0^N$ is given by

$$\pi_j^* = \frac{\pi_j(1 - v_0)\lambda^{-1}}{E(V)\pi_0 + E(B)(1 - v_0)}, \quad j = 0, \dots, N - 1, \quad (2.7.5)$$

$$\pi_N^* = 1 - \frac{(1 - v_0)\lambda^{-1}}{E(V)\pi_0 + E(B)(1 - v_0)}. \quad (2.7.6)$$

Proof: From $\rho' = \lambda(1 - \pi_N^*)E(B)$ and (2.7.2), we solve for π_N^* to get (2.7.6). Based on the PASTA property, we see that π_j^* is also the probability that there are j customers in the system just before an arrival. It

follows from the general version of Burke's theorem (see Cooper (1981)) that

$$\pi_j = \frac{\pi_j^*}{1 - \pi_N^*}, \quad j = 0, \dots, N-1. \quad (2.7.7)$$

Substituting (2.7.6) into (2.7.7) yields (2.7.5). \square

Note that in a finite buffer system, π_N^* in (2.7.6) is the probability that an arrival is lost, and is thus an important system performance measure.

We now derive the LST of the waiting time of this finite-buffer vacation system. The waiting time of an arriving customer depends on the number of customers in the system and on the residual service time if the server is attending the queue or the residual vacation time if the server is on vacation at this instant. We need the joint distribution of the residual service time (or the residual vacation time) and the number of arrivals during the backward-recurrence service time (or the backward-recurrence vacation time). Define \hat{B} as the residual service time, \hat{V} as the residual vacation time, $N_{\hat{B}}$ as the number of arrivals during the backward-recurrence service time, and $N_{\hat{V}}$ as the number of arrivals during the backward-recurrence vacation time. The quantities

$$\begin{aligned} \tilde{a}_n(s) &= P(N_{\hat{B}} = n)E(e^{-s\hat{B}}|N_{\hat{B}} = n), \\ \tilde{v}_n(s) &= P(N_{\hat{V}} = n)E(e^{-s\hat{V}}|N_{\hat{V}} = n), \end{aligned}$$

were derived in Lee (1984) as

$$\begin{aligned} \tilde{a}_n(s) &= \frac{1}{(\lambda - s)E(B)} \left\{ B^*(s) \left(\frac{\lambda}{\lambda - s} \right)^n - \sum_{j=0}^n a_j \left(\frac{\lambda}{\lambda - s} \right)^{n-j} \right\}, \\ \tilde{v}_n(s) &= \frac{1}{(\lambda - s)E(V)} \left\{ V^*(s) \left(\frac{\lambda}{\lambda - s} \right)^n - \sum_{j=0}^n v_j \left(\frac{\lambda}{\lambda - s} \right)^{n-j} \right\}. \end{aligned} \quad (2.7.8)$$

Theorem 2.7.2. The LST of the waiting time, denoted by W_v , is given by

$$\begin{aligned} W_v^*(s) &= B^*(s)^{N-1} \sum_{j=0}^{N-1} \pi_j \left(\frac{\lambda}{\lambda - s} \right)^{N-j} \\ &\quad + \lambda \frac{(1 - (\lambda/(\lambda - s))^N)(B^*(s))^N \pi_0 / (1 - v_0)(V^*(s) - 1)}{\lambda - s + \lambda B^*(s)}. \end{aligned} \quad (2.7.9)$$

Proof: An arriving customer sees the server either serving with probability ρ' or on vacation with probability $1 - \rho'$. If the server is serving, the probability that the actual service epoch started with k customers is $\pi_k + \pi_0\varphi_k$. Thus the LST of the waiting time is given by

$$W_v^*(s) = \frac{1}{1 - \pi_N^*} \left\{ \sum_{j=1}^{N-1} \sum_{k=1}^j \rho'(\pi_k + \pi_0\varphi_k) \tilde{a}_{j-k}(s) (B^*(s))^{j-1} + \sum_{j=0}^{N-1} (1 - \rho') \tilde{v}_j(s) (B^*(s))^j \right\}. \quad (2.7.10)$$

Substituting (2.7.8) into (2.7.10) gives (2.7.9) after some algebraic simplification. \square

Remark 2.7.1. Using a transform-free method, Niu and Cooper (1993) presented the waiting time distribution in terms of the stationary probability that there are k customers in the system immediately after a service-start epoch σ_k . The relation between σ_k and π_k is given by

$$\sigma_k = \pi_{k+1} + \pi_0\varphi_{k+1}, \quad k = 0, \dots, N-1.$$

2.8 Bibliographic Notes

A large number of studies in vacation models focus on the M/G/1 systems with exhaustive service and single or multiple vacations. These studies include those by Levy and Yechiali (1975), Scholl and Kleinrock (1983), Fuhrmann (1984), Fuhrmann and Cooper (1985), Levy and Kleinrock (1986), Keirson and Servi (1987), Harris and Marchal (1988), Takine and Hasekawa (1992), Brill and Harris (1992, 1997), Fery and Takahashi (1998), and Madan and Saleh (2001). Doshi (1990) and Takagi (1991) provided a systematic treatment of the exhaustive service M/G/1 vacation model. There is also some work on the transient behavior of this class of vacation models: see Keilson and Ramaswamy (1988), Takagi (1992), and Tang (1994). Kella (1990) presented a more general vacation policy, namely, at the completion of the $(i-1)$ th vacation ($i \geq 1$), if there are waiting customers, the server starts serving customers; otherwise, the server takes a vacation with probability p_i and enters the idle period with probability $q_i = 1 - p_i$. Clearly, the case of $p_i = 1$ corresponds to the multiple vacation policy and the case of $p_i = 0$ for $i \geq 2$ corresponds to the single vacation policy. The multiple adaptive vacation policy described in section 2.1 and introduced by Tian (1992) is another generalization of the multiple vacation, single vacation, and setup time models. Li and Zhu (1995) suggested a hybrid vacation policy that is also a generalization of the multiple and single vacation

policies. Yadin and Noar (1963) first studied the N -policy, which shuts down the server when the system is empty and turns on the server when the number of customers reaches a critical value N . The N -policy was later introduced in the vacation models. Some research work related to the N -policy includes that by Heyman (1968), Balachandran (1973), Shanthikumar (1981), Borthakur et al. (1987), Rubin and Zhang (1988), Lee and Srinivasan (1989), Tian et al. (1991), Federgruen and So (1991), Medhi and Templeton (1992), Mhu (1993), Takagi (1993b), Lee et al. (1994a, 1994b, 1996), Lee et al. (1995), Chae and Lee (1995), Park and Lee (1997), Artalejo (1998), Hur and Park (1999), Lee et al. (2001), etc.. Similar to the N -policy, are two other popular control policies in queueing systems, namely, T -policy and D -policy, which can be found in Heyman (1997), Balachandran and Tijms (1975), Artalejo (2001a, 2001b), Feinberg and Kella (2002). The N -policy was generalized to the two-threshold (r, N) policy by Dshalalow (1998). With an (r, N) policy, called the *hysteretic control*, the server is shut down when the number of customers is reduced to r (≥ 1). This policy is also related to the batch service system, where the server starts serving the customers in batch when the number of the customers reach a minimum batch size. For the studied on the (r, N) policy systems, see Chaudhry and Templeton (1981), Easton and Chaudhry (1982), Chaudhry et al. (1987), Jacob and Madhusoodanan (1987), Gold and Tran-Gia (1993), Dshalalow (1991, 1997). For comprehensive treatment of the (r, N) policy model, see Dshalalow (1998).

There are many research works on M/G/1 type vacation models with batch arrivals; see Baba (1986), (1987), Lee and Srinivasan (1989), Rosenberg and Yechiali (1993), Lee (1995) and Chaudhry (2000) etc. An extension of the Poisson arrivals is to introduce the nonrenewal arrival process in the vacation models. The general theory of the nonrenewal arrival processes can be found in Neuts (1979) and Lucantoni et al. (1990). If the nonrenewal arrival process is a Markov arrival process (MAP), the queueing system can be treated by using the matrix geometric method. Most results for the M/G/1 vacation models have been extended to the MAP/G/1 vacation models. Blondia (1991) studied the vacation model with a nonrenewal arrival process and a finite buffer. Scholl and Kleinrock (1994) discussed the MAP/G/1 vacation system with batch arrivals. Ferrandiz (1993) treated the BMAP/G/1 queue with either setup times or vacations. Martendo (1993) presented the vacation model with batch nonrenewal arrivals. Schellhaas (1994) studied the MAP/G/1 system with batch arrivals, multiple vacations, and a finite buffer. Other work related to nonrenewal arrival or MAP arrival vacation models includes that by Meier-Hellstern and Neuts (1990), Takine and Hasegawa (1993),

Takine (2001), Lee (2001), Alfa (1995), and Niu et al. (1999), (2003). There is also some research work that extends the M/G/1 type vacation model to the batch service case: see Nadarajan and Subranabuvam (1984), Madan (1991), Reddy and Anitha (1998), Sikdar and Gupta (2005), etc. For a variety of M/G/1 type vacation models with the finite buffer, see Courtois (1980), Lee (1994), Jacob et al. (1987), Takagi (1992, 1994), Bruneel (1994), Lee (1995), etc. It is worth noting that there are also some studies on the M/G/1 vacation models with retrials that have applications in the performance evaluation of computer and communication networks. This class of model can be found in Li and Yang (1995), Artalejo (1997), Choi (1999), Kumar and Arivudainambi (2002), etc.

Discrete-time models are more appropriate for modeling computer and telecommunication systems. Compared with research on continuous-time vacation models, there are fewer studies on discrete-time vacation models. The Geo/G/1 queue with multiple adaptive vacations was analyzed by Zhang and Tian (2001). For other related works on discrete-time vacation models, see Bruneel (1984, 1994), Isgizaki et al. (1995), Fiems and Bruneel (2001, 2002), etc.. Alfa (1995, 1998) presented an analysis of the discrete-time MAP/PH/1 type vacation model. Using the matrix analytical method, Alfa (2003) treated a more general type of Markov-based-representation discrete-time vacation models. Takagi (1993) provided a complete analysis of the discrete-time Geo/G/1 queue with and without vacations.



<http://www.springer.com/978-0-387-33721-0>

Vacation Queueing Models

Theory and Applications

Tian, N.; Zhang, Z.G.

2006, XII, 386 p. 8 illus., Hardcover

ISBN: 978-0-387-33721-0