Theorem 1: The working-vacation-breakdown-repair queueing model with the generator given in (7.2) is stable if and only if

$$\lambda < \mu \left[\frac{\xi + \theta\gamma}{\xi + \gamma} \right]. \tag{7.13}$$

Proof: Based on Theorem $3.2.1$ in the book by He [168], the $QBD-$process is stable if and only if

$$\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}, \tag{7.14}$$

where $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$ is given in equation (7.6).
Substituting the vector $\boldsymbol{\pi}$ from equation (7.6) into inequality (7.14) gives

$$\mu(\boldsymbol{\beta}(-S)^{-1} \otimes \boldsymbol{\delta})(I_n \otimes D_1)\boldsymbol{e} < \left(\frac{\xi\mu + \theta\gamma\mu}{\xi + \gamma} \right) (\boldsymbol{\beta}(-S)^{-1} \otimes \boldsymbol{\delta})(\boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m)\boldsymbol{e}. \tag{7.15}$$

Note that $\lambda = \boldsymbol{\delta} D_1 \boldsymbol{e}$ and $\boldsymbol{\delta}\boldsymbol{e} = 1$. Using $\mu = [\boldsymbol{\beta}(-S)^{-1}\boldsymbol{e}]^{-1}$, the above inequality yields

$$\lambda < \left(\frac{\xi\mu + \theta\gamma\mu}{\xi + \gamma} \right) \boldsymbol{\beta}(-S)^{-1} \boldsymbol{S}^0 \boldsymbol{\beta}\boldsymbol{e}. \tag{7.16}$$

The stated result in (7.13) now follows using

$$\boldsymbol{\beta}\boldsymbol{e} = 1 \ \text{ and } \ S\boldsymbol{e} + \boldsymbol{S}^0 = 0. \tag{7.17}$$

□

The stability condition in above theorem can be explained intuitively. Towards this end, we rewrite the equation in (7.13) as

$$\lambda < \frac{\mu\xi}{\xi + \gamma} + \frac{\theta\mu\gamma}{\xi + \gamma}. \tag{7.18}$$

The effective service rate is the quantity on the right-hand side of (7.18), which is obtained as the sum of the rates of the services offered by the main and the backup servers. Obviously, the arrival rate has to be less than the effective service rate in order for the queue to be stable.

7.3 The Stationary Probability Vector

Let $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots)$, where $\boldsymbol{x}_0$ is of dimension $2m$ and $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots$ are of dimension $3mn$, satisfy

$$\boldsymbol{x}Q = \mathbf{0}, \quad \boldsymbol{x}\boldsymbol{e} = 1. \tag{7.19}$$

When the working-vacation-breakdown-repair queue under study is stable (i.e., condition in (7.13) is satisfied), the stationary probability vector $\boldsymbol{x}$ is derived as (see, Neuts [184]):

$$\boldsymbol{x}_i = \boldsymbol{x}_1 R^{i-1}, i \geq 1. \tag{7.20}$$

The matrix R is the minimal non-negative solution to

$$R^2 A_2 + RA_1 + A_0 = 0, \tag{7.21}$$

and $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ are found by solving

$$\begin{aligned} &\boldsymbol{x}_0 B_1 + \boldsymbol{x}_1 B_2 = \mathbf{0}, \\ &\boldsymbol{x}_0 B_0 + \boldsymbol{x}_1 (A_1 + RA_2) = \mathbf{0}, \end{aligned} \tag{7.22}$$

subject to the (normalizing) condition

$$\boldsymbol{x}_0 \boldsymbol{e} + \boldsymbol{x}_1 (I_{3mn} - R)^{-1} \boldsymbol{e} = 1. \tag{7.23}$$

For later use, the vectors are partitioned as $\boldsymbol{x}_0 = (\boldsymbol{v}_0, \boldsymbol{w}_0)$, and $\boldsymbol{x}_i = (\boldsymbol{u}_i, \boldsymbol{v}_i, \boldsymbol{w}_i), i \geq 1$, such that $\boldsymbol{v}_0$ and $\boldsymbol{w}_0$ are of dimension m and $\boldsymbol{u}_i$, $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$ are of dimension mn.

The following are the interpretations of the vectors in steady-state.

- $\boldsymbol{v}_0$: the main server is on vacation with no one in the system and the arrival is in one of m phases.
- $\boldsymbol{w}_0$: the main server is under repair with no one in the system and the arrival is in one of m phases.
- $\boldsymbol{u}_i$: the system has exactly i customers with the main server busy serving with the

arrival and services are in various phases.

- $\boldsymbol{v}_i$: the system has exactly i customers with the main server on vacation with the arrival and services are in various phases.
- $\boldsymbol{w}_i$: the system has exactly i customers with the main server under repair with the arrival and services are in various phases.

The steady-state equations given in (7.19) can be rewritten as

$$\boldsymbol{v}_0 D_0 + \xi \boldsymbol{w}_0 + \boldsymbol{u}_1(\boldsymbol{S}^0 \otimes I_m) + \boldsymbol{v}_1(\theta \boldsymbol{S}^0 \otimes I_m) = \boldsymbol{0}, \tag{7.24}$$

$$\boldsymbol{w}_0(D_0 - \xi I_m) + \boldsymbol{w}_1(\theta \boldsymbol{S}^0 \otimes I_m) = \boldsymbol{0}, \tag{7.25}$$

$$\boldsymbol{u}_1(S \oplus D_0 - \gamma I_{mn}) + \eta \boldsymbol{v}_1 + \xi \boldsymbol{w}_1 + \boldsymbol{u}_2(\boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m) = \boldsymbol{0}, \tag{7.26}$$

$$\boldsymbol{v}_0(\boldsymbol{\beta} \otimes D_1) + \boldsymbol{v}_1(\theta S \oplus D_0 - \eta I_{mn}) + \boldsymbol{v}_2(\theta \boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m) = \boldsymbol{0}, \tag{7.27}$$

$$\boldsymbol{w}_0(\boldsymbol{\beta} \otimes D_1) + \gamma \boldsymbol{u}_1 + \boldsymbol{w}_1(\theta S \oplus D_0 - \xi I_{mn}) + \boldsymbol{w}_2(\theta \boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m) = \boldsymbol{0}, \tag{7.28}$$

$$\boldsymbol{u}_{i-1} D_1 + \boldsymbol{u}_i(S \oplus D_0 - \gamma I_{mn}) + \eta \boldsymbol{v}_i + \xi \boldsymbol{w}_i + \boldsymbol{u}_{i+1}(\boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m) = \boldsymbol{0}, \;\; i \geq 2, \tag{7.29}$$

$$\boldsymbol{v}_{i-1} D_1 + \boldsymbol{v}_i(\theta S \oplus D_0 - \eta I_{mn}) + \boldsymbol{v}_{i+1}(\theta \boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m) = \boldsymbol{0}, \;\; i \geq 2, \tag{7.30}$$

$$\boldsymbol{w}_{i-1} D_1 + \gamma \boldsymbol{u}_i + \boldsymbol{w}_i(\theta \boldsymbol{S}^0 \oplus D_0 - \xi I_{mn}) + \boldsymbol{w}_{i+1}(\theta \boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m) = \boldsymbol{0}, \;\; i \geq 2, \tag{7.31}$$

along with the (normalizing) restriction

$$(\boldsymbol{u}_0 + \boldsymbol{v}_0)\boldsymbol{e} + \sum_{i=1}^{\infty} (\boldsymbol{u}_i + \boldsymbol{v}_i + \boldsymbol{w}_i)\boldsymbol{e} = 1. \tag{7.32}$$

The rate matrix R can be computed using a number of well-known techniques in the queueing literature. For example, with m and n to be relatively reasonable, R can be obtained using, say, logarithmic reduction (see e.g., Latouche and Ramaswami [170]). Otherwise, one should employ (block) Gauss-Siedel iteration. For this, it may be worth to exploit the special structure of the coefficient matrices A_0, A_1, and A_2, considering that these are of dimension $3mn$. For example, the structure of the matrix R as presented in the following theorem can be exploited when m and n are reasonably large.

Theorem 2: The structure of R is such that

$$R = \begin{bmatrix} R_{11} & 0 & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & 0 & R_{33} \end{bmatrix}. \tag{7.33}$$

Proof: From the probabilistic interpretation of R matrix, it is clear that R should have the form as given in (7.33). For example, noting that away from the boundary states (which is what one is interested when computing R), the (main) server when busy serving customers cannot go on vacation without visiting the boundary states. However, the (main) server can go under repair through an arrival of a shock but in this case, the server has to get back to being busy through repair completion. This results as zero in the intersection of the first (block) row and second (block) column. Similarly one sees the other zero block. However, another proof that is constructive in nature is given here.

First, equation (7.21) can be rewritten as

$$R = (R^2A_2 + A_0)(-A_1)^{-1}. \tag{7.34}$$

From the structure of A_1 (see equation (7.5)), it is easy to verify that

$$(-A_1)^{-1} = \begin{bmatrix} C_{11} & 0 & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & 0 & C_{33} \end{bmatrix}, \tag{7.35}$$

where

$$C_{11} = \left[\gamma I_{mn} - S \oplus D_0 - \xi\gamma(\xi I_{mn} - \theta S \oplus D_0)^{-1}\right]^{-1}, \;\; C_{22} = (\eta I_{mn} - \theta S \oplus D_0)^{-1}, \tag{7.36}$$

$$C_{33} = \left[\xi I_{mn} - \theta S \oplus D_0 - \xi\gamma(\gamma I_{mn} - S \oplus D_0)^{-1}\right]^{-1}, \tag{7.37}$$

$$C_{13} = \gamma(\gamma I_{mn} - S \oplus D_0)^{-1} C_{33}, \;\; C_{21} = \eta(\eta I_{mn} - \theta S \oplus D_0)^{-1} C_{11}, \tag{7.38}$$

$$C_{23} = \eta(\eta I_{mn} - \theta S \oplus D_0)^{-1} C_{13}, \;\; C_{31} = \xi(\xi I_{mn} - \theta S \oplus D_0)^{-1} C_{11}. \tag{7.39}$$

Note that pre-multiplying $(-A_1)^{-1}$ by a diagonal (block) matrix will not destroy the structure as seen in (7.35). Thus, the matrices $A_2(-A_1)^{-1}$ and $A_0(-A_1)^{-1}$ will have the same structure as $(-A_1)^{-1}$. Also, the structure is retained by matrix powers. That is, if F has that structure, then $F^n,\ n \geq 1$, also retains that structure. These are the keys to the proof of this theorem.

It is well-known (see e.g., Neuts [184]) that the sequence $\{R^{(k)}, k = 0, 1, 2, ..\}$ defined by

$$R^{(k+1)} = [(R^{(k)})^2 A_2 + A_0](-A_1)^{-1}, \;\; k = 0, 1, 2, ..., \tag{7.40}$$

with $R^{(0)} = 0$ monotonically converges to the minimal non-negative solution to (7.21). Noting that at each iterate the structure of $R^{(k)}$ remains same as that of $-(A_1)^{-1}$, the stated result is clear. □

Suppose, the state space is reordered such that the first two sets of states (of order $2mn$) correspond to the server is either serving or under repair, and the last set of states (of order mn) corresponds to the server being on vacation. Accordingly, we rearrange the entries of A_0, A_1, A_2, and R. Note that the entries of A_0 and A_2 remain the same, while R and A_1 are to be rearranged.

Suppose, the matrix R is rewritten based on the reordering of the states and if $\tilde{R}$ denotes the rate matrix of the reordered one. That is,

$$\tilde{R} = \begin{bmatrix} R_1 & 0 \\ R_2 & R_{22} \end{bmatrix}, \text{ with } R_1 = \begin{bmatrix} R_{11} & R_{13} \\ R_{31} & R_{33} \end{bmatrix}, \text{ and } R_2 = \begin{bmatrix} R_{21} & R_{23} \end{bmatrix}. \tag{7.41}$$

Similarly, if $\tilde{A}_1$ denotes the rearranged A_1, such that

$$\tilde{A}_1 = \begin{bmatrix} \tilde{A}_{11} & 0 \\ \tilde{A}_{12} & \theta S \oplus D_0 - \eta I_{mn} \end{bmatrix},$$

$$\text{with } \tilde{A}_{11} = \begin{bmatrix} S \oplus D_0 - \gamma I_{mn} & \gamma I_{mn} \\ \xi I_{mn} & \theta S \oplus D_0 - \xi I_{mn} \end{bmatrix}, \text{ and } \tilde{A}_{12} = \begin{bmatrix} \eta I_{mn} & 0 \end{bmatrix}. \tag{7.42}$$

In the following theorem, it will shown that R can be decomposed into three components. One component corresponds to the rate matrix, say, R_1, for the queueing model in which the server is busy serving at a normal rate or the server is under repair in which case the backup server is serving at a lower rate; the second component, say, R_2, corresponds to connecting R_1 and R_{22} matrices; and finally, the third component, R_{22}, satisfies the matrix-quadratic equation of the corresponding to vacation model.

Theorem 3: The rate matrix R has the following decomposition. The matrices R_1 and R_{22} are obtained as solutions to matrix-quadratic equations and R_2 is obtained explicitly in terms of R_1 and R_{22} as follows.

$$R_1^2 \tilde{A}_{21} + R_1 \tilde{A}_{11} + I_{2n} \otimes D_1 = 0,$$

$$\theta R_{22}^2 (\boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m) + R_{22}(\theta S \oplus D_0 - \eta I_{mn}) + (I_n \otimes D_1) = 0, \tag{7.43}$$

$$\tau(R_2) = -\tau(R_{22}\tilde{A}_{12}) \Big[(I_{mn} \otimes R_1 \tilde{A}_{21}) + (R_{22}' \otimes \tilde{A}_{21}) + (I_{mn} \otimes \tilde{A}_{11}) \Big]^{-1},$$

where $\tau(B)$ denotes the direct sum of the rows B (see e.g., Horn and Johnson [172]).

Proof: First note that we can rewrite (7.21) as

$$R_1^2\tilde{A}_{21} + R_1\tilde{A}_{11} + I_{2n} \otimes D_1 = 0,$$

$$(R_2R_1 + R_{22}R_2)\tilde{A}_{21} + R_2\tilde{A}_{11} + R_{22}\tilde{A}_{12} = 0, \tag{7.44}$$

$$\theta R_{22}^2(\boldsymbol{S}^0\boldsymbol{\beta} \otimes I_m) + R_{22}(\theta S \oplus D_0 - \eta I_{mn}) + (I_n \otimes D_1) = 0.$$

With the knowledge of R_1 and R_{22}, the matrix R_2 is explicitly obtained using the Sylvester matrix equation of the form $BXC + EXF + H = 0$, where only the matrix X is unknown. The stated result follows immediately by applying the result $\tau(BXC) = \tau(X)(B' \otimes C)$. □

Note: (a) It is worth pointing out that such a decomposition result for R has been reported in the past (see e.g., Alfa [265]; Chakravarthy et al. [266]).
(b) In view of the decomposition result, the computation of R may be reduced to dealing with matrices of smaller dimension. This is accomplished by first computing the submatrices R_1 and R_{22}, and then obtaining R_2 through taking advantage of the sparsity of the coefficient matrices.
The results in the following three lemmas have nice probabilistic interpretations. Further, they help to verify the accuracy in numerical computation.

Lemma 2: The following result holds

$$\boldsymbol{v}_0 + \boldsymbol{w}_0 + \Big[\sum_{i=1}^{\infty}(\boldsymbol{u}_i + \boldsymbol{v}_i + \boldsymbol{w}_i)(\boldsymbol{e} \otimes I_m)\Big] = \boldsymbol{\delta}, \tag{7.45}$$

where $\boldsymbol{\delta}$ is given in (7.1).

Proof: Post-multiply each one of the equations in (7.24 -7.31) by $\boldsymbol{e} \otimes I_m$ and verify (after some elementary manipulations) that

$$\begin{aligned}
&\boldsymbol{v}_0 D_0 + \boldsymbol{w}_0 D_0 + \boldsymbol{u}_1(\boldsymbol{S}^0 \otimes I_m) + \theta \boldsymbol{v}_1(\boldsymbol{S}^0 \otimes I_m) + \theta \boldsymbol{w}_1(\boldsymbol{S}^0 \otimes I_m) = \boldsymbol{0}, \\
&\boldsymbol{v}_0 D_1 + \boldsymbol{w}_0 D_1 - \boldsymbol{u}_1(\boldsymbol{S}^0 \otimes I_m) + \boldsymbol{u}_1(\boldsymbol{e} \otimes D_0) - \theta \boldsymbol{v}_1(\boldsymbol{S}^0 \otimes I_m) + \boldsymbol{v}_1(\boldsymbol{e} \otimes D_0) - \theta \boldsymbol{w}_1(\boldsymbol{S}^0 \otimes I_m) \\
&\quad + \boldsymbol{w}_1(\boldsymbol{e} \otimes D_0) + \boldsymbol{u}_2(\boldsymbol{S}^0 \otimes I_m) + \theta \boldsymbol{v}_2(\boldsymbol{S}^0 \otimes I_m) + \theta \boldsymbol{w}_2(\boldsymbol{S}^0 \otimes I_m) = \boldsymbol{0}, \\
&\boldsymbol{u}_{i-1}(\boldsymbol{e} \otimes D_1) + \boldsymbol{v}_{i-1}(\boldsymbol{e} \otimes D_1) + \boldsymbol{w}_{i-1}(\boldsymbol{e} \otimes D_1) - \boldsymbol{u}_i(\boldsymbol{S}^0 \otimes I_m) + \boldsymbol{u}_i(\boldsymbol{e} \otimes D_0) - \theta \boldsymbol{v}_i(\boldsymbol{S}^0 \otimes I_m) \\
&\quad + \boldsymbol{v}_i(\boldsymbol{e} \otimes D_0) - \theta \boldsymbol{w}_i(\boldsymbol{S}^0 \otimes I_m) + \boldsymbol{w}_i(\boldsymbol{e} \otimes D_0) + \boldsymbol{u}_{i+1}(\boldsymbol{S}^0 \otimes I_m) \\
&\quad + \theta \boldsymbol{v}_{i+1}(\boldsymbol{S}^0 \otimes I_m) + \theta \boldsymbol{w}_{i+1}(\boldsymbol{S}^0 \otimes I_m) = \boldsymbol{0}, \quad i \geq 2,
\end{aligned} \tag{7.46}$$

from which we immediately obtain

$$\Big[\boldsymbol{v}_0 + \boldsymbol{w}_0 + \Big(\sum_{i=1}^{\infty}(\boldsymbol{u}_i + \boldsymbol{v}_i + \boldsymbol{w}_i)(\boldsymbol{e} \otimes I_m)\Big)\Big](D_0 + D_1) = \boldsymbol{0}. \tag{7.47}$$

The above equation coupled with the uniqueness of $\boldsymbol{\delta}$ yields the claimed result. □

Lemma 3: The following result holds

$$\sum_{i=1}^{\infty}(\boldsymbol{u}_i + \theta \boldsymbol{v}_i + \theta \boldsymbol{w}_i)(\boldsymbol{S}^0 \otimes \boldsymbol{e}) = \lambda, \tag{7.48}$$

and hence

$$\sum_{i=1}^{\infty}(\boldsymbol{u}_i + \theta \boldsymbol{v}_i + \theta \boldsymbol{w}_i)(I_n \otimes \boldsymbol{e}) = \lambda \boldsymbol{\beta}(-\boldsymbol{S})^{-1}. \tag{7.49}$$

Proof: The result (7.48) is intuitively obvious since in steady-state the rate at which the customers depart the system should be equal to the arrival rate. Now, post-multiply each one of the equations in (7.24) through (7.31) by $I \otimes \boldsymbol{e}$ and verify, after some simple manipulations, that

$$\boldsymbol{v}_0 D_0 \boldsymbol{e} + \boldsymbol{w}_0 D_0 \boldsymbol{e} + \boldsymbol{u}_1(\boldsymbol{S}^0 \otimes \boldsymbol{e}) + \theta \boldsymbol{v}_1(\boldsymbol{S}^0 \otimes \boldsymbol{e}) + \theta \boldsymbol{w}_1(\boldsymbol{S}^0 \otimes \boldsymbol{e}) = 0,$$

$$\boldsymbol{v}_0(\boldsymbol{\beta} \otimes D_1\boldsymbol{e}) + \boldsymbol{w}_0(\boldsymbol{\beta} \otimes D_1\boldsymbol{e}) + \boldsymbol{u}_1(S \otimes \boldsymbol{e}) + \theta \boldsymbol{v}_1(S \otimes \boldsymbol{e}) + \theta \boldsymbol{w}_1(S \otimes \boldsymbol{e}) + \boldsymbol{u}_1(I_n \otimes D_0\boldsymbol{e}) +$$

$$\boldsymbol{v}_1(I_n \otimes D_0\boldsymbol{e}) + \boldsymbol{w}_1(I_n \otimes D_0\boldsymbol{e}) + \boldsymbol{u}_2(\boldsymbol{S}^0\boldsymbol{\beta} \otimes \boldsymbol{e}) + \theta \boldsymbol{v}_2(\boldsymbol{S}^0\boldsymbol{\beta} \otimes \boldsymbol{e}) + \theta \boldsymbol{w}_2(\boldsymbol{S}^0\boldsymbol{\beta} \otimes \boldsymbol{e}) = 0,$$

$$(\boldsymbol{u}_{i-1} + \boldsymbol{v}_{i-1} + \boldsymbol{w}_{i-1})(I_n \otimes D_1\boldsymbol{e}) + (\boldsymbol{u}_i + \theta \boldsymbol{v}_i + \theta \boldsymbol{w}_i)(S \otimes \boldsymbol{e}) + (\boldsymbol{u}_i + \boldsymbol{v}_i + \boldsymbol{w}_i)(I_n \otimes D_0\boldsymbol{e})$$

$$+ (\boldsymbol{u}_{i+1} + \theta \boldsymbol{v}_{i+1} + \theta \boldsymbol{w}_{i+1})(\boldsymbol{S}^0\boldsymbol{\beta} \otimes \boldsymbol{e}) = 0, \;\; i \geq 2. \tag{7.50}$$

Adding the above two equations and noting that

$$(\boldsymbol{v}_0 + \boldsymbol{w}_0)(\boldsymbol{\beta} \otimes D_1\boldsymbol{e}) = (\boldsymbol{u}_1 + \theta \boldsymbol{v}_1 + \theta \boldsymbol{w}_1)(\boldsymbol{S}^0\boldsymbol{\beta} \otimes \boldsymbol{e}), \tag{7.51}$$

we get

$$\sum_{i=1}^{\infty}(\boldsymbol{u}_i + \theta \boldsymbol{v}_i + \theta \boldsymbol{w}_i)(\boldsymbol{S}^0\boldsymbol{\beta} \otimes \boldsymbol{e}) = \sum_{i=1}^{\infty}(\boldsymbol{u}_i + \theta \boldsymbol{v}_i + \theta \boldsymbol{w}_i)(-S \otimes \boldsymbol{e}). \tag{7.52}$$

The claimed result in (7.49) follows by using (7.48). □

7.4 Distribution of the Effective Service

In the working-vacation-breakdown-repair queueing model under study in this chapter, a customer can be served by the main server or by the backup server or by both, depending on various scenarios. Hence, it is of interest to see the distribution of the effective service time of a customer. Noting that there is no need to track the arrival process once the service begins, the following theorem establishes an explicit result for this distribution.

Theorem 4: The distribution of the effective service time of a customer is of phase type with representation $(\boldsymbol{\alpha}, T)$ of order $3m$, where

$$\boldsymbol{\alpha} = (a_1, a_2, a_3) \otimes \boldsymbol{\beta}, \;\; T = \begin{bmatrix} S - \gamma I_m & 0 & \gamma I_m \\ \eta I_m & \theta S - \eta I_m & 0 \\ \xi I_m & 0 & \theta S - \xi I_m \end{bmatrix}, \tag{7.53}$$

where

$$a_j = \begin{cases} \frac{1}{\lambda} \sum_{i=2}^{\infty} \boldsymbol{u}_i (\boldsymbol{S}^0 \otimes \boldsymbol{e}), \; j = 1, \\ \frac{1}{\lambda} \Big[\Big(\boldsymbol{u}_1 + \theta \sum_{i=1}^{\infty} \boldsymbol{v}_i \Big) (\boldsymbol{S}^0 \otimes \boldsymbol{e}) + \theta \xi \boldsymbol{w}_1 \Big(\boldsymbol{S}^0 \otimes (\xi I_m - D_0)^{-1} \boldsymbol{e} \Big) \Big], \; j = 2, \\ \frac{\theta}{\lambda} \Big[\sum_{i=2}^{\infty} \boldsymbol{w}_i (\boldsymbol{S}^0 \otimes \boldsymbol{e}) + \boldsymbol{w}_1 \Big(\boldsymbol{S}^0 \otimes (\xi I_m - D_0)^{-1} D_1 \boldsymbol{e} \Big) \Big], \; j = 3. \end{cases} \tag{7.54}$$

Proof: First, observe that the initial probability vector, $\boldsymbol{\alpha}$, has three components, each of dimension n. The quantities, a_1, a_2, and a_3, respectively, give the probabilities that the service is initiated by the main server, by the backup server while the server is on vacation, and by the backup server while the main server is under repair. These probabilities are multiplied by $\boldsymbol{\beta}$ to initiate a service in one of n phases. Now looking at the various possibilities of a customer to begin a service the claimed result follows. □

Corollary: (a) The mean, μ_{ST}, of the effective service time of a customer is calculated as

$$\mu_{ST} = \boldsymbol{\alpha}(-T)^{-1}\boldsymbol{e}. \tag{7.55}$$

(b) The mean, μ_{MST}, time spent by a customer with the (main) server is calculated as

$$\mu_{MST} = \boldsymbol{\alpha}(-T)^{-1}(\boldsymbol{e_1}(3) \otimes \boldsymbol{e}(mn)). \tag{7.56}$$

(c) The mean, μ_{BST}, time spent by a customer with the (backup) server is given by

$$\mu_{BST} = \boldsymbol{\alpha}(-T)^{-1}(\boldsymbol{e} - \boldsymbol{e_1}(3) \otimes \boldsymbol{e}(mn)) = \mu_{ST} - \mu_{MST}. \tag{7.57}$$

(d) The fraction, f_{MS}, of time a customer is with the main server is given by

$$f_{MS} = \frac{\mu_{MST}}{\mu_{ST}}. \tag{7.58}$$

7.5 Performance Measures of the System

Qualitative analysis of any stochastic model requires establishing key system performance measures. In this section, we list a few in addition to the ones pointed out earlier.

1. P_{idle}^{System}: This is the probability of the system having no customers and is calculated as

$$P_{idle}^{System} = \boldsymbol{x}_0\boldsymbol{e}.$$

2. P_{NB} : This is the probability that the main server is busy serving and is calculated as

$$P_{NB} = \sum_{i=1}^{\infty} \boldsymbol{u}_i\boldsymbol{e}.$$

3. P_{WV} : This is the probability that the backup server is busy while the (main) server is on vacation and is obtained as

$$P_{WV} = \sum_{i=1}^{\infty} \boldsymbol{v}_i\boldsymbol{e}.$$

4. P_{WB} : This is the probability that the backup server is busy while the (main) server is under repair and is obtained as

$$P_{WB} = \sum_{i=1}^{\infty} \boldsymbol{w}_i\boldsymbol{e}.$$

5. μ_{NS} : The average number in the system is derived as

$$\mu_{NS} = \sum_{i=1}^{\infty} i(\boldsymbol{u}_i + \boldsymbol{v}_i + \boldsymbol{w}_i)\boldsymbol{e} = \boldsymbol{x}_1(I - R)^{-2}\boldsymbol{e}.$$

7.6 Special Cases

The purpose of this section is to point out a number of models in the literature that can be obtained as special cases of our model in this chapter.

7.6.1 $M/M/1$ model

For an $M/M/1$-type working-vacation-breakdown-repair queues, we set in our model: $m = n = 1, D_0 = -\lambda, D_1 = \lambda, \boldsymbol{\beta} = 1$, and $S = -\mu$. This reduces R to be a 3×3 matrix and the scalar quantity R_{22} (see the third equation in (7.44)) is the unique solution in $(0, 1)$:

$$R_{22} = \frac{\lambda + \eta + \theta\mu - \sqrt{(\theta\mu + \lambda + \eta)^2 - 4\lambda\theta\mu}}{2\theta\mu}. \tag{7.59}$$

(i) If $\eta \to \infty$

In this limiting case, there is no vacation for the server and hence η plays no role. This case reduces to the classical $M/M/1$ queue with working breakdowns studied by Kalidass and Kasturi [146].

(ii) If $\xi \to \infty$

In this limiting case, there are no breakdowns and hence no repairs, and thus both ξ and γ play no role. This case reduces to the classical $M/M/1$ queue with working vacations studied by Servi and Finn [137].

(iii) If $\theta = 0$

This case corresponds to no backup server during a vacation or a repair. Thus, the cases in (i) and (ii) above reduce, respectively, to the classical $M/M/1-$type vacation-breakdown-repair studied extensively (see e.g., Shoukry et al. [267]).

7.6.2 $MAP/M/1$

By letting $\eta \to \infty$ and $n = 1, \beta = 1, S = -\mu$, the model analyzed here reduces to the one studied by Ye and Liu [157].

7.6.3 $MAP/PH/1 : \eta \to \infty, \theta = 0$

In this case, the server never takes a vacation and also there is no backup server to offer services during the times the main server is incapable to serve. This case reduces to the classical $MAP/PH/1-$type breakdown-repair queues. It is necessary and sufficient that $\lambda < \mu$ in this case for the stability of the queue.

7.7 Decomposition Results for $M/M/1$ Case

The decomposition results for μ_{NS} in the special case of $M/M/1-$type working-vacation-breakdown-repair queues are proved here. Hereafter, $L(z)$ will stand for the probability generating function (PGF) of the number in the system. Then, using (7.20), it is easy to see that

$$L(z) = x_0 + \sum_{k=1}^{\infty} z^k \boldsymbol{x}_k \boldsymbol{e} \ = x_0 + z\boldsymbol{x}_1(I - zR)^{-1}\boldsymbol{e}, \ \ |z| < 1. \tag{7.60}$$

Case 1: *The $M/M/1$ queue with exponential (non-working) vacations*

If there is no backup server during vacation mode and no breakdown or repair, then by letting $\gamma = 0, \theta = 0$, the elements in Q (see (7.2)) are:

$$B_0 = \begin{bmatrix} 0 & \lambda \end{bmatrix}, \ B_1 = \begin{bmatrix} -\lambda \end{bmatrix}, \ B_2 = \begin{bmatrix} \mu \\ 0 \end{bmatrix}, \tag{7.61}$$

$$A_0 = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \ A_1 = \begin{bmatrix} -\lambda - \mu & 0 \\ \eta & -\lambda - \eta \end{bmatrix}, \ A_2 = \begin{bmatrix} \mu & 0 \\ 0 & 0 \end{bmatrix}. \tag{7.62}$$

When the stability condition, $\rho = \frac{\lambda}{\mu} < 1$, holds good, then R (see (7.21)) is given by

$$R = \begin{bmatrix} \frac{\lambda}{\mu} & 0 \\ \frac{\lambda}{\mu} & \frac{\lambda}{\lambda+\eta} \end{bmatrix}, \tag{7.63}$$

from which the solution to the equations (7.22) and (7.23) is obtained as

$$u_1 = \rho x_0, \quad v_1 = \frac{\lambda}{\lambda+\eta} x_0, \quad x_0 = \frac{(1-\rho)\eta}{\lambda+\eta}. \tag{7.64}$$

The special structure of R and $(I - zR)^{-1}$ for this case yields

$$\begin{aligned} L(z) &= \frac{(1-\rho)\eta}{\lambda+\eta}\Big[1 + \frac{\rho z}{1-\rho z} + \frac{\rho\lambda z^2}{(1-\rho z)(\lambda+\eta-\lambda z)} + \frac{\lambda z}{\lambda+\eta-\lambda z}\Big] \\ &= \frac{1-\rho}{1-\rho z} \cdot \frac{\eta}{\lambda+\eta-\lambda z} \\ &= L_0(z) L_d(z), \end{aligned} \tag{7.65}$$

where $L_0(z)$ is the PGF of the stationary system size in corresponding classical $M/M/1$ queue without vacation and $L_d(z)$ is the PGF of the additional system size during a vacation. It follows that μ_{NS} is decomposed as the sum of the mean number in the system in the corresponding classical $M/M/1$ queue and the mean number of arrivals during a vacation. That is,

$$\mu_{NS} = \frac{\rho}{1-\rho} + \frac{\lambda}{\eta}, \tag{7.66}$$

which is in agreement with the results obtained by Fuhrmann and Cooper [268].

Case 2: *The $M/M/1$ queue with exponential (non-working) vacations and breakdowns*

For this case, let $\theta = 0$ and verify that the (block) matrices in (7.2) are given by

$$B_0 = \begin{bmatrix} 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}, \quad B_1 = \begin{bmatrix} -\lambda & 0 \\ \xi & -\lambda-\xi \end{bmatrix}, \quad B_2 = \begin{bmatrix} \mu & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{7.67}$$

$$A_0 = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}, \quad A_1 = \begin{bmatrix} -\lambda-\mu-\gamma & 0 & \gamma \\ \eta & -\lambda-\eta & 0 \\ \xi & 0 & -\lambda-\xi \end{bmatrix}, \quad A_2 = \begin{bmatrix} \mu & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{7.68}$$

Using the facts that (a) $RA_2\boldsymbol{e} = \lambda\boldsymbol{e}$ and (b) the entries of R are scalars, the structure of R in (7.33) yields $R_{11} = R_{21} = R_{31}$. Hence, under the condition that $\lambda < \frac{\mu\xi}{\gamma+\xi}$, the entries of R are given by

$$R_{11} = R_{21} = R_{31} = \frac{\lambda}{\mu}, R_{13} = R_{23} = \frac{\lambda\gamma}{\mu(\lambda+\xi)}, R_{22} = \frac{\lambda}{\lambda+\eta}, R_{33} = \frac{\lambda(\gamma+\mu)}{\mu(\lambda+\xi)}. \tag{7.69}$$

Verify that the solution to the equations (7.22) and (7.23) is

$$w_0 = 0, \quad u_1 = \frac{\lambda}{\mu}v_0, \quad v_1 = \frac{\lambda}{\lambda+\eta}v_0, \quad w_1 = \frac{\gamma\lambda}{\mu(\lambda+\xi)}v_0, \quad v_0 = \frac{\eta(\mu\xi - \lambda\xi - \lambda\gamma)}{\mu\xi(\lambda+\eta)}. \tag{7.70}$$

Once again, the special structure $(I - zR)^{-1}$ for this case also yields the PGF in a compact form as

$$L(z) = \frac{\boldsymbol{v}_0(1 + zR_{23} - zR_{33})}{[1 - zR_{22}][1 - zR_{11}(1 + zR_{23} - zR_{33}) - zR_{33}]}. \tag{7.71}$$

Observe that μ_{NS} is calculated as

$$\mu_{NS} = \left.\frac{dL(z)}{dz}\right|_{z=1} = l_1 + l_2 + l_3 \cdot l_4, \tag{7.72}$$

where

$$l_1 = \frac{\lambda}{\eta}, \quad l_2 = \frac{\rho}{1-\rho}, \quad l_3 = \frac{\lambda}{\xi}, \quad l_4 = \frac{\rho - (\lambda/\mu)}{1-\rho}, \quad \rho = \frac{\lambda(\xi+\gamma)}{\mu\xi}. \tag{7.73}$$

It is worth pointing out that μ_{NS} is now decomposed as the sum of (i) the mean number (l_2) in the system of the corresponding classical $M/M/1$ queue; (ii) the mean number (l_1) in the system due to a vacation; and (iii) an additional quantity ($l_3 \cdot l_4$) obtained as the product of the mean number (l_3) arriving during a repair and the mean number (l_4) that arrived during a non-service time. This decomposition is in agreement with the one given by Mary et al. [269] for the single arrival case.

Case 3: *The $M/M/1$ queue with multiple working vacations*

For this case, we set $\gamma = 0,\ \xi \to \infty$. Verify the elements in (7.2) are obtained as

$$B_0 = \begin{bmatrix} 0 & \lambda \end{bmatrix}, \quad B_1 = \begin{bmatrix} -\lambda \end{bmatrix}, \quad B_2 = \begin{bmatrix} \mu \\ \theta\mu \end{bmatrix}, \tag{7.74}$$

$$A_0 = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \quad A_1 = \begin{bmatrix} -\lambda-\mu & 0 \\ \eta & -\lambda-\theta\mu-\eta \end{bmatrix}, \quad A_2 = \begin{bmatrix} \mu & 0 \\ 0 & \theta\mu \end{bmatrix}. \tag{7.75}$$

If $\rho = \frac{\lambda}{\mu} < 1$, then from Theorem 2, R (see (7.21)) is of the form

$$R = \begin{bmatrix} r_1 & 0 \\ r_2 & r_3 \end{bmatrix}, \tag{7.76}$$

where $r_1 = \rho$, $r_2 = \rho - \theta r, r_3 = r$, and r, the unique root in $(0,1)$ of the equation

$$\theta\mu r^2 - (\lambda + \theta\mu + \eta)r + \lambda = 0, \tag{7.77}$$

is obtained as

$$r = \frac{1}{2\theta\mu}\Big(\lambda + \theta\mu + \eta - \sqrt{(\lambda + \theta\mu + \eta)^2 - 4\lambda\theta\mu}\Big). \tag{7.78}$$

The solution to the equations (7.22) and (7.23), is given by

$$u_1 = r_2 x_0, \quad v_1 = r_3 x_0, \quad x_0 = \frac{(1-r_1)(1-r_3)}{1-r_1+r_2}. \tag{7.79}$$

The mean, μ_{NS}, is given by

$$\mu_{NS} = \boldsymbol{x}_1(I - R)^{-2}\boldsymbol{e} = \frac{r_1}{1 - r_1} + \frac{r_2 + r_3 - r_1}{(1 - r_3)(1 - r_1 + r_2)}, \tag{7.80}$$

which can be further simplified as

$$\mu_{NS} = \frac{\rho}{1 - \rho} + \frac{r_3(1 - \theta)}{(1 - r_3)(1 - \theta r_3)}. \tag{7.81}$$

The above shows that the decomposition is obtained as the sum of the mean number in the system of the corresponding classical $M/M/1$ queue and the mean number in the system due to the server on vacation. This again coincides with the results of Liu et al. [270].

Case 4: *The $M/M/1$ queue with (non-working) breakdowns*

By letting $\theta = 0$ and $\eta \to \infty$, the (block) matrices in (7.2) are given by

$$B_0 = \begin{bmatrix} \lambda & 0 \end{bmatrix}, \; B_1 = \begin{bmatrix} -\lambda \end{bmatrix}, \; B_2 = \begin{bmatrix} \mu \\ 0 \end{bmatrix}, \tag{7.82}$$

$$A_0 = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \; A_1 = \begin{bmatrix} -\lambda - \mu - \gamma & \gamma \\ \xi & -\lambda - \xi \end{bmatrix}, \; A_2 = \begin{bmatrix} \mu & 0 \\ 0 & 0 \end{bmatrix}. \tag{7.83}$$

Assuming the stability condition, $\lambda < \frac{\mu\xi}{\xi+\gamma}$, holds good, we have with the fact that $RA_2\boldsymbol{e} = \lambda\boldsymbol{e}$, and with Theorem 2, R is of the form

$$R = \begin{bmatrix} r_1 & r_2 \\ r_1 & r_3 \end{bmatrix}. \tag{7.84}$$

Verify that the solution (see (7.21)) yields the entries of R to be

$$r_1 = \frac{\lambda}{\mu}, r_2 = \frac{\lambda\gamma}{(\lambda + \xi)\mu}, r_3 = \frac{\lambda(\gamma + \mu)}{\mu(\lambda + \xi)}. \tag{7.85}$$

The solution to the equations (7.22) and (7.23), is then given by

$$u_1 = r_1 x_0, \quad v_1 = r_2 x_0, \quad x_0 = \frac{d}{1 + r_2 - r_3}, \tag{7.86}$$

where $d = (1 - r_1 - r_3 - r_1(r_2 - r_3)) = \dfrac{\mu\xi - \lambda\xi - \lambda\gamma}{\mu(\lambda + \xi)}$.

It is easy to verify for this case, the PGF is given by

$$L(z) = \frac{d}{1 + r_2 - r_3} \cdot \frac{1 + zr_2 - zr_3}{1 - zr_1(1 + zr_2 - zr_3) - zr_3}. \tag{7.87}$$

Thus, we have

$$\mu_{NS} = \left.\frac{dL(z)}{dz}\right|_{z=1} = \frac{\rho}{1-\rho} + \frac{\lambda}{\xi}\frac{\rho - (\lambda/\mu)}{1-\rho}, \tag{7.88}$$

where $\rho = \frac{\lambda(\xi+\gamma)}{\mu\xi}$. Note that this decomposition result agrees with the one by Kalidass and Kasturi [146].

7.8 Illustrative Numerical Examples

In this section, to understand the qualitative aspects of the working-vacation-breakdown-repair queueing model under study, we illustrate a few numerical examples. We analyze different scenarios by varying the parameters of the model including the arrival processes and service time distributions.

For the arrival process, we consider the following five sets of values for D_0 and D_1, which are considered as input data in many published works in the literature (see e.g., Chakravarthy et al. [266]; Chakravarthy [180, 271, 272]). For sake of completeness, we display them here.

ERLA: This is Erlang distribution of order 2.

$$D_0 = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}, D_1 = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}.$$

EXPA: This is the classical exponential distribution.

$$D_0 = \begin{pmatrix} -1 \end{pmatrix}, D_1 = \begin{pmatrix} 1 \end{pmatrix}.$$

HEXA: This is the mixture of two exponentials.

$$D_0 = \begin{pmatrix} -1.90 & 0 \\ 0 & -0.19 \end{pmatrix}, D_1 = \begin{pmatrix} 1.71 & 0.19 \\ 0.171 & 0.019 \end{pmatrix}.$$

MNA: This one has a negative correlation (with a value of -0.4889) between two successive inter-arrival times.

$$D_0 = \begin{pmatrix} -1.00222 & 1.00222 & 0 \\ 0 & -1.00222 & 0 \\ 0 & 0 & -225.75 \end{pmatrix}, D_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0.01002 & 0 & 0.9922 \\ 223.4925 & 0 & 2.2575 \end{pmatrix}.$$

MPA: This one has a positive correlation (with a value of 0.4889) between two successive inter-arrival times.

$$D_0 = \begin{pmatrix} -1.00222 & 1.00222 & 0 \\ 0 & -1.00222 & 0 \\ 0 & 0 & -225.75 \end{pmatrix}, D_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0.9922 & 0 & 0.01002 \\ 2.2575 & 0 & 223.4925 \end{pmatrix}.$$

Note that $ERLA$, $EXPA$, and $HEXA$, correspond to renewal processes. All five arrival processes are qualitatively different. Verify that the ratio of the standard deviations of the times between two successive inter-arrival times of these five processes with respect to the Erlang one are, respectively, 1, 1.4142, 3.1745, 1.9933, and 1.9933.

For service times with representation $(\boldsymbol{\beta}, S)$, the following three phase type distributions are considered. Normalization of these representations will be done to get a desired μ.

ERLS : This is Erlang of order 2.

$$\boldsymbol{\beta} = (1, 0), \quad S = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}.$$

EXPS : This is an exponential one.

$$\boldsymbol{\beta} = 1, \quad S = (-1).$$

HEXS : This is the mixture of two exponentials.

$$\boldsymbol{\beta} = (0.9, 0.1), \quad S = \begin{pmatrix} -10 & 0 \\ 0 & -1 \end{pmatrix}.$$

Irrespective of how these are normalized, the standard deviation of $EXPS$ and $HEXS$ are, respectively, about 1.4142 and 3.1745 of that of $ERLS$.

Next, we define

$$\rho = \frac{\lambda(\xi + \gamma)}{\mu(\xi + \theta\gamma)}. \tag{7.89}$$

Example 1: The effect of the vacation rate, η, on two performance measures, f_{MS} and $P_{BB} = P_{WB} + P_{WV}$ is of focus here. Recall that f_{MS} gives the fraction of time a customer is served by the main server, and P_{BB} is the probability that the backup server is busy serving. We fix $\lambda = 1, \mu = 1.372, \theta = 0.8, \xi = 0.5, \gamma = 0.4$ and vary η. In Fig. 7.2, we display the two selected measures under various scenarios. Analyzing the graphs in the figure, we notice the following.

- An increase in η causes an increase in f_{MS}, which is to be expected. However, the rate of increase is higher for small values of η. The type of arrival processes and the service times distribution play a role in the rate of increase.

- Observe that among the renewal arrivals, the $HEXA$ process produces the highest