Applied Mathematics and Computation 213 (2009) 92–101

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Controlling arrivals for a queueing system with an unreliable server: Newton-Quasi method

Kuo-Hsiung Wang [a,*], Dong-Yuh Yang [b]

[a] Department of Applied Mathematics, National Chung-Hsing University, Taichung 402, Taiwan
[b] Department of Industrial of Engineering and Management, National Chiao Tung University, Hsinchu 30050, Taiwan

ARTICLE INFO

Keywords:
F-policy
Matrix analytical method
Optimization
Newton-Quasi method
Startup
Server breakdowns

ABSTRACT

This paper deals with the control policy of a removable and unreliable server for an M/M/1/K queueing system, where the removable server operates an *F*-policy. The so-called *F*-policy means that when the number of customers in the system reaches its capacity *K* (i.e. the system becomes full), the system will not accept any incoming customers until the queue length decreases to a certain threshold value *F*. At that time, the server initiates an exponential startup time with parameter γ and starts allowing customers entering the system. It is assumed that the server breaks down according to a Poisson process and the repair time has an exponential distribution. A matrix analytical method is applied to derive the steady-state probabilities through which various system performance measures can be obtained. A cost model is constructed to determine the optimal values, say (F^*, μ^*, γ^*), that yield the minimum cost. Finally, we use the two methods, namely, the direct search method and the Newton-Quasi method to find the global minimum (F^*, μ^*, γ^*). Numerical results are also provided under optimal operating conditions.

1. Introduction

One of the most investigated topics of queueing problem is the control of queue. The aim of controllable queueing models is to find the optimal operating policy which enables a decision-maker to turn the server on or off at a minimum cost. A comprehensive review on the controllable queues can be found in Tadj and Choudhury [12]. Past work regarding controllable queues may be divided into two parts, one aims at controlling service whereas the other aims at controlling arrivals. In the controlling service category, the *N*-policy M/M/1 queueing system without startup was first introduced by Yadin and Naor [22]. Tadj [11] applied the matrix analytic methodology to study an *r*-quorum queueing system under *N*-policy discipline. The *N*-policy M/M/1 queueing system with exponential startup times was first proposed by Baker [1]. Borthakur et al. [2] extended Baker's model to general startup time. Queueing models with an unreliable server closely resemble to practical situations. The server may meet unpredicted breakdowns while providing service. Considering queueing systems with server breakdowns, Wang [14,15], Wang et al. [16] derived analytic steady-state solutions of the *N*-policy M/M/1, the *N*-policy M/E_k/1 and *N*-policy M/H_2/1 queueing systems, respectively. Moreover, Wang et al. [17] generalized Wang et al. [16] to the *N*-policy M/H_K/1 queueing system. Ke and Pearn [7] developed the closed-form solutions for the *N*-policy M/M/1 queue with server breakdowns and multiple vacations. The expected number of customers in the controllable M/G/1 queueing system with server breakdowns operating under three control policies were developed by Wang and Ke [18]. In some actual situations, the server often requires a startup time before starting the service. As for the *N*-policy M/G/1 queue with startup

* Corresponding author.
E-mail address: khwang@amath.nchu.edu.tw (K.-H. Wang).

0096-3003/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2009.03.002

times, it was investigated by numerous authors, including Lee and Park [8], Medhi and Templeton [9], Takagi [13], etc. Lately, Ke [6] examined some important system characteristics for the *N*-policy M/G/1 queue with server vacations, startup and breakdowns. Wang et al. [21] developed the approximate probability distribution of the queue length for the *N*-policy M/G/1 queue with server breakdown and startup by using the maximum entropy approach. In the controlling arrivals category, Gupta and Melachrinoudis [4] derived complementary relationships between *N*-policy and *F*-policy in finite source queueing models with spares. Through a series of propositions, the relationship between the *N*-policy and the *F*-policy is established by Gupta [3]. Karaesmen and Gupta [5] applied the duality relationship to obtain the stationary queue length distributions for the two queueing systems under *N*-policy and *F*-policy. Recently, Wang et al. [19,20] investigated the optimal management problem of an M/G/1/K and G/M/1//K queueing systems with combined *F*-policy and an exponential startup time, respectively.

On the whole, controlling arrivals queueing systems with server breakdowns have seldom been investigated by existing research works. In this paper, we deal with a single unreliable server in an M/M/1/K queueing system with combined *F*-policy and an exponential startup time. The *F*-policy addresses the issue of controlling arrivals in a queueing system. The policy of controlling arrivals focuses on reducing the number of customers in the system. There are many real-life situations which can be fit into this model

There are three primary objectives in this paper. Firstly, to present a matrix analytical method for developing the steady-state solutions for the *F*-policy M/M/1/K queueing system with server breakdowns and an exponential startup time. Secondly, we develop the cost model to determine the joint optimal values of F, μ and γ that will yield the minimum cost. Thirdly, we use the two methods, namely, the direct search method and the Newton-Quasi method to find the global minimum.

2. Description of system

Gupta [3] first introduced the concept of an *F*-policy. The definition of an *F*-policy is described as follows: when the number of customers in the system reaches its capacity K (i.e. the system becomes full), no further arriving customers are allowed to enter the system until there are enough customers in the system being served so that the number of customers in the system decreases to a threshold value F ($0 \leqslant F \leqslant K-1$). At that time, the server is required to take an exponential startup time with parameter γ to start allowing customers entering the system. Thus, the system operates normally until the number of customers in the system reaches its capacity at which time the above process is repeated all over again.

In this paper, we consider controlling the arrivals to a finite capacity M/M/1/K queueing system under *F*-policy subject to server breakdowns and an exponential startup time. It is assumed that customers arrive according to a Poisson process with parameter λ and the service times of the customers are independent random variables having a common exponential distribution with mean $1/\mu$. Assume that the server may encounter break down at any time with breakdown rate α. Whenever the server breaks down, it is immediately repaired at a repair rate β. Breakdown and repair time distributions of the server are assumed to be exponentially distributed. Customers arriving at the server are assumed to form a single waiting line and are served in the order of their arrivals; that is, according to the first-come, first-served (FCFS) discipline. Suppose the server can serve only one customer at a time, and that the service is independent of the arrival of the customers. Customers entering into the service facility and finding that the server is busy have to wait in the queue until the server is available.

3. Steady-state solutions

For the *F*-policy M/M/1/K queueing system with server breakdowns and an exponential startup time, we define the following notations that will be used in the subsequent derivation:

$N(t) \equiv$ the number of customers in the system at time t,
$Y(t) \equiv$ the server state at time t,

where

$$Y(t) = \begin{cases} 0 & \text{if the arrivals are not allowed to enter the system and the server is broken down,} \\ 1 & \text{if the arrivals are not allowed to enter the system and the server is busy,} \\ 2 & \text{if the arrivals are allowed to enter the system and the server is busy,} \\ 3 & \text{if the arrivals are allowed to enter the system and the server is broken down.} \end{cases}$$

Then $\{Y(t), N(t);\ t \geqslant 0\}$ is a continuous time Markov process on state space

$$S = \{(0,n)|n = 1,2,\ldots,K\} \cup \{(1,n)|n = 0,1,2,\ldots,K\} \cup \{(2,n)|n = 0,1,2,\ldots,K-1\} \cup \{(3,n)|n = 1,2,\ldots,K-1\}.$$

We define

$$P_{0,n}(t) = \Pr\{Y(t) = 0, N(t) = n\},$$
$$P_{1,n}(t) = \Pr\{Y(t) = 1, N(t) = n\},$$
$$P_{2,n}(t) = \Pr\{Y(t) = 2, N(t) = n\},$$

and

$$P_{3,n}(t) = \Pr\{Y(t) = 3, N(t) = n\}.$$

Furthermore, we let

$$P_{i,n} = \lim_{t\to\infty} P_{i,n}(t), \quad i = 0, 1, 2, 3.$$

The derivation of the following steady-state equations for a finite capacity M/M/1/K queueing system combined with *F*-policy, server breakdowns and an exponential startup time are given in the Appendix.

3.1. Steady-state equations

$$\beta P_{0,n} = \alpha P_{1,n}, \quad 1 \leqslant n \leqslant K, \tag{1}$$
$$\gamma P_{1,0} = \mu P_{1,1}, \tag{2}$$
$$(\mu + \alpha + \gamma)P_{1,n} = \mu P_{1,n+1} + \beta P_{0,n}, \quad 1 \leqslant n \leqslant F, \tag{3}$$
$$(\mu + \alpha)P_{1,n} = \mu P_{1,n+1} + \beta P_{0,n}, \quad F + 1 \leqslant n \leqslant K - 1, \tag{4}$$
$$(\mu + \alpha)P_{1,K} = \lambda P_{2,K-1} + \beta P_{0,K}, \tag{5}$$
$$\lambda P_{2,0} = \gamma P_{1,0} + \mu P_{2,1}, \tag{6}$$
$$(\lambda + \alpha + \mu)P_{2,n} = \gamma P_{1,n} + \lambda P_{2,n-1} + \mu P_{2,n+1} + \beta P_{3,n}, \quad 1 \leqslant n \leqslant F, \tag{7}$$
$$(\lambda + \alpha + \mu)P_{2,n} = \lambda P_{2,n-1} + \mu P_{2,n+1} + \beta P_{3,n}, \quad F + 1 \leqslant n \leqslant K - 2, \tag{8}$$
$$(\lambda + \alpha + \mu)P_{2,K-1} = \lambda P_{2,K-2} + \beta P_{3,K-1}, \quad K - 1 = n \neq F, \tag{9}$$
$$(\lambda + \beta)P_{3,1} = \alpha P_{2,1}, \tag{10}$$
$$(\lambda + \beta)P_{3,n} = \alpha P_{2,n} + \lambda P_{3,n-1}, \quad 2 \leqslant n \leqslant K - 2, \tag{11}$$
$$\beta P_{3,K-1} = \alpha P_{2,K-1} + \lambda P_{3,K-2}. \tag{12}$$

3.2. Matrix analytical solutions

Matrix analytical method is a useful tool for constructing stochastic models. This method is widely used to deal with various queueing models that desires exact analysis. The matrix analytical method is proposed by Neuts [10] for analyzing the embedded Markov chains of many practical queueing systems. By using the matrix analytical method, we develop the steady-state probabilities $P_{0,n}$ $(1 \leqslant n \leqslant K)$, $P_{1,n}$ $(0 \leqslant n \leqslant K)$, $P_{2,n}(0 \leqslant n \leqslant K - 1)$, $P_{3,n}(1 \leqslant n \leqslant K - 1)$. The corresponding transition rate matrix $\mathbf{Q}$ of this Markov chain has the block-tridiagonal form:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A_0} & \mathbf{B_0} & \mathbf{0} & \mathbf{0} \\ \mathbf{C_1} & \mathbf{A_1} & \mathbf{B_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_2} & \mathbf{A_2} & \mathbf{B_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{C_3} & \mathbf{A_3} \end{bmatrix}.$$

The rate matrix $\mathbf{Q}$ of this state process is similar to the quasi birth and death type, and this class of Markov process has been extensively studied by Neuts [10]. Each element of the matrix $\mathbf{Q}$ is a square matrix of order $(K + 1)$ which takes the following form:

$$\mathbf{A}_j = \begin{bmatrix} x_{j,0} & y_{j,0} & 0 & 0 & \cdots & 0 & 0 & 0 \\ z_{j,1} & x_{j,1} & y_{j,1} & 0 & \cdots & 0 & 0 & 0 \\ 0 & z_{j,2} & x_{j,2} & y_{j,2} & \cdots & 0 & 0 & 0 \\ 0 & 0 & z_{j,3} & x_{j,3} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & x_{j,K-2} & y_{j,K-2} & 0 \\ 0 & 0 & 0 & 0 & \cdots & z_{j,K-1} & x_{j,K-1} & y_{j,K-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & z_{j,K} & x_{j,K} \end{bmatrix} \quad \text{for } j = 0, 1, 2, 3.$$

Before representing the elements of $\mathbf{A}_j$ ($j = 0,1,2,3$), we need to define an indicator function $\Lambda_{[a,b]}(i)$

$$\Lambda_{[a,b]}(i) = \begin{cases} 1, & \text{if } a \leqslant i \leqslant b, \\ 0, & \text{o.w.} \end{cases}$$

Consequently, $x_{j,i}$, $y_{j,i}$ and $z_{j,i}$ can be represented as:

$$x_{j,i} = \begin{cases} \phi\Lambda_{[0,0]}(i) + \beta\Lambda_{[1,K]}(i) & \text{if } j = 0,\ 0 \leqslant i \leqslant K, \\ \gamma\Lambda_{[0,0]}(i) + (\gamma+\mu+\alpha)\Lambda_{[1,F]}(i) + (\mu+\alpha)\Lambda_{[F+1,K]}(i), & \text{if } j = 1,\ 0 \leqslant i \leqslant K, \\ \lambda\Lambda_{[0,0]}(i) + (\lambda+\alpha+\mu)\Lambda_{[1,K-1]}(i) + \phi\Lambda_{[K,K]}(i), & \text{if } j = 2,\ 0 \leqslant i \leqslant K, \\ \phi\Lambda_{[0,0]}(i) + (\lambda+\beta)\Lambda_{[1,K-2]}(i) + \beta\Lambda_{[K-1,K-1]}(i) + \phi\Lambda_{[K,K]}(i), & \text{if } j = 3,\ 0 \leqslant i \leqslant K, \end{cases}$$

$$y_{j,i} = \begin{cases} 0, & \text{if } j = 0,1,\ 0 \leqslant i \leqslant K-1, \\ -\lambda\Lambda_{[0,K-2]}(i), & \text{if } j = 2,\ 0 \leqslant i \leqslant K-1, \\ -\lambda\Lambda_{[1,K-2]}(i), & \text{if } j = 3,\ 0 \leqslant i \leqslant K-1, \end{cases}$$

and

$$z_{j,i} = \begin{cases} 0, & \text{if } j = 0,3,\ 1 \leqslant i \leqslant K, \\ -\mu\Lambda_{[1,K]}(i), & \text{if } j = 1,\ 1 \leqslant i \leqslant K, \\ -\mu\Lambda_{[1,K-1]}(i), & \text{if } j = 2,\ 1 \leqslant i \leqslant K, \end{cases}$$

$$\mathbf{B}_j = \begin{bmatrix} b_{j,0} & 0 & 0 & \cdots & 0 & 0 \\ 0 & b_{j,1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & b_{j,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{j,K-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & b_{j,K} \end{bmatrix}, \quad j = 0,1,2$$

are $(K+1)$ square matrices. For $0 \leqslant i \leqslant K$

$$b_{j,i} = \begin{cases} \phi\Lambda_{[0,0]}(i) - \beta\Lambda_{[1,K]}(i), & \text{if } j = 0, \\ -\gamma\Lambda_{[0,F]}(i), & \text{if } j = 1, \\ -\alpha\Lambda_{[1,K-1]}(i) + \phi\Lambda_{[K,K]}(i), & \text{if } j = 2, \end{cases}$$

$$\mathbf{C}_j = \begin{bmatrix} c_{j,0} & d_{j,0} & 0 & \cdots & 0 & 0 \\ 0 & c_{j,1} & d_{j,1} & \cdots & 0 & 0 \\ 0 & 0 & c_{j,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_{j,K-1} & d_{j,K-1} \\ 0 & 0 & 0 & \cdots & 0 & c_{j,K} \end{bmatrix}, \quad j = 1,2,3$$

are $(K+1)$ square matrices and

$$c_{j,i} = \begin{cases} -\alpha\Lambda_{[1,K]}(i), & \text{if } j = 1, \\ \phi\Lambda_{[K,K]}(i), & \text{if } j = 2, \\ \phi\Lambda_{[0,0]}(i) - \beta\Lambda_{[1,K-1]}(i) + \phi\Lambda_{[K,K]}(i), & \text{if } j = 3, \end{cases}$$

$$d_{j,i} = \begin{cases} -\lambda, & \text{if } j = 2, i = K-1, \\ 0, & \text{o.w.} \end{cases}$$

It is noted that ϕ is a non-zero real number in the above expressions. Let $\mathbf{\Pi}$ be the corresponding steady-state probability vector of $\mathbf{Q}$. By partitioning the vector $\mathbf{\Pi}$ as $\mathbf{\Pi} = \{\Pi_0, \Pi_1, \Pi_2, \Pi_3\}$, where $\Pi_j = \{P_{j,0}, P_{j,1}, P_{j,2}, \ldots, P_{j,K-1}, P_{j,K}\}$ is an $1 \times (K+1)$ row vector for $j = 0,1,2,3$, where $P_{0,0}$, $P_{2,K}$, $P_{3,0}$ and $P_{3,K}$ are equal to zero. By solving the steady-state equations $\mathbf{\Pi Q} = \mathbf{0}$, we obtain

$$\Pi_0 A_0 + \Pi_1 C_1 = 0, \tag{13}$$

$$\Pi_0 B_0 + \Pi_1 A_1 + \Pi_2 C_2 = 0, \tag{14}$$

$$\Pi_1 B_1 + \Pi_2 A_2 + \Pi_3 C_3 = 0, \tag{15}$$

$$\Pi_2 B_2 + \Pi_3 A_3 = 0. \tag{16}$$

Thus after routine substitutions, we get

$$\Pi_3 = -\Pi_2 B_2 A_3^{-1}, \tag{17}$$

$$\Pi_2 = -\Pi_1 B_1 (A_2 - B_2 A_3^{-1} C_3)^{-1}, \tag{18}$$

$$\Pi_1 = -\Pi_0 B_0 \left[A_1 - B_1 (A_2 - B_2 A_3^{-1} C_3)^{-1} C_2 \right]^{-1}, \tag{19}$$

$$\Pi_0 (A_0 + \Psi_1 C_1) = 0. \tag{20}$$

If we set $\Psi_j = -B_{j-1}(A_j + \Psi_{j+1} C_{j+1})^{-1}$ $(j = 1,2)$, Π_1 and Π_2 can be simplified to $\Pi_1 = \Pi_0 \Psi_1$ and $\Pi_2 = \Pi_1 \Psi_2$, respectively. In addition, one can observe that $\Psi_3 = -B_2 A_3^{-1}$. Eq. (20) determines Π_0 up to a multiplicative constant. The other Eqs. (17)–(19) determine Π_1, Π_2 and Π_3, up to the same constant, which is uniquely determined by the following normalizing equation

$$\sum_{j=0}^{3} \Pi_j \mathbf{e} = 1, \tag{21}$$

where $\mathbf{e}$ represents a column vector with each component equal to one. An efficient computer program was developed to solve Π_j and $P_{j,n}$ for $0 \leqslant j \leqslant 3$. If we set $\alpha = 0$, the server in this queueing model is reliable. It is worth to mention that the results coincided with those in Gupta [3, p. 1007].

4. Cost function structure

In this section, we first obtain some system performance measures from steady-state probabilities. Based on these system performance measures, we derive the total expected cost function per unit time.

4.1. System performance measures

Some important system performance measures of the *F*-policy M/M/1/K queueing system with combined server breakdowns and an exponential startup time are defined as follows:

$L_S \equiv$ the expected number of customer in the system;
$P_B \equiv$ the probability that the server is busy;
$P_I \equiv$ the probability that the system is idle;
$P_D \equiv$ the probability that the server is broken down.

The expressions for L_S, P_B, P_I, and P_D are given by:

$$L_S = \sum_{n=1}^{K} nP_{0,n} + \sum_{n=1}^{K} nP_{1,n} + \sum_{n=1}^{K-1} nP_{2,n} + \sum_{n=1}^{K-1} nP_{3,n}, \tag{22}$$

$$P_B = \sum_{n=1}^{K} P_{1,n} + \sum_{n=1}^{K-1} P_{2,n}, \tag{23}$$

$$P_I = \sum_{j=1}^{2} P_{j,0}, \tag{24}$$

$$P_D = \sum_{n=1}^{K} P_{0,n} + \sum_{n=1}^{K-1} P_{3,n}. \tag{25}$$

4.2. Total expected cost function

We develop an expected cost function per unit time for the *F*-policy M/M/1/K queue with server breakdowns and an exponential startup time. In this cost model, three decision variables F, μ and γ are considered. The discrete variable F is required to be a natural number. The continuous variables μ and γ are positive numbers. Let us define the following cost elements:

$C_h \equiv$ holding cost per unit time for each customer present in the system;
$C_b \equiv$ busy cost per unit time for a busy server;
$C_d \equiv$ breakdown cost per unit time for a failed server;
$C_i \equiv$ idle cost per unit time for an idle server;
$C_\mu \equiv$ fixed cost of providing a service rate μ for customers;
$C_\gamma \equiv$ fixed cost of providing a startup rate γ when allowing customer to enter the system at each time.

Based on the definitions of each cost element listed above and its corresponding system characteristics, the total expected cost function per unit time is given by

$$TC(F,\mu,\gamma) = C_h L_S + C_b P_B + C_i P_I + C_d P_D + C_u \mu + C_\gamma \gamma. \tag{26}$$

The cost parameters in Eq. (26) are assumed to be linear in the expected number of the identical quantity. Due to the highly non-linear and complex nature of the optimization problem, it is extremely difficult to develop analytic results for the optimum value (F^*,μ^*,γ^*). In the next section, we will utilize the direct search method to find the optimal threshold F, say F^*, when μ and γ are fixed. Next, we fix F^* and use the Newton-Quasi method to find the optimal value of (μ,γ), say (μ^*,γ^*).

5. Cost minimization

The aim of this section is to designate optimal values for this queueing system. In the cost function (26), three variables F, μ and γ are considered. The discrete variable F is required to be a natural number. The continuous variables μ and γ are positive numbers. Our objective is to determine the optimal value of (F,μ,γ), say (F^*,μ^*,γ^*), so as to minimize this function. In the following subsections, we use the direct search method and the Newton-Quasi method to find three optimum values F, μ and γ that will minimize the cost function. Some numerical results are also provided for the purpose of illustration.

5.1. Direct search method

By using Eqs. (22)–(25), explicit expression for formula (26) can be obtained. However, it is tediously complex. Therefore, we perform numerical computations to demonstrate that the cost function is indeed convex, so the solution obtained gives a global minimum. By exploiting the property of F, we first use direct substitution of successive values of F into the cost function. Numerical results are provided by considering the following cost parameters:

$$C_h = \$5, \quad C_b = \$100, \quad C_d = \$300, \quad C_i = \$200, \quad C_\mu = \$3, \quad C_\gamma = \$1.$$

The cost minimization problem can be illustrated mathematically as:

$$TC(F^*,\mu,\gamma) = \underset{F}{Minimize} \quad TC(F,\mu,\gamma). \tag{27}$$

Since F is a discrete variable, we employ the following inequalities to find F^*:

$$TC(F^*-1,\mu,\gamma) \geqslant TC(F^*,\mu,\gamma),$$

and

$$TC(F^*+1,\mu,\gamma) \geqslant TC(F^*,\mu,\gamma).$$

We fix the number of system capacity $K = 15$, $(\mu,\gamma) = (1.5, 0.5)$, $\alpha = 0.05$, $\beta = 3.0$, vary the value of threshold F from 0 to 14, and choose different values of $\lambda = 0.6, 0.8, 1.0, 1.2, 1.4$. From Table 1, we observe that (i) the optimal threshold value F^* decreases as λ increases; (ii) $TC(F^*,\mu,\gamma)$ decreases as λ increases. Next, we fix $K = 15$, $\lambda = 1.2$, $(\mu,\gamma) = (1.5, 0.5)$, vary the value of threshold F from 0 to 14, and choose different values of $(\alpha,\beta) = (0.01, 1.0), (0.05, 1.0), (0.10, 1.0), (0.10, 2.0), (0.10, 6.0)$. The results in Table 2 make it obvious that (i) the optimal threshold value F^* decreases as α increases or β decreases; (ii) $TC(F^*,\mu,\gamma)$ increases as α increases or β decreases.

Moreover, we fix $K = 15$, $\lambda = 1.4$, $\alpha = 0.05$, $\beta = 3.0$, vary the value of threshold F from 0 to 14, and choose $(\mu,\gamma) = (1.5, 0.1), (1.8, 0.1), (2.0, 0.1), (2.0, 0.5), (2.0, 1.0)$. One can see from Table 3 that (i) the optimal threshold value F^* increases as μ increases or γ decreases; (ii) $TC(F^*,\mu,\gamma)$ increases as μ increases or γ increases. Numerical results of the minimum expected cost $TC(F^*,\mu,\gamma)$ and the system performance measures L_s, P_I, P_B and P_D, at the optimal threshold value F^* are also listed in Tables 1–3.

Table 1
System performance measures of the F-policy M/M/1/K queueing system with server breakdowns and an exponential startup time under optimal operation conditions ($K = 15$, $\mu = 1.5$, $\gamma = 0.5$, $\alpha = 0.05$, $\beta = 3.0$).

λ	0.6	0.8	1.0	1.2	1.4
F^*	14	11	8	6	5
$TC(F^*,\mu,\gamma)$	169.105	158.500	149.855	144.928	144.244
L_S	0.688	1.187	2.039	3.352	4.792
P_B	0.007	0.533	0.665	0.781	0.862
P_I	0.593	0.458	0.324	0.206	0.124
P_D	0.400	0.009	0.011	0.013	0.014

Table 2
System performance measures of the *F*-policy M/M/1/K queueing system with server breakdowns and an exponential startup time under optimal operation conditions ($K = 15$, $\lambda = 1.2$, $\mu = 1.5$, $\gamma = 0.5$).

(α,β)	(0.01,1.0)	(0.05,1.0)	(0.1,1.0)	(0.1,2.0)	(0.1,6.0)
F^*	6	5	4	5	6
$TC(F^*,\mu,\gamma)$	144.054	150.049	157.500	149.813	144.892
L_S	3.296	3.664	4.088	3.64728	3.348
P_B	0.782	0.771	0.755	0.773	0.782
P_I	0.210	0.190	0.169	0.188	0.205
P_D	0.008	0.039	0.076	0.039	0.013

Table 3
System performance measures of the *F*-policy M/M/1/K queueing system with server breakdowns and an exponential startup time under optimal operation conditions ($K = 15$, $\lambda = 1.4$, $\alpha = 0.05$, $\beta = 3.0$).

(μ,γ)	(1.5,0.1)	(1.8,0.1)	(2.0,0.1)	(2.0,0.5)	(2.0,1.0)
F^*	11	13	14	8	7
$TC(F^*,\mu,\gamma)$	146.673	146.847	149.684	149.740	150.213
L_S	4.655	3.074	2.314	2.331	2.333
P_B	0.826	0.753	0.691	0.695	0.696
P_I	0.160	0.234	0.297	0.293	0.292
P_D	0.014	0.013	0.012	0.012	0.012

5.2. Newton-Quasi method

We fix F^* and use the Newton-Quasi method to globally search (μ,γ) until the minimum value of $TC(F^*,\mu,\gamma)$, say $TC(F^*,\mu^*,\gamma^*)$ is attained. The cost minimization problem can be illustrated mathematically as

$$TC(F^*,\mu^*,\gamma^*) = \underset{\mu,\gamma}{Minimize} \quad TC(F^*,\mu,\gamma). \tag{28}$$

The essence of the Newton-Quasi method is to find a search direction in each iteration. Then try different step length along this direction for a better solution until the tolerance is small enough. We designate the vector $\vec{\Omega}$ consisting of μ and γ. We construct the respective gradient $\vec{\nabla}TC(\vec{\Omega}_0)$, which consists of $\partial TC/\partial\mu$ and $\partial\ TC/\partial\gamma$. Next, we use the Newton-Quasi method to find the global minimum expected cost. Let the corresponding solution be denoted by (μ^*,γ^*). The steps of the Newton-Quasi method can be summarized as follows:

1. Let $\vec{\Omega}_0 = [\mu,\gamma]^T$.
2. Set the initial trial solution for $\vec{\Omega}_0$, and compute $TC(\vec{\Omega}_0)$.
3. Compute the cost gradient $\vec{\nabla}TC(\vec{\Omega}_0) = [\partial TC/\partial\mu, \partial TC/\partial\gamma]^T|_{\vec{\Omega}_0}$ and the cost Hessian matrix

$$H(\vec{\Omega}) = \begin{bmatrix} \partial^2 TC/\partial\mu^2 & \partial^2 TC/\partial\mu\partial\gamma \\ \partial^2 TC/\partial\gamma\partial\mu & \partial^2 TC/\partial\gamma^2 \end{bmatrix}.$$

4. Find the new trial solution $\vec{\Omega}_{n+1} = \vec{\Omega}_n - [H(\vec{\Omega})]^{-1}\vec{\nabla}TC(\vec{\Omega}_n)$.
5. Set $n = n + 1$ and repeat steps 2–4 until $|\partial TC/\partial\mu| < \varepsilon_1$ and $|\partial TC/\partial\gamma| < \varepsilon_2$, where $\varepsilon_1 = \varepsilon_2 = 10^{-6}$ are the tolerances.
6. Find the global minimum value $TC(\vec{\Omega}_n^T) = TC(\mu^*,\gamma^*)$.

Table 4
Newton-Quasi method in searching the optimal solution ($\lambda = 1.2$, $\alpha = 0.05$, $\beta = 3.0$).

Number of iterations	$TC(F^*,\mu,\gamma)$	F^*	μ	γ	Max. tolerance
0	144.9280850	6	1.50000000	0.50000000	–
1	143.6933638	6	1.27089080	0.30859284	13.680
2	143.5658694	6	1.31542201	0.36403603	3.430
3	143.5570658	6	1.31318749	0.39981980	0.443
4	143.5568332	6	1.31340494	0.40684769	0.064
5	143.5568333	6	1.31341033	0.40708021	1.989×10^{-3}
6	143.5568333	6	1.31341033	0.40708039	1.549×10^{-5}
7	143.5568328	6	1.31341034	0.40708033	5.14×10^{-7}

Max. tolerance = Max{$|\partial(TC)/\partial\mu|, |\partial(TC)/\partial\gamma|$}.

Table 5
Newton-Quasi method in searching the optimal solution ($\lambda = 1.2$, $\alpha = 0.1$, $\beta = 2.0$).

Number of iterations	$TC(F^*,\mu,\gamma)$	F^*	μ	γ	Max. tolerance
0	149.8131619	5	1.50000000	0.500000000	–
1	149.2752325	5	1.36276454	0.383504737	8.212
2	149.2617250	5	1.37244177	0.423341664	0.741
3	149.2613561	5	1.37236902	0.433307170	0.072
4	149.2613557	5	1.37237423	0.433715952	0.003
5	149.2613559	5	1.37237424	0.433716671	4.83×10^{-6}
6	149.2613555	5	1.37237422	0.433716533	1.708×10^{-6}
7	149.2613556	5	1.37237424	0.433716552	1.655×10^{-6}
8	149.2613556	5	1.37237425	0.433716641	6.050×10^{-7}

Max. tolerance = Max$\{|\partial(TC|)/\partial\mu|, |\partial(TC)/\partial\ \gamma|\}$.

Using the results shown in the second last column of Table 1, we select $\lambda = 1.2$, $\alpha = 0.05$, $\beta = 3.0$ and the initial trial solution $(F^*,\mu,\gamma) = (6, 1.5, 0.5)$ with initial value $TC(F^*,\mu,\gamma) = 144.928$. We apply the algorithm of the Newton-Quasi method as mentioned above. After seven iterations, Table 4 clearly shows that the minimum expected cost converges to the solution $(F^*,\mu^*,\gamma^*) = (6, 1.31341034, 0.40708033)$ with value 143.5568328. Next, we utilize the results of Table 2, we select $\lambda = 1.2$, $\alpha = 0.1$, $\beta = 2.0$ and the initial trial solution $(F^*,\mu,\gamma) = (5, 1.5, 0.5)$ with initial value $TC(F^*,\mu,\gamma)$=149.813. The algorithm of the Newton-Quasi method is used. The numerical results after eight iterations are shown in Table 5. The minimum expected cost converges to the solution $(F^*,\mu^*,\gamma^*) = (5, 1.37237425, 0.433716641)$ with value 149.2613556.

Finally, we employ the results of Table 3, that is, we select $\lambda = 1.4$, $\alpha = 0.05$, $\beta = 3.0$ and the initial trial solution $(F^*,\mu,\gamma) = (13, 1.8, 0.1)$ with initial value $TC(F^*,\mu,\gamma)$=146.847. Again, we use the Newton-Quasi method as mentioned above. After nine iterations, Table 6 shows that the minimum expected cost converges to the solution $(F^*,\mu^*,\gamma^*) =$

Table 6
Newton-Quasi method in searching the optimal solution ($\lambda = 1.4$, $\alpha = 0.05$, $\beta = 3.0$).

Number of iterations	$TC(F^*,\mu,\gamma)$	F^*	μ	γ	Max. tolerance
0	146.8465179	13	1.8000000	0.10000000	–
1	145.4015473	13	1.50828140	0.19512111	11.841
2	145.0161744	13	1.60459867	0.22598991	7.130
3	145.0072276	13	1.60203890	0.24449565	0.864
4	145.0070504	13	1.60256213	0.24762375	0.107
5	145.0070498	13	1.60257251	0.24770760	0.003
6	145.0070498	13	1.60257251	0.24770769	3.170×10^{-6}
7	145.0070498	13	1.60257250	0.24770765	1.420×10^{-6}
8	145.0070499	13	1.60257254	0.24770766	2.890×10^{-6}
9	145.0070496	13	1.60257253	0.24770768	6.720×10^{-7}

Max. tolerance = Max$\{|\partial(TC)/\partial\mu|, |\partial(TC)/\partial\gamma|\}$.

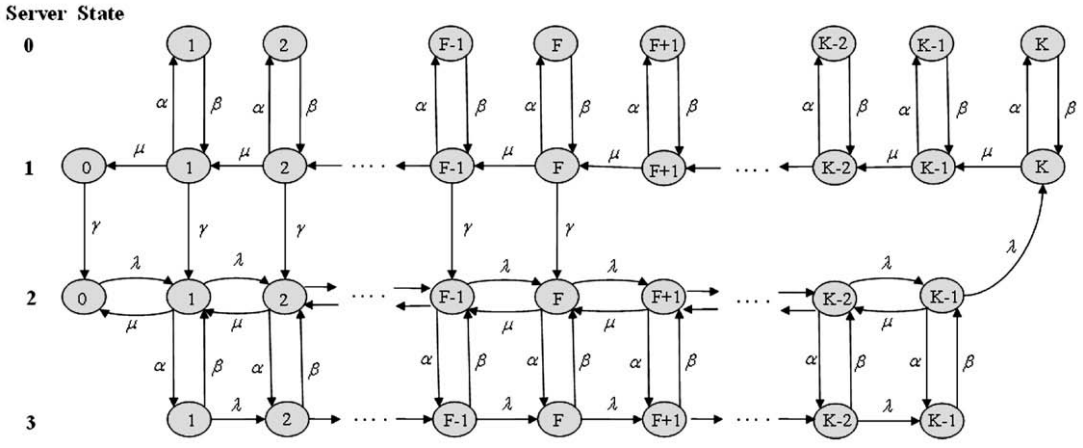

Fig. 1. State-transition-rate diagram for the *F*-policy M/M/1/K queueing system with server breakdowns and an exponential startup time.

(13, 1.60257253, 0.24770768) with value 145.0070496. Tables 4–6 show that the cost $TC(F^*, \mu^*, \gamma^*)$ is actually lower than the initial cost. Furthermore, we may conclude that the Newton-Quasi method is quite useful and easy to implement in finding the optimum values μ^* and γ^*.

6. Conclusions

In this paper, the matrix analytical technique is used to derive the steady-state probabilities. System performance measures can be obtained from steady-state results. Following the construction of the total expected cost function per unit time, we employ two methods to obtain the optimum values F^*, μ^* and γ^* that yield the minimum cost. One method is the direct search method which is used to find the optimal threshold F. The other is the Newton-Quasi method which is applied to obtain the optimal values of the continuous variables μ and γ if F^* has been decided. Specifically, an efficient procedure (Newton-Quasi method) is developed for searching the optimum values (F^*, μ^*, γ^*) that minimizes the cost function. We also provide some numerical experiments in which the system performance measures are evaluated under optimal operating conditions.

Appendix. Differential equations for the *F*-policy M/M/1/K queueing system with server breakdowns and an exponential startup time

Referring to the state-transition-rate diagram for the F-policy M/M/1/K queueing system with server breakdowns and an exponential startup time shown in Fig. 1, we consider the transitions occurring during $[t, t + dt]$. Using the birth and death process, we shall obtain the differential equations for the controlling arrivals systems:

$$\frac{dP_{0,n}(t)}{dt} = -\beta P_{0,n}(t) + \alpha P_{1,n}(t), \quad 1 \leqslant n \leqslant K, \tag{A.1}$$

$$\frac{dP_{1,0}(t)}{dt} = -\gamma P_{1,0}(t) + \mu P_{1,1}(t), \tag{A.2}$$

$$\frac{dP_{1,n}(t)}{dt} = -(\mu + \alpha + \gamma) P_{1,n}(t) + \mu P_{1,n+1}(t) + \beta P_{0,n}(t), \quad 1 \leqslant n \leqslant F, \tag{A.3}$$

$$\frac{dP_{1,n}(t)}{dt} = -(\mu + \alpha) P_{1,n}(t) + \mu P_{1,n+1}(t) + \beta P_{0,n}(t), \quad F + 1 \leqslant n \leqslant K - 1, \tag{A.4}$$

$$\frac{dP_{1,K}(t)}{dt} = -(\mu + \alpha) P_{1,K}(t) + \lambda P_{2,K-1}(t) + \beta P_{0,K}(t), \tag{A.5}$$

$$\frac{dP_{2,0}(t)}{dt} = -\lambda P_{2,0}(t) + \gamma P_{1,0}(t) + \mu P_{2,1}(t), \tag{A.6}$$

$$\frac{dP_{2,n}(t)}{dt} = -(\lambda + \alpha + \mu) P_{2,n}(t) + \gamma P_{1,n}(t) + \lambda P_{2,n-1}(t) + \mu P_{2,n+1}(t) + \beta P_{3,n}(t), \quad 1 \leqslant n \leqslant F, \tag{A.7}$$

$$\frac{dP_{2,n}(t)}{dt} = -(\lambda + \alpha + \mu) P_{2,n}(t) + \lambda P_{2,n-1}(t) + \mu P_{2,n+1}(t) + \beta P_{3,n}(t), \quad F + 1 \leqslant n \leqslant K - 2, \tag{A.8}$$

$$\frac{dP_{2,K-1}(t)}{dt} = -(\lambda + \alpha + \mu) P_{2,K-1}(t) + \lambda P_{2,K-2}(t) + \beta P_{3,K-1}(t), \quad K - 1 = n \neq F, \tag{A.9}$$

$$\frac{dP_{3,1}(t)}{dt} = -(\lambda + \beta) P_{3,1}(t) + \alpha P_{2,1}(t), \tag{A.10}$$

$$\frac{dP_{3,n}(t)}{dt} = -(\lambda + \beta) P_{3,n}(t) + \alpha P_{2,n}(t) + \lambda P_{3,n-1}(t), \quad 2 \leqslant n \leqslant K - 2, \tag{A.11}$$

$$\frac{dP_{3,K-1}(t)}{dt} = -\beta P_{3,K-1}(t) + \alpha P_{2,K-1}(t) + \lambda P_{3,K-2}(t). \tag{A.12}$$

If the solutions for the steady-state exist, it must satisfy

$$\lim_{t \to \infty} \frac{dP_{i,n}(t)}{dt} = 0 \quad \text{where } i = 0, 1, 2, 3;$$

that is, $P_{i,n}(t)$ is independent of t, and let us define

$$P_{i,n} = \lim_{t \to \infty} P_{i,n}(t),$$

then we obtain the steady-state equations for the F-policy M/M/1/K queueing system with server breakdowns and an exponential startup time, which is corresponding to Eqs. (1)–(12).

References

[1] K.R.A. Baker, Note on operating policies for the queue M/M/1 with exponential startups, INFOR 11 (1973) 71–72.
[2] A. Borthakur, J. Medhi, R. Gohain, Poisson input queueing systems with startup time and under control operating policy, Computers and Operations Research 14 (1987) 33–40.
[3] S.M. Gupta, Interrelationship between controlling arrival and service in queueing systems, Computers and Operations Research 22 (1995) 1005–1014.
[4] S.M. Gupta, E. Melachrinoudis, Complementarity and equivalence in finite source queueing models with spares, Computers and Operations Research 21 (1994) 289–296.
[5] F. Karaesmen, S.M. Gupta, Duality relations for queues with arrival and service control, Computers and Operations Research 24 (1997) 529–538.
[6] J.-C. Ke, The optimal control of an M/G/1 queueing system with server vacations, startup and breakdowns, Computers and Industrial Engineering 44 (2003) 567–579.
[7] J.-C. Ke, W.L. Pearn, Optimal management policy for heterogeneous arrival queueing systems with server breakdowns and vacations, Quality Technology and Quantitative Management 1 (2004) 149–162.
[8] H.W. Lee, J.O. Park, Optimal strategy in *N* policy production system with early setup, Journal of the Operational Research Society 48 (1997) 306–313.
[9] J. Medhi, J.G.C. Templeton, A Poisson input queue under *N*-policy and with a general start up time, Computers and Operations Research 19 (1992) 35–41.
[10] M.F. Neuts, Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach, The John Hopkins University Press, Baltimore, 1981.
[11] L. Tadj, A matrix analytic solution to a hysteretic queueing system with random server capacity, Applied Mathematics and Computation 119 (2001) 161–175.
[12] L. Tadj, G. Choudhury, Optimal design and control of queues, Top 13 (2005) 359–412.
[13] H. Takagi, M/G/1/K queues with *N* policy and setup times, Queueing Systems 14 (1993) 79–98.
[14] K.-H. Wang, Optimal operation of a Mokovian queueing system with a removable and non-removable server, Microelectronics Reliability 35 (1995) 1131–1136.
[15] K.-H. Wang, Optimal control of an $M/E_k/1$ queueing system with removable service station subject to breakdowns, The Journal of the Operational Research Society 48 (1997) 936–942.
[16] K.-H. Wang, K.-W. Chang, B.D. Sivazlian, Optimal control of a removable and non-reliable server in an infinite and a finite $M/H_2/1$ queueing system, Applied Mathematical Modelling 23 (1999) 651–666.
[17] K.-H. Wang, H.-T. Kao, G. Chen, Optimal management of a removable and non-reliable server in an infinite and a finite $M/H_k/1$ queueing system, Quality Technology and Quantitative Management 1 (2004) 325–339.
[18] K.-H. Wang, J.-C. Ke, Control policies of an M/G/1 queueing system with a removable and non-reliable server, International Transactions in Operational Research 9 (2002) 195–212.
[19] K.-H. Wang, C.-C. Kuo, W.L. Pearn, Optimal control of an M/G/1/K queueing system with combined *F* policy and startup time, Journal of Optimization Theory and Applications 135 (2007) 285–299.
[20] K.-H. Wang, C.-C. Kuo, W.L. Pearn, A recursive method for the *F*-policy G/M/1/K queueing system with an exponential startup time, Applied Mathematical Modeling 32 (2008) 958–970.
[21] K.-H. Wang, T.-Y. Wang, W.L. Pearn, Maximum entropy analysis to the *N* policy M/G/1 queueing system with server breakdowns and general startup times, Applied Mathematics and Computation 165 (2005) 45–61.
[22] M. Yadin, P. Naor, Queueing systems with a removable service station, Operational Research Quarterly 14 (1963) 393–405.