Controlling arrival for the machine repair problem with switching failure

Fu-Min Chang, Jau-Chuan Ke and Cheng-Hwai Liou
Chaoyang University of Technology, Department of Finance, Taichung, Taiwan
fmchang@cyut.edu.tw, jauchuan@nutc.edu.tw, jhliou@nutc.edu.tw

Abstract. This study investigates the warm-standby machine repair problem which involves a controlling arrival policy and switching failure probability. It involves operating machines with S warm standbys and one single server. The failure times and repair times are assumed to follow an exponential distribution. For such system, some system performance measures are derived and a steady-state expected cost function per unit time is developed. By using Quasi-Newton method followed direct search method, we can find the joint optimal parameter values at maximum profit such that the availability constraint is satisfied.

1. Introduction

Controllable queueing models aims to find the optimal operating policy, which are rules for turning the server on and off that result in the lowest long-run cost. The pioneering work in the issue of controlling the arrivals, also called F-policy, was first investigated by Gupta [1]. The F-policy aims to control the arrival process when service control is not possible through N-policy. Ke and Wang [2] analyzed machine repair problems with constant balking probability, negative exponential distributed reneging, and unreliable servers. The system steady-state availability, MTTF, and some system performance measures were presented. Wang, Ke, and Ke [3] investigated the profit analysis of the M/M/R machine repair problem with balking, reneging, and standby switching failures. A comprehensive and exhaustive discussion of machine repair problems was given by Haque and Armstrong[4].

In this paper, we study the issue for the F-policy M/M/1/K machine repair system consisting of M operating machines with S warm standbys. In this system, each of the operating machine fails independent of the state of the others and has an exponential time-to-failure distribution with parameter $\lambda$. Whenever an operating machine breaks down, it is instantly replaced by a standby machine. Each of the available standby machine fails independent of the state of all the others and has an exponential time-to-failure distribution with parameter $\alpha (0 \leq \alpha \leq \lambda)$. Whenever a machine fails, it is immediately sent to the repair facility where repair work is provided in the order of breakdowns, with a time–to-repair which is exponential distribution with parameter $\mu$. We consider that there is always the possibility of failures during the switching process from standby state to operating state. We assume that there is a significant probability $q$ of a switching failure. The server requires an exponential startup time with parameter $\gamma$ to start allowing failed machines in the queue. Arriving failed machines form a single waiting line based on the order of their arrivals; that is, they are queued according to the first-come, first-served (FCFS) policy. We assume that when a standby machine moves into an operating state, its characteristics will be that of an operating machine. If a machine repair system reaches its finite capacity $K (K = M + S < \infty)$, no further failed machines are allowed to enter the system until enough machines who are already in the system have been repaired so that the number of failed machines decreases to a predetermined threshold $F (0 \leq F \leq K - 1)$. 
2. Steady-state equations and matrix-geometric solutions

The states of the presented system are described by the pair \((i, n)\), where \(i\) represents server’s state and \(n\) represents the number of failed machines in the system, respectively. Let \(i=0\) denote the failed machine is not allowed to enter into the system, \(i=1\) denote the failed machine is allowed to enter into the system. Thus \((i, n)\) is a state of the system. Let \(\lambda_n\) for this system is given by

\[
\lambda_n = \begin{cases} 
M\lambda(1 - q) + (S - n)\alpha & n = 0, 1, \ldots, S - 1 \\
(K - n)\lambda & n = S, \ldots, K - 1 \\
0 & \text{otherwise}
\end{cases}
\]

In steady-state, let \(P_0(n)\) denotes the probability that there are \(n\) failed machines in the system when the arrivals are not allowed to enter into the system, where \(n = 0, 1, 2, \ldots, K\). We also let \(P_1(n)\) denotes the probability that there are \(n\) failed machines in the system when the arrivals are allowed to enter into the system, where \(n = 0, 1, 2, \ldots, K - 1\).

Referring to the state-transition rate diagram for the F-policy M/M/1/K machine repair system with standby switching failures for \(F \leq S\) case, the equations of the machine repair model are constructed as follows:

\[
\gamma P_0(0) = \mu P_0(1),
\]

\[
(\mu + \gamma) P_0(n) = \mu P_0(n + 1), \quad 1 \leq n \leq F
\]

\[
\mu P_0(n) = \mu P_0(n + 1), \quad F + 1 \leq n \leq K - 1
\]

\[
\mu P_0(K) = \lambda P_1(K - 1)
\]

\[
(M\lambda + S\alpha) P_1(0) = \gamma P_0(0) + \mu P_1(1)
\]

\[
[M\lambda + (S - n)\alpha + \mu] P_1(n) = \gamma P_0(n) + [M\lambda(1 - q) + (S - n + 1)\alpha] P_1(n - 1) + \mu P_1(n + 1) + \sum_{i=1}^{n-2} \Phi_{n-1-i} P_1(i), \quad 2 \leq n \leq F
\]

\[
[M\lambda + (S - n)\alpha + \mu] P_1(n) = [M\lambda(1 - q) + (S - n + 1)\alpha] P_1(n - 1) + \mu P_1(n + 1) + \sum_{i=1}^{n-2} \Phi_{n-1-i} P_1(i), \quad F + 1 \leq n \leq S
\]

\[
[(M - 1)\lambda + \mu] P_1(S + 1) = M\lambda P_1(S) + \mu P_1(S + 2) + \sum_{i=1}^{S-1} \Theta P_1(i)
\]

\[
[(K - n)\lambda + \mu] P_1(n) = (K - n + 1)\lambda P_1(n - 1) + \mu P_1(n + 1) \quad S + 1 \leq n \leq K - 2
\]

\[
(\lambda + \mu) P_1(K - 1) = 2\lambda P_1(K - 2)
\]

Similar to \(F \leq S\) case, we can obtain the following steady-state equations for \(F > S\) case.

\[
\gamma P_0(0) = \mu P_0(1),
\]

\[
(\mu + \gamma) P_0(n) = \mu P_0(n + 1), \quad 1 \leq n \leq F
\]

\[
\mu P_0(n) = \mu P_0(n + 1), \quad F + 1 \leq n \leq K - 1
\]

\[
\mu P_0(K) = \lambda P_1(K - 1)
\]

\[
(M\lambda + S\alpha) P_1(0) = \gamma P_0(0) + \mu P_1(1)
\]

\[
[M\lambda + (S - n)\alpha + \mu] P_1(n) = \gamma P_0(n) + [M\lambda(1 - q) + S\alpha] P_1(n - 1) + \mu P_1(n + 1) + \sum_{i=1}^{n-2} \Phi_{n-1-i} P_1(i), \quad 2 \leq n \leq F
\]

\[
[M\lambda + (S - n)\alpha + \mu] P_1(n) = [M\lambda(1 - q) + S\alpha] P_1(n - 1) + \mu P_1(n + 1) + \sum_{i=1}^{n-2} \Phi_{n-1-i} P_1(i), \quad F + 1 \leq n \leq S
\]

\[
[(M - 1)\lambda + \mu] P_1(S + 1) = M\lambda P_1(S) + \mu P_1(S + 2) + \sum_{i=1}^{S-1} \Theta P_1(i)
\]

\[
[(K - n)\lambda + \mu] P_1(n) = (K - n + 1)\lambda P_1(n - 1) + \mu P_1(n + 1) \quad S + 1 \leq n \leq K - 2
\]

\[
(\lambda + \mu) P_1(K - 1) = 2\lambda P_1(K - 2)
\]
\[ M \alpha (S-n) + \mu P_1(n) = \gamma P_0(n) + \left[M \beta (1-q) + (S-n+1) \alpha \right] P_1(n-1) + \mu P_1(n+1) + \sum_{i=0}^{k-2} \phi_{n-i} P_1(i) \] \quad 2 \leq n \leq S \tag{18}

\[ (M-1) \lambda + \mu P_1(S+1) = \gamma P_0(S+1) + M \alpha P_1(S) + \mu P_1(S+2) + \sum_{i=1}^{S-1} \Theta_i P_1(i) \tag{19} \]

\[ [(K-n) \lambda + \mu] P_1(n) = \gamma P_0(n) + (K-n+1) \lambda P_1(n-1) + \mu P_1(n+1) \quad S+2 \leq n \leq F \tag{20} \]

\[ [(K-n) \lambda + \mu] P_1(n) = (K-n+1) \lambda P_1(n-1) + \mu P_1(n+1) \quad F+1 \leq n \leq K-2 \tag{21} \]

\[ (\lambda + \mu) P_1(K-1) = 2 \lambda P_1(K-2) \tag{22} \]

Since the closed-form probability solutions of equations (1)-(22) are too complicated to develop explicit expressions by using a recursive method. Hence we use matrix-geometric methods to analysis this problem. We find that the equations (1)-(22) of the present model can be expressed in the matrix form. Then the steady-state probabilities, \( P_0(n), P_1(n) \), can be computed by using the following normalizing conditions:

\[ \sum_{n=0}^{K} P_0(n) + \sum_{n=0}^{K-1} P_1(n) = 1. \tag{23} \]

3. System performance measures and cost analysis

Using the steady-state probabilities derived in the previous section, we compute some system performance measures. The assumptions and notations are defined as follows: \( L \) denotes the expected number of failed machines in the system; \( P_s \) denotes the probability that the server is in the busy state; \( P_i \) denotes the probability that the server is in the idle state; \( P_{bl} \) denotes the probability that the server is blocked. \( E(O) \) denotes the expected number of operating machines in the system; \( E(S) \) denotes the expected number of standby machines in the system; \( M.A. \) denotes machine availability (the fraction of the total time that the machines are working). \( S.R. \) denotes the average switching failure rate; \( W \) denotes the expected waiting time in the system. The expressions for \( L, P_B, P_I, P_{bl}, E(O), E(S) \), \( M.A. \), \( S.R. \), and \( W \) are given by:

\[ L = \sum_{n=0}^{K} nP_0(n) + \sum_{n=0}^{K-1}nP_1(n), \quad P_B = \sum_{n=1}^{K} P_0(n) + \sum_{n=1}^{K-1} P_1(n), \quad P_I = 1 - P_B, \quad P_{bl} = \sum_{n=0}^{K} P_0(n), \]

\[ E(O) = M - \left[ \sum_{n=1}^{M} nP_0(S+n) + \sum_{n=1}^{M-1}nP_1(S+n) \right], \quad E(S) = \sum_{n=0}^{S} (S-n)P_0(n) + \sum_{n=0}^{S} (S-n)P_1(n) \]

\[ M.A. = 1 - \frac{L}{K}, \quad S.R. = \sum_{n=1}^{S} M \lambda q P_1(n-1), \quad \text{Let } W \text{ denote the expected waiting time in the system.} \]

By using the Little’s formula we obtain \( W = L / \lambda_e \), where the effective arrival rate in the system is \( \lambda_e = \sum_{n=0}^{K-1} \lambda_n P_1(n) \).

We construct a total expected cost function per unit time for the F-policy M/M/1/K repair system with standby switching failures, in which \( F, S \) and \( \mu \) are decision variables. The discrete variables \( F \) and \( S \) are required to be natural numbers, and the continuous variable \( \mu \) is positive numbers. Our main objective is to search the optimal values of \( F, S, \mu \) so as to minimize the total expected cost.

We define the following cost elements: \( C_h \) denotes the holding cost per unit time when one failed machine joins the system; \( C_i \) denotes the cost per unit time of a failed machine after all standbys are exhausted; \( C_2 \) denotes the cost per unit time when one machine is acting as a warm standby state; \( C_s \) denotes the cost per unit time when the server is idle; \( C_t \) denotes the cost per unit time when the server
is busy; $C_4$ denotes fixed cost for each lost failed machine when the system is blocked; $C_5$ denotes the cost per unit time of providing the service rate $\mu$; $C_6$ denotes the waiting cost per unit time when one machine is waiting for services. Using the definitions of these elements listed above, the total expected cost function per unit time is determined by:

$$TC(F, S, \mu) = C_1 E + C_2 \lambda E + C_3 P_1 + C_4 P_b + C_5 \lambda P_{bl} + C_6 \mu + C_7 W$$

It is extremely difficult to develop the useful analytical results for the optimum value $(F^*, S^*, \mu^*)$, because that the values $F$ and $S$ are discrete quantities, $\mu$ is continuous quantity, and the cost function is highly complex. To determine the optimal value $(F^*, S^*, \mu^*)$, we apply the direct search method to obtain optimal value. Firstly, we obtain optimal value $(F^*, S^*, \mu^*)$ when $F$ and $S$ are initially fixed. Next, based on this solution, we then use direct search method to obtain the optimal value $(F^*, S^*, \mu^*)$, which minimize the cost function $TC(F, S, \mu)$. For fixed values $F$ and $S$, the direct search algorithm is applying to search $\mu$ until the minimum value of $TC(F, S, \mu)$ is achieved, say $(F, S, \mu^*)$ and constraints are satisfied. In practice, the server rate could be adjusted to further minimize the total cost. It is assumed that the value of service rate has the upper bound of $\mu_U$. The investment budget is restricted by a given budget $C$. After obtaining the joint optimal value $\mu^*$ of the continuous variable $\mu$, we use direct search method again to obtain the optimal value $(F^*, S^*)$ such that the expected cost function $TC(F, S, \mu^*)$ attains a minimum value, say $TC(F^*, S^*, \mu^*)$.

4. Conclusions

In this paper, we studied an $F$-policy M/M/1/K machine repair system with standby switching failures, and obtain the steady-state analytic solutions. The stationary probability vectors were obtained using matrix-geometric analysis and the technique of matrix recursive. We have provided an efficient method to determine the optimum repair policy, including the optimal threshold, the number of standbys and the optimal server rate simultaneously, in order to minimize the expected cost function when system availability is maintained at given level, and calculated various system performance measures under optimal operating conditions. This research presents an extension of the Markovian model theory and the analysis of the model will provide a useful performance evaluation tool.

References

[1] Gupta, S.M. (1995). Interrelationship between controlling arrival and service in queueing systems. Computers and Operations Research, 22, 1005–1014.
[2] Ke, J. C., & Wang, K. H. (1999). Cost analysis of the M/M/R machine repair problem with balking, reneging, and server breakdowns. Journal of the Operational Research Society, 50(3), 275–282.
[3] Wang, K. H., Ke, J. B., & Ke, J. C. (2007). Profit analysis of the M/M/R machine repair problem with balking, reneging, and standby switching failures. Computers and Operations Research, 34(3), 835 – 847.
[4] Haque, L., & Armstrong, M. J. (2007). A survey of the machine interference problem. European Journal of Operational Research, 179, 469–482.