International Journal of
***O**perations **R**esearch*

International Journal of Operations Research Vol. 8, No. 2, 1-7 (2011)

Analysis of Queues with an Imperfectly Repairable Server

Zhe George Zhang[1,3*], **Naishuo Tian**[2], **Ernie Love**[3,4] **and Zhanyou Ma**[2]

[1] Department of Decision Sciences, Western Washington University, WA98225, USA

[2] Department of Mathematics, Yanshan University, Qinhuangdao, Hebei, China

[3] Faculty of Business Administration, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

[4]School of Business and Management, American University of Sharjah, UAE

Received October 2010; Revised January 2011; Accepted January 2011

Abstract— We consider a queueing system where the server is subject to failures and can be repaired within a random period of time. After an "imperfect" repair, the server (machine) will be in a state between "as good as new" and "as bad as old". The server state will determine either the failure rate or the service rate or both. A Quasi Birth Death process is developed for modeling such a system. We investigate the effect of maintenance policy on the queue and on overall performance. Under a cost structure, we illustrate the search for the optimal maintenance policy to minimize the costs. Queueing effects on the average cost of operating such a system are examined with numerical analysis. Many examples of such structures exist in practice.

Keywords—Imperfect repair, quasi birth death process, stationary distribution, threshold policy, average operating cost.

1. INTRODUCTION

We consider a queueing system where the server is subject to failures and can be repaired within a random period of time. After the repair, the server (machine) will be in a state between "as good as new" and "as bad as old". The server state will determine either the failure rate or the service rate or both. In the maintenance literature, this state can be represented as the server's "virtual age". (Kijima 1988, 1989). Thus, the state of the server can be interpreted as the condition of the server (or machine). In the maintenance literature, the post-repair status of the machine is characterized by this virtual age. Kijima posited two forms of repair effects. In a Kijima I repair, a repair improves any damage incurred between the last repair and the current failure. Hence, virtual age monotonically increases over time. A Kijima II repair however can improve all cumulative damage incurred since the beginning of the last "perfect" repair (or overhaul). Thus, with an increasing failure intensity, the virtual age of a machine is at least asymptotically bounded (Guo, et al. 2001). In this paper Kijima II repairs are assumed. By discretizing virtual age we can represent the server's "virtual age" as a series of increasing states. (see Shirmohammadi et al. 2007). Thus, the state of the server can be interpreted as the condition (or "virtual age") of the server (or machine). The control of three costs is of concern in this paper. First are the costs of customers waiting for service. Second are the costs associated with repairs incurred upon failure. Finally, should the virtual state of the machine reach a preset maximum value, the server is automatically overhauled to a "good-as-new" state and a replacement cost is incurred. We seek to find the policy that minimized average total cost per unit time as a function of the virtual state of the server.

Many examples of such structures exist in practice. A simple example is an ATM machine wherein customers arrive to conduct financial transactions. The machine is subject to failures. At a failure instant, based on the virtual age of the ATM, a decision of either replacing or repairing has to be made. Industrially, ships arriving for unloading at a port facility can experience a similar situation. If ships arrive and the facility is working then unloading occurs. If the unloading facility fails then, provided its virtual age is below an acceptable level, it is repaired. The post-repair virtual

* Corresponding author's email: gzhang@sfu.ca

age is (presumed) reduced by this action. Meanwhile ships continue to arrive and queue. Above a predetermined virtual age at a failure instant, the unloading facility is replaced. Queuing continues at all times.

The rest of the paper is organized as follows: In Section 2, we formulate the system as a queueing model with server's interruptions due to failures and repairs. Section 3 presents the solution procedure and algorithms. Numerical examples are shown in Section 4. Finally Section 5 concludes.

2. MODEL FOMULATION

Assume that customers arrive at a service facility according to a Poisson process with rate l. A single server offers service to customers in a First-Come-First-Served order. The service times are independent and identically distributed random variables with the exponential distribution with rate m (which can be dependent on the server state). The server's condition is represented by a server-state variable which takes non-negative integers $\boldsymbol{i} = 0,1,2,...,\boldsymbol{M}$, with 0 representing a new state and $\boldsymbol{M}$ the state that a replacement must be made if a failure results in this state. The replacement time is exponentially distributed with rate d. It is assumed that the failure rate, denoted by a_i, of the machine is an increasing function of $\boldsymbol{i}$. To represent the operating status of the system, a further indicator variable is introduced with $\boldsymbol{J} = 0$ representing a working state and $\boldsymbol{J} = 1$ representing a failure state. Whenever the server fails and the server-state $\boldsymbol{i} < \boldsymbol{M}$, a repair action is performed right away requiring a random period of time, exponentially distributed with rate b. Without loss of generality, in this paper, we assume $b > d$, meaning that on average the replacement time is longer than repair time. Let $\boldsymbol{L}$ be the state variable for the number of customers in the system. Now the process $\{\boldsymbol{L}(\boldsymbol{t}),\boldsymbol{I}(\boldsymbol{t}),\boldsymbol{J}(\boldsymbol{t})\}$ becomes a quasi-birth-death (QBD) process with state space:

$$W = \{(0,\boldsymbol{i},0) \cup (\boldsymbol{k},\boldsymbol{i},\boldsymbol{j}) : \boldsymbol{k} \geq 1, \boldsymbol{i} = 0,1,...,\boldsymbol{M}, \boldsymbol{j} = 0,1\}$$

When this QBD process is positive recurrent (which we demonstrate with a stability condition presented later), the stationary distribution can be reached and denoted by

$$p_{\boldsymbol{kij}} = \boldsymbol{P}\{\boldsymbol{L} = \boldsymbol{k}, \boldsymbol{I} = \boldsymbol{i}, \boldsymbol{J} = \boldsymbol{j}\} = \lim_{\boldsymbol{t} \to \infty} \boldsymbol{P}\{\boldsymbol{L}(\boldsymbol{t}) = \boldsymbol{k}, \boldsymbol{I}(\boldsymbol{t}) = \boldsymbol{i}, \boldsymbol{J}(\boldsymbol{t}) = \boldsymbol{j}\}, \quad (\boldsymbol{k},\boldsymbol{i},\boldsymbol{j}) \in W.$$

Writing the probability distribution in the form of vector segments, we have

$$p_0 = \boldsymbol{p}_0$$
$$p_{\boldsymbol{k}} = [\boldsymbol{p}_{\boldsymbol{k}}, \boldsymbol{q}_{\boldsymbol{k}}] \qquad \textit{for } \boldsymbol{k} \geq 1,$$

and

$$\boldsymbol{p}_{\boldsymbol{k}} = (\boldsymbol{p}_{\boldsymbol{k}0}, \boldsymbol{p}_{\boldsymbol{k}1}, ... \boldsymbol{p}_{\boldsymbol{kM}}), \quad \boldsymbol{k} \geq 0,$$
$$\boldsymbol{q}_{\boldsymbol{k}} = (\boldsymbol{q}_{\boldsymbol{k}0}, \boldsymbol{q}_{\boldsymbol{k}1}, .. \boldsymbol{q}_{\boldsymbol{kM}}), \quad \boldsymbol{k} \geq 1,$$

where

$$\boldsymbol{p}_{\boldsymbol{ki}} = p_{\boldsymbol{ki}0}, \quad \boldsymbol{k} \geq 0, \boldsymbol{i} = 0,1,...,\boldsymbol{M};$$
$$\boldsymbol{q}_{\boldsymbol{ki}} = p_{\boldsymbol{ki}1}, \quad \boldsymbol{k} \geq 1, \boldsymbol{i} = 0,1,...,\boldsymbol{M}.$$

Here we use symbol $p_{ki}(q_{ki})$ to denote the stationary probability for the working (failure) state with k customers and server condition i. Note that $\boldsymbol{k} \geq 1$ for the failure states implies that the server cannot fail during the idle time. It is assumed that after each repair, the server-state reaches one of the three possible states: newer, same, or, older with probabilities $h, j,$ and g. Thus we have the transition probabilities as follows:

$$\boldsymbol{P}\{(\boldsymbol{k},\boldsymbol{i},1) \to (\boldsymbol{k}',\boldsymbol{i}-1,0)\} = h, \text{ for } \boldsymbol{k} \geq 1, \text{ and } \boldsymbol{k}' \geq \boldsymbol{k};$$
$$\boldsymbol{P}\{(\boldsymbol{k},\boldsymbol{i},1) \to (\boldsymbol{k}',\boldsymbol{i},0)\} = j, \text{ for } \boldsymbol{k} \geq 1, \text{ and } \boldsymbol{k}' \geq \boldsymbol{k};$$
$$\boldsymbol{P}\{(\boldsymbol{k},\boldsymbol{i},1) \to (\boldsymbol{k}',\boldsymbol{i}+1,0)\} = g, \text{ for } \boldsymbol{k} \geq 1, \text{ and } \boldsymbol{k}' \geq \boldsymbol{k},$$

where $h + j + g = 1$. It is easy to prove that the expected server-state is between "as good as new" and "as bad as old" if $h > g$; is "as bad as old" if $h = g$; and is "worse than old" if $h < g$. Therefore, with proper choice of h and g values, we can model different repair effects. We need to point out that the server-state can determine both the service rate and the failure rate. A reasonable assumption is that the service rate is a decreasing function of $\boldsymbol{i}$ and the failure rate is an increasing function of $\boldsymbol{i}$. To simplify our analysis, we present the case where the service rate is constant for all operating states. Due to the complex transition structure of this three-dimensional state QBD process we will not present the state-transition diagram which will be complex and messy. Instead, we directly present the infinitesimal generator as

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{A}_0 & \boldsymbol{C}_0 & & & & \\ \boldsymbol{B}_1 & \boldsymbol{A} & \boldsymbol{C} & & & \\ & \boldsymbol{B} & \boldsymbol{A} & \boldsymbol{C} & & \\ & & \boldsymbol{B} & \boldsymbol{A} & \boldsymbol{C} & \\ & & & \mathbf{O} & \mathbf{O} & \mathbf{O} \end{pmatrix},$$

where

$$\boldsymbol{A}_0 = \begin{pmatrix} -\lambda & & & \\ & -\lambda & & \\ & & \mathbf{O} & \\ & & & -\lambda \end{pmatrix}_{(M+1)\times(M+1)}, \quad \boldsymbol{B}_1 = \begin{pmatrix} \mu & & & \\ & \mu & & \\ & & \mathbf{O} & \\ & & & \mu \end{pmatrix}_{(2M+2)\times(M+1)} = \begin{pmatrix} \mu\boldsymbol{I} \\ 0 \end{pmatrix},$$

where $\boldsymbol{I}$ is the unit matrix of order $(M+1)$ (with 1's on the diagonal),

$$\boldsymbol{C}_0 = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \mathbf{O} & \\ & & & \lambda \end{pmatrix}_{(M+1)\times(2M+2)} = \begin{pmatrix} \lambda\boldsymbol{I} & 0 \end{pmatrix},$$

$$\boldsymbol{A} = \begin{pmatrix} -(\mu+a_0+\lambda) & & & & a_0 & & & \\ & -(\mu+a_1+\lambda) & & & & a_1 & & \\ & & \mathbf{O} & & & & \mathbf{O} & \\ & & & -(\mu+a_M+\lambda) & & & & a_M \\ (h+j)b & gb & & & -(\lambda+b) & & & \\ hb & jb & gb & & & -(\lambda+b) & & \\ & \mathbf{O} & \mathbf{O} & \mathbf{O} & & & \mathbf{O} & \\ d & & & & & & & -(\lambda+d) \end{pmatrix}_{(2M+2)\times(2M+2)},$$

$$\boldsymbol{B} = \begin{pmatrix} \mu\boldsymbol{I} & 0 \\ 0 & 0 \end{pmatrix}_{(2M+2)\times(2M+2)}$$

where $\boldsymbol{I}$ is the $(M+1)\times(M+1)$ unit matrix, and

$$\boldsymbol{C} = \lambda\boldsymbol{I},$$

where $\boldsymbol{I}$ is the $(2M+2)\times(2M+2)$ unit matrix.

Letting $\boldsymbol{W} = \boldsymbol{B} + \boldsymbol{C} + \boldsymbol{A}$, we have

$$\boldsymbol{W} = \begin{pmatrix} -a_0 & & & & a_0 & & & \\ & -a_1 & & & & a_1 & & \\ & & \mathbf{O} & & & & \mathbf{O} & \\ & & & -a_M & & & & a_M \\ (h+j)b & gb & & & -b & & & \\ hb & jb & gb & & & -b & & \\ & \mathbf{O} & \mathbf{O} & \mathbf{O} & & & \mathbf{O} & \\ d & & & & & & & -d \end{pmatrix}.$$

Now we solve for the stationary probability vector of $\boldsymbol{W}$, denoted by $\boldsymbol{X} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_M, \boldsymbol{x}_{M+1}, ..., \boldsymbol{x}_{2M+1}\}$. The equation system is

$$-a_0\boldsymbol{x}_0 + (h+j)b\boldsymbol{x}_{M+1} + hb\boldsymbol{x}_{M+2} + d\boldsymbol{x}_{2M+1} = 0, \tag{1}$$

$$-a_1\boldsymbol{x}_1 + gb\boldsymbol{x}_{M+1} + jb\boldsymbol{x}_{M+2} + hb\boldsymbol{x}_{M+3} = 0, \tag{2}$$

$$-a_2\boldsymbol{x}_2 + gb\boldsymbol{x}_{M+2} + jb\boldsymbol{x}_{M+3} + hb\boldsymbol{x}_{M+4} = 0, \tag{3}$$

$$\vdots$$

$$-a_{M-3}x_{M-3} + gbx_{2M-3} + jbx_{2M-2} + hbx_{2M-1} = 0, \tag{4}$$

$$-a_{M-2}x_{M-2} + gbx_{2M-2} + jbx_{2M-1} + hbx_{2M} = 0, \tag{5}$$

$$-a_{M-1}x_{M-1} + gbx_{2M-1} + jbx_{2M} = 0, \tag{6}$$

$$-a_{M}x_{M} + gbx_{2M} = 0, \tag{7}$$

$$a_0x_0 - bx_{M+1} = 0, \tag{8}$$

$$a_1x_1 - bx_{M+2} = 0, \tag{9}$$

$$a_2x_2 - bx_{M+3} = 0, \tag{10}$$

$$\vdots$$

$$a_{M-1}x_{M-1} - bx_{2M} = 0, \tag{11}$$

$$a_{M}x_{M} - dx_{2M+1} = 0. \tag{12}$$

From (12), letting $x_{2M+1} = s$, we have

$$x_M = \frac{d}{a_M}x_{2M+1} = \left(\frac{d}{a_M}\right)s,$$

Using (7), we obtain

$$x_{2M} = \frac{a_M}{gb}x_M = \frac{a_M}{gb}\left(\frac{d}{a_M}\right)s = \left(\frac{d}{gb}\right)s,$$

Then using (11), we get

$$x_{M-1} = \frac{b}{a_{M-1}}x_{2M} = \frac{b}{a_{M-1}}\left(\frac{d}{gb}\right)s = \left(\frac{d}{a_{M-1}g}\right)s.$$

Now from (6), we have

$$\begin{aligned} x_{2M-1} &= \frac{a_{M-1}}{gb}x_{M-1} - \frac{j}{g}x_{2M} \\ &= \left(\frac{a_{M-1}}{gb}\right)\left(\frac{d}{a_{M-1}g}\right)s - \left(\frac{j}{g}\right)\left(\frac{d}{gb}\right)s \\ &= \left(\frac{d}{g^2b}\right)(1-j)s. \end{aligned}$$

Next we can find x_{M-2} from $a_{M-2}x_{M-2} - bx_{2M-1} = 0$, that is

$$\begin{aligned} x_{M-2} &= \left(\frac{b}{a_{M-2}}\right)x_{2M-1} = \left(\frac{b}{a_{M-2}}\right)\left(\frac{d}{g^2b}\right)(1-j)s \\ &= \left(\frac{d}{a_{M-2}g^2}\right)(1-j)s. \end{aligned}$$

Using (5), we get

$$\begin{aligned} x_{2M-2} &= \left(\frac{d}{g^3b}\right)(1-j)^2 s - \left(\frac{d}{g^2b}\right)hs \\ &= \left(\frac{d}{g^2b}\right)s\left(\frac{(1-j)^2}{g} - h\right) \\ &= \left(\frac{d}{g^3b}\right)g^2 + gh + h^2)s. \end{aligned}$$

We also obtain

$$\boldsymbol{x}_{\boldsymbol{M}-3} = \frac{b}{a_{\boldsymbol{M}-3}}\boldsymbol{x}_{2\boldsymbol{M}-2} = \left(\frac{d}{a_{\boldsymbol{M}-3}g^3}\right)\left(g^2 + gh + h^2\right)s.$$

Keep doing this, we can obtain the general solution:

$$\boldsymbol{x}_{\boldsymbol{M}-\boldsymbol{i}} = \left(\frac{d}{a_{\boldsymbol{M}-\boldsymbol{i}}g^{\boldsymbol{i}+1}}\right)\left(\sum_{\boldsymbol{j}=0}^{\boldsymbol{i}-1} g^{\boldsymbol{i}-\boldsymbol{j}}h^{\boldsymbol{j}}\right)s, \qquad \boldsymbol{i} = 0,1,2,\ldots,\boldsymbol{M};$$

$$\boldsymbol{x}_{2\boldsymbol{M}-\boldsymbol{i}} = \left(\frac{d}{g^{\boldsymbol{i}+1}b}\right)\left(\sum_{\boldsymbol{j}=0}^{\boldsymbol{i}} g^{\boldsymbol{i}-\boldsymbol{j}}h^{\boldsymbol{j}}\right)s, \qquad \boldsymbol{i} = 0,1,2,\ldots,\boldsymbol{M}-1;$$

and it follows from the normalization condition that

$$s = \left[\sum_{\boldsymbol{i}=0}^{\boldsymbol{M}} \frac{d}{a_{\boldsymbol{M}-\boldsymbol{i}}g^{\boldsymbol{i}+1}}\left(\sum_{\boldsymbol{j}=0}^{\boldsymbol{i}-1} g^{\boldsymbol{i}-\boldsymbol{j}}h^{\boldsymbol{i}}\right) + \sum_{\boldsymbol{i}=0}^{\boldsymbol{M}-1} \frac{d}{g^{\boldsymbol{i}+1}b}\left(\sum_{\boldsymbol{j}=0}^{\boldsymbol{i}} g^{\boldsymbol{i}-\boldsymbol{j}}h^{\boldsymbol{i}}\right)\right]^{-1}.$$

Now we have obtained the solution $\boldsymbol{X}$. Using the mean drift condition (Neuts, 1981), one can demonstrate a stability condition as

$$\frac{l}{m} < \sum_{\boldsymbol{i}=0}^{\boldsymbol{M}} \boldsymbol{x}_{\boldsymbol{M}-\boldsymbol{i}} = \sum_{\boldsymbol{i}=0}^{\boldsymbol{M}} \left(\frac{d}{a_{\boldsymbol{M}-\boldsymbol{i}}g^{\boldsymbol{i}+1}}\right)\left(\sum_{\boldsymbol{j}=0}^{\boldsymbol{i}-1} g^{\boldsymbol{i}-\boldsymbol{j}}h^{\boldsymbol{j}}\right)s.14 \tag{13}$$

Under this condition, we have a mean recurrent system and steady state can be reached. The stationary distribution can be obtained as a matrix-geometric solution.

3. SOLUTION PROCEDURE AND PERFORMANCE MEASURES

Due to the structure of the sub-matrices in this model, it is not possible to obtain the explicit expression for the stationary probability vector, $\boldsymbol{W}$. In Figure 1 below, we present an algorithm to numerically evaluate a rate matrix $\boldsymbol{R}$ which we can use to obtain a solution to our model.

After $\boldsymbol{R}$ is obtained, the boundary state stationary probability vector, $[p_0, p_1]$ can be solved from

$$[p_0, p_1]\begin{Bmatrix}\boldsymbol{A}_0 & \boldsymbol{C}_0\\ \boldsymbol{B}_1 & \boldsymbol{H}\end{Bmatrix} = 0.$$

where $\boldsymbol{H} = \boldsymbol{A} + \boldsymbol{RB}$. Obviously, the matrix-geometric solution is given as

$$p_{\boldsymbol{k}} = p_1\boldsymbol{R}^{\boldsymbol{k}-1}, \qquad \text{for } \boldsymbol{k} \geq 2,$$

and normalization condition must be used to completely determine the stationary distribution of the system.
With the stationary distribution, we can numerically evaluate the major performance measures for the system. First, we can determine the average number in the system is computed as

$$\boldsymbol{E}(\boldsymbol{L}) = \sum_{\boldsymbol{k}=0}^{\infty}\boldsymbol{k}\left(\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}} \boldsymbol{p}_{\boldsymbol{ki}}\right) + \sum_{\boldsymbol{k}=1}^{\infty}\boldsymbol{k}\left(\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}} \boldsymbol{q}_{\boldsymbol{ki}}\right),$$

and the average time in the system can be also obtained by using Little's Law. The probabilities of system being in failure/repair, failure/replacement, and working states are given by

$$\boldsymbol{P}(\text{failure / repair}) = \frac{b^{-1}\sum_{\boldsymbol{k}=1}^{\infty}\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}-1}\boldsymbol{q}_{\boldsymbol{ki}}}{\sum_{\boldsymbol{k}=0}^{\infty}\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}} a_{\boldsymbol{i}}^{-1}\boldsymbol{p}_{\boldsymbol{ki}} + b^{-1}\sum_{\boldsymbol{k}=1}^{\infty}\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}-1}\boldsymbol{q}_{\boldsymbol{ki}} + d^{-1}\sum_{\boldsymbol{k}=1}^{\infty}\boldsymbol{q}_{\boldsymbol{kI}}},$$

$$\boldsymbol{P}(\text{failure / replacement}) = \frac{d^{-1}\sum_{\boldsymbol{k}=1}^{\infty}\boldsymbol{q}_{\boldsymbol{kI}}.}{\sum_{\boldsymbol{k}=0}^{\infty}\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}} a_{\boldsymbol{i}}^{-1}\boldsymbol{p}_{\boldsymbol{ki}} + b^{-1}\sum_{\boldsymbol{k}=1}^{\infty}\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}-1}\boldsymbol{q}_{\boldsymbol{ki}} + d^{-1}\sum_{\boldsymbol{k}=1}^{\infty}\boldsymbol{q}_{\boldsymbol{kI}}},$$

$$\boldsymbol{P}(\text{working}) = \frac{\sum_{\boldsymbol{k}=0}^{\infty}\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}} a_{\boldsymbol{i}}^{-1}\boldsymbol{p}_{\boldsymbol{ki}}}{\sum_{\boldsymbol{k}=0}^{\infty}\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}} a_{\boldsymbol{i}}^{-1}\boldsymbol{p}_{\boldsymbol{ki}} + b^{-1}\sum_{\boldsymbol{k}=1}^{\infty}\sum_{\boldsymbol{i}=0}^{\boldsymbol{I}-1}\boldsymbol{q}_{\boldsymbol{ki}} + d^{-1}\sum_{\boldsymbol{k}=1}^{\infty}\boldsymbol{q}_{\boldsymbol{kI}}}.$$

Under a cost structure, we can now calculate the average operating cost per time unit. Let $\boldsymbol{C_f}, \boldsymbol{C_r}$, and $\boldsymbol{C_h}$ be the repair cost per time unit when the system is in failure/repair state, the replacement cost per time unit when in failure/replacement state, and waiting cost per customer per time unit, respectively. We have the average operating cost as follows:

$$\boldsymbol{g(M)} = \boldsymbol{C_f P(failure\ /\ repair\ )} + \boldsymbol{C_r P(failure\ /\ replacement\ )} + \boldsymbol{C_h E(L)}.$$

The average operating cost is a function of the decision variable $\boldsymbol{M}$, the threshold virtual age for replacing the machine. In the next section, we present some numerical results to show the $\boldsymbol{g(M)}$ curve and locate the cost minimizing virtual age for this imperfect repair service system. One can see from Figure 2 below, $\boldsymbol{g(M)}$ yields a minimum virtual age at which point replacement is optimal.

The average operating cost is a function of the decision variable, $\boldsymbol{M}$, the threshold virtual age for replacing the machine. In the next section, we conduct numerical experiments to show the shape of the $\boldsymbol{g(M)}$ curve for one set of parameters as well as locating the cost minimizing virtual age for this (imperfect repaired) service system.

4. NUMERICAL ILLUSTRATIONS

Example 1: We assume system parameters of:

$$l = 1, m = 4, a_i = 1 + \boldsymbol{i}, b = 1, d = 0.5, h = 0.5, j = 0.3, g = 0.2.$$

Assumed cost parameters are: $\boldsymbol{C_f} = \$30, \boldsymbol{C_r} = \$200, \boldsymbol{C_h} = \$3.$

Note that this is a case where a repair can improve the system state as $h > g$. We search the cost minimization $\boldsymbol{M}$, the threshold virtual age at which the system must be replaced. Results are shown in Figures 2-6 below.

In Figure 2 one can see a minimum $g(M)$ at a virtual age of approximately 5.5 time units. Beyond that time, $g(M)$ rises again but levels off to an equilibrium cost per unit time. Large values of M mean no replacement and then the system is only kept operating by repairing as needed. Even with no replacements, the system reaches a stable number of customers in the system as noted in Figure 3 due to the condition improving effects of repairs. Likewise the probability of failing (and needing repair) reaches a stable asymptotic limit in the absence of replacement (Figure 4). Of course, if M is allowed to increase, as seen in Figure 5, the probability of a failure leading to a replacement falls to zero. Finally, as seen in Figure 6, with large M, the probability that the system is in a working state stabilized to approximately 0.651 (using these parameters). That is, roughly 65% of the time the system would be observed serving customers and 35% of the time it would be in a state of repair.

5. CONCLUDING REMARKS

In this paper we have developed a model that includes the effect of server failures on queueing systems. The basic framework is to incorporate a Kijima type-II virtual age model to capture the impact of repairs. If repair effects are significant enough then the virtual age of the server will remain low and the need for replacement becomes unnecessary. Typically however, with a aging system, a virtual age can be reached in which it is more economical to replace the unit rather than repair it. We have demonstrated in this paper that an optimal (virtual age) replacement time can be identified. Our model allows for extensive parameterization to explore scenarios of interest. Several interesting features of the model should be highlighted. First, in this paper, we have assumed the mean replacement time ($1/d$) is presumed $>$ than the mean repair time ($1/b$) although the stability condition established in this paper does not require such a constraint. In fact as often in practice replacements, while more costly, can be carried out instantaneously in comparison to a repair and appropriate parameter selection in our model would allow this. Finally, if both mean replacement and mean repair times ($1/d$, $1/b$) $<<$ average service time ($1/m$) of customers then our model reverts to one of standard queuing.

A variety of extensions of this model can be formulated to better reflect actual operating environments. For example, it is well know that many failure systems can be modeled using Weibull or other non-memoryless distributions. Phase-type models are easily amenable to the matrix structures in this paper and can be designed to reflect non-memoryless failure process. Models involving multiple servers, albeit computationally burdensome, are also straightforward.

ACKNOWLEDGEMENTS

The financial support from NSERC Grant RGPIN197319 of Canada is acknowledged by the first author.

REFERENCES

1. Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied Probability*, 26: 89-102.
2. Kijima, M., Morimura, H. and Suzuki Y. (1988). Periodical replacement problem without assuming minimal repair. *European Journal of Operational Research*, 37: 194-203.
3. Guo, R., Ascher H. and Love C. E. (2001). Generalized Models of Repairable Systems: A Survey via Stochastic Processes Formalism. *ORiON*, 16(2): 87-128.
4. Shirmohammadi, A. H., Zhang Z. G. and Love, C. E. (2007). A Computational Model for Determining the Optimal Preventive Maintenance Policy with Random Breakdowns and Imperfect Repairs. *IEEE Transactions on Reliability*, 56(2): 332-339.
5. Neuts, M. (1981). *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach.* The Johns Hopkins University Press, Baltimore, 1981

INPUT B, A, C, I is the identity matrix
e is the column of 1's
and △ is the error tolerance
OUTPUT approximate solution to R
Step 1 G = $(-A)^{-1}$ B
Step 2 while ‖ e - G.e ‖ ≥ △ do Steps 3 - 5
Step 3 set U = A + CG
Step 4 and G = $(-U)^{-1}$B
Step 5 Set R = C$(-U)^{-1}$
Step 6 OUTPUT

Figure 1: Linear Progression Algorithm for Computing Rate Matrix

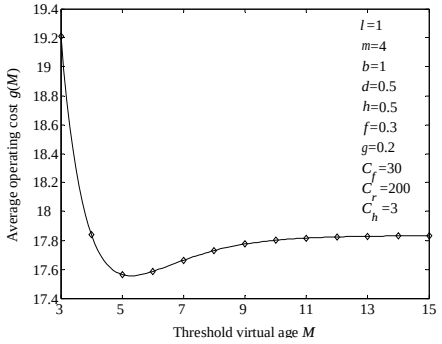

Figure 2: $g(M)$ versus M

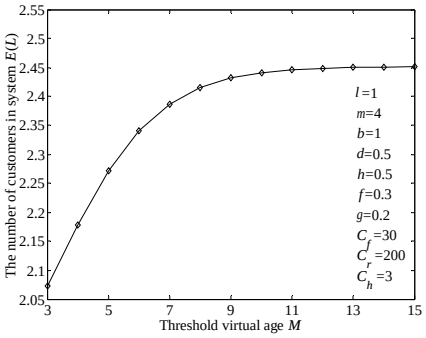

Figure 3: $E(L)$ versus M

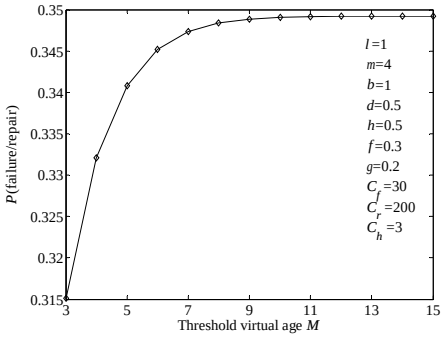

Figure 4: P(failure/repair) versus M

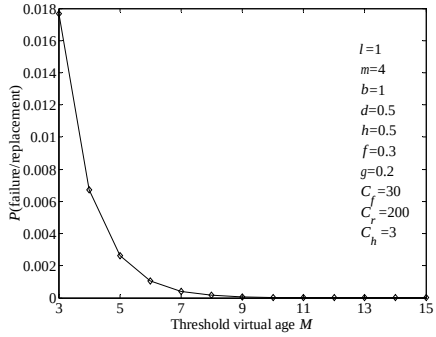

Figure 5: P(failure/replacement) versus M

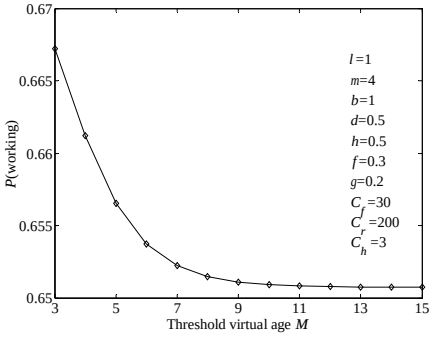

Figure 6: P(working) versus M