



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Computers & Industrial Engineering 46 (2004) 87–99

computers &
industrial
engineering

www.elsevier.com/locate/dsw

Sensitivity analysis of the optimal management policy for a queuing system with a removable and non-reliable server[☆]

Wen Lea Pearn^a, Jau-Chuan Ke^{b,*}, Ying Chung Chang^c

^aDepartment of Industrial Engineering and Management, National Chiao Tung University, Taiwan, ROC

^bDepartment of Statistics, National Taichung Institute of Technology, Taiwan, ROC

^cDepartment of Industrial Engineering and Management, Ching Yun University, Taiwan, ROC

Accepted 10 November 2003

Abstract

The management policy of an $M/G/1$ queue with a single removable and non-reliable server is considered. The decision-maker can turn the single server on at any arrival epoch or off at any service completion. It is assumed that the server breaks down according to a Poisson process and the repair time has a general distribution. Arrivals form a Poisson process and service times are generally distributed. In this paper, we consider a practical problem applying such a model. We use the analytic results of the queueing model and apply an efficient Matlab program to calculate the optimal threshold of management policy and some system characteristics. Analytical results for sensitivity analysis are obtained. We carry out extensive numerical computations for illustration purposes. An application example is presented to display how the Matlab program could be used. The research is useful to the analyst for making reliable decisions to manage the referred queueing system.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Analytical results; Management policy; Non-reliable; Sensitivity analysis

1. Introduction

In this paper we study the operational characteristics of an $M/G/1$ queueing system in which a removable and non-reliable server operates with an N policy. The term ‘removable server’ is just an abbreviation for the system of turning on and turning off the server, depending on the number of

[☆] This manuscript was processed by Area Editor Gary L. Hogg.

* Corresponding author.

E-mail addresses: jauchuan@mail.ntit.edu.tw (J.-C. Ke); roller@cc.nctu.edu.tw (W.L. Pearn); ycchang@cyu.edu.tw (Y.C. Chang).

customers in the system. A non-reliable server means that the server is typically subject to unpredictable breakdowns. The server is removable and applies the N policy: turn the server on whenever N ($N \geq 1$) or more customers are present, turn the server off only when no customers are present. After the server is turned off, the server may not operate until N customers are present in the system.

For a reliable server, the N policy M/M/1 queueing system was first developed by [Yadin and Naor \(1963\)](#), and the N policy M/G/1 queueing system was developed by several researchers such as [Bell \(1971, 1972\)](#), [Heyman \(1968\)](#), [Kimura \(1981\)](#), [Teghem \(1987\)](#), [Tijms \(1986\)](#), [Artalejo \(1998\)](#), and [Wang and Ke \(2000\)](#). For a non-reliable server, [Avi-Itzhak and Naor \(1963\)](#) studied the ordinary M/M/1 queueing system where the service rule does not depend on the number of customers in the queue. The ordinary M/E_k/1 queueing system with arrival rate depending on server breakdowns, was investigated by [Shogan \(1979\)](#). [Neuts and Lucanton \(1979\)](#) studied a Markovian queueing system with multiple servers subject to breakdowns and repairs. The explicit solutions for the N policy Markovian queueing systems with a non-reliable server may be used to obtain the results for the N policy M/M/1 queueing system with a reliable server (see [Sivazlian & Stanfel, 1975](#)), or the ordinary M/M/1 queueing system with a non-reliable server (see [Wang, 1990](#)), or the ordinary M/M/1 queueing system with a reliable server (see [Sivazlian & Stanfel, 1975](#)) as a special case.

The purpose of this paper is threefold. First, an efficient Matlab program is used to calculate the optimal policy value N and some system characteristics. Second, the analytical results of the sensitivity analysis are derived. We then carry out extensive numerical computation for sensitivity analysis purpose. Third, we present an application example showing the way in which the Matlab program is used to calculate system characteristics, the optimum value of N and its minimum expected cost for various system parameters, while maintaining the maximum service quality.

Note that existing research works for the queueing system have never investigated the analytic solutions for the sensitivity analysis. In this paper, we will completely and successfully perform the sensitivity analysis for the M/G/1 queueing system with a removable and non-reliable server. Through this sensitivity analysis, we will be able to analyze the complex but exact solutions for a practical and general queueing system.

2. The queueing service model

2.1. System description and assumptions

Referring to [Wang and Ke \(2002\)](#), we consider the following model formulation. A cycle of the model consists of an idle period and a completion period. The completion period is composed into busy period and the breakdown period. As the system is empty, one cycle begins. The server is turned off until there are N customers in the system. We call this the idle period. The busy period is initiated when the server starts serving the waiting customers. While providing the service, the server may break down and be sent for repair immediately. This is called the breakdown period. As soon as the server is repaired, he returns to service again until all customers in the system are serviced. Since the completion period starts when the idle period is over and terminates when there are no customers in the system, the completion period may be represented as the sum of the busy period and the breakdown period. In addition, we consider the model under the following assumptions:

(1) the arrival process is a Poisson process with rate λ . Arriving customers at the server form a single waiting line and are serviced in the order of their arrivals (FCFS). The server may serve only one customer at a time. The service times constitute a set of independent and identically distributed random variables with a common distribution function $F_S(t)$, a mean μ_S and a finite variance σ_S^2 . (2) Whenever the system is empty the idle period starts. When the server finds at least N customers waiting in the system, he begins the service immediately until the system is empty again. (3) When the server is working, the server may break down at any time with a Poisson breakdown rate α . (4) When the server fails, it is immediately repaired in a repair facility, where the repair times are independent and identically distributed random variables with a common distribution function $F_R(t)$, a mean μ_R and a finite variance σ_R^2 . (5) If the server fails or a customer is in service, then newly arriving customers or waiting customers have to wait in the queue until the server is free. The service is interrupted if the server breaks down, and the server is immediately sent for repair. When the repair is completed, the server immediately returns for service.

2.2. Steady-state results

2.2.1. The expected number of customers in the system

We define the expected number of customers in the system for the M/G/1 queueing system under the N policy with server breakdowns as $E[N_s]$. From Wang and Ke (2002) we have the following expression.

$$E[N_s] = \frac{(N-1)}{2} + \rho(1 + \alpha\mu_R) + \frac{\lambda^2[(1 + \alpha\mu_R)^2(\mu_S^2 + \sigma_S^2) + \alpha\mu_S(\mu_R^2 + \sigma_R^2)]}{2[1 - \rho(1 + \alpha\mu_R)]}, \quad (1)$$

where $\rho = \lambda\mu_S$.

2.2.2. Other system characteristics

Define $E[I] \equiv$ the expected length of the idle period; $E[B] \equiv$ the expected length of the busy period; $E[D] \equiv$ the expected length of the breakdown period; $E[C] \equiv$ the expected length of a cycle. Using the results stated in Wang and Ke (2002), we have the long-run fraction of time, that the server is idle, busy, or broken down, respectively, are:

$$\frac{E[I]}{E[C]} = 1 - \rho(1 + \alpha\mu_R), \quad (2)$$

$$\frac{E[B]}{E[C]} = \rho, \quad (3)$$

$$\frac{E[D]}{E[C]} = \alpha\lambda\mu_S\mu_R, \quad (4)$$

and the number of cycles per unit time is

$$\frac{1}{E[C]} = \frac{\lambda[1 - \rho(1 + \alpha\mu_R)]}{N}. \quad (5)$$

3. Optimal management policy

3.1. Total expected cost function

We develop the total expected cost function per unit time for the N policy M/G/1 queueing system with a non-reliable server, in which N is a decision variable. With the cost structure being constructed, our objective is to determine the optimal management policy to minimize this cost function, while maintaining the maximum service quality to the customers. Let $C_h \equiv$ holding cost per unit time for each customer present in the system; $C_o \equiv$ cost per unit time for keeping the server on; $C_f \equiv$ cost per unit time for keeping the server off; $C_b \equiv$ breakdown cost per unit time for a broken server; $C_s \equiv$ start-up cost for turning the server on. Based on the definitions of each cost element, and its corresponding system characteristics, the total expected cost function per unit time is given by

$$F(N) = C_h E[N_s] + C_o \frac{E[B]}{E[C]} + C_f \frac{E[I]}{E[C]} + C_b \frac{E[D]}{E[C]} + C_s \frac{1}{E[C]}. \quad (6)$$

It is to be noted that the second and third terms of Eq. (1) are independent of the decision variable N . Likewise, we note from Eqs. (2)–(4) that, terms $E[B]/E[C]$, $E[I]/E[C]$, and $E[D]/E[C]$ do not involve the decision variable N . Discarding those cost terms that are not a function of the decision variable N , the optimization problem reduces to

$$\text{minimize } \tilde{F}(N) = C_h \frac{N-1}{2} + C_s \frac{\lambda[1 - \rho(1 + \alpha\mu_R)]}{N}. \quad (7)$$

Omitting the fixed cost— $(1/2) C_h$ of the first term, Eq. (7) reduces to

$$\text{minimize } \hat{F}(N) = C_h \frac{N}{2} + C_s \frac{\lambda[1 - \rho(1 + \alpha\mu_R)]}{N}, \quad (8)$$

subject to $0 < \rho < 1$, and $N = 1, 2, \dots$

3.2. Determining the optimal management policy

Since N is a positive integer, $N = 1, 2, \dots$, the optimal value of N , N^* , to minimize $F(N)$ is determined by satisfying the following inequalities

$$\begin{aligned} \hat{F}(N^* - 1) &\geq \hat{F}(N^*), \\ \hat{F}(N^* + 1) &\geq \hat{F}(N^*). \end{aligned} \quad (9)$$

From Eq. (8), the necessary conditions for N^* to be optimal reduces to

$$(N^* - 1)N^* \leq \frac{2\lambda C_s [1 - \rho(1 + \alpha\mu_R)]}{C_h} \leq N^*(N^* + 1). \quad (10)$$

Note that a double solution is possible.

Differentiating $F(N)$ with respect to N and setting the result equal to zero yields

$$\frac{C_h}{2} - C_s \frac{\lambda[1 - \rho(1 + \alpha\mu_R)]}{N^2} = 0.$$

Thus, the optimal value of N is approximately given by

$$N^* = \sqrt{\frac{2\lambda C_s [1 - \rho(1 + \alpha\mu_R)]}{C_h}}. \quad (11)$$

Differentiate $F(N)$ with respect to N twice and then substitute $N = N^*$ to obtain

$$\frac{d^2F(N^*)}{dN^2} = \frac{2\lambda C_s [1 - \rho(1 + \alpha\mu_R)]}{N^3} > 0, \quad (12)$$

which implies that $F(N)$ is a concave upward (convex) function and achieves a global minimum when $N = N^*$. We note that if N^* is not an integer, the optimum value of N is one of the integers closest to N^* , the expression may rewrite as

$$N^* = \sqrt{\frac{2\lambda C_s [1 - \rho(1 + \alpha\mu_R)]}{C_h}} + \varepsilon, \quad (13)$$

where $\varepsilon \in (-1, 1)$ is a constant.

For the special case, N -policy M/M/1 queueing system with a removable and non-reliable server, where the service and repair times are assumed to follow the exponential distribution with means μ_S and μ_R , respectively (see Wang, 1995), with some similar algebraic manipulations. It is interesting to note that we would obtain the same expression for the optimal value N^* as stated in Eq. (13).

4. Analytical results for sensitivity analysis

We now perform a sensitivity analysis on the optimum value N^* based on changes in considerable values of the cost parameters C_h , C_o , C_f , C_b , C_s and system parameters λ , μ_S , α , and μ_R . It is to be noted that the terms $E[B]/E[C]$, $E[I]/E[C]$, and $E[D]/E[C]$ do not involve the decision variable N . We may set the relative cost parameters C_o , C_f , and C_b to be some fixed constants. Eq. (13) suggests that $N^* \propto \sqrt{C_s/C_h}$, which is straightforward. Differentiate N^* with respect to λ to obtain

$$\frac{\partial N^*}{\partial \lambda} = \frac{\sqrt{C_s/C_h} [1 - 2\rho(1 + \gamma)]}{\sqrt{2\lambda} [1 - \rho(1 + \gamma)]}, \quad (14)$$

where $\gamma = \alpha\mu_R$.

On putting this last expression equal to 0 and solving for λ , we find $\lambda = 1/\mu_S [2(1 + \gamma)]$. (Note that $\lambda < 1/\mu_S [2(1 + \gamma)]$ is required.) Differentiate $\partial N^*/\partial \lambda$ with respect to λ again and substitute $\lambda = 1/\mu_S [2(1 + \gamma)]$ to get

$$\frac{\partial^2 N^*}{\partial \lambda^2} \Big|_{\lambda=1/\mu_S [2(1+\gamma)]} = -2\sqrt{\frac{2C_s \mu_S^3 (1 + \gamma)^3}{C_h}} < 0, \quad (15)$$

which implies N^* is a concave downward function with respect to λ and obtains its maximum at $\lambda = 1/\mu_S[2(1 + \gamma)]$. Differentiating N^* with respect to μ_S , we have

$$\frac{\partial N^*}{\partial \mu_S} = \frac{-(1 + \gamma)\lambda^2\sqrt{C_s}}{\sqrt{2\lambda C_h[1 - \rho(1 + \gamma)]}} < 0, \quad (16)$$

for $\lambda < 1/\mu_S[(1 + \gamma)]$, $\forall \mu_S$. This means that N^* decreases in μ_S . Differentiate N^* with respect to α to find

$$\frac{\partial N^*}{\partial \alpha} = \frac{-\rho\sqrt{\lambda C_s}}{\mu_R\sqrt{2C_h[1 - \rho(1 + \gamma)]}} < 0, \quad (17)$$

for $\lambda < 1/\mu_S[(1 + \gamma)]$, $\forall \alpha$. This means that N^* decreases in α . Differentiate N^* with respect to μ_R we get

$$\frac{\partial N^*}{\partial \mu_R} = \frac{-\lambda\rho\alpha\sqrt{C_s}}{\sqrt{2C_h[1 - \rho(1 + \gamma)]}} < 0, \quad (18)$$

for $\lambda < 1/\mu_S[(1 + \gamma)]$, $\forall \mu_R$. This means that N^* decreases in μ_R . We have the following analytic sensitivity analysis results:

1. N^* increases in λ for $\rho < 1/[2(1 + \gamma)]$ and decreases in λ for $\rho > 1/[2(1 + \gamma)]$.
2. N^* decreases in μ_S . For the queueing system with service times described by an arbitrary probability distribution with mean μ_S , it is to be noted that only the mean service time would affect the sensitivity results.
3. N^* decreases in α for large ρ and γ (for $\rho(1 + \gamma)$ closes to 1), whereas it is insensitive to α for small ρ and γ ($\rho(1 + \gamma) \ll 1$).
4. N^* decreases in μ_R for large ρ and γ (for $\rho(1 + \gamma)$ closes to 1), whereas it is insensitive to μ_R for small ρ and small γ (for $\rho(1 + \gamma) \ll 1$).
5. C_o , C_f , and C_b do not affect N^* .
6. N^* is proportional to $\sqrt{C_s/C_h}$. In other words, N^* increases in C_s whereas it decreases in C_h .

It is noted that the results of the sensitivity investigation in Wang (1995) were not correctly interpreted.

5. Numerical computations

5.1. Parameters setting

We now perform a numerical illustration of the sensitivity analysis on the optimum value N^* based on changes in considerable values of system parameters λ , μ_S , α , and μ_R . It is to be noted that the terms $E[B]/E[C]$, $E[I]/E[C]$, and $E[D]/E[C]$ do not involve the decision variable N . We may set the relative cost parameters C_o , C_f , and C_b to be some fixed constants. Additionally, incremental, rather than accounting costs are considered, since the latter often include such nonincremental elements as overhead. Eq. (13) suggests that $N^* \propto \sqrt{C_s/C_h}$. And we note that the N policy is applied to control the queueing system due to expensive start-up cost per cycle (relative to holding cost). In purpose to

Table 1

Parameters settings for various system parameters combinations with fixed costs $C_o = 100$, $C_f = 20$, and $C_b = 200$

C_h	C_c	λ	μ_S^{-1}	α	μ_R^{-1}	Parameter setting (1)	Parameter setting (2)
5	1600	(1)	(2)	0.2	1	0.05(0.0125)0.95	1(0.125)10
5	1600	(1)	(2)	0.2	1	0.05(0.0125)0.95	1, 1.5, 3
5	1600	(1)	(2)	0.2	1	0.35, 0.65, 0.95	1(0.125)10
5	4000	0.3	1	(1)	(2)	0.01(0.018)0.91	1(0.18)10
5	4000	0.3	1	(1)	(2)	0.01(0.018)0.91	1, 1.9, 3.7
5	4000	0.3	1	(1)	(2)	0.28, 0.64, 0.91	1(0.18)10

demonstrate some of the sensitivity analysis results, $C_s = \$1600$, $\$4000$ have been set to represent two levels of cost relationship and C_h is selected to be $\$5$.

The sensitivity calculations demonstration may now focus on the four critical input parameters: λ , μ_S , α and μ_R . We group them into two pairs: λ and μ_S ; α and μ_R , under consideration simultaneously in order to study the interaction of the key factors. Individual affection on the optimal solution is examined as well. We now consider the following experimental design of system parameters for sensitivity analysis on the optimum value N^* based on changes in considerable input values. Note that $0 < \rho < 1$ and $0 < \gamma < 1$ are sufficient for stationary. We calculate the optimal value N^* for the parameters settings summarized in Table 1, which cover a widespread range of applications dealing with the referred queueing model.

Rows 2–4 list the parameters settings for various combinations of λ and μ_S . The specified range $\lambda = 0.05$ (0.0125) 0.95 and $\mu_S^{-1} = 1$ (0.125) 10 is considered in which $\rho = \lambda\mu_S$ covers widespread traffic intensities in real application. Rows 3 and 4 are chosen to examine the sensitivity of N^* versus λ or μ_S once at a time. In row 3, $\lambda = 0.05$ (0.0125) 0.95 and three levels of $\mu_S^{-1} = 1, 1.5, \text{ and } 3$ are selected. In row 4, three levels of $\lambda = 0.35, 0.65, 0.95$ and $\mu_S^{-1} = 1(0.125)10$ are considered. Rows 5–7 list the parameters settings for various combinations of α and μ_R . The range $\alpha = 0.01$ (0.018) 0.91 and $\mu_R^{-1} = 1(0.18)10$ are considered in which $\gamma = \alpha\mu_R$ covers widespread real applications. In rows 6 and 7, the sensitivity of N^* versus α or μ_R are examined once at a time. Row 6 demonstrates the relationship of N^* versus α for three levels $\mu_R^{-1} = 1, 1.9, 3.7$. Row 7 illustrates the relationship of N^* versus μ_R for three levels $\alpha = 0.28, 0.64, 0.91$.

Figures are arranged in the following way: Fig. 1(a) plots the surface of N^* versus $\lambda = 0.05$ (0.0125) 0.95 and $\mu_S^{-1} = 1$ (0.125) 10 for various parameters settings given in Table 1. Fig. 1(b) and (c) are the cross-section, which plot the curves of N^* versus λ and μ_S , respectively. Three levels of the other system parameter are picked (see Table 1). Fig. 2(a) plots the surface of N^* versus $\alpha = 0.01$ (0.018) 0.91 and $\mu_R^{-1} = 1(0.18)10$ for various parameters settings given in Table 1. Fig. 2(b) and (c) are the cross-section plot the curves of N^* versus α and μ_R , respectively. Three levels of the other system parameter are picked as listed in Table 1.

5.2. Interpretation of the results in figures

Fig. 1(a)–(c) reveal that: (i) N^* increases in λ for $\lambda < 1/\mu_S[2(1 + \gamma)]$ and decreases in λ for $\lambda > 1/\mu_S[2(1 + \gamma)]$. The ‘local maximum’ $1/\mu_S[2(1 + \gamma)]$ is moving from left to right as μ_S

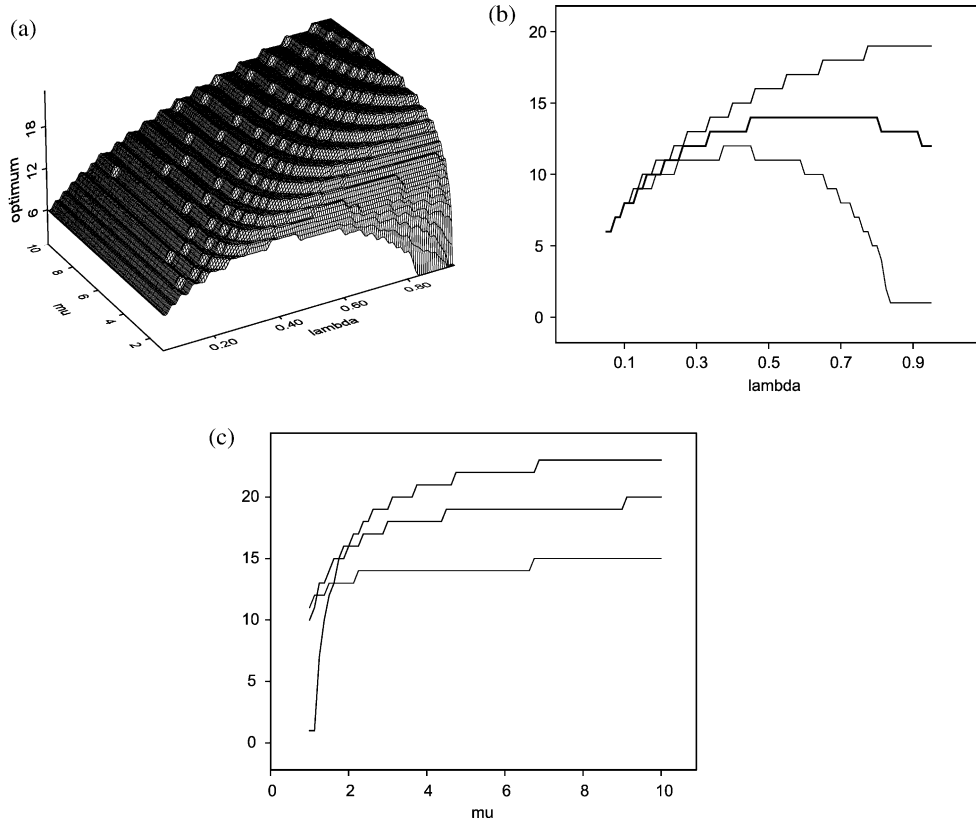


Fig. 1. (a). Surface plot of N^* versus $\lambda = 0.05(0.0125)0.95$ and $\mu_S^{-1} = 1(0.125)10$. (b). Plots of N^* versus $\lambda = 0.05(0.0125)0.95$ for $\mu_S^{-1} = 1, 1.5, 3$ (bottom to top in plot). (c). Plots of N^* versus $\mu_S^{-1} = 1(0.125)10$ for $\lambda = 0.35, 0.65, 0.95$ (bottom to top in plot).

decreases. If μ_S is small enough, one could see that N^* increases in λ (see Fig. 1(b)). (ii) N^* decreases in μ_S (see Fig. 1(c)). One can see from Fig. 1(a) that N^* increases in λ but decreases in μ_S for small μ_S .

We observe from Fig. 2(a) and (c) that: (i) N^* decreases in α for large γ , whereas it is insensitive to α for small γ (see Fig. 2(a) and (b)). (ii) N^* decreases in μ_R for large γ (see Fig. 2(a) and (c)), whereas it is insensitive to μ_R for small γ (see Fig. 2(a) and (c)). It is interesting to see that for a wide range of combinations of α and μ_R in Fig. 2(a), N^* is insensitivity to both parameters.

6. An application example

M/G/1 queueing systems arise naturally as models for many computer communication systems, production systems and integrated manufacturing systems. This section gives an example of the way in which the Matlab program can be used by an analyst to calculate system characteristics, the optimum value of N and its minimum expected cost. The application illustrates the levels of detail that are

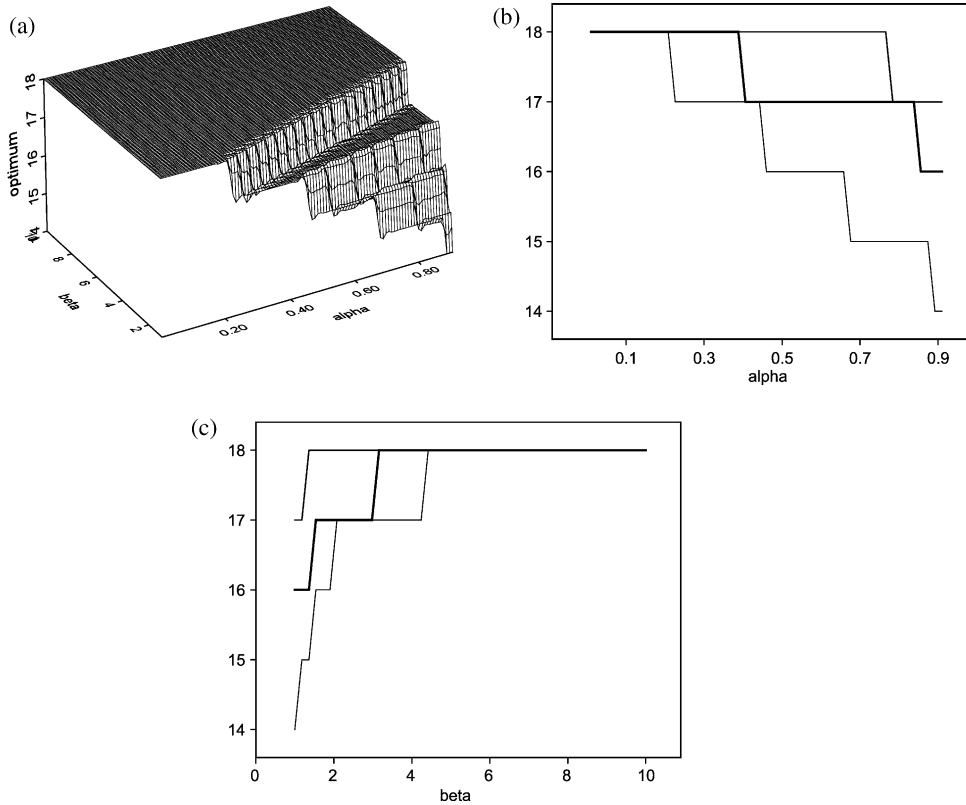


Fig. 2. (a). Surface plot of N^* versus $\alpha = 0.01(0.018)0.91$ and $\mu_R^{-1} = 1(0.18)10$. (b). Plots of N^* versus $\alpha = 0.01(0.018)0.91$ for $\mu_R^{-1} = 1, 1.9, 3.7$ (bottom to top in plot). (c). Plots of N^* versus $\mu_R^{-1} = 1(0.18)10$ for $\alpha = 0.28, 0.64, 0.91$ (top to bottom in plot).

appropriate for building a model and using that model for performance projection. The example illustrates the relationship between modeling concepts, evaluation algorithms, and modeling software. The example also indicates how such software can save the cost by the analyst.

Consider the following example. Computer communication networks use a variety of flow control policies to achieve high throughput, low delay, and stability. Here, we model the flow control policy of IBM’s System Network Architecture (SNA). SNA routes messages from sources to destinations by way of intermediate nodes which temporarily buffer the messages. Messages buffers are a scarce resource. The flow control policy regulates the flow of messages between source/destination pairs in an effort to avoid problems such as deadlock and starvation, which could result from poor buffer management.

SNA has a window flow control policy. The key control parameter is the window size, N . There is a ‘message generation center’ (server) and a ‘pacing box’. Together, the message generation center and the pacing box mimic the flow control policy, in the following way. When a source starts sending messages to a particular destination, a pacing count at the source is initialized to the value of zero. This pacing count is incremented every time a message is received. The pacing box ‘stores’ up to $N - 1$ messages. When the N th message arrives, it triggers the discharge of all

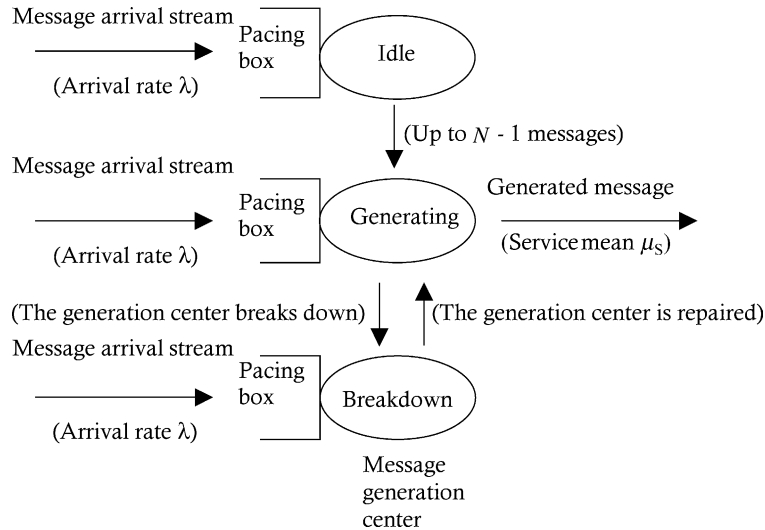


Fig. 3. A flowchart for the message generating in computer communication networks.

N messages into the queue of the message generation center to be served. The service time of the message generation center has a general distribution with mean μ_S ; as long as its queue is non-empty, it will generate message traffic at this rate. One assumption made by this modeling approach is that ‘message generation center’ is subject to breakdown at any time while it is working. Whenever the ‘message generation center’ fails, it is immediately repaired. After the repair the interrupted service is resumed. Breakdowns occur only when the service is in progress, and never when the system is empty. As soon as the server (message generation center) is repaired, it functions as good as a new one. Our objective is to model the ‘pacing level’ of messages between a single source/destination pair—the optimum window size N to minimize the total expected cost. Customers, which represent messages, arrive at the source node at rate λ . They flow node to node, requiring the message generation center to service at a mean time μ_S . A single coaxial cable is used to interconnect stations. The source continues to transmit, regardless of the number of outstanding messages. This model is shown in Fig. 3.

The calculations for the model do not require complicated intermediate functions to be implemented, and most of the system characteristics usually of interest can be calculated in a straightforward way. In the example investigated, input system parameters the message stream arrival rate $\lambda = 0.4$ message/s, the mean time of message generating (service) $\mu_S = 1.0$ s/message, the standard deviation of message generating (service) $\sigma_S = 1.0$ s/message, the breakdown rate $\alpha = 0.05$ unit/s, the mean time to repair $\mu_R = 0.2$ s/unit, the standard deviation of repair $\sigma_R = 1.0$ s/message, and cost element the holding cost per second for each message present in the system set to $C_h = \$5$, the cost per second for keeping the ‘message generation center’ (server) working set to $C_o = \$50$, the cost per second for keeping the ‘message generation center’ off set to $C_f = \$10$, the breakdown cost per second for a broken ‘message generation center’ set to $C_b = \$100$, and the start-up cost for turning the ‘message generation center’ on set to $C_s = \$200$.

System parameters	Notation	Value
Message stream arrival rate	λ	0.4
Mean time of message generating (service)	μ_S	1.0
Standard deviation of message generating (service)	σ_S	1.0
Breakdown rate	α	0.05
Mean time to repair	μ_R	0.2
Standard deviation of message generating (service)	σ_R	1.0
Holding cost per second for each message present in the system	C_h	5
Cost per second for keeping the server working	C_o	50
Cost per second for keeping the server off	C_f	10
Breakdown cost per second for a broken server	C_b	100
Start-up cost for turning the server on	C_s	200

The program output is shown in the following:

The output is:

$$E[N_s] = 2.1848$$

$$E[B] = 6.7114$$

$$E[I] = 10$$

$$E[D] = 0.0671$$

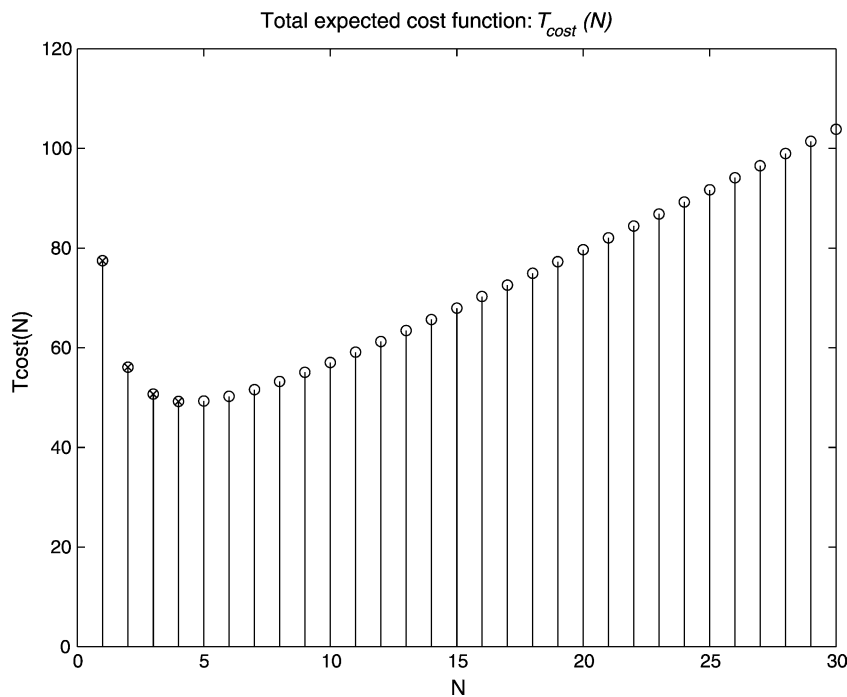


Fig. 4. Plot of $T_{cost}(N)$ versus N for $N = 1(1)30$.

$$N^* = 4$$

$$T_{\text{cost}}(N^*) = 49.2042$$

The MATLAB computer program gives the expected number of messages in the system $E[N_s] = 2.18$ messages, the expected length of generating (busy) period $E[B] = 6.71$ s, the expected length of idle period $E[I] = 10$ second and the expected length of breakdown period $E[D] = 0.07$ s. The value of N for the optimal management policy, is $N^* = 4$ units, and the corresponding minimum expected cost is found to be $T_{\text{cost}}(N^*) = \$49.20$. Fig. 4 plots the minimum expected cost $T_{\text{cost}}(N)$ versus $N = 1(1)30$. The plot shows that that the minimum expected cost indeed occurs when $N = 4$.

System characteristics	Notation	Value
Expected number of messages in the system	$E[N_s]$	2.18
Expected length of busy period	$E[B]$	6.71
Expected length of idle period	$E[I]$	10
Expected length of breakdown period	$E[D]$	0.07
Optimal management policy	N^*	4
Minimum expected cost	$T_{\text{cost}}(N^*)$	49.20

Appendix A

```

Appendix
%-----%
% Input system parameter values
% lambda = lambda (arrival rate), mu = mus (service mean),
% os=sigmas (service standard deviation)
% alpha = alpha (breakdown rate), muR = mur (repair mean)
% sigmaR=sigmas (service standard deviation)
% ch = holding cost per unit time for each customer
% co = cost per unit time for keeping the server on
% cf = cost per unit time for keeping the server off
% cb = breakdown cost per unit time for a broken server)
% cs = start-up cost for turning the server on
% L = the upper bound of N considered.
%-----%
function output=NMG1(lambda,mus,sigmas,alpha,mur,
    sigmar,L,ch,co,cf,cb,cs)
hold off
format short;
for th=1:L
rho=lambda*mus;
gamma=alpha*mur;
%-----%
% Calculate system characteristics:
% E[Ns](th) (= ENs(th)), (E[B]/E[C])(th) (= EB(th))
% E[I]/E[C](th) (= EI(th)), (E[D]/E[C])(th) (= ED(th))
% 1/E[C](th) (= ECI(th)), th=1, 2, 3, ... , L.
% Calculate the total expected cost function
% Tcost(th) (= F(th)), th=1,2,3,...,L.
%-----%
nom=lambda^2*(1+gamma)^2*(mus^2+sigmas^2)
+alpha*mus*(mur^2+sigmar^2);
denom=2*(1-rho*(1+gamma));
ENs(th)=(th-1)/2+ rho*(1+gamma) +(nom/denom);
EB(th)=rho;
EI(th)=1-rho*(1+gamma);
ED(th)=rho*(gamma);
ECI(th)=lambda*(1-rho*(1+gamma))/th;
F(th)=ch*ENs(th)+co*EB(th)+cf*EI(th)+cb*ED(th)
+cs*ECI(th);
%-----%
% Stem(X,Y) plots the data sequence Y for given X.
% If F(th)=min(F), mark (th,min(F)) as 'x'.
% Print E[Ns] (expected number of customers in the system)
% E[B] = EB (expected length of busy period)
% E[I] = EI (expected length of idle period)
% E[D] = ED (expected length of breakdown period)
% N = optimal policy on the command window.
% Print the minimum expected cost Tcost.
%-----%
stem(th,F(th))
xlabel('N');
ylabel('Tcost(N)');
title('Total expected cost function: \it{T}_{cost}(N)');
hold on
if (F(th)==min(F))
plot(th,min(F),'r*x')
N=th;
ENs=ENs(th);
EB=EB(th)/ECI(th);
EI=EI(th)/ECI(th);
ED=ED(th)/ECI(th);
end
Tcost=min(F);
Tcost

```

References

- Artalejo, J. R. (1998). Some results on the M/G/1 queue with N-policy. *Asia-Pacific Journal of Operations Research*, 15, 147–157.
- Avi-Itzhak, B., & Naor, P. (1963). Some queueing problems with the service stations subject to breakdowns. *Operations Research*, 11, 303–320.
- Bell, C. E. (1971). Characterization and computation of optimal policies for operating an M/G/1 queueing system with removable server. *Operations Research*, 19, 208–218.
- Bell, C. E. (1972). Optimal operation policies for M/G/1 queueing systems. *Operations Research*, 21, 1281–1289.
- Heyman, D. P. (1968). Optimal operation policies for M/G/1 queueing systems. *Operations Research*, 16, 362–382.
- Kimura, T. (1981). Optimal control of an M/G/1 queueing system with removable server via diffusion approximation. *European Journal of Operations Research*, 8, 390–398.
- Neuts, M. F., & Lucanton, D. M. (1979). A Markovian queue with N servers subject to breakdowns and repairs. *Management Science*, 25, 849–861.
- Shogan, A. W. (1979). A single server queue with arrival rate dependent on server breakdowns. *Naval Research Logistics*, 26, 487–497.
- Sivazlian, B. D., & Stanfel, L. E. (1975). Analysis of system in operations research. Englewood Cliffs, NJ: Prentice-Hall.
- Teghem, J., Jr. (1987). Optimal control of a removable server in an M/G/1 queue with finite capacity. *European Journal of Operations Research*, 31, 358–367.
- Tijms, H. C. (1986). Stochastic modelling and analysis. New York: Wiley.
- Wang, K.-H. (1990). Infinite source M/M/1 queue with breakdown. *Journal of the Chinese Institute of Industrial Engineers*, 7, 47–55.
- Wang, K.-H. (1995). Optimal operation of a Markovian queueing system with a removable and non-reliable server. *Microelectronics and Reliability*, 35, 1131–1136.
- Wang, K.-H., & Ke, J.-C. (2000). A recursive method to the optimal control of an M/G/1 queueing system with finite capacity and infinite capacity. *Applied Mathematical Modelling*, 24, 899–914.
- Wang, K.-H., & Ke, J.-C. (2002). Control policies of an M/G/1 queueing system with a removable and non-reliable server. *International Transactions in Operations Research*, 9, 195–212.
- Yadin, M., & Naor, P. (1963). Queueing systems with a removable service station. *Operational Research Quarterly*, 14, 393–405.