K. Bidhi Singh

Cost analysis of a queueing system with two service facilities

<http://www.numdam.org/item?id=RO_1988__22_1_67_0>

COST ANALYSIS OF A QUEUEING SYSTEM WITH TWO SERVICE FACILITIES (*)

by K. Bidhi Singh (¹)

Abstract. — *This paper studies the economic behaviour of a Markovian queueing system based upon the assumptions: (i) there is a regular service facility serving the units one by one, (ii) search for an additional service facility for the service of a group of units is started when the queue length increases to K (0 < K < M, where M is the maximum waiting space), (iii) the search is dropped when the queue length reduces to a tolerable fixed size M − N, (iv) the time for arranging an additional service facility is a random variable. The model is directed towards finding the optimal operating policy (N, K) for a queueing system with a linear cost structure.*

Keywords : Cost analysis; queueing system.

Résumé. — *Cet article étudie le comportement économique d'un système markovien de files d'attente basé sur les hypothèses suivantes : (1) il existe un serveur régulier servant les unités une à une; (2) on démarre la recherche d'un serveur supplémentaire pour un groupe d'unités quand la longueur de la file d'attente atteint K (0 < K < M où M est l'espace maximal d'attente); (3) on abandonne la recherche quand cette longueur se réduit à une taille fixe tolérable M − N; (4) le temps nécessaire pour mettre à disposition un serveur supplémentaire est une variable aléatoire. L'objectif du modèle est de trouver la stratégie optimale (N, K) pour un système de files d'attente ayant une structure de coût linéaire.*

INTRODUCTION

Some work has been done on the optimal operating policy for a queueing system with linear cost structure. Among the studies made on this phenomenon, a mention may be made of Hillier [1-3], Hillier and Lieberman [4], Morse [9], Sobel [6], Kabak [8] and Stidham, Sh., Jr. [7] etc. Hilliers model 3 a, concerns a $M/M/C$ system in which the decision variables are the number of servers (denoted by C) and the mean service rate of each server (denoted by μ). The objective is to minimize the steady state expected total cost for operation of the system. Morse has shown that the optimal value of C for

this model is 1 and that for any fixed value of the product $C\mu$ (as indicated by the graph), the expected number in the system is minimized by setting C equal to 1. Recently Stidham, Sh. Jr. has shown that a single-server system is optimal for a wide class of arrival processes and service-time distributions, and for a wide variety of waiting cost functions.

These studies relate to the practical situation in a cottage industry godown-cum-sale depot (g. c. s. d.) having a limited space for M units and/or limited resources for the production of M units. Because of the limited waiting space and/or limited resources (i. e. when there is no space in the g. c. s. d. and/or all the ressources are exhausted), the production in the industry has to be stopped, which results in a big loss. In order to avoid the loss, the manager of the industrial concern would like to sell some of the units on a reduced price to another business concern. The manager of the industrial concern plans in such a way that when the number of units in the g. c. s. d. increases to K, $(0 < K < M)$ and/or the resources for the production of $(M-K)$ units are left, a search is started for the sale of N units, the search is stopped when the number of the units in the g. c. s. d. reduces to $M-N$ $(M-N < K < M)$ and/or the resources for the production of N units are left. And if the services of a business concern are procured, N units are supplied to it. The search for another business concern for the sale of N units is again started when the number of units increases to K.

THE MATHEMATICAL MODEL BASED ON G. C. S. D.

1. $M(M > 4)$ is assumed to be the maximum waiting space, i. e. if at any time the queue length is M, an arriving unit is considered lost for the system.

2. Units arrive at the system in a poisson stream with mean rate λ and form a queue.

3. The queue discipline is first come, first served.

4. There is one regular service facility (r. s. f.) serving the units one by one and the service time of a unit is exponentially distributed with parameter μ.

5. When the queue length increases to K, $(K > (M/2) + 1)$ a search for an a. s. f. for the service of N, $(M - K < N < (M/2))$ units is started. The time for arranging an a. s. f. is exponentially distributed with parameter V.

6. The search is dropped when the queue length reduces to $M - N$. If during the search time the a. s. f. becomes available, N units are sent to it for service.

FORMULATION OF EQUATIONS AND THEIR SOLUTION

Define

$P_{n,\,A}(t) \equiv$ Probability that at time t, there are $M - n$ unoccupied waiting spaces in the g. c. s. d., (i. e., n units in the queue) a search for an additional service facility for the service of N units is in progress. $M - N < n \leqq M$.

$P_{n,\,B}(t) \equiv$ Probability that at time t, there are $M - n$ unoccupied waiting spaces in the g. c. s. d. (i. e. n units in the queue) and no search for an a. s. f. is in progress. $0 \leqq n < K$.

$R_n(t) \equiv$ Probability that at time t, there are n units in the queue.

Keeping in view the transition, the state $(K - 1, B)$ will change to the (K, A) through an arrival, the state $(M - N + 1, A)$ will change to the $M - n, B)$ through a service, the state $(n + N, A)$, $M - N < n + N \leqq M$, will change to the (n, B) when an a. s. f. for N unit becomes available.

The Laplace transform of probability generating functions for various queue lenghts are obtained and steady state results are explicitly derived therefrom. In particular, we obtain the explicit expressions for the probabilities, $P_{0,\,B}(t)$, $P_{K-1,\,B}(t)$, $P_{M-N+1,\,A}(t)$ and $P_{M,\,A}(t)$ in the limiting case.

Kolmogorov's forward equations governing the system are:

$$\frac{d}{dt} P_{n,\,A}(t) = -(\lambda + \mu + V) P_{n,\,A}(t) + \mu P_{n+1,\,A}(t)$$

$$+ \lambda P_{n-1,\,A}(t)(1 - \delta_{n,\,M-N+1}) + \lambda P_{K-1,\,B}(t) \cdot \delta_{n,\,K}$$

$$M - N + 1 \leqq n \leqq M - 1 \tag{1}$$

$$\delta_{n,\,M-N+1} = 0 \quad \text{for} \quad n \neq M - N + 1$$
$$= 1 \quad \text{for} \quad n = M - N + 1$$

$$\delta_{n,\,k} = 0 \quad \text{for} \quad n \neq K$$
$$= 1 \quad \text{for} \quad n = K$$

$$\frac{d}{dt} P_{M,\,A}(t) = -(\mu + V) P_{M,\,A}(t) + \lambda P_{M-1,\,A}(t), \qquad n = M \tag{2}$$

$$\frac{d}{dt} P_{0,\,B}(t) = -\lambda P_{0,\,B}(t) + \mu P_{1,\,B}(t), \qquad n = 0 \tag{3}$$

$$\frac{d}{dt} P_{n,\,B}(t) = -(\lambda + \mu) P_{n,\,B}(t) + \lambda P_{n-1,\,B}(t) + \mu P_{n+1,\,B}(t) \tag{4}$$

$$0 < n \leqq M - 2N$$

$$\frac{d}{dt} P_{n,\,B}(t) = -(\lambda+\mu) P_{n,\,B}(t) + \lambda P_{n-1,\,B}(t)$$

$$+ \mu P_{n+1,\,B}(t) + V P_{n+N,\,A}(t) + \mu P_{M-N+1,\,A}(t) \delta_{n,\,M-N} \tag{5}$$

$$M-2N+1 \leq n \leq M-N$$

$$\delta_{n,\,M-N} = 0 \quad \text{for} \quad n \neq M-N$$
$$= 1 \quad \text{for} \quad n = M-N$$

$$\frac{d}{dt} P_{n,\,B}(t) = -(\lambda+\mu) P_{n,\,B}(t) + \lambda P_{n-1,\,B}(t) + \mu P_{n+1,\,B}(t)(1-\delta_{n,\,K-1})$$

$$M-N+1 \leq n \leq K-1 \tag{6}$$

$$\delta_{n,\,K-1} = 0 \quad \text{for} \quad n \neq K-1$$
$$= 1 \quad \text{for} \quad n = K-1$$

Let the time reckoned from the instant when the queue length is zero, so that the initial condition becomes

$$P_{0,\,B}(0) = 1 \tag{7}$$

Define the probability generating functions

$$P_A(t, \alpha) = \sum_{n=M-N+1}^{M} \alpha^n P_{n,\,A}(t) \tag{8}$$

$$P_B(t, \alpha) = \sum_{n=0}^{K-1} \alpha^n P_{n,\,B}(t) \tag{9}$$

$$R(t, \alpha) = \sum_{n=0}^{M} \alpha^n R_n^{(t)} \tag{10}$$

Let $\bar{F}(s)$ denote the Laplace transform (L.T.) of $F(t)$ defined by

$$\bar{F}(s) = \int_0^\infty e^{-st} F(t)\, dt$$

Multiplying (1)-(6) by appropriate powers of α, using (8)-(9) and taking L.T. s, we get

$$\bar{P}_A(s, \alpha) = \frac{\left\{ \begin{array}{c} \lambda \alpha^{M+1}(1-\alpha) \bar{P}_{M,\,A}(s) - \mu \alpha^{M-N+1} \bar{P}_{M-N+1,\,A}(s) \\ + \lambda \alpha^{K+1} \bar{P}_{K-1,\,B}(s) \end{array} \right\}}{\alpha'(\lambda+\mu+V+s) - \mu - \lambda \alpha^2} \tag{11}$$

$$[\alpha(\lambda+\mu+V+s)-\mu-\lambda\alpha^2][\alpha-\mu(1-\alpha)\bar{P}_{0,B}(s)]$$
$$+\lambda V_\alpha^{M-N+2}(1-\alpha)\bar{P}_{M,A}(s)$$
$$+[\alpha(\lambda+\mu+V+s)-\mu-\lambda\alpha^2-V\alpha^{-N+1}]$$
$$\times[\mu\alpha^{M-N+1}\bar{P}_{M-N+1,A}(s)-\lambda\alpha^{K+1}\bar{P}_{K-1,B}(s)]$$

$$\cdot\ \bar{P}_B(s,\alpha)=\frac{}{[\alpha(\lambda+\mu+V+s)-\mu-\lambda\alpha^2][\alpha(\lambda+\mu+s)-\mu-\lambda\alpha^2]}\quad(12)$$

$$\bar{R}(s,\alpha)=\bar{P}_A(s,\alpha)+\bar{P}_B(s,\alpha)$$

The denominator in (12) has four roots in α, since $\bar{P}_B(s,\alpha)$ is a polynomial, these roots must vanish the numerator in (12), giving rise to four equations involving four unknowns namely $\bar{P}_{0,B}(s)$, $\bar{P}_{M-N+1,A}(s)$, $\bar{P}_{K-1,B}(s)$ and $\bar{P}_{M,A}(s)$. Solving these four equations, we can determine all the unknowns. Thus $\bar{P}_A(s,\alpha)$, $\bar{P}_B(s,\alpha)$ and $\bar{R}(s,\alpha)$ are completely determined.

STEADY STATE SOLUTION

The steady state solution can be obtained by the well known property of L.T. viz.

$$\lim_{s\to0}\ S\bar{F}(s)=\lim_{t\to\infty}\ F(t)=F\quad(13)$$

if the limit on the right hand side exists, thus if

$$P_A(\alpha)=\sum_{n=M-N+1}^{M}\alpha^n P_{n,A}$$

$$P_B(\alpha)=\sum_{n=0}^{K-1}\alpha^n P_{n,B}$$

$$R(\alpha)=\sum_{n=0}^{M}\alpha^n R_n$$

Applying the property (13) to equations (11), (12), we have

$$P_A(\alpha)=\frac{\left\{\begin{array}{c}\rho_1\rho_2\alpha^{M+1}(1-\alpha)P_{M,A}-\rho_2\alpha^{M-N+1}P_{M-N+1,A}\\+\rho_1\rho_2\alpha^{K+1}P_{K-1,B}\end{array}\right\}}{\alpha(\rho_1+\rho_2+\rho_1\rho_2)-\rho_2-\rho_1\rho_2\alpha^2}\quad(14)$$

$$-(1-\alpha)[\alpha(\rho_1+\rho_2+\rho_1\rho_2)-\rho_2-\rho_1\rho_2\alpha^2]P_{0,\,B}$$
$$+\rho_1^2(1-\alpha)\alpha^{M-N+2}P_{M,\,A}$$
$$+[\alpha^{M-N+1}P_{M-N+1,\,A}-\rho_1\alpha^{K+1}P_{K-1,\,B}]$$

$$P_B(\alpha)=\frac{[\alpha(\rho_1+\rho_2+\rho_1\rho_2)-\rho_2-\rho_1\rho_2\alpha^2-\rho_1\alpha^{-N+1}]}{\rho_1\rho_2(1-\alpha)(\alpha\rho_1-1)(\alpha-\alpha_1)(\alpha_2-\alpha)} \qquad (15)$$

Where $\rho_1=\lambda(1/\mu)\equiv$ Mean number of arrivals during the mean service time of the regular service channel.

$\rho_2=\lambda(1/\bar{V})\equiv$ Mean number of arrivals during the mean procurability time of the random additional service facility.

$$\alpha_1+\alpha_2=\frac{\rho_1+\rho_2+\rho_1\rho_2}{\rho_1\rho_2},\ \alpha_1\alpha_2=\frac{1}{\rho_1} \qquad (16)$$

$$R(\alpha)=P_A(\alpha)+P_B(\alpha)$$

$P_B(\alpha)$ is a polynomial, the roots of the denominator in $P_B(\alpha)$ must vanish its numerator, giving rise to the set of three equations. Solving these equations, we have

$$P_{M,\,A}=\frac{\alpha_1^K\alpha_2^{M-N}-\alpha_1^{M-N}\alpha_2^K}{\rho_1[\alpha_1^K\alpha_2^M(1-\alpha_2)-\alpha_1^M\alpha_2^K(1-\alpha_1)]}P_{M-N+1,\,A} \qquad (17)$$

$$P_{K-1,\,B}=\frac{\alpha_1^M\alpha_2^M[\rho_1(\alpha_1^{-N}-\alpha_2^{-N})-(\alpha_1^{-N-1}-\alpha_2^{-N-1})]}{\rho_1^2[\alpha_1^K\alpha_2^M(1-\alpha_2)-\alpha_1^M\alpha_2^K(1-\alpha_1)]}P_{M-N+1,\,A} \qquad (18)$$

$$P_{0,\,B}=\frac{1}{(1-\rho_1)\rho_1^{M-N+1}}$$
$$\times\left[\frac{(1-\rho_1)(\alpha_1^K\alpha_2^{M-N}-\alpha_1^{M-N}\alpha_2^K)}{\alpha_1^K\alpha_2^M(1-\alpha_2)-\alpha_1^M\alpha_2^K(1-\alpha_1)}-\rho_1(1-\rho_1^N),\right.$$
$$+\frac{\alpha_1^M\alpha_2^M\rho_1^{M-N-K}(1-\rho_1^N)}{\alpha_1^K\alpha_2^M(1-\alpha_2)-\alpha_1^M\alpha_2^K(1-\alpha_1)}$$
$$\left.\times\rho_1(\alpha_1^{-N}-\alpha_2^{-N})-(\alpha_1^{-N-1}-\alpha_2^{-N-1})\right]P_{M-N+1,\,A} \qquad (19)$$

The normalizing condition

$$R(1)=P_A(1)+P_B(1)=1$$

gives

$$(1-\rho_1) = P_{0,\;B} - \rho_1 P_{M,\;A} - N(\rho_1 P_{K-1,\;B} - P_{M-N+1,\;A}). \tag{20}$$

Thus all the four unknowns $P_{M,\;A}$, $P_{K-1,\;B}$, $P_{0,\;B}$ and $P_{M-N+1,\;A}$ can be explicitly determined from (17)-(20). The probability $P_A(1)$ of the procurement of an a. s. f. is ensured by setting $\alpha = 1$ in (14).

$$P_A(1) = \frac{\rho_1 \rho_2 P_{K-1,\;B} - \rho_2 P_{M-N+1,\;A}}{\rho_1} \tag{21}$$

BALANCE BETWEEN SERVICE COST OF AN A.S.F. AND CUSTOMERS LOST

Let T_M be the fraction of the customers who go elsewhere for the $M/M/1$ model with limited space for M units, it is given by

$$T_M = \frac{\rho_1^M (1-\rho_1)}{(1-\rho_1^{M+1})}$$

Where ρ_1 is the traffic intensity,

$C_1 \equiv$ cost of providing an additional service facility per unit time per customer.

$C_2 \equiv$ Loss or cost per unit time per customer when the queue length is maximum.

$NC_1 VP_A(1) =$ Expected cost of providing an additional service facility per unit time.

$C_2 \lambda (T_M - P_{M,\;A}) =$ Expected gain due to the provision of an additional service facility per unit time.

Employing an a. s. f. is profitable if

$$NC_1 VP_A(1) < C_2 \lambda (T_M - P_{M,\;A})$$

$$C^* < \frac{\rho_2 (T_M - P_{M,\;A})}{NP_A(1)} \quad \text{where} \quad C^* = \frac{C_1}{C_2}.$$

Upper bounds for C^* are given by

$$UC^*(M, \rho_1, \rho_2, N, K) = \frac{\rho_2 (T_M - P_{M,\;A})}{NP_A(1)}.$$

It is obvious that if $C^* > UC^*$ $(M, \rho_1, \rho_2, N, K)$ for a given value of $(M,$ $\rho_1, \rho_2, N, K)$, then certainly we are going to lose by employing an a. s. f. If $C^* = UC^*$ $(M, \rho_1, \rho_2, N, K)$ then there would be no profit, but the provision of an a. s. f. would help in increasing the production. Thus the provision of an a, s. f. is profitable if

$$C^* \leqq UC^* (M, \rho_1, \rho_2, N, K)$$

The optimization has been done in two phases. In the first phase we find for given value of (M, ρ_1, ρ_2) the least value of UC^* $(M, \rho_1, \rho_2, N, K)$ over all feasible values of N and K and the least of the upper bounds of UC^* $(M, \rho_1, \rho_2, N, K)$ is denoted as

$$LUC^* (M, \rho_1, \rho_2) = \min_{N, K} UC^* (M, \rho_1, \rho_2)$$

If we take $C^* \leqq LUC^*$ (M, ρ_1, ρ_2), then the provision of an a. s. f. is profitable for all values of N and K.

In the second phase we minimize the expected total cost.

$$E (T.C.) = N. \ LUC^* (M, \rho_1, \rho_2) [P_A (1)] + \rho_2 P_{M, A}$$

over all feasible values of N and K.

Given

$$M = 20, \qquad \rho_1 = 0.3, \ 0.5, \ 0.7, \ 0.9,$$

$$\rho_2 = 0.3, \ 0.5, \ 0.7, \ 0.9,$$

the optimal values of N and K for different values of ρ_1 and ρ_2 are given in Table.

DISCUSSION OF TABLE

The table indicates the optimal size N of the units to be sold on reduced price and the optimal size K of the inventory/resources left for running the cottage industry on long term planning for different service levels. The table also indicates the following:

(1) For fixed ρ_1 or ρ_2.

As ρ_2 or ρ_1 increases the optimal batch of the units sold on reduced price increases, and the optimal inventory/resources left for running the cottage

Optimal values of N and K for M = 20 $\overline{N/K}$

ρ_1	0.3	0.5	0.7	0.9
0.3.	5/16	6/15	7/14	8/13
0.5.	6/15	7/14	8/13	9?12
0.7.	6/15	8/13	9/12	9/12
0.9.	7/14	8/13	9/12	9/12

industry on long term planning (the optimal queue size, where the search for the random additional service facility is started) decreases.

(2) For increase in both ρ_1 and ρ_2, the optimal size of the units sold on reduced price increases and the optimal size of the inventory decrease. Both stabilize for higher values of ρ_1 and ρ_2.

Graph shows the objective function as a function of service levels. It clearly indicates that when either ρ_1 or ρ_2 or both increase, the total expected cost also increases

ACKNOWLEDGEMENT

REFERENCES

1. F. S. HILLIER, *Economic Models for Industrial Waiting Line Problems*, Management Sci., Vol. 10, (1), 1963, pp. 119-130.
2. F. S. HILLIER, *The Application of Waiting Line Theory to Industrial Problems*, J. Indust. Engg., Vol. 15, (1), 1964, pp. 3-8.
3. F. S. HILLIER, *Cost Models for the Application of Priority Waiting Line Theory to Industrial Problems*, J. Indust. Engg., Vol. 16, (3), 1965, pp. 178-185.
4. F. S. HILLIER, and G. J. LIEBERMAN, *Introduction to Operations Research*, San Francisco, Holden-Day, 1967.
5. D. HEYMAN, *Optimal Operating Policies for Stochastic Service Systems*, Opns. Res., Vol. 16, 1968, pp. 362-382.
6. M. G. SOBEL, *Optimal Average Cost Policy for a Queue with Start-Up and Shut-Down Costs*, Opns. Res., Vol. 17, 1969, pp. 145-162.
7. S. STIDHAM, Jr., *On the Optimality of Single Server Queueing System*, Opns. Res., Vol. 18, 1970, pp. 708-732.
8. KABAK, W. IRWIN, *Optimal Allocations to M/G/1: ∞ Queueing Systems*, OPSE-ARCH, Vol. 12, sept-Déc. 1975.
9. P. MORSE, *Queues, Inventories and Maintenance*, Wiley, New York, 1958.

10^8

10^7

$P_1 = 0.9$

10^6

$P_1 = 0.7$

10^5

$P_1 = 0.5$

10^4

10^3

10^2

10^1

$P_1 = 0.3$

OBJECTIVE FUNCTION $\times 10^{13}$

0.2   0.3   0.4   0.5   0.6   0.7   0.8   0.9

EXPECTED NUMBER OF ARRIVALS DURING MEAN SEARCH TIME