Searching for Optimum Service Rate In a Queuing System

N.N.Elkhadar, M.M.Mekaeil*

ABSTRACT

The use of queuing theory in practice involves two major aspects:

1. Selection of the appropriate mathematical model that will represent the real system adequately with the objective of determining the systems measures of performance.

2. Implementation of a decision model based on the systems measures of performance for the purpose of designing the service facility.

This work is merely a simulation study using a celebrated simulation programs in operations research namely TORA, LINGO, LINDO, Process Model plus EXCEL.

In this study we invstigated the optimal service rate in the queueing system (M/M/1) : (GD/∞ /∞) , plus finding the optimal service rate for a given arrival rate , several service rates are tried to obtain the performance measurements which are important in determining the minimal cost by using the cost function

$$TC(\mu) = C_S \mu + C_w L_s \ .$$

Promising results are obtained which are consistent with all the theoretical results available in the literature concerning the steady state behavior of a queueing system.

*Statistics Department, Faculty Of Science, Benghazi University ,Benghazi Libya

Key words: arrival rate (λ); service rate (μ) ; steady-state probability (p_n); customers; server.

Introduction

Waiting lines, or queues, are a common occurrence in our everyday lives. All of us have waited in line to purchase a ticket at a sporting event or waited to pay for groceries at a supermarket. Waiting lines are also common in a variety of operations management situations. Waiting lines form whenever the arrivals, or demand, for service from a facility exceeds the capacity of that facility. Queues can be deterministic; that is, arrival intervals and service times are constant. If the arrivals and service times are not constant, as is often the case in production systems, the queue is a realization of a stochastic process. Deterministic queues are easy to handle mathematically, the more difficult cases of stochastic processes. Most queuing results are based on a family of distributions containing the negative exponential for service rates and the Poisson for arrival rates because these distributions have the useful property of a single parameter and are often associated with random events

Queuing models with Poisson input—exponential service

The state probabilities are found by formulating and solving balance equations that equate the rate at which transitions occur into a state (or group of states) to the rate at which transitions occur out of the state (or group of states). This method relies on the memory less property of the exponential distribution and does not apply to systems that do not have exponential service times and Poisson arrivals

A more general model have exponential service times and Poisson arrivals, but allow λ and μ to be different for each state.

Infinite queue-infinite source, single-server model (M/M/ 1): (GD/∞/∞)

Assume

(a) The average arrival rate is constant; $\lambda_n = \lambda$ for all n.

(b) The average service rate is constant; $\mu_n = \mu$ for all n

(c) The average arrival rate is less than the average service rate; $\lambda < \mu$.

(d) The transition rates can be summarized in a diagram, as in Fig. (taken from Hillier and Lieberman [1974]). This leads to the following set of balance equations:

Balance Equations for (M/M/1):(GD/∞/∞) Queue

Line

0,1 $\quad \mu P_1 = \lambda P_0$

1,2 $\quad \mu P_2 = \lambda P_1$

2,3 $\quad \mu P_3 = \lambda P_2$

⋮

$n, n+1 \quad \mu P_{n+1} = \lambda P_n$ (1)

The general relationship between state probabilities should be apparent:

$$P_n = \frac{\lambda}{\mu} P_{n-1} = \rho P_{n-1} \quad (2)$$

Furthermore, it should also be apparent that all of the state probabilities can be expressed in terms of a single state probability:

$$P_n = \rho^n P_0 \qquad (3)$$

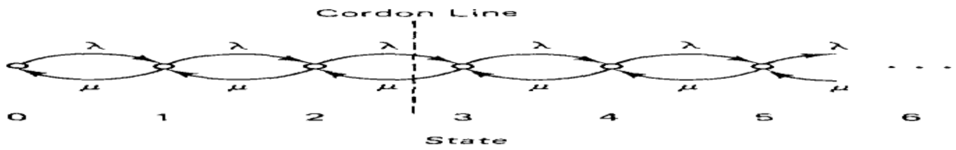


Figure 1 Transition rate diagram for the (M/M/1):(GD/ ∞ / ∞) queue.

All that is needed to complete the calculations for P_n is to determine P_0, this is easily accomplished by noting one additional characteristic of P_n

To be a proper probability distribution, the sum of the state probabilities must equal 1. That is

$$\sum_{n=0}^{\infty} P_n = 1 \qquad (4)$$

Therefore, substitution of Eq. (3) into Eq. (4) gives

$$\sum_{n=0}^{\infty} \rho^n P_0 = 1 \qquad (5)$$

Equation (5) is the sum of a geometric series, then we can write Eq.(5) as

$$P_0 \frac{1}{1-\rho} = 1 \ \rightarrow P_0 = 1 - \rho \qquad \rho < 1 \qquad (6)$$

Substitution of P_0 in Eq. (3) provides all of the state probabilities:

$$P_n = (1-\rho)\rho^n \qquad \rho < 1 \qquad (7)$$

Performance measures:

The expected number of customers in the system and the expected number of customers in the queue are found Little's formula

$$L_s = \sum_{n=0}^{\infty} n(1-\rho)\rho^n = (1-\rho)\sum_{n=0}^{\infty} n\rho^n = \frac{\rho}{1-\rho} \qquad \rho < 1 \qquad (8)$$

$$L_q = L_s - \rho = \frac{\rho}{1-\rho} - \frac{\rho(1-\rho)}{1-\rho} = \frac{\rho^2}{1-\rho} \qquad \rho < 1 \qquad (9)$$

The expected time in queue and time in system are not uniquely defined by ρ ; they also depend on the arrival rate. From Little's formula

$$W_s = \frac{1}{\lambda}\frac{\rho}{1-\rho} \qquad \rho < 1 \qquad (10)$$

$$W_q = \frac{1}{\lambda}\frac{\rho^2}{1-\rho} \qquad \rho < 1 \qquad (11)$$

Optimum service rate μ

Consider the single-channel model with arrival rate λ and service rate μ presented in Infinite queue-infinite source, single-server model (M/M/1):(GD/∞/∞)

if C_s is the cost per unit increase in μ per unit time and C_w is the cost of waiting per unit waiting time per customer and TC(μ) is the expected cost of waiting and service per unit time given μ.

Thus

$$TC(\mu) = C_s\mu + C_w L_s$$

$$= C_s\mu + C_w \frac{\lambda}{\mu-\lambda} \qquad (12)$$

Notice that cost of service per unit time is directly proportional to μ , and that cost of waiting per unit time is equal to the expected number of customers in the system, multiplied by the waiting cost per customer per unit time.

The minimum cost service rate can be found by finding the effect upon TC(μ) of a small change in μ and forcing this effect to approach zero.

$$\frac{\Delta TC(\mu)}{\Delta\mu} = C_s - C_w[\,\lambda(\mu-\lambda)^{-2}]$$

$$C_s - C_w[\,\lambda(\mu-\lambda)^{-2}] = 0$$

$$\frac{\lambda C_w}{(\mu-\lambda)^2} = C_s$$

$$\frac{1}{(\mu-\lambda)^2} = \frac{C_s}{\lambda C_w}$$

$$(\mu-\lambda)^2 = \frac{\lambda C_w}{C_s}$$

$$\mu-\lambda = \sqrt{\frac{\lambda C_w}{C_s}} \qquad (13)$$

Thus optimum μ is given by

$$\mu = \lambda + \sqrt{\frac{\lambda C_w}{C_s}}\,. \qquad (14)$$

Simulation study

we simulated data from the model (M/M/1) : (GD/∞/∞) and obtained the steady state probabilities and hence the performance measures were calculated (See, table 1.1, figure1.1)

Table 1.1 State probabilities for model (M/M/1) : (GD/∞/∞) with λ=4 and μ=6.4

n	Probability, p_n	Cumulative, P_n	n	Probability, p_n	Cumulative, P_n
0	0.37500	0.37500	12	0.00133	0.99778
1	0.23438	0.60938	13	0.00083	0.99861
2	0.14648	0.75586	14	0.00052	0.99913
3	0.09155	0.84741	15	0.00033	0.99946
4	0.05722	0.90463	16	0.00020	0.99966
5	0.03576	0.94040	17	0.00013	0.99979
6	0.02235	0.96275	18	0.00008	0.99987
7	0.01397	0.97672	19	0.00005	0.99992
8	0.00873	0.98545	20	0.00003	0.99995
9	0.00546	0.99091	21	0.00002	0.99997
10	0.00341	0.99432	22	0.00001	0.99998
11	0.00213	0.99645			

The above step was done for different values of service rate μ and fixed arrival rate λ (See, table 1.2 and figures 1.2_1.6).

Table 1.2.The performance measures for model (M/M/1) : (GD/ ∞/∞) .

λ	μ	P0	Ls	Lq	Ws	Wq
4.00000	4.20000	0.04762	20.00000	19.04762	5.00000	4.76190
4.00000	4.50000	0.11111	8.00000	7.11111	2.00000	1.77778
4.00000	4.70000	0.14894	5.71429	4.86322	1.42857	1.21581
4.00000	5.00000	0.20000	4.00000	3.20000	1.00000	0.80000
4.00000	5.30000	0.24528	3.07692	2.32221	0.76923	0.58055
4.00000	5.50000	0.27273	2.66667	1.93939	0.66667	0.48485
4.00000	5.80000	0.31034	2.22222	1.53257	0.55556	0.38314
4.00000	6.00000	0.33333	2.00000	1.33333	0.50000	0.33333
4.00000	6.20000	0.35484	1.81818	1.17302	0.45455	0.29326
4.00000	6.40000	0.37500	1.66667	1.04167	0.41667	0.26042
4.00000	6.80000	0.41176	1.42857	0.84034	0.35714	0.21008
4.00000	7.00000	0.42857	1.33333	0.76190	0.33333	0.19048
4.00000	7.50000	0.46667	1.14286	0.60952	0.28571	0.15238
4.00000	8.00000	0.50000	1.00000	0.50000	0.25000	0.12500
4.00000	8.30000	0.51807	0.93023	0.44830	0.23256	0.11208
4.00000	8.50000	0.52941	0.88889	0.41830	0.22222	0.10458
4.00000	9.00000	0.55556	0.80000	0.35556	0.20000	0.08889
4.00000	9.50000	0.57895	0.72727	0.30622	0.18182	0.07656
4.00000	10.00000	0.60000	0.66667	0.26667	0.16667	0.06667
4.00000	10.50000	0.61905	0.61538	0.23443	0.15385	0.05861

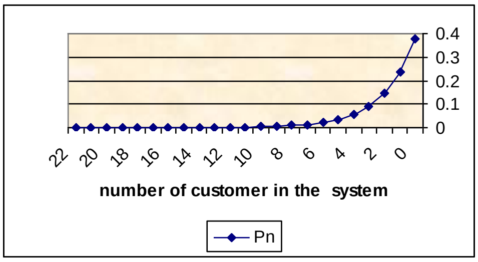


Figure 1.1: The probability of number customers in system.

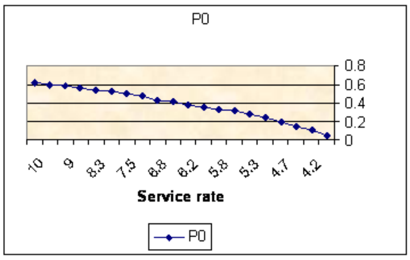
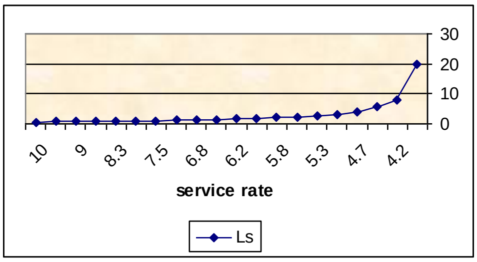


Figure 1.2: The probability the servers not busy (no customers in system)

.



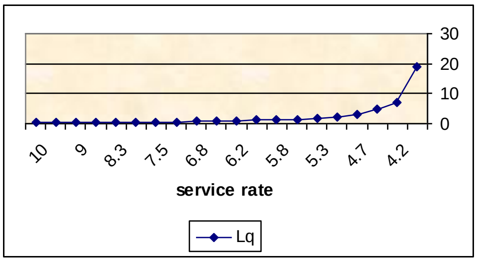
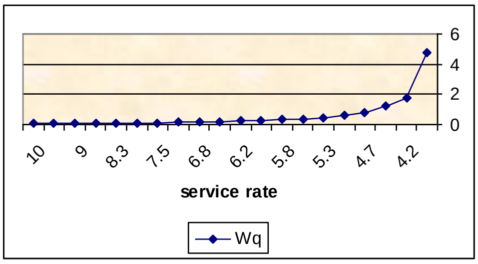


Figure 1.3: Expected number of customers in system.

Figure 1.4: Expected number of customers in queuing.



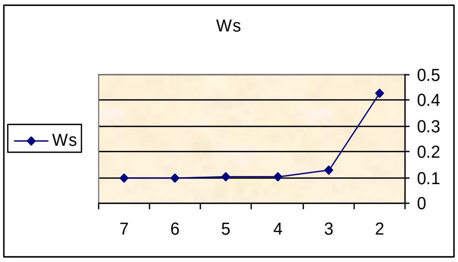


Figure 1.5: Expected waiting time in queuing.

Figure 1.6: Expected waiting time in system.

The values of performance measures (L_s, L_q) and μ obtained in the previous step were used to simulate data from the cost function and cost of the increase in service rate per unit time (Cs =5) and customer waiting cost per unit time (C_w =7) and obtained the optimal service rate of μ = 6.4, which occurred at the possible minimal cost (See, table 1.3, figure 1.7) .

Table 1.3.The simulation data from the cost function of (M/M/1): (GD/∞/∞).

μ	Ls	μ Cs	Cw Ls	TC(μ)
4.20000	20.0000	21	140	161
4.50000	8.00000	22.5	56	78.5
4.70000	5.71429	23.5	40.00003	63.50003
5.00000	4.00000	25	28	53
5.30000	3.07692	26.5	21.53844	48.03844
5.50000	2.66667	27.5	18.66669	46.16669
5.80000	2.22222	29	15.55554	44.55554
6.00000	2.00000	30	14	44
6.20000	1.81818	31	12.72726	43.72726
6.40000	1.66667	32	11.66669	43.66669
6.80000	1.42857	34	9.99999	43.99999
7.00000	1.33333	35	9.33331	44.33331
7.50000	1.14286	37.5	8.00002	45.50002
8.00000	1.00000	40	7	47
8.30000	0.93023	41.5	6.51161	48.01161
8.50000	0.88889	42.5	6.22223	48.72223
9.00000	0.80000	45	5.6	50.6

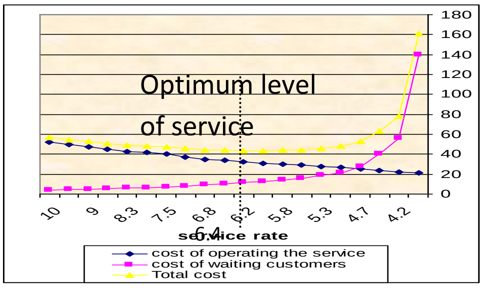


Figure 1.7: A cost function for optimum service rate.

Discussion and Conclusions

The simulated data from the model (M / M / 1) : (GD/ ∞ / ∞) ,

Seems to suggest that the probability of n customers in the system decreases with the increase of number of people in the system .

Moreover, our simulation results appear to indicate that the probability of non-existence of any customer in the system (no service) becomes slightly larger as the service rate increases.

It is also evident from our results that the expected number of people and waiting time in both the queue and system increase with the decrease of the service rate.

As far as the cost function is concerned our results show that the optimal service rate attains at minimal cost which is as anticipated consistent with mathematical relationship

$$\mu = \lambda + \sqrt{\frac{\lambda C_w}{C_s}}$$

The significance the of above results is that when studying a real system it is possible to locate the optimal service rate for given arrival rate, service level cost (C_s), waiting cost for the service (C_w).

References

1. Bose, K.S. (2002); An introduction to the queueing system, Springer, New York.
2. Carter, W. M. Price, C. C. (2000) ; A practical introduction in operations research, CRC Press Boca Raton.
3. Cooper, R. E. (1980); Introduction to queueing theory, Elsevier-North Holland Publishing, New York.
4. Ecker, G. J. (2004); An introduction to operations research, Michael kupferschmid Krieger publication, New York.
5. Gillett, B. E. (1976); Introduction to operations research, McGraw-Hill. New York
6. Hall, W. R. (1991); Queueing methods for service and manufacturing, Englewood Cliffs, Hall,Inc.
7. Hillier, F. S. and Lieberman G. J. (1974); Introduction to operations research, McGraw-Hill, London.
8. Taha, A .H. (2003); Introduction to operations research, Macmillan CO., New Jersey.
9. Wagner, H . M. (1969); Principles of operations research , Prentice Hall , Englewood Cliffs, New Jersey.
10. Wayne, L . W. (2004); Applications and algorithms in operations research, Thomson Learning, New York.