Analysis of Multi-Server Single Queue System with Multiple Phases

Ekpenyong, Emmanuel John
Department of Mathematics, Statistics and Computer Science
Michael Okpara University of Agriculture, Umudike, Nigeria
ekpesstat@yahoo.com

Udoh, Nse Sunday
Department of Mathematics and Statistics
University of Uyo, Nigeria
nseudoh07@yahoo.com

Abstract

This paper extends and improves on the performance measures of the Single-Server Single Queue System with Multiple Phases. The extension results in a new Queuing System of Multi-Server with Multiple Phases under the conditions of First Come First Served, infinite population source, Poisson arrivals and Erlang service time. Queuing properties such as expected total service time, its variance and some performance measures like the expected number of phases in the system, expected number of phases in the queue, expected number of customers in the queue, expected waiting times in the queue and in the system as well as the number of customers in the system have been derived in this work for this Multi-Server with Multiple Phases ($M/E_k/s$: (∞/FCFS)) queuing model with *k* identified stages in series. These performance measures so obtained were compared with those of the already existing Single-Server with Multiple Phases ($M/E_k/1$: (∞/FCFS) model. Numerical illustration indicates the efficiency and effectiveness of the latter over the former.

Keywords: Multi-server, Single queue, Multiple phase queue model, Erlang distribution, Performance measures.

1.0 Introduction

In many areas of real life, for instance, in the multi-specialty outpatient clinic, patients first form the queue for registration and then is triaged for assessment, then for diagnostics, review, treatment, intervention or prescription and finally exit from the system or triage to different providers. This describes a multi-queue multi-server model. We also have cases in a queue where customers are served in multi-channels and in stages at different points. Many queue models have been developed to solve varying congestion or queue problems; but this cannot be achieved because of the varying and stochastic characteristics of queues. In this light, more models still need to be developed to help solve queue problems on the bases of their features. Consequently, this paper develops a new queue model referred to as a multi-server single queue system with each service channel having *k* identical phases in series, (see figure 1.0), with Poisson arrivals and Erlang service time ($M/E_k/s:(\infty/FCFS)$), which is an extension of the $M/E_k/1:(\infty/FCFS)$ model with multiple service channels; see Sharma (2007) and Gupta (2008).

Maurya (2009) analyzed multiple service channels queuing with practical situations arising at those places where phase service is provided under priority queue discipline. He added that the models become more significant when the source of input customers is to be classified into two or more categories. Dimitriou and Langaris (2010) analyze a repairable queuing model with a two-phase service in succession provided by a single server. They considered a case where customers arrive in a single queue and after the completion of the first phase service, either proceed to the second phase or join a retrial box from where they retry after a random amount of time and independently of the other customers in orbit, to find a position for service in the second phase. Moreover, the server is subject to breakdowns and repairs in both phases, while a start-up time is needed in order to start serving a retrial customer. When the server upon a service or a repair completion finds no customers waiting to be served, he departs for a single vacation of an arbitrarily distributed time. Luh (2004) considered a queuing model of general servers in tandem with finite buffer capacities. He studied the probability of blocking and in order to obtain the steady state probability distribution of this model, he constructed an embedded Markov chain at the departure points. He further solved the model analytically and its analysis is extended to semi-Markovian representation of performance measures in queuing networks.

2.0 Method

The Model: The model is made up of 's' multiple service channel with k identical stages in series, each with average service time of $\frac{1}{s\mu}$. The distribution of total servicing time of customers in the system is some joint distribution of time in all these stages.

Customers arrive in a single queue and a set of them, whose number is based on the number of servers in each phase, enters the system to be served in the first phase before proceeding to the second phase, up to the kth phase. The assumptions are that each set of customers is served in k-phases set-by-set and a new service does not start until all k-phases have been completed. Moreover, the queue discipline is first come first served with infinite source. The arrivals follow a Poisson distribution and the service times follow Erlang distribution. Symbolically, the model can be expressed as $M/E_k/s: (\infty/FCFS)$. The queue facility is represented in figure 1.0 (see Reid and Sanders, 2010).

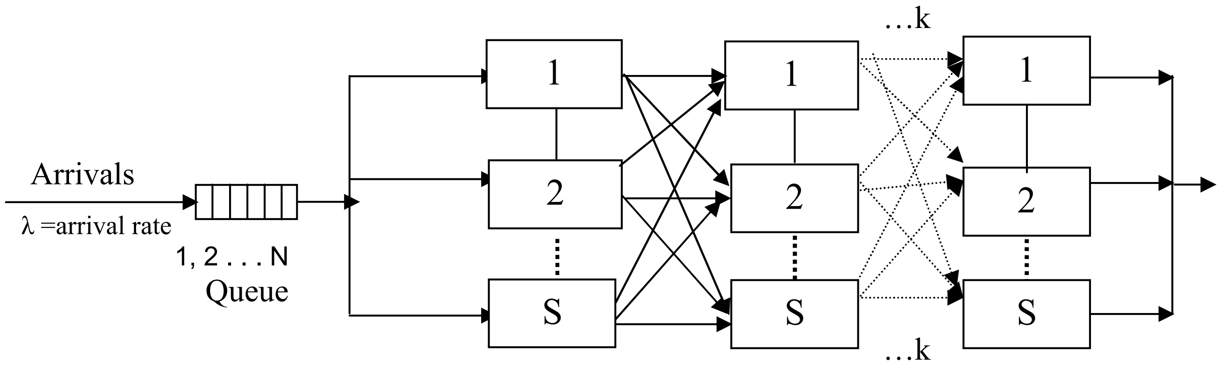

Figure 1.0: Model of single queue with S servers in k-phases

3.0 Determination of Expected Total Service Time and its Variance

Let $s\mu$ denote the number of customers served per unit time, then $ks\mu$ will be the number of phases served per unit time. The probability density function for Erlang distribution is;

$$f(t) \;=\; \frac{(ks\mu)^k}{(k-1)!}\, t^{k-1}\, e^{-ks\mu t}, \quad t \geq 0$$

$$E(T) \;=\; \int_o^{\infty} t\, f(t)\, dt$$

$$= \int_o^{\infty} t\, \frac{(ks\mu)^k}{(k-1)!}\, t^{k-1}\, e^{-ks\mu t}\, dt \;=\; \frac{1}{s\mu}$$

with variance $\dfrac{1}{k(s\mu)^2}$

4.0 Derivation of Performance Measures of the M/E$_K$/s:(∞/FCFS) Model

Consider $P_n\,(t)$ to be the probability that there are *n* customers in the *k* phases in the system at time *t*. Let $P_n(t+\Delta t)$ be the probability that there are n customers in the system, $\lambda \Delta t$ be the probability that a customer is in the queue at time Δt, $\mu \Delta t$ the probability that a customer is served and $ks\mu\Delta t$ is the probability that the customers in the s service points in all the phases. $P_{n+1}(t)$ and $P_{n-1}(t)$ are probabilities that there are n+1 and n-1 customers in the system respectively. The probability that the system will contain n customers at time (t+Δt) in the k-phases can be expressed as the sum of joint probabilities of mutually exclusive and collectively exhaustive cases. The difference equations for the model are given as follows:

$$P_n(t+\Delta t) = -(\lambda + ks\mu)P_n(t)\,\Delta t + ks\mu P_{n+1}(t)\,\Delta t + P_{n-k}(t)\,\Delta t + P_n(t) + (\Delta t)^2,\; n \geq 1 \quad \ldots(4.1)$$

$$P_o(t+\Delta t) = P_o(t)[1-\lambda\Delta t] + P_1(t)\, ks\mu\Delta t,\; n = o \quad \ldots(4.2)$$

where λ is the arrival rate and μ is the service rate.

The steady state system of difference equations are obtained as:

$$-(\lambda + ks\mu)\pi_n + ks\mu\pi_{n+1} + \lambda\pi_{n-k} = 0; \;\; n > 0 \qquad \ldots (4.3)$$

$$-\lambda\pi_0 + ks\mu\pi_1 = 0; n = 0 \qquad \ldots (4.4)$$

$Let \; \rho = \frac{\lambda}{ks\mu}$ and dividing (4.3) and (4.4) by $ks\mu$, we have:

$$(1+\rho)\pi_n = \rho\pi_{n-k} + \pi_{n+1}; n \geq 1 \qquad (4.5)$$

$$\pi_1 = \rho\pi_0, n = 0 \qquad (4.6)$$

To solve (4.5) and (4.6), the method of generating function is used:

Let $$G(y) = \sum_{n=0}^{\infty} \pi_n y^n; / y / \leq 1 \qquad (4.7)$$

Multiplying (4.5) by y^n and summing over the range, we obtain

$$(1+\rho)\sum_{n=1}^{\infty} \pi_n y^n = \rho\sum_{n=1}^{\infty} \pi_{n-k} y^n + \sum_{n=1}^{\infty} \pi_{n+1} y^n \qquad (4.8)$$

An addition of $\rho\pi_0$ to LHS of (4.8) and $\pi_1 = \rho\pi_0$ to RHS of (4.8), yields:

$$(1+\rho)\sum_{n=1}^{\infty} \pi_n y^n + \rho\pi_0 = \pi_1 + \rho\sum_{n=1}^{\infty} \pi_{n-k} y^n + \sum_{n=1}^{\infty} \pi_{n+1} y^n \qquad (4.9)$$

$$\Rightarrow (1+\rho)\left[\pi_0 + \sum_{n=1}^{\infty} \pi_n y^n\right] - \pi_0 = \rho\sum_{n=1}^{\infty} \pi_{n-k} y^n + \left[\pi_1 + \sum_{n=1}^{\infty} \pi_{n+1} y^n\right]$$

$$\Rightarrow (1+\rho)\sum_{n=0}^{\infty} \pi_n y^n - \pi_0 = \rho\sum_{n=k}^{\infty} \pi_{n-k} y^n + \frac{1}{y}\sum_{n=0}^{\infty} \pi_{n+1} y^{n+1}$$

Since $\overline{\Lambda}_{n-k} \neq 0$ for $n - k = 0,$ we have

$$(1+\rho)\sum_{n=0}^{\infty} \pi_n y^n - \pi_0 = \rho\sum_{m=0}^{\infty} \pi_m y^{m+k} + \frac{1}{y}\sum_{i=1}^{\infty} \pi_i y^i, \;\; n-k=m; \; n+1=i$$

$$= \rho\, y^k \sum_{m=0}^{\infty} \pi_m y^m + \frac{1}{y}\left[\sum_{i=0}^{\infty} \pi_i y^i - \pi_0\right]$$

$$or \quad (1+\rho)\, G(y) - \pi_0 = \rho\, y^k G(y) + \frac{1}{y}[G(y) - \pi_0]$$

$$or \quad G(y) = \frac{\pi_0(1-y)}{(1-y) - \rho y(1-y^k)} = \pi_0\left[1 - \rho y\left\{\frac{1-y^k}{1-y}\right\}\right]^{-1}, / y / \leq 1$$

$$= \pi_0 \sum_{n=0}^{\infty} (yp)^n \left(\frac{1-y^k}{1-y}\right)^n \qquad (4.10)$$

$$Therefore,\ G(y) = \pi_0 \sum_{n=0}^{\infty} (y\rho)^n \left(1 + y + y^2 + \ldots + y^{k-1}\right)^n$$

$$= \pi_0 \sum_{n=0}^{\infty} \rho^n \left(y + y^2 + \ldots + y^k\right)^n \qquad (4.11)$$

Setting y=1 in (4.11), the terms in the bracket would be $\sum_{i=1}^{k} 1^k = k$, therefore

$$G(1) = \pi_0 \sum_{n=0}^{\infty} \rho^n k^n = \pi_0 \left(\frac{1}{1 - k\rho}\right) \qquad (4.12)$$

$$For\ \ y = 1, (4.7)\ gives\ G(1) = \sum_{n=0}^{\infty} \pi_n = 1 \qquad (4.13)$$

$$1 = \pi_0 \left(\frac{1}{1 - k\rho}\right) or\ \ \pi_0 = 1 - k\rho \qquad (4.14)$$

Putting the value of π_0 in (4.10), we obtain

$$G(y) = (1 - k\rho) \sum_{n=0}^{\infty} (y\rho)^n (1 - y^k)^n (1 - y)^{-n} \qquad (4.15)$$

$$= (1 - k\rho) \sum_{n=0}^{\infty} (y\rho)^n \left\{\left[\sum_{z=0}^{\infty} (-1)^z\ {}^nC_z y^{zk}\right]\left[\sum_{i=0}^{\infty}\ {}^{n+i-1}C_i y^i\right]\right\}; (-1)^{2i} = 1$$

$$\Rightarrow \sum_{n=0}^{\infty} \pi_n y^n = (1 - k\rho) \sum_{n=0}^{\infty} \rho^n \left[\sum_{z=0}^{\infty} \sum_{i=0}^{\infty} (-1)^z\ {}^nC_z\ {}^{n+i-1}C_i y^{i+zk+n}\right] \qquad (4.16)$$

Comparing the coefficient of y^n in both sides of (4.16), we obtain

$$\pi_n = (1 - k\rho) \sum_{n,z,i} \rho^n (-1)^z\ {}^nC_z\ {}^{n+i-1}C_i \qquad (4.17)$$

5.0 Results

To determine the expected number of phases in the system $L_s(k) = \sum_{n=0}^{\infty} n\pi_n$, we consider one of the difference equations as:

$$(1 + \rho)\pi_n = \rho\pi_{n-k} + \pi_{n+1};\ n \geq 1$$

Multiplying both sides by n^2 and summing gives:

$$(1 + \rho) \sum_{n=1}^{\infty} n^2 \pi_n = \rho \sum_{n=1}^{\infty} n^2 \pi_{n-k} + \sum_{n=1}^{\infty} n^2 \pi_{n+1}$$

$$(1 + \rho) \sum_{n=1}^{\infty} n^2 \pi_n = \rho \sum_{n=k}^{\infty} n^2 \pi_{n-k} + \sum_{n=1}^{\infty} n^2 \pi_{n+1}$$

If $n-k \equiv n \text{ and } n+1 \equiv n$, then RHS becomes;

$$(1+\rho)\sum_{n=1}^{\infty} n^2\pi_n = \rho\sum_{n=0}^{\infty}(n+k)^2\ \pi_n + \sum_{n=1}^{\infty}(n-1)^2\ \pi_n$$

$$= \rho\sum_{n=0}^{\infty}(n+k)^2\pi_n + \left\{\sum_{n=0}^{\infty}(n-1)^2\ \pi_n - \pi_0\right\}$$

$$= \sum_{n=0}^{\infty}\left[\rho(n+k)^2 + (n-1)^2\right]\pi_n - \pi_0$$

$$= \sum_{n=0}^{\infty}\left[\rho(n^2+k^2+2nk) + (n^2+1-2n)\right]\pi_n - \pi_0$$

$$(1+\rho)\sum_{n=1}^{\infty} n^2\pi_n = (1+\rho)\sum_{n=0}^{\infty} n^2\pi_n - 2(1-k\rho)\sum_{n=0}^{\infty} n\,\pi_n + 2(1+\rho k^2)\sum_{n=0}^{\infty}\pi_n - \pi_0$$

$$hence, \quad 2(1-\rho k)\sum_{n=0}^{\infty} n\,\pi_n = \rho k^2 + 1 - \pi_0$$

$$\mathrm{L_s}(k) = \sum_{n=0}^{\infty} n\,\pi_n = \sum_{n=0}^{\infty} n\,\pi_n = \frac{pk^2+1-\pi_0}{2(1-\rho k)} = \frac{\rho k^2+1-(1-\rho k)}{2(1-\rho k)}$$

$$\therefore \mathrm{L_s}(k) = \frac{\rho k^2+1-1+\rho k}{2(1-\rho k)}$$

$$= \frac{\rho k\,(k+1)}{2(1-\rho k)} = \frac{k(1+k)}{2}\left(\frac{\rho}{1-\rho k}\right)$$

$$= \frac{k(1+k)}{2}\left(\frac{\lambda}{ks\mu}\right)\left(\frac{s\mu}{s\mu-\lambda}\right)$$

$$= \frac{(1+k)}{2}\left\{\frac{\lambda}{s\mu-\lambda}\right\},\ \rho = \frac{\lambda}{ks\mu}, \lambda < s\mu \text{ and } \rho k < 1 \qquad (5.1)$$

The other operating characteristics of the model have been obtained as follows:

Expected Number of Phases in the Queue

$$\mathrm{L_q}(k) = \frac{\mathrm{L_s}(k)}{s\mu} = \left(\frac{1+k}{2}\right)\left[\frac{\lambda}{s\mu\,(s\mu-\lambda)}\right] \qquad (5.2)$$

Expected Number of customers in the Queue

$$\mathrm{L_q} = \frac{\mathrm{L_s}(k)\ - Average\ Number\ of\ Phases\ in\ Service}{k}$$

$$= \frac{1}{k}\left[\left(\frac{k+1}{2}\right)\left(\frac{\lambda}{s\mu-\lambda}\right) - \left(\frac{k+1}{2}\right)\frac{\lambda}{s\mu}\right]$$

$$= \left(\frac{k+1}{2k}\right)\left(\frac{\lambda}{s\mu\,(s\mu-\lambda)}\right) \qquad (5.3)$$

Expected Waiting Time in the Queue

$$W_q = \frac{L_q}{\lambda} = \left(\frac{k+1}{2k}\right)\left[\frac{\lambda}{s\mu\,(s\mu-\lambda)}\right] \tag{5.4}$$

Expected Waiting Time of a customer in the system

$$W_s = W_q + \frac{1}{s\mu} = \left(\frac{k+1}{2k}\right)\left(\frac{\lambda^2}{s\mu\,(s\mu-\lambda)}\right) + \frac{1}{s\mu} \tag{5.5}$$

Expected Number of customers in the system

$$L_s = L_q + \frac{\lambda}{s\mu} = \left(\frac{k+1}{2k}\right)\left(\frac{\lambda^2}{s\mu\,(s\mu-\lambda)}\right) + \frac{1}{s\mu}$$

$$or \quad L_s = \lambda W_s \tag{5.6}$$

6.0 Applications

Illustration 1

In a National Youth Service Camp in Nigeria, there are four counters through which all Corp members must pass before Registration and accreditation are completed. At the first counter, registration is done; in the second counter, there is a presentation of credentials for authentication. At the third counter, traveling allowances are paid and at the fourth counter, assignment of Corp members to various hostels is done. If arrival of Corp members follows a Poisson process with a mean arrival of 9/hr, the service times follow an Erlang distribution with a mean of 1.5 minutes per Corp member, and the Queue discipline is first come first served with infinite population; the performance measures can be obtained as follows:

CASE 1:

We note that this is a case of Multiple Phases Queue model with a single server in each phase, that is, $M/E_K/1$

λ = *9/hr,* μ = 1.5mins/Corp member, = 10 Corp members/hr
k = *4*

The following results are obtained:

$$L_q = \frac{k+1}{2k}\left[\frac{\lambda^2}{\mu(\mu-\lambda)}\right] = \frac{5}{8}\left[\frac{9^2}{10(10-9)}\right] = 5.0625 \approx 5\ \textit{Corp members},$$

$$W_q = \left[\frac{k+1}{2k}\left[\frac{\lambda}{\mu(\mu-\lambda)}\right]\right] = \frac{5}{8}\left[\frac{9}{10(10-9)}\right] = 0.5625 hrs = 33.75 \min s,$$

$$W_s = W_q + \frac{1}{\mu} = 0.5625 + \frac{1}{10} = 0.6625hrs = 39.75\min s,$$

$$L_s = \lambda W_s = 9 \times 0.6625 = 5.9625 \approx 6\ Corp\,members,$$

and $E(T) = \frac{1}{10} = 0.1hr\ or\ 6\min s$

Case 2:

Now, suppose each counter has 2 service points, we can obtain the performance measures as is the case in this work using *M/E$_K$/2*

From (5.3), $L_q = \frac{5}{8}\left[\frac{9^2}{20(20-9)}\right] = \frac{405}{17600} = 0.023$;

(5.4) gives $W_q = \frac{5\times 9}{17600} = 0.00256hrs = 0.153\min s = 9s$;

(5.5) is $W_s = W_q + \frac{1}{s\mu} = 0.00256 + \frac{1}{20} = 0.05256hr = 3.15\min s$;

Also, (5.6) gives $L_s = \lambda W_s = 9 \times 0.05256 = 0.473$

and $E(T) = \frac{1}{s\mu} = \frac{1}{20} = 0.05hr\ or\ 3\min s$

In contrast, the performance measures in case 2 give improved results in handling congestion through queuing model. For instance, in case 2, W_q = 9_S indicates that in adopting the two-server multi-phase queuing model, a Corp member can only wait in the queue for 9 seconds; the result L_q = 0.023 shows that by adopting this model there will be little or no congestion at all. Whereas in case 1, a Corp member can spend 33.75mins in a queue before he is being served.

Illustration 2

Suppose the Corp members arrive at the mean rate of 18 per hour while the mean service time is still 1.5mins per Corp member. We find out in this case, that the application of case 1 will be impracticable, because arrival out-matches the service rate, such that

$$L_q = \left(\frac{k+1}{2k}\right)\left[\frac{\lambda^2}{\mu(\mu-\lambda)}\right] < 0$$

This implies that the single server will not be able to contain or control the congestion.

On the other hand, case 2 which is Multiple-Server Multi-Phase Queuing System becomes more practicable, since it involves more service point at each phase. Thus:

(5.3) gives $L_q = \frac{k+1}{2k}\left[\frac{\lambda^2}{s\mu(s\mu-\lambda)}\right] = \frac{5}{8}\left[\frac{(18)^2}{20(20-18)}\right] = 5.0625 \approx 5\ Corp\ members$;

(5.4) gives $W_q = \frac{L_q}{\lambda} = \frac{5.0625}{18} = 0.281hr = 16.875\min s$;

(5.5) is $W_s = W_q + \frac{1}{s\mu} = 0.281 + \frac{1}{20} = 0.33125hr = 19.875\min s$;

and (5.6) gives $L_s = \lambda W_s = 18 \times 0.33125 = 5.9625 \cong 6\ Corp\ members.$

Also, $E(T) = \frac{1}{s\mu} = \frac{1}{20} = 0.05hr\ or\ 3\min s$

7.0 Conclusion

In this paper, we have extended the *M/E$_k$/1: (∞/FCFS)* model with single server to *M/E$_k$/s:(∞/FCFS)* model with multiple server and obtained its performance characteristics which is an improvement of the *M/E$_K$/1: (∞/FCFS)* model for handling congestion especially at peak period which in turn sustains customer's goodwill. Numerical illustrations have shown that the extended *M/E$_k$/s: (∞/FCFS)* model outperformed the single server *M/E$_K$/1: (∞/FCFS)* model when the arrival rate of customers is very high with a fixed mean service time. In the first illustrations, it is observed that with M/E$_k$/1: (∞/FCFS), an average of five corps members would wait in the queue for service and each corp member would spend 33.75 mins waiting for service. In contrast, the use of M*/E$_k$/s: (∞/FCFS)* model indicates that little or no queue would be obtained and a corp member can only wait for 9seconds only for service. These results show the efficiency and effectiveness of the latter model over the former. Further, in the second illustration, case 1 cannot be applied here because the traffic intensity here could not contain this model. Therefore the second case becomes suitable and practicable.

References

1. Dimitriou, I. and Langaris, C. (2010). A Repairable Queueing Model with Two-phase Service, Start-up Times and Retrial Custom, *Journal of Computers and Operations Research, 37(7), 1181-1190*.
2. Gupta, P. K. (2008). Operations Research. S. Chand & Co. Ltd, New Delhi.

3. Luh, H. (2004). A Queuing Model of General Servers in Tandem with Finite Buffer Capacities. *International Journal of Operations Research, 1(1), 71-76*.
4. Maurya, V. N. (2009). Analysis of a Multiple Service Channel Queuing Model with Priority Queue Discipline and Phase Service-Part I. *International J. of Maths. Sc. & Eng. Appls. (IJMSEA), 3 (IV), 357-364*.
5. Reid, R. D and Sanders, N. R. (2010): Operations Management. Wiley and Sons. 3rd Edition. New York.
6. Sharma, J. K. (2007). Operations Research: Theory and Applications. Macmillan India Ltd, New Delhi.