TITLE: Comparison of Efficiency Measures of Non-Priority and Non-Pre-Emptive Priority Queues

AUTHORS: Abdullah ÇELIK,Vedat SAGLAM

PAGES: 86-90

ORIGINAL PDF URL: https://dergipark.org.tr/tr/download/article-file/1599563

Comparison of Efficiency Measures of Non-Priority and Non-Pre-Emptive Priority Queues

Öncelikli Olmayan ve Sınırlı Olmayan Öncelikli Kuyrukların Performans Ölçülerinin Karşılaştırılması

Abstract

When there are different customer classes in a queue system, the j-th class customers have their services before the $j + 1, j = 1, 2, ...$ class customers. Such queues are named as queues with priority scheduling. In this study a Markov priority queue system with two customer classes is analyzed both under non-pre-emptive and nonpriority scheduling and the efficiency measures (the expected number of customer in system, the average waiting time in system) are obtained using Little's Law. The efficiency criteria are compared according to the priority situation. In addition, parameter estimates were compared with simulation results. The simulation was performed using the R program.

Öz

Bir kuyruk sisteminde farklı müşteri sınıfları olduğunda, j-inci sınıf müşteriler hizmetlerini $j + 1, j = 1, 2, ...$ sınıfı müşterilerden önce alırlar. Bu tür kuyruklar, öncelikli kuyruklar olarak adlandırılır. Bu çalışmada, iki müşteri sınıfına sahip bir Markov öncelikli kuyruk sistemi hem sınırlı olmayan hem de önceliksiz planlama kapsamında analiz edilmiş ve performans ölçümleri (sistemdeki beklenen müşteri sayısı, sistemdeki ortalama bekleme süresi) Little yasası kullanılarak elde edilmiştir. Performans kriterleri öncelik durumuna göre karşılaştırılmıştır. Ayrıca gerçek parametre tahminleri simülasyon sonuçlarıyla karşılaştırılmıştır. Simülasyon, R programı kullanılarak yapılmıştır.

Abdullah Çelik
Ondokuz Mayıs University,
abdullahcel@gmail.com,
Orcid: 0000-0003-4288-5567

Vedat Sağlam
Professor, Ondokuz Mayıs University,
vsaglam@omu.edu.tr,
Orcid: 0000-0002-8586-1373

Introduction

When analyzing some certain queues, customers are evaluated according to their customer class. In such queuing systems, customer priorities are determined by the importance of the customer. When there are different customer classes in the queue system, the *j-th* class customers have their services before the $j + 1$, $j = 1, 2, ...$ class customers. Customers in each class receive service according to FCFS discipline. There are different situations that apply when a customer comes to the system with a higher priority than the customer in the system. One of them is the non-pre-emptive priority queue system. In this case, when the customer comes to the system with a higher priority than the customer in the system, he expects the customer who sees the service in the system to complete the service. Barberis (1980) gave a method for calculating the average queue length in a priority queue system with two customer classes. Alfa (1997) presented a matrix-geometric solution of a discrete time $MAP/PH/1$ priority queueing system. Brodal et al. (1998) analyzed a fixed-time parallel queueing system with priority scheduling. Sanders (1998) developed an algorithm for faster operation of a random priority queuing system. Laevents and Bruneel (1998) examined the efficiency measure of a discrete time multiservice priority queuing system. Bitran and Caldentey (2002) investigated the efficiency measure of a two-customer priority queue system. Ali and Song (2004) performed a performance analysis of a tailored decimated timed priority queuing system. Pearl and Yechiali (2010) in their study, they examined the $M/M/c$ queuing system in a two-phase Markovian random enviroment, fast and slow, for non-patient customers. Jolai et al. (2016) studied a multi-purpose priority fuzzy queuing system. Kim et al. (2016) adapted a priority tandem queuing system to a customer service model. Nazarov and Paul (2016) analyzed a cyclic priority queue system.

Atencia (2017) studied the $GEO/G/1$ priority queuing system in which no waiting is allowed. Chaudhry et al. (2020) in their study starts with m customers $Geo/G/1$ queue and the distribution of the number of customers served in a busy period for exceptions to the early arrival and late arrival system with system latency access policies under the assumptions of the system are investigated.

In this study a Markov priority queue system with two customer classes is analyzed both under non-pre-emptive and non-priority scheduling and the efficiency measures (the expected number of customer in system, the average waiting time in system) are obtained using Little's Law. The obtained efficiency measures are then compared. In Section 2, it is investigated to obtain the efficiency measures of the queue systems using Little's law. In Section 3, the system with two customer class single channel priority queuing discipline is analyzed. In Section 4, it is compared the efficiencies of non-priority queue system and non-pre-emptive priority queue system. The results of numerical example is reported in Section 5. Finally, we arrive at a conclusion from these findings in the last section.

1. Obtaining Efficiency Measures of the System

The Little Law is the most commonly used formula in the Queuing Theory because it does not require a very broad acceptance of the ease of expression, the broad application possibilities and the features of the queuing system (Stewart, 2009).

This law equates the number of customers in the queue system to the product of the effective arrival rate and the system standby time. Let the random variables $a(t)$ and $d(t)$ be the number of arrivals and number of departures in the time interval $(0, t]$, respectively (Stewart, 2009). Thus, the number of customers in system at any t time is $a(t) - d(t)$. If $g(t)$ is the total time that all customers spend in system in interval $(0, t]$, then;

$$\lambda_t \equiv \frac{a(t)}{t} \tag{1}$$

is the average arrival rate in interval $(0, t]$. In the same interval the avarage waiting time per customer is,

$$R_t \equiv \frac{g(t)}{a(t)} \tag{2}$$

and finally,

$$L_t \equiv \frac{g(t)}{t} \tag{3}$$

is the average customer number in $(0, t]$ time interval. As a results it is seen that,

$$L_t = \frac{g(t)}{t} = \frac{g(t)}{a(t)} \cdot \frac{a(t)}{t} = R_t \lambda_t \tag{4}$$

is obtained.

When the following limits,

$$\lim_{t \to \infty} L_t = L = E(N) \tag{5}$$

$$\lim_{t \to \infty} \lambda_t = \lambda \tag{6}$$

$$\lim_{t \to \infty} R_t = E(R) \tag{7}$$

are assumed for $t \to \infty$, then

$$E(N) = \lambda E(R). \tag{8}$$

Little's Law (equation-8) equals the product of the average number of customers in the system with the average arrival time in the system and the average waiting time. This formula is mostly shown as:

$$L = \lambda W \tag{9}$$

If the average number of customers in the queue is L and the average waiting time in the system is W, Little's Law applies.

This means that, this law can be applied to different parts and service units of the queue system according to the requirement. If L_q is the average number of customer waiting in queue, L_s is the average number of customer receiving service, W_q is the average waiting time in queue and x is the average service time then using Little Formula,

$$L_q = \lambda W_q \quad \text{ve} \quad L_s = \lambda \bar{x} \tag{10}$$

finally we get,

$$L = L_q + L_s = \lambda W_q + \lambda \bar{x} = \lambda(W_q + \bar{x}) = \lambda W \tag{11}$$

as can be seen with this result, the Little Law can be applied in the more general cases, in the sub parts of the wider queuing systems, or in the complex queuing networks in which the different queuing systems come together (Stewart, 2009; Hayati, 2017; Burruni et al. 2019).

2. Single Channel Priority Queuing System with two Customer Class

In order to facilitate analysis of priority queuing systems, we begin by analyzing single channel queuing system in which there are only two customer classes and the scheduling policy is pre-emptive priority. In this system class-1 and class-2 customers arrive the queue according the Poisson distribution with means λ_1 and λ_2 respectively. Both class-1 and class-2 customers have exponentially distributed service times with rate μ, hence the traffic densities of class-1, class-2 and the system are as following respectively:

$$\rho_1 = \frac{\lambda_1}{\mu} \tag{12}$$

$$\rho_2 = \frac{\lambda_2}{\mu} \tag{13}$$

$$\rho = \rho_1 + \rho_2 \tag{14}$$

Since the service time μ exponentially distributed, because of the memorylessness property of the exponential distribution, the priority scheduling discipline will not affect the analysis whether the discipline is preempt-resume or preempt-restart (Stewart, 2009).

In fact, since the service times of all customers in the system have the same distribution with the same parameters, the total number of customers will be independent of the queue discipline. Therefore, the average number of customer in the system is

$$E(N) = \frac{\rho}{1-\rho}. \tag{15}$$

Let us now go on with the non-pre-emptive priority. In this case, when a class-1 customer arrives in the system and finds a class-2 customer receiving service, class-1 customer waits for the class-2 customer to complete service. When a class-1 customer arrives the system there are $E(N_1^*)$ class-1 customer in the system and each of these class-1 customers need a time of $1/\mu$ to complete their services. Also, this incoming class-1 customer needs a time of $1/\mu$. When a class-1 customer arrives in the system and finds a class-2 customer receiving service, an extra time of $1/\mu$ needs to be added to the total spent time of class-1 customer. Furthermore when a class-1 customer arrives the system, the probability that there is a class-2 customer receiving service is ρ_2. Sum of all service times class-1 gives a customers average waiting time

$$E(R_1^*) = E(N_1) \frac{1}{\mu} + \frac{1}{\mu} + \rho_2 \frac{1}{\mu}. \tag{16}$$

In equation (8), we obtain $E(N_1) = \lambda_1 E(R_1)$, and writing this equality in equation (16), then

$$E(R_1^*) = \lambda_1 E(R_1) \frac{1}{\mu} + \frac{1}{\mu} + \rho_2 \frac{1}{\mu}$$

$$E(R_1^*) = \frac{\frac{1+\rho}{\mu}}{\frac{\mu-\lambda_1}{\mu}} = \frac{1+\rho_2}{\mu-\lambda_1} = \frac{1+\rho_2}{\mu-\mu\rho_1} = \frac{1+\rho_2}{\mu(1-\rho_1)}. \tag{17}$$

The expected number of class-1 customers is calculated as

$$E(N_1^*) = \lambda_1 E(R_1)$$

$$E(N_1^*) = \lambda_1 \frac{1+\rho_2}{\mu(1-\rho_1)} \tag{18}$$

or

$$E(N_1^*) = \frac{\rho_1(1+\rho_2)}{(1-\rho_1)}. \tag{19}$$

With similar calculations we find

$$E(N_2^*) = E(N) - E(N_1^*)$$

$$E(N_2^*) = \frac{\rho_1+\rho_2}{1-\rho_1-\rho_2} - \frac{\rho_1(1-\rho_2)}{1-\rho_1}. \tag{20}$$

As a result, using Little's Law again, the average waiting time for a class-2 customer in the system is as follows

$$E(R_2^*) = \frac{E(N_2^*)}{\lambda_2} = \frac{\frac{[1-\rho_1(1-\rho_1-\rho_2)\}}{\mu}}{(1-\rho_1)(1-\rho_1-\rho_2)}. \tag{21}$$

3. Comparison of Efficiency Measures of Non-Priority and Non-Pre-Emptive Priority Queues

Waiting times will be equal because customers in a non-prioritized queue system are received equally. However, in the non-pre-emptive priority queue system, first-class customers will spend less time on the system than second-class customers, as the priority is to make the customer more efficient than the system. This will increase the waiting times of second class customers in the system. The customer class we are interested in is first class customer. Therefore, as seen from the comparisons, the waiting times of the first class customers are smaller than the waiting times in a non-prioritized queuing system. In the case of second-class customers, waiting times in the queue are slightly higher than in the non-prioritized system.

Theorem 1. $E(R_1^*) \le E(R)$

Proof 1. Since,

$$\frac{1+\rho_2}{\mu(1-\rho_1)} \le \frac{\rho}{\rho(1-\rho)} \tag{22}$$

under condition $0 < \rho_2 < 1$ it is easy to see $E(R) \ge E(R_1^*)$. In other words, the average waiting times in the system of first-class customers in a non-pre-emptive priority queue discipline are smaller than in a non-prioritized system.

Theorem 2. $E(R) \le E(R_2^*)$

Proof 2. When necessary simplificarions are made in following equations,

$$\frac{\rho}{\lambda(1-\rho)} \le \frac{\frac{[1-\rho_1(1-\rho_1-\rho_2)}{\mu}}{(1-\rho_1)(1-\rho_1-\rho_2)} \tag{23}$$

is obtained. Where $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$. That is, in the case of a non-pre-emptive priority queue system, a second-class customer has an average waiting time in the queue, which is slightly higher than a non-prioritized system (Çelik, 2007).

4. Numerical Results

Let's consider in this queuing system, $E(R_1^*)$, $E(R_2^*)$, $E(R)$ arrival parameters and average service times by changing the exact results and simulation results, as well as comparing the results of the non-priority queue system is given in the following Table 1. When calculating the average waiting time $E(R)$ of the non-priority queue system, $\lambda = \lambda_1 + \lambda_2$ is taken. Since $X_1 \sim Poiss(\lambda_1)$ and $X_2 \sim Poiss(\lambda_2)$ are present, $X = X_1 + X_2 \sim Poiss(\lambda)$. The reason for this is that the Poisson distribution is closed to collect. Note that customer numbers taken as N=50,100,500,1000.

Tablo 1. Comparison of Exact and Simulation Values of Non-Priority and Non-Pre-Emptive Priority Queue System

| λ_1 | λ_2 | μ | $E(R_1)$ | | | | | $E(R_2)$ | | | | | Non-priority exact |
| | | | Non-pre-emptive exact | Non-pre-emptive simulation expected waiting times | | | | Non-pre-emptive exact | Non-pre-emptive simulation expected waiting times | | | | |
				N=50	N=100	N=500	N=1000		N=50	N=100	N=500	N=1000	
0.3	0.5	1	2.142	2.012	2.079	2.178	2.143	6.714	4.469	5.312	6.449	6.580	5,000
0.3	0.5	0.9	2.600	2.385	2.473	2.572	2.580	14.520	6.339	8.290	12.261	13.410	10.100
0.1	0.2	1	1.330	1.315	1.324	1.333	1.336	1.476	1.442	1.478	1.479	1.477	1.428
0.1	0.2	0.9	1.523	1.521	1.523	1.528	1.523	1.726	1.703	1.739	1.736	1.726	1.658
0.1	0.2	0.8	1.786	1.763	1.764	1.784	1.789	2.107	2.033	2.056	2.095	2.102	2,000
0.1	0.2	0.7	2.143	2.089	2.141	2.137	2.142	2.678	2.537	2.637	2.670	2.685	2.499
0.1	0.2	0.6	2.671	2.584	2.651	2.662	2.660	3.674	3.459	3.630	3.668	3.644	3.330
0.1	0.2	0.5	3.500	3.433	3.497	3.464	3.491	5.750	5.307	5.494	5.641	5.735	5,000
0.1	0.2	0.4	5,000	4.818	4.928	4.951	4.980	12.500	9.398	11.038	12.173	12.324	10.000
0.5	0.3	1	2.600	2.369	2.459	2.586	2.582	9,000	5.314	6.543	8.455	8.620	5,000
0.2	0.1	0.4	6.250	5.784	6.089	6.235	6.197	17.500	12.054	14.366	17.078	16.968	10,000
0.25	0.25	1	1.670	1.630	1.647	1.660	1.661	2.330	2.211	2.269	2.319	2.312	2,000
0.4	0.4	1	2.330	2.221	2.252	2.318	2.327	7.670	4.962	6.034	7.392	7.418	5,000

Conclusion

In this study, an $M/M/1$ priority queue with two customer classes is analyzed both under non-priority and non-pre-emptive priority scheduling and the efficiency measures (the average waiting

time in system) are obtained using Little's Law. Obtained efficiency measures are then compared. Comparison results for both non-priority and non-pre-emptive priority scheduling are given as two theorems. The exact results of the average waiting times of the non-priority and non-pre-empty priority systems are compared with the simulation results. From Table 1, it is observed that the simulation results approached the final results when N increased. In addition, by changing the value of the arrival parameters (λ_i, $i = 1, 2$) and the service parameter (μ), it is seen that the first-class customers have less waiting times in the average queue. In this study, the average waiting times of the non-priority class of the non-pre-emptive priority system decreased significantly compared to the non-priority system. For further studies, efficiency measures of queues with non-priority and non-pre-emptive policies can be analyzed in sophisticated queuing systems.

References

Alfa, A. S. 1998. Matrix-Geometric solution of discrete time MAP/PH/1 priority queue, *John Wiley & Sons, Inc. Naval Research Logistics,* 45, pp. 23-50.

Atencia, I. 2017. A Geo/G/1 retrial queuing system with priority services, *European Journal of Operational Research* 256, pp. 178-186.

Barberis, G. 1980. A Useful Tool in the Theory of Priority Queueing, *IEEE Transactions on Communications*, 28, pp. 1757-1762.

Bitran, G. and Caldentey, R. 2002. Two-Class Priority Queueing System with State-Dependent Arrivals, *Queueing Systems,* 40, pp. 355-382.

Brodal, G. S., Träff, J. L. and Zaroliagis, C. D. (1998). A parallel priority queue with constant time operations. *Journal of Parallel and Distributed Computing*, 49 (1), pp. 4-21.

Burruni, R., Cuany, B., Valerio, M., Jichlinski, P. and Kulik, G. 2019. Reduction and follow-up of hospital discharge letter delay using Little's law. *International Journal for Quality in Health Care*, 31 (10), pp. 787-792.

Chaudhry, M. L., Goswami, V. and Mansur, A. Analytically closed-form solutions for the distribution of a number of customers served during a busy period for special cases of the GEO/G/1 queue. *Probability in the Engineering and Informational Sciences*, pp. 1-30.

Çelik, A. 2017. Priority queuing systems with two customer classes and a numerical example, Master thesis, Ondokuz Mayıs University, Samsun, Turkey.

Hayati, F. 2017. Some perspectives on the application of Little's law: L= λW. *International Journal of Modelling in Operations Management*, 6 (3), pp. 141-152.

Jolai, F., Asadzadeh, S. M., Ghodsi, R. and Bagheri-Marani, S. 2016. A multi-objective fuzzy queuing priority assignment model. *Applied Mathematical Modelling*, 40 (21-22), pp. 9500-9513.

Kim, C., Klimenok, V. I. and Dudin, A. N. 2016. Priority tandem queueing system with retials and reservation of channels as a model of call center, *Computers & Industrial Engineering,* 96, pp. 61-71.

Leavens, K. and Bruneel, H. 1998. Discrete-time multiserver queues with priorities, *Performance Evaluation,* 33, pp. 249-275.

Mehmet Ali, M. and Song, X. 2004. A performance analysis of a discrete-time priority queuing system with correlated arrivals". *Performance Evaluation,* 57, pp. 307-339.

Nazarov, A. and Paul, S. A. 2016. Cyclic Queuing System with Priority Customers and TStrategy of Service, *Distributed Computer and Communication Networks,* pp. 182-193.

Perel, N. and Yechiali, U. (2010). Queues with slow servers and impatient customers. *European Journal of Operational Research*, *201*(1), 247-258.

Sanders, P. 1998. Randomized Priority Queues for Fast Parallel Access, *Journal of Parallel and Distributed Computing* 49, pp. 86-97.

Stewart, W. J. 2009. *Probability, Markov Chains, Queues, and Simulation,* Princeton University Press, United Kingdom.