

## A novel approach for optimization in a fuzzy finite capacity queuing model with system cost and expected degree of customer satisfaction

Mohammad Bagher Shahin<sup>a</sup>, Ali Doniavi<sup>a</sup>, Maghsoud Solimanpur<sup>a</sup> and Mahdi Shahin<sup>b\*</sup>

<sup>a</sup>Department of Industrial Engineering, urmia University, Iran

<sup>b</sup>Department of Industrial Engineering, Science and Technology University, Tehran, Iran

### CHRONICLE

*Article history:*  
Received February 9, 2015  
Received in revised format:  
May 12, 2015  
Accepted June 2, 2015  
Available online  
June 2 2015

*Keywords:*  
*Finite-capacity queuing model*  
*Expected degree of customer satisfaction*  
*System cost*  
*Productivity rate index*  
*Optimal number of servers*

### ABSTRACT

From a wide variety of queuing models, the finite-capacity queuing models are the most commonly used, where arrival and service rates follow an exponential distribution. Based on two criteria of system cost and expected degree of customer satisfaction, the present study defines a new productivity rate index and evaluates the optimization of a queuing model with finite capacity. In queuing models, obviously, as the number of servers increases, the length of waiting lines decreases, the expected degree of customer satisfaction enhances, and obviously the system cost increases. This study deals with the mathematical relationships involved in the computations of these two criteria, and proposes a novel approach to determine an optimal number of servers by considering a decision-maker's priority and establishing a trade-off between criteria.

© 2015 Growing Science Ltd. All rights reserved.

## 1. Introduction

In basic queuing models, unit arrivals are assumed to follow a Poisson distribution and service times are exponential. Some models have infinite capacity, while others are with finite capacity queues. Queuing models with finite capacities have been widely used for the real setting and, hence, they are of great importance. Such models deal with systems of finite capacity; meaning that the units involved in any system could not be exceeded by a given number, and if the system is at full capacity, the arrival of a new unit would not be authorized and it will be lost. Finite-capacity queuing models play an important role with many practical applications. To name a few studies, Balsamo et al. (2003) studied queuing network models to estimate the performance of software architecture. Takagi and Yoshino (2005) investigated the system capacity calculation for wireless systems. Moreover, Heragu and Srinivasan (2011), Smith (2003), Bocharov and Viskova (2005), Jain (2005), Gupta and Sikdar (2006), and Thomas (2006) provided practical researches in queuing models with finite-capacity.

\* Corresponding author.  
E-mail address: [mahdi.Shahin@ind.iust.ac.ir](mailto:mahdi.Shahin@ind.iust.ac.ir) (M. Shahin)

Many studies have been conducted on the optimization of finite-capacity queuing models. Through the introduction of a multi-agent based bank queuing model, Zhang (2012) sought on optimization problem. Lariviere and Porteus (2001) presented an analysis of price-only contracts for a special case of the newsvendor problem. Zhao et al. (2014) proposed a group of optimization models that aim at determining the number of servers, which is needed for net jobs with random process times and then they have achieved the optimal total expected costs for each model. Ziya et al. (2006) developed a lower bound to determine the optimal price for each service in blocking systems, and proved how optimal prices alter with changes in the size of waiting room or the system capacity. In this field and for optimal pricing policies, worthy efforts have also been made by Paschalidis and Liu (2002), Paschalidis and Tsitsiklis (2000), and Sumita et al. (2001).

Additionally, for the optimization of queuing models with finite-capacity, Ke and Wang (2002), and Wang and Ke (2000) applied a recursive method for, respectively, the study of N policy and the optimal control of an M/G/1 queuing system with finite capacity. Here, the N-policy means that when one customer is found in a system, then the server will stop its operations until N customer ( $N \geq 1$ ) will enter the system. Chen (2007) proposed a mixed-integer nonlinear programming method for queuing problems and, then, applied it to calculate a proper service rate. Wesołek et al. (2006) optimized queuing models by using a multi-objective approach, in which the objective was to increase the level of customer satisfaction and minimize the number of servers. Sindal and Tokekar (2012) interested in analyzing of two queuing based call admission control schemes with a fuzzy based call admission controller model. The models developed were based on channel reservation scheme for handoff calls. Finite queuing scheme for handoff call was implemented in the first model whereas finite queuing for both handoff as well as new call was taken into account for the second model. The fuzzy call admission controller was used based on Mamdani inference scheme taking relative mobility of user and number of reserved channels for handoff calls as input parameters.

Prado and La Fuente (2007, 2008, and 2009) dealt with the optimization of queuing models, by considering priorities, selection of service rates, and profit. Also, for the optimized selection of the number of servers in finite-capacity queuing models, Prado and La Fuente (2008) developed a novel technique based on the expected degree of customer satisfaction. Note that the method proposed here has great limitations, since it relies only on the expected degree of customer satisfaction, thereby restricting its applications. Through the study of the optimization of finite capacity queuing systems, the current research provides an alternative approach to determine the optimal number of servers by considering two criteria, including the level of customer satisfaction and the total cost in a queuing system. It should be noted that this is more simple, comprehensive, and applicable method, from a practical viewpoint. The paper is presented as following:

Section 2 deals with a finite-capacity queuing model, describing the calculations for the probabilities of steady states. Section 3 explains how to determine the optimal number of servers. Next, Section 4 provides a numerical example, and ultimately, Section 5 draws the conclusion.

## 2. Finite-Capacity Queuing Model (M/M/m/K)

A finite-capacity queuing model ( $\tilde{M}/\tilde{M}/m/K$ ) is composed of  $m$  servers, where all servers have the same service rate, equal to  $\mu$ , and independent from the system status – i.e. the number of units involved in the system. In fact, intervals between the arrival and service times follow an exponential distribution with parameters  $\lambda$  and  $\mu$ , respectively. An additional assumption is that when the number of units present in a queuing system equals the capacity of the system (denoted by  $K$ ), then the arrival of a new unit will be impossible since the system is at full capacity, and it will be considered as a missing unit. The exit rate for this model is different from the arrival pattern; to put it differently, if the number of units is less than  $m$ , the exit rate is equal to  $n\mu$ . However, when the number of units in the system is more than  $m$ , the exit rate equals  $m\mu$ . Hence, we have:

$$\mu_n = \begin{cases} n\mu & n=1, 2, \dots, m-1 \\ m\mu & n=m, m+1, \dots, K \\ 0 & n > K \end{cases}, \quad \lambda_n = \begin{cases} \lambda & n=0, 1, \dots, K-1 \\ 0 & n \geq K \end{cases} \quad (1)$$

The relations given in Eq. (2) were introduced by Taha (2003) to compute the probabilities of steady states and performance indicators in a queuing model:

$$\pi_0 = \left[ \sum_{n=0}^m \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{n!} + \left(\frac{\lambda}{\mu}\right)^m \cdot \frac{1}{m!} \sum_{n=m+1}^K \left(\frac{\lambda}{m \cdot \mu}\right)^{n-m} \right]^{-1}$$

$$\pi_n = \begin{cases} \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{n!} \cdot \pi_0 & 1 \leq n \leq m \\ \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{m!} \cdot \frac{1}{m^{n-m}} \cdot \pi_0 & m \leq n \leq k \end{cases} \quad (2)$$

where  $(\pi_0)$  represents the probability of a null system, while  $(\pi_n)$  represents the probability of involving  $n$  units in the system, shown as below:

The average queue length ( $Lq$ ) and the average waiting time in the queue ( $Wq$ ) are computed as follows,

$$Lq = \frac{\pi_0}{m!} \cdot \left(\frac{\lambda}{\mu}\right)^m \sum_{n=m+1}^K (n-m) \cdot \left(\frac{\lambda}{m \cdot \mu}\right)^{n-m}, \quad (3)$$

$$Wq = \frac{Lq}{\lambda(1 - \pi_n)}.$$

The average number of units ( $L$ ) and the average waiting time in the system ( $W$ ) are as follows,

$$L = \sum_{n=0}^{m-1} n \cdot \pi_n + Lq + m \left(1 - \sum_{n=0}^{m-1} \pi_n\right), \quad (4)$$

$$W = \frac{L}{\lambda(1 - \pi_n)}.$$

### 3. Determination of Optimal Number of Servers

The present study aims to determine the optimal number of servers with regards to two criteria; namely, the expected degree of customer satisfaction and the total cost of utilization of the server. Hence, we will first discuss the methodology for calculating these two criteria. Next, two criteria will be used to develop a method in order to determine the optimal number of servers.

#### 3.1 Expected Degree of Customer Satisfaction

The underlying idea for the degree of customer satisfaction is that when the number of units waiting in a queue or system is low at the arrival time, the degree of customer satisfaction is high, and vice versa. Prado and La Fuente (2008) proposed a novel method to obtain the degree of customer satisfaction, where it is determined based on the queue length seen by an arriving customer. In this case, the arriving customers show higher levels of satisfaction when they see short queues, a normal satisfaction for moderate length of the queue, and lower levels of satisfaction for long queues. These terms “short”,

“moderate” and “long” used to describe the queue length have comparative and theological concepts, so the authors defined the queue by using the fuzzy set theory as the following fuzzy states:

$\tilde{A}$  = Short Queue,  $\tilde{B}$  = Moderate Queue,  $\tilde{C}$  = Long Queue.

$\tilde{A}$ ,  $\tilde{B}$ , and  $\tilde{C}$  represent the fuzzy states for the fuzzy partition  $X = (X_0, X_1, \dots)$ , where  $X_i$  shows an state of  $i$  customers involved in the system. Based on the principles of Zadeh Extension (Zadeh, 1965, 1971, 1973, 1974, 1975, 1983, 1997), the fuzzy states of  $\tilde{A}$ ,  $\tilde{B}$ , and  $\tilde{C}$  are defined as below:

$$\tilde{A} = \{(0, \mu_{\tilde{A}}(0)), (1, \mu_{\tilde{A}}(1)) \dots (K, \mu_{\tilde{A}}(K))\}$$

$$\tilde{B} = \{(0, \mu_{\tilde{B}}(0)), (1, \mu_{\tilde{B}}(1)) \dots (K, \mu_{\tilde{B}}(K))\}$$

$$\tilde{C} = \{(0, \mu_{\tilde{C}}(0)), (1, \mu_{\tilde{C}}(1)) \dots (K, \mu_{\tilde{C}}(K))\}$$

where,  $\mu_{\tilde{A}}$ ,  $\mu_{\tilde{B}}$ , and  $\mu_{\tilde{C}}$  represent the membership functions for the fuzzy states  $\tilde{A}$ ,  $\tilde{B}$ , and  $\tilde{C}$ , respectively (Zimmermann, 1978, 1984, 1985, 2001). Hence,  $\mu_{\tilde{A}}(i)$  shows the degree of possibility for  $\tilde{A}$  (the short queue) when there are  $i$  customers in the system. According to the rules of fuzzy partitions, the following relation is to be met for definitions of short, moderate, and long queues:

$$\mu_{\tilde{A}}(i) + \mu_{\tilde{B}}(i) + \mu_{\tilde{C}}(i) = 1, i = 1, 2 \dots K \quad (5)$$

where  $\pi_{\tilde{A}}$ ,  $\pi_{\tilde{B}}$ , and  $\pi_{\tilde{C}}$  denote the chance of arriving in a short, moderate and long queue for a new customer. Prado and La Fuente (2007) presented the following formula to calculate the probabilities of fuzzy states:

$$\pi_{\tilde{A}} = \sum_{n=0}^K \pi_n \cdot \mu_{\tilde{A}}(n)$$

$$\pi_{\tilde{B}} = \sum_{n=0}^K \pi_n \cdot \mu_{\tilde{B}}(n) \quad (6)$$

$$\pi_{\tilde{C}} = \sum_{n=0}^K \pi_n \cdot \mu_{\tilde{C}}(n)$$

Finally, they defined the utility values of  $u_1$ ,  $u_2$ , and  $u_3$  as to represent the degrees of customer satisfaction with – respectively - short, moderate, and long queues. Also, the relation below was provided to calculate the expected degree of customer satisfaction:

$$EDCS = (u_1 \cdot \pi_{\tilde{A}}) + (u_2 \cdot \pi_{\tilde{B}}) + (u_3 \cdot \pi_{\tilde{C}}) \quad (7)$$

### 3.2 Total Cost of Server Utilization

Obviously, increasing the number of servers leads to reducing the number of waiting units in the queuing system, and thereby increasing system costs. Selecting an optimal number of servers without taking the cost into consideration will not necessarily result in accurate and applicable outcomes, as this criterion is among the most important parameters which affect decisions. This paper proposes the following relation to compute the cost of server utilization:

$$\text{Total cost} = \text{Fixed cost} + \text{Variable cost} + \text{Opportunity cost} \quad (8)$$

The “fixed cost” reflects the cost which is not directly dependent to the number of servers; it is defined per all servers or some by the system manager. However, the variable cost is the cost with exactly a linear relationship to the number of servers. As stated before, when the system is at full capacity, no new arrival is allowed and this charges the system with a lost opportunity cost. Here, the following relation is suggested to compute the variable and lost opportunity costs:

$$\text{Variable Cost} = m \times Y_1, \text{ Opportunity Cost} = \lambda \times \pi_k \times Y_2 \quad (9)$$

where,  $Y_1$  is the cost of server utilization per hour, and  $Y_2$  is the average opportunity cost per customer.

In the real world, note that it is not quite simple to estimate total costs, and sometimes impossible by using this relation. Through describing the criterion of total costs, the purpose is not to develop Eq. (8) to estimate the cost of server utilization; rather, this study seeks to introduce a methodology to obtain an optimal number of servers by estimating the expected degrees of customer satisfaction and total costs per different numbers.

### 3.3 Optimal Number of Servers

The queuing model presented here has a finite capacity; i.e. the optimal number of servers can take the values 1, 2, 3, up to  $K$  (the size of system capacity) and no constraint is allowed. The values of expected degree of customer satisfaction are estimated per different numbers of servers in order to obtain an optimal number of servers. Next, the products are normalized. To do this, the values are divided by the largest number calculated. Hence, the number of servers with greatest degree of customer satisfaction is found the highest value; it is 1.

Then, the system costs are estimated in terms of various numbers of servers (from 1 to  $K$ ) and the results are normalized. In this case, the smallest cost value obtained in the calculations per different number of servers is divided by each value of the system costs. So, the number with lowest system costs is given the highest value; it is 1. These two criteria are, indeed, normalized so that the normalized criteria would be given the highest value when the system cost is contributed to the lowest rate and the customer satisfaction is at the greatest level.

In order to determine the optimal number of servers, the current study applies two criteria of normalized system cost and normalized degree of customer satisfaction to present the productivity rate (PR) as a decision-making index as below:

$$\text{Productivity Rate} = (W_1 \times \text{Normalized EDCS}) + (W_2 \times \text{Normalized SC}) \quad (10)$$

where,  $W_1$  and  $W_2$  denote the weights of the expected degree of satisfaction and the system cost.

The PR index shows a direct relationship with both normalized criteria. It means that higher degrees of customer satisfaction and lower system costs will result in a greater value for this decision-making index. Therefore, the optimal number of servers is a definite number and it is equal to the number per which the ER index receives the largest value. Moreover, according to Eq. (10), it is clear that we can assume different weights for customer satisfaction and system costs criteria as to obtain an optimal number of servers.

## 4. Numerical Example with Sensitivity Analysis

This section presents a numerical example to further explain how the optimal number of servers is determined, discussing the exact steps. Then, the sensitivity analysis is proposed to determine the parameters affecting greatly the model. Consider the M/M/m/10 finite-capacity queuing model, where the arrival and service rates are 40 and 10 units per hour, respectively. The system has a limited capacity by 10 units. The cost of server utilization is equal to 2000 monetary units per hour, the lost opportunity cost is equal to 15 monetary units per hour, and the fixed cost of server utilization is defined as below:

$$\text{Fixed cost} = \begin{cases} 5000 & 1 \leq n < 5 \\ 5500 & 5 \leq n < 9 \\ 7000 & 9 \leq n \leq 10 \end{cases}$$

Also, suppose the utility values of  $u_1 = 0.95$ ,  $u_2 = 0.5$ , and  $u_3 = 0.05$ . The fuzzy states of  $\tilde{A}$ ,  $\tilde{B}$ , and  $\tilde{C}$  are defined as the following fuzzy partitions:

$$\tilde{A} = \{(0,1), (1,1), (2, 1), (3,0.7), (4, 0.3)\}, \tilde{B} = \{(3,0.3), (4,0.7), (5,1), (6,1), (7,0.6), (8,0.2)\}$$

$$\tilde{C} = \{(7, 0.4), (8, 0.8), (9, 1), (10, 1)\}$$

Moreover, the decision maker intends to allocate the weighting values of 0.4 for the expected degree of customer satisfaction, and of 0.6 for the system cost. The data are summarized as below:  
 $\lambda = 40$ ,  $\mu = 10$ ,  $K = 10$ ,  $Y_1 = 200$ ,  $Y_2 = 15$ ,  $u_1 = 0.95$ ,  $u_2 = 0.5$ ,  $u_3 = 0.05$ ,  $W_1 = 0.4$  and  $W_2 = 0.6$ .

To obtain the optimal number of servers, it is necessary to estimate the probabilities of steady states ( $\pi_0, \pi_1 \dots \pi_{10}$ ) per all serves by Eq. (2). Table 1 shows the computational results.

**Table 1**  
Limiting Probabilities of System per Multiple Servers

	m = 1	m = 2	m = 3	m = 4	m = 5	m = 6	m = 7	m = 8	m = 9	m = 10
$\pi_0$	0.0000	0.0002	0.0033	0.0102	0.0152	0.0173	0.0181	0.0183	0.0184	0.0184
$\pi_1$	0.0000	0.0010	0.0133	0.0407	0.0608	0.0693	0.0723	0.0732	0.0734	0.0735
$\pi_2$	0.0000	0.0020	0.0266	0.0814	0.1216	0.1387	0.1446	0.1464	0.1469	0.1469
$\pi_3$	0.0000	0.0039	0.0355	0.1085	0.1621	0.1849	0.1928	0.1952	0.1958	0.1959
$\pi_4$	0.0002	0.0078	0.0473	0.1085	0.1621	0.1849	0.1928	0.1952	0.1958	0.1959
$\pi_5$	0.0007	0.0156	0.0631	0.1085	0.1297	0.1479	0.1542	0.1561	0.1566	0.1567
$\pi_6$	0.0029	0.0313	0.0841	0.1085	0.1037	0.0986	0.1028	0.1041	0.1044	0.1045
$\pi_7$	0.0117	0.0625	0.1121	0.1085	0.0830	0.0657	0.0588	0.0595	0.0597	0.0597
$\pi_8$	0.0469	0.1251	0.1495	0.1085	0.0664	0.0438	0.0336	0.0297	0.0298	0.0299
$\pi_9$	0.1875	0.2502	0.1993	0.1085	0.0531	0.0292	0.0192	0.0149	0.0133	0.0133
$\pi_{10}$	0.7500	0.5004	0.2658	0.1085	0.0425	0.0195	0.0110	0.0074	0.0059	0.0053

Then,  $\pi_{\tilde{A}}$ ,  $\pi_{\tilde{B}}$ , and  $\pi_{\tilde{C}}$  are calculated by using Eq. (6), and Eq. (8) is used to determine and normalize the expected degrees of customer satisfaction in terms of  $m = 1, \dots, m$  servers. Table 2 shows the summarized data of the estimations.

**Table 2**  
Values of  $\pi_{\tilde{A}}$ ,  $\pi_{\tilde{B}}$ , and  $\pi_{\tilde{C}}$ , and Normalized Degrees of Customer Satisfaction

	m = 1	m = 2	m = 3	m = 4	m = 5	m = 6	m = 7	m = 8	m = 9	m = 10
$\pi_{\tilde{A}}$	0.0001	0.0083	0.0823	0.2407	0.3596	0.4103	0.4277	0.4331	0.4344	0.4347
$\pi_{\tilde{B}}$	0.0202	0.1161	0.2881	0.4122	0.4585	0.4797	0.4918	0.4971	0.4987	0.4990
$\pi_{\tilde{C}}$	0.9797	0.8756	0.6296	0.3471	0.1819	0.1101	0.0805	0.0699	0.0669	0.0663
EDCS	0.0592	0.1097	0.2537	0.4521	0.5800	0.6351	0.6562	0.6634	0.6654	0.6658
N EDCS	0.0889	0.1647	0.3811	0.6791	0.8711	0.9539	0.9857	0.9965	0.9995	1.0000

According to Table 2, N EDCS represents the normalized values for the expected degrees of customer satisfaction. Now, the system costs are to be determined per all servers by using Eq. (15) and Eq. (16). Table 3 provides the results.

**Table 3**  
Normalized Values of System Cost

	m = 1	m = 2	m = 3	m = 4	m = 5	m = 6	m = 7	m = 8	m = 9	m = 10
<i>Fixed cost</i>	5000	5000	5000	5000	5500	5500	5500	5500	7000	7000
<i>Variable cost</i>	200	400	600	800	1000	1200	1400	1600	1800	2000
<i>Opportunity</i>	450.00	300.22	159.48	65.08	25.49	11.69	6.58	4.46	3.54	3.18
<i>Total cost</i>	5650.0	5700.2	5759.4	5865.0	6525.4	6711.6	6906.5	7104.4	8803.5	9003.1
N total cost	1.0000	0.9912	0.9810	0.9633	0.8658	0.8418	0.8181	0.7953	0.6418	0.6276

**Table 4**  
PR Index for All Numbers of Servers with Equal Weights

	m = 1	m = 2	m = 3	m = 4	m = 5	m = 6	m = 7	m = 8	m = 9	m = 10
PR Index	0.6356	0.6606	0.7410	0.8496	0.8680	0.8867	0.8851	0.8758	0.7849	0.7765

Finally, the productivity rate is estimated for all numbers of servers. As seen from Table 4, the optimal number of servers is equal to 6, because this value results in the highest PR index (0.8867). Also, according to Table 1, the rate of missing customers per six servers ( $\pi_{10}$ ) is 1.95 percent, which is an acceptable rate. Meanwhile, the performance indicators for the optimal model M/M/6/10 include;

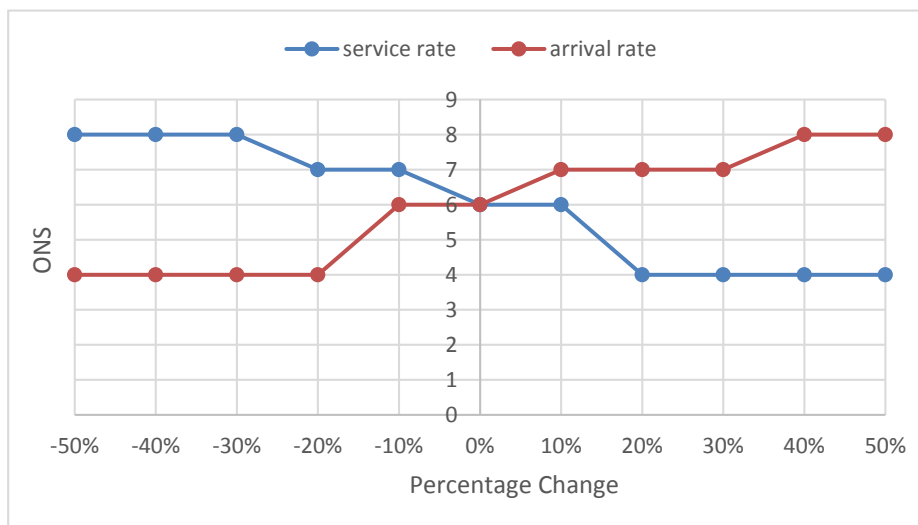
$Lq = 0.3190, L = 4.2411, Wq = 0.081$  and  $W = 0.2163$ .

Now, in order to carry out the sensitivity analysis and to determine sensitive parameters, we change the parameters of the arrival rate, service rate, and variable cost separately in the range of 10 to 50 percent, and estimate again the optimal number of servers. As seen from Table 5, the more sensitive parameters include the service and arrival rates; with 50% change of two criteria, the number of servers changes by 2 units.

**Table 5**  
Sensitivity Analysis for Arrival Rate, Service Rate, and Variable Cost

Percentage change	$\mu$	ONS	$\lambda$	ONS	$Y_1$	ONS	$Y_2$	ONS
-%50	5	8	20	4	100	7	7.5	6
-%40	6	8	24	4	120	7	9	6
-%30	7	8	28	4	140	7	10.5	6
-%20	8	7	32	4	160	7	12	6
-%10	9	7	36	6	180	6	13.5	6
%0	10	6	40	6	200	6	15	6
%10	11	6	44	7	220	6	16.5	6
%20	12	4	48	7	240	6	18	6
%30	13	4	52	7	260	6	19.5	6
%40	14	4	56	8	280	6	21	6
%50	15	4	60	8	300	6	22.5	6

Fig. 1 shows the optimal number of servers per 0 to 50 % changes in the arrival and service rates.



**Fig. 1.** Sensitivity Analysis of Arrival and Service Rates

It should be noted that the model demonstrated no sensitivity to changes in the variable and lost opportunity costs per customer, since the number of servers is just increased by 1 unit for almost 50% changes in the variable cost. Moreover, 50 % changes in the lost opportunity cost requires no changes in the number of servers.

## 5. Conclusion

The present study has aimed to determine an optimal number of servers with simultaneous regarding the system cost and degree of customer satisfaction. To this end, the productivity rate index was introduced which demonstrated a direct relationship with the normalized values of these two criteria. In order to normalize the values, these criteria are assigned greater values (max. 1) with decreased system costs and increased customer satisfaction. Further, in the relation proposed to estimate the PR index, unequal values can be allocated to the criteria, and this is considered as an advantage for the current methodology. In fact, to improve the flexibility and adaptability of the proposed method with thoughts of a decision maker, the decision maker is responsible to select weighting values (relative significance) for the criteria. Finally, the study provided the sensitivity analysis for the queuing system in order to determine which parameters were sensitive. The results indicated that the arrival and service rates were the most sensitive parameters, while the rest showed almost equal, albeit not effective, impact on the queuing system.

As a final note, this study provided several advantages, including: investigating two criteria of system cost and expected degree of customer satisfaction at the same time, allocation of unequal weights for decision criteria, and introducing the PR index.

## Acknowledgement

The authors would like to thank the anonymous referees for constructive comments on earlier version of this paper.

## References

- Balsamo, S., Personè, V. D. N., & Inverardi, P. (2003). A review on queueing network models with finite capacity queues for software architectures performance prediction. *Performance Evaluation*, 51(2), 269-288.
- Bocharov, P. P., & Viskova, E. V. (2005). A single-server finite-capacity queueing system with Markov flow and discrete-time service. *Automation and Remote Control*, 66(2), 233-248.
- Ch Paschalidis, I., & Liu, Y. (2002). Pricing in multiservice loss networks: static pricing, asymptotic optimality, and demand substitution effects. *IEEE/ACM Transactions on Networking (TON)*, 10(3), 425-438.
- Chen, S. P. (2007). Solving fuzzy queueing decision problems via a parametric mixed integer nonlinear programming method. *European Journal of Operational Research*, 177(1), 445-457.
- Gupta, U. C., & Sikdar, K. (2006). Computing queue length distributions in MAP/G/1/N queue under single and multiple vacation. *Applied Mathematics and Computation*, 174(2), 1498-1525.
- Heragu, S. S., & Srinivasan, M. (2011). Analysis of manufacturing systems via single-class, semi-open queueing networks. *International Journal of Production Research*, 49(2), 295-319.
- Jain, M. (2005). Finite capacity M/M/r queueing system with queue-dependent servers. *Computers & Mathematics with Applications*, 50(1), 187-199.
- Ke, J. C., & Wang, K. H. (2002). A recursive method for the N policy G/M/1 queueing system with finite capacity. *European Journal of Operational Research*, 142(3), 577-594.
- Lariviere, M. A., & Porteus, E. L. (2001). Selling to the newsvendor: An analysis of price-only contracts. *Manufacturing & Service Operations Management*, 3(4), 293-305.
- Nakamura, Syouji. (2014). Optimal number of web server resources for job applications. *International Journal of Industrial Engineering: Theory, Applications and Practice*, 22(1).



- Pardo, M. J., & de la Fuente, D. (2007). Optimizing a priority-discipline queueing model using fuzzy set theory. *Computers & Mathematics with Applications*, 54(2), 267-281.
- Pardo, M. J., & De la Fuente, D. (2008). Optimal selection of the service rate for a finite input source fuzzy queueing system. *Fuzzy Sets and Systems*, 159(3), 325-342.
- Pardo, M. J., & de la Fuente, D. (2008). Design of a fuzzy finite capacity queueing model based on the degree of customer satisfaction: analysis and fuzzy optimization. *Fuzzy Sets and Systems*, 159(24), 3313-3332.
- Pardo, M. J., & de la Fuente, D. (2009). A new technique to optimize the functions of fuzzy profit of queueing models: Application to a queueing model with publicity and renouncement. *Computers & Mathematics with Applications*, 57(5), 850-864.
- Paschalidis, I. C., & Tsitsiklis, J. N. (2000). Congestion-dependent pricing of network services. *Networking, IEEE/ACM Transactions on*, 8(2), 171-184.
- Sindal, R., & Tokekar, S. (2012). A comparative analysis of queueing and fuzzy logic based admission control schemes in CDMA cellular Network. *Procedia Technology*, 4, 145-150.
- Smith, J. M. (2003). M/G/c/K blocking probability models and system performance. *Performance Evaluation*, 52(4), 237-267.
- Sumita, U., Masuda, Y., & Yamakawa, S. (2001). Optimal internal pricing and capacity planning for service facility with finite buffer. *European Journal of Operational Research*, 128(1), 192-205.
- Taha, H. A. (2003). *Operations Research: An Introduction*, 7<sup>th</sup> ed., Prentice-Hall, NJ.
- Takagi, H., Yoshino, H., Matoba, N., Azuma, M., & Shirakabe, M. (2005). System capacity calculation for packet-switched traffic in the next generation wireless systems, part II: Batch arrival M/G/1 nonpreemptive priority queueing model for transmission over a radio channel. Submitted to ITC-19.
- Thomas, N. (2006). Approximation in non-product form finite capacity queue systems. *Future Generation Computer Systems*, 22(7), 820-827.
- 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(12), 899-914.
- Wesołek, P., & Kubiak, M. (2006). Multiobjective fuzzy approach to the vehicle routing problem with time windows. *Diagnostyka*, 21-24.
- Zhang, Q. (2012). Multi-agent based bank queueing model and optimization. *Advances in Electronic Engineering, Communication and Management, Lecture Notes in Electrical Engineering*, 140(2), 121-127.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3), 338-353.
- Zadeh, L. A. (1971). Similarity relations and fuzzy orderings. *Information sciences*, 3(2), 177-200.
- Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *Systems, Man and Cybernetics, IEEE Transactions on*, (1), 28-44.
- Zadeh, L. A. (1974). *The concept of a linguistic variable and its application to approximate reasoning* (pp. 1-10). Springer US.
- Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—II. *Information sciences*, 8(4), 301-357.
- Zadeh, L. A. (1983). A computational approach to fuzzy quantifiers in natural languages. *Computers & Mathematics with applications*, 9(1), 149-184.
- Zadeh, L. A. (1997). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy sets and systems*, 90(2), 111-127.
- Zimmermann, H. J. (1978). Fuzzy programming and linear programming with several objective functions. *Fuzzy sets and systems*, 1(1), 45-55.
- Zimmermann, H. J. (1984). Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Decision Analysis, North-Holland, Amsterdam*.
- Zimmermann, H. J. (1985). Applications of fuzzy set theory to mathematical programming. *Information sciences*, 36(1), 29-58.
- Zimmermann, H. J. (2001). *Fuzzy set theory—and its applications*. Springer Science & Business Media.

Ziya, S., Ayhan, H., & Foley, R. D. (2006). Optimal prices for finite capacity queueing systems. *Operations Research Letters*, 34(2), 214-218.