

# A Multi-Stage Queue Approach to Solving Customer Congestion Problem in a Restaurant

Adegoke S. Ajiboye, Kayode A. Saminu

Department of Statistics, The Federal University of Technology, Akure, Nigeria

Email: [asajiboye@futa.edu.ng](mailto:asajiboye@futa.edu.ng), [asajib11@gmail.com](mailto:asajib11@gmail.com)

**How to cite this paper:** Ajiboye, A.S. and Saminu, K.A. (2018) A Multi-Stage Queue Approach to Solving Customer Congestion Problem in a Restaurant. *Open Journal of Statistics*, 8, 302-316.

<https://doi.org/10.4236/ojs.2018.82019>

**Received:** October 19, 2017

**Accepted:** April 10, 2018

**Published:** April 13, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The multistage queue model was developed for a situation where parallel and unrelated queues exist at the first stage only. These queues merged into single queues at the remaining stages. The parallel queues offer services that are different from one another and customers arrive to join the queue that offer services that they need. The mathematical model was developed assuming an M/M/1 queue system and the measures of effectiveness were derived. The model was applied to solve the problem of customer congestion in a restaurant in the city of Ibadan, Nigeria that serves three different local delicacies. The three local delicacies constitute three different queues at the first stage. The second stage consists of only one queue which is for purchase of drinks and the third stage which is the last stage is for payment. Every customer in the restaurant passes through the three stages. Utilization factors for the five queues were determined and found to range from 70% to 97%. The average time spent by customers in the system was found to be 543.04 minutes. A simulation study using what-if scenario analysis was performed to determine the optimum service configuration for the system. The optimum configuration reduced average time for customers in the system from 543.04 minutes to 13.47 minutes without hiring new servers.

## Keywords

Multistage Queues, Customer Congestion, What-If Scenario Analysis, Queue Network

## 1. Introduction

Queue is a common phenomenon. It occurs in all aspects of life. Queue involves waiting in line. We wait in line in our cars in traffic jams or at toll booths; we wait on hold for an operator to pick up our telephone calls; we wait in line at

supermarkets to check out; we wait in line at fast-food restaurants; and we wait in line at banks and post offices. Customers do not, in generally, like these situations, and the managers of the establishments at which they wait also do not like it since it can lead to loss of business through customer dissatisfaction. Why then do we have waiting-in-line?

Gross *et al.* [1] provides an answer to the question. According to him, demands for services are usually more than facilities available for service delivery. This is due to many factors. Some of the factors are shortage of available servers, inability of a business to provide the level of service necessary to prevent waiting-in-line, limitation of space for service facilities and poor arrangement of servers and number of stages a customer would have to pass through before he/she complete service in the system may be another factors.

### 1.1. Multistage Queue System/Network

Multistage queue system is a queue in which customers are served at more than one station, arranged in a network structure. It is a collection of nodes connected by a set of paths. In a queue network, a group of servers operating from the same facility is identified as a node or a station. Examples of multi stage queues are found in Computer programming, communication, banks and manufacturing systems. In a multi stage queue network, customers demand service from more than one server. Queue networks better represents the real world situation than a single system. A multi stage network is represented diagrammatically as shown in Figure 1.

### 1.2. Characteristics of a Queue System

Queue processes are describe by six basic characteristics, which provide adequate representation of the queue system. The characteristics are 1) arrival pattern of customers, 2) service pattern of servers, 3) queue discipline, 4) system capacity, 5) number of service channels, and 6) number of service stages [1].

Sundarapandian [2] gave the followings as the basic characteristics of a queue system. 1) the input or arrival pattern 2) the service mechanism 3) the queue discipline and 4) the system capacity. Nevertheless, these two researchers are talking about the same thing. Gross *et al.* [1] provides an excellent explanation of

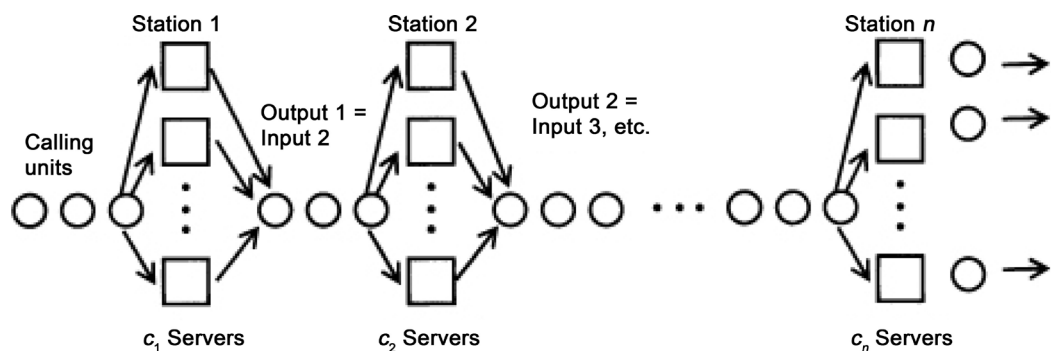


Figure 1. Multistage queue network with  $n$  stations.

these basic characteristics.

### 1.3. Performance Measures of a Queue System

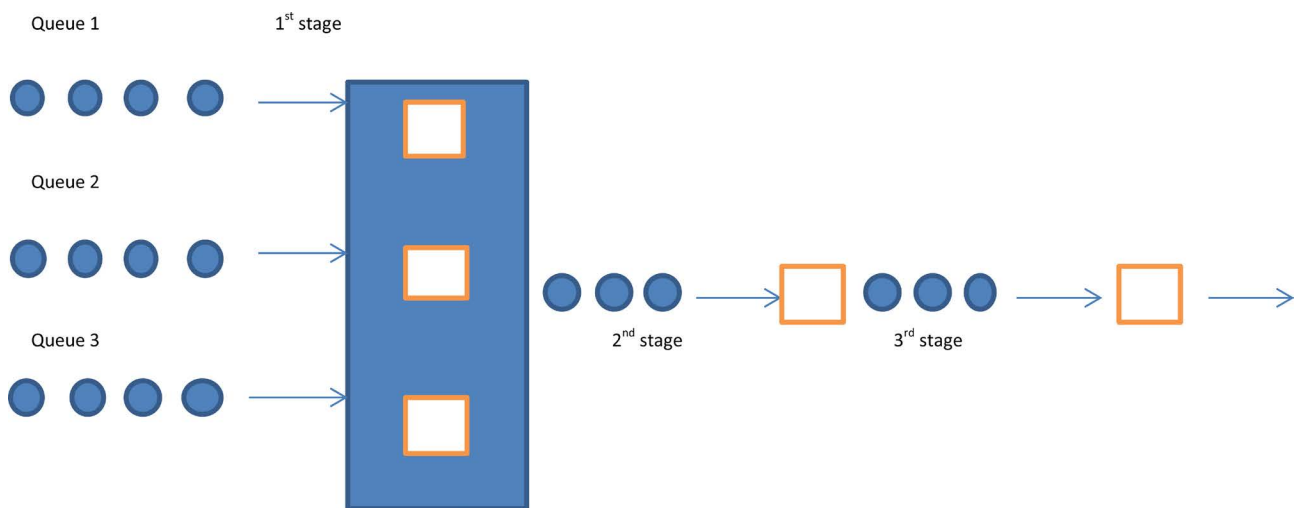
In this work we study a multistage queue system with a certain number of independent parallel servers and multiple queues at all or some of the stages. *i.e.* multiple waiting lines with independent parallel servers in series. This research work reviews the application of queue theory in analyzing performance measures of this type of system and develop discrete-event simulation of customers' flows in the system purposely to carry out what-if-scenario analysis to improve efficiency of the system.

There are various performance measures of a multistage queue system. The ones we are interested in are the total expected number of customers in the system, total expected waiting time of an arbitrary customer in the queues formed at all the stages and the total expected waiting time or lead time (cycle time) that an arbitrary customer would spend in the system.

The main aim of this study is to analyze the customer-service situation in a multistage queue system.. The problem arose in a restaurant in the city of Ibadan, south-western Nigeria that serves three classes of local delicacies during lunch/ dinner services. Because of limited space and large customer pool, congestions frequently occur in the restaurant. A multistage queue model was designed for the restaurant. The three-stage queue model has 3 service points in stage 1 and single service point in each of the remaining two stages. **Figure 2** shows the diagrammatic illustration of the system.

## 2. Methodology

The multistage queue system consists of various stations through which a customer requiring service from such system would have to pass. This contains networks of queues which can be described as a group of nodes (say,  $k$  of them) that is we have  $k$  number of nodes, where each node represents a service facility



**Figure 2.** Diagrammatic illustration of the three-stage queue model for the restaurant.

of some kind with  $s_j$  servers at node  $j, j = 1, 2, \dots, k$ . In the most general cases, customers may arrive from outside the system at any node and may depart from the system from any node. Thus customers may enter the system at some node, traverse from node to node in the system, and depart from some node. Not all customers necessarily enter and leave at the same nodes, or take the same path once they entered the system. Customers may return to nodes previously visited, skip some nodes entirely, and even choose to remain in the system forever [1].

A multistage queue system or queues in series as described in [3] are open networks where

$$\gamma_j = \begin{cases} \gamma, & j = 1, 2, \dots, k \\ 0, & \text{elsewhere} \end{cases} \quad (1)$$

and

$$\delta_{i,j} = \begin{cases} 1, & (j = i + 1, 1 \leq i \leq k - 1) \\ 1, & (i = k, j = 0) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

since the nodes can be viewed as forming a series system with flow always in a single direction from node to node. This is supported by a remarkable theorem propounded by Jackson [3] that if 1) inter-arrival times for a series queue system are exponential with rate  $\gamma$  2) service times for each stage  $j$  server are exponential and 3) each stage has an infinite-capacity waiting room, then inter-arrival times to each stage of the queue system are exponential with rate  $\gamma$ , the arrival rate of each customer from node to node.

## 2.1. Operating Parameters and Notations

The following system operating parameters and notations were used in this research work.

- $N$  = total number of customers in the system.
- $\beta$  = mean arrival rate at a single-channel, single server queue.
- $\tau$  = mean service rate of a single channel, single server queue.
- $\rho$  = system intensity or load, utilization factor.
- $P_0$  = steady-state probability of an idle server.
- $P_n$  = steady-state probability of exactly  $n$  customers in the system.
- $\beta_i$  = mean arrival rate of customers at the  $i^{\text{th}}$  queue of the first stage.
- $\gamma_j$  = effective mean or expected arrival rate of customers at the  $j^{\text{th}}$  stage.
- $\pi_i$  = probability that an arbitrary customer joins the  $i^{\text{th}}$  queue at the first stage.
- $\tau_i$  = mean service rate of the  $i^{\text{th}}$  queue server at the first stage.
- $\alpha_j$  = mean service rate of the  $j^{\text{th}}$  stage server.
- $\rho_j$  = system utilization factor of the  $j^{\text{th}}$  stage server.
- $v$  = total number of customers in the first stage of the queue system.
- $\omega_i$  = system utilization factor of the  $i^{\text{th}}$  queue server at the first stage.
- $\delta_{j-1,j}$  = transition probability of a customer after completing service at the  $(j - 1)^{\text{th}}$  stage, proceeding to the  $j^{\text{th}}$  stage.

- $T$  = waiting time of a customer in the multistage queue system.
- $T_q$  = total waiting time of a customer in queues formed at different stages.
- $T_{qi}$  = waiting time of a customer in the  $i^{\text{th}}$  queue of the first stage.
- $T_{qj}$  = waiting time of a customer in queue at the  $j^{\text{th}}$  stage.
- $W_q$  = expected total waiting time of customers in queues formed at different stages.
- $W_{qi}$  = average or expected waiting time of a customer in the  $i^{\text{th}}$  queue at the first stage.
- $W_{qj}$  = average or expected waiting time of a customer at the  $j^{\text{th}}$  stage.
- $W_s$  = expected waiting time of a customer in the multistage queueing system.
- $P_c$  = probability that there are exactly C customers at 2-k stages.
- $P_v$  = probability that there are exactly V customers at the first stage.

### 2.2. The Multi-Stage Queue Model

We assume there are  $k$  stages in the queue model and stage 1 has  $h$  parallel single-server queues which are assumed to be independent. That is, they offer different services. It is also assumed that there is no balking, reneging or jockeying in the system; the  $i^{\text{th}}, i=1,2,\dots,h$  queue follows a  $M/M/1$  queue system; the system is a feed-forward queue network and bypassing is not allowed. Let the arrival rate at the  $i^{\text{th}}, i=1,2,\dots,h$  queue be  $\beta_i$  and the service rate be  $\tau_i$ . Suppose the probability that an arriving customer joins queue  $i$  is  $\pi_i$ . We want to determine the effective rate of arrival and service in the first stage.

Let  $\gamma_1$  = expected customer arrival rate into stage 1. Then  $\gamma_1 = \sum_{i=1}^h \pi_i \beta_i$

We estimate  $\pi_i$  using  $\pi_i = \frac{\beta_i}{\sum_{i=1}^h \beta_i}, \sum_{i=1}^h \pi_i = 1,$

Similarly  $\alpha_i = \sum_{i=1}^h \pi_i \tau_i.$

The general traffic intensity for stage one is given as  $\omega = \gamma_1 / \alpha_1.$

The arrival rate to stage 2 is given by  $\gamma_2 = \delta_{12} \gamma_1.$  This holds since the system is a feed-forward network and bypassing is not allowed. Recursively,

$$\gamma_j = \delta_{j-1,j} \gamma_j, j = 1, 2, \dots, k.$$

The joint equilibrium distribution of the number of customers in the system is

$$P_{v,c} = P_n = \prod_{i=1}^x (1 - \omega_i) \omega_i^{v_i} \cdot \prod_{j=2}^k (1 - \rho_j) \rho_j^{n_j}. \tag{3}$$

The total expected number of customers in the system is

$$L_s = \sum_{j=1}^k L_{sj} = \sum_{i=1}^x \frac{\omega_i}{1 - \omega_i} + \sum_{j=2}^k \frac{\rho_j}{1 - \rho_j} \tag{4}$$

The total expected waiting time of an arbitrary customer in the queues formed at all stages ( $W_q$ ) is

$$W_q = \frac{\beta_i^2}{\tau_i (\tau_i - \beta_i) \gamma_1} + \frac{1}{\tau_i} - \frac{1}{\beta_i} + \sum_{j=2}^k \frac{\gamma_j^2}{\alpha_j (\alpha_j - \gamma_j) \gamma_j} \tag{5}$$

for  $i = 1, 2, \dots, x.$

The total expected waiting time of an arbitrary customer in the system ( $W_s$ ) is

derived as

$$W_s = \frac{\beta_i^2}{\tau_i(\tau_i - \beta_i)\gamma_1} + \frac{1}{\tau_i} - \frac{1}{\beta_i} + \frac{1}{\tau_i} + \sum_{j=2}^k \frac{\rho_j^2}{1 - \rho_j} + \frac{1}{\alpha_j}, \quad (6)$$

for  $i = 1, 2, \dots, x$ .

For details of the derivations of these measures go to **Appendix**.

### 2.3. Simulation

Simulation is a powerful tool for the analysis of new system designs, retrofits to existing systems and proposed changes to operating rules. Conducting a valid simulation is both an art and a science [4].

Stewart [5] defines simulation as imitation of real-world scenario systems. The system might be manufacturing system, communication system, health-care system, restaurant system etc. There are many types and kinds of simulation namely: Monte Carlo-type simulation, discrete-event simulation to mention but few. In this research work, we limit ourselves to discrete, stochastic process oriented simulation *i.e.* a discrete-event simulation.

Simulation models are abstractions of a real-world, or proposed real-world, system. In other words, the models do not contain every aspect and detail of the system that they are attempting to imitate [6].

The majority of simulation models could be described as being visual interactive simulations (VIS). This was a concept first introduced in [7]. The idea is that the model provides a visual display showing an animation of the model as it runs.

We use SIMUL8, a software package fortified with dynamic, powerful Graphical User Interface (GUI) to carry out a visual interactive discrete-event simulation of customers flow in the multistage queue system. The *what-if* scenario analysis to compare a number of system alternatives was also performed. It allows for the identification of problems, bottlenecks and design shortfalls before building or modifying a system.

## 3. Results and Discussion

Data used for the analysis of the model were collected from *Iyaduuni* Restaurant in the city of Ibadan Nigeria. The restaurant has about 25 tables and chairs for customers and more than 10 servers work on shift basis. At least 5 servers are on duty per shift. There are 4 to 5 waiters and waitresses whose jobs are to secure comfortable places for customers to seated and clean the tables when customers leave.

Data were collected on customer arrival time on each of the three starting queues in the first stage and the other queues in the second and third stages. The time service begins and time service ends in all the queues were also recorded. The queues are assumed to be infinite capacity and customers visit all the stages. Data were collected for 5 days. The average arrival rates of customers and average service rates, together with the utilization factors of servers are displayed in

Table 1 and Table 2 respectively.

Total expected number of customers in the system is computed using (2.4) as

$$\frac{0.97}{1-0.97} + \frac{0.93}{1-0.93} + \frac{0.7}{1-0.7} + \frac{0.93}{1-0.93} + \frac{0.94}{1-0.94} = 76 \text{ customers}$$

The total expected waiting time of an arbitrary customer for a complete service in the system who joined 1<sup>st</sup>, 2<sup>nd</sup> or 3<sup>rd</sup> queue at the first stage is obtained using (2.6) as  $W_s = (10.014 \text{ mins}, 5.874 \text{ mins}, 3.594 \text{ mins})$  respectively.

### 3.1. Simulation Runs

A simulation model was fitted to the system using the inter arrival times and the service times. Both are exponentially distributed. The basic model as displayed in Figure 3 was simulated for a time equivalent of 2 weeks of operation. The average or expected waiting time of customers at each node and average or expected lead-time were obtained and displayed in Table 3.

Based on a two-week simulation run of the restaurant system, the average time spent on the 1<sup>st</sup> queue at stage 1 for local delicacy 1 (Amala) by an arbitrary customer was 46 mins, likewise 3 mins on the 2<sup>nd</sup> queue for local delicacy 2 and 2

Table 1. Mean arrival rates of customers into the system and mean service rates at the queues.

	1 <sup>ST</sup> QUEUE AT STAGE 1	2 <sup>ND</sup> QUEUE AT STAGE 1	3 <sup>RD</sup> QUEUE AT STAGE 1	2 <sup>ND</sup> STAGE QUEUE	3 <sup>RD</sup> STAGE QUEUE
MEAN ARRIVAL RATE (IN 10 MINUTES INTERVALS)	6.96	2.52	1.59	9.58	9.68
MEAN SERVICE RATES (IN 10 MINUTES INTERVAL)	7.15	2.7	2.26	10.31	10.36

Table 2. Utilization factors of servers.

	1 <sup>ST</sup> QUEUE AT STAGE 1	2 <sup>ND</sup> QUEUE AT STAGE 1	3 <sup>RD</sup> QUEUE AT STAGE 1	2 <sup>ND</sup> STAGE QUEUE	3 <sup>RD</sup> STAGE QUEUE
UTILIZATION FACTOR	0.97	0.93	0.70	0.93	0.94

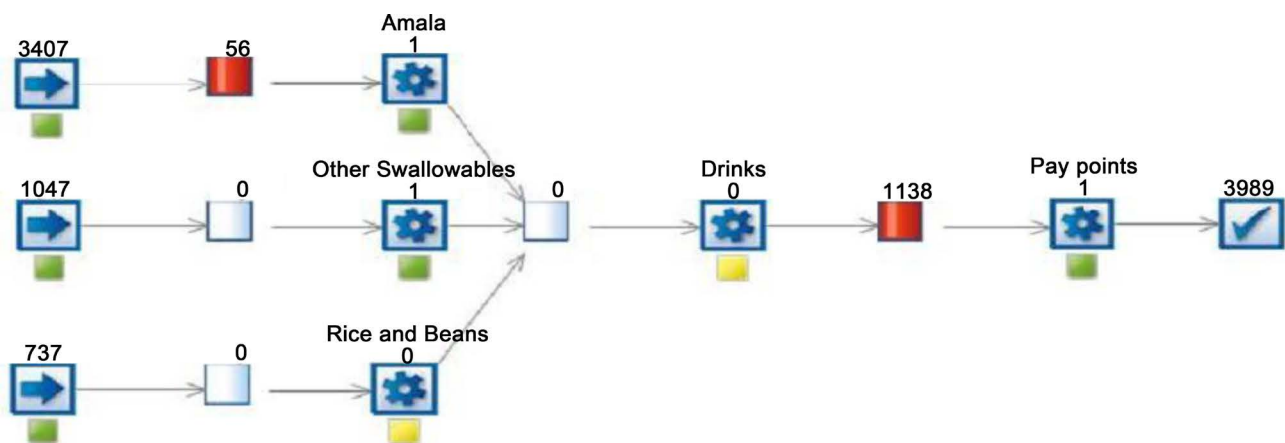


Figure 3. A Visual Interactive display of the simulation for the basic model.

mins on the 3<sup>rd</sup> queue for local delicacy 3. There were lumps or bottlenecks on the 1<sup>st</sup> queue of the 1<sup>st</sup> stage and last stage (Paypoint) of the system.

### 3.2. Performing What-If Scenario Analysis

We noticed that there were bottlenecks at category I (local delicacy group 1) service point of stage I and Paypoint service point. We, therefore, perform a *what-if scenario* analysis to compare alternatives and reduce both time spent in queues and time spent in the system as a whole. This was done by shifting the server attending to customers interested in local delicacy 3, being the least utilized activity to join the one at Paypoint where a bottleneck was observed. The server at local delicacy 2 service centre was assigned to attend to those who would be interested in local delicacy 2. The model is shown in **Figure 4**.

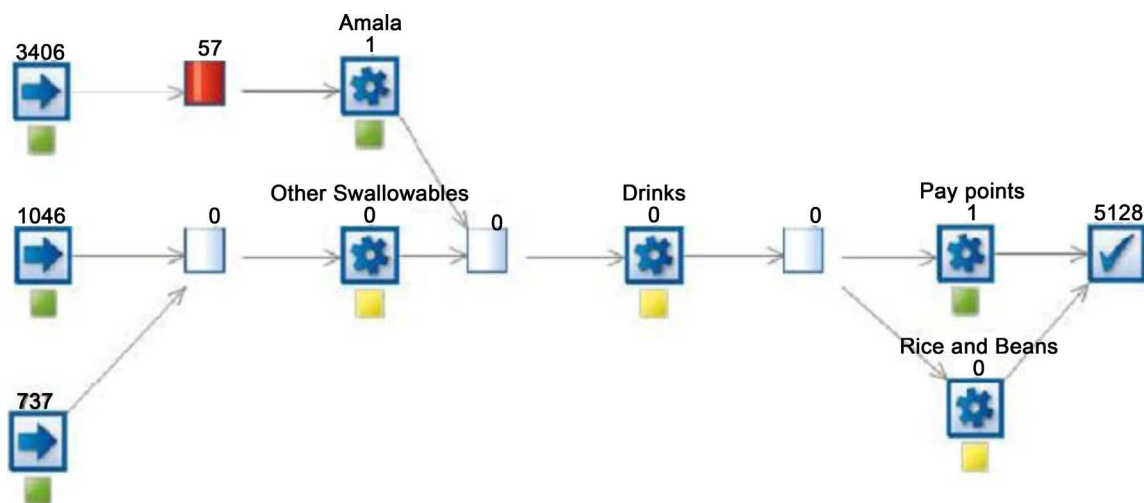
The summary of the analysis is given **Table 4**.

The average time spent in the system has been reduced from 543.02 mins to 42.81 mins and the average queue time at Paypoint has also been reduced (**Table 4**). This is a great improvement but we observed that large queue still occurred at local delicacy 1 service point. Time spent here is still 46 min. To improve on this we used the configuration in **Figure 5**.

The average time spent by a customer in the system has reduced further to

**Table 3.** Result of the simulation for the basic model.

Queue	Average Queue Time
Stage 3	511 mins (Pay-point)
Stage 2	1 min (Drinks)
Stage 1, Queue 1	46 mins (local Delicacy 1)
Queue 2	3 mins (Local Delicacy 2)
Queue 3	2 mins (Local Delicacy3)
Average Time in System	543.02 mins

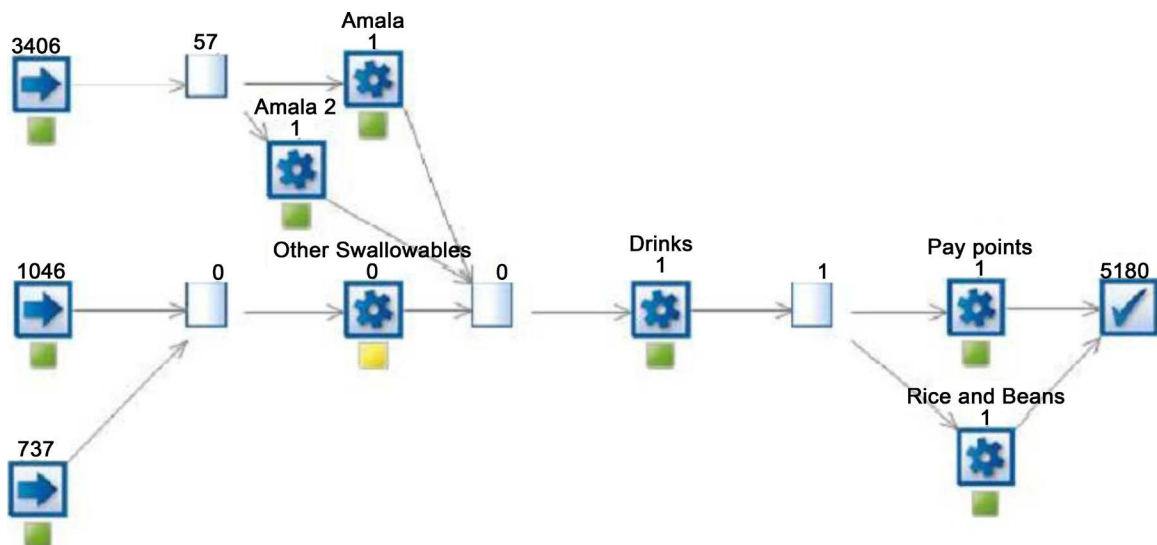


**Figure 4.** A visual interactive simulation run for the first what-if scenario analysis.



**Table 4.** Average queueing time the first *what if scenario* analysis.

Name	Average Queueing Time
Paypoint	11 mins
Drink	3 min
Local delicacy 3	2 mins
Local delicacy 2	3 mins
Local delicacy 1	46 mins
Average Time in System	42.81 min



**Figure 5.** A visual interactive simulation run for the second *what if scenario* analysis.

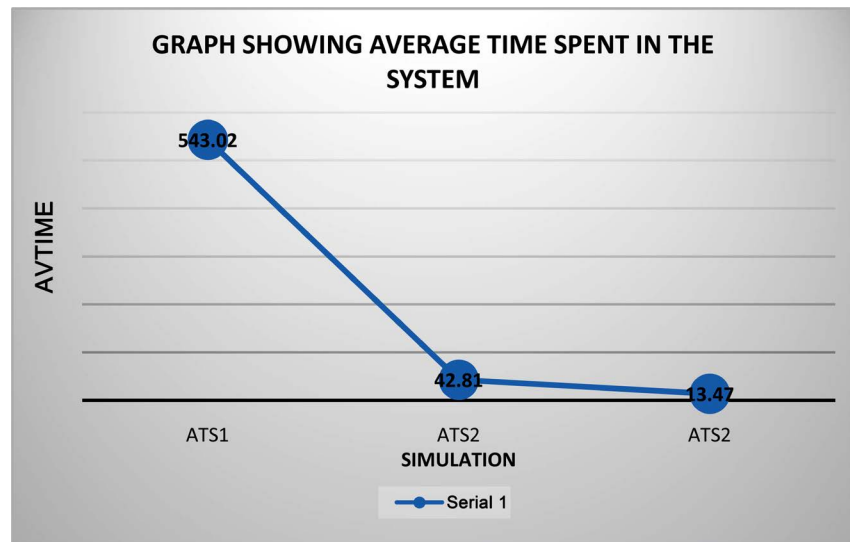
**Table 5.** Average queueing time after the second *what-if scenario* analysis.

Name	Average Queueing Time
Paypoint	1 min
Drink	1 min
Local Delicacy 2 & 3	11 mins
Local Delicacy 1	0 min
Average Time in System	13.47 min.

13.47 mins (**Table 5**). Also local delicacy 1 now enjoys zero waiting time.

### 3.3. Discussion

The analytical results show that at any point in time as many as 76 customers are present in the system. The utilization factors for the five queues in the system range from 70% to 97% (**Table 2**). The time spent by customers on queue, starting from stage 1 and ending in stage 3 from any of the three queues of stage 1 are respectively 10.4 minutes for local delicacy 1, 5.87 minutes for local delicacy 2 and 3.594 minutes for local delicacy 3. Customers requiring local delicacy 1



**Figure 6.** Decrease in average time spent in the system.

spend longer time on queue, reflecting the high demand for this meal. The utilization factor for this queue stands at 97% which is the highest.

The simulation study provides different operation scenarios to the management of the restaurant. Through the study the average time spent by customers in the system reduced from 543.02 minutes to as little as 13.47 minutes. This was achieved by considering different configurations of servers in the system without adding or subtracting any server. This means the reduction was achieved with no additional cost to management. Yet as customers spend less time in the restaurant more space are freed for addition customers to enjoy the services of the restaurant. This will translate to higher profit, improved customer satisfaction and customer goodwill. This improvement is illustrated in **Figure 6**.

#### 4. Conclusion

This study has shown an application of queue theory in analyzing a multistage queue system consisting of M/M/1 (Poisson arrivals, exponential service times, a single server and infinite room capacity) queue models. Evaluation of performance measures for the queues were done after deriving expressions for the effectiveness measures. We assessed the level of service delivery through utilization factors of each M/M/1 queue. We found out that majority of the servers are well utilized as revealed by the values of the utilization factors obtained. Also, discrete-event simulation that gave close imitation of the system was developed and run to visualize and ascertain the feasible operations of the system. What-if scenario analysis was carried out in order to make vital and necessary suggestions for decision-making that will improve the efficiency of the system. The method led to significant reduction in the waiting time of an arbitrary customer in the system by resolving areas where bottlenecks were observed in the system. It also provides an effective mean of managing the queues for maximum customer satisfaction at no additional cost.

## References

- [1] Gross, D., Shortle, J.F., Thompson, J.M. and Harris, C.M. (2008) Fundamentals of Queueing Theory. 4th Edition, John Wiley and Sons Inc., New Jersey, 1.  
<https://doi.org/10.1002/9781118625651>
- [2] Sundarapandian, V. (2009) Probability, Statistics and Queueing Theory. PHI Learning Private Ltd., New Delhi, 686-709.
- [3] Jackson, J.R. (1957) Networks of Waiting Lines. *Operations Research*, **5**, 518-521.  
<https://doi.org/10.1287/opre.5.4.518>
- [4] John, S.C. (2003) Introduction to Modelling and Simulation. *Proceedings of the 2003 Winter Simulation Conference*, Marietta.
- [5] Stewart, R. (2004) Simulation: "The Practice of Model Development and Use". John Wiley and Sons Ltd., West Sussex, 1-33.
- [6] Roger, B., Stewart, R. and Kathy, K. (2011) Conceptual Modelling for Discrete-Event Simulation. CRC Press, Taylor & Francis Group, London, New York, 1-26.
- [7] Hurrion, R.D. (1976) The Design, Use and Required Facilities of an Interactive Visual Computer Simulation Language to Explore Production Planning Problems. PhD Thesis, University of London, London.

## Appendix

### Derivation of the Joint Equilibrium Distribution of the Number of Customers in the System

The joint probability of  $N$  customers in the system is given by

$$P_n = P(N_1 = n_1)P(N_2 = n_2) \cdots P(N_k = n_k)P(0)$$

Let  $C_j = N_j$ , for  $j = 2, 3, \dots, k$

Recall that  $N_1 = V = \sum_{i=1}^x v_i$ , Therefore,  $P_n = P_v P_c P(0)$

$$\begin{aligned} P_{v,c} &= P(V_1 = v_1, V_2 = v_2, \dots, V_x = v_x)P(C_2 = c_2) \cdots P(C_k = c_k)P(0) \\ &= P(v_1) \cdot P(v_2) \cdots P(v_x) \cdot P(c_2) \cdot P(c_3) \cdots P(c_k)P(0) \end{aligned}$$

where  $P(0)$  is a normalizing constant.

From a single channel M/M/1 queue model,

$$P_n = P_0 \rho^n$$

where

$$P_0 = 1 - \rho,$$

Following this, Equation (3) becomes

$$P_{v,c} = \omega_1^{v_1} \omega_2^{v_2} \cdots \omega_x^{v_x} \cdot \rho_2^{c_2} \rho_3^{c_3} \cdots \rho_k^{c_k} P(0) \tag{7}$$

Using normalizing condition,

$$\left( \sum_{v_x=0}^{\infty} \cdots \sum_{v_2=0}^{\infty} \sum_{v_1=0}^{\infty} \omega_1^{v_1} \omega_2^{v_2} \cdots \omega_x^{v_x} \cdot \sum_{c_k=0}^{\infty} \cdots \sum_{c_3=0}^{\infty} \sum_{c_2=0}^{\infty} \rho_2^{c_2} \rho_3^{c_3} \cdots \rho_k^{c_k} \right) P(0) = 1$$

Making

$$P(0) = \frac{1}{\sum_{v_x=0}^{\infty} \cdots \sum_{v_2=0}^{\infty} \sum_{v_1=0}^{\infty} \omega_1^{v_1} \omega_2^{v_2} \cdots \omega_x^{v_x} \cdot \sum_{c_k=0}^{\infty} \cdots \sum_{c_3=0}^{\infty} \sum_{c_2=0}^{\infty} \rho_2^{c_2} \rho_3^{c_3} \cdots \rho_k^{c_k}}$$

Since each term of the denominator is a geometric series,

Therefore,

$$\begin{aligned} P(0) &= (1 - \omega_1)(1 - \omega_2) \cdots (1 - \omega_x) \cdot (1 - \rho_2)(1 - \rho_3) \cdots (1 - \rho_k) \\ &= \prod_{i=1}^x (1 - \omega_i) \prod_{j=2}^k (1 - \rho_j) \end{aligned} \tag{8}$$

Substituting RHS of Equation (8) for  $P(0)$  in Equation (7)

We get,

$$\begin{aligned} P_{v,c} &= \omega_1^{v_1} \omega_2^{v_2} \cdots \omega_x^{v_x} (1 - \omega_1)(1 - \omega_2) \cdots (1 - \omega_x) \\ &\quad \cdot \rho_2^{c_2} \rho_3^{c_3} \cdots \rho_k^{c_k} (1 - \rho_2)(1 - \rho_3) \cdots (1 - \rho_k) \\ P_{v,c} &= P_n = \prod_{i=1}^x (1 - \omega_i) \omega_i^{v_i} \cdot \prod_{j=2}^k (1 - \rho_j) \rho_j^{n_j} \end{aligned} \tag{9}$$

### The Total Expected Number of Customers in the System

This is done by decomposing the system into  $k$  individual stages. Let  $L_{sj}$  ( $j = 1, 2, 3, \dots, k$ ) be the expected number of customers at the  $j$ th stage. Therefore,

$$L_s = \sum_{j=1}^k L_{sj}$$

Let  $L_{s1}$  be the expected number of customers at the first stage. Remember that  $N_1 = V = \sum_{i=1}^x V_i$

Therefore,

$$\begin{aligned} L_{s1} &= E(V) = E\left(\sum_{i=1}^x V_i\right) = \sum_{i=1}^x E(V_i) \\ &= E(V_1) + E(V_2) + \dots + E(V_x) \\ &= \sum_{v_1=0}^{\infty} v_1 P_{v_1} + \sum_{v_2=0}^{\infty} v_2 P_{v_2} + \dots + \sum_{v_x=0}^{\infty} v_x P_{v_x} \end{aligned}$$

where  $P_{v_i}$  is the probability of  $V_i (i = 1, 2, \dots, x)$  customers in the  $i$ th queue of the first stage.

$$\begin{aligned} L_{s1} &= \sum_{v_1=0}^{\infty} v_1 (1 - \omega_1) \omega_1^{v_1} + \sum_{v_2=0}^{\infty} v_2 (1 - \omega_2) \omega_2^{v_2} + \dots + \sum_{v_x=0}^{\infty} v_x (1 - \omega_x) \omega_x^{v_x} \\ &= (1 - \omega_1) \omega_1 \sum_{v_1=1}^{\infty} v_1 \omega_1^{v_1-1} + (1 - \omega_2) \omega_2 \sum_{v_2=1}^{\infty} v_2 \omega_2^{v_2-1} + \dots + (1 - \omega_x) \omega_x \sum_{v_x=1}^{\infty} v_x \omega_x^{v_x-1} \\ &= \frac{(1 - \omega_1) \omega_1}{(1 - \omega_1)^2} + \frac{(1 - \omega_2) \omega_2}{(1 - \omega_2)^2} + \dots + \frac{(1 - \omega_x) \omega_x}{(1 - \omega_x)^2} \\ &= \frac{\omega_1}{1 - \omega_1} + \frac{\omega_2}{1 - \omega_2} + \dots + \frac{\omega_x}{1 - \omega_x} = \sum_{i=1}^x \frac{\omega_i}{1 - \omega_i} \end{aligned}$$

Since each of the remaining stages has a single server with single queue, then

$$\begin{aligned} L_{s2} &= E(N_2) = \sum_{n_2=0}^{\infty} n_2 P_{n_2} = \frac{\rho_2}{1 - \rho_2} \\ L_{s3} &= E(N_3) = \sum_{n_3=0}^{\infty} n_3 P_{n_3} = \frac{\rho_3}{1 - \rho_3} \\ &\vdots \\ L_{sk} &= E(N_k) = \sum_{n_k=0}^{\infty} n_k P_{n_k} = \frac{\rho_k}{1 - \rho_k} \end{aligned}$$

Therefore, the total expected number of customers in the system is

$$\begin{aligned} L_s &= L_{s1} + L_{s2} + \dots + L_{sk} \\ L_s &= \sum_{i=1}^x \frac{\omega_i}{1 - \omega_i} + \sum_{j=2}^k \frac{\rho_j}{1 - \rho_j} \end{aligned} \tag{10}$$

**Total Expected Waiting Time of an Arbitrary Customer in the Queues Formed at All Stages ( $W_q$ )**

Starting with the first stage having  $x$  independent M/M/1 queues, we evaluate average or expected waiting time of an arbitrary customer in the  $i$ th queue of the first stage using the following relation:

$$\begin{aligned} E(T_{qi}^{y+1}) &= E(T_{qi}^y) + E(S_i^y) - E(T_i^y) \\ W_{qi}^{y+1} &= W_{qi}^y + \frac{1}{\tau_i} - \frac{1}{\beta_i} \end{aligned}$$

where

$T_{qi}^{y+1}$  = the waiting time of an arbitrary customer in the  $i^{th}$  queue.  
 $W_{qi}^{y+1}$  = the expected waiting time of an arbitrary customer in the  $i^{th}$  queue.  
 $T_{qi}^y$  = the waiting time of the  $y^{th}$  customer in the  $i^{th}$  queue.  
 $W_{qi}^y$  = expected waiting time of  $y^{th}$  customer in the  $i^{th}$  queue.  
 $E(S_i^y)$  = expected service time of the  $y^{th}$  customer in the  $i^{th}$  queue.  
 $E(T_i^y)$  = expected interarrival time between the  $y^{th}$  customer and  $(y + 1)^{st}$  customer in the  $i^{th}$  queue

We assume that there are  $V_i$  customers in the  $i^{th}$  ( $i = 1, 2, \dots, x$ ) queue and that service has started before the arbitrary customer joins the  $i^{th}$  queue.

Let  $L_{qi}$  be the queue length or expected number of customers in the  $i^{th}$  queue of stage 1

$$L_{qi} = E(V_i - 1) = \sum_{v_i=1}^{\infty} (v_i - 1) P_{v_i} = \sum_{v_i=1}^{\infty} (v_i - 1) \omega^{v_i} (1 - \omega_i)$$

Let  $d = (v_i - 1)$  and substitute in the summation.

Thus, it follows that

$$\begin{aligned} &= \sum_{v_i=1}^{\infty} d \omega_i^{d+1} (1 - \omega_i) \\ &= \omega_i^2 (1 - \omega_i) \sum_{d=1}^{\infty} (d) \omega_i^{d-1} \\ &= \frac{\omega_i^2 (1 - \omega_i)}{(1 - \omega_i)^2} \end{aligned}$$

Hence we have,

$$L_{qi} = \frac{\omega_i^2}{1 - \omega_i} = \frac{\beta_i^2}{\tau_i (\tau_i - \beta_i)}$$

By Little's formulae,

$$L_{qi} = \gamma_1 W_{qi}$$

Therefore,

$$W_{qi}^m = \frac{L_{qi}}{\gamma_1} = \frac{\beta_i^2}{\tau_i (\tau_i - \beta_i) \gamma_1}$$

$$E(S_i^m) = \frac{1}{\tau_i}$$

and

$$E(T_i^m) = \frac{1}{\beta_i}$$

Substituting these three terms into the relation, we get the expected waiting time of an arbitrary customer in the  $i^{th}$  queue of the first stage as:

$$W_{qi}^{y+1} = \frac{\beta_i^2}{\tau_i (\tau_i - \beta_i) \gamma_1} + \frac{1}{\tau_i} - \frac{1}{\beta_i}$$

At the  $j^{th}$  stage, the arbitrary customer is the  $n$ th customer in the  $j^{th}$  queue

(  $j = 2, 3, \dots, k$  ) station. Therefore,  $L_{qj}$ , the queue length of  $n$ th customer in the  $j^{\text{th}}$  queue is given as:

$$L_{qj} = E(N_j - 1) = \sum_{n_j=1}^{\infty} (n_j - 1) P_{n_j}$$

where  $P_{n_j}$  is the probability of  $n_j$  customers at the  $j^{\text{th}}$  stage.

It follows from equation \* that

$$L_{qj} = \frac{\rho_j^2}{1 - \rho_j} = \frac{\gamma_j^2}{\alpha_j (\alpha_j - \gamma_j)}$$

Hence,

$$W_{qj} = \frac{\gamma_j^2}{\alpha_j (\alpha_j - \gamma_j) \gamma_j} \quad \text{for } j = 2, 3, \dots, k$$

The total average waiting time of an arbitrary customer in the queues formed at all the stages is obtained as

$$W_q = \frac{\beta_i^2}{\tau_i (\tau_i - \beta_i) \gamma_1} + \frac{1}{\tau_i} - \frac{1}{\beta_i} + \sum_{j=2}^k \frac{\gamma_j^2}{\alpha_j (\alpha_j - \gamma_j) \gamma_j} \tag{11}$$

for  $i = 1, 2, \dots, x$

**The total expected waiting time of an arbitrary customer in the system ( $W_s$ )**

We obtained equation for the expected lead-time  $E(T)$  (or cycle time) for an arbitrary customer, including waiting times and service times spent in the system from the first arrival until the final departure.

By Little’s formulae, the expected waiting time of a customer in a single channel system is given as:

$$W_s = W_q + \frac{1}{\alpha}$$

Therefore,

$$E(T) = E(T_{qi}^{y+1}) + E(S_i) + \sum_{j=2}^k E(T_{qj}) + E(S_j) \tag{12}$$