Faculty of Natural and Applied Sciences Journal of Computing and Applications
Print ISSN: 3026-8133
www.fnasjournals.com
Volume 3; Issue 1; March 2026; Page No. 7-18.
DOI: https://doi.org/10.63561/jca.v3i1.1197

An Improved Routing Protocol in Queue Networks with Heterogeneous Service Rates

***[1]Gbadebo, A.D., [2]Raji-Lawal, H.Y., [1]Abimbola, O.G., & [3]Odupe, T.A.A.**
[1]Department of Computer Science Education, Lagos State University of Education, Oto/Ijanikin, Lagos State, Nigeria
[2]Department of Computer Science, Lagos State University, Ojo, Lagos State, Nigeria
[3]Department of Mathematics Education, Lagos State University of Education, Oto/Ijanikin, Lagos State, Nigeria

***Corresponding author email**: gbadeboad@lasued.edu.ng

Abstract
While there are enormous studies on methods of attaining optimality in packets' routing in queue networks in literature, none of these had sufficiently addressed the attainment of optimality and network stability concurrently. Since network stability influences network optimality, it is necessary to ensure optimal admissibility of customers to ensure optimal servers' usage and network stability. The proposed model is a fuzzified routing protocol for enhanced system performance and network cost reduction. The model was benchmarked with Elementary Job Routing Problem with heterogeneous servers and general cost structures. Simulation was done using OMNeT++ as framework while dataset were randomly generated. Simulation results indicated that the proposed model has higher throughput with minimal customers' losses and memory consumption. With these results, it was concluded that the proposed model is more optimal in customers' routing as well as queue network control.

Keywords: Optimality, Fuzzy Control, Servers, Network System, Defuzzifier

Introduction

Queues are common daily events. This is so as people wait in parks, banks, event centres, supermarkets, airports, hospitals, etc. These are visible queues. There are other types of queues such as voice calls or data packets in communication networks. These are invisible queues. More often, queues are undesirable as it come with costs. Queues exist as a result of inadequate resources to satisfy demand (Chen et al. 2023). This can be attributed to a number of factors. Abubakar et al. (2022) stated that servers might be unavailable as a result of cost limitations. In other cases, it may not always be cost effective to give the level of service needed to prevent waiting. Congestion control had become an important factor needed to ensure optimal network performance. This is as a result of recent and widespread use and application of computer networks. Congestion occurs in buffer of network router when arrivals exceed the capacity of the available network resources, including buffer space, router processing speed, etc (Agarwal et al. 2022). More importantly, multiple queue networks are applied to solving practical network congestion problems. It therefore becomes imperative to ensure there are reliable queue network models to ensure optimal network performance. Simulations are often applied in studying performance of multiple queue network systems (Abdali et al. 2023). Other researchers had channeled efforts at applying mathematical models of multiple queue systems as they are cheaper and simplified than simulation and experimental applications which are capable of analyzing dynamic approaches that apply to network systems.

With recent advances in communications technology, it becomes expedient to determine how best to route customers in queues to achieve optimal network performance. With the routing of data traffic in the Internet for instance, it is possible to model alternative routes as parallel service providers, i.e. servers. The same applies in cloud computing in which each customer need to be assigned to specified servers. The task of routing customers for service is an important

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

operation in contemporary network systems. While the common approach of join-the-shortest-queue routing technique reduces average mean response time in a general setting environment, this trend had changed recently as a result of complexities that characterize network behaviour. This had resulted in minimal advantage as a result of complexities emanating from infinite state spaces. The present study aims at applying fuzzy-based approach to ensuring optimal customers' routing thereby ensuring improved network performance.

Related works

The problem of customers' routing to servers had been of great concern in queue theory. In contemporary queue networks, this problem had lingered in varied forms. The problem was first considered by Haight (1958). Hyytia et al. (2017) studied a simple routing scenario involving heterogeneous service system with general structures of costs. The authors minimized state space finitely thereby estimating network mean performance while also optimizing policies of the routing system involving a wide range of cost structures. In a similar study, Sakalauskas et al. (2024) studied a model which is concerned with queue stalling involving infinite heterogeneous service system. It was possible to derive the state graph as well as corresponding linear parameters applicable in the steady-state probabilities derivable by means of a standard Markov chain. The solution obtained from steady-state probabilistic model was numerically stable while the complexities of corresponding expressions varied with the amount of queue states. Armony (2005) considered a service system having multiple servers with a single customer class. The author suggested a routing instruction that minimizes its steady-state queues asymptotically as well as waiting time which involves a fastest-server-first scheme. The author considered the Halfin-Whitt multiple-server large-traffic system which achieved high performance in the quality of service as well as efficiency of the system. Related to this, Natsheh and Buragg (2010) studied a queue model with batch Poisson arrivals and two heterogeneous service providers which are exponentially distributed. In this study, while the faster server is in ready state at all times, the slower server is not. Rather it is only switched on when the queue length reaches a given threshold. The authors suggested a model named scheduling policy which applied fuzzy logic reasoning in order to attain the gains of heterogeneity in service rates. Nourbakhsk and Tuner (2022) modeled a mathematical variant of the general routing assignment in which the authors were able to construct dynamic policies which outperformed the fastest-server-first as well as its extensions such as the fastest-server-first static block and the fastest-server-first dynamic block. Lidiya and Julia (2024) analyzed the efficiency of a one-server queue involving heterogeneous customers with many types of service breakdowns and working vacations. In this study, arrivals follow a Poisson process with rates that vary based on network features. While the servers are both in busy and working vacation states, they provide services using exponential distributions. At peak times, it was possible for the server to breakdown as a result of its unavailability.

Sani and Daman (2015) considered an M/G/2 queue system with exponential service system and a general server using a supervised queue discipline. The model represents a system in which servers are assigned to customers. The first-come first-served discipline was violated in order to attain minimal waiting time of arrivals in the network. With the second server, the authors derived a distribution of the steady state for the customers in the queue network. The authors were able to formulate the closed form expressions of the: mean waiting time, mean queue length and blocking probability of the system. Mean performance metrics were numerically computed. Jali et al. (2024) considered efficiency in job routing involving a common queue to heterogeneous service providers. A threshold system involving routing jobs to slow server(s) when certain threshold regarding queue length is attained is assumed to be optimal via the one-fast-one-slow dual-server network. This is unlike what is obtainable in a homogenous service system. However, in this study, an optimal system involving multiple server system is not known. Efrosinin and Stepanova (2021) studied an optimal routing problem involving a dual-server heterogeneous system being operated in parallel to ensure optimal assignment of customers to servers in each queue. An algorithm to derive optimal routing system using Markov-modulated Poisson approach was formulated. The researchers considered an optimal Bernoulli splitting approach which defined the optimal assignment probabilities. It was shown that the optimal assignment policy between servers in each queue had a threshold form based on the length of queue as well as arrival pattern. Legros (2017) considered a dual-server queue model involving the formulation of a job routing problem. This is aimed at managing the stationary sum of expected time elapsed in the system as well as the volume of unsatisfied arrivals. With a Markov decision approach, the author was able to declare that the optimal routing system of customers to servers involved a threshold which is determined by the length of the queue.

Efosinin et al. (2023) simulated an annealing algorithm to optimize weights and biases of a multiple layer neural network which was initially trained using arbitrary heuristic control system. This was meant to minimize average cost. The optimal solutions and scheduling policies were derived using a Markov decision problem. Numerical results

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

indicated that the efficiency of the proposed approach to finding optimal deterministic control system for routing, scheduling or resource assignments in general queue systems was established. A queue network which was activated at given time when arrivals are of heterogeneous classes was studied by Bandyopadhyay (2023). In this study, each class of arrival had its routes, the costs of which are linear functions of waiting and service completion times. The author restricted the study to a dual-class, dual-queue network which is a function of problem parameters that are substantial. This suggests a combinatorial increase as the number of queues, customer classes and routes increase. Mahanta et al. (2024) performed a network traffic prediction which was based on a flexible exponential smoothing system that optimized the coefficient of the model. This was achieved by using a differential evolution algorithm which was a cubic function that was based on traffic prediction. It was formulated to improve the existing adaptive random early detection queue control algorithm that uses the cubic function to do nonlinear processing on its packet dropping probabilistic function. Experimental results indicated that the prediction model using triple exponential system achieved a high prediction level as well as a reduction in packets' losses and improvement in throughput.

Materials and Methods

This section is discussed under the following subsections: problem definition and model, proposed fuzzy logic model and benchmark.

Problem definition and model

A simplified routing scenario is one involving a buffer of unlimited size which accommodates one stream of arrivals while feeding two parallel servers. This system is depicted in figure 1.

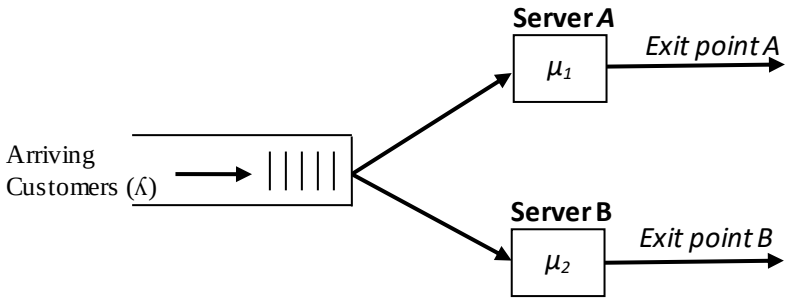

Figure 1: A queue network with two heterogeneous servers arranged in parallel

This study is geared at optimal routing of customers in queueing systems with heterogeneous servers and functions. The scenario considered involves the study of systems in which servers have a common queue with heterogeneous service rates and functions. The objective is to dynamically assign customers to servers which are idle on the basis of the state of the system in order to minimize the cost of customer delays using fuzzy logic approach. The Mamdani implication is adopted in representing the "if-then" rules while height defuzzification is used in the transformation of fuzzy outputs to control systems. In the Mamdani implication, the "if-then" rule is represented by relevant membership functions. In fuzzy control system, the implication is defined by:

$$\mu_C(x,\ y) = \min\left[\mu_A(x), \mu_B(y)\right]$$

for the rule

"*if X is A, then Y is B*".

Defuzzification is a method for changing fuzzy control decisions to non-fuzzy ones. This is necessary since only crisp values of controls are applicable in practice. In addition, the defuzzifier also de-normalizes given control parameters if normalized values are applied. The height defuzzification method was adopted for the transformation of fuzzy outputs to control systems. By convention, the nucleus of a fuzzy set *A* comprises of values *x* for which $\mu_A(x) = 1$. The choice of the height method is based on the fact that it takes account of *k* outputs μ_i which has a height f_i with given peak value being e_i. Consequently:

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

$$\mu_c = \frac{\sum_{i=1}^{k} e_i f_i}{\sum_{i=1}^{k} f_i}$$

The Markov decision system is a continuous time function. It is shown in Viniotis and Ephremides (1988) that a continuous time function is more optimal in situation where it is possible to use a faster server for service while only activating the low performance server when the population of customers awaiting service turn is greater than the critical threshold *n*. This process is of the threshold form and often referred to as the t_n policy. According to Lin and Kumar (1984), the function to derive this optimal threshold value is:

$$\rho = \frac{\lambda}{(\mu_1 + \mu_2)}$$

Consequently:

$$a = (\lambda + \mu_1 + \mu_2)$$

while

$$b = \frac{\lambda}{\mu_1}$$

Simplifying further, we have the following:

$$b_1 = \frac{a - \sqrt{a^2 - 4\mu_1\lambda}}{2\mu_1}$$

$$b_2 = \frac{a + \sqrt{a^2 - 4\mu_1\lambda}}{2\mu_1}$$

$$c_1 = \frac{1 - b_1}{b_2 - b_1}$$

$$c_2 = \frac{b_2 - 1}{b_2 - b_1}$$

In this case, the network state could be described as: y, z_1, z_2 where

$y = 0, 1, ..., \alpha$ (This defines the populations of customers in the network).

$z_1 = 0, 1$ (This defines the state of the server whether idle or busy).

Consequently, $y + z_1 + z_2$ describes the amount of customers in the queue network at this time

Since there is always a change in system's state for every arrival and / or completion of service, the service offered by the system is non-preemptive. This implies that a server that is busy will not be able to accept a new arrival until service is completed. The decision epochs can be defined as the time an arrival is in the queue while there is at least one idle server. This is the situation when an arrival finds no queue and consequently idle servers. It could also be that an arrival is departing the system after service completion while there are other arrivals in the queue network. Since service in the queue network is non-preemptive, the state of the network is such that servers are idle as there are no

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

arrivals in the network, i.e. $x = 0$. In this case, customers are not assigned to any server. However when $\mu_1 > \mu_2$, it is optimal for the system to apply the faster server in the service of customers arriving the network. This implies that when $y > 0$ and $z_1 = 0$, a customer is assigned to the fast server. The objective now is to derive an optimal assignment policy whenever the following conditions exist:

1. $y > 0$;
2. $z_1 = 1$; and
3. $z_2 = 0$.

In these cases, the slower service remain idle even while there are customers awaiting service turns in the queue network.

Proposed fuzzy logic model

The proposed model is a Fuzzified Routing Protocol for Enhanced System Performance (FRESP). In this model, the size of customers and respective arrival rates in the queue network are considered while deciding on whether or not to assign customer to sever B. Two scenarios are considered in this case:

Scenario one: There are no arrivals

In a case, there are no arrivals i.e. $\lambda = 0$, the choice of using the slower server is based on the sum of customers in the queue network. In this case, it is essential to determine the least buffer level i.e. y_0, when it is cost effective to allocate one customer to the slower server i.e. server B. If at time $t = 0$, the number of customers in the system is $y+1$, while server B is idle at this time, then $J(y+z)$ is the incurred cost until $y+1$ customers are assigned to server B while the remaining customers are allocated to server A. In this case, y_0 denotes the least buffer level at which the optimality of $z = 1$ is attained, i.e.:

$$J(y_0, 1) \leq J(y_0, 0) \qquad (1)$$

Consequently, $J(y, z)$ depicts the total mean of travel times of z customer which are assigned to server B while $y - z + 1$ customers are assigned to server A. In this case, the initial customer to be assigned to server *A* will stay in the queue network for an average time of $\frac{1}{\mu_i}$. The second, third, fourth customers will be $\frac{2}{\mu_i}$, $\frac{3}{\mu_i}$, $\frac{4}{\mu_i}$, etc respectively. This is given in the following expressions:

$$J(y, \mathrm{z}) = \frac{1}{\mu_1} + \ldots + \frac{y - \mathrm{z} + 1}{\mu_1} + \frac{1}{\mu_2} + \ldots + \frac{z}{\mu_2} \qquad (2)$$

Expanding this further result in:

$$J(y, \mathrm{z}) = \frac{(\mathrm{y} - \mathrm{z} + 1)(\mathrm{y} - \mathrm{z} + 2)}{2\mu_{\mathrm{i}}} + \frac{y(y+1)}{2\mu_2} \qquad (3)$$

Substituting (1) in (3) results in:

$$\frac{\mu_1}{\mu_2} - 1 \leq y$$

Therefore

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

$$y \geq y_0 = \left\lceil \frac{\mu_1}{\mu_2} - 1 \right\rceil$$

Scenario two: There are arrivals

In this case, $\lambda \geq 1$. At this point, it is optimal to begin service in the system using server B. The threshold optimality y_λ is bounded by y_0. The rule base and membership functions are given in order to activate server B when either y or λ or both are large. Fuzzy inputs are customers awaiting service turns i.e. $y = 0, 1, ..., \alpha$ while the mean inter-arrival is $\lambda \in [0, \mu_1 + \mu_2)$. Decision (*dec.*) is the fuzzy output i.e. $dec. = 1, 0$, which gives the basis of assigning customer to server B. The discourse universes for y, λ and *dec.* are $[0, \alpha), [0, 6)$ and $[0, 1]$ respectively. The following notations are used correspondingly: zero (NL), positive (PO), very positive (VP), highly positive (HP) and extremely positive (EP). In this case, the least positive value is NL, followed by PS in this order while EP is the highest positive value. Similarly, the decision to 'admit customer' is YES while the decision 'do not admit customer' is NO. These are the corresponding *dec.* outputs '1' and '0' respectively as depicted in table 1.

Table 1

Rule number	λ	y	*dec.*
1	NL	NL	NO
2	NL	PO	NO
3	NL	VP	NO
4	NL	HP	NO
5	NL	EP	YES
6	PO	NL	NO
7	PO	PO	NO
8	PO	VP	NO
9	PO	HP	YES
10	PO	EP	YES
11	VP	NL	NO
12	VP	PO	NO
13	VP	VP	YES
14	VP	HP	YES
15	VP	EP	YES
16	HP	NL	NO
17	HP	PO	YES
18	HP	VP	YES
19	HP	HP	YES
20	HP	EP	YES

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

Table 1: Output (dec) for input λ and y

The corresponding membership function for fuzzy input y is given in figure 2.

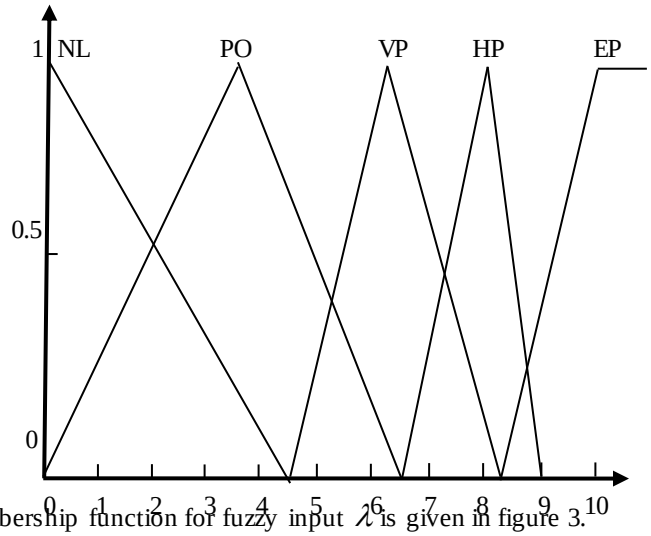

Figure 2: Memberships function for y

The corresponding membership function for fuzzy input λ is given in figure 3.

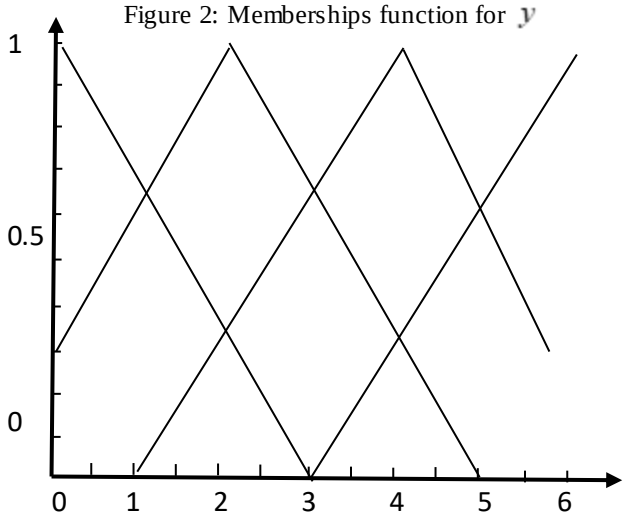

Figure 3: Memberships function for λ

The corresponding membership function for fuzzy output *dec.* is given in figure 4.

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

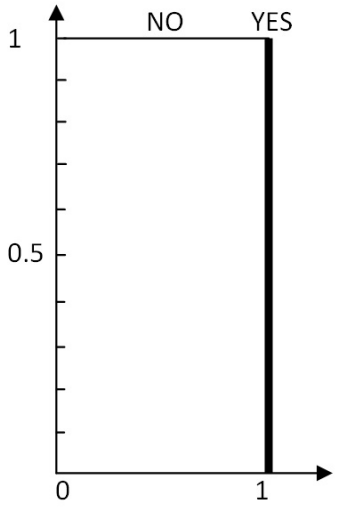

Figure 4: Memberships function for *dec*.

Benchmark

The benchmark adopted is the model proposed by Hyytia et al. (2017). This model is the Elementary Job Routing Problem with heterogeneous servers and general cost structures (EJRP). The cost function in the model is defined as follows:

> *Let* a_{ij} *represent the probabilistic function that an arriving customer is routed to server A with state* (i, j). *Hence for the probabilistic function* $1-a_{ij}$, *the arrival is routed to server B.*

Hyytia et al. (2017) defined the equivalent cost in state (i, j) as follows:

$$r_{ij} = \lambda\left(a_{ij}1(i>2)+(1-a_{ij})1(j>2)\right) \quad (5)$$

The condition in (5) is true for system A as well as modified system D while $\bar{r}_i = \lambda$, given $n, m > 2$. Regarding system D, it is:

$$q_{n^*},(n-1,m) = \mu_1(1-\rho),$$

$$q_{n^*},(n,m-1) = \mu_2(1-\rho),$$

$$r_{n^*} = \lambda$$

Similarly, let $m = (m_1,...,m_k)$ be the dimensions of finite state space while m_k denotes the highest number of customers in server k. In this case, the state function is given as:

$$M = \prod_{k=1}^{K}(m_k+1)$$

The arbitrary state function is described using $n = (n_1,...,n_K)$ while n_k denotes the number of customers in server k. Therefore it is possible to map the K-dimensional function state space in a dimension using:

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

$$s(n) = \sum_{i=1}^{K} \left(n_i \prod_{k=1}^{i-1} (m_k + 1) \right) \quad (7)$$

In (7), the arbitrary probabilistic routing function is given using an MxK matrix α while α_{ik} is the fraction of customers routed to server k in state i.

Simulation

The topology comprises of a system with controller's software and switches which are open-flow configured as well as finite servers and terminals. NED language was used to define the topology of the network. The software was used on a server with core i7 and 64 gigabytes of random access memory. OMNeT++ was applied to simulate processes. A packet generator (PktGen) was used to generate service requests, i.e. customers. The PktGen had been applied in similar studies by Chen et al. (1994) as well as Thakur et al. (2021). The descriptions of customers' properties were done using algorithmic processes. Corresponding modules of customers' properties were created, defined and implemented in the OMNet++ project. The in-built statistical as well as visualization resources in OMNet++ were applied in analyzing metrics.

Results

The performance metrics that were considered in evaluation of FRESP and EJRP include: network throughput, queue length, degree of customers' losses and memory usage.

a. Network throughput

Throughput refers to the output of the queue network. A comparison of the performance of FRESP and EJRP regarding the throughput of the system was made using varying number of customers as shown in figure 5.

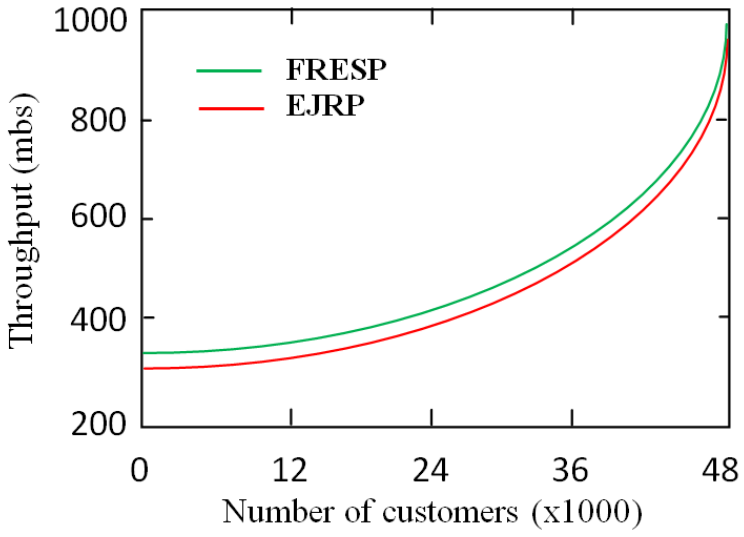

Figure 5: Comparison of network throughput

It is obvious from figure 5 that with more arrivals i.e. customers, system throughput increased linearly. In figure 5, with 12k customers in the network, the corresponding system throughput are 342mbs and 325mbs for FRESP and EJRP respectively. Similarly, with 24k, 36k and 48k customers, the corresponding system throughput for FRESP are 412mbs, 584mbs and 972mbs while EJRP has 387mbs, 557mbs and 961mbs respectively. The percentage analysis of

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

system throughput are 52.3% and 47.8% for FRESP and EJRP respectively. This indicates a comparative performance of FRESP over EJRP.

b. Queue size

A comparison of the performance of FRESP over EJRP for queue size was made. Experimental results are indicated in figure 6. From this, it is observed that as more customers get into the system, the queue length increases linearly.

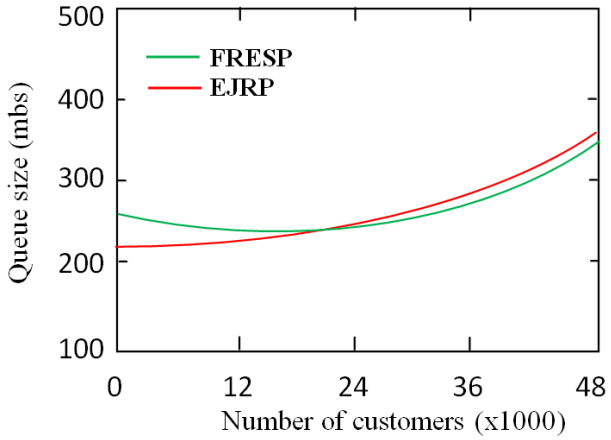

Figure 6: Comparison of queue size

From figure 6, with 12K customers in the system, corresponding size length of FRESP and EJRP are 248mbs and 235mbs respectively. This indicates a fair comparative performance of EJRP over FRESP. However, as more customers arrive the system, the performance of FRESP begins to be better than EJRP. For instance, with 24k, 36k and 48k customers in the system, the queue length for FRESP are 252mbs, 287mbs and 348mbs while that of EJRP are 255mbs, 291mbs and 371mbs respectively. This indicates that FRESP has an advantage of more optimal performance in queue length management that EJRP which tends to be more optimal in queue length management with lesser customers.

c. Degree of customers' losses

The comparison of degree of customers' losses to droppings upon congestion was made. Experimental results are presented in figure 7.

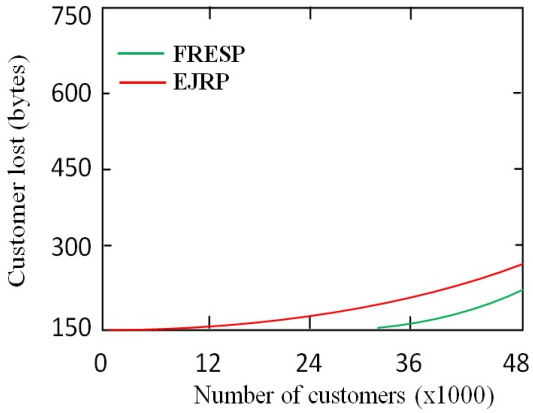

Figure 7: Comparison of de

From figure 7, with 12k and 24k customers, the corresponding numbers of customers' losses to droppings as a result of network congestion are 163 bytes and 197 bytes respectively for EJRP while none exists for FRESP. Similarly, with 36k and 48k customers, the corresponding number of customers' losses to droppings as a result of network

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

congestion are 161 bytes and 223 bytes for FRESP while EJRP has 227 bytes and 261 bytes respectively. These results show a significant difference in performance of both models. Consequently, the percentage of customers' losses to droppings are 94.7% and 5.3% and for EJRP and FRESP respectively.

d. Degree of memory usage

A comparison of the extent of memory consumed regarding both models is made as presented in figure 8.

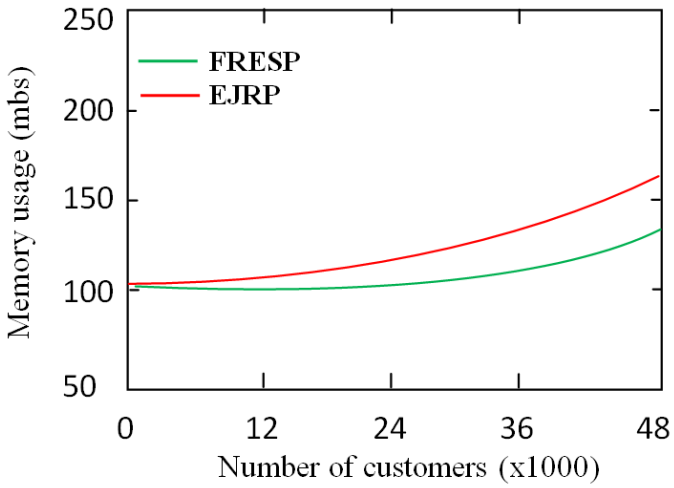

Figure 8: Comparison of memory usage

The amount of memory consumed increases as more customers arrive the buffer. From figure 8, with 12k customers, the amount of memory consumed are 102mbs and 111mbs for FRESP and EJRP respectively. Similarly, with 24k, 36k and 48k customers, the amount of memory used are 104mbs, 109mbs and 136mbs for FRESP while EJRP consumed 119mbs, 145mbs and 163mbs respectively. These results indicate that FRESP consumed lesser memory compared to EJRP.

Conclusion

Simulation results from the study indicated that FRESP has higher throughput than EJRP. On the other hand, regarding queue size, FRESP performed fairly poor with regards to minimal customers than EJRP, thereby giving the later an edge in this regards. This means that the queue length is more with minimal customers than it is for EJRP. However as the number of customers' increases, FRESP tends to perform better than EJRP. The amount of customers' losses as well as memory usage tends to be more minimal for FRESP than EJRP. Consequently, FRESP is more optimal than EJRP.

References

Abdali, N., Heidari, S., Alipour-Vaezi, M., Jolai, F. & Aghsami, A. (2023). A priority queueing - inventory approach for inventory management in multi-channel service retailing using machine learning algorithms. *Kybernetes*. 10(9). https://doi.org/10.1108/K-07-2023-1281.

Abubakar, I. A., Arora, G, Kumar, B. & Danjuma, M. (2022). The optimal number of servers in a many server queueing system. *Journal of Physics: Conference Series*. IOP Publishing. doi:10.1088/1742-6596/2267/1/012105.

Agarwal, S., Upadhyaya, F. & Ahmad, Z. (2022). Optimization of a stochastic model with immediate or delayed repair of servers. *International Journal of Recent Technology and Engineering* (IJRTE). 9(9).1112 - 1126.

Armony, M. (2005). Routing in large-scale service system with heterogeneous servers. *Queueing Systems*. 51. 287-329. https://doi.org/10.1007/s11134-005-3760-7

Cite this article as:
Gbadebo, A. D., Raji-Lawal, H. Y., Abimbola, O. G., & Odupe, T. A. A. (2026). An improved routing protocol in queue networks with heterogeneous service rates. *FNAS Journal of Computing and Applications, 3*(1), 7-18. https://doi.org/10.63561/jca.v3i1.1197

Bandyopadhyay, A. (2023). Game of arrivals at a two-queue network with heterogeneous route choice. *Performance*. Cornell University. https://doi.org/10.48550/arXiv.2310.18149

Bie, Y.L.Z., Hu, Z. & Chen, A. J. (2022) Queue management algorithm for satellite network-based on traffic prediction. *IEEE Access*. 10. 54313 – 54324. https://doi.org/10.1109/ACCESS.2022.3163519

Chen, H., Duenyai, S. & Iravani, S. (2023). Admission and routing control of multiple queues with multiple types of customers. *IEEE/ACM Transactions on Networking (TON)*. 44(3). 998–1011.

Chen, H., Yang, P. & Yao, D. (1994) Control and scheduling in a two-station queueing network: optimal policies and heuristics. *Queueing System Theory and Applications*. 18:301-332. doi: 10.1007/BF01158766

Efrosinin, D., Vishnevsky. V. & Stepanova, N. (2023) Optimal scheduling in general multi-queue system by combining simulation and neural network technologies. *Sensors*. 23(12). doi.org/10.3390/s23125479.

Efrosinin, D. & Stepanova, N. (2021) Optimal open loop routing and threshold-based allocation in two parallel queueing systems with heterogeneous servers. *Mathematics*. 9(21) 2766. Doi.org/10.3390/math9212766.

Haight, F. A. (1958) Two queues in parallel. *Biometrika*. 45. 401–410.

Hyytia, E., Righter, R. & Samuelsson, S. G. (2017). Beyond shortest queue routing with heterogeneous servers and general cost functions. Research Paper on Research Gate. doi: 10.1145/3150928.3150946

Jali, N., Qu, G, Wang, W. & Joshi, G. (2024) Efficient reinforcement learning for routing jobs in heterogeneous queueing Systems. *Performance*. Cornell University. https://doi.org/10.48550/arXiv.2402.01147

Legros, B. & Jouini, O. (2017) Routing in a queueing system with two heterogeneous servers in speed and quality of resolution. *Stochastic Models*. 33(3). 392-410. https://doi.org/10.1080/15326349.2017.1303615

Lidiya, P. & Julia, R.M. (2024) A study on the performance of a queuing system with heterogeneous arrivals and various types of breakdowns under multiple working vacations. *Operations Research and Decisions*: 34(4), 125-140. Doi: 10.37190/ord240408

Lin, W. & Kumar, P. R. (1984). Optimal control of a queueing system with two heterogeneous servers. *IEEE Automatic Control* AC-29:696–703.

Mahanta, S., Kumar, N. & Choudhury, G. (2024) Study of a two types of general heterogeneous service queueing system in a single server with optional repeated service and feedback queue. *Hacettepe Journal of Mathematics and Statistics*. 53(3). 851-878. doi: 10.15672/hujms.1312795

Natsheh, E. & Buragga, K. A. (2010). Optimizing scheduling policy of queueing systems in heterogeous environment using fuzzy reasoning. *International Journal of Computer Science and Network Security*. 10(4) 111-130

Nourbakhsh, V. & Turner, J. (2022) Dynamized routing policies for minimizing expected waiting time in a multi-class multi-sever system. *Computers and Operations Research*. 137. https://doi.org/10.1016/j.cor.2021.105545

Sakalauskas, L., Kaklauskas, L. & Macaitiene, H. (2024). Stalling in queueing systems with heterogeneous channels. *Applied Sciences*.14(2). doi.org/10.3390/app14020773

Sani, S., & Daman, O. A. (2015). The M/G/2 queue with heterogeneous servers under a controlled service discipline: Stationary Performance Analysis. *International Journal of Applied Mathematics*. 45(1). 31-40. https://www.academia.edu/17446225

Thakur, S., Jain, A. & Jain, M. (2021). ANFIS and cost optimization for Markovian queue with operational vacation. *International Journal of Mathematical Engineering and Management Science*. 6(3):894-910. 10.33889/IJMEMS.2021.6.3.053

Viniotis, I. & Ephremides, A. (1988). Extension of the optimality of the threshold policy in heterogeneous multi-server queueing systems. IEEE on Automatic Control 33:104–109