Congestion Analysis of Continuous System through Multiple Server System

ISSN (e) 2520-7393
ISSN (p) 2521-5027
www.estirj.org

Mukesh[1], Wajiha Shah[2], Akash Manwani[3],

[1,2,3]*Department of Electronic Engineering,MUET Jamshoro*

Abstract: Queuing theory is a quantitative technique that entails creating mathematical models of many kinds of queuing systems. The evaluation of a server's performance in the areas of computer communications and networking, transportation, and manufacturing relies heavily on the information provided by queueing models with numerous servers. Congestion is a significant impact in many systems and can lead to reduced performance, slowing or blockage, increased delays, and decreased user satisfaction. The purpose of this research is to overcome the congestion of finite capacity continuous system with variable thresholds of storing capacity through variety of server phase type distribution rates. The System model combines the queuing theory and Markov chains to examined congestion in a continuous system using various system thresholds both internally and externally, as well as varied service rates. The proposed model allows for the calculation of key performance including queue length, waiting time, and service rate. The model is then extended to include multiple servers, and various congestion mitigation strategies are evaluated. The multiple server limited queue length phase type approach is best for customers who want to wait less in both the system and the line. Simulation results show that the proposed model and mitigation strategies are effective in reducing congestion and improving system performance. This study provides valuable insights into the design and optimization of systems with continuous service demands and can be used to guide the development of congestion mitigation strategies in real-world systems.

Keywords: *Queuing Model, Multiple Server, finite Queue Length, Phase Distribution, Markov chain*

1. Introduction

Queuing theory also known as waiting line theory is the mathematical study of behavior of waiting lines, or queues. It deals with the analysis and modeling of the arrival of customers or entities to a system, the way they are served, and the wait times they experience. It is often used to optimize service processes and systems, such as call centers, airports, hospitals, and manufacturing facilities. It helps to predict and manage queues, improve customer satisfaction, reduce wait times, and increase efficiency. [1,2,3]

A queue is formed whenever service requests or customers arrive at a service facility and are forced to wait while the server is busy working on other requests. The basic operations in a queuing system are input or arrival process, output or service process, number of servers or service channels, queue discipline, system Capacity and customer behavior. [4] In this paper the idea of continuous system through multiple server phase type distribution, service rate is explored which can be applied to the different available queuing systems to improve the performance of the system. The idea is applied on sample system of $M/M_k/C$: FCFS/N/f where customers arrival process is according to a markov process at rate λ and the service customers receive is also markov distributed with a mean service time of μ with multiple phases from a multiple server, served on basis of first come first served with limited queue size and finite population. Congestion occurs when the system or server has more customers than its ability to handle which results in balking reneging and delays. The service and flow process analysis of the system provide the detailed information about the customer in a system that which path and how much time it will take to flow in the system.

Service and flow process analysis determines the congestion occurrence in the queuing system in order to switch the service rate to the second probabilistic service rate which is increased service rate. This results in less congestion by handling traffic efficiently, increases system efficiency and ability to provide service at maximum possible level. [5]

2. Methodology

2.1 Queuing System

A queuing system essentially happened when there are people or entities that arrived (arrival), which required a kind of service from another entity (service). The arrival rate, the service rate, the number of servers, the queuing capacity, the number of clients, and the queueing discipline are the characteristics that define a queuing system. [7]

Queuing systems are shorthand by Kendall's notation. Kendall's notation *a / b / c: d / e / f* is used to show entire structure of queuing system where *a* denotes the arrival distribution, *b* represents the service distribution, *c* denotes the number of servers. *d* denotes the queuing discipline FIFO (FCFS), LIFO (LCFS), SIRO (Service in random order), Priority (service in priority bases), *e* represents the system size (queuing capacity) finite or infinite, *f* denotes the population of customers finite or infinite. [10]

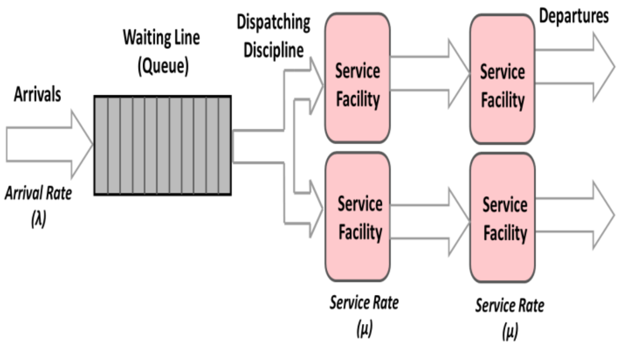


Figure.1. An illustration of a single queue multi server channel & multi phases queuing model

There is first come first serve (FCFS) discipline that a queueing system follows according to the application. Model is consisting of two scenarios as defined bellow.

1. Single Queue + two servers in series with phase type distribution.
2. Single Queue + two parallel service lines with two phase distribution.

Scenario#01

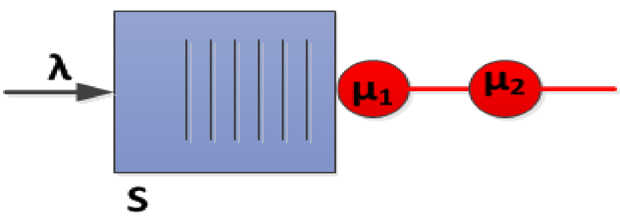


Figure.2. I. Single Queue + two servers in series with phase type distribution

In scenario one the Queuing system model has single λ rate of arrival, and single channel with two phases each has service rate of μ_1, μ_2 respectively and S is the system size.

Scenario#02

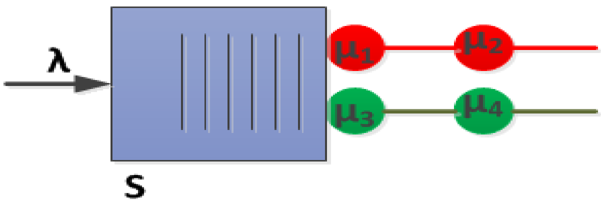


Figure.3. II. Single Queue + two parallel service lines with two phase distribution.

In scenario two the Queuing system model has single λ rate of arrival, and two channel each channel have two phases each channel one has service rate of μ_1, μ_2 and channel 2 have service rate of μ_3, μ_4 respectively and S is the system size.

2.2 Markov Chain

The Markov chain is a stochastic process in which the evolution of a system is thought to be governed only by the current state, without any dependence on the preceding series of events. The transition from one state to another is represented by the chain [8,9]. It actually describes a collection of prospective outcomes, the likelihood of which depends on how the prior event turned out. The state space, which contains all conceivable future states, is one particularly useful tool in this area. Regardless of how the process arrived at the current state, the transition or jumping occurs as a result of the potential future states being fixed.

It also evaluates different scenarios and decides how to manage a "queue" of clients. Internal protocols, service levels, the number of servers, the queueing capacity, the number of clients, and queue discipline are among the features. This enables a more accurate description and analysis of the system's internal behaviours.

2.2.1 Markov chain of scenario#01

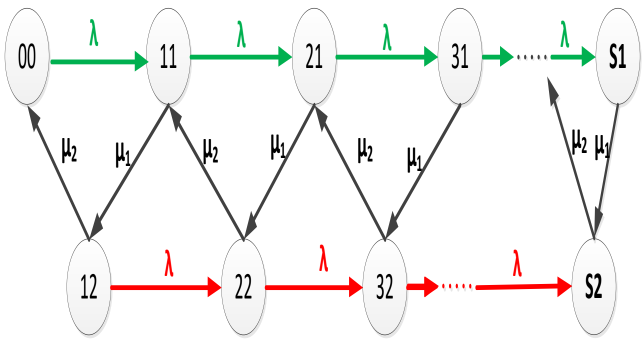


Figure 4. Markov chain of scenario#01

The chain consists of S number of states. State (00) represents no arrival in the system and when one arrival comes it goes to state (11) which is 1st arrival in 1st phase after getting service μ_1 from phase one it jump to phase 2 and get μ_2 service and leave the system and state (21) represent that 2nd arrival in queue getting μ_1 service and goes to next phase for μ_2 service this process continue up to state of (S1) means S number of customer are arriving with rate of arrival λ and getting service of μ_1. As we have observed that state (12) to (22) so on till state (S2), arrival in second phase are also arriving with rate of arrival of λ and getting μ_2 service.

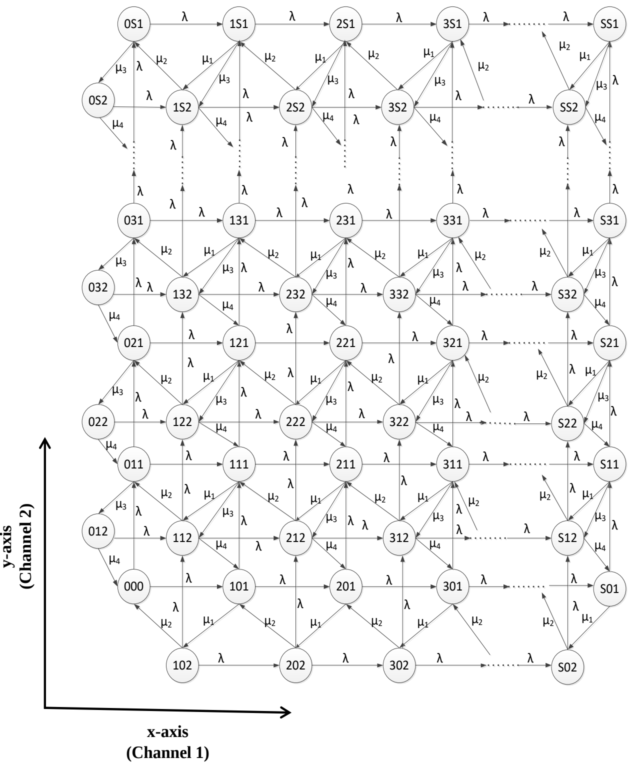


Figure.5. Markov chain of scenario#02

In 2nd scenario λ is the rate of arrival and two parallel line (channels) each line with two phase type services in figure 5 the 1st channel represented by x-axis and arrival is getting service of μ_1,μ_2 and 2nd channel with two phases with services rate μ_3 & μ_4 show by y-axis. The state (000) represent no customer in the system, (101) shows that 1st arrival reach at channel 1 and phase 1 getting μ_1 service and arrive with λ in phase 2 state (102) of channel 1 and get μ_2 service, when 1st arrival is getting services from 1st phase channel 1 and now chances of arrival 2nd arrival. Now 2nd arrival arrive to 2nd channel of 1st phase and behave as a 1st arrival for 2nd channel state shows (011) and get service of μ_3 and arrive with arrival rate λ in 2nd phase state (012) of 2nd channel and get μ_4 service.

2.3 Flow Process and Service Process

The flow process provides a visual depiction of the route taken by each new arrival (consumer). The flow process illustrates all viable customer flow paths, including entrance stages where customers may enter the system and depart after the service has been provided. The whole process flow is analyzed, beginning with the time the test arrival enters the system receiving the service and ending when the test departure departs the system. A customer's possibility of arriving in a state other than the blocked state is calculated by the flow process of the system. In a transition diagram for a flow process, there are three major categories of states.

1) The procedure begins with the customer's entry into the system. We call them "initial states."
2) It is not the current condition that initiates the flow process. They are classified as subsidiary states.
3) The point at which the flow process is completed. They are called "absorbing states" for obvious reasons.

The flow process starts when a client is recognized and admitted into the system. Only three structural circumstances in the overloaded system may allow trial data collecting. The Markov chain of flow process is constructed as shown in Figure 5, which is the combination of both system state process and service process Markov chains. The horizontal axis of Markov chain of the flow process shows the system state process and the vertical axis of the Markov chain of the flow process shows the service process. The inclined states of the Markov chain of the flow process shows the flow process of the arrival, which can enter any of these states at first arriving phase. Once the arrival arrives and accepted by the system, then its flow process begins. There are three main possible entering states of the arrival in the Markov chain of the flow process, where its flow process begins. The state labeled 8 (110), represents starting state when arrival entered in the system and get μ_1 service and go to second phase and get μ_2 service and leave (depart) from system. In this case, arrival gets service immediately and leaves the system. Flow time of the arrival at this state is small because it served immediately. When an arrival arrives at state labeled 14 (211), it finds one arrival already present in the system and it wait until the arrival in front of it is not served. Therefore the flow time at this state for the arrival is comparatively more than the state 18 (312). Similarly, at state labeled 19 (322), flow time of the arrival entering at this state is comparatively more than states labeled 8 and 14 due to the two arrival s already in the system. The states labeled from 1 to 7 are absorbing states, where arrival leaves the system.

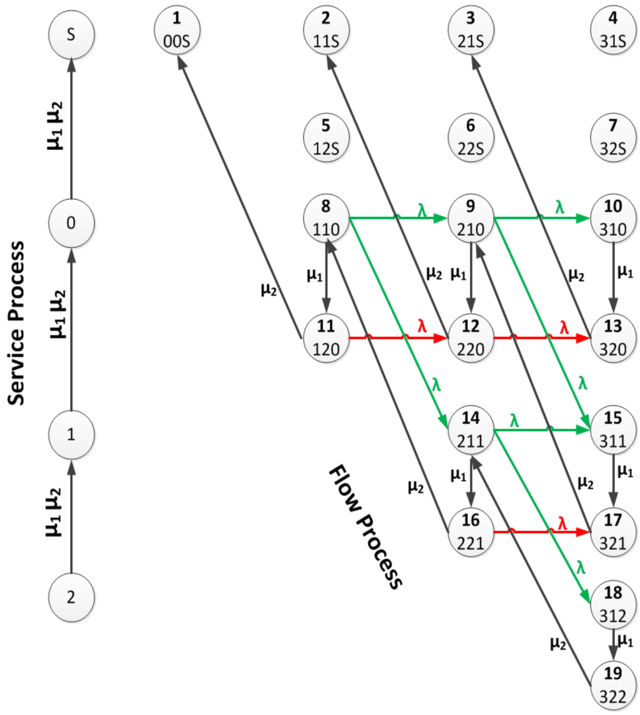


Figure 6. Flow process and service process

2.4 Analytical Equations

At state 8 (110), When arrival arrives at empty System then the starting state probability to begin the Flow process is computed as under

$$\sigma_8 = \left[\frac{\pi_8}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}} + \frac{\pi_{14}}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}}\mu_1\right] \times \left[\frac{1}{1-\frac{\pi_8}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}}}\right] \quad \ldots\ldots\ldots (1)$$

At state 14 (211), When arrival arrives in system and finds already one arrival is in system then the starting state probability of this state to begin the flow process is computed as under

$$\sigma_{14} = \left[\frac{\pi_{14}}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}}\mu_1 + \frac{\pi_{18}}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}}\mu_2\right] \times \left[\frac{1}{1-\frac{\pi_{14}}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}}}\right] \quad \ldots\ldots (2)$$

At state 18 (312), when arrival arrives in system and finds already two arrival s are in system then the starting state probability of this state to begin the flow process is computed as under

$$\sigma_{18} = \left[\frac{\pi_{18}}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}}\mu_1 + \frac{\pi_{19}}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}}\mu_2\right] \times \left[\frac{1}{1-\frac{\pi_{18}}{\pi_8+\pi_{11}+\pi_{14}+\pi_{16}+\pi_{18}+\pi_{19}}}\right] \quad \ldots\ldots\ldots\ldots (3)$$

6. Results and Discussion

Analytical and simulated calculations, such as the probability density function and the flow time, are made possible using MATLAB programs. Flow procedure estimates the probabilities of customer appearances in all states other than the blocked one.

We looked at different situations where the service rate changes based on how many people show up. You can figure out the total probability density function with different arrival rates and the probability density function of starting states.

Scenario#01: When finite queue length with arrival rate of customers is λ=0.4 and service rate of customers are μ_1=0.1, μ_2=0.2.

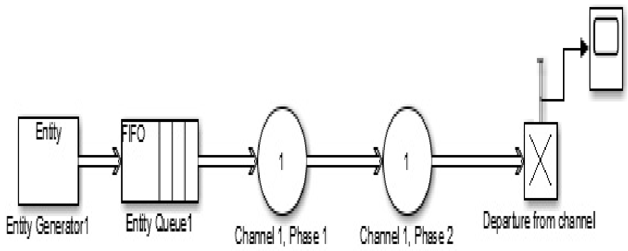


Figure 7. Scenario#1 Simulink Model

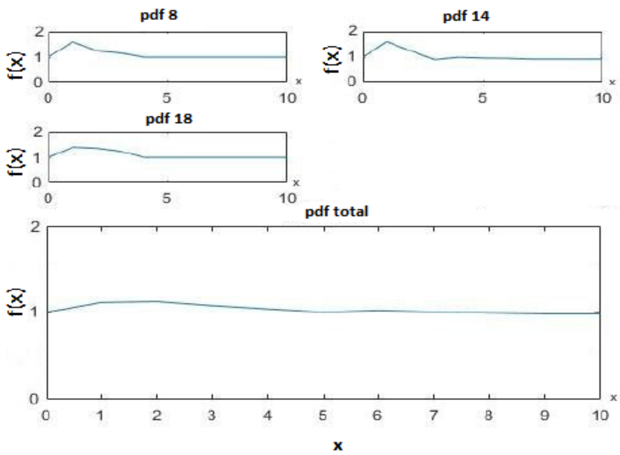


Figure 8. Scenario#1 arrival λ=0.4 so service rate μ_1=0.1, μ_2=0.2.

Probability density function of entering states 8, 14 and 18 are shown in the upper parts of figure while lower parts shows the total probability density function.

Scenario#02: When finite queue length at arrival rate λ=0.8 the congestion occur so system increases the service rate and service rate of customers are μ1=0.1, μ_2=0.2, μ_3=0.1, μ_4=0.2

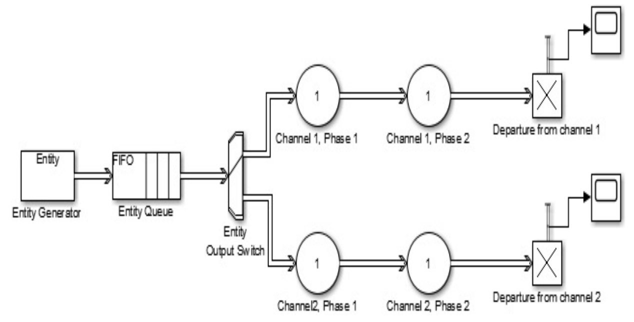


Figure 8. Scenario#2 Simulink Model

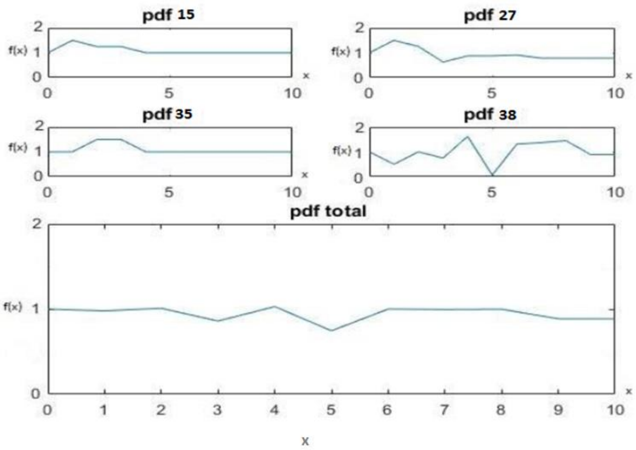
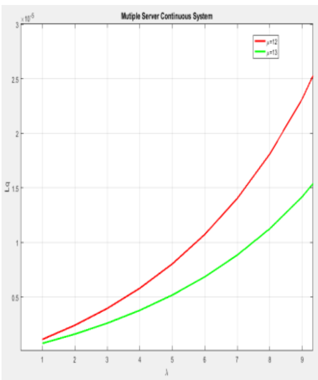


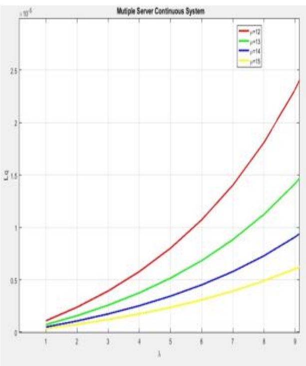
Figure 9. Scenario#2 arrival λ=0.8 and service rate µ1=0.1, µ2 =0.2, µ3=0.1, µ4 =0.2

The figure's top halves display the probability density function for transitioning between states 15, 27, 35, and 38, while the figure's bottom portions display the overall probability density function. The obtained analytical and simulation findings demonstrate that the modeled system offers a more evenly distributed service rate phase type when congestion occurs, hence reducing the severity of the congestion.

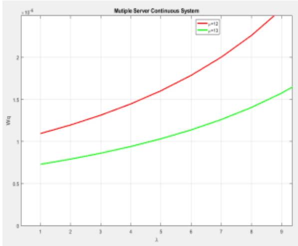
Comparing Results of scenario I & II



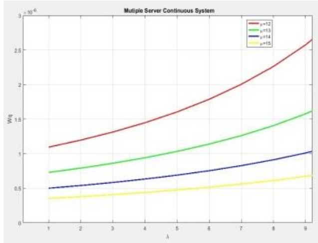
Scenario #01 Graph shows the relationship of rate of arrival (λ) with length of queu (Lq)



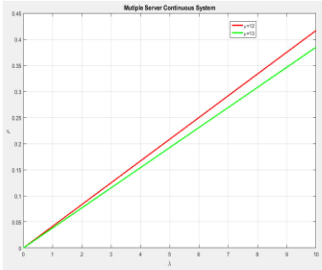
Scenario #02 Graph shows the relationship of rate of arrival (λ) with length of queu (Lq)



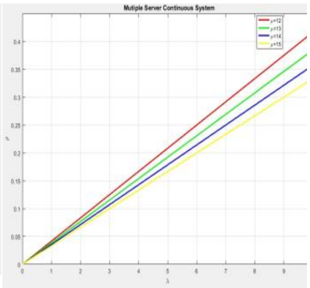
Scenario #01 Graph of rate of arrival (λ) and waiting time in the queu (Wq)



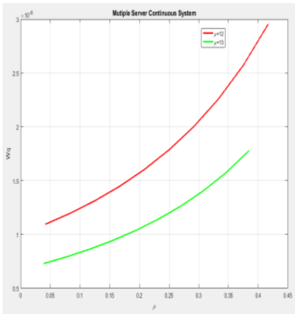
Scenario #02 Graph of rate of arrival (λ) and waiting time in the queu (Wq)



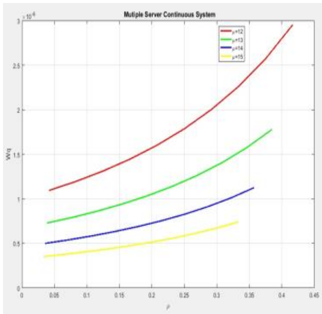
Scenario #01 Graph of rate of arrival (λ) and utilization factor (rho)



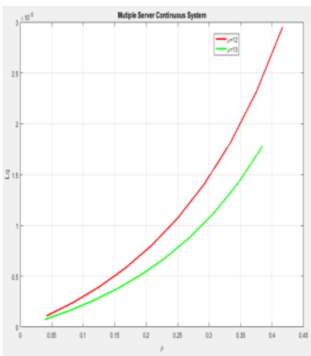
Scenario #02 Graph of rate of arrival (λ) and utilization factor (rho)



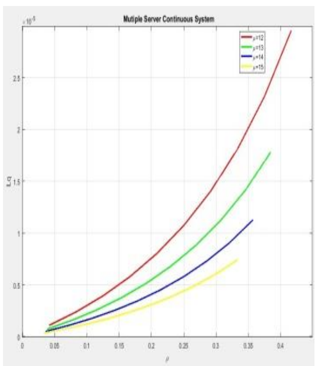
Scenario #01 Graph of utilization factor (rho) and waiting time in the queue (Wq)



Scenario #02 Graph of utilization factor (rho) and waiting time in the queue (Wq)



Scenario #01 Graph of utilization factor (rho) and length of queue (Lq)



Scenario #02 Graph of utilization factor (rho) and length of queue (Lq)

7. Conclusion

In this paper congestion analysis of continuous system through multiple server system is discussed. The finite continuous system with variable thresholds of storing capacity with multiple servers phase type distributions was designed.

Markov chain and Markov flow process constructed of multiple server system to represent the overall flow system at each arrival and at each entry of the phase type system. To assess system congestion through various system thresholds both internally and externally by modifying service rates, the simulation results were used to depict the probability density function of initial states and the total probability density function. Simulation results show that the proposed model and mitigation strategies are effective in reducing congestion and improving system performance. The implementation of a higher service rate led to a reduction in congestion by effectively managing traffic and delivering services at their maximum capacity. The management of system congestion becomes more manageable as the quantity of clients reaches a specific threshold. Based on the findings obtained through analytical and simulation analyses, it is evident that a rise in the service rate, when accompanied by a phase type distribution, leads to a reduction in congestion. This is achieved by effectively managing traffic and ensuring that the maximum level of service is delivered whenever the number of customers reaches a specific threshold.

References

[1] Zeng, G., Chlamtac, I., Fang, Y., & Wu, Y. (2001). *Modeling and Analysis for the Continuous-time Finite Queueing System M/M/1/L/c.*

[2] *Mélange, W., Bruneel, H., Steyaert, B., Claeys, D., & Walraevens, J. (2014). A continuous-time queueing model with class clustering and global FCFS service discipline. Journal of Industrial and Management Optimization, 10(1), 193-206.*

[3] *Nityangini Jhala, Pravin Bhathawala, Analysis and Application of Queuing Theory in Supermarkets, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 9, September 2017*

[4] *Sandhiya, N., & Varadharajan, R. (2018, April). A study on single and multi server queuing models using interval number. In Journal of Physics: Conference Series (Vol. 1000, No. 1, p. 012133). IOP Publishing.*

[5] *Elliriki, M., Reddy, C. S., Anand, K., & Saritha, S. (2021). Multi server queuing system with crashes and alternative repair strategies. Communications in Statistics-Theory and Methods, 51(23), 8173-8185.*

[6] *P.Chithra & K. Srinivasan, An Approach to Minimize Traffic Congestion Using Queueing Theory Techniques, Journal of Algebraic Statistics Volume 13, No. 2, 2022, p. 2263-2266*

[7] *Hurtado-Lange, D., & Maguluri, S. T. (2022). A load balancing system in the many-server heavy-traffic asymptotics. Queueing Systems, 101(3-4), 353-391*

[8] *Choudhary, P., & Dwivedi, R. K. (2022, March). A Real-Time Preliminary Analysis of Traffic Congestion on Indian Roads. In 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 881-884). IEEE.*

[9] *Lange, V., & Daduna, H. (2023). The Weber problem in logistic and services networks under congestion. EURO Journal on Computational Optimization, 100056.*

[10] *Alam, S. T., Azrin, F. A., Uddin, A. S. M. N., & Sagor, E. (2023). Formulation and Performance Assessment for Multiple Server Queueing Models*

About Authors

Mukesh received his B.E degree from Mehran University of Engineering and Technology, Jamshoro Sindh and is enrolled in M.E at Mehran University of Engineering and Technology, Jamshoro Sindh. His research area is performance modeling.

Dr. Wajiha Shah received her B.E and M.E degree from Mehran University of Engineering and Technology, Jamshoro, Pakistan, and her Ph.D. degree from Vienna University of Engineering and Technology, Vienna, Austria.

Akash received his B.E degree from Quaid-E-Awam University Of Engineering, Science & Technology Nawabshah, Sindh, Pakistan and is enrolled in M.E at Mehran University of Engineering and Technology, Jamshoro Sindh. His research area is performance modeling.