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

PERFORMANCE ANALYSIS OF PRIORITY QUEUING SYSTEM WITH DUAL SERVICE USING VECTOR DOMAIN METHOD

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

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

Abstract: The ever-increasing demand for efficient service delivery has prompted the exploration of novel queuing systems to optimize resource utilization and enhance user experience. This study presents a comprehensive performance analysis of a priority queuing system with dual service, employing the innovative Vector Domain Method (VDM) to investigate its effectiveness in diverse operational scenarios. The priority queuing system is designed to accommodate two distinct service classes, each with specific priorities and service requirements. The system is comprised of a solitary server that possesses a finite queue capacity, along with two thresholds denoted as λ1 and λ2. The queuing system is comprised of two distinct forms of service that are correlated with certain queue criteria. When the size of the queue is less than or equal to S1, the system functions at a rate of simple service. Nevertheless, once the size of the queue surpasses S1, the system transitions to a dual service rate to efficiently alleviate congestion inside the system. The queuing system is characterized by a model with a solitary server. The process of generating the Markov chain and analyzing the flow process of the system include the identification and calculation of the transition matrix, as well as the formulation of equations to describe the starting states. VDM's ability to model complex systems with multiple parameters and classes allows for a more thorough assessment of the queuing system's performance under diverse real-world conditions. This dual service approach aims to enhance system efficiency and reduce waiting times for high-priority customers while maintaining satisfactory service levels for standard-priority customers.

Keywords: *Queuing system, Markov Chain,* Vector Domain Method, *Priority system*

1. Introduction

Queuing theory is a mathematical probabilistic modeling tool that is utilized to gain a comprehensive understanding of and devise effective strategies for managing a "queue" of clients in various scenarios. The priority queuing system is a crucial aspect of various applications, such as communication networks, manufacturing systems, and healthcare facilities, where the efficient allocation of resources and service differentiation is paramount. [2, 3] Appropriate delay statistics are raised recently and in the past to provide network traffic limit that is sensitive to delays.[1]

In the system, packages achieved to build a buffer are still distributed i.e. in high-priority or in low-priority class. The proposed model integrates the concept of dual service, where each customer is served by two parallel service channels simultaneously. This dual service approach aims to enhance system efficiency and reduce waiting times for high-priority customers while maintaining satisfactory service levels for standard-priority customers.[7,8]

2. Methodology

2.1 Queuing System

A queuing system consists primarily of one or more servers that provide services to customers as they enter the system. Customers who arrive to discover that all servers in the system are occupied typically form one or more lines in front of the servers, hence the term "queuing systems."[9]

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 FCFS,LCFS, SIRO, Priority (service in priority bases), e represents the system size (queuing capacity) finite or infinite, f denotes the population of customers finite or infinite.[4,10]

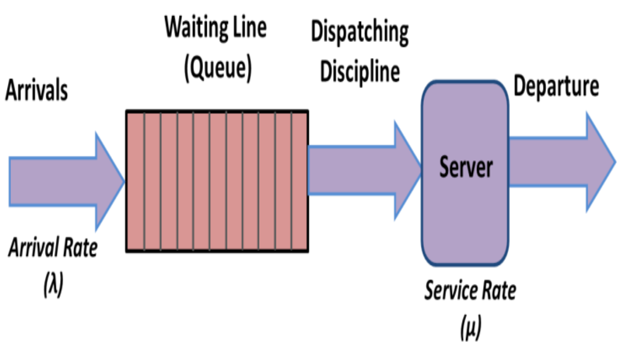

Figure.1 Basic Queuing System

There is four queue service disciplines that a queueing system follows according to the application, here in this system the customers will stick to the priority technique to ensure that all the necessary information should be served first.[5,6]

Corresponding author Email address*: ssss@gmail.com*

2.2 System Model

This research study analyzes queuing system service and flow. Double service rate is used in M/M/1/C/Priority queuing. The system assumes clients come at a rate λ, following a geometric process. This study analyzes service reception using a geometric distribution of service rate (μ). This service is offered by a single first-come, first-served server. Note that the server's queue size, C, is restricted. Double service rate uses two service probability levels to control system load. System internal thresholds define these thresholds.

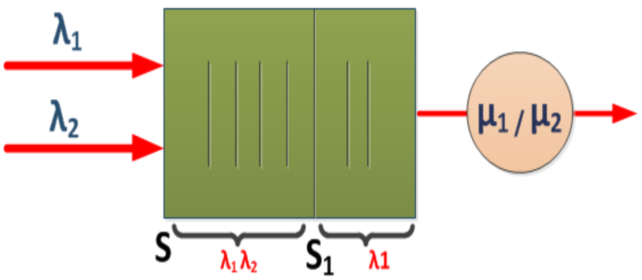

Figure 2. Priority Queuing System with Dual Service Rate

- Finite Queue with two types of data
- Two type of priority arrivals λ_1 & λ_2
- Two Variable service rate μ_1 & μ_2
- System has two thresholds/limits/capacity
  - Internal S_1
  - External S
- Data type λ_1 is priority data
- System provides service to λ_1 up to internal threshold with μ_1 (Normal)
- System provides service to both λ_1 & λ_2 after S_1 and up to S with service μ_2 (Increased)

λ represents the clients' geo-arrivée rate per time unit, whereas μ represents their mean service rate per time unit. Geographically dispersed, In the system, L denotes the mean number of users (server plus queue), whereas Lq displays the mean number of users per queue. Whereas Wq is the average queue delay, W is the average system delay (server + queue). Customers in the line receive standard service rate μ when their care threshold is < λ1, which is the first of two thresholds in the queuing system. Upon reaching the λ1 barrier, consumers in the queue will be sent to the second probability service, which has a service rate of double that of the first probability.

2.3 Markov Chain

It is a graphical representation of overall system behavior each state has incoming and outgoing transitions and each system entry represents the status of the system at that particular point. A mathematical and probabilistic instrument for understanding and determining how to manage a "queue" of customers under various circumstances. The present discourse endeavors to offer a more precise depiction and comprehensive examination of the internal functioning of the system. The concept of the Markov Chain in the context of system flow processes clarify that the forthcoming state of the system is exclusively contingent upon its present state. This implies that it is feasible to anticipate the future state of the system by solely considering its current state, without necessitating access to or analysis of the system's complete historical data.

The chain consists of S states 0, 1, 2, 3…..… S1,…..,S there are S number of customers. The arrival rate of customers is λ and service rate of customers are μ and 2μ.

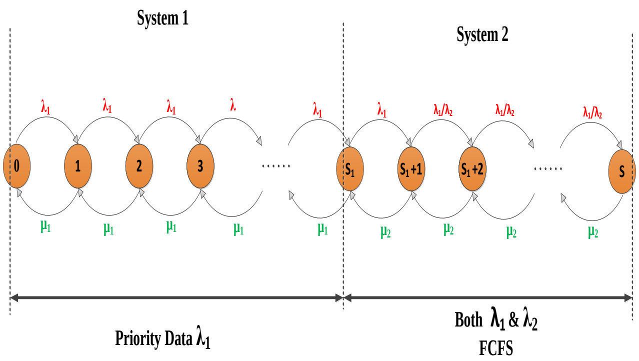

Figure 3. Markov Chain System Model

At state 0, when the queue is vacant and the server is inactive, the customer is routed directly to the server. At state S1 another customer arrives in the system it finds the server occupied serving the first customer the second customer waits in the queue at state 1 total 2 customers in the system1 of which is in the queue and 1 in the server. At state 2, the total number of consumers in the system is three, with two customers standing in line and one on the server. At state 3, there are a total of four clients in the system, with three in the queue and one on the server. At state 4, there are a total of five clients, four of whom are in the queue and one on the server. At state 5, there are a total of six consumers, five of whom are in line and one on the server, and so on. As long as the system is not overloaded, the service rate for states 0 through S1 will be normal. If the system is at state S1+1 and another customer arrives, the system will begin serving at full capacity with a service rate of 2 for the transition from state S1+1 to S.

The preceding system's state transition matrix is computed as

B=

	0	1	2	3	...	S_1	S_1+1	S_1+2	...	S
0	$\lambda_1' * \lambda_2'$	λ_1	0	0	...	0	0	0	...	0
1	μ_1	0	λ_1	0	...	0	0	0	...	0
2	0	μ_1	0	λ_1	...	0	0	0	...	0
3	0	0	μ_1	0	...	0	0	0	...	0
.	.	.	.	.	...	.	.	.	...	.
.	.	.	.	.	...	.	.	.	...	.
.	.	.	.	.	...	.	.	.	...	.
S_1	0	0	0	0	...	0	$(\lambda_1+\lambda_2)$	0	...	0
S_1+1	0	0	0	0	...	μ_2	0	$(\lambda_1+\lambda_2)$	...	0
S_1+2	0	0	0	0	...	0	μ_2	0	...	0
.	.	.	.	.	...				...	
.	.	.	.	.	...	0	0	0	...	0
.	.	.	.	.	...				...	
S	0	0	0	0	...	0	0	0	...	0

2.3.1 MGM (Matrix Geometric Method)

It is a method for the analysis of quasi-birth-death process. The system Markov chain has transition rate matrices. Develop a finite generator matrix. Develop a boundary and repetitive sub matrix. System equation having R matrix.

B=

	System 1						System 2			
	0	1	2	3	...	S_1	S_1+1	S_1+2	...	S
0	$[\lambda_1' * \lambda_2']$	λ_1	0	0	...	0	0	0	...	0
1	μ_1	0	λ_1	0	...	0	0	0	...	0
2	0	μ_1	0	λ_1	...	0	0	0	...	0
3	0	0	μ_1	0	...	0	0	0	...	0
.	.	.	.	.	...	.	.	.	...	.
.	.	.	.	.	...	.	.	.	...	.
.	.	.	.	.	...	.	.	.	...	.
S_1	0	0	0	0	...	0	$(\lambda_1+\lambda_2)$	0	...	0
S_1+1	0	0	0	0	...	μ_2	0	$(\lambda_1+\lambda_2)$	...	0
S_1+2	0	0	0	0	...	0	μ_2	0	...	0
.	.	.	.	.	...				...	
.	.	.	.	.	...	0	0	0	...	0
.	.	.	.	.	...				...	
S	0	0	0	0	...	0	0	0	...	0

MGM Finitesimal Generator Matrix of Over All System

System Decomposition

- System is decomposed into two sub-system.
- System 1 Provide service to the highest priority arrivals (λ_1)
- System 2 system provides service to both arrivals λ_1 and λ_2

System 1

- Arrival are accepted up to S1 as a high priority λ1.
- Only high priority arrivals are served up to S1
- Service provide to these arrival with normal service rate μ1 normal.

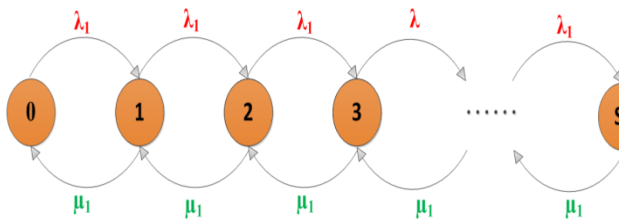

Figure 4: Markov chain of System 1

	0	1	2	3	...	S_1
0	$[\lambda_1' * \lambda_2']$	λ_1	0	0	...	0
1	μ_1	0	λ_1	0	...	0
2	0	μ_1	0	λ_1	...	0
3	0	0	μ_1	0	...	0
.	.	.	.	.	...	.
.	.	.	.	.	...	.
.	.	.	.	.	...	.
S_1	0	0	0	0	...	0

Finitesimal generator Matrix of system 1.

Develop a boundary and repetitive sub matrix.

$$B_{oo} = [\ \lambda_1' * \lambda_2'\ ]$$
$$Ao = [\ \lambda_1\ ]$$
$$A1 = [\ \mu_1\ ]$$

System 2

- Arrivals will be considered as non priority of the threshold S1.
- Arrivals serviced first come first service.
- Service provides to these customer as increased services rate μ2.

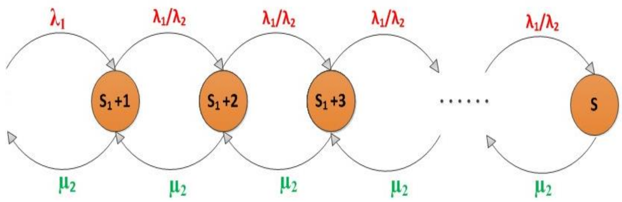

Figure 5: Markov chain of System 2

	S_1+1	S_1+2	S_1+3	...	S
S_1+1	(λ_1)	$(\lambda_1+\lambda_2)$	0	...	0
S_1+2	μ_2	0	$(\lambda_1+\lambda_2)$	...	0
S_1+3	0	μ_2	0	...	0
.				...	
.	0	0	0	...	0
.				...	
S	0	0	0	...	0

Finitesimal generator Matrix of system 2

Analytical Equation

- Calculate the rate R matrix

$$(\pi_0 \quad \pi_1)\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & A_1 + RA_0 \end{pmatrix} = (0 \quad 0)$$

- Calculate the steady state probabilities and measures using R matix

$$\pi_i = \pi_1 R^{i-1}$$

6. Results and Discussion

MATLAB programs are made to do calculations for analysis and simulations. For a long time, the system is saved for the first customer, who starts their service. The second customer, on the other hand, takes less time than the first. due to that system time will decrease as each customer arrives.

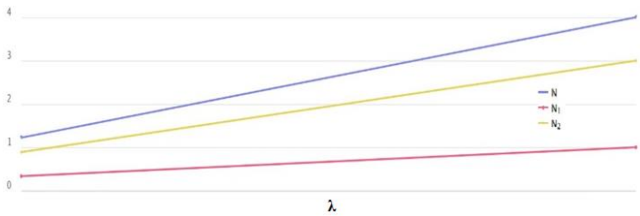

Figure 5: Average number of customers in the system N, Average number of customers of the first priority class in the system N1 Average number of customers of the second priority class in the system N2

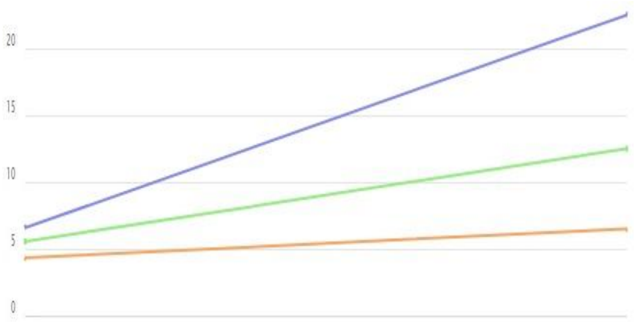

Figure 6:Mean response time T, Mean response time for first priority class customers T1, Mean response time for second priority class customers T2

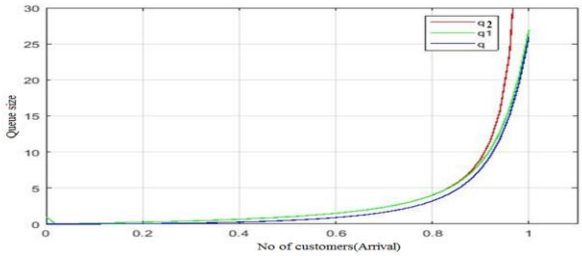

Figure 7: Queue size versus no: of Customer (Arrival)

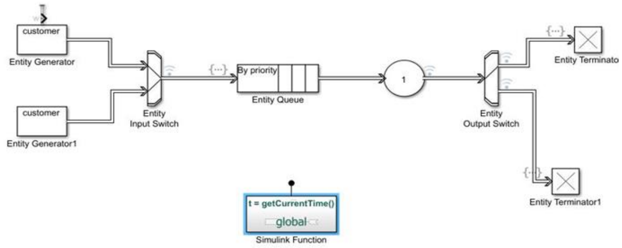

Figure 8: Simulink Model of System

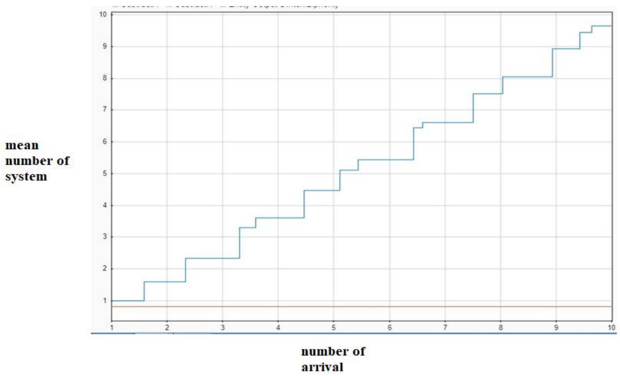

Figure 9: number of arrival versus mean number of system

7. Conclusion

The aim was to investigate the efficiency and effectiveness of the proposed system in allocating resources and providing service differentiation to customers based on their priority levels. The findings obtained from simulations and analytical derivations offer valuable insights into the advantages and capabilities of the dual-service priority queuing system. Our results demonstrate that the integration of dual service in the priority queuing system significantly enhances performance metrics compared to traditional single-service priority systems. High-priority customers experience reduced waiting times and expedited service, thereby improving their overall satisfaction. The proposed model integrates the concept of dual service, where each customer is served by two parallel service channels simultaneously. This dual service approach aims to enhance system efficiency and reduce waiting times for high-priority customers while maintaining satisfactory service levels for standard-priority customers. The vector domain method, a powerful mathematical tool for analyzing complex queuing systems, is employed to evaluate the performance metrics of interest.

The key performance indicators include average waiting time, system throughput, resource utilization, and queue length distribution for both priority classes. Additionally, the impact of varying arrival rates, service rates, and customer priorities on the system's performance is thoroughly investigated.

References

[1] Jin, X., & Min, G. (2007). Performance analysis of priority scheduling mechanisms under heterogeneous network traffic. Journal of Computer and System Sciences, 73(8), 1207-1220.

[2] Kamoun, F. (2012). Performance analysis of two priority queuing systems in tandem. American Journal of Operations Research, 2, 509.

[3] Chuprikov, P., Nikolenko, S. I., Davydow, A., & Kogan, K. (2018). Priority queueing for packets with two characteristics. IEEE/ACM Transactions on Networking, 26(1), 342-355.

[4] Antonelli, D., Litwin, P., & Stadnicka, D. (2018). Multiple System Dynamics and Discrete Event Simulation for manufacturing system performance evaluation. *Procedia CIRP*, *78*, 178-183.

[5] Meddeb, N., Makhlouf, A. M., & Ayed, M. A. B. (2019, June). Priority based safety management and slot reservation for authenticated vehicle. In 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC) (pp. 1977-1982). IEEE.

[6] Mirtchev, S., Goleva, R., Atamian, D., & Ganchev, I. (2018, April). Investigation of priority queue with peaked traffic flows. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (pp. 1017-1019).

[7] Liu, P., Wang, C., Lei, M., Li, M., & Zhao, M. (2020, May). Adaptive priority-threshold setting strategy for statistical priority-based multiple access network. In 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring) (pp. 1-5). IEEE

[8] Horváth, G. (2020). Waiting time and queue length analysis of Markov-modulated fluid priority queues. Queueing Systems, 95(1), 69-95.

[9] Xu, J., & Gautam, N. (2021). Peak age of information in priority queuing systems. IEEE Transactions on Information Theory, 67(1), 373-390.

[10] Zhang, Y., Sun, H., He, Y., Zhang, Z., Wang, X., & Quek, T. Q. (2023). A Spatio-Temporal Analytical Model for Statistical Priority-Based Multiple Access Network. IEEE Wireless Communications Letters, 12(1), 153-157.

About Authors

Akash received his B.E degree from QUEST Engineering and Technology, Nawabshaha 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.

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.