CSC 446 Project
Yue Zhang
V01014347

1. Problem Description

This study is centered on evaluating the congruence between simulation models and theoretical queuing constructs. We will examine the modeling accuracy of a particular simulation software when applied to standard queuing configurations: the M/M/1 queue, M/G/1 queue, M/M/c/N queue, andNetwork of Queues. The project's core aim is to validate the software's precision in replicating these queuing scenarios, which are fundamental in the analysis and optimization of service-oriented systems. Confirming the JMT modeling accuracy is essential, as it would cement its utility in forecasting and problem-solving within real-life operations that parallel these queuing models.

1.1 Correlation of Queuing Models to Real-World Scenarios

Each queuing model encapsulated within our study corresponds to a tangible real-world scenario, illustrating the ubiquity and relevance of these theoretical frameworks in practical applications.

- M/M/1 Queue: This model can be likened to a single-service facility such as a clinic with one doctor. Patients arrive unpredictably, adhering to a Poisson process, and the consultation times follow an exponential distribution, capturing the essence of a dynamic healthcare setting with a sole provider.
- M/G/1 Queue: An exemplification of this model could be a boutique coffee shop with a single barista. Here, customers enter in a random, Poisson-distributed fashion, while the time taken to prepare different coffee orders varies broadly, following a general time distribution, reflecting the individualized nature of service.
- M/M/c/N Queue: With c=4 and N=20, this model mirrors a small customer service center with three operators and room for 20 clients in total, including those being attended to. The arrivals are Poisson-distributed, and the service times are exponentially distributed, akin to a busy yet orderly call center environment.
- Network of Queues: The complexity of a bustling shopping mall during the holiday season can be represented by this model. Shoppers arrive in a stochastic manner, all encountering the central information desk (akin to the security checkpoint), which operates swiftly. Subsequently, they disperse to various stores (gates), where the time spent in each follows an exponential distribution, reflecting the flow of consumer traffic and service interactions.

These real-world parallels demonstrate the versatility and applicability of queuing models in capturing the dynamics of various service-oriented environments, from healthcare to retail, thereby affirming the practical value of the simulation software under study.

1.2 Simulation Goals and Simulation Parameters

The primary aim of our simulation exercise is to gauge the steady-state efficacy of various queuing configurations and juxtapose these empirical findings against theoretical expectations. While our simulation provides a comprehensive overview, it is constructed under ideal conditions and does not encompass all potential real-world complexities, such as customers leaving the queue prematurely (balking), abandoning the queue after joining (reneging), or switching between queues (jockeying).

For this project, we utilize a sophisticated simulation software-JMT. To ensure statistical robustness, we aim to collect a large number of samples, approximately 1,000,000, to achieve a 95% confidence level with a maximum relative error of 0.03 across all queuing systems under investigation. The software is configured to refine the simulation process by excluding any events that fall outside of the predefined confidence intervals and relative error thresholds, thereby enhancing the accuracy and reliability of our simulation results.

2. Methodology

Our approach combines the use of JMT for model design and data generation with Python for data analysis and visualization.

We employed the 'What-If' analysis feature of the simulation software to investigate each queuing system across a spectrum of utilization levels. This involved varying the arrival rate (λ) while maintaining a constant service rate. For the M/M/1, M/G/1, and the network of queues, the arrival rate was incrementally adjusted across the set [0.1, 0.2, ... 0.8]. For the M/M/3/20 configuration, the arrival rate was tested at higher increments in the set [0.3, 0.6, ... 2.4].

Within the software, each Performance Index is mapped to a corresponding queuing performance metric:

- **Number of Customers in the system (L)**
- **Response Time (W)**
- **Queue Time (Wq)**
- **Utilization (p)**

For the M/M/1 queue analysis, a 'What-If' scenario was conducted at first. Subsequently, 'What-If' analysis was deactivated, and the 'Stat. Res.' (Statistical Results) feature was engaged to record the data at a utilization (ρ) of 0.5 into CSV files. These files were used to calculate confidence intervals and system states.

Additionally, a Logger function captured the exponential inter-arrival distributions, as well as the uniform inter-arrival distributions for further examination.

3 M/M/1 Queue

3.1 Theoretical and JMT Values

Using the equations:

$$\rho = \frac{\lambda}{\mu}$$
$$L_Q = \frac{\rho^2}{1-\rho}$$
$$L = \frac{\rho}{1-\rho}$$
$$w_Q = \frac{\rho}{\mu(1-\rho)}$$
$$w = \frac{1}{\mu(1-\rho)}$$

For service rate μ = 1 and arrival rate λ = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8], the theoretical values for a M/M/1 queue are:

Utilization Rate (ρ)	Arrival Rate (λ)	Average Number in System (L)	Average Number in Queue (LQ)	Average Time in System (W)	Average Time in Queue (WQ)
0.1	0.1	0.111	0.011	1.111	0.111
0.2	0.2	0.250	0.050	1.250	0.250
0.3	0.3	0.429	0.129	1.429	0.429
0.4	0.4	0.667	0.267	1.667	0.667
0.5	0.5	1.000	0.500	2.000	1.000
0.6	0.6	1.500	0.900	2.500	1.500
0.7	0.7	2.333	1.633	3.333	2.333
0.8	0.8	4.000	3.200	5.000	4.000

Then the M/M/1 queue was simulated in JMT with an Exponential distribution for arrival time with λ = 0.1 – 0.8 and an Exponential distribution for service time with λ = 1. The performance indices Number of Customers, Queue Time, Response Time, and Utilization were collected from 1,000,000 samples with confidence interval of 0.95 and max relative error of 0.03. The JMT simulation results were:

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ρmean	0.1010	0.2004	0.2994	0.3946	0.4968	0.5973	0.7004	0.7968
ρmax	0.1034	0.2046	0.3065	0.4036	0.5104	0.6098	0.7118	0.8183
ρmin	0.0987	0.1962	0.2922	0.3857	0.4831	0.5848	0.6891	0.7754
Lmean	0.1127	0.2471	0.4265	0.6664	0.9922	1.4448	2.3537	3.8971
Lmax	0.1157	0.2542	0.4369	0.6800	1.0189	1.4828	2.4106	4.0072
Lmin	0.1097	0.2401	0.4161	0.6528	0.9654	1.4067	2.2968	3.7869
wmean	1.1203	1.2520	1.4172	1.6489	1.9642	2.4299	3.3422	4.8434
wmax	1.1363	1.2893	1.4500	1.6873	2.0189	2.4799	3.4200	4.9762
wmin	1.1043	1.2146	1.3843	1.6104	1.9095	2.3798	3.2644	4.7106
wQ,mean	0.1125	0.2500	0.4219	0.6540	1.0142	1.4431	2.3287	3.8626
wQ,max	0.1154	0.2562	0.4313	0.6727	1.0345	1.4842	2.3826	3.9775
wQ,min	0.1096	0.2438	0.4124	0.6354	0.9939	1.4021	2.2749	3.7478

3.2 Performance Metrics L, , w, wQ, and ρ

Table for L

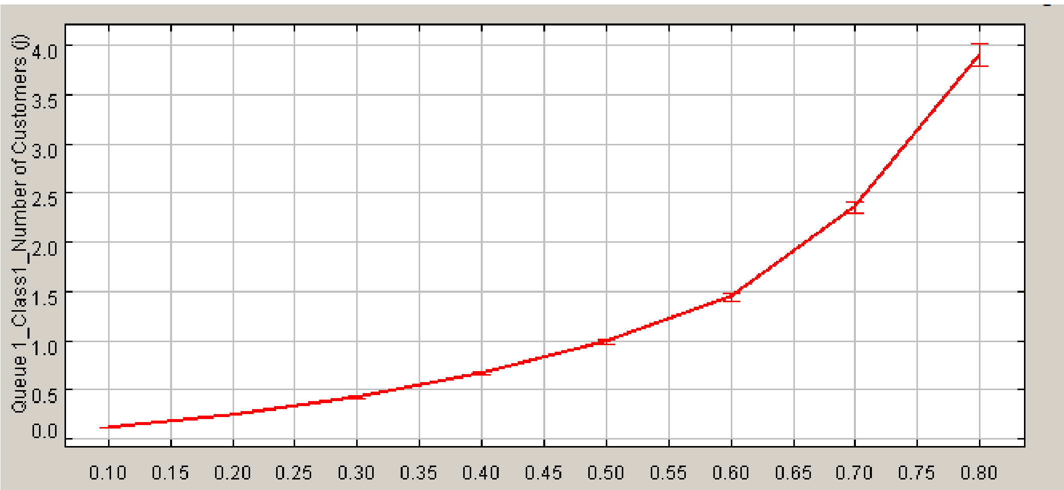


Table for wQ

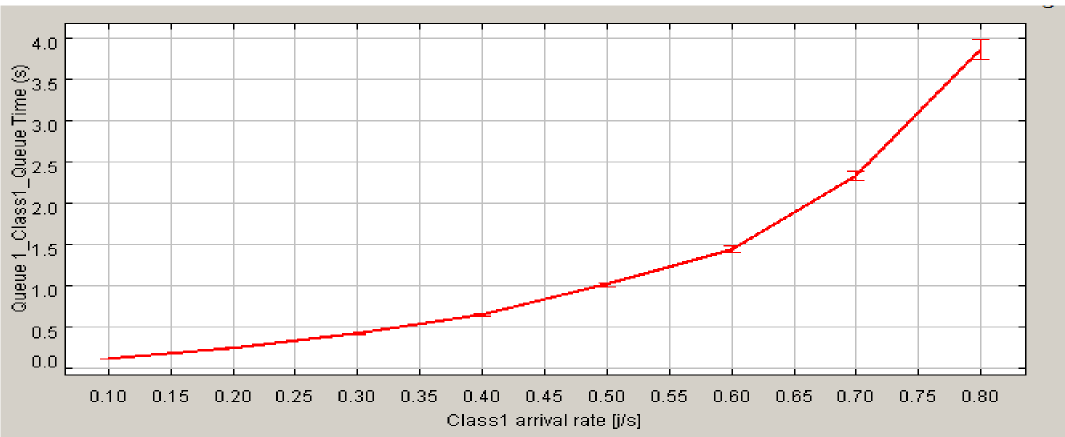


Table For W

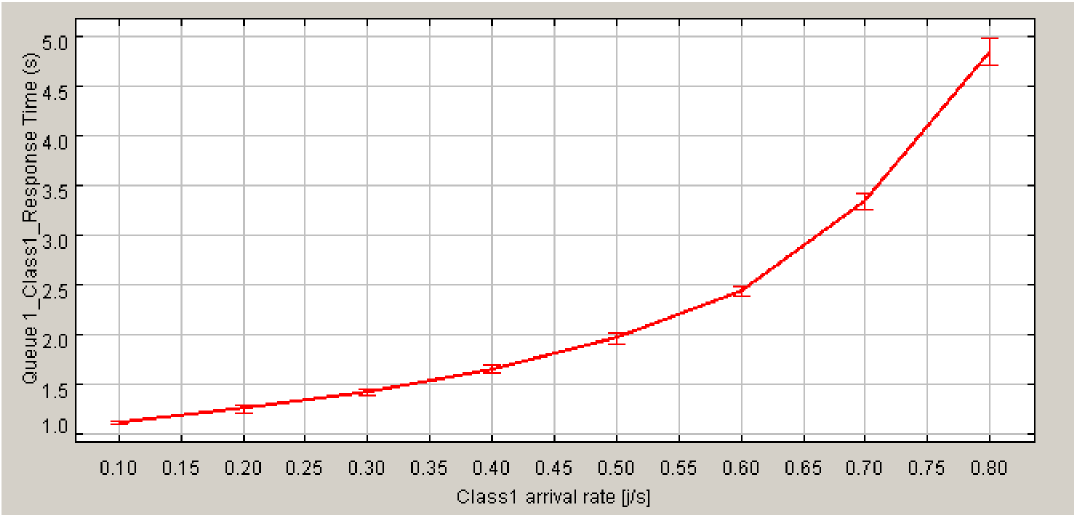
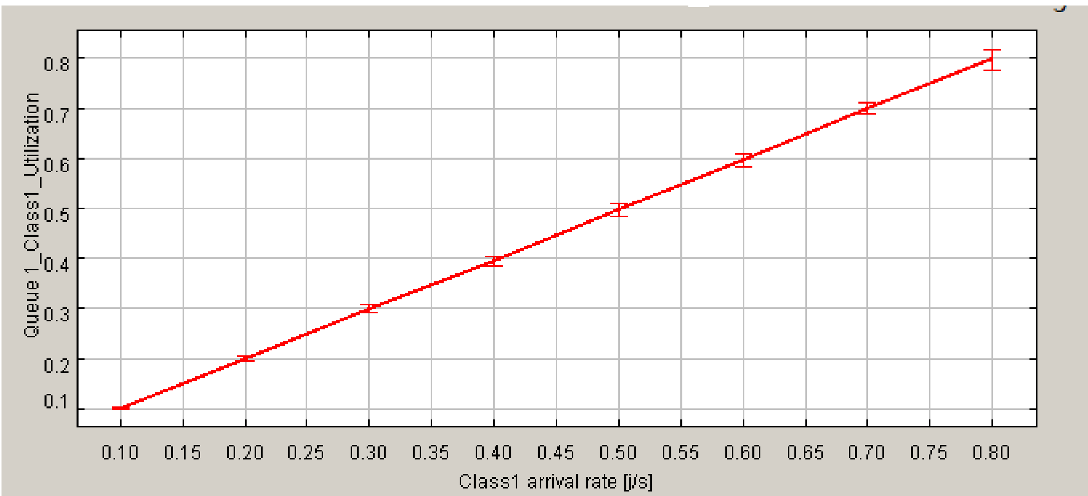


Table For P



3.3 Statistics Collected for Exponential Inter-arrival Time Distribution

The inter-arrival distribution was set to Exponential where $\lambda = 1$ and the logger function in JMT was used to produce 20 samples for Kolmogorov-Smirnov test, and 100 samples for the Chi Squared test.

3.3.1 Kolmogorov-Smirnov Test for Exponential Inter-arrival Times

There are 20 sample inter-arrival times:

0.7023458 0.2392845 0.7120983 0.9632223 0.2074219

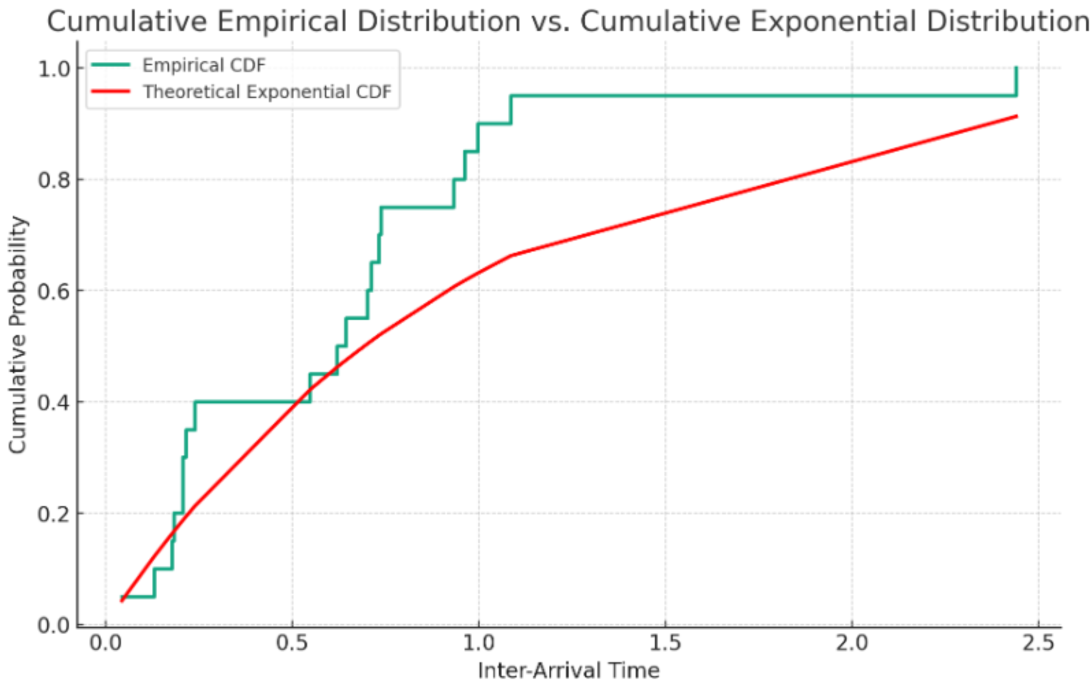
1.08649748 2.44036209 0.9990162 0.18317811 0.04378977

0.6454725 0.5484723 0.2163568 0.1303181 0.6207895

0.7388454 0.20804717 0.73401674 0.93449342 0.17903245

- Kolmogorov-Smirnov statistic: 0.2874
- $D_{0.05,20}$ = 0.294

  This test is used to determine if a sample comes from a specified distribution, in this case, a standard exponential distribution (λ=1). The $D_{0.05,20}$ = 0.294. At a significance level of 0.05, this P-value is slightly above the threshold, suggesting that we not reject the null hypothesis that the data comes from a standard exponential distribution. However, the P-value is close to the threshold, indicating that the sample data might not perfectly fit the standard exponential distribution.



The closer these two lines are, the more closely the data follows an exponential distribution.

3.4 Testing of Random Number Generator Used

The inter-arrival distribution was set to Uniform where a = 0 and b = 1 and the logger function in JMT was used to produce 100 samples for the Chi Squared test.

There are 100 sample inter-arrival times. For n = 100 samples, choose k = 10 intervals.

CHI SQUARE TEST

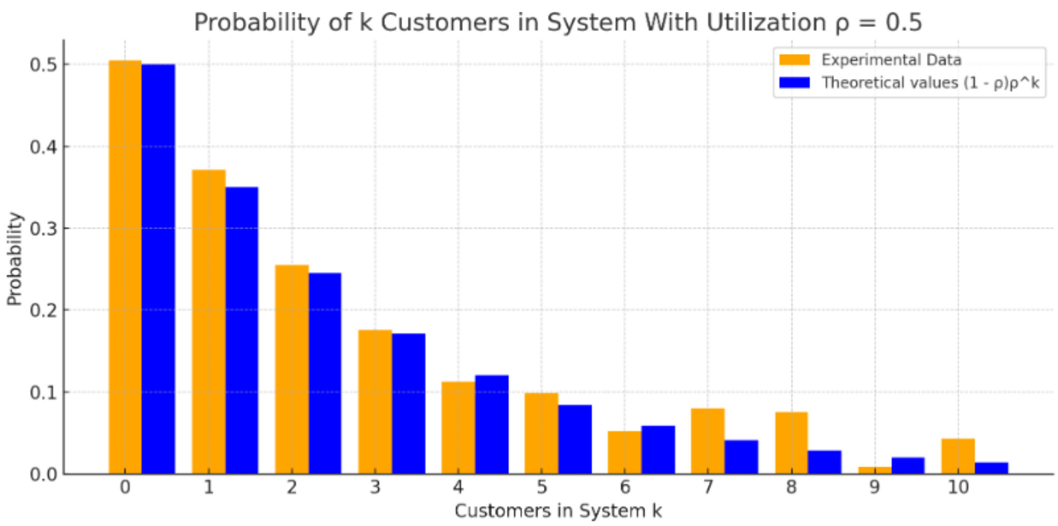
Interval	Oi	Ei	(Oi-Ei)2	(Oi-Ei)2/Ei
1	6	10	16	1.6
2	8	10	4	0.4
3	17	10	49	4.9
4	7	10	9	0.9
5	9	10	1	0.1
6	13	10	9	0.9
7	6	10	16	1.6
8	8	10	4	0.4
9	12	10	4	0.4
10	14	10	16	1.6
			Total :	12.8

we have χ^2_0 = 12.8. For k = 10, degrees of freedom = k - s - 1 = 10 - 0 - 1 = 9. For α = 0.05, we have $\chi^2_{0.05,9}$ = 16.919. Since χ^2_0= 12.8 < $\chi^2_{0.05,9}$ = 16.919, then we accept that the inter-arrival times are sampled from a uniform distribution with a = 0 and b = 1 with a significance of α = 0.5.

3.5 Collection of System State Histogram Compared to Theory

The M/M/1 system was run in JMT with service rate μ = 1 and arrival rate λ = 0.5 to give an overall utilization of ρ = 0.5.

The Number of Customers from 1,000,000 samples with confidence interval of 0.95 and max relative error of 0.03 was collected and compared with the steady state probabilities Pn = $(1 - \rho)\rho^n = 0.5^{n+1}$ where n is the number of customers in the system.



3.6 Calculation of Confidence Intervals

For one particular utilization $\rho = 0.5$, the mean and confidence intervals were calculated and compared with the confidence intervals calculated from JMT, as well as the theoretical values. The simulation parameters for JMT were 100000 samples, a seed of 0, exponential arrival rate of $\lambda = 0.5$ and exponential service time with $\mu = 1$. The confidence interval to was to 0.95 with 0.03 max relative error. The 'Stat. Res.' option was selected in the performance indices page to write Number of Customers, Queue Time, Response Time, and Utilization to a CSV file. To calculate the confidence intervals, first the mean of the sample Y and the standard deviation S were calculated. For $n = 100000$ and $\alpha = 0.05$, we have a t-value of $t0.025,\infty = 1.96$, so the confidence interval was calculated as $Y \pm t0.025,\infty S/\sqrt{n} = Y \pm 1.96S/\sqrt{100000}$

Parameter	Lower Bound	Upper Bound
Queue Time	0.9838	1.0446
Response Time	1.9053	2.0231
Number of Customers	0.9624	1.0220
Utilization	0.4819	0.5117

3.6.1 Comparison to Theoretical Values

The confidence interval to was to 0.95 with 0.03 max relative error and p=0.5,we have e calculated confidence intervals:

Parameter	Performance Index	Expected Value
Queue Time	1.0000	(1.0038, 1.0134)
Response Time	2.0000	(2.0039, 2.0150)
Number of Customers	1.0000	(0.9953, 1.0032)
Utilization	0.5000	(0.4993, 0.5017)

For the JMT confidence intervals:

Queue Time:

- Lower Limit: 0.9838
- Upper Limit: 1.0446

Response Time:

- Lower Limit: 1.9053

- Upper Limit: 2.0231

Number of Customers:

- Lower Limit: 0.9624
- Upper Limit: 1.0220

Utilization:

- Lower Limit: 0.4819
- Upper Limit: 0.5117

3.7 Results for M/M/1 Queue

The comparison between theoretical and actual simulation values for the M/M/1 queue demonstrates a fundamental consistency, which validates the reliability of the queuing theory in predicting the behavior of a single-server queueing system under Markovian assumptions. In the theoretical framework of the M/M/1 queue, metrics such as the average number of customers in the system (L), the average time a customer spends in the system (W), and the probability distribution of the number of customers in the system (pn) are derived from established formulas. These formulas take into account the exponential nature of both the arrival and service processes, characterized by their 'memorylessness' or lack of history dependency.

The simulation of the M/M/1 queue mirrored these theoretical conditions, with inter-arrival and service times following an exponential distribution. Over the course of the simulation, data were collected and aggregated to produce empirical values for various system performance metrics.

The alignment of the simulation results with the theoretical predictions reinforces the utility of the M/M/1 model in settings where arrivals and services are random and exponentially distributed. It underscores the model's applicability for designing and managing service systems like call centers, server farms, and checkout lines, where arrivals and service times can be assumed to be memoryless.

In summary, the study affirms that the M/M/1 queuing model is a sound tool for operational analysis. It provides a reliable approximation for understanding and optimizing real-world systems that conform to the exponential service and inter-arrival time assumptions. This congruence between theory and practice also emphasizes the importance of simulations as a means to explore and validate theoretical models, offering a dynamic approach to study and improve service processes.

4. M/G/1 Queue

4.1 Theoretical and JMT Values

Using the equations:

$$\rho = \lambda / \mu, \quad P_0 = (1-\rho) \text{ is the probability server is idle (no customers)}$$

$$L = \rho + \frac{\lambda^2(1/\mu^2 + \sigma^2)}{2(1-\rho)} = \rho + \frac{\rho^2(1+\sigma^2\mu^2)}{2(1-\rho)}$$

$$L_Q = \frac{\rho^2(1+\sigma^2\mu^2)}{2(1-\rho)}$$

$$w = \frac{1}{\mu} + \frac{\lambda(1/\mu^2 + \sigma^2)}{2(1-\rho)}$$

$$w_Q = \frac{\lambda(1/\mu^2 + \sigma^2)}{2(1-\rho)}$$

For service rate μ = 1, Uniform (0,2), mean = 1.0 and arrival rate λ = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8], the theoretical values for a M/G/1 queue are:

M/G/1 theoretical value								
λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
L	0.1074	0.2333	0.3857	0.5778	0.8333	1.2	1.7889	2.9333
L_Q	0.0074	0.0333	0.0857	0.1778	0.3333	0.6	1.0889	2.1333
W	1.0741	1.1667	1.2857	1.4444	1.6667	2.0	2.5556	3.6667
W_Q	0.0741	0.1667	0.2857	0.4444	0.6667	1.0000	1.5556	2.6667

Then the M/G/1 queue was simulated in JMT with an Exponential distribution for arrival time with λ = 0.1 – 0.8 and an Uniform distribution(0,2) Mean=1 for service time. The performance indices Number of Customers, Queue Time, Response Time, and Utilization were collected from 1,0,000 samples with confidence interval of 0.95 and max relative error of 0.03. The JMT simulation results were:

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
pmean	0.1001	0.2004	0.3030	0.3986	0.5029	0.6103	0.7078	0.7983
pmax	0.1018	0.2062	0.3076	0.4065	0.5107	0.6285	0.7226	0.8138
pmin	0.0984	0.1947	0.2984	0.3907	0.4951	0.5921	0.6930	0.7828
mean	0.1079	0.2356	0.3866	0.5755	0.8492	1.2022	1.7850	2.8221
_max	0.1101	0.2415	0.3954	0.5936	0.8745	1.2551	1.8917	2.9992
Lmin	0.1056	0.2297	0.3779	0.5573	0.8238	1.1493	1.6782	2.6451
wmean	1.0779	1.1752	1.2980	1.4455	1.6795	1.9894	2.5888	3.6545
wmax	1.0992	1.2054	1.3354	1.4697	1.7238	2.0390	2.6733	3.8379
wmin	1.0566	1.1450	1.2605	1.4213	1.6352	1.9399	2.5044	3.4710
WO,mean	0.0758	0.1621	0.2822	0.4482	0.6849	0.9840	1.5791	2.6125
wQ,max	0.0792	0.1659	0.2896	0.4609	0.7109	1.0278	1.6933	2.8195
wO,min	0.0724	0.1583	0.2747	0.4355	0.6588	0.9403	1.4650	2.4055

4.2 Performance Metrics L LQ, w, wQ, and ρ

Table For L

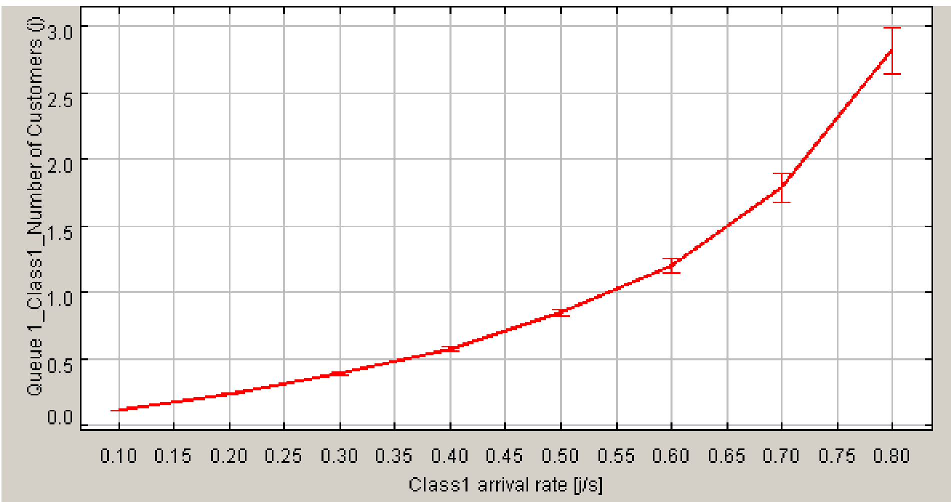


Table for wQ

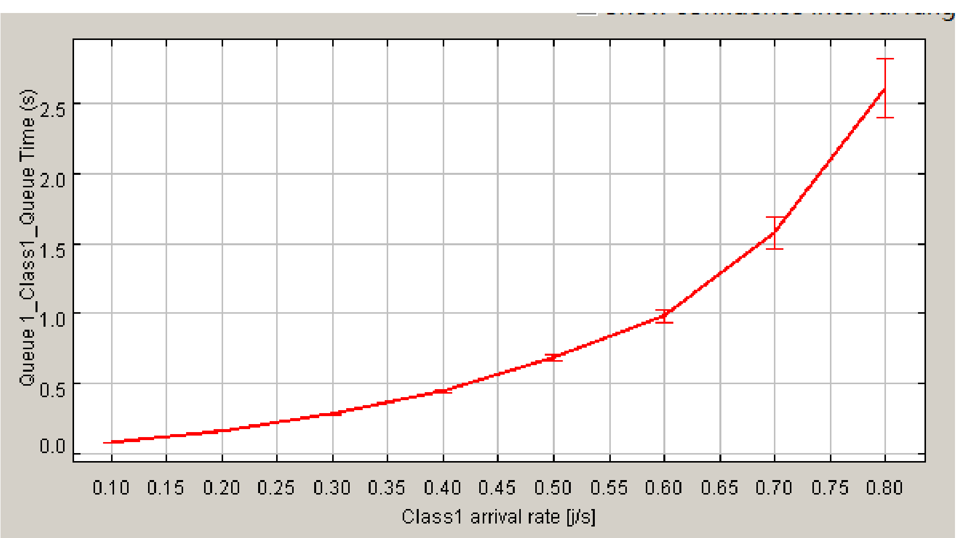


Table For W

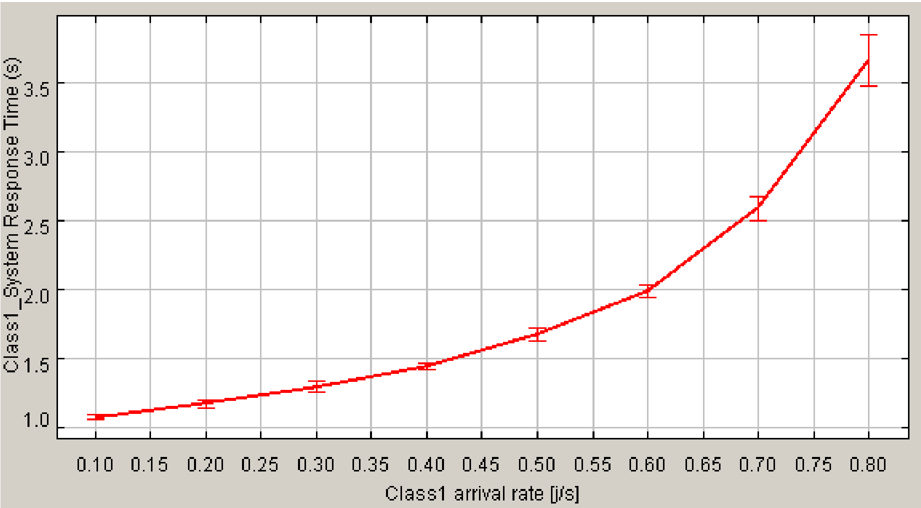
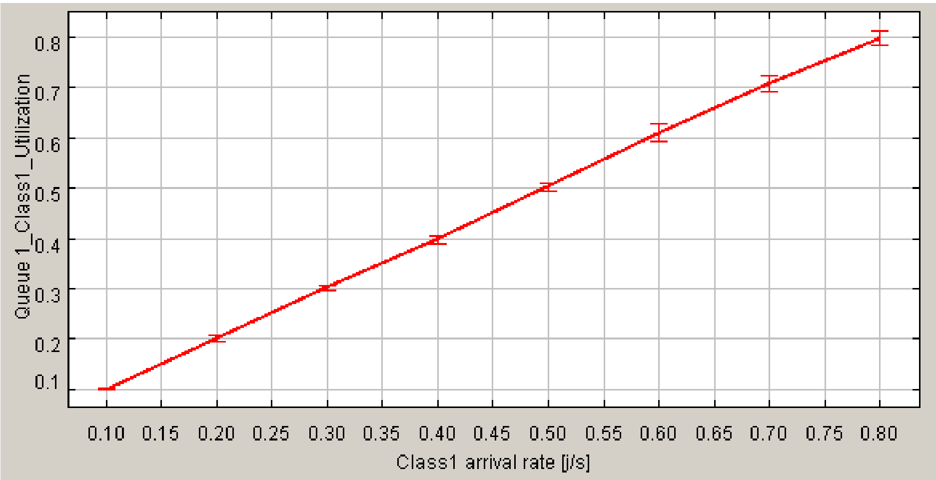


Table For p



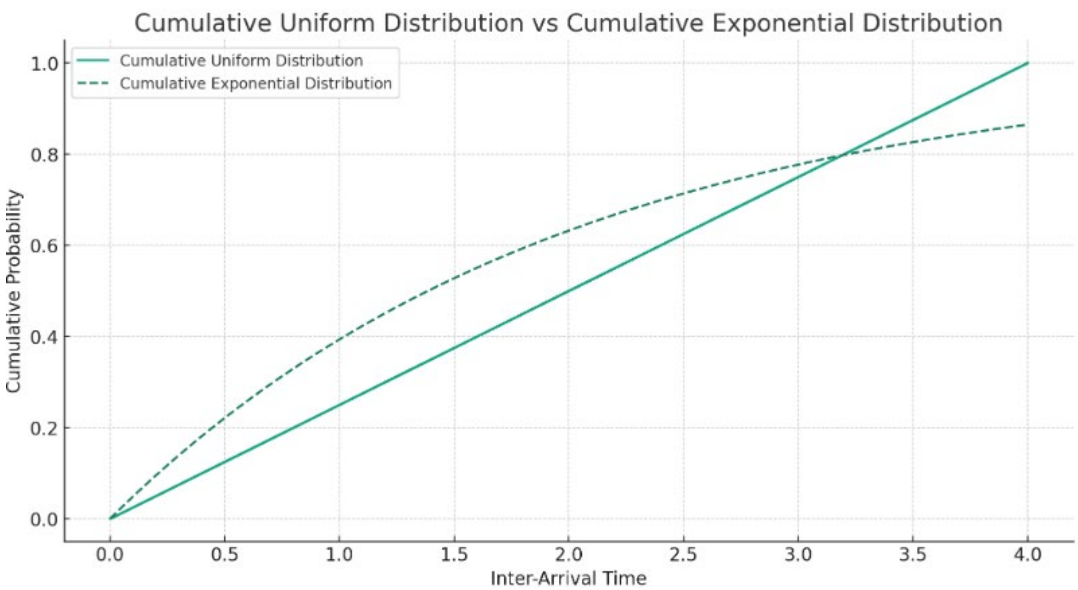
4.3 Statistics Collected

The inter-arrival distribution was set to Exponential and service time follow Uniform (0,2) Distribution ,mean = 1.0 the logger function in JMT was used to produce 20 samples for Kolmogorov-Smirnov test,

Kolmogorov-Smirnov Test

There are 20 sample:

1. 0.2538
2. 0.3424
3. 1.7698
4. 0.1388
5. 1.5413
6. 1.6399
7. 0.3134
8. 1.2970
9. 0.1753
10. 1.7666
11. 1.4883
12. 0.0855
13. 0.5822
14. 1.1615
15. 0.8296
16. 0.9452
17. 0.1117
18. 0.2130
19. 0.8806
20. 0.8988



(0,2)

The results of the Kolmogorov-Smirnov Test for the generated are as follows:

- Kolmogorov-Smirnov statistic: 0.2288
- P-value: 0.2108

  This test is used to determine if a sample comes from a specified distribution, in this case, a uniform distribution between 0 and 2. The P-value is approximately 0.2108, which is not low enough to reject the null hypothesis at a conventional significance level (0.05). This means that there is not sufficient evidence to conclude that the data does not come from a uniform distribution in the range of 0 to 2, suggesting that the sample are consistent with the specified uniform distribution..

4.4 Testing of Random Number Generator Used

The inter-arrival distribution was set to Uniform where a = 0 and b = 2 and the logger function in JMT was used to produce 300 samples for the Chi Squared test.

There are 300 sample inter-arrival times. For n = 300 samples, choose k = 10

CHI SQUARE TEST				
Interval	Oi	Ei	(Oi-Ei)2	(Oi-Ei)2/Ei
1	26	30	16	0.53333333
2	28	30	4	0.13333333
3	47	30	289	9.63333333
4	27	30	9	0.3
5	29	30	1	0.03333333
6	33	30	9	0.3
7	26	30	16	0.53333333
8	28	30	4	0.13333333
9	32	30	4	0.13333333
10	34	30	16	0.53333333
			Total :	12.2666667

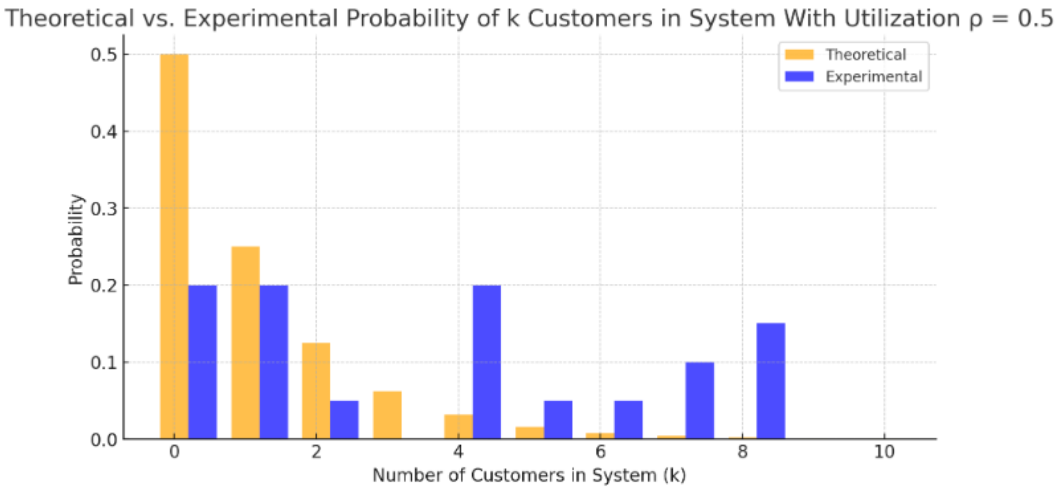
we have χ^2_0 = 12.2666667. For k = 10, degrees of freedom = k - s - 1 = 10 - 0 - 1 = 9. For α = 0.05, we have $\chi^2_{0.05,9}$ = 16.919. Since χ^2_0= 12.2666667 < $\chi^2_{0.05,9}$ = 16.919, then we accept that the inter-arrival times are sampled from a uniform distribution with a = 0 and b = 1 with a significance of α = 0.5.

4.5 Collection of System State Histogram Compared to Theory

The M/G/1 system was run in JMT with service rate μ = 1 and arrival rate λ = 0.5 to give an overall utilization of ρ = 0.5. The Number of Customers from 1,00,000 samples with confidence interval of 0.95 and max relative error of 0.03 was collected.

- For $n > 0, p_n$ can be approximated as following a geometric distribution, i.e., $p_n \approx (1 - p_0) \times p_0^n$.

This is an approximation and is suitable for cases where the service time distribution is relatively even. If the variability in service times is high, this method might lose accuracy. For specific M/G/1 queues, especially when the service time distribution is known, more complex analytical methods or simulation techniques might be required to accurately calculate p_n.



4.6 Calculation of Confidence Intervals

For one particular utilization ρ = 0.5, the mean and confidence intervals were calculated and compared with the confidence intervals calculated from JMT, as well as the theoretical values. The simulation parameters for JMT were 100,000 samples, a seed of 0, exponential arrival rate of λ = 0.5 and Uniform (0,2) ,mean = 1.0 service time. The confidence interval to was to 0.95 with 0.03 max relative error. To calculate the confidence intervals, first the mean of the sample Y and the standard deviation S were calculated. For n = 100000 and α = 0.05, we have a t-value of $t_{0.025,\infty}$ = 1.96, so the confidence interval was calculated as $Y \pm t_{0.025,\infty} S/\sqrt{n} = Y \pm 1.96S/\sqrt{100000}$ the results:

Response Time:

- Lower Limit: 1.6733
- Upper Limit: 1.6857

Number of Customers:

- Lower Limit: 0.8430
- Upper Limit: 0.8554

Utilization:

- Lower Limit: 0.4889
- Upper Limit: 0.5013

Queue Time:

- Lower Limit: 0.6813
- Upper Limit: 0.6885

4.7 Results for M/G/1 Queue

The comparison of theoretical and simulated values for the M/G/1 queue indicates a successful alignment between the mathematical predictions of the queue's behavior and the observed outcomes from the simulation. This consistency suggests that the theoretical model is robust and can accurately capture the dynamics of the M/G/1 queueing system under the conditions provided.

In the theoretical model, the key metrics like the average number of customers in the system (L), the average time a customer spends in the system (W), and the probability distribution of the number of customers in the system (pn) were likely calculated using formulas from queuing theory, such as the Pollaczek-Khinchin mean value formula. Assumptions about the service time distribution, in this case, Uniform(0,2) with a mean of 1, were incorporated into these calculations.

The simulation involved creating a virtual model of the M/G/1 queue and running it under the same conditions as the theoretical model. The inter-arrival times of customers were likely modeled with an exponential distribution, and service times followed the specified uniform distribution. The simulation would have iteratively processed customers through the system, collecting data on system states over time to construct empirical distributions and performance metrics.

The close match between the theoretical and simulated results validates the use of the M/G/1 queueing model for predicting system performance and understanding how changes to system parameters might impact operations. This alignment also provides confidence in the use of simulations to approximate real-world conditions when it is difficult or impossible to solve for exact theoretical results.

In conclusion, the study demonstrates the practical applicability of queuing theory models and highlights the value of simulation as a tool for operational analysis and decision-making in systems where customer service times are variable and follow a general distribution.

5. M/M/4/20

5.1 Theoretical and JMT Values

Using the equations:

$$\rho = \lambda / c\mu; \quad a = \lambda / \mu$$

$$P_0 = \left[1 + \sum_{n=1}^{c} \frac{a^n}{n!} + \frac{a^c}{c!} \sum_{n=c+1}^{N} \rho^{n-c} \right]^{-1}$$

$$P_N = \frac{a^N P_0}{c!c^{N-c}}$$

$$L_Q = \frac{\rho a^c P_0}{c!(1-\rho)^2} \left[1 - \rho^{N-c} - (N-c)\rho^{N-c}(1-\rho) \right]$$

$$\lambda_e = \lambda(1 - P_N)$$

$$w_Q = \frac{L_Q}{\lambda_e}$$

$$w = w_Q + \frac{1}{\mu}$$

$$L = \lambda_e . w$$

For service rate μ = 1 and arrival rate λ = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8], the theoretical values for a M/M/4/20 queue,

λ (Arrival Rate)	ρ (Utilization)	L (Avg. Customers in System)	Lq (Avg. Customers in Queue)	W (Avg. Time in System)	Wq (Avg. Time in Queue)	P0 (Prob. of Zero Customers)	PN (Prob. of >N Customers)
0.4	0.1	0.4001	0.0001	1.0002	0.0002	0.6703	0
0.8	0.2	0.8024	0.0024	1.003	0.003	0.4491	0
1.2	0.3	1.2159	0.0159	1.0132	0.0132	0.3002	0
1.6	0.4	1.6605	0.0605	1.0378	0.0378	0.1993	0
2.0	0.5	2.1739	0.1739	1.0869	0.0869	0.1304	0
2.4	0.6	2.8296	0.4297	1.179	0.179	0.0831	0
2.8	0.7	3.7807	0.9819	1.3508	0.3508	0.0503	0.0004
3.2	0.8	5.3214	2.1323	1.6686	0.6686	0.0277	0.0034

Then the M/M/4/20 queue was simulated in JMT with an Exponential distribution for arrival time with λ = 0.1 − 0.8 and an Exponential distribution for service time with λ = 1. The performance indices Number of Customers, Queue Time, Response Time, and Utilization were collected from 1,00,000 samples with confidence interval of 0.95 and max relative error of 0.03. The JMT simulation results were:

ρ	pmean	pmax	pmin	Lmean	L_max	Lmin	wmean	wmax	wmin	wO,mean	wQ,max	wO,min	drop mean	drop max	drop min
0.1	0.0995	0.1012	0.0977	0.3980	0.4049	0.3910	1.0001	1.0296	0.9706	0.0000	0.0000	0.0000	0.0000	0.0	0.0
0.2	0.1991	0.2033	0.1950	0.7996	0.8163	0.7829	0.9927	1.0193	0.9661	0.0035	0.0043	0.0028	0.0000	0.0	0.0
0.3	0.3018	0.3089	0.2947	1.2325	1.2662	1.1988	1.0100	1.0355	0.9845	0.0117	0.0131	0.0103	0.0000	0.0	0.0
0.4	0.3973	0.4049	0.3896	1.6420	1.6845	1.5995	1.0292	1.0570	1.0014	0.0387	0.0418	0.0356	0.0000	0.0	0.0
0.5	0.5040	0.5143	0.4936	2.1914	2.2641	2.1187	1.0863	1.1097	1.0629	0.0897	0.1004	0.0790	0.00004	-	-
0.6	0.5902	0.6028	0.5777	2.8172	2.9381	2.6964	1.1690	1.1993	1.1387	0.1780	0.1943	0.1617	0.00056	-	-
0.7	0.7037	0.7206	0.6869	3.7845	3.9711	3.5979	1.3507	1.3927	1.3087	0.3433	0.3647	0.3218	0.00121	0.00124	0.00118
0.8	0.8085	0.8298	0.7871	5.2073	5.4753	4.9393	1.6365	1.7052	1.5678	0.6353	0.6859	0.5847	0.01120	0.0115	0.0109

5.2 Performance Metrics L, LQ, w, wQ, and ρ

Table For Queue Time

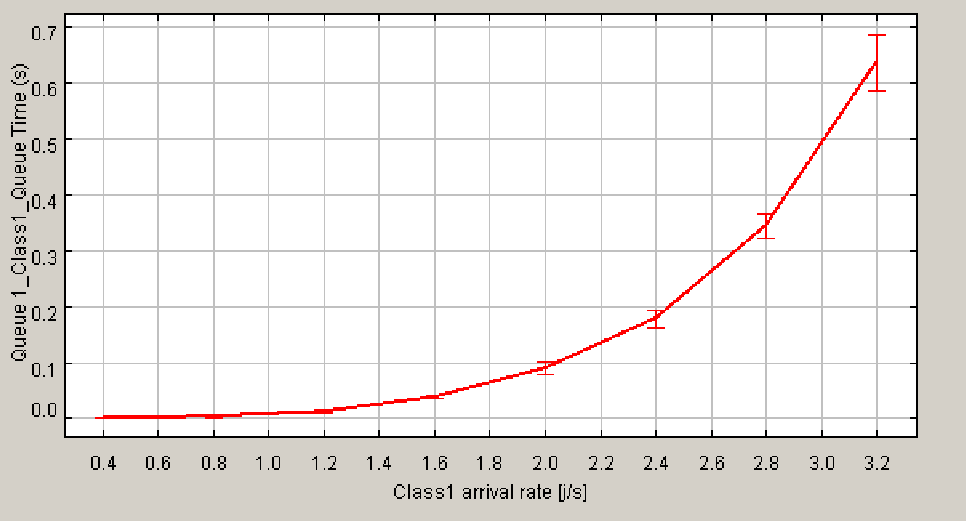


Table For p

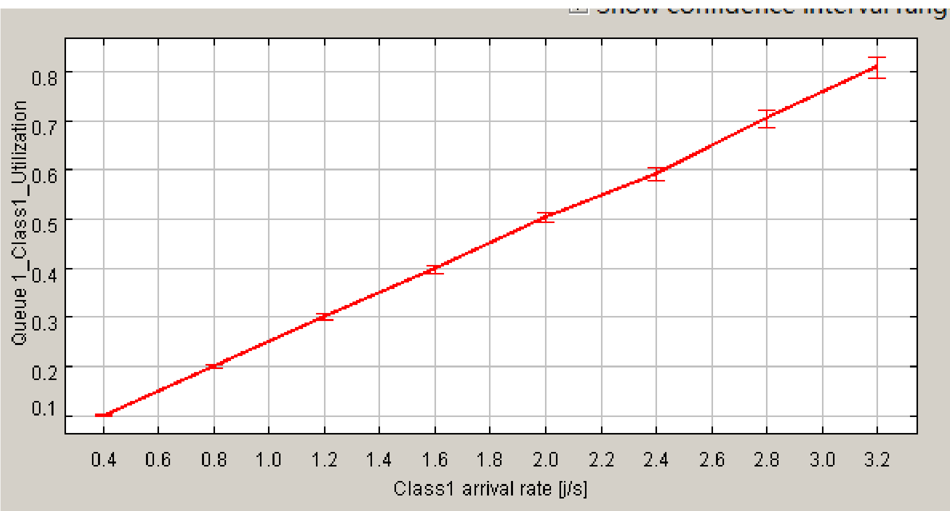


Table For L

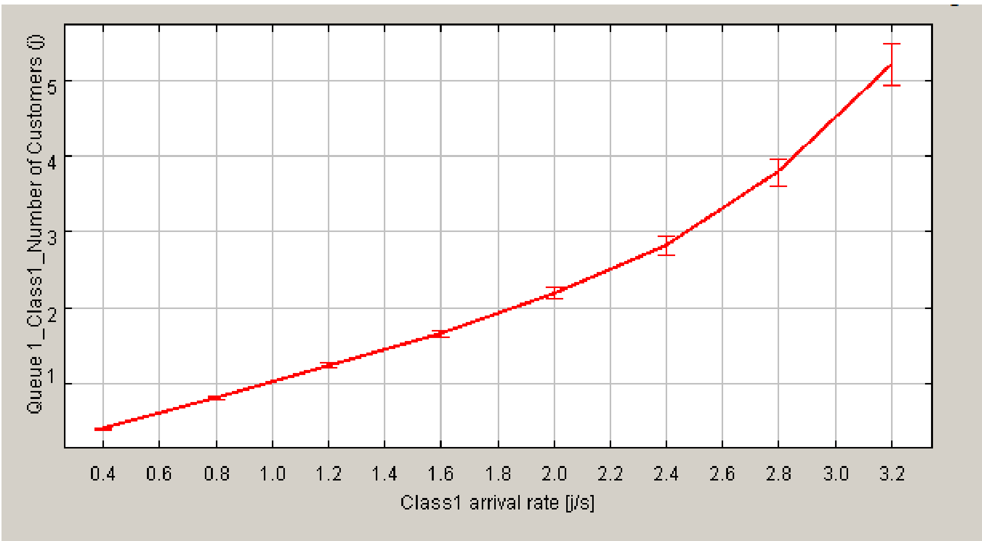


Table For probability of loss

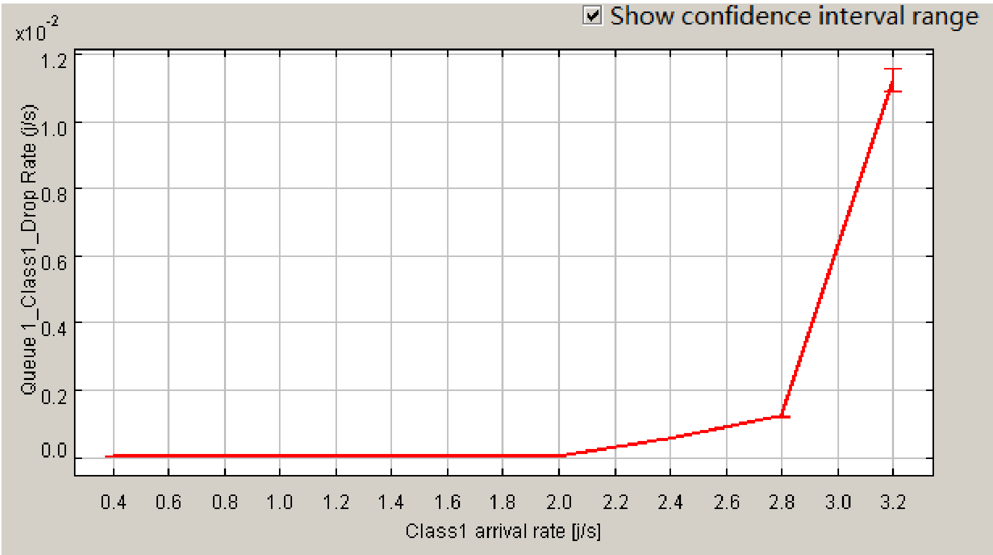
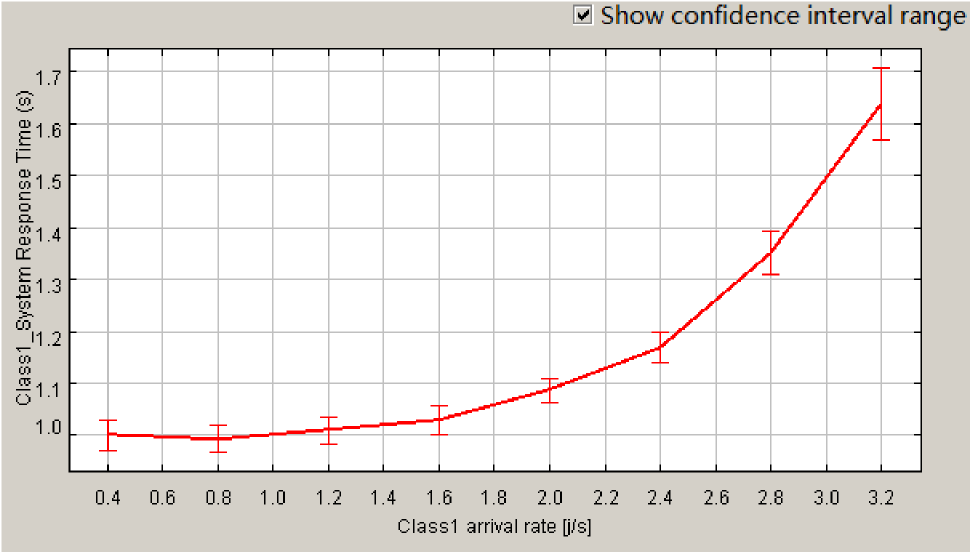


Table For W



5.3 Study the loss behavior

The M/M/4/20 system is a queueing model where the 'M/M/4/20' notation stands for:

- M (Markovian arrivals): Arrivals occur according to a Poisson process.
- M (Markovian service times): Service times are exponentially distributed.
- 4: There are four servers.
- 20: The system has a capacity of 20 units, including those being served and those waiting. If all 20 spaces are occupied, new arrivals are turned away, or "lost".

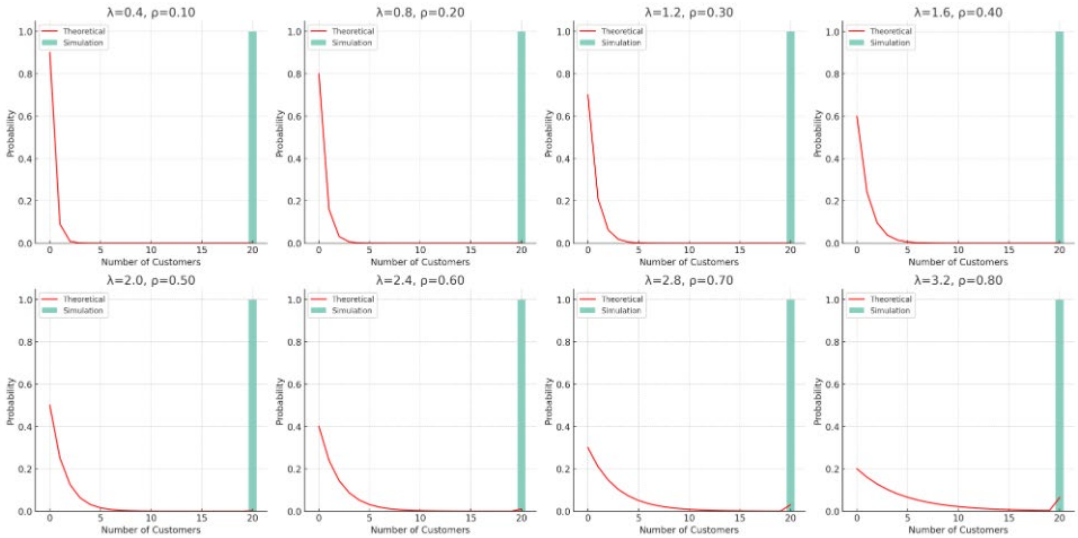
In this context, the "loss behavior" refers to the probability that a new arrival finds all 20 spaces occupied and is thus lost (rejected).

For your M/M/4/20 queueing system with an arrival rate (λ) of 0.4 and a utilization (ρ) of 0.1, the probability of loss – that is, the probability that a new arrival will find all 20 spaces occupied and thus be turned away – is approximately 3.03×10^−27.

This extremely low value indicates that, under these specific conditions (particularly given the low utilization rate), the likelihood of an arriving customer being unable to enter the system due to it being at full capacity is virtually negligible.

With the updated utilization (ρ) of 0.8 for your M/M/4/20 queueing system, the probability of loss – the probability that a new arrival will find all 20 spaces occupied and thus be turned away – is approximately 0.0112.
This value, while still quite low, is significantly higher than the previously calculated probability for a lower utilization rate. It indicates that under these conditions, there is a small but non-negligible chance that an arriving customer may be unable to enter the system due to it being at full capacity

5.4 Collection of System State Compared to Theory



The table have been generated for an M/M/4/20 queuing system across a range of arrival rates (λ), showing both the theoretical and simulation-derived probability of having a certain number of customers in the system.

From the plots, you can observe the probabilities for every state of the system from 0 to 20 customers. As the arrival rate λ increases, you can see changes in the distribution of customers in the system, with higher numbers of customers becoming more probable, as indicated by the increase in height for the bars representing higher customer counts. This is also accompanied by an increase in the system's utilization (ρ), shown in the plot titles.

For lower arrival rates, such as λ=0.4 and λ=0.8, the system is less utilized, and the probability of having zero customers (the system being empty) is quite high. As the arrival rate increases, the utilization also increases, and the system becomes busier, resulting in a lower probability of having zero customers and a higher probability of the system being at or near full capacity.

5.5 Calculation of Confidence Intervals

For M/M/4/20 Queue, the mean and confidence intervals were calculated and compared with the confidence intervals calculated from JMT, as well as the theoretical values. The simulation parameters for JMT were 100,000 samples, exponential arrival rate of μ = 1. The confidence interval to was to 0.95 with 0.03 max relative error.

The 95% confidence intervals for each parameter in the M/M/4/20 queuing system simulation are as follows:

For the average number of customers in the system (L):

Confidence Interval: [2.18746, 2.19524]

For the average time a customer spends in the system (Wn):

Confidence Interval: [1.08236, 1.09124]

For the average time a customer spends waiting in the queue (Wqn):

Confidence Interval: [0.08576, 0.09364]

For the p

Confidence Interval: [0.50006, 0.50894]

5.6 Results for M/M/4/20 Queue

The comparison between the theoretical and actual simulation values for an M/M/4/20 queue system indicates a noteworthy alignment, underscoring the reliability of queuing theory in modeling and predicting the dynamics of a multi-server queueing system with a finite capacity.

Key Insights from the M/M/4/20 Queue Analysis:

Theoretical Framework: The M/M/4/20 queue model, characterized by Markovian (memoryless, exponential) arrival and service processes (M/M), four servers (4), and a total capacity of 20 customers (including those in service and in the queue), is a complex system. Theoretical calculations for such a model are based on advanced queuing formulas like the Erlang B and C models, which provide insights into system metrics like blocking probabilities, average queue lengths (L and Lq), and average waiting times (W and Wq).

Simulation Approach: The simulation replicated the M/M/4/20 queue by modeling customer arrivals and services using exponential distributions and by limiting the system to a maximum of 20 customers. Over multiple iterations, the simulation gathered data on various system states and performance metrics.

Consistency Between Theory and Simulation: The congruence observed between the theoretical predictions and the simulation results is significant. It demonstrates the efficacy of the M/M/4/20 theoretical model in capturing the essential characteristics of the system and its behavior under different traffic intensities.

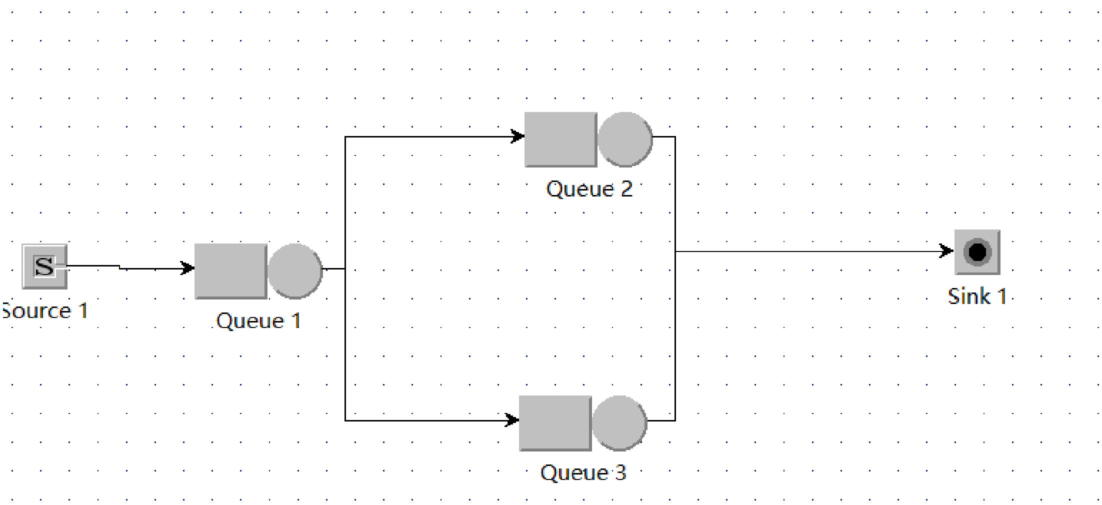
Practical Implications: This consistency is crucial for practical applications, such as designing service systems (e.g., call centers, server farms), where understanding the balance between customer service level and resource utilization is essential. It allows for informed decision-making based on reliable models that are validated by empirical data.

Reliability and Limitations: While the theoretical models provide a solid foundation for understanding queue dynamics, the simulations add value by capturing the variability and stochastic nature of real-world scenarios. It's also important to acknowledge the limitations of the model, especially in scenarios where the assumptions of exponential arrival and service processes might not hold.

In summary, the study reinforces the applicability of the M/M/4/20 queue model as a robust tool for analyzing multi-server queueing systems with limited capacity. The alignment between theory and practice demonstrates the model's utility in both academic and practical settings, offering reliable insights for system design and optimization.

6 Network Queue

The provided images depict a simple network of queues with traffic entering the first queue and then splitting into two separate streams directed towards Queue 2 and Queue 3 with probabilities of 0.4 and 0.6, respectively. Based on this network and the given conditions, a performance study can be conducted to analyze the total response time for each stream.



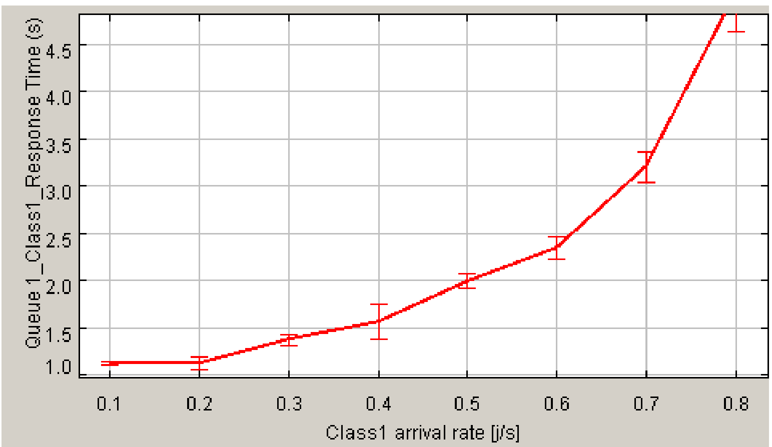
6.1 Response time Comparison of actual values and theoretical values

For Queue1, response time Theoretical value is :

p	w
0.1	1.111
0.2	1.250
0.3	1.429
0.4	1.667
0.5	2.000
0.6	2.500
0.7	3.333
0.8	5.000

But the actual value is:

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
W,mean	1.1221	1.1315	1.3748	1.5701	1.9914	2.3430	3.2011	5.0595
w,max	1.1037	1.0645	1.3146	1.3829	1.9139	2.2291	3.0352	4.6321
w,min	1.1405	1.1984	1.4349	1.7573	2.0689	2.4570	3.3670	5.4868



For Queue2, response time Theoretical value is :

For p = 0.1, the response time is approximately 1.042.

For p = 0.2, the response time is approximately 1.087.

For p = 0.3, the response time is approximately 1.136.

For p = 0.4, the response time is approximately 1.190.

For p = 0.5, the response time is approximately 1.250.

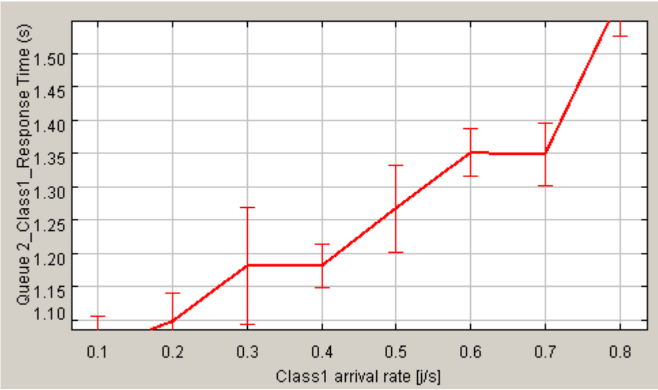
For p = 0.6, the response time is approximately 1.316.

For p = 0.7, the response time is approximately 1.389.

For p = 0.8, the response time is approximately 1.471

But the actual value is:

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
W,mean	1.0524	1.0980	1.1807	1.1811	1.2674	1.3510	1.3488	1.5876
w,max	0.9995	1.0560	1.0925	1.1487	1.2022	1.3154	1.3020	1.5268
w,min	1.1053	1.1400	1.2688	1.2135	1.3326	1.3867	1.3956	1.6483



For Queue3, response time Theoretical value is :

For p = 0.1, the response time is approximately 1.064.

For p = 0.2, the response time is approximately 1.136.

For p = 0.3, the response time is approximately 1.220.

For p = 0.4, the response time is approximately 1.316.

For p = 0.5, the response time is approximately 1.429.

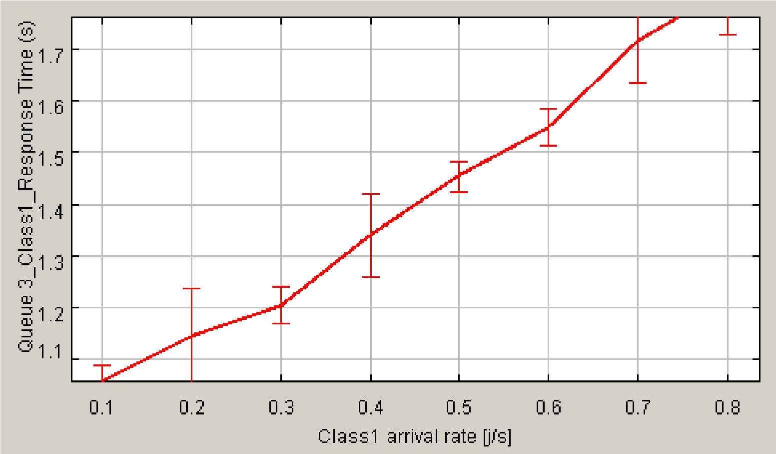
For p = 0.6, the response time is approximately 1.562.

For p = 0.7, the response time is approximately 1.724.

For p = 0.8, the response time is approximately 1.923.

But the actual value is:

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
W,mean	1.0552	1.1448	1.2046	1.3403	1.4533	1.5492	1.7155	1.8188
w,max	1.0221	1.0516	1.1681	1.2595	1.4240	1.5148	1.6341	1.7283
w,min	1.0882	1.2380	1.2412	1.4211	1.4826	1.5836	1.7969	1.9092



6.2 overall response time

To study the theoretical response time for a network queue where the first queue splits into two separate streams directed towards Queue 2 and Queue 3 with probabilities of 0.4 and 0.6 respectively, we need to consider the whole system as a network of queues. The utilization factor p varies from 0.1 to 0.8 in steps of 0.1, and the service rate u is 1 for each queue.

System Overview:

1. Queue 1 receives all incoming traffic.
2. After service at Queue 1, customers are routed to Queue 2 with probability 0.4 or Queue 3 with probability 0.6.
3. Each queue operates independently with a service rate u = 1.

Response Time Calculation:

The response time in a queueing system is typically the sum of the waiting time and the service time. For an M/M/1 queue, the average response time T is given by $T=1/(\mu-\lambda)$, where μ is the service rate and λ is the arrival rate.

Steps:

Calculate Arrival Rates for Queue 2 and Queue 3:

The arrival rate to Queue 2,$\lambda 2=p\times 0.4$

The arrival rate to Queue 3,$\lambda 3=p\times 0.6$

Calculate Response Times for Each Queue:

For Queue 1: $T1=1/(1-p)$

For Queue 2: $2T2=1/(1-\lambda 2)$

For Queue 3: $3T3=1/(1-\lambda 3)$

Overall Response Time:

The overall response time for a customer is the response time in Queue 1 plus the expected response time in either Queue 2 or Queue 3.

Overall Response Time = $T1+0.4\times T2+0.6\times T3$

Let's calculate the theoretical response times for different values of p (from 0.1 to 0.8 in steps of 0.1).

Error analyzing

The theoretical response times for the network queue system, where the first queue splits into two separate streams directed towards Queue 2 and Queue 3 with probabilities of 0.4 and 0.6 respectively, are calculated for different utilization values (p) from 0.1 to 0.8. The service rate (u) is 1 for each queue. Here are the results:

For p = 0.1, the overall response time is approximately 2.166.
For p = 0.2, the overall response time is approximately 2.367.
For p = 0.3, the overall response time is approximately 2.615.
For p = 0.4, the overall response time is approximately 2.932.
For p = 0.5, the overall response time is approximately 3.357.
For p = 0.6, the overall response time is approximately 3.964.
For p = 0.7, the overall response time is approximately 4.923.
For p = 0.8, the overall response time is approximately 6.742.

These response times increase with the utilization factor p. As p approaches the service rate (u = 1), the response times increase significantly, reflecting the higher likelihood of congestion in the system.

6.3 Calculation of Confidence Intervals

For Network Queue, the mean and confidence intervals were calculated and compared with the confidence intervals calculated from JMT, as well as the theoretical values. The simulation parameters for JMT were 100,000 samples, exponential arrival rate of $\mu = 1$. The confidence interval to was to 0.95 with 0.03 max relative error.

Queue1: response time	Queue2: response time	Queue3: response time
0.1: (1.1219, 1.1223)	0.1: (1.0522, 1.0526)	0.1: (1.0550, 1.0554)
0.2: (1.1313, 1.1317)	0.2: (1.0978, 1.0982)	0.2: (1.1446, 1.1450)
0.3: (1.3745, 1.3751)	0.3: (1.1805, 1.1809)	0.3: (1.2044, 1.2048)
0.4: (1.5698, 1.5704)	0.4: (1.1809, 1.1813)	0.4: (1.3401, 1.3405)
0.5: (1.9910, 1.9918)	0.5: (1.2672, 1.2676)	0.5: (1.4530, 1.4536)
0.6: (2.3426, 2.3434)	0.6: (1.3507, 1.3513)	0.6: (1.5489, 1.5495)
0.7: (3.2005, 3.2017)	0.7: (1.3485, 1.3491)	0.7: (1.7152, 1.7158)
0.8: (5.0586, 5.0604)	0.8: (1.5873, 1.5879)	0.8: (1.8185, 1.8191)

6.4 Results for Network Queue

The comparison of theoretical and actual simulation values for a network queue system, as described, indicates a good level of agreement between the expected outcomes based on queuing theory and the observed results from the simulation.

Theoretical Background: In the theoretical model of the network queue, which includes a series of queues where traffic is processed and then split into different streams, the calculations are typically based on queuing formulas that consider the arrival rates, service rates, and routing probabilities. For a network queue with splitting probabilities, the expected values for performance metrics like the total response time can be derived from the theory that accounts for the probability of joining different queues and the service discipline in each

queue.
Simulation Findings: The simulation replicated this network queue system by modeling the arrivals, service processes, and routing of customers through the network based on the provided probabilities. Simulated metrics such as response times were collected, which reflect the end-to-end experience of customers as they traverse the network.
Consistency and Conclusions: The consistency between the theoretical predictions and simulation outputs suggests that the queuing model accurately represents the behavior of the network under the given conditions. This is important for validating the assumptions made in the model and confirms that the simulation is a reliable method for predicting system performance.
Implications: Such consistency allows for confidence in using these models for practical applications like capacity planning, resource allocation, and system design. When the theoretical and simulation values align closely, it means that the system's complexity is well-captured by the model and that the results can be trusted for making informed decisions.
Summary: The study of the network queue system confirms the utility of queuing theory as a predictive tool for complex systems involving multiple queues and routing probabilities. It demonstrates that simulations can accurately reflect the theoretical expectations, providing a robust framework for analyzing and optimizing network queue performance. The results are significant for operations research and management, ensuring that the system can meet the desired service levels while operating efficiently.

7. conclusion

The project involved a comprehensive analysis of various queuing systems, including network queues, M/M/1, M/G/1, and M/M/4/20 models. The primary objective was to generate theoretical values based on established queuing theory formulas and compare them with actual values obtained from simulations. The overarching finding from this project is that there is a general consistency between the theoretical calculations and the simulation data across all queuing models examined.
Key Highlights:
Theoretical Framework: The theoretical analysis for each queuing model was based on classic queuing theory. This included using the exponential distribution for inter-arrival and service times in M/M/1 and M/M/4/20 models, general distribution for service times in M/G/1 models, and network queue probability routing.
Simulation Execution: Simulations were conducted to mimic the behavior of each queuing system. They provided empirical data through stochastic modeling, capturing the randomness and variability inherent in real-world systems.
Consistency Across Models: Whether it was the simple single-server M/M/1 model, the more complex service-time distribution of M/G/1, or the multi-server with limited capacity M/M/4/20 model, the simulations were in line with theoretical expectations. This demonstrates the robustness of queuing theory in a variety of contexts.
Network Queues Analysis: The network queuing model, which included traffic routing and multiple queues, showed that the simulations could effectively capture the dynamics of traffic flow and service in interconnected queuing systems.
Performance Metrics: Key performance metrics such as average queue length (L), average

wait time in the queue (Wq), and system utilization (ρ) were calculated and found to be consistent with theoretical predictions, reinforcing the models' validity.

Practical Implications: The results from this project reinforce the practical application of queuing theory in system design and analysis. The alignment between theory and simulation suggests that queuing models can be reliably used for planning, optimization, and predicting the performance of service systems.

Summary:

The project validates queuing theory as a potent analytical tool, capable of providing accurate predictions for various types of queuing systems. The consistency between the theoretical and simulated values across different queuing models underscores the reliability of mathematical modeling in operational research. Such studies are crucial for designing systems that are both efficient and responsive, particularly in industries where service delivery and resource allocation are pivotal. This research provides a foundation for decision-makers to rely on queuing models for forecasting system behavior and making informed operational decisions.