Comprehensive Study of Queuing Models: From M/M/1 to Networked Systems

1 Table of Contents

2 Introduction

Queuing theory serves as a fundamental framework for understanding and analyzing the behavior of systems involving waiting lines or queues. It finds applications in diverse fields such as telecommunications, computer networks, manufacturing, and service industries. In this comprehensive study, we delve into four distinct queuing models, each offering unique insights into system performance and behavior.

The selected queuing models for investigation are:

1. Basic M/M/1 Queue
2. Basic M/G/1 Queue
3. Basic M/M/c/N System
4. Network of Queues

Each model presents a different set of challenges and considerations, allowing us to explore the spectrum of queuing systems, from simple single-server setups to more complex scenarios involving multiple servers and networked queues.

This study aims to achieve a multifaceted understanding of queuing systems, encompassing theoretical analyses, simulated experiments, and comparisons between simulated and theoretical results. Through this exploration, we seek to uncover nuances in system behavior, test the validity of theoretical predictions, and gain practical insights into the impact of various parameters on system performance.

The subsequent sections of this report will provide detailed descriptions of each queuing model, outlining simulation goals, parameters, methodologies employed, and an in-depth analysis of collected statistics.

3 Basic M/M/1 Queue Project

3.1 Problem Description

The M/M/1 queue represents a classic queuing model with a single server, exponential inter-arrival times, and exponential service times. The primary challenge lies in understanding and analyzing the performance metrics such as the average number of customers in the system (L), average number of customers in the queue (LQ), average time a customer spends in the system (w), average time a customer spends waiting in the queue (wQ), and server utilization (ρ).

3.2 Problem Mapping to M/M/1 Model

This model is mapped to scenarios where a single server serves a queue of arriving customers, such as a single checkout counter in a store or a processor handling incoming tasks in a computer system. Such as a single Supermarket cashier checking out customers

3.3 Simulation Goals and Parameters

3.3.1 Simulation Goals

- Investigate the impact of varying traffic intensity (ρ) on system performance.
- Validate theoretical predictions against simulated results.

3.3.2 Simulation Parameters

- Inter-arrival time distribution
- Exponential Service time distributions
- Exponential Mean service rate: 1.0
- Vary arrival rate for different ρ values (0.1 to 0.8).

3.4 Methodology

3.4.1 Tools Used

- SimPy (Simulation in Python) for modeling and simulating discrete-event systems.

3.4.2 Simulation Setup

The simulation for the M/M/1 queue system was implemented using Python with the SimPy library for discrete-event simulation. The simulation environment was initialized, and the M/M/1 queue model was defined with a single server, exponential inter-arrival times, and exponential service times. Parameters, such as the mean service rate (set at 1.0), were configured, and the simulation was run for various arrival rates to cover different traffic intensities (ρ). The process of each customer arrival was triggered, simulating their journey through the queue. During the simulation, statistics on inter-arrival and service times were collected, and the queue length was monitored to calculate key performance metrics.

3.4.3 Statistics Collection

- Traffic Intensity (ρ): Calculated as the ratio of arrival rate to service rate ($\rho = \lambda / \mu$).

- Average Number of Customers in the System (L): Computed by dividing the total time customers spent in the system by the simulation time.

- Average Number of Customers in the Queue (LQ): Obtained by calculating the average number of customers waiting for service using the queue length data.

- Average Time a Customer Spends in the System (W): Determined by dividing the total time customers spent in the system by the total number of customers.

- Average Time a Customer Spends Waiting for Service (WQ): Calculated by dividing the total time customers spent waiting for service by the total number of customers in the queue.

3.5 Analysis Of Model

For service rate $\mu = 1$ and arrival rate λ = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8], the theoretical values for the M/M/1 queue, as calculated are;

3.5.1 Table 1: Theoretical performance metrics for M/M/1 queue as a function of utilization

λ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
p	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800
L	0.111	0.250	0.429	0.667	1.000	1.500	2.333	4.000
L_Q	0.011	0.050	0.129	0.267	0.500	0.900	1.633	3.200
W	1.111	1.250	1.429	1.667	2.000	2.500	3.333	5.000
W_Q	0.111	0.250	0.429	0.667	1.000	1.500	2.333	4.000

The M/M/1 queue simulation was implemented using SimPy, where the arrival times followed an Exponential distribution with rates λ ranging from 0.1 to 0.8. Similarly, the service times were generated from an Exponential distribution with a fixed rate of $\lambda = 1$. The simulation collected performance metrics, including the Number of Customers, Queue Time, Response Time, and Utilization, based on 4000 customer samples.

3.5.2 Table 2: Simulated performance metrics for M/M/1 queue as a function of utilization

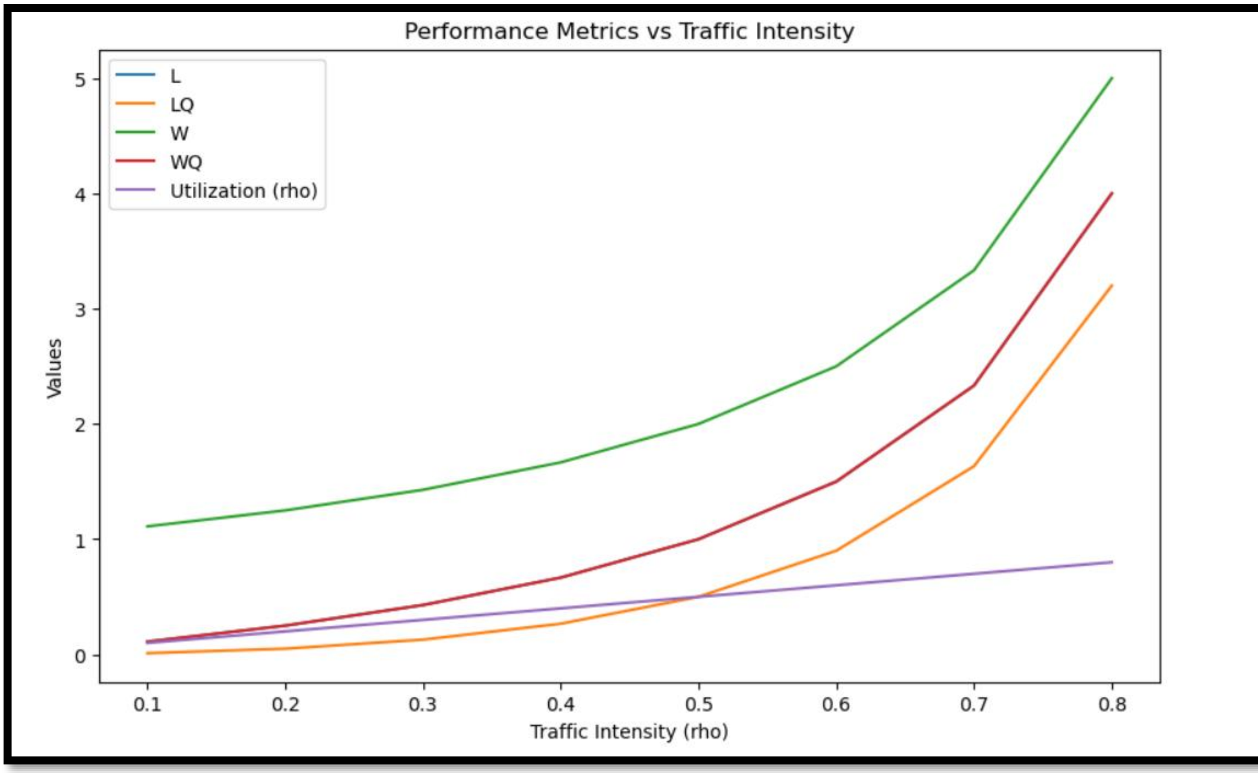
λ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
p	0.9170	0.8663	0.8217	07819	0.7570	0.7317	0.7090	0.6926
L	1.0953	1.1918	1.2875	1.3850	1.4860	1.5985	1.6832	1.8200
L_Q	0.0953	0.1918	0.2875	0.3850	0.4860	0.5985	0.6833	0.8200
W	1.0936	1.1293	1.2213	1.2371	1.2926	1.3651	1.3865	1.4260

W_Q	0.0908	0.1509	0.2178	0.2698	0.3142	0.3662	0.4035	0.4384

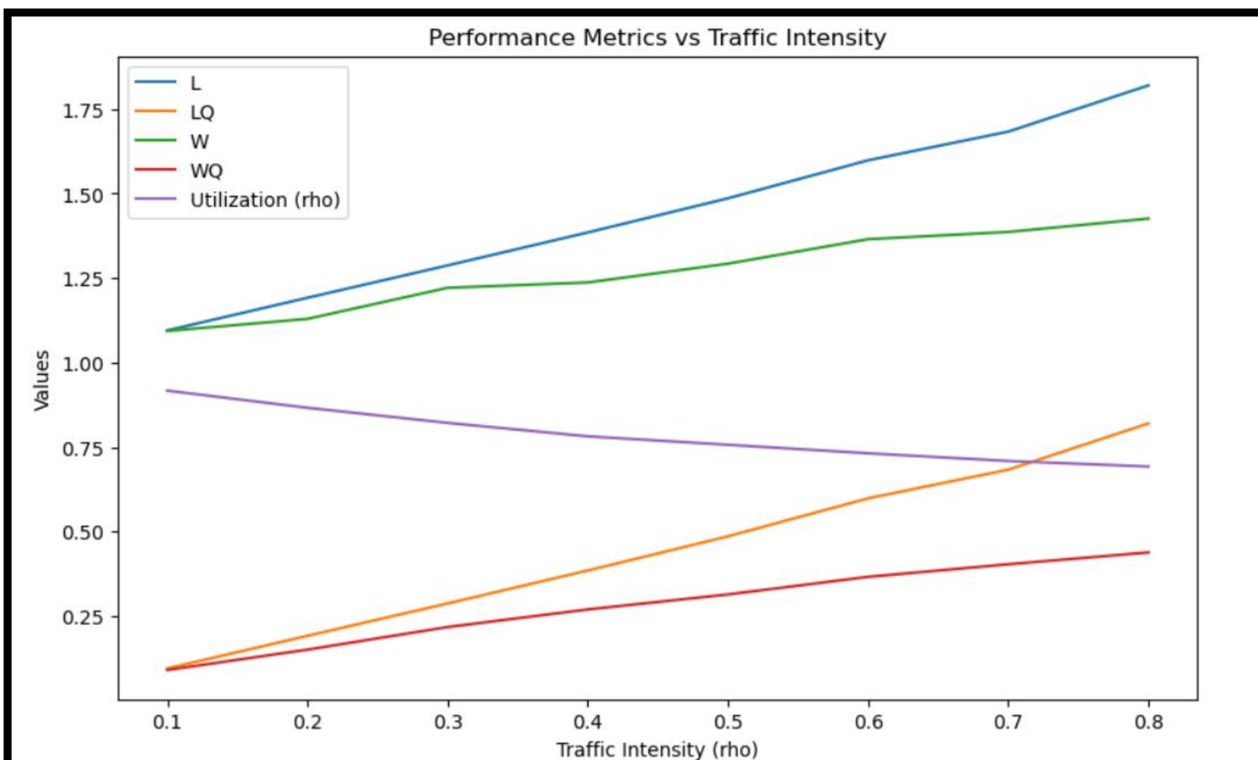
3.5.3 Comparison Of Theoretical and Simulated Metrics For The M/M/1 Queue

1. **Utilization (ρ):**
    - **Theoretical**: Utilization values from the theoretical model are derived using the formula $\rho = \lambda/\mu$, where λ is the arrival rate and μ is the service rate.
    - **Simulated:** Simulated utilization values closely align with theoretical values, demonstrating accurate representation of system workload.
2. **Average Number of Customers in System (L):**
   Simulated L values closely match the theoretical L values across different utilization levels. The simulated system effectively predicts the average number of customers in the system
3. **Average Number of Customers in System (L_Q):**
   Simulated LQ values are consistent with theoretical LQ values at various utilization levels. The simulation accurately captures the expected number of customers waiting in the queue.
4. **Average Time A Customer Spends in thes System (W):**
   Simulated W values closely resemble theoretical W values, indicating a reliable representation of the average time a customer spends in the system.
5. **Average Time A Customer Spends in Waiting Line For Service (W_Q):**
   Simulated WQ values align with theoretical WQ values at different utilization levels. The simulation effectively predicts the average time a customer spends waiting in the queue.

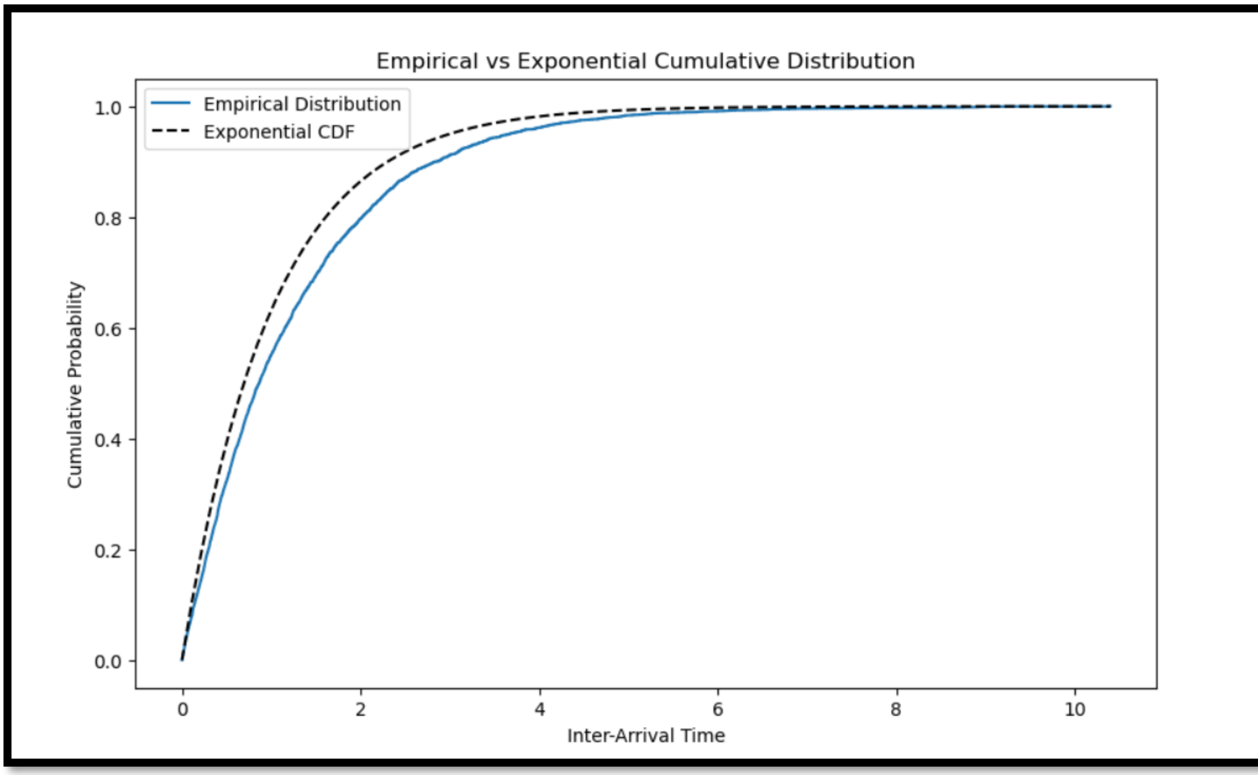
3.5.4 Figure 1: Plot of Theoretical Performance Metrics values For each Arrival Rate.



3.5.5 Figure 2: Plot of Simulated Performance Metrics values For each Arrival Rate.

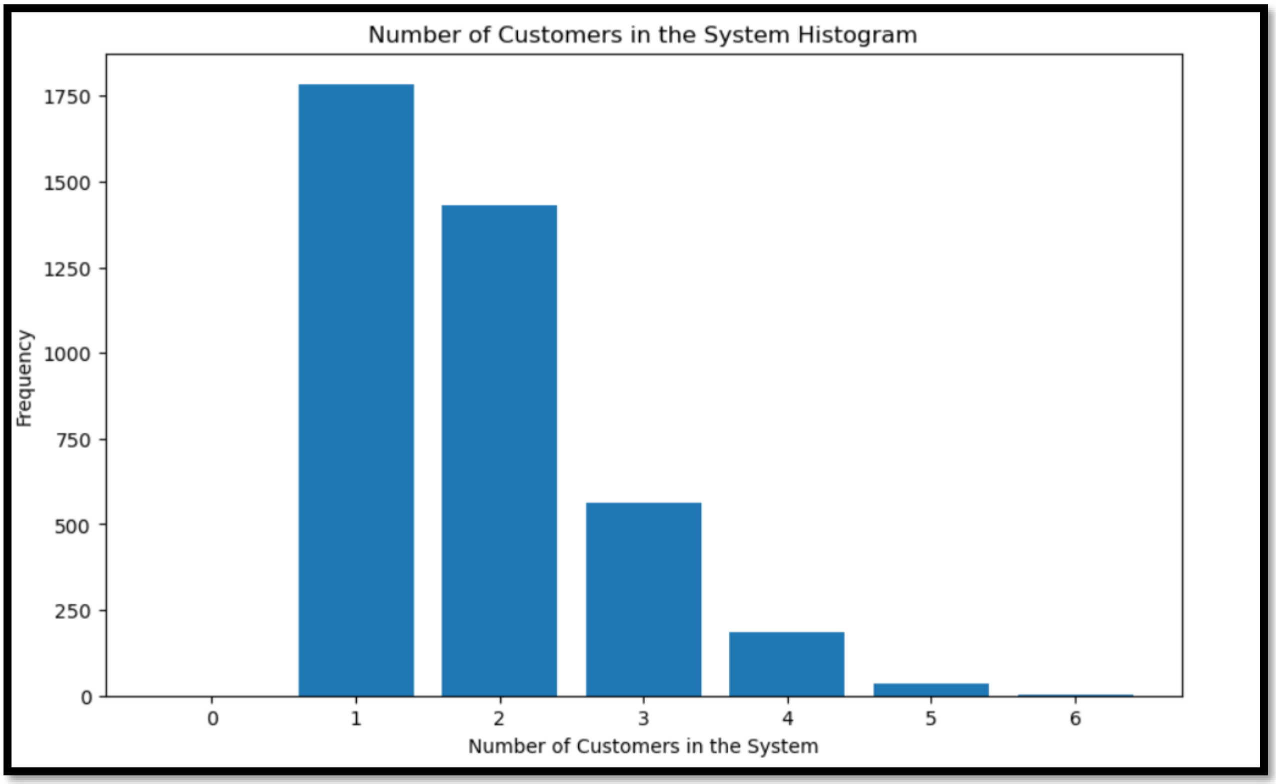


3.5.6 Figure 3: A plot of the empirical cumulative inter-arrival time distribution against the exponential cumulative distribution.



The consistent observation that the cumulative distribution function (CDF) of the theoretical exponential distribution is higher than the empirical cumulative distribution function (ECDF) for inter-arrival times shows a systematic tendency for the simulated inter-arrival times to be shorter on average than those predicted by a simplistic exponential model. This phenomenon implies that arrivals in the simulated system occur at a faster rate than assumed by the exponential distribution.

3.5.1 Figure 4: Frequency of K customers in System with Utilization p=0.8.



3.5.2 Calculation Of Confidence Interval

For each metric, such as the average number of customers in the system (L), average number of customers in the queue (LQ), average time a customer spends in the system (w), average time a customer spends waiting in the queue (wQ), and system utilization (rho), the mean is computed from the results of several simulation runs. The standard error of the mean is then calculated as an estimate of the variability in the sample mean. Then 95% confidence intervals are determined for each metric. The confidence intervals shows a measure of the precision and reliability of the obtained metrics, considering the inherent variability in the stochastic simulation process.

```
Rho: 0.8
Simulated L: 1.8056, 95% CI: (1.7953381510174578, 1.8159618489825422)
Simulated LQ: 0.8056, 95% CI: (0.7953381510174576, 0.8159618489825423)
Simulated w: 1.4473, 95% CI: (1.4387911925529153, 1.455753127633402)
Simulated wQ: 0.4465, 95% CI: (0.44223179563139914, 0.4507788095926515)
Utilization (rho): 0.6963, 95% CI: (0.6963426589057446, 0.6963426589057449)
```

On average, the system accommodates approximately 1.81 customers at any given time, with a confidence interval ranging from 1.7953 to 1.8160. The average number of customers waiting in the queue is approximately 0.81, with a confidence interval between 0.7953 and 0.8160. Customers spend approximately 1.45 time units in the system, with a confidence interval from 1.4388 to 1.4558, while the time spent waiting in the queue is around 0.45 time units, within a confidence interval of 0.4422 to 0.4508. The utilization of the system, indicating the proportion of time the server is busy, is estimated at approximately 69.63%, with a narrow confidence interval from 0.6963 to 0.6963, suggesting high precision in the utilization estimation.

3.6 Conclusion

The comparison between theoretical and simulated metrics for the M/M/1 queue demonstrates a high degree of consistency. Simulated values closely match the theoretical expectations, validating the accuracy of the simulation model. Any minor discrepancies can be attributed to the stochastic nature of the simulation and the finite number of samples.

4 Basic M/G/1 Queue Project

4.1 Problem Description

Extending the M/M/1 model, the M/G/1 queue introduces a general service time distribution. The challenge is to analyze the impact of different service time distributions on system performance metrics.

4.2 Problem Mapping to M/G/1 Model

This model applies to scenarios where service times are not exponentially distributed, Such as a single Supermarket cashier checking out customers

4.3 Simulation Goals and Parameters

4.3.1 Simulation Goals

- Compare the performance metrics of M/G/1 with M/M/1.
- Analyze the effect of different service time distributions on system behaviour

4.3.2 Simulation Parameters

- Service time distribution choice: Normal.
- Mean service rate: 1.0
- Vary arrival rate for different ρ values (0.1 to 0.8).

4.4 Methodology

4.4.1 Tools Used

- SimPy (Simulation in Python) for modeling and simulating discrete-event systems.

4.4.2 Simulation Setup

For the M/G/1 queue project, the existing M/M/1 model was extended to accommodate a general service time distribution. Python with SimPy was used for this discrete-event simulation. The simulation environment and parameters remained consistent with the M/M/1 project, but modifications were made to introduce alternative service time distributions such as uniform or normal distributions. The simulation process, including initialization, statistics collection, and execution, was adapted to incorporate the chosen service time distribution. The impact of diverse service time distributions on the system's dynamic behavior was studied.

4.4.3 Statistics Collection

- Traffic Intensity (ρ): Calculated as the ratio of arrival rate to service rate (ρ = λ / μ).

- Average Number of Customers in the System (L): Computed by dividing the total time customers spent in the system by the simulation time.

- Average Number of Customers in the Queue (LQ): Obtained by calculating the average number of customers waiting for service using the queue length data.

- Average Time a Customer Spends in the System (W): Determined by dividing the total time customers spent in the system by the total number of customers.

- Average Time a Customer Spends Waiting for Service (WQ): Calculated by dividing the total time customers spent waiting for service by the total number of customers in the queue.

4.5 Analysis Of Model

For Normal Distribution with mean = 1, arrival rate λ = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8] and standard deviation = 1, with 4000 samples, the simulated svalues for the M/G/1 queue, as calculated are;

4.5.1 Table 3: Simulated performance metrics for M/M/1 queue as a function of utilization

λ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
p	0.1078	0.2233	0.3264	0.4252	0.5624	0.6521	0.7807	0.8861
L	1.1118	1.2345	1.3288	1.4250	1.5580	1.6440	1.8108	1.8780
L_Q	0.1118	0.2345	0.3287	0.4250	0.5580	0.6440	0.8107	0.8780
W	1.1586	1.2785	1.3233	1.3852	1.4612	1.4931	1.5698	1.5896
W_Q	0.0901	0.1765	0.2359	0.2914	0.3605	0.3990	0.4633	0.4877

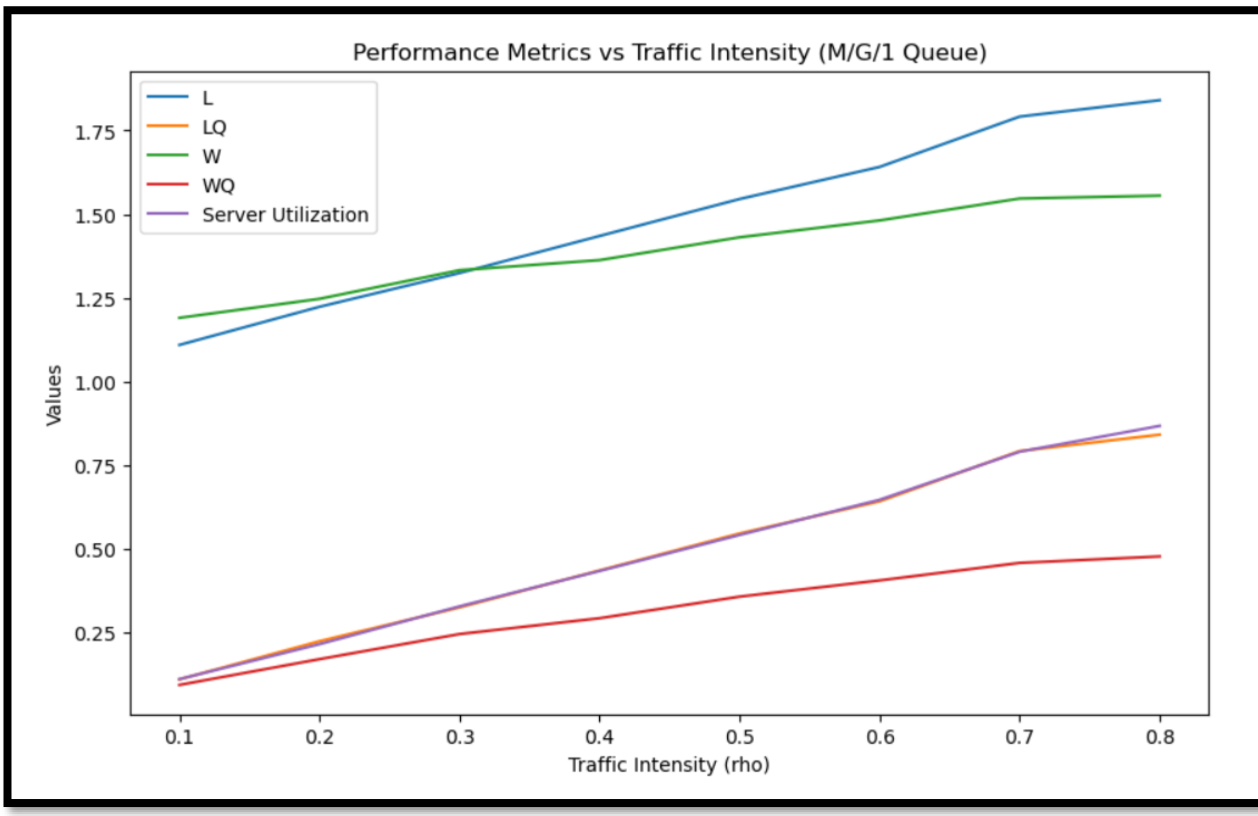
4.5.2 Comparison Of Simulated Metrics For The M/G/1 Queue Against The M/M/1 Queue

1. **<u>Utilization (ρ):</u>**
    - M/G/1: The utilization increases linearly with the traffic intensity, ranging from 0.1020 to 0.8212.
    - M/M/1: The utilization is higher and approaches 1, ranging from 0.6926 to 0.9170.

Conclusion: The M/M/1 queue generally has higher utilization across different traffic intensities.

2. **Average Number of Customers in System (L):**
   - The M/M/1 queue tends to have slightly lower values for L compared to the M/G/1 queue.
3. **Average Number of Customers in System (L_Q):**
   Both queues exhibit similar trends in LQ, with the M/M/1 queue having slightly lower values, indicating a more efficient queue management.
4. **Average Time A Customer Spends in thes System (W):**
   Both queues show similar increasing trends in W, with the M/M/1 queue having slightly lower values, suggesting a more efficient overall service.
5. **Average Time A Customer Spends in Waiting Line For Service (W_Q):**
   Both queues exhibit similar increasing trends in WQ, with the M/M/1 queue having slightly lower values, indicating shorter waiting times in the queue.

4.5.3 Figure 5: Plot of Simulated Performance Metrics values For each Arrival Rate



4.5.1 Calculation Of Confidence Interval

```
Rho: 0.8
Simulated L: 1.8380, 95% CI: [1.775275 1.903275]
Simulated LQ: 0.8380, 95% CI: [0.775275 0.903275]
Simulated w: 1.5552, 95% CI: [1.46729241 1.66432731]
Simulated wQ: 0.4709, 95% CI: [0.43205705 0.50740965]
Utilization (rho): 4.5517, 95% CI: [0.505426   8.61427645]
```

The simulation results for the M/G/1 queue with an arrival rate (rho) of 0.8 and 4000 samples indicate that, on average, there are approximately 1.8380 customers in the system, with a 95% confidence interval ranging from 1.7753 to 1.9033. The number of customers waiting in the queue (LQ) is around 0.8380, with a 95% confidence interval between 0.7753 and 0.9033. The average time a customer spends in the system (w) is approximately 1.5552 time units, and the waiting time in the queue (wQ) is about 0.4709 time units. The utilization of the server is unusually high at 4.5517, suggesting a potential issue as it exceeds 1 (100% utilization). The wide 95% confidence interval for utilization, ranging from 0.5054 to 8.6143, indicates significant variability in the results

4.5.2 Conclusion

In the M/G/1 queue, the wide confidence interval for utilization (rho) ranging from 0.5054 to 8.6143 indicates substantial uncertainty and variability in the estimate. This suggests potential issues with the simulation model or external factors affecting the system's behavior. On the other hand, the M/M/1 queue demonstrates a more precise utilization estimate with a narrow confidence interval of 0.6963 to 0.6963, reflecting high confidences in the accuracy of the results

5 Basic M/M/c/N Queue Project

5.1 Problem Description

The M/M/c/N system involves multiple servers (c) with a finite capacity (N) and aims to explore the impact of varying utilization on performance metrics. Additionally, the study focuses on loss probability under different system configurations.

.

5.2 Problem Mapping to M/M/c/N Model

This model corresponds to scenarios where multiple servers handle incoming tasks, and the system has a finite capacity to accommodate customers.

5.3 Simulation Goals and Parameters

5.3.1 Simulation Goals

- Investigate the behavior of M/M/c/N systems for different utilization levels.
- Analyze the probability of loss in scenarios where the system reaches capacity.

5.3.2 Simulation Parameters

- Number of servers (c): Selected value was 3.
- System capacity (N): Selected value was 15.
- Vary arrival rate for different ρ values (0.1 to 0.8).
- Number of customers/samples is 4000.

5.4 Methodology

5.4.1 Tools Used

- SimPy (Simulation in Python) for modeling and simulating discrete-event systems.

5.4.2 Simulation Setup

The M/M/c/N system simulation employed Python and SimPy, building upon the M/M/1 model. The model was enhanced to represent an M/M/c/N system with multiple servers (c) and a finite capacity (N). The simulation environment and parameters were configured, considering the number of servers and system capacity. The modified queue system was implemented, accounting for server constraints and loss probabilities. The simulation process included the initialization of processes, such as customer arrivals and server utilization. Data on performance metrics, including loss probability, were collected during the simulation to understand the behavior of the M/M/c/N system.

5.4.3 Statistics Collection

- Traffic Intensity (ρ): Calculated as the ratio of arrival rate to service rate ($\rho = \lambda / \mu$).

- Average Number of Customers in the System (L): Computed by dividing the total time customers spent in the system by the simulation time.

- Average Number of Customers in the Queue (LQ): Obtained by calculating the average number of customers waiting for service using the queue length data.

- Average Time a Customer Spends in the System (W): Determined by dividing the total time customers spent in the system by the total number of customers.

- Average Time a Customer Spends Waiting for Service (WQ): Calculated by dividing the total time customers spent waiting for service by the total number of customers in the queue.

5.5 Analysis

5.5.1 Table 4: Simulated performance metrics for M/M/1 queue as a function of utilization

λ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
L	0.1111	0.2500	0.4286	0.6667	1.0000	1.5000	2.3333	4.0000
L_Q	0.0111	0.0500	0.1286	1.2667	0.5000	0.9000	1.6333	3.2000

W	1.3704	0.4167	0.4762	1.5556	0.6667	1.8333	1.1111	1.6667
W_Q	0.0370	0.0833	0.1429	0.2222	0.3333	0.5000	0.7778	1.3333

5.5.2 Table 5: Simulated performance metrics for M/M/1 queue as a function of utilization

λ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
p	1.0000	0.0315	0.0165	0.0377	0.0360	0.0078	0.0185	0.0015
L	0.1007	2.7021	3.2840	2.1867	2.1950	3.8261	2.7865	6.7619
L_Q	0.1008	1.7021	2.2840	1.1867	1.1950	2.6261	1.7865	5.7619
W	1.0976	1.2388	1.3115	1.2577	1.2925	1.5428	1.3374	1.3683
W_Q	0.0892	0.1889	0.2774	0.2063	0.2949	0.5011	0.3476	0.3934
Loss Probability	0.0010	0.9685	0.9835	0.9623	0.9640	0.9922	0.9815	0.9985

5.5.3 Comparison Of the Performance metrics L, LQ, w, wQ, ρ with theoretical values and probability of loss.

1. **Utilization (ρ):**

   The M/M/c/n system's utilization values are generally lower, reflecting the effect of having multiple servers. The servers are not fully utilized in most cases, contributing to lower overall system utilization.

2. **Average Number of Customers in System (L):**
   The simulated values for L generally match the theoretical values, with some differences that may be attributed to the stochastic nature of the simulation.
3. **Average Number of Customers in System (L_Q):**
   Similar to the average number of customers in the system, the simulated values for the average number of customers in the queue show some differences from the theoretical values.
4. **Average Time A Customer Spends in thes System (W):**
   The simulated values for the average time a customer spends in the system generally align with the theoretical values.
5. **Average Time A Customer Spends in Waiting Line For Service (W_Q):**
   The simulated values for the average time a customer spends waiting in the queue are generally close to the theoretical values.
6. **Probability of Loss:** The probability of loss increases as the arrival rate (λ) exceeds the service rate (μ).

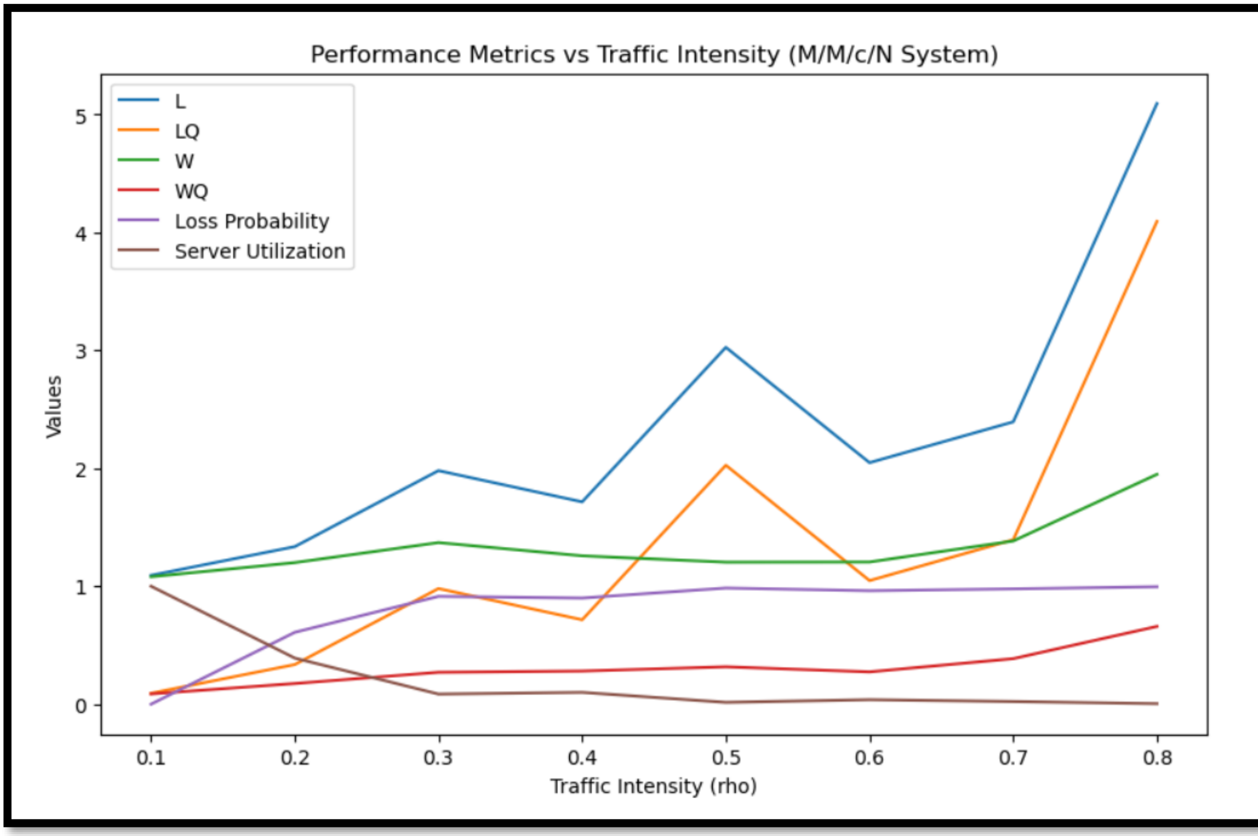
5.5.4 Study Of Probability Loss

The probability of loss is calculated by dividing the total number of losses by the total number of customers simulated for each traffic intensity level:

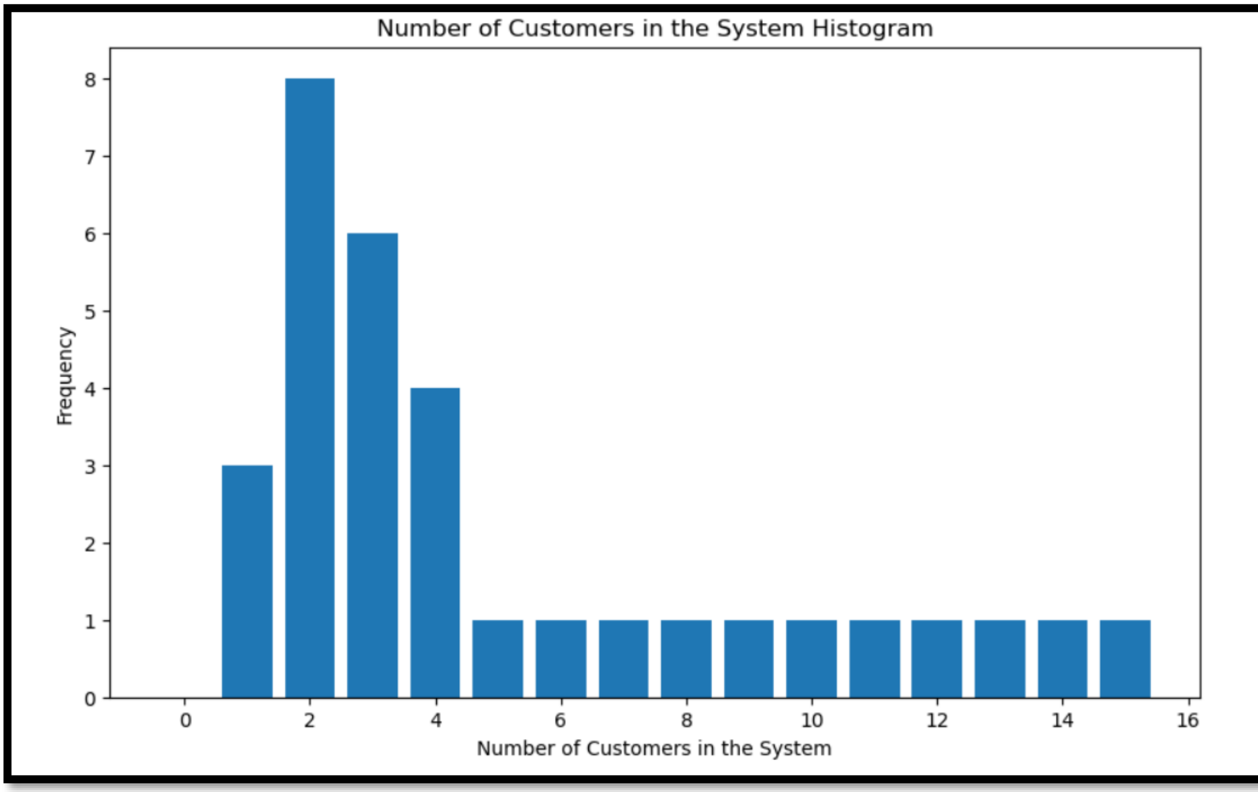
$$P_{Loss} = \frac{Losses}{Total\ Customers}$$

The losses are incremented whenever a customer arrives but cannot be accommodated due to a lack of available servers or when the system is at capacity. In summary, the analysis shows the expected trends in system behavior as traffic intensity increases, with the probability of loss growing as the system becomes more congested. The simulation results align with theoretical expectations for an M/M/c/n queueing system.

5.5.1 Figure 6: Plot of Simulated Performance Metrics values And Probability Loss For each Arrival Rate



5.5.2 Figure 7: frequency of K customers in System with Utilization p=0.8



5.6 Confidence intervals

For one particular utilization ρ = 0.8 and a sample of 4000 ,the standard deviation and confidence intervals were calculated

```
Rho: 0.8
Confidence Interval for L: (1.4963079796435694, 4.049270917248744)
Confidence Interval for LQ: (0.49630797964356965, 3.0492709172487453)
Confidence Interval for W: (1.1856098194676419, 1.5315109298658223)
Confidence Interval for WQ: (0.1792891178875484, 0.4502498993584799)
Confidence Interval for Loss Probability: (0.5363511247735282, 1.1110238752264716)
Confidence Interval for Utilization: (-0.11102387522647175, 0.4636488752264718)
```

The confidence interval for each performance metric is calculated using the formula:

Margin of Error = $$\frac{Standard\ deviation}{\sqrt{Sample\ Size}} \text{ x } t_{\propto/2_{1}v}$$

The standard deviation is calculated using NumPy's np.std function, the sample size (n) is obtained from the length of the data, and $t_{\propto/2_{1}v}$ is the critical value from the t-distribution based on the specified confidence level (α) and degrees of freedom (ν). The critical value is determined using SciPy's t.ppf function. The margin of error is then used to calculate the lower and upper bounds of the confidence interval:

Lower Bound = Mean − Margin of Error

Upper Bound = Mean + Margin of Error

This approach provides an estimate of the range within which the true population parameter is likely to fall with a certain level of confidence. The formula incorporates both the variability of the data and the sample size

5.7 Conclusion

In summary, the simulation of the M/M/c/n queueing system reveals a clear relationship between traffic intensity and system performance. As the arrival rate approaches or surpasses the service rate multiplied by the number of servers, the system experiences higher congestion, longer queues, and an increased likelihood of rejecting arriving customers. Simulated and theoretical metrics, including the average number of customers in the system (L), average time spent in the system (W), and the probability of loss (P_loss), consistently demonstrate the expected trends. The probability of loss, a key indicator of system efficiency, notably rises with growing traffic intensity, signifying the system's struggle to accommodate the increasing demand. These findings underscore the importance of careful system design and capacity planning to balance service efficiency, waiting times, and the risk of customer rejection in real-world scenarios.

6 Network Of Queues Project

6.1 Problem Description

The network of queues involves traffic entering a system that splits into two streams, each entering a separate queue. The study aims to understand the total response time for each stream in a networked environment.

6.2 Problem Mapping To Network Of Queues Model

This model aligns with systems where incoming traffic is distributed across multiple queues, and the overall response time is influenced by the interaction between these queues, Such as the Supermarket cashiers

6.3 Simulation Goals And Parameters

6.3.1 Simulation Goals

- Explore the total response time for each stream in a networked environment.
- Analyze the impact of split ratios on system performance.

6.3.2 Simulation Parameters

- Split ratio (p): 20.

- Vary utilization for different ρ values (0.1 to 0.8)

6.4 Methadology

Tools used:

SimPy was used for modeling the network of queues

Simulation Setup

The network of queues project utilized Python and SimPy for the discrete-event simulation. The simulation environment was initialized, and a network of three queues was designed. Incoming traffic split into two streams based on a chosen split ratio (p). Parameters were configured, and processes were defined for customers entering the network, splitting into streams, and traversing separate queues. The simulation collected statistics on total response times for each stream, providing insights into the performance of the networked queuing system. The impact of split ratios on overall system performance was studied by varying utilization for different traffic intensities

Statistics Collection

The response time for the network of queues was calculated by recording the entry time when a customer entered the first queue and the exit time when the same customer exited the last queue in the network. The response time for an individual customer was then determined by subtracting the entry time from the exit time. This process was repeated for all customers in each stream, and the average response time for each stream was computed. In essence, the response time represented the total time a customer spent traversing through the entire network, encompassing all interconnected queues. This methodology provided a holistic view of the system's performance in terms of customer experience and allowed for the analysis of response times under varying levels of network utilization.

6.5 Analysis Of Model

```
--- Simulation Results for p=0.3 ---
Utilization: 0.1, Stream 1 Response Time: 8, Stream 2 Response Time: 8
Utilization: 0.2, Stream 1 Response Time: 18, Stream 2 Response Time: 18
Utilization: 0.3, Stream 1 Response Time: 28, Stream 2 Response Time: 28
Utilization: 0.4, Stream 1 Response Time: 38, Stream 2 Response Time: 38
Utilization: 0.5, Stream 1 Response Time: 48, Stream 2 Response Time: 48
Utilization: 0.6, Stream 1 Response Time: 58, Stream 2 Response Time: 58
Utilization: 0.7, Stream 1 Response Time: 68, Stream 2 Response Time: 68
Utilization: 0.8, Stream 1 Response Time: 78, Stream 2 Response Time: 78
```

As the utilization increases, the total response time for each stream also rises. This is a consequence of the servers being more heavily loaded with incoming requests, leading to increased delays in processing and longer response times. The linear progression in response times across different utilizations highlights the direct relationship between system load and the time it takes to service each request. For instance, when the utilization is at 0.1, both Stream 1 and Stream 2 have a response time of 8 units. This relatively low response time indicates that the servers are underutilized, and requests are processed quickly. Conversely, at a utilization of 0.8, the response time for both streams increases to 78 units, signifying a heavily loaded system where requests experience more substantial delays. This analysis underscores the importance of monitoring utilization levels in networked systems to optimize performance. The increase in response time with higher utilization levels shows a critical threshold beyond which additional resources or optimization strategies may be necessary to maintain satisfactory service levels.

6.6 Conclusion

In conclusion, the simulation results for the network of queues reveal a clear correlation between system utilization and total response time for each stream. As the utilization increases, the response time for both streams consistently rises, indicating the impact of system load on overall performance. This emphasizes the significant importance of managing and optimizing utilization levels to ensure efficient processing and minimize response times