Assignment - 3

Queueing Systems

Team members

22010848 - شهد محمد مصطفى

22010691 - جنا محمد فتحي أحمد

22011265 - منن محمد ابودهب

Overview:

Queueing theory is the mathematical study of waiting lines or queues. A queueing model is constructed so that queue lengths and waiting times can be predicted. Queueing theory is generally considered a branch of operations research because the results are often used when making business decisions about the resources needed to provide a service.

A queueing system can be described as a system having a service facility at which units of some kind (generically called "customers") arrive for service; whenever there are more units in the system than the service facility can handle simultaneously, a queue (or waiting line) develops. The waiting units take their turn for service according to a preassigned rule, and after service they leave the system. Thus, the input to the system consists of the customers demanding service, and the output is the serviced customers. A queueing system is usually characterized by the following terms:

- Ls : Average number of items in the system
- Lq : Average number of items in the queue
- Ws : Average time in system (wait + service)
- Wq : Average waiting time in queue
- ρ : the fraction of time servers are busy
- P0 : Probability the system is empty

Theoretical calculations:

1. **ρ (Utilization)** = λ / μ
2. **Ls (Average number in system)** = ρ / (1 - ρ)
3. **Lq (Average number in queue)** = ρ^2 / (1 - ρ)
4. **Ws (Average time in system)** = 1 / (μ - λ)
5. **Wq (Average waiting time in queue)** = λ / [μ(μ - λ)]
6. **P_0 (Probability of 0 customers)** = 1 - ρ
7. **P_1, P_2, P_3, … (State probabilities)** = (1 - ρ) × ρ^n
8. Higher λ → lower **P_0**, higher **Ls**, **Lq**, **Ws**, **Wq**

Comparing theoretical and simulation:

First Scenario λ = 4 , μ = 12

Metric	Theoretical	Simulation
ρ	0.3333	0.3391
Ls	0.5	0.5123
Lq	0.1667	0.1731
Ws	7.5	7.5892
Wq	2.5	2.5642
P0	0.6667	0.66
P1	0.2222	0.2194
P2	0.0741	0.0826
P3	0.0247	0.0248

Second Scenario λ = 6 , μ = 12

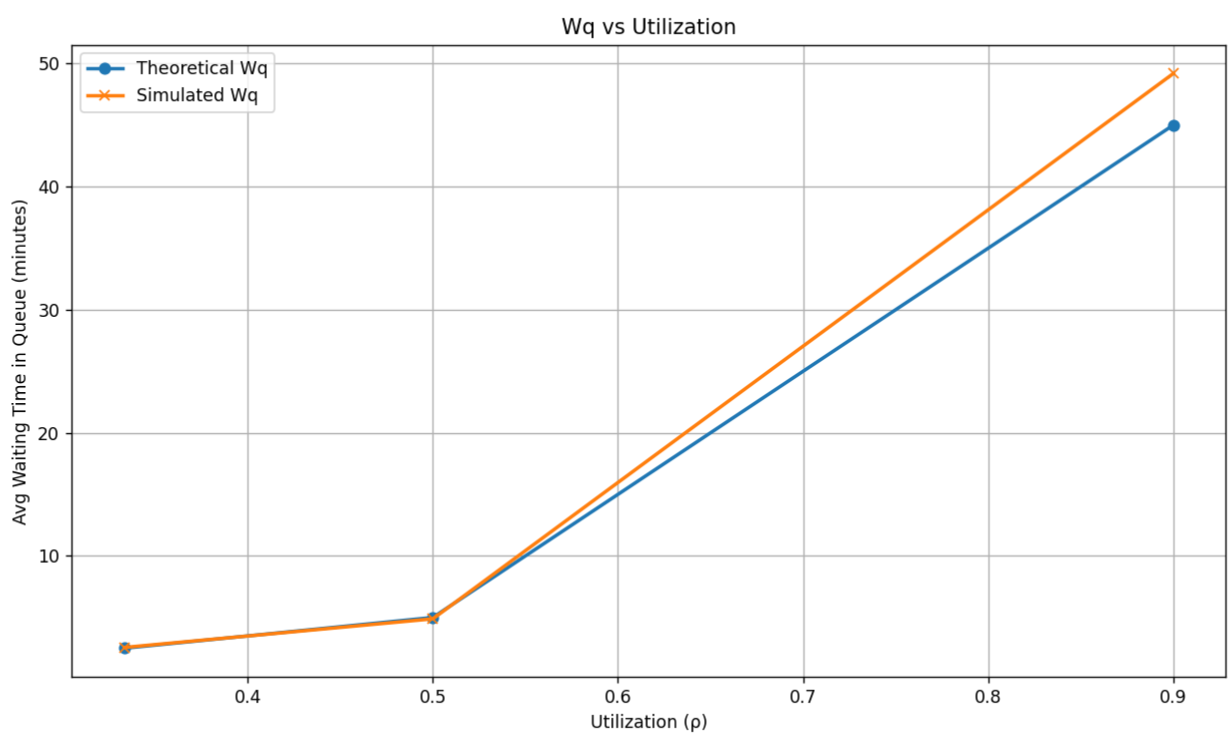
Metric	Theoretical	Simulation
ρ	0.5	0.4939
Ls	1	0.9878
Lq	0.5	0.4936
Ws	10	9.7849
Wq	5	4.8892
P0	0.5	0.5061
P1	0.25	0.243
P2	0.125	0.1257

P3	0.0625	0.0619

Third Scenario $\lambda = 10.8$, $\mu = 12$

Metric	Theoretical	Simulation
ρ	0.9	0.8971
Ls	9	9.8821
Lq	8.1	8.9849
Ws	50	54.0934
Wq	45	49.1895
P0	0.1	0.1029
P1	0.09	0.0866
P2	0.081	0.08
P3	0.0729	0.0775

Plotting



Discussion of discrepancies:

the discrepancies between theoretical and simulated values are natural and expected due to:

1. Randomness and Sample Size in Simulation:

in our simulation we used 20,000 tics which virtually corresponds to minutes ,so If the number of events (customers) is not sufficiently large, the simulation results may deviate from theoretical expectations due to statistical variation. Theoretical results assume an infinite time , while simulations are time-bounded and not provide perfect random of time events .

Randomization used in our code implementation :

```
next_arr = np.random.exponential(1 / self.lam)
next_dep = t + np.random.exponential(1 / self.mu)
```

2. **Warm-up Period (Initial Bias)**

Theoretical model assume the system is in **steady state**, but simulations start from an empty system (cold start). This causes early data to be biased (fewer customers in the system at the beginning), which skews metrics like Ls , Lq , Ws , Wq

3. Precision and Rounding Errors

Minor discrepancies can also arise from floating-point precision or rounding in simulation calculations.

E.g. outputs

ρ **:** 0.3333 vs 0.3391

Ls : : 0.5000 vs 0.5123

4. Queue Saturation at High Traffic Intensity ($\rho=\lambda/\mu$)

the long waiting times and queue lengths. This is visible in:

- Wq: for $\lambda=10.8$, $\mu=12$: 45.0000 (theoretical) vs 49.1895 (simulation)
- Ws : 50.0000 vs 54.093

Description of our code:

This Python implementation provides a comprehensive analysis of M/M/1 queueing systems through both theoretical calculations and discrete-event simulation. The code validates theoretical queueing models against simulation results and visualizes the relationship between system utilization and waiting times.

Code Structure:

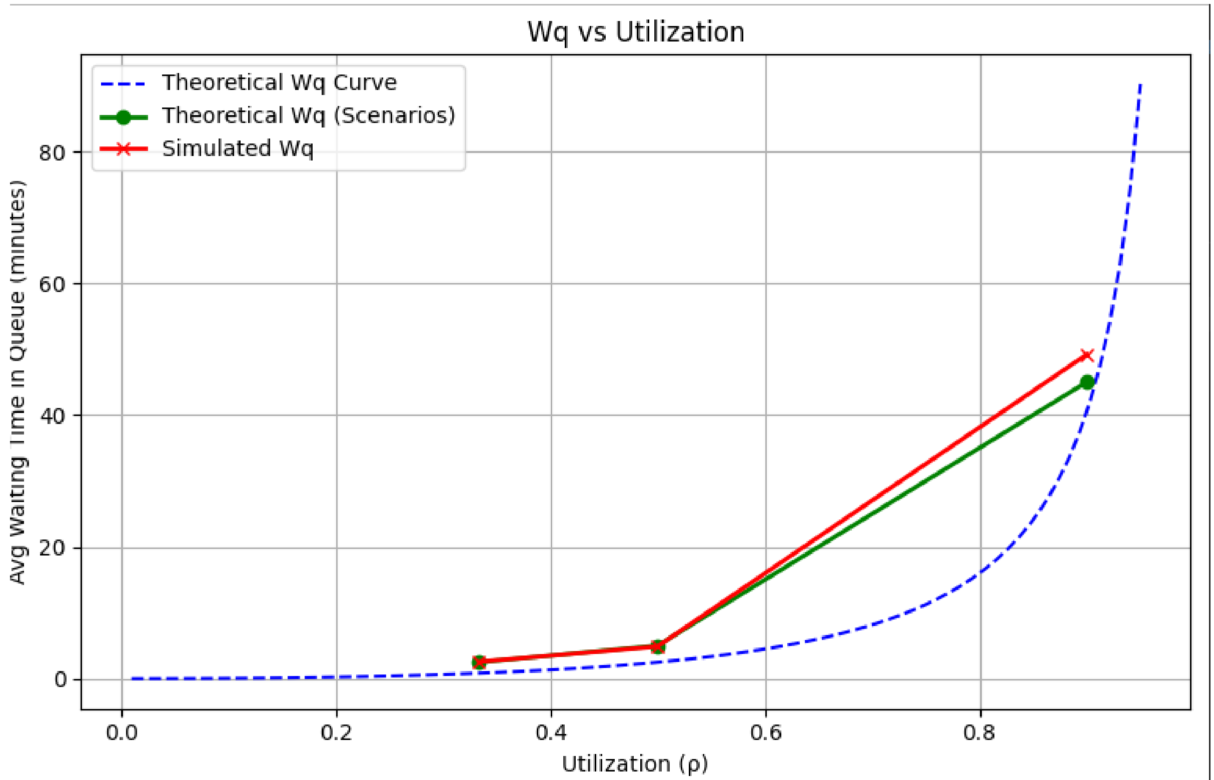
1) Theoretical Analysis Module

   calculates exact M/M/1 queue performance metrics using established queueing theory equations, ensures $\rho = \lambda/\mu < 1$ for system stability

2) Simulation Module
   implements a discrete-event simulation approach to model the M/M/1 queue behavior, converts hourly rates to per-minute rates, sets simulation duration (default: 20,000 minutes for statistical stability)
   manages two primary events: customer arrivals and service completions, uses exponential distributions for inter-arrival and service times, maintains queue state using Python's deque data structure, tracks system metrics throughout the simulation run
   records time-weighted statistics for queue length distributions, calculates cumulative waiting times for both queue and system
3) Analysis and Validation
   The main execution section performs comparative analysis across three scenarios

Bonus :

1. We reran the code and found the plotting differences between the scenarios demanded (done) : $\rho = 0.5$ (e.g., $\lambda = 6$, $\mu = 12$), $\rho = 0.9$ (e.g., $\lambda = 10$, $\mu = 12$)
2. We Plotted the average waiting time in the queue (Wq) from our simulations as a function of the utilization factor (ρ). We plotted the theoretical Wq curve.



3.

4. Describe as utilization factor (ρ) approaches to 1 from graph :
   - As ρ approaches 1, the average waiting time in the queue Wq increases sharply in both the theoretical and simulated results.
     This reflects the system becoming overloaded, causing delays to grow significantly as ρ nears full capacity.

Repo Link:

https://github.com/janamohamed2304/queueing-system-simulation