CAPACITY ANALYSIS PROJECT REPORT

PROJECT BY

Madhukar Reddy Putta

1. Introduction:

In today's fast-paced and customer-centric food service industry, optimizing operational efficiency is critical for maintaining high service quality and customer satisfaction. Restaurants, particularly those offering a variety of menu items, face the challenge of managing demand fluctuations while ensuring that customers experience minimal waiting time. This report presents a simulation-based approach to understanding and improving service efficiency in a restaurant setting by dynamically reallocating servers between different tasks based on real-time demand.

The simulated restaurant environment in this project focuses on serving three primary product categories: pizzas, dishes (main courses other than pizza), and drinks. Each product category experiences its unique demand pattern, and the arrival of customers can vary unpredictably over time.

1.2 . Goal and Objectives

The goal of this study is to evaluate the impact of demand-responsive server allocation on customer waiting times and overall service system performance. By developing a flexible simulation model, this project allows the restaurant to adapt staffing based on the flow of customer orders, thus minimizing bottlenecks and ensuring a more balanced utilization of available resources.

Objectives:

- To analyse sojourn times (the total time a customer spends in the system) as a measure of overall service quality and efficiency.
- To evaluate customer waiting times for pizzas and dishes under dynamic server allocation, assessing whether flexible staffing reduces delays effectively.

2. System Overview

This project focuses on simulating and analyzing the operations of a restaurant service system where customers can order pizzas, dishes, or both. The system is designed to handle two main types of orders—pizza-only and dishes-only, with additional handling for customers who order both pizza and dishes together. A key feature of this system is its dynamic server reallocation, which allows servers to switch between handling pizza and dishes based on real-time demand.

2.1 Queuing System classification

This system can be classified as a Flexible M/M/c Queuing System with Dynamic Server Allocation.

1. **Arrival Process (M)**:
   The system assumes that customer arrivals for pizza, dishes, and combined orders follow a **Poisson process**. This is typical for many real-world applications where customer arrivals are random. The arrival rates (lambda values) for each order type are used to simulate these Poisson arrivals.
2. **Service Process (M)**:
   The service times for pizza and dish orders follow an exponential distribution, as shown by the use of exprnd(1 / mu_pizza) and exprnd(1 / mu_dishes) in the code. Exponential service times are typical in Markovian queuing models, which exhibit a "memoryless" property; this means that the duration of service for a customer is independent of the times taken for previous customers.
3. **Number of Servers (c)**:
   The restaurant initially allocates a set number of servers to pizza and dishes orders. However, this allocation is **dynamic** rather than fixed. For example, if there are initially two pizza servers and three dishes servers, a server from the dishes side can temporarily join the pizza side if pizza demand rises, and vice versa. This dynamic adjustment based on demand represents a flexible server structure rather than the fixed number of servers seen in traditional M/M/c systems.

Simulation Analysis:

DIMENSIONING AND OPTIMISATION

Dimensioning:

Here we determine the average Sojourn time of the customers in the system.

- The simulation time considered is 600 minutes, which is equivalent to a day activity at the restaurant.
- The arrival rate is kept constant, while the service rates are varied such that the load is less than one (underloaded) and greater than one (overloaded) and observed for different number of servers.

Optimisation:

Here we determine the average Sojourn time of the customers in the system.

- The simulation time considered is 600 minutes, which is equivalent to a day activity at the restaurant.
- The arrival rates of customers are varied, while the service rates are kept constant such that the load is less than one (underloaded) and greater than one (overloaded) and observed for different number of servers.

Dimensioning Results

C = 3, 1(pizza), 2 (dishes)

Lambda	Mu	Sojourn times (Pizza, Dishes)	Load (rho)
3	2	30.66, 20.62	1.50
3	3	18.27, 11.11	1.00
3	4	3.81, 0.63	0.75
3	5	0.92, 0.49	0.60
3	6	0.59, 0.27	0.50

C = 4, 1 (pizza), 3 (dishes)

Lambda	Mu	Sojourn time	Load (rho)
3	2	18.69, 11.93	1.50
3	3	5.87, 1.77	1.00
3	4	0.83, 0.52	0.75
3	5	0.42, 0.24	0.60
3	6	0.31, 0.17	0.50

C = 5, 2 (pizza), 3 (dishes)

Lambda	Mu	Sojourn time	Load (rho)
3	2	12.55, 5.12	1.50
3	3	1.17, 0.75	1.00
3	4	0.35, 0.29	0.75
3	5	0.28, 0.21	0.60
3	6	0.25, 0.15	0.50

Optimisation Results

C = 3, 1 (pizza), 2 (dishes)

Lambda	Mu	Sojourn time	Load (rho)
3	6	0.62, 0.39	0.50
4	6	1.65, 0.71	0.75
5	6	13.68, 3.89	0.83
6	6	15.86, 11.11	1.00
7	6	25.87, 11.95	1.16

C = 4, 1 (pizza), 3 (dishes)

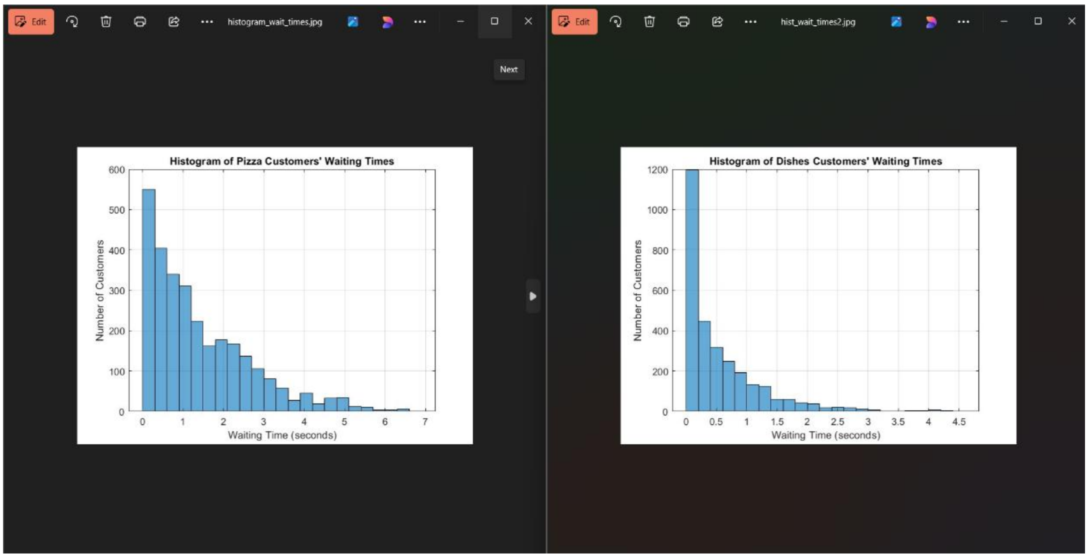
Lambda	Mu	Sojourn time	Load (rho)
3	6	0.32, 0.19	0.50
4	6	0.50, 0.30	0.75
5	6	1.50, 0.61	0.83
6	6	10.03, 0.77	1.00
7	6	14.71, 3.03	1.16

C = 5, 2(pizza), 3 (dishes)

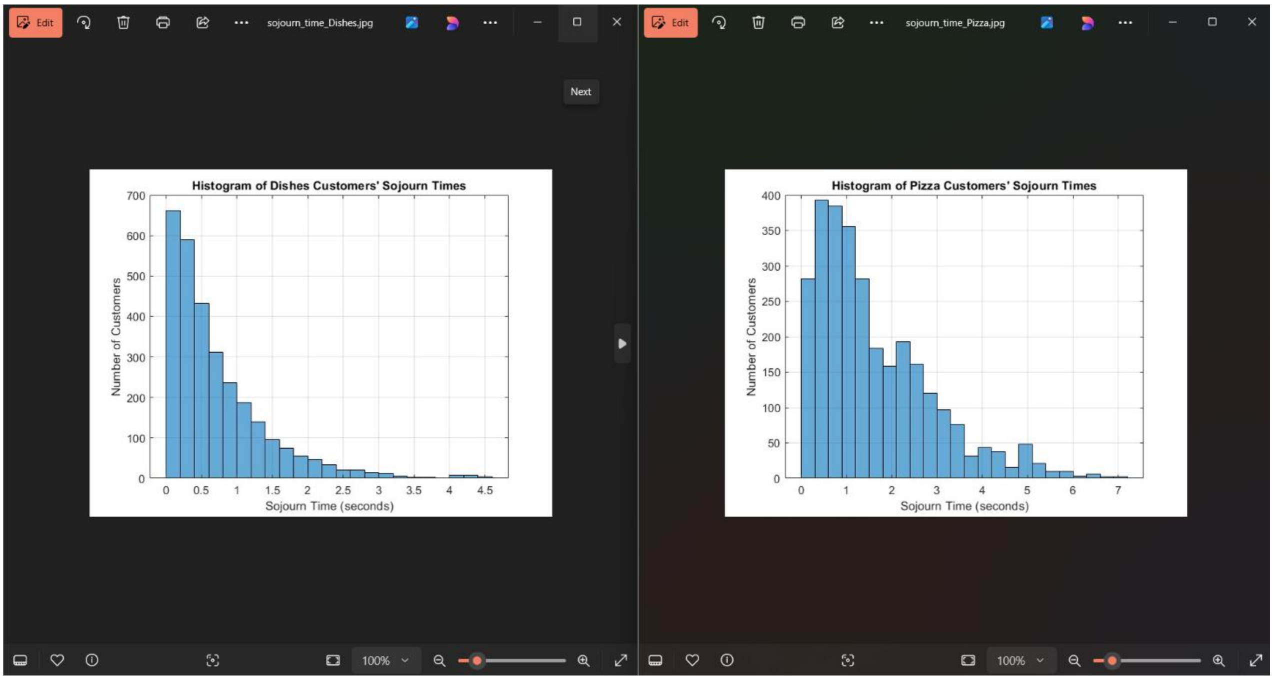
Lambda	Mu	Sojourn time	Load (rho)
3	6	0.20, 0.13	0.50
4	6	0.23, 0.16	0.75
5	6	0.33, 0.24	0.83
6	6	0.58, 0.59	1.00
7	6	6.76, 0.83	1.16

Graphs:

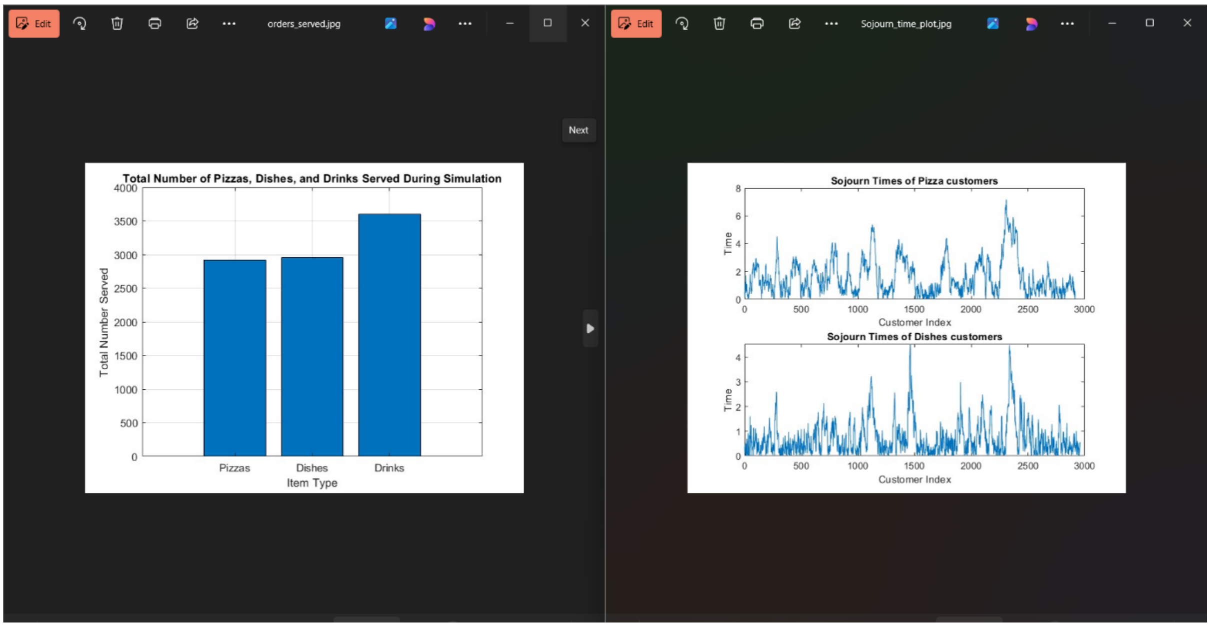
Case 1 : 50 percent load



These short waiting times reflect effective service rates and server allocation at 50% load, where the system handles demand well. Peaks in waiting time likely occur when server reallocation delays service.

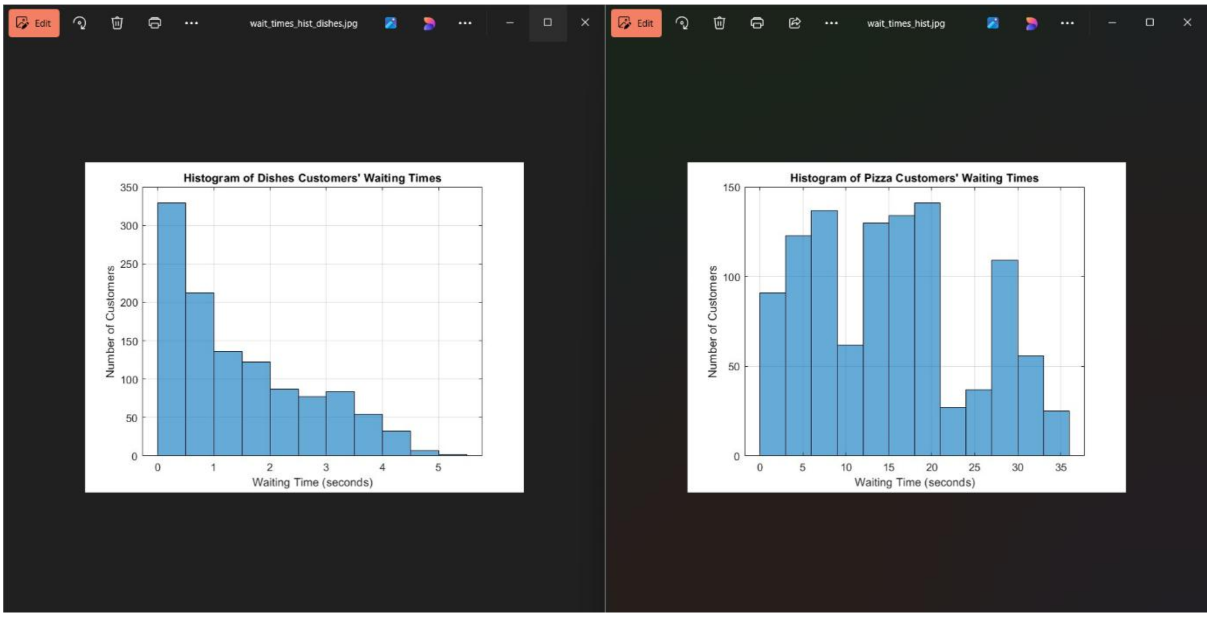


Sojourn time trends are indicative of the system's adaptability. While most customers spend minimal time, some face longer durations, possibly due to temporary demand surges.



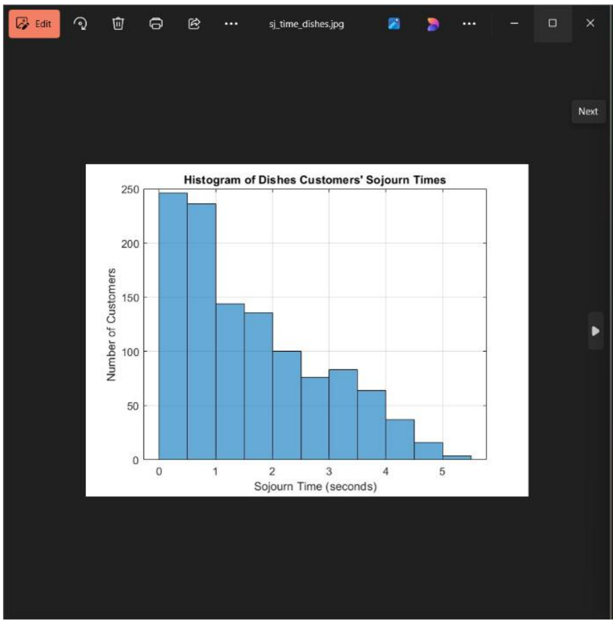
At 50% load, the restaurant system is highly efficient, with most customers experiencing minimal waiting and sojourn times. This efficiency reflects the model's effective design, balancing customer demand and server reallocation to reduce congestion. The server reallocation mechanism plays a critical role in optimizing wait and service times, particularly during demand fluctuations. This dynamic adjustment ensures that resources are allocated where needed most, preventing significant delays.

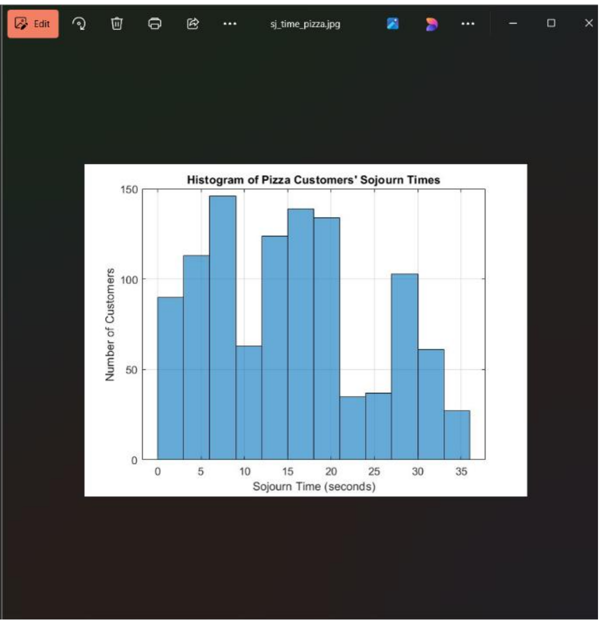
Case 2: 75 percent load



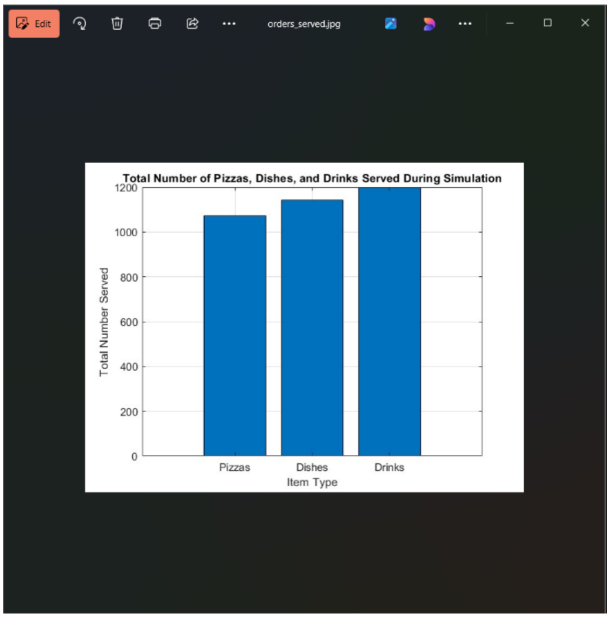
The histogram of waiting times for pizza customers shows a broader range compared to the 50% load, with waiting times now extending up to 35 seconds. This increase suggests that as the load approaches higher levels, more customers experience longer waits, likely due to periods of server congestion or delays in reallocation.

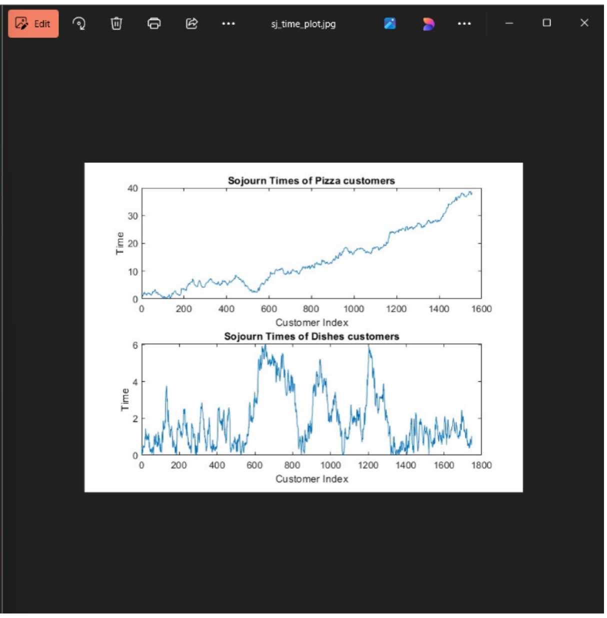
Waiting times for dishes still peak around 0–1 second, indicating many customers are served quickly. However, a noticeable tail in the distribution shows some customers waiting up to 5 seconds, longer than at 50% load. This distribution suggests that dishes still benefit from the initial larger server allocation but are now experiencing some delays as well.





The system shows resilience, particularly in handling dishes orders, yet the increased sojourn times for pizza customers illustrate the effect of high load on customer experience, where more customers face extended durations from order to fulfilment.

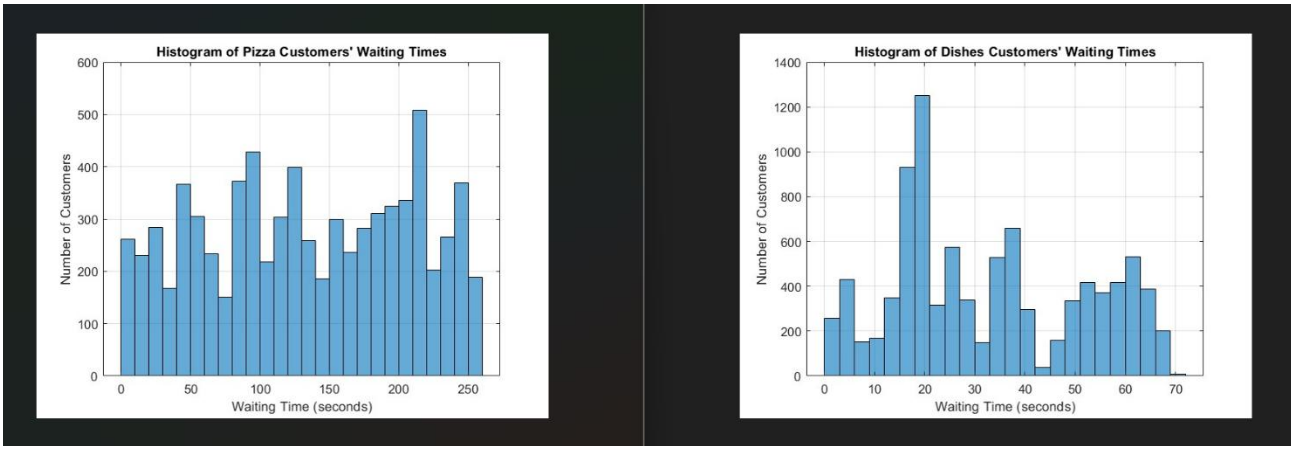




The server reallocation mechanism is evidently effective at managing up to 75% load but shows limitations under this higher demand. The increased waiting and sojourn times for pizzas reflect the impact of demand on the system, while dishes remain comparatively efficient due to their initial allocation and shorter preparation requirements.

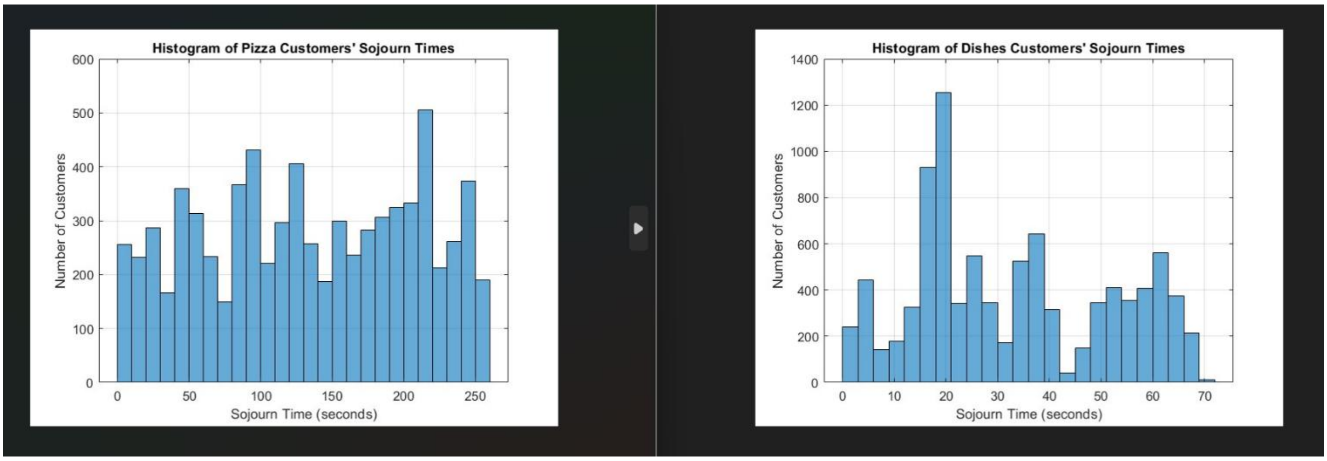
To enhance performance at higher loads, the system could benefit from a re-tuned allocation, perhaps increasing initial servers for pizzas or implementing faster reallocation to avoid the accumulation of delays in sojourn times.

Case 3: 83 percent load

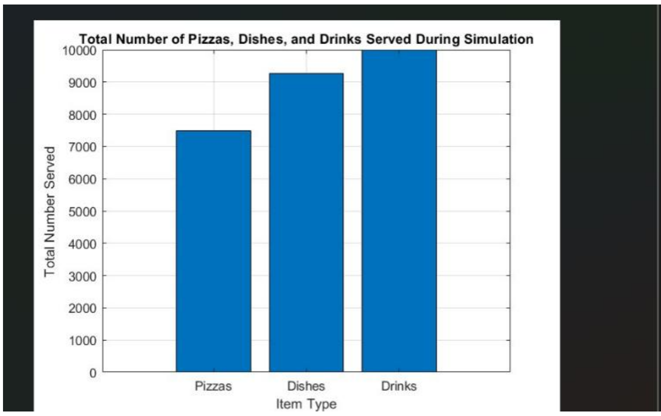


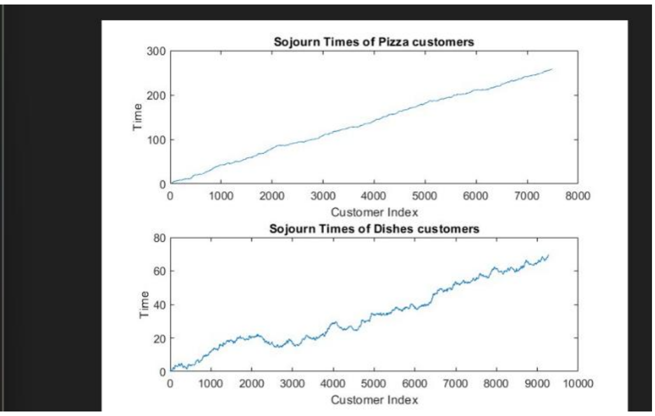
The histogram for pizza waiting times shows a substantial increase, with waiting times spread up to around 250 seconds. This distribution indicates that under the high load, many pizza customers face extended waits, likely due to limited server availability for pizzas or delayed server reallocation in response to growing queues.

Dishes waiting times have also increased, with a broad range extending up to around 70 seconds. However, the concentration around shorter waiting times compared to pizzas shows that dishes are still served relatively faster.



The high sojourn times for pizzas are symptomatic of the system reaching its operational limits. Dishes experience delays, but they are comparatively less affected, likely due to their faster service time and initial server advantage.





The restaurant service system at 83% load demonstrates signs of reaching its operational threshold. Both waiting and sojourn times are significantly higher across the board, with pizzas facing particularly prolonged delays. This scenario reveals the limitations of the current dynamic reallocation approach under near-peak load conditions and suggests that further scaling measures are necessary for optimal performance.

CONCLUSION:

In conclusion, the dynamic server reallocation system provides flexibility and efficiency under moderate demand but requires further optimization for high-load scenarios. The analysis reveals that while the system is effective at balancing resources and minimizing wait times at lower loads, higher loads expose limitations that affect customer wait times, particularly for pizzas. Enhancing the server configuration, optimizing reallocation strategies, and considering load-specific adjustments can improve the system's robustness and scalability, ensuring a smoother customer experience even under peak demand conditions.