July 19, 2024

1 Customer Priority in Queuing Systems

1.1 Goal

To explore the impact of customer priority rules on the performance of queueing systems. Two scenarios are compared: one where high-priority customers preempt low-priority customers (Preempt/Resume) and another where they do not (Priority without Preemption). I investigate the mean delay experienced by different customer types and the overall average queue length under various traffic intensities.

1.2 Methodology

Simulation Framework: The Python-based simulation models customer arrivals with two separate exponential distributions and processes departures from the system. In the Preempt/Resume model, high-priority customers interrupt low-priority service, which resumes once the high-priority customer has been served. The Priority without Preemption model serves customers in a first-come-first-served manner within their priority class.

Experimental Setup:

- Arrival rates λ_1 and λ_2 for high and low-priority customers, respectively.
- Service rate μ.
- The total number of customers served before recording system performance was 3000 with a warm-up period of 800 customers.
- Each scenario was simulated 30 times to average out random variability.

1.3 Results and Observations

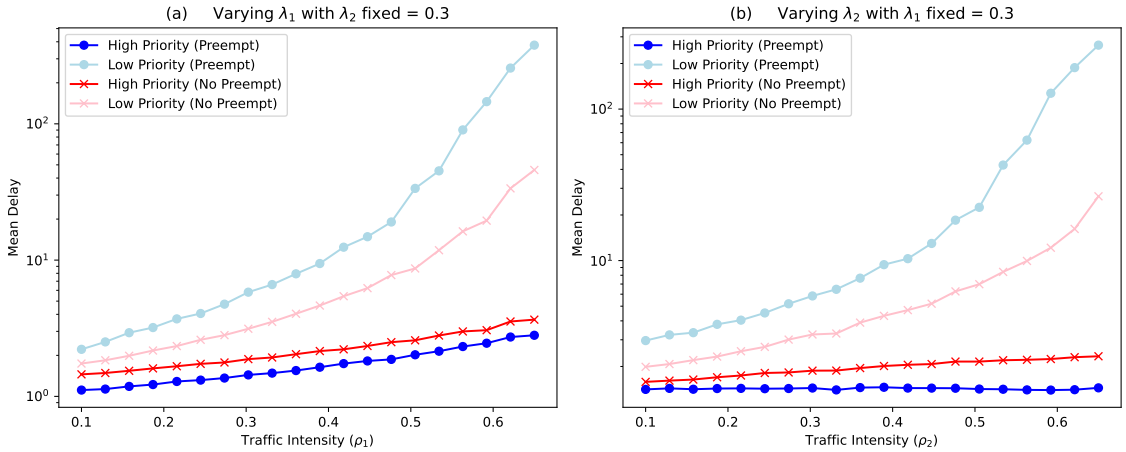

Figure 1: Comparison of mean delay for high and low-priority customers under varying traffic intensities for both Preempt/Resume and Priority without Preemption scenarios.

Mean Delay Analysis: The simulation results depicted in Figure 2 illustrate the system's mean delay for high and low-priority jobs. Figure 2(a) presents the scenario where the arrival rate for low-priority jobs is fixed, clearly demonstrating that the mean delay for high-priority jobs is significantly lower under the preemption scheme compared to low-priority jobs. Additionally, due to preemption, low-priority jobs experience extended delays, resulting in a higher mean delay when preempted than when not preempted. A same trend is observed when varying the arrival rate or traffic intensity of low-priority jobs, with a service rate set at 1, as shown in Figure 2(b).

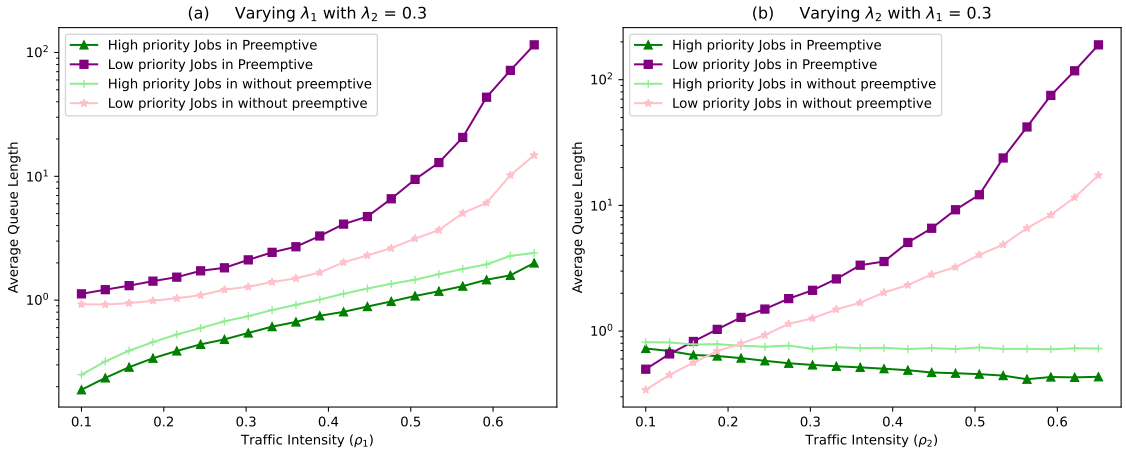

Figure 2: Average queue length comparison highlighting the effects of priority rules on the buildup of customers in the system.

Queue Length Dynamics: The behavior of queue lengths (Figure 3) follows a similar pattern to that of the mean delay. In Figure 3(a) and 3(b), the Priority with Preemption model, queue lengths for low-priority jobs experience a more noticeable increase. Conversely, high-priority jobs in the preemptive scheme endure lower wait times, as they must enter service without any wait if low priority job is running. These dynamics persist regardless of whether the arrival rates for high or low-priority jobs are held constant and the other varied. Notably, one slight deviation is observed in Figure 3(b) when the high-priority job arrival rate is fixed: the average queue length tends to decrease as traffic intensity for high-priority jobs increases.

1.4 Conclusions

This study offers a comprehensive examination of the influence of customer priority rules on the performance of queuing systems. It reveals that although Preempt/Resume effectively addresses the urgency of high-priority jobs, it consequently increases the wait for low-priority jobs. System designers are thus faced with the critical task of balancing swift service for high-priority tasks against the increased latency for low-priority ones.

One should align priority strategies with specific operational needs and customer expectations. Regular adjustments based on customer arrival patterns and service demands are essential.

Recommendations based on the simulations: 1. Deploy the Preempt/Resume strategy in environments where the immediacy of high-priority job completion is essential and the system's overall efficiency is a secondary concern. 2. Adopt Priority without Preemption when fairness is pivotal, and the increased waiting time for high-priority jobs does not drastically affect the outcome. 3. Regularly review and adjust customer arrival patterns and service requirements, as this will help refine priority rules to meet evolving service demands and expectations.