Weicong Feng

Professor Zhanyang Zhang

CSC 86120

9/14/2023

Lab2 – Queueing Theory and Its Applications

Objective: Through this lab exercise, students should learn: (1) how to apply queueing theory to model a simple web application system; (2) how to use fundamental laws to investigate the relations of system quantities and gain insights about the system performance; (3) how to design and implement a simulation study based on the queueing model; and (4) how to analyze simulation results statistically and comparing them to analytical results derived from queueing theory.

Part A: Consider a system modeled as shown in the figure below. A web request send to web server must wait in queue if the server is busy, otherwise it will begin the processing at the server. When the server finishes processing the request, it sends a response to the user. A user will think for a while before generating next request.



Fig. L2 - a model of a simple web application system.

(a) If there are 100 active users each with 20 seconds average think time (IID). The average system

response time (IID) is 10 seconds which is the sum of queueing time and processing time. What is the throughput of the web server (responses/second)?

(b) If average queueing time is 9.75 second, what is the average processing time? What is the web server utilization (busy time/total time)?

(c) Based on your study, without increasing system response time, what is the maximum number of active users the system can support?

Provide analytical solutions to questions (a) to (c) using queueing theory. Show the calculation steps and which fundament laws you applied.

Ans:

1. The information we have:

- There are 100 active users.

- Each user has an average think time of 20 seconds

- The average system response time (W) is 10 seconds.

So, requests per second

Then,

So, on average, there are 50 requests in the system at any given time.

To calculate the throughput (responses per second), we need to consider that the system is not overloaded. Therefore, the throughput is equal to the arrival rate:

Throughput X = responses per second.

So, the throughput of the web server is 5 responses per second.

We know that the average system response time is 10 seconds, including queueing time and processing time. If the average queueing time is 9.75 seconds, then the average processing time S = W – d = 10 – 9.75 = 0.25 seconds.

During 20 seconds, all 100 activate users make requests, and the average system response time is 10 seconds, means that during total 30 seconds, all 100 activated users send once requests and get responses. So, U = 100\*0.25/30 = 0.833 = 83.3%.

Without increasing system response time, the maximum number of activate users the system can support means to figure out the maximum number so that . So, N\*0.25/30 = 100%, N = 1\*30/0.25 = 120 activate users.