Queueing Theory

LA Math Circle
High School II
Dillon Zhi

November 1, 2015

Consider the "M/M/1/K" queue: The system has one server and has a capacity of K customers. (That is, at any time, at most one customer is being served and at most $K-1$ customers are waiting.) The customers arrive at the rate λ and the server serves customers at the rate μ.

1. Find the probabilities $P_0, P_1, ..., P_K$, where P_j is the probability that the system has j customers in the "steady state". Express your answer in terms of K and $a = \lambda/\mu$, the ratio of the arrival rate to the service rate. The parameter a is often called the **offered load**.

2. **Utilization**: What fraction of the time is the server actually being used?

    a. Find a formula for the utilization in terms of K and a.

    b. What is the formula for utilization when $a = \frac{1}{2}$? Now plot a graph of utilization versus capacity for $a = \frac{1}{2}$. What is the utilization like for large K?

c. Plot a graph of utilization versus offered load for $K = 1$. Also plot a graph of utilization versus offered load for $K = 3$.

d. What happens as $K \to \infty$? Plot a graph of utilization versus offered load when the capacity for customers is infinite ("$K = \infty$").

3. **Loss probability**: What fraction of customers arrive to find that the queue is full?

   a. Find a formula for the loss probability in terms of K and a.

   b. Plot a graph of loss probability versus capacity for $a = \frac{1}{2}$.

   c. How much capacity do you need if you don't want any customers to be lost?

   d. Plot a graph of loss probability versus offered load for $K = 3$.

4. **Mean number of customers**: On average, how many customers are in the system?

   a. Show that the mean number of customers in the system is

   $$\frac{a}{1-a} + \frac{K+1}{1-a^{K+1}} a^{K+1}$$

   if $a \neq 1$, and the mean number of customers in the system is $K/2$ if $a = 1$.

b. Plot a graph of the mean number of customers versus offered load for $K = 3$.

c. Suppose the system has infinite capacity. What is the formula for the mean number of customers in terms of a? Plot a graph of this relationship.

5. Now consider a system with s servers and a capacity of K customers. (That is, at any time, at most s customesr are being served and at most $K-s$ customers are waiting.) The servers share a queue. The customers arrive at the rate λ and *each* server serves its customers at the rate μ.

   Find the probabilities $P_0, P_1, ..., P_K$, where P_j is the probability that the system has j customers in the "steady state".

6. Which of these systems is "best"?

   - s servers with each server serving at the rate μ, sharing a queue
   - s servers with each server serving at the rate μ, each with its own separate queue
   - one fast server serving at the rate $s\mu$