Stochastic Processes

Lecture 22: queueing networks

Hailun Zhang@SDS of CUHK-Shenzhen

April 19, 2021

$M/M/1$ queue

- $X(t) =$ number of customers in system at time t.
- When $\lambda < \mu$, the stationary distribution is given by

$$\pi_i = (1-\rho)\rho^i, \quad i = 0, 1, 2, \ldots,$$

where

$$\rho = \lambda/\mu.$$

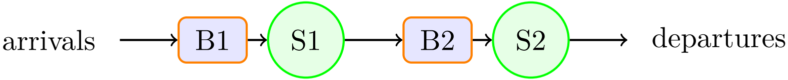
- Average waiting time per customer is equal to

$$W_q = m\frac{\rho}{1-\rho},$$

where $m = 1/\mu$ is the mean service times. Non-linear in load ρ.

2-Station Tandem Queue

One M/M/1 queue feeds another queue:



“Station” k = buffer k + server k

Some questions:

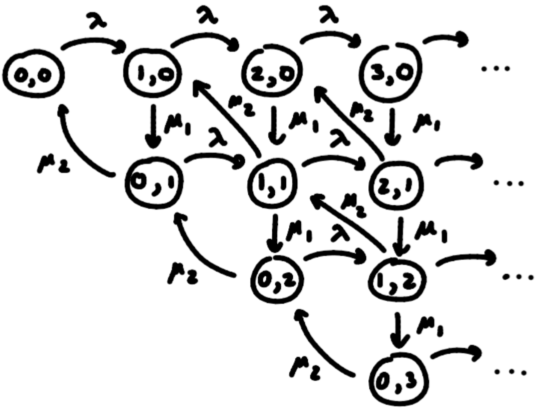
- Utilization of each server, assuming $\lambda = 2$, $m_1 = 0.45$, $m_2 = .40$?
- Throughput of the system?
- Fraction of time there are, respectively, 2 and 3 customers at the upstream and downstream stations.

Stationary Distribution

State $= (i_1, i_2)$

- i_k = number of customers at station k

Rate Diagram:



Stationary Distribution

Do we really have to solve the balance equations?

Alternative: Guess & Check!

Recall that

$$\text{CTMC is irreducible} \implies \text{has at most one stationary distribution}$$

So, if we can find a π such that

$$\pi G = 0 \qquad \text{and} \qquad \sum_i \pi_i = 1,$$

this π is the *unique* stationary distribution.

How can we come up with a guess for π?

Stationary Distribution

The 2-station tandem queue can be viewed as two independent M/M/1 queues. (Really?)

Theorem

The stationary distribution π of the 2-station tandem queue is given by

$$\begin{aligned}\pi_{(i_1,i_2)} &= \left(1-\frac{\lambda}{\mu_1}\right)\left(\frac{\lambda}{\mu_1}\right)^{i_1}\left(1-\frac{\lambda}{\mu_2}\right)\left(\frac{\lambda}{\mu_2}\right)^{i_2} \\ &= (1-\rho_1)\rho_1^{i_1}(1-\rho_2)\rho_2^{i_2}\end{aligned}$$

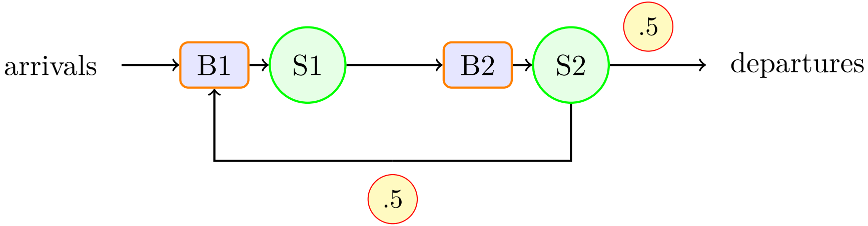
for $i_1, i_2 \in \{0, 1, \ldots\}$, $\rho_k = \lambda/\mu_k$, $k = 1, 2$.

Proof. Check that it satisfies "rate out = rate in" and sums to 1!

$$(\lambda+\mu_1+\mu_2)\pi(2,3) = \lambda\pi(1,3) + \mu_2\pi(2,4) + \mu_1\pi(3,2)$$

Exercise for the reader ☺.

Two queues with feedback



- External Poisson arrival process with rate $\alpha = 1$ job/minute.
- $m_1 = .45$ minutes, and $m_2 = 0.40$ minutes.
- Each job leaving station 2 has probability $p = 50\%$ going back to station 1.
- Find

$$\mathbb{P}\{X_1(\infty) = 2, X_2(\infty) = 3\}$$

- What is the utilization of server 1?

Traffic equation

- Assume that system is stable. $\lambda_2 = \lambda_1$,

$$\lambda_1 = \alpha + 0.5\lambda_2 \quad \Rightarrow \quad \lambda_1 = \alpha + 0.5\lambda_1 \quad \Rightarrow \quad \lambda_1 = 2\alpha = 2.$$

Therefore,

$$\rho_1 = \lambda_1 m_1 = 2(0.45) = 90\%, \quad \rho_2 = \lambda_2 m_2 = 2(0.40) = 80\%.$$

- Average number of jobs in the system

$$\begin{aligned}
&= \text{average \# of jobs in station 1} + \text{average \# of jobs in station 2} \\
&= \frac{\rho_1}{1-\rho_1} + \frac{\rho_2}{1-\rho_2} = \frac{0.9}{1-0.9} + \frac{0.8}{1-0.8} = 9 + 4 = 13 \text{ jobs.}
\end{aligned}$$

- What is the average time in system per job? Use Little's Law

$$L = \lambda W \quad \Rightarrow \quad 13 = 1W \quad \Rightarrow \quad W = 13 \text{ minutes.}$$

Reducing the failure rate

- From 50% to 40%

$$\rho_1 = \frac{5}{3}(0.45) = (0.15)5 = 0.75 = \frac{3}{4}, \quad \rho_2 = \frac{5}{3}(0.4) = \frac{2}{3}$$
$$W = \frac{\rho_1}{1-\rho_1} + \frac{\rho_2}{1-\rho_2} = \frac{3/4}{1/4} + \frac{2/3}{1/3} = 3 + 2 = 5 \text{ minutes.}$$

- Just with 10% of decrease of the failure rate, the lead time drops more than a half.

Open Jackson Network

J stations

For each station $j \in \{1, \ldots, J\}$:

- external arrivals $\sim \mathrm{PP}(\alpha_j)$
- n_j server; iid exponential service times with rate μ_j at station j
- unlimited waiting room
- When a job completes service, it is either
  - routed to station k with probability P_{jk}, or
  - exits the system with probability

$$P_{j0} = 1 - \sum_{k=1}^{J} P_{jk}$$

- The routing matrix P is transient or $(I - P)$ is invertible.

A three-station open Jackson queueing network

- external arrival to station 1 with rate α.
- $P_{21} = .3$ and $P_{32} = .2$

$$\lambda_1 = \alpha + .3\lambda_2$$
$$\lambda_2 = \lambda_1 + .2\lambda_3$$
$$\lambda_3 = .7\lambda_2$$

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & .3 & 0 \\ 1 & 0 & .2 \\ 0 & .7 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}$$

- $\rho_k = \lambda_k m_k < 1$, $k = 1, 2, 3$
- $\pi(n_1, n_2, n_3) = (1 - \rho_1)\rho_1^{n_1}(1 - \rho_2)\rho_2^{n_2}(1 - \rho_3)\rho_3^{n_3}$.

Closed Jackson Network

J stations

For each station $j \in \{1, \ldots, J\}$:

- No external arrivals; the number of jobs in system N remains fixed
- n_j server; iid exponential service times with rate μ_j at station j
- unlimited waiting room
- When a job completes service, it is either
  - routed to station k with probability P_{jk}, or
  - exits the system with probability

$$0 = 1 - \sum_{k=1}^{J} P_{jk}$$

- The routing matrix P is stochastic.

A two-station closed network

- $J = 2$; two-station closed network

$$\mu_1 = 1 \text{ job per minute}, \quad \mu_2 = 1/2 \text{ job per minute}.$$

- $N = 2$, state space

$$S = \Big\{(2,0), \quad (1,1), \quad (0,2)\Big\}.$$

- Stationary distribution

$$\pi(2,0) = \frac{1}{7}, \quad \pi(1,1) = \frac{2}{7}, \quad \pi(0,2) = \frac{4}{7}. \tag{1}$$

- Server 2 utilization:

$$U_2 = \pi(1,1) + \pi(0,2) = \frac{6}{7}.$$

- Throughput: rate at which the completed jobs leaves the production system;

$$U_2 \times \mu_2 = \frac{6}{7}\Big(\frac{1}{2}\Big) = \frac{3}{7}.$$

A two-station closed network

- $J = 2$; two-station closed network
- $N = 2$
- Average time in the production system per job:
  Little's law

$$L = \lambda W.$$

- $L = 2$, $\lambda = \frac{3}{7}$,

$$W = 2/(3/7) = \frac{14}{3} \text{ minutes.}$$