Chapter 12
Introduction to Queueing Theory

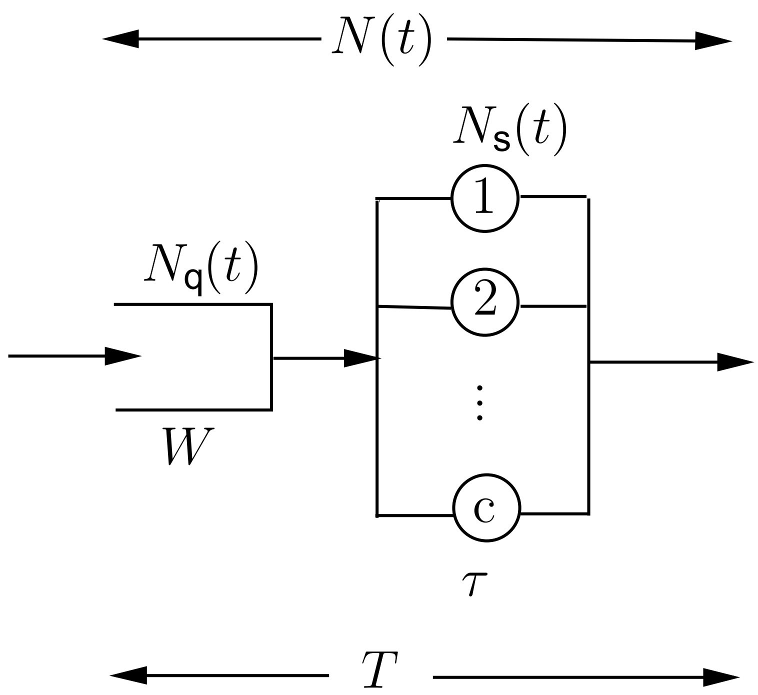
ENCS6161 - Probability and Stochastic Processes

Concordia University

Elements of a Queueing System

- A queueing system is defined by $a/b/m/k$, where
  - a: type of arrival process.
    $$a = M \quad \text{Poisson Process}$$
  - b: service time distribution.
    $$b = \begin{cases} M & \text{exponential} \\ D & \text{deterministic} \\ G & \text{general} \end{cases}$$
  - m: number of servers
  - k: max number of customers allowed in the system

Little's Formula



$T = W + \tau$

W : waiting time

τ : service time

Little’s Formula

- Little’s Formula: $E[N] = \lambda E[T]$, where
  - $E[N]$: average number of customers in the system
  - λ: arrival rate
  - $E[T]$: average time that a customer stays in the system
- Little’s Formula is very general. The system here could be the whole queueing system, the waiting buffer, or the seervice system, etc.
- Read the proof on your own.

$M/M/1$ or $M/M/1/1$ Queue

- Let $N(t)$ be the number of customers.
- Poisson arrival. In a small period δ, probability of one arrival is

$$P[A(\delta) = 1] = \frac{\lambda\delta}{1!} e^{-\lambda\delta} = \lambda\delta + o(\delta)$$

$$P[A(\delta) \geq 2] = o(\delta)$$

$$\Rightarrow \quad p_{n,n+1}(\delta) = P[A(\delta) = 1] = \lambda\delta + o(\delta)$$

$$\Rightarrow \quad r_{n,n+1} = \lambda$$

$$r_{n,n+2} = 0, \; r_{n,n+3} = 0, \; \cdots$$

$M/M/1$ or $M/M/1/1$ Queue

- Exponential service time. So the probability that there is one departure in a small period δ is

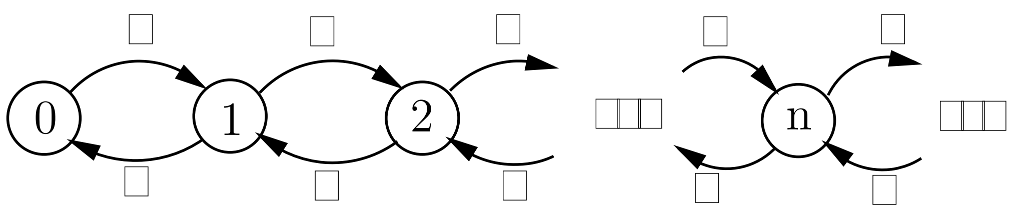
$$P[\tau < \delta] = 1 - e^{-\mu\delta} = \mu\delta + o(\delta)$$

$$\Rightarrow \quad p_{n,n-1}(\delta) = \mu\delta + o(\delta)$$

$$\Rightarrow \quad r_{n,n-1} = \mu \quad (n \geq 1)$$

$$r_{n,n-2} = 0, \; r_{n,n-3} = 0, \; \cdots$$

- State transition diagram



$M/M/1$ or $M/M/1/1$ Queue

- Global Balance Equations are

$$\lambda P_{j-1} = \mu P_j \quad \text{for all } j = 1, 2, \cdots$$

$$\Rightarrow \quad P_j = \frac{\lambda}{\mu} P_{j-1}$$

Let $\rho = \frac{\lambda}{\mu}$, then

$$P_j = \rho P_{j-1} = \rho^2 P_{j-2} = \cdots = \rho^j P_0$$

$$\text{Since} \quad \sum_j P_j = 1 \quad \Rightarrow \quad \sum_j \rho^j P_0 = 1$$

$$\text{i.e.} \quad \frac{1}{1-\rho} P_0 = 1 \quad \Rightarrow \quad P_0 = 1 - \rho$$

$$P_j = (1-\rho)\rho^j \quad j = 0, 1, 2, \cdots$$

Note: we need $\rho = \frac{\lambda}{\mu} < 1$ for stability.

$M/M/1$ or $M/M/1/1$ Queue

- The expected number of customers

$$E[N] = \sum_{j} jP_j = \sum_{j} j(1-\rho)\rho^j = \frac{\rho}{1-\rho}$$

From Little's Formula

$$E[T] = \frac{E[N]}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu-\lambda}$$

Expected waiting time in the queue is

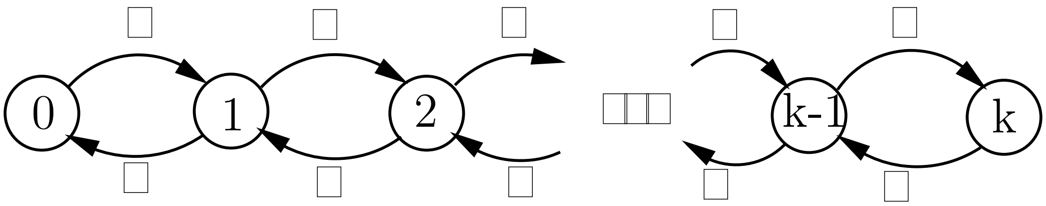
$$E[W] = E[T] - E[\tau] = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)}$$

Expected number in the queue is

$$E[N_q] = \lambda E[W] = \frac{\rho^2}{1-\rho}$$

Server utilization is $1 - P_0 = 1 - (1-\rho) = \rho = \frac{□}{□}$

$M/M/1/K$ Queue



- Similarly to $M/M/1$ queue, we have

$$\lambda P_{j-1} = \mu P_j \Rightarrow P_j = \rho P_{j-1} \text{ for } j = 1, 2, \cdots, K$$

$$\text{so} \quad P_j = \rho^j P_0, \quad j = 0, 1, 2, \cdots, K$$

$$\text{From} \quad \sum_{j=0}^{K} P_j = 1 \quad \Rightarrow \quad P_0 = \frac{1-\rho}{1-\rho^{K+1}}$$

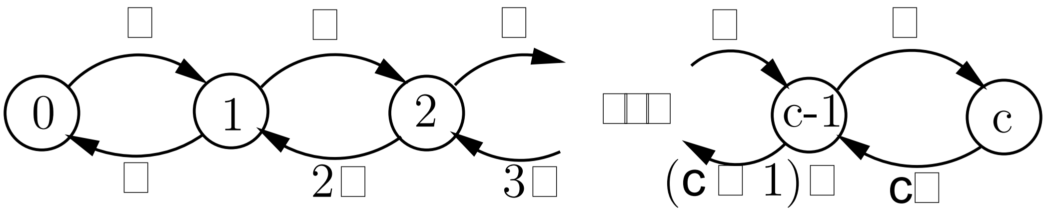
$$\Rightarrow \quad P_j = \frac{(1-\rho)\rho^j}{1-\rho^{K+1}}$$

$M/M/1/K$ Queue

- Blocking Probability: prob. of rejecting a customer

$$P[N = K] = P_K = \frac{(1 - \rho)\rho^K}{1 - \rho^{K+1}}$$

$M/M/c/c$ Queue



$$\lambda P_{j-1} = j\mu P_j \qquad j = 1, 2, \cdots, c$$

$$\Rightarrow \qquad P_j = \frac{1}{j}\left(\frac{\lambda}{\mu}\right) P_{j-1} \quad \text{let } \alpha = \frac{\lambda}{\mu}$$

Then $\quad P_j = \dfrac{\alpha^j}{j!} P_0 \qquad j = 1, 2, \cdots, c$

$$\sum_{j=0}^{c} P_j = \sum_{j=0}^{c} \frac{\alpha^j}{j!} P_0 = 1 \quad \Rightarrow \quad P_0 = \frac{1}{\sum_{j=0}^{c} \frac{\alpha^j}{j!}}$$

$M/M/c/c$ Queue

- The blocking probability

$$
\begin{aligned}
P_B &= P[N = c] = P_c \\
&= \frac{\frac{\alpha^c}{c!}}{1 + \alpha + \frac{\alpha^2}{2!} + \cdots + \frac{\alpha^c}{c!}}
\end{aligned}
$$

$M/M/c/c$ Queue

- Example: A company has 4 telephone lines with arrival rate of one call every two minutes. The average duration of a call is 4 minutes. What is the P_B?

$$\lambda = \frac{1}{2},\ \mu = \frac{1}{4} \quad) \quad \alpha = \frac{\lambda}{\mu} = 2$$

$$P_B = \frac{\frac{2^4}{4!}}{1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!}} = 0.095 \text{ or } 9.5\%$$