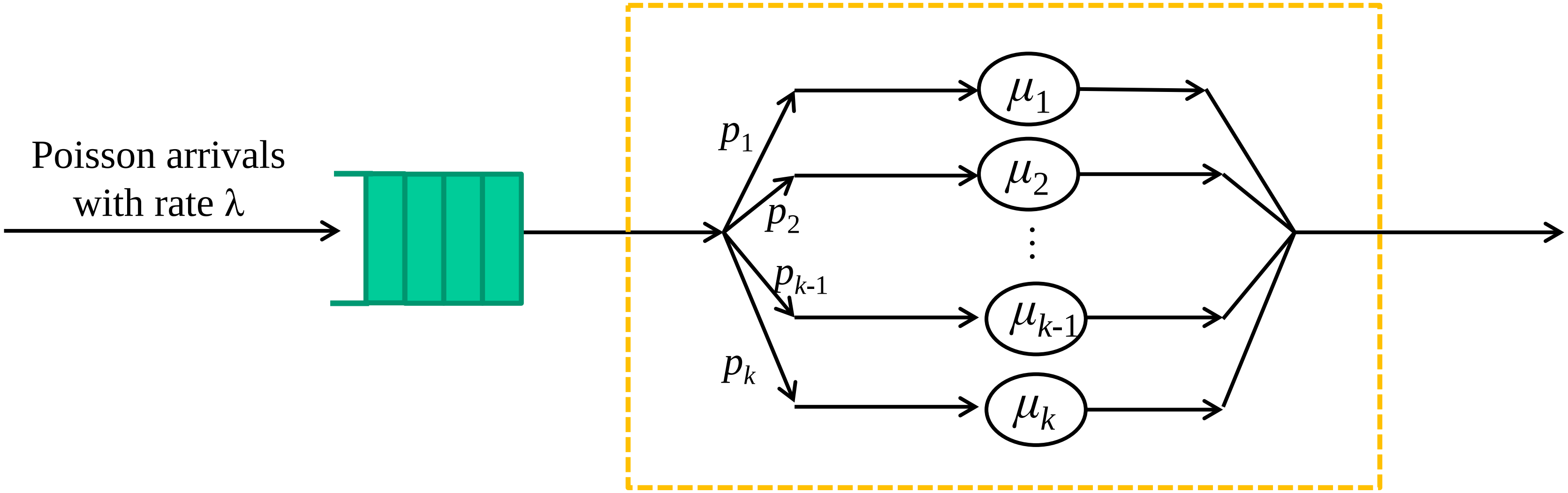
Queuing System III

Wha Sook Jeon

Mobile Computing & Communications Lab.

M/H$_k$/1 (1)

- A single server, infinite-capacity queueing system with hyperexponentially distributed service time of order k



A k-stage hyperexponential server

- Since M/H$_k$/1 is a M/G/1, we can obtain its performance measures, by using the performance measure equations of M/G/1

M/H$_k$/1 (2)

- Remind that $\bar{N} = \lambda E[S] + \frac{\lambda^2 E[S^2]}{2(1-\lambda E[S])}$ for M/G/1
- M/H$_k$/1
  - Probability density function of a *k*-stage hyperexponential random variable *S*: $f_S(x) = \sum_{i=1}^{k} p_i \mu_i e^{-\mu_i x}$
  - Laplace Transform of *S*

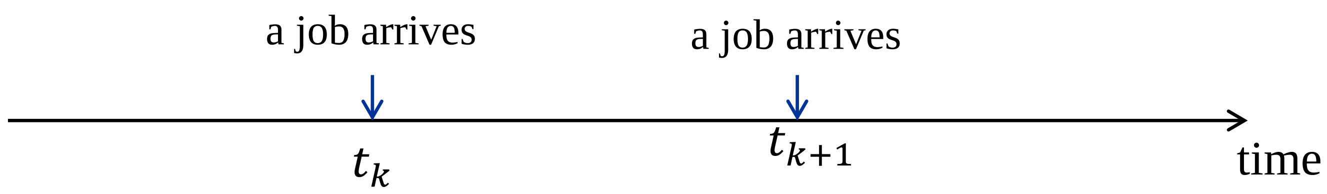
$$F^*(\theta) = E\left[e^{-\theta S}\right] = \int_0^\infty e^{-\theta x}\left(\sum_{i=1}^{k} p_i \mu_i e^{-\mu_i x}\right) dx$$

$$= \sum_{i=1}^{k} p_i \mu_i \int_0^\infty e^{-\theta x} e^{-\mu_i x} dx = \sum_{i=1}^{k} p_i \left(\frac{\mu_i}{\mu_i + \theta}\right)$$

  - $E[S] = -\lim_{\theta \to 0} \frac{dF^*(\theta)}{d\theta} = \sum_{i=1}^{k} p_i \frac{1}{\mu_i}$
  - $E[S^2] = \lim_{\theta \to 0} \frac{d^2 F^*(\theta)}{d\theta^2} = \sum_{i=1}^{k} p_i \frac{2}{\mu_i{}^2}$
  - $\bar{N} = \lambda E[S] + \frac{\lambda^2 E[S^2]}{2(1-\lambda E[S])}$, $\bar{T} = E[S] + \frac{\lambda E[S^2]}{2(1-\lambda E[S])}$

GI/M/1 (1)

  - General Independent Arrival Process
  - Exponential service time distribution
- Embedded Markov Chain Approach
  - We observe the system at an instant that a job arrives
  - Then, the time elapsed since the last arrival job does not need to be considered.
  - Thus, the system at arrival epochs has Markovian property



- Let N_k be a random variable representing the number of jobs in the system at an arrival epoch t_k
- Embedded Markov chain is described as $\{N_k\ , k = 1, 2, 3, \cdots\}$

GI/M/1 (2)

- State probability distribution in EMC
  - Let π_i be the limiting probability of i jobs in system at an arrival epoch (seen by an arrived job)

    $$\pi_i = \sum_j \pi_j \, p_{ji}$$

    - where $p_{ji} = Pr\{N_{k+1} = i \mid N_k = j\}$ and is the probability that $(j + 1 - i)$ jobs are served between two arrival epochs
  - We need P_{ji} ($P_{ji} = 0$ for all $i > j+1$ since $j + 1 - i \geq 0$)

    : P_{ji} depends on the number of departing jobs during an interarrival time
- Let r_m be a probability that m jobs are served during an inter-arrival time

  $$r_m = \int_0^{\infty} \frac{(\mu t)^m}{m!} e^{-\mu t} a(t) dt$$

  - $a(t)$: probability density function of random variable representing inter-arrival time

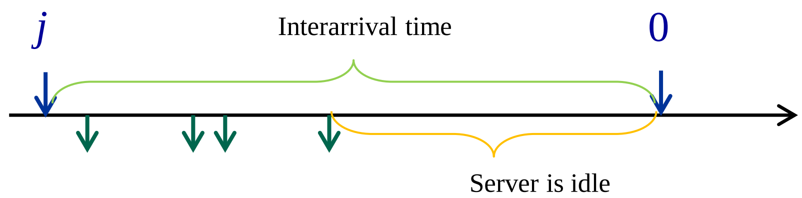
GI/M/1 (3)

- $\pi_i = \sum_{n=0}^{\infty} \pi_{i-1+n}\, r_n$ for $i > 0$

  $= \pi_{i-1}\, r_0 + \pi_i\, r_1 + \pi_{i+1}\, r_2 + \pi_{i+2}\, r_3 + \cdots$

- $\pi_0 = \sum_{j=0}^{\infty} \pi_j\, p_{j0}$,

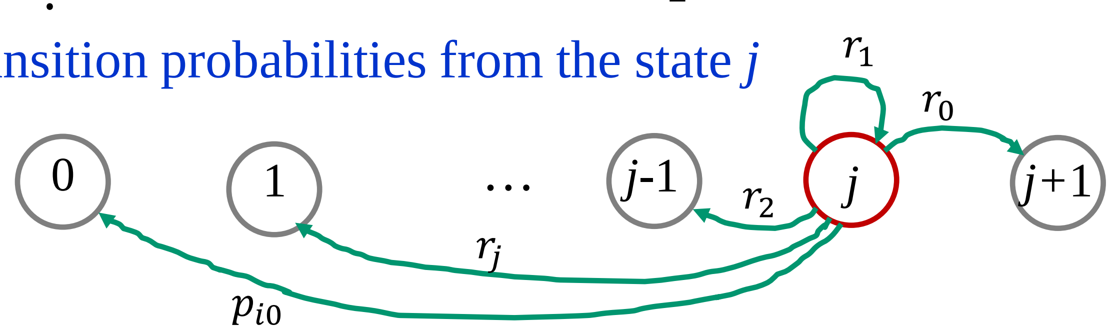
  where $p_{j0} = 1 - \sum_{n=0}^{j} r_n$



- One-step State Transition Matrix

$$
\mathrm{P} = \begin{bmatrix}
1 - r_0 & r_0 & 0 & 0 & 0 & \cdots \\
1 - \sum_{n=0}^{1} r_n & r_1 & r_0 & 0 & 0 & \cdots \\
1 - \sum_{n=0}^{2} r_n & r_2 & r_1 & r_0 & 0 & \cdots \\
1 - \sum_{n=0}^{3} r_n & r_3 & r_2 & r_1 & r_0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \cdots
\end{bmatrix}
$$

- Outgoing transition probabilities from the state j



GI/M/1 (4)

- We use the operator method to find the solution
  - ✓ Let $\pi_i = \beta\, \pi_{i-1}$.
  - ✓ Then, $\pi_i = \beta\, \pi_{i-1} = \beta^2\, \pi_{i-2} = \beta^3\, \pi_{i-3} = \cdots = \beta^i\, \pi_0$
  - ✓ $\pi_i = \pi_0\, \beta^i$ $(0 < \beta < 1)$
- From $\sum_{i=0}^{\infty} \pi_i = 1$, $\quad \pi_i = (1\text{-}\beta)\, \beta^i$
- We need β

  - $\pi_i = \sum_{n=0}^{\infty} \pi_{i-1+n}\, r_n \rightarrow (1\text{-}\beta)\beta^i = \sum_{n=0}^{\infty} (1-\beta)\beta^{i-1+n}\, r_n \rightarrow \beta = \sum_{n=0}^{\infty} \beta^n\, r_n$

- $\beta = \sum_{n=0}^{\infty} \beta^n \int_0^{\infty} \frac{(\mu t)^n}{n!} e^{-\mu t} a(t) dt$

$$= \int_0^{\infty} \sum_{n=0}^{\infty} \frac{(\beta \mu t)^n}{n!} e^{-\mu t}\, a(t) dt$$

$$= \int_0^{\infty} e^{-(\mu - \beta\mu)t} a(t) dt$$

$$= F_A^{\,*}(\mu - \beta\mu)$$

GI/M/1 (5)

- We should calculate β for an arrival process, using

$$\boxed{\beta = F_A^{*}(\mu - \beta\mu)}$$

- Example
  - Poisson Arrivals (M/M/1): $F_A^{*}(s) = \frac{\lambda}{\lambda+s}$
    - $\beta = \frac{\lambda}{\lambda+\mu-\beta\mu} \quad \Rightarrow \quad (\lambda + \mu - \beta\mu)\,\beta = \lambda \quad \Rightarrow \quad (\mu\beta - \lambda)(\beta - 1) = 0$
    - Since $0 < \beta < 1, \quad \beta = \frac{\lambda}{\mu}$
  - Erlang-k arrivals (E_k/M/1): $F_A^{*}(s) = \left(\frac{k\lambda}{k\lambda+s}\right)^k$
    - $\beta = \left(\frac{k\lambda}{k\lambda+\mu-\beta\mu}\right)^k$
    - When $k = 2$, $\beta = \left(\frac{2\lambda}{2\lambda+\mu-\beta\mu}\right)^2 \quad \Rightarrow \quad \beta = \frac{2\lambda}{\mu} + \frac{1}{2} - \sqrt{\frac{2\lambda}{\mu} + \frac{1}{4}}$

GI/M/1 (6)

- ❖ The mean sojourn time of a job in EMC is the same as the mean sojourn of a job in the original GI/M/1.
- T : Sojourn time of a job in the system

  $$\mathrm{E}[T] = \sum_{i=0}^{\infty}(i+1)\frac{1}{\mu}\pi_i = \sum_{i=0}^{\infty}(i+1)\frac{1}{\mu}(1-\beta)\beta^i = \frac{1}{\mu(1-\beta)}$$

- W: Waiting time of a job in queue
  - $\mathrm{E}[W] = \sum_{i=0}^{\infty} i\,\frac{1}{\mu}\pi_i = \sum_{i=0}^{\infty} i\,\frac{1}{\mu}(1-\beta)\beta^i = \frac{\beta}{\mu(1-\beta)}$
  - $\mathrm{E}[W] = \mathrm{E}[T]\text{-}\mathrm{E}[S] = \frac{1}{\mu(1-\beta)} - \frac{1}{\mu} = \frac{\beta}{\mu(1-\beta)}$
- N_A: the number of jobs in system seen by an arriving job
  - $\mathrm{E}[N_A] = \sum_{i=0}^{\infty} i\,\pi_i = \sum_{i=0}^{\infty} i\,(1-\beta)\beta^i = \frac{\beta}{1-\beta}$
  - Since $\mathrm{E}[W] = \mathrm{E}[N_A] \times \mathrm{E}[S]$, $\mathrm{E}[N_A] = \frac{\beta}{\mu(1-\beta)} \times \mu = \frac{\beta}{1-\beta}$
- ❖ The number of jobs in server seen by an arriving job: $1 - \pi_0 = \boldsymbol{\beta}$

GI/M/1 (7)

❖ Derivation using Laplace Transform

- Let T_i be the sojourn time of a job arriving while seeing i jobs in system
  - $T_i = X_1 + X_2 + \ldots + X_{i+1}$
    - X : a random variable for the service time of a job
      - ✓ an exponential random variable with rate μ
    - $F_X{}^*(s) = \frac{\mu}{\mu+s}$
  - $F_{T_i}{}^*(s) = (F_X{}^*(s))^{i+1} = \left(\frac{\mu}{\mu+s}\right)^{i+1}$
- T: the sojourn time of a job in the system
  - $F_T{}^*(s) = \sum_{i=0}^{\infty} F_{T_i}{}^*(s)\ \pi_i = \sum_{i=0}^{\infty} \left(\frac{\mu}{\mu+s}\right)^{i+1} (1\text{-}\beta)\ \beta^i = \frac{\mu(1-\beta)}{\mu(1-\beta)+s}$
  - $\mathrm{E}[T] = -\lim_{s\to 0} \frac{dF_T{}^*(s)}{ds} = \frac{1}{\mu(1-\beta)}$

The sojourn time of a job is exponentially distributed with mean $\frac{1}{\mu(1-\beta)}$

GI/M/1 (8)

- N: the number of jobs in system, seen by an outside observer
  - By Little's Law, $\boxed{E[N] = \lambda\, E[T] = \frac{\lambda}{\mu(1-\beta)}}$
    - ❖ The number of jobs in server seen by an outside observer: $\boldsymbol{\lambda/\mu}$ (by Little's law)

- Since an arrival process is not Poisson,
  the PASTA (Poisson Arrivals See Time Averages) property
  does not hold
  - $E[N] \neq E[N_A]$, this is $\frac{\lambda}{\mu(1-\beta)} \neq \frac{\beta}{1-\beta}$

- If and only if Poisson arrivals (GI $\rightarrow$ M**),** $E[N] = E[N_A]$

That is, since $\beta = \frac{\lambda}{\mu}$ in M/M/1

- $E[N] = \frac{\lambda}{\mu(1-\beta)} = \frac{\lambda}{\mu-\lambda}$, $E[N_A] = \frac{\beta}{1-\beta} = \frac{\lambda}{\mu-\lambda}$

H_2/M/1 (Example of GI/M/1)

- GI/M/l
  - $\beta = F_A{}^*(\mu - \beta\mu)$
  - $\mathrm{E}[N] = \frac{\lambda}{\mu(1-\beta)}$, $\mathrm{E}[T] = \frac{1}{\mu(1-\beta)}$, $\mathrm{E}[W] = \frac{\beta}{\mu(1-\beta)}$
- H_2/M/1
  - Remind that a k-stage hyperexponential random variable (H_k) is

    $F_A{}^*(\theta, k) = \sum_{i=1}^{k} p_i \left(\frac{\lambda_i}{\lambda_i+\theta}\right)$
  - $F_A{}^*(\theta, 2) = \frac{p\,\lambda_1}{\lambda_1+\theta} + \frac{(1-p)\,\lambda_2}{\lambda_2+\theta}$
  - $\beta = \frac{p\,\lambda_1}{\lambda_1+\mu-\beta\mu} + \frac{(1-p)\,\lambda_2}{\lambda_2+\mu-\beta\mu}$

$\therefore \quad \cancel{\beta}=1, \quad \beta = \frac{\mu+\lambda_1+\lambda_2}{2\mu} \cancel{+} \sqrt{\left(\frac{\mu+\lambda_1+\lambda_2}{2\mu}\right)^2 - \frac{p\lambda_1+(1-p)\lambda_2}{\mu}}$,

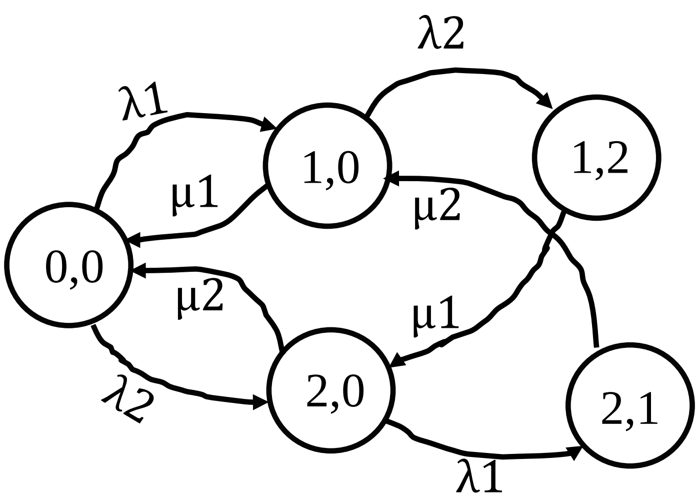
$$\beta = \frac{\mu+\lambda_1+\lambda_2}{2\mu} - \sqrt{\left(\frac{\mu+\lambda_1+\lambda_2}{2\mu}\right)^2 - \frac{p\lambda_1+(1-p)\lambda_2}{\mu}}$$

Exercises

1. There are two machines and one repairman in a small factory. For the machine i (i=1,2), its lifespan is exponentially distributed with mean $1/\lambda_i$ and the time taken to fix the machine i is exponentially distributed with mean $1/\mu_i$.

- This system can be modeled as a continuous time Markov chain. Define the system state.
- Draw the transition rate diagram

System state (수리중인 기계, 수리대기중인 기계 개수)



$$(\lambda 1+\lambda 2)\pi_{0,0} = \mu 1\pi_{1,0} + \mu 2\pi_{2,0}$$

$$(\mu 1+\lambda 2)\pi_{1,0} = \lambda 1\pi_{0,0} + \mu 2\pi_{2,1}$$

$$(\mu 2+\lambda 1)\pi_{2,0} = \lambda 2\pi_{0,0} + \mu 1\pi_{1,2}$$

$$\mu 1\pi_{1,2} = \lambda 2\pi_{1,0}$$

$$\mu 2\pi_{2,1} = \lambda 1\pi_{2,0}$$

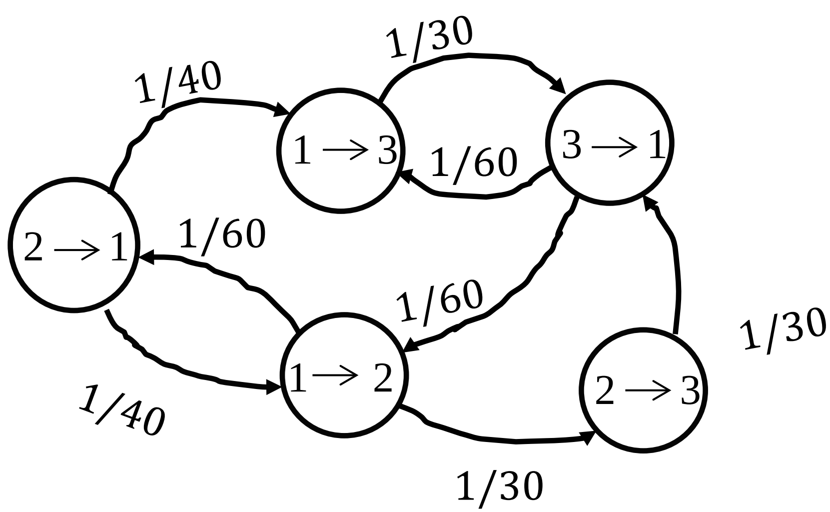
$$\pi_{0,0} + \pi_{1,0} + \pi_{2,0} + \pi_{1,2} + \pi_{2,1} = 1$$

2. A taxi alternates between three locations. When it reaches location 1, it is equally likely to go next to either 2 or 3. When it reaches 2 it will next go to 1 with probability 1/3 and to 3 with probability 2/3. From 3 it always goes to 1. The moving time from location i to location j is t_{ij} according to exponential distribution. And, $t_{12}= t_{21}=20$, $t_{13}= t_{31=}30$, $t_{23}= 30$.

① What is the (limiting) probability that the taxi's most recent stop was at location i, i=1,2,3 ?

② What is the (limiting) probability that the taxi is heading for location 2.

- System state: (departure location $\longrightarrow$ arrival location)



$$\pi_{2,1} = \frac{1}{19}, \ \pi_{1,3} = \frac{9}{38}$$

$$\pi_{3,1} = \frac{25}{38}, \pi_{1,2} = \frac{3}{19}, \pi_{2,3} = \frac{3}{19}$$

① Last stop location:

1: $\pi_{1,3} + \pi_{1,2} = \frac{15}{38}$

2: $\pi_{2,1,} + \pi_{2,3} = \frac{4}{19}$

3: $\pi_{3,1} = \frac{25}{38},$

② $\pi_{1,2}= \frac{3}{19}$

3. Consider an M/G/1 system in which a departing job immediately joins the queue with probability p or departs the system with probability (1-p). The queuing discipline is FCFS, and the service time for a returning job is independent of its previous service times. Let $B^*(s)$ be the Laplace transform for the service time and let $B^*_T(s)$ be the Laplace transform for the total service time of a job.

① Show that $B^*_T(s) = (1\text{-}p)\, B^*(s) / \{1\text{-}p\, B^*(s)\}$

② Find the first moment and second moment of the total service time, using p and the first moment and second moment of the service time.

③ Find the average number of jobs in the system

– $B_T{}^*(\theta) = \mathrm{E}\left[e^{-\theta(S_1+S_2+\cdots+S_N)}\right]$

$$= \sum_{k=1}^{\infty} \mathrm{E}\left[e^{-\theta(S_1+S_2+\cdots+S_N)} \middle| N = k\right] \times \Pr\{N = k\}$$

$$= \sum_{k=1}^{\infty} \mathrm{E}\left[e^{-\theta(S_1+S_2+\cdots+S_k)}\right] \times p^{k-1}(1-p)$$

$$= \sum_{k=1}^{\infty} \left(\mathrm{E}\left[e^{-\theta S}\right]\right)^k \times p^{k-1}(1-p) = \frac{(1-p)B^*(\theta)}{1-B^*(\theta)}$$

$$= (1-\mathrm{p})B^*(\theta) \sum_{k=1}^{\infty} \left(pB^*(\theta)\right)^{k-1} = \frac{(1-p)B^*(\theta)}{1-pB^*(\theta)}$$

② Find the first moment and second moment of the total service time, using p and the first moment and second moment of the service time.

- $B_T{}^*(\theta) = \frac{(1-p)B^*(\theta)}{1-pB^*(\theta)}$
- $E[T] = -\lim_{\theta \to 0} \frac{dB_T{}^*(\theta)}{d\theta} = \frac{E[S]}{1-p}$
- $E[T^2] = \lim_{\theta \to 0} \frac{d^2B_T{}^*(\theta)}{d\theta^2} = \frac{(1-p)E\left[S^2\right]+2p(E[S])^2}{(1-p)^2}$

③ Find the average number of jobs in the system

- $\bar{N} = \lambda E[T] + \frac{\lambda^2 E\left[T^2\right]}{2(1-\lambda E[T])}$