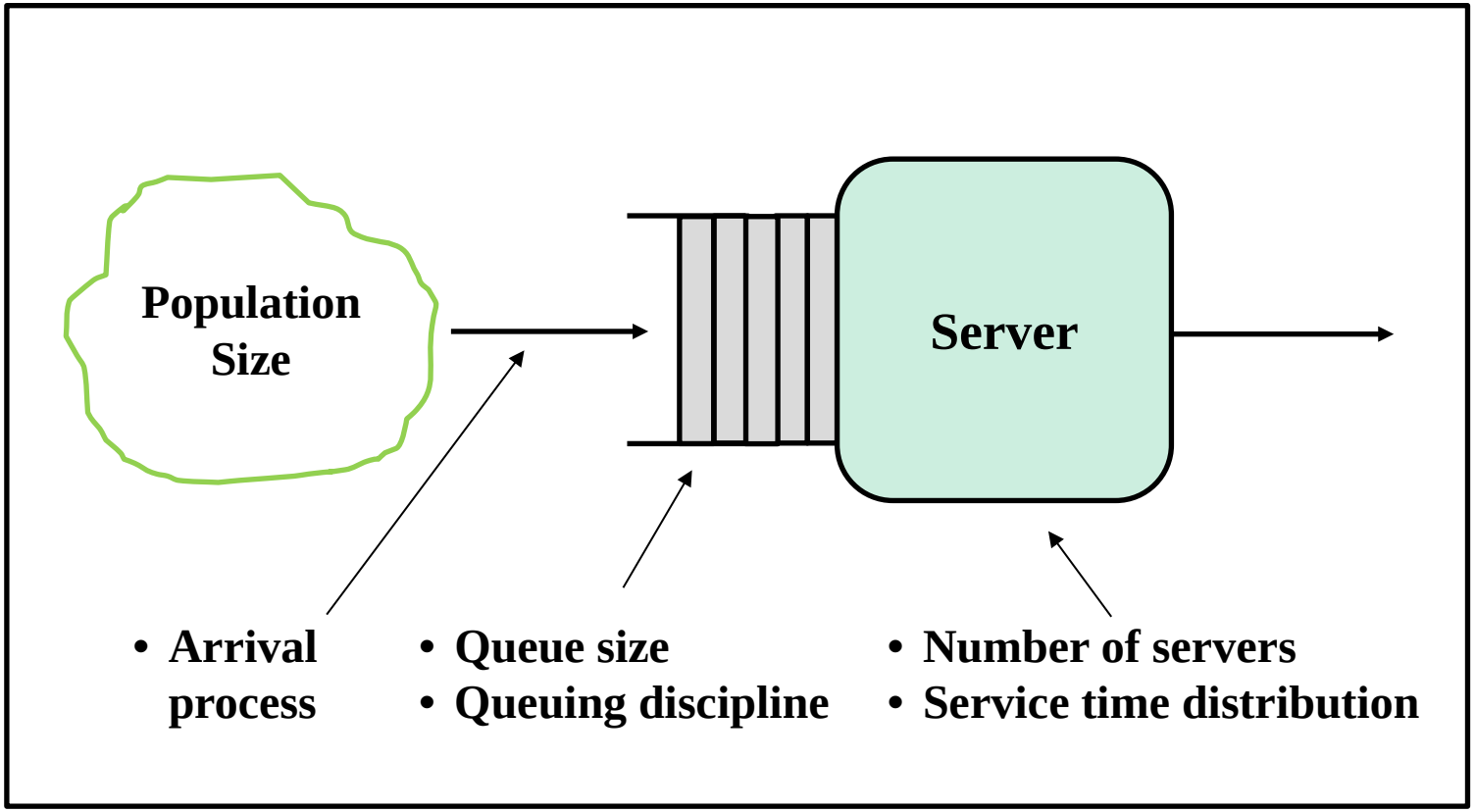
Single Queue System I

Wha Sook Jeon

Mobile Computing & Communications Lab.

Kendall's Notation (1)

- One way to describe queuing system



Queuing System

Kendall's Notation (2)

$$\underbrace{\text{Arrival process} / \text{Service time distribution} / \text{Number of servers}}_{\text{mandatory}} / \text{Size of queue} / \text{Size of population} / \text{Queuing discipline}$$

- Arrival process/service time distribution
  - M: Poisson arrivals or exponential service time distribution
  - E_k: k-stage Erlang distribution
  - D: deterministic distribution
  - G: general distribution (GI - general & independent arrival)
- Size of queue (including jobs in the server): ∞ (default)
- Size of population: ∞ (default)
- Queuing discipline
  - FCFS (default)
  - LCFS
  - RS (random selection)
  - PR (priority): preemptive or non-preemptive

Kendall's Notation (3)

- Kendall's notation is only for continuous time stochastic process
- Examples of Kendall's notation
  - M/M/1
  - M/G/1
  - E_k/M/1
  - M/M/c/c : (no queueing)
  - M/M/c/K: K is typically larger than c
  - M/M/∞ : no waiting
  - M/G/∞
  - GI/M/1

M/M/1 (1)

- Poisson arrival process (exponentially distributed inter-arrival time)
- exponentially distributed service time
- One server

- A special case of Birth & Death process
  - Birth & Death process with fixed arrival rate λ and fixed service rate μ, which are independent of system state
- Stochastic process representation of M/M/1:
  - $\{N(t), t \geq 0\}$
    - $N(t)$ is a random variable that represents the number of jobs in the system at time t

M/M/1 (2)

- Derivation of state probability distribution

Let $P_n(t) \triangleq \Pr\{N(t)=n\}$

Since M/M/1 is a birth and death process,

$$P_0(t+h) = P_0(t)(1-\lambda h) + P_1(t)\mu h$$

$$P_n(t+h) = P_{n-1}(t)\lambda h + P_{n+1}(t)\mu h + P_n(t)(1-\lambda h-\mu h), \quad \text{for } n \geq 1$$

$$\lim_{h\to 0} \frac{P_0(t+h) - P_0(t)}{h} = P_0'(t) = -\lambda P_0(t) + \mu P_1(t) \qquad \cdots\cdot (1)$$

$$\lim_{h\to 0} \frac{P_n(t+h) - P_n(t)}{h} = P_n'(t) = \lambda P_{n-1}(t) + \mu P_{n+1}(t) - \lambda P_n(t) - \mu P_n(t)$$

$$\text{for } n \geq 1 \qquad \cdots\cdot (2)$$

M/M/1 (3)

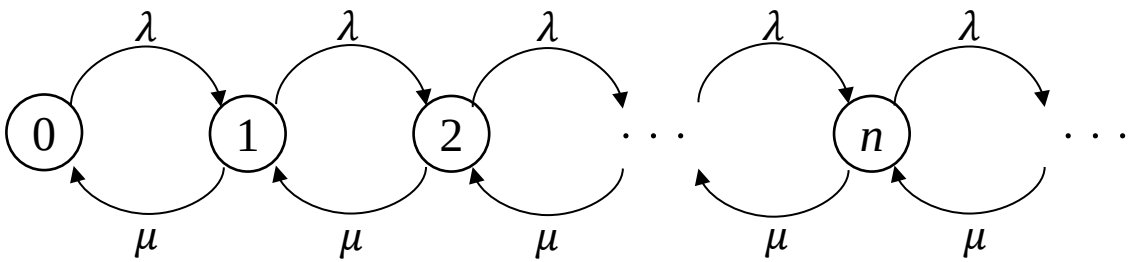
- In steady state, since $P_n'(t) = 0,\ P_n(t) = P_n$
- From (1) and (2),

$$-\lambda P_0 + \mu P_1 = 0$$

$$\lambda P_{n-1} + \mu P_{n+1} - \lambda P_n - \mu P_n = 0, \quad \text{for } n \geq 1$$

$$\Rightarrow \quad \lambda P_n = \mu P_{n+1}, \quad \text{for } n \geq 0 \qquad \cdots\cdot (3)$$

- From (3), state transition rate diagram is



M/M/1 (4)

- From the state rate transition diagram

$$\begin{aligned} \lambda P_0 &= \mu P_1 \\ \lambda P_1 &= \mu P_2 \\ &\vdots \\ \times \quad \lambda P_{n-1} &= \mu P_n \end{aligned}$$

$$\lambda^n P_0 P_1 \cdots P_{n-1} = \mu^n P_1 P_2 \cdots P_{n-1} P_n \qquad \Rightarrow \lambda^n P_0 = \mu^n P_n$$

- Let $\rho = \frac{\lambda}{\mu}$. For stable system, $\lambda < \mu \quad (\rho < 1)$
- $P_n = \rho^n P_0$
- Since $\sum_{n=0}^{\infty} P_n = 1$

  $P_0 = 1 - \rho$ from $P_0(1 + \rho + \rho^2 + \rho^3 + \cdots) = 1$

  - P_0: the probability that the server is idle
  - ρ **:** the probability that the server is busy
- $P_n = \rho^n (1 - \rho)$

M/M/1 (5)

- Performance Measures
  - $\bar{N}$: the mean number of jobs in the system
    - $\bar{N} = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n\rho^n(1-\rho) = (1-\rho)\sum_{n=0}^{\infty} n\rho^n$

      $= \frac{\rho}{1-\rho}$ (since $\sum_{n=0}^{\infty} n\rho^n = \frac{\rho}{(1-\rho)^2}$)
  - $\bar{M}$: the mean number of jobs in the server
    - $\bar{M} = 1\cdot \rho + 0\cdot (1-\rho) = \rho$
  - $\bar{Q}$: the mean number of jobs in the queue
    - $\bar{Q} = \sum_{n=1}^{\infty}(n-1)P_n = \sum_{n=1}^{\infty}(n-1)\rho^n(1-\rho)$

      $= \rho(1-\rho)\sum_{n=0}^{\infty} n\rho^n = \frac{\rho^2}{1-\rho}$
    - By Little's Law, $\bar{Q} = \bar{N} - \rho = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$

M/M/1 (6)

- $\bar{T}$: the mean sojourn time of a job in the system
  - $\bar{T} = \sum_{n=0}^{\infty} (n+1) \frac{1}{\mu} P_n = \frac{1}{\mu} (\sum_{n=0}^{\infty} nP_n + \sum_{n=0}^{\infty} P_n)$

    $= \frac{1}{\mu} \left(\frac{\rho}{1-\rho} + 1 \right) = \frac{1}{\mu} \frac{1}{1-\rho}$
  - By Little's Law, $\bar{T} = \frac{1}{\lambda} \bar{N} = \frac{1}{\lambda} \frac{\rho}{1-\rho} = \frac{1}{\mu} \frac{1}{1-\rho}$
- $\bar{W}$: the mean waiting time of a job in the queue
  - $\bar{W} = \sum_{n=0}^{\infty} \frac{1}{\mu} nP_n = \frac{1}{\mu} \frac{\rho}{1-\rho}$
  - mean service time: $\bar{S} = \frac{1}{\mu}$
  - By Little's Law, $\bar{W} = \bar{T} - \bar{S} = \frac{1}{\mu} \frac{1}{1-\rho} - \frac{1}{\mu} = \frac{1}{\mu} \frac{\rho}{1-\rho}$

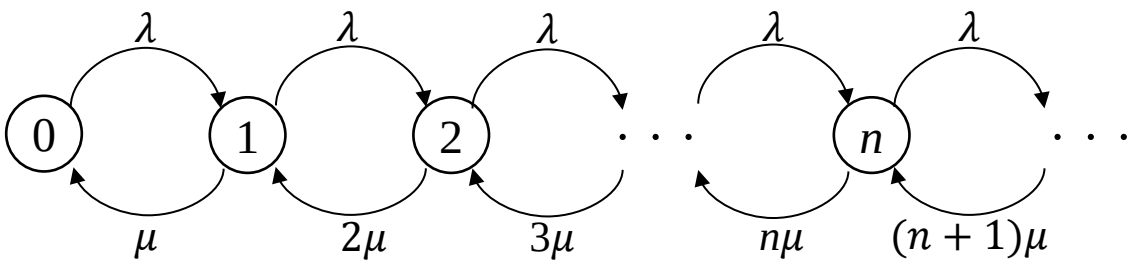
M/M/1 (7)

- Throughput:
  - Mean departure rate: $\Lambda = 0\,P_0 + (1 - P_0)\mu = \rho\mu = \lambda$: arrival rate
- Sever utilization: ρ
  - Server busy probability
  - Mean number of jobs in the server

M/M/∞ (1)

- arrival rate: $\lambda(n) = \lambda$
- load dependent service rate: $\mu(n) = n\mu$

- State transition rate diagram



$$\lambda P_0 = \mu P_1$$
$$\lambda P_1 = 2\mu P_2$$
$$\vdots$$
$$\times \ \lambda P_{n-1} = n\mu P_n$$

$$\lambda^n P_0 P_1 \cdots P_{n-1} = n!\, \mu^n P_1 P_2 \cdots P_{n-1}\, P_n$$

$$\Rightarrow \quad P_n = \frac{1}{n!}\rho^n P_0, \quad \text{where } \rho = \frac{\lambda}{\mu}$$

M/M/∞ (2)

- Since $\sum_{n=0}^{\infty} P_n = 1$,

  $P_0 \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = 1 \quad \Rightarrow \quad P_0 = e^{-\rho}$

- $P_n = \frac{\rho^n}{n!} e^{-\rho} = \frac{(\lambda \frac{1}{\mu})^n}{n!} e^{-\lambda \frac{1}{\mu}}$ pdf of Poisson distribution

- Performance Measures
  - Mean number of jobs in system

    $\bar{N} = \sum_{n=0}^{\infty} n \frac{\rho^n}{n!} e^{-\rho} = \rho e^{-\rho} \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = \rho$

    - When using Little's Law, $\bar{N} = \bar{T} \times \lambda = \frac{1}{\mu} \times \lambda = \rho$
  - System utilization: $1 - P_0 = 1 - e^{-\rho}$
  - Mean response time: $\bar{T} = \frac{1}{\mu}$
  - Throughput: $\Lambda = 0\, P_0 + \sum_{n=1}^{\infty} n\mu\, P_n$

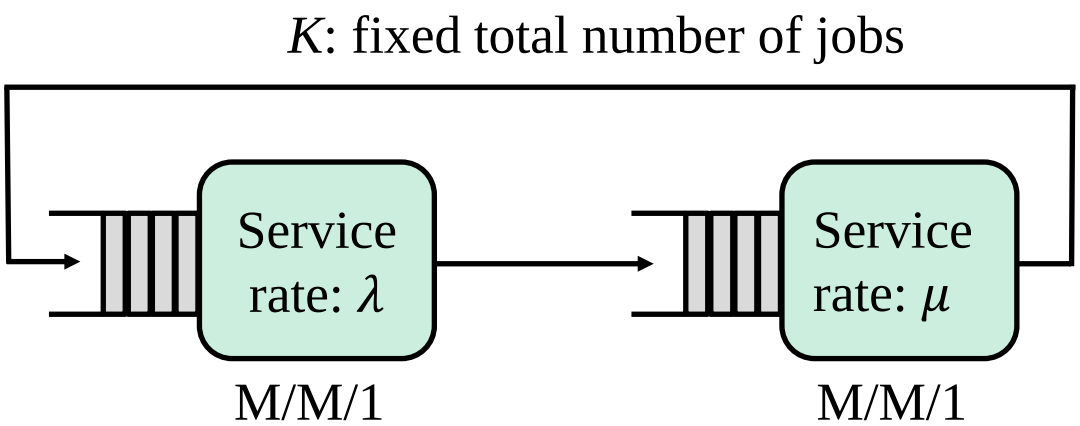
    $= \sum_{n=1}^{\infty} n\mu \frac{\rho^n}{n!} e^{-\rho} = \mu\rho e^{-\rho} \sum_{n=1}^{\infty} \frac{\rho^{n-1}}{(n-1)!} = \mu\rho = \lambda$

M/G/∞ :

$$P_n = \frac{(\lambda \mathrm{E}[s])^n}{n!} e^{-\lambda \mathrm{E}[S]}$$

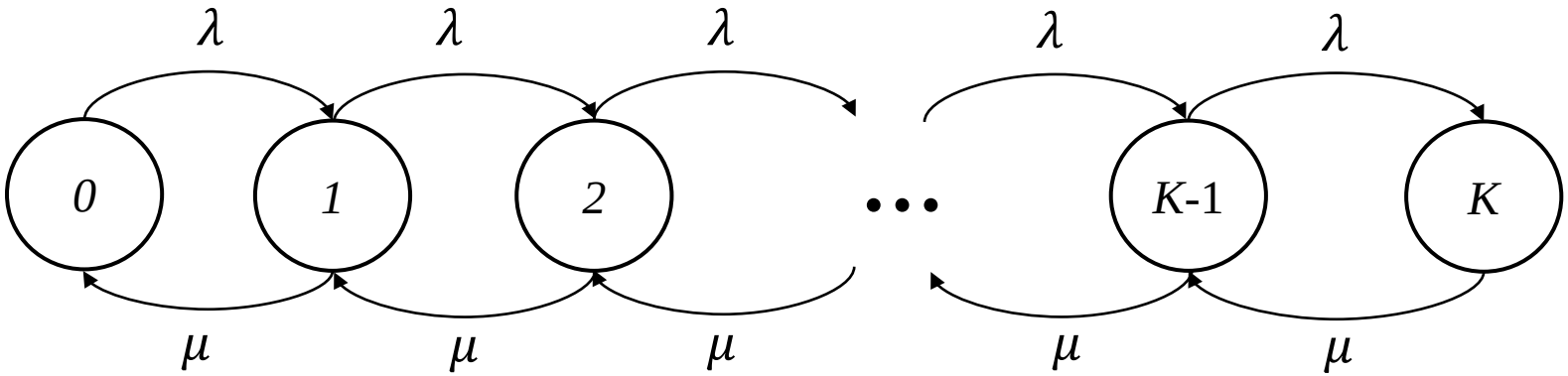
M/M/1/K (1)

- Arrival rate $\lambda(n) = \lambda, \quad \text{for } n \leq K$
- Service rate $\mu(n) = \mu, \quad \text{for } n \leq K$
- One server
- The number of jobs in system: at maximum K



M/M/1/K (2)

- A state transition rate diagram



$$\lambda P_0 = \mu P_1$$

$$\lambda P_1 = \mu P_2$$

$$\lambda P_2 = \mu P_3$$

$$\vdots$$

$$\times \quad \lambda P_{n-1} = \mu P_n$$

$$\lambda^n P_0 P_1 \cdots P_{n-1} = \mu^n P_1 P_2 \cdots P_{n-1} P_n$$

$$\Rightarrow \lambda^n P_0 = \mu^n P_n \qquad \therefore P_n = \left(\frac{\lambda}{\mu}\right)^n P_0 \quad (0 \le n \le K)$$

M/M/1/K (3)

- Since $\sum_{n=0}^{K} P_n = 1$

$$1 = P_0\left\{1 + \left(\frac{\lambda}{\mu}\right) + \left(\frac{\lambda}{\mu}\right)^2 + \left(\frac{\lambda}{\mu}\right)^3 + \ \dots + \left(\frac{\lambda}{\mu}\right)^K\right\}$$

$$= P_0\left(\frac{1-\left(\frac{\lambda}{\mu}\right)^{K+1}}{1-\frac{\lambda}{\mu}}\right) \quad \Rightarrow \quad P_0 = \left(\frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{K+1}}\right)$$

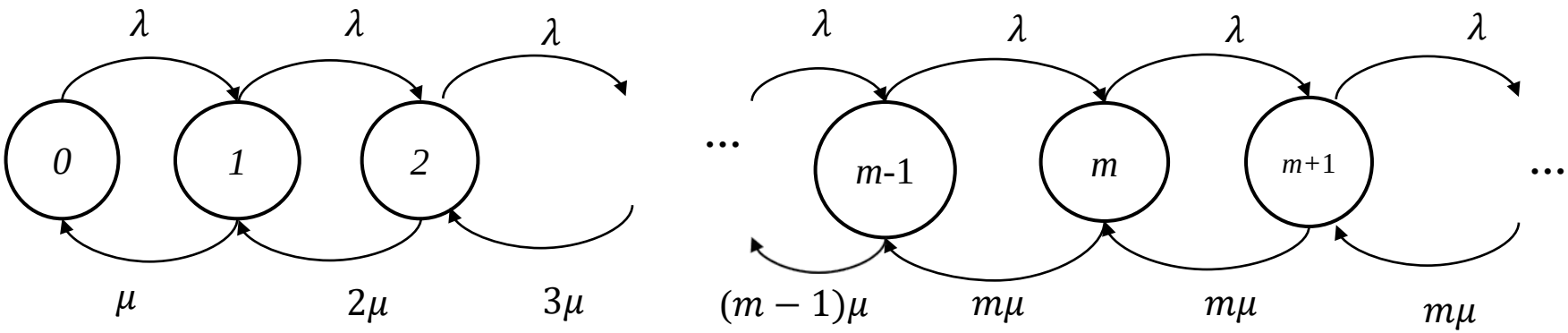
$$\therefore\ P_n = \left(\frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{K+1}}\right)\left(\frac{\lambda}{\mu}\right)^n \qquad (0 \le n \le K)$$

M/M/1/K (4)

- Performance Measures
  - Probability that an arriving job is blocked: P_K
  - Throughput: Λ
    - Arrival rate: $\Lambda = \lambda(1 - P_K)$
    - Departure rate: $\Lambda = (1-P_0)\mu$
  - The number of jobs in server: N_s
    - Server busy probability: $N_s = 1-P_0$
    - By Little's law: $N_s = \textit{arrival rate} \times \textit{service time of a job} = \Lambda \frac{1}{\mu}$
  - The number of jobs in system: N
    - $N = \sum_{n=0}^{K} n\, P_n$
  - Mean response time: T
    - By Little's law: $T = N/\Lambda$

M/M/m (1)

- State transition rate diagram



- If $n \leq m$

$$\lambda P_0 = \mu P_1$$
$$\lambda P_1 = 2\mu P_2$$
$$\vdots$$
$$\times \quad \lambda P_{n-1} = n\mu P_n$$

$$\lambda^n P_0\, P_1 P_2 \ldots P_{n-1} = n!\, \mu^n P_1 P_2 P_3 \ldots P_n$$

$$\Rightarrow \quad \lambda^n P_0 = n!\, \mu^n P_n$$

$$\therefore\ P_n = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 \quad (n \leq m)$$

M/M/m (2)

- If $n > m$,

$$\begin{aligned} \lambda P_m &= m\mu P_{m+1} \\ \lambda P_{m+1} &= m\mu P_{m+2} \\ \lambda P_{m+2} &= m\mu P_{m+3} \\ &\vdots \\ \times \quad \lambda P_{n-1} &= m\mu P_n \end{aligned}$$

$$\lambda^{n-m} P_m P_{m+1} P_{m+2} \cdots P_{n-1} = (m\mu)^{n-m} P_{m+1} P_{m+2} P_{m+3} \cdots P_n$$

$$\Rightarrow \quad \lambda^{n-m} P_m = (m\mu)^{n-m} P_n$$

$$\therefore \; P_n = \left(\frac{\lambda}{m\mu}\right)^{n-m} \frac{1}{m!} \left(\frac{\lambda}{\mu}\right)^m P_0 \quad \text{for } n > m$$

M/M/m (3)

- Since $\sum_{n=0}^{\infty} P_n = 1$

$$1 = \sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 + \sum_{n=m+1}^{\infty} \left(\frac{\lambda}{m\mu}\right)^{(n-m)} \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m P_0$$

$$= P_0 \left\{ \sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \sum_{k=1}^{\infty} \left(\frac{\lambda}{m\mu}\right)^k \right\}$$

$$= P_0 \left\{ \sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \left(\frac{\lambda}{m\mu-\lambda}\right) \right\}$$

$$\therefore\ P_0 = \frac{1}{\sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \left(\frac{\lambda}{m\mu-\lambda}\right)}$$

- Erlang's C formula
  - The probability that an arriving customer enters the queue

$$\Pr\{\text{queuing}\} = \sum_{n=m}^{\infty} P_n = \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{m\mu-\lambda}{m\mu} P_0$$

M/M/m (4)

- Performance Measures
  - The number of jobs in server: ρ
    - By Little's law: ρ = arrival rate × mean service time of a job = $\lambda \cdot \frac{1}{\mu}$
    - $\rho = \sum_{n=0}^{m} n \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 + \sum_{n=m+1}^{\infty} m \left(\frac{\lambda}{m\mu}\right)^{(n-m)} \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m P_0$

$$= \frac{\lambda}{\mu} P_0 \left\{ \sum_{n=1}^{m} \frac{1}{(n-1)!}\left(\frac{\lambda}{\mu}\right)^{n-1} + \frac{1}{(m-1)!}\left(\frac{\lambda}{\mu}\right)^{m-1} \sum_{k=1}^{\infty}\left(\frac{\lambda}{m\mu}\right)^k \right\}$$

$$= \frac{\lambda}{\mu} P_0 \left\{ \sum_{n=0}^{m-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^{n} + \frac{1}{(m-1)!}\left(\frac{\lambda}{\mu}\right)^{m-1} \left(\frac{\lambda}{m\mu-\lambda}\right) \right\}$$

$$= \frac{\lambda}{\mu} P_0 \left\{ \sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^{n} + \frac{1}{(m-1)!}\left(\frac{\lambda}{\mu}\right)^{m-1} \left(\frac{\lambda}{m\mu-\lambda}\right) - \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^{m} \right\}$$

$$= \frac{\lambda}{\mu} P_0 \left\{ \sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^{n} + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^{m} \left(\frac{\lambda}{m\mu-\lambda}\right) \right\} \leftarrow \frac{1}{P_0}$$

$$= \frac{\lambda}{\mu}$$

  - The number of jobs in system: $N = \sum_{n=0}^{\infty} n P_n$
    - $N = \sum_{n=0}^{m} n \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 + \sum_{n=m+1}^{\infty} n \left(\frac{\lambda}{m\mu}\right)^{(n-m)} \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m P_0$

$$= P_0 \left\{ \sum_{n=0}^{m-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^{n+1} + \frac{1}{(m-1)!}\left(\frac{\lambda}{\mu}\right)^{m} \frac{\lambda}{(m\mu-\lambda)} \left(1 + \frac{\mu}{m\mu-\lambda}\right) \right\}$$

M/M/m (5)

- Performance Measures
  - Mean response time: $T = N/\lambda$
  - Throughput: Λ
    - Arrival rate: λ
    - Departure rate:

$$\Lambda = 0\, P_0 + \sum_{n=1}^{m} n\mu\, P_n + \sum_{n=m+1}^{\infty} m\mu\, P_n$$

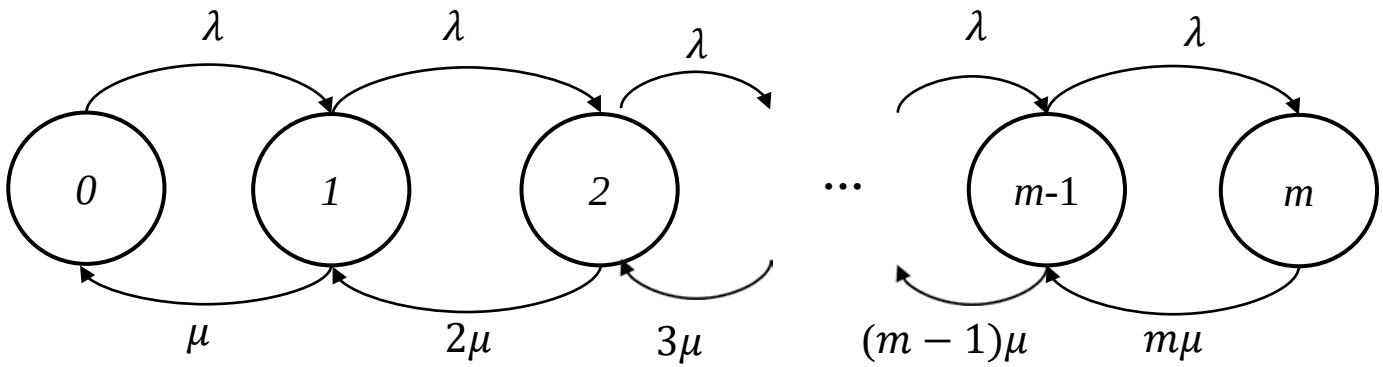
$$= \sum_{n=1}^{m} n\mu \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 + \sum_{n=m+1}^{\infty} m\mu \left(\frac{\lambda}{m\mu}\right)^{(n-m)} \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m P_0$$

$$= \mu \left(\sum_{n=1}^{m} n \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 + \sum_{n=m+1}^{\infty} m \left(\frac{\lambda}{m\mu}\right)^{(n-m)} \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m P_0\right)$$

$$= \rho\mu = \lambda$$

M/M/m/m (1)

- State transition rate diagram



- $n = 0, 1, 2, \ldots, m$

$$\lambda P_0 = \mu P_1$$
$$\lambda P_1 = 2\mu P_2$$
$$\vdots$$
$$\times \quad \lambda P_{n-1} = n\mu P_n$$

$$\lambda^n P_0 \, P_1 P_2 \ldots P_{n-1} = n! \, \mu^n P_1 P_2 P_3 \ldots P_n$$

$$\Rightarrow \quad \lambda^n P_0 = n! \, \mu^n P_n$$

$$\therefore \; P_n = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0$$

M/M/m/m (2)

- Since $\sum_{n=0}^{m} P_n = 1$

$$\sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 = 1 \quad \Rightarrow \quad P_0 = \frac{1}{\sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n}$$

- Erlang's *B* formula
  - Probability that an arriving customer is blocked: Erlang's loss formula

$$P_m = \frac{\frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m}{\sum_{n=0}^{m} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n}$$

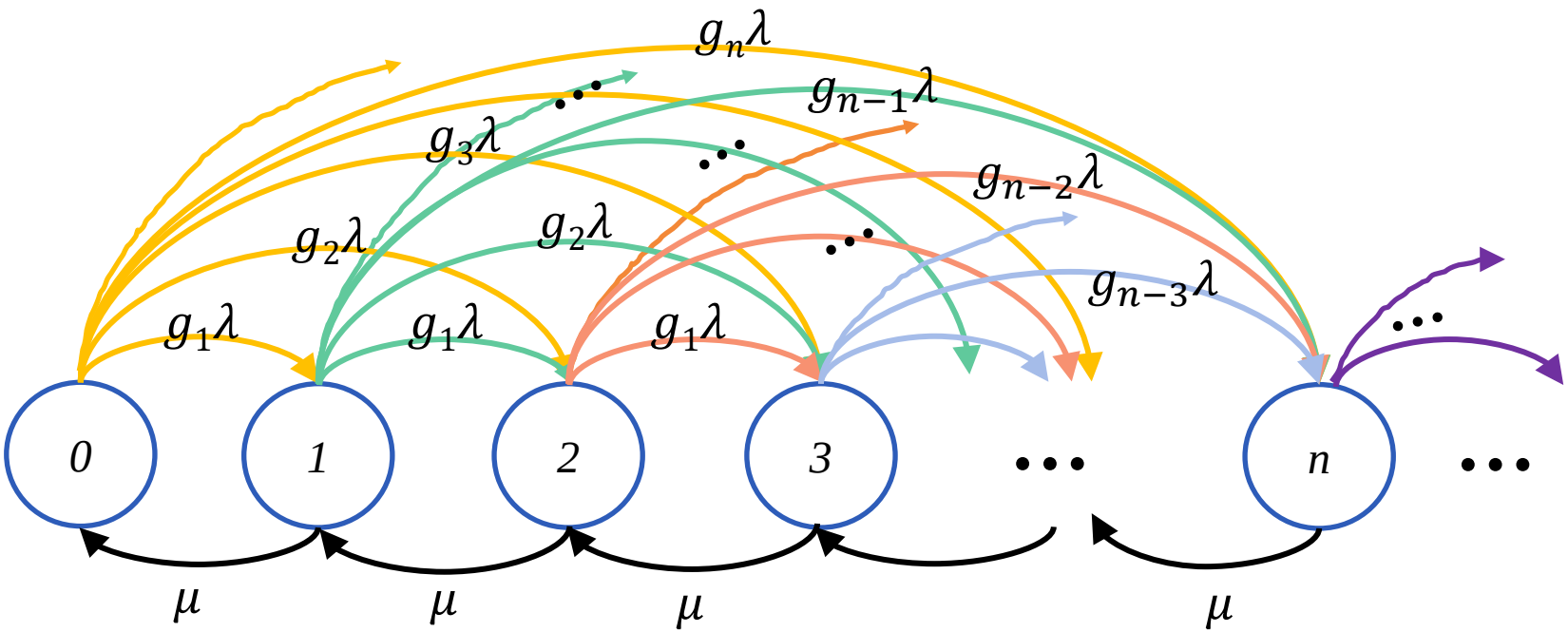
- Performance Measures
  - Throughput: $\lambda(1- P_m)$
    - $\sum_{n=0}^{m} n\mu \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 = \mu\frac{\lambda}{\mu}\sum_{n=0}^{m-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 = \lambda(1- P_m)$
  - The number of jobs in server (system) : $\lambda(1- P_m)\frac{1}{\mu}$
    - $\sum_{n=0}^{m} n \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 = \frac{\lambda}{\mu}\sum_{n=0}^{m-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 = \frac{\lambda}{\mu}(1- P_m)$

$M^X/M/1$ (1)

- Compound Poisson arrival process (CPP): Batch (Bulk) arrival
- Size of a batch: a random variable G with general distribution

$$\Pr\{G = n\} = g_n \quad (n \geq 1)$$

- State transition rate diagram



$M^X/M/1$ (2)

- $\lambda P_0 = \mu P_1 \quad \ldots \ (1)$
- $(\lambda+\mu)P_1 = \lambda g_1 P_0 + \mu P_2$

  $(\lambda+\mu)P_2 = \lambda g_2 P_0 + \lambda g_1 P_1 + \mu P_3$

  $(\lambda+\mu)P_3 = \lambda g_3 P_0 + \lambda g_2 P_1 + \lambda g_1 P_2 + \mu P_4$

  $\vdots$

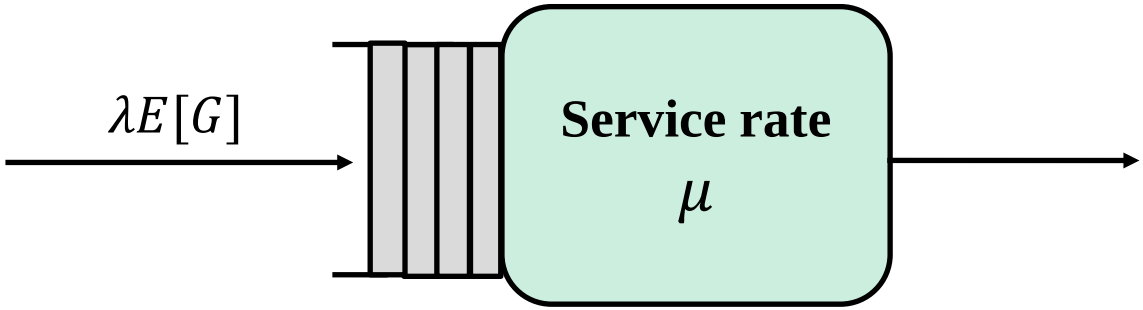
  $\Rightarrow \quad (\lambda+\mu)P_n = \mu P_{n+1} + \lambda \sum_{k=1}^{n} g_k P_{n-k} \qquad (n \geq 1) \qquad \ldots (2)$

- How to calculate the mean number of jobs in the system, $\bar{N}$

1. Calculate $P_0, P_1, P_2 \ldots$ until $(nP_n < \varepsilon)$ and approximate $\bar{N} = \sum_{k=0}^{n} kP_k$

2. Derive $\bar{N}$ using probability generating function

$M^X/M/1$ (3)



Mean number of jobs in server: $\rho = \frac{\lambda E[G]}{\mu}$
(server busy probability)

1. Calculation of state probability distribution: $\{P_n\}$

$$P_0 = 1 - \rho = 1 - \lambda E[G]/\mu$$

$$P_1 = \frac{\lambda}{u} P_0$$

$$P_2 = \left(1 + \frac{\lambda}{\mu}\right) P_1 - \frac{\lambda}{u} g_1 P_0$$

$$\vdots$$

until $(nP_n < \varepsilon)$

$M^X/M/1$ (4)

2. Derivation of $\bar{N}$ using probability generating function

- Using the probability generating function, we define

$$\mathrm{Y}(z) = E[z^X] = \sum_{n=0}^{\infty} z^n P_n$$

$$\mathrm{B}(z) = E[z^G] = \sum_{k=1}^{\infty} z^k g_k$$

$$\lambda P_0 = \mu P_1 \quad (1)$$

$$(\lambda+\mu)P_n = \mu P_{n+1} + \lambda \sum_{k=1}^{n} g_k P_{n-k} \qquad (n \geq 1) \quad (2)$$

- Using (2)

$$\sum_{n=1}^{\infty} z^n (\lambda+\mu) P_n = \sum_{n=1}^{\infty} z^n (\mu P_{n+1} + \lambda \sum_{k=1}^{n} g_k P_{n-k})$$

$$(\lambda+\mu) \sum_{n=1}^{\infty} z^n P_n = \frac{\mu}{z} \sum_{n=1}^{\infty} z^{n+1} P_{n+1} + \lambda \sum_{n=1}^{\infty} \sum_{k=1}^{n} z^n g_k P_{n-k}$$

$$(\lambda+\mu)(\sum_{n=0}^{\infty} z^n P_n - P_0)$$

$$= \frac{\mu}{z}(\sum_{n=0}^{\infty} z^n P_n - P_0 - zP_1) + \lambda \sum_{n=1}^{\infty} \sum_{k=1}^{n} z^k g_k z^{n-k} P_{n-k}$$

$$\rightarrow \quad (\lambda+\mu)(Y(z) - P_0) = \frac{\mu}{z}(Y(z) - P_0 - zP_1) + \lambda \sum_{n=1}^{\infty} \sum_{k=1}^{n} z^k g_k z^{n-k} P_{n-k}$$

$M^X/M/1$ (5)

- $\sum_{n=1}^{\infty}\sum_{k=1}^{n} z^k g_k\, z^{n-k} P_{n-k}$

$$
\begin{array}{llll}
= & z^1 g_1 z^0 P_0 & & \\
+ & z^1 g_1 z^1 P_1 & + z^2 g_2 z^0 P_0 & \\
+ & z^1 g_1 z^2 P_2 & + z^2 g_2 z^1 P_1 & + z^3 g_3 z^0 P_0 \\
+ & z^1 g_1 z^3 P_3 & + z^2 g_2 z^2 P_2 & + z^3 g_3 z^1 P_1 \quad + z^4 g_4 z^0 P_0 \\
 & \vdots & \vdots & \vdots \qquad\qquad \vdots \qquad\qquad \cdots \\
 & ① & ② & ③
\end{array}
$$

$\leftarrow \sum_{k=1}^{1} z^k g_k\, z^{1-k} P_{1-k}$

$\leftarrow \sum_{k=1}^{2} z^k g_k\, z^{2-k} P_{2-k}$

$\leftarrow \sum_{k=1}^{3} z^k g_k\, z^{3-k} P_{3-k}$

$\leftarrow \sum_{k=1}^{4} z^k g_k\, z^{4-k} P_{4-k}$

① $= z^1 g_1 (z^0 P_0 + z^1 P_1 + z^2 P_2 + \cdots) = z^1 g_1 Y(z)$

② $= z^2 g_2 (z^0 P_0 + z^1 P_1 + z^2 P_2 + \cdots) = z^2 g_2 Y(z)$

③ $= z^3 g_3 (z^0 P_0 + z^1 P_1 + z^2 P_2 + \cdots) = z^3 g_3 Y(z)$

$\vdots$

① + ② + ③ + $\cdots = Y(z)(z^1 g_1 + z^2 g_2 + z^3 g_3 + \cdots) = \boldsymbol{Y}(z)\,\boldsymbol{B}(z)$

$M^X/M/1$ (6)

$$\Rightarrow (\lambda+\mu)z(Y(z)-P_0)=\mu(Y(z)-P_0-zP_1)+\lambda zY(z)B(z) \quad \ldots (3)$$

- From (1) and (3), $Y(z)=\frac{\mu P_0(1-z)}{\mu(1-z)-\lambda z+\lambda zB(z)}$ … (4)

- $Y(1)=1$, but $Y(1)=\frac{\mu P_0(1-1)}{\mu(1-1)-\lambda+\lambda B(1)}=\frac{0}{0}$ from (4)

By using L'Hopital's rule,

$$Y(1)=\lim_{z\to 1}Y(z)=\lim_{z\to 1}\frac{(\mu P_0(1-z))'}{(\mu(1-z)-\lambda z+\lambda zB(z))'}$$

$$=\lim_{z\to 1}\frac{-\mu P_0}{-\mu-\lambda+\lambda B(z)+\lambda zB'(z)}=\frac{-\mu P_0}{-\mu-\lambda+\lambda B(1)+\lambda B'(1)} \quad \ldots (5)$$

- Since $B(1)=1$ and $B'(1)=E[G]$, from (5)

$$Y(1)=\frac{-\mu P_0}{-\mu+\lambda E[G]}=1 \qquad \therefore P_0=1-\frac{\lambda E[G]}{\mu}$$

$M^X/M/1$ (7)

- Performance Measures
  - $\bar{N}$: the mean number of jobs in the system

    : by applying the L'Hospital rule twice after calculating $\frac{dY(z)}{dz}$,

    $$\bar{N} = \mathrm{E}[X] = \mathrm{Y}'(1) = \frac{\lambda(\mathrm{E}[G]+\mathrm{E}[G^2])}{2(\mu-\lambda\mathrm{E}[G])}$$
  - ρ: the mean number of jobs in server

    $$\rho = \lambda\mathrm{E}[G]\mathrm{E}[S] = \frac{\lambda}{\mu}\mathrm{E}[G]$$
  - $\bar{Q}$: the mean number of jobs in the queue

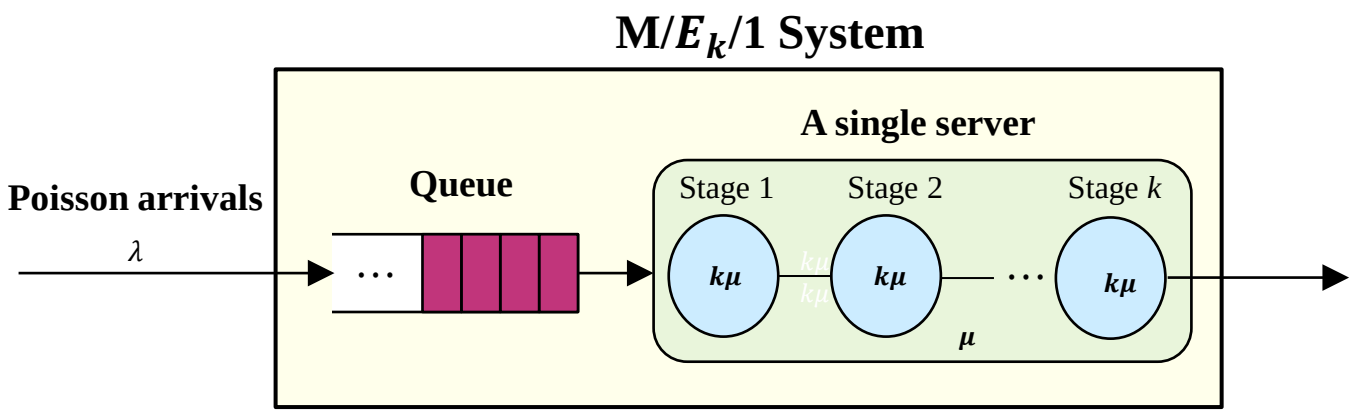
    $$\bar{Q} = \bar{N} - \rho$$
  - $\bar{T}$: the mean sojourn time of a job in system

    $$\bar{T} = \frac{\bar{N}}{\lambda\mathrm{E}[G]} = \frac{\mathrm{E}[G]+\mathrm{E}[G^2]}{2\mathrm{E}[\mathrm{G}](\mu-\lambda\mathrm{E}[G])}$$ by Little's law
  - Λ: Throughput

    $$\Lambda = P_0\, 0 + (1- P_0)\, \mu = \frac{\lambda}{\mu}\mathrm{E}[G]\, \mu = \lambda\mathrm{E}[G]$$

M/E$_k$/1 (1)

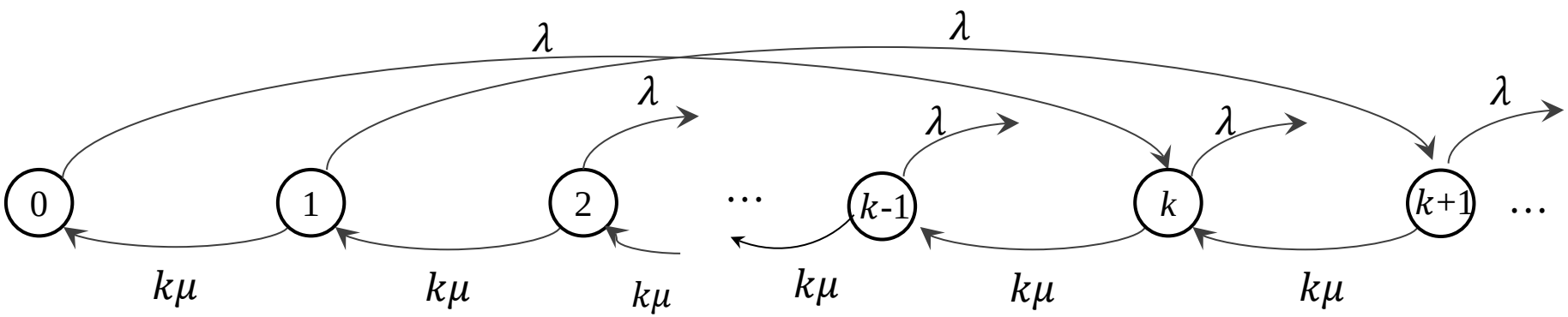
- $Y = X_1 + X_2 + \cdots + X_k$
  - Y: k-stage Erlang r.v. with service rate μ
    - mean service time $= \frac{1}{\mu}$
  - X: exponential r.v. with service rate $k\mu$
    - mean service time $= \frac{1}{k\mu}$



- Each stage is modeled as exponential sever with service rate $k\mu$

M/E$_k$/1 (2)

- We will take the system state as the total number of service stages left in the system (not the number of jobs in the system)



$$\lambda\pi_0 = k\mu\pi_1$$

$$(\lambda + k\mu)\pi_n = k\mu\pi_{n+1} \qquad (1 \le n \le k-1)$$

$$(\lambda + k\mu)\pi_n = \lambda\pi_{n-k} + k\mu\pi_{n+1} \qquad (n \ge k)$$

M/E$_k$/1 (3)

- **Solution approach 1**

$$\pi_1 = \frac{\lambda}{k\mu}\pi_0$$

$$\pi_n = \frac{\lambda + k\mu}{k\mu}\pi_{n-1} \qquad (2 \le n \le k)$$

$$\pi_n = -\frac{\lambda}{k\mu}\pi_{n-k-1} + \frac{\lambda + k\mu}{k\mu}\pi_{n-1} \qquad (n \ge k+1)$$

  - If we know π_0, we can calculate π_n based on the above equations
    - Stopping criteria: $\left|1 - \sum_{n=0}^{N} \pi_n\right| \le \epsilon$
  - π_0: the probability that the server is idle

$$\pi_0 = 1 - Sever\ utilization\ = 1 - \frac{\lambda}{\mu}$$

M/E$_k$/1 (4)

- **Solution approach 1**
  - P_n: the probability of *n* jobs in the system

    $$P_n = \sum_{i=(n-1)k+1}^{nk} \pi_i$$

  - $\bar{N}$: Mean number of jobs in the system

    $$\bar{N} = \sum_{n=0}^{\infty} nP_n$$

  - $\bar{T}$: Mean sojourn time of a job in the system

    $$\bar{T} = \frac{\bar{N}}{\lambda}, \qquad \text{by the Little's law}$$

M/E$_k$/1 (5)

- **Solution approach 2**
  - Using a probability generating function (PGF)

$$A(z) = E[z^{\mathrm{x}}] = \sum_{n=0}^{\infty} z^n \pi_n$$

$$\boldsymbol{\lambda\pi_0 = k\mu\pi_1}$$

$$\boldsymbol{(\lambda + k\mu)\pi_n = k\mu\pi_{n+1} \quad (1 \le n \le k-1)}$$

$$\boldsymbol{(\lambda + k\mu)\pi_n = \lambda\pi_{n-k} + k\mu\pi_{n+1} \quad (n \ge k)}$$

$$\sum_{n=1}^{k-1} z^n (k\mu + \lambda)\pi_n = \sum_{n=1}^{k-1} z^n k\mu\pi_{n+1}$$

$$+ \quad \sum_{n=k}^{\infty} z^n (k\mu + \lambda)\pi_n = \sum_{n=k}^{\infty} z^n \lambda\pi_{n-k} + \sum_{n=k}^{\infty} z^n k\mu\pi_{n+1}$$

$$\sum_{n=1}^{\infty} (k\mu + \lambda) z^n \pi_n = \sum_{n=k}^{\infty} \lambda z^n \pi_{n-k} + \sum_{n=1}^{\infty} k\mu z^n \pi_{n+1}$$

M/E$_k$/1 (6)

- **Solution approach 2**

$$\sum_{n=1}^{\infty} z^n (k\mu + \lambda)\pi_n = \sum_{n=k}^{\infty} z^k z^{n-k} \lambda \pi_{n-k} + \sum_{n=1}^{\infty} \frac{z^{n+1}}{z} k\mu \pi_{n+1}$$

$$(k\mu + \lambda) \sum_{n=1}^{\infty} z^n \pi_n = \lambda z^k \sum_{n=k}^{\infty} z^{n-k} \pi_{n-k} + \frac{k\mu}{z} \sum_{n=1}^{\infty} z^{n+1} \pi_{n+1}$$

$$(k\mu + \lambda)(A(z) - \pi_0) = \lambda z^k A(z) + \frac{k\mu}{z}(A(z) - \pi_0 - z\pi_1)$$

$$\mathrm{z}(k\mu + \lambda)(A(z) - \pi_0) = \lambda z^{k+1} A(z) + k\mu(A(z) - \pi_0 - z\pi_1)$$

M/E$_k$/1 (7)

- **Solution approach 2**

Since $\lambda\pi_0 = k\mu\pi_1$,

$$A(z) = \frac{k\mu(z-1)\pi_0}{k\mu(z-1)-\lambda z\left(z^k-1\right)} = \frac{k\mu\pi_0}{k\mu-\lambda(z+z^2+z^3+\cdots+z^k)}$$

Since $A(1) = \dfrac{k\mu\pi_0}{k\mu-\lambda k} = 1,$ $\pi_0 = 1-\lambda/\mu$

$$A'(z) = \frac{k\lambda(\mu-\lambda)\,(1+2z+3z^2+4z^3\cdots+kz^{k-1})}{\{k\mu-\lambda(z+z^2+z^3\cdots+z^k)\}^2}$$

$$A'(1) = \frac{k\lambda(\mu-\lambda)\dfrac{k(k+1)}{2}}{k^2(\mu-\lambda)^2} = \frac{\lambda(k+1)}{2(\mu-\lambda)}$$

: average number of stages ($\bar{N}_{\text{stage}}$)

M/E$_k$/1 (8)

- **Solution approach 2**
  - Mean number of stages in the server

$$0 \cdot \pi_0 + (1-\pi_0)\left(\frac{1}{k}\cdot 1 + \frac{1}{k}\cdot 2 + \cdots + \frac{1}{k}\cdot k\right) = \frac{1-\pi_0}{k}\cdot\frac{k(k+1)}{2}$$

$$= \frac{\lambda}{\mu}\cdot\frac{(k+1)}{2}$$

  - Mean number of stages in the queue

$$\bar{Q}_{\text{stage}} = \bar{N}_{\text{stage}} - \frac{\lambda}{\mu}\cdot\frac{(k+1)}{2} = \frac{\lambda(k+1)}{2(\mu-\lambda)} - \frac{\lambda}{\mu}\cdot\frac{(k+1)}{2}$$

  - Mean number of jobs in the queue

$$\bar{Q}_{\text{job}} = \frac{\bar{Q}_{\text{stage}}}{k}$$

  - Mean number of jobs in the system

$$\bar{N}_{\text{job}} = \bar{Q}_{\text{job}} + \frac{\lambda}{\mu}$$

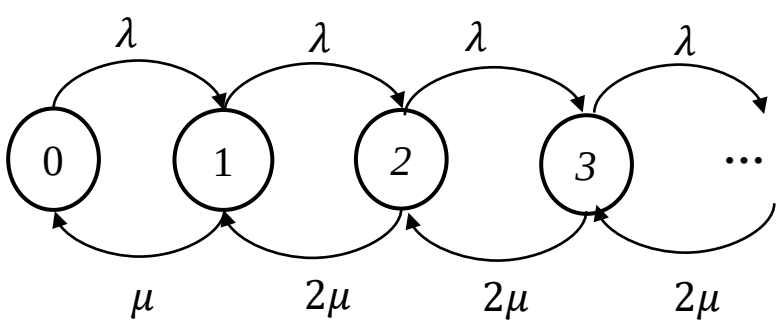
Exercise1: M/M/m

- 타자수를 고용하고자 한다. 고용 방법은 다음 두 가지 안이 있다.
  - 1안: 서비스율이 2μ 인 한 명의 타자수를 고용
  - 2안: 각각 서비스율이 μ 인 두 명의 타자수를 고용
- 한 문서를 타이핑하는데 걸리는 시간은 지수분포를 따르고, 타이핑할 문서는 rate λ 인 포아송과정으로 발생한다.
- 문서를 돌려받는데 걸리는 평균시간을 기준으로 할 때 어떤 안이 좋겠는가?

[1안]: M/M/1 (service rate: 2μ), [2안]: M/M/2 (service rate of each server: μ)

[M/M/1]: $\bar{T}_1 = \frac{1}{2\mu}\frac{1}{1-\rho} = \frac{1}{2\mu-\lambda}$

[M/M/2]:



$$\left.\begin{aligned}\lambda P_0 &= \mu P_1\\ \lambda P_{n-1} &= 2\mu P_n,\ n\geq 2\end{aligned}\right\}\quad P_0 = \frac{2\mu-\lambda}{2\mu+\lambda},\quad P_n = \frac{\lambda}{\mu}\left(\frac{\lambda}{2\mu}\right)^{n-1} P_0$$

$$\bar{T}_2 = \frac{1}{\lambda}\sum_{n=1}^{\infty} nP_n = \frac{4\mu}{4\mu^2-\lambda^2}$$

$$\bar{T}_2 - \bar{T}_1 = \frac{1}{2\mu+\lambda} > 0$$

→ 1안이 좋다.

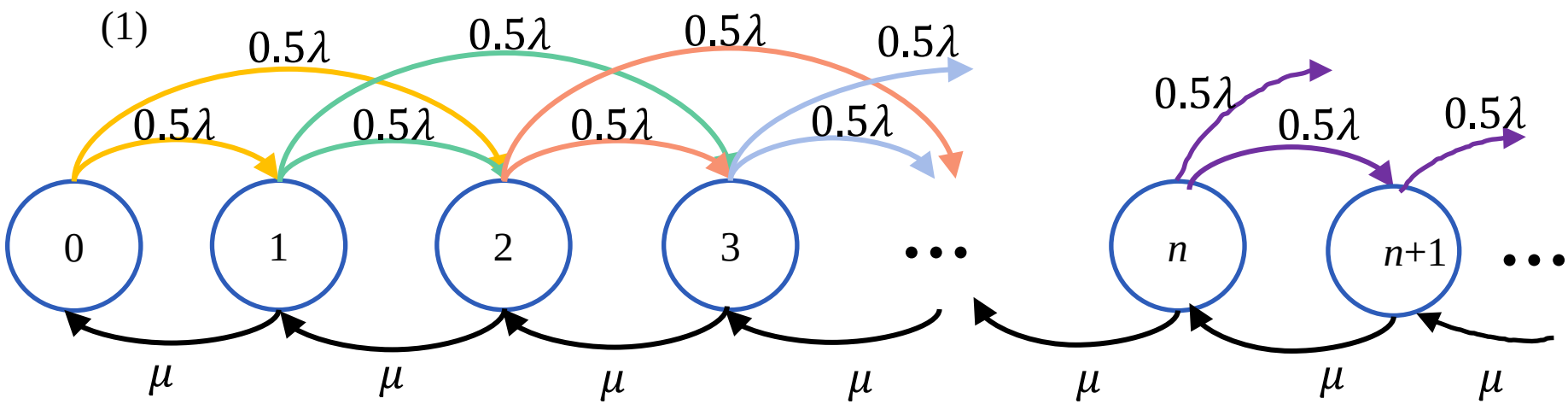
Exercise2: $M^X/M/1$

- Consider an M/M/1 system with arrival rate λ and service rate μ. At each of the arrival instants, one new customer will enter the system with probability 0.5 or two new customers will enter simultaneously with probability 0.5.

  (1) Draw the state transition rate diagram for this system

  (2) Find the expected number of customers in the system

(1)



- $\lambda P_0 = \mu P_1$
- $(\lambda + \mu)P_1 = 0.5\lambda P_0 + \mu P_2$
- $(\lambda + \mu)P_n = 0.5\lambda P_{n-2} + 0.5\lambda P_{n-1} + \mu P_{n+1}$ $\quad (n \geq 2)$

Exercise2: $M^X/M/1$

- $\lambda P_0 = \mu P_1$
- $(\lambda + \mu)P_1 = 0.5\lambda P_0 + \mu P_2$
- $(\lambda + \mu)P_n = 0.5\lambda P_{n-2} + 0.5\lambda P_{n-1} + \mu P_{n+1} \qquad (n \geq 2)$

Let $P(z) = \sum_{n=0}^{\infty} z^n P_n$.

$$\sum_{n=2}^{\infty} z^n(\lambda + \mu)P_n = \sum_{n=2}^{\infty} z^n\, 0.5\lambda P_{n-2} + \sum_{n=2}^{\infty} z^n 0.5\lambda P_{n-1} + \sum_{n=2}^{\infty} z^n\, \mu P_{n+1}$$

$$(\lambda + \mu)\sum_{n=2}^{\infty} z^n P_n = 0.5\lambda z^2 \sum_{n=2}^{\infty} z^{n-2} P_{n-2} + 0.5\lambda z \sum_{n=2}^{\infty} z^{n-1} P_{n-1} + \frac{\mu}{z}\sum_{n=2}^{\infty} z^{n+1} P_{n+1}$$

$$(\lambda + \mu)(P(z) - P_0 - zP_1) = \frac{1}{2}\lambda z^2 P(z) + \frac{1}{2}\lambda z(P(z) - P_0) + \frac{\mu}{z}(\mathrm{P(z)} - P_0 - zP_1 - z^2 P_2)$$

$$P(z) = \frac{2\mu P_0(1-z)}{\lambda z^3 + \lambda z^2 - 2\lambda z + 2\mu - 2\mu z} = \frac{2\mu P_0}{-\lambda z(z+2) + 2\mu}$$

Since $P(1) = 1$, $\frac{2\mu P_0}{-3\lambda + 2\mu} = 1$. Thus, $P_0 = 1 - \frac{3}{2} \times \frac{\lambda}{\mu}$.

$$\bar{N} = P'(1) = \frac{4\lambda}{2\mu - 3\lambda}$$

We can directly use the equation $\bar{N} = \frac{\lambda(\mathrm{E}[G] + \mathrm{E}[G^2])}{2(\mu - \lambda \mathrm{E}[G])}$ in slide 30.
In the case of this exercise, $\mathrm{E}[G] = 3/2$, $\mathrm{E}[G^2] = 5/2$