Lecture 5

Prof. Krishna R. Pattipati

Dept. of Electrical and Computer Engineering
University of Connecticut
Contact: krishna@engr.uconn.edu (860) 486-2890

EE 336

Stochastic Models for the Analysis of Computer Systems And Communication Networks

Outline

- Phase type (general Markovian) queues
  - Quasi-Birth-Death (QBD) Processes
- Why Markovian queues are simple to solve?
  - Time reversibility
- Burke's Theorem
- Introduction to Open networks

Erlang Density as Exponential Stages

- What if arrival and service processes have memory ?
- Recall sum of *m* exponential random variables is Erlang (*m*)

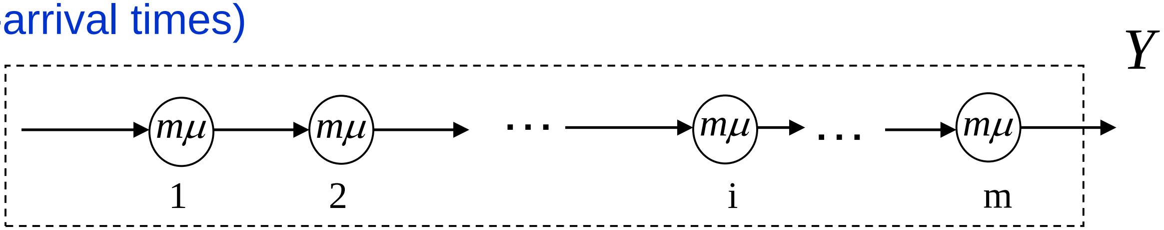
-- $x_1 \; x_2 \; \ldots \; x_m$ are i.i.d exponential with rate parameter $m\mu$

-- $Y = x_1 + x_2 + \ldots + x_m \Rightarrow f_Y(t) = \dfrac{m\mu (m\mu t)^{m-1} e^{-m\mu t}}{(m-1)!}; t \geq 0$

-- Moment generating function (MGF)

$$L_Y(s) = \left(\frac{m\mu}{s + m\mu} \right)^m \Rightarrow \bar{Y} = \frac{1}{\mu}, \sigma_Y^2 = \frac{1}{m\mu^2} \Rightarrow C_Y = \frac{1}{\sqrt{m}} < 1$$

-- Can be viewed as a ***serial combination*** of exponential stages of service (or inter-arrival times)



-- If viewed as a Markov chain, the system has (*m+1*) states, where the end state, (*m+1*) is an ***absorbing state***.

Phase Type Distributions -1

-- State transition rate matrix, Q: $(m+1)$ by $(m+1)$ matrix

$$Q = \begin{bmatrix} -m\mu & m\mu & 0 & \dots & 0 \\ 0 & -m\mu & m\mu & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & -m\mu & m\mu \\ 0 & & \cdots & & 0 \end{bmatrix} = \begin{matrix} & m & 1 \\ m & \left[A \right. & \left. \underline{c} \right] \\ 1 & \left[\underline{0}^T \right. & \left. 0 \right] \end{matrix}, \quad \underline{c} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ m\mu \end{bmatrix}$$

Since

$$Q\underline{e} = \underline{0} \Rightarrow A\underline{1} + \underline{c} = \underline{0} \Rightarrow A^{-1}\underline{c} = -\underline{1}$$

$$\underline{1}^T = [1\ \ 1\ \ 1 \cdots 1\ \ 1\ \ 1]; \quad m\ vector$$

Valid for any Markov chain with absorbing states

-- Time to absorption = inter-arrival time or service time

-- Let $\begin{pmatrix} \underline{p} \\ p_{m+1} \end{pmatrix}$ be the probability distribution of $(m+1)$ states

$$\begin{pmatrix} \underline{p}(t) \\ p_{m+1}(t) \end{pmatrix} = e^{Q^T t} \begin{bmatrix} \underline{p}(0) \\ p_{m+1}(0) \end{bmatrix} = \begin{bmatrix} e^{A^T t} & 0 \\ \underline{c}^T \int_0^t e^{A^T \sigma} d\sigma & 1 \end{bmatrix} \begin{bmatrix} \underline{p}(0) \\ p_{m+1}(0) \end{bmatrix}$$

Phase Type Distributions - 2

-- Probability of absorption at time *t*

$$p_{m+1}(t) = \underline{c}^T \left[\int_0^t e^{A^T \sigma} d\sigma \right] \underline{p}(0) + p_{m+1}(0)$$

$$= \underline{p}^T(0) \left[\int_0^t e^{A\sigma} d\sigma \right] \underline{c} + p_{m+1}(0)$$

$$= -\underline{p}^T(0) e^{At} \underline{1} + 1$$

Valid for any Markov chain with absorbing states

Since $\int_0^t e^{A\sigma} d\sigma = A^{-1}\left(e^{At} - I\right) \text{ and } A^{-1}\underline{c} = -\underline{1}$

-- The density of time to absorption is:

$$f_Y(t) = \underline{p}^T(0) e^{At} \underline{c}$$

Erlang Density as Exponential Stages - 2

-- For the Erlangian state transition matrix Q, easy to show that

$$e^{At} = \begin{bmatrix} e^{-m\mu t} & m\mu t e^{-m\mu t} & \frac{(m\mu t)^2}{2!} e^{-m\mu t} & \cdots & \frac{(m\mu t)^{m-1}}{(m-1)!} e^{-m\mu t} \\ 0 & e^{-m\mu t} & m\mu t e^{-m\mu t} & \cdots & \frac{(m\mu t)^{m-2}}{(m-2)!} e^{-m\mu t} \\ 0 & 0 & e^{-m\mu t} & \cdots & \frac{(m\mu t)^{m-3}}{(m-3)!} e^{-m\mu t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & e^{-m\mu t} \end{bmatrix}$$

$$\underline{p}^T(0) = [1 \quad 0 \quad \cdots \quad 0]$$

$$\Rightarrow f_Y(t) = \frac{m\mu (m\mu t)^{m-1} e^{-m\mu t}}{(m-1)!}; t \geq 0$$

$\Rightarrow$ By appropriately selecting Q, we can approximate any *pdf*!

Phase Type Distributions - 3

-- Moment generation function (MGF)

$$L_Y(s) = \underline{p}^T(0)(sI - A)^{-1}\underline{c}$$

$$E[Y] = \bar{Y} = -\left.\frac{dL_Y(s)}{ds}\right|_{S=0} = \underline{p}^T(0)A^{-2}\underline{c} = -\underline{p}^T(0)A^{-1}\underline{1}$$

For Erlangian case,

$$A^{-1} = -\frac{1}{m\mu}\begin{bmatrix} 1 & 1 & \cdots & \cdots & 1 \\ 0 & 1 & \cdots & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix} \Rightarrow \bar{Y} = \frac{1}{\mu} \text{ when } \underline{p}^T(0) = [1 \quad 0 \quad \cdots \quad 0]$$

In general,

$$E\left[Y^k\right] = (-1)^k k!\left(\underline{p}^T(0)A^{-k}\underline{1}\right), \quad k = 1, 2, \ldots$$

Valid for any Markov chain with absorbing states

Phase Type Distributions - 4

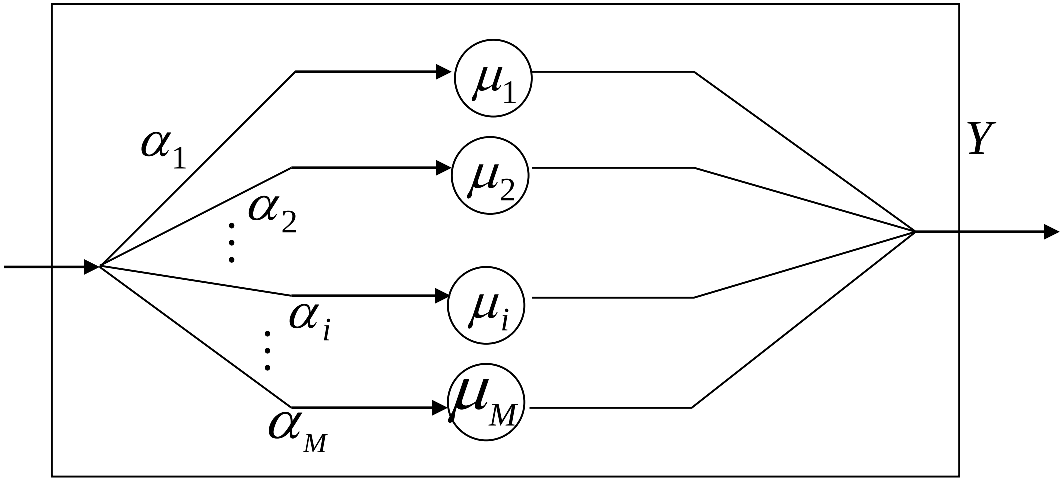
-- A computational trick

suppose
$$\left(A^k\right)^T \underline{v}_k = (-1)^k k!\,\underline{p}(0)$$
$$\left(A^{k+1}\right)^T \underline{v}_{k+1} = (-1)^{k+1}(k+1)!\,\underline{p}(0)$$
$$\Rightarrow A^T \underline{v}_{k+1} = -(k+1)\underline{v}_k;\quad \underline{v}_0 = \underline{p}(0)$$

Can solve for $\underline{v}_k$ via LU decomposition recursively

So, $$E\left[Y^k\right] = \underline{v}_k^T \underline{1}$$

- Erlang(*m*) has $C_Y < 1$. Can we generate densities with $C_Y > 1$?
- **Parallel stages**



Hyperexponential Density -1

$$L_Y(s) = \sum_{i=1}^{M} \alpha_i \left(\frac{\mu_i}{s+\mu_i} \right)$$

$$E[Y] = \sum_{i=1}^{M} \frac{\alpha_i}{\mu_i}$$

$$E[Y^2] = 2\sum_{i=1}^{M} \frac{\alpha_i}{\mu_i^2}$$

$$Q = \begin{bmatrix} -\mu_1 & 0 & 0 & \cdots & \mu_1 \\ 0 & -\mu_2 & \cdots & \cdots & \mu_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & -\mu_M & \mu_M \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

$$\underline{p}(0) = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_M]^T$$

$$C_Y^2 = \frac{E[Y^2] - \{E[Y]\}^2}{\{E[Y]\}^2} = \frac{2\sum_{i=1}^{M} \frac{\alpha_i}{\mu_i^2}}{\left(\sum_{i=1}^{M} \frac{\alpha_i}{\mu_i} \right)^2} - 1$$

$$A = Diag(\mu_i) \Rightarrow e^{At} = Diag\left(e^{-\mu_i t}\right); \underline{c} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{pmatrix}$$

$$\Rightarrow f_Y(t) = \sum_{i=1}^{M} \alpha_i \mu_i e^{-\mu_i t}; t \geq 0 \; Hyper\exp onential \; Density$$

Hyperexponential Density -2

- From Cauchy-Schwarz inequality

$$\left(\underline{a}^T \underline{b}\right)^2 \le \left(\underline{a}^T \underline{a}\right)\left(\underline{b}^T \underline{b}\right)$$

$$\underline{a} = \begin{bmatrix} \sqrt{\alpha_1} & \sqrt{\alpha_2} & \cdots & \sqrt{\alpha_M} \end{bmatrix}^T; \quad \underline{b} = \begin{bmatrix} \frac{\sqrt{\alpha_1}}{\mu_1} & \frac{\sqrt{\alpha_2}}{\mu_2} & \cdots & \frac{\sqrt{\alpha_M}}{\mu_M} \end{bmatrix}^T$$

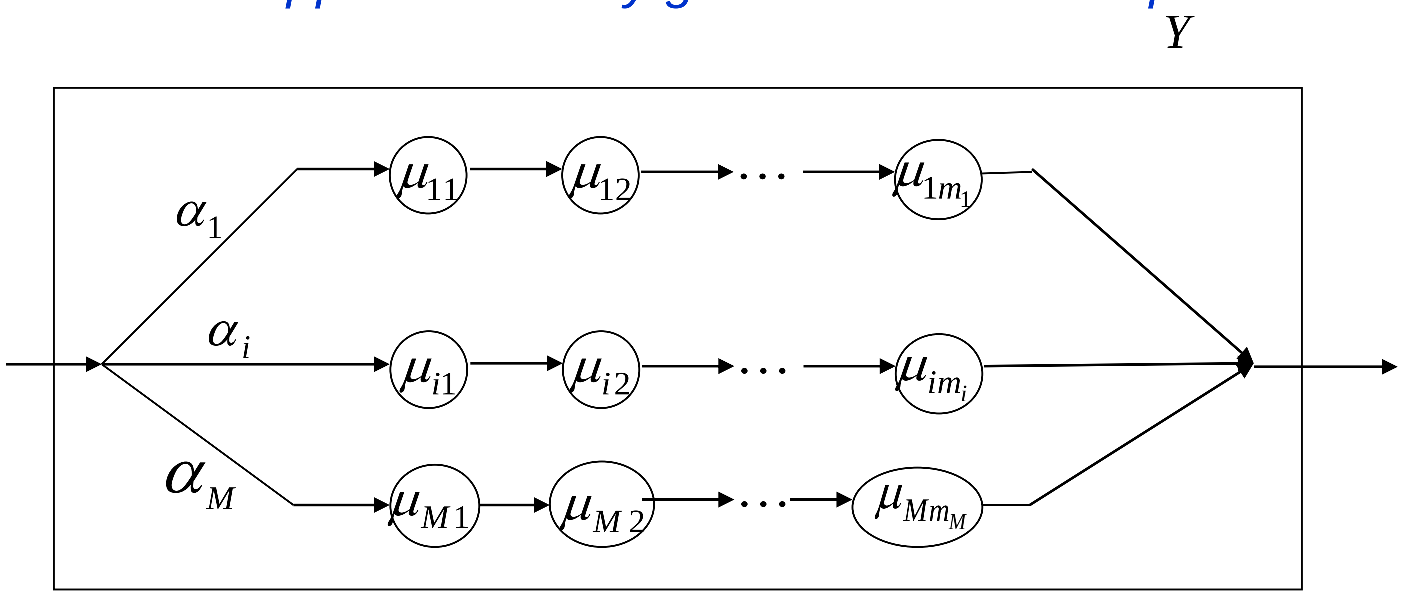
and

$$\sum_{i=1}^{M} \alpha_i = 1, \text{ we have}$$

$$C_Y^2 \ge 1$$

Series - Parallel Stages - 1

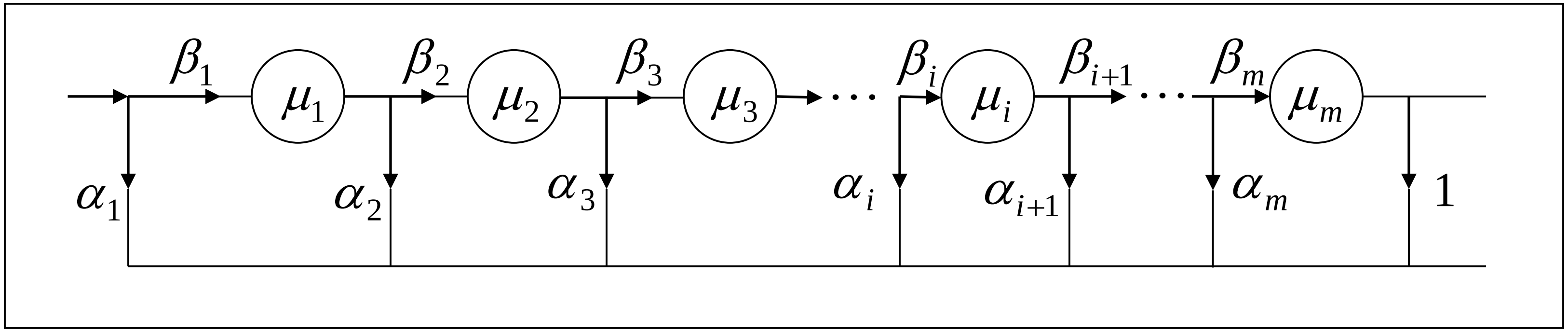
- *Indeed, series-parallel combination of exponential random variables can approximate any general distribution quite accurately*



$$L_Y(s) = \sum_{i=1}^{M} \alpha_i \prod_{j=1}^{m_i} \left(\frac{\mu_{ij}}{s+\mu_{ij}} \right)$$

Series - Parallel Stages - 2

- **Coxian representation** ~ minimal representation (~ the standard controllable form, standard observable form in control theory) Y



$$L_Y(s) = \alpha_1 + \sum_{i=1}^{m} \alpha_{i+1} \prod_{k=1}^{i} \left(\frac{\beta_k \mu_k}{s + \mu_k} \right); \ \alpha_{m+1} = 1$$

- This implies that we can solve general queues with general arrival and service processes …… but the size of state space will increase
- **The state is no longer the number of customers, n…… It should include the phase of the arrival process and the phase of the service process.**

$$(n, q, s)$$

Number of customers (n) — Phase of Arrival Process (q) — Phase of Service Process (s)

Quasi-birth-death (QBD) process

Phase Type Queues -1

- Let us consider several special cases

$$M/E_{m_S}/1,\ E_{m_a}/M/1,\ E_{m_a}/E_{m_S}/1,\ PH/PH/1$$

- $M/E_{m_S}/1$ can be viewed in one of two ways:
  - -- service time is Erlangian
  - -- bulk arrival process
  - -- state (*n, s*)
- $E_{m_a}/M/1$ can also be viewed in one of two ways:
  - -- Inter-arrival time is Erlangian
  - -- bulk service system
  - -- state (*n, a*)
- $E_{m_a}/E_{m_S}/1$ system
  - -- state (*n, a, s*)
- $PH/PH/1$ system
  - -- state (*n, a, s*)
  - -- arrival process represented by $\left(\underline{\alpha}(0), A, \underline{c}\right)$, where *A* is an $m_a \times m_a$ matrix
  - -- service process represented by $\left(\underline{\beta}(0), B, \underline{d}\right)$, where *B* is an $m_s \times m_s$ matrix

Phase Type Queues - 2

-- State space, S

$$S=\{(0,1,1),(0,2,1),\cdots,(0,m_a,1),(1,1,1),\cdots,(1,1,m_S),(1,2,1),\cdots,(1,2,m_S),$$
$$\cdots,(1,m_a,1),\cdots,(1,m_a,m_S),\cdots,(2,m_a,m_S),\cdots,(\infty,m_a,m_S)\}$$

-- Let us denote states with the same number of customers, *n* in the system are said to belong to *level n*. $m_a \times m_s$ vector

$$\underline{p}_n=\left[p_{(n,1,1)},\cdots,p_{(n,1,m_S)},p_{(n,2,1)},\cdots,p_{(n,2,m_S)},\cdots,p_{(n,m_a,1)},\cdots,p_{(n,m_a,m_S)}\right]^T$$

-- The transition rate matrix has a special *block tri-diagonal structure*

$$Q=\begin{bmatrix} B_{00} & B_{01} & 0 & \cdots & \cdots & \cdots & 0 \\ B_{10} & A_1 & A_0 & \cdots & \cdots & \cdots & 0 \\ 0 & A_2 & A_1 & A_0 & \cdots & \cdots & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Phase Type Queues - 3

- Matrix B_{00} describes state changes internal to level 0 (n=0). Since there is no customer, only changes in the arrival process is needed.

$$\boxed{B_{00} = A} \qquad m_a \times m_a \text{ matrix}$$

- A_0 describes transitions to the next higher level and includes $\underline{c}$ which indicated the rate at which the arrival process completes (i.e., transitions to absorbing state of the arrival process).

$$\boxed{A_0 = \underline{c}\,\underline{\alpha}^T_{(0)} \otimes I_{m_s}} \qquad m_a m_s \times m_a m_s \text{ matrix}$$

the factor $\underline{\alpha}(0)$ accounts for the possible change in phase in the next arrival interval

Aside:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \mathrm{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \qquad \boxed{\otimes \quad \text{Kronecker (tensor) product}}$$

$$A \otimes B = \begin{bmatrix} a & 2a & b & 2b \\ 3a & 4a & 3b & 4b \\ 5a & 6a & 5b & 6b \\ c & 2c & d & 2d \\ 3c & 4c & 3d & 4d \\ 5c & 6c & 5d & 6d \end{bmatrix}; \qquad B \otimes A = \begin{bmatrix} a & b & 2a & 2b \\ c & d & 2c & 2d \\ 3a & 3b & 4a & 4b \\ 3c & 3d & 4c & 4d \\ 5a & 5b & 6a & 6b \\ 5c & 5d & 6c & 6d \end{bmatrix}$$

$$\boxed{\begin{array}{l} A\ m\,by\,n\ and\ B\ q\,by\,r \\ \Rightarrow A \otimes B\ and\ B \otimes A\ are\ mq\,by\,nr \\ A \otimes B \neq B \otimes A \end{array}}$$

Phase Type Queues - 4

- B_{01} has a similar structure to A_0. Since it represents the arrival of the first customer after the system has been empty for some time, it needs an extra factor taking into account the phase at which the next service is started, which explains the factor $\underline{\beta}^T(0)$.

$$\boxed{B_{01} = \underline{c}\,\underline{\alpha}^T(0) \otimes \underline{\beta}^T(0)} \qquad m_a \times m_a m_s \text{ matrix}$$

- A_2 describes the rate at which services complete multiplied by the vector $\underline{\beta}^T(0)$ to account for the next service epoch.

$$\boxed{A_2 = I_{m_a} \otimes \underline{d}\,\underline{\beta}^T(0)} \qquad m_a m_s \times m_a m_s \text{ matrix}$$

- B_{10} is similar to A_2 except that it represents transition to the empty system, no new service epochs can start, so that a factor $\underline{\beta}^T(0)$ is missing.

$$\boxed{B_{10} = I_{m_a} \otimes \underline{d}} \qquad m_a m_s \times m_a \text{ matrix}$$

- A_1 describes changes in arrival or service process phase within a single level.

$$\boxed{A_1 = I_{m_a} \otimes \underline{d}\,\underline{\beta}^T(0)} \qquad m_a m_s \times m_a m_s \text{ matrix}$$

Matrix Geometric Solution -1

- The steady state equation:

$$B_{00}^T \underline{p}_0 + B_{10}^T \underline{p}_1 = \underline{0}$$

$$B_{01}^T \underline{p}_0 + A_1^T \underline{p}_1 + A_2^T \underline{p}_2 = \underline{0}$$

$$A_0^T \underline{p}_n + A_1^T \underline{p}_{n+1} + A_2^T \underline{p}_{n+2} = \underline{0} \;\; \forall n > 0$$

- A technique, termed *matrix-geometric solution*, assumes a solution of the form:

$$\underline{p}_{n+1} = R^T \underline{p}_n \qquad R : m_a m_s \times m_a m_s \text{ matrix}$$

For n>0

$$A_0^T \underline{p}_n + A_1^T R^T \underline{p}_n + A_2^T \left(R^T\right)^2 \underline{p}_n = \underline{0}$$

$$\Rightarrow R^2 A_2 + R A_1 + A_0 = 0 \quad \ldots \quad \text{Quadratic matrix equation}$$

$$\therefore R = -\left(A_0 + R^2 A_2\right) A_1^{-1}$$

or

$$R = -A_0 \left(A_1 + R A_2\right)^{-1}$$

- Iterative Algorithm

$$R(k+1) = -\left(A_0 + R^2(k) A_2\right) A_1^{-1} \text{ or } R(k+1) = -A_0\left(A_1 + R(k) A_2\right)^{-1}; R(0) = -A_0 A_1^{-1}$$

Matrix Geometric Solution - 2

- ∃Better algorithms. See books by Daigle & Haverkort and

1. A. Latouche and V. Ramaswami, " A logarithmic reduction algorithm for quasi-birth and death processes", J. of Applied Probability, vol. 30, pp. 650-674, 1993.
2. D.M Lucantoni and V. Ramaswami, "Efficient algorithms for solving the nonlinear matrix equations arising in phase-type queues", Stochastic models, vol. 1, pp. 29-51, 1996.
3. A different approach: L. Lipsky, Queuing Theory: A Linear Algebraic Approach, McMillan, 1992.

- Once *R* is known, can get $\underline{p}_n$ as follows:

Know

$$\begin{bmatrix} B_{00}^T & B_{10}^T \\ B_{01}^T & A_1^T + A_2^T R \end{bmatrix} \begin{bmatrix} \underline{p}_0 \\ \underline{p}_1 \end{bmatrix} = \begin{bmatrix} \underline{0} \\ \underline{0} \end{bmatrix}$$

also

$$\sum_{n=0}^{\infty} \underline{p}_n^T \underline{1} = \underline{p}_0^T \underline{1} + \sum_{n=1}^{\infty} \underline{p}_n^T \underline{1}$$

$$= \underline{p}_0^T \underline{1} + \underline{p}_1^T \sum_{n=0}^{\infty} R^n \underline{1} = \underline{p}_0^T \underline{1} + \underline{p}_1^T (I - R)^{-1} \underline{1} = 1$$

Perf. Measures of Phase Type Queues -1

- Performance measures
  - -- Average inter-arrival time: $E(\tau) = -\underline{\alpha}^T(0)A\underline{1}, \quad \lambda = \frac{1}{E(\tau)}$
  - -- Average service time: $E(S) = -\underline{\beta}^T(0)B\underline{1}, \quad \mu = \frac{1}{E(S)}$
  - -- Utilization: $U = \frac{\lambda}{\mu} = \rho = \frac{E(s)}{E(\tau)}$
  - -- Average system length:

$$E(n) = Q = \sum_{n=1}^{\infty} n\,\underline{p}_n^T \underline{1} = \underline{p}_1^T \left(\sum_{n=1}^{\infty} nR^{n-1} \right) \underline{1}$$
$$= \underline{p}_1^T (I - R)^{-2} \underline{1}$$

  - -- Average response time:

$$R = \frac{Q}{\lambda} = \frac{-\underline{p}_1^T (I-R)^{-2}\underline{1}}{\underline{\alpha}^T(0)A\underline{1}}$$

Perf. Measures of Phase Type Queues - 2

- Performance measures

  -- Average waiting queue length:

$$Q_w = Q - \rho$$

  -- Average waiting time: $W = R - \frac{1}{\mu} = R + \underline{\beta}^T(0)B\underline{1}$

  -- $p_n = \underline{p}_n^T \underline{1}$

  -- $P(n \geq k) = \sum_{n=k}^{\infty} \underline{p}_n^T \underline{1} = \underline{p}_1^T \left(\sum_{n=k}^{\infty} R^{n-1} \right) \underline{1}$

$$= \underline{p}_1^T R^{k-1} (I - R)^{-1} \underline{1}$$

Time Reversibility - 1

- Why is M/M/1, M|M|1|N, M/M/1, M/M/m analysis so simple? Because
  - they are special cases of birth-death processes
  - they satisfy local (detailed) balance equations

$$\lambda_{n-1} p_{n-1} = \mu_n p_n$$

- What does it all mean?
  - B-D processes are time-reversible Markov chains. Consider a discrete-time Markov chain $\{X_n\}$ where $X_n \in (0, 1, 2, ...N),\ n = 0, 1, 2, ...$ with stationary transition probabilities $P\{X_{n+1} = j | X_n = i\} = P_{ij}$
  - Let $p_j(n) = P(X_n = j)$ and $p_j = \lim_{n \to \infty} p_j(n),\ j = 0, 1, 2, ..., N$
  - Markov chain is homogeneous, irreducible and aperiodic) $\{p_j\}$ exists
  - This Markov chain is running forward in time. Suppose we want to run this chain backward in time $\overleftarrow{n-2 \quad n-1 \quad n}$
  - How do we describe the evolution of $X_n, X_{n-1}, ...,$ the reversed chain?
  - Key: The reversed chain is also a DTMC. Suppose it is Markov:

Time Reversibility - 2

$$P^*_{ij} = P\{X_n = j | X_{n+1} = i\}$$
$$= \frac{P\{X_{n+1} = i | X_n = j\} P\{X_n = j\}}{P\{X_{n+1} = i\}}$$

$$P^*_{ij} = \frac{p_j}{p_i} \cdot P_{ji}$$

Note that time reversibility requires homogeneity (or stationarity)

- If Markov, we need

$$P\{X_n = j | X_{n+1} = i, X_{n+2} = i_2, ..., X_{n+k} = i_k\} = P^*_{ij}$$

Proof:

$$P\{X_n = j | X_{n+1} = i, X_{n+2} = i_2, ..., X_{n+k} = i_k\}$$

$$= \frac{P\{X_n = j, X_{n+1} = i, X_{n+2} = i_2, ..., X_{n+k} = i_k\}}{P\{X_{n+1} = i, X_{n+2} = i_2, ..., X_{n+k} = i_k\}}$$

$$= \frac{P\{X_n = j\} \cdot P\{X_{n+1} = i | X_n = j\} ... P\{X_{n+k} = i_k | X_{n+k-1} = i_{k-1}\}}{P\{X_{n+1} = i\} \cdot P\{X_{n+2} = i_2 | X_{n+1} = i\} ... P\{X_{n+k} = i_k | X_{n+k-1} = i_{k-1}\}}$$

Time Reversibility - 3

$$P_{ij}^* = \frac{p_j}{p_i} \cdot P_{ji}$$

$\Rightarrow$ Reversed process is also a Markov chain with transition probabilities

$$\boxed{P_{ij}^* = \frac{p_j}{p_i} \cdot P_{ji}}$$

- Definition: A Markov chain is time-reversible if $\boxed{P_{ij}^* = P_{ij}}$

  Transition probabilities of backward & forward chains are identical

$$\Rightarrow \quad p_i P_{ij} = p_j P_{ji}$$

Proportion of transitions from state i to state j = Proportion of transitions from state j to state i

detailed balance or local balance eqns.

- Properties:

  1. The reversed chain is *irreducible and aperiodic*. *It has the same steady state distribution as the forward chain*.

Time Reversibility - 4

$$p_j^* = \sum_i p_i^* P_{ij}^* = \sum_i p_i^* \frac{P_{ji} p_j}{p_i}$$

$$= \sum_i p_i^* P_{ij}$$ **From time-reversibility,** $P_{ij} = P_{ij}^*$

$$\Rightarrow p_j^* = p_j \ \forall j$$

2. If we can find $\{p_i, i \geq 0 | p_i \geq 0; \sum_i p_i = 1\}$ & find a transition probability matrix $[P_{ij}^*]$ such that:

$$p_i P_{ij}^* = p_j P_{ji}, \quad i \geq 0$$

Then $\{p_i\}$ is the steady state distribution of the forward **and** reversed chains. $[P_{ij}^*]$ is the transition probability matrix of the reversed chain.

Proof:

$$\sum_j p_j P_{ji} = p_i \sum_j P_{ij}^* = p_i$$

Note that we did not assume time-reversibility

- The above ideas extend naturally to continuous-time Markov chains. Transition rates

$$q_{ij} = \frac{\partial P_{ij}}{\partial t}; \quad q_{ii} = -\frac{\partial P_{ii}}{\partial t}$$

Time Reversibility - 5

- The reversed chain is Markov with the same steady-state distribution with transition rates:

$$q_{ij}^{*} = \frac{p_j q_{ji}}{p_i} \quad \forall i, j \geq 0$$

- If we can find p_i & q_{ij}^{*} $\ni$

$$p_i q_{ij}^{*} = p_j q_{ji} \;\; and \sum_i p_i = 1$$

then

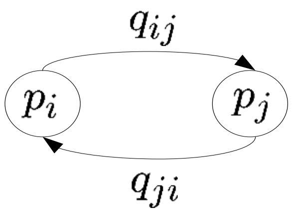
q_{ij}^{*} are the transition rates of the reversed chain &

$\{p_i | i \geq 0\}$ is the steady-state distribution of **both forward & reversed chains**

- The forward chain is time-reversible iff its steady-state distribution & transition rates satisfy the detailed (local) balance equations

$$p_i q_{ij} = p_j q_{ji} \quad \forall i, j$$

$$\Rightarrow \boxed{q_{ij}^{*} = q_{ij}}$$



In particular, for B-D processes

$$p_n \mu_n = p_{n-1} \lambda_{n-1} \;\; \Rightarrow q_{n,n+1} = \lambda_n ; \; q_{n+1,n} = \mu_{n+1}$$

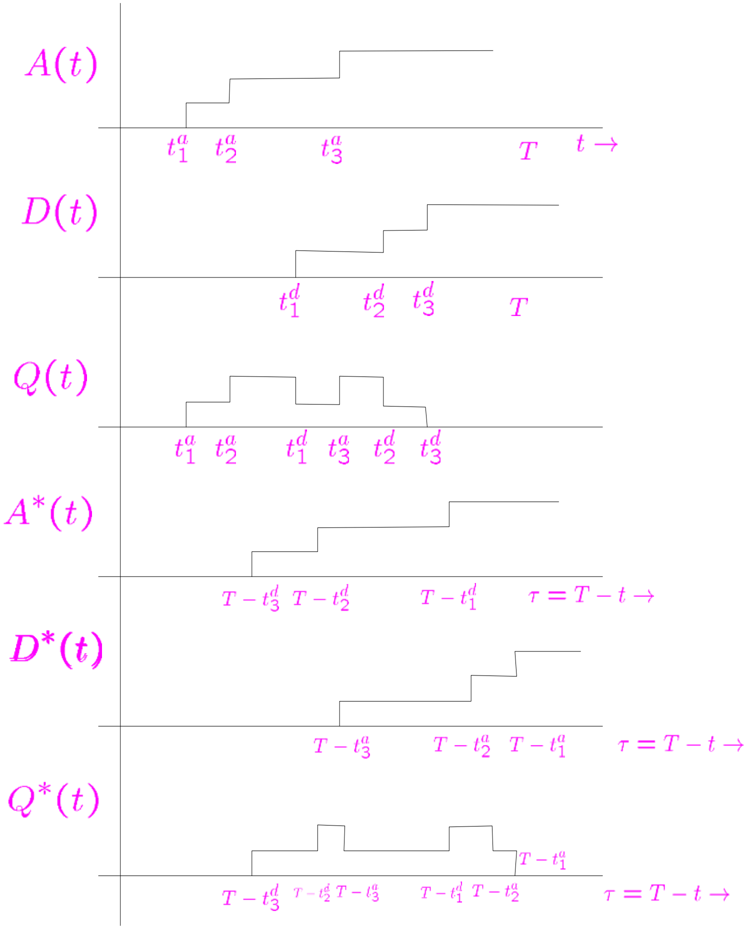
Markovian Queues and Time Reversibility -1

- Since M/M/1, M/M/m, M/M/1/N, M/M/1 , M/M/m/m etc. are all B-D processes, they are time reversible. Physically, what this means is that:

Departure process of forward system =	Arrival process of reversed system
Arrival process of forward system =	Departure process of reversed system

Forward and reversed processes are indistinguishable in the steady-state

Markovian Queues and Time Reversibility -2



Burke's Theorem -1

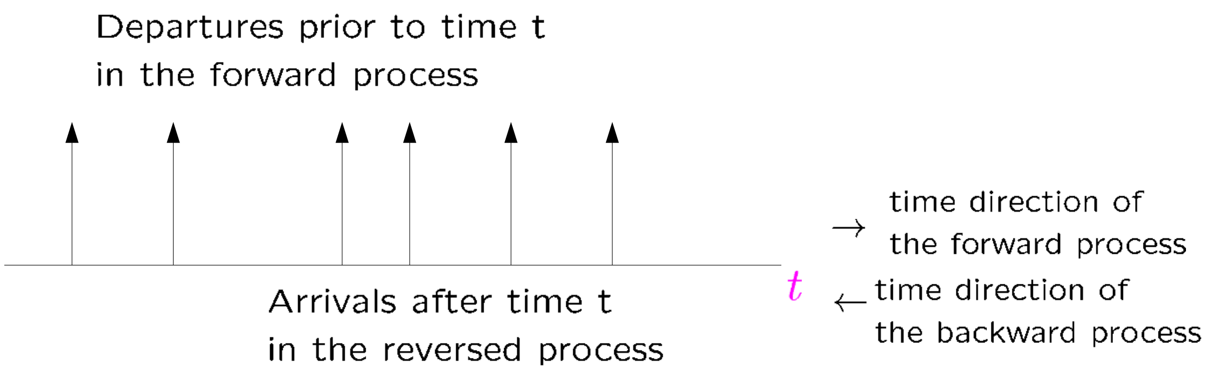
- Burke's Theorem: For M/M/1, M/M/m and M/M/1 systems with arrival rate λ , suppose we start the system in the steady state, then
  - The departure process is Poisson with rate λ
  - At each time *t*, the number of customers in the system is independent of the sequence of departure times prior to time *t*
  - If customers are served in the order they arrive, then given that a customer departs at time *t*, the arrival time of that customer is independent of the departure process prior to time *t*
- Proof of 1): Departure process is Poisson with rate λ
  - Processes in M/M/1, M/M/M & M/M/1 are time-reversible Markov chains. Know that:
    - the forward and reverse chains are statistically indistinguishable in the steady state.
    - the *departure process in the forward chain is the arrival process in the reverse system*. The only way this can happen is if the departure process is Poisson with rate λ.

Burke's Theorem - 2

- Alternative Proof for M/M/1:

$$
\begin{aligned}
D^*(s)|_{\text{nonempty}} &= \frac{\mu}{s+\mu}; \quad D^*(s)|_{\text{empty}} = \frac{\lambda}{s+\lambda}\cdot\frac{\mu}{s+\mu} \\
\therefore\ D^*(s) &= \frac{\lambda}{\mu}\cdot\frac{\mu}{s+\mu} + \frac{\lambda}{s+\lambda}\cdot\frac{\mu}{s+\mu}\left(1-\frac{\lambda}{\mu}\right) \\
&= \frac{\lambda}{\mu}\cdot\frac{\mu}{s+\mu}\left[1-\frac{\lambda}{s+\lambda}\right] + \frac{\lambda\mu}{(s+\mu)(s+\lambda)} \\
&= \frac{\lambda s}{(s+\mu)(s+\lambda)} + \frac{\lambda\mu}{(s+\mu)(s+\lambda)} = \frac{\lambda}{s+\lambda}!!
\end{aligned}
$$

- Proof of 2: # of customers independent of prior departure times

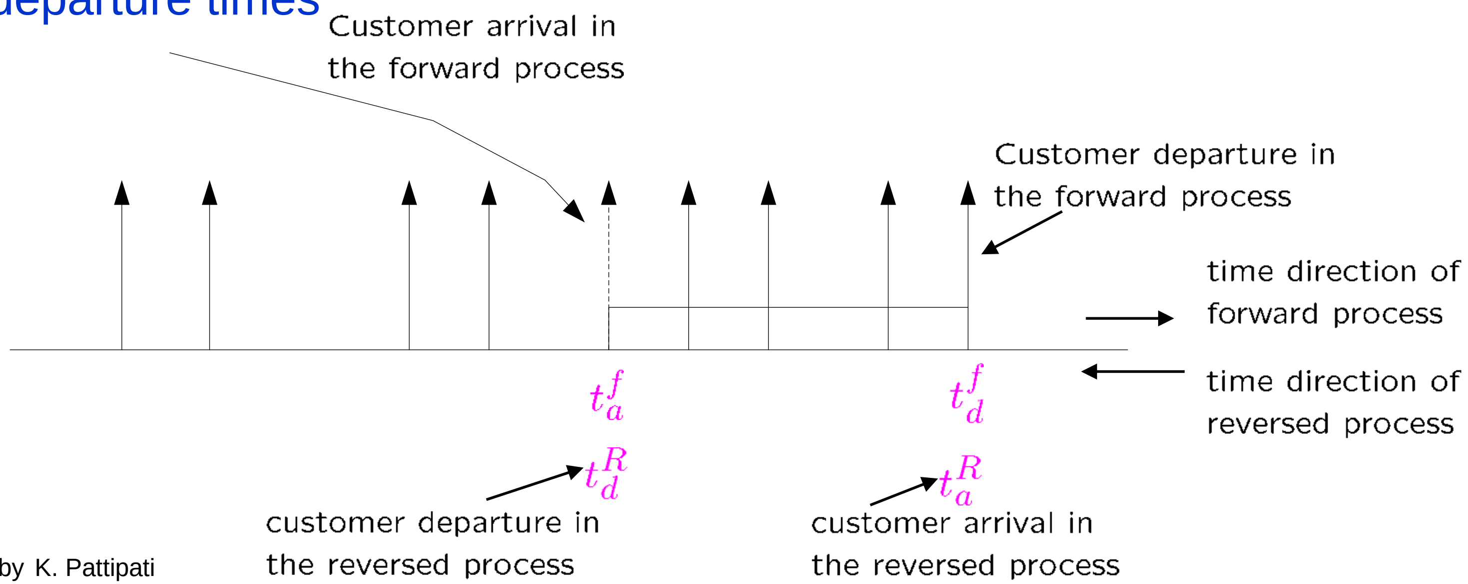


Burke's Theorem - 3

- Departures prior to time *t* in the forward process $=$ Arrivals after time *t* in the reversed process
- Arrivals in the reversed system are independent & Poisson $\Rightarrow$ Future arrival process is independent of # of customers in the system in the reversed process

$\Rightarrow$ • Past departure process is independent of the # of customers in the system

❑ Proof of # 3: Arrival time of a departing customer is independent of prior departure times



Burke's Theorem - 4

- In the reversed system, the time spent by a customer arriving at $t_a^R (= t_d^f)$ is not affected by customers arriving after time t_a^R (i.e, to the left of t_a^R). In the forward system, this means that the time spent in the system $(t_d^f - t_a^f)$ is independent of the departure process prior to the customer's departure

- M|M|m queue (cute proof)
  - Let d_n be the *mean* number of departures over a small interval Δ when there are n customers in the system
  - Clearly,

$$d_n = E\{number\ of\ departures\ over\ \Delta t \mid n\} \begin{cases} 0; n = 0 \\ n\mu\Delta t; 1 \le n \le m \\ m\mu\Delta t; n > m \end{cases}$$

  - Expected number of departures, d

$$d = \sum_{n=0}^{\infty} d_n p_n = \Delta t [\sum_{n=1}^{m-1} n\mu p_n + m\mu \sum_{n=m}^{\infty} p_n]$$

$$= \lambda \Delta t [\sum_{n=1}^{m-1} p_{n-1} + \sum_{n=m}^{\infty} p_{n-1}] = \lambda \Delta t \Rightarrow Poisson$$

Departure Process of of M|M|1|N Queue

- Because of blocking, output is not Poisson!
- Recall a departing customer sees a system with himself removed.

$$D^*(s)|_{n=0} = \left(\frac{\lambda}{s+\lambda}\right)\left(\frac{\mu}{s+\mu}\right); D^*(s)|_{n=1,2,..,N-1} = \left(\frac{\mu}{s+\mu}\right)$$

$$D^*(s) = \frac{1-\rho}{1-\rho^N}\left(\frac{\lambda}{s+\lambda}\right)\left(\frac{\mu}{s+\mu}\right) + \frac{\rho(1-\rho^{N-1})}{1-\rho^N}\left(\frac{\mu}{s+\mu}\right)$$

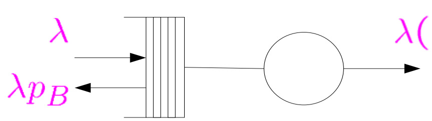
$$= \frac{1}{1-\rho^N}[(1-\rho)\left(\frac{\lambda}{s+\lambda}\right)\left(\frac{\mu}{s+\mu}\right) + \rho\left(\frac{\mu}{s+\mu}\right)] - \frac{\rho^N}{1-\rho^N}\left(\frac{\mu}{s+\mu}\right)$$

$$= \frac{1}{1-\rho^N}[\left(\frac{\lambda}{s+\lambda}\right) - \rho^N\left(\frac{\mu}{s+\mu}\right)]$$

$$f_D(t) = \frac{1}{1-\rho^N}\lambda e^{-\lambda t} - \frac{\rho^N}{1-\rho^N}\mu e^{-\mu t}; t \ge 0$$

$$E(D) == \frac{1}{1-\rho^N}\frac{1}{\lambda} - \frac{\rho^N}{1-\rho^N}\frac{1}{\mu} = \frac{1}{\lambda}\left(\frac{1-\rho^{N+1}}{1-\rho^N}\right) = \frac{1}{\lambda(1-P_B)}$$

$$\sigma_D^2 == \frac{1}{(1-\rho^N)^2}\frac{1}{\lambda^2}[(1+\rho^{N+1})^2 - 2\rho^N(1+\rho^2)]$$

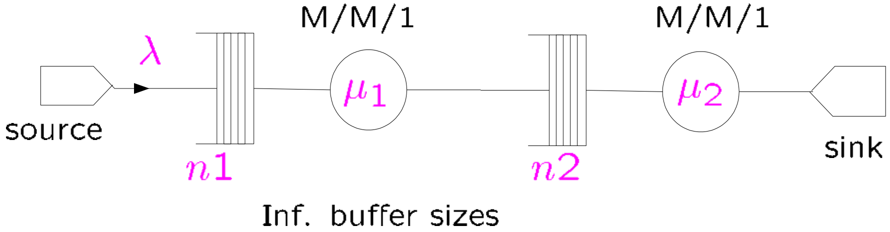


$\lambda(1-p_B) = throughput = \mu(1-p_0)$

Two-stage Tandem Queue

Two stage tandem queue

- simple network or sequential (pipeline or assembly) network



- The service times at the two nodes (servers) are exponentially distributed and mutually independent
- Can show that

$$p(n_1, n_2) = p(n_1)\cdot p(n_2) = \left(1-\frac{\lambda}{\mu_1}\right)\left(\frac{\lambda}{\mu_1}\right)^{n_1}\left(1-\frac{\lambda}{\mu_2}\right)\left(\frac{\lambda}{\mu_2}\right)^{n_2}$$
$$= (1-\rho_1)\rho_1^{n_1}(1-\rho_2)\rho_2^{n_2}$$

- We provide only intuitive proof. We will provide proof in a more general context later.

Product Form of Joint pmf -1

- Node 1 is an M/M/1 queue. Using Burke's theorem, the departure process from node 1 is Poisson with rate λ and independent of service time of node 2.

$$\begin{aligned} p(n_1 \text{ at node } 1) &= (1-\rho_1)\rho_1^{n_1} \\ p(n_2 \text{ at node } 2) &= (1-\rho_2)\rho_2^{n_2} \end{aligned}$$

- Burke's theorem also tells us that the number of customers at node 1 is independent of the sequence of earlier departures from node 1 (or equivalently, the sequence of earlier arrivals at node 2)

$\Rightarrow$ the number of customers at node 1 are independent of the number of customers at node 2

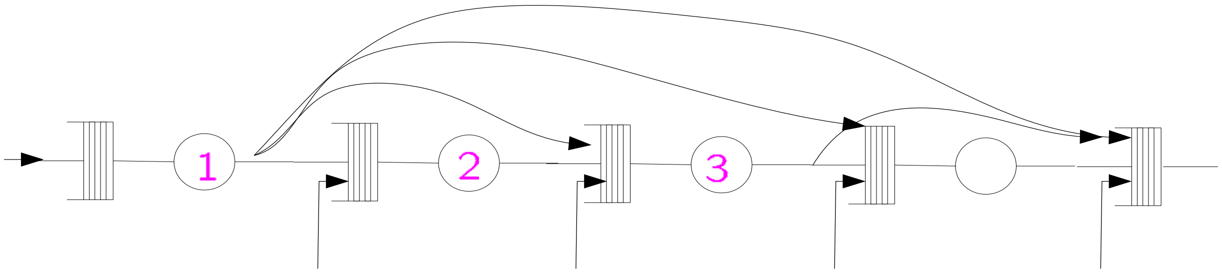
$$p(n_1, n_2) = p(n_1) \cdot p(n_2) = (1-\rho_1)\rho_1^{n_1}(1-\rho_2)\rho_2^{n_2}$$

- Performance measures

node 1	node2	overall
$Q = \frac{\rho_1}{1-\rho_1}$	$Q_2 = \frac{\rho_2}{1-\rho_2}$	$Q = Q_1 + Q_2$
$X = \lambda$	$X = \lambda$	$X = \lambda$
$R_1 = \frac{1}{\mu_1(1-\rho_1)}$	$R_2 = \frac{1}{\mu_2(1-\rho_2)}$	$R = R_1 + R_2$
$U_1 = \rho_1$	$U_2 = \rho_2$	$bottleneck = max(U_1, U_2)$

Product Form of Joint pmf - 2

- Result is similar to M/M/1 queue, node with larger λ/μ is the bottleneck in the system
- the result above extends to any feedforward network, since the outputs of M/M/m queues are Poisson. Indeed the result is true for any **acyclic** network.



no feedback paths

$\Rightarrow$ what happens when we have feedback paths?

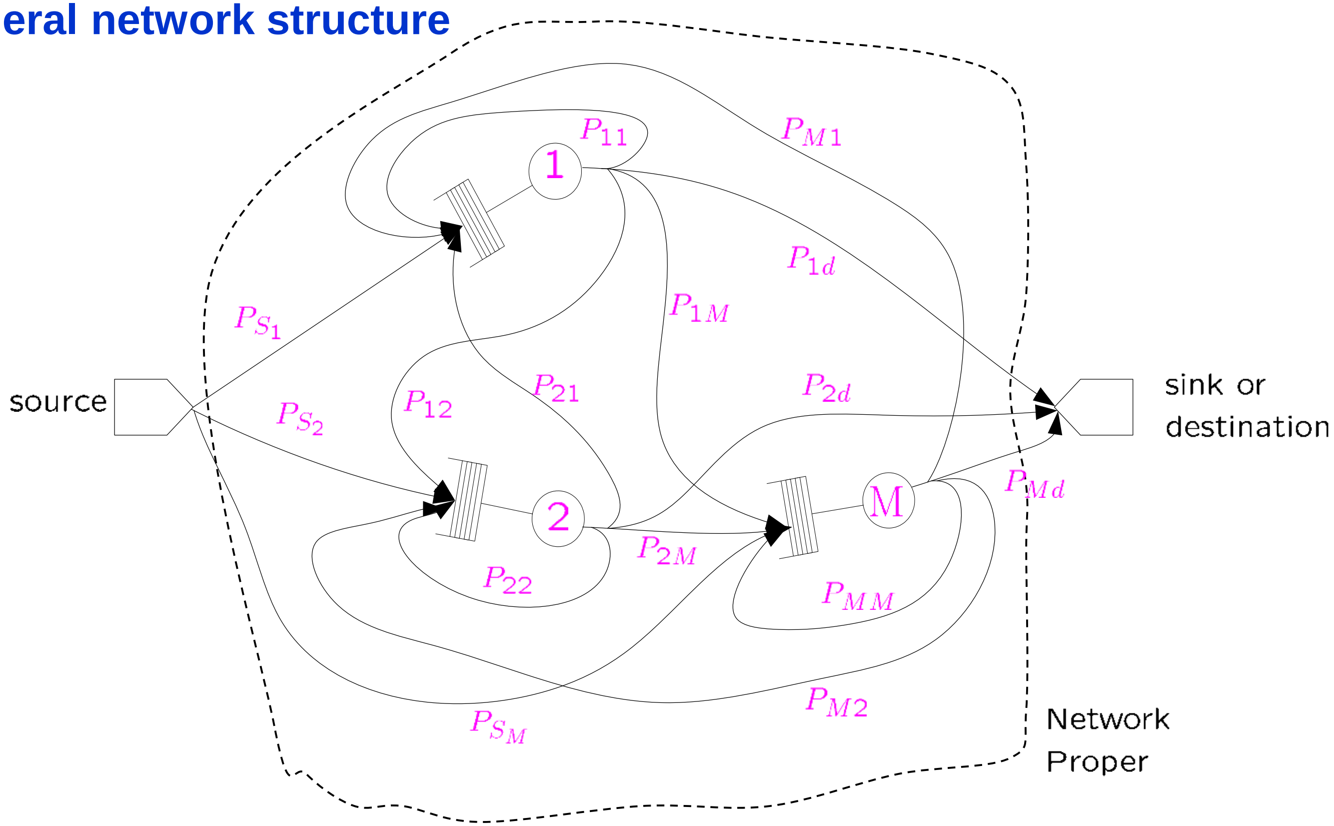
JACKSON NETWORKS

Arrivals at each node are not Poisson, but they behave as if they were!!

PRODUCT FORM still holds!!

General Network Structure -1

❑ General network structure



- A general network is a <u>directed</u> graph G=(V,E)
  - V={1,2,…,M} is the set of vertices or nodes or service stations
  - E = V £ V is the set of edges (arcs) that link the vertices

Inf. server
single server
multiple server
state dep. node

General Network Structure - 2

= ordered pairs of the form (i, k) where $i \in V,\ k \in V$ and (i, k) denotes a link from node i to node k

- Stochastic process: queue lengths at each node i

$$[n_1(t), n_2(t), ..., n_M(t)]$$

$$n_i(t) = \text{number of customers at node } i \text{ at time } t.$$

We are interested only in the steady state.

$$[n_1, n_2, ..., n_M]$$

- **Closed (Gordon-Newell) Network:** $N = n_1 + n_2 + ..., +n_M = \text{Constant}$
- **Open (Jackson) network**:
  - customers enter the network proper from a fictitious Poisson source node. Arrival rate at node $i,\ r_i = p_{s_i}\lambda$. Since the merging and decomposition of a Poisson process is Poisson, no loss of generality. Have a sink that absorbs all customers who are departing from the network proper.
  - Total arrival rate λ. Individual stream at each node i is also Poisson with rate $r_i = \lambda p_{s_i}$
  - Service demand at each node i , s_i has Exponential distribution.
  - Service rates $\mu_i(n)$

Routing in a Jackson Network

$$\mu_i(n) = \begin{cases} \mu_i & \text{single server} \\ n\mu_i & \text{infinite server} \\ min(n, m_i)\mu_i & \text{multiple server} \end{cases}$$

- Any work conserving queuing discipline → serve when customer is present; → service demand is independent of queuing disciplline

- The routing of customers in the network proper is governed by a first-order Markov chain

$$P_{ij} = \text{Prob}\{\text{customer departing node } i \text{ will go next to node } j\}$$

$$j = 1, 2, ..., M, d$$ → absorbing state

Markov chain is irreducible and aperiodic $\Longrightarrow$ steady state distribution exists.

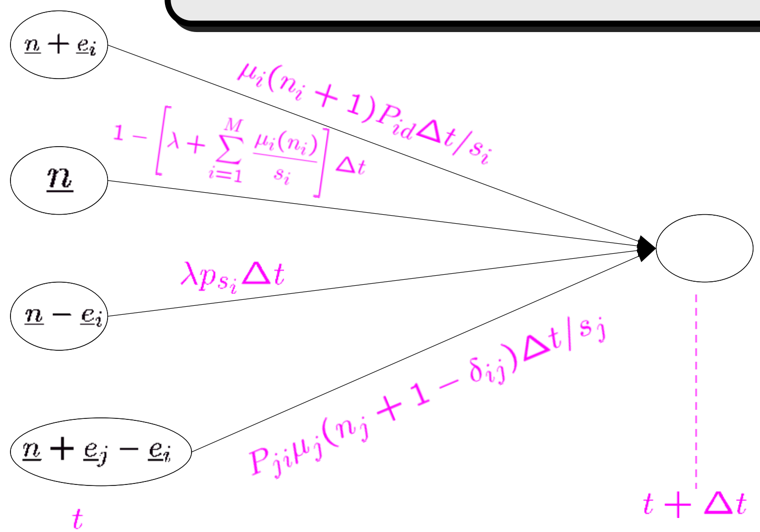
$$v_i = p_{s_i} + \sum_{j=1}^{M} P_{ji} v_j; \quad 1 \le i \le M$$

$$\underline{v} = \underline{p}_s + P^T \underline{v} \quad \Rightarrow \quad \underline{v} = (I - P^T)^{-1} \underline{p}_s$$

$$\Rightarrow p_{s_i} = v_i - \sum_{j=1}^{M} P_{ji} v_j; \quad 1 \le j \le M$$

$$p(\underline{n}) = \text{Probability that the system is in state } \underline{n} = (n_1, n_2, ..., n_M)$$

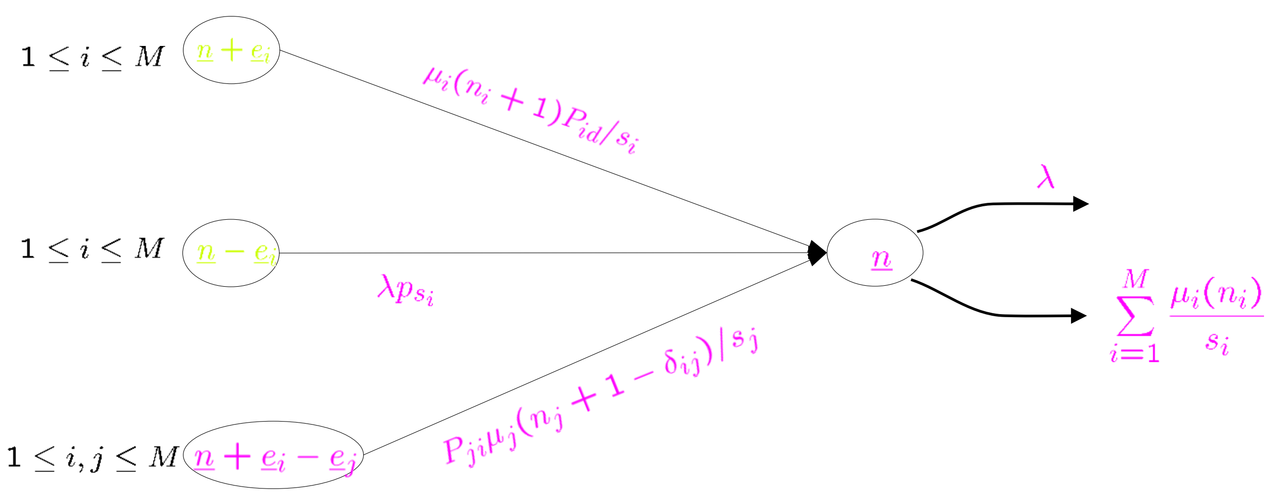
Transition Rate Diagram



i^{th} place

$$\underline{e}_i = (0, 0, ..., 1, ...0)$$

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0; \textit{otherwise} \end{cases}$$



Steady State Probability Equations

$$\left[\lambda + \sum_{i=1}^{M} \frac{\mu_i(n_i)}{s_i}\right] p(\underline{n}) = \sum_{i=1}^{M} P_{id} \frac{\mu_i(n_i+1)}{s_i} p(\underline{n}+\underline{e}_i)$$
$$+ \lambda \sum_{i=1}^{M} p_{s_i} p(\underline{n}-\underline{e}_i)$$
$$+ \sum_{i=1}^{M} \sum_{j=1}^{M} P_{ji} \frac{\mu_j(n_j+1-\delta_{ij})}{s_j} p(\underline{n}+\underline{e}_j-\underline{e}_i)$$

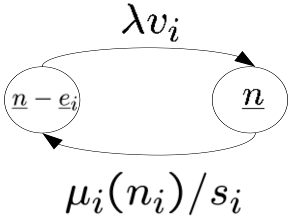
Substitute:

$$p_{s_i} = v_i - \sum_{j=1}^{M} P_{ji} v_j; \quad 1 \leq i \leq M$$

$$\underbrace{\lambda p(\underline{n}) - \sum_{i=1}^{M} P_{id} \frac{\mu_i(n_i+1)}{s_i} p(\underline{n}+\underline{e}_i)}_{R(\underline{n})}$$
$$= -\sum_{i=1}^{M} \left\{ \underbrace{\frac{\mu_i(n_i)}{s_i} p(\underline{n}) - \lambda v_i p(\underline{n}-\underline{e}_i)}_{B_i(\underline{n})} \right\}$$
$$+ \sum_{i=1}^{M} \sum_{j=1}^{M} P_{ji} \underbrace{\left[\frac{\mu_j(n_j+1-\delta_{ij})}{s_j} p(\underline{n}+\underline{e}_j-\underline{e}_i) - \lambda v_j p(\underline{n}-\underline{e}_i) \right]}_{B_j(\underline{n}+\underline{e}_j-\underline{e}_i)}$$

Product Form of S.S. Probabilities -1

$$B_i(\underline{n}) = 0 \quad \forall \underline{n}$$
$$\Downarrow$$
$$R(\underline{n}) = 0 \quad \forall \underline{n}$$



$\mu_i p(\underline{n}) = \lambda v_i s_i p(\underline{n} - \underline{e}_i)$...Local (detailed) balance equations

$$\Rightarrow \quad p(\underline{n}) = p(n_1, n_2, ..., n_M) = p(\underline{o}) \prod_{i=1}^{M} \frac{(\lambda v_i s_i)^{n_i}}{\prod_{k=1}^{n_i} \mu_i(k)}$$
$$= p(\underline{o}) \prod_{i=1}^{M} Y_i(n_i) = \prod_{i=1}^{M} p_i(n_i)$$

$$Y_i(n_i) = \prod_{k=1}^{n_i} \frac{\lambda v_i s_i}{\mu_i(k)}$$

λ = total arrival rate

v_i = # of visits to node i

$v_i s_i$ = **total** service demand at node i

Stable if $p(\underline{0})$ is non-zero $\Longrightarrow$ ergodic Markov process. Steady state solution exists $\Rightarrow p_i(0)$ is nonzero.

$$p_i(0) = \left[\sum_{n=0}^{\infty} \frac{(\lambda v_i s_i)^n}{\prod_{k=1}^{n} \mu_i(k)} \right]^{-1} > 0$$

Product Form of S.S. Probabilities - 2

- The *stationary distribution of the network state is the product of the marginal distributions at each node i* $\Rightarrow$ Product form
  - Known as *Jackson's Decomposition Theorem*

- Individual nodes behave as if they are M|M|SD queues with rate λv_i and service time per visit $\frac{s_i}{\mu_i(n)}$

Summary

- Phase type (general Markovian) queues
  - Quasi-Birth-Death (QBD) Processes
- Why Markovian queues simple to solve?
  - Time reversibility
- Burke's Theorem
- Product form of steady state distribution in open networks