LECTURE #12
BULK SYSTEM – NETWORK OF Q

204528

Queueing Theory and Applications in Networks

Assoc. Prof. Anan Phonphoem, Ph.D. (รศ.ดร. อนันต์ ผลเพิ่ม)

Computer Engineering Department, Kasetsart University

Outline

- Bulk System
  - Bulk arrival
  - Bulk service
- Parallel and serial servers
- Network of Queue

3

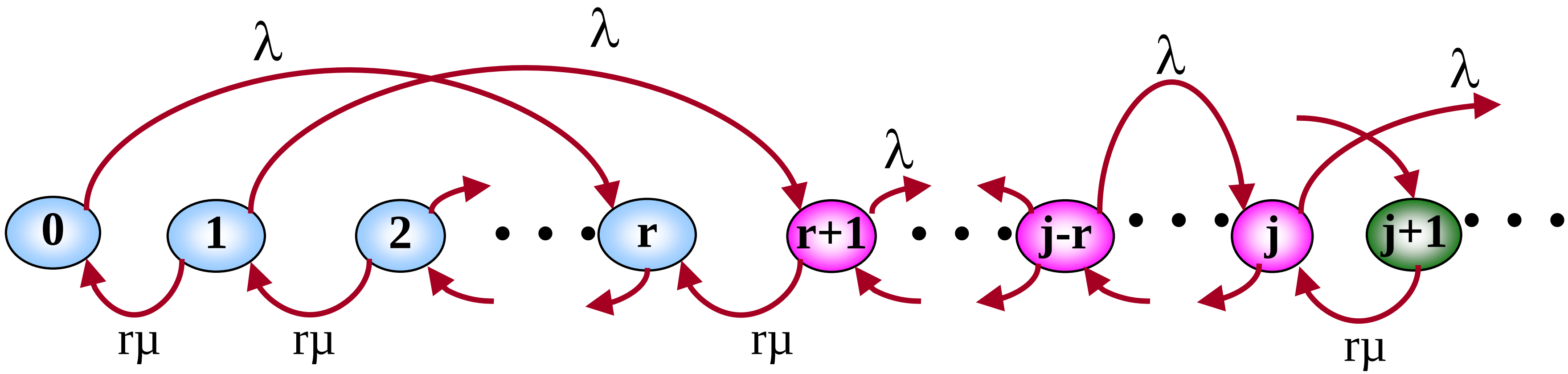
Bulk System

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Arrival Systems

- The arrival of **r** customers
- Each of **r** customers requires only a single stage of service
- M/M/1 with bulk arrivals of size **r**

Bulk Arrival State diagram



- Same as $M/E_r/1$
- Differences:
  - For $M/E_r/1$
    - State variable = total # of service stage yet to be completed
  - For Bulk Arrival
    - State variable = the number of customers in the system

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Arrival Systems

- The result from $M/E_r/1$

$$P(z) = \frac{r\mu(1-\rho)(1-z)}{r\mu + \lambda z^{r+1} - (\lambda + r\mu)z}$$

$$p_j = (1-\rho)\sum_{i=1}^{r} A_i(z_i)^{-j} \qquad j = 1, 2, \ldots, r$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Arrival Systems

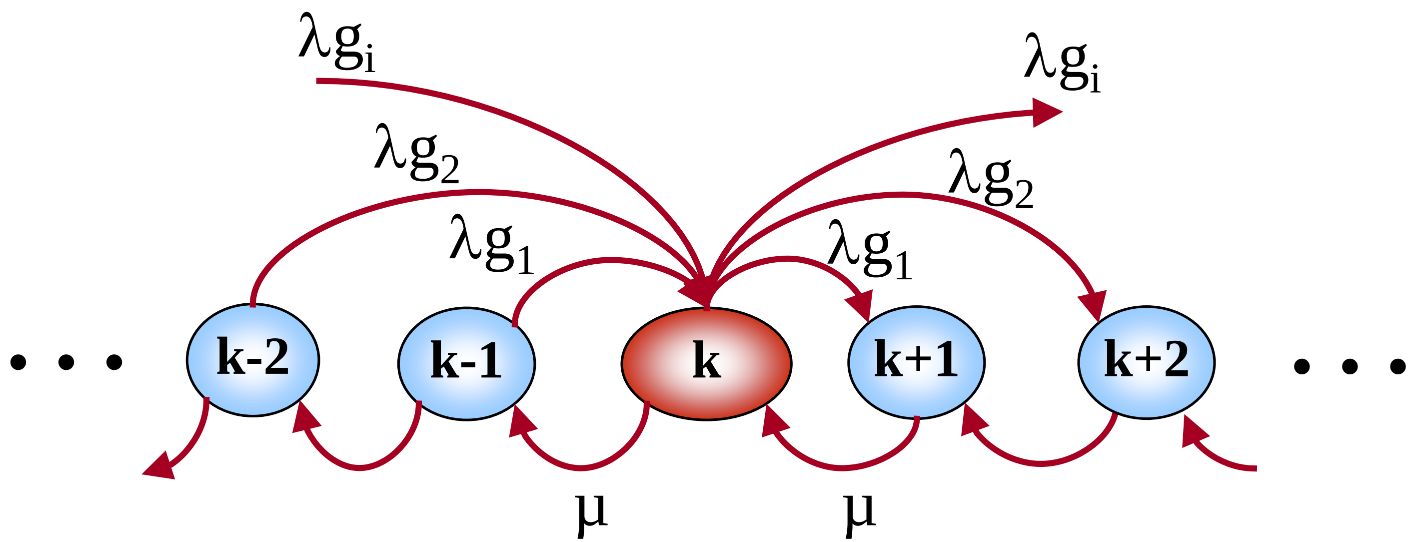
- Previous is fixed size bulk arrival systems
- For general random size bulk arrival systems
- Example:
  - Family arrival for a dentist
- Define

$$g_i = P[\text{ bulk size is } i \text{ }]$$

$$\lambda = \text{the arrival rate of bulks}$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Arrival (general) State diagram



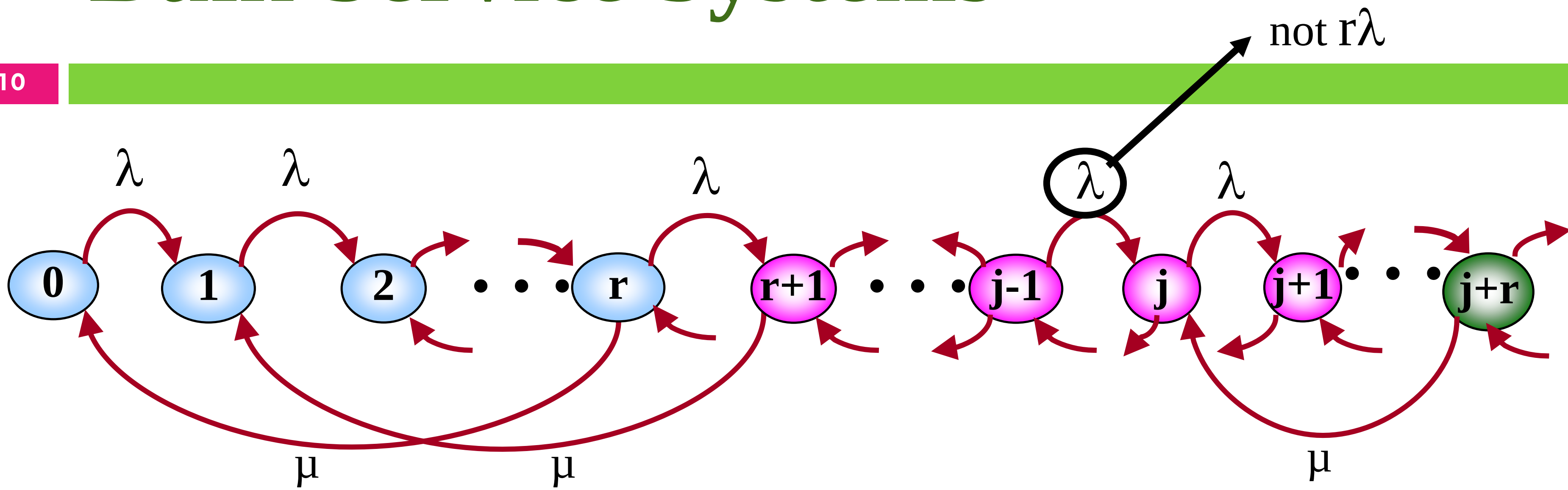
Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Service Systems

- When server is free, it will accept “bulk” of r customers from the queue
- The service time for the group is exponential with μ
- If less than r customers in the queue, the server will wait until r customers
- Example:
  - Shared taxi, Shared Van

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Service Systems



- Same as $E_r/M/1$
- Difference:
  - <u>For $E_r/M/1$</u>
    - State variable = total # of service stage yet to be completed
  - <u>For Bulk Service</u>
    - State variable = the number of customers in the system

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Service Systems

- The result from E_r/M/1

$$P_j = \begin{cases} \frac{1}{r}(1 - z_0^{-j-1}) & 0 \le j < r \\ \rho(z_0 - 1)z_0^{r-j-1} & j \ge r \end{cases}$$

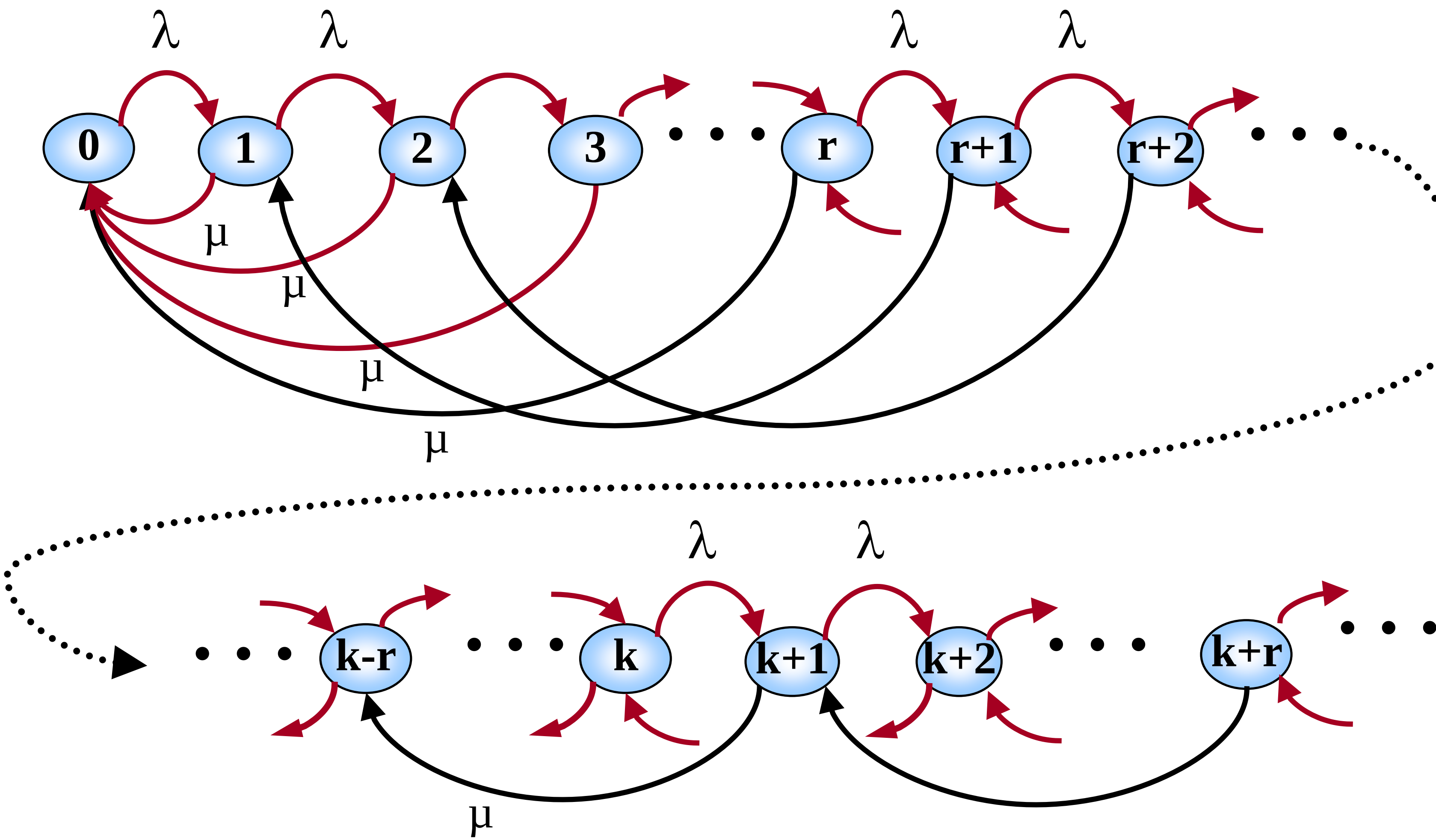
Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Service Systems (no idle)

- Previously, the server is idle if less than r customers
- For system that, if available, accept r customers
- Otherwise, it will accept less than **r** customers

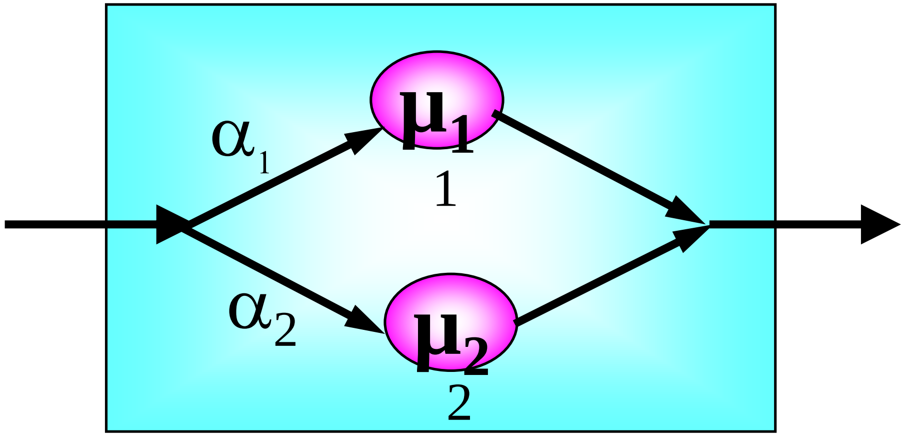
Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Bulk Service (no idle) State diagram



Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

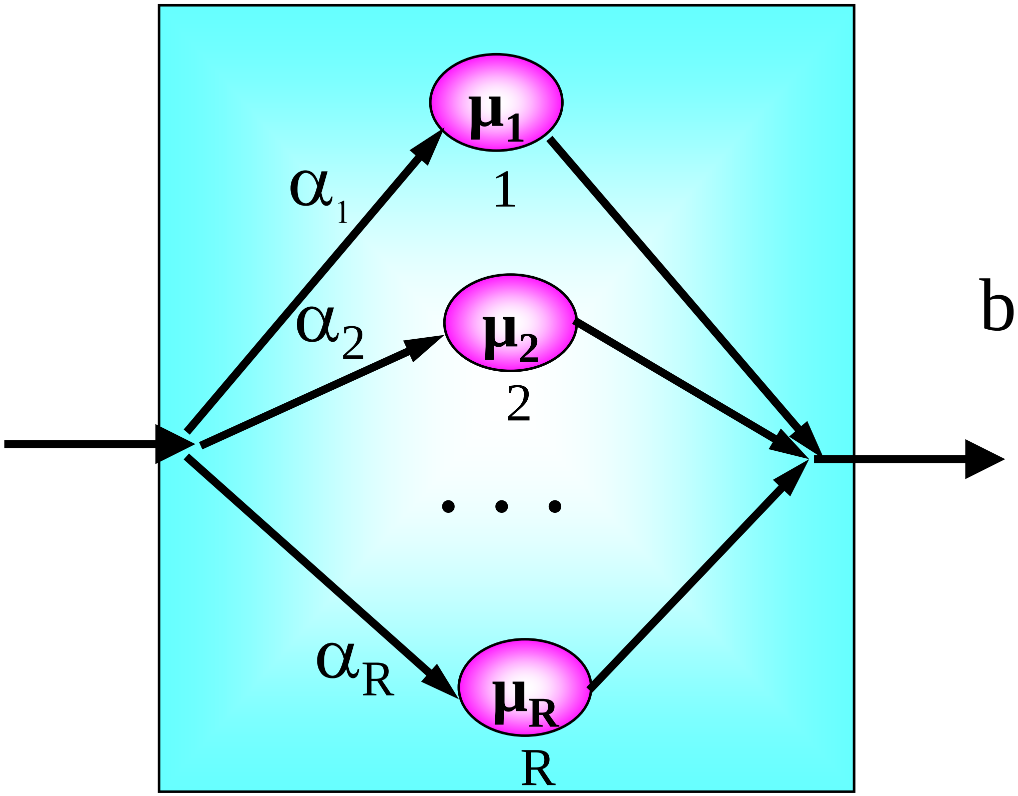
Parallel Servers



Service Facility

- α_i = Probability to select path i
- $\alpha_1 + \alpha_2 = 1$
- Only one customer in the service facility
- $b(x) = \alpha_1 \mu_1 e^{-\mu_1 x} + \alpha_2 \mu_2 e^{-\mu_2 x} \qquad x \geq 0$

R-Stage Parallel Servers (H_R)



Service Facility

$$\sum_{i=1}^{R} \alpha_i = 1$$

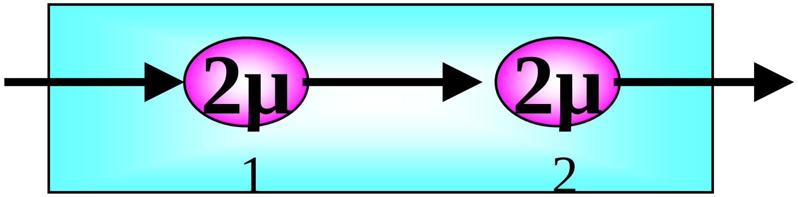
$$b(x) = \sum_{i=1}^{R} \alpha_i \mu_i e^{-\mu_i x} \qquad x \geq 0$$

The pdf is called "***Hyperexponential***" distribution (H_R)

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Serial Servers

- The system with faster exponential stages in series ➔ decreases the variability of service time



Service Facility

- For Mean and Variance of the service time

$$E[x] = 2E[y] = \frac{1}{\mu}$$

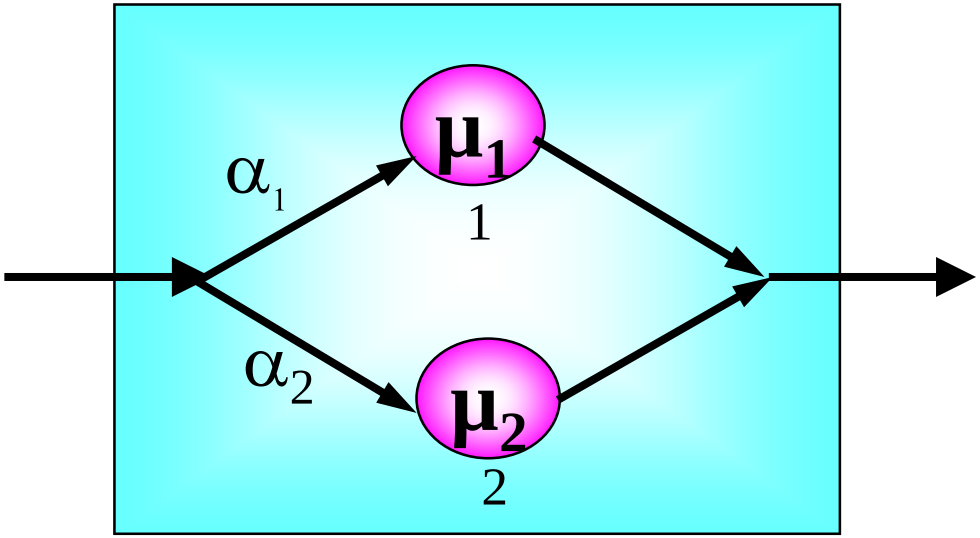
$$\sigma_b^2 = \sigma_h^2 + \sigma_h^2 = \frac{1}{2\mu^2}$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Parallel Servers

- For the parallel system

  ➔ increases the variability of the service time



Service Facility

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Networks of Queues

Outline

- Network of Queues
- Burke’s Theorem
- Application Example 1
- Jackson’s Theorem
- Cyclic Network
- Multidimensional Markov Chains

Networks of Queues

- Multiple-node systems
- Network of nodes =

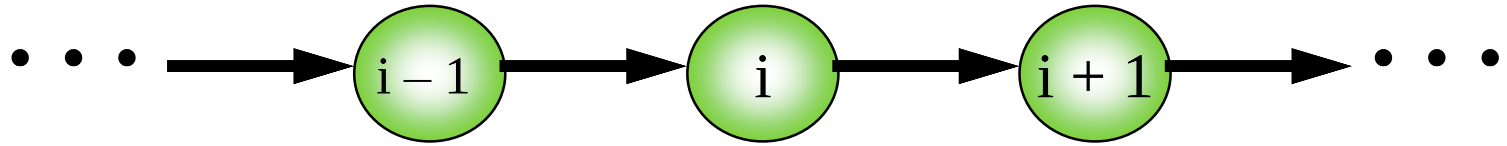
  "**Customers enter system at various points, queue for service, then proceed to other nodes**"
- Topology structure is important

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Tandem network

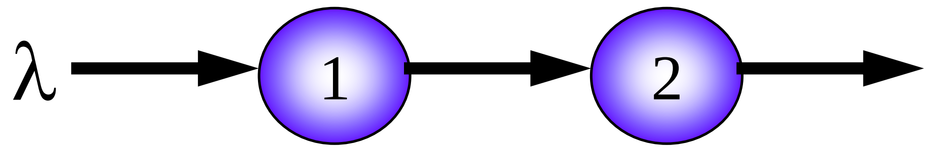
- For a network of k nodes
- Customers depart from node i will immediately enter node i + 1
- Interdeparture times from node i

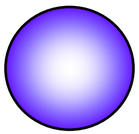
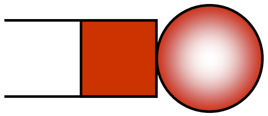
  = Interarrival times for node i + 1



Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Two-Tandem Network



 = queueing system = 

- Assumption:
  - The arrival only enters node 1
  - The arrival is Poisson process with rate λ
  - Each node is exponential server rate μ

Two-Tandem Network

- Therefore,
  - M/M/1 system
- Find
  - Interarrival time of Node 2 (Interdeparture time of node 1)
- Let
  - d(t) = the pdf of the interdeparture process of node 1
  - D*(s) = Laplace transform of d(t)

Two-Tandem Network

- Two cases of Interdeparture Time:
  - A customer enters node 1

    ➔ the next customer departs as exactly service time
  - No customer enters node 1

    ➔ the next customer departs has to wait for the arrival and service time of the new customer
  - (Both independent → Convolution)
- $D^*(s)|\text{non-empty@node } 1 = B^*(s)$
- $D^*(s)|\text{empty@node } 1 = \dfrac{\lambda}{s+\lambda} B^*(s)$

Two-Tandem Network

- For exponential server

$$B^*(s) = \frac{\mu}{s + \mu}$$

- For Interdeparture Time

$$D^*(s) = \rho D^*(s)|\text{non-empty@node1} + (1 - \rho) D^*(s)|\text{empty@node1}$$

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

$$= \left(\frac{\lambda}{s+\lambda}\right)$$

Two-Tandem Network

- For Interdeparture Time

$$D^*(s) = \left(\frac{\lambda}{s + \lambda}\right)$$

$$D(t) = 1 - e^{-\lambda t} \qquad t \geq 0$$

- **Note**: Interdeparture time = exponential distribution with the same parameter as Interarrival time

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Burke’s Theorem

- “A **Poisson** process driving an **exponential** server generates a **Poisson** process for **departures**”
- “The steady-state **output** of M/M/m with λ and service-time with μ for each m channels is a **Poisson process at rate λ**”

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Application Example 1

(from *Gross*: Fundamentals of Queueing Theory 3rd Edition)

- A department store
- Instead of check-out counter, the new design is to provide a check-out lounge
- After complete the shopping, enter the lounge
- If all checkers are busy, receive the number and take a seat
- The next one will be called in sequence
- There is No limit space (even during peak period) on both Shopping area and Waiting area (lounge)

Application Example 1

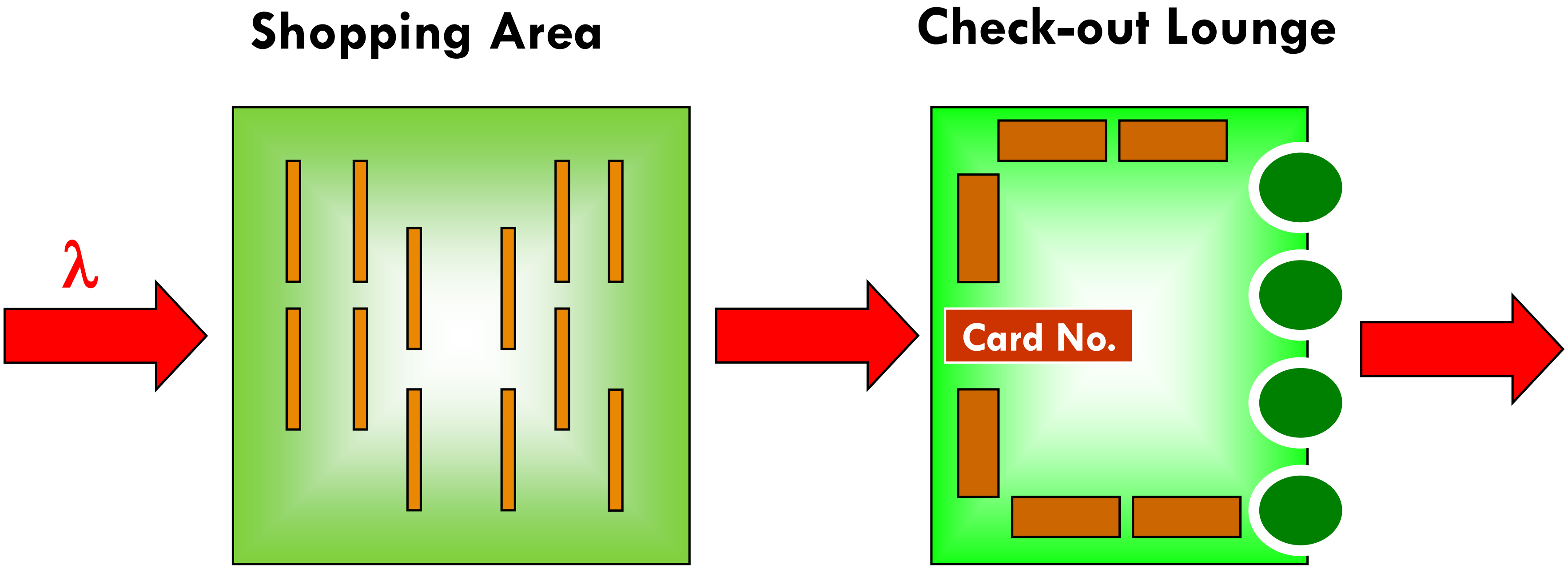
- During peak hour
  - Customers arrive as a Poisson process rate 40/Hr
  - Each customer take ¾ Hr for shopping
- The checkout time at the checker is approximately exponential distributed with a mean 4 min. (there are a bagger for each counter checkout)

Application Example 1

- The manager wants to know :
- **Q1**: What is the minimum number of checkout counters during the peak period?
- **Q2**: If it is decided to add one more than the minimum number of counters,
  - What is the average waiting time in the lounge?
  - On average, how many people will be in the lounge?
  - On average, how many people will be in the department store?

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

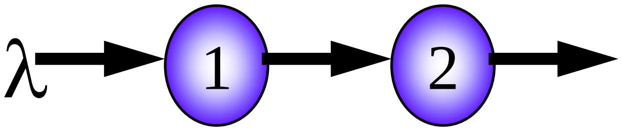
Drawing a Scenario



Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Application Example 1

- Modeling
  - Two-Tandem Network



- Define the first node
  - Shopping area
  - It is self-service
  - Arrival is Poisson

➔ M/M/∞ with $\lambda = 40$ and $\mu = 4/3$

Application Example 1

- Define the second node
  - At Lounge
  - Arrival is the interdeparture of the node 1 = Poisson process
  - Each counter mean service time = 4 min.

  ➔ M/M/c

  with $\lambda = 40$ and service rate = $c\mu$

Application Example 1

- Q1: What is the minimum number of checkout counters during the peak period?
- For steady-state
  - $c\mu > \lambda$
  - So, $c > \lambda/\mu$ ➔ $c > 40/15 = 2.67$
  - c ➔ 3 =The number of check-out counters
  - The system becomes **M/M/3**

Application Example 1

- Q2: If it is decided to add one more than the minimum number of counters
- The system becomes
  - M/M/4/∞
  - With $\lambda = 40$ and $\mu = 15$
  - $\rho = 40/(4*15) = 2/3$

Application Example 1

- M/M/4/∞

$$p_0 = \left(\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \left(\frac{(m\rho)^m}{m!} \right) \left(\frac{1}{1-\rho} \right) \right)^{-1}$$

$$p_0 = \left(\sum_{k=0}^{3} \frac{(4*2/3)^k}{k!} + \left(\frac{(4*2/3)^4}{4!} \right) \left(\frac{1}{1-2/3} \right) \right)^{-1}$$

$$= 0.06$$

Application Example 1

- The Expected queue size, N_q (for $k \geq m$)

$$N_q = \sum_{k=m+1}^{\infty} (k-m)\, p_k = \sum_{k=m+1}^{\infty} (k-m) \frac{m^m \rho^k}{m!} p_0$$

$$= p_0 \frac{m^m \rho^{m+1}}{m!\,(1-\rho)^2}$$

$$= 0.06 \frac{4^4 (2/3)^{4+1}}{4!\,(1-2/3)^2} = 0.759$$

Application Example 1

- The average waiting time in the lounge, W_q
- $N_q = \lambda W_q$

  ➔ $W_q = N_q / \lambda = 0.759/40$

  $= 0.019$ Hr $= 1.14$ min.
- The average number of customer in node 2, N_2
- $N_2 = \lambda W_2 = \lambda (W_q + 1/\mu) = 40 (0.019 + 1/15)$

  $= 3.43$

Application Example 1

- The total number of customer in the system,

  $N = N_1 + N_2$
- $N_1 = \lambda\, W_1 = 40/(4/3) = 30$ (M/M/∞)
- $N = 30 + 3.43 = 33.43$

Jackson's Theorem

- For N nodes
- i^{th} node consists of m_i exponential servers
- With parameter = μ_i
- The i^{th} node receives arrivals from outside with Poisson rate = γ_i
  - If N =1 ➔ M/M/m

Jackson's Theorem

- After leave the i^{th} node, a customer proceeds to the j^{th} node with prob. = r_{ij}, $r_{ij} > 0$
- The total average arrival rate of customers at a given node = λ_i

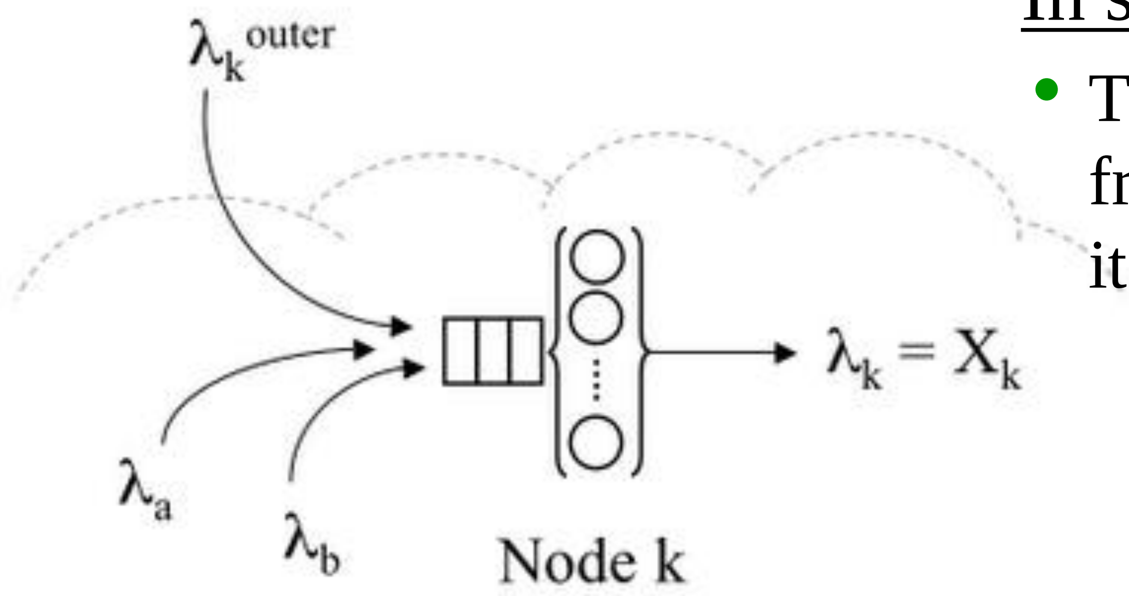
$$\lambda_i = \text{arrivals from outside + inside}$$

$$\lambda_i = \gamma_i + \sum_{j=1}^{N} \lambda_j r_{ji} \qquad i = 1, 2, \ldots, N$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Jackson's Theorem

http://perfdynamics.blogspot.com/2010/05/jacksons-theorem-for-cloud.html



- <u>In steady state</u>
  - The total departure rate from node k is the same as its local throughput X_k

$$\lambda_k = \lambda_k^{outer} + \sum_k P_{jk} \lambda_j \equiv X_k$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Jackson's Theorem

- To be ergodic Markov chains ➔ $\lambda_i < m_i\mu_I$
- Jackson shown that
  - Each node behaves as independent M/M/m with a Poisson rate λ_i
- The Joint distribution for all nodes becomes

$$p(k_1,k_2,\ldots,k_N) = p_1(k_1)p_2(k_2)\ldots p_N(k_N)$$

- $p_i(k_i)$ = solution of M/M/m

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Jackson's Theorem

- From the λ_i

$$\lambda_i = \gamma_i + \sum_{j=1}^{N} \lambda_j r_{ji} \qquad i = 1, 2, \ldots, N$$

- Can rewrite in the vector-matrix form

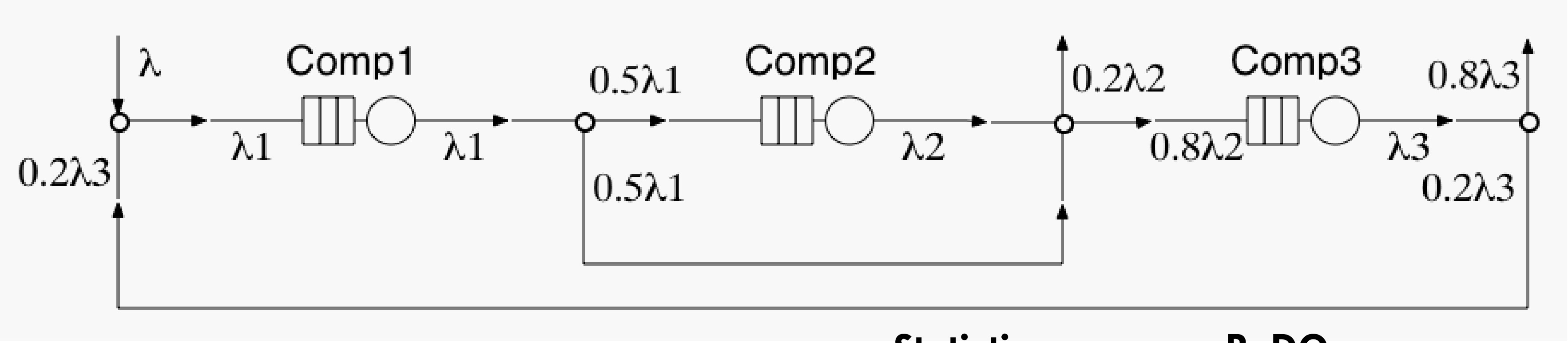
$$\lambda = \gamma + \lambda R$$

- R = Routing matrix

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Communications Network

http://www.perfdynamics.com/Tools/PDQpython.html#tth_sEc4.2



global call rate is $\lambda = 0.50$

Traffic equations:

$\lambda_1 = \lambda + 0.2\,\lambda_3$ (1)

$\lambda_2 = 0.5\,\lambda_1$ (2)

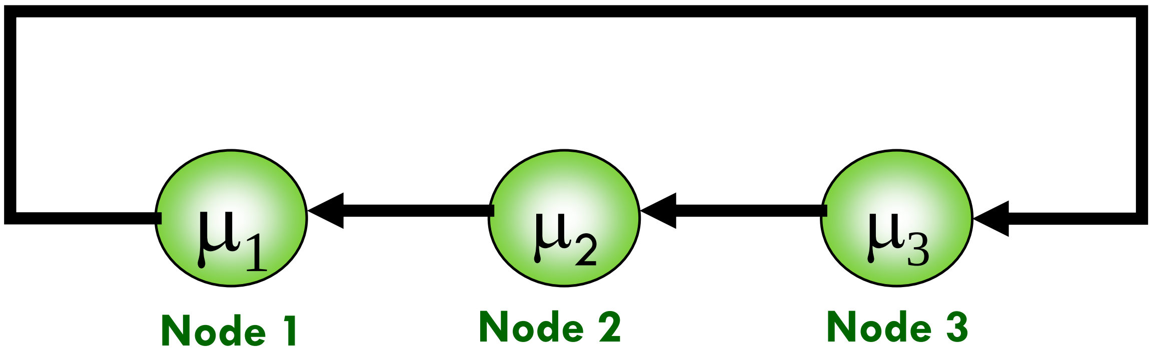
$\lambda_3 = 0.5\,\lambda_1 + 0.8\,\lambda_2$ (3)

Statistic	PyDQ
Avg. queue length at Comp1	1.5625
Avg. queue length at Comp2	1.5625
Avg. queue length at Comp3	1.2162
Avg. number in the system	4.3412
Average response time	8.6824
Average service time	1.1994
Average throughput	0.5000

PDQ (*Pretty Damn Quick*) is open source software associated with the books *Analyzing Computer System Performance with Perl::PDQ* (Springer 2005) http://www.perfdynamics.com/Tools/PDQ.html

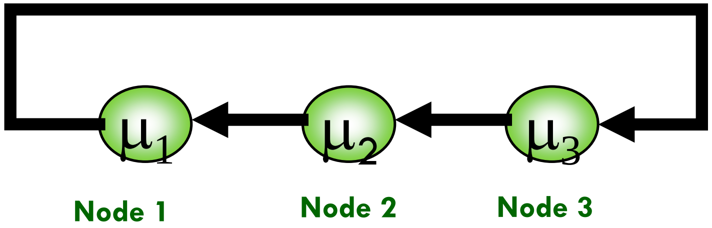
Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Closed Cyclic Network



- N = The number of node = 3
- K = Customers in the system = 2
- Service rate = μ_i
- $r_{13} = r_{32} = r_{21} = 1$; otherwise 0

State Transition



- Let k = # customers in node i
- State description (k_1, k_2, k_3)
- $k_1 + k_2 + k_3 = K = 2$
- How many states can it be?

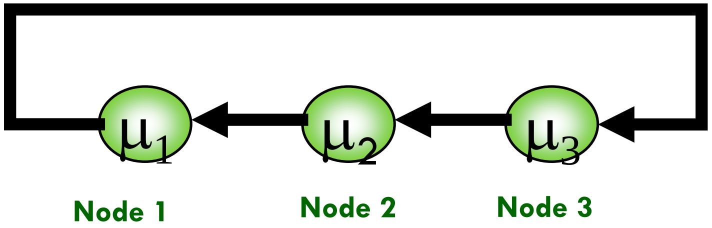
$$\binom{N+K-1}{N-1} = \binom{3+2-1=4}{3-1=2} = 6$$

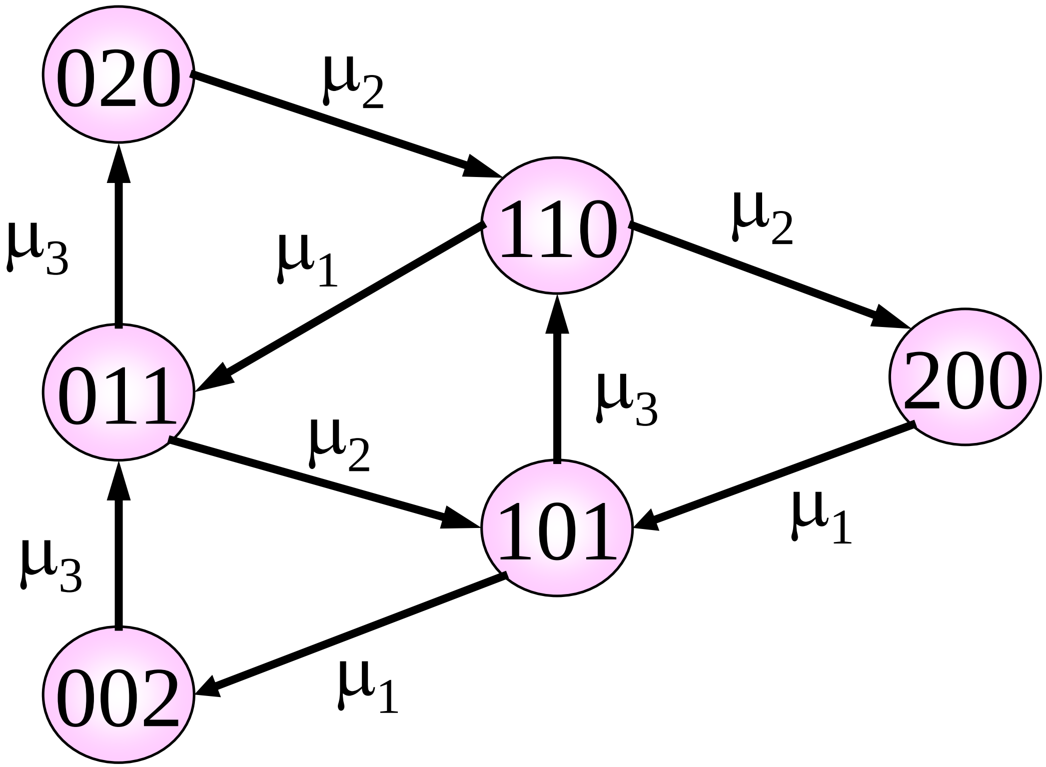
- Can you list them ?

(002) (020) (200) (011) (101) (110)

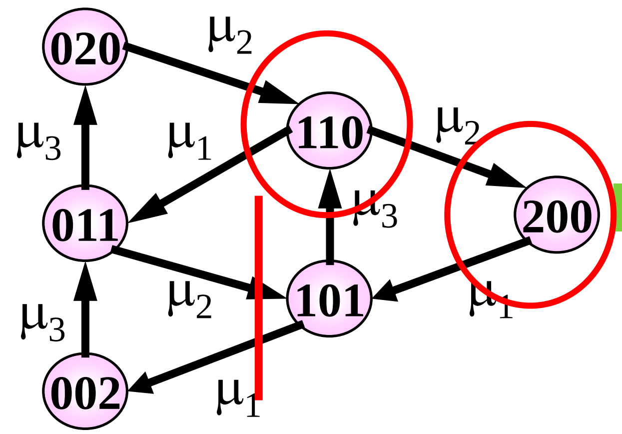
Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

State transition





Global Balance Eq.



$$\mu_1\, p(2,0,0) = \mu_2\, p(1,1,0)$$

$$\mu_2\, p(0,2,0) = \mu_3\, p(0,1,1)$$

$$\mu_3\, p(0,0,2) = \mu_1\, p(1,0,1)$$

$$(\mu_1 + \mu_2)\, p(1,1,0) = \mu_2\, p(0,2,0) + \mu_3\, p(1,0,1)$$

$$(\mu_2 + \mu_3)\, p(0,1,1) = \mu_3\, p(0,0,2) + \mu_1\, p(1,1,0)$$

$$(\mu_1 + \mu_3)\, p(1,0,1) = \mu_1\, p(2,0,0) + \mu_2\, p(0,1,1)$$

$$p(0,0,2) + p(0,2,0) + p(2,0,0) + p(0,1,1) + p(1,0,1) + p(1,1,0) = 1$$

$$p(2,0,0) = \left(1 + \frac{\mu_1}{\mu_3} + \frac{\mu_1}{\mu_2} + \frac{(\mu_1)^2}{\mu_2\,\mu_3} + \left(\frac{\mu_1}{\mu_3}\right)^2 \left(\frac{\mu_1}{\mu_2}\right)^2\right)^{-1}$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Multidimensional Markov Chains

(from “Data Network”, Bertsekas & Gallager)

- Normally, only a single type of customer
- How about many classes of customers?
  - Each with different statistical characteristics
  - Cannot lump together
- Multidimensional Markov Chains

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Example :

Two session classes in a circuit switching system

- A transmission media consists of **m** independent circuits with equal capacity
- 2 sessions arrive with Poisson rate λ_1 and λ_2
- A session will be blocked if all circuits busy
- The sessions holding times are exponential with means $\mathbf{1/\mu_1}$ and $\mathbf{1/\mu_2}$
- **Find the steady-state blocking probability**

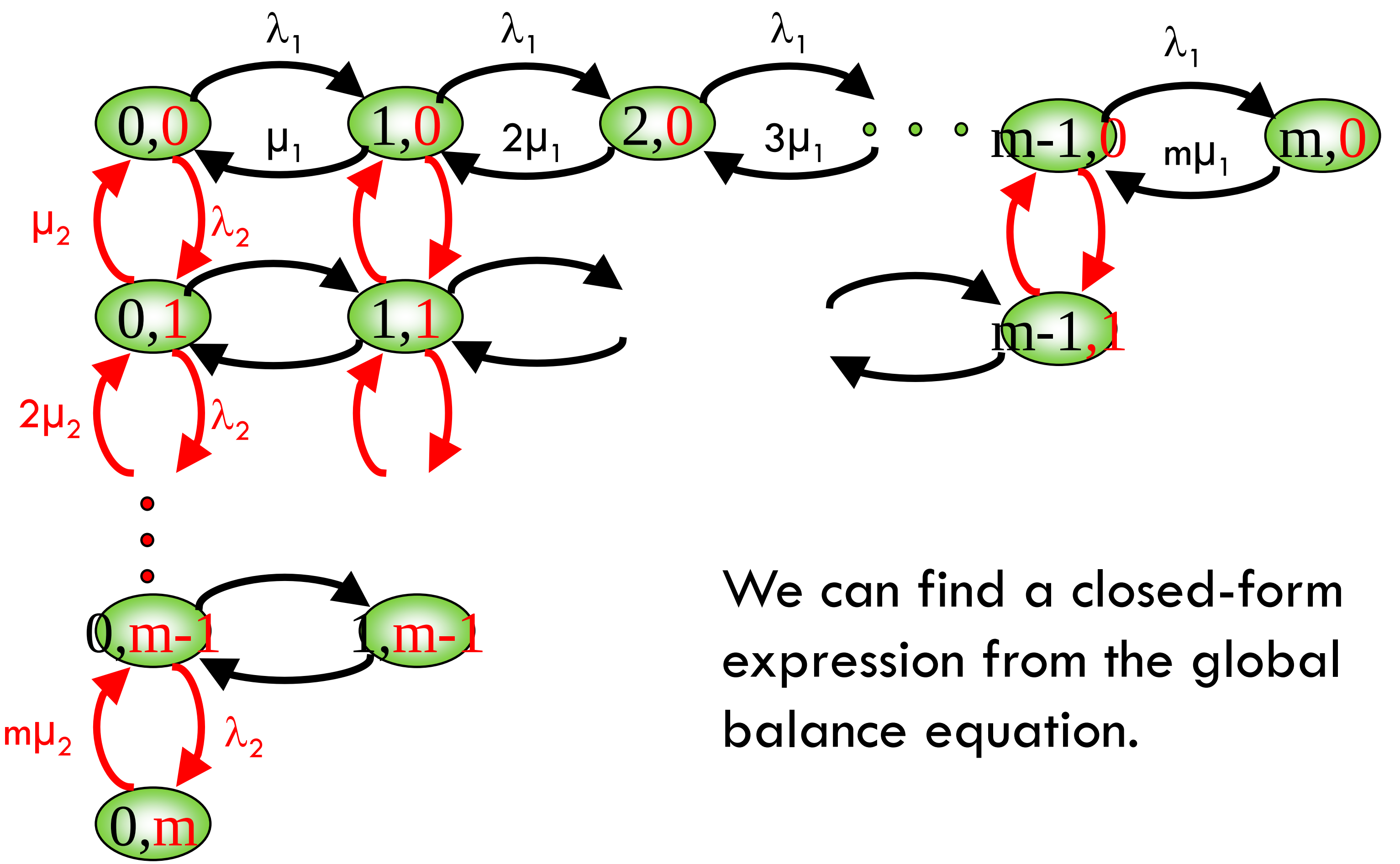
Example :

Two session classes in a circuit switching system

- If $\mu_1 = \mu_2$
  - Two sessions are the same
  - Can be lumped and modeled as M/M/m/m
  - With arrival rate $(\lambda_1 + \lambda_2)$
  - # of states = total number of busy circuits
- If $\mu_1 \neq \mu_2$
  - Total number of busy circuits cannot specify each behavior
  - Needs **Two-dimensional state (n_1, n_2)**
  - n_i = # of circuits used by each type **i**

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Transition Probability diagram



We can find a closed-form expression from the global balance equation.

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Homework

- The insurance company
- Three-node telephone system
- Calls coming into the toll-free number = Poisson with rate 35/Hr
- The caller has 2 options
  - Press 0 for claims service
  - Press 1 for Policy service

Homework

- Each caller will listening and making decision with exponential mean 25 sec.
- Only one call at a time will be proceeded
- Others will hear the nice background and how important the call is
- 60% of calls go to claims, the reminder goes to the policy

Homework

- The claims processing node has 3 parallel servers, each service is exponential with mean 6 min.
- The policy service node has 5 parallel servers, each service is exponential with mean 20 min.
- All buffers in front of the nodes can hold unlimited
- About 1% of customers finishing the claims then go on to the policy service
- About 2% of customers finishing the policy services go on to the claims service

Homework

- What is the average queue sizes in front of each node ?
- What is the total average time a customer spends in the system?

Assignment

- 10 % of grade
- Find a paper that apply/implement/analysis using the Queueing System
- Prepare a summarized report (3 pages of A4-12pt)
  - The goal of the paper
  - The queueing techniques used in the paper
  - The explanation of the queueing system
  - Etc.
- Prepare a presentation for 15 min. / Q&A 10 min.