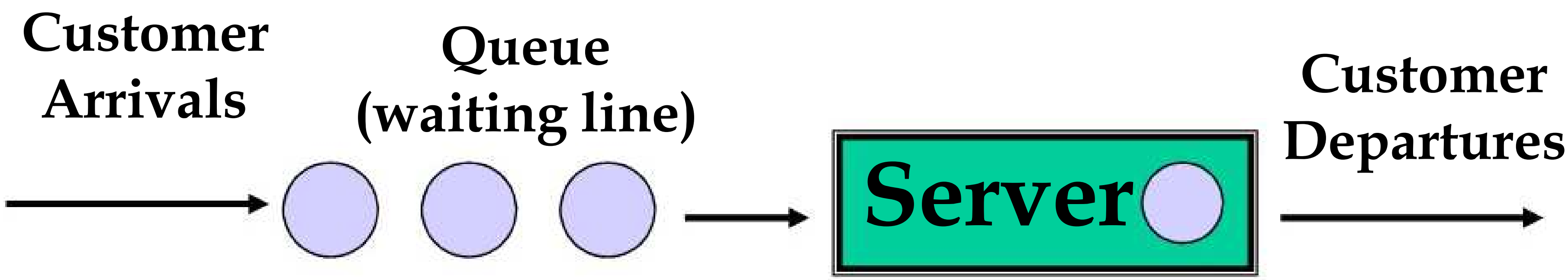
Queuing Networks Modeling Virtual Laboratory

Dr. S. Dharmaraja
Department of Mathematics
IIT Delhi
http://web.iitd.ac.in/~dharmar

Outline

- Introduction
- Simple Queues
- Performance Measures
- Case Study

Queuing System



Kendall Notation: A/B/C/D/X/Y

A: Distribution of inter arrival times

B: Distribution of service times

C: Number of servers

D: Maximum number of customers in system

X: Population

Y: Service policy

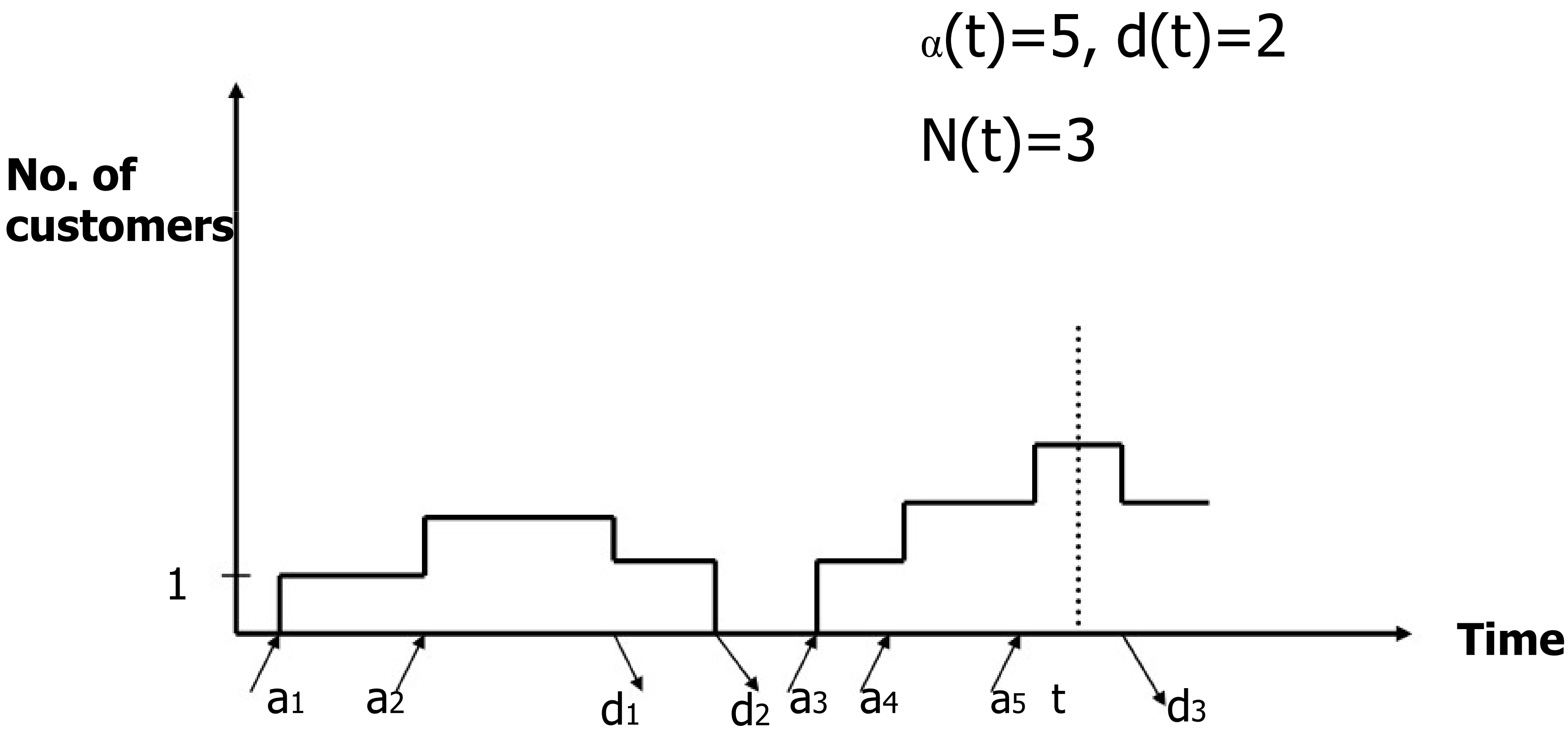
Queuing Models

- M/M/1/∞/FCFS, M/E_k/c/∞/LCFS, M/D/c/k/FCFS, GI/M/c/c/FCFS, MMPP/PH/1/FCFS are some of the examples of the queuing models.
- The notation M/M/1/∞/FCFS indicates a queuing process with the following characteristics:
  - Exponential Inter Arrival Times,
  - Exponential Service Times
  - One Server
  - Infinite System Capacity,
  - First Come First Served Scheduling Discipline.

Some Basic Relations

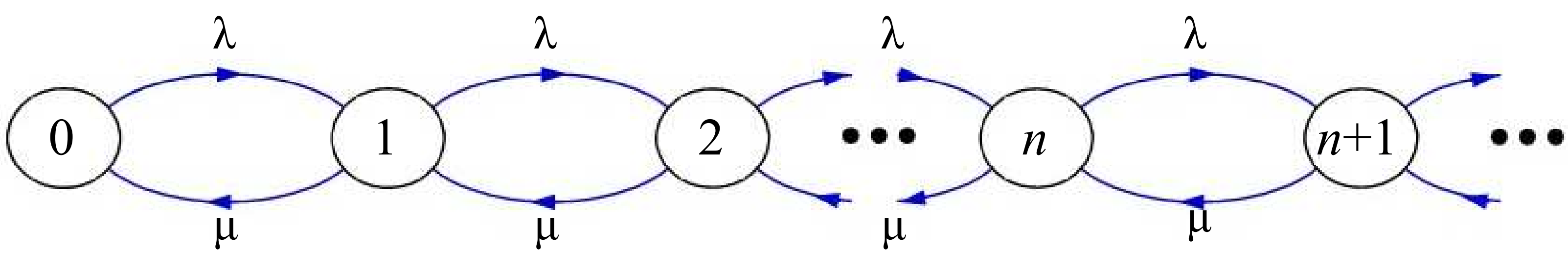
- C_n: n_{th} customer, n=1,2,…
- A_n: arrival time of C_n
- D_n: departure time of C_n
- $\alpha(t)$: no. of arrivals by time t
- $\delta(t)$: no. of departures by time t
- N(t): no. of customers in system at time t,
  $N(t) = \alpha(t) - \delta(t)$.

Diagrammatic Representation



The M/M/1 Queue

- Arrival process: Poisson with rate λ
- Service times: iid, exponential with parameter μ
- Service times and interarrival times: independent
- Single server
- Infinite waiting room
- N(t): Number of customers in system at time *t* (state)



M/M/1 Queue: Markov Chain Formulation

- Transitions due to arrival or departure of customers
- Only nearest neighbors transitions are allowed.
- State of the process at time t: $N(t) = i \; (i \geq 0)$.
- $\{N(t): t \geq 0\}$ is a continuous-time Markov chain with

$$\begin{aligned} q_{i,i+1} &= \lambda \\ q_{i,i-1} &= \mu \\ q_i &= -(\lambda + \mu) \\ q_{i,j} &= 0 \quad \text{for} \quad |i - j| > 1 \end{aligned}$$

M/M/1 Queue: Stationary Distribution

- Birth-death process

$$\mu p_n = \lambda p_{n-1} \Rightarrow$$

$$p_n = \frac{\lambda}{\mu} p_{n-1} = \rho p_{n-1} = \quad = \rho^n p_0$$

- Normalization constant

$$\sum_{n=0}^{\infty} p_n = 1 \Leftrightarrow p_0 \left[1 + \sum_{n=1}^{\infty} \rho^n \right] = 1 \Leftrightarrow p_0 = 1 - \rho, \text{ if } \rho < 1$$

- Stationary distribution

$$p_n = \rho^n (1 - \rho), \quad n = 0, 1, \ldots$$

Performance Measures

1. N: Number of customers in the system at steady state.
   L: Average number of customers in the system.

$$L = E[N] = \frac{\lambda}{\mu - \lambda} \tag{1}$$

2. N_q: Number of customers in the queue at steady state.
   L_q: Average number of customers in the queue.

$$L_q = E[N_q] = \frac{\lambda^2}{\mu(\mu - \lambda)} \tag{2}$$

Performance Measures

3. T_q: Waiting time in the queue.
   W_q: Average waiting time in a queue.

$$W_q = E[T_q] = \frac{\lambda}{\mu(\lambda - \mu)}. \tag{3}$$

4. T: Time spent in the system, including service.
   W: Average time spent in the system.

$$W = E[T] = \frac{1}{\mu - \lambda}. \tag{4}$$

Blocking Probability

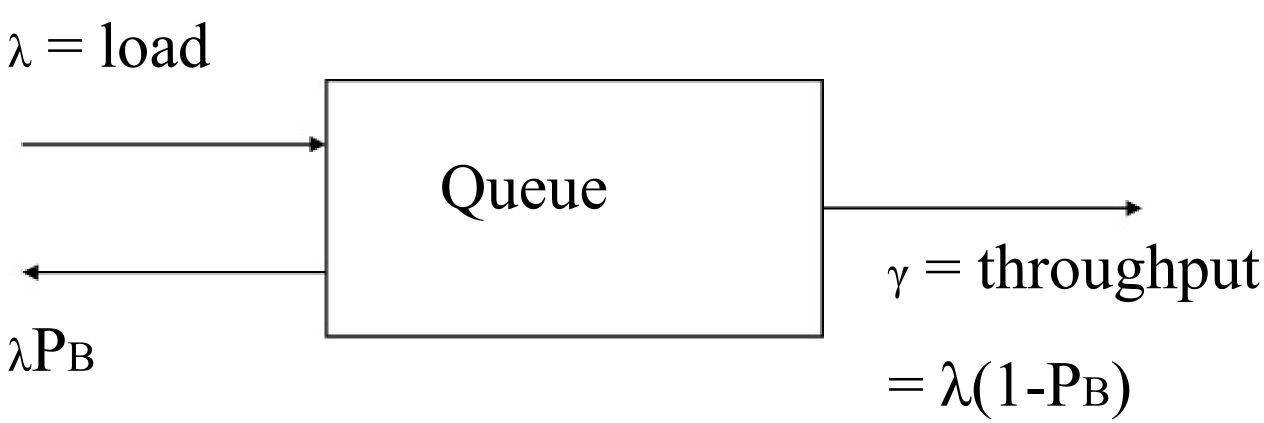
- An important design criterion is the blocking probability of a queuing system
- Would we be happy to lose one packet in 100?, 1000? 1000,000?
- How much extra buffer space must we put in to achieve these figures?

Blocking Probability

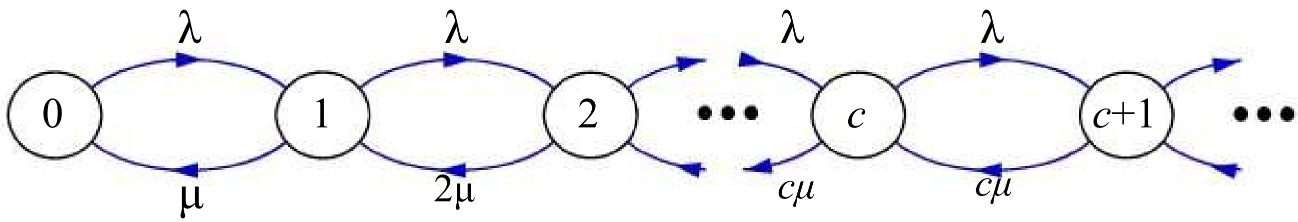
- If a queue is full when a packet arrives, it will be discarded, or "blocked"
- So the probability that a packet is blocked is exactly the same as the probability that the queue is full
- That is, $P_B = p_N$

Blocking Probability

• Schwartz has this useful diagram to describe throughput and blocking



M/M/c Queue



- Poisson arrivals with rate λ
- Exponential service times with parameter μ
- $c$servers
- Arriving customer finds n customers in system
  - $n < c$: it is routed to any idle server
  - $n \geq c$: it joins the waiting queue - all servers are busy
- Birth-death process with state-dependent death rates

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq c \\ c\mu, & \end{cases}$$

M/M/c Queue

- Steady state solutions

$$p_n = \begin{cases} \dfrac{\lambda^n}{n!\mu^n} p_o & 1 \le n \le c \\ \dfrac{\lambda^n}{c^{n-c}c!\mu^n} p_0 & n > c \end{cases}$$

- Normalizing

$$\sum_{n=0}^{\infty} p_n = 1 \Rightarrow p_0 = \left[\sum_{n=0}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n + \frac{1}{c!} \left(\frac{\lambda}{\mu} \right)^c \left(\frac{c\mu}{c\mu - \lambda} \right) \right]^{-1}$$

Performance Measures

1. L_q: Average number of customers in a queue.

$$L_q = \left[\frac{(\lambda/\mu)^c \lambda\mu}{(c-1)!(c\mu-\lambda)^2} p_0 \right] \tag{1}$$

2. L: Average number of customers in the system.

$$L = \sum_{n=1}^{c-1} \frac{n\lambda^n}{n!\mu^n} p_0 + \sum_{n=c+1}^{\infty} \frac{n\lambda^n}{c^{n-c}c!\mu^n} p_0 \tag{2}$$

Performance Measures

3. W_q: Average waiting time in a queue.

$$W_q = \frac{L_q}{\lambda} = \left[\frac{(\lambda/\mu)^c \lambda}{(c-1)!(c\mu - \lambda)^2} p_0 \right] \tag{3}$$

4. W: Average time spent in the system.

$$\begin{aligned} W &= W_q + \frac{1}{\mu} \\ &= \frac{1}{\mu} + \left[\frac{(\lambda/\mu)^c \lambda}{(c-1)!(c\mu - \lambda)^2} p_0 \right] \end{aligned} \tag{4}$$

M/M/c/K Queues

- Poisson arrivals with rate λ
- Exponential service times with parameter μ
- $c$servers with system capacity K
- Arriving customer finds n customers in system
  - $n < c$: it is routed to any idle server
  - $n \geq c$: it joins the waiting queue - all servers are busy
- Customers forced to leave the system if already K present in the system.

M/M/c/K Queues

- Birth death process with state dependent death rates

$$\mu_n = \begin{cases} n\mu_n, & 1 \le n < c \\ c\mu, & c \le n \le K \end{cases}$$

- Stead

$$p_n = \begin{cases} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n p_0 & 0 \le n < c \\ \frac{1}{c^{n-c}c!}\left(\frac{\lambda}{\mu}\right)^n p_0 & c \le n \le K \end{cases}$$

Performance Measures

1. L_q: Average length of the queue.

$$L_q = \frac{p_0 (c\rho)^c \rho}{c!(1-\rho)^2}[1-\rho^{K-c+1}-(1-\rho)(K-c+1)\rho^{K-c}]$$

2. L: Average number of customers in the system.

$$L = L_q + c - p_0 \sum_{n=0}^{c-1} \frac{(c-n)(c\rho)^n}{n!}$$

Performance Measures

3. W_q: Average waiting time in a queue.

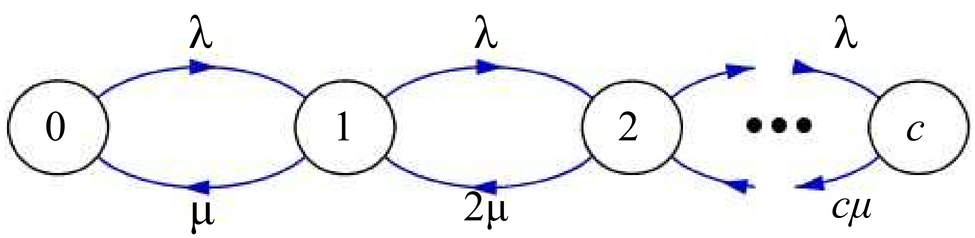
$$W_q = \frac{L_q}{\lambda'}, \qquad \lambda' = \lambda(1 - p_K)$$

where λ' is the effective rate at which the jobs enter the system.

4. W: Average time spent in the system.

$$W = W_q + \frac{1}{\mu}$$

M/M/c/c Queue



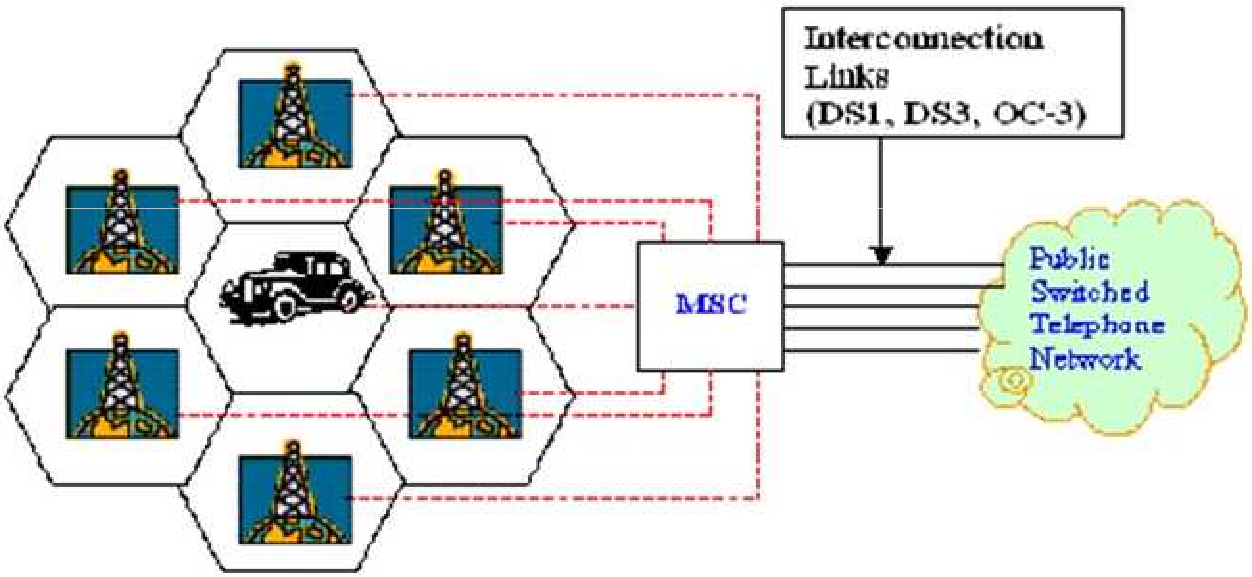
- c servers, no waiting room
- An arriving customer that finds all servers busy is blocked
- Stationary distribution:

$$p_n = \frac{(\lambda/\mu)^n}{n!}\left[\sum_{k=0}^{c}\frac{(\lambda/\mu)^k}{k!}\right]^{-1}, \quad n = 0, 1, \ldots, c$$

Case Study 1

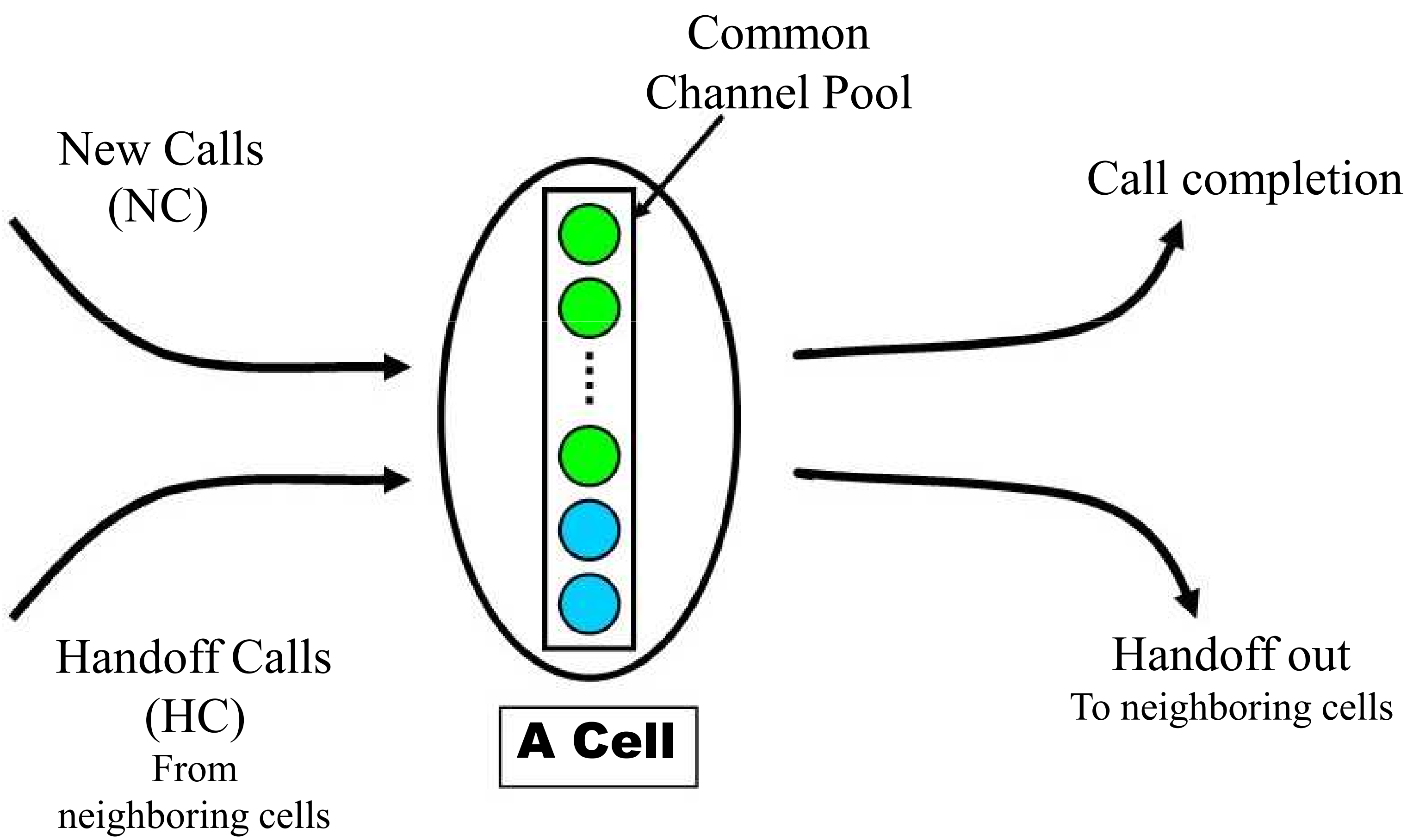
Cellular Networks

Components of Cellular Systems



(1) Base transceiver station (BTS)
(2) Mobile switching center
(3) Mobile unit
(4) Fixed network
(5) Interconnection to the PSTN

Wireless Handoff Performance Model



System Description

- Handoff Phenomenon:
  - A call in progress handed over to another cell due to user mobility
  - Channel in old base station is released and idle channel given in new base station
- Dropping Probability
  - If No idle channel available, the handoff call is dropped.
  - Percentage of calls forcefully terminated while handoff.
- Blocking Probability
  - If number of idle channels less than or equal to `g', new call is dropped
  - Percentage of new calls rejected.

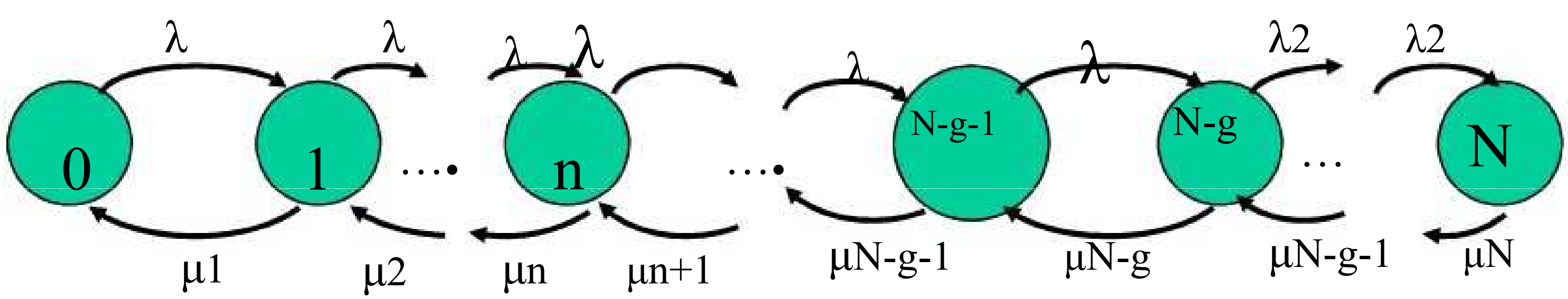
Basic Model

- Calls arrives in Poisson manner.
  - λ_1:Rate of Poisson arrivals for NC
  - λ_2:Rate of Poisson arrivals for HC
- Exponential service times
  - μ_1:Rate of ongoing service
  - μ_2:Rate of handoff of call to neighboring cell.
- N: Total number of channels in pool
- g: No. of guard channels.

Basic Model

- C(t): No. of busy channels,
- $\{C(t), t \geq 0\}$: continuous time, discrete state Markov process,
- Only nearest neighbor transitions allowed
- Variation of M/M/c/c queuing model.

State Transition Diagram



Markov Chain Model of Wireless Handoff

Steady State Equations

$$
\begin{aligned}
0 &= -\lambda p_0 + \mu_1 p_1 \\
0 &= \lambda p_{n-1} - (\lambda + n\mu_n) p_n + \mu_{n+1} p_{n+1} \quad n = 1, 2, \ldots N - g - 1. \\
0 &= \lambda p_{N-g-1} - (\lambda_2 + \mu_{N-g}) p_{N-g} + \mu_{n+1} p_{N-g+1} \quad n = N - g \\
0 &= \lambda p_{n-1} - (\lambda_2 + \mu_n) p_n + \mu_{n+1} p_{n+1}, \quad n = N - g + 1, \ldots, N - 1 \\
0 &= \lambda_2 p_{N-1} - \mu_N p_N.
\end{aligned}
$$

Steady State Solutions

$$p_n = p_0 \begin{cases} \frac{A^n}{n!}, & n \leq N - g \\ \frac{A^{N-g}}{n!} A_1^{n-(N-g)}, & n \geq N - g \end{cases}$$

where

$$p_0 = \frac{1}{\sum_{n=0}^{N-g-1} \frac{A^n}{n!} + \sum_{n=N-g}^{N} \frac{A^{N-g}}{n!} A_1^{n-(N-g)}}$$

Loss Probabilities

- Dropping Probabilities

$$P_d(N,g) = p_N$$
$$= \frac{\frac{A^{N-g}}{N!}A_1^g}{\sum_{n=0}^{N-g-1}\frac{A^n}{n!} + \sum_{n=N-g}^{N}\frac{A^{N-g}}{n!}A_1^{n-(N-g)}}$$

- Blocking Probabilities

$$P_b(N,g) = \sum_{n=N-g}^{N} p_n$$
$$= A^{N-g}\frac{\sum_{k=0}^{g}\frac{A_1^k}{(k+N-g)!}}{\sum_{n=0}^{N-g-1}\frac{A^n}{n!} + \sum_{n=N-g}^{N}\frac{A^{N-g}}{n!}A_1^{n-(N-g)}}$$

Text and Reference Books

- **Main Text Books**
  - Data Networks, Dimitri P. Bersekas and Gallager, Prentice Hall, 2nd edition, 1992 (chapters 3 and 4).
  - Probability and Statistics with Reliability, Queuing and Computer Science Applications, Kishor S. Trivedi, John Wiley, second edition, 2001 (chapters 8 and 9).
- **Reference Books**
  - The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling, Raj K. Jain, Wiley, 1991.
  - Computer Applications, Volume 2, Queuing Systems, Leonard Kleinrock, Wiley-Interscience, 1976.