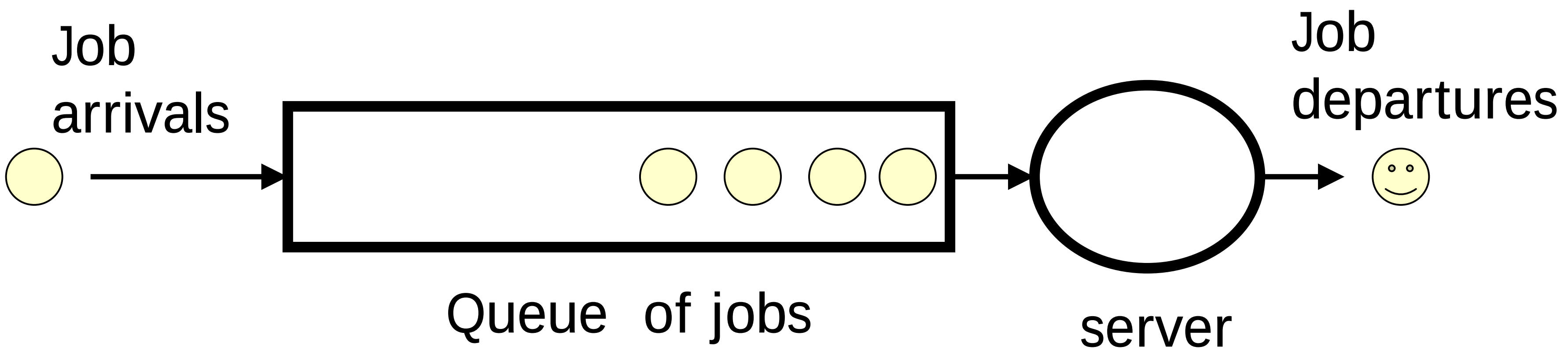
Performance Evaluation Methods and Their Applications in Wireless and Optical Networks

Mostafa Bassiouni
University of Central Florida

August 3, 2005

Queuing Systems



Kendall's Notation: A/B/C/K/m/Z

A= Arrival process

B= Server process

C= Number of servers

K= Queue capacity

m= Job population

Z= Queuing discipline

Kendall's Notation: A/B/C/K/m/Z

- Notation:
  - M= Markovian (exponential)
  - D= Deterministic (constant)
  - G= General
  - E= Erlang
  - H= Hyper Exponential
- Example: M/M/1/∞/ ∞/FIFO
  - This is a single-server FIFO queue with Markovian arrival and service processes
  - The default values of K/m/Z are truncated and the queue is simply denoted by M/M/1

The Poisson Process

A non-negative integer random variable X is said to be Poisson if it has the following distribution:

$$\text{Prob}\ \{X \ \text{is equal to}\ k\} = \frac{\lambda^k e^{-\lambda}}{k!}$$

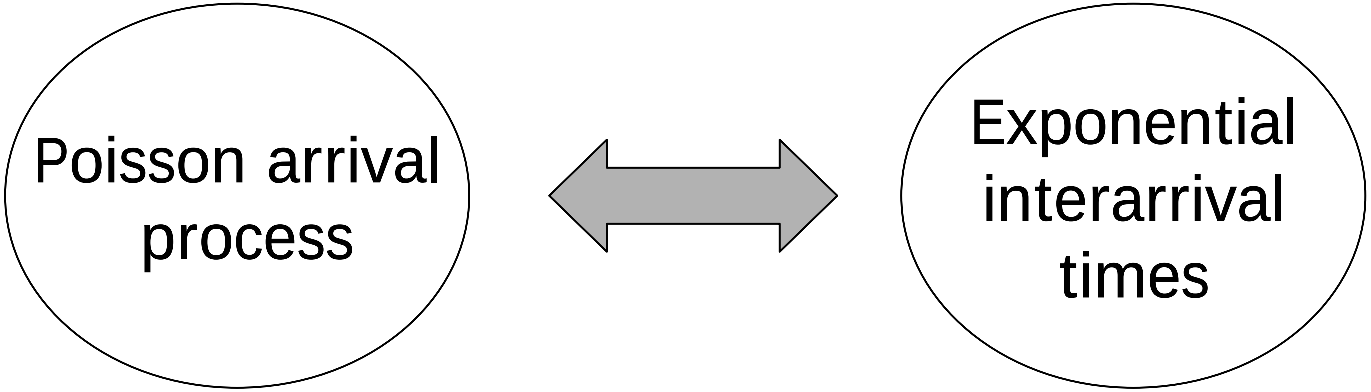
The mean of X indicates the rate of events (number of arrivals per unit of time) and is given by

$$E(X) = \lambda$$

Given a Poisson process of rate λ, the distribution of the number of arrivals in a period of length τ is given by

$$\Pr ob(k\ arrivals) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$$

Interarrival Times



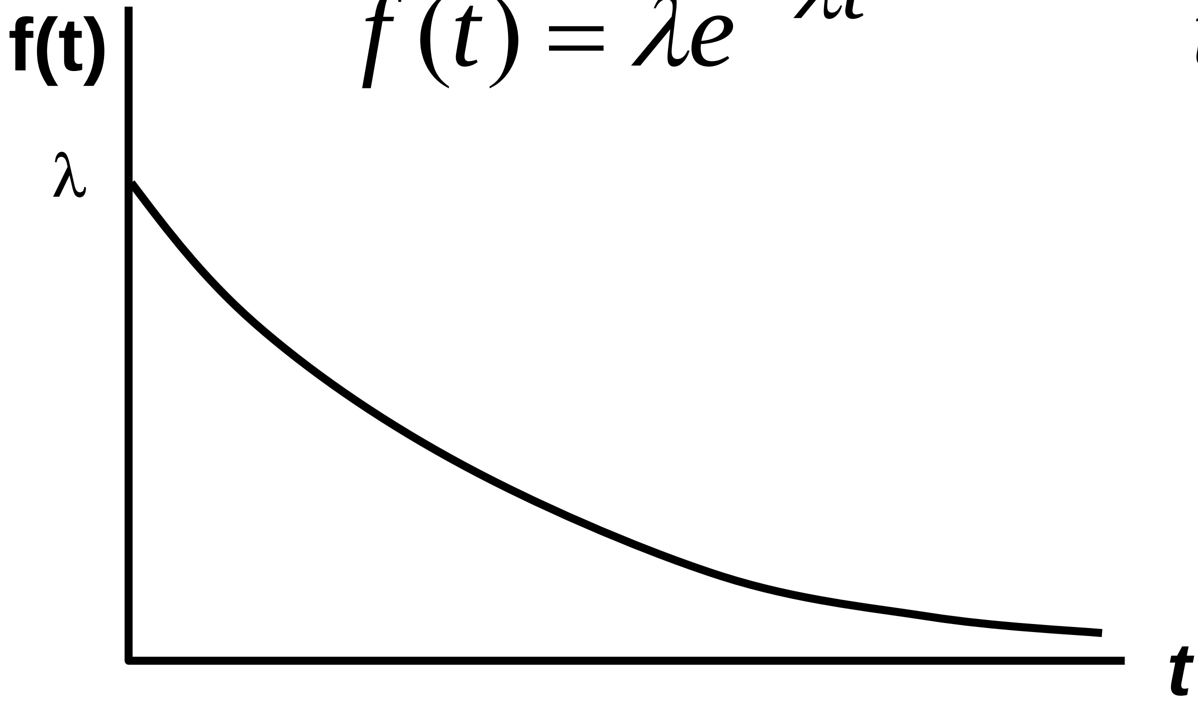
- The interarrival times of a Poisson arrival process are exponentially distributed with the following probability density function

$$f(t) = \lambda e^{-\lambda t}$$

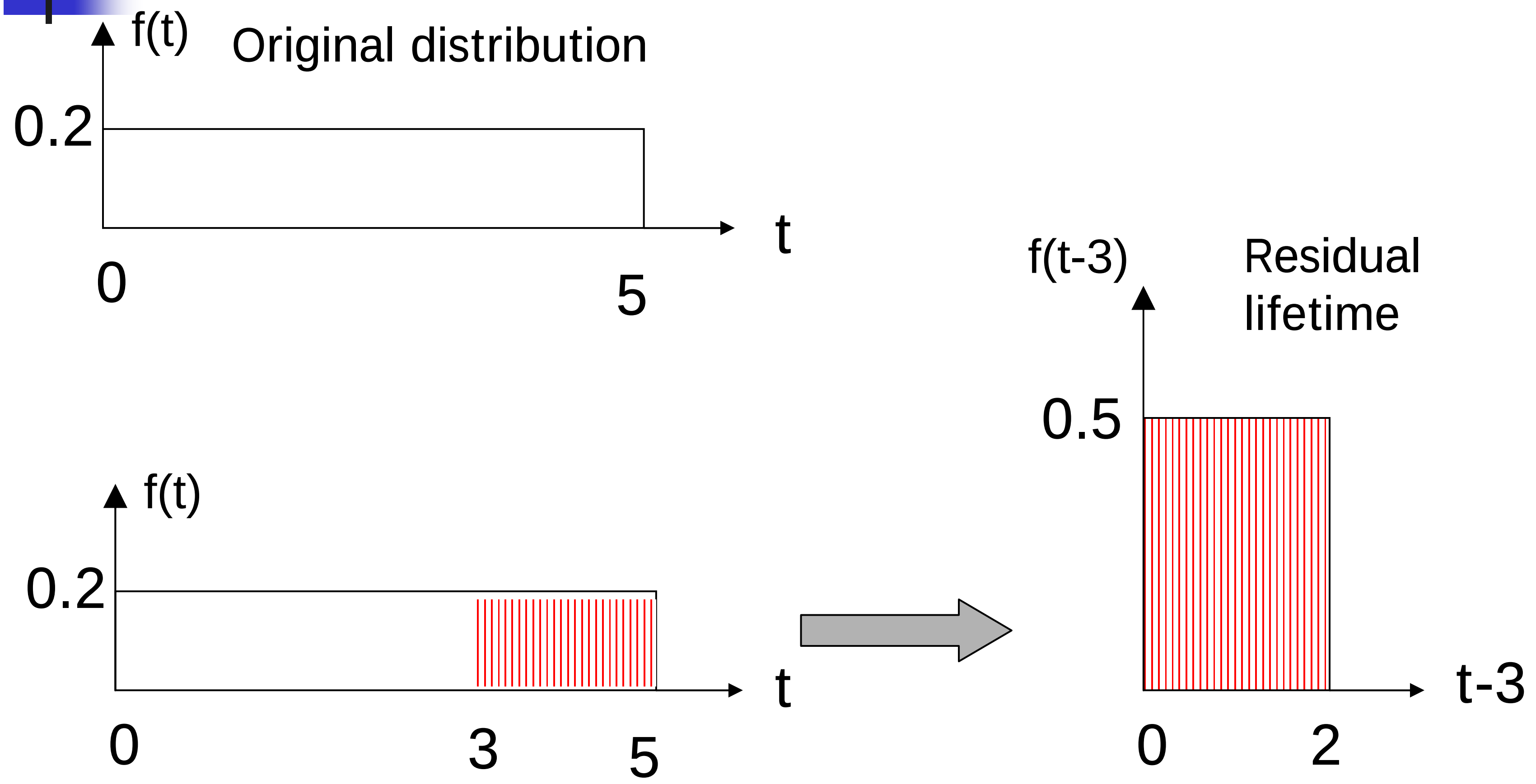
The Exponential Distribution

- Graph of the Exponential pdf

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

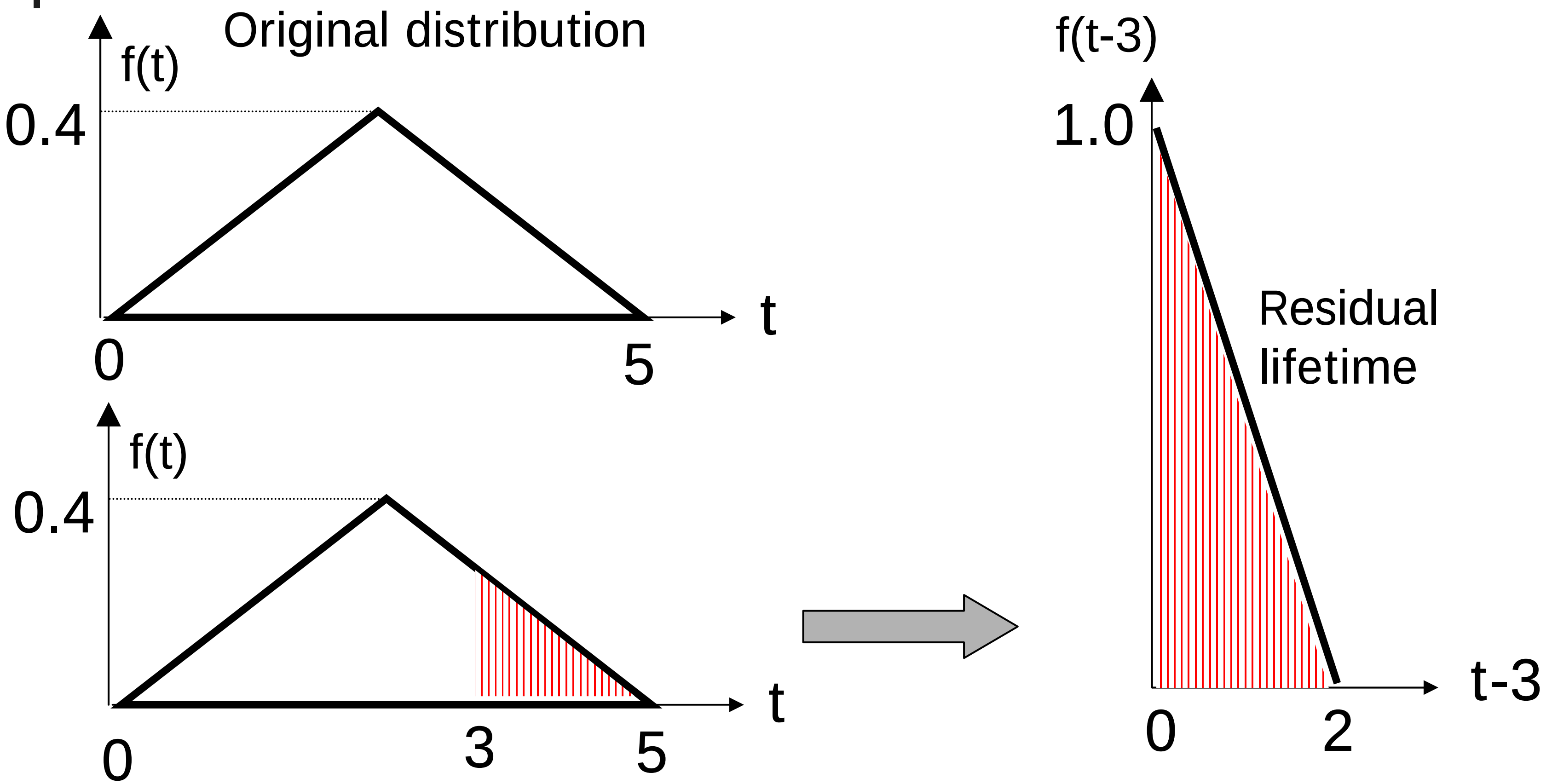


Distribution of Residual Lifetime Example 1



After waiting for 3 units of time, the residual lifetime distribution is not exactly the same as the original distribution

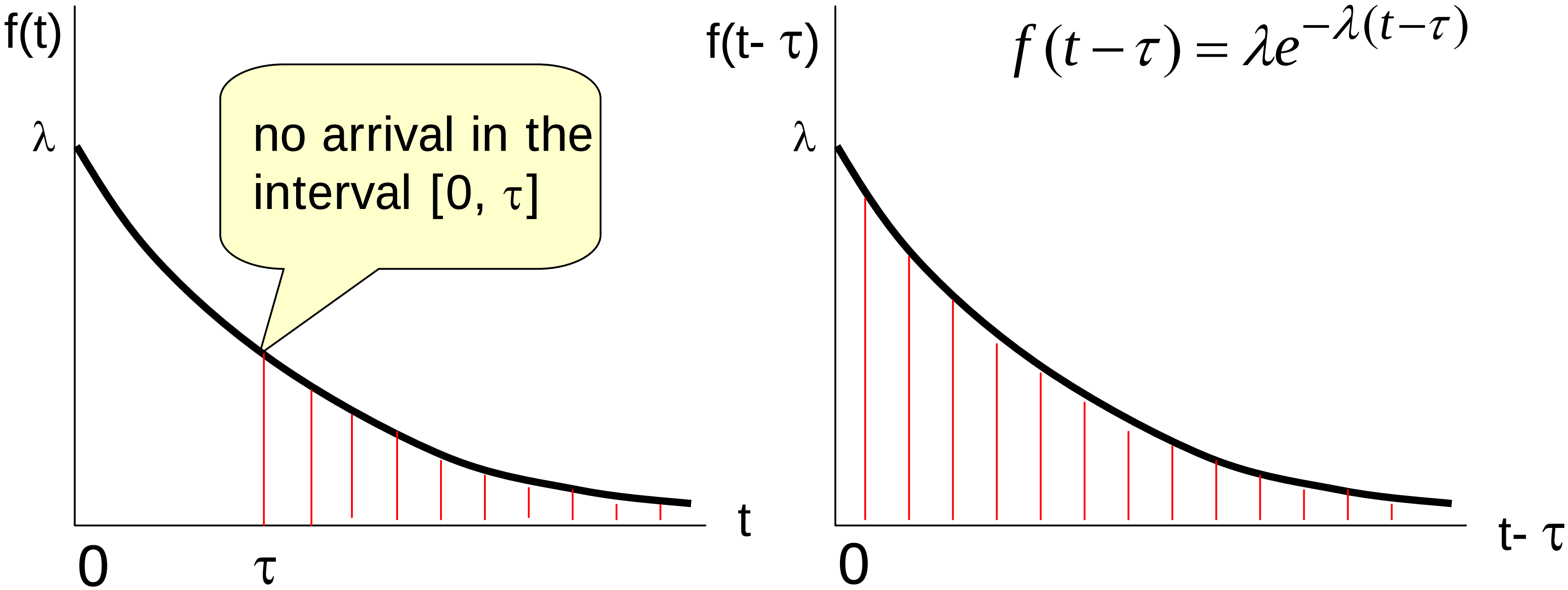
Distribution of Residual Lifetime Example 2



After waiting for 3 units of time, the residual lifetime distribution is not exactly the same as the original distribution

The Memoryless Property

After waiting for a period of length τ, the Poisson process forgot that any time had been expended



$$f(t-\tau) = \lambda e^{-\lambda(t-\tau)}$$

The distribution of residual lifetime is precisely the same as the original exponential distribution

Properties of the Poisson Process

- The Memoryless Property
  - The distribution of the time to next arrival starting from any time would be just the same as if we start from the moment of the previous arrival
- Approximating a Poisson Process
  - During a small interval of time δ,

    Prob{one arrival} $\approx \lambda\delta$

    Prob{no arrival} $\approx 1- \lambda\delta$

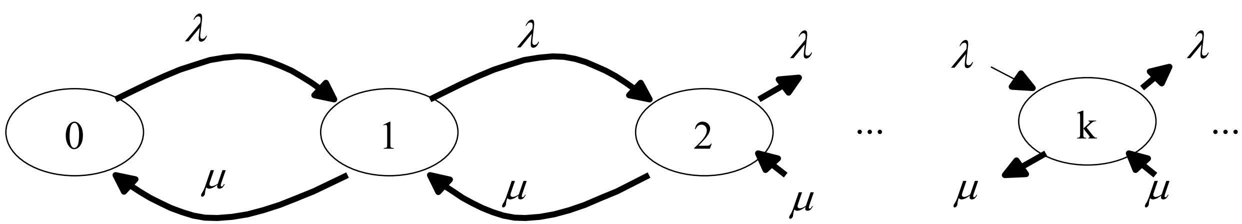
    Prob{two or more arrivals} ≈ 0
- Birth-Death Chain for Poisson Arrival/Service
  - Transition to state k can only be made from state k-1 via a birth (arrival) or from state k+1 via a death (departure)

M/M/1 Birth-Death Markov Chain

Service rate = μ

Average service time = 1/ μ

The state of an M/M/1 queuing system is the number of jobs in the system.



Condition for Stability

$$\lambda < \mu \quad \text{or} \quad \lambda/\mu < 1$$

Solving the M/M/1 Queue

- Equilibrium Balance-of-flow Equations
  - For a stable system in the long run, the number of transitions from state j to state j+1 is equal to the number of transitions from state j+1 to state j
  - If π_j denotes the steady state probability that the system is in state j, then $\lambda\pi_j$ is the number of transitions per unit of time from state j to state j+1. Consequently,

$$\lambda\pi_j = \mu\pi_{j+1}$$

Solving the M/M/1 Queue (continued)

- Normalization Condition: $$\sum_{j=0}^{\infty} \pi_j = 1$$
- Solving the balance of flow equations and the normalization condition, we get

$$\pi_k = \rho^k \pi_0$$

$$\pi_0 = 1 - \rho$$

where $\rho=\lambda/\mu$ is the utilization factor

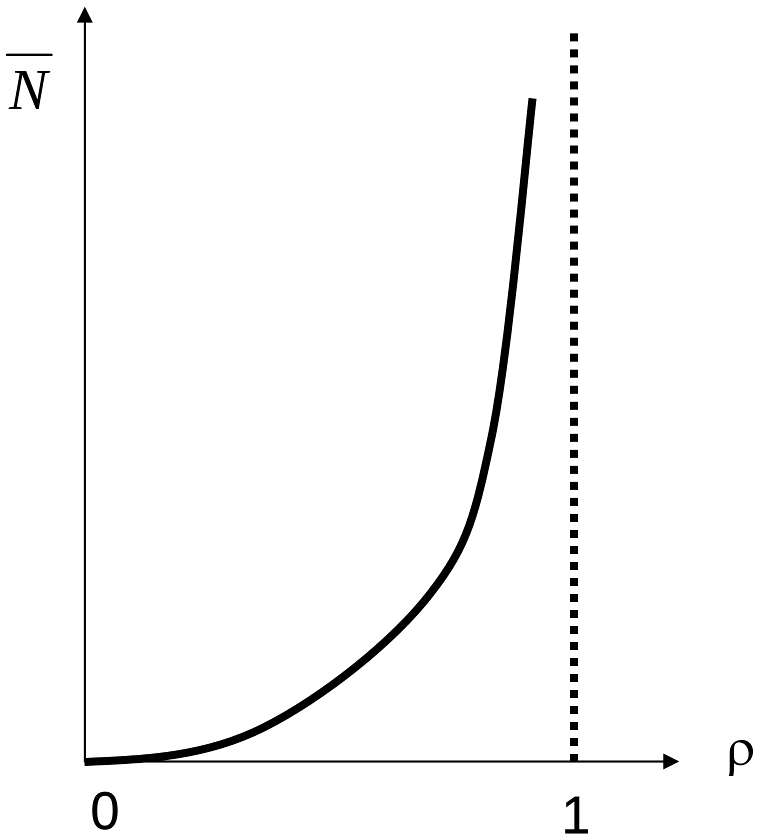
Solving the M/M/1 Queue (continued)

- Average number of customers in system

$$\overline{N} = \sum_{k=0}^{\infty} k\pi_k$$

$$\overline{N} = \frac{\rho}{1-\rho}$$

where $\rho=\lambda/\mu$



Little's Result: $\overline{N} = \lambda T$

- The average number of customers in a queuing system is equal to the average arrival rate of customers to that system, times the average time spent by a customer in that system
- Little's result is applicable to the entire queuing system or to any of its subsystems.
- Applying Little's result to the M/M/1 system

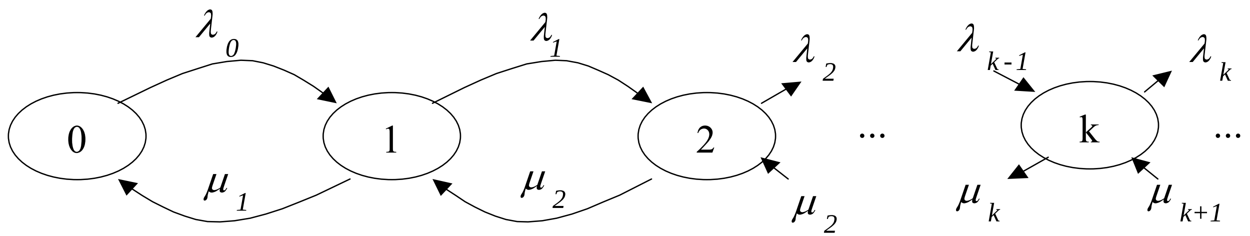
$$T = \overline{N} / \lambda = \frac{1/\mu}{1-\rho}$$

- Applying Little's result to the server

$\overline{N}_{server}$ = λ * average service time

= λ* $(1/\mu)$ = ρ

State-dependent Arrival/Service Rates



Arrival rate in state k = λ_k

Service rate in state k = μ_k

The system can be solved using the same approach used in the M/M/1 queue

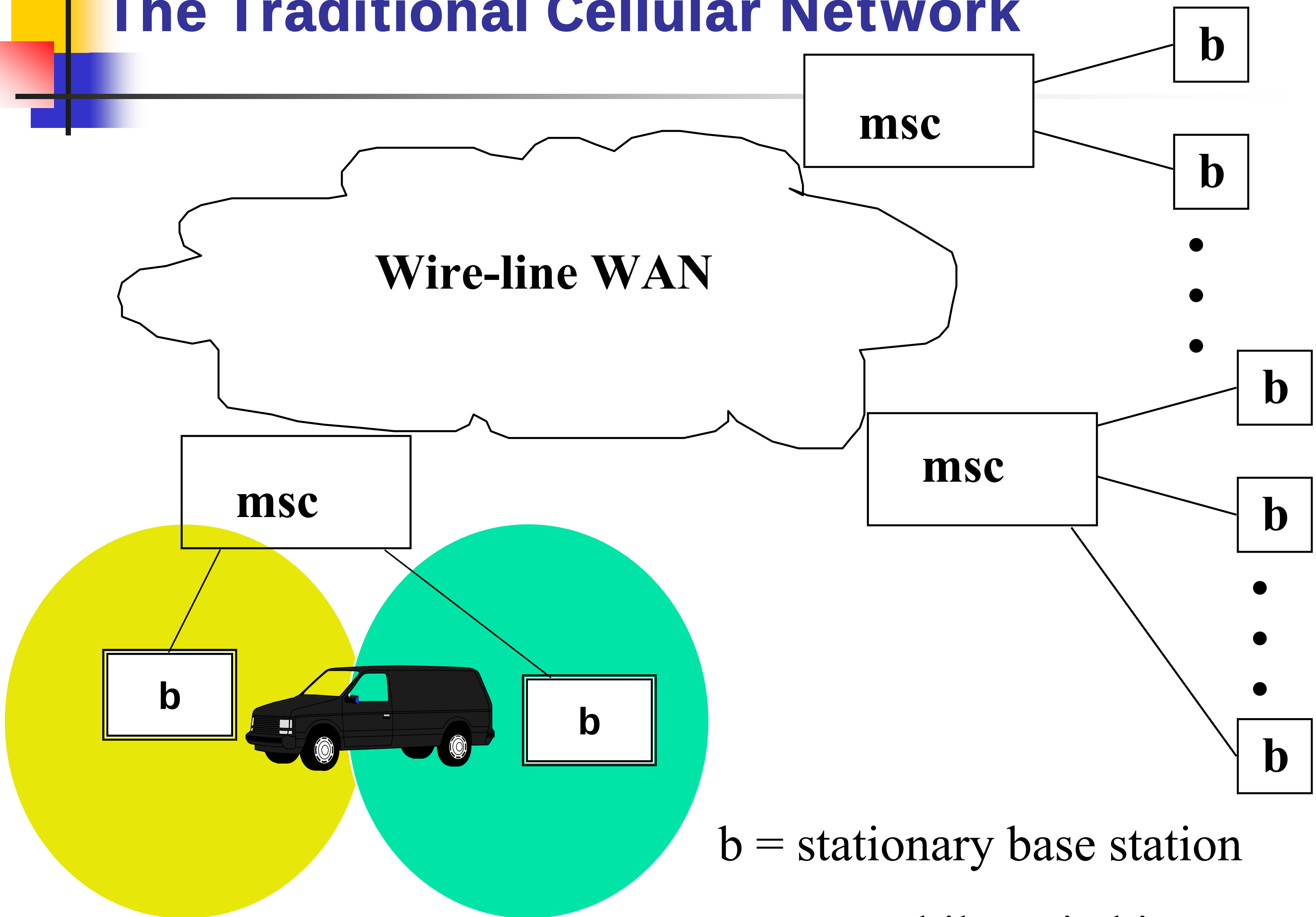
Some Other Queuing Systems

- Systems with multiple servers
- Systems with finite capacity
- Systems with finite population
- Networks of Markovian Queues
- Jackson Networks
- The M/G/1 Queue

Application # 1

Predictive Schemes for Handoff Prioritization in Cellular Networks Based on Mobile Positioning

The Traditional Cellular Network

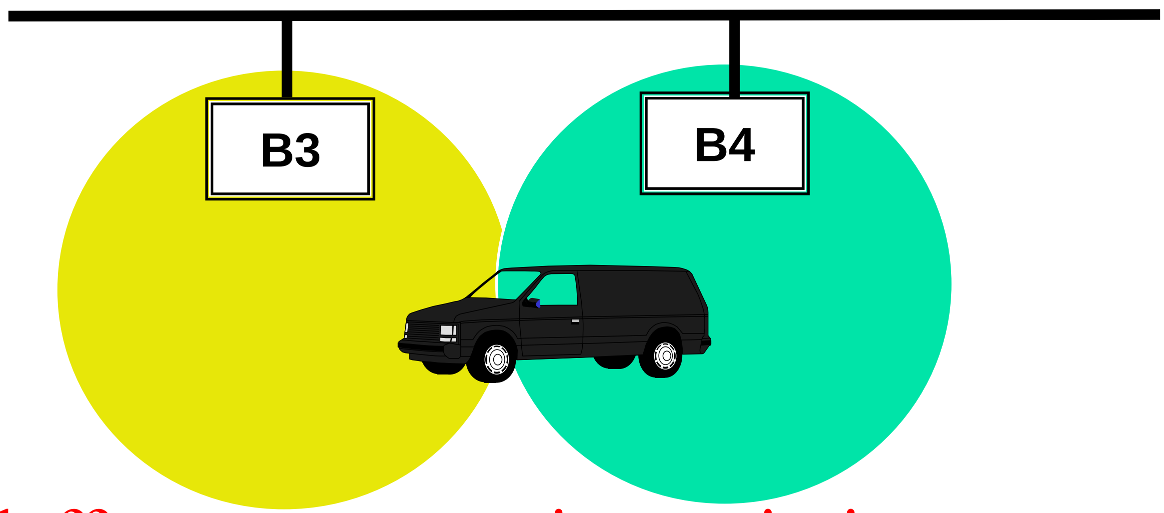


b = stationary base station

msc = mobile switching center

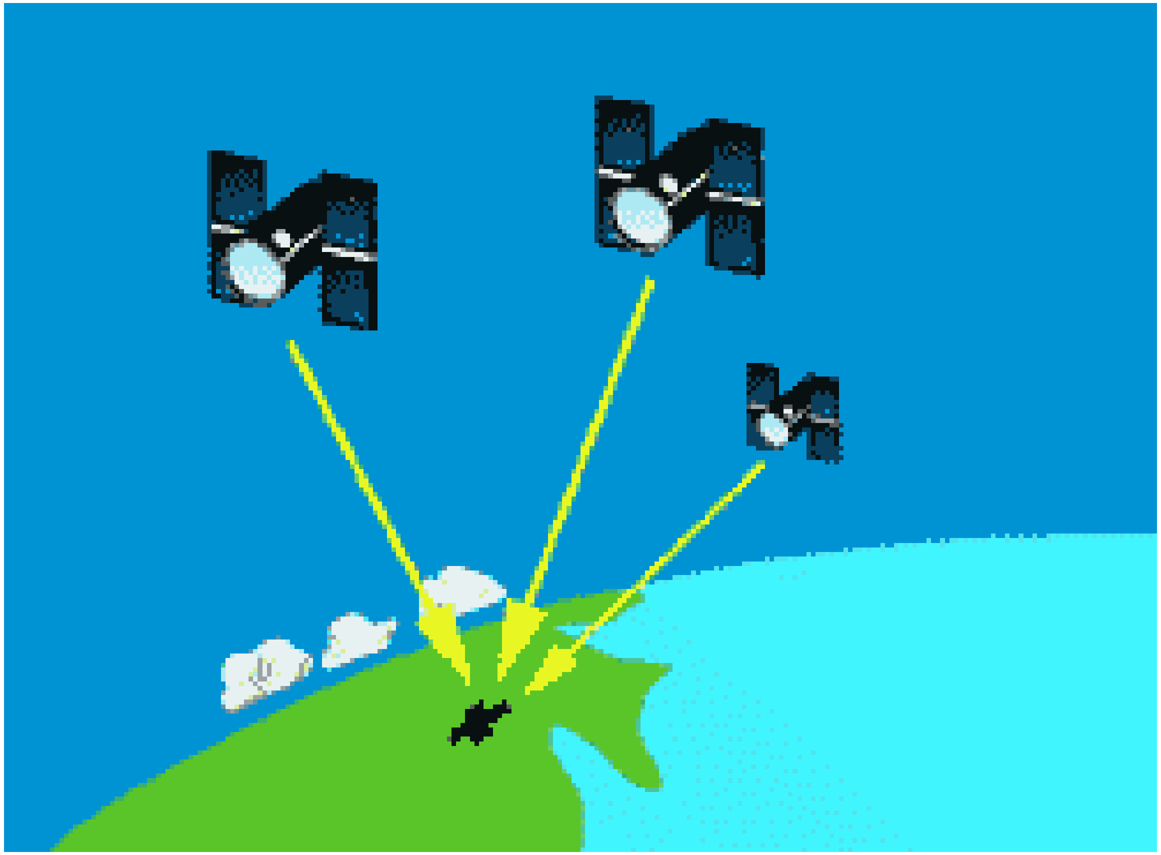
Cellular Networks

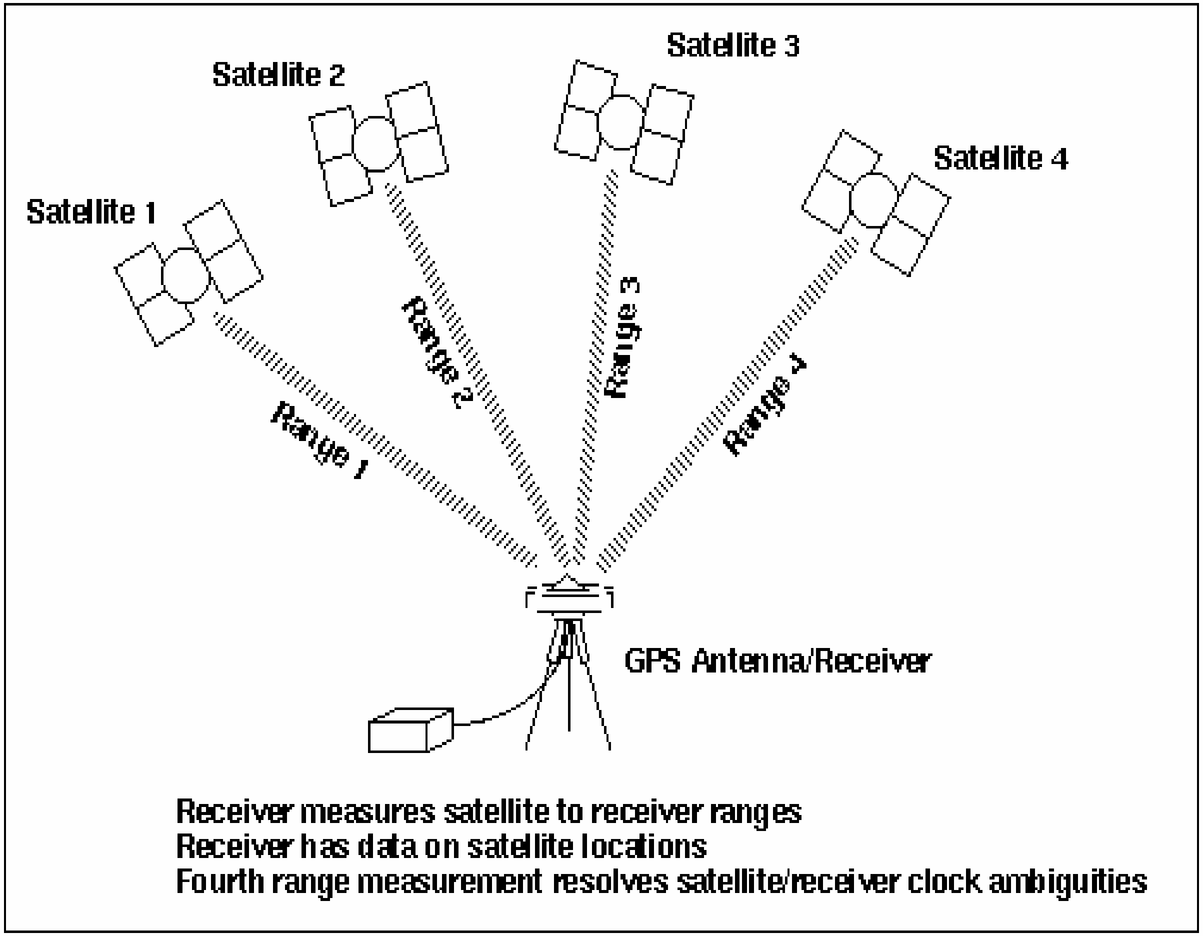
- Handoff Dropping
  - **the cell does not have a free channel and declines to accept the incoming handoff request; the call is disconnected.**
- New Call Blocking
  - **the cell does not have a free channel and declines to accept a new call generated inside the cell.**



Handoff requests are given priority over new calls

Global Positioning System (GPS)



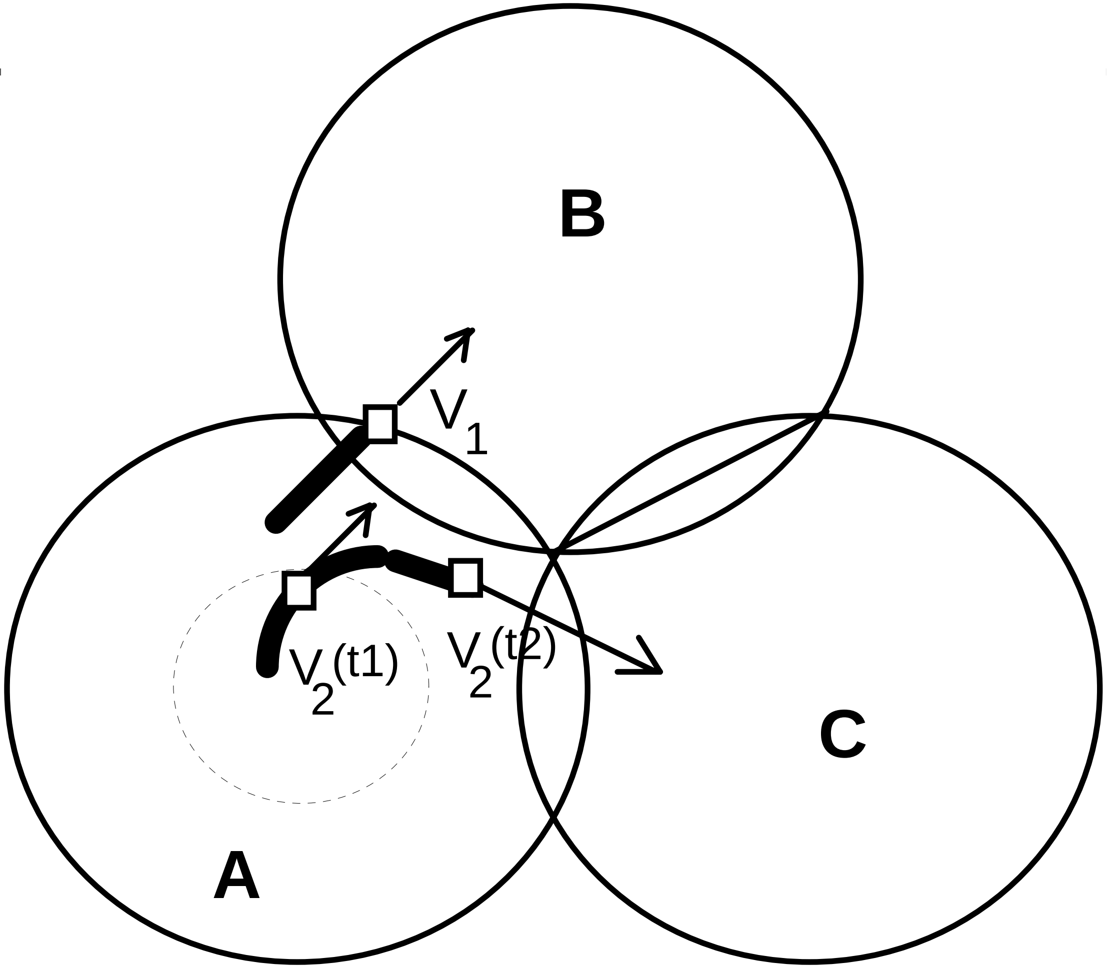


Four imperfect measurements give good accuracy

Geolocation-based Predictive Channel Reservation (Handoff Prioritization)

- Real-time GPS measurements of position and orientation are periodically performed
- The current base station extrapolates the path of the mobile to determine the expected future cell.
- The current base station informs the new base station in order to pre-allocate a channel for the expected handoff request.
- Cancellation of reservation is also sent if the mobile changes its direction.

Predictive Channel Reservation



•At time t1, a reservation for V2 is sent from A to B
•At time t2, A sends a cancellation for V2 to B and a reservation to C

Simple Predictive Channel Reservation (SPCR)

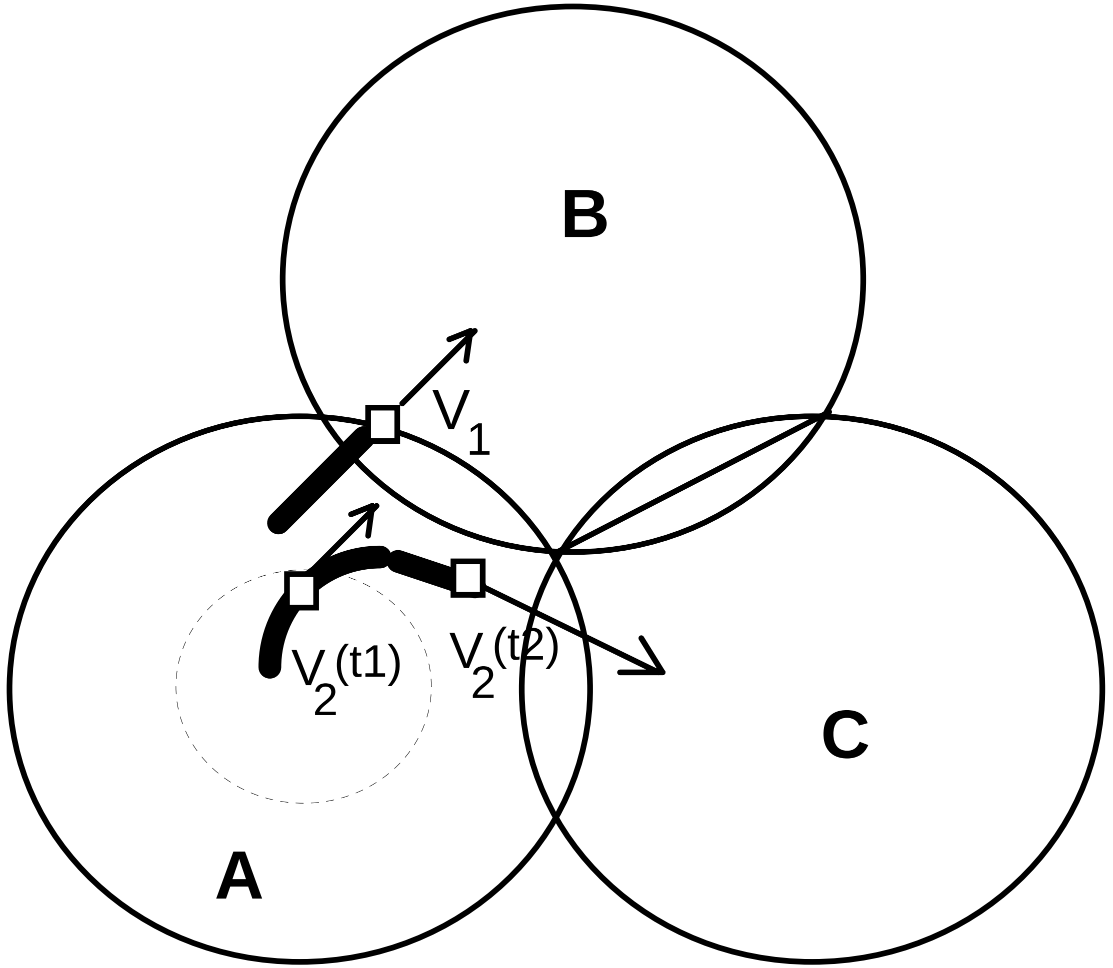
The base station uses the position/orientation measurements to make extrapolation for the projected future path of the mobile station. Channel reservations and cancellations are sent according to these predictions.

False Reservations

Motion Cancellation: the mobile deviates from its predicted path and heads to a different neighboring cell.

Termination Cancellation: the ongoing call terminates before the mobile reaches the new cell boundary.

False reservation



False reservation in cell B from time t1 to time t2

Design Enhancements

- Reservation Pooling.
- Incorporating Guard Channels.
- Queuing of Reservation Requests.
- Careful Selection of the Threshold Distance.

These design enhancements have produced a highly efficient PCR scheme that surpasses previous proposals in achieving significant handoff prioritization while only incurring remarkably small increases in the new call blocking rates.

Reservation Pooling

- The set of reserved channels at any moment is used as a generic pool to serve handoff requests but not new calls.
- Reservation pooling has given good performance improvement.
- Handoff Algorithm with Pooling

```
If  (there is a reserved channel)
     {allocate the reserved channel}
elseif  (there is a free channel)
          {allocate the free channel}
else  {drop call}  // blocking
```

Guard Channels

- Notation
  - **C** = Total number of channels
  - **C-T** = Number of Guard Channels
- A new call is accepted only if the number of free channels is equal to or less than T. A handoff call is accepted if there is one or more free channel.
- Example:

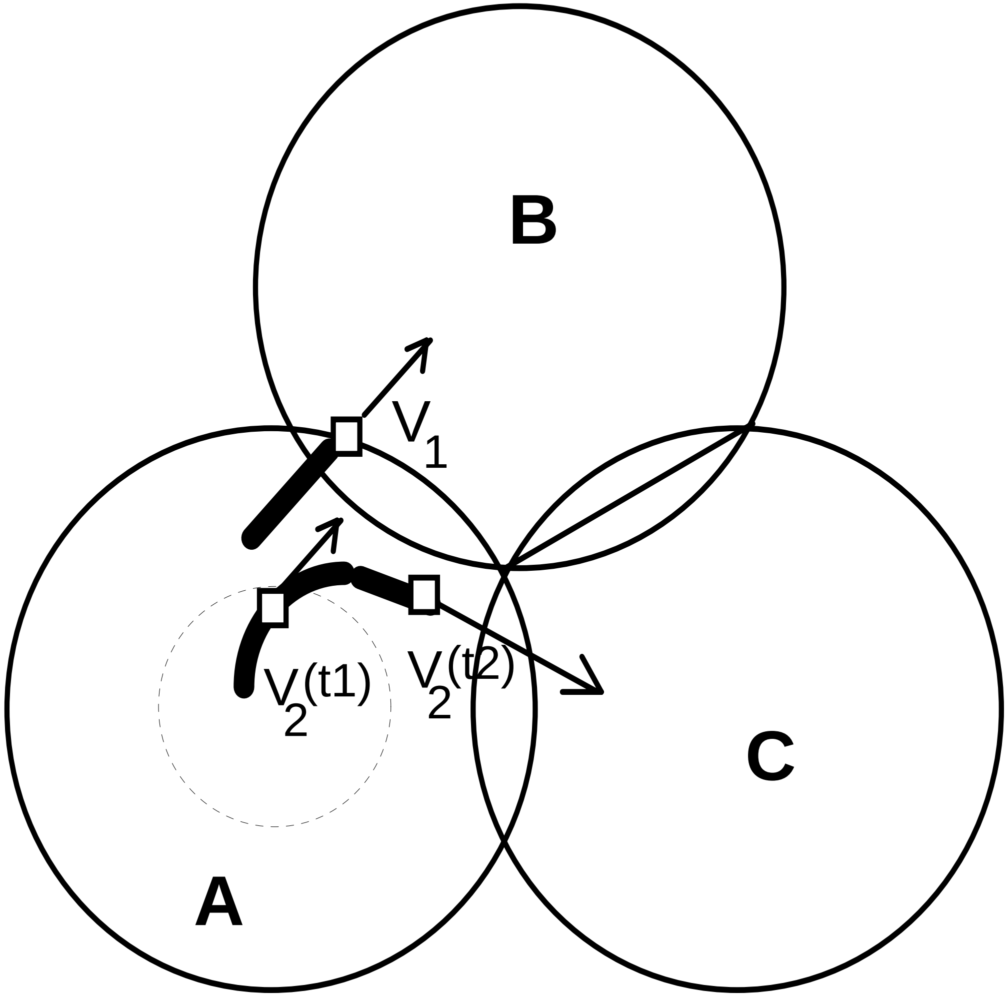
  C=50

  Number of Guard Channels = 5

  New Call Acceptance Threshold= T= 45

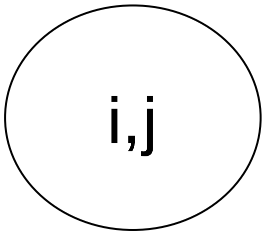
Analytical Markov Model

- Poisson processes
  - New call arrivals
  - Handoff requests
  - Reservation requests
  - Cancellations
- Channel holding time in a cell is assumed to be exponentially distributed
- Reservation pooling is used with C channels and C-T guard channels.



System State and Transition Rates

- Two-dimensional Markov chain with state (i, j) where
  i = numbers of used channels
  j= number of reserved channels



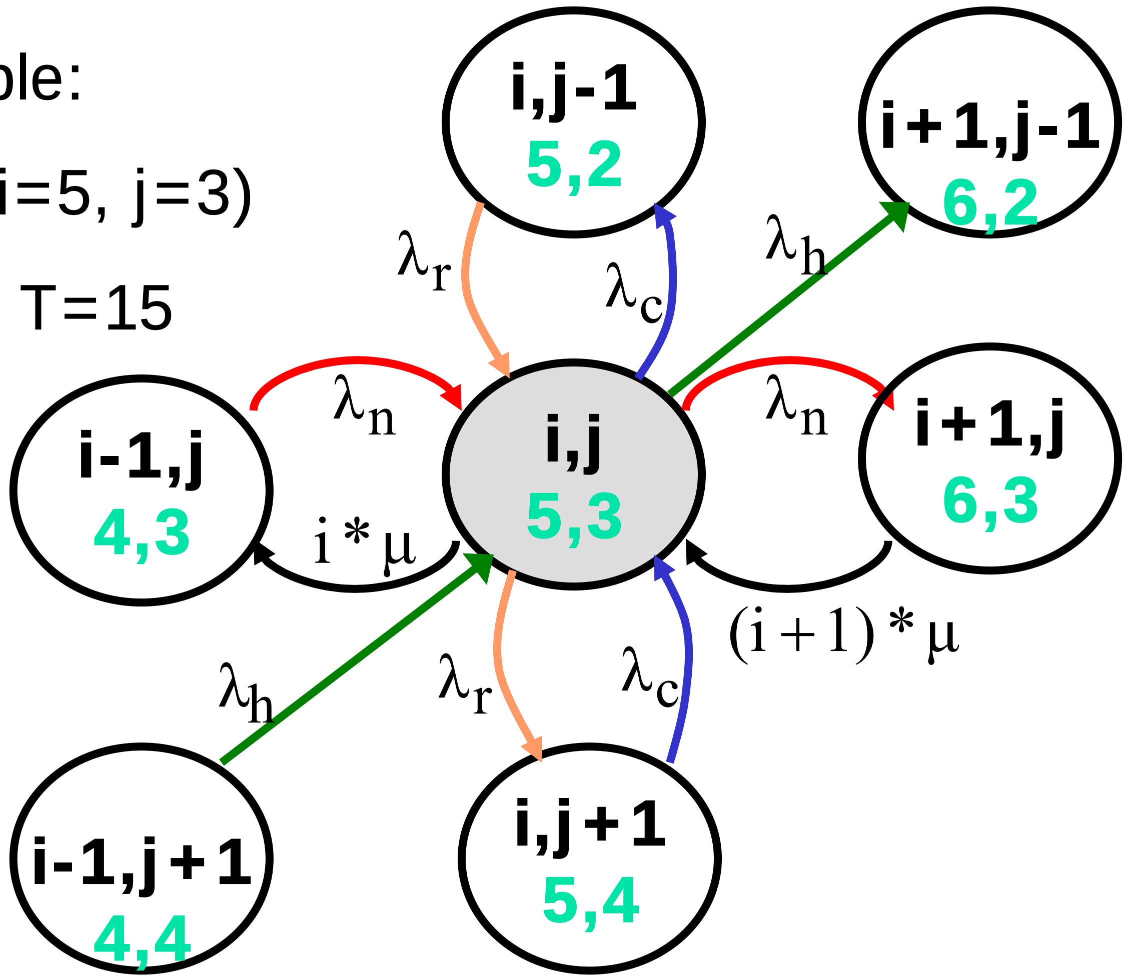
- Poisson Processes
  call termination rate = μ
  new call arrival rate = λ_n
  handoff request rate = λ_h
  reservation request rate = λ_r
  cancellation request rate = λ_c

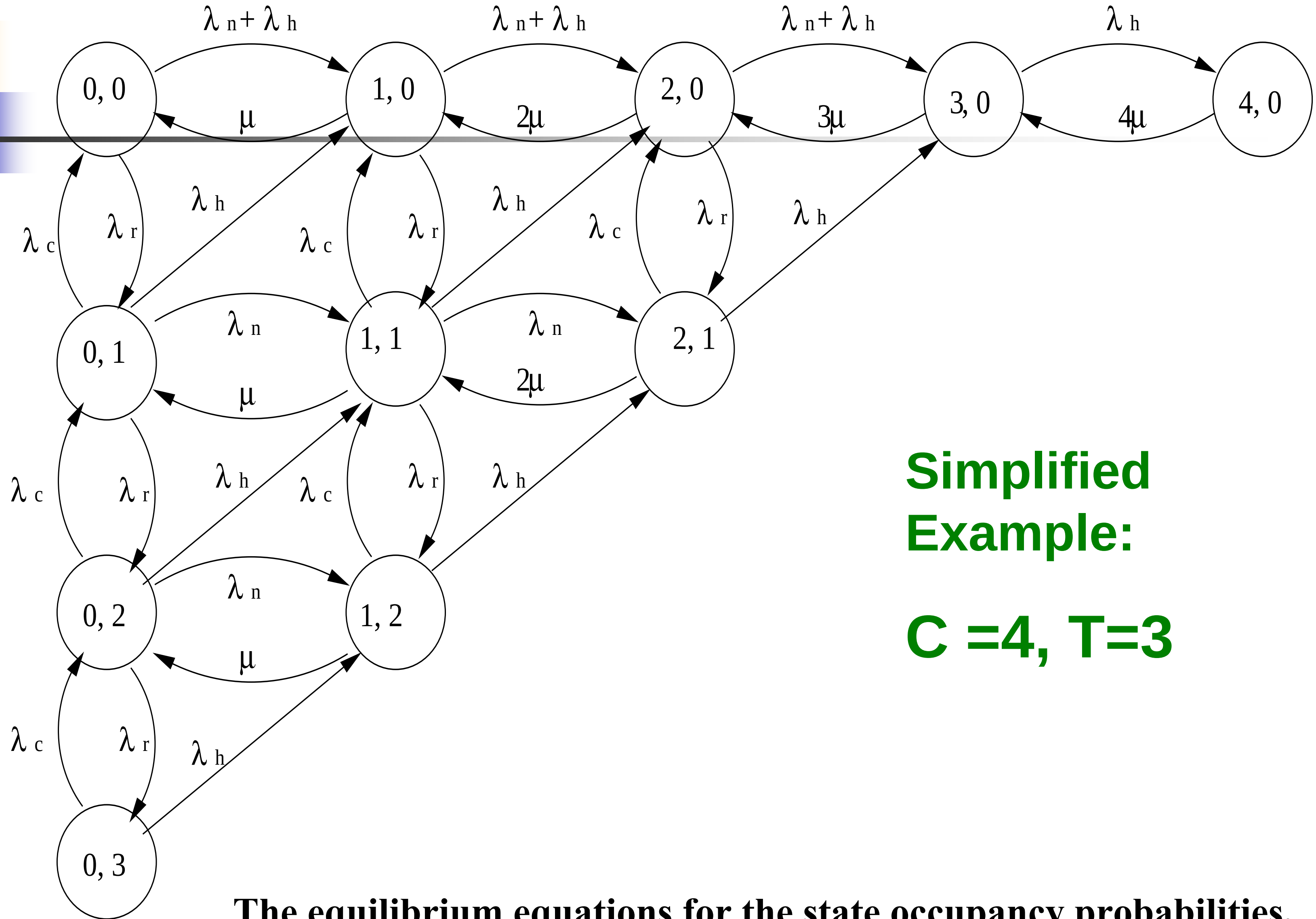
Non-Boundary State Transitions

Example:

State(i=5, j=3)

C=20, T=15

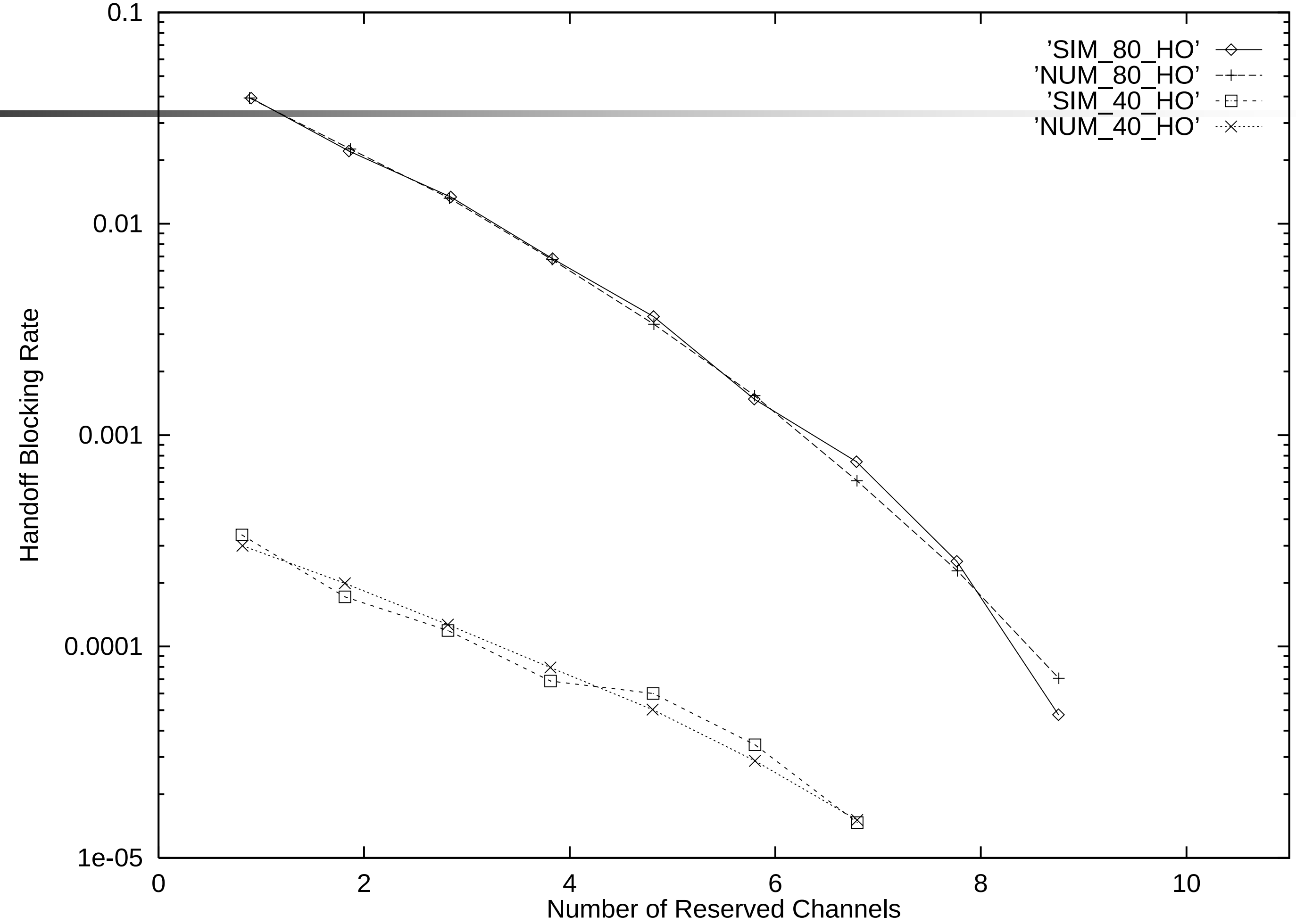




Simplified Example:

C =4, T=3

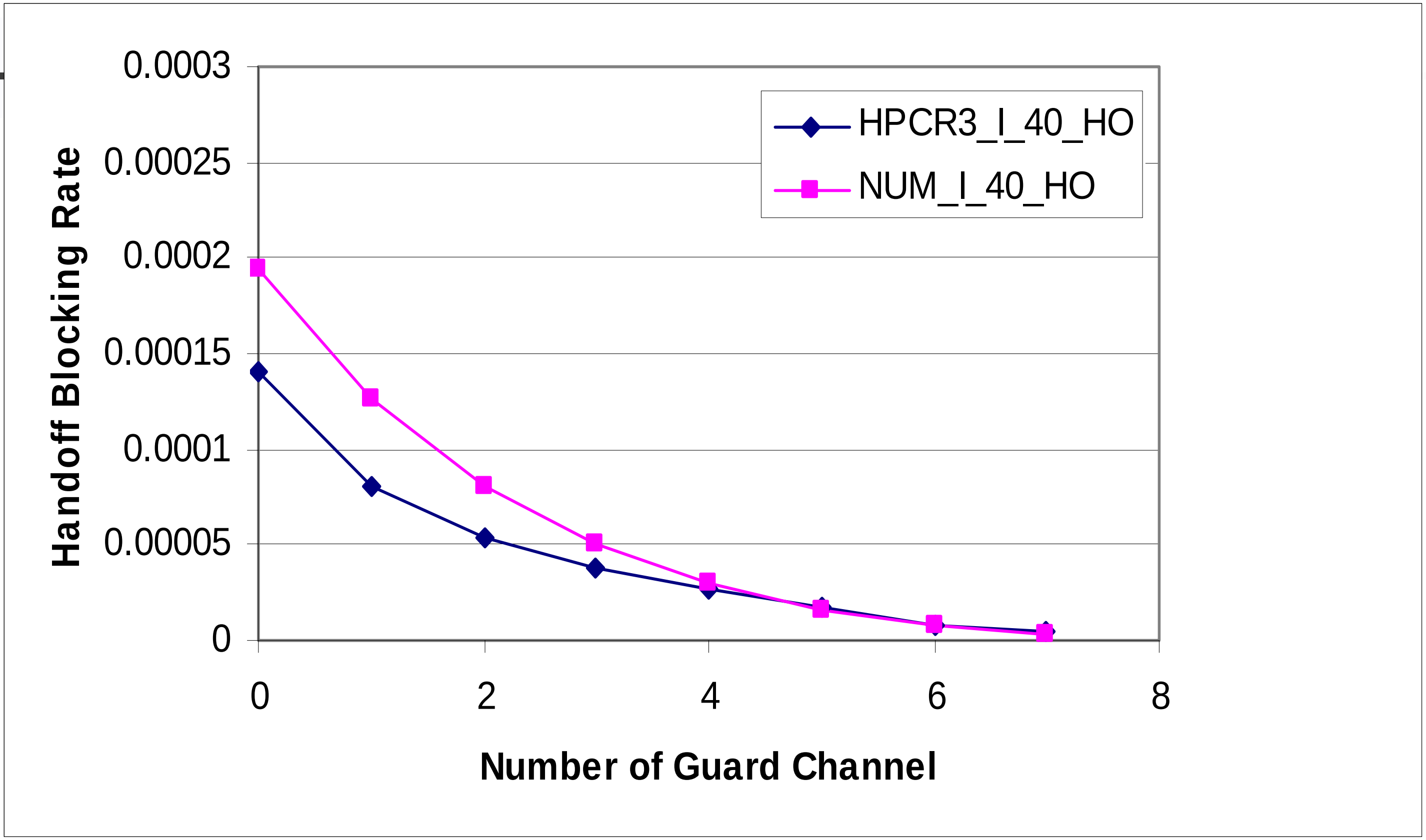
The equilibrium equations for the state occupancy probabilities, boundary conditions and the normalization condition are solved to give the blocking probabilities.



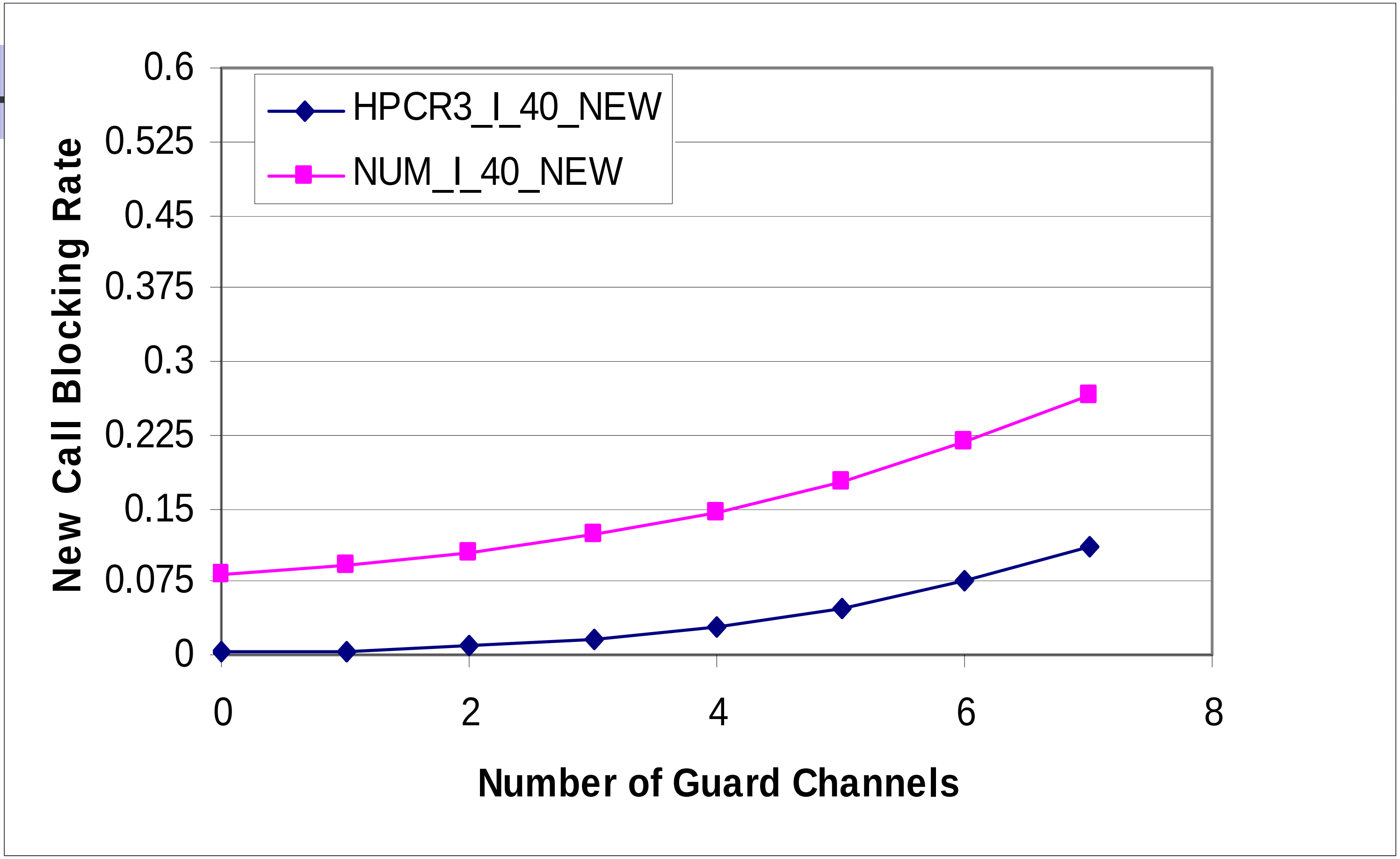
Validation at 40% and 80% traffic Load

NUM →Results from Analytical Model

SIM → Results from Simulation with Poisson Reservation/Cancellation



Discrepancy on Handoff Blocking Rate Between the Simulation Model (HPCR3) and the Analytical Model (NUM)

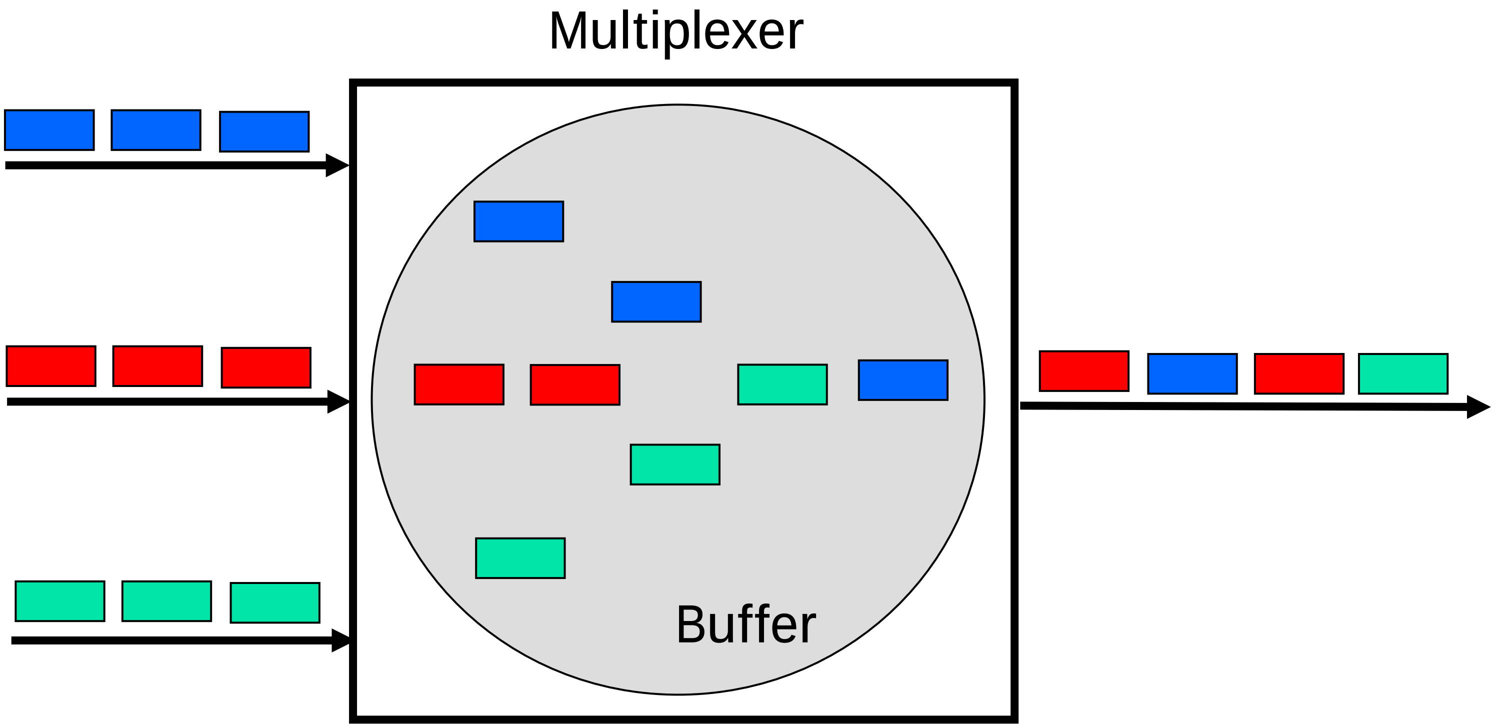


Discrepancy on New Call Blocking Rate Between the Simulation Model (HPCR3) and the Analytical Model (NUM)

Application # 2

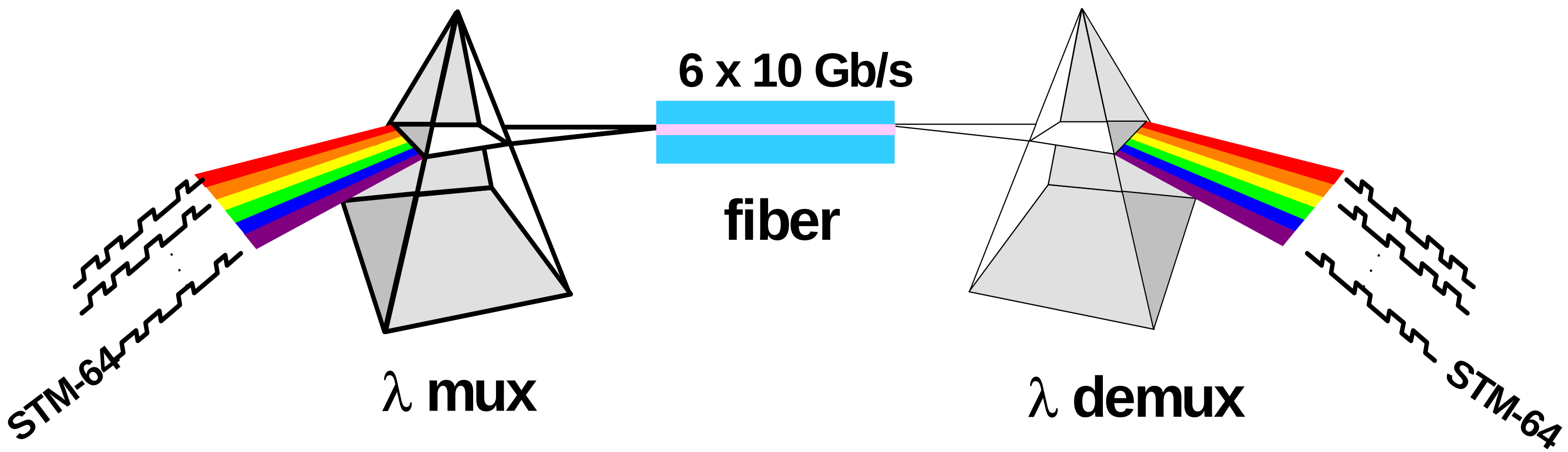
Performance Evaluation and Design of Optical Networks

Multiplexing in Electronic Networks



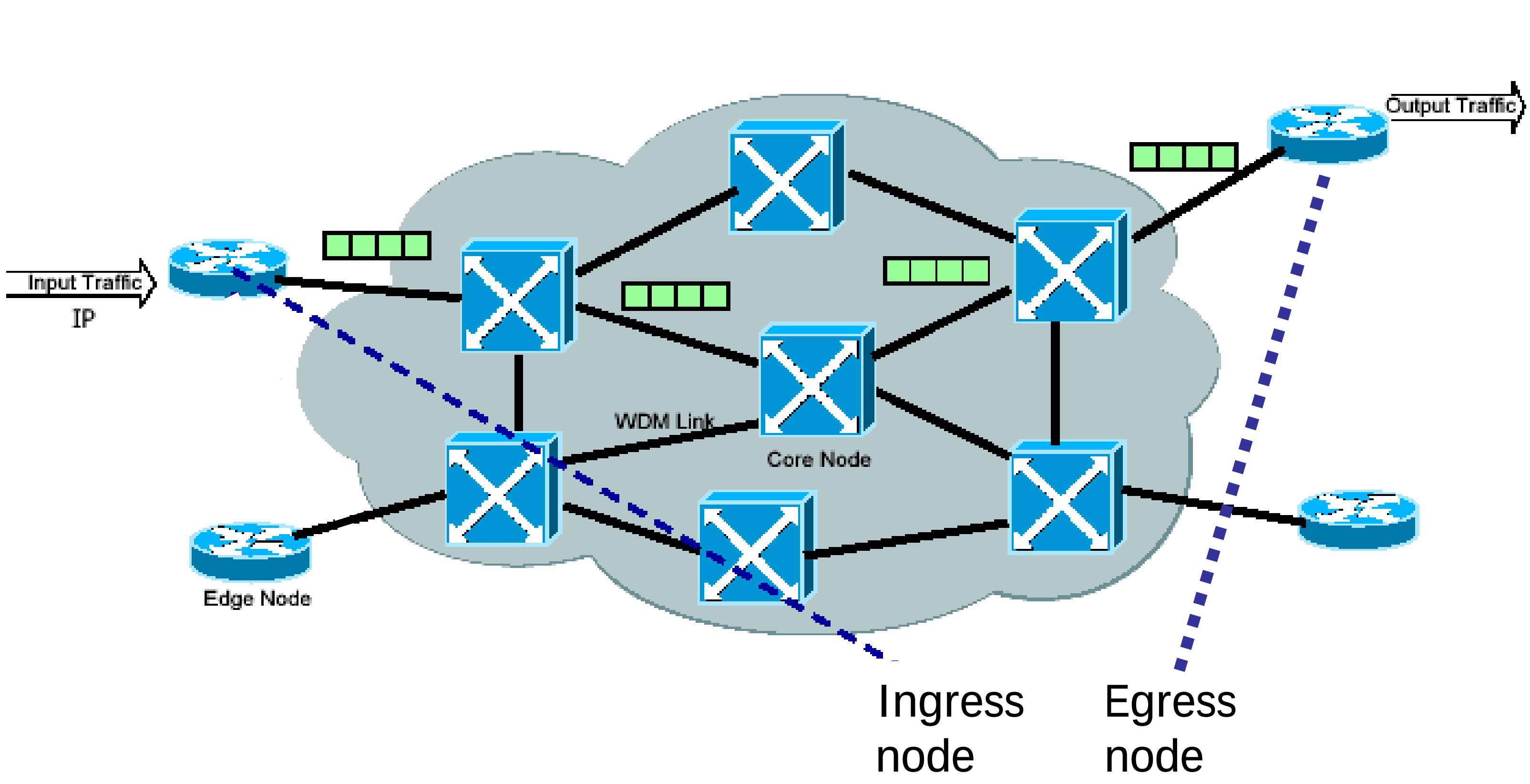
Time-Division-Multiplexing (TDM): Any input or output line can be transmitting at most one packet at a time

Multiplexing in Optical Networks



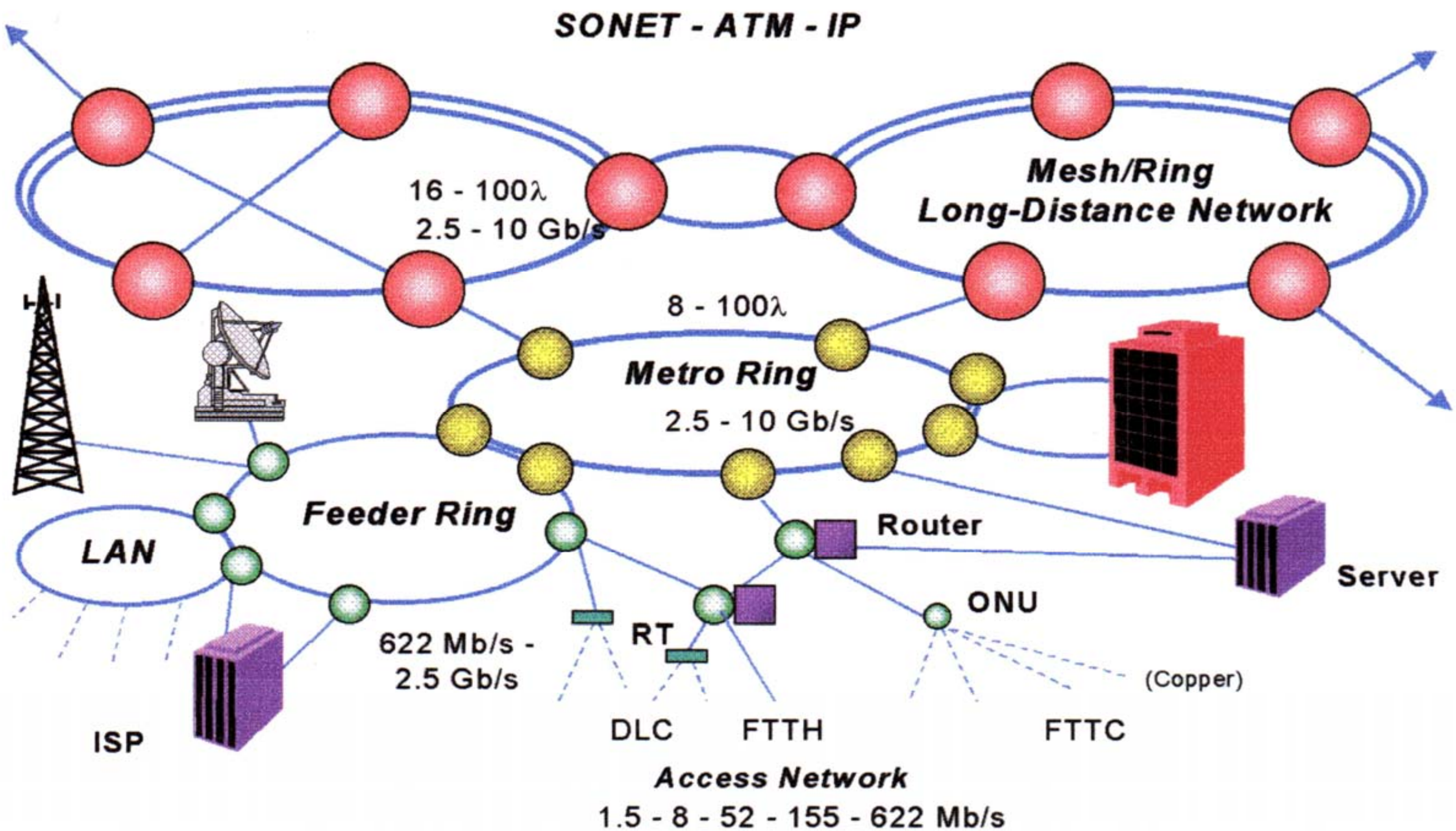
Wavelength-Division Multiplexing (WDM): multiple light signals with different wavelengths are transmitted on the same fiber link simultaneously

Routing in Optical Networks

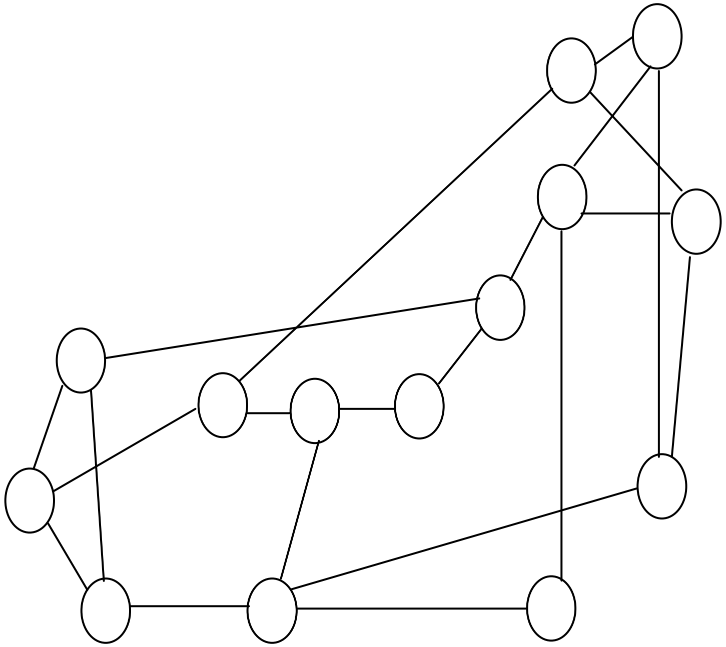


Optical technology is used primarily for transport

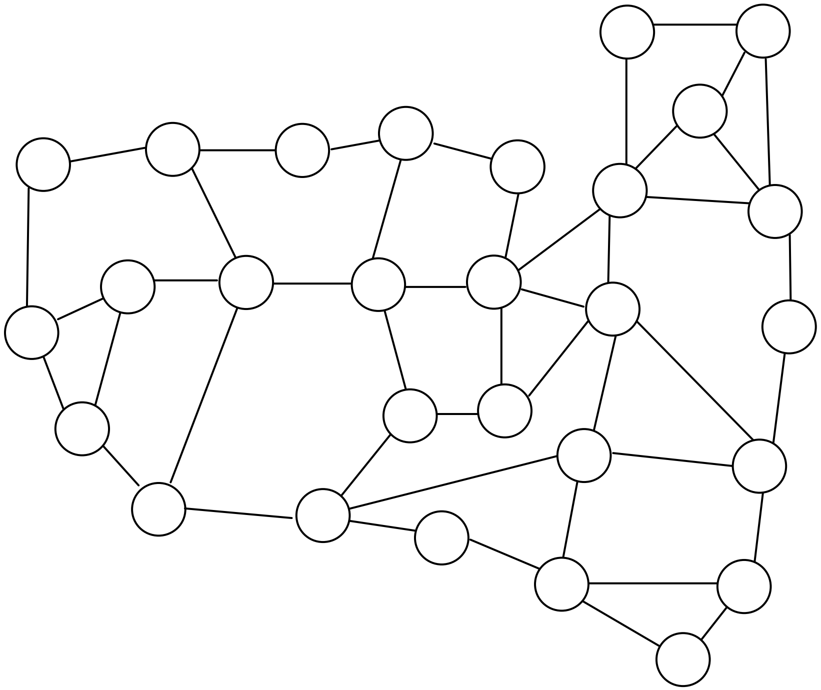
Core, Metro and Access Optical Networks



Optical Network Topologies



NSFNet



U.S. LongHaul

Alarm-based Path Protection in WDM Networks*

- **Network service providers have made significant investments in building operational surveillance and trouble management tools.**

- **The alarm-analysis engines of next-generation networks will have the capability to:**

    - **Process alarms, alerts, and warning messages.**

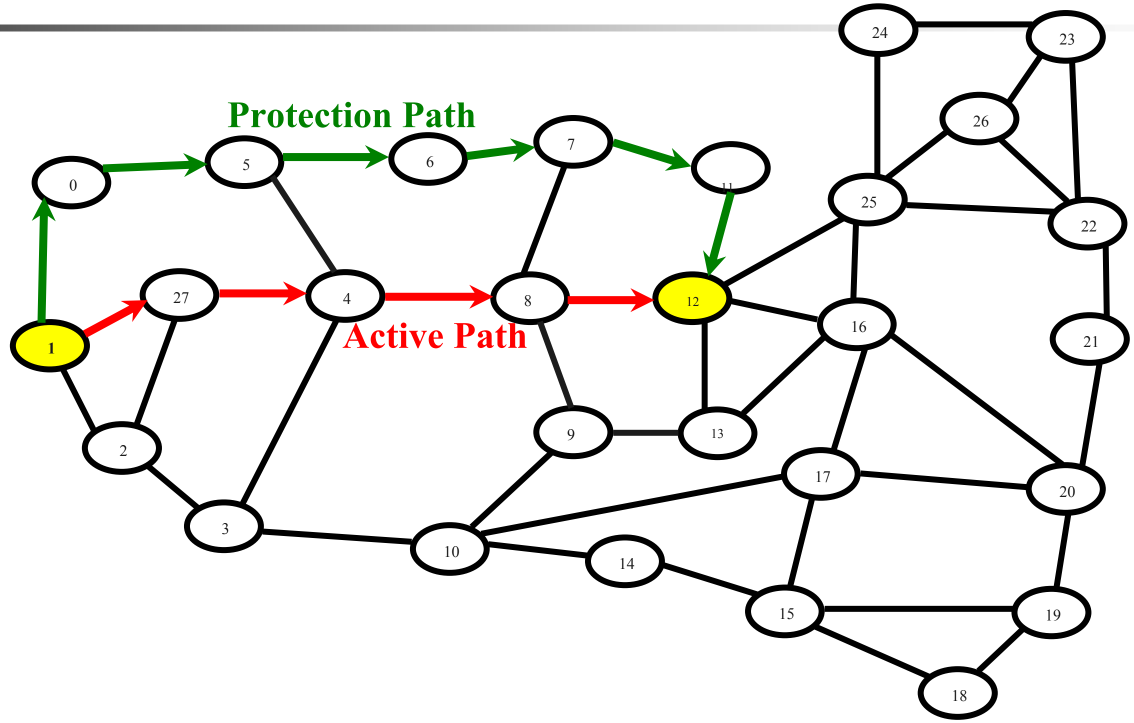
    - **Precisely pinpoint the root-cause problems of failures (e.g., fiber cuts, laser failure, signal cross-talk).**

* Source: paper by M. El Houmaidi et al, Journal of Optical Networking, March 2005.

Network Monitoring & Alarms

- **Network interface controllers (e.g., Cisco NM-AIC-64) remotely monitor all network elements permitting the detection and reporting of alarms on:**
  - **Building security (door or window open/close).**
  - **Fire and smoke indication.**
  - **Building environmental state (temperature and humidity) and utility power readings.**
  - **Flooding or excessive vibration.**

Path Protection in WDM Networks



- **Dedicated Path Protection (DPP) reserves a link-disjoint PP. The wavelength of PP is not shared with any other APs or PPs.**
- **Disjoint Shared Path Protection (DSPP) allows protection links to be multiplexed among multiple APs at no cost.**
- **Joint Shared Path Protection (JSPP) similar to DSPP as long as a link failure does not entail the activation of more than one PP on any wavelength on any link.**

Network Failure Model

- $T_V(t)$ and $T_L(t)$ are the corresponding "tolerance to failure" probabilities for a single node V and a single link L, respectively.
- If a path P is composed of M vertices (V_1, V_2… V_M) and M-1 links (L_1, L_2… L_{M-1}), the tolerance to failure function of path P will be:

$$T_P(t) = \prod_{i=1}^{i=M} T_{V_i}(t) \times \prod_{i=1}^{i=M-1} T_{L_i}(t) \qquad (1)$$

Failure inter-arrival times are exponentially distributed

The expected failure rate for node V and link L are μ_V and μ_L, respectively.

Network Failure Model (continued)

T_P (t) = $e^{-\beta t}$ **and β =** $\sum_{i=1}^{i=M} \mu_{V_i} + \sum_{i=1}^{i=M-1} \mu_{L_i}$ **(2)**

Source node V_1 is assumed to be functioning properly

T_P (t) = $\prod_{i=1}^{i=M-1} TC_{L_i}(t)$ **(3)**

T_P (t) = $e^{-\beta t}$ **and β =** $\sum_{i=1}^{i=M-1} \mu c_{L_i}$ **(4)**

Alarm-based Routing/Path Protection

- **Cost function for link L_i:** $F_C(L_i) = w_1.\mu c_{Li} + w_2.U_W + w_3.\ U_P$

  μc_{Li}: Alarm-based node-link failure rate.

  U_W: Number of wavelengths used by active and protection paths in link L_i.

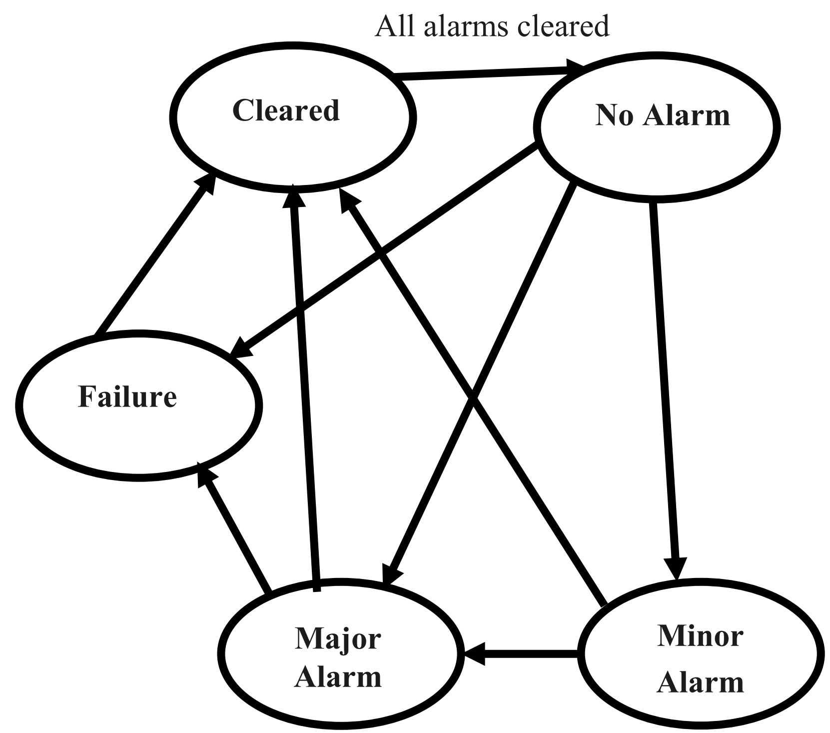
  U_P: Number of wavelengths reserved for protection in link L_i.

- **One failure state and three alarm states: No alarm, Minor alarm, and Major alarm.**

- **Each link (node) has two counts A_{MI} and A_{MA} that keep track of the number of minor and major alarms that have been posted but not yet cleared for that link (node).**

Alarm-based Routing/Path Protection

- **A Minor alarm could escalate to a Major alarm or could get cleared with certain probability.**
- **Similarly, a Major alarm could ultimately cause failure or could get cleared.**
- **At the "No Alarm" state, some failures may occur (suddenly) with no prior alarms and some occur after an alarm escalation.**



FTPP Routing Algorithm

I. Initialization:

1. **No failures in the network: All alarm counts A_{MI}, A_{MA} set to 0.**

2. **All links have W wavelengths available: U_W= 0 and U_P= 0.**

***II. On connection Request between source node s and destination d*:**

1. **Given the alarm counts for all nodes/links in the network, compute μ_{Vi} and μ_{Li}**

2. **For a link L_i (V_i, V_{i+1}), compute $F_C(L_i)$**

3. **Based on the link's cost function, determine a candidate active path (CAP).**

4. **Search for a wavelength, λ_a, that is free on all links of the candidate active path CAP. If there is no free common wavelength on all links of CAP, find the link L_H in CAP with the highest cost, then set $F_C(L_H) = \infty$ and go to step 3.**

FTPP Routing Algorithm

5. **Eliminate all links of AP as follows: for all links Li in AP, set $F_C(L_i) = \infty$.**

6. **Find the set $\Im$ of all active paths that share a link with the new AP.**

7. **Determine a candidate protection path (CPP): shortest path between s and d.**

8. **Find wavelength λp such that: 1) λp is not used by any active path that shares a link with CPP and 2) for each AP $P_i \in \Im$, if the PP for P_i uses wavelength λ_{ip} and shares a link with CPP, then λp≠λip. Otherwise, eliminate link in CPP with the highest cost and go to step 7.**

III. On new alarm event posted for link Li or node Vi:

Based on the type of the alarm (minor or major), increment the appropriate alarm counter (AMI or AMA) for link Li or node Vi.

FTPP Routing Algorithm (continued)

IV. On failure of link L or node V:

1. **Set $\mu_L = \infty$ or $\mu_V = \infty$, re-compute the combined failure rates based on μ_L or μ_V.**

2. **Find the set $\Im$ of all active paths (connections) affected by this failure.**

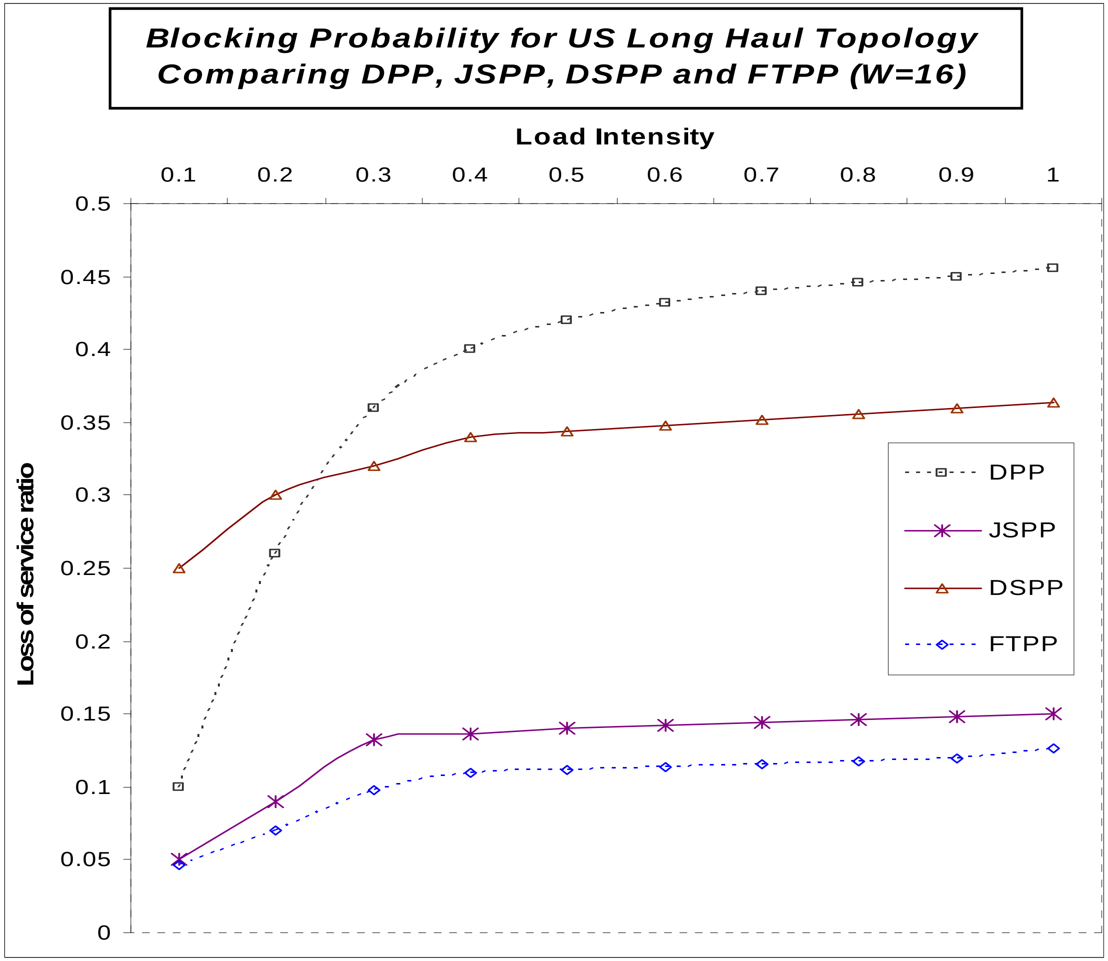
3. **For each active path $P_i \in \Im$, activate the PP of P_i and switch the connection to it. Compute a new PP for the switched connection.**

V. Heuristic for path elimination:

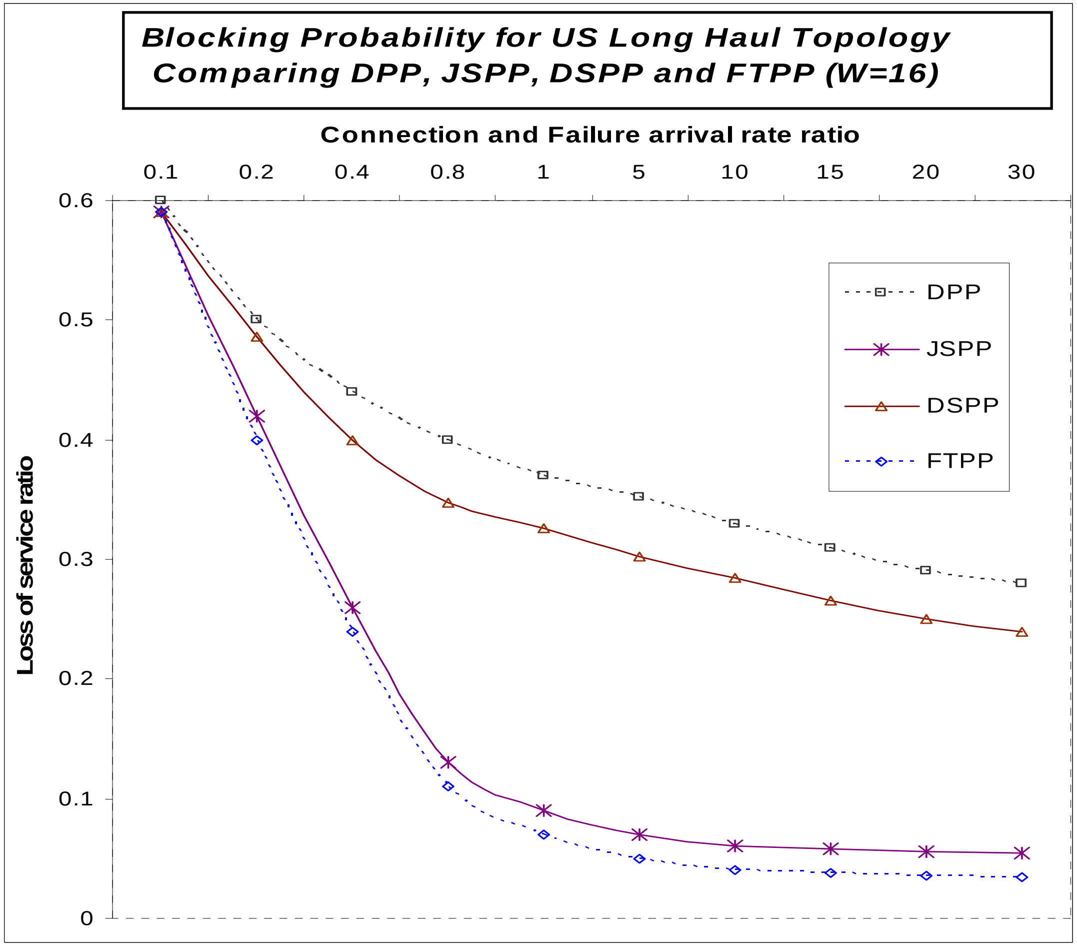
Override the link cost and setting it to infinity in the following two cases:

- **U_W is equal to the total number of wavelengths: $U_W = W$.**
- **Link suffers from crowdedness of PPs: $U_P > 0.5 .W$.**

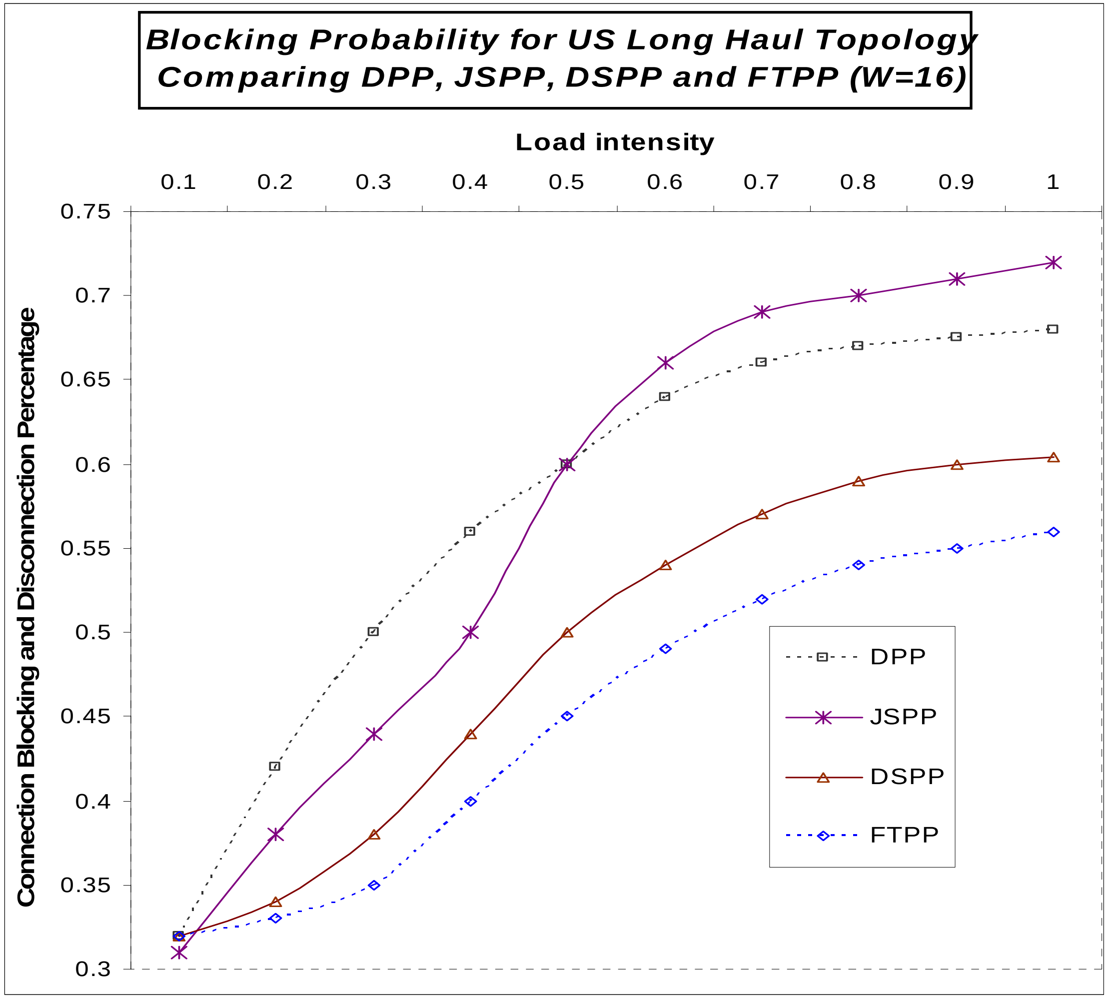
Loss of Service Ratio



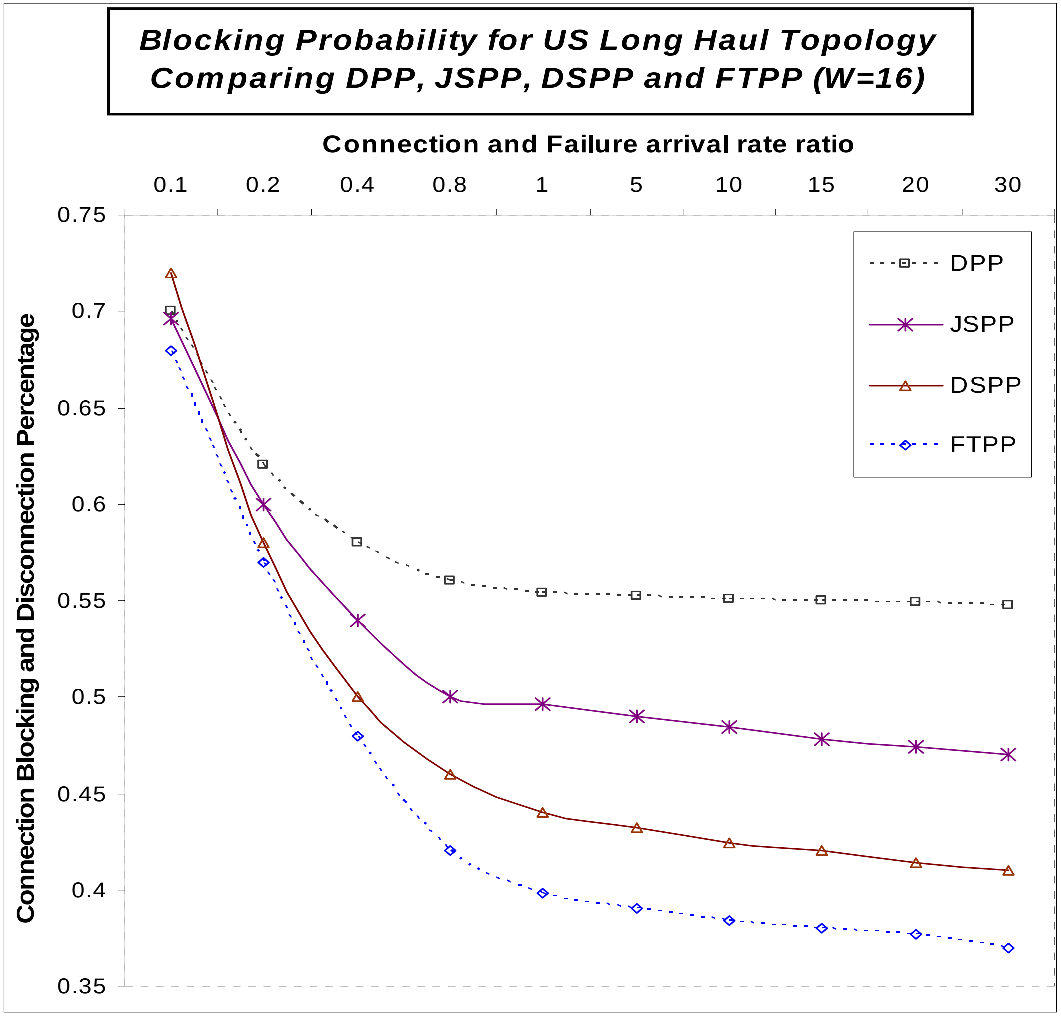
Loss of Service Ratio (continued)



Connection Blocking and Disconnection



Connection Blocking and Disconnection



QoS Enhanced FTTP

(QEFTPP) uses a class based connection admission policy and path protection with priority. We define four classes of connections:

1. Preempted with no protection (PNP): active and backup paths can be pre-empted by higher priority connections.

2. Preempted with shared protection (PSP): Pre-empted by higher priority connections but has a PP fully shared with other PPs.

3. Non-preempted with shared protection (NPSP): Never pre-empted but can get disconnected only if failures affect both the AP and PP.

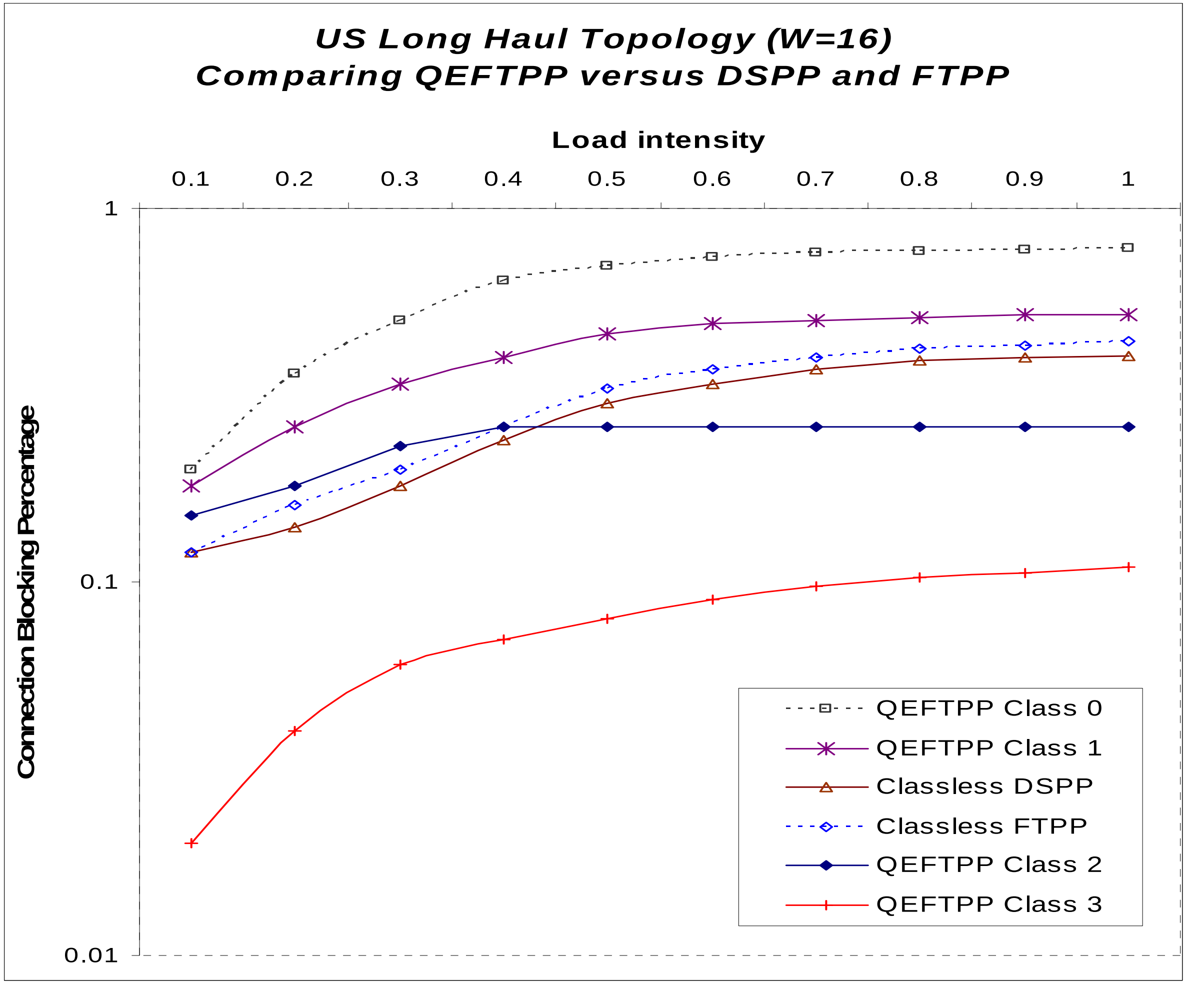
4. Non-preempted with guaranteed protection (NPGP). Highest priority connections. PP can change as a result of newly posted link/node alarms.

Can be preempted

Cannot be preempted

Increasing Priority

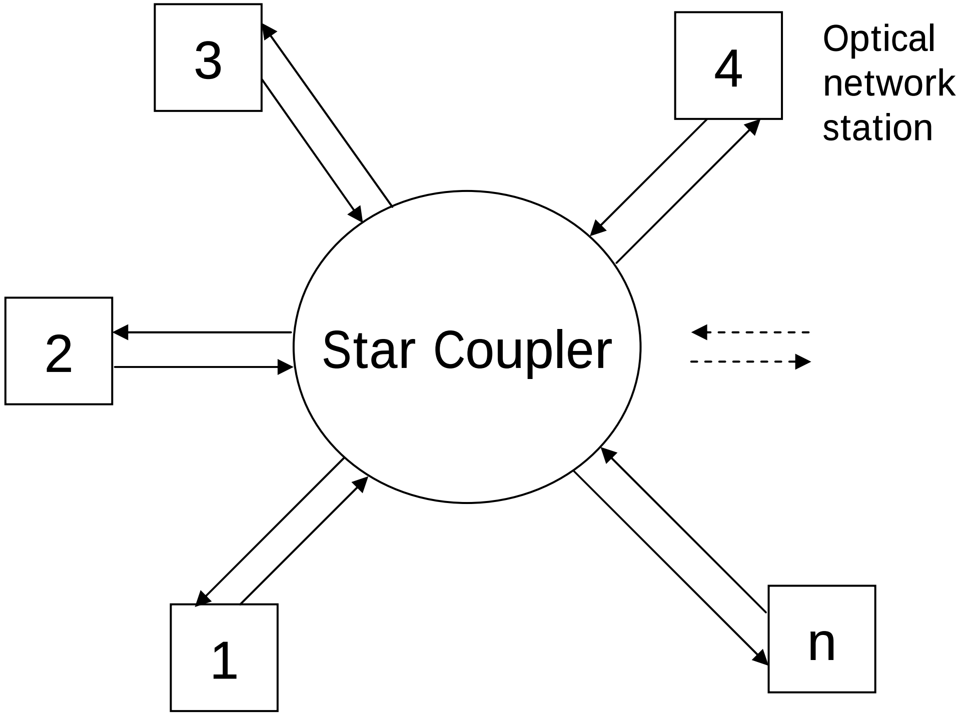
QEFTPP: QoS Differentiation



Optical Network Station

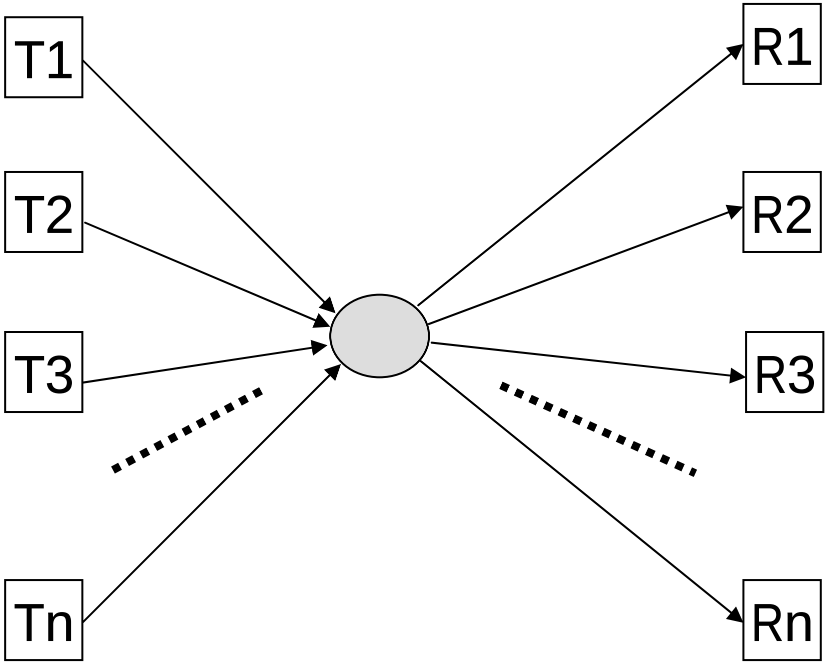
- Transmitter (Laser diode)
  - Fixed Transmitter (FT): operates only on a specific wavelength
  - Tunable Transmitter (TT): can be tuned to one of multiple distinct wavelengths
- Receiver (Photodetector)
  - Fixed Receiver (FR): operates only on a specific wavelength
  - Tunable Receiver (TR): can be tuned to one of multiple distinct wavelengths

Simple Optical LAN: Broadcast Star Coupler



The star coupler combines all inbound signals and broadcasts them on each outbound access fiber

Logical Representation of the Star Coupler



source stations

destination stations

Star Coupler: Example 1

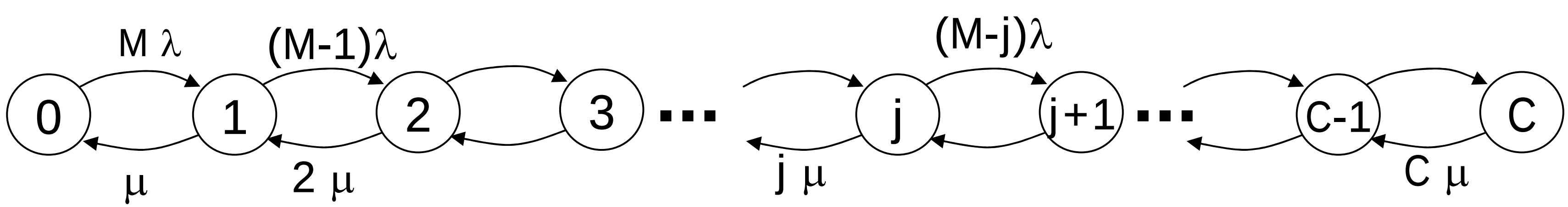
- Number of network stations = M
- There are C wavelengths operational in this LAN, where C << M
- Each source station has a single tunable laser
- Each destination station has C fixed receivers covering all C wavelengths
- Connection requests from a source station are a Poisson process with rate λ. The connection is destined to any station with equal likelihood
- The duration of a connection is exponentially distributed with mean $1/\mu$

Station has one TT and C FR's

Birth-Death Chain of Example 1

- In state j, the system has j active connections. This means that there are C-j free wavelengths in state j.
- Total arrival rate of connections in state j is (M-j)λ and total departure rate is j μ

Station has one TT and C FR's



Solving the balance of flow equations, we get

$$\pi_j = \binom{M}{j} \rho^j \pi_0 = \binom{M}{j} \rho^j \left[\sum_{k=0}^{C} \binom{M}{k} \rho^k \right]^{-1}$$

where $\rho = \lambda/\mu$

Star Coupler: Example 2

- Number of network stations = M
- There are C wavelengths operational in this LAN, where C << M
- Each source station has a single tunable laser
- Each destination station has a single tunable receiver
- Connection requests from a source station are a Poisson process with rate λ. The connection is destined to any station with equal likelihood
- The duration of a connection is exponentially distributed with mean $1/\mu$

Station has one TT and one TR

Birth-Death Chain of Example 2

- In state j, the system has j active connections. This means that there are C-j free wavelengths in state j.

- Total arrival rate of connections in state j is $\lambda(M-j)(\frac{M-j}{M})$
  and total departure rate is $j\mu$

Station has one TT and one TR