

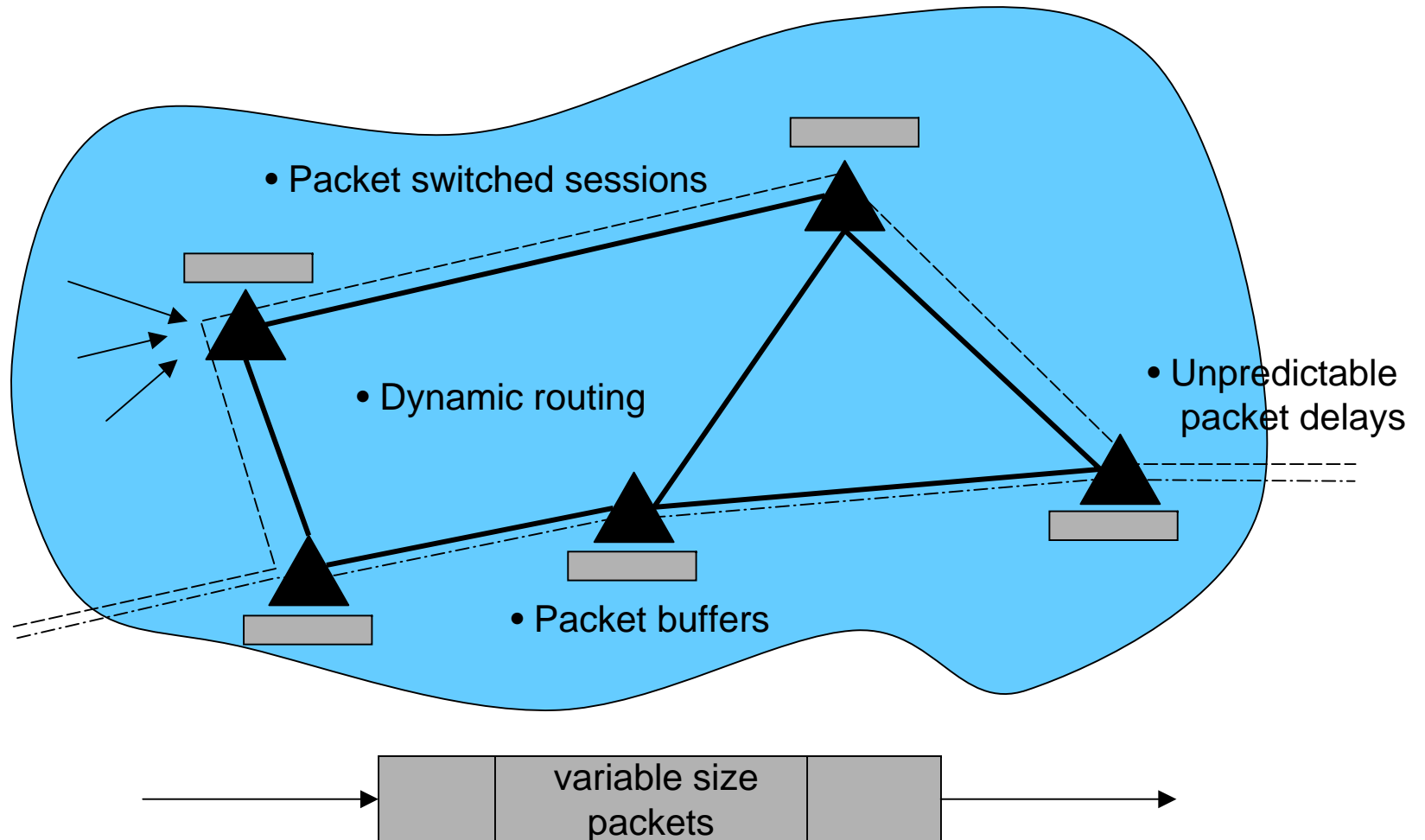
NIS/CpE 691A

Information System Security

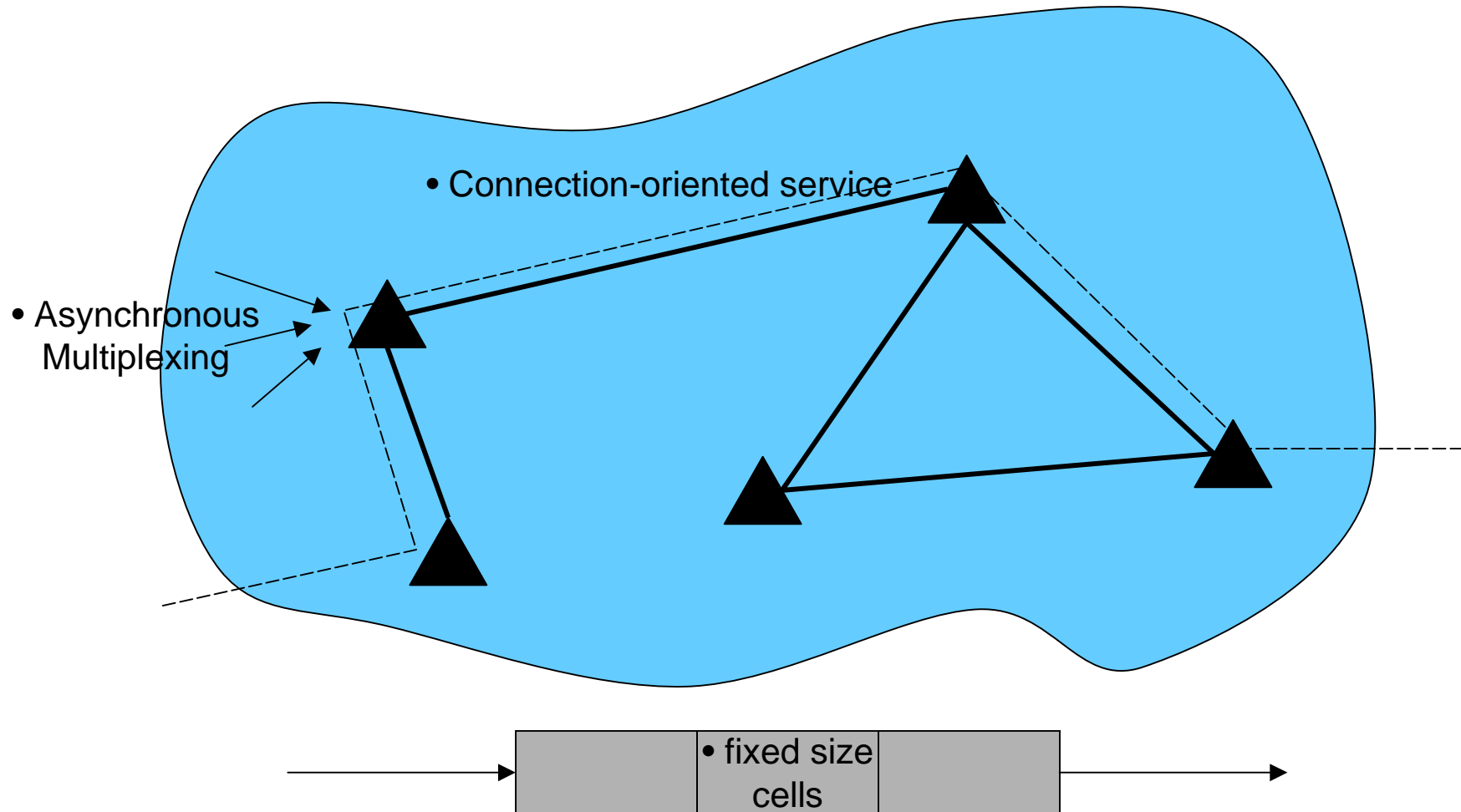
**Stevens Institute of Technology
Spring 2006**

Class 8 – 3/9/06

Characteristics of IP networks



Characteristics of ATM networks



QoS = f(Source Traffic Statistics)

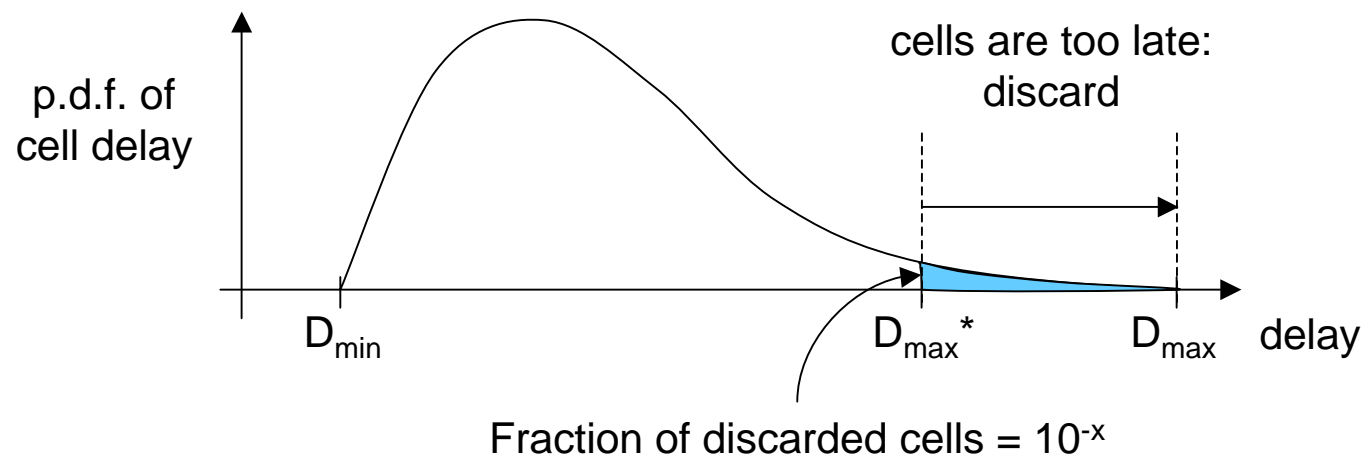
- Survivability was the original key attraction for IP networks
- ATM networks' guaranteed QoS parameters have been their draw:
 - end-to-end delay
 - jitter
 - cell loss probability
- These parameters are important for voice, video, and other delay sensitive services
 - All are guaranteed on a per-call basis if cell-level traffic controls are enforced at UNI
 - QoS is based on peak cell rate (PCR) and sustainable cell rate (SCR)
- However:
 - PCR, SCR are hard to obtain - the user parameter is based on PCR, SCR and maximum burstiness of traffic
 - What interval is PCR measured over?
 - The QoS guarantees are not **analytic** guarantees
- A large number of low-bandwidth calls are easier to predict and control.
- Few high-bandwidth calls can strain network severely
 - This is the Central Limit Theorem working against the network operator

Overview of Traffic Controls

- Traffic controls can only be enforced at the edges of the network:
 - The network switches are moving large volumes of traffic at high speeds; measuring and enforcing traffic flows would slow switches down
 - ATM cells are statistically multiplexed on high-capacity links in the core of the network; enforcing per-user controls at this level is problematic
 - The potentially enormous number of QoS contracts between end users and the network cannot be known at every point in the network
- Traffic controls can **only** realistically be enforced at the edge of the network

QoS Underlying Assumptions

- The primary ATM applications are delay-sensitive: cell delay is the primary QoS metric
- Delay depends on traffic volume, buffer architecture and buffer management



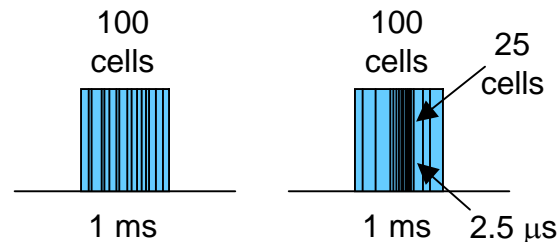
Limitations in Measuring PCR

- Peak Cell Rate:

$$PCR = \lim_{T \rightarrow 0} \frac{N}{T}$$

- The minimum interval, T , depends on the time resolution of the system doing the measuring

- Consider:



- $100 \text{ cells} * 53 \text{ bytes} * 8 \text{ bits} / 1 \text{ ms} = 42.4 \text{ Mb/s}$
- $25 \text{ cells} * 53 \text{ bytes} * 8 \text{ bits} / 2.5 \mu\text{s} = 4.24 \text{ Gb/s}$
- Over the shorter measurement window, the PCR is a factor of 100 greater: The user equipment could readily think that it is conforming to its traffic agreement while the network equipment sees a much higher peak rate.

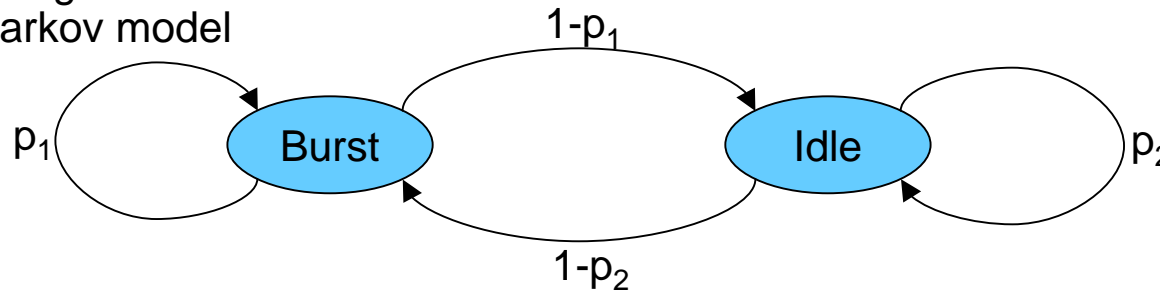
QoS Guarantees and Traffic

- Controlling a traffic source is a key requirement to being able to offer QoS guarantees
- Traffic controls that have been used:
 - resource reservation
 - switch/multiplexor scheduling
 - proactive congestion control
 - call admission control
 - burst admission control
 - peak-rate, sustainable rate policing
 - traffic shaping
 - reactive congestion control
 - window based controls
 - source rate controls
 - assigning cell-loss priorities
 - deliberate blocking of bursts
 - congestion notification

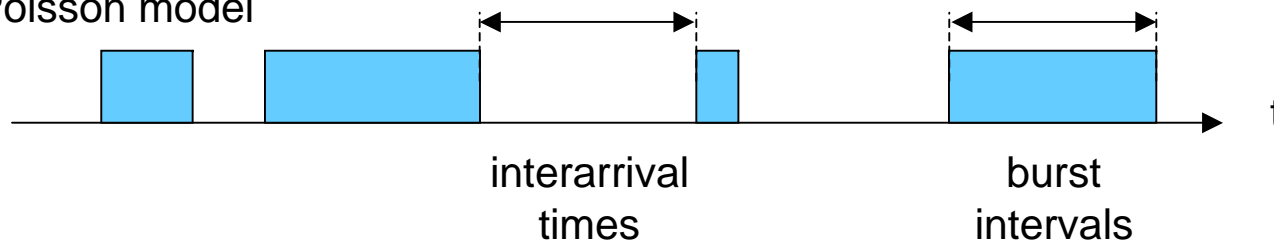
Traffic Characterization

- ATM traffic types:
 - Continuous Bit Rate (CBR) - characterized by a constant rate over long period (e.g., 64 kbps PCM)
 - Variable Bit Rate (VBR) - characterized by its peak bit rate with a lower average bit rate (e.g., compressed video)
 - Unspecified Bit Rate (UBR)

- Modeling burst characteristics:
 - Markov model



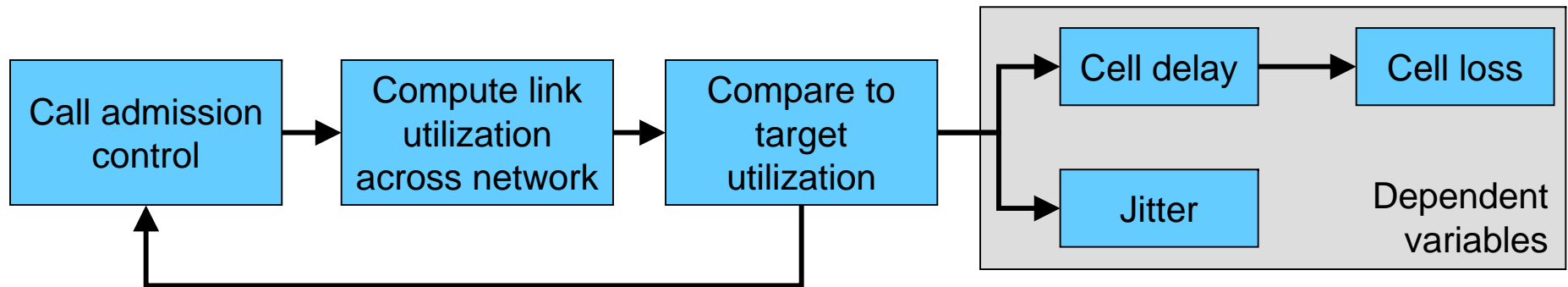
- Poisson model



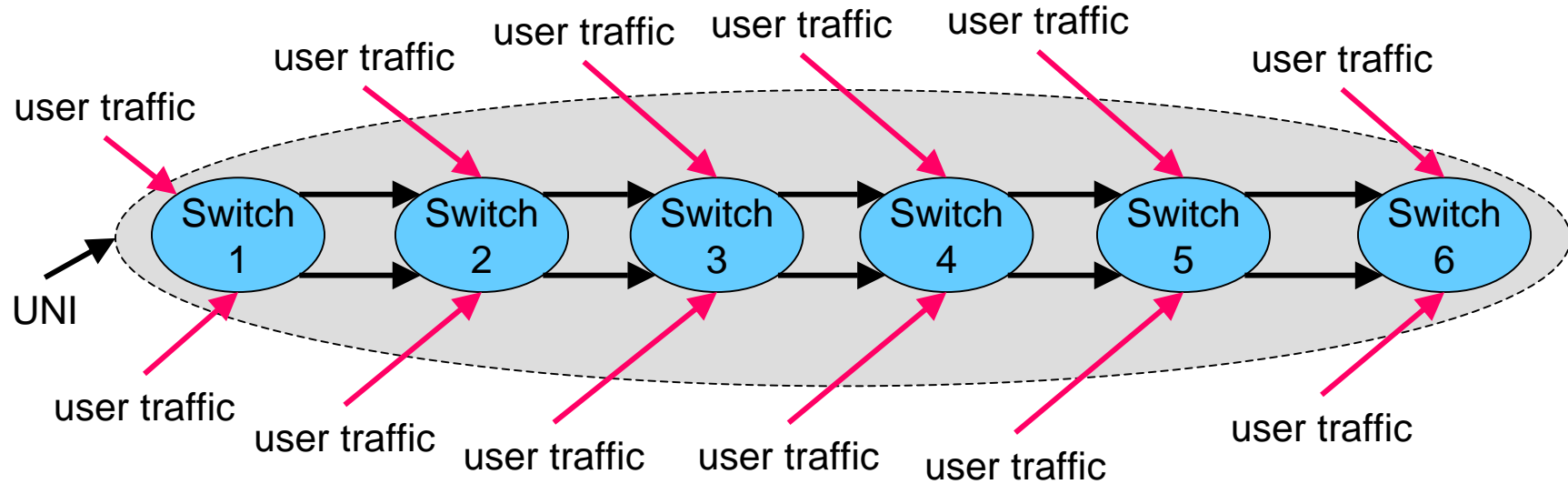
- Limitations: Lack of experimental validation; assumption that source can specify PCR/SCR/burst length; cell level models vs. source level traffic

QoS vs. Traffic Bandwidth Distribution

- Network performance:
 - Users are only interested in QoS for **their** traffic
 - Network operator is concerned about
 - link utilization - to recover investment
 - Every users' QoS - for customer retention
- Cell-level traffic controls do not guarantee QoS performance
 - Must look at call-level traffic parameters
 - Focus on traffic bandwidth requirement



Network Topology and User Traffic

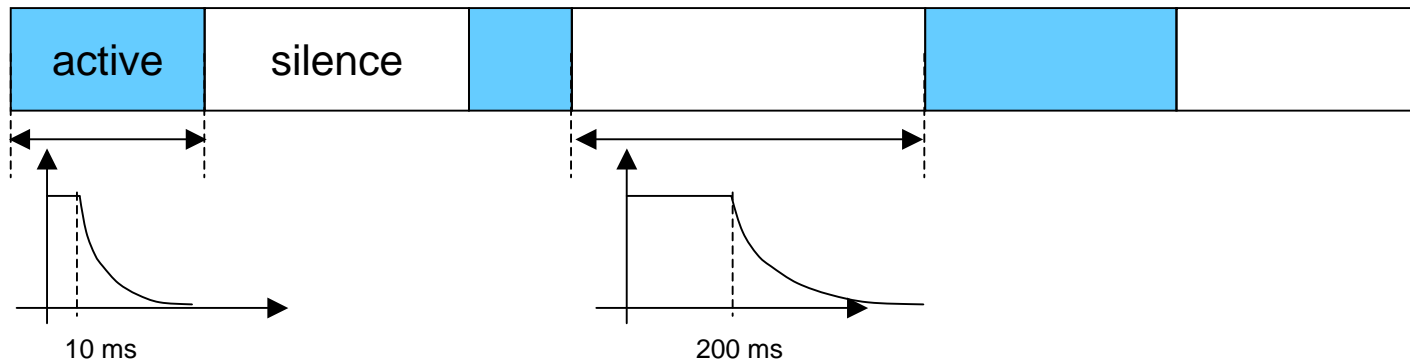


- User Traffic Model:

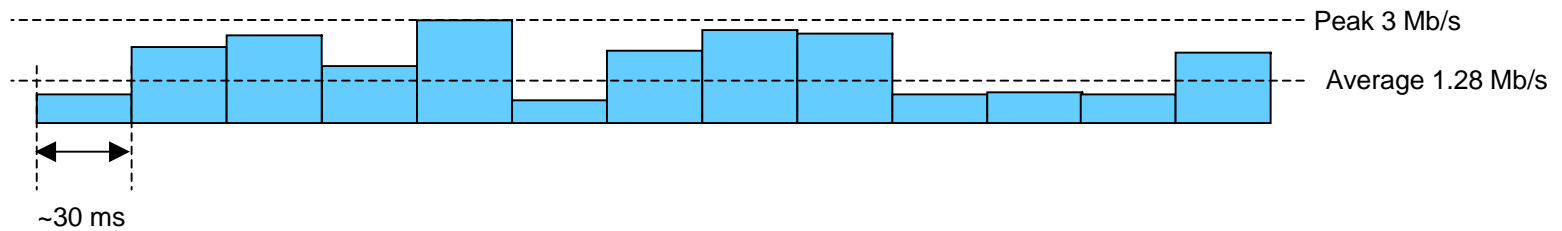
- voice (64 kb/s), video (95 kb/s - 55 Mb/s), and data (50 kb/s - 30 Mb/s)
- 25 - 2040 traffic sources
- 80% average link utilization
- traffic patterns from constant bit rate to highly bursty
 - on/off burstiness
 - variable rate burstiness
- various frame patterns:
 - variable bit rate with random frame duration
 - variable bit rate with fixed frame duration
 - fixed bit rate with constant inter frame gaps

Traffic Source Models

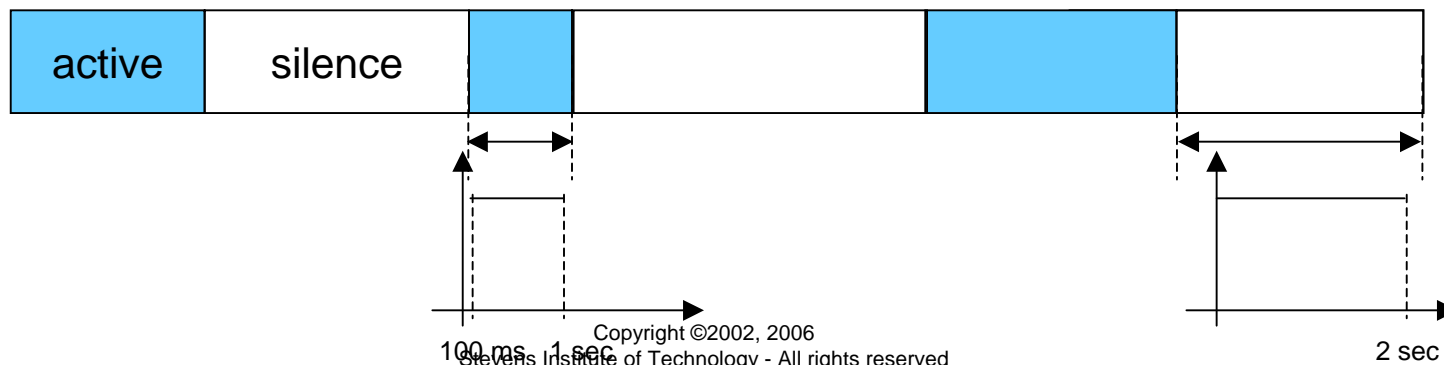
- Speech - peak bit rate 64 kb/s, average bit rate 25 kb/s



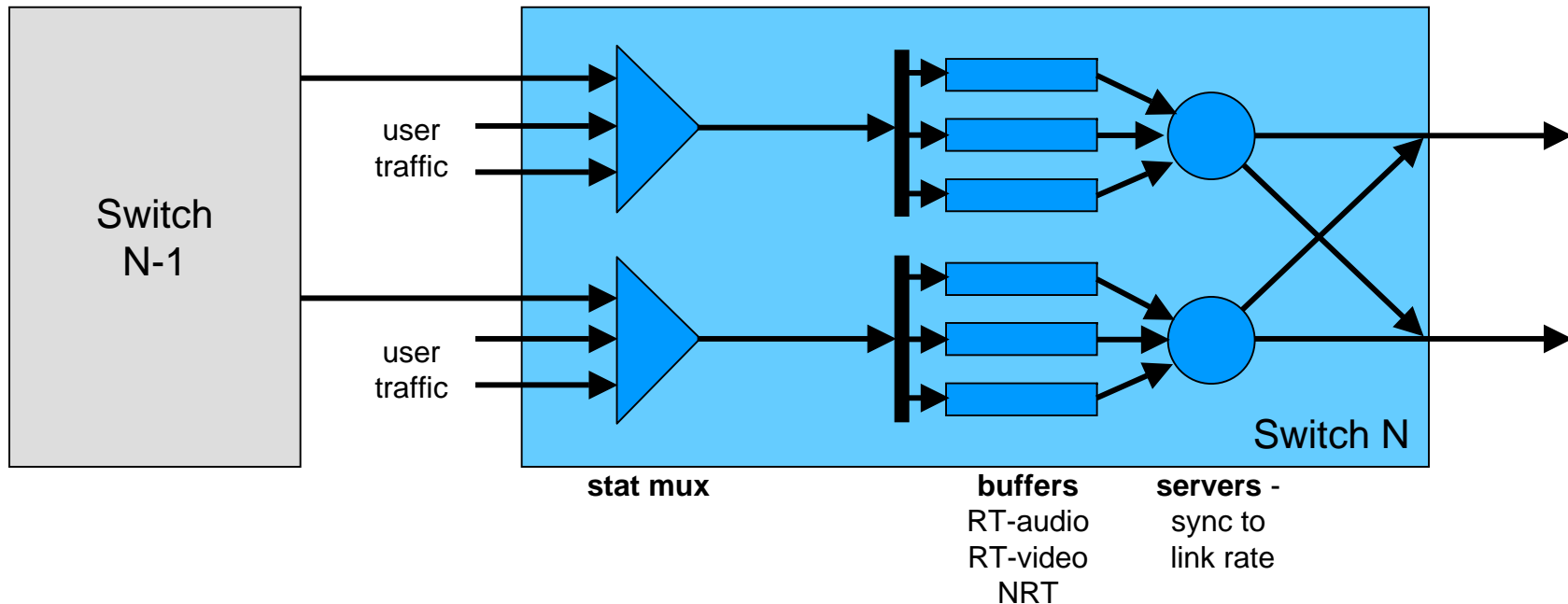
- H.261 video



- Data - peak bit rate 2 Mb/s, average bit rate 1 Mb/s



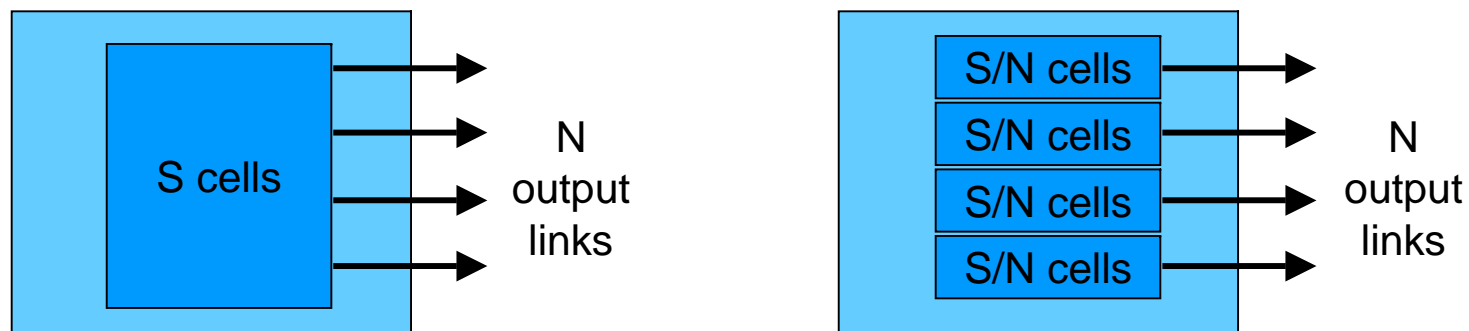
ATM Switch Model



- Arriving cells are assigned to a buffer, depending on their traffic type on space-available basis
- Server
 - gives first priority to real-time traffic over non-real-time,
 - attempts to balance delay between the real-time buffers,
 - periodically services non-real-time traffic

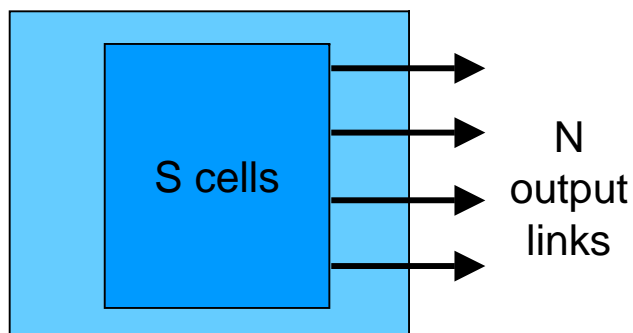
Attacking the ATM Output Buffer

- Assume:
 - To accept a call, adequate resources (e.g., bandwidth) must be available at each link of call
 - User traffic must comply with negotiated parameters at the UNI
- Examine how QoS guarantees can be broken, including severe cell loss
- Underlying issue is due to:
 - Bursty traffic
 - Inability to estimate network traffic
 - Means to determine buffer sizes from traffic
 - Playing on weakness of buffer architecture
- Buffer architectures - fixed at time of switch fabric design:

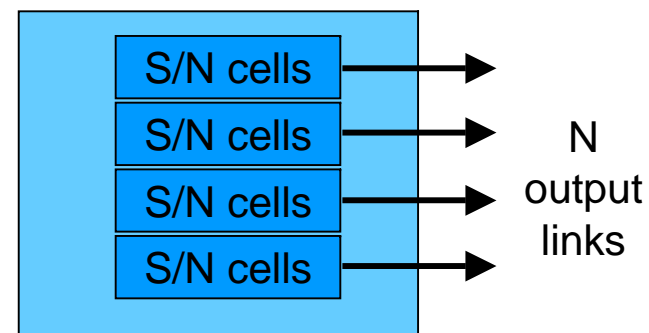


Attacking the ATM Output Buffer

- Assume:
 - To accept a call, adequate resources (e.g., bandwidth) must be available at each link of call
 - User traffic must comply with negotiated parameters at the UNI
- Examine how QoS guarantees can be broken, including severe cell loss
- Underlying issue is due to:
 - Bursty traffic
 - Inability to estimate network traffic
 - Means to determine buffer sizes from traffic
 - Playing on weakness of buffer architecture
- Buffer architectures - fixed at time of switch fabric design:

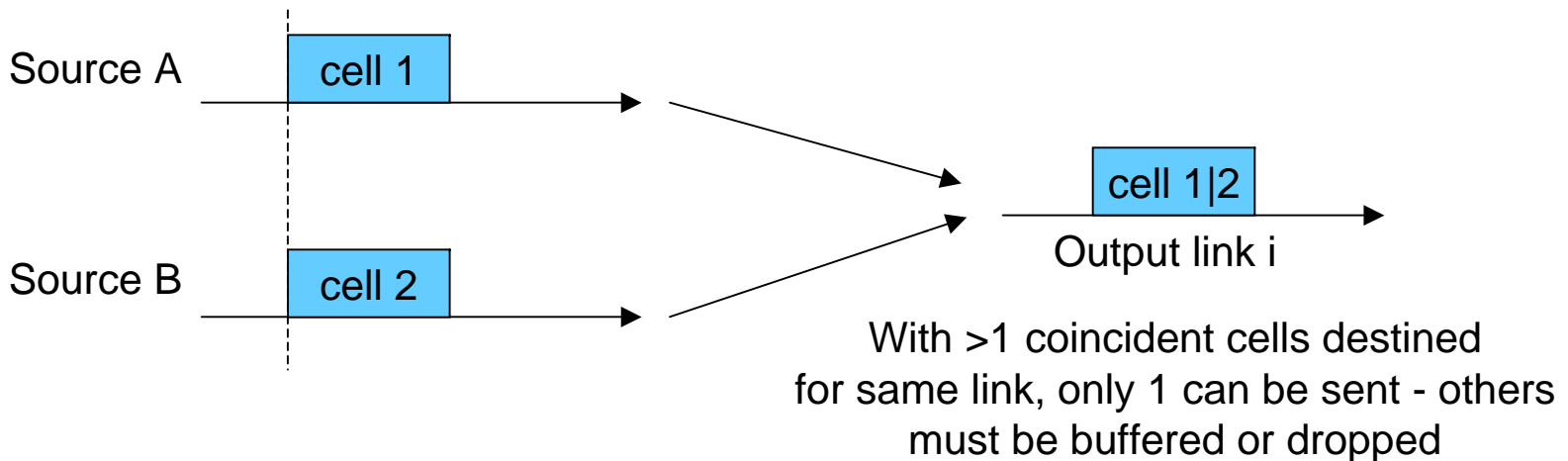


**All links are susceptible
to abuse of one**



**Susceptible to the
server-splitting problem**

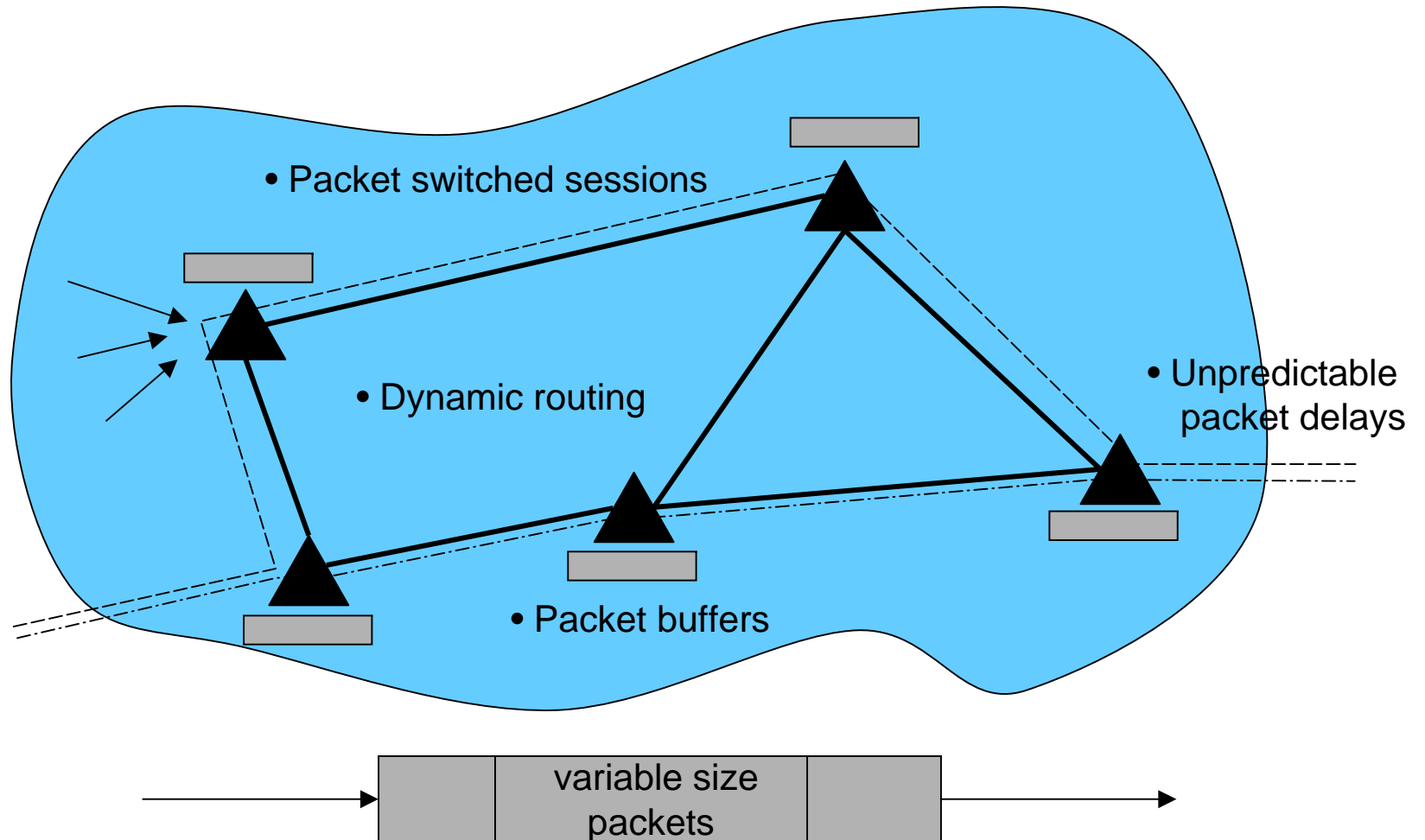
The Buffering Requirement



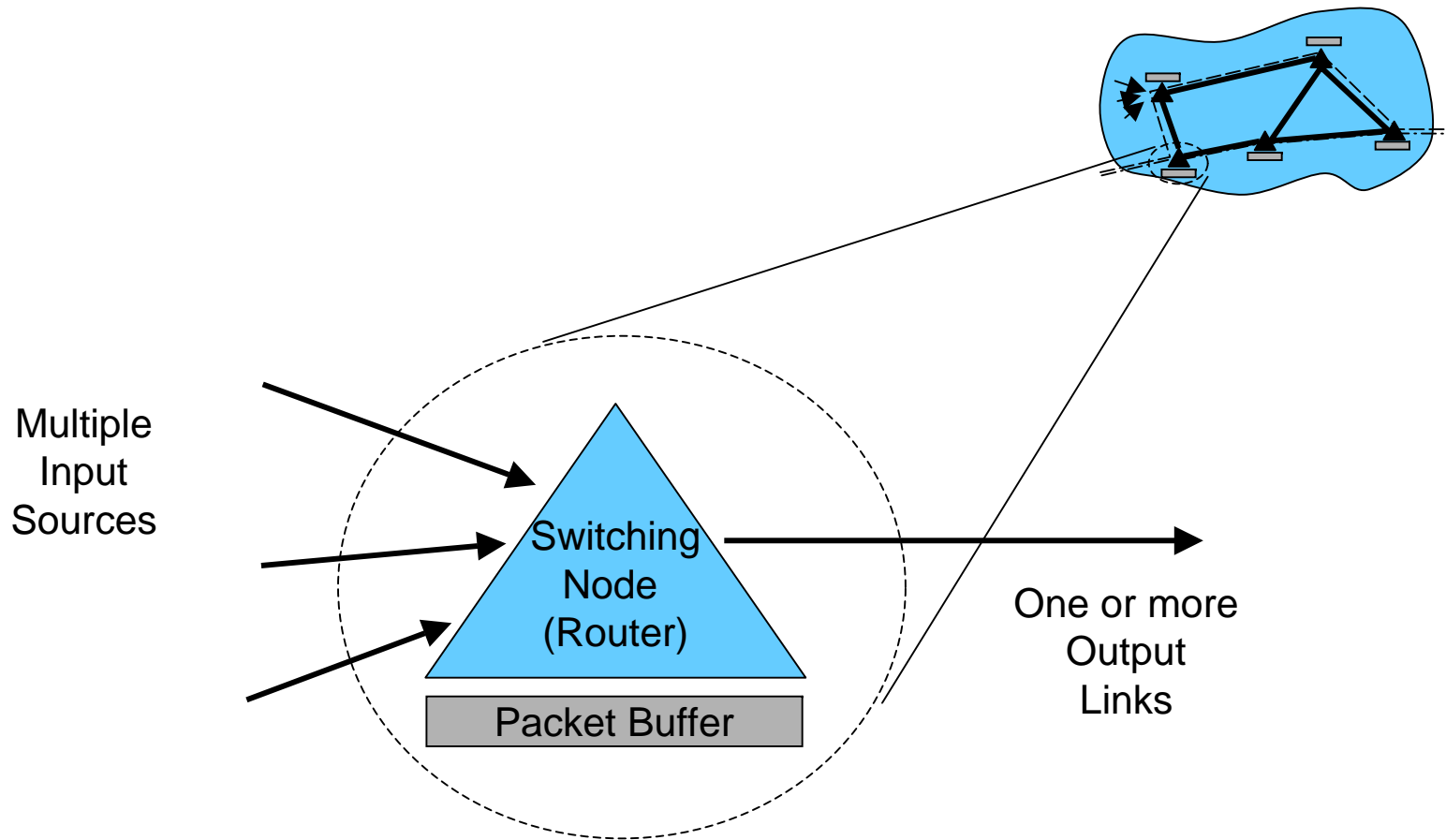
- Example:

- 8x8 switch with 8000 cell buffer, average burst length = 5, 10% cell loss probability:
 - Single shared buffer sustains 88% of the load
 - Separate buffers sustain 45% of load
- When burst length (relative to buffer cells per output link) increases, shared buffer may be 80-90% occupied by one output link's traffic - this causes congestion for all links
- But separate buffers would encounter unacceptable cell loss rates on individual links much earlier

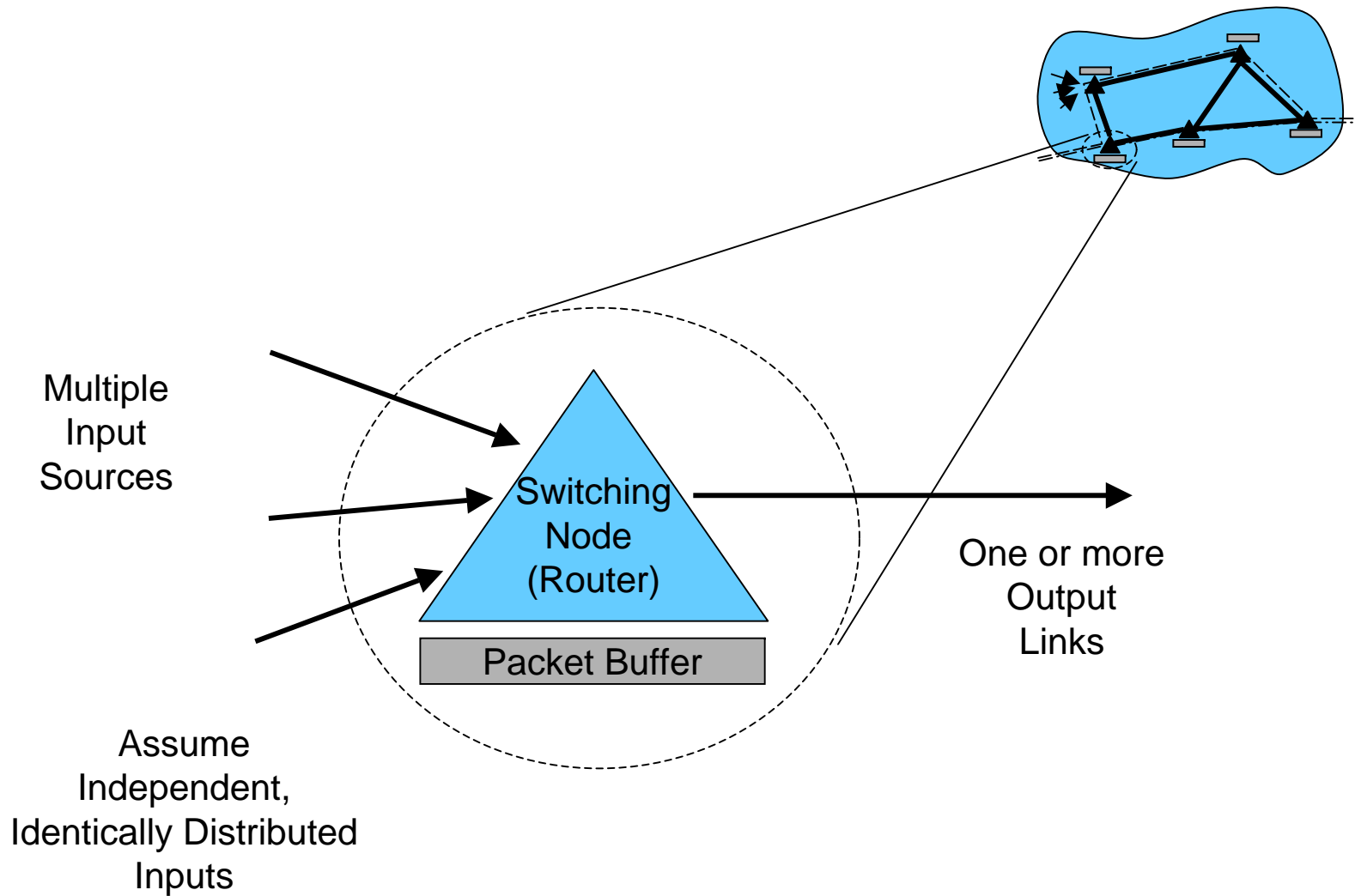
Characteristics of IP networks



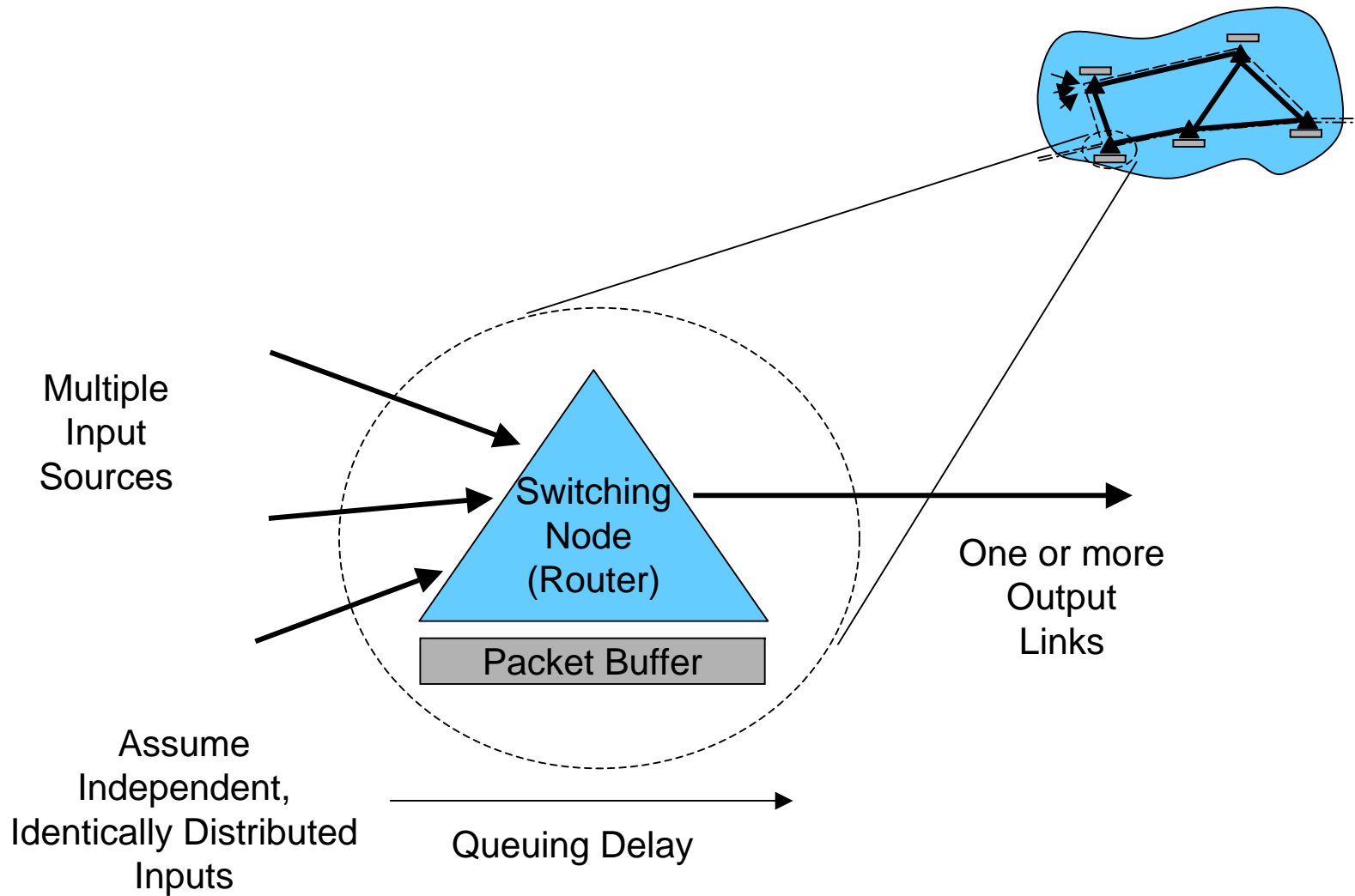
Characteristics of IP networks



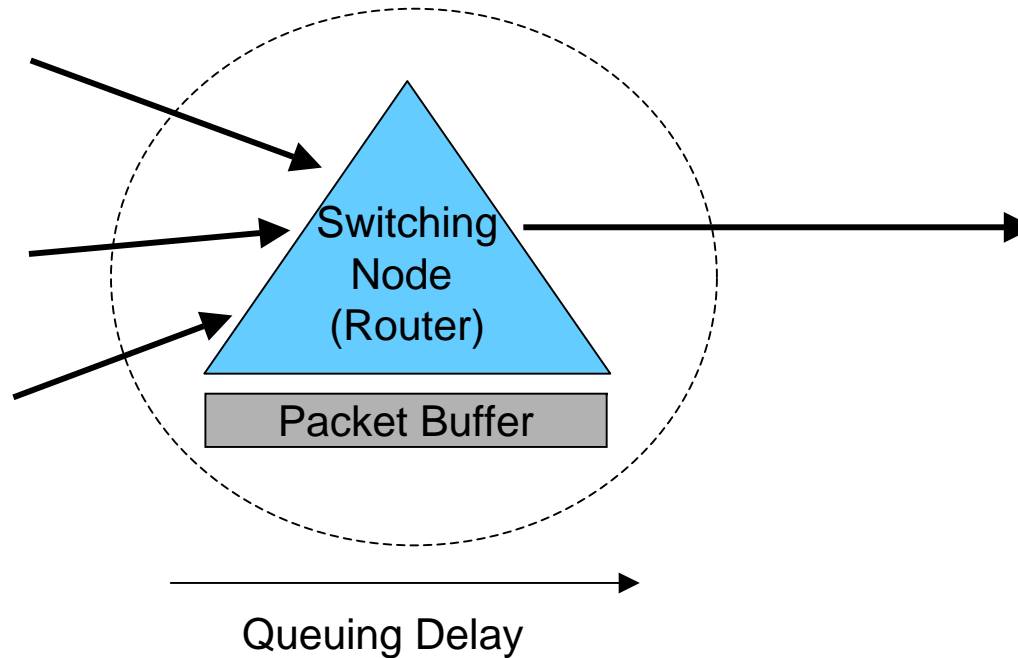
Characteristics of IP networks



Characteristics of IP networks



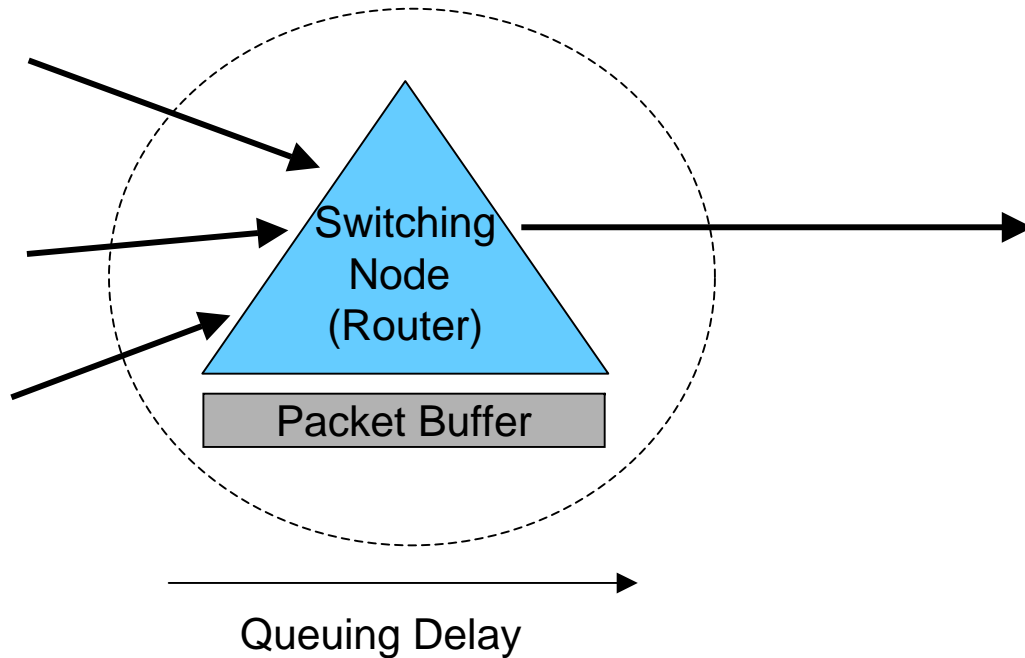
Basic Queuing Theory



Assume:

- Exponentially distributed inter-arrival times with average λ
- Exponentially distributed packet lengths (service time) with average μ
- Single output link with capacity 1 (normalized)

Basic Queuing Theory

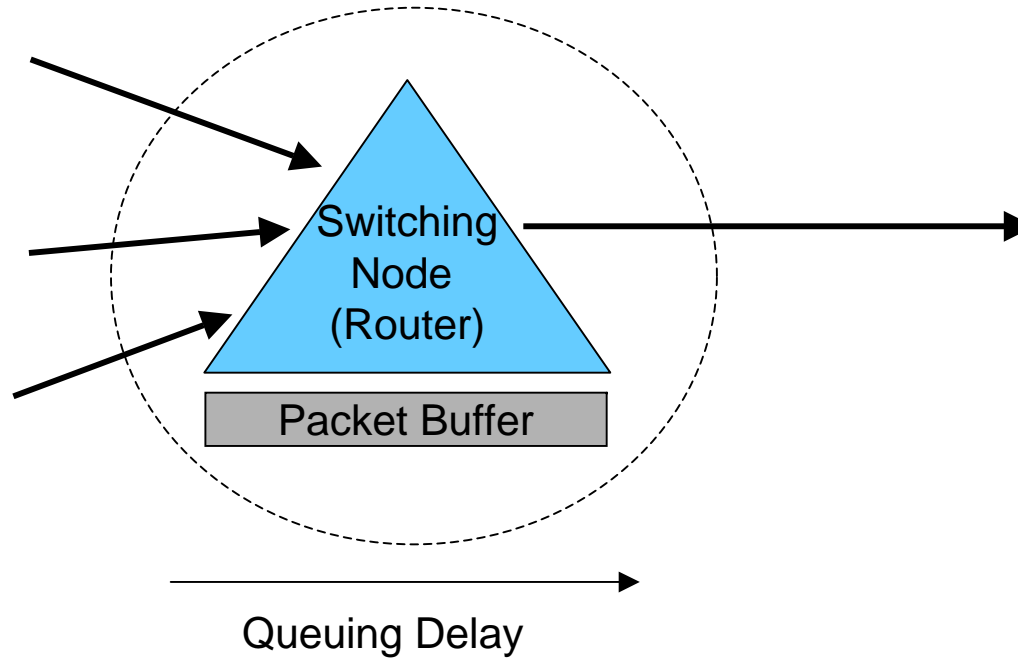


Assume:

- Exponentially distributed inter-arrival times with average λ
- Exponentially distributed packet lengths (service time) with average μ
- Single output link with capacity 1 (normalized)

This is an M/M/1 queue

Performance of M/M/1 Queue



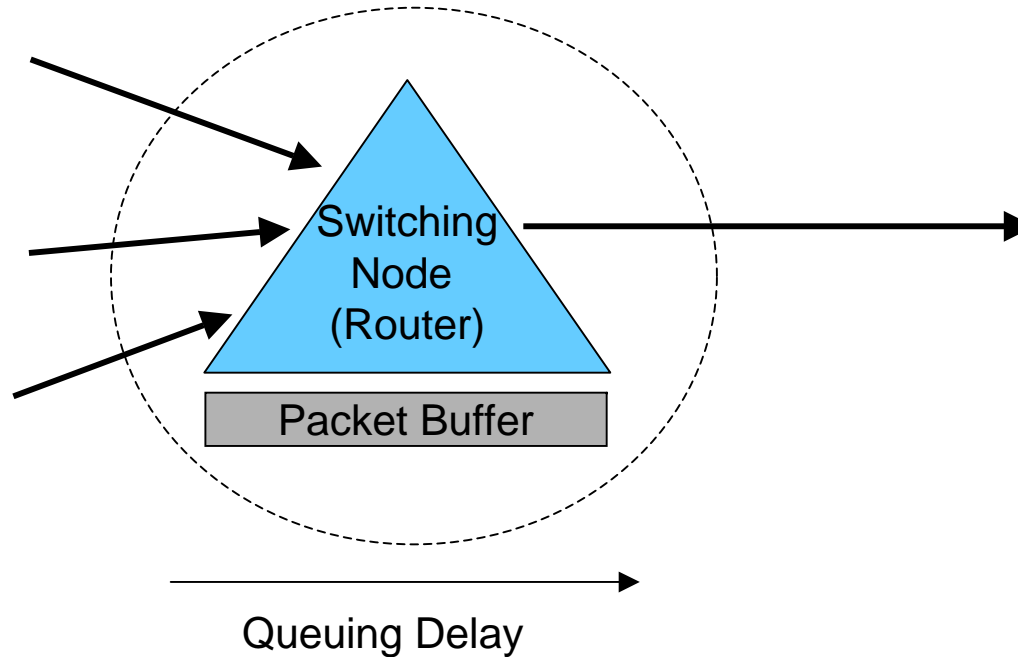
Parameters:

Average utilization $\rho = \frac{\lambda}{\mu}$

Average waiting time $W = \frac{\rho}{\mu(1-\rho)}$

Average queue length $L_q = \frac{\rho^2}{(1-\rho)}$

Performance of M/M/1 Queue



Average utilization $\rho = \frac{\lambda}{\mu}$

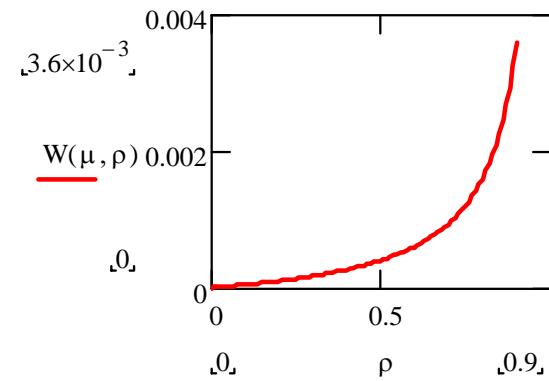
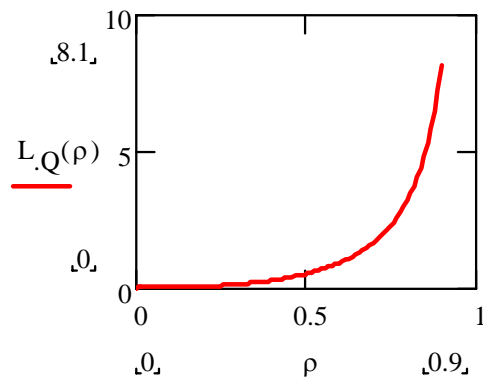
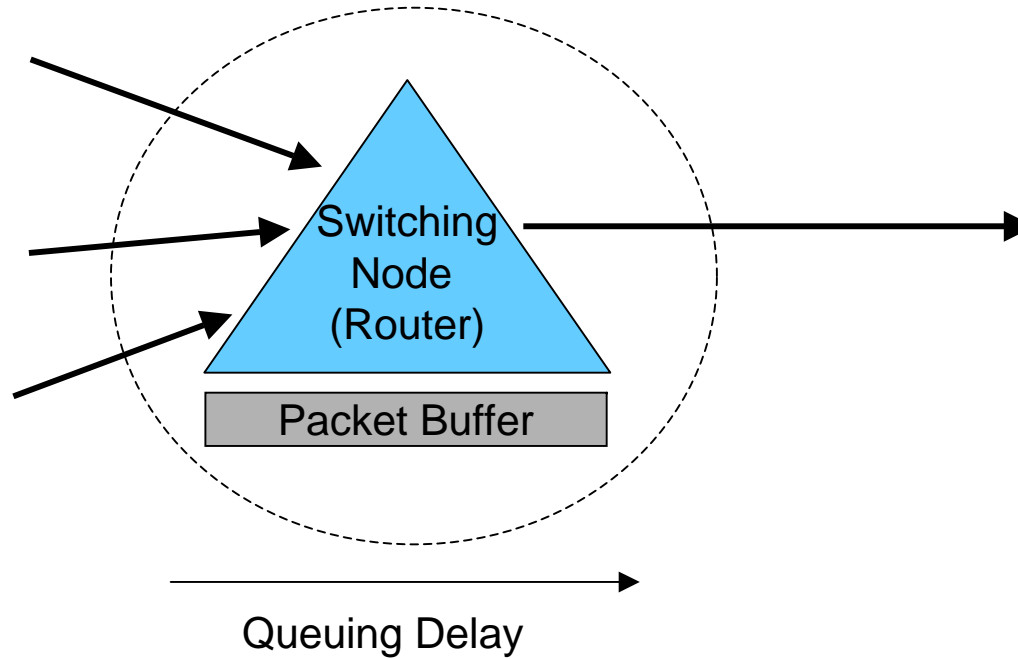
Average waiting time $W = \frac{\rho}{\mu(1-\rho)}$

Average queue length $L_Q = \frac{\rho^2}{(1-\rho)}$

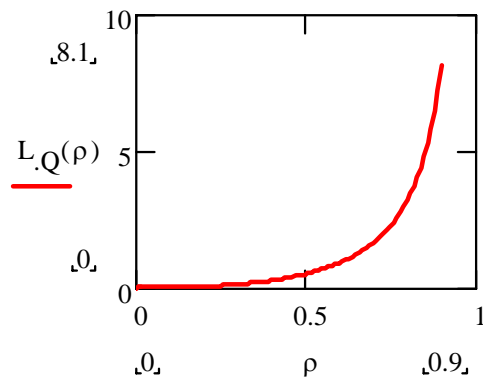
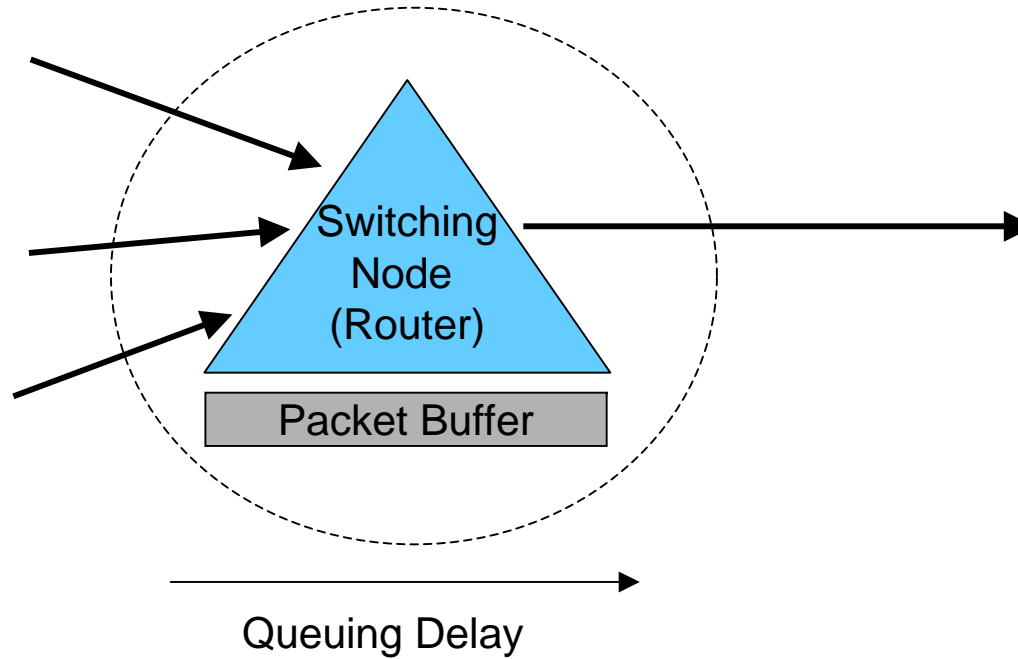
Illustrative values:

- Packet arrival rate: 1000 packets/second
- Average (output) service time for 500 byte packets, 10 Mb/s:
 $4000 \text{ bits} / 10,000,000 \text{ bits/sec} = .0004 \text{ sec} = 2500/\text{sec}$
- $\rho = 1000/2500 = .4$, $W = 266 \mu\text{sec}$, $L_Q = .266$

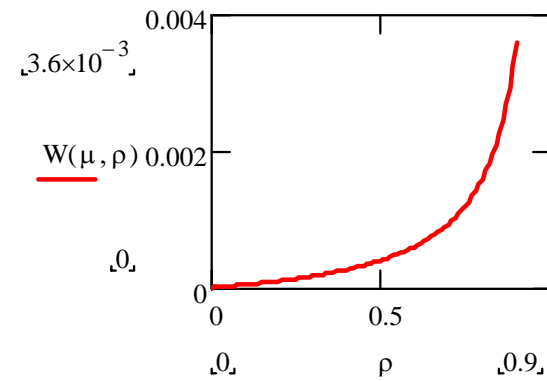
Packet Switches in Overload



Packet Switches in Overload

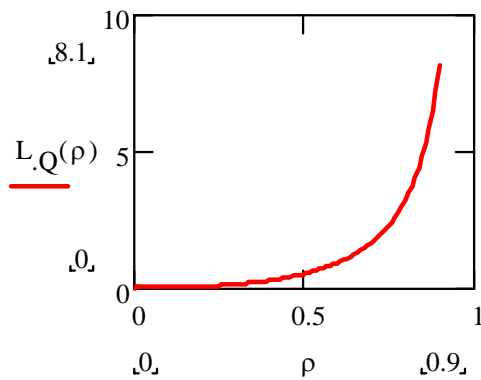
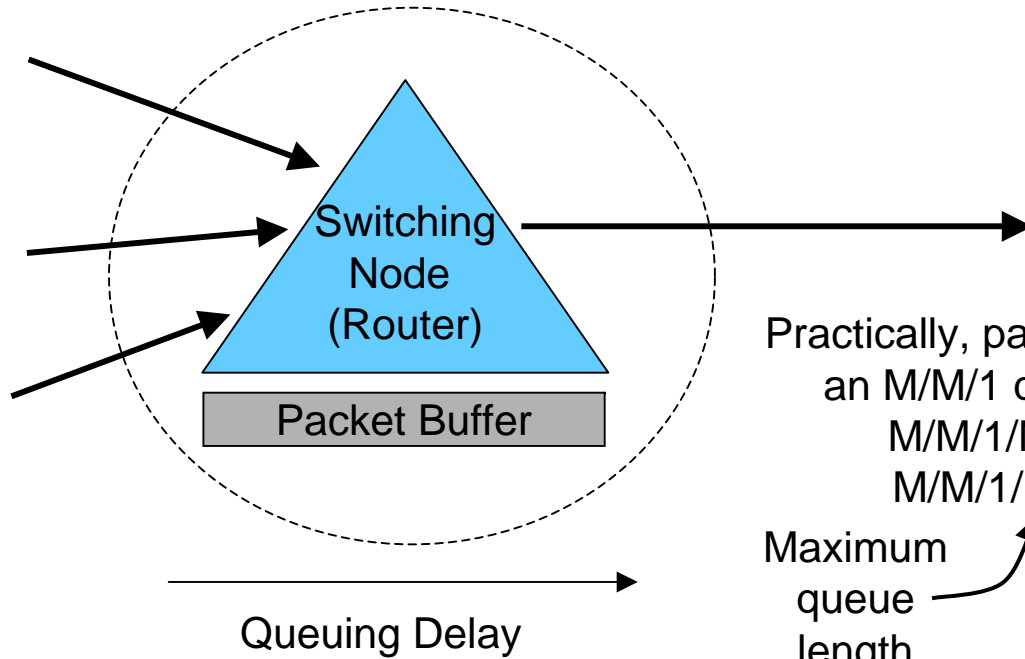


Packet
buffer
length
grows

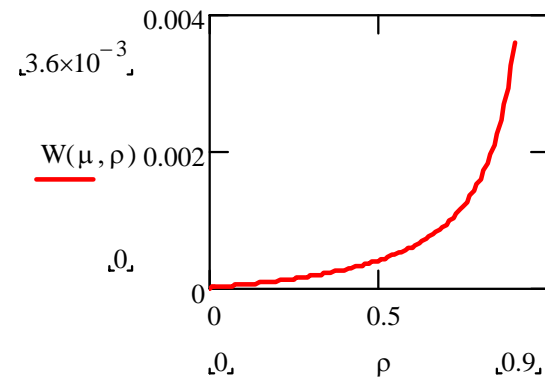


Delays
increase

Packet Switches in Overload

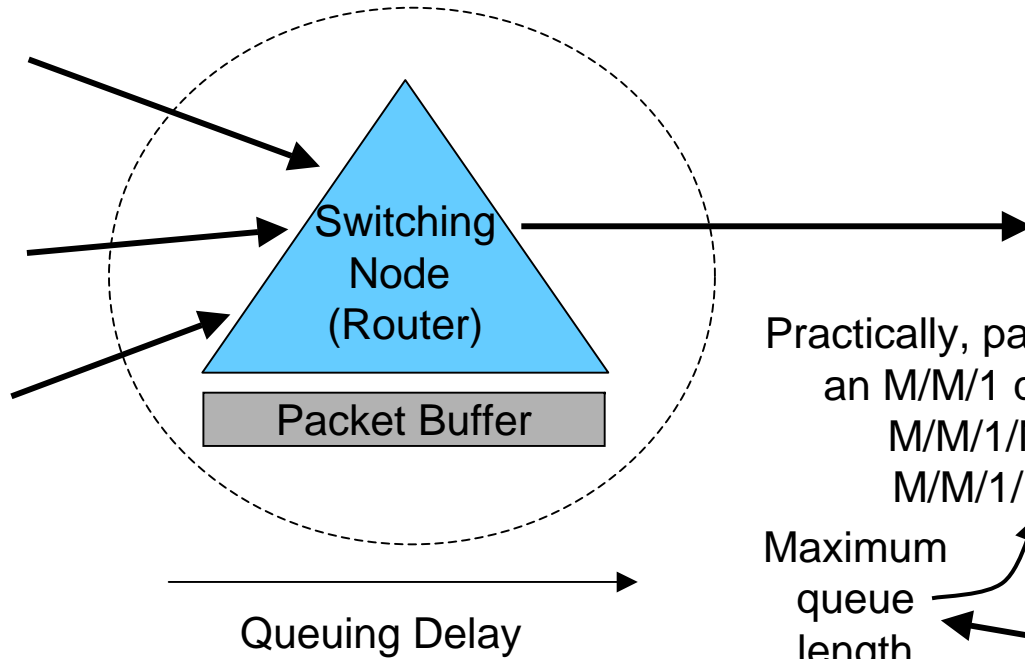


Packet buffer length grows



Delays increase

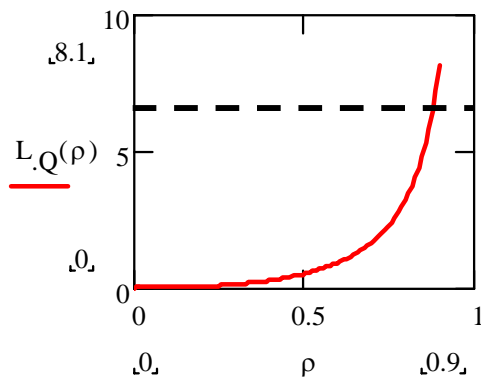
Packet Switches in Overload



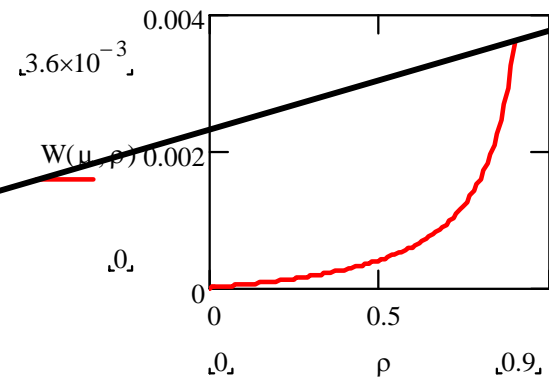
Practically, packet switch is not an M/M/1 queue. It is an M/M/1/N queue or M/M/1/N/K queue

Maximum queue length Population size

Packets are lost



Packet buffers overflow



Delays increase