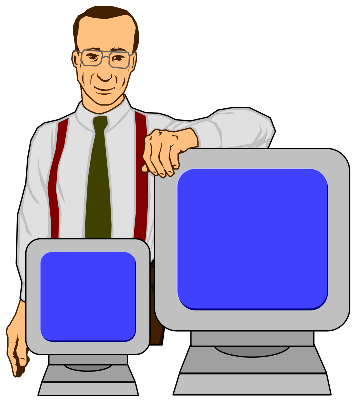
ATM Traffic Management



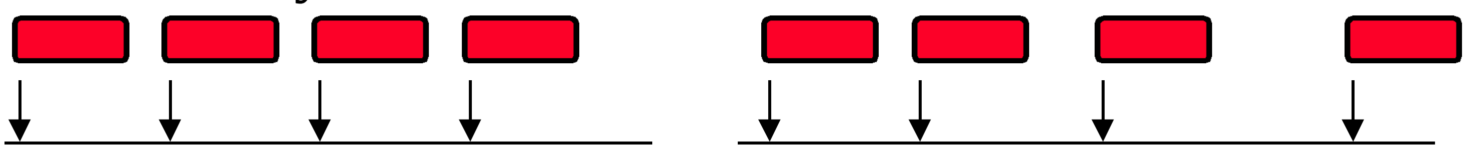
One Megabit memory, One Megabyte disk, One Mbps link, One MIP processor, one dollar each.....

Raj Jain

Raj Jain is now at
Washington University in Saint Louis
Jain@cse.wustl.edu
http://www.cse.wustl.edu/~jain/

ATM Networks: Overview

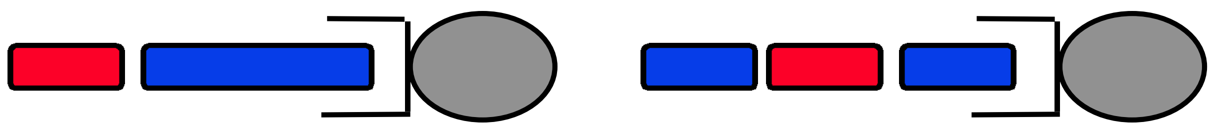
- STM = Synchronous Transfer Mode,
  ATM = Asynchronous Transfer Mode



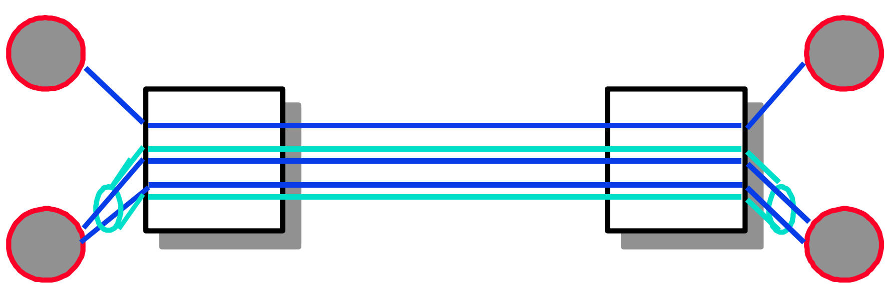
Allows any-speed and even variable rate connection

Broadband = Rate greater than primary rate (1.5 Mbps)

- ATM = Short fixed size 53-byte cells



- Connection oriented ⇒ Virtual Channels (VC)



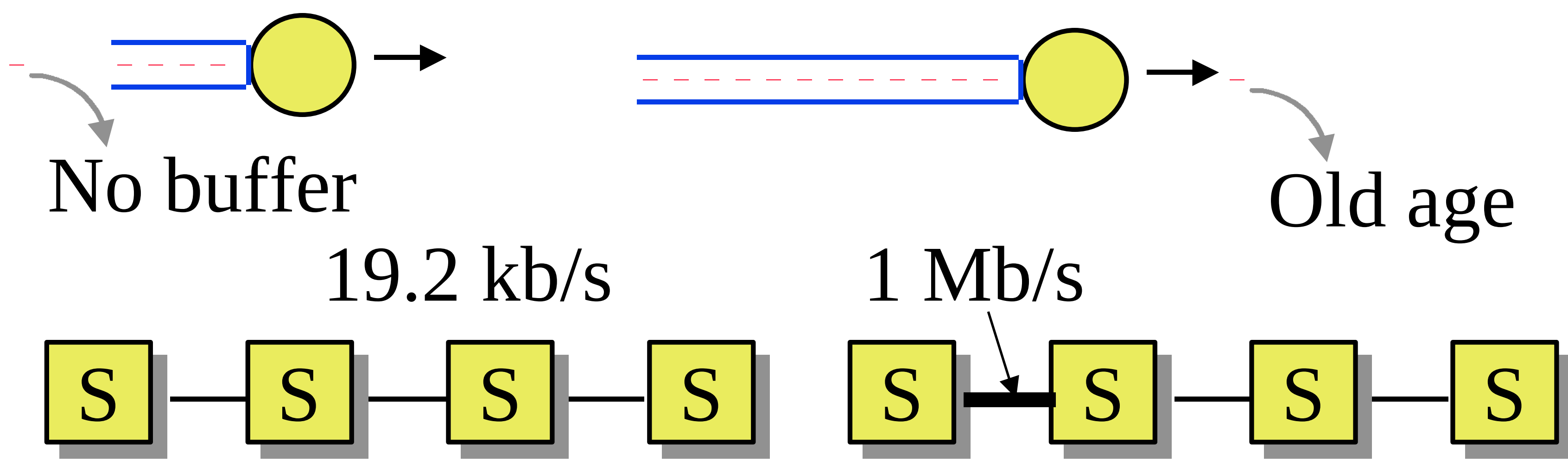
- Why worry about congestion?
- Congestion schemes for ATM
- Rate vs Credit: Key issues
- Explicit Rate-based Control
- ABR Traffic Management

Why Worry About Congestion?

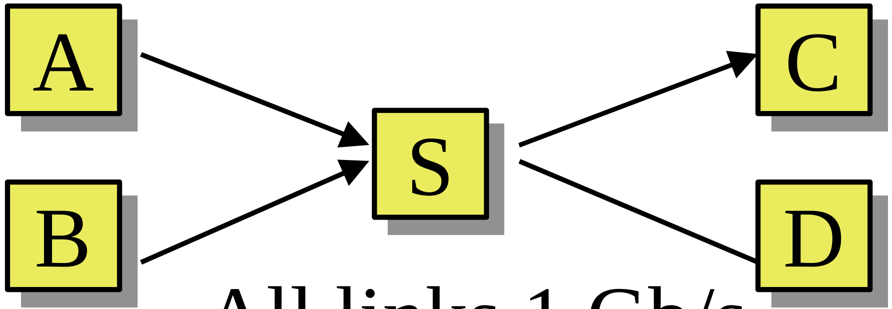
Q: Will the congestion problem be solved when:

- Memory becomes cheap (infinite memory)?
- Links become cheap (very high speed links)?
- Processors become cheap?

A: None of the above.



File transfer time = 5 mins   Time = 7 hours



All links 1 Gb/s

Conclusions:

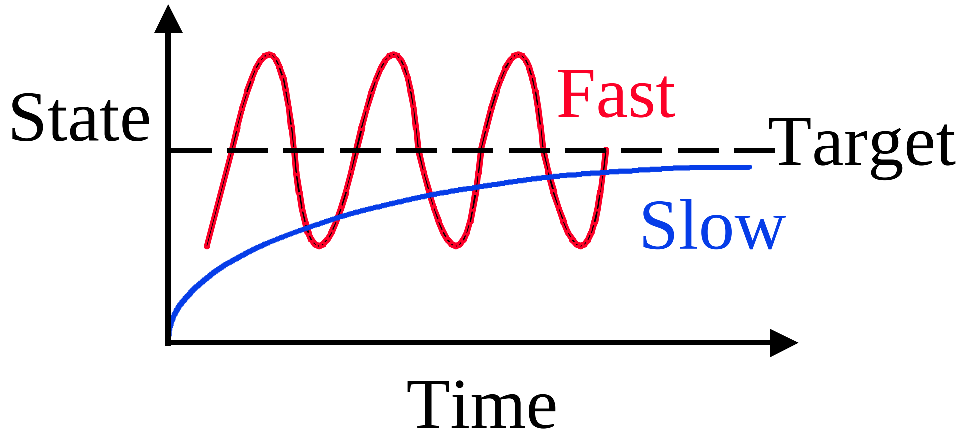
- Congestion is a dynamic problem. Static solutions are not sufficient
- Bandwidth explosion
  $\Rightarrow$ More unbalanced networks
- Buffer shortage is a symptom not the cause.

Economic Reasons

- Network is a shared resource
  Because it is expensive and needed occasionally
  (Like airplanes, emergency rooms)
- Most costs are fixed.
  Cost for fiber, switches, laying fiber and maintaining them does not depend upon usage
  $\Rightarrow$ Underutilization is expensive
- But overutilization leads to user dissatisfaction.
- Need a way to keep the network maximally utilized

One Scheme or Many?

- Fundamental principle of control theory:



- Control faster than feedback ⇒ Instability
  Control slower than feedback ⇒ non-responsiveness
  Ideal: Control rate ≈ Feedback rate
- Lesson: No scheme can cure a congestion lasting less than its feedback delay.

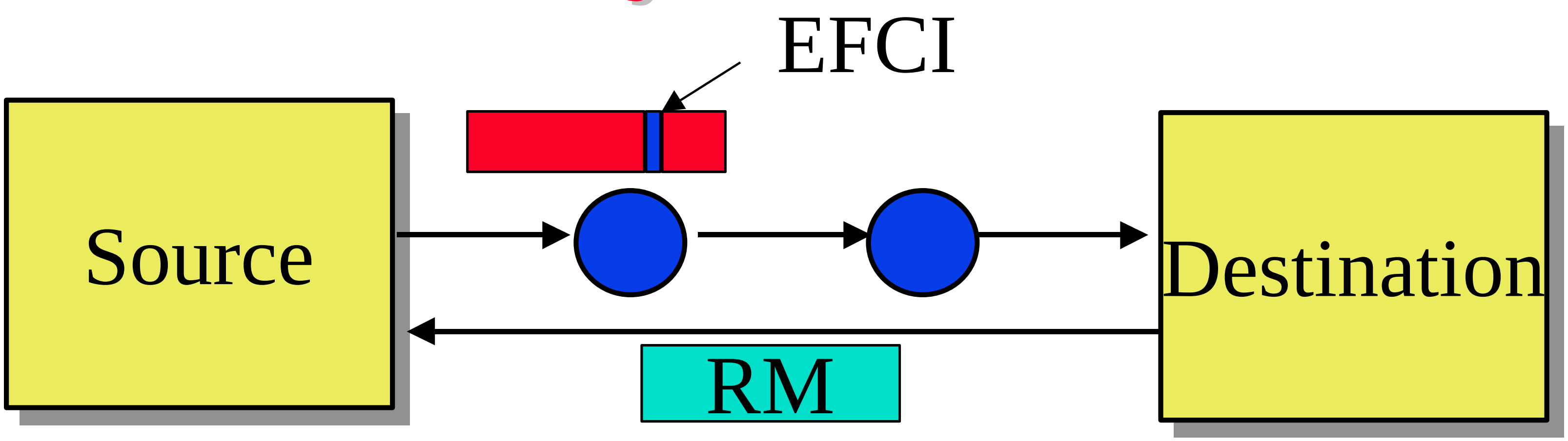
Classes of Service

- CBR (Constant bit rate): User dec8 Tcres required rate. Throughput, delay and delay variation guaranteed.
- VBR (Variable bit rate): User dec8ares average and max rate.
  - rt-VBR (Real-time variable bit rate): Conferencing. Max delay and delay variation guaranteed.
  - nrt-VBR (non-real time variable bit rate): Stored video.
- ABR (Available bit rate): Follows feedback instructions. Network gives maximum throughput with minimum loss.
- UBR (Unspecified bit rate): User sends whenever it wants. No feedback mechanism. No guarantee. Cells may be dropped during congestion.

Traffic Management Functions

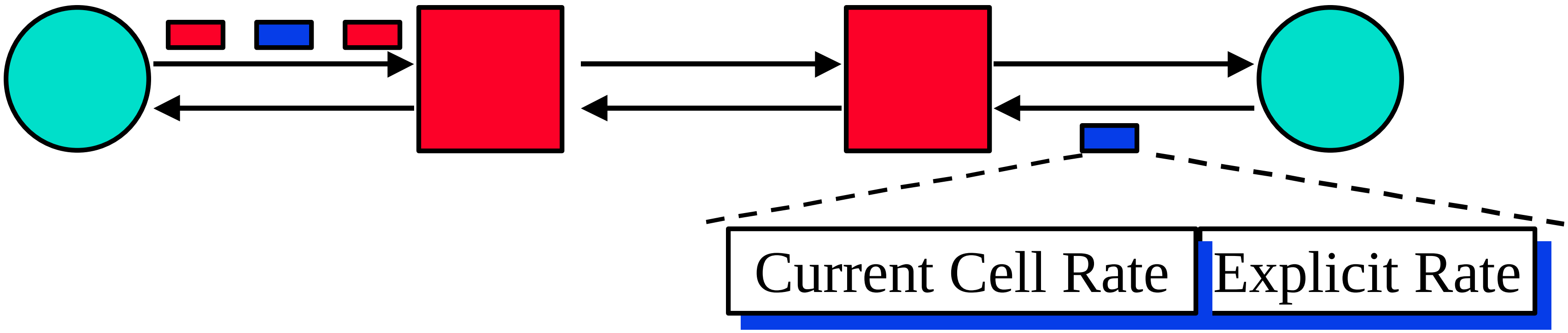
- Connection Admission Control (CAC):
  Can requested bandwidth and quality of service be supported?
- Traffic Shaping: Limit burst length. Space-out cells.
- Usage Parameter Control (UPC):
  Monitor and control traffic at the network entrance.
- Network Resource Management:
  Scheduling, Queueing, virtual path resource reservation
- Selective cell discard:
  Cell Loss Priority (CLP) = 1 cells may be dropped
  Cells of non-complient connections may be dropped
- Frame Discarding
- Feedback Controls: Network tells the source to increase or decrease its load.

Initial Binary Rate-based Scheme



- *We invented DECbit scheme (1986). Implemented in many standards since 1986.*
- *Forward explicit congestion notification (FECN) in Frame relay*
- Explicit forward congestion indicator (EFCI) set to 0 at source. Congested switches set EFCI to 1
- Every *n*th cell, destination sends an resource management (RM) cell to the source indicating increase amount or decrease factor

The Explicit Rate Scheme

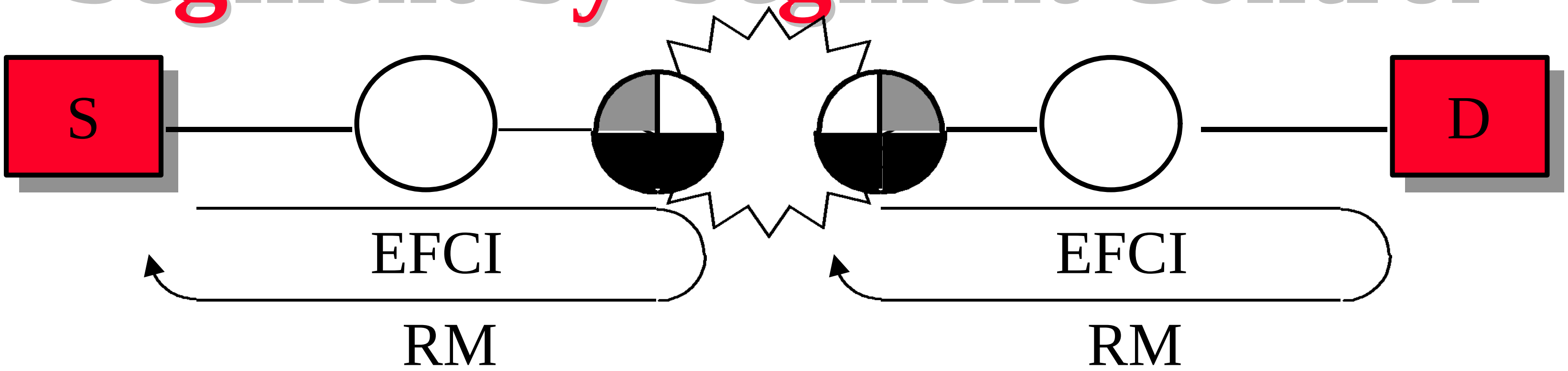


- Sources send one RM cell every n cells
- The RM cells contain "Explicit rate"
- Destination returns the RM cell to the source
- The switches adjust the rate down
- Source adjusts to the specified rate

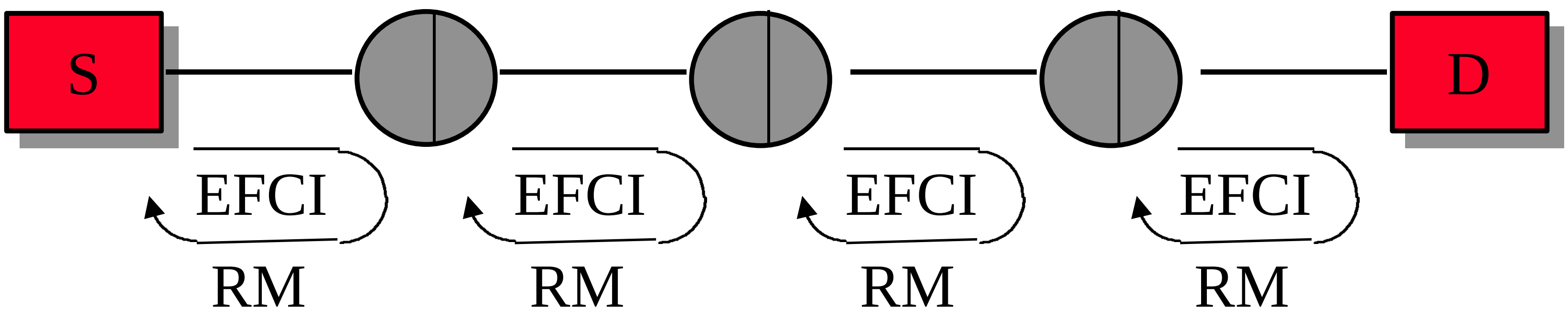
ERICA Switch Algorithm

- Each manufacturer will have its own explicit rate switch algorithm
- Explicit Rate Indication for Congestion Avoidance (ERICA) is the most thoroughly analyzed algorithm among disclosed algorithms
- Shown to be efficient, fair, fast transient response, able to handle bursty TCP traffic
- ERICA+ allows low delay even at 100% utilization and provides stability in the presence of high frequency VBR background traffic
- Being implemented by several vendors

Segment-by-Segment Control

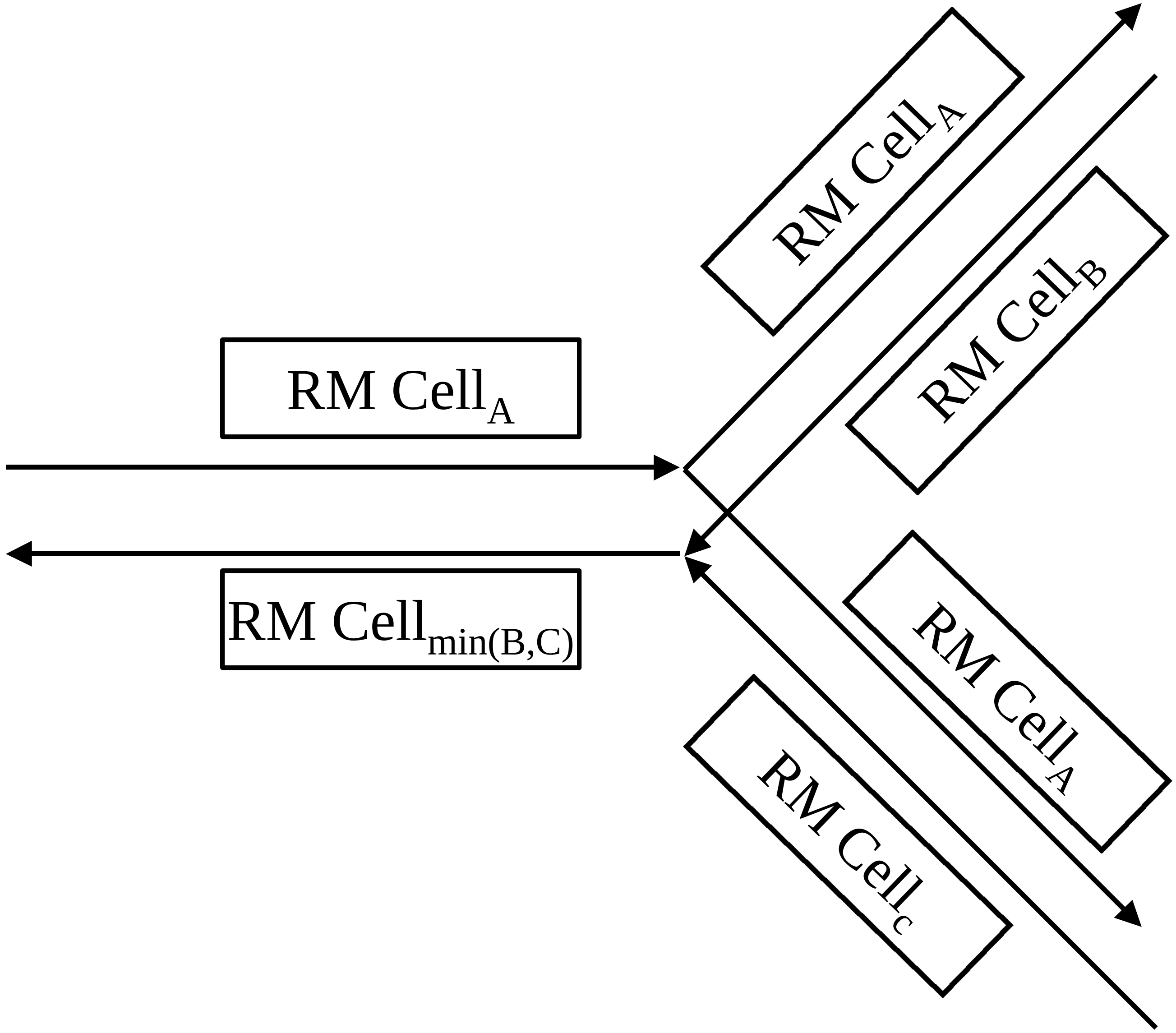


- Virtual source/virtual destinations follow all notification/control rules
- Can be hop-by-hop



- Virtual dest/sources maintain per-VC queues.

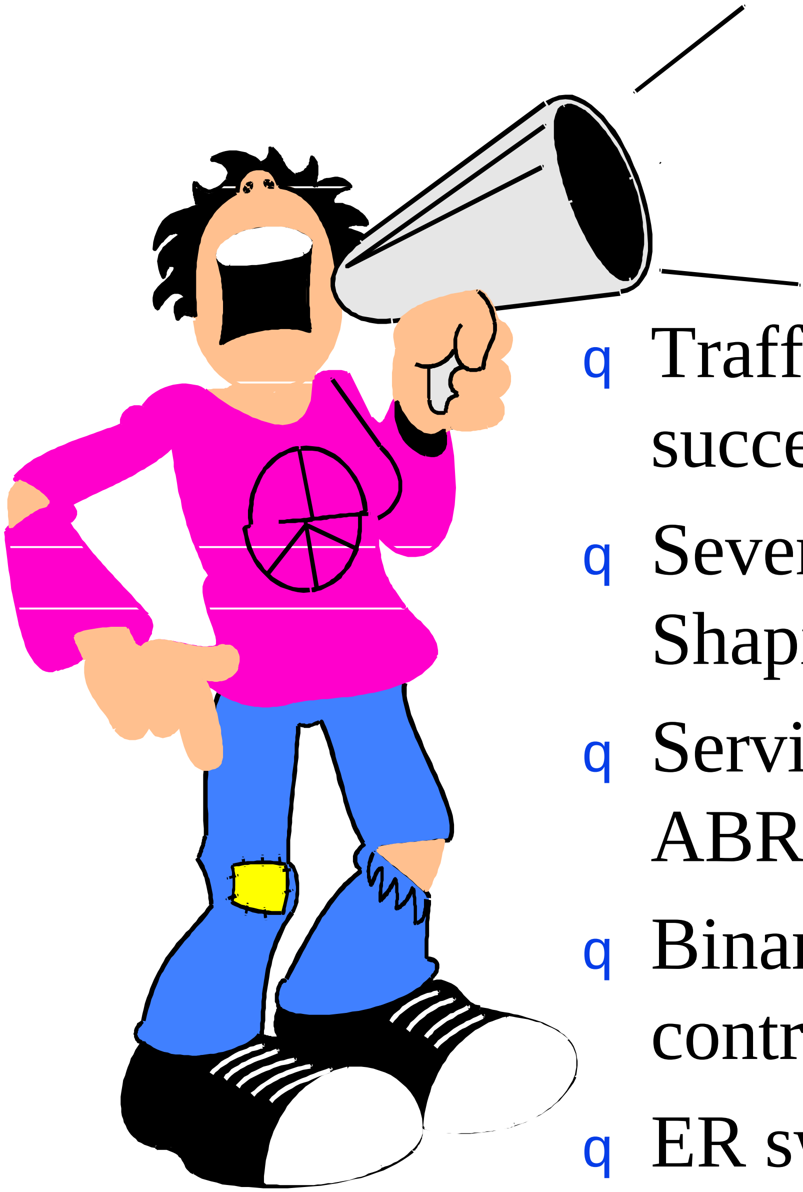
Multicast



Outstanding Issues

- Bursty sources: Client server, transactions, WWW
- Effect of parameters: Optimal parameter values
- Priority service for RM cells
- Multicast
- Connection admission control (CAC)
- TCP/IP over UBR
- Non-conforming sources
- Optimal Source Strategy: Parameter + Out-of-rate cells
- Virtual Source/destination
- Implicit feedback schemes: Heterogeneous Networks

Congestion: Summary



- Traffic Management is key to success of ATM
- Several different methods: CAC, Shaping, UPC, Scheduling, ...
- Service categories:CBR, VBR, ABR, UBR
- Binary feedback too slow for rate control. Especially for satellites.
- ER switches provide much better performance than EFCI.

References

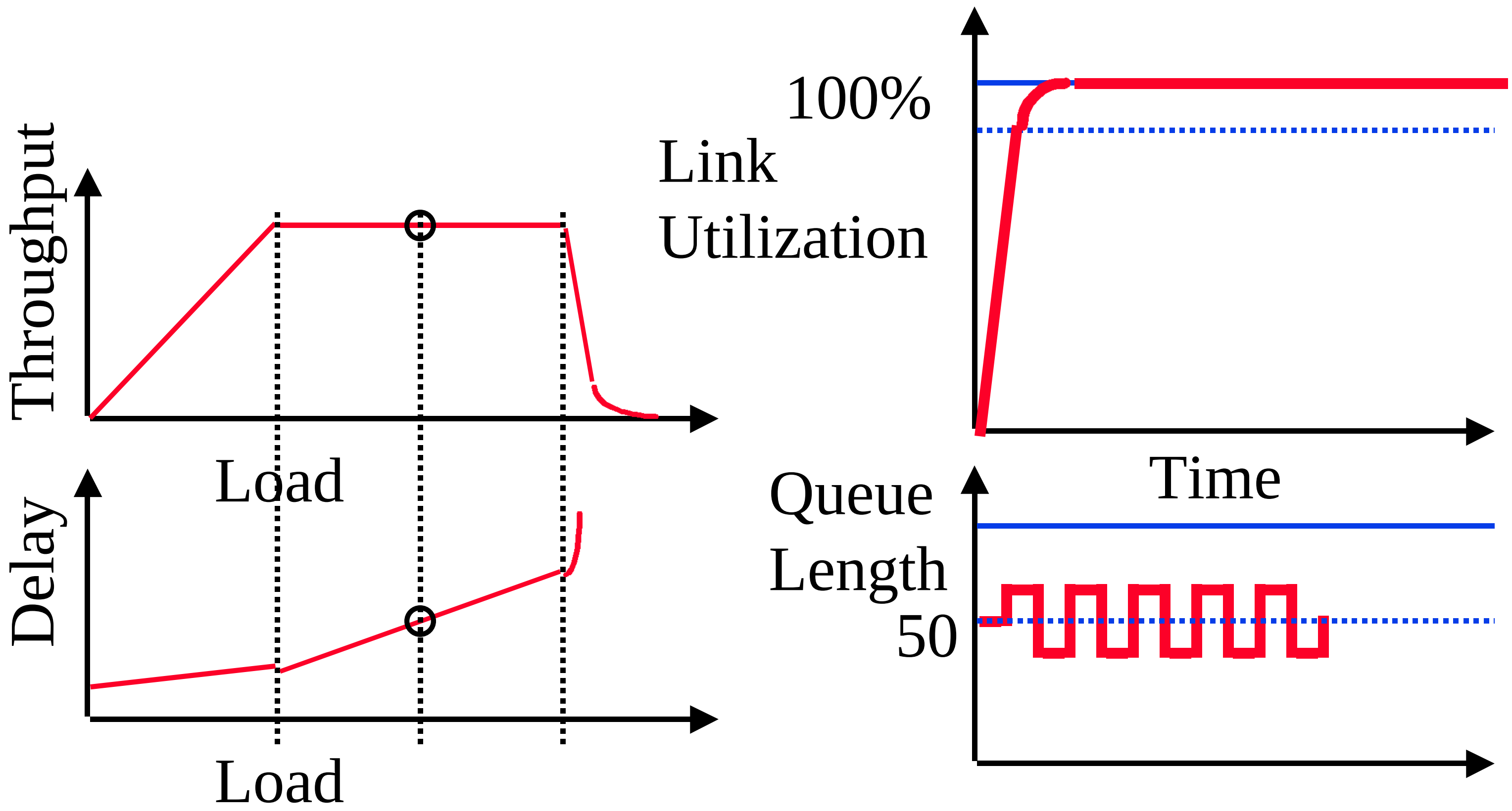
- R.Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and A Survey", Invited submission to Computer Networks and ISDN Systems, February 1995.
  Available at http://www.cis.ohio-state.edu/~jain/
- *User-Network Interface Specifications, V3.0*, Prentice-Hall, September 10, 1993., (515)-284-6751
  V3.1 released by ATM Forum, January 1995.
- K. Siu and R. Jain, "A Brief Overview of ATM: Protocol Layers, LAN Emulation, and Traffic Management," Computer Communications Review (ACM SIGCOMM), April 1995.
- "ATM Forum Traffic Management Specification, Version 4.0," ATM Forum/95-0013R13, April 1996.

- Craig Partridge, *Gigabit Networking*, Addison-Wesley, 1993.
- S. Akhtar, *Congestion Control in a Fast Packet Switching Network*, Master's Thesis, Washington University, 1987. [Originated the concept of leaky bucket.]
- Allyn Romanow, *Packet Discard Strategies for Controlling AM Congestion*, ATM Forum Traffic Management Subworking Group submission #94-0107, January 17-24, 1994.
- S. Kalyanaraman, R. Jain, S. Fahmy, R. Goyal, F. Lu and S. Srinidhi, ``Performance of TCP/IP over ABR," Submitted to Globecom'96.
- R. Jain, S. Kalyanaraman, R. Viswanathan, "The OSU Scheme for Congestion Avoidance in ATM networks Using Explicit Rate Indication," Proceedings WATM'95, Paris, December 1995, Also available an extended OSU technical report and an extended conference version.

- "Killer App? It's the high-speed backbone," An interview with Raj Jain. Government Computer News, October 16, 1995, pp. 24.
- R. Jain, "ABR Service on ATM Networks: What is it?" Network World, June 24, 1995.
- A. Charny, D. Clark, R. Jain, "Congestion Control with Explicit Rate Indication," Proc. ICC'95, June 1995, 10 pp.

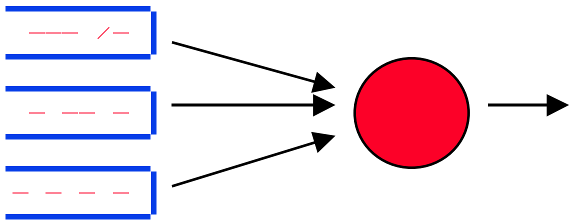
ERICA+: Full Utilization

- Allows operation at any point between the knee and the cliff
- The queue time can be set to any desired value.
- Allows utilization to be 100%



Fair Queueing

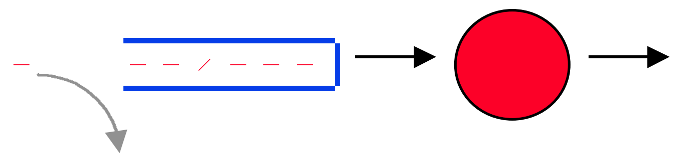
- Goal: Isolation of flows.
- Round Robin:



- Problem: Flows with large packets get more bandwidth.
- Bit-by-bit Round Robin
- Fair Queueing: Compute end-times of packets under bit-by-bit round robin.
- Weighted Fair Queueing: Different flows are allowed different bandwidth.

Selective Cell Discarding

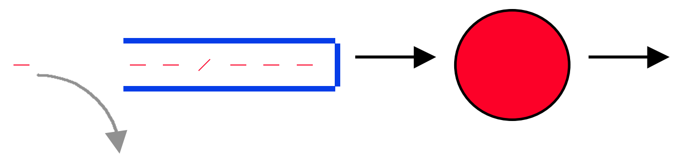
- **Goal**: Discard cells of VCs not following contract
- **Normal Discard**: CLP=1 cells are discarded if buffer is full.



- **Selective Cell Discard**: May discard CLP =1 or CLP =0 cells of non-complient VCs.

Early Packet Discard

- **Goal**: To minimize packet loss rate
- **Normal Discard**: Cells are discarded if buffer is full.



  Problem: Unfair to late-comers.

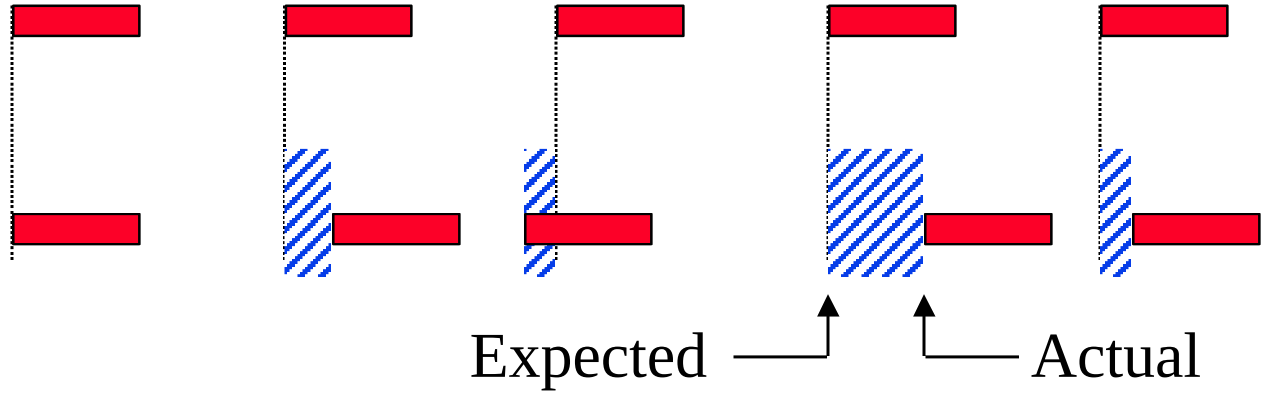
- **Random Discard**: Randomly select a cell in the queue and discard it.
  Problem: Need to look inside the queues.
- **Early Packet Discard**: When a queue threshold exceeds,
  - First cell of the next arriving packet is discarded.
  - Drop all cells of discarded packets.

Cell Delay Variation

- Cell Transfer Delay (CTD): First bit out to last bit in
- Cell Delay Variation (CDV) = CTD_{max} - CTD_{min}
  - Peak-to-peak CDV
  - Instantaneous CDV

Instantaneous CDV

- I-CDV= Actual-Expected arrival time
- Expected = Emission + Nominal delay
- Cell Delay Variation Window (CDV-W)
  CDV-W = |I-CDV(Max)| + |I-CDV(Min)|
- Cells arriving outside window are considered lost
- Large CDV ⇒ Large buffers ⇒ Higher cost



Generic Flow Control (GFC)

- ATM header format designed from DQDB cell headers. The first byte in DQDB is access control for contention.
- Some countries wanted to keep this in ATM for compatibility and others saw no use for it.
- Compromise ⇒ One half of the octet was reserved for "Generic Flow Control" to be defined at some future point.
- British Telecom adapted multiservice flow control (MSFC) from Orwell ring and proposed to use it for GFC

Fast Resource Management

- User requests to send a burst.
- Network reserves the resources and grants the request
- User sends the bursts
- If network rejects, ???

ERICA Features

- Measured overload/load at switch
- Small queue lengths during steady state
- Fast response due to optimistic design
- Parameters: Few, insensitive, easy
- Insensitive to source not using their allocated rates
- Several options: Backward Explicit Congestion Notification
- Simplified switch algorithm
- Optimized all steps. Eliminated unncessary steps. Eliminated many parameters

ERICA+: Switch Algorithm

- Target cell rate = Target Utilization × Link Capacity
- Target Utilization
  = fn(Current load, Queue length, Queue drain time goal)
- Rest is similar to ERICA
- Features:
  - Queue length is bounded during overload
  - No queue underflow ⇒ Switches keep ABR cells waiting to be transmitted as soon as the bandwidth becomes available.
  - 100% Utilization even with VBR

Control Mechanisms

- nrt-VBR: Open-Loop + Optional closed loop component
  - Traffic shaping, CLP
  - EFCI optional
    ⇒ No switch or end-system behavior specified
- UBR: Local policy. CLP=0 or 1. Not subject to CAC.
- ABR:
  - CLP=1 data cells not allowed.
  - No EFCI in RM cells
  - All parameters are negotiated independently for the two directions
  - MCR=0 not subject to CAC due to load

Rate Representation

Reserved	Nonzero	Exponent	Mantissa
1	1	5	9 ← Size in bits

- Rate in cells/second = $[2^e(1+m/512)]*nz$
- Example:

  0-1-01010-0 1100 1010 = $2^{18}(1+202/512)$

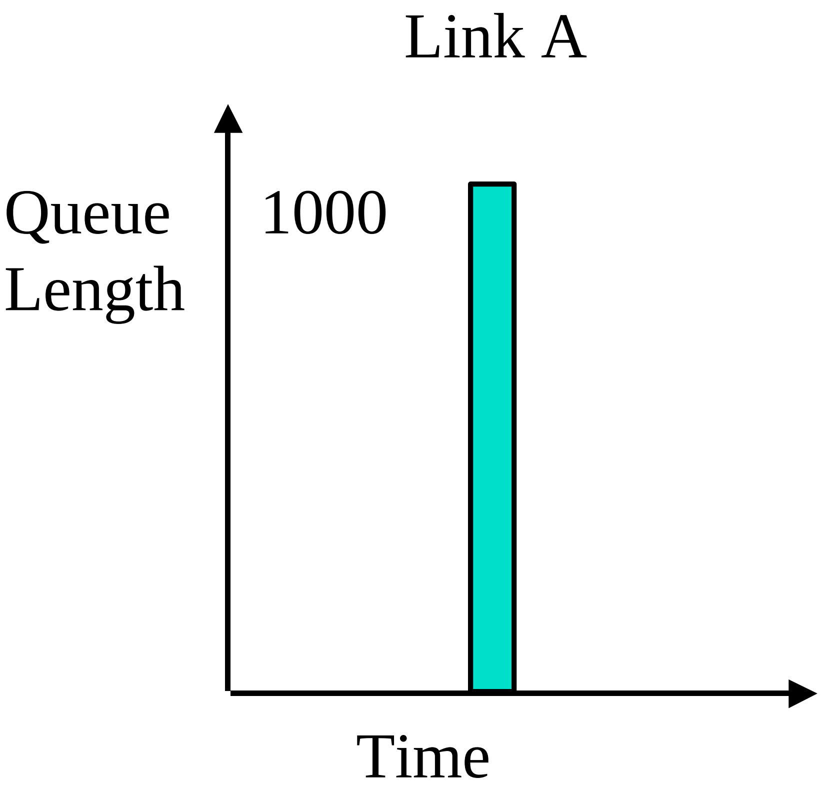
  = 262144×1.394523

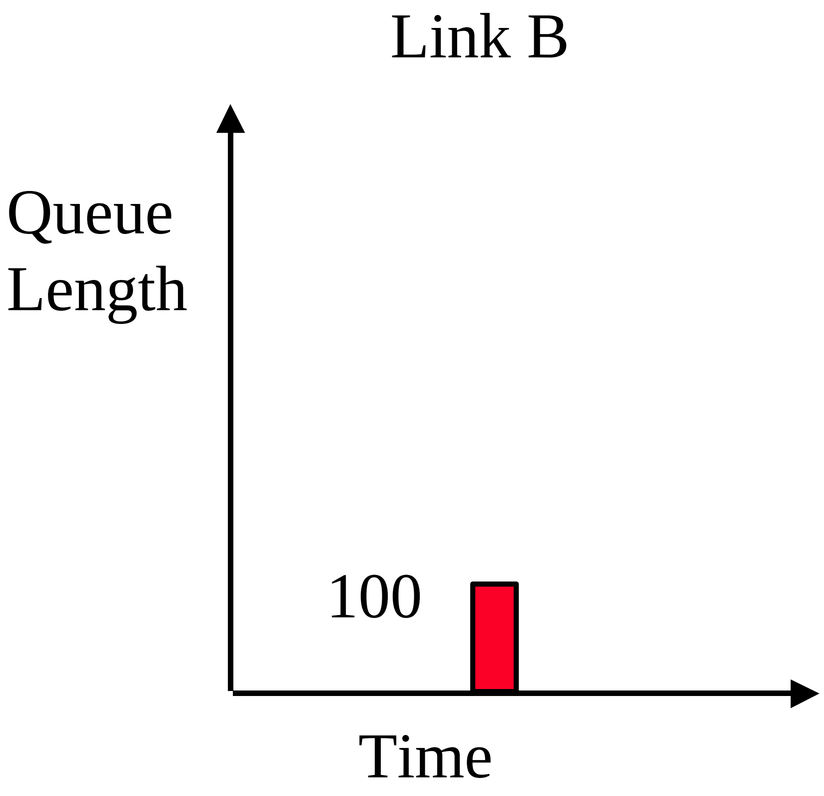
  = 365,566 cells/sec = 155 Mbps

OSU Congestion Principles

- Input rate (and not queue length) is the load measure
- Congestion avoidance (and not congestion control) should be the goal
- Transient performance (and not the steady state performance) is more important

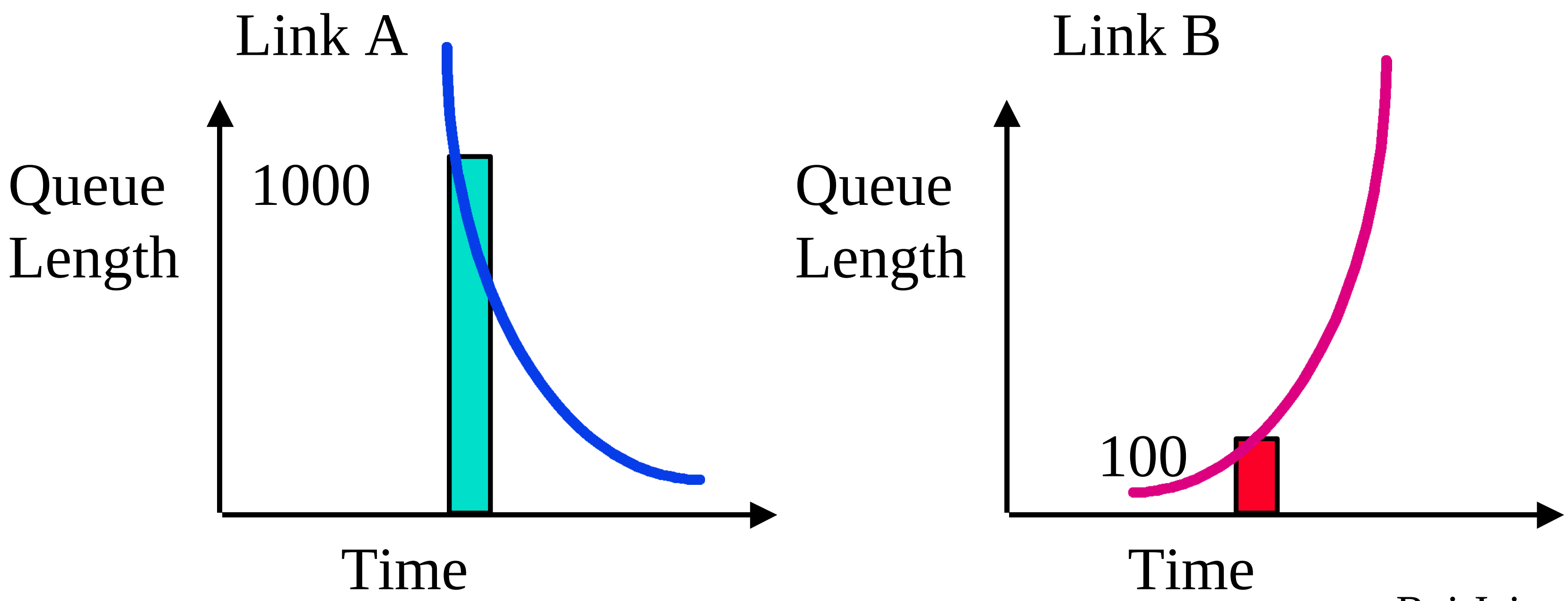
Which Link is More Overloaded?





Answer: It Depends!

- Link Speed: OC-12 or T1?
- Control: Rate or Window?
  Q = Window, dQ/dt = Rate
- For Rate Control: Monitor input rate

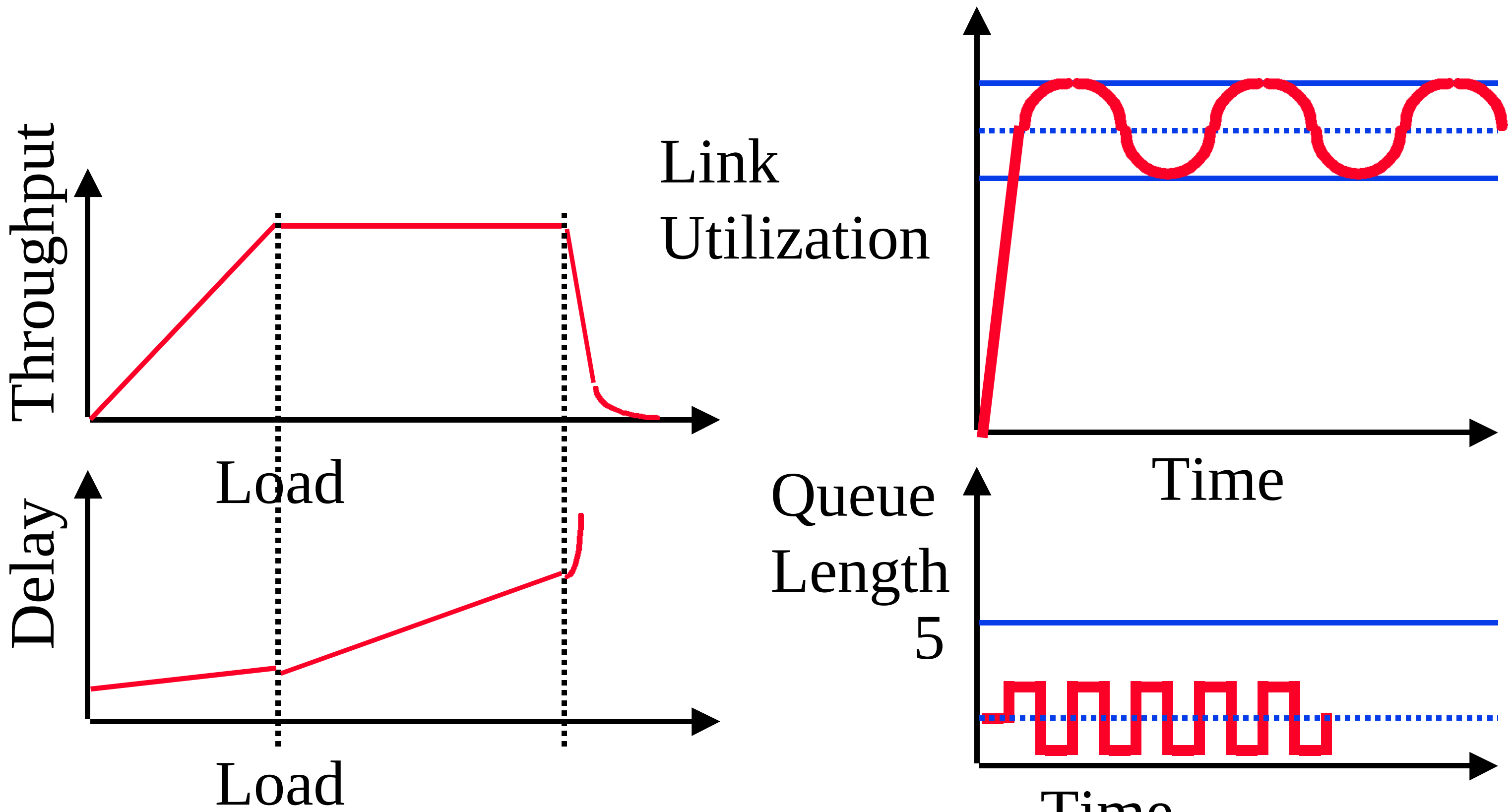


Conclusions I

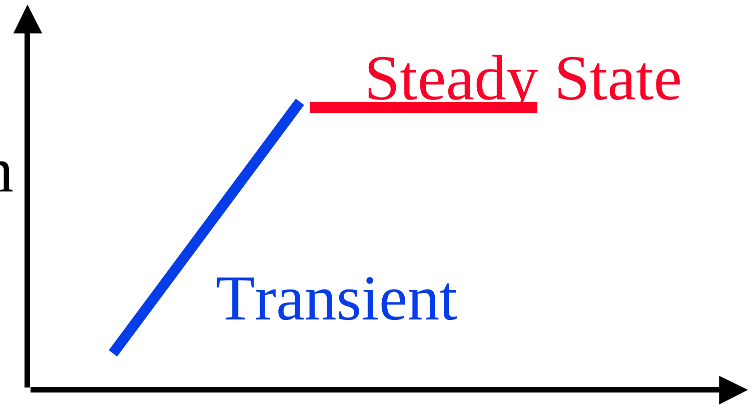
+ Instantaneous queue length is not a good indicator of load for a rate controlled system.
  Q(t) = Q(t-1) + Input rate(t) - Service rate(t)
+ Using queue length as the load indicator in a rate controlled system leads to unnecessary oscillations.
+ Input rate monitoring not only correctly tells whether the system is overloaded, it also tells by what factor.
+ Queue = n is not a good goal. Input rate=service rate is.

Congestion Avoidance

- High throughput, Low delay
- Small queues
- Bounded oscillations ⇒ Good for Video traffic



Why Worry About Transients?

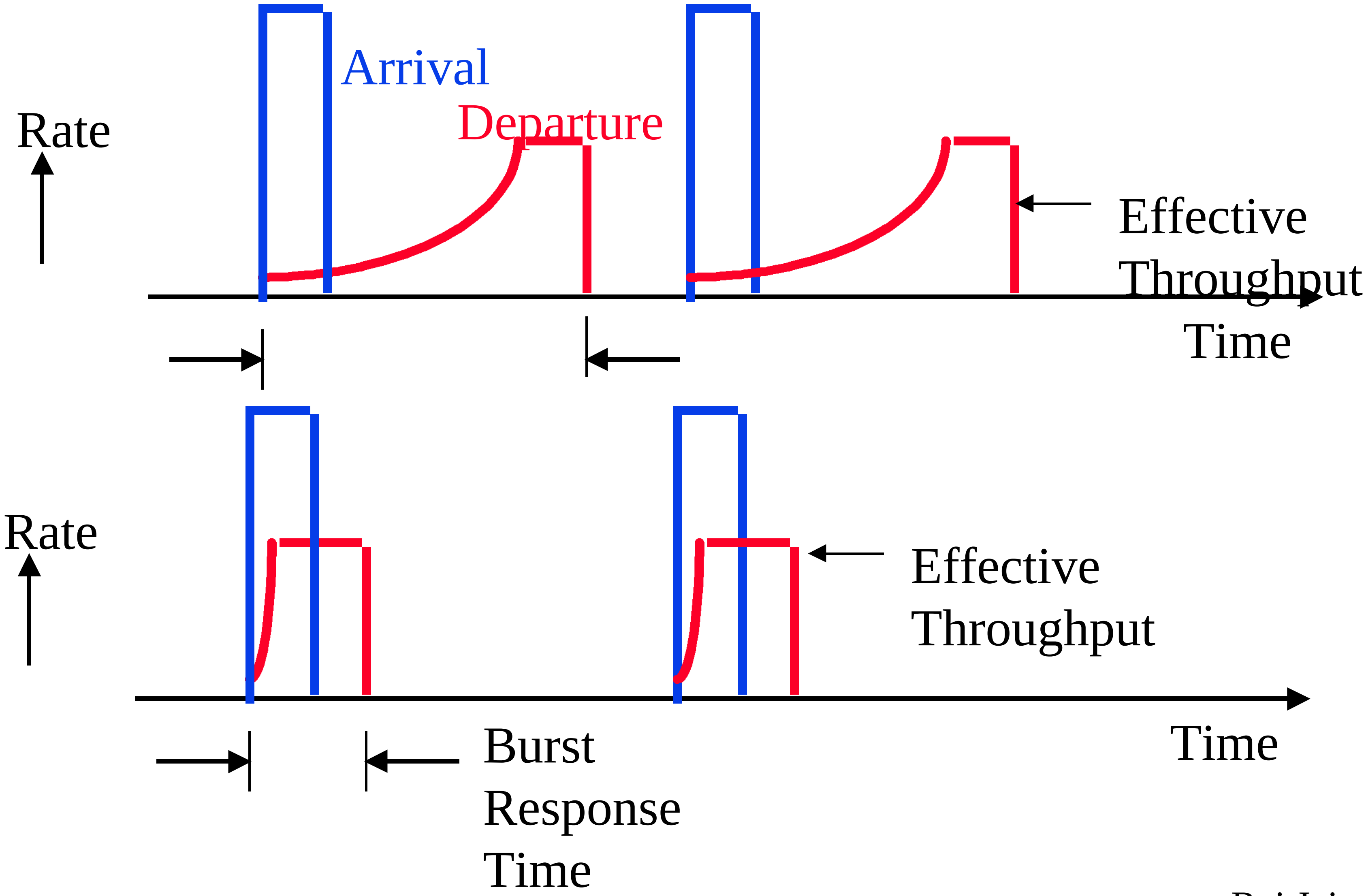


On most networks:

- There are no infinite sources.
- Sources come and go
- VCs may stay but are mostly inactive
- Traffic is highly bursty

⇒ Networks are operating in the transient region, most of the time.

Burst Performance



Legacy LANs vs ATM

- Today's LANs have a very fast transient response. Can get to the peak rate within a few microseconds
- On ATM LANs:
  Wait for connection setup and then...
  Everytime, a burst arrives, take several milliseconds to ramp up
- Q: Given 100 Mbps Switched Ethernet and 155 Mbps ATM at the same price, which one would you buy?

Quiz

T F Please check True/False

1. ❑ ❑ Congestion is not a problem in high-speed networks
2. ❑ ❑ User parameter control (UPC) allows a user to set its parameters
3. ❑ ❑ CDVT measures the cell delay variation caused by the network
4. ❑ ❑ ABR users do not have to specify CDVT
5. ❑ ❑ GCRA allows a network to determine conforming and non-conforming cells
6. ❑ ❑ All non-conforming cells are dropped at the source
7. ❑ ❑ Credit based scheme requires per-VC queueing
8. ❑ ❑ EFCI is better than explicit rate for high-speed networks
9. ❑ ❑ VSVD allows a network operator to use proprietary control scheme inside its networks.

Thank You!