Quality of Service In Data Networks: Problems, Solutions, and Issues

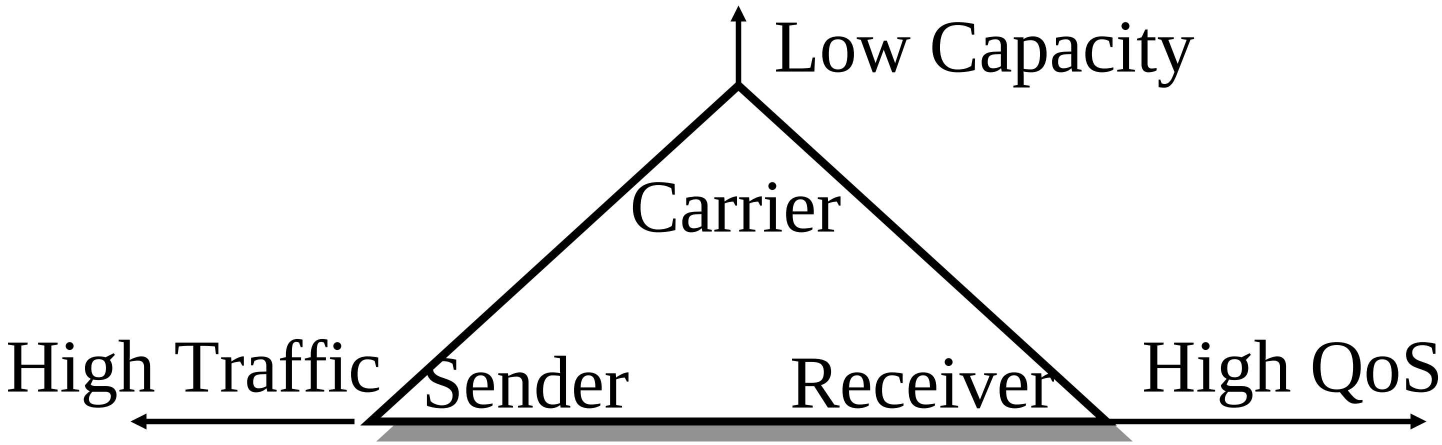
Raj Jain
The Ohio State University

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

state.edu/~jain/talks/qos9906.htm

- ATM QoS and Issues
- Integrated services/RSVP and Issues
- Differentiated Services and Issues
- QoS using MPLS
- End-to-end QoS
- This is an update to the May'98 talk
  http://www.cis.ohio-state.edu/~jain/talks/ipqos.htm

QoS Triangle



- Senders want to send traffic any time with high load, high burstiness
- Receivers expect low delay and high throughput
- Since links are expensive, providers want to minimize the infrastructure
- If one of the three gives in ⇒ no problem

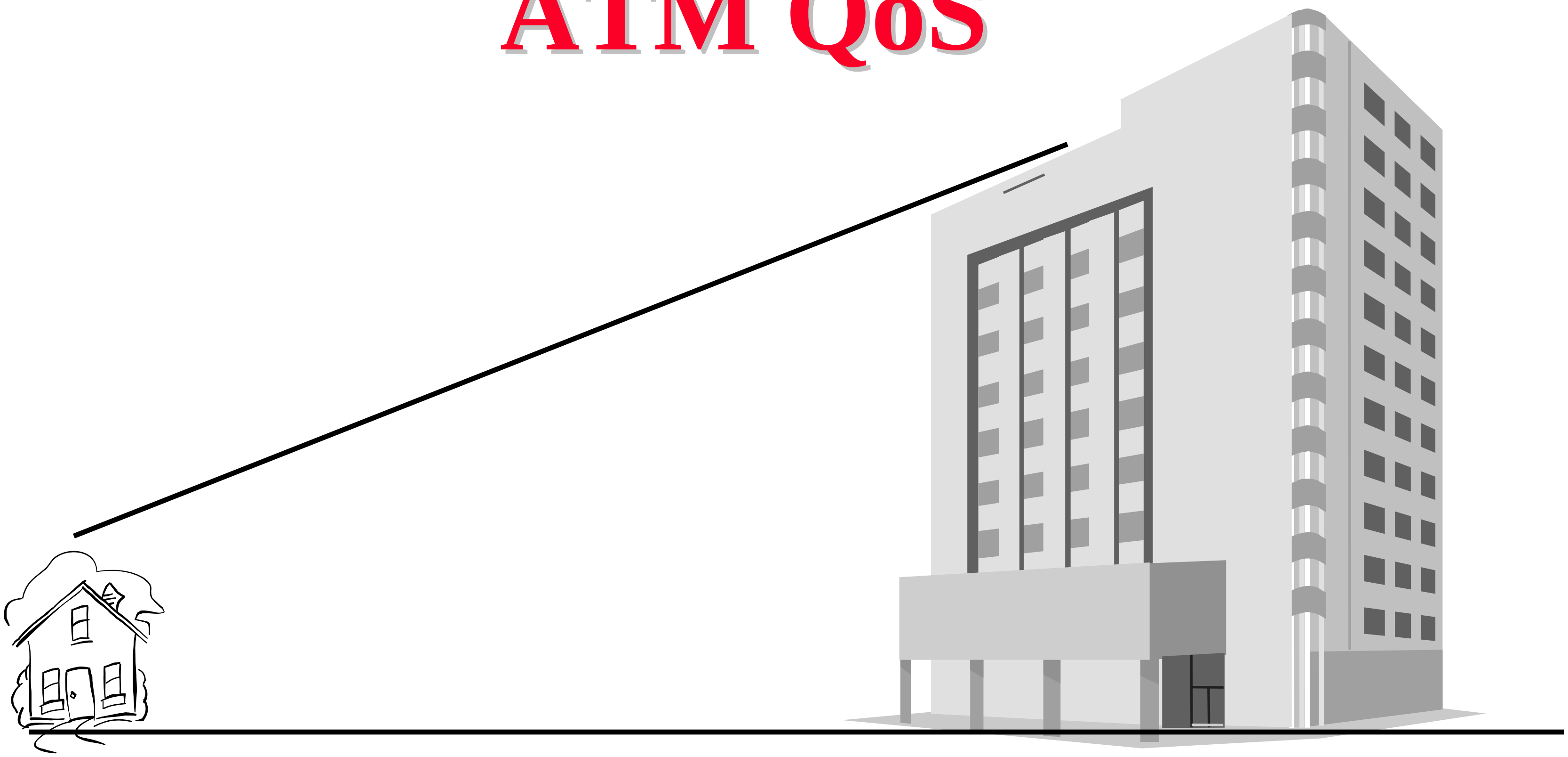
What is QoS?

- Predictable Quality: Throughput, Delay, Loss, Delay jitter, Error rate
- Opposite of best effort = Random quality
- Mechanisms:
  - Capacity Planning
  - Classification, Queueing, Scheduling, buffer management
  - QoS based path determination, Route pinning
  - Shaping, policing, admission control
  - Signaling

ATM Service Categories

- **CBR**: Throughput, delay, delay variation
- **rt-VBR**: Throughput, delay, delay variation
- **nrt-VBR**: Throughput
- **UBR**: No Guarantees
- **GFR**: Minimum Throughput
- **ABR**: Minimum Throughput. Very low loss. Feedback.
- ATM also has QoS-based routing (PNNI)

ATM QoS



Today ATM

Too much too soon

ATM QoS: Issues

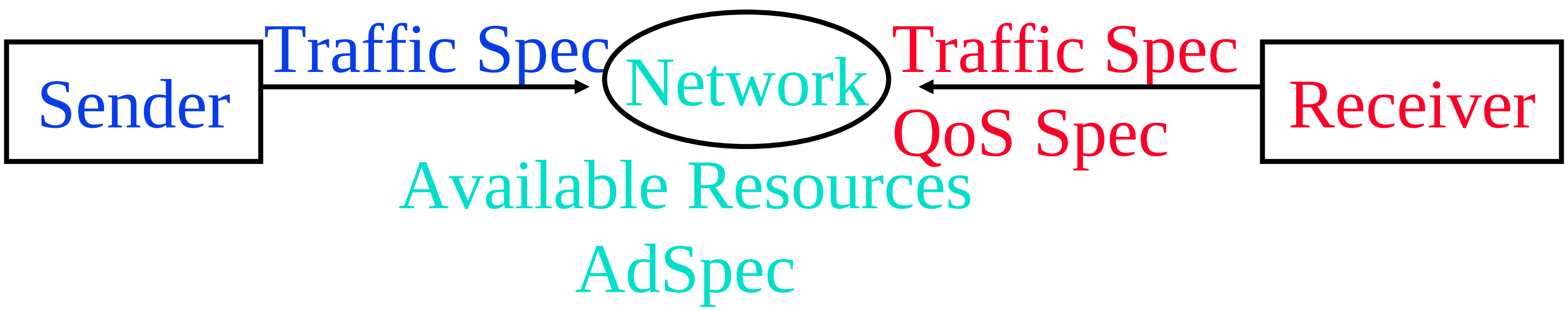
- Can't easily aggregate QoS: VP = Σ VCs
- Can't easily specifiy QoS: What is the CDV required for a movie?
- Signaling too complex ⇒ Need Lightweight Signaling
- Need Heterogeneous Point-to-Multipoint: Variegated VCs
- Need QoS Renegotiation
- Need Group Address
- Need priority or weight among VCs to map DiffServ and 802.1D

Integrated Services

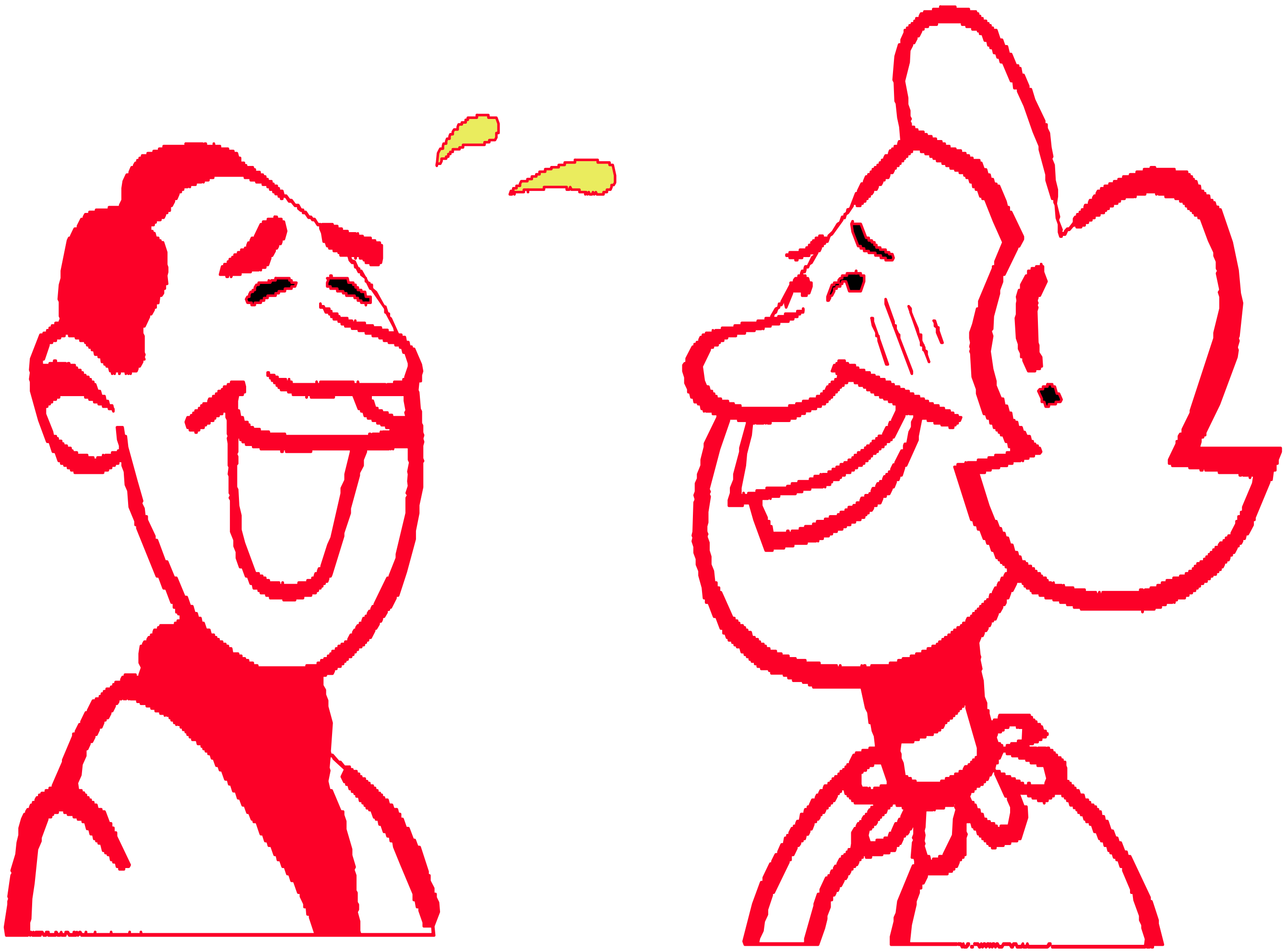
- Best Effort Service: Like UBR.
- Controlled-Load Service: Performance as good as in an unloaded datagram network. No quantitative assurances. Like nrt-VBR or UBR w MCR
- Guaranteed Service: rt-VBR
  - Firm bound on data throughput and delay.
  - Delay jitter or average delay not guaranteed or minimized.
  - Every element along the path must provide delay bound.
  - Is not always implementable, e.g., Shared Ethernet.
  - Like CBR or rt-VBR

RSVP

- Resource ReSerVation Protocol
- Internet signaling protocol
- Carries resource reservation requests through the network including traffic specs, QoS specs, network resource availability
- Sets up reservations at each hop



Before



After

Problems with RSVP and Integrated Services

- Complexity in routers: packet classification, scheduling
- Scalable in number of receivers per flow but
  Per-Flow State: O(n) ⇒ Not scalable with # of flows.
  Number of flows in the backbone may be large.
  ⇒ Suitable for small private networks
- Need a concept of "Virtual Paths" or aggregated flow groups for the backbone
- Need policy controls: Who can make reservations?
  Support for accounting and security.
  ⇒ RSVP admission policy (rap) working group.

Problems (Cont)

- Receiver Based:
  Need sender control/notifications in some cases.
  Which receiver pays for shared part of the tree?
- Soft State: Need route/path pinning (stability).
  Limit number of changes during a session.
- RSVP does not have negotiation and backtracking
- Throughput and delay guarantees require support of lower layers. Shared Ethernet ⇒ IP can't do GS or CLS. Need switched full-duplex LANs.
- Can't easily do RSVP on ATM either
- Most of these arguments also apply to integrated services.

Differentiated Services

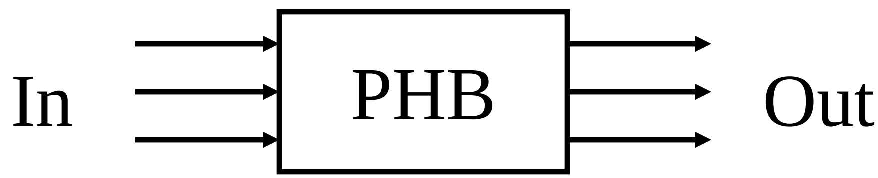
Ver	Hdr Len	Precedence	ToS	Unused	Tot Len
4b	4b	3b	4b	1b	16b

- IPv4: 3-bit precedence + 4-bit ToS
- OSPF and integrated IS-IS can compute paths for each ToS
- Many vendors use IP precedence bits but the service varies ⇒ Need a standard ⇒ Differentiated Services
- DS working group formed February 1998
- Charter: Define ds byte (IPv4 ToS field)
- Mail Archive: http://www-nrg.ee.lbl.gov/diff-serv-arch/

Service

- Service: Offered by the protocol layer
  - Application: Mail, FTP, WWW, Video,...
  - Transport: Delivery, Express Delivery,...
    Best effort, controlled load, guaranteed service
  - DS group will not develop services
    They will standardize "Per-Hop Behaviors"

Per-hop Behaviors

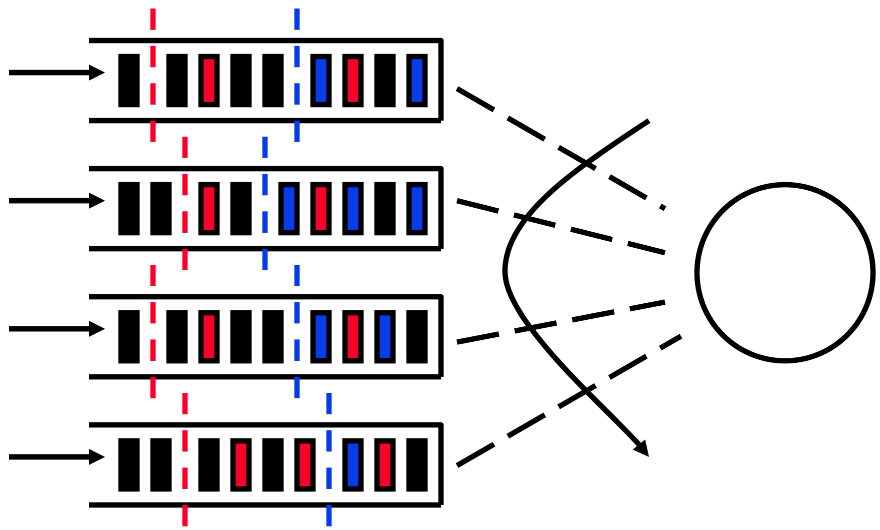


- Externally Observable Forwarding Behavior
- x% of link bandwidth
- Minimum x% and fair share of excess bandwidth
- Priority relative to other PHBs
- PHB Groups: Related PHBs. PHBs in the group share common constraints, e.g., loss priority, relative delay

Expedited Forwarding

- Also known as "Premium Service"
- Virtual leased line
- Similar to CBR
- Guaranteed minimum service rate
- Policed: Arrival rate < Minimum Service Rate
- Not affected by other data PHBs
  $\Rightarrow$ Highest data priority (if priority queueing)
- Code point: 101 110

Assured Forwarding



- PHB <u>Group</u>
- Four Classes: No particular ordering
- Three drop preference per class

Assured Forwarding (Cont)

- DS nodes SHOULD implement all 4 classes and MUST accept all 3 drop preferences. Can implement 2 drop preferences.
- Similar to nrt-VBR/ABR/GFR
- Code Points:

Drop Prec.	Class 1	Class 2	Class 3	Class 4
Low	010 000	011 000	100 000	101 000
Medium	010 010	011 010	100 010	101 010
High	010 100	011 100	100 100	101 100

- Avoids 11x000 (used for network control)

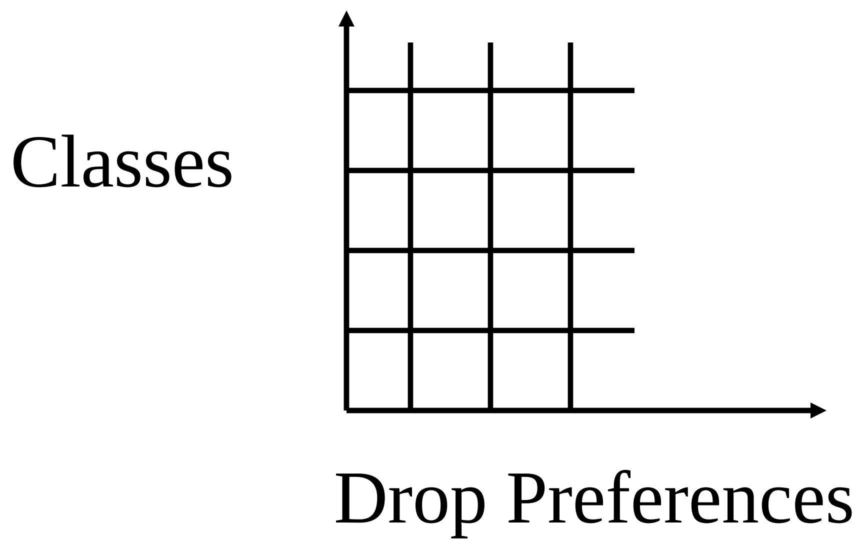
AF Simulation Results

1. W/O DPs, TCP is punished for good behaviour
2. Fairness is also poor.
3. Three DPs give better performance for TCP flows when there is considerable unused bandwidth.

Reason: TCP does not get any share of excess bandwidth in presence of UDP.

Reference: M. Goyal, et al, "Effect of Number of Drop Precedences in Assured Forwarding," IETF draft-goyal-dpstdy-diffserv-00.txt, March+July 1999, http://www.cis.ohio-state.edu/~jain/ietf/dpstdy2.htm

On Drop Preferences



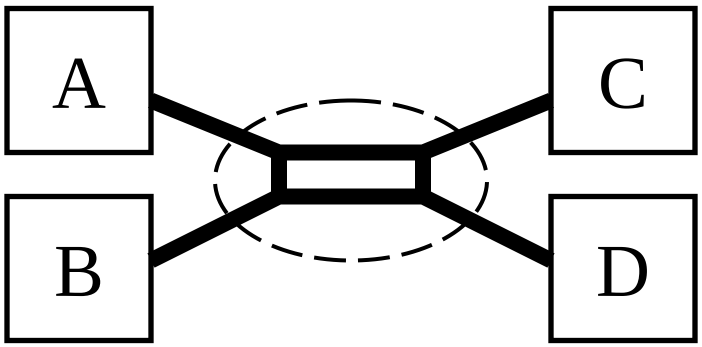
- We have two dimensions of control
  - Classes = Queues
  - Drop Preferences = Right to enter the queue
- Classes ⇒ Directly controls bandwidth allocation

Drop Preferences (Cont)

- DPs $\Rightarrow$ Controls buffer allocation
  $\Rightarrow$ Indirectly affects bandwidth allocation
  - Depends upon the arrival pattern
    $\Rightarrow$ Random $\Rightarrow$ Not Reliable
- Given a limited number of PHB's, it is better to have more classes than more DPs

DiffServ Problems (Cont)

- DiffServ is unidirectional ⇒ No receiver control
- Modified DS field ⇒ Theft and Denial of service. Ingress node should ensure.
- How to ensure resource availability inside the network?
- QoS is for the aggregate not per-destination. Multi-campus enterprises need inter-campus QoS.



DiffServ Problems (Cont)

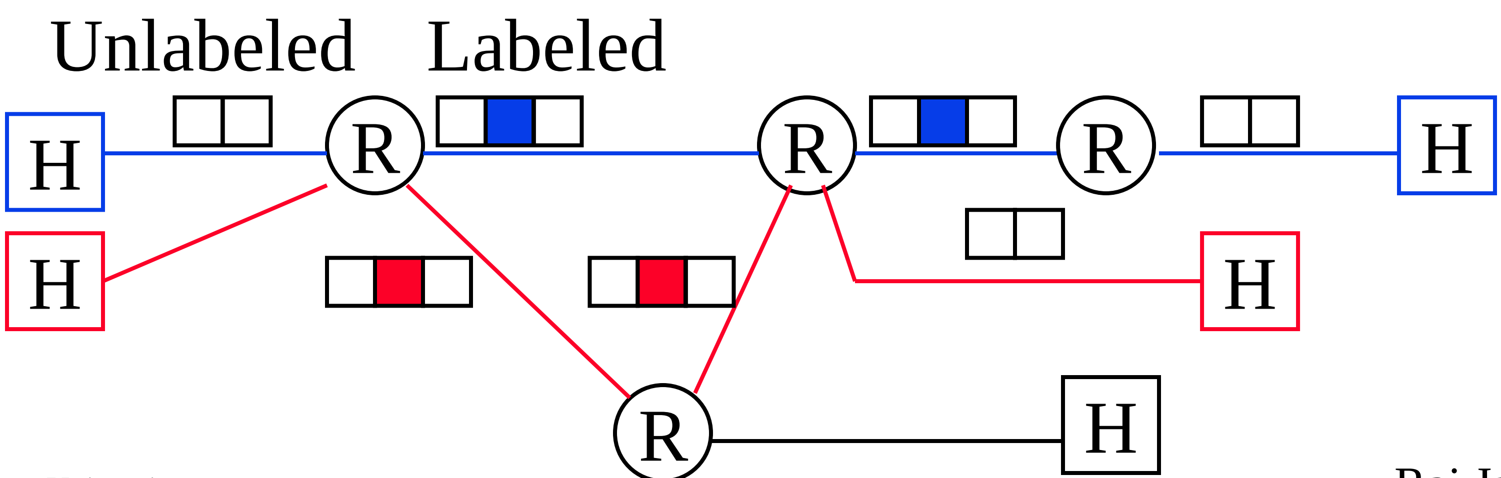
- QoS is for the aggregate not micro-flows.
  Not intended/useful for end users. Only ISPs.
  - Large number of short flows are better handled by aggregates.
  - Long flows (voice and video sessions) need per-flow guarantees.
  - High-bandwidth flows (1 Mbps video) need per-flow guarantees.
- All IETF approaches are open loop control ⇒ Drop
  Closed loop control ⇒ Wait at source
  Data prefers waiting ⇒ Feedback

DiffServ Problems (Cont)

- Guarantees $\Rightarrow$ Stability of paths
  $\Rightarrow$ Connections (hard or soft)
  Need route pinning or connections.

Multiprotocol Label Switching

- Label = Circuit number = VC Id
- Ingress router/host puts a label.
  Exit router strips it off.
- Switches switch packets based on labels.
  Do not need to look inside ⇒ Fast.

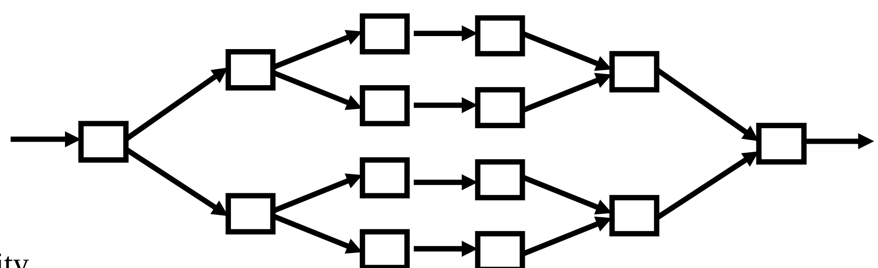


Traffic Engineering Objectives

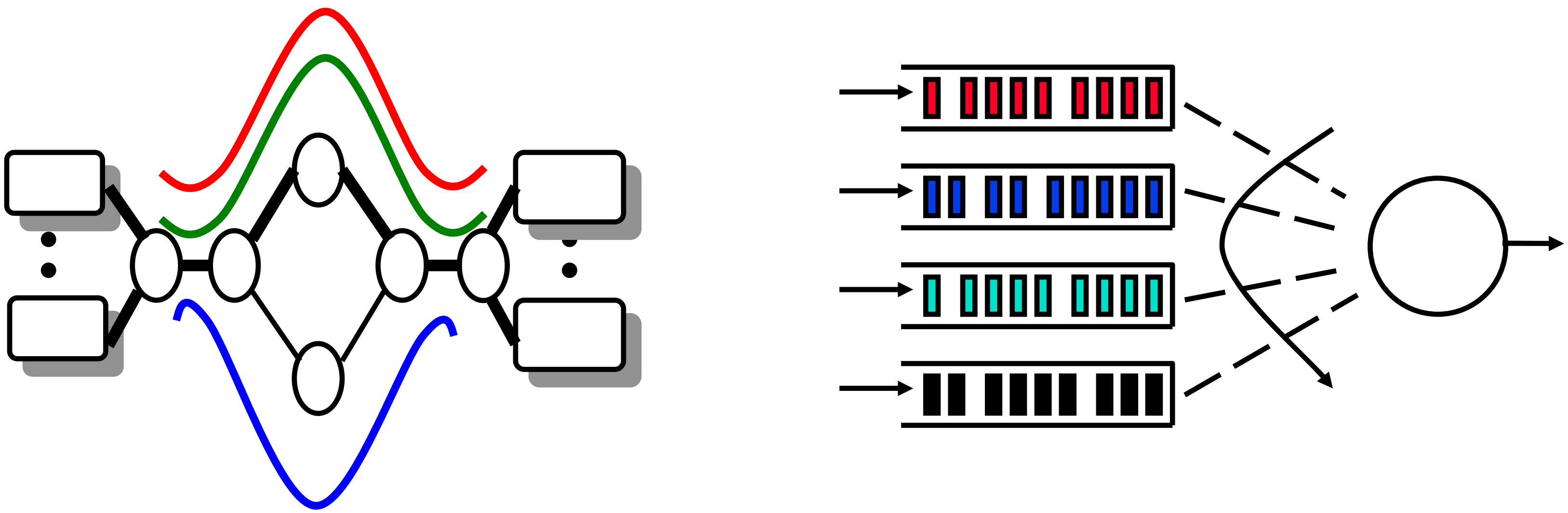
- User's Performance Optimization
  $\Rightarrow$ Maximum throughput, Min delay, min loss, min delay variation
- Efficient resource allocation for the provider
  $\Rightarrow$ Efficient Utilization of all links
  $\Rightarrow$ Load Balancing on parallel paths
  $\Rightarrow$ Minimize buffer utilization
  - Current routing protocols (e.g., RIP and OSPF) find the shortest path (may be over-utilized).
- QoS Guarantee: Selecting paths that can meet QoS
- Enforce Service Level agreements
- Enforce policies: Constraint based routing $\supseteq$ QoSR

MPLS Mechanisms for TE

- Signaling, Admission Control, Routing
- Explicit routing of LSPs
- Constrained based routing of LSPs
  Allows both Traffic constraints and Resource Constraints (Resource Attributes)
- Hierarchical division of the problem (Label Stacks)
- Traffic trunks allow aggregation and disaggregation (Shortest path routing allows only aggregation)

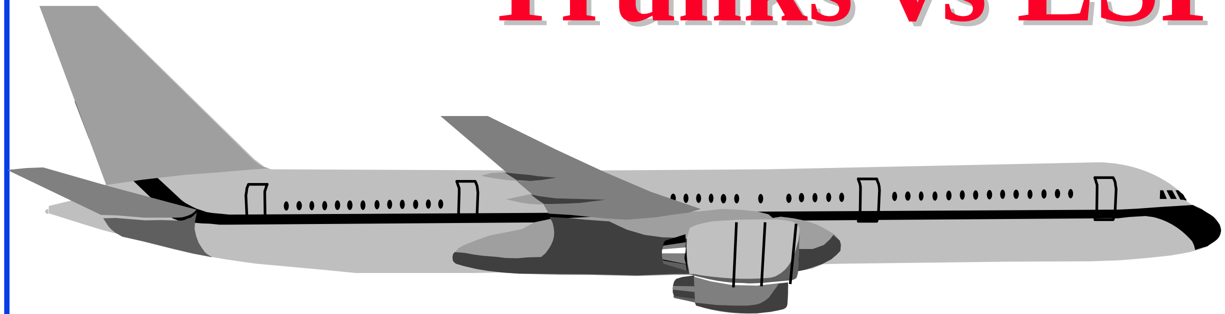


Traffic Trunks



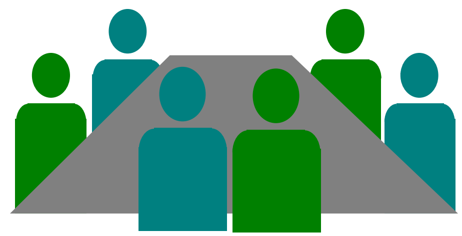
- Trunk: Aggregation of flows of same class on same LSP
- Trunks are routable
  ⇒ LSP through which trunk passes can be changed
- Class ⇒ Queue, LSP ⇒ Next hop
  Class can be coded in Exp or Label field. Assume Exp.

Trunks vs LSPs

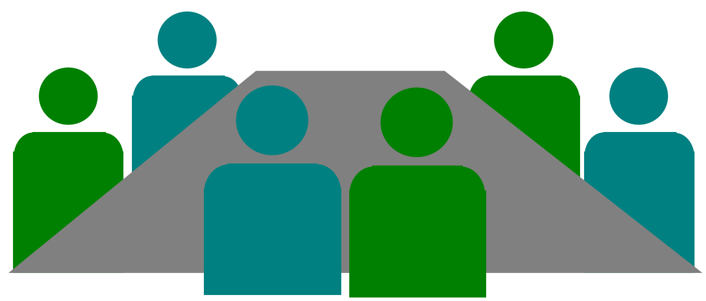


Tour Group

First Class



Business Class



Coach Class

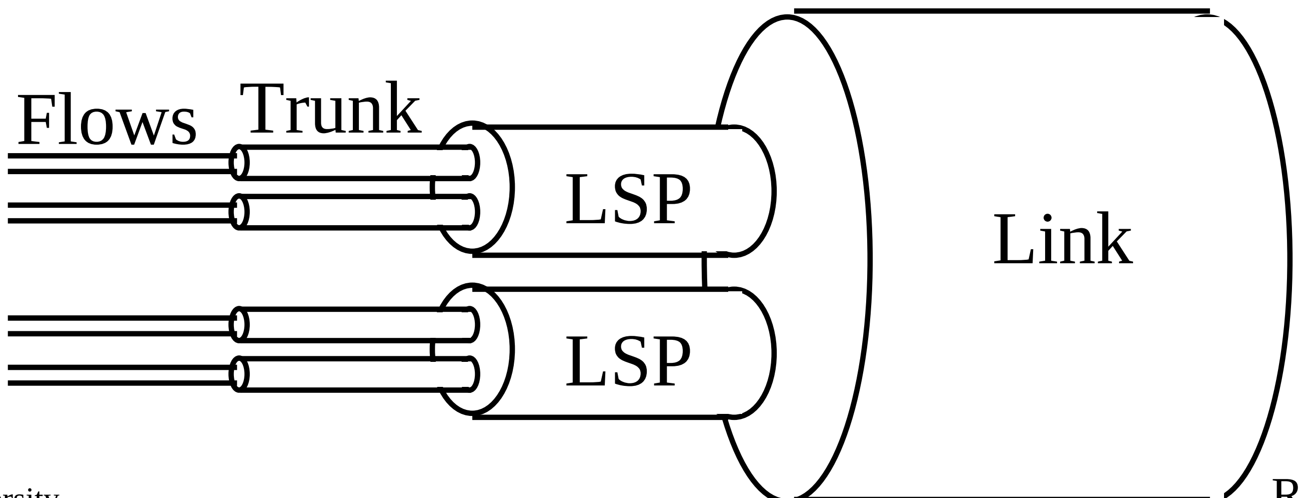
Flights = LSP

Tour Groups = Trunks

Flows, Trunks, LSPs, and Links

- Label Switched Path (LSP):
  All packets with the same label
- Trunk: Same Label+Exp
- Flow: Same MPLS+IP+TCP headers

DL	Label	Exp	SI	TTL	IP	TCP	



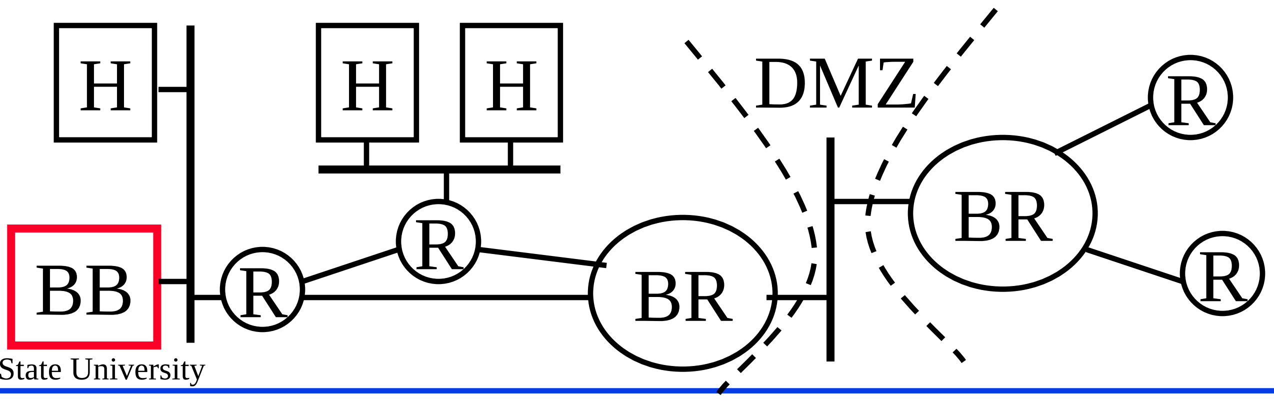
MPLS Simulation Results

- Total network throughput improves significantly with proper traffic engineering
- Congestion-unresponsive flows affect congestion-responsive flows
  - Separate trunks for different types of flows
- Trunks should be end-to-end
  - Trunk + No Trunk = No Trunk

Reference: P. Bhaniramka, et al, "*QoS using Traffic Engineering over MPLS: An Analysis*," IETF draft-bhani-mpls-te-anal-00.txt, March 1999, http://www.cis.ohio-state.edu/~jain/teanal.htm

Bandwidth Broker

- Repository of policy database. Includes authentication
- Users request bandwidth from BB
- BB sends authorizations to leaf/border routers
  Tells what to mark.
- Ideally, need to account for bandwidth usage along the path
- BB allocates only boundary or bottleneck



IEEE 802.1D Model

Dest Addr	Src Addr	Tag Prot ID	Pri	CFI	VLAN ID

← 802.1Q header → (Tag Prot ID, Pri, CFI, VLAN ID)

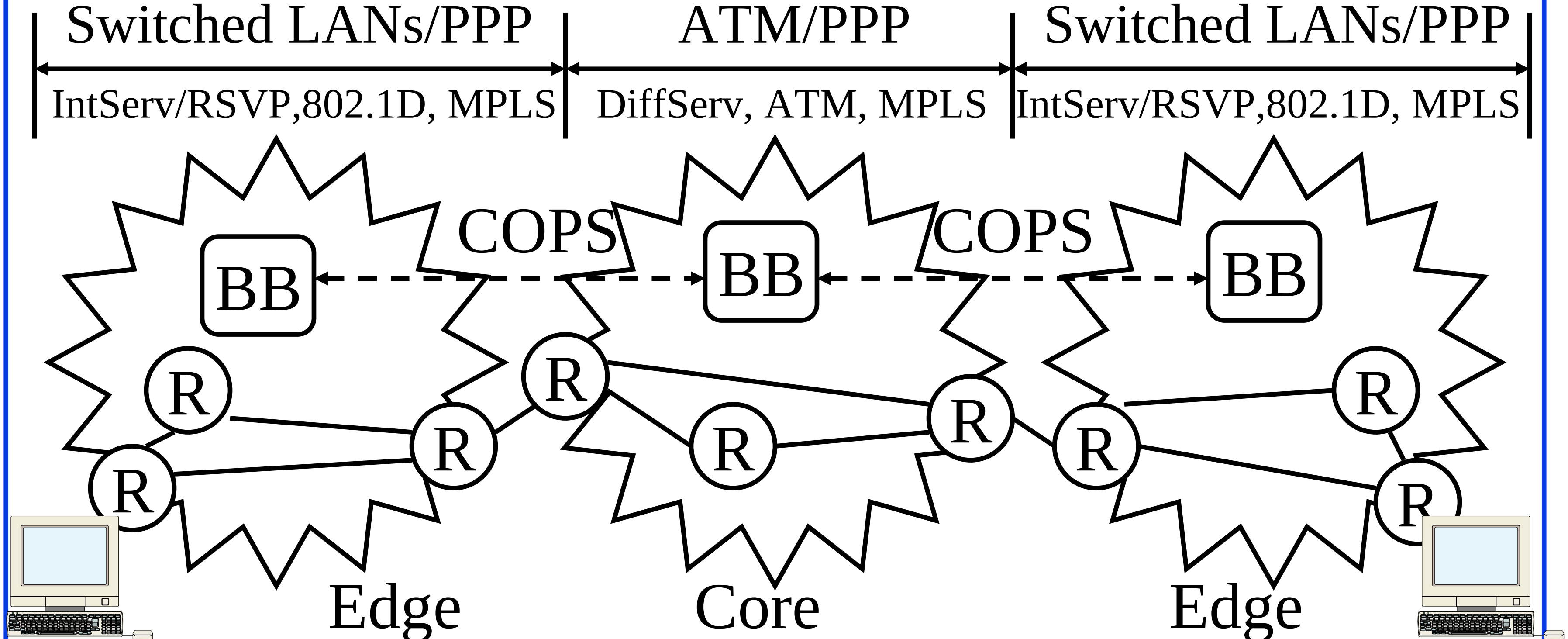
	Prot Type	Payload	FCS

CFI = Canonical Format Indicator (Source Routing)

- **Up to eight priorities:** Strict.
  1 Background
  2 Spare
  0 Best Effort
  3 Excellent Effort
  4 Control load
  5 Video (Less than 100 ms latency and jitter)
  6 Voice (Less than 10 ms latency and jitter)
  7 Network Control

End-to-end View

- ATM/PPP backbone, Switched LANs/PPP in Stub
- IntServ/RSVP, 802.1D, MPLS in Stub networks
- DiffServ, ATM, MPLS in the core



QoS Debate Issues

- Massive Bandwidth vs Managed Bandwidth
- Per-Flow vs Aggregate
- Source-Controlled vs Receiver Controlled
- Soft State vs Hard State
- Path based vs Access based
- Quantitative vs Qualitative
- Absolute vs Relative
- End-to-end vs Per-hop
- Static vs Feedback-based
- One-way multicast vs n-way multicast
- Homogeneous multicast vs heterogeneous multicast
- Single vs multiple bottlenecks: Scheduling

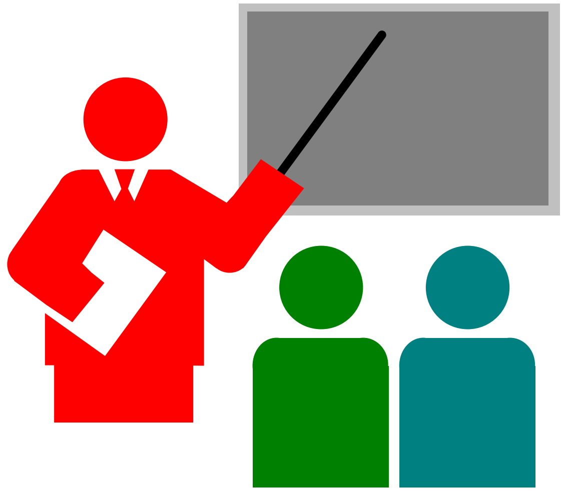
Comparison of QoS Approaches

Issue	ATM	IntServ	DiffServ	MPLS	IEEE 802.3D
Massive Bandwidth vs Managed Bandwidth	Managed	Managed	Massive	Managed	Massive
Per-Flow vs Aggregate	Both	Per-flow	Aggregate	Both	Aggregate
Source-Controlled vs Receiver Controlled	Unicast Source, Multicast both	Receiver	Ingress	Both	Source
Soft State vs Hard State	Hard	Soft	None	Hard	Hard
Path based vs Access based	Path	Path	Access	Path	Access
Quantitative vs Qualitative	Quantitativ e	Quantitativ e+Qualitati ve	Mostly qualitative	Both	Qualitative
Absolute vs Relative	Absolute	Absolute	Mostly Relative	Absolute plus relative	Relative

Comparison (Cont)

Issue	ATM	IntServ	DiffServ	MPLS	IEEE 802.3D
End-to-end vs Per-hop	e-e	e-e	Per-hop	e-e	Per-hop
Static vs Feedback-based	Both	Static	Static	Static	Static
One-way multicast vs n-way multicast	Only one-way				
Homogeneous multicast vs heterogeneous multicast	Homogeneous	Heterogeneous	N/A	Homogeneous	N/A
Single vs multiple bottlenecks: Scheduling	Multiple bottleneck	Multiple		Multiple	

Summary



- ATM: CBR, VBR, ABR, UBR, GFR
- Integrated Services: GS = rtVBR, CLS = nrt-VBR
- Signaling protocol: RSVP
- Differentiated Services will use the DS byte
- MPLS allows traffic engineering and is most promising
- 802.1D allows priority

References

- For a detailed list of references see: refs/ipqs_ref.htm
- Additional papers and presentations on QoS are at: http://www.cse.ohio-state.edu/~jain/

Thank You!