DiffServ, IntServ, MPLS

Udo Payer

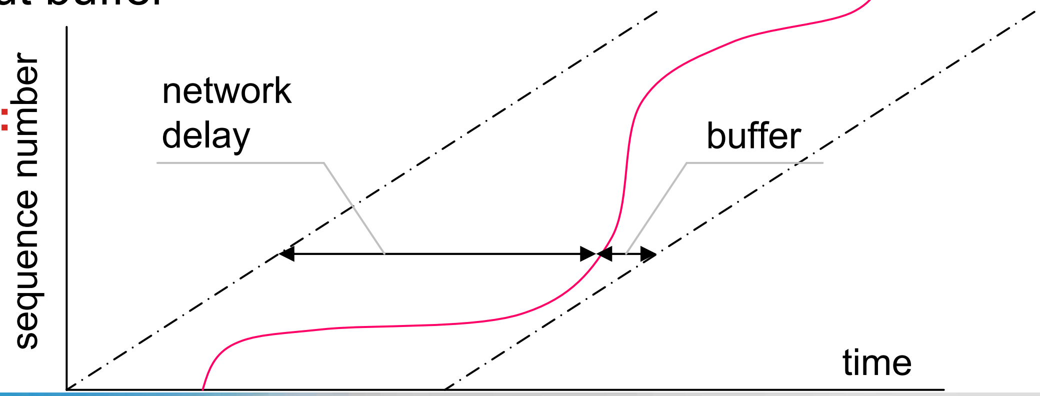
SS 2005

recapitulation RKN2004

Quality of Service (QoS)

- (Network) application requirements:
  - Realtime applications
  - Non-realtime application (*elastic-application* – since they can be stretched gracefully in the face of increasing delay)
- Voice applications:
  - At a rate of one per 125µs
  - Queue-length in switches an routers vary with time
    - If packets are arriving in time, the are stored in a *playback buffer*
    - If packets are late, they do not have to be stored very long
    - If packets are to late (arriving after their *playback time*) → draining of input buffer

For audio applications: maximal Mouth-to-Ear delay of 300ms



Quality of Service (QoS)

- Approaches to QoS support
  - Fine grained approaches: provides QoS in individual applications or flows:

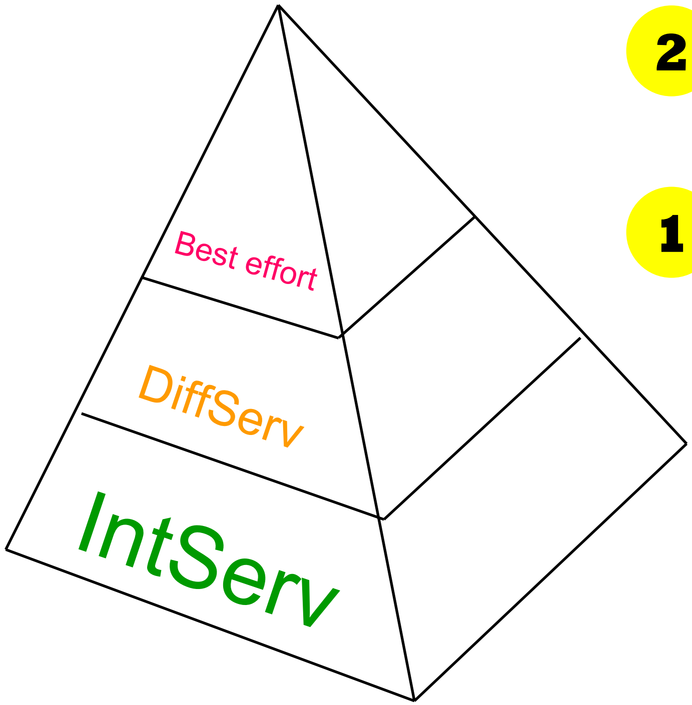
    … here we find "Integrated Services" (developed in the IETF) and often associated with the *Reservation Protocol* (RSVP)
  - Coarse grained approach: provides QoS to large classes of data or aggregated traffic

    … here we find "Differentiated Services" (undergoing standardization at the time of writing)

ATM is known to have a rich set of QoS capabilities and is considered in the fine-grained category (since resources are associated with individual VCs).

ATM is often used to interconnect routers – and may choose to send a highly aggregated traffic down a single VC → so ATM can be used for coarse grained QoS as well.

Quality of Service (QoS)



- **0** **Best-Effort**—Best-Effort does not provide QoS, because there is no reordering of packets.
- **2** **Differentiated Services (DiffServ)**—DiffServ, as the name suggests, differentiates between multiple *traffic flows*.
- **1** **Integrated Services (IntServ)**—IntServ is often referred to as "Hard QoS," because it can make strict bandwidth reservations. Needs signaling first. Must be configured on every router along a path. The main drawback of IntServ is its lack of scalability. Specifically, packets are "marked," and routers and switches can then make decisions. Bandwidth reservation based on application level.

Quality of Service (QoS)

1 Integrated Services (RSVP)

- The "Integrated Service Working Group" developed specifications of a number of service classes, designed to meet the needs of some application types (1995-97)
- It also defines, how **RSVP** can be used to make reservations, based on these service classes

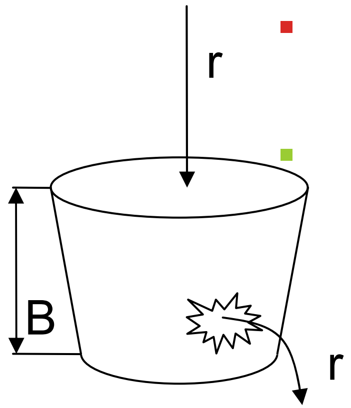
- Service classes:
  - Intolerant applications: maximum delay of any packet is guaranteed by the network → playback point can be set
  - Controlled load: By using queuing mechanisms, we can isolate the controlled load traffic from the other traffic
  - …
- Overview of Mechanisms:
  - Provide the network with a set of information → flowspec
  - Ask the network to provide particular services → admission control

?

Quality of Service (QoS)

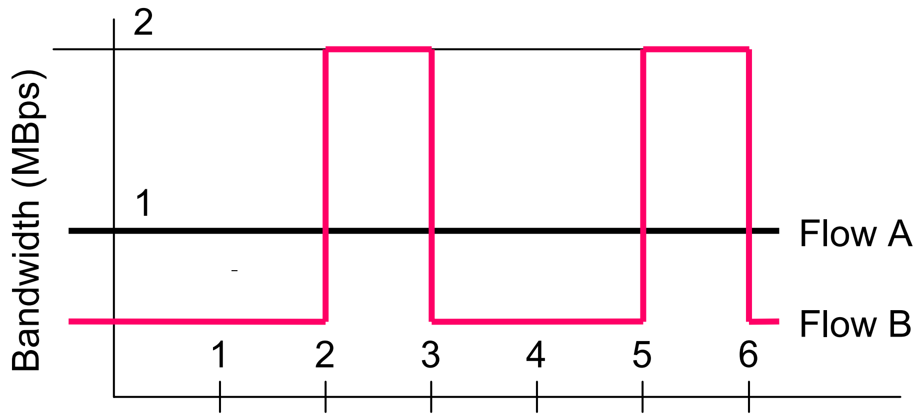
- User (application) and network have to exchange information to request services, *flowspecs and admission control* → resource reservation (*signalling*)
- Manage the way how packets are queued and scheduled in the switches and routers → packet scheduling (*traffic policing*)

- <u>Flowspecs</u>
  - TSpec: describing the flow's traffic characteristics – give the network enough information about the needed bandwidth → for intelligent admission control decisions.
  - RSpec: describing the service request (ex. request for controlled load, or delay bound)
  - Token Bucket filters are used for admission control
    - Token rate **r**, and token depth **B**
    - We can send a burst of **B** bytes, but we cannot send more than **r** bytes per second for a longer period



Quality of Service (QoS)

- Token Bucket



- Flow A generates a steady rate of 1MBps (r=1MBps, B=1B)
- Flow B sends an average rate of 1MBps (r=1MBps, B=1MB)

!

A single flow can be described by many different token buckets – but we have to avoid over-allocation of resources in the network.

Quality of Service (QoS)

- Admission Control
  - Admission control looks at TSpec and RSpec and decides, if a desired service can be provided to that amount of traffic.
  - For a controlled load service, the decision may be based on heuristics *“last time I allowed a flow with this TSpec into this class, but the delay for this class exceeded the acceptable bound – so I had better say no”* or *“my delay are so fair inside the bounds, that I should be able to admit another flow”*
  - Admission control := per-flow decision to admit a new flow …
  - Policing := per-packet decision – to make sure that all flows are conform to their TSpec’s

?

Quality of Service (QoS)

- Reservation Protocol (IETF, May 1993)
  - RSVP is trying to maintain the robustness of a network, by two highly innovative features:
    - receiver initiation and "soft state"

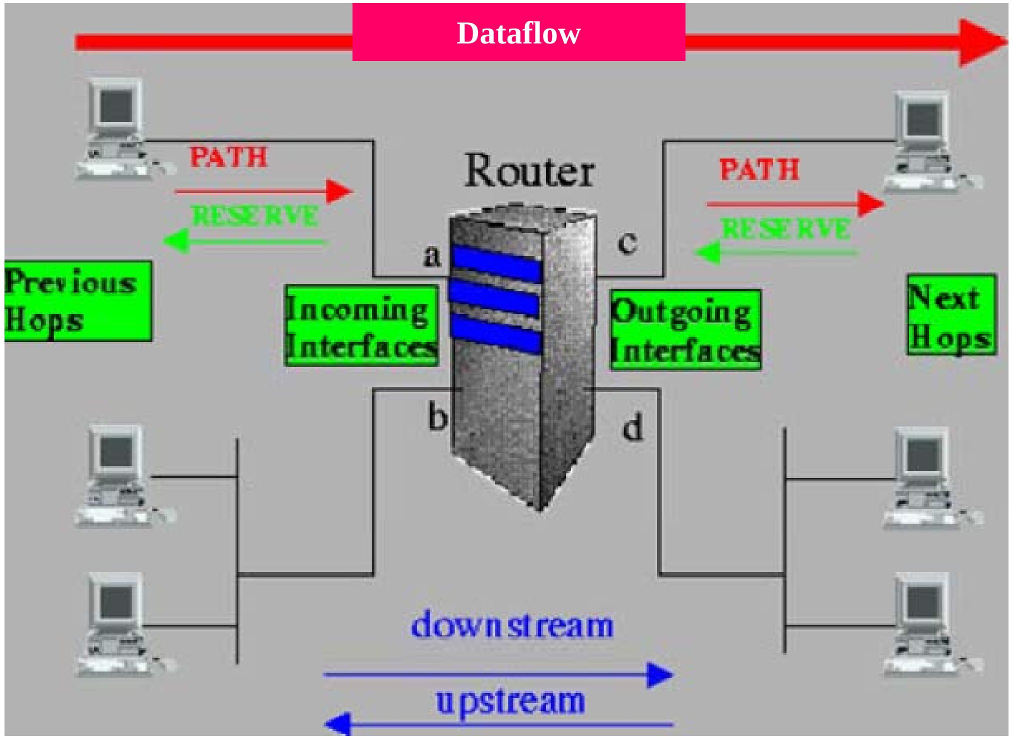
      (*Hard states – found in connection oriented networks*)
  - Soft states do not need to be explicit deleted when it is no longer needed
  - RSVP → Receiver-oriented approach (in contrast to connection oriented networks → resource reservation is done by the sender)
  - Simple to increase or decrease the level of resource allocation:
    - Each receiver is periodically sending refresh messages to keep the soft-state in place

… an IntServ QoS-example

Quality of Service (QoS)

- Simplex link: for up- and downstream two logical RSVP-sessions are needed:

Sender



Receiver

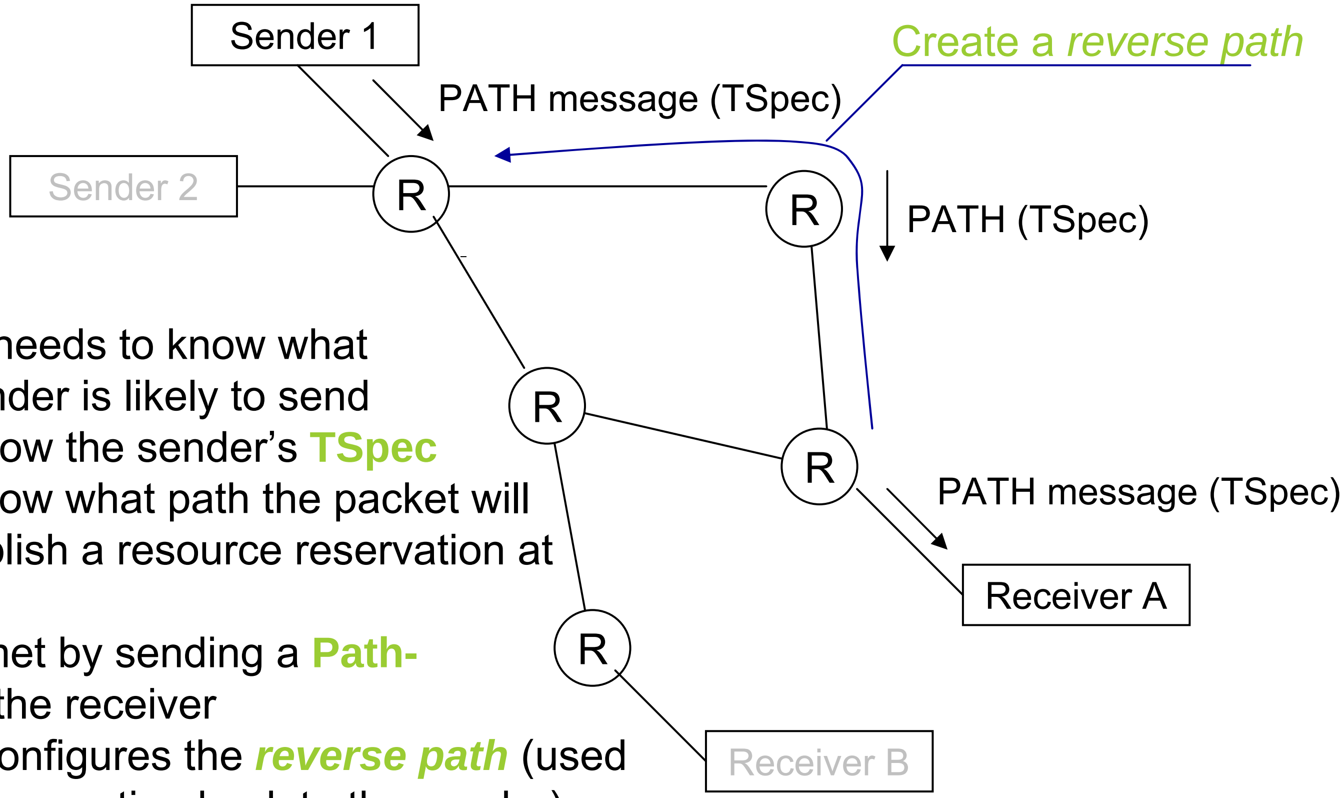
Quality of Service (QoS)

- Reservation Protocol
  1. The receiver needs to know, what traffic the sender is likely to send → it need to know the sender's TSpec.
  - … it needs too know what path the packet will go …
  - Establish a resource reservation at each router on the path
  2. Send a message from sender to receiver (containing the *TSpec* → **Path Message**)
     - Each router is looking at this packet
  - Reservations are sent back to the sender (**RESV Message** → containing the *RSpec's*)

  … if the reservation can be made → the RESV request is passed to the next router

Quality of Service (QoS)

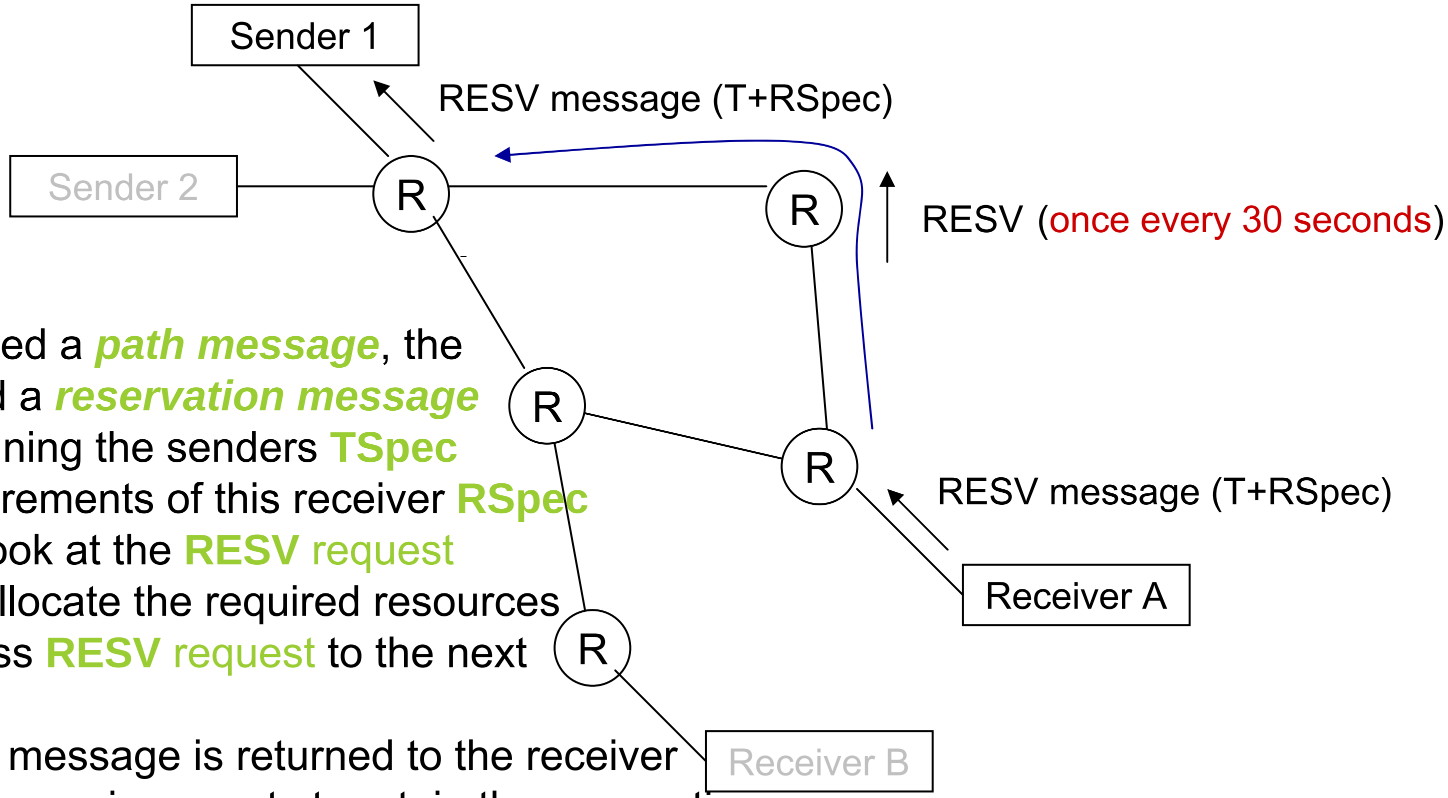
- One sender – one receiver



- The receiver needs to know what traffic the sender is likely to send
- It needs to know the sender's **TSpec**
- It needs to know what path the packet will flow (to establish a resource reservation at each router)
- This can be met by sending a **Path-message** to the receiver
- Each router configures the ***reverse path*** (used to send the reservation back to the sender)

Quality of Service (QoS)

- One sender – one receiver



•Having received a ***path message***, the receiver send a ***reservation message*** back – containing the senders **TSpec** and the requirements of this receiver **RSpec**

•Each router look at the **RESV** request and tries to allocate the required resources

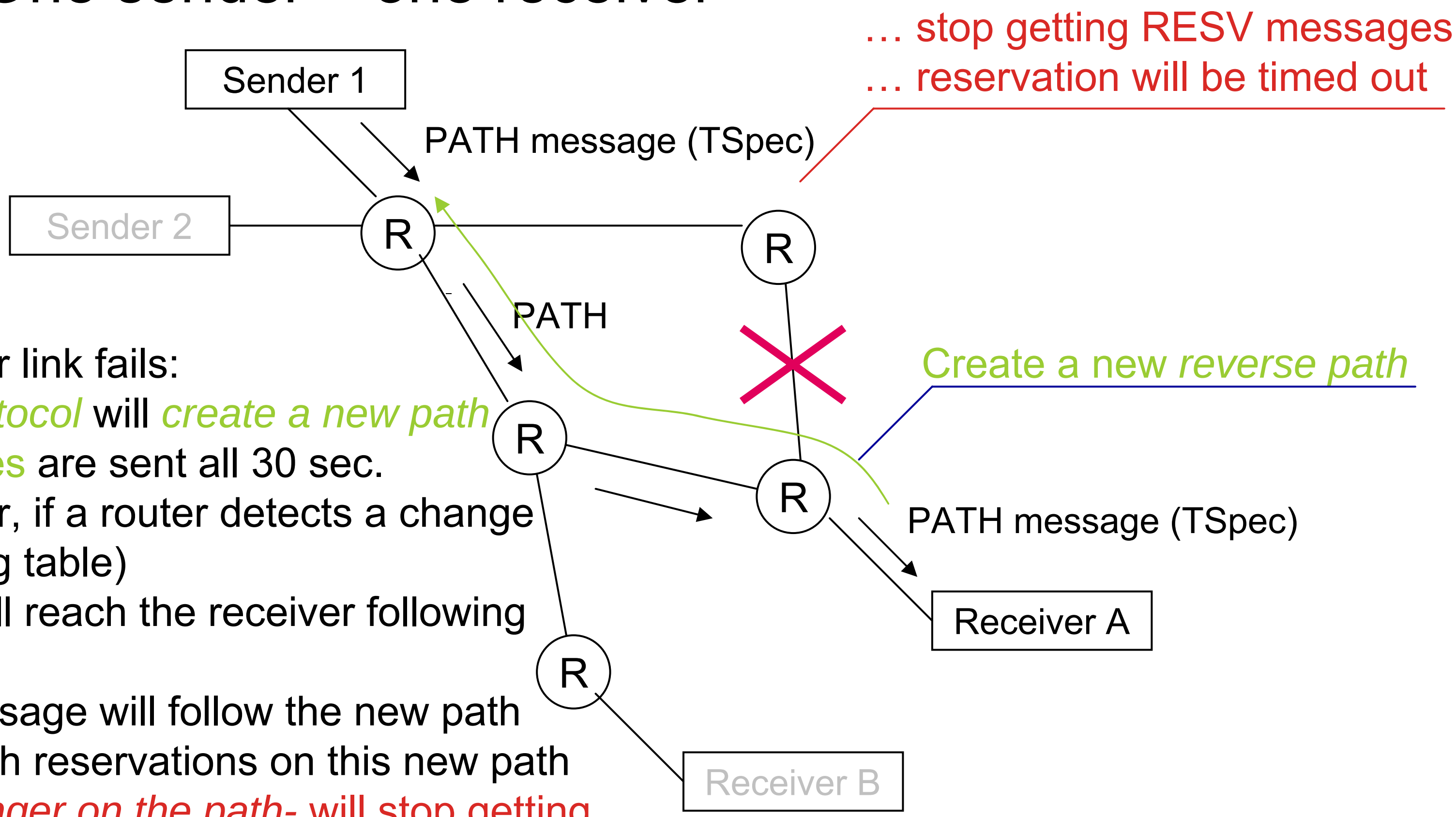
If possible: pass **RESV** request to the next router

If not: an error message is returned to the receiver

•As long as the receiver wants to retain the reservation it sends the same **RESV** message all 30 seconds

Quality of Service (QoS)

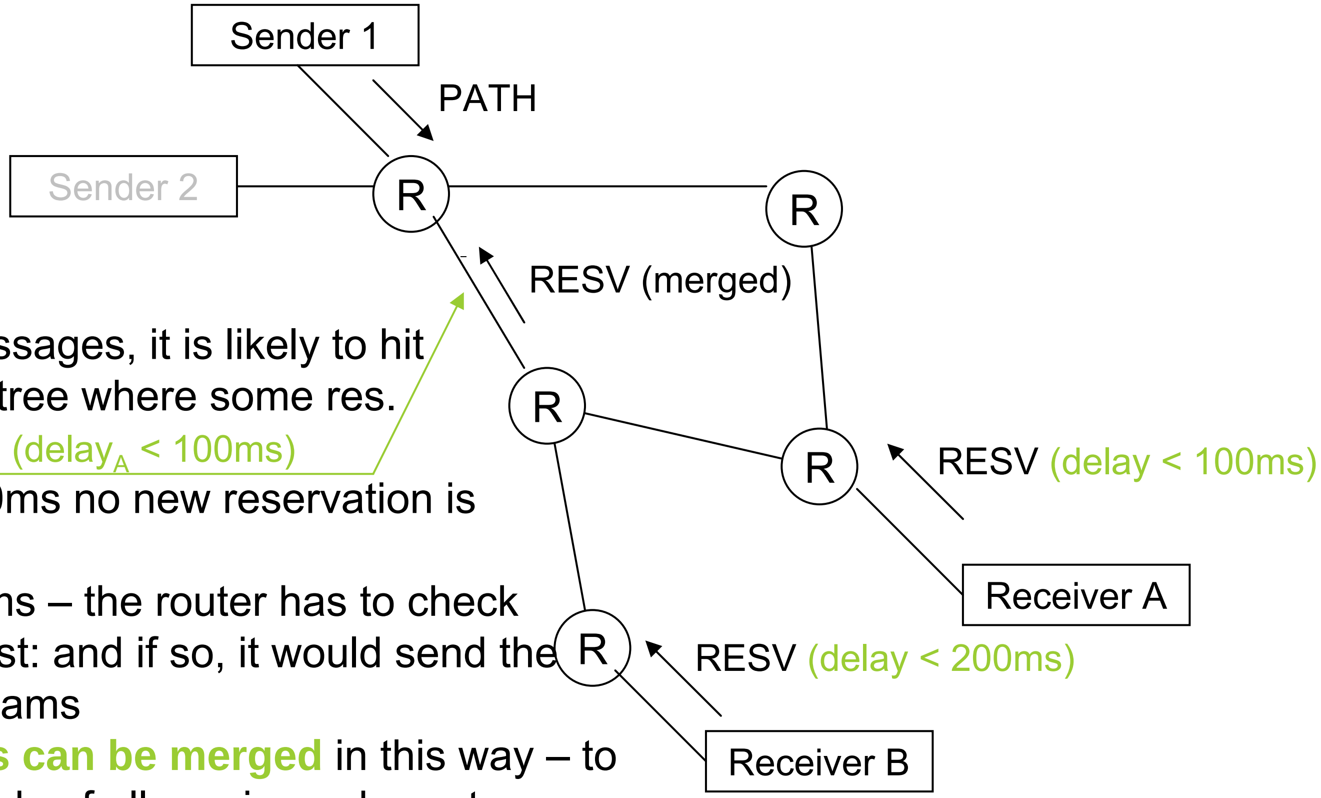
- One sender – one receiver



When a router or link fails:
•The *routing protocol* will *create a new path*
•**PATH** messages are sent all 30 sec.
(may be sooner, if a router detects a change in its forwarding table)
•The first one will reach the receiver following the new path
•The RESV message will follow the new path
• … and establish reservations on this new path
•Routers *–no longer on the path-* will stop getting RESV messages – reservation will be timed out and released **(soft state)**

Quality of Service (QoS)

- Single sender to multiple receivers



For RESV messages, it is likely to hit a piece of the tree where some res. already exists. (delay$_A$ < 100ms)

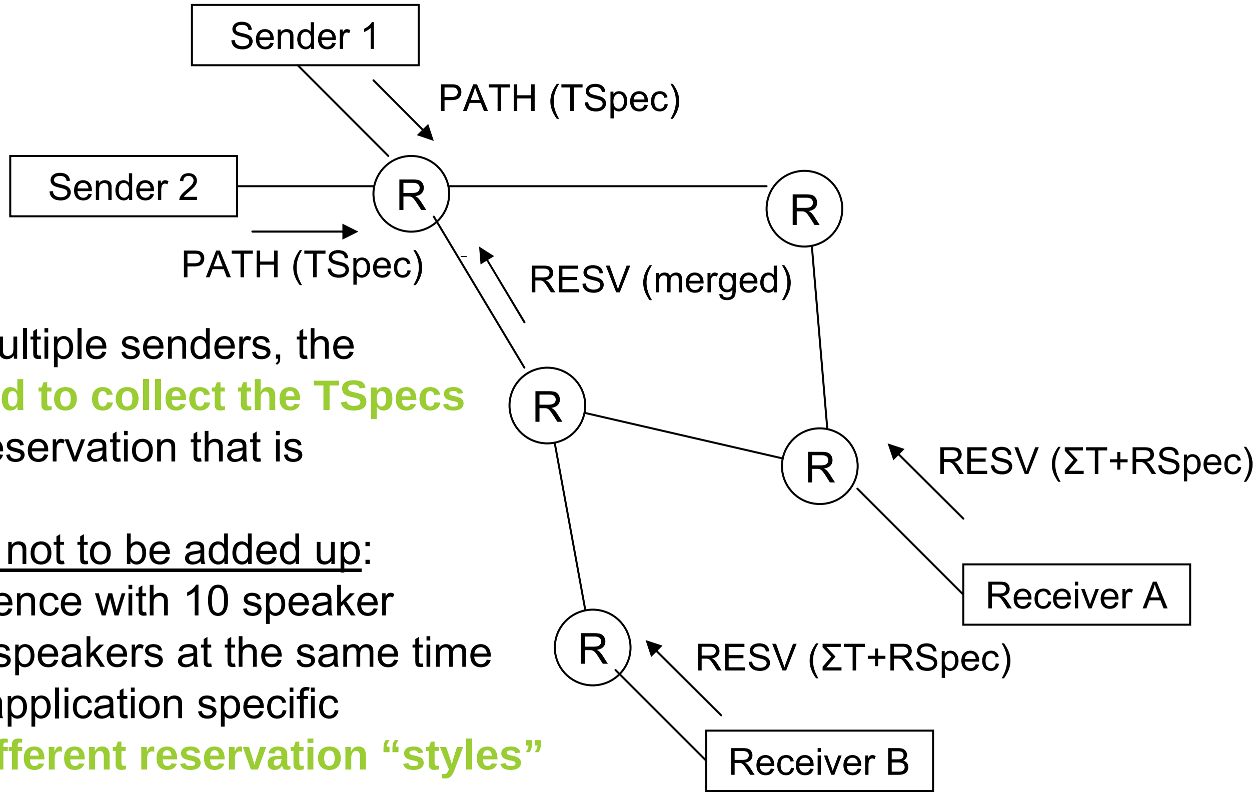
If delay$_B$ = 200ms no new reservation is Required

If delay$_B$ = 50ms – the router has to check this request first: and if so, it would send the request upstreams

- **Reservations can be merged** in this way – to meet the needs of all receivers downstream

Quality of Service (QoS)

- Multiple sender to multiple receivers



•If there are multiple senders, the **receivers need to collect the TSpecs** And make a reservation that is large enough

•TSpecs <u>need not to be added up</u>:
  Audio conference with 10 speaker
  … but just 2 speakers at the same time

•TSpec's are application specific

•RSVP has different reservation "styles"

Quality of Service (QoS)

- Packet scheduling
  1. Associate each packet with the appropriate reservation (*classifying packets*)
     - Done by examining: source address, destination address, protocol number, source port, destination port.
     - … mapped to a class identifier: determining how the packet is handled
  2. Manage the packets in the queues (*packet scheduling*)

- For guaranteed services a weighted fair queue (one queue per flow) will provide a guaranteed end-to-end delay
- For controlled load, simpler schemes may be used

Quality of Service (QoS)

- Scalability Issues
  - Integrated services and RSVP represents a *significant enhancement of the best effort service model of IP*
  - … but one of the fundamental IP design-goals is not supported: **scalability**
  - In the "best-effort" service model, no flow specific states are stored at the routers
  - Thus, as the Internet grows, the only thing routers have to do is, to keep up with this links speeds.
  - Example: 64kbps audio channels on an OC-48 link (2.5Gbps):

$$2.5x10^9/64x10^3 = 39{,}000$$

  - Each of this reservations needs some memory
  - … and has to be refreshed periodically

Quality of Service (QoS)

- Scalability Issues
  - Routers have to classify, police, and queue each flow
  - Admission control decisions have to be made
  - "Push-back" mechanisms are required – *to prevent long term reservations*
  - The scalability problem has prevented the widespread of "Integrated Services"
  - Other approaches do not require a per-flow state
  - The next section discusses such approaches:
    - Differentiated Services
    - ATM
    - …

Quality of Service (QoS)

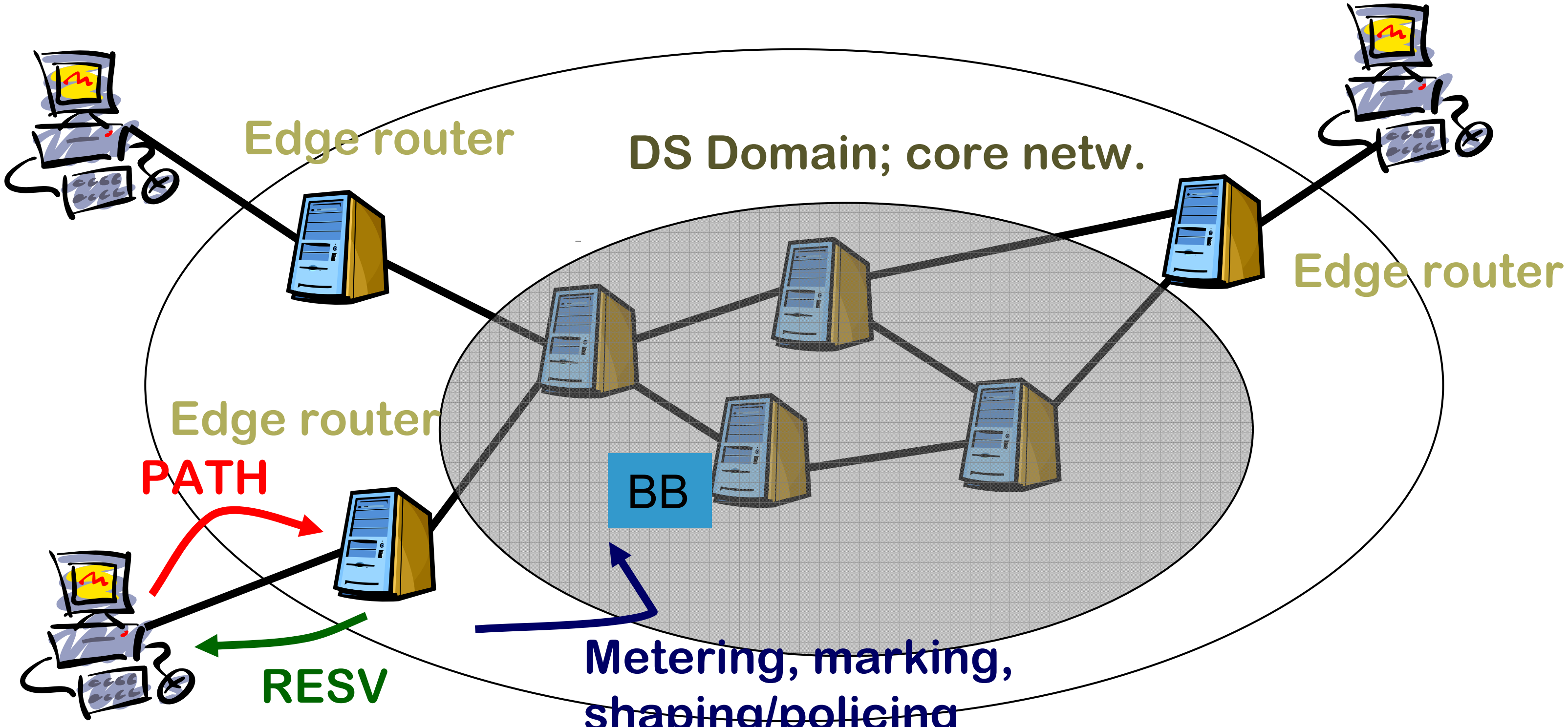
2

- **Differentiated Services**
  - The DiffServ model allocates resources to a small number of classes of traffic.
  - Usually just two classes: Premium, Regular
  - We are using RSVP to tell the routers that some flow is sending "*premium*" packets.
  - *It would be much easier, if the packet identifies themselves at the routers*
    - Can be done, using a single bit in the packet header
  - Two questions:
    - Who sets the "premium" bit, and under what circumstances?
      - Set the bit as an administrative boundary
    - What does a router differently when it sees a packet with a bit set?
      - Different router behaviors are specified (*by the DiffServe IETF WG*)

?

Quality of Service (QoS)

- **Differentiated Services (Schema)**



?

Quality of Service (QoS)

- **Differentiated Services (Schema)**
  - Step 1

    Source sends request message to first hop router
  - Step 2

    First hop router sends request to BB (bandwidth broker), which sends back either a accept or reject
  - Step 3

    If accept, either **source** or **first hop** router will mark DSCP (*differentiated services code point*) and start sending packets
  - Step 4

    Edge router checks compliance with SLA (*service level agreement*) and does policing. Excess packets are either discarded or marked as low priority to comply with the SLA
  - Step 5

    Core routers will just look at DSCP and decide **PHB**

Quality of Service (QoS)

- Differentiated Services
  1. Router behaviors are called: "per-hop behaviors" (PHB's)
     - Indicates the behavior of individual routers *rather than* end-to-end services
     - … more than 1 bit is required
     - The old TOS byte of the IP header was taken:
       - Six bits of this byte have been allocated for the "DiffServ Code Points" (DSCP) identifying a particular PHB (applied to a packet).
       - The simplest PHB: "*expedited forwarding*" (EF) → should be forwarded with *minimal delay and loss*

         express service
       - EF-arrival rate is limited by the link speed of the router
       - EF can be implemented by:

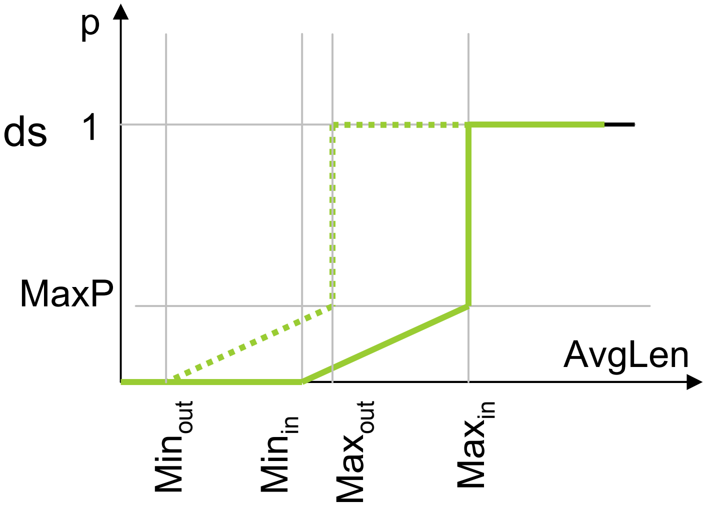
         a.) strict priority over all other packets

         b.) weighted fair queuing → *can prevent routing packets from being locked out*

Quality of Service (QoS)

- Differentiated Services
  2. Another PHB: "*assured forwarding*" (AF) → has its roots in "RED (*random early detection*) with in and out" (RIO) or "weighted RED"

"out" has lower tresholds



- Two classes of traffic → two drop probability curves (in, out)
- In- or out- of the profile (guaranteed by the edge router – service provider)

Quality of Service (QoS)

- Differentiated Services
  - There must be enough bandwidth so that "in"-packets alone are rarely able to congest the link (RIO starts dropping "in"-packets)
  - RIO does not misorder "in"- and "out"-packets → "*fast retransmit*" can be falsely triggered
  - RIO can be generalized to more than two drop probabilities
    - *DSCP field is used to pick one drop probability curve*

3. DSCP can also be used to determine a queue to put a packet into a "*weighted fair queuing*" scheduler (WFQ).
   - One code point to indicate "*best-effort*" services (queues)
   - A second to select the "*premium queue*" (higher weighted than best-effort)

Example: $B_{premium} = W_{premium} / (W_{premium} + W_{best\text{-}effort})$ … *premium bandwidth*
$= 1/(1 + 4) = 0.2$ … 20% *reserved for premium traffic*

Quality of Service (QoS)

- Differentiated Services
  - The premium class can be kept low, since WFQ will try to transmit premium packets as soon as possible.
  - If premium load ~10% → behaves as if premium traffic is running on an under-loaded network
  - If premium load ~30% → behaves like a highly loaded network
  - Just as in WRED, we can generalize this WFQ-based approach to allow more than two classes represented by different code points
  - We can also combine the idea of a queue selector with a drop preference

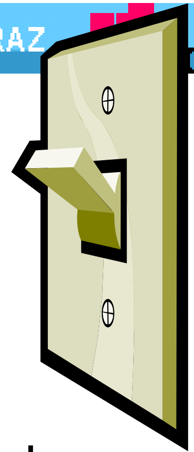
Quality of Service (QoS)

- Differentiated Services
  - Advantages:
    - QoS++
    - Because DS specifies QoS at Layer 3 it can run without modification on any Layer 2 infrastructure that supports IP
    - No need for per flow state and signaling at every hop
    - Core network is (can be) an optimized network on routing
      - e.g. MPLS
    - Avoid complexity and maintenance of per-flow state information in core nodes
    - Push unavoidable complexity to the network edges
    - Capabilities can be added via incremental firmware or software upgrades.

Quality of Service (QoS)

- Comparison of IntServ & DiffServ Architectures

	IntServ (access network)	DiffServ (core network)
Coordination for service differentiation	End-to-End	Local (Per-Hop)
Scope of Service Differentiation	A Unicast or Multicast path	Anywhere in a Network or in specific paths
Scalabilty	Limited by the number of flows	Limited by the number of classes of service
Network Accounting	Based on flow characteristics and QoS requirement	Based on class usage
Network Management	Similar to Circuit Switching networks	Similar to existing IP networks
Interdomain deployment	Multilateral Agreements	Bilateral Agreements

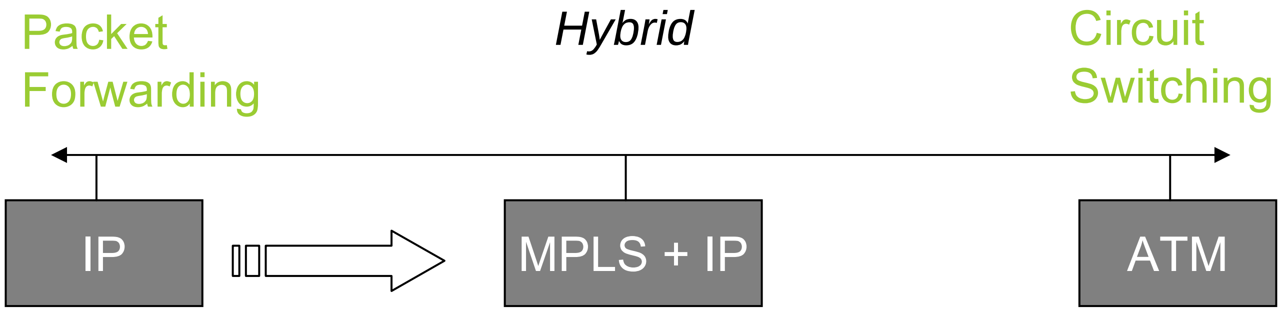


IP ATM and MPLS

- **IP**, the “only” protocol for global internet working
- … but there are other contenders: most notable **ATM**
- ATM could not displace other technologies in the local area network
- An interesting development in the relationship between IP and ATM is the appearance of **MPLS** (based on IP routing and Tag switching)
- “Merge” the forwarding algorithm used in ATM with IP control protocols
  - A Label switching router (LSR) forwards packets by looking at fixed length labels – using this label to find the output interface
  - LSR has to **rewrite the label** before it can send the packet
  - This is exactly how ATM switches are forwarding cells
    - Significantly simpler than longest-match
    - LSRs will be cheaper at a given performance (than conventional router)
    - Forwarding decisions can be based on more complex criteria
    - Supporting VPs to be merged into one single VP

IP ATM and MPLS

- MPLS + IP form a middle ground that combines the best of IP and the best of circuit switching technologies.
- ATM and Frame Relay cannot easily come to the middle so IP has!!



Problems of IP Routing

- **Hop-to-hop routing**
  - In connectionless IP, each router has to make independent forwarding decisions based on the IP header (32 bit v4, 128 bit v6)
  - But the header contains much more information than the router needs to find the next hop
- **Basically there are two routing functions:**
  - dividing the whole address space in *forward equivalence classes* (FEC)
    - e.g. *longest address prefix match*
  - allocate the FEC to the next hop

MPLS

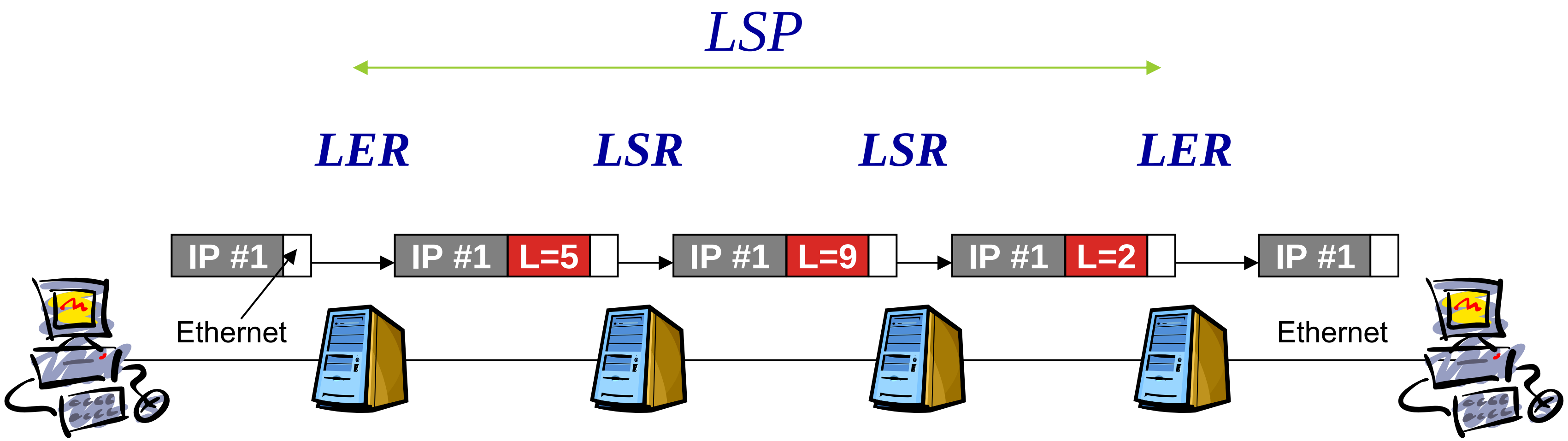
3 MPLS performs the following functions:

- specifies mechanisms to manage traffic flows of various granularities, such as flows between different hardware, machines.
- remains independent of the Layer-2 and Layer-3 protocols
- provides a mechanism to map IP addresses to simple, fixed-length labels
- interfaces to existing routing protocols (RSVP, OSPF)
- supports the IP, ATM, and frame-relay Layer-2 protocols

MPLS

- data transmission occurs on label-switched paths (LSPs).
- The labels are distributed using label distribution protocol (LDP) or RSVP or piggybacked on routing protocols like border gateway protocol (BGP) and OSPF.
- Each data packet encapsulates and carries the labels during their journey from source to destination
- Hardware can be used to switch packets quickly between links.
- LSRs and LERs
  - An **LSR** is a high-speed router device in the core of an MPLS network
  - An **LER** is a device that operates at the edge of the access network and MPLS network - supports multiple ports connected to dissimilar networks (such as frame relay, ATM, and Ethernet)

Example



IP Addr	Out Label
192.4/16	5
Layer 2 Transport	**Assign init label**

In Label	Out Label
5	9
Label Swapping	

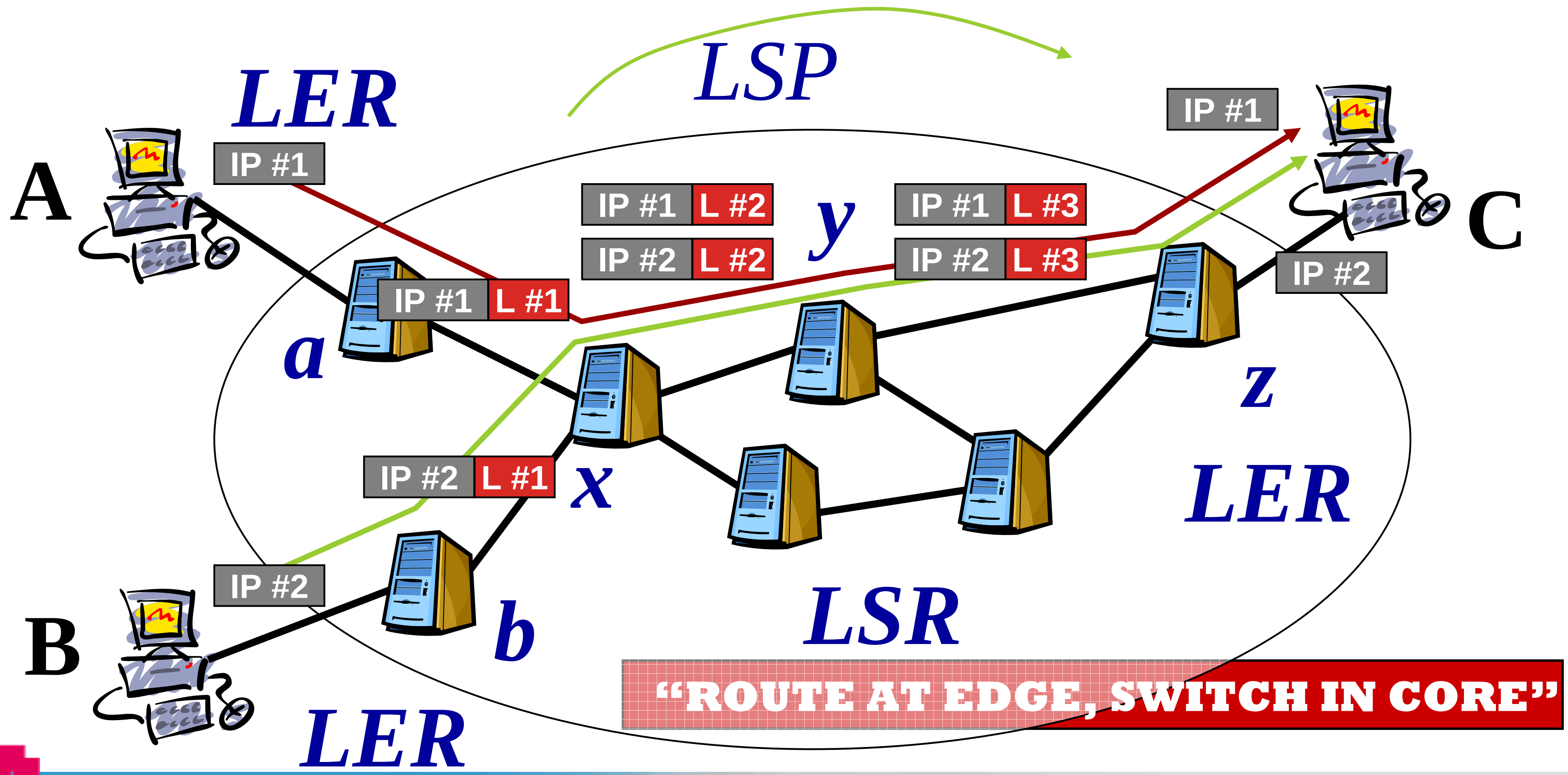
In Label	Out Label
9	2
Label Swapping	

In Label	Next Hop
2	192.4/16
Remove Label	Layer 2 Transport

“ROUTE AT EDGE, SWITCH IN CORE”

Example

- Communication A => C resp. B => C
  - mapped onto the same FEC at x, y, z

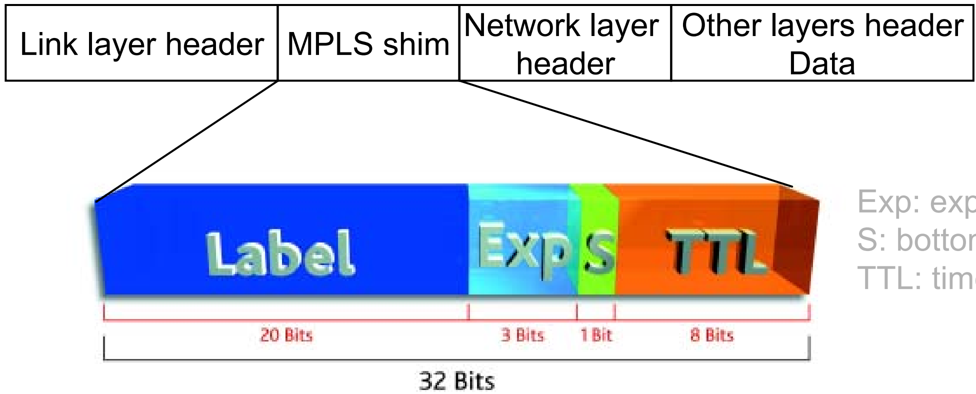


MPLS

- **FEC**
  - The forward equivalence class (FEC) is a representation of a group of packets that share the same requirements for their transport.
  - the assignment of a particular packet to a particular FEC is done just once, as the packet enters the network
- **Labels and Label Bindings**
  - A label is carried or encapsulated in a Layer-2 header along with the packet
  - The label values are derived from the underlying data link layer.
  - The receiving router examines the packet for its label content to determine the next hop
  - Once a packet has been labelled, the rest of the journey of the packet through the backbone is based on label switching.
  - The label values are of local significance

MPLS

- Label assignment decisions may be based on forwarding criteria such as the following:
  - destination unicast routing
  - “traffic engineering”
  - multicast
  - virtual private network (VPN)
  - QoS
- Generic label format:

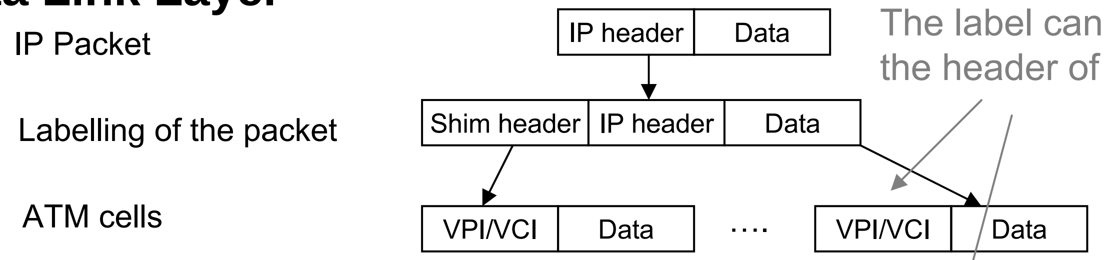


Exp: experimental
S: bottom of stack (first label in stack)
TTL: time to live

Source: http://www.spirentcom.com/documents/654.pdf

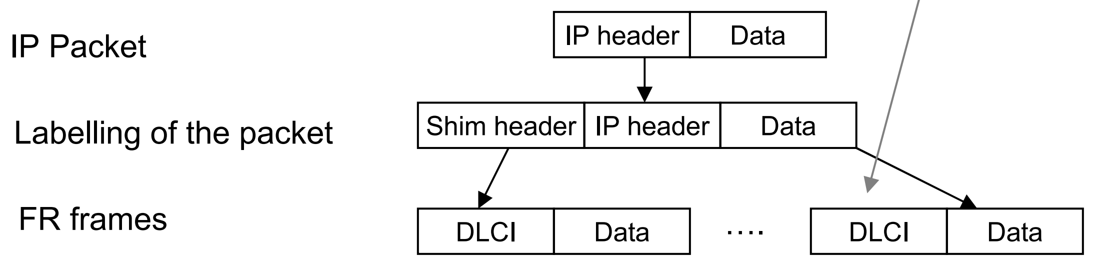
MPLS

ATM as the Data Link Layer

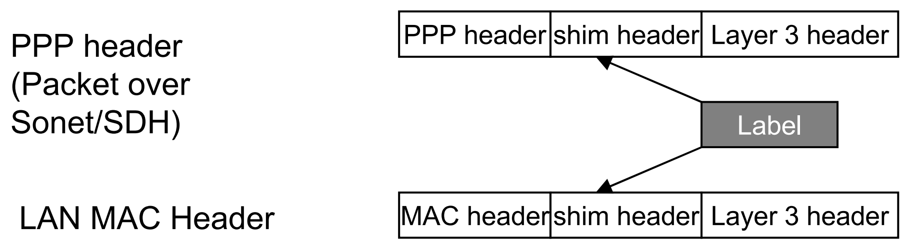


The label can be embedded in the header of the data link layer…

Frame Relay as the Data Link Layer



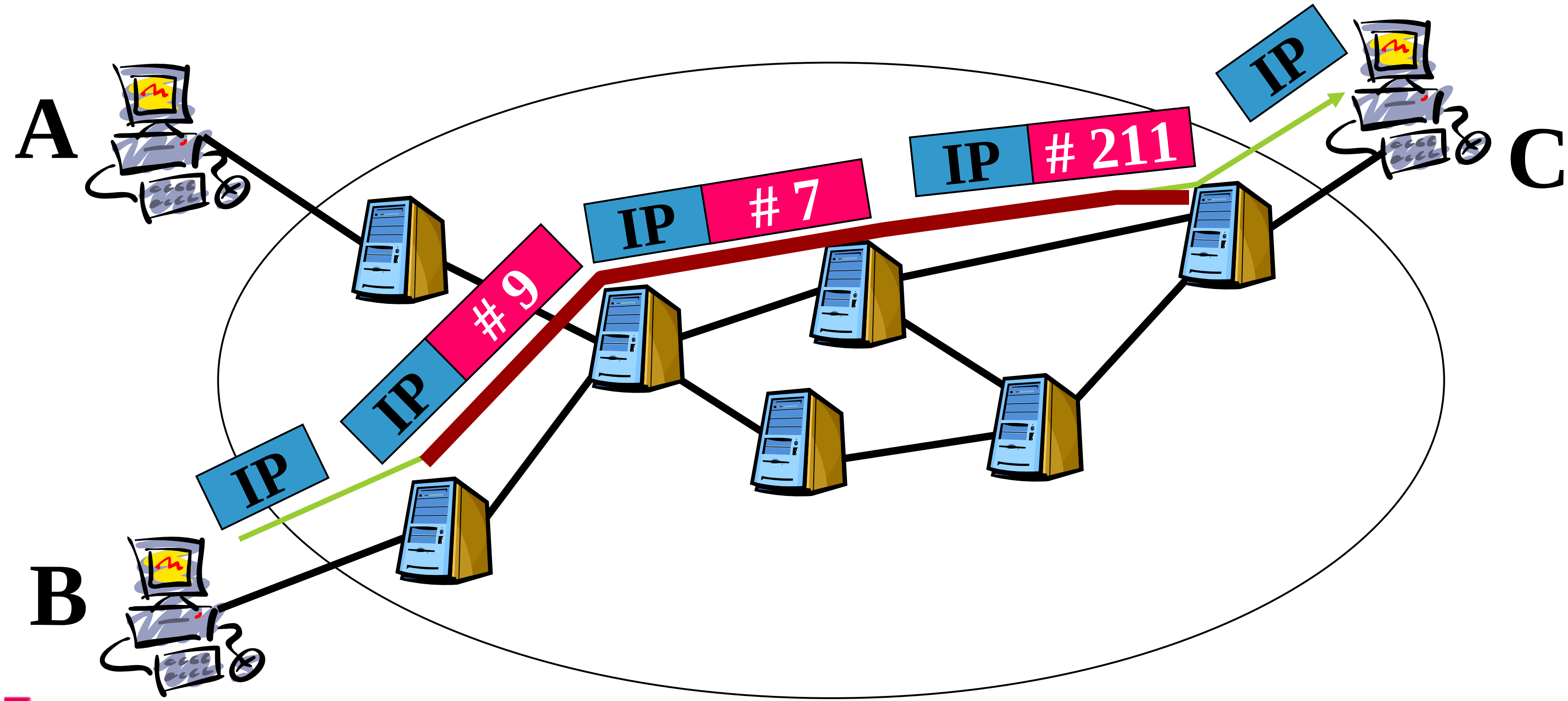
Point-to-Point (PPP)/Ethernet as the Data Link Layer



… or in the shim (between the Layer-2 data-link header and Layer-3 network layer header,

Basic Idea (II)

- “route at edge”, and “switch in core“



MPLS

- **Label distribution:**
  - MPLS does not mandate a single method of signaling for label distribution
  - BGP has been enhanced to piggyback the label information within the contents of the protocol
  - RSVP has also been extended to support piggybacked exchange of labels.
  - IETF has also defined a new protocol known as the label distribution protocol (LDP) for explicit signaling and management
  - Extensions to the base LDP protocol have also been defined to support explicit routing based on QoS requirements.

MPLS

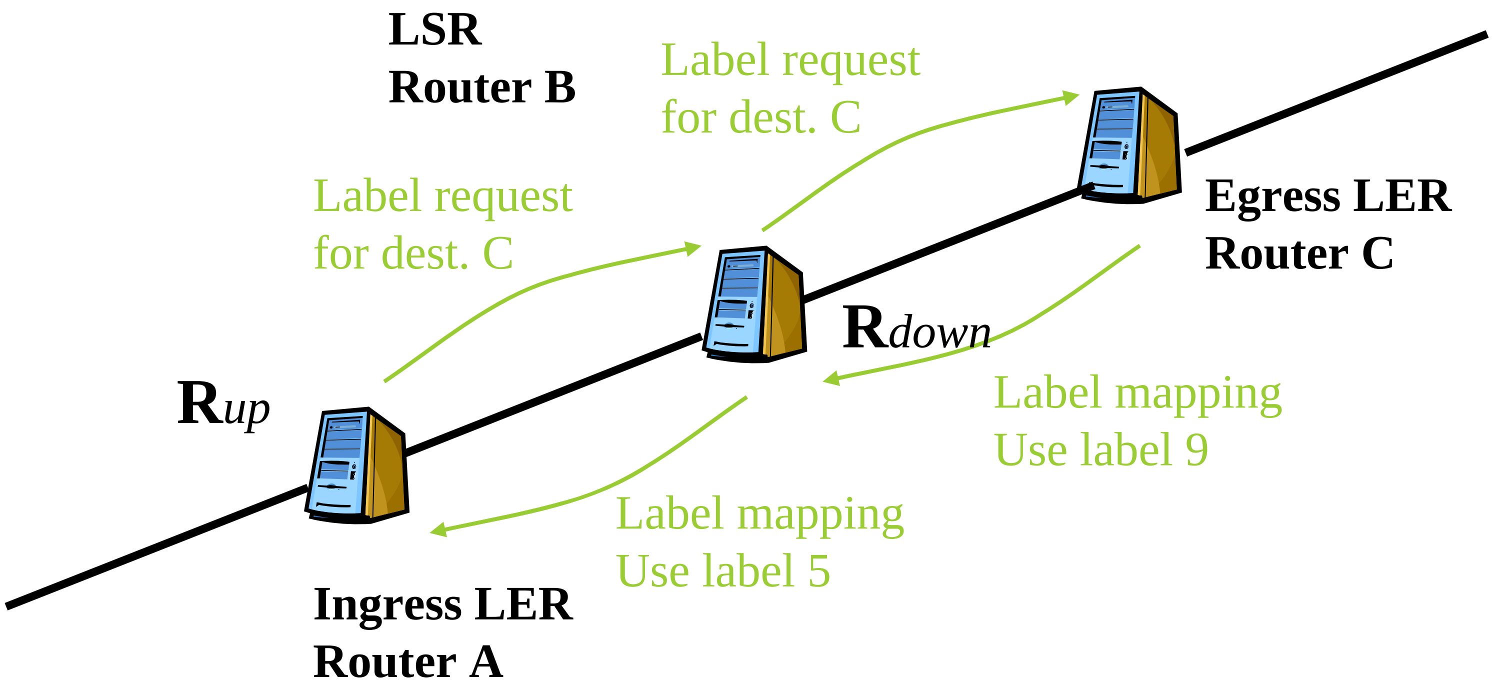
- **Label-Switched Paths (LSPs)**:
  - A collection of MPLS—enabled devices represents an MPLS domain. Within this domain, a path is set up for a given packet to travel on an FEC. The LSP is set up prior to data transmission. MPLS provides the following two options to set up an LSP:
    - **hop-by-hop routing:** Each LSR independently selects the next hop for a given FEC. The LSR uses any available routing protocols, such as OSPF, ATM private network-to-network interface (PNNI), etc.
    - **explicit routing:** is similar to source routing. The ingress LSR (i.e., the LSR where the data flow to the network first starts) specifies the list of nodes through which the ER–LSP traverses.

  The LSP setup for an FEC is unidirectional in nature. The return traffic must take another LSP.

MPLS

- **Label Merging:**
  - Incoming traffic from different interfaces can be merged together and switched using a common label if they are traversing the network toward the same destination (*stream merging* or *aggregation of flows*)
  - If the underlying transport network is ATM, LSRs could employ VP- or VC-merging. (avoid cell interleaving problems)
- **Signaling Mechanisms**
  - **label request:** A LSR requests a label from its downstream neighbor so that it can bind to a specific FEC. This mechanism can be employed down the chain of LSRs up to the egress LER
  - **label mapping**: In response to a label request, a downstream LSR will send a label to the upstream initiator using the label mapping mechanism

Upstream vs. downstream



MPLS

- **Label Distribution Protocol (LDP)**
  - The LDP is a new protocol for the distribution of label binding information to LSRs in an MPLS network. It is used to map FECs to labels, which, in turn, create LSPs. LDP sessions are established between LDP-peers in the MPLS network (not necessarily adjacent). The peers exchange the following types of LDP-messages:
    - discovery messages— announce and maintain the presence of a LSR in a network
    - session messages— establish, maintain, and terminate sessions between LDP peers
    - advertisement messages— create, change, and delete label mappings for FECs
    - notification messages— provide advisory information and signal error information

MPLS Applications

- Enabling IP over ATM
  - The LER devices are responsible for IP flow classification and label imposition.
  - The LSR devices located in the core are responsible for forwarding at Layer 2 while participating in the exchange of Layer 3 routing information.
- Traffic Engineering
  - Best effort delivery is not (always) sufficient.
  - MPLS provides for explicit routing
- Class of Services
  - The head end LSR could place high-priority traffic in one LSP, medium-priority traffic in another LSP, best-effort traffic in a third LSP, and less-than-best-effort traffic in a fourth LSP.
- VPN´s
  - Each ingress LSR places traffic into LSPs based on a combination of a packet's destination address and VPN membership information