QUALITY OF SERVICE

Static Content Applications

- Web, file transfer, and electronic mail
- Text and images.
- Sent from one host to another, desirable for to arrive at the destination as soon as possible.
- Long end-to-end delays, up to tens of seconds, are often tolerated.

Multimedia Networking Applications

- highly sensitive to delay;
- some are loss tolerant;
  - occasional loss only causes occasional glitches in the audio/video playback,
  - losses can be partially or fully concealed.

Multimedia applications features

- Streaming: e.g. playing back audio or video while the file is being received.
- user interactivity: e.g. pause/resume and temporal jumps to the future and past of an audio/video file
- real-time: e.g. Internet radio stations.

Internet Multimedia Today

- multimedia over the Internet has achieved significant but limited success
- main reason: IP provides a best-effort service to all the datagrams it carries -
- Delay and Packet jitter (the variability of packet delays within the same packet stream) -> deteriorates quality.

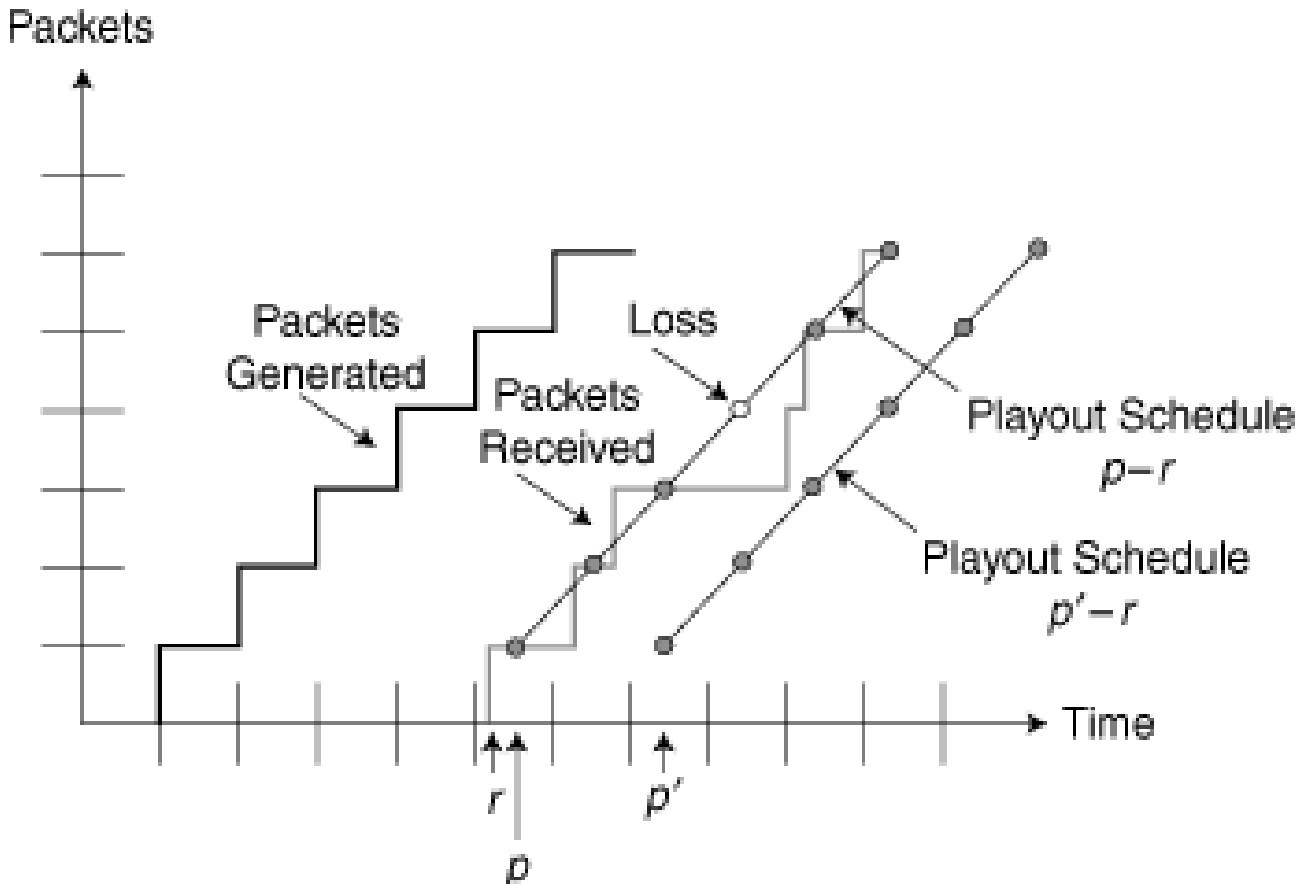
Solutions

- Two extremes:
  - No fundamental changes - just more bandwidth
  - Fundamental changes to the network -No more best-effort.
- Differentiated services - between the two extremes - small changes at the network and transport layers.

Practical Solutions Today

- Delay playback at the receiver by 100 msecs or more $\rightarrow$ diminish the effects of network-induced jitter.
- Timestamp packets at the sender -> the receiver knows when the packets should be played back.
- Prefetch data during playback when client storage and extra bandwidth is available.
- Send redundant information -> mitigate the effects of network-induced packet loss.

Playback delay and packet loss



Adaptive adjustment of playout delays algorithm

- t_i = timestamp of the ith packet = the time packet was generated by sender
- r_i = the time packet i is received by receiver
- p_i = the time packet i is played at receiver

- End-to-end network delay of the ith packet = $r_i - t_i$.
- d_i = estimate of average network delay upon reception of the ith packet.

$$d_i = (1 - u)\, d_i - 1 + u\,(r_i - t_i)$$

u = fixed constant (e.g. 0.01)

- v_i = estimate of average deviation of the delay from d_i.

$$v_i = (1 - u)\, v_i - 1 + u \mid r_i - t_i - d_i \mid$$

- p_i = the playout time: $p_i = t_i + d_i + Kv_i$

K = fixed constant (e.g. 4)

Loss recovery schemes

- Loss= packet never arrived at the receiver *or* arrived after its scheduled playout time.
- Retransmission - not a good solution
- Loss-anticipation schemes:
  - Forward error correction (FEC)
  - Interleaving.

Forward Error Correction (FEC)

- Add redundant information to the original packet stream
- Use redundant information to reconstruct "approximations" or exact versions of some of the lost packets

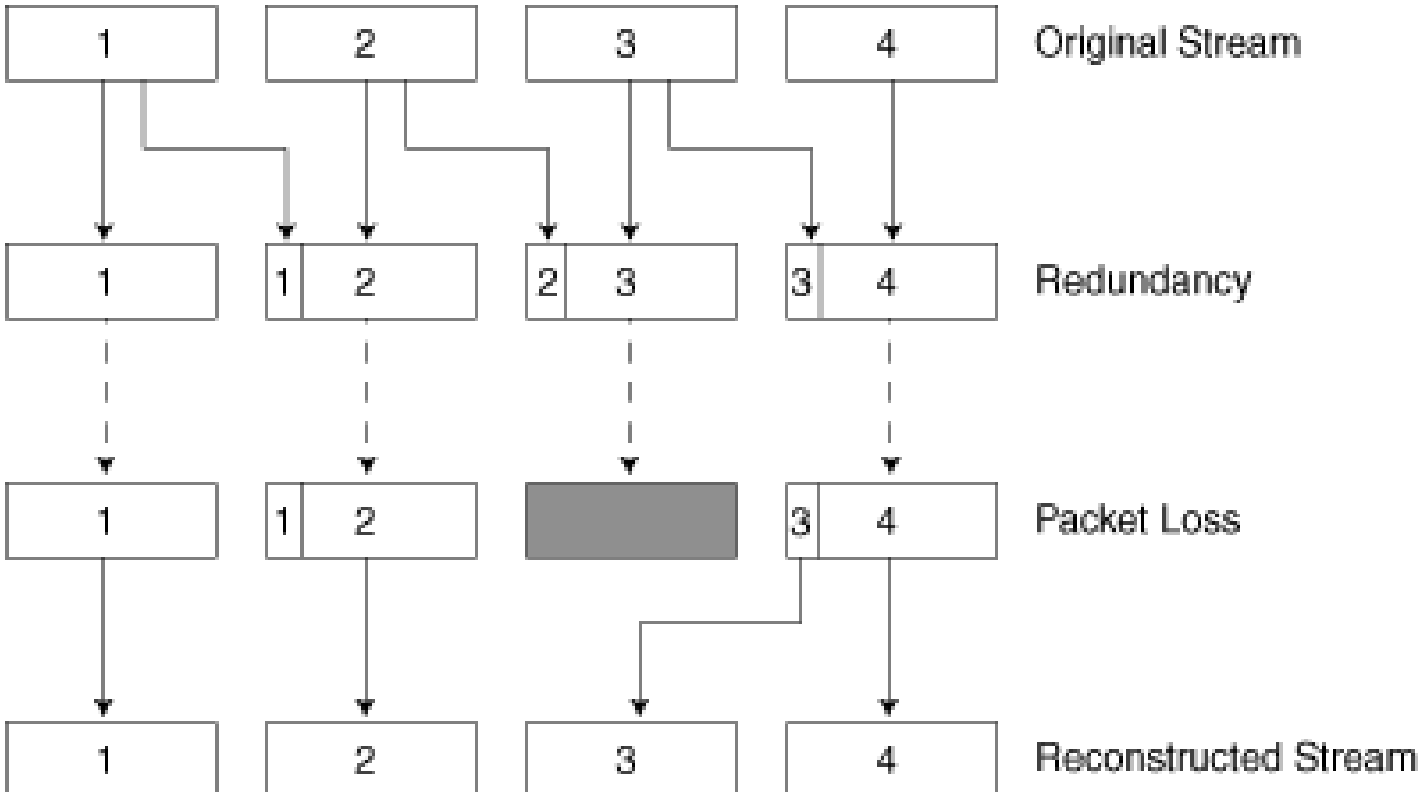
1. Send a redundant encoded chunk after every n chunks.

(e.g. by exclusive OR-ing the n original chunks).

2. send a lower quality audio stream as the redundant information (e.g. a low-bit rate audio stream).

→Loss is concealed loss by playing out the redundant chunk that arrives with the subsequent packet.

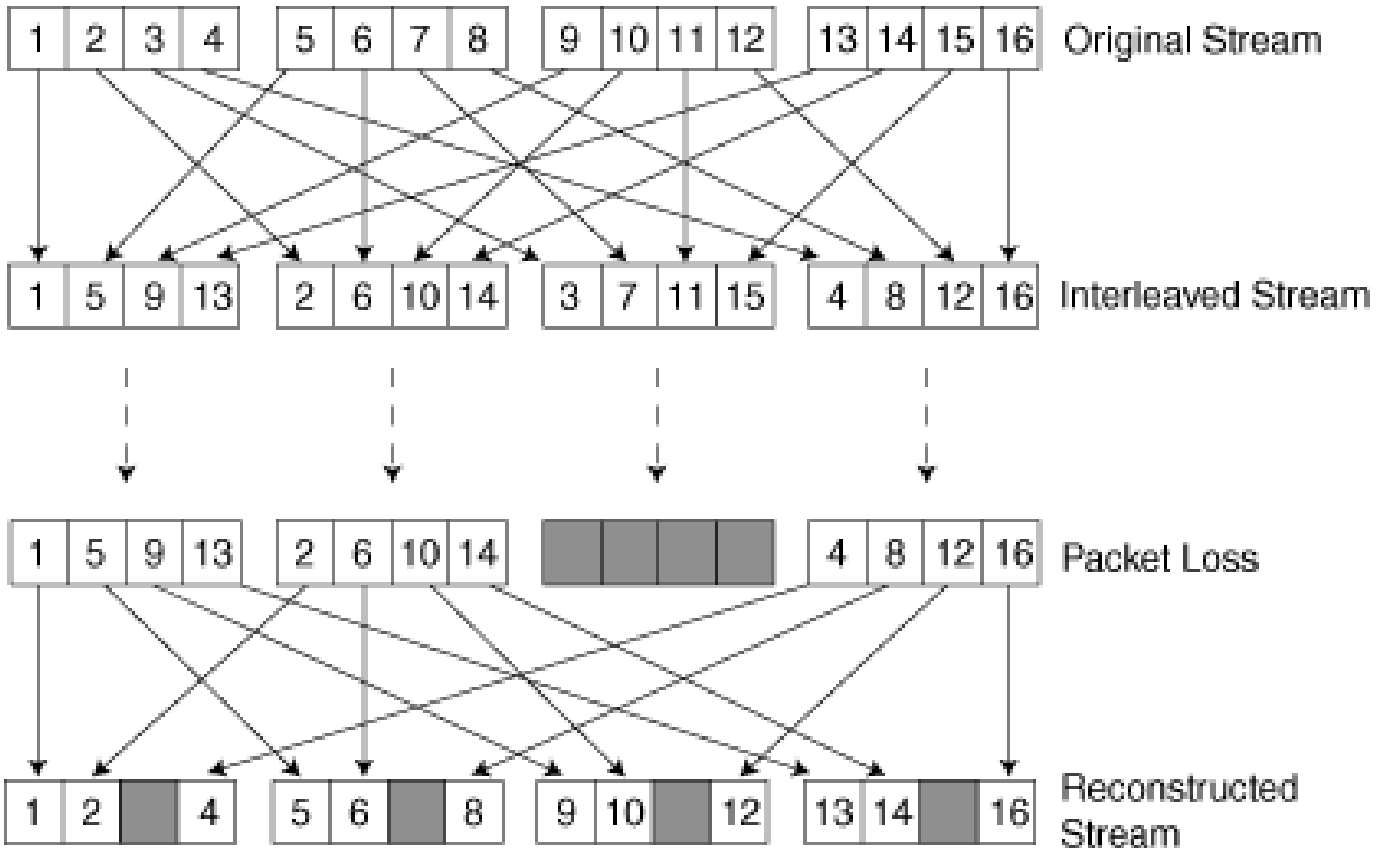
Piggybacking lower-quality redundant information



Interleaving

- Resequence units of data before transmission - originally adjacent units are separated by a certain distance in the transmitted stream.
- Loss of a single packet from an interleaved stream → multiple small gaps in the reconstructed stream
- disadvantage: interleaving increases latency → limits its use for interactive applications
- advantage: interleaving does not increase the bandwidth requirements of a stream

Sending interleaved audio



Receiver-based recovery

1. Packet repetition : replace lost packets with copies of the packets that arrived mediately before the loss.

→ low computational complexity + reasonable performance

2. Interpolation : use packets before and after the loss to interpolate a suitable packet to cover the loss.

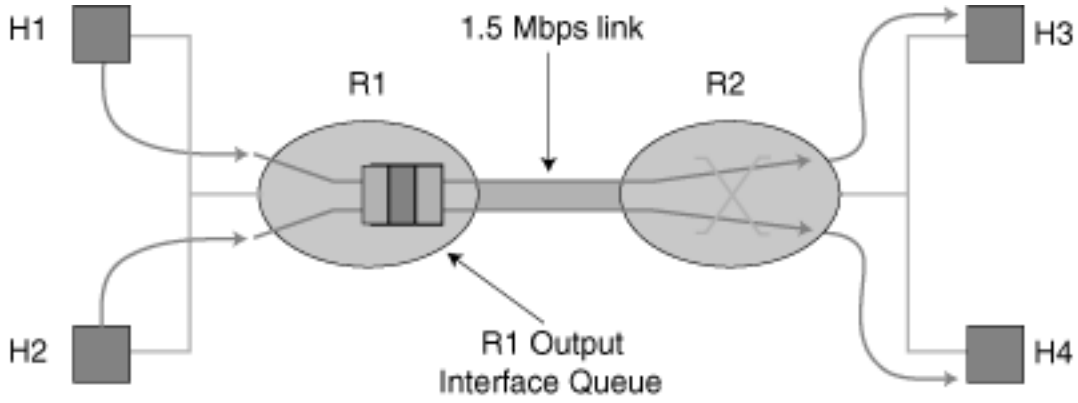
→ better performance, but more computationally intensive

RTP

- standard packet structure : includes fields for audio/video data, sequence number, timestamp, and more.
- used for transporting common formats (WAV, GSM, MPEG1, MPEG2)
- runs on top of UDP: code in the application encapsulates chunks of data in RTP packets, and sends RTP packets to a UDP socket interface, where it is encapsulated in a UDP packet.
- RTP encapsulation is only seen at the end systems--it is not seen by intermediate routers

New Solutions

- Explicit support for the QoS needs of multimedia applications

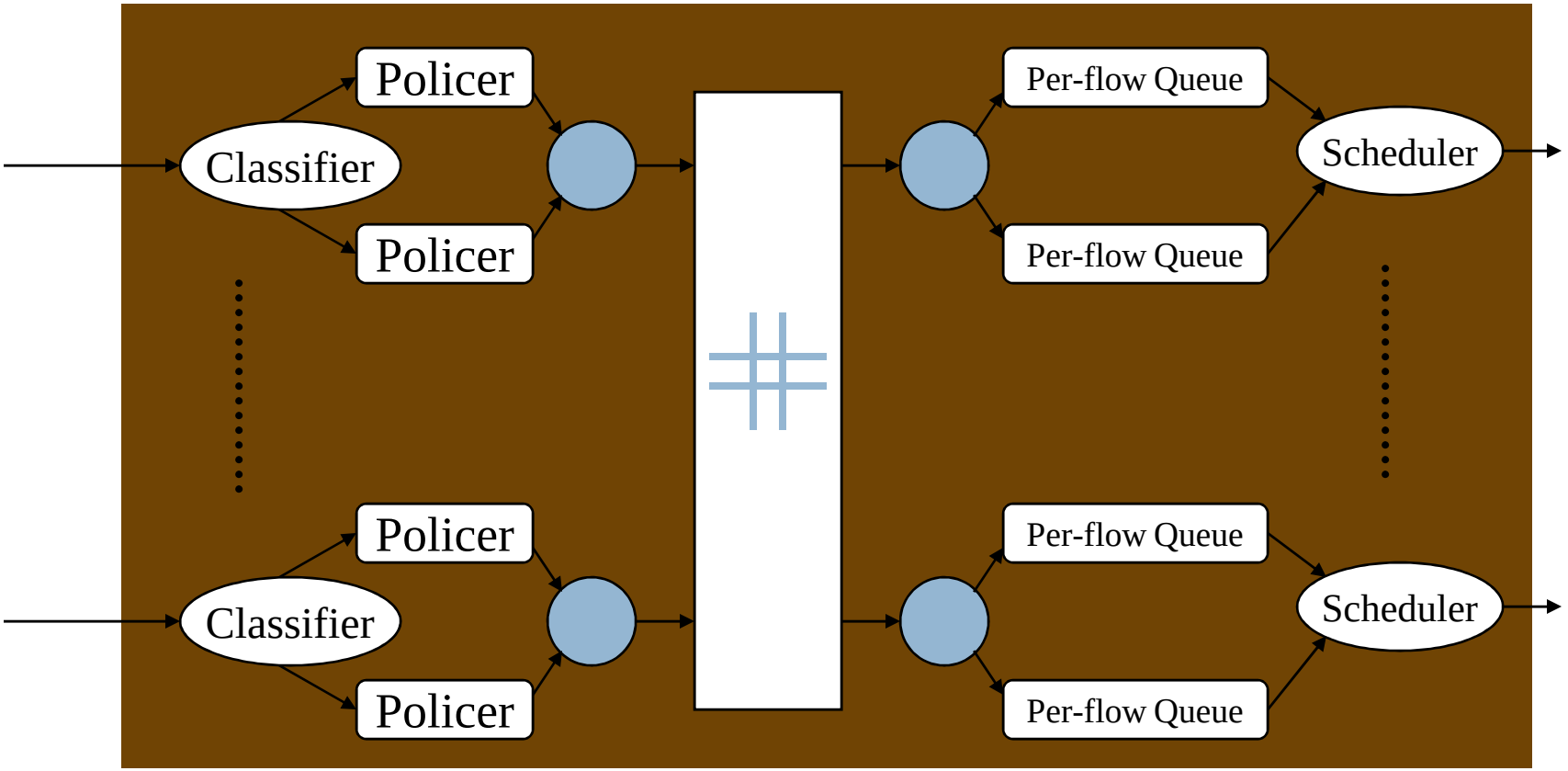


Problem: sending rate of H1 and H2 exceeds 1.5 Mbps and they need different QoS

Support for QOS

1. Packet classification and marking : allows a router to distinguish among packets belonging to different classes of traffic.
2. degree of isolation among traffic flows: one flow will not be adversely affected by another misbehaving flow.
3. Efficient use of resources (e.g. link bandwidth and buffers)
4. Call admission process: flows declare their QoS requirements and are either admitted to the network or blocked from it.

The QoS Router

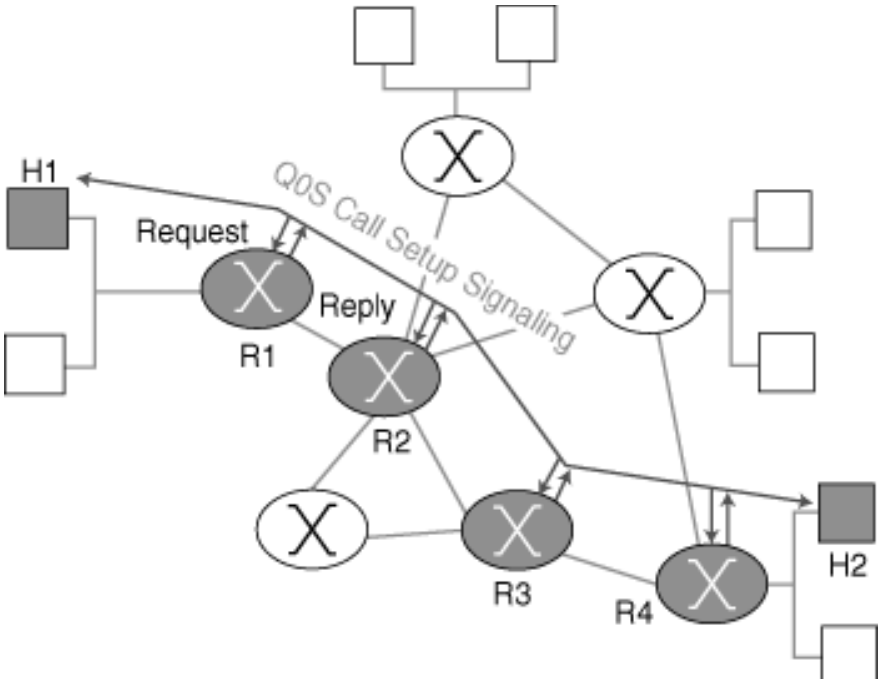


Integrated Services Principals

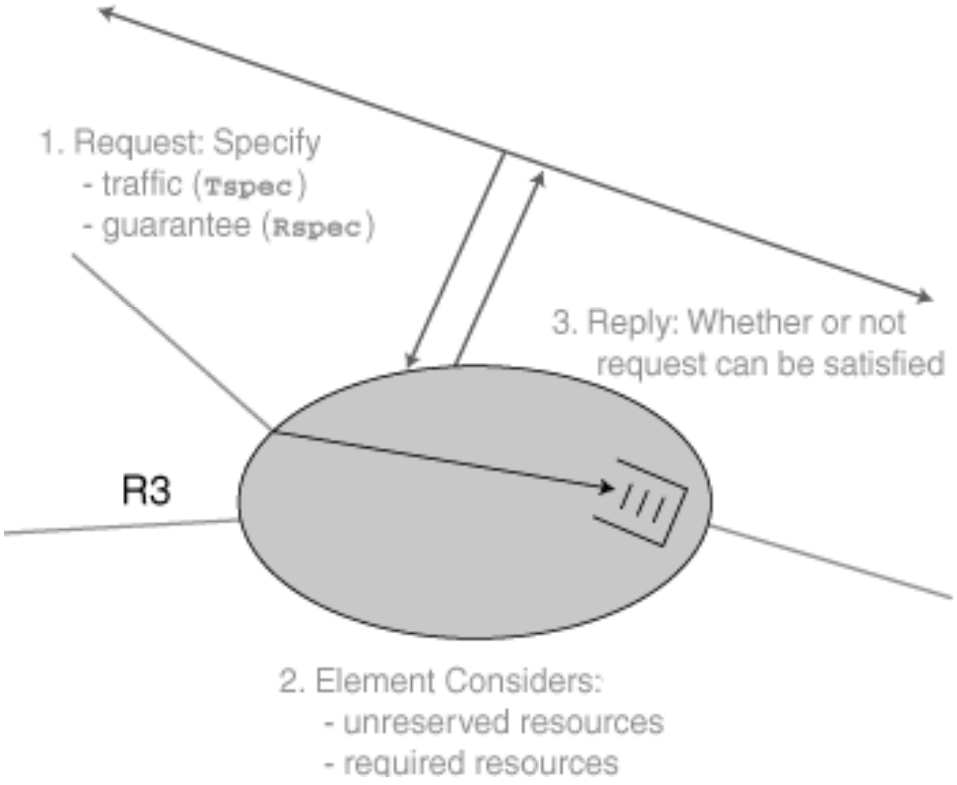
- Provides individualized quality of service guarantees to individual application sessions.
- Reserved Resources: A router should know what amounts of its resources (buffers, link bandwidth) are already reserved for on-going sessions.
- Call Setup: participation of each router on the path to ensure that its end-to-end QoS requirement.

Call setup process

1. session must first declare its QoS requirement, and characterize its traffic (Rspec and Tspec).
2. signaling protocol carries Rspec and Tspec to routers.



3. Router determines if it can admit the call



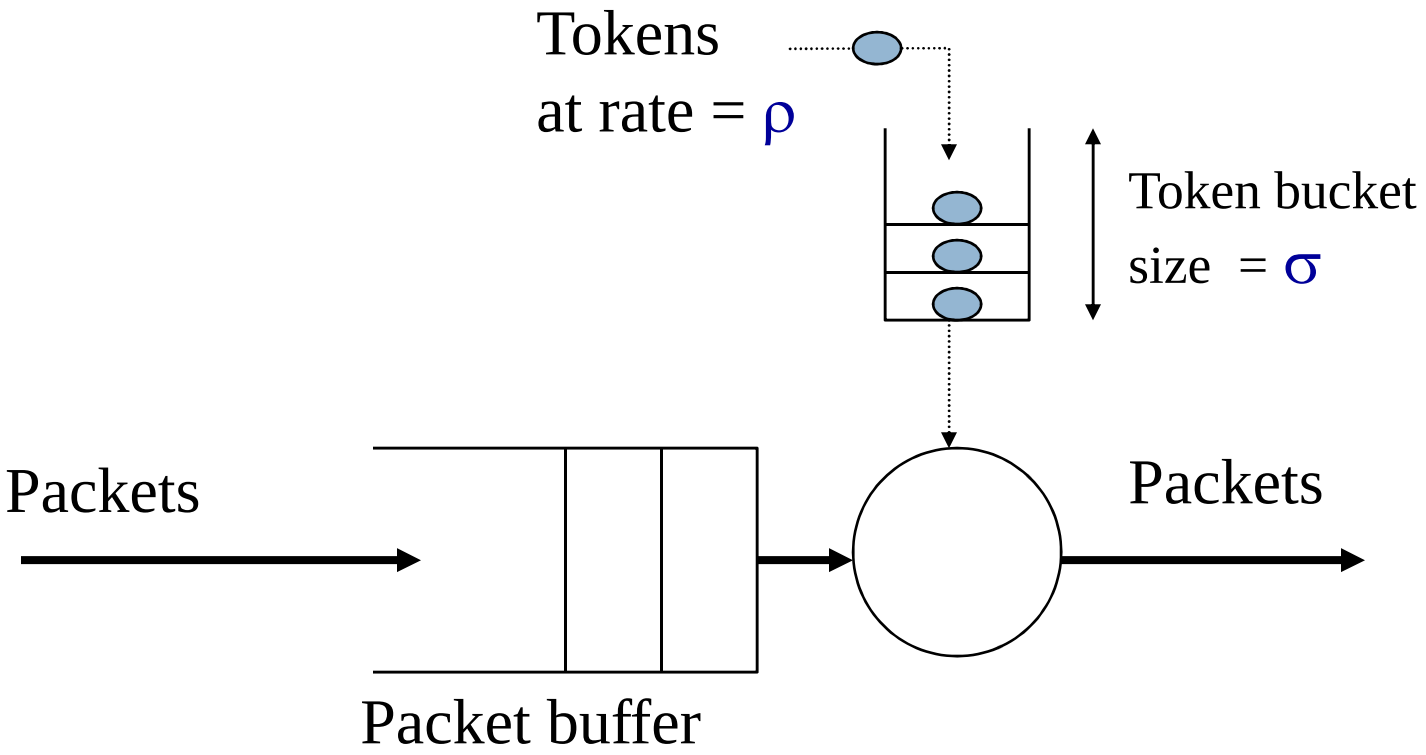
Two major classes of service

- Guaranteed service : provides (mathematically proven) bounds on queuing delay in routers

[forwarding rate of R bits/sec is guaranteed and use of (ρ,σ) model - amount of traffic (in bits) generated over any interval of length t is bounded by $\rho t+\sigma$ - :
queuing delay is bounded by σ/R ; if $\rho<R$]

- Controlled-Load service : provides a QoS like the one in an unloaded network element.

The (σ,ρ) traffic regulator

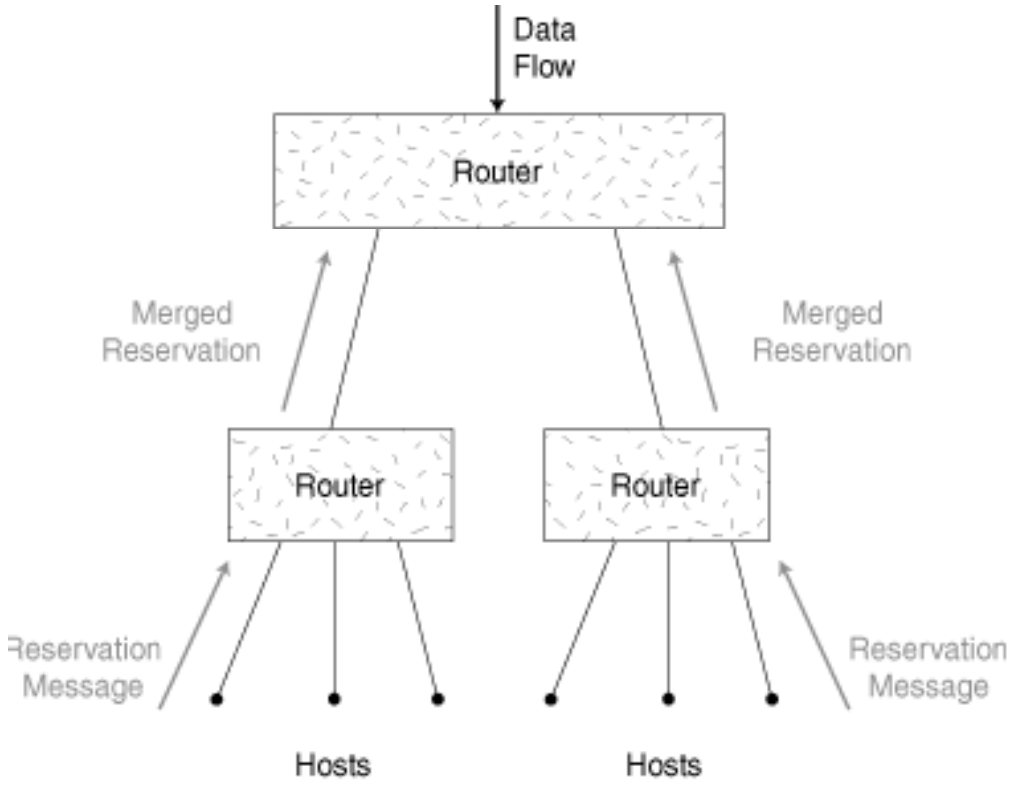


Number of bytes that can arrive in any period of length *t* is bounded by:

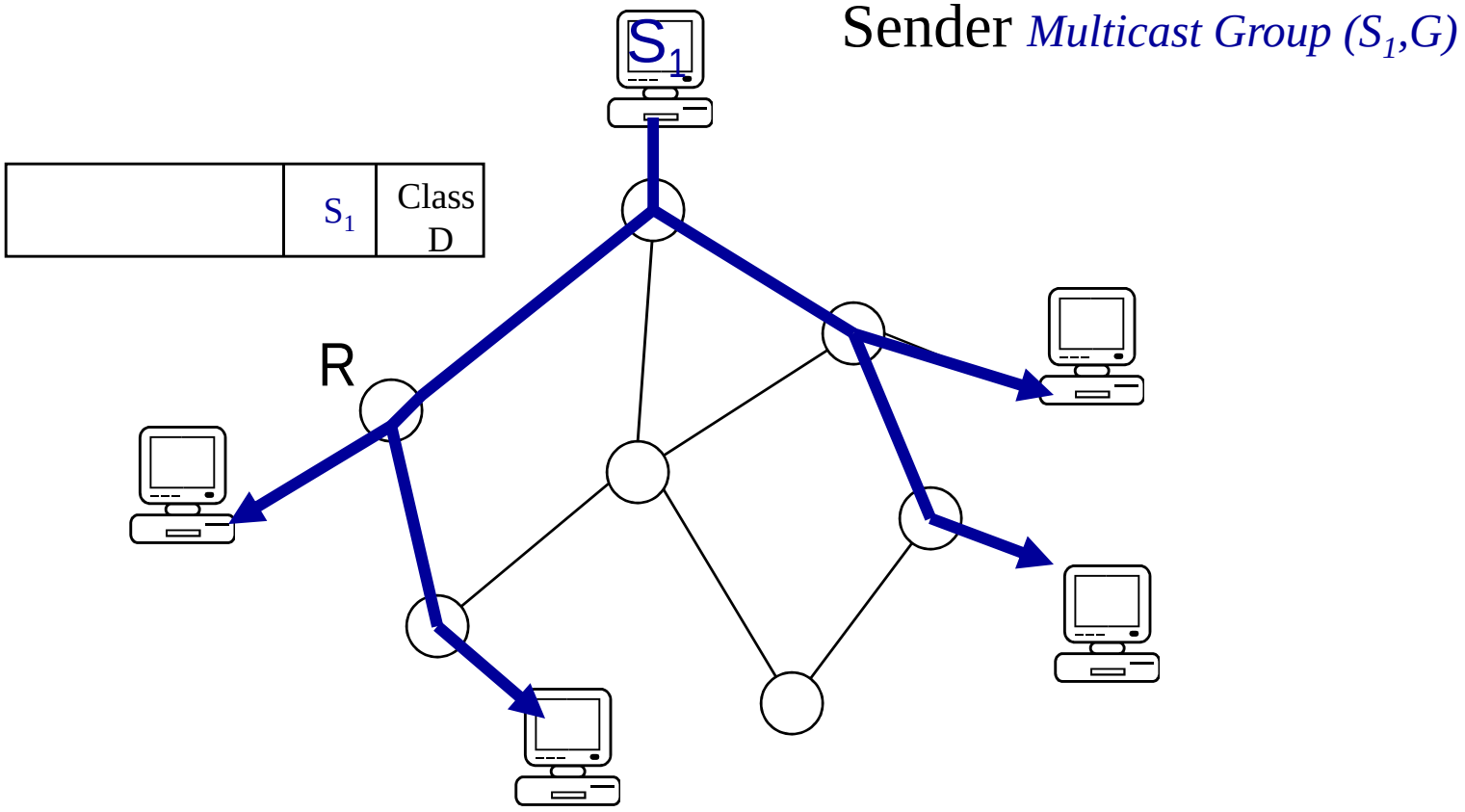
$$\sigma + \rho t$$

Bandwidth Reservation Protocol

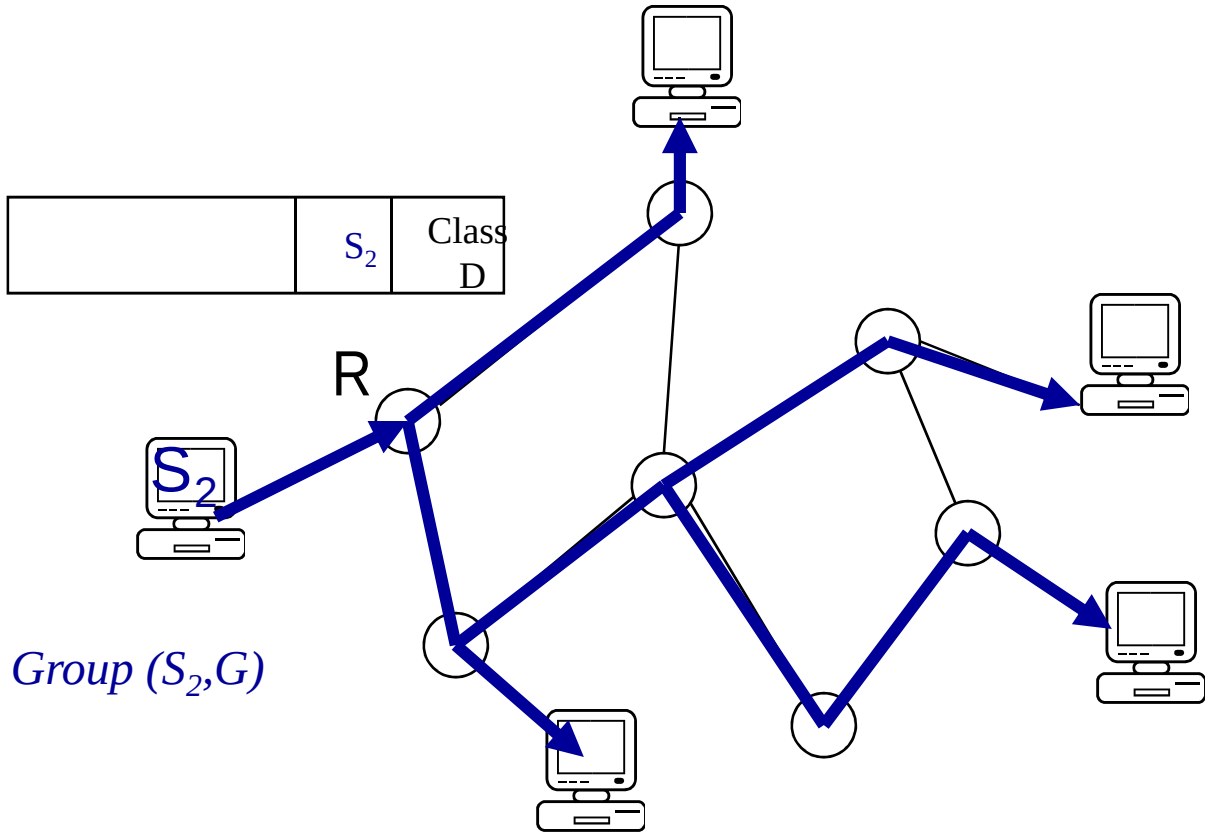
- Multicast and Receiver-oriented



Multicast Trees



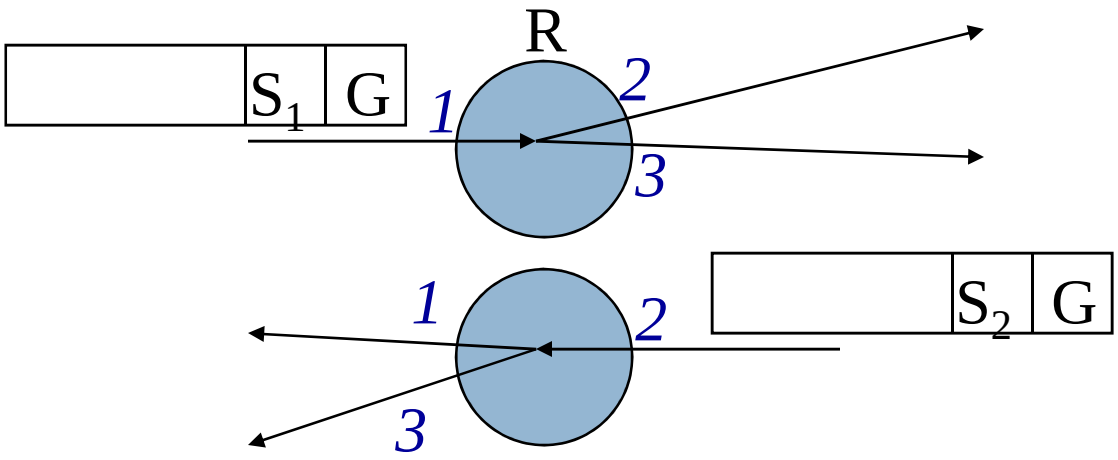
Multicast Trees



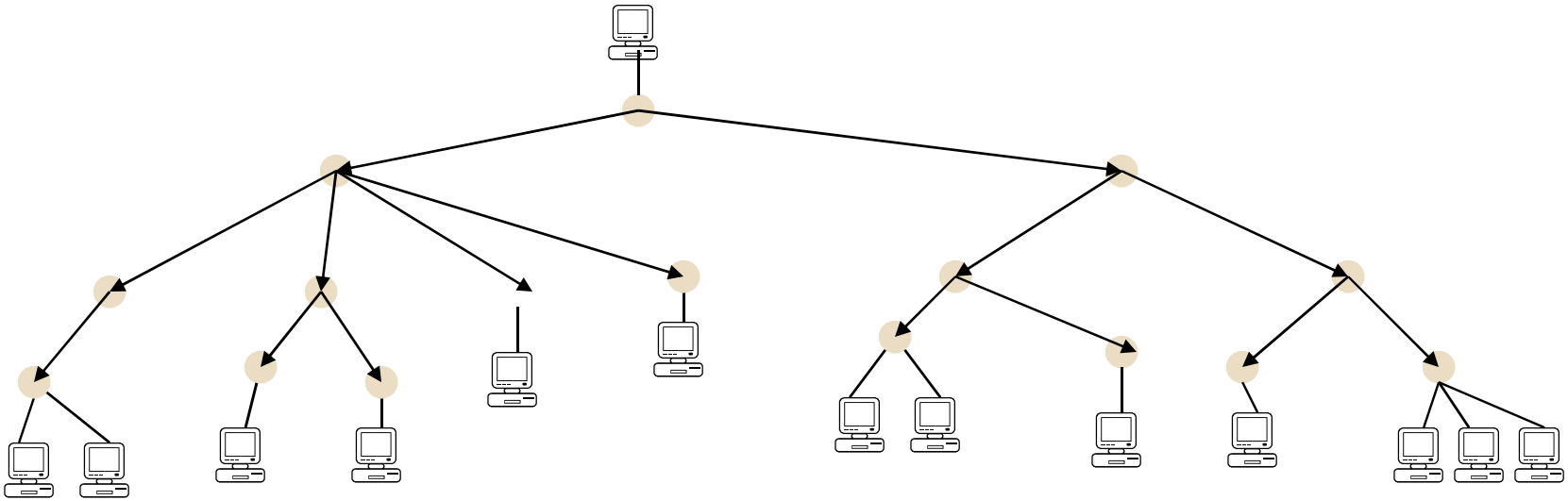
Sender *Multicast Group (S_2,G)*

Multicast Forwarding is Sender-specific

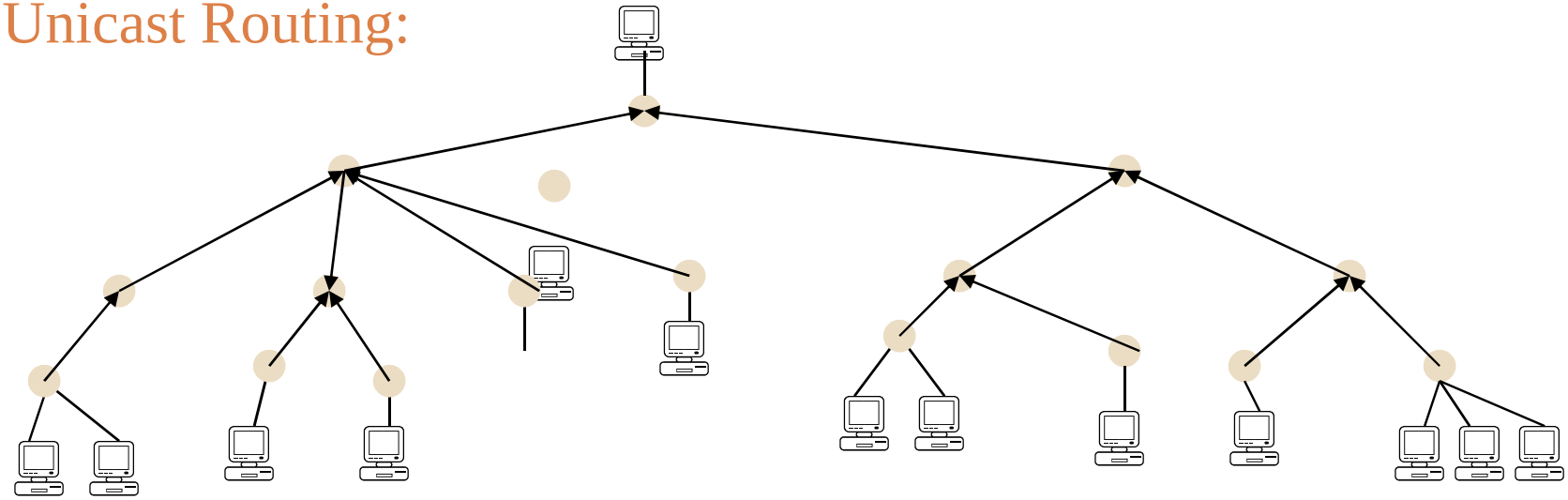
Group Address	Src Address	Src Interface	Dst Interface
G	S_1	1	2,3
	S_2	2	1,3
⋮	⋮	⋮	⋮



Multicast Routing:



Unicast Routing:



Intserv Problems

- per-flow reservation processing is a considerable overhead in large networks.

(256,000 source-destination pairs might be seen in one minute in a backbone router)

- provides for a small number of pre-specified service classes

- Few hosts in today's Internet can express Rspec and Tspec and/or signal them

Differentiated Services Principals

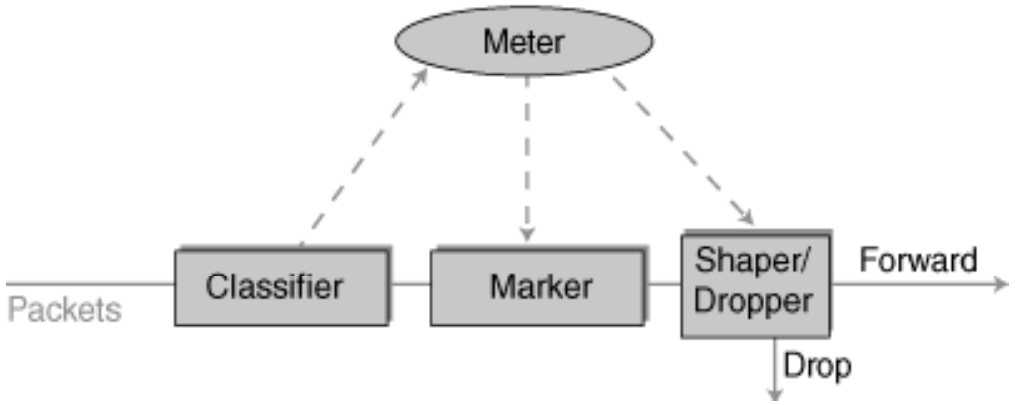
- scalable and flexible service differentiation:

the ability to handle different "classes" of traffic in different ways within the Internet

- packet classification and traffic conditioning onlyat the incoming "edge" of the network.

- Core routers only forward the traffic

Edge function

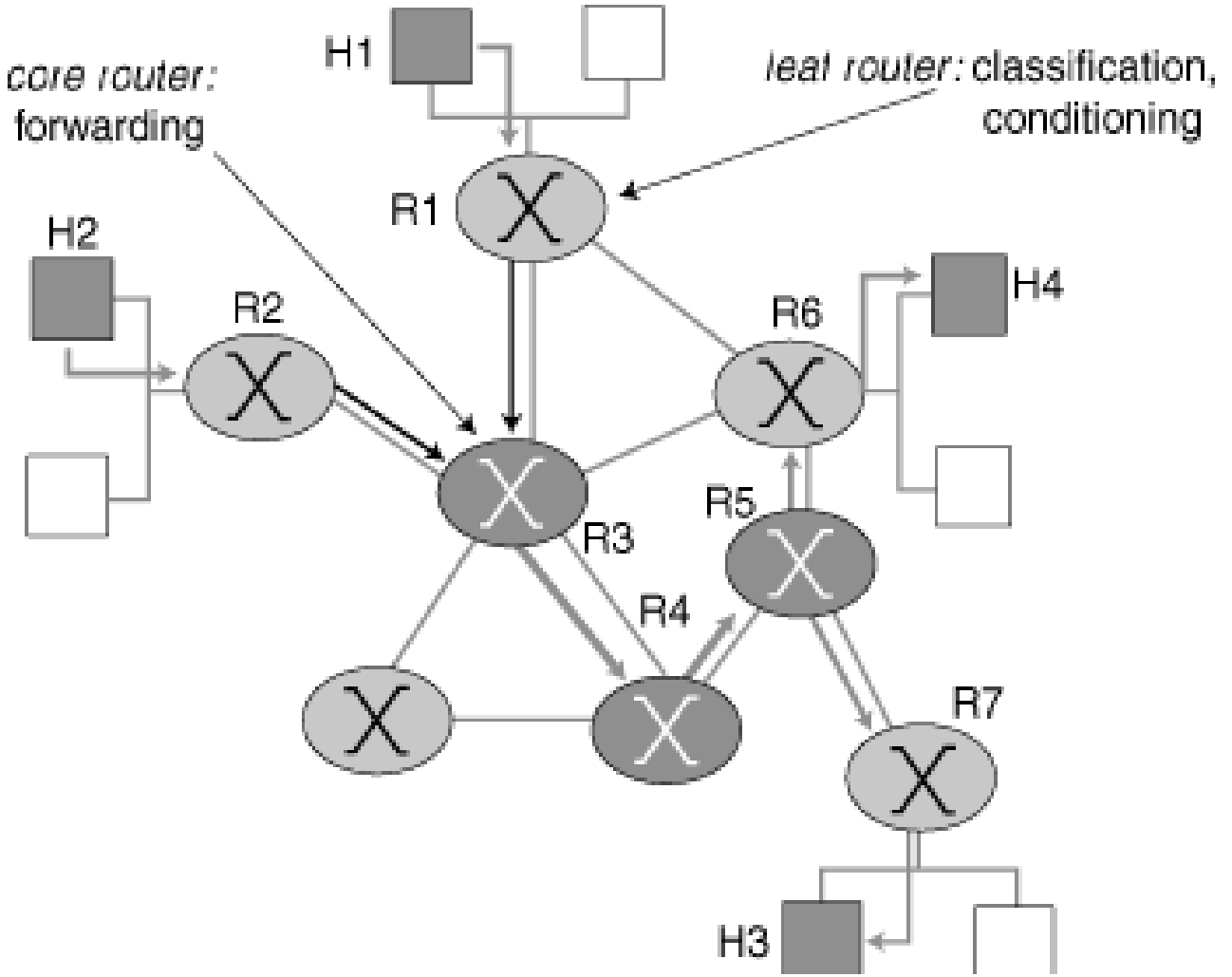


Core (per-hop) function

- Expedited Forwarding - bound on departure rate (minimum guaranteed link bandwidth) of a class of traffic from a router.

- Assured Forwarding - each class is guaranteed to be provided with some minimum amount of bandwidth and buffering : different "drop preference" categories in each class, enable router to drop packets based on their drop preference values

An example


core router:
forwarding
H1
edge router: classification,
conditioning
R1
H2
R2
R6
H4
R5
R3
R4
R7
H3