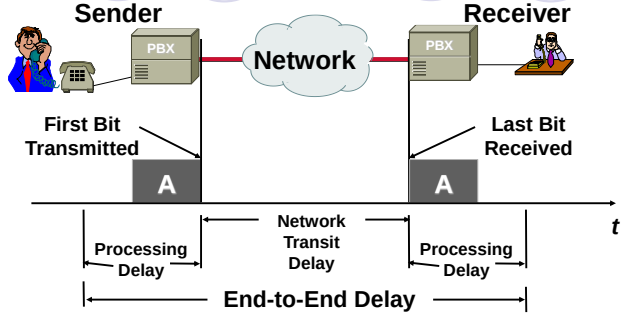
IP QoS Techniques and Timings Through the Network

Quality of Service (QoS)

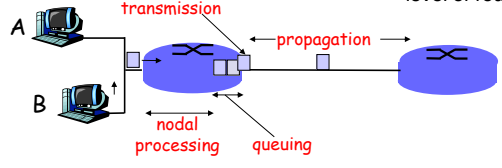
- applications: IP telephony, video conference, distributed interactive worlds
- end-end delay requirements:
  - audio: < 150 msec good, < 400 msec OK
    - includes application-layer (packetization) and network delays
    - higher delays noticeable.

Delay



Delay in packet-switched networks

- packets experience delay on end-to-end path
- four sources of delay at each hop
- nodal processing:
  - check bit errors
  - determine output link
- queuing
  - time waiting at output link for transmission
  - depends on congestion level of router



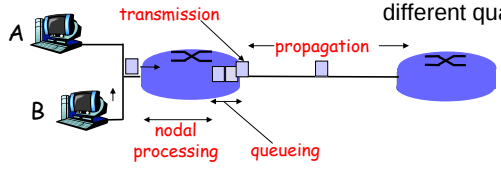
Delay in packet-switched networks

- Transmission delay:
- R=link bandwidth (bps)
- L=packet length (bits)
- time to send bits into link = L/R

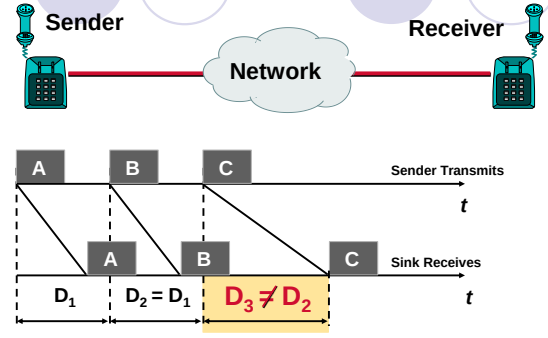
Propagation delay:

- d = length of physical link
- s = propagation speed in medium (~$2x10^8$ m/sec)
- propagation delay = d/s

Note: s and R are *very* different quantities!

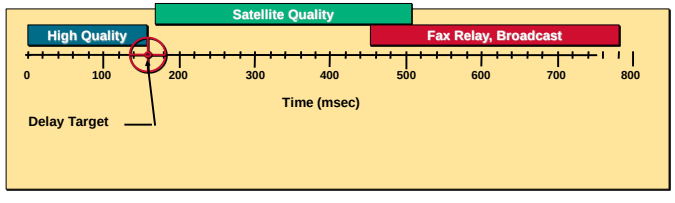


Delay Variation—"Jitter"



Voice Transport and Delay

Cumulative Transmission Path Delay



Voice Parameters & Classes

Parameter \ Class	Class 4: Best (better than today's PSTN/ISDN)	Class 3: High (equivalent to today's PSTN/ISDN)	Class 2: Medium (equivalent to today's wireless networks)	Class 1: Low (significantly impaired, but usable)
End-to-End Delay (ms)	< 100	< 100	< 150	< 400

What is *Quality of Service (*QoS)

- The capability to provide resource assurance and service differentiation so that delay, jitter or loss sensitive applications can perform satisfactorily is often referred to as *quality of service* (QoS).

- For Many Years Internet was primarily used for networking research.
- File transfer, email were the most popular application:
- They do not need any performance guarantee from underlying network.
- New applications such as VoIP, video conferencing, e-commerce applications are sensitive to network performance.

- Internet cannot provide any resource guarantees : the service is *best effort*
- Internet does not provide service differentiation : all packets are treated equal.
- But applications such as VoIP require low delay, jitter and packet loss; whereas file transfer application can tolerate fair amount of delay and loss.
- Thus there is a need to differentiate between packets of such applications.

Current State of Internet

- Uses best-effort service model
- No guarantee of timeliness or delivery
- No service discrimination
- Bandwidth and network congestion problems
- Unpredictable network response time

QoS in circuit-switched network

Quite Easy

QoS in packet-switched network

Difficult (why?)

QoS parameters

Delay
Delay Jitter
Packet loss

Resource Allocation

- Many problems in the Internet come down to issue of resource allocation.
- Packets get delayed or dropped because network resource cannot meet the traffic demands.
- A network consists of shared resources : bandwidth, buffer, serving traffic from competing users.
- To support QoS network must allocate resources and decide who should get how much resources.

Motivation

- Quality of Service (QoS)
- Provide circuit-switched features in a packet-switched network
- Guarantees on rate, latency, and jitter
- Two approaches
  - Differentiated Services
  - Integrated Services

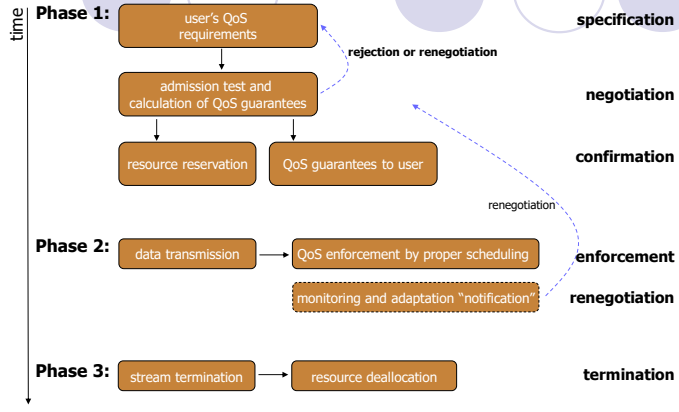
Integrated Services

- Framework by IETF to provide individualized QoS guarantees to individual application sessions
- Goals:
  - Efficient Internet support for applications which require service guarantees
  - Fulfill demands of multipoint, real-time applications (like Voice or video conferences)

Integrated Services

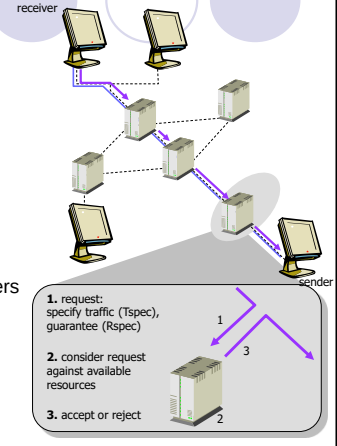
- Two key features
  - reserved resources – the routers need to know what resources are available (both free and reserved).
  - call setup (admission call) – reserve resources on the whole path from source to destination

Resource Management Phases



Integrated Services (IntServ)

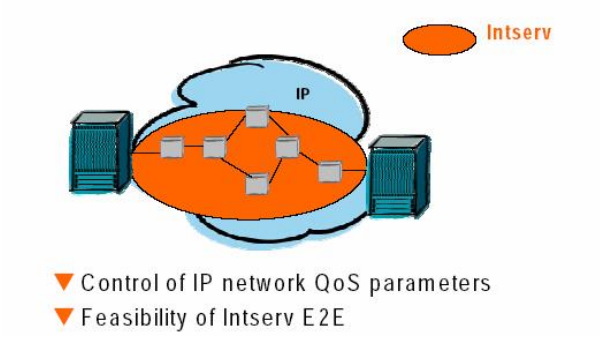
- Admission call:
  - traffic characterization and specification
    - one must specify the traffic one will transmit on the network (Tspec)
    - one must specify the requested QoS (Rspec – reservation specification)
  - signaling for setup
    - send the Tspec and Rspec to all routers
  - per-element admission test
    - each router checks whether the requests specified in the R/Tspecs can be fulfilled
    - if YES, accept; reject otherwise



Integrated Services

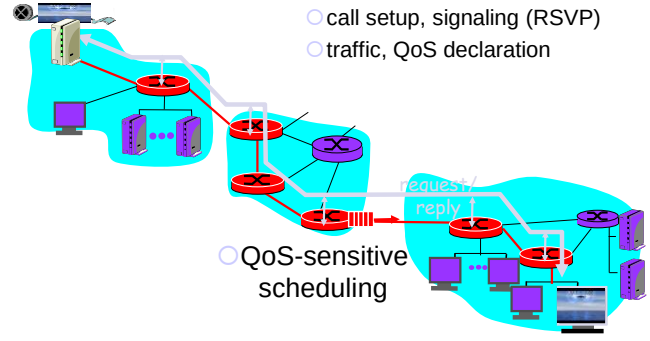
- Based on per flow resource reservation.
- Applications must make a reservation before transmitting traffic.
- Applications characterize its traffic and resource requirement.
- Network uses routing protocol to find a path.
- A reservation protocol is used to install the reservation state along that path.

QoS Architectures



Intserv: QoS guarantee scenario

- Resource reservation
  - call setup, signaling (RSVP)
  - traffic, QoS declaration



Integrated services philosophy:

- Architecture for providing QoS guarantees in IP networks for individual flows
- Fundamental changes in Internet so that applications can reserve end-to-end bandwidth
- Components of this architecture are:
  - Admission control
  - Reservation protocol
  - Routing protocol
  - Classifier and route selection
  - Packet scheduler

Differentiated Service

DiffServ aims at providing support for differential service in the Internet, but is focusing more on traffic aggregates...

- Scalability: signaling, maintaining per-flow router state difficult with large number of flows
- Diffserv approach:
- simple functions in network core, relatively complex functions at edge routers (or hosts)

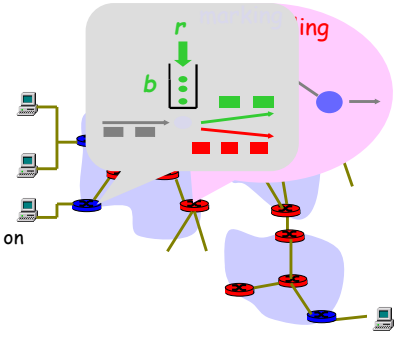
Diffserv Architecture

Edge router:

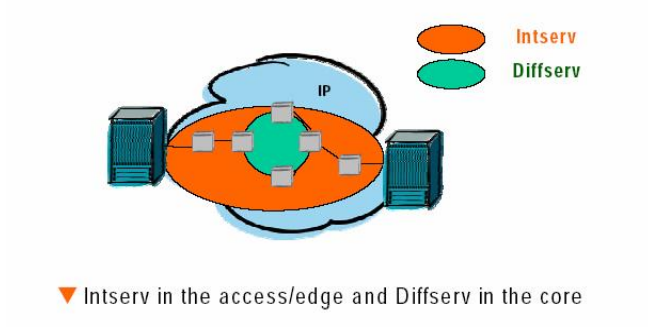
- per-flow traffic management
- Set the DS field; value determines type of service

Core router:

- buffering and scheduling based on marking at edge
- per-class traffic management



QoS Architectures



Intserv in the access/edge and Diffserv in the core