Multimedia Applications

- Multimedia requirements
- Streaming
- Phone over IP
- Recovering from Jitter and Loss
- RTP
- Diff-serv, Int-serv, RSVP

Application Classes

- Typically sensitive to delay, but can tolerate packet loss (would cause minor glitches that can be concealed)
- Data contains audio and video content ("continuous media"), three classes of applications:
  - Streaming
  - Unidirectional Real-Time
  - Interactive Real-Time

Application Classes (more)

- Streaming
  - Clients request audio/video files from servers and pipeline reception over the network and display
  - Interactive: user can control operation (similar to VCR: pause, resume, fast forward, rewind, etc.)
  - Delay: from client request until display start can be 1 to 10 seconds
  - Example: RealAudio/RealVideo

Application Classes (more)

- Unidirectional Real-Time:
  - similar to existing TV and radio stations, but delivery on the network
  - Non-interactive, just listen/view
  - Example, online course broadcast

- Interactive Real-Time:
  - Phone conversation or video conference
  - More stringent delay requirement than Streaming and Unidirectional because of interactive real-time nature
  - Video: < 150 msec acceptable
  - Audio: < 150 msec good, <400 msec acceptable

Challenges

- TCP/UDP/IP suite provides best-effort, no guarantees on expectation or variance of packet delay
- Streaming applications delay of 5 to 10 seconds is typical and has been acceptable, but performance deteriorates if links are congested (transoceanic)
- Real-Time Interactive requirements on delay and its jitter have been satisfied by over-provisioning (providing plenty of bandwidth), what will happen when the load increases?...
Challenges (more)

- Most router implementations use only First-Come-First-Serve (FCFS) packet processing and transmission scheduling.
- To mitigate impact of “best-effort” protocols, we can:
  - Use UDP to avoid TCP and its slow-start phase...
  - Buffer content at client and control playback to remedy jitter
  - Adapt compression level to available bandwidth

Solution Approaches in IP Networks

- Just add more bandwidth and enhance caching capabilities (over-provisioning)!
- Two Camps:
  - Need major change of the protocols (Integrated Services):
    - Incorporate resource reservation (bandwidth, processing, buffering), and new scheduling policies
    - Set up service level agreements with applications, monitor and enforce the agreements, charge accordingly
    - Need moderate changes (“Differentiated Services”):
      - Use two traffic classes for all packets and differentiate service accordingly
      - Charge based on class of packets
      - Network capacity is provided to ensure first class packets incur no significant delay at routers

Streaming

- Important and growing application due to reduction of storage costs, increase in high speed net access from homes, enhancements to caching and introduction of QoS in IP networks
- Audio/Video file is segmented and sent over either TCP or UDP.
  - public segmentation protocol: Real-Time Protocol (RTP)

Using a Streaming Server

- User interactive control is provided
  - public protocol Real Time Streaming Protocol (RTSP)
- Helper Application: displays content, which is typically requested via a Web browser; e.g. RealPlayer; typical functions:
  - Decompression
  - Jitter removal
  - Error correction: use redundant packets to be used for reconstruction of original stream
  - GUI for user control

Options When Using a Streaming Server

- Use UDP and Server sends at a rate (Compression and Transmission) appropriate for client; to reduce jitter, Player buffers initially for 2-5 seconds, then starts display
- Use TCP and sender sends at maximum possible rate under TCP: retransmit when error is
Real Time Streaming Protocol (RTSP)

- For user to control display: rewind, fast forward, pause, resume, etc...
- Out-of-band protocol (uses two connections, one for control messages (Port 554) and for media stream)
- RFC 2326 permits use of either TCP or UDP for the control messages connection, sometimes called the RTSP Channel
- As before, meta file is communicated to web browser which then launches the Player; Player sets up an RTSP connection for control messages in addition to the connection for the streaming media

Real-Time (Phone) Over IP’s Best-Effort

- Internet phone applications generate packets during talk spurs
- Bit rate is 8 KBytes, and every 20 msec, the sender forms a packet of 160 Bytes + a header to be discussed below
- The coded voice information is encapsulated into a UDP packet and sent out; some packets may be lost; up to 20% loss is tolerable; using TCP eliminates loss but at a considerable cost: variance in delay; FEC is sometimes used to fix errors and make up losses

Real-Time (Phone) Over IP’s Best-Effort

- End-to-end delays above 400 msec cannot be tolerated; packets that are that delayed are ignored at the receiver
- Delay jitter is handled by using timestamps, sequence numbers, and delaying playout at receivers either a fixed or a variable amount
- With fixed playout delay, the delay should be as small as possible without missing too many packets; delay cannot exceed 400 msec

Internet Phone with Fixed Playout Delay

Adaptive Playout Delay

- Objective is to use a value for playout delay that tracks the network delay performance as it varies during a phone call
- The playout delay is computed for each talk spurt based on observed average delay and observed deviation from this average delay
- Estimated average delay and deviation of average delay are computed in a manner similar to estimates of RTT and deviation in TCP
- The beginning of a talk spurt is identified from examining the timestamps in successive and/or sequence numbers of chunks

Recovery From Packet Loss

- Loss is in a broader sense: packet never arrives or arrives later than its scheduled playout time
- Since retransmission is inappropriate for Real Time applications, FEC or Interleaving are used to reduce loss impact.
- FEC is Forward Error Correction
- Simplest FEC scheme adds a redundant chunk made up of exclusive OR of a group of n chunks; redundancy is 1/n; can reconstruct if at most one lost chunk; playout time schedule assumes a loss per group
Recovery From Packet Loss

- Mixed quality streams are used to include redundant duplicates of chunks; upon loss a lower quality redundant chunk is available.
- With one redundant chunk per chunk can recover from single losses.

Piggybacking Lower Quality Stream

Interleaving

- Has no redundancy, but can cause delay in playout beyond Real Time requirements
- Divide 20 msec of audio data into smaller units of 5 msec each and interleave
- Upon loss, have a set of partially filled chunks

Improving QOS in IP Networks

- IETF groups are working on proposals to provide better QOS control in IP networks, i.e., going beyond best effort to provide some assurance for QOS
- Work in Progress includes RSVP, Differentiated Services, and Integrated Services
- Simple model for sharing and congestion studies:

Principles for QOS Guarantees

- Consider a phone application at 1Mbps and an FTP application sharing a 1.5 Mbps link.
  - bursts of FTP can congest the router and cause audio packets to be dropped.
  - want to give priority to audio over FTP
- PRINCIPLE 1: Marking of packets is needed for router to distinguish between different classes; and new router policy to treat packets accordingly

Principles for QOS Guarantees (more)

- Applications misbehave (audio sends packets at a rate higher than 1Mbps assumed above);
- PRINCIPLE 2: provide protection (isolation) for one class from other classes
- Require Policing Mechanisms to ensure sources adhere to bandwidth requirements; Marking and Policing need to be done at the edges:
Principles for QOS Guarantees (more)

- Alternative to Marking and Policing: allocate a set portion of bandwidth to each application flow; can lead to inefficient use of bandwidth if one of the flows does not use its allocation
  - PRINCIPLE 3: While providing isolation, it is desirable to use resources as efficiently as possible

- Cannot support traffic beyond link capacity
  - PRINCIPLE 4: Need a Call Admission Process; application flow declares its needs, network may block call if it cannot satisfy the needs

Summary

QoS for networked applications

Integrated Services

- An architecture for providing QOS guarantees in IP networks for individual application sessions
- relies on resource reservation, and routers need to maintain state info (Virtual Circuit?), maintaining records of allocated resources and responding to new Call setup requests on that basis

Call Admission

- Session must first declare its QOS requirement and characterize the traffic it will send through the network
- R-spec: defines the QOS being requested
- T-spec: defines the traffic characteristics
- A signaling protocol is needed to carry the R-spec and T-spec to the routers where reservation is required; RSVP is a leading candidate for such signaling protocol

Differentiated Services

- Intended to address the following difficulties with Intserv and RSVP:
  - Scalability: maintaining states by routers in high speed networks is difficult due to the very large number of flows
  - Flexible Service Models: Intserv has only two classes, want to provide more qualitative service classes; want to provide 'relative' service distinction (Platinum, Gold, Silver, Lead ...)
  - Simpler signaling: (than RSVP) applications and users may only want to specify a more qualitative notion of service
Differentiated Services

- **Approach:**
  - Only simple functions in the core, and relatively complex functions at edge routers (or hosts)
  - Do not define service classes, instead provides functional components with which service classes can be built