Multimedia Networking

The majority of the slides in this course are adapted from the accompanying slides to the books by Larry Peterson and Bruce Davie and by Jim Kurose and Keith Ross. Additional slides and/or figures from other sources and from Vasos Vassiliou are also included in this presentation.

Multimedia Over Today's Internet

TCP/UDP/IP: "best-effort service"

- *no* guarantees on delay, loss

? ? ? ? ? ?

But you said multimedia apps requires QoS and level of performance to be effective!

? ? ? ? ?

Today's Internet multimedia applications use application-level techniques to mitigate (as best possible) effects of delay, loss

A few words about audio compression

- Analog signal sampled at constant rate
  - telephone: 8,000 samples/sec
  - CD music: 44,100 samples/sec
- Each sample quantized, i.e., rounded
  - e.g., 2^8=256 possible quantized values
- Each quantized value represented by bits
  - 8 bits for 256 values
- Example: 8,000 samples/sec, 256 quantized values --> 64,000 bps
- Receiver converts it back to analog signal:
  - some quality reduction

Example rates

- CD: 1.411 Mbps
- MP3: 96, 128, 160 kbps
- Internet telephony: 5.3 - 13 kbps

A few words about video compression

- Video is sequence of images displayed at constant rate
  - e.g. 24 images/sec
- Digital image is array of pixels
- Each pixel represented by bits
- Redundancy
  - spatial
  - temporal

Examples:

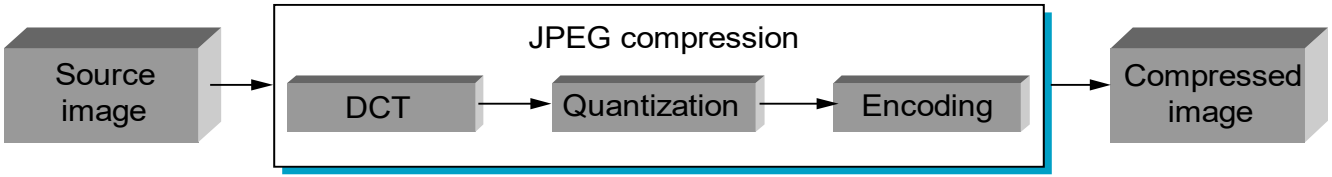
- MPEG 1 (CD-ROM) 1.5 Mbps
- MPEG2 (DVD) 3-6 Mbps
- MPEG4 (often used in Internet, < 1 Mbps)

Research:

- Layered (scalable) video
  - adapt layers to available bandwidth

Image Compression

- JPEG: Joint Photographic Expert Group
- Lossy still-image compression
- Three phase process



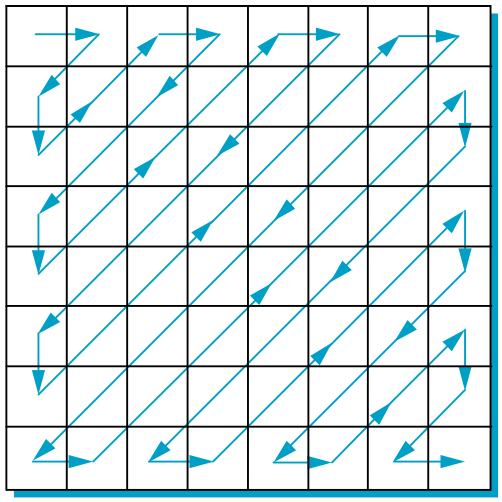
  - process in 8x8 block chunks (macro-block)
  - grayscale: each pixel is three values (YUV)
  - DCT: transforms signal from spatial domain into and equivalent signal in the frequency domain (loss-less)
  - apply a quantization to the results (lossy)
  - RLE-like encoding (loss-less)

Quantization and Encoding

- Quantization Table

3	5	7	9	11	13	15	17
5	7	9	11	13	15	17	19
7	9	11	13	15	17	19	21
9	11	13	15	17	19	21	23
11	13	15	17	19	21	23	25
13	15	17	19	21	23	25	27
15	17	19	21	23	25	27	29
17	19	21	23	25	27	29	31

- Encoding Pattern



Example of coding an image and tradeoff between quality and image size

256x256 grey scale

original (GIF: 54749)

JPEG Q=20, 0.54 bits/pixel (4442 bytes)

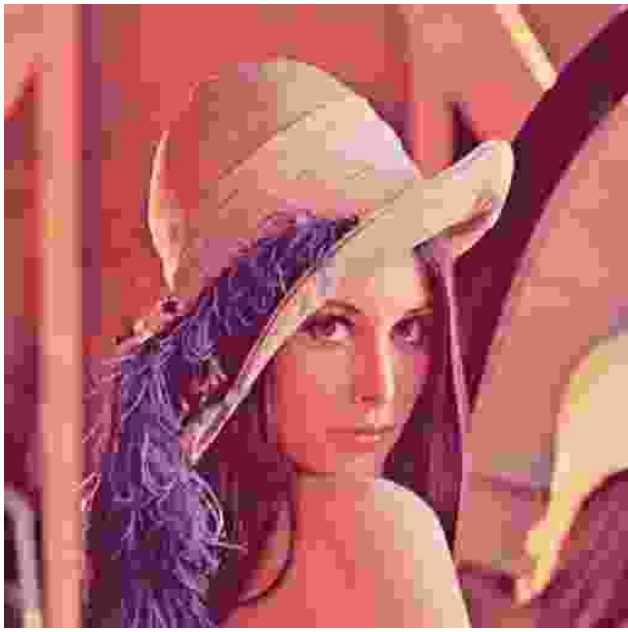
JPEG Q=5, 0.25 bits/pixel (2080 bytes)

Example of coding an image and tradeoff between quality and image size (cont.)

512 x 512 colour

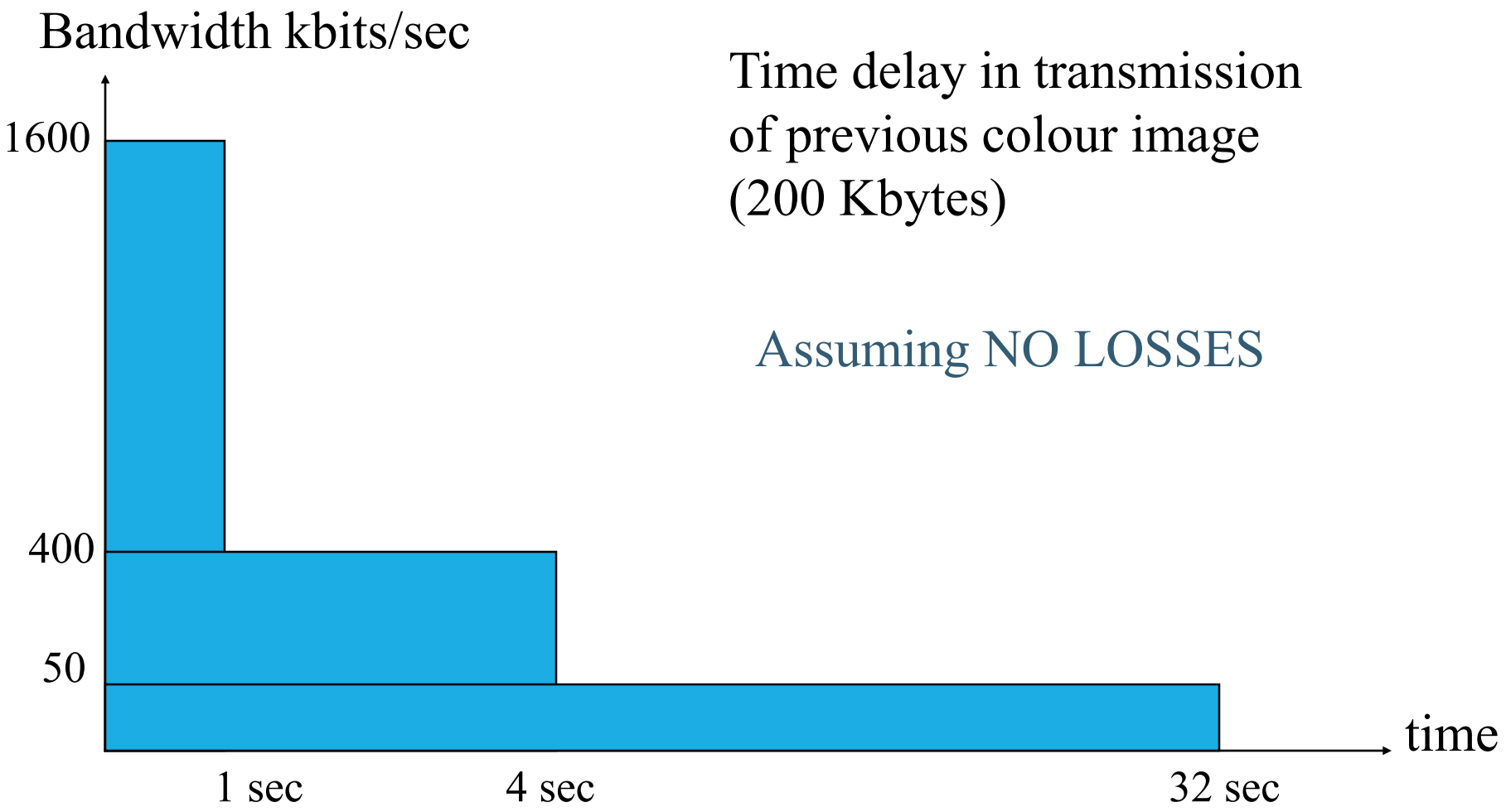
original
(GIF: 202749)

JPEG, Q=5, 0.22 bits/pixel (7128 bytes)



JPEG, Q=20, 0.43 bits/pixel (13984 bytes)

Example of coding an image and tradeoff between quality and image size (cont.)



Time delay in transmission of previous colour image (200 Kbytes)

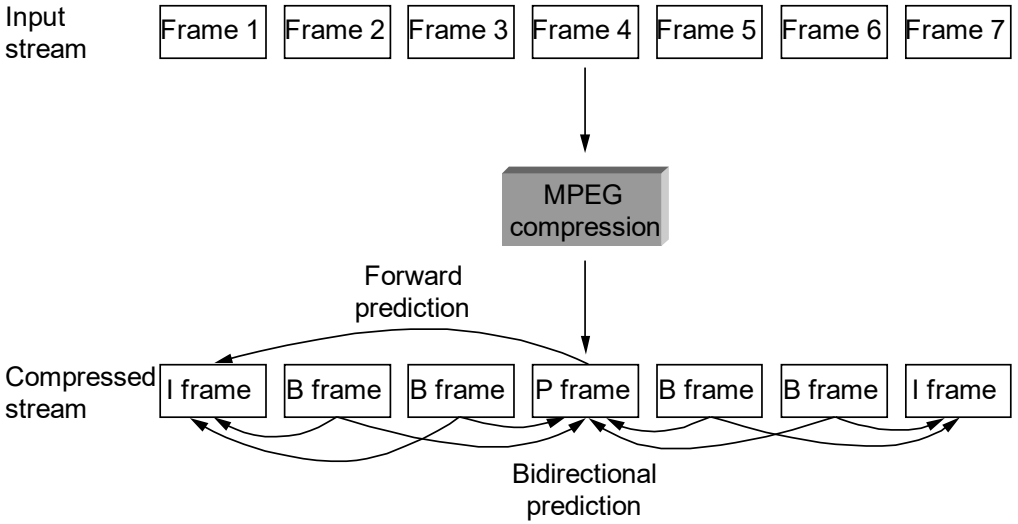
Assuming NO LOSSES

MPEG

- Motion Picture Expert Group
- Lossy compression of video
- First approximation: JPEG on each frame
- Also remove inter-frame redundancy

MPEG (cont)

- Frame types
  - I frames: intrapicture
  - P frames: predicted picture
  - B frames: bidirectional predicted picture



- Example sequence transmitted as I P B B I B B

MPEG (cont)

- B and P frames
  - coordinate for the macroblock in the frame
  - motion vector relative to previous reference frame (B, P)
  - motion vector relative to subsequent reference frame (B)
  - delta for each pixel in the macro block
- Effectiveness
  - typically 90-to-1
  - as high as 150-to-1
  - 30-to-1 for I frames
  - P and B frames get another 3 to 5x

Transmitting MPEG

- Adapt the encoding
  - resolution
  - frame rate
  - quantization table
  - GOP mix
- Packetization
- Dealing with loss
- GOP-induced latency

Layered Video

- Layered encoding
  - e.g., wavelet encoded
- Receiver Layered Multicast (RLM)
  - transmit each layer to a different group address
  - receivers subscribe to the groups they can “afford”
  - Probe to learn if you can afford next higher group/layer
- Smart Packet Dropper (multicast or unicast)
  - select layers to send/drop based on observed congestion
  - observe directly or use RTP feedback

Streaming Stored Multimedia

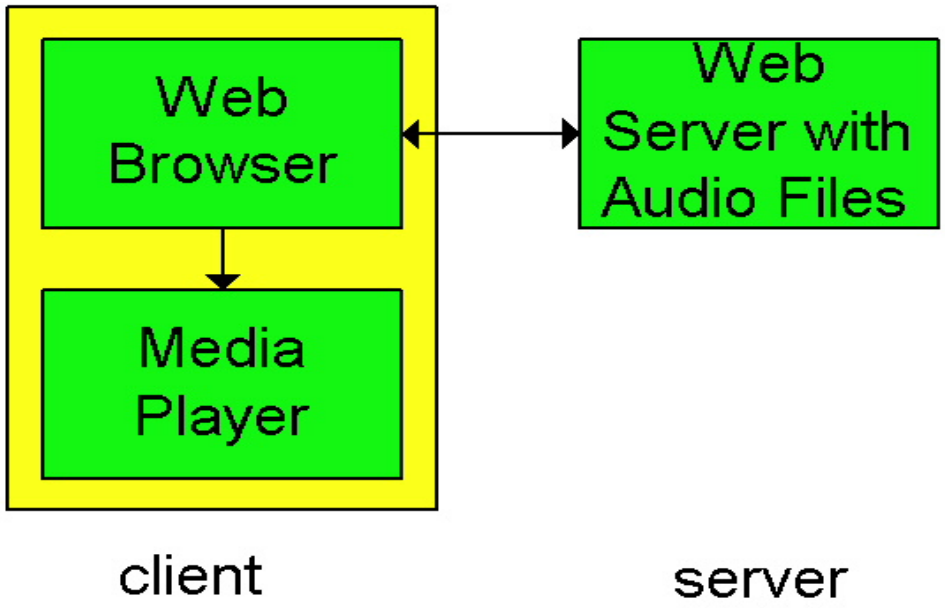
Application-level streaming techniques for making the best out of best effort service:

- client side buffering
- use of UDP versus TCP
- multiple encodings of multimedia

Media Player

- jitter removal
- decompression
- error concealment
- graphical user interface w/ controls for interactivity

Internet multimedia: simplest approach

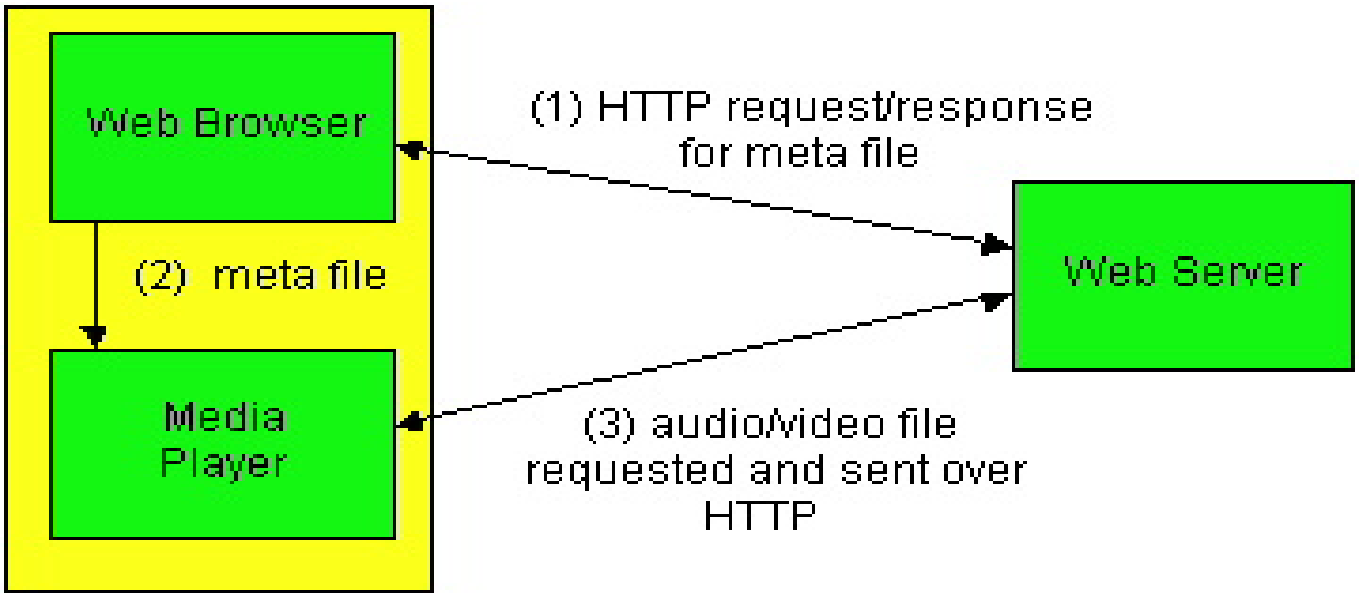


- audio or video stored in file
- files transferred as HTTP object
  - received in entirety at client
  - then passed to player

audio, video not streamed:

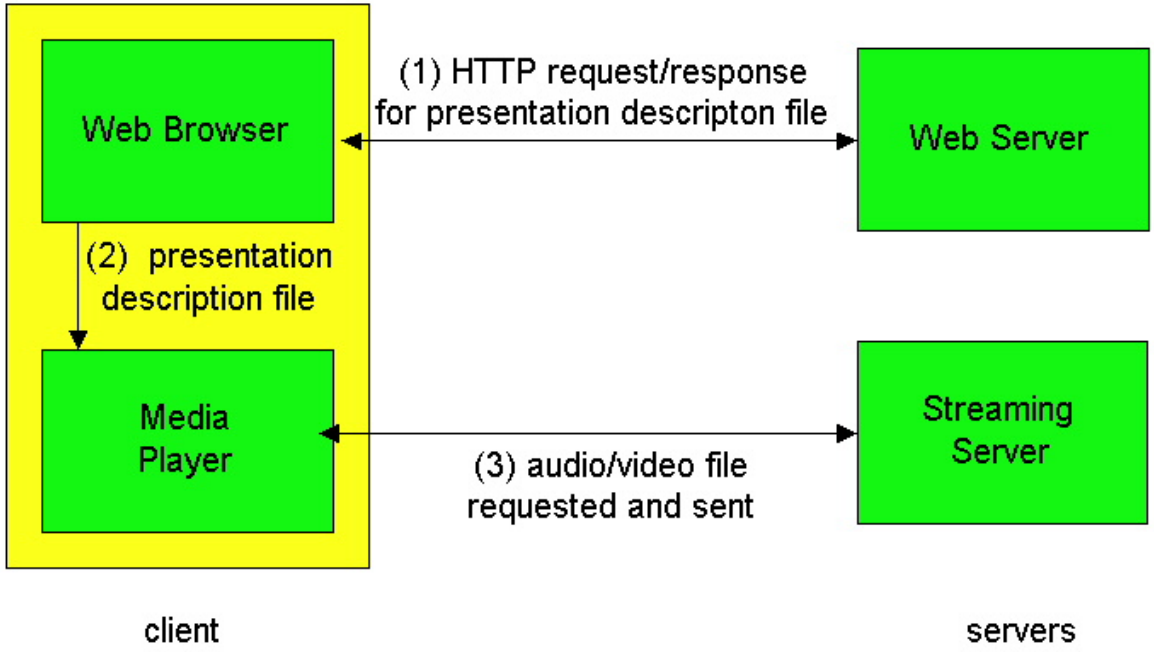
- no "pipelining," long delays until playout!

Internet multimedia: streaming approach



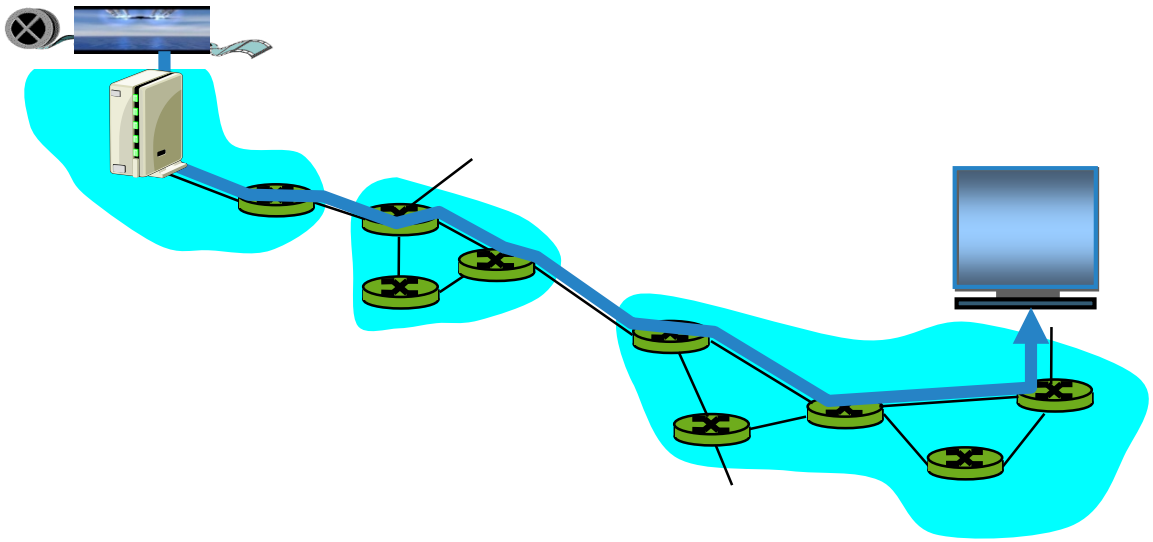
- browser GETs **metafile**
- browser launches player, passing metafile
- player contacts server
- server **streams** audio/video to player

Streaming from a streaming server



- This architecture allows for non-HTTP protocol between server and media player
- Can also use UDP instead of TCP.

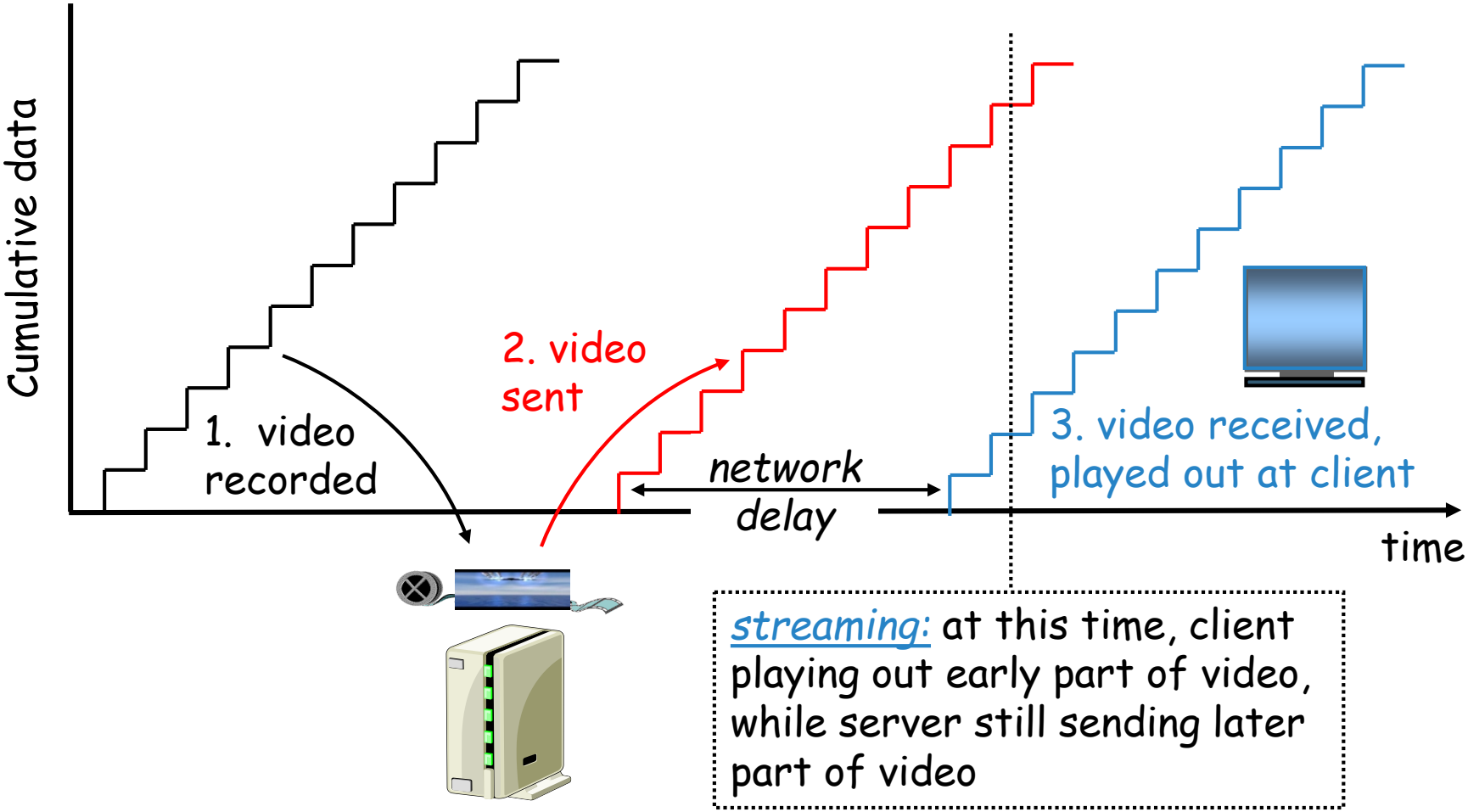
Streaming Stored Multimedia



Streaming:

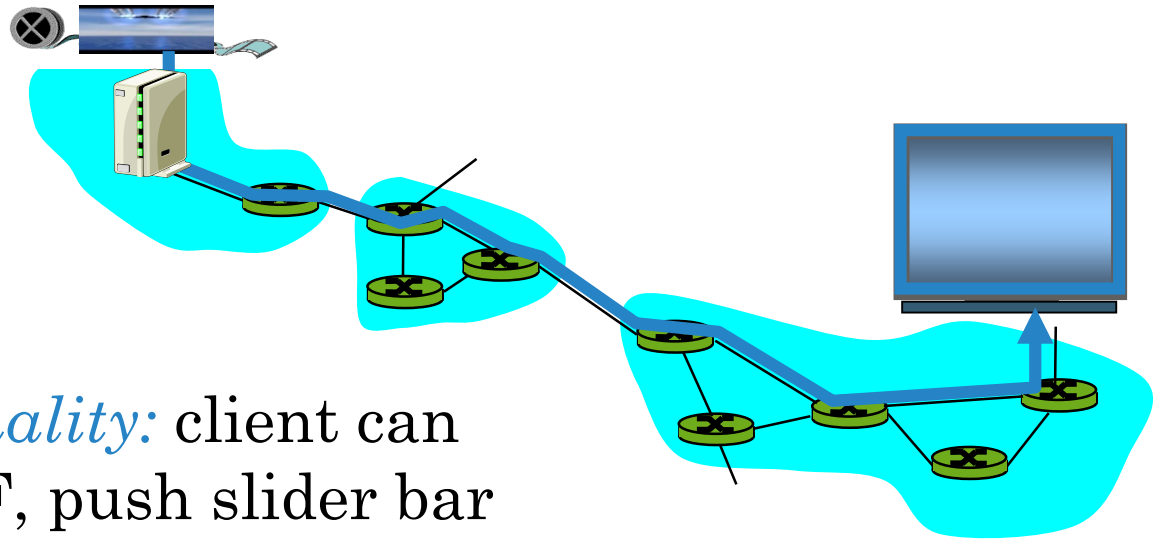
- media stored at source
- transmitted to client
- streaming: client playout begins before all data has arrived
- timing constraint for still-to-be transmitted data: in time for playout

Streaming Stored Multimedia: What is it?



streaming: at this time, client playing out early part of video, while server still sending later part of video

Streaming Stored Multimedia: Interactivity



- *VCR-like functionality:* client can pause, rewind, FF, push slider bar
  - 10 sec initial delay OK
  - 1-2 sec until command effect OK
  - RTSP often used (more later)
- timing constraint for still-to-be transmitted data: in time for playout

Streaming Live Multimedia

Examples:

- Internet radio talk show
- Live sporting event

Streaming

- playback buffer
- playback can lag tens of seconds after transmission
- still have timing constraint

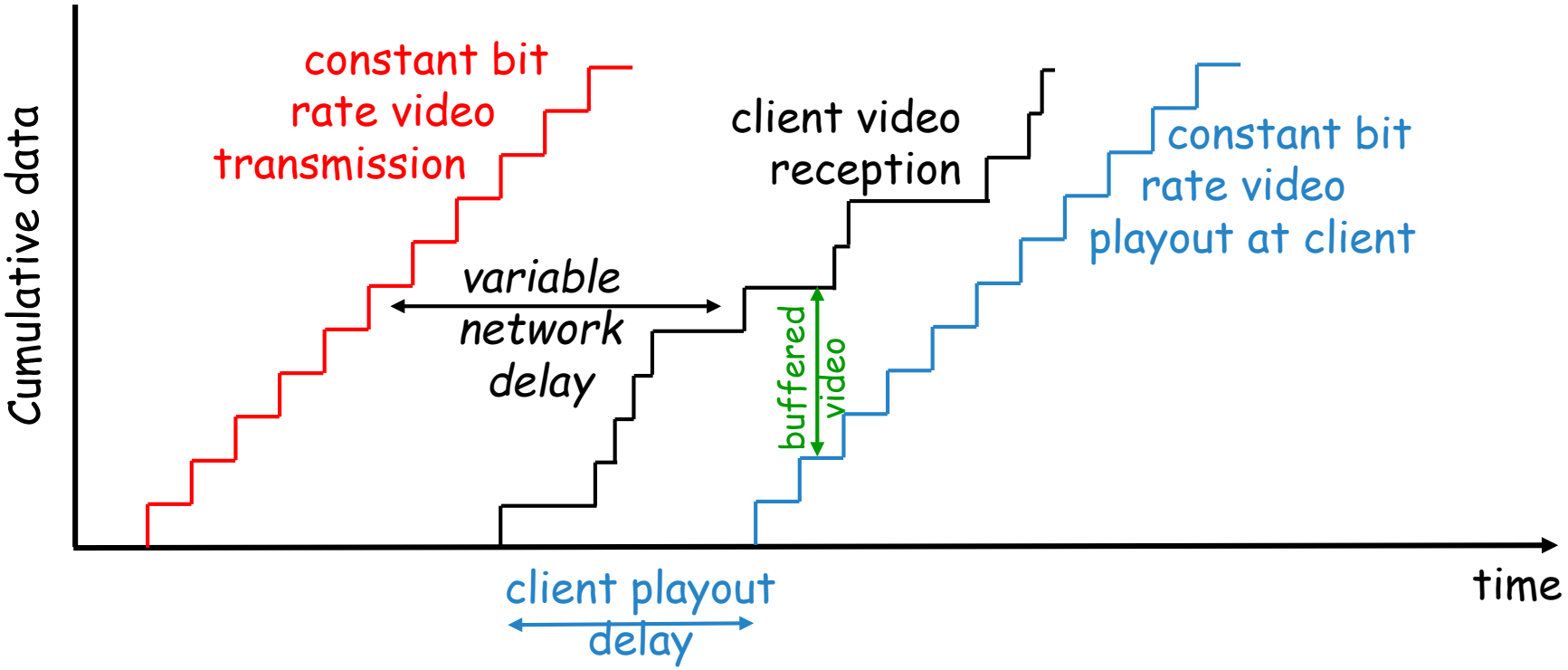
Interactivity

- fast forward impossible
- rewind, pause possible!

Interactive, Real-Time Multimedia

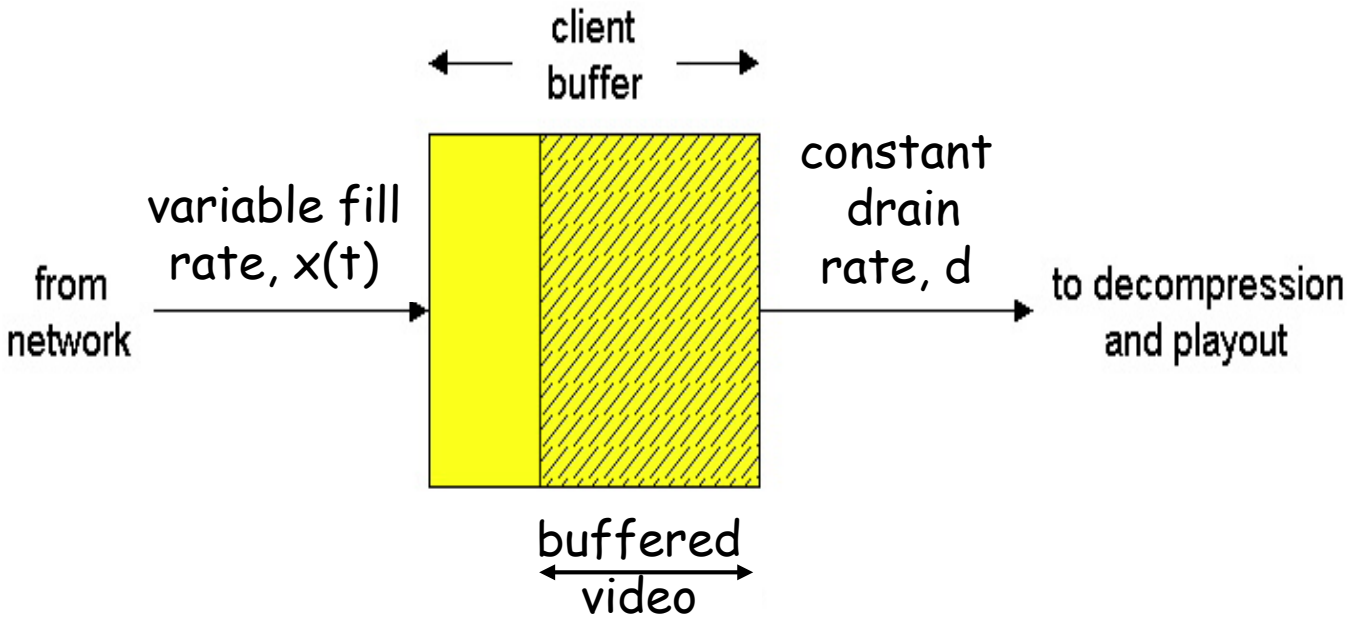
- end-end delay requirements:
  - audio: < 150 msec good, < 400 msec OK
    - includes application-level (packetization) and network delays
    - higher delays noticeable, impair interactivity
- session initialization
  - how does callee advertise its IP address, port number, encoding algorithms?
- applications: IP telephony, video conference, distributed interactive worlds

Streaming Multimedia: Client Buffering



- Client-side buffering, playout delay compensate for network-added delay, delay jitter

Streaming Multimedia: Client Buffering

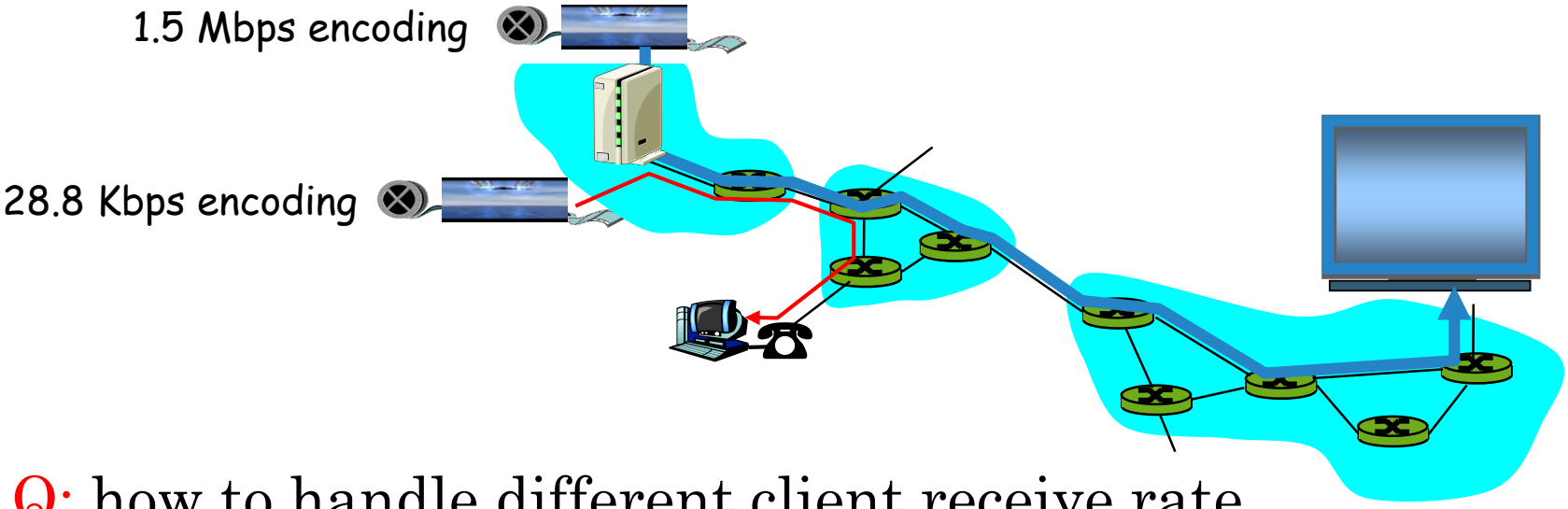


- Client-side buffering, playout delay compensate for network-added delay, delay jitter

Streaming Multimedia: UDP or TCP?

- UDP
  - server sends at rate appropriate for client (oblivious to network congestion !)
    - often send rate = encoding rate = constant rate
    - then, fill rate = constant rate - packet loss
  - short playout delay (2-5 seconds) to compensate for network delay jitter
  - error recover: time permitting
- TCP
  - send at maximum possible rate under TCP
  - fill rate fluctuates due to TCP congestion control
  - larger playout delay: smooth TCP delivery rate
  - HTTP/TCP passes more easily through firewalls

Streaming Multimedia: client rate(s)



Q: how to handle different client receive rate capabilities?

- 28.8 Kbps dialup
- 100Mbps Ethernet

A: server stores, transmits multiple copies of video, encoded at different rates

Interactive Multimedia: Internet Phone

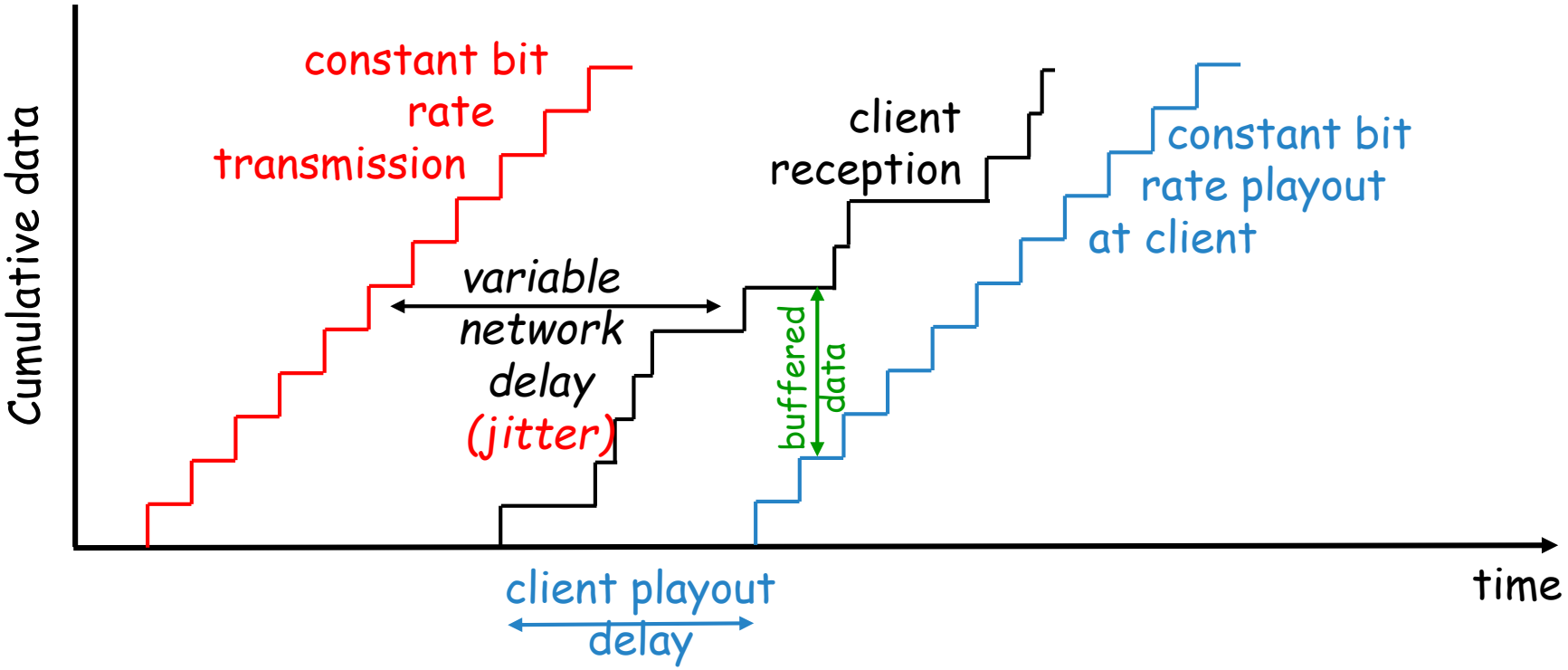
Introduce Internet Phone by way of an example

- speaker's audio: alternating talk spurts, silent periods.
  - 64 kbps during talk spurt
- pkts generated only during talk spurts
  - 20 msec chunks at 8 Kbytes/sec: 160 bytes data
- application-layer header added to each chunk.
- Chunk+header encapsulated into UDP segment.
- application sends UDP segment into socket every 20 msec during talkspurt.

Internet Phone: Packet Loss and Delay

- network loss: IP datagram lost due to network congestion (router buffer overflow)
- delay loss: IP datagram arrives too late for playout at receiver
  - delays: processing, queueing in network; end-system (sender, receiver) delays
  - typical maximum tolerable delay: 400 ms
- loss tolerance: depending on voice encoding, losses concealed, packet loss rates between 1% and 10% can be tolerated.

Delay Jitter



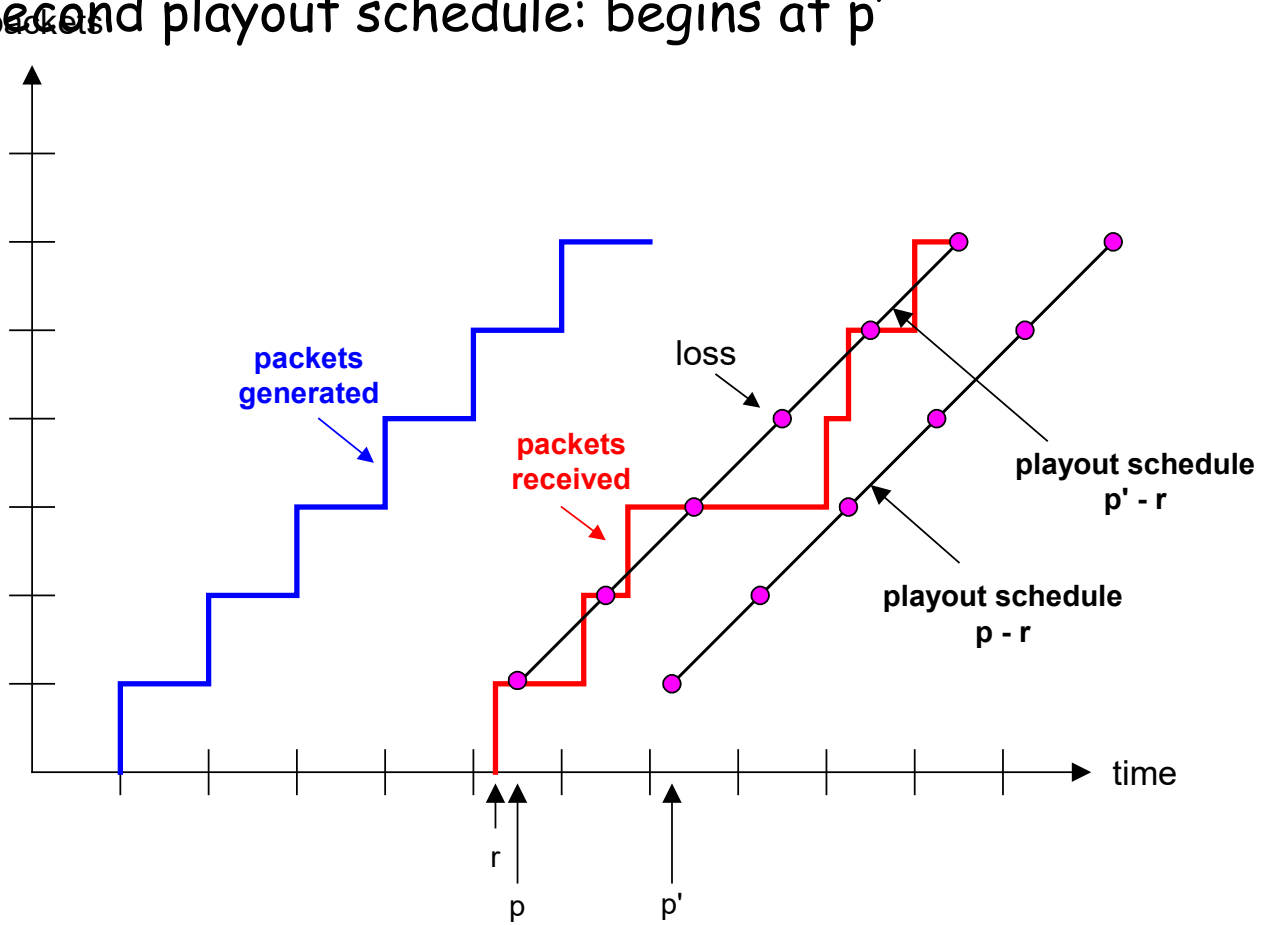
- Consider the end-to-end delays of two consecutive packets: difference can be more or less than 20 msec

Internet Phone: Fixed Playout Delay

- Receiver attempts to playout each chunk exactly q msecs after chunk was generated.
  - chunk has time stamp t: play out chunk at t+q .
  - chunk arrives after t+q: data arrives too late for playout, data "lost"
- Tradeoff for q:
  - large q: less packet loss
  - small q: better interactive experience

Fixed Playout Delay

- Sender generates packets every 20 msec during talk spurt.
- First packet received at time r
- First playout schedule: begins at p
- Second playout schedule: begins at p'



Adaptive Playout Delay, I

- Goal: minimize playout delay, keeping late loss rate low
- Approach: adaptive playout delay adjustment:
  - Estimate network delay, adjust playout delay at beginning of each talk spurt.
  - Silent periods compressed and elongated.
  - Chunks still played out every 20 msec during talk spurt.

t_i = timestamp of the ith packet

r_i = the time packet i is received by receiver

p_i = the time packet i is played at receiver

$r_i - t_i$ = network delay for ith packet

d_i = estimate of average network delay after receiving ith packet

Dynamic estimate of average delay at receiver:

$$d_i = (1-u)d_{i-1} + u(r_i - t_i)$$

where u is a fixed constant (e.g., $u = .01$).

Adaptive playout delay II

Also useful to estimate the average deviation of the delay, v_i :

$$v_i = (1-u)v_{i-1} + u\,|\,r_i - t_i - d_i\,|$$

The estimates d_i and v_i are calculated for every received packet, although they are only used at the beginning of a talk spurt.

For first packet in talk spurt, playout time is:

$$p_i = t_i + d_i + Kv_i$$

where K is a positive constant.

Remaining packets in talkspurt are played out periodically

Adaptive Playout, III

- Q: How does receiver determine whether packet is first in a talkspurt?
- If no loss, receiver looks at successive timestamps.
  - difference of successive stamps > 20 msec -->talk spurt begins.
- With loss possible, receiver must look at both time stamps and sequence numbers.
  - difference of successive stamps > 20 msec and sequence numbers without gaps --> talk spurt begins.

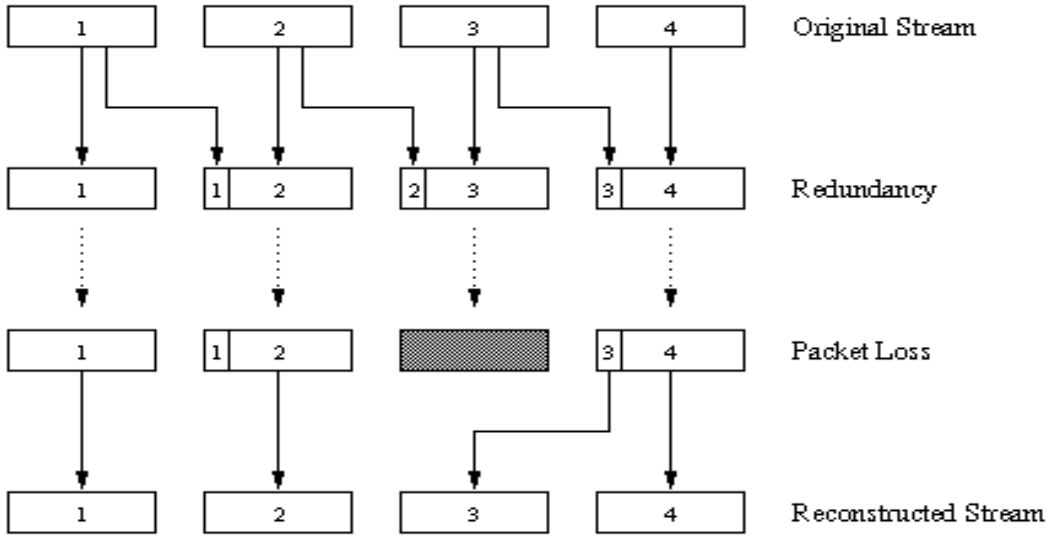
Recovery from packet loss (1)

forward error correction (FEC): simple scheme

- for every group of n chunks create a redundant chunk by exclusive OR-ing the n original chunks
- send out n+1 chunks, increasing the bandwidth by factor 1/n.
- can reconstruct the original n chunks if there is at most one lost chunk from the n+1 chunks
- Playout delay needs to be fixed to the time to receive all n+1 packets
- Tradeoff:
  - increase n, less bandwidth waste
  - increase n, longer playout delay
  - increase n, higher probability that 2 or more chunks will be lost

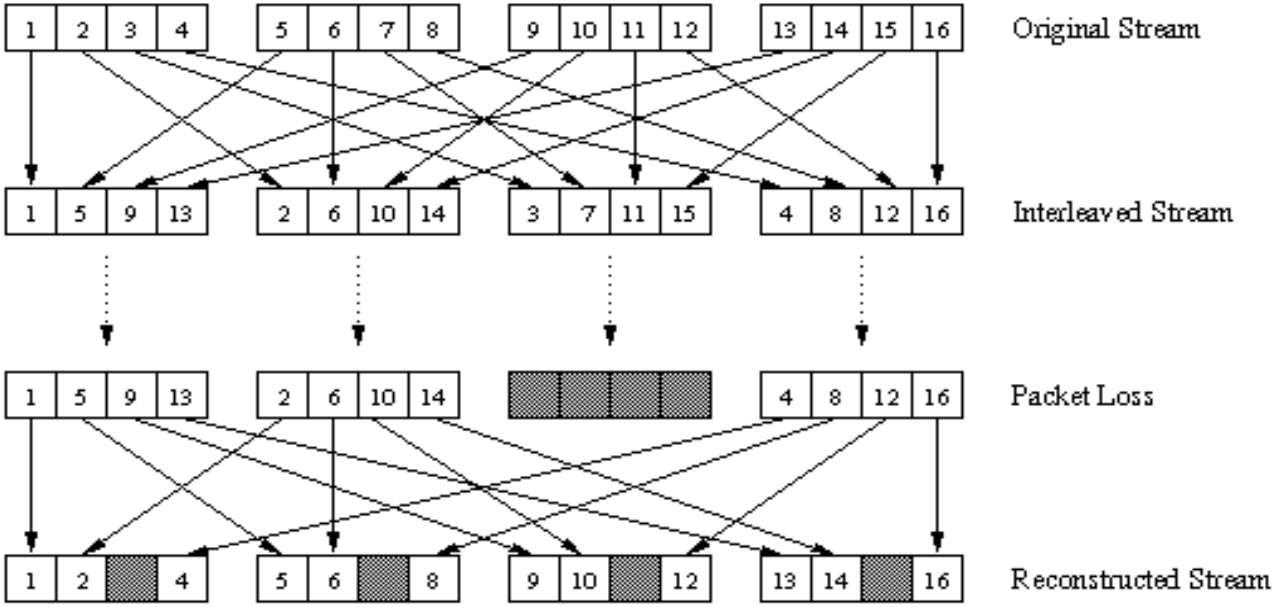
Recovery from packet loss (2)

- 2nd FEC scheme
- "piggyback lower quality stream"
- send lower resolution audio stream as the redundant information
- for example, nominal stream PCM at 64 kbps and redundant stream GSM at 13 kbps.



- Whenever there is non-consecutive loss, the receiver can conceal the loss.
- Can also append (n-1)st and (n-2)nd low-bit rate chunk

Recovery from packet loss (3)



t of

Interleaving

- chunks are broken up into smaller units
- for example, 4 5 msec units per chunk
- Packet contains small units from different chunks

Summary: Internet Multimedia: bag of tricks

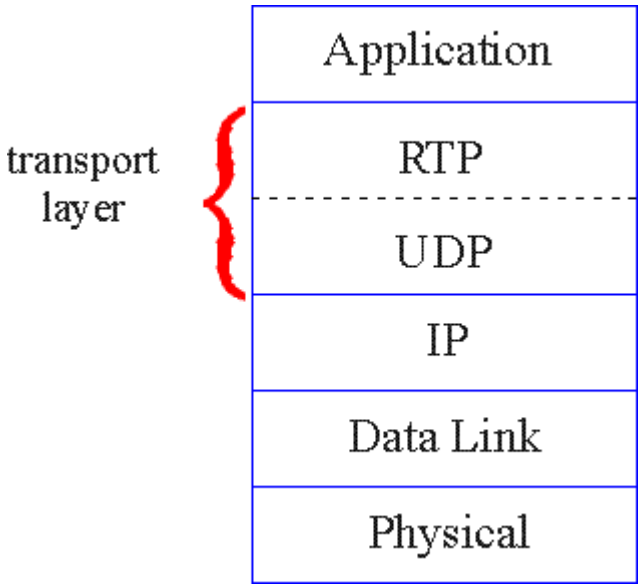
- use UDP to avoid TCP congestion control (delays) for time-sensitive traffic
- client-side adaptive playout delay: to compensate for delay
- server side matches stream bandwidth to available client-to-server path bandwidth
  - chose among pre-encoded stream rates
  - dynamic server encoding rate
- error recovery (on top of UDP)
  - FEC, interleaving
  - retransmissions, time permitting
  - conceal errors: repeat nearby data

Real-Time Protocol (RTP)

- RTP specifies a packet structure for packets carrying audio and video data
- RFC 1889.
- RTP packet provides
  - payload type identification
  - packet sequence numbering
  - timestamping
- RTP runs in the end systems.
- RTP packets are encapsulated in UDP segments
- Interoperability: If two Internet phone applications run RTP, then they may be able to work together

RTP runs on top of UDP

- RTP libraries provide a transport-layer interface
- that extend UDP:
  - port numbers, IP addresses
  - payload type identification
  - packet sequence numbering
  - time-stamping



RTP Example

- Consider sending 64 kbps PCM-encoded voice over RTP.
- Application collects the encoded data in chunks, e.g., every 20 msec = 160 bytes in a chunk.
- The audio chunk along with the RTP header form the RTP packet, which is encapsulated into a UDP segment.
- RTP header indicates type of audio encoding in each packet
  - sender can change encoding during a conference.
- RTP header also contains sequence numbers and timestamps.

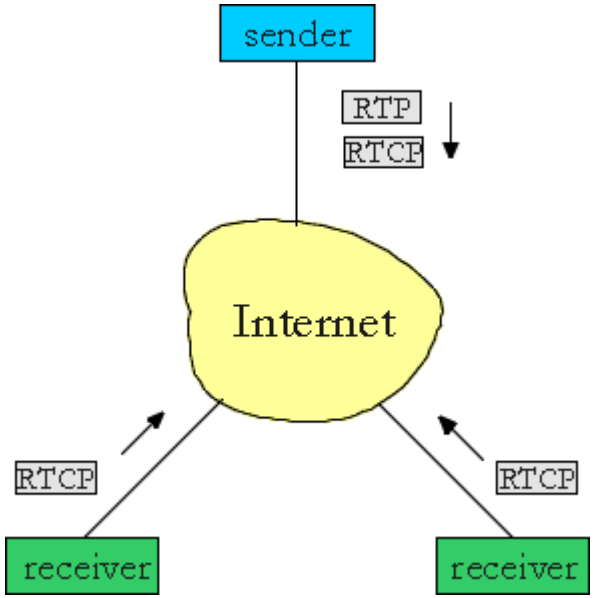
RTP and QoS

- RTP does **not** provide any mechanism to ensure timely delivery of data or provide other quality of service guarantees.
- RTP encapsulation is only seen at the end systems: it is not seen by intermediate routers.
  - Routers providing best-effort service do not make any special effort to ensure that RTP packets arrive at the destination in a timely matter.

Real-Time Control Protocol (RTCP)

- Works in conjunction with RTP.
- Each participant in RTP session periodically transmits RTCP control packets to all other participants.
- Each RTCP packet contains sender and/or receiver reports
  - report statistics useful to application
- Statistics include number of packets sent, number of packets lost, interarrival jitter, etc.
- Feedback can be used to control performance
  - Sender may modify its transmissions based on feedback

RTCP - Continued



- For an RTP session there is typically a single multicast address; all RTP and RTCP packets belonging to the session use the multicast address.

- RTP and RTCP packets are distinguished from each other through the use of distinct port numbers.

- To limit traffic, each participant reduces his RTCP traffic as the number of conference participants increases.

RTCP Packets

Receiver report packets:

- fraction of packets lost, last sequence number, average interarrival jitter.

Sender report packets:

- SSRC of the RTP stream, the current time, the number of packets sent, and the number of bytes sent.

Source description packets:

- e-mail address of sender, sender's name, SSRC of associated RTP stream.
- Provide mapping between the SSRC and the user/host name.

Synchronization of Streams

- RTCP can synchronize different media streams within a RTP session.
- Consider videoconferencing app for which each sender generates one RTP stream for video and one for audio.
- Timestamps in RTP packets tied to the video and audio sampling clocks
  - not tied to the wall-clock time
- Each RTCP sender-report packet contains (for the most recently generated packet in the associated RTP stream):
  - timestamp of the RTP packet
  - wall-clock time for when packet was created.
- Receivers can use this association to synchronize the playout of audio and video.

RTCP Bandwidth Scaling

- RTCP attempts to limit its traffic to 5% of the session bandwidth.

Example

- Suppose one sender, sending video at a rate of 2 Mbps. Then RTCP attempts to limit its traffic to 100 Kbps.
- RTCP gives 75% of this rate to the receivers; remaining 25% to the sender
- The 75 kbps is equally shared among receivers:
  - With R receivers, each receiver gets to send RTCP traffic at 75/R kbps.
- Sender gets to send RTCP traffic at 25 kbps.
- Participant determines RTCP packet transmission period by calculating avg RTCP packet size (across the entire session) and dividing by allocated rate.