Computer Networks
-Multimedia communication, streaming-

College of Information Science and Engineering
Ritsumeikan University

W12 short test (1)

- What is the difference between full-duplex, half-duplex, and simplex links?
  - Full duplex: can send and receive data simultaneously (so two-directional, at the same time), Half-duplex: can send and receive data, but not simultaneously (so two-directional, but not at the same time), Simplex: can send data but can't receive (one-directional).
- Give two examples of twisted pair type of cables!
  - Cat 5e UTP, Cat 7e
- What are the advantages of fiber optics for guided transmission media?
  - Huge bandwidth, less repeaters are needed compared to copper. Fiber is not affected by power surges, electromagnetic interference, power failures, corrosive chemicals in the air. Fiber is thin and lightweight: useful in situations when there are space or weight restrictions.

W12 short test (2)

- What are the three main components of fiber optics?
  - light source: a pulse of light indicated a 1 bit, the absence of light refers to a 0 bit, transmission medium: the transmission medium is an ultra-thin fiber of glass, detector: it generates an electrical pulse when light falls on it
- What is the difference between baseband transmission and passband transmission?
  - To send digital information, analog signals must be represented in bits. In baseband transmission, the signal occupies frequencies from zero up to a maximum that depends on the signaling rate, there is no shift to higher frequency ranges. Schemes which regulate the amplitude, phase, or frequency of a carrier signal to convey bits are called passband transmission schemes (so the baseband frequencies are shifted). Baseband is mainly used for wired, and passband is used for wireless (but not exclusively).

W12 short test (3)

- What is multiplexing and what are the three main ways to multiplex?
    - Multiplexing is about sharing lines among many signals; managing multiple conversations on the same transmission medium at the same time without those interfering with one another. Practically speaking, it allows a single physical medium to carry multiple simultaneous transmissions.

W12 recap (1)

- The mobile phone system is used for wide area voice and data communication
- In mobile phone systems, a geographic region is divided up into cells
- At the center of each cell is a base station to which all the telephones in the cell transmit
- 1G technology: analog voice, 2G technology: digital voice + GSM standard
- Data traffic began to exceed voice traffic on the fixed network
  - 3G: digital voice and data, 4G is a completely packet-switched network technology
  - even higher data rates and lower latency for 5G by improving the area capacity
- The main wireless LAN standard for over two decades has been the 802.11 (infrastructure and ad hoc modes)

W12 recap (2)

- For wireless, the data link layer is split into two or more sublayers
  - the MAC sublayer determines how the channel is allocated
  - the logical link control sublayer's job is to unify the different 802 variants for the network layer
- All of the 802.11 techniques use short-range radios to transmit signals in either the 2.4GHz or the 5GHz frequency bands
  - 5GHz has a shorter range, but the 2.4GHz band tends to be more crowded
  - different generations, but most modern mobile devices use the 802.11ac
- 802.11 tries to avoid collisions with CSMA/CA, and solve the hidden and exposed terminal problem using the NAV field in each frame

W12 recap (3)

- Clients should not waste power when they have neither information to send nor to receive: variety of power saving methods
- The association service is used by mobile stations to connect themselves to APs, and the data delivery service lets stations transmit and receive data
- Stations must also authenticate before they can send frames via the AP (nowadays mostly WPA2)

Agenda

- Streaming audio and video
- Streaming stored media
- Real-time streaming
- Summary

Background

- **Multimedia** usually means two or more continuous media (e.g., video and audio)
- Since about 2000, real-time audio and real-time video traffic has started to grow (streaming)
- Must be played out at some predetermined rate to be useful (different from normal Web traffic)
- Computers became more powerful, with microphones and cameras
- Increasing bandwidth: ISPs can carry huge levels of traffic across their backbones
- The telephone service is expensive: carry voice traffic over the Internet
- Video takes up a large amount of bandwidth: no problem with broadband access
- There is enough bandwidth, but the key issue for streaming applications is network delay
- Requirements for QoS differ (absolute delay vs. jitter)

Digital audio

- An audio wave is a one-dimensional acoustic (pressure) wave
- The (usual) frequency wave of the human ear runs from 20Hz to 20,000Hz, perceiving loudness logarithmically (in decibels)
- The human ear is sensitive to sound variations lasting only a few milliseconds: even a really short jitter affects the sound quality (unlike with image quality - eyes)
- Digital audio is the digital representation of an audio wave that can be used to recreate it: **ADC** (Analog-to-Digital Converter)
- Taking digital values to produce an analog electrical voltage: **DAC** (Digital-to-Analog Converter)
- A speaker can covert the analog signal to acoustic waves so people can hear sounds

Audio compression (1)

- Audio is usually compressed to reduce bandwidth needs and transfer times
- Compression systems require two algorithms
  - **encoding**: compressing the data at the source
  - **decoding**: decompressing it at the destination
- Asymmetry 1: the encoding can be slow and require expensive hardware, but the decoding need to be fast and require relatively inexpensive hardware
- Asymmetry 2: the decoded output is not exactly equal to the original input, it is **lossy** (vs. **lossless**)

Audio compression (2)

- **MP3** (MPEG audio layer 3) and **AAC** (Advanced Audio Coding) carried in MP4 files
- Audio compression can be done in two ways
- **Waveform coding** where the signal is Fourier transformed into its frequency components, then the amplitude of each component is encoded
- **Perceptual coding** which is about encoding a signal in such a way so it sounds the "same" to the human ear (e.g., MP3 and AAC)

Digital video

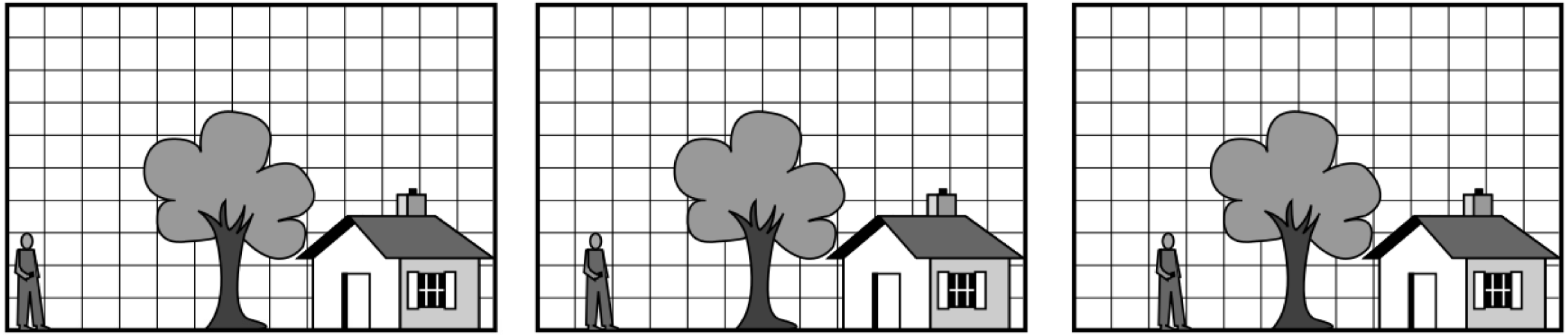
- If an image appears on the retina, it is retained for a few milliseconds
  - if the images are drawn at 50 images/sec, the eye does not notice that it is looking at discrete images
- Digital representation of video: sequence of frames, each consisting of a rectangular grid of picture elements: **pixels** (1280x720 - 720p, 1920x1080 - 1080p, 3840x2160 - 4K, 7680x4320 - 8K)
- Most systems use 24 bits per pixel, with 8 bits for the RGB components
- PAL used to be 25 frames/sec, NTSC 30 frames/sec
- **Interlacing**: to reduce the required bandwidth to broadcast television signals, frames are divided to two fields broadcasted alternately (one with odd, one with even numbered rows), so 25 frames/sec=50 fields/sec
- **Progressive video**: modern video just sends entire frames in sequence, 50 frames/sec (PAL) or 60 frames/sec (NTSC)

Video compression (1)

- Video compression is even more critical than audio to send video over the Internet
- Worldwide standard from the **MPEG** (Motion Picture Experts Group)
    - every few seconds a complete video frame is transmitted (compressed similarly to the JPEG algorithm for still pictures)
    - then, the transmitter sends out differences between the current frame and the base (full) frame it most recently sent out
- **JPEG** (Joint Photographic Experts Group): instead of working with the RGB components, it converts the image into luminance (brightness) and chrominance (color) components (fewer bits is used to encode the chrominance)
    - image is broken up to 8x8 pixel blocks, all processed separately
    - luminance and chrominance are Fourier transformed to get the spectrum, and high-frequency amplitudes are discarded
    - the more amplitude are discarded, the fuzzier the image (and smaller)
    - finally, lossless compression techniques are applied

Video compression (2)

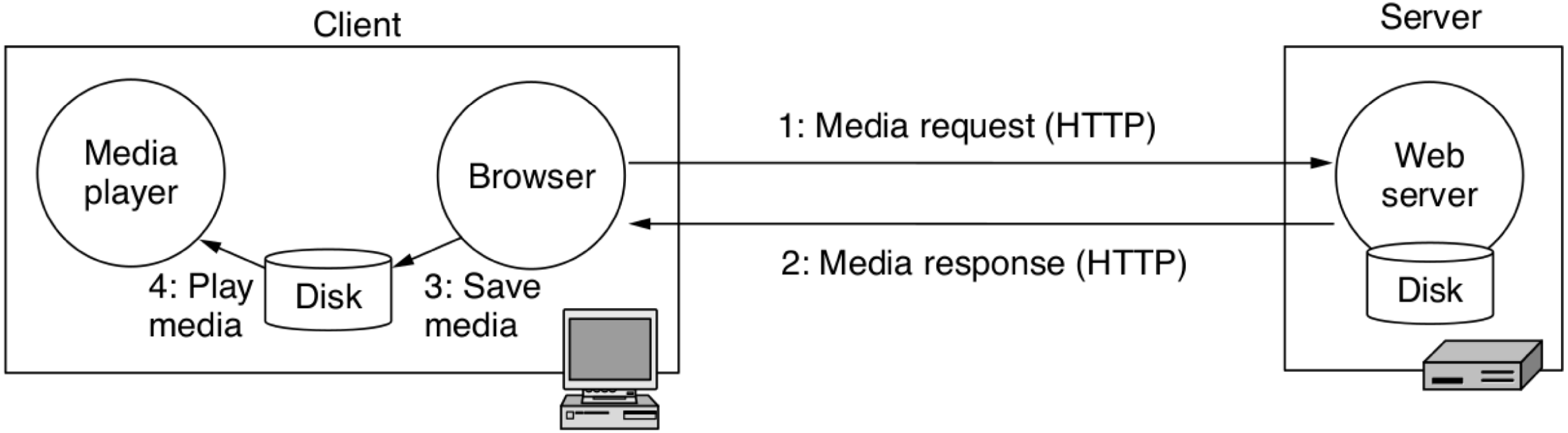
- With MPEG it is very similar to the usual JPEG method, but only blocks that differ from the base frame are transmitted



- Three different kinds of frames:
  - I (Intracoded) frames that are self-contained compressed still images
  - P (Predictive) frames that are difference with the previous one
  - B (Bidirectional) frames that code differences with the next I-frame
- One of the most recent formats is MPEG-4, or MP4 (standard H.264, but H.265 is already up to 8K)

Over the Web via downloads

- **VoD** (Video on Demand): streaming a video that is already stored on a server (e.g., watching a movie on Netflix)
- The easiest way is just make the file (video, audio) available as a pre-encoded file, and let the browser download it
- No real-time network issue from the "streaming" standpoint, but the entire video must be transmitted before the movie starts

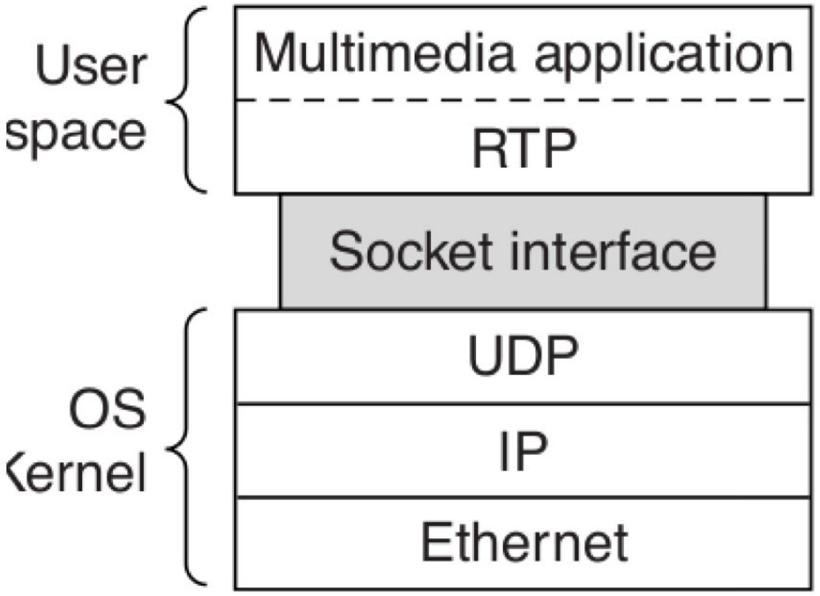


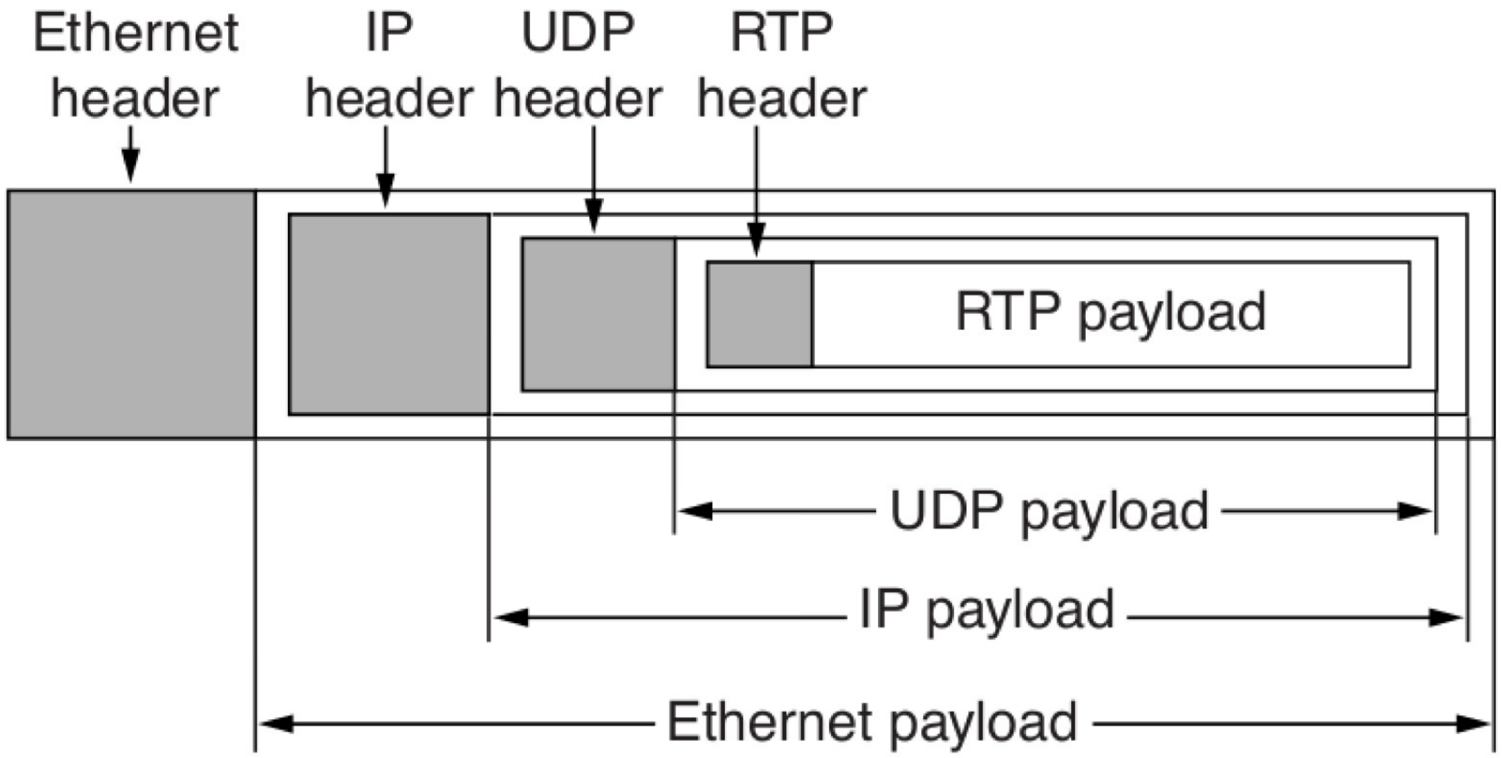
Media player designed for streaming, handling errors

- A media player designed for streaming is needed:
  - manage the user interface: buttons, sliders, panels, skins, etc.
  - decrypt the file: commercial videos are encrypted to prevent piracy
  - handle transmission errors, decompress the content, eliminate jitter
- Dealing with errors depends on weather a TCP-based transport like HTTP is used or UDP-based like **RTP** (Real-time Transport Protocol)
  - TCP already provides reliability by using retransmissions (but uncertain and variable delays - jitter problem)
  - no retransmissions with UDP: packet loss due to congestion or transmission errors mean that some of the media does not arrive
  - possible to use FEC, such as encoding the file with some redundancy (e.g., with Hamming code), but redundancy makes files bigger
  - another approach is selective retransmission of parts of the stream that are the most "important"

RTP

- RTP is widespread for multimedia applications
- The basic function of RTP is to multiplex several real-time data streams onto a single stream of UDP packets
- Position of RTP in the protocol stack and packet nesting:



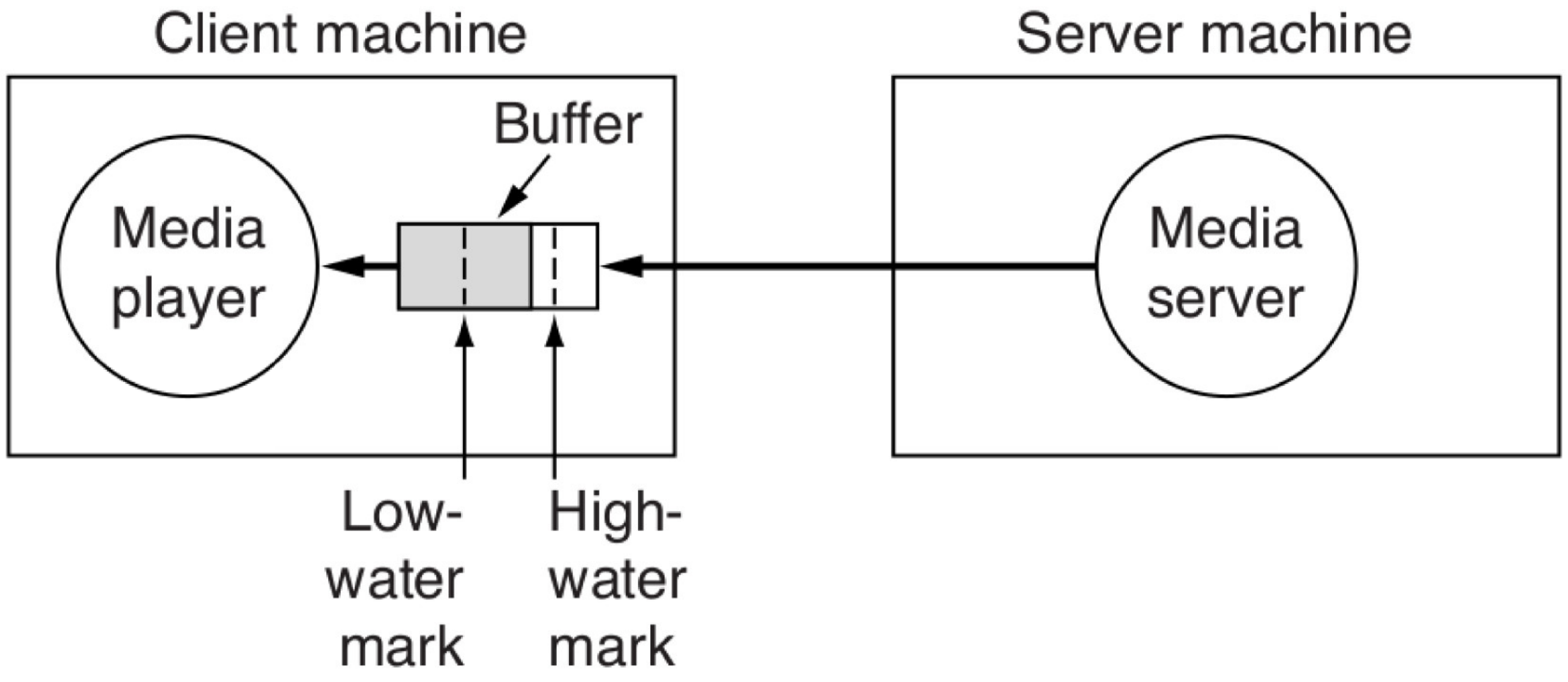


RTCP

- **RTCP** (Real-time Transport Control Protocol) it was defined along RTP, but it handles feedback, synchronization, and the user interface
- Provide feedback on delay, jitter, bandwidth, congestion, and other network properties
  - can be used to increase the data rate when the network is functioning well and to cut back the data rate when there is a trouble provide the best quality possible under the current circumstances
  - also provides a way for naming various sources (e.g., display who is talking at the moment in ASCII text)

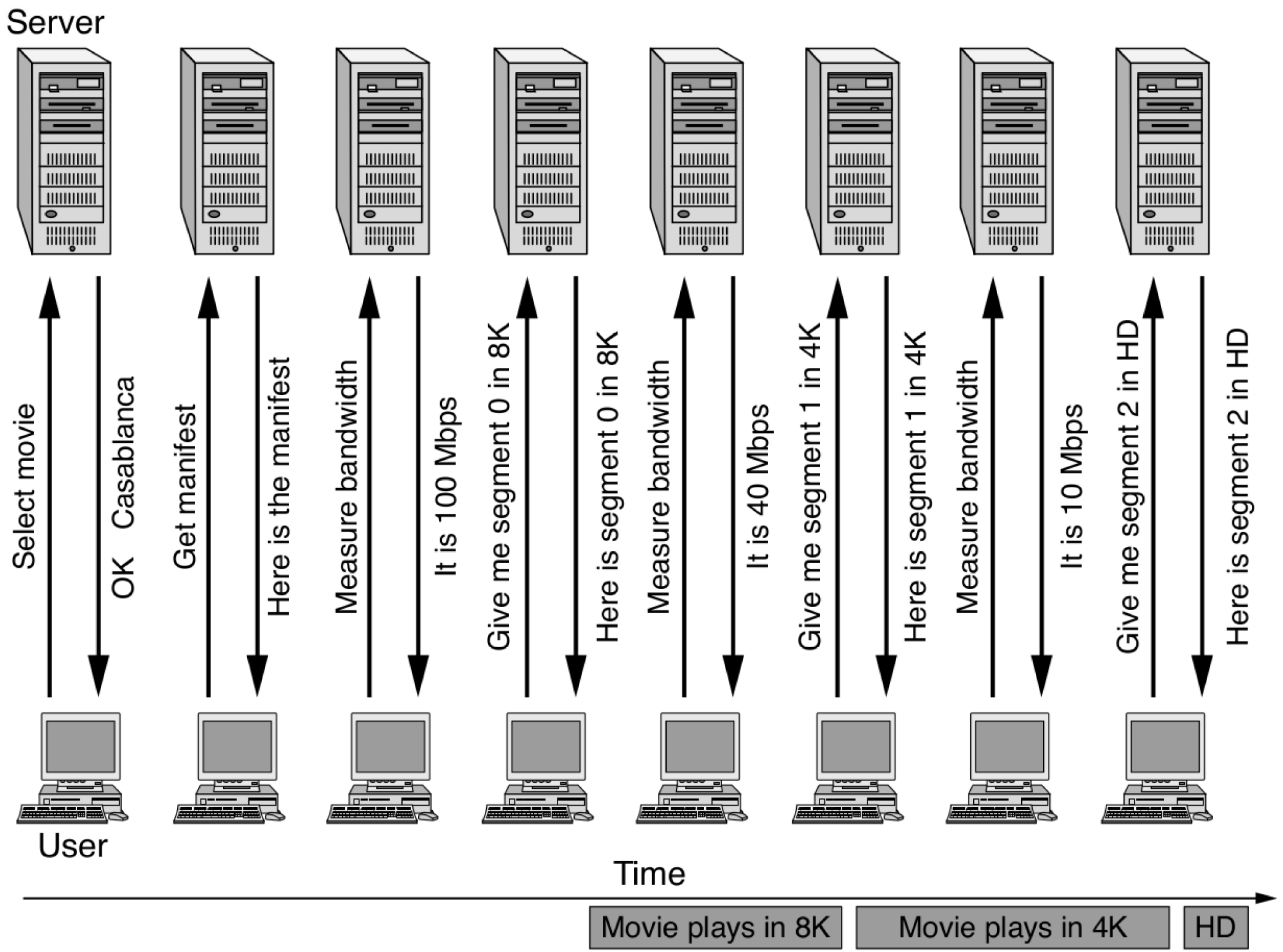
Decompress the content, eliminate jitter

- How to decode media if the network protocol does not correct transmission errors?
- Many compression schemes, later data cannot be decompressed until earlier data has been decompressed (later data is encoded relative to the earlier data)
- The general solution that all streaming systems use is a playout buffer, where the idea is that the data should arrive regularly enough that the buffer is never completely emptied



DASH and HLS

- How to handle the problem of the large variety of devices?
  - encode it in (almost) every combination of screen resolution and frame rates
- What to do when there is a network congestion, and the varying bandwidth cannot always support the full resolution?
- **DASH** (Dynamic Adaptive Streaming over HTTP): the streaming server first encodes the movies at multiple resolutions and frame rates, and has them stored in many files (each stores few secs of audio and video)
- Apple's **HLS** (HTTP Live Streaming) supported by browsers, game consoles, TVs: similar idea to DASH, with additional features (e.g., fast forward and backward)
- DASH is agnostic and allows inserting ads to the video stream, while HLS works only with Apple supported algorithms (H.264 and H.265 are ok), and no ads in the stream


Server
Select movie
OK Casablanca
Get manifest
Here is the manifest
Measure bandwidth
It is 100 Mbps
Give me segment 0 in 8K
Here is segment 0 in 8K
Measure bandwidth
It is 40 Mbps
Give me segment 1 in 4K
Here is segment 1 in 4K
Measure bandwidth
It is 10 Mbps
Give me segment 2 in HD
Here is segment 2 in HD
User
Time
Movie plays in 8K
Movie plays in 4K
HD

Streaming over IP

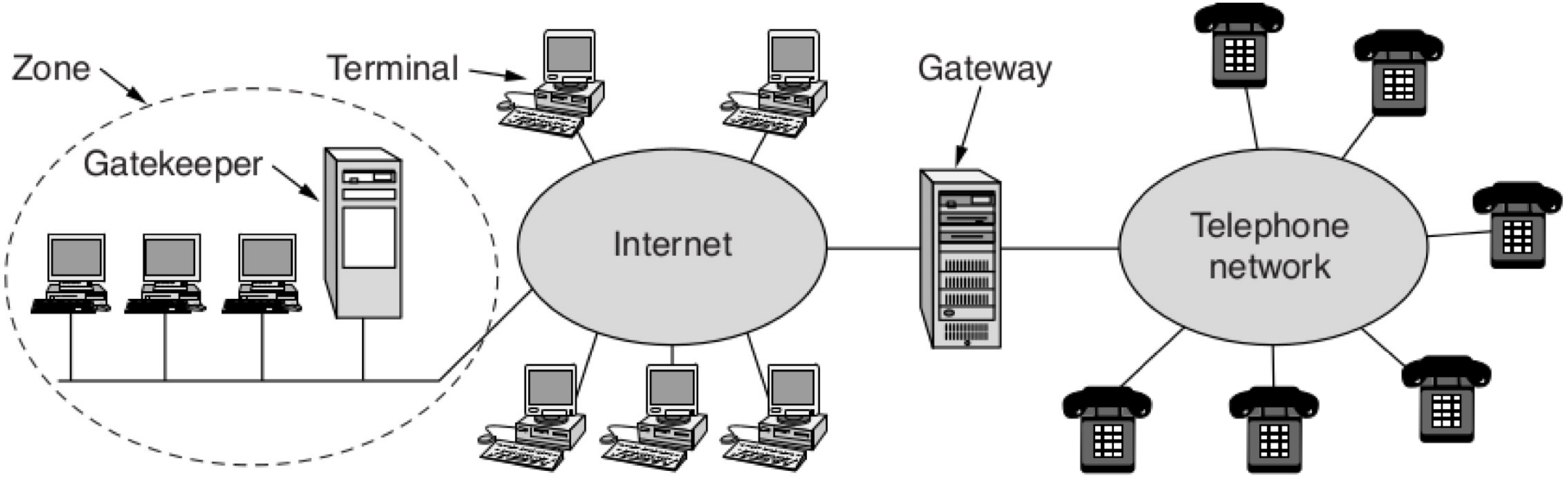
- Live streaming using **IPTV** (IP TeleVision) is widespread nowadays
- Cable providers do live streaming over IP to build their own broadcast systems, used by radios, zoos, etc.
- **Podcast**: record programs to disk, then viewers connect to he server's archives and download it
- When streaming live events, news broadcast, etc. when the user logs onto the site covering the live event, no video is shown until the buffer fills
  - in this case, the only difference between a movie and the live stream is that the server cannot load a lot of data even if the connection is fast

Voice over IP

- We use the Internet also to transmit telephone or video calls where buffering is not possible (e.g., Skype, Zoom, Line, etc.)
- Biggest difference streaming a movie and Voice over IP/Internet telephony is the need for low latency that is difficult to achieve
    - up to 150 msec is the acceptable one-way window, amount of buffering is kept small
    - using UDP makes more sense as TCP retransmissions introduce at least one round-trip delays
    - the physical distance itself can be a problem: transmission delays as each IP router stores and forwards a packet along the long distance
    - using larger packets is not the best: voice over IP systems use short packets to reduce latency at the cost of bandwidth efficiency
    - additional software overheads (e.g., encoder and decoder must be fast)

H.323 (1)

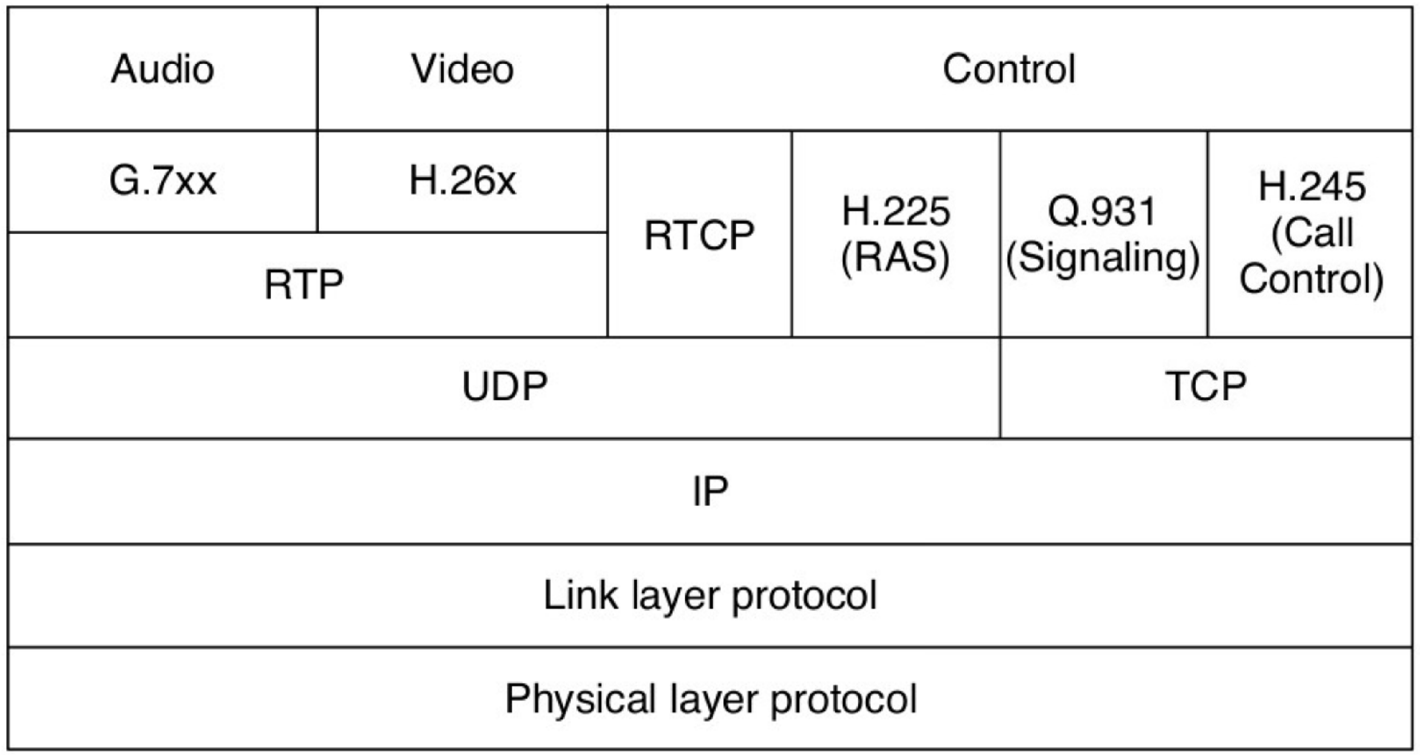
- **H.323** "Packet-based Multimedia Communications Systems" was the basis for the first widespread Internet conferencing systems
- Is more of an architectural overview of Internet telephone than a specific protocol: references a large number of protocols for speech coding, call setup, signaling, data transport, etc.
- The *gateway* connects the Internet to the telephone network, the communicating devices are the *terminals*, and the *gatekeeper* controls the end points under its *zone*

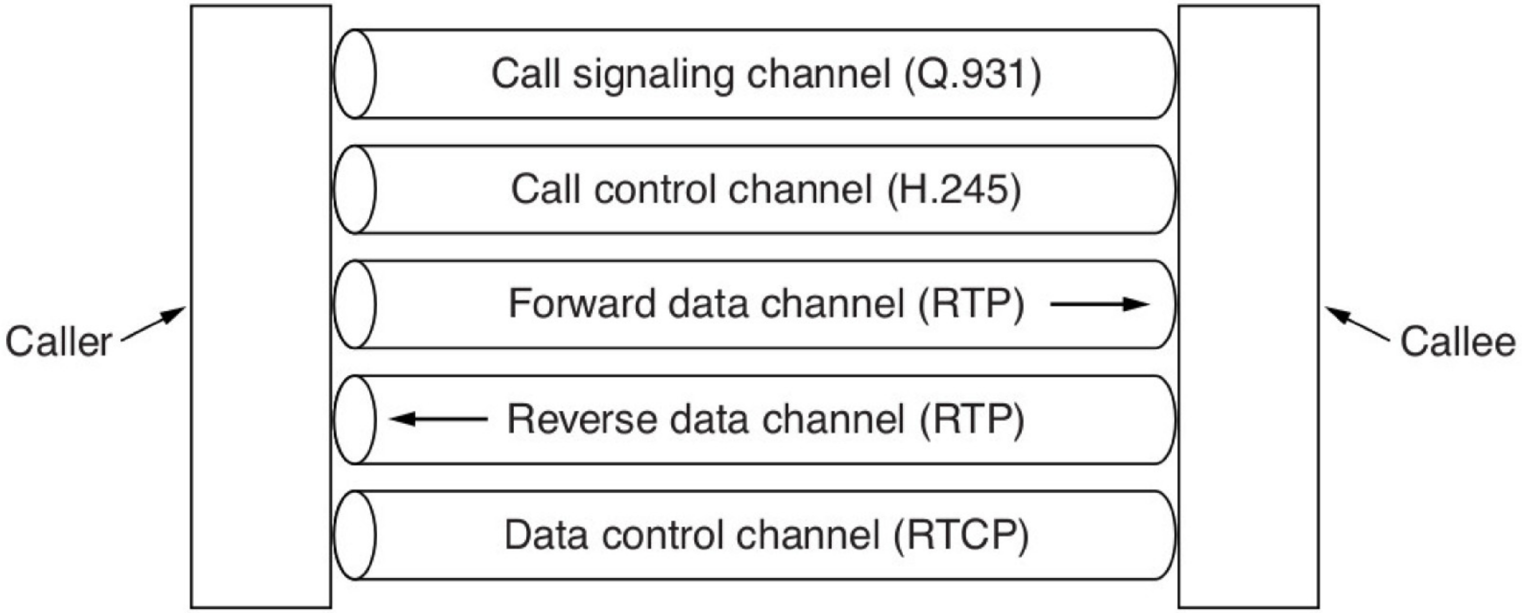


H.323 (2)

- Protocol for encoding and decoding audio and video: G.711 for audio, for example H.264 for video
- Protocol to allow the terminals to negotiate which compression algorithms to use, and the bit rate: H.245
- Control the RTP channels: RTCP
- Establish and release connections, providing dial tones, ringing sounds, etc.: Q.931
- Terminals talking to the gatekeeper: H.225 to allow terminals to join and leave the zone, request and return bandwidth, provide status updates, etc.
- Protocol for the actual data transmission: RTP over UDP (managed by RTCP)

H.323 protocol stack and logical channels





SIP

- **SIP** (Session Initiation Protocol) describes how to set up Internet telephone calls, video conferences, and other multimedia connections
  - a more modular way for voice over IP compared to H.323
  - H.323 is a complete protocol suite, but SIP is a single module - interworks well with existing Internet applications
- Can establish two-party sessions, multiparty sessions and multicast sessions
- May contain audio, video, data
- SIP handles setup, management, and termination of sessions
- It is an application layer protocol and can run over UDP or TCP

Item	H.323	SIP
Designed by	ITU	IETF
Compatibility with PSTN	Yes	Largely
Compatibility with Internet	Yes, over time	Yes
Architecture	Monolithic	Modular
Completeness	Full protocol stack	SIP just handles setup
Parameter negotiation	Yes	Yes
Call signaling	Q.931 over TCP	SIP over TCP or UDP
Message format	Binary	ASCII
Media transport	RTP/RTCP	RTP/RTCP
Multiparty calls	Yes	Yes
Multimedia conferences	Yes	No
Addressing	URL or phone number	URL
Call termination	Explicit or TCP release	Explicit or timeout
Instant messaging	No	Yes
Encryption	Yes	Yes
Size of standards	1400 pages	250 pages
Implementation	Large and complex	Moderate, but issues
Status	Widespread, esp. video	Alternative, esp. voice

W13 summary (1)

- Real-time audio and video must be played out at some predetermined rate to be useful (different from normal Web traffic)
- There is enough bandwidth, but the key issue for streaming applications is network delay
- Digital audio is the digital representation of an audio wave that can be used to recreate it: ADC
- Taking digital values to produce an analog electrical voltage: DAC
- Audio is usually compressed to reduce bandwidth needs and transfer times (loss vs lossless encoding and decoding)
- Audio compression can be done in two ways: waveform coding and perceptual coding
- Worldwide standard for video compression comes from the MPEG, using the JPEG algorithm and three different kind of frames (I, P, B)

W13 summary (2)

- VoD: streaming a video that is already stored on a server
- A media player designed for streaming is needed: manage user interface, decrypt file, handle transmission errors, decompress content, eliminate jitter
- Dealing with errors depends on weather a TCP-based transport like HTTP is used or UDP-based like RTP
- The basic function of RTP is to multiplex several real-time data streams onto a single stream of UDP packets
- RTCP handles feedback, synchronization, and the user interface
- Streaming using DASH: the streaming server first encodes the movies at multiple resolutions and frame rates, and has them stored in many files
- Apple's HLS supported by browsers, game consoles, TVs: similar idea to DASH, with additional features

W13 summary (3)

- Live streaming using IPTV is widespread nowadays
- When streaming live events, news broadcast, etc. when the user logs onto the site covering the live event, no video is shown until the buffer fills
- Biggest difference streaming a movie and Voice over IP/Internet telephony is the need for low latency that is difficult to achieve
- Internet conferencing systems using H.323 and SIP
  - both compatible with the Internet
  - SIP has a modular architecture
  - H.323 includes a full protocol stack
  - both use RTP/RTCP and supports encryption