Chapter 28

Multimedia and IP Telephony

A Few Definitions

- **Multimedia** combines two or more forms of information, such as
  - Photos and music
  - Audio and video
- **Real-time** refers to information that must be presented in a predetermined timed sequence, such as
  - Audio
  - Video
- An individual source provides one particular sequence of real-time information

A Few Definitions (continued)

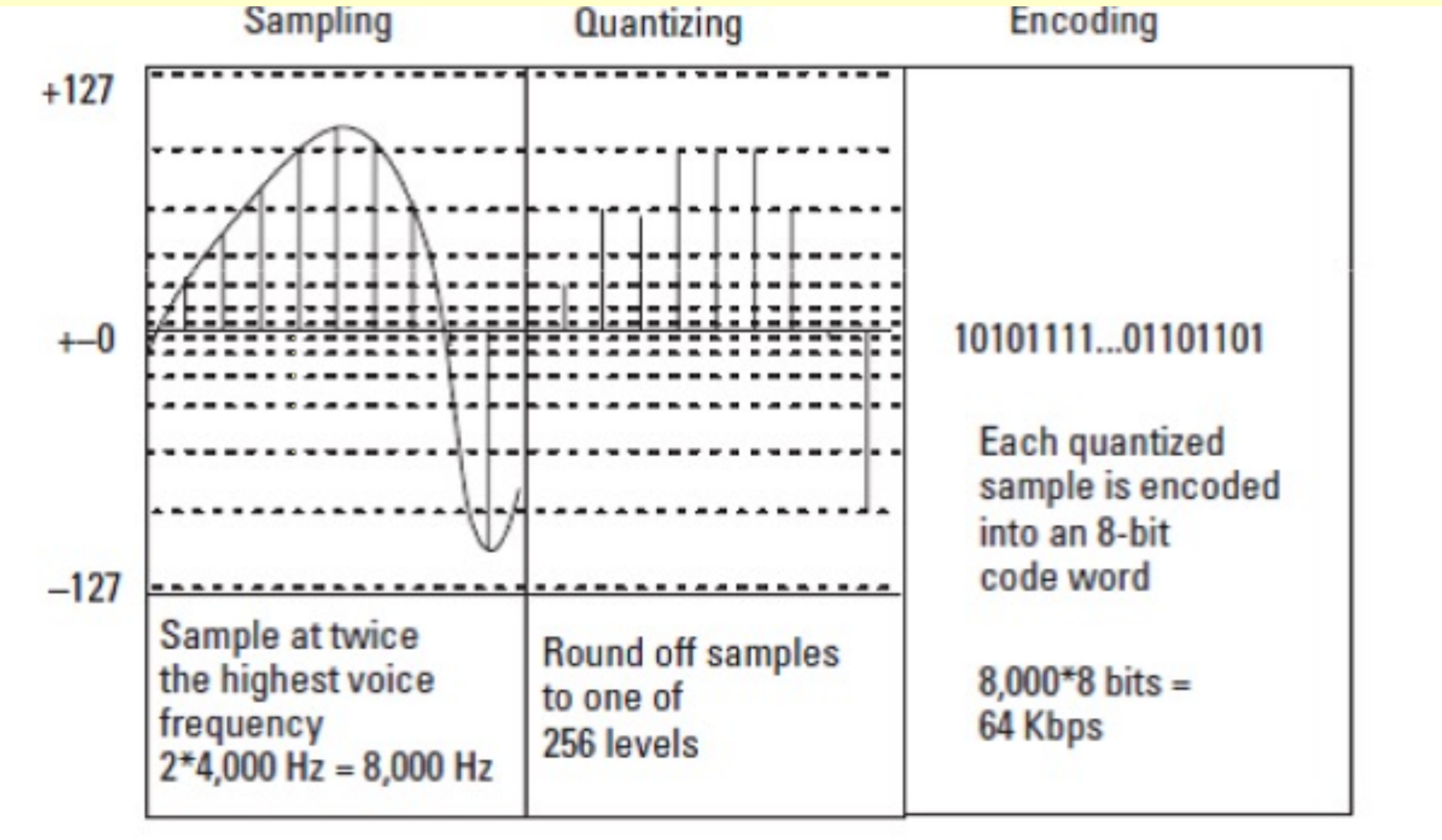
- **Sampling** is converting analog media streams to a digital form
- **Sample rate** refers to the rate at which real-time information has been converted (e.g., audio sampled 8000 times per second)
- **Quantization** is the process of constraining an input from a continuous value (such as the real numbers) to a discrete set (such as the integers)
- **Playback** refers to the output of real-time information for a user (e.g., video display or audio output)
- **Synchronization** refers to the coordination of playback information from multiple sources (e.g., a movie requires synchronization between audio and video)

Real-Time Sample Rates

- Each source of real-time data can choose a sample rate and encoding
- Examples
  - A video stream might contain 30 frames per second, with an encoding that uses compression
  - An audio stream might contain 8000 audio samples per second using a PCM encoding
- Important concept

  Because each source of real-time information can choose a sample rate, playback and synchronization must know the sample rate and encoding that was selected

PCM encoding



Transfer Of Streamed Real-Time Data

- Source
  - Samples information at regular intervals
  - Generates data continuously
  - Prepares data for transmission
- An ideal transmission channel would
  - Accepts input at rate source produces
  - Delivers output at same rate as input

Quantitative Network Performance Needed For Real-Time Streaming

- QoS type: Constant Bit Rate (CBR)
  - Throughput sufficient to accommodate sender's data rate (known in advance)
  - Latency within a specified bound, usually 200 msec
  - Jitter of zero or near-zero

Output Buffering

- Especially important in a packet transmission system
- Combines multiple samples into a single transmission
- Advantage
  - Increases transmission efficiency
- Disadvantage
  - Introduces delay

Buffering Example

- Consider PCM audio
- One eight-bit audio sample taken every 125 μseconds
- Ethernet has 1500 octet payload
- Waiting to fill an entire frame takes

  125×10^{-6} seconds/byte $\times$ 1500 bytes = 0.188 seconds

- Filling a packet incurs delay at the source

Buffering Compromise

- Choose buffer size according to application
- Example: send 128 audio samples in each packet
- Trade-offs
  - Packet size is larger than one sample per packet, but generates more packets than absolutely necessary
  - Header overhead is a smaller percentage of total bits than with one sample per packet, but a greater percentage than for larger packets
  - Latency is better than with many samples per packet, but not as good as with one sample per packet

Streaming Of Real-Time Data Across The Internet

- We don’t have a perfect channel (with CBR QoS), thus we must handle
  - Lost packets
  - Duplicated packets
  - Variance in delay (jitter)
  - Packets delivered out of order
- Key facts
  - Conventional retransmission is (normally) useless
  - Jitter is unavoidable

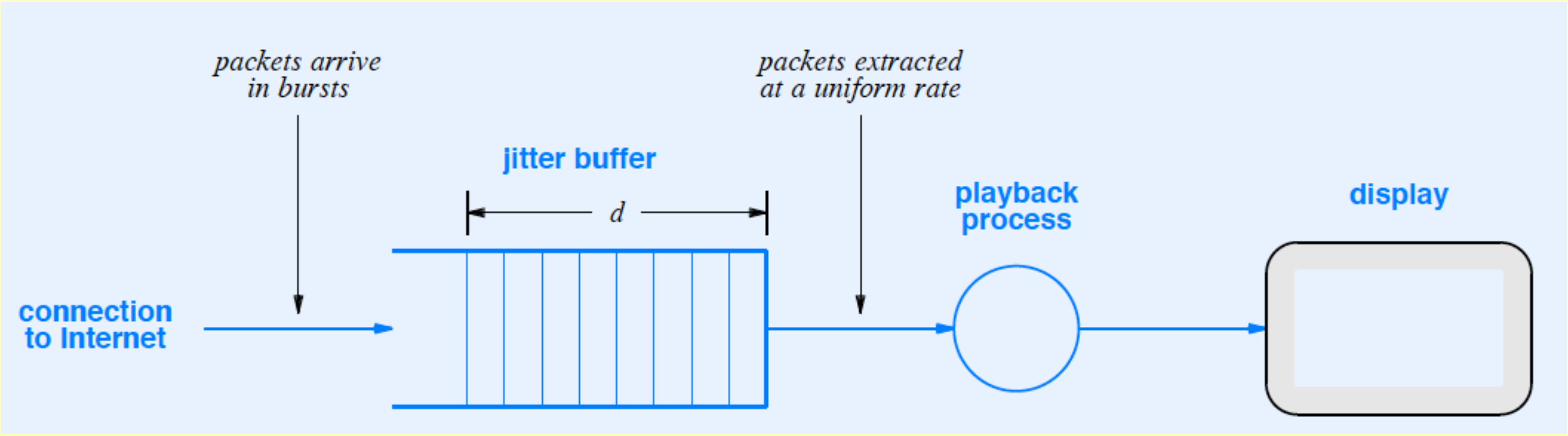
Two Useful Techniques

- Timestamps
  - Provided by sender
  - Assigned to each piece of data
  - Allow receiver to know when data should be played
  - Use relative values to avoid need for clock synchronization
- Jitter buffer
  - Used by receiver
  - Accommodates small variance in delay

Jitter Buffer

- Used by receiver to assemble incoming real-time data
- Timestamp on an item determines where item is placed in the playback sequence
- General principle: ensure information will be available in time to play without delay
- Trick: to compensate for maximum jitter of d, delay playback for d time units
- Result: jitter buffer holds just enough data so playback can proceed uninterrupted

Illustration of a Jitter Buffer in the Receiver



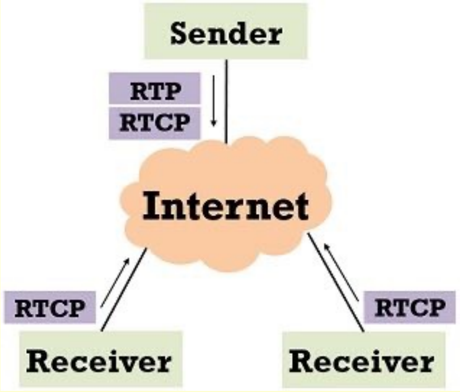
- During normal operation, playback can continue for d time units while waiting for delayed packets

Real-Time Transport Protocol (RTP)

- Widely used for voice and video
- Despite the name, not really a transport protocol
- Does not contain a jitter buffer and does not control playback
- It is to providing three basic mechanisms:
  - Sequence number on each packet that allows a receiver to handle loss and out-of-order delivery
  - Timestamp used for playback of the data
  - Series of source identifiers that tell a receiver the source(s) of the data

RTP Details

- Allows sender and receiver to choose sample rate and encoding
- Specifies a header for each message transferred
- Uses UDP for transport
- Separates timestamp from packet sequence number
- Includes a marker bit that allows some frames to be marked
- RTP has a companion protocol known as RTP Control Protocol (RTCP) which provides quality of service feedback to the participants of an RTP session.



Motivation For RTP Design

- Marking
  - Permits differential encoding with a full frame followed by incremental changes
  - Example use: video I-frame followed by B-frames
- Separation of timestamp and packet sequence
  - Means timestamps do not need to be linearly related to packets
  - Allows compression schemes that vary the rate at which data is sent

RTP Header Format

0 1		3		8		16 31
VER	P	X	CC	M	PAYTYPE	SEQUENCE NUMBER
TIMESTAMP						
SYNCHRONIZATION SOURCE IDENTIFIER						
CONTRIBUTING SOURCE IDENTIFIER . . .						

- TIMESTAMP is interpreted by sender and receiver
- PAYTYPE specifies the payload type
- Initial SEQUENCE NUMBER chosen at random
- CONTRIBUTING SOURCE IDENTIFIERS allow sender to mix streams from multiple sources

RTP Payload types

Payload Type Number	Audio Format	Sampling Rate	Throughput
0	PCM mu-law	8 KHz	64 Kbps
1	1016	8 KHz	4.8 Kbps
3	GSM	8 KHz	13 Kbps
7	LPC	8 KHz	2.4 Kbps
9	G.722	8 KHz	48-64 Kbps
14	MPEG Audio	90 KHz	-
15	G.728	8 KHz	16 Kbps

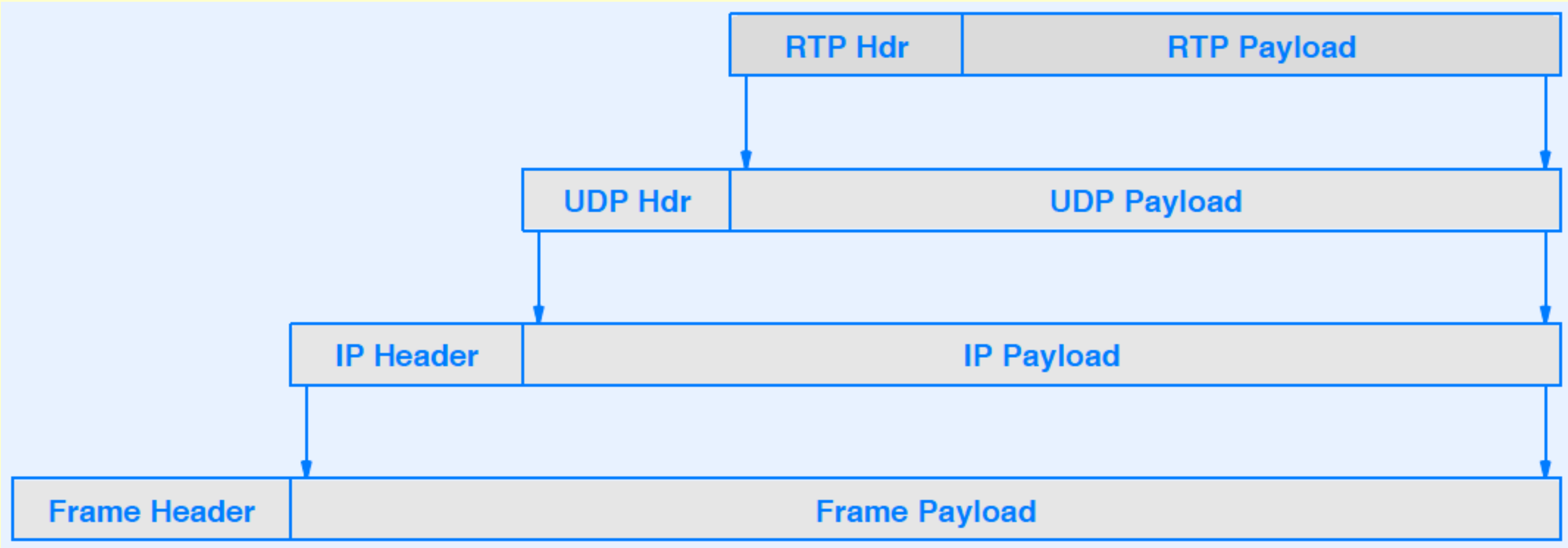
Some audio payload types supported by RTP

Payload Type Number	Video Format
26	Motion JPEG
31	H.261
32	MPEG1 video
33	MPEG2 video

Some video payload types supported by RTP

RTP Encapsulation

- Three levels of encapsulation



- Use of UDP permits sending one multicast instead of multiple unicast copies

Real-Time Streaming Protocol (RTSP)

- IETF standard
  - RFC 2326
- Real-Time Streaming Protocol
  - acts as a « network remote control »
- supports the following operations:
  - retrieval of a media from a server
  - invitation of a media server to a conference
  - recording of a conference

RTSP methods

- Major methods
  - SETUP: server allocates resources for a stream and starts an RTSP session
  - PLAY: starts data tx on a stream
  - PAUSE: temporarily halts a stream
  - TEARDOWN: free resources of the stream, no RTSP session on server any more
- Additional methods
  - OPTIONS: get available methods
  - ANNOUNCE: change description of media object
  - DESCRIBE: get low level descr. of media object
  - RECORD: server starts recording a stream
  - REDIRECT: redirect client to new server
  - SET_PARAMETER: device or encoding control

Session Description Protocol (SDP)

- The Session Description Protocol (SDP) is a format for describing multimedia communication sessions for the purposes of session announcement and session invitation
- The set of properties and parameters is called a *session profile*

Session description
v= protocol version number
o= originator and session identifier
s= session name
u=* URI of description
e=* email address
p=* phone number
c=* connection information
b=* bandwidth information lines)
Time descriptions *("t=" and "r=" lines)*
z=* time zone adjustments)
k=* encryption key
a=* session attribute lines)
Media descriptions *(each one starting by an "m=" line)*

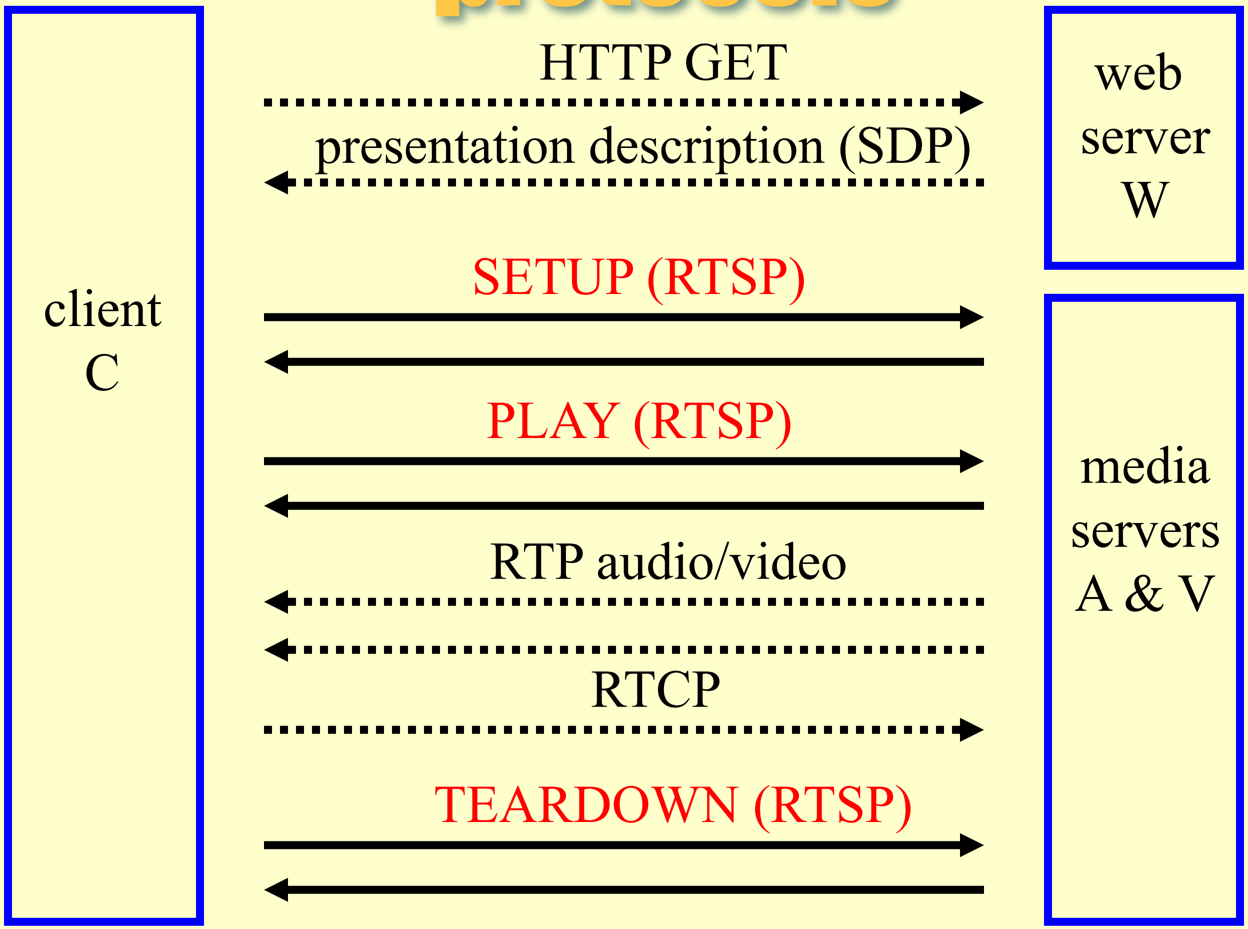
Synchronized Multimedia Integration Language (SMIL)

- Synchronized Multimedia Integration Language is an Extensible Markup Language (XML) markup language to describe multimedia presentations
- It defines markup for timing, layout, animations, visual transitions, and media embedding, among other things. SMIL allows presenting media items such as text, images, video, audio, links to other SMIL presentations, and files from multiple web servers

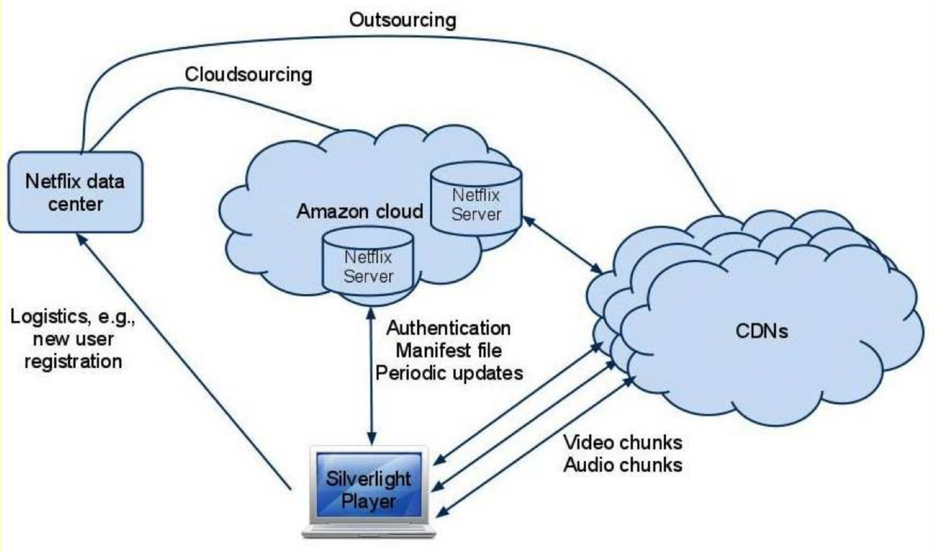
A combined view of the real-time protocols

- the protocols and their application field…
  - **stream description:** **SDP and/or SMIL**
    describe the session and content
  - **stream control:** **RTSP**
    remote control the session
  - **media transport:** **RTP/RTCP**
    send data and metadata
  - **resource reservation (if any!):** **RSVP, DiffServ**
    make sure the communication path offers appropriate guaranties…
    …otherwise Best-Effort transmissions!

RTSP Example with supporting protocols



Real world example - Netflix

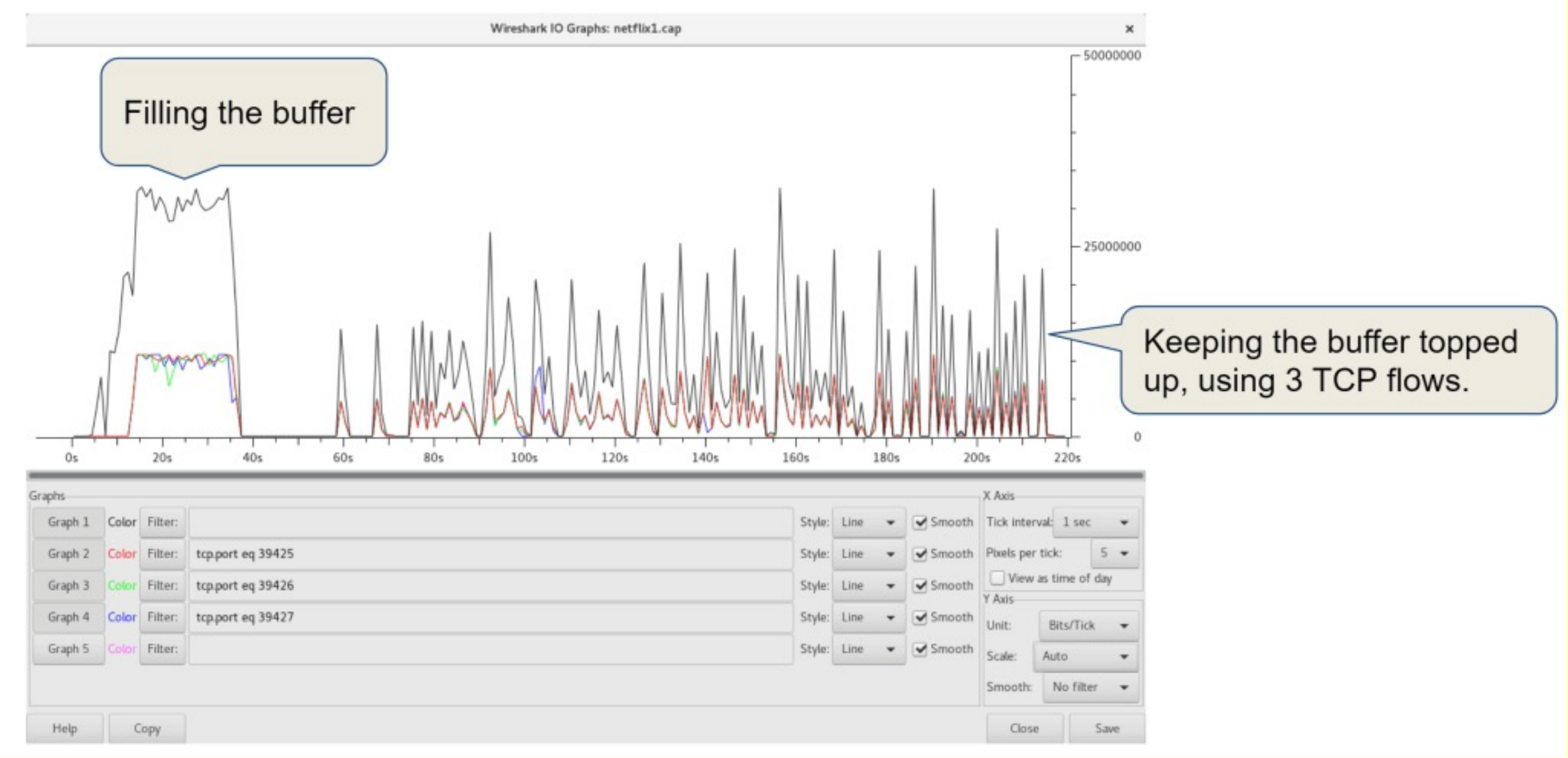


Netflix uses the DASH (*Dynamic Streaming over HTTP*) protocol for streaming. In DASH, each video is encoded at several different quality levels, and is divided into small 'chunks' - video segments of no more than a few seconds in length. The client requests one video chunk at a time via HTTP. With each download, it measures the received bandwidth and runs a *rate determination algorithm* to determine the quality of the next chunk to request. DASH allows the player to freely switch between different quality levels at the chunk boundaries.

Source: "Measurement Study of Netflix, Hulu, and a Tale of Three CDNs" by Vijay K Adhikari et al

Real world example - Netflix

- Measured behaviour of Netflix streaming:



Source: https://preseem.com/2018/04/netflix-behavior-wireless-network/

IP Telephony (VoIP)

IP Telephony

- Known as Voice over IP (VoIP)
- Two groups have created standards
  - International Telecommunications Union (ITU)
  - Internet Engineering Task Force (IETF)
- Standards agree on two basics
  - Audio encoded using Pulse Code Modulation (PCM)
  - RTP used to transfer digitized audio
- Standards disagree on
  - Signalling
  - Public Switched Telephone Network (PSTN) interaction

Signalling

- Telecommunication term for the process of establishing and terminating a call
- Includes
  - Mapping a phone number to a location
  - Finding a route to the called party
  - Recording information used for accounting and billing
  - Handling functions such as call forwarding
- Standard call management facility for the traditional telephone system is known as Signalling System 7 (SS7)

IETF Approach

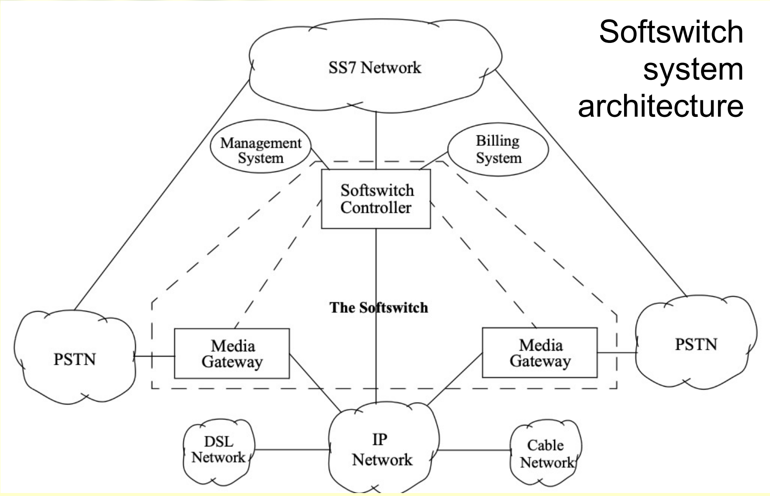
- Known as Session Initiation Protocol (SIP)
- Domain Name System used to map a telephone number to an IP address
- SIP signalling system
  - User agent makes or terminates calls (e.g., an IP phone)
  - Location server consults a database of users, services to which they subscribe, and preferences
  - Proxy server forwards requests and optimizes routing
  - Redirect server handles tasks such as call forwarding and 800-number connections (toll free numbers)
  - Registrar server allows users to register for service

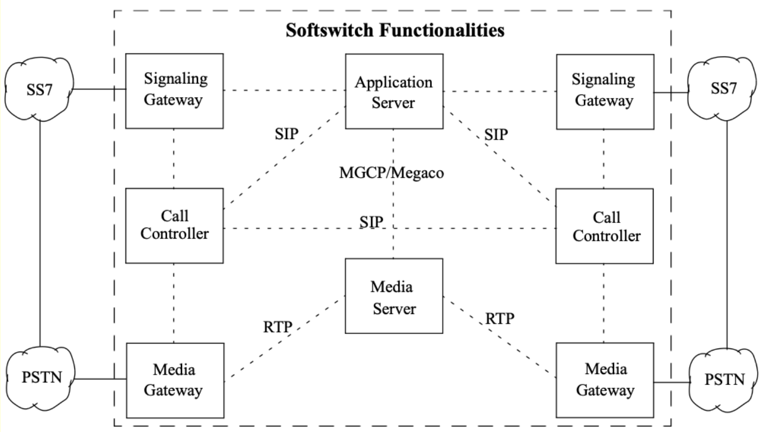
ITU Approach

- Standard is H.323
- Differs substantially from terminology used by SIP
- Terminal provides IP telephone functions and may also include facilities for video and data transmission
- Gatekeeper provides location and signalling functions, and establishes connections to the public switched telephone network (PSTN)
- Gateway interconnects the IP phone system and PSTN, and handles both signalling and media translation
- Multipoint Control Unit (MCU) provides services such as multipoint conferencing

International Softswitch Consortium (ISC)

- Formed by vendors to consolidate terminology from multiple standards and create a single conceptual model
- Defined a list of 10 functions that are sufficient to explain all others
- Invented new terms for each function





List of VoIP Protocols and Layering

Layer	Call Process.	User multimedia	User Data	Support	Routing	Signal Transport
5	H.323 Megaco MGCP SIP	RTP	T.120	RTCP RTSP NTP SDP	ENUM TRIP	SIGTRAN
4	TCP UDP	UDP	TCP	TCP UDP		SCTP
3	IP, RSVP, and IGMP					

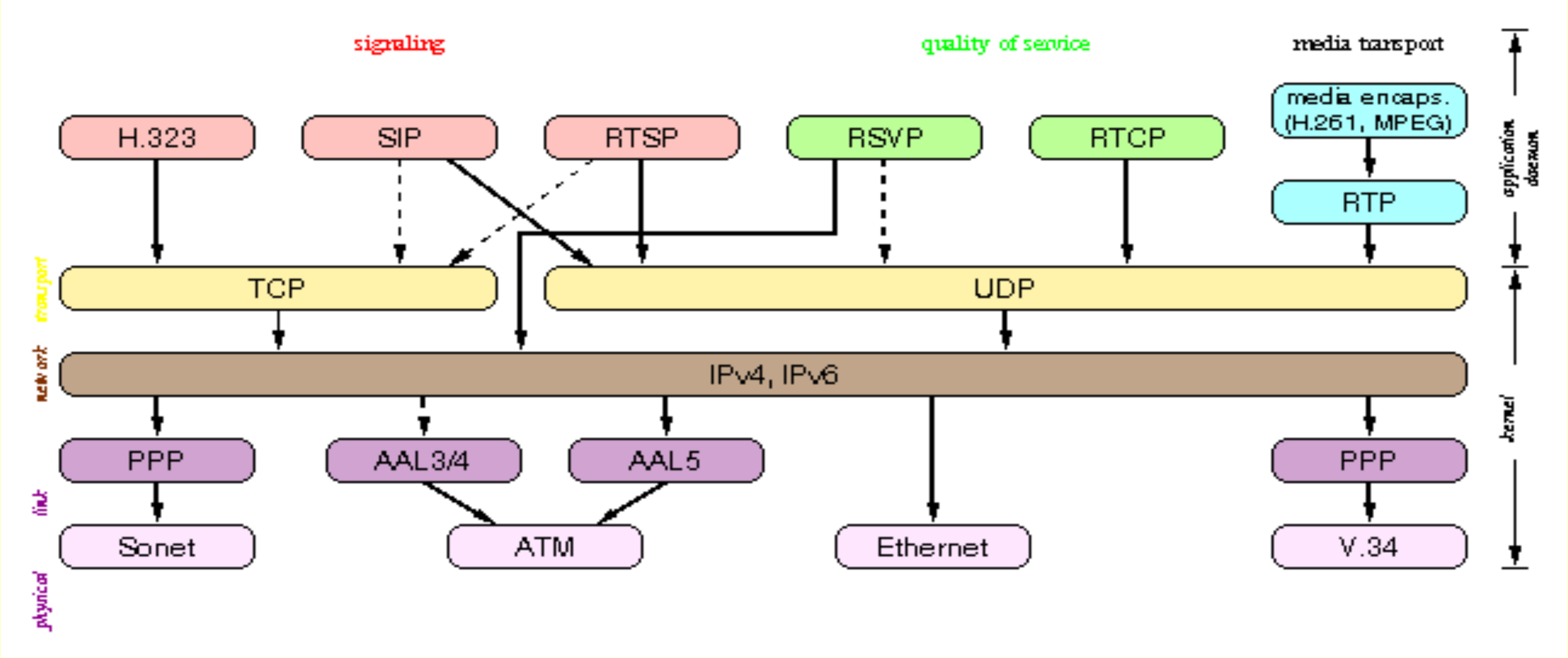
- Each protocol can be complex
- H.323 is an umbrella (see next slide)
- SCTP is sometimes referred to as "next generation TCP - or TCPng

H.323

- Large set of protocols collected together
- Provides voice, video, and data transfer
- Summary of major protocols

<table>
<tr><th>Layer</th><th>Signaling</th><th>Registration</th><th>Audio</th><th>Video</th><th>Data</th><th>Security</th></tr>
<tr><td rowspan="2">5</td><td rowspan="2">H.225.0-Q.931
H.250-Annex G
H.245
H.250</td><td rowspan="2">H.225.9-RAS</td><td>G.711
H.263
G.722
G.723
G.728</td><td>H.261
H.323</td><td rowspan="2">T.120</td><td rowspan="2">H.235</td></tr>
<tr><td colspan="2">RTP, RTCP</td></tr>
<tr><td>4</td><td>TCP, UDP</td><td colspan="3">UDP</td><td>TCP</td><td>TCP, UDP</td></tr>
<tr><td>3</td><td colspan="6">IP, RSVP, and IGMP</td></tr>
</table>

How some of the protocols relate to each other



Telephone Number Mapping And Routing

- Two standards proposed by IETF
  - TRIP relies on location servers to exchange information
  - ENUM (E.164 NUMbers) uses arpa top-level domain in the Domain Name System

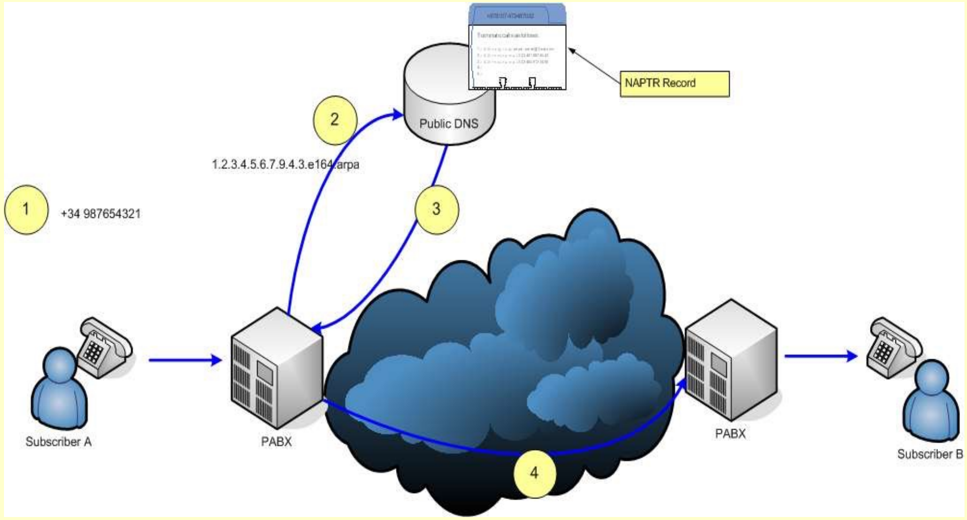
TRIP

- TRIP is a protocol for advertising the reachability of telephony destinations between location servers, and for advertising attributes of the routes to those destinations.
- TRIP's operation is independent of any signalling protocol, hence TRIP can serve as the telephony routing protocol for any signalling protocol.
- It is built similar as BGP, but with the purpose to route telephone calls on internet rather than datagrams.
- In reality, TRIP isn't used anywhere. Telephony routing in the different inter-operator peering solutions is achieved using ENUM

Source: https://www.voip-info.org/trip/

ENUM

- ENUM (E.164 Number to URI Mapping) translates telephone numbers into Internet addresses. You can dial a telephone number and reach a SIP, H.323 or any other Internet Telephony user. This all happens in the background; you don't need to do anything special while calling someone.
- A server with ENUM support will lookup a dialled telephone number in the ENUM tree of the DNS to see if there are alternate ways to set up the call instead of just calling out on the PSTN telephone line. ENUM may contain a reference to a SIP URI, a telephone number to dial, a web page or an e-mail address.
- ENUM example
  - Phone number is 1-800-555-1234
  - Domain name is constructed as the string *4.3.2.1.5.5.5.0.0.8.1.e164.arpa*



Sources:
https://www.voip-info.org/enum/
https://en.wikipedia.org/wiki/Telephone_number_mapping

Summary

- Streamed transfer of real-time data incompatible with Internet's best-effort delivery
- Two approaches
  - Isochronous network
  - Timestamps and jitter buffer
- Real-Time Transport Protocol (RTP) uses timestamps and sequence numbers

Summary (continued)

- Many IP telephony standards proposed
- Connection to PSTN causes debate
- H.323 and SIP standards are most widely used
- ENUM system uses DNS to convert phone number to IP address