Multimedia Networking

Raj Jain

Washington University in Saint Louis

Saint Louis, MO 63130

Jain@wustl.edu

Audio/Video recordings of this lecture are available on-line at:

http://www.cse.wustl.edu/~jain/cse473-21/

Student Questions

Overview

❑ Multimedia Networking Applications

❑ Skype

❑ Real-Time Transport Protocol (RTP)

❑ Session Initiation Protocol (SIP)

Note: This class lecture is based on Chapter 9 of the textbook (Kurose and Ross) and the figures provided by the authors.

Student Questions

Multimedia Applications

Overview

1. Audio Digitization
2. Playout Buffers
3. Streaming Using UDP
4. Streaming Using HTTP

Student Questions

©2021 Raj Jain

Audio Digitization

- **Sampling**: Analog audio signal sampled at constant rate
  - Telephone:
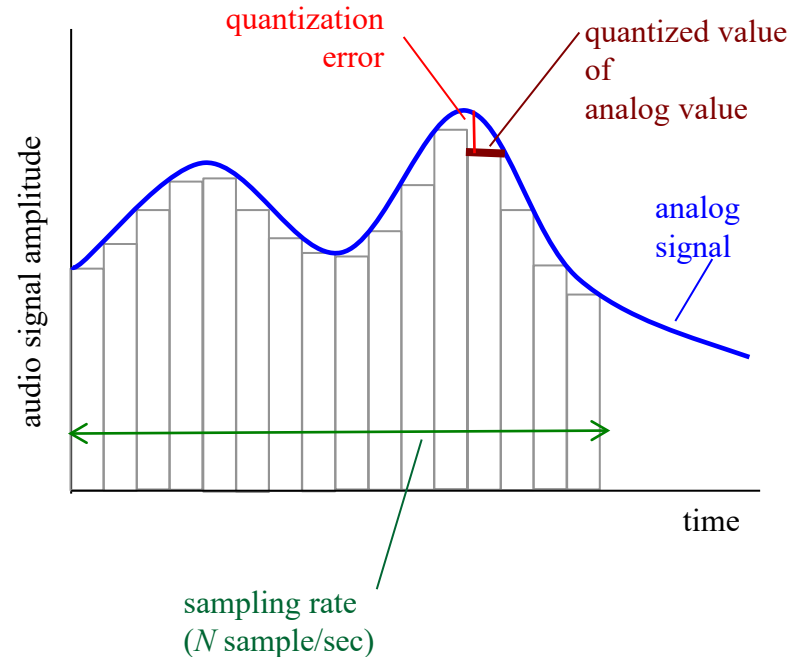   8,000 samples/sec
  - CD music:
    44,100 samples/sec
- **Quantization**: Each sample
  - 8 bits: $2^8=256$ values
  - 16 bit: 2^{16} values
- 8 k samples/s each 8 bit
  $\Rightarrow$ 64 kbps
- **Compression**: Compress to 5-16 kbps, e.g., using differences
- Lower bits/sample, samples/sec, or higher compression $\Rightarrow$ Lower Quality

Student Questions

quantization error

quantized value of analog value

audio signal amplitude

analog signal

sampling rate
(N sample/sec)

time

Multimedia Networking Applications

❑ **Streaming Stored Multimedia**

  ➢ Stored Media: Fast rewind, pause, fast forward

  ➢ Streaming: simultaneous play out and download

  ➢ Continuous play out: Delay jitter smoothed by playout buffer

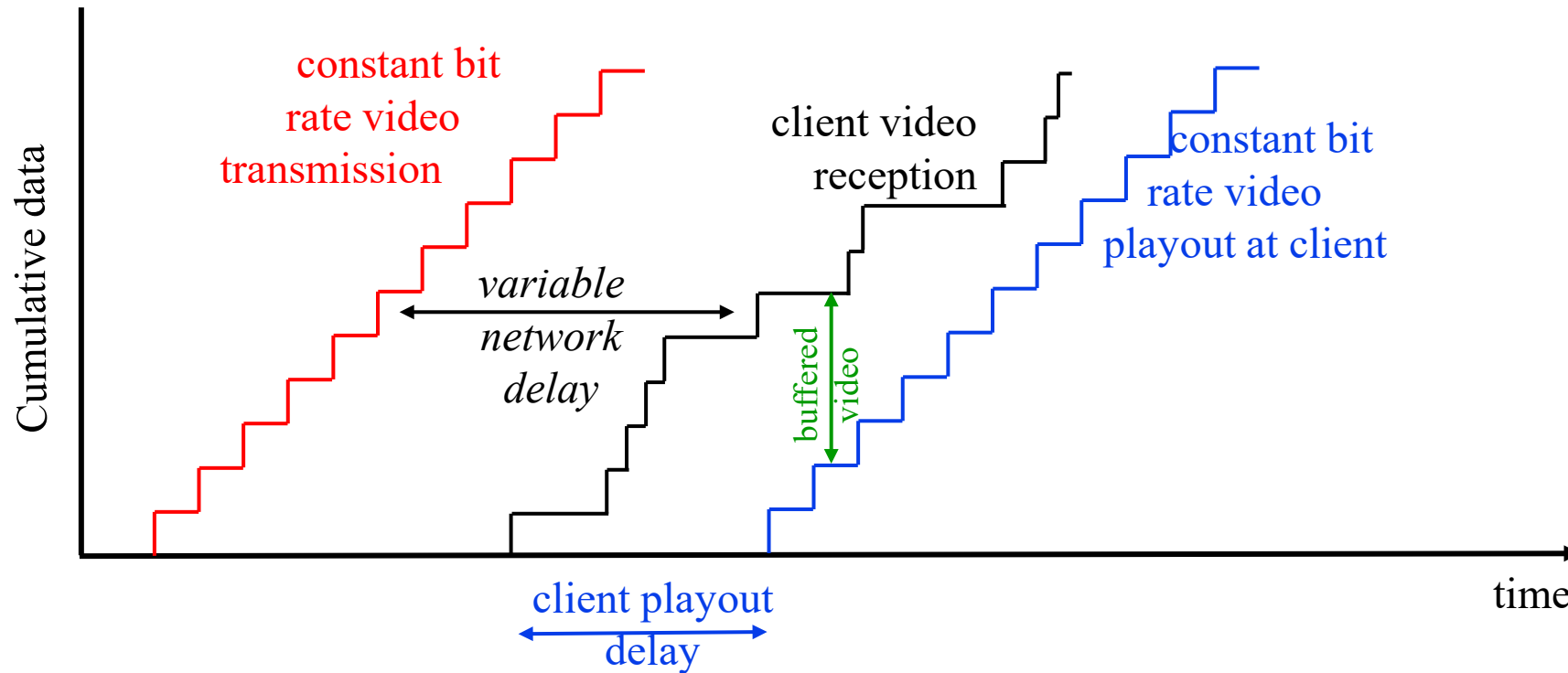❑ **Streaming Live Multimedia**: IPTV and Internet Radio

  ➢ No fast-forward

  ➢ High data rate to large number of users
    $\Rightarrow$ multicast or P2P,

  ➢ Delay jitter controlled by caching,

❑ **Real-Time Interactive Multimedia**: Internet Telephone, Video Conferencing
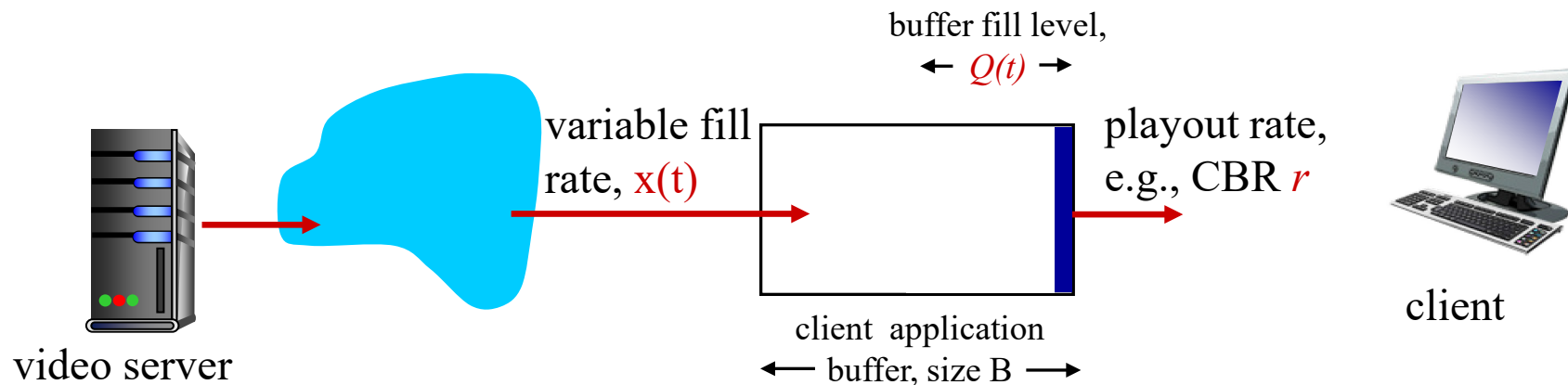
  ➢ Delay<400 ms.

Student Questions

http://www.cse.wustl.edu/~jain/cse473-21/ ©2021 Raj Jain

Playout Buffers

- ❑ Playout delay compensates for network delay, delay jitter
- ❑ Delay > Playout Delay ⇒ Packet late ⇒ Same as a lost packet

Client-side Buffering

1. Initial fill of buffer until playout begins at t_p
2. Fill rate $x(t)$ varies and playout rate r is constant
3. *$x < r$*: Buffer eventually empties causing freezing of video
4. *$x > r$*: buffer will not empty, Flow control to avoid overflow
5. *Tradeoff: Large initial playout delay $\Rightarrow$ Buffer starvation less likely but Larger delay until user begins watching*

Student Questions

buffer fill level,
$\leftarrow Q(t) \rightarrow$

variable fill rate, x(t)

playout rate, e.g., CBR *r*

client

video server

client application
$\leftarrow$ buffer, size B $\rightarrow$

http://www.cse.wustl.edu/~jain/cse473-21/
©2021 Raj Jain

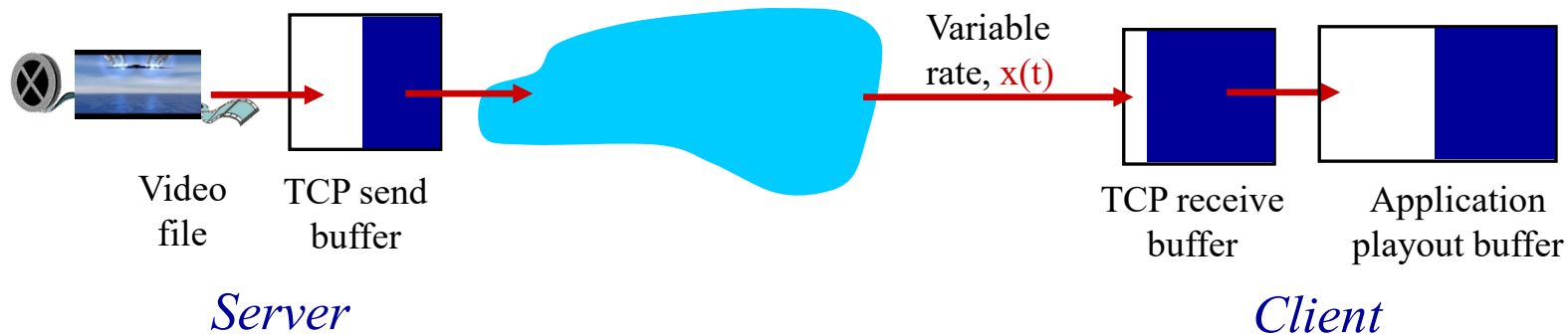# Streaming Using UDP

❑ Server sends at rate appropriate for client

  ➢ Often: Send rate = Encoding rate = Constant

  ➢ Transmission rate can be oblivious to congestion levels

❑ Short playout delay (2-5 seconds) to remove network jitter

❑ Application level error recovery

❑ UDP may *not* go through firewalls

Student Questions

Streaming Using HTTP

- Multimedia file retrieved via HTTP GET
- Send at maximum possible rate under TCP

Video file → TCP send buffer → ☁ → Variable rate, $x(t)$ → TCP receive buffer → Application playout buffer

Server   *Client*

- Fill rate fluctuates due to TCP congestion control, retransmissions (in-order delivery)
- Larger playout delay to smooth TCP delivery rate
- HTTP/TCP passes more easily through firewalls

Student Questions

- Are there ways to get around having to use TCP/HTTP to avoid firewalls?

Yes. That's why most video service providers have their own server-client apps that use specialized protocols and ports. E.g., You can't watch Netflix without Netflix app.

http://www.cse.wustl.edu/~jain/cse473-21/
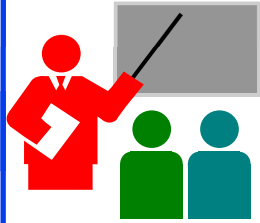
Multimedia Applications

1. Audio is sampled, digitized, and compressed

2. Initial playout delay helps overcome the jitter in delay

3. UDP results in lower jitter but may not go through firewall

4. HTTP uses TCP and so the delay variation can be large

Student Questions

❑ How does one choose whether to use UDP or HTTP/TCP? Is this a developer decision?

Server and client software developer decision

Ref: Section 9.1 and 9.2, Review Questions R1-R8, Problems P1-P5

http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

Overview
Voice Over IP

1. VoIP Packet Losses
2. VoIP with Fixed Playout Delay
3. VoIP with Adaptive Playout Delay
4. Recovering From Packet Loss
5. Skype

Student Questions

Voice-over-IP (VoIP)

❑ **End-end-delay Requirement**: needed to maintain "conversational" aspect

  ➢ Higher delays noticeable, impair interactivity

  ➢ < 150 ms: good

  ➢ > 400 ms: bad

  ➢ Includes application-level (packetization, playout), network delays

❑ Alternating talk spurts, silent periods.

  ➢ 64 kbps during talk spurt

  ➢ Packets generated only during talk spurts

  ➢ 20 ms chunks at 8 Kbytes/sec: 160 bytes of data

❑ Application sends a segment every 20 ms during talk spurt

Student Questions

http://www.cse.wustl.edu/~jain/cse473-21/   ©2021 Raj Jain

VoIP Packet Losses

- **Network Loss**: IP datagram lost due to network congestion (router buffer overflow)

- **Delay Loss:** IP datagram arrives too late for playout

  - typical maximum tolerable delay: 400 ms

- **Loss Tolerance:** *P*acket loss rates up to 10% can be concealed

Student Questions

VoIP with Fixed Playout Delay

❑ Example: Packets sent every 20 ms during talk spurt.

❑ First packet received at time r

❑ If playout begins at p, 4th packet will arrive too late

❑ If playout begins at p', all packets can be played on time

Packets

Sent   **Received**   **Played**

loss

time

r
p   p'
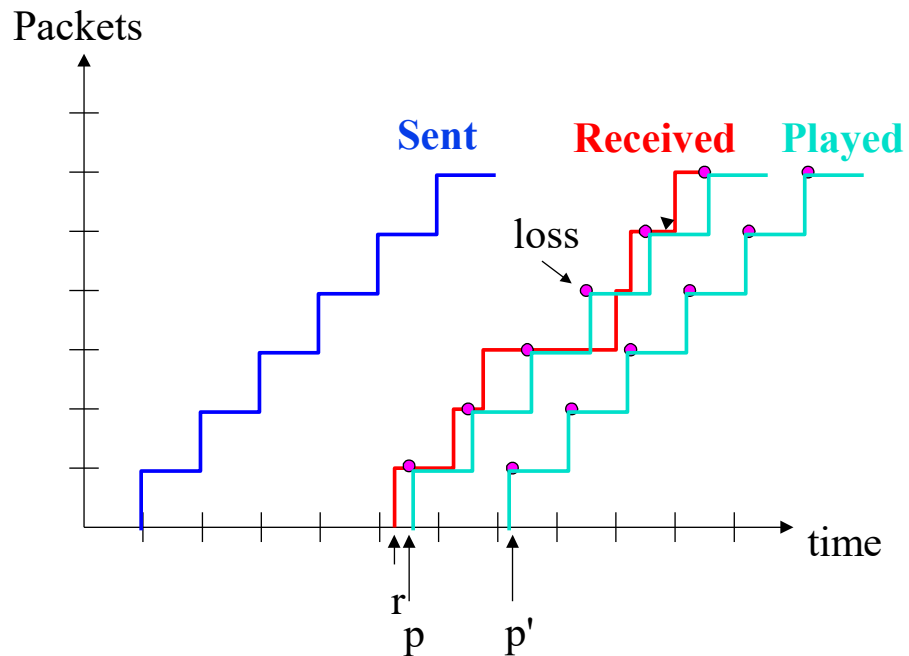
Homework 9: Playout Delay

Student Questions
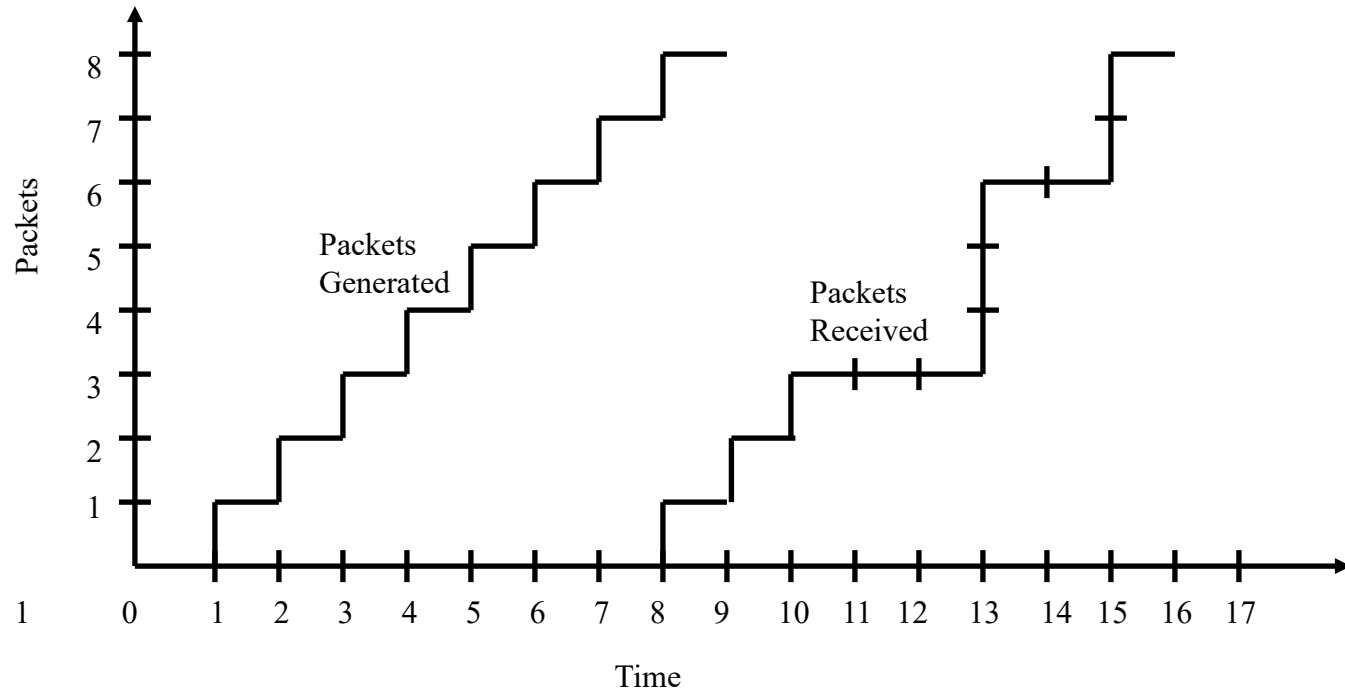
❑ [4 points] Consider the packet generation and reception sequence shown above. The first packet is generated at t=1 and is received at t=8.

❑ A. If Playout delay is zero and playout begins at t=8, which of the packets will not arrive in time? Show in a table.

❑ B. What is the minimum playout delay at the receiver that result in all of the first eight packets arriving in time for their playout?

Adaptive Playout Delay

❑ Estimate network delay, adjust playout delay at beginning of each talk spurt

❑ Silent periods compressed and elongated

❑ Chunks still played out every 20 ms during talk spurt

❑ Adaptively estimate packet delay: Similar to TCP RTT estimate

Student Questions

Adaptive Playout Delay

- t_i =Sending time
- r_i = Receiving time
- Measured delay sample = r_i-t_i
- d_i = Average network delay

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

- v_i = Variation of the delay

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

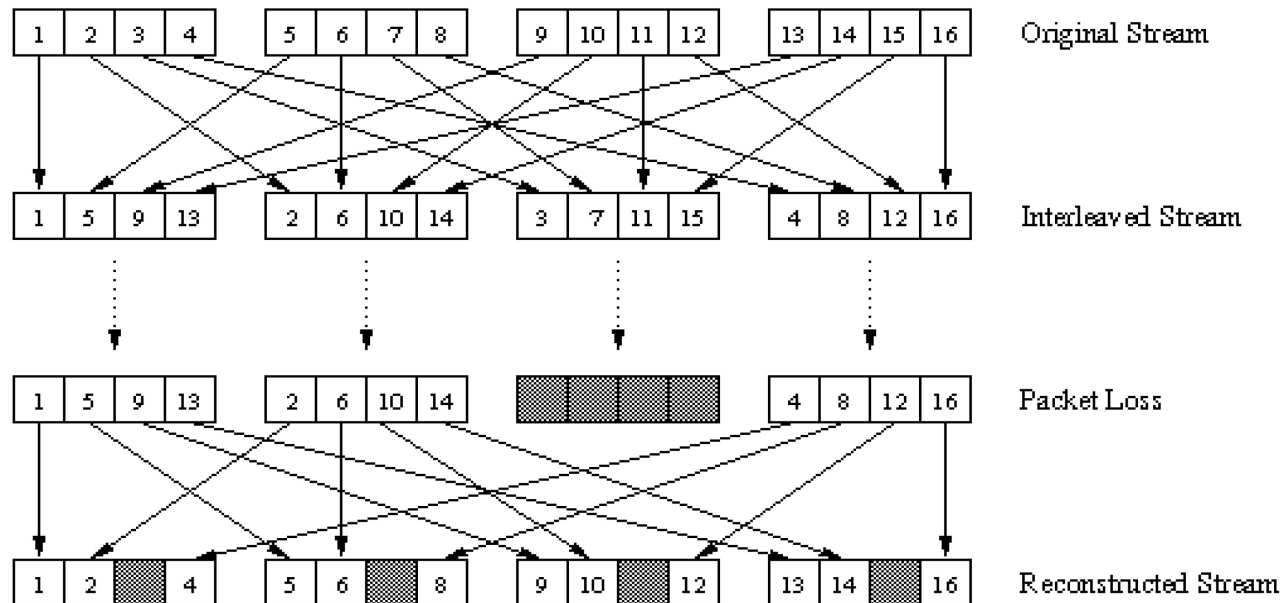- p_i = Playout time

$$p_i = t_i + d_i + Kv_i$$

- Here K is a constant, say 4.
- Sequence numbers and timestamps used to determine talk spurts and silence

Student Questions

Recovering From Packet Loss

- **Forward Error Correction**
- Send n+1 packets in place of n packets
- Send a lower-resolution stream in addition
- Play out the old syllable
- Busty Loss $\Rightarrow$ **Interleave** audio/video frames

| 1 | 2 | 3 | 4 | | 5 | 6 | 7 | 8 | | 9 | 10 | 11 | 12 | | 13 | 14 | 15 | 16 | Original Stream |

| 1 | 5 | 9 | 13 | | 2 | 6 | 10 | 14 | | 3 | 7 | 11 | 15 | | 4 | 8 | 12 | 16 | Interleaved Stream |

| 1 | 5 | 9 | 13 | | 2 | 6 | 10 | 14 | | ▓ | ▓ | ▓ | ▓ | | 4 | 8 | 12 | 16 | Packet Loss |

| 1 | 2 | ▓ | 4 | | 5 | 6 | ▓ | 8 | | 9 | 10 | ▓ | 12 | | 13 | 14 | ▓ | 16 | Reconstructed Stream |

http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

Student Questions

- For HTTP/3 I found that QUIC transport layer protocol is used in lieu of TCP - at high level QUIC seems to leverage multiplexed connections between two endpoints using UDP.

Yes. QUIC is being standardized in IETF. Proposed by Google, it is already widely used.

- Is the following idea correct? Because TCP is implemented explicitly in some OS kernels and middlebox firmwares, it's difficult to make significant changes to it, whereas because QUIC ~~builds from UDP~~ it can be iterated upon.

Because it runs in user-space.

Voice-over-IP: Skype

- **Proprietary** application-layer protocol (inferred via reverse engineering)
- **Encrypted** messages
- **P2P**: Media does not go through a central server
- **Clients**: Skype peers connect directly to each other for VoIP call
- **Super Nodes** (SN): Skype peers with special functions
- **Overlay Network**: Among SNs to locate clients
- **Login server**

Skype login server

Skype clients (SC)

Super node (SN)

Super node overlay network
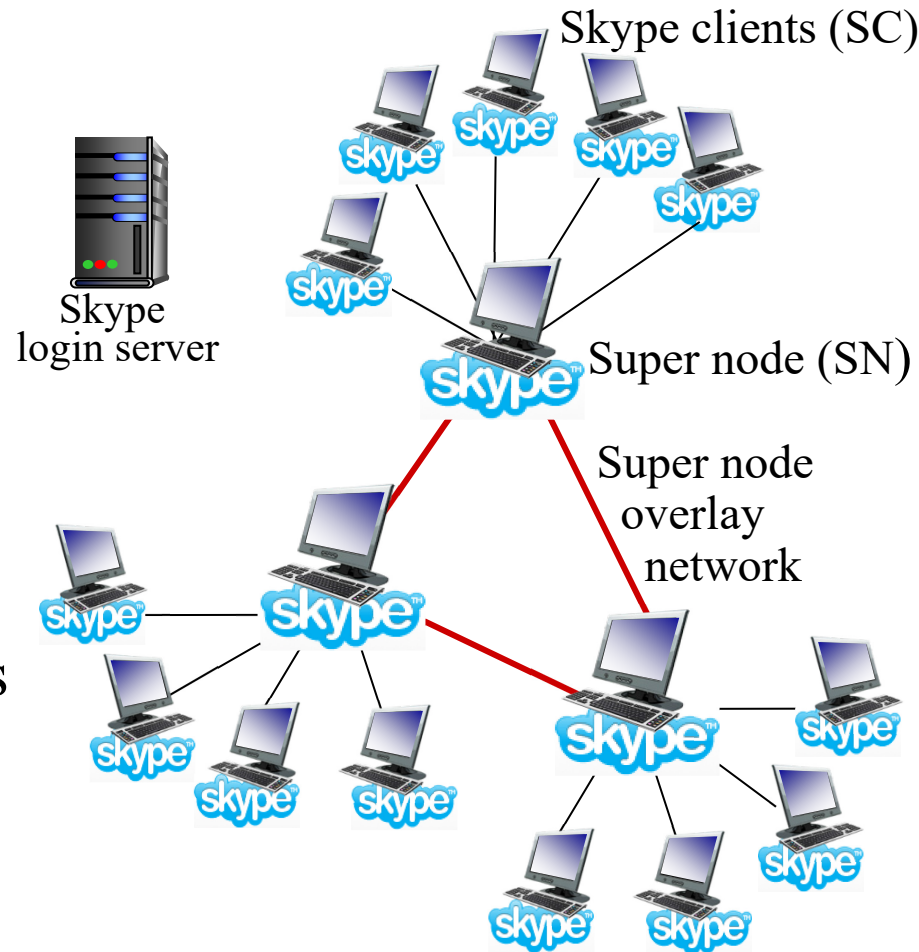
Student Questions

- Are applications like Microsoft Teams and Zoom repackaged Skype applications? How are they different, if so?

Sorry. I have no idea about Teams and Zoom.

http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

P2P voice-over-IP: Skype

Skype client operation:
1. Joins Skype network by contacting SN (IP address cached) using TCP
2. Logs-in: Username, password to centralized Skype login server

3. Obtains IP address for callee from SN, SN overlay
4. Initiate call directly to callee via SN
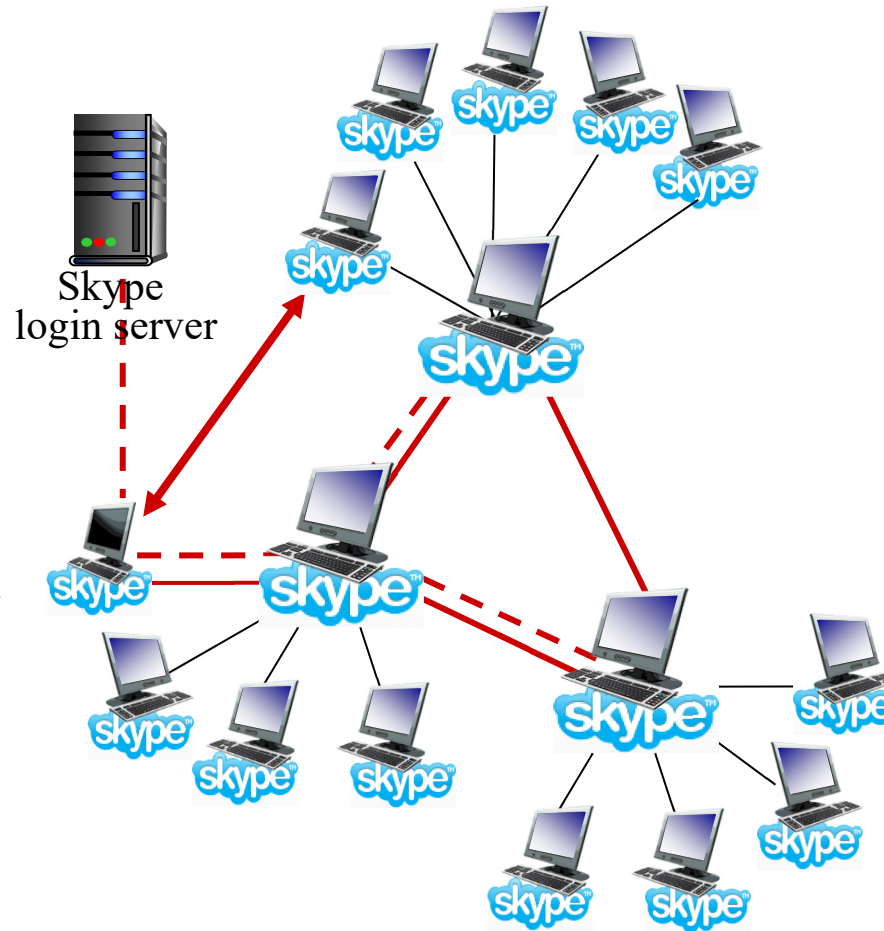
Skype
login server

Skype: Super Nodes as Relays

- **Problem**: both Alice, Bob are behind "NATs"
  - NAT prevents outside peer from initiating connection to inside peer
  - Inside peer *can* initiate connection to outside
- **Relay solution**: Alice, Bob maintain open connection to their SNs
  - Alice signals her SN to connect to Bob
  - Alice's SN connects to Bob's SN
  - Bob's SN connects to Bob over open connection Bob initiated

http://www.cse.wustl.edu/~jain/cse473-21/ ©2021 Raj Jain
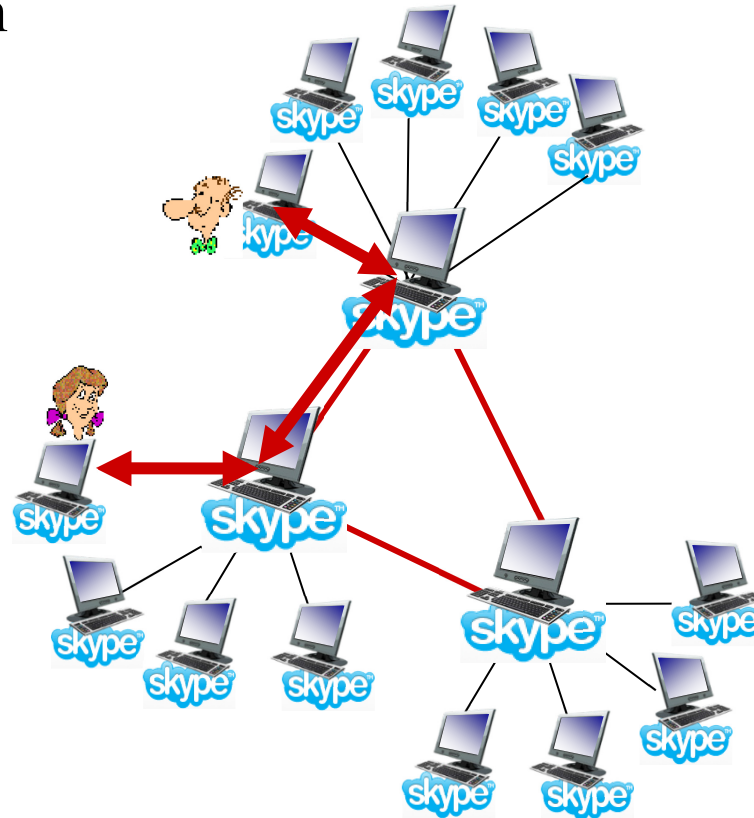
Voice over IP

Student Questions
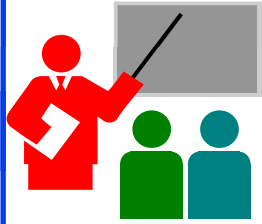
1. Talk spurts are delayed to overcome jitter

2. Playout delay is estimated adaptively using mean and standard deviation

3. Forward error correction and interleaving is used to overcome losses and burst errors

4. Skype uses super nodes to help connect peers. A login server is used for authentication.

5. Skype nodes maintain an outgoing connection with the super nodes. These connections are used for incoming VoIP packets.

Protocols for Real-Time Applications

1. Real-Time Transport Protocol (RTP)
2. Session Initiation Protocol (SIP)
3. H.323 Protocols

Student Questions

Real-Time Transport Protocol (RTP)

❑ Common sublayer between applications and UDP

❑ Provides sequence numbers, timestamps, and other facilities

❑ Supports both unicast and multicast

transport layer {

| Application |
| RTP |
| UDP |
| IP |
| Data Link |
| Physical |
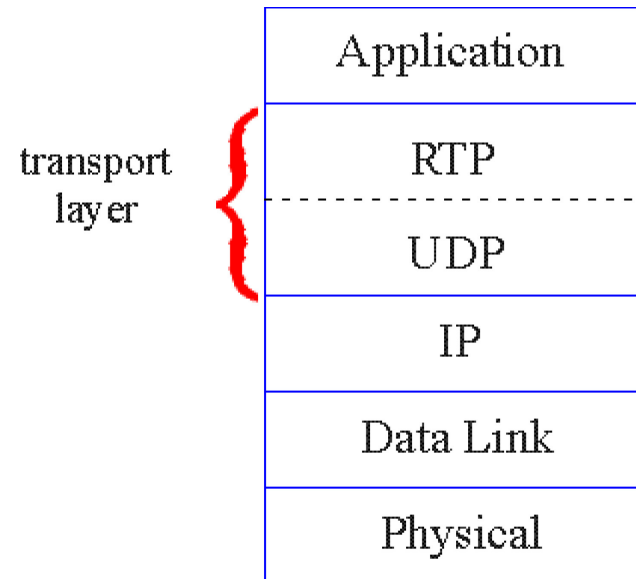
RTP Packet Format

Marker	Payload Type	Sequence Number	Time-stamp	Synchronization Source Identifier	Miscellaneous Fields
1b	7b	16b	32b	32b	

❑ Marker indicates that the packet contains special data required by some applications

❑ SSRC = Synchronization Source = Stream #

Payload Type	Coding	Rate
0	PCM mu-law	64 kbps
3	GSM	13 kbps
7	LPC	2.4 kbps
26	Motion JPEG	
31	H.261	
33	MPEG2 video	

Student Questions

http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

Session Initiation Protocol (SIP)

❑ Application level signaling protocol for voice and video conferencing over Internet

❑ Allows creating, modifying, terminating sessions with one or more participants

❑ Carries session descriptions (media types) for user capabilities negotiation

❑ Supports user location, call setup, call transfers

❑ Supports mobility by proxying and redirection

Student Questions

SIP (Cont)

❑ SIP Uniform Resource Identifiers (URIs):
Similar to email URLs
sip:jain@wustl.edu
sip:+1-614-292-3989:123@wustl.edu?subject=lecture

❑ SIP can use UDP or TCP

❑ SIP messages are sent to SIP servers:
  ➢ Registrar: Clients register and tell their location to it
  ➢ Location: Given name, returns possible addresses for a user. Like Directory service or DNS.
  ➢ Redirect: Returns current address to requesters
  ➢ Proxy: Intermediary. Acts like a server to internal client and like a client to external server

Student Questions

❑ How does SIP decide on TCP or UDP?

SIP Server and clients decide.

Locating using SIP

❑ Allows locating a callee at different locations

❑ Callee registers different locations with Registrar

❑ SIP Messages: Ack, Bye, Invite, Register, Redirection, ...
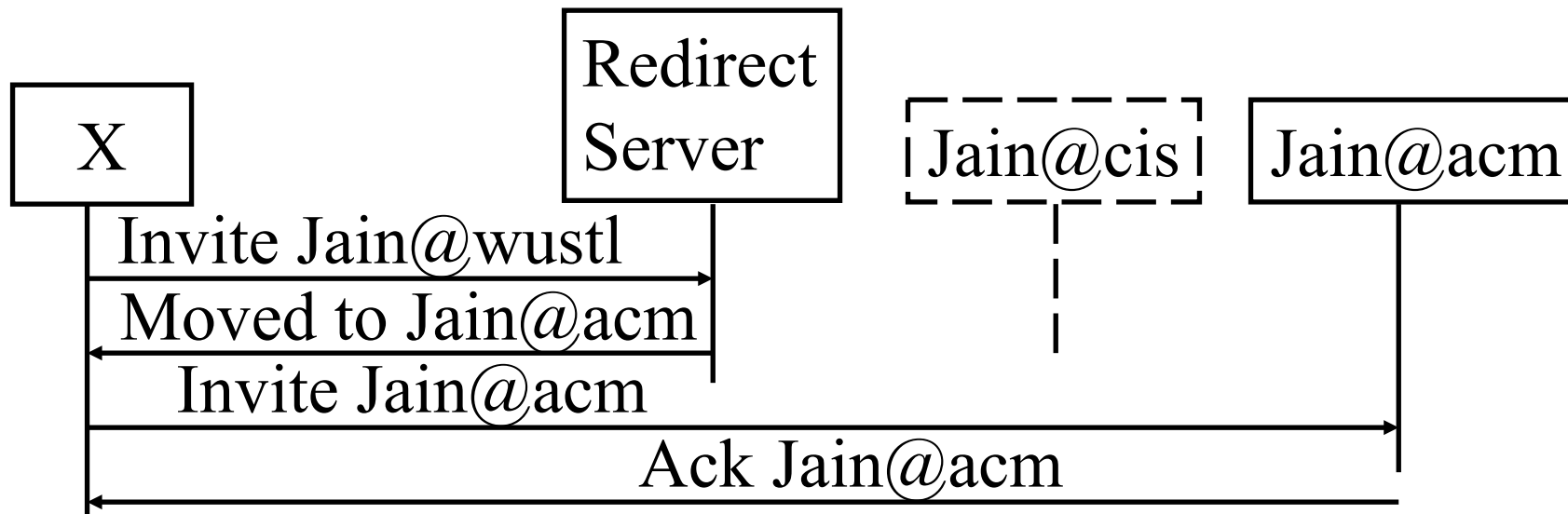
Student Questions

SIP Proxy

SIP Invite

SIP Invite

SIP Trying

SIP 180 Ringing

SIP 180 Ringing

SIP 200 OK

SIP 200 OK

SIP Ack

SIP Ack

Conversation using RTP

SIP Bye

SIP Bye

SIP OK

SIP OK
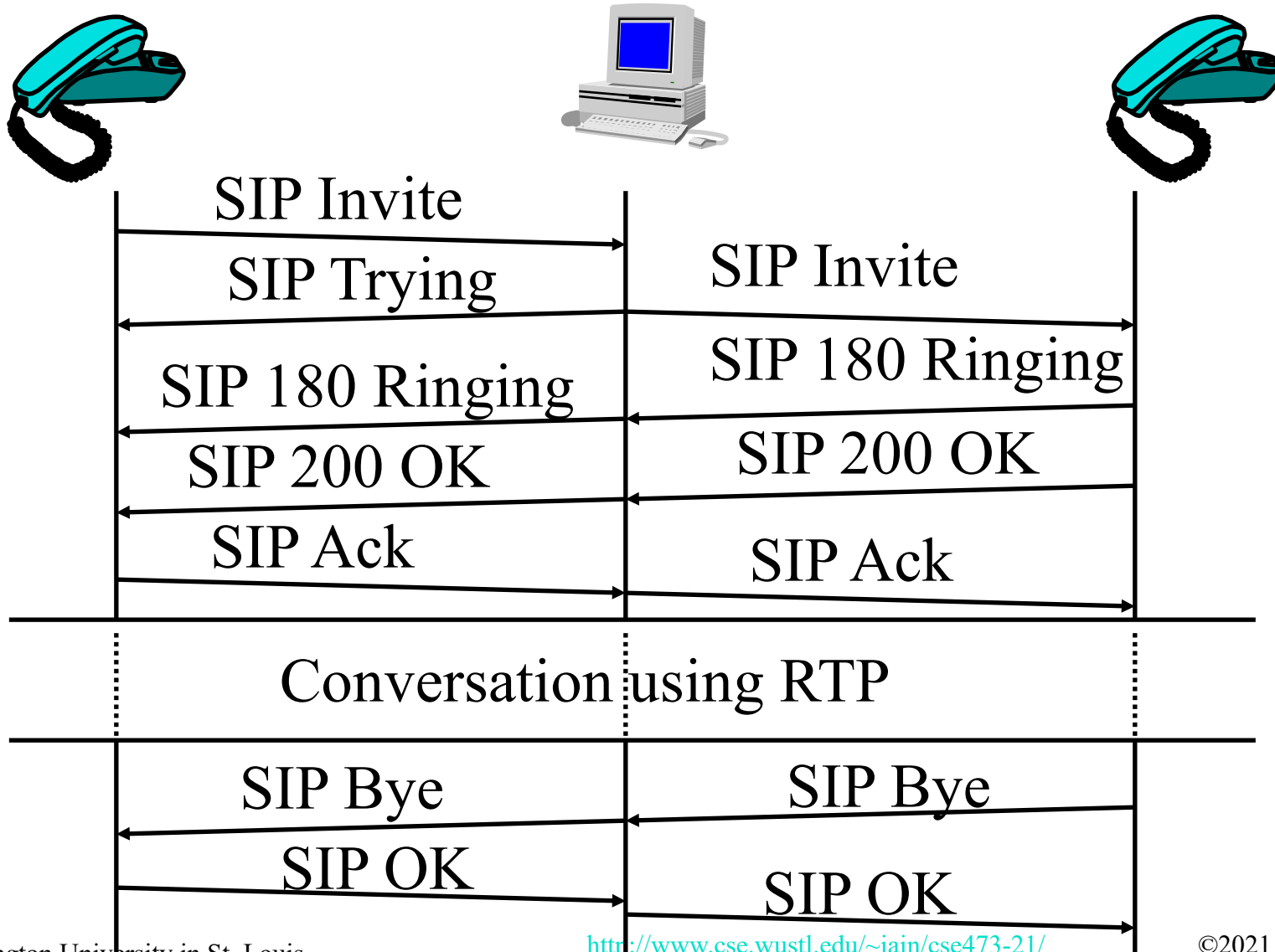
Student Questions

H.323 Protocols

❑ Multimedia over LANs, V1 (June 96), V2(Feb 98)

❑ Provides component descriptions, signaling procedures, call control, system control, audio/video codecs, data protocols

Video	Audio		Control and Management			Data
H.261 H.263	G.711, G.722, G.723.1, G.728, G.729	RTCP	H.225.0 RAS	H.225.0 Signaling	H.245 Control	T.124
RTP			X.224 Class 0			T.125
UDP			TCP			T.123
Network (IP)						
Datalink (IEEE 802.3)						

Protocols for Real-Time Applications

1. RTP is used to transmit multimedia over UDP

2. SIP is a signaling (control) protocol to establish multimedia connections

3. H.323 is a framework for a group of protocols used for multimedia

Student Questions

❑ Can you please explain what bandwidth provisioning and network dimensioning are? Are they used in conjunction?

Provisioning: Allocation

Dimensioning: Capacity planning

Networking Support for Multimedia

1. QoS Components
2. Traffic Shaping
3. Token Bucket Shaper
4. Traffic Policing
5. Differentiated Services

Student Questions

http://www.cse.wustl.edu/~jain/cse473-21/ ©2021 Raj Jain
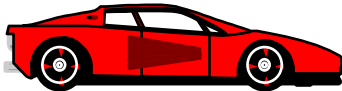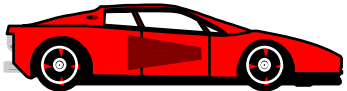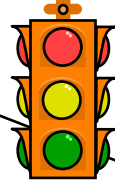
QoS Components

① Signaling: What I need

② Admission control

③ Shaping

④ Policing

⑤ Classification

Scheduling ⑥
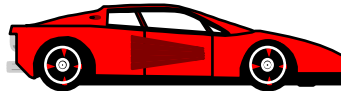
⑧ Routing
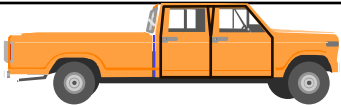
⑦ Buffer Mgmt Drop Policies
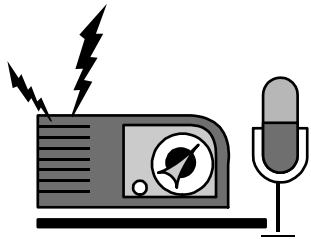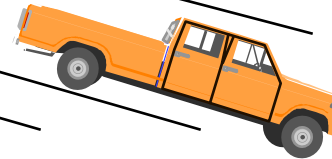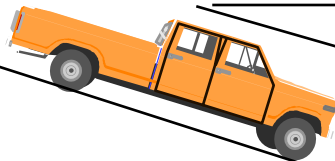
⑨ Traffic Monitoring and feedback

Student Questions

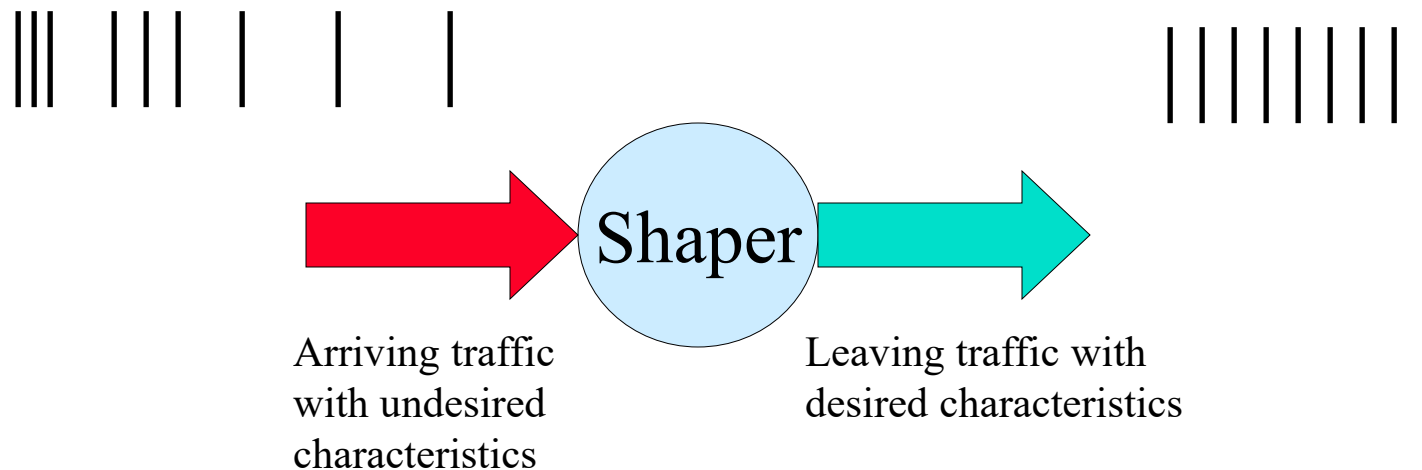http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

QoS Components (Cont)

1. **Signaling**: Users need to tell/negotiate their QoS requirements with the network
2. **Admission Control**: Network can deny requests that it can not meet
3. **Shaping**: Traffic is smoothed out so that it is easier to handle
4. **Policing**: Ensuring that the users are sending at the rate they agreed to
5. **Marking/Classification**: Packets are classified based on the source, destination, TCP ports (application)
6. **Scheduling**: Different flows get appropriate treatment
7. **Drop Policies**: Low priority packets are dropped.
8. **Routing**: Packets are sent over paths that can meet the QoS
9. **Traffic Management**: Sources may be asked to reduce their rates to meet the loss rate and delay guarantees

Student Questions

http://www.cse.wustl.edu/~jain/cse473-21/
©2021 Raj Jain

Traffic Shaping

❑ Altering the traffic characteristics of a given flow is called traffic shaping

❑ The source must shape its traffic prior to sending it to network so it does not violate traffic contract

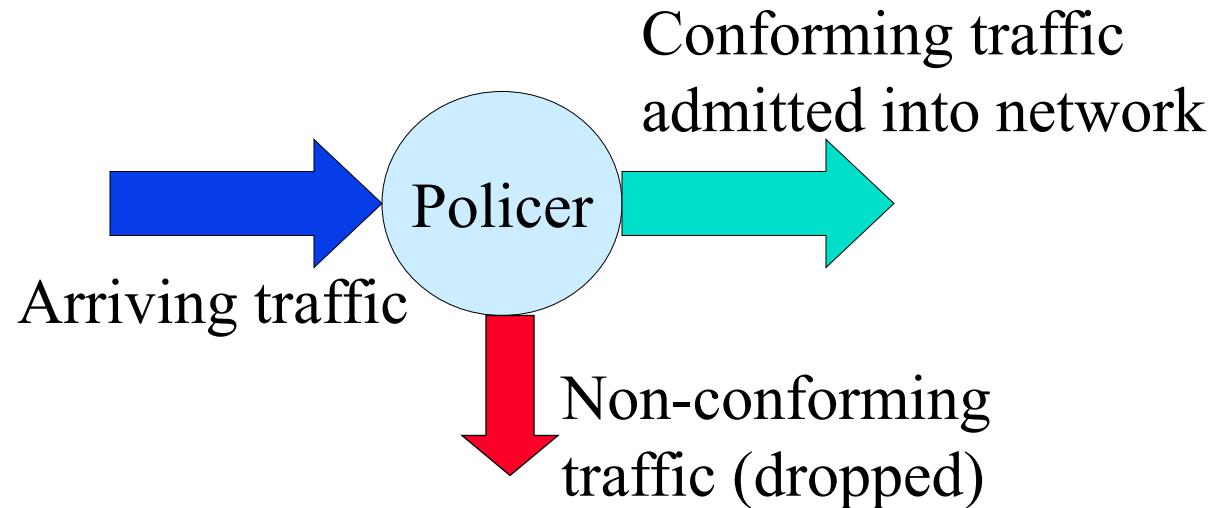Arriving traffic with undesired characteristics

Leaving traffic with desired characteristics

Token Bucket Shaper

Tokens arrive periodically at Average Rate

Token → Bucket Size K

Incoming Traffic → ▮ ▮ Server → Shaped Traffic
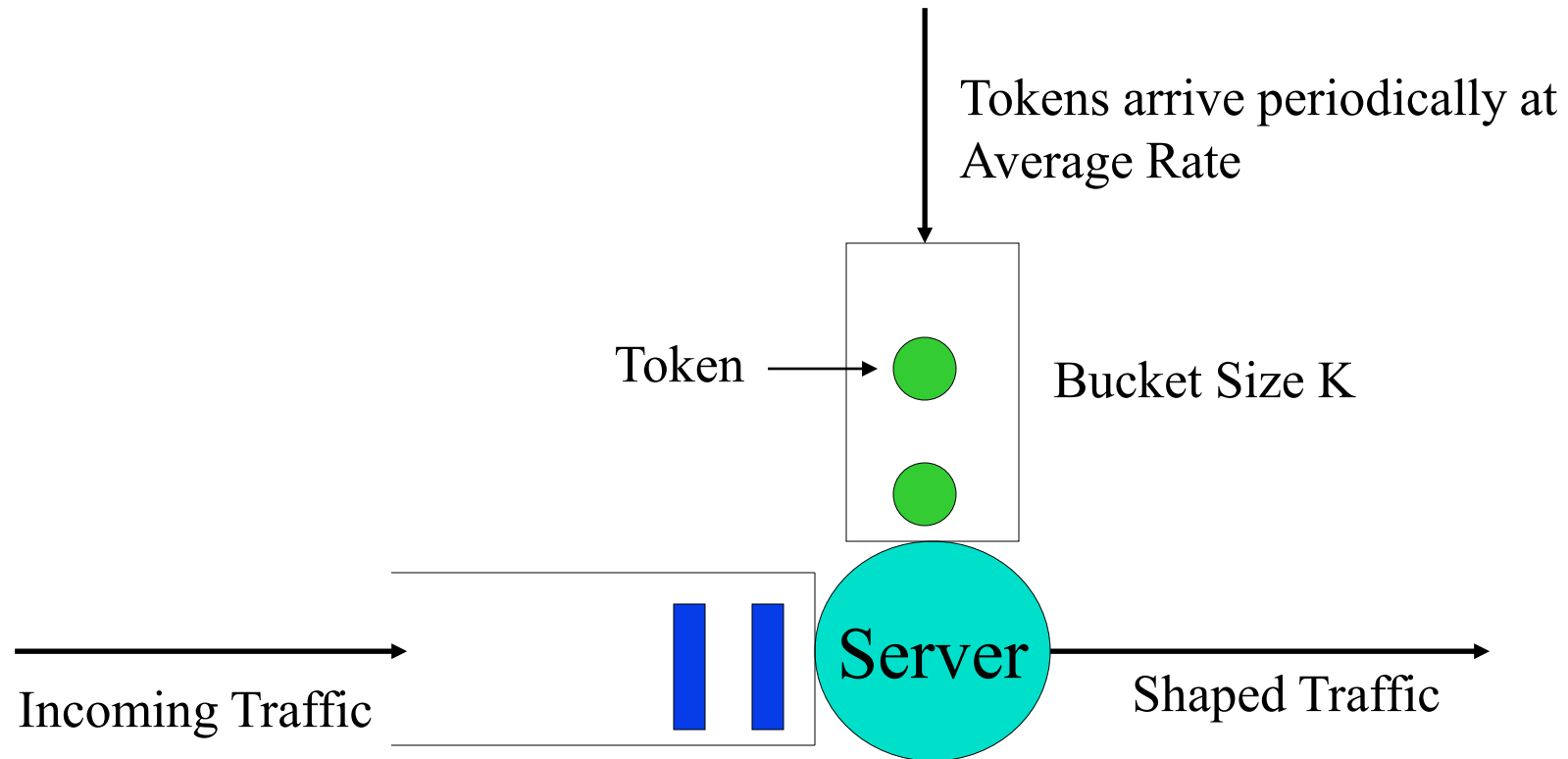
Traffic Policing

❑ Users violating the traffic contract can jeopardise the QoS of other connections

❑ The network must protect well behaving users against such traffic violations

❑ Policing functions are deployed at the edge (entry) of the network

Conforming traffic admitted into network

Policer

Arriving traffic

Non-conforming traffic (dropped)

Student Questions

❑ Under what circumstances would a user violate the traffic contract?

When the bit rate produced is higher than the contracted rate, e.g., too much action can result in higher bit rates in video

http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

# Peak Rate Policing with Leaky Bucket

- Enforces sustained rate and maximum burst size
- Requires only one counter
  - counter is decremented, to a minimum of zero, at the avg rate
  - counter is incremented by one, to a maximum of a limiting value, for each packet arrival
- An arriving packet is non-conforming if counter is at its limit

Incoming Packets

Rejected

Accepted
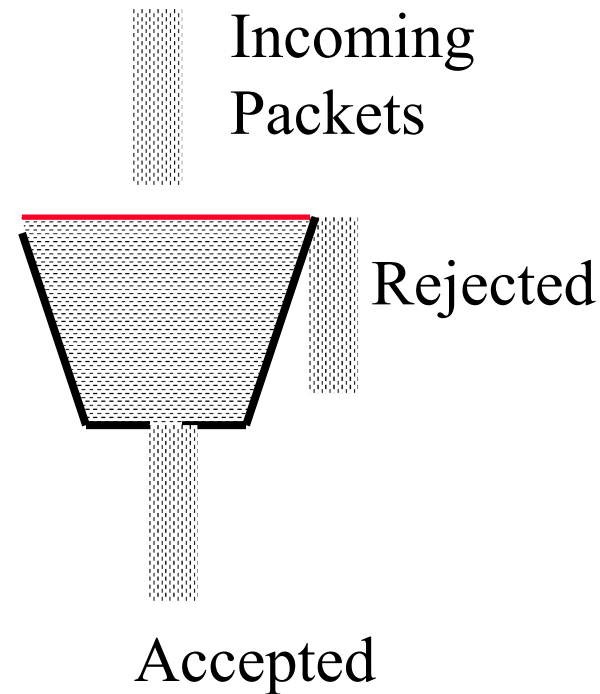
Student Questions

Fair Queueing
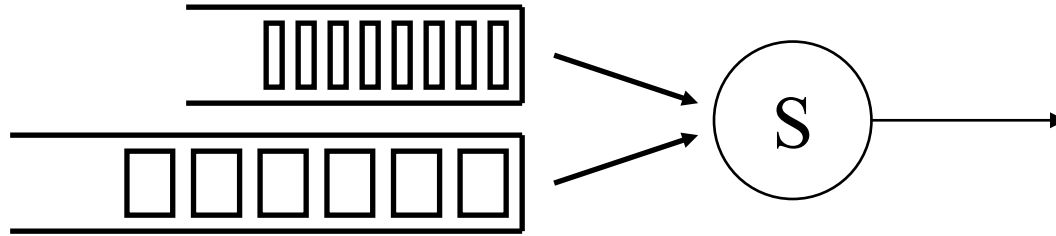
❑ Bit-level round robin but packet level scheduling

❑ Count the packet size and determine which packet would finish first. Serve that packet.

❑ Each flow gets the same number of bits/sec

http://www.cse.wustl.edu/~jain/cse473-21/

Weighted Fair Queueing (WFQ)
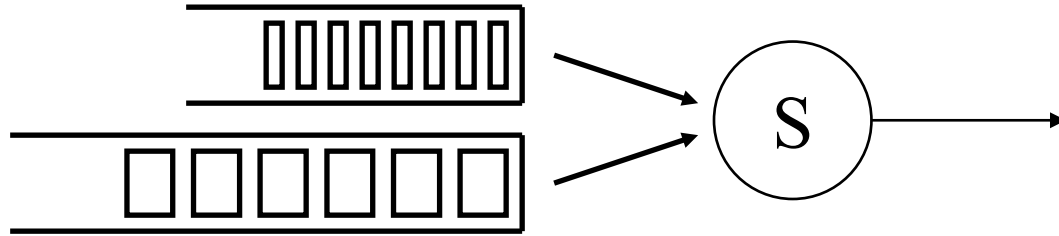
- Fair queueing with different weight for each queue
- Flow 1 gets x bit/sec
- Flow 2 gets y bit/sec
- Flow n gets z bit/sec
- Here, x, y, z are weights

Student Questions

http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

Maximum Delay with WFQ and Policing

- Max Delay $d_{max} = b_i/(R\, w_i/\Sigma\, w_j)$
- Here,
- b_i=Burst size of ith flow
- R=Service Rate
- W_i=Weight of ith flow

arriving traffic

token rate, r

bucket size, b

per-flow rate, R

WFQ

$D_{max} = b/R$

r_1

b_1

w_1

r_n

b_n

w_n

http://www.cse.wustl.edu/~jain/cse473-21/

Differentiated Services

Ver	Hdr Len	Precedence	ToS	Unused	Tot Len
4b	4b	3b	4b	1b	16b

DSCP	ECN
6b	2b
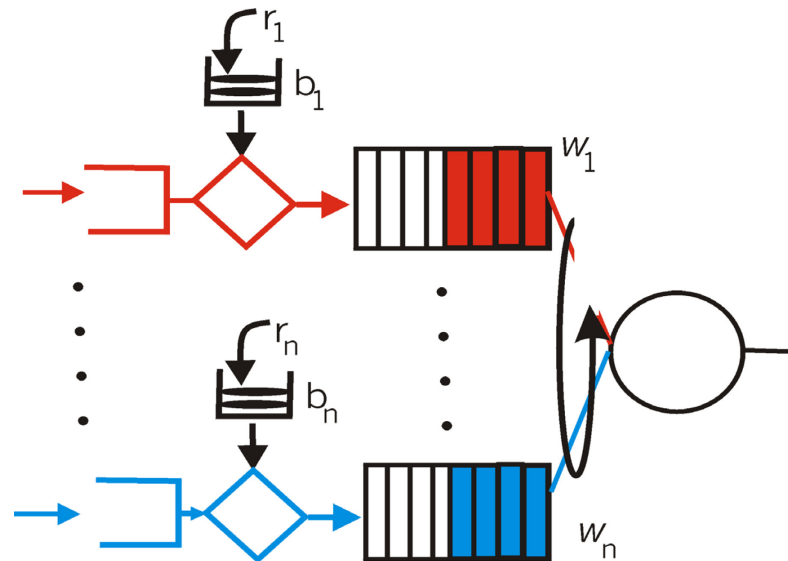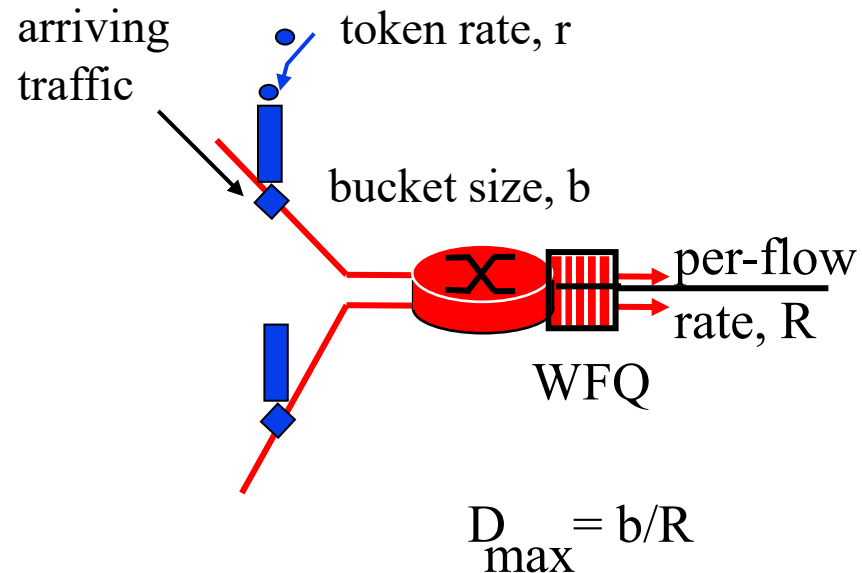
- IPv4: 3-bit precedence + 4-bit ToS
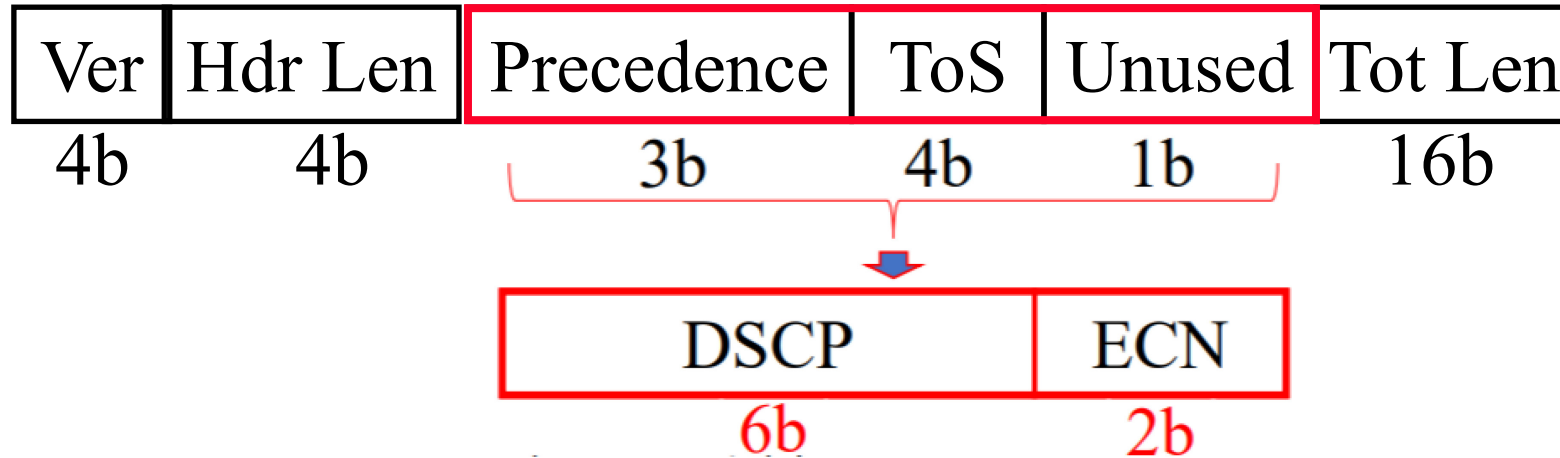- OSPF and integrated IS-IS can compute paths for each ToS
- Many vendors use IP precedence bits but the service varies $\Rightarrow$ Need a standard $\Rightarrow$ Differentiated Services
- Edge routers can mark the packets $\Rightarrow$ Set ToS field
- Core routers use ToS field to provide "Per-Hop-Behavior"

Student Questions

Per-hop Behaviors

❑ Externally Observable Forwarding Behavior

❑ x% of link bandwidth

❑ Minimum x% and fair share of excess bandwidth

❑ Priority relative to other PHBs

Student Questions

http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

Expedited Forwarding

❑ Also known as "Premium Service"

❑ Virtual leased line

❑ Guaranteed minimum service rate

❑ Policed: Arrival rate < Minimum Service Rate

❑ Not affected by other data PHBs
   $\Rightarrow$ Highest data priority (if priority queueing)

❑ Code point: 101 110

Student Questions

Assured Forwarding
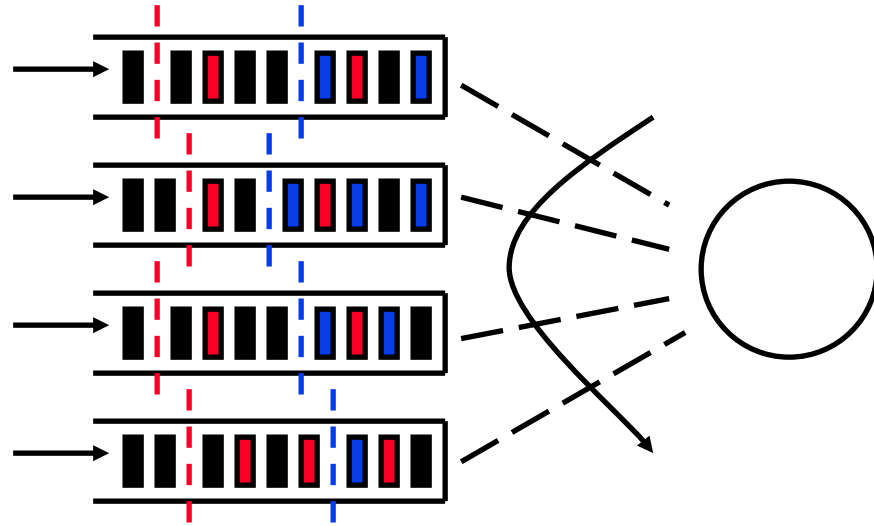
- PHB <u>Group</u>
- Four Classes: No particular ordering
- Three drop preference per class

Assured Forwarding (Cont)

- ❑ DS nodes SHOULD implement all 4 classes
  and MUST accept all 3 drop preferences. Can implement 2 drop
  preferences.
- ❑ Similar to nrt-VBR/ABR/GFR
- ❑ Code Points:

Drop Prec.	Class 1	Class 2	Class 3	Class 4
Low	010 000	011 000	100 000	101 000
Medium	010 010	011 010	100 010	101 010
High	010 100	011 100	100 100	101 100

- ❑ Avoids 11x000 (used for network control)

Student Questions

- ❑ Does each class always have the same QoS attributes or are they used differently in different situations?

Only relative drop and service priorities. No absolute values have been specified ⇒ No real guarantees.

Network Support for Multimedia

Approach	Granularity	Guarantee	Mechanisms	Complex	Deployed?
Making best of best effort service	All traffic treated equally	None or soft	No network support (all at application)	low	everywhere
Differentiated service	Traffic "class"	None of soft	Packet marker, scheduling, policing.	med	some
Per-connection QoS	Per-connection flow	Soft or hard after flow admitted	Packet marker, scheduling, policing, call admission	high	little to none
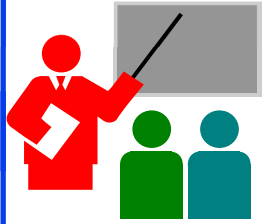
Network Support for Multimedia

1. QoS is obtained using several components including shaping, policing, differentiated services

2. Shaping is done by a token bucket

3. Policing is done using a leaky bucket

4. Differentiated services specifies per-hop behaviors

   1. Expedited Forwarding: min service rate

   2. Assured Forwarding: 4 classes, 3 drop precedence's

Ref: Section 9.5, Problems P17-P22
Washington University in St. Louis

http://www.cse.wustl.edu/~jain/cse473-21/

©2021 Raj Jain

Summary
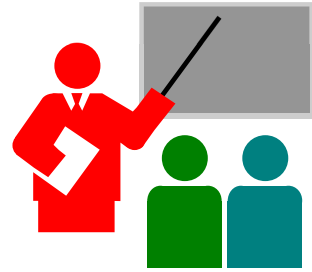
1. Multimedia applications require bounded delay, delay jitter, and minimum throughput
2. RTP allows sequencing and time stamping
3. SIP allows parameter negotiation and location
4. QoS requires shaping, policing, scheduling, etc.
5. Diffserv allows different packets to get different service

Student Questions

Acronyms

- ABR   Available Bit Rate
- CBR   Constant Bit Rate
- CD   Compact Disk
- DNS   Domain Name System
- DS   DiffServe
- GFR   Guaranteed Frame Rate
- HTTP   HyperText Transfer Protocol
- IEEE   Institution of Electrical and Electronics Engineers
- IP   Internet Protocol
- IPTV   Internet Protocol Television
- IPv4   Internet Protocol Version 4
- IS   Integrated Services
- LAN   Local Area Network
- NAT   Network Address Translator
- OSPF   Open Shortest Path First
- PHB   Per-Hop Behavior

Student Questions

Acronyms (Cont)

- QoS — Quality of Service
- RAS — Registration, Admission, and Status
- RTCP — Real-Time Transport Protocol Control Protocol
- RTP — Real-Time Transport Protocol
- RTSP — Real-Time Streaming Protocol
- RTT — Round Trip Time
- SC — Skype Clients
- SIP — Session Initiation Protocol
- SN — Super Node
- SSRC — Synchronization Source
- TCP — Transmission Control Protocol
- ToS — Type of Service
- UDP — User Datagram Protocol
- URI — Uniform Resource Identifiers
- URL — Uniform Resource Locator
- VBR — Variable Bit Rate

Student Questions

Acronyms (Cont)

❑ VoIP   Voice over IP

Student Questions

Washington University in St. Louis   http://www.cse.wustl.edu/~jain/cse473-21/   ©2021 Raj Jain

Scan This to Download These Slides

THANK YOU

Raj Jain

http://rajjain.com

Student Questions

http://www.cse.wustl.edu/~jain/cse473-21/i_9mmn.htm

http://www.cse.wustl.edu/~jain/cse473-21/

Related Modules

CSE 567: The Art of Computer Systems Performance Analysis
https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypJWOSPMh8Azcgy5e_10TiDw

CSE 570: Recent Advances in Networking (Spring 2013)

https://www.youtube.com/playlist?list=PLjGG94etKypLHyBN8mOgwJLHD2FFIMGq5

CSE571S: Network Security (Spring 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyyg93u

Video Podcasts of Prof. Raj Jain's Lectures,
https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw

Student Questions