Chapter 15
Network Services and Protocols for Multimedia Communications

15.1 Protocol Layers of Computer Communication Networks

15.2 Local Area Network and Access Networks

15.3 Internet Technologies and Protocols

15.4 Multicast Extension

15.5 Quality-of-Service and Quality-of-Experience

15.6 Protocols for Multimedia Transmission and Interaction

15.7 Case Study: Internet Telephony

15.8 Further Exploration

15.1 Protocol Layers of Computer Communication Networks

- Computer networks are essential to modern computing environment.
- Multimedia communications and networking share all major issues and technologies of computer communication networks.
- The ever-growing demands from numerous conventional and new generation multimedia applications have made networking one of the most active areas for research and development.
- Various network services and protocols are becoming a central part of most contemporary multimedia systems.

OSI Network Layers

- **OSI Reference Model has the following network layers:**

1. **Physical Layer**: Defines electrical and mechanical properties of the physical interface, and species the functions and procedural sequences performed by circuits of the physical interface.

2. **Data Link Layer**: Species the ways to establish, maintain and terminate a link, e.g., transmission and synchronization of data frames, error detection and correction, and access protocol to the Physical layer.

3. **Network Layer**: Defines the routing of data from one end to the other across the network, using circuit switching or packet switching. Provides services such as addressing, internetworking, error handling, congestion control, and sequencing of packets.

OSI Network Layers (Cont'd)

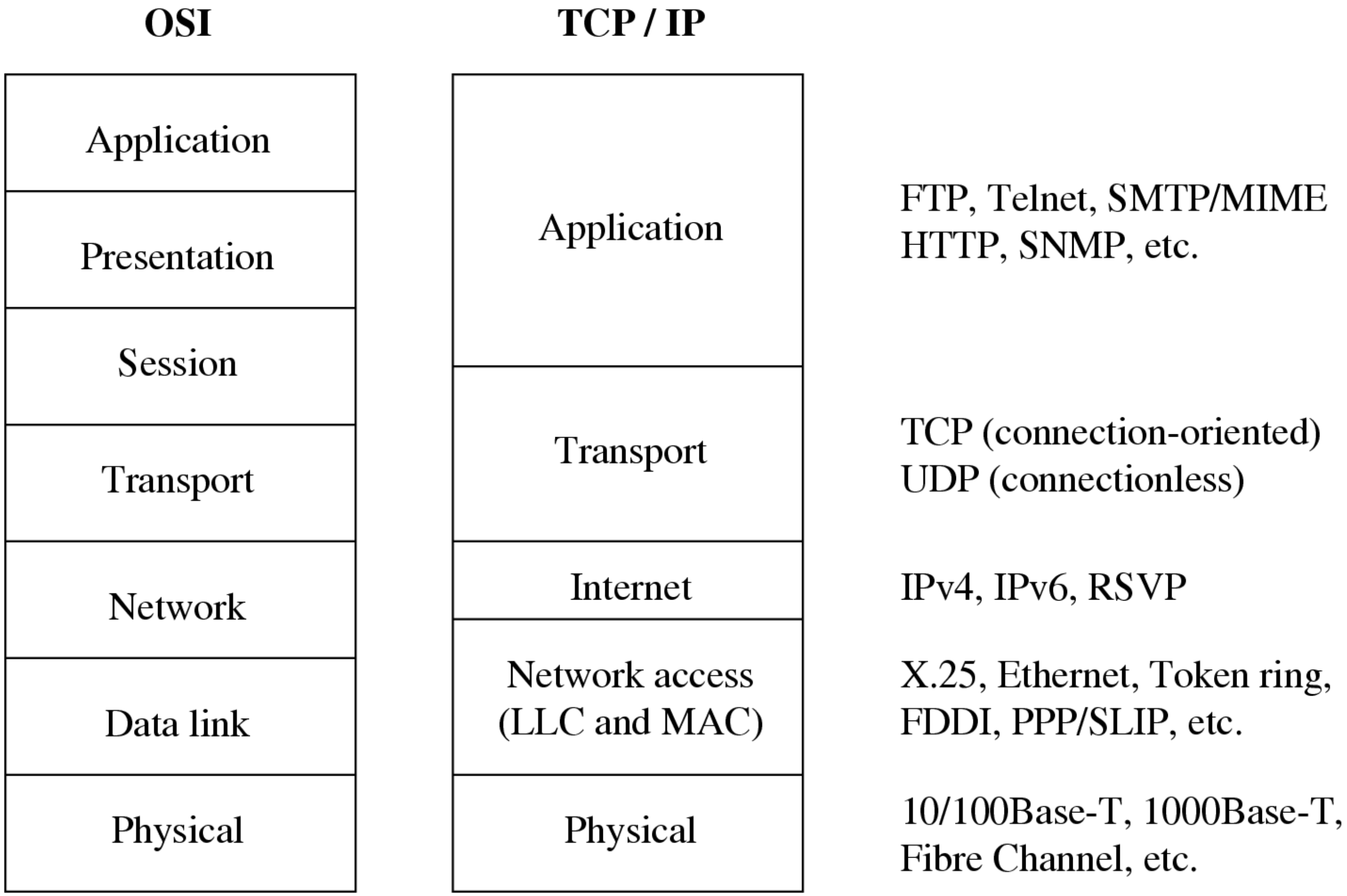
4. **Transport Layer**: Provides end-to-end communication between *end systems* that support end-user applications or services. Supports either *connection-oriented* or *connectionless* protocols. Provides error recovery and flow control.

5. **Session Layer**: Coordinates interaction between user applications on different hosts, manages sessions (connections), e.g., completion of long file transfers.

6. **Presentation Layer**: Deals with the syntax of transmitted data, e.g., conversion of different data formats and codes due to different conventions, compression or encryption.

7. **Application Layer**: Supports various application programs and protocols, e.g., FTP, Telnet, HTTP, SNMP, SMTP/MIME, etc.

TCP/IP Protocols



OSI

Application

Presentation

Session

Transport

Network

Data link

Physical

TCP / IP

Application — FTP, Telnet, SMTP/MIME HTTP, SNMP, etc.

Transport — TCP (connection-oriented) UDP (connectionless)

Internet — IPv4, IPv6, RSVP

Network access (LLC and MAC) — X.25, Ethernet, Token ring, FDDI, PPP/SLIP, etc.

Physical — 10/100Base-T, 1000Base-T, Fibre Channel, etc.

Fig. 15.1: Comparison of OSI and TCP/IP protocol architectures

Access Network Connected to an ISP

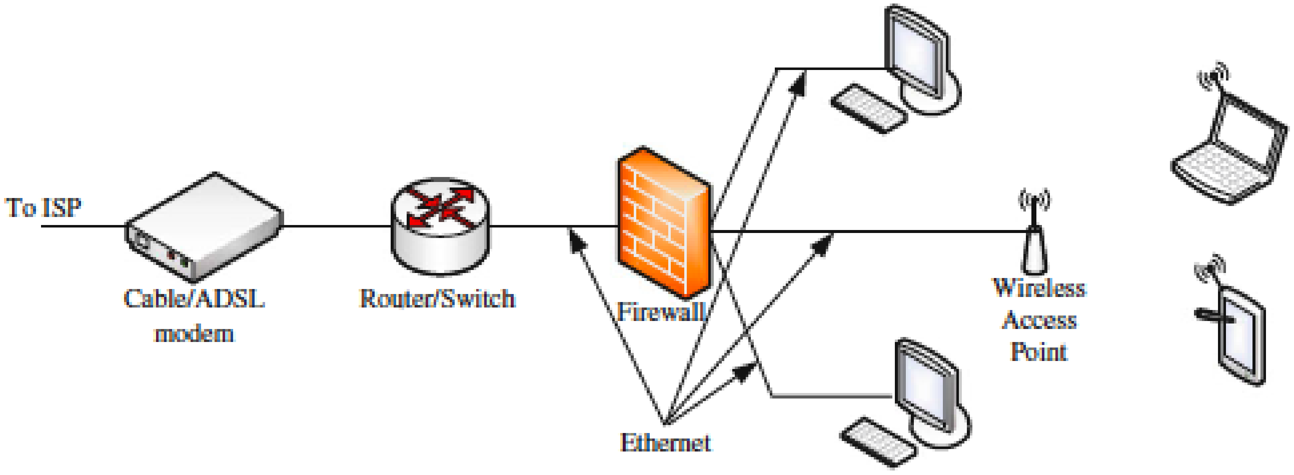


Fig. 15.2: A typical home/office network setup

- The users inside the network are then able to access diverse multimedia services in the public Internet.
- A firewall can protect users from malicious attack.

15.2 Local Area Network and Access Networks

- For home or office users, the networks of direct use is generally a LAN, which isrestricted to a small geographical area.

- The physical links that connect an end system inside a LAN toward the external Internet is referred to as the Access Network.

- It is also known as the “last mile” for delivering network services.

15.2.1 LAN Standards

- The IEEE 802 committee developed the IEEE 802 reference model for LANs, with a focus on the lower layers, namely, the Physical and the Data Link layers.

- In particular, the Data Link layer’s functionality is enhanced, and the layer has been divided into two sublayers:

  1. **Medium Access Control (MAC) layer** This sublayer assembles or disassembles frames upon transmission or reception, performs addressing and error correction, and regulates access control to a shared physical medium.

  2. **Logical Link Control (LLC) layer** This sublayer performs flow and error control and MAC-layer addressing. It also acts as an interface to higher layers. LLC is above MAC in the hierarchy.

LAN Standards (Cont'd)

- **Following are some of the important IEEE 802 subcommittees and the areas they define:**

  1. **802.1 (Higher Layer LAN Protocols)** It concerns the overall 802 LAN architecture, the relationship between the 802.X standards and wide area networks (WAN).

  2. **802.2 (LLC)** The general standard for LLC, which provides a uniform interface to upper layer protocols, masking the differences of various 802.X MAC layer implementations.

  3. **802.3 (Ethernet)** It defines the physical layer and the data link layer’s MAC of the wired Ethernet, in particular the CSMA/CD method.

LAN Standards (Cont'd)

4. **802.11 (Wireless LAN)** It defines the medium access method and physical layer specifications for wireless LAN (WLAN, also known as Wi-Fi).

5. **802.16 (Broadband wireless)** It defines the access method and physical layer specifications for broadband wireless networks. One commercialized product is WiMAX (Worldwide Interoperability for Microwave Access), which targets the delivery of last mile wireless broadband access as an alternative to cable and DSL.

15.2.2 Ethernet Technology

Preamble	Start of frame delimiter	MAC destination	MAC source	Type or Length	Payload Data	CRC
7 bytes	1 bytes	6 bytes	6 bytes	2 bytes	46-1500 bytes	4 bytes

Fig. 15.3: Ethernet frame structure

- Ethernet is a LAN technology initially developed in 1970s, which soon defeated many other competing wired LAN technologies and has since become dominating in the market.

- The basic Ethernet uses a shared bus. Each Ethernet station is given a 48-bitMAC address. The MAC addresses are used to specify both the destination and the source of each data packet, referred to as a *frame*.

Ethernet Technology (Cont'd)

- To send a frame, the recipient's Ethernet address is attached to the frame, which is then broadcast to everyone on the bus.
- For a LAN with multiple stations, often a star topology is used, in which each station is connected directly to a hub (and recently a switch).
- The maximum data rate for the early Ethernet is 10 Mbps, using unshielded twisted pairs.
- The link layer has also evolved to meet new bandwidth and market requirements.

15.2.3 Access Network Technologies

- An access network bridges the LAN in a home or office to the external Internet.

- To save cost for laying a new network line, an existing network that is already in the home is often used, in particular, the telephone or cable TV networks.

- Direct fiber optics connections have been popular nowadays for new buildings.

Dial-Up and Integrated Services Digital Network

- The very earlier Internet accesses are often using the telephone line to establish a dialed connection to an ISP.

- In the 1980s, the *International Telecommunication Union (ITU)* started to develop the *Integrated Service Digital Network (ISDN)* to meet the needs of various digital services.

Digital Subscriber Line

- DSL is the telephone industry's newer answer to the last mile challenge.
- One important technology is Discrete Multi-Tone (DMT), which, for better transmission in potentially noisy channels, sends test signals to all subchannels first.
- DSL uses FDM (Frequency Division Multiplexing) to multiplex three channels:

  1. **The high-speed (1.5-9Mbps) downstream channel** at the high end of the spectrum.
  2. **A medium speed (16-640 kbps) duplex channel.**
  3. **A voice channel** for telephone calls at the low end (0–4 kHz) of the spectrum.

Digital Subscriber Line (Cont'd)

Table. 15.1: Maximum distances for DSL using Twisted-Pair Copper Wires

Data rate (Mbps)	Wire size (mm)	Distance (km)
1.544	0.5	5.5
1.544	0.4	4.6
6.1	0.5	3.7
6.1	0.4	2.7

- Because signals attenuate quickly on twisted-pair lines, and noise increases with line length, even with DMT, the SNR will drop to an unacceptable level after a certain distance. DSL thus has the distance limitations shown in Table 15.1 when using only ordinary twisted-pair copper wires.

Digital Subscriber Line (Cont'd)

Table. 15.2: Different types of Digital Subscriber Lines

Name	Meaning	Data rate	Mode
HDSL	High data rate	1.544Mbps	Duplex
		or 2.048Mbps	
SDSL	Single line	1.544Mbps	Duplex
		or 2.048Mbps	
ADSL	Asymmetric	1.5-9Mbps	Down
		16 – 640kbps	Up
VDSL	Very high data rate	13 - 55Mbps	Down
		1.5 – 3Mbps	Up
VDSL2		200 – 300 Mbps	Down
		100 Mbps	Up

Hybrid Fiber-Coaxial Cable Networks

- Besides telephone lines, another network access that is readily available in many homes is the Cable TV network.

- A cable modem can be used to provides bi-directional data communication via radio frequency channels on this Hybrid Fiber-Coaxial (HFC) network.

- The peak connection speed of a cable modem can be up to 30 Mbps, which is faster than most DSL accesses that are up to 10 Mbps.

- In most areas, both DSL and cable accesses are available, although some areas may have only one choice.

Fiber-To-The-Home or Neighborhood

- Optical fibers can be laid to connect home networks to the core network directly.

- Such fiber-based accesses are considered to be "future-proof" because the data rate of a connection is now only limited by the terminal equipment rather than the fiber.

- It also offers good support for high-quality multimedia services.

  1. For example, AT&T Fiber's FTTH is expected to reach 7 million homes by 2022.

  2. Another example is Google Fiber, which provides Internet connection speeds around 1Gbps (gigabit per second) for both download and upload.

15.3 Internet Technologies and Protocols

- Through the access networks, the home and office users are connected to the external wide area Internet.

- The TCP/IP protocol suite plays the key roles in the Internet, interconnecting diverse underlying networks and serving diverse upper-layer applications.

- The *Internet Engineering Task Force* (IETF) and the *Internet Society* are the principal technical development and standard-setting bodies for the Internet.

15.3.1 Network Layer: IP

- The network layer provides two basic services: **packet addressing** and **packet forwarding.**

- The forwarding is guided by *routing tables* that are collectively built and updated by the routers using *routing protocols*.

- There are two common ways to move data through a network of links and routers:

  1. **Circuit Switching** The *PSTN* is a good example of circuit switching, in which an end-to-end circuit must be established, which is dedicated for the duration of the connection at a guaranteed bandwidth.

  2. **Packet Switching** Packet switching is used for many modern data networks, particularly today’s Internet, in which data rates tend to be variable and sometimes bursty.

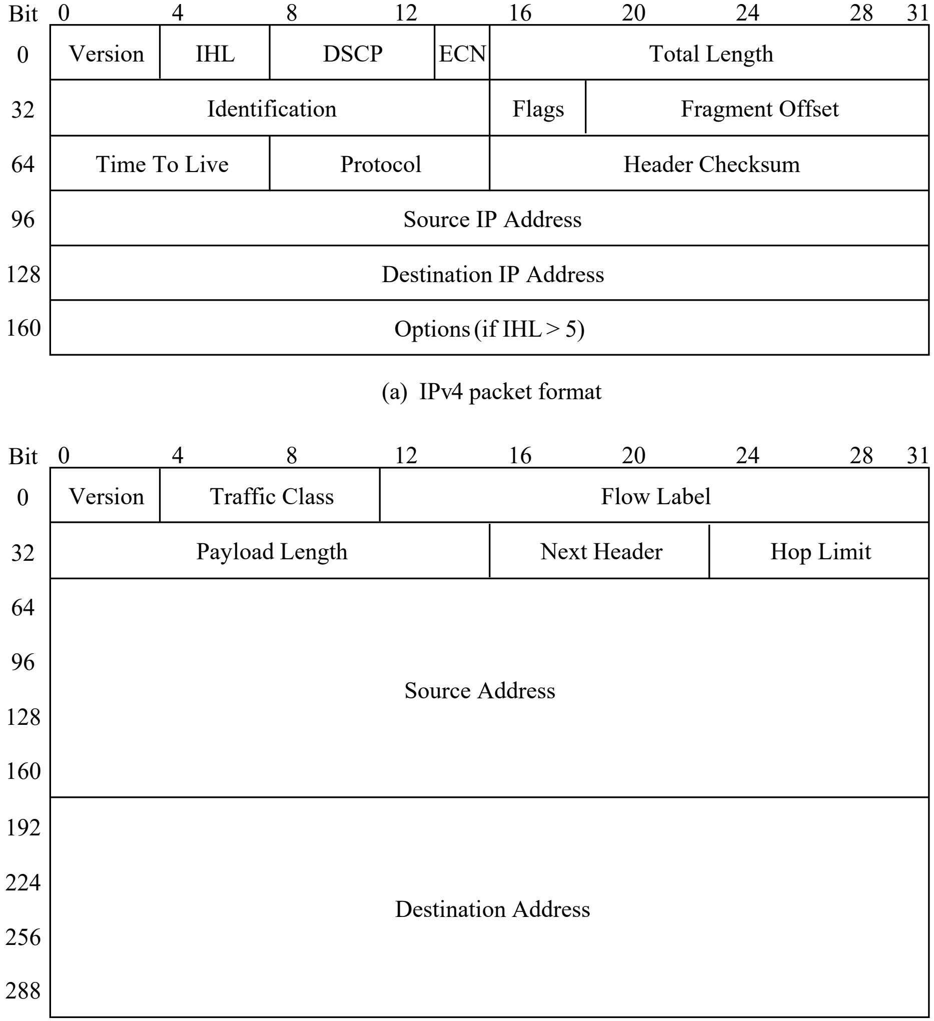
Network Layer: IP (Cont'd)

- For packet switching, two approaches are available to switch and route the packets: ***datagram*** and ***virtual circuit*.**

- In virtual circuits, a route is predetermined through *request* and *accept* by all nodes along the route.

- As a datagram service, IP is *connectionless* and provides no end-to-end control.

- The IP protocol also provides global addressing of computers across all interconnected networks, where every networked device is assigned a globally unique *IP address*.



Bit	0	4	8	12	16
0	Version	IHL	DSCP	ECN	Total Length
32	Identification			Flags	Fragment Offset
64	Time To Live		Protocol		Header Checksum
96	Source IP Address				
128	Destination IP Address				
160	Options (if IHL > 5)				

(a) IPv4 packet format

Bit	0	4	16	24
0	Version	Traffic Class	Flow Label	
32	Payload Length		Next Header	Hop Limit
64–160	Source Address			
192–288	Destination Address			

(b) IPv6 packet format

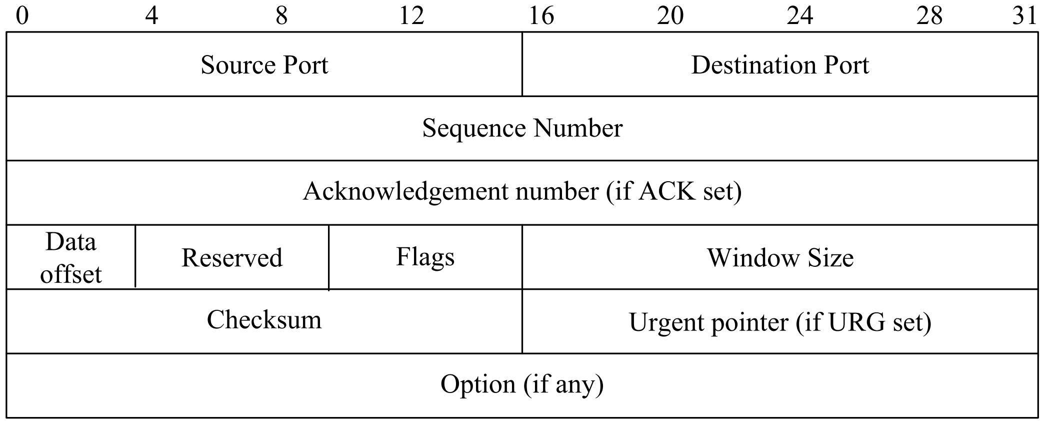
Fig. 15.4: Packet formats of IPv4 and IPv6

15.3.2 Transport Layer: TCP and UDP

- **TCP** and **User Datagram Protocol (UDP)** are two transport layer protocols used in the Internet to facilitate host-to-host (or end-to-end) communications.

1. **TCP** offers a reliable byte pipe for sending and receiving of application messages between two computers, regardless of the specific types of applications.
2. **UDP** is *connectionless* with no guarantee on delivery: if a message is to be reliably delivered, it has to be handled by its own application in the application layer.

Transmission Control Protocol



<table>
<tr><td colspan="3">Source Port</td><td>Destination Port</td></tr>
<tr><td colspan="4">Sequence Number</td></tr>
<tr><td colspan="4">Acknowledgement number (if ACK set)</td></tr>
<tr><td>Data offset</td><td>Reserved</td><td>Flags</td><td>Window Size</td></tr>
<tr><td colspan="3">Checksum</td><td>Urgent pointer (if URG set)</td></tr>
<tr><td colspan="4">Option (if any)</td></tr>
</table>

Fig. 15.5: Header format of a TCP packet

- TCP is ***connection-oriented***: a connection must be established through a *3-way handshake* before the two ends can start communicating.
- To ensure reliable transfer, TCP offers such services as message packetizing, error detection, retransmission, and packet resequencing.

Transmission Control Protocol (Cont'd)

- Each TCP datagram header contains the source and destination ports, sequence number, checksum, window field, acknowledgment number, and other fields.
  - The *source* and *destination ports* are used for the source process to know where to deliver the message and for the destination process to know where to reply to the message.
  - A sequence number reorders the arriving packets and detects whether any are missing.
  - The checksum verifies with a high degree of certainty that the packet arrived undamaged, in the presence of channel errors.

Transmission Control Protocol (Cont'd)

- The checksum verifies with a high degree of certainty that the packet arrived undamaged, in the presence of channel errors.
- The window field specifies how many bytes the destination’s buffer can currently accommodate.
- Acknowledgment (ACK) packets have the ACK number specified—the number of bytes correctly received so far in sequence (corresponding to a sequence number of the first missing packet).

Transmission Control Protocol (Cont'd)

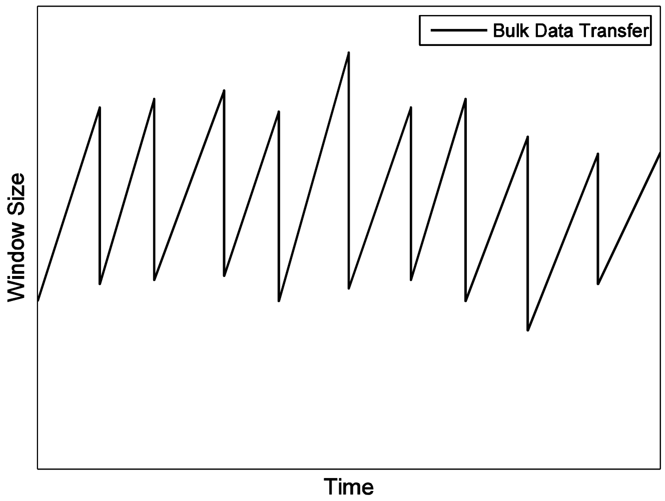


Fig. 15.6: Sawtooth behavior in TCP data transfer

- TCP also implements a *congestion control* mechanism in response to network congestion, which can be observed by packet losses.

- TCP congestion window will grow linearly when there is no congestion, but when there is a packet loss, it can instantly reduce to half of the window size that is before congestion.

User Datagram Protocol

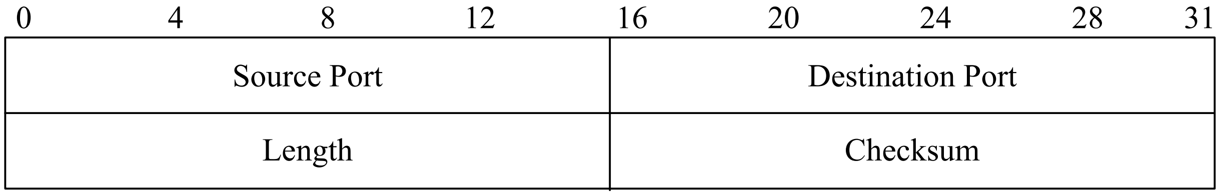


Fig. 15.7: Header format of a UDP datagram

- UDP is *connectionless* with no guarantee on delivery
- Essentially, the only thing UDP provides is multiplexing using port numbers and error detection through a checksum.
- Given the low header overhead and the removal of connection setup, UDP data transmission can be faster than TCP. It is however unreliable, especially in a congested network.

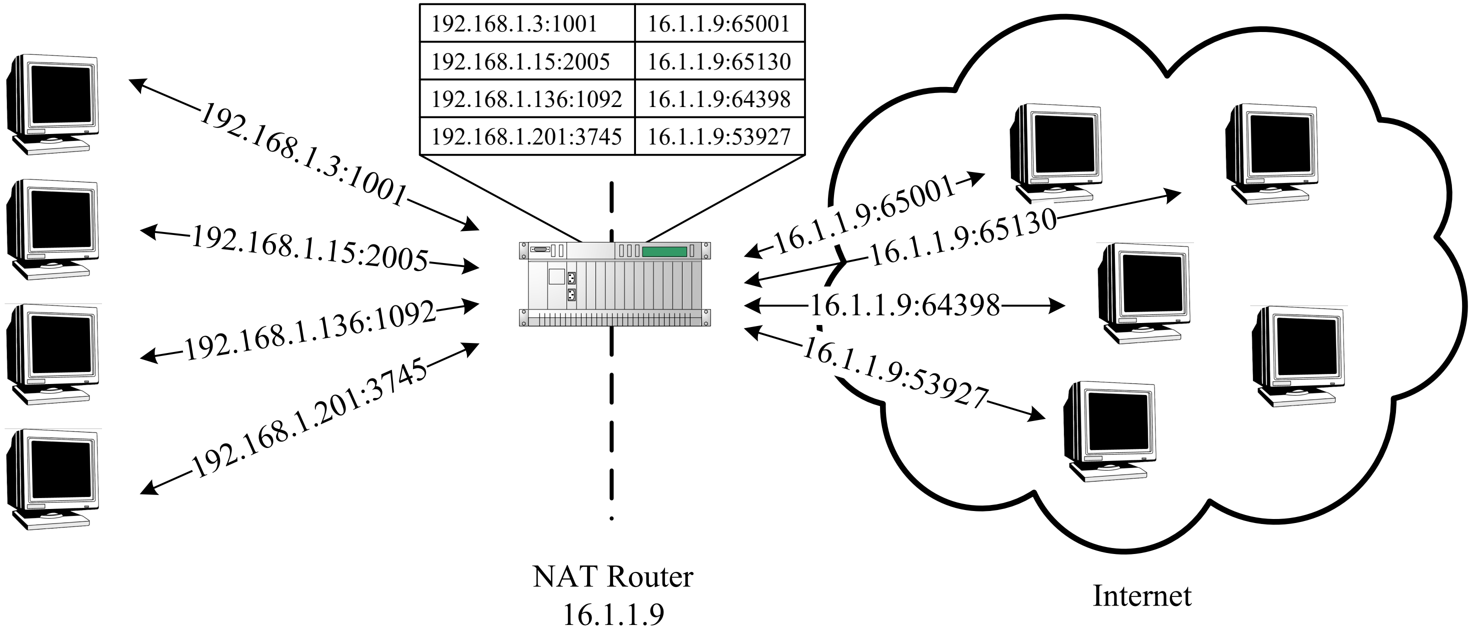
TCP-Friendly Rate Control

- The sawtooth behavior of the window-based TCP congestion control is not well suited for media streaming, but an uncontrolled UDP flow can be too aggressive.

- *TCP-Friendly Rate Control (TFRC)* has been introduced, which ensures a UDP flow to be reasonably fair when competing for bandwidth with TCP flows.

- TFRC is generally implemented by estimating the equivalent TCP throughput over the same path using parameters that are observable by the sender or the receiver.

15.3.3 Network Address Translation and Firewall



192.168.1.3:1001	16.1.1.9:65001
192.168.1.15:2005	16.1.1.9:65130
192.168.1.136:1092	16.1.1.9:64398
192.168.1.201:3745	16.1.1.9:53927

Fig. 15.8: An illustration of Network Address Translation (NAT)

- Even though The 32-bit IPv4 addressing in principle allows $2^{32} \approx 4$ billion addresses, in reality, it has already largely been exhausted.
- To solve the IPv4 address shortage, a practical solution is *Network Address Translation* (NAT).

NAT and Firewall (Cont'd)

- While NAT alleviates the IP address shortage problem, it imposes fundamental restrictions on pair-wise connectivity of nodes, and may prohibit direct communication with one another.

- Similar penetration problem happens for a *firewall*, which is a software or hardware-based network security system that controls the incoming and outgoing network traffic based on a rule set.

- The connectivity constraints are a significant challenge to the viability for multimedia content distribution mechanisms over the Internet, particularly for peer-to-peer sharing.

15.4 Multicast Extension

- In network terminology, a *broadcast* message is sent to all nodes in a domain, a *unicast* message is sent to only one node, and a *multicast* message is sent to a set of specified nodes.

- A large number of emerging applications require support for broadcast or multicast, i.e., simultaneous content delivery to a large number of receivers.

- In the Internet environment, the primary issue for multicast is to determine at which layer it should be implemented.

  1. pushed to higher layers if possible
  2. implemented at the lower layer can achieve significant performance benefits that outweigh the cost of additional complexity.

15.4.1 Router-Based Architectures: IP Multicast

- For much of the 1990s, the research and industrial community mainly focused on the router-based *IP Multicast* architecture.

- IP multicast has open anonymous group membership.

- The *Internet Group Management Protocol (IGMP)* was designed to help the maintenance of multicast groups.

- Multicast routing is generally based on a shared tree: once the receivers join a particular IP multicast group, a multicast distribution tree is constructed for that group.

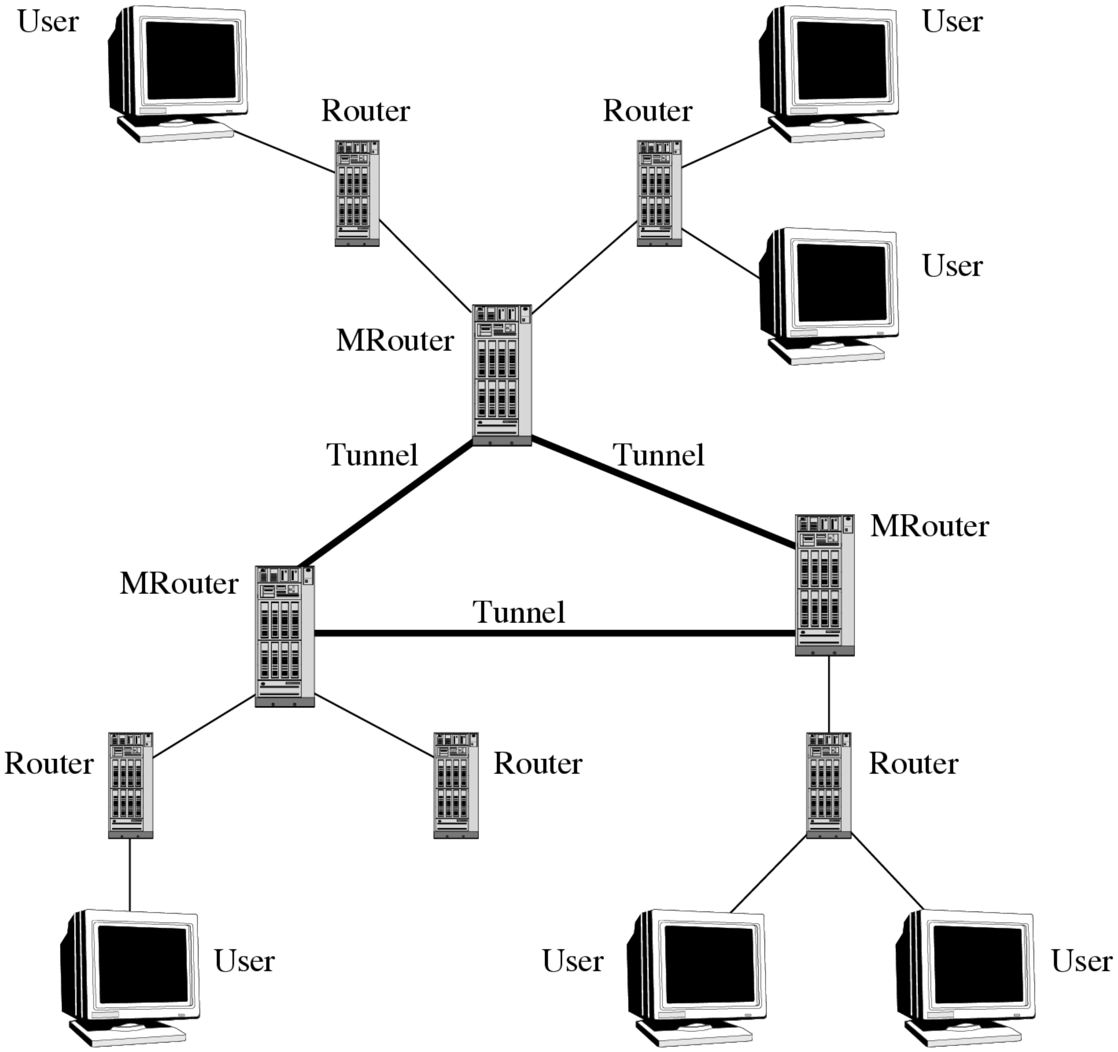


Fig. 15.9: Tunnels for IP Multicast in MBone

IP Multicast (Cont'd)

- **Benefits**
  - IP multicast is a loosely coupled model that reflects the basic design principles of the Internet.
  - Given that the network topology is best-known in the network layer, multicast routing in this layer is also the most efficient.

- **Drawbacks**
  - Providing higher level features such as error, flow, and congestion control has been shown to be more difficult than in the unicast case.
  - In general, UDP (not TCP) is used in conjunction with IP multicast, so as to avoid too many ACKs from TCP receivers.

15.4.2 Non Router-Based Multicast Architectures

- IP multicast calls for changes at the infrastructure level, i.e., in network routers. This introduces high complexity and serious scaling constraints.

- Moving multicast functionality to end systems has the potential to address many of the problems associated with IP multicast.

- Given that nonrouter-based architectures push functionality to the network edges, there are several choices in instantiating such an architecture.

- While the application layer solutions have the promise to enable ubiquitous deployment, they often involve a wide range of autonomous users that may not provide as good performance and easily fail or leave at will.

15.5 Quality-of-Service for Multimedia Communications

- **Challenges in multimedia network communications arise due to a series of distinct characteristics of audio/video data:**
  - **Voluminous and Continuous** They demand high data rates, and often have a lower bound to ensure continuous playback.
  - **Real-Time and Interactive** They demand low startup delay and synchronization between audio and video for “lip sync”.
  - **Rate fluctuation** The multimedia data rates fluctuate drastically and sometimes bursty.

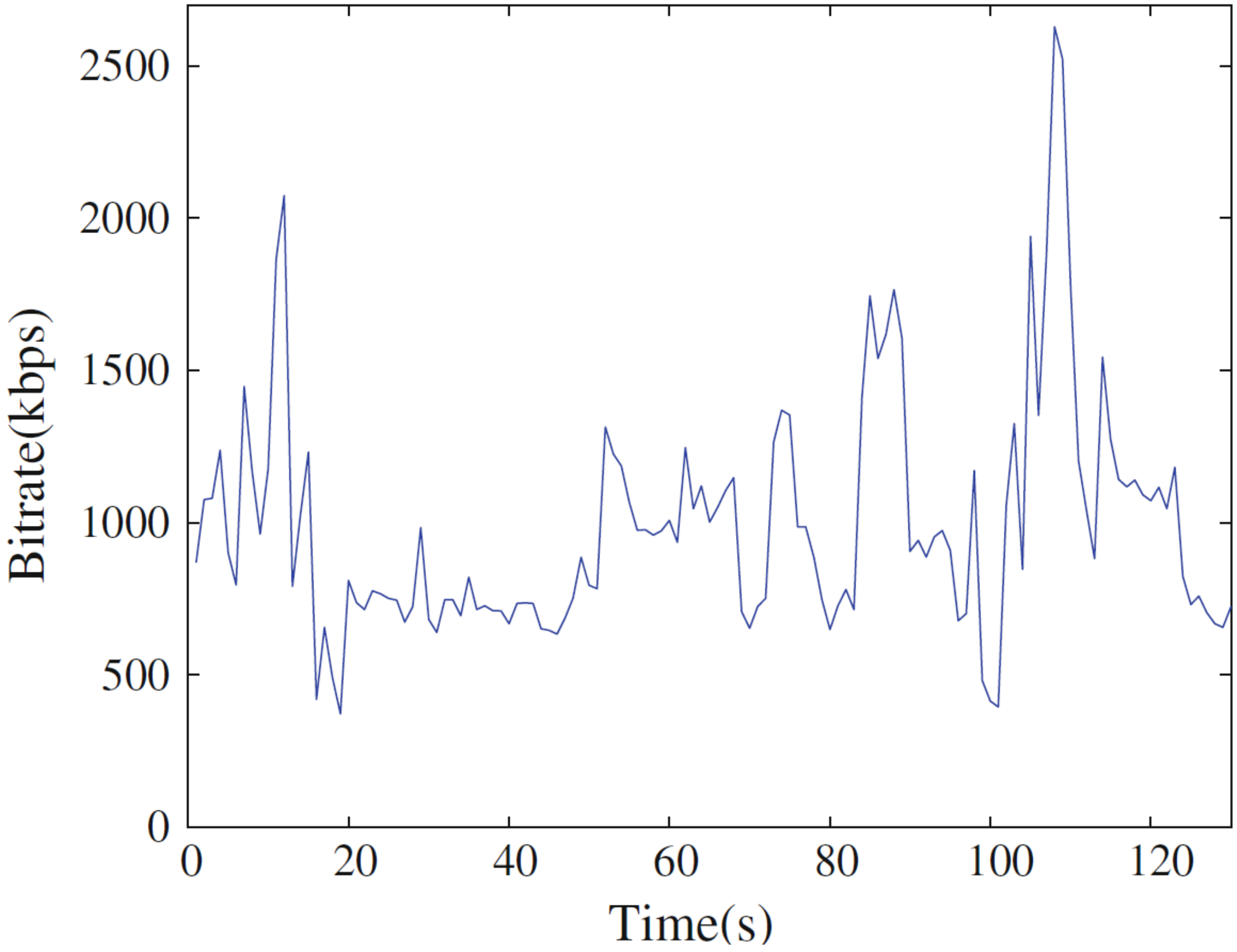


Fig. 15.10: The bitrate over time of an MPEG-4 video

15.5.1 Quality of Service

- **QoS for multimedia data transmission depends on many parameters.**
  - **Bandwidth** A measure of transmission speed over digital links or networks.
  - **Latency (maximum frame/packet delay)** The maximum time needed from transmission to reception.
  - **Packet loss or error** A measure (in percentage) of the loss- or error rate of the packetized data transmission.
  - **Jitter (or delay jitter)** A measure of smoothness (along time axis) of the audio/video playback.
  - **Sync skew** A measure of multimedia data synchronization.

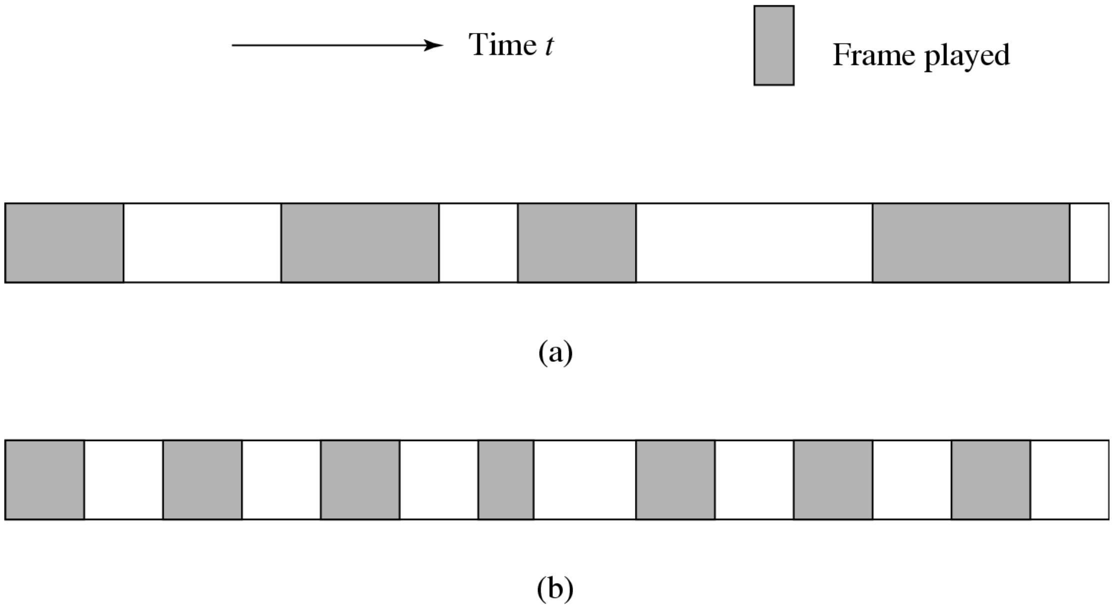


Fig. 15.11: Jitters in frame playback: (a) high jitter (b) low jitter

Multimedia Service Classes

- **We now list a set of the typical multimedia applications of different QoS demands:**
  - Two-way traffic, low latency and jitter, possibly with prioritized delivery, such as voice telephony and video.
  - Two-way traffic, low loss and low latency, with prioritized delivery, such as e-commerce applications.
  - Moderate latency and jitter, strict ordering and sync.
  - No real-time requirement, such as downloading or transferring large files (movies).

Table. 15.4: Requirement on network bandwidth/bitrate

Application	Speed requirement
Telephone	16 kbps
Audio conferencing	32 kbps
CD-quality audio	128-192 kbps
Digital music (QoS)	64-640 kbps
H.261	64 kbps-2 Mbps
H.263	<64 kbps
H.264	1-12 Mbps
MPEG-1 video	1.2-1.5 Mbps
MPEG-2 video	4-60 Mbps
MPEG-4 video	1-20 Mbps
HDTV (compressed)	>20 Mbps
HDTV (uncompressed)	>1 Gbps
4K (compressed)	> 4 Gbps
MPEG-4 video-on-demand (QoS)	250-750 kbps
Videoconferencing (QoS)	384 kbps-2 Mbps

Table. 15.5: Tolerance of latency and jitter in digital audio and video

Application	Average latency tolerance (msec)	Average jitter tolerance (msec)
Low-end videoconference (64 kbps)	300	130
Compressed voice (16 kbps)	30	130
MPEG NTSC video (1.5Mbps)	5	7
MPEG audio (256 kbps)	7	9
HDTV video (20Mbps)	0.8	1

Quality-of-Experience (QoE): User Perceived QoS

- Although QoS is commonly measured by the above technical parameters, it itself is a "collective effect of service performances that determine the degree of satisfaction of the user of that service."

- Many issues of perception can be exploited in achieving the best perceived QoS in networked multimedia.
  - For example, in real-time multimedia, regularity is more important than latency and temporal correctness is more important than the sound and picture quality.
  - Humans also tend to focus on one subject at a time; a user's focus is usually at the center of a screen, and it takes time to refocus.

Quality-of-Experience (QoE): User Perceived QoS

- *Quality of Experience (QoE)* is a measure of the delight or annoyance of a customer’s experiences with a service, e.g., video streaming. More formally, ITU adopted the following definition in its Recommendation ITU-T P.10 (2016):

The degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the users personality and current state.

15.5.2 Internet QoS

- The conventional IP provides the “best-effort” service only, which does not differentiate among different applications.
- There have been significant efforts toward data networking with better or even guaranteed QoS, and a representative is the ATM network.
- There are two common approaches.
  - ***IntServ*** or **integrated services** is an architecture that specifies the elements to guarantee QoS in fine-grains for each individual flow.
  - ***DiffServ*** or **differentiated services** specifies a simple, scalable, and coarse-grained class-based mechanism for classifying and managing aggregated network traffic and providing specific QoS to different classes of traffic.

Integrated Service and Resource ReSerVation Protocol

- In IntServ, *Flow Specs* describe what the resource reservation is for a flow, while the *Resource ReSerVation Protocol* (RSVP) is used as the underlying mechanism to signal it across the network.

- RSVP is a setup protocol for Internet resource reservation, which targets a multicast setup for general multimedia applications.
  - ***RSVP is receiver-initiated*** A receiver (at a leaf of the multicast tree) initiates the reservation request Resv, and the request travels back toward the sender but not necessarily all the way.

  - ***RSVP creates only soft state*** The receiver host must maintain the soft state by periodically sending the same Resv message; otherwise, the state will time out.

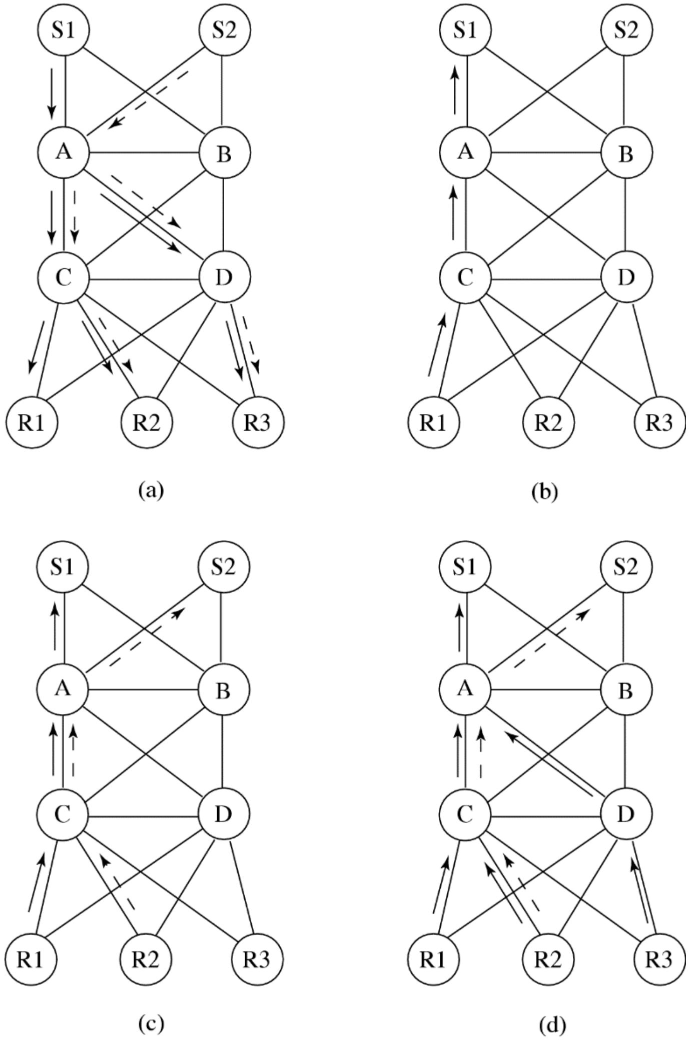


Fig. 15.12: A scenario of network resource reservation with RSVP

Differentiated Service

- Differentiated Service (DiffServ) operates on the principle of traffic aggregation and classification.
- In DiffServ, network routers implement *per-hop behaviors* (PHBs), which define the packet-forwarding properties associated with a class of traffic.
- Different PHBs may be defined to offer different services within a multimedia application data:
  - **Uncompressed audio** PCM audio bitstreams can be broken into groups of every nth sample—prioritize and send k of the total of n groups (k ≤ n) and ask the receiver to interpolate the lost groups if so desired.
  - **JPEG image** The different scans in Progressive JPEG and different resolutions of the image in hierarchical JPEG can be given different services.
  - **Compressed video** To minimize playback delay and jitter, the best service can be given to the reception of I-frames and the lowest priority to B-frames.

Differentiated Service (Cont'd)

- In practice, most networks use the following commonly defined PHB:
  - **Default PHB**, which is typically the best-effort service.
  - **Expedited Forwarding (EF)**, which is dedicated to low-loss, low-latency traffic.
  - **Assured Forwarding (AF)**, which achieves assurance of delivery under prescribed conditions.
  - **Class Selector PHBs**, which maintain backward compatibility with non-DiffServ traffic

- One implementation may divide network traffic in AF into the following categories and allocate bandwidth accordingly:
  - **Gold**: Traffic in this category is allocated 50% of the available bandwidth.
  - **Silver**: Traffic in this category is allocated 30% of the available bandwidth.
  - **Bronze**: Traffic in this category is allocated 20% of the available bandwidth.

Differentiated Service (Cont'd)

- Compared with IntSev, DiffServ has coarser control granularity (in aggregated classes, rather than individual flows), and is therefore simpler and scales well.

- However, DiffServ is not necessarily exclusive to each other.

- In real-world deployment, IntServ and DiffServ may work together to accomplish the QoS targets with reasonable costs.

15.5.3 Network Softwarization and Virtualization: SDN and NVF

- IntServ and DiffServ have been implemented in many of today's Internet routers; however, their use in wide-area networks have been limited.
  - The complexity of maintaining these services in large-scale dynamic networks can be quite high;
  - Complete end-to-end QoS guarantee is generally difficult, so for service differentiation.

- More importantly, in the traditional Internet design, the control plane and the data plane, as well as the software and hardware for them, are tightly coupled, making any change to the network core very difficult to be deployed.

Software-Defined Networking (SDN)

- Decouple network control and forwarding functions.
- Enabling network control to become directly programmable and the underlying infrastructure to be abstracted from applications and network services.
- Representative: OpenFlow
  - Highly programmable.
  - Centrally manageable.
  - Openly available.

Network Functions Virtualization (NFV)

- Virtualize the entire classes of network node functions into building blocks that may connect, or chain together, to create communication services.

- Greatly reduce capital and operational expenditures, and accelerate service and product deployment network control and forwarding functions.

- Closely related to SDN, but does not necessarily depends on SDN.

15.5.3 Rate Control and Buffer Management

- IntServ and DiffServ functions have been implemented in many of today's Internet routers; however, their use in wide area networks remain limited.
  - First, the complexity of maintaining these services in large-scale dynamic networks can be quite high, particularly for flow-based RSVP;
  - Second, the scale and heterogeneity of Internet terminals and routers make a complete end-to-end QoS guarantee generally difficult, so for service differentiation.

- As such, most of the time, a networked multimedia application still has to assume that the underlying network is of the best-effort service, and adaptive transmission and control are to be used

Rate Control and Buffer Management (Cont'd)

- A key concern here is rate fluctuation with multimedia data, and VBR coding is often used.

- To cope with the variable bitrate and network load fluctuation, buffers are usually employed at both sender and receiver ends.

- A *prefetch buffer* can be introduced at the client side to smooth the transmission rate (reducing the peak rate).

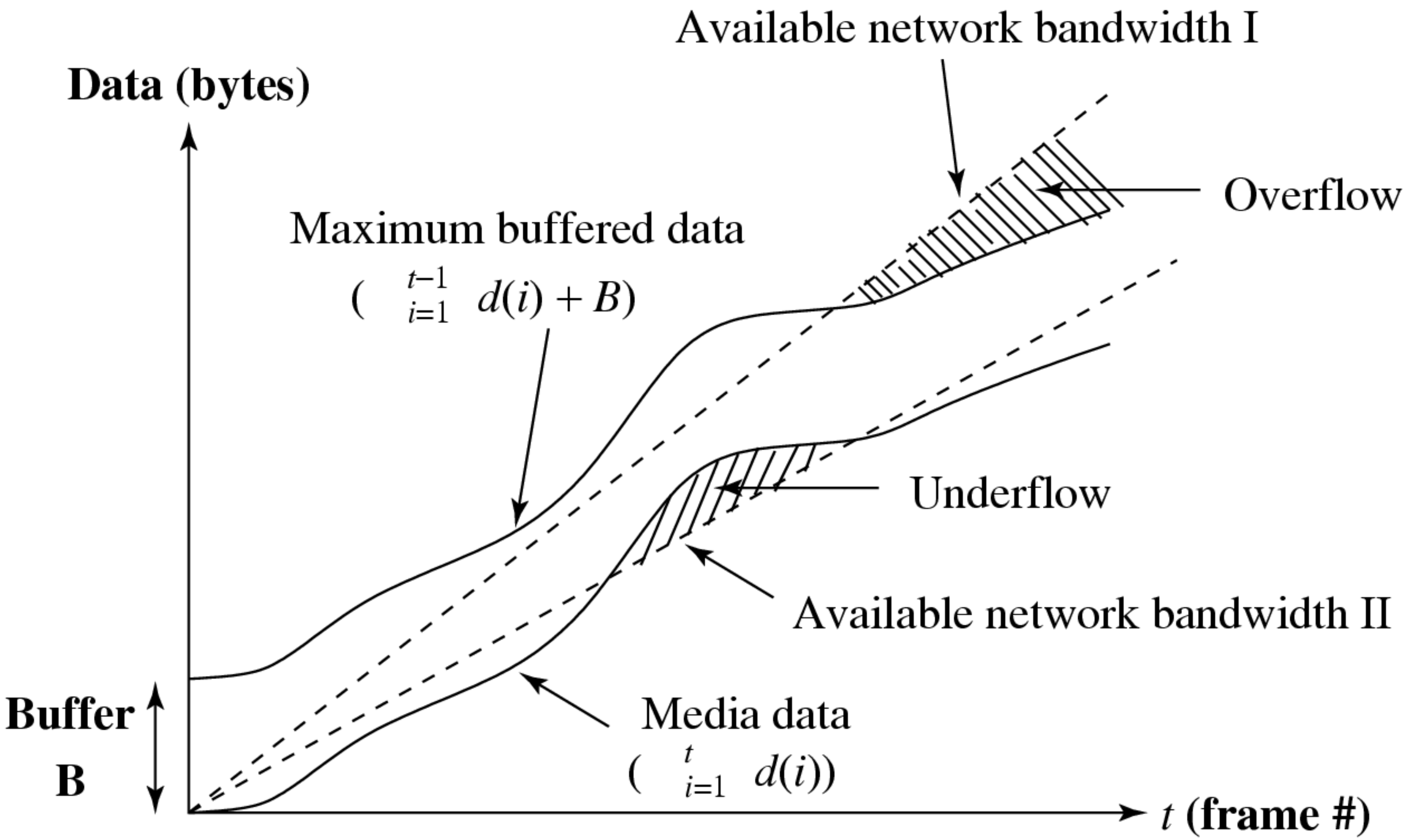


Fig. 15.13: The data that a client can store in the buffer assists the smooth playback of the media when the media rate exceeds the available network bandwidth

Rate Control and Buffer Management (Cont'd)

- If the media is sent as fast as possible without buffer considerations (as in normal file downloads), then toward the end of the video, the data received will be greater than the buffer can store at the time.
  - To address this, we need to prefetch video data to fill the buffer and try to transmit at the mean video bitrate
  - to keep the buffer full without exceeding the available bandwidth, which can be estimated as the TCP-friendly bandwidth.

- If the data rate characteristics are known in advance, it is possible to use the prefetch buffer more efficiently for the network.
- The media server can plan ahead for a transmission rate such that the media can be viewed without interruption and with minimized bandwidth.

15.6 Protocols for Multimedia Transmission and Interaction

- Review the protocols for multimedia communications
- Build on top of UDP or TCP
- Work with the best-effort Internet or with IntServ or Diff-Serv to provide quality multimedia data transmission, particularly in the streaming model
- Enable various interactions between a media server and its clients

15.6.1 HyperText Transfer Protocol

- HTTP is a protocol that was originally designed for transmitting Web content, but it also supports transmission of any file type.

- HTTP is a “stateless” request/response protocol.

- The *Uniform Resource Identifier* (URI) identifies the resource accessed, such as the host name, always preceded by the token “http://” or “https://”.

- HTTP builds on top of TCP to ensure reliable data transfer.

15.6.2 Real-Time Transport Protocol

- Real-Time Transport Protocol (RTP), is designed for the transport of real-time data, such as audio and video streams.

- RTP's design follows two key principles, namely
  - **application layer framing**, i.e., framing for media data should be performed properly by the application layer.
  - **integrated layer processing**, i.e., integrating multiple layers into one to allow efficient cooperation.

- RTP usually runs on top of UDP, which provides an efficient (albeit less reliable) connectionless transport service.

Real-Time Transport Protocol (Cont'd)

- There are three main reasons for using UDP instead of TCP.
  - First, TCP is a connection-oriented transport protocol; hence, it is more difficult to scale up in a multicast environment.
  - Second, TCP achieves its reliability by retransmitting missing packets. As mentioned earlier, multimedia data transmissions is loss-tolerant and perfect reliability is not necessary.
  - Last, the dramatic rate fluctuation (sawtooth behavior) in TCP is often not desirable for continuous media.

Real-Time Transport Protocol (Cont'd)

- RTP introduces the following additional parameters in the header of each packet:
  - **Payload type** indicates the media data type as well as its encoding scheme.
  - **Timestamp** is the most important mechanism of RTP.
  - **Sequence number** is to complement the function of time stamping.
  - **Synchronization source (SSRC) ID** identifies the sources of multimedia data.
  - **Contributing Source (CSRC) ID** identifies the source of contributors, such as all speakers in an audio conference.

15.6.3 RTP Control Protocol

- RTP Control Protocol (RTCP) is a companion protocol of RTP.
- RTCP provides a series of typical reports:
  - **Receiver report (RR)** provides quality feedback.
  - **Sender report (SR)** provides information about the reception of RR, number of packets/bytes sent, and so on.
  - **Source description (SDES)** provides information about the source.
  - **Bye** indicates the end of participation.
  - **Application-specific functions (APP)** provides for future extension of new features.

15.6.4 Real-Time Streaming Protocol

- The Real-Time Streaming Protocol (RTSP) is a signaling protocol to control streaming media servers and is used for establishing and controlling media sessions between end points.

- Four typical RTSP operations:
  - **Requesting presentation description:** the client issues a DESCRIBE request to the Stored Media Server to obtain the presentation description.
  - **Session setup:** the client issues a SETUP to inform the server of the destination IP address, port number, protocols, and TTL (for multicast).
  - **Requesting and receiving media:** after receiving a PLAY, the server starts to transmit streaming audio/video data, using RTP.
  - **Session closure:** TEARDOWN closes the session.

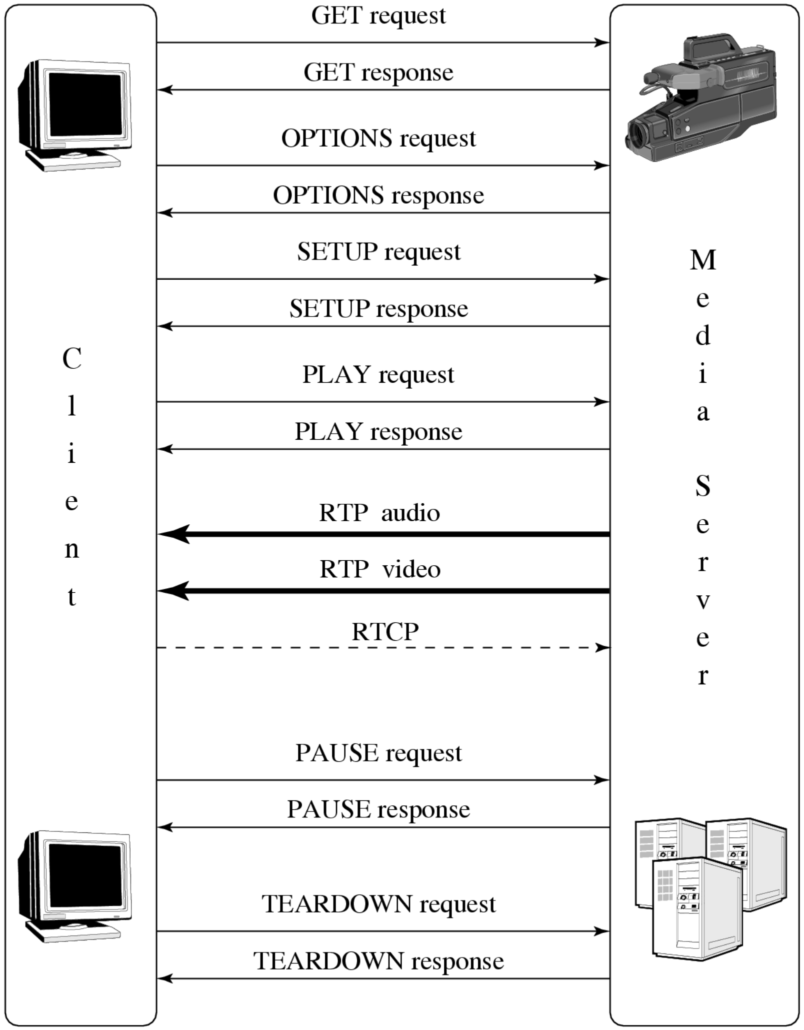


Fig. 15.15: A scenario of RTSP operations

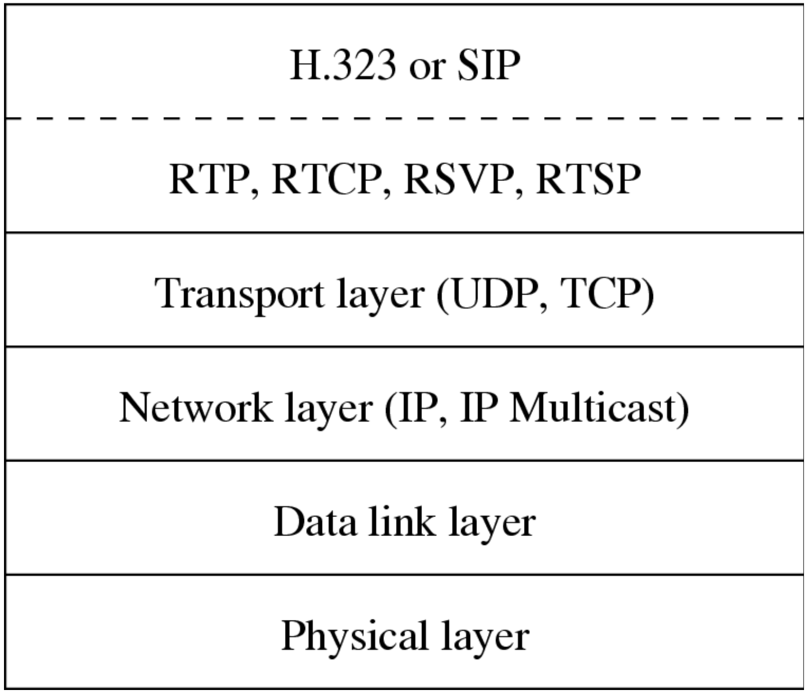
15.7 Case Study: Internet Telephony

- With ever-increasing network bandwidth and the ever-improving quality of multimedia data compression, **Internet telephony** has become a reality.
- Main advantages of Internet telephony over *POTS* (*Plain Old Telephone Service*):
  - Provides great flexibility and extensibility in accommodating IntServ such as voicemail, video conversations, live text messages, etc.
  - Uses packet switching, network usage is much more efficient (voice communication is bursty and VBR-encoded).

- With the technologies of multicast or multipoint communication, multiparty calls are not much more difficult than two-party calls.
- With advanced multimedia data-compression techniques, various degrees of QoS can be supported and dynamically adjusted according to the network traffic.
- Richer graphical user interfaces can be developed to show available features and services, monitor call status and progress, etc.

Case Study: Internet Telephony (Cont'd)

- As shown in Fig. 15.16, the transport of real-time audio (and video) in Internet telephony is supported by RTP (whose control protocol is RTCP).
- Streaming media is handled by RTSP and Internet resource reservation is taken care of by RSVP.



H.323 or SIP

RTP, RTCP, RSVP, RTSP

Transport layer (UDP, TCP)

Network layer (IP, IP Multicast)

Data link layer

Physical layer

Fig. 15.16: Network protocol structure for Internet telephony

15.7.1 Signaling Protocols: H.323 and Session Initiation Protocol

- Acceptance of a call via Internet telephony depends on the callee’s current location, capability, availability, and desire to communicate, which requires advanced signaling protocols.
  - **H.323 Standard**
    - H.323 is an ITU standard for packet-based multimedia communication services.
  - **Session Initiation Protocol (SIP)**
    - SIP is IETF’s recommendation (RFC 3261) for establishing and terminating sessions in Internet telephony.

H.323 Standard

- **H.323 standard** specifies signaling protocols and describes terminals, multipoint control units (for conferencing), and gateways for integrating Internet telephony with General Switched Telephone Network (GSTN)4 data terminals. The H.323 signalling process consists of two phases :
  - **Call setup**
    - The caller sends the gatekeeper (GK) a Registration, Admission and Status (RAS) Admission Request (ARQ) message.
  - **Capability exchange**
    - An H.245 control channel will be established, for which the first step is to exchange capabilities of both the caller and callee

H.323 Standard (Cont’d)

- **Signaling and Control**
  - **H.225** Call control protocol, including signaling, registration, admissions, packetization, and synchronization of media streams.
  - **H.245** Control protocol for multimedia communications—for example, opening and closing channels for media streams, obtaining gateway between GSTN and Internet telephony.
  - **H.235** Security and encryption for H.323 and other H.245-based multimedia terminals.
- **Audio Codecs**
  - **G.711** Codec for 3.1kHz audio over 48, 56, or 64 kbps channels. G.711 describes PCM for normal telephony.
  - **G.722** Codec for 7kHz audio over 48, 56, or 64 kbps channels.

Session Initiation Protocol

- An application-layer control protocol in charge of the establishment and termination of sessions in Internet telephony.
  - SIP is a text-based protocol, also a client-server protocol.
- SIP can advertise its session using email, news group, web pages or directories, or SAP — a multicast protocol.

Session Initiation Protocol

- The methods (commands) for clients to invoke:
  - **INVITE**: invites callee(s) to participate in a call.
  - **ACK**: acknowledges the invitation.
  - **OPTIONS**: enquires media capabilities without setting up a call.
  - **CANCEL**: terminates the invitation.
  - **BYE**: terminates a call.
  - **REGISTER**: sends user's location info to a Registrar (a SIP server).

Session Initiation Protocol (Cont'd)

- Scenario of a SIP Session
  - **Step 1** Caller sends an INVITE john@home.ca to the local Proxy server P1.
  - **Step 2** The proxy uses its Domain Name Service (DNS) to locate the server for John@home.ca and sends the request to it.
  - **Steps 3,4** john@home.ca is not logged on the server. A request is sent to the nearby location server. John's current address, john@work.ca, is located.
  - **Step 5** Since the server is a redirect server, it returns the address john@work.ca to the proxy server P1.

Session Initiation Protocol (Cont’d)

– **Step 6** Try the next proxy server P2 for john@work.ca.

– **Steps 7,8** P2 consults its location server and obtains John’s local address, john_doe@my.work.ca.

– **Steps 9,10** The next-hop proxy server P3 is contacted, which in turn forwards the invitation to where the client (callee) is.

– **Steps 11-14** John accepts the call at his current location (atwork) and the acknowledgments are returned to the caller.

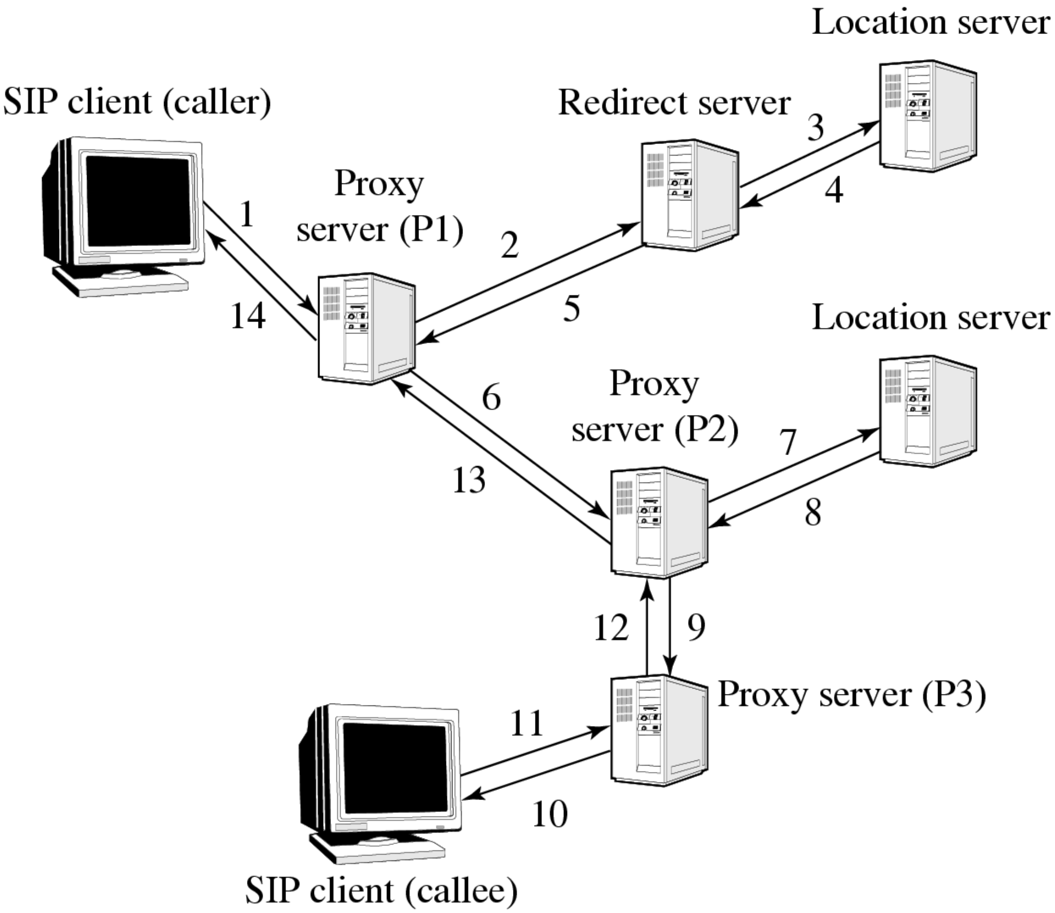


Fig. 15.17: A possible scenario of SIP session initiation

15.8 Further Exploration

- **Text books:**
  - Computer Networks (5th ed.) by A.S. Tanenbaum and D. J. Wetherall
  - Data and Computer Communications(10th ed.) by W. Stallings
  - Computer Networking: A Top-Down Approach (6th edn) by J.F. Kurose and K.W. Ross

- **RFCs (can be found from IETF):**
  - Criteria for evaluating reliable multicast transport protocols.
  - Protocols for real-time transmission of multimedia data (RTP, RTSP, and RSVP).
  - Protocols for VoIP (SIP, SDP, and SAP).
  - DiffServ and MPLS IETF