MULTIMEDIA COMMUNICATION

Dr. Quang Duc Tran
About ME

• Contact Information
  ▫ Room B1-801 (10 am - 11 am Every Tuesday)
  ▫ Department of Data Communication and Computer Network
  ▫ School of Information and Communication Technology
  ▫ Hanoi University of Science and Technology
  ▫ E-mail: ductq@soict.hust.edu.vn
  ▫ Tel: (+84) (4) 38682596

• Education
  ▫ PhD in Information Engineering, 2014
  ▫ MSc in Electrical Engineering, 2008
Aim of this Module

- Describe the theory and operation of the major technologies and equipment of relevance to the media and information industries
- Introduce you to the multimedia communications and its range of applications and networking infrastructures
- Learn about different media types (text, images, speech, audio and video) and applications (VoIP, Video on Demand, multimedia electronic mail, interactive television and others)
Syllabus

- Topic 1: Introduction to Multimedia Communication
- Topic 2: Multimedia Signal Processing
- Topic 3: Multimedia Signal Processing (Cont.)
- Topic 4: Multimedia Communications
- Topic 5: Multimedia Applications
- Week 6: Mini Project (1)
- Week 7: Mini Project (2)
References


References


INTRODUCTION

Dr. Quang Duc Tran
Multimedia Concepts

• Multimedia:
  ▫ Multimedia = Integration of multimedia data (text, speech, audio, video and images)
• Multimedia Communication is originated in a confluence of two technological trends, i.e., Multimedia computing and Networking.
• Applications: Video/Audio conferencing, education, entertainment, and monitoring of remote locations, etc.
Data vs. Signal

• To be transmitted, data must be transformed to electromagnetic signals. Signals is continuously varying electromagnetic wave that may be propagated over a variety of medium, depending on spectrum.

• Data can be analog or digital. **Analog data** takes on continuous values on some interval (e.g., voice and video). **Digital data** takes on discrete values (e.g., text, integer).

• Signals can be analog or digital. Analog signal have an infinite number of values in a range; Digital signals can have only a limited number of values.
Periodic vs. non-periodic signal

- A periodic signal completes a pattern within a measurable time frame, called period, and repeats such a pattern over subsequent identical periods. One full pattern is known as a cycle.

- A non-periodic signal changes without exhibiting a pattern or cycle that repeats over time.

- In data communications, we commonly use periodic analog signals and non-periodic digital signals.
Multimedia Data

- **Text**: Block of characters, each represented by a fixed number of binary digits (bit), known as *codeword*.
- **Still images** (e.g., graphics, pictures): Two dimensional blocks of picture elements represented by a fixed number of bits.
- **Sound and Video**: Type of signal is known as an analogue signal and varies continuously with time (e.g., telephone conversation can last for several minutes while a movie (audio + video) can last for a number of hours).
- **Animation**: Sequence of graphics.
Media Type Classification

- **Captured vs. synthesized media**
  - Captured media (natural): information captured from the real world (examples?).
  - Synthesized media (artificial): information synthesized by the computer (examples?).

- **Discrete vs. continuous media**
  - Discrete media: spaced-based, media involves the space dimension only (examples?).
  - Continuous media: time-based, media involves both the space and the time dimension (examples?)
Media Type Classification

- Sound
- Video
- Animation
- Image
- Text
- Graphics

Captured from real world
Synthesized by computer
Text

- Plain text
  - Unformatted
  - Characters coded in binary form
  - ASCII code
  - All characters have the same style and font
- Rich text (RTF)
  - Formatted
  - Contains format information besides codes for characters
  - No predominant standards
  - Characters of various size, shape and style, e.g., bold, colorful
Text Coding & Compression

- **Text coding**
  - **ASCII**
    - Standard code table (7 bit – 128 characters)
    - Extended code table (8 bit – 256 characters)
  - **Unicode**
    - 16 bit system (65,536 characters)
    - > 110,187 graphical, formatting and control characters

- **Text compression**
  - Statistical compression: Huffman code
  - Compression using dictionary: Lempel-Ziv
  - Compression rate: 1/2 - 2/3 document size
Graphics

- Revisable document that retains structural information
- Consists of objects such as lines, curves, circles, etc.
- Usually generated by graphic editor of computer programs, such as Corel Draw, Adobe Illustrator
Still Image

- 2D matrix consisting of pixels
  - Pixel: smallest element of resolution of the image
  - One pixel is represented by a number of bits
  - Pixel depth: the number of bits available to encode the pixel
- Have no structural information
- Possibly scanned and synthesized still image
Still Image (Cont.)

- Examples of images
  - Binary image: pixel depth 1
  - Grayscale image: pixel depth 8
  - Color image: pixel depth 24
Graphics vs. Still Image

- **Graphics**
  - Revisable documents
  - Document format retains structural information
  - Semantic content is preserved in presentation
  - Described by objects

- **Images**
  - No revisable
  - Document format is unaware of any structural information
  - Semantic content is NOT preserved
  - Described as bitmaps formed of individual pixels
Image Compression

• **Lossless compression**
  ▫ Run-Length coding (RLC)
  ▫ Lempel-Ziv coding
  ▫ GIF, BMP, TIFF

• **Lossy compression**
  ▫ Transform coding
  ▫ Chroma sub-sampling
  ▫ JPEG, JPEG2000
Video

- Video - moving images or moving pictures
  - Captured or Synthesized
  - Consists of a series of bitmap images
- Frame rate: the speech to playback the video
  - NTSC (US, Japan): 30 frames/s
  - PAL (EU): 25 frames/s
  - SECAM (France): 24 frames/s
  - HDTV: 50/60 frames/s
  - UHDTV: 120 frames/s
Video vs. Animation

- Both images and graphics are displayed as a succession of view, which creates an impression of movement.
- Video - moving images or moving pictures
  - Captured or Synthesized
  - Consists of a series of bitmap images
- Animation - moving graphics
  - Generated by computer program
  - Consists of a set of objects
  - The movement of the objects are calculated
Video Compression

- **International Telecommunication Union (ITU-T)**
  - H.261: ISDN Video Phone (px64 kb/s)
  - H. 263: PSTN Video Phone (<64 kb/s)
  - H.26L: A variety of applications (<64 kb/s)
    - Internet Video Application, VOD, Video Mail

- **International Organization for Standard (ISO)**
  - MPEG-1 Video: CD-ROM (1.2 Mb/s)
  - MPEG-2 Video: SDTV, HDTV (4-80 Mb/s)
  - MPEG-4 Video: A variety of applications (24-1024 kb/s)
Sound

- Sound (e.g., speech, music, noise)
  - Analog 1-D time-based signal
  - Vibration that propagates as a typically audible mechanical wave of pressure and displacement through a medium such as air or water.
- Sound wave: Sinusoidal planes wave, which has the following properties
  - Frequency
  - Wavelength
  - Amplitude
- Human ear: 20-20,000 Hz. The upper limit decreases with age
Digital Audio

- Audio signal that is encoded in digital form
  - Sampling
  - Quantization
- Sampling rate
  - Telephone: 8 kHz
  - CD-audio: 44.1 kHz
- Quantization
  - Speech: 8 bit
  - CD-audio: 16 bit
- Audio Compression
  - Lossless compression: FLAC, Apple Lossless, MPEG-4 ALS
  - Lossy compression: MP3 (50-60% of original size)
A Simple Data Communication Model

- Source generates the data to be transmitted (Examples?)
- Transmitter transforms and encodes the information to produce electromagnetic signals that can be transmitted across some sort of transmission system.
- Transmission system can be transmission line or a complex network connecting source and destination
- Receiver accepts the signals and convert them into a form that can be handled by the destination device.
- Destination takes the incoming data from the receiver.
## Communication Tasks

<table>
<thead>
<tr>
<th>Transmission system utilization</th>
<th>Flow control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interfacing</td>
<td>Addressing</td>
</tr>
<tr>
<td>Signal generation</td>
<td>Routing</td>
</tr>
<tr>
<td>Synchronization</td>
<td>Recovery</td>
</tr>
<tr>
<td>Exchange management</td>
<td>Security</td>
</tr>
<tr>
<td>Error detection and correction</td>
<td>Network management</td>
</tr>
</tbody>
</table>
Distributed Multimedia System

- Distributed Multimedia System involves transmission and distribution of multimedia information on the network.

- Multimedia Servers
  - Silicon Graphic
  - HP Media Server

- Multimedia Networks
  - PSTN/Data Network etc.

- Multimedia Clients
Distributed Multimedia System (Cont.)

• Live media transmission system
  ▫ Capture, compress, and transmit the media on the fly (example?)

• Send stored media across the network
  ▫ Media is pre-compressed and stored at the server. This system delivers the stored media to one or multiple receivers (example?)

• Differences between the two systems
  ▫ **Media capture:** Real-time media capture vs. pre-stored media.
  ▫ **Media compression:** Real-time/off-line compression?, compression can be adjusted during transmission?
Distributed Multimedia System (Cont.)

• Requirements of networked multimedia applications
  ▫ Delay requirements
  ▫ Quality requirements
    • Satisfactory quality of media presentation
    • Synchronization requirement
    • Continuous requirement
    • Can tolerate some degree of information loss

• Challenges of distributed multimedia system
  ▫ Conflict between media size and bandwidth limit
  ▫ Conflict between the user requirement of multimedia application and the best-effort network.
  ▫ How to meet different requirement of different users?
Multimedia Server

- Multimedia Servers
  - Data storage and retrieval
  - Media synchronization
  - Video/Audio Compression
- Requirements
  - High performance
  - High reliability

![Diagram of Multimedia Server System](image)
Multimedia Equipments

- Digital equipments:
  - Computers
  - A/D, D/A converters for Video and Audio.

- Analogue equipments:
  - Video camera,
  - Microphone, Speaker,
  - Tape player/recorder.

- Synchronization equipments:
  - Support time synchronization.

- Interactive equipments:
  - Display, mouse, keyboard etc.
Multimedia Networks

• Telephone Networks:
  ▫ Public switched telephone networks (PSTNs): initially designed to provide speech services, now can support multimedia applications due to the advances in Digital Signal Processing hardware and software.

• Data Networks: initially supported data applications (e-mail/ftp), now support much complex multimedia applications.

• Broadcast Television Networks: Broadcast TV

• Broadband Multiservice Networks: Multiservice.
Multimedia Operating System

- Multimedia operating system provides a comfortable environment for the execution of programs, and ensures effective utilization of the computer hardware.
Multimedia Operating System (Cont.)

• MMOS Requirements:
  ▫ **Soft real-time applications**: statistical guarantees
  ▫ **Interactive applications**: no absolute performance guarantees, but low average response times
  ▫ **Throughput-intensive application**: no performance guarantees, but high throughput.
  ▫ **Fair, Proportionate resource allocation**: Divide resources according to application requirements.
  ▫ **Application Isolation**: Preventing misbehaving or overloaded application from effecting others (e.g., overloaded web server should not affect streaming media server).
Multimedia Operating System (Cont.)

CPU Scheduler

Packet Scheduler

Disk Scheduler

Class-specific Schedulers

Class-Independent Scheduler

FCFS
Multimedia Applications

- **Multimedia Information Systems:** Multimedia Database, Information hypertexts, Hypermedia, Electronic books, Multimedia expert systems, etc.

- **Multimedia Communication Systems:** VOIP, Audio-video Communication, Computer-supported collaborative works, Videoconferencing, IPTV, Streaming media and Multimedia Teleservices, etc.

- **Multimedia Entertainment Systems:** Game, 3D computer games, Multimedia design, Multiplayer networks, Interactive audiovisual productions, etc.
Multimedia Applications (Cont.)

- **Multimedia Educational Systems**: E-Books, E-learning, Flexible teaching materials, Simulation education systems, etc.

- **Multimedia Business Systems**: Electronic commerce, Marketing, Multimedia presentation – PR, Virtual shopping, etc.
Multimedia Application Environment

[Diagram showing network connections and bandwidths: Home user (U1) connected to LAN at 1-10 Mb/s, Business user (U2) connected to Multimedia Server at 1-10Mb/s, Wide area network with Customer access network at 28.8-128 Kb/s, and LAN connected to Multimedia Server.]
Example (Video on Demand)

- Media Server
- Authentication/Billing Server
- Streaming Server
- Router
- ADSL
  - STB
- CATV
  - STB
- Modem
  - STB
Project Topics

1) Audio coding: PCM, DPCM, ADPCM
2) Audio coding: CELP, ACELP, Apple Lossless, MP3
3) Image coding: JPEG
4) Image coding: JPEG 2000
5) Video coding: ITU-T H.261, H.263, H26L
6) Video coding: ISO MPEG-1, MPEG-2
7) Video coding: ISO MPEG-4
8) Video coding: Scalable Video Coding (SVC)
Project Topics (Cont.)

1) Systems: ITU-T H.320
2) Systems: ITU-T H.323
3) Systems: Video on Demand
4) Systems: IPTV
5) Systems: Video Conferencing
6) Systems: VoIP
Evaluation Method

• Final written exam (70%)
  ▫ Held after the end of the lecturing period

• Project (30%)
  ▫ Written in group of 2 students
  ▫ Shall not have more than 10 pages (printed)
  ▫ Deadlines: 18/05/2015
  ▫ Includes presentation and discussion
DIGITIZATION PRINCIPLES

Dr. Quang Duc Tran
Analog Signals

• As mentioned earlier the amplitude of the signal varies continuously with time

• The **Fourier transform** decomposes a signal into the frequencies that make it up

• The Fourier analysis can be used to show that any time varying signal is made up of infinite number of single-frequency sinusoidal components
Analog Signals (Cont.)

- A band-limited signal \( x(t) \) is one whose Fourier Transform is non-zero on only a finite interval of the frequency axis.
- There exists a positive number \( B \) such that \( X(f) \) is non-zero only in \(-B \leq f \leq B\). \( B \) is called the signal bandwidth.
- Speech bandwidth: \( 50\text{Hz} – 10\text{kHz} \)
- Music Bandwidth: \( 15\text{Hz} – 20\text{kHz} \)
- To transmit an analogue signal through a network the bandwidth of the transmission channel should be equal to or greater than the signal bandwidth.
Encoder Design

- The encoder consists of AAF (Anti-Aliasing Filter) and an Analogue-to-Digital Converter (ADC)
Encoder Design (Cont.)

- Anti-Aliasing Filter (AAF) removes the selected higher frequency components from the source signal.

- Sample and Hold samples amplitude of the filtered signal at regular intervals and holds the sampled amplitudes between samples.

- Quantizer converts the samples into their corresponding binary form
Data Representation

• The most significant bit of the codeword represents the sign of the sample.

• A binary 0 indicates a *positive value* and a binary 1 indicates a *negative value*.

• The signal must be sampled at a much higher rate than the maximum rate of change of the signal amplitude.

• The number of quantization levels should be as large as possible to represent the signal accurately.
Sampling Rate

• Nyquist theorem: If \( x(t) \) is a band-limited signal containing no frequencies higher than \( B \). A sufficient sampling rate (i.e., \( f_s \)) is \( 2B \) or anything larger. \( 2B \) is called the Nyquist rate.

• Nyquist rate is represented either in Hz or more correctly in \textit{samples per seconds}
Sampling Rate (Cont.)

- For a band-limited signal \((X(f) = 0 \text{ for all } |f| \geq B)\), and \(f_s \geq 2B\), it is possible for the copies to remain distinct from each other. But when the Nyquist criterion is not satisfied, adjacent copies overlap. Any frequency component above \(f_s/2\) is indistinguishable from a lower-frequency component, called an alias.
Quantization

- **Quantization** is the process of mapping an analogue value to a number of digits
- Three bits are used to represent each sample (1 bit for the sign and two bits to represent the magnitude)

If $V_{\text{max}}$ is the maximum positive and negative signal amplitude and $n$ is the number of binary bits used then the *quantization interval*, $q$, is defined as

$$q = V_{\text{max}}/2^n$$
Quantization Error

- Quantization error is the difference between the *actual signal amplitude* and the corresponding *nominal amplitude*. Consequently, a difference of $\pm q/2$ from the actual signal level is present. Quantization error is also known as *quantization noise* since values vary randomly.

$$\text{SQNR} = 6.02 \times n \text{ (dB)}$$

$$\text{SQNR} = 1.761 + 6.02 \times n \text{ (dB)}$$
A signal decoder is an electronic circuit that performs the conversion prior to their output back again into their analogue form through a digital-to-analogue converter and a low pass filter.
Zero-Order Hold describes the effect of converting a discrete-time signal to a continuous-time signal by holding each sample value for one sample interval.
Dr. Quang Duc Tran
Various Types of Text

- *Unformatted text* (i.e., plain text) enables pages to be created, which consist of strings of fixed-sized characters from a limited character set.
- *Formatted Text* (i.e., rich text (RTF)) enables pages to be created, which consist of strings of characters of different styles, sizes and shape with tables, graphics, and images inserted at appropriate points.
- *Hypertext* enables an integrated set of documents (Each consisting of a formatted text) to be created, which have defined linkages between them.
**ASCII Character Coding**

33 control characters
Back space, Delete, Escape

95 printable characters
Alphabetic, Numeric, Punctuation

A – 1000001 (65)

- The American Standard Code for Information Interchange is one of the most widely used character sets. Each character is represented by a 7-bit codeword.
ISO/IEC 8859

- ISO/IEC 8859 is a standard for 8 bit character encodings. It allows positions for another 96 printable characters (Latin alphabets)
- ISO/IEC 8859 is divided into the following parts:
  - Part 1: Latin-1 Western European
  - Part 2: Latin-2 Central European
  - ...
  - Part 16: Latin-10 South-Eastern European
- Although Vietnamese uses Latin based characters, it doest not fit into 96 positions.
Two mapping methods

- **Unicode Transformation Format (UTF)**
  - UTF-8: 8 bit, variable-width length which maximizes compatibility with ASCII
  - UTF-16: 16 bit, variable-width length
  - UTF-32: 32 bit, fixed-width length

- **Universal Character Set (UCS)**
  - UCS-2 is a subset of UTF-16
  - UCS-4 and UTF-32 are functionally equivalent

**UTF-8 and UTF-16 are the most commonly used encodings** (XML, HTML, Microsoft also recommends applications use UTF-8 or UCS-2/UTF-16)
Text Compression

- Lossless compression
  - Statistical compression (e.g., Huffman coding)
  - Compression using dictionary (e.g., Lempel-Ziv)

- These are intended for compressing natural language text and other data with a similar sequential structure.

- These are used by general purpose compressors such as zip, bzip2, 7zip, etc.

- Compression rate: approximately $\frac{1}{2}-\frac{2}{3}$ document size
Huffman Coding

- **Huffman coding**: A statistical compression technique, which needs a probability distribution as an input. The method for determining the probabilities is called a *model*, which can be *static*, *semi-adaptive* or *adaptive*.

- A **static model** is a fixed model that is known by both the compressor and decompressor.

- A **semi-adaptive model** is a fixed model that is constructed from the data to be compressed.

- An **adaptive model** changes during the compression.
The Basic Algorithm

• Not all characters occur with the same frequency.

• Not all characters are allocated the same amount of space.

• Codeword lengths are no longer fixed like ASCII.

• Codeword lengths vary and will be shorter for the more frequently used characters.
The Basic Algorithm (Cont.)

1) Scan text to be compressed and tally occurrence of all characters.

2) Sort or prioritize characters based on number of occurrences in text.

3) Build Huffman code tree based on prioritized list.

4) Perform a traversal of tree to determine all codewords.

5) Scan text again and create new file using the Huffman codes.
Examples

- Consider the following text:
  BCAACADBDADDEABACDBACDADBADBEABEAA
- A: 15; B: 7; C: 6; D: 6; E: 5

How about

Text Compression for Web Developers?
Lempel-Ziv Coding

- Lempel-Ziv Coding relies on reoccurring patterns to save data space.

- **Example**: Extended ASCII code – Every character is stored with 8 binary bit, allowing up to 256 unique characters for the data.

- Lempel-Ziv tries to extend the library to 9 to 12 bits per character. The new symbols are made up of combination of symbols that occurred previously in the string. Lempel-Ziv do not compress well with short, diverse strings.
Examples

- Consider the following text: ABCBCABCABCD

<table>
<thead>
<tr>
<th>Previous Input</th>
<th>Input</th>
<th>Output</th>
<th>Symbol</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIL</td>
<td>A</td>
<td>A</td>
<td>AB</td>
<td>256</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>AB</td>
<td>256</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>B</td>
<td>BC</td>
<td>257</td>
</tr>
<tr>
<td>C</td>
<td>B</td>
<td>C</td>
<td>CB</td>
<td>258</td>
</tr>
<tr>
<td>BC</td>
<td>A</td>
<td>BC</td>
<td>BCA</td>
<td>259</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>C</td>
<td>AB</td>
<td>ABC</td>
<td>260</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
<td>C</td>
<td>CA</td>
<td>261</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC</td>
<td>D</td>
<td>ABC</td>
<td>ABCD</td>
<td>262</td>
</tr>
<tr>
<td>D</td>
<td>EOL</td>
<td>D</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Previous Input</th>
<th>Input</th>
<th>Output</th>
<th>Symbol</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIL</td>
<td>A</td>
<td>A</td>
<td>AB</td>
<td>256</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>AB</td>
<td>256</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>C</td>
<td>BC</td>
<td>257</td>
</tr>
<tr>
<td>C</td>
<td>257</td>
<td>BC</td>
<td>CB</td>
<td>258</td>
</tr>
<tr>
<td>BC</td>
<td>256</td>
<td>AB</td>
<td>BCA</td>
<td>259</td>
</tr>
<tr>
<td>AB</td>
<td>C</td>
<td>C</td>
<td>ABC</td>
<td>260</td>
</tr>
<tr>
<td>C</td>
<td>260</td>
<td>ABC</td>
<td>CA</td>
<td>261</td>
</tr>
<tr>
<td>ABC</td>
<td>D</td>
<td>D</td>
<td>ABCD</td>
<td>262</td>
</tr>
</tbody>
</table>
Image

- All types of images are displayed in the form of a two-dimensional matrix of individual picture elements (pixels).
Color Models

- RGB (Red, Green, and Blue) – Computer Monitors. RGBA is RGB with an additional channel, alpha, to indicate transparency.
- CMYK (Cyan, Magenta, Yellow, and Black) – Photo printing.
- YCbCr – Video capture system. YUV stores a luminance value along with two chroma values.
- HSV – Artist. It is often more natural to think about color in terms of hue and saturation than additive and subtractive color components
Color Models (Cont.)

HSV

RGB

YCbCr (Y=0.5)

CMYK
Color Models (Cont.)

- YCbCr is not an absolute color space, it is a way of encoding RGB information. YCbCr contains perceptually meaningful information.

\[
Y = 16+219(0.299R+0.587G+0.114B)/255 \\
Cb = 128+224(-0.169R-0.331G+0.5B)/255 \\
Cr = 128+224(0.5R-0.419G-0.081B)/255
\]
Image Compression

• **Lossless Compression**
  ▫ Run-Length coding (RLC)
  ▫ Huffman coding
  ▫ Lempel-Ziv coding

• **Lossy compression**
  ▫ Transform coding
  ▫ **Chroma sub-sampling**
  ▫ **JPEG**
Chroma Sub-sampling

- 4:4:4
- 4:2:2
- 4:1:1
- 4:2:0

- Pixel with only Y value
- Pixel with only Cr and Cb values
- Pixel with Y, Cr and Cb values
Chroma Sub-sampling (Cont.)

- **4:4:4 (1:1)**
  - High end film scanners
  - Cinematic postproduction.
- **4:2:2 (3:2)**
  - High end digital video formats and interfaces
- **4:1:1 (2:1)**
  - DVCPRO (e.g., NTSC, PAL)
- **4:2:0 (2:1)**
  - MPEG, H.26X video coding
  - DVD, Blue-ray Disc
  - JPEG, MJPEG
JPEG Standard

• JPEG is an image compression standard that was developed by the “Joint Photographic Experts Group”.

• JPEG is the most common format for storing and transmitting photographic images.

• JPEG works with both color and grayscale images

• The compression rate is typically 10:1 with little perceptible loss in image quality
JPEG Standard (Cont.)

• JPEG is a **lossy** compression technique. It employs 2D-DCT (Discrete Cosine Transform) transform coding, whose effectiveness replies on 3 major observations:

  ▫ **Observation 1**: Useful image contents change relatively slowly across the image.

  ▫ **Observation 2**: Humans are much less likely to notice the loss of very high spatial frequency components than the loss of low frequency components.

  ▫ **Observation 3**: Visual acuity (accuracy in distinguishing closely spaced lines) is much greater for gray than for color.
JPEG Coding

Source Image \(8 \times 8\) blocks

FDCT \(\rightarrow\) Quantization \(\rightarrow\) Entropy Encoding

Quant. Table \(\uparrow\) Huffman Table

Compressed Image

IDCT \(\leftarrow\) Dequantization \(\leftarrow\) Entropy Decoding

Quant. Table \(\uparrow\) Huffman Table

Source Image \(8 \times 8\) blocks
Discrete Cosine Transform

• DCT transforms a signal of image from the spatial domain to the frequency domain.

• DCT is similar to the Fast Fourier Transform (FFT) but can approximate lines well with fewer coefficients.

• For most images, much of the signal energy lies at low frequencies, which appear in the upper left corner of the DCT.

• Compression is achieved since the lower right values represent higher frequencies, and are often small – small enough to be neglected with little visible distortion.
Both DC (i.e., $F(0,0)$) and AC (i.e., $F(u,v)$ $(u,v \neq 0)$) coefficients contain integers, these can range from -1024 to 1023.
Quantization

• Quantization discards information, which is not visually significant. It is fundamentally lossy and is the principal source of lossiness in DCT-based encoders.

\[ F^Q(u,v) = \text{Integer Round} \left( \frac{F(u,v)}{Q(u,v)} \right) \]

• Quantization step size should be chosen as a perceptual threshold, which is a function of the source image characteristic, display characteristic and viewing distance.

• For a particular application, the best thresholds should be determined by performing psycho-visual experiments.
• It is typically the case that many of the higher frequency components are rounded to zero, and many of the rest become small positive or negative numbers, which take many fewer bits to represent.
Entropy encoding is a special form of lossless data compression.

Entropy encoding involves arranging the image components in a zigzag order employing Run Length Coding (RLC) algorithm, inserting length coding zeros, and then using Huffman coding on what is left.
Zigzag Order

<table>
<thead>
<tr>
<th>-26</th>
<th>-3</th>
<th>-6</th>
<th>2</th>
<th>2</th>
<th>-1</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-2</td>
<td>-4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-3</td>
<td>1</td>
<td>5</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-4</td>
<td>1</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- All of the quantized coefficients are ordered into the zigzag sequence. It places low frequency coefficients (which are more likely to be non-zero) before high-frequency coefficients.
DPCM & RLC

- There are strong correlation between DC coefficients of adjacent blocks. Hence, the quantized DC is encoded as the difference from the DC term of the previous block in the encoding order by using DPCM.

\[ \text{DIFF} = \text{DC}_i - \text{DC}_{i-1} \]

- RLC is applied to encode other AC coefficients
DPCM & RLC (Cont.)

- DC coefficient: \(+3 \rightarrow (2),(3)\)
- AC coefficients:
  - \( (0,2) (-3) \)
  - \( (1,2) (-3) \)
  - \( (0,2) (-2) \)
  - \( (0,3) (-6) \)
  - \( (0,2) (2) \)
  - \(...\)
  - \( (0,0) \rightarrow \text{EOB} \)
- AC coefficient is made up of a pair of symbols \((\text{run length, size}) (\text{amplitude})\). Run length is the number of zeros in the run. Size is the number of bits used to encode amplitude.
Huffman Coding

• Significant levels of compression can be obtained by replacing long string of a binary digit by a string of much shorter codeword.

• The length of each codeword is a function of its relative frequency of occurrence.

• JPEG provides general purpose Huffman tables, encoders may also choose to generate Huffman tables optimized for the actual frequency distributions in images being encoded.
## Huffman Coding (Cont.)

<table>
<thead>
<tr>
<th>Category</th>
<th>Values</th>
<th>Bits for the value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1,1</td>
<td>0,1</td>
</tr>
<tr>
<td>2</td>
<td>-3,-2,2,3</td>
<td>00,01,10,11</td>
</tr>
<tr>
<td>3</td>
<td>-7,-6,-5,-4,4,5,6,7</td>
<td>000,001,010,011,100,101,110,111</td>
</tr>
<tr>
<td>4</td>
<td>-15,...,-8,8,...,15</td>
<td>0000,...,0111,1000,...,1111</td>
</tr>
<tr>
<td>5</td>
<td>-31,...,-16,16,...,31</td>
<td>00000,...,01111,10000,...,11111</td>
</tr>
<tr>
<td>6</td>
<td>-63,...,-32,32,...63</td>
<td>000000,...,011111,100000,...,111111</td>
</tr>
<tr>
<td>7</td>
<td>-127,...,-64,64,...,127</td>
<td>0000000,...,0111111,1000000,...,1111111</td>
</tr>
<tr>
<td>8</td>
<td>-255,...,-128,128,...,255</td>
<td>...</td>
</tr>
<tr>
<td>9</td>
<td>-511,...,-256,256,...,511</td>
<td>...</td>
</tr>
<tr>
<td>10</td>
<td>-1023,...,-512,512,...,1023</td>
<td>...</td>
</tr>
<tr>
<td>11</td>
<td>-2047,...,-1024,1024,...,2047</td>
<td>...</td>
</tr>
</tbody>
</table>

### Values and bits for values
### Huffman Coding (Cont.)

<table>
<thead>
<tr>
<th>Run, category</th>
<th>Code Length</th>
<th>Codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,0</td>
<td>4</td>
<td>1010</td>
</tr>
<tr>
<td>0,1</td>
<td>2</td>
<td>00</td>
</tr>
<tr>
<td>0,2</td>
<td>2</td>
<td>01</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0,10</td>
<td>16</td>
<td>1111111110000011</td>
</tr>
<tr>
<td>1,1</td>
<td>4</td>
<td>1100</td>
</tr>
<tr>
<td>1,2</td>
<td>5</td>
<td>11011</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>15,10</td>
<td>16</td>
<td>1111111111111110</td>
</tr>
</tbody>
</table>

Luminance AC coefficients
Huffman Coding (Cont.)

<table>
<thead>
<tr>
<th>Category</th>
<th>Code Length</th>
<th>Codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>00</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>010</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>011</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>101</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1110</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>11110</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>111110</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>1111110</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>11111110</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>11111110</td>
</tr>
</tbody>
</table>

Luminance DC coefficients
Huffman Coding (Cont.)

- DC coefficient: $+3 \rightarrow (2),(3)$
- Encoded DC coefficient: 011 11
- AC coefficients: $(0,2) (-3), (1,2) (-3) \ldots$ EOB
- Encoded AC coefficient: 01 00 11011 00 ... 00
Example

1. $Q = 100$ - 83,2 bytes
2. $Q = 50$ - 15,1 bytes
3. $Q = 25$ - 9,5 bytes
4. $Q = 10$ - 4,7 bytes
5. $Q = 5$ - 1,5 bytes
Why another standard?

- At low bit-rates (e.g., 0.25 bpp), the distortion in JPEG becomes unacceptable.

- JPEG has 44 modes, many of them are not used by the majority of the JPEG encoders.

- JPEG quality suffers dramatically, when bit errors are encountered during transmission.

- JPEG is optimized for natural images and performs badly on graphics.

- JPEG fails to compress bi-level (text) imagery.
JPEG2000 Standard

• **Key features**
  ▫ Wavelet transform
  ▫ Superior low bit-rate performance
  ▫ Bi-level compression
  ▫ Lossy and lossless compression
  ▫ Robustness to bit errors
  ▫ Scalability
  ▫ Region of Interest

• **Sample Applications**
  ▫ Web browsing and printing
  ▫ Digital camera
  ▫ Medical Imagery
  ▫ Satellite Imagery
JPEG2000 Standard (Cont.)

Original image

Lossless

Lossy

Lower resolution

ROI

1.89 bpp

5.2 bpp
JPEG2000 Standard (Cont.)

0.125 bpp

JPEG vs. JPEG2000

0.25 bpp
JPEG2000 Standard (Cont.)

- Disadvantages
  - Encoder and Decoder are complex and computationally demanding
  - When the compression ratio is below 25:1, the wavelet-based algorithm produces less blocky but somewhat less detailed images, compared to JPEG.
AUDIO

Dr. Quang Duc Tran
Sound Facts

• Sound is a continuous wave that travels through the air. The wave is made up of pressure differences

• Sound waves have normal wave properties (reflection, refraction, diffraction etc.)

• Frequency represents the number of periods in a second and is measured in hertz (Hz) or cycles per second.

• Amplitude is the measure of displacement of the air pressure wave from its means. It is related to but not the same as loudness.
Digital Audio

- Audio signal that is encoded in digital form
  - Sampling
  - Quantization
- Sampling rate
  - Telephone: 8 kHz
  - CD-audio: 44.1 kHz
- Quantization
  - Speech: 8 bit
  - CD-audio: 16 bit
- Number of sound tracks
  - Stereo: 2 channels
  - Professional: 16, 32 or more.
Digital Audio (Cont.)

- **Example 1:** Sampling a four second sound of a speaking voice
  - Because the voice is in the lower range of the audio spectrum, we could sample the sound at 8 kHz with a mono channel and 8 bit resolution

- **Example 2:** Sampling a four second sound of a musical selection
  - In order to capture the complex musical sound, the sampling rate should be 44.1 kHz. We will record a stereo sample with 16 bit resolution.
Audio Compression

• Differential Pulse Code Modulation (DPCM)

• Adaptive Differential Pulse Code Modulation (ADPCM)

• Linear Predictive Coding (LPC)

• Perceptual Coding

• MPEG Audio Compression
DPCM

- The range of differences in amplitudes between successive samples of the audio waveform is less than the range of the actual sample amplitudes. Hence, fewer bits to represent the difference signals.

- Designing of DPCM is easy. Prediction is then based on the previously coded/transmitted samples.

- Because the co-relation between successive are very good, so the output sample is far better than that of PCM.
DPCM (Cont.)

Speech Input Signal

AAF → ADC → Subtractor → Parallel-to-serial converter

Timing + Control

Register R → Adder

Network

Speech Output Signal

AAF → ADC → Register R → Adder → Serial-to-parallel converter

Timing + Control
**DPCM (Cont.)**

- **Encoder**
  - Previous digitized sample is held in the register (R).
  - The DPCM signal is computed by subtracting the current content (R) from the new output by the ADC.
  - The register value is then updated before transmission.

- **Decoder**
  - Decoder simply adds the previous register contents (PCM) with the DPCM.
  - Since ADC will have noise there will be cumulative errors in the value of the register signal.
Third-order predictive DPCM

Speech Input Signal

AAF → ADC → Subtractor → Parallel-to-serial converter

Timing + Control

C1 → X → C2 → X → C3 → X

R1 → X → R2 → X → R3 → X

Network

Speech Output Signal

AAF → ADC → Adder → Serial-to-parallel converter

Timing + Control

C1 → X → C2 → X → C3 → X

R1 → X → R2 → X → R3 → X
Third-order predictive DPCM (Cont.)

• To eliminate the noise effect, predictive techniques are used to predict a more accurate version of the previous signal (i.e., using not only the current signal but also varying proportions of a number of the preceding estimated signals).

• These proportions used are known as predictor coefficients.

• Different signal is computed by subtracting varying proportion of the last three predicted values from the current output by the ADC.
Third-order predictive DPCM (Cont.)

- R1, R2, R3 will be subtracted from PCM.
- The values in the R1 register will be transferred to R2 and R2 to R3 and the new predicted value goes into R1.
- Decoder operates in a similar way by adding the same proportions of the last three computed PCM signals to the received DPCM signal.
Adaptive DPCM (ADPCM)

- Saving of bandwidth is possible by varying the number of bits used for the difference signal depending on its amplitude.

- An international standard for this is defined in ITU-T Recommendation G.721.

- The principle is similar to that of DPCM except an eight-order predictor is used and the number of bits used to quantize each difference is varied.

- Larger step-size is used to encode differences between high frequency samples and smaller step-size for differences between low frequency samples.
Sub-band ADPCM

Speech Input Signal
- Lower sub-band bandlimiting filter (50Hz-3.5 kHz)
- Upper sub-band bandlimiting filter (3.5kHz-7kHz)

48 kbps
Multiplexer

Network

Lower sub-band low-pass filter (50Hz-3.5 kHz)
Upper sub-band low-pass filter (3.5kHz-7kHz)

48 kbps
Demultiplexer

Speech Output Signal
- Lower sub-band ADPCM encoder
- Upper sub-band ADPCM encoder
- Lower sub-band ADPCM decoder
- Upper sub-band ADPCM decoder

16 kbps
Sub-band ADPCM

- The sub-band ADPCM is defined in ITU-T Recommendation G.722 (better sound quality in VoIP).

- This uses sub-band coding, in which the input signal prior to sampling is passed through two filters: one which passes only signal frequencies in the range 50 Hz through to 3.5 kHz and the other only frequencies in the range 3.5 kHz through to 7 kHz.

- Each is then sampled and encoded independently using ADPCM. 6 digits are used to encode the lower sub-band signal, while 2 digits are used to encode the upper sub-band signal.
Linear Predictive Coding

- Linear Predictive Coding derives its name from the fact that current speech sample can be closely approximated as a linear combination of past samples. The approximation is done on short chunks of the speech signal, called frames. Generally 30 to 50 frames per second give intelligible speech with good compression.

\[ x[n] = \sum_{k=1}^{p} a_k x[n-k] + e[n] \]

- Order of the model
- Prediction coefficient
- Previous speech samples
- Prediction error
Linear Predictive Coding (Cont.)

• The predictor coefficients are determined by minimizing the sum of squared differences (over a finite interval) between the actual speech samples and the linearly predicted ones.

• The higher the coefficient order $p$, the closer the approximation is.

• Uncompressed speech is transmitted at 64 kbps. The LPC transmits speech at a bit rate of 2.4 kbps. There is a noticeable loss of quality, however, the speech is still audible.

• LPC is used by phone companies (e.g., GSM standard).
Code-excited LPC (CELPC)

- In CELPC model instead of treating each digitized frame independently for encoding purposes, just a limited set of frames are used, each known as a wave template.

- A pre-computed set of templates are held by the encoder and the decoder in what is known as the template codebook.

- Each of the individual digitized samples that make up a particular template in the codebook are differently encoded.
Perceptual Coding

• LPC and CELPC are used for telephony applications. PC is designed for compression of general audio associated with a digital television broadcast.

• Sampled frames of the source audio waveform are analyzed – but only those features that are perceptible to the ear are transmitted.

• Although the human ear is sensitive to signals in the range 15 Hz to 20 kHz, the level of sensitivity to each signal is non-linear, i.e., the ear is more sensitive to some signals than others.
Sensitivity of the ear

- Sensitivity of the ear varies with the frequency of the signal. The ear is most sensitive to signals in the range 2-5 kHz.
- The dynamic range of ear is defined as the loudest sound it can hear to the quietest sound.

Signal A is above the hearing threshold and B is below the hearing threshold.
Frequency Masking

Signal B is larger than signal A. This causes the basic sensitivity curve to be distorted.

Signal A will no longer be heard as it is within the distortion band.

- When multiple signals are present, a strong signal may reduce the level of sensitivity of the ear to other signals which are near to it in frequency.
Frequency Masking (Cont.)

If the magnitude of the frequency components that make up an audio sound can be determined, it is possible to determine those frequencies that will be masked and do not need to be transmitted.

The width of each curve at a particular signal level is known as the critical bandwidth for that frequency.
Temporal Masking

After the ear hears a loud sound, it takes a further short time before it can hear a quieter sound. This is known as the *temporal masking*.

- During this time, signals whose amplitudes are less than the decay envelope will not be heard and hence need not be transmitted.
MPEG Audio Coder

MPEG Audio Coder

Speech Input Signal

PCM Encoder

Analysis filter bank (DFT)

Masking Threshold

Signal-to-mask ratios + Bit allocations

Format frame for transmission

Q1

Q2

Q32
The bandwidth available for transmission is divided into 32 frequency sub-bands using a bank of analysis filters.

Processing associated with both frequency and temporal masking is carried out by the psychoacoustic model. A set of signal-to-mask ratios (SMR) is determined. This indicates the frequency components whose amplitude is below the audible components.

More bits are assigned for highest sensitivity regions compared with less sensitivity regions.
Analog Video

• A video signal is a sequence of two dimensional (2D) images projected from a dynamic three dimensional (3D) scene onto the image plane of a video camera.

• A video records the emitted and/or reflected light intensity from the objects. The intensity changes both in time and space.
Composite vs. Component Video

• Ideally, a color video is specified by three function or signals, each describing one color component. A video in this format is known as component video.

• In composite video, the three color signals are multiplexed into a single signal. A composite signal has bandwidth that is significantly lower than the sum of the bandwidth of three component signals, and hence, can be stored and transmitted efficiently. This is achieved at the expense of video quality.

• S-video consists of two components, the luminance component and a single chrominance component.
Analog Video Raster

- The analog TV systems of today use raster scan for video capture and display, which can be interlaced or progressive.

- The progressive scanning scans an image sequentially from line 1 to the final line of the raster to create a video frame.

- The interlaced scanning scans odd lines in field 1 and even lines in field 2. Together field 1 and 2 constitute one frame.
Refresh rate

- The frequency between the display of two still images is known as the refresh rate and is expressed in Hz or frames per second.

- The movie industry uses a refresh rate of 24 frames per second.

- For a fixed bandwidth, the interlaced scan provides a video signal with twice the display refresh rate for a given line count as compared to the progressive scan. This higher refresh rate improves the appearance of objects motion.

- However, interlaced scan causes interlacing effects if recorded objects move fast to be in different positions when each individual field is captured.
Analog Television System

<table>
<thead>
<tr>
<th>Parameters</th>
<th>NTSC</th>
<th>PAL</th>
<th>SECAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field Rate</td>
<td>59.94</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Line No./Frame</td>
<td>525</td>
<td>625</td>
<td>625</td>
</tr>
<tr>
<td>Line Rate (Line/s)</td>
<td>15,750</td>
<td>15,625</td>
<td>15,625</td>
</tr>
<tr>
<td>Luminance Bandwidth (MHz)</td>
<td>4.2</td>
<td>5.0, 5.5</td>
<td>6.0</td>
</tr>
<tr>
<td>Chrominance Bandwidth (MHz)</td>
<td>1.5 (I), 0.5 (Q)</td>
<td>1.3 (U,V)</td>
<td>1.0 (U,V)</td>
</tr>
<tr>
<td>Audio Subcarrier</td>
<td>4.5</td>
<td>5.5, 6.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Composite Signal Bandwidth (MHz)</td>
<td>6.0</td>
<td>8.0, 8.5</td>
<td>8.0</td>
</tr>
</tbody>
</table>

- HDTV enhances the visual impact by employing a wider screen (16:9) and sampling resolution of 60 frames/s, and 720 line/frame.
- For computer display, much higher temporal and spatial sampling rates are needed (e.g., SVGA has 72 fps and a resolution of 1024x720 pixels).
Digital Video

- A digital video can be obtained either by sampling a raster scan, or directly using digital video camera.

- In the BT. 601 standard, a sampling rate of 13.5 MHz is used for both the NTSC and PAL/SECAM systems.

- BT.601 also defines a digital color coordinate, known as YCbCr (see Chroma Sub-sampling).

- In addition to BT.601, other standards exists. For example, CIF (Common Intermediate Format) has about half the resolution of BT.601.
Video Compression

- **International Telecommunication Union (ITU-T)**
  - H.261: ISDN Video Phone (px64 kb/s)
  - H.263: PSTN Video Phone (<64 kb/s)
  - H.26L: A variety of applications (<64 kb/s)
    - Internet Video Application, VOD, Video Mail

- **International Organization for Standard (ISO)**
  - MPEG-1 Video: CD-ROM (1.2 Mb/s)
  - MPEG-2 Video: SDTV, HDTV (4-80 Mb/s)
  - MPEG-4 Video: A variety of applications (24-1024 kb/s)

- **MJEG (Moving JPEG)** applies JPEG algorithm to each frame independently.
Video Compression (Cont.)

- A video stream has 2 spatial and 1 temporal dimensions, video compression is usually done independently in the spatial dimensions and, respectively, in the time dimension.

- In the spatial dimensions – encoder tries to eliminate spatial redundancy (like in JPEG) and typically works on 8x8 pixel image blocks.

- In the time dimension – encoder tries to eliminate temporal redundancy (i.e. motion of objects) and typically works on 16x16 pixel image blocks.
Motion Estimation

- The references between the different types of frames are realized by a process called *motion estimation*. The correlation between two frames in terms of motion is represented by a *motion vector*. 

![Diagram](image-url)
Motion Estimation (Cont.)

- Good estimation of the motion vector results in higher compression ratios and better quality of the coded video sequence.

- The actual frame is divided into non-overlapping blocks (macro blocks) usually 16x16 pixels. The smaller the block sizes are chosen, the more motion vectors need to be calculated.

- Motion vectors are only calculated if the difference between two blocks at the same position is higher than a threshold.
Motion Estimation (Cont.)

- *Block Matching* tries to “stitch together” an actual predicted frame by using blocks from previous frames.

- Each blocks of the current frame is compared with a past frame within a search area.

- Rectangular search area is used, which takes into account that horizontal movements are more likely than vertical ones.

- Only luminance information is used to compare the blocks, but color information will be included in the encoding.
Motion Compensation

- Video motions are often complex. A simple “shifting in 2D” is not a perfectly suitable description of the motion in the actual scene, causing so called prediction errors.

- Since the estimation is not exact, additional information must also be sent to indicate any small differences between the predicted and actual positions of the moving segments involved. This is known as the motion compensation.

- Generally, less data is needed to store the differences.
Schematic Process of Motion Estimation

1. Current Frame
2. Motion Estimation
3. Motion Compensation
4. Reference Frames
5. Frame N-1
6. Frame N
7. Motion Vector
8. DCT
9. Quantization
10. Entropy Coding
11. Bit stream
MPEG Standards

- **MPEG-1**: Initial Audio/Video Compression Standard
  - Total bit rate: 1.5 Mbps
  - Video: 352x240 pixels/frame, 30 frames/s
  - Audio: 2 channels, 48,000 samples/s, 16 bits/sample

- **MPEG-2**: for better quality audio and video
  - Total bit-rate: 4-80 Mbps
  - Video: 720x480 pixels/frame, 30 frames/s
  - Audio: 5.1 channels, Advanced Audio Coding (AAC)

- **MPEG-4**: for a variety of applications with a wide range of quality and bit rate
A Typical MPEG Frame Display Order

**I-frame**: Intra-coded frame
**P-frame**: One directional motion prediction from a previous frame
**B-frame**: Bi-directional motion prediction from a previous or future frame
I-frames

• I-frames are encoded without reference to any other frames. Each frame is treated as a separate picture and the Y, C_b and C_r matrices are encoded separately using JPEG.

• I-frames must be repeated at regular intervals to avoid losing the whole picture as during transmission it can get corrupted and hence loses the frame.

• The number of frames/pictures between successive I-frames is known as a group of pictures (GOP). Typical values of GOP are 3 - 12.
P-frames

- The encoding of the P-frame is relative to the contents of either a preceding I-frame or a preceding P-frame.

- P-frames are encoded using a combination of motion estimation and motion compensation.
  - If the two contents are the same, only the address of the macro block in the reference frame is encoded.
  - If the two contents are very close, both the motion vector and the difference matrices associated with the macro block in the reference frame are encoded.
  - If no close match is found, then the target macro block is encoded in the same way as a macro block in an I-frame.

- Number of P frames between I-frames is limited to avoid error propagation.
B-frames

- Motion estimation works well in slow moving applications like video telephony.

- For fast moving video it will not work effectively. Hence **B-frames (Bi-directional)** are used. Their contents are predicted using the *past* and the *future* frames.

- B-frames provides highest level of compression and because they are not involved in the coding of other frames they do not propagate errors.
MPEG-1 Encoder

Input Video

DCT

Q

Q⁻¹

IDCT

Reference Frames

Motion Compensation

Motion Estimation

Coding Control

Entropy Coding

Buffer

Coded Bitstream

+ +

Q: Quantization

Q⁻¹: Inverse Quantization

Motion Vector
MPEG-1 Decoder

Coded Bit Stream

Buffer → Entropy Decoding → Q^-1 → IDCT

+ → Motion Compensation

Motion Vector

Reference Frames

Ordering

Output Video
**MPEG-1 Performance**

- Compression for I-frames are similar to JPEG for Video typically 10:1 through to 20:1 depending on the complexity of the frame contents.

- P and B frames are higher compression and in the region of 20:1 through to 30:1 for P frame and 30:1 to 50:1 for B-frames.
MPEG-1 Bit stream structure
MPEG-2 Standard

- MPEG-2 is a higher-quality video coding standard at a bit rate of more than 4 Mbps for video on demand (VOD), standard definition (SD) and high-definition (HD) digital TV broadcasting and for storing video on digital storage media like the DVD

- MPEG-2 supports interlaced video and alternative scan order.

- MPEG-2 have scalable coding and include error resilience techniques
MPEG-2 Interlaced Scan

Progressive Scan

Interlaced Scan
MPEG-2 Alternative Scan Order

- Alternative scan order may improve the effectiveness of DCT on prediction errors. This is due to the fact that in interlaced video, the consecutive rows in the blocks are from different fields. Hence there is less correlation between them than between the alternative rows.
MPEG-2 Scalable Coding

- Scalable coding divides the audio-video stream into a base layer and some enhancement layers.

- When the base layer is decoded basic quality is achieved, but if the transmission channel allows it, decoding enhancement layers brings additional quality to the decoded stream.

- The types of scalability can be SNR scalability, spatial scalability, temporal scalability or hybrid (combination of the above).
MPEG-4 Standard

• MPEG-4 is used for interactive multimedia applications over the Internet and over various entertainment networks.

• MPEG standard contains features to enables the manipulation of the individual elements that make up a scene within a video.

• In MPEG-4 each video frame is segmented into a number of video object planes (VOP) each of which will correspond to an AVO (Audio visual object) of interest.
MPEG-4 Standard (Cont.)

- Each audio and video object has a separate **object descriptor** associated with it, which allows the object to be manipulated by the viewer prior to it being decoded and played out.

- Before being compressed each scene is defined in the form of a background and one or more foreground audio-visual objects (AVOs).

- The audio associated with an AVO is compressed depending on the available bit rate of the transmission channel and the sound quality required.
MPEG-4 Standard (Cont.)

![Diagram of MPEG-4 Standard]

The diagram illustrates the process of encoding VOPs (Visual Object Parts) and audio input into a transmission multiplexer. The steps include:

1. Video Input:
   - VOP Identification + Definition
   - VOP 0 encoding
   - VOP 1 encoding
   - VOP 2 encoding

2. Audio Input:
   - Audio encoding

The output from these processes is multiplexed for transmission.
MULTIMEDIA NETWORKING SYSTEM

Dr. Quang Duc Tran
User Requirements

- Fast preparation and presentation of the different multimedia types of interest, taking into account the capabilities of available terminals and services.

- Dynamic control of multimedia applications with respect to connection interactions and quality on demand, combined with user-friendly interfaces.

- Intelligent support of users, taking into consideration their individual capabilities.
Network Requirements

• High speed and changing bandwidth
  ▫ Multimedia applications, particularly those using video and images demand large bandwidth. However, bandwidth for the foreseeable future will be limited. The limitation arise from the cost of installing optical fiber transmission, terminal equipment complexity and speed, etc.

• Quality of Services
  ▫ The availability of multimedia sources places demands on the service that a network must provide. The most importance of these are the bit error rate, the packet or cell loss, delay and delay variation.
Network Requirements (Cont.)

• Synchronization of different information types
  ▫ Multimedia synchronization refers to temporal relationships between the media objects. A common example of temporal relationship is movie or television, where both audio and video objects are involved.

• Reliable security features and firewalls
  ▫ Security features include digital watermarking, data hiding, multimedia content protection, biometrics, multimedia human-computer interface.
  ▫ Firewall refers to a system, which controls the incoming and outgoing network traffic based on an applied rule set.
Multimedia Networks

• Wide-Area Network (WAN)
  ▫ Typically, a WAN consists of a number of interconnected switching nodes. It covers a broad area using leased telecommunication lines. Conventionally, WANs have been implemented using one of two technologies, i.e., circuit switching and packet switching.

• Local Area Network (LAN)
  ▫ A LAN interconnects computers within a limited area. Hence, the scope of LAN is typically a single building or a cluster of buildings. The internal data rates of LANs are much greater than those of WANs.
## LAN Technology

<table>
<thead>
<tr>
<th>Features</th>
<th>FDDI</th>
<th>Ethernet</th>
<th>Token Ring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission rate</td>
<td>125 MBAUD</td>
<td>20 MBAUD</td>
<td>8 &amp; 32 MBAUD</td>
</tr>
<tr>
<td>Data rate</td>
<td>100 Mbps</td>
<td>10 Mbps</td>
<td>4 &amp; 16 Mbps</td>
</tr>
<tr>
<td>Signal encoding</td>
<td>4B/5B</td>
<td>Manchester</td>
<td>Diff. Manchester</td>
</tr>
<tr>
<td></td>
<td>80% Efficient</td>
<td>50% Efficient</td>
<td>50% Efficient</td>
</tr>
<tr>
<td>Maximum coverage</td>
<td>100 km (UTP, Fiber)</td>
<td>2.5 km (Coax., UTP, Fiber)</td>
<td>Configuration dependent</td>
</tr>
<tr>
<td>Maximum nodes</td>
<td>500</td>
<td>1024</td>
<td>250</td>
</tr>
<tr>
<td>Maximum distance between nodes</td>
<td>2 km (MMF)</td>
<td>2.5 km</td>
<td>300 m recommended 100 m</td>
</tr>
<tr>
<td></td>
<td>40 km (SMF)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FDDI: Fiber Distributed Data Interface
FDDI Properties

• FDDI uses a ring topology of multimode or single mode optical fiber transmission links operating at 100 Mbps to span up to 100 km and permits up to 500 stations.

• To accommodate a mixture of stream, FDDI is designed to handle two types of traffic:
  ▫ **Synchronous** frames that typically have tighter delay requirements (e.g., voice and video).
  ▫ **Asynchronous** frames that have greater delay tolerances (e.g., data traffic).

• FDDI uses TTRT (Target Token Rotation Time) to ensure that token rotation time is less than some value.
## FDDI Layers

<table>
<thead>
<tr>
<th>Layer</th>
<th>Sub-Layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Layer</td>
<td>Physical (PHY)</td>
</tr>
<tr>
<td></td>
<td>Physical medium dependent (PMD)</td>
</tr>
<tr>
<td>Data Link Layer</td>
<td>Logical Link Control (LLC)</td>
</tr>
<tr>
<td></td>
<td>Media Access Control (MAC)</td>
</tr>
<tr>
<td>Station Management (SMT)</td>
<td></td>
</tr>
</tbody>
</table>
FDDI Layers

- Physical Layer provides the media independent functions associated with the OSI physical layer. Its functionalities also include decoding and encoding (PMD to a symbol stream and data and control symbols to PMD).

- MAC layer provides fair (i.e., no node has higher priority than others in accessing the medium) and deterministic access (i.e., under error-free conditions, the time a node has to wait to access the medium can be predicted).

- SMT is a sophisticated, build in network monitoring and management capabilities. It is not an OSI-RM specification.
FDDI Advantages

• High Bandwidth (100 Mbps)

• Large distance between FDDI nodes because of very low attenuation (≤0.3 dB/km) in fibers.

• Improved Signal-to-Noise ratio because of no interference from external ratio frequencies and electromagnetic noise.

• Bit error rate ($10^{-11}$) is substantially better than that in Copper ($10^{-5}$) and microwave system ($10^{-7}$).

• Very difficult to tap signal from a fiber cable.
FDDI Limitations

- High cost of optical components required for transmission/reception of signals (especially for single mode fiber network)

- More complex to implement than existing low speech LAN technologies such as Ethernet and Fast Ethernet.
WAN Technology

- Public Switched Telephone Network (PSTN)
- Integrated Service Digital Network (ISDN)
- Broadband ISDN
- Wireless Network (note: wireless LAN)
- Broadcast Channel: Terrestrial, Cable, Satellite.
PSTN

• Now known as *Plain Old Telephone Service* (POTs)

• The term *switched* means a subscriber can make a call to any other telephone on the ‘total’ network.

• The copper wire between the home and central office has a limited bandwidth (56 Kbps). This is still too low for carrying video with pleasing quality.

• Higher data rates, up to 6 Mbps in the downlink direction are possible with ADSL (Asymmetric Digital Subscriber Loop) modems. ADSL is one of the main transport media for streaming MPEG-1, MPEG-2 movies through VOD services.
ISDN

- ISDN is the first network using digital transmission. The data rate can be multiples of basic channels (B channels) of 64 Kbps each, with multiplying factor ranging from 1 to 24.

- The basic subscription of ISDN comes with a 2B+D channel, where D channel is 16 Kbps and is used for return signaling. At 128 Kbps, very low quality video can be achieved.

- ISDN connection is very reliable. H.320 (i.e., the first recommendation series for audio-video conferencing) was developed for ISDN.
Broadband ISDN

- B-ISDN services offer H.0 channels (384 Kbps), H.11 channels (1.536 Mbps) and H.12 channels (1.920 Mbps). This requires higher bandwidth coaxial cable or optical fiber.

- B-ISDN uses Asynchronous Transfer Mode (ATM) packet switching with fixed size packets (known as cells). The very short cell size (53 Bytes with 48 Bytes payload) makes B-ISDN suitable for real-time applications with low delay requirements.

- Cell loss can occur due to traffic congestion, although the loss ($10^{-6}$-$10^{-4}$) is quite low for video services.
Why the Cell Size is 53 Bytes?

• The cell size is determined as a trade-off between packetizing delay and cell overhead. Generally, cell overhead (waste) is smaller at larger payload sizes.

• However, small cell size is preferred because of cell loss. For example, a 53 Bytes cell contain 48 voice samples, which is only $48 \times 125 \, \mu s = 6 \, ms$ of voice. The loss of a cell would be almost unnoticed. However, a loss of cell with 32 ms of voice would be very disruptive.

• 32 Byte is the maximal payload size, at which there is no noticeable packetizing delay, but the overhead would be 13.5%. 64 Bytes payload would produce a small echo in voice communication, but the overhead would be 7.1%.
ATM Service Categories

- **Constant Bit Rate (CBR)**
  - CBR supports real-time applications that require tightly constrained delay and delay variation (e.g., video conferencing, telephone call, video/audio distribution).

- **Variable Bit Rate (VBR)**
  - Real-time VBR supports real-time applications, but may allow more efficient use of a network.
  - Non Real-time VBR supports real-time applications, which are more tolerant of network delays (e.g., airline reservations, banking transactions, process monitoring)
ATM Service Categories (Cont.)

• Available Bit Rate (ABR)
  ▫ ABR is a best effort service, which supports non real-time applications that allow congestion control because the sender can be informed to slow down the traffic in congestion periods (e.g., critical data transfer, text/data/image retrieval and distribution, remote terminal).

• Unspecified Bit Rate (UBR)
  ▫ UBR is a best effort service without any performance requirement. UBR is equivalent to Internet.
## Wireless Network

<table>
<thead>
<tr>
<th>Network</th>
<th>Data Rate</th>
<th>Mobility</th>
<th>Range</th>
<th>Channel Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular Network</td>
<td>&lt;20 Kbps</td>
<td>High</td>
<td>2.5 km</td>
<td>Poor</td>
</tr>
<tr>
<td>Wireless Data Network</td>
<td>64-384 Kbps</td>
<td>High</td>
<td>2.5 km</td>
<td>Poor</td>
</tr>
<tr>
<td>Wireless LAN</td>
<td>2-25 Mbps</td>
<td>Low (Indoor)</td>
<td>50 m</td>
<td>Location dependent</td>
</tr>
<tr>
<td>Wireless IP Network</td>
<td>1-600 Mbps downlink</td>
<td>Low (Indoor)</td>
<td>3-30 miles</td>
<td>Good</td>
</tr>
</tbody>
</table>

Cellular Network: GPRS, EDGE (8.8-59.2 Kbps)

Wireless Data Network: 3G (BER <10^{-6})

Wireless IP Network (2.15-40 GHz): MMDS (1 Mbps), LMDS (600 Mbps)
• Difficulties for video transport over wireless networks lie in the low bandwidth, high error rate, and most importantly, the fluctuation of available bandwidth and error characteristics.

• For real-world applications, quite sophisticated FEC (Forward Error Check) codes are used to reduce the bit error rates significantly.

• Most wireless interactive multimedia communication systems employ H.223. H.223 conveys media data including compressed video in the form of packets of variable size. Typical packet sizes are around 100 bytes to ensure good delay characteristics.
3G Wireless System

- 3G provide higher data rates by using higher carrier frequencies, wider bandwidth and more sophisticated techniques for multiple access, error control, and signal detection.

- BER for data transmission is below $10^{-6}$.

- Generally, wireless channel is quite noisy with high bit error rates. But the use of rate adaptation, FEC and ARQ yield an almost error-free environment for data transmission. For video transmission, where ARQ has to be limited, one has to cope with quite high bit error rates and packet loss rates.
Broadcast Channels

• This is used for broadcasting of digital TV including HDTV using the MPEG-2 video coding and transport streams.

• The compressed data are carried over packets of 188 bytes each. The MPEG-2 transport layer ensure an almost error free environment.

• For SDTV, the available bandwidth is between 3-10 Mbps. For HDTV, 20 Mbps is typically allocated.
# Video Application Characteristics

<table>
<thead>
<tr>
<th>Application and Standard Family</th>
<th>Multiplex Protocol</th>
<th>Video Coding Standard</th>
<th>Bit Rate</th>
<th>Error Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISDN Video Phone (H.320)</td>
<td>H.221</td>
<td>H.261 and H.263</td>
<td>64-384 Kbps</td>
<td>Error free</td>
</tr>
<tr>
<td>PSTN Video Phone (H.324)</td>
<td>H.223</td>
<td>H.263</td>
<td>20 Kbps</td>
<td>Very low</td>
</tr>
<tr>
<td>Mobile Video Phone (H.324 wireless)</td>
<td>H.223</td>
<td>H.263</td>
<td>10-300 Kbps</td>
<td>BER = $10^{-5}$-$10^{-3}$ Occasional packet loss</td>
</tr>
<tr>
<td>Video Phone over Packet network (H.323)</td>
<td>H.225/RTP/UDP/IP</td>
<td>H.261, H.262, H.263</td>
<td>10-1000 Kbps</td>
<td>BER = 0, 0-30% packet loss</td>
</tr>
<tr>
<td>Satellite TV</td>
<td>MPEG-2 System</td>
<td>MPEG-2</td>
<td>6-12 Mbps</td>
<td>Almost error free</td>
</tr>
<tr>
<td>Video conferencing over ATM (H.320, H.321)</td>
<td>H.222</td>
<td>H.262</td>
<td>1-12 Mbps</td>
<td>Almost error free</td>
</tr>
</tbody>
</table>

H.262 is identical to MPEG-2 video, H.222 is identical to MPEG-2 system
QUALITY OF SERVICE

Dr. Quang Duc Tran
## End-user QoS Categories Mapping

<table>
<thead>
<tr>
<th>Error Tolerant</th>
<th>Conversation Voice and Video (FER &lt;3%)</th>
<th>Voice/Video Messaging (FER &lt; 3%)</th>
<th>Streaming Audio and Video (FER &lt; 1%)</th>
<th>Fax (BER &lt; 10^{-6})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Intolerant</td>
<td>Command Control (e.g., Telnet)</td>
<td>Transactions (e.g., E-commerce)</td>
<td>Messaging Downloads (e.g., FPT)</td>
<td>Background (e.g., Email arrival)</td>
</tr>
<tr>
<td>Interactive</td>
<td>Interactive (delay &lt;&lt; 1s)</td>
<td>Responsive (delay ~ 2s)</td>
<td>Timely (delay ~ 10s)</td>
<td>Non-critical (delay &gt;&gt; 10s)</td>
</tr>
</tbody>
</table>

ITU-T G.1010 Standard
## End-user Performance Expectations (Conversational Services)

<table>
<thead>
<tr>
<th>Media</th>
<th>Application</th>
<th>Degree of Symmetry</th>
<th>Data Rate</th>
<th>Performance Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>One-way Delay</td>
</tr>
<tr>
<td>Audio</td>
<td>Conversation Voice</td>
<td>Two-way</td>
<td>4-25 Kbps</td>
<td>&lt; 150 ms preferred &lt; 400 ms</td>
</tr>
<tr>
<td>Video</td>
<td>Video Phone</td>
<td>Two-way</td>
<td>32-384 Kbps</td>
<td>&lt; 150 ms preferred &lt; 400 ms</td>
</tr>
<tr>
<td>Data</td>
<td>Interactive Games</td>
<td>Two-way</td>
<td></td>
<td>&lt; 250 ms</td>
</tr>
<tr>
<td>Data</td>
<td>Telnet</td>
<td>Two-way</td>
<td></td>
<td>&lt; 250 ms</td>
</tr>
</tbody>
</table>
Conversation Voice

• The preferred range of audio transfer delay is 0-150 ms (below 30 ms the user does not notice any delay at all).

• There are three types of satellite systems: LEO, MEO, and GEO. For LEO and MEO, the propagation delay for transmitted signal varies from 10 ms to 250 ms. A GEO system cannot achieve an end-to-end delay below 250 ms.

• The human ear is highly intolerant to short-term delay variation (jitter) so it should be kept really low (< 1 ms).
Video Phone

- Video Phone requires a full-duplex system, carrying both video and audio. The same delay requirements of conversation voice will be applied with added requirement that audio and video must be synchronized within certain limits to provide ‘lip-synch’.

- Human eye is tolerant to some information loss, and hence, some degree of packet loss is acceptable.
# End-user Performance Expectations (Interactive Services)

<table>
<thead>
<tr>
<th>Media</th>
<th>Application</th>
<th>Degree of Symmetry</th>
<th>Data Rate</th>
<th>Performance Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>Voice Messaging</td>
<td>Primarily One-way</td>
<td>4-13 Kbps</td>
<td>One-way Delay: &lt; 1 s playback, &lt; 2 s record, &lt; 1 ms, &lt; 3% FER</td>
</tr>
<tr>
<td>Data</td>
<td>Web-browsing HTML</td>
<td>Primarily One-way</td>
<td>&lt; 0.5 s/page preferred, &lt; 4 s/page</td>
<td>Delay Variation: N/A, Information Loss: Zero</td>
</tr>
<tr>
<td>Data</td>
<td>Transaction services</td>
<td>Two-way</td>
<td></td>
<td>N/A, Zero</td>
</tr>
<tr>
<td>Data</td>
<td>Email (Server Access)</td>
<td>Primarily One-way</td>
<td>&lt; 4 ms</td>
<td>N/A, Zero</td>
</tr>
</tbody>
</table>
# End-user Performance Expectations (Streaming Services)

<table>
<thead>
<tr>
<th>Media</th>
<th>Application</th>
<th>Degree of Symmetry</th>
<th>Data Rate</th>
<th>Performance Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>One-way Delay</td>
</tr>
<tr>
<td>Audio</td>
<td>Speech, Music</td>
<td>Primarily One-way</td>
<td>5-128 Kbps</td>
<td>&lt; 10 s</td>
</tr>
<tr>
<td>Video</td>
<td>Movie clips, Real-time video</td>
<td>Primarily One-way</td>
<td>1-12 Mbps</td>
<td>&lt; 10 s</td>
</tr>
<tr>
<td>Data</td>
<td>Data Transfer</td>
<td>Primarily One-way</td>
<td>&lt; 384 Kbps</td>
<td>&lt; 10 s</td>
</tr>
<tr>
<td>Data</td>
<td>Still image</td>
<td>Primarily One-way</td>
<td></td>
<td>&lt; 10 s</td>
</tr>
</tbody>
</table>
Queuing and Scheduling

• First In First Out (FIFO)
  ▫ There is a single queue and data is served according to its own arrival time. Hence, high priority packets may get stuck behind low priority packets.
  ▫ Aggressive flows obtain better performances because they fill more the queue.

• Priority Scheduling
  ▫ Delay, jitter and loss are reduced for the high priority traffic at the cost of starving the lower priority traffic.
  ▫ A parameter may be assigned to each priority queue, which determines the extend, to which the priority queue is served.
Queuing and Scheduling (Cont.)

- **Weighted Fair Queuing (WFQ)**
  - All traffic is classified into the so-called traffic classes, which can be either individual flows or a bunch of flows with similar transmission requirements.
  - A share of bandwidth for each class is provided in proportion to its specified rates.

- **Round Robin**
  - Buffer is organized in separate queues (each implemented FIFO) for each flow and a single packet is selected at time from queues with a circular mode.
Congestion Control & Queue Discard

- **Tail Drop**
  - Tail Drop drops arriving packets when buffers in queue are full. It may lead to network meltdown due to TCP global synchronization.

- **Random Early Discard (RED)**
  - RED is more fair than tail drop because it does not posses a bias against traffic that uses only a small portion of the bandwidth. The more host transmits, the more likely it is that its packets are dropped.

- **Weighted Random Early Discard (WRED)**
  - A variant of RED, which attempts to weight queues for random early discard.

- **Tri-Color Marking**
Best-Effort

- Best-effort does not provide QoS because there is no ordering of packets.

- In a best-effort network, all users obtain unspecified variable bit rate and delivery time, depending on the current traffic load.

- The internet protocol offers a best-effort service of delivering data between hosts, which can be lost, delayed, corrupted or duplicated.
Integrated Services

- Integrated Services (IntServ) makes strict bandwidth reservations. It must be configured on every router along a path.

- Each application that requires a service guarantee has to make a reservation by using Resource Reservation Protocol (RSVP) signaling.

- When bandwidth is reserved for a certain application, it cannot be reassigned for another application.
Integrated Services (Cont.)

- Routers between the sender and the receiver determine whether they can support the reservation made by the application.

- The task of reserving paths would be very tedious in a busy network such as the Internet.

- The main drawback of IntServ is its lack of scalability (Routers have to classify, police and queue each flow). Moreover, many Internet flows are short-lived, and hence, not worth setting up Virtual Channel.
IntServ Mechanism

• Provide the network with a set of information, known as Flow Specs. Flow Specs include Traffic Specification (TSPEC) and Request Specification (RSPEC).

• TSPEC describes the flow traffic characteristics. RSPEC describes the service request (request for controlled traffic and/or delay bound).
Inserv Mechanism (Cont.)

- Admission Control asks the network to provide particular services. The decision may be based on heuristics (e.g., “Last time I allowed a flow with the TSPEC but the delay exceeded the acceptable bound, hence I would say no”).

- For admission control, the so-called Token Bucket Filter only passes packets arriving at a rate which is not exceeding some administratively set rate, but with possibility to allow short bursts in excess of this rate.
Inserv Mechanism (Cont.)

- Application and network exchange information to request services, flow specs and admission control by using the Resource Reservation Protocol (RSVP).

- Each packet is mapped into a class service, which determine how the packet is scheduled and handled. Typically, the WFQ should be considered to provide a guaranteed end-to-end delay.
Soft State

- Per Session State (Path state, reservation state) has a timer associated with it. State lost when timer expires.
- Senders and Receivers periodically refresh the state, resend PATH/RESV messages, reset timer.
- State can be explicitly deleted by a Teardown message.
- Soft-State is useful for efficiency, but not essential. It supports dynamic automatic adaptation to network changes.
RSVP

- RSVP allows applications running in hosts to reserve resources in the Internet for their data flows.

- RSVP must be present in the receivers, senders and routers.

- RSVP provides reservations for bandwidth in multicast trees. It is also receiver-oriented, i.e., receiver initiates and maintains the resource reservation for data flows.

- RSVP is not routing protocol, sometimes referred to as a signaling protocol.
RSVP Mechanism

- Sender sends PATH message to let the routers know on which links they should forward the reservation (RESV) message. PATH message contains TSPEC and specifies source traffic characteristics (max. bandwidth, token bucket size and max. packet size).

- Receiver requests for resources using RESV message. There are three reservation styles, which can be Fixed-filter, Wildcard-filter and Shared-explicit.
RSVP Mechanism (Cont.)

- **Fixed-filter style**
  - It specifies a list of senders from which it wants to receive a data flow along with a single bandwidth reservation. This reservation is not shared.

- **Wildcard-filter style**
  - It tells the network that it wants to receive all flow from all senders and the bandwidth reservation is shared among the senders.

- **Shared-explicit style**
  - It specifies a list of senders from which it wants to receive a data flow along with a single bandwidth reservation. This reservation is shared among all the senders in the list.
**Differential Services**

- DiffServ does not require nodes in the network to remember state information in the routers. Remembering state information and reserving paths in a busy network (Internet) may be a tedious task.

- Differential Services (DiffServ) provides a improved level of service in the existing best-effort environment by differentiating traffic flow.

- DiffServ reduces the latency in traffic containing voice and streaming video, while providing best-effort service to traffic containing file transfer. It discards more packets in low priority traffic class upon congestion.
DiffServ Architecture

- **DiffServ Policy** specifies which traffic receives a particular level of service in the network. It also marks packets as IN when the measure traffic rate complies to its profile. Otherwise OUT, which are dropped upon congestion.

- **Edge Router** classifies incoming services according to policy specified and measurement. It marks packets with a code point reflecting the desired level of services.

- **Core Router** differentiates incoming packets based on code point and entries in Per-Hop-Behavior (PHB) tables.
Per-Hop-Behavior

• PHB defines differences in performance among classes such that class A packets have strict priority over class B packets.

• PHBs can be as follow:
  ▫ Expedited Forwarding (EF) has the characteristics of low delay, low loss and low jitter. EF traffic is given strict priority queuing above all other traffic classes. Typical networks will limit EF traffic to no more than 30% of the capacity of a link.
  ▫ Assured Forwarding (AF) defines four separate classes. When congestion occurs, the traffic in the higher class is given higher priority (WFQ), and the packets with the higher drop precedence are discarded (RED).
SYNCHRONIZATION

Dr. Quang Duc Tran
Multimedia Protocol Stack

- DASH
- HTTP
- TCP
- DCCP
- IP Version 4, IP Version 6
- AAL3/4
- AAL5
- MPLS
- ATM/Fiber Optics
- Ethernet/WIFI

Synchronization Service
- SIP
- RTSP
- RSVP
- RTCP

Media encaps (H.264, MPEG-4)
- RTP

UDP

- RTP
Synchronization Issues

• Content Relations
  ▫ It defines a dependency of media objects on some data. An example of a content relation is two graphics that are based on the same data but show different interpretations of the data.

• Spatial Relations
  ▫ It defines the space used for the presentation of a media object on an output device at a certain point of time in a multimedia presentation.

• Temporal Relations
  ▫ It defines the temporal dependencies between media objects. They are of interest whenever time-dependent media objects exist.
Intra- and Inter-Object Synchronization

• Intra-Object Synchronization
  ▫ It refers to the time relation between various presentation units of one time-dependent media object. An example is the time relation between the single frames of a video sequence. For a video with a rate of 25 frames per second, each of the frames must be displayed for 40 ms.

• Inter-Object Synchronization
  ▫ It refers to the synchronization between media objects. An example is a multimedia synchronization that starts with an audio/video sequence, followed by several pictures and an animation that is commented by an audio sequence.
Live and Synthetic Synchronization

- **Live Synchronization**
  - The goal of the synchronization is to exactly reproduce at a presentation the temporal relations as they existed during the capturing process. An example is a video conference for person-to-person discussion, which demands lip synchronization of the audio and video.

- **Synthetic Synchronization**
  - The temporal relations are artificially specified and assigned to media objects that were created independently of each other. For example, 4 audio messages are recorded as a part of engine in animation. The time relations between animation and matching audio sequences are specified.
Presentation Requirements

• Lip Synchronization
  ▫ It refers to the temporal relationship between an audio and video stream for the particular case of human speaking.
    • The “in sync” region spans a skew of +/-80 ms.
    • The “out of sync” area spans a skew of +/-160 ms.

• Pointer Synchronization
  ▫ The speakers use a pointer to point out individual elements of the graphics which may have been relevant to the discussion taking place.
    • The “in sync” region spans a skew of +750/-500 ms.
    • The “out of sync” area spans a skew of +1250/-1000 ms.
Each layer implements synchronization mechanisms, which are provided by an appropriate interface. Each interface defines services, offering the user a means to define his requirements. Each interface can be used by an application or by the next higher layer.
• **Media Layer**
  ▫ An application operates on a single continuous media stream, which is treated as a sequence of Local Data Units. Using this layer, the application is responsible for the intra-stream synchronization by using flow-control mechanisms.

• **Stream Layer**
  ▫ It operates on continuous media streams as well as on groups of media streams. In a group, all streams are presented in parallel by using mechanisms for inter-stream synchronization.
  ▫ The streams are executed in a real-time environment, where all processing is constrained by well-defined time specifications. The applications requiring the stream layer service are executed in a non real-time environment.
Reference Model (Cont.)

- **Object Layer**
  - It operates on all types of media and hides the differences between time-independent and time-dependent media. To the application, it offers a complete and synchronized media. This layer takes a synchronization specification as input and is responsible for the correct schedule of the overall presentation.

- **Specification Layer**
  - It is an open layer, which contains application and tools that allow one to create synchronization. For example, MODE system offers a graphical interface to select video and text objects to use, to select suitable points where the subtitles have to be shown, to specify the temporal relations of these points and to store synchronization specification.
Synchronization Specification

• Synchronization specification describes all temporal dependencies of the included object in the multimedia object. It should be comprised of inter- and intra-object synchronization for the media objects of the presentation and QoS for inter- and intra-object synchronization.

• In the case of live synchronization, the temporal relations are implicitly defined during capturing. QoS requirements are defined before starting the capture.

• In the case of synthetic synchronization, the specification must be created explicitly.
Synchronization Specification (Cont.)

Interval-Based Specification

Axes-Based Specification

Control Flow-Based Specification

Event-Based Specification

- Start a presentation
- Stop a presentation
- Prepare a presentation
Distributed Environment

- Synchronization in a distributed environment is more complex than in a local environment. This is caused by the distributed storage of synchronization information and the distributed locations of source and the sink (receiver). Even different media objects involved in the presentation may be located at different places.

- The communication between the storage and presentation site introduces additional delays and jitter.

- Often, we also encounter multi-party communication patterns.
Transport of the Sync Specification

• Delivery of the complete synchronization information before the start of the presentation.
• Use of an additional channel.
  ▫ +: No addition delay
  ▫ -: Errors caused by delay or loss of synchronization units
  ▫ -: Difficult to handle multiple source nodes
• Multiplexed data streams
  ▫ +: No additional channel and delay
  ▫ +: MPEG bit stream combines audio, video and sync info.
  ▫ -: Difficult to select an appropriate QoS
  ▫ -: Difficult to handle multiple source nodes
Location of Sync Operations

Synchronization at the sink node

Synchronization at the source node
Clock Synchronization

• It is possible to allocate buffer capacities at the sink to guarantee that the required media units are available.

• By assuming limited buffer capacity, it is necessary to limit the maximal offset (i.e., \( \max(O_a, O_v) \)). Network Time Protocol allows the synchronization of the clocks with an accuracy in the range of 10 ms.

\[
T_a = T_{av} - N_{la} - O_a \\
T_v = T_{av} - N_{lv} - O_v \\
O_a \text{ and } O_v \text{ are unknown}
\]
Multi-Step Synchronization

1) Synchronization during object acquisition, e.g., during digitizing video frames.
2) Synchronization of retrieval, e.g., synchronized access to frames of a stored video.
3) Synchronization during delivery of the Local Data Units to the network.
4) Synchronization during transport, e.g., by isochronous protocols.
5) Synchronization at the sink, i.e., synchronized delivery to the output devices.
6) Synchronization within the output device.
Multimedia Traffic

The production, transmission, and use of data take place at the same time

Real-time Multimedia Traffic

The production, transmission, and use of data take place at different times

Streaming Live A/V
(Broadcast TV/radio via Internet)
Can not pause, rewind. The time between request and display is from 1 to 10 seconds.

Real-Time Interactive A/V
(IP Phone, Video conferencing)
Can not pause, rewind. The time between request and display is small (video<150 ms and audio<400 ms)

Streaming Stored A/V
(Like VoD)
May pause, rewind... The time between request and display is from 1 to 10 seconds.
The use of timestamps may overcome the jitter problem. Each packet has a time of the packet with respect to the first packet.

The playback is delayed 7s after receiving the first packet.

Playback buffer is needed to separate playback time from the arrival time.
Playback Point

- \( D = ED + \beta \ EV \)
  
  Where
  - \( D \): Playback Point
  - \( ED \): Estimated average packet delay
  - \( EV \): Estimated average packet delay variation
  - \( \beta \): Safety Factor (\( \beta = 4 \))

- \( ED_i = \alpha ED_{i-1} + (1 - \alpha) (r_i - t_i) \)
- \( EV_i = \alpha ED_{i-1} + (1 - \alpha) (r_i - t_i - ED_i) \)

  Where
  - \( \alpha \): Weighting Factor (\( \alpha = 0.998 \))
  - \( r_i \): Time the packet \( i \) is received
  - \( t_i \): Timestamp of the packet \( i \)
Why Real-Time Data Can Not Be TCP?

- TCP forces the sink application to wait for retransmission(s) in the case of packet loss, causing large delays.

- TCP cannot support multicast, which is a basic requirement of video conferencing applications.

- TCP congestion control mechanisms decreases the congestion window when packet losses are detected. Audio and video on the other hand have bitrates that cannot be suddenly decreased.
Why Real-Time Data Can Not Be TCP?

- TCP headers are larger than a UDP header.
- TCP does not contain the timestamp and encoding parameters, needed by the receiver.
- TCP does not allow packet loss. In A/V, a loss of 1-20% is tolerable. The loss can be compensated by FEC.
Multimedia Protocol Stack

Synchronization Service

HTTP

TCP

DCCP

UDP

IP Version 4, IP Version 6

AAL3/4

AAL5

MPLS

ATM/Fiber Optics

Ethernet/WIFI

Media encaps (H.264, MPEG-4)

RTP

RTCP

RSVP

RTSP

SIP

DASH
Real-Time Transport Protocol

• RTP is a network protocol for delivering audio and video over IP network. RTP is used in conjunction with the Real-Time Control Protocol (RTCP). While RTP carries the media streams, RTCP is used to monitor transmission statistics and QoS and aids synchronization of multiple streams.

• RTP does not ensure real-time delivery, but it provide means for
  ▫ Jitter elimination/reduction by using playback buffer.
  ▫ Synchronization of several audio and video streams.
  ▫ Multiplexing of audio and video streams.
  ▫ Translation of audio and video streams.
Real-Time Transport Protocol

<table>
<thead>
<tr>
<th>Ver.</th>
<th>P</th>
<th>X</th>
<th>Contr. Count</th>
<th>M</th>
<th>Payload Type</th>
<th>Sequence Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Timestamp**

**Synchronization Source Identifier**

**Contributor Identifier**

---

Version Number (2)

Padding Bit (1 – Packet contains padding)

Marker bit (1 – The frame boundary is marked)

Incremented for each RTP packets (it is used to indicate packet loss and packet sequence)

Extension bit (1 - Fixed header is followed by an extension header)

UDP Header | RTP Header | RTP Payload | Padding | Pad Count
Timestamp and Sequence No.

- **Audio**
  - RTP packet carries 20 ms of audio samples. Timestamp clock rate for audio is 8000 Hz. Hence, timestamp increments by 160.
  - No. of bits per RTP payload for uncompressed audio is 160x8=1280. That for compressed audio is typically 8 times less.

- **Video**
  - RTP packet carries one video frame. RTP packet rate is 25 or 30 Hz. Timestamp clock rate for video is 90,000 Hz. Hence, timestamp increments by 3600 or 3000.
  - No. of bits per RTP payload for uncompressed video conferencing is 352x240x12 = 10000
Real-Time Control Protocol

- RTCP provides out-of-band statistics (e.g., packet loss, packet delay variation, round-trip delay time) and control information for an RTP session.

- The functionalities of RTCP include:
  - Gathering statistics on quality aspects of the media distribution and transmitting this data to the session media source and other session participant.
  - Provisioning session control functions. RTCP is a convenient means to reach all session participants. RTP is only transmitted by a media source.
Real-Time Control Protocol (Cont.)

- **UDP Header**
- **RTCP Packet**
- **RTCP Packet**
- **RTCP Packet**

**RTCP Header**

- **Version**
- **P**
- **RR Count**
- **Packet Type**
- **Message Length**

- **Padding bit**

- **Number of Reception Report Blocks, contained in the packet**

- 200: SR (Sender Report)
- 201: RR (Receiver Report)
- 202: SDES (Source Destination)
- 203: BYE
- 204: APP (Application Specific Message)
- 207: XR (RTCP Extension)
Real-Time Control Protocol (Cont.)

• **Sender Report (SR)**
  ▫ It is sent periodically by the active senders to report transmission and reception statistics. The report include an absolute timestamp, allowing the receiver to synchronize RTP messages. (Note: video and audio streams use independent relative timestamps).

• **Source Description (SDES)**
  ▫ It is used to send CNAME item to session participant that provides additional information such as the name, e-mail address, telephone number of the owner of the source.
Real-Time Control Protocol (Cont.)

- Receiver Report (RR)
  - It informs the sender and other receivers about the QoS.

- Goodbye (BYE)
  - A source sends a BYE message to shut down a stream. It also allow an endpoint to announce that it is leaving the conference.

- Application Specific Message (APP)
  - The application-specific message provides a mechanism to design application-specific extensions to the RTCP protocol.
• The volume of RTCP traffic may exceed the RTP traffic during a conference session involving large number of participants. This is because RTCP packets are sent regardless whether participant is talking or not.

• RTCP traffic is dynamically changed depending of the number of participants. Typically, it is designed to be no more than 5% of the RTP traffic (1.25% allocated to sender, and 3.75% allocated to receivers). As number of receivers increases, frequency of response per receiver decreases.
Forward Error Correction

\[ P_1 = \text{XOR}(D_1, D_2, D_3, \ldots, D_{N-1}) \]
\[ D_3 = \text{XOR}(D_1, D_2, \ldots, D_{N-1}, P_1) \]

The Parity Packets can help to recover the loss of any N-k out of N packets (Reed Solomon Erasure Code).

FEC increases the required bandwidth and latency.
Interleaving

Original RTP packets, each contains 20 ms of voice samples

Lost packet causes 5 ms gaps in the audio stream, which can not be noticed. Interleaving does not increase the bandwidth, but increases delays.
Receiver-Based Repair does not increase the bandwidth requirement nor delays. It works with small packets and is based on the assumption that there is a small difference between two neighboring packets (voice packets).

Packet is recovered by interpolation, which can be computationally expensive and small delay.
Real-Time Streaming Protocol

- RTSP is a network control protocol (port number is 554), designed for controlling streaming media servers. It is used to establish and control media session between end-points. Most RTSP servers use the RTP and RTCP for media stream delivery.

- Similarly to HTTP, RTSP defines control sequences useful in controlling multimedia playback and uses TCP to maintain end-to-end connection. Unlike HTTP, RTSP has state. Request can be made by both the streaming server and client.
# Real-Time Streaming Protocol (Cont.)

<table>
<thead>
<tr>
<th>Control Request</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SETUP</td>
<td>Asks the server to allocate resources for a stream and start an RTSP session.</td>
</tr>
<tr>
<td>PLAY</td>
<td>Starts data transmission on the allocated stream.</td>
</tr>
<tr>
<td>RECORD</td>
<td>Initiates recording a range of media data.</td>
</tr>
<tr>
<td>PAUSE</td>
<td>Temporarily halts a stream without freeing the allocated resources.</td>
</tr>
<tr>
<td>TEARDOWN</td>
<td>Frees resources associated with the stream.</td>
</tr>
<tr>
<td>ANNOUNCE</td>
<td>Changes description and media object.</td>
</tr>
<tr>
<td>REDIRECT</td>
<td>Redirects request to inform the client to connect to another server location.</td>
</tr>
<tr>
<td>SET_PARAMETER</td>
<td>Sets the value of a parameter for a presentation or stream.</td>
</tr>
<tr>
<td>DESCRIBE</td>
<td>Gets description of media object.</td>
</tr>
</tbody>
</table>
Real-Time Streaming Protocol (Cont.)

1) The Web Server or HTTP Server serves Web pages. The Streaming Server or Media Server, serves the audio/video files.

2) The two servers can run on the same end system or on two distinct end systems.

3) Media Player and Streaming Server can interact using RTSP.
H.323

Dr. Quang Duc Tran
H.323 Definition

- H.323 is a ITU-T Recommendation that specifies the components, protocols and procedures that provide multimedia communication services, including real-time audio, video and data communication.

- It is widely used within various Internet real-time applications, such as NetMeeting.

- H.323 is a part of the ITU-T H.32x series of protocols, which address multimedia communication over ISDN, PSTN and 3G mobile networks.
H.323 Elements

- H.323 Terminal
- H.323 Gatekeeper
- H.323 Gateway
- H.323 Terminal
- H.323 Multipoint Control Unit (MCU)

Networks:
- PSTN
- LAN
- N-ISDN
- B-ISDN

Terminals:
- V.70 Terminal
- H.324 Terminal
- Speech Terminal
- H.322 Terminal
- Speech Terminal
- H.320 Terminal
- H.321 Terminal
H.323 Terminal

• H.323 terminal can be either a personal computer (PC) or a stand-alone device, running H.323 and the multimedia application.

• H.323 terminal supports audio communication and optionally supports video and data communications. It can be used in multipoint conferences.

• H.323 terminal is compatible with H.324 terminals on PSTN, H.310 terminals on B-ISDN, H.320 terminals on ISDN, H.321 terminals on B-ISDN and H.322 terminals on QoS LAN.
H.323 Gateway

• H.323 gateway connects two dissimilar networks. For example, a gateway can connect and provide communication between an H.323 terminal and PSTN networks.

• A connectivity of dissimilar networks is achieved by translating protocols for call setup and release, converting media formats between different networks and transferring information between the networks.

• A gateway is not required for communication between two H.323 terminals.
H.323 Gatekeeper

- H.323 gatekeeper is an optional component in the H.323 network that provides a number of services to terminals, gateways and MCU devices.

- Address resolution is the most important service as it enables two endpoints to contact each other without either endpoint having to know the IP address of the other endpoint.

- H.323 endpoints use RSA protocol to communicate with a gatekeeper. Likewise, a gatekeeper uses RSA to communicate with other gatekeepers.
H.323 Multipoint Control Unit

- H.323 MCU is responsible for managing multipoint conferences. It consists of two logical entities, i.e., Multipoint Controller (Call signaling, conference control) and Multipoint Processor (switching/mixing of media streams).

- An MCU manages conference resources, negotiates between terminals in order to determine the audio and video CODEC, and may handle the media stream. Some MPs can do real-time transcoding of the received audio/video streams.

- The gatekeepers, gateways and MCU can be implemented in a single physical device.
H.323 Multipoint Control Unit

- Multimedia Applications, User Interface
- Terminal Control & Management
  - H.225 Call Signaling
  - H.245
  - H.225 RAS
- Media Control
  - RTP
  - RTCP
  - CODEC
- TCP/UDP
- TCP
- UDP
- UDP
- IP

Text: T.140
Video: H.26x
Audio: G.711, G.729, G.723, G.726
H.225

- H.225 Call Signaling is used to establish a connection between two H.323 endpoints. This is achieved by exchanging H.225 protocol messages between two H.323 endpoints or between an endpoint and the gatekeeper.

- H.225 RSA (Registration, Admission, and Status) is the protocol between endpoints and gatekeepers. RSA messages perform registration, admission control, bandwidth changes, status, and disengage procedures between endpoints and gatekeeper. The signaling channel is opened prior to the establishment of any other channels.
### H.225 Call Signaling

<table>
<thead>
<tr>
<th>Message</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setup</strong></td>
<td>This is sent by H.323 caller to indicate its desire to set up a connection to the callee.</td>
</tr>
<tr>
<td><strong>Call Proceeding</strong></td>
<td>This is sent by the callee to indicate that it has received all the information it needs to route the call to its destination.</td>
</tr>
<tr>
<td><strong>Alerting</strong></td>
<td>This message might be sent by the callee to indicate that the callee is being alerted of the incoming call, i.e., the phone is ringing.</td>
</tr>
<tr>
<td><strong>Connect</strong></td>
<td>This is sent by the callee to the caller to indicate that the call has been accepted.</td>
</tr>
<tr>
<td><strong>Release Complete</strong></td>
<td>This is sent by either callee or caller to indicate the call’s release.</td>
</tr>
<tr>
<td><strong>Notify</strong></td>
<td>This is used to notify a device of a change that has occurred in the call.</td>
</tr>
<tr>
<td><strong>Status Inquiry</strong></td>
<td>This can be used to request call status.</td>
</tr>
<tr>
<td><strong>Status</strong></td>
<td>This is used to respond to an unknown call signaling or a Status Inquiry message.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
H.225 RSA

Gatekeeper Discovery
- Gatekeeper Request (GRQ)
- Gatekeeper Confirm (GCF)
- Gatekeeper Reject (GRJ)

Registration
- Registration Request (RRQ)
- Registration Confirm (RCF)
- Registration Reject (RRJ)

Admission Control
- Admission Request (ARQ)
- Admission Confirm (ACF)
- Admission Reject (ARJ)

Unregistration
- Unregistration Request (URQ)
- Unregistration Confirm (UCF)
- Unregistration Reject (URJ)

Disengage
- Disengage Request (DRQ)
- Disengage Confirm (DCF)
- Disengage Reject (DRJ)

Endpoint Location
- Location Request (LRQ)
- Location Confirm (LCF)
- Location Reject (LRJ)
H.245

H.245 control signaling is used to exchange end-to-end control messages, governing H.323 operation.

<table>
<thead>
<tr>
<th>Master/Slave Determination</th>
<th>Logical Channel Signaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master-Slave Determination</td>
<td>Open Logical Channel</td>
</tr>
<tr>
<td>Master-Slave Determination Ack.</td>
<td>Open Logical Channel Ack.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Close Logical Channel Signaling</th>
<th>Bidirectional Logical Channel Signaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close Logical Channel</td>
<td>Open Logical Channel</td>
</tr>
<tr>
<td>Close Logical Channel Ack.</td>
<td>Open Logical Channel Ack.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ending a session</th>
<th>Capabilities Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td>End Session</td>
<td>Terminal Capability Set</td>
</tr>
<tr>
<td></td>
<td>Terminal Capability Set Ack.</td>
</tr>
<tr>
<td></td>
<td>Terminal Capability Set Reject</td>
</tr>
</tbody>
</table>
H.323 Call Flow

1. Setup
2. Call Proceeding
3. Alerting
4. Connect
5. Capability Exchange
6. Master-Slave Determination
7. Open Logical Channel
8. Open Logical Channel Ack.
9. Open Logical Channel
10. Open Logical Channel Ack.
11. End Session
12. End Session
13. Release Complete
H.323 Call Flow (Cont.)

1. ARQ
2. ACF
3. Setup
4. ARQ
5. ACF
6. Alerting
7. Connect
8. Release Complete
9. DRQ
10. DCF
11. DRQ
12. DCF
H.323 Call Flow (Cont.)

1. ARQ
2. LRQ
3. LCF
4. ACF
5. Setup
6. ARQ
7. ACF
8. Alerting
9. Connect
10. Release Complete
11. DRQ
12. DCF
13. DRQ
14. DCF
H.323 Call Flow (Cont.)

1. H.225 Setup
2. Q.931 Setup
3. Q.931 Call Proceeding
4. H.225 Call Proceeding
5. H.225 Alerting
6. Q.931 Connect
7. H.225 Connect
8. Q.931 Connect Ack.
9. Q.931 Disconnect
10. H.245 End Session
11. H.245 End Session
12. Q.931 Release
13. Q.931 Release Complete
14. H.225 Release Complete
H.323 Ad-Hoc Conference

1. Setup
2. Connect
3. Capability Exchange
4. Master-Slave Determination
5. Logical Channel Establishment
6. Setup
7. Connect
8. Capability Exchange
9. Master-Slave Determination
10. Logical Channel Establishment
Session Initiation Protocol

Dr. Quang Duc Tran
SIP Definition

- Session Initiation Protocol (IETF RFC 3261) is an application layer control (signaling) protocol for creating, modifying and terminating multimedia sessions across packet networks.

- The main signaling functions of SIP are as follows:
  - Location of an endpoint
  - Contacting an endpoint to establish a session
  - Exchange of media information to establish a session
  - Modification of existing media sessions
  - Teardown of existing media sessions
SIP Definition (Cont.)

• SIP is highly extensible. It can be extended to accommodate features and services such as call control services, interoperability with existing telephony systems, and more.

• A SIP network consists of four types of logical SIP entities, i.e., User Agent (UA), Gateway, Back-to-back UA, Proxy Server, Redirect Server and Registrar. Each entity has specific functions and participates in SIP communication as a client (initiating requests), as a server (responding to requests), or as both.
A UA takes input from a user and acts as an agent on his behalf to set up and tear down media sessions with other UAs. The user can be a human or another protocol (e.g., gateway). UAs typically register with a proxy server in their domain.

RFC 2543 defines UA as an application that contains both UA Client (UAC) and UA Server (UAS). UAC is a client application that initiates SIP requests. UAS is a server application that contacts the user when a SIP request is receiving, and returns a response on behalf of the user.
SIP Server

- **A SIP Proxy Server** receives a request from a UA or another proxy and acts on behalf of the UA in forwarding or responding to the request. It does not issue requests and has no media capabilities. Stateful proxy server (e.g., forking proxy server) keeps track of requests and responses received in the past.

- **A SIP Redirect Server** accepts a SIP request, maps the SIP address of the callee into zero (if there is no known address) or a new address and returns it to the UA. It does not pass the request to other servers.

- **A SIP Registrar Server** accepts REGISTER requests for the purposes of updating a location database with the contact information of the user, specified in the request.
SIP Back-to-back User Agent

• A B2BUA is a type of UA that receivers a SIP request, then reformulates the request and sends it out as a new request.

• B2BUAs act like a proxy but do not follow proxy routing rules. B2BUAs are call-stateful while proxy servers are transaction-stateful. Proxy server keeps state only during SIP transactions (i.e., at the beginning and end of a call) and do not keep any state during the whole call.

• B2BUAs can be a part of many devices (e.g., PBX). Conference bridge and mixers also use B2BUA logic.
A SIP gateway is an application that interfaces a SIP network to a network utilizing another signaling protocol. SIP gateway is a special type of UA, there the UA acts on behalf on another protocol.

A SIP gateway may be decomposed into (1) a media gateway (MG) that manages call control protocols, and (2) a media gateway controller (MGC) that manages the media connection.

A SIP gateway can support hundreds of users.
## SIP Requests

<table>
<thead>
<tr>
<th>Message</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>INVITE</td>
<td>Initiates a call, changes call parameters (re-INVITE)</td>
</tr>
<tr>
<td>ACK</td>
<td>Confirms a final response for INVITE</td>
</tr>
<tr>
<td>BYE</td>
<td>Terminates a call</td>
</tr>
<tr>
<td>CANCEL</td>
<td>Cancels searches and “ringing”</td>
</tr>
<tr>
<td>OPTIONS</td>
<td>Queries the capabilities of the other side</td>
</tr>
<tr>
<td>REGISTER</td>
<td>Registers with the Location Service</td>
</tr>
<tr>
<td>INFO</td>
<td>Send mid-session information that does not modify the session state</td>
</tr>
<tr>
<td>UPDATE</td>
<td>Modifies the state of a session without changing the state of the dialog</td>
</tr>
</tbody>
</table>
# SIP Responses

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1xx</td>
<td>Informational (Indicates the status of the call prior to completion)</td>
</tr>
<tr>
<td>2xx</td>
<td>Success (The request has succeeded)</td>
</tr>
<tr>
<td>3xx</td>
<td>Redirection (The server has returned possible locations)</td>
</tr>
<tr>
<td>4xx</td>
<td>Client error (The request has failed due to an error by the client)</td>
</tr>
<tr>
<td>5xx</td>
<td>Server failure (The request has failed due to an error by the server)</td>
</tr>
<tr>
<td>6xx</td>
<td>Global failure (The request has failed, and should not be tried at this or other server)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Trying</td>
</tr>
<tr>
<td>180</td>
<td>Ringing</td>
</tr>
<tr>
<td>200</td>
<td>OK</td>
</tr>
<tr>
<td>202</td>
<td>Accepted</td>
</tr>
<tr>
<td>300</td>
<td>Multiple Choices</td>
</tr>
<tr>
<td>301</td>
<td>Move Permanently</td>
</tr>
<tr>
<td>302</td>
<td>Move Temporarily</td>
</tr>
<tr>
<td>400</td>
<td>Bad request</td>
</tr>
<tr>
<td>401</td>
<td>Unauthorized</td>
</tr>
<tr>
<td>402</td>
<td>Payment Required</td>
</tr>
<tr>
<td>403</td>
<td>Forbidden</td>
</tr>
<tr>
<td>404</td>
<td>Not found</td>
</tr>
<tr>
<td>503</td>
<td>Service Unavailable</td>
</tr>
<tr>
<td>604</td>
<td>Does Not Exist Anywhere</td>
</tr>
</tbody>
</table>
SIP Call Flow

UA A
- INVITE
- 180 - Ringing
- 200 - OK
- ACK
- (RTP/RTCP)
- BYE
- 200 - OK

UA B

UA A
- REGISTER
- 401 Unauthorized
- REGISTER
- 200 - OK

Registrar
Sip: name@hust.com

Location
Credentials
SIP Call Flow (Cont.)

UA A  Redirect Server  UA B  UA A  Proxy Server  UA B

INVITE 302 - Moved ACK INVITE 100 - Trying 180 - Trying 200 - OK INVITE
180 - Ringing 200 - OK ACK
200 - OK

(RTP/RTCP)

INVITE 180 - Ringing 200 - OK
ACK
200 - OK

BYE 200 - OK

200 - OK
SIP Call Flow (Cont.)

UA A

Gateway

Proxy Server

UA B

INVITE

100 - Trying

180 - Ringing

200 - OK

ACK

INVITE

180 - Ringing

200 - OK

ACK

(RTP/RTCP)

INVITE

180 - Ringing

200 - OK

ACK

INVITE

180 - Ringing

200 - OK

ACK

(Q.931 Setup)

(Q.931 Call Pro.)

(Q.931 Connect)

(Q.931 Con. Ack.)

(Q.931 Dis.)

(Q.931 Release)

(Q.931 Re. Com.)

PSTN

Ring

Off-Hook

On-Hook

TDM Stream
SIP Call Flow (Cont.)
VOICE OVER INTERNET PROTOCOL

Dr. Quang Duc Tran
Voice Over Internet Protocol

- VoIP is a methodology and group of technologies for the delivery of voice communications and multimedia sessions over Internet Protocol (IP) network. Other terms associated with VoIP are IP Telephony, Internet Telephony, Broadband Telephony.

- The principles involved in originating VoIP telephone calls are similar to traditional digital telephony and involve signaling, channel setup, digitization of the analog voice signals, and encoding. However, the digital information in VoIP is packetized and transmission occurs as IP packets over a packet-switched network.
# VoIP vs. PSTN

<table>
<thead>
<tr>
<th>Feature</th>
<th>VoIP</th>
<th>PSTN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connectivity type</td>
<td>Internet connectivity</td>
<td>Dedicated telephone lines</td>
</tr>
<tr>
<td>Required bandwidth</td>
<td>~ 10 Kbps in each direction</td>
<td>~ 64 Kbps in each direction</td>
</tr>
<tr>
<td>Pricing</td>
<td>Free VoIP calling (local and international), Calls to mobile and landline phones have nominal subscription fees</td>
<td>No free calls can be made, Costly international calling and monthly phone plans.</td>
</tr>
<tr>
<td>Scalability</td>
<td>More bandwidth and simple software updates</td>
<td>More dedicated lines and hardware</td>
</tr>
<tr>
<td>Remote extensions</td>
<td>Typically standard</td>
<td>Require dedicated lines for each extension and is very pricey</td>
</tr>
<tr>
<td>Disaster recovery</td>
<td>Service terminates when internet connectivity is lost</td>
<td>Service remains active during power outages. But cordless phones would be unusable</td>
</tr>
</tbody>
</table>
VoIP Protocols

- **H.323**
  - An ITU Recommendation that defines packet-based multimedia communication systems, including Video Conferencing, VoIP.

- **Session Initiation Protocol (SIP)**
  - IETF RFC 2543. SIP defines a distributed architecture for creating multimedia applications, including VoIP.

- **Megaco/H.248**
  - An ITU Recommendation that defines Gateway Control Protocol. It is the result of a joint-collaborate with the IETF. H.248 defines a centralized architecture known as Megaco.
VoIP Scenarios

PC to PC

PC to Phone

Phone to Phone
An IP PBX is a private branch exchange that switches calls between VoIP users on local lines while allowing all users to share a certain number of external phone lines.

A typical IP PBX can also switch calls between a VoIP user and a traditional telephone user, or between two traditional telephone users in the same way that a conventional PBX does.
VIDEO CONFERENCING

Dr. Quang Duc Tran
Video Conferencing

- Video Conferencing consists of a set of telecommunication technologies, which allow two or more locations to communicate by simultaneous two-way video and audio transmissions.

- Video Conferencing differs from video phone calls in that it is designed to serve a conference or multiple locations rather than individuals.

- Simultaneous video conferencing among three or more remote points is possible by means of a MCU.
Multipoint Control Unit

- Multipoint Control Unit (MCU), also referred to as “conference bridge”, is a central gateway in a multipoint video conferencing system.

- Some of the functions an MCU can provide
  - Transcoding
  - Call Signaling
  - Conference Setup
  - Audio Mixing
  - Video Composition
Types of VC

• Ad-hoc Conference
  ▫ The conference initiator creates the conference and invites participants to join without sending any prior notification about the conference information. Endpoints that do not have the Conference Key cannot initiate the conference. Resources for an ad-hoc conference cannot be reserved.

• Scheduled Conference
  ▫ The conference organizer schedules the conference using the scheduling tool or calendar applications. The conference resources are reserved at the time the conference is scheduled. The conference cannot be scheduled if there are insufficient resources.
Types of VC

• Video Conferencing over ISDN
  ▫ Dedicated and direct channels between participants are necessary to ensure a minimum quality throughout the development of the event. It is advisable to have at least 3 ISDN lines to have a good user experience.
  ▫ Very few conferencing that uses this type of connection due to its cost, low quality and the need to have additional ISDN lines.

• Video conferencing over IP networks (e.g., ADSL, FTTH)
  ▫ It is the most used system to carry out video conferencing since the speed of connection to Internet are high and stable enough to replace the connections over ISDN. The bandwidth available in IP networks might fluctuate due to packet loss, but it is easy to upgrade to higher bandwidth.
Centralized vs. Distributed VC

Main Campus

IP WAN

MCU

Campus B

IP WAN

MCU
VC Enterprise

- Home Users
- Partner Org.
- IP WAN
- Main Campus
- ISDN Gateway
- MCU
- ISDN PSTN