UNIT – I

AUDIO COMPRESSION

- Audio compression
- DPCM
- Adaptive PCM
- Adaptive predictive coding
- Linear predictive coding
- Code excited LPC
- Perpetual coding

1.AUDIO COMPRESSION:

- For digitization of audio signals pulse code modulation is used. Where sampling rate is twice the maximum frequency of the audio signal.

- Alternatively if the frequency (BW) of the communication channel to be used is less than that of the signal, then the sampling rate is determined by the bandwidth of the communication channel. This is known as a "band limited signal"

- In most Multimedia applications, the band width of the communication channel is less than that of the signal rate,

  In this case, two ways are adopted.

  1. The audio signal is sampled at a lower rate.
  2. Compression algorithm used.

  In the first technique the following draw back occurs,

  1. Quality of the decoded signal is reduced owing to loss of high frequency components.
  2. Use of fever bits per samples introduces high level of quantization noise.

  The second technique achieves a comparable perceptual quality that obtained with a higher sampling rate but with reduced BW requirements.

2. DIFFERENTIAL PULSE CODE MODULATION:-

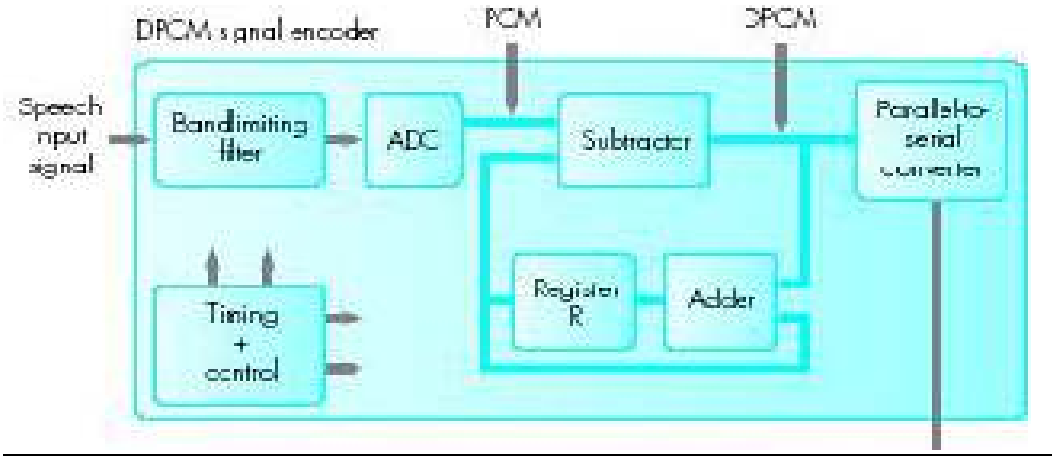
- DPCM is a derivative of standard PCM.
- In DPCM the difference in Amplitude of the current sample with the previous sample is encoded to achieve fever bit rate requirements.
- But the same sampling rate as of PCM is achieved.

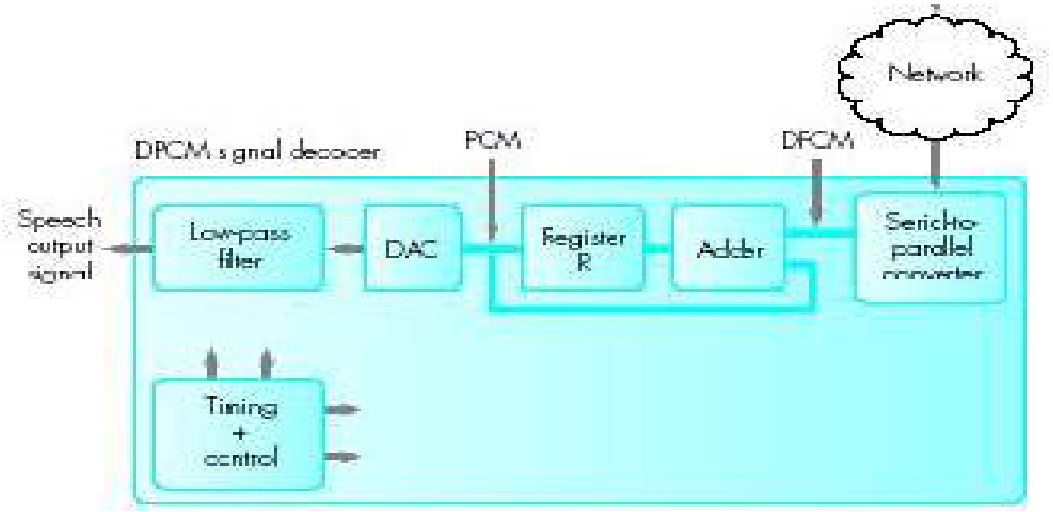
DPCM Encoder:-

- Here the previous digitized sample of the analog signal is held in a temporary register R.

- The difference signal (DPCM) is computed by the subtracting the current content of the register Ro from the new digitized sample output by the ADC
- The value in the register is then up dated by adding the current register content and signal output from the subtractor.
- Parallel to serial converter. Convert parallel data to serial data and send to network.

DPCM Signal Encoder



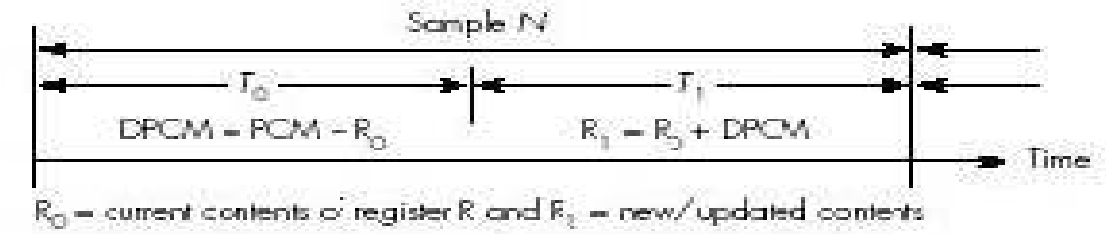


DPCM Signal Decoder

DPCM PRINCIPLES: (a) Encoder / Decoder Schematics.

DPCM decoder

- In the decoder the parallel signal is converted into serial signal.
- Now the DPCM signal is added with the previously computed signal held in Register R.
- In DPCM the bit rate requirements (o) reduced to 56kbps in DPCM where in PCM is 64kbps.

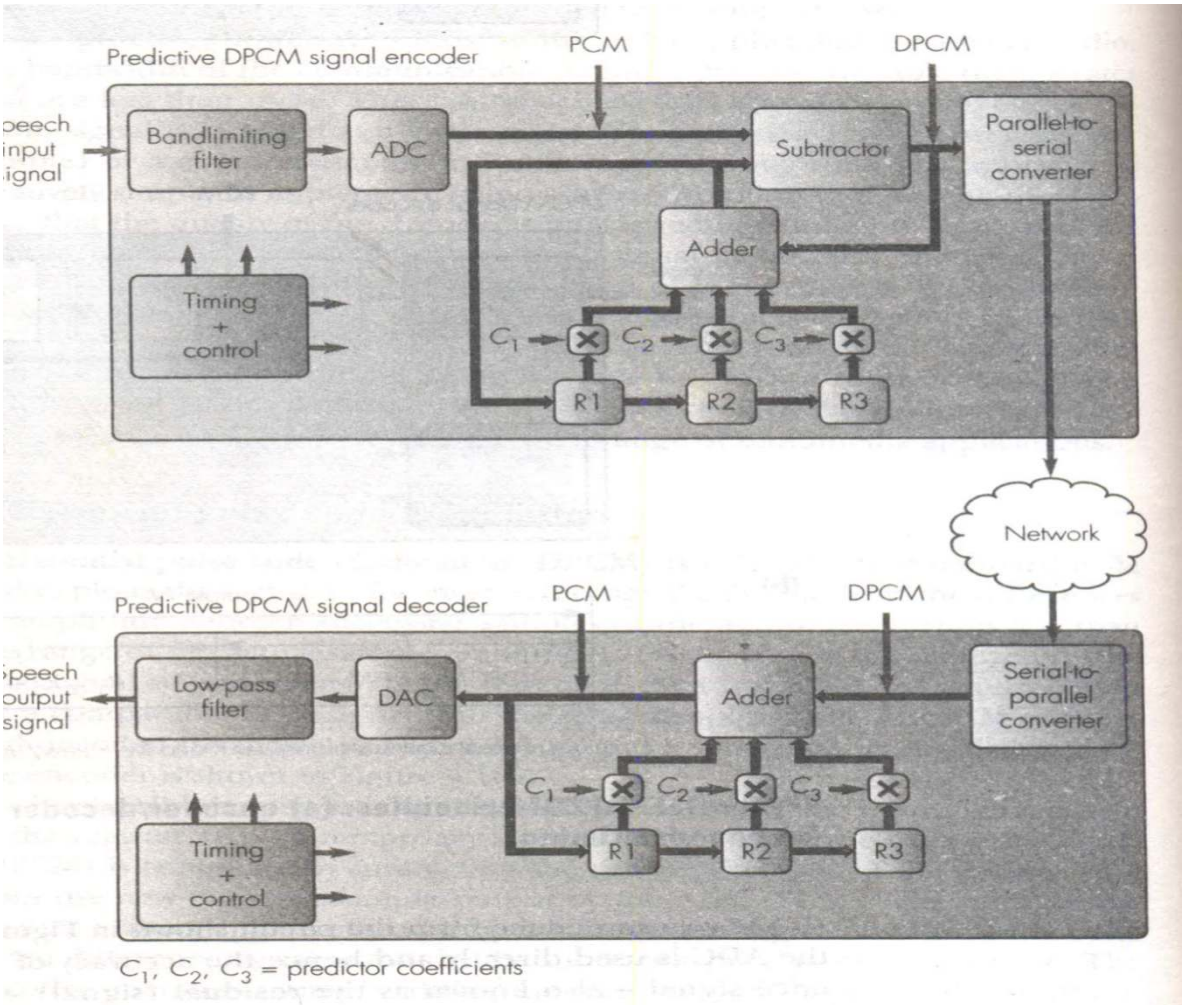


R0 – current content of register R

R1 – new / updated contents

Encoder Timing

- The output of ADC is used directly and hence the accuracy of each computed difference signal (residue signal) is determined by the accuracy of the previous signal value present in register.
- To overcome this predictive technique is used here the previous signal is predicted not only from current signal but also varying proportions of a number of the immediately preceding estimated signals. The proportions used are determined by predictor coefficients.



Third order predictive encoder and decoder

- The difference signal is computed by subtracting varying proportion of last three predicted values from current digitized value output from ADC.
- Ex, if predictor coefficient C1= 0.5 and C2 = C3 = 0.25 then the content of Reg R1 multiplied by 0.5, R2 and R3 with 0.25.
- Now all these values are added and subtracted from current digitized value output from ADC.
- Then content of R1 moved to R2, R2 moved to R3. The new predicted value is loaded to R1.
- The decoder operates in a similar way by adding same proportion of the last three PCM computed signals to the received DPCM signal.
- This technique approach a similar performance level to standard PCM by using only 6 bits for difference signal with bit rate 32 kbps.

3. ADAPTIVE DIFFERENTIAL PCM (ADPCM) PRINCIPLE:

- Additional savings in bandwidth or improved quality can be obtained by varying the number of bits used for the difference signal depending on its amplitude;

(ie) using fever bits to encode (and hence transmit) smaller difference values than for larger values.

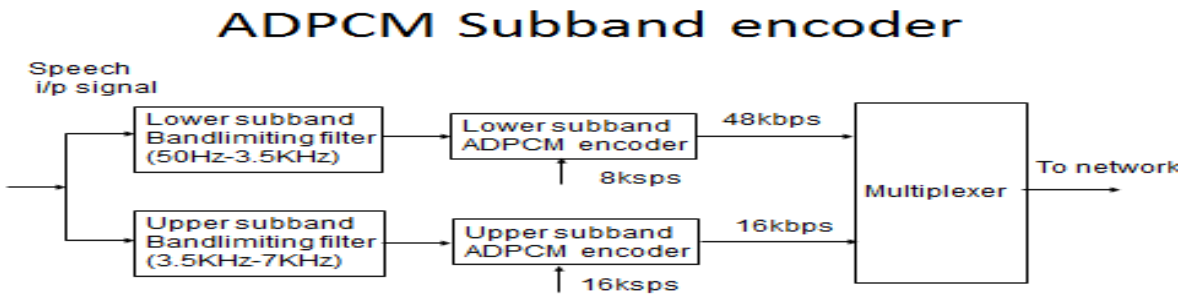
- An international standard for ADPCM is G.721 (ITUT – Recommended).
- Its principle is same as DPCM except an eight order predictor is used and the number of bits used for quantization of difference value is varied.

  It can be either, (ie) 6bits producing 32 kbps (better quality)

  → 5 bits producing 16 kbps (if layer band width is more important).
- Another standard for ADPCM is G.722.This provides better sound quality than G.721.
- The technique adapted in G.722 is sub band coding.
- The input signal band width is extended to be 50HZ through] KHZ [for PCM only 3.4 KHZ] Hence wider bandwidth produces a higher fidelity speech signal.

  Ex:- Conferencing

ADPCM Subband encoder:-



ADPCM Sub band decoder:-

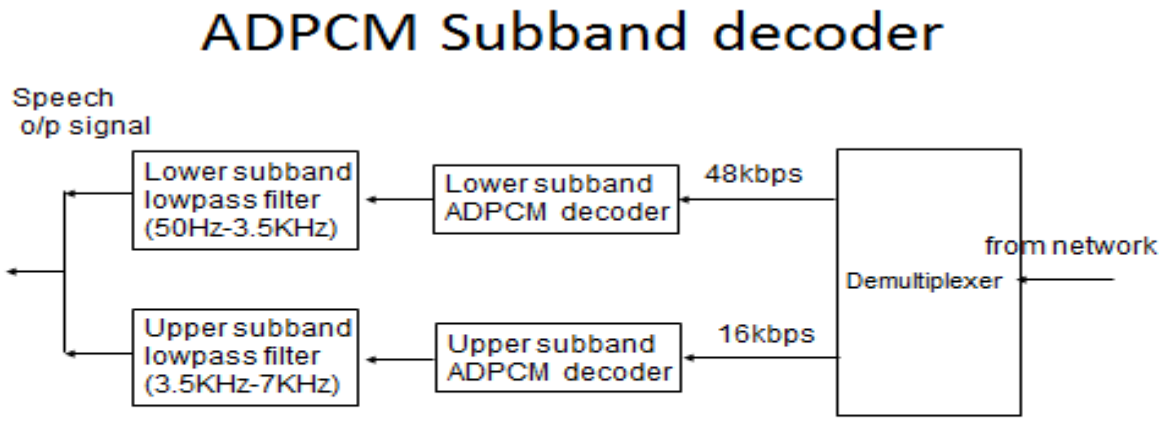


Fig 4.ADPCM Sub band encoder / Decoder:-

The audio Input signal passed through two fillers.

Lower Sub band Band limiting filler:

→ It passes only signal frequency of range 50HZ – 3.5 KHZ.

Upper Sub band Band limiting Filler:

→ It passes signal having frequency in range 3.5 KHZ. 7KHZ

The signal is divided into separate equal bandwidth signals known as

- Lower sub band signal
- Upper sub band signal

Each signal is sampled and encoded separately using ADPCM. The sampling rate of upper sub band signal is 16KSPS to allow high frequency components.

Advantages of two Sub bands:-

- Different bit rate can be used for both the side bands.
- Frequency components in lower sub band have high perceptual importance than higher sub band.
- The operating bit rate are 64, 56 or 48 kbps.
- If bit rate is 64 kbps, Lower sub band ADPCM encode at 48 kbps and upper sub band at 16 kbps.
- The two bit streams are then multiplexed to produce 64 kbps signal.
- The decodes in the receiver divide them back into two separate streams for decoding.

ITU – T Recommendation G.726 [Third standard for ADPCM]

- This also use sub band coding BW = 3.4KHZ
- Operating bit rate can be 40, 32, 24 or 16 kbps.

4. ADAPTIVE PREDICTIVE CODING:

Principle:-

- Higher level of compression can be achieved by making the predictor. Co efficient adaptive. So the predictor co efficient continuously changes.
- Optimum set of the prediction co efficient continuously vary since they are a function of the characteristics of the audio signal being digitized.
- To exploit this property, the input speech signal is divided into fixed time segments.
- And for each segment the currently prevailing characteristic are determined.
- The optimum set – of co efficient are then computed and they are used to predict more accurately the previous signal.
- This type of compression can reduce the band width requirement to 8 kbps.

Advantage:-

Higher level of compression

Disadvantage:-

Complexity is high.

5. LINEAR PREDICTIVE CODING:

Basic:
In Linear predictive coding the source simply analyse the audio wave form to determine certain understandable features it contain.

- They are then quantized and sent to the destination. At the receiver these features are used to regenerate the signal by using sound synthesizer.

Three features that determine perception of ear,

Pitch:-

This is closely related to frequency. This is more important because ear is More important because ear is more sensitive to frequency in range 2-5 KHZ.

Period:-

This is duration of the signal.

Loudness:-

This is determined by the amount of energy in the signal.

Vocal track excitation parameters are,

→ Voiced Sound

→ Unvoiced Sound

Voiced Sound : These are generated through the vocal chords Ex: Sound related to the Letters M1V! I.

Unvoiced Sound: With these the vocal chords are open.

Ex: Sound related to F and S.

Linear predictive coding signal Encoder / Decoder schematics: LPC Signal Encoder

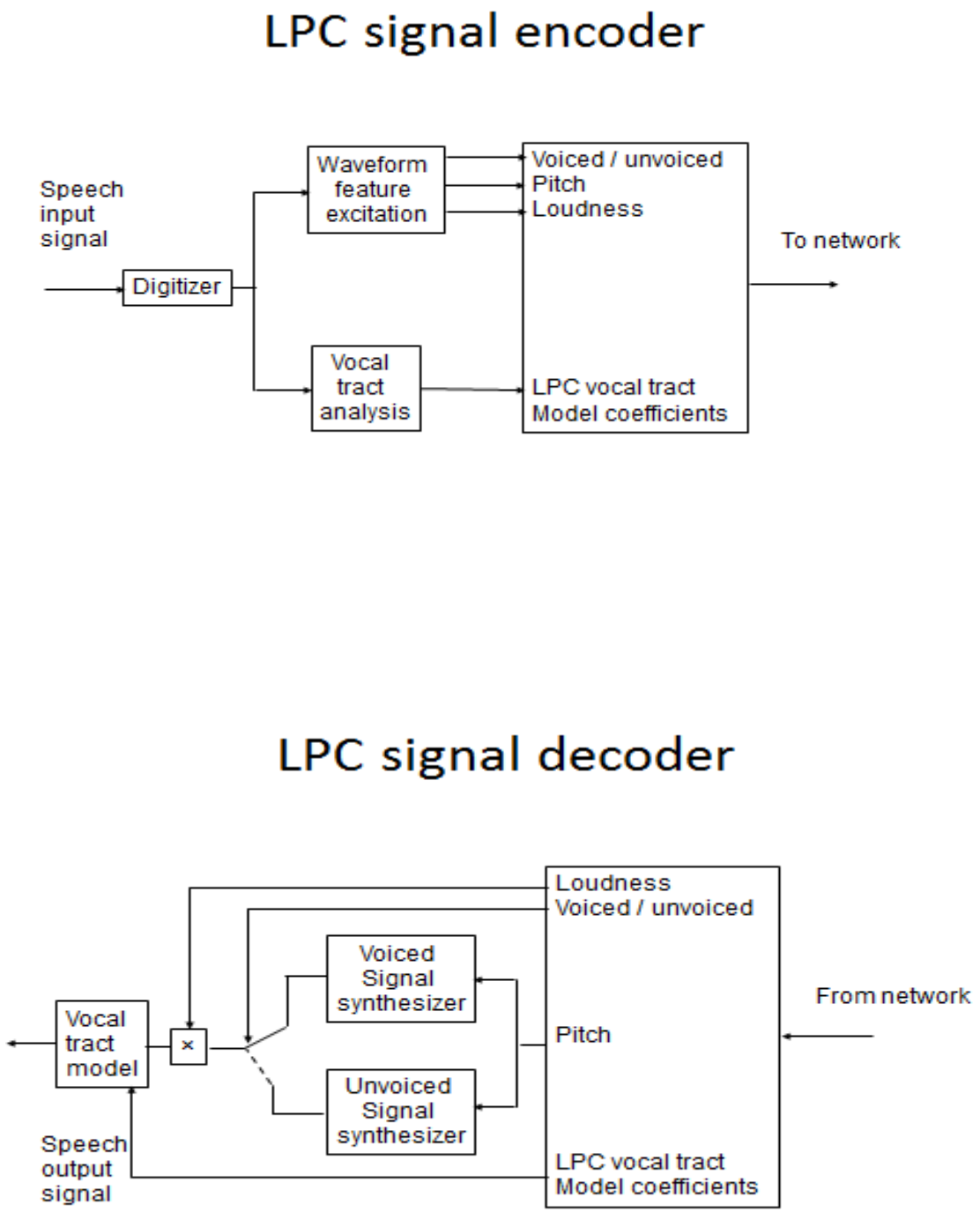


Fig 5 LPC encoder and decoder

LPC Encoder:-

- The input speech wave form is first sampled and quantized at a defined rate.
- This digitized signal is then analysed to determine the perceptual feature it have.

(Perceptual feature = pitch, loudness, period, voiced / unvoiced).

- The output of the encoder is a string of frames. Each frame contain fields for pitch and loudness [the period is determined by the sampling rate used],
- Whether the signal is voiced (or) unvoiced and the co efficient are transmitted across the N/W.

LPC Decoder:-

- Pitch determines voiced and unvoiced sound. Accordingly voiced (or) unvoiced sound is selected and filtered by vocal track model. Once speech segment is fallen for duration of 10ms.
- Some LPC encoders use upto 10 set of previous model co efficient to predict the O/P sound (LPC – 10). They use bit rate as low as 2.4 kbps and 1.2kbps.

Application:-

- The generated sound is more syntherric and is used in military applications where Band width is more important.

6. CODE EXCITED LPC: (CELP)

- The synthesizers used in most LPC decoders are based on a very basic model of the vocal track. A more sophisticated version of this known as code excited linear prediction (CELP) model, and is an example of enhanced excitation (LPC) model.
- These are used in application where amount of band width available is limited.
- In the CELP model, the standard set audio segments are stored as wave form templates.

- The encodes as well as decodes stores the same set of wave form templates. It is known as template code books.
- Each digitized audio segment is compared with the waveform templates in the code book.
- Each digitized audio segment is compared with the wave form templates in the code books.
- The matching template is selected from the code book. It is differentially encodes of and transmitted.
- At the receiver the differentially encoded code word selects a matching template. This produces more natural sounding speech.

This type of encoder having delay, when each block of digitized sample is analysed at the encoder and when the speech signal is reconstructed at the decoder.

- The combined delay value of encoder & decoder is known as the coders processing delay.
- Before analysing the speech signal, the block of samples are to be stored in memory. The time to accumulate the block of samples is known as algorithm delay.
- Look ahead delay:- If the samples from next successive block is considered these algorithm delay change to look ahead delay.
- These delay occur in addition to the end to end delay transmission delay.

However processing delay is an important parameter which determines suitability of a coder for specific application.

Ex:- Conventional telephone

- In conventional telephone application a low delay coder is required since a large delay can affect the flow of a conversation.

→ In an interactive applications that involves the O/P of speech stored in a file.

Ex. A delay of several seconds before the speech starts to be O/P I is often acceptable and hence coder delay less important other parameter of the coder have tobe consider, complexity and perceived quality of O/P speech.

Table of CELP – based standards.

Standard	Bit rate	Total coder delay	Application domain
G.728	16 kbps	0.625ms	low bit rate Telephony
G.729	8kbps	25ms	Telephony in cellular W/w's
G.729(A)	8kbps	25ms	(DSVD)
G.723.1	5.3/6.3kbps	67.5ms	Voice & Internet Telephony

7. PERCEPTUAL CODING:-

- Perceptual coders are special coders designed for the compression of a general audio such as that associated with a digital television broadcast.
- The perceptual coding techniques are mainly based on audio perception mechanism.
- The perceptual coders use psychoacoustic model which exploit the limitation of ear. (ie) the strong signal mask the weak signal.

Principle of perceptual coding:-

→ The human ear can hear very small sound when there is complete silence. But if other big sounds are present, then human ear cannot listen very small sound. These characteristics of human ear are used in perceptual coding. The strong signal reduces level of sensitivity of the ear to other signal which has low strength.

→ This effect is called Frequency Masking.

Sensitivity of the ear:-

→ Dynamic range of a sound signal is defined by the ratio of max Amplitude of the signal and minimum amplitude of the signal.

$$\text{Sensitivity of ear} = \frac{\text{Loudest sound it can hear}}{\text{Quietest sound it can hear}}$$

Sensitivity of ear is around about 96db.

- The sensitivity of the ear varies with the frequency of signal.
- Assume that a single frequency signal is present at a time, the perceptual threshold of the ear (minimum level of sensitivity) as function of frequency is shown below fig 6 a.

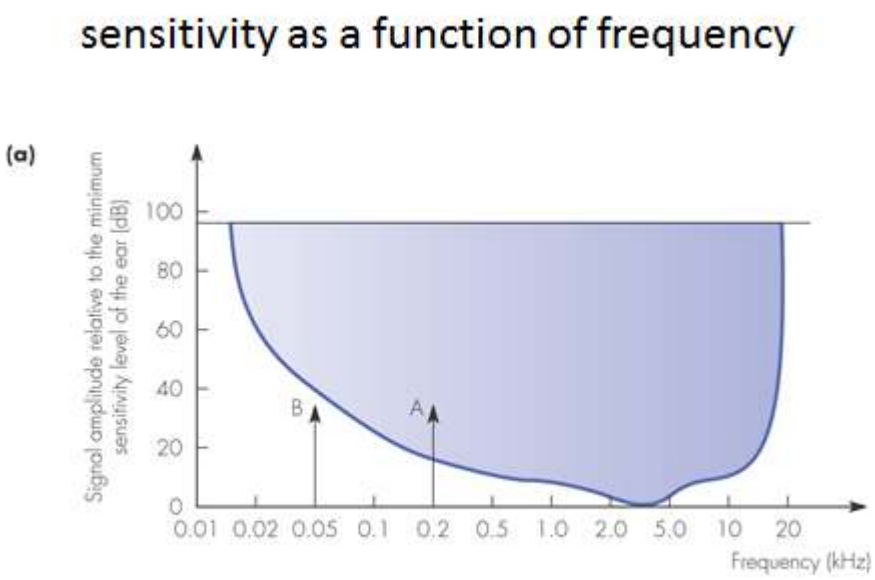


Fig 6 a) Sensitivity as a function of frequency

- The sensitivity is very good between 2.5 KHZ.
- In figure although the two signals A and B have the same relative amplitude.
- A signal would be heard since it is above the hearing threshold. Where signal B cannot be heard.

Frequency Masking:

- When multiple frequencies are present in an audio signal the sensitivity of the ear changes with the relative amplitude of the signals.
- In the following example signal B is large in amplitude than A. So B can be heard.
- Due to high amplitude of B, the basic sensitivity curve change as shown below fig 6 (b) and cannot be heard even though it is above the hearing threshold.

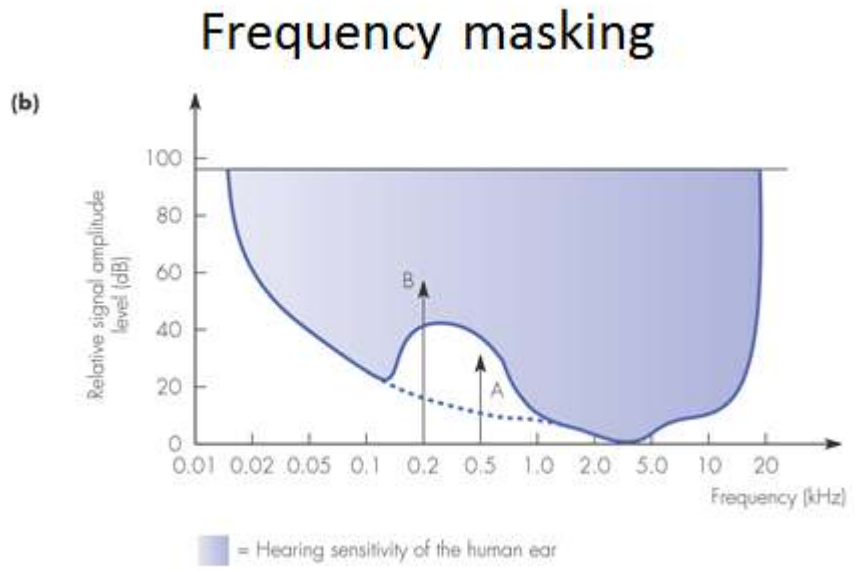
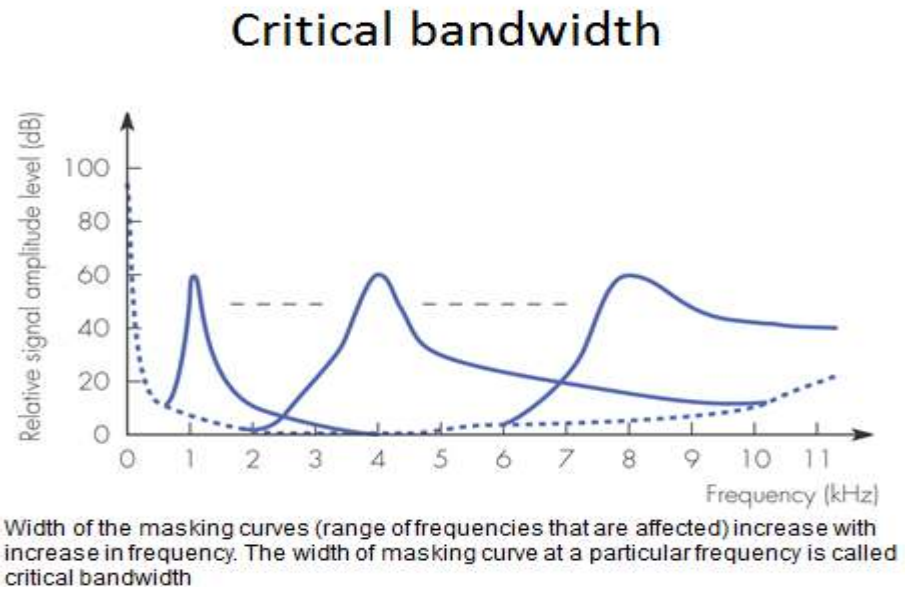


Fig 6 b (Frequency Masking)

Masking Effect:

- The width of the masking curve (ie) the range of frequencies that are affected increase with increase in frequency. The width of each curve at a particular signal level is known as critical band width.

The masking affect is shown below fig:-



Width of the masking curves (range of frequencies that are affected) increase with increase in frequency. The width of masking curve at a particular frequency is called critical bandwidth

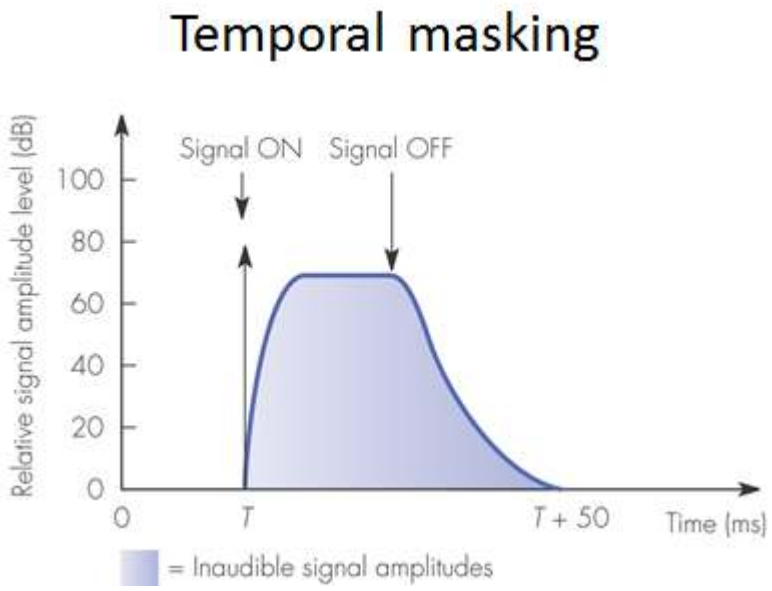
- For frequencies less than 500 HZ the C1 band width is constant and is about 100 HZ.
- For frequencies greater than 500 HZ the critical BW increases linearly in multiple of 100 HZ.

  Ex: For a signal of 1 KHZ (2*500 HZ) critical BW is (2*100 HZ) while at 5 v KHZ (10*500 HZ) it is about 1000 (10*100 HZ).

- Hence the magnitude of frequency components that make up an audio sound can be determined, once it determined frequencies that will be masked and do not therefore need to be transmitted.

Temporal Masking:-

When ear hears the loud signal certain time has to be passed before it hears quieter sound. This is called Temporal Masking.



Temporal masking caused by a loud signal

- After the loud sound ceases it takes a short period of time for the signal amplitude to decay.
- During this time signals, whose amplitudes are less than the decay envelop will not be heard and hence need not be transmitted.
- In order to make full use of this phenomenon it is necessary to process the i/p audio wave form over a time period (ie) comparable with that associated with temporal masking.

UNIT – III

IMAGE AND VIDEO COMPRESSION

VIDEO COMPRESSION

Video with sound find application in a number of applications like.

Inter personal: - Video telephony & video conferencing

Interactive:- Access to stored video in various forms

Entertainment:- Digital television and movie / video on demand.

The quality of video used in these applications vary and is determined by the digitization format and frame refreshing rate.

Digitization format: - It defines the sampling rate that is used for the Luminance Y and two chrominance Cb & Cr and their relative frame position.

Video Compression principles:-

- Video is simply a group of digitized pictures. Video is also referred to as moving pictures.
- Video source can be compressed using JPEG algorithm. The compression ratio of JPEG is between 10:1 and 20:1 which is not the sufficient compression rate.

Frame Types:-

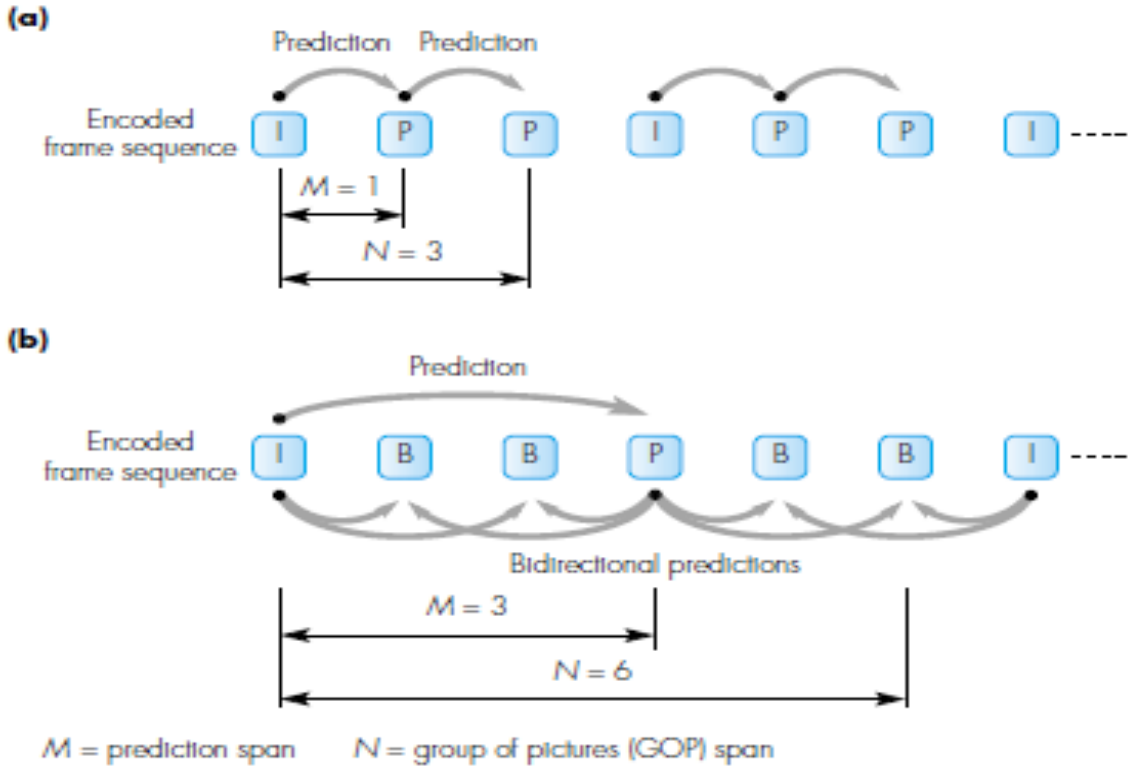
There are three types of frames,

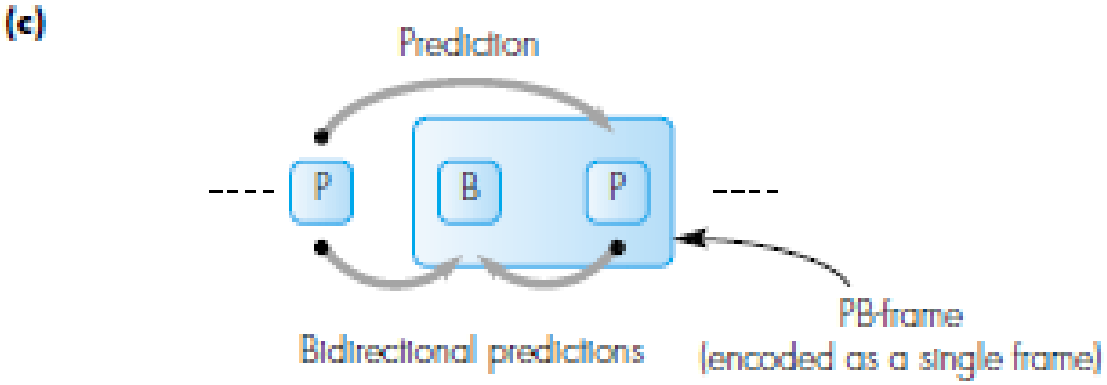
- Independently encoded frames – I frames.
- Predicted frames – P frames
- Bidirectional frames – B frames.

→ I frames also known as (or) called as Intra coded frames.

→ P- frames and bidirectional (01) B – frames are known as predictive frames. These frames are also called as Interpolation frames.

A frame sequence involving I,P and B frames is shown below





I – Frames

- I – frames are encoded without reference to any other frames.
- Each frame is treated as a separate picture and the Y1 cb and C1 matrices are encoded independently using JPEG Algorithm.
- Compression obtained with I – frames is relatively small.

- So this used for first frame (new scene in Movie).
- frames should be presented at regular interval in output stream, because if 'I' frame is corrupted, the predicted frames will be corrupted the no. Of frame / pictures between successive I – frames is called Group of Pictures (GOP).
- It is given by the symbol N and typical values for N are from 3 through to 12.

P– Frames

- The encoding of P- frame is relative to the content of either a preceding P- frames (or) an I – frames.
- P – frames are encoded using a combination of motion estimation and motion compensation and hence high level of compression can be achieved.
- The number of P – frames between successive I – frames is limited to avoid propagation of error from P- frames if any.
- The number of frames between a P – frame and the immediately preceding I (or) P – frame is called the prediction span (Rep M = 1-3).

B – Frames

- Motion estimation involves comparing small segments of two consecutive frames for finding the difference between them.
- To minimize time needed for it only few neighbouring segments are selected.
- For slowly varying picture it is works by for fast varying (moving) picture like movie the search may be outside the selected segments.
- To allow for this possibility in applications such as movies in addition to P – frames a second type of prediction is used (ie) B – frames.
- B frame contents are predicted using search regions in post and future frames (preceding and succeeding I (or) P frames).
- It provide better motion estimation for fast moving picture as well as for objects moving infront and behind other object.
- Here the encoding delay is high, due to waiting for future I (or) P frame.
- B – frames provide highest level of compression and no propagation error occur.

Uncoded frame sequence:-

I B B P B B P B B I

Reordered Sequence:-

I P B B P B B I B B

D - Frames

- Used in Movie / Video on demand application.
- These applications require decompression at much higher speed. D – frame support these functions, the encoded video stream also contains D – frames which are inserted at regular intervals throughout the stream.

PB – Frames

- Here the two neighbouring P & B frames and encoded as if they were a single frame.

MOTION ESTIMATION AND COMPENSATION:-

- The encode content of both P and B frames are predicted by estimating any motion that has taken place between the frame being encoded and the preceding I (or) P – frame.
- In this case of B – frame it depend on the successor I (or) P – frametoo.

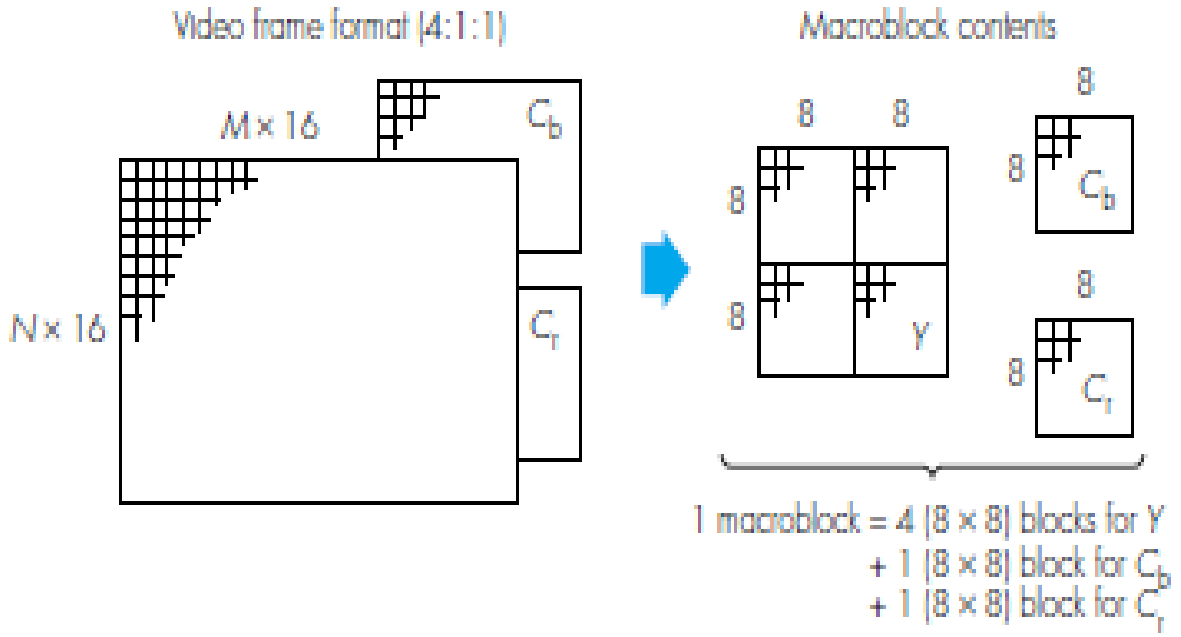
Steps in encoding P – frame

- The digitized Y – matrix for each frame are divided into 2 D matrix of 16*16 pixels known as macro block.
- The macro block consist of four blocks of 8*8 pixels. The macro block is in RGB form. The RGB macro block is then converted into Y Cr Cb macro block.

Y – Luminance signal,

Cr, Cb →Chrominance signal

Macro Structure:-

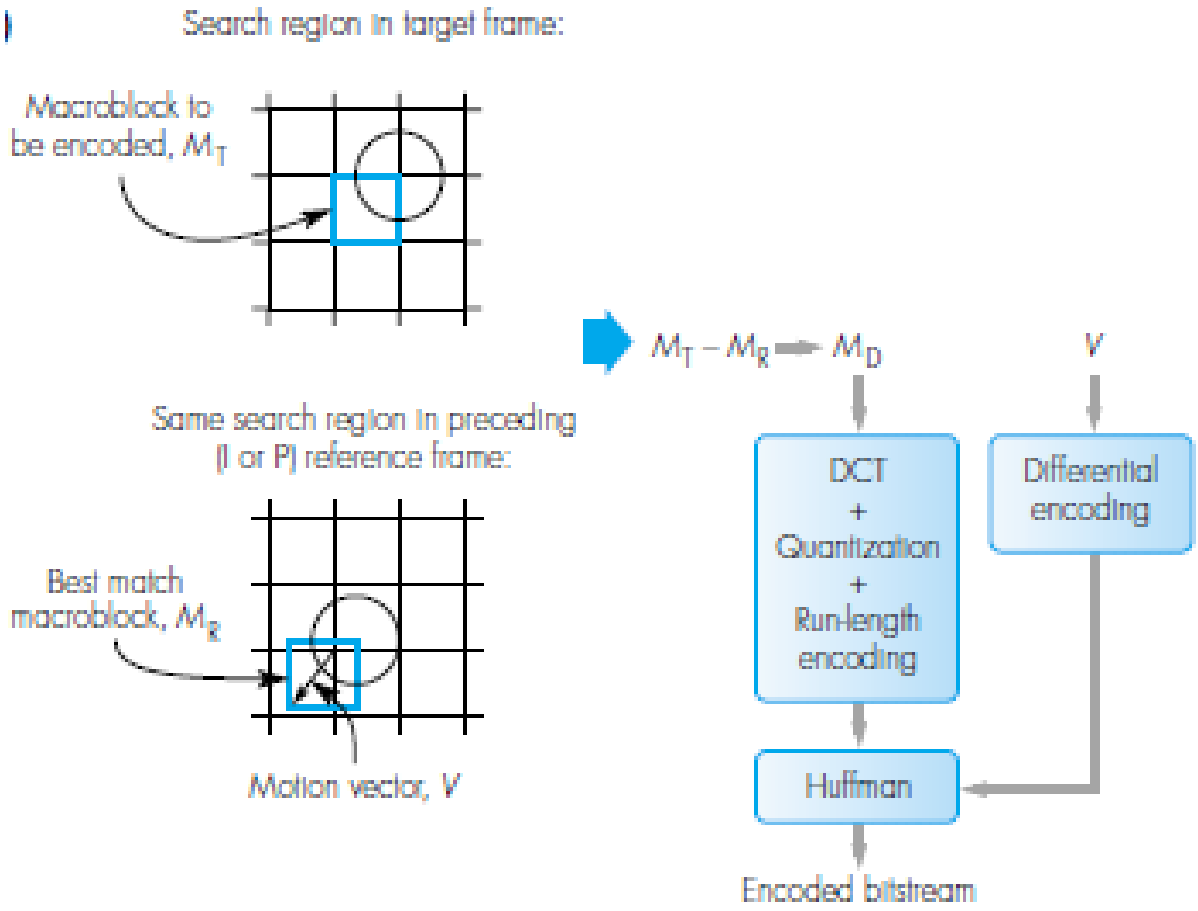


- Then Y macro block has 4,(8*8) size block, Cb, Cr have one (8*8) blocks. The chrominance signals Cr and Cb are sub samples.

- For identification purpose each macro block has an address associated with it. DCT block size is 8*8.

So a macro block has 4 DCT for Luminance and one for each chrominance.

- To encode a P – frame, the content of each macro block in the frame known as the target frame are compared on a pixel b pixel basis with the content of the corresponding macro block in the preceding I (or) P – frame (ie) reference frame.
- If both the contents match then address of the macro block is encoded.
- If there is no match between target frame and preceding frame, then search is extended around macro blocks in reference frame. Then search is extended around macro blocks in reference frame.
- Only the content of Y – matrix are used in the search. The search is said to be found if mean of absolute error between reference frame is below threshold.
- All the possible macro blocks in the search area in reference frame are compared with target frame.
- And if match is found motion vector and prediction error are to be encoded.



encoding procedure

- The difference of all 6(8*8) pixel blocks of the best matching macro block and the macro block to be coded are transformed using a two dimensional DCT.
- For further data reduction blocks that only have DCT co efficient with all values of zero are not processed further.

- These are stored using 6 bit values which are added to the encoded data stream.
- In the next step, a run length encoding and the determination of a variable length code is applied. Here DPCM encoding is used. The result is again transformed using a table leading to a variable length encoded word.

Motion Vector:-

Motion vector indicates the offset (x,y) of the target frame macro block from reference from macro block.

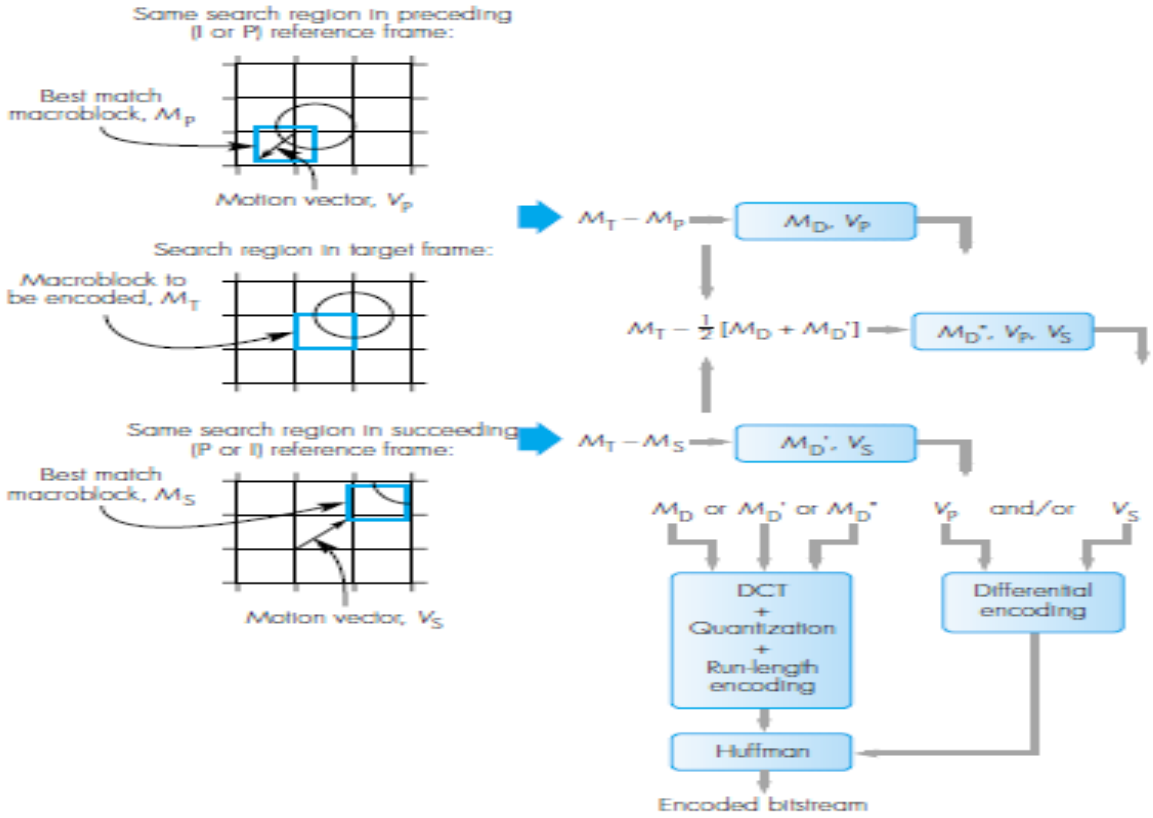
Prediction Error:-

- Prediction error consist of error matrices for each Y1 cb and cr. It contains the pixel to pixel difference B/W the matching macro block of target and reference frames.

- If match is not found with any of the macro block in the reference frame. Then macro block of the target frame is encoded independently in the same way as macro blocks in I – frames.

B – Frame encoding procedure:-

- To encode a B – frame, any motion is estimated with reference to both the immediately preceding I (or) P – frame and the immediately succeeding P (or) I frame. The general scheme is given by below.

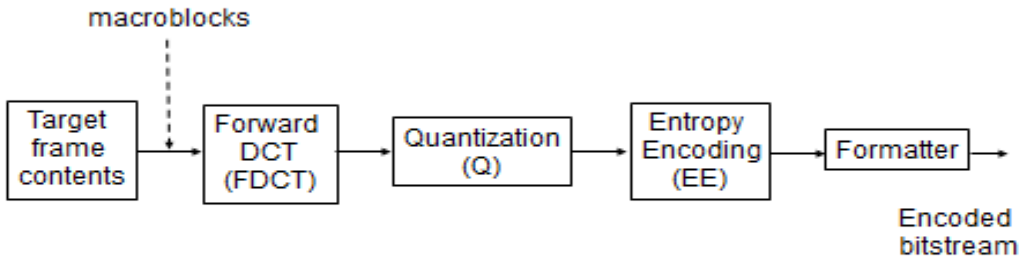


B – Frame Encoding procedure

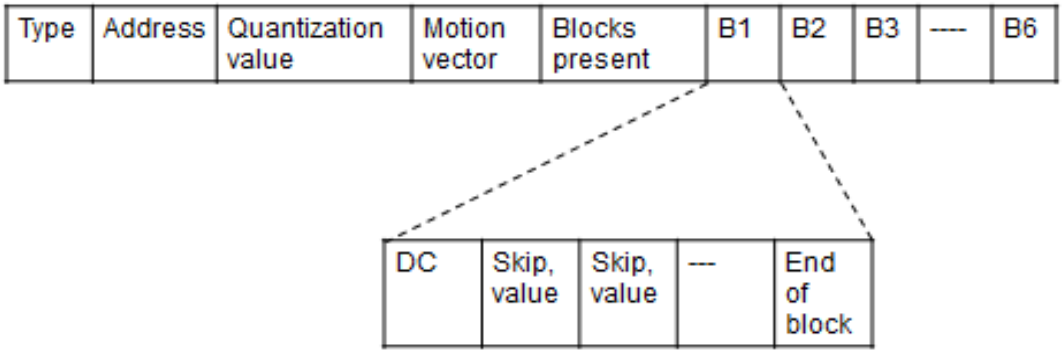
- The motion vector and different matrices are computed using first the preceding frame as the reference and then the succeeding frame as reference.
- The third motion vector and set of difference matrices are then computed using the target and near of the two other predicted set of values.
- The set with the lowest set of difference matrices is then chosen and these are encoded in the same way as for P – Frames.

The schematic diagram showing the essential units associated with the encoding of I, P and B frames is given below fig 13.

I-frames implementation schematics

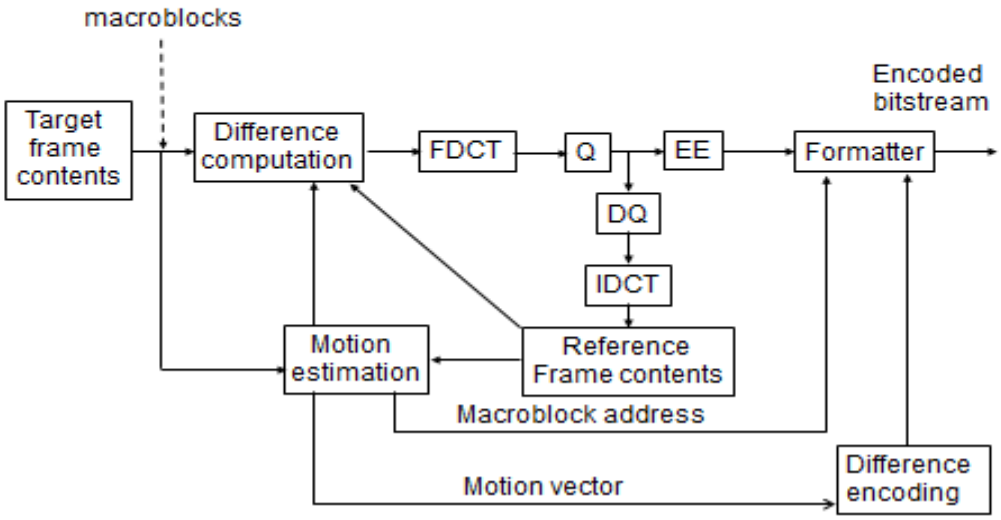


Example macroblock encoded bitstream format

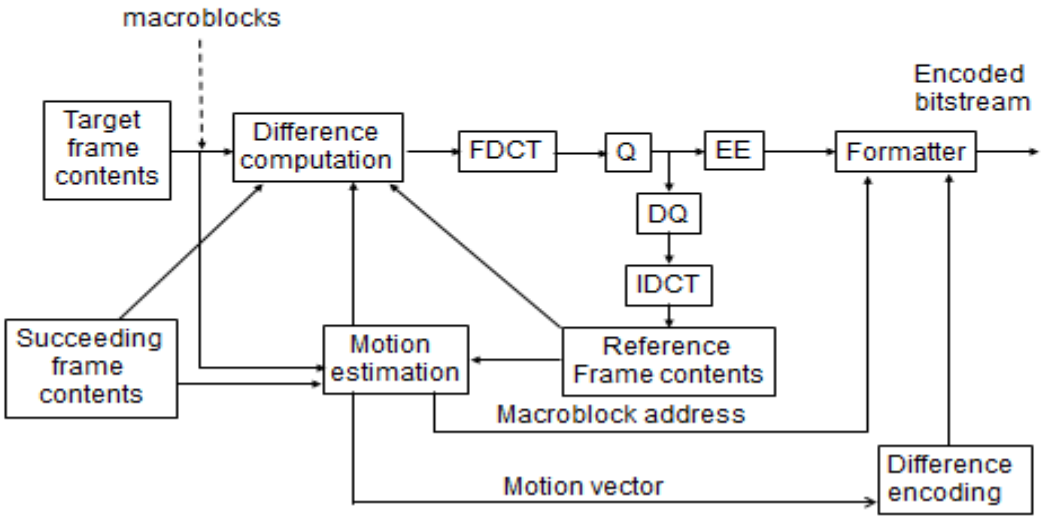


Type: I or P or B frames
Address: identifies the location of the macroblock in the frame
Quantization value: threshhold value used for quantization
Motion vector: encoded vector
Blocks present: indicates which of the six 8 x 8 pixel blocks that make up the macroblock are present

P-frames implementation schematics



B-frames implementation schematics



H 261:-

- ITU – T has been defined the video compression standard for the provision of video telephony and video conferencing services over an ISDN. That standard has been named as H.261.
- It is assumed that the N/W offers transmission channels of multiples of 64 kbps. The standard also known as P*64. Where P can be 1 through 30.

Under H.261, we have two digitization formats.

CIF – Common Intermediate Format. (used for video conferencing).

QCIF - Quarter Common Intermediate format.

(used for video telephony)

- In both the digital formats each frame divided into macro blocks of 16*16 pixels for compression.
- The horizontal resolution is reduced from 360 to 352 pixels to produce an integral no of 22 macro blocks.
- Both the formats use sub sampling of the two chrominance signals at half the rate used for the Luminance signal.

The spatial resolution of each format is,

	Luminance	**Chrominance**
CIF	Y=352*288	cb = cr =176*144
QCIF	Y=176*144	cb = cr 88*72

Non – interfaced scanning is used with a frame refresh rate 30fps for the CIF and either 15 (or) 7.5 fps for the QCIF.

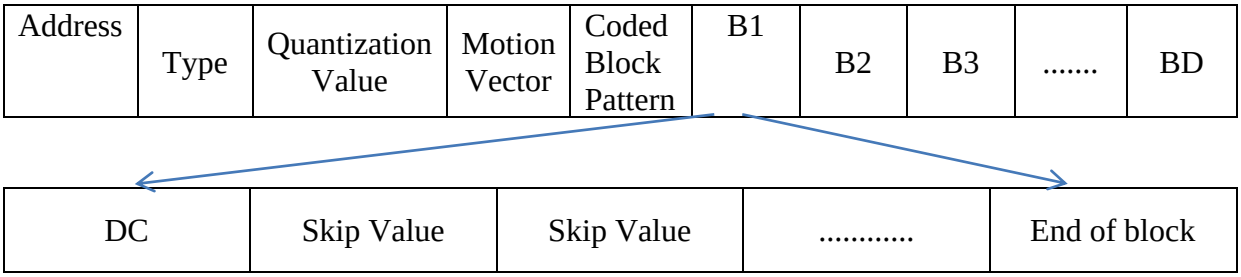
H.261 Macro block format:

H.261 uses only I and P.- frames with three P - frames B/W each pair of I – frames.

→ The encoding of each of the six 8** pixel blocks that make up each macro block in both I and P – frames.

(8*8) Pixel of 6 blocks; 4 blocks doe Luminance

1 for cb and 1 for cr

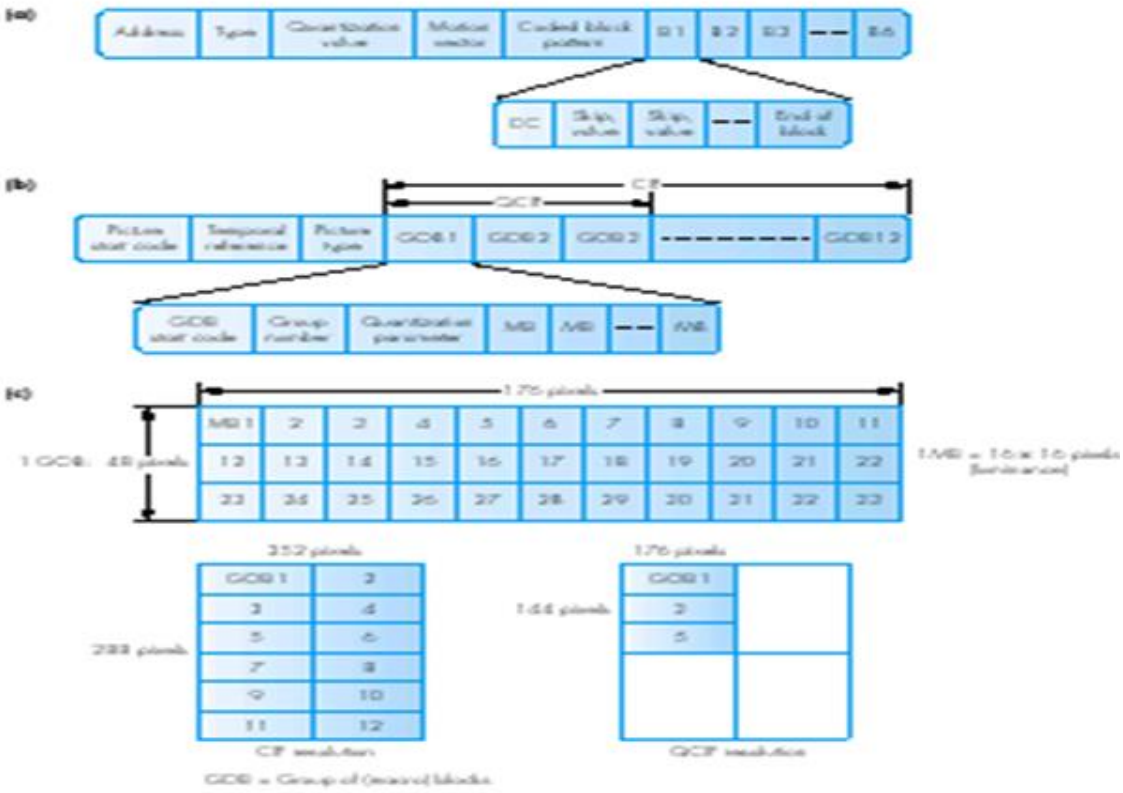


H.261 MACRO BLOCK FORMAT

Macro block Format Fields:

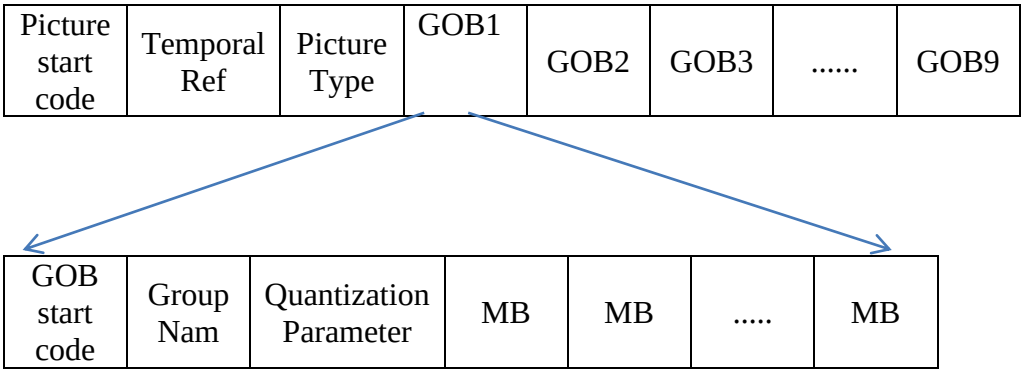
1. **Address:**
   Each macro block has an address associated with it for identification purposes.
2. **Type:**
   Type field indicates whether the macro block has been encoded in dependently intra coded (or) interceded.
3. **Quantization Value:**
   It is the threshold value that has been used to quantize all the DCT co efficient in the macro block.
4. **Motion Vector:**
   Motion vector is the encoded vector if one is present.
5. **Coded block pattern:**
   It indicates which of the 6(8*8) pixel blocks that make up the macro block are present – if λ, λfor those present. The JPEG encoded DCT co eff are given in each block.

Figure 15 H.261 encoding formats: (a) macroblock format; (b) frame/picture format; (c) GOB structure.



CIF

QCIF



H.261 Frame / Picture Formats.

Frame Format Field:-

Picture start code:

The start of each New video format picture is indicated by the picture start codes.

Temporal Reference:-

This field is a time stamp to enable the decoder to synchronize each video block with an associated audio block containing the same time stamp.

Picture Type:-

It indicates whether theframe is an I (or) P- Frames.

GOB:-

Group of Blocks – encoding operation is carried out on individual macro block.

The above Fig shows 11*3 macro blocks, size GOB -12 in the case of CIF (2*6) & 3 in the case of QCIF (1*3).

176 Pixels

MBI	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22

MBI	23	24	25	26	27	28	29	30	31	32

352 pixel

GOBI	2
3	4
5	6
7	8
9	10
11	12

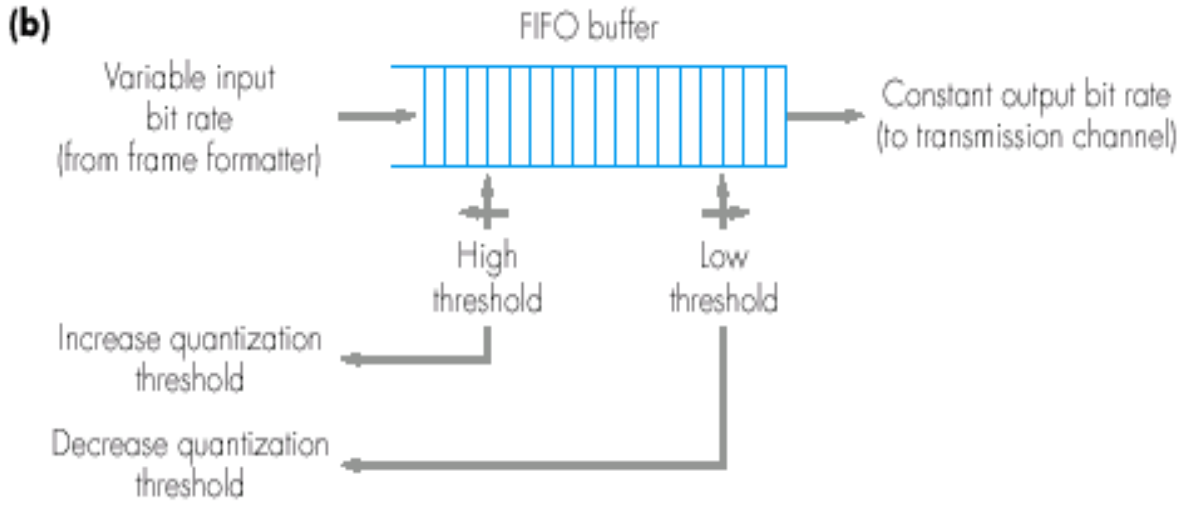
CIF – resolution GOB – Group of Macro block

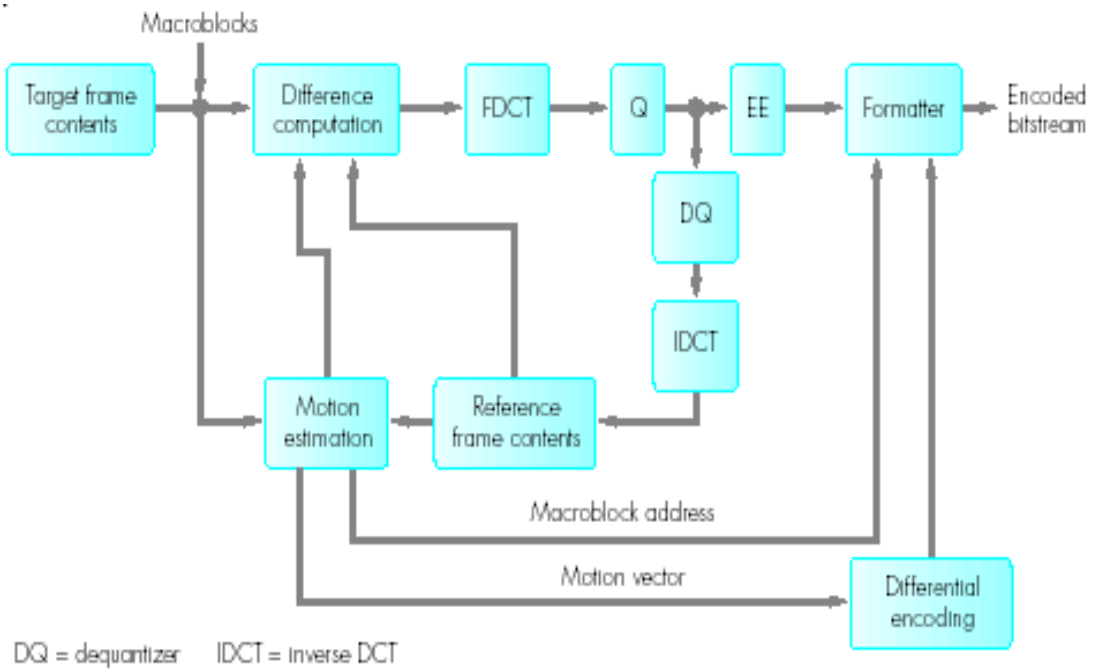
GOB structure

FiFO Buffer Operation:-

- To convert the variable bit rate produced by the basic encodes into a constant bit rate optimize the use of band width.
- The transmission BW of H.261 is fixed 64 kbps. (or) multiple of this.
- The optimization is achieved by FIFO (First –in-first out buffer).

The buffer operation is shown in Fig 16(b). The general scheme of H.261 video encoder is shown.





H.261 VIDEO ENCODER

- The O/P bit rate produced by the encoder is determined by the quantization threshold values.
- The higher the threshold, the lower the accuracy and hence the lower is the O/P bit rate.
- Same compression logic is used for macro blocks in video encodes, it is possible to obtain a constant O/P bit rate from the encoder by dynamically varying the quantization threshold used. This is the role of FIFO buffer.

FIFO the order the O/P from a FIFO buffer is same as that of I/P.

If the i/P data rate to the buffer is exceeds the O/P data rate will start to fill .Similarly data rate falls below the O/P data rate it decrease.

To exploit this property two threshold used.

- Lower threshold :- The amount of information in the buffer is continuously monitored and the contents. Fall below the low threshold (Fig 1)
- Then the quantization threshold is reduced, thereby increasing the O/P rate, from the encoder conversely should the contents increase beyond the higher threshold.
- Then the quantization threshold is increased in order to reduce the O/P rate from the encoder.

H.263:-(Video compression standard)

- ITU-7 defined the special video compression standard for use in a range of video applications over wireless and PSTNs, name as H.263.

- The applications include video telephony, video conferencing, security survillence and so on of all which require O/P of the video encoder to be transmitted across the N/W connection as it is output by the encoder.
- PSTNs operates in analog mode and to transmit a digital signal over modem. Typical maximum bit rates ranges from 28.8 kbps to 56 kbps.
- H.263 video encoder based on that used in H.261 standard.

1. Digitization Formats:

Under H.263 two digitization formats.

→ QCIF

→ Sub-ACIF (S-QCF).

As like an H.261 each frame divided into 16*16 pixels for compression.

→ Horizontal resolution reduced from 180*175 pixels to produce 11macro blocks.

Spatial resolution of the two formats are,

	Luminance	Chrominance
QCIF	Y = 176*144	cb = cr = 88*72
S – QCIF	Y = 128*96	cb = cr = 64*68

In H.263 standard progressive scanning is used refreshing rate 15 or 75 fps.

2. Frame Types:-

- H.263 uses all I-P_B Frames to obtain the higher levels of compression.
- PB – Frames handles the high frame rates. A PB frame comprises a B – Frame and immediately succeeding P – frame.
- In a macro block the encoded information of both these frames is interleaved.
- Hence at the decodes the macro block for P – frame is reconstructed first using received information relating to the P – macro block and then preceding P- Frame.

3. Unrestricted motion vectors:-

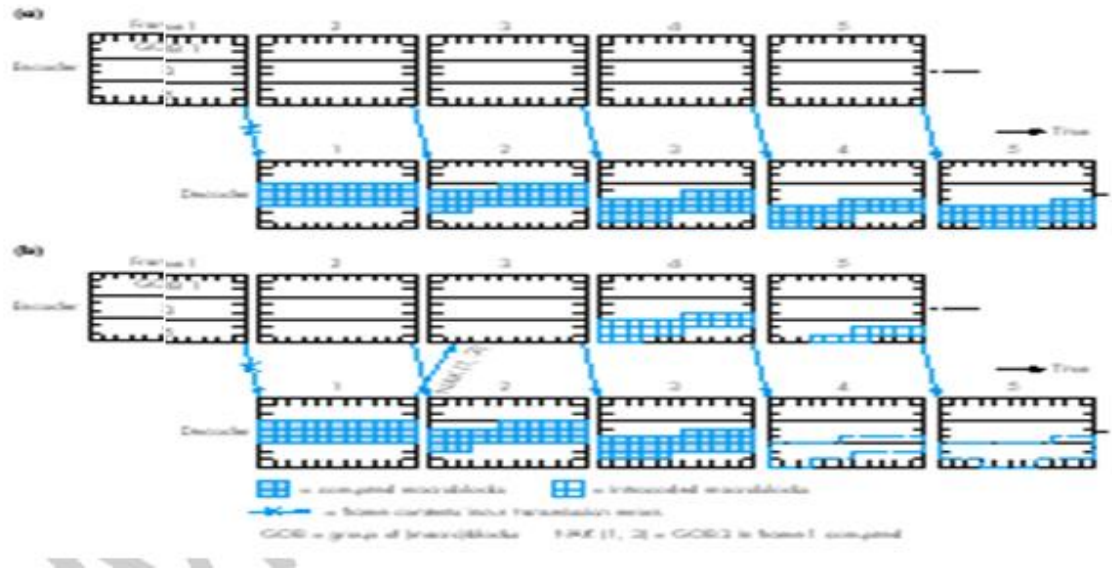
- The motion vectors associated with predicted macro blocks are normally restricted to a defined area in the ref frame of the macro block being encoded.
- Because of this restricted area portion of a potential close match macro block fall outside of the frame boundry.
- To overcome this limitation the edge pixels themselves are used instead and should be resulting macro blocks produce a close match then motion vector.

4. Error Recovery:-

- H.263 standard for wireless & PSTN N/WS with this type of N/W high probability that the transmission errors will be present in the bit stream received by the decoder.

- When a string of error - free frames is received followed by a short burst of errors which corrupt a string of macro blocks in a frame.
- In practice it is not possible to identify the specific macro block that are corrupted rather than group of blocks (GOB) are in error.
- When error ina GOB is detected the decoder skips the remaining macro blocks in the affected GOB.
- And searches for the unique resynchrotion marker (start code) at the head of next GOB.
- With digitization formats such as QCIF which has only 3 GOBs/Frame, resulting error can be annoying to the viewer.

Figure 17 H.263 error tracking scheme: (a) example error propagation; (b) same example with error tracking applied.



The typical effect is shown diagram forms below:-

(a) H.263 error recovery / example error

(b) error tracking scheme with example:

In Fig, the initial error occurs in one GOB position, it rapidly spreads to other neighbouring GOBs.

- H.263 standard, different schemes are employed to minimized the effect of errors are,

1. Error tracking
2. Independent segment decoding
3. Reference picture selection

1. Error Tracking:

In application such as video telephony, a two way communication channel is required for the exchange of compressed audio/video information generated by the code in each terminal.

In such band of transmission, errors are detected in a no of ways:-

- One (or) more out of range motion vectors
- One (or) more out of range DCT co.eff.
- One (or) more invalid (variable – length code word)
- On excessive co. efficient with in a macro block.

In the error tracking scheme,

→ The encoder retains the error prediction information for all GOBs in each of the most recently transmitted.

→ When an error is detected return channel is used by the decoder to send a negative acknowledgement (NAK) message back to the encoder in the source code containing both the frame number and the location of the GOB in the frame that is in error.

→ The encoder then uses the error prediction information relating to this GOB to identify the MB in those GOBs frames are affected.

→ Hence on receipt of NAK (1,3) it is assumed that the encoder has predicted and retained error prediction information relating to Frame 1.

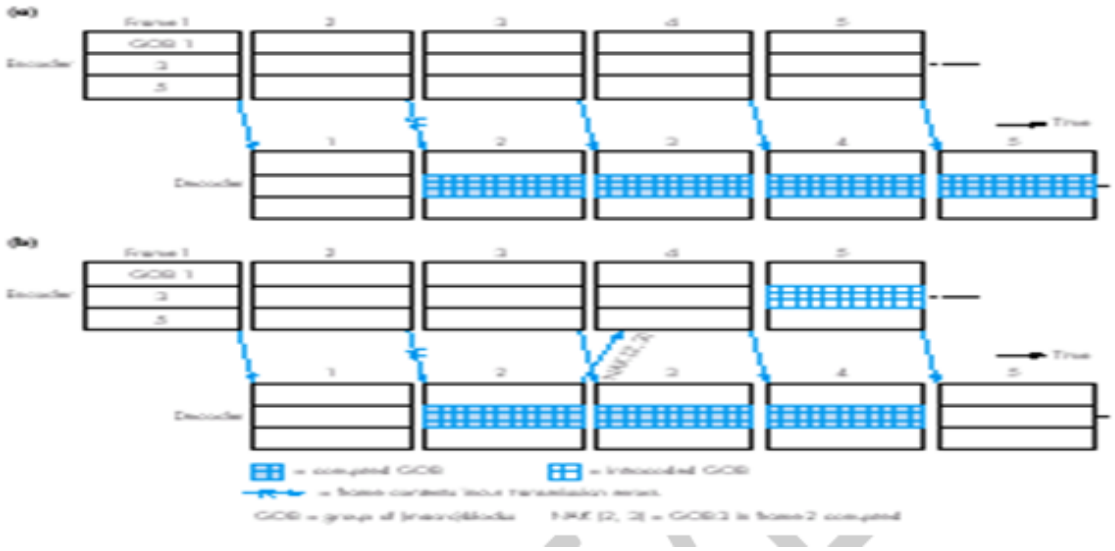
2. Independent segment decoding:

The aim of this scheme is not to overcome errors that occur within a GOB but rather to prevent these errors from affecting neighbouring GOBs in succeeding frames.

- To achieve this each GOB is treated as a separate sub video which is independent of the other GOBs in the Frame.

The operation schematic is shown in below Fig 18 (a) and (b)

Figure 18 Independent segment decoding: (a) effect of a GOB being corrupted; (b) when used with error tracking.



effect of a GOB being corrupted Fig 18 b) when used with error tracking

- In Fig (18 a) although when an error in GOB occur the same GOB in each successive frame is affected until a new intracoded GOB is sent by the encoder neighbouring GOBs are not affected.
- The limitation of this scheme is that of the efficiency of motion estimation and compensation in the vertical direction is reduced significantly.

3. Reference picture selection:

This scheme is to the error tracking scheme is as much as it endeavours to stop errors propagating by the decoder returning Ack messages when an error in a GOB is detected.

This scheme can be operated into different modes.

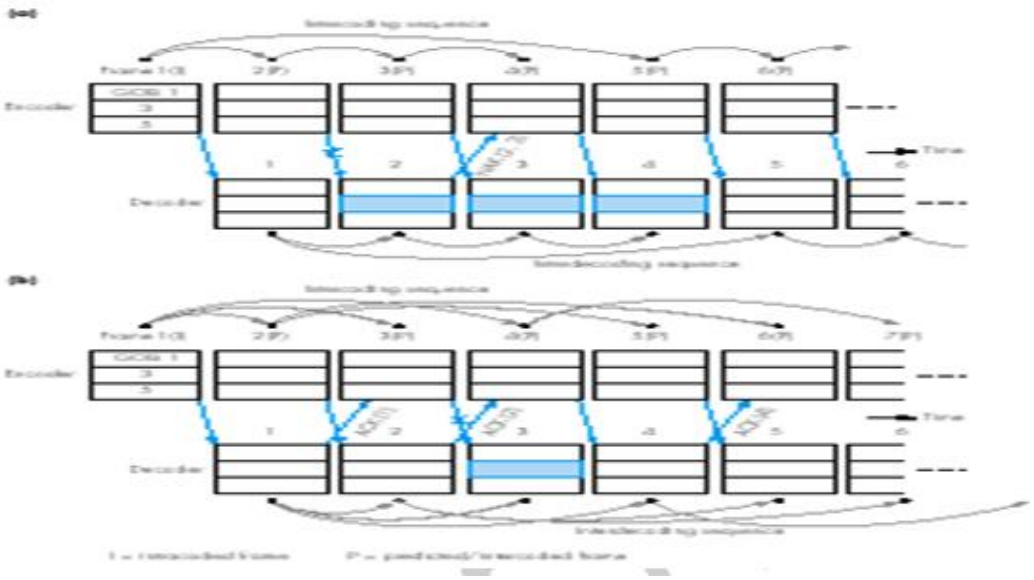
1. NAK mode (Negative Acknowledgement)
2. ACK mode (Acknowledgement)

- **NAK Mode:-**

In this mode (ref Fig 19 a) only GOBs in error are signated by the decoder returning a NAK message.

- In Fig (4) when the NAK relating to Frame 2 is received the encodes selects GOB 3 of frame 1 as the ref to encode GOB 3 of next frame - frame 5..

Figure 19 Reference picture selection with independent segment decoding: (a) NAK mode; (b) ACK mode.



a) MAK mode b) ACK mode

With this scheme the GOB in error will propagate for a no of frames, the number being determined by the round – trip delay of the communication channel, (ie) the time delay b/w NAK being sent by the decoder and an interceded frame derived from the initial I frame being received.

ACK mode:

- Ref Fig (5) with this mode all frames received without errors are acknowledged by the decoder returning an ACK message.
- Only frames that have been acknowledged are used as reference frames.
- In Fig, the lack of an ACK for frame 3 means, that frame 2 must be used to encode frame B in addition to frame 5.
- At this point the ACK for frame 4 is received and hence the encoder then uses this to encode frame 7.
- This ACK mode performs best when the round trip delay of the communication channel is short.

MPEG 1, 2, 4:- (Video compression standard):

MPEG 1:-

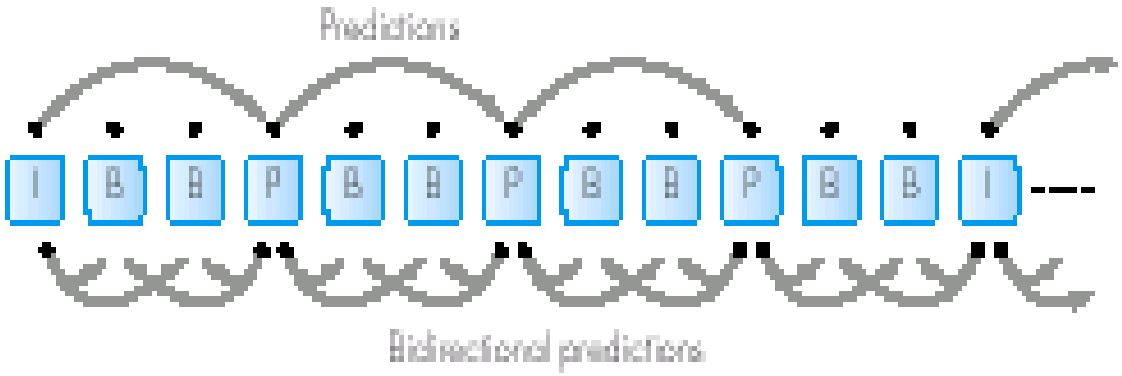
- The MPEG – 1 video standard uses a similar video compression technique used in H.261 but MPEG-1 uses the SIF (Source Intermediate Format) digitization format.
- Hence the two chrominance signals are sub sampled at half the rate of the Luminance signal. The spatial resolution for the two types of video source are,

NTSC	PAL	
352*240	352*288	Y – Luminance
176*120	176*144	cb = cr = 176*144

In MPEG – 1 progressive scanning is used with a refresh rate of 30 HZ(NTSC) and 25HZ (PAL).

- ✓ The MPEG standard allows the use of I frames only, I & P frames only or I,P – B frames.
- ✓ No D – Frame are supported for MPEG standards.
- ✓ In the case of MPEG – 1, I – frames must be used for the various random access functions associated with VCRs.

The below figure shows the standard frame sequence used in both NTSC & PAL system.



MPEG -1 Example Frame

The example sequence is,

I B B P B B P B B P B B I.....

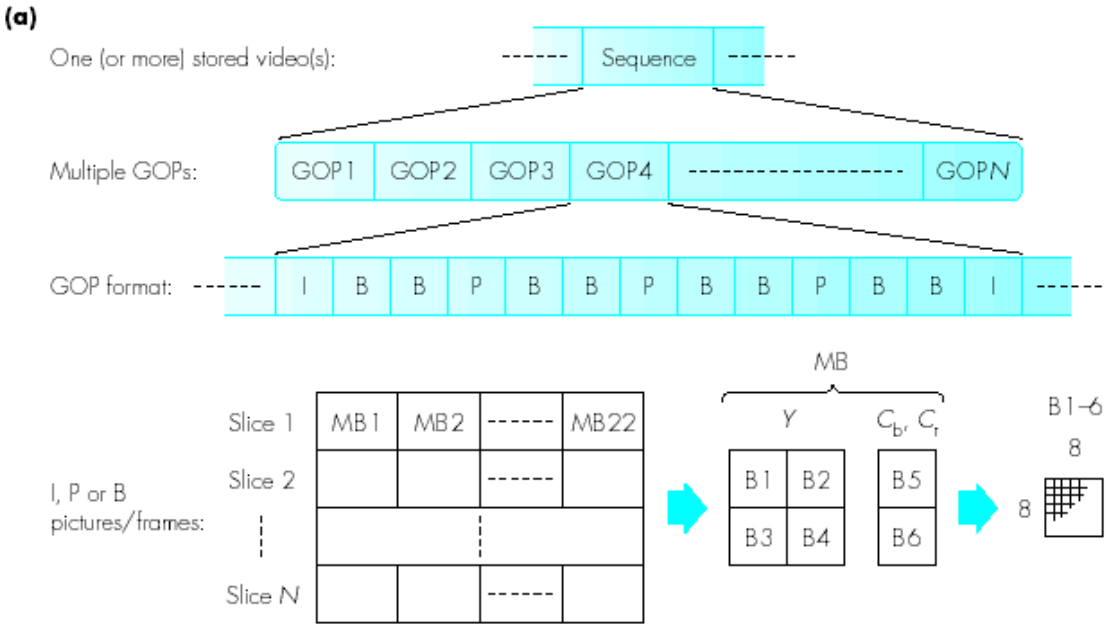
- ✓ H.261 standard compression techniques is used in MPEG – 1.
  Hence each macro block is made up of 16*16 pixels is the Y – Plane & 8*8 pixels in the cb&cr plane.

Difference B/W H.261 & MPEG:

1. The temporal references time stamps can be inserted within a frame to enable the decoder to resynchronize more quickly in the event of on (or) more corrupted (or) missing macro blocks.
    - The no. Of MBs b/w 2 time stamps is known as slice and a slice can comprise from1 to max no of MBs in a frame.
    - Typically slice is equal to 22 which no of macro blocks in a line.
2. The difference arises because of the introduction at B – Frames, which increases the time interval b/w I & P frames.

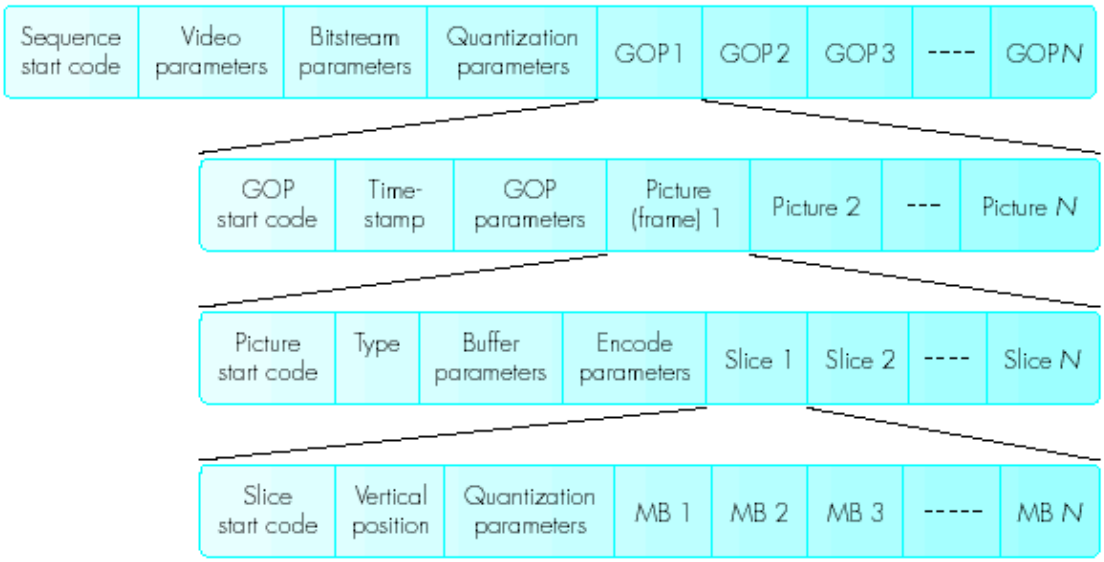
   Typical compression ratio vary from about 10.1 for I – frames 20.1 for P – frames 50.1 for B – Frames.

   MPEG – 1 video structure – composition bit stream.



At the top level, the complete compressed video is known as a sequence which in turn consists of group of pictures. (GOBs)

- ✓ Each comprising a string of I, P or B Pictures / Frames in the defined sequences.
- ✓ Each picture / frame is made up of N slices each of which comprises multiple macro block and so on down to 8*8 pixel block.
- ✓ Hence in order for the decoder to decompress the received bit stream. The format of bit stream is given by below in Fig (22).
- ✓ Each picture / frame is made up of N slices each of which comprises multiple macro block and so down to 8*8 pixel block.
- ✓ Hence in order for the decoder to decompress the received bit stream. The format of bit stream is given by below in Fig (22).



MPEG – 1 VIDEO BIT STEAM FORMAT

- **Sequence start code:-**

The start of a sequence is indicated by this.

This is followed by three parameters each of which applies to the complete video sequence.

- Video Parameter specifies the screen size & aspect ratio.
- Bit stream parameter indicates the bit rate and size of memory / frame buffers that are required.
- The quantization parameter contains the content of the quantization tables that are to be used for the various F/P types.

- ✓ These are followed by the encoded video stream which is in form of a string of GOPs.
- ✓ Each GOP (IBBPBB.....) is separated by a (GOP) short code.
- ✓ Type :- I, P (or) B frame type.
- ✓ Buffer Parameter:- Mention Buffer status.
- ✓ Encoder Parameter:- Sags about resolution used for motion vector.
- ✓ Slice:- Group of macro blocks
- ✓ Slice start code:- Each slice is separated by slice start code.
- ✓ Vertical position:- Define scan line apply to slices.
- ✓ Quantization Parameter:- Specify the threshold value

MPEG – 2

MPEG -2 supports four levels of video resolution low, main, high 1440, high each targeted at a particular application.

- ✓ These have been defined that the four levels and five profiles collectively form a 2 – D table which acts as a frame work for all standards activities with MPEG – 2.

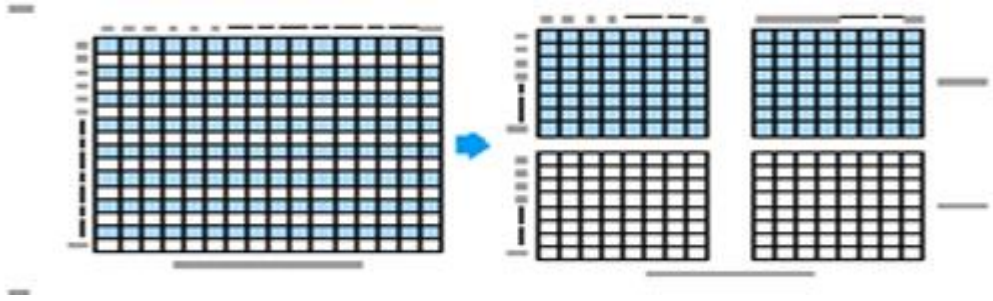
MP @ ML (Main profile at main level)

- ✓ MP @ ML standard is for digital television broadcasting.
- ✓ Hence interlaced scanning used with a resulting frame refresh rate of either 30 HZ (NTSC) (or) 25HZ (PAL).
- ✓ The 4:2:0 digitization format is used with resolution of either 720*480 pixels at 30HZ (or) 720*576 pixels at 25 HZ.
- ✓ The O/P bit rate from the system multiplexer can range from 4MBPS through to 15 Mbps
- ✓ For a frame refresh rate – temporal resolution of 30/25 HZ the corresponding field refresh rate is 60 / 50 HZ.
- ✓ Fig (23) will produce a higher compression ratio owing to the shorter time interval b/w successive fields.
- ✓ If there is little movement the frame mode can be used the longer time interval b/w successive frames. Fig (c)

MPEG – standard allows either mode to be used.

For live sporting event → Field Mode

Studio based program → Frame Mode



MPEG -2 DCT block derivation with I frames:

Field Mode.

In MPEG – 2 video coding scheme in filler to used in MPEG – 1 the main difference resulting from the use of Interlaced scanning instead of progressive scanning.

Interlaced Fields are shown below: Fig

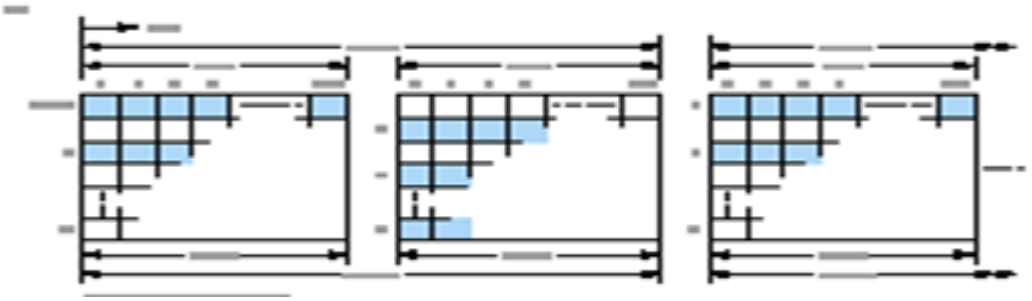


Fig MPEG -2 Effect of interlaced scanned.

→ In Fig alternative lines are present in each field.

- **Two Modes:-**

1. Field Mode

2. Frame Mode

As in Fig (23) & Fig (24), two alternatives are possible depending on whether the DCT blocks are derived from the lines in a field – field mode or the lines in a frame – frame mode.

MPEG – 2 DCT block derivation with I frames.

Frame Mode:-

Mode of Motion estimation:-

Field Mode:-

→The motion vector for each macro block is computed using the search window around the corresponding macro blocks in the immediate preceding (I or P) field both P & B frames.

Frame Mode:-

→A macro block in an odd field is encoded relative to that in the preceding / succeeding odd fields & similarly for the macro blocks in even fields.

→The motion vectors relates to the amount of movement that has taken place time to scan two fields.

Mixed Mode:-

In the mixed mode, the motion vector for both field and frame modes are computed and the one with the smallest value is selected.

MPEG – 2 in HDTV:-

- There are three standards associated with HDTV.
  1. Advanced television (ATV)
  2. Digital video broadcast (DVB) – Europe
  3. Multiple sub Nyquist sampling encoding (MUSE) – Japan & Asia.

1.) ATV – Aspect ratio, 16/9:
Resolution = 1280*720
It has lower resolution format

2.) DVB – Aspect ratio 4/3
Resolution of 1152 (1080 visible) lines / frames, 1440 S/LIN
PAL digitization format 720*576.

3.) MUSE –
→ Aspect ratio = 16/9.
→ Digitization format 1920 samples / line and 1035 Lines / Frame.
→ Video compression algorithm used in MPEHL.

MPEG – 4:

MPEG -4 standard mainly used with interactive multimedia applications over internet and the various types of entertainment N/W's.

→ The standard contains features access a video sequence but also to access & manipulate the individual / elements.

Scene Composition:-

The main difference b/w MPEG -4 and the other standards, it has a no. Of content based functionalities.

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

- Each AVO is defined in the form of more video object /or audio objects.
- Each AVO has sub subjects.
- Each AVO has a separate object description.
- The language used for to describe and modify objects is called the Binary format for scenes (BIFs).
- This has commands to delete an object and in case of video objects change its shape, appearance &color.
- Audio objects have a filler set of commands to change it volume.
- At higher level, the composition of a scene interns of the AVOs it contains is defined in a separate scene descriptor.

AUDIO & VIDEO COMPRESSION:-

- The audio associated with an AVO is compressed using algorithms, the selection of algorithms depends on the available bit rate of the Transmission channel.
- G723.1 (CEIP) algorithm used for interactive multimedia applnover the internet and video telephony over the wireless N/W's, PSTNs through dolby AC -3 Or MPEG – Layer 2 for interactive TV application over entertainment N/Ws.

An overview structure of the encoder associated with the audio / video of a frame / scene is shown by Fig (25) & essential features of each VOP encodes are shown in Fig (26).

Fig (25) MPEG – 4 Encodes / Decoder – schematics:-

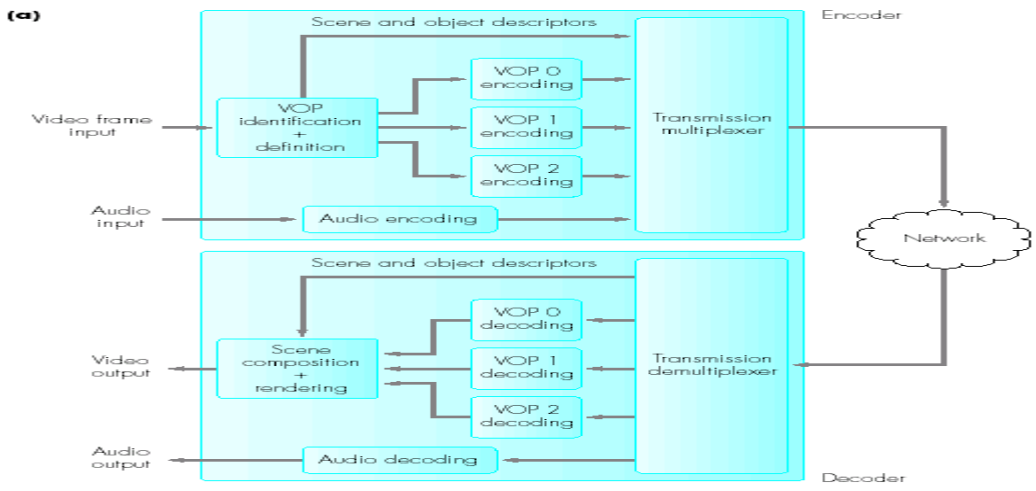
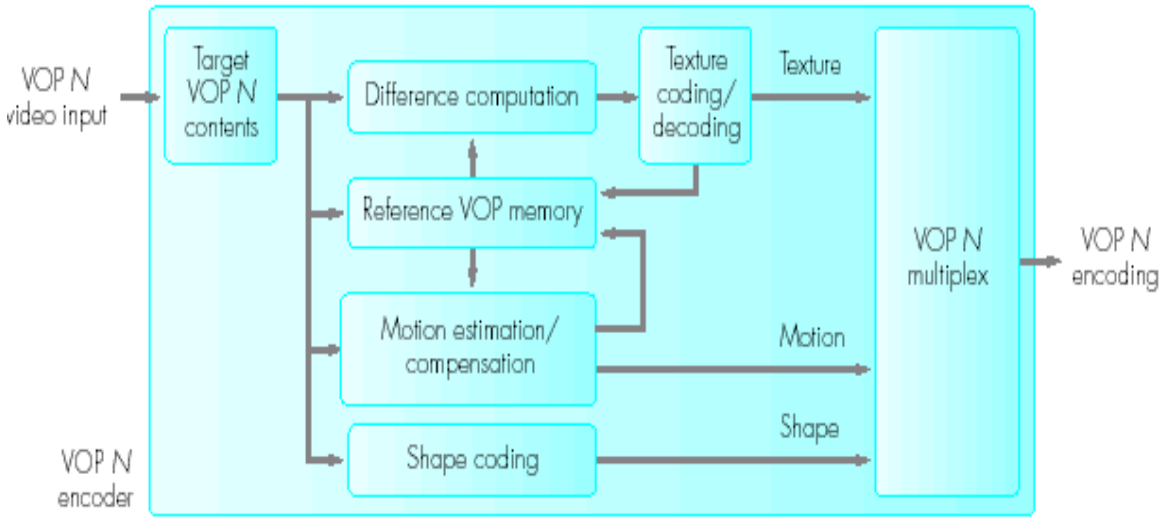


Fig (26) MPEG – 4 VOP encoder schematic



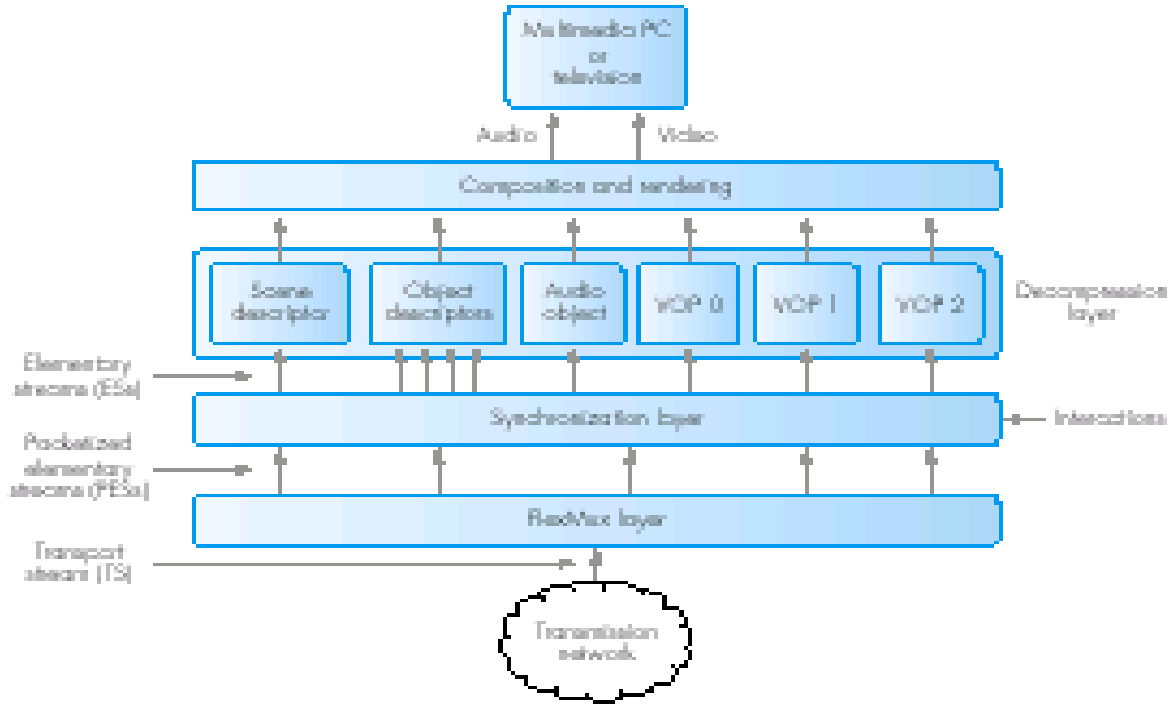
- As shown in Fig. First each VOP is identified and defined and is then encoded separately.
- In practice it involves identifying regions of a frame that have similar properties such as color, telephony or brightness.
- Each of the resulting object shapes as then bounded by a rectangle to form the related VOP.
- The VOP(s) which move often occupy only a small portion of scene / frame, the bit rate of the multiplexed video stream is much lower than that obtained with the other standards.
- Particular VOP has a no of advanced coding algorithm.

TRANSMSSION FORMAT:-

All the information relating to a frame / scene encoded in MPEG – 4 is transmitted over a N/W in the form of a transport stream (Is) consisting of a multiplexed stream of pack zed elementary streams (PES).

- In most cases, MPEG -4 transport stream uses the 188 – byte packet format since this helps interworking with the encoded bit streams associated with the MPEG ½ standards.

MPEG – 4 DECODER SCHEMATICS:-



- The compressed audio and video information relating to each AVO in the scene is called an elementary system.
- Each PES packet contains a type field in the packet header and this is used by the Flexmax layer to identify & route the PEC to the related synchronization block in the synchronization layer.
- The compressed audio & video associated with each AVO is carried in a separate ES.
- Associated with each object descriptor is an elementary stream descriptor (ESD). This is used by the synchronization layer to route each ES to its related decoder.
- The output from the decoders associated with each AVO making up a frame, together with the related object scene descriptor information is then passed to the composition and rendering object.

IMAGE COMPRESSION

For image lossy as well as lossless compression methods can be applied. Images are of two types.

(a) Graphical images

(b) Digitized images

Graphics interchange format (GIF)

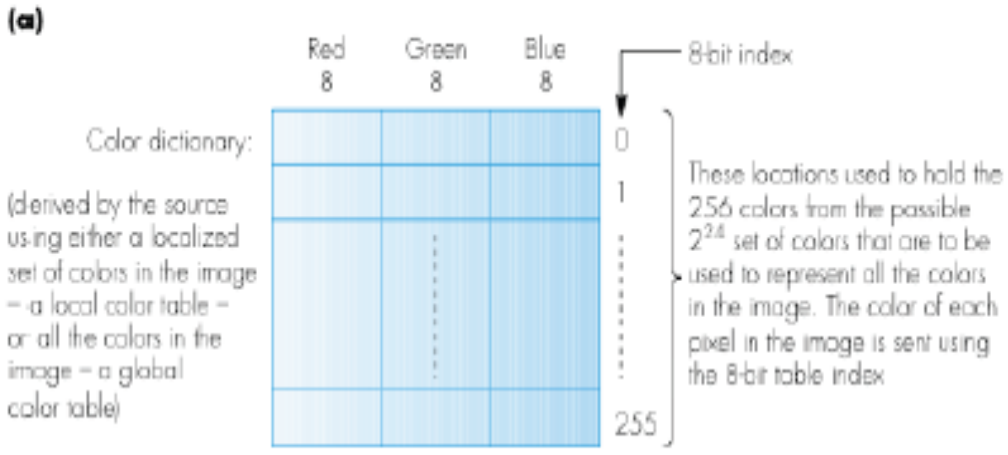
The graphics interchange format (GIF)is used extensively with the internet for the representation and compression of graphical images.

Although colour images comprising 24-bit pixels are supported-8 bits each for R,G and B-256 colours from the original set of 224 colour that match most closely those used in the original image. The resulting table of colors therefore consists of256 entries each of which contains a 24bit color valve.

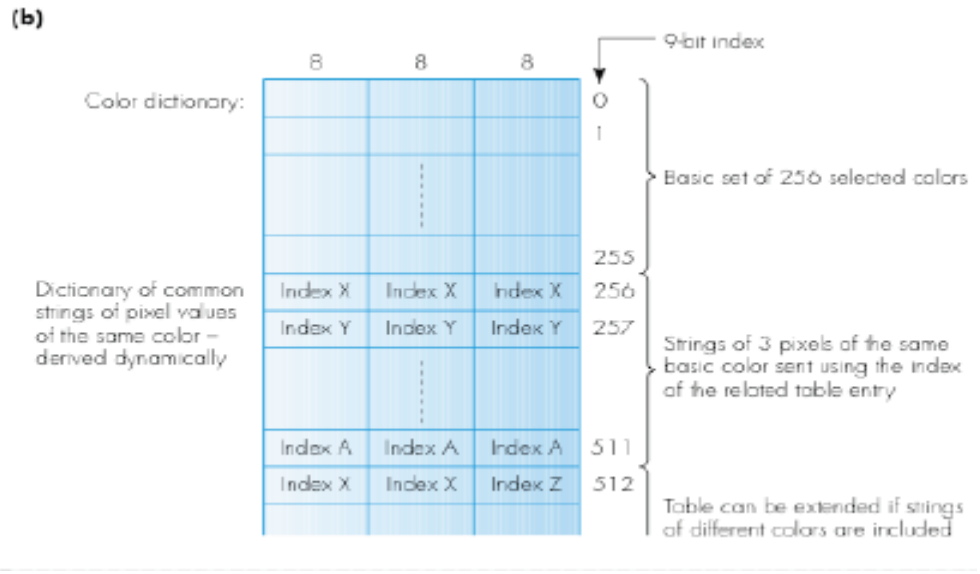
Hence instead of sending each pixel as a24bit valve, only the8bit index to the table entry that contains the closest match color to the original is sent.

The result in a compression ratio-3:1.The table of colors can relate either to the whole image in which case it is referred to as the globle color table- or to portion of the image, when it is referred to as a local color table.

The contents of the table are sent across the network- together with the compressed image data and other information such as the screen size and aspect ratio in a standardized format. The principles of the scheme are show in Figure 5(a)



The color dictionary, screen size, and aspect ratio are sent with the set of indexes for the image.



GIF compression principles (a) basic operation mode (b) dynamic mode using LZW coding

As shown in the figure 5 (b), the LZW coding algorithm can be used to obtain further levels of compression by extending the basic colour table dynamically as the compressed image data is being encoded and decoded.

As with text compression, the occurrence of common strings of pixel values –such as long strings of the same colour-are detected and these are entered into the colour table after the 256 selected colours.

However in this application, since each entry in the colour table comprises 24 bits, in order to save memory, to represent each string of pixel values just the corresponding string of 8 bit indices to the basic colour tables are used.

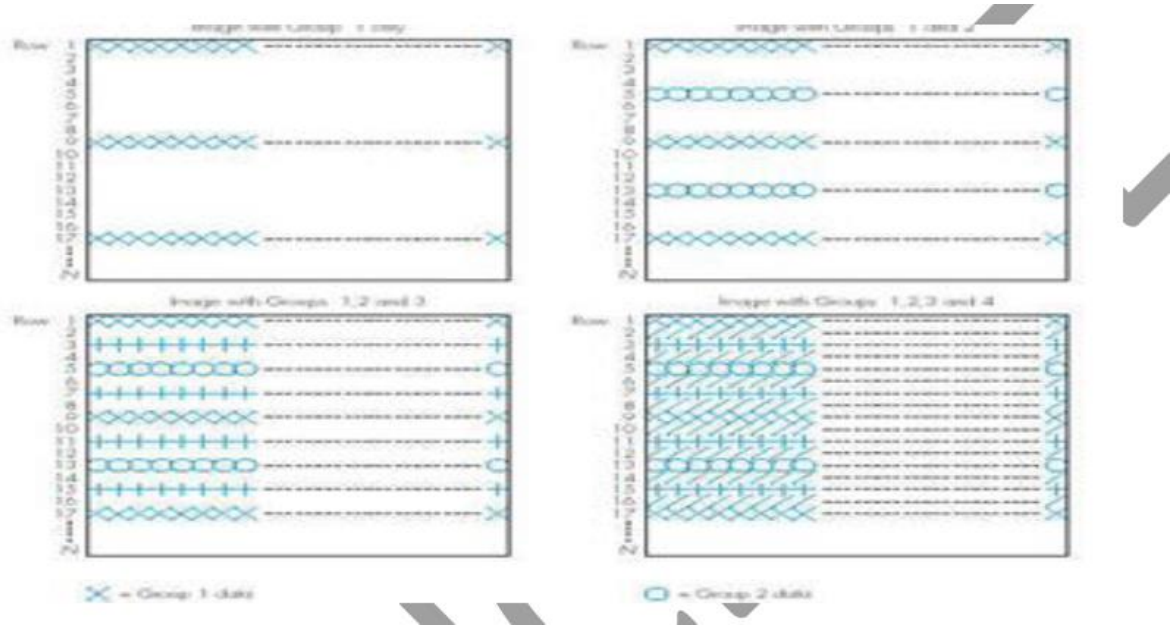
If we limit each entry in the table of 24 bits, then this will allow common strings comprising three pixel values to be stored in each location of the extended table.

Normally, since the basic table contains 256 entries, an initial table size of 512 entries is selected which allows for up to 256 common strings to be stored

. As with text compressions, however, should more strings be found, then the number of entries in the table is allowed to increase incrementally by extending the length of the index by 1 bit.

GIF also allows an image to be stored and subsequently transferred over the network in an interlaced mode. This can be useful when transferring images over either low bit rate

channels or the internet which provides a variable transmission rates. With this mode, the compressed image data is organized so that the decompressed image is built up in a progressive way as the data arrives. To achieve this, the compressed data is divided into four groups as shown in figure 6



GIF interlaced mode

and, as we can see, the first contains 1/8 of the total compressed image data, the second a further 1/8,the third a further ¼,and the last remaining ½. GIF also allows an image to be stored and subsequently transferred over the network in an interlaced mode.

This can be useful when transferring images over either low bit rate channels or the internet which provides a variable transmission rates.

With this mode, the compressed image data is organized so that the decompressed image is built up in a progressive way as the data arrives. To achieve this, the compressed data is divided into four groups as shown in figure 3, the first contains 1/8 of the total compressed image data, the second a further 1/8,the third a further ¼,and the last remaining ½

Tagged image file format (TIFF)

The tagged image file format (TIFF) is also used extensively. It supports pixel resolutions ofupto 48 bits-16 bits each for R,G and B-and is extended for the transfer of both the images and digitized documents.

The image data, therefore, can be stored-and hence transferred over the network-in a number of different formats. The particular format being used is indicated by a code number and these range from the uncompressed format (code number 1)through to LZW-compressed which is code number 5.code number 2,3.and 4 are intended for use wit digitized documents.

The LZW compression algorithm that is used is the same as that used with gif.it starts with a basic colour table containing 256 colors and the table can be extended to contain upto4056 entries containing common strings of pixels in the image being transferred. Again a standard format is used for the transfer of both the color table and the compressed image data.

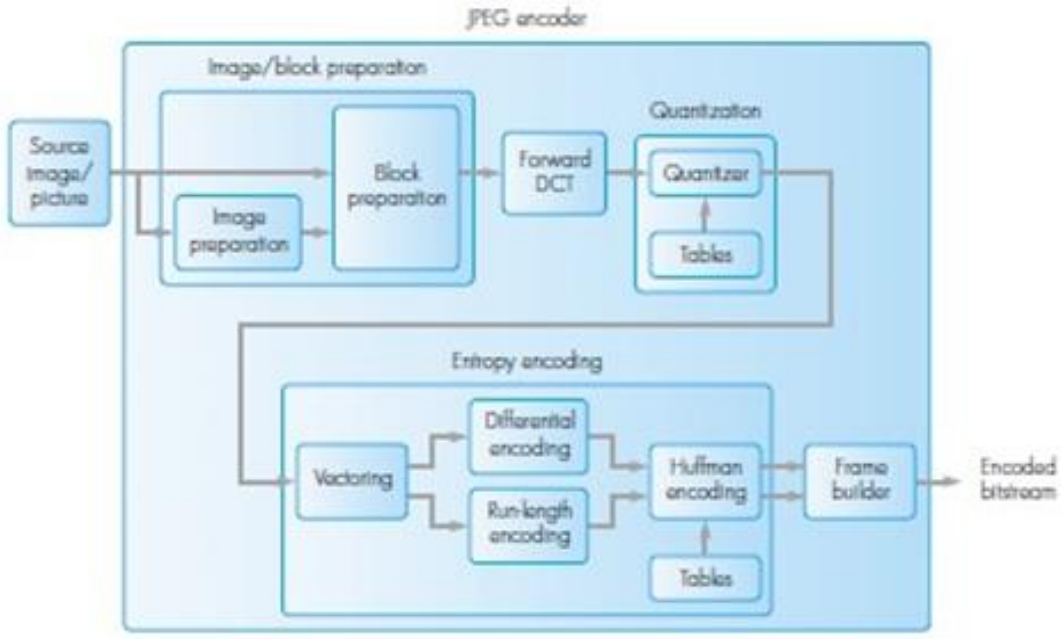
JPEG

In order to reduce the time to transmit digitized pictures, compression is normally applied to thetwo-dimensional array of pixel values that represents a Digitized picture before it is transmitted over the network.

The most widely adopted standard relating to the compression of digitized pictures has been developed by an international standards body known as the Joint Photographic Experts Group (JPEG).

JPEG also forms the basis of most video compression algorithms and hence we shall limit our discussion of the compression digitized pictures to describing the main principles of the JPEG standard.

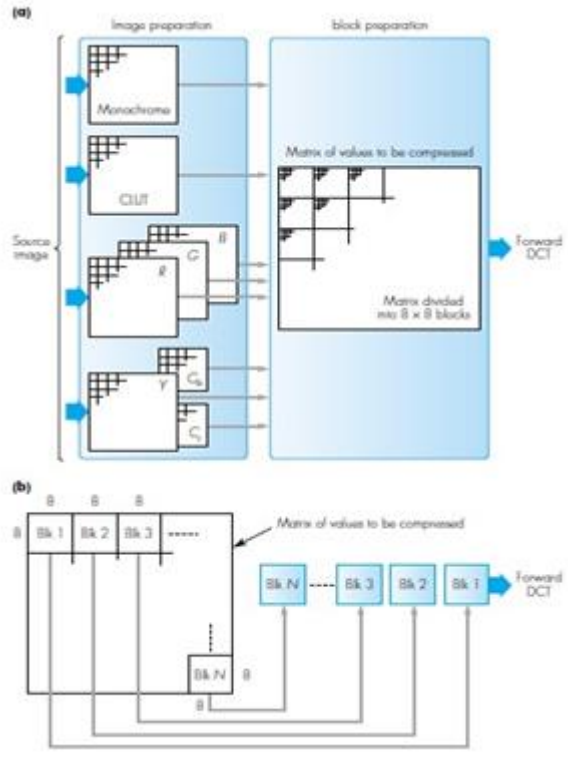
There are five main stages associated with the lossy sequential mode (baseline mode): image/block preparation, forward DCT, quantization, entropy encoding and frame building. These are shown in Figure 7 and their roles are discussed separately.



JPEG encoder schematic

(i) Image/block preparation

In the case of a continuous tone monochromatic image, just a single 2-D matrix is required to store the set of 8-bit gray-level values that represent the image. Similarly, for a color image, if a CLUT is used just a single matrix of values is required. Alternatively, if the image is represented in an R, G, B format three matrices are required, one each for the R, G, B quantized values. For color images, the alternative form of representation known as Y, Cb , Cr can optionally be used. This is done to exploit the fact that the two chrominance signals, Cb and Cr, require half the bandwidth of the luminance signal, Y. This in turn allows the two matrices that contain the digitized chrominance components to be smaller in size than the Y matrix so producing a reduced form of representation over the equivalent R,G,B form of representation. For example in the 4:2:0 format ,groups of four neighbouring chrominance values are averaged to produce a single value in the reduced matrix so reducing the size of cb and cr matrices by a factor of four. The four alternative forms of representation are as shown in the figure 8(a). Once the source image format has been selected and prepared ,the set of values in each matrix are compressed separately using the DCT. Before performing the DCT on each matrix however a second step known as block preparation is carried out. This is necessary since to compute the transformed value in all the locations of the matrix to be processed. It would be too time consuming to compute the DCT of the total matrix in a single step so each matrix is first divided into a set of smaller8*8 sub matrices. Each is known as a block and as we can see in part (b) of the figure 8, these are then fed sequentially to the DCT which transforms each block separately.



Image/block preparation (a) image preparation (b) block preparation

Forward DCT

Each pixel value is quantised using 8 bits which produce a value in the range 0 to 255 for the intensity/luminance values –R,G,B or Y- and a value in the range -128 to +127 for the two chrominance values –Cb and Cr. In order to compute the forward DCT however all the values are first centred around zero by subtracting 128 from each intensity/luminance value. Then ,if the input 2-D matrix is represented by :P[x,y] and the transformed matrix by F[i,j], the DCT of each

8*8 block of values is computed using the expression:

$$F[i,j]=\frac{1}{4}C(i)C(j)\sum_{x=0}^{7}\sum_{y=0}^{7}P[x,y]\cos\frac{(2x+1)i\pi}{16}\cos\frac{(2y+1)j\pi}{16}$$

Where C(i)and C(j)=$1/\sqrt{2}$ for i=j=0

=1 for all other values of i and j

and x,y, i and j all vary from 0 through 7.

However we can deduce a number of points by considering the expressions above:

All the 64 values in the input matrix P[x,y] contribute to each entry in the transformed matrix[x,y]

For i=j=0, the two cosine terms are both zero. Also, since cos(0)=1, the value in location F[0,0] of the transformed matrix is simply a function of the summation of all other values in the input matrix. Essentially it is the mean of all the 64 values in the matrix and is known as the dc coefficients.

Since all the values in all the other locations of the transformed matrix have a frequency coefficient associated with them either horizontal ,vertical or both they are known as ac coefficients.

For j=0,only horizontal frequency are present which increase the frequency for i=1 to7.

For i=0,only vertical frequency coefficients are present which increase the frequency for j=1 to 7.

In all other locations in the transformed matrix, both horizontal and vertical frequency coefficients are present to varying degrees.

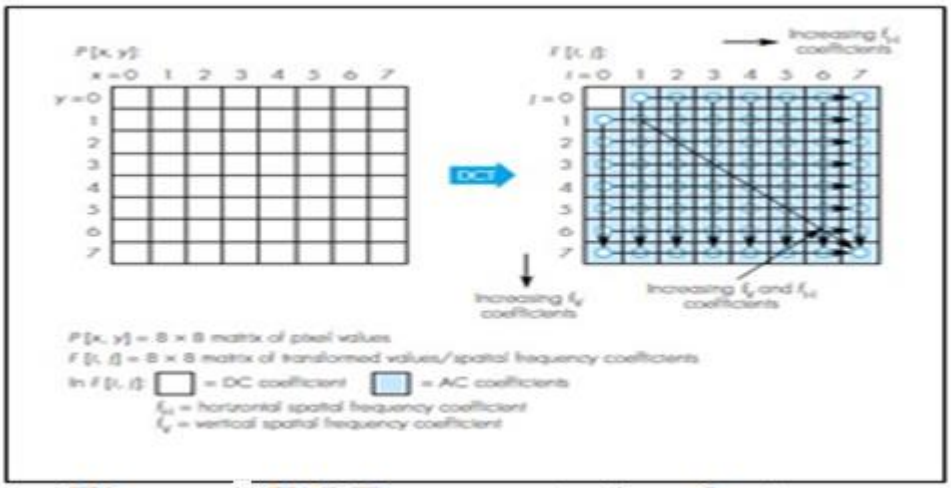


Figure DCT computation features

The above points are summarised in figure 9.

Quantisation:

In theory, providing the forward DCT is computed to a high precision using, say, floating pointarithmetic, there is very little loss of information during the DCT phase .although in practice small losses occur owing to the use of fixed point arithmetic, the main source of information loss occurs during the quantization and entropy encoding stages where the compression takes place. In transform encoding, the human eye responds primarily to the DC coefficient and the lower spatial frequency coefficients . Thus if the magnitude of a higher frequency coefficient is below a threshold ,the eye will not detect it. This property is exploited in the quantisation phase by dropping -in practice ,setting to zero - those spatial frequency coefficient in the transformed matrix whose amplitudes are less than a defined threshold value. It should be noted that although the eye is less sensitive to these frequency coefficients, once dropped the same frequency coefficients cannot be retrieved during the decoding procedure.

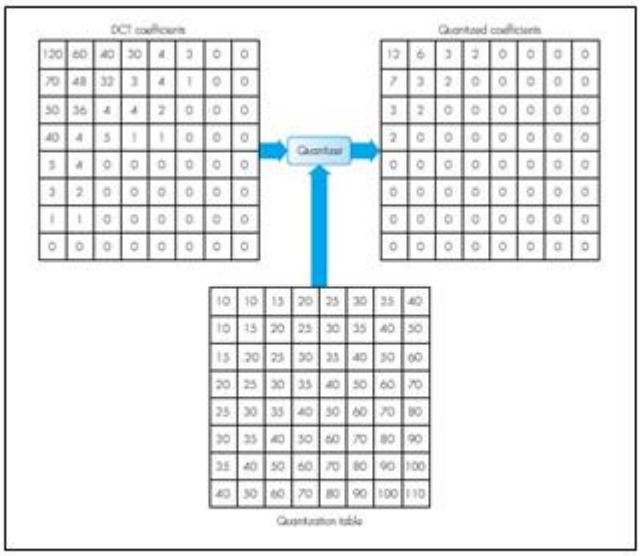


Figure Example computation of a set of quantized DCT coefficients

In addition to determining whether a particular spatial frequency coefficient is above a threshold ,the quantisation aims to reduce the size of AC and DC coefficients so that less bandwidth is required for their transmission. Instead of simply comparing each coefficient with the corresponding threshold value, a division operation is performed using the defined threshold value as the divisor. If the resulting quotient is zero, the coefficient is less the threshold value while if it is non zero, this indicates the number of times the coefficient is greater than the threshold rather than its absolute value. The sensitivity of the eye varies with spatial frequency, which implies that the amplitude threshold value below which the eye will detect a spatial frequency also varies. In practice, therefore threshold values used vary for each of the 64 DC coefficients. These are held in a two dimensional matrix known as quantisation table with the threshold value used with a particular DCT coefficient in the corresponding position in the matrix. Clearly the choice of the threshold values is important and in practice is a compromise between the level of compression that is required and the resulting amount of information loss that is acceptable. Although the JPEG standard includes two default quantisation table values - one for use with luminous coefficients and the other for chrominance coefficients -it also allows a customised table to be used and sent with compressed image .an example set of threshold values is given in the figure 10 together with a set of DCT coefficients and their corresponding quantised values.

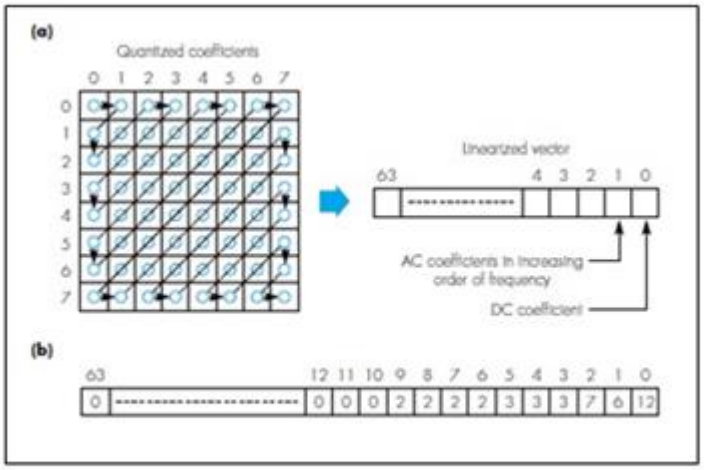
We can conclude a number of points from the values shown in the tables:

The computation of the quantised coefficients involves rounding the coefficients ti the nearest integer value.

The threshold values used in general increase in magnitude with increasing spatial frequency. The DC coefficient in the transformed matrix is largest. Many of the higher frequency coefficients are zero.

Entropy encoding:

The entropy encoding stage comprise s four steps: differential encoding, run-length encoding ,vectoring and Huffman encoding.



Vectoring using a Zig-zag scan (a) principle (b) vector for example shown in

Vectoring:

Entropy encoding operate on a one dimensional string of values, that is a vector. As

we have just seen ,however, the output of the quantisation stage is a 2D matrix of values .Hence before we can apply any entropy encoding to set of values in the matrix ,we must first represent the values in the form of a single dimensional vector. This operation is called vectoring.

As we saw in the figure , the output of a typical quantisation is a 2-D matrix of values/coefficients which are mainly zeros except for a number of non-zero values in the top left hand corner of the matrix. Clearly if we simply scanned the matrix using a line by line approach, then the resulting vector would contain a mix of zero and non-zero values. In general however this type of information structure does not lend itself to compression.

Differential encoding:

The first element in each transformed block is the DC coefficient which is a measure of luminance/chrominance associated with the corresponding 8 x 8 block of pixel values. Hence it is the largest coefficient and because of its importance, its resolution must be kept as high as possible during quantisation phase. Because of the small physical area covered by each block, the DC coefficient varies only slowly from one block to next. The most efficient type of compression with this form of information structure is differential encoding since this encodes only the difference between each pair of values in a string rather than the absolute values.

Run length encoding

The remaining 63 values in the vector are the AC coefficients and, because of the zig-zag scan,the vector contains long strings of zeros within it. To exploit this feature, the AC coefficients are encoded in the form of a pairs of values. Each pair is made up of (skip, value) where skip is the number of zeros in the run and the value the next non-zero coefficient.

(0,6) (0,7) (0,3) (0,3) (0,3) (0,2) (0,2) (0,2) (0,2) (0,0)

Note that the final pair (0,0) indicate the end of the string for this block and that all the remaining coefficients in the block are zero. Also that the value field is encoded in the form SSS /value.

Huffman encoding

For the encoding of digitized documents , significant levels of compression can be obtained by emplacing long strings of binary digits by a string of much shorter code words ,the length of ach code word being a function of this relative frequency of occurrence. Normally, a table of code words is used with the set of code words pre-computed using the Huffman coding algorithm.

Frame building

Typically the bit stream output by a JPEG encoder –corresponding to, say, compressed version of a printed picture is stored in the computer ready for either integrating without the media if necessary or accessing from a remote computer . As we can see from the above , in order for the decoder in the remote computer to be able to interpret all the different fields and table that make up the bit stream .it is necessary to delimit each field and set up table values in a defined way .the JPEG standard therefore also includes a definition of the structure of the total bit stream relating to the particular image or picture this is known as the frame and the outline structure is shown in the figure 12. The role of frame builders is to encapsulate all the information relating to an encoded image or

picture in this format and, as we can see, the structure of a frame is hierarchical. At the top level the complete frame plus header is encapsulated between a start of frame and end of frame delimiter which allows the receiver to determine the start and end of all the information relating to a complete picture or image. The frame header contains a number of fields that include:

The overall width and height of the pixels;

The number and type of the components that are used to represent the image (CLUT, R/G/B, Y/C/C);

The digitalization format used (4:2:2, 4:2:0 etc)

At the second level, frame consists of a number of components each of which is known as a scan. These are preceded by a header which contains fields that include:

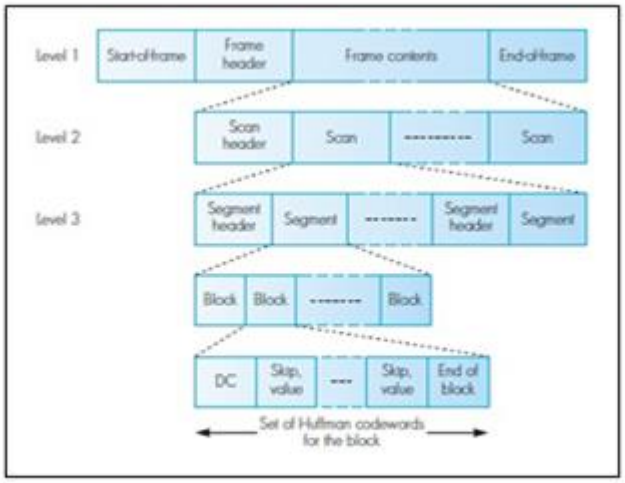
The identity of the components (R/G/B etc)

The number of the bits used to digitalize each component;

The quantization table of values that have been used to encode each component.

Typically, each scan /component comprises of one or more segments each of which can contain a group of (8X8) blocks preceded by a header . This contains the Huffman table of values that has been used to encode each block in the segment should the defaults tables not to be used. In

this way, each segment can be decoded independently of the others which overcomes the possibility of the bit errors propagating and affecting other segments. Hence each complete frame contains all the information necessary to enable the JPEG decoder to identity each field in a received frame and then perform the corresponding decoding operation.



JPEG encoder output bitstream format

JPEG decoding

A JPEG decoder is made up of a number of stages which are simply the corresponding decoder sections of those used in the encoder as shown in the figure 13.

Hence the time to carry out the decoding function is similar to that used to perform the encoding, on receipt of the encoded bit stream the frame decoder first identifies the control information and tables within the various headers. If then loads the contents of each table into the related table and passes the control information to the image builder. It then starts to pass the compressed bit stream to the Huffman decoder which carries out the corresponding decompression operation using either the default or the preloaded table of code words. The two decompressed streams containing the DC and AC coefficients of each block are then passed to the differential and run-length decoders respectively. The resulting matrix of values is then dequantized using either the default or the preloaded values in the quantization table. Each resulting block of 8X8 spatial frequency coefficients is passed in turn to the inverse DCT which transforms them back in to their spatial form using the expression:

$$P[x,y]=(1/4)\sum_{i=0}^{7}\sum_{j=0}^{7} C(i)\,C(j)\,F[i,j]\cos\left((2x+1)i\pi\right)/16)\cos\left((2y+1)j\pi/16\right)$$

Where C(i) and C(j)=$1/\sqrt{2}$ for i,j=0

=1 for all other values of i and j

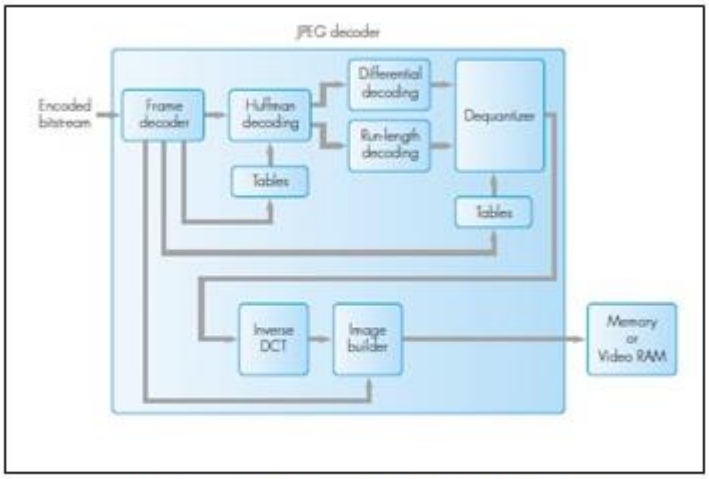


Figure JPEG decoder schematic

The image builder then reconstructs the original image from these blocks using the control information passed to it by the frame decoder. Although the JPEG standard is relatively complicated owing to the number of encoding /decoding stages associated with it , compression ratios in excess of 20:1 can be obtained while still retaining a good quality output image/picture. This level of compression, however, applies to pictures whose content is relatively simple-that is, have relatively few colour transitions-and, for more complicated pictures, compression ratios nearer to 10:1 are more common. These

figures, however, assume each pixel location has three associated with it – R/G/B or Y/Cb/Cr-and hence if a CLUT is used , then both figures can be multiplied by a factor of three. Nevertheless, even with a compression ratio of 10:1, the amount of memory required with various types of display tabulated nit table 2.1 is reduced to a range of from 30 kB to 240Kbytes. More importantly, the time delay incurred in accessing these images is reduced by a factor of 10. Finally, as with the GIF it is also possible to encode and rebuilt the image in a progressive way by first sending a an outline of the image and then progressively adding more detail to it. This can be achieved in following ways

Progressive mode: in this mode, first the DC and low frequency coefficients of each block are send and then the higher frequency coefficients.

Hierarchical mode: in this mode the total image is first sent using a low resolution-for example

320X240 –then at a higher resolution such as 640X480.

UNIT III

TEXT COMPRESSION

- **COMPRESSION PRINCIPLES**
- **SOURCE ENCODERS & DESTINATION DECODERS**
- **LOSSLESS & LOSSY COMPRESSION**
- **ENTROPY ENCODING**
- **SOURCE ENCODING**
- **TEXT COMPRESSION**
- **STATIC HUFFMAN CODING**
- **DYNAMIC HUFFMAN CODING**
- **ARITHMETIC CODING**
- **LEMPEL –ZIV- WELSH COMPRESSION**

1. COMPRESSION PRINCIPLES:

The compression principles comprises of the following parameters

- Source encoders and destination decoders
- Lossless and lossy compression
- Entropy encoding
- Source encoding

2. SOURCE ENCODER AND DESTINATION DECODER

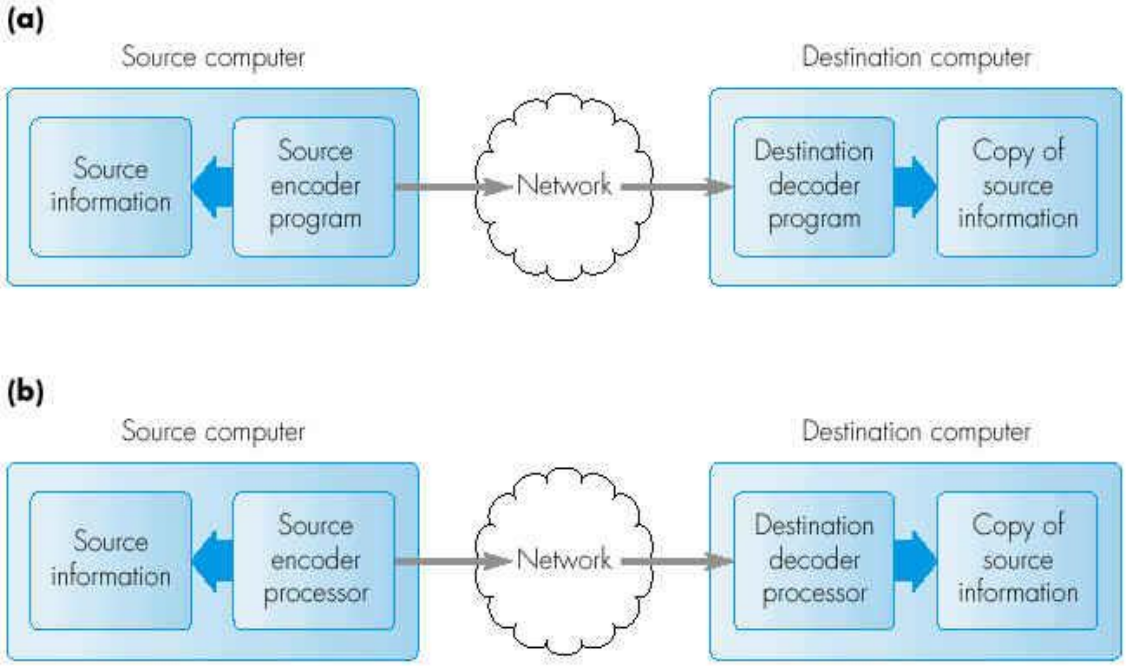


Fig 1

(1) When we transmit the source information we compress that using compression algorithm.(i.e) the information is given to the source encoder the source encoder removes redundant information and produces compressed information .hence data rate is reduced.

(2) It is then transmitted across network

(3) At the receiver side, the destination decoder uncompressed this information. It generates replica of the original information and gives it to the receiver.

(4) Here, the compression /decompression can be performed by the software (or) special purpose processor can be used.

(5) Normally DSP processors are used for this purpose.

(6) In application involving two computers in case of text and image file compression fig (a) is employed

(7) The time required to perform the compression and decompression algorithms in software is not acceptable and instead the two algorithms must be performed by special processors in separate units ref fig (b)

3. LOSS LESS AND LOSSY COMPRESSION

Lossless compression algorithm, when the compressed information is decompressed, there is no loss of information to be reversible

Lossy compression algorithms, is normally not to reproduce an exact copy of the source information after decompression

Example application of lossy compression is for the transfer of digitized images and audio and video streams

LOSSLESS COMPRESSION	**LOSSY COMPRESSION**
1.No information is lost	Some information is lost
2.completely reversible	It is not reversible
3.used for text and data	Used for speech and video
4.compression ratio is less	High compression ratio
5.compression is independent of human response	Compression depends upon sensitivity of human ear and eyes etc.
Ex Huffman, run length coding	Ex: Transform coding

4. ENTROPY ENCODING

Lossless and independent of the type of information that is compressed
Two examples:

- Run-length encoding
- Statistical encoding

Run-length encoding

When the source information comprises long substrings of the same character or binary digit.

Instead of transmitting these directly, they are sent in the form of a string of codewords, each indicating both the bit - 0 or 1 - and the number of bits in the substring

000000011111111110000011…

This could be represented as:0,7,1,10 0,5,1,2…

Alternatively, if we ensure the first substring always comprises binary 0s, then the string could be represented as 7,10,5,2…

Statistical encoding

Statistical encoding exploits this property by using a set of variable length codewords with the shortest codewords used to represent the most frequently occurring symbols

Ensure that a shorter codeword in the set does not form the start of a longer codeword otherwise the decoder will interpret the string on the wrong codeword boundaries prefix property

EX: Huffman coding algorithm

Minimum average number of bits that are required to transmit a particular source stream is known as the entropy of the source

Theoretical minimum average numbers of bits that are required to transmit (represent)

information is known is entropy Computed using Shannon's formula of Entropy

$$\text{Entropy, } H = -\sum_{i=1}^{n} P_i \log_2 P_i$$

n number of different symbols Pi the probability of occurrence of the symbol i

Efficiency of a particular encoding scheme isoften computed as a ratio of entropy of the source

- To the average number of bits per codeword that are required for the scheme

$$= \sum_{i=1}^{n} N_i P_i$$

n number of different symbols Pi the probability of occurrence of the symbol i, Ni number of Bits to represent this symbol

EXAMPLE

A statistical encoding algorithm is being considered for the transmission of a large number of long text files over a public network. Analysis of the file contents has shown that each file comprises only the six different characters M, F, Y, N, 0, and 1 each of which occurs with a relative frequency of occurrence of 0.25, 0.25, 0.125, 0.125, 0.125, and 0.125 respectively. If the encoding algorithm under consideration uses the following set of codewords:

M = 10, F = 11, Y = 010, N = 011, 0 = 000, 1 = 001

compute:

(i) the average number of bits per codeword with the algorithm,

(ii) the entropy of the source,

(iii) the minimum number of bits required assuming fixed-length codewords.

ANSWER:

(i) Average number of bits per codeword

$$= \sum_{i=1}^{6} N_i P_i = (2(2 \times 0.25) + 4(3 \times 0.125))$$

$$= 2 \times 0.5 + 4 \times 0.375 = 2.5$$

(ii) Entropy of source

$$= \sum_{i=1}^{6} P_i \log_2 P_i = -(2(0.25\log_2 0.25) + 4(0.125\log_2 0.125))$$

$$= 1 + 1.5 = 2.5$$

(iii) Since there are 6 different characters, using fixed-length codewords would require a minimum of 3 bits (8 combinations).

5. SOURCE ENCODING

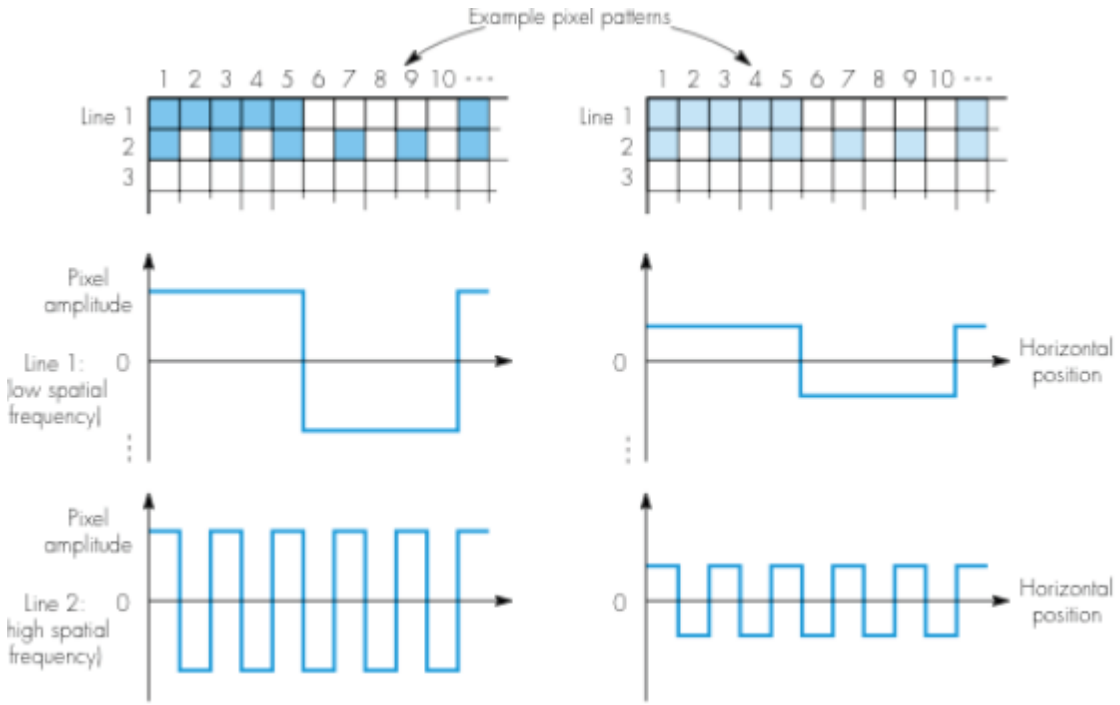
Differential encoding

-- Instead of using a set of relatively large codewords to represent the amplitude of each value/signal, a set of smaller codewords can be used each of which indicates only the difference in amplitude between the current value/signal being encoded

-- For example, 12 bits to obtain the required dynamic range but the maximum difference in amplitude between successive samples of the signal requires only 3-bits

Transform encoding

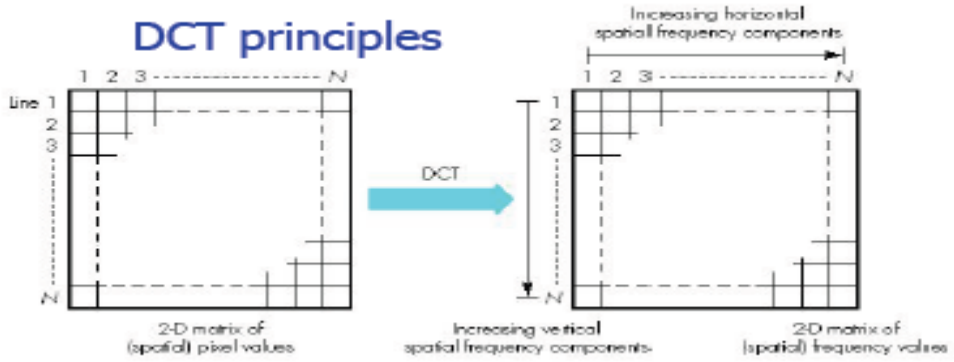
-- As we scan across a set of pixel locations
-- The rate of change in magnitude will vary from zero, if all the pixel values remain the same
-- A high rate of change it each pixel magnitude changes from one location to the next
-- The rate of change in magnitude as one traverses the matrix gives rise to a term known as spatial frequency



The human eye is less sensitive to the higher spatial frequency components

-- If we can transform the original spatial form of representation into an equivalent representation involving spatial frequency components, then we can more readily identify and eliminate those higher frequency components which the eye cannot detect thereby reducing the volume of information

The transformation of a two-dimensional matrix of pixel values into an equivalent matrix of spatial frequency components discrete cosine transform (DCT)- Figure 3



6. TEXTCOMPRESSION

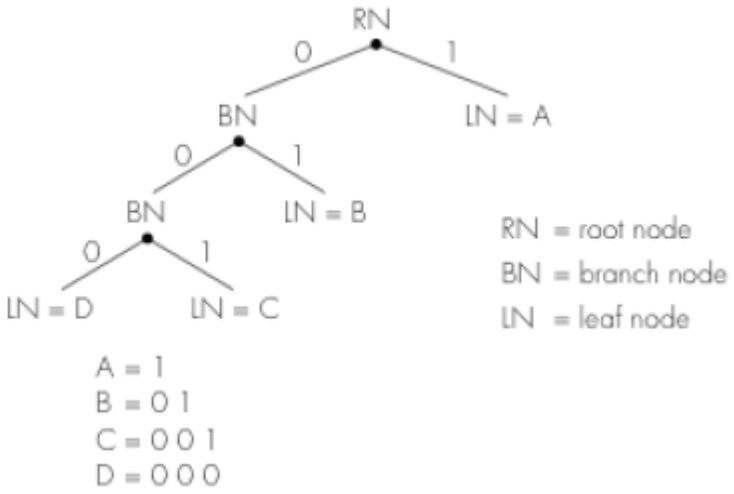
- Static Huffman coding
- The character string to be compressed is analysed
- The character types and their relative frequency are determined
- Coding operation by a Huffman code tree

• Binary tree with branches assigned the values 0 and 1

• Base of the tree is the root node, point at which a branch divides is called a branch node

• Termination point of a branch is the leaf node

An example of the Huffman code tree that corresponds to the string of characters AAAABBCD



Each branch divides a binary value 0 or1 is assigned for the new branch

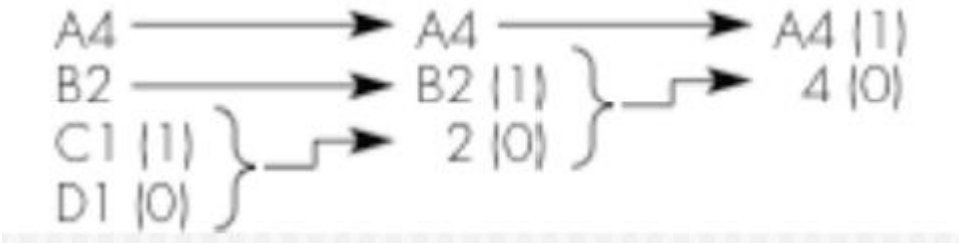
- The binary code words are determined by tracing the path from the root node out to each leaf
- Code has a **prefix property**
- A shorter code word in the set does not form a start of a longer code word

To code AAAABBCD by the Huffman code tree we need 14 bits

4*1+2*2+1*3+1*3=14 bits

- For 7-bits ASCII code words we need 8*7=56bits
- Which 56% of the Huffman code tree
- • 56%=14/56*100

Building a Huffman code tree

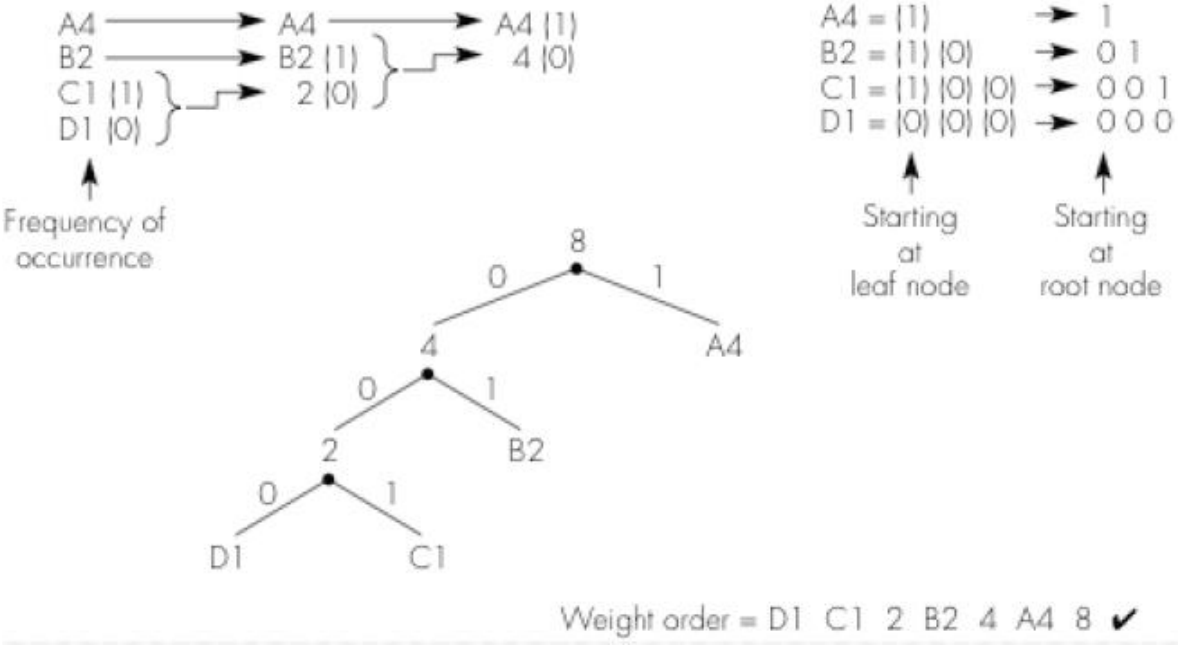


The first two less frequent characters C and D with their frequency 1 (C1, D1) are assigned to the (1) and (0) branches

- The two leaf nodes are then replaced by a branch

node whose weight is the sum of the weights of the two leaf nodes (sum is two)

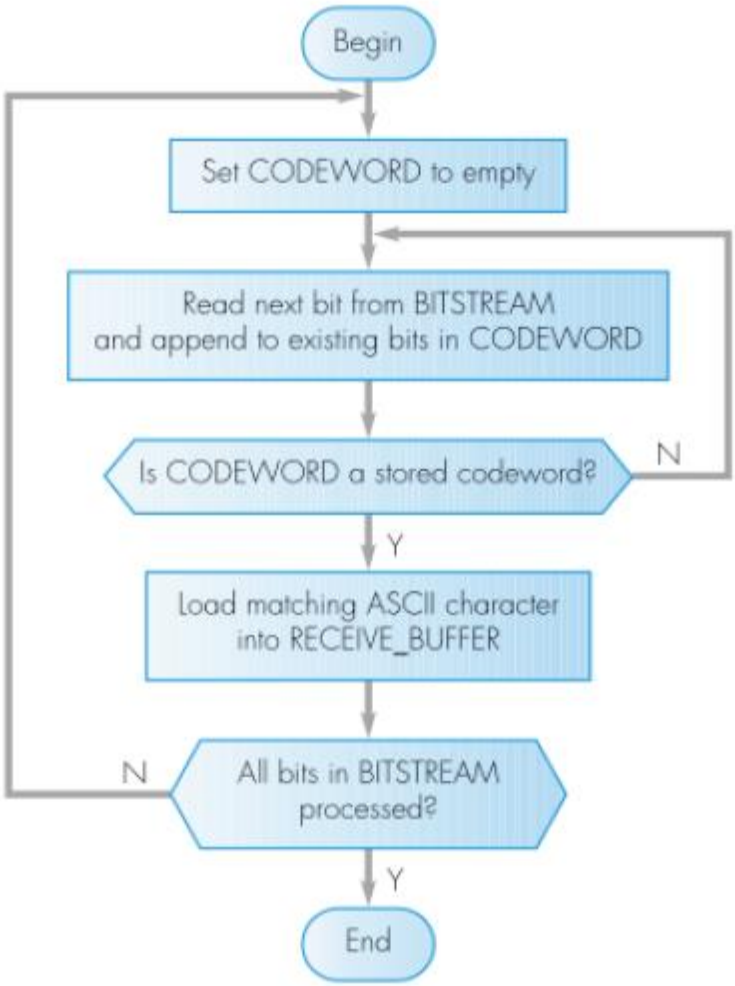
- This procedure is repeated until two nodes remain



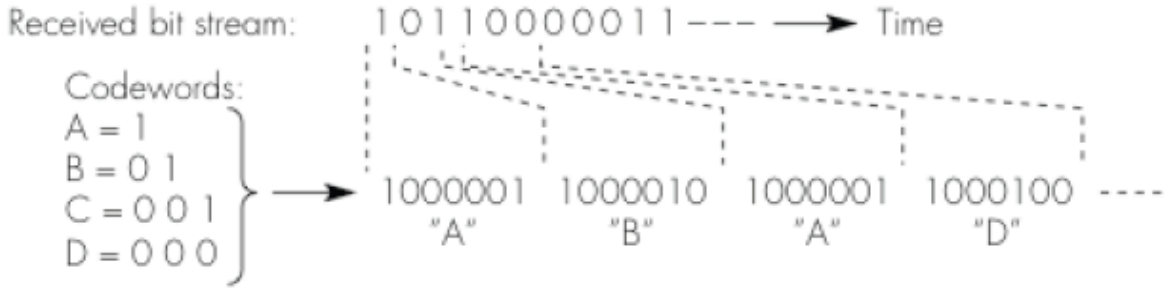
We check that his is the optimum tree - and -hence the code words

- List the resulting weights
- The code words are optimum if the resulting tree increments in weight order

Because of the order in which bits are assigned during the encoding procedure
Huffman code words have the unique property that shorter code words will never form the start of a longer code word Prefix property



Decoding into ASCII



Static Huffman coding

Messages comprising even different characters, A through G, are to be transmitted over a data link.

Analysis has shown that the relative frequency of occurrence of each character is:

Symbol	A	B	C	D	E	F	G	H
Probability	0.25	0.25	0.14	0.14	0.055	0.055	0.055	0.055

(i) Derive the entropy of the message (6)
(ii) Use static Huffman coding (10)

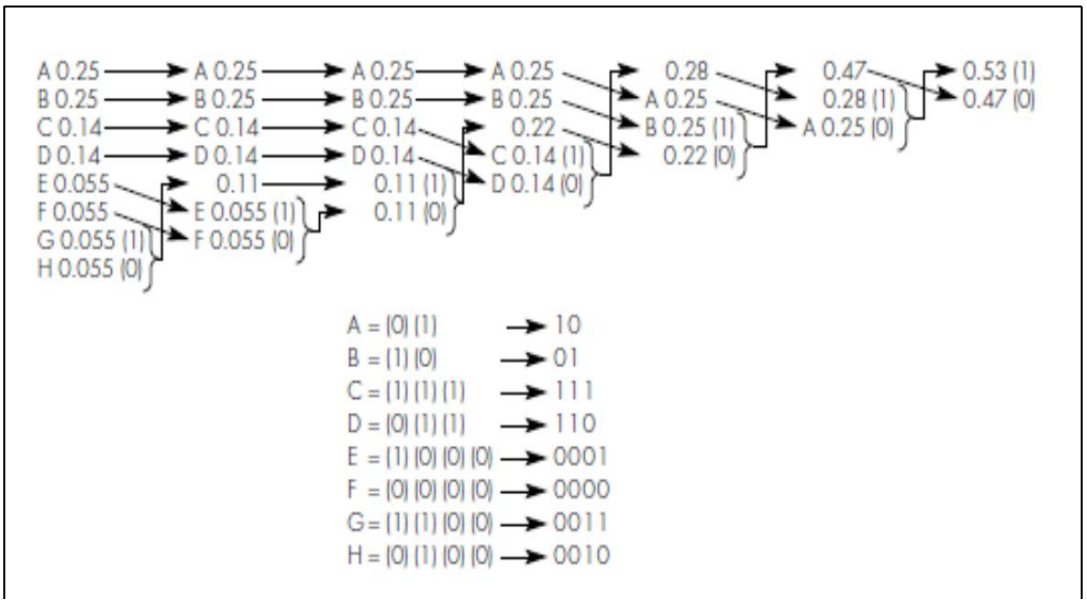
(i)

$$\text{Entropy, } H = -\sum_{i=1}^{8} P_i \log_2 P_i \text{ bits per codeword}$$

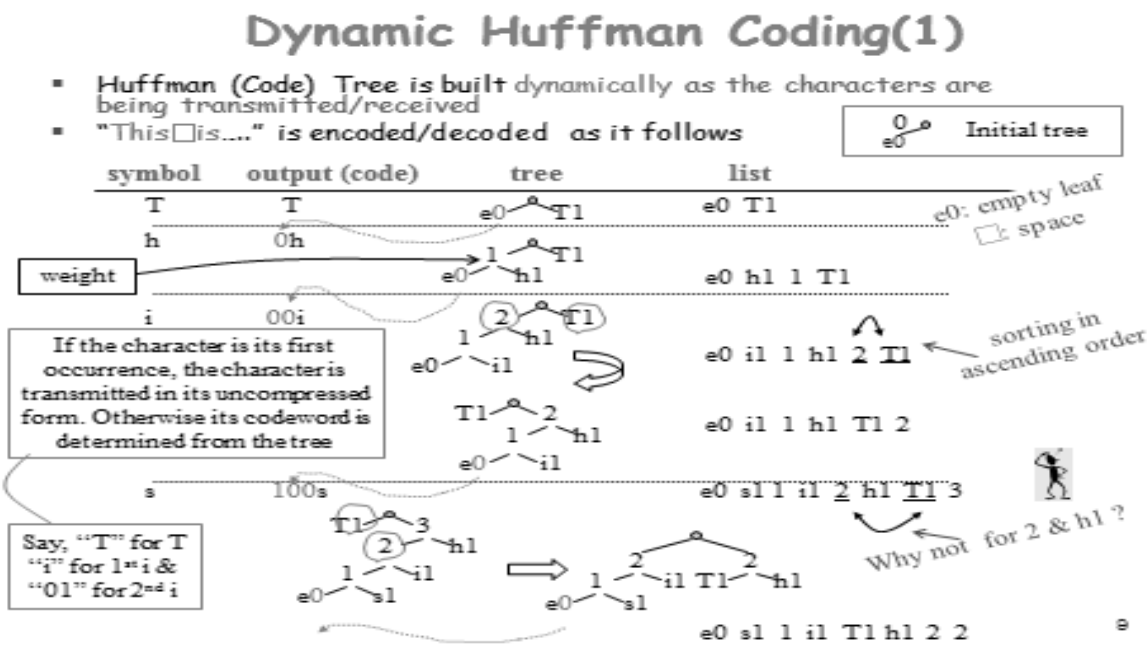
Therefore:

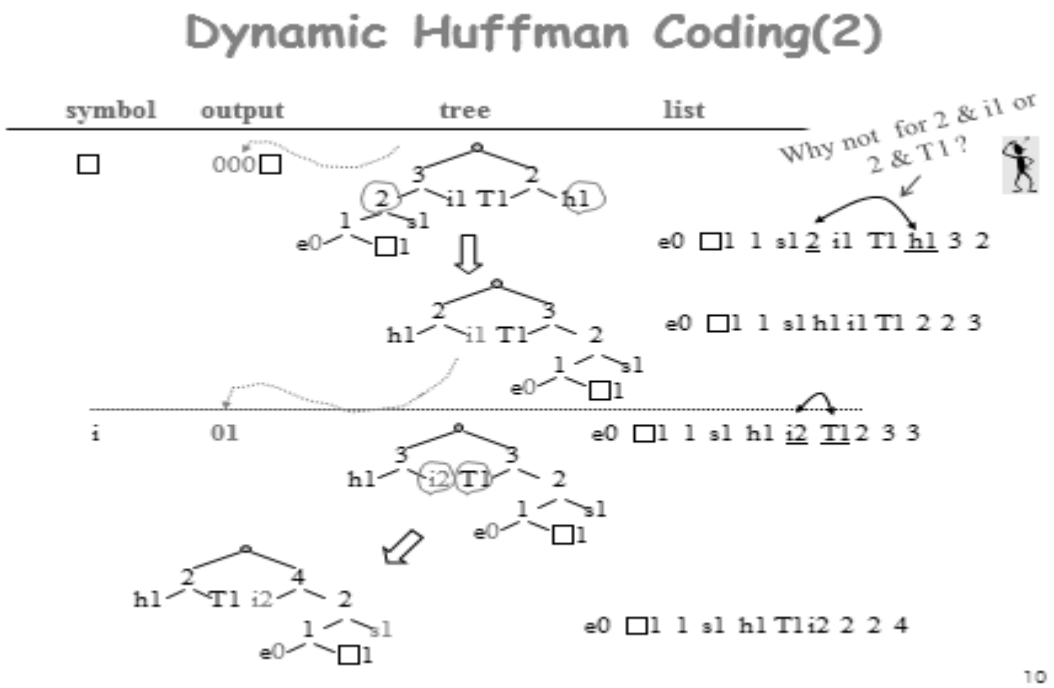
$$H = -(2(0.25 \log_2 0.25) + 2(0.14 \log_2 0.14) + 4(0.055 \log_2 0.055))$$
$$= 1 + 0.794 + 0.921 = 2.175 \text{ bits per codeword}$$

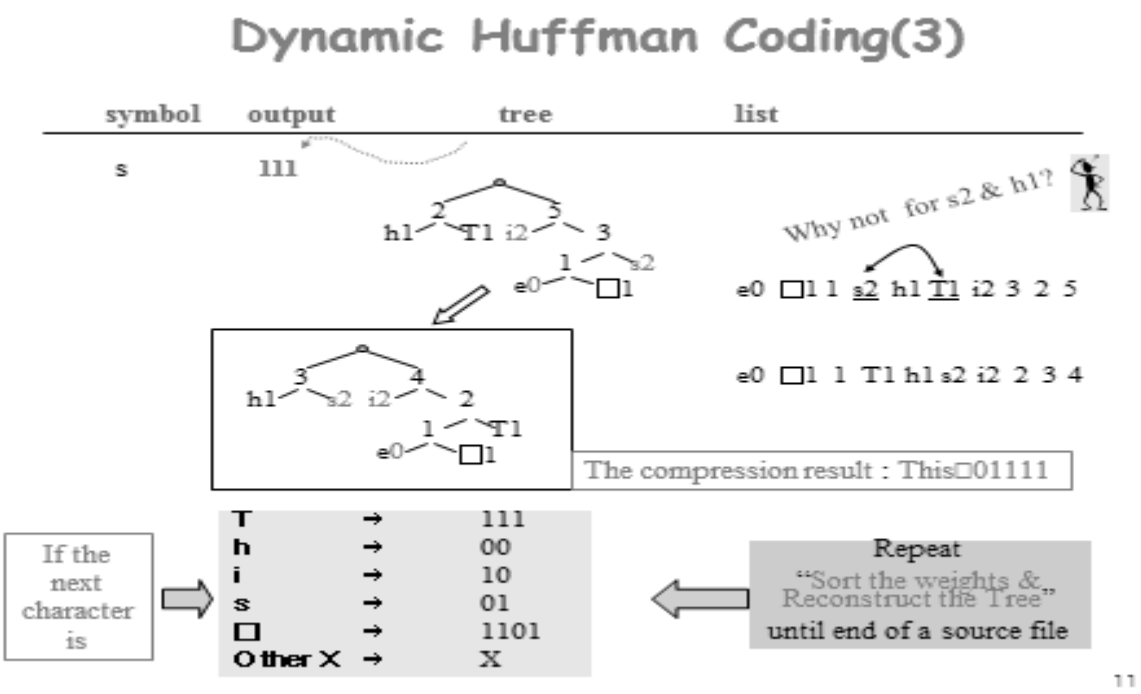
(ii)



DYNAMIC HUFFMAN CODING



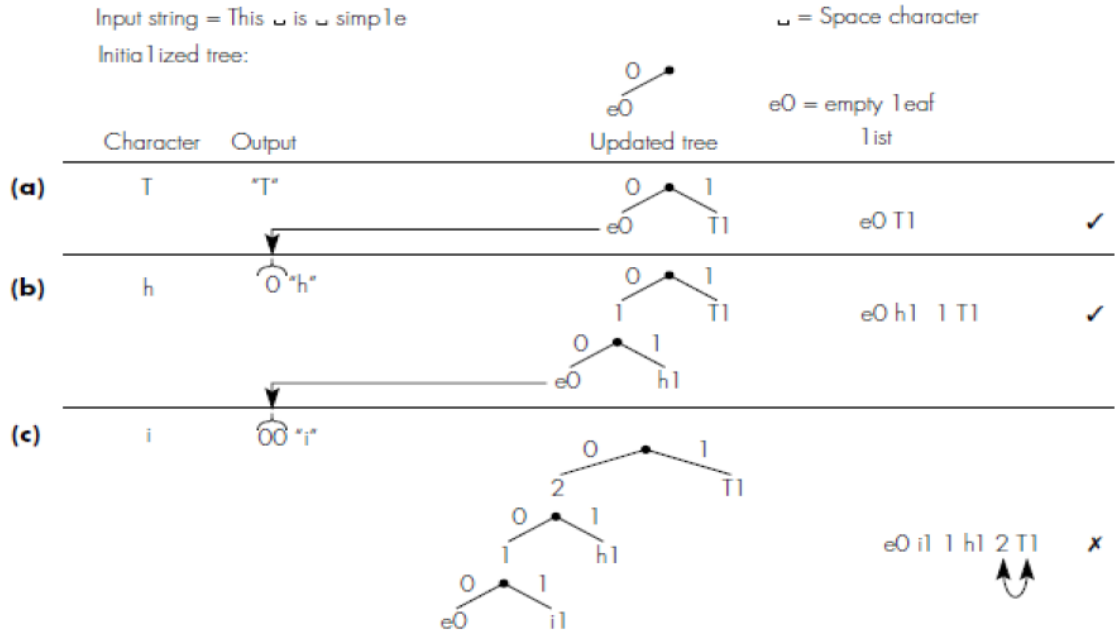


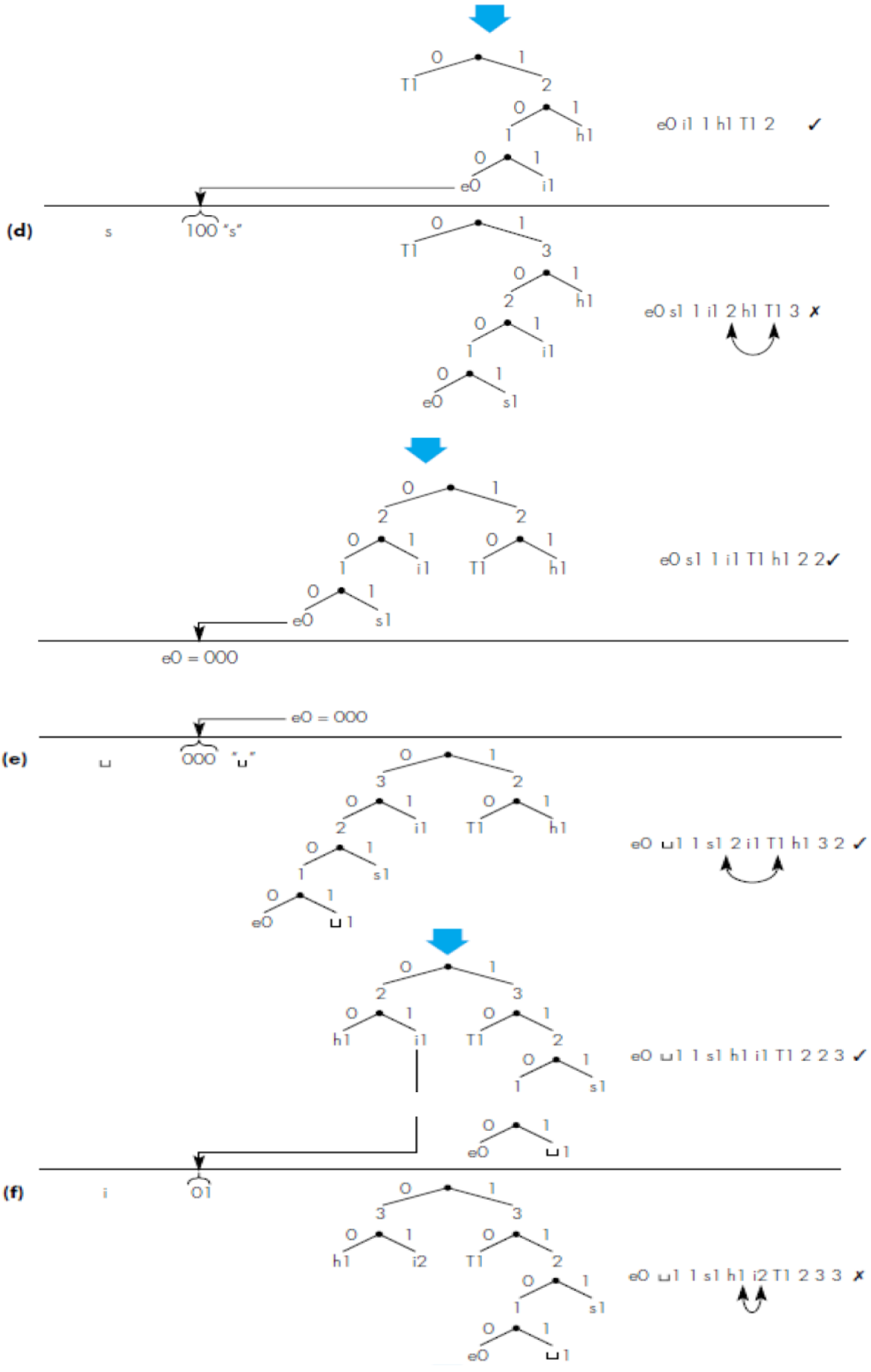


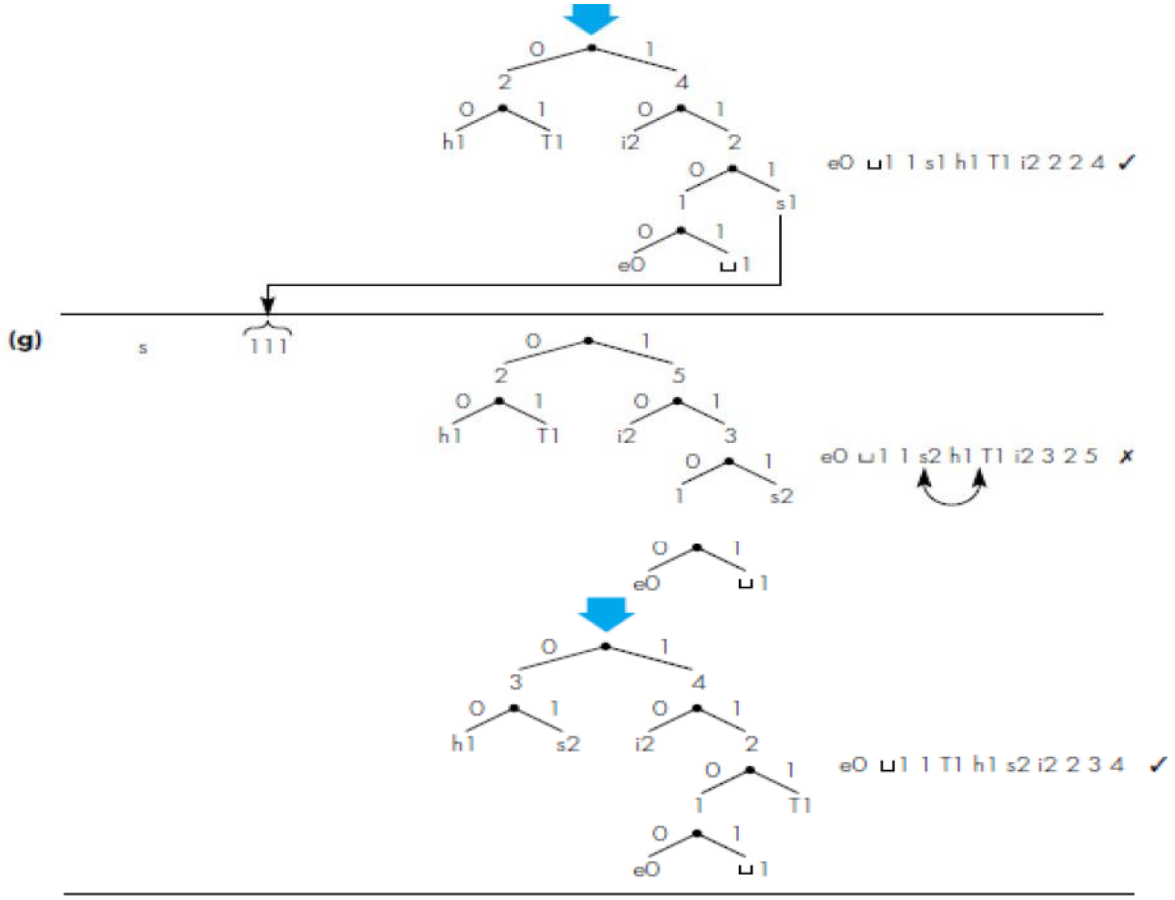
The following character string is to be transmitted using Huffman coding:

This is simple

Derive the Huffman code tree.




(d)
s
100 "s"
e0 i1 1 h1 T1 2 ✓
e0 s1 1 i1 2 h1 T1 3 ✗
e0 = 000
(e)
␣
000 "␣"
e0 ␣1 1 s1 2 i1 T1 h1 3 2 ✓
e0 ␣1 1 s1 h1 i1 T1 2 2 3 ✓
(f)
i
01
e0 ␣1 1 s1 h1 i2 T1 2 3 3 ✗

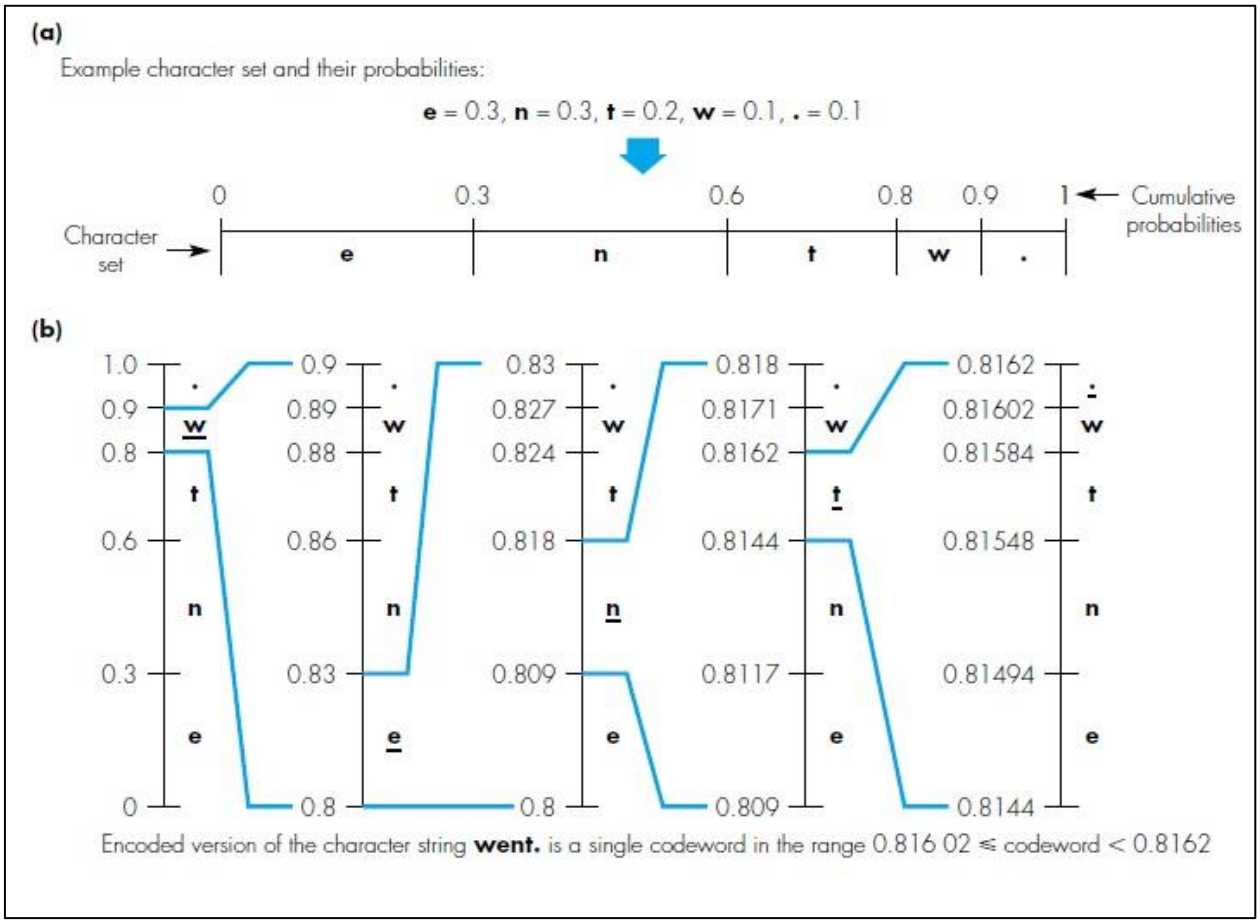

(g)
s
1 1 1
eO ␣1 1 s1 h1 T1 i2 2 2 4 ✓
eO ␣1 1 s2 h1 T1 i2 3 2 5 ✗
eO ␣1 1 T1 h1 s2 i2 2 3 4 ✓

7.Arithmetic coding

Assume the following character set and their probabilities:

Character	e	n	t	w	.
Probability	0.3	0.3	0.2	0.1	0.1

Using, determine the codeword for the character string ***went*.** Assuming this is received by the destination, explain how the decoder determines the original string from the received codeword value.



8. LEMPEL ZIV CODING

(i)The Lempel-Ziv (LZ) compression algorithm, instead of using single characters as the basis of the coding operation, uses strings of characters. For eg, for the compression of text, a table containing all the possible character strings.

(ii) for eg words-that occur in the text to be transferred is held by both the encoder and decoder.

(i) As each word occurs in the text instead of sending the word as a set of individual –say, ASCII- code words the encoders sends only the index of where the word is stored in the table

(ii) and on receipt of each index, the decoder uses this to access the corresponding word/string of characters from the table and proceeds the reconstruct the text into its original form.

(iii) Thus the table is used as a dictionary and the LZ algorithm is known as a dictionary – based compression algorithm.

(iv) Most word – processing packages have a dictionary associated with them which is used for both spell checking and for the compression of text. Typically, they contain in the region of 25,000 words and hence 15 bits which has 32768 combinations – are required to encode the index.

(v) To send the word "multimedia" with such a dictionary would require just 15 bits instead of 70 bits with 7- ASCII bit codewords.

(vi) These results in a compression ratio of 4:7:1. Clearly, shorter words will have a lower compression ratio and longer words a higher ratio.

(vii) Hence a variation of the LZ algorithm has been developed which allows the dictionary to be built up dynamically by the encoder and decoder as the compressed text is being transferred.

LEMPEL ZIV WELSH CODING:

(i) The principle of Lempel-Ziv-welsh (LZW) coding algorithm is for the encoder and the decoder to build the contents of the dictionary dynamically as the text is being transferred.

(ii) Initially, the dictionary held by both the encoder and the decider contains only the character set – for example ASCII-

(iii) The remaining entries in the dictionary are then built up dynamically by both the encoder and the decoder and contains the word that occur in the text.

(iv) for example, if the character set comprises 120 characters and the dictionary is limited to say4096 entries,

(v) Then the first 128 entries would contain the single character that make up the character set and the remaining 3968 entries would each contain strings of two or more characters that make up the words in the text being transferred.

(vi) The more frequently the words stored in the dictionary occur in the text, the higher the level of compression, let us assume that the text to be compressed starts with the string. This is simple as it is….

(vii) The dictionary to contain only words, then only the strings of character that consists of alphanumeric characters

(viii) Initially, the dictionary held by both the encoder and decoder contains only the individual characters form the character set being used; for eg, the 128 characters in the ASCII character set.

(ix) Hence the first word in the eg text sent by the encoder using the index of each of the four characters T,h,i and s. at this point , when the encode reads the next character from the string- the first space (SP) character-it determines that this is not an alphanumeric character.

(x) in addition, interprets it as terminating the first word and hence stores the proceeding four characters in the next available (free) location in the dictionary.

(xi) Similarly the decoder, on receipt of the first five indices /code words, reads the characters stored at each index and commences to reconstruct the text. When it determines that the fifth character is a space character, it interprets this as a word delimiter and proceeds to store the word this in its dictionary.

(xii) Same procedure is followed by both the encoder and decoder for transferring the otherwords except the encoder , prior to sending each word in the form of single characters , first checks to determine the word is currently stored in its dictionary and ,if it is , it sends only the index for the word.

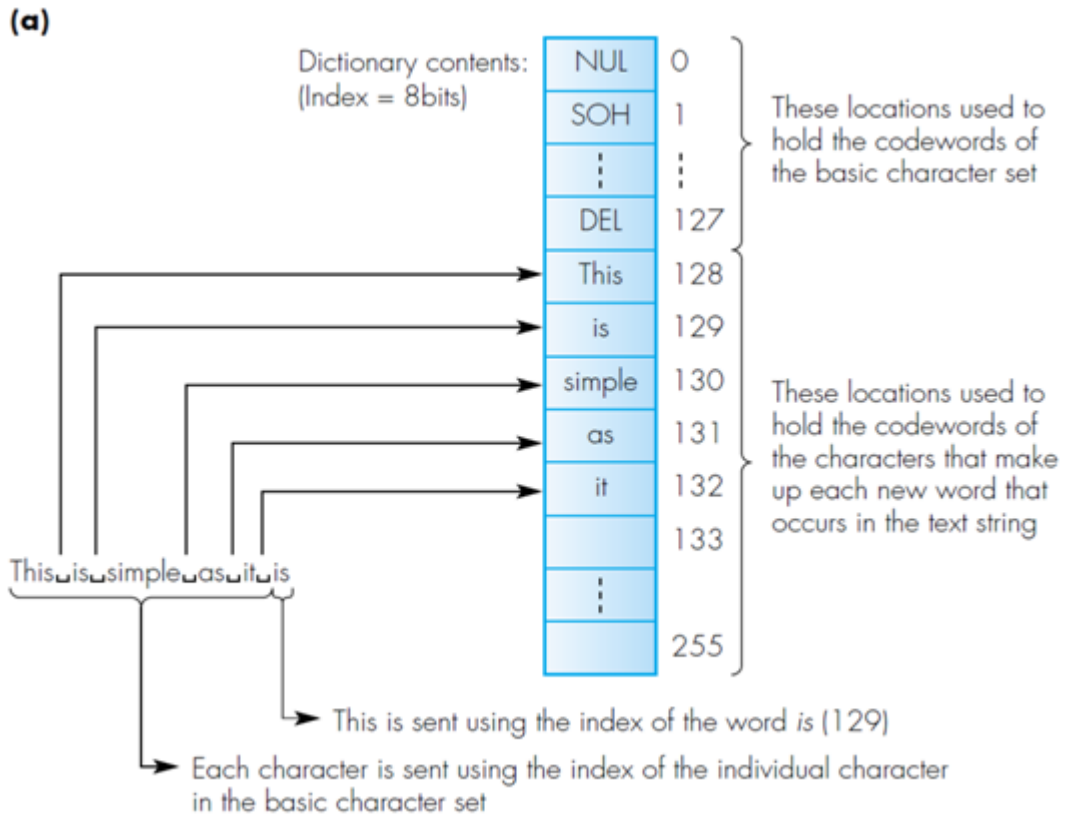


Fig 4 a LZW compression basic principle

(xiii) Similarly the decoder, since it also has the word stored in its dictionary, uses the index to access the string of characters that make up the word.

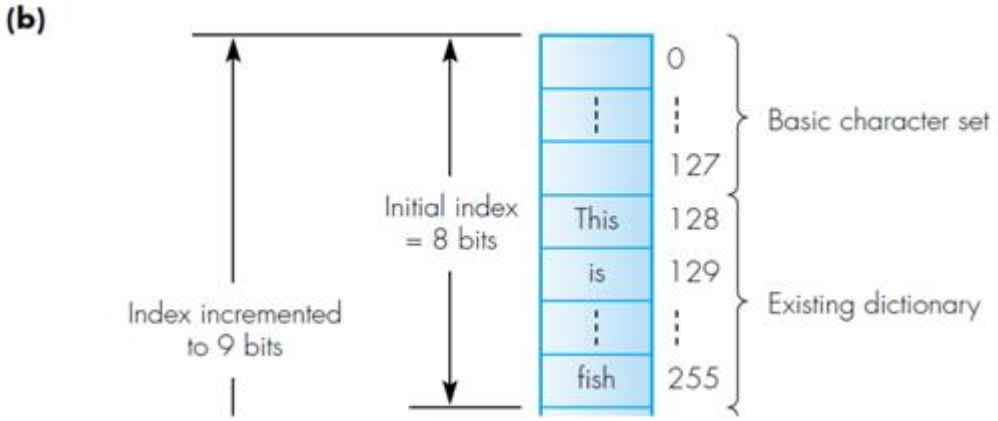
(xiv) After the space character following the second occurrence of the word is, the contents of the dictionary held by both the encoder and the decoder will be as shown in the Figure 1 (a).

(xv) Since this is the second occurrence of the word is, it is transferred using only the index of where it is stored in the dictionary (129).

(xvi) from this example , a key issue in determining the level of compression that is achieved , is the number of entries in the dictionary since this, in turn, determines the number if bits that are required for the index.

(xvii) . For eg , in an application that uses 128 characters in the basic character set , then both the encoder and decoder would start with, say, 256 entries in the dictionary

(xviii) an index/code word length of 8 bits and the dictionary would provide space for the 128 characters in the character set and a further 128 locations for words that occur in the text.



Dynamically extending the number of entries in the dictionary

this number of locations become insufficient, on detecting this, both the encoder and the decoder would double the size of their dictionary to 512 locations.

(xix) An index length of 9 bits and so from this point, the encoder uses the 9 bit code words. However, since the decoder has also doubled the size of its own directory, it expects 9 bit code words from this point. In this way, the number of entries in the dictionary more accurately reflects the number of different words in the text being transferred and hence optimizes the number of bits used for each index/code word. The procedure is shown in diagrammatic form in Figure (b).

- In this example it is assumed that the last entry in the existing table at location 255 is the word fish and the next word in the text that is not currently in the dictionary is pond.

.

UNIT IV – Guaranteed Service Model

Part A

1. **What are the responsibilities of interface and information designers in the development of a multimedia project?**
   - ✓ An interface designer is responsible for: i) creating software device that organizes content, allows users to access or modify content and present that content on the screen, ii) building a user friendlyinterface.
   - ✓ Information designers, who structure content, determine user pathways and feedback and select presentationmedia.
2. **List the features of multimedia.**
   - ✓ A Multimedia system has four basiccharacteristics:

- Multimedia systems must be computercontrolled.
- Multimedia systems areintegrated.
- The information they handle must be representeddigitally.
- The interface to the final presentation of media is usuallyinteractive.

3. **What are the multimedia components?**

   ✓ Text, Audio, Images, Animations, Video and interactive content are the multimediacomponents.

   ✓ The first multimedia element is text. Text is the most common multimediaelement.

4. **Definemultimedia.**

   ✓ 'Multi' means 'many' and 'media' means 'material through which something can be transmitted or send'.

   ✓ Information being transferred by more than one medium is called asmultimedia.

- ✓ It is the combination of text, image, audio, video, animation, graphic & hardware, that can be delivered electronically / digitally which can be accessedinteractively.
- ✓ It is of two types: Linear & Non –Linear.

5. Describe the applications ofmultimedia.

- ✓ Multimedia in Education: It is commonly used to prepare study material for the students and also provide them proper understanding of differentsubjects.
- ✓ Multimedia inEntertainment:

a) Movies: Multimedia used in movies gives a special audio and videoeffect.

b) Games: Multimedia used in games by using computer graphics, animation, videos has changed the gamingexperience.

- ✓ Multimedia inBusiness:

a) Videoconferencing: This system enables to communicate using audio and video between two different locations through theircomputers.

b) Marketing and advertisement: Different advertisement and marketing ideas about any product on television and internet is possible withmultimedia.

6. Write the difference between multimedia andhypermedia.

S.No	Multimedia	Hypermedia
1.	Multimedia is the presentation of media as text, images, graphics, video & audio by the use of computers or the information contentprocessing devices.	Hypermedia is the use of advanced form of hypertext like interconnected systems to store and present text, graphics & other media types where the content is linked to each other by hyperlinks.

2.	Multimedia can be in linear or non-linear content format, but the hypermedia is only in non-linear content format.	Hypermedia is an application of multimedia, hence a subset ofmultimedia.

7. Why multimedia networking is needed

The importance of communications or networking for multimedia lies in the new applications that will be generated by adding networking capabilities to multimedia computers, and hopefully gains in efficiency and cost of ownership and use when multimedia resources are part of distributed computing systems.

8. Mention the applications of multimedianetworking.

- StreamingVideo
- IPtelephony
- Internet Radio
- Teleconferencing
- Interactive games
- Virtualworlds
- Multimediaweb

9. What are the scheduling mechanisms used in multimedia networking?

1. First In First Out (FIFO)
2. Priority Queuing
3. Round Robin and Weigthed Fair Queuing

10. What are the policing criteria used in multimedia networking?

Three policing criteria are used and each differing from the other, according to the time scale over which the packet flows is policed.

1. Average Rate
2. Peak Rate
3. Burst Size

11. Name the two key features of integrated services architecture and define them.

Two key features of integrated service architecture is

1. Reserved Resources
2. Call setup

Reserved Resources:

A router is required to know what amounts of its Resources are already Reserved for ongoing sessions.

Call setup:

Call setup also known as call admission it requires the participation of each router on the path.

12. Define multicast overlay networks.

Multicast overlay network consists of servers scatted throughout the ISP network. These servers and the logical links between them collectively form an overlay network which multicasts traffic from the source to the millions of users.

13. Write the principal characteristics of RSVP.

1. It provides reservations for bandwidth in multicast trees.
2. It is receiver oriented that is, the receiver of a data flow initiates and maintains the resource reservation used for that flow.

14. What are the drawbacks of RTSP?

RTSP does not define compression schemes for audio and video.

RTSP does not define how audio and video are encapsulated in packets for transmission over a network.

RTSP does not restrict how the media player buffers the audio/video.

RTSP does not restrict how streamed media is transported.

15. Difference between RTSP and RSVP.

Sl. No.	RTSP	RSVP
1	It is an out of band protocol	It is a signaling protocol
2	It is used for media player	It is used for the internet
3	It sends a message over TCP or UDP	It sends a message using router buffers

PART B

1. **Discuss about the Best effort service model and its drawback in detail.**

Real-time conversational voice over the Internet is often referred to as **Internet telephony**, since, from the user's perspective, it is similar to the traditional circuit-switched telephone service. It is also commonly called **Voice-over-IP (VoIP)**. In this section we describe the principles and protocols underlying VoIP. Conversational video is similar in many respects to VoIP, except that it includes the video of the participants as well as their voices. To keep the discussion focused and concrete, we focus here only on voice in this section rather than combined voice and video.

The Internet's network-layer protocol, IP, provides best-effort service. That is to say the service makes its best effort to move each datagram from source to destination as quickly as possible but makes no promises whatsoever about getting the packet to the destination within some delay bound or about a limit on the percentage of packets lost. The lack of such guarantees poses significant challenges to the design of real-time conversational applications, which are acutely sensitive to packet delay, jitter, and loss.

In this section, we'll cover several ways in which the performance of VoIP over a best-effort network can be enhanced. Our focus will be on application-layer techniques, that is, approaches that do not require any changes in the network core or even in the transport layer at the end hosts. To keep the discussion concrete, we'll discuss the limitations of best-effort IP service in the context of a specific VoIP example. The sender generates bytes at a rate of 8,000 bytes per second; every 20 msecs the sender gathers these bytes into a chunk. A chunk and a special header (discussed below) are encapsulated in a UDP segment, via a call to the socket interface. Thus, the number of bytes in a chunk is (20 msecs)·(8,000 bytes/sec) = 160 bytes, and a UDP segment is sent every 20 msecs.

If each packet makes it to the receiver with a constant end-to-end delay, then packets arrive at the receiver periodically every 20 msecs. In these ideal conditions, the receiver can simply play back each chunk as soon as it arrives. But unfortunately, some packets can be lost and most packets will not have the same end-to-end delay, even in a lightly congested Internet. For this reason, the receiver must take more care in determining (1) when to play back a chunk, and (2) what to do with a missing chunk.

Packet Loss

Consider one of the UDP segments generated by our VoIP application. The UDP segment is encapsulated in an IP datagram. As the datagram wanders through the network, it passes through router buffers (that is, queues) while waiting for transmission on outbound links. It is possible that one or more of the buffers in the path from sender to receiver is full, in which case the arriving IP datagram may be discarded, never to arrive at the receiving application.

Loss could be eliminated by sending the packets over TCP (which provides for reliable data transfer) rather than over UDP. However, retransmission mechanisms are often considered unacceptable for conversational real-time audio applications such as VoIP, because they increase end-to-end delay [Bolot 1996]. Furthermore, due to TCP congestion control, packet loss may result in a reduction of the TCP sender's transmission rate to a rate that is lower than the receiver's drain rate, possibly leading to buffer starvation. This can have a severe impact on voice intelligibility at the receiver. For these reasons, most existing VoIP applications run over UDP by default. [Baset 2006] reports that UDP is used by Skype unless a user is behind a NAT or firewall that blocks UDP segments (in which case TCP is used).

But losing packets is not necessarily as disastrous as one might think. Indeed, packet loss rates between 1 and 20 percent can be tolerated, depending on how voice is encoded and transmitted, and on how the loss is concealed at the receiver. For example, forward error correction (FEC) can help conceal packet loss. We'll see below that with FEC, redundant information is transmitted along with the original information so that some of the lost original data can be recovered from the redundant information. Nevertheless, if one or more of the links between sender and receiver is severely congested, and packet loss exceeds 10 to 20 percent (for example, on a wireless link), then there is really nothing that can be done to achieve acceptable audio quality. Clearly, best-effort service has its limitations.

End-to-End Delay

End-to-end delay is the accumulation of transmission, processing, and queuing delays in routers; propagation delays in links; and end-system processing delays. For real-time conversational applications, such as VoIP, end-to-end delays smaller than 150 msecs are not perceived by a human listener; delays between 150 and 400 msecs can be acceptable but are not ideal; and delays exceeding 400 msecs can seriously hinder the interactivity in voice conversations. The receiving side of a VoIP application will typically disregard any packets that are delayed more than a certain threshold, for example, more than 400 msecs. Thus, packets that are delayed by more than the threshold are effectively lost.

Packet Jitter

A crucial component of end-to-end delay is the varying queuing delays that a packet experiences in the network's routers. Because of these varying delays, the time from when a packet is generated at the source until it is received at the receiver can fluctuate from packet to packet, as shown in Figure 7.1. This phenomenon is called **jitter**. As an example, consider two consecutive packets in our VoIP application. The sender sends the second packet 20 msecs after sending the first packet. But at the receiver, the spacing between these packets can become greater than 20 msecs. To see this, suppose the first packet arrives at a nearly empty queue at a router, but just before the second packet arrives at the queue a large number of packets from other sources

arrive at the same queue. Because the first packet experiences a small queuing delay and the second packet suffers a large queuing delay at this router, the first and second packets become spaced by more than 20 msecs. The spacing between consecutive packets can also become less than 20 msecs. To see this, again consider two consecutive packets. Suppose the first packet joins the end of a queue with a large number of packets, and the second packet arrives at the queue before this first packet is transmitted and before any packets from other sources arrive at the queue. In this case, our two packets find themselves one right after the other in the queue. If the time it takes to transmit a packet on the router's outbound link is less than 20 msecs, then the spacing between first and second packets becomes less than 20 msecs.

The situation is analogous to driving cars on roads. Suppose you and your friend are each driving in your own cars from San Diego to Phoenix. Suppose you and your friend have similar driving styles, and that you both drive at 100 km/hour, traffic permitting. If your friend starts out one hour before you, depending on intervening traffic, you may arrive at Phoenix more or less than one hour after your friend.

If the receiver ignores the presence of jitter and plays out chunks as soon as they arrive, then the resulting audio quality can easily become unintelligible at the receiver. Fortunately, jitter can often be removed by using **sequence numbers**, **timestamps**, and a **playout delay**, as discussed below.

2. Explain in detail about the Scheduling policies.

First-In-First-Out (FIFO)

Figure 7.17 shows the queuing model abstractions for the FIFO link-scheduling discipline. Packets arriving at the link output queue wait for transmission if the link is currently busy transmitting another packet. If there is not sufficient buffering space to hold the arriving packet, the queue's **packet-discarding policy** then determines whether the packet will be dropped (lost) or whether other packets will be removed from the queue to make space for the arriving packet. In our discussion below, we will ignore packet discard. When a packet is completely transmitted over the outgoing link (that is, receives service) it is removed from the queue.

The FIFO (also known as first-come-first-served, or FCFS) scheduling discipline selects packets for link transmission in the same order in which they arrived at the output link queue. We're all familiar with FIFO queuing from bus stops (particularly in England, where queuing seems to have been perfected) or other service centers, where arriving customers join the back of the single waiting line, remain in order, and are then served when they reach the front of the line.

Figure 7.18 shows the FIFO queue in operation. Packet arrivals are indicated by numbered arrows above the upper timeline, with the number indicating the order

in which the packet arrived. Individual packet departures are shown below the lower timeline. The time that a packet spends in service (being transmitted) is indicated by the shaded rectangle between the two timelines. Because of the FIFO discipline, packets leave in the same order in which they arrived. Note that after the departure of packet 4, the link remains idle (since packets 1 through 4 have been transmitted and removed from the queue) until the arrival of packet 5.

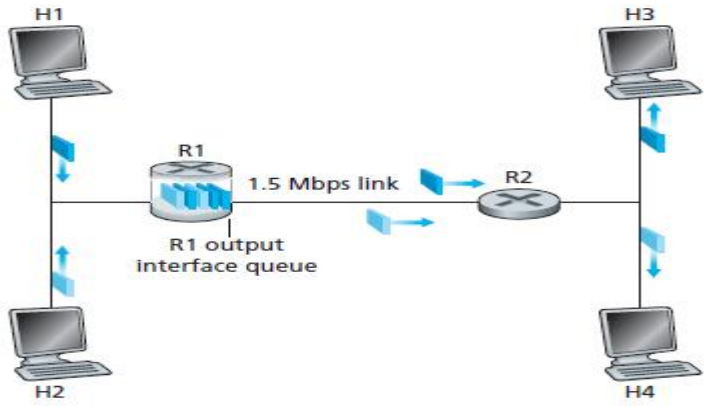


Figure 7.17 ◆ FIFO queuing abstraction

Priority Queuing

Under **priority queuing**, packets arriving at the output link are classified into priority classes at the output queue, as shown in Figure 7.19. As discussed in the previous section, a packet's priority class may depend on an explicit marking that it carries in its packet header (for example, the value of the ToS bits in an IPv4 packet), its source or destination IP address, its destination port number, or other criteria. Each priority class typically has its own queue. When choosing a packet to transmit, the priority queuing discipline will transmit a packet from the highest priority class that has a nonempty queue (that is, has packets waiting for transmission). The choice among packets *in the same priority class* is typically done in a FIFO manner.

Figure 7.20 illustrates the operation of a priority queue with two priority classes. Packets 1, 3, and 4 belong to the high-priority class, and packets 2 and 5 belong to the low-priority class. Packet 1 arrives and, finding the link idle, begins transmission. During the transmission of packet 1, packets 2 and 3 arrive and are queued in the low- and high-priority queues, respectively. After the transmission of packet 1, packet 3 (a high-priority packet) is selected for transmission over packet 2 (which, even though it arrived earlier, is a low-priority packet). At the end of the transmission of packet 3, packet 2 then begins transmission. Packet 4 (a high-priority packet) arrives during the transmission of packet 2 (a low-priority packet). Under a nonpreemptive priority queuing discipline, the transmission of

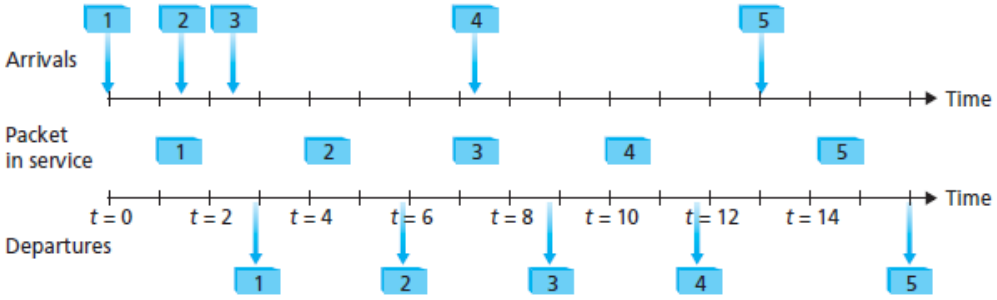


Figure 7.18 ♦ The FIFO queue in operation

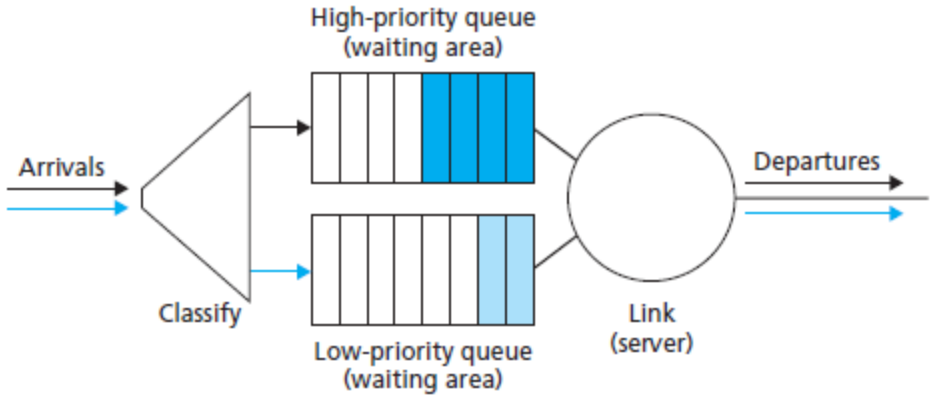


Figure 7.19 ♦ Priority queuing model

a packet is not interrupted once it has begun. In this case, packet 4 queues for transmission and begins being transmitted after the transmission of packet 2 is completed.

Round Robin and Weighted Fair Queuing (WFQ)

Under the **round robin queuing discipline**, packets are sorted into classes as with priority queuing. However, rather than there being a strict priority of service among classes, a round robin scheduler alternates service among the classes. In the simplest form of round robin scheduling, a class 1 packet is transmitted, followed by a class 2 packet, followed by a class 1 packet, followed by a class 2 packet, and so on. A so-called work-conserving queuing discipline will never allow the link to remain idle whenever there are packets (of any class) queued for

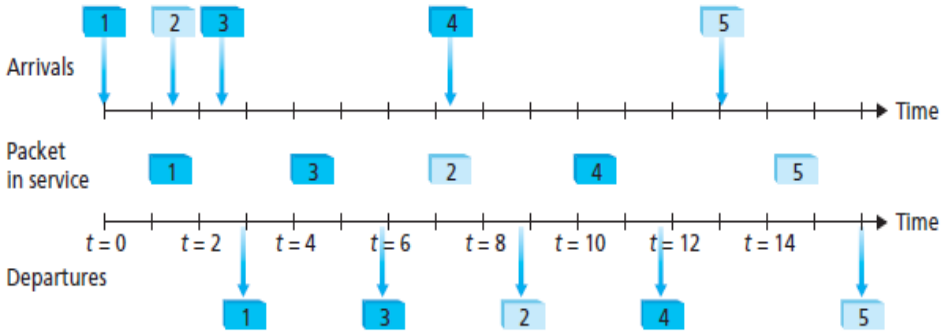


Figure 7.20 ♦ Operation of the priority queue

transmission. A **work-conserving round robin discipline** that looks for a packet of a given class but finds none will immediately check the next class in the round robin sequence.

Figure 7.21 illustrates the operation of a two-class round robin queue. In this example, packets 1, 2, and 4 belong to class 1, and packets 3 and 5 belong to the second class. Packet 1 begins transmission immediately upon arrival at the output queue. Packets 2 and 3 arrive during the transmission of packet 1 and thus queue for transmission. After the transmission of packet 1, the link scheduler looks for a class 2 packet and thus transmits packet 3. After the transmission of packet 3, the scheduler looks for a class 1 packet and thus transmits packet 2. After the transmission of packet 2, packet 4 is the only queued packet; it is thus transmitted immediately after packet 2.

A generalized abstraction of round robin queuing that has found considerable use in QoS architectures is the so-called **weighted fair queuing** (WFQ) discipline [Demers 1990; Parekh 1993]. WFQ is illustrated in Figure 7.22. Arriving packets are classified and queued in the appropriate per-class waiting area. As in round robin scheduling, a WFQ scheduler will serve classes in a circular manner—first serving class 1, then serving class 2, then serving class 3, and then (assuming there are three classes) repeating the service pattern. WFQ is also a work-conserving queuing discipline and thus will immediately move on to the next class in the service sequence when it finds an empty class queue.

WFQ differs from round robin in that each class may receive a *differential* amount of service in any interval of time. Specifically, each class, i, is assigned a weight, w_i. Under WFQ, during any interval of time during which there are class i packets to send, class i will then be guaranteed to receive a fraction of service equal to $w_i/(\sum w_j)$, where the sum in the denominator is taken over all classes that also have packets queued for transmission. In the worst case, even if all classes have queued packets, class i will still be guaranteed to receive a fraction $w_i/(\sum w_j)$ of the

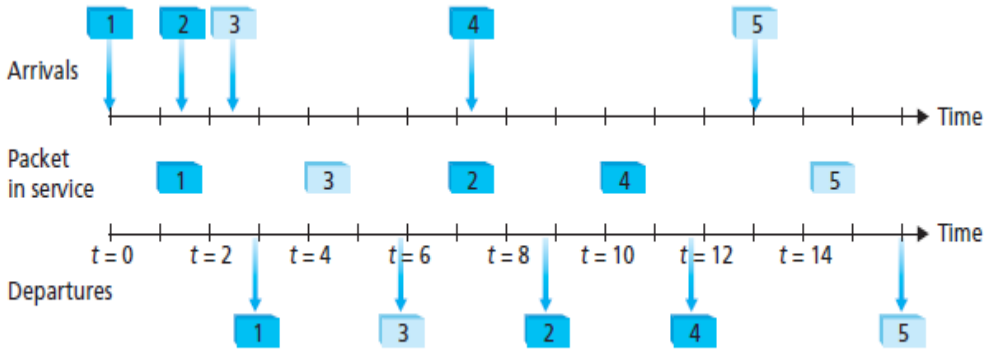


Figure 7.21 ♦ Operation of the two-class round robin queue

bandwidth. Thus, for a link with transmission rate R, class i will always achieve a throughput of at least $R \cdot w_i/(\sum w_j)$. Our description of WFQ has been an idealized one, as we have not considered the fact that packets are discrete units of data and a packet's transmission will not be interrupted to begin transmission of another packet; [Demers 1990] and [Parekh 1993] discuss this packetization issue. As we will see in the following sections, WFQ plays a central role in QoS architectures. It is also available in today's router products [Cisco QoS 2012].

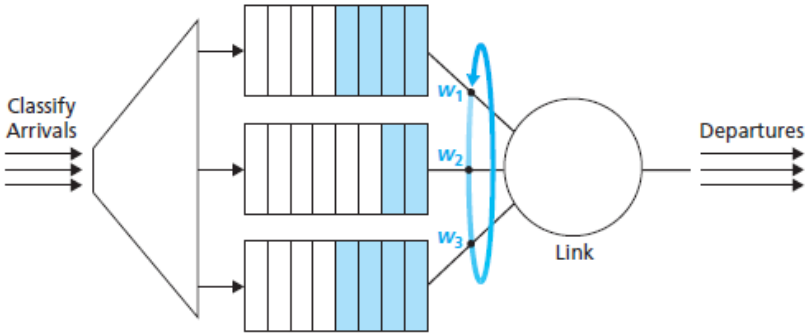


Figure 7.22 ♦ Weighted fair queuing (WFQ)

3. Explain in detail about the Leaky Bucket policies

Policing: The Leaky Bucket

One of our earlier insights was that policing, the regulation of the rate at which a class or flow (we will assume the unit of policing is a flow in our discussion below) is allowed to inject packets into the network, is an important QoS mechanism. But what aspects of a flow's packet rate should be policed? We can identify three important policing criteria, each differing from the other according to the time scale over which the packet flow is policed:

- *Average rate.* The network may wish to limit the long-term average rate (packets per time interval) at which a flow's packets can be sent into the network. A crucial issue here is the interval of time over which the average rate will be policed. A flow whose average rate is limited to 100 packets per second is more constrained than a source that is limited to 6,000 packets per minute, even though both have the same average rate over a long enough interval of time. For example, the latter constraint would allow a flow to send 1,000 packets in a given second-long interval of time, while the former constraint would disallow this sending behavior.

- *Peak rate.* While the average-rate constraint limits the amount of traffic that can be sent into the network over a relatively long period of time, a peak-rate constraint limits the maximum number of packets that can be sent over a shorter period of time. Using our example above, the network may police a flow at an average rate of 6,000 packets per minute, while limiting the flow's peak rate to 1,500 packets per second.
- *Burst size.* The network may also wish to limit the maximum number of packets (the "burst" of packets) that can be sent into the network over an extremely short interval of time. In the limit, as the interval length approaches zero, the burst size limits the number of packets that can be instantaneously sent into the network. Even though it is physically impossible to instantaneously send multiple packets into the network (after all, every link has a physical transmission rate that cannot be exceeded!), the abstraction of a maximum burst size is a useful one.

The leaky bucket mechanism is an abstraction that can be used to characterize these policing limits. As shown in Figure 7.23, a leaky bucket consists of a bucket that can hold up to b tokens. Tokens are added to this bucket as follows. New tokens, which may potentially be added to the bucket, are always being generated at a rate of r tokens per second. (We assume here for simplicity that the unit of time is a second.) If the bucket is filled with less than b tokens when a token is generated, the newly generated token is added to the bucket; otherwise the newly generated token is ignored, and the token bucket remains full with b tokens.

Let us now consider how the leaky bucket can be used to police a packet flow. Suppose that before a packet is transmitted into the network, it must first remove a

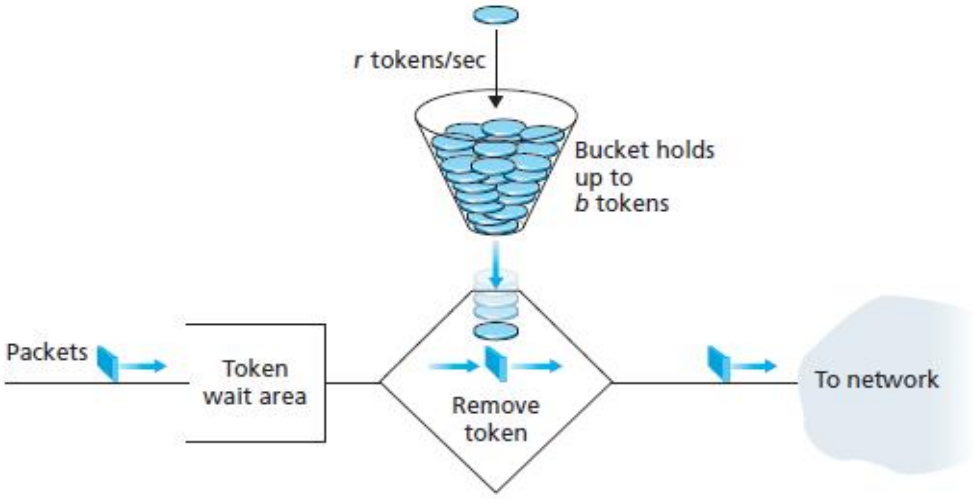


Figure 7.23 ♦ The leaky bucket policer

token from the token bucket. If the token bucket is empty, the packet must wait for a token. (An alternative is for the packet to be dropped, although we will not consider that option here.) Let us now consider how this behavior polices a traffic flow. Because there can be at most b tokens in the bucket, the maximum burst size for a leaky-bucket-policed flow is b packets. Furthermore, because the token generation rate is r, the maximum number of packets that can enter the network of *any* interval of time of length t is $rt + b$. Thus, the token-generation rate, r, serves to limit the long-term average rate at which packets can enter the network. It is also possible to use leaky buckets (specifically, two leaky buckets in series) to police a flow's peak rate in addition to the long-term average rate; see the homework problems at the end of this chapter.

Leaky Bucket + Weighted Fair Queuing = Provable Maximum Delay in a Queue

Let's close our discussion of scheduling and policing by showing how the two can be combined to provide a bound on the delay through a router's queue. Let's consider a router's output link that multiplexes n flows, each policed by a leaky bucket with parameters b_i and r_i, $i = 1, \ldots, n$, using WFQ scheduling. We use the term *flow* here loosely to refer to the set of packets that are not distinguished from each other by the scheduler. In practice, a flow might be comprised of traffic from a single end-to-end connection or a collection of many such connections, see Figure 7.24.

Recall from our discussion of WFQ that each flow, i, is guaranteed to receive a share of the link bandwidth equal to at least $R \cdot w_i/(\sum w_j)$, where R is the transmission

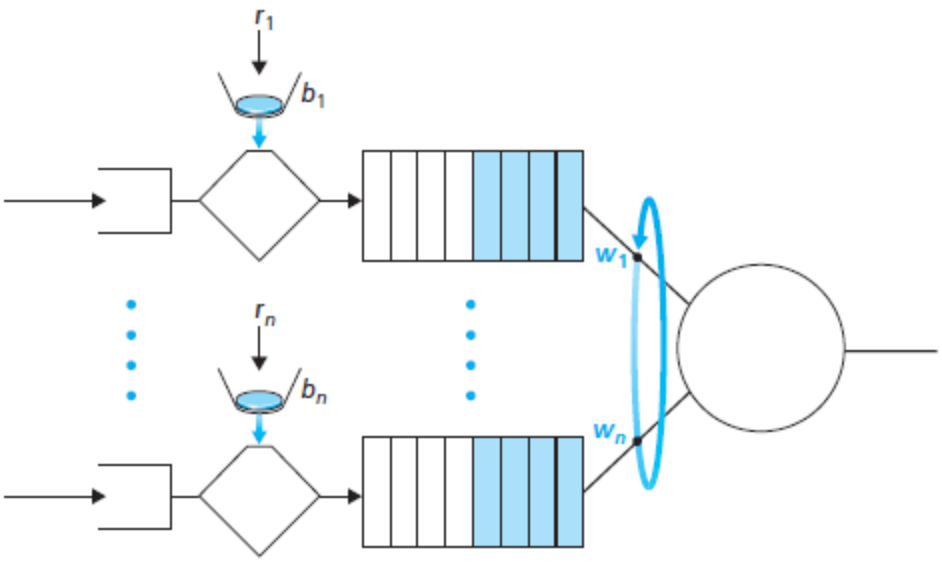


Figure 7.24 ♦ *n* multiplexed leaky bucket flows with WFQ scheduling

rate of the link in packets/sec. What then is the maximum delay that a packet will experience while waiting for service in the WFQ (that is, after passing through the leaky bucket)? Let us focus on flow 1. Suppose that flow 1's token bucket is initially full. A burst of b_1 packets then arrives to the leaky bucket policer for flow 1. These packets remove all of the tokens (without wait) from the leaky bucket and then join the WFQ waiting area for flow 1. Since these b_1 packets are served at a rate of at least $R \cdot w_i/(\sum w_j)$ packet/sec, the last of these packets will then have a maximum delay, $d_{\max,}$ until its transmission is completed, where

$$d_{\max} = \frac{b_1}{R \cdot w_1 / \sum w_j}$$

The rationale behind this formula is that if there are b_1 packets in the queue and packets are being serviced (removed) from the queue at a rate of at least $R \cdot w_1/(\sum w_j)$ packets per second, then the amount of time until the last bit of the last packet is transmitted cannot be more than $b_1/(R \cdot w_1/(\sum w_j))$. A homework problem asks you to prove that as long as $r_1 < R \cdot w_1/(\sum w_j)$, then $d_{\max}$ is indeed the maximum delay that any packet in flow 1 will ever experience in the WFQ queue.

4. Explain in detail about the Resource Reservation and Call admission.

we have seen that packet marking and policing, traffic isolation, and link-level scheduling can provide one class of service with better performance than another. Under certain scheduling disciplines, such as priority scheduling, the lower classes of traffic are essentially "invisible" to the highest-priority class of traffic. With proper network dimensioning, the highest class of service can indeed

achieve extremely low packet loss and delay—essentially circuit-like performance. But can the network *guarantee* that an ongoing flow in a high-priority traffic class will continue to receive such service throughout the flow's duration using only the mechanisms that we have described so far? It cannot. In this section, we'll see why yet additional network mechanisms and protocols are required when a hard service guarantee is provided to individual connections.

Let's return to our scenario from Section 7.5.2 and consider two 1 Mbps audio applications transmitting their packets over the 1.5 Mbps link, as shown in Figure 7.27. The combined data rate of the two flows (2 Mbps) exceeds the link

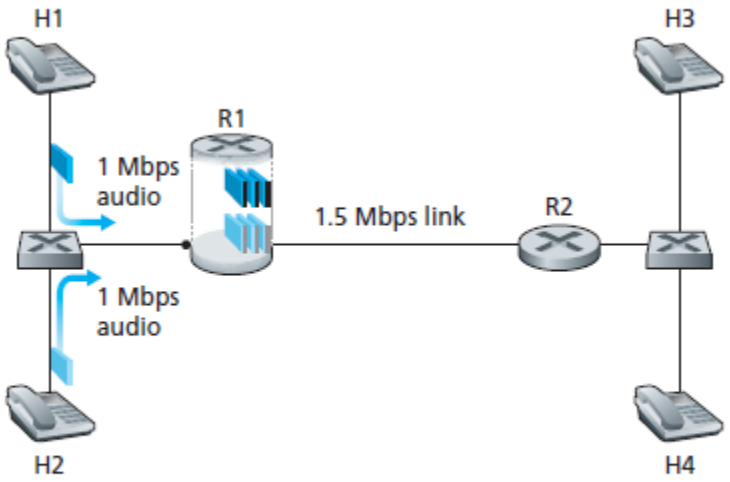


Figure 7.27 ♦ Two competing audio applications overloading the R1-to-R2 link

capacity. Even with classification and marking, isolation of flows, and sharing of unused bandwidth (of which there is none), this is clearly a losing proposition. There is simply not enough bandwidth to accommodate the needs of both applications at the same time. If the two applications equally share the bandwidth, each application would lose 25 percent of its transmitted packets. This is such an unacceptably low QoS that both audio applications are completely unusable; there's no need even to transmit any audio packets in the first place.

Given that the two applications in Figure 7.27 cannot both be satisfied simultaneously, what should the network do? Allowing both to proceed with an unusable QoS wastes network resources on application flows that ultimately provide no utility to the end user. The answer is hopefully clear—one of the application flows should be blocked (that is, denied access to the network), while the other should be allowed to proceed on, using the full 1 Mbps needed by the application. The telephone network is an example of a network that performs such call blocking—if the required resources (an end-to-end circuit in the case of the telephone network) cannot be allocated to the call, the call is blocked (prevented from entering the network) and a busy signal is returned to the user. In our example, there is no gain in allowing a flow into the network if it will not receive a sufficient QoS to be considered usable. Indeed, there is a cost to admitting a flow that does not receive its needed QoS, as network resources are being used to support a flow that provides no utility to the end user.

By explicitly admitting or blocking flows based on their resource requirements, and the source requirements of already-admitted flows, the network can guarantee that admitted flows will be able to receive their requested QoS. Implicit in the need to provide a guaranteed QoS to a flow is the need for the flow to declare its QoS requirements. This process of having a flow declare its QoS requirement, and then having the network either accept the flow (at the required QoS) or block the flow is referred to as the **call admission** process. This then is our fourth insight (in addition to the three earlier insights from Section 7.5.2) into the mechanisms needed to provide QoS.

- *Resource reservation.* The only way to *guarantee* that a call will have the resources (link bandwidth, buffers) needed to meet its desired QoS is to explicitly

allocate those resources to the call—a process known in networking parlance as **resource reservation**. Once resources are reserved, the call has on-demand access to these resources throughout its duration, regardless of the demands of all other calls. If a call reserves and receives a guarantee of x Mbps of link bandwidth, and never transmits at a rate greater than x, the call will see loss- and delay-free performance.

Call admission. If resources are to be reserved, then the network must have a mechanism for calls to request and reserve resources. Since resources are not infinite, a call making a call admission request will be denied admission, that is, be blocked, if the requested resources are not available. Such a call admission is performed by the telephone network—we request resources when we dial a number. If the circuits (TDMA slots) needed to complete the call are available, the circuits are allocated and the call is completed. If the circuits are not available, then the call is blocked, and we receive a busy signal. A blocked call can try again to gain admission to the network, but it is not allowed to send traffic into the network until it has successfully completed the call admission process. Of course, a router that allocates link bandwidth should not allocate more than is available at that link. Typically, a call may reserve only a fraction of the link's bandwidth, and so a router may allocate link bandwidth to more than one call. However, the sum of the allocated bandwidth to all calls should be less than the link capacity if hard quality of service guarantees are to be provided.

- *Call setup signaling.* The call admission process described above requires that a call be able to reserve sufficient resources at each and every network router on its source-to-destination path to ensure that its end-to-end QoS requirement is met. Each router must determine the local resources required by the session, consider the amounts of its resources that are already committed to other ongoing sessions, and determine whether it has sufficient resources to satisfy the per-hop QoS requirement of the session at this router without violating local QoS guarantees made to an already-admitted session. A signaling protocol is needed to coordinate these various activities—the per-hop allocation of local resources, as well as the overall end-to-end decision of whether or not the call has been able to reserve sufficient resources at each and every router on the end-to-end path. This is the job of the **call setup protocol**, as shown in Figure 7.28. The **RSVP protocol** [Zhang 1993, RFC 2210] was proposed for this purpose within an Internet architecture for providing quality-of-service guarantees. In ATM networks, the Q2931b protocol [Black 1995] carries this information among the ATM network's switches and end point.

Despite a tremendous amount of research and development, and even products that provide for per-connection quality of service guarantees, there has been almost no extended deployment of such services. There are many possible reasons. First and foremost, it may well be the case that the simple application-level mechanisms that we studied in Sections 7.2 through 7.4, combined with proper

network dimensioning (Section 7.5.1) provide "good enough" best-effort network service for multimedia applications. In addition, the added complexity and cost of deploying and managing a network that provides per-connection quality of service guarantees may be judged by ISPs to be simply too high given predicted customer revenues for that service.

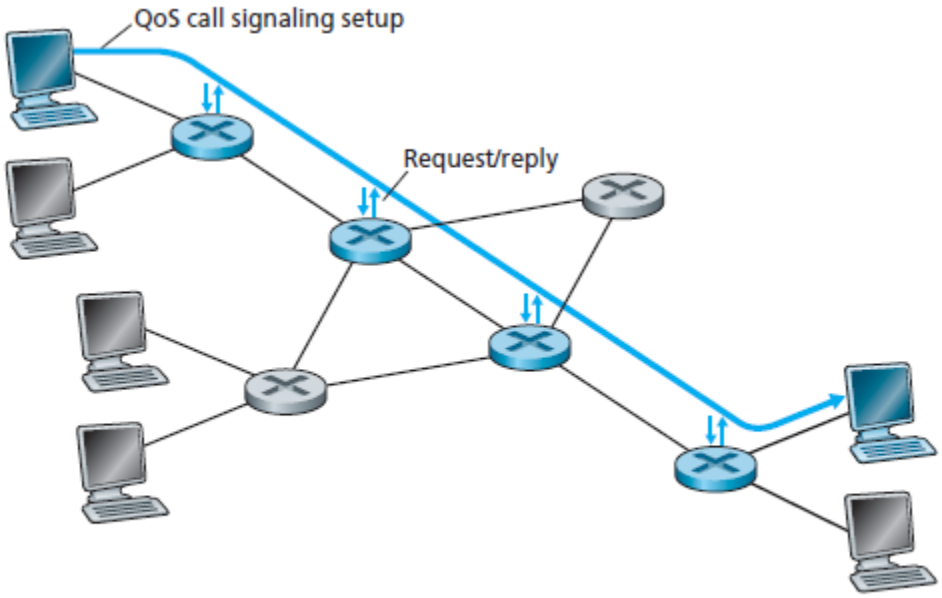


Figure 7.28 ♦ The call setup process

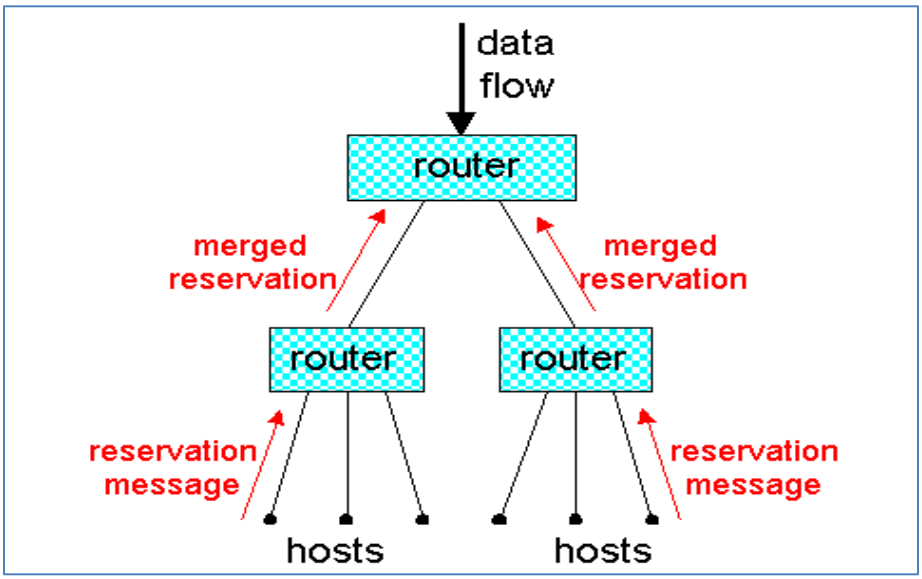
5. Explain in detail about the RSVP Protocol with suitable diagram.

The RSVP protocol allows applications to reserve bandwidth for their data flows. It is used by a host, on the behalf of an application data flow, to request a specific amount of bandwidth from the network. RSVP is also used by the routers to forward bandwidth reservation requests. To implement RSVP, RSVP software must be present in the receivers, senders, and routers. The two principle characteristics of RSVP are:

1. It provides reservations for bandwidth in multicast trees (unicast is handled as a special case).

2. It is receiver-oriented, i.e., the receiver of a data flow initiates and maintains the resource reservation used for that flow.

These two characteristics are illustrated in Figure:.



The above diagram shows a multicast tree with data flowing from the top of the tree to six hosts. Although data originates from the sender, the reservation messages originate from the receivers. When a router forwards a reservation message upstream towards the sender, the router may merge thereservation message with other reservation messages arriving from downstream.Before discussing RSVP in greater detail, we need to recall the notion of a **session**. As with RTP, a session can consist of multiple multicast data flows. Each sender in a session is the source of one or more data flows; for example, a sender might be the source of a video data flow and an audio data flow. Each data flow in a session has the same multicast address. To keep the discussion concrete, we assume that routers and hosts identify the session to which a packet belongs by the packet's multicast address. This assumption is somewhat restrictive; the actual RSVP specification allows for more general methods to identify a session. Within a session, the data flow to which a packet belongs also needs to beidentified. This could be done, for example, with the flow identifier field in IPv6.

WhatRSVPisNot

We emphasize that the RSVP standard does not specify how the network provides the reserved bandwidth to the data flows. It is merely a protocol that allows the applications to reserve the necessary link bandwidth. Once the reservations are in

place, it is up to the routers in the Internet to actually provide the reserved bandwidth to the data flows. This provisioning is done with the scheduling mechanisms. It is also important to understand that RSVP is not a routing protocol -- it does not determine the links in which the reservations are to be made. Instead it depends on an underlying routing protocol (unicast or multicast) to determine the routes for the flows. Once the routes are in place, RSVP can reserve bandwidth in the links along these routes. (We shall see shortly that when a route changes, RSVP re- reserves resources.) And once the reservations are in place, the routers' packet schedulers can actually provide the reserved bandwidth to the data flows. Thus, RSVP is only one piece - albeit an important piece - in the QoS guaranteee puzzle.RSVP is sometimes referred to as a *signaling protocol*. By this it is meant that RSVP is a protocol that allows hosts to establish and tear-down reservations for data flows. The term "signaling protocol" comes from the jargon of the circuit-switched telephony community.

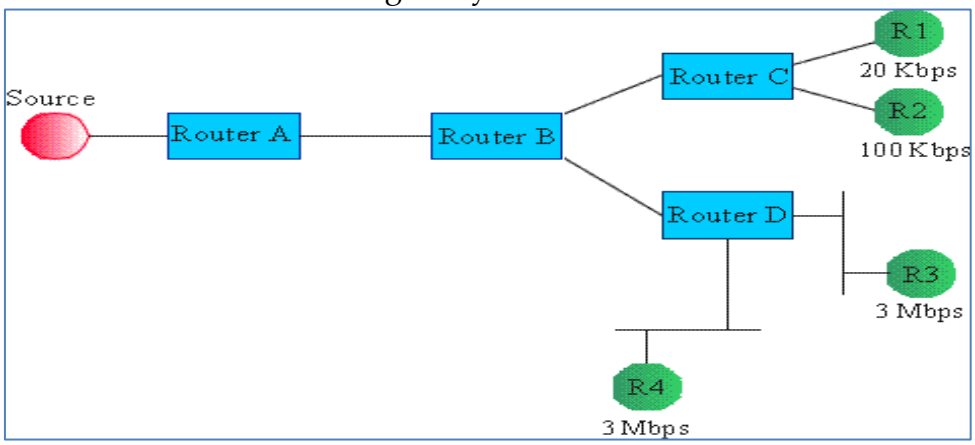
HeterogeneousReceivers

Some receivers can receive a flow at 28.8 Kbps, others at 128 Kbps, and yet others at 10 Mbps or higher. This heterogeneity of the reservers poses an interesting question. If a sender is multicasting a video to a group of heterogeneous receivers, should the sender encode the video for low quality at 28.8Kbps, for medium quality at 128 Kbps, or for high quality at 10 Mbps? If the video is encoded at 10Mbps, then only the users with 10 Mbps access will be able to watch the video. On the other hand, if the video is encoded at 28.8 kbps, then the 10 Mbps users will have to see a low-quality image when they know they can something much better.To resolve this dilemma it is often suggested that video and audio be encoded in layers. For example, a video might be encoded into two layers: a base layer and an enhancement layer. The base layer could have a rate of 20 Kbps whereas the enhancement layer could have a rate of 100 Kbps; in this manner receivers with 28.8 access could receive the low-quality base-layer image, and receivers with 128 Kbps could receive both layers to construct a high-quality image.

We note that the sender does not have to know the receiving rates of all the receivers. It only needs to know the maximum rate of the all its receivers. The sender encodes the video or audio into multiple layers and sends all the layers up to the maximum rate into multicast tree. The receivers pick out the layers that are appropriate for their receiving rates. In order to not excessively waste bandwidth in the network's links, the heterogeneous receivers must communicate to the network the rates they can handle. We shall see that RSVP gives foremost attention to the issue of reserving resources for heterogeneous receivers.

Example:

Let us first describe RSVP in the context of a concrete one-to-many multicast example. Suppose there is a source that is transmitting into the Internet the video of a major sporting event. This session has been assigned a multicast address, and the source stamps all of its outgoing packets with this multicast address. Also suppose that an underlying multicast routing protocol has established a multicast tree from the sender to four receivers as shown below; the numbers next to the receivers are the rates at which the receivers want to receive data. Let us also assume that the video is layered encoded to accommodate this heterogeneity of receiver rates.



Crudely speaking, RSVP operates as follows for this example. Each receiver sends a reservation message upstream into the multicast tree. This reservation message specifies the rate at which the receiver would like to receive the data from the source. When the reservation message reaches a router, the router adjusts its packet scheduler to accommodate the reservation. It then sends a reservation upstream. The amount of bandwidth reserved upstream from the router depends on the bandwidths reserved downstream. In the example in Figure 6.9-2, receivers R1, R2, R3 and R4 reserve 20 kbps, 120 kbps, 3 Mbps and 3 Mbps, respectively. Thus router D's downstream receivers request a maximum of 3Mbps. For this one-to-many transmission, Router D sends a reservation message to Router B requesting that Router B reserve 3 Mbps on the link between the two routers. Note that only 3 Mbps is reserved and not 3+3=6 Mbps; this is because receivers R3 and R4 are watching the same sporting event, so there reservations may be merged. Similarly, Router C requests that Router B reserve 100 Kbps on the link between routers B and C; the layered encoding ensures that receiver R1's 20 Kbps stream is included in the 100 Mbps stream. Once Router B receives the reservation message from its downstream routers and passes the reservations to its schedulers, it sends a new reservation message to its upstream router, Router A. This message reserves 3 Mbps of

bandwidth on the link from Router A to Router B, which is again the maximum of the downstream reservations.

We see from this first example that RSVP is **receiver-oriented**, i.e., the receiver of a data flow initiates and maintains the resource reservation used for that flow. Note that each router receives a reservation message from each of its downstream links in the multicast tree and sends only one reservation message into its upstream link.

UNIT V – Multimedia Communication

Part A

1. **Define packet jitter.**

   JitterinIPnetworksisthevariationinthelatencyonapacketflowbetweentwosystems, when some packets take longer to travel from one system totheother. A jitterbuffer (or de-jitter buffer) can mitigate the effects of jitter, either in the network on a router or switch, or on a computer.

2. **What is meant by RSVP.**

   RSVP (Resource Reservation Protocol) is a set of communication rules that allows channels or paths on the Internet to be reserved for the multicast(one source to many receivers) transmission of video and other high-bandwidthmessages. RSVP is part of the Internet Integrated Service (IIS) model, which ensures best-effort service, real-time service, and controlled link-sharing.

3. **Define any 4 quality of service parameters related to multimedia data transmission.**

   Decompressin , Jitter removal

   Error correction , GUI

4. **What are the limitations of best effort service**

   The limitations of best-effort service are packet loss, excessive end-to-end delay and packet jitter.

5. **What is meant by streaming**

   Streaming media is video or audio content sent in compressed form over the internet & played immediately. It avoids the process of saving the data to the hard. By streaming, a user need not wait to download a file to play it.

6. **Give the applications of real time streaming protocol**

Real Time Streaming Protocol(RTSP) is used by the client application to communicate to the server information such as the requesting of media file, type of clients applications, mechanism of delivery of file & other actions like resume, pause, fast- forward & rewind. It is mostly used in entertainment & communication system to control streaming mediaservers.

7. **Write the shortcomings of integrated services**

The shortcomings of integrated service(intserv) is that, the per-flow resource reservation may give significant workload to routers & also it does not allow more qualitative definitions of service distinctions.

PART - B

1. **Explain in detail the network architecture and protocol design of SIP.**

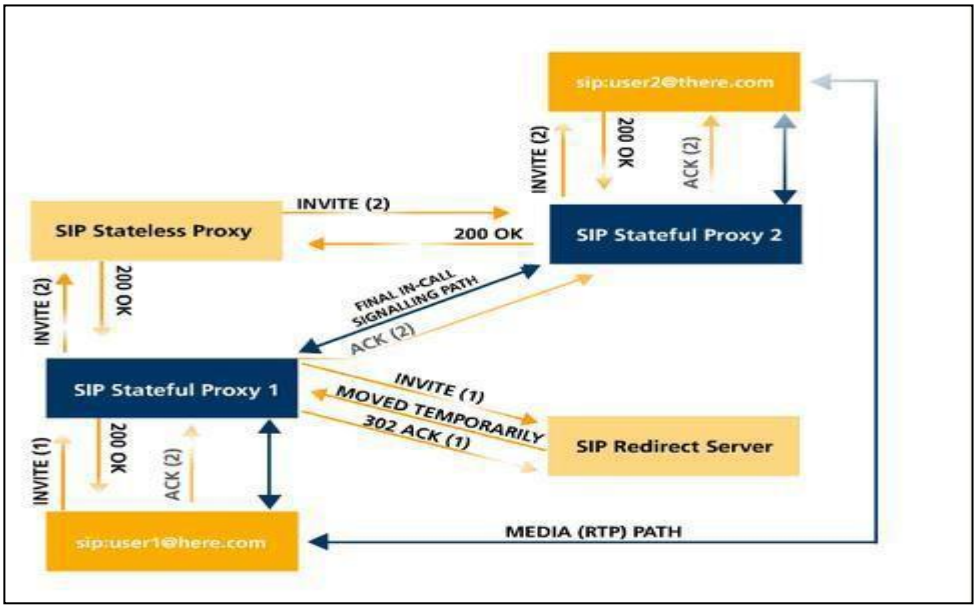
SIP

SIP (Session Initiation Protocol) is a protocol to initiate sessions

- It is an application layer protocol usedto
  - establish
  - modify
  - terminate
- It supports name mapping and redirection servicestransparently

Used by a UA to indicate its current IP address and the URLs for which it would like to receive calls.

- INVITE
- initiatesessions(sessiondescriptionincludedinthemessagebody encoded usingSDP)
- ACK
- confirms sessionestablishment
- BYE
- terminatessessions
- CANCEL
- cancels a pendingINVITE
- REGISTER
- binds a permanent address to a currentlocation
- OPTIONS
- capabilityinquiry
- Other extensions have beenstandardized
  - e.g. INFO, UPDATE, MESSAGE, PRACK, REFER,etc.

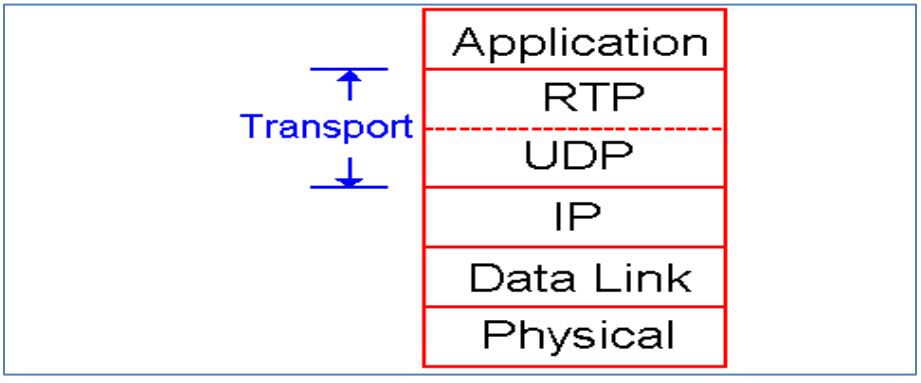


2. Explain in detail about the RTP with necessary diagram.

Welearnedthatthesendersideofamultimediaapplicationappendsheaderfieldstotheaudio/videochunksbeforepassingthechunkstothetransportlayer.Theseheaderfieldsincludesequencenumbersandtimestamps.Sincemostallmultimedianetworkingapplicationscanmakeuseofsequencenumbersandtimestamps,itisconvenienttohaveastandardizedpackettstructurethatincludesfieldsforaudio/videodata,sequencenumberandtimestamp,aswellasotherpotentiallyusefulfields.RTPcanbeusedfortransportingcommonformatssuchasWAVorGSMforsoundandMPEG1andMPEG2forvideo.Itcanalsobeusedfortransportingproprietary soundandvideoformats.ToprovideareadableintroductiontoRTPandtoitscompanionprotocol,RTCP.Wealsodiscusstheroleof RTPinthe H.323standardforreal-timeinteractiveaudioandvideoconferencing.

RTPBasics

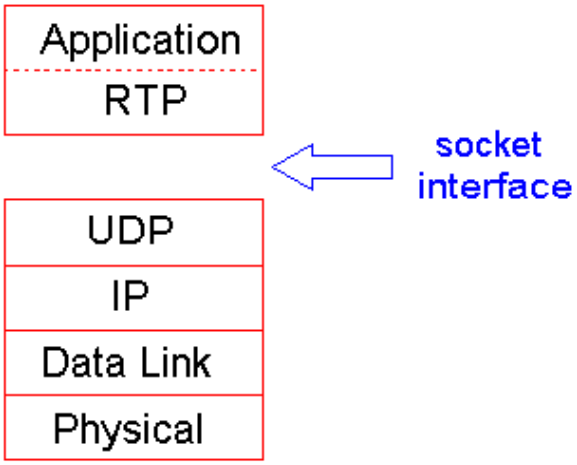
RTPtypicallyrunsontopofUDP.Specifically,audioorvideochunksofdata,generatedbythesendingsideofamultimediaapplication,areencapsulatedinRTPpackets,andeachRTPpacketisinturnencapsulatedinaUDPsegment.BecauseRTPprovidesservices(timestamps,sequence numbers,etc.)tothemultimediaapplication,RTPcanbeviewedasasublayerofthetransportlayer,asshowninFigure.



FigureRTPcanbeviewedasasublayerofthetransportlayer.

Fromtheapplicationdeveloper'sperspective,however,RTPisnotpartofthetransportlayerbutinsteadpartoftheapplicationlayer.Thisisbecause

thedevelopermustintegrateRTPintotheapplication.Specifically,forthesendersideoftheapplication,thedevelopermustwritecodeintothe applicationwhichcreatestheRTPencapsulatingpackets;theapplicationthensendstheRTPpacketsintoaUDPsocketinterface.Similarly,atthe receiversideoftheapplication,theRTPpacketsentertheapplicationthroughaUDPsocketinterface;thedeveloperthereforemustwritecodeinto theapplicationthatextractsthemediachunksfromtheRTPpackets.ThisisillustratedinFigure .



Fromadeveloper'sperspective,RTPispartoftheapplicationlayer.

AsanexampleconsiderusingRTPtotransportvoice.SupposethevoicesourceisPCM encoded(i.e.,sampled,quantized,anddigitized)at64kbps. Furthersupposethattheapplicationcollectstheencodeddatain20msecchunks,i.e,160bytesinachunk.Theapplicationprecedeseachchunkof theaudiodatawithan**RTPheader,**whichincludesthetypeofaudioencoding,asequencenumberandatimestamp.Theaudiochunkalongwith the RTPheaderformthe**RTPpacket**.TheRTPpacketisthensentintotheUDPsocketinterface,whereitisencapsulatedinaUDPpacket.Atthe receiverside,theapplicationreceivestheRTPpacketfromitssocketinterface.Theapplicationextractstheaudiochunkfromthe RTPpacket,and usestheheaderfieldsoftheRTPpackettoproperlydecodeandplaybacktheaudiochunk.

IfanapplicationincorporatesRTP--insteadofaproprietaryschemetoprovidepayloadtype,sequencenumbersortimestamps--thenthe applicationwillmoreeasilyinteroperatewithothernetworkingapplications.Forexample,

iftwodifferentcompaniesdevelopInternetphone softwareandtheybothincorporateRTPintotheirproduct,theremaybesomehopethatauserusingoneoftheInternetphoneproductswillbeable tocommunicatewithauserusingtheotherInternetphoneproduct.Attheendofthissection weshallseethatRTPhasbeenincorporatedintoan importantpartofanInternettelephonystandard.

ItshouldbeemphasizedthatRTPinitselfdoesnotprovideanymechanismtoensuretimelydeliveryofdataorprovideotherqualityofservice guarantees;itdoesnotevenguaranteedeliveryofpacketsorpreventout-of-orderdeliveryofpackets.Indeed,RTPencapsulationisonlyseenatthe endsystems--itisnotseenbyintermediaterouters.Routersdonotdistinguishbetween IPdatagramsthatcarryRTPpacketsandIPdatagramsthat don't.RTPallowseachsource(forexample,acameraoramicrophone)tobeassigneditsownindependentRTPstreamofpackets.Forexample,fora videoconferencebetweentwoparticipants,fourRTPstreamscouldbeopened:twostreams fortransmittingtheaudio(oneineachdirection)andtwo streamsforthevideo(again,oneineachdirection).However,manypopularencodingtechniques--includingMPEG1andMPEG2--bundlethe audioandvideointoasinglestreamduringtheencodingprocess.Whentheaudioandvideo arebundledbytheencoder,thenonlyoneRTPstreamisgeneratedineachdirection.

RTPpacketsarenotlimitedtounicastapplications.Theycanalsobesentoverone-to-manyandmany-to-manymulticasttrees.Foramany-to-many multicastsession,allofthesendersandsourcesinthesessiontypicallysendtheirRTPstreamsintothesamemulticasttreewiththesamemulticast address.RTPmulticaststreamsbelongingtogether,suchasaudioandvideostreamsemanatingfrommultiplesendersinavideoconference application,belongtoanRTPsession.

RTPPacketHeaderFields

AsshownintheFigure,thefourprinciplepacketheaderfieldsarethepayloadtype,sequencenumber,timestampandthesourceidentifier.

RTP Header

ThepayloadtypefieldintheRTPpacketisseven-bitslong.Thus2^7or128differentpayloadtypescanbesupportedbyRTP.Foranaudiostream, thepayloadtypefieldisusedtoindicatethetypeofaudioencoding(e.g.,PCM,adaptivedeltamodulation,linearpredictiveencoding)thatisbeing used.Ifasenderdecidestochangetheencodinginthemiddleofasession,thesendercaninformthereceiverofthechangethroughthispayload typefield.ThesendermaywanttochangetheencodinginordertoincreasetheaudioqualityortodecreasetheRTPstreambitrate.Figure listssomeoftheaudiopayloadtypescurrentlysupportedbyRTP

PayloadTypeNumber	AudioFormat	SamplingRate	Through
0	PCMmu-law	8KH	64Kbps
1	101	8KH	4.8Kbps
3	GS	8KH	13Kbps
7	LP	8KH	2.4Kbps
9	G.72	8KH	48-
14	MPEGAudio	90KH	-
15	G.72	8KH	16Kbps

FigureSomeaudiopayloadtypessupportedby RTP.

Foravideostreamthepayloadtypecanbeusedtoindicatethetypeofvideoencoding(e.g., motionJPEG,MPEG1,MPEG2,H.231).Again,the sendercanchangevideoencodingon-the-flyduringasession.Figure6.4-5listssomeofthevideopayloadtypescurrentlysupportedbyRTP.

PayloadTypeNumber	VideoFormat
26	MotionJPEG
31	H.261

32	MPEG1video
33	MPEG2video

FigureSomevideopayloadtypessupported byRTP.

SequenceNumberField

Thesequencenumberfieldis16-bitslong.Thesequencenumberincrementsbyoneforeach RTPpacketsent,andmaybeused bythereceiverto detectpacketlossandtorestorepacketsequence.Forexampleifthereceiversideoftheapplication receivesastreamofRTPpacketswithagap betweensequencenumbers86and89,thenthereceiverknowsthatpackets87and88werelost.Thereceivercanthenattempttoconcealthelost data.

TimestampField

Thetimestampfieldis32byteslong.Itreflectsthesamplinginstantofthefirstbyteinthe RTPdatapacket.Aswesawinthepreviussection,the receivercanusethetimestampsinordertoremovepacketjitterintroducedinthenetworkand toprovidesynchronousplayoutatthereceiver.The timestampisderivedfromasamplingclockatthesender.Asanexample,foraudiothetimestampclockincrementsbyoneforeachsamplingperiod(forexample,each125usecsfora8KHzsamplingclock);iftheaudioapplicationgenerateschunksconsistingof160encodedsamples,then thetimestampincreasesby160foreachRTPpacketwhenthesourceisactive.Thetimestamp clockcontinuestoincreaseataconstantrateevenif thesourceisinactive.

SynchronizationSourceIdentifier(SSRC)

TheSSRCfieldis32bitslong.ItidentifiesthesourceoftheRTPstream.Typically,eachstreaminaRTPsessionhasadistinctSSRC.TheSSRCis nottheIPaddressofthesender,butinsteadanumberthatthesourceassignsrandomlywhenthe newstreamisstarted.Theprobabilitythattwo streamsgetassignedthesameSSRCisverysmall.

3. Discuss about the RTCP with necessary diagrams.

AprotocolwhichamultimedianetworkingapplicationcanuseinconjunctionwithRTP. TheuseofRTCPis particularlyattractivewhenthenetworkingapplicationmulticastsaudioorvideotomultiplereceiversfromoneormoresenders.AsshowninFigure,RTCPpacketsaretransmittedbyeachparticipantinanRTPsessiontoallotherparticipantsinthesession.TheRTCPpacketsaredistributedtoalltheparticipantsusingIPmulticast.ForanRTPsession,typicallythereisasingle multicastaddress,andallRTPandRTCPpacketsbelonging tothesessionusethemulticastaddress.RTPandRTCPpacketsaredistinguishedfromeachotherthroughtheuseofdistinctportnumbers.

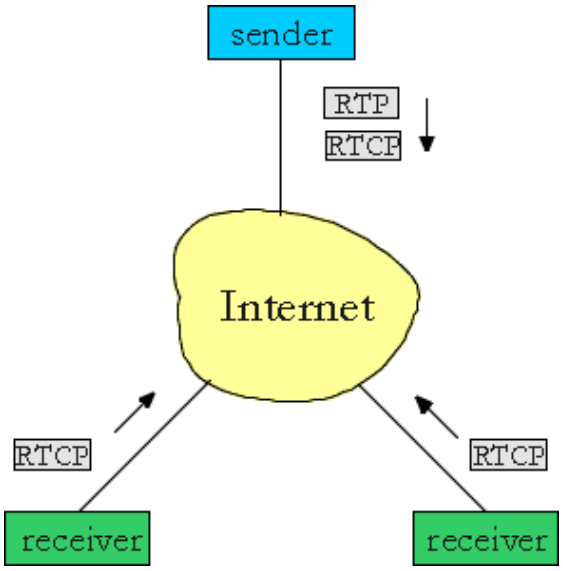


Figure. BothsendersandreceiverssendRTCPmessages.

RTCPpacketsdonotencapsulatechunksofaudioorvideo.Instead,RTCPpacketsare sentperiodicallyandcontainsenderand/orreceiverreports thatannouncestatisticsthatcanbeusefultotheapplication.Thesestatisticsincludenumberofpacketssent,numberofpacketslostandinterarrival jitter.TheRTPspecification[RFC1889]doesnotdictatewhattheapplicationshoulddowiththeseedbackinformation.Itisuptotheapplication developertodecidewhatitwantstodowiththefeedbackinformation.Senderscanusethefeedbackinformation,forexample,tomodifytheir

transmissionrates.Thefeedbackinformationcanalsobeusedfordiagnosticpurposes;fore xample,receiverscandeterminewhetherproblemsare local,regionalorglobal.

RTCPPacket Types

Receiverreceptionpackets

ForeachRTPstreamthatareceiverreceivesaspartofasession,thereceivergeneratesar eceptionreport.Thereceiveraggregatesitsreception reportsintoasingleRTCPpacket.Thepacketisthensentintomulticasttreethatconnectstoge therallttheparticipantsinthesession.Thereception reportincludesseveralfields,themostimportantofwhicharelistedbelow.

- ❍TheSSRCoftheRTPstreamforwhichthereceptionreportisbeinggenerated.
- ❍Thefractionofpacketslostwithin theRTPstream.Eachreceivercalculatesthenu mberofRTPpacketslostdividedbythenumberof RTPpacketssentaspartofthestream.Ifasenderreceivesreceptionreportsindica tingthatthereceiversarereceivingonlyasmall fractionofthesender'stransmittedpackets,thesendercanswitchtoalowerenco dingrate,therebydecreasingthecongestioninthe network,whichmayimprovethereceptionrate.
- ❍Thelastsequencenumberreceivedinthestream ofRTPpackets.
- ❍Theinterarrivaljitter,whichiscalculatedastheaverageinterarrivaltimebetweens uccessivepacketsintheRTPstream.

Senderreportpackets

ForeachRTPstreamthatasenderistransmitting,thesendercreatesandtransmitsRTCPsen der-reportpackets.Thesepacketsincludeinformation abouttheRTPstream,including:

- •TheSSRCoftheRTPstream.
- •Thetimestampandwall-clocktimeofthemostrecentlygeneratedRTPpacketinthestream

- Thenumberofpacketssentinthestream.
- Thenumberofbytessentinthestream.

Thesenderreportscanbeusedtosynchronizedifferentmediastreamswithina RTPsession.Forexample,consideravideoconferencingapplication forwhicheachsendergeneratestwoindependentRTPstreams,oneforvideoandoneforaudio.Thetimestampsintheses RTPpacketsaretiedtothe videoandaudiosamplingclocks,andarenottiedtothe*wall-clocktime*(i.e.,torealtime).EachRTCPsender-reportcontains,forthemostrecently generatedpacketintheassociatedRTPstream,thetimestampoftheRTPpacketandthewall-clocktimeforwhenthepacketwascreated.Thusthe RTCPsender-reportpacketsassociatethesamplingclocktothereal-timeclock.ReceiverscanusethisassociationintheRTCPsenderreportsto synchronizetheplayoutofaudioandvideo.

Sourcedescriptionpackets

ForeachRTPstreamthatasenderistransmitting,thesenderalsocreatesandtransmits source-descriptionpackets.Thesepacketscontain informationaboutthesource,suchase-mailaddressofthesender,thesender'snameandtheapplicationthatgeneratestheRTPstream.Italso includestheSSRCoftheassociatedRTPstream.Thesepacketsprovideamappingbetweenthesourceidentifier(i.e.,theSSRC)andtheuser/host name.RTCPpacketsarestackable,i.e.,receiverreceptionreports,senderreports,andsourcedescriptorscanbeconcatenatedintoasinglepacket.The resultingpacketisthenencapsulatedintoaUDPsegmentandforwardedintothemulticasttree.

RTCP Bandwidth Scaling

TheastutereaderwillhaveobservedthatRTCPhasapotentialscalingproblem.Considerforexample anRTPsessionthatconsistsofonesender andalargenumberofreceivers.IfeachofthereceiversperiodicallygenerateRTCPpackets,thentheaggregatetransmissionrateofRTCPpackets cangreatlyexceedtherateofRTPpacketssentbythesender.Observethattheamountoftrafficsentintothemulticasttreedoesnotchangeasthe numberofreceiversincreases,whereastheamountofRTCPtrafficgrowslinearlywiththen

umberofreceivers.Tosolvethisscalingproblem, RTCPmodifiestherateatwhichaparticipantsendsRTCPpacketsintothemulticasttreeasaf unctionofthenumberofparticipantsinthesession. Observethat,becauseeachparticipantsendscontrolpacketstoeveryoneelse,eachparticip antcankeeptrackofthetotalnumberofparticipantsin thesession.

RTCPattemptstolimititstrafficto5%ofthesessionbandwidth.Forexample,supposeth ereisonesender,whichissendingvideoatarateof2

Mbps.ThenRTCPattemptstolimititstrafficto5%of2Mbps,or100Kbps,asfollows.The protocolgives75%ofthisrate,or75Kbps,tothe receivers;itgivestheremaining25%oftherate,or25Kbps,tothesender.The75Kbpsdevotedt othereceiversisequallysharedamongthe receivers.Thus,iftherearеRreceivers,theneachreceivergetstosendRTCPtrafficatarateof75 /RKbpsandthesendergetstosendRTCPtraffic at arateof25Kbps.Aparticipant(asenderorreceiver)determinestheRTCPpackettransmissio nperiodbydynamicallycalculatingthetheaverage RTCPpacketsize(acrosstheentiresession)anddividingtheaverageRTCPpacketsizebyitsal locatedrate.Insummary,theperiodfortransmitting RTCPpacketsforasenderis

$$T = \frac{\text{number of senders}}{.25*.05*\text{session bandwidth}} (\text{avg. RTCP packet size})$$

AndtheperiodfortransmittingRTCPpacketsforareceiveris

$$T = \frac{\text{number of receivers}}{.75*.05*\text{session bandwidth}} (\text{avg. RTCP packet size})$$

4. Explain in detail about the H.323.

H.323isastandardforreal-timeaudioandvideoconferencingamongendsystemsontheInternet.AsshowninFigure6. 4-7,italsocovershowend systemsattachedtotheInternetcommunicatewithtelephonesattachedtoordinarycircuit-switchedtelephonenetworks.Inprinciple,if manufacturers ofInternettelephonyandvideoconferencingallconformtoH.323,thenalltheirproductssho uldbeabletointeroperate,andshould be

abletocommunicatewithordinarytelephones.WediscussH.323inthissection,asitprovide sanapplicationcontextforRTP.Indeed,weshall seebelowthatRTPisanintegralpartoftheH.323standard.

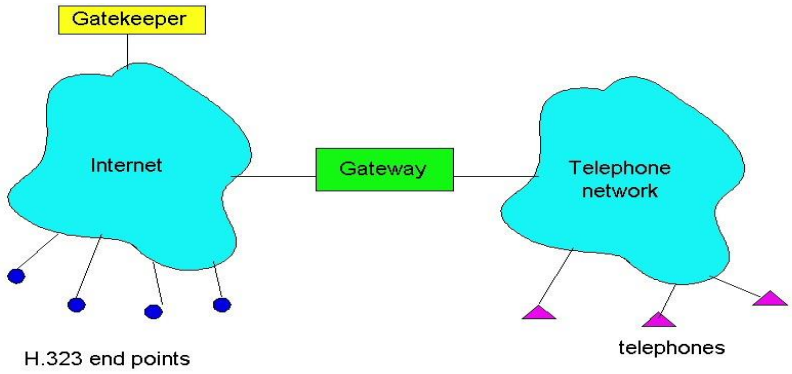


Figure. H.323endsystemsattachedtotheInternetcancommunicatewithtelephonesattachedtoa circuit-switchedtelephonenetwork.

H.323**endpoints**(a.k.a.terminals)canbestand-alonedevices(e.g.,WebphonesandWebTVs)orapplicationsinaPC(e.g.,Internetphoneor video conferencingsoftware).H.323equipmentalsoincludes**gateways**and**gatekeepers**.Gatew ayspermitcommunicationamongH.323endpoints andordinarytelephonesinacircuit-switchedtelephonenetwork.Gatekeepers,whichareoptional,provideaddresstranslation ,authorization, bandwidthmanagement,accountingandbilling.Wewilldiscussgatekeepersinmoredetail attheendofthissection.

TheH.323isanumbrellaspecificationthatincludes:

- Aspecificationforhowendpointsnegotiatecommonaudio/videoencodings.Becaus eH.323supportsavarietyofaudioandvideoencoding standards,aprotocolisneededtoallowthecommunicatingendpointstoagreeonaco mmonencoding.
- Aspecificationforhowaudioandvideochunksareencapsulatedandsentovernetwor k.Asyoumayhaveguessed,thisiswhereRTPcomes intothepicture.

- AspecificationforhowendpointscommunicatewiththeirrespectivegatekeepersAspecificationforhowendpointscommunicatewiththeirrespectivegatekeepers.

- AspecificationforhowInternetphonescommunicatethroughagatewaywithordinaryphonesinthepubliccircuit-switchedtelephone network.AspecificationforhowInternetphonescommunicatethroughagatewaywithordinaryphonesinthepubliccircuit-switchedtelephone network.

Minimally,eachH.323endpoint*must*supporttheG.711speechcompressionstandard.G.711usesPCMtogeneratedigitizedspeechateither56 kbpsor64kbps.AlthoughH.323requireseveryendpointtobevoicecapable(throughG.711),videocapabilitiesareoptional.Becausevideo supportisoptional,manufacturersofterminalscansellsimplerspeechterminalsaswellasmorecomplexterminalsthatsupportbothaudioand video.

AsshowninFigure6.4-8,H.323alsorequiresthatallH.323endpointsusethefollowingprotocols:

- RTP-thesendingsideofanendpointencapsulatesallmediachunkswithinRTPpackets.SendingsidethenpassestheRTPpacketstoUDP.

- H.245-an"out-of-band"controlprotocolforcontrollingmediabetweenH.323endpoints.Thisprotocolisusedtonegotiateacommon audioorvideocompressionstandardthatwillbeemployedbyalltheparticipatingendpointsinasession.

- Q.931-asignalingprotocolforestablishingandterminatingcalls.Thisprotocolprovidestraditionaltelephonefunctionality(e.g.,dialtones andringing)toH.323endpointsandequipment.

- RAS(Registration/Admission/Status) channelprotocol-aprotocolthatallowsendpointstocommunicatewithagatekeeper(ifgatekeeperispresent).

Figure6.4-8showstheH.323protocolarchitecture

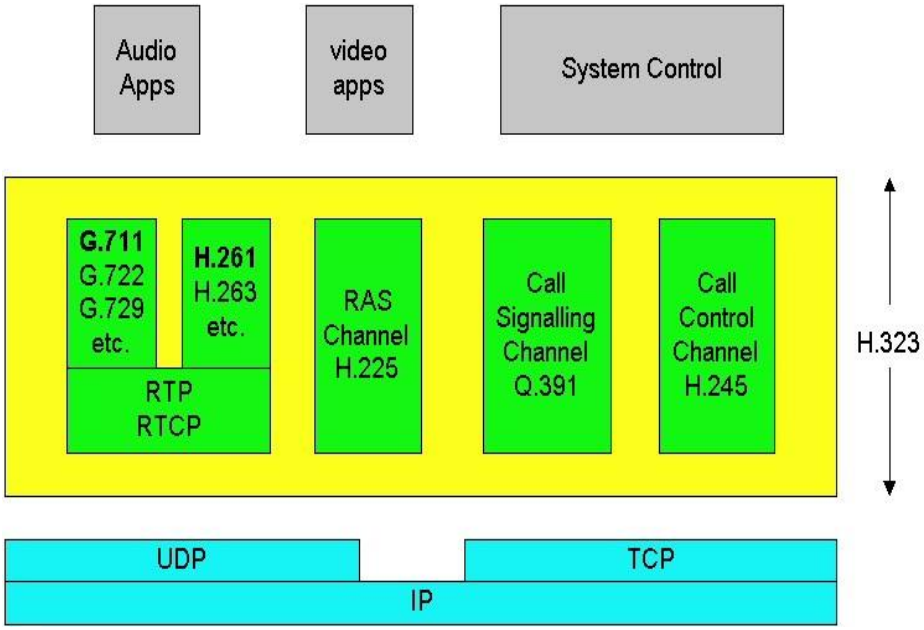


Figure6.4-8H.323protocolarchitecture.

Audio and Video Compression

TheH.323standardsupportsaspecificsetofaudioandvideocompressiontechniques.Let'sfirstconsideraudio.Aswejustmentioned,allH.323 endpointsmustsupporttheG.711speechencodingstandard.BecauseofthisrequirementtwoH.323endpointswillalwaysbeabletodefaultto G.711andcommunicate.ButH.323allowsterminalstosupportavarietyofotherspeechcompressionstandards,includingG.723.1,G.722,G.728 andG.729.Manyofthesestandardscompressspeechtoratesthatwillpassthrough28.8Kbpsdial-upmodems.Forexample,G.723.1compresses speechtoeither5.3kbpsor6.3kbps,withsoundqualitythatiscomparabletoG.711.

AswementionedearliervideocapabilitiesforanH.323endpointareoptional.However,ifanendpointdoessupportsvideo,thenitmust(atthe veryleast)supporttheQCIFH.261(176x144pixels)videostandard.Avideocapableendpointmyoptionallysupportotherh.261schemes, includingCIF,4CIFand16CIF.,andtheH.263standard.AstheH.323standardevolves,it willlikelysupportalongerlistofaudioandvideo compressionschemes.

H.323 Channels

WhenaendpointparticipatesinanH.323session,itmaintainsseveralchannels,asshowninFig ure

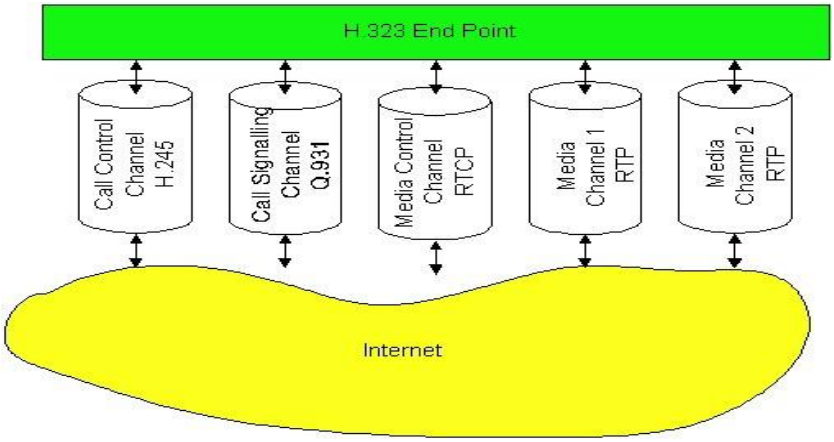
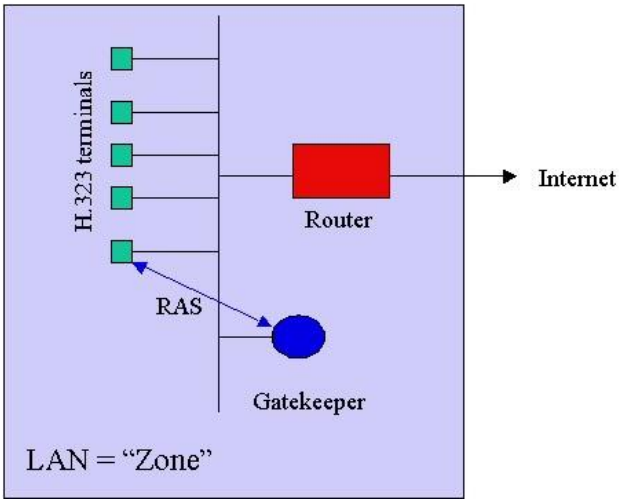


Figure6.4-9H.323channels

ExaminingFigure6.4-9,weseethatanendpointcansupportmanysimultaneousRTPmediachannels.Foreachmediatype,therewilltypicallybeonesendmediachannelandonereceivemediachannel;thus,ifaudioandvideoaresentinseparateRTPstreams,therewilltypicallybefourmediachannels.AccompanyingtheRTPmediachannels,thereisoneRTCPmediacontrolchannel,asdiscussedinSection6.4.3.AlloftheRTPandRTCPchannelsrunoverUDP.InadditiontotheRTP/RTCPchannels,twootherchannelsarerequired,thecallcontrolchannelandthecallsignalingchannel.TheH.245callcontrolchannelisaTCPconnectionthatcarriesH.245controlmessages.Itsprincipletasksare(i)openingandclosingmediachannels;and(ii)capabilityexchange,i.e.,beforesendingmedia,endpointsagreeonandencodingalgorithm.H.245,beingacontrolprotocolforreal-timeinteractiveapplications,isanalogoustoRTSP,whichisacontrolprotocolforstreamingofstoredmultimedia.Finally,theQ.931callsignalingchannelprovidesclassicaltelephonefunctionality,suchasdialtoneandringing.

Gatekeepers

ThegatekeeperisanoptionalH.323device.EachgatekeeperisresponsibleforanH.323zone.AtypicaldeploymentscenarioisshowninFigure6.4. Inthisdeploymentscenario,theH.323terminalsandthegatekeeperareallattachedtothesameLAN,andtheH.323zoneistheLANitself.Ifa zonehasagatekeeper,thenallH.323terminalsinthezonearerequiredtocommunicatewithitusingtheRASprotocol,whichrunsoverTCP. Addresstranslationisoneofthemoreimportantgatekeeperservices.Eachterminalcanhaveanaliasaddress,suchasthenameofthepersonatthe terminal,theemailaddressofthepersonattheterminal,etc.Thegatewaytranslatesthesealiasaddressestoaddresses.Thisaddresstranslation serviceissimilartotheDNSservice,coveredinSection2.5.Anothergatekeeperserviceisbandwidthmanagement:thegatekeepercanlimitthe numberofsimultaneousrealtimeconferencesinordertosavesomebandwidthforotherapplicationsrunningovertheLAN.Optionally,H.323calls canberoutedthroughgatekeeper,whichisusefulforbilling



H.323terminalmustregisteritselfwiththegatekeeperinitszone.WhentheH.323application isinvokedattheterminal,theterminalusesRASto senditsIPaddressandalias(providedbyuser)tothegatekeeper.Ifgatekeeperispresentinazone,eachterminalinthezonemustcontact gatekeepertoaskpermissiontomakeacall.Onceithaspermission,theterminalcansendthegatekeeperane-mailaddress,aliasstringorphone extensionfortheterminalitwantstocall,whichmaybeinanotherzone.Ifnecessary,agatekeeperwillpollothergatekeepersinotherzonesto resolveanIPaddress.