Chapter 2

IP based Multimedia Services

2.1 Introduction

Wireless systems and Internet are the two major forces driving the telecommunication Industry. The convergence of these increase the demand for new services, increasing bandwidth and ubiquitous connectivity continuously grows. The next-generation mobile systems will be based solely, (or in a large extent), on Internet Protocol(IP). The basic intent of IP is to provide seamless multimedia services to users who access an all-IP infrastructure through a variety of heterogeneous access technologies, meeting the demands of both enterprise and public environments anywhere and anytime. IP is assumed to act as an adhesive to provide global connectivity, mobility among networks, and a common platform for service provisioning across different types of access networks.

IP based Multimedia Services (IMS) relates to the integration of mobile communications and Internet technologies which will bring the features of Internet services to the mobile environment. These services provide inter-operability between fixed and mobile networks which will allow users to experience seamless converged services. To ease the integration with the Internet, IMS uses Internet Engineering Task Force(IETF) protocols wherever possible, e.g. Session Initiation Protocol (SIP)[RFC 3261]. According to the 3rd Generation Partnership Project(3GPP), IMS [RFC3113] is not intended to standardize applications but rather to aid the access of multimedia and voice applications from wireless and wireline terminals, i.e. create a form of fixed-mobile convergence(FMC).
Multimedia has already covered and is covering various categories of applications design and services and hence, forms a major part of Internet traffic on the web. This scenario will lead to increase in the number of data and users. However, the network resources will be limited. This lays a great potential of research work in the area of image and video compression, optimization of bandwidth utilization and quality of service management of multimedia data under transmission or storage. Examples of recent efforts to define new international compression standards, with higher coding efficiency than the state of the art solutions, are the JPEG XR image compression standard, approved and published as ITU-T recommendation and ISO/IEC standard in 2010, and the on-going activities of the Joint Collaborative Team on Video Coding (JCT-VC) of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) to define the next generation video coding standard, called High Efficiency Video Coding (HEVC).

2.1.1 Present scenario of Audio/Video standardization

Proprietary technologies such as Apple QuickTime, Microsoft Silverlight and Adobe Flash, which allow video visualizations on the Web, have become popular. Also the World Wide Web Consortium is currently developing a standard, HTML 5, that provides the enhanced functionality to embed non-proprietary video formats in a web page. This will allow users to view video streams, that is embedded in a web page without a specific video player, simplifying the access to video resources. Google recently proposed an open, royalty-free, image and video file format, called WebP for the images and WebM for the audio visual sequences. This format has been specifically designed for the Web and is quickly gaining popularity, being natively supported in an increasing number of Web browsers. The visual data contained in a WebP or WebM file is encoded using the VP8 open source video compression [Simone et al. 2012] algorithm.

An article related to performance analysis of VP8 image and video compression based on subjective evaluations, released by Moscow State University (MSU) for comparing different H.264/AVC implementations and VP8, using both compressed and uncompressed source video sequences. At present, the only study directly comparing both VP8 and HEVC with H.264/AVC is that conducted by Ohworiole and
Andreopoulos, which considers uncompressed source video sequences and PSNR and structural similarity index measure (SSIM) for performance evaluation. Considering the image compression, an extensive study of the performance of WebP in comparison to JPEG has been performed by Google, using SSIM and uncompressed source images, as well as PSNR and JPEG compressed source images.

The next section will cover the basics related to the format of Multimedia Data, which is vital for understanding the efforts related to the scheduling and routing of these traffic over communication networks.

2.2 Multimedia Standards

Multimedia stands for more than one media including text, audio, video, graphics, animation. It also refers to two or more continuous media being played during some well defined time interval with user interaction, most demanding among them is audio and video. Some of the popular multimedia format standards are governed by

- ITU Telecommunication Standardization Sector (ITU-T) for H.26x standard

- International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), popularly known as (ISO/IEC) for MPEG-x (Moving Picture Experts Group) standards.

Various factors that affects the quality of MPEG compressed video have been referred from [Jeremiah Golston (Distinguished Member Technical Staff 2004] and ISO/IEC 14496-2: Information Technology: Coding of audio-visual objects, 2001 for MPEG-4 Part 2, MPEG-4 Visual. Some of them have been summarized as under:

- The resolution of the original video source: If the original video source has poor resolution, nothing much can be done on the video output.

- The bit-rate(channel bandwidth) allowed after compression: For videos with higher bit rates (channel bandwidth), more information can be transmitted, allowing fewer motion artifacts to be present or a higher-resolution image to be displayed. e.g the video quality rapidly degrades, with the 8 x 8 blocks becomes clearly visible once the bit-rate drops below a given threshold.
• Effectiveness of Motion Estimator: This determines the reduction in video size, quality when movement starts or when the amount of movement is above a certain threshold. Poor motion estimation will contribute to a general degradation of video quality.

2.2.1 Video Codec Standards

Some of the prominent Video codecs popularly used for multimedia data is H.261, MPEG-1, H.263, H.264(AVC), MPEG-4 Part 2 (Visual), MPEG-4 Part 10. A brief insight into these codes are mentioned below:

• H.261: The major features of H.261 are:
  – Designed primarily for video phone and video conference over Integrated Services Digital Network(ISDN)
  – Uses Discrete Cosine Transform(DCT) based compression to reduce spatial redundancy (similar to Joint Photographic Experts Group (JPEG))
  – Includes Block based motion compensation to reduce temporal redundancy

• MPEG-1: It is designed for storage/retrieval of VHS quality video on CD-ROM. The major features of MPEG-1 are:
  – Support for applications based on storage and retrieval of moving pictures and audio on digital media such as video CDs using SIF resolution 352 x 240 at 30 fps
  – The targeted output bit-rate is around 1.15 Mbps, which produces effectively 25:1 compression
  – It has similar coding scheme to H.261 with Random access support for fast forward/backward support

In addition, level 3 of MPEG-1 is the most popular standard for digital compression of audio known as MP3.

• MPEG-2 was designed for broadcast quality video storage and transport. It is used primarily in DVB, DirecTV, Digital CATV. The major features of MPEG-2 are:
- Scales well to High Definition Television (HDTV) resolution
- Supports bit rate of 2Mbps or higher (Constant Bit Rate (CBR)/Variable Bit Rate (VBR))
- Supports two system bit streams: Packet Stream and Transport Stream

The most significant enhancement from MPEG-1 is its ability to efficiently compress interlaced video. It is designed for videos with bit-rate between 1.5-15.0 Mbps.

- **H.263**: Primarily designed for low bit-rate video applications. It is a legacy codec that is used by existing H.323 systems and has been kept for compatibility. It was further enhanced to codec’s such as H.263v2 (a.k.a. H.263+ or H.263 1998) and H.263v3 (a.k.a. H.263++ or H.263 2000). The codec is very similar to H.261, but more efficient. Its features are:
  - Used for Video over ordinary telephone modems that ran at 28.8 Kbps
  - Target resolution is from Sub Quarter Common Intermediate Format (SQ-CIF) (128 x 96) to 16CIF (1408 x 1152)

- **H.264/AVC**: This is one of the most advanced video compression standard developed by Joint Video Team (JVT) in 2003 and formally standardized as ITU-T H.264 and ISO/IEC MPEG-4 Part 10. As compared to MPEG-2 and MPEG-4 Part 2, a bit rate gain of more than 50% is achieved through a set of advanced coding tools such as multi-frame variable block size motion compensated prediction, advanced context-based entropy coding, advanced temporal prediction structures and adaptive in-loop deblocking filter. It is currently one of the most commonly used video codecs for recording, compression and distribution of high definition video in a large variety of applications including TV broadcast, video conferencing, web video, Tele-medicine, Satellite telecast, and Blu-Ray.

  H.264 defines a set of three profiles each supporting a particular set of coding functions and each specifying what is required of an encoder or decoder that complies with the corresponding profile. The profiles for H.264 is displayed in Figure 2.1 [Atul, Chen, and Luthra 2004]. A brief overview into the details is mentioned below:
The Baseline Profile supports intra and inter-coding (using I-slices and P-slices) and entropy coding with context-adaptive variable-length codes (CAVLC). e.g. Video-telephony, video-conferencing and wireless communications.

The Main Profile includes support for interlaced video, inter-coding using B-slices, inter coding using weighted prediction and entropy coding using Context Based Arithmetic Coding (CABAC). e.g. Television broadcasting and video storage.

The Extended Profile does not support interlaced video or CABAC but adds modes to enable efficient switching between coded bitstreams (SP and SI-slices) and improved error resilience (Data Partitioning). e.g. Streaming media applications.

- MPEG-4 [TellaSonera 2004] absorbs many of the features of MPEG-1 and MPEG-2 and other related standards, adding new features such as (extended) Virtual Reality Markup Language (VRML) support for 3D-rendering, object-oriented composite files (including audio, video and VRML objects), support for externally-specified Digital Rights Management and various types of interactivity. Advanced Audio Codec (AAC) was standardized as an adjunct to MPEG-2 (as Part 7) before MPEG-4 was issued. MPEG-4 Standard is predominantly used for multimedia and Web compression. MPEG-4 involves object-based compression, similar in nature to the Virtual Reality Modeling Language. Individual objects within a scene are tracked separately and compressed together to create an MPEG-4 file. This leads to a efficient compression that scalable, from low to high bit rates. It allows developers to access objects in a scene independently, and therefore, introduce interactivity. Most of the features included
in MPEG-4 are left to individual developers to decide whether to implement them, which is why it is divided into many parts ranging from Part 1 - Part 22. The major features of MPEG-4 can be summarized as under:

- Increased error robustness in supporting wireless networks
- Better support for low bit-rate applications
- Support both low bandwidth connections (wireless/mobile) and high bit rates (fixed/wire line)
- Support for a variety of new tools to support merging graphic objects with video and on the fly composition
- Zoom-in and zoom-out (remote monitoring) features
- Fast-forward and fast-backward (video on demand)
- Change viewing point used in applications like online shopping, sports

- WebM: It is an audio-visual format recently developed and sponsored by Google. A WebM file consists of a VP8 coded video stream and a Vorbis coded audio stream multiplexed into a Matroska container. The VP8 video codec [Simone et al. 2012] includes similar coding tools than H.264/AVC and some alternative tools such as adaptive mixing strategies for artificial reference frames, processor adaptive real time encoding and a low complexity loop filter.

- HEVC: Joint Collaborative Team on Video Coding (JCT-VC) has started to develop a new video coding standard known as high efficiency video coding (HEVC), due to the increasing demand for more efficient and flexible video coding solutions beyond H.264/AVC. The first evaluations have already shown that bit rate gains up to 50% can be achieved with respect to H.264/AVC. This gain is achieved due to extended or new coding tools, such as larger block sizes with flexible subpartitioning, intra picture prediction from adjacent prediction units, motion vector competition and hierarchical variable length coding.

### 2.2.2 Multimedia codec comparisons

A comparison of different compression techniques as referred from [Kwon, Tamhankar, and Rao 2005] and [Atul, Chen, and Luthra 2004] is mentioned in Table 2.1 and 2.2.
Table 2.1: Comparison of Multimedia Codecs (Part-I)

<table>
<thead>
<tr>
<th>Feature / Standard</th>
<th>MPEG-1</th>
<th>MPEG-2</th>
<th>MPEG-4 part 2 (visual)</th>
<th>H.264/MPEG-4 part 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macroblock size</td>
<td>16x16</td>
<td>16x16 (frame mode) 16x8 (field mode)</td>
<td>16x16</td>
<td>16x16</td>
</tr>
<tr>
<td>Block Size</td>
<td>8x8</td>
<td>8x8</td>
<td>16x16, 16x8, 8x8</td>
<td>16x16, 16x8, 8x16, 8x8, 4x8, 8x4, 4x4</td>
</tr>
<tr>
<td>Transform</td>
<td>8x8 DCT</td>
<td>8x8 DCT</td>
<td>8x8 DCT/Wavelet</td>
<td>4x4, 8x8 Int DCT, 4x4,2x2 Hadamard</td>
</tr>
<tr>
<td>Entropy coding</td>
<td>VLC</td>
<td>VLC</td>
<td>VLC</td>
<td>VLC, CAVLC, CABAC</td>
</tr>
<tr>
<td>Quantization</td>
<td>Scalar Quantization with step size of constant increment</td>
<td>Scalar Quantization with step size of constant increment</td>
<td>Vector Quantization</td>
<td>Scalar Quantization with step size increase at the rate of 12.5%</td>
</tr>
<tr>
<td>Motion Estimation &amp; Compensation</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>Yes, more flexible Up to 16 MVs per MB</td>
</tr>
<tr>
<td>Reference Picture</td>
<td>one</td>
<td>one</td>
<td>one</td>
<td>multiple</td>
</tr>
<tr>
<td>Picture Types</td>
<td>I, B, P, D</td>
<td>I, B, P</td>
<td>I, B, P</td>
<td>I, B, P, SP, SI</td>
</tr>
</tbody>
</table>
### Table 2.2: Comparison of Multimedia Codecs (Part-II)

<table>
<thead>
<tr>
<th>Feature / Standard</th>
<th>MPEG-1</th>
<th>MPEG-2</th>
<th>MPEG-4 part 2 (visual)</th>
<th>H.264/MPEG-4 part 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Robustness</td>
<td>Sync &amp; concealment</td>
<td>Data partitioning, FEC for important packet transmission</td>
<td>Sync, Data partitioning, Header extension, Reversible VLCs</td>
<td>Data partitioning, Parameter setting, Flexible macroblock ordering, Redundant slice, Switched slice</td>
</tr>
<tr>
<td>Transmission Rate</td>
<td>Up to 1.5 Mbps</td>
<td>2-15 Mbps</td>
<td>64 Kbps-2 Mbps</td>
<td>64Kbps-150Mbps</td>
</tr>
<tr>
<td>Compatibility with previous standards</td>
<td>N/A</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Encoder complexity</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>MEDIUM</td>
<td>HIGH</td>
</tr>
</tbody>
</table>

### 2.3 Multimedia Artifacts

The basic categorization of various streams of MPEG [Pinnacle 2000] includes Elementary Stream, Packet Elementary Stream, Program Stream and Transport Stream. The Figure 2.2 displays the relationship between these categories. The details related to this is are as under:

- Elementary Stream (ES) - the output of a single MPEG audio or video coder is called an elementary stream. They may be video, audio or data and there may be several elementary streams for each type of media (such as multiple audio channels for surround sound).

- Packet Elementary Stream (PES) - is a component media stream (such as audio or video) that has been converted to a sequence of packets. This packetization process involves segmenting a group of bits in an elementary stream and adding packet header information to the data. This packet header includes a packet identification code (PID) that uniquely identifies the packetized element-
CHAPTER 2. IP BASED MULTIMEDIA SERVICES

Figure 2.2: The Categorization of various Streams in MPEG

- Program Stream (PS) - is a combination of PES streams (such as video and audio) that compose a media program (such as a television program). A program stream is called a single program transport stream - SPTS. All the packets in a program stream share the same time reference system time clock (STC).

- Transport Stream (TS) - is a communications protocol for audio, video, and data, multiplexing of multiple PES that compose program channels (typically digital video channels) onto a signal communication channel. It is a type of digital container format that encapsulates packetized elementary streams and other data. MPEG-TS is also called a multi-program transport stream (MPTS).

This Figure 2.2 shows the pictorial representation of the components of MPEG. It shows that the MPEG system allows multiple media types to be used (voice, audio and data), codes and compresses each media type, adds timing information and combines (multiplexes) the media channels into an MPEG program stream. It also shows that multiple program streams (e.g. television programs) can be combined into a transport channel. When the MPEG signal is received, the program channels are separated (demultiplexed), individual media channels are decoded and decompressed and they are converted back into their original media form. Hence, one video PES and a number of audio PES can be combined to form a Program Stream (PS), provided that all of the coders are locked to a common clock. Time stamps in each PES can be used to ensure lip-sync between the video and audio.
CHAPTER 2. IP BASED MULTIMEDIA SERVICES

It is worth noting that, the packets in the same elementary stream all have the same PID, so that the decoder (or a demultiplexer) can select the elementary stream(s) it wants and reject the remainder. Packet continuity counts ensure that every packet that is needed to decode a stream is received. An effective synchronization system is needed so that decoders can correctly identify the beginning of each packet and deserialize the bit stream into words.

At the decoder the basic functionality is to the demultiplex various streams contained within a Transport Stream into audio and video streams which needs to be synchronized as shown in Figure 2.3.

![MPEG Decoder Block diagram](image)

Figure 2.3: MPEG Decoder Block diagram

### 2.3.1 MPEG Frame Types

MPEG uses three fundamental techniques to achieve compression:

- **Sub-sampling**: This process is used to reduce color information that is less sensitive to the eye. i.e. it converts each pixel in RGB format to YUV format.

- **Spatial compression or intra coding**: This is related to removal of redundant information within frames using the property that pixels within a single frame are related to their neighbors.

- **Temporal compression or inter-frame coding**: This process refers to removal of redundant information between the frames.
The details about the intra-coding and inter-coding is mentioned in the sections below. Each MPEG frame contains block, macro-block and slice information as shown in the Figure 2.4 referred from [Greengrass, Evans, and Begen 2009].

![Figure 2.4: Blocks, macroblocks, and slices](image)

A brief overview into the basic building blocks of a MPEG frame is presented below:

- **Block**: It is a $8 \times 8$ matrix of pixels or corresponding discrete cosine transform information that represents a small chunk of brightness(luma) or color(chroma) within the frame.

- **Macroblock**: It contains several blocks that define a section of the frame’s brightness component and spatially corresponding color components. Typically, pictures (frames) are segmented into macroblocks, and individual prediction types can be selected on a macroblock basis rather than being the same for the entire picture, as mentioned below:
  - I-frames contain only intra macroblocks
  - P-frames contain either intra macroblocks or predicted macroblocks
  - B-frames contain intra, predicted, or bi-predicted macroblocks

- **Slice**: In H.264/MPEG-4 AVC, the granularity of the establishment of prediction types is brought down to a lower level called slice. Every slice contains at least one macroblock, although the number of macroblocks within a slice
can vary, and their position might change from picture to picture. A slice is a spatially distinct region of a frame that is encoded separately from any other region in the same frame. In H.264/MPEG-4 AVC, instead of I-frames, P-frames, and B-frames, there are I-slices, P-slices, and B-slices. Also, instead of using I, B and P-frame type selections, the encoder can choose the prediction style distinctly on each individual slice. Some of the additional types of frames/slices used are:

- SI-frames/slices (Switching I): Facilitates switching between coded streams; contains SI-macroblocks (a special type of intra coded macroblock).
- SP-frames/slices (Switching P): Facilitates switching between coded streams; contains P and/or I-macroblocks
- Multi-frame motion estimation (up to 16 reference frames, or 32 reference fields)

Multi-frame motion estimation will allow increase in the quality of the video while allowing the same compression ratio. SI and SP-frames (defined for Extended Profile) will allow for increase in the error resistance. When such frames are used along with a smart decoder, it is possible to recover the broadcast streams of the damaged DVDs.

As referred from [Greengrass, Evans, and Begen 2009], a digital snapshot or frame of a black and white 640 × 480 pixel standard-definition (SD) television image that uses 8 bits to represent each pixel’s grayscale consumes 2.45 Mbits of memory. If we updated this image with a frame rate of 30 frames per second (fps), the resulting video bandwidth requirement would be around 70 Mbps. This requirement will grow, if we add more bits to encode the image and add color and audio channels.

The three basic frame types of MPEG-4/AVC is I, P and B frames/slices. The details for the same is mentioned as follows:

- Intra or I-frames/slices: The I-frames carry fully specified picture, like a conventional static image file. They’re coded without reference to other frames and might use spatial compression but don’t use temporal compression. The I-frames may be generated by an encoder to create a random access point (to
allow a decoder to start decoding properly from scratch at that picture location). Spatial compression uses the property that pixels within a single frame are related to their neighbors; by removing spatial redundancy. The size of the encoded frame can be reduced and prediction can be used at the decoder to reconstruct the frame. These frames are the least compressible and also do not require any frame to decode their information. Hence, they require more bits to encode than other frame types. The intra refresh periods of a half-second are common on such applications as digital television broadcast and DVD storage. Longer refresh periods may be used in some environments. For example, in video-conferencing systems it is common to send I-frames very infrequently.

- Predictive-coded or P-frames/slices: The P-frames/slices holds only the changes in the image from the previous frame. They predict the frame to be coded from a preceeding I-frame or P-frame using temporal compression. Hence, they require the prior decoding of some other picture(s) in order to be decoded. The P-frames may contain both image data and motion vector displacements and combinations of the two. They can reference previous pictures in decoding order. MPEG-2 uses only one previously decoded picture as a reference during decoding, and require that picture to also precede the P-picture in display order. H.264 can use multiple previously decoded pictures as references during decoding, and can have any arbitrary display-order relationship relative to the picture(s) used for its prediction. The P-frame require fewer bits for encoding than I-pictures.

- Bi-directional predicted frames/slices (B-frames/slices): These frames saves even more space by using differences between the current frame and both the preceding and following frames to specify its content. They may contain both image data and motion vector displacements or combinations of the two. They use the prior decoding of the I-frame or P-frame as their reference points for motion compensation. B-frames can use both previous and forward frames for data reference to get the highest amount of data compression. For theoretical analysis, the B-frames are assumed to be 1/4th of the size of I-frames. These frames include some prediction modes that form a prediction of a motion region
(e.g. a macroblock or a smaller area) by averaging the predictions obtained using two different previously decoded reference regions. The B-frames are never used as references for the prediction of other pictures. As a result, a lower quality encoding (resulting in the use of fewer bits than would otherwise be the case) can be used for such B-frames because the loss of detail will not harm the prediction quality for subsequent pictures.

However, in H.264, they may or may not be used as references for the decoding of other pictures (at the discretion of the encoder). In older standard designs (such as MPEG-2) two previously decoded pictures are used as references during decoding, and require one of those pictures to precede the B-frame in display order and the other one to follow it. In H.264, we can use one, two, or more than two previously decoded pictures as references during decoding, and can have any arbitrary display-order relationship relative to the picture(s) used for its prediction. The B-frames/slices typically require fewer bits for encoding than either I or P-frames.

As per the MPEG Video Coding format the 3-bit coding format for Picture Coding Type in the Picture/Frames contained within a GoP has different values for I, P and B frames. This is displayed in Table 2.3:

Table 2.3: Picture Coding Type Values for I, P and B Frames

<table>
<thead>
<tr>
<th>Frame Type</th>
<th>Picture Coding Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-Frame</td>
<td>001</td>
</tr>
<tr>
<td>P-Frame</td>
<td>010</td>
</tr>
<tr>
<td>B-Frame</td>
<td>011</td>
</tr>
</tbody>
</table>

### 2.3.2 Group of Pictures (GoP)

In MPEG encoding, frames are arranged into groups of pictures (GoPs). A GoP includes the I-frame and all subsequent frames leading up to the next I-frame. GoPs typically have 12 or 15 frames, which support National Television System Committee (NSTC) at 30fps(interlaced) and Phase Alternating Line (PAL) standards at 25 fps. Many possible GoP structures exist, and the number of I, P, and B-frames within
a GoP is determined by the source video signal’s format, any bandwidth constraints on the encoded video stream (which determine the required compression ratio), and possible constraints on the encoding or decoding delay. A typical 15-frame GoP structure has 1 I-frame, 4 P-frames, and 10 B-frames. We can describe a regular GoP structure by its total number of frames (that is, the GoP size) as well as by how many B-frames occur between its P-frames. A typical 15:2 GoP structure is shown in the Figure 2.5.

![Figure 2.5: A typical 15:2 Group of Pictures](image)

The typical part of this structure is that the decoding of 2\textsuperscript{nd} and 3\textsuperscript{rd} (B-frame) frame requires the 4\textsuperscript{th} (P-frame) frame. Hence, the fourth frame would be sent prior to the 2\textsuperscript{nd} and 3\textsuperscript{rd} frame. This may delay the transmission, since, it will be necessary to keep the P-frame. The basic advantage of this structure is that it minimizes the problem of possible uncovered areas. Also, the P-frames and B-frames need less data as compared to I-frames, so less data is transmitted. However, the structure has certain weaknesses, like; it increases the complexity of the decoder, which can mean more memory is needed to rearrange the frames. Also, the interpolated frames (namely B-frames) require more motion vectors which means an increased bit rate.

In this research work we have used four different types of benchmark videos for study and analysis. This includes the following:

- Claire-news reader: that displays only facial movement and very slow motion speed
- Silent-Deaf news reader: that displays facial movement, hand movement and medium motion speed
- Coastguard: moving boat and steamer in opposite direction with high motion speed
- Foreman: site man explaining site

The Figure 2.6 covers, the distribution of I, P and B frames in the various types of video types. In Claire-news reader video, we observe that the number of I-frames is
very less as compared to the number of P and B frames. Similar observations were made for other types of videos like Silent-Deaf news reader, Coastguard and Foreman. This study is vital to understand the wide variation in the size and distribution of various types of videos. This makes the QoS based provisioning of multimedia a very challenging domain to work on. Multimedia data over communication network is mainly as Transport Stream. Transport streams differ from the similarly named program streams, since, program streams are designed for reasonably reliable media, such as discs (like DVDs), while transport streams designed for less reliable transmission. Also, a transport stream may carry multiple programs.

2.4 MPEG 4.10/H.264 over MPEG-2 TS

Instead of MPEG-2 video, MPEG-2 transport and program streams can carry MPEG-4.10 (H.264) video in PES packets. This enables existing infrastructures and equipment to accommodate the H.264 video codec easily. The PES packet stream_id = "1110 ××××" for MPEG-4.10 (H.264) video. Stream_type = 0×1B within the PMT or PSM. Carriage of the H.264 stream must need to be signaled by using the MPEG-2 AVC Video Descriptor and/or the MPEG-2 AVC Timing and HRD Descriptor. The basic structure of MPEG-TS is shown in Figure 2.7 as referred from [Jack 2007].

The details related to various fields in the MPEG TS is covered in Chapter
Figure 2.7: Block Diagram of MPEG TS

on MPEG-2 in [Jack 2007]. As mentioned previously, a TS combines one or more Programs (with its own Time Base for temporal complexity) into a single stream. The transport stream consists of one or more 188-byte packets. The data for each packet is from PES packets, PSI (Program Specific Information) sections, stuffing bytes, or private data. The PID is used to identify which Packets are part of the same program.

It is worth noting here that a value of ”1” in the Transport Priority field indicates that this packet is of higher priority than other packets having the same PID. A typical IP packet for transporting MPEG video contains seven 188 byte MPEG-TS packets.

2.5 Summary

MPEG-4 Visual and H.264 are popular standards for the coded representation of visual information that supports low bit rate applications. The compression ratio in MPEG-4 is much higher than that in MPEG-2. Hence, they are more preferred over bandwidth constrained applications. However, a higher level of compression also affects network performance adversely, since, it requires a lot of decoding overhead. Also, packet losses become more advertinent, because loss of fewer frames can lead to a higher loss in perceptual quality in the decoded video. Hence, it is important to strike a balance between multimedia compression and broadcast requirements, for various multimedia application types. The next Chapter focuses on the issues related to QoS provisioning of multimedia traffic over various layers of communication networks.