CHAPTER 1

Transmission of High Quality Audio and Video over IP Network

Transmission of high quality multimedia over the internet has always been a challenging task. For efficient transmission, video and audio data have to be compressed. It is often a standard procedure to use compression algorithms to compress audio and video and then store the data in storage devices for transmission. The application layer QoS module which is very important in multimedia transmission is used to align the audio/video stream bits in accordance with the status of the network. In order to send the compressed data streams to the internet the transport protocols first need to put the above mentioned compressed data streams into packets. Incidentally, it is worth noting that the packets that have reached the end receiver successfully will pass through the transport layers and will then be processed by the application layer before being decoded by the decoder. However, we shouldn't be oblivious of the fact that even streaming and downloading tools have their pros and cons. If you want to deliver live event and the connection is fast that is (broadband) then streaming will be the best technology to use. On the other hand, to convey stored movies, the best method to use is progressive download.

1.1 Media Capture and Compression

1.1.1. Audio capture and compression

Storing it into an audio input buffer and considering the specifics of audio capture, Fig 1.1 be speaks the sampling procedure on a general aim workstation, as it involves the process of sound capture, digitization. It is only after a bushel number of samples have been gathered that this input buffer is appended to the application, usually. Obstructing the process until enough samples have been gathered in order to create a complete frame, for the most component part audio capture application programming interfaces are come back data from the input buffer in bushel duration frames. Not provided until the last sample has been gathered and giving this procedure it should not be surprising that some delays will be

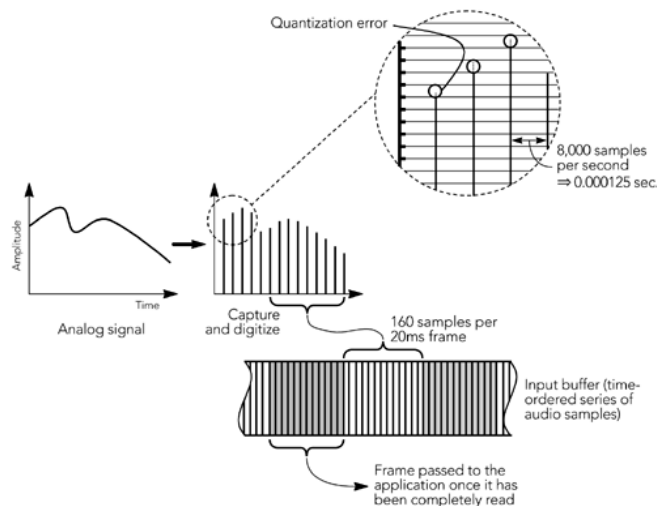encountered since the first sample in a frame is held.

Fig 1.1 Audio capture, digitization and framing [14]

With an ambit of certain sample types and at one of several sampling rates, it is possible to come back the decompressed audio frames from the capture device. With the application of audio capture devices that are the same it is possible to come back samples with 8 bit, 16 bit, or, 24 bit resolution, using linear, μ-law or A-law quantization, at rates between 8,000 and 96,000 samples per second, and in mono or stereo [14]. The audio data samples are rounded to a finite value, this method is named quantization, after the sampling. Each quantization value is signified by a bushel number of bit and the number of finite values is distinctively a power of two.

Lastly, the quantized audio bit streams are sampled at regular intervals by a process known as beat code modulation. Speech ciphering applies pulse code modulation, with sampling charge per unit of 8000 samples per second and 8-bits per sample, giving a rate of 64kbps. A Compact disc, mini-disc, 3/2 stereo all uses code pulse modulation. Due to large storage space occupied by both stereo and mono music file i.e. 1.411Mbps and their size exceeds the access rate of dial-up modem; PCM-encoded speech is obsolete in the internet. Also known as music player 3, a recognized compression procedure for CD quality stereo system music is mpeg 1 layer 3. Encoders of the music player 3 compress audio file to rates of 96kbps, 128kbps and produces miniature sound degradation. Mp3 is well accepted in multimedia industry because of the nature of format is header less which countenances mp3 music files to be streamed via the Net. Bit reservoir buffering [32] and the mp3 compression standard is nonetheless a complex procedure as it involves psycho acoustic masking, redundancy reduction.

1.1.2 Video Capture and Compression

It's not unusual for video capture devices to operate on complete frames of video instead of reverting human scan Lines, also frames may have an ambit of sizes, and capture devices may come back frames in an assortment of formats, color spaces, depths, and sub sampling. Most contemporary world remainder on the codex used, it could be necessary to convert from the device format before the frame can be used. The most commonly enacted changeover is probably between RGB and YUV color spaces, color dithering and sub-sampling are needed.
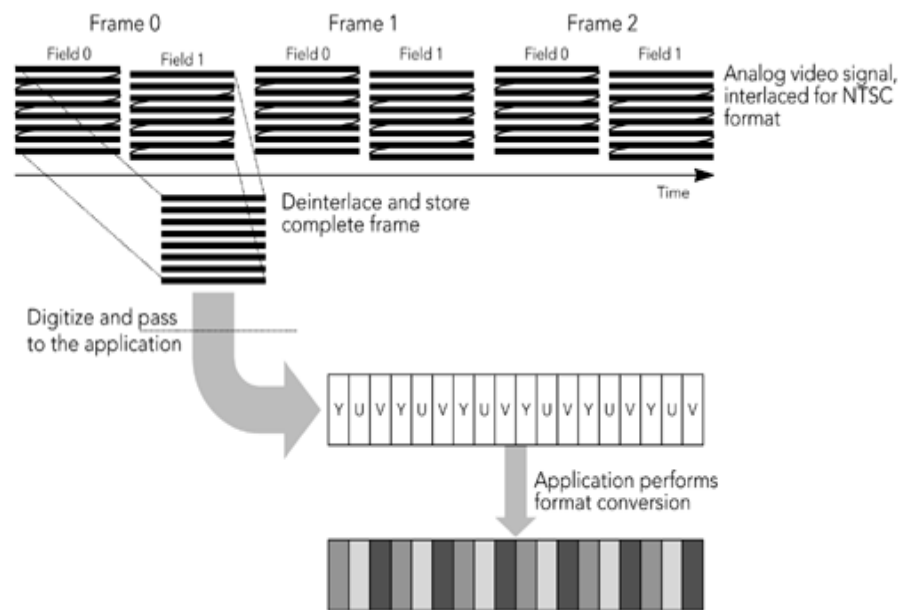

Fig 1.2   Video capture and framing [14]

Once captured, for compression video frames are first buffered before being go a crossed to the encoder. The quantity of buffering necessitated usually depends on the compression strategy that will be used at the compression stage. Most video codecs performs entomb frame compression, in which each frame relies on the encircling frames [14].

There are two types of redundancy in video, and it is possible to put up for compression. Spatial redundancy is the redundancy within a given mental image that eliminates unwanted frames and can be easily compressed. In temporal redundancy, the human visual system is not equally sensitive to all patterns. Therefore, the compression algorithm can discard information that is not visible to the observer. This is referred to as lossy compression [11]. Mpeg compression algorithms are the among the most well-known compression procedures. These includes mpeg 1 for CD read only storage quality (1.5mbps), mpeg 2 for high quality DVD video (3 6mbps) and H.264 video compression are all deployed in net today [32].Compression in MPEG is compress data in five steps, these are:

3

.

1.1.2.1 Resolution Reduction

Since the JPEG standard prescribes no color model, the image can be compressed in principle without prior conversion. It has been found that certain color models are better than others. For example, the human eye is able to detect small differences in brightness better than color changes. For this reason, we used mostly so-called brightness-color models, such as the YUV model or the YCbCr model. During the RGB color model, the brightness and color of a pixel is determined by its share of red, green and blue, is the YUV model saved the brightness Y (luminance), the color and the U F V (chrominance). The YUV can be easily calculated using weighted sums of the RGB values:
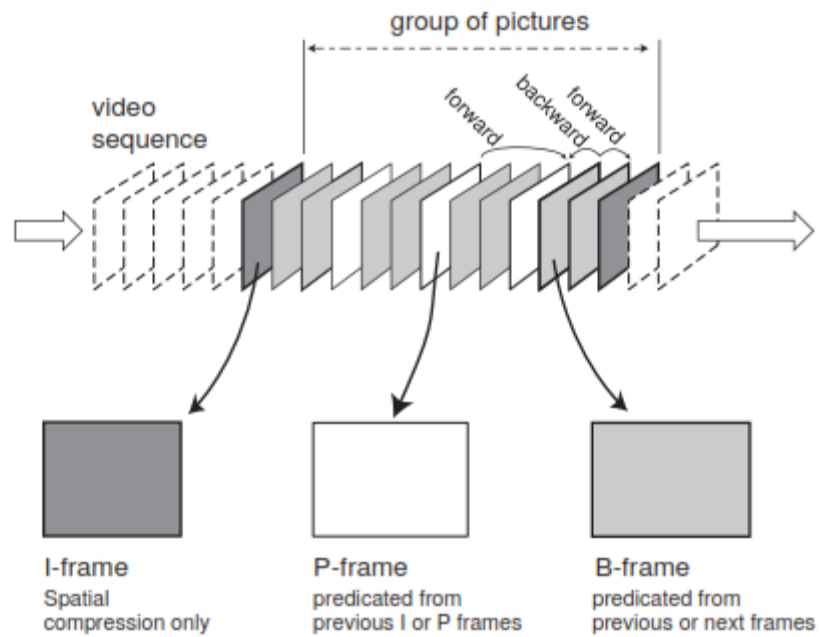
$$Y = 0.299R + 0.587G + 0.114B$$
$$V = R - Y$$ …………………………………….(1.1)
$$U = B - Y$$

One can see that the colors are not equally in the formula, which is because the human eye is very sensitive to the example of green. For this reason, the factor of the green component in the calculation of Y is also the highest. With the same Y also operates the YCrCb color model. In the 2 components, Cb and Cr, however, in each case the deviation from gray to blue or red value is stored.

1.1.2.2 Motion Estimation

It is the mpeg standard that offers a fashion of limiting this temporal redundancy and an mpeg video can be understood as a stream of frames because two successive frames of a video sequence often have small disparity. It entails the utilization of three types of frames I frames (intra), P frames (predicted) and B-frames (bidirectional) [21]. The I -frame designates an "intra-frame" that has been compressed by itself, without reference to other frames. It is also encoded using DCT with Huffman encoding of the run lengths of zeroes followed by the values of the first nonzero DCT coefficients. These coefficients are listed in zig-zag form

F (1, 0), F (0, 1), F (2, 0), F (1, 1) to determine the run lengths of zeros and the first non-

zeroes.

Fig: 1.3 I, P and B frames [20]

The P designates a frame that has been predictive-coded, using motion compensation. The B designates a frame that has been bi-directionally predictive coded. That, is the prediction is based on past and future I and P frames.

Motion estimation is elusive. It examines how each block in the image has moved from one frame to the next instead of transmitting the next block. MPEG encodes the motion vectors differently, because such vectors tend to be similar across a picture. Estimating the motion is the time consuming part of the MPEG algorithm.

1.1.2.3 Discrete Cosine Transforms (DCT)

"The Fast Fourier Transform permits, Indistinguishable to the Discrete Cosine Transform A depiction of an image data in footing of frequencies elements. So it is feasible that the frame blocks (8x8 or 16x16 pixels) can be depicted as frequency elements. Note that the changeover into the frequencies arena is described by the following formula [21]".

$DCT(i,j)$

$$= \frac{1}{\sqrt{2N}} C(i)C(j) \sum_{X=0}^{N-1} \sum_{Y=0}^{N-1} Pixel(x,y) cos\left[\frac{(2x+1)i\pi}{2N}\right] cos\left[\frac{(2x+1)J\pi}{2N}\right] \dots\dots\dots\dots (1.2a)$$

$$Pixel(x, y)$$

$$= \frac{1}{\sqrt{2N}} \sum_{X=0}^{N-1} \sum_{Y=0}^{N-1} C(i)C(j)DCT(i,j)cos\left[\frac{(2x+1)i\pi}{2N}\right]cos\left[\frac{(2x+1)J\pi}{2N}\right] \dots\dots\dots\dots (1.2b)$$

As a consequence of spatial redundancy, many of the coefficients consequence in aught or near aught values. It is possible to do away with these from the sequence of coefficients so the video frame will be evinced in as few bit as possible. The consequence is lossy compression that acquires dis embarrass of some detail, but the degree of detail rendered redundant or Useful is so fine that it is undetectable by the human optic. Sometimes, the word length of the remainders of the coefficients must be indicated in even fewer bit however, even greater compression is necessitated, so. Al though trim downing additional bit compromises the precision of the digitized video stream and appends to the lowering of perceived quality at the end user [31].

"Some of the other important observations about the images used in DCT are that across a movie tend to have similar features when changed by DCT. Another study shows that DCT coefficients tend to be zero. A good programming consists in counting the zeroes instead of listing them. This method called run length encoding is merged with Huffman encoding. Huffman encoding is useful to the stream of zeroes followed by the value of the first nonzero discrete cosine transform value [32]".

1.1.2.4 **Quantization**

"The technique of allocating concealed numeric values to the theoretically infinite possible values of each sample is named quantization. After the motion estimation, the numerical accuracy of the transformed coefficients is compact in order to reduce the quantity of bits in the stream. The degree of quantization applied to each coefficient is often determined by the visibility of the resulting distortion to a human observer; an example is that high-frequency coefficient can be more roughly quantized than low-frequency coefficient. Quantization is the stage that is responsible for the 'lossy' part of compression [51]". This is achieved in that the range of values decreased. Many coefficients are also zero.

F^G(u, v) = Integer Round$\left[\frac{F(u,v)}{Q(u,v)}\right]$ u, v………………………………….. (1.3)

The type of the quantization is not prescribed or set at this point, and it is choice of quantization table that determines the quality of the image. Also the quantization table used will be stored and sent to the decoder. It should be also known that it is possible to quantize a discrete cosine transform values by carve uping it with a non aught positive integer named quantization value and then proceeding to the discrete cosine transform values to the closest whole number. If the quantization value is big then the gibing precision of the quantized discrete cosine transform values becomes smaller. It is possible to transmit bring down precision values to a decoder with lesser bits. Often the utilization of large quantization values for high spatial frequencies countenances the encoder to selectively cast aside high spatial frequencies activity that the human oculus cannot readily perceive [11].

1.1.2.5 **Entropy coding**

"Entropy coding re scales coefficients based on the number of contemporary world they happen. Frequently ingeminated coefficients are evinced in the fewest number bit, hence hugely decreasing the total bandwidth necessitated to transmit the coefficients [2]. The entropy coding entails two staircase run length ciphering (RLE) and Huffman coding. These are popular loss less compression techniques, which could compress data, depending on its flexibility, by an appended factor of 3 to 4 [21]".
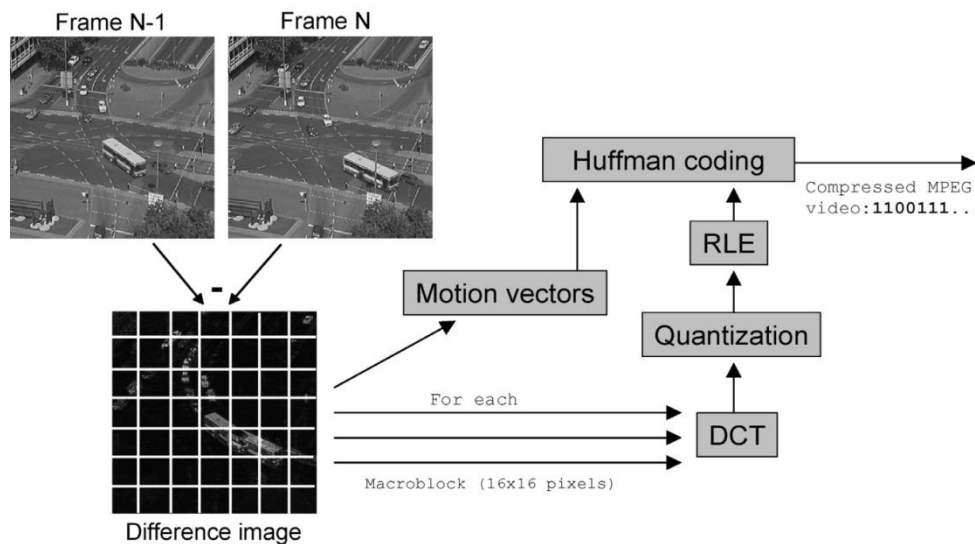

Fig 1.4 Diagram of five steps for MPEG compression [21]

The Huffman coding is the process of "using short codes for frequent symbols and longer ones for rare symbols". For Reconstruction of the mental image is available and of the

entropy encoded data is transferred to the decoder if 100%." The transfer route may need some physical reduction in the amount of data being delivered and so a truncation may happen.

This truncation of entropy-coded data reduces detail in the image and the codec is therefore lossy [13]. It is possible for entropy coding to produce a much shorter image representation with average content by utilizing short coded words for more likely bit sequences, and longer code words for less likely sequences. One of the initial examples of entropy coding was Morse code [20].

Meanwhile, "mpeg video compression lie ins of numerous changeover and compression formats. Always descriptor a trade offs between quality, data volume and computational complexity [21] and at every step other critical compression issues come about. However, the area of application of the video will finally decide which compression standard will be deployed. Almost the same procedure to attain an optimal compression with an excellent possible quality most of the other compression standards apply".

1.2. Application layer QoS control

This process consists of application layer techniques such as congestion control, error control and differentiated services which are implemented at the application level. Consequently congestion control is utilized to stop packet loss and limit delays. Also the process of error control is employed to develop the display quality in the incidence of packet loss. The mechanisms of error control include Forward Error Correction (FEC), retransmission, error- resilient encoding and error concealment. Differentiated services provide quality of services by dividing traffic into classes.

1.2.1. Congestion Control

Congestion happens frequently in communication networks when the offered load goes beyond the designed limit, leading to deterioration in network performance such as throughput [34]. Network congestion usually caused by packet

loss, excessive delay, jitter and collision in multiple access networks has disturbing effect on the video presentation quality. To reduce these effects congestion control mechanisms are employed. For multimedia streaming, congestion control uses rate control and rate shaping.

- Rate control

"Control procedures are often used to bring down the amount of network load, in order to foreclose the unwanted signs of congestion. Like these are called rate control although they are also named congestion control measures. It should be noted that different network engineering may employ charge per unit control in different degrees, such as hop to hop degree [34]".

"Load, ascertaining an appropriate transmission charge per unit can be difficult and for situation such as the Internet where little can be assumed about the network scenario. The charge per unit control method applied in the transmittal control protocol has been empirically proven to be sufficient in most cases, nonetheless. In the Internet, transmission control protocol is the machine in the bringing of web pages, e mails, being the dominant traffic type and some streaming media "[30].

TCP-friendly rate control is based on emulation of the steady-state response function for TCP; the response function is a mathematical model for the throughput of a TCP connection, a predication of the average throughput, given the loss rate and round-trip time of the network: The TCP rate control formula is given as:

$$T = \frac{s}{R \sqrt{\frac{2p}{3} + 3p(1 + 32p^2) * T_{\beta\left(\frac{3p}{8}\right)}}} \dots \dots \dots \dots \dots \dots \dots \dots . \dots \dots \dots \dots . (1.4)$$

In this formula, s is the packet size in octets, R is the round-trip time [14] between sender and receiver in seconds, p is the loss event rate, and T_{rto} is the TCP retransmit time out in seconds.

- Rate shaping

Charge per unit defining techniques moderate the charge per unit of traffic generated by the video encoder according to the present network circumstances. The network, with which the encoder's end product charge per unit can be corresponded to the available network bandwidth and a charge per unit chiseler is an interface (or filter) associating the encoder. Is applicable to both domicile and stored video [19] and since charge per unit defining does not necessitate interaction with the encoder, charge per unit defining is applicable to any video coding scheme. The transportation of video over connections with constant bit rate (e.g., telephone or cable TV networks) is a method that has long been in use, though t he rate of the video sequence alter s very often as a result of the scene content and motion. The changing charge per unit video is established to a buffer which is emptied at a constant charge per unit. In such condition, The video encoder can accomplish constant charge per unit by moderating its compression parameters based on feedback info such as the buffer occupancy degree [5]. This multiple diminish, additive increase performances that conforms well to the condition of the network[13].

"If median loss > tolerable loss

Max_rate = max(max_rate / GAIN .min_rate)

Else

Max_rate = max(max_rate / INC .min_rate) [13]"

"This multiple decrease, additive increase mechanism adapts well to network changes [13]".

1.2.2. Error control

The mechanisms of error control include Forward Error Correction (FEC), retransmission, error concealment and interleaving.

- Forward error correction (FEC)

The principle of FEC is to lay appended (redundant) data to a compressed video bit stream so that the original video can be rebuilt in cases of bundle loss [19]. However, bundle losses in a bundle network one characteristically applies block code (e.g. Reed Solomon or Tornado codes) that take K data bundles in which N-K of the packets are superfluous packets. As far as any K of the N bundles is correctly gotten the original data can be recovered, for certain code. There is increment of required bandwidth by a factor of N/K [34] "if redundant packets are added. Because media applications are sensitive to loss, FEC schemes have been proposed and standardized for real transmission protocols applications and because the

Internet is a lossy medium, The bit streams are reconstructed from these methods, the type of FEC used also determines the quantity and also on the loss type [14]".The "the bundles are generated from each independent frames(I, P, or B) as shown in Figure .Thus, when K packets are transmitted with (N-K) redundant FEC packets with packet loss" [25] probability P_{avg} the successfully probability after encoding is defined by a binomial trial.

$$P_{avg}(N, K, P_{avg}) = \sum_{i=K}^{N} \binom{N}{i} (1 - P_{avg})^i . P_{avg}^{N-1} \dots \dots \dots \dots \dots \dots \dots \dots \dots (1.5)$$

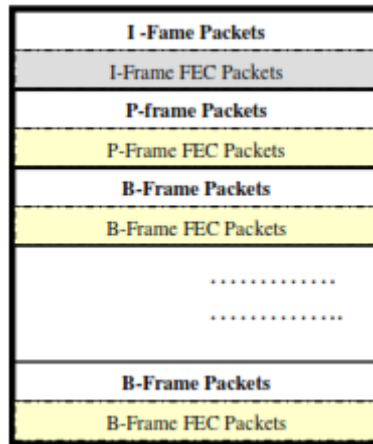Fig 1.5 Arrangement of source and FEC Packets in frame-level FEC technique [25]

"In practice, several factors indicate that FEC can provide only probabilistic repair. Key among those is the fact that adding FEC increases the bandwidth of a stream. In particular, adding bandwidth to the stream may increase congestion, worsening the loss that the FEC was supposed to correct."
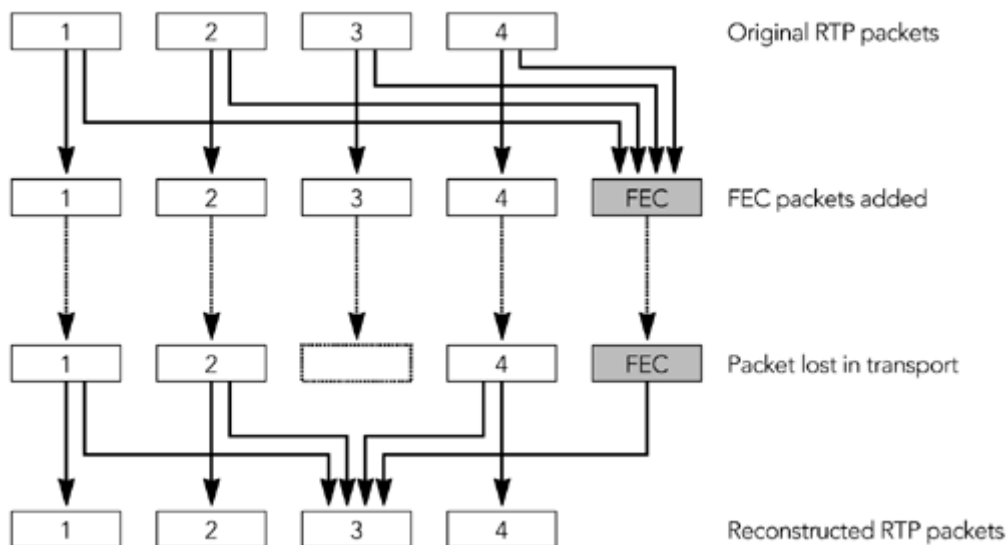

Fig 1.6 Repair using parity FEC [14]

FEC creates certain merits and demerits. In comparing to re transmissions, FEC does not need a back channel and may offer bring down delay since it is not dependent on the round trip time of retransmit. "Demerits of FEC involve high max end-to-end delay taken to recover lost packets and the overhead when there are no of packet losses. Most importantly, FEC based approaches are contrived to defeat a predetermined amount of loss and they are efficient if they are appropriately matched to the transmission channel [34]".

- Retransmission

In addition, retransmission necessitates a feedback channel, and this may not be feasible or attainable in various applications such as broadcast, multicast, or point to point without a feedback channel. The delay that occurs in most applications such as browsing, telnet and FTP because of retransmission is allowable. In some situations, when definite delivery is needed (and a backchannel is available) then feedback-based retransmits offer a powerful solution to channel losses" [34] . Retransmission is a natural approach to error correction, and it works well in some   scenarios. It is, however, not without problems that can limit its applicability. Retransmission is not a part of standard RTP; however, an RTP profile is under development that provides an RTCP-based framework for retransmission requests and other immediate feedback [14]. "numerous schemes of retransmission do exist which are. e.g for video streaming of time-sensitive data one may use delay-subdued retransmission where packets are re-transported if they could reach before deadline, or priority-based retransmission, where more vital packets are re-transported before less priority packets." [12].

- Error concealment

Error concealment is a viable technique to handle bundle loss, since human eyes can brook a certain degree of deformation in video signaling. Hence, the basic draw near in mistake concealment is to exploit the correlation by performing some descriptor of spatial and or temporal interpolation and to quantify information loss from carefully received multimedia data. "Considering the situation where a single 16x16 block of pixels (a macro block in MPEG terminology) is drop, [56] this clearly shows basic idea of concealment of error. However the above example does not show the exact data originally lost during video streaming; hence it is a very important example for conveying the basic concepts" [34]. There are two techniques for mistake concealment, spatial and temporal interpolation. Whereas in temporal interpolation lost info is reconstructed from the data in previous frames, in spatial interpolation lost pixel numbers is recovered by applying

adjacent spatial interpolation. Both temporal and spatial interpolation almost serve the same purpose but spatial interpolation recover loss frames in I-frames while temporal recover lost frames from P and B-frames respectively. The gap that is supposed to be replaced with content from bundle normality needs to be replaced with something [23].
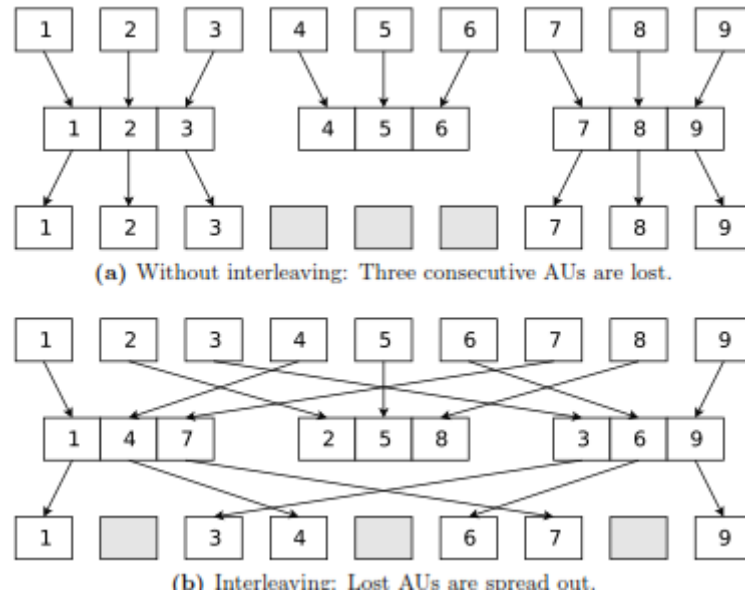

(a) Without interleaving: Three consecutive AUs are lost.

(b) Interleaving: Lost AUs are spread out.

Fig 1.7: Illustration of the consequences of packet loss both with and without interleaving [23]

If packet N is lost, the receiver cannot just jump directly from packet N- 1 to N + 1 since this will lead to time synchronization issues. The gap that is supposed to be replaced with content from packet N needs to be replaced with something [23]. "There are many concealment error algorithms developed in order to be used with spatial and/or temporal interpolation. In general compensation of motion rules presents the best concealment (assuming an accurate motion vector estimate). Inverse problem, leading to the contrive of sophisticated algorithmic programs (typically iterative algorithms) that furnish improved error concealment in many cases [34]" or this problem can also be created as a signal recovery. Error concealment lets the application to maneuver even if the network misbehaves. "An excellent error concealment scheme provides the disparity between a device that can be used in the actual world and one that is continually non functional when exposed to the wide deviation in loss rates intrinsic in the Internet. For this reason, all applications are supposed to put into service some form of error concealment [14]".

- Interleaving

The interleaving process reorders information prior to transmission so that initially adjacent data is separated by a guaranteed distance during transport. Interleaving is useful because it makes bursts of consecutive packet loss in the transport stream appear as isolated losses when the original order is restored. In figure 1.8 the loss of four consecutive packets in the interleaved stream is transformed into four single-packet losses when the original order is reconstructed [14]. The actual loss rate is unchanged, but it is typically easier for a receiver to conceal a series of single-packet losses than it is to conceal a longer burst of loss.
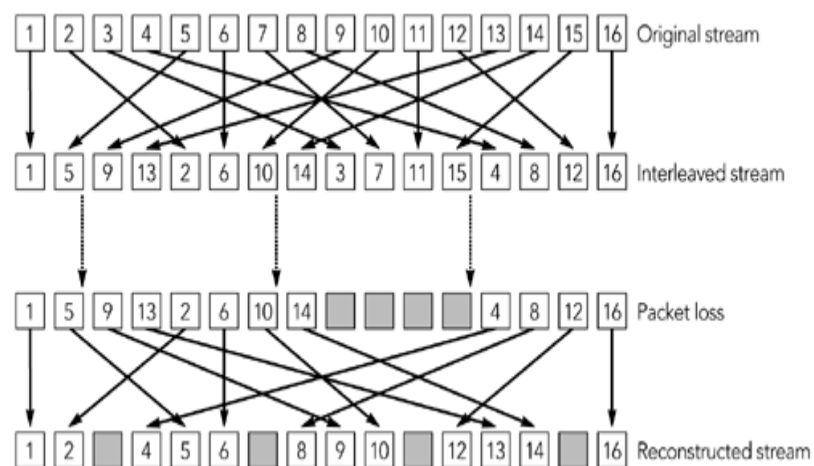

Fig 1.8 Interleaving, transforming burst Loss to isolated loss [14]

Interleaving adds considerable latency to the transmission process. The same number of packets as in the interleave group at both sender and receiver will be buffered [13]. Interleaving has a certain disadvantage, in that it increases latency. "Its benefits are limited for interactive applications such as internet phone, even though it can function well for streaming stored audio and video. As an advantage interleaving does not hike the bandwidth requirements of a stream".

1.2.3. Differentiated Services

Currently the architecture of the internet is called Best Effort (BE). BE gives no guarantees and treats all packets equally. While this is a reasonable solution for regular data traffic and file transfer, multimedia traffic could benefit from a service differentiation due to requirements for low delay and a low packet loss ratio [23]. "In the sense that it does not define particular services, the Diffserv structure is adjustable or service classes as in the case

14

of Intserv. The Diffserv architecture furnishes the functional components, i.e, the "pieces" of a network architecture, with which such services can be developed, instead". [4]

Diffserv marks traffic at the input to a network and is unmarked when the packet leaves that network. To operate, Diffserv requires a service level agreement at network peering points. There is no holy grail for improving the quality of service for media streams crossing the Internet. There are a number of solutions that deployed together can improve the streaming experience, without requiring a complete swap-out of the network infrastructure. Diffserv can assign packet priorities [20]. Given these two architectures it is important to evaluate the performance benefits from using the more advanced approach. The delay and jitter for voice traffic in Diffserv and BE environments are compared in [54],and it is found out that Diffserv leads to better performance especially for the codecs using high bit rate.

1.3. Continuous media distribution services

Application layer back up is not enough, in order to achieve high quality multimedia display at the end product. Adequate network back up is necessitated to cut down bundle losses and transport delays. Achieving efficacy for streaming video and audio over the Internet and they are contrived to support QOS. The network back up includes network filtering and multicasting and caching.

❖ Multicasting

Unicasting a single stream is presented to the Internet as a multicasting and multitasking offers an optional to conventional streaming. To the same stream the client initiates a multicast the server just delivers the stream to the network and all the viewing audience then can attach. Moreover TV audience just attach to the same stream [20]. The server has no cognition of where the stream is heading, in contrast the normal TCP client server handclasp interactions of a Net connection. A client will be doed aware of a multicast by some out of band channel it could be by mail or via internet. The watcher

then call for the multicast at the appropriate date and time and optionally session protocol is considered. The working principle is briefly describe as follows: Each multicast able node (called media bridge) [24]) performs routing at the application layer, in the media multicast networks. More neighboring media bridge through explicit configuration, which defines the application degree, overlay topology or in addition, each media bridge is completed with one.

The media bridges in a media multicast network apply a distributed application degree multicast routing algorithmic program, collectively and to ascertain the virtual paths for propagating content throughout the network. Becomes overly clog uped, the media multicast network automatically and intermittently re routes content via alternate paths according to application degree routing policies or when the underlying network fails [19]. Only when a downstream client calls for it and in addition, media bridges dynamically subscribe to multicast content. Hence, multitasking is not curtailed to streaming it also can be used for wide ambit data bringing (like database upgrades via a separate endeavor, or for video call).

❖ Network filtering

Network filter outing is a method of congestion control that aids in stabilizing video quality during network peak time. There is peculiar sort of network filtrating methods such as applying TCP congestion control algorithmic programs and network filter. Transmission control protocol is the well known example of sliding window protocol. The origin includes sequence numbers with each data bundle it sends, and these are reflected back to it in ACK (positive acknowledgment) packets from the receiving system. The easiest sliding window protocol necessitates each bundle to be known immediately, before next packet could be send. This is referred to as a stop and waits protocol because the sender must wait for the notification before it can send the next bundle. Distinctly, a stop and wait rules "preventing the sender from infesting the receiver but it also hinders performance, wait protocol furnishes flow, Receiving system is long and especially if the round trip time between sender is lengthy.

❖ Content replication (Caching)

.   A key technique for ameliorating quantifiability of the media delivery system is content media replication. The content replication takes two descriptors, namely, caching and

mirroring, which are deployed by publishers, service providers. Both caching and mirroring seek to lay content closer to the clients and both share the following merits.

- Consumption of bandwidth across the link is decreased.
- Streaming servers load is decreased.
- Latency for clients is also decrease.
- Accessibility is also increased.

"Mirroring is to lay copies of the original multimedia files on other machines break up around the Net. While replica of the original multimedia files are to be found on the replica servers the recent multimedia data files are on the innermost server, in this fashion, the close replica server, which gives the clients the most advantageous performance clients can be recovered from multimedia data [19]".

On the other hand, "caching is based on the dependence that numerous clients will load many of the equivalent contents, does local copies of contents that the client's restore. Ordinarily, clients in a single group restore all contents from a single neighboring workhorse, named a cache. From the source server, storing a copy locally, the cache recaptures a video file and frontwards it to the user who beseeches it. If a user request for a video file which the server has already in its database, the cache retreats the replica copy rather than going all far to the parent server where the video file domiciles [41]".

.

1.4. Protocols for streaming media

Standardized for communicating between end users and streaming servers are called contrived media streaming protocols. These protocols are Network layer protocol such as internet protocol Transport layer protocol such as user datagram protocol and transmission control protocol Session layer protocol such as real-time streaming protocol Brief verbal descriptions of the above media streaming protocols are given below

➢ Network layer

On the network layer, net protocol is the most known protocol. But is essentially an unreliable transmission system medium IP is not as complex as OSI model. IP is the de-facto network rule which is situated at layer three communication theory protocol. There are three main kinds of setbacks with information science variable network latency.

Bundles arrive in different order; bundles may be lost along the route to destination or disfigured.

- Variable network latency.
- Packets arrive in different order
- Packets may be lost along the way or disfigured.

These setbacks could not be bushel by IP itself but can be remedied by the upper layer protocols, i.e the transport layer and seventh layer of the OSI model. [38] It ought to be noted that IP packages can be integrated in any network support either via physical link or radio frequency signals. The layer 3 that applies IP could be considered a convergence point towards which all multimedia networks may efficiently relocate in the nearest future.

➢ Transport layer

Of the whole protocol hierarchy the layer 4 is not just another layer it is the heart. Its undertaking is to furnish reliable, cost efficient data packets transmission from the source station to the receiving station, not dependant of the physical network or network in use. The whole concept of layered protocols would do little sense, without the transport layer [53]. On the transport layer for regular data transport, the transmission control protocol is often used. Signifying that the connection has to be lain up before transmission begins; this is a connection oriented protocol. It is assured that the sent data points are actually gotten and this is also a dependable protocol. The receiver will convey back a notification for each package received [23]. Within a certain time frame the bundle will be retransmitted if the sender does not get this notification. The received bundles are rearranged and in correct manner delivered to the upper layer. With TCP is that this protocol adjusts its sending charge per unit to the current network conditions which is another important feature. TCP will increase the sending ratio as long as notifications are received from the receiving stations. When notification got missing, it is assumed that this has take placed due to variation of network and load on the network and this results in reduction of sending ratio. By doing this all TCP flowing will be struggling for the network resources on equal footing.

Furnishes a minimal lay of extensions to internet protocol Layer 4 protocol such as user datagram protocol furnishes the raw IP service. It does not furnish any enhanced

dependability to the transport (although the checksum does allow for detection of payload errors that IP does not detect), neither does it affect the delivery of timing packets. To the transport layer an application utilizing layer 4 protocols such as UDP supplies data packages, which convey them to a port on the receiving station machine (or to a group of receiving station machines if multicast is used). Those bundles may be drop, scattered and delayed as the same as traditional IP network [14].

➢ Session layer

The session layer allows users on different machines to establish Roger Huntington Sessions between them. This layer offer services such as "dialog control (keeping track of whose turn it is to transmit), token management (preventing two parties from attempting the same critical operation at the same time), and synchronization (check pointing long transmissions to allow them to continue from where they were after a crash) [53]".

The real time streaming protocol is an application degree protocol for the moderating of real time multimedia data. An extended model rather than a protocol RSTP furnishes. It allows videocassette recorder like control of the playback play, pause, and so forth. Changing the media bandwidth to case the available capacity, a streaming server also can respond to network congestion [38]. Intentionally to be similar in syntax RSTP was developed, operation to HTTP version 1.1. It does differ in several important aspects. However, Server can issue call for during interaction with hypertext transfer protocol the client always issues the requests and with this protocol whereas hypertext transfer protocol is stateless, this protocol has to retain the state of a session.

CHAPTER 2

Transmission of high quality audio and video over ATM network

      In this chapter H.264 compression method is discussed, as well as a brief introduction about ATM technology; ATM classes of services such as constant bit ratio, Available bit ratio, variable real time bit rate and an unspecified bit rate that determines bandwidth allocation and call admission control for each service category that passes through ATM network. Video and audio Compression/coding (refer back to chapter one for explanation), mapping of bit streams into ATM cells, specifications of QOS such as delay of cell, cell loss rate, and jitter that makes an ATM very suited for conveying of multimedia traffic. selection of AAL (ATM adaptation layer), Issues in media transport over ATM and effectiveness of ATM network technology as regards to conveying high quality multimedia and perceived quality will be discussed in this chapter.

2.1 H.264 / SVC compression method

2.1.1 Brief History

H 264 is component part of the advanced video coding (AVC) project aimed at furnishing good video quality with reduced bit rates than the previous standards without increasing the intricacy of the contrive. In this manner it would be easy to implement resulted in H.264 standards. The H. 264 codex offers substantially upgraded coding effectuality at the same video quality as the mpeg 4 and H.263 schemes [52]. The major aim of this new ITU-T h 264 standardization try was to better the compression rate, at the same time supplying a network packet video representation addressing both time video telephony and non-conversational i.e storage, broadcast or streaming type applications [52]. Amidst the vital constructs of H.264 is the separation of contrive into two person layers, namely: the video coding layer and the network adaptation layer. The video coding layer manages expert depiction of the video content, while the network adaptation layer deals with the packaging of the encrypted data in a suitable method for transmittal via the network [3].

2.1.2 How H.246 Compression Works.

      Video compression deals with the diminishing and removing of unwanted video data points, so as to successfully send and shop digital video file. The process embraces applying

an algorithm to the origin video to create a compressed file that is laid for transport. "To play the compressed video data, an inverse algorithm is used to bring forth a video that demoes virtually the exact content as the original source video. The time taken to compact, transport, decompress and display a video data is named as latency [3]". Given same power of allotment, the more sophisticated the compression algorithmic rule is, the higher the latency's pair of algorithms that work together is named a video codex (encoder/decoder). Video codecs which utilize different standards are specifically not well suited to each other that is, video content that is compressed applying a particular standard may be decompressed employing a peculiar standard. For instance, an MPEG-4 Part 2 decoder is incompatible with an H.264 encoder. This is simply because one algorithm cannot properly decipher the end product from a peculiar algorithmic program. But it can apply several peculiar algorithmic programs in the same software or hardware, which would then speed up numerous formats to be compressed. Several video compression standards manipulate different methods of lessening data points, and thus, results vary in bit ratio, perceived quality and latency. Because decorators of an encoder can make up one's mind to apply different lays of tools defined by a standard, the consequences from encoders that employ similar compression standard may also vary. If the end product of an encoder comply to a standards format and decoder, it is likely to do numerous implementations.

This can be of importance because several implementations have different purposes and scheme. Expert non real time software package encoders for understanding optical media ought to have the alternative of being able to convey enhanced encoded video compared to a real time hardware encoder for video conferencing that is in a mitt held device. [3] A gave standard, thence, may not guarantee a given bit charge per unit. A decoder necessitates all the necessary parts of a standard for it to decipher a submissive bit stream, contrary to an encoder. This is because a standard demoes exactly how a decompression algorithm ought to have recover each bit of a compressed video.

2.1.3 Types of Frames Used in H.246 Standards

Three types of frames are used in H.246 compression format:

i. The Intra- frame (i frame) is a self sufficient frame that can be autonomously decoded without any mention to new images. In a video sequence is always an I frame the first image. Resynchronization points if the conveyed bit stream is deformed or I frames are required as starting points for new audiences. To utilize fast forward, rewind and random access functions I frames can be employed. At time intervals, an encoder will routinely infix I frames or on demand if new clients are likely to join in viewing a stream. The setback of I frames is that they utilize much more bits; they do not result in many artifacts, but on the other hand.

ii. The predictive inter frame (p frame) does references to parts of earlier I and/or P frames to code the frame. A P-frame usually needs littler number bits than an I-frame, but one of its setbacks is that it is very prone to transmission errors as an effect of compounded dependence on earlier P and I -frame,

iii. The bi-predictive inter frame is a frame that makes references to both preceding reference frames and a prospective frame [3].
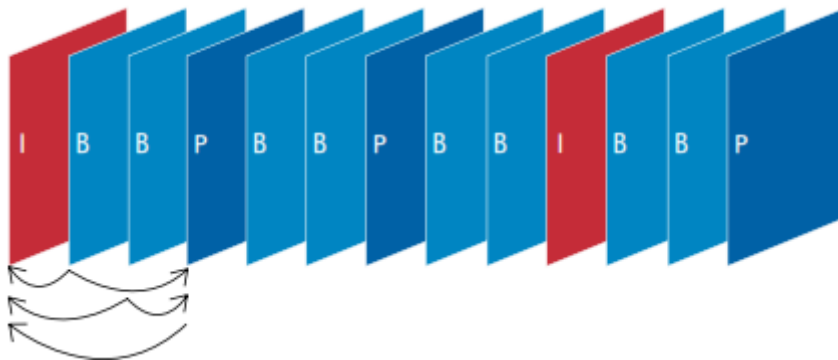
Fig 2.1 Diagram showing I, P and B frames [3]

2.1.4 Entropy Coding

It depends on future infinite length codeword lays and the first of which is UVLC [33]. A different VLC table for each phrase structure component the mapping to the single UVLC table is customized according to the statistics of the data to be encoded instead of contriving. The CABAC [40] can be used to ameliorate the efficacy of entropy coding. Moreover, the implementation of arithmetic coding gives room for the assignment of non integer number of bit to each symbol of an alphabet [37], The application of adaptive code helps adaptation to

non stationary symbol statistics, this could be particularly advantageous in relation with symbol probabilities higher than 0 5.

2.1.5 De-blocking Filter

A de-blocking filter is enclosed in the H.264 codex contrive, so as to extenuate the degraded visual effects of block edge artifacts. In comparison to the MPEG-4 codecs, where the de-blocking filter is an optional portion of the recommendation [51], hence descriptors an integral component part of the recommendation and in the H.264 codecs the filter is applied within the motion prediction loop. The cut off frequency of the filter outing is familiarize controlled by the values of several coding phrase structure elements [37].

2.2 Brief Introduction

The ATM is being designed by a consortium of a few hundred companies regrouped under the ATM forum. The ATM forum publishes recommendations that define the ATM protocols. The development of ATM started back in the 80's when there was not for a technology that can transport multiple multimedia data without experiencing bandwidth constraints, delay, and congestion to the respective end stations. One good aspect about the technology is that it can transport voice, audio and data at changing speed from T1, kilo bit per second (Kbps) and Gbps with an acceptable quality. It was developed by the ATM standard body to cater for   real and non-real time services making it very convenient for conveying multiple media applications. For this reason, ATM is widely accepted not only by the multimedia company but also academia. The ATM technology was finally standardized by ITU-T in 1987. ATM is based on switching of bundles of data cells and connection oriented that is connection has to be established before transmission can take place. The packet used in ATM network is called a cell which is a small size packet with payload of 48 bytes and a 5 byte header. ATM does have an inbuilt mechanism that allows it to attain to numerous quality of service to several traffic types.

2.3. Issues in Media Transport over ATM

In order to successfully transmit audio and video over ATM network and also to ensure satisfactory quality, a number of design strategies have to be addressed. These strategies are discussed as follows:

2.3.1. Choice of ATM adaptation layer (AAL)

This is sand witch layer that is in charge of dealing with external data traffic across the ATM network. It consists of two layers which are: convergence sub layer and reassembly sub layer. The convergence sub layer is responsible for recuperating timing information and covers generation e.g it deals with issues like cell delay, cell loss and bandwidth constraints. The -reassembly sub layer, which is at the sending side deals with the division of higher layer PDUs into appropriate size for the ATM cell. At the end station, it rearranges the ATM cells into upper layers of personal data unit [27].

Traffic Class	Class A	Class B	Class C	Class D
Timing relation between source and destination	Required		Not required	
Bit Rate	CBR	VBR		
Connection Mode	Connection-Oriented			Connectionless
AAL Type	AAL1	AAL2	AAL3/4 or AAL5	
Example Application	T-1, E-1 circuit emlation	Packet video, audio	FR, X.25	IP, SMDS

Fig 2.2 Traffic classes [39]

There five versions of the AAL that are designed for five classes of traffic. These layers are used to convey data to the users at the receiving end. AAL-1 for rt-CBR, AAL-2 for rt-VBR, AAL-3 for link-oriented packet streams, AAL-4 for datagram's, and AAL-5 for IP packets and can be used for datagram applications and for variable bit rate applications such as high definition television [32].

⬩ ATM adaptation layer 1 (AAL 1)

In [27] this layer was meant for servicing class A traffic, and the application of this layer is found in video conferencing, interactive media storage, interactive audio and video in which the CBR has to be kept constant for the duration of the session. It handles CDV, detect drop and misdirected cells. AAL 1 consists of two layers namely: a SAR sub layer and a

convergence sub layer (CS). The SAR sub layer detects, correct and transmit blocks of codes received from convergence sub layer. However, The CS sub handles many of functions, such as variation of cell delay, dealing out of the sequence count, ordered and unordered data transfers, and conveying of timing data [27]. The convergence sub-layer of AAL-1 groups the bit streams into 47-byte packets. The SAR sub-layer adds a 1-byte header that consists of a convergence sub-layer indication (CSI), a number sequence (SN), and a sequence number protection (SNP). The use of CSI is optional. The destination uses the SN to detect missing cells. The SNP can correct solitary bit errors and detect numerous bit errors in the SN.
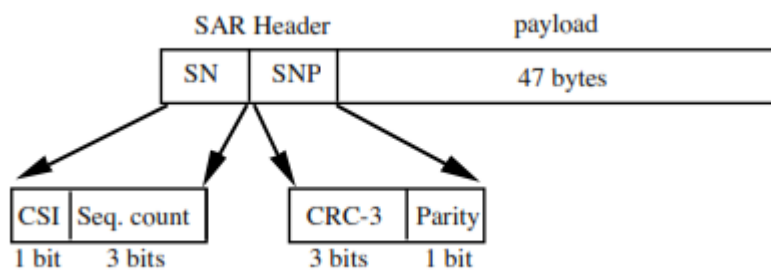

Fig 2.3 The SAR encapsulation for AAL-1[27]

.

🞣 ATM adaptation layer 2 (AAL 2)

This layer was meant for servicing class B data traffic types. It has been used to render an effective means of conveying numerous applications that are responsive to delay and minimal bit rate over ATM network such as voice, fax, and voice and data transfer [27]. The AAL-2 is also used for compressed audio and video data (segmented isochronous traffic). Hence the compression of class B audio and video data does not obviate the need for latency. The functions of AAL-2 are as follows:

🞣 Dealing with cell delay variation (jitter)
🞣 Packetization and reassembly of data
🞣 Alerting the higher layers of uncorrectable errors in AAL
🞣 Conveying of timing information between source and destination
🞣 Handles PDU from variable bit rate source
🞣 Observance and dealing with AAL PCI bit errors
🞣 Recovering of drop and mis-directed cells
🞣 Regeneration of source frequency
🞣 It compressed audio and video data without losing quality
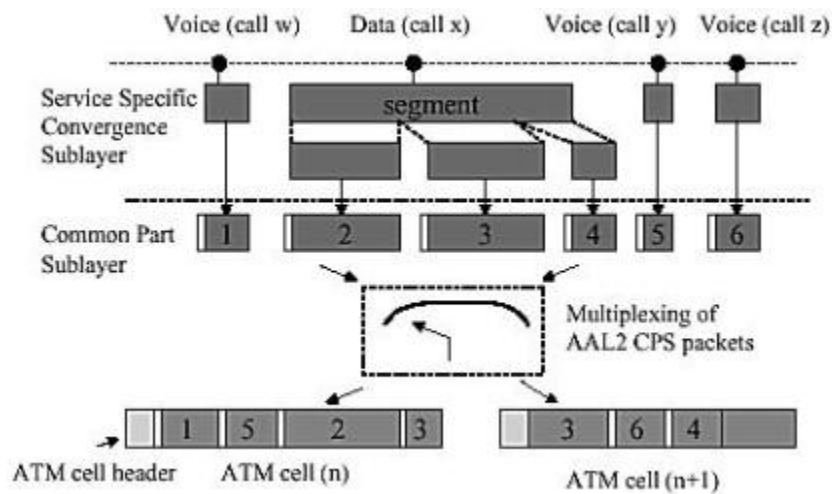🞣 Observing and changing of bit errors in user information field.

Fig 2.4 AAL-2 Multiplexing [26]

In [26] AAL2 has the ability to multiplex application packets from diverse users over similar asynchronous transfer mode link. Packets from different user connections can be categorically multiplexed over the same ATM connection. Each ATM cell is filled with packets from one or more user, such as multimedia packets and data packets. In order to avoid unnecessary padding and thereby increasing the efficacy, the AAL2 packet can be splice over two ATM cells. The working principle is depicted in the Figure 2.3. Above

✦ ATM adaptation layer 3/4 (AAL 3/4)

AAL3/4 has been separated before but is now combined and support class C and class D services. The primary purpose of this stratum is to furnish sufficient data to allow variable length bundles to be conveyed over the asynchronous transfer mode network as a flow of bushel size cells i.e Reorganizing procedure and the categorizing the AAL backups. The IEEE 802.6 standard is congenial with this stratum.

The SAR sub stratum and by the concurrence sub stratum the AAL 3/4 purpose are rendered. The common part sub stratum (CPS) and the concurrence sub stratum is further categorize into the service exact concurrence sub stratum (SSCS).The CPS is named common part concurrence sub-stratum (CPCS)[27].. This layer uses CPS, and SAR sub-layers.

Fig 2.5 Segmentation of CPS and SARs layers of AAL ¾ [39]

The definitions of the above layers are given below:

- ❖ CPI – Common Part Indicator which is the unit of measurement in both BA sizes and bits.

- ❖ BA Size- Buffer Allocation size which is the CPCS payload

- ❖ The Alignment field (AL) (consist of the trailer a full 32 bit's field to facilitate the receiver design)

- ❖ Length – Length of CPCS PDU

- ❖ ST - Segment Type (BOM = beginning of message, COM = continuation of message EOM = end of message, SSM = single segment message)

- ❖ Sequence Number (SN) (used to detect lost SAR PDUs)

- ❖ MID - Multiplex ID (allows multiplexing of up to 1024 different CPCS PDUs over the same VCC)

- ❖ LI - Length Indicator (= 44 for BOM or COM, can be less than 44 in EOM, SSM)

- ❖ CRC - Protects complete SAR PDU, including header, payload and trailer

🔸 .ATM adaptation layer 5 (AAL 5)

This stratum was initially designed by IBM but later adopted by ITU-T and accepted also by ATM forum. This layer supports LAN, FR, VBR and CBR audio and video data. It is

suitable for a connectionless service (class D). It was planned in reaction to AAL 3/4 which is needlessly complicated. [27] The concurrence sub-stratum (CS) and the SAR sub-stratum render AAl-5 services. CS is further separated into the SSCS and CPS. Originally CPS is named the common part of a concurrence sub-layer (CPCS).
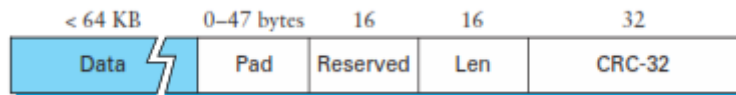

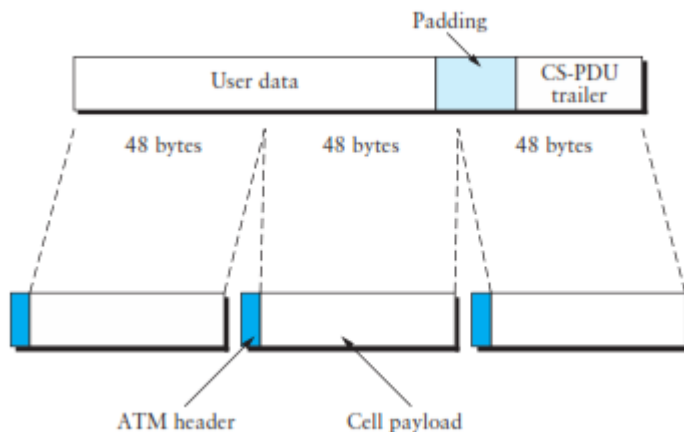Fig 2.6 ATM Adaptation layer packet format [10]

Fig 2.7 Encapsulation and segmentation for AAL 5 [10]

AAL-5 carries IP packets or other payloads with little overhead. The CS packages the information into CSU-SDUs with a length equal to 48 bytes. The CS-SDU contains a header whose use is under study. The 2-byte alignment field (AL) is filler. The padding (PAD) and length (LEN) are required to accommodate variable length payloads. The CRC is calculated over the complete CS-SDU. The SAR then puts the CS-SDU into back-to-back 48-byte cells with no additional control information. Lost cells are detected by the higher layers (e.g TCP or UDP).

2.4. Service class selection

In many respects, the QoS capabilities that are provided in ATM networks are related to those provided in an IP network using Integrated Services. However, [10] the ATM standards bodies came up with a total of six service classes compared to the IETF's three. There are six classifications of services provided by ATM namely: invariable bit ratio known as (CBR), real-time changing bit ratio known as (RT-VBR), non-real-time changing bit ratio known as (NRT-VBR), accessible bit ratio known as (ABR), undecided bit rate known as (UBR), and assured frame ratio known as (GFR). The CBR and RT-VBR are for real-time applications, whereas the other service divisions are meant for non-real-time services [27].

❖ Constant bit rate (CBR)

It is planned for concurrent services which transmit at continuous bit ratio, such as interactive media services such as live video call, audio and video interaction. Also given the continuous bit ratio remains constant at a specified period of time and the cell peak ratio quantifies the amount of data traffic across the network [27].The delay variation tolerance of cell is also given in this service.. A continuous bit ratio service is for actual time services, and therefore, the end-to-end delay has to be minimal. This service is the most costly of all the ATM services and also guaranteed bandwidth to the connection. The following applications used CBR:

- ✓ Video retrieval (Video on demand)
- ✓ Audio retrieval (audio library)
- ✓ Video distribution (TV broadcast)
- ✓ Interactive video such as video conferencing

❖ Variable bit rate (RT-VBR)

This is used for real time applications that are bursty in nature and allow efficient use of the network. Both delay variation of cell and delay tolerance of cell are strongly managed. [27] The parameters used in this service type are ratio of peak cell, the cell delay variation tolerance; the sustained cell rate the maximum burst size Peak-to peak cell delay variation and the maximum delay of cell tolerance .As in the continuous bit ratio service, the actual

time changing bit ratio service is also used for actual-time applications. The cell sustained ratio is an upper bound on the long term arrival of the rate cells. The tolerance of burst allows for the cells to arrive faster than at rate cell sustained ratio for some time. This type of service is mostly used in video coders and application that generate variable size frame.

o   Non-real-time variable bit rate (NRT-VBR)

This type of service is basically used in applications with more tolerant network delays. It is designed for non-delay sensitive applications which convey at a changing bit rate. The parameters used in NRT-VBR are: ratio of peak cell, [27] the variation delay of cell tolerance, the cell sustained ratio (SCR) and the maximum burst size (MBS). Since this service is not meant for delay prone services only the ratio of cell loss is stated. This type of service is applicable in the following areas:

- Baking transactions
- Process monitoring
- Reservation of Airline tickets
- Video storing and forwarding
❖ Unspecified bit rate (UBR)

This type of service for non-real-time applications with a changing bit rate. It is intended for applications that engage in conveying of data, such as file transfer, web-browsing, and email. None of the quality-of-service parameters are stated. The ratio of peak cell and the delay tolerance cell variation parameters can be stated but a network may disregard it. Ratio of unspecified bit type of service is the same to internet. [47] It does not specify any form of QoS guarantees from the network. Applications here can handle delay and latency without too much trouble. The network provider will make a best effort to deliver the data within a realistic amount of time.

➕ Available bit rate (ABR)

This type of service is basically used in non real time applications that endure network delays and cell loss. It also allows numerical gain without guarantee for throughput and allows congestion control in which the sender would be informed to slowdown the sending rate of the traffic. The ABR is suited for applications that can definitely tolerate longer

delays on the network. The ABR adapts to whatever resources the network has available, rather than a specified amount of throughput. Typically, the ABR also uses a minimum cell rate (MCR) as a means of controlling the traffic. The network continually feeds back information on the regular traffic and therefore allocates the MCR to hold the connection alive, while other traffic runs across the network in a priority over the ABR rate [47]. The Minimum cell rate for bandwidth subdued ratio of cell availability link to assure no underflow for a dormant user is:

$$McR = Max\left(\frac{F}{n-nT}\left(\frac{\beta + \alpha K}{\alpha + \beta}\sum_{i=1}^{n} xi - \frac{IcR\eta T}{F}\right)\right) \dots \dots \dots \dots \dots \dots \dots (2.1a)$$

Where $\eta \epsilon [\eta_T + 1, N]$

The likelihood condition for an ratio available bit link to subsist at a given quality of service preferred by a user for a given interactive level if only buffer size and cell

loss   are   given $Max\left(\frac{F}{n-nT}\left(\frac{\beta + \alpha K}{\alpha + \beta}\sum_{i=1}^{n} xi - \frac{IcR\eta T}{F}\right)\right) \leq Min\left(\frac{F}{n-nT}\left(\frac{\beta + \alpha}{\alpha}B + \right.\right.$

$\frac{\beta + \alpha K}{\alpha}\sum_{i=1}^{n} xi - \frac{\beta + \alpha}{\alpha}\frac{I_{CR}\eta T}{F} - $

$\left.\left.\frac{\beta}{\alpha}\sum_{i=\eta_T+1}^{n} x_i\right)\right) \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (2.1b)$

For all $\eta \epsilon [\eta_{T,} N]$

Examples of ABR applications include existing UBR applications (news and weather pictures, LAN interconnection, telecommuting, file transfer and electronic mail) that require more predictable behavior. [27] Additional examples are defense applications and banking applications with critical data transfer requirements. Such use of three types of traffic over ATM optimizes usage of capacity of the physical link.

↓ Guaranteed frame rate (GFR)

This service is meant for non delay sensitive applications that involve a cell minimum ratio assurance, but they can transport in surplus of their demanded cell minimum ratio. This

service arranges data in frames and the adaptation layer five convey the frames. The network does not warranty deliverance of the surplus traffic. When congestion happens, the network tries to drop all AAL 5 CPS-personal data units rather than each and every cell. [27] The frame guaranteed ratio does not provide any feedback to the user concerning the present intensity of congestion in the network. Hence, the user has to decide network congestion via a means such as transmission control protocol, and adjust its transfer ratio. The commonly specified parameters used in this service are: cell peak ratio, cell minimum ratio , maximum attainable size of the burst ,and maximum frame size .

Type of Service Parameters	Parameters During connection setup	Bandwidth guarantee	Burstyness	Network Priority
CBR	PCR and CDVT	yes	None	1
RT-VBR	PCR, CDVT, SCR, and MBS	yes	Small	2
NRT-VBR	PCR, CDVT, SCR, and MBS	yes	Small	3
ABR	MCR,PCR, ICR, TBE, RIF and RDF	optional	High	4
UBR	PCR, CDVT, and MCR	no	High	5

Table 2.1 ATM service characteristics suitable for multimedia [27]

2.5. Transport of packet encapsulation

In [30] after the AAL has been chosen, the next issue is how MPEG-2 transport stream packets are mapped into AAL-5 Service Data Units. Basically, 1 to n transport packets can be mapped into one AAL-5 SDU. For AAL-5 with a "null" service specific convergence sub-layer, ATM Forum requires that n = 2 must be supported for all Conformable equipment and with the following constraints:

- An AAL-5 PDU shall contain two TS Packets, except it has the last TS Packet of the only Program Transport Stream.

- An AAL-5 PDU shall have only one MPEG-2 Transport Packet, if that MPEG-2 transmit packet is the last transfer packet of the single program transport stream.
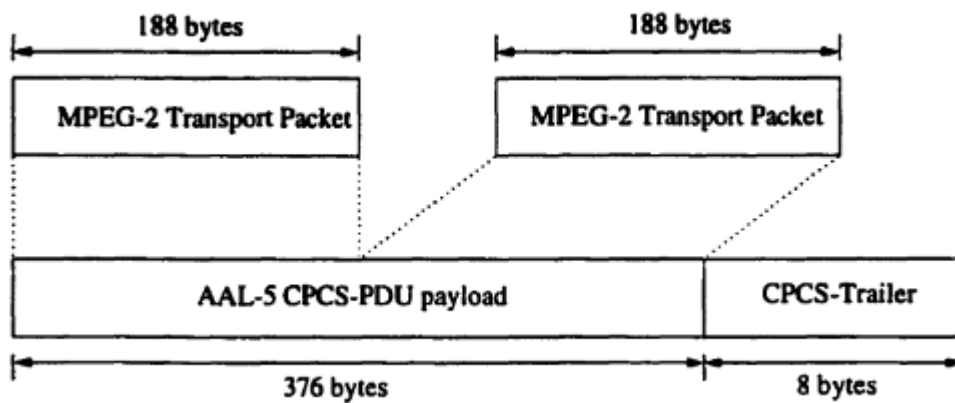

Fig 2.8 Mapping of MPEG-2 packets stream according to ATM forum [49]

We can see that the transport packets need 376 byte, which are mapped together with the CPCS trailer of 8 byte into the payload of exactly 8 ATM cells.. N > 2 is also allowed, as long as the stuffing byte used in SAR sub-layer is minimized.

2.6. Quality of service parameters

There are different guidelines that can be used to state the QOS of a link, such as, Traffic shaping, congestion control, ratio of cell loss, ratio of transmission delay, and variation of cell delay.

2.6.1 The Cell delay variation (CDV)

Variation of cell delay is the total of time between the maximum and the minimum end-to-end cell transit time. Variation of cell delay can merged multiple links onto a single physical link and the changeability in ATM switching queuing delays. In the codex mpeg 2 video systems apply a 27 MHz system clock. The trading operations at the decoder with those at the encoder this clock is used to synchronize. Audio data to be properly corresponded and this enables video and the retrieval of frames from the decoder buffer to avoid well over also moderates. Decoder in coordination with each other the encoder lay program clock references (PCR) sporadically in the Ts and to keep the encoder. These are used to regulate the system clock at the decoder as necessary. "If there is jitter in the cells of asynchronous transfer mode the peak cell loss ratio will also have jitter? This will lead to picture quality disfiguration [18]".

2.6.2 Cell Loss Rate (CLR)

The CLR is a means of determining the ratio of drop cells to the overall number of cells that have been transmitted. There are many reasons why cells get lost, yet the goal is to hold the amount of lost cells to a minimum. [47] An ATM switch can discard calls that have been corrupted, especially the header. Moreover, cells can be explicitly marked (tagged) as eligible for discard if the network gets congested. The formula for the ratio is as follows:

Cell Loss Rate (CLR) = Lost Cells/Total cells transmitted

"The ratio of cell loss relies on diverse conditions which include: the physical media in use, the switching technique, the switch buffer size, the quantity of switches traversed in a link, the QoS class used for the service, and whether the video stream is continuous ratio bit and changing ratio bit [18]".

2.6.3 Cell transfer delay (CTD)

The CTD is calculated as the total number of elapsed time from when a cell enters a network (a switch) until it exits the network. [47] This takes into account the total amount of inter-nodal processing time, buffering time, and propagation delays across whatever the medium is used.

Cell Transfer Delay (CTD) = Node Processing + Buffer + Propagation.

The fixed cell transport impediment is the total of all fixed delays that a cell come across from the transmission station to the receiving station, such as, delay of propagation, fixed delays influenced by conveying systems, and permanent switch processing times. The variable transfer of cell delay, called peak-variation of cell delay, is the total of all variable delays that a cell comes across from the sending receiving station to the transmission station. These delays are primarily as a result of inline delays in the switches all along the cell's route [27].
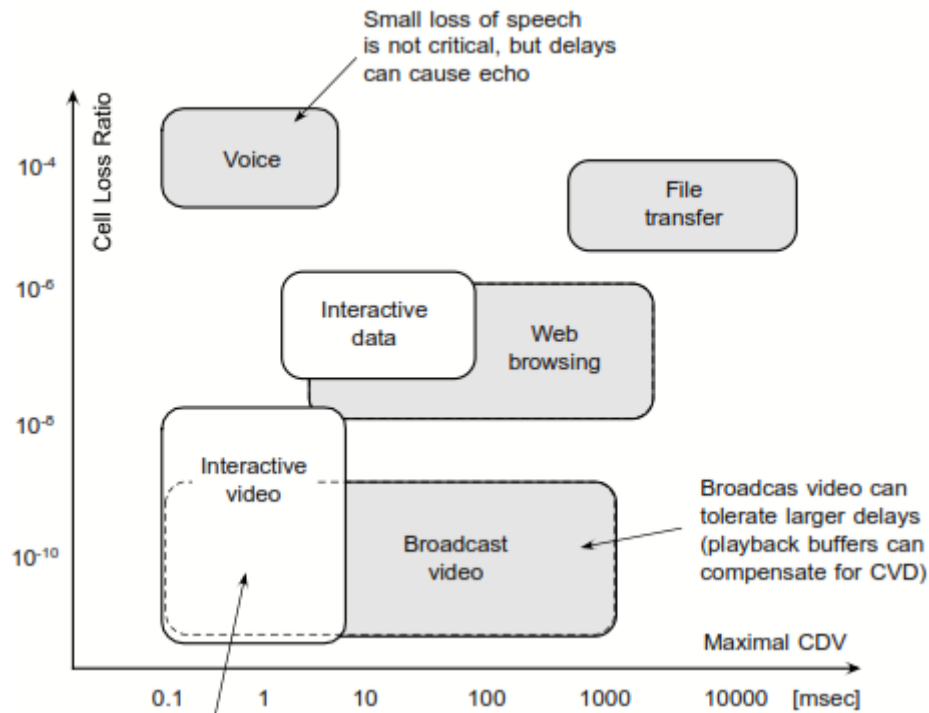
Fig 2.9 Diagram showing basic quality of service parameters [39]

2.6.4. Traffic shaping

When traffic does not conform to the footing of a traffic contract, the traffic can be routed through a defining mathematical function to bring it into conformance with the contract. Other parameters associated with the contract remain valid for traffic defining, CTD and the CDV are important parameters and for example, real time VBR traffic whose peak charge per unit bursts are higher than the contracted peak cell charge per unit can be modified to extend the duration of the burst by spreading the cells in time as long as the burst tolerance. For example, more CTD for less CDV during the traffic defining process, it is possible to trade off CTD against CD. The sources of CTD and CDV are the ATM switch over the ATM access concentrators [1].

The transfer is shaping function changes the ATM transfer flow to attain better network efficiency, for interworking to ATM switches with small buffers and for virtual paths (VPs) originating at the ATM switch. Egress side of an ATM switch or traffic defining can be done either at the entering. Traffic defining at the entering side is used to modify the traffic flow of terminal equipment and ATM switch over that are not able to keep the traffic contract sending ATM cells with a higher peak bit rate than allowed. In ratio of peak cell traffic shaping includes diminish; traffic defining includes decrease in ratio of peak cell [27].

2.6.5. Congestion control and flow control

Allows the management of traffic when many users contend for finite network resources congestion control . Congestion management is one of the most important mathematical functions that act as standard ATM switches over will have to perform. Ratio of unspecified bit service traffic and especially for required bit ratio . The asynchronous transfer manner continuous ratio bit .To avoid network congestion conditions when implemented with robust connection admission controls changing ratio bit service classes have been intended and usage parameter controls. Ratio of unspecified bit service classifications have been defined to take advantage of the excess capacity beyond that required for continuous ratio bit and changing ratio bit connections [35] and required bit ratio. The network components are assumed not to do any congestion control, in the current Internet. As such, the endpoints rely on end-to-end protocols, such as TCP/IP, to regulate the flow of traffic, based upon packet loss. This framework can also apply for TCP/ IP over an asynchronous transfer fashion transport infrastructure.

In high speed ATM networks is to control the user traffic at the user network interface [16] one of the effective methods of congestion control. At the user network interface is also called rate- based access control this type of control. [41] The excess traffic from the user is cast aside by the network, if the user exceeds the allocated input rate. The charge per unit based access control serves as a traffic chiseler at the edge of the network, in a sense.

CHAPTER 3

Methods of Evaluation of Audio and Video Quality

In this chapter subjective and objective quality assessment methods are discussed. The reasons for discussing the methods are to evaluate the quality of service of the transported audio and video data according to the user's perception. The procedures for carrying out the subjective quality assessment methods are also the explained. In order to accurately calculate the data obtained from the subjective assessment method, objective quality metrics are going to be discussed. The reason for using the objective quality metrics to get the result of the data is because subjective assessment metrics are time consuming and also expensive to calculate the data. Hence, different objective quality metrics will be also discussed.

3.1. Subjective quality evaluation method

3.1.1. Introduction

The multimedia streams can be affected by distorted video frames due to conversions between analog and digital formats, coding and transmission channels errors. Hence, methods are required to evaluate the user's QoE (Quality of experience) of multimedia streaming application. The two types of assessment method are: subjective and objective quality assessments methods. In subjective quality assessment method human viewers are used which watch a video clip to determine the quality scores. Measurements are based on how the user perceives the video quality. Quality scores are then averaged to produce the Mean Opinion Score (MOS). [48]. The MOS is extracted from each individual by giving a sample video or audio material. They either listen or watch the audio or video or both. The individual who serves as a sample statistic is given a chart represented by the table below.

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very Annoying

Table 3.1 MOS scores [36]

"There are different types of subjective quality assessment which are: Single Stimulus Continuous Quality Evaluation, Absolute Category Rating, Double Stimulus Continuous Quality Evaluation, Double Stimulus Impairment Scale, Double Stimulus Comparison Scale and Subjective Assessment Methodology for Video Quality (which is currently being standardized). SAMVIQ has been designed specifically for multimedia content, taking into account codec types, image format, bitrates, packet loss, etc. However, Objective assessment method used algorithm to evaluate the quality of the video and audio data. There are different types of objective quality assessment methods such as peak signal to noise ratio, Mean Square Error , structural similarity and image quality , and video quality metric are the most widely used objective quality metrics because it is simple to implement and has clear substantial definition".

3.1.2. Subjective assessment requirement description

To successfully carry out subjective quality assessment, the following requirements and Conditions given below have to be considered. The requirements are:

➢ Viewing conditions

There are different viewing conditions in case of referring to a laboratory environment than the ones used in home environment. Assuring the best conditions is a basic aim in order to allow a significant evaluation to be conducted derived from the assessment. [36] The test should be conducted under these viewing circumstances.

i.   Viewing distance should be 1.8
ii.   Peak luminance of the screen should be 100-200cd
iii. The ratio of dormant screen to peak luminance should be less than 0.05
iv.   Chromaticity background should be D_{65}
v.   Background room illumination should be less than 20lux

➢ Condition of the room

The room must be comfortable and well conditioned with light and there should be no reflection, because the human viewers (subject) will view the videos for considerable amount of time. All these conditions must be put in place to avoid getting substandard data.

Observers

The minimum number of observers that could be is 15 but non-expert and if you really want to obtain good result, the number of observers should be 40. However, you should use people that are not really professionals i.e those that could not easily identify impairment in the video.

Materials / test sequences

There must be a selection of sufficient test material, including different type of contents emitted by a conventional TV channel (sports, movies, news, documentary, etc.), and with different settings of spatial (more or less level of detail and high frequencies) and temporal (faster or slower contents) complexity [36].

Each sequence must have a duration of between 10 and 20 seconds due to human memory and perception, to assure a correct viewing, neither too short than the viewers do not have time to observe the image in detail, nor too long to avoid causing fatigue in the observer.

Presentation of result

In the presentation of result, all the information that will be used must be valid and must be presented in details: The data given must include:

- Test configuration details;
- List of test materials;
- Monitor display and picture type.
- Different type of assessors used.
- Reference systems used.
- The experiment grand mean score.

Sound

Audio quality assessment should not be combined with video assessment, it would be better if the tests are carried out separately, In general, it is recommended that sound or audio should not be used when carrying out video quality test to avoid distraction.

3.2. Methods of subjective assessments

There are different methods of subjective assessment, but the most important ones will be discussed in this thesis:

➕ Double Stimulus Continuous Quality Scale (DSCQS)

The main aim of this method is to appraise the quality of systems in relation to a mention. Human audiences watch pairs of video sequences (the mention sequence and the impaired sequence) in an irregular fashion. To context effects it is popularly acknowledged as a perfect run method with modest understanding, [36] as audiences are shown videos two times. They are asked to give opinion on the quality of each sequence in the pair after the second demonstrating. The quality of stereoscopic mental image coding is also measured.

Fig 3.1 Scheme of a DSCQS system [36]

Only a single quality rack up for a given video sequence is given in this technique, Inquiries have been raised as to the pertinence of these testing methods, where a usual video sequence may be 10 seconds long,for appraising the public presentation of objective real time video quality monitoring systems.

➕ Double Stimulus Impairment Scale (DSIS) (the EBU method)

This is the method used by the European broadcasting union, the mention is always bring out before the run sequence and its not retried. Beholders are asked to give opinion about the amount of damage in the run sequence applying a five degree graduated table. [29] The classification in the graduated table are imperceptible, perceptible, but not disturbing, slightly disturbing, disturbing, and very disturbing. For appraising visible distortions this method is paramount.

Fig 3.2 Scheme of a DSIS system [36]

The run sequences are shown only once and then mention. One to the other subjects charge per unit the amount of damage in the run sequence is compared.

➕ Single stimulus continuous quality evaluation (SSCQE)

In [29] this process, people are shown a video for about 20 30 minutes and no mention is given and the video content is refined applying the conditions given. The perceiver employs a grader to constantly rate the quality, during the presentation as it changes. The graduated table (ruler) moves from worst to best..

Fig 3.3 Scheme of a SSCQE system [36]

.

➕ Simultaneous double stimulus for continuous evaluation (SDSCE) method

Two projection screens are necessary for this method of evaluation, which are parallel placed in front of the user [36],While the right one plays the distorted sequence that TV audience must rack up, the go forth ed projection screen plays the mention sequence.
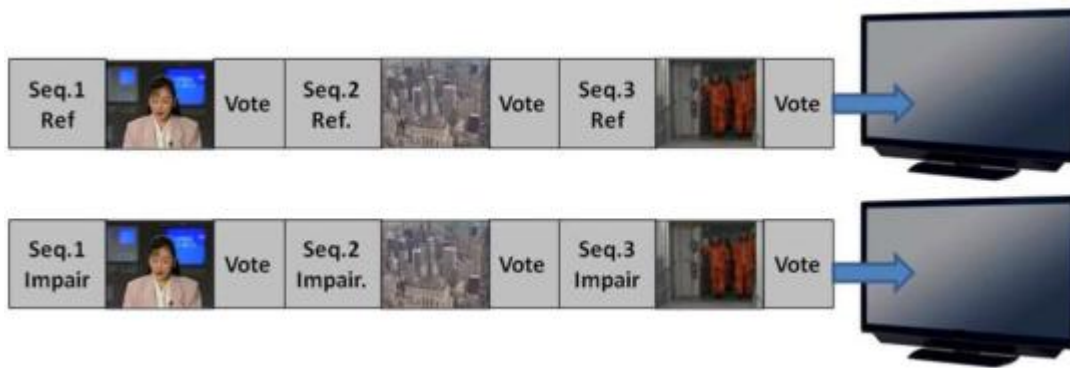
Fig 3.4 Scheme of a SDSCE system [36]

The main purpose of this method is to measure the fidelity between two video sequences. It is also used to compare different error resilience tools. Each video pair is shown once or twice. The duration of the test session is shorter, and allows evaluating a higher amount of quality parameters.

These methods discussed above are the important subjective quality assessment method. Despite this, there are still some methods that are recommended by [29]. The methods are discussed below:

🞣 Absolute category rating (ACR)

In [29] under run viewing audience only see the video, the rating is presented for its general quality applying a distinct five degree graduated table from worst to best. The evidence that the reference is not shown with every test snip does absolute category rating a very effective technique in contrast to DSIS, which take about 2 or 4 contemporary times as extension, respectively. Several TV audiences run concurrently from a tape if a constant voting time is used. Then the casting time ought to be below or equal to 10 s. The presentation time may be decrease or increased due to the content of the run stuff. The grading level quality scale is shown below:

5 Best

4 Good

3 Fair

2 Poor

1 Worst

- Absolute category rating with hidden reference (ACR-HR)

This technique is a class judgment where the run sequences are shown one at a time and are judged individually on a group ordered series. The present run method must involve a mention version of each run sequence bring out as any other run stimulant. This is known as a encrypted reference condition [29].Its corresponding (hidden) mention and during the data analysis, a derivative quality rack up (difference of mean opinion scores) will be measured between each run series. This technique is referred to as "Hidden reference". After each demonstration, the subjects are asked to appraise the quality of the sequence shown, the method shows that. Extended according to the content of the test material or the demonstration time may be decrease. The grading level quality scale is shown below:

5 Best

4 Satisfactory

3 Fair

2 Poor

1 Worst

Differential viewer scores (DV) are calculated on a per subject per processed video sequence (PVS) basis. The appropriate hidden reference (REF) is used to calculate DV using the following formula:

DV (PVS) = V(PVS) – V(REF) + 5,……….…………………………………..(3.1a)

Where

V is the viewer's ACR score. A dv of 1 bespeaks bad quality and in applying this formula, a dv of 5 bespeaks excellent quality. Any DV values higher than 5(i.e where the refined series is graded better quality than its related hidden reference series) will generally be viewed legitimate. Consequently it may be applied to ACR-HR to avoid the perceivers from enticing the mean opinion scores

Crushed DV = (7*DV)/ (2+DV) when DV > 5…………………………. (3.1b)

A nine degree ACR graduated table may be used, if higher discriminatory power is required. Best quality on the above five degree graduated table or the ACR 60 minutes method should only be used with mention video that an expert in the field considers to be of satisfactory [28].The ACR 60 minutes method may not be suitable for analyzing unusual damages that come about in the first and last 60 msec of the video series.

🔸 Degradation category rating (DCR)

In pairs the degradation category rating connotes that the run sequences are presented. The first stimulant presented in each pair is always the origin mention; while the second stimulant is the same origin presented through one of the systems under run (This method is also called the double stimulus impairment scale method.)when reduced picture formats are used for example QCIF, SIF. The run sequence concurrently on the same monitor and it could be helpful to display the mention. In annex c guidelines on this demonstration procedure are discoursed. Then if stable voting time is applied, the voting time should be below 10 seconds. Increased due to the content of the run material or the demonstration time may be decreased. However, the subjects are asked to charge per unit the damage of the second stimulant in relation to the mention. The grading level quality scale is shown below:

5 Imperceptible

4 Perceptible but not disturbing

3 Slightly disturbing

2 Disturbing

1 Very disturbing

The necessary number of replications is obtained for the DCR method by ingeminating the same run conditions at different points of time in the run [28].

🔸 Pair comparison methods (PC)

For this method, test clips from the same scene but different conditions are paired in all possible combinations, [50] and perceivers make a preference judgment for each pair. This allows very fine quality discrimination between clips.

3.3. Objective quality assessment method

Objective video quality metrics can be engaged to:

- To optimize the rules and factors settings for a video processing systems.
- Comparison of video processing systems and rules
- Observe video quality.

Objective quality metrics can be categorized into three different groups according to original reference video.

1) Full Reference metrics (FR); in this method both the original and the distorted videos that are used are made available.
2) Reduced Reference Metrics; in this method, description of the original video as well as some parameters are made available. The distorted videos are also given.
3) No Reference Metrics: Only distorted video is made available.

3.3.1. Types of Objective Quality Metrics

There are many types of objective quality metrics based on full reference, reduced reference and no reference. Under this categories there are still further classifications based on statistics and pixel difference, based on structural similarity, based on artifact detection, and based on vision models. In this chapter the important types of the quality metrics will be discussed which are

.

❖ MSE (Mean Square Error)

This is one the most common disparity metrics in image and video processing. [36] The MSE is the mean of the squared differences involving the luminance level values of pixels between two images (X and Y), normally the original frame used as a reference and an impaired image, obtained by processing of the first image.

$$MSE = \frac{1}{M.N} \sum_{i=0,j=0}^{M-1,N-1} (X_i - Y_J)^2 \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (3.1)$$

M and N are the horizontal and vertical sizes of each frame of the sequence.

❖ SNR (Single-To-Noise Ratio)

The relation including the new mental image also based on pixel by pixel comparison, this metric calculates and degraded image, [36] in order to evaluate the degradation on image.

$$PSNR = 10log_{10}\left[\frac{\sum_{I=0,j=0}^{M-1,N-1}|X_i|^2}{\sum_{i=0,j=0}^{M-1,N-1}(X_i - Y_J)^2}\right] \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.(3.2)$$

SNR was widely substituted by its evolution PSNR, because it offers a higher efficiency than SNR, and a global extension more easy to compare in studies with different signals.

❖ PSNR (Peak Single-To-Noise Ratio)

PSNR is a derivative of the mean squared error (MSE) relative to the maximum likely value of the luminance (for a typical 8-bit value this is -1 = 255) as follows:

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[f(i,j) - F(i,j)]^2}{M.N} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.(3.3\ a)$$

$$PSNR = 20.log_{10}\left[\frac{255}{\sqrt{MSE}}\right]\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots..\dots\dots(3.3b)$$

Where

F (i,j) is the new signal at pixel (i, j), F(i, j) is the regenerated signal, and M x N is the picture size. [59] Spaning 30 to 40 for average to high quality video, the consequence is a single number in dB. In malice of several objective video quality frameworks that have been created in the past times decades, To be the .most accepted assessment of the quality difference among pictures PSNR maintain its standard.

❖ Blockiness or tiling

Subjective blocking consequence in video sequence media lab measure outs and the metric defined by MSU graphics , Gradients and based on energy computation. But in smooth areas

these edges are conspicuous; in direct contrast areas of the frame blocking is not appreciable. Other metrics are based on the structure of the picture elements image. [36]The model included. Research, first extracts edge pixels and computes horizontal (H (t, i, j)) and vertical (V(t, i, j)) gradient component of the edge pixels. The gradient is calculated employing the sobel operators. Vertically and horizontally gradient images, the magnitude (R) and angle (Φ) are

obtained$R(t, i, j) = \sqrt{H(t, i, j)^2 + V(t, i, j)^2}$ … … … … … … … … … … … … … … … … … .. (3.4a)

$$\theta(t, i, j) = \tan^{-1}\left[\frac{V(t, i, j)}{H(t, i, j)}\right], \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (3.4b)$$

Gradient parallel to the icon frame are considered as belonging to blocking region the pixels with, analyzing the angles of the gradient. if they have a determined magnitude, In comparison to the original image, errors are avoided due to real edge pixels.[2].

❖ Blurring

Analyses of edges and their spread and on the analysis of energy in high frequencies blurring metrics are based. As it suggests the step down of edge energy between the original and as it proposes the reduction of edge energy between the original. In [50]. Research, utilizes the gradient calculated in every pixel of the image to detect the blur artifact by analyzing the diminution of this

$$BL(x) = \frac{1}{\sqrt{FxC}}\left[\sqrt{\sum_{i=0}^{f-1}\sum_{j=0}^{C-1} SG^2_{ref(i,j)}} - \sqrt{\sum_{i=0}^{F-1}\sum_{j=0}^{C-1} SG^2_{impaired}}\right] \dots \dots \dots \dots \dots \dots \dots (3.5)$$

Magnitude between the original and the impaired image. SI is the root mean square of the spatial gradient (SG), so blurring is computed as above.

➢ Structural Similarity and Image Quality (SSIM)

The most representative metric reliant upon Structural Similarity is SSIM. The metric projected by Wang et al. is dependent on the combination of three properties of the image luminance, contrast and structure; comparing the original and the distorted image, three

conditions must be met: symmetry, roundedness and being unique maximum [36]. To calculate the similarity between a reference image and the related reference, the SSIM algorithm quantifies the luminance l(x , y), dissimilarity c(x ,y), and configuration s(x ,y) of the test image y and the equivalent mention image x, using the formulas below:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C'} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.6a)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2'} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.6b)$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3'} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.6c)$$

Where $C_1, C_2,$ and C_3 are small constants given by $C_1 = (K_1.L)^2$, $C_2 = (K_2,.L)^2$, and $C_3 = C_2/2$.

L is the dynamic range of the pixel values (for 8 bits/pixel gray scale images, L = 255),

$K_1 \ll 1$, and $K_2 \ll 1$.

The general formula of the SSIM metric is given by

SSIM(x, y) = [l(x, y)]$^\alpha$ * [c(x, y)]$^\beta$ *[s(x, y)]$^\Upsilon$ ……………………………………………………..(3.6d)

Where α, β and Υ are parameters that describe the relative significance of the luminance, contrast, and structure components. If α= β= Υ =1, the above equation is summarized to:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.6e)$$
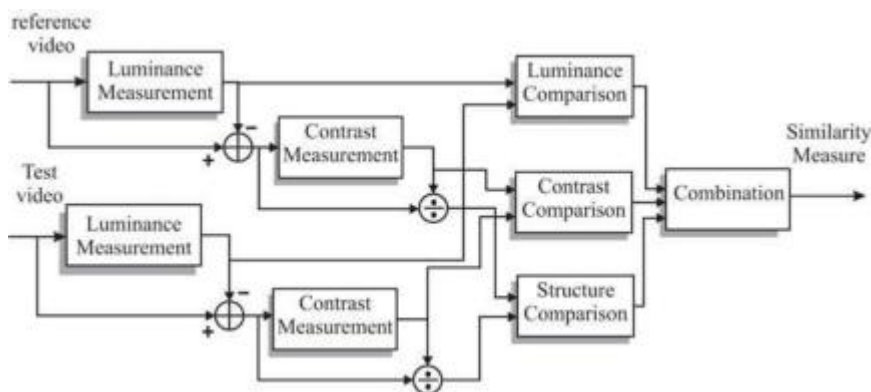

Fig 3.5 Block diagram of the SSIM algorithm [2]

The SSIM has a span of values spanning between 0 and 1, with 1 being the best value possible and 0 as worst. A block diagram of this algorithm is shown in Fig 3.5. A research of the performance of SSIM has revealed that this simple metric illustrate good results [42].

➢ Video quality metric (VQM)

This is a metric suggested by NTIA [44]. This rule has been newly adopted by ANSI as a model for objective video quality assessment. In VQEG Phase II (VQEG, 2003), VQM offered a very good relationship with subjective scores. VQM offered one of the best efficiency amongst other subjective and objective assessment methods. The VQM algorithm can detect visualizing effects of numerous video degradations, such as blurring, unnatural motion, global noise, block distortion, and color distortion.

Chapter 4

Results Analysis and Discussion

In order to ascertain the theories mentioned so far, attempt was made to find out from earlier studies how H .264 has fared when passed through an IP Network. This study is used as a case to understand the significance of the transmission of H.264 through an IP network.

In an earlier discussion, mention was made of ATM. Based on the capacity provided by ATM in the transmission of data to the network layer; one can say that high capacity broadband TCP/IP layer 1 networks like ATM would be needed to convey real-time H.264 video/audio in an efficient manner. However at the Internet layer where IP routes the H.264 Bit stream, it was important to carry out a desktop research to find out if what quality of video can be achieved.

- H.264 over IP.

The case of H.264 transmitted via IP was documented a recent study published in 2010 by K Seshadrinathan et al, reference H.264 and MPEG2 were compressed and evaluated. Then H.264 was transmitted via both IP and wireless Networks to evaluate the video quality [7]. They employed both the subjective and the objective methods to evaluate the Video quality.

Unfortunately, the raw data from the subjective and objective tests were not published, rather the correlation between the subjective results and the objective results were published.

However, focus will be placed more on the correlation of the objective and subjective data gathered. This computation produces the Objective Video Quality Metric. MPEG 2 isn't important in this case because it is used as a reference in the video assessment. Secondly MPEG2 can't be compared to H.264. H.264 can only be compared to another AVC in which category MPEG4 falls under. This will be done in this chapter.

The video sample size of 150 were used for both the subjective and objective testing

- Objective Video Quality evaluation

The Video Quality Assessment Algorithm was used.

"The VQA parameters tested; were Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), Multi-scale SSIM (MS-SSIM), Speed SSIM, Visual Signal to Noise Ratio (VSNR), Video Quality Metric (VQM), V-VIF and Motion-based Video Integrity Evaluation (MOVIE)"[8].

The Spearman's Rank Order Correlation Coefficient (SROCC) was used to measurate the robust nature of the VQA algorithm, hence aiding in comparing the performance of the VQA algorithms in relation to the subjective test. The Pearson's Linear Correlation Coefficient (LCC) was used to ascertain the accuracy of the results to the subjective tests that is carried out. The result focus here will be on the SROCC data.

The main take for this thesis from these results is how a compressed H.264 video quality fares when transmitted via an IP network. The Objective Evaluation data isn't used in the observation for this thesis because it is better than the subjective method; rather it is used because that was the data that could be found as at the time the thesis was ready.

Table 4.1: The full result of the SROCC was as follows [7]:

Prediction Model	Wireless	IP	H.264	MPEG-2
PSNR	0.4334	0.3206	0.4296	0.3588
SSIM	0.5233	0.4550	0.6514	0.5545
MS-SSIM	0.7285	0.6534	0.7051	0.6617
Speed SSIM	0.5630	0.4727	0.7086	0.6185
VSNR	0.7019	0.6894	0.6460	0.5915
VQM	0.7214	0.6383	0.6520	0.7810
V-VIF	0.5507	0.4736	0.6807	0.6116
Spatial MOVIE	0.7927	0.7046	0.7066	0.6911
Temporal MOVIE	0.8114	0.7192	0.7797	0.8170
MOVIE	0.8109	0.7157	0.7664	0.7733

What is not clear from the table and their methodology is what wireless technology was used. This is because the results for transmitting H.264 through a WLAN may be different from

values obtained if it passed through WMAN. The second thing that is not clear is; what TCP/IP suite network layer technology enabled the IP isn't also very clear. Hence one can't generalize that the data here would be the same for all wireless and IP networks. Nevertheless, what makes the IP test significant here is that one can have a general idea of the distortions created if a H.264 video is passed through an IP network? Of course these distortions will vary depending on the type of Network Layer supporting the IP. This is why this thesis recommends ATM.

From the method carried out by Seshadrinathan et al (2010), MPEG 2 is used as a control. They are neither passed transmitted via the reference IP or wireless networks [8]. Hence is a bit difficult to make an argument for an MPEG standard on how they would fare if transmitted via a wireless or IP network.

In streamlining the result as seen in the table below, one can see the outcomes when a H. 264

Prediction Model	IP	H.264
PSNR	0.3206	0.4296
SSIM	0.4550	0.6514
MS-SSIM	0.6534	0.7051
Speed SSIM	0.4727	0.7086
VSNR	0.6894	0.6460
VQM	0.6383	0.6520
V-VIF	0.4736	0.6807
Spatial MOVIE	0.7046	0.7066
Temporal MOVIE	**0.7192**	**0.7797**
MOVIE	0.7157	0.7664

Table 4.2 Performance of H.264 algorithm [7]

It is clear that at a glance, when one looks at the Video Quality Metric in the table above; a compressed H.264 not transmitted via any transport medium is of a better quality than the compressed H.264 transmitted via an IP network. This is because the correlation

between the Objective scares and the subjective scores are strong. But the correlation between the objective evaluation and the subjective evaluation vary when the H.264 packets are transmitted via IP. One would say that subjectively, the data gathered is based on human judgment. Human judgment on color, sharpness, contrasts and distortions vary. Hence what is good to one may be fair to the other and even excellent to the next person. But objectively, the signals are processed via systems and the results gathered and computed.

But as seen above the various objective measurements video quality parameters parameters have various degrees of correlation to the subjective measurement. This makes the exclusive use of one Objective evaluation insufficient to determine video quality.

However, when one looks at the data above, it is clear that the variance in the evaluation of quality and the degree of variance between the subjective and the objective is as a result of packet loss when the H.264 is transmitted via IP. Else the correlation would be strong.

Let's take a closer look at the evaluation of each VQA parameter. Below are the parameters used in evaluating the results of the SROCC

For values of 0.9 to 1, the correlation is very strong.

For values of 0.7 and 0.89, correlation is strong.

For values of 0.5 and 0.69, correlation is moderate.

For values of 0.3 and 0.4.9, correlation is moderate to low.

For values of 0.16 and 0.29, correlation is weak to low.

For values below .16, correlation is too low to be meaningful.

PSNR: The PSNR is a function of the Mean Square Error (MSE). In the test reported here the MSE between the reference (Control) H.264 and the H.246 transmitted via the IP is evaluated. Based on the SROCC evaluation parameters mentioned above, the correlation of the picture quality for the PSNR for both the H.264 transmitted via IP and the reference is between moderate and low compared to whatever data was gathered subjectively for both tests.

SSIM: But for the Structural Similarity (SSIM), the objective evaluation was much better than that of PSNR. The SROCC for reference H.264 was moderate; hence it wasn't much

different when compared to the subjective test. However, the video sample passed through IP was recorded between moderate and weak.

MS SSIM: The Multi-Scale SIM (MS SSIM) data for both the reference video and the H.264 transmitted via IP were close to data obtained from the subjective evaluation. This makes this objective method of evaluating VQA much better than other Structural Similarity evaluation methods.

Speed SSIM: the Speed SIM, the reference video had a strong correlation to the subjective test but the H.264 transmitted via IP had a moderate to low correlation.

VSNR: The Video Signal to Noise Ratio (VSNR) data were both moderate to that obtained from the subjective test.

VQM: The Video Quality Metric (VQM) also had moderate correlation to the subjective test.

V-VIF: For the Video Visual Information Fidelity (V-VIF), The H.264 reference had a moderate correlation to the subjective evaluation of the same video, but the video transmitted via IP had a weak correlation to that of the subjective test.

The test for the Spatial MOVIE for both samples had a strong correlation with the result of the subjective evaluation. Same could be said for the Temporal MOVIE and the MOVIE.

Hence the Objective evaluation can be used to evaluate MOVIE, Temporal MOVIE, Spatial MOVIE as well as the MS SIM. The data for these parameters were close to that of the subjective results.

In looking at these results, one can see objectively that the probability of having a moderate correlation to the subjective result is higher for the reference H.264 video except for the video that is transmitted via IP. The debate on whether the objective evaluation or subjective evaluation of video quality rises at this point. This is still being researched. However, the correlation from both sides points to packet loss.

- H.246 over ATM

This transmission mode of multimedia including H.264 over ATM network has been discussed in Chapter 2. In chapter 1 mention was made of delays in the transmission of

packets. These delays arise as a result of congestion in the network, the data rates and the frequency (if wireless) on which the network is operating.

However, H.264 as mentioned earlier can be transmitted at lower bit rates than previous standards like MPEG-2, H.263 and MPEG-4. Hence it can use for multimedia development for a network with any data rates.

The advantages of ATM are its greater bandwidth, high QoS, low delay and its convergence abilities. Convergence in this case refers to the merging of voice, video and data into a single network.

Looking at the positive attributes of H.264 and its compatibility with ATM as expressed in chapter 2, one can say that H. 264 would be transmitted with ease using an ATM network that isn't congested beyond its capacity.

- IP over ATM

In the first result, mention was made of packet loss when a compressed H.264 was transmitted via an IP network. As mentioned in chapter one, the Network at the Network layer of the TCP/IP suite determines the throughput delivered to the Internet layer. In the previous section (H.264 over ATM), mention was made of H.264's ability to be transmitted with low bit rates and ATM being able to transmit such a codec with low delay and high QoS. If an IP network is the routing network for an ATM backbone network, then it is expected that the IP packets will transmitted at a more rapid rate than if it were transmitted with Ethernet.

- Why H.264

Much has been said about H.264 and the rational for transmitting it over an IP network. It has also been argued that for the video quality to be improved, then an ATM network should be the backbone of the network. But the lingering question is on why H. 264, Why not other Codecs like MPEG 4, the popular MPEG 2 and H.263.

H.264	MPEG 4	MPEG 2	H. 263
Advance Video Coding (AVC)	AVC	Normal Video coding	Normal video coding
Ability to encode voice audio and speech	Ability to encode video, audio and speech	The audio compression stream limited to 2 channels	Ability to encode voice audio and speech
Optimized for low bit rates (1.5 Mbit/s)	Optimized for low bit rate but later expanded to accommodate multimedia coding	Not optimized for Low bit rates (3.5 Mbits/s)	Optimized for Low bit rates
Used for Videoconferencing, Blue-ray Disks, Internet Streaming (youtubeetc)	Enables animations, graphics, digital TV, WWW and extensions.	Used for Digital Video broadcast and DVDs	Used for videoconferencing
Less complex design	It is still a developing standard	It is a complex set of standards	Less complex design
Adaptable for multimedia transmission over IP networks	Multiplexes and synchronizes data in such a way that they can be transported via networks	Not much in use due to lack of interest by industry.	It has now given way to H.264.

Table 4.3: Comparison of Compression algorithms [34]

The table above doesn't encompass all the characteristics of the various compression standards. However as made evident in the table, the closest competitor to H .264 are the

MPEG 4 set of standards and the H.263, which has now given way to H. 264. One may say that the advantage of MPEG 4 is that is contains the attributes of MPEG 2 and MPEG 1 But the only disadvantage would be that as the byte of the multimedia increases, so will the need for capacity to transport the MPEG 4. But H. 264 is designed to transport multimedia at lower bit rates. H.264 is an improvement of the H. 26x set of standards. The EU telecom standards body ETSI in conjunction with 3GPP used the H.263 video for the 3rd generation mobile telephony standard, 3GP container format. The 3GP is the multimedia container format defined by 3GPP for UMTS networks. The container format is a Metafile format whose requirement specifications explain how different data elements and the metadata coexist in a computer file.

With the flexibility of H.264 over MPEG 4 set of standards, which give H.246 the edge. However in practice, industry experts use either MPEG 4 set of standards or H.264.

<u>Conclusion and Recommendation</u>

The thesis was designed to research into how high quality video can be transmitted over IP networks and ATM networks. From the case study conducted via desktop research, the conclusion is that high quality video can be transmitted without many distortions through an IP network, if the video is transmitted through a network with high capacity hence leading to QoS and less delay.

Theoretical considerations were taken as regard high quality video codec H. 264. In the thesis, there has been comparison between H.26X of which H.264 belongs and MPEG set of standards. Theoretical considerations also touched on ATM and it prospects to the delivery of a high quality video using the IP routing network. In the process of this research a set of relations was developed to depict a process in which a high quality video can be delivered on an IP network.
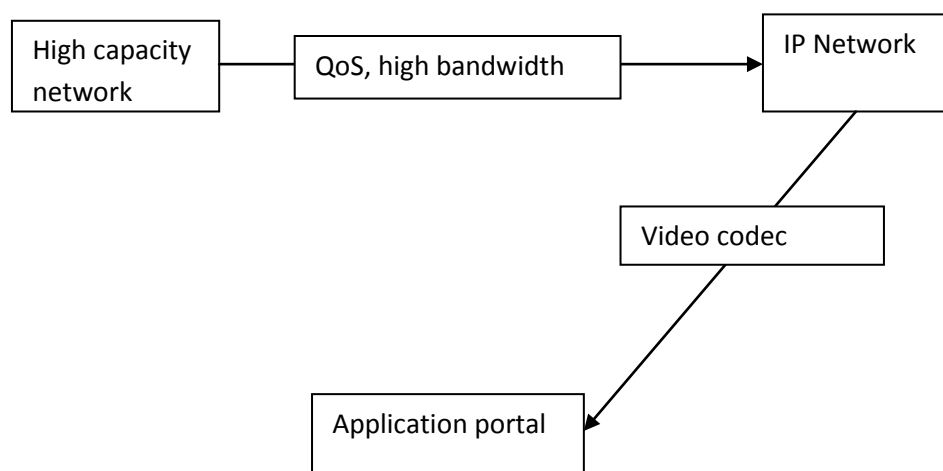

Fig 5.1: Depicted network diagram to transmit high quality multimedia data

The high capacity network here is ATM, and the application portal is the device used by the end user to access the video quality.

In chapter 4 attempts was made to investigate how H.264 video would fare when transmitted via an IP network. Although the level of distortion couldn't be determined, it was discovered that distortion occur. Based on the theory posited in chapter 2, attempt was made to reason on

how a H.264 video would fare over an IP network. Reasoning based on theory was made on how H.264 would fare over an ATM network. It has been difficult to actually tell the nature of the video but from theory, one can say that the distortions found in the case experiment would be minimized if the H.264 were transmitted via an IP network that was routing in an ATM network.

However, one major finding was that although H.264 is an AVC. MPEG standards can also be transmitted if the conditions in the figure above hold. IP on ATM can transmit any codec with an acceptable QoS. The only limitation would be what the Codec can be used for.

Recommendations

1. To transmit a High Quality audio and video data over a network, I suggest a high capacity back bone with low delay and QoS parameter. This will lead to the minimization of packet loss.

2. To transmit a voice or video data over any IP network, I also propose a Codec that provide high quality like H.264 or MPEG 4, MPEG-7. H.264 or MPEG-4 is more preferable as it can be transmitted with low data rates. This can't be achieved with MPEG 2.

3. I recommend ATM as a network technology for video/ audio transmission because of its ability to transfer converged data simultaneously and still provide QoS. Hence congestion, unless the capacity of data is more than that designed for the ATM network, is not likely.

4. I also propound ATM network in combination with optical fiber communication link can transmit uncompressed audio and video without degradation of video quality. This is because fiber optics cable could carry multimedia data traffic up to 20 Gbps. So I believe combination of these two things will solve the problem of transmitting high quality audio and video though a lot of research has been going on to devise different means to convey high quality multimedia over the network.

5. To achieve high quality audiovisual presentation at the end user, I suggest MPEG-7 compression format. This is because MPEG-7 allows description of audio visual content in multimedia environment, add functionality to audiovisual data and MPEG standards use this compression format to enhance their facilities of content description.

➕ Delimitations of the research:

This greatest delimitation of the research was the inability of the researcher to test practically. The analysis made is theoretical but also logical and can't be discarded.

Bibliography

1. Abhijit S. P, Ercan, S.: ATM Technology for Broadband Telecommunications Networks, Jan 1998, CRC Press LLC, Pp 177-182

2. Alan, C. B, Hamid R. S and Zhou, W.: Objective video quality assessment, 2003, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 13, NO. 4, Pp 2-5

3. Axis, Comm.: H.264 video compression standard. New possibilities within video surveillance, 2008, AB communications, Pp 3-7

4. Black D, E. Davies, M. Carlson, S. Blake, W. Weiss and Z. Wang.: An Architecture for Differentiated Services, December 1998, RFC 2475

5. Bobby, V. Raj, J. Sonia, F. and Wu-chi, F.: A Survey of Application Layer Techniques for Adaptive Streaming of Multimedia, 1999-08-16,

6. Bolot J.C and T. Turletti.: A rate control mechanism for packet video in the internet, November 1994, IEEE INFOCOM.

7. Bovik A. C,K. Seshadrinathan, L. K. Cormack, and R. Soundararajan.: A Subjective Study to Evaluate Video Quality Assessment Algorithms, Jan. 2010,SPIE Proceedings Human Vision and Electronic Imaging

8. Bovik A. C,K. Seshadrinathan, L. K. Cormack, and R. Soundararajan.: Study of Subjective and Objective Quality Assessment of Video, June 2010,IEEE Transactions on Image Processing, vol.19, no.6, Pp 1427-1441.

9. Broder A. Z, J. Almeida, L. Fan, and P. Cao.: Summary cache :a scalable wide-area web cache sharing protocol, June 2000 ,IEEE/ACM Trans. on Networking, vol. 8, no. 3, pp. 281-293,

10. Bruce S. D, and Larry L. P.: Computer Networks a Systems Approach, 2003, Morgan Kaufmann Publishers, Pp 162-517

11. Chad, E. Didier, J. Joan, L. and William, B.: Mpeg video compression standard, 2002, Kluwer Academic Publishers, Pp 26-29

12. Chou P. and Z. Miao.: Rate-distortion optimized streaming of packetized media, Feb. 2001, IEEE Trans. on Multimedia, vol. 3, no. 1, pp. 108–122

13. Cliff, W.: A Practical Guide to Video and Audio Compression From Sprockets and Rasters to Macroblocks, 2005, Elsevier Inc, Pp 211-213

14. Collin, P: Audio and Video for the internet, June 12, 2003, Addison Wesley, Pp 25-231

15. Collins, P, O. Hodson , and V. Hardman.: A Survey of Packet Loss Recovery Techniques for Streaming Audio,1998, IEEE Network, vol. 12, no. 5, pp. 40-48

16. Cooper C. A. and K. Park.: Towards a Broadband Congestion Control Strategy, May 1990, IEEE Network Magazine, Pp 18-23

17. Crowcroft, J, L. Rizzo and L. Vicisano.: TCP-like congestion control for layered multicast data transfer, 1998.Proc. IEEE INFOCOM, Pp 98

18. Daniel, M.: Video dial tone technology: digital video over ADSL, HFC, FTTC, and ATM, 7 Dec 2007, McGraw-Hill, Pp 123-234

19. Dapeng Wu,Jon M. Peha, Wenwu Zhu, Ya-Qin Zhang and Yiwei Thomas Hou: Streaming Video over the Internet: Approaches and Directions,2-1-2001, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 11, NO. 3, Pp 286-291

20. David, A.: The Technology of Video and Audio Streaming, 2005, Focal Press 200 Wheeler Road, Burlington, MA 01803, USA, Pp 83-91

21. Djordje, M: Video Compression, 2005, University of Edinburgh, Pp 3-7

22. Eleftheriadis A. and D. Anastassiou.: Meeting arbitrary QoS constraints using dynamic rate shaping of coded digital video, Apr. 1995, in Proc. 5th Int. Workshop Network and Operating System Support for Digital Audio and Video (NOSSDAV'95), pp. 95–106.

23. Erik, Hellerud .: Transmission of High Quality Audio over IP Networks, April 2009, Centre for Quantifiable Quality of Service in Communication Systems, Pp 17-23

24. Fast Forward networks' broadcast overlay architecture," URL: http://www. net.com/ pdfs /boa-whitepaper. pdf.

25. Ghaida, Al Suhail and Sarah A. Subber.: ROBUST NETWORK FEC-EMBEDDED VIDEO TRANSMISSION OVER ERROR-PRONE WIRELESS CHANNELS, June 2012, International Journal of Wireless & Mobile Networks (IJWMN) Vol. 4, No. 3, Pp 39-40

26. Hang, Zhao: MPEG2 BIT RATE CONTROL FOR NATIVE ATM, 2001, San Diego State University, Pp 48-54

27. Harry, P.: An introduction to ATM networks, 2000, Harry Perros, Pp 63-147

28. ITU-R BT. 500-11: Methodology for the subjective assessment of the quality of television pictures, Pp 3-5

29. ITU-T P.910: Subjective video quality assessment methods for multimedia applications, 04/2008, Pp 3-23

30. Jacobsen V,: Congestion Avoidance and Control, August 1988, ACM SIGCOMM

31. JDSU: Mpeg-2 reference guide to digital video technology, testing, and monitoring, 2008, JDS Uniphase Corporation, Pp 9-13

32. Jean, W: Communication networks a first course, 1998, WCB/McGraw-Hill, Pp 127-651.

33. Jeon B. and W. Choi.: Dynamic UVLC codeword remapping with fixed re-association table for H.264, April2001, in Picture Coding Symposium (PCS) (Seoul, Korea), Pp 167-170

34. John G. Apostolopoulos, Wai- tian Tan, Susie J. Wee Mobile.: Video Streaming: Concepts, Algorithms, and Systems, 2002, Hewlett Packard company, Pp 11-29

35. Jonathan H, Chao .: Architecture Design for Regulating and Scheduling User's Traffic in ATM Networks, August 1992,in Proc. ACM SIGCOMM'92, Pp 77-87

36. Juan, Pedro. Video Quality Assessment, March 23, 2012, Universidad Politécnica de Madrid Spain InTech , Pp 134-146

37. Kot A, G. Bi, L. Cheng, and Y. Zeng,: H.26L test model long term number 7 (TML-7),April 2001,ITU-T/SG 16/VCEG(formerly Q.15 now Q.6), ", Doc., Pp 81

38. Kozamernik, F.: Media streaming over the internet an overview of delivery technologies, October 2002, EBU technical department, Pp 9-13

39. Marko, V.: ATM networking basics, 1999, CS596, Pp 4-41

40. Marpe D., G. Biattermann, and T. Wiegand,: Improved codec, Dec 2001, in ITU-T Q.6/SG16, VCEG-O18,Pp 4–6

41. Monterio J, L. Fratta and M. Gerla.: Leaky Bucket Analysis for ATM Networks, September 1990, in Proc. SBT/IEEE International Telecommunications Symposium Rio de Janerio Brazil , Pp 482-502

42. Mylène C. Q. F.: Video Quality Metrics, February 1, 2010, Floriano De Rango, ISBN 978-953-7619-70-1,Pp 337-354

43. Perkins C, O. Hodson, and V. Hardman, "A Survey of Packet Loss Recovery Techniques for Streaming Media," IEEE Network Magazine, September/October 1998. © 1998 IEEE. (Diagram interleaving and FEC)

44. Pinson H, and Stephen, W.: A New Standardized Method for Objectively Measuring Video Quality, 2004, Pp 1-4

45. Postel J.: Transmission Control Protocol, Sept. 1981, RFC 793

46. Ramsey J. L..: Realization of optimum inter-leavers, May 1970, IEEE Trans. Info. Theory, vol. IT-16, pp. 338–45

47. Rejis, J.: Broadband Telecommunications Handbook second edition, 2002, Mc-Graw Hill Telecomms, Pp 159-165

48. Samuel, P.: Next Generation Mobile Networks and Ubiquitous Computing, August 31, 2010, IGI Global, Pp 64-69

49. Song. Pu.: MPEG-2 Transport over ATM Networks with Best Effort Service,1999, Bell & Howells lnformation and Learning, Pp 33-36

50. Stefan, W.: Digital video quality vision models and metrics, 2005, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, Pp 35-70

51. Streit J, L. Hanzo and P.J. Cherriman.: Video Compression and Communications: From Basics to H.261, H.263, H.264, MPEG2,MPEG4 for DVB and HSDPA-Style Adaptive Turbo-Transceivers, 2007, john Wiley & Sons Ltd, Pp 414-415

52. Sullivan G. J, T. Stockhammer and T. Wiegand.: Draft H.264 video coding standard for mobile applications, October 2001,in Proceedings of IEEE International Conference on Image Processing, vol. 3, (Thessaloniki, Greece),pp. 573–576

53. Tanenbaum A. S..: Computer Networks, Fourth Edition, March 17, 2003, Prentice Hall, Pp 400-407.

54. Tyagi A, J.K. Muppala,and T. Bancherdvanich.: VoIP Performance on Differentiated Services Enabled Network, 2000, in IEEE International Conference on Networks (ICON00) ,Pp 419-423

55. Wang, Y.: Survey of objective video quality measurements, Technical Report, (2006), Worcester Polytechnic Institute T1A1.5,Pp 96- 110

56. Wang,Y. and Q.-F. Zhu.: Error control and concealment for video communication A review, May 1998, Proceedings of the IEEE, vol. 86, no. 5, Pp 974-997

57. Wu D, Y. T. Hou, and Y.Q. Zhang,: Transporting real-time video over the Internet: Challenges and approaches, Dec. 2000, Proc. IEEE, vol. 88, pp.1855–1875

58. Wu H, M. Claypool and R. Kinicki.: A Model for MPEG with Forward Error Correction and TCP-Friendly Bandwidth,(2003), Proceedings of Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV), Monterey, California, USA.

59. Yubing, W.: Survey of Objective Video Quality Measurements, EMC Corporation Hopkinton, MA 01748, Pp 1-6