Error Tolerant Video Streaming in IEEE WLAN

by

Saima Shaheen
(PhD-076)

Supervised by

Dr. Muhammad Usman Akram

THESIS
submitted in partial fulfilment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY
in
COMPUTER SOFTWARE ENGINEERING

Department of Computer Engineering
COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING
NATIONAL UNIVERSITY of SCIENCES & TECHNOLOGY
2015
In the name of Almighty Allah
The Most Beneficent The Most Merciful
Abstract

Broadband network proliferation at higher bit rates has spurred a host of multimedia streaming applications like internet video surveillance, live broadcast of international events, distant learning and entertainment etc. Wireless networks have been recognized as the most popular networking technology due to their flexible infrastructure, portability and easy deployment. Today’s wireless internet user is truly global and can access content free of space or time constraints. But the caveat is that efficient streaming of multimedia contents over wireless links has proven to be an open research problem. In-time packet delivery and successful play out on receiver end are stringent quality requirements of any bandwidth intense, delay-sensitive real time multi-media application. However, wireless link characterized by fluctuating channel bandwidth and route obstacles results in transmitted signal attenuations or loss thus adversely affect the played back video quality.

Ultimately, efficient streaming of multimedia contents over error prone wireless links has emerged as one of the most challenging problems of the current era of digital communication. However, applying unequal error protection strategies and avoiding unnecessary packet discard at various network levels yield valuable outcomes. In this research work, an error control mechanism has been proposed that enables a multimedia application to adapt the video bit rate gracefully with changing channel conditions based on intelligent network estimation. Intelligent link estimation triggers the rate adaptation on sender side through reducing the video bit rate by re-encoding at lower quantization scale. In addition, an error frame forwarding approach based on the idea of discriminating video streaming calls from the data packeting over IEEE WLAN channels through bit demarcation in network packet headers. Error computation at various network levels are evaluated and disabled in order to attain efficient utilization of channel data rates characterized by minimized re-transmissions and reduced delays with least error checksum computations and packet re-transmissions, increased throughput characterized by the higher number of packets available for decoding and enhanced multimedia visual quality due to gap elimination (appears as a consequence of some frame loss).

Results have been supported through experimentation involving numerous video statistics as well as tools and techniques that can possibly be deployed in standard video codecs. Objective and Subjective quality assessments have yielded the encouraging outcomes for a mix of standard video clips (foreman, mother daughter, tennis) with variable mobility level (degree of variations in background and foreground frame contents). Reconstructed video assessment has further been improved through deploying perceptual quality tool with a rich set of parameters available for precise and complete QoE based video analysis. It has been observed that proposed error control strategy has outperformed the traditional error elimination techniques(particularly Automatic Repeat reQuest (ARQ)) for all Group Of Picture (GOP) structures of standard video codecs in terms of efficient utilization of available WLAN channel data rates even at higher Bit Error Rates (BERs). Selective re-transmission events have been incorporated for inevitable part of video bit stream without causing a noticeable drop in objective or subjective media quality. Moreover, proposed error frame forwarding approach exhibits higher link saving even at higher BER with acceptable visual quality.
# Table of Contents

**CHAPTER 1** .................................................................................................................. 1  
Introduction .................................................................................................................. 1  
1.1 Motivation .................................................................................................................. 2  
1.2 Proposed Solution .................................................................................................... 3  
1.2.1 Objectives ........................................................................................................... 3  
1.3 Thesis Organization ................................................................................................. 4  

**CHAPTER 2** .................................................................................................................. 5  
Multimedia Streaming over Wireless Link .................................................................... 5  
2.1 Digital Video ............................................................................................................. 5  
2.1.1 MPEG Video Structure ...................................................................................... 6  
2.2 Video Compression .................................................................................................. 8  
2.2.1 Intra-frame Coding ......................................................................................... 9  
2.2.2 Inter-frame Coding ......................................................................................... 11  
2.3 Video Compression Standards ............................................................................... 13  
2.4 Multimedia Streaming ......................................................................................... 14  
2.4.1 Streaming Architecture .................................................................................... 15  
2.4.2 Video Streaming Applications ......................................................................... 16  
2.4.3 Performance Constraints of Video Streaming .................................................. 18  
2.5 Wireless Networks .............................................................................................. 19  
2.5.1 Design Challenges of Wireless Networks ....................................................... 20  
2.6 Network Error Control Techniques ...................................................................... 22  

**CHAPTER 3** .................................................................................................................. 23  
Literature Survey .......................................................................................................... 23  
3.1 Prevalent Error Control Techniques ...................................................................... 23  
3.2 Accessing Corrupted Network Packets .................................................................. 26  
3.2.1 Reliable UDP .................................................................................................... 27  
3.2.2 UDP-Lite .......................................................................................................... 27  
3.2.3 Complete UDP .................................................................................................. 28  
3.3 Exploring Video Statistics ..................................................................................... 28  
3.4 Cross Layer Solutions ............................................................................................. 31
CHAPTER 4 .................................................................................................................. 35

Proposed Streaming Design ....................................................................................... 35

4.1 Design Architecture .......................................................................................... 35
4.2 ILE Interaction Design ...................................................................................... 38
4.3 Layer Coordination Mechanism ........................................................................ 39
4.4 Layer Re-configuration ...................................................................................... 40
  4.4.1 Application Layer ......................................................................................... 40
  4.4.2 Transport Layer .......................................................................................... 41
  4.4.3 Network Layer ........................................................................................... 42
  4.4.4 Data Link Layer ......................................................................................... 42
  4.4.5 Physical Layer ............................................................................................ 44

CHAPTER 5 .................................................................................................................. 47

Results & Discussion ................................................................................................. 47

5.1 Data Set ............................................................................................................. 47
5.2 GOP Structure ................................................................................................... 48
5.3 Media Quality Analysis ..................................................................................... 49
  5.3.1 Objective and Perceptual Quality Analysis ............................................... 50
  5.3.2 Subjective Quality Analysis ....................................................................... 53
5.4 ILE Modeling Accuracy ..................................................................................... 54
5.5 Overhead of Feedback Mechanism .................................................................... 56
  5.5.1 Computation Overhead ............................................................................... 56
  5.5.2 Delay Overhead .......................................................................................... 56
5.6 Performance Analysis ......................................................................................... 57
  5.6.1 Channel Analysis ....................................................................................... 58
  5.6.2 Media Quality Trends ................................................................................ 59
  5.6.3 Link Saving ................................................................................................ 69
  5.6.4 Visual Perception ....................................................................................... 74
5.7 Comparison with other techniques ................................................................... 79
  5.7.1 Throughput Comparison ............................................................................ 79
  5.7.2 Channel Usage Comparison ....................................................................... 80

CHAPTER 6 .................................................................................................................. 82

Conclusion & Future Directions ................................................................................. 82
6.1 Conclusion .................................................................................................................. 82
6.2 Contribution ............................................................................................................... 84
6.3 Future Research Directions ..................................................................................... 84
  6.3.1 Media Statistics ...................................................................................................... 84
  6.3.2 Compression Parameters & Techniques ................................................................. 84
  6.3.3 Media Visual Perception ....................................................................................... 84
  6.3.4 Utilizing Layer Potentials ..................................................................................... 85
  6.3.5 Dynamic ILE Parameters ..................................................................................... 85

REFERENCES.................................................................................................................... 86
APPENDIX A Subjective Assessment Method ................................................................. 92
LIST OF FIGURES

2.1 A GOP of video sequence ........................................................................................................... 7
2.2 MPEG frame dependencies ......................................................................................................... 8
2.3 Block Matching .......................................................................................................................... 12
2.4 Streaming Architecture .............................................................................................................. 16
2.5 Infrastructure mode IEEE 802.11 wireless network for video streaming ............................. 20
2.6 Multi-path fading with three reflection paths ............................................................................. 21
3.1 Header Description for UDP and UDP-Lite Segments .............................................................. 28
4.1 Proposed Streaming Design ....................................................................................................... 35
4.2 System Block Diagram .............................................................................................................. 36
4.3 ILE Interaction Design ............................................................................................................. 38
4.4 Layer Coordination Design ...................................................................................................... 40
4.5 IEEE 802.11 MAC Frame Header .............................................................................................. 43
4.6 Description of the route taken by 802.11 video frame .............................................................. 44
4.7 IEEE 802.11 PHY Packet .......................................................................................................... 45
4.8 Bit demarcation in video packet at various layers ...................................................................... 45
5.1 Multimedia quality trends with various GOP structures .......................................................... 49
5.2 VQMT input and output ............................................................................................................. 50
5.3 ILE comparison result ............................................................................................................... 54
5.4 Decision tree generated for output generation of ILE module ................................................. 55
5.5 Play out time comparison for different streaming strategies .................................................... 57
5.6 PSNR estimation for IBP GOP .................................................................................................. 61
5.7 PSNR estimation for IPI GOP .................................................................................................... 63
5.8 PSNR estimation for M-JPEG GOP ........................................................................................... 64
5.9 PSNR estimation for IBP GOP .................................................................................................. 66
5.10 PSNR estimation for IPI GOP .................................................................................................. 68
5.11 SSIM Indexing ......................................................................................................................... 75
5.12 VQM Plot with average VQM value = 0.422017 ....................................................................... 76
5.13 Description of B5 frame as a bad frame .................................................................................... 79
5.14 Throughput drop per station with increasing number of wireless users .............................. 79
5.15 Throughput comparison .......................................................................................................... 80
LIST OF TABLES

2.1 Frame quality and compression gain as a function of quantization step ($Q$).................. 11
2.2 Video Codec Development Progress ........................................................................... 13
2.3 Major applications of video compression .................................................................... 14
3.1 Objective Quality Metrics Comparison ........................................................................ 30
4.1 Adaptation strategy with respect to MAC control frame bits specification ............... 37
4.2 ILE input parameter categories ................................................................................. 39
4.3 MAC frames types through 'Type' field ...................................................................... 43
5.1 Video samples with different mobility levels ............................................................. 47
5.2 ILE training and testing accuracies ............................................................................ 55
5.3 Adaptation strategy with respect to MAC control frame bits specification ............... 56
5.4 Video bit rate for different GOP structures ............................................................... 58
5.5 Effects of various impairment levels ......................................................................... 59
5.6 IBP Quality Assessment for PB-EFF .......................................................................... 62
5.7 IPI Quality Assessment for PB-EFF .......................................................................... 63
5.8 M-JPEG Quality Assessment for PB-EFF ................................................................... 65
5.9 IBP Quality Assessment for A-EFF ............................................................................ 67
5.10 IPI Quality Assessment for A-EFF ............................................................................ 69
5.11 Link Saving Trend of GOP M-JPEG for IEEE 802.11 g.......................................... 70
5.12 Objective and subjective quality estimates ............................................................... 71
5.13 Link Saving Trend of GOP IBP for IEEE 802.11 g................................................. 72
5.14 Link Saving Trend of GOP IPI for IEEE 802.11 b................................................... 73
5.15 Objective and subjective quality estimates ............................................................... 74
5.16 Original and reconstructed frames ............................................................................ 76
5.17 Description of frame 635 with VQM value=7.9205............................................... 77
5.18 Objective and subjective VQA algorithms profile .................................................... 78
5.19 Channel usage trends of various streaming options.................................................. 81
Chapter 1

Introduction

Portable network devices have promoted the integration of digital communication and complex computation services in a rapidly flourishing domain of frameworks, known as multimedia applications. These applications involve processing and transmission of High Definition (HD) video contents like video surveillance, live broadcast of international events, internet video gaming, video telephony, telemedicine and distant learning services are to name a few. This domain of applications has facilitated the internet user to get aware of global updates as well as provides access to the world of uninterrupted entertainment. On the other hand, wireless networks have been recognized as the most popular networking technology due to their flexible infrastructure, portability and easy deployment. Consequently, a large number of applications have been transited from wired to wireless communication medium. Demand of multimedia contents by the viewer without time and space constraints, can be better fulfilled through broad band network proliferation offering higher bit rates.

Both the domains are intermingled to offer a luxurious aspect of digital computing through portable accessibility to multimedia contents. But the caveat is that efficient streaming of multimedia contents over wireless links has proven to be an open research problem. Performance challenges of wireless infrastructure and stringent quality requirements of multimedia applications have blurred the idea of wireless media streaming, and it has emerged as one of the most challenging problems of the current era of telecommunication.

Multimedia applications require a guaranteed quota of channel bandwidth for a uninterrupted smooth flow of multimedia contents. Normally, digital images or video frames are displayed at a certain rate for human eye to perceive the fluid motion in a way that an ultimate threshold delay of 100 milli seconds could be tolerated. This characterizes the delay sensitive nature of video applications that becomes more crucial in case of real time interactive media applications like video conferencing. Though, a video file is highly redundant with repeated patterns and useless information, that is why all the prevalent video codecs put best efforts in reducing video transmission bit rates. However, ultimate compressed media stream still demands a guaranteed bandwidth share, particularly for HD videos with higher degree of mobility and details involved. Video frames should be delivered in time to assure successful play back on receiver end even at the cost of media quality. This is attributed to the fact that
unlike data applications, a video application may tolerate few bit errors and ultimate degraded media quality, depending upon the type of video frame affected.

Streaming constraints of multimedia applications become worst when deployed in wireless medium. Due to frequency selective nature of radio signals involved, wireless communication poses various design challenges like phenomena of free space propagation, multi-path fading and shadowing adversely affect the source radio signal resulting in error burst or even packet losses. Normally, such higher Bit Error Rate (BER) is mitigated through deploying various error control techniques like the implication of Forward Error Correction (FEC) coupled with interleaving, but it appears as possible causes for the long delay. For instance, round-trip propagation delay in the 3G wireless networks is in the order of 100 ms without involving the link-level retransmission. Whereas involving link-level retransmission, the delay for the wireless link alone becomes significant. These long round-trip delays tend to lessen the efficacy of numerous error control services. Ultimately in real time streaming applications, end to end re-transmission events are significantly reduced. Since most of real time multimedia applications prefer a short interval glitch in video display rather than a noticeable pause that may emerge as a consequence of a single or few frames loss.

1.1 Motivation

In best effort networks, error correction and adjustment services at various network levels try to mitigate the effect of channel induced errors. Usually, at MAC and transport layers, network packets are discarded upon getting the wrong checksums and are requested to be re-sent from the respective sender application. These typical error operations do not seem suitable for video streaming applications specially when deployed in an error prone wireless medium where bit error induction becomes significantly high as a consequence of sender PHY layer or reflection from the route obstacles. So, wireless channels have been proven to induce higher degree of distortion in video signal originally sent. Ultimately a large number of corrupted packets appear after getting evaluated through Cyclic Redundancy Check (CRC) at MAC layer in wireless protocol stack. This triggers a high rate of re-transmissions that tends to throttle the channel bandwidth and adds-up to net transmission latency.

However, as a precautionary measure of higher BER in radio communication, a threshold BER of up to $10^{-5}$ - $10^{-6}$ is acceptable without causing a major difference in the reconstructed media quality. Experimentally, we have proven that this factor can further be exceeded safely (depending upon the objective of a video application and sensitivity of contents affected) in
order to avoid unnecessary packet discards and costly re-transmissions, since in time contents delivery is of at most concern for a real time streaming application. IEEE 802.11 e has been proposed as a typical standard for multimedia streaming, but it lacks in major aspects as it mainly comprises of MAC layer services whereas open research issues involved in wireless video streaming demand for a joint layer coordination mechanism in order to exploit the useful channel and video information for efficient link resources usage. In addition to this, a major fraction of corrupted network packets contains error free contents that can efficiently be utilized to reduce the channel effective throughput wastage through eliminating the re-transmission latency. On the other hand, disabling the error concealment practices at error control layers completely may result in disastrous consequences especially when critical video data and essential header information (like source and destination addresses) is involved. This paves the way for our proposal of a partial error computation and adaptability strategy to facilitate error tolerant, efficient video streaming mechanism to attain higher packet throughput through efficiently utilizing the link bandwidth.

1.2 Proposed Solution
A blend of techniques and design strategies has been proposed based on channel and video statistics in order to facilitate an optimized wireless streaming scenario. Basic idea proposed is based on an intelligent agent that is able to get aware of changing channel error states and to trigger a feedback event in a way to adapt the video bit rate with respect to available channel throughput. Intelligent link estimation is based on channel statistics as well as outcomes of QoE based video analysis which reflects the video appropriateness for end user visual perception. Ultimate feedback event directs towards various probable bit rate adaptation scenarios through deploying differential error control strategies. For instance, one of the proposed error control strategies promotes the usage of error free contents of corrupted network packets based on error tolerance factor of redundant video data. Basic objective is to make maximum number of video frames available for decoding on receiver end (to bridge the gap created due to packet discard) at the cost of minor quality degradation. Eliminating the packet discard greatly enhances the video throughput upon decoding, as well as efficient usage of link capacity rather than throughput wastage in re-sending the already sent data. This reduces the transmission latency greatly. Through intelligent link estimation, various control switches in a radio communication system have been nominated in order to tune a multimedia transmission against higher channel distortion.
The proposed idea is based on the significance of QoE based analysis has been decoded of reconstructed media through subjective quality assessments methods, that exhibit potential correlation with human perception of visual quality. Distorted video stream is passed up to the application decoder and rather than estimating the objective profile a reconstructed video sample is assessed through subjective quality metrics to judge its suitability for visual perception.

**Solution statement:** A network aware media streaming mechanism has been proposed, based on channel and video statistics analysis that facilitates the bit rate adaptation on sender APP with respect to available channel throughput. Rate adaptation further triggers various error control strategies that exploit video scalability information as well as the error free contents of corrupted network packets.

1.2.1 Objectives
Proposed streaming design has helped to:

- Efficiently utilize the link resources
- Improve viewer visual experience
- Enhance video throughput for interactive multimedia applications
- Reduce transmission latency

1.3 Thesis Organization
Rest of the thesis is organized as follows:
Chapter 2 explains an MPEG video structure and collaboration of video frames in a GOP. It also includes the performance challenges of a radio communication system and a brief survey of error control schemes. Chapter 3 gives a detailed and critical review of already presented schemes and algorithms for wireless multimedia streaming. In chapter 4, the proposed technique and algorithms are discussed. It includes error control mechanism as well as layer re-configuration design. Experimental setup and results have been discussed in chapter 5 and Chapter 6 concludes the whole research and highlight the future directions.
Chapter 2

Multimedia Contents over Wireless Link

In this chapter, different components of the work domain have been discussed, beginning from the basics of digital video to video streaming mechanism along with details of video compressions steps have been explained. Infrastructure of video streaming in WLAN has been elaborated and the identified error causing phenomenon of wireless channel have been discussed with sufficient level of clarity and understanding. Finally, it summarizes the prevalent error control techniques deployed to detect and correct various error scenarios at certain network levels.

2.1. Digital Video

Digital video is a series of frames or digital images, presented at regular time interval \( t \) to create the perception of motion. For instance, the NTSC (National Television System Committee) has specified a temporal sampling rate of 30 frames/second and interlace 2 to 1. A digital image is consequent through the process of quantization applied on a continuous image both spatially and in amplitude. Digitizing the spatial coordinates and amplitude are called image sampling and gray-level quantization, respectively. For instance, a continuous image is represented through \( g(x,y) \), where the amplitude at the coordinates \((x,y)\) is the intensity value or brightness of an image [53]. The transformation of a continuous image to a digital image can be expressed in equation (2.1) as:

\[
f(m,n) = Q[g(x_0 + m \delta x, y_0 + n \delta y)]
\]  

(2.1)

where \( Q \) is a quantization operator, \( x_0 \) and \( y_0 \) are the origin of image plane, \( m \) and \( n \) are the discrete values \( 0,1,2,\ldots \), and \( \delta x \) and \( \delta y \) are the sampling intervals in the horizontal and vertical directions, respectively. If this sampling process is extended to a third temporal direction, a sequence, \( f(m, n, t) \) is obtained through equation (2.2) as:

\[
f(m,n,t) = Q[g(x_0 + m \delta x, y_0 + n \delta y, t_0 + t \delta t)]
\]  

(2.2)

Where \( \delta t \) is the time interval with \( t \) ranges from 0, 1, 2,.....and each of the image is a video frame [1]. As a result of this spatio-temporal sampling, video frame exhibits high spatial and temporal correlation, just as the analog signals did before video data compression.

A video frame captured from a camera can be represented as a combination of highly correlated RGB (red, green, blue) signals, visualized in a YUV colour space where \( Y \) denotes
the frame brightness or luminance and UV collectively show colour or chrominance components. This colour space is large in size and highly redundant, so requires huge bandwidth for transmission. For instance, a 5 minute YUV video of resolution 352 x 288 pixels, displayed at frame rate of 30 fps will have a file size of 1.27GB. Applying a video compressor will significantly reduce its size, it is reduced to 7.5 Mb when compressed at a bit rate of 200 kbps using MPEG. Standard video codecs are applied to reduce the storage and streaming media to an economical and manageable size.

2.1.1 MPEG Video Structure

MPEG encoding involves partitioning of a video sequence into Groups of Pictures (GOP) (shown in figure 2.1), normally include three types of pictures: Intra-coded (I) pictures, Predictive-coded (P) and bidirectional predictive-coded (B) pictures [2]. These frame types serve differently in encoding and decoding processes, most important are the I-frames.

**I-frames**

I-frame directs towards the process of intra-coding, a way in which a frame is independently compressed without incorporating any other frame type assistance, rather it acts as an anchor frame for subsequent frames. This category is effectively identical to baseline JPEG images, hence spatial correlation is eradicated. I-frames facilitate to provide high-speed seeking through an MPEG video to the nearest I-frame. So after cutting a video, video segment is possibly played back at a point with the first I-frame in the segment.

**P-frames**

P-frames or inter-frames exist to improve compression through exploiting the temporal correlation in a video over time interval t. P-frames are meant to store the frame difference in image from the anchor frame (either an I-frame or P-frame) immediately preceding it in the form of pixel coordinates called motion vectors. A motion vector is computed for each macro block of the frame with respect to its anchor frame. Consequent motion vector data is embedded in the P-frame for future use by the decoder. A P-frame usually contain a mix of intra-coded and forward predicted blocks in various ratios depending upon the nature of video being compressed.
B-frames

B-frames or backwards-predicted frames are basically similar to P-frames; except that they undergo predictions using both the previous and future frames as two anchor frames in forward and backward directions respectively. B-frames processing involve complex computation overhead because it is mandatory for the player to decode the dependent I- or P-anchor frame (sequentially after the B-frame) prior to decode and display the depending B-frames. This increases the computation overhead since it requires large data buffers, and causes an increased delay in encoding and decoding processes. But it greatly adds to compression gain characterize through the lower bit rates since B-frame can be inserted to control the bit rate whenever desired. In addition to backwards-predicted or bidirectional predicted blocks, a B-frame may contain a certain number of intra-coded and forward-predicted blocks.

Figure 2.1: A GOP of video sequence [1]

The distance between two nearest I-frames is denoted by N, which is the size of GOP. The distance between two nearest anchor frames is denoted by M. MPEG-1 most commonly uses a GOP size of 15-18. i.e. 1 I-frame for every 14-17 non-I-frames. In some intelligent encoders, video parameters like GOP size are dynamically adjusted up to some pre-selected threshold value. Limits are placed on the maximum number of frames between I-frames in accordance with decoding complexity, decoder buffer size, recovery time after data errors,
and seeking ability. However, large values of M and N result in error propagation and practically not recommended for real time streaming applications like video surveillance.

Due to the frame dependency, frames are decoded and displayed in different order, as before decoding a depending frame its anchor frames must be decoded. Similarly, errors induced in dependent frames due to channel impairments affects the depending frames and propagate accordingly. Frame dependencies derive a far different pattern than the frame display sequence, as shown in figure 2.2. It is apparent from frame dependencies in MPEG GOP structure that an I-frame error propagates through all the subsequent frames of the whole GOP whereas a P-frame error affects the related P and B-frames only and restores the quality level on getting the next anchor frame. However, B-frames errors are isolated.

![MPEG video frames in display order](image1)

(a) MPEG video frames in display order

![MPEG video frames in transmission order](image2)

(b) MPEG video frames in transmission order

**Figure 2.2:** MPEG frame dependencies [18]

### 2.2 Video Compression

A video compression standard works to eliminate spatial (pixel correlation inside an individual frame) and temporal redundancies (pixel correlation among the frames) present in a video, called intra and inter-frame coding respectively. A sequence of video frames exhibit content similarity or resemblance because normally a sequence of frames displays same event or stages of a motion involved. This yields higher degree of pixel correlation in subsequent frames, characterized by minor changes in background or foreground contents. After eliminating these spatial and temporal redundancies, a series of code words is obtained which is further compressed to yield the compressed bit stream through variable length coding. Following are the details of sub-processes involved in the process of video compression as:
2.2.1 Intra Frame Coding

Nearby pixels within a single video frame possess similar amplitude level and higher RGB correlation, spatial encoding has been attributed to the fact that human eye is unable to differentiate small colour differences as easily as it can perceive changes in brightness. It is advantageous in a way that resembling areas of colour can be averaged out, as in jpeg images. Intra-coding or still image coding is based on an efficient transformation coding technique called Discrete Cosine Transform (DCT). DCT technique has been approved as the most suitable transformation technique for video compression as compared to all the newly emerged wavelet based transformation techniques [50], that's why all the standard codecs are based on DCT. Natural imagery contains most of its energy in low frequencies components, exploited through applying cosine transforms. DCT is based on a lossy compression algorithm that deploys the techniques of image sampling at regular intervals. Frequency components of the resultant sample are assessed for usability and less significant (components that merely affect the Human Visual System (HVS)) components are discarded. In order to exploit the maximum correlation present among the neighbouring pixels, a video frame is divided into smaller blocks of 8x8 size. 2-D DCT is applied on these 8x8 smaller blocks that restore their energies in a smaller fraction of DCT coefficients. It is a lossless transformation technique and merely changes the pixel representation of a block set.

A two–dimensional DCT for an N ×N block of pixels can be represented in equation (2.3) as $f(i,j)$ denotes the pixel values and $F(u,v)$ shows the transform coefficients obtained.

$$F(u, v) = \frac{2}{N} \cdot C(u) \cdot C(v) \cdot \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos \left( \frac{2(2i+1)u\pi}{2N} \right) \cos \left( \frac{2(2j+1)v\pi}{2N} \right)$$

(2.3)

where C(u) and C(v) are two consecutive one-dimensional DCTs in horizontal and vertical directions respectively.

The lowest order coefficients are referred to as DC-component and other components are named as AC-components. These coefficients are quantized and processed further through applying video compression steps like zigzag scanning, run-length coding and Huffman coding and that yield an ultimate compressed bit stream.

After DCT, a video frame shows the frequency component representation of the pixel values, where lower frequency components contains maximum information contents and consequently higher frequency components are small. Spatial encoding involves a process called quantization to eliminate this frequency components redundancy, where a quantizer is
characterized through a quantization threshold $T$ and quantization step size $Q$, that determine the compression gain and ultimate media quality as:

- $F(u,v)$ coefficients having values less than $T$ are assigned value zero an
d referred to as dead zone.
- $F(u,v)$ coefficients with values greater than or equal to $T$ are divided by twice of $Q$ and rounded to nearest integer value.

DCT coefficients are converted to quantized coefficients through the equation (2.4) as:

$$I(u, v) = \begin{cases} 0 & \text{for } |F(u, v)| < T \\ \left[\frac{F(u, v)}{2Q}\right] & \text{for } |F(u, v)| \geq T, \end{cases} \quad (2.4)$$

Quantization step size $Q$ determines the decoded video quality and compression gain or bit rate required to transmit the compressed bit stream, illustrated in Table 2.1 as video quality is degraded with increasing the quantization step [7]. This also reflects the higher compression gain achieved for larger $Q$ at the cost of reconstructed media quality. Two types of coding are Constant Bit Rate (CBR) and Variable Bit Rate (VBR) in accordance with quantization step size $Q$. In CBR, process of quantization is controlled in a way that $Q$ is adjusted after feedback from the resultant bit rate required to be suitable for available channel bandwidth. Consequently, a constant bit rate is required for the whole video transmission, since required variations are adjusted through varying $Q$. In contrast for VBR, quantization step $Q$ is kept constant and required bit rate varies according to the video frames details under compression [51,52].

Quantization phase is followed by zig-zag scanning that prepares the video contents for variable length coding. In this step, values of quantized video frames are retrieved in a way that first lower frequency coefficients are scanned followed by AC-components. This scanning has a purpose of attaining higher compression gain through picking zero-values first and it can be stopped before reading all the non-zero values (less significant higher frequency coefficients i.e. AC components) to attain further compression.
Table 2.1: Frame quality and compression gain as a function of quantization step ($Q$)

<table>
<thead>
<tr>
<th>Quantization step ($Q$)</th>
<th>Quality (PSNR-dB)</th>
<th>Frame size (byte)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>44.95</td>
<td>4640</td>
</tr>
<tr>
<td>35</td>
<td>33.93</td>
<td>135</td>
</tr>
<tr>
<td>40</td>
<td>30.27</td>
<td>79</td>
</tr>
</tbody>
</table>

2.2.2 Inter-frame Coding

This technique is meant to exploit the similarity of the contents in successive frames, also called temporal redundancy. For instance consider a frame sequence capturing a book reader, the only difference among the upcoming frames is characterized by the motion in face region of the reader. The similarity among the video frames is incorporated to compute the frame differences in order to attain maximum compression. A video frame is viewed as a grid of macro blocks that are considered to predict the most similar block in previous or next coming frame through deploying the Block Matching Algorithms (BMA) [54]. BMAs work to estimate the displacement of the prediction block from the actual current block, referred to as Motion Vectors (MV) computed through the process of Motion Estimation (ME), as shown in Figure 2.3. Prediction macro block search is carried out in a proximity of $n$ pixels within a video frame, where full search to get best possible match is carried out through $(2n+1)^2$ computations. Researchers have worked out various BMAs like New Three Step Search (NTSS), Exhaustive Search (ES), Simple and Efficient Search (SES) etc to facilitate the efficient search with less computation cost.
Size of macro block and the search space are the control parameters that drive decoded frame quality and computation cost to get the prediction block. We have worked on the application level to attain the optimized quality reconstruction in a video sensor network that has defined a trade-off between frame quality and macro block size searched in a confined space to improve the quality further [55].

After estimating the MV for the video frame, frame difference is computed which is fed to compression steps of intra frame coding to achieve the compression further, called Motion Compensation (MC). So through inter-frame coding, a frame is represented through:

- Compressed frame difference from the previous or next coming frame.
- A set of MVs for predicted blocks in the searched frame.

These contents are utilized in the process of decoding on receiver end to reconstruct the original frame. Such frame dependencies prediction directly affect the performance of video streaming applications in terms of compression efficiency and error adjustment services involved. In a GOP, first frame is intra-coded through spatial encoding and it acts as the anchor frame for subsequent frames of the GOP, that undergo temporal encoding using ME and saves the bandwidth greatly. Video compression is a trade-off between storage space, video quality, and the cost of hardware required to decompress the video in a reasonable time.

**Figure 2.3:** Block Matching
2.3 Video Compression Standards

A video compression standard is not the standardization of encoding or decoding processes, rather it reflects a way to clearly identify the syntax of a compressed bit stream obtained through these steps such that to decode it properly. This deliberate approach paves a way for encoder and decoder process improvement, the only constraint is that an encoder must generate a syntactically correct bit stream that is efficiently decoded by a decoder to produce the original stream compressed.

Despite a large variety of video coding and decoding systems, two work bodies have put efforts in producing the video compression standards, International Telecommunication Union-Telecommunication (ITU-T) and International Standardization Organization (ISO).

The early H.261 codec of the ITU–T was focused on delivering video over Integrated Services for Digital Networks (ISDN) with a fixed bit rate of $n \times 64\text{kbit/s}$, where $n$ denotes the number of multiplexed ISDN lines. From this starting point, codecs were developed for different purposes such as the storage of digital media or delivery over packet oriented networks.

The evolving standards achieved better quality with lower bit rates and thus better rate distortion performance. Table 2.2 sketches an overview of the video standards along with the techniques deployed in each year, with computation complexity increases with time.

<table>
<thead>
<tr>
<th>Video codec</th>
<th>Technique used</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCT</td>
<td>Intra frame coding</td>
<td>Ahmed et. al., 1974</td>
</tr>
<tr>
<td>JPEG</td>
<td>Conditional replenishment,</td>
<td>1994</td>
</tr>
<tr>
<td></td>
<td>DPCM, scalar quantization</td>
<td>1984</td>
</tr>
<tr>
<td>H.120</td>
<td>Frame difference coding</td>
<td>1988</td>
</tr>
<tr>
<td>H.120 ver. 2</td>
<td>Motion compensation</td>
<td>1991</td>
</tr>
<tr>
<td>H.261</td>
<td>Half-pel accurate motion</td>
<td>1993</td>
</tr>
<tr>
<td>MPEG-1</td>
<td>compensation</td>
<td>1994</td>
</tr>
<tr>
<td>MPEG-2/H.262</td>
<td>Variable block-size motion</td>
<td>1996</td>
</tr>
<tr>
<td>H.263/MPEG-4</td>
<td>compensation</td>
<td>1999</td>
</tr>
</tbody>
</table>

Various standard video codecs have been used for different streaming applications according to the suitability of the bitrates offered and techniques used. Table 2.3 shows the standard
video codec deployment in various multimedia applications along with their supported bit rates.

**Table 2.3:** Major applications of video compression

<table>
<thead>
<tr>
<th>Applications</th>
<th>Bit rate (bps)</th>
<th>Video Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital television broadcasting</td>
<td>(2-6) Mbps</td>
<td>MPEG-2</td>
</tr>
<tr>
<td></td>
<td>(10-20) Mbps for HDTV</td>
<td></td>
</tr>
<tr>
<td>DVD video</td>
<td>(6-8) Mbps</td>
<td>MPEG-2</td>
</tr>
<tr>
<td>Internet video streaming</td>
<td>(20-200) kbps</td>
<td>Proprietary, similar to H.263, MPEG-4, or H.26L/JVT</td>
</tr>
<tr>
<td>Videoconferencing, video telephony</td>
<td>(20-320) kbps</td>
<td>H.261, H.263</td>
</tr>
<tr>
<td>Video over 3G wireless</td>
<td>(20-100) kbps</td>
<td>H.263, MPEG-4</td>
</tr>
</tbody>
</table>

### 2.4 Multimedia Streaming

Multimedia streaming is a mechanism that enables a viewer to access the media contents in real time from a remote media server without a prior download in a way that media contents are streamed and played as they arrive. Media quality varies as a function of system processor speed and the link bandwidth. Video streaming can be summarized as a set of following steps:

1) Fragment the compressed video into packets

2) Begin to deliver these packets

3) Packets are delivered and decoded at the receiver while it is being played

The approach of streaming facilitate the viewer with simultaneous transmission and play back of a video, a viewer has just to wait for initial buffering of the video prior to the beginning of play back called pre-roll delay (usually in the order of 5-15 seconds) and rest of the contents are delivered while it is being played. This approach has overcome the limitations of huge storage space and video start-up delay, normally faced in file download. Streaming begins as a browser sends request for metafile to a location where it resides. After finishing with metafile download, browser delivers the video contents to the media player that tends to extract the location as well as protocol from the metadata. Media player tries to contact the
specialized media server mentioned in the metafile and after the connection establishes, streaming commences using the pre-specified protocol.

Streaming uses specialized protocol suite to accomplish the streaming process mainly including Real-Time Streaming Protocol (RTSP) [3] and Real Time Protocol (RTP) [4]. RTSP is used to establish and control the mechanism of streaming using remote control-like functions e.g. Play, Rewind, Fast Forward etc. Whereas video contents are delivered using RTP which is an application level protocol and facilitates the transmission of media files through deploying sequence numbering, time stamping and delivery supervision. RTP usually works in collaboration with any transport protocol to support media transmission.

2.4.1 Streaming Architecture

A media streaming architecture may constitutes a variety of components that exchange the information during the process of streaming. However, a baseline streaming architecture consists of simple components as encoders, media servers and clients as shown in Figure 2.4.

2.4.1.1 Encoder

Raw video files are captured through a video camera, which are not suitable for transmission due to large sizes. Prior to streaming, a media file has to be converted to transmittable format through a media encoder that eliminates the fine details and yields a comparatively low quality video in accordance with the channel capacity.

For efficient transmission, a video must be encoded at a bit rate lower than the link bandwidth but it is not the case all the time and poses many challenges specially for time varying bandwidth channel.

2.4.1.2 Server

Servers are meant for data storage and responding to clients request for stored contents. Two categories of servers are incorporated in media streaming, web servers and specialized media servers in different transmission scenarios. In first scenario, a web server is used alone in a way that it considers the client request for media file as a data object request and wraps up the media file inside an HTTP message to deliver it to the requesting browser. This approach is advantageous in a way that web servers are easy to deploy without adding any specialized infrastructure for downloading the media contents.
In another scenario, specialized media servers are used in association with common web servers. Web servers are limited to transmission of metafile for the requested video to the requesting browser. Afterwards, media servers with specialized protocol suit directly comes in contact with media player and delivers the video.

2.4.1.3 Client

Media players are standalone applications and act as the client in media streaming. A media player is designed for playing streamed or stored media files, usually launched by a browser as its extended functionality (plug-in). Streaming client uses RTSP for establishment of streaming session with the server and controls the whole process through RTSP services pause, rewind, fast forward etc.

![Streaming Architecture](image)

Figure 2.4:  Streaming Architecture [18]

2.4.2 Video Streaming Applications

A communication system is mainly influenced by numerous components, particularly when streaming of diverse quality multimedia contents is involved. In a video streaming scenario, significant factors involved should also be analyzed like application performance constraints, nature of communication link, nature of contents transmitted, and certain channel statistics.

2.4.2.1 Application Type

Video applications can be interactive or non-interactive in nature, where interactive applications enlist video surveillance, internet gaming, teleconferencing to name a few. These are highly delay-sensitive, where maximum tolerable latency depends upon the nature of application but it is approximately in order of 150 msec [5], otherwise contents arriving late
are useless. On the other hand, non-interactive applications category involves streaming of live events like multicast of academic lectures or broadcast of global events. These applications also deal with live captured video contents but do not exhibit that severe response to transmission delays as the interactive applications do.

2.4.2.2 Communication link

Video communication is being carried out in various ways like one-one (videophone), one-many communications that further includes multicast and broadcast communications. Each of these categories has been characterized through its design challenges and deployed solutions. Example of one-one or point to point communication is video phone on internet where receiver must have a back channel to sender for the purpose of feedback. In the absence of back channel, a sender is unaware of the receiver capabilities to receive the contents properly.

A well-known example of broadcast communication is digital television that delivers multimedia contents to number of users with different channel conditions. To assure the delivery of sufficient quality signal to all the recipients, appropriate source channel coding designs are chosen in order to equip the communication system for worst signal scenarios. However, involvement of large number of recipients create an unfeasibility to provide receiver feedback to sender that hinders the system adaptation.

Another one-many communication design is carried out through multicast transmission which is comparatively efficient then combining number of one-one transmission links with dedicated path for each connection.

2.4.2.3 Nature of Video Contents

Video applications may deal with two categories of video contents as stored video and real-time/online captured video, transmitted in different scenarios. For instance, Interactive video applications like videophone, video conferencing and video games require on the fly encoding of the real time captured contents. Other non-interactive applications may demand the live video contents too like live streaming of some global event etc. This category is advantageous in the sense that number of methods have been observed in the literature that incorporate ways to adapt the video parameters (like quantization scale, bit rate etc) in accordance with channel conditions for on the fly encoding.
Whereas other applications make use of pre-encoded or stored video contents from remote streaming servers or locally on DVDs. This set of application do not require at the spot encoding constraint that limits the freedom of adaptation in current channel conditions.

2.4.2.4 Channel statistics

A transmission channel of a communication system is normally characterized by its transmission delay, capacity or bandwidth and packet loss rate. Stability of these parameters define the nature of a communication channel as static or dynamic [5], where static channel is attributed to fixed bandwidth, transmission delay and negligible loss rate like ISDN. While examples of dynamic channel are wireless and best effort network with time varying characteristics of delay, bandwidth and loss rate. video streaming on a dynamic channel is more difficult comparatively and faced with many challenges.

2.4.3 Performance Constraints of Video Streaming

Video streaming poses certain performance constraints of delay sensitivity and guaranteed bandwidth that afflict the transmitted media quality, because for any communication medium it provides no guarantees on bandwidth, delay jitter, or loss rate. Specifically, these characteristics are unknown and dynamic for wireless networks. Therefore, a key goal of wireless video streaming is to design a system to reliably deliver high-quality video when dealing with unknown and dynamic:

Bandwidth

Delay jitter

In any network, bandwidth or channel capacity between two points is time-varying. In exploiting such link capacity, if a sender transmits the data faster than the available bandwidth then it will result in congestion and packet drop. Inversely, if data is transmitted at a slower rate than the available link speed, degraded media quality is consequent. The way to attain the optimal control over channel bandwidth is to precisely estimate the link capacity and then match the video bit rate with the estimated value. Challenges involved are accurate estimation of available channel bandwidth, adjusting the bit rate of pre-encoded video to current link capacity and transmitting the contents in a fair way with other concurrent flows.

Other significant factor is the transmission delay experienced by the network packets, it may vary from packet to packet named as delay jitter. Consequently, packets are reached out of
order that creates a serious problem for the decoder that is able to decode the video frames only if the contents are received in accurate order. Solution lies in involvement of a play out buffer on receiver end to store the out of order packets awaited to be decoded using reference information in upcoming network packets. However, play out buffer itself tends to add up delay. Additional latencies are caused by error control techniques like re-transmissions etc.

2.5 Wireless Networks

Wireless networks have been established as a renowned class of networking that facilitates the user connectivity to internet without physical connecting wires. Their popularity has been characterized through immense portability and easy deployment. Internet user connecting through a wireless connection is free from time and space constraints and approaches the world where ever he is and whenever he wants. Basic component of a wireless network, designated to provide access to all mobile hosts with in a specific proximity is called a Base Station (BS) or Access Point (AP). A mobile host is disconnected from the network when signals from the BS are blocked due to buildings or other route obstacles.

Wireless communication is advantageous due to lower overhead of physical connections and wires instalment, however it is mainly affected with two loop holes. First is that wireless communication is carried out through radio signals that are highly sensitive to interference with other radio devices. Such interaction results in reduced signal strength, errors and even losses, and ultimate corrupted network packets are consequent that triggers the re-transmission very often. Second, transmission capacity of a wireless link is far lower than a wired link and varies as a function of number of users sharing a base station. With increased user number, each user feels a degradation in contents quality as well as transmission speed that can be compensated through adding more BSs. A BS is further connected with a internet server through a physical wired connection. Figure 2.5 shows an infrastructure modes of IEEE 802.11 WLAN for video streaming setup, comprises a streaming server connected to an AP through wired connection that provides wireless connectivity to wireless devices like media receivers/ clients through IEEE 802.11 channel.
2.5 Design Challenges of Wireless Networks

Wireless networks are time-varying and frequency selective in nature. Phenomena of free space propagation, reflection and scattering tend to induce the high degree of channel distortion in transmitted radio signal. Error prone nature of wireless link gives birth to two categories of error induction in video streaming as: (a) error caused by factors involved in radio communication (will be discussed later in the section) (b) error caused by burstiness of video contents. Major challenges in radio communication are due to the factors like free space propagation, multi-path fading and shadowing etc.

2.5.1 Free Space Propagation

In a wireless medium, transmitted signals travel in free space that tends to reduce its strength. Signal power is attenuated due to various obstacles in the route followed. This attenuation depends upon numerous factors, given by free space propagation expression [6] in equation (2.3) as:

\[ P_r(d) = \frac{P_s G_t G_r \lambda^2}{(4\pi d)^2 L} \]  

where \( P_r \), power of received signal depends upon transmitted signal power \( P_s \), antenna gain \( G_t \) and \( G_r \) for transmitter and receiver respectively, wavelength \( \lambda \), path loss \( L \) and inter-distance of transmitter and receiver \( d \). It is apparent from the equation (2.3) that received signal strength varies as a square of the distance \( d \) between sender and receiver.

2.5.1.2 Multi-Path Fading

In a wireless environment, omni-directional cameras of the transmitter send the signals in all directions in free space. Consequently, same copies of the original signal adopt different routes to reach the destination. Moreover, reflection, diffraction and scattering form the route...
obstacles result in Non-Line of Site (NLOS) replicas of the original signal that travel through indirect paths approaching the receiver end. A signal travelling through direct path to receiver is called the Line of Site (LOS) signal. Figure 2.6 shows the model for multi-path fading with one LOS and two NLOS signals approaching the receiver end. All the signals reaching at the destination have different amplitude, phase and delay, where latency between the LOS signal arrival and the last version of NLOS signal arrival is called delay spread that highly depends upon the environmental settings. As a result of this multi-path propagation, symbols of the same signal interfere called Inter-Symbol Interference (ISI) that induces quality distortion and the signal strength is reduced. Multi-path fading can severely damage the signal quality depending upon the number of fading paths and physical characteristics of wireless environment like urban and rural areas.

![Multi-path fading model](image)

**Figure 2.6:** Multi-path fading with three reflection paths [67]

### 2.5.1.3 Shadow Fading

It is another signal distorting effect of wireless channel, caused due to impediment produced by buildings, mountains or vegetation present in the route. Transmitted signals from the mobile terminals are blocked by these obstacles for large time periods that last from hundred milli seconds to seconds depending upon the velocity of signal transmitting mobile terminal. This results in error burst or packet losses in wireless channels, characterized by rough environment with thick vegetation and closely spaced buildings.

### 2.5.1.4 Node Mobility

In contrast to a wired network, communication hosts in a wireless network are mobile and it poses another challenge on radio communication and assuring the in time delivery of contents with sufficient quota of bandwidth required.
Such phenomena of radio communication tend to induce bit errors or error burst in transmitted contents and sometimes even result in entire packet loss. Prevalent error control methods designed for the wired medium are also applied for typical wireless transmission with some modifications in applied techniques and error control codes used.

2.6 Network Error Control Techniques

To combat the effects of error, network error correction and adjustment services are categorized as: (a) Forward Error Correction (FEC) scheme, based on Shannon’s channel coding theorem (if the channel capacity is larger than the data rate, a coding scheme can be found to achieve small error probabilities) involves additional parity bits embedded in packet payload to recover the data after being corrupted by channel error. But this adds to the bandwidth requirement and become useless in error burst scenario when it is probable to lose the error control information as well. Adaptive-FEC tends to overcome the problems of simple FEC through applying the error control services after analysing the channel error characteristics, but it fails due to additional computational hardware cost and signaling overhead. (b) Re-transmission scheme, to mitigate the effect of packet error corrupted packets are re-transmitted that results in additional latency in net transmission delay specially in real-time applications. Hybrid-ARQ scheme encapsulate the advantages of fixed throughput and delay of FEC and re-transmission services. (c) Decoder-based Error Concealment, applied on receiver end to hide the effect of error. MPEG-4 decoder has in-built error concealment property but it cannot be productively applied in a wireless network with high error rates i.e. error bursts. These techniques along with their variants will be discussed in next chapters as a critical assessment of their pros and cons.

Summary:

This chapter includes the explanation of the background of research domain. Beginning from the introduction of digital video, need for a video compressor, steps of video compressions and a quick overview of video standard has been discussed. Mechanism of the video streaming has been elaborated at the root level with its constraints identified. Finally it includes the performance challenges of a wireless networks with different fading effects involved. At the end, a brief overview of the prevalent error control techniques has been presented.
Chapter 3

Literature Survey

This chapter is based on the analysis of error control strategies used so far as well as a look on new propositions to cater for the challenges of efficient wireless video streaming. Literature survey conducted has been divided into certain categories in accordance to the streaming idea proposed, each discussed with its specification and a comparison with our proposed solution. We begin the discussion with the descriptions of prevalent error adjustment techniques along with their variants proposed in the literature. Following this survey is the summary of ways to utilize the error free contents of the corrupted network packets. It further includes a survey of proposed techniques that deal with video statistics and ways of perceptual analysis. Finally, a summary of cross layer approaches to cater for the performance challenges of wireless video streaming has been presented.

3.1 Prevalent Error Control techniques

This section includes the state of the art error control techniques along with the short descriptions of their variants proposed in the literature with their pros and cons discussed. Normally whenever a network packet fails to pass the CRC at receiver MAC layer, error is detected that leads to packet drop and ultimate error adjustment services are invoked. These error control techniques or methodologies either work to undo the error occurrence or to minimize its effect and disastrous consequences. These error adjustment methods have been categorized as prior transmission error control measures and post error detection measures where post control measures can either eliminate the error entirely or to lessen its severity.

Pre-transmission error control methods include Forward Error Correction (FEC) and error resilience techniques that work at source application to enhance robustness of video against channel impairment. FEC frequently use Bose–Chaudhuri–Hocquenghem (BCH) or Low Density Parity Check (LDPC) codes typically used in wireless transmission. These methods involve the padding of packet payload with the bit patterns later used to recover the data originally sent through detecting and sometime correcting the error induced. \textit{FEC is advantageous in terms of constant throughput with fixed transmission delays, independent of the error rate in wireless channel} [7]. However, adding redundant bits induces extra bandwidth load as well as computation complexity and a fixed delay is added to each packet transmission latency which is practically not suitable for delay sensitive applications [8].
method proves to be unsuccessful in error burst scenarios particularly when added bit patterns are also at danger of loss. However, number of FEC variants have been proposed since it depends upon the functional constraints of the source application and proves to be a trade-off between number of redundant data bits added and ultimate media quality. Afterwards, in order to overcome the drawbacks of classical FEC techniques, adaptive-FEC methods have been proposed, based on adaptive design of FEC encoder after long term estimation of the error patterns of wireless channel. Generally, adaptive FEC methods have proven to be effective error control strategies and productively used in conjunction with other layer error remedial methods like re-transmissions (discussed next). However, these methods tend to increase the computation complexity and add signaling overhead.

Error resilience techniques work at application level to make the media unsusceptible through resynchronization, adaptive and scalable encoding or data recovery etc. In re-synchronization approach, a bit stream is fed to decoding from both sides, facilitated in a way that code words are created as such that these can be decoded from both sides. A decoder jumps to next synchronization point on getting some error and begin to decode in backward direction to recover the code word in error. Adaptive and scalable encoding encapsulate the application level amendment in a compressed video bit stream either on the fly in accordance with the channel conditions or at encoding level. On the fly encoding involves the re-design of encoding parameters like quantization level that yields variable bit rate video streams to be fed to the channel with fluctuating bandwidth and distortion. Scalable encoding on the other hand, refers to the term Multiple Descriptive Coding (MDC) where multiple data streams are generated from one source video, each independently decodable and jointly refinable [9].

MDC is an effective strategy to mitigate the event of packet error and loss in video transmission over noisy channels, particularly for real-time applications in which re-transmission of lost information is not practical. **MDC is proficient to maintain a constant bit rate to feed the decoder and reduces the packet error rate near to zero** [10]. It is an efficient mechanism that greatly reduces the media susceptibility against channel distortion as well as assures the fairness among all the channel sharing applications. Single Descriptive Coding (SDC) follows a pre-defined coding steps and yields a single video description with no ability to cope with loss prone transmission channel. Contrarily, MDC is a kind of data partitioning producing more than one video version with different bit rates and error resilience that results in improved scalability. Researchers have proposed a joint collaboration of application and MAC layers coupled with MDC implemented on IEEE 802.11e networks [11]. They have compared the benefits of
MDC and Single Descriptive Coding (SDC) for a cross layer design and have shown an upgrade in media quality as well as reduced end to end transmission delay. A dark side of conventional MDC is its lower side distortion quality as compared to Single Description Coding (SDC), which can be mitigated at the cost of channel capacity utilized to make the streams more redundant and robust against the channel errors. This however leads to central quality degradation. Researchers have tried to work out this problem of MDC, where idea of mixed layer MDC has been proposed [12]. Similar to the idea of Scalable Video Coding (SVC), they proposed to generate two versions from the baseline video contents as two sets of DCT coefficients are generated as base coefficients (BC) and enhancement coefficients (EC) that further produce two stream descriptions. At decoder end, if one of the descriptor is lost, BC are estimated from the available descriptor as much as possible. However, this approach exhibit a trade-off between error resilience and video redundancy since higher bit rates are consequent.

A hybrid MDC approach has been proposed where MDC has been applied in spatial and frequency domains to split motion compensated residual data and quantized coefficients, respectively [13]. It has been evident from the results that the hybrid MDC method can improve error-resilient performance through coupling data splitting techniques in more than one domain.

Post detection error control services are being activated after detecting an error scenario i.e. on getting wrong CRC outcome, these services instantly drop the affected packet and either demand for their re-transmission from the sender application until these are received correctly, named as Automatic Repeat reQuest (ARQ) or invoke other error correcting measures like error concealment methods. Basic ARQ design has been presented in Send and Wait where sender waits for the acknowledgement for previously sent packet from the receiver prior to send the next packet. A receiver sends an acknowledgment to sender to notify the data delivery, and sends duplicate or negative acknowledgement on an event of missing data packet. A sender gets aware of the error on receiver end through getting duplicate or negative acknowledgement, probably due to a missing sequence number. Erroneous data contents received are requested to be re-transmitted by the sender applications. ARQ is effective in an environment with lower probability of channel errors. However, as the channel errors become severe, ARQ ends with extra delays and becomes impractical and intolerable for real time streaming applications. One of the solutions is to provide buffering support at receiver where video packets can be stored in order to meet their play out time [14]. However, in wireless streaming scenarios with error bursts, such error
control method does not seem successful and adds complexity and cost. Other variants of ARQ have been proposed with variable efficiency and complexity to overcome the basic drawback of induced delays like Go back N and Selective Repeat etc. Later on, combining fixed throughput benefit of FEC and conservative bandwidth utilization of ARQ resulted in Hybrid ARQ (HARQ) [15] for more effective video streaming. Researchers have exploited the benefits of HARQ for wireless video streaming for the first time in [16] and then numerous techniques have experimented with in this combination. Another category of post detection measure is error concealment service, these are meant to hide the error data within the affected video frame through spatial or temporal interpolation [17]. In spatial interpolation, a reference frame data is recovered from the neighbouring pixels whereas in temporal interpolation inter-coded frames data is recovered. This approach has an advantage over other error control methods that it does not require additional data information to recover the lost data since it takes benefit from the repeated pattern in the successive frames and replace the erasure (error pattern) with the information in previous accurately received frame [18].

### 3.2 Accessing Corrupted Network Packets:

Most of the video streaming applications deploy user datagram protocol (UDP) for end-end transmission at transport layer as it is a simple transport protocol with minimized processing overhead due to lack of congestion control, reliable and in order content delivery. A UDP packet header needs only source and destination port identifiers, length of packet and an error checksum field to monitor packets for channel induced bit errors. A UDP checksum (computed for packet header plus payload) simply discards a packet upon getting wrong checksum that may be a consequence of single bit error due to PHY layer of WLAN. Mathematically, the Packet Error Rate (PER) can be approximated in equation 3.1 as:

\[
PER = 1 - (1 - p)^m \approx mp
\]

where \( m \) represents the number of frames in a packet and \( p \) represents the residue frame error rate (FER) after channel coding and retransmission. PER is directly related to packet length and FER, for instance a typical scenario of 1% FER and ten frames per packet would results in a 10% PER. Moreover, with increase in PER, number of packet discard grows as well. This hard rule of packet drop increases the rate of discarded packets for a typical wireless channel with error bursts due to variations in channel frequency spectrum, interference with
route obstacles etc. This all tends to throttle the throughput that adversely affects the ultimate video quality. Contrarily, this does not suit well to the real-time applications, preferring errors in the payload to the loss of whole packets [19]. It has been noticed that a packet with a few bit errors appears as a short interval glitch in the reconstructed audio/video, whereas a lost packet appears as an annoying pause for audio or a noticeable disturbance for video [20].

A boundary solution may lie in completely de-activating the UDP checksum, but it may lead to disastrous consequences. At least header of the packet must be evaluated for bit alterations to assure accuracy of destination address. Moreover, UDP checksum is mandatory for IP V6 protocols where IP checksum has been disabled. Depending upon the sensitivity of video data lost in unnecessary packet discard, number of variations to the basic UDP have been proposed like UDP-Lite, Reliable UDP and Complete UDP [19, 21-22].

3.2.1 Reliable-UDP

A UDP variant was proposed as RUDP, layered on the UDP/IP Protocols to provide reliable, in-order delivery facilitated with a maximum number of retransmissions for virtual connections. Flexibility of UDP design has tried to be preserved in RUDP that makes it suitable for number of transport purposes, specifically to transport telecommunication signaling protocols.

3.2.2 UDP-Lite

One remarkable contribution in this scenario is UDP-Lite that was based on traditional UDP with increased flexibility and preservation of existing low overhead services. They have proposed a midway of error verification through partial checksum of UDP packet after fragmentation. A UDP packet is divided into sensitive and insensitive parts in a way that a corrupted packet is discarded only if it has errors in sensitive part, while packets with errors in insensitive part are saved. A UDP-Lite receiver just needs to evaluate the sensitive data contents against the induced errors, that results in less computation overhead. Length of sensitive data bytes are specified by the source application in accordance to the sensitivity of contents to channel distortion, defined through a coverage window containing bits from the start of the packet up to the end of some limit containing significant data as shown in Figure 3.1. In addition to this error length, UDP-Lite header and IP pseudo-headers are always monitored for induced bit errors to assure the validity of critical header information. This
error control approach is superior to simple UDP protocol in the sense that it tends to exploit
the error free contents of corrupted packets through preventing unnecessary packet discard. However, CRC checksum is computed at transport layer without any check on MAC CRC mechanism. Consequently, corrupted PHY frames with in an IP packet are misplaced at MAC layer. And if MAC layer is also configured to allow frame transmission without any error check then due to lack of information about the error localization, this approach fails to provide the probable performance gain. Ultimately error free contents are not utilized productively, since frame level CRC has not been taken into account.

![Figure 3.1: Header Description for UDP and UDP-Lite Segments [19]](image)

3.2.3 Complete-UDP

The approach of CUDP has re-worked the UDP in a way to localize the error information in the packets so that error may be recovered or its effect may be eliminated. They have tried to eradicate the confusion about error free frame placements within the UDP packets through marking the error location as start and end bits of the corrupted part. This error location may be represented through an erasure pattern to facilitate the application decoder to recognize it as error pattern and invoke error concealment strategies accordingly. Moreover, layers coordination facilitates the exchange of critical error information to assists in treating the packet.

3.3 Exploring Video Statistics

In MPEG video standards, video frames are ranked differently according to their usage and significance in the process of reconstruction, it protrudes out as an additional challenge in media transmission due to more sensitivity of some video parts against the channel distortion as compared to the remaining video data. On the other side, it provides a great potential in differential frame treatments in a way that significant video data is transmitted more safely on the expense of less critical video fragments. Video frames can be prioritized and be treated
differently on the basis of their contribution to overall multimedia visual quality, named as *Unequal Error Protection (UEP)* strategy.

A video categorization scheme for wireless networks has been proposed in [23]. This approach includes the frames nomination as I and P frames only without any B frame. As presented in the study, two quality levels or frames queues (wireless transmission buffers) are defined on MAC layer where high quality queue contains the I frames while P frames are subjected to low quality queue. Frames for high quality queue are assigned higher values for re-transmission tries while low quality queue data is re-transmitted less frequently. Proposed algorithm was simulated that showed encouraging results as prioritizing the I frames greatly improved the media quality through minimizing the errors in these pictures, further avoiding the re-transmission of less sensitive data reduced the viewer waiting time as well. This idea proved to be productive because network resources were utilized for more sensitive video contents along with selective usage of insensitive data. This paves a way for the work based on exploitation of critical video contents that pays much upon decoding through saving the resources consumption (bandwidth and computation cost) as well as improves the media quality. Our proposed approach is based on UEP strategy too, where critical video data is to be highlighted in order to attain classified treatment.

Another frame discriminating approach involves video adaptation depending upon channel rates and treat the video frames according to their significance [24]. They have focused on application level, where a 'cognitive layer' has been proposed to discriminate more significant video objects from the lesser one to apply unequal error protection strategy. But they mainly have targeted the issues of video streaming for wired networks.

In [25], a GOP frames prioritization approach has been proposed where video frames at receiver MAC are treated according to their significance and impact on subsequent frames. Similarity of this approach to our proposed idea is to signify the I frames, these pictures errors exhibit a long lasting effect on subsequent frames of the GOP. But the difference lies in approach to safe guard these pictures, as we primarily tend to embed the FEC-LDPC codes in I frames at the application level and then design the network layer stack for the purpose. Secondly they proposed selective frame drop strategies in different error scenario. For instance assigning a priority check on various video frames and dropping the subsequent frames of a GOP without any further re-transmission, if its I frame is corrupted. Consequently, a whole GOP is missed out in this way. Similarly, late arriving insignificant
video data like P and B frames are dropped to overcome transmission delays. Contrarily, our approach leaves the decision of packet discard up to the application decoder. This not only tends to raise the throughput, but also eliminates the visual gaps in video play out.

Reconstructed media quality is one of the critical factors in analysing the productivity of a streaming application. Researchers have proposed to judge the media quality through its visual perception [26], which is the major essence of our proposed idea. Our approach is based on the fact that for majority of real time media applications, it all depends upon the satisfaction of end user who wants to have a pleasant visual experience without long delays, frame gaps attributed to losses and jerky motion due to channel induced errors. A decoded video should be visually acceptable, no matter what objective analysis yields. Researchers have emphasized on subjective evaluation of reconstructed video for quality improvement or distortion through modelling the HVS [65], since objective measures do not take into account the perceptual properties of HVS. Since objective (de-facto standard PSNR) and subjective quality evaluation methods exhibit very limited correlation [57]. Objective quality evaluation fails to cover all the aspects of quality assessment and requires an in depth analysis to pin point each and every change in pixel intensities and perceptual effect [58-61].

Numerous valuable surveys have been conducted to enlist the Perceptual Visual Quality Metrics (PVQMs) proposed with their significance and applicability highlighted in multimedia domain [62-64]. These surveys have also been presented a comparison among the available objective quality measures and their correlation with subjective quality measures, as shown in Table 4.1.

<table>
<thead>
<tr>
<th>Quality Metric</th>
<th>Mathematical Complexity</th>
<th>Correlation with Subjective Methods</th>
<th>Accessibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>Simple</td>
<td>Poor</td>
<td>Easy</td>
</tr>
<tr>
<td>MPQM</td>
<td>Complex</td>
<td>Varying</td>
<td>Not Available</td>
</tr>
<tr>
<td>VQM</td>
<td>Very Complex</td>
<td>Good</td>
<td>Not Available</td>
</tr>
<tr>
<td>SSIM</td>
<td>Complex</td>
<td>Fairly Good</td>
<td>Available (MATLAB)</td>
</tr>
<tr>
<td>NQM</td>
<td>Complex</td>
<td>Unknown</td>
<td>Not Available</td>
</tr>
</tbody>
</table>

This has revealed the suitability of SSIM for quality perception of media contents with optimal correlation with subjective assessment methods.
An approach based on a performance analysis of robust streaming of MPEG-2 video over IEEE 802.11g in a frequency selective channel has made use of a perceptual quality measure through Perceptual Quality Metric (PQM) [26]. They have presented that PQM can better estimate the effect of channel impairments on transmitted media from the end user perspective than the traditional Bit Error Rate (BER) and Packet Error Rate (PER) metrics. They assessed system performance both as a function of signal to noise ratio (SNR) and range between the WLAN Access Point (AP) and the station (STA). This approach is more appealing attributed to the visual aspect of video due to its emphasis on visual perception of multimedia. However, BER and PER are major channel effects estimators that can never be neglected in simulating real world phenomenon, secondly this approach is more towards lower bound of inter-distance of WLAN AP and a work station which is practically not the case all the time.

PER tends to badly affect the multimedia quality and higher Packet Loss Rates (PLR) result in disastrous consequences. However, an approach proposed to exploit the packet loss correlation [27]. They have presented a performance model that takes into account the spatial and temporal packet loss characteristics. It has been shown that packet loss at various network nodes show a correlation that can be exploited to better estimate the loss effect and quality measures. However, the approach has not been monitored for the real world multimedia trends and needs to be evaluated with significant details.

3.4 Cross layer Solutions

Besides the error control streaming strategy, we have worked with ISO stack layer parameters to explore the hidden switches that can be used to enhance the media quality and throughput. So in this section, layer coordination schemes present in the literature have been discussed along with the comparison with our proposed parameters set to highlight similarities and differences.

Idea of layer isolation has proven to be very efficient and productive for traditional wired networks that mainly incorporate data applications. However, these stereotypic independent layer services are not suitable to meet the performance challenges of wireless media streaming. With the availability of broadband network connections and proliferation of delay and bandwidth constrained multimedia applications, an immense need for layer coordination mechanism has been protruded out, named as Cross Layering (CL). Research advances have been made to facilitate the coordination among ISO stack layers for exchanging information.
about the real time applications constraints and critical data contents involved. Numerous techniques are present in the literature to cater for various performance challenges of multimedia streaming over wireless networks, implemented through network layers coordination. Well-structured surveys with valuable information have been presented based on different aspects of efficient multimedia streaming over wireless networks against each layer of network stack as well cross layer solutions [28-31]. In these surveys, authors describe the performance challenges of a wireless video sensor network and assess the specification of each layer in order to construct a cross layer treatment for sensor network contents. Similarly, an evaluation of the individual layer protocol and schemes as well as joint optimization in a cross layer scenario has been presented. These surveys are very useful with respect to layer parameters that currently deployed protocols deal with. They highlighted the limitations of prevalent techniques as well as paved the way for future research. They notified the significance of layers coordination and quality gains achieved through deploying cross layer framework. Moreover, implementation principles derived through design challenges and performance constraints of radio communication systems are explained.

AdaptNet solution has been proposed about a decade ago, envisioning the advancements in broadband wireless connections, heterogeneity of wireless networks and large scale deployment of highly constrained multimedia applications [32]. Akyildiz, et al. have proposed a cross layer solution in terms of re-design of application, MAC and transport layers of the network stack in this scheme. Basic idea is to adapt the services of these layers according to the network environment with changing parameters like transmission rate, error trend and available channel bandwidth.

Mainly, wireless video streaming networks face two performance challenges: (1) Maintaining the encoder buffer at application layer and (2) Safeguarding the video contents against erroneous wireless channel through applying some reliable error resiliency technique. These two factors are inter related, since error resilience is generally provided through adding redundant bits in the transmission contents that result in disturbance of decoder buffer at receiving end. Solutions proposed include the video bit stream conversion to address the issues; called video transcoding. It is an application layer mechanism that involves re-compressing of video stream depending upon the factors like available channel bandwidth, BER of wireless link, desired media resolution and transmission latency etc. Number of transcoding techniques are successfully prevalent like varying the video bit rate after sensing the channel conditions. To mitigate the effects of error prone wireless link, video bit rate is
reduced so that to safeguard the ultimate media quality and to minimize the probable video losses. Video bit rate can be reduced through increasing the quantization step that tends to cover up the unnecessary video details and small sized bit stream is consequent. Some other techniques have proposed to transmit constant bit rate video (though highly susceptible to channel errors) and maintain the encoder buffer through altering the decoder buffer [33]. These are application level remedies to control the channel distortion of a wireless network up to some extent. However in real time video applications deployed in error prone wireless channels, it directly tends to increase the computation complexity as well as signaling overhead for very frequent channel sensing and video re-encoding.

Some other contributions involve video transcoding techniques through collaboration of application and MAC layers for controlling video transmission rates and hybrid error resilience techniques respectively. A joint control of application and MAC layers has been proposed through involving a Cross Layer Module (CLM) that interacts and controls the application and MAC layers through assessing the channel conditions, maintaining the encoder buffer, setting the transcoder parameters and configuring the FEC/ARQ parameters [34]. This approach is supported through an increase of 3 dB in quality but at the cost of extra packet processing time. So, it suggests a trade-off between media quality and processing time depending upon a video application performance constraints. Similarity to our idea is the utilization of the potentials available in layer statistics (like header unused bits) that can be incorporated to attain higher gains in video streaming calls. However, novelty of our approach lies in the bit demarcation in packet headers at each layer to get optimized treatment.

[35] depicts the collaboration of APP and PHY for IEEE 802.11 standards through joint source channel coding. They have proposed the usage of frame statistics captured at application levels to adapt the transmission rates at PHY layer, particularly to amplify the robustness of wireless video. Similarly, authors have enforced the deployment of cross-layer mechanism for wireless communication, shown through a relay network infrastructure [36]. They have defined various layer optimization scenarios depending upon different information levels about network status.

Other cross layer solutions include a recent cross layer framework proposed that includes application layer through varying quantization level at codec end, MAC layer to allocate the available time slots to different users and the physical layer to select the number of bits
utilized per channel use [37]. They have modelled a Memory In Memory Out-Time Division Multiple Access (MIMO-TDMA) system, where multiuser resource allocation is carried out through a central control. This framework is based on their previous work that was designed for multimedia transmission for a single user [38]. In [39], researchers have assessed the significant layer parameters at different network levels. They have proposed a Cross Layer Design (CLD) through joint optimization of application, MAC and physical layers. Their experimental results show trade-offs between performance gains and computation cost. Our streaming mechanism also assessed the application, MAC and physical layer performance metrics. But the factor that distinguish it from others is the way in which bits are marked in all the layer headers that facilitate the video packets to attain classified treatment. We primarily propose to disable the possible packet drops at various network layers that results in extra delay due to re-transmissions. And the proposed scenario is achieved through bit demarcation in layer headers, besides parameters assessment is a part of our proposal that depicts a survey on media control switches that are specifically tuned to attain quality benefits.

Our streaming approach presented in this article is based on utilization of the error free contents of corrupted video packets, implemented through de-activating the packet throw at various network levels. Let the video decoder at application layer decide what to do with corrupted video packets. However, disabling the error monitoring practices may yield the disastrous outcomes specifically in case of I frames and motion vector contents of P frames. Solution lies in partial check of errors for unavoidable packet parts that leads to error control scenario. Besides de-activating the packet discard, a survey on various layer parameters has been done to assure high reconstruction quality and transmission saving. Such classified streaming calls have been accommodated to attain: (a) increased throughput due to higher number of packets available for decoding (b) enhanced multimedia visual perception due to gap elimination (appears as a consequence of some frame loss) (c) efficient utilization of link bandwidth with no re-transmissions (d) reduced delays with least error checksum computations and packet re-transmissions etc.

**Summary:**

This chapter includes a survey of ideas and techniques, presented by various researchers based on different aspects of the wireless video streaming with an emphasis on channel error control techniques in particular. All the techniques analyzed, have been categorized according to the proposed streaming frame work, compared against the proposed idea with weaknesses and strength identified.
Chapter 4

Proposed Streaming Design

This chapter includes the detail discussion of proposed streaming mechanism, description of various functional blocks of system architecture and layers collaboration design with details of re-configuration schemes. An interlayer cross layer design has been used that facilitates the sharing of data with some (or all) of the layers in the OSI protocol stack [76-77] . A cross layer streaming mechanism has been proposed, where a joint collaboration of layers mainly APP, MAC and PHY layers has been worked out in order to adopt a differential error control strategy. This cross layer strategy tends to feed an intelligent agent, specialized for channel and media assessments. System block diagram exhibits the workflow among various architectural blocks to facilitate the deployment of proposed streaming scenario.

To begin the discussion of design details, first system architecture is explained that comprises of basic components of a radio communication system as in figure 4.1.

4.1 Design Architecture

Our proposed streaming mechanism has been deployed in wireless network setup through implementation of a multimedia radio communication system. Normally, a communication system comprises of a sender with source application running on it, data stream transformer (modulator), transmission channel, stream de-transformer (de-modulator) and a receiver.

![Proposed Streaming Design](image)

**Figure 4.1:** Proposed Streaming Design
Depending upon the nature of data contents and performance constraints of the source application, more design and functional components can be added in this baseline architecture, if required.

The proposed system architecture as shown in Figure 4.1 consists of various components beginning from the source to the receiving hop.

Encoder block includes the service routines of the multimedia application residing on application layer of the network stack, while WLAN PHY routines of modulation and demodulation are encapsulated in communication channel block along with the wireless link specifications. It shows the basic component of a radio communication system from source to destination ends including the channel connecting the both ends, and the way how the media contents flow through it. Beginning from the sender end, a video encoder encodes the video bit stream in accordance with the Sender Rate Adaptor (SRA) at application layer. Ultimate compressed bit stream is passed to lower layers, after applying FEC (convolutional) encoding and modulation at PHY layer, bit stream is packetized and sent on transmission channel. After passing through the error prone radio channel, video contents are demodulated and FEC decoding is applied to eliminate as much bit errors as possible. Afterwards, it is sent to receiver MAC where CRC evaluation leads to MAC frame identification as corrupted or non-corrupted frames estimated as Frame Error Rate (FER). Corrupted MAC frames are re-transmitted through ARQ.

**Figure 4.2 System Block Diagram**
At receiver MAC, a logical module named as Intelligent Link Estimator (ILE) works to estimate the current error state of the transmission link through assessing the significant layer parameters as BER from PHY, FER from MAC and video rating information form the APP layer, where QoE based analysis of reconstructed video exhibits its suitability for visual perception. Figure 4.2 further elaborates the individual functional blocks at various stages of transmission flow, significant service blocks on sender end are video encoder, packetizer and SRA. ILE as one of the service blocks on receiver end assesses the link and video statistics through joint collaboration of top APP layer and two WLAN lower layers. APP layer video assessment is based on objective, perceptual and subjective quality parameters. Based on the combined analysis of APP, PHY and MAC layer parameters, ILE triggers a feedback event on sender side using MAC control frames through a back channel. This stimulates the SRA at sender APP to initiate a differential error control strategy depending upon the available link capacity. Previously sent video QoE based analysis on receiver end and current channel error state information guides the sender APP layer to re-encode the video contents with different compression statistics (quantization scale Q) as well as to deploy a suitable error remedial scheme, most appropriate for the current distortion level.

2 bytes long 'Frame Control' field in MAC frame header drills down to 'Type' and 'sub-type' fields that are used to identify the type of MAC frames or nature of contents it carries. 'Type'-00 indicates a frame to be control frame with a definite 'sub-type' bit pattern to invoke specialized services for that frame category. An unused bits reservation here can be exploited to implement the feedback mechanism from receiver to SRA on sender side in order to modify the video bit rate.

The rate adaptation may protrude out in one of four possibilities as transmission classes as Low Quantization Video (LQV), All Error Frame Forwarding (A-EFF), P and B Error Frame Forwarding (PB-EFF) and High Quantization Video (HQV) specified by bit patterns as in Table 4.1.

| Table 4.1: Adaptation strategy with respect to MAC control frame bits specification |
|---------------------------------|-----------------|-----------------|
| Transmission Class | Bits Specification |
| | Type | Sub-type |
| LQV | 00 | 0000 |
| A-EFF | 00 | 0001 |
| PB-EFF | 00 | 0010 |
| HQV | 00 | 0011 |
4.2 ILE Interaction Design

ILE is an intelligent module, specialized in assessing the link and video statistics through joint collaboration of APP, MAC and PHY layers of protocol stack. Such channel estimation takes into account the two crucial factors involved in multimedia streaming that are viewer satisfaction and guarantee of minimal bandwidth share to transmit the sufficient video bit rate to assure the better QoE in available link capacity. Cross layer approaches present in the literature rely on these factors but to the best of my knowledge no streaming strategy has been found that has taken combined benefits of both the video statistics and channel error state. In the proposed design an intelligent module assesses the channel impairment level through two layer parameters BER estimated from channel error and signal strength at PHY layer and FER estimated from number of MAC frames that fail the CRC check at the MAC layer. Video statistics are measured after video contents are decoded at APP layer and subjected to combined analysis of objective, perceptual and subjective quality assessment methods.

ILE takes these three input parameters (Figure 4.3), that are categorical in nature so underneath classification has been carried out through an entropy based class, Decision Trees (DT).

![Figure 4.3: ILE Interaction Design](image)

Decision trees (DT) perform classification by dividing the data into different partitions on the bases of rules defined at each node of tree. The final decision is taken by taking a class label to test. In this framework, a data set is classified by sequentially subdividing it according to the decision framework defined by the tree, and a class label is assigned to each observation according to the leaf node into which the observation falls. As compared to other supervised classifiers, DT works well when the data is in form of different categories. Another benefit of decision trees is that they are nonparametric and they do not require any prior assumption regarding data distribution. DTs are also capable of handling nonlinear In addition, they
handle nonlinear relations between features and classes, allow for missing values, and are capable of handling both numeric and categorical inputs in a natural fashion.

All the input parameters (Table 4.2) exhibit a definite range of values with precise boundaries for each category, for instance FER shows the state zero for 50-100% MAC frames with wrong CRC. This category is inter-linked with 'Bad (0)' channel state. Similarly channel seems good when 2-10 % MAC frames show wrong CRC only, and vice versa. Subjective methods show five categories of visual perception that are being converted to three states for ease in state manipulation.

<table>
<thead>
<tr>
<th>Video rating</th>
<th>Channel state</th>
<th>FER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperceptible (0)</td>
<td>Bad (0)</td>
<td>50 -100 (0)</td>
</tr>
<tr>
<td>Satisfactory (1)</td>
<td>Medium (1)</td>
<td>10 -50 (1)</td>
</tr>
<tr>
<td>Good (2)</td>
<td>Good (2)</td>
<td>2 -10 (2)</td>
</tr>
</tbody>
</table>

4.3 Layer Coordination Design

Media streaming over an error prone wireless link demands collaboration among independent layers parameters and services. Proposed streaming design is based on joint collaboration of APP, PHY and MAC layers on receiver end in order to facilitate the graceful adaptation of video bit rate on sender side, through intelligent link and video statistics assessment.

On sender side, rate adaptation is carried out through re-encoding processes in terms of HQV and LQV as well as all the layers are re-configured to deploy an efficient error control strategy of erroneous frames forwarding. For A-EFF and PB-EFF transmission class, layer re-configuration is carried out through exploiting the unused bits in header of each layer that helps to notify the flow of video contents. Figure 4.4 depicts the layer coordination design on receiver end, where ILE interaction at various points is elaborated.
4.4 Layer Re-configuration

Traditionally, as a data packet passes through the network stack, each layer associates certain information called 'layer header' with the packet for its peer layer to treat the packet accordingly [40]. Conversely, peer layer removes the header and forward the packet to the next layer that does the same. A layer recognizes the services and protocols of its peer layer through the information encapsulated in concerned header in the form of designated bit patterns or flags marked to present a typical data scenario. These flags are normally used to control various transmission and error parameters like sender transmission rate, receiver buffer capacity, data packet priority and security levels etc. In error frame forwarding strategy of error control, these bit patterns have been assessed and utilized at each network layer to make peer layers aware of the sensitivity and usability of the video contents. A protocol stack has been proposed with respect to the header bit demarcation particularly on network, MAC and PHY layers to notify the video transmission. However, we have tried to re-configure each layer of the stack to tackle the network errors and optimize the productivity of a communication system.

4.4.1 Application Layer

a. Compression- After getting the feedback of a bad channel, video quantization scale is increased to reduce the size of compressed bit stream. This may result in quality degradation but helps in adapting for the reduced channel throughput. The same process is reversed for a
feedback of good channel where videos are highly quantized to display the better quality and
finer details in video frames with ultimate high bit rate video.

For EFF error control strategy, video contents are compressed at higher quantization scale to
prevent any quality degradation while encoding. However, no error control strategy is being
implemented through its journey from sender to receiver hop except the FEC encoding at
sender PHY that is a part of IEEE 802.11 WLAN standard. Each next layer is notified about
the video contents of incoming packets through setting the header bits in order to treat the
packets differently. Intra-coded frames of the video stream play a vital role in subsequent
frames re-generation, as a corrupted I frame tends to affect the whole GOP reconstitution. In
PB-EFF class, I frame contents are being recovered from the channel error through ARQ at
receiver MAC where an alternate option is using the Low Density Parity Check (LDPC)
codes that are standardized for high speed wireless transmission. LDPC codes are a re-
discovered class of linear error control codes that use parity check matrix at the encoding and
decoding ends and process long block lengths efficiently [41]. This parity check matrix
depicts relationships between source symbols, later used by the receiver to reconstruct the
original symbols even if some get erroneous. However, this may induce data redundancy and
does not seem a preferable approach in low throughput channel.

4.4.2 Transport Layer
a. UDP Adaptation- Most of the video streaming applications deploy user datagram protocol
(UDP) for end-end transmission at transport layer as it is a simple transport protocol with
minimized overhead due to lack of congestion control, reliable and in order content delivery.
UDP header includes source and destination port identifiers, length of packet and an error
checksum field to monitor packets for channel induced bit errors. A UDP checksum simply
discards a packet upon getting wrong checksum that may be a consequence of single bit error
due to PHY layer of WLAN. This hard rule increases the rate of packet discard for a typical
wireless channel with error bursts due to variations in channel frequency spectrum,
interference with route obstacles etc. This all tends to throttle the throughput that affects the
ultimate video quality.

One solution is to completely de-activate the UDP checksum, but it can lead to disastrous
consequences. At least header of the packet must be evaluated for bit alterations to assure
accuracy of destination addresses. Moreover, UDP checksum is mandatory for IP V6
protocols. One remarkable contribution in this scenario is UDP-Lite that incorporates partial
checksum of the UDP packet, based on fragmentation of the data segment into sensitive and
insensitive parts, a packet is discarded only if it has errors in sensitive part, defined through a ‘convergence window’ containing bits from the start of the packet up to the end of some limit containing significant data. This mechanism of UDP-Lite suits our design constraints for the transport layer since our design strategy is to check only the header of UDP packet in order to accurately identify a packet to be a video packet to treat it accordingly. The sensitive part in this case will be marked as the header of the packet that should be free of errors in any case while payload part should be forwarded unchecked to higher layers.

4.4.3 Network Layer
Network level requirements also highlight the need for a strategy to minimize packet loss particularly in case of I frames of MPEG video. One idea is to mark the Internet Protocol (IP) packets requiring special service through utilizing the bit patterns reserved in the frame header to identify the requested type of network service. Network paths vary widely as a function of type of service they provide; some exhibit a trade-off between transmission delay and throughput when shortest paths may not be reliable options for packets to reach in order to enhance the net throughput. It is the choice of sender application to follow an 'optimal ' path that may provide minimum delay or maximum throughput. Bit pattern for network service treatment has been given different names, any of which can be deployed depending upon the network router configuration.

Differentiated Services (DS)
RFC-2474 [42] has defined an octet 'DS', consists of a 6-bit Differentiated Services Code Point (DSCP) and 2-bit Currently Unused (CU) fields. DSCP field is marked to identify the Per Hop Behaviour (PHB) i.e. how each router will treat the respective network packet. PHB may be divided into certain packet forwarding classes e.g. Default Forwarding (DF), Assured Forwarding (AF) and Expedited Forwarding (EF). We mark the packets to select the EF behaviour for low-loss, low-delay, and low-jitter services. Two CU bits are further marked for our I-frames to prevent maximum possible packet loss.

4.4.4 Data Link Layer
Modified IEEE 802.11 Frame Header
IEEE 802.11 wireless LAN standard has well-defined specifications for MAC and physical layers. It describes a discrete set of attainable MAC services with header of MAC frame describing the fields that identify services a frame must be provided with as shown in Figure
4.5. Isolating two bytes of 'Frame Control' field paves a way to discriminate a frame as a video frame with specific bit pattern to define the required set of services. These two bytes are further re-arranged into 11 fields as standard Version, Type, Subtype and Miscellaneous (as To DS, From DS etc).

Figure 4.5: IEEE 802.11 MAC Frame Header [43]

Two bits of 'Type' field differentiate a frame as management, control or data frame while 'Subtype' defines the particular type in each specified frame class. Where the field 'Type' bit combinations differentiate a MAC frame as in Table 4.3.

<table>
<thead>
<tr>
<th>'Type' bits</th>
<th>MAC frame type</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>Management frame</td>
</tr>
<tr>
<td>01</td>
<td>Control frame</td>
</tr>
<tr>
<td>10</td>
<td>Data frame</td>
</tr>
</tbody>
</table>

Here an unused bit combination for these fields can be exploited for the video streaming as bit pattern '101000' could well indicate a frame containing video data to attain specialized service upon reaching the receiver end. Usually when a frame reaches the MAC layer on receiver end, the foremost step is to analyze it for channel induced errors through CRC and it is discarded if found corrupted. Modification proposed to this scenario is to propose the bit pattern for fields 'Type' and 'sub-type' such that to mark the packet as IEEE 802.11 video frames [43]. Such video frame recognition eliminates the CRC computation and allows direct forwarding of the frame to higher layers.

Figure 4.6 elaborates how the proposed 802.11 frame routes to receiver end and recognized as a video frame.
4.4.5 Physical Layer

IEEE 802.11 PHY Frame [44]- PHY layer of IEEE 802.11 WLAN deals with issues of reflection, external interference, bandwidth allocation etc. For instance, it tends to pass the transmitted data through FEC encoder to combat the effect of erroneous wireless channel. WLAN PHY packet header has been designed to cover all these service specifications as shown in Figure 4.7. It highlights the bit demarcation as initial 18 bytes make up Physical Layer Convergence Procedure (PLCP) preamble, followed by 3 bytes of PLCP header field. PLCP preamble bits are used for transmission synchronization purposes while PLCP header encapsulates the significant details about length of MAC PDU, transmission data rates, CRC etc. In the PLCP header field, 8-bits of ‘Service ’ field are reserved for future use that can be incorporated to notify PHY layer on receiving end about an incoming packet being a video packet. Moreover, these bits can further be drilled down to mark the critical data (like I frames) in a video packet, later used to compensate P and B frames that is meant to indicate the sensitivity of these frames against channel errors and its long term consequences. Idea proposed here is to split the 'Service' octet into two parts, lower 6 bits are named as 'data type' and assigned the bit pattern as '100000' to differentiate a packet containing video data which should be subjected to partial checksum in a way that header should only be monitored for bit errors and payload part should be forwarded unchecked to higher layers. Remaining two bits of the octet are proposed to specify the I frame data as bit combination '11' defines an I frame and '10' marks the non-I frame data to indicate information sensitivity against channel errors. All IEEE 802.11 PHY packets are fed to FEC encoder for error concealment purposes. FEC
decoder at PHY layer of receiving host removes the redundant bits and video pictures actually sent are recovered.

Figure 4.7: IEEE 802.11 PHY Packet [45]

Figure 4.8 shows the pictorial representation of the all the layer headers with their respective bit patterns proposed to differentiate the streaming calls from traditional data transmission.

Figure 4.8: Bit demarcation in video packet at various layers

The purpose of marking the header bits is to make network layers aware of video contents of the packet so that to save the extra error processing overhead and to avoid consequent delay due to error correcting services like re-transmissions. However, these bits are exclusively marked for transport, MAC and PHY layers in order to avoid the probable packet drop. Other parameter strategies are just the part of video application residing on the application layer of the network stack. Though such bits processing has been proposed as a new computation in the stack but on the other hand it saves prevalent computation of the error processing overhead. Additionally, link saving is another benefit of the proposed error control strategy. This all pays for the enhancement in end user visual experience where gap due to packet losses is eliminated through making the partly corrupted frames available on receiver end. Practically, this streaming strategy can be applied for different real time streaming scenarios of delay and throughput sensitive applications for instance, in video surveillance a viewer wants an un interrupted flow of information in a way for the video frames joining end to end
without any gap or delay. Similarly, in video streaming where video contents are directly transmitted from a streaming server, most of the time a viewer desires a quick availability of a video to decide about its need. In both of the cases, streaming throughput and delay should be minimized to meet the video application constraints.

Summary:

Proposed streaming design details have been discussed in this chapter, particularly the error control strategy has been elaborated through its implementation at various network stack layers and the respective header bits involved. ILE as a basic design logic in bit rate adaptation has been explained in details along with its input and output parameters interaction involved. System architecture has been explained with different services blocks and resultant video signal formats. Experimentation data set and various streaming test scenarios with the rationale behind them are explained in the next chapter.
Chapter 5

Results & Discussion

In order to assess the effectiveness of proposed error control scheme of wireless video streaming, a streaming scenario has been simulated along with the proposed error control strategy implemented. Extensive experimentation scenarios have been established involving various GOP structures and video mobility levels, ultimate media quality has been assessed based on mathematical, perceptual and subjective parameters. Moreover, link resource usage trend has been quantified that reflects the remarkable saving in terms of efficient usage of channel throughput.

5.1 Data set

Data set used in experimentation has been selected with at most care to cover a vast range of video statistic, particularly in terms of background and foreground frame contents. Based on this statistics, three video clip categories have been chosen as Low Mobility (LM), Medium Mobility (MM) and High Mobility (HM) as shown in Table 5.1.

These all video samples are available online [48], and their mobility levels have been marked and used as such in the literature in various proposed streaming scenarios.

<table>
<thead>
<tr>
<th>Video samples with different mobility levels</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low Mobility</strong></td>
</tr>
<tr>
<td>(LM)</td>
</tr>
<tr>
<td>Mother-daughter</td>
</tr>
<tr>
<td>Container</td>
</tr>
</tbody>
</table>

Table 5.1
5.2 GOP Structure

Multimedia throughput depends upon number of other factors than the channel bandwidth alone, like GOP structure etc. As explained earlier, MPEG video is a sequence of GOPs where a GOP is a combination of I-, B- and P-frames. GOP combination is a critical factor since it derives the effects of packet delay or even loss on multimedia throughput. For instance, considering the sequence of frames like:

\[ I_1 \ B_2 \ B_3 \ P_4 \ B_5 \ B_6 \ P_7 \ B_8 \ B_9 \ I_{10} \]

All the B-frames are dependent upon the prior and subsequent I- and P-frames while P-frames rely on prior I- and P-frames only. For instance, considering the above GOP structure, errors induced in P_4 frames will propagate in P_7 frame and all the B-frames of GOP, such propagation exacerbates the visual quality of subsequent frames later while decoding. Loss of any B-frame does not have serious consequences and can be replaced by an adjacent P- or B-frame, while P-frame loss affects dependent P- and B-frames. Loss of an I-frame results in whole GOP failure and quality degrades abruptly. A cunning solution to impede the error propagation among P-frames is to encode the whole video through JPEG, called motion-JPEG where all the video frames are intra coded without using any other frame type.

In motion-JPEG, each of the video frames stands alone as the independent picture and error induced in one frame retains to it. Ultimately visual quality is greatly enhanced but on the expense of excessive channel bandwidth utilized to transmit whole frame data rather than just the motion vectors and frame residue. However, for a video file with rich multimedia and high motion contents, it does not seem feasible. A middle way may be to send reference information too often, avoid bidirectional estimation (B frames) and reduce GOP size to half as:

\[ I_1 \ P_2 \ P_3 \ P_4 \ P_5 \ I_6 \]

Though error still propagates due to inter-frame motion estimation among P-frame but visual quality becomes smooth on getting next I-frame which is shielded against channel errors. Moreover, keeping GOP as small as possible is economical as well as renders the visual perception of the video fine [1]. Figure 5.1(a-c) is an illustration of various GOP structures and their contribution towards multimedia quality.
In this experiment, 'Mother_Daughter', Container and 'Tennis' video sequences have been selected in order to attain the GOP style analysis for LM, MM and HM video respectively. A video sequence is compressed to yield encoded video stream, fed to three path Rayleigh rural fading channel with sampling period of 100 K seconds and Doppler shift of 200 Hz. Each of the video samples has been encoded with all three GOP structures as discussed above. Results show that Motion-JPEG outperforms the other two GOP structures but on the expense of higher channel bandwidth consumed since all the pictures are intra-coded, hence channel errors are localized to infected frames only. Besides, HM video samples that exhibit degraded reconstructed quality shown through PSNR curve for 'tennis'. However, on assessing frames with some perceptual quality tool, it reveals that visual quality is acceptable for video samples with high mobility too. GOP structure with combination of IPPPPI frames show remarkable results with link saving as well.

5.3 Media Quality Analysis

The proposed error control strategy is mainly based on viewer QoE analysis, whereas reconstructed media quality has also been analysed for the de-facto quality parameters like Peak Signal to Noise Ratio (PSNR) and Mean Squared Error (MSE). Moreover, structural variations due to channel distortion have also been quantified through structural quality estimates of Structural SIMilarity (SSIM) indexing and its variants.
5.3.1 Objective and Perceptual Quality Analysis

MSU VQMT has been used as a visual perceptual tool for analyzing reconstructed video clips through numerous quality parameters. VQMT is an objective quality measurement tool, as well as equipped with subjective measurements metrics to facilitate more precise quality estimation of a media file [47]. MSU VQMT computes a metric estimation through comparison of a media file to its original version that serves as a reference, moreover it provides an opportunity of comparative analysis of two processed files with one original version. Figure 5.2 presents input and output collaboration of VQMT, along with the results illustration it provides, where 'original' video file is the one that acts as a reference for files being processed for quality estimation indicated as 'reference 1' and 'reference 2'. 'video with mask' is an optional functionality that assists in selective region estimation in a video frame. MSU VQMT generates the results in the form of comma separated files (.csv files) with metric estimates for each frame of each file processed (shown as 'per-frame values'), with averaged values for each metric computed. It draws plots to visualize the metric filter results and corrupted frames that are saved as bitmap images having maximum difference from the original or source frames named as 'bad frames'. Besides, it also creates an Audio Video Interleaved (AVI) file for visualization, showing luminance and chrominance values of the estimated metric for each frame of the video file.

![Figure 5.2: VQMT input and output](image-url)
**MSU VQMT Metrics:**

It offers a complete list of parameters for in depth estimation of video premises for objective as well as subjective analysis of quality degradation. We have used objective metrics as PSNR, MSE and subjective analysis of media files has been carried out through frame structure metrics like SSIM, SSIM variants including MS-SSIM, 3 Component-SSIM, VQM and metrics for noise estimation etc.

a. **Mean Squared Error (MSE)**

MSE defines the squared mean error induced in the video frame, it provides a numeric estimate of the difference of each pixel of the corrupted frame from the pixel of original frame. For instance for two MxN video frames f and f' where former is original frame and later is the reconstructed version or approximation of the original, MSE is defined in equation 5.1 as:

$$MSE = \frac{1}{MN} \left( \sum \sum \left[ f'(x,y) - f(x,y) \right] \right)$$  \hspace{1cm} (5.1)

MSE limitation lies in the fact that it gives a numeric estimate and does not take into account the biological aspects of the human vision system that other subjective estimation parameters consider like SSIM and VQM.

b. **Power Signal to Noise Ratio (PSNR)**

PSNR is an objective quality metric, used to measure the quality of reconstructed video frame. It provides an empirical or quantitative measure of the media quality since it describes the ratio of originally sent signal to the power of distorting noise that affects the quality. Higher values indicate better quality. It is computed as a logarithmic trends for a video frame because pixels show a wide range of values for this estimate.

PSNR is mathematically related to MSE in equation 5.2 as:

$$PSNR = 10 \log_{10} \left( 255 \ast \frac{255}{MSE} \right)$$  \hspace{1cm} (5.2)

Though PSNR and MSE are de-facto standards [75], widely used to quantify the media quality and can easily computed for the given video sample as Full Reference (FR) quality metrics but these metrics do not show sufficient correlation
to HVS [70]. It urges for some quality tool, characterized by HVS to facilitate subjective assessment.

c. **Structural SIMilarity (SSIM)**

SSIM index is a measure of resemblance between two images, as one image is compared against the other perfect quality (reference) image in three aspects as contrast, luminance and structural similarity [71]. SSIM is an absolute quality measure since different images with same MSE (distortion) values from the original image exhibit different perceptual quality; they actually differ on SSIM index with distinct values. SSIM index values ranges from -1 to 1. Higher the values, better is the image quality and more image resemblance, for instance, SSIM value 1 shows equal quality frames. Though involves complex computations but this metric is comparatively more close to the human vision system than PSNR. More brighten region in the SSIM image shows greater difference in two images in comparison.

d. **Multi-Scale Structural SIMilarity (MS-SSIM)**

Multi-Scale SSIM indexing is based on SSIM metric in a way to provide several downscaled levels of original images [72], result is computed as weighted average of all level metrics. Perception of image details mainly depends upon the sampling density of the image signal, inter-distance of the image plane and observer, and the perceptual capability of the observer visual system. Ultimately, subjective evaluation of a given image varies as these factors change. This metric has proven to be more effective approach as compared to single scale SSIM as well as outperforms other Video Quality Assessment (VQA) algorithms through incorporating the variations of image resolution and viewing conditions.

e. **3 Components-Structural SIMilarity (3-SSIM)**

3-Component SSIM Index based on region division of source frames, since there are 3 types of regions as edges, textures and smooth regions [73]. Rationale behind this metric involves a significant property of HVS as human eye can see difference on textured or edge regions precisely than on smooth regions. Result metric computed as weighted average of SSIM metric for these specified regions.
f. **Video Quality Metric (VQM)**

VQM is based on discrete cosine transformation of the video frames [74]. It better corresponds to human perception for image quality as compared to any of the MSE-based performance metric. VQM values are always above zero and lower value symbolizes better quality. However equal quality frames have same VQM value.

g. **Noise Estimation Metric (NE)**

This metric provides a precise estimation of noise induced in an impaired video. It is a FR VQA algorithm, offering flexibility in filter design settings. It is based on three algorithms for estimation of noise level within a video frame:

- **Mean Absolute Deviation (MAD)**
- **Block-based**
  
  Video frames are fragmented into a number of 8x8 blocks, standard deviation is computed for change in intensity value for each block. A block with minimum value of deviation shows insignificant intensity change and remains smooth. The intensity variation of a smooth block may be due to noise, where computed standard deviation of the block is attributed to Gaussian noise added.

- **Spatio-Temporal Gradients**
  
  Each video frame undergoes wavelet decomposition to compute temporal and spatial histograms. An initial noise level is estimated for a value temporal or spatial histogram achieves the maximal value. Decision of temporal or spatial histogram selection is based on its deviation from the Rayleigh distribution. Then this estimation is corrected, using Kolmogorov-Smirnoff test. The normalized corrected estimation is the final value of the metric.

5.3.2 **Subjective Quality Analysis**

**MSU Perceptual Video Quality Tool**

It is an effective way for subjective quality assessment, facilitated through a user friendly Graphical User Interface (GUI) that can be decomposed into two parts PVQ-task manager and PVQ-player. PVQ-task manager assists in designing the quality test and to gather the experts results while PVQ-player is meant to offer direct user interaction for video display.
and quality ranking. A quality test specification generally includes the details about number of tests conducted, evaluation methods and results processing etc.

5.4 ILE Modelling Accuracy

As explained in the previous chapter that ILE takes three input parameters at a time as video rating, channel state and FER each with three possible states respectively. Moreover, it works on a data set of six video samples, one pair form each mobility level. So total input samples are given as:

\[3^3 \times 6 = 162\text{ input samples}\]

These are 162 videos which are generated for different input settings and each video is assigned a label out of four possible ILE output values as described in figure 4.3. The whole dataset is of categorical type and we compared decision trees with Naïve Bayes and K nearest neighbour (Knn) classifier. We divided the data into two parts i.e. training and testing. 120 samples are selected from whole data as training such that there should be equal number of samples from each class. The remaining 42 samples are left for final testing. Now in order to validate the model, cross validation is performed on training data using 3-fold cross validation technique. Once the models are trained and validated, the final testing is performed on test data. Figure 5.3 shows the confusion matrices for all three classifier which are used. Theses matrices show that decision trees achieved an accuracy of 85.71% against 59.52% and 73.81% for naïve Bayes and Knn respectively. Decision trees outperformed other two classifiers and we used decision trees in our proposed system. Here four classes are increase quantization (0), decrease quantization (1), PB-EFF (2) and All-EFF (3). The main reason of good results for DT is that there are direct decisions against different attribute values.

![](image)

**Figure 5.3:** ILE comparison result. (a) Confusion matrix for Naïve Bayes; (b) Confusion matrix for Knn; (c) Confusion matrix for decision trees
The proposed ILE model uses decision trees for video rate adaptation. Figure 5.4 shows the final tree which is generated for classification of unseen test data.

As we have used different parameters along with decision trees to generate best possible output of ILE, so we have also tested the contribution of each parameter in decision making. Table 5.2 shows the resultant accuracies against different combinations of parameters and it is apparent that individual parameter fails to offer noticeable outcomes and accuracy lies in range of 23-42%. However, a combination of all the three input parameters exhibit average accuracy of 88% in training samples whereas blind testing reveals 85% accuracy.

**Table 5.2**  ILE training and testing accuracies

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Rating (VR)</td>
<td>0.30</td>
</tr>
<tr>
<td>Channel State (CS)</td>
<td>0.42</td>
</tr>
<tr>
<td>FER</td>
<td>0.23</td>
</tr>
<tr>
<td>VR + CS</td>
<td>0.57</td>
</tr>
<tr>
<td>VR + FER</td>
<td>0.42</td>
</tr>
<tr>
<td>CS + FER</td>
<td>0.47</td>
</tr>
<tr>
<td>VR + CS + FER</td>
<td>0.88</td>
</tr>
<tr>
<td>Testing</td>
<td>0.85</td>
</tr>
</tbody>
</table>
5.5 Overhead of Feedback Mechanism:
ILE works to trigger the bit rate adaptation through involving SRA at sender APP layer. This feedback information is transmitted to sender through a back channel from receiver to sender using MAC control frames. Rate adaptation may protrude out in one of four possibilities as transmission classes, specified by a bit pair combinations as:

Table 5.3: Adaptation strategy with respect to MAC control frame bits specification

<table>
<thead>
<tr>
<th>Transmission Class</th>
<th>Bits Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>LQV</td>
<td>00</td>
</tr>
<tr>
<td>A-EFF</td>
<td>01</td>
</tr>
<tr>
<td>PB-EFF</td>
<td>10</td>
</tr>
<tr>
<td>HQV</td>
<td>11</td>
</tr>
</tbody>
</table>

5.5.1 Computation Overhead:
ILE processing mainly adds to computation cost, triggered each time on completion of one video frame data transmission at receiver. FER, link noise and quality rating for the candidate video frame directs towards the adaptation class, through parsing the input parameters to ILE DT. For consequent decision, a specific bit pair is set in the control MAC frame (as shown in Table 5.3) and fed back to sender APP which re-encodes the video to be transmitted accordingly. For current transmission class as all EFF, ILE computation for link adaptation is carried out 25 times/sec for the video encoded at a rate of 25 fps. This is so because all video frames (reference and dependent) need the quality check due to unavailability of error control service for all. However in case of LQV transmission class, ILE tends to compute the new transmission class for quality change of reference video frames only, identified through 'Service' octet sub-bit combination of PHY packet header which have been pre-set by sender PHY to notify the type of video frame transmitted. Whereas for P and B EFF class, ILE is invoked for dependent frame quality drop only because reference frame error has been mitigated through ARQ at MAC of receiver in this scenario.

5.5.2 Delay Overhead:
Consequent delay due to ILE deployment is insignificant since the proposed technique has already eliminated the processing and transmission latencies at various points of radio communication, particularly in case of EFF transmission classes. For instance, in case of all
EFF all corrupted MAC frames belonging to any video frame type are passed as such up to the application decoder at APP layer and ARQ protocol is not triggered to recover the loss or error of any frame type. This results in noticeable reduction in net transmission latency as well as consequent viewer waiting time is highly reduced. However, ILE is triggered for link adaptation each time when MAC frames of every video frame arrive at the receiver MAC. Whereas in case of PB EFF class, I frame data is recovered through ARQ protocol but results in minor delay attributed to reference information recovery only, as shown in figure 5.5 where ILE is triggered only at the arrival of P and B frames data at receiver MAC.

Figure 5.5: Play out time comparison for different streaming strategies

5.6 Performance Analysis
Effectiveness of the proposed error control strategy has been analyzed through estimating the reconstructed video quality as well as the trend of channel throughput usage. Video samples with different mobility levels are encoded and passed through the impaired radio link where MAC FER and channel BER are used to analyze the channel error state. For instance, 200 frames of 'foreman' video sample have been encoded with different GOP structures at
different bit rate and ultimately divided into number of MAC frame of 2304 bytes, as shown in table 5.4 for 1 second video data with 25 fps.

<table>
<thead>
<tr>
<th>GOP Structure</th>
<th>Bit rate (kbps)</th>
<th>FEC-Bit rate (kbps)</th>
<th>Num. of MAC frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPI</td>
<td>915</td>
<td>1830</td>
<td>100</td>
</tr>
<tr>
<td>IBP</td>
<td>1074</td>
<td>2148</td>
<td>117</td>
</tr>
<tr>
<td>M-JPEG</td>
<td>1815</td>
<td>3630</td>
<td>197</td>
</tr>
</tbody>
</table>

5.6.1 Channel Analysis

In the simulation setting, distortion level of a wireless link has been estimated through a noise parameter, Signal to Noise Ratio (SNR). It has been categorized into three whole values to reflect the distinct channel states, where floating point numerals can also be incorporated but these values exhibit channel state transitions that fail to account for the absolute distortion effect on reconstructed media files. Table 5.5 describes the effects of various channel error states on the video bit stream where each BER and number of MAC frame retries corresponding to each SNR level are shown. SNR range (1-4) dB shows a Bad Channel (B.C.) where all the transmitted MAC frames get the bit errors and re-transmission does not seem effective here too. Moving from SNR range of (5-7) dB, BER reduces gradually and apparently becomes zero at SNR 8 dB. Number of MAC frame re-transmissions are also being reduced with an improvement of BER and becomes zero at SNR 8 dB. Different GOP structure and spacing of reference frames (N) within a single GOP result in different bit stream size and ultimate different number of MAC frames re-transmitted at different channel states. For instance, Motion JPEG results in low compression and large size bit stream as compared to IBP and IPI GOP structures.

Besides, table 5.5 exhibits the effect of PHY layer FEC encoding on channel induced bit errors, as for non FEC scenarios bad channel prolongs for SNR range of (1-9) dB and finally BER becomes zero at SNR 15 dB which is almost double of FEC enabled scenarios (8 dB).
Table 5.5  Effects of various impairment levels

<table>
<thead>
<tr>
<th>GOP-IBBPBBPBBBI(FEC)</th>
<th>GOP-IBBPBBPBBBI(No-FEC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SNR</strong></td>
<td><strong>BER</strong></td>
</tr>
<tr>
<td>1-4</td>
<td>0.096</td>
</tr>
<tr>
<td>5</td>
<td>2.05 e-4</td>
</tr>
<tr>
<td>6</td>
<td>1.73 e-5</td>
</tr>
<tr>
<td>7</td>
<td>1.54 e-6</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GOP-M-JPEG(FEC)</th>
<th>GOP-M-JPEG(No-FEC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SNR</strong></td>
<td><strong>BER</strong></td>
</tr>
<tr>
<td>1-4</td>
<td>0.098</td>
</tr>
<tr>
<td>5</td>
<td>3.2 e-4</td>
</tr>
<tr>
<td>6</td>
<td>1.67 e-5</td>
</tr>
<tr>
<td>7</td>
<td>1.31 e-6</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GOP-IPPPPI(FEC)</th>
<th>GOP-IPPPPI(No-FEC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SNR</strong></td>
<td><strong>BER</strong></td>
</tr>
<tr>
<td>1-4</td>
<td>0.098</td>
</tr>
<tr>
<td>5</td>
<td>3.2 e-4</td>
</tr>
<tr>
<td>6</td>
<td>1.67 e-5</td>
</tr>
<tr>
<td>7</td>
<td>1.31 e-6</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.6.2  Media Quality Trends

As ILE directs about the available channel throughput through assessing various significant layer parameters, SRA at sender node dictates the APP layer to re-encode the video stream in order to ensure efficient link usage as well as to optimize the video quality. Ultimate reconstructed media quality is assessed through objective, perceptual and subjective methods to guarantee an optimized QoE of the target media viewer. This section of the streaming scenarios and results analysis involves the video quality assessment for all the error control strategies triggered in various channel states.

- **Error Frame Forwarding (EFF)**

   Error prone wireless link induces high level impairments in compressed bit stream transmitting through it. However, in case of video data where spatial and temporal correlation of video frames may provide a relaxation in checking and correcting channel induced bit errors. This sets a rationale to tolerate few bit errors in
compressed bit stream rather than error inducing computation and link wastage overheads which are normally triggered for each error event at MAC layer. On getting the indication of a bad channel from ILE link estimation, an A-EFF strategy is invoked which is basically a mechanism to train all the stack layers of sender and receiver nodes about de-activating the error adjustment techniques and passing the corrupted video data up to the application decoder. Next, ILE triggers the current feedback event based on the video quality assessment from APP and link estimation from PHY and MAC layers. So a comparatively better channel state may lead to an error control strategy PB-EFF where inevitable video contents like reference pictures are recovered through ARQ at receiver MAC while less critical data is passed as such to application decoder.

**PB-EFF**

Video samples with three mobility levels have been selected and a compressed bit stream is generated for all three GOP structures. Reconstructed video sequence in all cases is subjected to de-facto PSNR estimation which is later verified by structural variations in video frames due to channel distortions. Figure 5.6 accounts for the objective video quality against all the three channel states as low, medium and high BER, where low BER channel may induce a single bit error as shown through a red spot in the figure for SNR=7 dB but in most cases it exhibits zero BER for this SNR. In this experiment, 100 frames 'Foreman' video sequence is selected from MM category, encoded with GOP structure of IBP at 25 fps.

Error control strategy followed in this scenario is based on error recovery for I frames only while P and B frames are passed up to the application decoder without checking or recovering bit errors. This partly eliminates a probable quality drop due to error propagation of reference pictures in the whole GOP as the curve touches quality upper bound on getting the next I frame and then it follows the minima for subsequent P and B frames. However despite the lack of error control for P and B frames, objective quality is apparently good as even for high BER channel it is above 24 dB which is acceptable for visual perception.
Table 5.6 accounts for the visual analysis reconstructed video pictures along with the estimates of objective and subjective quality analysis. As this error control strategy is based on UEP scheme so reference pictures of the GOP exhibit the same quality level in all channel distortion levels while subsequent frames of the GOP show an improvement in objective and subjective quality as the channel error decreases. However, frames show acceptable visual quality in high BER channel too.

In assessing the effect of channel error on the video bit stream that is compressed following the IPI GOP structure with 25 fps at a bit rate of 1830 kbps after applying FEC encoding at PHY layer, objective quality trend is almost same. Here, GOP used is of small size comparatively with M=4 as I frames is repeated after every fourth P frame but there is no error control for P frames again and only I frames error is recovered which results in a quality drop for every following P frame in the GOP. This is due to error propagation in P frames because each P frame is involved in estimation of the next P frame in the GOP. However, visual perception again resides in acceptable range as shown through objective quality above 25 dB for most of the video frames even in high BER channel as elaborated in figure 5.7 and table 5.7.

Figure 5.6:  PSNR estimation for IBP GOP
### Table 5.6: IBP Quality Assessment for PB-EFF

<table>
<thead>
<tr>
<th>Frame</th>
<th>I1</th>
<th>B2</th>
<th>B3</th>
<th>P4</th>
<th>B5</th>
<th>B6</th>
<th>P7</th>
<th>B8</th>
<th>B9</th>
<th>I10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Origin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>27.03</td>
<td>29.38</td>
<td>29.35</td>
<td>26.92</td>
<td>26.50</td>
<td>30.16</td>
<td>29.25</td>
<td>27.75</td>
<td>34.95</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>28.82</td>
<td>30.93</td>
<td>30.28</td>
<td>29.26</td>
<td>28.41</td>
<td>31.66</td>
<td>31.04</td>
<td>29.29</td>
<td>34.95</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>28.82</td>
<td>31.35</td>
<td>30.28</td>
<td>29.26</td>
<td>28.41</td>
<td>27.53</td>
<td>31.78</td>
<td>26.84</td>
<td>34.95</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 5.7: IPI Quality Assessment for PB-EFF

<table>
<thead>
<tr>
<th>Frame</th>
<th>I1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>I6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>High</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>32.16</td>
<td>29.97</td>
<td>27.93</td>
<td>26.93</td>
<td>35.04</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Medium</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
<td><img src="image16" alt="Image" /></td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>32.47</td>
<td>31.19</td>
<td>30.30</td>
<td>29.17</td>
<td>35.04</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Low</td>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
<td><img src="image21" alt="Image" /></td>
<td><img src="image22" alt="Image" /></td>
<td><img src="image23" alt="Image" /></td>
<td><img src="image24" alt="Image" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>33.54</td>
<td>31.94</td>
<td>30.90</td>
<td>29.63</td>
<td>35.04</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

**Figure 5.7:** PSNR estimation for IPI GOP
In order to eliminate the error propagation in subsequent frame of a GOP, an extreme solution is to encode and decode each frame independently as in Motion JPEG scenario. A 100 frame 'foreman' video sample is encoded following M-JPEG with 25 fps, at a bit rate of 3630 kbps after applying FEC encoding at PHY layer. Reconstructed video shows an upgrade in objective quality as shown in figure 5.8 as most of the frames show objective quality above 27 dB even in high BER channel but at the expense of channel bandwidth due to low compression for all the video pictures. However, channel throughput is saved in a way to pass all the video frames data without any error control (for reference pictures too) and it still possesses acceptable quality even for high BER channel.

![PSNR vs. SNR for GOP-Motion JPEG](image)

**Figure 5.8:** PSNR estimation for M-JPEG GOP

Table 5.8 accounts for the visual perception of the reconstructed videos after passing the compressed bit stream through various impairment levels, since bit errors in reference picture are apparent visually and also shown through objective and subjective quality drop. However, abrupt quality variations are characterized by frame structural statistics otherwise it exhibits a smooth quality trend throughout the GOP.
Table 5.8: M-JPEG Quality Assessment for PB-EFF

<table>
<thead>
<tr>
<th>Frame</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin</td>
<td><img src="image1" alt="Original Frame" /></td>
<td><img src="image2" alt="Original Frame" /></td>
<td><img src="image3" alt="Original Frame" /></td>
<td><img src="image4" alt="Original Frame" /></td>
<td><img src="image5" alt="Original Frame" /></td>
<td><img src="image6" alt="Original Frame" /></td>
<td><img src="image7" alt="Original Frame" /></td>
<td><img src="image8" alt="Original Frame" /></td>
<td><img src="image9" alt="Original Frame" /></td>
<td><img src="image10" alt="Original Frame" /></td>
</tr>
<tr>
<td>High</td>
<td><img src="image11" alt="High Frame" /></td>
<td><img src="image12" alt="High Frame" /></td>
<td><img src="image13" alt="High Frame" /></td>
<td><img src="image14" alt="High Frame" /></td>
<td><img src="image15" alt="High Frame" /></td>
<td><img src="image16" alt="High Frame" /></td>
<td><img src="image17" alt="High Frame" /></td>
<td><img src="image18" alt="High Frame" /></td>
<td><img src="image19" alt="High Frame" /></td>
<td><img src="image20" alt="High Frame" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>27.37</td>
<td>29.09</td>
<td>26.28</td>
<td>28.17</td>
<td>28.54</td>
<td>28.34</td>
<td>27.80</td>
<td>26.80</td>
<td>27.96</td>
<td>27.45</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Medium</td>
<td><img src="image21" alt="Medium Frame" /></td>
<td><img src="image22" alt="Medium Frame" /></td>
<td><img src="image23" alt="Medium Frame" /></td>
<td><img src="image24" alt="Medium Frame" /></td>
<td><img src="image25" alt="Medium Frame" /></td>
<td><img src="image26" alt="Medium Frame" /></td>
<td><img src="image27" alt="Medium Frame" /></td>
<td><img src="image28" alt="Medium Frame" /></td>
<td><img src="image29" alt="Medium Frame" /></td>
<td><img src="image30" alt="Medium Frame" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>31.04</td>
<td>31.84</td>
<td>31.75</td>
<td>31.65</td>
<td>31.86</td>
<td>31.84</td>
<td>30.60</td>
<td>31.80</td>
<td>31.09</td>
<td>31.77</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Low</td>
<td><img src="image31" alt="Low Frame" /></td>
<td><img src="image32" alt="Low Frame" /></td>
<td><img src="image33" alt="Low Frame" /></td>
<td><img src="image34" alt="Low Frame" /></td>
<td><img src="image35" alt="Low Frame" /></td>
<td><img src="image36" alt="Low Frame" /></td>
<td><img src="image37" alt="Low Frame" /></td>
<td><img src="image38" alt="Low Frame" /></td>
<td><img src="image39" alt="Low Frame" /></td>
<td><img src="image40" alt="Low Frame" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>31.81</td>
<td>31.29</td>
<td>31.86</td>
<td>31.85</td>
<td>31.86</td>
<td>31.84</td>
<td>31.83</td>
<td>31.80</td>
<td>31.79</td>
<td>31.77</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
• **A-EFF**

Major essence of the proposed streaming design is to analyze the video quality while passing all the channel impairments unchecked in fragments of compressed bit stream. To model the worst case scenario where channel throughput is extremely limited and there should be a mechanism to stream the video contents up to the target destination with reduced delay and maximum possible good quality. A-EFF has proven to be an efficient solution for this where all the video frames are passed unchecked up to the application decoder at receiver end that yields an acceptable quality level for reconstructed videos. In this experiment, different video samples of different mobility levels are encoded following all the GOP structures to exploit the upper and lower bound of media quality achieved in all channel states.

A 100 frames 'mother-daughter' video sample is encoded following IBP GOP with 25 fps, that yields a bit stream at a bit rate of 1615 kbps after applying FEC encoding at PHY layer. Figure 5.9 shows the reconstructed objective quality for all three channel states where it is apparent that despite the unchecked errors in reference data, most of the video frames show an acceptable quality as above 24 dB even in high BER channel.

![PSNR vs. SNR for GOP-IBBPBBPBBI](image)

**Figure 5.9:** PSNR estimation for IBP GOP

Table 5.9 displays the individual video frames of one GOP along with their objective and subjective quality measures. It is shown that in high BER channel, some video frames exhibit lowest objective quality as 24.14 and 24.93 dB which is below 25 dB.
<table>
<thead>
<tr>
<th>Frame</th>
<th>I1</th>
<th>B2</th>
<th>B3</th>
<th>P4</th>
<th>B5</th>
<th>B6</th>
<th>P7</th>
<th>B8</th>
<th>B9</th>
<th>I10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Origin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOS.</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>36.97</td>
<td>36.50</td>
<td>35.93</td>
<td>36.04</td>
<td>34.34</td>
<td>35.88</td>
<td>35.17</td>
<td>34.98</td>
<td>35.13</td>
<td>35.50</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>36.97</td>
<td>36.25</td>
<td>35.93</td>
<td>36.04</td>
<td>34.34</td>
<td>35.28</td>
<td>35.17</td>
<td>35.98</td>
<td>36.24</td>
<td>36.92</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
However, video pictures are still visually acceptable and do not end in an annoying viewer response (shown through MOS).

![PSNR vs. SNR for GOP-IPPPPI](image)

**Figure 5.10**: PSNR estimation for IPI GOP

In the experiment for IPI GOP, a 100 frame 'foreman' video sample is encoded with 25 fps, yields a bit stream at 1830 kbps after applying FEC encoding. The compressed bit stream is transmitted and received at destination through applying A-EFF strategy. Here as compared to IBP GOP, a slight quality drop is apparent due to sequential propagation of bit errors in subsequent frames of GOP as shown in figure 5.10. Overall objective quality lies above 23 dB for most of video frames in high BER channel, attributed to maximum error propagation due to GOP structure formed by video frames types.

Table 5.10 shows the video pictures for one GOP where maximum channel impairment can be seen as the GOP progresses for subsequent P frames. However, visual quality again lies in an acceptable range.

For M-JPEG, video bit rate and ultimate objective quality is same as in case of PB-EFF because both lacks in error control for all types of video frames.
Table 5.10: IPI Quality Assessment for A-EFF

<table>
<thead>
<tr>
<th>Frame</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>27.49</td>
<td>26.82</td>
<td>25.81</td>
<td>25.07</td>
<td>23.99</td>
<td>29.89</td>
</tr>
<tr>
<td>MOS.</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>33.52</td>
<td>31.94</td>
<td>30.90</td>
<td>29.63</td>
<td>35.04</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>33.54</td>
<td>31.94</td>
<td>30.90</td>
<td>29.63</td>
<td>35.04</td>
</tr>
<tr>
<td>MOS.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

5.6.3 Link Saving

To quantify the trend of link saving for all the error control strategies, different streaming scenarios have been implemented in accordance with error states of various IEEE 802.11 radio channels as predicted through ILE. Link saving trends have been shown for all the radio channels in bad state or high BER.

Scenario-1: Link Saving in IEEE 802.11 g

This is the streaming scenario where available channel throughput allows to encode the video sample in order to limit the error propagation through following M-JPEG GOP. Here 100 frame 'foreman' video sample is encoded at 25 fps and yields a compressed bit stream at a bit
rate of 1815 kbps and fragmented into 197 frames at MAC layer, after applying FEC encoding at PHY layer bit rate becomes 3630 kbps. In worst case scenario, a radio channel in bad state triggers 201 MAC re-transmissions to recover the total data loss and ultimate throughput becomes 7335 kbps, average objective and subjective quality scores for this video sequence estimated at receiver APP are 34.30 dB and 5, respectively. In case of transmitting this stream on an IEEE 802.11 g channel with typical throughput of 8 Mbps, available channel throughput will be reduced to 665 kbps. After estimating such channel state along with the video quality statistics, ILE triggers a feedback mechanism to dictate the sender APP to cut down the video bit rate through applying a smart error control strategy. Now sender APP will decide whether to send a low quantization video produced through re-encoding the same video contents with low quantization scale that yields a compressed bit stream with low bit rate comparatively at the cost of quality degradation. A second option may be to send already compressed (at high quantization scale) bit stream to the receiver node while advising it not to recover the channel errors through MAC ARQ for non-reference video data, through applying PB-EFF. And a third option to ensure graceful bit rate adaptation is to de-activate the error recovery for the whole bit stream, through applying A-EFF strategy. Table 5.11 shows the video quality and bit rate estimates for LQF and A-EFF for the M-JPEG GOP.

<table>
<thead>
<tr>
<th>Bit rate (kbps)</th>
<th>Avg. PSNR (dB)</th>
<th>MOS</th>
<th>Channel throughput (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LQV</td>
<td>5787</td>
<td>32.31</td>
<td>5</td>
</tr>
<tr>
<td>A-EFF</td>
<td>3630</td>
<td>27.66</td>
<td>4</td>
</tr>
</tbody>
</table>

According to table 5.11, for LQV bit stream is encoded at a bit rate of 2874 kbps after applying FEC encoding at PHY layer with 156 MAC frames. At receiver MAC, ARQ triggers 158 re-transmission events to recover all the lost video data, ultimate throughput becomes 5787 kbps and available link throughput 2213 kbps is left for other wireless users. The ultimate reconstructed LQV exhibits objective quality level of 32.31 dB at a drop of about 2 dB from the previous HQV but visually it is still good with MOS of 5. Same for second streaming option where error recovery is discouraged at receiver end, video bit stream is encoded at a bit rate of 3630 kbps with 197 MAC frames. Due to zero re-transmissions at MAC video throughput remains same and available channel throughput remains 4370 kbps for the whole transmission which is almost double as compared to link usage of a LQV even.
### Table 5.12: Objective and subjective quality estimates

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>31.16</td>
<td>31.20</td>
<td>31.23</td>
<td>31.20</td>
<td>31.19</td>
<td>31.20</td>
<td>31.20</td>
<td>31.16</td>
<td>31.18</td>
<td>31.13</td>
</tr>
<tr>
<td>MOS</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>27.37</td>
<td>29.09</td>
<td>26.28</td>
<td>28.17</td>
<td>28.54</td>
<td>28.34</td>
<td>27.80</td>
<td>26.80</td>
<td>27.96</td>
<td>27.45</td>
</tr>
<tr>
<td>MOS</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Besides the efficient link saving reconstructed media quality is still acceptable as average PSNR of 27.66 dB and MOS of 4. Table 5.12 shows the frame by frame objective and subjective quality estimates for a single GOP of resultant videos of both streaming strategies. Similar streaming scenario can also be implemented for IBP GOP for a 'foreman' video sample as:

- ‘Foreman’ video clip FEC encoded at 2340 kbps with 25 fps at APP layer
- partitioned into 128 frames at MAC layer
- Bad channel triggers total 127 MAC frame re-transmissions to recover the loss
- New video bit rate is 4693 kbps, PSNR=31.30, MOS=4
- IEEE 802.11 g channel effective throughput becomes 3307 kbps for other users

After ILE triggers a feedback event to sender APP, three streaming options are available with their channel throughput usage and an ultimate video quality, as shown in Table 5.13.

<table>
<thead>
<tr>
<th>Bit rate (kbps)</th>
<th>Avg. PSNR (dB)</th>
<th>MOS</th>
<th>Channel throughput (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LQV</td>
<td>3970</td>
<td>30.65</td>
<td>5</td>
</tr>
<tr>
<td>P and B EFF</td>
<td>2352</td>
<td>29.76</td>
<td>4</td>
</tr>
<tr>
<td>All EFF</td>
<td>3052</td>
<td>26.86</td>
<td>3</td>
</tr>
</tbody>
</table>

All the available streaming strategies show great link saving with a minimal objective quality for first two strategies, however All EFF produces still a perceptually acceptable media stream even in high BER channel.

**Scenario-2: Link Saving in IEEE 802.11 b**

A 100 frame 'foreman' video sample is encoded following the IPI GOP at a bit rate of 1481 kbps with 25 fps. At MAC layer this stream is fragmented into 161 MAC frames and passed to PHY layer for FEC encoding. After applying FEC encoding at PHY layer, compressed bit stream bit rate becomes 2962 kbps. An erroneous radio channel in bad state results in ARQ at receiver MAC where total 166 MAC frames are re-transmitted to recover the total loss. Ultimately new video throughput becomes 6021 kbps that throttles the channel bandwidth (with typical throughput of 6.5 Mbps) which is left as 479 kbps for other wireless users. Video stream received at receiver APP is decoded and subjected to objective and subjective assessments that reveal objective quality as 32.36 dB and MOS of 5, respectively. Based on the channel and reconstructed video analysis, ILE triggers a feedback mechanism to sender...
APP that has three options to gracefully adapt the video bit rate for efficient and possibly optimized quality video streaming, as shown in table 5.14.

### Table 5.14: Link Saving Trend of GOP IPI for IEEE 802.11 b

<table>
<thead>
<tr>
<th>Bit rate (kbps)</th>
<th>Avg. PSNR (dB)</th>
<th>MOS</th>
<th>Channel throughput (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LQV</td>
<td>3710</td>
<td>31.65</td>
<td>5</td>
</tr>
<tr>
<td>P and B EFF</td>
<td>2586</td>
<td>29.06</td>
<td>4</td>
</tr>
<tr>
<td>All EFF</td>
<td>1830</td>
<td>25.86</td>
<td>3</td>
</tr>
</tbody>
</table>

In case of streaming a LQV, 100 frames 'foreman' video sample is encoded following the IPI GOP at a bit rate of 915 kbps with 25 fps using a low quantization scale. At MAC layer, bit stream is partitioned into 100 fixed length MAC frames and then subjected to FEC encoding at PHY layer to yield a bit stream at 1830 kbps. At receiver MAC, due to re-transmission events ultimate video throughput becomes 3710 kbps and available throughput of IEEE 802.11 b channel will be 2790 kbps for other users. This shows comparatively a better link usage as compared to a HQV with a minimal compromise of about 1 dB on objective quality while MOS is still 5 for good perception. Leaving the channel throughput for more concurrent wireless users and still streaming the media contents is possible through deploying EFF strategies. In P and B EFF, a video stream at a bit rate of 1830 is transmitted to target destination and an increase of 755 kbps is observed in net video throughput characterized to I frames recovery. This strategy shows the remarkable link saving with available channel throughput of 3914 kbps for other wireless users with an objective quality drop of 2 dB. Moreover, media users can further be facilitated in worst channel scenarios through a quality drop of about 5 dB through deploying All EFF strategy but still it is visually acceptable with a video rating score of 3 but without any pause particularly attributed to MAC re-transmissions.

Table 5.15 shows the frame by frame objective and subjective quality estimates for the resultant videos of all the three streaming strategies.
Table 5.15: Objective and subjective quality estimates

<table>
<thead>
<tr>
<th>Frame</th>
<th>I1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>I6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>LQV</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>33.38</td>
<td>31.83</td>
<td>30.93</td>
<td>29.82</td>
<td>35.04</td>
</tr>
<tr>
<td>MOS</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>A-EFF</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
<td><img src="image16" alt="Image" /></td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>27.49</td>
<td>26.82</td>
<td>25.81</td>
<td>25.07</td>
<td>24.29</td>
<td>29.89</td>
</tr>
<tr>
<td>MOS</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>PB-EFF</td>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
<td><img src="image21" alt="Image" /></td>
<td><img src="image22" alt="Image" /></td>
<td><img src="image23" alt="Image" /></td>
<td><img src="image24" alt="Image" /></td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>35.01</td>
<td>32.16</td>
<td>29.97</td>
<td>27.93</td>
<td>26.93</td>
<td>35.04</td>
</tr>
<tr>
<td>MOS</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

5.6.4 Visual Perception

The proposed idea is about assessing a video for its visual perception without bothering about the channel impairments that produces bit errors. Since multimedia applications can tolerate a certain degree of information loss without making a disastrous change. For video streaming scenario in wireless networks, invoking error eradication techniques or retransmissions add into computation cost as well as induce delay in video play back.

Proposed error control strategy based on the nominated performance parameters at various levels of ISO network stack greatly improve the multimedia quality without inducing any error adjustment overhead, media quality analysis results have been shown graphically for real time streaming clips of LM and HM levels, respectively. (Fig. 5.11-5.13)
Visual estimation of results involves objective and subjective VQA algorithms, computed in Table 5.3. In this experiment, real time captured video sequences are MPEG encoded and ultimate compressed bit stream is transmitted through impaired wireless channel. On receiving end, corrupted video frames are decoded somehow and assessed for visual perception through subjective quality indexing like SSIM, SSIM variants, noise estimation and VQM etc. As explained earlier, MSU VQMT provides a quality plot for respective parameter estimation where damaged frames are shown through red triangles, named as 'bad frames'. We have selected two video sequences with LM and HM levels, after passing through WLAN channel these sequences are reconstructed and assessed on SSIM indexing and VQM scale. Figure 5.11 describes the SSIM indexing for first video sequence of LM level with an average SSIM value= 0.92932, highlighting visual quality with good acceptance. In the SSIM indexing plot, where red triangles show the bad frames with lesser SSIM values.

**Figure 5.11:** SSIM Indexing, bad frames are corrupted frames with minimum SSIM values shown as red triangles, Average value=0.92932.

We have further picked two frames from this video sequence; a 'bad' and a normal frame to quantify the amount of error induced through channel impairment. In table 5.16, two reconstructed frames are shown along with their respective SSIM error images where bright regions indicate the percentage of error induced in the frame. Since, SSIM indexing is more closer to human visual perception as compared to any other metric like PSNR, MSE etc; a bad frame exhibits lower SSIM value= 0.431157 but shows acceptable visual quality, shown through its reconstructed frame.
To assess a video corresponding to better human perception for image quality, second video sequence of HM level has been assessed through VQM. It is based on DCT of the video frames, its values are always above zero and lower value symbolizes better quality. VQM plot with average value=0.422017 shown in Fig. 5.12, elaborates high quality reconstructed video with one frame shown as a bad frame.

Table 5.16: Original and reconstructed frames are shown along with respective error images and SSIM index values.

<table>
<thead>
<tr>
<th>Original frame</th>
<th>Reconstructed frame</th>
<th>SSIM error image</th>
<th>SSIM value</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Original frame" /></td>
<td><img src="image2" alt="Reconstructed frame" /></td>
<td><img src="image3" alt="SSIM error image" /></td>
<td>0.431157</td>
</tr>
<tr>
<td><img src="image4" alt="Original frame" /></td>
<td><img src="image5" alt="Reconstructed frame" /></td>
<td><img src="image6" alt="SSIM error image" /></td>
<td>0.9464</td>
</tr>
</tbody>
</table>

Figure 5.12: VQM Plot with average VQM value = 0.422017. Frame 635 is shown as a bad frame with value = 7.9205.

This one bad frame (frame-837) is further elaborated in Table 5.17, where degraded quality after reconstruction is depicted through VQM image with value= 7.9205. Since this one bad
frame possesses high VQM value and causes a quality drop in the whole video sequence but ultimately lower average value shows the high perceptual quality of the video sequence. Conclusively, video sequences with less motion contents are well reconstructed with high numeric and perceptual quality, while video sequences with high fraction of motion contents show degradation in quality after reconstruction on receiver side but these still are perceptually acceptable.

Table 5.17: Description of frame 635 with VQM value=7.9205.

<table>
<thead>
<tr>
<th>Original frame</th>
<th>Reconstructed frame</th>
<th>VQM error image</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Original frame" /></td>
<td><img src="image2" alt="Reconstructed frame" /></td>
<td><img src="image3" alt="VQM error image" /></td>
</tr>
</tbody>
</table>

MSU VQMT offers a way to compute the maximum probable quality assessment correlation to HVS through deploying its rich set of subjective metrics like significant variants of SSIM and other noise measures. Table 5.18 depicts a quantitative profiles of subjective quality metrics for 'container' video sample encoded following the IBP GOP and passed through a medium error channel state, shown one GOP statistics. Table 5.18 exhibits the objective and subjective metric evaluation for 'container' video sample, where frame B5 of the GOP appears as a bad frame with lowest objective quality (PSNR-29 dB, MSE-72.16), all the subjective metrics also show a decline in respective readings for this particular frame and maximum noise induced up to 1.115 level. Figure 5.13 shows the description of B5 frame as a bad frame along with the error images generated by subjective VQA algorithms. In comparison with SSIM, MS-SSIM is a more refined estimation of error where dark regions are apparent due to multiplication of several metric values below 1.0. Results improvement gradually shifts from single scale SSIM to 3 Component-SSIM and finally MS-SSIM reflects the most precise value of correlation between perceived video quality to perceptual properties of HVS, shown through estimated values for all these metrics.
Table 5.18: Objective and subjective VQA algorithms profile

<table>
<thead>
<tr>
<th>GOP-frm</th>
<th>Original frame</th>
<th>Distorted frame</th>
<th>MSE</th>
<th>PSNR</th>
<th>SSIM</th>
<th>J Com-SSIM</th>
<th>MS-SSIM</th>
<th>VQM</th>
<th>NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td></td>
<td></td>
<td>19.72019</td>
<td>35.1817</td>
<td>0.93697</td>
<td>0.96094</td>
<td>0.97422</td>
<td>0.99812</td>
<td>0.42146</td>
</tr>
<tr>
<td>B2</td>
<td></td>
<td></td>
<td>54.84203</td>
<td>30.73957</td>
<td>0.91616</td>
<td>0.92628</td>
<td>0.95612</td>
<td>2.80187</td>
<td>0.95999</td>
</tr>
<tr>
<td>B3</td>
<td></td>
<td></td>
<td>41.77806</td>
<td>31.92144</td>
<td>0.92462</td>
<td>0.93695</td>
<td>0.96488</td>
<td>3.05643</td>
<td>0.77262</td>
</tr>
<tr>
<td>P4</td>
<td></td>
<td></td>
<td>40.72308</td>
<td>32.03253</td>
<td>0.92025</td>
<td>0.9363</td>
<td>0.96269</td>
<td>2.94291</td>
<td>0.74934</td>
</tr>
<tr>
<td>B5</td>
<td></td>
<td></td>
<td>72.16356</td>
<td>29.5475</td>
<td>0.90164</td>
<td>0.91115</td>
<td>0.94096</td>
<td>3.06073</td>
<td>1.11599</td>
</tr>
<tr>
<td>B6</td>
<td></td>
<td></td>
<td>48.75177</td>
<td>31.25097</td>
<td>0.91217</td>
<td>0.9313</td>
<td>0.95393</td>
<td>2.77826</td>
<td>0.94598</td>
</tr>
<tr>
<td>P7</td>
<td></td>
<td></td>
<td>54.30072</td>
<td>30.78271</td>
<td>0.90357</td>
<td>0.92145</td>
<td>0.94931</td>
<td>2.83984</td>
<td>0.86566</td>
</tr>
<tr>
<td>B8</td>
<td></td>
<td></td>
<td>42.56697</td>
<td>31.84019</td>
<td>0.92396</td>
<td>0.93638</td>
<td>0.9647</td>
<td>3.07077</td>
<td>0.77345</td>
</tr>
<tr>
<td>B9</td>
<td></td>
<td></td>
<td>63.10515</td>
<td>30.13003</td>
<td>0.89604</td>
<td>0.91808</td>
<td>0.93957</td>
<td>2.93152</td>
<td>1.00135</td>
</tr>
<tr>
<td>I10</td>
<td></td>
<td></td>
<td>19.89539</td>
<td>35.1433</td>
<td>0.93721</td>
<td>0.95652</td>
<td>0.9742</td>
<td>1.0087</td>
<td>0.36509</td>
</tr>
</tbody>
</table>

VQM metric value is 3.06073 shows a slight quality decline as compared to remaining frames of the presented GOP in table 5.18, since its higher values identify low quality.
5.7 Comparison with other proposed techniques

5.7.1 Throughput Comparison

Yang Xiao et al. have proposed an idea of prioritized video frame treatment (PMFT [25]) characterized by assigning variable frame drop probability and retry limits to I, P and B video pictures. Comparing the improvement in throughput of a streaming station through applying the prioritized frame treatment (PMFT) to a normal streaming scenario (802.11), a graphical representation is shown as figure 5.14.

It shows an improvement in throughput drop of a streaming station as compared to normal scenario, where all erroneous MAC frames are dropped at receiver end. However in the PMFT approach, retry limits for I, P and B frames is set as 7, 3 and 1 respectively and P and B frames are discarded by the MPEG decoder if corresponding I frame is lost at receiver MAC. GOP size is taken to be as N=9 and maximum size of an I frame is 260 Kb.

Figure 5.13: Description of B5 frame as a bad frame

Figure 5.14: Throughput drop per station with increasing number of wireless users [25]
In order to compare the throughput trend of the proposed error control streaming strategy to PMFT and original 802.11 schemes, frame size is set to 260 Kb with N=9 for A-EFF and PB-EFF schemes. Figure 5.15 elaborates the comparison, where A-EFF outperforms the PMFT and original 802.11 schemes in terms of improved throughput without any loss. Moreover, BP-EFF also shows the improvement to original scheme but the throughput drop is attributed to the inevitable video parts recovery which greatly improves the visual quality, mainly due to error adjustment of I frames.

**Figure 5.15:** Throughput comparison

### 5.7.2 Channel Usage Comparison

In comparison with other streaming options, proposed error control strategies exhibit excellent channel usage trends for worst channel states. Table 5.19 shows the performance and channel resource consumption for different streaming options for transmission link with higher BERs. Video streaming option under original 802.11 standard specifications that involves ARQ at receiver MAC for lost and erroneous MAC frames has been selected as the reference, compared against the streaming of low resolution and less quantized contents where transmission bit stream is being re-encoded at runtime to reduce the size in accordance with available channel throughput. In table 5.19, reference scheme requires almost double number of MAC frames to be re-transmitted to recover the channel error or loss that tends to throttles the link capacity. LQV and LRV as the probable streaming options show require lower data rate but still MAC frame re-transmissions demands some extra bandwidth usage. However, deploying the error control scheme for selective video pictures (PB-EFF) or for all video pictures (A-EFF) greatly saves the link capacity and bring an opportunity to stream the
media contents even in low bit rate channels with ultimate acceptable objective and subjective video quality. For instance, PB-EFF and A-EFF show reduced data rates for streaming the contents of same video clips without wasting the link bandwidth in recovering bit errors, particularly ineffective for video contents. Although, these schemes exhibit an obvious drop in media quality attributed to the uncontrolled error but still quality is acceptable and does not result in an unpleasant effect on visual perception.

### Table 5.19: Channel usage trends of various streaming options

<table>
<thead>
<tr>
<th>E.C.S.</th>
<th>Avg. PSNR (dB)</th>
<th>Diff. (dB)</th>
<th>Data rate (kbps)</th>
<th>No. of MAC retr.</th>
<th>Sub. Ass. (MOS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orig. 802.11</td>
<td>32.36</td>
<td>----</td>
<td>6021</td>
<td>166</td>
<td>5</td>
</tr>
<tr>
<td>LQV</td>
<td>31.65</td>
<td>0.71</td>
<td>3710</td>
<td>102</td>
<td>5</td>
</tr>
<tr>
<td>LRV</td>
<td>30.60</td>
<td>1.76</td>
<td>2546</td>
<td>71</td>
<td>4</td>
</tr>
<tr>
<td>PB-EFF</td>
<td>29.06</td>
<td>3.3</td>
<td>2586</td>
<td>41</td>
<td>4</td>
</tr>
<tr>
<td>A-EFF</td>
<td>25.86</td>
<td>6.5</td>
<td>1830</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

**Summary:**

This chapter includes the details of the experimentation setup and the data set used. Results have been shown both in numeric form as well as video pictures are displayed independently to present the probable quality level as a consequence of various streaming and error control strategies. As obvious from different experimentation outcomes, selective error control mode outperforms all the other streaming options, particularly in terms of channel usage. Although, these methodologies result in media quality drop but still it is visually acceptable, shown through the results of subjective analysis.
Chapter 6

Conclusion & Future directions

6.1 Conclusion

Wireless networks have thrived as a renowned class of digital communication, characterized by easy deployment, immense portability and flexible infrastructure. With the proliferation of broad band radio connections offering higher bit rates, multimedia applications are being rapidly transited from the wired communication medium to wireless domain. However, media streaming over wireless network has faced with numerous design and performance challenges and indirectly pave the way for incessant research in this domain. Video streaming applications exhibit variable level of sensitivity to transmission delay; interactive applications (video conferencing) require instant data delivery without any pause or break otherwise video contents are useless whereas non-interactive media applications (video streaming) may tolerate some latency. Some other application classes prefer video throughput over quality, this category requires an uninterrupted contents flow sometimes at the cost of quality like streaming applications where a pause or break seems unpleasing rather a distorted frame may maintain the ongoing sequence.

This forms the baseline for the proposed streaming design as header bits at various network levels are marked to notify the streaming calls in order to attain the specified treatment. Such specified treatment has been characterized through disabling the checksum computation at various network levels that tends to disable the packet throw and ultimate corrupted video packets are forwarded to receiving end for probable decoding and quality assessments. Besides this error control strategy, joint optimization of various layer parameters has also been proposed that define a cross layer mechanism [56]. Since the error control strategy directs to retain the induced bit errors to save time and bandwidth, whereas optimization strategy on the other side works to reduce the media susceptibility against channel impairments. The proposed error control strategy has been triggered trough involving an intelligent module, ILE that facilitates graceful adaptation of video bit rate in accordance with the time varying radio link.

Proposed idea has been evaluated through conducting a variety of experiments in order to estimate the various media statistics (like objective/subjective media quality, video throughput etc) as well as network resource consumption like bandwidth usage etc. A mix of online available standard media clips has been used in experimentation to assess the
productivity of proposed idea for all possible video mobility levels (variations in foreground and background frame contents). Results have been assessed through subjective tests as well as through deploying a perceptual quality tool in order to measure the video suitability for visual experience, which is of at most concern in majority of streaming applications. System error control behaviour has been measured for different error scenarios, characterized by low and high error channels and various packet loss rates. Experimental outcomes obtained are quite revealing and assure numerous benefits of proposed idea. As obvious from different experimentation outcomes, selective error control mode outperforms all the other streaming options, particularly in terms of channel usage. Although, these methodologies result in media quality drop but still it is visually acceptable, shown through the results of subjective analysis. Analysis of all the video clips through a video perceptual tool shows encouraging outcomes even for video clips with rapid variations exhibit good estimates for certain perceptual video statistics that are based on frame structure and luminance values, potentially correlated to HVS. This favours the idea of video analysis on the ground of visual perception through deploying subjective methods, rather than computing objective values that are far beyond the HVS. Through deploying the proposed error control strategy, resource utilization had also been quantified that shows an efficient pattern of link capacity usage that directs the forward contents transmission (new frames transmission) rather than wasting the link in re-transmission of already sent data.

6.2 Contributions

The main contributions of the proposed idea are:

1. This idea is novel in the sense that it has surveyed the role of checksumming in video streaming over wireless link through estimating the computation and transmission costs of unnecessary packet drop as well as savings in current deployment.
2. It has highlighted the potentials available in network stack layers in the form of available bit patterns, that can be marked to classify a transmission flow in accordance with the performance constraints of a video application.
3. Novel error control strategies have been proposed that partly deploy the UEP technique to safeguard the critical video contents through MAC ARQ.
4. In addition, an intelligent module ILE has been deployed that assists in channel analysis and graceful adaptation of video bit rate.
6.3 Future Research Directions
Based on the encouraging outcomes of the experimentation conducted through deploying the proposed modifications and specifications in wireless media streaming framework, certain valuable future research directions have been highlighted. Video objective and subjective assessments have unveiled the significant parameters and practices that should be adopted in wireless framework particularly and in network stack generally to attain the pre-specified benefits of the proposed idea. These highlighted research tracks are summarized as:

6.3.1 Media Statistics
A digital video possesses variable contents pattern attributed to variable channel capacity or bit rates required for transmission. Similarly, some video fragments are more critical than other and requires enormous care for transmission through highly impaired wireless communication channel. Video statistics like GOP size, GOP style, macro block dimensions etc have been worked out in various research studies, and should be exploited to enhance the video invulnerability against channel error. Moreover, context aware binary encoding algorithms should be deployed with built in capability to accommodate variations in compressed bit stream.

6.3.2 Compression Parameters
Standard video compression algorithms are under a constant development process, various compression parameters like quantization step Q, macro block size, search space size, and compression technique like DCT and other newly emerged wavelet transforms etc should be worked out. Similarly, coding techniques with efficient codes like convolutional codes, turbo codes and LDPC codes should be deployed to achieve the improved compression efficiency in different communication domains.

6.3.3 Media Visual Perception
As the main essence of this research is to analyze the media quality though subjective VQA algorithms, for this work MSU VQMT ver. 3 has been used containing a rich set of quality parameters to assist in objective and subjective quality evaluations. There is a tremendous need to model more automated methods with publically available quality tools. Besides, more algorithms should be designed and implemented based on video contents structure and usability of information in error-free part of the corrupted network packets.
6.3.4 Utilizing Layer Potential

The proposed error control strategy is based on demarcation of bit patterns available in layer headers, this provides a motivation for using the header bits to classify a transmission flow with unusual network resource constraints. Wireless video streaming has emerged as a renowned class of radio communication domain, it requires some significant modifications in the persistent network stack protocols and services. Challenging demands of this application demand can better be fulfilled through developing new tools and techniques, and available bit patterns in the layer headers may assist in these valuable deployments.

6.3.5 Dynamic ILE Parameters

More dynamic network parameters in ILE data set should be incorporated like changing number of current wireless users, user requirements and priorities, re-transmission limits, back off counter depth and channel idle time in order to assess the time varying radio channel in a better way.
References

[10] Li Li and Qi Han, Xiamu Niu, “Enhanced Adaptive FEC based Multiple Description Coding for Internet Video Streaming over Wireless Network”,2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing.


"Evolution of 802.11 (physical layer)", http://www.okob.net/texts/mydocuments/80211physlayer (viewed on 11/03/2014)


"Video Test Media", https://media.xiph.org/video/derf/ (viewed on 10/03/2013)


[68] Frank Fitzek and Patrick Seeling and Martin Reisslein, "Video Streaming in Wireless Internet".


Appendix A

SUBJECTIVE ASSESSMENT OF RECONSTRUCTED MEDIA FILES

1. Subjective Assessment Methods
Video data is highly redundant in nature, urges for some mechanism to reduce the media load into a manageable size for efficient storage and transmission purpose. Standard video compression steps tend to reduce the size of video stream but at the cost of quality in terms of various irreversible encoding steps like quantization. Video quantization characterizes the process of lossy compression that tends to vary the size of compressed video stream through altering the quantization scale where loss of quality cannot be reverted on decoding. Moreover, compressed video stream undergoes transformations into various formats while transmitting from sender to receiver nodes at each step of a communication domain. This all tends to induce a quality drop, primarily over a transmission link where the scenario becomes worst for an error prone radio channel. Finally, upon reaching the destination reconstructed video needs to be judged for its visual perception based on the criteria of HVS. Unfortunately, no precise metric or tool is capable of fulfilling the perceptual demands of HVS due to involvement of multi-dimensional premises of human perception of media quality. The optimum solution lies in scoring of reconstructed media quality by humans themselves in accordance with their perception of media quality, termed as subjective Video Quality Analysis (VQA). Subjective VQA proves to be very expensive and even becomes impossible for some real time multimedia applications, however it provides an extreme valuable and concrete analysis outcomes.

Subjective VQA needs to be carried out in a controlled environment due to various influencing factors like ambient illumination, display device, viewing distance etc. Moreover, ITU-R Recommendation BT.500-11 [78] has standardized a checklists for selection of subjects as the most appropriate representatives of viewing audience who must be capable of providing a quality score with in probable range of a typical video scenario. This all pay towards the accuracy and authenticity of a subjective VQA process.
In order to facilitate the QoE analysis of decoding media files in the proposed research, a structured subjective VQA process has been carried out through involving an automated tool to assist the subjects.

2. **Population Description:**
Subject pool selected is a mix of 40 male and female students from UG and PG classes of CEME, NUST. Students have not been tested for any visual impairments as well as color blindness test as facilitated by the tool used, since almost all the videos are gray scaled. Each subject was briefly introduced about the goal of experiment and viewed a short training session in order to get his/her meaningful opinion. All the subjects were encountered with both the subjective scales as DSIS and DSCQS covering all the videos and each video was analyzed by 40 selected subjects. Each subject record was saved in data repository as the task name and its corresponding MOS score, where Difference MOS (DMOS) for the M number of subjects with MOS for jth video can be computed as:

$$\text{DMOS} = \frac{1}{M} \sum_{i=1}^{M} \text{MOS}_j$$

3. **Automated tool Support:**
An automated tools, MSU Perceptual Video Quality tool 1.0 [79] with a rich set of video analysis formats has been incorporated to assist the human subjects in assigning their opinion video rating. This tool comprises of two parts, first is "MSU Perceptual Video Quality - task manager" that can be used to create task file and to collect expert's results. Here, video files (.avi, .avs) to be compared, type of test that a subject must go through and name of the task are assigned to save the task file.(Figure 1)
Second part of the tool is "MSU Perceptual Video Quality - player". It plays the video files specified in the task file and record the subject opinion. (Figure 2)

MSU PVQ tool provides a user friendly GUI, based on three easy steps to record the subject opinion as:

1. Task files are prepared using "MSU PVQ - task manager".
2. Subjects go through tasks using "MSU PVQ - player", where their MOS are recorded.
3. Results are calculated using "MSU PVQ- task manager".

MSU PVQ tool offers two scoring approaches as to assign the quality score to a video after a playback finishes and the second one is well applicable for getting the scene dependent
quality opinion which is video rating while playback. Besides, it provides six subjective comparison methods, where the two methods used in the process are Double Stimulus Impairment Scale (DSIS) and Double Stimulus Continuous Quality Scale (DSCQS).

a. **Double Stimulus Impairment Scale (DSIS)**

In this method two media files are played simultaneously, where first is the reference second is an impaired one. And after playback finishes, a subject is asked to mark his opinion using comparison scale as in step 3 of figure 3.

![Figure 3: Step sequence in DSIS](image)

This method provides the evaluation of an impaired video only, where subject assigns the score in accordance with the scale as in table 1.

<table>
<thead>
<tr>
<th>MOS</th>
<th>Visual perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Immmperceptible</td>
</tr>
<tr>
<td>4</td>
<td>Perceptible, but not annoying</td>
</tr>
<tr>
<td>3</td>
<td>Slightly annoying</td>
</tr>
<tr>
<td>2</td>
<td>Annoying</td>
</tr>
<tr>
<td>1</td>
<td>Very annoying</td>
</tr>
</tbody>
</table>

b. **Double Stimulus Continuous Quality Scale (DSCQS)**

This subjective method is used more often where videos are played in pairs, and both videos are shown simultaneously. Each pair is repeated a given amount of times ("repetition" parameter in the Task Manager). As one of videos is the reference one, but a subject is not informed about it. After playback subject is asked about the impairment scale as in step 2 of figure 4.
Figure 4: Step sequence in DSCQS

Figure 5: Subjective test case generation for LM videos
Figure 5 elaborates the test case and respective video rating generation for subjective VQA for one mobility level as for video clips with minor foreground and background variations. According to results, DSCQS suffices for quality results generation since it provides an accurate visual comparison of both the media clips simultaneously where channel impairments can be well analyzed.