Effect of Codec Performance on Video QoE for videos encoded with Xvid, H.264 and WebM/VP8

Raj Kiran Addu and Vinod Kumar Potuvardanam

This thesis is submitted to the School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering. The thesis is equivalent to 20 weeks of full time studies.

This master thesis is typeset using LaTeX

Contact Information:

Author(1):
Raj Kiran Addu
Address:Karlskrona, Sweden.
E-mail: rajkiranaddu@gmail.com

Author(2):
Vinod Kumar Potuvardanam
Address:Karlskrona, Sweden.
E-mail: vinod.potuvardanam@gmail.com

University advisor:
Tahir Nawaz Minhas
School of Computing

University Examiner:
Dr. Patrik Arlos
School of Computing

Abstract

In recent years, there has been a significant growth in multimedia services such as mobile video streaming, Video-on-Demand and video conferencing. This has led to the development of various video coding techniques, aiming to deliver high quality video while using available bandwidth efficiently. This upsurge in the usage of video applications has also resulted in making end-users more quality-conscious. In order to meet the users' expectations, the Quality of Experience (QoE) studies has gained utmost importance from both researchers and service providers.

This thesis aims to compare the performance of H.264/AVC, Xvid and WebM/VP8 video codecs in wired and wireless networks. The codec performance is evaluated for different packet loss and delay variation values. The evaluation of codec performance is done using both subjective and objective assessment methods.

In subjective assessment method, the evaluation of video codec performance is done using ITU-T recommended Absolute Category Rating (ACR) method. Using this method the perceptual video quality ratings are taken from the users, which are then averaged to obtain Mean Opinion Score. These obtained scores are used to analyze the performance of encoded videos with respect to users' perception.

In addition to subjective assessment method, the quality of encoded video is also measured using objective assessment method. The objective metric SSIM (Structural Similarity) is used to evaluate the performance of encoded videos.

Based on the results, it was found that for lower packet loss and delay variation values H.264 showed better results when compared to Xvid and WebM/VP8 whereas, WebM/VP8 outperformed Xvid and H.264 for higher packet loss and delay variation values. On the whole, H.264 and WebM/VP8 performed better than Xvid. It was also found that all three video codecs performed better in wired network when compared to the wireless network.

Keywords: Quality of Experience, SSIM, H.264, WebM/VP8, Xvid.

Acknowledgments

We would like to thank our mentor and supervisor, **Tahir Nawaz Minhas** for his valuable support and guidance throughout our thesis work. We are grateful to him for his continuous advice and encouragement throughout the thesis.

We would like to specially thank **Dr. Patrik Arlos** for his advice and support during every stage of the thesis.

We are thankful to our parents for their unceasing support and encouragement. We are thankful to our beloved friends for their help and support and we would also like to thank one and all who helped directly or indirectly for the completion of this thesis.

Raj Kiran Addu
Vinod Kumar Potuvardanam
August, 2014

Contents

List of Figures

List of Tables

List of Abbreviations

ACR Absolute Category Rating

AVC Advanced Video Coding

DAG Data Acquisition Generation

DPMI Distributed Passive Measurement Infrastructure

FIFO First In First Out

FPS Frames Per Second

FR Full-Reference

GOP Group Of Pictures

HD High Definition

HTTP Hypertext Transfer Protocol

ITU-R International Telecommunication Unit, Radio Communications Sector

ITU-T International Telecommunication Unit, Telecommunications Sector

MArC Measurement Area Controller

MOS Mean Opinion Score

MP4 MPEG-1 Audio Layer-4

MP Measurement Point

MPEG Moving Pictures Expert Group

MSE Mean Squared Error

NR No-Reference

NTP Network Time Protocol

PEVQ Perceptual Evaluation of Video Quality

PSNR	Peak Signal-To-Noise Ratio
QoE	Quality of Experience
QoS	Quality of Service
QoE	Quality of Experience
QVGA	Quarter Video Graphics Array
RR	Reduced-Reference
SI	Spatial Information
SSIM	Structural Similarity
TCP	Transmission Control Protocol
TI	Temporal Information
TS	Traffic Shaper
VLC	Video Lan Client
VoD	Video on Demand
VQEG	Video Quality Experts Group
VCEG	Video Coding Experts Group
VQA	Video Quality Assessment

Chapter 1

Introduction

In recent years, video streaming has gained immense popularity. There has been a huge development and research in the field of video streaming. This has led to an increase in the use of multimedia applications and services over the Internet. According to the latest available statistics, globally the internet video traffic is predicted to be approximately 69 percent of all consumer internet traffic in the year 2017, which would be a drastic increase from 57 percent in 2012 [1]. Especially, the Video-on-demand traffic is estimated to triple by 2017 which shall be equivalent to 6 billion DVDs per month [2]. Also, two-thirds of the global mobile data traffic will be video by 2017, which would be about 66 percent of world mobile data traffic. Thus accounting for a 16 fold increase of mobile video between 2012 and 2017. This popularity and availability of video services make the user quality-conscious. The service providers should satisfy and fulfill the user satisfaction or Quality of Experience constraints in order to lead the multimedia market. This fast-rising demand is getting attention of network providers as well as research communities to prioritize the concept of Quality of Experience (QoE) along with Quality of Service (QoS) in order to improve services. Generally, the end-user is more interested in the quality of video rather than quality of service, which makes the concept of Quality of Experience much more important for the network service providers to stay ahead of their competitors.

In video streaming, QoE is influenced by various factors such as video compression, video transmission and video reconstruction [3]. Video compression is a technology that is used to compress the videos from raw format, so that they can be streamed via network without consuming more bandwidth and resources. The modern video encoders and decoders are used for video compression in order to deliver the video without compromising on quality while consuming less bandwidth and resources. These video encoders and decoders use video codecs to achieve optimum compressed videos that can be delivered via network. There are various videos codecs that have been developed by various organizations and open source groups such as Theora, Dirac, Xvid, DivX, H.264 and VP8. Some of the most widely used video codecs amongst them are H.264/AVC [4], Xvid [5] and WebM/VP8 [6].

Another aspect that plays a key role in video streaming is the video transmission. The impact on the quality of streamed video also depends on the type of network used for transmission of videos. The transport layer protocols have a noticeable impact on the quality of streamed videos. Generally, the video streaming services are based on HTTP and TCP, in which the video player is embedded to the web browser [7]. Many popular video providers such as Youtube [8], Vimeo [9] and Dailymotion [10] use HTTP video streaming. Many studies have evaluated the user perceived quality for HTTP based applications [11], [12]. However, the knowledge of performance and quality of experience of HTTP video streaming is still limited, especially in the aspect of perception of video quality and its factors.

In this thesis, we study how network parameters such as packet loss and delay variation have an impact on the user perception of video quality with respect to video codecs H.264/AVC, Xvid and WebM/VP8. The experimentation is performed using wired and wireless emulated network scenarios. The required network conditions for the test environment are created using NetEM [13]. The NetEM is used, rather than other emulators such as KauNet and NISTNet. In [14], [15], the authors have shown that performance of NetEm is more reliable when compared to NISTNet and KauNet. The Video-on-Demand streaming is performed, where videos with different temporal and spatial aspects are streamed with varying network parameters such as packet loss and delay variation. Here, Flumotion is used as a streaming server and VLC player is used on the client side. The HyperText Transfer Protocol (HTTP) over TCP is used for video streaming over the emulated networks namely wired network and wireless network.

Further, the investigation of the users perception is done for the impact of delay variation and packet loss on the video Quality of Experience for different video sequences encoded with H.264/AVC, Xvid and WebM/VP8. The assessment of videos is done in two fold. Firstly, the objective assessment of videos were made using Structural Similarity (SSIM) index as a performance measurement metric. Secondly, the subjective assessment of videos were performed adhering to the International Telecommunication Union (ITU) recommendations P.910 [16].

1.1   Motivation

Advances in video coding technology along with rapid developments and improvements in the field of video streaming are enabling an increasing number of video applications. Due to this rise in video applications, there is an increased attention towards evaluation of video codec performance. Many video codec comparisons have been performed by various groups, but they were mostly based on either objective measures of video assessment or subjective measures of video assessment. However, the studies of video quality assessment for different video codecs on wired and wireless network is limited.

It is interesting to know how different codecs behave or perform in wired network and wireless network scenarios. Especially, when the network is affected by various delay variation and packet loss conditions. In this thesis, we try to figure out the performance of the three video codecs in wired and wireless network scenarios. A series of experiments are conducted to investigate the performance of video codecs. However, neither objective nor subjective tests alone gives a complete picture about the video quality. That is why, we have used two assessment methods, namely objective and subjective in our research work. These findings shall contribute to the video streaming service providers to get an understanding of the video codec behavior with respect to different networks.

1.2 Aims and Objectives

The main aim of this research is to study the impact of network parameters on the codec performance with respect to Quality of Experience (QoE). This research work is divided into following objectives:

- To get an understanding on the video quality of experience for encoded videos that are streamed over wired and wireless network scenarios.
- To investigate how these three codecs perform in wired and wireless network scenarios.
- To get an understanding on various video codecs, encoding techniques and network emulators.
- To analyze the results obtained from the experiment for effect of packet loss and delay variation using objective assessment method.
- To analyze user Quality of Experience by using Mean Opinion Score.

1.3 Research Question

The research questions that are addressed in this thesis are as follows:

1. How does the user perceive the quality of videos encoded by three codecs in the presence of packet loss and packet delay variation? This research question is further broken down into three sub-questions:

   (a) How does the user perceive the quality of video encoded with Xvid in the presence of packet loss and packet delay variation?

   (b) How does the user perceive the quality of video encoded with H.264 in the presence of packet loss and packet delay variation?

(c) How does the user perceive the quality of video encoded by Web-M/VP8 in the presence of packet loss and packet delay variation?

2. How does the user perceive the quality of encoded videos that are streamed over wired and wireless networks?

1.4   Research Methodology

To answer the above research questions, both qualitative and quantitative approaches have been used. In the early stage of our research, a literature study has been performed which gave us an insight to the previous works done by other researchers in the video streaming research area. This literature study provided us with information on video compression technology, video codec functionality, video quality assessment methods, emulated experimental setup and network scenarios. Based on the literature study, the required components and parameters of the experimental environment were selected.

After the completion of literature study, an emulated experimental environment was created and the experimentations were performed for two network scenarios, namely wired network and wireless network. During the experiment, packet loss and delay variation were introduced using a traffic shaper in a controlled environment. The selection of video sequences for the experiment were done based on their temporal and spatial characteristics. These selected videos were encoded and then streamed from server to client. The resulting videos were stored in the local database so that they could be used later in the assessment stage.

In the later stage, the evaluation of resultant videos was done using objective and subjective assessment methods. Firstly, the objective assessment of videos were performed based on the structural similarity (SSIM). And then the subjective assessment of videos was performed. In which the user perceptions were obtained by conducting a user rating survey of the videos, where videos with different parameters were shown to the subjects. The user perceptions were collected and analyzed as per the International Telecommunication Union recommendations [16] using subjective analysis methods. The single stimulus ACR methodology was used to perform subjective assessment of videos. As we wanted to measure users' perception of video quality for video sequences with different distortions with regard to three codecs H.264, Xvid, WebM/VP8. Also, ACR method is a recommended method for conducting assessment tests as per several studies [17], [18]. Finally, the results were calculated and presented by using the five-level quality scale mean opinion score (MOS) and conventional statistical methods.

RQs 1-2 are preliminarily answered by experiments and analysis of the results, where the network based impact of the video codec performance of the three specific codecs on video Quality of experience (QoE) is unveiled.

1.5   Thesis Outline

The remainder of the thesis is organized as follows. Chapter 2 describes the technical background on video streaming and video quality assessment. The concept of compression as well as detailing about video codecs and network scenarios are presented. An elaboration on how the video quality assessment is done by using objective and subjective methods is given. A brief description of widely used statistical methods such as MOS is also given. Chapter 3 describes the measurement environment. The test scenarios and parameters used are detailed explicitly. Chapter 4 details about the analysis and results. Finally, Chapter 5 summarizes the conclusion and future work of the thesis

Chapter 2

Technical Background

In this chapter, the basic concepts used in the evaluation of video codec comparison are introduced. An overview of video streaming, video compression and video codecs are provided. Later the main aspects related to video transmission and network scenario is presented. Finally, the concept of video quality assessment is demystified with a brief introduction.

2.1 Quality of Experience

Generally QoE refers to the perceived quality of service by the end-user. As per ITU-T, it is defined [19] as *"The overall acceptability of an application or service, as perceived subjectively by the end-user"*. It is also defined [20] as *"Degree of delight of the user of a service. In the context of communication services, it is influenced by content, network, device, application, user expectations and goals and context of use"*. It considers complete end-to-end system effects such as effects of network, infrastructure of services and client. It also considers the end user's satisfaction and acceptability to the services provided. The end user satisfaction involves various psychological factors such as user's mood, emotions and state of mind while using the service. In [21], the author states about the QoE parameters importance when compared to QoS parameters.

2.2 Video Streaming

Video streaming refers to the process of videos being transferred from a source to one or more destinations. The video is usually streamed from server to clients. The video streaming is basically comprised of two fundamental activities:

1. Creation of digital content using compression techniques.
2. Content transmission over the network.

The creation of digital content is done using compression techniques, as it is expensive to transmit a raw video over the network. Streaming a raw

video over the network consumes more network resources as well as storage resources [7]. Thus video compression plays a key role in the process of video streaming.

2.3   Video Compression

Video compression refers to the process, where the raw video is compressed using various compression based algorithms and mechanisms. This is done in order to reduce the size of the videos, so as to ensure transmission of videos over a network without consuming more network resources. Generally, in wired or wireless networks, an uncompressed video consumes more bandwidth and storage. Thus, increasing the end user cost with respect to bandwidth and data transmission capacity in the network. In order to maximize bandwidth utilization of the network while transmission of videos, it is indispensable to use video compression [22].

The video compression is done in two ways, namely lossless compression and lossy compression. Lossless compression is a technique in which no information is lost. The lossless compression technique can reduce the size of the video to a small extent. The videos compressed using lossless techniques are not suitable to be streamed over a network due to their large size. Thus, for greater compression of videos one must use lossy compression technique. Lossy compression is a technique that compresses a video by discarding information. Thus, in lossy compression, some information is lost which leads to reduction in the size of video along with degradation of video quality. The main goal of video compression is to maintain a fine balance between video quality and size of the video. To serve this purpose, various video codecs have been developed [23].

2.4   Video Codecs

Video compression involves two components, namely (Co)der and (Dec)oder, which is combinedly called as CODEC. Video codec is a software program capable of encoding and decoding. It compresses the raw video into a small-sized video that can be easily transmitted over network without consuming more bandwidth resources. Video codec is used to compress a video file where as an audio codec is used to compress an audio file. Due to increase in video communication nowadays, many efficient video coding techniques have been developed to provide high quality video streams using the available bandwidth [24].

2.5 H.264

H.264 is a video compression technology that was jointly developed and standardized by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG). It is also known as MPEG4 part 10 or AVC (Advanced Video Coding). It is quite complex and flexible. It has been developed to cover a wide range of video streaming services from mobile video streaming to hi-definition video streaming or broadcasts.

H.264 compresses the video more effectively, when compared to previous H.26x standards. It has some optimized features such as utilization of temporal and spatial correlation, usage of entropy coder, which were taken from previous encoding standards. Some of the important features of H.264 video codec are as follows [4]:

- It uses a 4x4 spatial block transform instead of 8x8 DCT as found in older video codecs.
- An additional Hadamard transform (2x2 on chroma and 4x4 on luma) can be usually performed to obtain more compression in smooth regions.
- Quarter sample motion vector accuracy.
- In-the-loop deblocking filtering.
- It houses latest enhanced entropy coding method namely Context Adaptive Variable Length Coding (CAVLC) and Context Adaptive Binary Arithmetic Coding (CABAC).
- Complex spatial prediction for intra frame compression has been introduced in H.264.
- Multiple reference frames, allowing up to 16 reference pictures to be used unlike previous standards where 1 or 2 reference frames were used.
- Network Abstraction Layer (NAL) which facilitates simple and effective video streaming over networks.

Special attention has been given by the developers to improve the robustness to data losses during video transmission [24]. H.264 codec has three types of frames namely I-frame, P-frame, and B-frame. I-frame (Intra-frame) is independent and does not reference any other frames. P-frame (Predicted frame) is decoded predictively from the closest previous reference frame that can be either I-frame or P-frame. P frames are more compressible when compared to I frames. Finally, B-frame (Bi-directional frame) depends on both past and future frames for decoding and is the most compressed frame amongst the three frames. Therefore, I-frame is the most important frame when compared to other two frames.

H.264/AVC standard defines mainly three encoding profiles which are frequently used namely Main profile, Baseline profile and High profile. The baseline profile has less amount of video data compared to other profiles due

to lower quality versions of the video. Baseline profile is suitable for video conferencing. Whereas the main profile and high profile are suitable for high end video broadcast and video storage [4]. These profiles are classified in levels indicating the limits of various parameters, namely video resolution, coding bit rate, max decoding speed and max frame size. These profiles and levels are used as limits to encode/decode videos which targets respective classes of video communication applications.

2.6   Xvid

Xvid is an open source implementation of MPEG-4 video standard. Xvid is one of the popular MPEG-4 codecs used for non-commercial purpose [25]. It is used in fields such as Digital Television and interactive multimedia applications. It provides high data compression while maintaining good video quality. The MPEG-4 compression is based on visual-objects coding [26]. The Xvid codec uses discrete cosine to transform functions developed by the Independent JPEG group. It provides a flexibility in coding techniques and visual data types which include synthetic and hybrid visual data, still images and arbitrary shaped regions. Similar to MPEG-4 AVC, Xvid also has a wide range of profiles and levels for encoding videos with various parameters for video streaming purposes [27].

2.7   WebM

VP8 [28] is an open source video codec which was initially developed by On2 technologies and released by Google. It was developed for serving video on web with high quality. It consists of Intra frame and inter frame which are similar to I-frame and P-frame of H.264 respectively. An intra-frame is the key frame which is independent and has no reference to other frames. Whereas, an inter frame is dependent on previous frames, including intra frame to decode. However, lack of B-frames is the most important difference between VP8 and MPEG-4. Generally, VP8 is comparable with H.264 Baseline. Some of the prominent features are as follows [6]:

- High performance sub-pixel interpolation.
- Adaptive in-loop deblocking filtering.
- Frame level adaptive entropy coding.
- Flexible reference frames.
- Efficient intra and inter prediction.
- 4x4 block based discrete cosine transform (DCT) for all luma and chroma residual signal.

2.8 Video Transmission

The streaming protocols are designed to provide data transmission, network addressing and service between the server and the client. The transport protocols are used for the purpose of establishing communication between streaming servers and clients. At the transport layer, two major techniques are used, namely TCP based video streaming and UDP based video streaming. Currently, TCP is the most widely used transport protocol in the internet. The TCP or transmission control protocol is a byte stream, connection oriented and reliable delivery transport layer protocol [7].

TCP is said to be reliable due to its mechanisms such as checksums, re-transmissions and sequencing. In the context of video streaming, there are benefits of using TCP due to its congestion control mechanism and flow control mechanism. On the other hand, TCP introduces undesirable transmission delays due to its re-transmission capabilities. Many video streaming services such as YouTube, Vimeo and Dailymotion are usually based on HTTP over TCP streaming methods [29]. HTTP has some advantages such as data integrity, omnipresence and firewall friendliness. Moreover the authors in [30], suggests the usage of TCP for optimum performance in video streaming especially when the available network bandwidth is twice to that of video rate.

2.8.1 Transmission Control Protocol (TCP)

TCP is a connection oriented and reliable transport layer protocol that uses a three-way handshake method to establish and maintain a session between sender and receiver. The three-way handshake method is described as follows:

- Initially, the client sends a synchronization packet (SYN) to the server, requesting the server to synchronize its sequence numbers with that of the client.
- Then the server responds by sending an acknowledgement (ACK) along with a SYN. Here, SYN is a request for client to synchronize its sequence numbers with that of the server.
- Finally, the client acknowledges the servers request by sending an acknowledgement (ACK) to the server.

TCP also employs a congestion control mechanism that adjusts the transmission rate by limiting each TCP connection to its fair share of network bandwidth. The TCP congestion control regulates the sending rate which forces the sender to transmit at a lower rate [29].

2.9 Video Quality Assessment

With the increase in the usage of various video based applications, has led to a fierce competition between service providers and also between developers to provide better quality than each other. This has given rise to the importance of assessment of video quality. Basically, the video quality assessment is done in two methods:

1. Objective video quality assessment method based on mathematical calculations.
2. Subjective video quality assessment method based on tests performed by users.

2.9.1 Objective Video Quality Assessment

There is a growing interest in the development of objective video quality assessment methods to provide instantaneous measurement of video quality. The subjective video quality assessment methods are usually time consuming, exhaustive and expensive. The Video Quality Experts Group (VQEG) validates the objective video quality metric models that result in ITU recommendations and standards of objective quality models [31]. The objective video quality assessment methods are classified into 3 classes namely Full Reference (FR), Reduced Reference (RR) and No-Reference (NR).

Generally, full-reference methods are mostly used for objective video quality assessments. A full-reference method is used when the original reference video is available for comparison. Some of the widely used objective quality metrics are SSIM, PEVQ, PSNR and MSE [32]
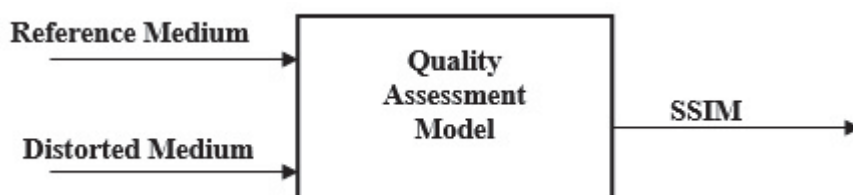
Figure 2.1: Full Reference Video Quality Assessment

SSIM (Structural Similarity)

SSIM is a widely used objective video quality assessment method which compares luminance, contrast and structural similarity between original reference video and processed video. SSIM uses a different approach than that of other metric such as PSNR. It is based on the idea that human visual perception is adapted for extracting structural information from the video [33]. The structures of objects in a particular scene are not influenced by luminance

and contrast. Thus, the measurement of structural similarity is aimed to give a better correlation with respect to subjective user perception.

The structural similarity index can be calculated as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{2.1}$$

Where μ_y and μ_y are estimates of the mean of x and mean of y respectively. σ_x^2 and σ_y^2 are estimates of variance of x and variance of y respectively. σ_{xy} is the co-variance of x and y.
c_1 and c_2 are constants.

The value of SSIM is between -1 and 1 and gets the best value of 1. According to Wang [33], the correlation between SSIM and subjective scores are good.

2.9.2 Subjective Video Quality Assessment

The subjective video quality assessment method is the most reliable video quality measurement method to evaluate the video service. It is the assessment of video quality as perceived by the user. In QoE the subjective results are different from user to user [34]. Human perception involves various aspects of human psychology and various factors such as illumination, vision ability and other viewing conditions.

MOS stands for Mean Opinion Score which is a metric obtained from the user ratings given by the subjects for the video quality perceived by them.

2.10 Related Works

In this section the relevant research work done in the field of video QoE and video codecs is discussed. Several studies have been presented with the quality of experience (QoE) with respect to the videos. In paper [35], the authors presented a conceptual model of QoE, which considers both measurable and non-measurable parameters in quality evaluations. They have used their model to measure QoE in mobile scenarios.

In [21] authors have proposed a QoE assessment model for video streaming service using QoS parameters in wired and wireless network through which the network operators can correspond to poor quality by monitoring the QoE of video streaming service. Additionally, in paper [36] authors address the video quality correlation with respect to QoE and QoS. In this study, a generic formula has been proposed in which QoE and QoS parameters are connected through an exponential relationship which has been validated for streaming servers. In paper [37], authors explain the technical aspects such as application and network level QoS and non-technical aspects such as user

experience. The importance of the relationship between QoE and technical parameters to manage the user perceived quality is explained.

In [38] authors have subjectively assessed the video quality of H.263, H.264 and Xvid for mobile devices. They have investigated the effects of codecs with respect to different combinations of bitrates and video content in mobile networks. Alternatively, In [28] authors have compared two codecs namely H.264 and VP8 in which they have measured the objective video quality of videos with different encoding related parameters.

The transport layer protocols have a noticeable impact on the quality of streamed video. The studies [39] depict the importance of transport layer protocols in assessing the video quality. In [40] authors have analyzed the performance of latest video codecs such as H.264/AVC, H.264/SVC, and VP8 over IEEE 802.11 wireless network. They have shown that reducing the coded video data, considering the queue size, implementation of error recovery features have facilitated a robust transmission over WLANs. They have compared the performance of the above video codecs on various mapping schemes.

In [23] the authors have compared the rate distortion performance of five video codecs in terms of peak signal to noise ratio (PSNR) and just noticeable difference. In [41] the authors have compared two video codecs namely H.264 and VP8 in terms of video bit rate-distortion performance and video network traffic variability. Here, the authors have used different long video sequences for comparing performance of video codecs. Studies [42], [43] have analyzed user's perception for streamed videos and the impact of packet loss and effect of various artifacts have also been analyzed.

In [44], authors have evaluated the sensitivity of mobile video to packet loss and packet delay. They have revealed that mobile video (H.264 baseline) is very sensitive to packet loss and packet delay variation. In [45], authors have made a comparative study of objective and subjective video quality for the codec H.323 in which they found that jitter had the biggest effect when loss, delay and jitter has been introduced to the same video sequence. In [46], authors have investigated the effect of artifacts on user perceived quality where the video quality assessment is done by analyzing the effects of artifacts and packet loss. Moreover, in [47], [48] the authors have investigated the impact of video freezes and video jump on user perception.

Although many works have investigated the performance comparison of video codecs, those comparisons were mostly limited to the comparison of various encoding based parameters with respect to either subjective or objective metrics. In this thesis work, we evaluate and compare the performance of three video codecs using objective and subjective assessment methods. The videos encoded with selected codecs are streamed in wired and wireless network in the presence of packet loss and delay variation. This enables us to compare the performance of codecs in wired and wireless networks.

Chapter 3
Experimental Design

This chapter discusses about the measurement environment. The technical aspects of the experiment design are discussed as well as the experimental setup and configurations under which the measurements were collected are epitomized.

3.1 Video Parameters

This section describes the video parameters that are selected for the experimentation.

Video Content Selection

The video sequences are selected from the standard video pool, such that their temporal and spatial properties are consistent. All the test sequences have different motion activities, where videos differ in their Temporal Information (TI) and Spatial Information (SI) characteristics. As per the suggestions of ITU-T [16], it is suggested to consider distinct video sequences to evaluate the video codec performance which is directly dependent on the TI and SI characteristics. For example, videos with high motion activities respond differently to the coding paradigms of the video codecs when compared to videos with slow motion activities.

The selected video sequences are taken from a commonly used repository that is used for video quality assessment studies as suggested by Simone Et al. [49]. These video sequences are widely used in the video testing community because of their varied content with regard to the spatial and temporal aspects. The selected video sequences are described in the Table 3.1.

The selected set of test sequences for the experiment which are described in the Table 3.1, covers a wide range of video contents in terms of motion and spatial details.

Table 3.1: Overview of the video content

Video Name	Number of frames	Duration (Seconds)	Description
Crew	300	10	A group of crew members walking through the corridor waving. The camera moves according to the motion of the group.
Foreman	300	10	The face of a foreman speaking is shown. The camera shakes a little and then at the end the camera moves towards the building under construction.
News	300	10	The scene consists of two news readers in the front with low movement and two dancers performing in the background with high movement.
Football	260	8	A high motion video sequence of a game where the camera follows the ball.
Hall-Monitor	300	10	Two persons walk in an office corridor in the opposite direction, where one person is carrying a briefcase and another person carries a TV. The Background has no activity and the focus is on the movement of the two subjects.

Video Resolution

The video resolution selected for the experimentation is QVGA (320 x 240) which is larger (higher resolution) than QCIF, but smaller than CIF. The Quarter Video Graphics Array (QVGA) is a common resolution for modern phone displays. The QVGA resolution is used in modern mobiles as this low resolution mode requires less data storage capacity than other higher resolutions and offers more detail than the standard QCIF resolution (177 x 144) which was used in previous video studies for mobile [44]. This resolution was chosen for the experiment, as a wide range of mobile phones (Eg. HTC, Nokia, Sony Ericsson, iPhone) support this display resolution.

Video Frame Rate and Bit Rate

Frame rate is basically the rate at which the video system projects or displays the images (frames) per second. It is measured in frames per second (fps). The frame rate used for this experiment is 30 fps as this is commonly used in various video assessment related studies [44] and also supported by various modern mobile phones.

The bit rate basically refers to the amount of information (number of bits) that the video will process in a given period of time. The bit rate chosen for the experimentation is 768 Kbps. This value is commonly used for mobile devices and also commonly used in various video assessment studies for mobile [44].

Video Sequence Duration

As per ITU-T [16], the length of each video sequence used for our experiment is between eight and ten seconds. According to [34], short sequences are considered more stable in terms of spatio-temporal properties and also lead to valid and consistent results. The results obtained are termed to be consistent as the subjects rate the video based on its quality and do not get immersed in the content. As for longer videos there is a chance that users focus may shift towards the content rather than video quality.

Video Codec

The video sequences for the experiment have been encoded using FFmpeg [50] (available in APPENDIX A), which is a cross platform tool.

H.264

The video sequences are encoded using H.264 baseline profile, level 1.3 in our thesis work. As these settings are recommended for mobile video [44] [5]. FFmpeg has been used for this purpose. The video parameters as shown in Table 3.2 are used to encode the video.

Xvid

The video sequences are encoded using the Xvid simple profile in our thesis work. FFmpeg has been used for this purpose. The video parameters as shown in Table 3.3 are used to encode the video.

WebM/VP8

The video sequences are encoded using VP8 codec using FFmpeg. The video parameters as shown in Table 3.4 are used to encode the video.

All the below encoding parameters have been kept constant for the whole experimentation.

Table 3.2: Video Parameters for H.264

Video Sequences	Crew, Foreman, News, Football, Hall-Monitor.
Video Codec	H.264/AVC baseline profile, level 1.3
Resolution	QVGA (320x240)
Bit rate	768kbps
Frame-rate	30fps
Aspect Ratio	4:3
Container	MP4
Encoder	FFmpeg (libx264)

Table 3.3: Video Parameters for Xvid

Video Sequences	Crew, Foreman, News, Football, Hall-Monitor.
Video Codec	Xvid simple profile
Resolution	QVGA (320x240)
Bit rate	768kbps
Frame-rate	30fps
Aspect Ratio	4:3
Container	MP4
Encoder	FFmpeg (libxvid)

Table 3.4: Video Parameters for WebM

Video Sequences	Crew, Foreman, News, Football, Hall-Monitor.
Video Codec	WebM/VP8
Resolution	QVGA (320x240)
Bit rate	768kbps
Frame-rate	30fps
Aspect Ratio	4:3
Container	MP4
Encoder	FFmpeg (libvpx)

3.2   Network Emulation

The network emulation is useful in scenarios where the traffic needs to be shaped by delay, packet loss and packet reordering. Network emulation is used to eliminate certain factors found in real networks such as, traffic interference and variation in load. Generally, an emulated network is used by the research community in performance measurement studies because of its features like reproducibility and controllability. The traffic shapers are used to emulate the network by varying parameters like delay, jitter and packet loss.

The NetEm traffic shaper has been chosen in our thesis for network emulation, to vary performance parameters like delay variation and packet loss. NetEm belongs to Traffic Control (TC) bandwidth provisioning package of linux [13]. In [14], [15] authors have assessed the performance of NetEm to be more reliable than other traffic shapers namely NIST Net and KauNet. NetEm uses FIFO Queuing Mechanism to make decision policy of packets to be sent based on given input parameters. The input parameters are given through a command line interface tool tc traffic control.

3.3   Packet Loss and Delay Variation

The packet loss refers to the number of packets that fail to reach the destination when being transmitted across the network. Packet loss has noticeable effects in various applications such as video streaming and video conferencing. The packet loss is basically calculated in terms of percentage (%) of packets lost from the overall transmitted packets.

The packet loss values used in our thesis are 0%, 2.5%, 5%, 10%, and 15%. In general, 1% packet loss means for every hundred packets transmitted there will be a drop of one packet. And also a similar range of values has been used in various studies [44], [51].

Delay variation is commonly used due to its inherent variability in arrival times of individual packets. The packets arrive to the destination with random time distortion where the time between packets at destination varies from that of time between packets at source [44]. The delay and delay variation values are expressed as D $\pm\Delta$ D, where D is the fixed delay and ΔD is the variable delay. The ITU-T G.114 [52] suggests that the delay must not exceed the 150ms limit for one-way transmission time. The above mentioned study defines the use of 150ms as a good and acceptable delay value.

The fixed delay (D) and variable delay (Δ D) values considered for our thesis are 150ms and {$\pm$50ms, $\pm$90ms, $\pm$110ms, $\pm$130ms, $\pm$150ms} respectively.

In our thesis the shaping values were chosen from the literature study and laboratory testing. Initially, tests were performed for various values

of packet loss and delay variation. Finally, values where the MOS ratings changed significantly were selected for the experimentation.

3.4   Experiment Test bed

This section describes the experimental scenarios used in our thesis and also elucidates the technical components of the setup along with the experimental procedure.

3.4.1   Wired Network Scenario

The wired experimental test bed depicted in Figure 3.1 is used to carry out the experiments. It consists of two Flumotion streaming servers (SA and SB), two traffic shapers (TS A and TS B) and two client systems (CA and CB). All these components are connected to the MArC via Measurement point (MP). All the systems are connected as shown in the below test bed. Two independent full duplex test networks have been used in the wired setup where each network consists of a set of server, traffic shaper, client and MP. The two independent test networks are illustrated using different color coding in Figure 3.1. The traffic shapers acts as bridges between server and client in their respective networks.

The test network also includes Measurement points (MP) equipped with DAG 3.6E cards which captures the packets using wiretaps as shown in the test bed. The MArC is used to control the measurements and the test network is controlled by a controller (XPS) as shown in Figure 3.1.
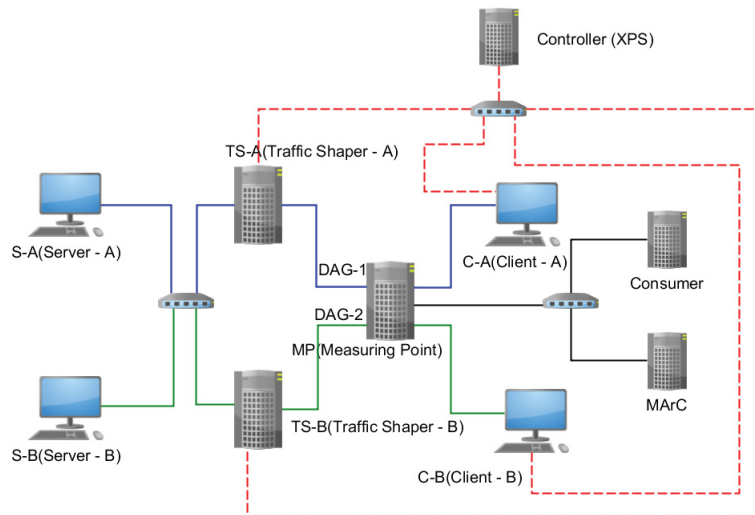

Figure 3.1: Wired Experimental Test-bed

The technical specifications of the components of the wired experimental test bed are shown in the following Table 3.5.

Table 3.5: Technical specifications (Wired Setup)

Devices	Technical Specifications	Description
Server A	HP MicroTower AMD Athlon X64, CPU @ 2.7GHz,2048MB DDR3 SDRAM, Ubuntu-12.04	Flumotion VoD Server
Server B	HP MicroTower AMD Athlon X64, CPU @ 2.7GHz,2048MB DDR3 SDRAM, Ubuntu-12.04	Flumotion VoD Server
Shaper A	Dell Optiplex, Pentium-IV Processor, Ubuntu-12.04	NetEm
Shaper B	Dell Optiplex, Pentium-IV Processor, Ubuntu-12.04	NetEm
Client A	HP MicroTower AMD Athlon X64, CPU @ 2.7GHz,2048MB DDR3 SDRAM, Ubuntu-12.04	VLC-2.0.6 Client
Client B	HP MicroTower AMD Athlon X64, CPU @ 2.7GHz,2048MB DDR3 SDRAM, Ubuntu-12.04	VLC-2.0.6 Client

3.4.2 Wireless Network Scenario

The wireless experimental test bed depicted in Figure 3.1 is used to carry out the experiments. It consists of a Flumotion streaming server (S), a NetEm traffic shaper (TS), a VLC client (C) and RF shielded test enclosure along with WLAN AP. All these components of test bed are connected to the MArC via measurement point as shown in Figure 3.2. A full duplex WLAN test network has been established using a WLAN AP (Access Point). The WLAN AP is enclosed in an RF shielded test enclosure in order to avoid external interference in the network. The traffic shaper acts as a bridge between server and client, where the client system receives the videos streamed from server via the shaper in a wireless network.

Similar to wired scenario, the test network consists of Measurement point (MP) equipped with DAG 3.6E cards which captures the packets using wire-taps as shown in the test bed. The MArC is used to control the Measurements and the whole test network is controlled by a controller (XPS) as shown in Figure 3.2.
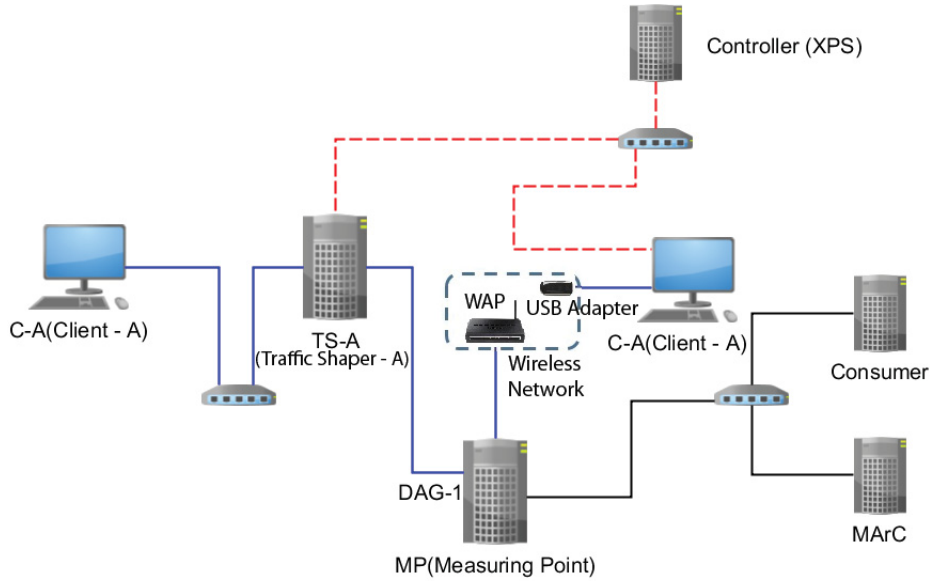
Figure 3.2: Wireless Experimental Test-bed

The technical specifications of the components of the wireless experimental test bed are shown in the following Table 3.6.

Table 3.6: Technical specifications (Wireless Setup)

Devices	Technical Specification	Description
Server	HP MicroTower AMD Athlon X64, CPU @ 2.7GHz,2048MB DDR3 SDRAM, Ubuntu-12.04	Flumotion VoD Server
Shaper	Dell Optiplex, Pentium-IV Processor, Ubuntu-12.04	NetEm
Client	HP MicroTower AMD Athlon X64, CPU @ 2.7GHz,2048MB DDR3 SDRAM, Ubuntu-12.04	VLC-2.0.6 Client
RF Shielded Test Enclosure	Ramsey STE3000FAV	It was used to avoid external signal interference.
WLAN	D-link DAP-1522, Firmware 1.21	Wireless Access point.
USB Network Adapter	D-link DWA-160, Firmware v1.70(E)	Used to connect the Access point and host

3.4.3 Technical Components

Further description of the main technical components of the experiment is elucidated in this section.

Measurement Point

Measurement point (MP) is used to passively capture the packets from the output of the traffic shaper (TS) to the Client (C). It is a Linux based system equipped with Endace (Digital Acquisition and Generation) DAG 3.6E cards. The MP captures the traffic passively with the help of wiretaps. Here, the wiretaps tap the packets, duplicates it and passes it to MP. The filtering of traffic is done according to the filter rules given by the Measurement Area Controller (MArC). The DAG cards are synchronized with respect to time using NTP (Network Time Protocol) server and Global Positioning System (GPS) to achieve time stamp accuracy of 60ns [53].

MArC

MArC (Measurement Area Controller) is the main controller of the measurement area which controls the Measurement points (MPs). It controls the MPs by applying a set of filtering rules, according to which the tapping or capturing of traffic is done [54].

Consumer

Consumer is a system that runs on a Linux based environment. It stores the duplicated packets captured by the DAG cards. It consists of Libcaputils which is used to convert the binary traces obtained from measuring point into text format. In our experiment, the role of the consumer is to copy the obtained traces from local storage to network storage (Database).

Server

Flumotion is an open-source based streaming server. It is a streaming platform that is based on the multimedia framework GStreamer [55]. It supports all the leading audio and video codecs such as H.264, VP8, Xvid, AAC, Ogg Theora, Ogg Vorbis and WMA. The Flumotion server offers good performance, stability and quality. The Flumotion server consists of two components, namely manager and worker. The user interface connects the manager which controls the workers.

In our thesis Flumotion streaming server (version 0.10.1) has been chosen to be used as a video-on-demand (VoD) server. The VoD server streams the content (video files) present on the hard drive and the client can request the content from the streaming server any time. The server is installed on a Linux based operating system (Ubuntu 12.04LTS), which is an HP desktop with AMD CPU @ 2.7GHz 2048 MB DDR3 SDRAM.

Shaper

NetEm traffic shaper has been used to emulate the required network conditions. It has been used to introduce desired packet loss and delay variation parameters in the network. The following tc commands have been used to create the required network conditions.

1. **Packet Loss**

   # tc qdisc add dev ethW root netem loss X%
   # tc qdisc change dev ethW root netem loss X%
   where W is the Ethernet card number and X is the desired packet loss value in %.

2. **Packet Delay Variation**

   # tc qdisc add dev ethW root netem delay Yms Zms
   # tc qdisc change dev ethW root netem delay Yms Zms
   where W is the Ethernet card number, Y is the fixed delay values in [ms] and Z is the delay variation in [ms].

   The NetEm traffic shaper runs on a Linux based operating system (Ubuntu 12.04LTS), which is a Dell optiplex desktop with Pentium IV processor. The packet loss values used for the experiment are 0%, 2.5%, 5%, 10% and 15% respectively, and the delay (D) and delay variation (ΔD) values are D $\pm\Delta$D = 150ms $\pm$ {50ms, 90ms, 110ms, 130ms, 150ms}.

Client

The client is a Linux based desktop computer which runs on Ubuntu 12.04 operating system. The VLC media player version 2.0.6 is installed on the client side which plays the streamed videos and saves the video files to the network storage.

   The Jitter buffer of VLC player is set to zero value to see the immediate effect of network impairments [44]. In our thesis, VLC player is used to receive HTTP/TCP stream from the server and to save the received video sequences to the local database. The VLC player does not support packet reordering, if the packets arrive late, they will be discarded. VLC player has been used through command line interface to save the streamed videos.

3.4.4   Experimental Procedure

The emulated networks are established as shown in Figure 3.1 and Figure 3.2 respectively to conduct the experiments in a controlled environment. The VoD streaming server (Flumotion) is responsible to stream the encoded videos via HTTP/TCP connection to the client (VLC player). The video

sequences are streamed from the server to the client via traffic shaper which is placed between server and client. The traffic shaper (Netem) is used to control the packet loss and variable delay shaping of traffic from server to client in the network. It introduces the desired packet loss and delay variation values in the network. The video files received at the client side are saved in the network storage.

Distributed Passive Measurement Infrastructure (DPMI) [53] is used, which is connected to the test network. During the experiment, the network traffic between Traffic shaper and the client is passively captured by the measurement points using wiretaps. The packet traces tapped by the measurement point are stored into the network storage.

A vast series of experiments were conducted to investigate the video performance through an automated setup (available in APPENDIX B) which was controlled by the controller which ran on Perl control script. In the experimentation, for each shaping parameter the video has been streamed for a series of forty times in order to achieve acceptable data. The streamed videos (with artifacts) are saved on the client host and corresponding network traces are captured at MP. The different videos collected at the client side for varying packet loss and delay variation are video sequences for different video codecs. The videos collected at the client side are further assessed using objective video quality assessment and subjective video quality assessment as described in section 3.6. Apart from objective and subjective assessment, the network level analysis is done for the selected videos. The respective stored network traces were selected and analyzed using a network analysis tool namely Captcp tool [56].

3.5 Data Collection

A limited number of videos were selected from the vast dataset obtained at the client side. The selection of videos was done on the basis of SSIM values obtained for the whole video dataset. Firstly, SSIM was calculated for each and every video. Then one video was stelected from the set of forty videos for each shaping parameter. The criterion used was that the SSIM value of the video to be selected must be near or equal to the average SSIM of a set of 40 videos (for each shaping value). Thus, the selected videos were used for subjective evaluation where these selected video set has been shown to the users to obtain the respective MOS ratings of the videos.

3.6 Assessment Methodology

This section presents the assessment methodology used in our thesis to investigate the performance of the video codecs. The assessment methodology is performed according to three different perspectives as shown below:

1. Objective video quality assessment

2. Subjective video quality assessment

3. Network level analysis

Objective Video Quality Assessment

Though subjective video assessment is a reliable means of video quality evaluation based on the perceived quality, they lack in terms of time consumption and expenditure. Thus, objective video quality assessment methods are used. The Objective video assessment method is based on mathematical calculations. These metrics produce the video quality scores that reflect the perceived video quality. These objective values are compared to the degree of closeness with the subjective user perceptions.

In our thesis SSIM (Structural similarity) metric has been chosen for the objective assessment of video quality. SSIM is based on the Human Visual System (HVS), which measures the structural similarity at frame level. SSIM prediction is known for its degree of closeness with respect to subjective video quality, as it follows a different approach when compared to other objective metrics. SSIM considers the quality degradations in the frames depending on the structural similarity of the two compared frames.

In our thesis, Evalvid2.7 has been used for objective video quality assessment. Evalvid is a framework and a tool-set used for assessment of quality of videos transmitted over a network [57]. It is a tool developed specifically for researchers who want to evaluate the video quality. It supports various video codecs such as H.263, MPEG-4, H.264, Xvid and VP8. It supports standard metrics such as PSNR and SSIM.

Evalvid has been used to generate the SSIM Values for the video sequences collected from the client. The Generated SSIM ratings for the videos are used for the analysis of video quality. The SSIM Scores obtained are correlated with subjective MOS ratings.

As neither Objective method nor subjective method alone can give a complete picture of video quality, thus both the methods are necessary to be included in the video assessment process in order to get reliable results [58].

Subjective Video Quality Assessment

The subjective video quality assessment is used to serve as a benchmark for the performance evaluation of objective video quality assessment. The perceptual video quality predicted by the objective assessment is compared for the degree of closeness with the perceptual quality obtained from the subjective video quality assessment.

As per the ITU-T recommendations, we have used the ACR 5-grade scale (ACR5). ACR is the method where the test sequences are presented one at

a time as shown in Figure 3.3. It uses a 5-grade Scale where each video is rated on a scale of 5 (1- Bad, 2-Poor, 3-Fair, 4-Good, 5- Excellent). Also in [18], the authors have rated ACR5 method as the most suitable method for quality assessment of mobile video services.
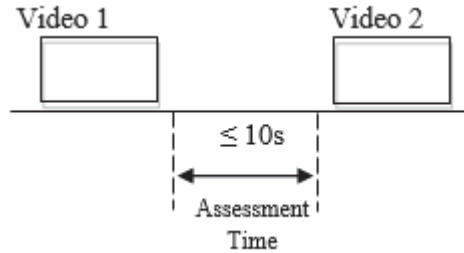

Figure 3.3: ACR 5 Method

The subjective quality assessment test has been performed for the mobile videos. The subjective quality assessment sessions were conducted in the User perception Lab conforming to the specifications of the ITU-T [59] at the premises of Blekinge Institute of Technology, Karlskrona, Sweden. The user study was conducted using 39 participants of which 65% were male and 35% were female. The number of subjects were chosen based on the recommendations of ITU-T [16], according to which the sample size is not recommended to be more than 40 members. Also, author in [60] asserts to use a sample of 25-30 individuals for statistical purpose.

As per the recommendations of ITU-T, the general viewing conditions were strictly regulated. The participants were seated with a viewing distance of 1-8 H in a room with background illumination of 20 lux. All the test conditions as specified by [16] were strictly maintained.

A subjective video quality assessment tool was developed for our thesis as shown in Figure 3.4 and 3.5. The tool was designed to have a graphical interface (Front end) to the users using which the users could view the video sequences and give their desired rating for the video quality. When the user clicks on the opinion score and submits the rating, the corresponding values are automatically stored in the database (Back end). The collected values from the database were used for the subjective assessment of the videos. The tool was specifically designed adhering to the recommendations of ITU-T [16], to implement ACR 5 subjective assessment method for videos.

The assessment session was divided into two phases, where initially the participants were given a brief introduction of the subjective assessment test and usage of the subjective assessment tool. Then they were asked to fill the online questionnaire at the beginning of assessment session and then in the next phase the videos were shown to users where they would rate the videos based on their perception of the video quality on a 5-grade scale.

Finally, the perceptual video quality ratings obtained from the subjects were averaged to obtain the Mean Opinion Score (MOS).

Figure 3.4: MOS Questionaire Screen Shot

Figure 3.5: MOS-Tool Screen Shot

Network Level Analysis:

In addition to objective level and subjective level assessment, we have analyzed the traffic over network level to calculate throughput and also to verify the applied delay variation and packet loss values. The measurement point with the help of wiretaps is used to passively capture the packets from the client-server communication on the network level. These network traces are stored locally on the consumer system with a cap format. The network level traces (cap files) obtained from the MP (Measurement Point) for selected videos are converted into Pcap, using Cap2pcap and those obtained Pcap files are analyzed using a network analysis tool, namely Captcp (APPENDIX D). Captcp is an extensive TCP based network protocol analyzer.

In the network level analysis, the application throughput is calculated and analyzed. The analysis is done for all the three video codecs in wireless network scenario. This throughput comparison gives us insights on how the video codecs are performing at network level in wireless network scenarios.

Chapter 4

RESULTS AND DISCUSSION

This chapter explains the detailed description of the obtained results. These results are based on the experiments conducted as described in the previous chapter. The collected data from subjective and objective video quality assessments are investigated for the effect of packet loss and delay variation.

4.1 Subjective Video Quality Assessment

The only reliable method to assess the video quality as perceived by a human observer is to ask human subjects for their opinion, which is termed as Subjective Video Quality Assessment (VQA). Subjective VQA is impractical for most applications due to the human involvement in the process. However, subjective VQA studies provide valuable data to assess the performance of objective or automatic methods of quality assessment. The Subjective VQA is used to compare the visual impression of video sequences based on the perception of users. The results in the following sections show the subjective video quality of three codecs in wired and wireless networks.

As per the ITU-T recommendations, the MOS user interface required for the thesis work has been developed on the Windows platform using C# .NET (available in APPENDIX E). Using this interface, the videos selected for user perception tests based on the SSIM values were shown to the users in a perception lab at BTH, Karlskrona. The respective user ratings were collected and stored in the database. To ensure perfect playback, we have used high performance systems with low applications and services installed, to avoid latencies due to slow access to Hard Disk.

The video sequences were rated on a 5 grade scale Excellent (5), Good (4), Fair (3), Poor (2) and Bad (1). Each video was assessed by 39 human subjects, but only data of 33 subjects were used. Since, Six human subjects within the original sample had to be discarded due to insufficient compliance with the test procedure. Also, some amongst the six discarded users gave strange ratings, such as least MOS ratings for original videos and highest MOS ratings for videos with high packet loss and packet delay variation

values. The average of collected MOS ratings was calculated for each video and the Mean MOS values of five videos (Crew, Foreman, News, Football and Hall-Monitor) for each codec, at each shaping value is calculated and the resulting graphs are plotted.

Based on the recommendation BT. 500 subjective assessment in the quality of television pictures, of the International Telecommunications Union Radio communications Sector (ITU-R) [59], the mean was calculated. The mean is defined as:

$$\bar{X}_{jk} = \frac{1}{N} \sum_{i=1}^{N} X_{ijk} \tag{4.1}$$

4.1.1 Comparison of H.264, Xvid and WebM/VP8 in Wired and Wireless Networks

Packet Loss

For all the videos, subjects feel that videos have better quality at low packet loss values. It is observed that as the packet loss value increases, the quality of video degrades steeply. The videos streamed without artifacts was graded with a MOS rating that lies between excellent (5) and good (4). This shows that the viewers in some cases are reluctant to score the videos as excellent. This behavior was also found in other studies [61].

Figure 4.1 represents the behavior of H.264, Xvid and WebM with respect to different packet loss values in the wired network scenario and Figure 4.2 represents the behavior of H.264, Xvid and WebM with respect to different packet loss values in wireless network scenario. For wired and wireless networks, the average MOS ratings of three codecs are plotted against respective packet loss values. In wired and wireless network scenarios, it can be observed that at 0% and 2.5% packet loss, user ratings for all the videos are high as there are no human perceptible artifacts in the videos. At 5% and 10% packet loss values the user MOS ratings for H.264 are higher than WebM and Xvid. But at 15% packet loss value, WebM has been rated higher than H.264 and Xvid. And for 5%, 10% and 15% packet loss values, Xvid has the least MOS rating when compared to WebM and H.264. These observations reveal that, subjects felt H.264 encoded videos had better quality at lower packet loss values and WebM encoded videos had better resistance for higher packet loss values. However the Xvid encoded videos also had similar quality in comparison with H.264 and WebM for low packet loss values, but these videos are feeble for moderate and high network disturbances.
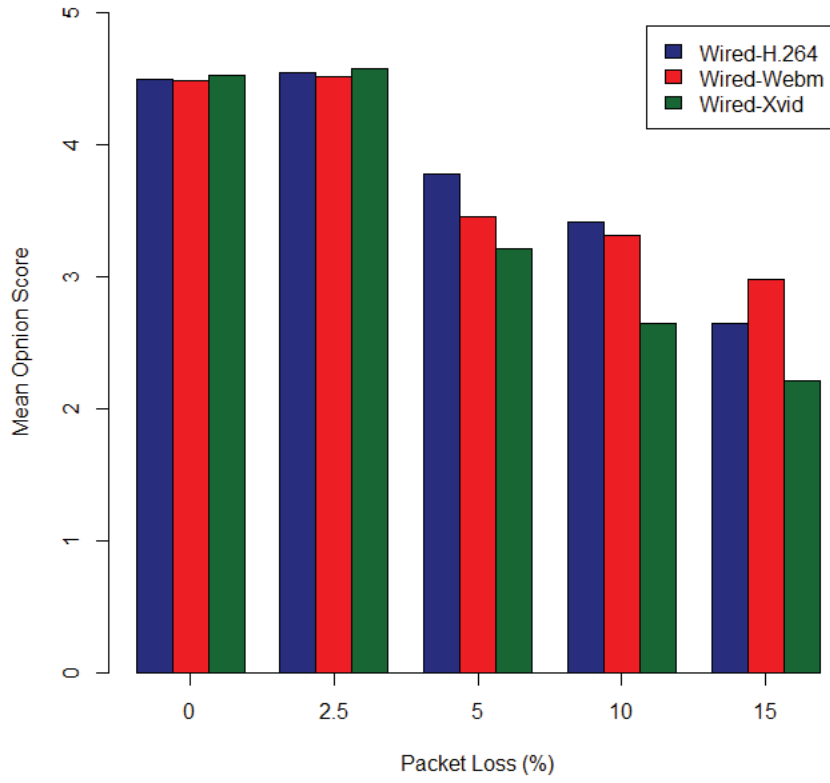
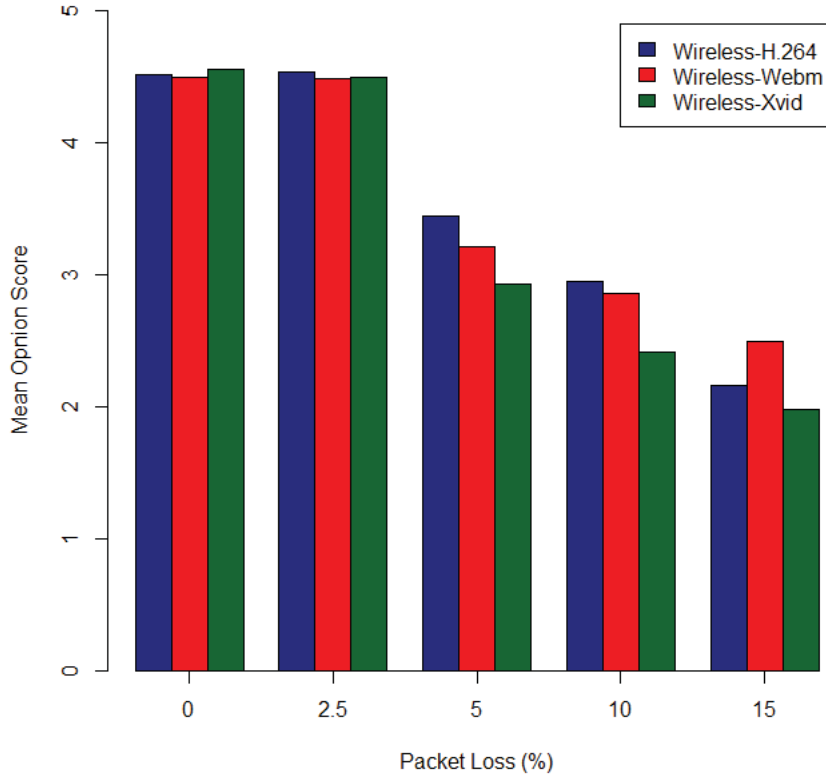**Figure 4.1:** Average MOS Ratings for Packet Loss in Wired Network

Figure 4.2: Average MOS Ratings for Packet Loss in Wireless Network

Packet Delay Variation

Figure 4.3 represents the behavior of H.264, Xvid and WebM/VP8 with respect to different packet delay variation values in the wired network scenario and Figure 4.4 represents the behavior of H.264, Xvid and WebM/VP8 with respect to different packet delay variation values in wireless network scenario. It can be observed that at 150±50 and 150±90, H.264 has MOS ratings in between good (4) and fair (3) whereas, WebM and Xvid are in between fair (3) and bad (2). At 150±110, 150±130 and 150±150 the WebM has better MOS ratings than H.264 and Xvid. This shows a similar trend to that of packet loss where the subjects feel that H.264 encoded videos has better quality at lower packet delay variation values and WebM encoded videos has better resistance for higher packet delay variation values. However, the Xvid encoded videos are feeble to network disturbances.
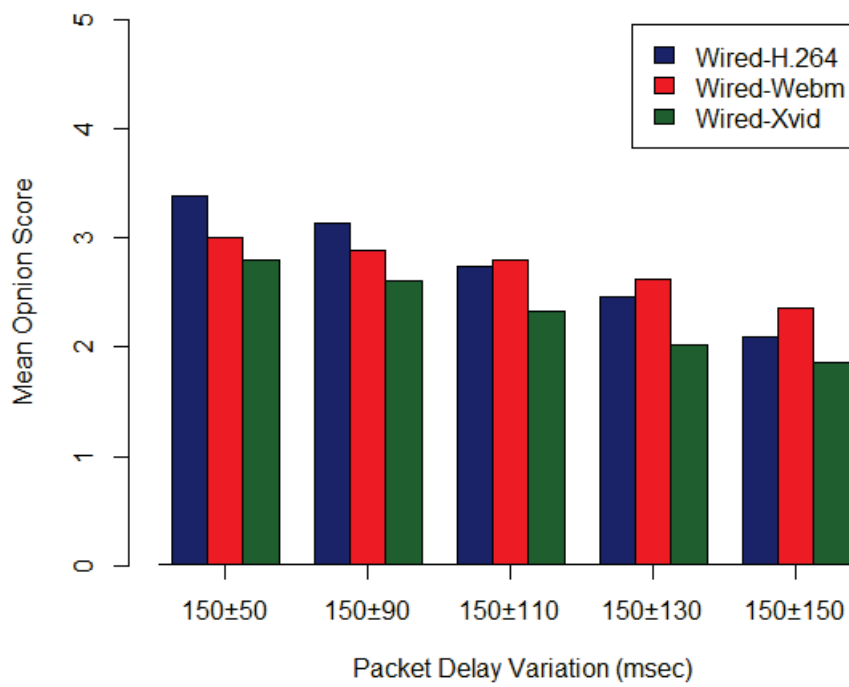
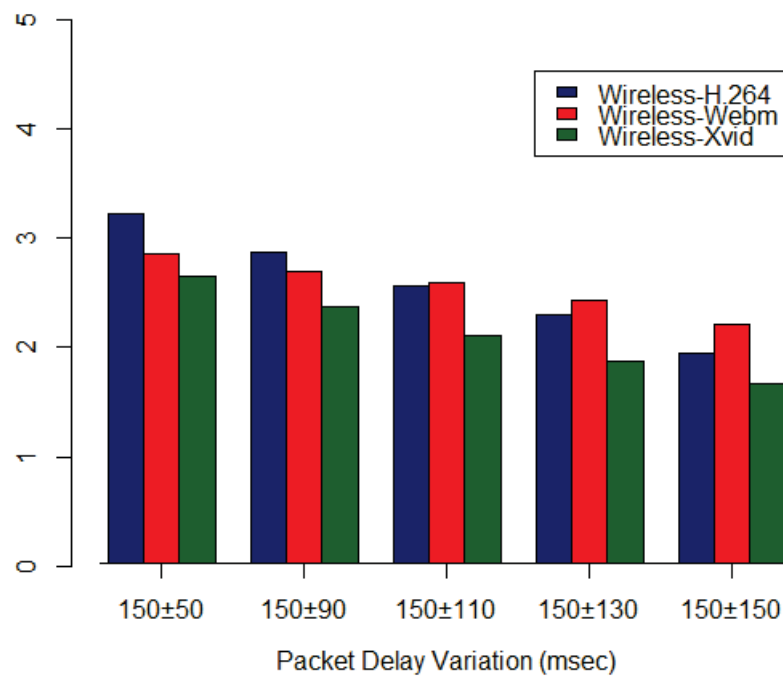*Figure 4.3:* Average MOS Ratings for Packet Delay Variation in Wired Network

Figure 4.4: Average MOS Ratings for Packet Delay Variation in Wireless Network

4.1.2 Overall Comparison of Video Codec Performance in Wired and Wireless Networks

The MOS for the packet loss and delay variation was plotted for Wired and Wireless networks in Figure 4.5 and Figure 4.6 respectively. From these graphs, we can notice that the results are almost similar for wired and wireless networks. The difference found is that the MOS ratings for videos streamed in wired network has a slightly higher MOS rating than that of videos streamed in a wireless network. Also, we can notice that the difference between wired MOS ratings and corresponding wireless MOS ratings are almost same at 0% and 2.5% packet loss, as there are no human perceptible artifacts in the streamed videos.
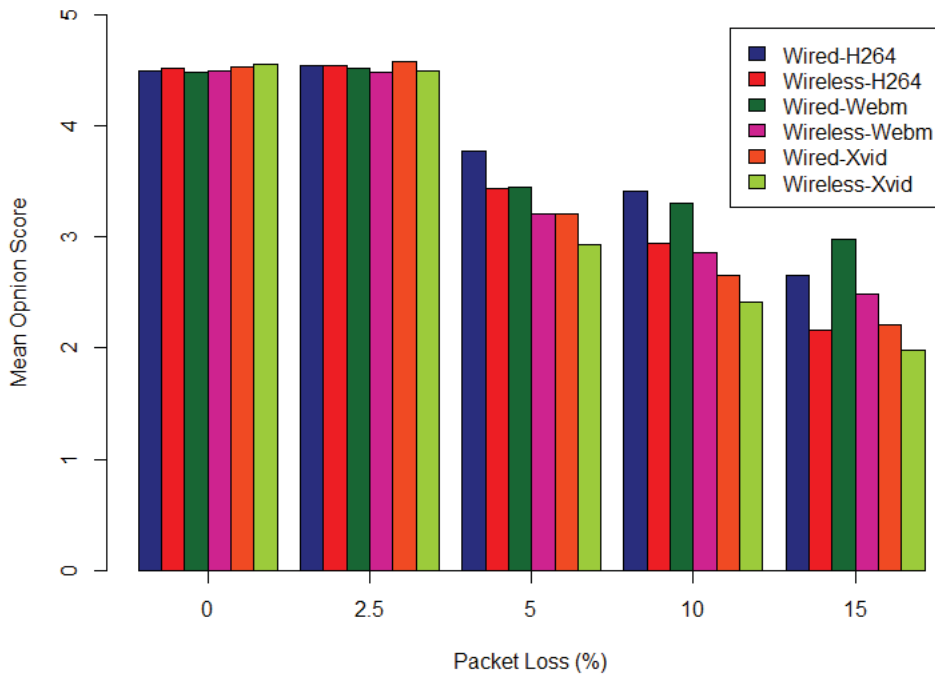

Figure 4.5: Network Based Comparison for Packet Loss with respective MOS Ratings
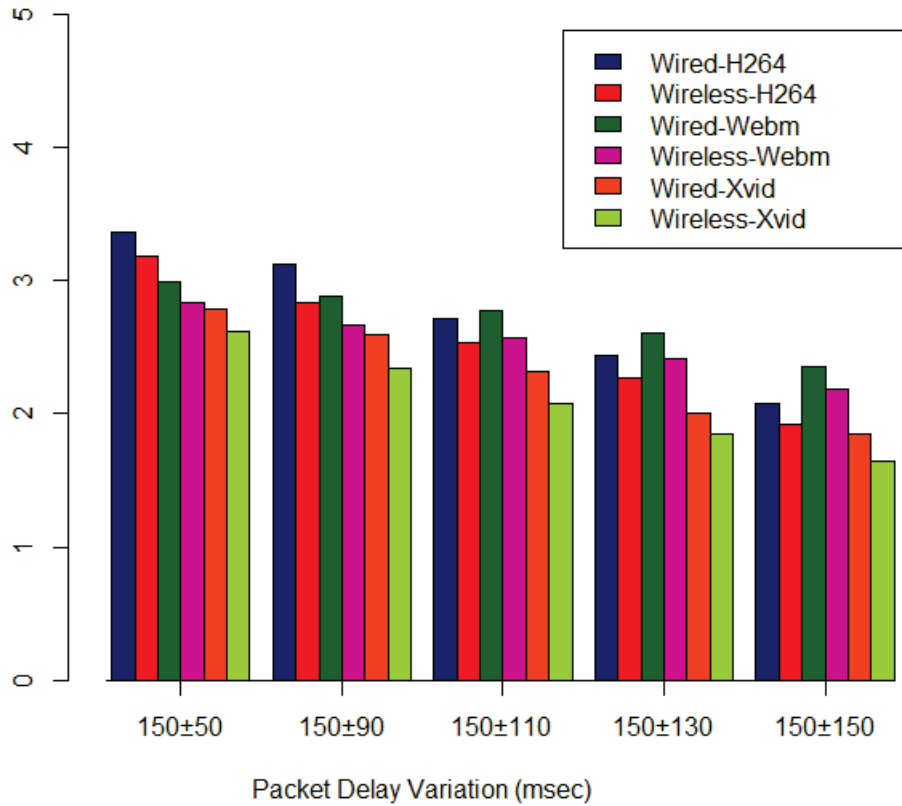
Figure 4.6: Network Based Comparison for Packet Delay Variation with respective MOS Ratings

4.2 Objective Video Quality Assessment

The performance of a perceptual video quality metric depends on its correlation with subjective results. The objective assessment is used for comparing the degree of closeness with the perceptual quality obtained from subjective methods. The SSIM is used as a performance metric for objective video quality assessment. SSIM index or structural similarity index is a popular method that is used for measuring the similarity between two still images [33], [62] and extended to video in [63]. The SSIM index was applied frame-by-frame on the luminance component of the video and the overall SSIM index for the video was computed as the average of the frame level quality scores. The Evalvid [64] tool has been used to get the SSIM index values (provided in APPENDIX C) of each frame in the video.

4.2.1 Comparison of H.264, Xvid and WebM/VP8 in Wired and Wireless Network

Packet Loss

Figure 4.7 represents the behavior of H.264, Xvid and WebM/VP8 with respect to different packet loss values in the wired network scenario and Figure 4.8 represents the behavior of H.264, Xvid and WebM/VP8 with respect to different packet loss values in wireless network scenario. For wired and wireless networks streamed videos, the average SSIM value of three codecs is plotted against respective packet loss values. In wired network it can be observed that at 0% and 2.5% packet loss, the SSIM values for all the videos are high i.e., 1 as there are no artifacts in the videos. But at 2.5% packet loss in wireless network, there is a slight degradation in the SSIM value. At 5% packet loss values the SSIM value of H.264 is higher than WebM and Xvid. But at 10% and 15% packet loss values, WebM has higher SSIM value than H.264 and Xvid. And for 5%, 10% and 15% packet loss values, Xvid has the least SSIM value when compared to WebM and H.264. We noticed that SSIM of WebM was changed very slightly as it has single stuck with change in duration.
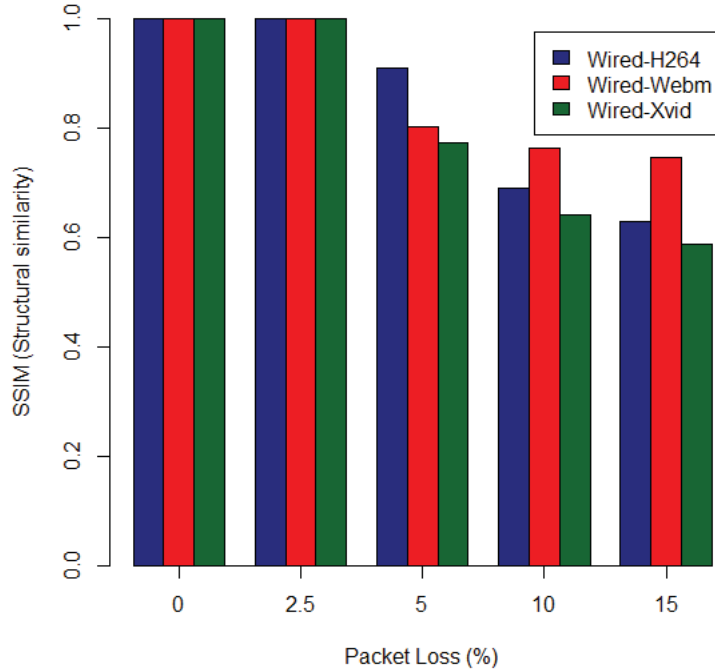

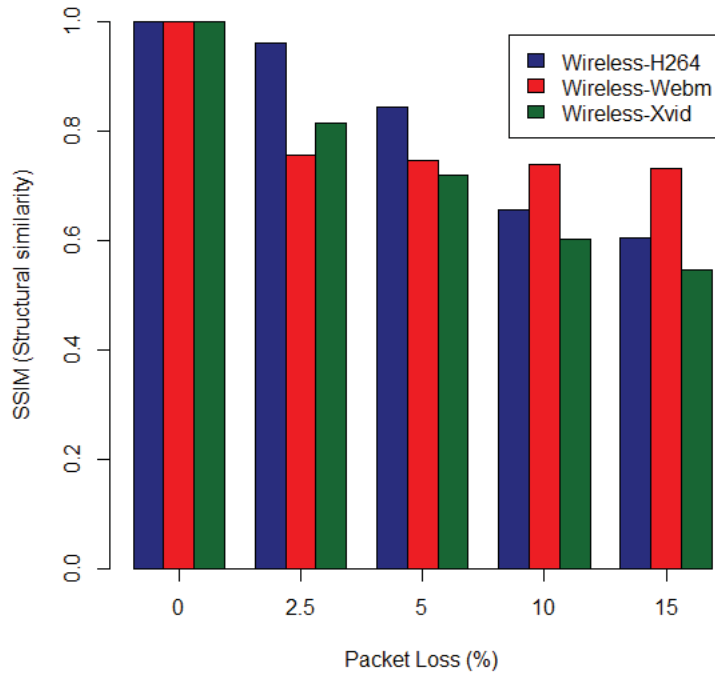***Figure 4.7:*** SSIM for Packet Loss in Wired Network

Figure 4.8: SSIM for Packet Loss in Wireless Network

Packet Delay Variation

Figure 4.9 and 4.10 shows the plotted results of SSIM for delay variation in wired and wireless network, respectively. In wired and wireless networks, at 150±50ms H.264 and WebM have higher value than Xvid. As the distortion in the network increases, the WebM has higher SSIM value than H.264 and Xvid.
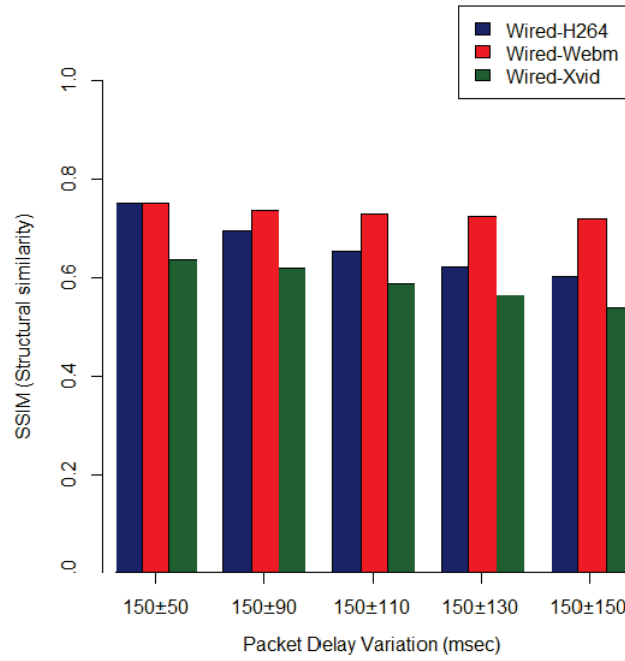
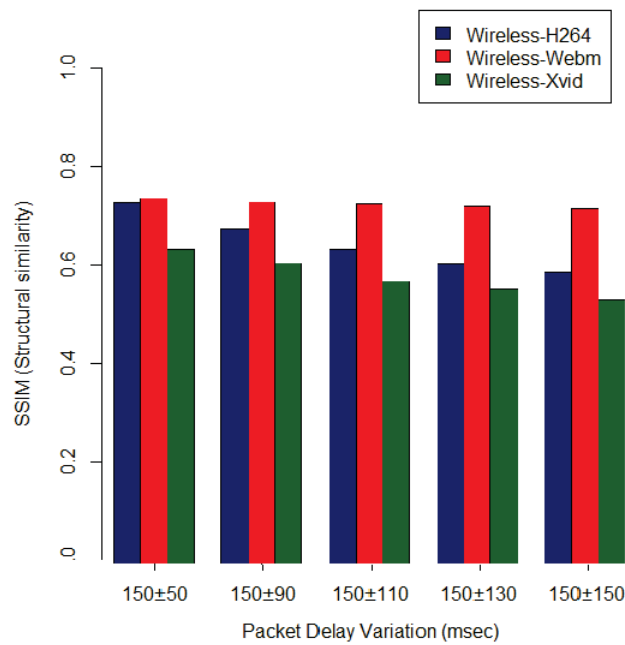*Figure 4.9:* SSIM for Packet Delay Variation in Wired Network

Figure 4.10: SSIM for Packet Delay Variation in Wireless Network

4.2.2   Overall Comparison of Video Codec Performance in Wired and Wireless Networks

Packet Loss

The SSIM value for the packet loss and delay variation was plotted for Wired and Wireless networks in Figure 4.11 and Figure 4.12 respectively. From these graphs we can notice that the results are almost similar for wired and wireless networks. The dierence found is that the SSIM ratings for videos streamed in wired network has a slightly higher rating than that of videos streamed in wireless network. These graphs reports that, behavior of results obtained from the MOS ratings (Subjective) are similar to the behavior of results from the SSIM values (Objective), except at wireless packet loss 2.5%, delay variation 150±50ms and 150±90ms. The dissimilarity for these values is because the SSIM was calculated at frame level by a machine where every distortion can be notified, but coming to MOS the videos which are moving pictures, were rated by humans where some distortion cannot be notified. As those minor distortions are not perceptible to the human eye and also the perception of each human is undefined.
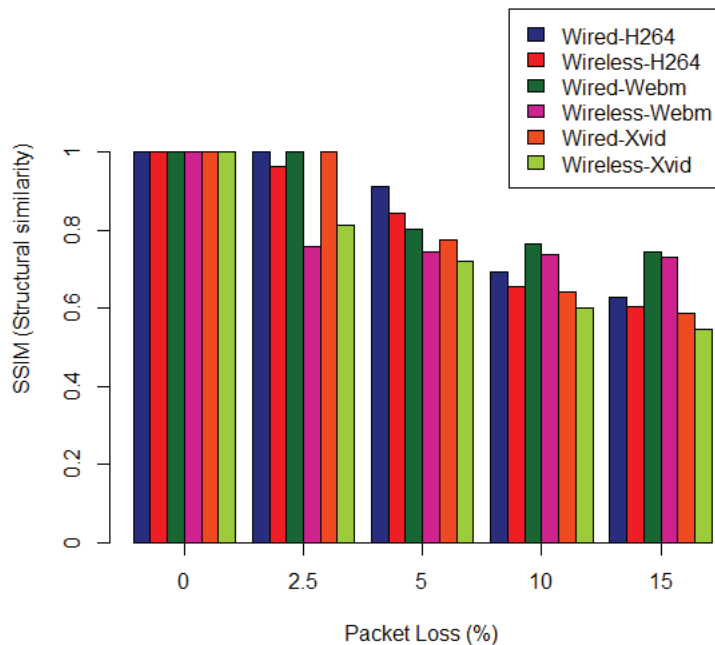

Figure 4.11: SSIM based comparison for Packet Loss in Wired and Wireless Network
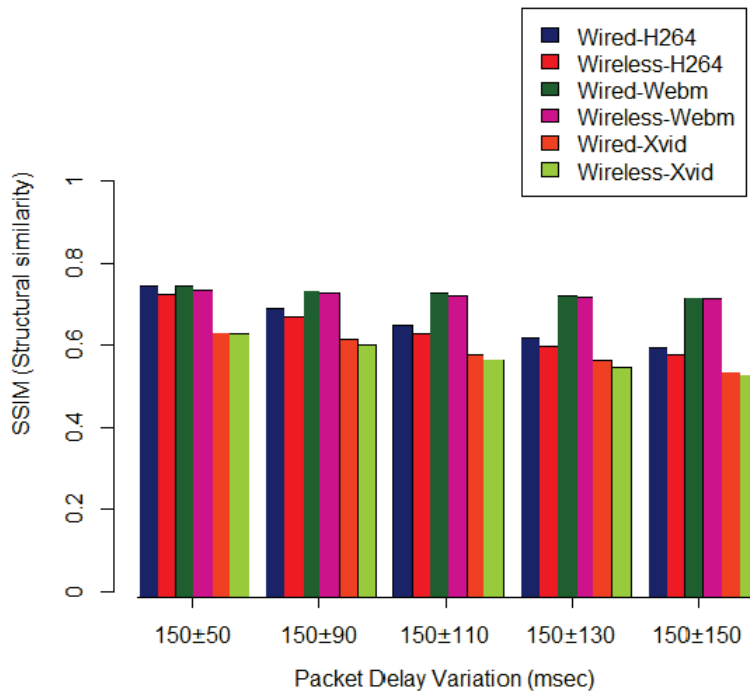
Figure 4.12: SSIM based comparison for Packet Delay Variation in Wired and Wireless Network

4.3 Network Trace Analysis using Application Throughput

The application throughput or goodput is the rate of successful bits received at application level. During the experiment we captured the network packets with the help of measuring point as described in previous chapter. We have then analyzed these packets with captcp tool to get the application throughput.

Figure 4.13 and 4.14 represents the behavior of application throughput of the network for videos streamed with respect to different packet loss and delay variation values in wireless network. From the graphs we can observe that as the packet loss or delay variation values increases the throughput decreases, but the behavior of throughput was not coordinated with the behavior of subjective or objective video quality assessment results. The reason is, In general low throughput resembles low network quality due to missing packets but in video perception the conventional network metric such as throughput may not judge the quality of video in a network, because in a network the quality of a video depends on many factors such as coding rate, type of codec and content of video [65], [66] and rather than the number of frames missed,

which type of frame was missed i.e., an I-frame or P-frame or B-frame has
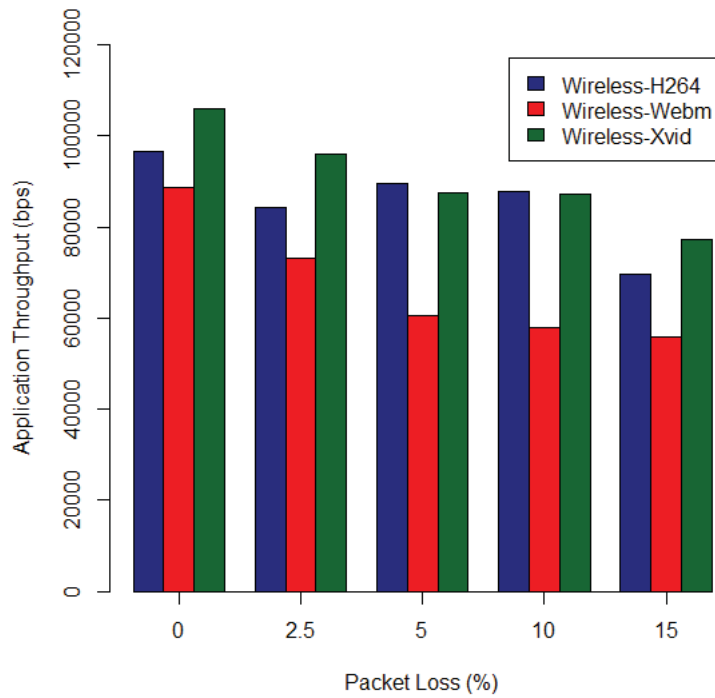more effect on the video quality.

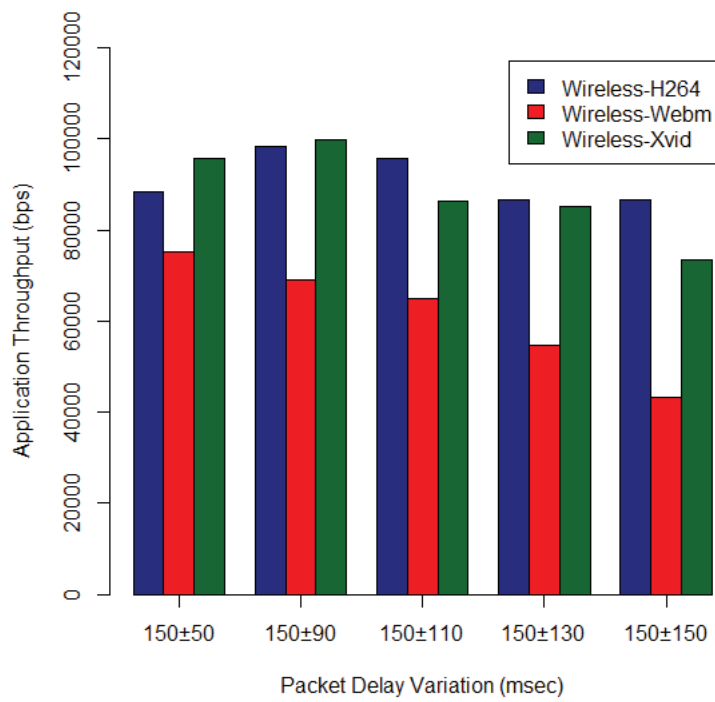***Figure 4.13:*** Application throughput for Packet Loss in Wireless Network

Figure 4.14: Application throughput for Packet Delay Variation in Wireless Network

Chapter 5

CONCLUSIONS

5.1 Conclusions

In this thesis work we investigated the video quality of experience of videos encoded with H.264, WebM and Xvid. The experimental results of our thesis work show the performance analysis of H.264, Xvid and WebM/VP8 encoded video sequences based on subjective assessment and objective assessment methodologies.

An emulated network setup was created in order to obtain streamed video sequences for video quality assessment. The video sequences were introduced with varying packet loss and delay variation values in the experimentation. The videos thus obtained were evaluated using subjective and objective assessment methodology. In subjective video quality assessment, the user experience survey has been performed. In the survey, the users perception of the video quality with respect to video impairments were collected. The collected perceptual video quality ratings from users were averaged to obtain the Mean Opinion Score (MOS). In objective quality assessment, the SSIM (Structural Similarity) performance metric has been used to calculate the video quality scores. The SSIM video quality scores have been generated using Evalvid tool. The obtained video quality scores are compared with that of subjective MOS values to get an understanding on the video quality.

Our first research question deals with the impact of packet loss and delay variation on the user perceived quality of the videos encoded using H.264, Xvid and WebM. This research question was divided into three sub-questions.

The research question RQ1.1, deals with how the user perceives the video quality with respect to packet loss and delay variation for the videos encoded using the Xvid video codec. We can observe that as the packet loss values increase the user perception ratings decreases. In case of low packet loss values Xvid is rated between good and excellent, As the packet loss increases, Xvid is given least ratings when compared to that of H.264 and WebM/VP8. Where as in the case of delay variation, Xvid is rated between fair and good for low delay variations and for higher delay variation the ratings decrease

too drastically when compared with the other two codecs.

The research question RQ1.2, sought to understand how users perceive the video quality in the presence of packet loss and delay variation for videos encoded using H.264. The subjective ratings for H.264 varies significantly. H.264 performs better for lower loss values when compared to higher packet loss values. H.264 also has been rated between good and excellent for lower packet loss values, but as the packet loss value increases the user ratings decreases but are higher than that of Xvid. In the case of delay variation H.264 gets higher ratings for lower values of delay variation, when compared to other two codecs. And for higher values of delay variation, H.264 performance is higher than that of Xvid but is lower than WebM/VP8.

The research question RQ1.3, sought to understand how users perceive the video quality for WebM/VP8 encoded video in the presence of packet loss and delay variation. For low packet loss values, WebM/VP8 performs similar to that of other two codecs, but for higher packet loss values WebM/VP8 comparatively performs best amongst the three codecs.

Our second research question deals with the impact of packet loss and delay variation for streamed videos in two different networks. This research question, sought to understand how the user perceives the quality of encoded videos that are streamed over wired network and wireless network. The user perception ratings for videos that were streamed over wired network were given higher ratings than that of wireless network for higher packet loss and delay variation values. Whereas, for lower packet loss and delay variation values, wired network performed similar to that of wireless network.

Apart from subjective results, the results obtained from the objective assessment were used to compare the degree of closeness with the subjective results. The results obtained from the objective assessment were mostly similar to that of subjective results. The findings from our thesis work reveal that H.264 exhibits better performance for lower values of delay variation and WebM performs better than other two codecs at higher values of delay variation. Whereas for low packet loss values the three codecs showed similar performance. And as the packet loss value increased, H.264 initially showed better performance than other two. But at higher values of packet loss, WebM/VP8 showed better performance than the other two codecs. Furthermore, our findings provide support that video codecs also have an impact on Quality of Experience.

5.2 Future Work

Further research needs to be done based on the comparison of GOP structures of these video codecs. The comparison of video codecs can be extended to different video streaming protocols. The network based comparison of codecs can be extended to codecs such as H.265 and VP9.

Appendix

APPENDIX A

Encoding commands

H264

ffmpeg -i <input.y4m> -vcodec libx264 -b:v 768000 -r 30 -profile:v baseline -level:v 1.3 -s 320X240 -aspect 4:3 -f output.mp4 <encoded.mp4>

Xvid

ffmpeg -i <input.y4m> -vcodec libxvid -b:v 768000 -r 30 -s 320X240 -aspect 4:3 -f mp4 <output.mp4>

Webm

ffmpeg -i <input.y4m > -vcodec libvpx -b:v 768000 -r 30 -s 320X240 -aspect 4:3 -f webm <output.webm>

APPENDIX B

Perl script for VLC automation

```perl
#!/usr/bin/perl/
#Incl. modules
use File::Copy;
#config
$VLCbin="/home/rvt-server/vlc-2.0.6/vlc";
#end config
$content = shift(@ARGV);
$logdir  = shift(@ARGV);
$experiment_run_num
$platform=`uname -a`;
        if ($platform =~ m/Linux/ ){
                printf "OK platform ; $platform . \n";
        } else {
                printf "Platfrom $platfrom not supported. \n";
                print "FAILURE";
                exit(0);
                        }
#Check if custom VLC is found.
#START VLC
`$VLCbin -vvv $content --
sout="#duplicate{dst=std{access=file,,dst='$experiment_run_num.mp4'},dst=display}" --play-and-exit`;
#copy files.
`echo "NEW EXPERIMENT" >> $logdir/vlc.log`;
`cat vlc.log >> $logdir/vlc.log`;
`echo "NEW EXPERIMENT" >> $logdir/vlcr.log`;
`cat vlcr.log >> $logdir/vlcr.log`;
`echo "NEW EXPERIMENT" >> $logdir/lossframes.log`;
`cat vlcs.log >> $logdir/lossframes.log`;
#`rm vlc.log vlcr.log`;
print "SUCCESS";
```

APPENDIX C

SSIM Commands

To calculate SSIM, the following command is used in evalvid:

$$psnr\ x\ y\ \text{<YUV format> <src.yuv> <dst.yuv> [ssim] > output.txt}$$

Here, x is the frame width
y is the frame height
YUV format: 420, 422, 444 etc.
src.yuv is the source video or the original video
dst.yuv is the distorted video
ssim is to be specified for calculating SSIM

APPENDIX D

Captcp Command

captcp statictis <file.pcap> > outputfile.txt

APPENDIX E

.NET Application for User Interface

The MOS user interface required for the thesis work has been developed on the Windows platform using C# .NET and the pre-compiled .NET code for MOS user interface is available at

`https://github.com/vipo12-raa12-100041/MOS-User_Interface-for-Windows/`

Bibliography

[1] "Cisco Visual Networking Index: Forecast and Methodology, 2012-2017 [Visual Networking Index (VNI)] - Cisco Systems." `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html`. [Online; Accessed: 09-Nov-2013].

[2] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017 - Cisco Systems," [Online; Accessed: 18-Nov-2013].

[3] T. N. Minhas and M. Fiedler, "Quality of experience hourglass model," in *Computing, Management and Telecommunications (ComManTel), 2013 International Conference on*, pp. 87–92, IEEE, 2013.

[4] "H.264 : Advanced video coding for generic audiovisual services." `http://www.itu.int/rec/T-REC-H.264-201304-I/en`. [Online; Accessed: 23-Sep-2013].

[5] O. Nemcic, M. Vranjes, and S. Rimac-Drlje, "Comparison of h. 264/avc and mpeg-4 part 2 coded video," in *ELMAR, 2007*, pp. 41–44, IEEE, 2007.

[6] J. Bankoski and P. Wilkins, "Yaowu xu;," technical overview of vp8, an open source video codec for the web",," in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pp. 1–6.

[7] A. C. Begen, T. Akgul, and M. Baugher, "Watching video over the web: Part 1: Streaming protocols," *Internet Computing, IEEE*, vol. 15, no. 2, pp. 54–63, 2011.

[8] "YouTube." `http://www.youtube.com/`. [Online; Accessed: 06-Oct-2013].

[9] "Choose the Vimeo membership thats right for you." `http://www.vimeo.com`. [Online; Accessed: 14-Oct-2013].

[10] "Dailymotion." `http://www.dailymotion.com/`. [Online; Accessed: 26-Oct-2013].

[11] J. Shaikh, M. Fiedler, D. Collange, P. Arlos, and T. N. Minhas, "Inferring user-perceived performance of network by monitoring tcp interruptions," *Network Protocols and Algorithms*, vol. 4, no. 2, pp. 49–67, 2012.

[12] R. K. Mok, E. W. Chan, and R. K. Chang, "Measuring the quality of experience of http video streaming," in *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, pp. 485–492, IEEE, 2011.

[13] S. Hemminger *et al.*, "Network emulation with netem," in *Linux Conf Au*, pp. 18–23, Citeseer, 2005.

[14] T. N. Minhas, M. Fiedler, J. Shaikh, and P. Arlos, "Evaluation of throughput performance of traffic shapers," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, pp. 1596–1600, IEEE, 2011.

[15] J. Shaikh, T. N. Minhas, P. Arlos, and M. Fiedler, "Evaluation of delay performance of traffic shapers," in *Security and Communication Networks (IWSCN), 2010 2nd International Workshop on*, pp. 1–8, IEEE, 2010.

[16] P. ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," 1999.

[17] "Methods for subjective determination of transmission quality." http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P. 800-199608-I!!PDF-E&type=items, 1996. [Online; Accessed: 13-Nov-2013].

[18] T. Tominaga, T. Hayashi, J. Okamoto, and A. Takahashi, "Performance comparisons of subjective quality assessment methods for mobile video," in *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pp. 82–87, IEEE, 2010.

[19] "ITU-T FG IPTV Focus group on IPTV standardization: Definition of quality of experience (QoE)." http://ties.itu. int/ftp/public/itu-t/fgiptv/readonly/PreviousMeetings/ 20070122MountainView/il/T05-FG.IPTV-IL-0050-E.htm. [Online; Accessed: 06-Sep-2013].

[20] P. Le Callet, S. Möller, A. Perkis, *et al.*, "Qualinet white paper on definitions of quality of experience," *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, 2012.

[21] H. J. Kim, D. G. Yun, H.-S. Kim, K. S. Cho, and S. G. Choi, "Qoe assessment model for video streaming service using qos parameters in wired-wireless network," in *Advanced Communication Technology (ICACT), 2012 14th International Conference on*, pp. 459–464, IEEE, 2012.

[22] K. J. Ma, R. Bartos, S. Bhatia, and R. Nair, "Mobile video delivery with http," *Communications Magazine, IEEE*, vol. 49, no. 4, pp. 166–175, 2011.

[23] P. Lambert, W. De Neve, P. De Neve, I. Moerman, P. Demeester, and R. Van de Walle, "Rate-distortion performance of h. 264/avc compared to state-of-the-art video codecs," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 1, pp. 134–140, 2006.

[24] I. E. Richardson, *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*. John Wiley & Sons, 2004.

[25] "Xvid.org: General Info ." `http://www.xvid.org/General-Info.19.0.html`. [Online; Accessed: 09-Oct-2013].

[26] J. Lloret, M. Garcia, A. J. Ferrer, and M. Morant, "Hdtv compression for storage and transmission over internet," in *Proceedings of the 5th WSEAS international conference on Data networks, communications and computers*, pp. 57–62, World Scientific and Engineering Academy and Society (WSEAS), 2006.

[27] R. Koenen, "Profiles and levels in mpeg-4: Approach and overview," *Signal Processing: Image Communication*, vol. 15, no. 4, pp. 463–478, 2000.

[28] C. Feller, J. Wuenschmann, T. Roll, and A. Rothermel, "The vp8 video codec-overview and comparison to h. 264/avc," in *Consumer Electronics-Berlin (ICCE-Berlin), 2011 IEEE International Conference on*, pp. 57–61, IEEE, 2011.

[29] A. Biernacki and K. Tutschku, "Performance of http video streaming under different network conditions," *Multimedia Tools and Applications*, pp. 1–24, 2013.

[30] B. Wang, J. Kurose, P. Shenoy, and D. Towsley, "Multimedia streaming via tcp: an analytic performance study," in *Proceedings of the 12th annual ACM international conference on Multimedia*, pp. 908–915, ACM, 2004.

[31] K. Brunnstrom, D. Hands, F. Speranza, and A. Webster, "Vqeg validation and itu standardization of objective perceptual video quality metrics [standards in a nutshell]," *Signal Processing Magazine, IEEE*, vol. 26, no. 3, pp. 96–101, 2009.

[32] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 165–182, 2011.

[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[34] P. Frohlich, S. Egger, R. Schatz, M. Muhlegger, K. Masuch, and B. Gardlo, "Qoe in 10 seconds: Are short video clip lengths sufficient for quality of experience assessment?," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pp. 242–247, IEEE, 2012.

[35] A. Perkis, S. Munkeby, and O. I. Hillestad, "A model for measuring quality of experience," in *Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic*, pp. 198–201, IEEE, 2006.

[36] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *Network, IEEE*, vol. 24, no. 2, pp. 36–41, 2010.

[37] T. Hosfeld, M. Fiedler, and T. Zinner, "The qoe provisioning-delivery-hysteresis and its importance for service provisioning in the future internet," in *Next Generation Internet (NGI), 2011 7th EURO-NGI Conference on*, pp. 1–8, IEEE, 2011.

[38] S. Jumisko-Pyykkö and J. Häkkinen, "Evaluation of subjective video quality of mobile devices," in *Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 535–538, ACM, 2005.

[39] A. C. Dalal, A. K. Bouchard, S. Cantor, Y. Guo, and A. Johnson, "Assessing qoe of on-demand tcp video streams in real time," in *Communications (ICC), 2012 IEEE International Conference on*, pp. 1165–1170, IEEE, 2012.

[40] Y. Yoon, M. Kim, S. Lee, B. Lee, S. J. Hyun, and K. Lee, "Performance analysis of h. 264/avc, h. 264/svc, and vp8 over ieee 802.11 wireless networks," in *Computers and Communications (ISCC), 2012 IEEE Symposium on*, pp. 000151–000156, IEEE, 2012.

[41] P. Seeling, F. H. Fitzek, G. Ertli, A. Pulipaka, and M. Reisslein, "Video network traffic and quality comparison of vp8 and h. 264 svc," in *Proceedings of the 3rd workshop on Mobile video delivery*, pp. 33–38, ACM, 2010.

[42] Q. Dai and R. Lehnert, "Impact of packet loss on the perceived video quality," in *Evolving Internet (INTERNET), 2010 Second International Conference on*, pp. 206–209, IEEE, 2010.

[43] M. Venkataraman, M. Chatterjee, and S. Chattopadhyay, "Evaluating quality of experience for streaming video in real time," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pp. 1–6, IEEE, 2009.

[44] T. N. Minhas, O. Gonzalez Lagunas, P. Arlos, and M. Fiedler, "Mobile video sensitivity to packet loss and packet delay variation in terms of qoe," in *Packet Video Workshop (PV), 2012 19th International*, pp. 83–88, IEEE, 2012.

[45] P. Calyam, M. Sridharan, W. Mandrawa, and P. Schopis, "Performance measurement and analysis of h. 323 traffic," in *Passive and Active Network Measurement*, pp. 137–146, Springer, 2004.

[46] S.-J. Kim, C.-B. Chae, and J.-S. Lee, "Quality perception of coding artifacts and packet loss in networked video communications," in *Globecom Workshops (GC Wkshps), 2012 IEEE*, pp. 1357–1361, IEEE, 2012.

[47] T. Minhas, M. Shahid, A. Rossholm, B. Lovstrom, H.-J. Zepernick, and M. Fiedler, "Assessment of the rating performance of itu-t recommended video quality metrics in the context of video freezes," in *Telecommunication Networks and Applications Conference (ATNAC), 2013 Australasian*, pp. 207–212, IEEE, 2013.

[48] T. N. Minhas and M. Fiedler, "Impact of disturbance locations on video quality of experience," *Quality of Experience for Multimedia Content Sharing, EuroITV*, 2011.

[49] F. De Simone, M. Tagliasacchi, M. Naccari, S. Tubaro, and T. Ebrahimi, "A h. 264/avc video database for the evaluation of quality metrics," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 2430–2433, IEEE, 2010.

[50] "FFMPEG Documentation." `https://www.ffmpeg.org/documentation.html`. [Online; Accessed: 15-Sep-2013].

[51] T. N. Minhas, "Network impact on quality of experience of mobile video," *Blekinge Institute of Technology*, 2012.

[52] I. Recommendation, "114, one-way transmission time," *International Telecommunication Union*, 1996.

[53] P. Arlos, M. Fiedler, and A. A. Nilsson, "A distributed passive measurement infrastructure," in *Passive and Active Network Measurement*, pp. 215–227, Springer, 2005.

[54] P. Arlos, *On the quality of computer network measurements*. 2005.

[55] "Flumotion: open source streaming media server." `http://www.flumotion.net/`. [Online; Accessed: 16-Dec-2013].

[56] "CAPTCP - TCP Analyzer." `http://research.protocollabs.com/`
`captcp/index.html`. [Online; Accessed: 18-Nov-2013].

[57] J. Klaue, B. Rathke, and A. Wolisz, "Evalvid–a framework for video
transmission and quality evaluation," in *Computer Performance Evalu-
ation. Modelling Techniques and Tools*, pp. 255–272, Springer, 2003.

[58] J. Biström, "Comparing video codec evaluation methods for handheld
digital tv," in *T-111.590 Research Seminar on Digital Media*, 2005.

[59] I. Recommendation, "500-11, methodology for the subjective assessment
of the quality of television pictures," *International Telecommunication
Union, Geneva, Switzerland*, vol. 4, pp. 53–56, 2002.

[60] D. Howell, *Statistical methods for psychology.* Cengage Learning, 2012.

[61] M. H. Pinson, S. Wolf, and G. Cermak, "Hdtv subjective quality of
h. 264 vs. mpeg-2, with and without packet loss," *Broadcasting, IEEE
Transactions on*, vol. 56, no. 1, pp. 86–91, 2010.

[62] Z. Wang and A. C. Bovik, "A universal image quality index," *Signal
Processing Letters, IEEE*, vol. 9, no. 3, pp. 81–84, 2002.

[63] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on
structural distortion measurement," *Signal processing: Image commu-
nication*, vol. 19, no. 2, pp. 121–132, 2004.

[64] V. R. Jonnalagadda and V. Musti, "Evaluation of video quality of ex-
perience using evalvid," *Electrical Engineering*, 2012.

[65] A. Salim and T. O'Farrell, "A novel strategy for data rate adaptation
in ieee 802.11 g in the case of video streaming," in *Computer and Com-
munication Engineering, 2008. ICCCE 2008. International Conference
on*, pp. 812–817, IEEE, 2008.

[66] R. MacKenzie, T. O'Farrell, and S. Abukharis, "Comparison of through-
put with a perceived video quality metric as performance indicators of
streamed video in ieee 802.11 g,"