Effect of Packet Drop and Jitter on Perceived Video Quality for Various Encoded Video over Streaming Network

Dipendra J. Mandal
Department of Electrical & Electronics Engineering
Kathmandu University
Dhulikhel, Nepal
dipendramandal@ku.edu.np

Asst. Prof Subodh Ghimire
Department of Electrical & Electronics Engineering
Kathmandu University
Dhulikhel, Nepal
sghimire@ku.edu.np

Abstract— Transmission of video application over the internet is increasing nowadays. Due to increasing user expectation for high quality multimedia contents, video quality has very important role nowadays. Network characteristics such as packet loss and variation in delayextremely influences the quality of video. Therefore, in this paper, we emulated the effect of packet loss and jitter for different video codec such as H.264 and H.265 to determine the impacts on the received video quality using the objective methods such as PSNR &SSIM. For this, network based emulation was conducted in laboratory using the tools such as OPNET and EvalVid. The result is important in order to understand how above mentioned factors impactthe video quality and also help to choose appropriate delay buffer size and packet repair techniques for various types of video, which will further help to improve the user experience in field of multimedia.

Keywords-*Video Quality; Packet loss; Jitter; PSNR; SSIM*

*****

I. INTRODUCTION

Video has been an important means for communication and entertainment for many decades. Nowadays with the increasing availability of broadband access to Internet, there is an increased interest in web-based and Internet based multimedia application. Digital videos are subjected to wide variety of distortion during the acquisition, processing, compression, storage, transmission and reproduction [1-2]. Dominant network factors that influence the final video quality are especially packet loss, delay variation and the capacity of the transmission link [3]. Therefore, in order to avoid the user distraction while viewing the video, effect of network parameters need to be studied in order to estimate the distortion level in video signals while transmission. In this paper we have presented the effect of packet drop and variation in delay on the quality of video in terms of objective evaluation. The aim of this work is to evaluate the impact of packet losses and jitter in network during the transmission for H.264 and H.265 encoded video.

Analysis of cumulative jitter is important to size the playout buffer for real time streaming video applications [4]. Additional parameter that may affect the quality of the video is the temporal aspect of the video. Video which has very few difference between the frames is considered as low temporal aspect e.g. news video with plain background, whereas video which has more differences between the frames has high temporal aspects such as sports video. Therefore, in order to understand the effect of temporal level on transmitted video under network packet loss and jitter, we have emulated the videos withto different levels (low &high) of motion.

Objective evaluation of video quality can be done by comparing the original or undistorted video (also known as reference video) with the distorted video file. Objective metrics can be classified by the amount of information available about the original signal, the received signal, or whether there is a signal present at all [5]. One of these called Full Reference (FR), where original transmitted file is available to compare. This matrix is usually the most accurate but the computational effort is higher. Another is Reduced Referenced (RR), in this some features from both the transmitted and received videos are extracted and compared. The third matrix is No-Reference, in which quality is evaluated without any information of original file, hence, this is less accurate compared to FR and RR.

We have used the full referenced approached for our work as it is simple and more accurate to predict the video quality. This approach used the Peak Signal to Noise Ratio (PSNR) for video quality measurement. PSNR is the ratio of the power in the signal to the power contained in the noise that is presented at the particular point in the transmission. It is a derivative of the signal to noise ratio (SNR). For convenience, this ratio is often measure in decibels. Equation (1) shows the mathematical expression for PSNR.

$$PSNR = 10.log_{10}\left(\frac{P^2}{\mathrm{MSE}}\right). \qquad (1)$$

Where,
P: is image maximum possible pixel value
MSE: is mean square error between the images, and is expressed as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}.\sum_{j=0}^{n-1}[I(i,j) - I'(i,j)]^2$$

Where,
I (i, j) represent the luminance value where i and j are the row and column index of pixel in the original video frame and I' (i, j) is that of the reconstructed video frame after transmission.
m*n is the dimension of the both the images.

Since PSNR values alone do not correlate well with perceived video quality due to the complex, highly non-liner behavior of human visual system [6], a new matrix called Structural Similarity Index (SSIM) was also computed in this work. It defines the structure of an object's attributes in the frame regardless of the average illuminance and contrast. The main function of the human visual system is to extract structural information from viewing field, and the human visual system is highly adapted for this purpose, therefore a measurement of structural information loss can provide us with a good approximation to perceived video frame distortion

 ISSN: 2321-8169
07 - 11

[7].

The rest of the paper is organized as follows: Section II presents the emulation framework design including the short description of tools used. In section III, we have described the detail experimental setup during the work and analysis of the result obtained. Finally, section IV ends up this paper with summary of the work.

II. EMULATION FRAMEWORK

In this section we have presented the detailed framework and design of emulation environment. As shown in Fig. 1, three computers were used for the experiment, PC 1 and PC 3 is Linux based acting as video server and video client respectively. Network environment was created in the PC 2 having WINDOWS as OS. All three computers are connected with each other using the cross Ethernet cable.

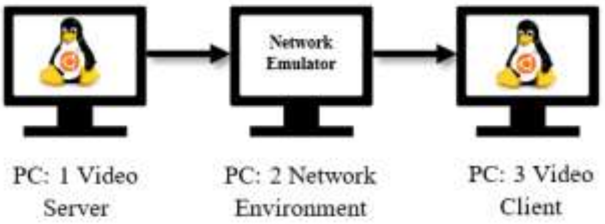

Figure 1. Topology of test Network

OPtimized Network Engineering Tools (OPNET modeler) and Video Evaluation Tool (EvalVid) are the two major tools used to simulate and evaluate the video quality in this experiment.OPNET modeler is one of the leaders in simulation environments specialized for complex modelling and simulation ofcommunications networks, devices and protocols [8].Overall process of the work has been illustrated in Fig.2.

Initially raw video based on eLearning platform was captured using the digital camera and further encoded to H.264 and H.265 using the publicly available ffmpeg encoder. In order to evaluate the video, we need to trace the IP packets

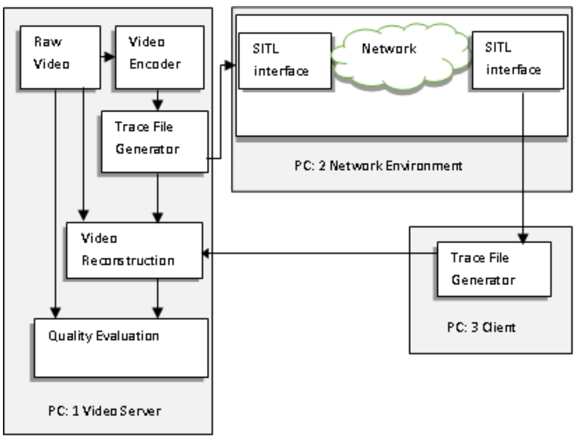

Figure 2. Video Quality Evaluation Methodology

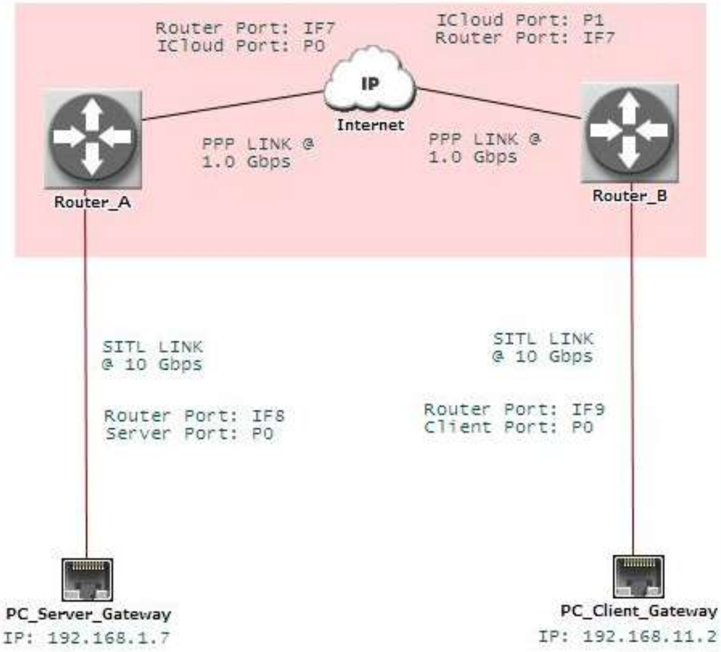

Figure 3. OPNET Simulation Model

at both sender and receiver end, for this we have used the tcpdump, a powerful command-line packet analyzer. Encoded video is streamed to client using the OPNET. Real packet is converted to simulation packet using the interface SITL (System in the Loop), SITL is separately distributed library for OPNET modeler, which provides an interface to real network hardware or software applications to the OPNET discrete event simulator. Simulation packets are then sent to cloud network where parameters such as packet drop ratio and latency level are varied according to the requirement of experiment. Further packets from network is again converted to real packet using another similar interface SITL. Above Fig. 3 is about the detailed simulation model of OPNET.

Trace file generated at both ends using tcpdump tools will be further needed to reconstruct the received video at client end using the tool EvalVid. It is a publicly available framework and tool-set for evaluation of the quality of video transmitted over a real or simulated communication network. Original video (referenced video) file and reconstructed video (distorted video) along with the sender / receiver trace file is required as input for the EvalVid in order to generate the PSNR value and SSIM. Jitter is calculated using the receiver and sender end trace file. Results from the EvalVid is used for the further analysis of the video quality. Fig. 4 shows the input and output for EvalVid block.

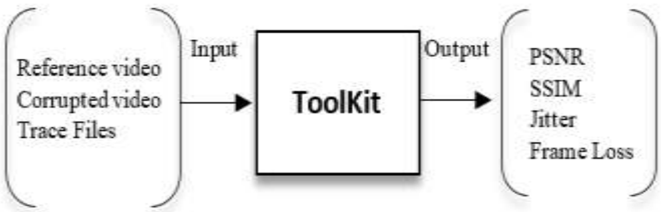

Figure 4. EvalVid I/O Model

III. EXPERIMENT, RESULT AND DISCUSSION

Two categories of video file were used in this experiment; low motion video and high motion video. Objective behind this was to determine the effect of temporal redundancy over the quality matrix. Characteristics of the video sequence is described in the Table 1. Original video sequence for both the video is shown in Fig. 5.

TABLE I. OVERVIEW OF USED VIDEOS

Description	Low Motion Video	High Motion Video
Codec	H.264 / Baseline	H264 / Baseline
Resolution	1280*720	1280*720
Duration	9.74 second	10 second
Temporal Variation	Low	High
Avg. Bitrate	23168 kb/s	22349 kb/s
Frame Rate	29.97 fps	29.97 fps
Size	28.2 MB	28 MB

Figure 5. (a): frame of Low motion video and (b) frame of high motion video

Fig. 6 shows the effect of packet drop over PSNR for both high and low level of temporal variation in video. It illustrates that for each level of packet drop, low temporal variation video has better PSNR over high temporal variation video. Same scenario using SSIM index have been plotted in Fig. 7, the result is similar with the case of PSNR.

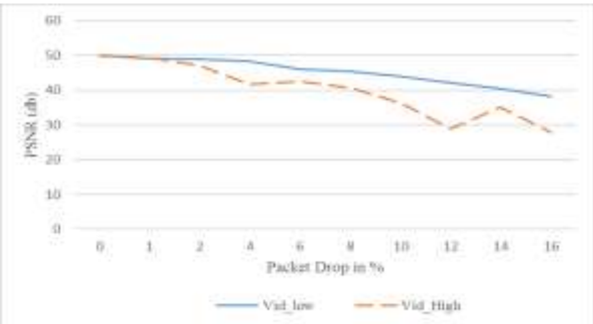

Figure 6. Packet drop Vs PSNR (low and High Motion Video)

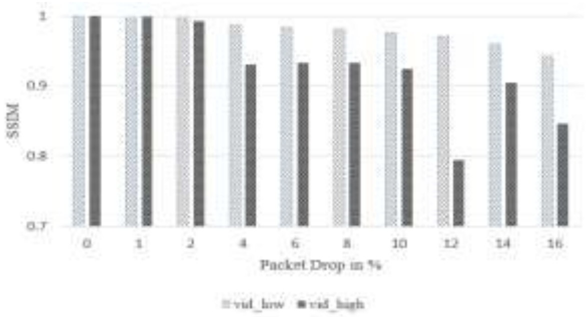

Figure 7. Packet drop Vs SSIM index (low and High Motion Video)

In order to understand the effects of different codecs on the quality of the video over the network, we further encoded the video sequence using the H.264 with profile constrained baseline, H.264 with profile high and H.265.Table II. shows the characteristics of encoded video sequence. We have plotted the effect of packet drop over PSNR for all three encoded video. As we can see in Fig. 8, H.265 is more sensitive to packet drop compared to other two profile of H.264. SSIM index for the same scenario is shown in the Fig. 9, both the PSNR and SSIM show the similar output.

TABLE II. CHARACTERISTICS OF ENCODED VIDEO

Description	H.264_Baseline	H.264_High	H.265
Resolution	1280*720	1280*720	1280*720
No of Frame	270	270	270
Frame Rate	29.97	29.97	29.97
Avg. Bitrate	22364 kb/s	1100 kb/s	151 kb/s
Duration	9.74 second	9.74 second	9.74 second
Size	25.0 MB	1.26 MB	196 KB

Result from this experiment also shows that occurrence of jitter for the received video varies with the types of codec used. Jitter is higher for the video encoded with H.264 with baseline profile and is in decresing order for H.264 main profile and H.265 respectively. The result is shown in the Fig. 10.

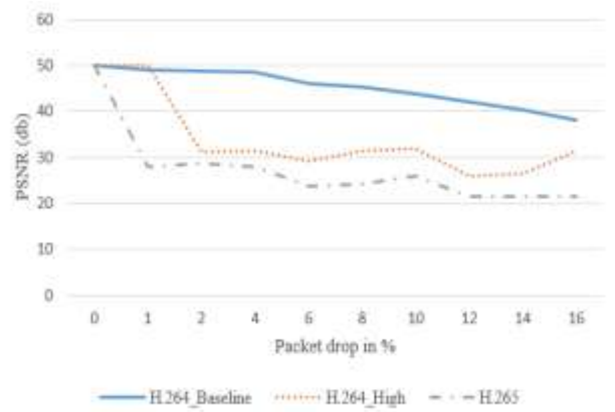

Figure 8. Packet drop Vs PSNR for different video codec

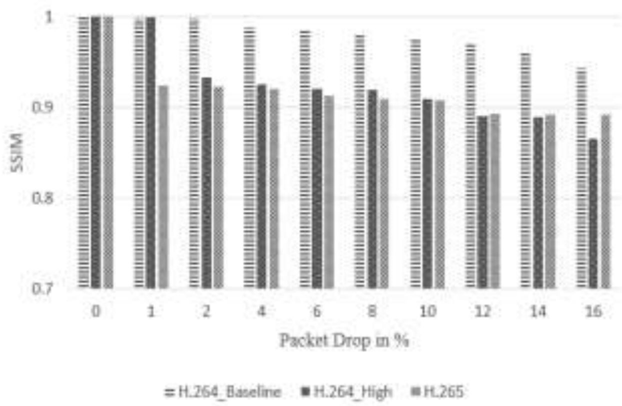

Figure 9. Packet drop Vs SSIM index for different video codec

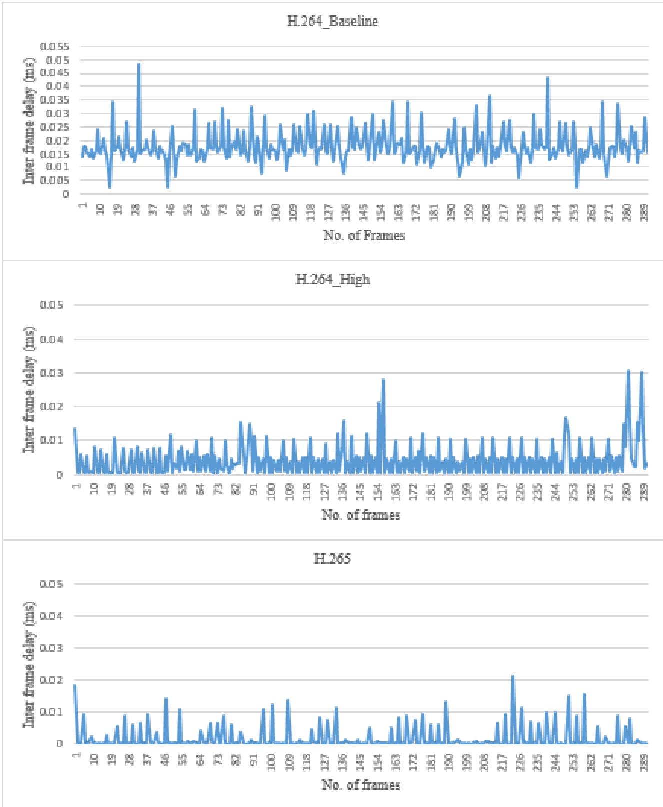

Figure 10. Comparison of Jitter performance for different video codec

To observe the effects of network jitter on the quality of video , we have conducted the experiments. This time , the value of packet drop was fixed and the amount of jitter was varied. Effect of this for both video temporal variation and types of codec used was observed. Fig. 11 . shows the effect of jitter on low and high motion video and types . Low motion video has good quality compared to high level of motion in video for the given amount of network jitter. Once again SSIM index for this was also calculated which showed the similliar result as PSNR. Effect of network jitter over SSIM is shown in Fig. 12.

Fig. 13 shows the result for effect of increased jitter over video quality for three codec described earlier. Result indicates that H.265 codec and H.264 with high profile has very minimum effect over quality compared to H.264 with profile baseline. High amount of network jitter can be tolerated for the transmission of video with H.265. SSIM index over netwok jitter for three codec is shown in Fig. 14.

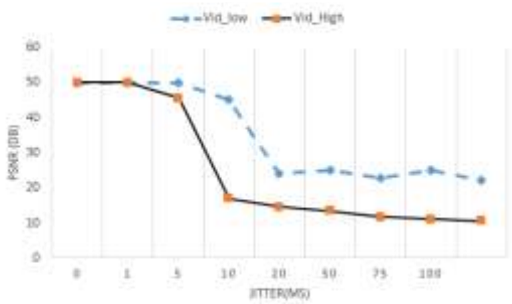

Figure 11. Effect of jitter over PSNR for low and high motion video

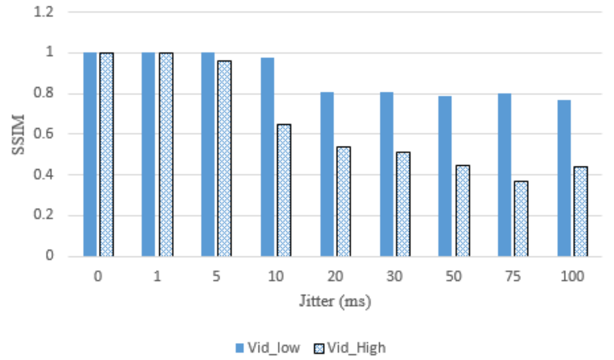

Figure 12. Effect of Jitter over SSIM index for low and high motion video

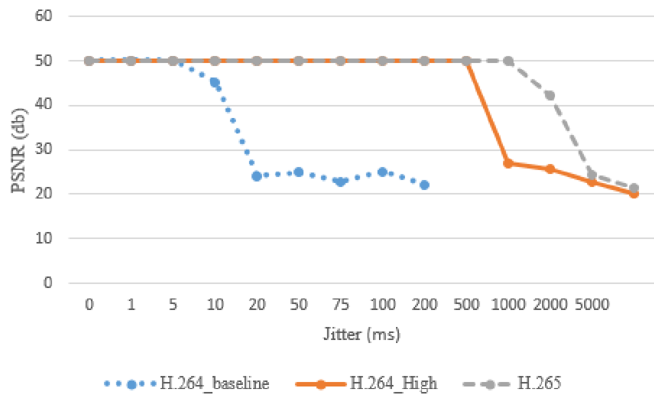

Figure 13. Effect of Jitter over video codec in terms of PSNR

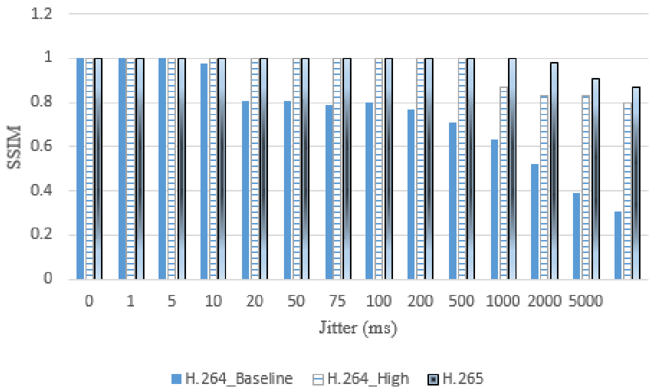

Figure 14. Effect of Jitter over video codec in terms of SSIM index

IV. CONCLUSION

In this paper, we have presented the impact of packet loss over the network on perceived video quality. In addition to the packet loss, we have also focused on another major network disruptive factor called variation in delay (jitter). Objective method was used for the evaluation of video quality over the emulated network. Video with different profile and codec was analyzed. The result from the experiment shows that, network performance affects the video quality differently depending upon the types of video codec used and level of temporal variation in the video with high level of motion is affected more compared to lower motion video. H.265 encoded video is more sensitive to network packet loss than H.264 encoded video, at the same time H.265 encoded videos are less affected in terms of network jitter. Obtained result for network jitter can be useful. to estimate the appropriate buffer size depending upon the characteristics of the video.

ACKNOWLEDGMENT

We would like to thank the E-learning Center of Kathmandu University for providing us the raw video content for our research work.

REFERENCES

[1] Z. Wang, A. C. Bovik, H.R. Sheikh and E. P. Simoncelli,“Image Quality Assessment: From Error Visibility to Structural Similarity”, IEEE Transaction on Image Processing. 2004, vol. 13, iss. 4, pp. 600-612.

[2] J. Klaue, b. Rathke, and A. wolisz,“Evalvid -A Framework for Video Transmission and Quality Evaluation”,13th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation,*Urbana, Illinois, USA, Sept 2003*

[3] CHIKKERUR, S., V. SUNDARAM, M. REISSLEIN and L. J. KARAM,“Objective Video Quality Assessment Methods: A Classi- fication, Review, and Performance Comparison”, IEEE Transactions on Broadcasting. 2011, vol. 57, iss. 2, pp. 165–182, ISSN 0018-9316

[4] O.Nukhet and T.Turhan,“On optimal receiver buffer size in adaptive Internetvideo streaming”, Journal of Zhejiang University. ISSN 1862-1775.Izmir, Turkey, Feb 2006

[5] Shahid, Muhammad, "No-reference image and video quality assessment: a classification and review of recent approaches", EURASIP Journal on Image and Video Processing

[6] Full Reference Image and Video Quality Assessment, Retrieved from hhttps://www.live.ece.utexas.edu/research/quality/frqa.htm

[7] Z. Wang and L. Lu, “Objective image and video quality assessment”, Technical Report: IBM T. J. Watson Research Center, Aug. 2001.

[8] Opnet Technologies. OPNET Modeler Product Documentation release 15.0,2009