



Two Level Optimization of MPEG-4 Compressed Video Transmission

Manish P Bhatt

ICT-CSE dept. Institute of Technology,
Nirma University, Amedabad, India

Abstract—Now a day, multimedia application like video conference, video streaming used very widely all over world. Multimedia application required large amount of bandwidth. We have bandwidth scarcity so such application faces video artifacts. In this paper, two level optimization is proposed to provide improve quality of service(QoS) and improve quality of experience(QoS). In first phase, packets are prioritize as per their importance and in second phase retransmission technique used to improve user perceive quality of video.

Keywords—PSNR; MOS; DiffServ; WRED; Delay; Jitter

I. INTRODUCTION

Multimedia file size is large compare to other type of file because it contains text, graphics, audio and video. Multimedia file should be compress at sender side and decompress at receiver end to transfer it over network or internet. Encoders are used to compress a multimedia files. In this paper, we used MPEG-4 encoders to compress multimedia file. MPEG-4 encoders convert video signal into a series of frames. Generally, only limited change occurs between one frame and the next, so an encoder can compress the video signal significantly by transmitting only the differences. Multimedia file has three frame types. Intra, or I-frames, carry a complete video picture and are coded without reference to other frames. I-frames compress using spatial compression. A received I-frame provides the reference point for decoding a received other frames. Predictive-coded or P-frames to be coded from a preceding I-frame or P-frame using temporal compression. P-frames can provide increased compression compared to I-frames, with a P-frame typically 20 to 70 percent the size of an associated I-frame. Finally, bi-directionally predictive-coded, or B-frames, use the previous and next I-frame or P-frame as their reference points for motion compensation. B-frames provide further compression, typically 5 to 40 percent the size of an associated I-frame.

In MPEG encoding, frames are arranged into groups of pictures (GoPs) that include an I-frame and all subsequent frames leading up to the next I-frame. GoP size is defined by number of frames in GoP. typically GoP size is 15 or 12.

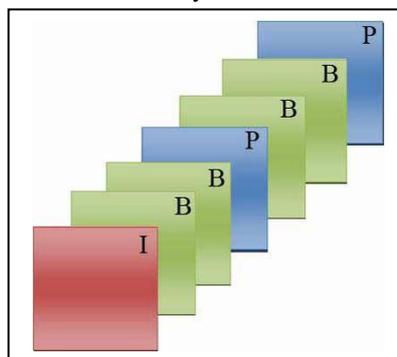


Fig. 1. Group of Picture (GoP)

This paper is organized as follows. Section II describes QoS assessment metric. Sections III describe QoS assessment framework. Section IV describes 1st level optimization (QoS). Section V describe simulation-I. Section VI describes simulation-II. Finally section VII concludes the paper.

II. QoS ASSESSMENT METRIC : PSNR AND MOS

There are basically two approaches to measure digital video quality, namely subjective quality measures and objective quality measures. Subjective quality metrics always grasp the crucial factor, the impression of the user watching the video while they are extremely costly. The human quality impression usually is given on a scale from 5 (best) to 1 (worst) as in Table 1. This scale is called Mean Opinion Score (MOS). The expensive and complex subjective tests can often not be afforded. Therefore, objective metrics have been developed to emulate the quality impression of the human visual system (HVS).

The most widespread method is the calculation of peak signal to noise ratio (PSNR) image by image. It is a derivative of the well known signal to noise ratio (SNR), which compares the signal energy to the error energy. The PSNR compares the maximum possible signal energy to the noise energy, which has shown to result in a higher correlation with the subjective quality perception than the conventional SNR.

TABLE 1

ITU-R quality and impairment scale	Scale	Quality	Impairment
	5	Excellent	Imperceptible
	4	good	Perceptible but not annoying
	3	fair	Slightly annoying
	2	poor	Annoying
	1	bad	Very annoying

Let p_n^{di} denote d_i^{th} pixel of the n^{th} reconstructed frame at the decoder, and p_n^{ei} denote e_i^{th} pixel of the n^{th} original coded frame at the encoder. The total frame error at frame n is defined as

$$e_n = \sum_{i=1}^M (p_n^{di} - p_n^{ei}) \quad (1)$$

Where M is the number of pixels in each frame. The Mean Square Error (MSE) associated with frame error e_n is given by

$$d_n = E[(p_n^{di} - p_n^{ei})^2] = \frac{1}{M} \sum_{i=1}^M [(p_n^{di} - p_n^{ei})^2] \quad (2)$$

Since the codec employs motion compensation and inter prediction to encode consecutive frames, distortion propagates to subsequent frames. Thus, the total distortion for a single lost frame at n is

$$D_n = \sum_{i \geq n} d_i \quad (3)$$

The PSNR of the video signal of frame n is given by

$$PSNR_{dB}[n] = 20 \log_{10} (V_{peak}/RMSE) = 20 \log_{10} (V_{peak}/\sqrt{d_n}) \quad (4)$$

Where V_{peak} is the maximum possible pixel value of the frame and RMSE is the root mean square error between received and original frames.

we consider luminance component Y of image and we represent pixel in row and column then equation 4 can be written as below

$$PSNR_{dB}[n] = 20 \log_{10} \left[\frac{V_{peak}}{\sqrt{\frac{1}{N_{col} N_{row}} \sum_{i=0}^{N_{col}-1} \sum_{j=0}^{N_{row}-1} [Y_S(n,i,j) - Y_D(n,i,j)]^2}} \right] \quad (5)$$

Where $V_{peak} = 2^k - 1$ and $k = \text{number of bits per pixel (luminance component)}$. Denominator is nothing but a mean square error (MSE). PSNR measures the error between a reconstructed image and the original one. Prior to transmission, it is possible to compute a reference PSNR value sequence on the reconstruction of the encoded video as compared to the original raw video. After transmission, the PSNR is computed at the receiver for the reconstructed video of the possibly corrupted video sequence received. Conversion of PSNR to MOS shown in table 2.

TABLE 2

Possible PSNR to MOS conversion	PSNR (dB)	MOS
	> 37	5 (Excellent)
	31-37	4 (good)
	25-31	3 (fair)
	20-25	2 (poor)
	< 20	1 (bad)

III. QOS ASSESSMENT FRAMEWORK : EVALVID FRAMEWORK WITH NS2

The QoS assessment framework for video traffic combines EvalVid Framework and NS2 as shown in figure-2. These framework are designed to evaluate the quality of delivered video. The evaluation process starts from encoding the source video either in the YUV QCIF or in the YUV CIF formats. MPEG4 encoder are used to compress source video and generate compressed file. The VS program will read the compressed file and generate the traffic trace file contain frame number, frame type, frame size, number of UDP packet and sender time. The MyTrafficTrace agent extracts the frame type and the frame size of the video trace file generated from the traffic trace file, fragments the video frames into smaller segments, and sends these segments to the lower UDP layer at the appropriate time according to the user settings specified in the simulation script file. MyUDP is an extension of the UDP agent. This new agent allows users to specify the output file name of the sender trace file and it records the timestamp of each transmitted packet, the packet ID, and the packet payload size. MyUDPSink is the receiving agent for the fragmented video frame packets sent by MyUDP.

This agent also records the timestamp, packet ID, and payload size of each received packet in the user specified receiver trace file. After simulation, the evaluation task begins. Based on these three trace files and the original encoded video, the ET program produces packet loss, delay, jitter report and the corrupted video file. Afterward, the corrupted video is decoded and generate reconstructed erroneous video. The total number of video frames at the receiver side, including the erroneous frames, must be the same as that of the original video at the sender side. If the codec cannot handle missing frames, the FV component is used to tackle this problem by inserting the last successfully decoded frame in the place of each lost frame as an error concealment technique. Finally, the reconstructed fixed YUV video can be compared with the original raw YUV video to evaluate the end-to-end delivered video quality. Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files

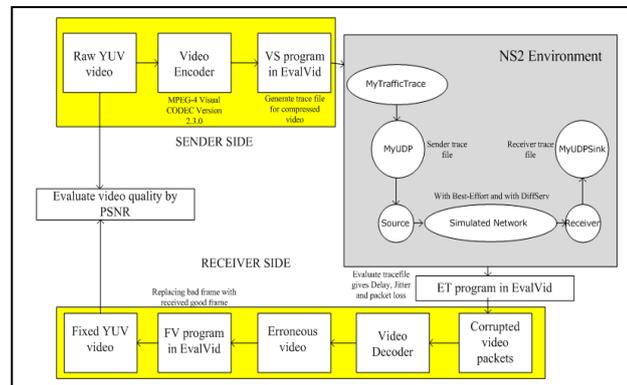


Fig. 2. QoS assessment framework

IV. 1ST LEVEL OPTIMIZATION (QOS)

DiffServ uses codepoints(DSCP) attached to a packet's IP-header to distinguish traffic with different PHBs (Per Hop Behavior). A PHBs defines a forwarding treatment of a single packet in a router. It is defining better treatment for some class of traffic than to another. Four standard PHBs are The Default PHB, Class-Selector PHBs, Expedited Forwarding PHB, Assured Forwarding PHB.

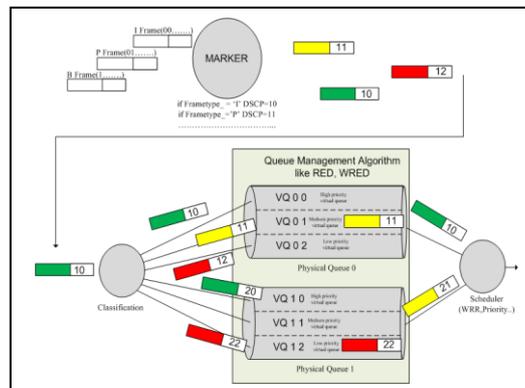


Fig. 3. DiffServ network with WRED

DiffServ manage packets inside a single queue or buffer using queue management technique like RED or WRED and scheduling between multiple queues, i.e. how bandwidth is shared between queues. In this paper, WRED queue management technique is used with DiffServ architecture. When the queue builds up and exceeds a given threshold, the WRED starts to drop packets following the specified drop probability parameters. In this paper we used mpeg4 compressed video file. In mpeg4 compressed video file, I-frame is more important compared to P-frame and B-Frame. As shown in figure-3, I-frame (green color) is pre-marked with DiffServ code point 10 (AF11) that is assured forwarding PHB class 1 low drop precedence. P-frame (yellow color) is pre-marked with DiffServ code point 11 (AF11 + 1). B-frame (red color) is pre-marked with DiffServ code point 12 (AF12) that is assured forwarding PHB class 1 medium drop precedence. Packets are classified as per DSCP and send to flow accordingly. As shown in Fig. 3, two physical queues having three virtual queues with different priority level. I-frame packets send through high priority virtual queue while B-frame packets send through low priority virtual queue.

V. 2ND LEVEL OPTIMIZATION (QOE)

When consecutive frames contain large amount of size then there is high probability to drop consecutive packets because of congestion. When consecutive packets drop or corrupted, consecutive frame have low PSNR value. When consecutive frames have low PSNR value or consecutive packets drop, user perceive poor quality of video. In 2nd level optimization, we propose retransmission technique to improve user perceive quality. In this technique, at receiver end, if consecutive packets are corrupted or lost during transmission then receiver initiate retransmission. Receiver send signal to sender to retransmit packets in odd/even manner so that drastic change in traffic can be control and also congestion can

be controlled i.e if frame no. 51,52,53,54,55,56 lost or corrupted then receiver initiate retransmission and send signal to sender, sender will resend frame 51, hold, 53, hold, 55 or 52, hold, 54, hold, 56. Simulation-II show improved users perceive quality using this technique.

```

if number_Cons_frames_lost >= 5,
receiver initiate retransmission,
$retransmit odd_number_frames.
    
```

Simulation-I for 1st level optimization(QoS)To evaluate the performance of the video quality over DiffServ network a number of experiments were conducted using NS2 simulator. Two standard sequences “Foreman”(QCIF, 30 fps, 300 frames) and "coastguard"(QCIF, 30 fps, 300 frames) are used for simulation. Fig. 4 presents the simple simulation topology, in which source s1 and s2 delivers a video traffic stream to destination D through routers R1, R2, R3 and R4. The bottleneck link has a capacity of 360 Kbps and is situated between router R3 and router R4. The queue limit at each router is set to 10 packets.

When video is transmitted with best-effort delivery, router R3 and R4 implement Droptail queue management technique.droptail queue management discards all the incoming packets when queue size reaches the queue limit.

When video is deliever over DiffServ network, I-, P- and B- frame packets are classified based on DSCP. I-frame packets are pre-marked with DSCP 10(AF10), P-frame packets are pre-marked with DSCP 11(AF10+1) and B-frame are pre-marked with DSCP 12(AF12).router R3 and R4 implement WRED queue management.when queue exceeds a given threshold value, the WRED starts to drop packets by specified drop probability parameters.The WRED parameters include a minimum threshold, a maximum threshold, and a maximum drop probability, i.e. minth, maxth, and Pmax.WRED parameter are specified respectively {4, 8, 0.025} for I-frame packets, {2, 4, 0.05} for P-frame packets and {1, 2, 0.1} for B-frame packets.

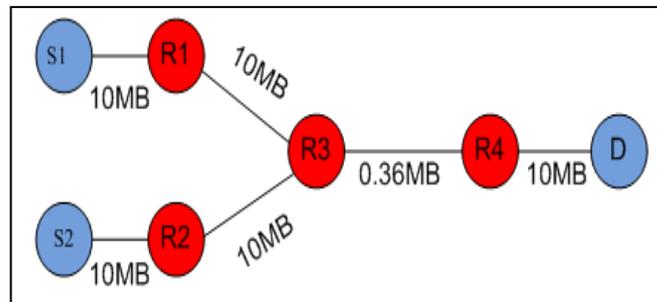


Fig. 4. Topology

Coastguard video is transmitted with best-effort delievery and DiffServ network.figure-5 shows PSNR(dB) values for each frames with Best-Effort delivery and DiffServ network. We can see that using DiffServ network PSNR values are good compare to Best-Effort delivery which shows good quality frames.

TABLE 3 QoS parameter for video sequences.

video	Delay	Jitter	PSNR(dB)
Best-effort	0.195504	0.212496	25.818733
DiffServ Network	0.136452	0.158187	28.7868

Table-3 shows that with DiffSrv network we can decrease avg.Delay and avg.Jitter value compare to Best-Effort delievery and with DiffServ network we have average PSNR value is good compare to Best-effort delievery that shows good quality frames with DiffServ network.

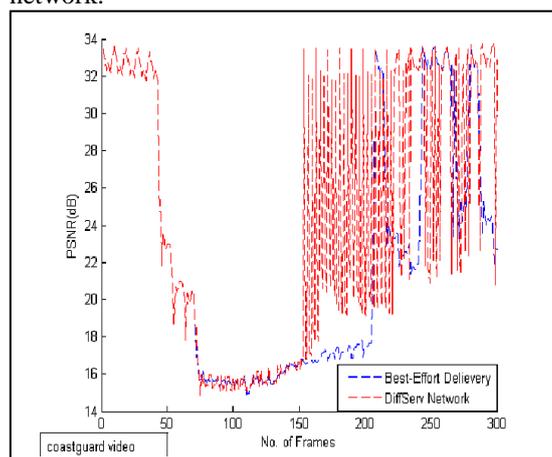


Fig. 5. PSNR value with DiffServ network and Best-effort for coastguard sequence.

VI. SIMULATION-II FOR 2ND LEVEL OPTIMIZATION(QOS)

This is extension of simulation-I. User perceive bad quality video when consecutive packets are corrupted or dropped. In this simulation we retransmit only packets where consecutive packets are corrupted or dropped. we retransmit packets in odd numbers or even numbers so that congestion can be controlled.

Figure 8.1 shows PSNR value are low for consecutive frame, from frame number 80 to frame number 150. Figure 8.3 shows user perceive quality of video, from frame number 80 to frame number 150.

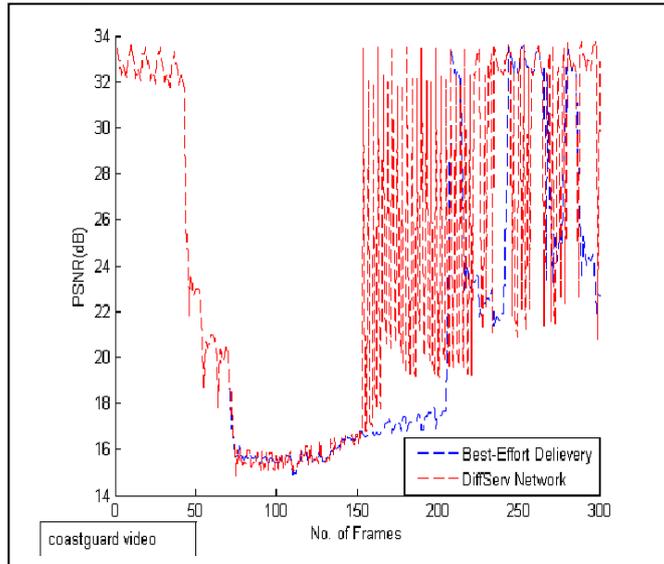


Fig. 6. PSNR value for coastguard video.

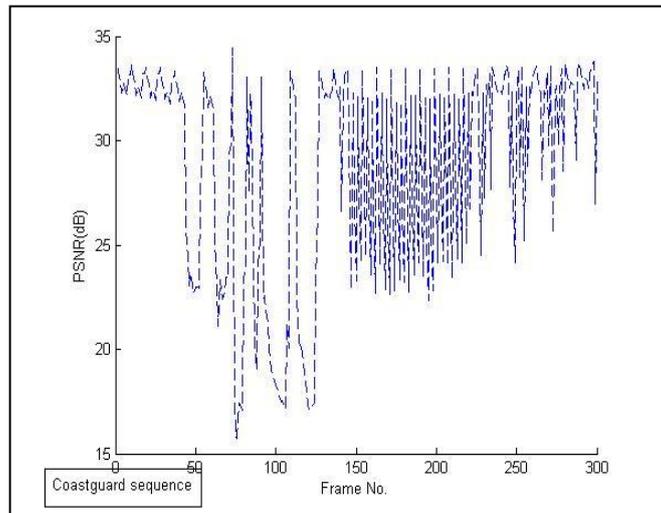


Fig. 7. PSNR value for coastguard video after retransmitting consecutive lost packets.

As shown in Figure 7 PSNR values are continuously low from frame number 80 to frame number 150, which deteriorates the perceive quality of video as shown in Figure 9. Now to improve the perceive quality of video, we are retransmitting consecutive drop packets. So, after this PSNR values varies from frame number 80 to frame number 150 as shown in Figure 8 and not continuously low as before also perceived quality of video is improved as shown in Figure 10.

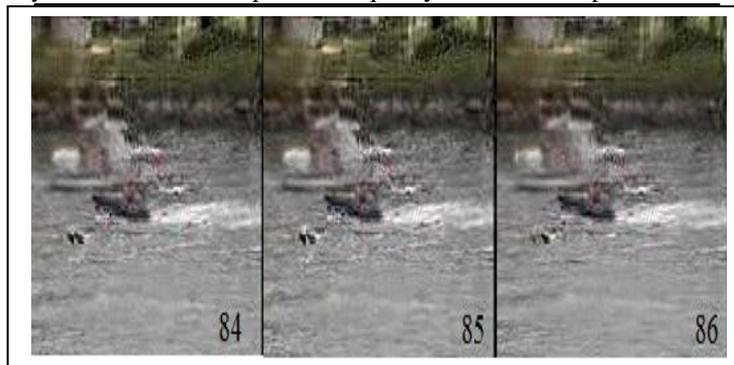


Fig. 8. Perceived poor quality of coastguard sequence

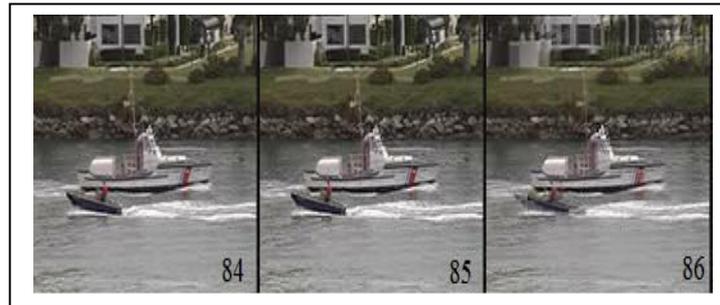


Fig. 9. Perceived improve quality of coastguard sequence.

VII. CONCLUSION

In this paper, we apply two level optimization to improve quality of service(QoS) and quality of experience(QoE) for video application. In 1st level optimization, we used DiffServ architecture, to classify packets and forwarding treatment as per their importance. simulation-I shows that QoS parameter like Delay,Jitter and packet loss are optimized using DiffServ architecture compare to Best-effort delievery. In 2nd level optimization, packets are retransmitted in odd/even manner where consecutive packets are corrupted or dropped. simulation-II shows that user perceive poor quality video when consecutive packets are corrupted or dropped. simulation result shows that user perceive quality is improved by retransmitting packets where consecutive packets are corrupted or dropped.

REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

- [1] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid A framework for video transmission and quality evaluation," in Proceedings of the International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, 2003, pp. 255-272.
- [2] Chih-Heng Ke, Ce-Kuen Shieh, Wen-Shyang Hwang, and Artur Ziviani,"An evaluation framework for more realistic simulations of MPEG video transmission", Journal of Information Science and Engineering,vol. 24, pp. 425-440, 2008.
- [3] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," RFC 2475, IETF, Dec. 1998.
- [4] J. Shin, J. Kim, and C.-C. J. Kuo, "Quality-of-service mapping mechanism for packet video in differentiated services network," IEEE Trans. Multimedia, vol. 3, no. 2, pp. 219-231, Jun. 2001.
- [5] C.-H. Ke, C.-K. Shieh, W.-S. Hwang, and A. Ziviani, "Two markers system for improved MPEG video delivery in a DiffServ network," IEEE Commun. Lett., vol. 9, no. 4, pp. 381-383, Apr. 2005.
- [6] L. Chen, G. Liu, and F. Zhao, "An improved marking mechanism for real-time video over DiffServ networks," in Proc. PCM, Dec. 2007, pp. 510-519.
- [7] Lishui Chen, and Guizhong Liu," A Delivery System for Streaming Video Over DiffServ Networks," IEEE trans. on circuits and system for video technology, vol. 20, no. 9, september 2010.
- [8] Manish P.Bhatt, Priyanka Sharma, and Purvi Kansara,"Streaming MPEG-4 Compressed Video Over DiffServ Network", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, 2014.
- [9] Manish P.Bhatt, Priyanka Sharma, and Purvi Kansara,"Analysis of Different Types of Video", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, 2014.