Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol 3, No.3, 2012

Performance Evaluation of MPEG-4 Video Transmission over IP-Networks: Best-Effort and Quality-of-Service

Md. Anwar Hossain[1*] A.N.M. Bazlur Rashid[2]

1. Department of Information and Communication Engineering, Pabna Science and Technology University, Pabna, Bangladesh
2. Department of Computer Science and Engineering, Comilla University, Comilla, Bangladesh

* E-mail of the corresponding author: manwar_ice@yahoo.com

Abstract

The demand for video communication over internet has been growing rapidly in recent years and the quality of video has become a challenging issue for video transmission. Different types of video coding standards like MPEG-2 and MPEG-4 have been developed to support application like video transmission. MPEG-2 which requires high bit rate transmission has been successful video standard for DVD and satellite digital broadcasting. On the other hand, MPEG-4 supports low bit rate and is suitable for transmitting video over IP networks. In this paper, MPEG-4 Video standard has been used for evaluating the performance of video transmission over two IP networks:- Best-effort and Quality of Service (QoS). For both of the best-effort and QoS IP networks, peak signal noise ratio (PSNR), throughput, frame and packet statistics have been considered as performance metrics. The calculated values of these performance metrics reflect that video transmission over QoS IP network is better than that of the best-effort network.

Keywords: video transmission, mpeg, ip networks, best-effort, quality of service, ns-2

1. Introduction

Few years ago internet services was not very popular media. Nowadays the demand of internet services increase day by day. People in anywhere depend on internet for any information such as books, news, audio, video, interactive TV *etc*. Before few years when we open any web pages that pages were simple that is those web pages contains only text and some images. But in this time most of the web pages contain video file that are playing automatically or play by user request. Video file is a large file so it difficult to transmit video data from one end to another. If one of the video data packets is lost then the corresponding video file may be corrupted. For that reason several video standard has been developed for different purposes such as MPEG-2, MPEG-4, MPEG-7, H.261, H.264 *etc*. For example MPEG-2 is used for high quality digital video (DVD), Digital high definition TV (HDTV), and cable TV (CATV) and MPEG-4 is used in Interactive TV over board or narrowband internet, small screen devices such as PDA's, mobile phones etc. MPEG-4 supports low data rates as compared to the MPEG-2. In this research, we have used two kinds of network, one is best-effort network and another is Quality of Service (QoS) network. MPEG-4 video has been used to transmit video data over both of those networks and the Peak Signal Noise Ratio (PSNR), throughput, packet and frame statistics have been measured. Depending on those data, we observed which network is better for MPEG-4 video transmission.

The term best-effort means that IP provides no error checking or tracking. IP assumes the unreliability of the underlying layers and does its best to get a transmission through to its destinations, but with no guarantees. For this network the queue management technique is drop tail.

The term Quality of Service (QoS) is defined as the set of parameters that define the properties of media streams. There are four layers of QoS: user QoS, application QoS, system QoS and network QoS. (i) The user QoS parameters describe requirements for perception of the multimedia data at the user interface. (ii) The application QoS parameters describe requirements for the application services, possibly specified in terms of media quality (high end-to-end delay) and media relations. Application layer QoS controls how to avoid congestion and to maximize video quality in the presence of packet loss. The application layer QoS control techniques include congestion control and error control. (iii) The system QoS parameters describe requirements on the communications services resulting from the application QoS. (iv) The network QoS parameters describe requirements on network services, like network load and network performance.

Here Weighted Random Early Detection (WRED) technique has been used for queue management. There are three virtual queues under one physical queue. Each of the queues is used for different types of video frame. The I frame packet is pre-marked with highest priority, P frame packet is pre-marked with medium priority and B frame packet is pre-marked with lowest priority.

There are some works that have developed by different researchers for transmission of video over networks. Some of the works are list in [1-6]. In this paper, we have evaluated performance of video transmission over IP networks of two type: Best-effort and QoS.

2. MPEG-4

MPEG-4 is an ISO/IEC (International Standardization Organization/International Electrotechnical Commission) standard developed by MPEG (Moving Picture Experts Group), the committee that also developed the Emmy Award winning standards known as MPEG-1 and MPEG-2 [7, 8].

First, it is worth mentioning that MPEG-4 deals with "media objects", that are a generalization for the visual and audio content. These media objects are used together to form the audiovisual scenes. The main parts of MPEG-4 are systems, visual, audio and DMIF (Delivery Multimedia Integration Framework). The basis is formed by systems (presentation, demux and buffer), audio and visual (decoding). DMIF is the transport interface between application and network (storage) [9].

2.2 MPEG-4 Structures

For structural presentation of MPEG-4, ISO model is deployed. Figure 1 shows the layered description of MPEG-4 standard.

2.2.1 Transmission/Storage Medium

Transmission/Storage Medium layer is very isolated from the actual logic behind the MPEG-4. It specifies the physical layer, digital storage requirements. But, still at this level the data is regarded as raw data. Normally it is the tasks of the upper network layer (like UDP/IP, ATM, MPEG-2 Transport Stream) to handle the actual physical network properties [9].

2.2.2 Delivery Layer

Media objects are transported in a streaming manner. Multiple elementary streams are used to convey the media object. These streams can contain a number of different information: audiovisual object data, scene description information, control information in the form of object descriptors, as well as meta-information that describes the content or associates intellectual property rights to it. The elementary streams themselves are created in the upper layers, and at this layer their meaning is not so important. The task of the delivery layer is to handle and relay these elementary streams [9].

2.2.3 Sync Layer

The layer receives the content from the media layer, process the data and pass the result to the delivery layer. No matter of the type of data conveyed in each elementary stream, it is important that they use a common mechanism for conveying timing and framing information.

Synch layer handles the synchronization of elementary streams and also provides the buffering. It is achieved through time stamping within elementary streams. It provides/recovers the timing information from the media object (or scene description). The layer is responsible with the synchronization of the elementary streams belonging to a particular presentation scene [9].

2.2.4 Multimedia Layer

The type of information identified in each stream must be retrieved at decoder (respectively the encoder must provide it). For this purpose object descriptors are used. These descriptors identify group of elementary streams to one media object (no matter if treating audio object, visual object, a scene description stream, or even point to an object descriptor stream). Briefly, the descriptors are the way decoder identifies the content being delivered to it [10].

2.2 Audio-Video Coding

MPEG-4 has an extensive set of audio features. It provides separate codecs for low-bit rate speech and general-purpose audio. The MP3 was one of the key elements in MPEG-1, but it seems unlikely that MPEG-4 audio (MP4) will become as important file format for consumers than MP3 because the needs for the consumers were well covered by MP3.

Basically for dynamic image coding, two coding models are used: Intra-Mode and Inter-Mode. In Intra-Mode both the spatial redundancy and irrelevancy are exploited with block based DCT coding, quantization, run length and huffman coding. Only information from the picture itself is used in Inter-Mode and thus every frame can be decoded independently. Afterwards the predicted image is subtracted from the original image. The resulting difference picture is DCT coded, quantized and VLC coded. The motion vectors describing the motion of the blocks in the picture are necessary side information for the decoder and are also encoded with VLC. Figure 2 shows the MPEG4 video coding scheme.

3. Simulation Methodology

3.1 Introduction to NS-2

The NS-2 simulator covers a large number of applications, in the protocols, different types of networks its elements and traffic models. Usually it is known as "simulated objects". NS-2 is an open-source freeware, and it is constantly maintained and updated by its large user base, and a small group of developers [10, 11].

NS-2 is a discrete event simulator targeted at networking research. NS-2 provides substantial support for simulation of Ipv4, UDP, TCP, routing and multicast protocols over wired and wireless networks.

NS-2 is a discrete event simulator which provides support for: (i) Various network protocols (transport, multicast, routing, MAC). (ii) Simple or complex topologies (including topology generation). (iii) Agents (defined as endpoints where network-layer packets are constructed or consumed). (iv) Various traffic generators. (v) Simulated applications (FTP, Telnet, and Web). (vi) Several queue management and packet scheduling schemes. (vii) Error models. (viii) Local area networks. (ix) Wireless networks. (x) Mobile Cellular Networks (wired-cum-wireless network) [11].

3.2 Video Quality Evaluation Tools (EvalVid)

The structure of the EvalVid framework is shown in Figure 3. The main components of the evaluation framework are described as follows [12]:

Source: The video source can be either in the YUV QCIF (176 x 144) or in the YUV CIF (352 x 288) formats.

Video Encoder and Video Decoder: Currently, EvalVid supports two MPEG4 codecs, namely the NCTU codec and ffmpeg. In the present investigation, the NCTU codec for video coding purposes has been used.

ET (Evaluate Trace): Once the video transmission is over, the evaluation task begins. The evaluation takes place at the sender side. Therefore, the information about the timestamp, the packet id, and the packet payload size available at the receiver has to be transported back to the sender. Based on the original encoded video file, the video trace file, the sender trace file, and the receiver trace file, the ET component creates a frame/packet loss and frame/packet jitter report and generates a reconstructed video file, which corresponds to the possibly corrupted video found at the receiver side as it would be reproduced to an end user.

FV (Fix Video): Digital video quality assessment is performed frame by frame. Therefore, the total number of video frames at the receiver side, including the erroneous ones, 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.

PSNR (Peak Signal Noise Ratio): PSNR is one of the most widespread objective metrics to assess the application-level QoS of video transmissions. The following equation shows the definition of the PSNR between the luminance component Y of source image S and destination image D:

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

where $V_{peak} = 2^k-1$ and k = number of bits per pixel (luminance component). PSNR measures the error between a reconstructed image and the original one.

MOS (Mean Opinion Score): MOS is a subjective metric to measure digital video quality at the application level. This metric of the human quality impression is usually given on a scale that ranges from 1 (worst) to 5 (best) [12].

3.3 New Network Simulation Agents

Three connecting simulation agents, namely MyTrafficTrace, MyUDP, and MyUDPSink, are implemented between NS2 and EvalVid. These interfaces are designed either to read the video trace file or to generate the data required to evaluate the video delivered quality. The following figure illustrates the QoS assessment framework for video traffic enabled by the new tool-set that combines EvalVid and NS2. Figure 4 shows the simulation framework [12].

MyTrafficTrace: The MyTrafficTrace agent is employed to extract the frame type and the frame size of the video trace file generated from the output of the VS component of EvalVid. Furthermore, this agent 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: Essentially, 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. The task of the MyUDP agent corresponds to the task that tools such as tcp-dump or win-dump do in a real network environment.

MyUDPSink: 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 file.

4. Results and Discussion

4.1Simulation Scenarios

The simulation model (network topology) used in the simulations is shown in Figure 5. As shown in figure, there are four nodes: one sender (S1), here we called video sender, one receiver (D1), here we called video receiver and two routers R1 and R2. The video receiver (D1) receives video from the video sender (S1) via the two routers.

We define the bandwidth between the video source (S1) and router (R1) is 10 Mb in each direction and also define the delay is 1 millisecond. The link bandwidth between the two routers is defining 0.18 Mb in each direction and delay is 10 milliseconds. The bandwidth between router (R2) and video destination is 10 Mb and delay is 1 millisecond.

The new UDP agent called myUDP is attached to the video sender (S1) and another agent myUdpSink2 is attached to the video receiver (D1). The new traffic source myTrace2 is attached to the node S1.

4.2Performance Metrics

Performance metric is one type of parameter. For measuring the performance of video we use important metric is call PSNR (Peak Signal Noise Ratio). PSNR is one of the most widespread objective metrics to assess the application-level QoS of video transmissions. Setting another parameter such as bandwidth, buffer size, delay we calculate PSNR in every cases. These metrics or parameters are briefly described in the following section

- Bandwidth: The number of packets in transit for every time instant (sec). It is measured in Mega Bits per Second.
- Buffer Size: The size of memory. It can be include in any node.
- Delay: Delay means the propagation delay of a packet.

In this research we have calculated the network throughput in every case. The throughput can be defined as

Throughput: How well does the network deliver packets from source to destination? *i.e.*,

$$\text{Throughput} = [\text{total sent data} - \text{total lost data}] / \text{time} \qquad (2)$$

Throughput generally represents in Mbps. It may also be expressed by Bps (Bytes per sec).

4.3Performance Analysis

For performance analysis we use a YUV video file forman_qcif.yuv. It is composed of 400 frames, including 45 I frames, 89 P frames, and 266 B frames. We have used a topology as shown in Figure 5 and write a tcl script refer to video_be.tcl for transmitting the video file from source to destination into the best effort network. Before going to transmit we encode this YUV video file to MPEG4 video format using mpeg4 encoder. After running the tcl script video_be.tcl we will get trace file that contain sending time, receiving time and other information of each packet. The erroneous video file from the receiver side is reconstructed using the original encoded video file using the video quality evaluation tools (Evalvid). Now we use a mpeg decoder to decode the video file to YUV format. Now we calculate the frame by frame and average PSNR from the trace file and record the data into another file. Here we use 30 frames per sec and single queue is used in router R1. The queue type used in router R1 is DropTail.

Now we write another tcl script for transmitting the video data into Quality of Services (QoS) network. For that reasons we add some QoS parameter into the tcl script, *i.e.*, we add three virtual queues under one physical queue each virtual queue is used for each type of frame. We add Per Hop Behavior (PHB) in each queue. We define the maximum dropping probability of each frame *e.g.*, we define maximum dropping probability of I frame is 2.5%, maximum dropping probability of P frame is 5% and maximum dropping probability of B frame is 10 %. This script file is referring to video_qos.tcl. After simulate this program we calculate the frame by frame and average PSNR and record the data into another file. The queue type used in router R1 is WRED (Weighted Random Early Detection). In the case of 30 Fps the simulation time is 13.33 second. Figure 6 shows the graph for PSNR verses frame number.

The average PSNR and throughput are shown in Table 1. The packet statistics for both cases are shown in Table 2. After simulation the video file forman_qcif.yuv using best effort and quality of services network, some of video frame are shown in Table 3. We now use 25 frames per second instead of 30 frames per second. In this case simulation time is 16.10 second. Again we calculate the PSNR for both cases and draw the graph shown as Figure 7. The average PSNR and throughput are shown in table 4. The packet statistics for both cases are shown in Table 5. The average PSNR with respect to the no of frame per second for both best-effort and QoS network are shown in the Figure 8. The throughput with respect to the no of frame per second for both best-effort and QoS network are shown in the Figure 9.

PSNR Vs Buffer size: Consider the same topology as shown in Figure 5. Using the QoS network and 30 frames per second, just we have changed the queue limit of router R1 and calculate the average PSNR and create a graph by using this data that is shown in Figure 10.

PSNR Vs Bandwidth: Consider the same topology as shown in Figure 5. Now we change the bandwidth of link between two routers R1 and R1 and calculate the PSNR in every time. Here we have used 30 frames per second and QoS network. Using this data we draw a graph as shown in Figure 11.

4.4Result and Discussion

From the Figure 6 it is observed that, for quality of services (QoS) network the Peak Signal Noise Ratio (PSNR) for each frame is higher than the best effort network. Because in the QoS network, the queue management is Weighted Random Early Detection (WRED) and I frame packet marked as highest priority, P frame packet marked as medium priority and B frame packet marked as lowest priority. From the Table 1 we see that the average throughput for best effort network is 22371.28 Bps which is greater than the average throughput of quality of services network is 21389.14 Bps. This simulation represents that throughput is not important factor for video transmission because video frame (I, P, B) are depend one another. The frame P and B depend on I frame, if I frame packet is lost then the corresponding P and B frame are not decodable even both of the frame packet is received. The average PSNR for best effort is 24.36 dB which is less than the quality of services network is 30.18 dB. From the above discussion we conclude that for video transmission, quality of services network is better than the best-effort services network.

Table 2 shows the packet statistics for both best-effort and quality of services network in case of 30 frames per second. From this table we have shown that the total frame lost for best-effort network is 43 but for quality of services network the total frame lost is 110, which is greater than the best-effort network. Here it is important that, I frame lost for quality of services network is only 3 but for best-effort network it is 23. We know that Intra (I) frame is the main frame in the video sequence, because another frame depends on I frame. If I frame lost is grater then the video quality going to be worst. From the above discussion we conclude that the quality of services network is better.

Some of the receiving video frame for best-effort network and quality of services network are shown in the Table 3. We observed that the frame no 173, 278 and 360 perceptually are not good for best-effort network. But in the case of Quality of Service network those three frames are perceptually better than the best-effort network.

Figure 7 shows the PSNR for the best-effort and quality of services network in case of 25 frames per second with respect to frame number. From Figure 8 we see that for 25 fps the average PSNR is greater than the 30 fps. Also the time required for simulation for 25 fps is greater than the 30 fps. In this case 25 fps the simulation time is 16.10 second. But in case of 30 frames per second the simulation time is needed 13.33 second. The time required for video data transmission is very important. So we use 30 frames per second for better reception.

6. Conclusion

In this research we have evaluated the performance of MPEG-4 video transmissions over the best-effort and quality of services network using NS-2. Network Simulator (NS-2) was chosen for this task, as it is an event driven network simulator, which is popular with the networking research community.

The performance of MPEG-4 video transmission over Quality of Services network is better than the best-effort network because QoS network uses Weighted Random Early Detection (WRED) queue management technique while best-effort network uses dropTail queue. In QoS network, three virtual queues have been added under one physical queue and each of the queues is used for I, P and B frame respectively. The I frame packet is pre-marked with lowest drop probability in the application layer at source, the P frame packet is pre-marked with middle drop probability and B frame packet is pre-marked with highest drop probability.

References

Ziviani, A., Wolfinger, B. E., Rezende, J. F., Duarte, O. C. M. B. & Fdida, S. (2005), "Joint Adoption of QoS Schemes for MPEG Streams", *Kluwer Academic Publishers* **26**(1), 59-80.

Heng, C. K., Shieh, C. K., Hwang, W. S. & Ziviani, A. (2005), "A Two Markers System for Improved MPEG Video Delivery in a DiffServ Network", *IEEE Communications Letters, IEEE Press* **9**(4), 381-383.

Sawaya, J. N., Ghaddar, B., Khawam, S., Safa, H., Artail, H. & Dawy, Z. (2005), "Adaptive Approach for QoS Support in IEEE 802.11e Wireless LAN", *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2005)*, 167-173.

Huang, H., Ou, J. & Zhang, D. (2005), "Efficient Multimedia Transmission in Mobile Network by using PR-SCTP", *Communications and Computer Networks (CCN 2005)*, 213-217.

Lo, A., Heijenk, G. & Niemegeers, I. (2005), "Performance Evaluation of MPEG-4 Video Streaming over UMTS Networks using an Integrated Tool Environment", *Proceedings of 2005 International Symposium on Performance Evaluation of Computer and Telecommunication Systems*, 676-682.

Vassiliou, V., Antoniou, P., Giannakou, I. & Pitsillides, A. (2006), "Requirements for the Transmission of Streaming Video in Mobile Wireless Networks", *International Conference on Artificial Neural Networks (ICANN)*, 528-537.

Rao, K. R., Bojkovie, Z. S. & Milovanovie, D. A. (2002), *Multimedia Communication System: Techniques, Standards and Networks*, Prentice Hall.

Halsall, F. (2000), *Multimedia Communication*, Addison Wesley Pub Co.

Burlacu, M. & Kangas, S. (2003), "MPEG-4 Technology Strategy Analysis", *T-109.551 Research Seminar on Telecommunications Business II*, *Telecommunications Software and Multimedia Laboratory*, Helsinki University of Technology.

Fall, K. & Varadhan, K., "NS-2 Manual: The NS Manual", The VINT Project Collaboration between Researchers at UC Berkeley, LBL, USC/ISI, and Xerox PARC.

"NS, The network simulator – NS-2", http://www.isi.edu/nsnan/ns [accessed 17 Jun 2007].

Klaue, J., Rathke, B. & Wolisz, A. (2003), "A Framework for Video Transmission and Quality Evaluation-EvalVid", *Proceedings of the 13th International Conference on Modelling Techniques and Tools for Computing Performance Evaluation*, 255-272.

Md. Anwar Hossain was born in 26th November 1983 in the district of Pabna, Bangladesh. He has got his B.Sc (Honours) and M.Sc degree in Information and Communication Engineering from the University of Rajshahi, Rajshahi, Bangladesh in 2007 and 2009 respectively. He has several years of experience in the field of programming and IT sector. Currently, he has been working as a Lecturer in the department of Information and Communication Engineering at Pabna Science and Technology University, Pabna, Bangladesh. His major field of study is radio resource management for wireless networks, network coding

and game theory in wireless communication, cognitive radio networks, wireless sensor network, mobile ad hoc network. E-mail: manwar_ice@yahoo.com

A.N.M. Bazlur Rashid born in 1984 at Rangpur, Bangladesh. He received his B.Sc. Engg. degree in Computer Science and Engineering from Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh in 2005 and M.Sc. Engg. degree in Information and Communication Technology from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 2010. He has several years of professional experience in the field of database having Oracle professional certification in database administration. Currently, he has been serving as a Lecturer in the Department of Computer Science and Engineering of Comilla University, Comilla, Bangladesh. His major field of study includes query processing, query optimization, materialized view, database and information system, data warehousing and data mining. E-mail: bazlur.rashid@yahoo.com

Appendix

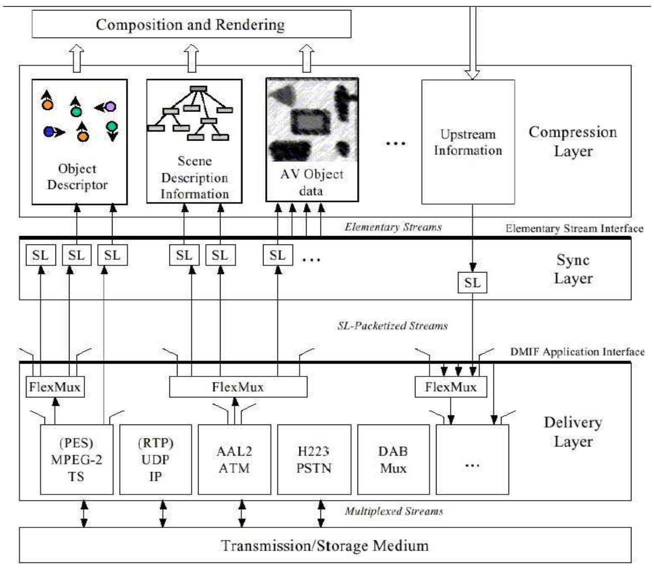


Figure 1. Layered description of MPEG-4 standard

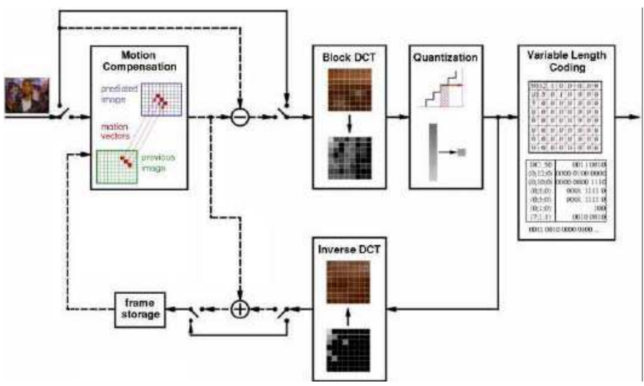


Figure 2. MPEG4- video coding scheme

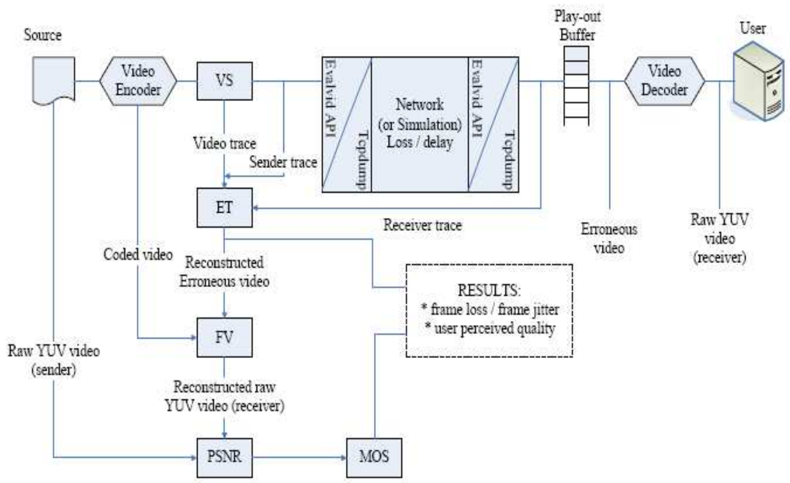


Figure 3. Scheme of Evaluation Framework

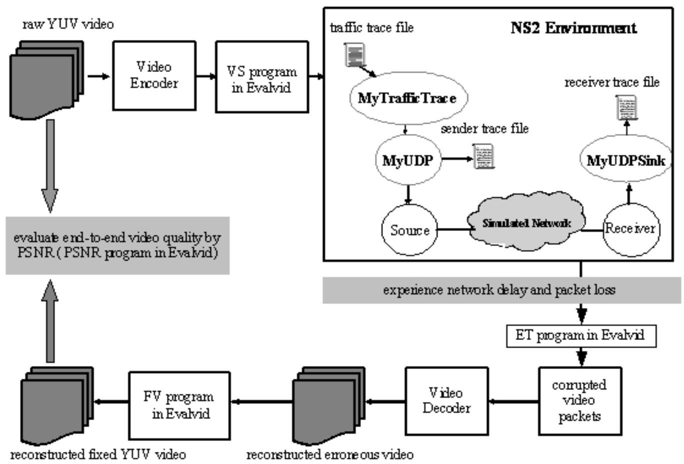


Figure. 4 Simulation Framework

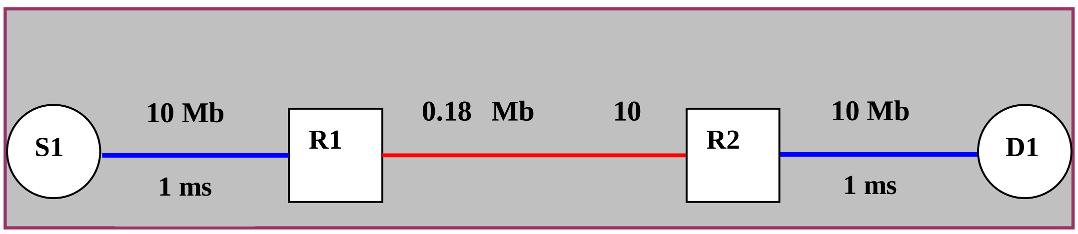


Figure 5. Simulation Scenario

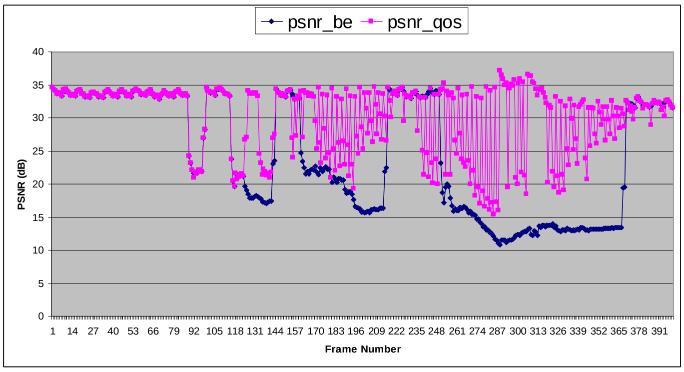


Figure 6. PSNR verses Frame Number in case of 30 fps

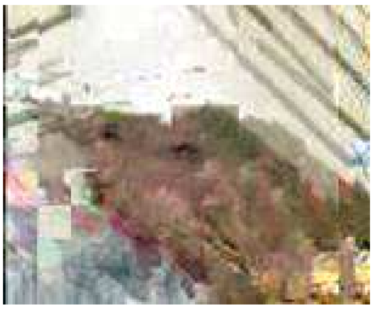
Table 1. Average PSNR and Throughput for 30 fps

30 FPS	Average PSNR (dB)	Throughput (Bps)
Best effort	24.3629	22371.2841
Quality of Services	30.17745	21389.1421

Table 2: Packet and Frame Statistics for 30 fps

30 FPS		I	P	B	Total
Best effort	Packet Sent	173	109	266	549
	Packet lost	48	14	7	69
Quality of Services	Packet sent	173	109	266	549
	Packet Lost	111	4	0	107
Best effort	Frame sent	45	89	266	401
	Frame lost	23	13	7	43
Quality of Services	Frame sent	45	89	266	401
	Frame lost	3	0	107	110

Table 3. Video Frame

BE173	BE278	BE360
		
QoS173	QoS278	QoS360

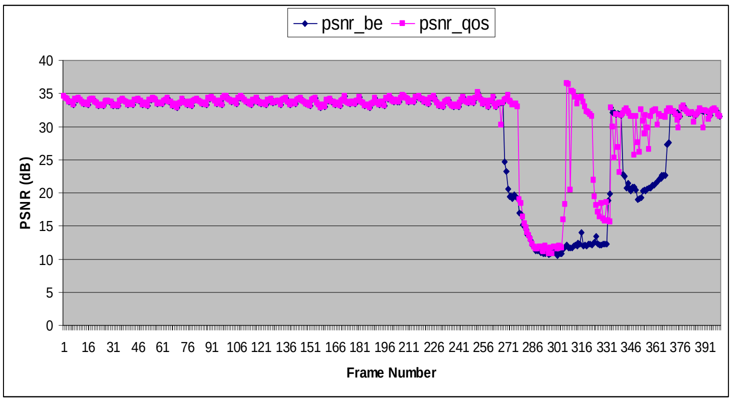


Figure 7. PSNR verses Frame Number in case of 25 fps

Table 4. Average PSNR and Throughput for 25 fps

25 FPS	Average PSNR (dB)	Throughput (Bps)
Best effort	29.43923	20884.7257
Quality of Services	31.3486	20405.1746

Table 5. Packet statistics for 25 frames per second

25 FPS		I	P	B	Total
Best effort	Packet Sent	173	109	266	549
	Packet lost	15	4	0	19
Quality of Services	Packet sent	173	109	266	549
	Packet Lost	6	2	30	38
Best effort	Frame sent	45	89	266	401
	Frame lost	10	4	0	14
Quality of Services	Frame sent	45	89	266	401
	Frame lost	4	2	30	36

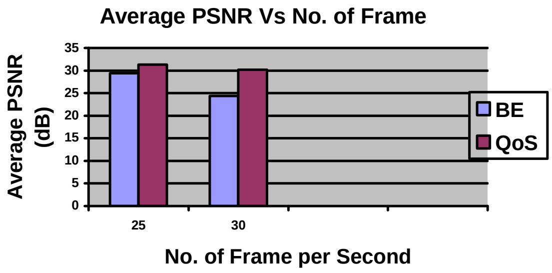


Figure 8. Average PSNR Vs No. of Frame per Second

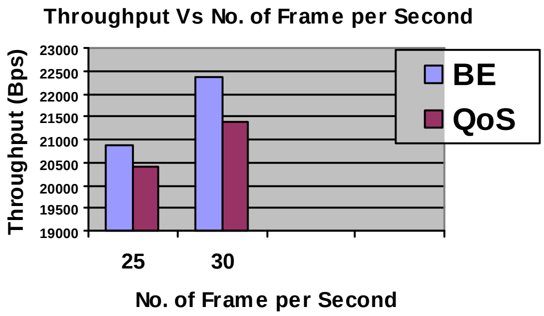


Figure 9. Throughput Vs No. of Frame per Second

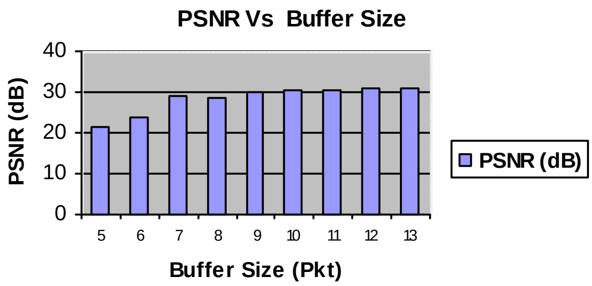


Figure 10. PSNR Vs Buffer Size

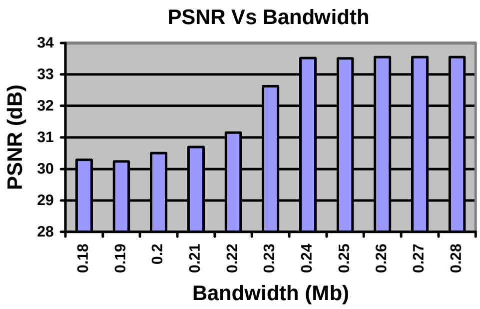


Figure 11. PSNR Vs Bandwidth