A Novel Recovery Technique For Lost Internet Video Frames Simulation and Evaluation

OMAR S. ESSA
Computer Science Dep., Faculty of Science,
Menoufia University,
EGYPT

NIKOS E MASTORAKIS
WSEAS European Office
Agiou Ioannou Theologou 17-13,
15773, Zografou, Athens,
GREECE

http://www.wseas.org/mastorakis

Abstract: Internet was designed to support traditional and multimedia applications. Multimedia applications suffer from unacceptable delay, jitter and data loss. Data loss has the largest impact on quality. When the retransmission technique is used to recover lost multimedia packets, more obstacles, like the delay time resulted from lost packets retransmission operation, network traffics overloading, and buffer management, are faced. All above demonstrated problems are due to the real-time communication features. In this paper, we extend our lost audio recovery technique to recover the video frames. Finally, NS2 is used to build a simulated environment for testing our new video technique as regards the MPEG transmission.

Key-Words: - Multimedia Loss, Internetworking, Video Frames, TCP/IP, Multimedia Protocols.

1 Introduction

Data loss in multimedia streams transmission can impact the continuity in the display. Data loss can occur involuntarily from network congestion or system buffer overflows, or voluntarily in order to avoid congestion at a client, server or network router. Too much data loss can result in unacceptable media quality. To compensate for data loss, much work has been done to find effective data loss recovery techniques. There are two categories of data loss recovery techniques: sender driven and receiver-based. Each of them has its own strength and weakness. These techniques have proven to be fairly effective (but not sufficient) for audio stream data loss, but have yet to be applied to video. In this paper, we extend our audio technique to recover the lost video frames [1], [2].

This paper is proceeding as follows; in section 2, our video technique is introduced. The detailed simulation is demonstrated in section 3.

2 Our Technique

The basic idea of our video recovery technique is to classify the video objects depending on there importance. We can define the importance of the object as its ability of providing the users with a helpful meaning that aid them in video understanding. To reach our target, we should analyze the video streams. This analysis provides us with the sensitive parts for the human eyes, and the data that can be understood by the video context. The obstacles that may be faced in our technique are; how we can determine the features of the video objects, how these objects are classified, and what are the parameters that controlled in the object importance (object prioritization). We will answer these questions through our technique description

In the following subsections, video contents analysis, video objects priority, and Real-Time Transport Protocol (RTP) header manipulation, are demonstrated. Finally, simulated experiments and results comments are showed.

2.1 Video contents analysis

To analyze and describe the video streams, we should answer two questions stated as follows: 1) How the video streams are transmitted after taken by a camera to the perception? 2) What are the video contents?

2.1.1 How the video streams are transmitted?

Video streams pass through four layers until viewed by the perception. The first layer is called a sensor layer. This layer takes video shots for the target place using a video camera. These video shots are passed to the second layer that is called a video-processing layer as a signal structure. The video-processing layer comprises a source media coding, alarms, and video database handling. Consequently, the video streams are transmitted within the network using the third layer that is called a network layer. This layer depends on a transmission network characteristics and used protocols (in our technique, we use the Internet as a transmission environment). Finally, the video data are decoded, retrieved, interfaced, and the man-to-machine relationship is constructed, see Fig. 1.

2.1.2 What are the video contents?

When we concentrated at the video film, we found that it constructed from a group of harmony stories. Each story is constructed from a group of scenes, and each scene can be considered as a group of shots. The video shot is constructed by a group of objects. Every object represents a meaning(s) transferred to a user perception. The video object has global and local related features. The global features are related to a fixed background in each video shot, and can be easy detected. The local object features are related to the colour, motion, event, and activates in the video streams. The object colour and motion are considered as a low level semantic and described qualitatively or quantitatively. The event or activity can be considered as a high level semantic. The event is defined by the video interpretation or context independent, hence the lost in the event structure affects the video meaning. The object behaviour depends on the video understanding and can be considered as a context dependant. So, the object event is more critical than the object behaviour. Before local features of the object are extracted, we should extract the moving objects and estimate each motion degree (fast, middle, low). The moving object extraction is related to the pixels-to-object relation. The object motion estimation and features extraction are related to object- to-feature relation at the low-level features level, and related to object-to-video object relation at the mid-level and high-level features level, see Figs. 2, 3, 4, 5.

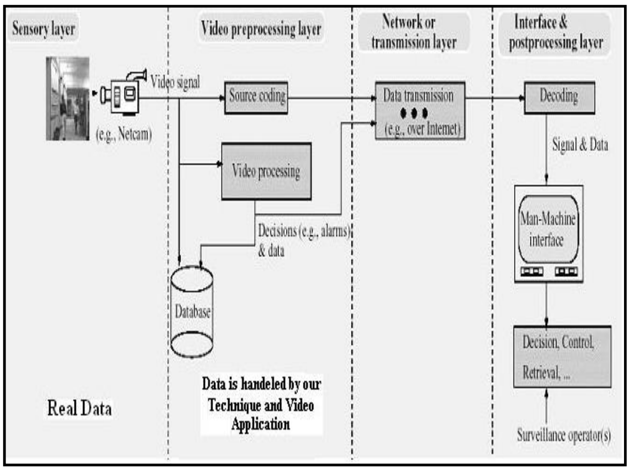

Figure (1): Source - Destination video layers

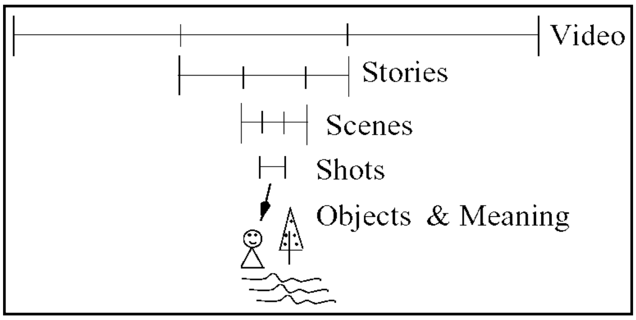

Figure (2): Video analysis

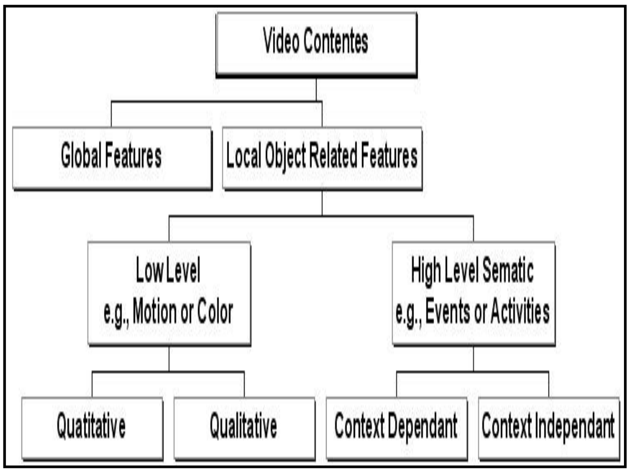

Figure (3): Video contents

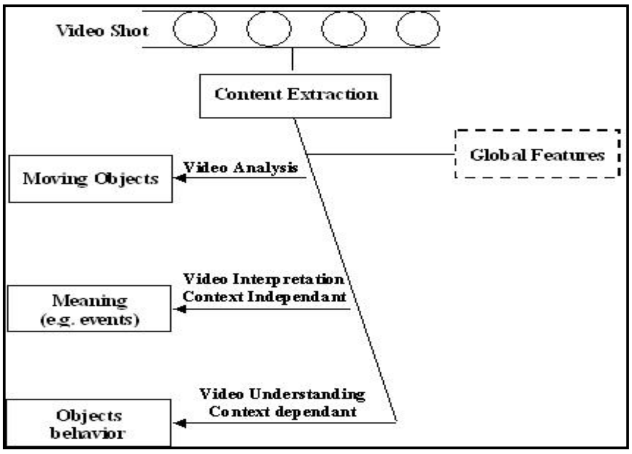

Figure (4): Video shot contents

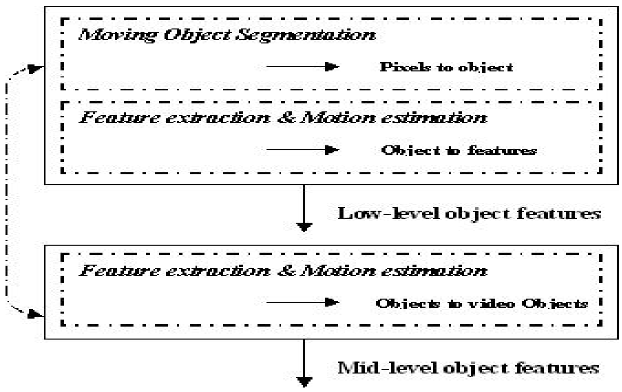

Figure (5): Video objects tracing

2.2 Objects Priority determination

After analysis of the video contents, we note that the infrastructure of video streams is the object. We can extract the object priority form above listed analysis. It's notable in table 1 that the first priority (P1) is given for the object motion property. This is because when the eye notes the object, its takes a time to observe it. Urgently, we should keep the moving object without errors to give an eye a chance for moving objects detection. The second priority (P2) is given for the object type. The object type is described as regards its quality or quantity. The object type takes the second priority, because the qualitative objects represent a main concept in the video story. Reversibly, the quantitative object can be easy detected as it has alternative(s) or huge view. The third priority (P3) is given to the color of the object. Some times, we face a colored object gives a special meaning for the perception. Non-colored object means that the object didn't contains more descriptions or contains neglected specifications (regardless some special cases). The forth priority (P4) is related to the object sharing i.e. how the object shares others in shot constructing. Is the object construct an event or describe another object behavior. There are two types of shared object; one can be understood from the context, called context dependent and another called context independent that is understood by the eye (i.e. seeing).

Note 1: This priority arrangement may be not fixed, but it depends on the used application. The user can change this priority or merge two priorities in one, if this is required.

	P1	P2	P3	P4	Id.	Id.
	Motion	Type	Color	Context Case		Code
Object	Fast	Qualitative	Colored	Event (C-I)	1	00000
				Behavior (C-D)	2	00001
			Not (W&&B)	Event (C-I)	3	00010
				Behavior (C-D)	4	00011
		Quantitative	Colored	Event (C-I)	5	00100
				Behavior (C-D)	6	00101
			Not (W&&B)	Event (C-I)	7	00110
				Behavior (C-D)	8	00111
	Middle	Qualitative	Colored	Event (C-I)	9	01000
				Behavior (C-D)	10	01001
			Not (W&&B)	Event (C-I)	11	01010
				Behavior (C-D)	12	01011
		Quantitative	Colored	Event (C-I)	13	01100
				Behavior (C-D)	14	01101
			Not (W&&B)	Event (C-I)	15	01110
				Behavior (C-D)	16	01111
	Low	Qualitative	Colored	Event (C-I)	17	01111
				Behavior (C-D)	18	10000
			Not (W&&B)	Event (C-I)	19	10001
				Behavior (C-D)	20	10010
		Quantitative	Colored	Event (C-I)	21	10011
				Behavior (C-D)	22	10100
			Not (W&&B)	Event (C-I)	23	10101
				Behavior (C-D)	24	10111
	Fixed	Qualitative	Colored	Event (C-I)	25	11000
				Behavior (C-D)	26	11001
			Not (W&&B)	Event (C-I)	27	11010
				Behavior (C-D)	28	11011
		Quantitative	Colored	Event (C-I)	29	11100
				Behavior (C-D)	30	11101
			Not (W&&B)	Event (C-I)	31	11110
				Behavior (C-D)	32	11111

Table 1: Video object priority.

2.3 RTP Heard Management

Our analysis comprises 32-importance priorities. In order to cover this target, we use 5 bits in the RTP header. To keep the RTP header simplicity, we can neglect some of the lowest priorities found in table 1. To keep standardization, we add our video technique management bits at the padding filed. Hence, no additional bits are added in the essential RTP header fields [3].

3 Experiments and Results

After developing our video technique, a network simulator called NS2 can be used to evaluate its performance. NS2 is an object oriented discrete-event network simulator written in C++ with an Otcl interpreter as a fronted. It provides support for simulation of TCP/IP, routing and multicast protocols over wired and wireless local and wide area networks [4]. The target of this simulation part is to compare the performance of the video transportation (MPEG) using our technique and only merged two recovery techniques as an old system The parameters that will be used in our evaluation are bandwidth utilization, start-up loss rate, transmission loss rates, and video latency. In addition, some experiments to get an average loss rate for our video technique.

Our simulation comprises the network performance test under our technique and old technique (system). Also, we demonstrate the MPEG specifications. Hence, we study the behavior of MPEG transmission over the Internet and test the bandwidth utilization.

3.1 Simulation Setup

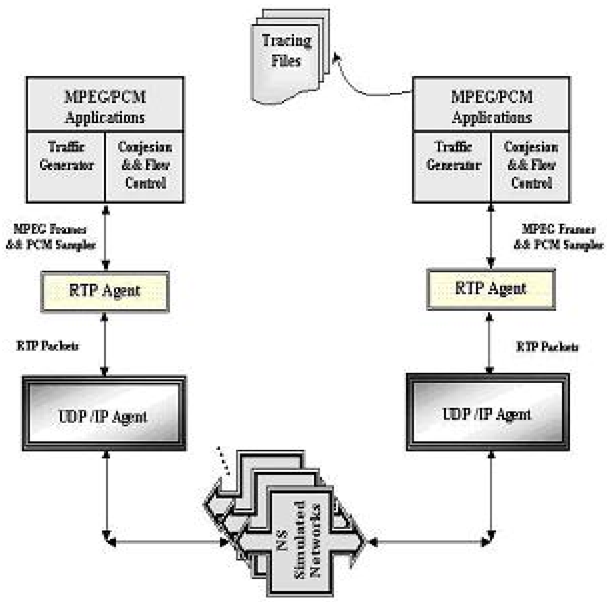

Figure (6): Simulation model structure.

In our simulation, we represent the old system by selecting different techniques to recover the lost video frames. But, merged two techniques are used during our simulation time. These recovery techniques are Forward Error Correction (FEC) and Retransmission, [5].

Definition	Value
Number of clients	5
Number of servers	1 Server
Session bandwidth	Minimally 512 kb/s
Video type	MPEG
Coding rate	20 Kb/s
Buffer size	10.000 bytes
Number of sent frames	14.000
Packet size	1Kb
Simulation time	100 seconds

Table 2: Simulation setup.

3.2 Network experiments and results

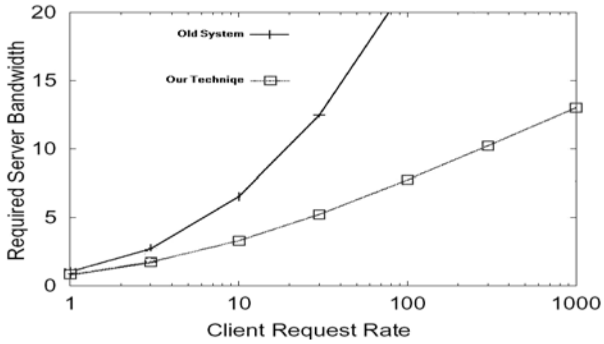

Figure (7): Requests required bandwidth

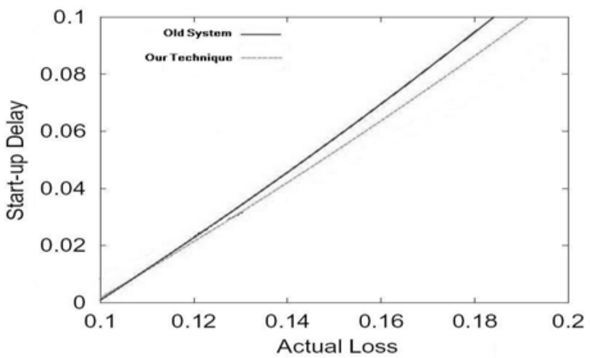

Figure (8): Start-up data loss rate

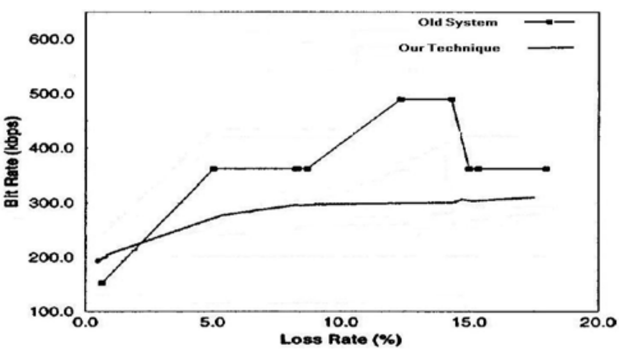

Figure (9): Transmission Loss rate

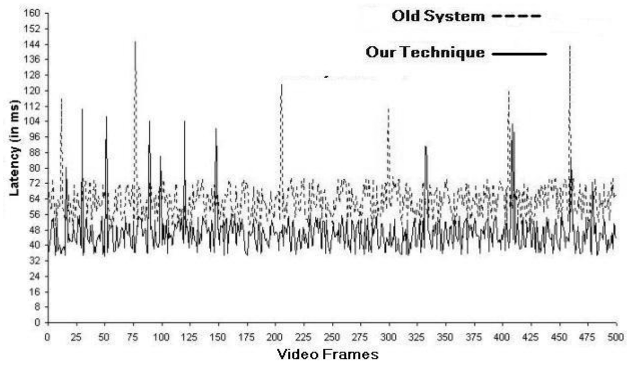

Figure (10): Video frames latencies rate.

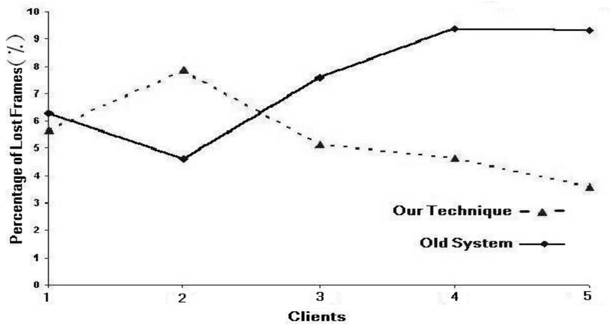

Figure (11): Client lost frames.

No	Avg. Loss Rate (Our Technique)	Avg. Loss Rate (Old System)	Avg. RTT (ms)
1	0.5	0.58	245
2	5.03	6.01	268
3	8.20	8.95	255
4	8.41	9.02	261
5	8.91	9.30	355
6	12.23	14.05	313
7	13.84	14.09	276
8	14.04	14.98	800
9	15.35	15.8	572
10	17.45	19.25	976

Table 3: Random video transmission trails with avg. loss rates for our technique and old system.

Fig. 7 represents the required server bandwidth when more clients requests video streams. As the video objects are arranged each depending on its importance, the bandwidth required for each request handling is reduced. It's notable in our technique plot that the required server bandwidth in our technique is less than the old system. This is due to mapping between client requests and importance of his requests. Fig. 8 shows the actual loss rates at the start of video streams transmission vs. the delay resulted from a start-up communication. The two plots in fig. 8 are narrow because the two systems have the same start-up delay. Fig 9 describes the loss rates during the video streams transmission. It's clearly notable that our technique is more efficient than the old system. And this results from the concentration on the important object, not the whole video streams like the old system. So, our technique bandwidth is utilized for basic video streams transmission or recovery of lost frames. Fig 10 demonstrates the latency time of video frames of our technique and the old system. Fig. 11 shows the number of lost frames at each client. It's notable that the second client has a starvation when compared with the old system. This is due to the faced bottlenecks and the huge fast changes that may cause congestion for that client. Table 3 contains a comparison between our proposed technique and other recovery techniques (i.e. old system) under different video applications and conditions.

3.3 MPEG experiments and results

3.3.1 MPEG specifications

The MPEG (Moving Pictures Expert Group) compression algorithm is a standard jointly developed by the International Organization of Standards (ISO) and the International Electro technical Commission (IEC). The MPEG algorithm utilizes the Discrete Cosine Transform (DCT) and motion entropy coding to obtain extremely high compression ratios [6]. For this project, a sample set of sixteen MPEG-1 compressed video sequences were studied to evaluate a number of measurement and traffic simulation techniques. MPEG compressed video streams are composed of a number of hierarchical elements. The highest layer is the Sequence layer. This layer is made up of an arbitrary number of Groups of Pictures (GOPs). In turn, each GOP is made up of a number of frames. The MPEG compression standard defines three frame types. The first frame type is the Intraor I frame. This frame is coded with reference to the current frame only. The second frame type, the Predictive or P frame differs from the I frame in that it is coded with reference to the current frame and a previous I or P frame. The third frame type, the Bidirectional or B frame, is even more complex because it is coded with reference to a past and a future I or P frame as well as the current frame. These complex interrelationships are shown in Fig. 12, [7].

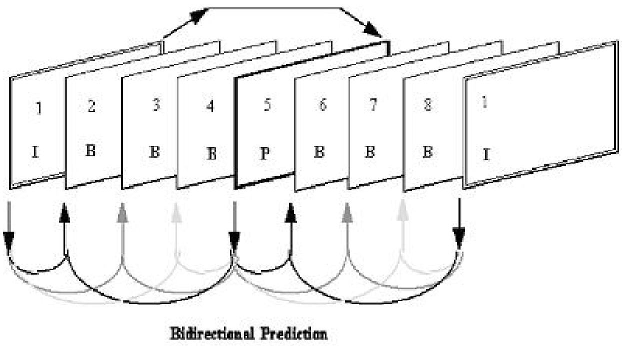

Figure (12): Group of MPEG frames (I, P, B)

Note 2: Through our experiments, we use I, P, B MPEG frames as a P2 instead of a type, see table 1.

3.3.2 MPEG performance study

In order to study the behavior of these MPEG video streams we simulate the link between the simulated multiplexer and the Internet connection, three traffic simulation experiments were conducted. First, in order to better observe the effect of highly correlated video traffic on the network, a traffic simulation consisting of multiplexed traffic of 16 identical and synchronized video streams was designed.

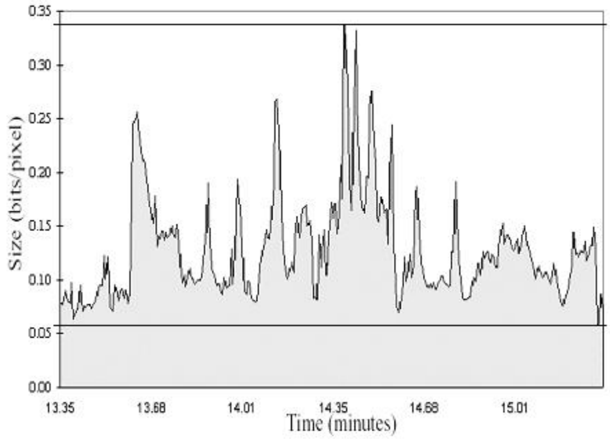

Figure (13): MPEG transmission (trace 1).

As can be seen in Fig. 13, the multiplexed traffic is quite bursty, ranging from approximately 0.35 bits per pixel to about 0.07 bits per pixel. Our simulated video average of 0.12 bits per pixel is greatly exceeded by the peaks in this graph, indicating potential difficulties for the network. Our simulation displays some of the longer correlations of the sample group. Sixteen synchronized copies of the movie were broadcast to facilitate comparison with multiplexed traffic simulations consisting of one copy of each video stream in the sample group. The second experiment used to measure the behavior of multiplexed MPEG traffic in the link differs only slightly from the first. In the second simulation, 16 copies of the same video sequence were again multiplexed and broadcast to the network. This time the video sequences were not synchronized. Instead, the start times of each video sequence were staggered 1.4 minutes from the previous start time. By starting a new movie every 1.4 minutes and running them in a circular manner, 16 copies of the video were always multiplexed into the network traffic. As shown in Fig. 14, staggering the video streams in this manner reduces the burstiness of the resulting network traffic, despite the fact that the video streams are identical and possess significant long-term correlations.

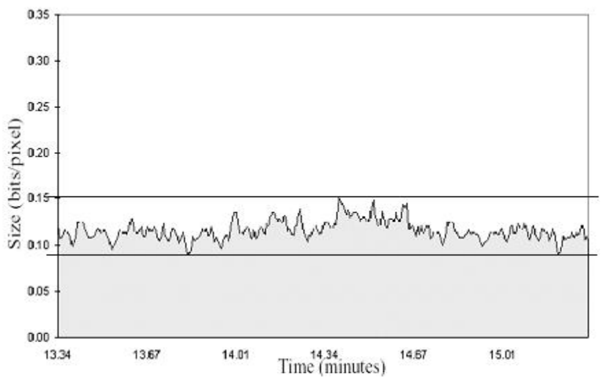

Figure (14): MPEG transmission (trace 2).

Comparing the traffic trace in Fig. 13 with that in Fig. 14 shows a significant decrease in burstiness when the start times are staggered. The same two-minute portion of the video trace in Fig. 14 shows a range from approximately 0.15 bits per pixel to 0.09 bits per pixel while the trace in Fig. 13 ranges from 0.35 to 0.07 bits per pixel. Staggering the traces decreases the effects of the correlations in the video stream and yields smoother traffic. To see how this result carries over to traffic, which consists of different video sources, the third experiment is a traffic simulation of 16 independent video sequences. As in the first experiment, the video sequences of the third experiment have synchronized start times. However, each stream is different from the others. The trace in Fig. 15 illustrates the behavior of the simulated traffic of the third experiment.

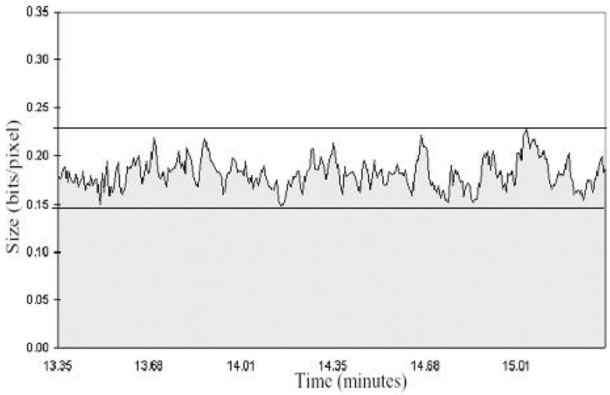

Figure (15): MPEG transmission (trace 3).

The traffic trace in Fig. 15 closely resembles the traffic trace in Fig. 14. There is little quantitative difference between the two traces. The trace of Fig. 15 yields a peak to average ratio of 2.4 while the trace of Fig. 14 yields a peak to average ratio of 2.0. The average of each of the three traces can be calculated from the characteristics of its component video streams. When compared to the trace in Fig. 13, the trace in Fig. 15 exhibits relatively smooth behavior.

Study of bandwidth utilization

Once the required bandwidth for the link has been calculated, it is necessary to determine the percentage of the bandwidth, which can be utilized for video transmission. Because of the burstiness of Variable Bit Rate (VBR) video sources, there will be periods of low bandwidth utilization [7]. During these periods, network efficiency can be increased by filling in the remainder of the available bandwidth with ABR or best effort traffic. However, the degree of utilization will remain somewhat less than one hundred percent or else increases in the bandwidth required by the video traffic will cause an unacceptable increase in cell loss in the filler traffic.

To study the limits of bandwidth utilization another traffic simulation experiment was conducted. In this experiment, a certain bandwidth was requested. With this bandwidth requested, the video traffic was then given priority over the filler traffic. Two levels of bandwidth utilization were then set. One, shown in Fig. 16, was set at 90 percent. The second, shown in Fig. 17, was set at 95 percent.

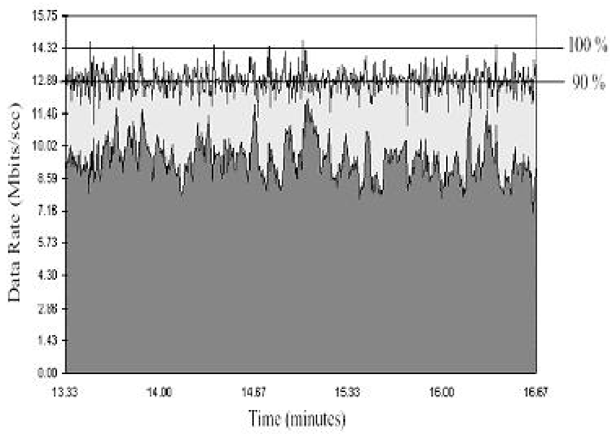

Figure (16): Bandwidth utilization: experiment 1.

Cell loss occurs in both Fig. 16 and Fig. 17 any time either or both of the data types exceed the requested bandwidth, in this case, 14.32 Mbit/sec. As can be seen in Fig. 17, with the bandwidth utilization set at 95 percent in this Data Rate (Mbits/sec) rough simulation, cell loss in the filler traffic becomes significant with over five percent of the peaks exceeding the requested bandwidth. Decreasing the bandwidth utilization to 90 percent greatly reduces the probability of cell loss, as shown in Fig. 16. However, this may present an exaggerated estimation of the cell loss in the actual link, because of the nature of the two data types. Pacing the video data into the network at the GOP level means each data point will have duration of approximately 0.4 seconds. Because of the higher temporal grain, which can be achieved by the filler traffic, the cell loss shown as lasting 0.4 seconds will actually last a fraction of that, reducing the duration of overly large peaks.

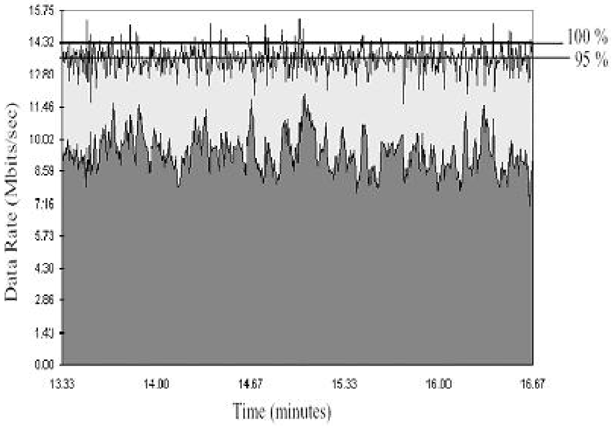

Figure (17): Bandwidth utilization: experiment 2.

4 Conclusion

In our paper, we demonstrated a new adaptive technique to recover the lost video frames. Our proposed technique arranges the video objects depending on there importance. It is adaptable as its sensitivity with the ideal recovery technique that should be used to recover the lost video frames at different video session times. Finally, we study our technique as regards a network performance and the transmitted MPEG behavior and showed the results.

5 Future Work

We shall try to test our technique as regards different video coding like JPEG. Also, we shall try to run our technique on a real system to show the video quality. Also, we shall test our technique under different network specifications.

References:

[1] Ze. Nian Li and Mark S Drew, *Fundamentals of Multimedia*, Prentice-Hall Inc. Upper Saddle River, NI, 2004.

[2] Omar S. Essa, NIKOS E MASTORAKIS," Adaptable Technique For Recovering Lost Internet Audio Packets, *WSEAS Int. conference of Data Networks, Communications, And Computers (DNCOCO'06)*, Bucharest, Romania, 16-18, October, 2006.

[3] H. Schulzrinne, S. Casner, and R. Frederick, and V. Jacobson, RTP: A Transport Protocol for Real-Time Applications, *Internet Draft*, March 2, 2003.

[4] McCanne, S. and Floyd, S. (McCanne, 1998), LBNL Network Simulator, *version*, 2003.

[5] G. Carle and E. Biersack, "Survey of error recovery techniques for ip-based audio-visual multicast applications," *IEEE Network*, Dec. 1997.

[6] B. Fuhrt, "Multimedia Systems: An Overview," *IEEE Multimedia*, Vol. 1, No. 1, Spring 1994, pp. 47-59.

[7] D. P. Heyman, A. Tabatabai, T. V. Lakshman, "Statistical Analysis of MPEG-2 CodedVBR Video Traffic," *Sixth International Workshop on Packet Video*, September 26-27, 1994, pp. B2.1-B2.5.