1999

A Model for Active Techniques for Compressed Video Transmission

Sheng-Yih Wang

Bharat Bhargava
Purdue University, bb@cs.purdue.edu

Report Number:
99-046

A MODEL FOR ACTIVE TECHNIQUES FOR COMPRESSED VIDEO TRANSMISSION

Sheng-Yih Wang
Bharat Bhargava

Department of Computer Sciences
Purdue University
West Lafayette, IN 47907

CSD TR #99-046
December 1999

Technical Report CS-99-046
A Model for Active Techniques for Compressed Video Transmission

Sheng-Yih Wang and Bharat Bhargava
Department of Computer Sciences
Purdue University
West Lafayette, IN 47907
E-mail: {swang,bb}@cs.purdue.edu

December 1999

1 Introduction

Compressed video transmission across IP networks have been a hot research topics for many years. Compressed video data are very sensitive to packet loss. Error control and concealment methods are often applied to alleviate the quality degrade due to data loss. The goal of the error control and concealment methods are to provide the best possible quality of video to the receiver under when packet loss are unavoidable.

The error control and concealment methods can be broadly classified into sender-based, receiver-based, or interactive methods. Sender-based methods try to preprocessing the data before sending it out to help the receiver recover as much data as possible when some of the data are lost. Sender-based methods include:

- Forward Error Correction (FEC). FEC adds redundancy data and/or applies error-correction codes before the data are sent. FEC mechanisms have been investigated by a lot of researchers [3].

- Layered Coding with Transport Prioritization. Layer coding can be designed in several different ways. For example, video frames can be encoded into a base layer and several enhancement layers. The base layer data is the most important data and need to be given a higher priority. Video frame can also be subsampled into several different sub-frames and each sub-frame is assigned a different priority.

Receiver-based methods try to postprocessing the data received and reconstruct the missing data using the data available. Receiver-based methods include:

- Spatial interpolation. The missing pixels are estimated using their neighbor pixels.

- Temporal prediction. The missing pixels are predicted from the previous or later frames.

Interactive error concealment methods involve both the sender and the receiver. Retransmission techniques are usually considered to be in this category. A survey of many of the error control and concealment techniques can be found in [6].

Although there are many researches going on on this topic, there are two important issues which have not been satisfactory resolved. The are:

- System-level integration of error control and concealment techniques. The source coding, transport protocol and postprocessing techniques should be designed together to achieve the optimal performance.

- Adaptable source-coding and transport-control mechanisms. In current error concealment approaches, there is very little interaction between the source coder and the transport layer. An optimal system should be adaptable to the error characteristic of the network and dynamically shift the burden of error concealment between the source coder and the transport layer.

In this paper we try to develop a general framework of evaluating the application of active techniques such as transcoding [1] and active fragmentation [4] to the compressed video transmission problem. We define a metric called *usefulness* which measures the portion of data received by the receiver that are *usable*. The fragmentation scheme proposed in [4] is used as an example application of the framework. The fragmentation scheme is based on an adaptable network architecture we proposed in [5]. In our proposed architecture, the packetized video data may go through a series of transformations before it reaches the destination. The objective of these transformation is to reduce the data size when congestion occurs. However, since each packet acts independently when travel through the network, the decoder must be able to deal with the packets from the same frame but are transformed differently when decoding the video frame. The active fragmentation scheme is very useful in this case.

2 The Model

Assume a data stream consist of n frames. Here we use the term frame to mean a complete data unit which can possibly be divided into smaller units (for example, a

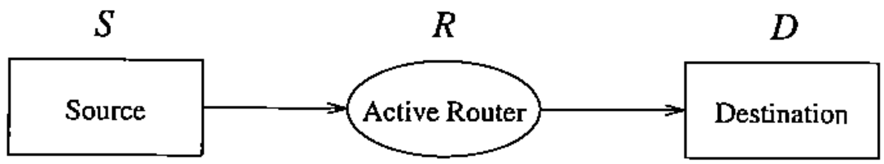

Figure 1: Network Setup

MPEG video frame can be divided into slices.). The size of frame i is denoted as f_i. Let M denote the maximum fragment size (or maximum transport unit, MTU) of the underlying transport network. For each frame, $\lceil \frac{f_i}{M} \rceil$ fragments will be generated when it is sent to the network. For simplicity, assume the network is configured as follows: the source S send the frames to the destination D through router R. Figure 1 shows the setup of the network. Each frame is assumed to be sent at time $t_{i,j}$ where i is the frame number and j is the fragment number of the frame i (Therefore i range from 1 to n and j range from 1 to $\lceil \frac{f_i}{M} \rceil$.). If no packet is lost and the router is not active (i.e. the router does not process the packets, it just forward them), the total amount of data received by D will be $\sum_{i=1}^{n} f_i$. Assume that the router applies a transform $\Gamma(s, l(t))$ to the packet of size s at time t with the load $l(t)$. In the simplest case, the router just forward the packet or drop it based on the load of the router (for example, the load can be derived from the queue size at time t.). Therefore the value of $\Gamma(s, l(t))$ can either be 0 or s depends on the size of the queue. In the case of active router, the possible value of $\Gamma(s, l(t))$ can be anything. Since we are interested in those cases where the network is congested, it make sense to restrict the value of $\Gamma(s, l(t))$ to be between 0 and s.

Based on the above definitions, the actual amount of data that will be received by D is

$$\sum_{i=1}^{n} \sum_{j=1}^{\lceil \frac{f_i}{M} \rceil} \Gamma(s(i,j), l(t_{i,j}))$$

the function $s(i,j)$ is the size of the packet received by the router and will take the value M most of the time except for the last fragment for each frame.

Not all data received by D is useful. For example, in TCP/IP protocol suite, the reassembly of the packet occurs at IP layer. If any one of the fragments is lost, all the other fragments from the same frame are useless. To model the usefulness of the received data, assume that the receiver D applies a function $U(\alpha, \beta)$ to the data fragments received for the same frame. α is the total data received for a particular frame. In other word, $\alpha_i = \sum_{j=1}^{\lceil \frac{f_i}{M} \rceil} \Gamma(s(i,j), l(t_{i,j}))$ for frame i. β is the frame size. For frame i, $\beta_i = f_i$. Since the importance of each frame is different, a weight W_i is associated with frame i.

Finally, the *usefulness* of the data received is defined as

$$\Upsilon = \frac{\sum_{i=1}^{n} W_i U(\alpha_i, \beta_i)}{\sum_{i=1}^{n} W_i f_i}$$

3 Analysis of the model

For simplicity, in the following analysis W_i is assumed to be 1 for all i.

3.1 Case 1: traditional router and IP fragmentation

In this case, the usability function is

$$U(\alpha, \beta) = \begin{cases} \beta & \text{if } \alpha \geq \beta \\ 0 & \text{if } \alpha < \beta \end{cases}$$

simply speaking, this means that if not all of the fragments are received, none of the fragments are useful.

Furthermore, assume that the packets are lost by a fixed probability p. The loss of packet could be the result of the router buffer overflow in the router or reassembly timeout at the destination. For simplicity of analysis, let's assume the usability function to be the identity function $U(\alpha_i, \beta_i) = \alpha_i$. Then the expected lost data will be

$$E_i = \min(\lceil \frac{f_i}{M} \rceil \times p, 1) \times \sum_{j=1}^{\lceil \frac{f_i}{M} \rceil} \Gamma(s(i,j), l(t_{i,j}))$$

and the *usefulness* of the data received is

$$\Upsilon_{IP} = 1 - \frac{\sum_{i=1}^{n} E_i}{\sum_{i=1}^{n} f_i}$$

A plot of Υ_{IP} with different p using the trace *Jurassic Park* is in figure 2. .

3.2 Case 2: traditional router and ACTP fragmentation

In this case, the usability function is different for different types of data. For text or numeric data, the usability model is still all-or-none. However, for multimedia data such as video or image, the usability function may look like figure 3.

For simplicity, let's assume the usability function is similar to figure 3. For example, the function

$$U(\alpha, \beta) = \beta \times e^{-(100\alpha/\beta - 100)^2/2000}$$

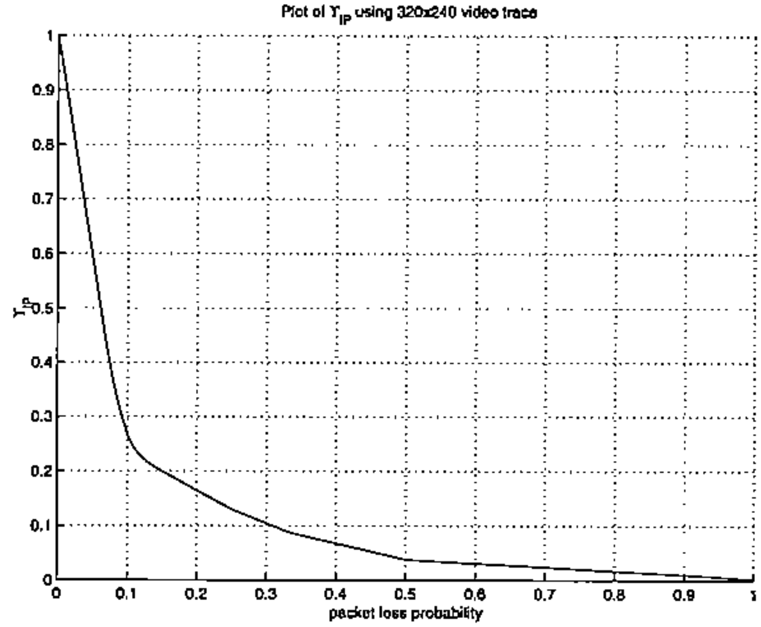

Figure 2: Plot of Υ_{IP} using the *Jurassic Park* trace (solid line: 3000 frames, dash line: 300 frames)

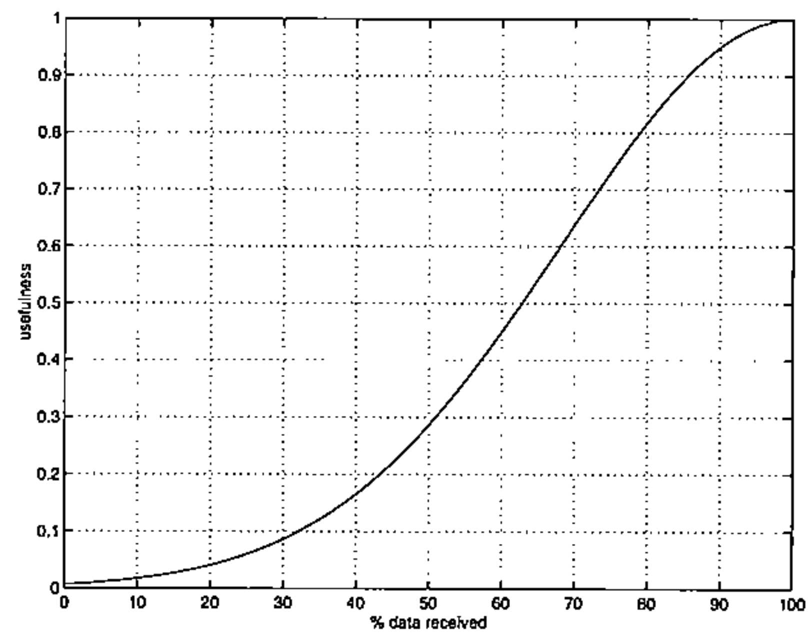

Figure 3: One possible usability function for multimedia data

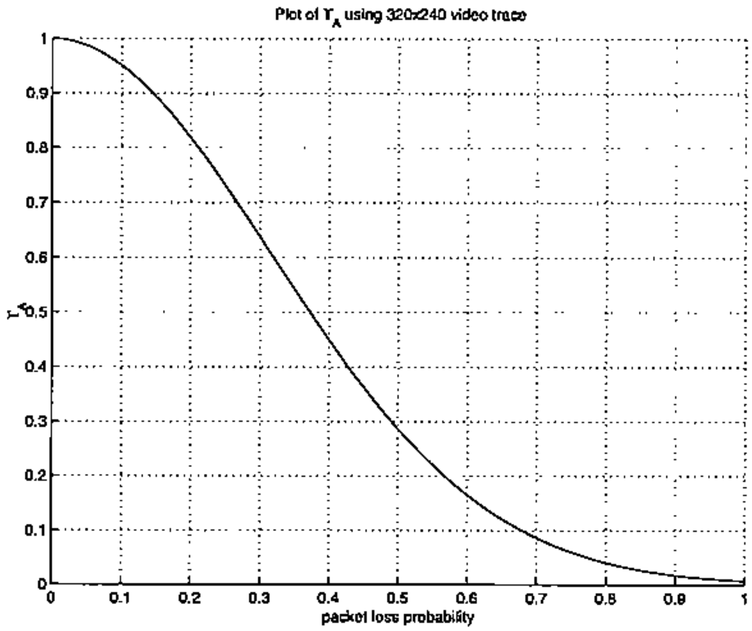

Figure 4: Plot of Υ_A using the *Jurassic Park* trace

could be used as a close approximation of the usability function in the figure 3. Then the expected lost data will be

$$E_i = p \times \sum_{j=1}^{\lceil \frac{f_i}{M} \rceil} \Gamma(s(i,j), l(t_{i,j}))$$

and the *usefulness* of the data received is

$$\Upsilon_A = \frac{\sum_{i=1}^{n} U(f_i - E_i, f_i)}{\sum_{i=1}^{n} f_i}$$

A plot of Υ_A with different p using the trace *Jurassic Park* is in figure 4.

4 Experiments

We have implemented a prototype video application suites that include a video encoder/server, an active router emulator, and a video decoder with error concealment modules. The application suites also includes a module to simulate ACTP (Active Transport Protocol) proposed in [5].

We have performed some preliminary experiments on the effect of ACTP fragmentation scheme on the improvement of the quality of service in adaptable video-on-demand applications. Figure 5 shows the setup of the experiments.

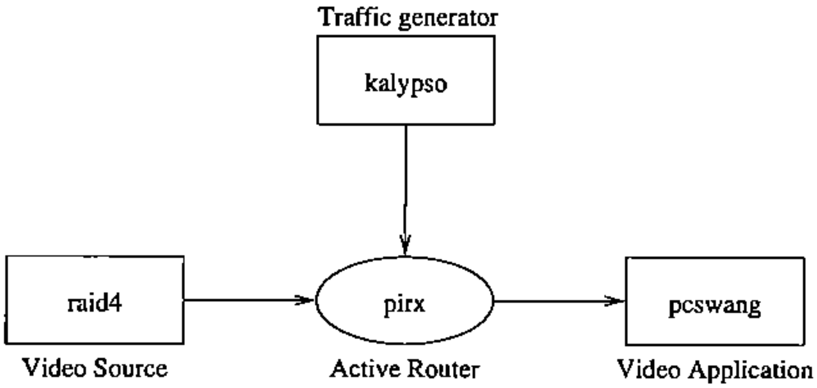

Figure 5: Experiment Setup

	frame rate	timer expired	playable frame	γ	perceptual quality
ACTP	3	237	287	0.80	good
	2	248	290	0.80	good
	1	235	286	0.83	good
UDP	3	182	119	0.17	bad
	2	187	114	0.16	bad
	1	185	115	0.16	bad

Table 1: Comparison of ACTP and UDP video session

In our experiments, we use a 300-frames video clip from the movie *Jurassic Park* (which we digitalized from the VHS tape to the resolution of 320x240 at 30 frames/sec.). Due to the limitation of the processing power in the client machine, we only conduct the experiments using the frame rate of 3, 2 and 1 frames/sec. For each experiment of frame rate n, a timer is set to expire at every $1/n$ second in the client program. When the timer expires, the expected frame number is increased by one and the data received for previous frames are considered late and discarded. We send the video data through a simulated active router program to the client. We conduct the experiments using our ACTP simulation and compare the result with regular (UDP/IP) approach. Some preliminary results are presented in table 1. The result shows that in UDP session for 3 frames/sec, the average timer expiration is 182 times, therefore in average only 119 out of the 300 frames are received and playable. In contrast, under ACTP session the timer expired 237 time in average, but there are 287 playable frames in average because most of the frame are partially received when the timer expired and most of these partially received frames can be made playable using simple error concealment techniques in the client program. By conducting real experiment and evaluate the subject perception quality, the ACTP session provide a very smooth playback, although there are some occasional block noise during the playback due to the lost slices. In contrast, regular playback session is very jittery because a lot of frame is not available. A runnable demo is available to visually see the impact of ACTP v.s. non-ACTP approach on video quality.

5 Discussion

- Our approach to the evaluation of usefulness in multimedia data is unique in the sense that we consider the data in the sub-frame level. Other works either consider the data at the packet level (each physical packet is a unit of data) or at the frame level (a frame is a data unit). For example, in [2] the author address the problem of MPEG stream packet loss by dropping related frames in the Group of Picture (GOP). The criteria used in other research works for the evaluation of the effectiveness of MPEG video transmission algorithm are frame-level parameters such as the percentage of I-frame received. Our view is that if the data are already sent and on the way, efforts should be made to deliver them to the destination even though the data may be partially lost. In our case, any data received by the receiver is useful, although they may have been transformed (reduced in size, etc.).

- Usability function U is very similar to Utility function in the literature. However, the usability function is applied to the packets at system level to determine the value of the data to the application. It is different in the sense that the function is calculated from the system point of view. Therefore the same amount of data received by the network software in OS can lead to very different usefulness level for the applications depending on the protocols used and other system parameters. For example, if different protocol suite are used, it is not easy to translate common QoS parameters such as the packet loss rate into the usefulness in the applications by a uniform measure (utility function).

- Currently we didn't include the timeliness parameter t in our analysis. In the future we plan to extend the model to explicitly incorporate the real-time constraint in multimedia data.

References

[1] Elan Amir, Steven McCanne, and Hui Zhang. An Application Level Video Gateway. In *ACM Multimedia'95, San Francisco*, 1995.

[2] Samrat Bhattacharjee, Kenneth L. Calvert, and Ellen W. Zegura. An Architecture for Active Networking. In *High Performance Networking (HPN'97), White Plains, NY*, April 1997.

[3] K. Park and W. Wang. AFEC: an adaptive forward error correction protocol for end-to-end transport of real-time traffic. In *Proceedings of International Conference on Computer Communications and Networks*, 1998.

[4] Sheng-Yih Wang and Bharat Bhargava. A fragmentation scheme for multimedia traffics in active networks. In *Proceedings of the 17th IEEE Symposium on Reliable Distributed Systems (SRDS'98)*, October 1998.

[5] Sheng-Yih Wang and Bharat Bhargava. An Adaptable Network Architecture for Multimedia Traffic Management and Control. Technical Report CSD-99-048, Purdue University, Department of Computer Sciences, November 1999.

[6] Yao Wang and Qin-Fan Zhu. Error Control and Concealment for Video Communication: A Review. *Proceedings of the IEEE*, 86(5):974–997, May 1998.