EFFICIENT PARTITIONING OF UNEQUAL ERROR PROTECTED MPEG VIDEO STREAMS FOR MULTIPLE CHANNEL TRANSMISSION

Wen Xu, Sheila S. Hemami

Cornell University, School of Electrical and Computer Engineering, Ithaca, NY 14853
email: {wxu, hemami}@ece.cornell.edu

ABSTRACT

Reliable transmission of video over wireless networks must address the limited bandwidth and the possibility of loss. When the bandwidth is insufficient on a single channel, the video can be partitioned over multiple channels with possibly unequal characteristics at the expense of more complex channel coding (error correction). This paper addresses the problem of efficiently partitioning forward error protected, pre-encoded video data for transmission over multiple channels. The assumption of pre-encoding precludes adjustment of source rates to the channels, since it is assumed that channel characteristics are not known until immediately prior to the start of transmission. The proposed partitioning exploits the structure of MPEG video, and frames in each group-of-picture are reordered based on their decoding dependence. To be spectrally efficient, the frames of different types are unequally error protected taking different channel reliabilities into account. A pruned tree search algorithm is implemented to efficiently solve the problem. Simulation results are presented.

1. INTRODUCTION

Wireless systems are characterized by limited bandwidths and high loss rates, while streaming video requires dedicated bandwidth and very reliable transmission. Bandwidth limitations can be overcome when multiple channels are aggregated into one higher bandwidth channel for transmission of a single video sequence [1]. Forward error correction (FEC) allows recovery from loss without retransmission by introducing controlled redundant data. Applying FEC across video data packets has been shown to be effective in an ATM environment [2]. Due to limited available bandwidth the error control mechanisms must also be spectrally efficient. Unequal error protection (UEP) is especially well suited to video. In previous work, video bitstreams have been divided into multiple partitions with different decoding importances at the macroblock level [3], the slice level [4] or the frame level [5] (further partitioning increases complexity and decreases efficiency). However, each of these techniques is applied to transmitting video over a single channel or can be extended to transmit video over multiple channels with different quality of service (QoS) guarantees. Bandwidth aggregation and error control can be combined to achieve higher rate reliable transmission while allowing more flexible data partitioning. Optimization of this combination is addressed in this paper.

This paper proposes a framework that uses UEP to balance multiple channels while retaining UEP on the frame level. The framework is designed for pre-encoded video sources where adjustment of source rates to the channels is precluded. It is assumed that channel characteristics are monitored and estimated at intervals during the transmission process, giving bandwidth, latency, and packet loss rate (PLR) (e.g., [6]). The variations in PLRs are considered in FEC selection for each channel. An algorithm is proposed which efficiently partitions FEC-protected video data across multiple channels to minimize the bandwidth consumption when the aggregate bandwidth is sufficient to transmit the video or maximize the decodable frame rate when the aggregate bandwidth is limited. Both vaying channel conditions and frame decoding dependence are considered. Although the proposed approach is illustrated using MPEG video, it can be extended to MPEG-2 or any other layered coding schemes.

This paper is organized as follows. Section 2 outlines the transmission model on which the partitioning problems are based. In Section 3, the optimal-partitioning problems are formulated and solved. Experimental results are presented in Section 4. Section 5 concludes the paper.

2. THE TRANSMISSION MODEL

The partitioning problem and solution will be illustrated using the MPEG compression standard [7], where the I-P-B structure provides an implicit transmission priority for the different types of frames. For example, the entire GOP cannot be decoded without the I-frame, while non-reception of a B-frame does not affect decodability of any other frames.

While the low bandwidth provided by wireless systems is often insufficient to support streaming video over a single channel, multiple channels can be used to provide a logical channel with a sufficient (or nearly sufficient) bandwidth. Assume that there are N ($N \geq 2$) channels having various characteristics (bandwidth, latency, PLR) available for streaming one video sequence. It is assumed that channel characteristics are monitored and estimated at intervals during the transmission process (e.g., [6]). Each channel is thus parameterized by its bandwidth $BW_i(t)$, delay $T_i(t)$ and PLR $p_i(t)$ ($i = 1, 2, \ldots N$). A packet is considered lost either if it is physically lost due to buffer overflow or channel fading, or if it is delayed and arrives at the decoder after its decoding time.

FEC across packets is applied to each coded frame in order to recover lost packets. Because the multiple channels have different PLRs, data transmitted over each channel requires a different protection level. Because the MPEG frames have priorities in terms of GOP decoding, UEP is applied at the frame-level. Assume a UEP encoding scheme encodes packets by adding redundancy by factor $R_i(A, t)$ ($i = 1, 2, \ldots N$, $A \in \{\text{I, P, B}\}$). $R_i(A, t)$ is determined by the frame type A and the current estimate of PLR $p_i(t)$. Higher redundancy (larger factor R) is given to more important frames. Furthermore, the same frame would assume different sizes after

FEC encoding in channels with different PLRs. After FEC protection the channel reliabilities are assumed to be approximately the same for a given frame type, though they vary for different types. Implementation of a UEP encoding scheme requires a relation between PLRs, decoding visual quality, and various error correction codes (ECC), which is beyond the scope of this work; rather this paper will address how to partition the video data assuming that the redundancy factors are known (note that this follows once specific ECCs are selected). Regardless of the specific ECCs selected, the following inequalities hold based on the above assumptions:

$$R_i(\mathrm{I}, t) \geq R_i(\mathrm{P}, t) \geq R_i(\mathrm{B}, t) \tag{1}$$

$$R_i(A, t) \geq R_j(A, t) \text{ if and only if } p_i(t) \geq p_j(t) \tag{2}$$

where $i,\ j = 1,\ 2,\ \ldots N$ and $A \in \{\mathrm{I, P, B}\}$.

The video data for each GOP is partitioned as a group, though the group could consist of integer numbers of GOPs or in general a segment of video data that is coded independently. The transmission process is thus described in discrete time with step size of the duration of a GOP, denoted as T_G. Assume the channel conditions vary in time at the order of T_G seconds and the interval of channel estimates is of multiple GOP durations. Then, bandwidth, delay, and PLR are approximated to be constants within a step time T_G though they can vary throughout the duration of the entire video sequence; i.e., $BW_i(nT_G)$, $T_i(nT_G)$, and $p_i(nT_G)$ ($n = 0,\ 1,\ 2,\ \ldots$) (simplified as $BW_i(n)$, $T_i(n)$ and $p_i(n)$ in the following) represent bandwidth, delay and PLR in the $(n+1)$-st GOP duration, respectively. Redundancy factors are simplified as $R_i(A, n)$ hereafter.

The physical decoder buffer size is assumed to be large enough so that channel bandwidths are the primary constraints [8].

3. THE OPTIMAL PARTITIONING PROBLEM

In the frame-level video partitioning problems a GOP consisting of Q frames is to be transmitted over N channels with differing bandwidths, PLRs, and latencies. In this section the optimization problems are formulated and the algorithm is presented.

3.1. Problem Formulation

Based on the transmission model, there are N^Q ways to divide a GOP into N non-overlapping sets corresponding to N channels at each time step. Let $S_i(n)$ denote the set of frames that are allocated to channel i and S be the set of all frames (or all transmitted frames) in a GOP. These sets have the following properties:

$$S_i(n) \cap S_j(n) = \phi \qquad \text{if } i \neq j \tag{3}$$

$$S_1(n) \cup S_2(n) \cup \ldots \cup S_N(n) = S \tag{4}$$

where $i, j = 1, 2, \ldots N$.

$F(A_j)$ refers to the non-FEC-coded data size of frame A_j of type A. When a frame A_j is assigned to channel i, redundancy is added in the fractional amount of factor $R_i(A, n)$, which increases its frame size to $(1 + R_i(A, n)) \cdot F(A_j)$. Thus the total amount of data to be allocated to channel i ($i = 1,\ 2,\ \ldots N$), denoted as E_i, is

$$E_i = \sum_{A_j \in S_i, A_j \text{ is of type } A} (1 + R_i(A, n)) \cdot F(A_j) \tag{5}$$

The partitioning optimization is performed for each GOP in a video sequence. To simplify the notation below, the time index n is omitted.

3.1.1. *Achieving Minimum Bandwidth Consumption*

Assume the channels are sorted in ascending order of reliability, i.e., channel 1 has the largest PLR while channel N has the smallest. To formulate the optimization problem, assume that the bandwidth of channel 1 is unconstrained. The more reliable channels 2 through N are preferred but unable to transmit the entire FEC-coded GOP within T_G seconds. The goal is to minimize the required bandwidth of channel 1 with the constraints of channel 2 to N bandwidths. Let the solution be partition $S = S_1^* \cup S_2^* \cup \ldots \cup S_N^*$ that gives:

$$\min E_1 \tag{6}$$

subject to $E_i \leq BW_i \cdot T_G$, $i = 1,\ 2,\ \ldots N$, where E_i are given by (5). This formulation achieves the highest possible utilization of reliable channels under the constraints.

3.1.2. *Achieving Maximum Frame Rate*

It is desirable to transmit the entire GOP, but in many situations even the aggregate bandwidth is so limited that some of the frames have to be dropped. In this case the objective is to transmit as many frames as possible in one GOP within the limited bandwidth-time product, $BW \cdot T_G$. In order to effectively use the bandwidths, there is no need to deliver a frame if any one of the frames required to decode it is not delivered. At playback the missing frames can be filled in by replaying the latest decoded one, or by using a more sophisticated frame reconstruction technique [9].

The frames in a GOP are therefore reordered based on their encoding dependencies. The reordering occurs prior to the partitioning process at each time step and completed by software processing based on the *a priori* knowledge of the pre-encoded video sequence. In the reordered sequence the j-th ($1 < j \leq Q$) frame is decodable on the condition that all frames from 1 to $j-1$ are transmitted. Therefore, frames from the end of the reordered sequence can be removed without affecting decodability of the remaining frames. I- and P-frames are placed in early positions obviously, but B-frames can be ordered, for example, to minimize the interval of unplayable frames, or, to maximize the playable frame rate with reasonable gaps. The reordering pattern can be determined prior to the start of transmission for fixed GOP patterns.

Following reordering, the optimization problem is to maximize the number q ($1 \leq q \leq Q$) of transmitted frames along the reordered sequence such that the bandwidth constraints remain satisfied, i.e.

$$\max_{\{q: 1 \leq q \leq Q,\, q = \sum_{i=1}^{N} |S_i|\}} q \tag{7}$$

subject to $E_i \leq BW_i \cdot T_G$, $i = 1, 2, \ldots N$, where E_i are given by (5).

The MPEG-2 video-compression standard [7] allows scalability. By combining certain frames' base and enhancement layers and rearranging the frame orders, the above-described formulation can be easily extended to work with scalable MPEG coding. The following discussions assume single layer MPEG encoding; the extension to scalable encoding is straightforward.

3.2. Incorporation of a Delay Constraint

In typical video transmission, the end-to-end delay each frame experiences consists of several components, such as en/decoder buffering delay, and propagation delay. Here, only components that vary among multiple channels are considered.

The delay is defined as the time interval between the start of a partitioned GOP transmission and the moment when all delivered frames of the GOP have arrived. It consists of two components: transmission delay, given by (total bits)/bandwidth, and channel propagation delay T_i as estimated. Let T_D be an upper bound to the delay of all channels. Following the previous notations, the delay in channel i is

$$D_i = \frac{E_i}{BW_i} + T_i \leq T_D \qquad i = 1,\, 2,\, \ldots N \tag{8}$$

By (8), the constraints in the problem formulations (6) and (7) are modified to be

$$E_i \leq BW_i \cdot (T_D - T_i) \qquad i = 1,\, 2,\, \ldots N \tag{9}$$

It is desirable to avoid the actual determination of T_D in order to simplify the formulation and algorithm. Equation (9) can be shown to coincide with the previous constraints in (6) and (7) where ($T_D - T_i$) is replaced by the GOP duration T_G under certain assumptions. This replacement is justified when the processing time associated with partitioning a GOP and performing the FEC encoding is negligible. While the latter is true, the former requires a fast partitioning algorithm, which is proposed in the next section. If the variation in propagation delays T_i is large, however, the replacement of ($T_D - T_i$) with T_G results in an overestimated solution for these channels. Therefore, it is preferable that the propagation delays have small variations.

3.3. Pruned-Tree-Search Algorithm

As noted above, the search space is $O(N^Q)$ for the solution to the optimization problems, which grows exponentially when the number of channels or frames in a GOP increases. A Pruned-Tree-Search (PTS) algorithm is implemented to find the optimal solution more efficiently than a full search. Every node in the tree represents a possible solution. Tree depth corresponds to number of frames and branches represent channel assignments. Constraints, as given in (6) and (7), are associated with each node. By evaluating constraints starting from the root and pruning impossible subtrees, the search space is greatly reduced. While the PTS algorithm has a worst-case complexity of $O(N^Q)$, experiments indicate that it finds the optimal solution at a significant speed-up due to the data profile of typical video traces.

When channel estimates are inaccurate or delayed at the time they are needed, greedy adjustment of the solutions are made according to the actual channel conditions. Frames at the end of the reordered sequence are either added to the solution or dropped from it greedily when the actual bandwidth deviates from the estimates.

4. EXPERIMENTS AND RESULTS

A number of experiments have been performed to evaluate the performance of the partitioning algorithm under various channel conditions. Real-life MPEG video traces varying in content including Movies (*Star Wars*), Sports (*Soccer*), *News* and *MTV* were used in the performance evaluation. The frame size traces were extracted from MPEG-1 sequences (384 x 288, 25 frames/s, 40,000 frames) encoded with the Berkeley MPEG- encoder (version 1.3) with the following parameters [1]: YUV(4:2:0,8bits), Quantization value (I=10, P=14, B=180), 'Logarithmic' / 'Simple' motion vector search, 1 slice and half pel motion estimation. The GOP's are of pattern IBBPBBPBBPBB with fixed size 12 [10].

In the following experiments, $N = 3$ and the redundancy factors are taken as ($R_1(\mathrm{I}) = 0.48$, $R_1(\mathrm{P}) = 0.16$, $R_1(\mathrm{B}) = 0.03$), ($R_2(\mathrm{I}) = 0.35$, $R_2(\mathrm{P}) = 0.10$, $R_2(\mathrm{B}) = 0$), ($R_3(\mathrm{I}) = 0.24$, $R_3(\mathrm{P}) = 0.08$, $R_3(\mathrm{B}) = 0$), respectively, based on observations in [11].

Occasionally extremely large I-frames can cause low bandwidth utilizations in solving the problems. In this case, I-frames are further segmented into smaller parts, each of which is considered the same way as a frame. Each segment introduces additional overhead; therefore, the number of segments should be kept as low as possible. The number of segments is determined based on factors such as number of channels, channel bandwidths, and the ratio of data amount in an I-frame to that of the entire GOP. In the experiments below, 3 segments are used when needed.

4.1. Minimizing Bandwidth Consumption

For the first optimization problem, experiments show that the formulation achieves a high bandwidth utilization of channels 2 to 3 and the PTS finds the optimal solution as found by a full search with a reduced computation time. In the experiments channels 2 and 3 have bandwidth-time products given by a fixed factor of the current GOP data amount to remove the effect of varying video bit rate and obtain an average performance of the algorithm. In practice channel bandwidths may be allocated by a reservation plan or vary independently of video bit rate, but the evaluation holds whenever the bandwidth of the reliable channels is limited as assumed in the formulation.

Table 1 lists the normalized bandwidth utilization of channels 2 and 3 and the speedup in computation time relative to that of a full search for various video traces. The algorithm is the slowest when the bandwidths are equally distributed because the pruned tree has more balanced branches. The computation time reduces significantly (by a factor of 5-10) when the channel bandwidths are very unequal. The resulting utilization varies within $\pm 2\%$.

Table 1. Normalized bandwidth utilization and the speedup relative to a full search. Channels 2 and 3 each have bandwidth as a factor (0.35) of GOP data size. The utilization is normalized over the two channels. Results are averaged over 20 consecutive GOPs randomly selected from the sequences. T_G = 0.5 sec.

Video clips (bits/s)	Normalized BW utilization (%)	Speedup
Star Wars (233k)	99.5	3.18
Soccer (678k)	98.8	3.44
MTV (615k)	99.0	3.60
News (517k)	98.7	3.67

This problem can also be formulated as the minimization of total bandwidth consumption of the N channels, rather than that of channel 1. The formulation presented here can be computed more quickly and achieves a higher utilization of reliable channels 2 to N by up to 7%. The differences in the total bandwidth consumed is under 2% of the original GOP size per T_G seconds.

4.2. Maximizing Frame Rate

Three experiments are reported here due to space limitations. The computation time required to solve this problem is reduced relative

---

[1] (MPEG traces available via anonymous FTP from ftp-info3.informatik.uni-wuerzburg.de/pub/MPEG/)

to that of the first by one or two orders of magnitude.

4.2.1. Comparison with Greedy Partitioning

A greedy algorithm alone allocates the frames to the channels, beginning with the best and ending with the worst. However, without frame reordering some received frames are not decodable because other frames needed to decode them are not transmitted, resulting in a reduced frame rate on average.

The average decodable frame rates achieved by the proposed and a greedy algorithm are compared, with total available bandwidth ranging from 70% to 150% of the video average bit rate and three channels of equal bandwidth. For the four sequences, the proposed algorithm outperforms a greedy algorithm, with maximum differences ranging from 0.64 to 2.12 frames/GOP and minimum differences ranging from -0.07 to 0.64 frames/GOP. The differences decrease at the two extremes of available bandwidth. The only negative number appears in *MTV* because the greedy algorithm transmits multiple B-frames prior to some large P-frames, resulting in more frames transmitted. In contrast, however, the proposed scheme provides an evenly distributed spacing of frames within a GOP due to the delivery of P-frames, which can help reduce the effect of time aliasing when frame rate up-conversion is performed at the receiver.

4.2.2. Effects of Differing Video Content

Various video content generates different patterns of frame size traces. For example, *News* has a large variation across both frames and GOPs, while the variation of *Star Wars* is small. These characteristics impact the achieved frame rates when the total available bandwidth varies. Figure 1 illustrates their average decodable frame rate over 100 GOPs when the total bandwidth ranges from 50%- 200% of the video average bit rate. *News* achieves the highest average frame rate when the available bandwidth is equal to or greater than its average bit rate. At lower bandwidths, however, the average frame rate drops significantly because *News* has very large I-frames in most GOPs which cannot be transmitted. The average frame rates of the other three traces increase more smoothly when the normalized channel bandwidth increases and reach above 10 frames/GOP when the total bandwidth is 150% or higher.

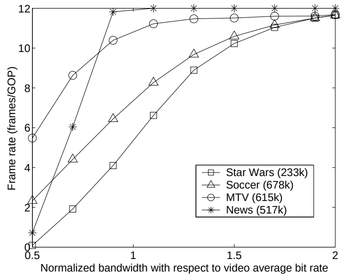

Fig. 1. Average frame rate (frames/GOP) over 100 GOPs vs. total channel bandwidths (normalized with respect to video bit rate) for various video traces. 3 channels have equal bandwidth.

4.2.3. Effects of Channel Parameter Mismatch

The performance of the transmission model depends on the accuracy of the channel parameters provided. Experiments reveal that at 130% estimated normalized bandwidth the average frame rate drops by 6-7 frames/GOP when the actual bandwidth is -15% lower than the estimates and increases by 1-2 frames/GOP when the actual value is +5% higher. Modifications can be made to the partitioning scheme to increase the robustness to parameter mismatch at the expense of reduced bandwidth utilization.

5. CONCLUSIONS

This paper has proposed formulations with a pruned tree search algorithm to optimally partition pre-encoded video sequences for transmission over multiple channels with various conditions. The available bandwidths of the channels are the primary constraints in the partitioning process. Minimum channel bandwidth consumption is achieved after the partition when the aggregate bandwidth is sufficient and maximum decodable frame rate is achieved if only limited bandwidth is available.

The partitioning for maximum frame rate experiments provide useful insight for source encoder design and it can be combined with rate control to some extent when the video source is not pre-encoded and achieves a smoother playback frame rate. The partitioning requires between milliseconds to seconds on a 1GHz processor in Matlab depending on video content and available bandwidth, so it is expected that it can be reasonably implemented in dedicated software or hardware at the sender.

6. REFERENCES

[1] P. H. Fredette, “The past, present, and future of inverse multiplexing,” *IEEE Commun. Magazine*, vol. 32, no. 4, pp. 42–6, Apr. 1994.

[2] E. W. Biersack, “Performance evaluation of forward error correction in an ATM environment,” *IEEE JSAC*, vol. 11, no. 4, pp. 631–40, May 1993.

[3] H. Gharavi and C. I. Richards, “Partitioning of MPEG coded video bitstreams for wireless transmission,” *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 153–55, June 1997.

[4] T. Tuan and R. M. A. P. Rajatheva, “Unequal error protection for MPEG-2 video transmissions over frequency-selective Rayleigh fading channels,” in *Proc. SPIE*, Perth, WA, Australia, June 2000, vol. 4067, pp. 152–63.

[5] I. E. G. Richardson and M. J. Riley, “MPEG coding for error-resilient transmission,” in *Proc. IEE Image Processing and Its Applications*, Edinburgh, UK, July 1995, pp. 559–63.

[6] Q. Zhang et al., “Resource allocation with adaptive QoS for multimedia transmission over W-CDMA channels,” in *Proc. WCNC*, Chicago, IL, Sept. 2000, vol. 1, pp. 179–84.

[7] B. Haskell etc., *Digital Video: An Introduction to MPEG-2*, Chapman & Hall, London, UK, 1997.

[8] A. R. Reibman and B. G. Haskell, “Constraints on variable bit rate video for ATM networks,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, no. 4, pp. 361–72, Dec. 1992.

[9] R. Castagno et al., “A method for motion adaptive frame rate up-conversion,” *IEEE Trans. CSVT*, vol. 6, no. 5, pp. 436–46, Oct. 1996.

[10] O. Rose, “Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems,” in *Proc. LCN*, Minneapolis, MN, Oct. 1995, pp. 397–406.

[11] M. Gallant and F. Kossentini, “Rate-distortion optimized layered coding with unequal error protection for robust internet video,” *IEEE Trans. CSVT*, vol. 11, no. 3, pp. 357–72, Mar. 2001.