Bandwidth Allocation and Analysis of VBR MPEG Video Bit Streams
R. Egorov, S. Gringeri, K. Shuaib, A. Lewis, and B. Basch
GTE Laboratories Incorporated
40 Sylvan Road, Waltham MA 02454

Abstract - This paper presents the benefits and challenges in the encoding and delivery of variable bitrate (VBR) MPEG video. The network resources required to transmit stored variable rate MPEG can be reduced by properly analyzing and smoothing the video stream before transmission. A scheduling technique is presented which selects a traffic contract for a pre-encoded MPEG video stream with the criteria of minimizing network resources and maintaining video quality. Several effective bandwidth metrics are discussed and used to model the potential saving in network resources for the shaped streams.

I. Introduction

Transmission of real time video streams is resource intensive even when the video is compressed using sophisticated algorithms like MPEG [4,5]. Variable bitrate encoding can achieve improved coding efficiency by better matching the encoding rate to the video complexity. Variable rate encoding is currently used in storage applications such as Digital Versatile Discs (DVDs) to achieve significant storage savings.

This paper discusses several issues related to shaping VBR video to control its burstiness and produce streams that are more suitable for transmission over switched ATM networks. Simulations were conducted to determine the required ATM network resources and traffic contracts for various video categories (action video, talking head, etc). The potential savings are analyzed for single streams and groups of streams. The bandwidth savings were quantified using several effective bandwidth metrics [1,2,3].

In this paper a video shaping method that consists of two phases, an analysis phase, which is done before transmission, and a transmission phase is proposed.

In our simulation and analysis several MPEG video clips were used. The MPEG-1 clips used were approximately 30 minutes long, encoded at 24fps with a resolution of 384x288. Variable rate encoding was achieved for the MPEG-1 clips by using a fixed quantization scale for each picture type, 10 for I frames, 14 for P frames, and 18 for B frames. The MPEG-2 video clips used were approximately 15 minutes long encoded at main profile/main level with a resolution of 720x480 and a frame rate of 30fps using a variable rate MPEG-2 commercial hardware encoder. The MPEG-4 video clips were 15 minutes long with a resolution of 160x120, at 30fps, generated using a fixed quantization scale for each picture type, 8 for I frames, 10 for P frames and 12 for B frames. All specified rates, shown in Table 1, include ATM and AAL-5 overheads. The peak rate was calculated as the maximum rate over half a second for the MPEG-1 clips and over a second for the MPEG-2 and MPEG-4 clips. The MPEG recommended decoder buffer sizes of 1.8 Mbits for MPEG-2 and 328 kbits for MPEG-1 were used. Since the MPEG-4 standards do not specify a decoding buffer size, we selected (somewhat arbitrary) a decoding buffer size of 164 kbits for our analysis.

<table>
<thead>
<tr>
<th>Video Clip</th>
<th>Type</th>
<th>Peak (Mbps)</th>
<th>Mean (Mbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action Video2</td>
<td>MPEG-2</td>
<td>6.65</td>
<td>3.5</td>
</tr>
<tr>
<td>Talking Head</td>
<td>MPEG-2</td>
<td>4.4</td>
<td>2.6</td>
</tr>
<tr>
<td>talk2</td>
<td>MPEG-1</td>
<td>1.576</td>
<td>0.45</td>
</tr>
<tr>
<td>video2</td>
<td>MPEG-1</td>
<td>1.62</td>
<td>0.45</td>
</tr>
<tr>
<td>Action4</td>
<td>MPEG-4</td>
<td>0.329</td>
<td>0.0835</td>
</tr>
<tr>
<td>Talk4</td>
<td>MPEG-4</td>
<td>0.302</td>
<td>0.123</td>
</tr>
</tbody>
</table>

II. Analysis of Video to Determine Transmission Requirements

The goal of the pre-transmission video analysis phase is to determine the needed traffic control parameters such as the Peak Cell Rate (PCR), the Sustainable Cell Rate (SCR) and the Maximum Burst Size (MBS) for the transmission of a VBR stored video clip. This traffic contract is based on achieving the best utilization of bandwidth while guaranteeing the desired QoS requirements by the application (no decoder buffer
underflow or overflow). This approach will be referred to as just-in-time scheduling analysis. With just in time scheduling, a model of the decoder buffer is kept at the source, the system will operate at the selected sustained rate unless buffer underflow or overflow conditions occur. In both cases the rate is modified to prevent loss of data. In the case of overflow the rate is simply reduced so that the buffer never fills up. For underflows the just-in-time scheduling determines the rate that is required to guarantee arrival of the data just in time for processing. For a given sustained cell rate the just-in-time scheduling analysis simulation produces the peak cell rate and sustainable burst size (MBS) \[2\] values needed to guarantee no decoder buffer underflows or overflows, when the video clip being analyzed is transported. The sustainable burst size can be related to the maximum burst size using the following formula:

\[
MBS_s = \frac{(PCR - SCR) \times MBS}{PCR}
\]  

(1)

Performing the analysis on a frame by frame basis produces a large PCR value. Therefore, the just-in-time scheduling simulation can look ahead a selectable number of frames, average their rates and obtain a smaller PCR. From the definition of sustainable burst size it is apparent that its value exhibits little variation for different look ahead periods since its value greatly depends on the SCR and a lower PCR value will be offset by a larger MBS. In general, the PCR of a specific video stream is reduced for larger look ahead periods. Larger look ahead periods imply smoother transmission, which should require fewer network resources. In practice, as the look ahead period approaches one second of video, little variations are noticed in the traffic contract parameters and therefore, the required network resources are about the same. Therefore, our analysis was generally limited to a look-ahead of approximately one second. Table 2 shows the effect on the ATM traffic contract found for the MPEG-2 Action Video2 clip when different look ahead periods are used in the just-in-time scheduling analysis for a specific SCR. To show how the just-in-time scheduling analysis adjusts the rate based on the decoding buffer occupancy to prevent overflows or underflows, a snapshot of the relationship between the decoder buffer occupancy and the selected rates is shown in Figure 1.

Using the just-in-time scheduling simulation with look ahead, a traffic contract for a stream with user-specified sustained cell rate, can be obtained based on maximizing the decoder buffer utilization and minimizing the required PCR without increasing the needed MBSs. This method promotes an effective utilization of the decoder buffer. This is so since during easy video passages, transmission at the sustained rate will tend to completely fill the buffer, whereas for difficult passages the rate will be selected so that the data arrives just in time, and the buffer is potentially as empty as possible without underflow.

Table 2: Effect of look ahead on determining a traffic contract for Action video2

<table>
<thead>
<tr>
<th>SCR (Mbps)</th>
<th>Look ahead</th>
<th>PCR (Mbps)</th>
<th>MBS (cells)</th>
<th>MBS, s (cells)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.85</td>
<td>1</td>
<td>1.1</td>
<td>22805</td>
<td>12840</td>
</tr>
<tr>
<td>15</td>
<td>6.48</td>
<td>51136</td>
<td>12863</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>6.1</td>
<td>62562</td>
<td>12820</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>1</td>
<td>7.97</td>
<td>2632</td>
<td>882</td>
</tr>
<tr>
<td>15</td>
<td>5.99</td>
<td>8313</td>
<td>958</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>5.85</td>
<td>10153</td>
<td>955</td>
<td></td>
</tr>
</tbody>
</table>

The analysis process can be performed as follows:

1. Pre-fill the decoder buffer at the selected SCR with the maximum number (N) of video frames, which will fit in the decoder buffer.
2. For the next frame to be transmitted after the pre-fill of the decoder buffer calculate the transmission rate R.

\[
R = \frac{frame\_size(N + 1)}{N \times frame\_period}
\]

(2)

where a frame_period = 1/frames_per_second.

a) If (R <= SCR), set R=SCR.
b) Test if the rate selected will cause buffer overflow, that is, if (R*frame_period) > (decoder_buffer_size – decoder_buffer_occupancy) . If so, overflow will occur and R must be calculated as follows.
3. Every time a frame is removed from the decoder buffer a new transmission rate for the next frame period is calculated according to the above constraints with N replaced with the number of current frames in the decoder buffer.

\[
R = \frac{\text{decoder_buffer_size} - \text{decoder_buffer_occupancy}}{N \times \text{frame_period}}
\]

\[R = \frac{\text{frame_size}(N+i) + \text{delta}}{\text{current_frames} \times \text{frame_period}} \]

where \(\text{delta}\) is the untransmitted data portion from the previous frame period.

If a look ahead of K frame periods is specified, and \(R >= \text{SCR}\), a procedure is called to calculate and return a rate equal to

\[
\text{max} \left( \left( \sum_{i=1}^{k} R_i \right) / j \right)
\]

This rate then will be used for the next frame period unless the decoder buffer is to overflow where the rate then will be set to \(R\) as follows:

\[
R = \frac{\text{decoder_buffer_size} - \text{decoder_buffer_occupancy}}{\text{current_frames} \times \text{frame_period}}
\]

For simplicity, the analysis process of a video clip is performed assuming a constant delay network, however it can easily be extended to yield a traffic contract that incorporates any assumed network jitter.

**III. Transmission of Video**

Using the just-in-time scheduling procedure to transport a video stream would be complicated. However, since from the point of view of an ATM traffic contract, transmitting at the sustained rate does not consume tokens, transmitting above the sustained rate consumes tokens and transmitting below the sustained rate frees tokens, a simple algorithm can be used to transport a video stream once its traffic contract has been determined using the analysis phase. Using the pre-determined traffic contract from the analysis phase and the hypothetical model of the decoder buffer at the source, a video clip can be transported as follows:

- Pre-fill the decoder buffer by transmitting data at the SCR.
- Transmit at the PCR as long as the sustain token bucket has enough tokens to transmit a cell at the PCR and the decoder buffer is not full. For every cell transmitted at the PCR, the sustain token bucket is decremented by \((\text{PCR} - \text{SCR})/\text{PCR}\).
- If not enough tokens exist in the sustain token bucket and the decoder buffer is not full, transmit at the SCR.
- If the decoder buffer is full, stop transmitting until a frame is removed from the decoder buffer and then resume transmission at the PCR or at the SCR. For the zero transmission period \((T_{\text{zero}})\) the sustain token bucket is incremented by \((\text{SCR} \times T_{\text{zero}})\).

**IV. Effective Bandwidth Metrics**

Modeling the benefits of implementing VBR encoding in video transmission systems is quantified using an effective bandwidth (EBW) metrics. Effective bandwidth or equivalent capacity (EC) is defined as the amount of link capacity or transmission bandwidth needed to serve a traffic source with the objective to satisfy a pre-specified cell loss probability (CLP) and delay-variation criteria.

In deriving a specific effective bandwidth formulation, a particular traffic distribution (i.e. bursts or on-off periods), is usually assumed. The shape of this distribution can drastically affect the effective bandwidth value and therefore, understanding the burstiness of video traffic is crucial in selecting and tuning an effective bandwidth metric. For example, the aggregate burstiness for a large number of similar streams is often modeled with a Gaussian distribution [1], which produces significant lower estimates than the exponential distribution, used commonly for more bursty sources.

The most commonly used EBW metrics are those based on the statistical characterization of traffic sources. In Guerin et.al [1], two EBW approaches were discussed. One was based on adopting a two-state fluid-flow model to capture the basic behavior of a data source. The model assumed that the transmission state of each individual source alternate between "idle, or off", zero transmission periods, and “active, or on”, transmission at the source peak rate, period. The approach estimated the upper bound of the needed effective bandwidth \((\lambda_{\text{eff}})\) for a source as

\[
\lambda_{\text{eff}_{\text{Guerin}}} = \frac{C - B_{\text{max}} + \sqrt{(C - B_{\text{max}})^2 + 4 \rho B_{\text{max}} C}}{2C/\text{PCR}}
\]

where,

\[
C = \alpha b (1 - \rho) \text{PCR }, \alpha = \ln ( 1 / \text{CLP})
\]

\(B_{\text{max}}\) = buffer size of the statistical multiplexer.
\(\rho\) = source utilization, \(b\) = mean burst period

The effective bandwidth given by this flow approximation for N multiplexed sources is just the sum
of their individual effective bandwidths. The other approach presented in [1] was based on the Gaussian approximation to model the total traffic. In this technique it is assumed that a large number of sources (N) with similar characteristics (mean rate ($\mu_i$), standard deviation ($\sigma_i$)) and having the same or very similar quality of service requirements are multiplexed together. The effective bandwidth ($\lambda_{\text{eff, Gauss}}$) for N sources was estimated, using the Gaussian approximation, as

$$\lambda_{\text{eff, Gauss}} = \mu + \eta \sigma$$

(7)

where,

$$\eta = \sqrt{-2 \ln(\text{CLP}) - \ln(2\pi)}$$

$$\sigma^2 = \sum_{i=1}^{N} \sigma_i^2, \ \mu = \sum_{i=1}^{N} \mu_i$$

Mark et.al [2] based their method of effective bandwidth estimation on the diffusion process approximation proposed by Kobayashi et.al [3] which depends on the mean and variance of the on-off periods of a fluid source. Mark et.al [2] assumed that the on-off periods of a source are distributed according to Erlang-k distributions. For k=1 the source model is a Markov on-off fluid source and constitutes an upper bound on the effective bandwidth calculation. As k increases the variability in the on-off periods decreases to give a lower estimate of the effective bandwidth. The effective bandwidth estimation for a source was given in [2] as

$$\lambda_{\text{eff, Mark}} = \text{SCR} + \theta \frac{(\text{PCR} - \text{SCR}) \text{MBS}}{\text{PCR}}$$

(8)

where,

$$\theta = \frac{\ln(1/\text{CLP})}{2kB_{\text{max}}}$$

$$\text{MBS} = \frac{(\text{PCR} - \text{SCR}) \times \text{MBS}}{\text{PCR}}$$

V. Using Effective Bandwidth Metrics to Determine Bandwidth Saving of VBR Video

The effectiveness of smoothing VBR video clips to reduce the required network resources while maintaining a certain QoS is assessed using the effective bandwidth concept as a measure. The video sequences were analyzed and transported using the schedule analysis and transport schemes described in the previous section. Table 3 shows the obtained traffic contracts for different VBR MPEG video clips. The CLP used was 10-6 and the traffic contract parameters (SCR, PCR, MBS) where chosen to optimize the bandwidth saving when a look ahead period of one second is applied in the analysis process. The multiplexing buffer sizes were chosen as follow: 5000 cells for MPEG-1, 10000 cells for MPEG-2 and 1000 cells for MPEG-4.

To model the most conservative allocated bandwidth for $\lambda_{\text{eff, Mark}}$ the variability factor k was chosen as one and for $\lambda_{\text{eff, Guerin}}$ the maximum burst size was used instead of the average burst size. By comparing the EBW needed to transport a smoothed VBR video stream to its peak rate as shown in the last two columns of Table 3, depending on the video content, 12%-31% savings in network resources can be achieved when smoothed VBR video is transmitted under the most conservative assumptions.

Since the selection of the SCR changes the MBS value used in the leaky bucket policing mechanism of the ATM call admission control (CAC), finding the right tradeoff between the SCR and MBS values greatly affects the estimation and calculation of the EBW. Table 4 shows the tradeoff between SCR and MBS as SCR is varied between the mean and peak of a video stream using the analysis process with a look ahead of one frame period.

Relaxing the assumed constraints of the EBW formulation used to estimate the needed network resources to transport a VBR video clip will result in more savings. For example, the effect of variability factor k on $\lambda_{\text{eff, Mark}}$ and the multiplex buffer size on $\lambda_{\text{eff, Guerin}}$ are shown in Figure 2 for Action4 MPEG-4 clip. In Figure 2a, when an SCR value of 231 kbps was used, the estimated EBW needed decreased from 81% of the peak rate for a k value of one to 71% for a k value of ten translating into an additional 10% savings. Similar savings can be seen in Figure 2b when $B_{\text{max}}$ is increased.

<table>
<thead>
<tr>
<th>Video Clip</th>
<th>MPEG Type</th>
<th>Peak (Mbps)</th>
<th>Mean (Mbps)</th>
<th>SCR (Mbps)</th>
<th>PCR (Mbps)</th>
<th>MBS (Cells)</th>
<th>$\lambda_{\text{eff, Mark}}$ / Peak</th>
<th>$\lambda_{\text{eff, Guerin}}$ / Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action video 2</td>
<td>2</td>
<td>6.65</td>
<td>3.5</td>
<td>5.19</td>
<td>5.23</td>
<td>13458</td>
<td>.82</td>
<td>.83</td>
</tr>
<tr>
<td>Talking Head</td>
<td>2</td>
<td>4.4</td>
<td>2.6</td>
<td>3.3</td>
<td>3.78</td>
<td>9579</td>
<td>.80</td>
<td>.81</td>
</tr>
<tr>
<td>video2</td>
<td>1</td>
<td>1.62</td>
<td>0.45</td>
<td>1.01</td>
<td>1.39</td>
<td>3215</td>
<td>.71</td>
<td>.69</td>
</tr>
<tr>
<td>talk2</td>
<td>1</td>
<td>1.576</td>
<td>0.45</td>
<td>1.353</td>
<td>1.526</td>
<td>3698</td>
<td>.88</td>
<td>.88</td>
</tr>
<tr>
<td>Action4</td>
<td>4</td>
<td>0.329</td>
<td>0.0835</td>
<td>0.231</td>
<td>0.305</td>
<td>843</td>
<td>.81</td>
<td>.82</td>
</tr>
<tr>
<td>Talk4</td>
<td>4</td>
<td>0.302</td>
<td>0.123</td>
<td>0.249</td>
<td>0.292</td>
<td>447</td>
<td>.85</td>
<td>.86</td>
</tr>
</tbody>
</table>

To study the effect of statistically multiplexing a number of video streams, the Gaussian effective bandwidth calculation can be used. Figure 3 shows the
effect of multiplexing a number of video sources on the needed EBW for different chosen SCR values. For simplicity N identical video streams, Action4, were used for the calculation. From Figure 3, it is noted that the Gaussian approximation tends to be optimistic in estimating the required bandwidth when N is large or when the value of the SCR approaches the mean.

Table 4: Trade off between SCR and MBS

<table>
<thead>
<tr>
<th>Video Clip</th>
<th>Peak (Mbps)</th>
<th>Mean (Mbps)</th>
<th>MBS (cells) given SCR=(1-X)∗Mean+X∗Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action2</td>
<td>6.65</td>
<td>3.5</td>
<td>X=0  X=.2  X=.4  X=.5  X=.6  X=.7</td>
</tr>
<tr>
<td>Talking</td>
<td>4.4</td>
<td>2.6</td>
<td>38309 10828 24669 8063 1688 1</td>
</tr>
<tr>
<td>video2</td>
<td>1.62</td>
<td>0.45</td>
<td>36556 4772 2584 1519 965 489</td>
</tr>
<tr>
<td>talk2</td>
<td>1.576</td>
<td>0.45</td>
<td>195161 22977 15956 11956 7438 3296</td>
</tr>
<tr>
<td>Action4</td>
<td>0.329</td>
<td>0.0835</td>
<td>26177 4415 1073 1 1 1</td>
</tr>
<tr>
<td>Talk4</td>
<td>0.302</td>
<td>0.123</td>
<td>28839 2097 1295 430 1 1</td>
</tr>
</tbody>
</table>

VI. Conclusions

In this paper we have discussed the applications, benefits, and challenges of using variable bitrate (VBR) MPEG video encoding. Using an analysis method such as just-in-time scheduling, a traffic contract can be found for a particular video stream which guarantees VBV buffer compliance. This traffic contract and an effective bandwidth metric can then be used to measure the potential savings. As was shown in Table 3, 12% to 31% bandwidth savings were achieved based on conservative effective bandwidth metrics, an additional 5% to 10% savings are possible if less conservative assumptions about the video traffic are made as was shown in Figure 2.

References


The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official pollicies, either expressed or implied, of the Army Research Laboratory or the U.S. government.