GUIDELINES FOR SELECTING PRACTICAL MPEG GROUP OF PICTURES

Huahui Wu, Mark Claypool, and Robert Kinicki
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
{flashine,claypool,rek}@cs.wpi.edu

ABSTRACT
The repeated pattern of I, P and B frames in an MPEG stream is known as the Group of Pictures (GOP). Current GOP choices are made using intuition and informal guidelines without the support of theoretical or practical evidence. This paper studies the impact of the choice of GOP by evaluating the effects of GOP on both static MPEG videos and MPEG videos streaming over a lossy network. The static analysis involves encoding raw video images into MPEG files with various GOP patterns to compare and contrast static properties such as the frame size, file size and quality. The streaming analysis varies the GOP length and pattern to study the impact of GOP on a model of the streaming bitrate and playable frame rate. The results consistently suggest two guidelines: 1) the number of B frames between two reference frames should be close to 2, except when limited to less than 2 by time constraints; 2) the number of P frames should be 5 or fewer as there is little performance gain in setting the number of P frames in the GOP larger than 5.

KEY WORDS
MPEG, GOP, Forward Error Correction, Temporal Scaling

1 Introduction

In order to support both inter- and intra-frame compression within the MPEG stream of frames, MPEG uses three frame types: I, P and B frames. Typically, an MPEG-encoded video flow consists of a repeated pattern of I, P, and B frames, known as the Group of Pictures (GOP). The GOP choice combined with MPEG compression techniques determine inherent properties of the encoded MPEG file such as size of the three frame types, overall file size and image quality. GOP and compression also determine the MPEG streaming performance in terms of streaming bitrate and perceived quality.

Currently the choice of GOP is mostly an intuitive process. Some researchers use the default GOP pattern that comes with an MPEG encoder. Other researchers have varied the GOP pattern with little concern for the practical ramifications of the specific GOP pattern on delivery of an MPEG video over a lossy network. In [6], the author searches a large range of GOPs to find the optimal GOP for MPEG streaming, which can result in a large number of P frames in one GOP (e.g., 35 P frames). Such a large GOP is seldom seen in real MPEG encoding [3]. In [2], the authors find the number of B frames between two reference frames should be from 1 to 4 while [9] concludes that the number should be varied from 0 to 2. However, the advantage of these proposed dynamic GOP length mechanisms is not significant. To the best of our knowledge, guidelines on how to *practically* choose a GOP has not been presented in any systematic fashion.

The goal of this paper is to investigate practical GOP considerations with respect to performance of MPEG encoded video streams, using a a network model with packet loss and capacity constraints. This research consists of two main components – the study of static MPEG video and analysis of streaming MPEG video. In the static MPEG analysis, the GOP length and pattern are varied to observe the properties of the resultant MPEG file, noting file size, frame sizes and video quality (measured by Peak Signal-to-Noise Ratio, PSNR). In the streaming MPEG analysis, the GOP is varied to provide insight on the impact of these practical GOP choices on the behavior of the streaming MPEG with Forward Error Correction (FEC) [8] and Pre-Encoding Temporal Scaling (PETS) in terms of bitrate and video quality (measured by playable frame rate). The two major recommendations from both components of this study are: 1) the number of B frames between two reference frames should be set to two when the video stream does not have severe delay constraints, and 2) the number of P frames should be 5 or fewer as there is little performance gain in setting the number of P frames in the GOP larger than 5.

This paper is organized as follows: Section 2 studies static MPEG; and Section 3 analyzes the behavior of MPEG streaming; and Section 4 summarizes the paper's contributions and recommendations.

2 Static MPEG Files

2.1 Methodology

This section considers the impact of GOP length on static MPEG file properties and suggests guidelines for GOP considerations. The analysis uses the following steps:

1. Study the impact of the number of B frames (denoted as N_{BP}) between two reference (P or I) frames on

frame size and frame quality (measured by PSNR). This provides a guideline for choosing N_{BP}.

2. Given the N_{BP} guideline, study the impact of the number of P frames in one GOP (denoted as N_P) on frame sizes and frame quality (measured by PSNR). This provides a guideline for choosing N_P.

Motion	Video	Description
Low	Container(CT)	A working container ship
Low	Hall(HL)	A hallway
Low	News(NW)	Two news reporters
Medium	Foreman(FM)	A talking foreman
Medium	Paris(PR)	Two people talking with high-motion gestures
Medium	Silent(SL)	A person demonstrating sign language
High	Coastguard(CG)	Panning of a moving coastguard cutter
High	Mobile(ML)	Panning of moving toys
High	Vectra(VT)	Panning of a moving car

Table 1. Video Clips

Nine video clips are used for the experiments, where each video clip has 300 raw images that play out at 30 fps for 10 seconds. The size of each frame is 352x288 pixels (CIF). For each video clip, Table 1 provides an approximate motion classification, an identifying name with an abbreviation code in parentheses, and a short description of the video content. The abbreviations identify the clips in subsequent graphs. All the experiments use the Berkeley MPEG encoder and decoder[1]. However, the results should hold for other MPEG encoders since the choice of encoder has little impact on compression relative to the impact on compression due to the choice of quantization level and GOP pattern. The quantization values for I, P and B frames are all 3 to yield a high picture quality in every frame.

2.2 Study of N_{BP}

Increasing the number of B frames decreases the correlation between the B frames and the frames they reference [4]. Although the exact tradeoff depends upon the nature of the video scene, for a large class of videos a reasonable spacing of references frames is every 1/10th second. This results in a frame pattern of 'IBBPBBPBB...IBBPBBPBB...' and more generally implies that N_{BP} commonly has a value of no more than two. Mayer-Patel [6] used the frame rate of 30 fps and a minimum ratio of reference frames to all frames of 1/3, which also implies N_{BP} is less than three. Feng et al. [3] extracted video data from DVDs and also found the most common value of N_{BP} is no more than two.

[1] http://bmrc.berkeley.edu/frame/research/mpeg/

Experiments were conducted by encoding raw images into MPEG videos with different values of N_{BP} and checking the impact on file size (in Mbytes), frame sizes (in Kbytes) and the quality (measured by PSNR, in decibels).

N_P=1	Frame Size (KB)		PSNR (dB)		File Size
N_{BP}	S_P	S_B	Q_P	Q_B	(MB)
0	11.97	N/A	41.1	N/A	5.18
1	14.22	7.65	41.1	36.7	3.87
2	15.22	8.66	41.1	34.7	3.57
3	16.14	9.46	41.1	33.8	3.53
5	17.36	10.60	41.1	32.7	3.57
11	19.89	12.84	41.1	30.9	3.97

a. N_P=1

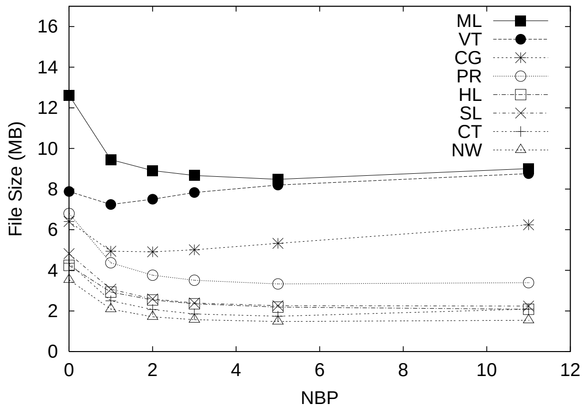
N_P=4	Frame Size (KB)		PSNR (dB)		File Size
N_{BP}	S_P	S_B	Q_P	Q_B	(MB)
0	12.05	N/A	41.0	N/A	4.19
1	14.17	7.57	41.0	36.6	3.45
2	15.31	8.60	41.1	34.7	3.33
3	15.93	9.42	41.1	33.9	3.35
5	17.35	10.56	41.1	32.6	3.48
11	19.17	12.81	41.1	30.9	3.93

b. N_P=4

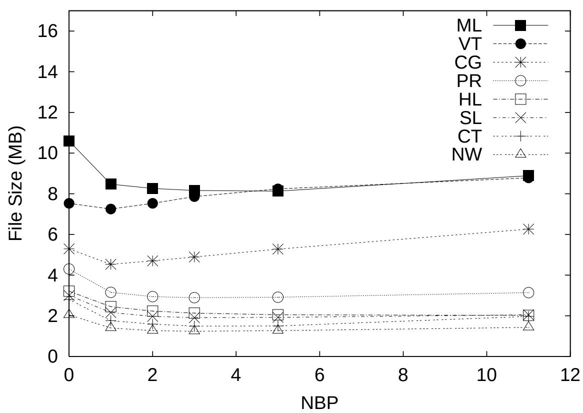
Table 2. Impact of N_{BP} on MPEG files for *Foreman*

Table 2 depicts the frame sizes and PSNR of the *Foreman* video for different N_{BP} sizes with a fixed number of P frames ($N_P = 1$ in Table 2.a and $N_P = 4$ in Table 2.b). Information on the I frames is not provided since they are intra-compressed only and do not change with GOP pattern. The data in the two tables are very similar. This suggests that the impact of N_P is small (the next Section, Section 2.3, explores N_P in more detail). As N_{BP} increases, the quality of the B frames decreases quickly. For example, in Table 2.a, the PSNR of the B frames drops dramatically from 36.7 dB to 30.9 dB. Notice that when N_{BP} increases, the sizes of the P and B frames also increase. In both tables, the sizes of the B frames nearly double as N_{BP} goes from 1 to 11 and this also causes the MPEG file size to grow when N_{BP} is above 2. In theory, having more B frames can reduce the MPEG file size since B frames are usually smaller than I frames. However, since the average size of a B frame increases when there are more B frames, the MPEG file does not necessarily have a higher compression rate for larger numbers of B frames. In fact, note that the size of the MPEG file is the lowest when $N_{BP} = 2$. These facts suggest that although B frames have the highest compression rate, a large number of B frames in a GOP introduces low inter-frame compression and lower quality. Thus a guidelines is to have N_{BP} close to or equal to two.

Similar experiments were conducted with the other eight videos in Table 1. Figure 1 shows the impact of N_{BP} on encoded MPEG file size ($N_P = 1$ in Figure 1.a and $N_P = 4$ in Figure 1.b). In the figures, the x-axis is N_{BP} and the y-axis is the encoded file size in Mbytes. The figures show $N_{BP} = 2$ provides a small file, very close to the



a. $N_P = 1$



b. $N_P = 4$

Figure 1. Impact of N_{BP} on MPEG files for the other videos

minimum size, for all videos. This result does support previous research [2, 9] which discussed that content-based dynamic GOP length can increase MPEG performance. However, the graphs imply the performance improvement is not significant when more B frames are added to the GOP. The PSNR data is not presented for these videos because the results in all cases are very similar to those in Table 2 in that the PSNR of the B frames drops dramatically by around 5dB for N_{BP} of three or larger. These results clearly suggest a practical GOP guideline of keeping N_{BP} close to two.

Another practical constraint for N_{BP} is that for streaming MPEG, B frames can not be decoded until after the arrival of the subsequent I or P frame. This implies latency increases linearly with the number of B frames. For interactive applications, such as a videoconference, the added latency contributes to the end-to-end delay. For typical full-motion streaming (30 fps frame rate), each B frame contributes about 33 ms of delay. In studies of streaming video on the Internet [1] and network delays in general [5], the median round-trip times for a variety of network configurations are around 100 ms. Thus, compared to the round-trip time, one or possibly two B frames may not represent a significant increase the end-to-end delay, while the use of three B frames could double the end-to-end delay. Thus, a GOP guideline for streaming MPEG is to have N_{BP} as high as the latency tolerates, but no more than 2.

In summary, the number of B frames between two reference frames should be less then or equal to two. This guideline is used in informing all subsequent experiments.

2.3 Study of N_P

Similar to section 2.2, experiments were run by encoding the raw *Foreman* images into MPEG videos with different N_P values and analyzing the impact on file size, frame sizes and PSNR.

N_{BP}=2	Frame Size(KB)		PSNR (dB)		File Size
N_P	S_P	S_B	Q_P	Q_B	(MB)
0	N/A	8.83	N/A	34.8	4.02
1	15.22	8.66	41.1	34.7	3.57
5	15.31	8.60	41.1	34.7	3.30
9	15.30	8.59	41.0	34.7	3.25
14	15.17	8.60	41.0	34.7	3.23
29	15.22	8.60	41.0	34.7	3.20

Table 3. Impact of N_P on MPEG files for *Foreman*

Table 3 presents frame sizes of the *Foreman* video clip for different values of N_P ($N_{BP} = 2$). These results show N_P increases, the sizes of the P and B frames do not significantly change, nor does the frame quality. Since increasing the GOP length does not impact the frame size and typical P frames are smaller than their referenced I frames, more P frames can reduce the MPEG file size, as shown in the last column of table 3. However, the reduction in file size is not significant.

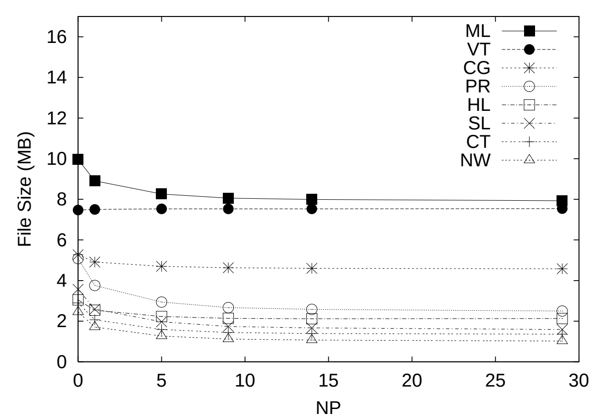


Figure 2. Impact of N_P on MPEG files for the other videos

Similar experiments were conducted with the other eight videos in Table 1. Figure 2 presents the impact of

N_{BP} on encoded MPEG file size ($N_{BP} = 2$). In the figure, the x-axis is N_P and the y-axis is the encoded file size in Mbytes. More P frames can reduce the MPEG file size, but the reduction is not significant after $N_P = 5$. The corresponding PSNR data is presented, but the results are very similar to Table 3, with the frame quality changing little with increases in N_P.

Another practical constraint associated with the number of P frames is the need to support VCR-like functions (pause, rewind, fast-forward, etc.). Since response to these functions require access to the I frames, this suggests GOP length should not be long. For example, if a user wants to pause a movie with a precision of 3 seconds, the GOP length should be no more than 90, and therefore the number of P frames should be at most 90, and more likely at most 30 if N_{BP} is 2.

As a summary, while there are no specific constraints concerning the number of P frames, as a guideline, the number of P frames should be no more than 30. Moreover, while having more P frames can improve the compression ratio, the benefit is not significant compared to the compression ratio obtained with five P frames per GOP. This guideline is used in informing all subsequent experiments and analysis.

3 Streaming MPEG

3.1 Methodology

To protect streaming MPEG from packet loss, Forward Error Correction (FEC) is often used to recover video data by adding redundancy. At the application layer, if the video frames are transmitted in K packets, then FEC consists of adding $(N - K)$ redundant packets to the K original packets and sending the N packets as the frame. If any K or more packets are successfully received, the frame can be completely reconstructed [8].

To adjust the streaming bitrate to the available bitrate, streaming systems use media scaling, and often *temporal scaling* where carefully selected video frames are discarded at the sender before transmission. In this section, one form of temporal scaling is introduced: Pre-Encoding Temporal Scaling (PETS).[2]

This section studies the impact of the GOP pattern on MPEG streaming under conditions of packet loss and limited capacity. Using the guidelines obtained in the static MPEG analysis, the streaming analysis uses the following steps:

1. Develop a model for streaming MPEG with Forward Error Correction (FEC) and Pre-Encoding Temporal Scaling (PETS) to estimate the video quality (measured by playable frame rate).
2. Use the model in conjunction with a model of network packet loss and capacity limit to study the impact of GOP length on streaming performance.

The system parameters and variables used in the model are provided in alphabetic order:

- δ: the gap distance between two neighboring encoded images.
- G: the GOP rate, or the number of GOPs sent each second for an MPEG stream.
- p: the packet loss probability used in the model.
- R_F: the maximum playable frame rate achieved when there is enough available capacity and no packet loss (typical full-motion video rates have $R_F = 30fps$).
- S_I, S_P, S_B: the size of an I, P or B frame respectively, in fixed-size packets.
- S_{IF}, S_{PF}, S_{BF}: the number of FEC packets added to each I, P or B frame, respectively.
- T: the modeled capacity constraint.

3.2 Pre-Encoding Temporal Scaling (PETS)

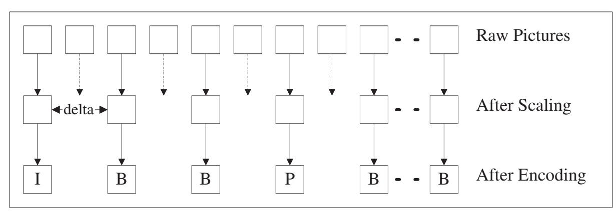


Figure 3. Pre-Encoding Temporal Scaling (PETS)

As depicted in Figure 3, Pre-Encoding Temporal Scaling reduces bitrates by discarding some of the raw pictures before encoding and compressing the remaining pictures into MPEG frames. The more raw pictures that are discarded, the lower the bitrate, but the lower the video quality. Note, with PETS the GOP pattern does not change with the amount of scaling, but the effectiveness of compression for P and B frames may decrease as their distance from their original I frame reference increases because of discarded frames.

To measure the discarding rate, the variable δ is defined as the gap distance between two neighboring encoded images. For example, in Figure 3, where every other image is discarded, δ is one since the gap distance between two neighbor encoded pictures is one.

Since the GOP pattern in PETS is never altered, the GOP rate needs to be adjusted to keep the playout rate at the receiver side the same as the original video to preserve

[2] Temporal scaling can also be done after encoding, otherwise known as POst encoded Temporal Scaling (POTS), but is not presented here.

the real-time aspects. For example, if every other image is discarded, the GOP rate needs to be reduced to half of that of the original GOP rate. Knowing the original full-motion frame rate is R_F and the GOP length is $1+N_P+N_B$, only $R_F/(1+\delta)$ of the raw pictures will be encoded into MPEG frames, so the GOP rate, as a function of δ, is:

$$G(\delta) = \frac{R_F/(1+\delta)}{(1+N_P+N_B)} \tag{1}$$

Notice, also, that when the raw pictures are discarded before encoding, the similarities among the encoded pictures decreases and, hence, the sizes of P and B frames increases. At the extreme, when δ is large (say, Δ), the P and B frames effectively become the same as I frames. Assuming the frame sizes increase linearly with increasing δ, one can determine the sizes of P and B frames as functions:

$$\begin{array}{l} S_P(\delta) = S_{P0} + (\delta/\Delta) \cdot (S_I - S_{P0}) \\ S_B(\delta) = S_{B0} + (\delta/\Delta) \cdot (S_I - S_{B0}) \end{array} \tag{2}$$

where S_{P0} and S_{B0} are the sizes of the P and B frames, respectively, in the MPEG video without PETS. Experiments (not shown here) show curves up to $\Delta = 9$ fit Equation 2 well. Notice that the sizes of the I frames do not change with δ since I frames use intra-image compression only.

3.3 Model of Playable Frame Rate

When PETS and FEC are used in streaming MPEG, the model for playable frame rate is similar to the model presented in our previous work [7]. However, two changes to the model are required. First, the frame size is no longer fixed but instead is a function of the scaling level δ, as in Equation 2. Second, since with PETS, some of the images may be discarded before encoding, the GOP rate must be decreased as in Equation 1 to keep real-time playout. After these changes, the model can be used to estimate the playable frame rate.

For given values of p, (N_P, N_B) and (S_I, S_{P0}, S_{B0}), the total playable frame rate R varies with the temporal scaling level and the amount of FEC as a function R(δ, (S_{IF}, S_{PF}, S_{BF})). However, the streaming bitrate is limited by the capacity constraint. This extended model can be used to optimize the playable frame rate, R:

$$\left\{ \begin{array}{l} Maximize: \\ \quad R = R(\delta, (S_{IF}, S_{PF}, S_{BF})) \\ Subject\ to: \\ \quad G(\delta) \cdot ((S_I(\delta) + S_{IF}) + N_P \cdot (S_P(\delta) + S_{PF}) \\ \qquad + N_B \cdot (S_B(\delta) + S_{BF})) \leq T \end{array} \right. \tag{3}$$

Unfortunately, finding a closed-form solution for the non-linear function R is difficult since there are many saddle points. However, given that the optimization problem is expressed in terms of integer variables over a restricted domain, an exhaustive search of the discrete space is feasible. With fixed input values for (p, T), (N_P, N_B) and (S_I, S_{P0}, S_{B0}), the space of possible values for δ and (S_{IF}, S_{PF}, S_{BF}) can be searched to determine the temporal scaling level and FEC pattern that yield the maximum playable frame rate under the capacity constraint.

3.4 Analysis

The GOP pattern is varied with different values of N_P and N_{BP} used to encode the MPEG stream. For each stream, the frame sizes is extracted and fed into our model (Equation 3) to find the optimal playable frame rate. By comparing the playable frame rates of different streams, the impact of the GOP pattern on streaming MPEG is analyzed.

Three different FEC choices are considered:

- Non-FEC: The sender adds no FEC to the video.
- 5% Fixed FEC: The sender protects each frame with FEC the size of 5% of the original frame size.
- Adjusted FEC: Before transmitting, the sender uses our model (Equation 3) to determine the FEC pattern and temporal scaling level that produce the maximum playable frame rate and uses these for the entire video transmission.

In all cases, the bitrates used by the MPEG video with added FEC are scaled by PETS to meet the capacity limits.

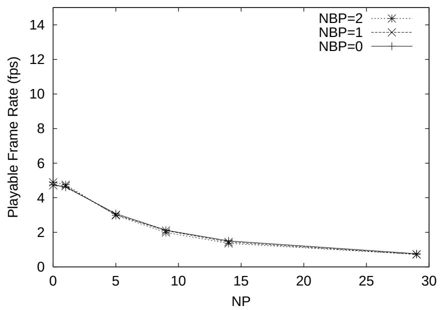
Figures 4 show performance results for a set of experiments with a 1.5 Mbps capacity constraint and with 2% induced modeled packet loss for the video *Foreman*. Fixed FEC is more effective than non-FEC when there is considerable loss since it repairs the loss, preventing degradation in the video quality. In all cases, the mechanism for adjusting FEC searches the space of choices for the best value of FEC and thus yields the best quality.

More importantly for the focus of this paper, the impact of GOP on streaming MPEG, these figures show results similar to those in the previous sections. All three graphs demonstrate that larger values of N_{BP} yield better quality (although delay constraints for interactive applications still limit N_{BP} to be no larger than 2) and there is little to be gained by having N_P greater than 5.

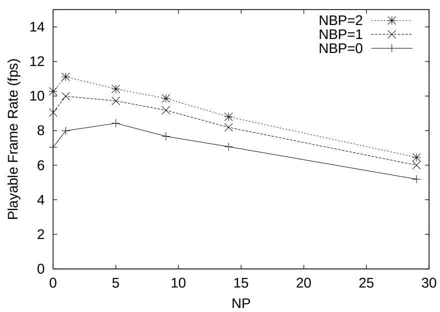
Figures 5 depicts the impact of N_P ($N_{BP} = 2$) on streaming MPEG with adjusted FEC for the other 8 videos, where the network model has a 1.5 Mbps capacity constraint and a 2% packet loss is modeled. These results also suggest there is little to be gained by having $N_P > 5$.

4 Conclusion

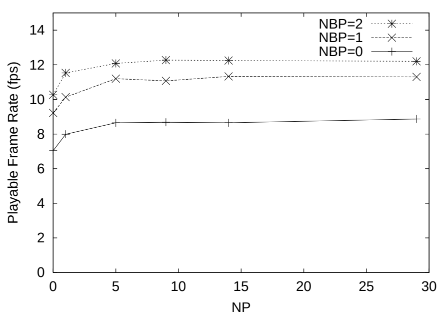
This paper presents an organized methodology to better understand the practical impact of both the GOP length and the detailed GOP pattern on static and streaming MPEG. Utilizing results from experiments and analytic modeling, practical guidelines are put forth for setting the GOP length



a. Non-FEC



b. 5% Fixed FEC



c. Adjusted FEC

Figure 4. Streaming *Foreman* with FEC and PETS. Network model has 2% loss and 1.5 Mbps capacity constraint

and selecting an appropriate GOP pattern over a range of MPEG conditions.

In the first set of experiments raw video images were encoded to MPEG files. These results suggest two guidelines: 1) The number of B frames between two reference frames should not exceed 2; and 2) while there were no specific limitations to the number of P frames in a GOP pattern, there should be no more than 30 P frames in the GOP pattern to support VCR-like functions.

The second phase of our investigation considered the GOP impact when MPEG was sent over a lossy network with Forward Error Correction, which protects packet loss, and Pre-Encoding Temporal Scaling, which satisfies capacity constraint. The optimal MPEG quality always occurs when $N_{BP} = 2$ and $N_P \leq 5$. The results suggest two guidelines: 1) The number of B frames between two reference frames should be kept at 2 except when constrained

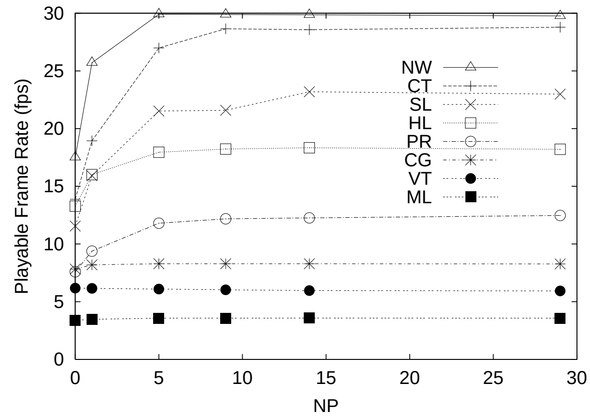


Figure 5. Streaming the other 8 videos with adjusted FEC and PETS, 2% loss and 1.5 Mbps capacity constraint ($N_{BP} = 2$).

lower by delay constraints; and 2) the number of P frames need not be more than 5.

References

[1] J. Chung, M. Claypool, and Y. Zhu. Measurement of the Congestion Responsiveness of RealPlayer Streaming Video Over UDP. In *Proceedings of the Packet Video Workshop (PV)*, Nantes, France, Apr. 2003.

[2] A. Dumitras and B. G. Haskell. I/P/B frame type decision by collinearity of displacements. In *Proceedings of ICIP 2004*, Singapore, Oct. 2004.

[3] W.-C. Feng, J. Choi, W.-C. Feng, and J. Walpole. Under the Plastic: A Quantitative Look at DVD Video Encoding and Its Impact on Video Modeling. In *Proceedings of the Packet Video Workshop (PV)*, Nantes, France, Apr. 2003.

[4] D. L. Gall. MPEG: A Video Compression Standard for Multimedia Applications. *Communications of the ACM*, 34(4):46 – 58, 1991.

[5] S. Jaiswal, G. Iannaccone, C. Diot, J. Kurose, and D. Towsley. Inferring TCP Connection Characteristics Through Passive Measurements. In *Proceedings of IEEE Infocom*, Hong Kong, Mar. 2004.

[6] K. Mayer-Patel, L. Le, and G. Carle. An MPEG Performance Model and Its Application To Adaptive Forward Error Correction. In *Proceedings of ACM Multimedia*, December 2002.

[7] H. Wu, M. Claypool, and R. Kinicki. A Model for MPEG with Forward Error Correction and TCP-Friendly Bandwidth. In *Proceedings of Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, Monterey, CA, USA, June 2003.

[8] H. Wu, M. Claypool, and R. Kinicki. Guidelines for Selecting Practical MPEG Group of Pictures. Technical Report WPI-CS-TR-05-18, CS Department, Worcester Polytechnic Institute, Aug. 2005.

[9] Y. Yokoyama. Adaptive GOP structure selection for real-time MPEG-2 video encoding. In *Proceedings of ICIP 2000*, Vancouver, Canada, Sept. 2000.