Guidelines for Selecting Practical MPEG Group of Pictures

ABSTRACT

MPEG is one of the most popular open standards for video on the Internet. MPEG uses intra-frame and inter-frame compression with three types of frames: I, P and B frames. The repeated pattern of I, P and B frames in an MPEG stream is known as the Group of Pictures (GOP). The choice of GOP affects static MPEG properties such as frame size and file size and also impacts the streaming MPEG in terms of network bitrate and playable quality. 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 on 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 MPEG streaming analysis considers 3 distinct cases over a network model with packet loss: normal streaming with no repair and no capacity limit; streaming with Forward Error Correction (FEC) but no capacity limit; and streaming with FEC and a capacity limit. 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 the encoding and 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.

1. INTRODUCTION

Accessing digital video over the Internet continues to grow in popularity. Network video products use video compression techniques to support delivery of digital video from video servers over the Internet to the home. Although there are many video compression standards, MPEG[7] has emerged as one of the most popular international standards for motion picture compression.

MPEG was originally designed to encode moving pictures and the associated audio for digital storage media on CD-ROM [11] with transmission rate targets at 1.5 Mbps. Over time, MPEG capabilities have been increased to support higher bitrates and larger picture sizes as popular digital video applications moved beyond simple storage devices to streaming multimedia applications over the Internet. The most recent expectation is for MPEG to support digital video over lossy wireless networks.

MPEG uses three types of frames (I, P and B frames) to implement different compression methods and support inter-frame dependences within the MPEG packet stream. 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 specifies the specific number and pattern of I, P and B frames. Hence, GOP choice combined with MPEG compression techniques determine inherent properties of the encoded MPEG file such as size of the three frame types, file size and image quality. GOP and compression also determines the MPEG streaming performance in terms of streaming bitrate and perceived user 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 [10] 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 [6]. In [5], the authors find the number of B frames between two reference frames could be from 1 to 4 and [16] and conclude 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 streamed over 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 MPEG stream 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.

The contributions of this paper are three-fold. First, the impact of GOP on both static MPEG and streaming MPEG is investigated. Second, an analytical model is derived and extended to handle a variety of network scenarios and treatments. The extended model is used to provide a determination of playable frame rate that can be used as an estimate of user perceptual quality. Third, the concept of Pre-Encoding Temporal Scaling is introduced and briefly evaluated as a viable method of media scaling to be used in tandem with forward error correction (FEC) for media repair.

This paper is organized as follows: Section 2 briefly presents background information on MPEG, MPEG streaming, and FEC. Section 3 presents the methodology and system settings. Section 4 studies static MPEG and Sections 5-7 analyze the behaviors of MPEG streaming with GOP choices under three distinct situations; and Section 8 summarizes the paper's contributions and recommendations.

2. BACKGROUND

2.1 MPEG and Group of Pictures

MPEG has three types of frames: I (intra-coded) frames, P (predictive-coded) frames and B (bi-directionally predictive-coded) frames. I and P frames are also called as *reference frames*. MPEG-encoded video typically repeats the pattern of I, P, and B frames (known as a Group of Pictures or GOP) for the duration of an individual video stream. Figure 1 shows a sample GOP, where the second I frame in the figure marks the beginning of the next GOP. The arrows indicate frame dependency relationships which show that I frames are more important than P frames, and P frames are more important than B frames. To limit the cascading effect that video transmission errors create due to frame dependencies, new I frames must be sent periodically. However, the increased frequency of I frames must be traded off against the higher compression rates afforded by the P and B frames.

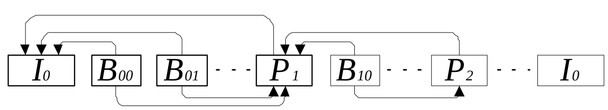


Figure 1: An MPEG Group Of Pictures Example

Let N_P represent the number of P frames in a GOP and N_{BP} represent the number of B frames in between an I and a P frame or two P frames[1]. Using these two terms, a specific GOP pattern can be identified uniquely by G(N_P,N_{BP}). For example, GOP(2,2) signifies the GOP pattern 'IBBPBBPBB'. Let N_B represent the number of B frames in a GOP and N_G represent the length of the GOP. Then $N_B = (1+N_P)\times N_{BP}$ and $N_G = 1 + N_P + N_B$.

As in Figure 1, for the remainder of the paper, subscripts will be used to identify individual frames within a GOP. The single I frame of a GOP is referred to as I_0, while P frames are indexed as P_i, where $1 \leq i \leq N_P$, and B frames are expressed as B_{ij}, where $0 \leq i \leq N_P$ and $0 \leq j < N_{BP}$. For example, P_i is the (i)th P frame, and B_{ij} is the $j+1$th B frame in the $i+1$th interval of I and P frames.

[1] As in most MPEG videos, B frames are assumed to be distributed evenly in the intervals between I and P frames.

2.2 Streaming MPEG

MPEG was originally designed for CD-ROMs and initially supported a data rate of 1.5 Mbps. On the current Internet, which provides high capacity to the end user (over 1 Mbps for home users with broadband and around 10 Mbps or more for universities and corporations), MPEG is also used as a video streaming standard due to its quality and compression rates. Prior to studying MPEG behavior over a network, it is worthwhile to review how MPEG flow characteristics differ from those of more common network traffic such as Web flows or bulk file downloads.

Unlike conventional flows, MPEG flows can tolerate some loss but are sensitive to delay and jitter. While one missing bit can make an executable file useless, the loss of a couple of video frames may not be perceivable to the user. Users will wait many seconds for a Web page to download but they will not watch a movie which plays fast for a few seconds then stops or plays slow for the next few seconds. TCP, the de facto standard transmission protocol, uses retransmission to recover packet loss but introduces delay to the application. TCP also uses an Additive Increase Multiplicative Decrease mechanism to frequently adjust its bitrate in response to congestion and thereby does not provide the smooth bitrate which is preferred by MPEG flows. So UDP is often used in MPEG streaming rather than TCP, especially for interactive applications.

However, an MPEG object is typically large and, consequentially, needs a high bitrate for streaming. For example, the size of a 1-minute movie with moderate visual quality and 320x240 resolution, is around 9 MBytes. When streamed in real time, this video needs requires a network capacity of over 1 Mbps. A couple of such flows can easily saturate the network and, if unresponsive to congestion, create congestion collapse. As an unresponsive protocol, UDP does not reduce its data rate when an Internet router drops packet to indicate congestion. To support other types of flows and avoid congestion collapse, there is a growing belief that Internet applications must be TCP-Friendly [9]. This implies limiting the flow's bitrate to that of an equivalent TCP flow. Moreover a streaming video flow may also be constrained due to the client's final network connection.

While some loss for an MPEG stream can be tolerated, too much data loss will cause unacceptable video quality for the user. The random dropping of packets by routers can seriously degrade video quality [3]. Since retransmission has the potential problem of high delay, especially for network connections with large round-trip times, streaming MPEG can use FEC (Forward Error Correction) to repair the video after packet loss [2, 12]. The impact of a packet loss varies with frame type due to the different importance of I, P and B frames in MPEG. A packet loss in an I frame makes the whole GOP useless while a packet loss in B frame damages only that B frame. Thus, it is more desirable to add different levels of FEC to each type of frame.

Conversely, if a streaming video is to operate within network capacity limits, additional FEC data reduces the effective transmission rate of the original video content. To preserve real-time streaming media playout, an MPEG server must scale back its streaming data rate to match the constrained data rate using *Media Scaling* [1, 14]. With temporal scaling, a widely used form of media scaling, the MPEG server discards some frames before transmission. Increasing the scaling level reduces the number of MPEG frames which in turn saves capacity for the FEC. Hence, selecting the optimal amount of FEC and the optimal scaling level in a network environment can be cast as a constrained optimization problem that optimizes video stream quality.

Previous work [15] derived an analytic model to characterize the performance of scaled MPEG video with Forward Error Correction in the presence of packet loss. Given network loss and MPEG frame types and sizes, the model allows specification of temporal scaling level and number of FEC packets for each type of MPEG frame and computes the total playable frame rate. Then, using the model, an optimization algorithm exhaustively searches all possible combinations of FEC and temporal scaling to find the configuration that yields the best video quality under the capacity constraint. The computation required by the search can be done in real-time, making the determination of optimal choices for adaptive FEC feasible for most streaming multimedia connections. The model is employed in Sections 5 to 7 and the optimization algorithm under a bandwidth constraint is used in Section 7.

2.3 Forward Error Correction (FEC)

As a low latency repair approach, Forward Error Correction is often used to recover from packet loss by adding redundancy. Since streaming video frames are often larger than a single Internet packet and Internet congestion results in lost packets, FEC is often applied at the packet level. Thus, an application level models video frames as being transmitted in K packets where K varies with frame type, encoding method, and media content. Media independent FEC [13] then 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.

To analyze the success rate of FEC on application layer frames, the sending of packets are modeled as a series of Bernoulli trials. Thus, the probability $q(N, K, p)$ that a K-packet data frame is successfully transmitted with $N - K$ redundant FEC packets in a lossy network with packet loss probability p is:

$$q(N,K,p) = \sum_{i=K}^{N} \left[\binom{N}{i} (1-p)^i * p^{N-i} \right] \quad (1)$$

Note, when FEC is not used, this equation reduces to:

$$q(K,K,p) = (1-p)^K \quad (2)$$

3. METHODOLOGY AND SYSTEM SETTINGS

This investigation has two main components - the study of static MPEG (as recorded to a file) and the study of streaming MPEG. The static MPEG analysis varies the GOP length and observe the properties of the resultant MPEG file, noting file size, frame sizes and video quality (measured by PSNR). The streaming MPEG analysis varies the GOP length and evaluates the dynamic behavior of streaming MPEG over a model of a lossy network, in terms of bitrate and video quality (estimated by playable frame rate).

The static MPEG analysis proceeds using the following steps:

1. Study the impact of N_{BP} on frame size and frame quality (measured by PSNR) to provide a guideline for choosing N_{BP}.

2. Given the N_{BP} guideline, study the impact of N_P on frame sizes and frame quality (measured by PSNR) to provide a guideline for choosing N_P.

Using the guidelines obtained in the static MPEG analysis, the streaming MPEG analysis uses the following steps:

1. Develop a model for streaming MPEG which can analytically estimate the video quality (measured by playable frame rate). Then, use the model in a network that models packet loss to study the impact of GOP length on streaming performance.

2. Extend the model by adding fixed FEC to the MPEG stream and study the impact of GOP length on bitrate and playable frame rate.

3. Incorporate a capacity constraint into the model, including adjusting the streaming bitrate using temporal scaling. Then, study the results using this model.

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 around a table 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[2] on Linux. 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 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.

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

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

- G: the GOP rate, or the number of GOPs sent each second for an MPEG stream.
- N_{BP}: the number of B frames between two reference (P or I) frames.
- N_P, N_B: the number of P or B frames, respectively, in one GOP.
- N_G: the length of a GOP, which is $1{+}N_P{+}N_B$.
- p: the packet loss probability used in the model.
- q_I, q_P, q_B: the successful transmission probability of an I, P and B frame, respectively.
- R_F: the maximum playable frame rate achieved when there is enough available capacity and no packet loss (typical full-motion video rates have $R_F = 30fps$).
- R_I, R_P, R_B: the playable frame rate of I, P and B frames, respectively.
- R: the total playable frame rate of the streaming MPEG, which is a summation of R_I, R_P, and R_B.
- 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 modeling capacity constraint, if applicable.

4. STATIC MPEG FILES

This section considers the impact of GOP length on static MPEG file properties and suggests guidelines for GOP considerations. First, N_{BP} is varied with specific values of N_P, to study the role of N_{BP} and demonstrate the impact of N_{BP}. Then N_{BP} is fixed to study the impact of N_P.

4.1 Study of N_{BP}

Increasing the number of B frames decreases the correlation between the B frames and the frames they reference [7]. 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 [10] 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. [6] extracted video data from DVDs and also found the most common value of N_{BP} is no more than two.

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

Tables 2 and 3 depict 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 and $N_P = 4$ in Table 3). 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 4.2, explores N_P in more detail). As N_{BP} increases, the quality

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

Table 2: Impact of N_{BP} on MPEG files for *Foreman* (N_P=1)

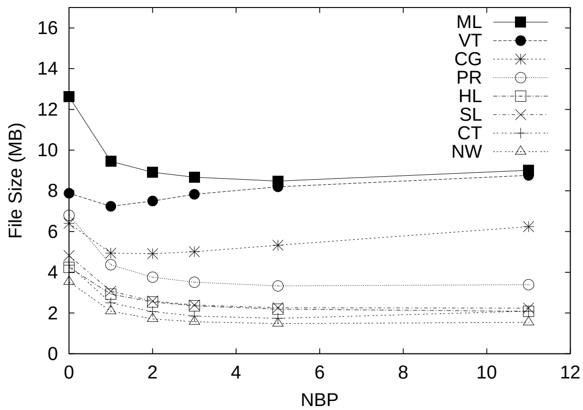
N_P=4	Frame Size (KB)		PSNR (dB)		File Size
N_{BP}	S_P	S_B	Q_P	Q_B	(MB)
0	12.05	N/A	41.0	N/A	4.19
1	14.17	7.57	41.0	36.6	3.45
2	15.31	8.60	41.1	34.7	3.33
3	15.93	9.42	41.1	33.9	3.35
5	17.35	10.56	41.1	32.6	3.48
11	19.17	12.81	41.1	30.9	3.93

Table 3: Impact of N_{BP} on MPEG files for *Foreman* (N_P=4)

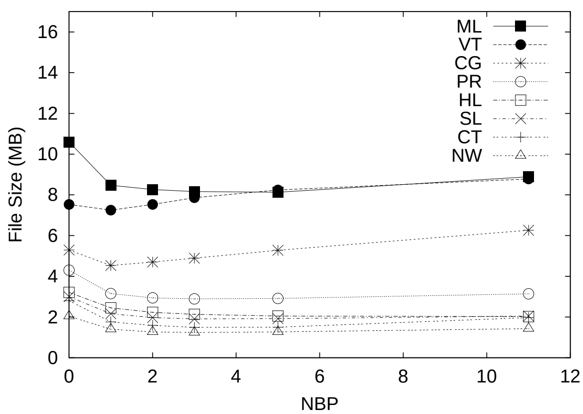
of the B frames decreases quickly. For example, in Table 2, the PSNR of the B frames drops dramatically from 36.7 dB to 30.9 dB. Notice that when N_{BP} increases, the sizes of the P and B frames also increase. In both tables, the size of the B frame nearly doubles as n_{BP} goes from 1 to 11 and this 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 always 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.

Similar experiments were conducted with the other eight videos in Table 1. Figure 2 shows the impact of N_{BP} on encoded MPEG file size ($N_P = 1$ in Figure 2.a and $N_P = 4$ in Figure 2.b). In each figure, 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 minimum size, for all videos. This result does support previous research [5, 16] 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 Tables 2 and 3 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



a. $N_P = 1$



b. $N_P = 4$

Figure 2: Impact of N_{BP} on MPEG files for the other eight videos

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 [4] and network delays in general [8], the median round-trip times for a variety of network configurations are all 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 will tolerate, but no more than two.

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.

4.2 Study of N_P

Similar to section 4.1, 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.

Table 4 and 5 present frame sizes of the *Foreman* video clip for different N_P sizes ($N_{BP} = 1$ in Table 4 and $N_{BP} = 2$ in Table 5). These tables show when 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

N_{BP}=1	Frame Size (KB)		PSNR (dB)		File Size
N_P	S_P	S_B	Q_P	Q_B	(MB)
0	N/A	7.87	N/A	36.7	4.59
1	14.22	7.65	41.1	36.7	3.87
5	14.17	7.57	41.0	36.6	3.42
9	14.16	7.55	41.0	36.7	3.33
14	14.16	7.56	41.0	36.6	3.27
29	14.19	7.55	41.0	36.6	3.23

Table 4: Impact of N_P on MPEG files for *Foreman* ($N_{BP} = 1$)

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 5: Impact of N_P on MPEG files for *Foreman* ($N_{BP} = 2$)

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 5. However, the reduction in file size is not significant.

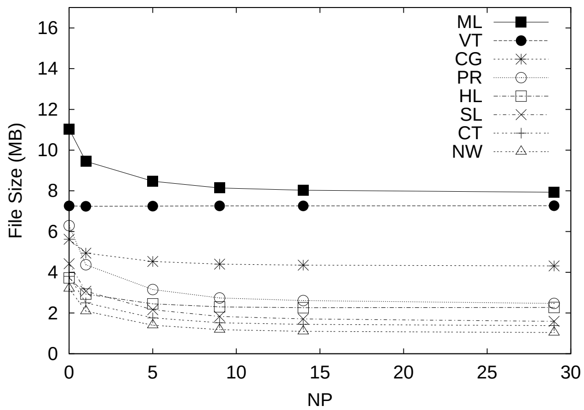
Similar experiments were conducted with the other eight videos in Table 1. Figure 3 presents the impact of N_{BP} on encoded MPEG file size ($N_{BP} = 1$ in Figure 3.a and $N_{BP} = 2$ in Figure 3.b). In each figure, the x-axis is N_P and the y-axis is the encoded file size in Mbytes. More P frames can 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 2 and 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.

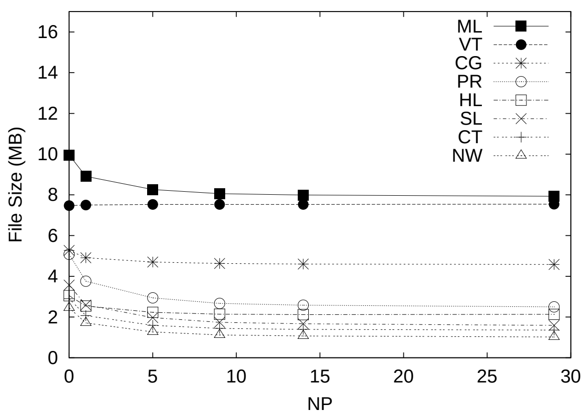
As a summary, while there are no specific constraints concerning the number of P frames, as a guideline, the number of P frames should be no more than 30. Moreover, while having more P frames can increase the compression ratio, the increase is not significant compared to the compression ratio obtained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

5. STREAMING MPEG

The next study is of the impact of GOP length on MPEG streaming over a network model with packet loss (namely, the Internet or a typical wireless network). In previous work [15], an analytical model was first introduced to estimate the MPEG quality at the receiver, measured by playable frame rate. This model can be used to evaluate the impact



a. $N_{BP} = 1$



b. $N_{BP} = 2$

Figure 3: Impact of N_P on MPEG files for the other 8 videos

on streaming MPEG quality with varying combinations of N_P and N_{BP}. For each GOP pattern, with previous guidelines, the raw images are encoded into an MPEG stream, its streaming bitrate and frame sizes are extracted, and then the frame sizes are used as input to the model that determines the playable frame rate. This methodology facilitates analysis of the impact of the GOP pattern on streaming MPEG.

This section begins with the simplest case: no repair and no capacity constraint. Section 6 evaluates streaming with forward error correction and no capacity constraint. Section 7 studies streaming with forward error correction and capacity constraints.

5.1 Model of Playable Frame Rate

Given specific I, P, and B frame sizes, Equation 2 can compute the successful frame transmission probabilities as:

$$\begin{aligned} q_I &= (1-p)^{S_I} \\ q_P &= (1-p)^{S_P} \\ q_B &= (1-p)^{S_B} \end{aligned} \tag{3}$$

Given R_F, the full motion frame rate, to maintain the real-time playout at receiver side, the GOP rate, G, must be kept to:

$$G = \frac{R_F}{(1 + N_P + N_B)} \tag{4}$$

Since I frames are independently encoded, the playable I frame rate is simply the number of I frames transmitted successfully over the network:

$$R_I = G \cdot q_I \tag{5}$$

Since each P frame depends on the success of its previous reference frame and its own successful transmission, the playable P frame rate is:

$$\begin{aligned} R_{P_i} &= R_I \cdot {q_P}^i \\ R_P &= \textstyle\sum_{i=1}^{N_P} R_{P_i} = G \cdot q_I \cdot \frac{q_P - q_P^{1+N_P}}{1-q_P} \end{aligned} \tag{6}$$

Since each B frame depends on the success of its preceding and succeeding reference frames and its own successful transmission, the playable B frame rate is:

$$\begin{aligned} R_{B_{ij}} &= \begin{cases} R_{P_{i+1}} \cdot q_B & when\ 0 \le i \le N_P - 1 \\ R_{P_i} \cdot q_B \cdot q_I & when\ i = N_P \end{cases} \\ R_B &= \textstyle\sum_{i=0}^{N_P} (N_{BP} \cdot R_{B_{ij}}) \end{aligned} \tag{7}$$

So the total playable frame rate of an MPEG flow is:

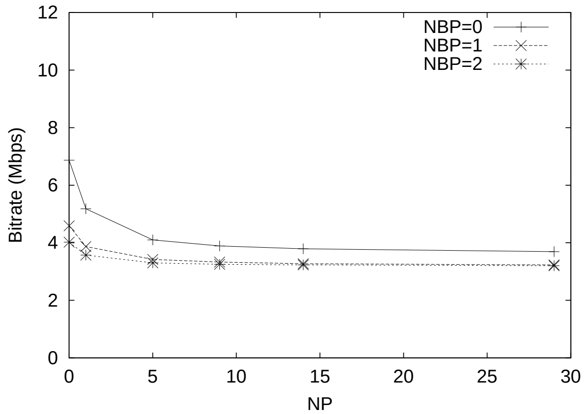
$$\begin{aligned} R &= R_I + R_P + R_B \\ &= G \cdot q_I \cdot (1 + \tfrac{q_P - q_P^{N_P+1}}{1-q_P} \\ &\qquad + N_{BP} \cdot q_B \cdot (\tfrac{q_P - q_P^{N_P+1}}{1-q_P} + q_I \cdot q_P^{N_P})) \end{aligned} \tag{8}$$

5.2 Analysis

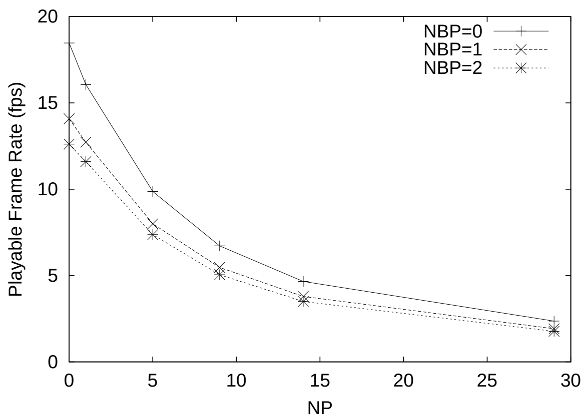
Within the ranges suggested by the guidelines in the previous section ($N_P \le 30$ and $N_{BP} \le 2$), the GOP pattern is varied using different values of N_P and N_{BP} to encode the MPEG stream. For each stream, the frame sizes are extracted, rounded to integer values, and fed into our model (Equations 3 to 8). By comparing the streaming bitrates extracted from the MPEG encoder against the playable frame rates computed by our model over different streams, the impact of GOP pattern on MPEG streaming is analyzed.

Figure 4 shows results from the encoder and the model for the *Foreman* video. In both graphs, the x-axis is the number of P frames in each GOP, and each curve represents a specific N_{BP}. Figure 4.a depicts the streaming bitrate of each MPEG stream over a range of different GOP patterns and Figure 4.b graphs the playable frame rate of each stream with the packet loss rate in the model set at 2%. These figures show that a small N_{BP} yields a high bitrate and a high quality. For a network with sufficient available capacity, a reasonable choice would be to have no B frames at all. However, for typical networks with capacity constraints, some number of B frames are typically preferred to reduce the bitrate. On the other hand, a smaller N_P also yields a high bitrate and a high quality. However, unlike the N_{BP} effect, the bitrate of the MPEG stream decreases faster when N_P decreases up to 5 and does not significantly decrease after N_P=5, while the quality drops rapidly with an increase N_P over the whole range.

Similar experiments were conducted with the other eight videos. Figure 5 depicts the impact of N_P ($N_{BP} = 2$) on

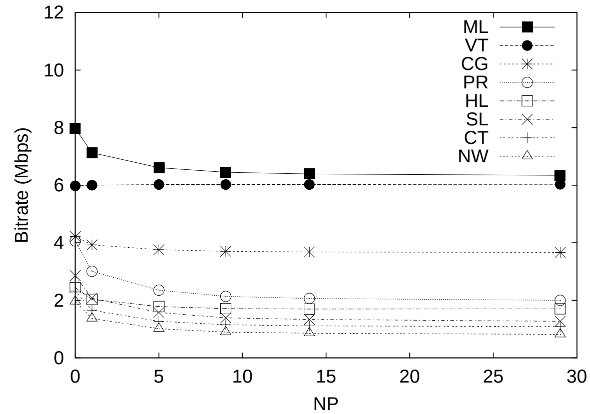


a. Streaming bitrate

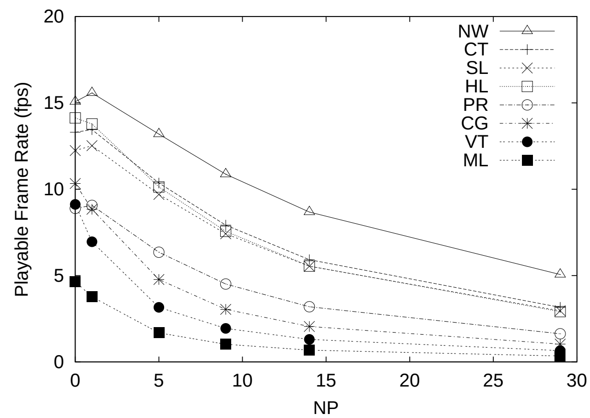


b. Playable frame rate, Network model has 2% loss and no capacity constraint.

Figure 4: Streaming *Foreman* with no repair.



a. Streaming bitrate



b. Playable frame rate, Network model has 2% loss and no capacity constraint.

Figure 5: Streaming the other eight videos with no repair. $(N_{BP} = 2)$

streaming MPEG with no repair. The results are very similar to Figure 4, where the streaming bitrate decrease fast when N_P is smaller than 5 and does not significantly change after $N_P = 5$ but the quality drops linearly with N_P over the whole range. This data suggests an additional guideline of having N_P no more than 5 for a good compromise between bitrate and quality.

6. STREAMING WITH FORWARD ERROR CORRECTION (FEC)

This section evaluates the effectiveness of FEC in repairing packet loss for streaming MPEG over a model of a network that experiences packet loss but is not subject to a capacity constraint. First, the model developed in the last section is extended to incorporate FEC. Then, experiments are run to study the impact of the GOP pattern on MPEG streaming when FEC is employed.

6.1 Extended Model

When FEC is used to repair lost packets, the model for playable frame rate is very similar to the model already presented. Equations 4 to 8 are unchanged. However, in place of Equation 3, the successful transmission probability of each type frame is expressed by:

$$\begin{aligned} q_I &= q(S_I + S_{IF}, S_I, p) \\ q_P &= q(S_P + S_{PF}, S_P, p) \\ q_B &= q(S_B + S_{BF}, S_B, p) \end{aligned} \tag{9}$$

6.2 Analysis

Experiments similar to those in the Section 5 were set up. Figure 6 depicts the results of these experiments with 2% packet loss and 5% fixed FEC for the MPEG frames for the video *Foreman*. The figure of streaming bitrate is not presented since it is very similar to Figure 4.a, the only difference being that this approach adds 5% overhead to the streaming bitrate. Compared against Figure 4 performance, FEC significantly improves the playable frame rate while adding 5% overhead to the streaming bitrate. However, the impact of the GOP pattern on bitrate and quality are very similar to those in the previous section. A small N_{BP} produces a high quality video, but also yields a high bitrate. On the other hand, the bitrate of the MPEG stream decreases fast when N_P decreases up to 5 and does not change much after N_P=5, while the quality decreases quickly with N_P over the entire range.

Figure 7 depicts the impact of N_P ($N_{BP} = 2$) on stream-

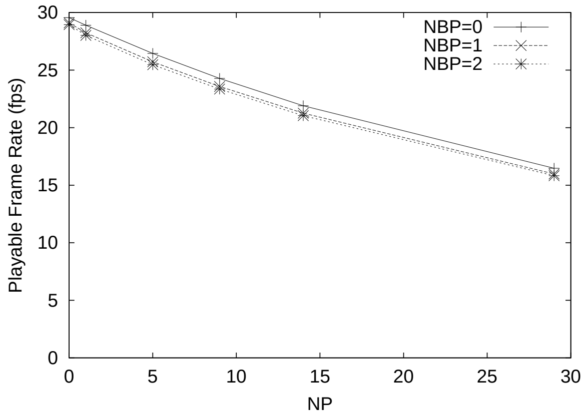


Figure 6: Streaming *Foreman* with 5% forward error correction. Network model has 2% loss and no capacity constraint.

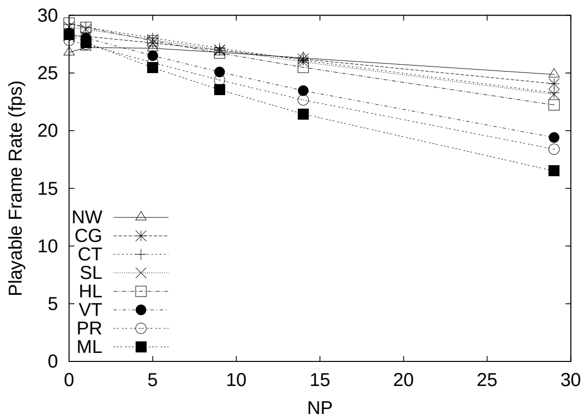


Figure 7: Streaming the other eight videos with 5% forward error correction. The network model has 2% loss and no capacity constraint. ($N_{BP} = 2$)

ing MPEG with 5% fixed FEC for the other eight videos. The streaming bitrate is not presented since it is very similar to Figure 5.a. These results parallel the result of the *Foreman* video, where the streaming bitrates decrease fast when N_P is smaller than 5 and do not significantly change after $N_P = 5$. The quality drops quickly with N_P over the whole range.

Thus, for a network model with loss and no capacity constraints, streaming with FEC still carries the guidelines that N_P should be no more than 5 and N_{BP} should be 1 or 2.

7. STREAMING WITH FEC AND CAPACITY CONSTRAINT

In typical conditions, the network capacity available for streaming can be less than the encoding capabilities. In many cases, researchers have promoted the need for TCP-Friendly bitrates, the capacity is constrained by the Internet Service Provider, or the last-mile connection has limited capacity. This section studies the impact of the GOP pattern on MPEG streaming under conditions of limited capacities

To adjust the streaming bitrate to the available bitrate, streaming systems use media scaling, and often *temporal scaling* where select video frames are discarded at the sender before transmission. In this section, one form of temporal scaling is introduced: Pre-Encoding Temporal Scaling (PETS).[3] Our model is extended again to incorporate PETS and to facilitate experiments that consider the effect of GOP pattern with PETS.

7.1 Pre-Encoding Temporal Scaling (PETS)

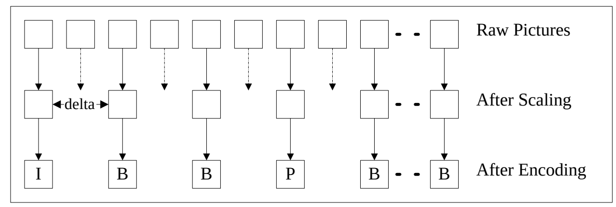


Figure 8: Pre-Encoding Temporal Scaling (PETS)

As depicted in Figure 8, 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 distance between two neighboring encoded images. For example, in Figure 8, where every other image is discarded, δ is one since the 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 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:

$$G(\delta) = \frac{R_F/(1+\delta)}{(1+N_P+N_B)} = \frac{R_F}{(1+\delta)\cdot(1+N_P+N_B)} \quad (10)$$

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 will increase. 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 of δ:

$$\begin{aligned} S_P(\delta) &= S_{P0} + (\delta/\Delta)\cdot(S_I - S_{P0}) \\ S_B(\delta) &= S_{B0} + (\delta/\Delta)\cdot(S_I - S_{B0}) \end{aligned} \quad (11)$$

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

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 11 well. Notice that the sizes of the I frames do not change with δ since I frames use intra-image compression only.

7.2 Extended Model

When PETS and FEC are used in streaming MPEG, the model for playable frame rate is similar to the model presented in Section 6. Equation 4-8 and Equation 9 still hold. 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 11. Second, since with PETS, some of the images may be discarded before encoding, the GOP rate must be decreased as in Equation 10 to keep real-time playout. After these changes, Equation 8 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 only limited by the capacity constraint. This extended model can be used to optimize the playable frame rate, R:

$$\begin{cases} \mathit{Maximize}: \\ \quad R = R(\delta, (S_{IF}, S_{PF}, S_{BF})) \\ \mathit{Subject\ to}: \\ \quad G(\delta)\cdot((S_I(\delta)+S_{IF}) + N_P\cdot(S_P(\delta)+S_{PF}) \\ \qquad +N_B\cdot(S_B(\delta)+S_{BF})) \le T \end{cases} \quad (12)$$

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 exhaustively searched to determine the temporal scaling level and FEC pattern that yield the maximum playable frame rate under the capacity constraint.

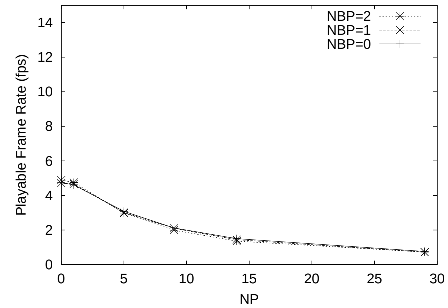
7.3 Analysis

Similar to the previous two sections, 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 12) 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 behavior is analyzed.

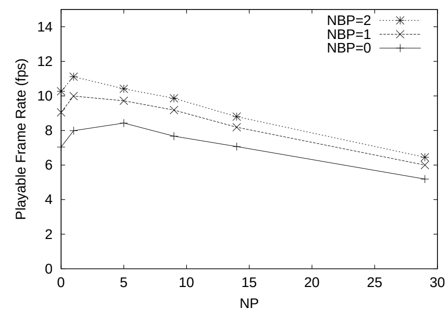
Three different FEC choices are considered:

- Non-FEC: The sender adds no FEC to the video, as in Section 5.
- 5% Fixed FEC: The sender protects each frame with FEC the size of 5% of the original frame size, as in Section 6.
- Adjusted FEC: Before transmitting, the sender uses our extended model (Equation 12) to determine the FEC pattern and temporal scaling level that produce the maximum playable frame rate and uses these for the entire video transmission.

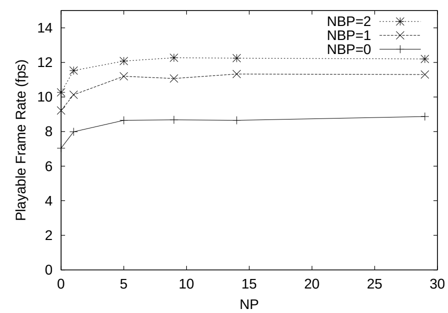
In all cases, the bitrates used by the MPEG video with added FEC are scaled by PETS to meet the capacity constraints.



a. Non-FEC



b. 5% Fixed FEC



c. Adjusted FEC

Figure 9: Streaming *Foreman* with FEC and PETS. Network model has 2% loss and 1.5 Mbps capacity constraint

Figures 9 show performance results for a set of experiments with a 1.5 Mbps capacity constraint and with 2% induced modeling 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 to 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 N_{BP}s 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 10 graphs the impact of N_P ($N_{BP} = 2$) on stream-

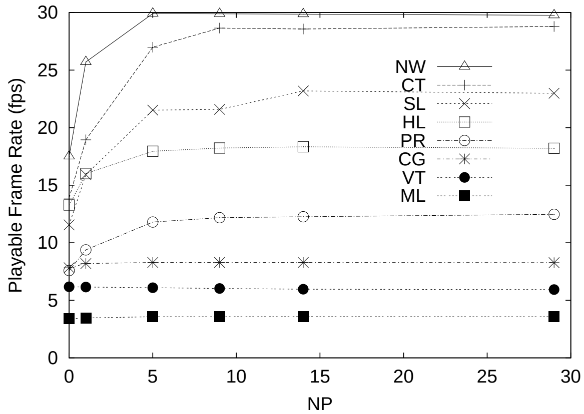


Figure 10: Streaming the other 8 videos with adjusted FEC and PETS, Network model has 2% loss and 1.5 Mbps capacity constraint. ($N_{BP} = 2$)

ing MPEG with adjusted FEC for the other eight videos, where the network model has a 1.5 Mbps capacity constraint and a 2% induced packet loss. These results also suggest there is little to be gained by having N_P greater than 5.

8. SUMMARY

This paper presents an organized methodology to understand the practical impact of both the GOP length and the detailed GOP pattern (the number of B frames between P frames) on static and streaming MPEG. Utilizing results from experiments and analytic modeling, practical guidelines are put forth for setting the GOP length 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. By deliberately varying GOP in the MPEG files, it was possible to observe and compare the properties of the encoded MPEG files. It was observed that although B frames have the highest compression rate, increasing the number of B frames decreases inter-frame compression and quality. These results suggest two guidelines: 1) The number of B frames between two reference frames should not exceed two; 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 the MPEG stream was sent under three distinct circumstances: streaming with no repair and no capacity constraint; streaming with forward error correction (FEC) and no capacity constraints; and streaming with FEC and capacity constraints. In the first scenario, a short GOP can produce a high quality video at the cost of a high bitrate, but having N_{BP} close to 2 and N_P no more than 5 provides a good compromise between bitrate and quality. In the second case, FEC greatly improves quality, but the impact of GOP is similar to that in the first case. Finally, when temporal scaling is used to adjust the bitrate to available network capacity, the optimal MPEG quality always occurs when $N_{BP} = 2$ and $N_P \leq 5$. The results from these three cases suggest two guidelines: 1) The number of B frames between two reference frames should be kept at 2 except when constrained lower by delay constraints; and 2) the number of P frames need not be more than 5.

9. REFERENCES

[1] P. Bocheck, A. Campbell, S.-F. Chang, and R. Lio. Utility-based Network Adaptation for MPEG-4 Systems. In *Proceedings of International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Basking Ridge, NJ, USA, June 1999.

[2] J.-C. Bolot, S. Fosse-Parisis, and D. Towsley. Adaptive FEC-Based Error Control for Internet Telephony. In *Proceedings of IEEE INFOCOM*, New York, NY, Mar. 1999.

[3] J. Boyce and R. Gaglianello. Packet Loss Effects on MPEG Video sent over the Public Internet. In *Proceedings of ACM Multimedia*, pages 181–190, Bristol, U.K., Sept. 1998.

[4] 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.

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

[6] 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.

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

[8] 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.

[9] R. Mahajan, S. Floyd, and D. Wetherall. Controlling High-Bandwidth Flows at the Congested Router. In *Proceedings of the 9th International Conference on Network Protocols (ICNP)*, Nov. 2001.

[10] 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.

[11] J. Mitchell and W. Pennebaker. *MPEG Video: Compression Standard*. Chapman and Hall, 1996. ISBN 0412087715.

[12] K. Park and W. Wang. QoS-Sensitive Transport of Real-Time MPEG Video Using Adaptive Forward Error Correction. In *Proceedings of IEEE Multimedia Systems*, pages 426 – 432, June 1999.

[13] I. S. Reed and G. Solomon. Polynomial Codes Over Certain Finite Fields. *Journal of the Society of Industrial and Applied Mathematics (SIAM)*, 8(2):300–304, June 1960.

[14] A. Tripathi and M. Claypool. Improving Multimedia Streaming with Content-Aware Video Scaling. In *Workshop on Intelligent Multimedia Computing and Networking (IMMCN)*, Durham, North Carolina, USA, Mar. 2002.

[15] 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.

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