Practical Group of Pictures Considerations

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

Abstract— **The MPEG standard is one of the most popular open standards for video on the Internet. MPEG uses intra-frame and inter-frame compression and has 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 pattern to be used by MPEG affects static MPEG properties such as frame size and file size and also impacts the delivery of the video to the receiver side of the network in terms of streaming bitrate and playable quality. Current GOP choices are made using intuition and rules-of-thumb that are not supported by theoretical or practical evidence.**

This paper studies the choice of GOP pattern by evaluating the impact of GOP on both static MPEG properties and the performance of MPEG streaming flows over a lossy network. The static analysis involves encoding raw video images into MPEG files and varying GOP patterns to compare and contrast static properties such as the frame size, file size and PSNR. In the MPEG streaming analysis the GOP length and pattern are varied to study the impact of GOP on streaming bitrate and playable frame rate. The MPEG streaming analysis considers 3 distinct cases over a network that encounters packet loss: normal streaming with no repair and no bandwidth limit; streaming with Forward Error Correction but no bandwidth limit; and streaming with FEC and bandwidth limit. All the results are consistent and suggest two guidelines: 1) the number of B frames between two reference frames should be close to 2, except when limited by the encoding and time constraints; 2) there is little performance gain in setting the number of P frames in the GOP larger than 5.

I. Introduction

Accessing digital video over the Internet continues to grow in popularity. However, sending these motion pictures as uncompressed streams would require raw data rates in the range of tens of Mbps. This is obviously impractical given current home last-mile connections to the Internet. Thus network video products need 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[6] has emerged as one of the most popular international standards for motion picture video compression.

MPEG was originally designed to encode moving pictures and associated audio for digital storage media on CD-ROM[10] with transmission rates at 1.5 Mbps. Over time, MPEG capabilities have been increased to support higher bit rates and larger picture sizes as popular digital video applications moved beyond simple storage devices to streaming multimedia applications over the Internet. Moreover, the near term expectation is for support of digital video over lossy wireless networks.

MPEG uses three types of frames (I, P and B frames) to implement differents compression methods and support inter-frame dependences within the MPEG packet stream. Typically, the MPEG-encoded video flow consists of a repeated pattern of I, P, and B frames that is 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 determines inherent properties of the encoded MPEG file such as size of the three frame types, file size and quality. It also decides the MPEG streaming performance in terms of streaming bitrate and user perceptual quality.

Currently GOP choice is mostly an intuitive process. Some researchers use the default GOP pattern that comes with the 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 [9] the author searchs a large range of GOPs to find the optimal GOP for MPEG streaming and the results use 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. The question of how to practically choose a reasonable GOP pattern has been little studied in any organized way.

The goal of this paper is to investigate practical GOP considerations with respect to performance of MPEG encoded video that is streamed over a network that encounters some packet loss and deal with capacity limitations due to either concurrent TCP flows along the end-to-end flow path or last-mile connection limits. This research consists of two main components - the study of static MPEG and analysis of streaming MPEG. 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 PSNR). In the MPEG streaming situation, 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: the number of B frames between two reference frames should be set to 2 when the video stream does not encounter severe delay constraint, and there is little advantage in setting the number of P frames in a GOP larger than 5.

The contributions of this paper are three-fold. First, the impact of GOP on static MPEG files and dynamic MPEG streaming performance are 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 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 temporal scaling to be used in tandem with FEC.

This paper is organized as follows. Section II briefly presents background knowledge about MPEG, MPEG streaming and forward error correction. The methodology and system settings are presented in section III. Section IV studies static MPEG and section V-VII analyzes the behaviors of MPEG streaming with GOP choices under three distinct situations. Finally, Section VIII summarizes the paper's contributions.

II. BACKGROUND

A. MPEG and Group of Pictures

MPEG is a popular compression standard for video. It 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. However, the increased frequency of I frames must be traded off against the higher compression rates due to P and B frames.

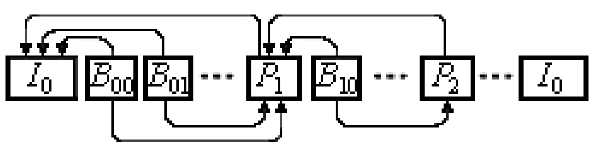


Fig. 1. A sample MPEG Group Of Pictures

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 GOP. Obviously, $N_B = (1 + N_P) \times N_{BP}$ 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.

B. Streaming MPEG

MPEG was originally designed for CD-ROM and initially supported a data rate of 1.5Mbps. On the current Internet, which provides high capacity to the end user (over 1Mbps for home user and around 10Mbps for the user in universities or companies), MPEG can also be used as a video streaming standard due to its excellent quality and high compression rate. Prior to studying MPEG behavior over a network, it is worthwhile to review how MPEG flow characteristics differ from those of traditional network traffic such as web flows.

First, unlike conventional flows, MPEG flows can tolerate some loss but are sensitive to delay and jitter. While one missing bit makes a executable file useless, the loss of a couple of video frames may not be perceivable. Users will wait many seconds for a web page to download but they will not watch a movie which plays fast for seconds then plays slow for another period. TCP, the de facto standard transmission protocol, uses retransmission to recover packet loss but introduces delay to the application. TCP also uses an AIMD mechanism to frequently adjust its bitrate 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 [14].

However, an MPEG object is typically large and, consequentially, needs a high bitrate for streaming. For example, the size of a movie with moderate quality, 1 minute length, 320x240 size, is around 9MB. When it is streamed in real time, the video needs a bandwidth over 1Mbps. A couple of such flows will easily saturate the network and create congestion collapse. UDP is an unresponsive protocol that 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 consensus that Internet applications must be TCP-Friendly [8]. This implies limiting the flow's bitrate to that of an equivalent TCP flow. Besides the constraint from congestion avoidance, the bandwidth of an end-user is also physically limited by the plan he/she purchased.

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 long RTT networks, streaming MPEG can use FEC (Forward Error Correction) to repair the video after packet loss [2], [11]. 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 I frame makes the whole GOP useless while a packet loss in B frame damages only that B frame. For this reason, it is more desirable to add different levels of FEC to each type of frame.

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

On the other hand, if a streaming video is to operate within bandwidth limits, additional FEC data reduces the effective transmission rate of the original video content. To preserve real-time streaming media playout, MPEG servers must scale back their streaming data rate to match the constrained data rate using *Media Scaling*[1], [13]. With media scaling, the MPEG server adjusts the scaling level before transmission, where increasing the scaling level can save capacity for the FEC overhead and vice versa. 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 attempts to optimize the quality of the video stream.

In previous works [15], we derive an analytic model that characterizes 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 scaling level and number of FEC packets for each type of MPEG frame and computes the total playable frame rate. The model then is used to exhaustively search all possible combinations of FEC and scaling level to find the combination of FEC and media scaling that yields the maximum distorted playable frame rate under the bandwidth constraint. The analytic calculations 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.

C. 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, we apply FEC at the packet level. Thus, we model an application level video frame as being transmitted in K packets where K varies with frame type, encoding method, and media content. Media independent FEC [12] 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:

$$q(N,K,p) = \sum_{i=K}^{N} \left[\begin{pmatrix} N \\ i \end{pmatrix} (1-p)^i * p^{N-i} \right] \tag{1}$$

Notice, when no FEC is used, this equation reduces to:

$$q(K,K,p) = (1-p)^K \tag{2}$$

III. Methodology and System Settings

This investigation has two main components - the study of static MPEG (as recorded to a file) and the study of MPEG streaming. In the static MPEG analysis, we vary the GOP length and observe the properties of the resultant MPEG file, noting file size, frame sizes and video quality (measured by PSNR). In the MPEG streaming analysis, we vary the GOP length and evaluate the dynamic behavior of MPEG streaming, in terms of bitrate and video quality (measured by playable frame rate).

The static MPEG analysis involves 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) Study the impact of N_P on frame sizes and frame quality (measured by PSNR) to suggest a guideline for N_P.

The guidelines obtained in the static MPEG analysis are applied to the MPEG streaming analysis which consists of the following steps:

1) Develop a model for streaming MPEG, which can analytically estimate the video quality (measured by playable frame rate). This model is then used in a network with packet loss to study the impact of GOP length on performance.
2) Extend our model to include adding fixed FEC to the MPEG stream, and study the impact of GOP length on playable frame rate and bitrate.
3) Incorporate a capacity constraint into our model. To adjust the streaming bitrate under the constraint, we assume using temporal scaling and modify our model to incorporate it. We set up experiments and study the results for this model.

All the experiments use the Berkeley MPEG encoder and decoder[2] on Linux. However, the results should hold for other MPEG encoders as well since the choice of encoder has little impact on compression compared to the specified quantization level and GOP pattern. We use the video clip *Tennis*, which comes with the Berkeley tools, for our analysis. Tennis, a short clip of two men playing ping-pong, has 150 raw images and lasts for 5 seconds, giving a frame rate of 30 fps. The size of each frame is 352x240 pixels.

The parameters used are provided in alphabetic order:

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.

q_I, q_P, q_B: the successful transmission probability of an I, P and B frame, respectively.

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

R_F: the maximum playable frame rate achieved when there is enough available capacity and no 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 frame.

R: the total playable frame rate of an MPEG stream, 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 capacity constraint, if applicable.

IV. Static MPEG Files

This section considers the impact of GOP length on static MPEG file properties and suggests guidelines for GOP considerations. First, we vary N_{BP} with certain 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.

A. Study of N_{BP}

Based on [6], increasing the number of B frames decreases the correlation between the B frames and the frames they reference. 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 interval. This results in a GOP of `IBBPBBPBB...IBBPBB..` and more generally implies that N_{BP} commonly has a value of no more than two. In Mayer-Patel's work [9], he used the frame rate of 30 fps and a minimum ratio of reference frames to all frames of 1/3, which implies N_{BP} is also less than three. Feng et al. [4] extracted video data from DVDs and also found the most common value of N_{BP} was no more than two.

We conducted experiments by encoding raw images into MPEG videos with different N_{BP} and check 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	9.74	N/A	36.3	N/A	2.13
1	11.17	6.72	36.3	33.3	1.63
2	11.77	7.45	36.3	31.8	1.51
3	13.67	8.05	36.1	31.3	1.52
5	14.08	9.01	36.2	30.7	1.54
11	15.04	10.70	36.2	29.9	1.69

TABLE I
IMPACT OF N_{BP} ON MPEG FILES (N_P=1)

Table I and II depict the frame sizes of the Tennis video clip for different N_{BP} sizes with a fixed number of P frames ($N_P = 1$ in table I and $N_P = 4$ in table II). We do not provide information on the I frames 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 (next subsection (IV-B) explores N_P in more detail). As N_{BP} increases, the quality of the B frames decreases quickly.

N_P=4	Frame Size (KB)		PSNR (dB)		Bitrate
N_{BP}	S_P	S_B	Q_P	Q_B	(MB)
0	9.25	N/A	36.3	N/A	1.67
1	10.61	6.50	36.3	33.3	1.41
2	11.27	7.27	36.3	31.8	1.37
3	13.09	7.88	36.2	31.2	1.43
5	12.72	8.88	36.2	30.7	1.46
11	15.43	10.66	36.1	29.9	1.68

TABLE II
IMPACT OF N_{BP} ON MPEG FILES (N_P=4)

For example, in table I, the PSNR of the B frames drops dramatically by 3.4dB. 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 size is nearly doubled and explains why the MPEG file size also increases when N_{BP} increases above 2. Ideally, more B frames, which are usually smaller than I frame, can reduce the MPEG file size. However since the size of the B frame increases when there are more B frames, the MPEG file can not have a higher compression rate with a large number of B frames. 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 introduce low inter-frame compression and lower quality. Thus it is better to have N_{BP} close to two.

Another practical constraint of N_{BP} is that during streaming, B frames can not be decoded until the subsequent I or P frame has arrived and imply latency increases that are linear with the number of B frames. For interactive applications, latency contributes to the end-to-end delay. For the typical full-motion streaming (frame rate of 30 fps), each B frame contributes 33 ms of delay. In studies of streaming video on the Internet [5] and for network RTTs in general [7], the median RTTs for different bandwidth configurations are all around 100 ms. Thus, compared to the RTT, a few B frames will not signficantly increase the end-to-end delay. However, since more than three B frames can more than double the end-to-end delay, the number of B frames between two reference frames should also be limited to under three.

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

B. Study of N_P

Similar to section IV-A, we run experiments by encoding the raw images into MPEG videos with different N_P values and check the impact on file size, frame sizes and PSNR.

Table III and IV presents frame sizes of the Tennis video clip for different N_P sizes ($N_{BP} = 1$ in table III and $N_{BP} = 2$ in table IV). These tables show when N_P increases, the sizes of the P and B frames do not signficantly 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 IV. However, the reduction in file size is not significant.

N_{BP}=1	Frame Size(KB)		PSNR (dB)		File Size
N_P	S_P	S_B	Q_P	Q_B	(MB)
0	N/A	7.16	N/A	33.5	1.93
1	11.18	6.72	36.3	33.3	1.62
4	10.61	6.50	36.3	33.3	1.40
9	10.68	6.45	36.3	33.3	1.34
14	10.33	6.39	36.3	33.2	1.30
29	10.53	6.40	36.3	33.2	1.29

TABLE III

IMPACT OF N_P ON MPEG FILES ($N_{BP} = 1$)

N_{BP}=2	Frame Size(KB)		PSNR (dB)		File Size
N_P	S_P	S_B	Q_P	Q_B	(MB)
0	N/A	7.85	N/A	31.9	1.70
1	11.78	7.45	36.3	31.8	1.49
4	11.27	7.27	36.3	31.8	1.37
9	11.04	7.19	36.3	31.9	1.31
14	11.18	7.18	36.3	31.8	1.30
29	11.01	7.16	36.3	31.8	1.28

TABLE IV

IMPACT OF N_P ON MPEG FILES ($N_{BP} = 2$)

Another factor associated with the number of P frames is that to support VCR-like functions (pause, rewind, etc.) for MPEG applications, GOP length should not be too long. For example, if a user wants to fast forward with a precision of 2 seconds, the GOP length should be no more than 60 and therefore the number of P frames should not be more than 20.

As a summary, we found that while there are no specific constraints about the number of P frames, we recommend as guidelines that the number of P frames should be no more than 20. We apply this guideline to experiments in all subsequent sections.

V. MPEG Streaming

We next study the impact of GOP length on MPEG streaming over a network with packet loss (namely, the Internet or a typical wireless network). This section begins with the simplest case: no repair and no bandwidth constraint. In Section VI we evaluate streaming with forward error correction and no bandwidth constraint. Then, in Section VII we study streaming with forward error correction and bandwidth constraints.

We first introduce an analytical model to estimate the MPEG quality at the receiver, measured by playable frame rate. Then we vary the combination of N_P and N_{BP} using the previous guidelines. For each GOP pattern, the raw images are encoded into an MPEG stream, its frame size and streaming bitrate are extracted, and then these results are used as input to our model that determines the playable frame rate. This methodology facilitates analysis of the impact of the GOP pattern on MPEG streaming.

A. The model of playable frame rate

Given specific I, P, and B frame sizes, one can use Equation 2 to compute the successful transmission probabilites of each frame type 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, we have:

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

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

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

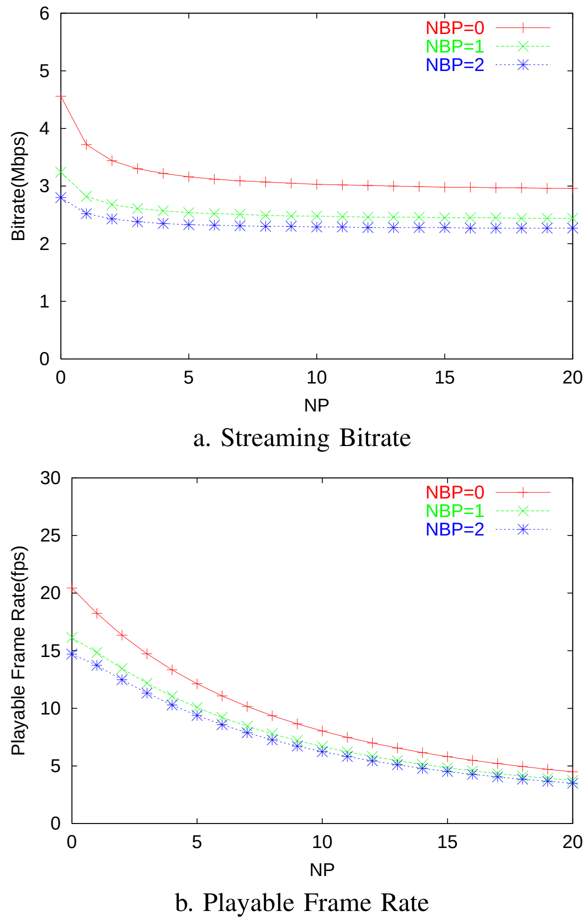
So the total playable frame rate of an MPEG flow is:

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

B. Analysis

Within the ranges suggested in the previous section ($N_P \le 20$ and $N_{BP} \le 2$), we vary the GOP pattern with different values of N_P and N_{BP} to encode the MPEG stream. For each stream, we extract the frame sizes and input them into our model (equation 3 to 8). By comparing the streaming bitrates extracted from the MPEG streams against the playable frame rates computed by our model over different streams, we analyse the impact of GOP pattern on MPEG streaming.

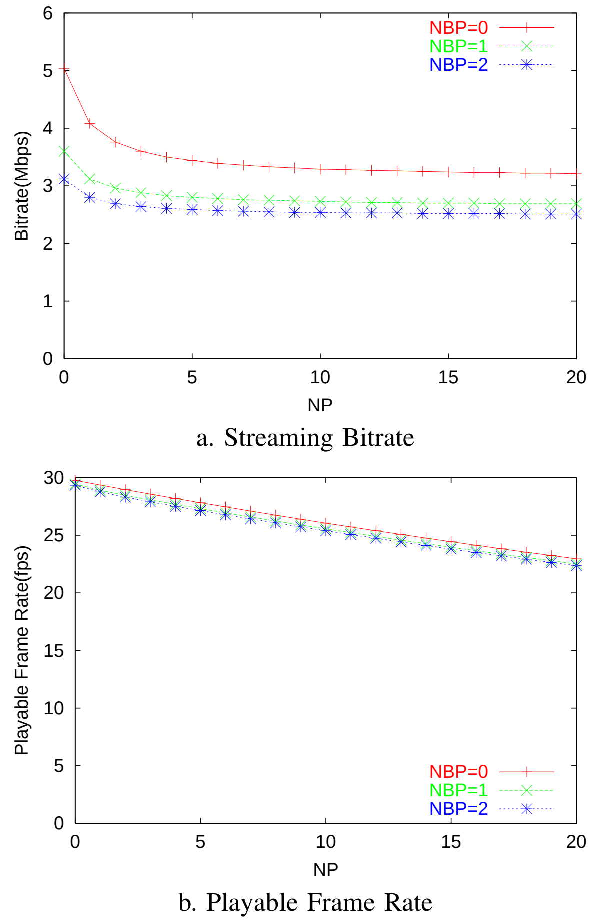
Figure 2 shows experimental results with packet loss set at 2 percent. In both figures, the x-axis is the number of P frames in each GOP, and each curve represents a specific N_{BP}. Figure 2a depicts the streaming bitrate of each MPEG stream over a range of different GOP pattern and Figure 2b graphs the playable frame rate of each stream seen at the receiver side. These figures show that a small N_{BP} yields a high bitrate and a high quality. For a network with sufficient available bandwidth, a reasonable choice would be to have no B frame at all. However, for typical networks with bandwidth constraints, some number of B frames are typically preferred to reduce



a. Streaming Bitrate

b. Playable Frame Rate

Fig. 2. MPEG streaming with no repair. Network with 2 percent loss and and no bandwidth constraint.



a. Streaming Bitrate

b. Playable Frame Rate

Fig. 3. MPEG streaming with forward error correction. Network with 2 percent loss and no bandwidth constraint.

bitrates. 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 signficantly change much after N_P=5, but the quality drops quickly with N_P over the whole range. These facts suggest an additional guideline of having N_P no more than 5 for a good compromise between bitrate and quality.

VI. Streaming with Forward Error Correction

This section evaluates the effectiveness of forward error correction to repair packet loss for an MPEG streamed flow over a network that experiences packet loss but is not subject to a bandwidth constraint. First, the model developed in the last section is extended to incoporate FEC. Then, experiments are run to study the impact of GOP pattern on MPEG streaming when FEC is employed.

A. Extended Model

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

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

B. Analysis

We set up experiments similar to those in the previous section. Figure 3 depicts the results of these experiments with 2% packet loss and 10% fixed FEC for the MPEG frames. Compared against Figure 2 performance, FEC significantly improves the playable frame rate while adding 10% 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} yields a high quality, but also a very 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.

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

VII. Streaming with FEC and Temporal Scaling

For typical streaming conditions, network bandwidth is constrained. In many cases, researchers have promoted the need for TCP-Friendly bandwidth use and the end-user is typically contrainted by the capacity provided by the Internet Service Provider. This section studies the impact of the GOP pattern on MPEG streaming under conditions of limited bandwidth.

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

A. Pre-Encoding Temporal Scaling

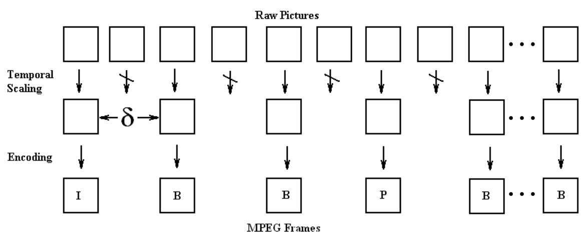


Fig. 4. Pre-Encoding Temporal Scaling (PETS)

As depicted in Figure 4, Pre-Encoding Temporal Scaling saves bandwidth 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, 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 increases because of discarded frames.

To measure the discarding rate, we introduce the variable δ defined as the distance between two neighboring encoded images. For example, in Figure 4, 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, PETS needs to adjust its GOP rate 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 should 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:

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

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:

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

where S_{P0} and S_{B0} are the sizes of the P and B frames in the MPEG video without pre-encoding temporal scaling, respectively. Experiments, not shown here due to lack of space, 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.

B. Extended Model

When PETS and FEC are used in MPEG streaming, the playable frame rate model is similar to the one in Section VI. 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 is a function of the scaling level δ, as in Equation 11. Second, since some of the images will 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 bandwidth constraint. This extended model can be used to optimize the playable frame rate, R:

$$\begin{cases} Maximize: \\ \quad R = R(\delta, (S_{IF}, S_{PF}, S_{BF})) \\ Subject\ to: \\ \quad G(\delta)\cdot((S_I(\delta) + S_{IF}) + N_P\cdot(S_P(\delta) + S_{PF}) \\ \qquad + N_B\cdot(S_B(\delta) + S_{BF})) \le T \\ \quad 0 \le \delta \le \Delta \\ \quad 0 \le S_{IF} \le S_I, 0 \le S_{PF} \le S_P, 0 \le S_{BF} \le S_B \end{cases} \tag{12}$$

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 bandwidth constraint.

[3]Temporal scaling can also be done after encoding, or POst encoded Temporal Scaling (POTS). Due to space limitations, POTS analysis is not presented here.

C. Analysis

Similar to the previous two sections, we vary the GOP pattern for different values of N_P and N_{BP} to encode the MPEG stream. For each stream, we extract the frame sizes and input them into our model (Equation 12) to find the optimal playable frame rate. By comparing the playable frame rates of different streams, we analyse the impact of the GOP pattern on MPEG streaming behavior.

We consider three different FEC choices:

- Non-FEC: The sender adds no FEC to the video, as in Section V.
- 10% Fixed FEC: The sender protects each frame with FEC the size of 10% of the original frame size, as in Section VI.
- 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, as in Section VII.

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

Figures 5 and 6 graph performance results for a large set of experiments with a 1.2 Mbps bandwidth constraint and with 0% and 2% induced packet loss, respectively. Non-FEC is more effective than fixed FEC when the packet loss rate is low since FEC introduces overhead and reduces the bitrate for the image data in the MPEG stream. Fixed FEC is more effective when there is considerable loss since it repairs the loss, preventing degradation in the video quality. In all cases, the adjusted FEC scheme searches the space of choices for the best value of FEC and thus always yields the best quality.

More importantly to the focus of this paper, the impact of GOP on MPEG streaming, these figures show results similar to those in the previous sections. All the six graphs clearly demonstrate that larger N_{BP}s yield better quality (although delay constraints for interactive applications still limit N_{BP} to be 2) and there is little to be gained by having $N_P > 5$.

VIII. 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, we have put forth practical guidelines for setting the GOP length and selecting an appropriate GOP pattern over a range of MPEG streaming 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. We 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 be not exceed two; and 2) we encountered no specific limitations to the number of P frames in a GOP pattern, but we recommend that there be no more than 20 P frames in the GOP pattern to support VCR-like functions for a streamed MPEG flow.

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 bandwidth constraint; streaming with FEC and no bandwidth limit; and streaming with FEC and bandwidth constraints. In the first scenario, a short GOP can produce a high quality video at the cost of high bandwidth. This strategy is not practical under most network streaming situations. 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 by delay constraints; and 2) the number of P frames needn't be more than 5.

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)*, 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*, 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] W. chi Feng, J. Choi, W. chang 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)*, Apr. 2003.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

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

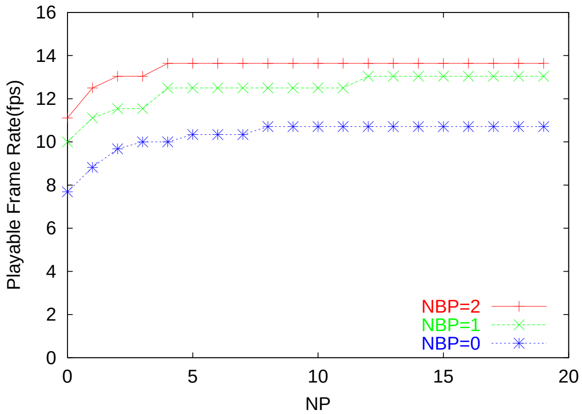
[11] K. Park and W. Wang. QoS-Sensitive Transport of Real-Time MPEG Video Using Adaptive Forward Error Correction. In *Proceedings of IEEE Multimedia Systems*, pages 426 – 432, June 1999.

[12] 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.

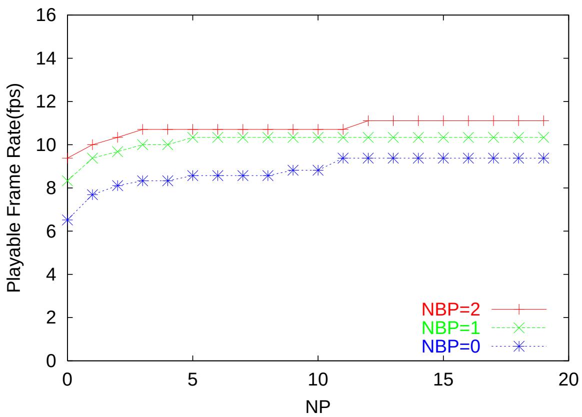
[13] A. Tripathi and M. Claypool. Improving Multimedia Streaming with Content-Aware Video Scaling. In *Workshop on Intelligent Multimedia Computing and Networking (IMMCN)*, Mar. 2002.

[14] Y. Wang, M. Claypool, and Z. Zuo. An Empirical Study of RealVideo Performance Across the Internet. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop (IMW)*, pages 295 – 309, San Francisco, California, USA, Nov. 2001.

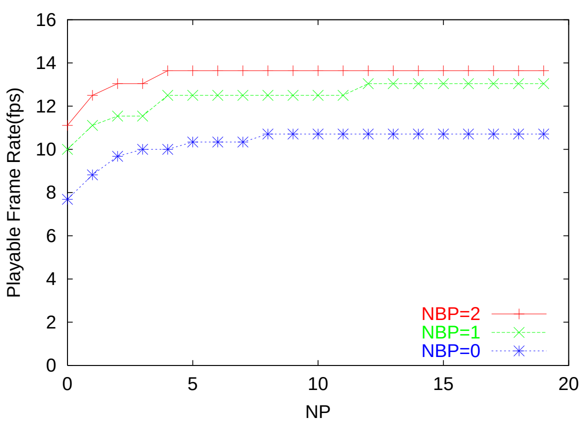
[15] H. Wu, M. Claypool, and R. Kinicki. Adjusting Forward Error Correction for TCP-Friendly Streaming MPEG. In *Proceedings of the Packet Video Workshop (PV), (Poster)*, Nantes, France, Apr. 2003.



a. Non-FEC

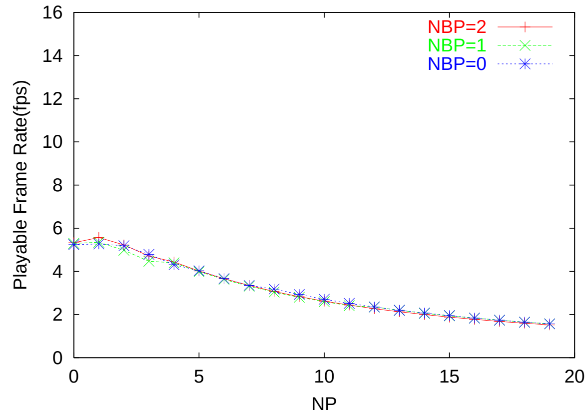


b. 10% Fixed FEC

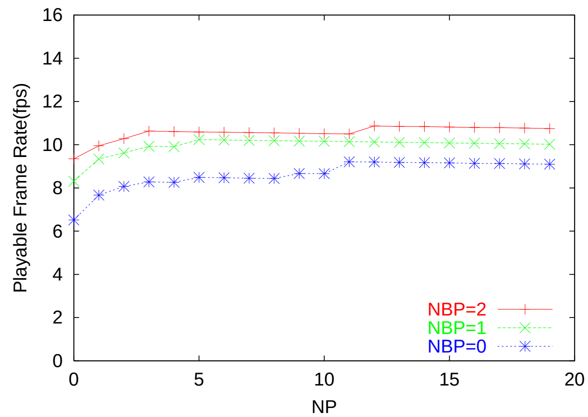


c. Adjusted FEC

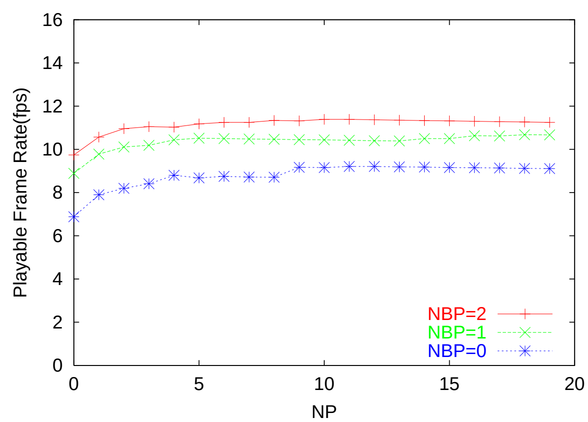
Fig. 5. Streaming with FEC and PETS, 0% loss and 1.2 Mbps bandwidth



a. Non-FEC



b. 10% Fixed FEC



c. Adjusted FEC

Fig. 6. Streaming with FEC and PETS, 2% loss and 1.2 Mbps bandwidth