An MPEG Performance Model And Its Application To Adaptive Forward Error Correction

Ketan Mayer-Patel
University of North Carolina
Sitterson Hall
Chapel Hill, NC 27599
kmp@cs.unc.edu

Long Le
University of North Carolina
Sitterson Hall
Chapel Hill, NC 27599
le@cs.unc.edu

Georg Carle
FhG FOKUS
Kaiserin-Augusta-Allee 31
D-10589 Berlin, Germany
carle@fokus.fhg.de

ABSTRACT

We present a general analytical model for predicting the reconstructed frame rate of an MPEG stream. Our model captures the temporal relationships between I-, P, and B-frames but is independent of the channel and media characteristics. We derive an adaptive FEC scheme from the general model and verify it by comparing it to the results of a simulation. The prediction error of the model compared to the simulation for a wide array of parameter values is less than 5%. We then use the derived adaptive FEC scheme to study the optimal rate allocation (i.e., between generating a higher frame rate or increasing the protection for a lower frame rate) when equation-based TCP rate control is used to couple packet rates to channel characteristics such as round trip time and packet loss probabilities. Surprisingly, we find that optimal protection levels for I- and P-frames are relatively static as loss rates increase from 1% to 4% while changes in the frame type pattern are used to ameliorate the effects of the increased loss. The study demonstrates how our model can be used to reveal joint source/channel coding tradeoffs and how they relate to encoding and transmission parameters.

1. INTRODUCTION

The concept of joint source/channel coding (JSCC) is to simultaneously solve the problems of efficiently representing a signal source (i.e., compression) and of transmitting the signal across a lossy channel (i.e., loss amelioration or recovery). While Shannon's separation principle [17] tells us that in theory an optimal joint solution can be found from optimal separate solutions to these two problems, in practice, these problems are often best considered jointly. For streaming continuous media, JSCC is often embodied as the problem of choosing media encoding parameters based on network conditions and/or designing network protocols that are informed by how the media is encoded.

Unfortunately, the effects of media encoding parameters, packet loss, and protocol mechanisms for loss amelioration are highly non-linear, often hard to predict, and encoding specific. Analytical models that capture these relationships are important tools for predicting performance and guiding adaptation decisions.

This paper presents an analytical model for the MPEG frame structure in which complex temporal dependencies exist between frames. Briefly, there are three types of frames in an MPEG stream. I-frames contain all data necessary for decoding and do not depend on any other frames. P-frames depend on the previous I- or P-frame transmitted. B-frames depend on both the previous I- or P-frame in display order as well as the subsequent I- or P-frame in display order. We review this structure in more detail in Section 3.1 in order to highlight a few important subtleties. A model that captures these relationships is important because some form of this frame structure exists in many different encoding schemes. This temporal structure exists in MPEG-1, MPEG-2, and to some degree in MPEG-4 and H.263.

The pattern of I-, P-, and B-frames is an important component in determining the effect of loss. If an I-frame is lost, all frames until the next I-frame are lost (or at least greatly damaged). Because the dependency of P-frames form a chain, the effects of losing a P-frame depend on the length of the dependency chain and the frame's position within the chain. In the presence of packet loss, a number of different strategies can be employed. One strategy is to reduce the length of the P-frame dependency chain by generating I-frames more often. The cost of this strategy is paid in the fact that I-frames are generally larger than P-frames and thus the total number of frames per second generated would decrease given a fixed bandwidth budget. Another strategy would be to use forward error correction (FEC) to protect I- or P-frames or soft-ARQ (i.e., retransmission) to recover lost packets of I- or P-frames. Again, the price of such techniques is a lower total number of frames per second generated as bandwidth must be redirected to these mechanisms. The first strategy employs source coding (i.e., changing the parameters of the encoding itself), the second strategy employs channel coding. The strategies are not mutually exclusive, and the most effective solution may lie in using both to some degree. Our model can be used to evaluate coding parameter changes and resource allocation decisions and predict performance.

The main contributions of this paper are:

- **A general model that captures MPEG frame-type relationships.** We first present a general model which expresses the expected number of frames per second of an MPEG stream that can be decoded and displayed in terms of abstract functions that determine the rate at which each frame type is generated and a frame type-specific probability of successful transmission.

- **A specific instantiation of the model for adaptive FEC.** We then show how this general model can be instantiated for a specific transmission scheme. In this case, we construct instantiations of the abstract functions for an FEC scheme in which bandwidth can be allocated toward protecting frames from loss in a frame-type specific manner. In other words, the level of protection for I-frames can be different than the level of protection of P-frames or B-frames.

- **A demonstration of how the model can be used.** With the FEC instantiation of the model, we demonstrate how the model can be used to explore the tradeoffs between source coding and channel coding for maximizing expected performance. In particular, we couple our model with equation-based TCP rate control to explore the optimal FEC allocation decision as a function of loss rates.

The rest of the paper is organized into 6 sections. Section 2 reviews related work. The general model is presented in Section 3 and its specific instantiation for adaptive FEC is presented in Section 4. We measure the accuracy of the FEC-specific model by comparing MPEG performance predicted by the model to the results of a simulation. These results are presented in Section 5. We demonstrate how the model can be used to explore joint source/channel coding tradeoff in Section 6. Finally, Section 7 summarizes the paper.

2. RELATED WORK

The problem of MPEG transmission over lossy packet-switched networks is well studied. The efforts have, for the most part, been concentrated on a few areas. One strategy is to adapt the output of a variable bit rate (VBR) MPEG-2 coder to dynamic network conditions ([7, 8, 19]). These investigations concentrate mainly on adapting the quality of the encoding and not on the mix of frame types. For pre-recorded MPEG-2 video, research has concentrated on smoothing techniques to ensure delivery of video streams given constraints on server capacity, bandwidth, and/or receiver buffer space ([3, 6, 15, 16]). The effects of packet loss on MPEG-2 video was studied in [2]. To ameliorate the effects of loss, FEC and soft-ARQ protocols have been proposed and studied by a number of researchers ([9, 10, 14, 18]). Using MPEG-2 scalability features to divide the video stream into two or more layers and transmitting each layer separately is studied in [1, 4, 5, 11, 13]. These studies are most relevant to the work presented here and we discuss a subset more thoroughly below.

Of the related work, Wolfinger in [18] takes an approach most similar to the approach we have taken. He also derives an analytical framework for assessing the frame loss probability given different frame type patterns and the use of FEC. The resulting expression has a very similar form as the expression we derive and involves similar terms. The main difference, however, is that Wolfinger assumes that the target rate is constant and independent of packet loss probability. In our work, the target transmission rate is derived from equation-based congestion control and thus is a function of the packet loss probability. The coupling of the loss rate and the target rate results in much more complex adaptation behavior. Furthermore, Wolfinger restricts himself to frame type patterns such that the distance between I frames is no more than 10 frames. This restriction is artificial and does not even allow for the most commonplace static pattern in use today in which the inter-I frame distance is 15.

In [7], the optimal allocation between FEC and MPEG-2 source coding is studied using a perceptual distortion metric and an analytical framework that allowed for partially complete frames. In this work, however, no distinction is made for FEC protection of I-frames versus that of P-frames versus that of B-frames. In fact, the encoding pattern of frame types and the total frame rate is held constant while only quality of the encoded frames is changed. In contrast, we maintain an average frame size for each frame type (implying a possibly changing quality factor) while allowing for variable frame rate and frame type patterns.

In [14], an adaptive FEC scheme that incorporates congestion control is described. They concentrate their investigation in coupling the rate allocated to FEC packets to a congestion-aware control law. The frame-type coding pattern as well as the source data rate is fixed and only the amount of FEC generated is allowed to vary. The scheme can be parameterized to allow separate FEC treatment of I-, P-, and B-frames, but these allocation tradeoffs are not made dynamically and instead are set statically. The scheme is evaluated empirically.

In [1], the MPEG video stream is split into two layers. One layer is considered high-priority and one layer is considered low-priority. While the high-priority data is required to recover any video, the low-priority data simply provides additional information to improve frame quality. The word layer is a bit misleading since information from both layers is packed into every packet. In other words, a receiver can not choose to receive only high-priority information. The purpose of the layering is to allow FEC to be applied only to the high-priority information. A fixed amount of FEC is applied and total transmission rate is kept constant. Again, the frame type was not a factor. The scheme was evaluated empirically, testing the effectiveness of different amounts of FEC and the tradeoff with video quality.

In summary, our work differs fundamentally from that of others primarily in two ways:

- We develop an analytical model that incorporates both the frame type pattern and the resulting reference structure as well as frame type-specific FEC.

Frames:	I	B	B	P	B	B	P	B	B	P	B	B	P	B	B	I
Display Order:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Transmission Order	1	3	4	2	6	7	5	9	10	8	12	13	11	15	16	14

Figure 1: Typical MPEG frame type pattern with display and transmission orders.

- We couple our rate allocation with a dynamic TCP-friendly equation-based congestion control scheme in which loss probability is the driving parameter.

3. GENERAL MODEL

This section presents the general model which captures the MPEG frame type relationships. We first review the MPEG frame structure and their temporal dependencies. We then proceed to build the model using abstract functions for the rate at which different frame types are produced and the probability with which they are successfully transmitted.

3.1 MPEG Frame Structure

As we briefly described in the introduction, MPEG frames can be classified as one of three types: I, P, or B. I-frames contain all information required to decode and display the frame. P-frames may employ motion compensation techniques that exploits information in the previous I- or P-frame. Thus, a P-frame is dependent on the previous I- or P-frame. B-frames further exploit motion compensation techniques by using information in both the previous I- or P-frame as well as the subsequent I- or P-frame in display order. Because information in an I- or P-frames may be used by other frames (i.e., other P-frames or B-frames), the term *reference frame* is used to describe a frame that is either an I-frame or a P-frame.

One subtlety that arises from these temporal relationships is that the transmission order of MPEG frames may not be the same as its display order. For example, Figure 1 uses the letters “I”, “P”, and “B” to symbolize a sequence of MPEG frames. Below the sequence, the frames are labeled with their display and transmission order. Notice that frame 4 (a P-frame) is sent ahead of frames 2 and 3 which precede it in display order. This is because frames 2 and 3 are B-frames which may depend on frame 4, the subsequent reference frame. The 15 frame sequence of frame types shown in Figure 1 is a de facto standard used by many MPEG encoders (the 16th frame in the sequence is the start of the next 15 frame sequence). Although this pattern is very common, it is not mandated by the standard and almost any pattern can be used. Later in the paper, we will refer to this 15 frame pattern as the “standard” pattern.

3.2 Model Components

To build the general model, we introduce a number of abstract functions to represent the rates at which different frame types are generated and the probabilities with which they are successfully transmitted. These abstract functions must be instantiated for specific encoding and transmission schemes. One such instantiation for an adaptive FEC scheme is given in Section 4.

First, we define functions for the number of frames of a particular frame type generated per unit time:

$f_I(.)$ The number of I-frames generated per second.

$f_P(.)$ The number of P-frames generated per second.

$f_B(.)$ The number of B-frames generated per second.

Now, we define functions for the probability that a frame of a specific type is successfully transmitted. Note that this is different from the probability of being able to reconstruct and display the frame because of the interframe dependencies of I-, P-, and B-frames. These functions only describe the probability of successfully transferring the data associated with a specific frame. Define these functions as:

$g_I(.)$ The probability of successful transmission for an I-frame.

$g_P(.)$ The probability of successful transmission for an P-frame.

$g_B(.)$ The probability of successful transmission for an B-frame.

Combining these functions, we can express the expected number of frames per second that will be successfully received for each frame type as:

$f_I^*(.) = f_I(.)g_I(.)$ The expected number of I-frames per second successfully transmitted.

$f_P^*(.) = f_P(.)g_P(.)$ The expected number of P-frames per second successfully transmitted.

$f_B^*(.) = f_B(.)g_B(.)$ The expected number of B-frames per second successfully transmitted.

3.3 Estimated Reconstructed Frame Rate

The general model provides an expression for the estimated number of frames per second successfully *reconstructed* given the interframe dependency structure of I-, P-, and B-frames. This estimate can be broken down into frame type-specific components. In other words, the expected reconstructed frame rate can be expressed as the sum of the expected reconstructed frame rates for I-, P-, and B-frames. Letting $E(.)$ represent the total expected reconstructed frame rate and $E_I(.)$, $E_P(.)$, and $E_B(.)$ type-specific reconstruction rates, the general model is expressed as:

$$E(.) = E_I(.) + E_P(.) + E_B(.) \tag{1}$$

To increase readability, we will omit the symbols “(.)” used to indicate an abstract function for the rest of this section. We now derive analytical expressions for each subcomponent of E.

3.3.1 E_I

The expression for E_I is simply the number of successfully transmitted I-frames since these frames have no dependency on any other frames. Thus we have:

$$E_I = f_I^* \tag{2}$$

3.3.2 E_P

An expression for E_P is not as easily constructed. The recovery of a P-frame depends in part on the recovery of the reference frame it depends on. This reference frame is either an I-frame or another P-frame.

For the subsequent analysis, we define the following symbols:

- p_I Probability of successful reconstruction of an I-frame.
- p_P Probability of successful reconstruction of a P-frame.
- p_R Probability of successful reconstruction of a reference frame (i.e., either a I- or P-frame) required for a P-frame.

We first note that for I-frames, the probability of successful reconstruction is the same as the probability of successful transmission. Thus:

$$p_I = g_I \tag{3}$$

We can rearrange the terms of the definition of f_I^* to provide an expression for g_I in terms of f_I and f_I^* giving us:

$$p_I = \frac{f_I^*}{f_I} \tag{4}$$

With these symbols defined, we can begin to find E_P by setting it to the P-frame generation rate multiplied by probability of successful reconstruction of a P frame.

$$E_P = f_P * p_P \tag{5}$$

The probability p_P is the joint probability that the P frame was both successfully *transmitted* (g_P) and the probability that the reference frame the P-frame depends on was successfully *reconstructed*. This gives us:

$$p_P = g_P * p_R \tag{6}$$

An expression for p_R depends on the relative frame rates of I- and P-frames. If the P-frame rate is less than or equal to the I-frame rate, then we expect at most one P-frame between any two I-frames, This means that all P-frames will simply depend on the I-frame that precedes it. This yields the expression:

$$p_R = p_I; \quad \text{when } f_P \leq f_I \tag{7}$$

Substituting the expression for p_I defined in equation 3:

$$p_R = \frac{f_I^*}{f_I}; \quad \text{when } f_P \leq f_I \tag{8}$$

This yields:

$$p_P = g_P \frac{f_I^*}{f_I}; \quad \text{when } f_P \leq f_I \tag{9}$$

We now turn to the case when $f_P > f_I$. If the P-frame rate is greater than the I-frame rate, the average number of P-frames between two I-frames is greater than 1. In this case some proportion of the P-frames will depend on an I-frame and the remaining P-frames will depend on P-frames. Since the first P-frame following every I-frame will depend on that I-frame, the proportion of the P-frames that depend on an I-frame will simply be the ratio of the I-frame rate to the P-frame rate. This proportion is simply $\frac{f_I}{f_P}$. The remaining proportion of P-frames will depend on a preceding P-frame. This proportion is simply $1 - \frac{f_I}{f_P}$ which can rewritten as $\frac{f_P - f_I}{f_P}$. Thus, the expression for p_R when $f_P > f_I$ is:

$$p_R = \frac{f_I}{f_P} p_I + \frac{f_P - f_I}{f_P} p_P; \quad \text{when } f_P > f_I \tag{10}$$

We can replace p_I with the expression defined in equation 3 and simplify to:

$$p_R = \frac{f_I^*}{f_P} + \frac{f_P - f_I}{f_P} p_P; \quad \text{when } f_P > f_I \tag{11}$$

Substituting this expression for p_R into equation 6 and solving for p_P yields:

$$p_P = \frac{g_P f_I^*}{f_P - g_P(f_P - f_I)}; \quad \text{when } f_P > f_I \tag{12}$$

The g_P term of the numerator can be pulled out in front, and the g_P term in the denominator can be replaced by $\frac{f_P^*}{f_P}$ (a rearrangement of the definition of f_P^*):

$$p_P = g_P \frac{f_I^*}{f_P - f_P^* + \frac{f_P^*}{f_P} f_I}; \quad \text{when } f_P > f_I \tag{13}$$

We now have a full expression for p_P as:

$$p_P = \begin{cases} g_P \frac{f_I^*}{f_I} & ; \quad \text{when } f_P \leq f_I \\ g_P \frac{f_I^*}{f_P - f_P^* + \frac{f_P^*}{f_P} f_I} & ; \quad \text{when } f_P > f_I \end{cases} \tag{14}$$

Substituting this expression for p_P back into equation 5 and combining the initial $f_P * g_P$ into f_P^* gives us the full expression for E_P as:

$$E_P = \begin{cases} \frac{f_P^* f_I^*}{f_I} & ; \quad \text{when } f_P \le f_I \\ \frac{f_P^* f_I^*}{f_P - f_P^* + \frac{f_P^*}{f_P} f_I} & ; \quad \text{when } f_P > f_I \end{cases} \tag{15}$$

3.3.3 E_B

We construct an expression for E_B in much the same way as for E_P. Define p_B as the probability of successful *reconstruction* of a B-frame giving us a starting point for E_B as:

$$E_B = f_B * p_B \tag{16}$$

The probability for successful reconstruction of a B-frame is the joint probability of successful transmission (i.e., g_B) and the probability of successful reconstruction of the reference frames that the B-frame depends on. If the B-frame precedes a P-frame temporally, the B-frame only depends on the future P-frame. The reference frame that preceded the B-frame does not need to be considered since the probability of its successful reconstruction is already factored into the probability of successful reconstruction of the future P-frame. If the B-frame precedes an I-frame, then the frame depends on both this I-frame and the reference frame that precedes it which may be either an I- or P- frame. In this case, the B-frame depends explicitly on both reference frames because there is no relationship between the I-frame that follows the B-frame and the reference frame that preceded the B-frame.

To arrive at an expression for p_B, we must be able to determine what portion of the B-frames are only dependent on a subsequent P-frame, what portion of the B-frames are dependent on both a subsequent I-frame and a preceding P-frame, and what portion of B-frames are dependent on two I-frames. To determine these portions, we will once again consider the two different cases of when $f_P \le f_I$ and $f_P > f_I$. We assume that B-frames are smoothly distributed between I- and P-frames.

When $f_P \le f_I$, at most one P-frame will exist between any two I-frames. All B-frames in the interval before a P-frame will depend only on that P-frame (again, because the successful reconstruction of the preceding I-frame is factored into the reconstruction of the P-frame). These intervals occur in the same proportion as P-frames occur within the total rate of I- and P-frames. Since we assume that B-frames are evenly inserted into the intervals between I- and P-frames, this allows us to define the portion of B-frames that depend on only a P-frame as:

$$\frac{f_P}{f_P + f_I} \tag{17}$$

Similarly, all B-frames in the interval after a P-frame will depend on both the preceding P-frame and the subsequent I-frame. Since the I- and P-frame are not related, these B-frames depend on both. The proportion of B-frames that follow a P-frame is the same proportion of B-frames that precede a P-frame. Thus, equation 17 also expresses what portion of B-frames are in this condition.

The remaining B-frames occur in intervals between two I-frames and thus depend on the successful reconstruction of two independent I-frames. We can find this portion by simply calculating what is left out of the two previous portions as in:

$$1 - 2\left(\frac{f_P}{f_P + f_I}\right) \tag{18}$$

Which reduces to:

$$\frac{f_I - f_P}{f_P + f_I} \tag{19}$$

Using p_I and p_P as the probability of successful reconstruction for I- and P-frames as in the previous subsection, we can express p_B when $f_P \le f_I$ as:

$$p_B = g_B\left(\frac{f_P}{f_P + f_I} p_P + \frac{f_P}{f_P + f_I} p_P p_I + \frac{f_I - f_P}{f_P + f_I} p_I^2\right) \quad \text{when } f_P \le f_I \tag{20}$$

Rearranging terms and substituting expressions for p_I (from equation 4) and p_P (from equation 14 under the same condition $f_P \le f_I$):

$$p_B = g_B\left(\frac{f_P}{f_P + f_I} g_B \frac{f_I^*}{f_I}\left(1 + \frac{f_I^*}{f_I}\right) + \frac{(f_I - f_P) f_I^{*2}}{(f_P + f_I) f_I^2}\right) \quad \text{when } f_P \le f_I \tag{21}$$

Combining $f_P * g_P$ into f_P^* and rearranging terms, we get:

$$p_B = g_B\left(\frac{f_I^*}{(f_P + f_I) f_I}\left(f_P^* + f_I^* + \frac{f_I^*}{f_I}(f_P^* - f_P)\right)\right) \quad \text{when } f_P \le f_I \tag{22}$$

Now we consider the case when $f_P > f_I$. In this case, at most one I-frame exists between any two P-frames. Thus, no B-frame exists between two I-frames. B-frames that precede a P-frame depend only on that P-frame. Again, this is because the probability of reconstruction of the preceding reference frame is already factored into the probability of reconstruction of the P-frame. B-frames that precede an

I-frame are dependent on the I-frame and the preceding P-frame. Again, we know that the preceding reference frame will be a P frame since $f_P > f_I$ meaning that at least one P-frame exists between any two I-frames.

The fraction of B-frames that precede a P-frame is the fraction of P-frames within the total rate of I- and P-frames, or:

$$\frac{f_P}{f_P + f_I} \tag{23}$$

Similarly, the fraction of B-frames that depend on both an I- and a P-frames is simply the fraction of B-frames that precede an I-frame, or:

$$\frac{f_I}{f_P + f_I} \tag{24}$$

Using p_I and p_P as the probability of successful reconstruction for I and P frames as in the previous sections, we can express p_B when $f_P > f_I$ as:

$$p_B = g_B \left(\frac{f_P}{f_P + f_I} p_P + \frac{f_I}{f_P + f_I} p_I p_P \right) \quad \text{when } f_P > f_I \tag{25}$$

Rearranging terms and substituting expressions for p_I (from equation 4) and p_P (from equation 14 under the same condition $f_P > f_I$):

$$p_B = g_B \left(\frac{f_P g_P f_I^*}{(f_P + f_I)(f_P - f_P^* + \frac{f_P^*}{f_P} f_I)} \right) \left(1 + \frac{f_I f_I^*}{f_P f_I} \right) \quad \text{when } f_P > f_I \tag{26}$$

Canceling terms and replacing $f_P g_P$ with f_P^*, we arrive at:

$$p_B = g_B \left(\frac{f_P^* f_I^*}{(f_P + f_I)(f_P - f_P^* + \frac{f_P^*}{f_P} f_I)} \right) \left(1 + \frac{f_I^*}{f_P} \right) \quad \text{when } f_P > f_I \tag{27}$$

Our full expression for p_B is then:

$$p_B = \begin{cases} g_B \left(\frac{f_I^*}{(f_P+f_I) f_I} \left(f_P^* + f_I^* + \frac{f_I^*}{f_I}(f_P^* - f_P) \right) \right) & \text{when } f_P \le f_I \\ g_B \left(\frac{f_P^* f_I^*}{(f_P+f_I)(f_P - f_P^* + \frac{f_P^*}{f_P} f_I)} \right) \left(1 + \frac{f_I^*}{f_P} \right) & \text{when } f_P > f_I \end{cases} \tag{28}$$

Substituting this expression for p_B back into equation 16 and combining the initial $f_B g_B$ into f_B^* gives us the full expression for E_B as:

$$E_B = \begin{cases} f_B^* \left(\frac{f_I^*}{(f_P+f_I) f_I} \left(f_P^* + f_I^* + \frac{f_I^*}{f_I}(f_P^* - f_P) \right) \right) & \text{when } f_P \le f_I \\ f_B^* \left(\frac{f_P^* f_I^*}{(f_P+f_I)(f_P - f_P^* + \frac{f_P^*}{f_P} f_I)} \right) \left(1 + \frac{f_I^*}{f_P} \right) & \text{when } f_P > f_I \end{cases} \tag{29}$$

3.3.4 The Full Expression for E

Combining the previous results for E_I, E_P, and E_B, the full expression for the estimated reconstructed frame rate is:

$$E() = \begin{cases} f_I^* + \frac{f_P^* f_I^*}{f_I} + f_B^* \left(\frac{f_I^*}{(f_P+f_I) f_I} \left(f_P^* + f_I^* + \frac{f_I^*}{f_I}(f_P^* - f_P) \right) \right) & \text{when } f_P \le f_I \\ f_I^* + \frac{f_P^* f_I^*}{f_P - f_P^* + \frac{f_P^*}{f_P} f_I} + f_B^* \left(\frac{f_P^* f_I^*}{(f_P+f_I)(f_P - f_P^* + \frac{f_P^*}{f_P} f_I)} \right) \left(1 + \frac{f_I^*}{f_P} \right) & \text{when } f_P > f_I \end{cases} \tag{30}$$

The general model has a number of notable features.

- *When $f_I == f_P$, the two halves of the model reduce to each other.* This property serves as a sanity check that the two halves of the model properly capture the temporal relationships when f_I and f_P are nearly the same. This is important because the construction of the model relies on inferring the relationship a B-frame has with its reference frames from the relative generation rates of the two reference frame types.

- *The model relies only on frame generation rates and successful frame transmission rates.* To instantiate the model for a specific transmission scheme, the relationship between frame generation rates and successful frame transmission rates must be captured by specific forms for g_I, g_P, g_B. However, the general model could be used as a way to predict the effect of changes in frame generation rates by measuring f_I^*, f_P^*, and f_B^* for an MPEG stream and then extrapolating values for g_I, g_P, and g_B.

- *The model provides for type-specific behavior.* By using separate abstract functions to represent the frame generation rates and transmission probabilities for each frame type, the model can be used capture the effects of frame type-specific features that the transmission scheme may include. An example that we will develop in the next section is an FEC scheme that affords different levels of protection to the different types of frames. One drawback of the model is that pattern position-specific behavior can not be captured. In other words, the effects of a transmission scheme which treats the first P-frame after an I-frame differently than the second P-frame after an I-frame, and so on, can not be captured.

4. INSTANTIATING THE GENERAL MODEL FOR ADAPTIVE FEC

To instantiate the general model, we must provide instantiations of the abstract functions $f_I(.)$, $f_P(.)$, $f_B(.)$, $g_I(.)$, $g_P(.)$, and $g_B(.)$. This section provides one such instantiation for an adaptive FEC scheme. In this specific case, the problem is cast as a rate allocation decision. Given the available bandwidth budget in terms of packets per second, we assign a specific packet rate to the generation of each frame type and to the generation of FEC for each frame type. Given the frame size (in packets) for each frame type, the bandwidth budget, the rate allocation, and the probability of packet loss, we construct expressions for each of the abstract functions required. Packet losses are assumed to be independent.

4.1 Input Parameters

The parameters to our problem are:

R The total number of packets produced per second.

l The probability of loss for any given packet.

s_I The size of an I-frame in packets.

s_P The size of a P-frame in packets.

s_B The size of a B-frame in packets.

a_{code} The percentage of packets allocated for coding MPEG picture data (i.e., not allocated to FEC).

a_{ref} The percentage of coding packets (i.e., of $R * a_{code}$) that are allocated to coding reference frames.

a_I The percentage of reference frame coding packets that are allocated to coding I-frames.

$a_{FEC_{ref}}$ The percentage of FEC packets (i.e., of $R * (1.0 - a_{code})$) that are allocated to protecting reference frames.

a_{FEC_I} The percentage of reference FEC packets that are allocated to protecting I-frames.

In this formulation, the packet rate R is divided by the percentages a_{code}, a_{ref}, a_I, $a_{FEC_{ref}}$, and a_{FEC_I}. Each percentage represents a different dimension of the allocation decision. The size of I-, P-, and B-frames is assumed to be constant within a picture type. We can think of these static frame sizes as representing the scene content complexity relative to compression. Thus, a particular set of values for these sizes reflects assumptions about scene complexity.

4.2 Frame Rates

First, we recombine the input parameters to define type-specific frame rates and determine the average number of FEC packets generated per frame. Thus defining f_I, f_P, and f_B as:

$$f_I = \frac{(R * a_{code} * a_{ref} * a_I)}{s_I} \tag{31}$$

$$f_P = \frac{(R * a_{code} * a_{ref} * (1.0 - a_I))}{s_P} \tag{32}$$

$$f_B = \frac{(R * a_{code} * (1.0 - a_{ref}))}{s_B} \tag{33}$$

4.3 Frame Transmission Probabilities

From the above expressions for the frame type-specific generation rates, we can calculate the average number of FEC packets generated per I, P, or B frame as:

$$s_I^* = \frac{(R * (1.0 - a_{code}) * a_{FEC_{ref}} * a_{FEC_I})}{f_I} \tag{34}$$

$$s_P^* = \frac{(R * (1.0 - a_{code}) * a_{FEC_{ref}} * (1.0 - a_{FEC_I}))}{f_P} \tag{35}$$

$$s_B^* = \frac{(R * (1.0 - a_{code}) * (1.0 - a_{FEC_{ref}}))}{f_B} \tag{36}$$

For each I-frame, s_I packets are used to code the frame and on average s_I^* packets are provided as FEC protection. Similarly, s_P is the P-frame size, s_P^* is the average number of FEC packets per P-frame, s_B is the B-frame size, and s_B^* is the average number of FEC packets per B-frame.

In general, an FEC scheme in which n packets of data are protected by an additional k packets of redundancy, requires any n of these $n+k$ packets to successfully recover the original data. Given n, k, and a specific independent packet loss probability l, we can define the function $g(n,k,l)$ as the probability of recovering the original data using a Bernoulli model:

$$g(n,k,l) = \sum_{q=n}^{n+k} \binom{n+k}{q} (1-l)^q l^{n+k-q} \tag{37}$$

The first thing to note is that (37) requires that n and k be integers. Since n will correspond to s_I, s_P, or s_B depending on the frame type, we can restrict ourselves to input parameters where this is the case. The number of FEC packets per frame, however, is not an input to the problem. Instead, these values (s_I^*, s_P^*, and s_B^*) are calculated from the rate allocation as per equations 34, 35, and 36. These equations

Scenario	I Size	P Size	B Size	Standard Bit-rate
A	20	10	5	2.52 Mb/s
B	40	15	5	3.43 Mb/s
C	30	20	10	4.8 Mb/s

Table 1: Frame sizes for scenarios used for verification.

calculate the number of FEC packets per frame *on average*. To accommodate this, we can modify equation 37 to calculate the average probability of successful transmission of a frame of integer n packets with on average non-integer k FEC packets as:

$$\hat{g}(n,k,l) = (\lceil k \rceil - k) g(n, \lfloor k \rfloor, l) + (k - \lfloor k \rfloor) g(n, \lceil k \rceil, l) \quad (38)$$

Thus, we can define g_I, g_P, and g_B as:

$$g_I = \hat{g}(s_I, s_I^*, l) \quad (39)$$

$$g_P = \hat{g}(s_P, s_P^*, l) \quad (40)$$

$$g_B = \hat{g}(s_B, s_B^*, l) \quad (41)$$

With definitions for the abstract functions of the general model in place, we now have a specific model for an adaptive FEC scheme. Given a measure of loss, the reconstructed frame rate achieved by a specific set of values for the rate allocation variables can be predicted. This model could be used to find the global optimal rate allocation for a static loss probability. Or, the model could be used as part of an adaptive on-line algorithm to find a local maxima in the rate allocation parameter space as the loss probability is continually measured and estimated.

5. MODEL VERIFICATION

To measure the accuracy of the FEC model, we wrote a simulator which simulated a frame generation process given a target packet rate, the size of I-, P-, and B-frames in terms of packets, and specific rate allocation values for a_{code}, a_{ref}, a_i, $a_{FEC_{ref}}$, and $aFEC_i$. Each packet generated by the frame generation process was chosen to be "lost" randomly with probability l. A simulated decoding process recorded which frames could be successfully decoded taking into account dependencies between a frame and any reference frames it may depend on. We ran the simulation for three different frame size scenarios. Table 1 summarizes the three different scenarios. Frame sizes are given in terms of packets per frame. The column labeled "Standard Bit-rate" contains the corresponding bit-rate associated with each scenario if the standard 15 frame pattern described in Section 3.1 was used with 1500 byte packets at 30 frames per second.

For each scenario, six different target packet rates were simulated. For each target packet rate, eight different loss probabilities were simulated. Each rate allocation parameter was allowed to take one of five values. All combinations of possible rate allocation parameters were tested. In all, 150000 different combinations were tested for each of the three different scenarios. Table 2 summarizes these parameter settings.

Parameter	Values
Packet Rate	220, 260, 300, 340, 380, 420
Loss Probability	.001, .005., .01, .02, .04, .06, .08, .10
Rate Allocation Variables	.1, .3, .5, .7, .9

Table 2: Parameter settings used for verification.

		Prediction Error	Variance
Scenario A	I-Frame Rate	0.7%	0.0002
	P-Frame Rate	3.1%	0.004
	B-Frame Rate	3.4%	0.005
	Total Rate	2.9%	0.003
Scenario B	I-Frame Rate	1.4%	0.001
	P-Frame Rate	4.3%	0.007
	B-Frame Rate	5.7%	0.013
	Total Rate	4.9%	0.009
Scenario C	I-Frame Rate	0.6%	0.0002
	P-Frame Rate	2.7%	0.004
	B-Frame Rate	3.5%	0.007
	Total Rate	2.5%	0.003
Overall		3.4%	0.005

Table 3: Model Prediction Error From Simulation

Each simulation was allowed to run until 5000 frames were generated. For each simulation, we calculate the difference between the reconstructed frame rate of the simulator and the reconstructed frame rate predicted by the model. This difference is the prediction error. We express this difference as a percentage of the prediction. Table 3 lists the resulting prediction error broken out by scenario and frame type as well as the overall average prediction error for all the simulations. Overall, the model has extremely good predictive power relative to the simulation. Error, however, does seem to increase with the size of a frame in packets.

6. COUPLING ADAPTIVE FEC TO TCP-FRIENDLY RATE CONTROL

In this section we demonstrate how we might use our adaptive FEC model to explore joint source/channel coding tradeoffs. We use the TCP-equation given in [12] to calculate the TCP-compatible available bandwidth. The equation is expressed as:

$$X = \frac{s}{R\sqrt{\frac{2bp}{3}} + t_{RTO}(3\sqrt{\frac{3bp}{8}})p(1+32p^2)} \quad (42)$$

where X is the transmit rate (bytes/sec), s is the packet size (bytes), R is the round trip time (sec), p is the loss event rate on the interval [0,1.0], t_{RTO} is the TCP retransmission timeout (sec), and b is the number of packets acknowledged by a single TCP acknowledgment.

Parameter	Value
Round Trip Time	50 ms
Round Trip Time Variance	.005
I-frame Size	20 packets
P-frame Size	12 packets
B-frame Size	3 packets

Table 4: Parameters for experiment coupled with equation-based TCP rate control.

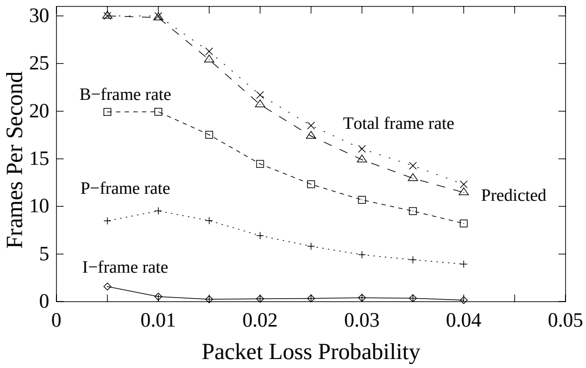

Figure 2: Frame rates for optimal allocation vs. loss probability.

The TCP-equation allows us to relate the available packet rate to the channel characteristics of loss and round trip time. To explore JSCC tradeoffs with our adaptive FEC model, we establish a hypothetical situation by setting values for round trip time, round trip time variance, and frame sizes. These values are listed in Table 4. Using these values, we calculate the packet rate associated with different loss rates and for each loss rate, search the allocation parameter space for an allocation that maximizes our predicted reconstructed frame rate. This optimization was done in two steps. In the first step, local maximums were searched in the entire parameter space using a coarse-grained brute force. These local maximums were then used as starting points for a fine-grained gradient climbing optimization in the second step. The optimal rate allocation is chosen from the results in the second step that gives the maximum reconstructed frame rate. The solution is constrained to a maximum generated frame rate of 30 frames per second and a minimum ratio of reference frames (i.e., I- and P- frames) to all frames of 1/3.

The results are shown in Figure 2, Figure 3, and Figure 4. Figure 2 shows the frame type-specific generated frame rates (i.e., f_I, f_P, and f_B from the model) as well as the total generated frame rate and the predicted reconstructed frame rate. Figure 3 shows the ratio of redundant data to frame data for each frame type (i.e., $\frac{s_I^*}{s_I}$, $\frac{s_P^*}{s_P}$, and $\frac{s_B^*}{s_B}$ from the model). Finally Figure 4 shows the ratio of P-frames generated to I-frames generated (i.e., $\frac{f_P}{f_I}$).

From this simple experiment we can see a number of interesting effects. Below, 1% loss, the TCP-equivalent rate is much higher than the number of packets required to generate 30 frames per second. Because we constrain the generation

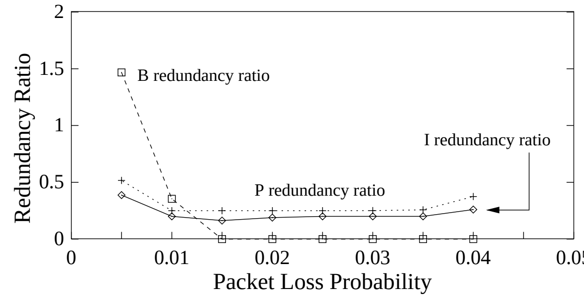

Figure 3: Ratio of redundant data to frame data by type vs. loss probability.

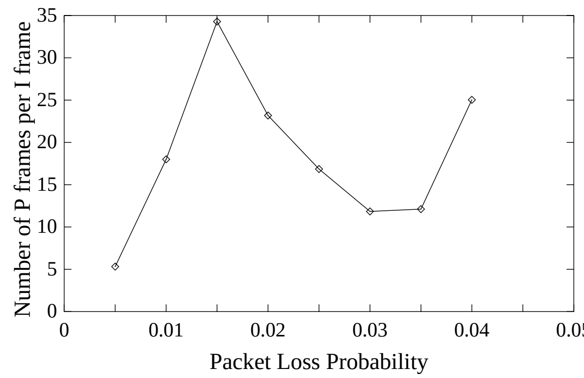

Figure 4: Ratio of P-frames to I-frames vs. loss probability.

processes to a combined rate of 30 frames per second, all of the excess bandwidth is used for redundancy. When excess bandwidth does not exists (i.e. loss is greater than 1%), B-frames are afforded no FEC protection, while I- and P-frames are still protected. Surprisingly, the levels of protection given to I- and P- frames remains relatively flat, even as loss increases. Instead, the number of P-frames generated between I-frames is reduced. We can infer that in this regime, frame pattern (i.e., source coding) is more effective than additional redundancy (i.e., channel coding).

We were also surprised to see that the level of protection for I-frames is lower than that for P-frames. Given the importance of I-frames[1], we had expected that I-frames would be given more protection than P-frames. This again demonstrates the efficiency of the joint source/channel coding concept. Instead of using more redundant data to protect large I-frames, it seems to be more efficient in some regimes to change the interframe dependencies by changing the frame type pattern. The packet rate that would have been allocated for protecting I-frames is spent on P- and B-frames to get more frames to the receiver. As the loss probability rises to 4%, it is necessary to increase the level of protection for I- and P-frames to cope with packet loss. Source coding benefits from the increased level of protection for I-frames and prompts changes to the frame-type pattern by increasing the number of P-frames between I-frames.

[1] If an I-frame is lost, all following P-frames and B-frames cannot be decoded until the next I-frame is received.

Of course, this experiment is somewhat simplified. Round trip time and loss are not completely independent. Also the values chosen for I-, P-, and B-frame size are somewhat ad hoc. Our goal, however, was to demonstrate how the model can be used to explore joint source/channel coding tradeoffs that arise from the complex temporal dependencies of the MPEG frame structure. Without this model, these tradeoffs are not easily seen. The model exposes these tradeoffs and reveals their shape, allowing them to be mapped against the parameters of the MPEG stream and of the specific transmission scheme in use.

7. SUMMARY

We presented a general model for predicting the reconstructed frame rate of an MPEG stream which captures the temporal dependencies between I-, P-, and B-frames. This model relies on abstract functions to represent type-specific frame generation and transmission processes. The general model is instantiated by providing specific functions for a particular encoding and transmission scheme. We provide one such instantiation for an adaptive FEC scheme. We verify the predictive power of this model by comparing it to the results of a simulation. We found that the prediction error between the model and the simulation was less than 5% over a wide array of possible parameter values. Finally, we demonstrated how the model can be used by searching for optimal FEC allocations for a range of loss probabilities given a TCP-equivalent packet rate and specific channel and media characteristics. The model reveals that the optimal redundancy levels for I- and P-frames remains relatively static as the loss probability increases from 1% to 4%, while the optimal ratio of P-frames to I-frames decreases. In this particular case, we learn that for this range of loss probabilities, additional redundancy is not as effective as changes to the frame-type pattern. We believe the value of our model is in providing a basis for investigating these types of joint source/channel coding tradeoffs.

8. REFERENCES

[1] J.M. Boyce. Packet loss resilient transmission of mpeg video over the internet. *Signal Processing: Image Communication*, 15:7–24, September 1999.

[2] J.M. Boyce and R.D. Gaglianelle. Packet loss effects on mpeg video sent over the public internet. *Proceedings of ACM Multimedia*, pages 181–190, 1998.

[3] G. Cao, W.-C. Feng, and W. Singhal. Online vbr video traffic smoothing. *Proceedings of the Eighth International Conference on Computer Communications and Networks (IC3N'99)*, 1999.

[4] M. Domanski, A. Luczak, S. Mackowiak, and R. Swierczynski. Hybrid coding of video with spatio-temporal scalability using subband decomposition. *Proceedings of the SPIE: Visual Communications and Image Processing*, 3653:1018–1025, 1999.

[5] L. Favalli. Hierarchical transmission of mpeg-2 sequences in noisy environments. *European Transactions on Telecommunications*, 12(3):175–180, May 2001.

[6] W.-C. Feng and J. Rexford. Performance evaluation of smoothing algorithms for transmitting prerecorded variable bitrate video. *IEEE Transactions on Multimedia*, 1(3):302–312, September 1999.

[7] P. Frossard and O. Verscheure. Content-based mpeg-2 structuring and protection. *Proceedings of SPIE: Multimedia Systems and Applications II*, 3845, 1999.

[8] S. Jacobs and A. Eleftheriadis. Streaming video using dynamic rate shaping and tcp flow control. *Journal of Visual Communication and Image Representation*, 9(3):211–222, 1998.

[9] H. Ma and M. El-Zarki. Broadcast/multicast mpeg-2 video over wireless channels using header redundancy fec strategies. *Proceedings of the SPIE: Multimedia Systems and Applications*, 3528:69–80, 1999.

[10] H. Ma and M. El-Zarki. Mpeg-2 encoded vod services over fixed wireless channels using arq schemes. *Proceedings of the SPIE: Multimedia Systems and Applications*, 3528:373–384, 1999.

[11] R. Mathew and J.F. Arnold. Efficient layered video coding using data partitioning. *Signal Processing: Image Communication*, 14(9):761–782, July 1999.

[12] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling tcp throughput: A simple model and its empirical validation. *Proceedings of ACM SIGCOMM*, 1998.

[13] P. Pancha and M. El Zarki. Prioritized transmission of variable bit rate mpeg video. *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, 2:1135–1139, 1992.

[14] K. Park and W. Wang. Qos-sensitive transport of real-time mpeg video using adaptive redundancy control. *Computer Communications*, 24(1):78–92, January 2001.

[15] J. Rexford and D. Towsley. Smoothing variable bitrate video in an internetwork. *IEEE/ACM Transactions on Networking*, 7(2):202–215, April 1999.

[16] S. Sen, J.L. Rexford, J.K. Dey, J.F. Kurose, and D.F. Towsley. Online smoothing of variable bitrate streaming video. *IEEE Transactions on Multimedia*, 2(1):37–48, March 2000.

[17] C.E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423, July 1948.

[18] B.E. Wolfinger. On the potential of fec algorithms in building fault-tolerant distributed applications to support high qos video communications. *Proceedings of the Sixteenth Annual ACM Symposium on Principles of Distributed Computing (PODC 97)*, pages 129–138, 1997.

[19] W. Zeng and B. Liu. Rate shaping by block dropping for transmission of mpeg-precoded video over channels of dynamic bandwidth. *Proceedings of ACM Multimedia*, pages 385–394, 1996.