Error-Resilient Unequal Protection of Fine Granularity Scalable Video Bitstreams

Hua Cai*, Bing Zeng†, Guobin Shen*, and Shipeng Li*
* Internet Media Group, Microsoft Research Asia, Beijing, China 100080
Email: t-huacai@microsoft.com, jackysh@microsoft.com, and spli@microsoft.com
† Department of Electrical and Electronic Engineering,
The Hong Kong University of Science and Technology
Email: eezeng@ee.ust.hk

***Abstract*— This paper deals with the optimal packet loss protection issue for streaming the fine granularity scalable (FGS) video bitstreams over IP networks. Unlike many other existing protection schemes, we develop an error-resilient unequal protection (ER-UEP) method that adds redundant information optimally for loss protection and, at the same time, cancels completely the dependency among bitstream after loss recovery. In our ER-UEP method, the FGS enhancement-layer bitstream is first packetized into a group of independent data packets, while each packet can be truncated to represent the original video signal at any fidelity (i.e., scalability). Parity packets are then created with intrinsic UEP capabilities that can easily adapt to the current channel conditions. Unlike conventional UEP schemes that suffer from bitstream contamination due to the dependency among packets, our method guarantees the successful decoding of all received bits, thus leading to a better error resilience as well as higher robustness (under varying and/or unclean channel conditions).**

I. Introduction

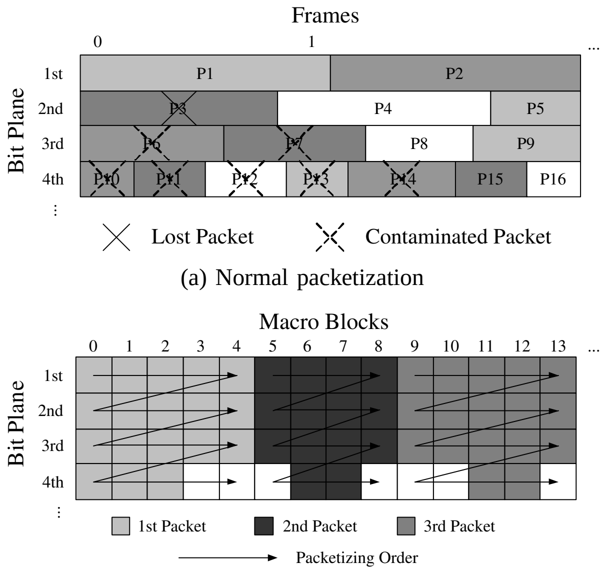
Streaming multimedia contents over the Internet is becoming more and more popular in recent years. However, network heterogeneity and dynamics are well-known hindering factors to successful streaming service. To cope with the varying bandwidth more efficiently, the fine granularity scalable (FGS) video coding [1][2] was proposed recently. An FGS video bitstream consists of two layers: a thin non-scalable base layer (to fit the lowest available bandwidth) and a scalable enhancement layer (can be truncated arbitrarily to adapt to any available bandwidth).

Besides the bandwidth fluctuation, packet losses also affect the streaming quality significantly. Automatic repeat request (ARQ) based methods are often adopted to rescue packet losses. However, they are usually not acceptable for real-time streaming applications because the excessive end-to-end delay caused. On the other hand, forward error correction (FEC) techniques can correct packet losses promptly without any further intervention from the sender [3]. Furthermore, *unequal* protection (UEP) has been widely adopted in many transmission schemes, noticing that different bits are of different importance. In particular, a general and flexible method called priority encoding transmission (PET) was proposed in [4] to cope with packet losses. In a PET system, a whole bitstream is first partitioned into several segments $m_0, ..., m_{K-1}$ with decreasing importance. Unequal FEC is then applied according to the importance so as to provide different levels of error recovering capability. The PET philosophy has also been used in formulating a systematic, fast, and end-to-end R-D optimized transmission scheme called FEC-based multiple description coding (MD-FEC) for scalable multimedia contents [5].

The good performance obtained by the conventional UEP schemes such as PET and MD-FEC depends on the clear knowledge of channel conditions. In practice, however, channel conditions can only be predicted before transmission, and noise-free prediction cannot be guaranteed since the packet loss behavior may variate over time [6]. Moreover, because of the intrinsic feature of scalable coding, the produced scalable bitstream is causally dependent. For example, the segments in PET and MD-FEC schemes are causally dependent, i.e., segment m_i depends on segments $m_0 \sim m_{i-1}$. Clearly, an error in a certain segment will contaminate all the dependent segments. Consequently, these conventional UEP schemes are still sensitive to channel degradations.

By considering both of the dependency feature in FGS enhancement-layer bitstream and the UEP principle for prioritized data, in this paper, we propose a new, general and flexible error-resilient unequal protection (ER-UEP) method to efficiently and robustly cope with packet loss. In our ER-UEP scheme, the FGS enhancement-layer bitstream is first packetized into a group of independent and scalable data packets. Parity packets are then created for error protection. According to channel conditions and based on the UEP principle, the number of parity symbols is selected optimally for symbols at different positions within each data packet. As a result, these numbers would be different from one symbol to another. Compared with the conventional UEP schemes that suffer from bit contamination, our new scheme achieves much better error resilience and can guarantee the successful decoding of all received bits (i.e., no bits are wasted due to contamination).

The rest of the paper is organized as follows. Section II briefly reviews the optimal packetization strategy proposed in [7] that is used to create independent and scalable data packets for FGS video bitstreams. Section III presents our ER-UEP framework. Extensive experimental results are shown in Section IV, with comparisons against the MD-FEC scheme. Finally, some conclusions are drawn in Section V.



(a) Normal packetization

(b) Optimal packetization

Fig. 1. Two packetization strategies.

II. Optimal Packetization for FGS Bitstreams

FGS [1], among other its improved versions such as PFGS [2], generates a very thin base layer and a fully embedded enhancement layer, thanks to the bit-plane coding technique adopted. It is the embedded enhancement layer that provides excellent fine scalability.

Normally, the enhancement-layer bits of each video frame are sequentially ordered on a bit-plane by bit-plane and MB by MB basis, i.e., from the most significant bit-plane of all MBs to the least significant bit-plane of all MBs. For such normal-ordered bitstream, the bits after truncation are put into packets continuously with the constraint of maximum packet length, as depicted by the *normal packetization* in Fig. 1-(a). This packetization strategy has two drawbacks: 1) it is difficult to choose the most efficient bits for transmission; 2) it introduces strong dependency among packets as bits from different bit-planes of the same MB, which are causally dependent, are often packetized into different packets. It is easy to see from Fig. 1-(a) that one packet loss (marked as $P3$) will contaminate many other packets (marked as $P6$, $P7$, and $P10$-$P14$) and render them useless even if they are received successfully.

To overcome the drawbacks of normal packetization, an R-D optimal packetization strategy for FGS enhancement-layer bits was developed in [7]. It first performs a R-D optimal bit allocation across bit-planes and MBs. It then packetizes those selected bits into packets by grouping bits from the same MBs into one packet, and thus packet dependency is completely eliminated. Fig. 1-(b) shows one scenario of performing this optimal packetization. Clearly, any lost packet will not contaminate the decoding of other received packets. Note that each packet is still fine scalable due to the MB by MB and bit-plane by bit-plane sequential scan ordering, as depicted by the *packetizing order* in Fig. 1-(b). That is, each packet can still be truncated arbitrarily. In our ER-UEP framework, we use this strategy to generate data packets.

III. Error-Resilient Unequal Protection

The key difference between this contribution and other UEP schemes is that we achieve UEP and complete cancellation of dependency among bitstreams simultaneously.

A. Performance Metric

The streaming service quality can be quantitatively measured by the expected distortion the end-user perceives. For an FGS video bitstream the base layer is usually very small and of high importance, error-free transmission could be achieved through high-priority protection. As a result, the overall streaming quality is mostly dominated by the enhancement layer. Therefore, we only study the problem of protecting the FGS enhancement-layer bitstream in this paper. All notations such as *bitstream*, *packet*, and *rate* refer to the enhancement-layer bitstream hereafter.

For the bits from the f^{th} frame, the i^{th} MB, and the l^{th} bit-plane, the corresponding *expected* distortion reduction is:

$$\begin{aligned}&\mathcal{E}\{\Delta D(f,i,l)\} = \Delta D(f,i,l) \times (1 - p_e(f,i,l)) \\ &\quad \times p((f,i,0) \sim (f,i,l-1) \text{ received} \mid (f,i,l) \text{ received})\end{aligned} \quad (1)$$

where (f,i,l) represents the bits for the f^{th} frame, the i^{th} MB, and the l^{th} bit-plane; $\Delta D(f,i,l)$ is the *actual* distortion reduction contributed by successfully receiving and decoding all bits in (f,i,l); $p_e(f,i,l)$ is the loss probability of (f,i,l) after FEC recovery (to simplify representation, we assume that all bits of (f,i,l) are transmitted and protected as a whole); and the conditional probability $\underline{p((f,i,0) \sim (f,i,l-1) \text{ received} \mid (f,i,l) \text{ received})}$ represents the impact of bitstream dependency after FEC recovery (since decoding (f,i,l) requires the correctness of all bits from bit-plane 0 to bit-plane $l-1$ in the same MB).

After accumulating all expected distortion reductions, the performance metric is as follows:

$$\mathcal{J} = D_{BL} - \sum_{(f,i,l)\in\mathcal{I}} \mathcal{E}\{\Delta D(f,i,l)\} \quad (2)$$

with the rate constraint

$$\sum_{(f,i,l)\in\mathcal{I}} R_S(f,i,l) + \sum_{(f,i,l)\in\mathcal{I}} R_{FEC}(f,i,l) \leq R_E. \quad (3)$$

In above, D_{BL} denotes the distortion when only the base layer is sent; $\mathcal{I}$ denotes the set of bits that will be transmitted in the current time-slot; $R_S(f,i,l)$ and $R_{FEC}(f,i,l)$ denote respectively the data rate and the channel rate of (f,i,l); and R_E denotes the total available rate in the current time-slot.

The performance metric defined above encourages two efforts to achieve the best streaming quality, namely, 1) unequal protection should be applied to different bits according to their contribution in distortion reduction, and 2) data dependency should be eliminated as much as possible, which is the key focus of this paper.

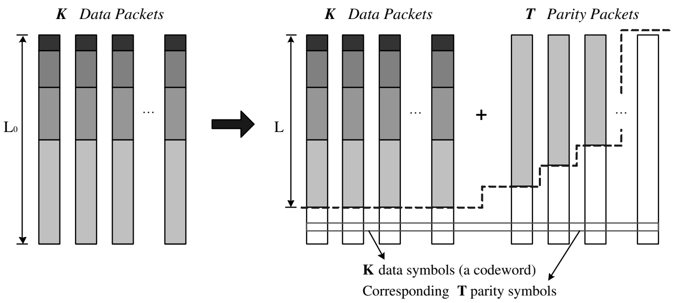


Fig. 2. The Error-Resilient Unequal Protection scheme (ER-UEP).

B. System-Level Description

Fig. 2 shows the principle diagram of our ER-UEP method. It is important to notice that bits in each data packet are processed on the symbol-by-symbol basis. For example, the k^{th} data packet is composed of symbols $s_k(0), s_k(1), \ldots$, where $s_k(i)$ denotes the i^{th} data symbol of the k^{th} data packet. Next, all K data symbols at the same position within their individual data packets are grouped to form a codeword. In this way, all bits in K data packets are organized into a list of codewords, and the i^{th} codeword consists of data symbols $s_1(i)$, $s_2(i)$, ..., $s_K(i)$. Channel coding is finally applied to generate T parity symbols for each codeword via the Reed-Solomon code $RS(K+T, K)$[1]. The generated T parity symbols for the same codeword are then filled into T parity packets, one symbol for each packet. In the end, K data packets and T parity packets together form an FEC block.

According to the UEP principle, different number of parity symbols is desired for different codewords, given the rate budget R_E. This critical issue will be solved via an optimization procedure in the next subsection. As a result, some codewords and parity symbols have to be discarded (and indeed, they can, thanks to the fact that both data packets and parity packets are scalable) as indicated by the dashed line in Fig. 2. Note that after truncation some packets (especially parity packets) may be shorter than the allowed maximum packet length, which may reduce the transmission efficiency. The transmission efficiency can be improved by merging more small-sized packets from different FEC blocks into one packet, at the cost of some extra system delay. It can also be done in conjunction with the packet-level interleaving, which is usually applied in practice to alleviate burst packet losses.

C. Optimization Problem Statement and the Solution

Suppose that the distortion of the i^{th} codeword (which consists of K symbols at the same position of all K data packets) is $D(i)$, which can be computed through accumulating the distortion of each data packet:

$$D(i) = D_{BL} - \sum_{k=1}^{K} \Delta D_k(i) \tag{4}$$

[1]The maximum value of T equals to $2^m - 1 - K$, where m is the symbol length in bits. Normally, one byte is chosen as a symbol, thus the maximum T is $255 - K$.

where $\Delta D_k(i)$ is the distortion reduction if the first i data symbol of the k^{th} data packet is decoded. Moreover, as $\Delta D_k(i)$ in each data packet is decreasing, the corresponding codewords importance is also in a decreasing fashion.

Clearly, one way to quantify the loss protection performance is to measure the probability that any data symbol out of k data symbols protected by t parity symbols is lost, denoted as $p_e(k,t)$. This function can be either obtained in the transmission system or calculated through some mathematical approaches [8]. Because of the complete dependency cancellation achieved by our ER-UEP scheme, the expected distortion can be calculated based on $p_e(k,t)$:

$$\mathcal{E}\{D\} = D_{BL} - \sum_{i=1}^{L} (1 - p_e(K, t_i)) \times (D(i-1) - D(i)) \tag{5}$$

where L is the number of selected data symbols for each data packet after the optimization, assuming that all K data packets are uniformly truncated to meet the rate budget R_E that is allocated to the current time-slot

$$\sum_{i=1}^{L} (K + t_i) \times m \leq R_E, \tag{6}$$

where m is the symbol length in bits.

Now, the optimization problem is formulated as follows: given the number of data packets K (each data packet has L_0 symbols), the R-D function $D(i)$ (i stands for the i^{th} codeword), and the loss-recovery performance function $p_e(k,t)$, find the most important codewords and the number of their corresponding parity symbols, that minimizes $\mathcal{E}\{D\}$ subject to the rate constraint given in Eqn. (6).

Lagrangian principle is applied to solve this optimization problem. For the i^{th} codeword, two vectors representing the protection efficiency (slope) and the corresponding rate can be obtained as follows:

$$S_i = [s(i,t)]_{t=0,\cdots,T} \qquad R_i = [r(i,t)]_{t=0,\cdots,T} \tag{7}$$

where

$$\begin{aligned} r(i,t) &= (K + t) \times m \\ s(i,t) &= (D(i-1) - D(i)) \cdot \frac{p_e(K, t-1) - p_e(K,t)}{r(i,t) - r(i,t-1)}. \end{aligned} \tag{8}$$

In Eqn. (8), we define $p_e(K, -1) = 1$ and $r(i, -1) = 0$ for completeness. Moreover, S_i can be represented as a projection of distortion reduction by the i^{th} codeword and a common vector V of length $T + 1$:

$$S_i = (D(i-1) - D(i)) \cdot \begin{bmatrix} v(0) & v(1) & \ldots & v(T) \end{bmatrix} \tag{9}$$

where

$$v(t) = \frac{p_e(K, t-1) - p_e(K,t)}{r(i,t) - r(i,t-1)}. \tag{10}$$

Notice that the Lagrangian optimization principle requires a monotonically decreasing sequence, which means that the slope vector S_i (or equivalently, the common vector V) should be strictly monotonous. However, this strict monotonicity of S_i or V cannot be guaranteed in general. To make V convex, a post-processing stage is required for merging those non-decreasing coefficients in V.

After post-processing, we obtain a strictly monotonous common vector V^* of length $\hat{T}+1$:

$$V^* = \left[v^*(t(0)) \quad v^*(t(1)) \quad ... \quad v^*(t(\hat{T}))\right] \tag{11}$$

where

$$v^*(t(j)) = \frac{p_e(K,t(j-1)) - p_e(K,t(j))}{r(i,t(j)) - r(i,t(j-1))}, \tag{12}$$

and $t(j)$ is the corresponding protection strength of the j^{th} element in V^*. Next, the strictly monotonous slope matrix S^* and the corresponding rate matrix R^* can be easily obtained from V^*, each of size $L_0 \times (\hat{T}+1)$:

$$S^* = \begin{bmatrix} S_1^* \\ \vdots \\ S_{L_0}^* \end{bmatrix} = \begin{bmatrix} s^*(1,t(0)) & \cdots & s^*(1,t(\hat{T})) \\ \vdots & \ddots & \vdots \\ s^*(L_0,t(0)) & \cdots & s^*(L_0,t(\hat{T})) \end{bmatrix} \tag{13}$$

$$R^* = \begin{bmatrix} R_1^* \\ \vdots \\ R_{L_0}^* \end{bmatrix} = \begin{bmatrix} r^*(1,t(0)) & \cdots & r^*(1,t(\hat{T})) \\ \vdots & \ddots & \vdots \\ r^*(L_0,t(0)) & \cdots & r^*(L_0,t(\hat{T})) \end{bmatrix} \tag{14}$$

where

$$\begin{aligned} r^*(i,t(j)) &= r(i,t(j)) = (K+t(j)) \times m \\ s^*(i,t(j)) &= (D(i-1) - D(i)) \times v^*(t(j)). \end{aligned} \tag{15}$$

Finally, the solution is very simple: by using some efficient searching algorithms such as bisection, the optimal solution can be found iteratively through looking for the best protection strength $t_i = t(j_i^\dagger)$ for the i^{th} codeword that satisfies $s^*(i,t(j_i^\dagger+1)) < \lambda \leq s^*(i,t(j_i^\dagger))$, with constraint of total rate budget for that time-slot:

$$\sum_{i=1}^{L} r^*(i,t(j_i^\dagger)) \leq R_E. \tag{16}$$

where λ is a given Lagrangian multiplier; and L is the maximum i that satisfies $s^*(i,t(0)) \geq \lambda$ (or in other words, it is the number of selected codewords).

After the optimization, we only select the first L data symbols for each data packet, whereas throwing away data symbols from position $L+1$ to L_0. This can be done easily since each data packet is scalable. Similarly, the parity packets are also selected and truncated according to the determined optimal protection strength t_i.

IV. Experimental Results

We tested our ER-UEP scheme against various packet loss cases extensively to simulate FGS video transmission over the Internet. The MPEG-4 standard test sequence *Coastguard* in CIF format and 10 Hz are used in our experiments. Only the first frame is encoded as **I** frame and all others as **P** frames. The video source is encoded with the PFGS [2] coder to generate 96 kbps base-layer bitstream and maximum $5,000$ kbps enhancement-layer bitstream. We assume that the base layer is transmitted without any loss.

To simulate the bandwidth fluctuation in the Internet, the total available enhancement-layer rate is assumed to be uniformly distributed within the range of $[512, 1024]$ kbps for each 1-second time-slot. Meanwhile, to simulate the burst loss in the Internet, a two-state Gilbert model, characterized by the global packet loss rate (PLR) and the average burst length (ABL), is used in our experiments. Furthermore, in order to evaluate the performance and robustness of our ER-UEP scheme under degraded channel conditions, the enhancement-layer bitstream is first protected at two Gilbert models with different (PLR, ABL): $(0.01, 1.5)$, $(0.05, 2.0)$, and then transmitted over channel with a varying PLR (over a wide range) but the fixed ABL (as given in the two models selected above). Finally, to randomize the burst packet loss, packets from two FEC blocks are interleaved before being transmitted out.

To demonstrate the effectiveness of our proposed scheme, the MD-FEC method [5] mentioned before is chosen as the benchmark for comparison. Note that to improve error resilience of both the MD-FEC scheme and the scheme without error protection, a 23-bits resynchronization marker followed by 9-bits MB address information is inserted at the MB boundary for any bits interval greater than 1000 bits.

We first evaluate the quality of the proposed scheme under a wide range of degraded channel conditions. As shown in Fig. 3 (all the numbers are averaged over 1000 experiments), the proposed ER-UEP scheme achieves better performance and higher robustness as compared with the MD-FEC scheme, especially when the channel seriously degrades. This is because in our ER-UEP scheme any received data bits can be decoded, whereas this cannot be guaranteed in MD-FEC. Meanwhile, as our ER-UEP scheme can achieve rather good error resilience, neither resynchronization marker nor MB address information is needed, which also saves quite a lot of redundant bits.

We then evaluate the PSNR values of individual frames on channels with prediction errors. This kind of channel is simulated by adding a Gaussian noise on the PLR of the Gilbert loss process. That is, for the predicted PLR on which the loss protection is based, the actual packet loss rate equals $PLR+w$, where w is an additive Gaussian noise (updated every time-slot) with zero mean and $\sigma^2 \times PLR^2$ variation (σ equals to 0.2 in our experiments). It can be seen from Fig. 4 that the MD-FEC scheme improves the quality of the No-EP scheme a lot, but still leaves the robustness unacceptable (i.e., very poor quality has been observed at some frames). On the contrary, our ER-UEP scheme provides quite robust quality for its strong error resilience.

V. Conclusions and Future Work

In this paper, we studied the optimal loss protection issue for streaming FGS video bitstreams over IP. A very general performance metric for quantifying the streaming quality was first defined and analyzed. This metric revealed that the quality degradation in conventional UEP schemes mainly comes from bitstream contamination. We then proposed a new error-resilient unequal protection (ER-UEP) method. Very good results have been achieved by our proposed scheme.

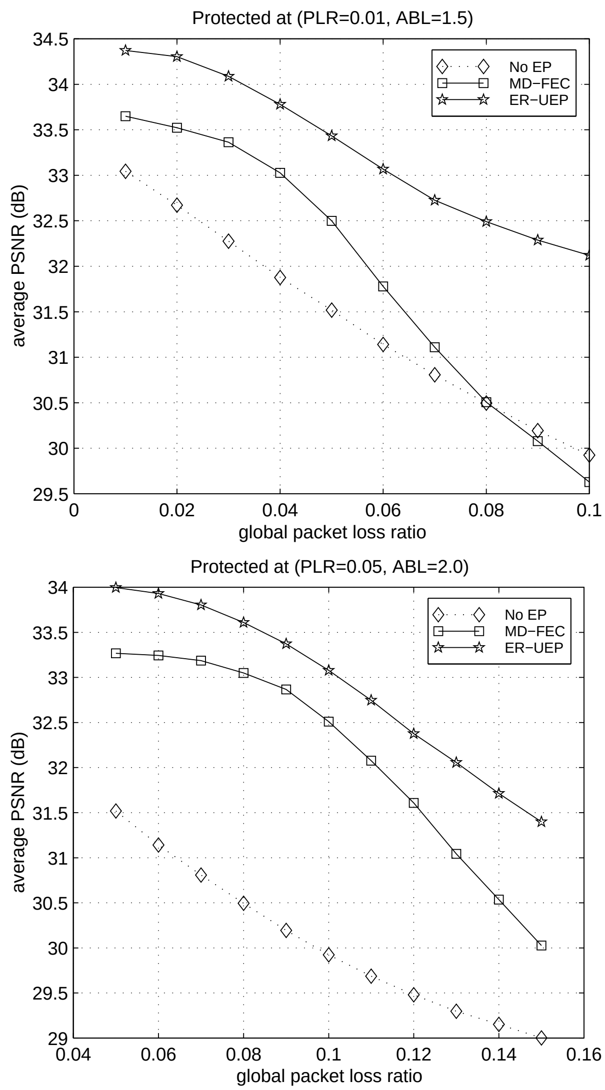


Fig. 3. Comparative evaluation of the proposed scheme at different packet loss ratios.

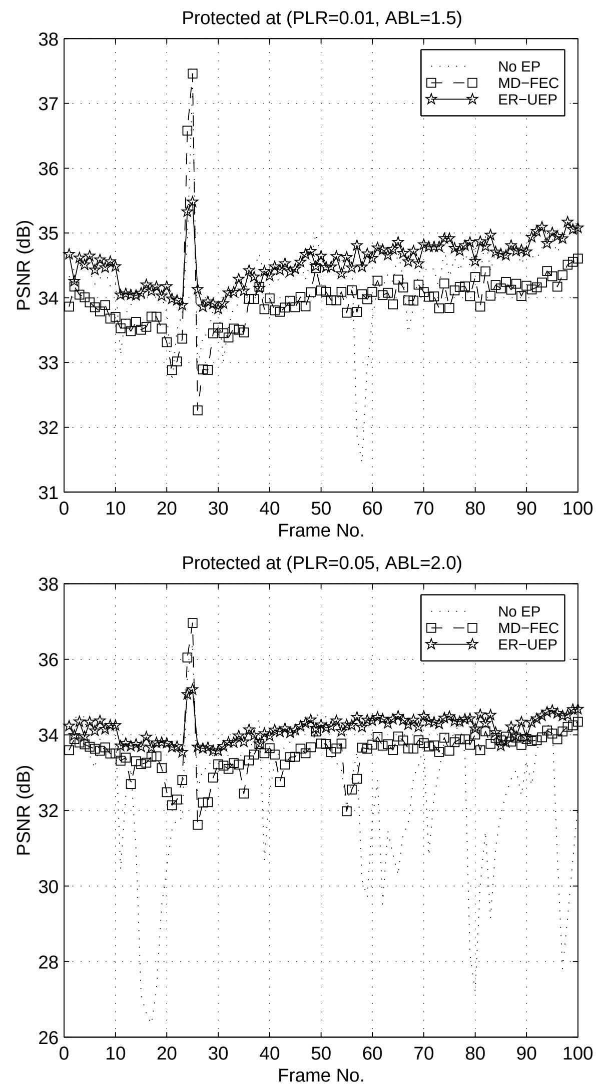


Fig. 4. Comparative evaluation on channels with prediction errors. The enhancement-layer rate is set to 768 kbps.

Besides FGS and PFGS, the proposed method in principle also works for other scalable image/video bitstreams such as those produced by JPEG-2000 [9], etc. Moreover, we believe that the unequal protection and error-resilience concept could give remarkable quality improvements for wireless videos, which is getting more and more interests recently. This is one focus in our future works.

ACKNOWLEDGMENT

The authors would like to thank Dr. Feng Wu from Microsoft Research Asia for many helpful discussions on the implementation of the proposed protection scheme for FGS video bitstreams.

REFERENCES

[1] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[2] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for Efficient Progressive Fine Granularity Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 332–344, Mar. 2001.

[3] E. Biersack, "Performance evaluation of forward error correction in an ATM environment," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 631–640, May 1993.

[4] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Transactions on Information Theory*, vol. 42, pp. 1737–1744, Nov. 1996.

[5] R. Puri, K.-W. Lee, K. Ramchandran, and V. Bharghavan, "An integrated source transcoding and congestion control paradigm for video streaming in the Internet," *IEEE Transactions on Multimedia*, vol. 3, no. 1, pp. 18–32, Mar. 2001.

[6] V. Paxson, "Measurements and analysis of end-to-end internet traffic," in *Ph.D. dissertation*, Univ. Calif. Berkeley, Feb. 1997.

[7] H. Cai, G. Shen, Z. Xiong, S. Li, and B. Zeng, "An optimal packetization scheme for fine granularity scalable bitstream," *ISCAS 2002*, vol. 5, pp. 641–644, May 2002.

[8] P. Frossard, "FEC performance in multimedia streaming," *IEEE Communications Letters*, vol. 5, no. 3, pp. 122–124, Mar. 2001.

[9] ISO/IEC 15444-1, "Information technology–JPEG2000–Image coding system–Part 1: Core coding system," 2000.