

PET Protection Optimization for Streaming Scalable Videos With Multiple Transmissions

Ruiqin Xiong, *Member, IEEE*, David S. Taubman, *Senior Member, IEEE*, and Vijay Sivaraman, *Member, IEEE*

Abstract—This paper investigates priority encoding transmission (PET) protection for streaming scalably compressed video streams over erasure channels, for the scenarios where a small number of retransmissions are allowed. In principle, the optimal protection depends not only on the importance of each stream element, but also on the expected channel behavior. By formulating a collection of hypotheses concerning its own behavior in future transmissions, limited-retransmission PET (LR-PET) effectively constructs channel codes spanning multiple transmission slots and thus offers better protection efficiency than the original PET. As the number of transmission opportunities increases, the optimization for LR-PET becomes very challenging because the number of hypothetical retransmission paths increases exponentially. As a key contribution, this paper develops a method to derive the effective recovery-probability versus redundancy-rate characteristic for the LR-PET procedure with any number of transmission opportunities. This significantly accelerates the protection assignment procedure in the original LR-PET with only two transmissions, and also makes a quick and optimal protection assignment feasible for scenarios where more transmissions are possible. This paper also gives a concrete proof to the redundancy embedding property of the channel codes formed by LR-PET, which allows for a decoupled optimization for sequentially dependent source elements with convex utility-length characteristic. This essentially justifies the source-independent construction of the protection convex hull for LR-PET.

Index Terms—Error correction, hybrid-ARQ, priority encoding transmission (PET), feedback, retransmission, scalable video.

I. INTRODUCTION

THIS paper addresses the robust transmission of scalably compressed video streams through lossy communication channels, for application scenarios where a small number of retransmissions is allowed. The channel is that of a packet-based “erasure channel,” in which each packet either arrives

intact or is entirely lost due to transmission errors or excessive delays. Traditionally, the possibility of packet losses is handled by *forward error correction* (FEC) or *automatic repeat request* (ARQ) retransmission.

FEC approaches are usually advocated for the transmission of real-time data, with the assumption that retransmission of lost data might cause excessive delays. In principle, unequal protection should be assigned to the stream elements based on their importance or utility. A well known FEC framework for unequal protection is the *priority encoding transmission* (PET) scheme of Albanese *et al.* [1]. Since then, unequal protection has been studied extensively [2]–[11], along with scalable coding methods. Typically, the video is compressed into a scalable stream with a number of quality layers, assigning stronger protection to the more important quality layers.

Retransmission of lost data may indeed not be possible in some cases. In other applications, however, the delivery time constraints allow for a limited number of retransmissions. This perspective has been adopted by a considerable body of literature [12]–[17]. The combination of both protection and retransmission has also been considered in a variety of settings [18]–[24]. In a general framework combining FEC and retransmission, there are two fundamental issues to consider. The first issue is what protection and re-protection scheme to use, since this sets up the recovery-probability and the redundancy-rate characteristics associated with each level of protection that can be offered to each quality layer in a single transmission or retransmission. The second issue is how to optimize the allocation of protection (or bandwidth) among the transmission opportunities, while keeping the potential for future transmissions in mind.

As to the first issue, we adopt the PET framework. The advantage of employing PET in this context has been studied, initially by Gan and Ma [24], and later by Taubman and Thie [25], [26]. In *Limited-Retransmission PET* (LR-PET) [25], [26], each source frame has a limited number of transmission opportunities, in each of which packets are protected by the PET coding scheme, using maximum distance separable (MDS) codes of a fixed length. At each feedback opportunity, the transmitter receives information about which packets are missing, based on which it can opt to retransmit some of the lost packet data, using PET for protection again. The important observation is that retransmission can take advantage of the packets (and hence codeword elements) that have already arrived, which strongly affects the effective redundancy-rate characteristic for retransmitted content. It is worth noting that there are other ways to achieve essentially the same behavior.

Manuscript received September 12, 2012; revised February 23, 2013 and June 7, 2013; accepted June 21, 2013. Date of publication July 9, 2013; date of current version September 12, 2013. This work was supported in part by the Australian Research Council Discovery under Grant DP0774447, the National Basic Research Program of China 973 Program under Grant 2009CB320904, the National Natural Science Foundation of China under Grants 61073083 and 61121002, the Beijing Natural Science Foundation under Grants 4112026 and 4132039, and the Research Fund for the Doctoral Program of Higher Education under Grants 20100001120027 and 20120001110090. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nikolaos V. Boulgouris. (*Corresponding author: R. Xiong.*)

R. Xiong is with the School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: rxiong@pku.edu.cn; rxiong@gmail.com).

D. Taubman and V. Sivaraman are with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney NSW 2052, Australia.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2013.2272516

One possibility is to use rateless codes for the PET encoding. One benefit of the rateless approach is that the transmitter need only know the number of lost packets (or the number of packets still required by the decoder) as opposed to the specific identities of these packets. This can also be achieved by using very long (much longer than needed) MDS codes (see Section II for more details). The main advantage of rateless codes is reduced complexity, in exchange for which there will be a small loss in coding efficiency. However, the design of the coding scheme itself is not the focus of this paper.

As to the second issue, greedy approaches have been proposed before, in various settings. A greedy PET protection approach based on MDS codes was proposed by Gan and Ma [24]. In LR-PET [25], [26], however, the behavior of future retransmission of lost data is hypothesized and formulated in the optimization objective. This effectively constructs channel codes which extend into future transmission slots. Of course, a longer code may offer better protection efficiency when properly optimized. Importantly, this extended channel code is progressively determined based on the feedback from earlier transmissions, so that it can adapt to dynamic channel conditions automatically. Nevertheless, the protection assigned at any specific transmission opportunity is based on hypotheses regarding future retransmissions, subject to the best available estimates of the future channel behavior. Indeed, the consideration of hypothetical retransmissions provides the major part of the performance gain of LR-PET over plain PET [26].

Unfortunately, the consideration of hypotheses also makes the optimization problem much more complicated, since it involves the joint optimization of a large number of hypothetical redundancy indices. Most notably, when the number of retransmission opportunities increases, the number of possible hypothetical retransmission paths increases exponentially, while the complexity of the existing optimization algorithm increases at an even greater rate. For this reason, the original work on LR-PET [25] considered only one retransmission to be possible.

Obviously, the restriction to only one retransmission may impair the efficiency of LR-PET protection in scenarios where a second or even more retransmissions is actually allowed by the delivery time constraint. The most obvious benefit of extra retransmission opportunities is the chance to recover the packets lost in previous transmissions. A more significant benefit is that part of the protection may be moved from an earlier transmission to future ones, which avoids over-protection and waste of bandwidth in the earlier transmission slots. This is especially useful for rapidly varying channels, for which the actual future channel behavior is quite uncertain to the streaming server and is difficult to forecast reliably.

The present paper proposes a general LR-PET optimization solution, which can efficiently find the optimal protection for each element in a scalable source stream, with any number of transmissions. The central achievement of this paper is to show that the recovery probability versus redundancy rate (P - R) characteristic and its convex hull, for the LR-PET procedure with any number of transmissions,

can be efficiently calculated beforehand so that the ultimate protection assignment procedure can be simplified to a quick search on an appropriate protection convex hull. Importantly, we demonstrate that the effective protection convex hulls can be constructed recursively; and our simulation results suggest that the complexity increases only approximately linearly with the number of transmissions, even though the number of hypothetical retransmission paths increases exponentially.

The second important contribution of this paper relates to the partitioning of the optimization problem for an entire scalably compressed data set into a collection of independent optimization problems, one for each element in the data source. The data source is considered to consist of a sequence of frames (these could be video frames, audio frames, groups of pictures, etc.), each of which is abstracted as a collection of compressed elements, with sequential decoding dependencies. In the PET context, this abstraction turns out to be sufficient to cover a broad range of scalably compressed data sources, including those which involve tree-structured dependencies [27]. The original work on LR-PET [25] provided a proof that the optimization can be carried out independently on each source element, but only in the limit as the channel codes become very long, so that a continuous approximation of the P - R characteristic can be used. In this paper, we provide a concrete proof of the redundancy embedding property required for independent optimization of the source elements, which is valid for channel codes of any length, with any number of retransmission opportunities.

While this paper considers multiple retransmission opportunities, it is also possible to extend LR-PET in other ways. In another work (initially presented in [28] and extended in [29]), we consider both packet loss and packet delay, both for each transmitted PET packet and for each acknowledgment message. The presence of stochastic packet delay processes can result in incomplete acknowledgment information at the streaming server, at the point when retransmission must be scheduled. The retransmission encoding scheme and optimization algorithm in this delay-sensitive LR-PET extension are more complicated. However, the development in this paper forms an important foundation for that work and potentially other extensions to the original LR-PET algorithm.

Many other techniques [30]–[35] have also been proposed in recent years to handle the challenges in predicting channel behavior. An important idea among these works is to utilize rateless (or fountain) codes, such as the LT codes [36] and Raptor codes [37], to cope with channel uncertainties. In Ahmad *et al.* [30], the sender continuously emits packets (possibly with temporary pauses to wait for feedback) until the receiver acknowledges that it has collected enough data. In such scheme, the sender optimizes its sending rate and waiting period to minimize bandwidth usage while ensuring successful decoding. Ahmad *et al.* [31] and Arslan *et al.* [35] also investigated the use of rateless codes for unequal error protection. As noted earlier, rateless codes have complexity advantages over the MDS (e.g., Reed-Solomon) codes considered in this paper. In fact, rateless codes can be used in our framework. Strictly speaking, however, the results developed in this paper

rely upon the MDS property, which is only approximately satisfied by rateless codes.

The rest of this paper is organized as follows. The PET and LR-PET schemes are reviewed in Section II. The optimization procedure and the method to calculate the effective P - R characteristic for LR-PET are described in Section III. Section IV proves the redundancy embedding property, which justifies the independent optimization of the protection assignment for sequentially dependent elements. Section V describes the implementation of P - R convex hull construction and briefly analyzes the computational complexity. Experimental results are reported in Section VI.

II. PET AND LR-PET

A. Background and Notations

For this paper, we model any scalable video as a sequence of independently compressed “source frames” $\mathcal{F}[n]$, $n = 1, 2, 3, \dots$, each of which consists of a collection of embedded elements $\mathcal{E}_q[n]$, $q = 1, 2, \dots, Q$, having lengths $L_q[n]$ and utilities $U_q[n]$, which progressively augment the quality of reconstructed $\mathcal{F}[n]$. In the simplest case, each source frame corresponds to one video frame, scalably compressed by a scalable image coder such as EZW [38], SPIHT [39] or JPEG2000 [40]. In more general cases, each so-called “source frame” may actually represent a group of video frames, jointly compressed by a scalable video coder, such as [41]–[46], to exploit the temporal correlation. Furthermore, we assume that the elements in each source frame exhibit: (i) a sequential dependency, $\mathcal{E}_1[n] < \mathcal{E}_2[n] < \dots < \mathcal{E}_Q[n]$, which means that an element $\mathcal{E}_q[n]$ in a frame cannot be decoded without obtaining all the preceding elements $\mathcal{E}_p[n]$, $p = 1, 2, \dots, q-1$ in that frame; (ii) a convex source utility-length characteristic, $\frac{U_1[n]}{L_1[n]} \geq \frac{U_2[n]}{L_2[n]} \geq \dots \geq \frac{U_Q[n]}{L_Q[n]}$, which means that the importance of the elements $\mathcal{E}_q[n]$ decreases with the index q in each $\mathcal{F}[n]$.

The transmission of the above stream is arranged in a series of “transmission slots” $\mathcal{T}[n]$, $n = 1, 2, 3, \dots$, in each of which a fixed number of packets are sent according to the budget of transmission bandwidth. To be specific, the sender performs the primary transmission of $\mathcal{F}[n]$ in the slot $\mathcal{T}[n]$ and then waits for feedback from the receiver. To deal with packet losses, extra transmissions for $\mathcal{F}[n]$ may be scheduled in the subsequent slots $\mathcal{T}[n+i]$, $i = 1, 2, \dots$, depending on the channel outcomes in transmissions. Therefore, the packets sent in the slot $\mathcal{T}[n]$ contain the primary data from the frame $\mathcal{F}[n]$, along with any retransmitted content from the previous frames $\mathcal{F}[n-i]$, $i = 1, 2, \dots$. Note that both the primary transmission and the retransmissions can be protected by FEC codes. Since each source frame can be processed independently, the frame index n will be omitted in the remainder of this paper when it brings no confusion.

The channel of interest in this paper is erasure channels, in which each packet either arrives intact or is lost entirely. To specify the channel loss characteristics for a transmission slot, we use the distribution of packet loss number instead of the average packet loss rate. This allows us to reveal the advantage of the LR-PET framework in the context of fast-varying

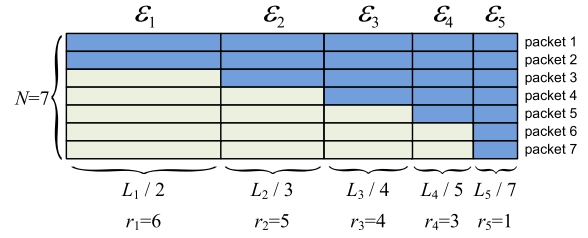


Fig. 1. An example of PET frame with $N = 7$. The elements \mathcal{E}_1 through \mathcal{E}_5 , having lengths L_1 through L_5 , are protected by (N, k) channel codes with $k = 2, 3, 4, 5$ and 7 , respectively. The dark and light shaded boxes correspond to source and parity symbols, respectively.

channels, in which the long-term channel behavior is known as a packet loss distribution (collected from feedback messages) but the instantaneous packet loss rate is highly uncertain. For this purpose, we introduce notation $\rho = (\rho_0, \rho_1, \dots, \rho_N)$, where ρ_k denotes the probability that k packets are received while other $N - k$ packets are lost in a transmission slot that sends N packets.

B. PET Transmission

In the PET [1] procedure, source elements are protected by a family of (N, k) codes, all of which have the same codeword length N but may have different source lengths k , as illustrated in Fig. 1. We only consider *maximum distance separable* (MDS) codes (e.g. Reed-Solomon codes), which have the key property that the receipt of any k coded symbols in an (N, k) codeword is sufficient to recover the k source symbols.

For convenience, we use redundancy index (r-index) r instead of k to indicate the amount of redundancy used in transmission. The value $r = 0$ is reserved for the special case that an element is not transmitted at all. For other integer r values, e.g. $r = 1, 2, \dots, N$, the $(N, N + 1 - r)$ channel code is used. Therefore, the function

$$k_{\min}(r) = \begin{cases} N + 1 - r, & r > 0 \\ +\infty, & r = 0 \end{cases} \quad (1)$$

indicates the minimum number of packets that a receiver must acquire in order to recover the content protected using r-index r . Note that a higher r value means a higher redundancy or stronger protection in transmission and $r = 1$ corresponds to the case that an element is transmitted without adding any redundancy. Once the r-index r_q of each element \mathcal{E}_q has been determined by an optimization procedure, the PET scheme packages the encoded elements into N network packets, which we call a “PET frame,” as illustrated in Fig. 1. For elements \mathcal{E}_q assigned with $r_q > 0$, $L_q/k_{\min}(r_q)$ source symbols are placed in each of the first $k_{\min}(r_q)$ packets, while each of the remaining $N - k_{\min}(r_q)$ packets in the PET frame contains $L_q/k_{\min}(r_q)$ parity symbols. To address the sequential dependency, the redundancy indices in PET is always assigned in such a way that $r_1 \geq r_2 \geq \dots \geq r_Q$. Fig. 1 shows an example of PET frame consisting of $N = 7$ packets.

During transmission, network packets may become lost and source symbols cannot be recovered when too many packet

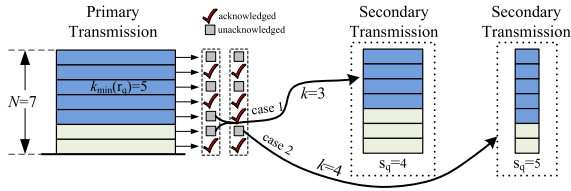


Fig. 2. An example of PET encoding for the primary transmission and the secondary transmission. In this case, the sender knows exactly which packets were lost in previous transmission. The symbols in $k_{\min}(r_q) - k$ out of the $N - k$ lost packets are regarded as a new element to send in the secondary transmission slot, protected by a new redundancy index s_q in the PET frame. Two hypothetical cases with $k = 3$ and $k = 4$ are illustrated here.

losses occur. We define a *utility realization* (UR) function $\varphi(r, k)$ as the utility realized for an element protected with r -index r , when k packets arrive at the receiver. In plain PET, each element has only one opportunity to be transmitted so that the UR function is

$$\varphi(r, k) = \begin{cases} 1, & k \geq k_{\min}(r) \\ 0, & k < k_{\min}(r). \end{cases} \quad (2)$$

Similarly, we define a *transmission rate* (TR) function $\psi(r, k)$ as the ratio of the encoded transmission length to the original uncoded length of an element, under the same conditions. The TR function for plain PET is

$$\psi(r, k) = \frac{N}{k_{\min}(r)} = \begin{cases} \frac{N}{N+1-r}, & r > 0 \\ 0, & r = 0. \end{cases} \quad (3)$$

C. LR-PET Transmission

Our interest in PET lies primarily in its extension to the more interesting cases of LR-PET [25], [26] with one or more retransmission opportunities. For an element \mathcal{E}_q to recover from a transmission slot in which \mathcal{E}_q is protected with r -index r_q , the number of packets which must be received from that slot is $k_{\min}(r_q)$. When the number of received packets k is no less than $k_{\min}(r_q)$, no retransmission is necessary for \mathcal{E}_q . When $k < k_{\min}(r_q)$ and retransmission is possible, the sender needs to resend the symbols of \mathcal{E}_q from $k_{\min}(r_q) - k$ of the $N - k$ lost packets. These symbols as a whole are regarded as a new source element and encoded into the PET frame of the retransmission slot using a new r -index. The PET encoding for retransmission is illustrated in Fig. 2. In this way, the effective uncoded length of \mathcal{E}_q is reduced from L_q to $L_q\theta(r_q, k)$, where the function $\theta(r, k)$ represents the fraction to resend and is defined by

$$\theta(r, k) = \begin{cases} 1 - \frac{k}{k_{\min}(r)}, & k < k_{\min}(r) \\ 0, & k \geq k_{\min}(r). \end{cases} \quad (4)$$

Note that the scheme in Fig. 2 assumes that the transmitter receives information about which packets were lost. In fact, this requirement can be relaxed to that the transmitter knows how many packets were lost, as opposed to the specific identities of these packets, by employing a slightly different PET scheme which uses a (N^+, k) MDS code (with $N^+ \gg N$) to generate N^+ packets but only transmits the first N (out

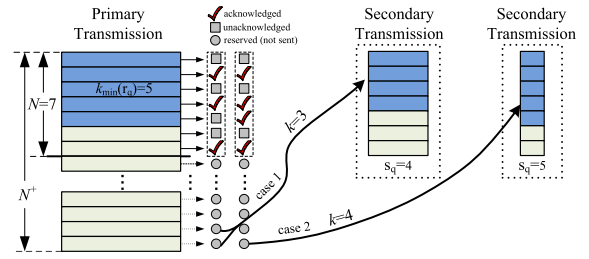


Fig. 3. An example of PET encoding for the primary transmission and the secondary transmission. (N^+, k) MDS codes (with $N^+ \gg N$) are used for PET encoding. In this case, the sender need only know the number of lost packets in previous transmission. The symbols in $k_{\min}(r_q) - k$ out of the $N^+ - N$ reserved packets are regarded as a new element to send in the secondary transmission slot, protected by a new redundancy index s_q in the PET frame. Two hypothetical cases with $k = 3$ and $k = 4$ are illustrated here.

of the N^+) packets in its primary transmission, leaving the other $N^+ - N$ packets for possible “retransmission”¹ in future transmission slots (see [28], [29], [47] for more details). This scheme is illustrated in Fig. 3. In both schemes, the length of data to “retransmit” is the same, although the data content is different. Therefore, the above “from L_q to $L_q\theta(r_q, k)$ ” statement on the effective uncoded length is still valid.

If the receiver still fails to receive sufficient packets for the successful decoding of an element, further retransmission may be performed, so long as the delivery time constraint allow it. As the number of performed transmissions increases, the total transmitted rate and the expected utility increase while the effective uncoded length decreases. For a general LR-PET communication process consisting of n transmissions, we use $\mathbf{r}_q = (r_q^1, r_q^2, \dots, r_q^n)$ to denote the redundancy indices assigned to element \mathcal{E}_q , $\mathbf{k} = (k_1, k_2, \dots, k_n)$ to denote the number of received packets, and $\Upsilon = (\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \dots, \boldsymbol{\rho}^n)$, $\boldsymbol{\rho}^i = (\rho_0^i, \rho_1^i, \dots, \rho_N^i)$ to denote the channel characteristics, in the n transmission slots, respectively. Then, with meanings similar to the ones for the plain PET case, we define $\varphi^n(r^1, k_1, \dots, r^n, k_n)$, $\psi^n(r^1, k_1, \dots, r^n, k_n)$ and $\theta^n(r^1, k_1, \dots, r^n, k_n)$ for the LR-PET procedure with n transmissions. Specially, we have $\varphi^1(r^1, k_1) = \varphi(r^1, k_1)$, $\psi^1(r^1, k_1) = \psi(r^1, k_1)$ and $\theta^1(r^1, k_1) = \theta(r^1, k_1)$. Note that the ensuing development applies uniformly to all elements \mathcal{E}_q , so we drop the explicit dependence of the r_q^i terms on q . According to the LR-PET transmission mechanism, we can easily derive the following recursive formulations:

$$\begin{aligned} \varphi^n(r^1, k_1, \dots, r^n, k_n) &= \begin{cases} 1, & k_1 \geq k_{\min}(r^1) \\ \varphi^{n-1}(r^2, k_2, \dots, r^n, k_n), & k_1 < k_{\min}(r^1), \end{cases} \end{aligned} \quad (5)$$

$$\begin{aligned} \psi^n(r^1, k_1, \dots, r^n, k_n) &= \psi^1(r^1, k_1) + \theta(r^1, k_1) \\ &\quad \cdot \psi^{n-1}(r^2, k_2, \dots, r^n, k_n), \end{aligned} \quad (6)$$

$$\theta^n(r^1, k_1, \dots, r^n, k_n) = \theta(r^1, k_1) \cdot \theta^{n-1}(r^2, k_2, \dots, r^n, k_n). \quad (7)$$

¹It is not really retransmission in this case, because the data sent in the retransmission slots to cope with packet losses is not created from any of the data sent in previous slots. It may be more accurate to call it “supplementary transmission”.

III. OPTIMAL PROTECTION ASSIGNMENT

A. Plain PET

For a single PET-protected transmission over channels with a packet loss distribution ρ , the functions

$$P_\rho(r) = \sum_{k=0}^N \rho_k \cdot \varphi(r, k) \quad (8)$$

and

$$R_\rho(r) = \sum_{k=0}^N \rho_k \cdot \psi(r, k) \quad (9)$$

formulate the recovery probability and the redundancy rate, respectively, of an element protected by redundancy r . Under the constraint $r_1 \geq r_2 \geq \dots \geq r_Q$, the expected total utility at the receiver is

$$U(\{r_q\}) = \sum_q U_q \cdot P_\rho(r_q) \quad (10)$$

and the used total transmission length is

$$L(\{r_q\}) = \sum_q L_q \cdot R_\rho(r_q). \quad (11)$$

The goal of the protection assignment procedure is to find a set of r-indices $\{r_q\}$, $q = 1, 2, \dots$, which maximizes $U(\{r_q\})$ subject to the length constraint $L(\{r_q\}) \leq L_{\max}$.

This length-constrained optimization problem can be converted to a family of unconstrained optimization problems, parameterized by a quantity $\lambda > 0$. We find a set of r-indices $\{r_q(\lambda)\}$ which maximize the Lagrangian-style function

$$J(\lambda) = U(\{r_q(\lambda)\}) - \lambda L(\{r_q(\lambda)\}) \\ = \sum_q \left\{ U_q \cdot P_\rho(r_q(\lambda)) - \lambda L_q \cdot R_\rho(r_q(\lambda)) \right\} \quad (12)$$

subject to the constraint $r_1(\lambda) \geq r_2(\lambda) \geq \dots \geq r_Q(\lambda)$. Under the assumption of convex source utility-length characteristics (i.e. $\frac{U_1}{L_1} \geq \frac{U_2}{L_2} \geq \dots \geq \frac{U_Q}{L_Q}$), the optimal $\{r_q(\lambda)\}$ can be obtained by separately maximizing

$$J_q(\lambda) = U_q \cdot P_\rho(r_q(\lambda)) - \lambda L_q \cdot R_\rho(r_q(\lambda)), \quad (13)$$

or equivalently

$$J'_q(\lambda) = P_\rho(r_q(\lambda)) - \lambda_q \cdot R_\rho(r_q(\lambda)), \quad (14)$$

for each element \mathcal{E}_q , $q = 1, \dots, Q$, without violating the redundancy constraint (see [25], [26] for the proof). Here,

$$\lambda_q = \lambda \cdot L_q / U_q. \quad (15)$$

For any given λ , the optimal solution to (14),

$$r_q(\lambda) = \operatorname{argmax}_r \left[P_\rho(r) - \lambda_q \cdot R_\rho(r) \right], \quad (16)$$

can be found by searching the upper convex hull of the $P_\rho(r)$ versus $R_\rho(r)$ characteristic for the point whose slope is closest to but no less than $\lambda_q = \lambda \cdot L_q / U_q$. With any packet loss characteristic ρ learned by collecting statistics from acknowledgment messages, the convex hull of the $P_\rho(r)$ versus $R_\rho(r)$ characteristic for plain PET can be pre-calculated from (8) and (9) so that the FEC code assignment procedure is very simple and fast. Fig. 4 illustrate the P-R convex hulls for PET transmission over IID channels (i.e. the loss probability of each packet is independent and identically distributed).

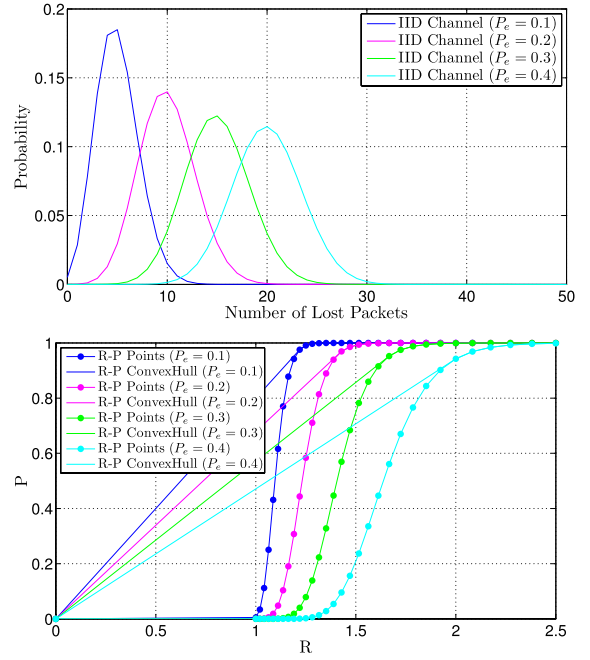


Fig. 4. Example P-R convex hulls for the PET transmission (with $N = 50$). Top: packet loss distribution of IID channels with different packet loss rate P_e . Bottom: the corresponding P-R convex hulls for PET transmission.

B. LR-PET With a Single Retransmission ($n = 2$)

In the following, we consider the protection assignment in general LR-PET procedure with n transmission opportunities. At the first of these opportunities, we need to determine only the primary redundancy index r_q^1 to use for each element \mathcal{E}_q . However, we still need to consider the effect of possible future retransmissions. The principle is that when feedback channel and retransmission opportunity exist, the sender may avoid using too much redundancy in its initial transmission opportunity and wait for the feedback channel to tell whether extra redundancy is really necessary. Of course, the optimal choice of redundancy index r_q^i for the i^{th} transmission depends on the protection indices assigned in the earlier transmission slots, identified by $\mathbf{r}_q^{1:i-1} = (r_q^1, \dots, r_q^{i-1})$, as well as the channel outcomes in these slots, identified by $\mathbf{k}_{1:i-1} = (k_1, \dots, k_{i-1})$, because they determine the successfully conveyed fraction and thus the effective uncoded length of \mathcal{E}_q at the time to start the i^{th} transmission. Importantly, at the time when we assign r_q^1 , the channel behavior identified by $\mathbf{k}_{1:i-1}$ can only be hypothesized, which leads to a collection of hypothetical retransmission redundancy indices $r_q^{i, \mathbf{k}_{1:i-1}}$ for each $i > 1$.

To begin, we make the discussion simple by limiting our attention to the case $n = 2$. For a specific element \mathcal{E}_q and a given λ value, the optimization objective is to find the optimal r_q^1 which maximizes

$$J_q^2(\lambda_q) = \sum_{\mathbf{k} \in \{0, N\}^2} \rho_{k_1}^1 \rho_{k_2}^2 \cdot \varphi^2(r_q^1, k_1, r_q^{2, k_1}, k_2) \\ - \lambda_q \cdot \sum_{\mathbf{k} \in \{0, N\}^2} \rho_{k_1}^1 \rho_{k_2}^2 \cdot \psi^2(r_q^1, k_1, r_q^{2, k_1}, k_2), \quad (17)$$

while for each hypothesis k_1 , corresponding to the number of packets which may be received from the first transmission slot, a separate hypothetical retransmission redundancy index r_q^{2,k_1} is determined by the optimization procedure to maximize (17). The λ_q here, formulated by (15), is the element-specific Lagrangian parameter, incorporating the element's utility-length characteristic. The optimization problem (17) involves $1 + (N + 1)$ free variables, i.e. r_q^1 and $\{r_q^{2,k_1}\}_{k_1}$, $k_1 \in [0, N]$, with up to $(N + 1)$ choices each.

In practice, the search can be considerably simplified by observing that each hypothetical redundancy index r_q^{2,k_1} can be obtained independently of the others, as a modified PET search which depends only on the selected value of r_q^1 and the hypothesis k_1 . To see this, observe that the objective (17) can be rewritten using (5) and (7) as

$$J_q^2(\lambda_q) = \left(P_{\rho^1}(r_q^1) + \sum_{k_1 \in \mathbb{K}(r_q^1)} \rho_{k_1}^1 \cdot P_{\rho^2}(r_q^{2,k_1}) \right) - \lambda_q \cdot \left(R_{\rho^1}(r_q^1) + \sum_{k_1 \in \mathbb{K}(r_q^1)} \rho_{k_1}^1 \cdot \theta(r_q^1, k_1) \cdot R_{\rho^2}(r_q^{2,k_1}) \right) \quad (18)$$

Here, the notation $\mathbb{K}(r) = [0, k_{\min}(r) - 1]$ is used to identify the set of hypotheses k_1 for which retransmission of an element protected with primary redundancy index r may be beneficial. For a selected r_q^1 value, and for each hypothesis $k_1 \in \mathbb{K}(r_q^1)$, maximizing (18) is reduced to maximizing

$$J'_q(\lambda_q) = P_{\rho^2}(r_q^{2,k_1}) - \lambda_q \theta(r_q^1, k_1) \cdot R_{\rho^2}(r_q^{2,k_1}). \quad (19)$$

The solution, $r_q^{2,k_1} = \arg\max_r [P_{\rho^2}(r) - \lambda_q \theta(r_q^1, k_1) \cdot R_{\rho^2}(r)]$, can be quickly obtained from the convex hull of the $P_{\rho^2}(r)$ versus $R_{\rho^2}(r)$ characteristic, using a slope threshold $\lambda_q \theta(r_q^1, k_1)$. This convex hull can be easily constructed, based on the packet loss distribution ρ^2 , in the same way as what we do to solve the optimization problem (16) for plain PET.

Evidently, this observation reduces the set of possibilities which must be examined to at most $(N+1)^3$. The optimization procedure developed in [25], [26] follows this idea, but it still has the following weakness: 1) it explicitly evaluates the ‘‘intermediate’’ hypothetical indices $\{r_q^{2,k_1}\}_{k_1}$, even though only the optimal choice of r_q^1 value is needed; and 2) the computation-intensive evaluation for the full set of hypothetical indices $\{r_q^{2,k_1}\}_{k_1}$ is performed repeatedly for every candidate value of r_q^1 , every element \mathcal{E}_q and every λ value tested to make $L^{(\lambda)}$ approach L_{\max} sufficiently closely.

A key observation in the present paper is that the complete space of potentially optimal combinations of r_q^1 and $\{r_q^{2,k_1}\}_{k_1}$ can be explored by using a separate Lagrangian parameter $\tilde{\lambda}$ to characterize the choice of the $\{r_q^{2,k_1}\}_{k_1}$ terms. In what follows, we will show how this observation allows us to create an effective convex hull which can be used to select the optimal r_q^1 directly without explicitly calculating the ‘‘intermediate’’ hypothetical indices $\{r_q^{2,k_1}\}_{k_1}$ each time.

For the convenience of discussion, we describe the behavior of a protected communication process by a set of triplet $T = (P, R, \mathbf{V})$, each representing one possible case of protection

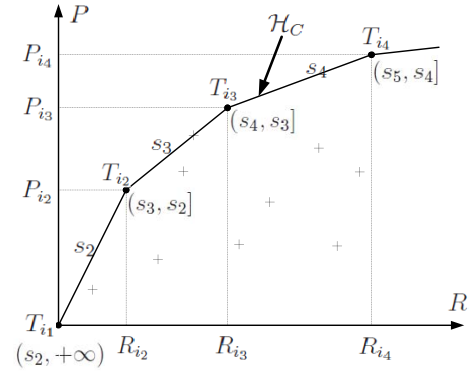


Fig. 5. The convex hull \mathcal{H}_C of an effective channel code C . The points marked with ‘‘+’’ indicate the elements in C but not in \mathcal{H}_C . The interval labeled under each T_{ij} is the range of λ values for which V_{ij} is the optimal protection assignment solution.

assignment. In each triplet, \mathbf{V} denotes a set of parameters to specify the protection; P and R represent the recovery probability and redundancy rate associated with it, respectively. In the simplest case, i.e. $n = 1$, \mathbf{V} consists of the single r-index r (i.e. r^1). For $n = 2$, \mathbf{V} includes the primary r-index r^1 and the hypothetical r-indices r^{2,k_1} for each k_1 . However, we are ultimately interested only in the value of the primary r-index r^1 . In all cases, a practical implementation of our proposed algorithm only needs to explicitly store the value of r^1 in \mathbf{V} for each T .

Among all the protection assignment possibilities, we are only interested in the cases which achieve an optimal trade off between the recovery probability P and the redundancy rate R . Therefore, we define $C = \{T_i\}_i$ to include all possible $T_i = (P_i, R_i, \mathbf{V}_i)$ (listed in the increasing order of R_i) and define a subset of C , $\mathcal{H}_C \subseteq C$, in such a way that $\{(R_i, P_i) | T_i \in \mathcal{H}_C\}$ forms the complete set of vertices on the upper convex hull of the coordinates $\{(R_i, P_i) | T_i \in C\}$. For this reason, we also call \mathcal{H}_C the ‘‘convex hull’’ of C . To be precise, a $T_i \in C$ belongs to \mathcal{H}_C if and only if there exists $0 < a < b$, s.t.

$$P_i - \lambda \cdot R_i > P_j - \lambda \cdot R_j, \quad \forall j \neq i, \quad \text{whenever } \lambda \in (a, b). \quad (20)$$

We call the maximal interval (a, b) satisfying (20) as the ‘‘optimal λ interval’’ for T_i . Fig. 5 illustrates the P-R characteristic for both \mathcal{H}_C and C . Suppose the j^{th} point in \mathcal{H}_C is T_{ij} . Obviously, the P-R slope values defined for the points in \mathcal{H}_C by

$$s_j = \begin{cases} \frac{P_{ij} - P_{i_{j-1}}}{R_{ij} - R_{i_{j-1}}}, & j > 1 \\ +\infty, & j = 1 \end{cases} \quad (21)$$

satisfy $s_1 > s_2 > \dots > 0$ and the optimal λ interval of T_{ij} is (s_{j+1}, s_j) . Since we are mainly interested in the elements in \mathcal{H}_C instead of C , we notationally define $P_{\mathcal{H}_C}(j) = P_{ij}$, $R_{\mathcal{H}_C}(j) = R_{ij}$, $r_{\mathcal{H}_C}(j) = r_{ij}$, $S_{\mathcal{H}_C}(j) = s_j$ and $\mathbb{S}^1(\mathcal{H}_C) = \bigcup_i \{S_{\mathcal{H}_C}(i)\}$. The subscript \mathcal{H}_C here is added to specify the relevant convex hull. We further define the following operator,

$$\mathfrak{S}(\mathcal{H}_C, \lambda) = \max \{i \mid S_{\mathcal{H}_C}(i) \geq \lambda\}, \quad (22)$$

to search \mathcal{H}_C for the protection assignment which is optimal for a given λ value. We then have:

Property 1: For all \mathcal{H}_C , $\mathfrak{S}(\mathcal{H}_C, \lambda)$ is a non-increasing piecewise-constant function of λ . In particular, $\mathfrak{S}(\mathcal{H}_C, \lambda) = i$ for all $\lambda \in (\mathcal{S}_{\mathcal{H}_C}(i+1), \mathcal{S}_{\mathcal{H}_C}(i)]$. These intervals are illustrated in Fig. 5. Note that in the pathological cases $\lambda = \mathcal{S}_{\mathcal{H}_C}(i)$, redundancy indices $r_{\mathcal{H}_C}(j)$ and $r_{\mathcal{H}_C}(j-1)$ both yield optimal solution, but our convention always selects $r_{\mathcal{H}_C}(j)$.

With this notation, the optimal hypothetical indices $\{r_q^{2,k_1}\}_{k_1}$ can be expressed as

$$r_q^{2,k_1}(\lambda_q) = r_{\mathcal{H}_C^1(\rho^2)}(\mathfrak{S}(\mathcal{H}_C^1(\rho^2), \lambda_q \cdot \theta(r_q^1, k_1))). \quad (23)$$

Here $\mathcal{H}_C^1(\rho)$ represents the *first order* protection convex hull, for a single transmission with packet loss distribution ρ . It is formed from the triplet set $\{(P_\rho(r), R_\rho(r), r)\}$ specified by (8) and (9) over all possible r .

Now we consider the optimal choice of primary index r_q^1 in (18). We introduce two functions,

$$P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda}) = P_{\rho^1}(r) + \sum_{k \in \mathbb{K}(r)} \rho_k^1 P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \tilde{\lambda}\theta(r, k))), \quad (24)$$

$$R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda}) = R_{\rho^1}(r) + \sum_{k \in \mathbb{K}(r)} \rho_k^1 \theta(r, k) R_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \tilde{\lambda}\theta(r, k))), \quad (25)$$

(with $\tilde{\mathcal{H}} \triangleq \mathcal{H}_C^1(\rho^2)$ for notational convenience), so that according to (23) the optimization problem (18) can be rewritten as

$$J_q^2(\lambda_q) = P_{\Upsilon_{1:2}}^2(r_q^1, \lambda_q) - \lambda_q R_{\Upsilon_{1:2}}^2(r_q^1, \lambda_q). \quad (26)$$

Here $P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ and $R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ are the expected recovery utility and transmission rate for the combined $n=2$ transmissions, when r is used in the first transmission and the second transmission chooses its protection using $\tilde{\lambda}$ as its Lagrangian parameter.

In (24) and (25), the $P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ versus $R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ characteristic includes a set of scaled P - R characteristics for the secondary transmission, one for each hypothesis $k \in \mathbb{K}(r)$. The benefit of using $\tilde{\lambda}$ in (24) and (25) to characterize the choice of $\{r_q^{2,k_1}\}_{k_1}$ is that it allows us to avoid explicitly evaluating the r_q^{2,k_1} values for each \mathcal{E}_q , k and λ during the final optimization procedure. Although the original optimization problem involves the choice $\tilde{\lambda} = \lambda_q$, as indicated by (26), it turns out that we can free up the parameter $\tilde{\lambda}$, forming an extended optimization procedure in which the objective is to maximize

$$J_q^2(\lambda_q) = P_{\Upsilon_{1:2}}^2(r_q^1, \tilde{\lambda}) - \lambda_q R_{\Upsilon_{1:2}}^2(r_q^1, \tilde{\lambda}) \quad (27)$$

over all possible combinations of r_q^1 and $\tilde{\lambda}$. As we will show, this extended optimization problem turns out to be equivalent to the original problem, in the sense that they both yield the same optimal values for r_q^1 ; indeed, the optimal value for $\tilde{\lambda}$ turns out to be essentially identical to λ_q .

First of all, the following property can be easily obtained from (24), (25) and Property 1.

Property 2: For any fixed r , $P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ and $R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ are non-increasing piecewise-constant functions of $\tilde{\lambda}$. That is, if we collect the slopes from the scaled convex hulls associated with each hypothesis into a set $\mathbb{S}^2(\tilde{\mathcal{H}}, r) = \bigcup_{k \in \mathbb{K}(r)} (\mathbb{S}^1(\tilde{\mathcal{H}}/\theta(r, k)))$, enumerating its elements in decreasing order as $s'_1 > s'_2 > \dots > s'_{m'}$, then $\mathfrak{S}(\tilde{\mathcal{H}}, \tilde{\lambda}\theta(r, k))$

returns the same index for all $\tilde{\lambda} \in (s'_{i+1}, s'_i]$, regardless of the hypothesis k . Thus $P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ and $R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ are constant over the interval $(s'_{i+1}, s'_i]$.

Therefore, we can restrict the parameter $\tilde{\lambda}$ in (24) and (25) to the discrete set $\mathbb{S}^2(\tilde{\mathcal{H}}, r)$. For any fixed r , there are only $m' = |\mathbb{S}^2(\tilde{\mathcal{H}}, r)|$ discrete points on the $P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ versus $R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ curve, with the i^{th} point generated by setting $\tilde{\lambda} = s'_i$. In Appendix I, we prove the following property:

Property 3: The P - R slope at the i^{th} point on the $P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ versus $R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ curve, defined by

$$\tilde{\mathcal{S}}_r^2(i) = \begin{cases} \frac{P_{\Upsilon_{1:2}}^2(r, s'_i) - P_{\Upsilon_{1:2}}^2(r, s'_{i-1})}{R_{\Upsilon_{1:2}}^2(r, s'_i) - R_{\Upsilon_{1:2}}^2(r, s'_{i-1})}, & 1 < i \leq m' \\ +\infty, & i = 1, \end{cases}$$

satisfies $\tilde{\mathcal{S}}_r^2(i) = s'_i$.

Now we can prove the following key observation in this paper:

Corollary 1: The optimization objectives (26) and (27) yield the same optimal choice for r_q^1 .

Proof: For any candidate value of r_q^1 , the $P_{\Upsilon_{1:2}}^2(r_q^1, \tilde{\lambda})$ versus $R_{\Upsilon_{1:2}}^2(r_q^1, \tilde{\lambda})$ characteristic is convex, since $\{s'_i\}$ strictly decreases with i . Therefore, a potentially optimal $\tilde{\lambda}$ which maximizes (27) must satisfy

$$\tilde{\lambda} = s'_v, v = \max \{i \mid s'_i \geq \lambda_q, s'_i \in \mathbb{S}^2(\tilde{\mathcal{H}}, r_q^1)\}. \quad (28)$$

This means $\lambda_q \in (s'_{v+1}, s'_v = \tilde{\lambda}]$. For any λ_q in this interval, we have $P_{\Upsilon_{1:2}}^2(r_q^1, \lambda_q) = P_{\Upsilon_{1:2}}^2(r_q^1, \tilde{\lambda})$ and $R_{\Upsilon_{1:2}}^2(r_q^1, \lambda_q) = R_{\Upsilon_{1:2}}^2(r_q^1, \tilde{\lambda})$ – c.f. Property 2. In this case, the objectives (26) and (27) become equal. Therefore they must yield the same optimal choice for r_q^1 . ■

Now we conclude that the solution to the original optimization objective (18) is identical to the solution to (27). The latter one can be found by directly searching the convex hull formed from the points $P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ and $R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda})$ generated by all combinations of r and $\tilde{\lambda}$. That is

$$r_q^1(\lambda_q) = r_{\mathcal{H}_C^2(\Upsilon_{1:2})}(\mathfrak{S}(\mathcal{H}_C^2(\Upsilon_{1:2}), \lambda_q)), \quad (29)$$

where $\mathcal{H}_C^2(\Upsilon_{1:2})$ is the *second order* protection convex hull for LR-PET with two transmissions, formed from the set of triplets $\{(P_{\Upsilon_{1:2}}^2(r, \tilde{\lambda}), R_{\Upsilon_{1:2}}^2(r, \tilde{\lambda}), (r, \tilde{\lambda}))\}$ according to (24) and (25).

C. LR-PET With Multiple Retransmissions ($n > 2$)

We can now generalize the above results to the case of $n > 2$ transmission opportunities. As before, the key challenge is to find r_q^1 which maximizes

$$J_q^n(\lambda_q) = \sum_{\mathbf{k} \in [0, N]^n} \rho_{k_1}^1 \cdots \rho_{k_n}^n \cdot \varphi^n(r_q^1, k_1, \dots, r_q^{n, \mathbf{k}_{1:n-1}}, k_n) - \lambda_q \cdot \sum_{\mathbf{k} \in [0, N]^n} \rho_{k_1}^1 \cdots \rho_{k_n}^n \cdot \psi^n(r_q^1, k_1, \dots, r_q^{n, \mathbf{k}_{1:n-1}}, k_n), \quad (30)$$

where for each $i > 1$ and each possible hypothesis $\mathbf{k}_{1:i-1}$, a separate hypothetical redundancy index $r_q^{i, \mathbf{k}_{1:i-1}}$ is determined by the optimization procedure to maximize (30). Based on (5) and (7), the objective (30) can be rewritten as (31), shown at

$$J_q^n(\lambda_q) = \left(P_{\rho^1}(r_q^1) + \sum_{k_1 \in \mathbb{K}(r_q^1)} \rho_{k_1}^1 \cdot \left\{ \sum_{\mathbf{k}_{2:n} \in [0, N]^{n-1}} \rho_{k_2}^2 \cdots \rho_{k_n}^n \cdot \varphi^n(r_q^{2,k_1}, k_2, \dots, r_q^{n, k_{1:n-1}}, k_n) \right\} \right) - \lambda_q \cdot \left(R_{\rho^1}(r_q^1) + \sum_{k_1 \in \mathbb{K}(r_q^1)} \rho_{k_1}^1 \theta(r_q^1, k_1) \cdot \left\{ \sum_{\mathbf{k}_{2:n} \in [0, N]^{n-1}} \rho_{k_2}^2 \cdots \rho_{k_n}^n \cdot \psi^n(r_q^{2,k_1}, k_2, \dots, r_q^{n, k_{1:n-1}}, k_n) \right\} \right). \quad (31)$$

the top of the next page. When r_q^1 and k_1 are fixed, optimizing the 1st through the n^{th} transmissions is reduced to optimizing the 2nd through the n^{th} transmissions, with the element's uncoded length multiplied by $\theta(r_q^1, k_1)$. Equivalently, we can change the element-specific Lagrangian parameter from λ_q to $\lambda_q \theta(r_q^1, k_1)$. Suppose the convex hull $\mathcal{H}_C^{n-1}(\Upsilon_{2:n})$ for $n-1$ transmissions (i.e. the 2nd through the n^{th} transmissions) with arbitrary packet loss characteristic $\Upsilon_{2:n}$ has been constructed. With the optimal assignment for $r_q^{2, k_{1:1}}, r_q^{3, k_{1:2}}, \dots, r_q^{n, k_{1:n-1}}$, the expected recovery probability and transmission rate for the 2nd through the n^{th} transmissions must be $P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda_q \theta(r_q^1, k_1)))$ and $R_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda_q \theta(r_q^1, k_1)))$ (with $\tilde{\mathcal{H}} \triangleq \mathcal{H}_C^{n-1}(\Upsilon_{2:n})$ for notational convenience), respectively. Therefore, by defining

$$P_{\Upsilon_{1:n}}^n(r, \tilde{\lambda}) = P_{\rho^1}(r) + \sum_{k \in \mathbb{K}(r)} \rho_k^1 P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \tilde{\lambda} \theta(r, k))), \quad (32)$$

$$R_{\Upsilon_{1:n}}^n(r, \tilde{\lambda}) = R_{\rho^1}(r) + \sum_{k \in \mathbb{K}(r)} \rho_k^1 \theta(r, k) R_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \tilde{\lambda} \theta(r, k))), \quad (33)$$

the problem (31) can be rewritten as:

$$J_q^n(\lambda_q) = P_{\Upsilon_{1:n}}^n(r_q^1, \lambda_q) - \lambda_q R_{\Upsilon_{1:n}}^n(r_q^1, \lambda_q). \quad (34)$$

The problem set (32), (33) and (34) has the same form as (24), (25) and (26), with the only difference that $\tilde{\mathcal{H}}$ changed from $\mathcal{H}_C^1(\Upsilon_{2:2})$ to $\mathcal{H}_C^{n-1}(\Upsilon_{2:n})$. Actually, the development in Section III-B applies to arbitrary $\tilde{\mathcal{H}}$. Therefore, all the conclusions for the case $n=2$ can be directly extended to the cases of $n > 2$. It follows that the protection convex hull $\mathcal{H}_C^n(\Upsilon_{1:n})$ can be constructed recursively from $\mathcal{H}_C^{n-1}(\Upsilon_{2:n})$ by considering all combinations of r_q^1 and $\tilde{\lambda} \in \mathbb{S}^2(\mathcal{H}_C^{n-1}(\Upsilon_{2:n}), r_q^1)$.

IV. REDUNDANCY EMBEDDING PROPERTY

The solution described in Section III assumes that the protection for each element \mathcal{E}_q can be optimized independent of the others. However, since the elements in each scalably compressed source frame have a sequential dependencies, to ensure the validity of the additive utility formulation (10) in the optimization objectives, the following redundancy constraints must be satisfied:

$$\begin{aligned} r_q^1 &\geq r_{q+1}^1, \quad \forall q, \\ r_q^{2, k_1} &\geq r_{q+1}^{2, k_1}, \quad \forall k_1, \forall q, \\ &\dots \\ r_q^{n, k_{1:n-1}} &\geq r_{q+1}^{n, k_{1:n-1}}, \quad \forall \mathbf{k}_{1:n-1}, \forall q. \end{aligned} \quad (35)$$

In this section, we show that this constraint is automatically satisfied for any source with convex utility-length characteristic. To see this, we first prove that the optimized LR-PET procedure exhibits *Redundancy Embedding Property* (see Theorem 1). Note that the optimality of the original LR-PET strategy (for $n=2$) was originally proved in the continuous case in [25] and later proved in the discrete case by S. Dumitrescu in [48]. This work, initially presented in [49], provides a different proof of the optimality while extending the approach to any number of retransmissions.

Theorem 1 (Redundancy Embedding Property): For any $n \geq 1$, the protection convex hull \mathcal{H}_C^n formed by LR-PET procedure with n transmissions always satisfies $r_{\mathcal{H}_C^n}(i) \geq r_{\mathcal{H}_C^n}(j)$, $\forall i > j$, i.e. the primary redundancy indices are ordered. We call this the redundancy embedding property of \mathcal{H}_C^n . In other words, among all the potentially optimal protection assignment candidates in \mathcal{H}_C^n , an assignment with a higher overall redundancy rate must have equal or stronger protection in its primary transmission.

Proof: For $n=1$, the result is an immediate consequence of our definition of redundancy rate – c.f. equations (9) and (3). We now consider $n > 1$, for which $\mathcal{H}_C^n(\Upsilon_{1:n})$ is constructed from the triplet set $\left\{ \left(P_{\Upsilon_{1:n}}^n(r, \tilde{\lambda}), R_{\Upsilon_{1:n}}^n(r, \tilde{\lambda}), (r, \tilde{\lambda}) \right) \right\}_{r, \tilde{\lambda}}$. Consider any two points on $\mathcal{H}_C^n(\Upsilon_{1:n})$ with indices $i_x < i_y$, which are generated with parameter pairs $(r_x, \tilde{\lambda}_x)$ and $(r_y, \tilde{\lambda}_y)$, respectively. According to the definition of \mathcal{H}_C , it must be possible to identify two disjoint, non-empty intervals (a_x, b_x) and (a_y, b_y) such that $(r_x, \tilde{\lambda}_x)$ maximizes

$$J_\lambda^n(r, \tilde{\lambda}) = P_{\Upsilon_{1:n}}^n(r, \tilde{\lambda}) - \lambda R_{\Upsilon_{1:n}}^n(r, \tilde{\lambda}) \quad (36)$$

for all $\lambda \in (a_x, b_x)$ and $(r_y, \tilde{\lambda}_y)$ maximizes $J_\lambda^n(r, \tilde{\lambda})$ for all $\lambda \in (a_y, b_y)$. Moreover, since $i_x < i_y$, and the convex hull has strictly decreasing slopes, $(a_x, b_x) > (a_y, b_y)$.

According to Corollary 1, however, any $\lambda \in (a_x, b_x)$ must also belong to one of the intervals (s'_{b+1}, s'_b) over which $\tilde{\lambda}_x$ can range without affecting the optimality of the choice $(r_x, \tilde{\lambda}_x)$. Accordingly, for all $\lambda \in (a_x, b_x)$, we can set $\tilde{\lambda}_x = \lambda$ and Corollary 1 tells us that r_x maximizes

$$J_\lambda(r) = P_{\Upsilon_{1:n}}^n(r, \lambda) - \lambda R_{\Upsilon_{1:n}}^n(r, \lambda). \quad (37)$$

Equivalently, r_x maximizes

$$\tilde{J}_\lambda(r) = (1/\lambda) \cdot P_{\Upsilon_{1:n}}^n(r, \lambda) - R_{\Upsilon_{1:n}}^n(r, \lambda). \quad (38)$$

for all $\lambda \in (a_x, b_x)$. Similarly, r_y maximizes $\tilde{J}_\lambda(r)$ for all $\lambda \in (a_y, b_y)$. It is useful to define

$$\tilde{D}_{r_1, r_2}(\lambda) = \tilde{J}_\lambda(r_1) - \tilde{J}_\lambda(r_2). \quad (39)$$

Then, for any $\lambda_x \in (a_x, b_x)$ and any $\lambda_y \in (a_y, b_y)$, we have $\lambda_x > \lambda_y$ and

$$\tilde{D}_{r_x, r_y}(\lambda_x) > 0, \quad \tilde{D}_{r_x, r_y}(\lambda_y) < 0, \quad (40)$$

which leads to $\tilde{D}_{r_x, r_y}(\lambda_x) > \tilde{D}_{r_x, r_y}(\lambda_y)$. In Appendix I, we prove that, $\tilde{D}_{r_1, r_2}(\lambda)$ is a non-increasing function of λ if $r_1 > r_2$. It follows that $r_x \leq r_y$. ■

Corollary 2: If the source exhibits a convex utility-length characteristic (i.e. $\frac{U_1}{L_1} \geq \frac{U_2}{L_2} \geq \dots \geq \frac{U_Q}{L_Q}$), the optimal redundancy indices chosen by the optimization procedure satisfy the redundancy constraints (35) automatically.

Proof: Property 1 tells us $\mathfrak{S}(\mathcal{H}_C, \lambda)$ is non-increasing with λ . From Theorem 1, we have

$$r_{\mathcal{H}_C^n}(\mathfrak{S}(\mathcal{H}_C^n, \lambda_1)) \geq r_{\mathcal{H}_C^n}(\mathfrak{S}(\mathcal{H}_C^n, \lambda_2)), \quad \forall \lambda_1 < \lambda_2, \forall n. \quad (41)$$

Since $\frac{U_q}{L_q} \geq \frac{U_{q+1}}{L_{q+1}}$, we must have $\lambda_q \leq \lambda_{q+1}, \forall q$ - c.f. equation (15). Applying (41) to $\mathcal{H}_C^n(\Upsilon_{1:n})$, we get $r_q^1 \geq r_{q+1}^1$. From (4) we get $\theta(r_q^1, k_1) \leq \theta(r_{q+1}^1, k_1), \forall k_1$, and hence $\lambda_q \theta(r_q^1, k_1) \leq \lambda_{q+1} \theta(r_{q+1}^1, k_1), \forall k_1$. Applying (41) again, we get $r_q^{2, k_1} \geq r_{q+1}^{2, k_1}, \forall k_1$. Applying the same argument iteratively, it is clear that all the constraints in (35) are satisfied. ■

V. IMPLEMENTATION OF P - R CONVEX HULL CONSTRUCTION

Now we describe the implementation of P - R convex hull construction. When only one transmission is allowed (i.e. $n = 1$), the process is very simple. We just need to evaluate $P_\rho(r)$ and $R_\rho(r)$ for $r = 0, 1, \dots, N$ according to (8) and (9), producing an effective P - R curve for PET, and then remove the points not lying on the convex hull of this curve. This process was illustrated in Fig. 4. In this section, we consider general LR-PET (with $n \geq 2$) and focus on the recursive construction of the P - R convex hull \mathcal{H}_C^n from \mathcal{H}_C^{n-1} .

A. Exact Construction

As discussed in Section III, the P - R characteristic of LR-PET is determined not only by its primary transmission, but also by the hypothesized behavior in the 2nd through the n th transmissions. To be specific, we need to evaluate $P_{\Upsilon_{1:n}}^n(r, \tilde{\lambda})$ and $R_{\Upsilon_{1:n}}^n(r, \tilde{\lambda})$ according to (32) and (33), for all choices of r and $\tilde{\lambda}$, and then remove the points not lying on the convex hull of these points.

Suppose $\tilde{\mathcal{H}} = \mathcal{H}_C^{n-1}(\Upsilon_{2:n})$ is already constructed and we now generate the $P_{\Upsilon_{1:n}}^n(r, \tilde{\lambda})$ vs. $R_{\Upsilon_{1:n}}^n(r, \tilde{\lambda})$ curve for a fixed value of r . According to (32) and (33), this P - R characteristic includes a set of scaled version of the P - R characteristic for $\tilde{\mathcal{H}}$ (i.e. $\rho_k^1 P_{\tilde{\mathcal{H}}}(\ast)$ versus $\rho_k^1 \theta(r, k) R_{\tilde{\mathcal{H}}}(\ast)$, where $\ast = \mathfrak{S}(\tilde{\mathcal{H}}, \tilde{\lambda} \theta(r, k))$) goes through the indices of all points in $\tilde{\mathcal{H}}$ as $\tilde{\lambda}$ varies), one for each hypothesis $k \in \mathbb{K}(r)$. To produce the curve, we only need to check the critical values of $\tilde{\lambda}$ around which the value of $\ast = \mathfrak{S}(\tilde{\mathcal{H}}, \tilde{\lambda} \theta(r, k))$ changes for one $k \in \mathbb{K}(r)$. These $\tilde{\lambda}$ values correspond exactly to the slopes on these scaled P - R curves of $\tilde{\mathcal{H}}$. What we need to do is to create a list of all these slopes, sort them and record which hypothesis index k each of these slopes corresponds to. Based on this list, the whole $P_{\Upsilon_{1:n}}^n(r, \tilde{\lambda})$ versus $R_{\Upsilon_{1:n}}^n(r, \tilde{\lambda})$ curve can be easily calculated, without the

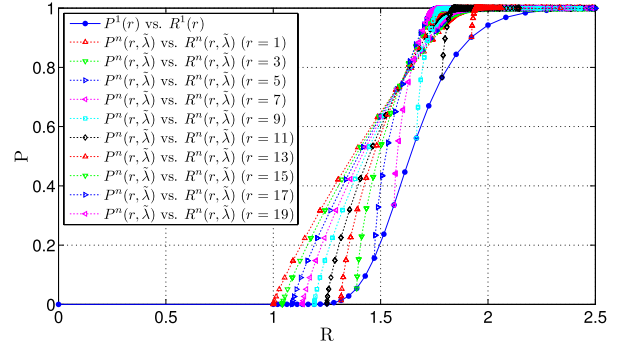


Fig. 6. Example of the $P^n(r, \tilde{\lambda})$ versus $R^n(r, \tilde{\lambda})$ curves (for $n = 2, r = 1, 3, 5, \dots$).

necessity to evaluate $\ast = \mathfrak{S}(\tilde{\mathcal{H}}, \tilde{\lambda} \theta(r, k))$ explicitly. Note that the slopes on each of these scaled P - R characteristic form a sub-list which is already sorted. Therefore, to create the whole list, we only need to merge the $\mathbb{K}(r)$ sub-lists, each having a length of $|\tilde{\mathcal{H}}|$. This can be done with a complexity of $|\mathbb{K}(r)| \cdot |\tilde{\mathcal{H}}| \cdot \log_2 |\mathbb{K}(r)|$ comparisons. To generate the $P_{\Upsilon_{1:n}}^n(r, \tilde{\lambda})$ vs. $R_{\Upsilon_{1:n}}^n(r, \tilde{\lambda})$ curve for all possible value of r , the total complexity is at most $\frac{1}{2} |\tilde{\mathcal{H}}| N^2 \log_2 N$ comparisons.

The complexity of above procedure can be further reduced if we taking the redundancy embedding property into consideration. Note that each r can maximize (37) or (38) only for a small range of λ values. In other words, for each $P^n(r, \tilde{\lambda})$ vs. $R^n(r, \tilde{\lambda})$ curve, only a short segment on it belongs to \mathcal{H}_C^n so that we do not need to evaluate the whole curve.

At first glance, we may think $|\mathcal{H}_C^n|$ would increase rapidly with the order n because of the large number of points on each $P^n(r, \tilde{\lambda})$ versus $R^n(r, \tilde{\lambda})$ curve. However, in our simulations, $|\mathcal{H}_C^n|$ is usually smaller than or close to $|\mathcal{H}_C^1|$, which is on the order of N , because most of the points on these curves turn out to be not lying on their convex hull. This empirical observation supports the conjecture that the complexity of LR-PET P - R convex hull construction grows roughly linearly with the number of transmissions, n .

Fig. 6 illustrates the $P^n(r, \tilde{\lambda})$ versus $R^n(r, \tilde{\lambda})$ characteristics obtained for several example r values, when constructing \mathcal{H}_C^2 from \mathcal{H}_C^1 . An IID channel with packet loss ratio $p_e = 0.4$ is considered here, where $N = 50$ is the number of packets in each PET frame. Comparing the $P^2(r, \tilde{\lambda})$ versus $R^2(r, \tilde{\lambda})$ curves to the $P^1(r)$ versus $R^1(r)$ curve, we note that extra recovery probability is obtained at the cost of extra transmission rate used by the secondary transmission. Importantly, the overall protection efficiency is improved, in the sense that the recovery probability P for \mathcal{H}_C^2 is always higher than P for \mathcal{H}_C^1 , subject to the same redundancy rate R .

B. Approximate Construction

One possible way to exploit the redundancy embedding property is the following procedure which exhibits much lower complexity, but still produces a very good approximation of \mathcal{H}_C^n with sufficiently fine granularity. The idea is to solve the optimal r directly from (37), for a sequence of λ values, $\lambda_1, \dots, \lambda_W$, which are densely sampled from the range

$[\lambda_{\min}, \lambda_{\max}]$ we are interested in, with step $\lambda_i/\lambda_{i+1} = \alpha$. For example, we may choose $\lambda_{\min} = 10^{-6}$, $\lambda_{\max} = 1$, $\alpha = \sqrt[4]{2}$, which leads to $W \approx 80$.

We start by solving (37) for $\lambda = \lambda_{i_1}$, $i_1 = W/2$. For each $r \in [0, N]$, to calculate $P^n(r, \lambda)$ and $R^n(r, \lambda)$, we need to find $s^{r,k} = \mathfrak{S}(\tilde{\mathcal{H}}, \lambda\theta(r, k))$ for each $k \in \mathbb{K}(r)$. Considering the fact that $\theta(r+1, k) \leq \theta(r, k)$ and hence $s^{r+1, k} \geq s^{r, k}$, $\forall r$, we can evaluate $s^{r,k}$, $r = 0, 1, \dots, N-k$ in sequence, which requires at most $(N-k+1)+|\tilde{\mathcal{H}}|-2$ comparisons. Therefore, solving (37) for λ_{i_1} requires at most $\sum_k ((N-k+1)+|\tilde{\mathcal{H}}|-2) \approx N^2/2 + N|\tilde{\mathcal{H}}|$ comparisons. We write the optimal solution for r as $r(\lambda_{i_1})$ and the optimal $s^{r,k}$ corresponding to the optimal r as $s^k(\lambda_{i_1})$. The above solution divides the search ranges of r and $s^{r,k}$ into two halves. For any $\lambda > \lambda_{i_1}$, the optimal indices must satisfy $r(\lambda) \leq r(\lambda_{i_1})$, hence $\theta(r(\lambda), k) \geq \theta(r(\lambda_{i_1}), k)$ and hence $s^{r,k}(\lambda) \leq s^k(\lambda_{i_1})$. In other words, for $\forall \lambda > \lambda_{i_1}$, we need to consider only $r \in [0, r(\lambda_{i_1})]$ and $s^{r,k} \in [0, s^k(\lambda_{i_1})]$. On the other hand, for $\forall \lambda < \lambda_{i_1}$, we need to consider only $r \in [r(\lambda_{i_1}), N]$ and $s^{r,k} \in [s^k(\lambda_{i_1}), |\tilde{\mathcal{H}}|]$.

We next solve (37) for $\lambda = \lambda_{i_2}, \lambda_{i_3}, i_2 = W/4, i_3 = 3W/4$. Obviously, we have $\lambda_{i_2} > \lambda_{i_1} > \lambda_{i_3}$. For each $k \in [0, N]$, we evaluate $s^{r,k}$, $r = 0, 1, \dots, r(\lambda_{i_2})$ for $\lambda = \lambda_{i_2}$ and $s^{r,k}$, $r = r(\lambda_{i_2}), r(\lambda_{i_2})+1, \dots, N-k$ for $\lambda = \lambda_{i_3}$ in sequence, which has a complexity of at most $(N-k+2)+|\tilde{\mathcal{H}}|-2$ comparisons. Therefore, solving (37) for λ_{i_2} and λ_{i_3} requires at most $\sum_k ((N-k+2)+|\tilde{\mathcal{H}}|-2) \approx N^2/2 + N|\tilde{\mathcal{H}}|$ comparisons. We proceed by solving (37) for $\lambda_{i_4}, \lambda_{i_5}, \lambda_{i_6}, \lambda_{i_7}, i_4 = W/8, i_5 = 3W/8, i_6 = 5W/8, i_7 = 7W/8$. Similarly, the total number of comparisons required to evaluate $s^{r,k}$ for the four λ values is at most $N^2/2 + N|\tilde{\mathcal{H}}|$. Continuing in this way, we can solve (37) for all $\lambda_1, \dots, \lambda_W$, with a complexity of at most $(N^2/2 + N|\tilde{\mathcal{H}}|) \log_2 W$ comparisons.

C. Complexity Comparison

Now we make a rough analysis on the computational complexity of the proposed protection optimization algorithms. Suppose M_λ is the number of λ values tried in an outer loop to satisfy the overall transmission length constraint. For each tested λ value, the original LR-PET optimization procedure (for $n = 2$) in [25] was shown to have a complexity on the order of $2N^2 \log_2(Q+1)$ to search for all the hypothetical redundancy indices $\{s_q^{k,r}\}$. The additional complexity to choose the actual primary redundancy index r is much lower than the above complexity. Therefore, the overall complexity per slot is on the order of

$$F_1 = 2M_\lambda N^2 \log_2(Q+1). \quad (42)$$

The optimization procedure proposed in this paper consists of two stages. In the first stage, the convex hulls \mathcal{H}_C^1 through \mathcal{H}_C^n are constructed. An important feature of this stage is that the construction of \mathcal{H}_C^1 through \mathcal{H}_C^n is independent of any specific elements, frames or λ values tried to achieve the target transmission rate and it needs to be constructed only when the channel statistics change. The complexity of this stage is on the order of

$$F_{2a} \approx (3N^2/2I_{\text{slot}}) \log_2 W. \quad (43)$$

Here I_{slot} is the interval, counted in number of slots, to update the convex hulls using the latest channel statistics. In the second stage, the redundancy index r_q is actually assigned for each element \mathcal{E}_q , by searching \mathcal{H}_C^m with an element-specific slope threshold λ_q , determined by λ , U_q and its current effective uncoded length, where m is the number of transmission opportunities still available for the frame to which \mathcal{E}_q belongs. This stage is equivalent to plain PET assignment in complexity. For each \mathcal{E}_q , searching for the optimal r_q from \mathcal{H}_C^m requires at most $|\mathcal{H}_C^m|$ comparisons. This bound can be easily reduced to $\log_2 |\mathcal{H}_C^m|$ by exploiting the monotonicity of slope values in \mathcal{H}_C^m , using a binary search for r_q . This leads to an overall complexity of $M_\lambda Q \log_2 |\mathcal{H}_C^m|$ per slot. The complexity may be further reduced by evaluating $r_q, q = 1, 2, \dots, Q$ in sequence, taking advantage of the fact that $r_1 \geq r_2 \geq \dots \geq r_Q$ will always be satisfied, whether we explicitly enforce it or not. This limits the search range for succeeding elements. At most $Q+|\mathcal{H}_C^m|-2$ comparisons are then required to determine all Q indices, leaving us an overall complexity of

$$F_{2b} = M_\lambda(Q+|\mathcal{H}_C^m|-2)$$

per slot, which is substantially lower than $M_\lambda Q |\mathcal{H}_C^m|$ and $M_\lambda Q \log_2 |\mathcal{H}_C^m|$ for large Q . Therefore, for LR-PET with two transmissions, the overall complexity for both convex hull construction and actual protection assignment is on the order of

$$F_2 = F_{2a} + F_{2b} \approx (3N^2/2I_{\text{slot}}) \log_2 W + M_\lambda(Q+N-2). \quad (44)$$

Both the $\log_2 W$ in (44) and M_λ in (42) reflect the granularity of λ we use, which should be on the same order.

In scalable coding schemes, each source frame can be compressed into an embedded stream consisting of a large number of elements, using block-wise bit-plane coding or similar techniques. In the extreme case, the embedded stream can be truncated and decoded at arbitrary position, which means that one element can be as short as a single byte. For practical purpose, however, it is usually enough to form only 20 ~ 100 quality layers for each frame, by merging several adjacent elements into a new larger element when necessary. In order to provide fine-granularity control on the FEC redundancy rate, the length N of channel codes for PET encoding should be reasonably large.² For example, to provide an increment step as low as 2% in the redundancy rate, we choose $N = 50$. For practical application settings (e.g. $Q = 20 \sim 100, N \gg 4$), the first term in (44) is much smaller than (42) (since $I_{\text{slot}} \log_2(Q+1) \gg 1$) and the second term in (44) is even more smaller (since $N^2 \gg Q/\log_2(Q+1)$ and $N^2 \log_2(Q+1) \gg N$). Therefore, the proposed LR-PET optimization procedure has a much smaller complexity than the original LR-PET procedure, even if the convex hulls must be reconstructed in each slot. Obviously, this complexity may be further reduced by constructing the convex hulls once a few slots.

²but not too large for Reed-Solomon codes, due to the complexity issue in decoding.

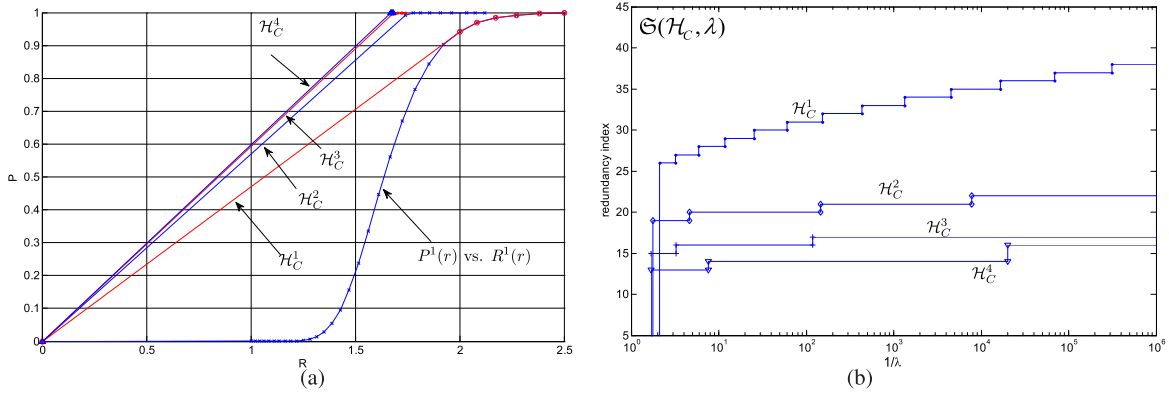


Fig. 7. Example of the effective channel codes for $n = 1, 2, 3, 4$ transmissions over IID channels with $p_e = 0.4$. (a) the P - R convex hull; (b) the optimal redundancy index corresponding to any specified λ .

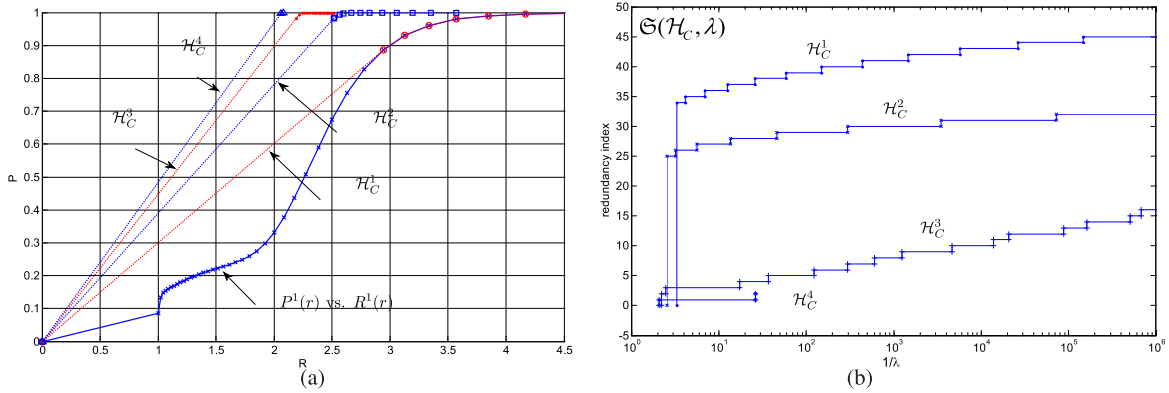


Fig. 8. Example of the effective channel codes for $n = 1, 2, 3, 4$ transmissions over a GE channel with $p_e = 0.33$, $M_{BAD} = 300$, $P_B = 0.6$, $P_G = 0.01$. (a) the P - R convex hull; (b) the optimal redundancy index corresponding to any specified λ .

VI. EXPERIMENTAL RESULTS

In this section, we report some experimental results on the performance of the LR-PET scheme and the complexity of the proposed optimization procedure.

We first show a few examples of the effective P - R convex hulls, \mathcal{H}_C^1 , \mathcal{H}_C^2 , \mathcal{H}_C^3 and \mathcal{H}_C^4 , corresponding to LR-PET with $n = 1, 2, 3$ and 4 transmissions. For this purpose, two kinds of channel are considered. Fig. 7 corresponds to an IID channel with packet loss ratio $p_e = 0.4$. Fig. 8 corresponds to a Gilbert-Elliott (GE) channel whose packet loss ratio is $P_G = 0.01$ for the GOOD state and $P_B = 0.6$ for the BAD state. In this example, the mean time in the BAD state is $M_{BAD} = 300$ and the mean time in the GOOD state is twice that in the BAD state. Fig. 7(a) and Fig. 8(a) illustrate the P - R characteristic of these two types of channels, with $N = 50$. We find that as the number of transmission opportunities increases, the efficiency in bandwidth utilization is improved to some extent. We can notice that the improvement from the first retransmission is most significant. For the example IID channel, LR-PET with only one retransmission opportunity is almost enough to achieve the channel's capacity. But for the example GE channel, the improvements from the second and the third retransmission are still noticeable. Beside the P - R characteristic, we also show the optimal redundancy index to choose at different λ values, as illustrated in Fig. 7(b) and Fig. 8(b). As we expect, the protection becomes progressively stronger as the value of $1/\lambda$ increases. Furthermore, the

optimal redundancy index from \mathcal{H}_C^m is consistently smaller than that from \mathcal{H}_C^{m-1} .

Now we investigate the overall performance of the LR-PET scheme with n transmission opportunities. Three optimization strategies are considered: *greedy* (G), *partial greedy* (PG) and *hypothetical* (H). The G strategy excludes all the future retransmissions from the optimization objective of the current transmission slot; the PG strategy considers only the first future retransmission, if it exists; the H strategy considers all possible future retransmissions. For $n > 2$, PG provides a reasonable alternative to H . The difference between PG and H is that when the number of future transmission opportunities exceeds 1, only one future retransmission is hypothesized when finding the redundancy indices in the current slot.

For source content, we use the CIF 30Hz sequences, *Bus*, *Mobile*, *Foreman* and *Football*, and the 4CIF 60Hz sequences, *City*, *Soccer* and *Harbour*. These are standard MPEG test sequences. Each sequence is compressed with Motion JPEG2000, using 5 levels of DWT decomposition and 20 quality layers. For each sequence, we cycle through the first 30 frames 10^3 times, effectively creating a much larger sequence.

For channel simulation and packet loss estimation, we consider two cases. In the first case, we assume that the packet loss behavior of the communication channel satisfies the Gilbert-Elliott model and the parameters required by this model, including P_B , P_G , M_{BAD} and M_{GOOD} , are already

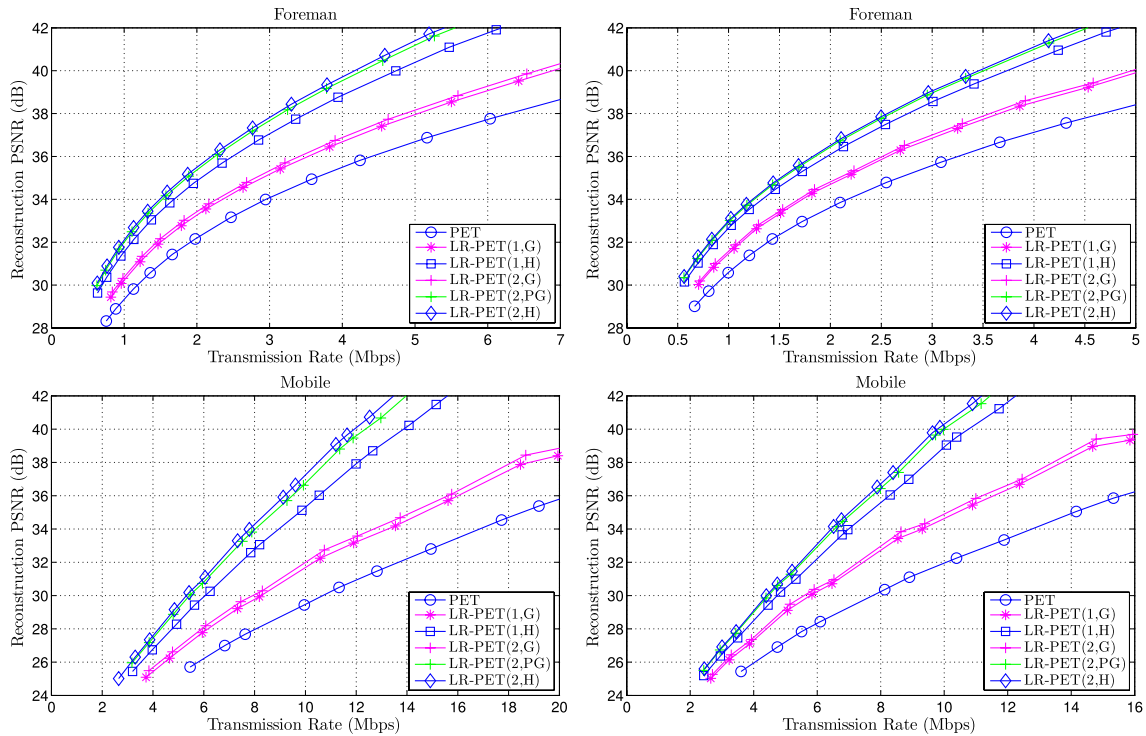


Fig. 9. Performance of LR-PET over GE channels, with up to two retransmission opportunities using different optimization strategies. Left: GE channel with $P_e = 0.207$ ($P_G = 0.01$, $P_B = 0.6$, $M_{BAD} = 300$, $M_{GOOD} = 600$). Right: GE channel with $P_e = 0.108$ ($P_G = 0.01$, $P_B = 0.6$, $M_{BAD} = 300$, $M_{GOOD} = 1500$).

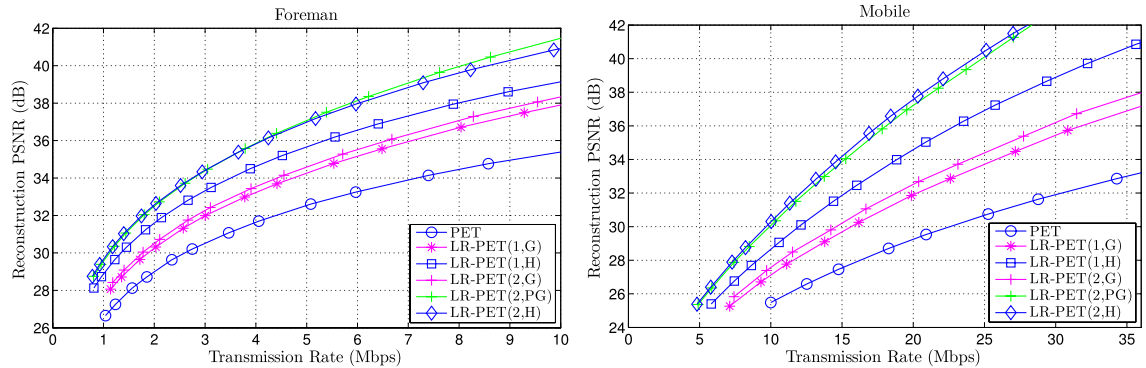


Fig. 10. Performance of LR-PET over a NS2-simulated channel, with up to two retransmission opportunities using different optimization strategies.

known. In this case, the packet loss probability is estimated dynamically, using the algorithm proposed in [50]. Two GE channels are considered: one channel has parameters $P_G = 0.01$, $P_B = 0.6$, $M_{BAD} = 300$, $M_{GOOD} = 600$, the other channel has the same parameters except that $M_{GOOD} = 1500$. The first channel has a probability of $1/3$ in the BAD state while the second channel has a probability of $1/6$ in the BAD state, producing effective packet loss probabilities of $P_e = 0.6 \times \frac{1}{3} + 0.01 \times \frac{2}{3} = 0.207$ and $P_e = 0.6 \times \frac{1}{6} + 0.01 \times \frac{5}{6} = 0.108$, respectively.

Fig. 9 shows the results for *Foreman* and *Mobile*. The number i in the legends “LR-PET(i ,X)” stands for the number of retransmission opportunities. From Fig. 9 we can see that, the performance of LR-PET increases with the number of retransmission opportunities, no matter which optimization strategy is used. However, the improvement from each extra

transmission opportunity is much larger when the hypothetical optimization strategy is used instead of more greedy approaches. For example, in Fig. 9, the first retransmission opportunity provides up to about 2dB gain with the greedy strategy, but up to 4 ~ 6dB gain with the hypothesis-based strategy. The second retransmission opportunity provides only 0.2~0.3dB gain with greedy optimization, but up to 1.2dB gain if hypotheses are formulated. The reason for this huge difference is that the greedy strategy ignores the possibility of future retransmissions so that it frequently overprotect streams during their primary transmission. We also notice that the PG strategy performs close to the H strategy, with a gap of only 0.1~0.3dB. This means including hypotheses on the second or additional future retransmissions into the optimization of the current transmission only provides limited improvement. This is because, in the above tested scenarios, most ele-

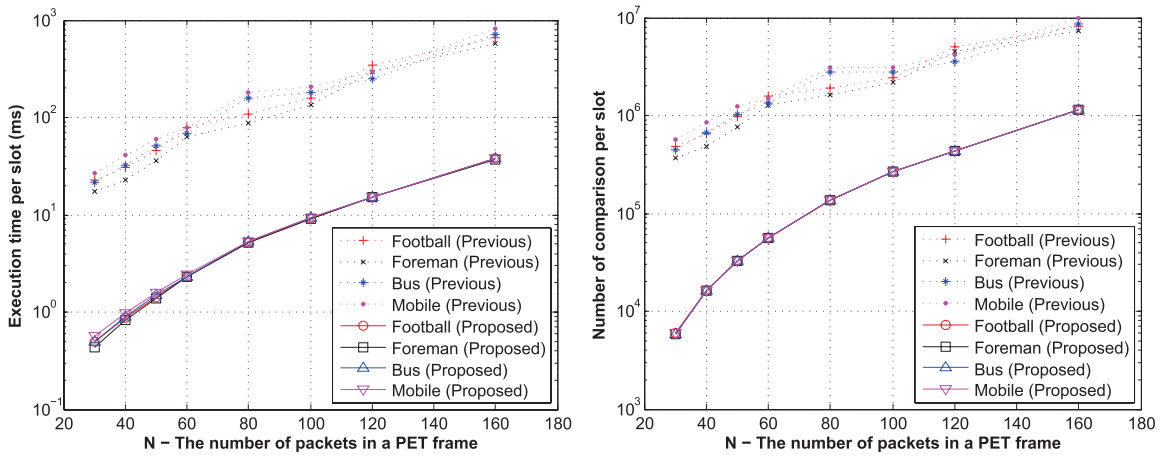


Fig. 11. Complexity comparison between the previous LR-PET optimization procedure and the new procedure proposed in this paper, measured in the execution time and the number of comparisons per transmission slot.

ments can get through the channel within two PET protected transmissions.

In the second case, instead of assuming a specific channel model, we simulate the LR-PET transmission process using NS2 (Net Simulator 2) [51] with a practical topology. We assume the delivery path from the server to the client goes through five shared links, each having a capacity of 100Mbps. We model the side traffic on each shared link as independent, and exhibiting long-range dependencies. Each link serves the packets in a first-in-first-out (FIFO) and drop-tail manner. In this situation, we do not have an accurate model for the LR-PET channel, so we use the general channel estimation algorithm proposed in [52] to learn packet loss probabilities. This algorithm is based on the assumption that the channel condition varies slowly so that the number of losses in previous slots can be used as a context to predict the number of losses in the current slot. Fig. 10 illustrates the performance of LR-PET over this NS2-simulated channel. The result for *Foreman* and *Mobile* are shown. Clearly, Fig. 10 is very similar to Fig. 9. The improvement from the second retransmission is even larger in Fig. 10 than in Fig. 9. This means the LR-PET scheme and proposed optimization procedure still work well for scenarios where we do not have an accurate channel model for estimating loss probabilities.

Finally, we compare the complexity of the proposed LR-PET optimization procedure with the existing LR-PET procedure in [25]. To make the comparison feasible, only two transmissions are considered. The results are obtained from a simulation consisting of 30000 transmission slots, on a computer with 2.66GHz Xeon CPU and 4GB memory. For the proposed procedure, the convex hull construction is performed in each slot, i.e. $I_{\text{slot}} = 1$. Two measures are used as indicators of the complexity. One is the execution time and the other is the number of comparisons performed. As illustrated in Fig. 11, both indicators exhibit the same trend. For both optimization procedures, the complexity increases rapidly with N – the channel code length. But with the proposed new procedure, the complexity has been reduced by about 95%. For example, for $N = 50$, the execution time has been reduced

from 30~60 ms per slot to 1~2 ms per slot. Furthermore, the complexity reduction ratios measured by both indicators are quite consistent. To make the comparison fair, we also evaluated the performance loss caused by the approximation in P - R convex hull construction. For the above scenarios, the performance loss in average PSNR is no more than 0.01dB for all the tested video sequences.

VII. CONCLUSION

As demonstrated in Section VI, LR-PET performs much better than PET. The performance gain of LR-PET over PET mainly comes from the consideration of hypotheses concerning the effect of possible future retransmissions. This effectively creates longer channel codes, spanning multiple transmission slots, which collect more packet loss events so that the channel capacity can be approached. When the size of a PET frame becomes very large, for example when each “source frame” is extended to include a group of pictures, the plain PET scheme may perform as well as LR-PET. However, a large PET frame either leads to very large packet sizes or very long channel codes. Large packet sizes usually involve high packet loss rates. Long channel codes, on the other hand, involve high decoding complexity. Both are undesirable.

Moreover, LR-PET is more flexible than PET. The extended channel code effectively created by LR-PET is progressively determined as the feedback from previous transmissions is received. It can also be adapted as channel conditions change. The long channel codes used in PET in the case of large frame size, on the other hand, must be determined completely ahead of the transmission.

This paper develops a method to efficiently derive the effective P - R characteristic for the extended channel codes formed by the LR-PET procedure with any number of transmissions. Importantly, we have shown that the protection convex hulls can be constructed using a recursive procedure whose empirically observed complexity grows only (approximately) linearly with the number of transmission opportunities. The proposed optimization algorithm offers several advantages. Firstly, the protection assignment procedure in the

original LR-PET scheme with only one retransmission can be accelerated significantly. Secondly, this algorithm makes it feasible to efficiently solve the optimal protection assignment problem for scenarios with more retransmissions.

Another contribution of this paper is the proof of the redundancy embedding property for the codes formed by LR-PET procedure. This property guarantees the automatic satisfaction of the order of redundancy indices for neighboring elements, given any source with convex utility-length characteristic. Of course, it is always possible to sequentially enforce redundancy constraints within a practical optimization procedure; however, without the proof provided by this paper, one could never be sure that the resulting solution would be optimal. The redundancy embedding property assures us that at any of the n transmission opportunities, the protection of each stream element can be optimized independently, without compromising the optimality of the obtained solution. This, in turn, justifies the construction of a single source-independent effective protection convex hull.

APPENDIX I PROOF OF PROPERTY 3

Recall that $\{s_j\}_j$ is an enumeration of slopes in $\mathbb{S}^1(\tilde{\mathcal{H}})$ and $\{s'_i\}_i$ is an enumeration of slopes in $\mathbb{S}^2(\tilde{\mathcal{H}}, r)$, both in decreasing order. For the trivial case $i = 1$, $\tilde{S}_r^2(i) = s'_1 = +\infty$. Therefore, we only need to consider the cases $i > 1$. According to the definition (24) and (25), we have

$$\begin{aligned} \tilde{S}_r^2(i) &= \frac{P_{\Upsilon_{1,2}}^2(r, s'_i) - P_{\Upsilon_{1,2}}^2(r, s'_{i-1})}{R_{\Upsilon_{1,2}}^2(r, s'_i) - R_{\Upsilon_{1,2}}^2(r, s'_{i-1})} \\ &= \frac{\sum_{k \in \mathbb{K}(r)} \rho_k^1 \left(P_{\tilde{\mathcal{H}}}^1(\mathfrak{S}(\tilde{\mathcal{H}}, s'_i \theta(r, k))) - P_{\tilde{\mathcal{H}}}^1(\mathfrak{S}(\tilde{\mathcal{H}}, s'_{i-1} \theta(r, k))) \right)}{\sum_{k \in \mathbb{K}(r)} \rho_k^1 \theta(r, k) \left(R_{\tilde{\mathcal{H}}}^1(\mathfrak{S}(\tilde{\mathcal{H}}, s'_i \theta(r, k))) - R_{\tilde{\mathcal{H}}}^1(\mathfrak{S}(\tilde{\mathcal{H}}, s'_{i-1} \theta(r, k))) \right)}. \end{aligned}$$

Actually, most terms in the above sum are zero because $x_1^k = s'_i \theta(r, k)$ and $x_2^k = s'_{i-1} \theta(r, k)$ lie within the same interval $(s_j, s_{j-1}]$ (depending on k). Non-zero contributions only occur when x_1^k and x_2^k lie in different intervals. If that occurs, suppose the interval in which x_2^k lies is $(s_j, s_{j-1}]$, we must have $x_1^k = s_j$, as illustrated in Fig. 12. To see this, consider the facts that $s_j/\theta(r, k) \in \mathbb{S}^2(\tilde{\mathcal{H}}, r)$ and s'_i is the largest element in $\mathbb{S}^2(\tilde{\mathcal{H}}, r)$ which is smaller than s'_{i-1} , we easily obtain $s_j/\theta(r, k) \leq s'_i$, i.e. $s_j \leq x_1^k$. Since $x_1^k < x_2^k$ and x_1^k is outside of $(s_j, s_{j-1}]$, we get $x_1^k = s_j$. Therefore, we have $\mathfrak{S}(\tilde{\mathcal{H}}, s'_i \theta(r, k)) = j$ and $\mathfrak{S}(\tilde{\mathcal{H}}, s'_{i-1} \theta(r, k)) = j-1$. The contribution of corresponding term in the sum is

$$\frac{\rho_k^1 \left(P_{\tilde{\mathcal{H}}}^1(\mathfrak{S}(\tilde{\mathcal{H}}, s_j)) - P_{\tilde{\mathcal{H}}}^1(\mathfrak{S}(\tilde{\mathcal{H}}, s_{j-1})) \right)}{\rho_k^1 \theta(r, k) \left(R_{\tilde{\mathcal{H}}}^1(\mathfrak{S}(\tilde{\mathcal{H}}, s_j)) - R_{\tilde{\mathcal{H}}}^1(\mathfrak{S}(\tilde{\mathcal{H}}, s_{j-1})) \right)} = \frac{\rho_k^1}{\rho_k^1 \theta(r, k)} s_j = s'_i.$$

The last equation follows from the fact that $s_j = s'_i \theta(r, k)$. It follows that all non-zero terms must have the same ratio s'_i , which is independent of k , even though s_j depends on k . Therefore, $\tilde{S}_r^2(i) = s'_i$.

Lemma A1: If $r_1 > r_2$, $P_{\Upsilon_{1,n}}^n(r_1, \lambda) \geq P_{\Upsilon_{1,n}}^n(r_2, \lambda)$ is true for all λ .

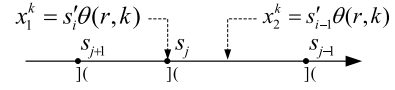


Fig. 12. The intervals to which $s'_i \theta(r, k)$ and $s'_{i-1} \theta(r, k)$ belong.

Proof: We define $\overline{\mathbb{K}}(r) = [0, N] - \mathbb{K}(r) = [k_{\min}(r), N]$. Since $r_1 > r_2$, we have $k_{\min}(r_1) < k_{\min}(r_2)$. Thus $\mathbb{K}(r_1) \subset \mathbb{K}(r_2)$ and $\overline{\mathbb{K}}(r_1) \supset \overline{\mathbb{K}}(r_2)$. Therefore,

$$\begin{aligned} &P_{\Upsilon_{1,n}}^n(r_1, \lambda) - P_{\Upsilon_{1,n}}^n(r_2, \lambda) \\ &= \left[\sum_{k \in \overline{\mathbb{K}}(r_1)} \rho_k^1 + \sum_{k \in \mathbb{K}(r_1)} \rho_k^1 P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda \theta(r_1, k))) \right] \\ &\quad - \left[\sum_{k \in \overline{\mathbb{K}}(r_2)} \rho_k^1 + \sum_{k \in \mathbb{K}(r_2)} \rho_k^1 P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda \theta(r_2, k))) \right] \\ &= \sum_{\substack{k \in \overline{\mathbb{K}}(r_1) \\ k \notin \overline{\mathbb{K}}(r_2)}} \rho_k^1 \left[1 - P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda \theta(r_2, k))) \right] \\ &\quad + \sum_{k \in \mathbb{K}(r_1)} \rho_k^1 \left[P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda \theta(r_1, k))) - P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda \theta(r_2, k))) \right] \end{aligned}$$

Since $r_1 > r_2$, we have $\theta(r_1, k) < \theta(r_2, k)$ and hence $P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda \theta(r_1, k))) \geq P_{\tilde{\mathcal{H}}}(\mathfrak{S}(\tilde{\mathcal{H}}, \lambda \theta(r_2, k)))$. Obviously, $P_{\tilde{\mathcal{H}}}(j) \leq 1$ is true for any j . Therefore, all items in the above equation are non-negative. ■

Lemma A2: For any r , $\tilde{J}(\lambda, r) = (1/\lambda) P_{\Upsilon_{1,n}}^n(r, \lambda) - R_{\Upsilon_{1,n}}^n(r, \lambda)$ is a continuous function of λ .

Proof: Property 2 tells us that $P_{\Upsilon_{1,n}}^n(r, \lambda)$ and $R_{\Upsilon_{1,n}}^n(r, \lambda)$ remain constant for all $\lambda \in (s'_{i+1}, s'_i]$, where $\{s'_i\}_i$ is an enumeration of items in $\mathbb{S}^2(\tilde{\mathcal{H}}, r)$ in decreasing order. Obviously, we only need to investigate the right-hand continuity of $\tilde{J}(\lambda, r)$ at $\lambda = s'_i$. When $\Delta \rightarrow 0^+$, we have $s'_i + \Delta \in (s'_i, s'_{i-1}]$, hence

$$\begin{aligned} &\lim_{\Delta \rightarrow 0^+} \left(\tilde{J}(s'_i + \Delta, r) - \tilde{J}(s'_i, r) \right) \\ &= \lim_{\Delta \rightarrow 0^+} \left\{ \left[\frac{1}{s'_i + \Delta} \cdot P_{\Upsilon_{1,n}}^n(r, s'_{i-1}) - R_{\Upsilon_{1,n}}^n(r, s'_{i-1}) \right] \right. \\ &\quad \left. - \left[\frac{1}{s'_i} \cdot P_{\Upsilon_{1,n}}^n(r, s'_i) - R_{\Upsilon_{1,n}}^n(r, s'_i) \right] \right\} \\ &= \frac{1}{s'_i} \cdot [P_{\Upsilon_{1,n}}^n(r, s'_{i-1}) - P_{\Upsilon_{1,n}}^n(r, s'_i)] - [R_{\Upsilon_{1,n}}^n(r, s'_{i-1}) - R_{\Upsilon_{1,n}}^n(r, s'_i)] \\ &= 0, \end{aligned}$$

where the last line follows from Property 3. Therefore, $\tilde{J}(\lambda, r)$ is continuous on the whole λ domain. ■

Corollary A1: If $r_1 > r_2$, the function $\tilde{D}_{r_1, r_2}(\lambda) = \tilde{J}(\lambda, r_1) - \tilde{J}(\lambda, r_2)$ is a non-increasing function of λ .

Proof: For any λ not in $\mathbb{S}^2(\tilde{\mathcal{H}}, r)$, the derivative of $\tilde{J}(\lambda, r)$ is

$$\begin{aligned} \frac{\partial \tilde{J}(\lambda, r)}{\partial \lambda} &= \frac{-1}{\lambda^2} P_{\Upsilon_{1,n}}^n(r, \lambda) + \frac{1}{\lambda} \frac{\partial P_{\Upsilon_{1,n}}^n(r, \lambda)}{\partial \lambda} - \frac{\partial R_{\Upsilon_{1,n}}^n(r, \lambda)}{\partial \lambda} \\ &= (-1/\lambda^2) \cdot P_{\Upsilon_{1,n}}^n(r, \lambda), \end{aligned}$$

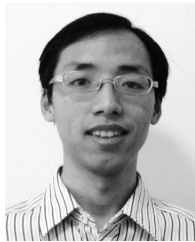
because $\partial P_{\Upsilon_{1,n}}^n(r, \lambda)/\partial \lambda = \partial R_{\Upsilon_{1,n}}^n(r, \lambda)/\partial \lambda = 0$ - c.f. Property 2. Therefore, if $r_1 > r_2$, according to Lemma A1 we have

$\partial \tilde{D}_{r_1, r_2}(\lambda) / \partial \lambda = (-1/\lambda^2) \cdot (P_{\gamma_{1:n}}^n(r_1, \lambda) - P_{\gamma_{1:n}}^n(r_2, \lambda)) \leq 0$. This means $\tilde{D}_{r_1, r_2}(\lambda)$ is at least piecewise decreasing. Since $\tilde{J}(\lambda, r_1)$ and $\tilde{J}(\lambda, r_2)$ are both continuous (c.f. Lemma A2), $\tilde{D}_{r_1, r_2}(\lambda)$ is also continuous. Therefore, the function $\tilde{D}_{r_1, r_2}(\lambda)$ is non-increasing on the whole λ domain. ■

REFERENCES

- [1] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.
- [2] A. Mohr, E. Riskin, and R. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 819–828, Jun. 2000.
- [3] W.-T. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 373–386, Mar. 2001.
- [4] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust Internet video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 357–372, Mar. 2001.
- [5] S. Dumitrescu, X. Wu, and Z. Wang, "Globally optimal uneven error-protected packetization of scalable code streams," in *Proc. IEEE Data Commun. Conf.*, Apr. 2002, pp. 73–82.
- [6] V. Stankovic, R. Hamzaoui, Y. Charfi, and Z. Xiong, "Real-time unequal error protection algorithms for progressive image transmission," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1526–1535, Dec. 2003.
- [7] J. Kim, R. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 121–131, Feb. 2003.
- [8] N. Thomos, N. Boulgouris, and M. Strintzis, "Wireless image transmission using turbo codes and optimal unequal error protection," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1890–1901, Nov. 2005.
- [9] R. Puri and K. Ramchandran, "Multiple description source coding using forward error correction codes," in *Proc. 33rd Asilomar Conf. Signals, Syst., Comput.*, vol. 1, Oct. 1999, pp. 342–346.
- [10] T. Stockhammer and C. Buchner, "Progressive texture video streaming for lossy packet networks," in *Proc. 11th Int. Packet Video Workshop*, May 2001, p. 57.
- [11] J. Thie and D. Taubman, "Optimal protection assignment for scalable compressed images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Sep. 2002, pp. 713–716.
- [12] M. Podolsky, M. Vetterli, and S. McCanne, "Limited retransmission of real-time layered multimedia," in *Proc. IEEE 2nd Workshop Multimedia Signal Process.*, Dec. 1998, pp. 591–596.
- [13] M. Podolsky, S. McCanne, and M. Vetterli, "Soft ARQ for layered streaming media," *J. VLSI Signal Process. Signal, Image, Video Technol.*, vol. 27, pp. 81–97, Feb. 2001.
- [14] V. Chande, N. Farvardin, and H. Jafarkhani, "Image communication over noisy channels with feedback," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Oct. 1999, pp. 540–544.
- [15] C.-Y. Hsu and A. Ortega, "Rate control for robust video transmission over burst-error wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 5, pp. 756–773, May 1999.
- [16] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," Microsoft Corp., Redmond, WA, USA, Tech. Rep. MSR-TR-2001-35, 2001.
- [17] G. Cheung, W.-T. Tan, and T. Yoshimura, "Rate-distortion optimized application-level retransmission using streaming agent for video streaming over 3G wireless network," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, Sep. 2002, pp. 529–532.
- [18] R. Puri and K. Ramchandran, "Joint source channel coding with hybrid FEC/ARQ for buffer constrained video transmission," in *Proc. IEEE 2nd Workshop Multimedia Signal Process.*, Dec. 1998, pp. 567–572.
- [19] P. Chou, A. Mohr, A. Wang, and S. Mehrotra, "FEC and pseudo-ARQ for receiver-driven layered multicast of audio and video," in *Proc. IEEE Data Commun. Conf.*, Snowbird, UT, USA, Mar. 2000, pp. 440–449.
- [20] J. Chakareski, P. Chou, and B. Aazhang, "Computing rate-distortion optimized policies for streaming media to wireless clients," in *Proc. IEEE Data Commun. Conf.*, Snowbird, UT, USA, Apr. 2002, pp. 53–62.
- [21] G. Wang, Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Channel-adaptive unequal error protection for scalable video transmission over wireless channel," *Proc. SPIE Visual Commun. Image Process.*, vol. 4310, pp. 648–655, Dec. 2001.
- [22] T. Stockhammer, H. Jenak, and C. Weiss, "Feedback and error protection strategies for wireless progressive video transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 465–482, Jun. 2002.
- [23] X. Zheng, S.-H. Chan, Q. Zhang, W.-W. Zhu, and Y.-Q. Zhang, "Feedback-free packet loss recovery for video multicast," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, Apr. 2003, pp. 870–874.
- [24] T. Gan and K.-K. Ma, "Sliding-window packetization for forward error correction based multiple description transcoding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Apr. 2003, pp. 756–759.
- [25] D. Taubman and J. Thie, "Optimal erasure protection for scalably compressed video streams with limited retransmission," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1006–1019, Aug. 2005.
- [26] J. Thie and D. Taubman, "Optimal erasure protection for scalably compressed video streams with limited retransmission on channels with iid and bursty loss characteristics," *Signal Process. Image Commun.*, vol. 20, pp. 697–709, May 2005.
- [27] J. Thie and D. Taubman, "Optimal erasure protection strategy for scalably compressed data with tree-structured dependencies," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2002–2011, Dec. 2005.
- [28] R. Xiong and D. Taubman, "Optimal LR-PET protection for scalable video streams over lossy channels with random delay," in *Proc. IEEE 10th Workshop Multimedia Signal Process.*, Oct. 2008, pp. 439–444.
- [29] R. Xiong, D. Taubman, and V. Sivaraman, "Optimal PET protection for streaming scalably compressed video streams with limited retransmission based on incomplete feedback," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2382–2395, Sep. 2010.
- [30] S. Ahmad, R. Hamzaoui, and M. Al-Akaidi, "Adaptive unicast video streaming with rateless codes and feedback," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 2, pp. 275–285, Feb. 2010.
- [31] S. Ahmad, R. Hamzaoui, and M. Al-Akaidi, "Unequal error protection using fountain codes with applications to video communication," *IEEE Trans. Multimedia*, vol. 13, no. 1, pp. 92–101, Feb. 2011.
- [32] P. Tourmoux, E. Lochin, J. Lacan, A. Bouabdallah, and V. Roca, "On-the-fly erasure coding for real-time video applications," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 797–812, Aug. 2011.
- [33] H.-P. Shiang and M. van der Schaar, "Online learning in autonomic multi-hop wireless networks for transmitting mission-critical applications," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 5, pp. 728–741, Jun. 2010.
- [34] D. Vukobratovic and V. Stankovic, "Unequal error protection random linear coding strategies for erasure channels," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1243–1252, May 2012.
- [35] S. Arslan, P. Cosman, and L. Milstein, "Generalized unequal error protection LT codes for progressive data transmission," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3586–3597, Aug. 2012.
- [36] M. Luby, "LT codes," in *Proc. 43rd Annu. IEEE Symp. Found. Comput. Sci.*, 2002, pp. 271–280.
- [37] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.
- [38] J. Shapiro, "An embedded hierarchical image coder using zerotrees of wavelet coefficients," in *Proc. IEEE Data Commun. Conf.*, Snowbird, UT, USA, Dec. 1993, pp. 214–223.
- [39] A. Said and W. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
- [40] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.
- [41] J. Ohm, "Three dimensional subband coding with motion compensation," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 559–571, Apr. 1994.
- [42] D. Taubman and A. Zakhor, "Highly scalable low-delay video compression," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, Nov. 1994, pp. 740–744.
- [43] S. Choi and J. Woods, "Motion compensated 3-D subband coding of video," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 155–167, Feb. 1999.
- [44] B. Pesquet-Popescu and V. Botreau, "Three-dimensional lifting schemes for motion compensated video compression," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2001, pp. 1793–1796.
- [45] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1530–1542, Dec. 2003.

- [46] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
- [47] M. Durigon and D. Taubman, "LR-PET optimization strategy for protection of scalable video with unreliable acknowledgement," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 9–12.
- [48] S. Dumitrescu, "Optimal erasure protection for scalably compressed video streams with limited retransmission," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 214–216, Jan. 2009.
- [49] R. Xiong and D. Taubman, "Optimal PET protection for scalably compressed video with multiple retransmissions," in *Proc. 5th Int. Conf. VIE*, Jul. 2008, pp. 795–800.
- [50] Y.-C. Su and C.-S. Y. C.-W. Lee, "The analysis of packet loss prediction for Gilbert-model with loss rate uplink," *Inf. Process. Lett.*, vol. 90, pp. 155–159, May 2004.
- [51] (2008). *The Network Simulator—ns-2*, Version ns-2.32 [Online]. Available: <http://www.isi.edu/nsnam/ns/>
- [52] M. Durigon, D. Taubman, J. Thie, and T. Moors, "Dynamically optimized protection for scalably compressed media with limited retransmission over channels with memory," *IEEE Trans. Multimedia*, to be published.



Ruiqin Xiong (M'08) received B.S. degree from the University of Science and Technology of China, Hefei, China, in 2001, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2007.

He was with Microsoft Research Asia as a Research Intern from 2002 to 2007 and the University of New South Wales, New South Wales, Australia, as a Senior Research Associate, from 2007 to 2009. He joined Peking University, Beijing, in 2010. His current research interests include image

and video compression, image and video restoration, joint source channel coding, and multimedia communication.



David Taubman (M'95–SM'06) received the B.S. and B.E. degrees from the University of Sydney, Sydney, Australia, in 1986 and 1988, respectively, and the M.S. and Ph.D. degrees from the University of California at Berkeley, Berkeley, CA, USA, in 1992 and 1994, respectively. From 1994 to 1998, he was with Hewlett-Packard's Research Laboratories, Palo Alto, CA, USA, joining the University of New South Wales, New South Wales, Australia, in 1998, where he is a Professor with the School of Electrical Engineering and Telecommunications. He is the author of *JPEG2000: Image Compression Fundamentals, Standards and Practice* with M. Marcellin. His current research interests include highly scalable image and video compression, inverse problems in imaging, perceptual modeling, and multimedia distribution systems.

He received the University Medal from the University of Sydney. He has received two Best Paper Awards: from the IEEE Circuits and Systems Society for the 1996 paper, "A Common Framework for Rate and Distortion Based Scaling of Highly Scalable Compressed Video"; and from the IEEE Signal Processing Society for the 2000 paper, "High Performance Scalable Image Compression with EBCOT."



Vijay Sivaraman (M'94) received the B.Tech. degree from the Indian Institute of Technology, Delhi, India, in 1994, the M.S. degree from North Carolina State University, Raleigh, NC, USA, in 1996, and the Ph.D. degree from the University of California at Los Angeles, Los Angeles, CA, USA, in 2000.

He was with Bell Labs as a Student Fellow, in a Silicon Valley start-up manufacturing optical switch-routers, and as a Senior Research Engineer with the CSIRO, Clayton, Australia. He is currently an Associate Professor with the University of New South Wales, Sydney, Australia. His current research interests include optical networking, packet switching and routing, network architectures, sensor networks for the environment, and healthcare and sports monitoring.