An end-to-end robust approach for scalable video over the Internet

WANG Guijin[1], ZHANG Qian[2], ZHU Wenwu[2] & LIN Xinggang[1]

1. Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;
2. Microsoft Research Asia, Beijing 100080, China
Correspondence should be addressed to Wang Guijin (email: wangguijin@tsinghua.org.cn; gj-wang@hotmail.com)

Received August 25, 2003

Abstract This paper introduces an end-to-end robust approach for scalable video over the Internet. The traditional method only considers congestion control, error control and is unable to achieve end-to-end high-quality video transmission in the error-prone environment like the Internet since it does not consider the packetization behavior, network conditions and the media characteristics simultaneously. This paper presents an end-to-end approach for scalable video over the Internet, combining network adaptive congestion control and unequal error control. Considering requirements of multimedia transmission, this paper introduces multimedia congestion control to estimate available bandwidth and smooth the media sending rate. Specially in the transport layer we propose unequal interleaving packetization method and unequal error protection scheme, which can alleviate the effect of the packet loss well. Further we develop the rate-distortion theory for the scalable video over the Internet. Thereafter the optimal bit allocation is presented to determine the bits budgets for the source part and error control part. Simulation shows our scheme can achieve good performance for scalable video over the Internet.

Keywords: congestion control, error control, video transmission, rate-distortion, scalable video.

DOI: 10.1360/02yf0215

With the growth of the Internet and abundant network resources, video streaming is becoming one of the increasingly important Internet applications. However, the current IP-based network provides only a single class best effort service. Video packet can be regarded as a packet loss by the video decoder either due to network congestion or due to exceeding the maximum delay threshold. It remains an open challenging task as to how to cope with the packet loss in the video streaming over the Internet and achieve high reconstructed-video quality.

Scalable video coding is of great interest recently because it is capable of coping with variability of bandwidth gracefully[1, 2]. Scalable source coder encodes input video into multiple layers. Base layer (BL), the most important layer that has to be successfully transmitted and highly protected, carries the basic information. Other layers are called enhancement layers (ELs) that can be dropped or transmitted based on the available network bandwidth.

So far two types of techniques have been proposed to address the issue of video over network in the literature: congestion control and error control. The basic idea of congestion control for video streaming case is to attempt to minimize the amount of packet loss by adjusting the delivery rate according to the network congestion status meanwhile maintaining the sending rate smoothly as much as possible[3−5]. However, they only considered the congestion control techniques for video streaming and did not discuss how to control the quality of video transmission when packet loss occurs. Besides, note that this technique does not guarantee there is no packet-loss in the streaming process. Joint work on congestion control and error control was discussed in ref. [6]. However, in this work we propose unequal interleaving packetization to enhance the usage efficiency of the network bandwidth. Joint work on layered scalable coding with Unequal Loss Protection (ULP) for robust Internet transmission has also been studied in refs. [7 — 9]. However, besides the interleaving packetization, we develop the rate-distortion theory for the scalable video over the Internet and give the global bit allocation to distribute the bit budgets between the source and channel so as to achieve good end-to-end system performance.

In this paper we propose an end-to-end robust approach for scalable video over the Internet, considering congestion, error control, data packetization simultaneously. In the transport layer, we present MSTFP (Multimedia Streaming Tcp-Friendly Protocol) to estimate the available bandwidth dynamically and smooth the video sending rate. In the application layer we present unequal error protection to alleviate the effect of the packet loss. Further unequal interleaving packetization scheme is introduced to enhance system robust performance. Particularly we analyze the rate-distortion relationship of the scalable video over the Internet and determine the protection degree of each video layer so as to achieve the minimal end-to-end distortion.

1 End-to-end approach for scalable video over the Internet

1.1 Problem formulation

Fig. 1 depicts the architecture for video streaming across the Internet. In this framework, RTP[10] above UDP is served as the transport protocol for the network multimedia application while RTCP is accompanied to convey the feedback information for Network Estimation module to track the network status. Also TCP-friendly Congestion Control module is with the transport layer to smooth the rate of the video data. In the application layer, unequal error control strategy is considered to utilize the unequal philosophy in the scalable bit-stream. The procedure of the video over the Internet can be described as follows. On the sender side, raw video has been compressed offline. Taking the unequal importance characteristics among the layers into account, these layers are protected with different protection degrees against packet losses according to network conditions. After this stage, the video data is pacetized and then passed transport layer before entering the network. These packets may be dropped by the router due to the network congestion or by the receiver due to excess delay. On the receiver side, after the inverse process of channel decoding, the reconstructed packets are directed to the source decoder. In the mean time, the receiver registered some statistical information, such as packet loss ratio and the transmission delay, by the Network Monitor module. The estimated available network bandwidth is dynamically updated using feedback information in the Network Estimation module. With the observed network conditions, the Target Bit Allocation module allocates resource between source and channel aiming to achieve high video transmission quality.

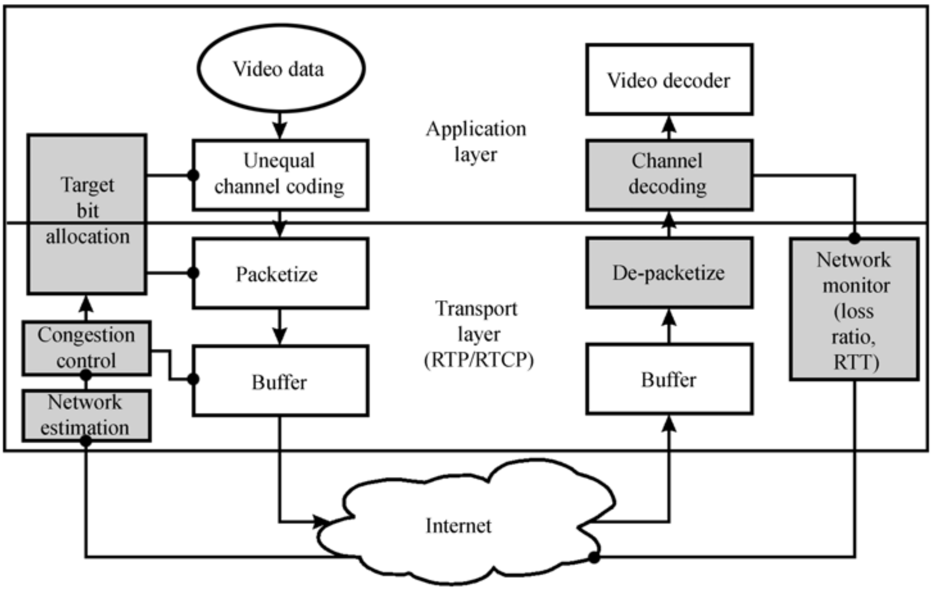

Fig. 1. End-to-end framework for scalable video over the Internet.

The key modules in the end-to-end video delivery system include Unequal Channel Coding, Target Bit Allocation, Packetize, Congestion Control, and Network Estimation, etc. In the following sections, we will introduce our algorithms of each key module in the proposed end-to-end framework: section 2 describes our TCP-friendly congestion control strategy; section 3 presents the proposed unequal data packetization, unequal error protection, and target bit allocation respectively.

1.2 Related background knowledge

1.2.1 Scalable video coder

PFGS (Progressive Fine Granular Scalable)[1] video is adopted as an example of scalable video in our work. PFGS video codec is designed based on MPEG-4 FGS[2] with a higher coding efficiency. PFGS generates bit rates anywhere from tens of kilobits to a few mega bits per second with arbitrarily fine granularity. BL (Base Layer) carries the most important information, such as motion vector information and etc. All the enhancement layers are coded mainly based on the base layer in the same frame and the reference frames. Thus, errors in the enhancement layers do not cause any drifting problem in the corresponding prediction layers. Note that layers of the same frame are correlated. Specifically, the higher layer information relies on the corresponding one in the lower layers. In the receiver, if any residual error occurs in the lower layers, the corresponding information in the higher layers will be discarded whether they are correct or not.

PFGS provides some tools to enhance the error resilience performance of the bit-stream. The tools adopted in the BL are Resynchronization Marker (RM), HEC, RVLC and Data Partition so that acceptable reconstructed video can be obtained under the error rate 10^{-5}. RM and HEC are added to the PFGS enhancement-layer bit-stream to improve its robustness and efficiency. With the resynchronization markers in the enhancement-layer bitstream, the decoder could just discard the part of the bitstream between two resynchronization markers and continue decoding the rest of the bitstream to minimize the error effects.

1.2.2 Path characteristics

Measures of packet loss in the Internet have shown that the behavior of the packet loss can be modeled well with 2-state Markov process (see fig. 2)[12]. The Markov chain is in state *R* if a packet is received in time and in state *L* if it is lost (see fig. 2), whose transition matrix is given by

$$A=\begin{bmatrix}1-p & p\\ q & 1-q\end{bmatrix}. \qquad (1)$$

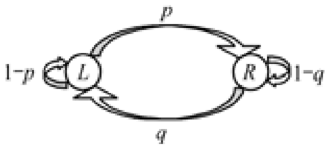

Fig. 2. Two-state Markov model of packet loss.

The parameters *p* and *q* are the transition probabilities between states *L* and *R*, which can be got by maximum likelihood (ML) estimators

$$p=\frac{N_1}{N_1+N_2} \quad \text{and} \quad q=\frac{N_3}{N_3+N_4}, \qquad (2)$$

where $N_1(N_2)$ is the number of successfully received (lost) packets when the previous packet is lost; $N_3(N_4)$ is the number of lost (successfully received) packets when the previous packet is successfully received. These parameters can be measured in the Network Monitor module (see fig. 1).

2 Network-adaptive congestion control

When the congestion occurs, all the network applications should adjust their sending rate to relieve the network burden so as to keep the network robustness and fairness, as known TCP offers most of the traffic in the current network[13]. To cope with packet dropping and bandwidth fluctuation in TCP we propose the multimedia streaming TCP-friendly protocol MSTFP to estimate available bandwidth, which can also share the bandwidth with the TCP fairly.

On the sender side, the *RTT* (Round Trip Time) can be estimated as

$$RTT = \alpha \times \overline{RTT} + (1-\alpha) \times (now - ST1 - \Delta RT), \tag{3}$$

where $\overline{RTT}$ is the current round trip time, *RTT* is the estimated round trip time, *now* is the timestamp of the arriving packet at the sender, *ST*1 indicates the last sending timestamp at the sender, ΔRT is the time spending in the receiver, and α is a weight parameter (here it is set to 0.75). Thus the available network bandwidth can be estimated as follows:

$$AvailBW = \frac{PacketSize}{RTT \times \sqrt{2P_L/3} + 3 \times TO \times P_L \times \sqrt{3P_L/8} \times \sqrt{1+32P_L^2}}, \tag{4}$$

where *PacketSize* is the length of the sending packet, P_L is the packet loss ratio, and time-out (TO) using the TCP model developed in ref. [14].

After estimating the available network bandwidth, the sender can dynamically adjust its rate as follows:

```
if ($AvailBW > \overline{currate}$)
    $multi = (now - lastchange)/RTT$
    constraint multi from 1 to 2
    $currate = \overline{currate} + (PacketSize/RTT) \times multi$
else
    $currate = \beta \times AvailBW + (1-\beta) \times \overline{currate}$,
```

where $\overline{currate}$ is the present sending rate, *currate* is the updated sending rate, *lastchange* is the timestamp of previous adjustment, β is the weight parameter which is set at 0.75 here. There are two advantages of our congestion control: 1) it can smooth the sending rate as much as possible, which is favorable for video transmission; 2) it is not sensitive to the packet loss.

3 Network-adaptive unequal error control

Although with our MSTFP congestion control, video data can still experience packet-loss in the network. This section presents unequal error control to alleviate the effect of the packet loss, including unequal interleaving packetization, unequal error protection and bit allocation.

3.1 Unequal interleaving packetization and error protection

To combat packet loss in the network, parities should be accompanied by the video data. In our work, FEC is adopted as the means of the error control. The idea of FEC across the packets is to transmit the redundant packets to help the receiver to recover lost packets. Reed-Solomon (RS) code is selected in our scheme, which is perfectly suited for error correction of the erasure errors like packet loss because they are the separable BCH codes with maximum distance[15]. With the known error position, RS (n, k) can recover even up to $t=n-k$ symbol errors. Across the packets, $RS(n, k)$ encodes k information symbols (each symbol per packet) into n symbols so as to construct the n-k parity packets (see fig. 3). Our symbols are bytes of 8 bits to favor the access the video data.

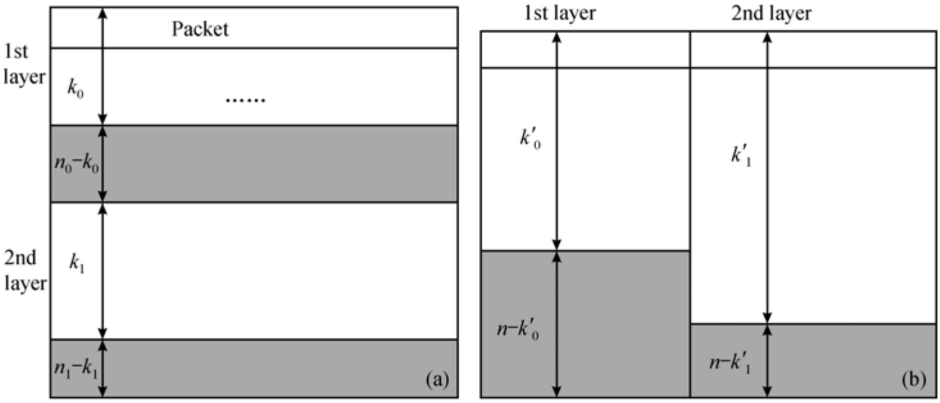

Fig. 3. Two types of packetization. (a) Traditional packetization; (b) unequal interleaving packetization.

For the scalable bit-stream, it is known that the first layer is more important than the second, which is more important than the next layer, etc. This unequal priority nature calls for ULP among different layers. Fig. 3 gives two packetization schemes. In fig. 3(a), after the n_0 packets of the first layer, the n_1 packets are delivered to the network. As known, the packet-loss in the network presents burst behavior. In the fig. 3(a) scheme, the case can occur that the higher layers can be reconstructed while the lower layers cannot, for example there are n_0-k_0+1 lost packets in the first layer. Considering the burst packet-loss behavior, we propose interleaving packetization scheme (see in fig. 3(b)). The basic idea is to interleave the information data among the layers and randomize the error so as to recover the loss information effectively. We multiplex all the layers in a group of frames, say 20 frames, into one new transmitting unit, block of packets (BOP), one for each layer, as shown in fig. 3(b). Meanwhile unequal loss protection is achieved by the varying k. For a given bit rate R (in bits), the number of packets n in one BOP equals R/P_L, where P_L is the packet size. Information bits in the l-th $(1\leq l\leq L)$ layer are distributed into k_l rows, each being k'_l bytes long. Then the remaining n-k_l blocks in layer l are filled with parity bits generated by channel coding.

In the next section, we will discuss how to decide the protection degree of each video layer, $RS(n, k)$.

3.2 Bit allocation between the source and the channel

In the end-to-end video delivery scenario, we consider the global distortion, D_r, which is quite different from the notion, the quantization distortion, in the traditional compression domain. Under lossy environment like the Internet, the distortion D_r is a random variable, consisting of the source distortion D_s (due to the quantization noise) and the channel distortion D_c (due to the packet loss). Assuming that D_s and D_c are uncorrelated[16], the end-to-end distortion can be represented as $D_r = D_s + D_c$. Define the global distortion as the expectation of the random variable D_r. That is

$$D = E\{D_r\} = E(D_s + D_r), \tag{5}$$

where D_r takes the value of D_s or D_c with certain probability, which is decided by the network conditions. It is known that under a given network condition, additional FEC increases video robustness; however, this reduces the available rate for source coding. Thus there is a tradeoff between source coding and FEC. Based on the current network conditions, the key issue is how to distribute the resource between the source and the channel coding so as to alleviate the effect of the packet loss and achieve little global distortion.

For a given target rate R, we pick a fixed packet length P_L, then the number of packets $n = R/P_L$ can be determined. Define bar $\overline{k} = [k_1, \cdots, k_L]$; $\overline{k}' = [k_1{}', \cdots, k_L{}']$, where k_i is the number of the information blocks in layer i while k_i' is the length of this block by the unit byte. Therefore bit allocation problem can be formulated to find the optimal sets of $\overline{k}$ and $\overline{k}'$ aiming to minimize the end-to-end distortion:

$$(\overline{K}, \overline{K}') = \underset{(\overline{k}, \overline{k}')}{\arg\min}(D_s(R_s) + D_c(R_c))$$

$$\text{s.t.}\quad R_s + R_c \le R,\ R_c = \sum_{i=1}^{L}\left(\frac{n}{k_i} - 1\right)R_i,\ R_s = \sum_{l=1}^{L} R_l, \tag{6}$$

where R_l is the number of bits of layer l. In expression (6), the number of the layers L and the rate for the highest layer R_L depend on the source rate R_s.

The source distortion can be roughly calculated by $D_s(R_s) = A2^{-2R_s}$ with A being a constant. For the calculation of the channel distortion, it is a very difficult task to determine the accurate dependency among the layers. To simplify the distortion analysis and catch the nature characteristics in the scalable bit-stream, we have two assumptions:

- Each block of the same layer is equally important.
- Each block in the same layer has the same behavior of error propagation.

Besides, the channel coding across the packets can be regarded as an extension of source coding. Thus we do not differentiate original video block and parity block in the following formulations. Simulations show that this simplification is very effective and reasonable, which can give people the insight of the distortion-rate of the scalable video over the Internet.

With the number of lost packets, we first check if the first layer could be recovered, then see if the second layer could be recovered, etc. Let the RS codes for all the L layers be parameterized by (n, k_1), (n, k_2), $\cdots$, (n, k_L) with $k_1 \leq k_2 \leq \cdots \leq k_L$. Depending on the number of lost packets r ($0 \leq r \leq n$), define

$$c(r) = \arg \max_{j=0,\cdots,L} \{(n-k_j) > r\}. \tag{7}$$

That is, given the number of lost packets r, the first $c(r)$ layers can be correctly decodable and result in none channel distortion while other layers contribute r lost blocks to the channel distortion. Then the channel distortion can be calculated as

$$D_c(R_c) = \sum_{r=0}^{n} \left\{ P(r,n) \times r \times \sum_{i=c(r)+1}^{L} D(i) \right\} \Big/ n, \tag{8}$$

where L is the number of layers to be transmitted, $D(i)$ represents the distortion resulting from a lost block in layer i in the BOP, and $P(r, n)$ is the probability of r lost packets out of n packets. Therefore the global rate-distortion relationship can be expressed as

$$D(R) = A2^{-2R_s} + \sum_{r=0}^{n} \left\{ P(r,n) \times r \times \sum_{i=c(r)+1}^{L} D(i) \right\} \Big/ n. \tag{9}$$

Replacing (9) by (6), the bit allocation can be formulated as the optimal problem to minimize the global distortion, $E\{D(R)\} = E\{D_s(R_s) + D_c(R_c)\}$, under the available bandwidth constraints. That is:

$$(\bar{K}, \bar{K}') = \arg\min_{(\bar{k}, \bar{k}')} \left(A2^{-2R_s} + \sum_{r=0}^{n} \left\{ P(r,n) \times r \times \sum_{i=c(r)+1}^{L} D(i) \right\} \Big/ n \right)$$

$$\text{s.t.} \quad R_s + R_c \leq R,\ R_c = \sum_{i=1}^{L} \left(\frac{n}{k_i} - 1 \right) R_i,\ R_s = \sum_{l=1}^{L} R_l,\ P_L = \sum_{i=1}^{L} n_i'. \tag{10}$$

Although some simplification is made, it remains difficult to obtain the optimal sets of the rate. With many experiments, we have found that the end-to-end distortion surfaces are not convex with respect to the channel rate. Therefore iterated search algorithm should be adopted to solve this problem[16, 17], whose complexity remains $O(n^L)$ time. In the following, we give a sub-optimal bit allocation method to quickly decide the channel rate of each layer.

Exchanging the position of index *r* and *l*, the channel distortion can be equivalent to

$$D_c(R_c)=\sum_{r=0}^{n}\left\{P(r,n)\times r\times\sum_{i=c(r)+1}^{L}D(i)\right\}\bigg/n=\sum_{i=1}^{L}\left\{\sum_{r=n-k_i+1}^{n}P(r,n)\times r\times D(i)/n\right\}$$

$$=\sum_{i=1}^{L}D_{channle}(i), \tag{11}$$

where $D_{channel}(i)$ is the part of channel distortion related to layer *i*. In this formulation, channel distortion is decomposed into several independent parts on the condition that all the constraints in (10) hold true. Accordingly the operational distortion-rate is equal to $D(R)=D_s(R_s)+\sum_{i=1}^{L}D_{channel}(i)$.

For the optimal sets of the channel rates, the necessary condition is

$$\sum_{i=1}^{L}\frac{\Delta D_{channel}(i)}{\Delta R_i}\le\frac{\Delta D_s(R_s)}{\Delta R_s}, \tag{12}$$

which reflects the case that the parities are enough to protect the video data. To favor the robustness of the video streaming, we strengthen the requirements of the left part in expression (12)

$$\sum_{i=1}^{L}\frac{\Delta D_{channel}(i)}{\Delta R_i}\le\frac{\Delta D_s(R)}{\Delta R}=D_s'(R)\le\frac{\Delta D_s(R_s)}{\Delta R_s}, \tag{13}$$

where $D_s'(R)=\dfrac{\Delta D_s(R)}{\Delta R}$ is a constant. Since $D_{channel}(i)$ is only affected by *RS(n, k*$_i$*)*, we can reduce and simplify (10) to several small optimal problems:

$$\overset{\wedge}{k_L}=\arg\left(\frac{\Delta D_{channel}(L)}{\Delta R_i}\le D_s'(R)/L,\forall k_L\le\overset{\wedge}{k_L}\right)\ \text{s.t.}\ \overset{\wedge}{k_L}\le n,$$

$$\overset{\wedge}{k_i}=\arg\left(\frac{\Delta D_{channel}(i)}{\Delta R_i}\le D_s'(R)/L,\forall k_i\le\overset{\wedge}{k_i}\right)\ \text{s.t.}\ \overset{\wedge}{k_i}\le\overset{\wedge}{k_j},\ \ \forall i\le j, \tag{14}$$

where $\Delta D_{channel}(i)=P(n-k_i+1,n)\times(n-k_i+1)\times D(i)/n$, and $\Delta R_i=\dfrac{n}{k_i-1}R_i-\dfrac{n}{k_i}R_l$.

The parameter *L* in (14) can be roughly determined when the source rate equals the total bit rate *R*. Therefore optimal protection level for each layer can be got as simply as searching all the possible values. The total complexity is reduced only *O*(*L*×*n*) fold, which is much lower than that of formulation (10).

4 Simulation results

This simulation is to demonstrate effectiveness of our proposed end-to-end network adaptive video transmission scheme. In this simulation four schemes are tested: (1) Fixed ULP without interleaving which is denoted as FULP/I (40% protection for BL, 17% protection for EL); (2) Fixed ULP with interleaving which is represented as FULPI (36% protection for BL, 17% protection for EL); (3) Our Network-adaptive ULP with optimal bit allocation represented as OULP, including unequal error protection, unequal interleaving and optimal bit allocation, where iterative algorithm is used to search the optimal sets of the channel rate; (4) Our network-adaptive sub-optimal scheme SOULP, including unequal error protection, unequal interleaving and optimal bit allocation, where sub-optimal algorithm is adopted. Note that in all the schemes, congestion control is used to enhance the bandwidth efficiency.

In all the tested cases, the first frame was intra-coded and the remaining frames were inter-coded. The testing CIF video sequence is *Coastguard* at a temporal resolution of 10 fps. Simulations have been conducted under varying simulated network environment with bandwidth from 320 kbps to 1 Mbps and the corresponding packet loss ratio varying from 0.5% to 25%. To build a big performance picture of each scheme, we classify the range of packet loss ratio into three types: (a) high packet loss ratio ranging from 15% to 25%; (b) medium packet loss ratio ranging from 5% to 10%; (c) low packet loss ratio from 0.5% to 1%. The following results are all obtained by averaging over 30 Monte Carlo simulations.

Figs. 4 and 5 show the average PSNR with different available bandwidth under high packet loss ratio, medium packet loss ratio and low packet loss ratio network, respectively. On one hand, we can observe that compared with OULP with complexity $O(n^L)$ time, SOULP scheme achieves rather high quality video transmission with much less complexity $O(L\times n)$ time, only no more than 0.2 dB under different network conditions (different bandwidth and different packet loss case). On the other hand, compared with FULP, FULP/I, OULPI and SOULP can achieve much better performance under different channel conditions. Furthermore, the higher the channel rate, the larger gain is achieved between SOULP and the FULPI scheme. The explanation is that SOULP/ OULP scheme adjusts the protection degree of each layer based on the observed network conditions and its impact on the overall quality. For the scalable bit-stream, if error occurs in the lower layers, the corresponding bits in the higher layers can be regarded useless. In this sense, the invalid probability becomes larger for the information bits in the higher layers. In SOULP/OULP scheme, when the available bandwidth increases, the protection in the lower layers can be strengthened. Thus the valid probability of the higher layers increases accordingly. Therefore the performance of our scheme turns better in a rate-distortion manner. For the fixed scheme, the invalid probability of the bits is fixed no matter how the bit budgets are. As a result, their performance becomes steady after a small bit budget bounds. Under (a) type channel, SOULP achieves much gain over FULP with even more than 5.5 dB and 2.7 dB gain for *Foreman* and *Coastguard* respectively (at 960 kpbs bandwidth). Under type (b) channel, all the schemes can achieve rather good performance. Specially OULPI, SOULPI and FULPI obtain almost the same performance, which indicates this FULPI is fit for this network condition. When the network conditions turn better (see fig. 4(c) and fig. 5(c)), SOULP allocates more bits to the source rate so that the high quality initial video is obtained on the sender side, resulting in up to 2.0 dB and 0.9 dB gain over FULPI for *Foreman* and *Coastguard* sequence respectively.

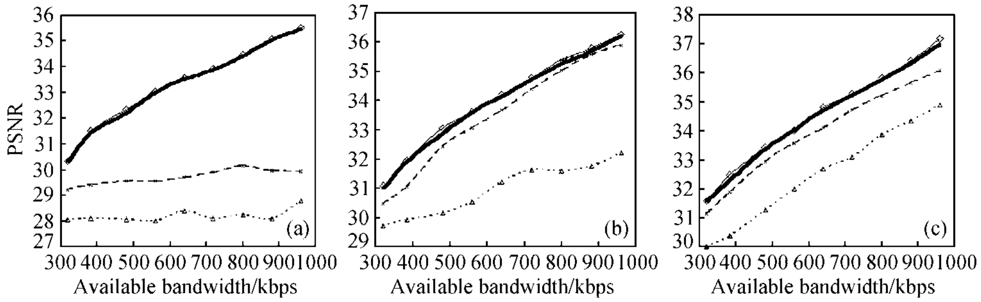

Fig. 4. Average PSNR for *Foreman* using different protection schemes under different bit rates. (a) High packet loss ratio; (b) medium packet loss ratio; (c) low packet loss ratio. ···△···, FULP/I; —∗—, FULPI; —◇—, OULPI; ——, SOULPI.

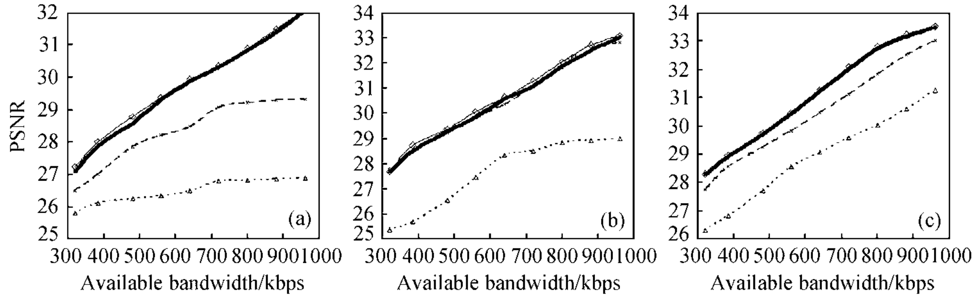

Fig. 5. Average PSNR for *Coastguard* using different protection schemes under different bit rates. (a) High packet loss ratio; (b) medium packet loss ratio; (c) low packet loss ratio. ···△···, FULP/I; —∗—, FULPI; —◇—, OULPI; ——, SOULPI.

In the following we analyze the performance of our interleaving packetization scheme. On one hand, by the interleaving packetization technique, the network noise can be randomized so that the receiver can recover the lost packets with FEC parities. On the other hand, each layer can get larger *RS* code than that in non-interleaving scheme so as to improve the efficiency of channel coding and system performance. To make it clear, let us study the two fixed protection schemes, FULPI and FULP/I. Compared with FULPI, the gain that FULPI achieved consists of two parts: source gain and channel gain. For the base layer bit-stream, with interleaving packetization technology FULPI can use 36% overhead in base layer bit-stream to guarantee nearly error-free transmission compared with that in FULP/I scheme, 40%. For the enhancement layers, the same protection overhead, 17%, is used in the two schemes. Thus the source gain is obtained from the 4% bit budgets of the base layer and the channel gain results from the interleaving packetization scheme. For example, in fig. 4(b), FULPI achieves 3.6 dB gain at 960 kbps bandwidth over FULP/I. Since the two schemes can combat the packet loss well under type (c) channel, we can refer the 1dB gain that FULPI achieved at any bandwidth to the source gain. Thus at this point (at 960 kbps bandwidth under type (b) channel), the gain resulting from the interleaving packetization is around 2.5 dB.

Several conclusions can be drawn observing the simulation results. First, the higher the network bandwidth, the more efficient our scheme compared with the other two fixed ULP schemes. The reason is that in our scheme we treat each layer with different priority/importance; meanwhile the varying channel condition is considered while the fixed ULP schemes cannot change with the channel conditions. Secondly, interleaving packetization introduces the benefits to the overall system. The gain resulting from this technology can be even up to 3 dB based on the network conditions. Thirdly, even in the good channel conditions, our scheme still can obtain much gain over the fixed scheme. The explanation is that SOULP can allocate more bits to the source rate so that the high quality initial video is obtained on the sender side.

Fig. 6 shows the comparison results of PSNR for the test sequence *Foreman* real-time streaming at 480 kbps channel rate using different error protection schemes. The horizontal axis is the frame number of the reconstructed video, which only lags the

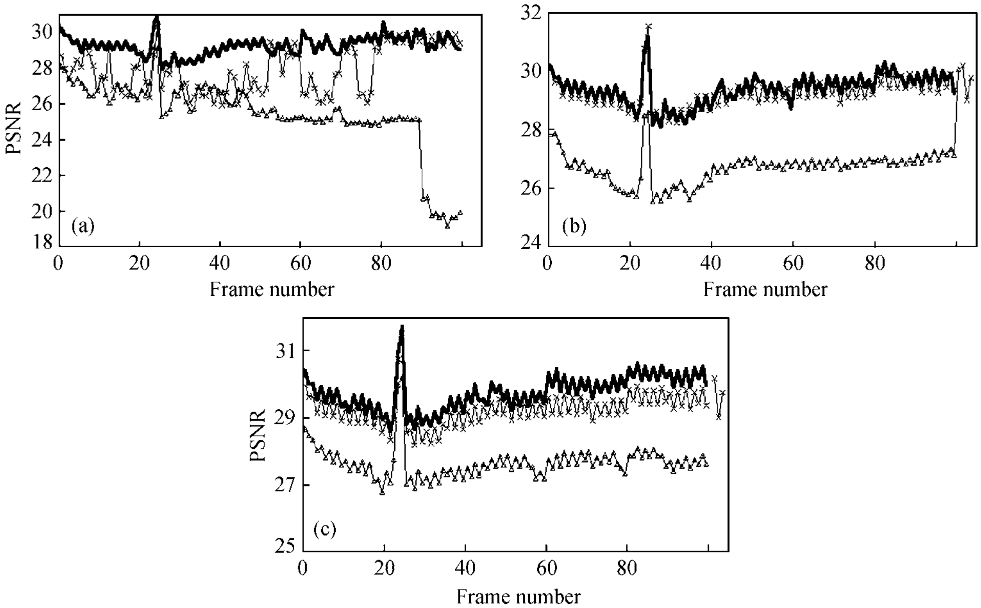

Fig. 6. The PSNR comparison results for *Foreman* real-time streaming using different protection schemes at 480 kbps. (a) High packet loss ratio; (b) medium packet loss ratio; (c) low packet loss ratio. —△—, FULP/I; —∗—, FULPI; ——, SOULPI.

sender by an application initial delay. In our scheme the initial delay is only one BOP packetization process, 2 seconds (20 frames in one BOP). It can be seen that our scheme outperforms the other three schemes. It also can be seen that the video quality achieved by our approach changes more smoothly, thereby being more comfortable for subjective feelings. In contrast, for the other fixed ULP schemes, the quality of the sequences is more fluctuant. Note that in our scheme we protect the BL strongly so that the video can be properly reconstructed. While for the fixed ULP schemes, the BL may be corrupted and cannot be concealed by the post processing. It may result in very poor quality and even lead to decoder crashing.

From figs. 4 to 6, it can be seen that our proposed end-to-end network-adaptive approach achieves rather good overall performance under different network conditions (different network bandwidth and different packet loss ratio) both subjectively and objectively.

5 Conclusions

It is becoming an important application for video streaming over the Internet, e.g. Video of Demand. The challenge in this application is how to cope with the network bandwidth fluctuation and packet loss so as to obtain the high-quality reconstructed video in the receiver. In this paper, we presented a new end-to-end network-adaptive robust approach for scalable video streaming over the Internet. The contributions of our work are:

● Consider all the key modules in the system simultaneously, including network-adaptive congestion control, unequal interleaving packetization, unequal error protection, and bit allocation.

● At the packet level, we present the unequal error protection with unequal interleaving packetization to alleviate the effect of the packet loss.

● Develop the rate-distortion theory for scalable video over the Internet, and give the optimal bit allocation scheme to decide the bit budgets between the source and channel so as to obtain high-quality reconstructed video.

Simulations show that our scheme can obtain global optimal and high-quality reconstructed video.

Acknowledgements This work is partially supported by NSFC (Grant No. 60072009), and was completed in MSRA.

References

1. Li, S. P., Wu, F., Zhang, Y.-Q., Study of a new approach to improve FGS video coding efficiency, ISO/IEC MPEG 50th Meeting, M5583, Mauri, USA, December 1999.
2. Li Weiping, Overview of fine granularity scalability in MPEG-4 video standard, IEEE Tran. on Circuits and Systems for Video Technology, 2001, 11(3): 301—317. [DOI]
3. Zhang, Q., Zhu, W., Zhang, Y.-Q., Network-adaptive rate control with TCP-friendly protocol for multiple video objects, in Proceedings of IEEE International Conference on Multimedia and Expo (ICME), July, 2000,

New York, vol. 2, 1055—1058.

4. Rejaie, R., Handley, M., Estrin, D., An end-to-end rate-based congestion control mechanism for realtime streams in the internet, in Proceedings of INFOCOMM, 1999, New York, vol. 3, 1337—1345.
5. Wu, D., Hou, Y. T., Zhu, W. et al., On end-to-end transport architecture for MPEG-4 video streaming over the Internet, IEEE Trans. on Circuits and Systems for Video Technology, 2000, 10(6): 923—941. [DOI]
6. Zhu, W., Zhang, Q., Zhang, Y.-Q., Network-adaptive rate control with unequal loss protection for scalable video over Internet, IEEE ISCAS, 2001, vol. 5, 109—112.
7. Mohr, A. E., Riskin, E. A., Ladner, R. E., Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction, IEEE Journal on Selected Area in Communications, June, 2000, 18(6): 819—828. [DOI]
8. Stuhlmuller, K., Link, M., Girod, G. et al., Scalable Internet Video Streaming With Unequal Error Protection, Packet Video Workshop, New York, 26/27, April 1999.
9. Zhang, T. T., Xu, Y., Unequal packet loss protection for layered video transmission, IEEE Trans. on Broadcasting, 1999, 45(2): 243—252. [DOI]
10. Schulzrinne, H., Casner, S., Frederich, R. et al., RTP: a transport protocol for real-time applications, RFC 1889, Internet Engineering Task Force, Audio-Video Transport Working Group, 1996.
11. Gringeri, S., Egorov, R., Shuaib, K. et al., Robust compression and transmission of MPEG-4 video, in Proceedings of the Seventh ACM International Conference on Multimedia, Orlando, Florida, USA, 1999, 113—120.
12. Bolot, J.-C., Turletti, T., Adaptive error control for packet video in the Internet, in Proc. IEEE ICIP’96, Sept. 1996, vol. 1, 25—28.
13. Thompson, K., Miller, G. J., Wilder, R., Wide-area Internet traffic patterns and characteristics, IEEE Network, 1997, 11(6): 10—23. [DOI]
14. Padhye, J., Firoiu, V., Towsley, D. et al., Modeling TCP throughput: A simple model and its empirical validation, in Proceedings of ACM SIGCOMM, Vancouver, CA, September 1998, 303—314.
15. Blahul, R. E., Theory and Practice of Error Control Codes, Reading, MA: Addison Wesley, 1983.
16. Stuhlmuller, K., Farber, N., Link, M. et al., Analysis of video transmission over lossy channels, IEEE Journal on Selected Areas in Communications, 2000, 18(6): 1012—1032. [DOI]
17. Cheung, G., Zakhor, A., Optimal bit allocation for joint source/channel coding of scalable video, IEEE Transactions on Image Processing, March 2000, 9(3): 340—357. [DOI]
18. Ramchandran, K., Ortega, A., Vetterli, M., Bit allocation for dependant quantization with applications to multiresolution and MPEG video coders, IEEE Trans. on Image Processing, Aug. 1994, 37: 533—545. [DOI]