An Adaptive Layered Hybrid ARQ Scheme for Scalable Video Transmission

Xi Zheng, Yonglin Xue, Hui Yang, and Jun Liu
Tsinghua National Laboratory of Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing, 100084, P. R. China
E-mail: zhengx06@mails.tsinghua.edu.cn

***Abstract*—In this paper, an adaptive layered hybrid ARQ scheme for scalable video transmission is proposed. The layered hybrid ARQ scheme is performed based on group of pictures (GOP). The sender dynamically adjusts the transmission decisions according to the feedback of transmission results and the channel condition. Meanwhile, the transmission time management is also adopted. The earliest start time and the latest end time of transmission procedure for each GOP are introduced to limit the buffer size and control the transmission delay. Simulation results show that the proposed scheme performs better than the conventional layered hybrid ARQ scheme in terms of average PSNR of reconstructed video in a wide range of packet loss rate over bandwidth-constrained networks.**

***Index Terms*—layered hybrid ARQ, scalable video, group of pictures, transmission time management**

I. INTRODUCTION

As the development of communication technology and popularization of personal terminal devices, video applications become ubiquitous in today's society. Thus, more and more researchers focus on the video coding and transmission technology. Considering the heterogeneous network conditions and different terminal capability, it remains a challenge to transmit video data reliably and effectively within limited bandwidth. Scalable video coding (SVC) provides a possible solution to this problem. The video data can be encoded to a bit stream containing several layers with temporal, spatial and fidelity flexibility. Partial bit streams with scalable video layers can also be decoded to videos of different temporal frame rate or spatial resolutions or quality fidelity. Consequently, SVC enables different quality of video service to be provided to different users according to the channel condition and the terminal capabilities and requirements [1].

Over networks without ARQ mechanism, the unequal error protection (UEP) method [2], [3] is naturally applied on scalable video layers of different importance. Forward error code (FEC) such as RS code is often used. In [2], unequal amount of FEC codes were allocated to different layers with different contributes to the video quality. The error propagation effect of scalable video layers is quantified by a performance metric namely layer-weighted expected zone of error propagation.

Compared to the conventional UEP method proposed in [2], the layered hybrid ARQ scheme [4], [5] is much more effective for scalable video transmission. On one hand, the layered hybrid ARQ scheme transmits scalable video layers sequentially from the base layer to the enhancement layers. The enhancement layers would not be transmitted until the sender is informed by an ACK from the receiver indicating that the former layer transmitted can be fully recovered, while the UEP scheme may transmit data of different layers in one packet alone, thus some data of enhancement layer received may not be recovered successfully over packet-lossy networks. On the other hand, the ACK from the receiver in the layered hybrid ARQ scheme indicates the real time packet loss rate, while the UEP method in [2] uses the predicated packet loss rate for FEC codes allocation. Therefore, the layered hybrid ARQ scheme is more suitable for the fast fading scenario. And this paper focuses on this scheme.

A dynamic hybrid UEP and ARQ method for scalable video transmission is proposed in [6]. It dynamically adjusts the transmission time budget of each GOP according to the feedback from the receiver, but the conventional UEP method was still used for scalable video layers. In this paper, an adaptive transmission scheme is proposed based on the layered hybrid ARQ mechanism. The number of scalable video layers in each GOP that would be transmitted is dynamically adjusted according to the feedback of transmission results and channel condition from the receiver. Meanwhile, limited buffer size and transmission delay are guaranteed by the transmission time management so that the proposed scheme is practical for video applications.

The rest part of this paper is structured as follows. An overview of the scalable video coding and layered hybrid ARQ scheme is presented in section II. In section III, an adaptive layered hybrid ARQ scheme is proposed for scalable video with transmission time management. Section IV gives and analyzes the simulation results. Finally the paper is concluded in section V.

Manuscript received April 12, 2013; revised May. 30, 2013.

doi: 10.12720/ijeee.1.3.160-164

II. Layered Hybrid ARQ Scheme for Scalable Video

A. Scalable Video Coding

SVC enables video data to be encoded to a single bit stream which support temporal, spatial and quality scalability. SVC extension of H.264/AVC standard is considered in this paper. Fig. 1 shows the case of combined temporal and quality scalability.

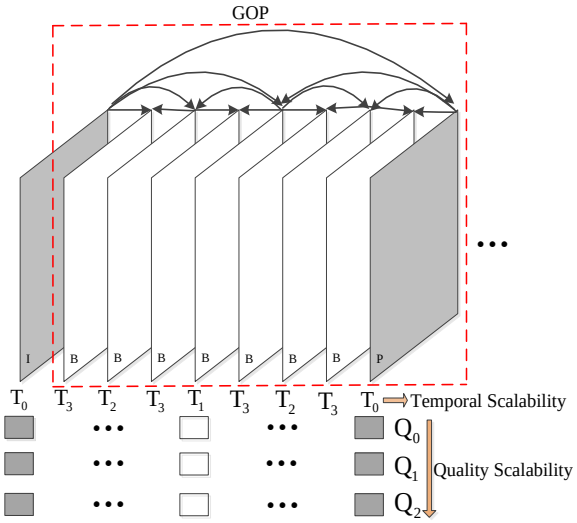


Figure 1. Structures of H.264 SVC with combined temporal and quality scalability

As illustrated in Fig. 1, a GOP consists of eight pictures, one of which is a key picture. The key pictures which are often coded as I or P-pictures form the base layer set T_0. Other pictures in a GOP, which are encoded as B or P-pictures according to the prediction relationship, form the enhancement layer set $T_k(k>0)$. The hierarchical prediction structures support the temporal scalability. Moreover, each picture in a GOP is encoded to several quality layers. The set Q_0 consists of the quality base layers and $Q_k(k>0)$ consists of the quality enhancement layers. The subscript k in T_k and Q_k represents the temporal and quality level identifier. The maximum temporal level is 4 and the maximum quality level is 3 respectively as shown in Fig. 1. The video sequence is thus encoded to scalable units (SU) of different temporal and quality levels. According to the hierarchical prediction structures, the temporal enhancement layer can only be decoded after the temporal base layer and enhancement layers with less temporal level identifier have been decoded. Similarly, the quality enhancement layers gradually improve the quality of the reconstructed video.

B. Layered Hybrid ARQ Scheme

In general, the layered hybrid ARQ scheme is shown as Fig. 2. The source video is encoded to L layers of different importance. The importance of source data decreases from layer 1 to layer L. Each layer consists of some source packets and additional FEC packets, assuming that the layer k $(1 \le k \le L)$ is encoded to l_k source packets and p_k FEC packets as shown in Fig. 2.

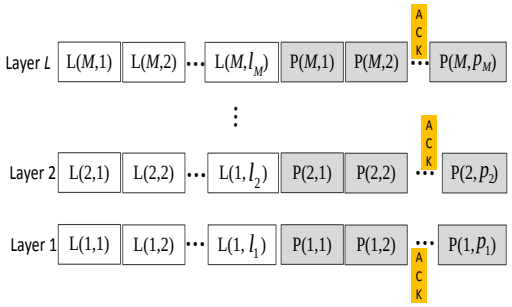


Figure 2. Layered hybrid ARQ scheme

The packets are sequentially sent from layer 1 to layer M. The receiver would responses with an ACK only if it has received enough packets to recover a video layer. When the sender receives an ACK, it will skip the remaining FEC packets of the current layer and start to send the packets of higher layer to the receiver.

The layered hybrid ARQ scheme is usually applied in a round fashion. The duration of a round is T. The round will be ended when the duration time of this round is exhausted, followed by the procedure of next round. When the sender receives an ACK indicating that all layers of this round can be fully recovered on the receiver side, the sender will stop transmitting the rest of the packets and stay idle till the next round start. It is obvious that the number of FEC packets transmitted for each layer depends on the explicit response from the receiver, i.e., ACK, and no packets of higher layer will be sent until all lower layers can be recovered, which means the bandwidth is utilized very effectively.

III. Adaptive Hybrid ARQ Scheme for Scalable Video Transmission

A. Analysis on the Conventional Layered Hybrid ARQ Scheme for Scalable Video Transmisson

As mentioned above, the conventional layered hybrid ARQ scheme sequentially transmits the packets according to the ascending order of layer level number until all layers are completely received by the user or the duration time is exhausted.

For convenience, the duration time of a round is defined as

$$T = G / R_f \tag{1}$$

where G represents the GOP size and R_f represents the frame rate of the video sequence. The target of a transmission round is to send all layers of a GOP to the receiver. The sender will start to transmit the packets of the next GOP when the current round is over no matter whether the user has received enough packets to recovery all layers of the GOP. In this paper, the encoded data of a GOP is divided into different layers according to the temporal and quality level identifier. To be specific, the SU of temporal level identifier i and quality level identifier j is defined as SU(i, j), then the layer level number for SU(i, j), denoted by l, can be calculated as

$$l = j \cdot L_T + i \tag{2}$$

where $0 \le i \le L_T - 1,\ 0 \le j \le L_Q - 1,\ i, j \in \mathrm{N}$, L_T and L_Q represents the maximum number of temporal and quality level respectively. Thus, the maximum number of layers of a GOP denoted as *L* can be calculated as

$$L = L_T L_Q \tag{3}$$

In the conventional layered hybrid ARQ scheme, the latest layer transmitted during a round may not be recovered on the receiver side because of packet loss, thus partial packets received will make no contribute to recovery the video layers. In addition, when the sender receives an ACK which indicates that the user can completely decode all video layers, it will stay idle till next round. This also means a waste of the bandwidth.

B. The Proposed Adaptive Transmission Method

Considering the defect of conventional layered hybrid ARQ scheme analyzed above, an adaptive transmission method for scalable video is proposed. In the proposed scheme, the sender makes a decision of whether the packets of next layer will be sent once an ACK is arrived. The decision depends on the probability of recovery of the next layer on the receiver side taking account of the channel condition and the bandwidth.

Assuming that the bandwidth is *B* and the length of a packet is *M*, then M / B represents the transmission time for a single packet and the maximum number of packets transmitted within duration time *t* can be calculated as

$$N_t = \left\lfloor \frac{B \cdot t}{M} \right\rfloor \tag{4}$$

As the sender transmits the packets of a layer, the decrease of the rest time of the round also means the decrease of the number of packets that can be transmitted within the current round. Suppose the FEC packets are always abundant for each layers, let *p*(*m*, *n*) represent the probability of *m* lost packets within *n* packets transmission, then the probability of recovery of layer *k* on the receiver side after the layers from layer 1 to layer *k*-1 have already be recovered, denoted by P_k, can be calculated as

$$P_k = \sum_{n=0}^{N_{left} - l_k} p(n, N_{left}) \tag{5}$$

where l_k represent the number of source packet of layer *k* and N_{left} derived from (4) represents the number of packets that can be transmitted within the rest of time of the current round.

The sender decides whether to sends the packets of layer *k* according to the probability P_k considering the channel condition and δ_k is used to indicate the transmission decision of the layer *k*. It is defined as

$$\delta_k = \begin{cases} 0 & (P_k \le p_T) \\ 1 & (P_k > p_T) \end{cases} \tag{6}$$

where $p_T \in [0,1]$ is introduced as the threshold probability. $\delta_k = 1$ means the packets of layer *k* will be sent whereas $\delta_k = 0$ means the opposite.

It is clear from the above that the duration time for a round is not fixed as *T* in the proposed scheme. Once the sender decides not to send the packets of a layer, no more packets of higher layer will be sent and the current round ends. Moreover, considering that the ACK indicating the user has received enough packets to recovery all video layers means no more packets are needed to be sent, the sender will start the next round instead of staying idle until the time of the current round runs out. Thus, the saved time of previous rounds can be used to increase the throughput of the following rounds, and more layers are probable to be recovered. To limit the buffer size and control the transmission delay, the transmission time management is also adopted for the transmission procedure. The earliest start time and the latest end time denoted as T_E and T_L respectively are introduced for each round. To be specific, the sender will not start the transmission procedure until the moment T_E and it will stop the procedure before the moment T_L. Of course, both T_E and T_L increase after each round.

To sum up, the procedure for each round works as follows:

(a)If $t < T_E$ now, the sender stays idle till the moment T_E. Otherwise, the sender should start a round immediately.

(b)At the beginning of a round or after the sender receives an ACK indicating that enough packets of the current layer (not the last layer)have been received for recovering the video layer, calculate the probability P_k as in (5), and decide whether to send the packets of the next layer according to (6) with given threshold probability p_T.

(c) Each round ends if (i) the sender decides not to send the packets of the next layer, or (ii) the receiver can fully recovery all the video layers, or (iii) the total time of this round is exhausted, i.e., $t = T_L$. Once a round is over, both T_E and T_L increase by *T* for the next round.

IV. Simulation Results

JSVM9.19.12 [7] is employed to conduct the simulations. Four different types of QCIF video sequences, 'Harbour', 'Soccer', 'Crew' and 'Foreman', are encoded to bit streams applying the combined temporal and quality scalability. The GOP size is set to eight and the number of quality enhancement layers to three. Thus L_T = 4 and L_Q = 4 respectively and there are total L = 16 layers during a round. The frame rate of all the video sequences is 30 frames per second. The average duration time *T* of a round for transmitting a GOP can be calculated from (1).

The two-state Markov model [8] which is widely used is employed to simulate the packet loss channel. Different channel condition represented by different packet loss rates (PLR) and average burst lengths of packet losses are tested as shown in Table I. For a given packet size *M*, the maximum number of packets which can be transmitted during a round is denoted as *N*. From (4), it is obvious that the constrained-bandwidth *B* also means that *N* is limited while *M* is fixed. Thus *N* is set to different numbers to represent different bandwidth. Table II shows the *M* and *N* set for the tested video sequences.

TABLE I. CHANNEL CONDITION

Channel condition index	1	2	3	4	5	6
PLR	1%	5%	10%	15%	20%	25%
Average burst length	2	3	4	4.5	5	5.5

TABLE II. SIMMULATION SETUP PARAMETERS OF UEP

Sequence	**Packet size (*M*)**	**Number of packets (N)**		**Average bytes per GOP**
		High bandwidth	**Low Bandwidth**	
Harbour	260	140	90	30466
Soccer	200	155	100	25992
Crew	200	145	95	23947
Foreman	200	125	80	20710

The initial earliest start time and the latest end time are set as $T_E = t - 4T$ and $T_L = t + T$ at the very beginning time *t* of transmission. T_E and T_L will increase by *T* at the end of each round. Specially, the initial state of $T_E = t$ and $T_L = t + T$ is equivalent to that of the conventional layered hybrid ARQ scheme in terms of transmission time management. In addition, the threshold probability introduced for transmission decisions in (6) is set as $p_T = 0.2$.

The average PSNR performance of the proposed method is compared with the conventional layered hybrid ARQ scheme. Both methods are tested over different channel condition and bandwidth with four types of video sequences. All situations are simulated for 100 times. The simulation results are shown in Fig. 3. The proposed scheme shows good adaptability in different transmission cases. The PSNR of reconstructed videos for 'Harbour', 'Soccer', 'Crew' and 'Foreman' is respectively improved by 0.46dB, 0.43dB, 0.40dB and 0.66dB on average over different bandwidth and packet loss rates. Additionally, the average amount of received layers of each GOP is compared. By applying the proposed scheme, the sender can transmit more layers of a GOP during a round. As shown in Fig. 4, compared with the conventional method, 1.23 more layers on average are received every GOP for 'Foreman' when the packet loss rate is 1% and *N* equals 80 if the proposed method is adopted, which also leads to 0.20dB gain of PSNR.

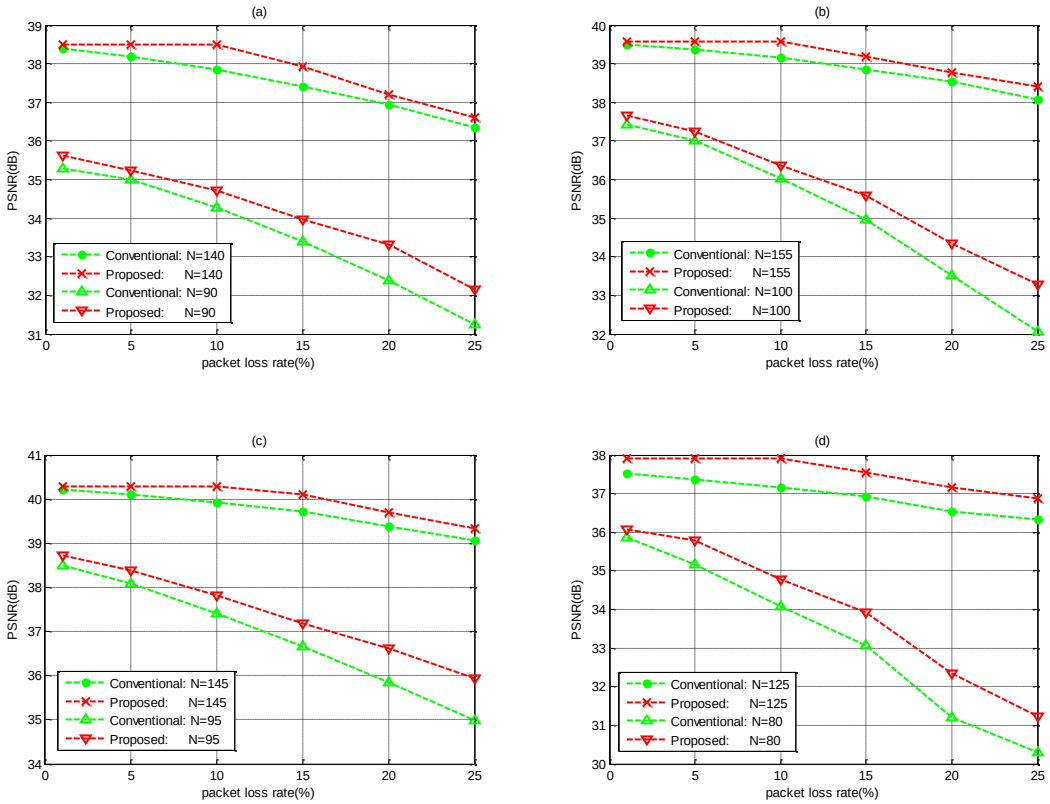


Figure 3. Average PSNR performance on four types of video sequences: (a) Harbour (b) Soccer (c) Crew (d) Foreman

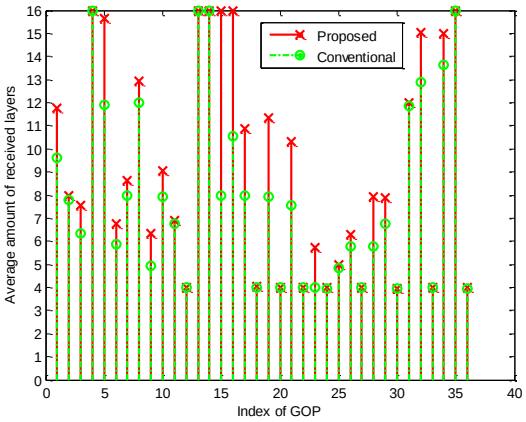


Figure 4. Average amount of received layers as a function of GOP index for 'Foreman' when the packet loss rate is 1% and N equals 80

V. Conclusion

In this paper, an adaptive transmission scheme for scalable video is proposed based on the layered hybrid ARQ scheme. The sender transmits the packets of a video layer according to the probability of recovery of this layer on the user side. And the duration time of each round is adjusted according to the transmission results of the former and the current round. Meanwhile, the transmission time management is adopted to limit the buffer size and control the transmission delay on the sender side for practical video applications. The proposed scheme was tested for different types of video sequences over a wide range of bandwidth and packet loss rates, and the simulation results show that the proposed adaptive transmission scheme for scalable video outperforms the conventional layered hybrid ARQ scheme in terms of average PSNR of reconstructed videos.

Acknowledgment

This work was supported by the National High Technology Research and Development Program of China (Grant No. 2012AA011704).

References

[1] 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.

[2] H. Ha and C. Yim, "Layer-weighted unequal error protection for scalable video coding extension of H.264/AVC," *IEEE Trans. Consumer Electron.*, vol. 54, no. 2, pp. 736-744, May. 2008.

[3] E. Maani and A. K. Katsaggelos, "Unequal error protection for robust streaming of scalable video over packet lossy networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 3, pp. 407–415, Mar. 2010.

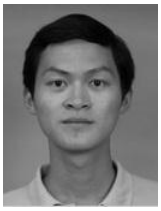
[4] Z. Liu, Z. Wu, H. Liu, M. Wu, and A. Stein, "A layered hybrid-ARQ scheme for scalable video multicast over wireless networks," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2007, pp. 914-919.

[5] Z. Liu, Z. Wu, P. Liu, H. Liu, and Y. Wang, "Layer bargaining: multicast layered video over wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 445 –455, Apr. 2010.

[6] J. Liu, Y. Zhang, X. Zheng, and J. Song, "A dynamic hybrid UXP/ARQ method for scalable video transmission," in *Proc. IEEE International Symp. on Personal, Indoor and Mobile Radio Comm.*, Sep. 2012, pp. 2566–2571.

[7] JSVM (Joint Scalable Video Model) software manual 9.19.12, March 29th, 2011.

[8] A. Majumdar, D. G. Sachs, I. V. Kozintsev, K. Ramchandran, and M. M. Yeung, "Multicast and unicast real-time video streaming over wireless LANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 524-534, June. 2002.



Xi Zheng was born in Hunan Province, PRC in October 1988. He received the Bachelor degree from Electronic Engineering Department of Tsinghua University, Beijing China, in 2010 and is now pursuing the Master degree at the DTV Technology R&D center of Tsinghua University. His research interests include video coding and communication.

Yonglin Xue was born in Shaanxi Province, PRC in November 1965. He got the Bachelor and Master degrees from Electronic Engineering Department of Tsinghua University and Chinese Academy of Sciences in 1989 and 1992. He is now associate professor of Research Institute of Information Technologies of Tsinghua University with major research interests in the digital TV broadcasting, video coding and communication. He has published more than 50 academic papers and received the first award from Chinese Institute of Electronics.

Hui Yang was born in Guangxi province, PRC in 1967. He got the Bachelor and Master degrees from Electronic Engineering Department of Tsinghua University in 1989 and 1992.He is now a senior engineer in Research Institute of Information Technology of Tsinghua University. His research interests include Digital TV transmission technology, power line communication (PLC) and Visible Light communication (VLC).

Jun Liu, born on February 23, 1988, received the B. E. and M. S. degrees from the School of Electronics and Information Engineering in Harbin Institute of Technology, Harbin China, in 2008 and 2010, respectively. Currently, he is pursuing the Ph. D. degree at the DTV Technology R&D Center, Tsinghua University. His research interests include wireless multimedia transmission, energy-efficient scheduling and resource allocation.