

# Receiver-Driven Adaptive Enhancement Layer Switching Algorithm for Scalable Video Transmission Over Link-adaptive Networks

Jian Yang, *Member, IEEE*, Quan Zheng, Hongsheng Xi, and Lajos Hanzo, *Fellow, IEEE*

**Abstract**—A receiver-driven adaptive layer switching algorithm is proposed for adapting the video bitrate to match the achievable network throughput. It relies on a QoS-constrained equivalent bandwidth estimator employed at the receiver, which is used for triggering the adjustment of video layers at the video source. Simulations are conducted to illustrate its efficiency by showing that it is capable of accommodating different channel qualities without their prior knowledge.

**Index Terms**—Quality-of-service (QoS), scalable video coding, video layer switching, wireless networks.

## I. INTRODUCTION

As benefit of the wide-spread roll-out of broadband wireless services and advances in video compression, it has become feasible to provide high-quality digital video services for mobile devices. However, supporting real-time video services under strict Quality-of-Service (QoS) constraints still poses challenges in the face of rapidly time-varying wireless channel qualities. Hence, prompt video bitrate adaptation is a fundamental functionality required for accommodating the time-varying network conditions. Scalable Video Coding (SVC) [1] provides a flexible way of controlling the video bitrate and it is deemed to be an effective solution for video streaming over a channel exhibiting a time-varying throughput.

Most previous works [2]–[5] concentrated on cross-layer optimization for transmitting scalable video streams from a base station (BS) to mobile clients by jointly considering both application layer information and the wireless channel conditions. However, in practical applications the video data may be conveyed both via wired networks like the Internet besides the wireless link. Hence, a bottleneck at any network nodes may influence the end-to-end QoS of the video. Hence it is a challenge to create an accurate network-induced-distortion model for such

end-to-end video transmission problems. Another challenge in applying cross-layer optimization to video transmission is that appropriate modification of the protocols is necessary, which is bound to increase the network's complexity and a cost required for upgrading the communication devices.

In order to guarantee the video quality on an end-to-end basis, feedback-based rate control is necessary for video streaming services. An end-to-end application-layer rate-control mechanism [6] was proposed based on the receiver's buffer-starvation probabilities analyzed using a discrete-time Markov chain. The work in [7] discussed the client intelligence required for adaptive video streaming based on progressive download over HTTP, like in HTTP Live streaming [8] and in 3GPP adaptive HTTP streaming [9]. In this letter, we conceive a receiver-driven adaptive layer-switching scheme for optimizing scalable video transmissions in the context of a state-of-the-art network architecture and protocols. Adaptive Modulation and Coding (AMC) as well as Hybrid Automatic Request (HARQ) are widely used for link adaptation in wireless communications, while ARQ-based error control may be applied for wired networks at the link layer. The employment of these link adaptation schemes results in a fluctuating throughput for each video streaming session, which requires dynamic layer control for scalable video transmission. In order to achieve this, we rely on instantaneous throughput measurements for formulating an estimation model for a QoS-constrained end-to-end equivalent bandwidth, which is applied to instruct the video source to appropriately configure the video layers for transmission. In comparison to previous works [6], [7], the advantage of such a measurement-driven scheme is that it can eliminate the need for a complex, yet potentially inaccurate model of the network traffic. Hence, it can be invoked where no accurate traffic model is available. This letter is organized as follows. Section II describes the system model. The layer-switching strategy based on equivalent throughput estimation is derived in Section III, while Section IV characterizes the performance of our scheme. Finally, our conclusions are summarized in Section V.

## II. SYSTEM MODEL

Consider the wireless streaming system depicted in Fig. 1, which consists of a streaming server, the Internet, base transceiver station (BTS), the wireless channel as well as the mobile station (MS). A multimedia streaming service requires not only maintaining the wireless connection between the MS and BTS, but also access to the Internet.

Manuscript received August 27, 2012; revised October 25, 2012; accepted October 30, 2012. Date of publication November 16, 2012; date of current version November 28, 2012. This work was supported by National Foundation of Natural Science of China under Grant 61174062. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. H. Vicky Zhao.

J. Yang, Q. Zheng, and H. S. Xi are with the School of Information Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China (e-mail: jianyang@ustc.edu.cn; qzheng@ustc.edu.cn; xihs@ustc.edu.cn).

L. Hanzo is with the School of Electronics and Computer Science (ECS), University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/LSP.2012.2227473

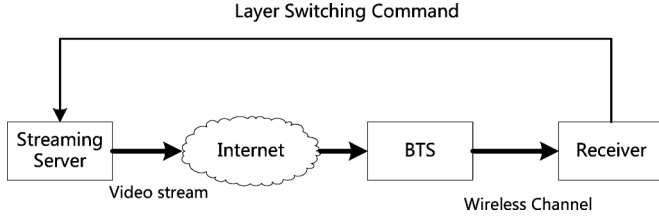


Fig. 1. System architecture.

We also consider a scalable video stream that consists of a base layer,  $l_1$ , and  $(n-1)$  enhancement layers,  $\{l_2, l_3, \dots, l_n\}$ . We assume that the available rates of the input video are  $\{R_1, R_2, \dots, R_n\}$ .

At the receiver, the transmission rate associated with each video frame is measured and recorded, which is referred to as the instantaneous network throughput available to the user. Let  $r_i$  denote the instantaneous throughput experienced, when conveying the  $i$ th video frame. Since the network resources are shared by a time-varying number of users, the dynamic resource scheduling scheme employed by the BTS may result in a burst-by-burst adaptive affordable user-throughput. Additionally, sophisticated link adaptation mechanisms based on AMC/HARQ/ARQ are widely applied, which lead to a burst-by-burst adaptive channel throughput. Hence, we assume the network throughput sequence  $r_i$  to be i.i.d. Without loss of generality, let  $r_1, r_2, \dots, r_m$  denote the most recent  $m$  measurements. Furthermore, let us denote the total throughput of the  $m$  most recent total throughput measurements and their average by  $S_m = \sum_{i=1}^m r_i$  and  $\bar{r} = S_m/m$ , respectively. We aim for finding a QoS-constrained equivalent throughput  $C$ , which satisfies the probability constraint:

$$P(\bar{r} \leq C) \leq \varepsilon \quad (1)$$

where  $\varepsilon$  is a desired probability threshold. If we apply the equivalent throughput value  $C$  as the video source bitrate in context of our video rate control problem,  $P(\bar{r} \leq C)$  can be used as a quantitative indicator of how much the video bitrate exceeds the channel bitrate. Naturally, having a smaller  $\varepsilon$  implies a lower risk of the video source rate exceeding the affordable channel rate.

### III. EQUIVALENT THROUGHPUT ESTIMATION BASED LAYER SWITCHING

In this section, we propose an equivalent throughput estimation model and then derive our video layer switching algorithm.

#### A. Equivalent Throughput Estimation

Below, we will apply the Hoeffding Theorem [10] for estimating the equivalent throughput  $C$  of the throughput sequence  $r_i$ . According to the Hoeffding theorem, if the independent random variable  $r_i$  satisfies  $a_i \leq r_i \leq b_i$ , for any  $t > 0$ , then we have:

$$P(|S_m - E[S_m]| \geq mt) \leq 2e^{-2m^2t^2 / \sum_{i=1}^n (b_i - a_i)^2}, \quad (2)$$

where  $E[\cdot]$  denotes expectation operator. Let  $\mu = E[S_m]/m$ . Then, (2) may be rewritten as

$$P(|\bar{r} - \mu| \geq t) \leq 2e^{-2m^2t^2 / \sum_{i=1}^n (b_i - a_i)^2}. \quad (3)$$

Let us set the expression at the right of (3) to  $\varepsilon$  as follows:

$$\varepsilon = 2e^{-2m^2t^2 / \sum_{i=1}^n (b_i - a_i)^2}. \quad (4)$$

Then, we have:

$$P(|\bar{r} - \mu| \geq t) \leq \varepsilon. \quad (5)$$

According to (4),  $t$  should satisfy:

$$t = \frac{1}{m} \sqrt{\frac{\ln\left(\frac{2}{\varepsilon}\right) \sum_{i=1}^m (b_i - a_i)^2}{2}}. \quad (6)$$

Let  $C = u - t$ . Then, we have:

$$P(\bar{r} \leq C) = P(\bar{r} - u \leq -t) \quad (7)$$

$$\leq P(|\bar{r} - u| \geq t) \leq \varepsilon, \quad (8)$$

which implies that (1) is satisfied. Hence, the equivalent throughput  $C$  can be estimated as follows:

$$\hat{C} = \mu - \frac{1}{m} \sqrt{\frac{\ln\left(\frac{2}{\varepsilon}\right) \sum_{i=1}^m (b_i - a_i)^2}{2}}. \quad (9)$$

Since  $r_i$  is i.i.d, we can estimate  $a_i$  and  $b_i$  as:

$$a_i = \min_{1 \leq j \leq m} r_j, \quad b_i = \max_{1 \leq j \leq m} r_j. \quad (10)$$

#### B. Adaptive Layer Switching Algorithm

Since the equivalent throughput  $\hat{C}$  satisfies (1), we can maintain a video QoS of  $\varepsilon$ , provided that the video bitrate does not exceed  $\hat{C}$ . Therefore,  $\hat{C}$  can be used as a threshold to trigger layer switching for the video transmission. In order to avoid substantial and hence objectionable video quality variations, we constrain the number of layers  $l$  to simply increase or decrease by one, say from the previous level  $l_k$  to  $l \in \{l_{k+1}, l_k, l_{k-1}\}$ .

Let us assume that the current video layer is  $k$ . Then, the adaptive layer switching algorithm operates in the context of three scenarios as follows:

- 1)  $k > 1$  and  $\hat{C} < R_k$ : Since  $\hat{C}$  is the estimated equivalent throughput, the current video source bitrate  $R_k$  cannot ensure the QoS-guarantee of  $\varepsilon$  for the transmission of the  $k$ -layer video. Hence, in order to satisfy the channel bitrate constraint, we may reduce the number of video layers according to  $k := k - 1$ , i.e., we discard one of the enhancement layers.
- 2)  $k < n$  and  $\hat{C} > R_{k+1}$ :  $\hat{C} > R_{k+1}$  implies that the wireless channel has the capacity to convey the  $(k+1)$ -layer video. Therefore, in order to improve the video quality and to beneficially exploit the currently high channel throughput, we

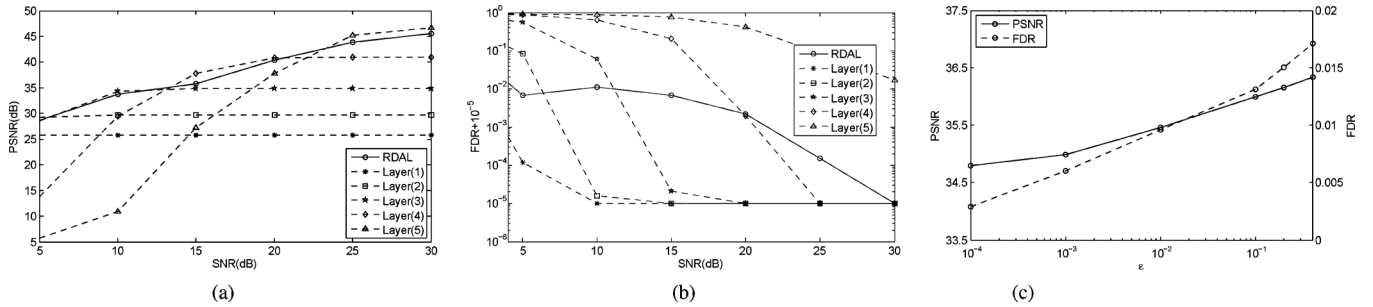


Fig. 2. (a) and (b): Simulation results for different average channel SNRs of Rayleigh channel. (c): Performance sensitivity at average channel SNR  $\gamma = 12$  dB. (a) PSNR. (b) FDR. (c) Performance sensitivity to  $\epsilon$ .

increase the number of video layers according to  $k := k + 1$  if we have  $k < n$ , while ensuring that  $P(\bar{r} \leq R_{k+1}) < \epsilon$ .

- 3) Otherwise: We maintain the current video bitrate without adjusting the number of video layers.

Therefore, the adaptive enhancement layer switching algorithm can be summarized as follows:

$$k = \begin{cases} (k - 1) & \text{if } k > 1 \text{ and } R_k > \hat{C} \\ (k + 1) & \text{if } k < n \text{ and } R_{k+1} < \hat{C} \\ k & \text{otherwise} \end{cases} \quad (11)$$

We can estimate  $\hat{C}$  and adjust the number of video enhancement layers during every  $T$  time slot. However, if its complexity is affordable, the algorithm may be activated more frequently for a prompt response to the channel-throughput variations.

### C. Discussions

Generally, the attainable video quality is significantly affected by network congestion in the Internet. In case of video streaming over HTTP, any potential network congestion may be relieved by the end-to-end flow control carried out by TCP or a TCP-friendly rate control protocol (TFRC) at the transport layer. However, this is a basic scheme conceived for packet-transmission rate control, not for video streaming. Our proposed enhancement layer switching scheme operates as an end-to-end rate control at the application layer, which may be viewed as a video rate control scheme designed for adapting the video streaming to the time-varying network conditions.

TFRC and Multiple TFRC (MULTFRC) utilize the parameters Round-Trip and the packet drop rate to estimate the achievable transmission rate, which may also be applicable to our proposed scheme. In this letter, we conceive a technique of estimating the available throughput: the video content is partitioned into a set of time-bounded segments referred to as *chunks*. Let  $B_i$  and  $T_i$  respectively denote the chunk size and the transmission duration (from the first bit of the chunk received by the receiver to the last one received). Then we have  $r_i = B_i/T_i$ .

## IV. PERFORMANCE EVALUATION

In this section, we characterize the performance of the proposed receiver-driven adaptive enhancement layer switching algorithm (RDAL). We consider a High Speed Packet Access (HSPA) network [11] relying both on AMC as well as on HARQ. The Channel Quality Indicator (CQI) mapping table

of the UE category 10 [12] is applied. According to the specification [12], the Transmission Time Interval (TTI) has 2 ms duration, which facilitates near-real-time HARQ transmission. The maximum number of concurrently communicating users was set to  $U = 32$  and the user arrival process follows a Markov process with an arrival rate of  $\lambda = 10^{-4}$  per TTI as well as a service rate of  $\nu = 1.25 \times 10^{-5}$  per TTI. A Round Robin scheme was applied for scheduling the transmissions of the users. Consider an i.i.d. Rayleigh channel for the target user, where the received SNR,  $s$ , has a probability density function (PDF) of  $f(s) = (1/\gamma)e^{-s/\gamma}$  having an average of  $\gamma$ . We applied the SNR(dB)-to-CQI mapping rule of [13], given by  $CQI = \lfloor SNR + 4.5 \rfloor$ . The video sequence used in our simulations is the Sony Demo clip from the Video Trace Library [14] scanned at 30 frame/s. It has one base layer and four enhancement layers. The duration of the video sequence is about 10 minutes. In order to satisfy the tolerable delay constraint of a specific video application, we dropped the video frames, when their delay exceeded 120 ms. The average Peak Signal-to-Noise ratio (PSNR) and Frame Dropping Ratio (FDR) are used for characterizing the performance of the proposed method. All the results were averaged over 50 independent simulation runs.

In our first experiment, we compared the proposed algorithm to the conventional fixed video profile strategy, where the number of video layers is constant during video transmissions. Let  $layer(n)$  denote the transmission policy associated with having one base and  $(n - 1)$  enhancement layers. In order to characterize the adaptive layer switching capability for different channel qualities, we conduct experiments at different average channel SNRs. We set  $\epsilon = 0.01$  for the proposed algorithm and used a sliding window size of  $m = 800$ . In Fig. 2(a), the average PSNR of the proposed algorithm approaches the highest average PSNR of the conventional fixed layer strategies, because our proposed algorithm has the capability of adjusting the number of enhancement layers according to the affordable network throughput: at a lower network throughput, RDAL may reduce the number of enhancement layers for achieving a lower FDR, as shown in Fig. 2(b), while keeping the video quality as high as possible. By contrast, at a higher network throughput, it may increase the number of video enhancement layers. Finally, for a fixed number of layers the achievable PSNR is forced to remain low at high channel throughputs. As a benefit, however, transmitting a reduced number of enhancement layers may achieve a lower FDR at a low channel SNRs, as shown in Fig. 2(b). Moreover, having a higher than

affordable number of enhancement layers implies a higher video bitrate, which results in more dropped frames at lower channel qualities, as evidenced by Fig. 2(b), which inevitably leads to a low average PSNR. Fig. 2(b) also shows that RDAL is capable of reducing the FDR below 0.01, thus meeting our QoS guarantee of  $\varepsilon = 0.01$ . Hence, these results demonstrate that the proposed algorithm is capable of accommodating different channel qualities by controlling the number of enhancement layers without any prior knowledge of the channel quality.

Finally, we conducted a further experiment for investigating the sensitivity of the proposed algorithm to the probability threshold  $\varepsilon$ . The channel SNR was set to 12 dB and the remaining parameters were the same as those in the above experiments. The results of Fig. 2(c) demonstrate that a reduced probability threshold  $\varepsilon$  provides a reduced FDR. This is because according to (9), a smaller  $\varepsilon$  leads to a lower equivalent throughput  $C$ , hence the number of enhancement layers is also reduced, which inevitably reduces the average PSNR, as seen in Fig. 2(c). Hence, having a lower  $\varepsilon$  implies a more conservative enhancement layer switching algorithm.

## V. CONCLUSIONS

In this letter, a receiver-driven adaptive enhancement layer switching algorithm has been proposed for scalable video transmission, which is a measurement-driven scheme capable of efficiently operating, where no traffic model was available. The estimated equivalent throughput assists in maintaining the QoS guarantee for video transmission. The simulation results demonstrate that the proposed scheme is capable of offering a near-unimpaired video quality, as a benefit of maintaining a low video FDR.

## 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. Zhang, Y. Zheng, M. A. Khojastepour, and S. Rangarajan, "Cross-layer optimization for streaming scalable video over fading wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 344–353, Apr. 2010.
- [3] J. Ding, D. Deng, T. Wu, and H. Chen, "Quality-aware bandwidth allocation for scalable on-demand streaming in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, Apr. 2010.
- [4] J. W. Huang, H. Mansour, and V. Krishnamurthy, "A dynamical games approach to transmission-rate adaptation in multimedia WLAN," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3635–3646, Jul. 2010.
- [5] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 749–759, May 2007.
- [6] M. Mitsumura, H. Masuyama, S. Kasahara, and Y. Takahashi, "Buffer-overflow and starvation probabilities for video streaming services with application-layer rate-control mechanism," in *Proc. QTNA*, Aug. 2011, pp. 134–138.
- [7] D. Jarnikov and T. Özçelebi, "Client intelligence for adaptive streaming solutions," *Signal Process.: Image Commun.*, vol. 26, no. 7, pp. 378–389, Aug. 2011.
- [8] R. Pantos, HTTP Live Streaming [Online]. Available: <http://tool.ietf.org/id/draft-pantos-http-live-streaming-00.txt>
- [9] Transparent End-to-End Packet Switched Streaming Service (PSS) 3GPP, 3GPP TS 26.244 v9.2.0.
- [10] M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, *Probabilistic Methods for Algorithmic Discrete Mathematics*. Berlin, Germany: Springer-Verlag, 1998.
- [11] L. Hanzo, J. Błogħ, and S. Ni, *3G, HSPA and FDD versus TDD Networking: Smart Antennas and Adaptive Modulation*. Hoboken, NJ: Wiley, 2008.
- [12] Physical Layer Procedures (FDD) 3GPP, 3GPP TS25.214 V6.2.0.
- [13] "R1-02-0675, Revised CQI Proposal," Tech. Rep. Motorola and Nokia, 2002, 3GPP RAN WG1.
- [14] *Video Trace Library*, [Online]. Available: <http://trace.eas.asu.edu/>