Joint Source-Channel Adaptation for Perceptually Optimized Scalable Video Transmission

Amin Abdel Khalek, Constantine Caramanis, and Robert W. Heath Jr.
Wireless Networking and Communications Group
The University of Texas at Austin
1 University Station C0803, Austin, TX 78712
Email: akhalek@utexas.edu, caramanis@mail.utexas.edu, rheath@ece.utexas.edu

***Abstract*—Providing perceptual quality guarantees for video transmission over wireless channels is important for the end-user experience. The paper proposes an algorithm for perceptual quality optimization of scalable H.264 video with temporal and quality scalability. The algorithm supports adaptive unequal error protection so that different video layers are protected according to their relevance to video quality. For each video layer, a target packet error rate (PER) is selected such that a perceptual quality guarantee, measured using the multi-scale structural similarity (MS-SSIM) index, is satisfied. Given the target PER per video layer, the algorithm selects the modulation and coding scheme and the number of temporal and quality layers to transmit adaptively based on the channel state information (CSI) and source rates per video layer. Results show that the algorithm provides immunity against short-term channel fluctuations, balances reliability and perceptual quality, reduces playback buffer starvation probability, and provides a convenient buffer management policy.**

I. Introduction

Delivering high quality video requires a cross-layer design that incorporates perceptual quality. High perceptual quality can be maintained by using link adaptation techniques at the PHY layer to adapt the modulation and coding scheme according to the variation in channel conditions. With scalable video coding, the set of video layers to transmit is another degree of freedom that can be jointly tuned for high quality continuous video playback at the receiver. Under conventional wisdom, link adaptation and source adaptation are performed separately at different layers. We propose to jointly adapt the source and channel to achieve high perceptual quality.

Previous work often relies on conventional rate-distortion optimization models [1], [2] to capture the tradeoff between the source rate and the video distortion, typically quantified by means of the peak signal-to-noise ratio (PSNR). This approach, however, does not capture the relation between resources and video quality in the context of multidimensional scalable coding [3]. Further, it is well known that the PSNR metric does not correlate well with perceived video quality [4]. Learning-based techniques using subjective quality assessment were explored in [3], however, subjective assessment is time-consuming and impractical for online network optimization [5]. There has been limited work thus far on perceptually-optimized online adaptation for scalable video.

In this paper, we propose an algorithm for joint source and channel adaptation of a temporally-scalable and quality-scalable H.264 compressed video bitstream over time-varying wireless channels. We jointly optimize the modulation and coding scheme (MCS) and the number of temporal/quality layers to transmit based on channel conditions and per-layer source rates. Perceptual quality is measured using the multi-scale structural similarity (MS-SSIM) index [6]. A model is developed to provide unequal error protection among layers to satisfy per-layer target packet error constraints. These constraints are designed to satisfy the overall perceptual quality guarantee. The quality guarantee is chosen as a fraction of the baseline quality under lossless transmission. Thus, it determines how much quality degradation can be tolerated due to channel-induced losses. Our proposed algorithm provides a buffer management policy that allows tradeoff between starvation probability and perceptual quality. It is shown that the quality guarantee is maintained irrespective of channel fluctuations as long as all temporal layers can be supported.

II. Video Coding Preliminaries

A. Scalable Video Coding

The H.264/AVC video coding standard [7] provides high coding efficiency. The video coding layer (VCL) encodes motion and texture information at the level of macroblocks. VCL units are encapsulated into network abstraction layer (NAL) units which are in turn passed to the lower protocol layers for transmission. In this work, we use the term NAL unit and packet interchangeably.

To provide a network-friendly design, the scalable video coding (SVC) extension of H.264/AVC [8] allows rate scalability at the bitstream level by generating embedded bitstreams that are partially decodable at different bitrates with degrading quality. The basic level of quality is supported by the base layer and incremental improvements are provided by the enhancement layers. Different NAL units can correspond to different video layers and are distinguished by layer identifiers.

1) Temporal Scalability: Temporal scalability allows scaling the temporal frame rate of the video sequence. The selected frame rate determines the frequency of motion-compensated prediction. A bitstream with one of the supported frame rates is obtained by removing all NAL units with a higher layer identifier [8].

2) Quality Scalability: Quality scalability can be realized by means of coarse-grain scalability (CGS) and/or medium-grain scalability (MGS). CGS refines texture information by quantizing the discrete cosine transform (DCT) coefficients of the enhancement layer with a smaller step size than that of lower layers. Some limitations of CGS are that only a few rate points can be extracted from the bitstream in addition to the fact that achieving quality scalability requires dropping

complete enhancement layers. An alternative is provided by MGS which achieves packet-level scalability so that individual NAL units can be dropped from the enhancement layers. This is made possible because MGS splits the quantization parameters of each macroblock over multiple MGS layers [8].

B. Perceptual Quality Assessment

Perceptual quality optimization relies on the selection of an appropriate tool for video quality assessment. We adopt a variation of the full reference (FR) frame-based MS-SSIM index [4], [6] in conjunction with frame copy error concealment and frame rate upconversion.

Traditional quality assessment metrics such as peak signal-to-noise ratio (PSNR) and the mean squared error (MSE) have poor correlation with user perceived quality [4], [5]. Different sources of distortions (e.g. blur, source compression, impulsive noise, contrast stretching, etc.) can be adjusted to yield the same PSNR/MSE while their perceived quality is drastically different. A new approach was pioneered in [4] based on the simple observation that each video frame is highly structured and spatially correlated. Thus, frame quality degradation can be quantified by the change in spatial image structure. Unlike PSNR and MSE, the structural similarity (SSIM) index proposed in [4] is able to accurately capture the perceptual degradation due to the different types of distortions. The SSIM index and can be expressed as

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (1)$$

where x is the reference image in the pixel domain and y is the corresponding distorted image, μ_x, μ_y, σ_x^2, σ_y^2, σ_{xy} are the mean intensity of x, the mean intensity of y, the variance of x, the variance of y, and the covariance of x and y respectively, and C_1 and C_2 are regularization constants.

Multi-scale SSIM [6] improves correlation with subjective quality by incorporating image detail at multiple resolutions. In the spatial domain, the frame is low-pass filtered and downsampled iteratively and the final score is a weighted average of the scale-specific SSIM scores.

III. System Model

In this section, we describe the proposed system model including the scalable video coding model, the channel model, and the PHY layer model.

A. Scalable Video Coding

We consider a scalable video bitstream composed of L_q quality layers and L_t temporal layers. The group of picture (GoP) size is r frames at a maximum frame rate of f_r frames/sec. A GoP consists of one intra (I) picture or predictively coded (P) picture and $r-1$ hierarchical bi-predictively coded (B) pictures so that $L_t = \log_2(r) + 1$.

The frame structure assumed in this work is shown in Figure 1. Two quality layers are encoded with coarse-grain scalability (CGS). The CGS enhancement layer is further split into three MGS layers. Thus, we have a total of $L_q = 4$ quality layers. We use a GoP size $p = 8$ so that $L_t = 4$. Macroblocks (MBs) of size 4×4 are used and the 16 coefficients per MB are split among the three MGS layers according to the MGS vector [3,3,10]. Frames corresponding to the i^{th} temporal layer are labeled T_i. We apply the concept of key pictures, enabled by MGS, so that reconstructed enhancement layers can be used as a reference picture for motion-compensated prediction [8].

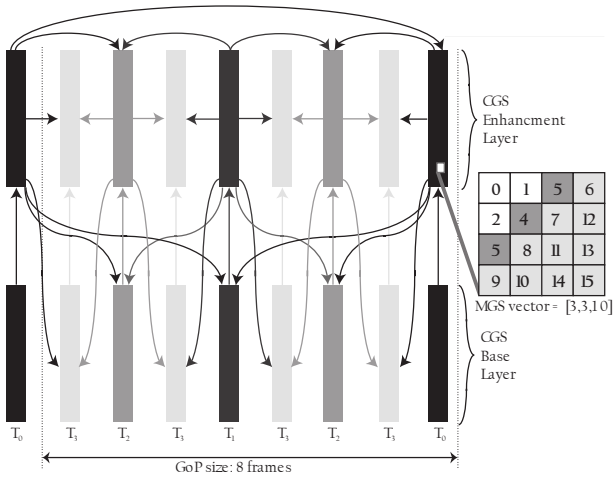

Fig. 1. Frame structure assumed in this paper: Temporal scalability is applied with four temporal layers while quality scalability splits the CGS enhancement into three MGS layers to provide packet-level scalability.

B. Channel and PHY Layer

We consider a flat fading channel with AWGN noise. The channel is assumed time-invariant over the duration of one packet. We use the instantaneous signal to noise ratio (SNR) to characterize the channel state information (CSI). Let $\gamma[p]$ denote the channel SNR in the time interval of transmission of the p^{th} packet. We assume Rayleigh fading so that $\gamma[p]$ is exponentially-distributed. We also assumed that the SNR is reliably fed back to the transmitter without delay.

At the PHY layer, a set of possible modulation and coding schemes (MCS) are available for transmission. The set of possible modulation schemes is denoted $\mathcal{R}$ and the set of possible coding rates is denoted $\mathcal{C}$. Per-layer packet error rates are used to provide unequal error protection (UEP) among layers. Thus, even for the same SNR, packets corresponding to lower layers are more protected than packets corresponding upper layers by using a lower coding rate. Consequently, MCS selection is performed at the packet level.

IV. Problem Formulation

In this section, the problem of adaptive video transmission is formulated with the objective of maximizing perceptual quality subject to packet error rate and buffer starvation constraints.

For the n^{th} GoP, adapting the scalable codec entails selecting the temporal scalability level $0 \le T[n] \le L_t - 1$ and the quality scalability level $0 \le Q[n] \le L_q - 1$. Let $B_{k,l}[n]$ denote the size of all encoded residual frames in the n^{th} GoP corresponding to temporal layer l and quality layer k.

Since a possibly different rate and channel code is selected independently for each packet, we can write the timeslot $t_{k,l}[n]$ allocated to the transmission of video layer (k,l) as

$$t_{k,l}[n] = \sum_{p \in \mathcal{G}^n} I(t(p) = k) I(q(p) = l) \frac{B[p]}{R[p]c[p]} \quad (2)$$

where $B[p]$ is the size of packet p, $\mathcal{G}^n$ is the set of packets belonging to the n^{th} GoP, $c[p] \in \mathcal{C}$ is the selected code rate for packet p, and $R[p] \in \mathcal{R}$ is the per-packet raw transmission rate specified by the selected modulation scheme. Each packet has a temporal layer identifier $t(p)$ and a quality layer identifier $q(p)$ and $I(.)$ is the indicator function.

One of the main objectives of the proposed link adaptation policy is to provide a high perceptual quality solution while avoiding playback buffer starvation which occurs when the video frame is requested by the application layer for playback before being received. This constraint can be translated as

$$\sum_{i=1}^{n}\sum_{k=0}^{T[i]}\sum_{l=0}^{Q[i]} t_{k,l}[i] \leq n\Delta t - t_b \tag{3}$$

where $\Delta t = r/f_r$ is the time to playback one GoP at the receiver and t_b is a "buffering margin" introduced to capture the interesting tradeoff between starvation probability and perceptual quality. Satisfying this constraint ensures continuity of playback at the receiver. Thus, it can be applied to guide the selection of $T[n]$ and $Q[n]$. Given that $n-1$ GoPs were already transmitted, the expression can be written as follows

$$\sum_{k=0}^{T[n]}\sum_{l=0}^{Q[n]} \tilde{t}_{k,l}[n] \leq n\Delta t - t_b - \sum_{i=1}^{n-1}\sum_{k=0}^{T[i]}\sum_{l=0}^{Q[i]} t_{k,l}[i]\cdot \tag{4}$$

Note that before transmitting the n^{th} GoP, $\tilde{t}_{k,l}[n]$ is not known since it depends on the channel response for the duration of the next GoP. We do not assume knowledge of the channel response for the entire GoP duration since it spans multiple coherence times. In §V-B1, we propose a method to estimate $\tilde{t}_{k,l}[n]$ and use it to make a decision on $T[n]$ and $Q[n]$. We denote the last term of (4) by $t_{\text{elapsed}}[n]$.

To guide the selection of different coding rates for different layers, we set a target packet error rate constraint for each video layer $\text{PER}_{k,l} \leq \rho_{k,l}$ where $\rho_{k,l}$ is the maximum tolerable PER that can be supported for layer (k,l). For a given video sequence, we select $\rho_{k,l}$ offline such that $\mathbb{E}[Q|\text{PER}_{k,l} = \rho_{k,l}] = q_l^{\text{target}}$ (See §V-A1).

We now formulate the problem of joint source-channel adaptation with the objective of maximizing perceptual quality

$$\max \quad Q\left(T[n], Q[n], \{c[p]\}, \{R[p]\}\right) \tag{5}$$

$$\text{s.t.} \quad \sum_{k=0}^{T[n]}\sum_{l=0}^{Q[n]} \tilde{t}_{k,l}[n] \leq n\Delta t - t_b - t_{\text{elapsed}}[n] \tag{6}$$

$$\text{PER}_{k,l} \leq \rho_{k,l}; 0 \leq k \leq T[n];\ 0 \leq l \leq Q[n] \tag{7}$$

$$c[p] \in \mathcal{C};\ R[p] \in \mathcal{R};\ p \in \mathcal{G}^n \tag{8}$$

where (5) is the quality measure corresponding to sending $T[n]$ temporal layers and $Q[n]$ quality layers with modulation and coding schemes specified per packet according to $c[p]$ and $R[p]$. The buffer starvation constraint is specified in (6) and the per-layer PER constraint is specified by (7).

V. Joint Source-Channel Adaptation Algorithm

In this section, we describe our proposed algorithm to solve the source-adaptive channel-adaptive perceptual video optimization problem. Since the channel is assumed invariant only over the duration of a single packet, the timescale of channel adaptation is smaller than that of source adaptation. Furthermore, because perceptual quality metrics operate in the pixel domain, it is hard to quantify the effect of channel distortions analytically. Thus, an offline process is developed to estimate the allowed per-layer PER for a target perceptual quality. Next, given these PER thresholds, online adaptation jointly adapts the MCS and the scalable codec.

A. Model Training

1) Modeling the effect of per-layer packet error rate on perceptual quality: Essentially, unequal error protection should be applied for different video layers in proportion to the degradation caused by packet losses corresponding to that layer. Thus, we need to quantify the loss in perceptual quality for a given packet error rate at some video layer. Although the empirical relations between PER and perceptual quality are content specific, they are similar for similar content (e.g. fast motion) and they can be learned offline. Additionally, they can be parameterized by per-layer source rates.

First, we develop a model that finds the dependencies among packets in the bitstream so that the error propagation effect can be quantified. Based on the H.264/SVC standard and the hierarchical encoding structure, we can deduce the packet dependency model. Assume packet p is lost. If $q(p) = L_q - 1$, packet p has no dependant packets. If $0 \leq q(p) < L_q - 1$, we drop the $L_q - 1 - q(p)$ packets corresponding to quality refinements of the same slice. If $q(p) = 0$ and $t(p) > 0$, we drop the entire frame. Finally, if $q(p) = 0$ and $t(p) = 0$, we drop the entire GoP. It can be seen that errors at $q(p) > 0$ are much less severe than those with $q(p) = 0$ because of MGS packet-level scalability. Thus, we expect to be able to support a higher level of PER for upper quality layers while maintaining good perceptual quality.

Given a set of video sequences to be used for training, we construct a Monte Carlo simulation where we drop a fixed percentage of packets from any given layer uniformly at random. Dropping these packets entails dropping their dependants. At each run, the bitstream is decoded and the quality index is measured. We apply the uniform packet drop for each PER value and for each layer and find the average quality if a given PER is allowed for some layer. We repeat over all layers and for a range of PERs. Thus, we find an empirical relation between the PER per layer and perceptual quality assuming all video layers are transmitted. For a given target perceptual quality $\mathbb{E}[Q|\text{PER}_{i,j} = \rho_{i,j}] = q_j^{\text{target}}$, we find $\rho_{i,j}$ that satisfies the constraint for each video layer.

2) PER Waterfall Curves: Since the packet size is fixed and the channel is considered time-invariant over the packet duration, the instantaneous SNR characterizes the channel. Using a Monte Carlo simulation, we can find waterfall curves for any combination of SNR and MCS. Thus, the choice of MCS and the instantaneous SNR uniquely identify the PER.

B. Source-Channel Adaptation Algorithm

For each GoP, the algorithm selects which video layers to transmit. This entails dropping the corresponding upper layer packets. For lower layer packets selected, a decision is made on the best MCS. GoP-level scheduling does not assume prior channel knowledge whereas packet-level MCS selection does.

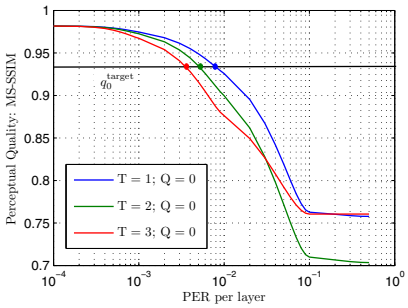

Fig. 2. Video quality vs. per-layer PER for CGS base layers.

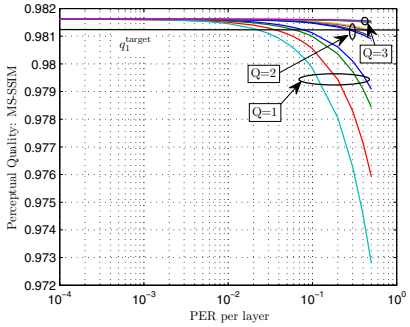

Fig. 3. Video quality vs. per-layer PER for MGS enhancement layers.

1) GoP-level video layer scheduling: For each GoP, the objective is to make a decision on which video layers to transmit. The scheduler always attempt to send the base layer (T=0,Q=0). For video layer (k,l), we cannot assume channel knowledge for the entire duration of transmission since it spans multiple coherence times of the channel. We know, however, the size of each packet $B[p]$ and its layer identifier $t(p)$ and $q(p)$. For any instantaneous SNR, we know its probability and the best corresponding MCS (See §V-B2). Based on the Rayleigh fading channel, a reasonable assumption is that the average SNR $\bar{\gamma}$ for the next video layer is the same as the previous video layer. Thus, the expected transmission time for video layer (k,l) is

$$\tilde{t}_{k,l}[n] = \sum_{\substack{p\in\mathcal{G}^n \text{s.t.} \\ t(p)=k,q(p)=l}} \int_0^\infty \frac{B[p]}{R(\gamma,k,l)c(\gamma,k,l)} \cdot \frac{e^{-\gamma/\bar{\gamma}}}{\bar{\gamma}} d\gamma \quad (9)$$

$$= \left(\sum_{\substack{p\in\mathcal{G}^n \text{s.t.} \\ t(p)=k,q(p)=l}} B[p] \right) \int_0^\infty \frac{e^{-\gamma/\bar{\gamma}}}{R(\gamma,k,l)c(\gamma,k,l)\bar{\gamma}} d\gamma \quad (10)$$

where $R(\gamma,k,l)$ and $c(\gamma,k,l)$ are uniquely specified by the SNR and the layer identifiers. The scheduler estimates $\tilde{t}_{k,l}[n]$ by numerical integration and picks the largest number of temporal layers and the largest number of quality layers such that the sum of average transmission times for all selected layers does not violate the buffer starvation constraint (6). Note that this does not fully avoid starvation. Since the decision is based on the Rayleigh fading assumption, the actual transmission time per layer will slightly vary. This provides the motivation behind the buffering margin defined in (3).

Next, the selected layers are transmitted. If the transmission time was underestimated and buffer starvation occurs, future enhancement layers are dropped, and transmission of the next GoP is started immediately.

This method reveals the buffering aspect of the problem. The algorithm is able to tradeoff quality and starvation probability by tuning the buffering margin parameter. Furthermore, it can be intelligently adapted over time based on past observations of buffer starvation and perceptual quality fluctuations.

2) Packet-level MCS selection: Assume a video layer is selected for transmission. For any given packet, a decision has to be made as to which MCS is used for transmission at the PHY layer. The layer identifier $t(p)$ and $q(p)$ is known and we assume perfect knowledge of $\gamma[p]$. Given the instantaneous SNR, for every possible modulation scheme, we find the maximum coding rate that can support the per-layer PER constraint based on the MCS PER empirical expressions. If the problem is feasible, i.e. at least one MCS can support the PER constraint, we pick the modulation scheme and its corresponding maximum coding rate such that net bit rate is maximized. If the problem is not feasible, we pick the MCS that maximizes reliability among all other MCSs.

VI. Results and Analysis

In this section, we present results on the mapping between per-layer PER and video quality as well as a case study for the joint source-channel adaptation algorithm. We consider a flat Rayleigh fading channel with a bell-shaped doppler spectrum suitable for an indoor environment. The speed is 3 m/s and the bandwidth is 500 KHz. The set of possible rates is $\mathcal{R} = [1, 2, 4, 8]$ bps corresponding to BPSK, 4-QAM, 16-QAM, and 64-QAM. The set of possible coding rates is $\mathcal{C} = [1/2, 2/3, 3/4, 5/6]$. Thus, the range of data rates is 0.5 Mbps to 2.5 Mbps. We use the LIVE video database [9], [10] to train the model. Based on q_1^{target} and q_0^{target}, we find $\rho_{k,l}$ for each video layer (k,l). The "riverbed" sequence is used for testing. The resolution is 384x224, the maximum frame rate is 25 fps, and the video sequence length is 10 sec. The corresponding per-layer source rates are shown in Table I.

TABLE I
Source rates (Kbps) for the riverbed sequence

$T\backslash Q$	0	1	2	3
0	275.0	377.9	464.6	602.8
1	469.5	666.2	828.0	1057.8
2	767.6	1129.5	1416.4	1771.0
3	1135.0	1802.0	2296.0	2825.0

Figure 2 shows the video quality achieved when a uniform packet drop is applied to each temporal enhancement layer in the base CGS layer of the "riverbed" video sequence. Losses in packets with layer identifier $Q = 0$ incur large video quality costs. We note that upper temporal layers may be perceptually more relevant since the corresponding number of frames is larger causing more data loss for the same PER.

Figure 3 shows the average video quality achieved when a uniform packet drop is applied to each MGS enhancement layer. Due to MGS packet-level scalability, the degradation in video quality has a significantly lower severity. Adjusting the quality threshold enables trading-off video quality and reliability. An interesting conclusion is that, for the same video

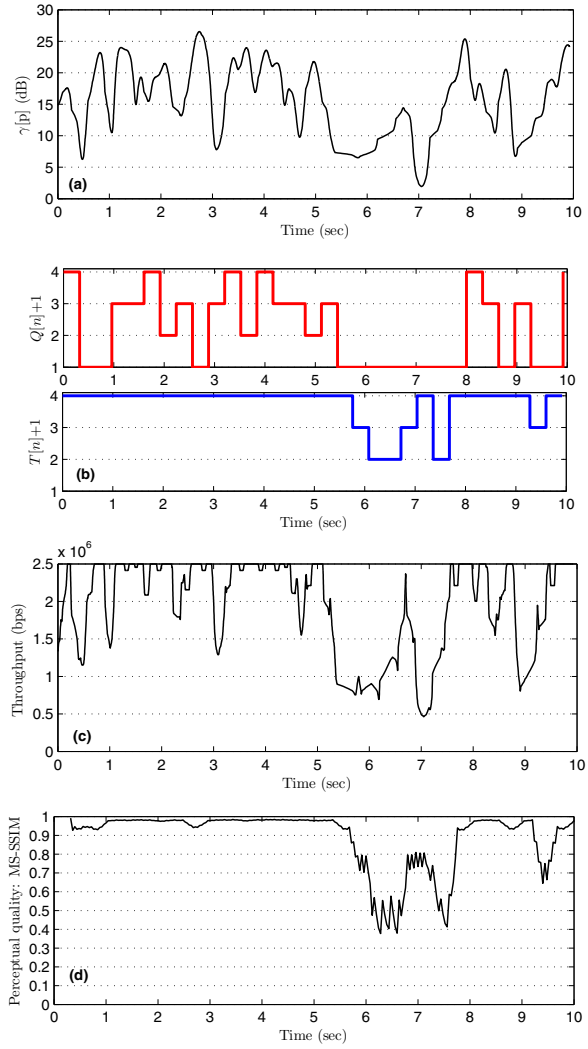

Fig. 4. Case study of the link adaptation algorithm: (a) Instantaneous SNR per packet $\gamma[p]$; (b) Number of temporal and quality layers selected; (c) Instantaneous throughput (averaged by a 50-packet-wide moving filter); (d) Perceptual video quality over time. Note the startup delay of 0.32 sec (1 GoP).

quality requirement, per-layer PER targets among different quality layers are an order of magnitude apart which clearly advocates unequal error protection.

A case study on the joint source-channel adaptation algorithm is provided in Figure 4. The instantaneous SNR process is shown in Figure 4(a). We consider a channel realization where the SNR stays in a deep fade for 2 seconds. The number of quality and temporal layers selected for transmission is shown in Figure 4(b). As per our design, temporal scale changes at a coarser pace than quality scale. Note that the number of layers is selected based on the average SNR in the previous interval. Thus, drastic variations may impact performance by causing the algorithm to overestimate or underestimate the time needed to transmit a video layer. Underestimation could cause buffer starvation. In this scenario, we set a buffering margin $t_b = 0.15$ sec. This margin is sufficient to fully avoid buffer starvation even in the deep fade from 5.5 to 7.5 sec because the difference between transmission time and playback time was large enough taking advantage of previous channel peaks.

Figure 4(c) and 4(d) show the instantaneous throughput and the perceptual quality respectively. Because the algorithm takes advantage of buffering on good channel conditions, at 6 seconds, we are successfully able to support layer (3,1) which requires 1.8 Mbps while the instantaneous throughput is consistently below 1 Mbps. We allow a startup time equal to 1 GoP (0.32 sec). Note that good perceptual quality is consistently maintained up to 5.5 seconds while the SNR and throughput fluctuate within reasonable ranges. Whenever all temporal layers are transmitted, the target quality metrics q_1^{target} and q_0^{target} are satisfied and the algorithm is immune to short-term fluctuations. When the channel is bad, reliability is favored over quality to ensure continuous video playback.

VII. Conclusion

We demonstrated an algorithm to adapt the use of wireless resources and the scalable codec for H.264 video bitstreams based on channel conditions and per-layer source rates. The algorithm takes advantage of buffering and unequal error protection to balance reliability and perceptual quality, and reduce buffer starvation probability. Immediate extensions of the work involve applying the framework to a MIMO-OFDM system and incorporating more robust learning-based techniques.

ACKNOWLEDGMENT

This work has been supported by the Intel-Cisco Video Aware Wireless Networks (VAWN) Program.

References

[1] Y. Fallah, H. Mansour, S. Khan, P. Nasiopoulos, and H. Alnuweiri, "A link adaptation scheme for efficient transmission of H. 264 scalable video over multirate WLANs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 875–887, July 2008.

[2] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for H. 264 video coding over error-prone networks," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 445–454, Apr. 2007.

[3] Y. Wang, M. van der Schaar, S. Chang, and A. Loui, "Classification-based multidimensional adaptation prediction for scalable video coding using subjective quality evaluation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 10, pp. 1270–1279, Oct. 2005.

[4] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Transactions on image processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[5] Z. Wang, L. Lu, and A. Bovik, "Video quality assessment based on structural distortion measurement," *Signal processing: Image communication*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[6] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *In Proceeding of Asilomar*, pp. 1398–1402, Nov. 2004.

[7] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H. 264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[9] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "A Subjective Study to Evaluate Video Quality Assessment Algorithms," *SPIE Proceedings Human Vision and Electronic Imaging*, Jan. 2010.

[10] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.