DISTORTION-OPTIMIZED VIDEO ENCODING AND STREAMING IN MULTI-RATE WIRELESS LANS

Antonios Argyriou

Philips Research Laboratories
Eindhoven, 5656 AE, The Netherlands

ABSTRACT

In this paper, we present a cross-layer methodology for video streaming in wireless LANs that employs joint source coding and rate adaptation at the wireless physical layer (PHY). More specifically, we investigate the impact of adapting the PHY transmission rate, and thus changing the throughput and packet loss channel characteristics, on the rate-distortion (RD) performance of a transmitted video sequence. We develop an algorithm that jointly optimizes source coding, application-layer channel coding, and the PHY rate adaptation. Simulation results obtained with the H.264/AVC codec, demonstrate video quality improvements when compared with a system that employs PHY rate adaptation independently from the application layer.

Index Terms— Wireless LAN, video coding, multimedia communication

1. INTRODUCTION

The need for high quality video services in existing and next generation wireless networks, dictates the design of video streaming algorithms that are flexible and highly adaptive to the rapid fluctuations of the wireless quality of service (QoS). It has been demonstrated that for the best performance, the functionality of protocols that belong to several layers should be jointly designed [1]. One of the most thoroughly studied mechanisms is joint source-channel coding (JSCC). The objective of JSCC is to allocate the available channel rate between source video coding and channel coding for protection, so that video quality is maximized. JSCC in the context of video transmission, has been studied in several works [2, 3]. For example in [3], the authors employ application-layer JSCC by using a real-time video encoder and a hybrid automatic repeat request (ARQ) and forward error correction (FEC) channel coding scheme. In that work, ARQ and FEC are used interchangeably depending on the channel conditions. For wireless networks, JSCC has been thoroughly studied mostly for scalable pre-encoded video since source rate adaptation can be achieved by adding or dropping layers of the bitstream [4]. Even though JSCC is a well-studied topic, the majority of physical layer (PHY) rate adaptation schemes for wireless networks have been devised for data transmission. For example, the work presented in [5], describes a mechanism for driving PHY rate adaptation in a 802.11a wireless LAN (WLAN) by using an analytical closed-form model of the effective throughput. In a more related work [1], the authors develop an analytical performance model that can be used for PHY adaptation when scalable encoded video is employed at the sender.

In this paper, we extended the JSCC principle by considering the packet loss rate and channel rate as another set of optimization parameters that can be controlled. To achieve that, we utilize the multi-rate adaptation mechanism available at the 802.11a WLAN PHY. The intuition behind our approach is that JSCC and PHY rate adaptation have conflicting objectives, that aim to maximize the rate-distortion (RD) performance of a transmitted video sequence and the effective throughput respectively. Therefore, employing them independently cannot lead to optimal performance.

2. SYSTEM OVERVIEW

Fig. 1 depicts the main components of the proposed video streaming system engineered to work with an adaptive WLAN PHY. The system consists of the media source, which is a real-time H.264/AVC baseline profile encoder. The task of estimating the decoder distortion is essentially performed at the encoder. Based on this estimate, our JSCC algorithm calculates the optimal source rate, application-layer FEC code, and PHY channel rate for each frame. Subsequently, it encodes each macroblock with the selected optimal encoding mode, and generates the source video packets. The next step is to sent the source packets to the application layer FEC Reed Solomon (RS) encoder. Next, a UDP/IP header can be added for Internet transmission.

This packet is sent next to the 802.11 medium access control (MAC) layer which appends a header (of 28 bytes), and creates a MAC protocol data unit (MPDU) for wireless transmission. Link layer retransmissions are not used in this work, since we wanted to simplify the derived analytical models. Nevertheless, the modular modeling approach could be easily extended to take into account link layer retransmissions based on existing works [6]. Subsequently, the 802.11a PHY

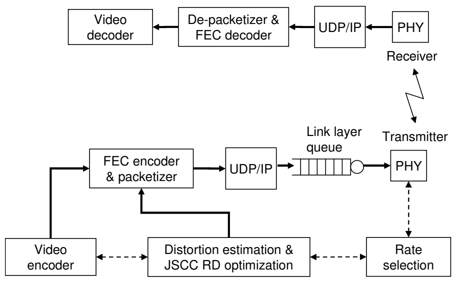

Fig. 1. The proposed cross-layer system for real-time video encoding and transmission in WLANs with adaptive PHY.

layer frame is created by encoding the MPDU with one of the eight available 802.11 transmission rates[1] ranging from 6 to 54 Mb/s [7].

Once the packets are transmitted at the physical layer, they pass through the application-layer RS decoder at the receiver. If RS decoding fails, the macroblocks that correspond to the partially received frames are not discarded but are placed in the decoder buffer (i.e. soft channel decoding). The decoder starts to decode video packets and display them, after an initial *startup delay* that is configured by the application. In case of packet loss, the decoder uses a simple temporal error concealment technique.

3. MODEL OF THE 802.11A PHY WITH MULTIPATH FADING

The channel model we selected for evaluating our system is widely used for indoor wireless environments [8]. With this model, the channel impulse response is captured through a tapped delay line, in which the distribution of the path amplitude is characterized by a Rayleigh fading path delay profile. Also, the average power of the different taps declines exponentially as delay is increased. Using the above channel model, we simulated several realizations of the wireless channel with the help of an existing simulator [9]. Subsequently, we obtained the performance curves of BER vs. SNR for the different PHY transmission rates of 802.11a. Therefore, for channel SNR γ and transmission rate r, the PER for packets of length L can be easily calculated as:

$$p_w(L,r) = 1 - (1 - BER(r,\gamma))^{8L} \quad (1)$$

For calculating the effective throughput at the sender, we assume that no other station is transmitting and therefore there is no packet loss due to MAC layer collisions. We also assume that ACK transmission on the reverse path is considered error free, which is something that can be easily achieved by applying strong error correcting codes. Therefore, if the video payload consists of S_d bytes, and the combined protocol overheads is S_{hdr} bytes, then the *effective throughput* for PHY rate r is given by:

$$T(r,\gamma) = \frac{S_d * R_r}{S_d + S_{hdr} + S_{dcf}(r)} * (1 - p_w(L,r)), \quad (2)$$

where S_{dcf} is the overhead for accessing the channel with the 802.11 DCF mechanism even when no other stations contend for the wireless medium. This value is calculated by considering the 802.11 PHY overhead between the transmission of two successive PHY frames [7]. The above equation is not directly used by any algorithm in the proposed system. However, the system that we used for comparison during our simulations, selects the optimal PHY data rate based on this formula. In our system, we are more concerned with the raw application layer data rate that can be achieved. To obtain this quantity, we just have to ignore packet losses in (2), which makes the maximum possible application data rate:

$$T_{max}(r,\gamma) = \frac{S_d * R_r}{S_d + S_{hdr} + S_{dcf}(r)} \quad (3)$$

4. DISTORTION ESTIMATION AT THE ENCODER

In this section, we will derive the the closed-form recursive model for the decoder distortion. Let $f_i(n)$ denote the value of pixel i in frame n, $\hat{f}_i(n)$ the reconstructed pixel at the encoder, and $\tilde{f}_i(n)$ is the encoder's estimate of the reconstructed pixel at the decoder. In this paper, the mean squared error (MSE) and the peak signal-to-noise ratio are used for calculating distortion. If the distortion of a single pixel $f_i(n)$, is denoted as $d(i,n)$, according to the previous definition it can be written as:

$$d(i,n) = [f_i(n) - \tilde{f}_i(n)]^2 \quad (4)$$

But since we follow a frame-level recursive approach, the overall expected distortion can be also written as the sum of the three contributing components, namely source, error propagation and channel distortion:

$$d(i,n) = (1-\rho)d_s(i,n) + (1-\rho)d_{ep}(i,n) + \rho d_{ec}(i,n) \quad (5)$$

In the above equation ρ is the residual PER after RS decoding is applied. When the distortion of particular macroblocks is calculated according to (5), it is stored in a structure called distortion map. The reason is that when calculating recursively the distortion for subsequent frames as given in (5), the appropriate distortion that will be propagated (d_{ep}) from specific sub-blocks in previous frames is identified and added to the overall expected distortion [10]. Since H.264/AVC supports encoding modes even for 4x4 pixel areas, a 4x4 sub-block is the minimum-sized element required for storing the calculated distortion components of previous frames. The second component in (5), is the error concealment distortion.

[1] Even though the appropriate term is transmission mode, we do not use it to avoid confusion with the encoding mode of video macroblocks.

We consider a simple error concealment approach at the decoder according to which the concealment motion vector for a lost macroblock is copied from the macroblock in the same spatial location of the previous frame [10]. Finally, the source distortion that is attributed to the use of different H.264/AVC encoding modes, can be easily calculated at the encoder.

5. RATE-DISTORTION FRAMEWORK FOR JSCC AND PHY RATE SELECTION

In this section we present the JSCC rate selection (RS) framework for jointly optimizing the selection of the source encoding mode of individual macroblocks, channel coding with application-layer FEC, and PHY transmission rate.

The optimization problem can be formally expressed as follows. Let us first define as **S** the set of all the available the source coding modes for H.264/AVC that also include the associated reference frame. Let also **C** be the set of the available channel coding (FEC) parameters. The vector of the candidate source and channel coding parameters for the specific frame n is denoted as $\mu(n)$ and $c(n)$ respectively. Let finally **M** be the set of 802.11a PHY transmission rates. The objective is therefore to minimize the overall distortion

$$\min_{\mu \in \mathbf{S}, c \in \mathbf{C}, r \in \mathbf{M}} E[d(\mu, c, r)], \tag{6}$$

such that the following rate constraint is satisfied:

$$R(k, n) \leq T_{max}(r_k) \tag{7}$$

The above constraint means that the transport packet k should not exceed the available maximum raw application data rate under transmission rate r. Our approach is to solve the above problem by considering a set of PHY transmission rates that correspond to throughput $T(r)$, instead using a fixed channel rate constraint like existing RD optimization techniques.

By using Lagrangian relaxation we convert the previous constrained optimization problem in a unconstrained one. The total Lagrangian cost for frame n can be expressed as

$$J_n = d(n) + \lambda_n \sum_{k=1}^{K} (R_k(n) - T_{max}(r_k)), \tag{8}$$

where λ is the Lagrange multiplier. Since the contribution of each macroblock in the overall cost has been shown to be additive, the Lagrangian cost is minimized for each macroblock individually. Bisection search is used for calculating the Lagrange multiplier for each frame. With this simple approach we observed very fast convergence to the global minimum.

For each possible transmission rate r, the sender calculates the estimated new residual PER $\rho(p_w)$ and throughput. Subsequently, the encoder calculates for each possible choice of H.264/AVC encoding mode μ the number of bits needed for the current macroblock and also the associated distortion. Now after the cost is estimated for a single macroblock, the

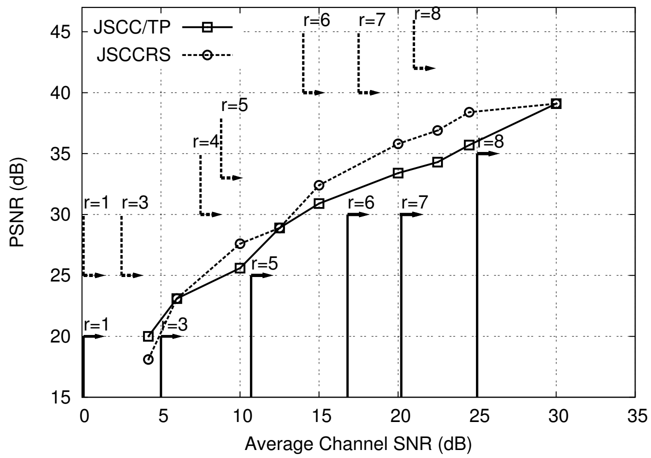

Fig. 2. Average PSNR of the Foreman sequence for JSCCRS and JSCC/TP with a multipath fading wireless channel model. Target frame rate is 30 fps.

algorithm proceeds to the same calculations for the complete frame. At the end of one execution run of this algorithm, the optimal values for μ, c and r for each macroblock of frame n are derived.

6. PERFORMANCE ANALYSIS

In this section, we present a comprehensive set of experiments that stress-test the performance of the proposed cross-layer video encoding and transmission system for IEEE 802.11a WLANs. For our experiments, we used the 12.2 JVT reference software implementation of H.264/AVC [11]. The luminance component of the QCIF sequence Foreman was used for real-time encoding. We used only I and P frames and we set the GOP size to 128 frames. The QP was set equal to 14 in order to get constant video quality, at the cost of VBR. However, the average bitrate throughout the sequence was 512Kbps. Furthermore, the maximum source video packet size was set to 1200 bytes. With this frame size, the encoded P frames are packetized into 3 source packets. The channel was simulated with an existing simulator [9], while the results were averaged for 50 realizations of the wireless channel.

The objective of our experiments is to demonstrate the advantage of exercising JSCC for video transmission jointly with the PHY rate selection (JSCCRS). For comparison we implemented the JSCC/TP scheme, where JSCC is applied independently of the PHY rate adaptation algorithm. With JSCC/TP the algorithm attempts to maximize throughput independently of JSCC and so the algorithm performs a rate downshift and upshift by using our analytical throughput formula in section 3.

The vertical lines in Fig. 2 indicate the wireless SNR for which the optimal transmission rate was changed for the two systems under test. The vertical lines in the lower part of the figure correspond to JSCC/TP, while the lines in the upper part to JSCCRS. The first striking observation from this figure is that once the optimal PHY transmission rate is selected for either system, it remains constant until the next change. This is something to be expected for the JSCC/TP algorithm since the rate that maximizes throughput is always selected Even though with JSCCRS the sender can select the transmission rate of each individual outgoing PHY frame, the similar behavior should be expected. The reason is that when a transmission rate has been selected as the currently optimal, an increasing for example channel SNR, leads to improvements both in throughput and PER that maintain the transmission rate as the optimal.

Another interesting observation is that the proposed system is switching to a higher rate at a lower channel SNR than the JSCC/TP algorithm. The reason is that with the extra available bandwidth from the selection of a higher PHY rate, a stronger FEC code is used and compensates for increased packet losses. The error concealment algorithm at the decoder is also in part responsible for this performance increase. Of course, for data traffic the performance would not increase since in that case the optimization objective is the actual effective throughput. We also observed that a drop the residual PER below 10% can be handled more effectively from the error concealment algorithm. For example when channel SNR is nearly 10 dB, JSCC/TP uses RS(11,9). However, JSCCRS has already switched to rate $r3$ that provides double data rate but lower effective throughput due to higher packet loss. For the same SNR, JSCCRS uses RS(16,9) but it has also increased the source encoding rate. With respect to the actual video quality results in terms of PSNR, we observe that JSCCRS can achieve a higher average PSNR compared to the JSCC/TP system. If we look more carefully into the same figure, we observe that JSCCRS outperforms JSCC/TP by a constant margin ranging from 1 to 4.5 dB depending on the channel SNR γ. The performance gain is slightly higher when switching from $r1 \rightarrow r3$, $r3 \rightarrow r5$, and $r5 \rightarrow r7$ since the raw data rate is doubled when compared with the case of $r7 \rightarrow r8$.

7. CONCLUSIONS

In this paper we presented a novel cross-layer approach for distortion-optimized video streaming in wireless LANs that support a multi-rate PHY. We considered the use of the following mechanisms by our system: error resilient video source encoding, application layer channel coding with FEC, error concealment, and PHY rate adaptation. We developed an algorithm that selects the optimal encoding mode of individual macroblocks (source coding), application-layer FEC (channel coding), and the 802.11a PHY data transmission rate. Our comparative performance analysis showed that video performance in terms of PSNR is considerably enhanced with JSCCRS, and more importantly for a realistic channel model. Furthermore, our analysis led us to the conclusion that significant performance gains can be achieved, if the PHY transmission rate is selected according to the RD characteristics of a transmitted video sequence.

8. REFERENCES

[1] M. van der Schaar and S. Shankar, "Cross-layer wireless multimedia transmission: Challenges, principles, and new paradigms," *IEEE Wireless Communications*, pp. 50–58, August 2005.

[2] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 340–356, March 2000.

[3] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 40–53, January 2006.

[4] Q. Zhang, W. Zhu, and Y. Zhang, "Channel-adaptive resource allocation for scalable video transmission over 3G wireless network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 8, pp. 1049–1063, August 2004.

[5] S. Choudhury and J. D. Gibson, "Payload length and rate adaptation for multimedia communications in wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 796–807, May 2007.

[6] A. Argyriou, "Cross-layer adaptive ARQ for uplink video streaming in tandem wireless/wireline networks," in *IEEE Multimedia Signal Processing Workshop (MMSP)*, October 2007.

[7] "Part 11: Wireless LAN Medium Access Control and Physical Layer (PHY) Specifications," IEEE 802.11-1999 Standard, August 1999.

[8] "Tentative criteria for comparison of modulation methods," IEEE P802.11-97/96, September 1997.

[9] J. Heiskala and J. Terry, *OFDM Wireless LANs: A Theoretical and Practical Guide*. Sams Publishing, 2001.

[10] 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, April 2007.

[11] "JVT reference software," Available from http://bs.hhi.de/ suehring/tml/download/.