Interactive Video Coding and Transmission over Heterogeneous Wired-to-Wireless IP Networks Using an Edge Proxy

Yong Pei

Computer Science and Engineering Department, Wright State University, Dayton, OH 45435, USA
Email: ypei@cs.wright.edu

James W. Modestino

Electrical and Computer Engineering Department, University of Miami, Coral Gables, FL 33124, USA
Email: jmodestin@miami.edu

Digital video delivered over wired-to-wireless networks is expected to suffer quality degradation from both packet loss and bit errors in the payload. In this paper, the quality degradation due to packet loss and bit errors in the payload are quantitatively evaluated and their effects are assessed. We propose the use of a concatenated forward error correction (FEC) coding scheme employing Reed-Solomon (RS) codes and rate-compatible punctured convolutional (RCPC) codes to protect the video data from packet loss and bit errors, respectively. Furthermore, the performance of a joint source-channel coding (JSCC) approach employing this concatenated FEC coding scheme for video transmission is studied. Finally, we describe an improved end-to-end architecture using an edge proxy in a mobile support station to implement differential error protection for the corresponding channel impairments expected on the two networks. Results indicate that with an appropriate JSCC approach and the use of an edge proxy, FEC-based error-control techniques together with passive error-recovery techniques can significantly improve the effective video throughput and lead to acceptable video delivery quality over time-varying heterogeneous wired-to-wireless IP networks.

Keywords and phrases: video transmission, RTP/UDP/IP, RS codes, RCPC codes, JSCC, edge proxy.

1. INTRODUCTION

With the emergence of broadband wireless networks and the increasing demand for multimedia transport over the Internet, wireless multimedia services are expected to be widely deployed in the near future. Many multimedia applications will require video transmission over links with a wireless first and/or last hop as illustrated in Figure 1. However, many existing wired and/or wireless networks cannot provide guaranteed quality of service (QoS), either because of congestion, or because temporally high bit-error rates cannot be avoided during fading periods. Channel-induced losses, including packet losses due to congestion over wired networks as well as packet losses and/or bit errors due to transmission errors on a wireless network, require customized error resilience and channel coding strategies that add redundancy to the coded video stream at the expense of reduced source coding efficiency or effective source coding rates, resulting in compromised video quality.

In this paper we quantitatively investigate the effects of packet losses on reconstructed video quality caused by bit errors anywhere in the packet in a wireless network if only error-free packets are accepted, as well as the effects of residual bit errors in the payload if errored packets are accepted instead of being discarded in the transport layer. The former corresponds to the use of the user datagram protocol (UDP) employing a checksum mechanism while the latter corresponds to the use of a transparent transport protocol, such as UDP-Lite [1], together with forward error correction (FEC) to attempt to correct transmission errors.

This work represents an extension of previous works [2, 3]. In particular, in [2] we described an approach using edge proxies which did not address the unique FEC requirements on the wired networks. This was followed by work reported in [3] where a concatenated channel coding approach was employed, but without an edge proxy, which attempted to address the distinct FEC requirements of both the wired and wireless networks.

A joint source-channel coding (JSCC) approach has been well recognized as an effective and efficient strategy to provide error-resilient image [4, 5, 6, 7, 8] and video [3, 9, 10, 11] transport over time-varying networks, such as wireless IP
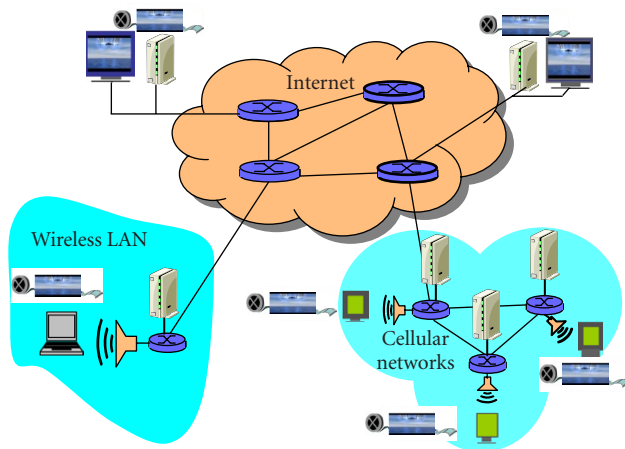
FIGURE 1: Illustration of heterogeneous wired-to-wireless networks.

networks. In this paper, we extend the work in [3] and provide a quantitative evaluation of a proposed JSCC approach used with a concatenated FEC coding scheme employing Reed-Solomon (RS) block codes and RCPC codes to actively protect the video data from the different channel-induced impairments associated with transmission over tandem wired and wireless networks. However, we demonstrate that this approach is not optimal since the coding overhead required on the wired link must also be carried on the wireless link which can have a serious negative effect on the ability of the bandwidth-limited wireless link to support high-quality video transport.

Finally, we will present a framework for an end-to-end solution for packet video over heterogeneous wired-to-wireless networks using an edge proxy. Specifically, the edge proxy serves as an agent to enable and implement selective packet relay, error-correction transcoding, JSCC, and inter-operation between different transport protocols for the wired and wireless networks. Through the use of the edge proxy located at the boundary of the wired and wireless networks, we demonstrate the ability to avoid the serious compromise in efficiency on the wireless link associated with the concatenated approach. More specifically, we employ RS codes only on the wired network to protect against packet losses while the RCPC codes are employed only on the wireless network to protect against bit errors. The edge proxy provides the appropriate FEC transcoding resulting in improved bandwidth efficiencies on the wireless network. We believe that the value of the proposed approach, employing an edge proxy with appropriate functionalities, lies in the fact that little or no change needs to be provided on the existing wired network while at the same time it addresses the distinctly different transport requirements for the wireless network. Furthermore, it uses fairly standard FEC approaches in order to support reliable multimedia services over the Internet with a wireless first and/or last hop.

The remainder of this paper is organized as follows. In Section 2, we provide some technical preliminaries describing an application level framing (ALF) approach employing RTP-H.263+ packetization. In Section 3, we briefly describe the background for packet video over wireless networks and provide a quantitative study of packet video performance over wireless networks based on the two different transport-layer strategies as discussed above. We also describe the RCPC codes, the channel-loss model, and the assumed physical channel model for the wireless networks under study. In Section 4, we introduce a concatenated FEC coding scheme for packet video transport over heterogeneous wired-to-wireless networks, and briefly describe the interlaced RS codes and packetization scheme employed. In Section 5, we present a framework for an end-to-end solution for packet video over heterogeneous wired-to-wireless network using edge proxies and provide a comparison of the performance achievable compared to the concatenated approach. Finally, Section 6 provides a summary and conclusions.

2. PRELIMINARIES

2.1. Application-layer framing

To provide effective multimedia services over networks lacking guaranteed QoS, such as IP-based wired as well as wireless networks, it is necessary to build network-aware applications which incorporate the varying network conditions into the application layer instead of using the conventional layered architecture to design network-based applications. A possible solution is through ALF as proposed in [12]. The principal concept of ALF is that most of the functionalities necessary for network communications will be implemented as part of the application. As a result, the underlying network infrastructure provides only minimal needed functionalities. The application is then responsible for assembling data packets, FEC coding and error recovery, as well as flow control. The protocol of choice for IP-based packet video applications is the real-time transport protocol (RTP) [13], which is an implementation of ALF by the internet engineering task force (IETF). Likewise, UDP-Lite [1] is a specific instance of ALF in the sense that the degree of transparency at the transport layer can be tailored to the application by allowing the checksum coverage to be variable, including only the header or portions of the packet payload as well. In this paper, we will consider the use of ALF-based RTP-H.263+ for video transmission over wired and wireless IP networks with a simplified transparent transport layer that does not require all the functionalities of UDP-Lite.

2.2. RTP-H.263+

In order to transmit H.263+ video over IP networks, the H.263+ bitstream must first be packetized. A payload format for H.263+ video has been defined for use with RTP (RFC 2429) [14]. This payload format for H.263+ can also be used with the original version of H.263. In our experiments, the group of block (GOB) mode was selected for the H.263+ coder and packetization was always performed at GOB boundaries, that is, each RTP packet contains one

or more complete GOBs. Since every packet begins with a picture or GOB start code, the leading 16 zeros are omitted in accordance with RFC 2429 [14]. The packetization overhead then consists only of the RTP/UDP/IP headers, which are typically 40 bytes per packet. This overhead can be significant at low bit rates for wireless network-based applications. It is important to improve the packetization efficiency in such cases [15]. To minimize the packetization header overhead, each RTP packet should be as large as possible. On the other hand, in the presence of channel impairments, the packet size should be kept small to minimize the effects of lost packets on reconstructed video quality.

3.   PACKET VIDEO OVER WIRELESS NETWORKS

Knowledge of the radio propagation characteristics is usually a prerequisite for effective design and operation of a communication system operating in a wireless environment. The fading characteristics of different radio channels and their associated effect on communication performance have been studied extensively in the past [16]. Despite the fact that Rayleigh fading is the most popular model, Rician fading is observed in mobile radio channels as well as in indoor cordless telecommunication (CT) systems [16]. In a cellular system, Rayleigh fading is often a feature of large cells, while for cells of smaller diameter, the envelope fluctuations of a received signal are observed to be closer to Rician fading. A slow and flat Rician fading model is assumed here,[1] where the duration of a symbol waveform is sufficiently short so that the fading variations cause negligible loss of coherence within each received symbol. At the same time, the symbol waveform is assumed to be sufficiently narrowband (sufficiently long in duration) so that frequency selectivity is negligible in the fading of its spectral components. As a result, the receiver can be designed and analyzed on the basis of optimal symbol-by-symbol processing of the received waveform, for example, by a sampled matched filter or other appropriate substitute in the same manner used in the nonfading case.

3.1.   Channel-induced loss models

In this work, we restrict our attention to a random loss model, that is, the wireless channel is characterized by uncorrelated bit errors. This is a reasonable model for a fairly benign wireless channel under the assumption of sufficient interleaving to randomize the burst errors produced in the decoder.

By means of FEC, some of these bit errors can be corrected. Depending on the FEC code parameters and the channel conditions, there will be residual bit errors. Generally, over existing wired IP networks, UDP is configured to discard any packet with even a single error detected in the entire packet including the header, although UDP itself need

not implement this error-detecting functionality. In the wireless video telephony system described by Cherriman et al. [17], such packets are also discarded without further processing. In this paper, we will define two channel-induced loss models. For the first model, we assume the same loss model as used in wired IP networks; that is, a packet is accepted only if there is no error in the entire packet including the header as well as the payload, otherwise, it is considered lost. This model corresponds to a transport scheme allowing only error-free packets (denoted as scheme 1 in this paper). So, for an interference-limited wireless channel, like the CDMA radio interface, the packet losses are primarily the results of frequent bit errors instead of congestion as in a wired network. The channel-induced impairment to the video quality is in the form of these packet losses. If a packet is considered lost, the RTP sequence number enables the decoder to identify the lost packets so that locations of the missing GOBs are known. The missing blocks can then be concealed by motion-compensated interpolation using the motion vector of the macroblock (MB) immediately above the lost MB in the same frame, or else the motion vector is assumed to be zero if this MB is missing. However, if too many packets are lost, concealment itself is no longer effective in improving the reconstructed video quality.

For the second model, we assume that the transport layer is transparent to the application layer; that is, a packet with errors only in the payload is not simply discarded in the transport layer. Such a transparent transport layer can be achieved by using, for example, UDP-Lite as proposed in [1]. However, UDP-Lite provides other functionalities not necessary for the work here and is not widely deployed. As a result, we employ a simplified transparent transport protocol which limits the use of the checksum only on the RTP/UDP/IP header and discards a packet only if there is an error detected in the header. In this case the application layer should be able to access the received data although such data may have one or more bit errors. This model corresponds to a transport scheme allowing bit errors in the payload (denoted as scheme 2 in this paper). The channel-induced impairment to the video quality is then in the form of residual bit errors in the video stream. It is the responsibility of the application layer to deal with these possible bit errors. Specifically, here we make use of the H.263+ coding scheme where, based on syntax violations, certain error patterns may be detected by the video decoder and the use of the corresponding errored data can be avoided by employing passive error-recovery (PER) techniques.

Our intention is to quantitatively compare these two channel-induced loss models, identify the different video data protection requirements for wired and wireless networks, and describe the corresponding appropriate transport schemes for packet video delivery over such networks.

3.2.   Physical channel model

The bitstreams are modulated before being transmitted over a wireless link. During transmission, the modulated bitstreams typically undergo degradation due to additive white

---

[1] The slow and flat Rician channel model is completely described in terms of the single parameter ζ^2 representing the ratio of specular-to-diffuse energy.

Gaussian noise (AWGN) and/or fading. At the receiver side, the received waveforms are demodulated, channel decoded, and then source decoded to form the reconstructed video sequence. The reconstructed sequence may differ from the original sequence due to both source coding errors and possible channel-error effects.

In this paper, the symbol transmission rate for the wireless links is set to be $r_S = 64$ Ksps, such that the overall bit rate employing QPSK modulation is constrained as $R_{\text{tot}} = 128$ Kbps. This in turn sets the upper limit for the bit rate over the wired networks to be $R_{\text{tot}} = 128$ Kbps as well. Since the total bit rate is limited by the wireless links, the use of RS and/or RCPC codes will result in a decrease of source coded bit rate proportional to the overall channel coding rates.

The transmission channel is modelled as a flat-flat Rician channel with ratio of specular-to-diffuse energy $\zeta^2 = 7$ dB.

3.3. RCPC channel codes

The class of FEC codes employed for the wireless IP network in this work is the set of binary RCPC codes described in [18]. With P representing the puncturing period of the code, the rates of the codes that may be generated by puncturing a rate $R_c = 1/n$ mother code are $R_c = P/(P + j)$, $j = 1, 2, \ldots, (n - 1)P$. Thus, it is easy to obtain a family of codes with unequal error correcting capabilities. In this work, a set of RCPC codes are obtained by making use of an $R_c = 1/4$ mother code with memory $M = 10$ and a corresponding puncturing period $P = 8$. Then the available RCPC codes are of rates, $R_c = 8/9, 8/10, \ldots, 8/32$.

3.4. Passive error recovery

If a packet is considered lost, the RTP sequence number enables the decoder to identify the lost packets so that locations of the missing data are known. The affected blocks can then be concealed by PER techniques. In this work, we make use of the error-detecting and recovery scheme described in Test Model 8 [19]. The major objective of this PER scheme is to detect the severe error patterns and prevent the use of such errors which may substantially degrade the video quality. The remaining undetected error patterns in the payload which are not detected by the H.263+ decoder will result in the use of incorrectly decoded image data which can cause quality degradation of the reconstructed video.

3.5. Selected simulation results

We present some selected results for a representative quarter common intermediate format (QCIF) video conferencing sequence, Susie at 7.5 frames per second (fps). These results were obtained using a single-layer H.263+ coder in conjunction with RCPC channel codes [18] together with quadrature phase shift keyed (QPSK) modulation. To decrease the sensitivity of our results to the location of bit errors, a sequence of $N_f = 30$ input frames is encoded, channel errors are simulated and the resulting distortion is averaged. Furthermore, each simulation was run N_t times. By taking empirical averages with N_t sufficiently large (i.e., $N_t = 1000$), statistical confidence in the resulting distortion can be achieved.
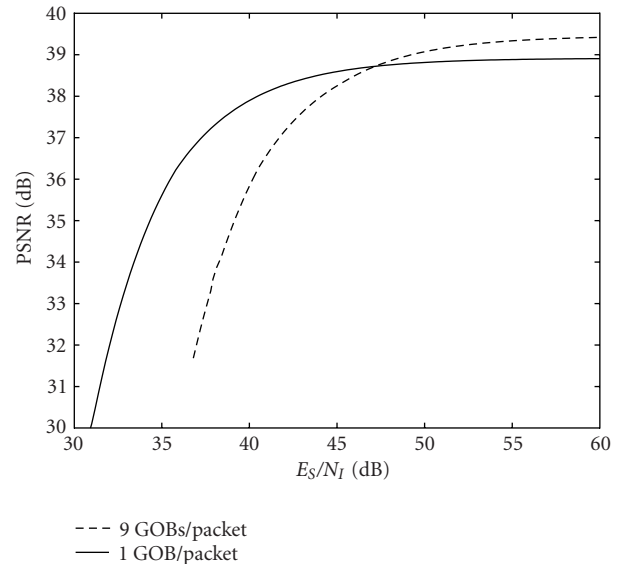

FIGURE 2: Performance of RTP-H.263+ packet video with 1 or 9 GOBs/packet over a wireless channel without channel coding and employing loss model 1; Rician channel with $\zeta^2 = 7$ dB.

Figure 2 demonstrates results for a system without channel coding under the assumption of the first loss model. Here, we plot the reconstructed peak signal-to-noise ratio (PSNR) versus the channel SNR, E_S/N_I.[2] In Figure 2, we provide results for two packetization choices which packetize either 1 or 9 GOBs (i.e., 1 frame for QCIF) into a single packet. It should be obvious that in the absence of channel impairments, the more GOBs contained in one packet, the better the quality should be as a result of the reduced overheads. This is clearly demonstrated in Figure 2 where for large E_S/N_I, the larger number of GOBs/packet results in improved PSNR performance. However, as the channel conditions degrade (i.e., the value of E_S/N_I decreases), a packetization scheme with fewer GOBs/packet can be expected to be more robust in the presence of the increasing channel impairments. This is because of the dependence of packet-loss rate upon the corresponding packet size. Although the bit-error rate remains the same, a larger packet size results in larger packet-loss rate. This is also demonstrated in Figure 2. It should also be noticed that under the first loss model, the video quality is extremely sensitive to packet losses due to the channel variation in E_S/N_I.

Next, we demonstrate the performance of the system with a transparent transport layer; that is, channel-loss model 2. We provide corresponding results in Figure 3 for both loss models for two packetization choices which again packetize 1 or 9 GOBs (i.e., 1 frame for QCIF) into a single packet. If a single GOB is packetized into a packet, the quality of the second transport scheme degrades somewhat

---

[2]The quality E_S/N_I represents the ratio of energy per symbol to the spectral density of the channel noise or interference level.
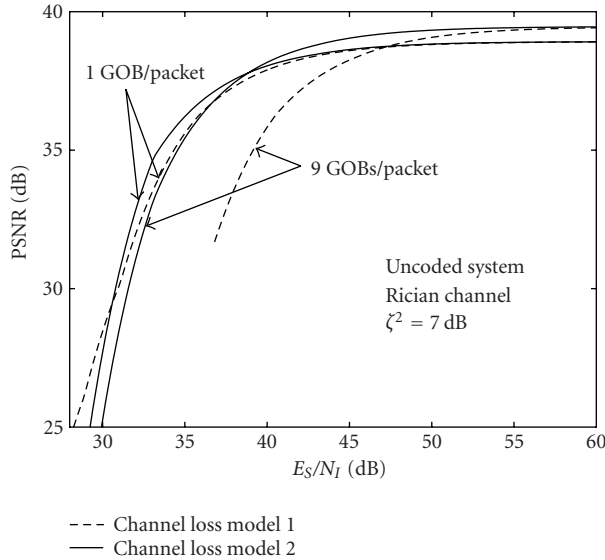
FIGURE 3: Performance of RTP-H.263+ packet video with 1 or 9 GOBs/packet over a wireless channel without channel coding for the two loss models.
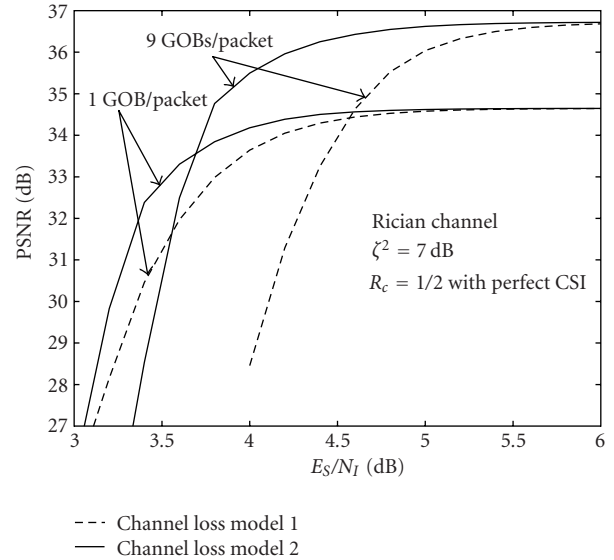

FIGURE 4: Performance of RTP-H.263+ packet video with 1 or 9 GOBs/packet over a wireless channel with a fixed $R_c = 1/2$, $M = 10$ convolutional code for the two loss models.

more gracefully compared to the first scheme as the channel E_S/N_I decreases. The relative disadvantage of the first scheme in this case is the result of discarding packets with even a single bit error in the payload. Instead, the second scheme makes use of the received data by selectively decoding those data without severely degrading the video quality. Since the packet size in this case is relatively small, as the bit error rate increases as a result of decreasing E_S/N_I, there is some advantage of the first scheme in the region $E_S/N_I < 31$ dB because it avoids the use of error-prone packets. For scheme 2, on the other hand, the remaining undetected errors in the payload begin to overwhelm the PER capabilities of the decoder as E_S/N_I decreases and substantially degrade the reconstructed video quality. This is also demonstrated in Figure 3. However, it should be noticed that in this region the video quality is already sufficiently degraded that the relative advantage of scheme 1 in this region does not make a significant difference for video users. Furthermore, as illustrated in Figure 3, if 9 GOBs are packetized into a packet, the quality of the second transport scheme substantially outperforms the first scheme as the channel E_S/N_I becomes smaller. As the packet size increases, the disadvantage of the first scheme is even more significant as a result of discarding packets with even single bit error in the payload. Based on these observations, it would appear that it is necessary to provide a transparent transport scheme for packet video over wireless networks. More specifically, packet video over wired and wireless IP networks may have to employ different transport-layer protocols.

FEC can be used to protect the video data against channel errors to improve the video delivery performance in the range of lower E_S/N_I, although, as we demonstrate, the

choice of channel coding rate must be carefully made. For example, the corresponding results for the previous two packetization choices are illustrated in Figure 4 for the two loss models where we somewhat arbitrarily employ an $R_c = 1/2$, $M = 10$ convolutional code to protect the packetized video data. In this case, the additional channel coding overheads force a decrease in the available source coding bit rate,[3] and this results in a corresponding decrease in the video quality in the absence of channel impairments. This can be seen if we compare the results in Figure 4 to the corresponding values in Figure 3 for large E_S/N_I. However, it should be noted that the coded cases can maintain the video quality at acceptable levels for considerably smaller values of E_S/N_I compared to the uncoded system. This is a good indication of the necessity of employing FEC coding in wireless networks.

It should also be observed in Figure 4, compared to the uncoded case illustrated in Figure 3, that the second loss model consistently and substantially outperforms the first loss model. For example, there is over 6 dB performance gain of the second model over the first model at $E_S/N_I = 4$ dB for the case of 9 GOBs/packet. This suggests the advisability of using FEC coding to constrain the bit-error rate in wireless networks together with the use of a transparent transport-layer scheme to provide acceptable packet video services. This provides further illustration that packet video transport over wireless IP networks may require a different transport-layer protocol from conventional wired networks in order to obtain more desirable error-resilient quality.

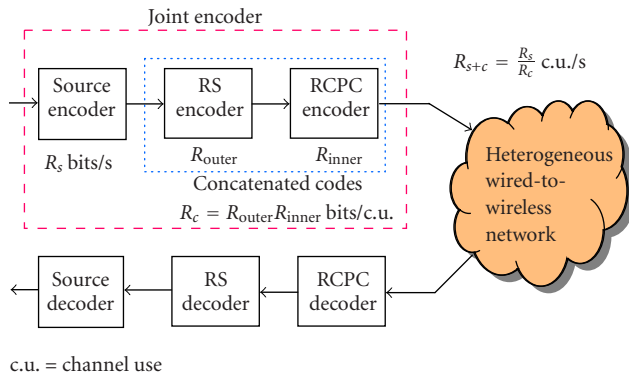[3]Recall that we are holding the total transmitted bit budget at $R_{tot} = 128$ Kbps.

FIGURE 5: Illustration of concatenated coding scheme.

4. PACKET VIDEO OVER WIRED-TO-WIRELESS IP NETWORKS

Many evolving multimedia applications will require video transmission over a wired-to-wireless link such as in wireless IP applications where a mobile terminal communicates with an IP server through a wired IP network in tandem with a wireless network as illustrated in Figure 1. We intend to address an end-to-end solution for video transmission over a heterogeneous network such as the UMTS third-generation (3G) wireless system, which provides the flexibility at the physical layer to introduce service-specific channel coding as well as the necessary bit rate required for high-quality video up to 384 Kbps.

Video quality should degrade gracefully in the presence of either packet losses due to congestion on the wired network, or bit errors due to fading conditions on the wireless channel. Due to the difference in channel conditions and loss patterns between the wired and wireless networks, to be efficient and effective the error-control schemes should be tailored to the specific characteristics of the loss patterns associated with each network. Furthermore, the corresponding error-control schemes for each network should not be designed and implemented separately, but jointly in order to optimize the quality of the delivered video.

Here, we present a possible end-to-end solution which employs an adaptive concatenated FEC coding scheme to provide error-resilient video service over tandem wired-to-wireless IP networks as illustrated in Figure 5. An H.263+ source coder encodes the input video which is applied to a concatenated channel encoder employing an RS block outer encoder and an RCPC inner encoder. The RS outer code operates in an erasure-decoding mode and provides protection against packet loss due to congestion in the wired IP network while the RCPC inner code provides protection against bit errors due to fading and interference on the wireless network. The RS coding rates can be selected adaptively according to the prevailing network conditions, specifically, packet-loss rate for the wired IP network. This channel rate matching is achieved by employing a set of RS codes with different erasure-correcting capabilities. The RCPC coding rates can also be selected adaptively to provide different levels of bit-

error-correcting capability according to the prevailing wireless network conditions, specifically, E_S/N_I for the wireless channels.[4] This end-to-end approach avoids the system complexities associated with transcoding in edge proxies located at the boundaries between the wired and wireless networks as treated in [2], for example. However, we will see that this reduction in complexity is at the expense of a considerable performance penalty.

4.1. Packet-level FEC scheme for wired IP networks

Packet loss is inevitable even in wired IP networks, and can substantially degrade reconstructed video quality which is annoying for users. Thus, it is desirable that a video stream be robust to packet loss. Regarding the tight delay constraints for real-time video applications, FEC should be applied to achieve error recovery when packet losses occur. For a wired IP network, packet loss is caused primarily by congestion, and channel coding is typically used at the packet-level [20, 21] to recover from such losses. Specifically, a video stream is first chopped into segments each of which is packetized into k packets, and then for each segment, a block code is applied to the k packets to generate an n-packet block, where $n > k$. To perfectly recover a segment, a user only needs to receive any k packets in the n-packet block. To avoid additional congestion problems due to channel-coding overheads, a JSCC approach to optimize the rate allocation between source and channel coding is necessary. One such approach employing interlaced RS coding with packet-loss-recovery capability has been described in [22].

In this paper, we will apply a form of concatenated FEC coding employing interlaced RS codes as illustrated in Figure 6, where FEC codes are applied across IP packets. Specifically, each packet is partitioned into successive m-bit symbols to form an encoding array, and individual symbols are aligned vertically to form RS codewords of block length n over $GF(2^m)$. As illustrated in Figure 6, each IP packet consists of w successive rows of m-bit symbols, then, the decoded packet-loss probabilities can be readily determined assuming erasure-only decoding.

4.2. Packetization for the interlaced RS coded video data

To quantitatively compare the performance between a coded system and an uncoded system, we have to maintain the same packet-generation rate. Specifically, for the QCIF video studied in this paper, in the uncoded system, each GOB is packetized into a single packet, resulting in 9 packets per video frame. For the coded system, network packets are obtained by concatenating successive rows of the encoding array illustrated in Figure 6. We maintain identical packet rate in the coded system as in the uncoded system. Specifically, with the use of $RS(63, k)$ codes, this results in packing 7 (i.e., $w = 7$ in Figure 6) coded symbols from the same RS codeword into the same packet together with other RS coded symbols from

---

[4]The RCPC rates should also depend on the Rician channel parameter ζ^2 which for purposes of this work we will assume is fixed and known.
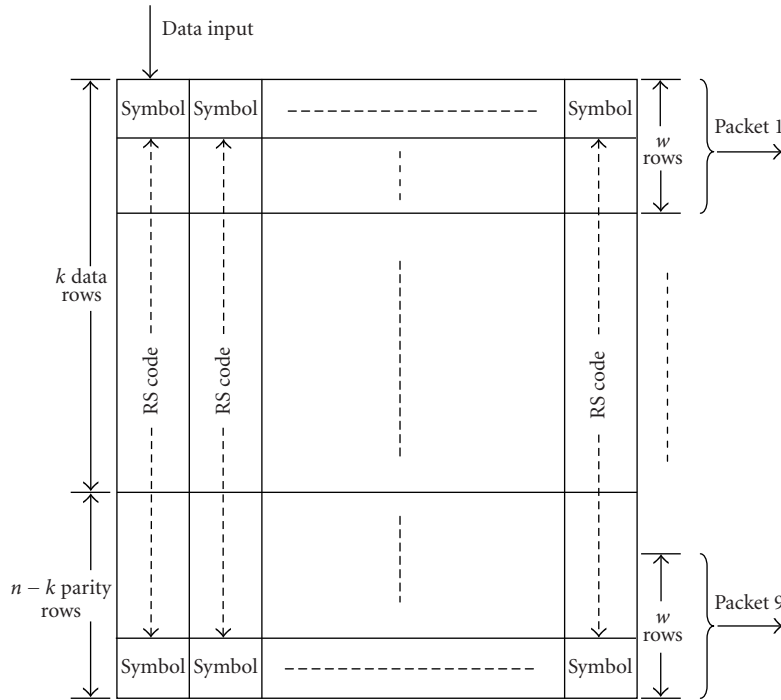
FIGURE 6: Illustration of interlaced RS codes.

the same video frame. As a result, both systems will generate 9 packets per frame.

4.3. Packet-loss correction using RS codes

Consider an $RS(n, k)$ code over $GF(2^m)$ applied in an interlaced fashion across the IP packets as described above and illustrated in Figure 6. Here, k symbols of m bits each are encoded into n m-bit symbols with d the minimum distance of the RS code given by

$$d = n - k + 1. \qquad (1)$$

For the proposed concatenated FEC scheme, it is possible that there are residual bit errors that cannot be corrected through the use of the inner RCPC codes. These residual bit errors may degrade the erasure-correction capability of the RS codes employing erasure decoding which attempts to correct the packet-loss-induced symbol erasures over the wired IP network. However, the probability of symbol errors for the RS coded symbols resulting from such residual bit errors will be very small compared to the symbol-erasure rate with appropriate choices of inner RCPC codes which maintain the residual bit-error rate low. For example, considering an $RS(63, k)$ code with a symbol size of 6 bits, a residual bit-error rate of 10^{-5} will result in a symbol-error rate of 6×10^{-5} which will have a negligible effect on the erasure correcting performance of the RS codes in a system where packet-loss-induced erasures are dominant. So, in this paper we assume the use of erasure-only decoding of RS codes with full erasure-correcting capability.

For an RS code with erasure decoding, $e \leq d - 1$ erasures can be corrected. Consider that w m-bit symbols from an RS codeword are packed into the same packet. A packet loss under this packetization scheme will result in w erasures for the corresponding RS coded symbols. Assume the symbol erasures are independent. For the coded system, the resulting packet-loss rate for the above specified packetization scheme then becomes

$$P_L = \sum_{i=W}^{9} \binom{9}{i} \lambda^i (1 - \lambda)^{9-i}, \qquad (2)$$

where λ is the corresponding uncoded packet-loss rate, and W is the maximum number of allowable packet losses that can be recovered through the use of RS codes, and is given by

$$W = \lceil e/w \rceil. \qquad (3)$$

It should be noted that a lost packet in the uncoded system as described above will result in a loss of 1 GOB. However, for the coded system, if there is a packet loss that cannot be recovered through the erasure-correcting capability of the corresponding RS codes, the whole frame, that is 9 GOBs, will be affected due to the interlaced RS coding scheme. In such a situation, PER, as will be described in Section 4.4, will be applied to conceal the errors.

4.4. Channel-induced loss models

In the previous section, we have shown the advantage of a transparent transport layer for video transmission over noisy wireless channels. In what follows, we will again assume that
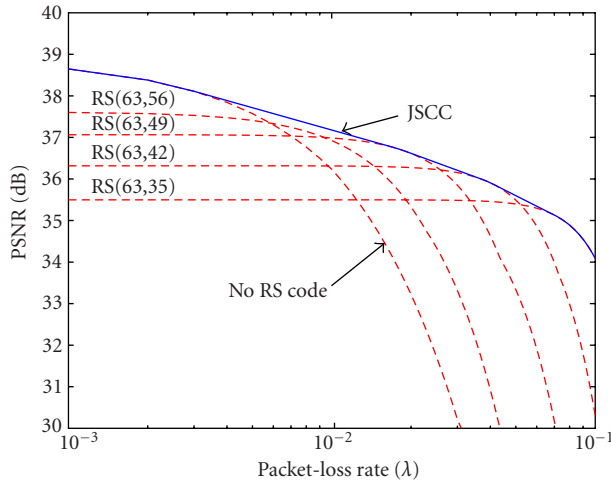
FIGURE 7: Performance of RTP-H.263+ packet video over wired IP networks using RS coding alone.
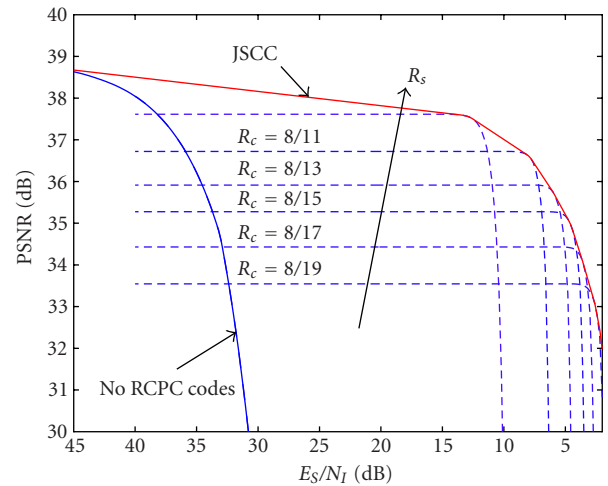

FIGURE 8: Performance of H.263+ coded video delivery over a wireless Rician fading channel with $\zeta^2 = 7$ dB using JSCC approach with RCPC coding only and employing perfect CSI. Performance results for a set of fixed channel coding rate schemes are also shown.

the transport layer is transparent to the application layer, that is, a packet with errors in the payload is not simply discarded in the transport layer. Instead, the application layer should be able to access the received data although such data may have one or more bit errors. It is the responsibility of the application layer to deal with the possible residual bit errors as described previously in Section 3.1.

4.5. JSCC approach

As has been demonstrated in the previous section, in order to protect against the channel impairments, some form of FEC coding must be employed. Since an arbitrarily chosen FEC design can lead to a prohibitive amount of overhead for highly time-varying error conditions over wireless channels, a JSCC approach for image or video transmission is necessary. The objective of JSCC is to jointly select the source and channel coding rates to optimize the overall performance due to both source coding loss and channel-error effects subject to a constraint on the overall transmission bit rate budget.

In [9, 10], it was shown that much of the computational complexity involved in solving this optimal rate allocation problem may be avoided through the use of universal distortion rate characteristics. Given a family of universal distortion rate characteristics for a specified source coder, together with appropriate bounds on bit-error probability P_b for a particular modulation/coding scheme as a function of channel parameters, the corresponding optimal distortion rate characteristics for a video sequence can be determined through the following procedure: for a specified channel SNR, E_S/N_I, we can find the associated P_b through the corresponding bit-error probability bounds for a selected modulation/coding scheme as discussed earlier. Then, for each choice of source coding rate R_s of interest, use the resulting P_b to find the corresponding overall PSNR from the universal distortion rate characteristics. This procedure is described in more detail in [9, 10].

4.6. Selected simulation results

We first consider the case where no channel error is introduced over the wireless links; that is, only the packet loss over the wired network will degrade the video quality. Figure 7 demonstrates the performance using a family of RS$(63, k)$ codes[5] with JSCC for RTP-H.263+ packet video over wired IP networks experiencing random packet loss. Here we illustrate PSNR results as a function of packet-loss rate λ for different values of source coding rate with the RS codes chosen to achieve the overall bit rate budget $R_{tot} = 128$ Kbps. In particular, the smaller values of R_s allow the use of more powerful low-rate RS codes resulting in improved performance for larger packet-loss rate. On the other hand, for small packet-loss rate performance, improvements can be obtained using larger values of R_s together with less powerful high-rate RS codes. The optimum JSCC procedure selects the convex hull of all such operating points as illustrated schematically in Figure 7. Clearly, compared to the system without using RS coding where video quality is substantially degraded with increasing packet-loss rate, the JSCC approach with RS coding provides an effective means to maintain the video quality as network-induced packet-loss rate increases.

Consider another case where now bit errors over the wireless links instead of packet loss over the wired network are dominant, and a JSCC approach using RCPC codes is employed. The results are illustrated in Figure 8 where we now plot PSNR versus E_S/N_I.[6] Again, as can be observed, the JSCC approach with RCPC coding alone clearly demonstrates significant performance improvements over either the uncoded case or the case where the channel coding rate is fixed at

---

[5]RS$(63, k)$ codes are used throughout the remainder of this paper.
[6]Observe the decreasing values of E_S/N_I used in plotting Figure 8.
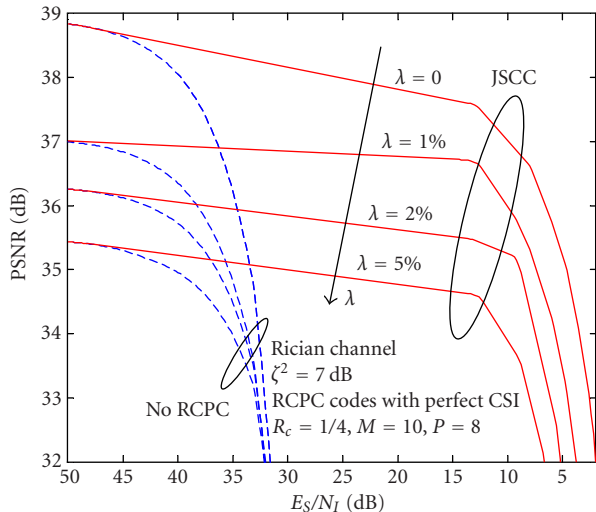
FIGURE 9: Performance of H.263+ coded video delivery over heterogeneous wired-to-wireless IP networks using JSCC employing concatenated RS and RCPC coding.

FIGURE 10: An end-to-end approach using an edge proxy.

an arbitrarily chosen value.[7] The use of JSCC can provide a more graceful pattern of quality degradation by keeping the video quality at an acceptable level for a much wider range of E_S/N_I. This is achieved by jointly selecting the channel and source coding rates based on the prevailing channel conditions, here represented by E_S/N_I.

In more general cases, packet loss due to congestion in the wired network and bit errors due to fading effects on the wireless networks coexist. We propose to jointly select the source coding rate, the RS coding rate, and the RCPC coding rate such that optimal end-to-end performance can be achieved with this concatenated coding scheme. Here, we demonstrate PSNR results for reconstructed video as a function of the wireless channel E_S/N_I for a set of packet-loss rates over the wired IP network with the RS codes and RCPC codes chosen to achieve the overall bit rate budget $R_{\text{tot}} = R_s/(R_c^{\text{RCPC}} \cdot R_c^{\text{RS}}) = 128$ Kbps [3]. In Figure 9, for a given packet-loss rate λ in the wired network, the optimal performance obtainable is demonstrated under the constraint of a fixed wireless transmission rate. It is clear that the RS coding rate has to be adaptively selected with the variation in the corresponding packet-loss rate. Meanwhile, the RCPC coding has to adapt to the change in the wireless link conditions, E_S/N_I in this case. Clearly, as shown by the dashed lines in Figure 9, for the system employing only adaptive RS codes selected according to the packet-loss rate on the wired network but no RCPC codes on the wireless network, video quality is substantially degraded with increasing bit errors as E_S/N_I decreases. In contrast, the JSCC approach with concatenated RS and RCPC coding provides an effective means

to maintain the video quality as network-induced packet-loss and/or bit-error rate increase.

5. PACKET VIDEO OVER WIRED-TO-WIRELESS IP NETWORK USING AN EDGE PROXY

In the previous section, we investigated a JSCC approach used with a concatenated FEC coding scheme employing interlaced RS block codes and RCPC codes to actively protect the video data from different channel-induced impairments over tandem wired and wireless networks. However, this approach is not optimal since, as noted previously, the coding overhead required on the wired link must also be carried on the wireless link.

As an alternative to the concatenated approach, we present an end-to-end solution with the use of an edge proxy operating at the boundary of the two networks as demonstrated in Figure 10. This end-to-end solution employs the edge proxy to enable the use of distinctly different error-control schemes on the wired and wireless networks. Specifically, we employ the interlaced RS codes alone on the wired network and the RCPC codes alone on the wireless network to provide error-resilient video service over tandem wired-to-wireless IP networks. As a result, under the constraint of a total bitrate budget R_{tot}, the effective video data throughput is given as $R_s = \min\{R_{\text{tot}} \cdot R_c^{\text{RS}}, R_{\text{tot}} \cdot R_c^{\text{RCPC}}\}$, where R_c^{RS} and R_c^{RCPC} are the channel coding rates for the RS and RCPC codes, respectively. In contrast, without the use of an edge proxy, these two codes have to work as a concatenated FEC scheme as described in the preceding section in order to provide sufficient protection against both congestion-caused packet loss in the wired network and fading-caused bit errors in the wireless network. The corresponding effective video data throughput in this case is then $R_s = R_{\text{tot}} \cdot R_c^{\text{RS}} \cdot R_c^{\text{RCPC}}$ and, because of the need to carry both overheads on both networks, this causes a serious reduction in achievable video quality. It is clear then that the reconstructed video quality can be improved through the use of an edge proxy. We will quantitatively investigate the resulting improvement for interactive video coding and transmission in what follows.

---

[7]For example, the arbitrary choice of $R_c = 1/2$ illustrated in Figure 4 would fall between the curves labelled $R_c = 8/15$ and $R_c = 8/17$ in Figure 8.
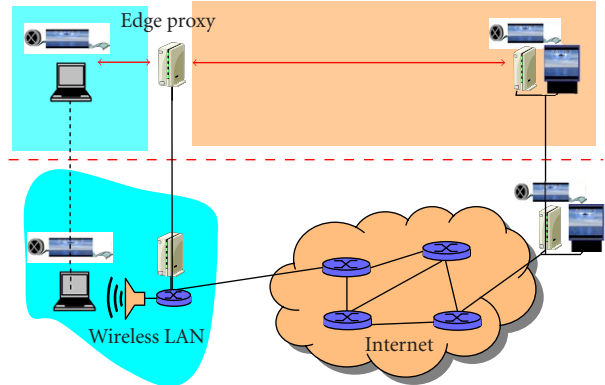
5.1. Edge proxy

To accommodate the differential error-control schemes as well as differential transport protocols for packet video over wired and wireless networks, appropriate middleware has to be employed to operate between the wired and wireless network to support the application layer solutions for video applications. Thus, we define an edge proxy here to accomplish these functionalities. The edge proxy should be implemented as part of a mobile support station. Furthermore, it should be application-specific; in our case it is video-oriented.

The use of edge proxies at the boundaries of dissimilar networks for a variety of functions have been used extensively in the networking community [23]. The uniqueness of the approach proposed here using edge proxies at the boundary between wired and wireless networks for video transport applications lies in its specific functionalities as defined above. Specifically, it serves as an agent to enable and implement

(1) selective packet relay,
(2) error-control transcoding,
(3) JSCC control,
(4) interoperation between different possible transport protocols for the wired and wireless network.

For the interactive applications we consider here, there exists two-way traffic including wired-to-wireless as well as wireless-to-wired. We assume that RS codes are employed to combat packet loss due to congestion in a wired network, and RCPC codes are used on the wireless network to combat bit errors. It is necessary for the edge proxy to do error-control transcoding if such a scheme is used.

Furthermore, the edge proxy should support the JSCC control scheme to adaptively adjust the source and channel coding rates. To avoid computation and time-expensive video transcoding in the edge proxy, an end-to-end adaptive coding control strategy is suggested here. The channel conditions including those for both the wired and wireless networks are collected in the edge proxy, and based on the prevailing channel conditions, video coding rates are adjusted accordingly using JSCC. For the wired network, the major channel condition parameter is the packet-loss rate, while for the wireless network, channel SNR as well as the fading parameters are used.

The edge proxy is also responsible for the interoperation between different possible transport protocols for the wired and wireless network. For a wireless network, the error-control scheme is implemented in the application layer, and erroneous packets should be delivered to the end user. However, for conventional wired networks, such as existing IP networks, no error is allowed. In this case, to achieve interoperation, the edge proxy has to repacketize the packet according to the appropriate transport protocol before relaying the packet in either direction.

5.2. Selected simulation results

Now we consider the system with the use of an edge proxy between the wired and wireless IP networks, such that error-
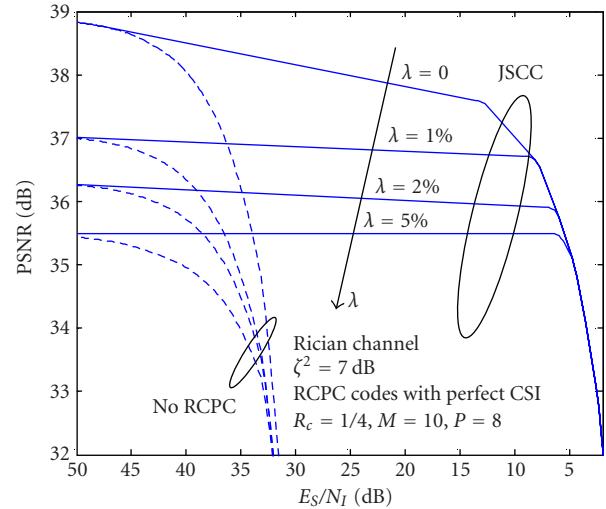

FIGURE 11: Performance of H.263+ coded video delivery over heterogeneous wired-to-wireless IP networks using JSCC with an edge proxy.
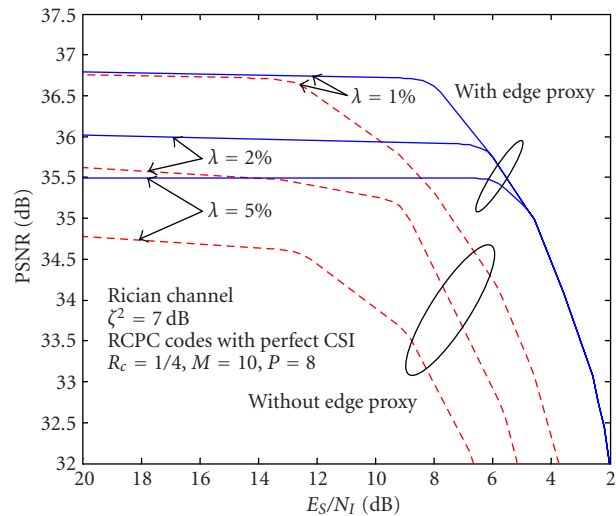

FIGURE 12: Relative performance improvement with and without the use of an edge proxy.

control transcoding can be done between the two heterogeneous networks each supporting different error-control approaches as described previously. With the use of an edge proxy, the corresponding optimal performance obtainable is demonstrated in Figure 11 under the constraint of the same fixed wireless transmission rate of 128 Kbps.

For comparison, we also present in Figure 12 the results for the systems with or without the use of an edge proxy under the same transmission rate limit, which have been shown previously in Figures 11 and 9, respectively. It clearly

demonstrates the substantial improvement using an edge proxy. For example, in the case that packet-loss rate over the wired IP network is $\lambda = 5\%$, there is a gain of over 6 dB in wireless channel E_S/N_I for a specified video quality of PSNR $= 34$ dB. This improvement is primarily due to the increase of effective video data throughput due to the error-control transcoding in the edge proxy. As a result, to meet the same error protection requirement for both wired and wireless network conditions, a larger effective video data throughput can be achieved through the use of an edge proxy compared to the case without an edge proxy.

6. SUMMARY AND CONCLUSIONS

We quantitatively demonstrate the requirements for different transport-layer schemes for packet video over wireless networks from the requirements for conventional wired networks. Then we described the possible end-to-end solutions with and without an edge proxy operating between the wired and wireless network for packetized H.263+ video over heterogeneous wired-to-wireless IP networks. A JSCC approach employing RS block codes and RCPC codes is studied for the two proposed architectures. The results quantitatively demonstrate the requirement for a joint design approach to address the special needs of error recovery for packet video over the wireless and wired network for acceptable end-to-end quality while exhibiting a graceful pattern of quality degradation in the face of dynamically changing network conditions. Furthermore, the results clearly demonstrate the advantage of using an edge proxy with clearly defined functionalities in heterogeneous wired-to-wireless IP networks for improved video quality.

REFERENCES

[1] L.-Å. Larzon, M. Degermark, and S. Pink, "UDP Lite for real time multimedia applications," in *Proc. IEEE International Conference of Communications*, Vancouver, BC, Canada, June 1999.

[2] Y. Pei and J. W. Modestino, "Robust packet video transmission over heterogeneous wired-to-wireless IP networks using ALF together with edge proxies," in *Proc. European Wireless Conference*, Florence, Italy, February 2002.

[3] Y. Pei and J. W. Modestino, "Use of concatenated FEC coding for real time packet video over heterogeneous wired-to-wireless IP networks," in *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 840–843, Bangkok, Thailand, May 2003.

[4] M. J. Ruf and J. W. Modestino, "Operational rate-distortion performance for joint source and channel coding of images," *IEEE Trans. Image Processing*, vol. 8, no. 3, pp. 305–320, 1999.

[5] P. C. Cosman, J. K. Rogers, P. G. Sherwood, and K. Zeger, "Combined forward error control and packetized zerotree wavelet encoding for transmission of images over varying channels," *IEEE Trans. Image Processing*, vol. 9, no. 6, pp. 982–993, 2000.

[6] P. G. Sherwood, X. Tian, and K. Zeger, "Efficient image and channel coding for wireless packet networks," in *Proc. IEEE International Conference on Image Processing*, pp. 132–135, Vancouver, BC, Canada, September 2000.

[7] J. Hua and Z. Xiong, "Optimal rate allocation in scalable joint source-channel coding for image transmission over CDMA networks," in *Proc. International Conference on Multimedia and Expo*, Baltimore, Md, USA, July 2003.

[8] V. Stankovic, R. Hamzaoui, and Z. Xiong, "Joint product code optimization for scalable multimedia transmission over wireless channels," in *Proc. International Conference on Multimedia and Expo*, Lausanne, Switzerland, August 2002.

[9] M. Bystrom and J. W. Modestino, "Combined source-channel coding schemes for video transmission over an additive white Gaussian noise channel," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 880–890, 2000.

[10] M. Bystrom and J. W. Modestino, "Combined source-channel coding for transmission of H.263 coded video with trellis-coded modulation over a slow-fading Rician channel," in *Proc. IEEE International Symposium on Information Theory*, MIT, Cambridge, Mass, USA, August 1998.

[11] T. Chu and Z. Xiong, "Combined wavelet video coding and error control for Internet streaming and multicast," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 1, pp. 66–80, 2003.

[12] D. D. Clark and D. L. Tennenhouse, "Architectural considerations for a new generation of protocols," *ACM Computer Communication Review*, vol. 20, no. 4, pp. 200–208, 1990.

[13] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, *RTP: A transport protocol for real-time applications*, RFC 1889, January 1996.

[14] C. Bormann, L. Cline, G. Deisher, et al., *RTP payload format for the 1998 version of ITU-T Rec. H.263 video (H.263+)*, RFC 2429, October 1998.

[15] Y. Pei and J. W. Modestino, "A joint source-channel coding approach for packet video transport over wireless IP networks," in *Proc. 11th International Packet Video Workshop*, pp. 41–50, Kyongju, Korea, April 2001.

[16] S. Stein, "Fading channel issues in system engineering," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 68–89, 1987.

[17] P. Cherriman, C. H. Wong, and L. Hanzo, "Turbo- and BCH-coded wide-band burst-by-burst adaptive H.263-assisted wireless video telephony," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1355–1363, 2000.

[18] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Communications*, vol. 36, no. 4, pp. 389–400, 1988.

[19] Intel Corporation, "Video Codec Test Model, TMN8," June 1997, ftp://standard.pictel.com/video-site/h263plus/draft13.doc.

[20] D. Wu, T. Hou, and Y.-Q. Zhang, "Scalable video coding and transport over broadband wireless networks," *Proceedings of the IEEE*, vol. 89, no. 1, pp. 6–20, 2001.

[21] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, and J. M. Peha, "Streaming video over the internet: Approaches and directions," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 282–300, 2001.

[22] R. Kurceren and J. W. Modestino, "A joint source-channel coding approach to scalable delivery of digital video over ATM networks," in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 1599–1603, Vancouver, BC, Canada, September 2000.

[23] R. Floyd, B. Housel, and C. Tait, "Mobile Web access using eNetwork Web Express," *IEEE Personal Communications*, vol. 5, no. 5, pp. 47–52, 1998.

Yong Pei is currently a tenure-track Assistant Professor in the Computer Science and Engineering Department, Wright State University. Previously he was a Visiting Assistant Professor in the Electrical and Computer Engineering Department, University of Miami. He received his B.S. degree in electrical power engineering from Tsinghua University, Beijing, in 1996, and M.S. and Ph.D. degrees in electrical engineering from Rensselaer Polytechnic Institute in 1999 and 2002, respectively. His research interests include information theory, wireless communication systems and networks, and image/video compression and communications. He is a member of IEEE and Association for Computing Machinery (ACM).
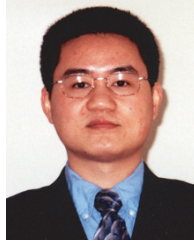
James W. Modestino received the B.S. degree from Northeastern University, Boston, Mass, in 1962, and the M.S. degree from the University of Pennsylvania, Philadelphia, Pa, in 1964, both in electrical engineering. He also received the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ, in 1968 and 1969, respectively. From 1970 to 1972, he was an Assistant Professor in the Department of Electrical Engineering, Northeastern University. In 1972, he joined Rensselaer Polytechnic Institute, Troy, NY, where until leaving in 2001 he was an Institute Professor in the Electrical, Computer and Systems Engineering Department and Director of the Center for Image Processing Research. In 2001 he joined the Department of Electrical and Computer Engineering at the University of Miami, Coral Gables, Fla, as the Victor E. Clarke Endowed Scholar, Professor and Chair. Dr. Modestino is a past member of the Board of Governors of the IEEE Information Theory Group. He is a past Associate Editor and book review editor for the IEEE Transactions on Information Theory. In 1984, he was corecipient of the Stephen O. Rice Prize Paper Award from the IEEE Communications Society and in 2000 he was corecipient of the Best Paper Award at the International Packet Video Conference.