H.263-Based Wireless Video Transmission in Multicode CDMA Systems

Mario R. Hueda and Carlos A. Marqués
Laboratorio de Comunicaciones Digitales – Universidad Nacional de Córdoba
Casilla de Correo 755 – Correo Central - Córdoba (5000) - Argentina
Email: mhueda@com.uncor.edu

Abstract – We present a simple and accurate semi-analytic methodology for performance evaluation of wireless video transmission in multicode DS-CDMA (MC-CDMA). Based on the novel approach (as well as on entire system simulations), we analyze performance of H.263 in an IS-95B system over slow fading channels at low bit rates. Peak signal-to-noise ratio (PSNR) values for various video test sequences demonstrate that important improvements of the video quality can be achieved by using feedback error control schemes. In addition, we investigate the effectiveness of the combined use of error correction and error resilience techniques for reliable video transmission in MC-CDMA systems.

I. INTRODUCTION

As a result of the evolution of wireless communications towards transmission of multimedia services, multicode direct sequence code division multiple access (MC-CDMA) has been proposed to support variable data rates [1]. In MC-CDMA (such as IS-95B [2]), high speed is provided through code aggregation. Up to eight codes may be assigned in IS-95B (one fundamental code channel (FCC) and seven supplemental code channels (SCC)), providing a maximum bit rate of 115.2 kilobits per second (kbps) [2]. However, it has been shown that multimedia services in MC-CDMA must operate at very low bit rates (<32 kbps), in order to provide an adequate capacity for voice users [3].

Video transmission is one of the most important requirements for next generation of wireless systems. H.263 is a compression standard developed to transmit video using the phone network at data rates less than 64 kbps [4]. A typical H.263 video stream is composed of inter-coded frames (I-frames) and predictive-coded frames (P-frames). An I-frame is an independently coded video frame that can be decoded by itself. A P-frame is composed of changes in the current image of the video stream relative to the last I-frame encoded. In general, an H.263 video stream is usually an I-frame followed by many P-frames, with an I-frame reintroduced to restore image quality in case of transmission errors. H.263+ is an extension of H.263 and its aim is to improve upon the compression efficiency of H.263 and broaden the range of applications for wireless transmission. However, additional protection schemes must be incorporated in transmission over slow fading channels as a result of the long block error bursts characteristic of those channels.

Recently, some papers have analyzed video transmission over MC-CDMA systems (e.g., [5]). However, none of them have considered MC-CDMA transmission over slow fading channels, as well the effects of important ingredients of the existing system architecture (e.g., RAKE receivers with maximal ratio combiner (MRC) and convolutional codes (CC) with soft-decision Viterbi decoding (SD-VD)). These important topics are addressed in this paper. We focus our study on H.263/H.263+ video transmission at low data rates (i.e., <32 kbps) in an IS-95B based system over slow fading Rayleigh channels. In order to evaluate performance of video transmission in MC-CDMA systems, a novel, simple, and accurate semi-analytic methodology is proposed. Based on the new approach, as well as on entire system simulations, we derive a variety of peak signal-to-noise ratio (PSNR) results for various H.263 video test sequences. In addition, we investigate the effectiveness of retransmission error correction and error resilience techniques to enhance the robustness of H.263/H.263+ video transmission over IS-95B systems.

II. BACKGROUND

A. H.263 Based Video Compression

The picture resolution is often QCIF (Quarter Common Intermediate Format, 176x144 pixels), which is the most common input format at low bit-rates. At QCIF resolution, each picture is divided into 11x9 macroblocks (MBs), which comprise 16x16 luminance samples, and two corresponding 8x8 blocks of chrominance samples. A fixed number of successive MBs is usually grouped into a group of blocks (GOB) and side information that is appropriate for a larger number of MBs, but not for an entire frame ("frame" refers to video), can be communicated efficiently on that level.

B. Error Concealment and Video Coding Mode

H.263 provides means to insert synchronization words at the picture and optionally at the GOB layers. We use the latter option to allow resynchronization in case of errors. We also assume that GOB sync words are inserted at the beginning of each macroblock row. This is exploited by the error concealment technique employed in the decoder, which discards corrupted GOBs and replaces the corresponding image content with data from the previously decoded frame. This technique works almost perfectly for non-moving parts of the sequence, but introduces severe distortion for moving image regions.

In transmission over mobile channels, transmission errors degrade the video quality at the receiver. In order to mitigate the effects of error transmission, periodic I-frame refresh has been proposed by several authors. However, since the overall image quality decreases due to the lower efficiency of I-frames in comparison to P-frames, and the MC-CDMA system has a severe bandwidth limitation, no periodic I-frame refresh is applied in the simulations in order to achieve the
highest coding performance for the error free case, which serves as a baseline.

C. Loss of Picture Quality

The average peak signal-to-noise ratio (PSNR) has been widely adopted as a distortion measure. For error-free transmission, the PSNR of the video reproduced at the receiver (i.e., the baseline) is given by

\[ \text{PSNR}_{\text{correct}}(t) = 10 \log_{10} \frac{255^2}{D_{SC}(t)} \]  

\[ D_{SC}(t) = \sum_{u=1}^{U} \left[ o_{u}(t) - r_{u}(t) \right]^2, \]

where \( U \) is the total number of samples of the picture at time \( t \) and \( o_{u}(t) \) and \( r_{u}(t) \) are the \( u \)-th amplitudes of the original and reconstructed pictures, respectively. Owing to transmission errors, an additional distortion \( D_{CH}(t) \) occurs at the video decoder. Thus, the overall decoded video distortion (i.e., after the error concealment stage) results in

\[ D_{o}(t) = D_{SC}(t) + D_{CH}(t). \]

Then, the PSNR of the video sequence at the decoder can be expressed as

\[ \text{PSNR}_{\text{loss}}(t) = 10 \log_{10} \frac{255^2}{D_{o}(t)}. \]

In this paper, the loss of picture quality \( \Delta \text{PSNR}(t) \), defined by

\[ \Delta \text{PSNR}(t) = \text{PSNR}_{\text{loss}}(t) - \text{PSNR}_{\text{correct}}(t) = 10 \log_{10} \frac{D_{SC}(t)}{D_{o}(t)}, \]

is used as a measure of the video degradation.

III. PERFORMANCE EVALUATION OF H.263 IN MC-CDMA TRANSMISSION USING ANALYTICAL MODELS

A. A Model for Block Error Process in MC-CDMA

Our performance studies are based on a novel Markov model for the block error process in MC-CDMA transmission which we introduced in [6]. Unlike previous work, this takes into account the presence of: (a) multicodes, (b) MRC RAKE receivers, (c) CC with SD-VD and (d) non-ideal interleaver performance (owing to transmission over slow fading channels). Consider a certain user that employs \( M \) multicodes in an MC-CDMA system. Let \( \beta_{i,n} \) be a binary process such that \( \beta_{i,n} = 1 \) if the data block \( i \) at the Viterbi decoder output of the \( n \)-th multicode is in error, and 0 otherwise. We define a new process \( \Psi_{i} \) as

\[ \Psi_{i} = \left( \beta_{i,0}, \beta_{i,1}, \ldots, \beta_{i,M-1} \right). \]

We showed in [6] that the process \( \Psi_{i} \) is well approximated by a \( 2^M \) state Markov model, with transition matrix \( \Pi_{M}(x) = \Pi_{M}(l)^x \), where

\[ \Pi_{M}(l) = \begin{bmatrix} m_{1,1} & \ldots & m_{1,2^M} \\ m_{2,1} & \ldots & m_{2,2^M} \\ \vdots & \ddots & \vdots \\ m_{2^M,1} & \ldots & m_{2^M,2^M} \end{bmatrix} \]

with \( m_{u,v} \) representing the probability of \( \Psi_{i} \text{ is in state } \Psi_{i} \text{ at time } t \), given that the superblock in slot \( i-x \) was \( \Psi_{i} \text{ at time } t \). The Markov model defined by (6) and (7) is called the "super-block Markov model" (SBMM). A method to estimate the elements \( m_{u,v} \) in transmission over slow fading channels is presented in [6].

B. A Model for the Loss of Picture Quality

Next we derive an analytical expression for the loss of P-frame picture quality owing to transmission errors, \( D_{CH}(t) \). Towards this end, we assume that an error signal is introduced at \( t=0 \) (there is no error in the video frames or \( D_{CH}(t) = 0 \) for \( t < 0 \)). The resulting error sequence is treated, before the video decoder, by the error concealment technique described previously. This introduces distortion for moving image regions, which propagates spatially and temporally until an I-frame refresh is applied at \( t = T_{max} \). Stuhlmüller et al. [7] have shown that the variance of the propagated error signal over a P-frame sequence owing to the error introduced at \( t = 0 \) can be well approximated at low error rates by

\[ \sigma_{pe}^2(t) = P_{b} \Gamma(1 + \theta t)^{-1}, \quad 0 \leq t < T_{max}. \]

\( P_{b} \) is the block error probability and depends on the transmission system and channel characteristics (e.g., channel code, number of RAKE fingers, etc.). Parameter \( \Gamma \) represents the sensitivity of the video decoder to an increase in error rate, and its value depends on several implementation issues, such as packetization, resynchronization, and error concealment, as well as the encoded video sequence. It can be considered as a constant that does not depend on the other model parameters. Note that \( P_{b} \Gamma \) is the error variance introduced in the video sequence at \( t = 0 \) [7]. The leakage \( \theta \) describes the efficiency of explicit and/or implicit (e.g., due to sub-pel motion compensation, overlapped block transmission) loop filtering to remove the introduced error. Its value depends on the strength of the loop filtering as well as on the shape of the power spectral density of the introduced error. The range of typical values is given by \( 0 < \theta < 1 \). From (8) note that the energy of the error signal decays over time due to spatial filtering in the prediction loop.

Unlike [7], in this work we are interested not only on the time averaged video distortion owing to transmission errors, but also on the "temporal progression" of the distortion value, \( D_{CH}(t) \). This model is very useful because it allows,
for example, to evaluate analytically the performance of numerous error resilience techniques. Assuming that the error process is stationary over successive frames (i.e., the effect of lost blocks is approximately constant for each transmitted data block) and the error signals are uncorrelated [7], from (8) we obtain:

\[ D_{\text{CH}}(t) = \sum_{r=0}^{T_{\text{max}}} \sigma_{\text{pe}}^2(\tau) = P_b \Gamma \sum_{r=0}^{T_{\text{max}}} \frac{1}{1+\theta \tau} = P_b \Gamma \Theta(t), \]

for \( 0 \leq t < T_{\text{max}} \), with

\[ \Theta(t) = \theta^{-1} \ln(1+\theta t) + 0.5(2+\theta t)(1+\theta t)^{-1}. \]

To derive (10) we approximated the sum in (9) by an integral using trapezoid based numerical integration).

\( P_b \) can be easily estimated from the SBMM (this analysis is not included in this paper). For a given packetization and video sequence, we estimate parameters \( \theta \) and \( \Gamma \) by fitting (5) (with \( D_{\text{SC}}(t) \) and \( D_{\text{CH}}(t) \) given by (2) and (9), respectively) to the measurement points obtained from SBMM based fast simulation. From experimental results we have verified that \( \Gamma \) and \( \theta \) are approximately constant over the range of the video rates considered in this work (i.e., <32 kbps).

C. A Model for the Overall Video Distortion

Using (9) in (3), the total distortion \( D_o(t) \) results in

\[ D_o(t) = D_{\text{SC}}(t) + P_b \Gamma \Theta(t), \quad 0 \leq t < T_{\text{max}}. \]

Then, from (11) we can derive the time averaged total video distortion:

\[ \bar{D}_o = \bar{D}_{\text{SC}} + P_b \Gamma \Theta, \]

where

\[ \Theta = (T_{\text{max}}+1)^{-1} \left[ \theta^{-1} \left( \ln(1+\theta T_{\text{max}}) - 1 \right) + 0.5(2+\theta T_{\text{max}})^{-1} \right]. \]

Note that the minimum value, \( \Theta = 1 \), occurs when there is no error propagation \( (T_{\text{max}} = 0) \). From (12) and (13) we can conclude that the minimum distortion \( \Theta = 1 \) of a given video sequence achieved by error resilience techniques (such as error-tracking [9]) in transmission over a channel with block error probability \( P_b \) is given by

\[ \min[D_o] = \bar{D}_{\text{SC}} + P_b \Gamma. \]

The distortion rate (DR) model is used to approximate \( \bar{D}_{\text{SC}} \). Thus, the average video source coding distortion can be expressed by

\[ \bar{D}_{\text{SC}} = \lambda_o (R_v - R_o)^{-1} + \zeta_o, \]

where \( R_v \) is the video source rate, while \( \lambda_o, R_o, \) and \( \zeta_o \) are the parameters of the DR model which depend on the encoded sequence [7].

IV. NUMERICAL RESULTS

To perform our simulation experiments, we use the University of British Columbia’s H.263+ Reference codec. Moreover, we utilize the rate control method discussed in TMN-8. All of the results represent around of 160 pictures of test sequences having QCIF resolution, coded at 6 frames per second (fps). Although numerous simulations have been performed, only a few results are discussed in this paper, because of space constraints. Specifically, we analyze the loss of picture quality for the test sequences MotherDaughter and Foreman. These sequences are selected because of their different characteristic in motion and spatial detail [7].

We use the parameters, interleaver/deinterleaver, and convolutional code of the downlink of the IS-95B standard [2]. Furthermore, we use soft-decision decoding of the rate \( 3/4 \), constraint length 9, convolutional code (i.e., rate set 2 of IS-95B). We set the number of RAKE fingers, \( L \), to four \( (L = 4) \). We assume that RAKE fingers have equal power. The block rate is \( 1/T_b = 50 \) block/s. A 16-bit cyclic redundancy code (CRC) is used for block error detection. Carrier frequency is 1800 MHz and the maximal Doppler frequency is \( f_d = 2 \) Hz. Since the rate variation of the channel is small, we assume ideal coherent demodulation at the RAKE receiver. The Rayleigh channel is simulated using Jakes’ model. We set the number of multicoecs to two \( (M=2) \). FCC and an SCC are used to transmit video. The rate of SCC is 14.4 kbps. The transmission rate of FCC is assumed 7.2 kbps (14.4 kbps is its maximal rate). 2.4 kbps are spent for overhead (i.e., packet headers + CRC + tail block) and the video net bit rate results in 19.2 kbps. Three channel states are analyzed in this paper: (a) “Bad”, (b) “Good” and (c) “Very Good”, which have an SCC average block error of 0.11, 0.035 and 0.017, respectively. These values correspond approximately to 15, 10 and 8 data users per cell (with one code per user) in a typical multicell environment.

A. Accuracy of SBMM

We use SBMM to evaluate H.263 video degradation in MC-CDMA transmission over slow fading channels. We also compare simulation results derived from both SBMM and entire system simulations in order to verify the accuracy of our channel modeling. It is important to realize that, although Markov models have been previously used to evaluate performance of wireless video transmission, comparisons with results obtained from simulation of the entire system have not been reported so far. These comparisons are important because it has been demonstrated that the first-order Markov approximation to model slow fading channels cannot be useful in several cases of interest (e.g., applications requiring a large number of consecutive samples)[8].

Fig. 1 shows the loss of picture quality for various video test sequences obtained from simulations. Fast simulation-based performance evaluation is achieved by using the SBMM defined by (7). We also present results from simulation of the entire system. The coded sequences are
transmitted 350 times using different: (i) realizations of the block error process (in SBMM simulations), and (ii) starting points in the fading simulator (in entire system simulations). The average $\Delta PSNR(t)$ over all the runs is presented. We can verify the excellent accuracy of the SBMM. Results also show that video quality is severely degraded in transmission over MC-CDMA. This is because, in slow fading environments, the probability that all $M$ received data blocks are in error in a deep fading is high [6], therefore long block-error bursts occur at the input of the video decoder. If we take into account that $P$-frames are being transmitted, it is easy to infer that the error propagation problem in the reproduced video is severe, as it can be seen in Fig. 1. Furthermore, we have verified that several modes incorporated in H.263+ to improve performance in wireless channels are not effective (these results are not included in this paper). From the above, we conclude that extra protection must be provided in order to improve the reliability of H.263 transmission in MC-CDMA.

B. Analysis of Error Correction Techniques

From Fig. 1 we noted that video quality is severely degraded in transmission over MC-CDMA. Improvements of video quality can be achieved by reducing the packet loss $P_B$ through error correction techniques such as automatic repeat request (ARQ) protocols. In this paper, we consider the Non-Selective Variable Bandwidth Retransmission (NSVBR) scheme, a novel bounded delay modified GBN retransmission-based scheme that uses FCC to transmit, at a proper rate (lower than its maximal), video information instead of voice. Because real-time services require a bounded delay, only one retransmission is allowed. Moreover, depending on the round trip delay and the number of MCCs, only a limited number of packets can be retransmitted. To retransmit data blocks with low time delay, the proposed approach increases the FCC data rate. Thus, our scheme not only avoids underutilizing and wasting the scarce bandwidth, but it also significantly reduces the interference and the complexity of the bandwidth allocation algorithm. Since slow fading and low data rates are assumed, it can be inferred that NSVBR provides a good tradeoff between efficacy and complexity.

Let $N$, $M$ and $R_{out}$ denote the round trip delay in slots, the number of MCCs assigned to a given user, and the total user bit rate for video, respectively. In “normal conditions” (i.e., no errors), the FCC works at a data rate, $R_{FCC_{max}}$, less than its maximal $R_{FCC_{max}}$. The value of $R_{FCC_{max}}$ will depend on the link quality (e.g., the system load), and is determined at the beginning of the communication. For example, if the average error rate of the link is moderately high, a low value of $R_{FCC_{max}}$ is selected (i.e., the video bandwidth (quality) is reduced), thus a “high” bandwidth $(R_{FCC_{max}} - R_{FCC_{min}})$ is reserved for retransmissions. When the receiver (e.g., a mobile station) detects only one of the $M$ received blocks in error, a negative acknowledgment (NAK) is sent to the transmitter (e.g., the base station). This approach simplifies implementation at the expense of efficiency. However, we have verified that the degradation is not important since the block error processes among MCCs are highly correlated [6]. Then, when the transmitter receives an NAK, the data rate of FCC is increased to $R_{FCC_{max}}$ and the retransmission (only one) of MCC blocks begins. Unlike classical GBN, not all the $NM$ blocks are sent again since the bandwidth assigned for retransmission is limited. Only a fraction of the information lost is retransmitted. For example, consider $N = 6$ and the transmission rates defined in Section IV (i.e., $R_{out} = R_{FCC_{max}} + R_{FCC_{min}} = 21.6$ kbps). Then, when an NAK is received, only 33% of the $NM = 12$ blocks can be retransmitted. When the retransmission process finishes, the system returns to the “normal” mode.

The ability of NSVBR to improve the reliability of H.263 wireless video transmission in MC-CDMA is analyzed in Fig. 2. We adopt $M$, $N$, $R_{out}$, $R_{FCC_{max}}$ and $R_{FCC_{min}}$ as given in the previous example. Moreover, we assume that no feedback errors occur. We consider the channel state “Very Good”. Results obtained from entire system simulation and theoretical values are presented. Residual $P_B$ can be estimated by using SBMM (this analysis is not included in this paper). We verify the excellent accuracy of both the distortion modeling approach (9) and the parameter estimation method based on SBMM. Moreover, note that NSVBR achieves significant gains for both test sequences with small additional complexity. These gains depend on the residual block error rate and the amount of motion present in the sequence. The biggest gain can be observed for Foreman [9].

C. Analysis of Error Resiliency Techniques

Residual transmission errors cannot be avoided with a mobile radio channel. Additional schemes to stop the error propagation have to be considered. In this sense, NSVBR can be efficiently combined with error resilience techniques to further improve the video quality in MC-CDMA transmission.
Here we evaluate video performance achieved by NSVBR combined with a theoretical stop error propagation technique we called ideal-error-tracking (I-ET). This approach is similar to that one presented in [9], that is, it utilizes intra-coded MBs (I-MBs) refresh to stop temporal error propagation. Using a feedback channel, the temporal and spatial occurrence of an error is reported to the transmitter. Thus, the location and extent of propagated errors is reconstructed at the encoder. However, unlike [9] we assume that in I-ET the video source coding distortion \( D_{SC}(t) \) is not affected by the I-MBs refresh, and the error signal at \( t = 0 \) is successfully canceled at \( t = T_d \). It can be verified that the video performance obtained from I-ET constitutes a theoretic upper bound of the video quality that can be achieved by any error resilience technique using the same conditions of operation (e.g., time delay \( T_d \)). The loss of the picture quality achieved by I-ET results in

\[
D_{CH-ET}(t) = \begin{cases} 
    P_b \Theta(t) & 0 \leq t \leq T_d \\
    P_b \Theta(T_d) & t > T_d
\end{cases},
\]

while the time averaged video distortion is given by

\[
\bar{D}_{CH-ET}(t) = P_b \bar{\Theta}(T_d).
\]

Table I presents results of the average distortion \( \bar{D}_a \) for NSVBR and I-ET. Channel state “Good” is used and the time delay between the video coder and decoder is assumed \( T_d = 0.8 \) s.

<table>
<thead>
<tr>
<th>Video Sequence</th>
<th>Average Video Distortion (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NP</td>
</tr>
<tr>
<td>M&amp;D</td>
<td>18.58</td>
</tr>
<tr>
<td>Foreman</td>
<td>23.86</td>
</tr>
</tbody>
</table>

Several cases are considered: no protection (NP), I-ET (only), and I-ET+NSVBR. The values are analytically derived from (12) with \( D_{SC} \) given by (15). \( \bar{\Theta} \) is calculated from (13) for NP, while \( \bar{\Theta} = \Theta(T_d) \) for I-ET and I-ET+NSVBR (see (17)). We also include the theoretical limit defined by (14), which represents the minimum distortion of the video sequence that can be achieved by NSVBR. Results show that significant gains can be achieved when NSVBR is used in combination with I-ET. We also verify that a residual average distortion around 0.75 dB is added to the theoretical minimum value achievable by NSVBR. This result suggests that new error resilience schemes could be designed to improve even more the performance of NSVBR in video transmission over error-prone channels.

V. CONCLUDING REMARKS

This paper has presented a novel semi-analytical methodology to estimate performance of wireless video transmission in MC-CDMA systems. Based on this approach, we analyzed H.263 video transmission in IS-95B at low bit rates over slow fading channels. Comparisons with values derived form entire system simulation have demonstrated the excellent accuracy of our technique. We found that extra protection must be provided in order to improve the video quality of the end user. We also introduced a new reduced complexity feedback error control scheme, which has been shown to significantly improve the reliability of video transmission. Furthermore, we investigated the effectiveness of error correction and error resilience techniques, and discussed theoretical limits of the video quality that can be achieved in transmission over error-prone channels. The methodology we proposed in this paper can be extended to performance evaluation of several video codecs in MC-CDMA transmission over generalized fading channels.

REFERENCES