MPEG 2 Video Services for Wireless ATM Networks

Jian Zhang, Student Member, IEEE, Michael R. Frater, Member, IEEE,
John F. Arnold, Senior Member, IEEE, and Terence M. Percival

Abstract—Audio-visual and other multimedia services are seen as an important source of traffic for future telecommunications networks, including wireless networks. In this paper, we examine the impact of the properties of a 50 Mb/s asynchronous transfer mode (ATM)-based wireless local-area network (WLAN) on moving picture experts group phase 2 (MPEG 2) compressed video traffic, with emphasis on the network’s error characteristics. The paper includes a description of the WLAN system used and its loss characteristics, a brief discussion of relevant aspects of the MPEG 2 standards and the associated error resilience techniques for minimizing the effect of transmission errors, and a description of the method by which the video data is organized for transmission on the network. We show results on the effect of cell loss due to transmission errors on the quality of the decoded video at the receiver, and demonstrate how error resilience techniques in both the systems and video layers of MPEG 2 can be used to improve the quality of service. Situations where up to 1% of the data is lost due to network transmission errors are examined. Most important among the findings are that error resilience experiments that do not take into account the effect of the MPEG 2 systems layer will tend to significantly overestimate the quality of received video, and that the error resilience techniques provided within the MPEG 2 standard are not sufficient to provide acceptable quality with acceptable overheads, but that this quality can be significantly increased by the addition of a small number of simple techniques.

I. INTRODUCTION

N recent years, there has been a trend toward providing mobility in telecommunications systems. The most obvious example of this has been the rapid growth of cellular telephone networks. One key disadvantage of these networks is that they provide only very limited capacity for each user, often less than 16 kb/s. This limitation will be addressed by a number of wireless local-area networks (WLAN’s) that are currently under development. However, there remains the significant issue of errors introduced by the wireless network, which occur very much more frequently than in fixed networks. These transmission errors can result in both packets of data being completely lost, as well as bit errors being introduced into transmitted data. It is this combination of bit errors and cell losses that makes the problems of providing error-resistant compression coding very different compared to conventional fixed networks.

In this paper, we examine the impact of the properties of a 50 Mb/s asynchronous transfer mode (ATM)-based WLAN on moving picture experts group phase 2 (MPEG 2) compressed video traffic. A brief description of the interface between ATM and video codecs is presented. A number methods for providing error resilience in compressed digital video bitstreams transmitted over cell-based telecommunication networks, such as ATM, are described. We show results on the effect of transmission errors, which can cause cell loss with probabilities as high as 1%, on the quality of the decoded video at the receiver, and demonstrate how error resilience techniques can be used to improve the quality of service. Most important among the findings is that the error resilience techniques provided within the MPEG 2 standard are not sufficient to provide acceptable quality, but that this quality can be significantly increased by the inclusion of a small number of simple additional techniques.

In principle, retransmission of errored cells in the WLAN can be used to overcome all transmission errors, except in a small number of cases where there is no significant signal at the receiver. A service that has a high level of resilience against the effect of transmission errors, however, will still be able to operate in those regions of a building where significant propagation errors occur, without a large amount of additional support from the network, thus reducing the complexity of the WLAN, and so making the network implementation less expensive. It is also likely that simpler algorithms for utilizing antenna diversity could be implemented where services are tolerant of transmission errors. Additionally, for real time conversational services, in which there are tight constraints on the permissible end-to-end delay, retransmission of lost or errored data is often not feasible.

The novel aspects of this paper are as follows.

1) Video performance in the presence of cell loss is assessed for loss streams based on network error/loss measurements.

2) In previous work, only the MPEG 2 video layer has been considered. In a real audio-visual application, the system layer is used to multiplex different streams, such as video and audio. In our work, the MPEG 2 systems layer has been implemented, and is shown to have a significant effect on the quality of the decoded video.

3) The cell loss probabilities that are examined in this paper range between 10⁻¹ and 10⁻². The upper end of this range is very much higher than is normally considered in video error resilience work, but is potentially very important in wireless applications.

The paper is organized as follows. In Section II, the ATM WLAN is described, followed by brief descriptions of relevant aspects of the MPEG 2 systems and video standards in

Manuscript received February 1, 1996; revised June 1996. This paper was supported by the CSIRO Division of Radiophysics.
J. Zhang, M. R. Frater, and J. F. Arnold are with the School of Electrical Engineering, University College, The University of New South Wales, Australian Defence Force Academy, Canberra, ACT 2600, Australia.
T. M. Percival is with the CSIRO Division of Radiophysics, Epping NSW 2121, Australia.

Publisher Item Identifier S 0733-8716(97)00059-0.

0733–8716/97$10.00 © 1997 IEEE
Section III and the interfacing between the video services and the ATM layer in Section IV. The details of the error resilience experiments are set out in Section V and the results and discussion in Section VI.

II. THE VERY HIGH PERFORMANCE WLAN

The Australian CSIRO Division of Radiophysics is developing an ATM-based very high performance WLAN (VHP-WLAN) which is designed to operate in the millimeter wave bands (nominally 60 GHz) and give a throughput of up to 155 Mb/s. The current prototype system operates at a link data rate of 50 Mb/s. The VHP-WLAN is designed specifically to facilitate the operation of mobile multimedia applications running over an ATM network. One important application is the transmission of real-time, digitally encoded audio-visual services.

A. The VHP WLAN Configuration

The WLAN is designed to have a hub which serves a number of users in a radio cell which covers a radius of about 15 m. A number of cells would be used to cover a floor of a building. Each hub is connected to a switch-port in a network of ATM switches, as shown in Fig. 1.

An in-building WLAN has to be designed to give reliable transmission in the hostile indoor environment. The major problem is caused by multipath which in general cannot be mitigated by the use of directional antennas, if a small portable system is required. The multipath signals can cause intersymbol interference (ISI) and/or fading. As the data rates being transmitted increase, the effects of ISI become harder to handle. This WLAN system uses a combination of multitone modulation and forward error correction (FEC) techniques to overcome the effects of ISI at high data rates. This technique is combined with antenna diversity and automatic repeat request (ARQ) to overcome the effects of broadband fading. (Detailed technical descriptions of the system may be found in [1]–[5].)

In a WLAN systems, ARQ is commonly used to reduce the errored packets by adding a cyclic redundancy check (CRC) at the end of each message. An acknowledgment is sent if the message is correctly transmitted, the message is retransmitted if no acknowledgment is received. This retransmission will only be useful if the characteristics of the radio channel have changed between the time of the original transmission and the retransmission. One way to alter the channel is to use a different antenna diversity setting for retransmissions. The other is to delay the retransmission for a significant period of time to allow the channel characteristics to change and hopefully allow for successful retransmission. For an in-building, 60 GHz channel, with objects moving at up to 1 m/s a delay of at least 5 ms, in addition to a doubling of the WLAN transmission time, would be required. Such an additional delay will be unacceptable for some real time services.

Antenna diversity can be achieved by changing the polarization or the beam pattern of the antenna, or by selecting between antennas which are physically separated by more than half a wavelength. This can either be done during the message preamble to select the best option or changed only after an unsuccessful transmission. The former method has the disadvantage of increasing the length of the preamble and hence reducing the bandwidth efficiency of the system.

Fig. 1. WLAN configuration.
As the VHP-WLAN is designed for ATM affinity, it will be possible to use different methods for different quality of service requirements. For real time services using MPEG 2 video transmission it is proposed to use antenna diversity but changing the selection only after an unsuccessful transmission with no retransmission. This will, however, result in errored or lost packets, but introduce no extra delay. Hence, the concealment techniques described below will be make an important contribution to the overall quality of service. For other services, a full ARQ system may be used, which results in almost no lost packets but an increase in both network delay and delay jitter.

B. System Performance

The above combination of techniques will give excellent performance for the radio system, but will not guarantee a 100% coverage of the indoor environment at very low bit error rates (BER’s). Measurements made using a VHP-WLAN testbed have shown that there are many areas where very low BER occur, some areas where moderate error rates occur and some areas where no reliable transmission is possible.

In order to determine the nature of the errors which might be expected, a series of measurements were conducted using a WLAN testbed. This device consists of a millimeter wave transmitter which generates the required waveform and a receiver and base-band unit which samples the received signal in real time and stores a total of up to 128k 8-b complex samples for later analysis. This analysis is performed by a software implementation of the multitone receiver running on a PC. For these tests the testbed was configured to operate at a throughput of 50 Mb/s after error correction. This allows continuous measurements to be taken over a period of almost 20 ms.

In a WLAN system, however, the transmission medium is shared between a number of users by time multiplexing the transmission between different users. For example, a mobile user transmitting an MPEG video stream at 4 Mb/s would only need access to the 50 Mb/s channel for around one tenth of the time. Multiplexing is achieved by transmitting data in packets. Each wireless packet consists of a wireless header plus four ATM cells (five bytes of header plus 48 bytes of payload each), giving a total length of 2120 b. One of these packets (of duration 42.4 ms) would need to be transmitted every 376 ms for a 4-Mb/s data stream.

A series of measurements was also conducted with the testbed capturing samples corresponding to packets transmitted at this rate. This gives a realistic estimate of the performance variation over time of the WLAN when transmitting video.

C. Error Characteristics

Errors in the received data occur frequently due to a number of causes, including noise and frequency-selective and broadband fades. FEC is applied to attempt to eliminate these errors. However, when a large number of errors occur in a wireless packet, not all can be corrected. Where a cell is received with uncorrectable errors, the receiver has the choice of delivering the errored data to the video decoder or deleting the cell. Our experience is that where a cell is known to contain errors, attempts to decode it tend to produce significantly more objectionable artifacts than simply applying concealment to the affected region.

The mean error rate in a given period of time depends on the propagation conditions that were experienced by the signal during this time. Hence, error rates measured at different times vary widely, and the error process is seen to be bursty. As was stated above, in good propagation conditions, there are almost no errors; in very bad propagation conditions, the receiver cannot synchronize to the incoming packets, and no data transmission is achieved. This suggests that, in order to assess the impact of packet loss on video quality, a range of packet loss probabilities should be examined.

In order to determine the type of temporal dependence required in the error model, autocorrelation functions for the number of errors occurring in a wireless packet were calculated. A sample autocorrelation function is shown in Fig. 2. It is clear that the autocorrelation decays approximately exponentially as the lag is increased. This is because the vertical axis is logarithmic, the horizontal axis is linear, and the curve is approximately a straight line. The deviation from a straight line for larger lags is due to the fact that only a small number of samples are available for estimating the autocorrelation function at these points. The fact that the autocorrelation does not instantly drop to zero for all nonzero lags indicates that errors in successive wireless packets do not occur independently. It has been found in many types of network traffic modeling that the traffic displays properties of long range dependence and self-similarity [6], requiring special classes of models. The exponential decay (as opposed to a polynomial-type behavior) of the autocorrelation function indicates that such phenomena do not occur here.

As outlined in the previous subsection, the length of measurements possible in the VHP-WLAN testbed (approximately 50 WLAN packets, or 20 ms) is much shorter than would be required for conducting video error resilience experiments (at least 5 s). Also, because the mean error rate varies widely with
time, it is hard to say whether a given set of measurements is typical of network performance. Hence, it is necessary to use models for the channel error characteristics.

From the analysis of the WLAN error data, we conclude that a simple Markovian model is an appropriate model for the cell/packet-loss process. The Gilbert model [7] is a two state Markovian model. In one state, the error probability (i.e., packet loss probability) is zero. In the other state, the error probability takes some predetermined value (in this case, 1). The transition rates between the states controls the lengths of the bursts of errors. There are two parameters to be controlled: the average error rate and the average length of a burst of errors. Reasonable ranges for these parameters are: packet loss probability between $10^{-4}$ and $10^{-2}$, and the average length of an error burst between 1 (i.e., independent errors) and 10.\footnote{While longer bursts of errors do occur, it is usually where the received signal is so weak that essentially no data is successfully received.}

This model could be applied to either the cell loss process, or the packet loss process. In practice, the loss process of the WLAN is found to be some combination of the two.

### III. Overview MPEG 2 Standard

#### A. Video Compression

For a standard size television picture (704 $\times$ 576 pixels) and frame rate (25 Hz), MPEG 2 is designed to provide distribution quality television at a bit rate between 4–9 Mb/s. Each frame in a video sequence is coded in one of three modes as follows [8].

1) **Intra (I) Frame:** Image data is transmitted using variable-length codes representing the coefficients derived by breaking the picture into $8 \times 8$ blocks, applying the discrete cosine transform (DCT) to these blocks, and quantizing the DCT coefficients to a user-defined level.

2) **Predictive (P) Frame:** The motion compensated prediction from the previous I or P frame is used, with the residual difference data being coded in much the same way as for an I frame.

3) **Bidirectional (B) Frame:** The motion compensated prediction is achieved by one of forward prediction from the previous I or P frame, backward prediction from the next (later) I or P frame or an interpolation between these two I or P frames. The prediction mode used can change for different parts of the picture.

The structure of the interframe prediction associated with each frame type is shown in Fig. 3. As indicated above, coding within a frame is based on blocks, each of which consists of 64 luminance or chrominance pixels in an $8 \times 8$ square. Luminance blocks are combined in groups of four, which, when combined with the associated chrominance information for this region of the picture, form macroblocks (MB’s), which are of size $16 \times 16$ pixels. Adjacent MB’s are grouped into a slice. A frame is composed of a number of slices. The bitstream syntax of coded data after variable length coding is shown in Fig. 4. A frame consists of a number of slices proceeded by a frame header. Similarly, a slice consists of a number of macroblocks proceeded by a slice header. Each macroblock also begins with a header, which includes information on the macroblock location (“MB address”), and motion vectors for use in the motion compensated prediction. In the first macroblock of each slice, the MB address and motion vector are coded absolutely. In each remaining macroblock in a slice, these parameters are coded differentially with respect to the corresponding values in the macroblock immediately before it. More detailed information on the MPEG 2 video compression standard can be found in [8].

#### B. MPEG 2 Systems Layer

For many audio-visual applications, it is necessary to transmit simultaneously streams of both audio and video data on a single channel. For example, in videoconferencing, there would usually be at least one video and one audio channel; in the distribution of entertainment-quality television, there would often be two or more audio channels. The MPEG 2 systems layer [9] allows multiple streams of audio and video data to be combined to produce a single output stream. The MPEG 2 systems layer can take one of two forms, known as program stream (PS) and transport stream (TS). The PS is very similar to the MPEG 1 systems layer. The TS has been used in this work.

The MPEG 2 Systems layer provides the following functionalities [9].

1) **Packet-Oriented Multiplex:** provides the ability to multiplex several input streams (e.g., audio, video or private data) onto a single output channel.
2) Error Resilience: provides features that assist a decoder receiving data that has had errors introduced by the transmission or storage medium.

3) Synchronization: is achieved by providing time stamps, and can be used for both buffer management in the decoder and synchronization between streams.

Packetization is carried out in two steps:

1) Each source bitstream to be transmitted is broken up into packets, known as packetized elementary stream (PES) packets. These packets are of variable length. For video, there would usually be one packet per slice. The packet header includes an optional CRC to allow for error detection.

2) PES packets are broken into TS packets, each of length 188 bytes. These TS packets are then time-multiplexed onto the output channel. TS packet headers include information that allows the decoder to channel the received data to the correct decoder (e.g., audio or video).

This hierarchy is illustrated in Fig. 8.

Error resilience is greatly aided by the fact that the TS packets are of fixed length. This means that the systems layer decoder’s ability to correctly decode one TS packet is not in any way affected by errors in the previous packet.

It can also be seen, however, that any error in the TS packet header that corrupts the source stream identification will result in the loss of a whole transport packet, even though the remainder of the data in that packet may be completely correct. As will be described below, almost twice as much video data is lost in some of the experiments reported here when the effect of the systems layer is taken into account, compared to looking at transmitting only the video layer.

C. Methods for Error Resilience

To achieve error resilience in video transmitted over wireless computer networks, the methods used in our experiments are divided into three categories, each of which is described in the following subsections. All methods used in the experiments that follow are compatible with the MPEG 2 standard, unless otherwise explicitly stated.

1) Concealment: These techniques attempt to conceal an error once it occurs by taking account of the remaining spatial and temporal correlation in the video sequence [8].

- a) In areas of the picture that do not change very much with time, it is effective to conceal the effect of cell loss by temporal replacement, i.e., using the corresponding information from the previous frame. This approach is not effective in high motion areas. In this situation, spatial interpolation tends to be more effective. Neither of these techniques are applied in our experiments.

- b) Motion compensated concealment, combining temporal replacement and motion estimation, can improve the concealment. In this technique, the motion vectors above and/or below the lost macroblock are used to predict the motion vector on the lost MB, and a motion-compensated concealment strategy is used. This improves the concealment in the moving areas of the screen, but is unable to conceal errors for a lost MB which is surrounded by intra coded macroblocks. This is because the intra-coding process does not supply motion vectors. To avoid this, the encoding process can be extended to include motion vectors for intra-coded MB’s. Of course, motion vectors for intra-coded MB’s are only used for error concealment. In addition, the motion vector and coded information for a particular intra-coded MB must be transmitted separately so that the motion vector is still available in the event that the coded data is lost [8]. This technique is applied in all the results described below.

2) Temporal Localization: One significant effect of cell loss and errored bitstreams on the decoded sequence is error propagation. By considering the processed frame order of the MPEG 2 coding scheme [8], it can be seen that errors occurring in I-frames can propagate into the following P or B frames, since these P and B-frames are predicated from the I-frame. Similarly, errors in P-frames can propagate into subsequent B and P frames. Temporal localization seeks to minimize error propagation from picture to picture in the temporal domain by providing early resynchronization of pictures that are coded differentially. Techniques include:

- a) Cyclic Intra-Coded Pictures: Extra intra-coded I-frames can be sent to replace B or P-frames as shown in Fig. 3. Thus error propagation is stopped at the expense of extra transmitted bits. This will reduce the coding efficiency compared with normal coding, and also increase the possibility of cell loss and erroneous bits within the same frame due to the sudden increase in bit rate.

- b) Cyclic Intra-Coded Slices: Instead of using intra-coded frames to limit the effect of errors in the image, extra intra-coded slices can be used periodically to refresh the frame from the top to the bottom over a number of P-frames. The disadvantage is that the partially updating of the screen in a frame period will produce the subjectively unpleasant “windscreen wiper” effect.

In the results described below, every 12th frame is an I frame. No other intra-coded slices or frames have been included. The reason is that it was found that, while extra I frames or slices could be used to reduce the effect of transmission losses, the overall video quality was degraded for a fixed video data rate.

3) Spatial Localization: Spatial localization encompasses those methods aimed at minimizing the extent to which errors propagate within a picture by providing early resynchronization of the elements in the bitstream that are coded differentially between MB’s. This resynchronization involves two features. The first is an unambiguous indication of the location in the bitstream of macroblocks where resynchronization is possible. The second is to code absolutely those quantities that are normally coded differentially with respect to the previous macroblock (e.g., motion vectors) for this resynchronizing macroblock. There are three methods applied in our work.

- a) The most basic method for achieving spatial localization of errors is to reduce the (fixed) number of MB’s in a
slice. Suppose that a slice of coded video data is divided into two small slices as shown in Fig. 5. If the first ATM cell is lost or received in error, the decoding procedure can resume at the slice start in the second ATM cell. This protects the second half of video slice data from being discarded. On the other hand, if both slice headers are packed in one cell, for example in the first ATM cell, once cell loss occurs in this cell, the simple slice method cannot resynchronize until the end of the entire video slice of data. Obviously, the coding efficiency will be lower than normal coding method due to the extra overheads in slice headers.

b) The small slice scheme does not take account of packing of bits into MPEG 2 TS packets, or of the packing of these packets into ATM cells. To address this point, an idea of improvement over the fixed small slice method can be developed [10], [11] as shown in Fig. 6. The first MB coded in every ATM cell will be coded absolutely by the encoder. That is to say, once a cell loss or erroneous bits occur within a cell, the decoding procedure can resume at the first macroblock in the next ATM cell. Where consecutive ATM cells are lost, the amount of information that becomes unavailable to the decoder will always be less than one ATM cell. This technique is known as the MB-resynchronization method. It is not compatible with the MPEG 2 standard, but could be implemented with quite low complexity and overheads. Typically, the overhead imposed would be two bytes per cell.

A slight improvement in efficiency can be obtained by observing that there is no need to insert this information in the second cell of a TS packet. This is because the MB-resynchronization information allows resynchronization in the bitstream when it has been lost due to errors in the previous cell. However, in the case of this second cell, loss of the previous cell means that the TS packet header has been lost, and hence that the whole of the remainder of the packet is unusable because the decoder does not know to which elementary stream (e.g., audio or video) it belongs.

This technique is not restricted to implementation at cell boundaries. In systems where boundaries of regions of cell loss are guaranteed to be aligned with, for example, TS packets, the resynchronization information would be required less often in the bitstream.

c) The use of adaptive slice sizes is similar to the MB-resynchronization method, but change the slice size based on the length of ATM cells as shown in Fig. 7. The encoder will trace the coding process to place the slice start code at the first opportunity in each ATM cell. This technique can achieve essentially the same results as MB-resynchronization, but at the cost of considerably higher overheads caused by the larger size of the additional slice headers that must be transmitted. The overhead of implementing this technique would be between five and eight bytes per cell, depending on the coding mode used.

IV. MPEG 2 VIDEO IN AN ATM NETWORK

In the wireless network studied here, all data is transmitted in cells with the same format as a standard ATM network. Data is first packetized in the ATM adaptation layer (AAL), and then packed into ATM cells. Fig. 8 illustrates a layered procedure in which data from a user’s application (such as a video slice) is packed into ATM cells. This is based on AAL I [12]. The first layer is specified by the user’s application and might be a coded video slice. The second layer is the AAL, which is divided into two sublayers called the convergence sublayer (CS) and the segmentation and reassembly layer (SAR). In the CS layer, data from the video slice is taken and CS-PDU is generated. In the SAR sublayer, a CS-PDU is segmented into a number of SAR-PDU. One SAR-PDU consists of 48 bytes including the SAR header and trailer. The third layer is the ATM stream layer. A SAR-PDU will have added the 5 byte ATM header information thus generating an ATM cell for transmission. The fourth layer is the physical layer which transmits the ATM cells over the wireless computer network.

In the results described below, it has been assumed that there is an exact alignment between wireless packets and MPEG 2 TS packets. In a practical system, this could be achieved by placing an interworking adaptor at the boundary of the wireless and fixed ATM networks. This interworking adaptor could be of relatively low complexity, and need not impose
significant additional transmission delay. The implications of not enforcing this alignment are discussed below.

V. ERROR RESILIENCE EXPERIMENTS

Experiments have been carried out in which the above-listed error resilience techniques were applied in the video coding process, and the resulting bitstream was corrupted as would occur during transmission across a WLAN. Two well-known video test sequences were used: “Bus” (length 150 frames) and “Flower Garden” (length 125 frames). Each video sequence was coded using MPEG 2 compatible software, with an output rate in all cases of 4 Mb/s. In all the results reported below, the full range of MPEG 2-compatible error resilience techniques described above are applied, with one slice per row of macroblocks (i.e., 44 macroblocks per slice). Two sets of results are shown. The first is simply the results from the MPEG 2-compatible bitstream. The second is with the addition of MB-resynchronization included at the beginning of each cell. Each experiment was repeated with 25 independently generated loss sequences. All comparisons are between the decoded sequence (with errors) and original sequence. Note that while the introduction of macroblock resynchronization increases the overhead, the reduction in PSNR is very small.

Based on the error characteristics obtained from measurements made on the WLAN testbed, three sets of results are presented:
1) cell losses occur independently (Table I);
2) wireless packet losses occur independently (Table II);
3) wireless packet losses occur in bursts with an average length of five wireless packets (Table III).

The reason for the choice of these three sets of test conditions is that losses in the wireless network consist of a combination of all three types. There are situations where single cells within wireless packets are discarded, and also those where whole wireless packets and even bursts of wireless packets are lost.
TABLE II
INDEPENDENT WIRELESS PACKET LOSSES: COMPARISON OF PSNR VALUES (IN dB)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Loss Prob</th>
<th>Single Slice Header per Row</th>
<th>Full MB-Resynchronization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>Std. Dev</td>
<td>Min</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Std. Dev</td>
</tr>
<tr>
<td>Flower Garden</td>
<td>$10^{-4}$</td>
<td>25.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>25.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>25.6</td>
<td>25.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.5</td>
<td>25.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>24.2</td>
<td>24.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.0</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td>Bus</td>
<td>$10^{-4}$</td>
<td>27.7</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.8</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.7</td>
<td>27.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>27.5</td>
<td>27.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.4</td>
<td>27.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>26.3</td>
<td>26.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.1</td>
<td>26.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>0.09</td>
</tr>
</tbody>
</table>

TABLE III
BURSTY WIRELESS PACKET LOSSES: COMPARISON OF PSNR VALUES (IN dB)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Loss Prob</th>
<th>Single Slice Header per Row</th>
<th>Full MB-Resynchronization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>Std. Dev</td>
<td>Min</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Std. Dev</td>
</tr>
<tr>
<td>Flower Garden</td>
<td>$10^{-4}$</td>
<td>25.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>25.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>25.6</td>
<td>25.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.6</td>
<td>25.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>24.9</td>
<td>24.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.1</td>
<td>24.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>Bus</td>
<td>$10^{-4}$</td>
<td>27.8</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.8</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.8</td>
<td>27.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>27.7</td>
<td>27.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.6</td>
<td>27.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>26.9</td>
<td>26.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.2</td>
<td>27.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.27</td>
<td>0.11</td>
</tr>
</tbody>
</table>

VI. RESULTS

It can be seen from the tables of results that, for all cases where the cell loss process is independent and the loss probability is $10^{-3}$ or greater, the use of MB-resynchronization results in a significant performance improvement, as measured by the PSNR. It is also true that the greatest benefit is gained in situations where the losses are independent. While it is true that the cell and packet loss processes will tend to be bursty in practical systems, the loss probability during bursts of loss will often be significantly less than one, perhaps between 10% and 50%. Under these conditions, the use of MB-resynchronization will allow some of the bitstream to be decoded, whereas without the use of this technique, essentially none of the bitstream would be decoded. The use of MB-resynchronization also results in a significant decrease in the variability of the quality of the decoded video, as measured by the standard deviation of the PSNR values.

The improvement in quality achieved by implementing MB-resynchronization becomes even more clear when the sequences are viewed. With MB-resynchronization, the video quality at $10^{-3}$ is perhaps roughly equivalent in annoyance to a moderate level of ghosting in terrestrial broadcast television, or replay from a VCR with worn heads or a worn tape. Without MB-resynchronization, the quality at $10^{-3}$ is extremely poor for all except the cases of bursty wireless packet losses. In general, viewing the decoded sequences suggests that MB-resynchronization will allow acceptable operation with cell loss probabilities up to $10 \times$ higher than the case of a single slice header per row of macroblocks. In cases where the loss probability was $10^{-4}$, there is very little degradation in PSNR caused by the loss. However, the effect of loss is still very clear in the video, and there is still a significant gain in quality achieved by implementation of MB-resynchronization. The reason for this is that the size of the regions in which concealment must be performed is greatly reduced.

In principle, the MPEG 2 Adaptive Slice size technique can be used to achieve the same results as the MB-resynchronization technique studied here, although the use of adaptive slice sizes can never result in better performance than MB-resynchronization. However, the performance is equivalent only for a certain narrow range of operating conditions. Roughly speaking, these conditions are those under which it is not necessary to have a slice header in every cell, but perhaps only every fourth or tenth cell. This can be seen as follows. Each ATM cell has a payload size of 48 bytes. Of these, one byte is taken by the AAL, and one more byte (on average) is used for the TS packet header. Without MB-resynchronization, the quality at $10^{-3}$ is extremely poor for all except the cases of bursty wireless packet losses. In general, viewing the decoded sequences suggests that MB-resynchronization will allow acceptable operation with cell loss probabilities up to $10 \times$ higher than the case of a single slice header per row of macroblocks. In cases where the loss probability was $10^{-4}$, there is very little degradation in PSNR caused by the loss. However, the effect of loss is still very clear in the video, and there is still a significant gain in quality achieved by implementation of MB-resynchronization. The reason for this is that the size of the regions in which concealment must be performed is greatly reduced.

In principle, the MPEG 2 Adaptive Slice size technique can be used to achieve the same results as the MB-resynchronization technique studied here, although the use of adaptive slice sizes can never result in better performance than MB-resynchronization. However, the performance is equivalent only for a certain narrow range of operating conditions. Roughly speaking, these conditions are those under which it is not necessary to have a slice header in every cell, but perhaps only every fourth or tenth cell. This can be seen as follows. Each ATM cell has a payload size of 48 bytes. Of these, one byte is taken by the AAL, and one more byte (on average) is used for the TS packet header. A slice header will require between five and eight bytes, leaving between 38 and 2In normal operating conditions, the PES packet headers require (on average) much less than one byte per cell.
41 bytes for video data. The MB-resynchronization header requires only two additional bytes, leaving 44 bytes. In other words, using adaptive slice sizes reduces the available data rate for video by approximately 10%. The reduction in video quality at 4 Mbs/s is quite noticeable.

It is clearly pertinent to ask, under which conditions the performance of adaptive slice sizes and MB-resynchronization are similar, i.e., when is it possible to reduce the number of slice headers such that they do not use up to much of the available network capacity, but the reduction in error-resilience is not significant. Roughly speaking, this will occur only where the loss process is very bursty, and there is some alignment between bursts of loss and groups of cells, for example those forming TS packets. Although we have assumed such an alignment here, there is a small cost in increased complexity associated with its enforcement.

Although not reported here, some experiments have been carried out in which the number of slices per row of macroblocks is increased, in a manner compatible with the MPEG 2 standard. While some improvement was observed with four slices per row of macroblocks, the performance did not come close to that obtained using MB-resynchronization. It is only when the number of slices approaches one per cell that the error-performance is comparable to MB-resynchronization. In this case, however, the video quality had been significantly degraded because a very high proportion of the channel capacity was being used for carrying the resynchronization information. For similar reasons, increasing the number of I frames does not result in a net performance increase.

As noted above, transmission errors do not necessarily cause the WLAN to discard complete wireless packets, but can often cause only individual cells to be lost. Hence, the actual losses experienced in a real network will be some combination of the cell loss and wireless-packet loss cases simulated here. Further, if alignment cannot be guaranteed between TS packets and WLAN packets, which must be the case in any network where the size of wireless packets is not an integer multiple of four cells, the loss characteristics will be much more like independent cell losses.

If the experiments described here had not taken into account the effect of the MPEG 2 Systems layer, those relating to independent cell losses would have underestimated the amount of video data lost in the transmission process by approximately a factor of two. This would have caused the PSNR to be increased by up to 2 dB, and the subjective video quality to be markedly greater than is realistic. The systems layer contributes a large portion of the difference in performance between the cases of independent cell losses and wireless packet losses. (Where MB-resynchronization is applied, the systems layer contributes almost the whole of this difference.)

Finally, we note that the complexity of implementing the macroblock resynchronization techniques described here in a practical system is low and that this could be achieved by the use a simple interworking adaptor at the boundary between the fixed and wireless networks. This is in contrast to many other available techniques, such as MUVLC [13], which impose higher computational and/or memory requirements on the decoder.

In summary, it has been found that the error resilience techniques defined within the MPEG 2 standard improved the quality of the decoded video significantly, but that the quality could be improved still further where the cell loss probability is high (say greater than 10^{-5}) by the addition of the MB-resynchronization technique. This finding is important for conversational services, such as video conferencing, in which low transmission delay is required, and hence the use of an automatic retransmission request protocol would not be appropriate.

VII. CONCLUSION

This paper reviews several error resilient video coding methods for cell-based transmission of MPEG 2 coded video. The significance of this paper is that it addresses the problem of transmitting compressed video data over wireless computer networks, which have different error characteristics from other networks, often exhibiting frequent cell loss. The results from our experiments show that with simple additions to the error-resilience techniques defined in the MPEG 2 standard, it is possible to significantly improve the quality of video decoded at the receiver, and hence increase the range of radio propagation conditions under which acceptable quality of service can be obtained.

ACKNOWLEDGMENT

The authors would like to thank J. Deane and J. Graham for their help in taking WLAN channel measurements and M. Batty and G. Daniels for many useful discussions.

REFERENCES


Jian Zhang (S’95) received the B.Sc. degree in electronic engineering from East China Normal University, China, in 1982, the Postgraduate Diploma in computer systems engineering from the Shanghai Institute of Mechanical Engineering, China, in 1989, and the M.Sc. degree in computer science from Flinders University of South Australia, Australia, in 1994.

Since 1994, he has been with the School of Electrical Engineering, University College, the University of New South Wales, Australian Defence Force Academy, Australia. He is currently working toward the Ph.D. degree in video communication.

Michael R. Frater (S’89–M’92) was born in Sydney, Australia, in 1965. He received the B.Sc. degree in mathematics and physics and the B.E. (Hons.) degree in electrical engineering from the University of Sydney, Australia, in 1986 and 1988, respectively. He received the Ph.D. degree from the Department of Systems Engineering at the Australian National University, Australia, in 1991.

He is a Senior Lecturer in Electrical Engineering at the University College, the University of New South Wales, Australian Defence Force Academy, Australia. His research interests include aspects of the coding and transmission of video services, teletraffic and queueing theory, stochastic processes, signal processing, and control.

John F. Arnold (S’77–M’85–SM’96) received the B.E. and M.Eng.Sc. degrees from the University of Melbourne, Australia in 1976 and 1979, respectively, and the Ph.D. degree from the University of New South Wales, Australia, in 1984.

Since 1978, he has been with the Department of Electrical Engineering, University of New South Wales, initially at the Royal Military College, Dunkeld and more recently at the Australian Defence Force Academy, Canberra. His research interests lie in the fields of video coding, ATM networks, and coding of remotely sensed data.

Terence M. Percival received the B.E. (Hons) and Ph.D. degrees from the University of Sydney, Australia, in 1976 and 1985, respectively.

His initial research interest was in microwave systems for radio telescopes. He worked on the construction of the University’s Fleurs radio telescope, the Very Large Array telescope in New Mexico and the Australia Telescope. In 1986, he joined the staff of the CSIRO National Measurement Laboratory where he worked on time and frequency standards and the use of satellites for time transfer. He also worked on high Tc superconducting microwave circuits. He then spent four years working with the Transmission Technology group at the OTC R&D Laboratory (now part of Telstra). During this time he worked on submarine optical fiber communications systems, the development of new satellite communications systems aimed at thin-route and VSAT applications, and the design of high-speed integrated circuits for switching and satellite communications. In 1991, he joined the CSIRO Division of Radiophysics to work on the development of high-speed wireless LAN’s. In 1995, he was appointed to the position of Director of CSIRO’s Telecommunications Engineering Initiative.