Engineering wireless broadband access to IPTV

Laith Al-Jobouri, Martin Fleury*, and Mohammed Ghanbari

University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

*Corresponding author: Email fleum@essex.ac.uk Tel. +44 (0)1245 381721

Abstract

IPTV is now extending to wireless broadband access. If video streaming across broadband links is to present competitive quality the video stream itself must be carefully engineered. This paper presents an example of how this can be done for H.264/AVC codec streaming across a WiMAX link. Packetization is an effective tool to govern error rates and in the paper source-coded data-partitioning serves to allocate smaller packets to more important data. A packetization strategy is insufficient in itself, as temporal error propagation should also be addressed by appropriate insertion of intra-coded data. For robust delivery it may also be necessary to include redundant packets, especially when channel conditions worsen. The whole should be protected by application-layer rateless coding. The paper concentrates on the configuration of the various elements of the protection scheme and their relative impact, as correct configuration is shown to be an important aspect of such schemes.

Keywords: H.264/AVC; IPTV; video codec; video streaming; WiMAX

1. Introduction

Due to the flexibility of network delivery, Internet Protocol TeleVision (IPTV) is attractive as an alternative to digital TV over terrestrial channels, though it may suffer: from delays due to congestion; and packet losses leading to fluctuations in video quality. It does represent an intrusion into personal privacy, as individual viewing habits can be tracked, although advertisers may benefit from this facility. Whereas terrestrial TV is limited to a fixed channel broadcast schedule, linear TV, IPTV can offer pause-TV (when a partially-viewed program is cached for later viewing), catch-up TV [1] and time-shifted TV (when a program is re-aired in a different time zone), as well as varieties of video-on-demand (VoD). It also opens up the possibility of interactive TV and hybrid TV (a combination of terrestrial broadcasting with network delivery). There are two basic forms of IPTV: 1) That delivered over closed or proprietary managed networks, including cable TV, often to set-top boxes
(STBs) that perform decoding and possibly decryption; and 2) That delivered over the unmanaged public Internet, often displayed directly on desktop computer screens. The latter may be called Web or Internet TV [2] or latterly Over-the-Top (OTT) TV, as a way of distinguishing it from the service that the telecommunications companies hoped to challenge terrestrial broadcasters with. Increasingly the industry trend, apart from cable TV, is towards the latter model of IPTV, i.e. OTT TV. There is no place for STBs when mobile IPTV [3] is delivered, hence, our interest in OTT TV. The two forms of IPTV may differ in: the picture quality of the stream offered (lower for unmanaged); the amount of content available (higher for unmanaged); the formats that the video is compressed to (the need to update the STB is an issue); the way content is secured (unmanaged delivery may not be encrypted or employ selective encryption [4] instead); and the frame resolutions presented (lower resolutions for unmanaged delivery). A flexible way to deliver IPTV [5] is to set-up (and tear-down) connections using the Real-Time Streaming Protocol (RTSP) (over reliable TCP transport), with subsequent data delivery using the Real-time Transport Protocol (RTP) over UDP transport. The Real-time Control protocol (RTCP) can also allow end-to-end feedback to the IPTV server to control the bitrate. RTSP also can support so-called trick mode functionality such as rewind and pause.

A crucial aspect of engineering IPTV is configuring the video codec in order to achieve good performance over a wireless link for mobile access. In this paper, H.264/Advanced Video Codec (AVC) [6] standard codecs are assumed, as though the High Efficiency Video Codec (HEVC) was standardised at the beginning of 2013, experience shows that it will take many years to be deployed, if indeed it is extensively used for streaming over wireless, given the absence of error resilience and concealment features [7]. The absence of error protection features may arise either because HEVC is intended to use Dynamic Adaptive Streaming over HTTP (DASH), a multi-streaming system for reliable TCP [8] or because HEVC’s considerable coding gain is achieved by abandoning macroblocks (MBs). HEVC is suitable for 720p High Definition (HD) network delivery, where the up to 50% lower bitrates over H.264/AVC compensate for the high frame rates, at least 50 frame/s at that resolution. Indeed, HEVC delivery chains have made quicker progress [9] than expected, though evaluation has
shown that, without persistent HTTP connections, streaming is subject to delays and interruptions when using DASH [10].

The current authors have introduced a robust scheme for streaming video over wireless broadband access, which has been simulated for IEEE 802.16 (WiMAX) systems [11]. WiMAX may have recently lost the technological competition with Long Term Evolution (LTE) [12] in developed Western countries, it is still being deployed in rural areas [13] and in countries where 3G cellular phone coverage is poor, including Africa and the Middle East. WiMAX is also attractive [14] for backhauling from local IEEE 802.11 networks. Though the authors’ scheme as a whole has been analysed in previous works such as in [15][16], the individual codec settings have not been analysed in isolation, whereas it has been suggested to the lead author that a very helpful service to the system developer community would be to analyze codec settings independently of each other. The authors’ scheme combines data-partitioning as a form of error resilience [17] with the addition of forward error correction (FEC) using rateless channel coding [18] at the application layer, together with retransmission of extra redundant data when required. Retransmission is limited to one round to avoid accumulating latencies. The current authors have introduced various error resilience measures [19]: different rates and types of Intra-refresh Macroblocks (MBs) [20], Constrained Intra Prediction (CIP) [21] and where necessary, redundant data-partitioned Network Abstraction Layer (NAL) units (codec level packets) [22]. The impact of each of these measures is analyzed in isolation from each other, for Constant BitRate (CBR) as well as Variable BitRate (VBR) streaming. CBR allows storage and bandwidth capacity to be planned in advance, at a cost in video quality fluctuations. VBR enables greater compression efficiency relative to CBR, which is why it is generally used for disc storage. VBR encoders can benefit from two or even three-pass encoders, which are unsuitable for live video compression. The relative merits of CBR and VBR are further discussed in [5] and [23]. In general, we have noticed that many researchers on similar topics have not explored the effect of video settings on their schemes and the authors believe that it is important to investigate this aspect of any protection scheme in order to understand the overall outcome of the scheme. In particular, it is important to determine how critical the video
Codec settings are to the success of a protection scheme. Is the scheme robust to changes or should certain settings be set within a given range?

To allow unmanaged delivery, OTT TV, to compete with Standard TV (SDTV) Quality-of-Experience (QoE), TV and video material can be locally cached [24], reducing latency to the access link, in this paper represented by a WiMAX base to subscriber station link. Whether streaming over managed or unmanaged networks, packets will be lost at the access network. This is the case, whether over broadband wireless or xDSL (Digital Subscriber Line), as the also latter suffers from burst errors [25]. Prior to that, whether managed or unmanaged networks are used, video streams are aggregated over high-capacity, optical networks, such as in Swisscom’s Bluewin IPTV service [26]. As only Passive Optical Networks (PONs) can reduce the error rates at the access link but PONs are not widely deployed. The approach adopted in this paper is potentially applicable across a range of access links, as it provides protection against error bursts and causes only moderate delays. If a multicast service is required, it can also be readily adapted from its unicast form by dropping acknowledgments to the base station and increasing the level of FEC. This possibility was tested in [27] and, therefore, this paper confines itself to unicast delivery.

The remainder of this paper is organized as follows. The following Section selects research from the literature that reveals the motivation behind the codec-based approach to video streaming used in this paper. Particular attention is paid to video streaming over WiMAX. Section 3 then outlines the approach taken to protection of IPTV streams. As the main focus of this paper is the configuration of the scheme rather than the scheme itself, the description is necessarily brief. Section 4 continues by detailing how the video codec settings were modelled in order to provide the evaluation that appears in Section 5. Section 6 summarizes the findings and rounds off with some observations about video streaming in this environment.

2. Related research

A number of research papers have considered some aspects of the video streamlining method used by us. In [28] a packetization method was presented for robust H.264 video transmission over an IEEE 802.11 wireless local area network (WLAN) configured as home network. Video robustness is enhanced by using small NAL
units and by retrieving possible error-free IP packets from the received MAC frame. An aggregation scheme with a recovery mechanism is deployed and evaluated via simulation. For fixed physical layer resources, the system provides a 2.5-dB gain in video quality (PSNR) compared to making no NAL packetization adjustments for similar throughput efficiency. Equally an 80% improvement in throughput is achieved for a similar video quality. However, data partitioning as a way of varying NAL sizes was not used.

Work in [29] used Forward Error Correction (FEC) and Automatic Repeat re-Quest (ARQ) to support streaming over WiMAX, exploiting features of the WiMAX standard. In particular, channel state information held at WiMAX stations served to dynamically construct the MAC packet data unit. The size of these units was thus determined such that the packet dropping probability was minimized without compromising goodput. Simulation results showed that the ARQ-enabled adaptive algorithm was always better than the non-adaptive algorithm.

In [17] the researchers compared non-scalable video coding with data partitioning using H.264/AVC under similar application and channel constructions for conversational applications over mobile channels. The experimental results showed that by using the data-partitioning scheme the number of entirely lost frames can be lowered and the probability of poor quality decoded video can be reduced. In the data-partitioning scheme of [17], differential protection was achieved by selecting from a set of discrete channel coding rates, through punctured convolutional codes. However, in order to determine the protection level, an optimization procedure was necessary to minimize potential distortion. This procedure depended on the quantization parameter (QP) and the coding rate for each partition. The wireless channel characteristics also had to be known in advance by the encoder. However, leaving aside the computational complexity of the optimization search in [17], there is another key difference between the scheme of [17] and this paper. In [17] no feedback occurs, so that it is not possible to request additional redundant data. In fact, when using punctured convolutional codes in [17] (rather than the rateless codes used herein) it is not possible to generate additional redundant data.

Data-partitioning in this paper can be viewed as a simplified form of SNR or quality layering. Extended quality layering can also be applied to video streaming across WiMAX. In [30], adaptive multicast streaming was
proposed for WiMAX using the Scalable Video Coding (SVC) extension for H.264. WiMAX channel conditions were monitored in order to vary the bitrate accordingly. Unfortunately, the subsequent decision of the JVT standardization body for H.264/AVC not to support fine-grained scalability (FGS) implied it will be harder to respond to channel volatility in the way proposed in [30]. Other work concerned with video streaming over WiMAX links has also investigated: combining scalable video with multi-connections in [31]; and compared [32] H.264/SVC with H.264/AVC for WiMAX. However, the data-dependencies between layers in H.264/SVC medium-grained scalability are a concern. Unlike in FGS, enhancement layer packets may successfully arrive but be unable to be reconstructed if key pictures also fail to arrive. Besides, for commercial one-way streaming, simulcast is now likely to be preferred to H.264/SVC for the reasons outlined in [33]. In [33], it was found that the extra overhead from sending an SVC stream compared to an H.264/AVC stream meant that the cost of bandwidth consumption outweighed the reduced storage cost of SVC once more than 64 sessions had occurred (assuming 16 simulcast streams or 16 video layers per session). In another comparison [34], it was proposed that scalable video with unequal error protection cannot provide any advantage over H.264/AVC with equal error protection in a wireless environment, due to the overhead of scalable video coding compared to that of single-layer coding.

3. Outline of streaming system

This Section outlines an effective video streaming system for WiMAX that provides error resilience through source-code data partitioning [17] and which works without the need to apply privileged protection to the high-priority partitions. As already mention in Section 1, the main point of this paper is to draw attention to the importance of correctly configuring the video codec parameters. Therefore, this Section provides a basic outline only and the interested reader is referred to the authors’ other works, such as [15] and [16], for more details and variants of our streaming system.
3.1. H.264 codec

Features of the H.264 codec are used in this work. H.264/Advanced Video Coding (AVC) standardization was initiated by the Video Coding Experts Group (VCEG), which is a working group of the International Telecommunication Union (ITU-T). The Joint Video Team carried out the final work of developing H.264/AVC [6], a co-operative effort of both Moving Picture Expert Group MPEG and VCEG in 2003. Figure 1 shows the video frame structure. A video-frame in H.264/AVC is divided into macroblocks (MBs) [35], which are the basic unit for motion prediction. Each MB can be further sub-divided into blocks that are transform-coded in order to de-correlate the data. For transmission, a video frame can be split into slices, each of which occupies a network packet. When data-partitioning is enabled, each slice can be further sub-divided in up to three partitions, data-partitions-A, -B and -C, each of which can occupy a separate network packet.

![Video frame structure](image)

Fig. 1. Video frame structure, DP-A = data-partition-A and similarly for DP-B and DP-C

The H.264/AVC Quantization Parameter (QP) is set by us in such a way that temporally-coded, texture data occupies a larger part of a frame’s data than that occupied by data in each of the other two partitions. If this texture data should be dropped, it can be replaced more easily through error concealment at the decoder than data in the other two partitions. Error concealment (backward error prediction) [36] is a non-normative feature [37] of H.26/AVC that nevertheless is present in the H.264/AVC JM reference code (found at http://iphome.hhi.de/suehring/tml/). The texture data in partition-C is packetized in a WiMAX MAC Service Data Unit (MSDU) within a MAC Protocol Data Unit (MPDU) [11] [38]. In contrast to the authors’ counter-
intuitive approach, the intuitive approach is to give special protection to the partition-A, which for predictively-code frames includes motion vectors as well as other important parameters. Partition-B, bearing intra-coded MBs, may also be given special protection, as it contains spatially-coded data whenever suitable references in other video frames are not available. As an example of the unequal error protection (UEP) approach, in [39] hierarchical modulation was employed to favour those partitions with more important data for the reconstruction of the video frame. For example, partition-A motion vectors were used in motion copy error concealment when partition-C data are lost and, therefore, the intuitive approach is to protect the partition-A. Readers should consider for themselves whether from the evaluation of this paper whether UEP measures are necessary.

3.2. Rateless channel coding

Though no special protection is given, it is still necessary to protect the bit-stream (without privileging partitions A and B) against the risk of packet loss. This was achieved through equal error protection (EEP) through FEC provision. Application-layer rateless coding [40] was selected for its flexibility and its linear computational complexity at both the decoder and the encoder. To avoid long latencies, which would occur if packet-level FEC were to be applied, redundant data were added to the packets themselves, treating the bytes within each packet as the data symbols. Again to reduce latency, a single Automatic Repeat reQuest (ARQ) was made if the available data were insufficient to reconstruct a corrupted packet. Additional rateless error data were added to the next available packet to be transmitted, according to the amount calculated in the way specified by us elsewhere [15].

We selected Raptor codes [41] as our rateless code. For a rateless code, even if there were no channel errors there is a very small probability that the decoding will fail, which failure can be modeled statistically by the scheme introduced in [25]. In compensation, these codes are completely flexible, have linear decode computational complexity, and generally their overhead is considerably reduced compared to fixed erasure codes. Furthermore, if the packets are pre-encoded with an inner code, a weakened Luby Transform [42] can be applied to the symbols and their redundant symbols. Figure 2 is a simple diagram of how a Raptor code works. k
represents source symbols and $m$ represents received symbols. A systematic Raptor code is arrived at [41] by first applying the inverse of the inner code to the first $k$ symbols before the outer pre-coding step.

Fig. 2. Raptor coding scheme

4. Simulation model

This Section details how we set about modeling the WiMAX system for the evaluation of Section 4.

4.1. Simulation system

Figure 3 shows the overall data flow of the simulation system. Raw video (.YUV) is encoded by H.264/AVC into compressed form. The compressed file is passed for later extraction of any dropped packets. At the same time, a video trace file is generated, which will become an input into the ns-2 simulator [43]. The trace file contains the size of each video packet and the transmission schedule. However, the trace file does not contain any video data. After simulation using the wireless channel model, the simulator outputs a list of sent and dropped packets. This is used to filter from the original compressed data file any data that did not arrive at the destination. The file is subsequently decoded by the H.264/AVC decoder. The decoder outputs a raw video file in YUV format (strictly in its digital form of YCbCr). This file would normally be displayed but, in this case, it is
compared with the original raw input video. This allows the objective video quality to be calculated as Peak
Signal-to-Noise Ratio (PSNR), through a pixel-by-pixel comparison.

Fig. 3. Simulation system

Two video clips with different source-coding characteristics were employed in the tests in order to judge the
dependency of the results upon video source-coding complexity. The first test sequence was Paris, which is a
studio scene with two upper body images of presenters and moderate motion. The background is of moderate to
high spatial complexity leading to larger slices. The other test sequence was Football, which has rapid
movements and consequently has high temporal coding complexity. Both sequences were encoded at Common
Intermediate Format (CIF) (352 × 288 pixel/picture). CIF resolution was used for ready comparison with the
prior work of others on video communication to mobile devices.
4.2. Wireless channel model

The Gilbert-Elliott channel mode employed in this work has been used by researchers in the wireless field [44] [45] because of its ability to model error burst patterns as experienced at the receiver. This channel model was introduced into the ns-2 event-driven simulator [43]. Ns-2 is an open-source simulator that since 1989 has been developed by the network research community. Simulations are controlled by a scripting language which interacts with simulation objects written in the C++ programming language. When an event occurs, such as a packet arrival, then a C++ event handler is called. The byte-level Gilbert-Elliott code was introduced into the appropriate event-handlers called by the simulator. In fact, this was the means that the modelled Raptor code response was also introduced into the simulator. Ns-2 also has a built-in Gilbert-Elliott channel model but this operates at the packet-level. When both byte-level and packet-level Gilbert-Elliott models were used then the built-in ns-2 model was employed by setting the appropriate parameters in the ns-2 configuration file.

The Gilbert-Elliott channel model itself is a two-state Markov chain. It is based on: good and bad states; the probabilities of these states; and the probabilities of the transition states between them. In the case of the bad state, losses happen with higher probability while in the good state losses happen with lower probability, PGG refers to the probability of being in the good state and PGB is the probability of a transition from the good state to the bad state. PBB is the probability of being in the bad state and PBG refers to the probability of a transition from the bad to good state. PGG (PBB) can be interpreted as the probability of remaining in the good (bad) state, given that the previous state was good (bad). Conversely, PGB represents the only other probability in this situation, that is, given that the previous state was good, a transition is made from the good to the bad state. By the law of total probability, all probabilities sum to one (certainty). Therefore, we have PGG + PGB = 1, resulting in (1). A similar argument for the bad state leads to (2).

\[ P_{GG} = 1 - P_{GB} \quad (1) \]

\[ P_{BB} = 1 - P_{BG} \quad (2) \]

For the stochastic process to remain stationary in time,

\[ \pi_G P_{GB} = \pi_B P_{BG} \quad (3) \]
where $\pi_G$ and $\pi_B$ are the steady state probabilities of being in a good or bad state respectively. Again by the law of total probability, $\pi_B = 1 - \pi_G$. Substituting this expression for $\pi_B$ into (3) easily leads to:

$$\pi_G = \frac{P_{BG}}{P_{BG} + P_{GB}}$$

(4)

Similarly, $\pi_B = 1 - \pi_G$. Substituting this expression for $\pi_G$ into (3) easily leads to:

$$\pi_B = \frac{P_{GB}}{P_{BG} + P_{GB}}$$

(5)

The Gilbert-Elliott model good and bad states have their own error distributions that are independent of the process of arriving or leaving those states. Suppose the probability of packet loss is $p_G$ and $p_B$ in the good and bad states respectively. Then the average packet loss rate produced by a Gilbert-Elliott Channel is given in (6) by the usual expression for the expectation of a probability distribution.

$$P = p_G\pi_G + p_B\pi_B$$

(6)

To model the effect of slow fading at the packet-level, the PGG was set to 0.96, PBB = 0.95, PG = 0.01 and PB = 0.02 for the Gilbert-Elliott model’s parameters. Additionally, it is still possible for a packet not to be dropped in the channel but nonetheless to be corrupted through the effect of fast fading (or other sources of noise and interference). This byte-level corruption was modelled by the second Gilbert-Elliott model, with the same parameters (applied at the byte level) as that of the packet-level model except that PB (now probability of byte loss) was increased to 0.165. Effectively, this second model emulates fast fading between good and bad conditions.

4.3. WiMAX simulation configuration

The physical (PHY) layer settings selected for WiMAX simulation on ns-2 [46] are given in Table 1. The antenna was modelled for comparison purposes as a half-wavelength dipole. The Time Division Duplex (TDD) frame length was set to 5 ms, is this is the only value supported by the WiMAX Forum [47]. Video was transmitted over the downlink with UDP transport (see Section 1). In order to introduce sources of traffic
congestion, an always available FTP (File Transfer Protocol) source was introduced with TCP transport to a Subscriber Station (SS) from the Base Station (BS). Likewise a CBR source with packet size of 1000 byte and inter-packet gap of 0.03 s was downloaded to a SS.

Table 1. IEEE 802.16 (WiMAX) parameter settings

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHY</td>
<td>OFDM</td>
</tr>
<tr>
<td>Frequency band</td>
<td>5 GHz</td>
</tr>
<tr>
<td>Bandwidth capacity</td>
<td>10 MHz</td>
</tr>
<tr>
<td>Duplexing mode</td>
<td>TDD</td>
</tr>
<tr>
<td>Frame length</td>
<td>5 ms</td>
</tr>
<tr>
<td>Max. packet length</td>
<td>1024 B</td>
</tr>
<tr>
<td>Raw data rate</td>
<td>10.67 Mbps</td>
</tr>
<tr>
<td>IFFT size</td>
<td>1024</td>
</tr>
<tr>
<td>Modulation</td>
<td>16-QAM 1/2</td>
</tr>
<tr>
<td>Guard band ratio</td>
<td>1/8</td>
</tr>
<tr>
<td>DL/UL ratio</td>
<td>3:1</td>
</tr>
<tr>
<td>Channel model</td>
<td>Gilbert-Elliott</td>
</tr>
<tr>
<td>SS transmit power</td>
<td>245 mW</td>
</tr>
<tr>
<td>BS transmit power</td>
<td>20 W</td>
</tr>
<tr>
<td>Approx. range to SS</td>
<td>1 km</td>
</tr>
<tr>
<td>Antenna type</td>
<td>Omni-directional</td>
</tr>
<tr>
<td>Antenna gains</td>
<td>0 dBD</td>
</tr>
<tr>
<td>MS antenna height</td>
<td>1.2 m</td>
</tr>
<tr>
<td>BS antenna height</td>
<td>30</td>
</tr>
</tbody>
</table>

5. Evaluation of codec settings

This Section evaluates in turn each of the codec configurations mentioned in Section 1.

5.1. Intra-refresh macroblocks

In H.264/AVC data partitioning [17] [48], motion vectors (MVs) are packed into partition-A bearing NAL units, allowing motion copy error concealment at the decoder to partially reconstruct a picture, despite missing partition-C NAL units (containing quantized transform coefficient residuals). Partition-B NAL units contain intra-coded (spatially encoded) MBs, which are substituted for inter-coded MBs according to encoder implementation (only the decoder input format is standardized in H.264/AVC). Therefore, when intra-refresh (IR) MBs are included alongside naturally intra-encoded MBs, partition-B slices grow in size. This means that
data-partitioned video compression provides a convenient way to examine the effect of various amounts of IR MB provision. Once H.264/AVC has formed a NAL unit, it can also provide a RTP header prior to encapsulation by IP/UDP network protocols.

A point to note is the different way that random IR MBs are specified in the H.264/AVC JM 14.2 implementation compared to that of cyclic IR line intra update. In random IR MB, a maximum percentage of IR can be specified, which percentage includes already encoded IR MB. If the given quota of IR MB is already largely occupied by naturally encoded MBs (those encoded for to cover newly revealed objects or to improve the quality of the video up to a given bit budget or some other encoder-dependent reason), then only a small amount of extra randomly inserted MBs will be added. In contrast, if a line of IR MBs is inserted then these MBs are added in addition to those intra-coded MBs that have already been included by the encoder, as shown in Figure 4.

Football was VBR encoded with a Group-of-Pictures (GoP) structure of IPPP..... at 30 frame/s. From Table 2, it is apparent that, as the percentage of IR MBs is increased, the size in bytes of partition-B increases for the same QP. Because more MBs are assigned to partition-B, the size of partition-C reduces. And because of the large

Fig. 4. Differences between random intra-refresh MBs (upper frames with 6%) and MB cyclic line intra update (lower frames).

Football was VBR encoded with a Group-of-Pictures (GoP) structure of IPPP..... at 30 frame/s. From Table 2, it is apparent that, as the percentage of IR MBs is increased, the size in bytes of partition-B increases for the same QP. Because more MBs are assigned to partition-B, the size of partition-C reduces. And because of the large
amount of naturally encoded intra MBs, this effect is gradual until 25% of random IR MBs are added. 25% of random IR MBs is shown in Table 2, as that amount approximately corresponds to the total partition-B size if cyclic line intra update is turned on instead (with approximately the same number of MBs).

Table 2
Mean size of different partitions in bytes for Football at various QPs

<table>
<thead>
<tr>
<th>QP</th>
<th>2% Intra-refresh MB</th>
<th>5% Intra-refresh MB</th>
<th>6% Intra refresh MB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>20</td>
<td>1842</td>
<td>2678</td>
<td>3889</td>
</tr>
<tr>
<td>25</td>
<td>1687</td>
<td>1697</td>
<td>2533</td>
</tr>
<tr>
<td>30</td>
<td>1459</td>
<td>1047</td>
<td>1496</td>
</tr>
<tr>
<td>35</td>
<td>1117</td>
<td>572</td>
<td>688</td>
</tr>
</tbody>
</table>

In Figure 5, increasing the provision of IR MBs from 2% to 5%, then to 6% and finally 25% (in the case of MB line intra update), increases the throughput and, hence, the bandwidth requirements in respect to co-existing traffic. A quantity of 6% rather than 5% IR MB refresh is preferable because, without naturally-encoded IR MBs, then one line of MBs corresponds to about 6% of a CIF picture. A 25% IR commitment is large due to the coding inefficiency of spatial reference coding. From the objective video quality (PSNR) results, it can be seen that reducing the IR MB percentage to 2% actually improves the PSNR at QPs 30 and 35. The size of packets is the most important factor affecting the percentage of dropped packets, as is evident from the decrease in dropped packet percentages when the QP increases (and video quality decreases).

Packet end-to-end delay is the mean delay of those packets unaffected by channel conditions. The results show that this is generally small in duration, though with a tendency to increase due to the propagation delay of larger packets at lower QP. However, there are larger percentages of corrupted packets. These are packets that have not been repaired completely by the adaptive channel coding scheme. Because of the additional
transmission time, the mean end-to-end delay of corrupted packets is higher than other packets. In fact, it is the extent of the delay that is the main contribution of corrupted packets to the quality of service.

One can see that the video quality is generally below 25 dB, and, hence, would probably be ranked as ‘poor’ according to the ITU P.800’s recommendation, originally intended for subjective testing but used in an objective as guide to quality in papers such as [49]. However, for the higher QPs of 30 and 35 under the ‘NAL redundant scheme’, the video quality is actually ‘good’ (above 31 dB) on the ITU scale. Comparing with the percentage of dropped packets for these QPs under ‘dropped packets’, the percentage of dropped packets is low.

The main effect of reducing the percentage of IR MBs is that the size of partition-B-bearing packets is reduced. In turn, this makes these packets less likely to be affected by channel conditions, especially burst errors arising from fast fading. During bursts it is possible that a packet and its redundant replacement are both affected by channel noise. Thus, extra redundant data are transmitted in an attempt to reconstruct the packet. However, if the retransmitted packet is itself dropped or corrupted then the original packet cannot be repaired.
Fig. 5. Mean performance metrics for Football with 2%, 5%, 6% IR MBs and MB line intra update.
5.2 Constrained Intra Prediction setting

In order to decode partition-B and -C, the decoder must know the location from which each MB was predicted, which implies that partitions B and C cannot be reconstructed if partition-A is lost. Though partition-A is independent of partitions B and C, CIP should be set in the codec configuration [21] to make partition-B independent of partition-C. (Though reference [21] refers to a proposal to add constrained inter prediction to H.264/AVC, it also describes constrained intra-prediction, which is already a part of the codec standard.) By setting this option, partition-B MBs are no longer predicted from neighbouring inter-coded MBs. This is because the prediction residuals from neighbouring inter-coded MBs reside in partition-C and cannot be accessed by the decoder if a partition-C packet is lost. There is a by-product of increasing packet sizes due to a reduction in compression efficiency but the increase in size may be justified in error-prone environments.

The two video clips (Paris and Football) were VBR encoded at CIF, with a GOP structure of IPPP….. at 30 Hz and with 5% IR MBs. Table 3 presents the results with and without CIP showing the increase in partition-B and -C sizes that results from the loss in encoding efficiency. Notice that NALs that might be above the maximum packet size of 1024 B (Table 3) were constrained to the maximum by the encoder when forming an RTP packet prior to encapsulation by network headers. This means that these NALs would not be segmented before reaching the link layer, avoiding the possible separation of header information from NAL data. At lower QPs, i.e. higher-quality video, the relative size of partition-C NALs means that the more important partition-A and –B packets are less likely to suffer channel error. However, the larger packet sizes mean that congestion may have more of an impact because longer packets take longer to transmit and free the channel. At higher QP, the advantage of differential packet sizes is lost but the generally smaller packet sizes compensate to some extent. There is also a small growth in partition-B and –C packet size when CIP is turned on.
From Table 4 for Football one sees that though the relative ranking of sizes between the partition types is similar, the actual sizes are larger than those for Paris (see Table 3). The larger sizes are due to the temporal coding complexity of Football. For high QP, the relatively larger size of partition-A NALs compared to the other partitions NALs may create a problem, as it does not result in a relatively reduced risk of loss of packets bearing partition-A NALs. Also of concern is the number of NAL units that are above the maximum packet size, causing more than one packet to be sent.

Table 4
Football video sequence: Mean NAL size according to partition type

<table>
<thead>
<tr>
<th>QP</th>
<th>Without CIP (B)</th>
<th>With CIP (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>20</td>
<td>1845</td>
<td>2767</td>
</tr>
<tr>
<td>25</td>
<td>1690</td>
<td>1763</td>
</tr>
<tr>
<td>30</td>
<td>1463</td>
<td>1082</td>
</tr>
<tr>
<td>35</td>
<td>1120</td>
<td>595</td>
</tr>
</tbody>
</table>

Figure 6 shows the effect of the various schemes on packet drops when streaming Paris. The Figure assesses for the presence of CIP or its absence. From Figure 6, the larger packet drop rates at QP = 20 will have a significant effect on the video quality. However, the packet size changes with and without CIP have little effect on the packet drop rate.
Figure 7 shows the pattern of corrupted packet losses arising from simulated fast fading. There is actually an increase in the percentage of packets corrupted if a completely redundant stream is sent (containing partitions A, B, and C), though this percentage is taken from corrupted original and redundant packets. Nevertheless, the effect of the corrupted packets on video quality only occurs if a packet cannot be reconstructed after application of the adaptive retransmission scheme.
Examining Figure 8 for the resulting PSNR, one sees that data partitioning with FEC protection is insufficient to bring the video quality to above 31 dB. This is similar in general terms to the experience reported in [50], though for multicast streaming and without feedback.
The impact of corrupted packets, given the inclusion of retransmitted additional redundant data, is largely seen in additional delay. There is an approximate doubling in per packet delay between the total end-to-end delay for normal packet end-to-end delay, Figure 9, and corrupted packets, Figure 10. Normal packets do not, of course, experience the additional delay of a further retransmission prior to reconstruction at the decoder. Therefore, the main penalty arising under the FEC protection scheme from an increased percentage of corrupted packets is an overall increase in delay. Nevertheless, the delays remain in the tens of millisecond range, except for when QP = 20, i.e. broadcast quality video. It should be recalled though that for the redundant schemes there is up to twice the number of packets being sent. Therefore, delay is approximately further doubled, still though with end-to-end delays remaining in the tens of millisecond range. This type of delay range is acceptable even for interactive applications but may contribute to additional delay if it forms part of a longer network path.

Fig. 8. Paris and Football video sequences: Protection scheme video quality (PSNR), with and without CIP.
Fig. 9. *Paris* and *Football* video sequences: Protection scheme mean end-to-end packet delay, with and without CIP

Fig. 10. *Paris* and *Football* video sequences: Protection scheme mean delay of corrupted packets, with and without CIP
5.3 Group of Pictures structure

Attention should be given to the Group of Pictures (GoP) structure. For the purposes of comparison, the authors tested two different GoP structures: IPPP… (i.e. one initial I-picture and all P) and IBBPBBP… (i.e. insertion of bi-predictive B-pictures for greater coding efficiency but still with one initial I-picture). Before starting the tests, the author examined the effect of the two different GoPs on the sizes of video frames. Table 5 shows one potential impact of the GoP structure on the scheme. B-pictures increase coding efficiency at a cost in coding complexity. However, with the inclusion of B-pictures, the mean size of P-pictures increases, as a result of the increased reference distance. For example, for QP = 20, the IBBPBBP… mean P-picture size is around 15 kB.

Table 5
Mean P-picture size (bytes) according to QP for two different GoP structures

<table>
<thead>
<tr>
<th>QP</th>
<th>Football</th>
<th></th>
<th>Paris</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IPPP…</td>
<td>IBBP…</td>
<td>IPPP…</td>
<td>IBBP…</td>
</tr>
<tr>
<td>20</td>
<td>8905</td>
<td>15520</td>
<td>5102</td>
<td>7590</td>
</tr>
<tr>
<td>25</td>
<td>6301</td>
<td>10311</td>
<td>2824</td>
<td>4249</td>
</tr>
<tr>
<td>30</td>
<td>4185</td>
<td>6381</td>
<td>1398</td>
<td>2238</td>
</tr>
<tr>
<td>35</td>
<td>2444</td>
<td>3756</td>
<td>647</td>
<td>1100</td>
</tr>
</tbody>
</table>

The tests were again performed on Paris and Football. Both sequences were VBR encoded video at 30 Hz, CIF, with 2% IR MBs randomly inserted. CIP was configured and different GoP types (IPPP… and IBBP…) were used for the video traces.

The effect of adding B-picture is evident in Figure 11, in which many more packets are dropped for Football at QP=20. ‘Dropped packets’ includes buffer overflow and outright channel drops. The GoP structure impacts the more temporally complex sequence. The general decay in packet drop numbers in Figure 11, results in an increase in video quality in Figure 12. An IPPP… GoP structure is preferable except at QP = 35. However, at QP = 35 all GoP structures result in a PSNR over 25 dB.
Fig. 11. Dropped packets for differing GoP structure and content

Fig. 12. Video quality (PSNR) for two different GoP structures and content
Corrupted packet levels are generally high, Figure 13, but in respect to GoP structure it seems that an IPPP… structure may be more favourable for temporally complex content (Football) and vice versa for less active sequences. The main consequence of higher levels of packet corruption, after the application of the proposed adaptive scheme, is in greater delay for a greater percentage of packets, Figure 14. Apart from the anomaly at QP = 20, corrupted packet delay is approximately twice that of normal packet end-to-end delay, Figure 15, reflecting the single retransmission of extra redundant that is permitted. It was also confirmed that for a moderate increase in mean packet size, making partition-B completely independent of partition-C through CIP resulted in a small (a few dB) improvement in video quality, whenever the QP setting allowed sufficient packets to be delivered.

![Fig. 13. Percentage of corrupted packets for two different GoP structures and content](image)
Fig. 14. Corrupted packet mean delay for two different GoP structures and content
There is a variety of ways of providing redundant packets. It is possible to use redundant picture slices [51] rather than duplicate slices. Redundant picture slices employ a higher QP and, hence, coarser quantization than the original slices. However, in our scheme using redundant picture slices can cause additional drift between the encoder and the decoder, if (say) a partition-A packet was matched with a partition-C packet’s data with a different QP. There is a penalty in user perception from switching between different fidelities but this is preferable to the freeze frame effect that would result if the primary redundant slice were to be lost. Even so, there is an implementation issue. Though both redundant slices and data-partitioned slices co-exist in the Extended profile, they are not jointly implemented in the JM implementation of H.264/AVC [52] and, in fact, appear to not to be implemented at all in most other software codec implementations such as QuickTime, Nero, and LEAD randomly to name a few. However, it is possible with data-partitioning through repeated runs of the encoder to create a duplicate stream of all partition-A slice packets or a duplicate stream consisting of partition-A and partition-B packets or, indeed, a complete replica of the original stream. Additionally, as the encoder is
unaware of a redundant slice substitution by the decoder there will be error drift when this occurs. Methods to refine the selection of redundant slices [53] have also been designed.

The redundant NAL method is a single slice per picture scheme and, hence, no advantage is gained from smaller packet sizes. 5% IR MB data were added to each picture, increasing the size of partition-B packets (refer to Table 2). The duplicate NAL scheme extends to all partitions. This does not amount to a change in bitrate because the packets are simply replicated. However, the end-to-end packet delay will obviously increase because of the interleaving of the duplicate slice packets. Notice also that the number of packets sent for the two-slice scheme is the same as for the redundant slice packet scheme.

In the redundant NAL scheme, retransmission of extra redundant data was scheduled for all corrupted packets, even if two packets duplicated each other. This is because it is not possible to know in advance whether the extra redundant data will arrive for any one of the two packets. This provision has a significant effect in improving the video quality at higher QPs. The reason is that retransmitting extra redundant data by two alternative means increases the chance that a packet can be reconstructed.

The same two video clips (Paris and Football) with different source coding characteristics were employed in the tests to judge content dependency. Both sequences were VBR encoded, with a GoP structure of IPPP….. at 30 Hz. 5% IR MBs (randomly placed) were included.

Figure 16 shows the effect of the various schemes on packet drops when streaming Paris. ‘Data-partition’ in the Figure legend refers to sending no redundant packets. ‘Redundant X’ refers to sending duplicate redundant packets containing data-partitions of partition type(s) X in addition to the rateless coded data-partition packets. From Figure 16, the larger packet drop rates at QP = 20 will have a significant effect on the video quality.
Figure 17 shows the pattern of corrupted packet losses arising from simulated fast fading. There is actually an increase in the percentage of packets corrupted if a completely redundant stream is sent (partitions A, B, and C), though this percentage is taken from corrupted original and redundant packets. However, the effect of the corrupted packets on video quality only occurs if a packet cannot be reconstructed after application of the adaptive retransmission scheme.

Examining Figure 18 for the resulting video quality (PSNR), one sees that data partitioning with FEC protection, when used without redundancy, is insufficient to bring the video quality to above 31 dB, that is to a good quality. However, it is important to note that sending duplicate redundant partition-A packets alone (without redundant packets from other partitions) is also insufficient to raise the video quality to a good rating (above 31 dB). Therefore, to raise the video quality to a good level requires not only the application of the adaptive rateless channel coding scheme but also the sending of duplicate data streams.
Fig. 17. Paris sequence protection schemes corrupted packets

Fig. 18. Paris sequence protection schemes video quality
The impact of corrupted packets, given the inclusion of retransmitted additional redundant data, is again largely seen in additional delay. As before, there is an approximate doubling in per packet delay between the total end-to-end delay for corrupted packets Figure 19 and normal packet end-to-end delay, Figure 20. It must be recalled, though, that for the redundant schemes there is up to twice the number of packets being sent. Therefore, delay is approximately further doubled, still though with end-to-end delays remaining in the tens of millisecond range. This type of delay range is acceptable even for interactive applications, but may contribute to additional delay if the WiMAX link forms part of a longer network path.

![Graph](image)

**Fig. 19.** Paris sequence protection schemes mean delay for corrupted packets
Fig. 20. Paris sequence protection schemes mean end-to-end delay for normal packets

The experimental results for Football are included in Table 5. Table 5 shows how packet drops and losses are reflected in video quality. Very large numbers of packets are dropped at QP = 20 because of the larger packet sizes. However, there is a threshold effect, as the numbers of dropped packets decline quickly with increasing QP (as packet sizes reduce). The protection pattern for redundant packets is accentuated compared to Paris, in the sense that providing duplicate redundant versions of more than just partition-A packets is now clearly seen to be preferable. Given that in quality-of-experience subjective tests for mobile devices [54], news scenes rather than sport are preferred by viewers, it may be advisable to favour content without rapid motion, especially if small footballs or similar sports’ balls need to be tracked by the viewer.

6. Conclusion

Both FEC and data-partitioning of IPTV video streams are a way of providing graceful quality degradation in a form that will work in good and difficult wireless channel conditions. This paper showed that video
configuration also affect the video quality, dropped packets and delay. In that respect, it was shown that it is better to include a small percentage of IR MBs that can build their effect over time than employ the cyclic IR line update scheme. Packet size, which is determined by content, video quality, and GoP structure is an important determinant of packet drops. The use of equal error protection is a way of taking advantage of the natural packet size differential, which is in inverse order of the priority of the data partitions. Thus, smaller packet lengths already confer a lower risk of channel error. However, the inverse size order of data partitions (larger partition-A and –B) was seen to occur when smaller QPs were chosen. It was also confirmed that for a moderate increase in mean packet size, making partition-B completely independent of partition-C resulted in a small but significant improvement in video quality.

An interesting observation is that there is a need to reduce packet size to reduce packet loss, despite the combined effect of redundant packets and application adaptive channel coding. This is because during ‘bursty’ error conditions (as was simulated by the Gilbert-Elliott channel model) it is possible that both the original packet and its redundant counterpart may be dropped or corrupted. However, this effect is dependent on choice of QP, as a low QP can lead to high packet drop rates with poor video quality. In general, in poor channel conditions with both slow and fast fading, it is not sufficient to employ just application-layer FEC unless stream replication also takes place.

Acknowledgments

Laith Al-Jobouri, who has recently completed his doctorate at the University of Essex, UK, would like to thank his financial sponsors at the Ministry of Science and Technology, Baghdad, Iraq.

References


