

Impact of Intra-Refresh Provision on a Data-Partitioned Wireless Broadband Video Streaming Scheme

Laith Al-Jobouri, Martin Fleury, and Mohammed Ghanbari
School of Computer Science and Electronic Engineering
University of Essex, Colchester, United Kingdom
{lamoha, fleum, ghan}@essex.ac.uk

Abstract—Intra-refresh macroblocks are normally provided in mobile broadband wireless access to avoid the effect of temporal error propagation. The questions then arise are: in what form should the refresh take place; and what percentage of refresh is necessary. This paper is a study of intra-refresh provision in the context of a robust video streaming scheme. The scheme combines data-partitioned video compression with adaptive channel coding and redundant packets. The main conclusions from a detailed analysis are that: because of the effect on packet size it is important to select a moderate quantization parameter; and because of the higher overhead from cyclic intra macroblock line update it is better to select a low percentage per frame of intra-refresh macroblocks. In harsh channel conditions from the combined effect of slow and fast fading producing ‘bursty’ errors, all the proposed measures are necessary but then periodic intra-refresh can be avoided with its sudden increases in the datarate if the proposed levels of intra-refresh macroblocks are applied.

I. INTRODUCTION

There is a growing interest in all forms of mobile TV, including networked Internet Protocol TV (IPTV) with access over broadband wireless technologies such as Long Term Evolution (LTE) [1] and IEEE 802.16e (mobile WiMAX) [2], and broadcast schemes such as DVB-H, and MediaFlo. The networked services will benefit from true video streaming, rather than progressive download, as this allows the buffer size on the mobile device to be reduced. This paper considers a robust video streaming scheme for mobile WiMAX mobile access. The scheme employs data-partitioned video compression for graceful degradation in the face of channel error and in worsening channel conditions, redundant packets are transmitted. This is because, though application-layer forward error correction (FEC) can protect against packet corruption, when the overhead from FEC is large or when packets are likely to be corrupted or dropped before reaching the application, it is then preferable to transmit redundant payload.

In this robust scheme, a key consideration is how to protect against the temporal error propagation that can occur whenever predictively-coded P-frames are lost. A traditional way to do this is to insert periodic, intra-coded I-frames, usually every 12 or 15 frames that is every half-second according to frame rate. The spatially-encoded macroblocks (MBs) of the I-frame halt the temporal error propagation and act as anchor points for a future set of frames. Unfortunately, the insertion of I-frames leads to sudden data transmission peaks due to the coding inefficiency of spatial

referencing. Therefore, distributed insertion of intra-refresh (IR) MBs can be considered. In the JM implementation of the H.264/AVC (Advanced Video Coding) codec [3], two main methods of distributed insertion are available: either random placement of IR MBs within each frame; or placing a line of intra-refresh MBs within each P-frame on a cyclic basis.

In the latter forced IR method, the line size can be increased to a region or slice [4] in order to control the rate that the total picture area is refreshed. Against this suggestion must be balanced the overhead from including a complete line or region of MBs, as such MBs are costly to encode. In fact, in this paper it is suggested that despite the apparent advantages of the cyclic line method, at least in respect to data-partitioned video compression, random selection of IR MBs is preferable. Random IR refresh may on occasion duplicate IR MBs but it has the advantage that in the JM implementation of H.264/AVC the overhead from intra-coded MBs is readily controllable. This is because IR MBs are not the only form of intra-coded MBs, as encoders will insert such MBs when new areas of a picture are revealed, as may especially happen when there is rapid motion within a sequence. Thus, the forced IR method does not account for areas of the region that may already have been intra-coded. It also has another weakness in that future motion prediction may occur from regions of a prior picture yet to be refreshed. This defect can be remedied, possibly by restricting the range of the motion vectors (MVs) within the refreshed region [5] or by observing the direction of motion within a sequence [6]. However, these alternatives [5] [6] add coding complexity to IR and may be unnecessary when processing a low-motion video sequence. Another possibility is to adaptively alter the extent of MB provision [7] according to scene content and channel conditions. This is most suited to live encoding and is not a general method.

One advantage of forced IR is that it provides a natural channel-swapping (or zapping) point at the start of each refresh cycle, just as periodic I-frames provided. In time-shifted TV systems such as the BBC’s iPlayer such a facility is not required, as no channels are swapped. If for channel swapping, gradual decoder refresh is performed instead of periodic refresh then growing the refresh area from an isolated region is preferable to random MB refresh, as the subjective visual effect is better. This facility was proposed to the Joint Video team developing the H.264/AVC codec [8] but is not currently implemented, perhaps because a

method of signaling a switching point is the subject of a patent.

The remainder of the paper is organized as follows. Section II, considers the complete video streaming scheme and includes an analysis of the impact of IR on packet size, prior to transmission over a WiMAX channel. This Section also describes the simulation model that is used in the evaluation of Section III. Finally, Section IV makes some concluding remarks.

II. METHODOLOGY

A. Data-partitioned video and the effect of IR MBs

In H.264/AVC data partitioning [9], MVs are packed into a partition-A bearing Network Adaptation Layer unit (NALU), allowing motion copy error concealment at the decoder to partially reconstruct a picture despite missing partition-C NALUs containing texture data (quantized transform coefficient residuals). Partition-B NALUs 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 IR MBs are included alongside naturally intra-encoded MBs referred to in Section I, 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 NALU, it can also provide a Real Time Protocol (RTP) header prior to encapsulation by IP/UDP.

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 MB, then only a small amount of extra randomly inserted MB will be added. In contrast, if a line of IR MB is inserted then these MBs are added in addition to those intra-coded MBs that have already been included by the encoder.

Table I is a comparison between the relative sizes of the partitions according to QP for the video clip which is employed in evaluation. The test sequence was *Football*, which is a scene with rapid movement and consequently has high temporal coding complexity. Because of this movement it is likely that the number of naturally intra-encoded MBs is higher than a sequence with less motion.

Football was Variable Bit Rate (VBR) encoded at Common Intermediate Format (CIF) (352×288 pixel/picture), with a Group of Pictures (GOP) structure of IPPP..... at 30 frame/s, i.e. one initial I-picture followed by all predictive P-pictures. This common arrangement removes the complexity of bi-predictive B-pictures at a cost

TABLE I. MEAN SIZE OF DIFFERENT PARTITIONS IN BYTES FOR FOOTBALL AT VARIOUS QP.

QP	2% Intra refresh MB			
	A	B	C	Total
20	1842	2678	3889	8409
25	1687	1697	2533	5917
30	1459	1047	1496	4002
35	1117	572	688	2377
QP	5% Intra refresh MB			
	A	B	C	Total
20	1845	2767	3867	8479
25	1690	1763	2511	5964
30	1463	1082	1482	4027
35	1120	595	682	2397
QP	6% Intra refresh MB			
	A	B	C	Total
20	1846	2810	3850	8506
25	1696	1793	2502	5991
30	1467	1098	1479	4044
35	1123	604	681	2408
QP	25% Intra refresh MB			
	A	B	C	Total
20	1893	3450	3669	9012
25	1746	2216	2379	6341
30	1505	1346	1405	4256
35	1146	729	646	2521
QP	MB Line Intra Update			
	A	B	C	Total
20	1885	3385	3683	8953
25	3683	2160	2400	8243
30	1498	1312	1414	4224
35	1143	716	652	2511

in increased bit rate. The range of QP in H.264/AVC is 0–51 with higher values corresponding to higher compression ratios and lower quality video.

From the Table, 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. Because of the large amount of naturally intra-encoded MBs, this effect is gradual until 25% of random IR MBs are added. The higher amount of random IR MBs is shown in the Table, as this amount, 25%, approximately corresponds to the total partition-B size if cyclic line intra update is turned on instead.

Another point to notice is that the total size of the stream declines significantly with coarser quantization due to a higher QP. Because intra-coding is less efficient, the total sizes increase as the percentage of IR MBs is increased.

B. Transmission protection scheme

To protect transmission of the IP/UDP/RTP packets an adaptive rateless channel coding scheme was devised. Notice that header compression is often employed over wireless links, which for the headers mentioned is able to achieve [10] up to 97.5% compression. The form of rateless coding [11] was Raptor code [12], which has linear decode computational complexity. Because rateless decoding is

TABLE II. IEEE 802.16E PARAMETER SETTINGS.

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

OFDMA = Orthogonal Frequency Division Multiple Access, QAM = Quadrature Amplitude Modulation, TDD = Time Division Duplex SS=subscriber station

probabilistic, the behavior was modeled statistically, according to the formulation in [13]. The adaptive rateless coding scheme relies on channel condition estimation but is robust against measurement noise. Further channel protection is afforded by an additional transmission of a fixed additional increment to the rateless code. However, retransmission only occurs once to reduce delay and the additional redundant data is piggybacked onto the next outgoing packet. Because the adaptive component is not the main focus of this paper, the reader is referred elsewhere [14] for details.

C. WiMAX simulation

To establish the behavior of the scheme under WiMAX the well-known ns-2 simulator augmented with a module from the Change Gung University, Taiwan [15] that has proved an effective way of modeling IEEE 802.16e's behavior.

The physical layer (PHY) settings selected for WiMAX simulation are given in Table II. The antenna is modeled for comparison purposes as a half-wavelength dipole. The Time Division Duplex (TDD) frame length was set to 5 ms in experiments as this is the only setting specified by the WiMAX Forum, though the IEEE 802.16e standard allows for a range of settings.

In order to introduce sources of traffic congestion, an always available FTP source was introduced with TCP transport to the SS. Likewise a Constant Bit Rate (CBR) source with packet size of 1000 B and inter-packet gap of 0.03 s was downloaded to the SS. While the CBR and FTP occupy the non-rtPS (non-real-time polling service) queue,

rather than the rtPS queue, they still contribute to packet drops in the rtPS queue for the video, if the 50 packet rtPS buffer is already full or nearly full, while the nrtPS queue is being serviced.

D. IEEE 802.16e channel model

A two-state Gilbert-Elliott channel model [16] was introduced into the physical layer of the simulation to simulate the channel model for WiMAX. To model the effect of slow fading at the packet-level, the PGG (probability of staying in a good state) was set to 0.96, PBB (probability of staying in a bad state) = 0.95, PG (probability of packet loss in a good state) = 0.01 and PB (probability of packet loss in a bad state) = 0.02 for the Gilbert-Elliott parameters. Notice that PGB (probability of leaving the good state) is $1 - \text{PGG}$, and similarly PBG = $1 - \text{PBB}$.

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 modeled by a 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.

III. EVALUATION

Before looking at the impact of different amounts of IR MBs, Table III examines the effect of differing forms of packet redundancy upon the transmission of data-partitioned packets over the IEEE 802.16e channel. 'NAL only' refers to the streaming without any packet redundancy, whereas 'NAL redundant' means that all NALU-bearing packets are duplicated in the stream. Other columns refers to the replacement of partition-A or both partition-A- and partition-B-bearing packets. The redundant packets were scheduled to be transmitted within the same sending interval as their original counterparts. This implies that there is no latency effect from including redundancy but there is an increase in throughput for a given QP. This might seem a heavy penalty but an important point to notice is that from Table I, going from QP 20 to 30 for 5% IR MB results in more than a halving of the size of the stream. Thus, duplicating the stream at QP = 30 results in a similar throughput to not duplicating the stream at QP = 20.

One reason that duplication is necessary is that dropped packets arising from buffer overflow at the sender or complete packet loss from fast fading cannot be redressed by application-layer FEC. Unfortunately, it is also possible in 'bursty' error conditions that both the original and the redundant replacement are lost. This can occur when the percentages of dropped packets is large. 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 as the QP increases. Packet end-to-end delay is the mean delay of those packets unaffected by

TABLE III. MEAN PERFORMANCE METRICS FOR VARIOUS REDUNDANT NALUS PROTECTION SCHEMES USED WITH 5% INTRA-REFRESH MBS

QP	Dropped Packets %			
	NAL only	A redundant	A and B redundant	NAL redundant
20	6.92	2.69	13.55	23.12
25	4.23	2.50	1.38	4.88
30	3.97	1.44	0.46	0.12
35	1.66	1.44	0.38	0.00
QP	Packet end-to-end delay (s)			
	NAL only	A redundant	A and B redundant	NAL redundant
20	0.012	0.032	0.083	0.112
25	0.010	0.009	0.019	0.071
30	0.008	0.008	0.008	0.009
35	0.007	0.007	0.007	0.007
QP	Mean PSNR (dB)			
	NAL only	A redundant	A and B redundant	NAL redundant
20	19.98	21.87	24.72	18.04
25	20.96	19.93	22.72	21.72
30	21.16	20.78	25.82	35.27
35	23.27	23.34	31.76	33.45
QP	Corrupted Packets %			
	NAL only	A redundant	A and B redundant	NAL redundant
20	30.76	31.37	22.72	14.51
25	30.64	30.79	30.27	25.11
30	27.56	30.79	27.42	26.97
35	21.92	16.55	24.73	22.35
QP	Corrupted packet mean delay (s)			
	NAL only	A redundant	A and B redundant	NAL redundant
20	0.022	0.042	0.095	0.133
25	0.020	0.019	0.030	0.083
30	0.018	0.018	0.018	0.019
35	0.017	0.021	0.016	0.017

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 and the effect of the extra packets in the ‘NAL redundant’ scheme as queuing delay is increased. Delays below 100 ms are acceptable even for interactive video applications if they are the only component of the path delay.

However, there are larger percentages of corrupted packets. These are packets that have not been repaired completely by the adaptive channel coding scheme. When redundant copies of the packet are not available then additional redundant data is requested, though this will not always be sufficient. Because of the additional transmission, 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.

Turning to the end result, one can see that the video quality expressed objectively as the Peak Signal-to-Noise Ratio (PSNR) is generally below 25 dB, and hence would probably be ranked as ‘poor’ under the ITU P.800’s [17] recommendation, originally intended for subjective testing.

TABLE IV. MEAN PERFORMANCE METRICS FOR REDUNDANT NALU PROTECTION SCHEME USED WITH 2%, 6% INTRA-REFRESH MBS AND MB LINE INTRA UPDATE

QP	Dropped Packets %		
	2% Intra refresh MB	6% Intra refresh MB	MB Line Intra Update
20	21.90	22.67	26.14
25	4.56	4.11	7.32
30	0.06	0.12	0.19
35	0.00	0.00	0.00
QP	Average Time Delay (end to end) (s)		
	2% Intra refresh MB	6% Intra refresh MB	MB Line Intra Update
20	0.109	0.113	0.110
25	0.063	0.069	0.074
30	0.010	0.009	0.010
35	0.007	0.007	0.007
QP	Mean PSNR (dB)		
	2% Intra refresh MB	6% Intra refresh MB	MB Line Intra Update
20	19.02	17.69	14.34
25	22.27	21.25	18.02
30	37.30	35.11	34.06
35	33.54	33.45	33.62
QP	Corrupted Packets %		
	2% Intra refresh MB	6% Intra refresh MB	MB Line Intra Update
20	14.51	15.28	12.26
25	24.91	24.08	22.86
30	26.23	26.65	27.16
35	22.20	23.05	23.95
QP	Corrupted Packets Average Delay (latency) (s)		
	2% Intra refresh MB	6% Intra refresh MB	MB Line Intra Update
20	0.128	0.135	0.134
25	0.075	0.085	0.090
30	0.020	0.020	0.020
35	0.017	0.018	0.017

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 actually low. This implies that the real gain of the redundant schemes is from replacement of corrupted packets, removing the risk of further loss after retransmission. Because the gain comes at the higher QP, the impact on throughput is limited. Thus, for these schemes it is preferable to avoid low QPs.

Table IV now retains the ‘NAL redundant’ scheme of Table III but varies the provision of IR MBs. Increasing the provision of IR MB above 5% to 6% and the equivalence of around 25% in the case of MB line intra update, increases the throughput and, hence, the bandwidth requirements in respect to co-existing traffic. 6% rather than 5% IR MB refresh is chosen 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

PSNR results it can be seen that reducing the IR MB percentage to 2% actually improves the PSNR at QPs 30 and 35. Other values in this Table do not differ noticeably from the ‘NAL redundant’ column in Table III. 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 the simulated 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.

Table V analyzes the numbers of dropped packets to illustrate the effect of packet size. Because the packet sizes are reduced for higher QPs, few packets are dropped at these QPs. From Table I, at high QPs, the packet size is in reverse order to the priority of the data. For example, at higher QP, partition-A packets are smaller than partition-B and -C packets. This affects the number of packets dropped of the different types. For the IR MB schemes between 2% and MB line intra update there is a definite increase in the number of partition-B packets dropped but statistical variation accounts for the counter figures going between 2% and 6% IR MBs at QP = 25. Where the higher numbers of dropped partition-B packets has an impact is when the original and its duplicate are both lost in the redundant NAL schemes.

IV. CONCLUSION

The main concentration of this paper has been on IR provision. In that respect it is 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. In fact, the random MB scheme used is more compatible with the isolated region based gradual decoder refresh, which allows for the full functionality of the replaced periodic I-frames. 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 simulated by the Gilbert-Elliott channel model) it is possible that both the original packet and its redundant counterpart may be dropped or corrupted. The packet sizes are controllable by varying the quantization parameter. Selecting a moderate quality initially may be better than selecting a higher quality video stream only to see its quality degraded by the harsh channel conditions, the effect of which is packet size dependent.

REFERENCES

[1] K. Ekström, et al. “Technical solutions for the 3G Long-Term Evolution,” *IEEE Commun. Mag.*, vol. 44, no. 3, pp. 38-45, 2006.

TABLE V. NUMBER OF DROPPED NALUS UNDER DIFFERENT SCHEMES

QP	No. of Dropped NAL UNITS								
	1 Slice A redundant			1 Slice A & B redundant			1 Slice NAL redundant		
	A	B	C	A	B	C	A	B	C
20	0	8	20	26	80	70	57	121	183
25	0	7	19	2	0	16	7	22	47
30	0	5	10	0	0	6	1	0	1
35	0	6	9	1	0	4	0	0	0
QP	2% Intra refresh MB			6% Intra refresh MB			MB Line Intra Update		
	A	B	C	A	B	C	A	B	C
	20	48	108	185	33	122	197	48	157
25	14	20	37	10	15	39	28	34	57
30	0	1	0	2	0	0	0	1	2
35	0	0	0	0	0	0	0	0	0

[2] L. Nuaymi, *WiMAX technology for broadband wireless access*, Wiley, Chichester, UK, 2007.

[3] S. Wenger, “H.264/AVC over IP,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 645-655.

[4] T. D. Tran, L.-K. Liu, and P.H. Westering, “Low-delay MPEG-2 video coding,” *Proc. SPIE - Int. Soc. Opt. Eng. (USA)*, vol.3309, pp. 510-16, 1997.

[5] E. Krause et al., “Method and apparatus for refreshing motion compensated sequential video images”, US 5,057,916, United States Patent Office, 1991.

[6] R.M. Schreier, and A. Rothermel, “Motion adaptive intra refresh for the H.264 video coding standard,” *IEEE Trans. Consumer Electronics*, vol. 52, no. 1, pp. 249-253, 2006.

[7] Y.J. Liang, K. El-Maleh, and S. Manjunath, “Upfront intra-refresh decision for low-complexity wireless video telephony,” *In Proc. of IEEE Int. Symposium on Circuits and Systems*, 2006.

[8] Y.K. Wang, and M.M. Hannuksela, “Gradual decoder refresh using isolated regions,” 3rd Meeting of the JVT, document JVT-C074, May 2002.

[9] S. Mys, P. Lambert, and W. De Neve, “SNR scalability in H.264/AVC using data partitioning,” *Proc. Pacific Rim Conf. in Multimedia*, 2006, pp. 329-338.

[10] Effnet AB, “An introduction to IP header compression,” White Paper, Bromma, Sweden, Feb. 2004.

[11] D. J. C. MacKay, “Fountain codes,” *IEE Proc.: Communications*, vol. 152, no. 6, pp. 1062-1068, 2005.

[12] A. Shokorollahi, “Raptor codes,” *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2551-2567, 2006.

[13] M. Luby, T. Gasiba, T. Stockhammer, and M. Watson, “Reliable multimedia download delivery in cellular broadcast networks,” *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 235-246, 2007.

[14] L. Al-Jobouri, M. Fleury, and M. Ghanbari, “Adaptive rateless coding for data-partitioned video streaming over a broadband wireless channel,” *IEEE Wireless Advanced*, 2010.

[15] F.C.D. Tsai, et al., “The design and implementation of WiMAX module for ns-2 simulator,” *Proc of Workshop on ns2: the IP network simulator*, 2006, article no. 5.

[16] G. Haflinger and O. Hohlfeld, “The Gilbert-Elliott model for packet loss in real time services on the Internet,” *Proc. of 14th GI/ITG Conf. on Measurement, Modelling, and Evaluation of Computer and Commun. Sys.*, 2008, pp. 269-283.

[17] ITU-T Rec. P.800 Methods for subjective testing of video quality, 1996.