A Prioritization Scheme for H.264/AVC Video Congestion Resilience with IEEE 802.11e

Sandro Moiron [1] [2], Ismail Ali [1], Martin Fleury [1] and Mohammed Ghanbari [1]
[1]School of Computer Science and Electronic Engineering - University of Essex, Colchester, United Kingdom
[2]Instituto de Telecomunicações - Leiria, Portugal
{smoiro, iaali,fleum,ghan}@essex.ac.uk

***Abstract*—Intra-refresh is an efficient technique to mitigate temporal error propagation resulting from network packet losses. However, while including intra-refresh macroblocks (MBs) can prevent error propagation, conversely when network congestion occurs, packets carrying this type of MBs should be dropped at wireless network access time, in preference to dropping packets containing data of other types of MBs. Exploiting this observation, the proposed network access prioritization scheme combines intra-refresh MB lines with the H.264/AVC codec's Flexible Macroblock Ordering (FMO) to enhance congestion resilience. The proposed scheme is shown to improve the resulting video quality by up to 4 dB during periods of congestion. As an application of the proposed algorithm, the paper applies a cross-layer mapping of video traffic to IEEE802.11e access categories. The potential gains are demonstrated both for random packet drops at input buffers and for a wireless sensor network scenario.**

I. Introduction

Though much attention has been given to providing error resilience during transmission for fragile compressed video streams, less attention has been given to congestion resilience. Suppose that some form of application layer error resilience has been generated through source coding measures. If congestion occurs, what should be the best policy when the video stream's packets must compete with coexisting traffic wishing to access a wireless network? Though a mechanism for regulating access through prioritized queues exists, IEEE 802.11e [1], it protects video traffic by a single access category. In this work, we provide further congestion resilience by prioritization that is internal to the video stream. Paradoxically, data that potentially helps to resist channel error, intra-refresh macroblocks (MBs), is given lower priority at network access time in the proposed scheme. Because the area occupied by this data corresponds to a comparatively small part of a video picture, the impact of this decision is reduced, though nevertheless intra-refresh remains an important error resiliency feature.

The H.264/AVC (Advanced Video Coding) standard [2] introduced a new set of error resilience tools to improve robustness to packet loss in transmission over error-prone networks. Flexible Macroblock Ordering (FMO) [3] allows arbitrary MB groupings into individual slices. One underused benefit of FMO is that different types of MBs can be selected to form a single slice. When channel packet loss occurs, error propagation results in temporally-predicted video frames. Introducing an intra-refresh MB line is a technique [4] that can mitigate error propagation, avoiding the need to periodically refresh the compressed bit-stream with intra-coded pictures (I-pictures). Additionally, the intra-refresh MB line technique also reduces the peak-to-average bitrate ratio compared to periodic I-pictures. Therefore, this paper proposes a prioritization scheme using an intra-refresh MB line combined with FMO.

The proposed scheme uses FMO to assemble MBs from the intra-refresh line into a single slice group. As a result, at network access time, enhanced congestion resilience can be achieved by assigning a lower priority to the packets containing intra-refresh line MBs. The logic behind this policy is further explained in Section II. Experimental results show that, for the same video data drop percentage, the proposed prioritization scheme significantly improves by up to 4 dB the video quality of the received video compared to other drop policies tested.

A number of related schemes exist. The architecture proposed in [4] is based round the data-partitioning error resilience technique of the H.264/AVC codec. In [5], a cross-layer architecture was proposed to improve H.264/AVC video delivery over IEEE 802.11e. This was accomplished by mapping each data partition of the compressed bit-stream to a different IEEE 802.11e access category (AC), allowing the prioritization of packets containing the more important data partitions.

Another general scheme was proposed in [6], where the authors analyzed the eventual influence of each MB upon picture quality in order to categorize the more important MBs and place them into their own separate slice group, representing high-priority packets for transmission. It was shown that prioritized transmission will improve the subjective and objective video quality when there is a high probability of transmission errors. Unlike our scheme, the work in [6] applied FMO to prioritize more important data for wireless channel resilience purposes, disregarding network congestion. In fact, the prioritization order was reversed in [6] compared to our scheme. A further content-aware packet-marking scheme was presented in [7] for streaming H.264/AVC coded video over a differentiated-service enabled network. Using FMO to group MBs by estimating their contribution to current frame, a relative priority index at MB level is calculated and used to mark video packets accordingly. In [8], the authors identify a number of ways to mark MBs according to their effect on bit-count, distortion-from-error concealment, distortion-from-

Fig. 1. Cyclic intra-refresh MB line technique for the Paris test sequence, showing successive MB lines in lighter shading, with some slice boundaries also shown.

error-propagation and so on. The method proposed in [8] takes into account the bit-count, distortion impact if lost, and the motion vectors of the MBs to group them into a separate FMO slice. In our paper, in contrast, a simple criterion is applied, reducing the implementation complexity. Again the work in [8] is concerned with transmission error resilience rather than congestion resilience.

The remainder of this paper is organized as follows: Section II describes the proposed algorithm and its potential gains. Section III presents the framework used to evaluate the several techniques being compared. The Section goes on to present results for a generalized congestion scenario in which random packet drops occur according to a Uniform distribution and a sensor network scenario in which simulated cross-traffic causes congestion at the IEEE 802.11e regulated access queues. Finally in Section IV, some concluding remarks are made and future development of the proposed scheme is outlined.

II. Proposed Scheme

A. Macroblocks and FMO

The proposed scheme exploits the periodic insertion of intra-refresh MB lines in a cyclic pattern within successive temporally predicted video frames. The aim of inserting intra-refresh MB lines [3] is to mitigate error propagation at the cost of lower coding efficiency than predictive inter coding. Using a vertical (or horizontal) sliding IR line, Fig. 1, reduces the error drift arising from packet loss without introducing high bitrate spikes, which occur when periodic intra-pictures are used instead. However, it should be carefully noted that an IR MB line within a temporally predicted frame represents a significant percentage of the bits devoted to compressing the whole frame. Nonetheless, a packet containing data from a line of intra-refresh MBs represents a small portion of the image area. Therefore, only a small potential quality penalty arises from the loss of packets containing intra-refresh MBs, due to the small image area affected. The scheme described herein exploits this feature and proposes selectively discarding the intra-refresh MB line data in preference to dropping data belonging to MBs coded in other ways within a video picture.

The proposed scheme combines intra-refresh MB line with FMO in order to group the corresponding intra-coded MBs into a single slice group by means of explicit FMO map slicing [3]. Explicit mapping is one of the seven possible slicing types introduced by FMO, whereby the user can specify which MBs belong to which slice. As the intra-refresh MB line slides down

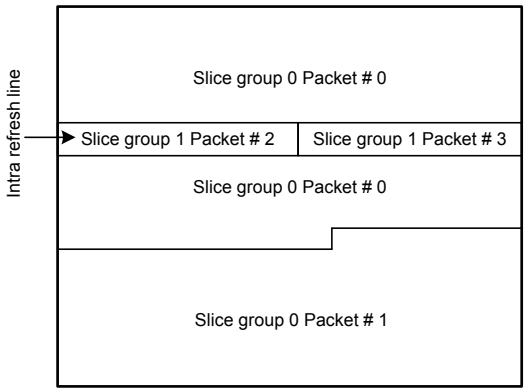

Fig. 2. Slice grouping under FMO with subsequent packetization. Notice that the intra-refresh line is likely to occupy proportionally more packets than the remainder of the compressed data.

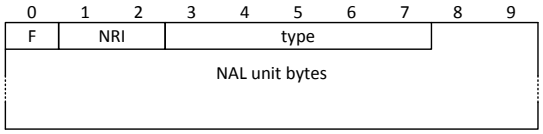

Fig. 3. RTP payload format for H.264/AVC.

one MB position per picture to guarantee a periodic whole picture refresh, an explicit FMO map is sent for each video picture to identify the current position of the intra-refresh line. Although each picture contains only two slice groups, Fig. 2, one with the intra-refresh line and the other with the remaining MBs, each slice group may be split into several packets due to network packetization constraints, which impose a maximum number of bytes per packet. The result is a variable number of packets per picture, the number of which depends on the content complexity and which makes impossible to know in advance which packets will contain intra-refresh lines. As a slice contains resynchronization markers to enable entropy encoding to be re-initialized, it is advisable to configure the codec itself to restrict the packet size rather than allow segmentation at the network level.

B. Packet Prioritization

In order to allow selective packet discard at the network nodes when congestion occurs, it is, therefore, necessary to identify which packets carry intra-refresh MB lines. Decoding the video stream to identify intra-refresh lines is not an option due to the delay introduced. Fortunately, in its network-friendly approach, H.264/AVC, in addition to the core compression algorithms that comprise the Video Coding Layer, also provides the Network Abstraction Layer (NAL) to support network delivery. Each NAL unit output by the NAL can be regarded as a packet prior to encapsulation with a Real-Time Protocol (RTP) header.

The proposed scheme modifies the packet priorities of the NAL unit header in order to simplify the task of a network node. This is accomplished by modification of the *First Byte* in

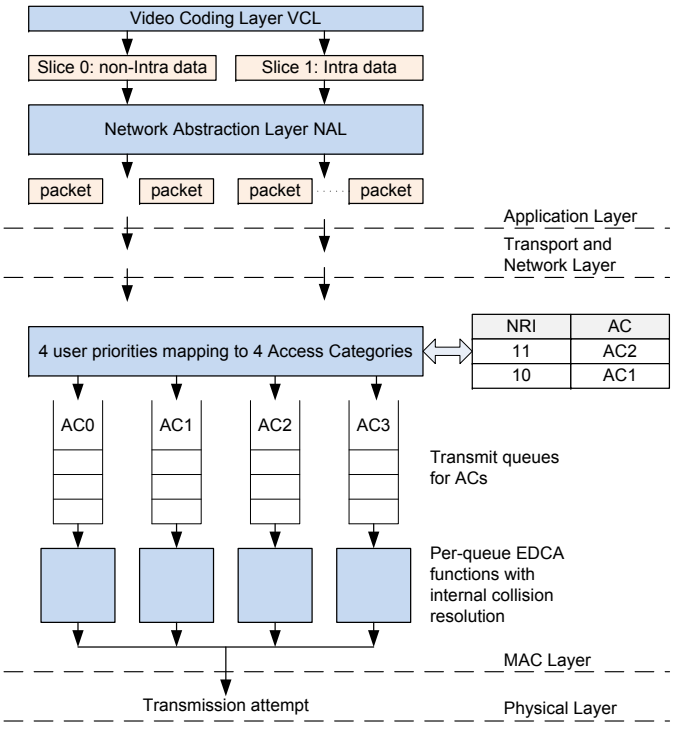

Fig. 4. IEEE 802.11e access category mapping for intra- and non-intra-coded data packets.

the NAL unit packet at the point where the *nal_ref_idc* (NRI) bits, shown in Fig. 3, identify the relative transport priority [10]. A packet carrying an intra-refresh MB line is assigned a lower priority than the others. This is done to make the packet more likely to be discarded in the event of network congestion.

C. IEEE 802.11e EDCA Access Category Mapping

Fig. 4 presents a practical implementation of the proposed prioritization scheme in which video packets are selectively mapped to IEEE 802.11e's *Enhanced Distributed Channel Access* (EDCA) Access Category (AC) queues. High-priority, non-intra-coded data bearing packets are assigned to AC2, as this is the default AC for video traffic. We mapped intra-coded data bearing packets to AC1 reserving AC3 for audio packets. The effect is to delay access to the channel for numerically lower indexed AC queues, thus increasing the probability that packets from these queues will be dropped through buffer overflow when additional packets arrive. The paper now examines the capability of the proposed scheme.

III. Evaluation

A. Simulation Model

In order to evaluate the performance of the proposed scheme, the Common Intermediate Format (CIF) test sequences *Paris* and *Stefan*, with 299 pictures, were coded using the H.264/AVC reference software JM16.1 [11] at 30 frame/s. The encoder was configured to use an IPPP... picture coding structure, RTP packetization, and a Constant Bit Rate (CBR) set to a target of 1 Mbps. FMO was configured with explicit

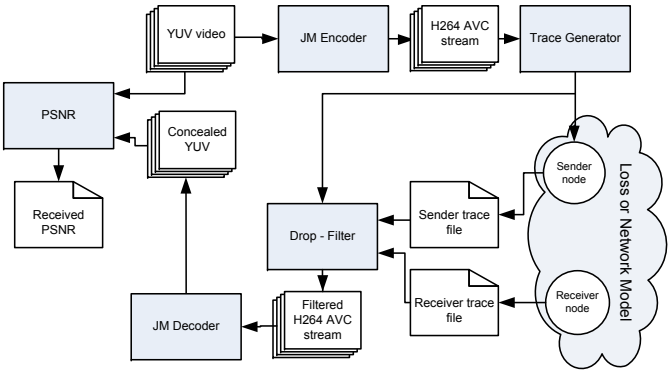

Fig. 5. Framework to obtain the video quality after network simulation.

maps and one intra-refresh MB line per picture. The original encoder was modified to handle the proposed slice grouping of the intra-refresh MB lines. The MB grouping technique splits a picture into one slice containing the intra-refresh MB line and another slice with the remaining MBs.

As a reference for comparison, the sequences were coded in two other ways. The first alternative used the same encoder settings but disabled FMO and replaced the intra-refresh MB line by an equivalent whole picture intra-refresh periodicity of 18 pictures (there are 18 MB lines per CIF picture; hence, the equivalence), i.e., a Group-of-Picture (GOP) size of 18, with intra-coded IDR-pictures inserted every 18 pictures. The second alternative was coded using a single IDR-picture at the beginning followed by predictively-coded P-pictures only, without intra-refresh lines, i.e., the same IPPP.... coding structure with an 'infinite' GOP size but without intra-refresh.

Four different simulations were carried out in order to evaluate the video quality obtained from different packet drop conditions. In two of them, the intra-refresh (IR in the graphs) stream from the proposed scheme was used, while the other two employed streams with GOP sizes 18 and 'infinite' respectively for comparison purposes. In the first set of simulations using the intra-refreshed stream, only packets carrying intra-coded MBs from the refresh MB line were randomly (Uniform distribution) dropped, labeled as 'IR Drop Intra' in the Figures. In the second set of simulations, the same stream was subjected to random packet losses regardless of its priority fields, i.e., any packet can be dropped, 'IR Drop Any' in the Figures. For the two reference sequences (GOP-size 18 and GOP size 'infinite'), packets were also randomly dropped regardless of their priority fields.

Fig. 5 shows the evaluation framework for assessing the impact on video quality. Raw YUV video from the test sequences is encoded to generate a compressed H.264/AVC bitstream which is parsed to form traces giving the packet sizes and transmission schedules. These traces are fed to a loss (network) model. By comparing the sender's trace file with the receiver's trace of video packets received, the dropped packets are found. The dropped packets that represent video slices are removed from the original compressed stream to generate

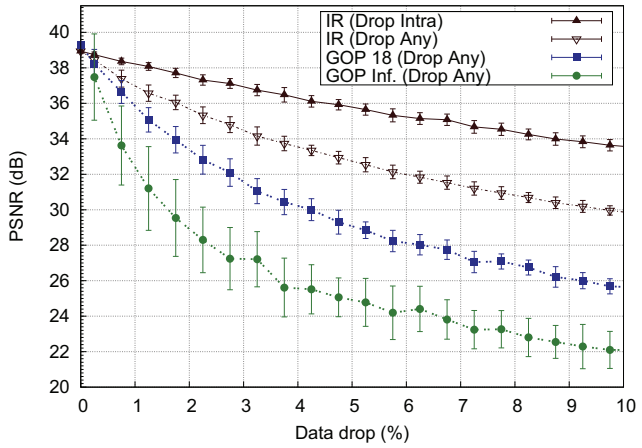

Fig. 6. PSNR vs. percentage video data drop for *Paris*, with standard error.

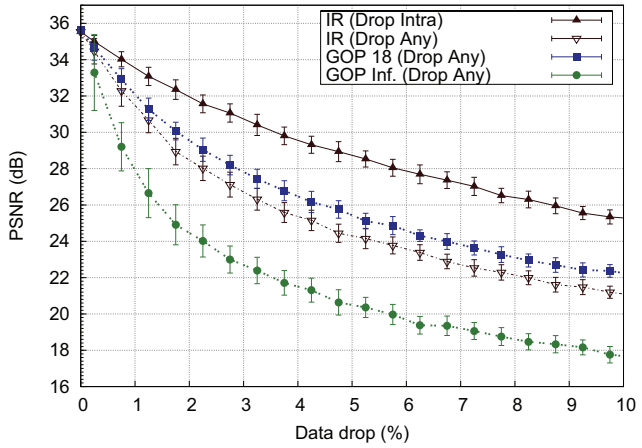

Fig. 7. PSNR vs. percentage video data drop for *Stefan*, with standard error.

the erroneous received stream. After error concealment at the decoder, the PSNR relative to the input YUV video is found.

When decoding the distorted video, frame copy error concealment was used. A total of 2000 runs were used for each set of simulations and the standard absolute Peak Signal-to-Noise Ratio (PSNR) deviations were calculated over intervals of 0.5 % packet drop. To make the results more accurate, the percentage of video data drop has been used instead of the packet drop rate. Since packet size depends heavily on the packet content (headers or transform coefficients), a high packet drop rate might not represent an equivalent high data drop.

B. Results for uniform packet drops

In this Section, network congestion is assumed to cause uniformly distributed packet drops. Arriving video packets are dropped according to the proposed prioritization scheme. In the results, the IEEE 802.11e model of Fig. 4 was not employed, as packets were simply separated into their two priorities and a number were randomly selected to be dropped up to a desired percentage. Figs. 6 and 7 show the video quality

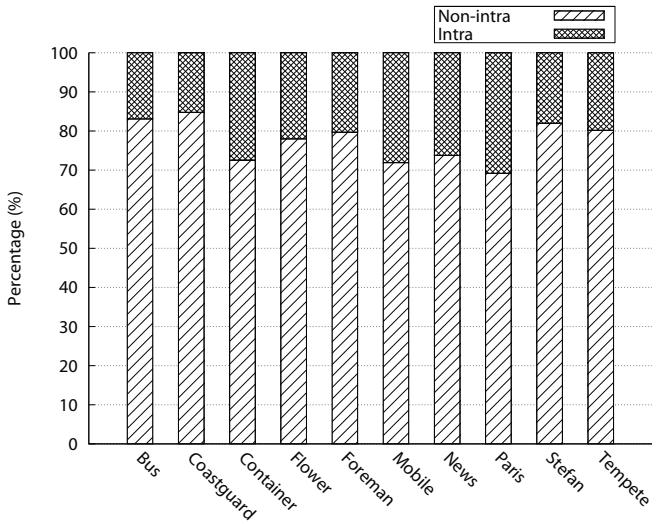

Fig. 8. Percentage of intra- and non-intra coded data in a variety of sequences.

TABLE I
VIDEO QUALITY GAIN FROM PROPOSED SCHEME ACROSS A RANGE OF DIFFERENCE SEQUENCES.

% Data Drop	Quality gain using proposed scheme (dB)									
	bus	coastguard	container	flower	foreman	mobile	news	paris	stefan	tempete
0.25	0.6	0.4	0.1	0.4	0.6	0.5	0.5	0.3	0.5	0.2
0.75	3.2	1.4	0.3	1.3	2.3	1.3	1.7	1	1.7	1.1
1.25	3.2	1.9	0.6	2.2	2.2	2	3.1	1.5	2.4	1.6
1.75	5.3	2.4	0.8	2.7	4.1	2.7	3.2	1.7	3.4	2
2.25	3.9	2.5	1.2	3.5	5	2.8	3.3	2	3.5	2.4
2.75	4.9	3.2	1.2	3.4	4.9	3.4	4	2.3	3.9	2.7
3.25	5.4	3.2	1.4	3.8	4.1	3.8	3.3	2.6	4.1	2.9
3.75	4.9	2.8	1.5	4	5.4	4	4.4	2.7	4.2	3.2
4.25	5.6	3.7	1.8	4.3	5.5	4.2	4.4	2.8	4.2	3.2
4.75	5.6	3.2	2.1	4.1	5.3	4.7	4.7	3	4.5	3.5
5.25	5.1	3.1	2.3	4.4	4.5	4.7	5.4	3.1	4.4	3.5
5.75	4.7	3.3	2.5	4.6	4.9	5	3.7	3.2	4.3	3.5
6.25	5.2	3.3	2.4	4.7	4	5.2	4.9	3.3	4.3	3.7
6.75	4.6	3.3	2.8	4.8	5.2	5.3	5.4	3.5	4.5	3.6
7.25	5.1	2.9	3.1	5	4.3	5.2	5.4	3.5	4.5	3.5
7.75	4.5	3.2	3.1	5.2	4.6	5.5	4.9	3.6	4.2	3.9
8.25	5.3	3.1	3	4.9	4	5.2	4.2	3.6	4.3	3.9
8.75	4	2.8	3.1	5.5	4.3	5.4	5.6	3.6	4.3	4
9.25	4.1	2.6	2.9	5	4.5	5.2	5.1	3.6	4.1	3.9
9.75	4.1	2.6	3.1	5.2	4.4	5.5	5.3	3.7	4.1	3.9

obtained from: (i) dropping packets from the intra-refresh MB slices only (IR-Drop intra); (ii) dropping from any slice (IR-Drop any). The quality obtained from the reference streams, GOP 18 and GOP 'infinite', is also shown.

From these Figures, it is clear that dropping the same percentage of data only from intra-refresh slices has significantly less impact (about 4dB) on the video quality than dropping randomly from any slice. Although such results suggest that intra refresh is not useful in minimizing overall data loss effects, these Figures show that not using any intra refresh at all, i.e., GOP 'infinite', makes the stream much more

TABLE II
IEEE 802.11E MAC PARAMETER VALUES FOR THE IEEE 802.11B RADIO.

Access Category	AIFSN	CW-min	CW-max	TXOP Limit (ms)
AC0 (BE)	7	31	1023	0
AC1 (BK)	3	31	1023	0
AC2 (Video: VI)	2	15	31	6.016
AC3 (Voice: VO)	2	7	15	3.264

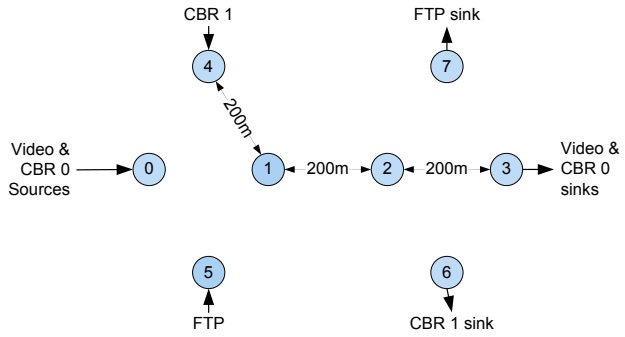

Fig. 9. Network topology used in our simulations.

vulnerable to packet losses.

Fig. 8 shows that an intra-refresh MB line can take up to 30% of the overall bitrate, even though it only represents a single row of MBs per frame. Although a significant portion of the whole bitrate is expended encoding intra MB slices, only a small area of the image is represented by such a high number of bits. Therefore, the penalty in picture quality when those MBs are dropped is low because error concealment is more efficient when small image areas are lost. The above procedures were repeated for many other sequences and the results are shown in Table I. The results show the quality gain when using the proposed scheme for different data drop percentages. The overall conclusion is that the proposed scheme can give quality gains for all the tested sequences.

C. Effect of congestion in a sensor network

In this Section we simulate the behavior of the scheme in the presence of congestion within a wireless sensor network in which one of the sources is providing video surveillance. To isolate the effect of the scheme upon congestion resilience, the wireless channel is assumed to be error-free.

To analyze the performance, we applied our mapping of Fig. 4 to the static scenario shown in Fig. 9, with all adjacent nodes 200m apart. IEEE 802.11e access control was applied on all nodes shown within Fig. 9. The IEEE 802.11b physical layer was modeled with a transmission range of 200 m and a data-rate of 11 Mbps and a basic rate of 1 Mbps. As IEEE 802.11b was in ad hoc mode, the well-known Ad-hoc On-Demand Distance Vector (AODV) routing protocol was deployed with control packets carried at higher-priority AC3.

The CBR video bitrate was set to 1 Mbps and packet size was limited to a maximum of 1 kB. The video under the proposed scheme was assigned either to AC1 and AC2 or no prioritization was used and the video stream was simply

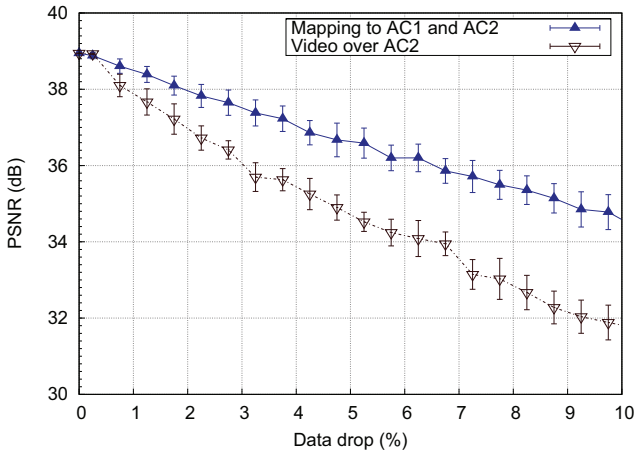

Fig. 10. PSNR vs. percentage video data drop for *Paris*, with standard error.

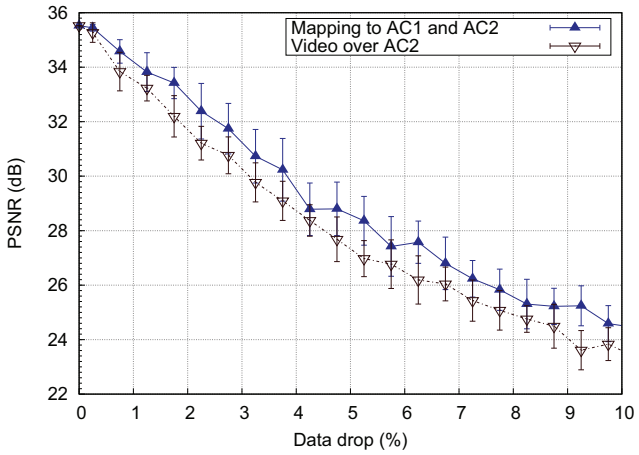

Fig. 11. PSNR vs. percentage video data drop for *Stefan*, with standard error.

assigned to IEEE 802.11e AC2. In these tests, frame copy was used as a simple form of error concealment.

We employed the IEEE 802.11e EDCA *Media Access Control* (MAC) model developed by the Technical University of Berlin [12] for the well-known ns-2 simulator. Table II gives the EDCA parameter values in for the IEEE 802.11b radio.

By repeating simulation runs, this arrangement enabled a wide range of total packet drop rates to be arrived at. Packet drops were totaled across the network path taken by the video from node 0 to node 3. Several CBR sources contributed to the congestion. A CBR traffic source from node 4 to node 6 was assigned to AC1 (to emulate background traffic), using UDP transport with a rate of 128 kbps and packet size of 500 B. Another CBR source from node 0 to node 3 was assigned to AC3 (to emulate voice traffic), also using UDP transport but with a variable rate (48 - 224 kbps) and a packet size of 160 B. The intention of varying the rate was to cause a range of data drop percentages (up to 10%) for the video as it crossed the network path. Additionally, an FTP stream, assigned to AC0 (to emulate best effort traffic), was sent from node 5 to node 7

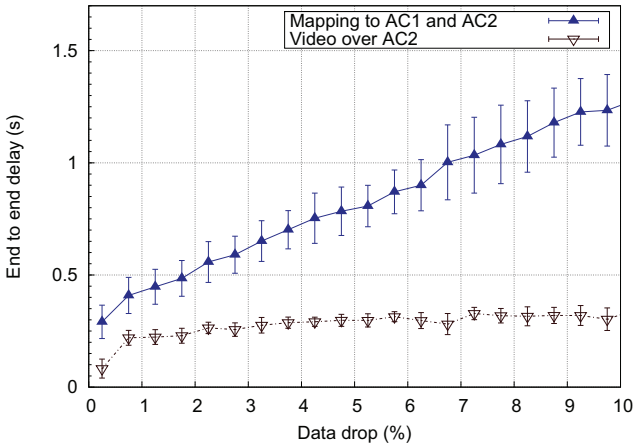

Fig. 12. Mean end-to-end delay vs. percentage video data drop for *Paris*, with standard error.

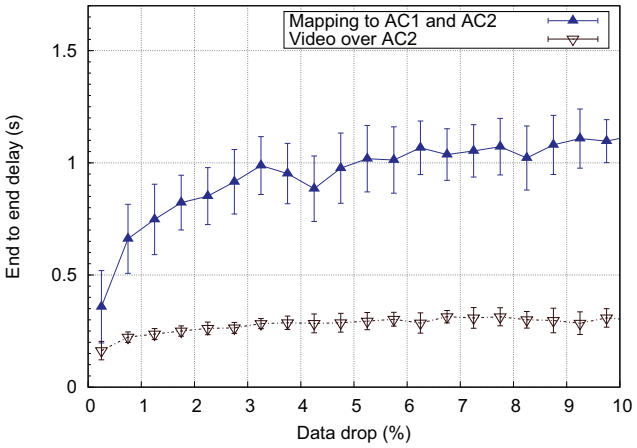

Fig. 13. Mean end-to-end delay vs. percentage video data drop for *Stefan*, with standard error.

with TCP transport, thus contributing to the congestion. 1400 simulation runs were conducted and the resulting packet drops were totaled across the video's network path for each run.

Fig. 10 and 11 show the video quality versus mean percentage video data drop. It is clear that, for both sequences, applying the proposed mapping algorithm will give better quality than when the video stream is entirely assigned to AC2. Fig. 12 and 13 show the end-to-end delay. As might be expected, one penalty of the scheme is that some video packets are assigned to the lower priority AC1 queue. As a result, their queuing time is increased. The increase is greater when there is more congestion, which is also when there are more packet drops.

IV. Conclusion

In this paper, the contrasting goals of congestion resilience and transmission error resilience are explored. The paper seems to show that these two goals may be irreconcilable. In fact, this would be a misreading of the results. It is better to gather together intra-coded MBs in the refresh line and accord this data lower priority (and a higher risk of being dropped from buffers). Nevertheless, the cyclic inclusion of intra-refresh lines does undoubtedly contribute to protection from error propagation. However, it is not necessary for the cycle to be continuous if data of some lines needs to be dropped in the interests of preserving a larger part of the image. Future investigations will quantify the joint effect of channel loss and congestion loss in order to more precisely quantify the trade-offs.

References

[1] IEEE 802.11e, *'Wireless LAN Medium Access Control and Physical Layer specifications Amendment 8: Medium Access Quality of Service Enhancements'*, 2005

[2] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *'Draft ITU-T recommendation and final draft international standard of joint video specification ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC'*, 2003, JVT-G050

[3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, *'Overview of the H.264/AVC video coding standard'*, IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560-576, July 2003.

[4] L. Liu, X.-J. Ye, S.-Y. Zhang, and Y. Zhang, *'H.264/AVC error resilience tools suitable for 3G mobile video services'*, J. of Zhejiang University SCIENCE, vol. 6, no. 4, pp. 1-46, 2005.

[5] A. Ksentini, M. Naimi, and A. Guroui, *'Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture'*, IEEE Communs. Mag., vol. 44, no. 1, pp. 107114, 2006.

[6] S.K. Im, and A.J. Pearman, *'Error resilient video coding with priority data classification using H.264 flexible macroblock ordering'*, IET Image Processing, no. 2, pp. 197-204, 2007.

[7] G. Sun, W. Xing, and D. Lu, *'A content-aware packets priority ordering and marking scheme for H.264 Video over DiffServ Network'* IEEE Asia Pacific Conf. on Circuits and Systems, Nov. 2008, pp. 1735-1738.

[8] T. H. Vu and S. Aramvith, *'An error resilience technique based on FMO and error propagation for H.264 video coding in error prone channels'*, IEEE Intl Conf. on Multimedia and Expo, pp. 205-208, 2009.

[9] 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.

[10] S. Wenger, M.M. Hanuksela, T. Stockhammer, M. Westerlund, and D. Singer, *'RTP payload format for H.264 video'*, IETF RFC 3984, Feb. 2005.

[11] JVT Reference Software Version JM16.1 Online *http://iphome.hhi.de/suehring/tml/download/old_jm/jm16.1.zip*

[12] An IEEE 802.11e EDCA and CFB Simulation Model for ns-2. Online *http://www.tkn.tu-berlin.de/research/802.11e_ns2*