PRIORITIZED PACKETIZATION FOR VIDEO WITH INTRA-REFRESH MACROBLOCK LINE

Ismail Ali, Sandro Moiron, Mohammed Ghanbari, Martin Fleury

University of Essex, United Kingdom
{iaali, smoiro, ghan, fleum}@essex.ac.uk

ABSTRACT

When transmitting video over low bandwidth networks a cyclic intra-refresh (IR) line is preferable to periodic intra frames as an error mitigation technique. This paper shows that IR lines introduce non-uniform error sensitivity among regions within a picture that are identifiable by the position of the IR line. To exploit this characteristic, a packet prioritization scheme is proposed in which priorities are assigned on a packet-level basis. Thus, higher priority packets can be given better protection. The proposed scheme is demonstrated for prioritized access to a congested IEEE 802.11e network. Experimental results indicate that, when compared with the classic non-prioritized packet access scheme, the proposed scheme can achieve quality gains of up to 3.6 dB for a 10% data loss rate.

***Index Terms*—** intra-refresh, access control, IEEE 802.11e, packet prioritization.

1. INTRODUCTION

Increasingly, video is streamed over bandwidth-limited, potentially congested wireless networks such as sensor networks [1]. Because of the predictive nature of video coding, temporal error propagation is likely to occur whenever packets are dropped. To halt temporal error propagation, periodic intra-coded frames (I-frames) are commonly inserted into compressed video streams. This allows the decoder to resynchronize with the encoder once an I-frame is received. However, when accessing network, I-frame packets can cause peaks in the data rate due to the lower compression efficiency of the spatial prediction (intra), compared to temporal prediction (inter). Such bursts can be problematic for bandwidth-limited networks. Therefore, intra-refresh (IR) macroblocks (MBs) can alternatively be spread across a picture sequence.

One practical way to accomplish this is by the cyclic insertion of an IR line of MBs on a per-picture basis so that the whole frame is refreshed after all the MB rows have been refreshed. However, a cyclic IR line introduces an unequal error sensitivity within packets from the same picture. Namely a packet containing data from recently intra-refreshed MBs should be better protected against drops due to their important role as a reference for upcoming pictures. On the other hand, MBs waiting to be refreshed in the short term represent a small threat to error propagation as they are soon to be intra-coded. Hence, it is clear that there are three regions of unequal error sensitivity within the intra-refreshed picture bounded by the intra-refresh line itself.

In [2], authors found that cyclic IR lines are less error sensitive than the remaining parts of the frame. This paper goes beyond the work in [2] and identifies more regions with distinct error sensitivities. Thus, the contribution of this paper is firstly to introduce a way of prioritizing the compressed video stream by the position of the IR line and secondly to demonstrate how the concept can effectively be used to counteract the effects of congestion. The scheme is applicable whenever unequal protection can be used.

One of the main advantages of the proposed scheme, when compared with other schemes, is its simplicity. The scheme can be applied with negligible computational increase, making it suitable for real-time video applications on low-powered devices. Additionally, the proposed scheme does not require Flexible Macroblock Ordering (FMO) [3], allowing therefore its use in both the widely-adopted H.264/AVC (Advanced Video Coding) standard's 'Main Profile' or the lightweight 'Constrained Baseline Profile'. Notice also that the proposed scheme is an alternative to periodic I-frames, as it limits error propagation within a fixed number of frames while other schemes do not provide this guarantee.

In [4], an adaptive IR scheme required the encoder to keep track of which parts of the image area were recently refreshed. The encoder would then refresh those MBs which had more impact on error propagation. Alternatively, [5] proposed a scheme in which FMO is combined with adaptive MB grouping. Because such schemes work at an individual MB level, they significantly increase the computational complexity and memory footprint arising from the required video content analysis. Moreover, methods using 'explicit' FMO imply sending additional packets for MacroBlock Allocation maps (MBAmaps) for every frame, which additionally increase the bitrate and degree of inter-packet dependency.

To test the effectiveness of the proposal, IEEE 802.11e is applied in the context of a bandwidth-limited sensor network

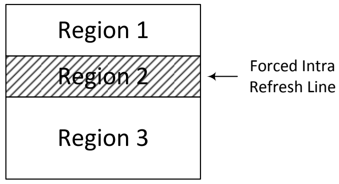

Fig. 1. Picture regions when using cyclic IR line.

subject to cross-traffic that causes congestion. For an assessment of the use of IEEE 802.11e to support real-time services in sensor networks refer to [6]. IEEE 802.11e Enhanced Distributed Channel Access (EDCA) [7] adds Quality-of-Service (QoS) support to legacy IEEE 802.11 wireless networks by introducing four Access Categories (ACs), viz. AC0, AC1, AC2, and AC3 for background (BK), best-effort (BE), video (Vi) and voice (Vo) traffic respectively in order of increasing priority. However, mapping video packets according to their priority across the 802.11e EDCA ACs is an effective alternative to sending the complete video stream over AC2, as it helps to counter the self-congestion within the AC2 queue that would otherwise lead to higher priority packets being dropped. In [8], priority sets and data-partitioned data were mapped to different ACs to improve the video quality. Work in [9] also combined video layering (but through the scalable video coding extension to H.264/AVC) with IEEE 802.11e EDCA allocation but none of these works created prioritized packets according to their intra-refresh status.

The H.264/AVC codec [10] was used due to its high compression efficiency relative to earlier standard codecs. As part of its network-friendly approach, H.264/AVC was designed to separate video compression from network delivery. Hence, all video-related tasks are handled by the Video Coding Layer (VCL) while network adaptation issues are handled by the Network Abstraction Layer (NAL). The content arriving from the VCL layer is encapsulated into NAL units (NALu). A NALu consists of a one byte header followed by the corresponding payload information. Subsequently, each NALu is encapsulated in an RTP header prior to output.

Section 2 demonstrates the basis of the proposed slicing scheme that exploits the unequal error sensitivity introduced by the IR line. Section 3 shows a refinement of the initial scheme according to the dynamic positioning of the IR line. Section 4 presents the implemented IEEE 802.11e access control mechanism to demonstrate the proposed packet-wise prioritization scheme and the results are presented in Section 5. Section 6 makes some concluding remarks.

2. ENHANCED SLICING SCHEME

2.1. Impact of intra refresh for error robustness

Including cyclic IR lines can be preferable to the periodic insertion of I-frames in bandwidth limited networks due to the smoother rate distribution of the coded stream. However, a side effect is an implicit division of the video frame into three regions in terms of the potential for temporal error propagation as represented in Fig. 1. The first region is located above the IR line and is the most error-sensitive region. This is because this region has recently been refreshed with intra-coded blocks and, thus, will likely play an important role as a reference for the following frames, without the prospect of being refreshed in the near future. As a result, an error affecting this region can propagate for a long period, introducing a significant loss of video quality. The second region is the IR line itself. Introducing a single IR line [2] plays an important role in mitigating error propagation. However, this line uses a significant portion of the bitrate share, while still representing a small portion of the image area (just 22 MBs of the 396 total MBs in a Common Intermediate Format (CIF) picture). Therefore, it was shown in [2] that if packets are to be discarded, packets carrying IR data introduce the least quality impact for the same percentage of video data loss. Hence, they are the best candidates to be discarded. Finally, the last region of the frame is located below the IR line. This region has the smallest potential error impact because an error in this region does not propagate for long. If the IR line is cycling from top to bottom, those MBs below the IR line will be intra coded in the following pictures. Hence, even if errors occur in this region they will be cleaned soon, limiting the number of pictures that are affected by prediction from the distorted reference slices. The same principle applies to a vertical IR column cycling horizontally. Overall, it can be said that cyclically intra refreshed video frame has three regions with unequal impact on error propagation.

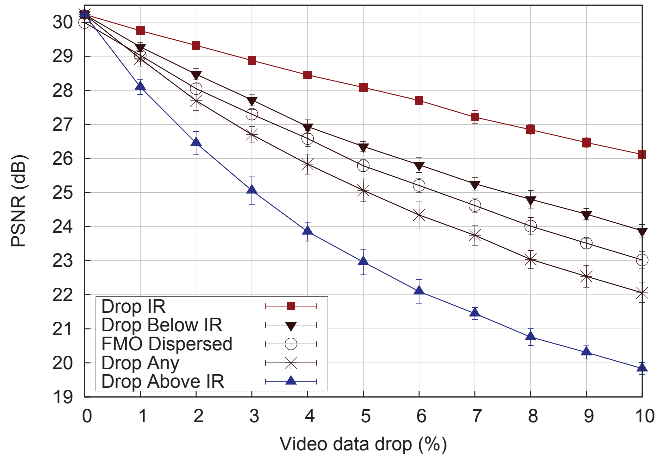

Fig. 2. Quality impact of dropping different picture regions for the *Mobile* sequence.

Figure 2 shows the unequal impact on video quality of each region by selectively dropping packets from different regions of the video picture, with increasing levels of random (but uniformly distributed) packet losses. 300 CIF frames from the *Mobile* sequence were encoded using the JM 16.1 [11] implementation of H.264/AVC. The 'Main Profile' was used, with a single I-frame followed by all P-frames and Context Adaptive Binary Arithmetic Coder (CABAC) entropy encoding at a constant bitrate of 1 Mbps. RTP packetization mode was set with a maximum packet size of 1 kB.

Five tests were conducted. In the first test, only packets belonging to the IR line slice were dropped (Drop IR). For the second test, only packets belonging to the region above the IR line were dropped (Drop Above IR) and for the third test those of the region below the IR line were dropped (Drop Below IR). The fourth test, used as a benchmark for comparison, packets were dropped randomly from all regions (Drop Any). For the last test, the *Mobile* sequence was encoded with the *Dispersed* mode (also known as Checkerboard) of FMO error resilient technique and then subject to random drops as in 'Drop Any'. In Fig. 2, each point is the average video quality assessed over 20 runs and the error bars represent the standard deviations from the mean. Motion copy error concealment technique was used to reproduce the lost slices.

Fig. 2 shows that dropping packets from Region 1 (Above IR) introduces a significant quality drop (up to 2 dB) when compared with the benchmark of dropping randomly. This is due to the fact that Region 1 was recently refreshed with intra coded blocks and therefore any errors occurring in this region are to be propagated for several frames until these blocks are intra coded again. Dropping packets from Region 2 (Drop IR) shows the least impact on video quality (4 dB better than the benchmark). Although IR lines play an important role in mitigating error propagation, these intra coded blocks are very costly in terms of data rate (20%), while only representing a small portion of the image area (5% for a CIF sequence). Dropping packets from Region 3 introduce a moderate quality gain (up to 2 dB) when compared with the benchmark. Overall, Fig. 2 illustrates that each region presents a very distinct quality impact for the same percentage of data loss. On the other hand, FMO dispersed scheme results in a quality gain of up to 1 dB when compared with the benchmark due to implicit improvements at the error concealment stage. However, this gain is significantly lower than the potential gain obtained by selectively dropping packets from low importance regions. Similar results (unreported herein for reasons of space) were obtained for other test sequences. In order to exploit this unequal packet importance, a new slicing scheme and a packet identification scheme are required to allow the identification of the region where each packet belongs to.

2.2. Slicing issues

In H.264/AVC, each frame is decomposed into MBs which are then grouped into slices. Each slice is independently decodable by the arithmetic decoder, thus acting as a resynchronization unit that prevents error propagation to the entire picture. When the coded frame size exceeds the maximum transport unit (MTU) supported by the underlying transport layer, slicing can be used to divide the coded frame into slices each

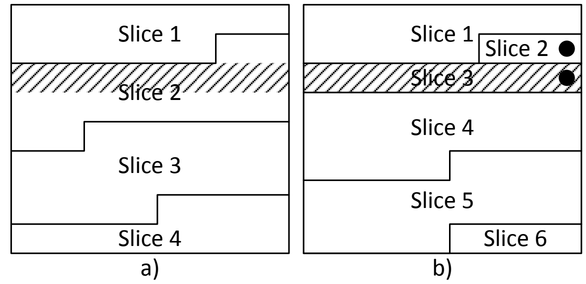

Fig. 3. a) Classic slicing b) Proposed enhanced slicing.

of which packed into a single packet smaller than MTU to prevent any further network segmentation.

During this process it is likely that some slices could contain MBs from more than one image region, as illustrated in 'Slice 2' in Fig. 3a. To avoid this and in order to slice an image into three distinct regions according to Fig. 1, the slice structuring procedure of the encoder was modified to accommodate a new set of rules for slicing. These rules prevent packing mixtures of MBs from different regions into the same slice by deciding when the current slice should be terminated or if more MBs are to be packed in. In addition to the original maximum packet size criterion, two new rules were introduced to terminate slices:

- If the current MB does not belong to the IR line and the following MB does, terminate the slice.
- If the current MB belongs to the IR line and the following MB does NOT, terminate the slice.

Figure 3b shows the check points (black dots) tested by the two new rules added to the slicing mechanism.

2.3. Identifying regions

Section 2.2 described a slicing scheme that prevents the mixture of macroblocks from different regions such that a different treatment can be applied to each region's packets. However, after packetizing the video into RTP packets [12], IR line packets are no longer identifiable in the bitstream, making it impossible to separate packets according to their associated regions. Thus, the transport network is not able to distinguish which packets should be given higher protection, which renders useless the previous slicing scheme.

Therefore, this section depicts a standard-compliant signaling strategy that allows a network layer to easily identify the importance of each packet. The H.264 NAL header defines two bits to indicate the relative importance of a video packet for decoding. However, most 2-bit codes are already reserved leaving the only free code used for low priority packets in data-partitioned video. The challenge here is how to identify packets from three different regions with only two different combinations.

The solution found is an implicit signaling technique. This technique starts by setting the low-priority NRI code to IR line packets (Region 2). This modification by itself only

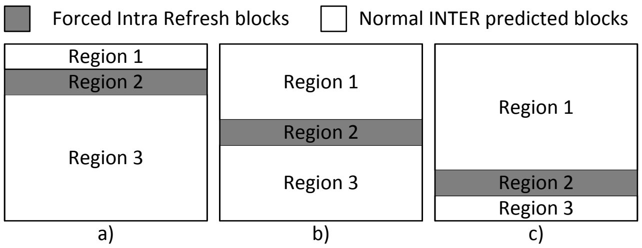

Fig. 4. Region size variation with position of cyclic IR line.

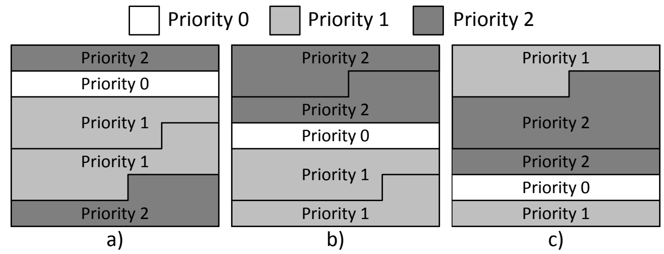

Fig. 5. Packet level priority based on the IR line closeness.

enables the identification of IR packets while Region 1 and Region 3 cannot yet be identified. On the network side, an implicit correspondence between the packet and region it belongs to is possible by analyzing the RTP's `TimeStamp` and packet number as well as the NALu NRI field. A new picture is detected by a change in the RTP `TimeStamp` field. Region 1 can be defined by all packets arriving: 1) after a `TimeStamp` change (meaning that a new picture is received) and 2) before receiving packets with the NRI low-priority code. Then, Region 2 consists of packets with the NRI low-priority code. Finally, Region 3 contains packets without the NRI low priority code, received after Region 2 packets but before a new `TimeStamp` is detected (new picture).

3. PACKET-WISE PRIORITIZATION

During the IR cycle, the IR line can be close to the center of the image, Fig. 4b, and the sizes of Regions 1 and 3 are close. As the IR approaches the bottom of a picture, Fig. 4c, the size of high-priority Region 1 grows and the size of lower-priority Region 3 shrinks. Consequently, disproportionately more high-priority packets are created, increasing the risk of self-congestion if these higher-priority packets are all directed to a high-priority access queue. (As a slice is packed into a single packet, one may therefore use packet to refer to a packed slice.) Additionally, as the IR line approaches the bottom of a picture, those slices at the top of the image become closer to being refreshed. Consequently, they are less deserving the high-priority status. The situation is reversed when the IR line is at the top of the picture, Fig. 4a, where more lower-priority packets are created than high priority packets.

To address this situation, we have developed an enhanced packet prioritization scheme which takes into account the potential quality impact in case of errors. Instead of simply defining Region 1 as high priority and Region 3 as lower priority, the proposed scheme defines three priorities on a packet level according to their proximity to the next refresh cycle. The lowest priority packets bearing the IR line slice remain unchanged. The overall effect is to change the prioritization scheme from region based to packet based according to the relative distance to the IR line. These changes mean that the method of identifying the packet's priority had to be modified, as a packet's importance is no longer defined by its position within the frame (above and below the IR line) but defined by its closeness to the next refresh cycle as shown in Fig. 5.

The packet priority scheme proposed in this paper defines three different priorities. The lowest priority (Pri0) is assigned to packets carrying IR data as in the previous scheme described in Section 2.

For the two remaining priorities, a temporary list is created containing packets below the IR line (Region 3) followed by packets above the IR line (Region 1), all with the same TimeStamp. Afterwards, medium priority (Pri1) is assigned to the first half (an integer division by two is performed) of packets and the remaining packets are assigned high priority (Pri2) as shown in Fig. 5c.

4. SIMULATION FRAMEWORK

Mapping video stream packets to multiple ACs according to their priorities is an effective alternative to placing all video packets in AC2 [9]. In fact, according to the packet-based prioritization scheme of Section 3, packets are placed in IEEE 802.11e ACs from lowest to highest priority in AC0, AC1 and AC2 respectively. In summary, packets bearing IR data are sent over AC0 (Prio0), packets to be refreshed soon (Prio1) are sent over AC1 and packets containing data which were recently refreshed (Prio2) are sent over AC2. IEEE 802.11e Medium Access Control (MAC) layer architecture is illustrated in Fig. 6, which shows the associated queues (set to 40 variable-sized packets in tests) with entry to the queues defined by a mapping function. Should several packets emerge simultaneously from the queues then contention is resolved by the virtual collision handler before a transmission attempt. Each AC has different Distributed Coordination Function (DCF) parameters for the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) back-off mechanism. In tests (Section 5), the default IEEE 802.11e MAC parameter values [7] for the IEEE 802.11b radio were employed. Packets mapped to numerically lower AC queues are delayed before accessing the channel, thus increasing the probability that they will be dropped through buffer overflow due to arriving packets. 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.

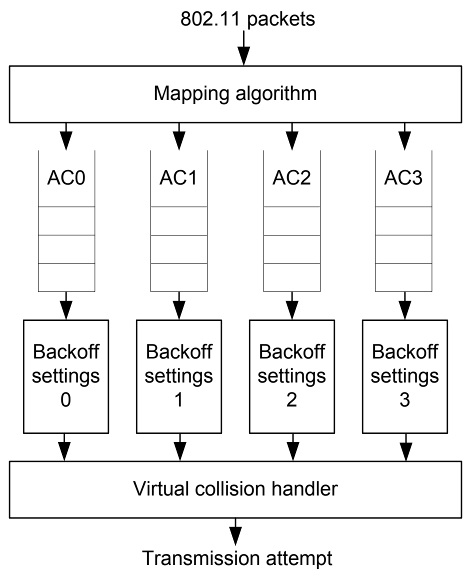

Fig. 6. IEEE 802.11e MAC layer architecture.

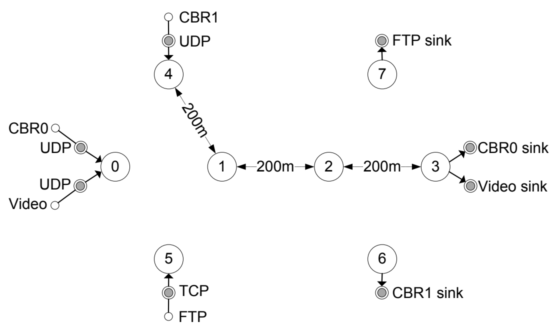

Fig. 7. Simulated IEEE 802.11b test scenario.

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed unequal packet priority scheme we encoded five sequences (*Akiyo*, *Foreman*, *Mobile*, *Paris*, and *Stefan*) with the video configuration already described in Section 2. The performance was measured in terms of Peak Signal-to-Noise Ratio (PSNR) for video data drop percentages ranging from 0 to 10%.

For each sequence, two tests were conducted. In the first test (Default), used as a reference for comparison purposes, all packets were assigned to AC2 (the default IEEE 802.11e AC defined for video [7]). For the second test (Proposed), the proposed priority scheme split packets over AC0, AC1 and AC2 according to their priority.

We employed the IEEE 802.11e EDCA MAC model developed by the Technical University of Berlin [13] for the NS-2 simulator. For each test, more than 200 independent simulation runs with differing packet drop percentages were performed. To analyze the performance, we applied our mapping to the static scenario shown in Fig. 7, with all adjacent nodes 200 m apart. The IEEE 802.11b physical layer was modeled with a transmission range of 250 m and a data-rate of 11 Mbps. The contributing traffic sources for this scenario and their mapping to different ACs are shown in Table 1.

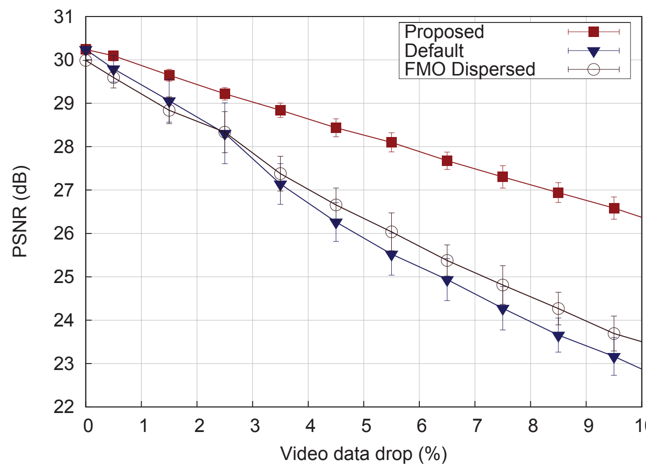

Fig. 8. PSNR versus percentage video data drop for *Mobile*.

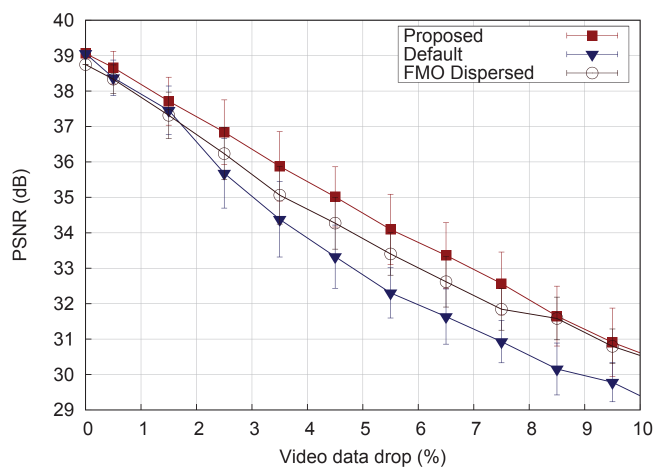

Fig. 9. PSNR versus percentage video data drop for *Foreman*.

Figure 8 shows the tests results for the *Mobile* test sequence. It is shown that when less important packets are assigned to lower priority ACs (AC0 and AC1) a significant quality gain (over 3.5 dB) can be achieved when compared with the reference scheme where all packets are sent over AC2. This is due to the fact that unloading AC2 from lower importance packets results in reduced congestion in this AC and lower packet drops for high-priority packets. Compared with (FMO's *Dispersed*) mode (also sent over AC2), the proposed scheme still presents a significant quality gain (over 3 dB). Figure 9 shows the case for the *Foreman* test sequence.

Although presenting moderate quality gains (around 2 dB) the proposed scheme shows a consistent superiority along the whole data drop range when compared with the reference scheme. It should be emphasized that FMO's *Dispersed* mode reduces the initial video quality when there is no data drop (0%) due to its implicit lower coding efficiency.

Table 1. Traffic in the scenario of Fig. 7.

Traffic	Source	Destination	AC	Transport	kbps
CBR0	0	3	3	UDP	0-192
Video	0	3	0, 1, 2	UDP	1000
CBR1	4	6	1	UDP	128
FTP	5	7	0	TCP	

Table 2. PSNR gain of the proposed prioritized access over the default access for several test sequences and loss rates.

Data Drop (%)	PSNR gain (dB)				
	Stefan	Paris	Mobile	Foreman	Akiyo
1	0.6	0.5	0.5	0.3	0.6
2	0.9	0.8	0.7	0.7	0.9
3	1.5	0.9	1.3	1.4	0.8
4	1.7	1.2	1.9	1.6	1.1
5	2.1	1.6	2.4	1.8	1.5
6	2.2	1.8	2.7	1.7	1.3
7	2.1	1.9	2.9	1.7	1.5
8	2.4	2.1	3.1	1.6	1.8
9	1.9	2.4	3.4	1.2	1.7
10	2.1	2.2	3.5	1.2	1.9

Similar results were obtained for the remaining test sequences as shown in Table 2. Overall, the proposed algorithm achieves better results for high motion activity sequences such as *Mobile* and *Stefan* than for less active ones like *Akiyo* and *Foreman*. This is due to the better performance of error concealment when sequences have low motion activity.

6. CONCLUSIONS

This paper has shown that when cyclic intra-refresh lines are used as an error mitigation technique, an unequal error sensitivity is created within packets from the same frame. An unequal packet prioritization scheme was presented and demonstrated for an IEEE 802.11e QoS access controlled scenario. Experimental results presented consistent quality gains (up to 3.6 dB) for multiple video sequences and over a wide range of video data loss percentages (0-10%).

Unlike previous attempts, the current implementation does not introduce any computational or bitrate increase, being therefore especially suitable for low-powered devices where energy is a serious constraint.

7. REFERENCES

[1] S. Misra, M. Reisslein, and G. Xue, "A survey of multimedia streaming in wireless sensor networks," *IEEE Communs. Surveys & Tutorials*, vol. 10, no. 4, pp. 18 –39, 2008.

[2] S. Moiron, I. Ali, M. Fleury, and M. Ghanbari, "A prioritization scheme for H.264/AVC video congestion resilience with IEEE 802.11e," in *6th Conf. on Wireless Advanced*, 2010, pp. 1 –6.

[3] P. Lambert, W. De Neve, Y. Dhondt, and R. Van de Walle, "Flexible macroblock ordering in H.264/AVC," *J. Visual Commun. and Image Representation*, vol. 17, no. 2, pp. 358 – 375, 2006.

[4] J. Jiang, B. Guo, and W. Mo, "Efficient intra refresh using motion affected region tracking for surveillance video over error prone networks," in *Int. Conf. on Intelligent Syst. Design and Applications*, 2008, vol. 2, pp. 242 –246.

[5] K. Tan and A. Pearmain, "An FMO based error resilience method in H.264/AVC and its UEP application in DVB-H link layer," in *IEEE Int. Conf. on Multimedia and Expo*, 2010, pp. 214 –219.

[6] L. Bononi, L. Donatiello, and M. Furini, "Real-time traffic in ad-hoc sensor networks," in *IEEE Int. Conf. on Commun.*, 2009, pp. 1 –5.

[7] "Wireless LAN medium access control and physical layer specifications amendment 8: Medium access quality of service enhancements," *IEEE Std 802.11e*, 2005.

[8] A. Ksentini, M. Naimi, and A. Gueroui, "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. 107 – 114, 2006.

[9] I. A. Ali, M. Fleury, and M. Ghanbari, "Congestion-resistant scalable media stream mapping for an IEEE 802.11e sensor network," in *IEEE Int. Conf. on Image Proc*, 2010, pp. 1–5.

[10] 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, 2003.

[11] JVT Reference Software Version JM16.1, "http://iphome.hhi.de/suehring/tml/download/," .

[12] S. Wenger, M.M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, "RTP Payload Format for H.264 Video," RFC 3984 (Proposed Standard), February 2005.

[13] An IEEE 802.11e EDCA and CFB Simulation Model for NS-2, "http://www.tkn.tu-berlin.de/research/802.11e_ns2," .