Broadband Video Streaming with Built-in Resiliency

I. Ali, S. M. Saleh Al-Majeed, S. Moiron1, M. Fleury, and M. Ghanbari School of Computer Science and Electronic Eng.
University of Essex
Colchester CO4 3SQ, United Kingdom
{iaali, ssaleha, smoiro, fleum, ghan}@essex.ac.uk

ABSTRACT: Mobile TV services are being actively developed for avariety of last hop, broadband wireless technologies. Application layer error control mechanisms such as Broadband Video Streaming seek to reduce packet loss from raw UDP transport. This paper goes further than existing streaming protocols by integrating source-coded error resilience through data-partitioning and intra-refresh macroblocks with the error control mechanism. Results show that for a temporally complex sequence, up to 6.23 dB gain in video quality (PSNR) can result, depending on burst error lengths across an IEEE 802.16e link.

Keywords – error resiliency; H.264/AVC; mobile TV; streaming protocols

I. INTRODUCTION

As mobile TV services proliferate, whether the delivery is by DVB-H [1], Long Term Evolution (LTE) or iMAX, robust ways of streaming video that can resist the congestion are sought. For IEEE 802.16e (mobile WiMAX) [2], capacity studies [3] suggest up to 16 mobile TV users per cell in a 'lossy' channel depending on factors such as the form of scheduling and whether MIMO is activated. True video streaming rather than progressive download or filecast, enables fast channel zapping [1] but requires error control when delivered from a remote server in the network core. To further resist error bursts, typically arising from slow fading or interference from neighbouring cells, source-coded error resiliency [4] is a promising alternative to application-layer forward error correction (FEC), as it does not overlap existing PHYsical layer FEC. This suggests that these two requirements can be combined in a way that integrates source-coded error resilience with a broadband video streaming (BVS) protocol.

BVS is a simple, single retransmission scheme [5] aimed at improving IPTV video streaming. Because it employs a single Negative Acknowledgement (NACK), it is most suitable for situations where the roundtrip time is not too long. In [6], ways to improve IPTV quality were discussed with the assumption that intelligent content management would bring popular video content nearer to the end viewer. As the server is closer to the access network, it is possible that a retransmission decision can be made at the server in response to short-term changes in the wireless channel. This paper introduces restricted retransmission to BVS in order to reduce the overhead arising from retransmission of all data. The central idea is to retransmit partition-A of an H.264/AVC

(Advanced Video Coding) data-partitioned compressed video stream [7] without retransmitting partitions B and C, which contain lower-priority data as far as decoder reconstruction of a video picture is concerned. In particular, partition-A contains the motion vectors (MVs), which in predictive coding identify replacement macroblocks (MBs) in correctly received reference pictures. This allows motion-copy error concealment at the decoder to partially reconstruct the picture without the texture data (transform coefficient residuals) contained in partition-C.

In [8] for an IEEE 802.16d live testbed, directly-applied UDP-transported streaming was implemented. While this approach may be feasible if the video source is at the base station (BS), it may not be adequate if the 'video hub office' [6] is placed on a core network feeding the BS via a 'video serving office'. Congestion can arise both on this network and from mobile subscriber stations (SSs) on the WiMAX access network. TCP-Friendly Rate Control (TFRC) has been proposed for LTE [9], provided packet losses are disguised from the application by repeated transmissions at the data-link layer. In general, in single connection TFRC, wireless channel packet loss is misinterpreted as congestion, causing the congestion controller to reduce its sending rate, resulting in poor utilization and lengthened streaming periods. The problem with data-link retransmissions is that delay management is taken out of the hands of the video streaming application designer. To further increase wireless channel utilization, multi-connection TFRC [10] can be considered. The main difficulty with this approach may be establishing the number of connections needed according to fluctuating channel conditions. In [11], cross-layer intervention occurred to mask channel packet loss from TFRC. This approach is most suitable for dedicated networks and not ones in which mixed traffic is present. This is because a BS must make a privileged intervention for one class of traffic. In [5], the more direct approach of BVS was shown to outperform the TFRC-based methods provided a single retransmission took place. An advantage that TFRC-based techniques present is that through DCCP [12] they are re-enforced by an industry-level standard.

Section II introduces data-partitioning in H.264/AVC and Section III outlines BVS-A, the version of BVS that includes restricted retransmission, while Section IV presents the simulation model for Section V's evaluation. Section VI concludes this paper.

II. ERROR RESILIENCY SCHEME

The H.264/AVC codec conceptually separates the Video Coding Layer (VCL) from the Network Abstraction Layer (NAL). The VCL specifies the core compression features, while the NAL supports delivery over various types of network. Table I is a summarized list of different NAL unit types. NAL units 1 to 5 contain different VCL data that will be described later. NAL units 6 to 12 are non-VCL units containing additional information such as parameter sets and supplemental information.

In the H.264/AVC codec, each video picture can be divided into one or more independently decodable slices; each of which contains a flexible number of MBs. The slices of an Instantaneous Decoding Refresh- (IDR-) are located in type 5 NAL units, while those belonging to a non-IDR (I- P- or B-pictures) are placed in NAL units of type 1, and in types 2 to 4 when data partitioning mode is active, as now explained. In type 1 NALs, MB addresses, motion MVs and the integer transform (IT) coefficient residuals of the blocks, are packed into the packet, in the order they are generated by the encoder. In type 5, all parts of the compressed bit stream are equally important, while in type 1, the MB addresses and MVs are much more important than the IT coefficients.

In H.264/AVC when data partitioning is enabled, every slice can be divided into three separate partitions and each partition is located in either of type-2 to type-4 NAL units, as listed in Table I. A NAL unit of type 2, also known as partition-A, comprises the most important information of the compressed video bit stream of P- and B-pictures, including the MB addresses, MVs, and essential headers. If any MBs in these pictures are intra-coded, their IT coefficients are packed into the type-3 NAL unit, also known as partition B. Type 4 NAL, also known as partition-C, carries the IT coefficients of the motion-compensated inter-picture coded MBs. 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, constrained intra prediction should be set [13] to make partition-B independent of partition-C. By setting this option, partition-B MBs are no longer predicted from neighbouring inter-coded MBs, the prediction residuals of which reside in partition-C.

This work employs a Group of Pictures (GOP) structure of IPPP…, which reduces the computational complexity involved in reconstructing bi-predictive B-pictures. The error resiliency scheme compensates for the absence of periodic intra-refresh pictures (except at the start of a sequence) by the periodic insertion of intra-refresh MB lines in a cyclic pattern within successive temporally predicted P-pictures. The aim of inserting intra-refresh MB lines [14] is to mitigate error propagation at a small cost of lower coding efficiency than all predictive inter coding.

III. OPERATION OF BVS-A

After observing that UDP at least succeeds in good wireless channel utilization in contrast to single connection TFRC (refer to Section I), without any protection from channel loss, the simple BVS scheme introduces a single NACK to UDP.

TABLE I. NAL UNIT TYPES

NAL unit type	Class	Content of NAL unit
0	-	Unspecified
1	VCL	Coded slice
2	VCL	Coded slice partition A
3	VCL	Coded slice partition B
4	VCL	Coded slice partition C
5	VCL	Coded slice of an IDR picture
6-12	Non-VCL	Suppl. info., Parameter sets, etc.
13-23	-	Reserved
24-31	-	Unspecified

Fig. 1 is a general representation of the processing involved, showing the NACK response of the receiver. The following describes the operation assuming downlink streaming from a server (node C in Fig. 1) to a mobile SS. At a mobile SS a record is kept of packet sequence numbers available through the RTP header and, if an out of sequence packet arrives, a NACK may be transmitted to the BS in the next sub-frame for forwarding to the video server. The SS only transmits a NACK if this is the first time that particular packet has been requested. To reduce the overhead at the SS, the decision as to whether to retransmit a packet is left to the server. The reorder buffer at the SS is a playout buffer which may change the sending order of video data to fit the decode order. The server prevents transmission from its input buffer until a single retransmission of the missing packet in the sequence has taken place. Not shown in Fig. 1, is a holding buffer that retains sent packets in the case of the need for a retransmission. However, the server will only retransmit if the NACK refers to a partition-A NAL bearing RTP packet is requested. This type of packet is the type-A packet referred to in Fig. 1.

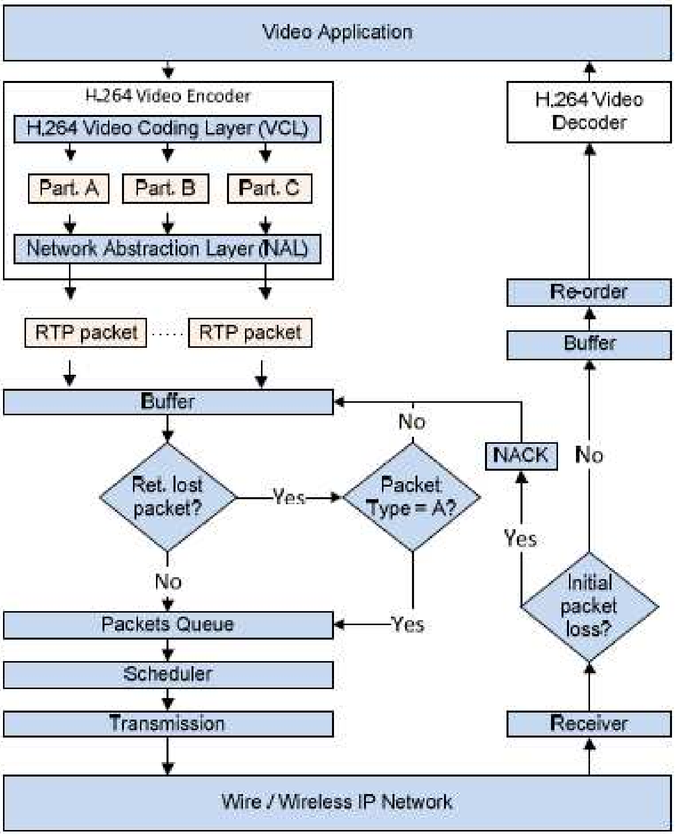


Figure 1. Operation of BVS-A.

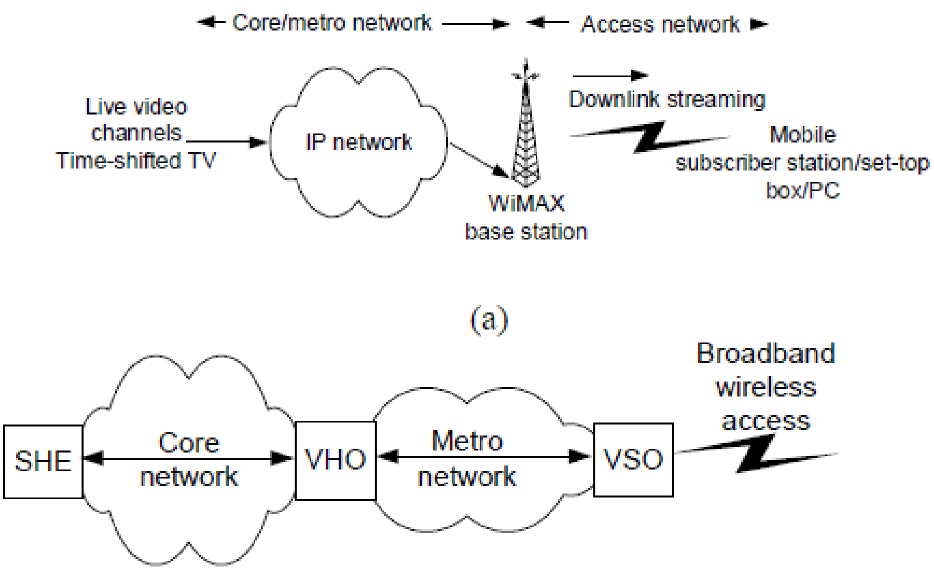
The server subsequently continues its transmissions with the next packet in sequence. Further retransmissions do not take place, as waiting packets could be delayed and because the failure of one retransmission may indicate continuing poor channel conditions.

IV. SIMULATION MODEL

In [8], IP/UDP/RTP IPTV streaming was evaluated on a WiMAX testbed for downlink (and uplink) delivery of TV channels, Fig. 2a. However, that research [8] did not consider the impact of the intervening core wired network connecting the WiMAX base stations. In [6], ways to improve IPTV quality were discussed with the assumption that intelligent content management would bring popular video content nearer to the end viewer. The typical IPTV architecture considered in [6], Fig. 2b, assumes a super head-end (SHE) distributor of content across a core network to regional video hub offices (VHOs). VHOs are connected to video serving offices VSOs) over a regional metro network.

Fig. 3 shows the tandem network simulated in which node C represents the source of downlink streaming according to Fig. 2. In the Figure, all links except a bottleneck link within the metro network are set to 100 Mbps to easily accommodate the traffic flows entering and leaving the network. A bottleneck link with capacity set to 5 Mbps is set up between the two routers. This arrangement is not meant to physically correspond to a network layout but to represent the type of bottleneck that commonly lies at a network edge. Node A sources to node B a Constant Bit-Rate (CBR) stream at 1.5 Mbps with packet size 1 kB and sinks a continuous TCP FTP flow sourced at node B. Node B also sources an FTP flow to the BS and a CBR stream at 1.5 Mbps with packet size 1 kB. This is a managed circuit-switched IP network in which statistical multiplexing still allows congestion to occur.

The WiMAX system operating in point-to-multipoint mode was simulated by well-known ns-2 simulator (v. .29) augmented by a WiMAX module [17]. Mean data points are the arithmetic mean of twenty-five runs. These points were found with 95% confidence to be statistically independent of equivalent points. The simulator was allowed to reach steady-state over 20s before commencing video streaming.



(b)

Figure 2. (a) Downlink and streaming scenarios, (b) schematic IPTV distribution network.

Figure 2. (a) Downlink and streaming scenarios, (b) schematic IPTV distribution network.

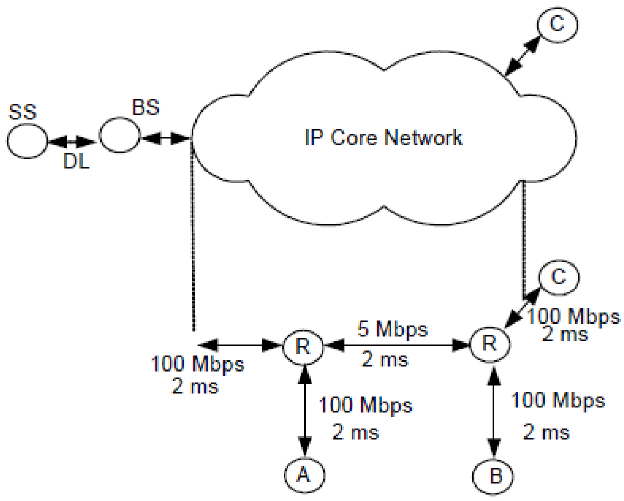


Figure 3. Video streaming scenario.

The PHY settings selected for WiMAX simulation are given in Table II, with additional MAC settings defaulted from [15]. In fact, the values shown are mandatory according to the Standard [2] to achieve a date-rate of 10.67 Mbps. The antenna is modelled for comparison purposes as a half-wavelength dipole but antenna heights and transmit powers are those recommended by the Standard. The TDD frame length is significant, as a longer frame reduces delay at SS by permitting more data to be removed from any queues at each polling time. The value of 20 ms is at the high end of the available durations in the Standard [2] in order to reduce this source of queuing delay. The buffer sizes at the BS and SS were set to fifty packets, as this value (representing about 1 s at 30 frame/s) can absorb jitter at the SS but still permit interactive TV (return of quiz answers and so on by a back channel). Similarly, router buffers were also set to fifty packets. In a WiMAX setting, a packet corresponds to a MAC Service Data Unit (MSDU) within a MAC Protocol Data Unit (MPDU).

TABLE II. SIMULATED WIMAX SETTINGS

Parameter	*Value*
PHY	OFDMA
Frequency band	5 GHz
Duplexing mode	TDD
Frame length	20 ms
Max. packet length	1024 B
Raw data rate	10.67 Mbps
IFFT size	1024
Modulation	16-QAM 1/2
Guard band ratio	1/16
DL/UL ratio	3:1
Path loss model	Two-ray ground
Channel model	Gilbert-Elliott
MS transmit power	250 mW
BS transmit power	20 W
Approx. range to MS	0.7 km
Antenna type	Omni-directional
Antenna gains	0 dBD
MS antenna height	1.5 m
BS antenna height	32 m

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

A trace file was input to ns-2 and packet losses recorded in the output. The output serves to calculate the objective video quality (PSNR). As a test, we used the Stefan sequence H.264/AVC CBR-encoded at 30 frame/s with Common Intermediate Format (CIF) (352 · 288 pixel/frame) with a target bitrate of 1 Mbps. CBR is common in broadcast distribution as it allows statistical multiplexing in a multimedia network pipe. The packet size was fixed at the encoder at 1 kB. In this way the risk of network segmentation of the packet was avoided, which could result in a loss of decoder synchronization when a packet loss occurs. Stefan is a sequence with fast motion, resulting in a high temporal coding complexity. Therefore, this sequence is a challenging sequence to reconstruct in the event of packet losses.

The well-known Gilbert-Elliott two-state, discrete-time, ergodic Markov chain [17] modelled the wireless channel error characteristics at the ns-2 physical layer. A two-state model reproduces conditions experienced during fast fading but does not model slow fades, implying the model is valid for mobile nodes within 1 km of the WiMAX mast, whereas the two-ray propagation loss model used is appropriate for line-of-sight at this distance and beyond. The probability of remaining in the good state was set to 0.95 and of remaining in the bad state was 0.94, with both states modelled by a Uniform distribution. The packet loss probability in the good state was fixed at 0.01 and the bad state probability (PB) was taken from {0.01, 0.05, 0.1, 0.15, 0.2, 0.25}.

V. EVALUATION

This Section compares BVS-A's performance to UDP, as essentially BVS-A is an improved version of baseline UDP, which was indeed used in wireless transport over WiMAX in a recent study [8]. Notice that Bell Labs introduced a reliable form of UDP, R-UDP (see RFC 1151) and there is also an RUDP protocol employed by Microsoft in their MediaRoom product for IPTV service delivery over multicast networks. By way of illustration the data-partitioned Stefan sequence (Section IV) performance is shown but to check the impact of coding complexity, results for other sequences are tabulated. In Fig. 4, UDP results in an almost linear increase in video packet loss rate (PLR) as channel conditions worsen. BVS-A's net PLR after retransmission is well-under the approximate 10% PLR beyond which experience shows [16] that adequate reconstruction of a sequence becomes difficult to achieve effectively. The increase in the PLR that does take place is because BVS-A only retransmits once, even if a partition-A bearing RTP packet is lost during retransmission.

From Fig. 5, the gain in video quality over UDP is approximately constant over the error rate range above 0.05, and given the logarithmic vertical scale represents a considerable gain in quality for BVS-A. At lower error rates (below 0.1), comparison with the input raw YUV video shows that the quality is approximately equivalent to a 'good' rating in the ITU-T's P.800's subjective mean opinion score scale. In fact, the PSNR is always approximately 'fair' (above 25 dB) according to this rating equivalence. The trade-off that BVS-A exploits is an increase in packet end-to-end delay. However, Fig. 6 illustrates that the increase in mean delay is moderate and almost asymptotic to around 30 ms. 30 ms is well within the one-way latency one would expect (circa 45 ms) across a medium-sized country such as the UK or France and in general would permit interactive applications such as videophone if the receiver was co-located with the server C in Fig. 1.

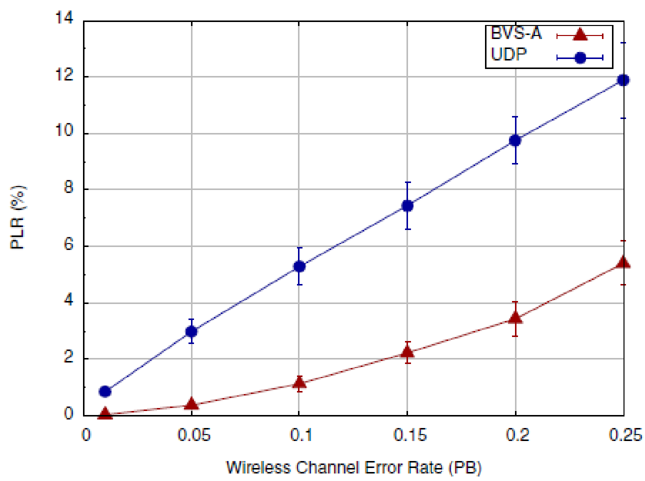


Figure 4. Video packet loss rate (PLR) for *Stefan* with increasing probability of error (PB) with 95% confidenceintervals.

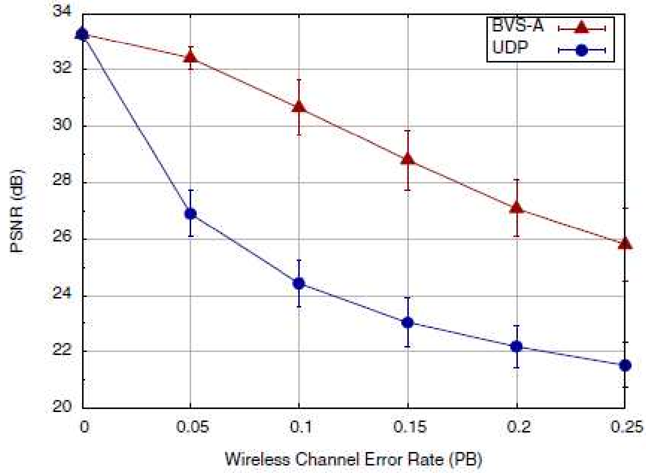


Figure 5. Mean video quality (luminance PSNR) for *Stefan* with increasing probability of error (PB) with 95% confidence intervals.

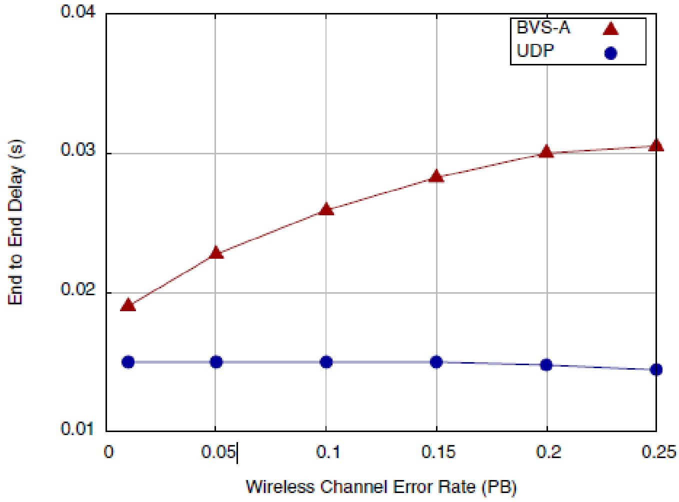


Figure 6. Mean packet end-to-end packet delay for *Stefan* with increasing probability of error (PB) with 95% confidence intervals.

Table III compares the video PLR response for other well-known sequences when encoded through datapartitioning. Akiyo is a largely static sequence with a TV announcer behind a desk filmed by a static camera. Foreman is filmed from a hand-held camera, with a fast pan towards the end of the sequence. Paris has two announcers with moderate motion but significant spatial complexity. As the sequences are CBR encoded, the different coding complexities are not reflected in the PLRs, as they would have been if Variable-Bit Rate sequences been employed. From Table III, the PLRs are consistent across the sequences for a particular channel error rate (PB). BVS-A's PLR is clearly always lower than for UDP transport.

From Table IV, the relative quality gain is determined by the coding complexity. Because of lower PLRs at lower channel error rates, the gain is generally less at these rates. The two sequences (Akiyo and Paris) with less motion benefit less from BVS, because, if partition-A bearing packets are lost, by default MVs of the MB row spatially above might be used to conceal lost MBs. When motion is limited, that MB row is more likely to be similar to its replacement. The quality gain also reduces at the highest channel error rate tested (PB = 0.25), because at that rate, the quality decline for UDP transport tends to flattens out (see Fig. 5 for an example).

TABLE III. PACKET LOSS RATES FOR DIFFERENT SEQUENCES

Channel Error Rate (PB)		Packet Loss Rate (PLR) (%)			
		Akiyo	Foreman	Paris	Stefan
0.05	UDP	2.8	2.7	2.8	2.9
	BVS	0.3	0.4	0.3	0.4
0.10	UDP	5.1	4.8	5.1	5.3
	BVS	1.0	1.1	1.1	1.1
0.15	UDP	7.4	6.9	7.4	7.5
	BVS	2.3	2.2	2.2	2.2
0.20	UDP	9.7	8.7	9.6	9.8
	BVS	3.6	3.5	3.6	3.4
0.25	UDP	11.8	10.9	11.9	11.9
	BVS	5.5	5.1	5.2	5.4

TABLE IV. QUALITY GAIN FOR DIFFERENT SEQUENCES

Channel Error Rate PB	Quality gain using BVS (dB)			
	Akiyo	Foreman	Paris	Stefan
0.05	2.75	5.30	3.32	5.52
0.10	4.00	5.72	4.34	6.23
0.15	4.09	5.53	4.66	5.76
0.20	4.02	5.25	4.34	4.91
0.25	3.94	3.45	4.35	4.28

VI. CONCLUSION

By symbiotically combining BVS with source-coded error resilience, end-to-end packet delay level is reduced as only a subset of packets is ever retransmitted. It is possible to retransmit by packet picture type (assuming periodic intrarefresh) but that would be less effective as updates would be less regular and no account of camera scene changes would be made. Future work will compare with other unequal error protection schemes. The alternative Jake's model for fast fading with the COST231 model for propagation loss, as mentioned in the Standard, can also be applied.

REFERENCES

[1] D. Gómez-Brown, D. Gozález, and N. Cardona, "Application layer FEC for mobile TV delivery in IP datacast over DVB-H systems," IEEE Trans. Broadcast., vol. 55, no. 2, pp. 396-406, 2010.

[2] IEEE, 802.16e-2005. IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, 2005.

[3] C. So-In, R. Jain, and A.-K Tamini, "Capacity evaluation of IEEE 802.16e WiMAX," J. of Computer Systems, Networks, and Communs., vol. 2010 [online], 12 pages.

[4] T. Stockhammer and W. Zia, "Error-resilient coding and decoding strategies for video communication," in P. A. Chou and M. van der Schaar (eds.) Multimedia in IP and Wireless Networks, Academic Press, Burlington, MA, pp. 13-58, 2007.

[5] S.S. Al-Majeed, and M. Fleury, "A Simple Wireless Broadband Video Streaming Scheme for IPTV," in Proc. IEEE Wireless and Mobile Networking Conf., Oct. 2010.

[6] N. Degrande, K. Laevens, and D. De Vleeschauwer, "Increasing the user perceived quality for IPTV services," IEEE Commun.. Mag., vol. 46, no. 2, pp. 94-100, Feb. 2008.

[7] T. Stockhammer, and M. Bystrom, "H.264/AVC data partitioning for mobile video communication,"in *Proc. of IEEE Int'l Conf. on* Image Processing, 2004, pp. 545-548.

[8] O. Issa, W. Li, and H. Liu, "Performance evaluation of TV over broadband wireless access networks," IEEE Trans. Broadcast., vol. 56, no. 2 , pp. 201-210, 2010.

[9] K. Tappayuthpijam, G. Liebl, T. Stockhammer, and T. Steinbach, “Adaptive video streaming over a mobile network wth TCP-Friendly Rate Control,” in *Proc. Int’l Conf. on Wireless Communs and Mobile Computing*, pp. 1325–1329, 2009.

[10] M. Chen, and A. Zakhor, “Multiple TFRC connections based rate control for wireless networks,” IEEE Trans. Multimedia, vol. 8, no. 5, pp. 1045-1061, 2006.

[11] Y. Fu, R. Hu, G. Tian, and T. Wang, “TCP-Friendly Rate Control for streaming service over 3G Network,” in *Proc. Int’l Conf. on* Wireless Commun., Networking and Mobile Computing, 2006.

[12] E. Kohler, M. Handley, and S. Floyd, “Datagram Congestion Control Protocol,” Internet Engineering Task Force, RFC 4340, 2006.

[13] P. Lambert, W. de Neve, Y. Dhondt, R. Van deWalle, ”Flexible macroblock ordering in H264/AVC,” J. of Visual Communication and Image Representation, vol. 17, pp. 358-375, 2006.

[14] 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.

[15] F. C. D. Tsai, et al. “The design and implementation of WiMAX module for ns-2 simulator,” in Proc. Workshop on NS2: The IP Network Simulator, article no. 5, 2006.

[16] F. Agboma, and A. Liotta, “Addressing user expectations in mobile content delivery,” Mobile Info. Syst., vol. 3, nos.3-4, pp. 153–164, 2007.

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