Adaptive Broadband Video Streaming using H.264 Scalability across a Tandem Network

Ismail Ali, Salah Al-Majeed, Martin Fleury, and Mohammed Ghanbari
School of Computer Science and Electronic Engineering
University of Essex, Colchester CO4 3SQ, UK
fleury.martin55@gmail.com {iaali, ssaleha, ghan}@essex.ac.uk

Abstract—As content management systems for IPTV reduce the latency between server and mobile device, real-time video streaming can tolerate single negative acknowledgments in the event of packet loss. However, the extent of packet retransmissions can be graduated and this paper proposes a scheme that is adaptive to whether network congestion or wireless channel degradation is most responsible for packet loss (or a mixture of both). By using the scalable variety of an H.264 codec, selective retransmission of base and enhancement layers is possible. The paper exhibits the trade-offs that arise as a result of adaptive wireless broadband video streaming. The scheme has been applied to a tandem network, i.e. a heterogeneous network with one part consisting of a wired IP network and the second part consisting of a broadband wireless access network.

Keywords- graceful degradation; H.264/SVC; NACKs; tandem network; video streaming; WiMAX

I. INTRODUCTION

As Internet Protocol TV (IPTV) extends to mobile devices, suitable transport protocols are sought that can adapt streaming to a tandem network that includes a final wireless access link. In previous work [1], we proposed a transport scheme that selectively negatively acknowledged (NACKed) lost packets according to their importance. The scheme leveraged the anticipated reduced latency between server and client arising from the widespread use of content delivery networks [2]. However, in such networks packet loss may occur both from poor channel conditions and congestion. Therefore, in this paper an adaptive extension is proposed that chooses between selective retransmission and complete retransmission of all lost packets according to whether congestion loss or channel loss is estimated to be dominant. The scheme has been applied to a tandem network, i.e. a heterogeneous network with one part consisting of a wired IP network and the second part consisting of a broadband wireless access network.

In the selective scheme, a NACK is only sent once per lost packet to avoid introducing excessive delay. Selective NACK is used in conjunction with the Scalable Video Coding (SVC) extension [3] of an H.264/AVC (Advanced Video Coding) codec. Though prior simple versions of BVS exist, such as that in [1], this is the first publication of the scheme with SVC. The current scheme allows only base-layer or selected quality enhancement layer packets to be acknowledged for retransmission. Importantly, the overall sending period is reduced compared to unselective NACKs and the throughput is clearly reduced. Consequently, there will be a smaller start-up delay at the receiver to avoid the risk of stream interruptions. This in turn will lead to a smaller playout buffer at the mobile receiver, contributing to a reduced energy budget through passive and active energy consumption from memory access and refresh. For ease of reference, the transport scheme is called Broadband Video Streaming (BVS) with the selective variety being BVS-selec. To further extend BVS so that it adapts to congestion, and channel conditions, we have used BVS according to the Spike scheme, first reported in [4] and then employed in [5]. However, application of BVS to the Spike scheme is the contribution of this paper's authors. The adaptive scheme does not gain as much over BVS with indiscriminant NACKs, BVS-all, as BVS-selec does. However, the adaptive variety of BVS, BVS-adap, does come close to the objective video quality (PSNR) delivered by BVS-all and still reduces throughput and end-to-end delay. Therefore, as in all engineering trade-offs BVS-adap is to be recommended for its all-round qualities. However, by providing three varieties of the BVS scheme it is possible to gracefully degrade between them resulting in lower video quality but also lower throughput and lower latency.

In [6], a survey was presented of the potential niches that exist within mobile networks for scalability and H.264/SVC in particular. Due to the widespread dissemination of H.264/AVC within (3rd Generation Partnership Project (3GPP) systems it is likely that H.264/SVC will easily transfer to such packet-switched streaming services [7]. The rate-distortion performance of H.264/SVC is also close [8] to that of its non-scalable pre-cursor. Because of the variety of mobile device types and the need for rate adaptation, scalability also finds a natural home in mobile TV broadcast systems such as DVB-H and DAB. The work in [9] is a survey of error resilience methods for H.264/SVC in 3G systems. Curiously, though retransmission strategies are identified as form of protection, no such strategy is cited. H.264/SVC was directly applied to fixed WiMAX accumulation-based congestion control for video-on-demand streaming in [10], in order to assess the capacity which was found to be about 20 mobile clients. In fact, the capacity may be larger as an early version of H.264/SVC was used with fine-grained scalability, which was later abandoned in favor of medium grained scalability (refer to Section II.B) for its coding efficiency. Various fairness metrics are compared in [10], which used standard Real-Time Transport Control Protocol (RTCP) feedback messages. Scalable video transmission can also utilize MIMO antennas [11] through parallel transmission of the layers and in [12] the layers were transmitted in parallel over multiple connections. Readers should also note another scheme for broadband video protection contributed by two of the authors, which is also adaptive and is presented in these proceedings as [13]. However, other than it is also applied to WiMAX transmission, it and related papers are entirely unrelated to this paper.

The remainder of this paper is organized as follows. Section II presents the proposed adaptive variety of BVS for scalable video. Section III details the simulation environment for the evaluation of the scheme in Section IV. Finally, some concluding remarks in Section V round-off the paper.

II. BVS ALGORITHM

This Section describes the operation of BVS-adap and the other varieties of BVS. This is followed by an outline of layer selection from H.264/SVC, followed by the adaptation extension details including the response to congestion and channel conditions.

A. *Basic operation of BVS*

Fig. 1 is a general representation of the processing involved in the adaptive version of the BVS algorithm (BVS-adap), showing the NACK responses of the receiver. Notice that H.264/SVC distinguishes between a Video Coding Layer (VCL), responsible for source coding, and a Network Abstraction Layer (NAL), responsible for video packetization. The following describes the operation assuming downlink streaming from a video server to a mobile device. At a mobile device, a record is kept of packet sequence numbers available through the Real-Time Protocol (RTP) header and, if an out-of-sequence packet arrives, a NACK may be transmitted to the server. The mobile device only transmits a NACK if this is the first time that particular packet has been lost. If it is the first time, then the channel condition is estimated. If the channel condition is judged good then only the base layer (B) and first enhancement (E1) layer packets are transmitted. If there is some doubt about the channel condition (intermediate) then for all prioritized packets, including those from enhancement layer 2 (E2), a NACK is sent.

A NACK is also sent if the channel conditions are definitely bad. Thus, a mobile device only transmits a NACK in BVS-adap if a prioritized packet has been lost for the first time. The reorder buffer at the mobile device is a playout buffer which may change the sending order of video data to fit the decode order. Upon receiving a NACK, the server prevents transmission from its input buffer until a single retransmission of the missing packet in the sequence has taken place. A holding buffer retains sent packets in the case of the need for a retransmission.

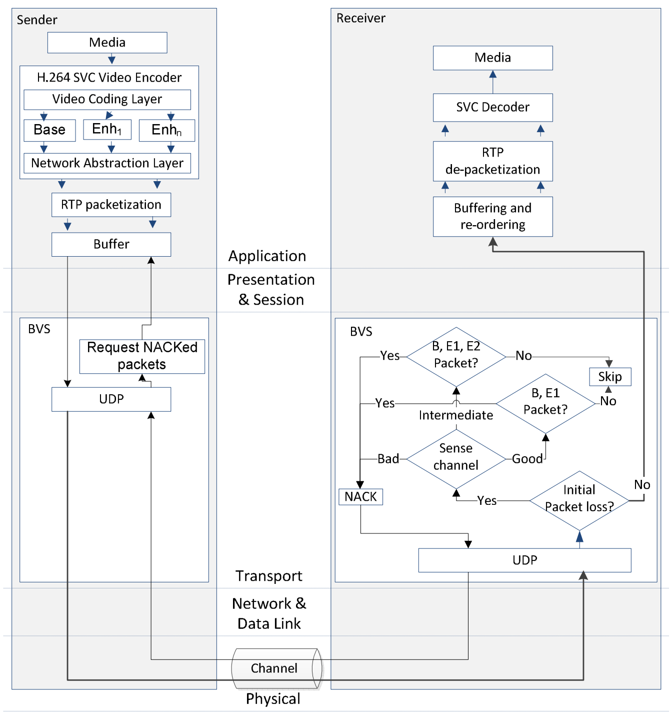

Fig. 1. Operation of BVS.

In Fig. 1, generic channel sensing is assumed. In many wireless technologies there is built in channel state estimation. For example, the IEEE 802.16e standard [14] specifies that a mobile station or device should provide channel measurements, which can either be received signal strength indicators or may be carrier-to-noise-and-interference ratio measurements made over modulated carrier preambles. However, in the adaptive scheme as described in Section II.B, a channel sensing scheme based on packet trip times is developed, which has the advantage that it is portable between different technologies.

B. *Application to H.264/SVC*

The scheme has been applied to H.264 scalability. Fig. 2 shows an illustrative size four SVC Group-of-Pictures (GOP) structure with hierarchical B-pictures (or P-pictures) between the key frames. In tests, a GOP size of 16 was employed. We have additionally utilized key pictures [3], which form the coarsest temporal layer and which serve to delimit the extent of drift within GOP borders. Partitioning the enhancement-layer transform coefficients into multiple slices, medium-grained scalability (MGS), increases the granularity of quality scalability and enables packet-based quality scalable coding. Unfortunately, this arrangement introduces inter-packet dependency which together with the predictive structure introduces complex dependencies. If not preserved, these dependencies can result in video quality reduction when the decoder drops packets that depend on non-received or erroneous ones. Our proposed adaptive BVS scheme

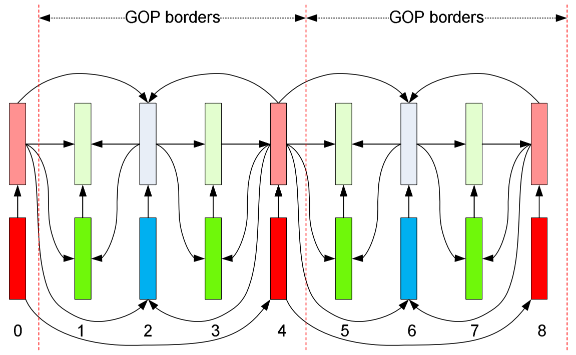

Fig. 2. Illustrative coding structure for GOP size of four with base and single enhancement layer, showing the prediction structure. Pictures appearing at 0, 4, and 8 are coded as key pictures

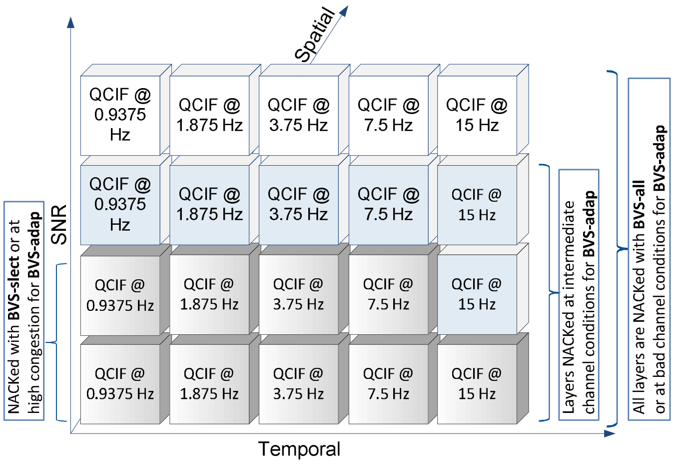

Fig. 3. H.264 SVC mapping to BVS-S. Packets belonging to darkly shaded layers are those that are negatively acknowledged

negatively acknowledges the non-received or erroneous important packets shown shaded in Fig. 3. These are either: the base SNR (quality scalability) layer packets (B1) together with selected packets from the first MGS sub-layer (E1); or additionally, packets from the second enhancement layer (E2) as well as B and E1.

If the non-adaptive scheme is used, BVS-all, then all packets from all layers in Fig. 3 are NACKed. If BVS-selec is used then only the base and enhancement layers are NACKed. Thus, BVS-selec is equivalent to BVS-adap during periods of high congestion. The paper now considers the behavior of BVS-adap.

C. BVS adaptation extension

BVS can be extended to adaptively react to congestion and channel conditions. As previously mentioned, we have used BVS according to the Spike scheme [4]. In the Spike scheme, a peak or spike in the Relative One-way Trip Time (ROTT) indicates the presence of congestion. Because of the problem of clock synchronization, one-way trip time can only be

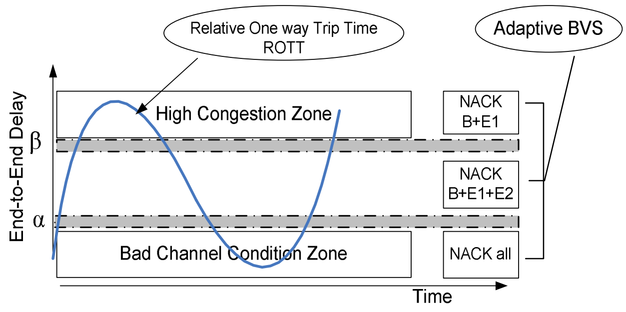

Fig. 4. Adaptive BVS NACK enhancement

relative to the clock rate at the sender and receiver and is not an absolute measure of end-to-end-delay. When the ROTT passes above a given threshold, packet loss is definitely from congestion. When it passes below a threshold, it is assumed to be definitely from wireless channel conditions.

In Fig. 4, classic BVS operates in the bad channel condition zone which can exist at various points in time. In this situation, packets from all layers are re-transmitted when necessary, in order to adequately reconstruct the video sequence. However, if there is limited congestion and moderate problems within the wireless channel then only B, E1, and E2 are retransmitted, while higher enhancement layer packets are excluded. If high congestion is declared then the second enhancement layer is not NACKed, thus limiting the number of retransmissions and as a by-product easing the congestion. Also shown in Fig. 4 are two shaded areas where there is a margin of measurement error as to which zone is occupied.

To judge the zone that is occupied the ROTT is calculated. This calculation simply consists of subtracting the receiver's clock time for receiving a packet from the RTP time stamp for when the packet was sent. Two threshold values, α and β, mark the extent of the zones in the adaptation of the Spike scheme for this paper. To find α the receiver continually monitors ROTT as recorded by the time stamped RTP packets to find the minimum ROTT. Clearly over time this estimate will be refined. The threshold is modified by a small percentage of the ROTT, which again is modified by a truncated normalized Gaussian random variable, P = G(-0.5, 0.5). The point of including this modification is to model the impact of measurement noise in the delay estimate, as in the simulations of Section IV measurement error would not otherwise occur. The lower threshold in Fig. 4 is calculated from:

$$\alpha = \text{Min. Delay Threshold} \pm (\text{ROTT} \times 5\% \times P) \quad (1)$$

In a similar manner, β is found as in (2). The maximum end-to-end delay threshold was set to 50 ms in tests, as this is the maximum recommended value for IPTV recommended by the ITU-T.

$$\beta = \text{Max.-End-to-End-Delay-Threshold} \pm (\text{ROTT} \times 5\% \times P) \quad (2)$$

Notice that adopting a constant value for the maximum end-to-end delay threshold avoids the problem of defining the separation between the upper and lower thresholds [5] that otherwise arises.

III. SIMULATION MODEL

To evaluate the scheme a tandem or heterogeneous network topology was designed with sources of congestion on the wired part and the risk of packet loss on the wireless channel. As IPTV is under active consideration [15] for IEEE 802.16e (mobile WiMAX) [14] this broadband wireless technology was modeled.

In Fig. 5's scenario, showing link delays and capacities, the server at position C streams video over the IP network, with routers designated as R. Various sources of congestion exist: node A sources to node B (obviously not to be confused with B = base layer !) constant bit-rate (CBR) data 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 CBR data at 1.5 Mbps with packet size 1 kB. Video is transferred to the WiMAX mobile subscriber station (SS) via the WiMAX base station (BS) across the downlink (DL). NACKs are sent from the SS to the server at position C.

A. WiMAX configuration

As is well-known, the PHYsical layer is the lowest layer in the seven-layer OSI standard model for protocol stacks. The PHY layer settings selected for WiMAX simulation are given in Table I. The antenna is modeled for comparison purposes as a half-wavelength dipole, whereas a sectored set of antenna on a mast might be used in practice. The antenna heights are typical ones taken from the standard [14]. The Time Division Duplex (TDD) frame length was set to 20 ms, as this value from the standard reduces polling access delay for real- time services. The data rate results from the use of one of the mandatory coding modes [14] for a TDD downlink/uplink sub-frame ratio of 3:1. The BS is assigned more bandwidth capacity than the uplink to allow the BS to respond to multiple mobile devices. Thus, the parameter settings in Table I such as the modulation type and physical layer coding rate are required to achieve a data-rate of 10.67 Mbps over the downlink.

A Gilbert-Elliott two-state, discrete-time, Markov chain modeled the wireless channel error characteristics at the ns-2 physical layer. In [16], it was shown that this model sufficiently approximates to Rayleigh fading, as occurs in urban settings during transmission from a base station to a mobile device. Moreover, in [17] it was shown that a first-order Markov chain can also model packet-level statistics.

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 embedded or hidden states modeled by a Uniform distribution. The packet loss probability in the good state was fixed at 0.01 and the bad state default was 0.05. However, the bad state

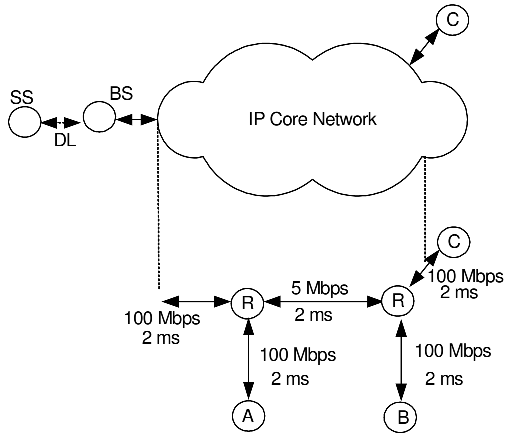

Fig. 5. Video streaming scenario

TABLE I. 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
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, IFFT = Inverse Fast Fourier Transform, DL/UL = Downlink/Uplink

packet loss probability, PB, was also varied as [0.01, 0.05, 0.1, ... , 0.25]. In this way, it is possible to judge the effect of worsening burst error channel conditions.

B. Video configuration

To judge the impact of the BVS schemes during streaming the well-known ns-2 simulator was augmented with a WiMAX module [18]. Data points are the mean of 25 runs, with 95% confidence intervals plotted in the results' graphs. As an input, 1065 frames of the Paris sequence at 30 Hz (display time 35.5 s) were encoded in Quarter Common Intermediate Format (QCIF) with the JSVM v. 9.19.1 encoder for H.264/SVC.

Paris consists of two figures seated around a table in a TV studio setting, with high spatial-coding complexity. For comparison with other results, previous frame replacement was used as a simple form of error concealment. The video packet size was limited to a maximum of 1 kB to avoid harmful network fragmentation. One SNR enhancement layer was used with transform coefficients written to three MGS layers and, hence, a finer scalable granularity was achieved. The Quantization Parameters (QP) was set to 35 and 31 for the base and enhancement layer respectively. The motion estimation and mode decision QP for the base layer was set to 33 and to 29 for the enhancement layer. Consequently, the composite SVC average bit-rate was 69.5 kbps, with up to layer 2, 1, and base layer amounting to 58.6, 47.9, and 33.1 kbps respectively, corresponding (before packet drops) to mean PSNRs of 36.2, 35.3, 34.6 and 33.6 dB.

IV. EVALUATION

The underlying protocol for BVS is UDP. Though UDP streaming has been used for fixed WiMAX [19], just as TCP is too reliable for video streaming, as some packet losses can be tolerated, UDP packet losses can harm a compressed video stream. This is due to the predictive nature of video coding which operates through motion compensation and entropy coding. Bell Labs introduced a reliable form of UDP, R-UDP, see [20], and there is also an R-UDP protocol employed by Microsoft in their MediaRoom product for IPTV service delivery over multicast networks.

The proposed BVS-Adap scheme was compared with raw UDP transport. It was also compared with sending NACKs for all packets generated by H.264/SVC and just those from the lower layers, as described in Section II.A.

In Fig. 6, the video quality of BVS-adap is close to that of BVS-all. However, at higher error rates there is about 0.5 dB difference in quality compared to the selective variety of BVS. Given the logarithmic vertical scale, this is a significance difference that would be apparent to viewers in the long term. However, for the limited differences in video quality BVS-adap reduces the bandwidth requirements compared to BVS-all. This is or could be an issue when streaming over the constrained bandwidths present at the final wireless hop. Furthermore, packet end-to-end delay is also reduced causing a general reduction in buffer memory requirements at the SS, and hence a reduction in energy consumption. Transmission is the largest consumer of energy and BVS-adap certainly achieves that compared to BVS-all.

It is possible that a reduced video quality and still further resource consumption might be desired and, in which case, BVS-selec may be preferred. However, the computational complexity of the adaptive scheme in Section II.C appears negligible. Finally, in Fig.6 and subsequent Figures there is some overlap of confidence intervals for the means of some data points and, hence, some caution is required in interpreting the plots. However, the trend of the data points in each plot taken as a whole is clear.

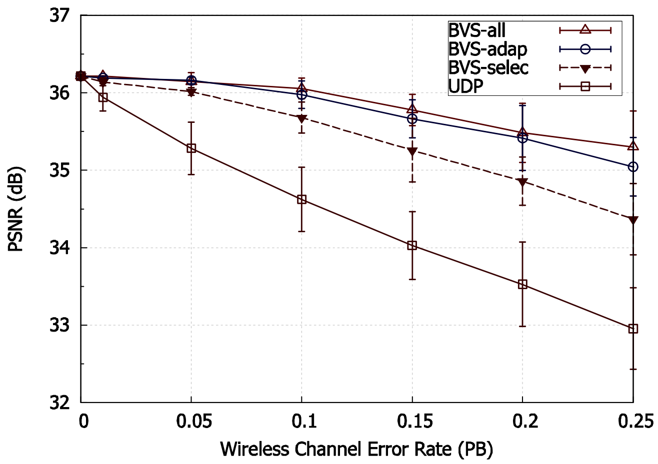

Fig. 6. Video quality (PSNR) of *Paris* after streaming at various channel error rates in the bad state

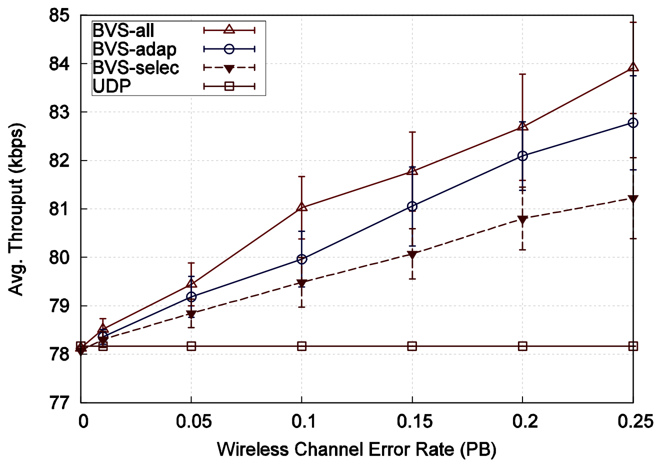

Fig. 7. Throughput at the sender from streaming *Paris* at various channel error rates, with changing error rates affecting the number of NACKs

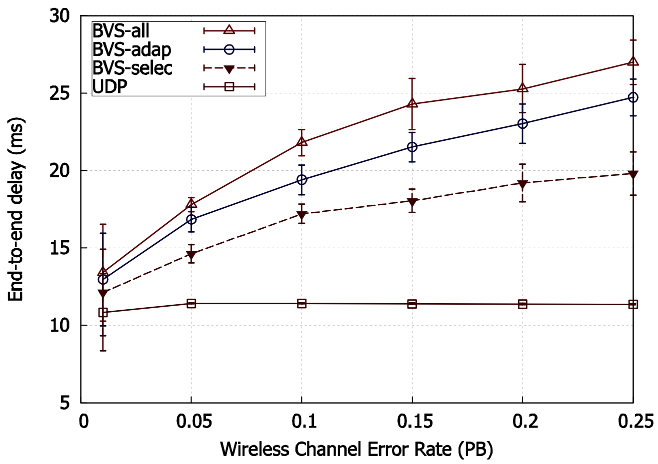

Fig. 8. Packet end-to-end delay from streaming *Paris* at various channel error rates.

V. CONCLUSION

Clearly there is a trade-off between BVS-adap, BVS-all and BVS-selec in terms of balancing video quality. In cases, where delay is important, especially in proposed interactive varieties of live IPTV (e.g. the viewer responding to quiz shows) then BVS-adap should be preferred. Start-up delay can also impact a consumer's preference for an IPTV scheme and BVS-adap is preferable. Equally, if there is competition for WiMAX capacity then BVS-adap should also be preferred. However, there is a form of graceful degradation operating such that if a service level agreement requires still further reductions in either throughput on latency then it is a simple matter to configure BVS-selec, or indeed to revert to BVS-all.

REFERENCES

[1] I. Ali, M. Al-Majeed, M. Fleury, and M. Ghanbari, "Semi-reliable transport protocol for IPTV over mobile WiMAX," in *Proc. of EURCON: 8th Conf. on Telecommunications*, Lisbon, 2011.

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

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circ. Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1119, 2007.

[4] Y. Tobe, Y.Tamura, A. Molano, S. Ghosh, and H. Tokuda, "Achieving moderate fairness for UDP flows by path-status classification," in *Proc. of 25th Ann. IEEE Conf. on Local Computer Networks*, 2000, pp. 252–261.

[5] S. Cen, P.C. Cosman, and G.M. Voelker, "End-to-end differentiation of congestion and wireless losses," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 703-715, 2003.

[6] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile video transmission using scalabale video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no.9, pp. 1204-1217, 2007.

[7] *Transparent End-to-End Packet-Switched Streamig Service (PSS) : Protocols and Codecs (Release 6)*, 3GPP TS, 26.234., Apr. 2006.

[8] M. Wien, T. Schwarz, and T. Obelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no.9, pp. 1194-1203, 2007.

[9] Y. Liu, S. Zhang, S. Xu, and Y. Zhang, "H.264/SVC error resilience strategies for 3G video service," in *Proc. of Int. Conf. on Image Analysis and Signal Proc.*, 2009, pp. 207-211.

[10] O.I. Hillestad, A. Perkis, V. Genc, S. Murphy, and J. Murphy, "Adaptive H.264/MPEG-4 SVC video over IEEE 802.16 broadband wireless networks," in *Proc. of Int. Packet Video Workshop*, 2007.

[11] D. Song, and C.W. Chen, "Scalable H.264/AVC video transmission over MIMO wireless systems with adaptive channel selection based on partial channel information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no.9, pp. 1218-1226, 2007.

[12] H.-H Juan, H.-C. Huang, C.Y. Huang, and T. Chiang, "Cross-layer system design for scalable video streaming over mobile WiMAX," in *Proc. of IEEE Wireless Communs. And Networking Conf.*, 2007, pp. 1860-1864.

[13] L. Al-Jobouri, M. Fleury, and M. Ghanbari, "Protection modes for segmented video streaming over broadband wireless," 4th Joint IFIP Wireless and Mobile Network Conf., 2011.

[14] 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.

[15] I. V. Uilecan, C. Zhou, and G. E. Atkin, "Framework for delivering IPTV services over WiMAX wireless networks," in *Proc. of IEEE Electro/Info. Technol. Conf.*, 2007, pp. 470-475.

[16] H. Wang and P. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 353–357, 1996.

[17] Ch. Jiao, L. Schwiebert, and B. Xu, "On modeling the packet error statistics in bursty channels**,"** in *Proc. of IEEE Conf. on Local Computer Networks*, 2002, pp. 534- 541.

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

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

[20] C. Partridge, and R. Hinden, "Version 2 of the Reliable Data Protocol," RFC 1151, 1990.