Bull, D. R., & Chung-How, J. T. H. (2002). Packet loss resilient videoconferencing using H.263+. In International Conference on Acoustics, Speech and Signal Processing (CASSP'02) Orlando, FL, USA. (Vol. 4, pp. IV-4052 - IV-4055). Institute of Electrical and Electronics Engineers (IEEE). 10.1109/ICASSP.2002.1004808

PACKET LOSS RESILIENT VIDEOCONFERENCING USING H.263+

D.R. Bull[a] and J.T.H. Chung-How[b]

[a]Image Communications Group, Centre for Communications Research
University of Bristol, Woodland Road, Bristol BS8 1UB, U.K.

[b]ProVision Communication Technologies Ltd.
University Gate, Park Row, Bristol BS1 5UB, U.K.

ABSTRACT

Real-time video transmission over the Internet is becoming increasingly desirable for videoconferencing and other interactive applications. The reliable transport protocols used by the Internet were designed mainly for non-real time data, and cannot be used for delay critical applications, which must therefore be able to cope with packet loss. Motion compensated video coding is especially sensitive to loss because of temporal error propagation. In this paper, the effect of packet loss on H.263+ video transmitted using the real-time transport protocol (RTP) is assessed. Various algorithms are described to minimise or prevent temporal error propagation. These algorithms do not rely on retransmissions and do not introduce more than one frame delay, and are therefore suitable for real-time and multicast applications. It is shown that the robustness of H.263+ video can be greatly improved using these techniques. Such techniques can also be applied to other motion compensated video coding standards such as MPEG-4 and H.26L.

1. INTRODUCTION

The Internet was designed mainly for non-time critical data, and is ill-suited for the transmission of time-critical data such as interactive video. Packets can be lost or dropped when intermediate links or routers become congested due to excess traffic. Reliable transport protocols like TCP/IP recover from loss by using acknowledgements and retransmissions. However, the resulting latency is generally too large for real-time interactive applications, where late packets are effectively lost. As a result, real-time multimedia applications typically use UDP/IP, which provides an unreliable packet delivery service. The Real-time Transport Protocol (RTP) [1] was defined to enable real-time multimedia applications over the Internet. A payload format for H.263+ video has been defined for use with RTP [2]. Video transmission over the Internet has received considerable attention. The most popular scheme for providing error-resilience in a packet video transport system is scalable or layered coding combined with some form of prioritisation, forward error correction (FEC) [3,4] or receiver-based rate control which is particularly well suited for the heterogeneous nature of the Internet [5,6]. Recently, a number of solutions for Internet video streaming have been developed [7].

In this paper, the problem of robust video transmission over the Internet is addressed. In order to make our solution applicable to the widest range of situations requiring video over IP, we assume that there is *no* feedback channel from the decoder to the encoder, as is the case for a typical multicast application because of the feedback implosion problem. We also assume a real-time environment with strict end-to-end delay requirements, such as in a typical two-way videoconferencing application. Further improvements are possible if these constraints are relaxed, e.g. in a video-streaming application where delay is not so critical. First, the robustness to packet loss of RTP-H.263+ video is investigated. The main problem is caused by temporal error propagation and two ways of minimising this propagation are proposed - the *selective use* of FEC on the motion information and the use of *periodic reference* frames. Using *periodic reference* frames is shown to be more efficient and robust than periodic intraframe coding when used with FEC. These two algorithms are then combined, resulting in a very efficient robust video-coding scheme that provides graceful degradation as the packet loss rate increases.

Generally, Internet packet loss rates vary widely and losses may occur in bursts, i.e. a lost packet is more likely to be followed by another loss [8,9]. However, no simple method exists for modelling the typical loss patterns likely to be seen over the Internet since the loss depends on so many untractable factors. Therefore, in all the loss simulations presented in this paper, random loss patterns are used. This is believed to be the best solution since our test sequences are of short duration and burst losses can be modelled as very high random loss.

2. RTP-H263+ WITH PACKET LOSS

In order to transmit H.263+ video over the Internet, the H.263+ bitstream must be packetised according to the RTP payload format specification [2], and then transmitted as RTP packets. In addition to the RTP/UDP/IP headers, the RTP packetisation also generates a RTP-H.263+ payload header, which is normally 16 bits. To minimise packetisation header overhead, each RTP packet should be as large as possible. In practice, to avoid IP fragmentation, the size of the packet must be kept below the maximum transmission unit (MTU) of the network, which is 1500 bytes for Ethernet. On the other hand, for maximum robustness to loss, packet size should be kept to a minimum.

In our experiments, the slice-structured mode was not used and packetisation was always performed at GOB boundaries, i.e. each RTP packet contains one or more complete GOBs. Since every packet begins with a picture or GOB start code, the leading 16 zeros are omitted in accordance with RFC 2429 [2]. The packetisation overhead then consists only of the RTP/UDP/IP headers, which is typically 40 bytes per packet. This overhead can be quite significant at low bit-rates.

The header overhead associated with using 1, 3 and 9 GOBs/packet for QCIF images at 5 fps is given in Table 1. Results for one simulation with the *foreman* sequence (QCIF, 125 frames, 5 fps) coded at 62 kbps (excluding RTP/UDP/IP header overhead) with 10% random packet loss are shown in Fig 1. Only the first frame was intracoded. The RTP sequence number enables the decoder to detect lost packets, so that the location of the missing GOBs is known. The missing blocks can then be concealed by using the motion vector of the MB immediately above the lost MB in the same frame or the motion vector is assumed to be zero if that MB is missing as well. As expected, the image quality degrades rapidly as the errors due to missing packets propagate from one frame to the next and the 1 GOB/packet scheme performs best, at the expense of increased header overhead. Other simulations have confirmed that for baseline H.263 video, the 1 GOB/packet approach provides the best trade-off between robustness to loss and packetisation overhead.

No. of GOBs/packet	1	3	9
Packetisation Overhead (kbps)	14.4	4.8	1.6

Table 1. RTP/UDP/IP header for QCIF at 5 fps.

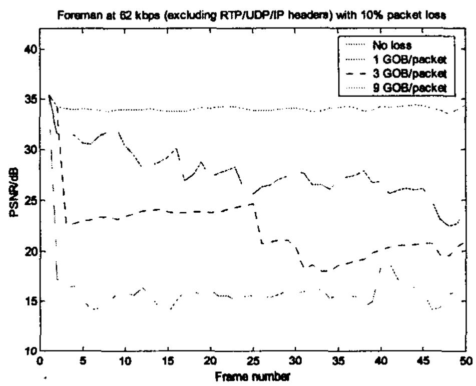

Fig. 1. H.263 with 10% loss and 1,3 and 9 GOBs/packet.

The main problem with motion compensated video in the presence of loss is the temporal propagation of the errors as a damaged frame is used as reference for the decoding of subsequent frames. The obvious way of limiting this error propagation is to use some form of intra-replenishment but these tend to be very expensive in bit-rate.

3. SELECTIVE FEC OF MOTION VECTORS

It is known that for a typical video sequence compressed with H.263, the correct decoding of the motion vectors play a crucial role in the proper reconstruction of the decoded video. This is especially true when considering the relatively small fraction of the total bit-rate occupied by the motion information, compared to the DCT coefficients and other side information. For example, for the *Foreman* sequence at 62 kbps, about 70% of the total bits is made up of the DCT coefficients whereas the motion vectors only take up 10% of the total. In order to increase the robustness to packet losses, we therefore propose to apply forward error correction across packets on the motion vector information only.

The Reed-Solomon Erasure (RSE) correcting code can be very effective in this case since lost packets in effect result in packet erasures, as the positions of the lost packets are known. With a RSE(n,k) code, k data packets are used to construct r parity packets, where $r=n-k$, resulting in a total of n packets to be transmitted. The k source packets can be reconstructed from any k packets out of the n transmitted packets. This provides for error-free decoding for up to r lost packets out of n.

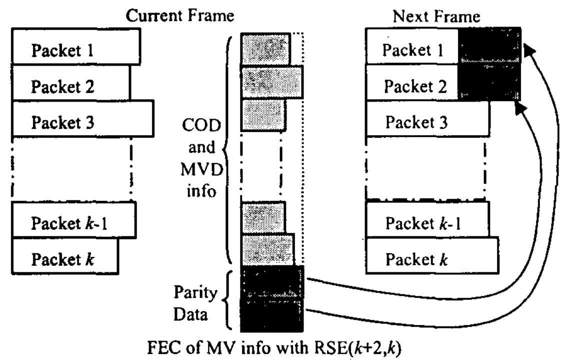

Fig. 2. FEC of motion information across packets with r=2.

No. of parity packets	0	1	3	6
Bit-rate/kbps (% increase)	60.31 (0%)	61.10 (1.3%)	62.69 (3.9%)	65.07 (7.9%)

Table 2. Increase in rate of MV-FEC for *foreman* at 62 kbps.

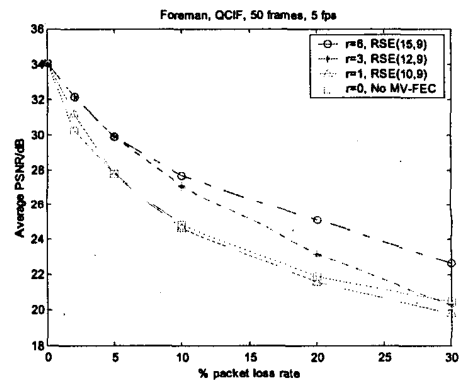

Fig. 3. Performance of MV-FEC.

In order to achieve a good trade-off between decoding delay and robustness to burst packet losses, the RSE code is applied across the packets of a single frame. This results in a RSE(n, k) where k is the number of data packets per frame. The motion information for each packet is contained in the *COD* and *MVD* bits of the H.263 bitstream. The RSE(n, k) encoding is therefore applied across these bits segments in each of the k packets, generating (n-k) parity bit segments. The length of the parity bit segments will be equal to the maximum length of the *COD* and *MVD* data segments among the k packets. When applying the RSE encoding, missing bits for shorter segments are assumed to be zero. The FEC data segments are then appended to the data packets of the following frame (Fig. 2), so that the number of packets per frame does not change. So, up to k parity packets (i.e. r=k) can be used by such a scheme, and if a data packet were to be lost, there would be an additional one frame delay at the decoder before the motion vectors could be recovered.

The increase in bit-rate resulting from the use of MV-FEC is detailed in Table 2. The average PSNR for the *Foreman* sequence at 62 kbps with 1 GOB/packet for different values of r and loss rates are shown in Fig. 3. FEC of the motion vectors considerably increases the robustness to loss at the expense of a very small increase in bit-rate. For example, with r=3, i.e. 3 parity packets, the degradation in PSNR caused by 10% packet loss is reduced by more than 4 dB for only about 4% increase in rate.

4. PERIODIC REFERENCE FRAME

In all motion-compensated video-coding schemes, the most recent previously coded frame is generally used as reference for the temporal prediction of the current frame. The *Reference Picture Selection* (RPS) mode (annex N) of H.263+ enables the use of any previously coded frame as reference. Using any other frame than the most recent one reduces the compression efficiency, but can be beneficial in limiting error propagation in error-prone environments, as we will show next. A scheme referred to as *Periodic RPS* is used here, where every n^{th} frame in a sequence is coded with the n^{th} previous frame as reference. Such frames are known as *periodic reference* (PR) frames, and n is the *frame period*, which is the number of frames between PR frames. All the other frames are coded as usual, i.e. using the previous frame as reference (Fig. 4). This idea is similar to [10], but here we do not rely on retransmissions and the scheme is applied explicitly with an H.263+ codec. The advantage of PR frames is that if any errors in a PR frame can be corrected through the use of FEC before it is referenced by the next PR frame, then this will effectively limit the temporal error propagation. FEC can be used on the PR frames in a similar fashion as for the motion vectors by using the RSE code across packets. In this case as well, the RSE encoding is applied across the packets of a single frame. However, this time the generated parity data is transmitted as separate RTP packets.

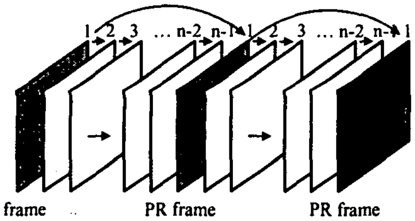

Fig. 4. PR frame scheme with frame period=n.

The amount of loss that the Periodic RPS with FEC (PR-FEC) scheme can tolerate depends on the amount of FEC used. For a frame period of 10 for the *foreman* sequence, the use of 4 parity packets (r=4) results in a total bit-rate similar to that required for an intraframe every 10 frames. For *akiyo*, 2 parity packets results in less than a quarter of the equivalent increase in rate for periodic intraframe coding with the same PSNR (Fig. 5). Fig. 6 compares the robustness of PR-FEC and intraframe coding for foreman at 62 kbps with 1 GOB/packet and a frame period of 10, i.e. for periodic RPS, there is a PR frame every 10 frames, and for the intracoding scheme, there is an intraframe every 10 frames. Results are shown for r=1, 3 and 6 parity packets for each PR frame. The amount of loss that the PR/FEC scheme can tolerate depends on the amount of FEC used. We observe that for low packet loss rates PR-FEC is more effective than intraframe coding in limiting temporal error propagation even with low values of r. For higher loss rates, larger values of r are required, e.g. with r=6, PR-FEC is still more robust than the use of intraframes for up to 25% loss rates.

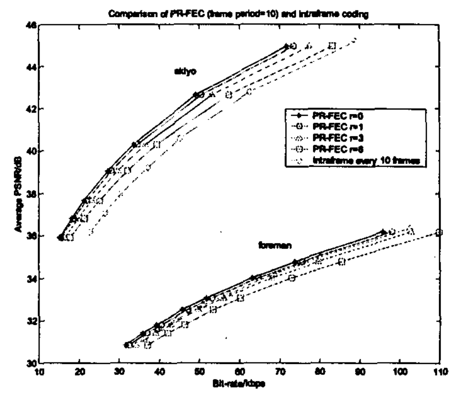

Fig. 5. Comparison of PR-FEC and intraframe coding.

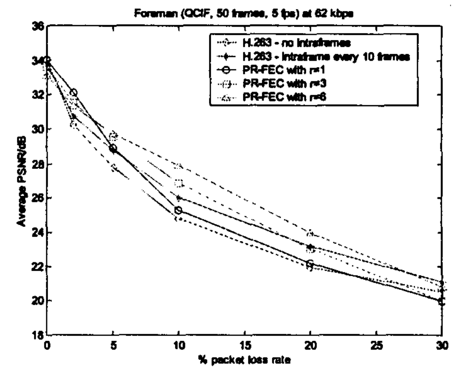

Fig. 6. PR-FEC and Intraframe coding for *foreman* at 62 kbps.

5. ROBUST RTP-H263+ VIDEO CODING

The two proposed robust techniques – MV-FEC and PR-FEC - are now combined into a single codec. Most practical applications require some form of resynchronisation in some way or another, e.g. in videoconferencing when a receiver joins a session halfway through. So we also include resynchronisation in our robust codec in the form of periodic intra-MB replenishment. This is applied to the PR frames only where a number of MBs in each PR frame are intracoded. The following schemes at roughly the same bit-rates are compared:

- RTP-H.263+: H.263+ packetised with RTP.
- RTP-H.263+ with intraframe: Same as previous but with an intracoded frame every 10 frames.
- PR-MV-FEC: H.263+ packetised with RTP together with PR-FEC with frame period of 10 as well as MV-FEC. FEC with r = 1,3, and 6 parity packets used on both motion vectors and PR frames. 5 MB per PR frame are also intracoded.

The results for different packet loss rates are shown in Fig. 7 for *Foreman* at 62 kbps and *Container* at 32 kbps with 1 GOB/packet. RTP-H.263+ without intra replenishment performs best for error free conditions but degrades catastrophically with loss. The use of intra replenishment provides slightly better loss performance at the expense of decreased coding efficiency. Depending on the amount of FEC used, our robust scheme can provide better coding efficiency than intraframe replenishment as well as greater resilience to packet loss. The image quality degrades gracefully with loss rates and increasing the amount of FEC provides greater robustness at high loss rates with only a minimal effect on coding efficiency for low loss rates. Typical decoded images obtained for PR-MV-FEC with r=3 at 62 kbps with 10 and 20% random loss are shown in Fig. 8. Similar results were also obtained with 3 GOBs/packet and for CIF sequences [11].

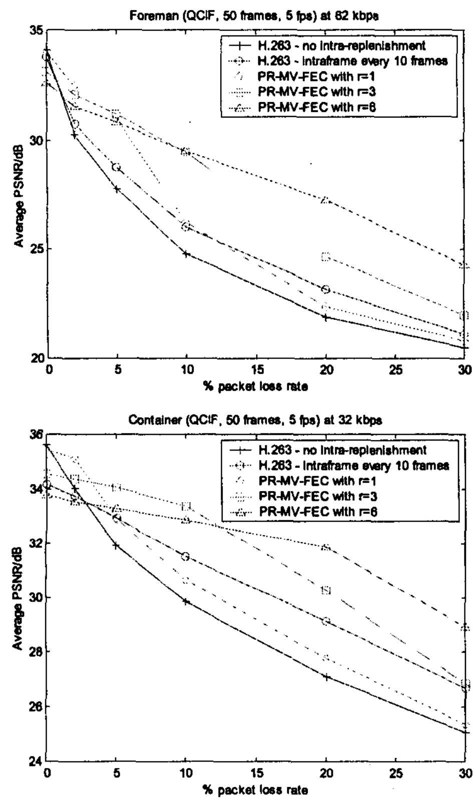

Fig.7. PR-MV-FEC with 1 GOB/packet (a) *foreman* at 62 kbps (b) *container* at 32 kbps.

Fig. 8. Frame 32 using PR-MV-FEC with 10 and 20 % loss.

The proposed robust coding scheme is fully compatible with the H.263+ standard and only requires minimal changes to the RTP specifications so that FEC packets and the amount of FEC used can be signalled to the decoder. In a typical multicast application, the scheme can be used in an adaptive fashion where the amount of FEC is varied at the encoder based on the loss rate received from RTCP reports.

6. SUMMARY

In this paper, we consider the robust transmission of H.263+ video over the Internet. We assume that there is no feedback channel, e.g. as in a multicast situation, and a strict end-to-end delay requirement, e.g. for a videoconferencing application. Two techniques, selective FEC of motion vectors and use of periodic reference frames, are described to minimise error propagation in H.263+ coded video. When these techniques are combined together, the robustness of H.263+ to packet loss is greatly improved. Simulation results show that acceptable image quality is still possible even with loss rates as high as 30%. The modified H.263+ codec has been implemented and integrated into the software videoconferencing tool *vic*, which can be used for real-time video multicast over the Internet.

REFERENCES

[1] Schulzrinne H., Casner S., Frederick R. and Jacobson V., "RTP: A transport protocol for real-time applications," *RFC 1889*, Jan. 1996.

[2] Bormann C., Cline L, Deisher G., Gardos T., Maciocco C., Newell D., Ott J., Sullivan G., Wenger S. and Zhu C., "RTP payload format for the 1998 version of ITU-T Rec. H.263 video (H.263+)," *RFC 2429*, Oct. 1998.

[3] Horn U., Stuhlmuller K., Link M. and Girod B., "Robust internet video transmission based on scalable coding and unequal error protection," *Signal Processing: Image Communication*, vol. 15, no. 1-2, pp. 77-94, Sept. 1999.

[4] Boyce J.M., "Packet loss resilient transmission of MPEG video over the Internet," *Signal Processing: Image Communication*, vol. 15, no. 1-2, pp. 7-24, Sept. 1999.

[5] Tan W. and Zakhor A., "Real-time internet video using error resilient scalable compression and TCP-friendly transport protocol," *IEEE Trans. Multimedia*, vol. 1, no. 2, pp. 172-186, June 1999.

[6] McCanne S., Vetterli M. and Jacobson V., "Low-complexity video coding for receiver-driven layered multicast," *IEEE J. Select. Areas Commun.*, vol. 16, no. 6, pp. 983-1001, Aug. 1997.

[7] Civanlar R., Luthra A., Wenger S. and Wenwu Z., "Introduction to the special issue on streaming video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 265 - 268, March 2001.

[8] Paxson V., "End-to-end internet packet dynamics," *IEEE/ACM Trans. Networking*, vol. 7, no. 3, pp. 277-292, June 1999.

[9] Yajnik M., Kurose J. and Towsley D., "Packet loss correlation in the MBone multicast networks," *IEEE Globecom '96*, London, U.K., 1996.

[10] Rhee I., "Error control techniques for interactive low-bit rate video transmission over the Internet," *Proc. SIGCOMM '98*, Vancouver, Sept. 1998.

[11] Chung-How J.T.H. and Bull D.R., "Loss resilient H.263+ video over the Internet," *Signal Process.: Image Commun.*, vol. 16, pp. 891-908, June 2001.