Error Resilient Packet-Switched Video Telephony with Adaptive Rateless Coding and Reference Picture Selection

Muneeb Dawood, Raouf Hamzaoui, Shakeel Ahmad, Marwan Al-Akaidi
Faculty of Technology, De Montfort University, Leicester, UK.
Email:muneebel@yahoo.com, {rhamzaoui,sahmad,mma}@dmu.ac.uk

Keywords: Video telephony, channel coding, rateless codes, H.264, LTE.

Abstract—**Providing high-quality video for packet-switched wireless video telephony on hand-held devices is a challenging task due to packet loss, limited available bandwidth, and complexity constraints. We propose a low-complexity channel-adaptive error resilience technique that combines application-layer forward error correction (FEC) with rateless codes, retransmission, and reference picture selection. Experimental results for H.264 video sequences show that the proposed technique achieves significant peak-signal-to-noise ratio (PSNR) and percentage degraded video duration (PDVD) improvements over previous techniques in networks involving two wireless links.**

I. Introduction

Real-time video communication over wireless packet networks typically relies on the User Datagram Protocol (UDP) at the transport layer. Since UDP is an unreliable transport protocol, packet loss can occur. Because of the encoding interdependencies, packet loss can significantly degrade the received video quality. For this reason, application-layer error resilience techniques have been proposed to protect the transmitted video. Surveys of these techniques can be found in [1], [2], [3]. In [4], we proposed a technique based on forward error correction (FEC) with Luby Transform (LT) codes [5] and reference picture selection. While our technique showed promising results, it lacked adaptivity to channel conditions as the redundancy of the rateless code was fixed in advance. In this paper, we show that the performance of our previous technique can be improved by adapting the rateless code redundancy to the channel conditions. Experimental results for standard H.264 [6] video sequences show that the new technique achieves significant peak-signal-to-noise ratio (PSNR) and percentage degraded video duration (PDVD) improvements over our previous technique in networks involving two wireless links.

The remainder of the paper is organized as follows. In Section II, we describe our previous technique [4]. In Section III, we present our new solution. In Section IV, we compare the peak signal to noise ratio (PSNR) and the percentage degraded video duration (PDVD) performance of our new solution to those of our previous technique for a simulated Long Term Evolution (LTE) network. Section V concludes the paper.

II. Previous work

Fig. 1 shows the block diagram of our previous system [4]. Live video frames are fed to the H.264 video encoder which compresses them at a source rate s_r and generates a sequence of Network Abstraction Layer (NAL) units. The NAL units corresponding to two input frames are combined into a source block. LT encoding is applied on the source block to generate encoded symbols. The LT encoded symbols produced by the LT encoder are packetized in RTP/UDP/IP packets. The RTP/UDP/IP header is compressed and a two-byte Packet Data Convergence Protocol (PDCP) header is appended to the resulting IP packet to form a Radio Link Control Service Data Unit (RLC-SDU). The RLC-SDUs are mapped into Radio Link Control Protocol Data Units (RLC-PDUs) for transmission. Due to bit errors in a received RLC-PDU, all IP packets that are partially or fully mapped to it are lost.

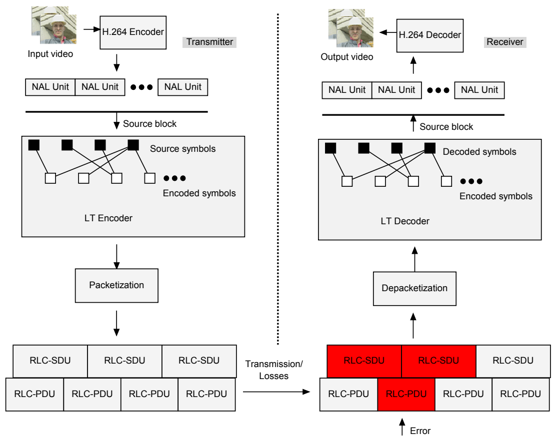

Fig. 1. System block diagram

IP packets that are received correctly are passed to the LT decoder. If enough encoded symbols are received, LT decoding is successful, and all NAL units associated to the source block are recovered. If LT decoding is not successful, all NAL units associated to the source block are considered to be lost and the video decoder uses frame freeze concealment to replace all frames in the failed source block by the last successfully decoded frame. In this case, a mismatch of reference frames between the sender and receiver occurs, which results in spatio-temporal error propagation. To mitigate it, a variant of the reference picture selection technique [7] is used. The receiver sends a feedback that contains the block ID of the last successfully received block, allowing the transmitter to update the reference frame (Fig. 2).

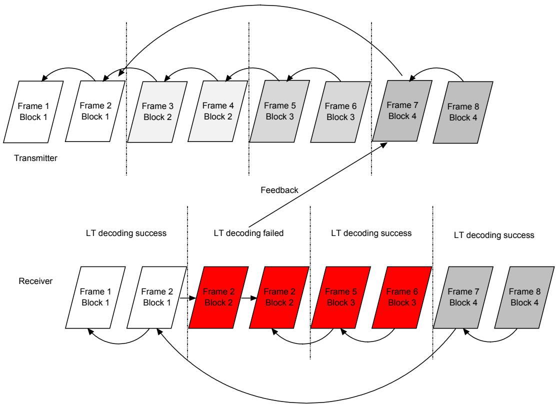

Fig. 2. Reference picture selection. LT decoding of block 2 has failed. Feedback is received by the encoder before encoding frame 7. Error propagation is stopped at frame 7.

III. Proposed System

In the method presented in Section II, the redundancy of the LT code for a source block is fixed in advance. The performance of this method can be improved by adapting the LT code redundancy to the channel conditions.

A. Adaptive scheme

Fig. 3 shows the proposed transmission strategy. It is assumed that the transmitter and receiver clocks are synchronized. This can be achieved, for example, by using the Network Time Protocol (NTP) [8]. The transmission start time for both end-users is the same. The transmission sending deadline for forward and backward transmission is set to $\frac{n_f}{f_r}$, where n_f is the number of frames corresponding to a source block and f_r is the frame rate. The transmission receiving deadline for forward transmission is set to $\frac{n_f}{f_r} + FTT$, where FTT is the maximum forward trip time. The transmission receiving deadline for backward transmission is set to $\frac{n_f}{f_r} + BTT$, where BTT is the maximum backward trip time.

When RTP/UDP/IP packets are mapped to RLC-PDUs for transmission, flexible RLC-PDU size [9] is used. Flexible RLC-PDU size allows an RTP/UDP/IP packet of any size to be mapped to exactly one RLC-PDU. A maximum size m of an RTP/UDP/IP packet is chosen such that the packet is mapped to an RLC-PDU of exactly 320 bytes.

In a first round, $k_s = k \times (1 + r)$ encoded symbols are sent. Here k is the number of information symbols in the source block and r is the initial LT code redundancy, which is chosen such that LT decoding has a high probability of success if all k_s encoded symbols are received. The k_s symbols are sent in p RTP/UDP/IP

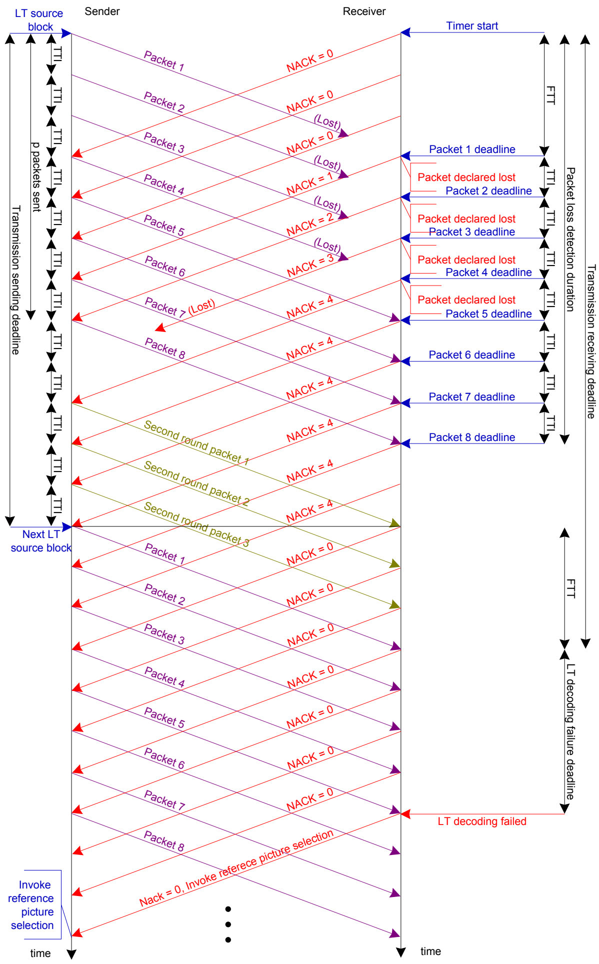

Fig. 3. Proposed adaptive transmission strategy. Packets are sent at fixed transmission opportunities separated by TTI. Cumulative NACK is used to indicate packet loss to the transmitter. There is no packet deadline for the second round packets. When there is no initial round or second round packet to send, the transmitter sends non-LT packet at the available TTI. These non-LT packets are not shown for simplicity.

packets with $p = \frac{k_s}{m}$. The first $p-1$ packets have the same size, equal to the maximum RTP/UDP/IP packet size m. The last packet may be smaller than the first p-1 ones. For each $n = 1, \ldots, p$, the nth packet is sent at time $t = (n-1) \times TTI$.

The receiver starts a timer at $t = 0$. It sets a deadline t_n for receiving the nth packet by

$$t_n = FTT + (n-1) \times TTI,\ n = 1 \ldots, \hat{p},$$

where $\hat{p}$ is an estimation of p. It is computed as $\frac{\hat{k} \times (1+r)}{m}$, where $\hat{k} = \frac{s_r}{f_r} n_f$ is an estimation of k. It is then updated to p as soon as a packet is received (this is possible since each packet contains the value of k). The receiver keeps a record of the number of packets received for the current source block. For each $n = 1, \ldots, p$, if by the nth deadline it does not receive at least n packets, it concludes that some packets have been lost.

The cumulative number of lost packets is sent as Negative Acknowledgment (NACK) in the backward packets. For example, in Fig. 3, the first three packets are not received by the receiver by the 3rd deadline, so the receiver sets NACK to 3 and sends it in the subsequent packet.

After the first round is completed, more packets are sent in a second round if not all p packets have been received. At each TTI, the sender determines the number of second round packets to send as the difference between the NACK and the cumulative number of second round packets already sent. The transmitter sends a second round packet only after it has received a packet containing a NACK and the number of second round packets already sent is smaller than this NACK.

The size of second round packets is set to the maximum size of the RTP/UDP/IP packet. Because the receiver does not know if or when a second round packet is sent, no receiver deadline for the second round packets is used.

When the transmitter has sent all p packets and there is no second round packet to send, it sends packets containing only NACK and reference picture selection information. These dedicated feedback packets are sent until the transmission sending deadline for the current source block is reached.

Fig. 4 illustrates the case where two backward packets are lost.

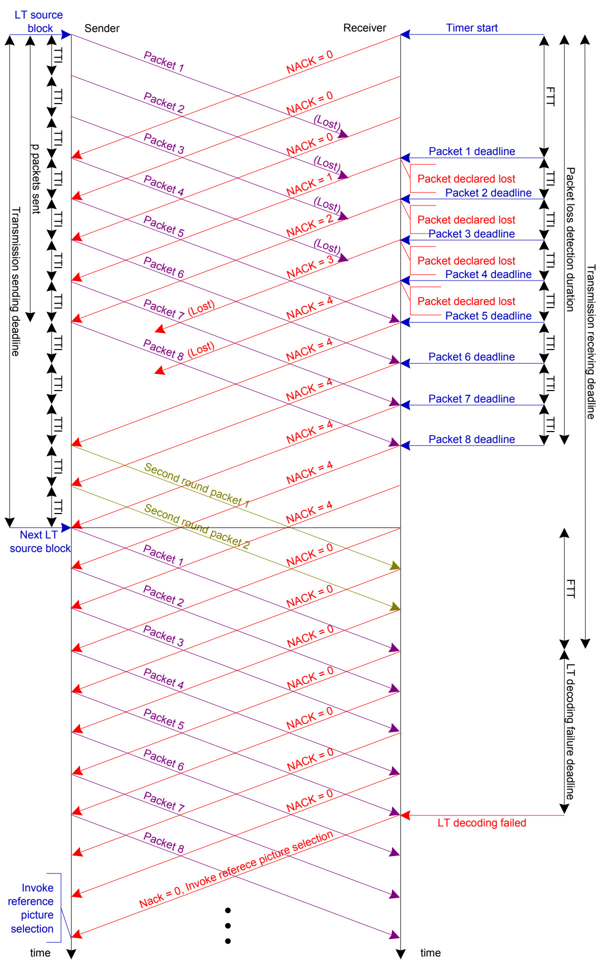

Fig. 4. Example when two backward packets are lost.

The initial round packets and second round packets contain LT encoded symbols, LT information, NACKs and reference picture selection feedback. The non-LT packets contain NACKs and reference picture selection feedback.

IV. Experimental Results

In this section, we compare the performance of the proposed adaptive system to that of our previous system [4] for Long Term Evolution (LTE) networks. Fig. 5 shows the considered network topology. Two user equipments (UEs) are connected to two base stations (eNodeBs) through a wireless channel. The eNodeBs are connected to the cellular network gateway through a wired link. The cellular gateways are connected to each other through a wired backbone link. The wired links are assumed to be error free. The source of packet loss are the two wireless links between the UEs and the eNodeBs.

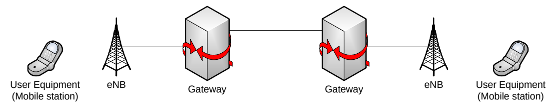

Fig. 5. Network topology.

The Nokia H.264 video coder was used to encode the QCIF (176×144) *Stunt* and *Party* [10] video sequences. *Stunt* contains 240 frames and has a frame rate of 15 frames per second (fps). *Party* contains 360 frames and has a frame rate of 12 fps. The first frame was encoded as an I frame and the remaining frames were encoded as P frames. The following coding parameters were used: one reference frame, one slice per frame, no rate-distortion optimization, no sub 8×8 coding modes, motion vector range of 8 pixels, CAVLC entropy coding. The resulting source bit rate was 89 kbit/s and 90 kbit/s for *Stunt* and *Party*, respectively.

The source block consisted of NAL units corresponding to two video frames. For the method of [4] the LT code redundancy was set to 35%. The initial LT code redundancy for the proposed adaptive method was set to 17%. The LT code symbol size was one bit. The robust soliton distribution [5] was used with $c = 0.1$ and $\delta = 0.5$. To improve the performance of the LT code, block duplication with an expanding factor [11] of 8 was used.

The maximum size of the RTP/UDP/IP packet was $40 + 7 + 306 = 353$ bytes. Here 40 bytes were for the RTP/UDP/IP header, 7 bytes for the LT code information, and 306 was the maximum number of LT encoded bytes. Each of the first p-1 RTP/UDP/IP packet contained exactly 2448 (306×8) LT encoded symbols (bits). The last of the p packets contained up to 2448 LT encoded symbols. The size of each second round RTP/UDP/IP packet was 353 bytes and contained 2448 symbols. A 353-byte RTP/UDP/IP packet yielded an RLC-SDU of 320 bytes after compression of the RTP/UDP/IP header and addition of the PDCP header. One RLC-SDU and hence one RTP/UDP/IP packet was mapped to one RLC-PDU. The TTI was set to 10 ms. One RLC-PDU was sent at each TTI. The maximum FTT and maximum

BTT were set to 40 ms. These values are realistic for 3GPP LTE [9].

We used the Rayleigh fading channel model proposed in [12] as the wireless channel model. In this model, the RLC-PDU loss is approximated by a two-state Markov process. In the good state, the RLC-PDU is received correctly. In the bad state, the RLC-PDU is lost. The mobile velocity was 3 km/h and the carrier frequency was 2 GHz. The fade margin parameter was selected to give steady state RLC-PDU loss rates of 0%, 0.5%, 1%, 1.5% and 5% over 10^8 iterations. The same RLC-PDU loss rate was used in both wireless channels.

Fig. 6 shows the PSNR results for the *Stunt* sequence. When there was no packet loss, the PSNR results of the method of [4] and the proposed method were similar. As the RLC-PDU loss rate increased, the adaptive method had a higher LT decoding success and consequently achieved better PSNR results. At RLC-PDU loss rate of 5% in each wireless link, the PSNR gain was about 4.13 dB.

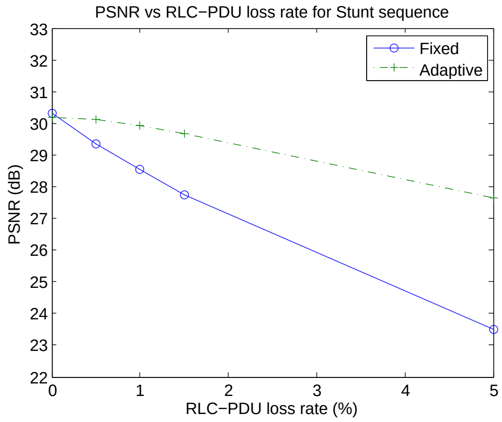

Fig. 6. PSNR vs. RLC-PDU loss rate for the *Stunt* sequence.

Fig. 7 shows the PDVD results for the *Stunt* sequence. The PDVD results of the two methods were similar when there was no packet loss. However, the adaptive method achieved better PDVD results as the RLC-PDU loss rate increased. At RLC-PDU loss rate of 5% in each wireless channel, the PDVD improvement reached 26.11%. The adaptive method had better PDVD results because it had fewer LT decoding failures and hence experienced less error propagation.

Fig. 8 shows the bit rate curves for the *Stunt* sequence. When there was no packet loss, the bit rate of the method of [4] was higher than that of the adaptive method. This is because the LT code redundancy of the previous method is fixed to 35% while the initial LT code redundancy of the adaptive method is 17%. As the RLC-PDU loss rate increased, the bit rate of the method of [4] did not change noticeably. In contrast, the bit rate of the adaptive method increased due to more retransmissions but remained smaller than that of the method of [4].

The PSNR, PDVD, and bit rate curves for the *Party* sequence are given in Fig. 9, Fig. 10, and Fig. 11, respectively. The results were similar to those of the *Stunt*

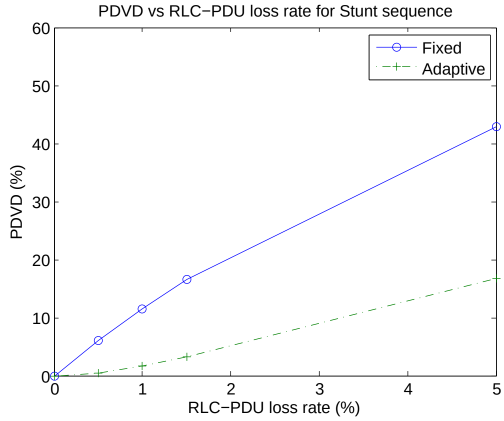

Fig. 7. PDVD vs. RLC-PDU loss rate for the *Stunt* sequence.

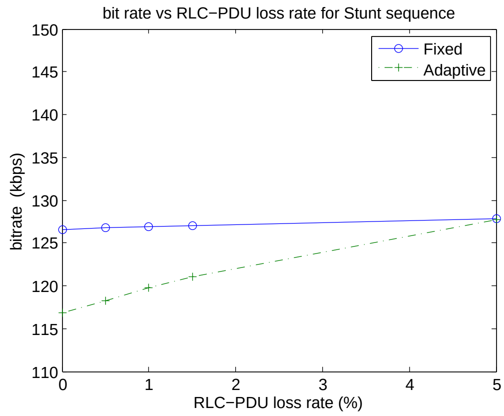

Fig. 8. Bit rate vs. RLC-PDU loss rate for the *Stunt* sequence.

sequence. At RLC-PDU loss rate of 5% in each wireless channel, the adaptive method had a PSNR improvement of 3.98 dB and PDVD improvement of 28.62%.

V. Conclusion

We proposed an error resilience technique for H.264 video telephony over LTE networks. Our technique combines rateless coding and reference picture selection. It is channel adaptive and exploits feedback and retransmission. Experimental results showed significant gains in PSNR, PDVD, and bit rates over the method proposed in [4]. Improvements to the proposed system can be obtained by making the transmitter predict when LT decoding will not be successful and stop sending packets earlier.

References

[1] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," Proceedings of the IEEE, vol. 86, no. 5, pp. 974–997, 1998.

[2] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Review of error resilient coding techniques for real-time video communications," IEEE Signal Processing Magazine, vol. 17, no. 4, pp. 61–82, 2000.

[3] J. Devadoss, V. Singh, J. Ott, C. Liu, Y.-K. Wang, and I. Curcio, "Evaluation of error resilience mechanisms for 3G conversational video," in: Proc. 10th IEEE International Symposium on Multimedia, Berkeley, CA, pp. 378–383, 2008.

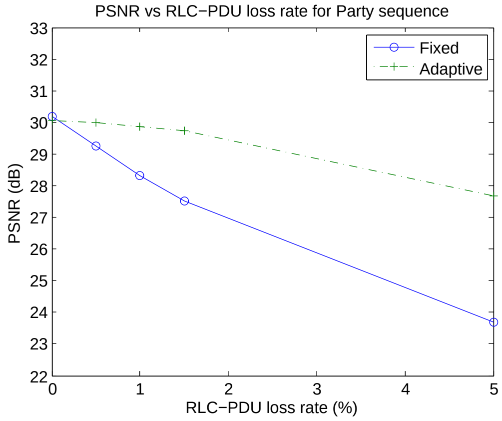

Fig. 9. PSNR vs. RLC-PDU loss rate for the *Party* sequence.

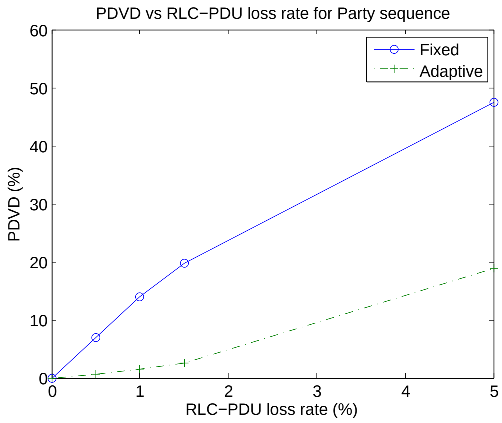

Fig. 10. PDVD vs. RLC-PDU loss rate for the *Party* sequence.

[4] M. Dawood, R. Hamzaoui, and M. Al-Akaidi, "Error-resilient packet switched H.264 mobile video telephony with LT coding and reference picture selection," in: Proc. EUSIPCO 09, Glasgow, Aug. 2009.

[5] M. Luby, "LT codes," in: Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002.

[6] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, pp. 560–576, July 2003.

[7] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," Proceedings of the IEEE, vol. 87, no. 10, pp. 1707–1723, 1999.

[8] D. L. Mills, Network Time Protocol (Version 3) Specification, Implementation and Analysis. Request for Comments 1305. Internet Engineering Task Force.

[9] H. Holma and A. Toskala (Eds.), LTE for UMTS - OFDMA and SC-FDMA Based Radio Access, John Wiley & Sons, Ltd. 2009.

[10] 3GPP, 3GPP Technical Recommendation 26.902 (v 7.1.2), *Video codec performance*, 2008.

[11] S. Ahmad, R. Hamzaoui, and M. Al-Akaidi, "Unequal error protection using LT codes and block duplication," in: Proc. Middle Eastern Multiconference on Simulation and Modelling, Amman, Aug. 2008.

[12] M. Zorzi, R.R. Rao, and L.B. Milstein, "On the accuracy of a first-order Markov model for data transmission on fading channels," in: Proc. Int. Conf. Universal Personal Communications, pp. 211–215, 1995.

BIOGRAPHIES

Muneeb Dawood received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2003 and the M.Sc. degree in

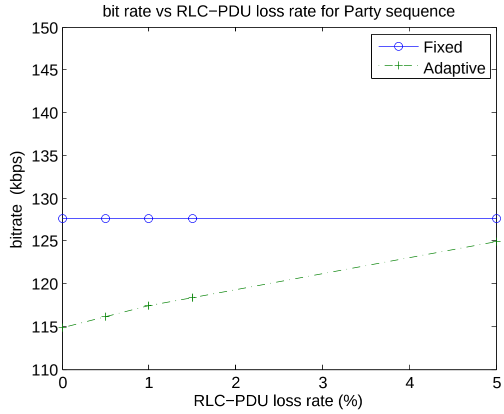

Fig. 11. Bit rate vs. RLC-PDU loss rate for the *Party* sequence.

telecommunications and computer networks engineering from London South Bank University, London, UK, in 2007. He is a PhD student at De Montfort University, Leicester, UK. His research interests include mobile cellular networks, video communication, and error resilient techniques.

Raouf Hamzaoui received the the Dr. rer. nat. degree from the Faculty of Applied Sciences, University of Freiburg, Freiburg, Germany, in 1997, and the Habilitation degree in computer science from the University of Konstanz, Konstanz, Germany, in 2004. He is a Professor of Media Technology in the Faculty of Technology, De Montfort University, Leicester, UK. His research interests include image and video compression, multimedia communications, error control systems, and algorithms.

Shakeel Ahmad received the Dr.-Ing degree from the University of Konstanz, Germany, in 2008. He is a Research Fellow with the Faculty of Technology, De Montfort University, Leicester, UK. His research interests include multimedia communication, error-resilient video streaming, and channel coding.

Marwan M. Al-Akaidi (PhD CEng CIP FIEE FBCS FIAP SMIEEE) is a Professor and Director of the Multimedia Communications and Signal Processing Group, De Montfort University, Leicester, UK. He led the school of Engineering & Technology at De Montfort University from Jan. 2002 to Sept. 2008, and currently is appointed as the Director of External Relations in the Faculty of Technology. He enjoys strong links with industry, including Siemens, O2, BT, Orange, Nokia, BMW, and Chronos. He is interested in all aspects of wireless communications and signal processing and he is the author of the "Fractal Speech Processing" book (Cambridge University Press). He is the SPC chair and conference coordinator of the IEEE UKRI. He is also the chair of the IET 3G Technology & beyond 3G technology conferences as well as the WiMAX08 London, WiMAX09 London conferences and the chair and founder of the MESM conferences since 99. He is also editor-in-chief of AWMC and editorial member of many national and international journals.