Application-Aware Dynamic Retransmission Control in Mobile Cellular Networks

Nadhir Ben Halima, Dzmitry Kliazovich, and Fabrizio Granelli

DISI - University of Trento
Via Sommarive 14, I-38050 Trento, Italy
{nadhir,kliazovich,granelli}@disi.unitn.it

Abstract. This paper proposes an application-aware cross-layer approach between application/transport layers on the mobile terminal and link layer at the wireless base station to enable dynamic control on the strength of per-packet error protection for multimedia and data transfers. Specifically, in the context of cellular networks, the proposed scheme allows to control the desired level of Hybrid ARQ (HARQ) protection by using an in-band control feedback channel. Such protection is dynamically adapted on a per-packet basis and depends on the perceptual importance of different packets as well as on the reception history of the flow[1].

Keywords: Hybrid ARQ (HARQ), cellular networks, service aware protocols.

1 Introduction

Nowadays, networking services are evolving to a “triple play” vision, implying delivery of data, voice and video to the end user using the same IP transport facility.

While no solution for end-to-end quality of service (QoS) assurance over heterogeneous networks is available, still several approaches exist for improving data transfer performance on the wireless access trunk [6].

In the specific framework of multimedia (e.g. voice and video), several works are available based on the Unequal Error Protection (UEP) paradigm [1-5]. The goal of UEP is to provide higher protection to the most perceptually relevant data, where protection can be achieved through means of adaptive power levels, forward error correction codes, retransmission control, etc. Nevertheless, since UEP is usually performed or managed at source level and thus without specific knowledge of the contingent operating scenario, such solutions (while increasing the complexity of multimedia codecs) can lead to non-optimal performance due to waste of available capacity in case network / channel conditions are good (and no packet drops are experienced) or for time-varying performance oscillations of the transport infrastructure (particularly true in the case of wireless networks).

[1] This work has been supported by the Italian National Project: Wireless multiplatfOrm mimo active access netwoRks for QoS-demanding muLtimedia Delivery (WORLD), under grant number 2007R989S, and by the PAT grant RObust Streaming Environment (ROSE).

F. Granelli et al. (Eds.): MOBILIGHT 2009, LNICST 13, pp. 45–52, 2009.

The proposed scheme represents a novel paradigm of dynamic and “link-level” UEP, focused on the access network and the actual “reception history” at the receiver. The core idea is to adaptively tune the level of HARQ protection based on the relative importance on the overall user experience of the packet being transmitted by the base station. Such approach enables to differentiate protection on the basis of the actual content of the packet: for voice and video flows, the impact of losing the current packet is estimated in terms of audio or visual quality as measurable by MOS or PSNR, respectively, for data transfer TCP throughput is chosen as the main quality feedback metric.

The structure of the paper is as follows: Section 2 describes in details the proposed framework, while performance evaluation is presented in Section 3. Finally, Section 4 concludes the paper with final remarks and outlines about future work on the topic.

2 Proposed Approach

The main idea of the proposed approach, called Application-Aware Dynamic Retransmission Control (A^2DRC), is to allow the mobile terminal receiver to control the level of HARQ protection applied by the base station for every frame transmitted on the radio link. The decision of the mobile terminal is based on the potential benefit in correctly receiving the next packet given the current reception history and the actual perceptual relevance of the packet itself.

As an evolution of Automatic Repeat Request (ARQ) approach used in UMTS, the Hybrid ARQ (HARQ) scheme employed in HSDPA allows incorrectly received data blocks to be soft-combined in the effort to correct propagated link errors [8].

In this paper, the level of HARQ protection (also indicated as “HARQ Strength” in the following) is considered in terms of the maximum number of stop-and-wait retransmission attempts taken for a packet delivery in case of failure.

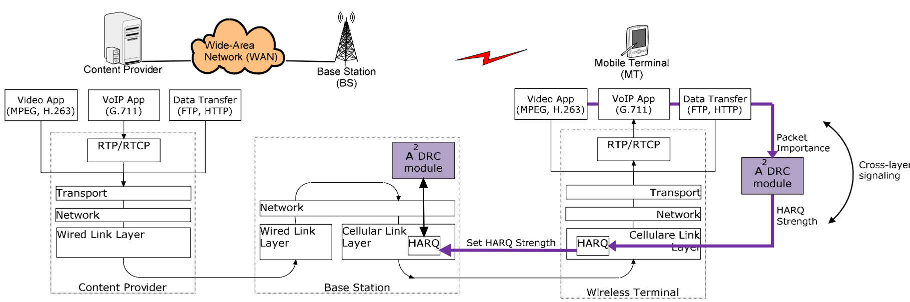

Fig. 1. Architectural principles of A^2DRC in a 3G cellular network. Blocks and links highlighted in “blue” underline the modules and signaling links employed by A^2DRC approach.

Fig. 1 illustrates architectural principles of the proposed A^2DRC approach. As outlined in the previous sections, A^2DRC operates on the wireless 3G link. At the mobile terminal side, whenever a packet is received by the application, the latter can specify packet importance for subsequent incoming packets for a given flow. The packet importance is then transferred into corresponding values of HARQ protection by the A^2DRC module (implemented within the protocol stack at the mobile terminal) and delivered to the HARQ entity at the link layer of the Base Station using cross-layer signaling. At the link layer, the specified HARQ protection parameter is sent along with HARQ acknowledgement, which is generated for every frame received according to stop-and-wait HARQ type.

The A^2DRC module implemented at the BS analyses incoming traffic and specifies the HARQ entity to use the requested HARQ protection on a per-packet basis.

In this paper, we address three different classes of services – voice, video, and data transfer – improving delivery performance by adapting network response to the relevance of packet being delivered over the radio link.

The level of HARQ protection in the proposed approach varies on the basis of a packet importance metric, which consists of two components:

- *Initial packet importance* corresponds to the level of quality reduction for a given flow in case the packet is lost during transmission or corrupted at the receiver [9]. The quality of the flow is determined by end-to-end application requirements and user demands. For example, commonly used metric for VoIP is Mean Opinion Score (MOS), for video is Peak Signal-to-Noise Radio (PSNR), and for TCP-based data is transfer throughput level.

- *Dynamic packet importance* component accounts for the "reception history" of the flow and adjusts initial packet importance. For example, the importance of frame i in a video sequence can be dynamically adjusted in case its decoding depends on the neighboring frames $i-1$ and $i+1$ and frame $i-1$ is not correctly received.

Packet Importance Metrics in Video Streams. For sake of a clear explanation, we consider a scenario with a mobile node receiving MPEG-4 video flows from a streaming server located in the wired Wide-Area Network (WAN). However, similar reasoning can be applicable to H.263 and H.264 encoded video streams, as well as embedded video streams.

An MPEG-4 video is composed of Groups of Pictures (GOPs), consisting of video frames of three types: I-Frames (Intra coded frames) which are encoded without reference to any other frame in the sequence, P-Frames (Predicted frames) which are encoded as differences from the last I- or P-frame, and B-Frames (Bidirectional frames) which are encoded as the difference from the previous or following I- or P-frames.

Due to the correlation property of P- and B-frames, the effective impact deriving from the loss of an I-frame can be clearly considered much higher than that of P- or B-frame. In addition, the loss of one I- or P-packet may generate error propagation: while the loss of a B-frame does not affect the quality of the consecutive frames, the loss of an I-frame may disable correct decoding of subsequent P- and B- frames. This leads to the conclusion that I-frames are more important than P-frames, which are more important than B-frames.

Fig. 2 shows the quality reduction of a real video flow transmitted using VideoLan software [12] in terms of PSNR measured at the receiver versus the loss of different types of packets within a GOP. The horizontal scale indicates which frame within the GOP was lost, while the first value (obtained with no losses) serves as a reference point.

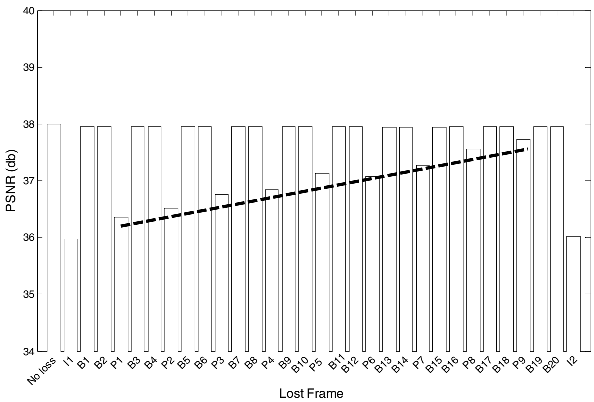

Fig. 2. Quality of the received video flow for different frames lost

Following such observation, the importance of P-frames P_{imp} is defined ranging linearly from I_{imp} to B_{imp}, where I_{imp} is the importance level of I-frames and B_{imp} is the importance level of B-frames with $I_{imp} \geq P_{imp} \geq B_{imp}$.

Packet importance metrics in VoIP flows. At the receiver side, speech frames are de-multiplexed and inserted into a playout buffer. The playout buffer plays an important role in perceived speech quality since it enforces speech frames delivery at the same interval at which they are generated by the encoder. This is done through reordering, delaying or even dropping the frames which arrive later than their expected playback time. However, whenever the frame is dropped it causes a relevant decrease of the quality of the voice stream.

Based on the above, initially, equal packet importance (i.e., "initial packet importance") is associated to all transmitted speech frames. However, in case the receiver detects frame losses after out-of-order frame reception, it increases importance (and error redundancy) for the subsequent packets of the stream (i.e. increases the "dynamic packet importance"). Summarizing, A^2DRC aims at avoiding bulk frame losses, which are critical for the quality of the speech stream, while single frame losses can be easily compensated or concealed by the decoder.

Packet importance metric in file transfer. Packet losses can severely decrease the data transfer performance of TCP which is the most widely used protocol in Internet.

The proposed A^2DRC scheme dynamically adapts the level of HARQ protection used on the radio link based on the value of the TCP congestion window computed at the receiver node.

The core idea is to provide higher protection on the radio link (and more retransmission attempts) when congestion window is small and lower protection for high window values. Indeed, when congestion window is small, any link error will trigger window reduction to its half – unnecessarily reducing the throughput of the TCP flow. In the opposite case, the impact of link errors becomes less significant, since the window will be possibly reduced due to congestion-related losses.

Fig. 3 presents congestion window evolution in TCP New Reno and the corresponding proposed variation of the packet importance metric. Specifically, the proposed approach assigns the highest importance ("High Imp") to TCP segments produced

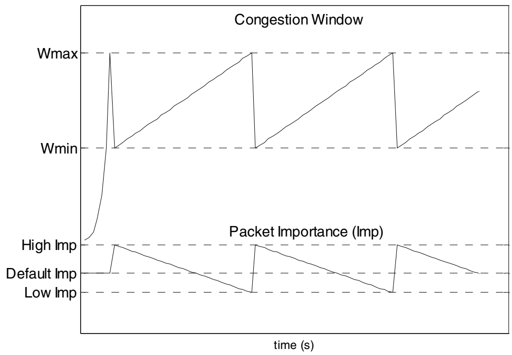

Fig. 3. Variation of HARQ strength used for TCP packet transmission based on the congestion window parameter estimated at the receiver

right after each window reduction and decreases it down to the "Low Imp" threshold following linear or any other monotonically decreasing function.

Summarizing, A^2DRC provides higher protection for low congestion window values or flow sending rates. This reduces the probability of packet losses due to link errors on the wireless channel, which is a well-known reason for TCP performance degradation [7].

3 Performance Evaluation

The proposed scheme is evaluated in the context of an UMTS/HSDPA cellular network. Network Simulator 2 (NS-2) [13] with the additional Enhanced UMTS Radio Access Network Extensions (EURANE) module [14] was used for the experiments. Figure 4 illustrates the reference scenario and the main parameters employed in the experiments.

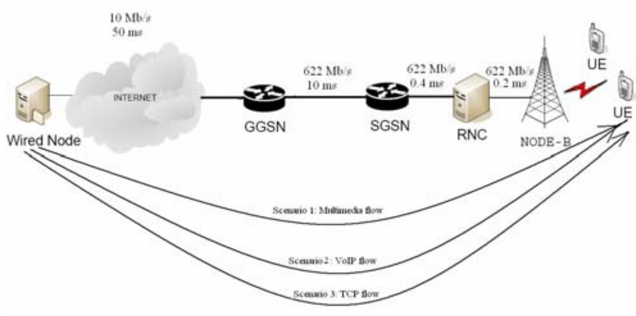

Fig. 4. Simulation scenario used for ns-2 experiments

In video transfer scenario, the FH is a video server which transmits video streams to the video receiver located at UE. Results are presented for the "Foreman" video sequence, using MPEG-4 (open-source ffmpeg [15]) video coding. The video format is Quarter Common Intermediate Format (QCIF, 176 * 144). The GOP structure is IBBPBBPBBPBBPB. The "Foreman" video trace is composed of 300 frames (10 I-frames, 102 P-frames and 188 B frames). An integrated environment methodology proposed and developed within the framework of EvalVid [11] was for experiments, enhanced as in [17] for including NS-2.

The crucial portion of the multimedia stream is retransmitted by A^{2}DRC with a higher HARQ strength, i.e. packets belonging to an I-frame are retransmitted with HARQ Strength = 8, while packets belonging to B-frames are retransmitted with a HARQ strength = 2. P-frames are retransmitted with a variable HARQ strength ranging from 8 to 3 depending on the position of the frame in the GOP. Default value of HARQ strength is set to 4 for all packets in the legacy scenario (i.e. without A^{2}DRC). Achieved results are illustrated in Fig.5, where A^{2}DRC increases the range of packet error tolerance to 10^{-2}-10^{-1}.

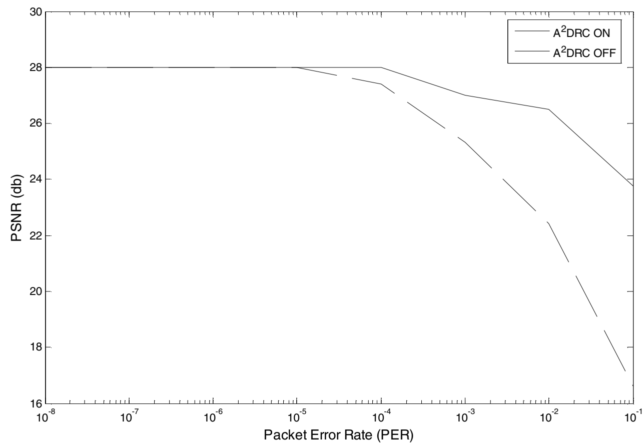

Fig. 5. Quality of "Foreman" video clip for different error rates

<u>*VoIP Transfer Performance*</u>. Experiments on VoIP flows are performed using the simulation model presented in [10]. Initially, equal HARQ strength equal to 3 is associated to all transmitted speech frames. However, in case the receiver detects frame losses after out-of-order frame reception, it increases HARQ strength linearly for the subsequent packets of the stream (with HARQ strength max equal to 8) in order to avoid bulk frame losses. Once no loss is detected, A^{2}DRC decreases the HARQ strength to the initial value.

Achieved results (Fig. 6) demonstrate that A^{2}DRC is able to provide a relevant improvement in terms of MOS both for G.711 and GSM AMR speech flows. In average, application of A^{2}DRC scheme enables the codec to deliver the same speech quality for error rate of 5% higher if compared with the case when A^{2}DRC is not enabled.

In File Transfer Performance scenario, FTP/TCP flow sent by FH is received by the UE. For the entire duration of the flow the receiver maintains up-to-date value of the congestion window (*cwnd*) computed by counting the number of packets received for the last RTT. Whenever the loss detection signal (three duplicate acknowledgements) is sent to the sender packet importance is increased according the function

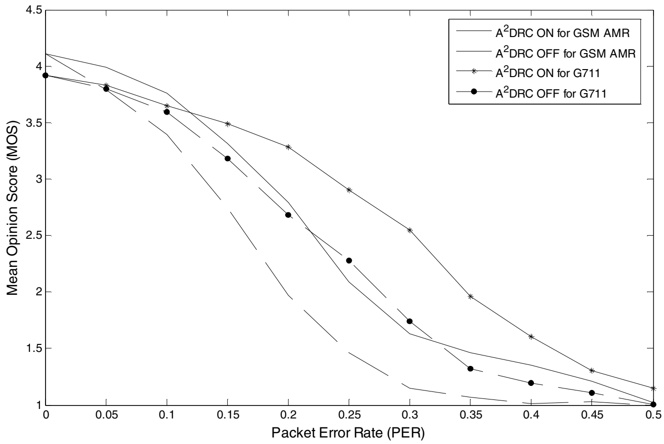

Fig. 6. G.711 and GSM AMR Voice MOS against Packet Error Rate

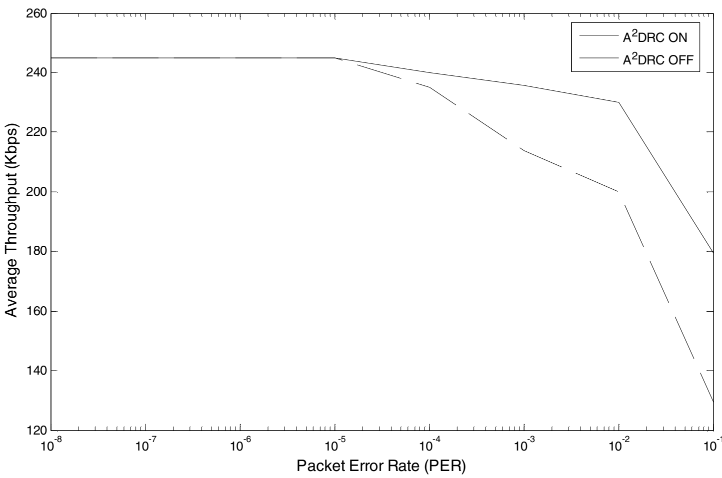

Fig. 7. Average TCP Throughput as a function of Packet Error Rate on the wireless channel

presented in Section 2 causing higher strength of HARQ process and, as a result, producing higher resistance to the link errors. Figure 7 presents TCP throughput achieved by the flows for different PERs of the wireless link. As expected, higher protection against the link errors for low congestion values of the congestion window brings evident performance improvement and underlines advantages of dynamic error protection techniques based on application awareness introduced by A^2DRC.

4 Conclusions

In order to enable dynamic control on the level of per-packet HARQ protection, the paper proposed a cross-layer solution between application/transport layers on a mobile terminal and link layer at the base station. Protection level is dynamically set on a per-packet basis and depends on the importance of different packets as well as on the reception history of the flow. Experimental results show that our scheme improves audio and video flows as well as TCP-based data transfers.

References

1. Horn, U., Stuhlmüller, K., Link, M., Girod, B.: Robust Internet video transmission based on scalable coding and unequal error protection. Image Communication 15(1-2), 77–94 (1999)
2. Wu, Z., Bilgin, A., Marcellin, M.W.: Unequal error protection for transmission of JPEG2000 codestreams over noisy channels. In: ICIP (2002)
3. van der Schaar, M., Radha, H.: Unequal packet loss resilience for Fine-Granular-Scalability video. IEEE Trans. on Multimedia 3, 381–394 (2001)
4. Yang, X.K., Zhu, C., Li, Z.G., Feng, G.N., Wu, S., Ling, N.: A degressive error protection algorithm for MPEG-4 FGS video streaming. In: Proc. of IEEE ICIP, Rochester, USA (September 2002)
5. Zhang, Q., Zhu, W., Zhang, Y.-Q.: Network-adaptive scalable video streaming over 3G wireless network. In: IEEE ICIP 2001, Thessaloniki, Greece (2001)
6. Granelli, F., Kliazovich, D., da Fonseca, N.L.S.: Performance Limitations of IEEE 802.11 Networks and Potential Enhancements. In: Xiao, Y., Pan, Y. (eds.) Wireless LANs and Bluetooth. Nova Science Publishers, Hardbound (2005)
7. Lim, H., Xu, K., Gerla, M.: TCP performance over multipath routing in mobile ad hoc networks. In: IEEE International Conference on Communications, vol. 2, pp. 1064–1068 (2003)
8. Chase, D.: Code combining: A Maximum-Likelihood Decoding App. for Comb. an Arbitrary Number of Noisy Packets. IEEE Trans. on Comm. (May 1985)
9. Hoene, C., Rathke, B., Wolisz, A.: On the Importance of a VoIP Packet. In: Proc. of ISCA Tutorial and Research Workshop on th Auditory Quality of Systems, Herne, Germany (April 2003)
10. Bacioccola, A., Cicconetti, C., Stea, G.: User-level Performance Evaluation of VoIP Using ns-2. In: Workshop on Network Simulation Tools (NSTools), Nantes, France (October 2007)
11. Klaue, J., Rathke, B., Wolisz, A.: EvalVid – A Framework for Video Transmission and Quality Evaluation. In: 13th International Conference on Modeling, Techniques and Tools for Computer Performance Evaluation, Urbana, Illinois (2003), http://www.tkn.tu-berlin.de/research/evalvid
12. VLC media player home page, http://www.videolan.org/vlc/
13. NS-2 simulator tool home page, http://www.isi.edu/nsnam/ns/
14. Enhanced UMTS Radio Access Extensions, http://www.ti-wmc.nl/eurane/
15. FFmpeg multimedia system, http://ffmpeg.mplayerhq.hu/