

Selective Packet Prioritization for Wireless Voice over IP

Henning Sanneck¹, Nguyen Tuong Long Le², Martin Haardt¹ and Werner Mohr¹

¹ Siemens AG, Information and Communication Mobile
Networks
D-81359 Munich, Germany

{henning.sanneck, martin.haardt, werner.mohr}@icn.siemens.de

² Department of Computer Science
University of North Carolina
Chapel Hill, NC 27599-3175, USA

le@cs.unc.edu

Abstract

Voice over IP (VoIP) has the potential to be integrated with other Internet applications to provide interactive multimedia communication services that are impossible (or at least very difficult) to deploy over the traditional circuit-switched wired and wireless networks. To fully exploit the benefit of service integration, it is necessary that VoIP services can be seamlessly provided with a good Quality-of-Service over several different network technologies. Due to their wide availability and significant user interest in mobile voice communications, particularly wireless network technologies will be crucial for the success of VoIP (and vice versa).

However, the legacy Internet architecture is based on the "best effort" principle which does not guarantee a minimum amount of packet loss and a minimum delay of packet transmission required for voice communication. Considering a wireless access network to the Internet, QoS assurance is even more difficult. This is due to the inherently shared nature of the medium as well as its high medium error rate which add to the effect of congestion at routers known from wireline networks.

In this paper we first analyze the concealment performance of the G.729 decoder as one prominent example of the CELP coder family which are typically employed for VoIP. Using this result, we then develop QoS support schemes which selectively mark packets to a higher (DiffServ) network priority at the sender dependent on the properties of the speech signal and the expected concealment performance. On the wireless access link the priorities are then mapped to a simple ARQ scheme. This layered approach is in full conformance with existing standards. Objective quality measures (ITU-T P.861A and EMBSD) show that almost the same speech quality as if all packets of the data stream would have been marked as eligible for retransmission can be achieved while significantly reducing the number of actual retransmissions.

Keywords

Wireless Voice over IP, CELP-Based Codecs, Differentiated Services, Selective Packet Marking, Selective ARQ, Objective Speech Quality Measurement

Parts of this work have been performed while Henning Sanneck and Nguyen Tuong Long Le were with GMD Fokus, Berlin, Germany.

1. Introduction

In recent years, both the general public and the research community have been showing significant interest in interactive speech transmission over the Internet (Voice over IP, Internet Telephony). A key advantage over circuit-switched voice is the potential for integration with other Internet services. In addition, the necessary transmission bandwidth can be reduced by silence detection techniques which adds to the statistical multiplexing gain inherent to a packet-switched network. The necessary high complexity speech encoding and decoding can be performed with inexpensive hardware in the end systems. Examples are CELP-based codecs like G.723.1 ([1]) and G.729 ([2]), which are very attractive for Voice over IP because they provide toll quality speech at much lower bit rates (5.3/6.3 kBit/s and 8 kBit/s respectively) than conventional PCM (64 kBit/s). Thus the network resource requirements for a large scale deployment can be reduced significantly, which is particularly important in a wireless networking environment.

However, today's packet-switched networks, like the Internet, are based on the "best effort" principle which does not guarantee any Quality-of-Service parameters. In addition to packet losses caused by the properties of the wireless link (contention for media access, bit errors), speech packets can be discarded when routers or gateways are congested as well as when they arrive late at the receiver (i.e. their playout time has already passed). Furthermore, considering the backward-adaptive coding schemes of the G.723.1 and G.729 source coders, packet loss results in loss of synchronization between the encoder and the decoder. Thus, degradations of the output speech signal occur not only during the time period represented by the lost packet, but also propagate into following segments of the speech signal until the decoder is resynchronized with the encoder. To alleviate this problem, both G.723.1 and G.729 decoders contain an internal (codec-specific) loss concealment algorithm.

In the Internet community, much research has been devoted recently on network layer schemes to enforce an acceptable Quality-of-Service (QoS) for the transmission of real-time multimedia streams (including voice). The work has covered per-flow reservation (Integrated Services, [4]), which allows the assurance of tight bounds on loss and delay, however needs the maintenance of state about every flow inside the network. Another research direction (Differentiated Services, [5]) focuses on only qualitative QoS assurance on a per-packet basis which has better scaling properties by only maintaining state and en-

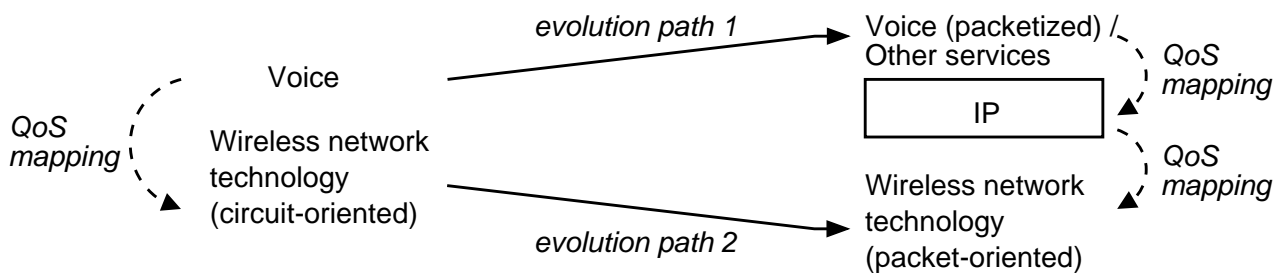


Figure 1: Dual evolution path for wireless Voice over IP.

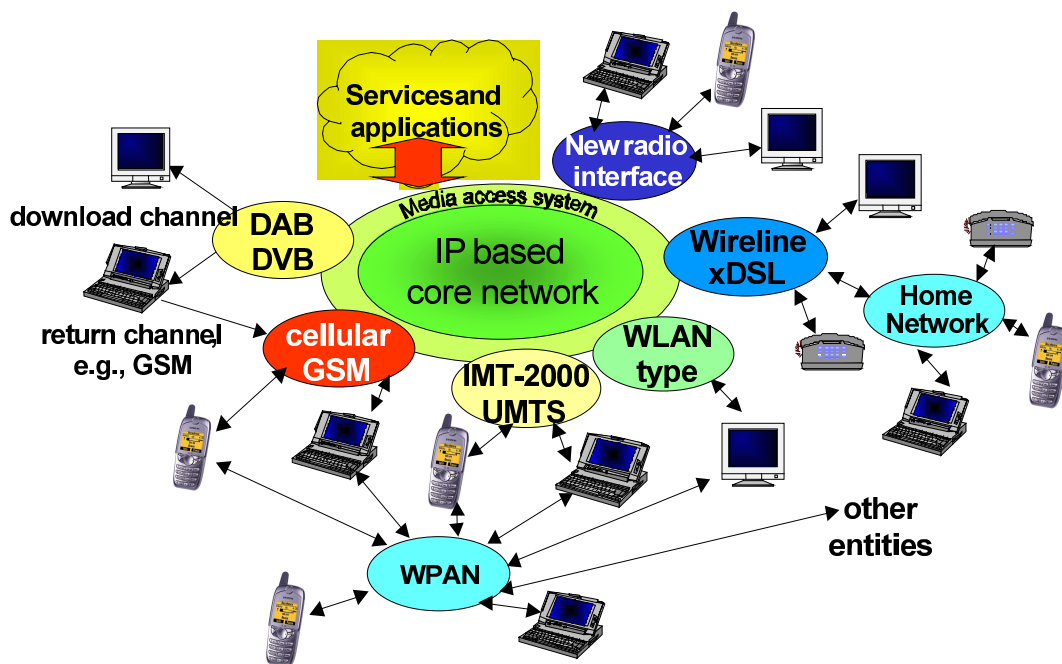


Figure 2: Vision of wireless communications beyond 3G: seamless network of complementary access systems [3].

forcing QoS for aggregated traffic. Besides the advantage of aggregation, per-packet QoS has also the desirable property that an application may control the desired QoS on a per packet (and thus per ADU - Application Data Unit) basis.

While generic network layer QoS schemes are important to ensure true *end-to-end* QoS, we believe that for wireless VoIP an integrated approach covering the codec level, the network layer QoS scheme as well as the considered wireless networking technology needs to be taken. On one hand (seen from a user point of view) this is due to the large user community who is already very much accustomed to using ubiquitous wireless voice communications in second generation cellular networks like GSM (see Fig. 1, evolution path 1). That means that VoIP over wireless must provide at least similar if not better perceived quality at the user level. On the other hand (seen from a perspective of network evolution), the third generation of wireless networks and probably even more the generation beyond ([6, 7], Fig. 2 [3]) are moving towards an IP-oriented architecture (see Fig. 1, evolution path 2).

The rest of the paper is structured as follows: Our approach is based on an analysis of the G.729 frame loss concealment algorithm in section 2. According to the obtained results, in section 3, we develop a selective packet marking / prioritization scheme. Section 4 presents our system architecture where we propose to employ two separate QoS mappings (between the codec and the network level as well as between the network and the wireless link level, cf. Fig. 1). In section 5 we present our simulation results based on linking a simple network model with a real speech codec and an evaluation using objective quality measures. Section 6 discusses related work. Finally, section 7 concludes the paper.

2. Analysis of the G.729 frame loss concealment

G.729 is also known as Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP) and operates at 8 kBit/s. Input data for the coder are 16-bit linear PCM data sampled at 8 kHz. G.729 is based on a model for human speech production. In this model, the throat and the mouth are represented by a linear filter (synthesis filter) and speech signals are produced by exciting this filter with an excitation vector. In G.729, a speech *frame*¹ is 10 ms in duration, corresponding to 80 PCM speech samples. For each frame, the G.729 encoder analyzes the input data and extracts the parameters of the Code Excited Linear Prediction (CELP) model such as linear prediction filter coefficients and excitation vectors. The approach for determining the filter coefficients and the excitation is called analysis by synthesis: The encoder searches through its parameter space, carries out the decode operation in each loop of the search, and compares the output signal of the decode operation (the synthesized signal) with the original speech signal. The parameters that produce the closest match are chosen, encoded, and then transmitted to the receivers. At the receivers, these parameters are used to reconstruct the original speech signal. The reconstructed speech signals are then filtered through a post-processing filter that reduces the perceived noise by emphasizing the spectral peaks (formants) and attenuating the spectral valleys ([8]).

The experiment which has been carried out ([9, 10]) is to

¹We use the term *frame* for the unit of the encoding/decoding operation and *packet* for the unit of transmission. One packet carries typically several frames.

measure the resynchronization time of the decoder after k consecutive frames are lost. The G.729 decoder is said to have resynchronized with the G.729 encoder when the energy of the error signal falls below one percent of the energy of the decoded signal without frame loss (this is equivalent to a signal-to-noise ratio (SNR) threshold of $20dB$). The error signal energy (and thus the SNR) is computed on a per-frame basis. Figure 3 shows the resynchronization time (expressed in the number of frames needed until the threshold is exceeded) plotted against the position of the loss for different values of k . The speech sample is produced by a male speaker where an unvoiced/voiced (w) transition occurs in the eighth frame.

We can see from Figure 3 that the position of a frame loss has a significant influence on the resulting signal degradation², while the degradation is not that sensitive to the length of the frame loss burst k . The loss of unvoiced frames seems to have a rather small impact on the signal degradation and the decoder recovers the state information fast thereafter. The loss of voiced frames causes a larger degradation of the speech signal and the decoder needs more time to resynchronize with the sender. However, the loss of voiced frames at an unvoiced/voiced transition leads to a significant degradation of the signal. We have repeated the experiment for different male and female speakers and obtained similar results. Taking into account the used coding scheme, the above phenomenon could be explained as follows: Because voiced sounds have a higher energy than unvoiced sounds, the loss of voiced frames causes a larger signal degradation than the loss of unvoiced frames. However, due to the periodic property of voiced sounds, the decoder can conceal the loss of voiced frames well once it has obtained sufficient information on them. The decoder fails to conceal the loss of voiced frames at an unvoiced/voiced transition because it attempts to conceal the loss of voiced frames using the filter coefficients and the excitation for an unvoiced sound. Moreover, because the G.729 encoder uses a moving average filter to predict the values of the line spectral pairs and only transmits the difference between the real and predicted values, it takes a lot of time for the decoder to resynchronize with the encoder once it has failed to build the appropriate linear prediction filter. Sun et al. ([11]) have confirmed that our results hold also for the G.723.1 and the AMR codec ([12]). Therefore it can be assumed that it applies to all voice codecs which employ similar coding and loss concealment schemes.

3. Selective packet marking / prioritization

The described result on the ability of the G.729 decoder to conceal packet loss is exploited to develop a new packet marking/prioritization scheme called Speech Property-Based Selective Packet Marking (SPB-MARK) which employs two priorities. The SPB-MARK scheme concentrates the high priority packets on the frames essential to the speech signal and relies on the decoder's concealment for other frames.

Figure 4 shows the simple algorithm written in a pseudo-code that is used to detect a w transition and protect the voiced frames at the beginning of a voiced signal. In the algorithm, the procedure *analysis()* is used to classify a block of k frames as voiced, unvoiced, or w transition. The procedure *send()* is used to send a block of k frames as a single packet with the appropriate priority (either "+1" or "0"). N is a pre-defined value and

²While SNR measures often do not correlate well with subjective speech quality, the large differences in the SNR -threshold-based resynchronization time clearly point to a significant impact on subjective speech quality.

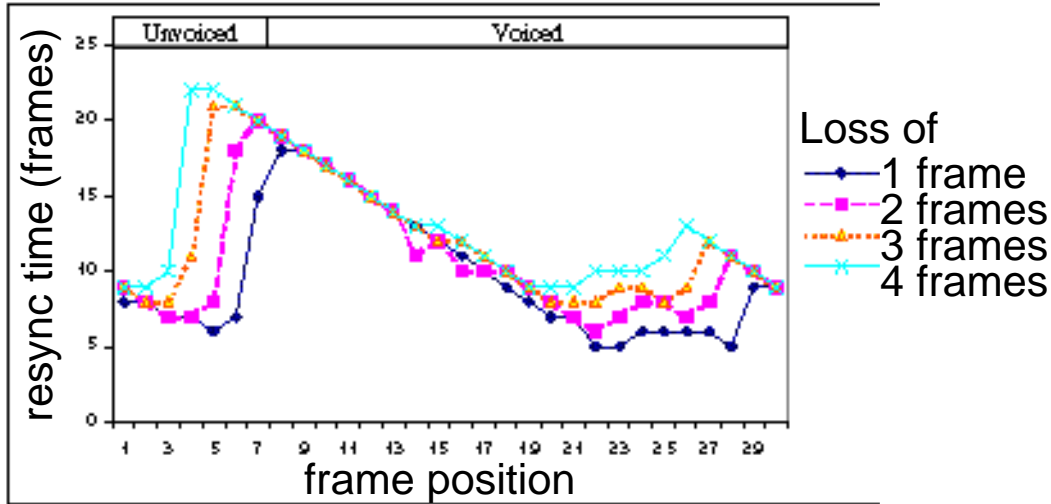


Figure 3: Resynchronization time (in frames) of the G.729 decoder after the loss of k consecutive frames ($k \in [1, 4]$) as a function of frame position.

```

protect = 0
foreach ( $k$  frames)
  classify = analysis( $k$  frames)
  if (protect > 0)
    if (classify == unvoiced)
      protect = 0
      send( $k$  frames, "0")
    else
      send( $k$  frames, "+1")
      protect = protect -  $k$ 
    endif
  else
    if (classify == uv_transition)
      send( $k$  frames, "+1")
      protect =  $N - k$ 
    else
      send( $k$  frames, "0")
    endif
  endif
endfor

```

Figure 4: SPB-MARK Pseudo Code

defines how many frames at the beginning of a voiced signal are to be protected. Our simulations have shown that the range from 10 to 20 are appropriate values for N (depending on the network loss condition). In the simulation presented in section 5, we choose $k = 2$, a typical value for interactive speech transmissions over the Internet (20ms of audio data per packet). A larger number of k would help to reduce the relative overhead of the protocol header but also increases the packetization delay and makes sender classification and receiver concealment in case of packet loss (due to a large loss gap) more difficult.

The network priorities are then enforced by the appropriate traffic shaping/policing mechanisms at the network nodes and/or are mapped to available lower layer traffic control mechanisms. Here we consider a mapping on error control mechanisms of a wireless link layer.

4. System Architecture

Section 2 has shown that some segments of the signal are essential to the speech quality while others, in the event of a packet loss, can be extrapolated well at the receiver from data received earlier. This knowledge has been exploited in the previous section by the design of the SPB-MARK algorithm. Thus, the requirements of the voice application from the network in terms of the reliability of packet delivery can be reduced. This appears to be particularly useful for wireless networks, where losses may be caused by channel fading in addition to congestion (due to contention for the media access).

Therefore, it is important that a voice application can make its QoS requirement known to the network on a per-packet (rather than per-flow) basis. However, this should be done in a generic way at the network layer. This allows to convey QoS requirements known only at the source to other QoS enforcement entities in the network (particularly a wireless last hop). Additionally it is possible to map the per-packet QoS requirements to different networking technologies. The Differentiated Service architecture ([5]) developed within the IETF provides such a QoS assurance on a per-packet basis.

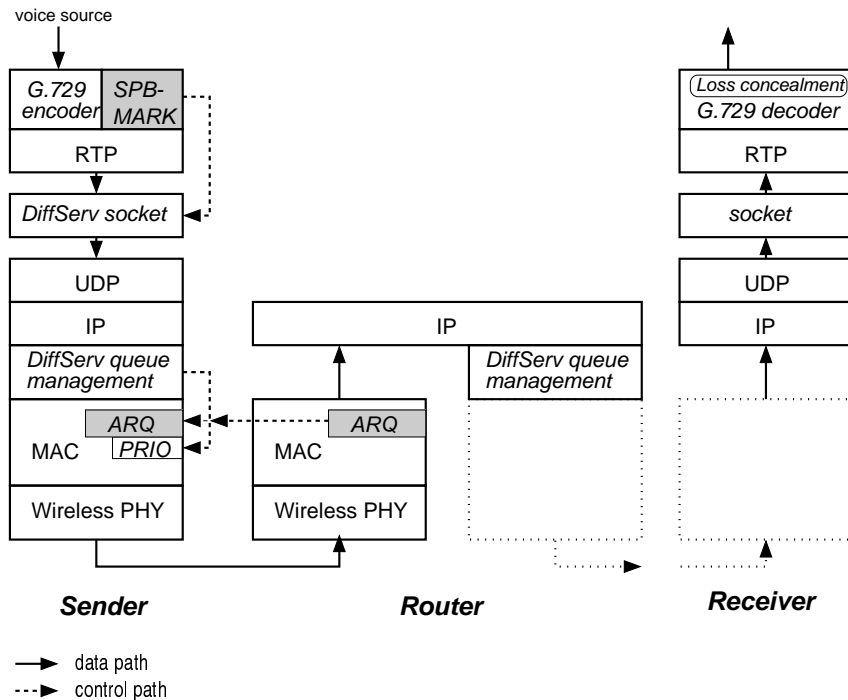


Figure 5: System architecture.

Figure 5 shows the considered system architecture³. We assume a “hot-spot” network scenario (e.g. using an IEEE 802.11 WLAN, [13]) where the interface above the MAC level is indistinguishable from an IP interface on a fixed network. Therefore (and due to the higher bandwidth compared to a cellular network) a conventional VoIP application and protocol stack can be employed.

In the architecture, DiffServ packet priorities are mapped to QoS control mechanisms of the layers below. This is particularly important for wireless link layers, because prioritized packets not only need to be protected against other flows using the shared medium (PRIO interface in Fig. 5) but also need to be protected against channel errors using ARQ and/or FEC⁴. This mapping then in turn has the implication that the real-time constraint of the application has to be taken into account (particularly when retransmitting packets). Thus, a comprehensive approach covering an application-, internetworking- and link-level view is enabled.

5. Results

Figure 6 shows the simulation setup consisting of a sender and receiver connected over a single wireless link. It comprises the following major components (note that only a small subset of the functionality of the components / protocols shown in brackets needs to be implemented for the simulation):

- the ITU-T reference implementation of the G.729 en-

³Note that the figure only shows the application of the proposed scheme (SPB-MARK, ARQ) at the first hop while the ARQ scheme will typically also operate at a wireless last hop.

⁴We only consider channel errors here (see e.g. [14, 15] for mapping DiffServ to link-level priority in a 802.11 WLAN to prioritize the media access).

coder and decoder⁵.

- the SPB-MARK algorithm.
- the ARQ algorithm: at the wireless link layer, only the packets selected by the prioritization scheme (SPB-MARK) are retransmitted. Therefore this scheme is called SPB ARQ. We compare the results to the FULL ARQ method, i.e. every packet is eligible for retransmission. For both schemes we consider a simple send-and-wait ARQ mechanism with a limit on the maximum possible number of retransmissions of a single packet due to the tight real-time constraint of voice.
- an error model for a single wireless link: we employ a simple Bernoulli model for bit errors which lead to frame corruptions and finally packet losses (the effects of packet loss due to buffer overflow in routers are studied in [16, 10]). The packet loss probability can be computed as follows:

$$p_L = 1 - (1 - BER)^{\frac{s}{hT}} \quad (1)$$

BER is the Bit Error Rate and s is the voice packet size at the PHY level. For our simulations we have used a value of 944 bit for s : the header sizes are 24, 34, 20, 8, 12 bytes at the PHY, MAC, IP, UDP and RTP-layer respectively (no header compression is used). The packet payload comprises 20 bytes corresponding to $k = 2$ voice frames, cf. Fig. 4.

- objective speech quality measurement for a comparison between the decoded voice data distorted by the voice encoding / decoding procedure with and without the simulated channel errors. We employ novel objective quality

⁵<http://www.itu.int/itudoc/itu-t/rec/g/g700-g799/software/g729>

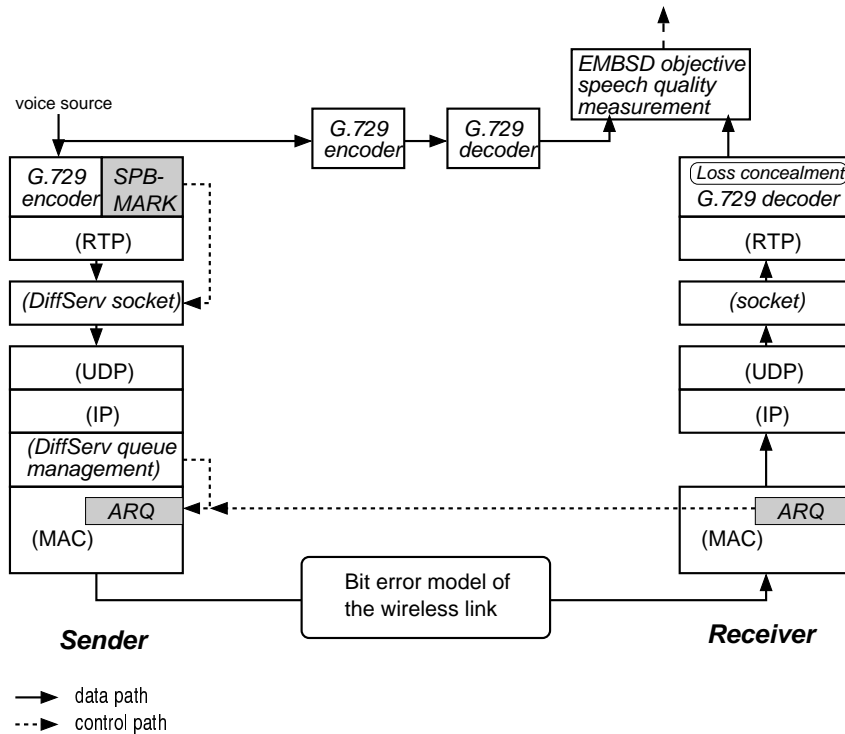


Figure 6: Simulation setup.

measures which attempt to estimate the subjective quality as closely as possible by modeling the human auditory system. Thus we can avoid the necessity to do extensive subjective testing which is a time consuming procedure yielding results which are difficult to reproduce. In our evaluation we use the Enhanced Modified Bark Spectral Distortion (EMBSD, [17]) and the Measuring Normalizing Blocks (MNB, [18]) described in the Appendix II of the ITU-T Recommendation P.861 ([19]). These two objective quality measures are reported to have a very high correlation with subjective tests ([20]), their relation to the range of subjective test result values (MOS) is close to being linear and they are recommended as being suitable for the evaluation of speech degraded by transmission errors in real network environments such as bit errors and frame erasures⁶.

The simulation setup allows us to map a specific PCM signal input⁷ together with network model parameters to a speech quality measure. While using a simple loss characterization for the wireless link, we generate a large number of loss patterns by using different seeds for the pseudo-random number generator (for the presented results we used 300 patterns for each simulated condition). This procedure takes into account that the input signal is not homogenous, i.e. a packet loss burst within one segment of the speech signal can have a largely different perceptual impact than a loss burst of the same size within another segment.

⁶We only show EMBSD results, however, as the results using the MNB measure are virtually identical.

⁷The speech material we have used contains different male and female voices. The length is 11.25s. The sample can be obtained at http://www.dvsinc.com/speech/orig_dam.zip.

Figure 7 shows results for the perceptual distortion (i.e. the user-level quality), where the Bit Error Rate (BER) and the maximum possible number of retransmissions (r_{max}) of a single packet are varied. Also, a scale is shown for the corresponding MOS (Mean Opinion Score) value for subjective tests ranging from 1 (unsatisfactory quality) to 5 (excellent quality).

Judging from the vertical distance between the curve for $r_{max} = 0$ (no retransmissions) and the curves for the ARQ schemes, we observe that generally retransmissions can improve the perceptual quality significantly. For $r_{max} = 1$, the performance of SPB ARQ is very similar to that of the FULL ARQ scheme. With increasing r_{max} the performance gap between SPB ARQ and FULL ARQ for higher bit error rates is increasing. This gap, however, must be seen in relation to the necessary number of retransmissions to achieve the perceptual quality figures shown. Fig. 8 therefore gives the number of actual retransmissions normalized to the total number of packets of a flow. (100% would mean that the total number of packets transmitted including retransmissions has doubled). Here it can be seen that the number of retransmissions for SPB ARQ for low bit error rates is similar to the FULL ARQ method, however for an increasing BER, the increase in the number of retransmissions is much lower than with FULL ARQ. Also, when increasing r_{max} the necessary overall number of retransmissions for SPB ARQ is much lower than for the FULL ARQ scheme.

Summarizing we suggest that for both ARQ schemes at least one retransmission should be allowed for the higher priority packets. For SPB ARQ, r_{max} can be safely set to a value of 3 also for higher BER values (close to 10^{-3}). For such high bit error rates this avoids the significant decrease in perceptual quality (which occurs for low values of r_{max} with SPB ARQ) as well as the explosion in the number of retransmissions (which

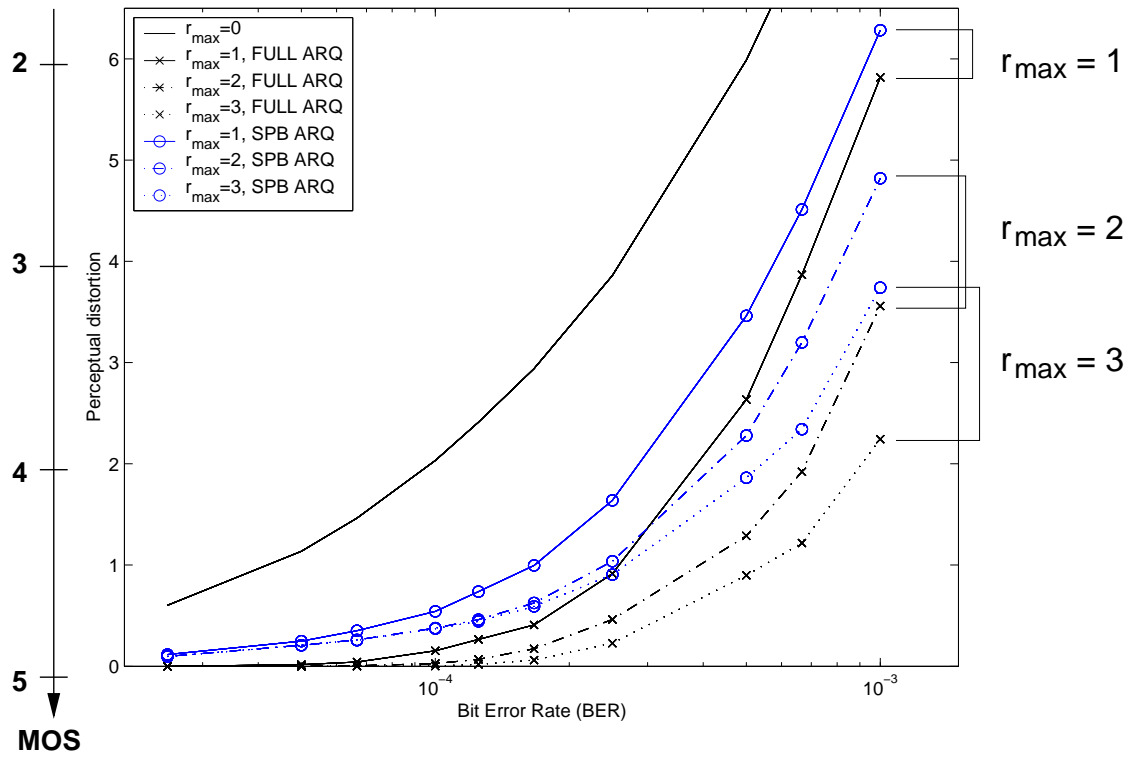


Figure 7: Perceptual distortion (EMBSD) of the ARQ schemes.

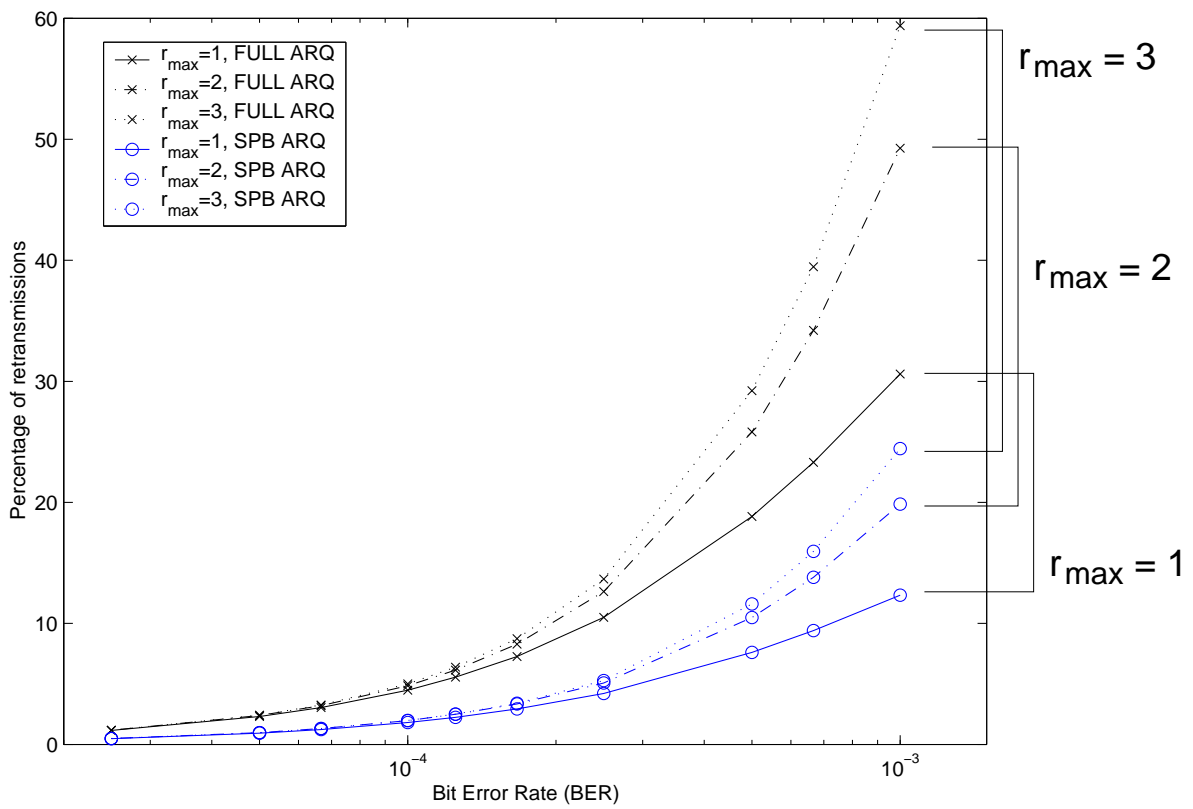


Figure 8: Relative number of retransmissions for the ARQ schemes.

occurs for high values of r_{max} with FULL ARQ). Thus SPB ARQ is offering a reasonable cost (additional processing to derive the packet priority, necessary retransmissions) versus quality tradeoff, particularly for high bit error rates. It should also be noted that a lower number of retransmissions reduces the probability of congestion in the wireless network and improves the delay performance by avoiding unnecessary retransmissions which would block the transmission of the next high priority packet.

6. Related work

While there is a rich literature on loss recovery for packet voice, in particular for VoIP ([21]), only recently similar investigations in the area of wireless VoIP have begun.

In [22] the authors present an analytical evaluation of the performance of packet voice transmission using the point coordination function (PCF) of an 802.11 WLAN ([13]). They show that (while the PCF being a CBR transmission mode which does not exploit voice activity detection schemes) a reasonable number of concurrent voice calls can be accommodated at an access point. However, they clearly also point out the need for error recovery by FEC, ARQ or loss concealment (in fact our approach aims at a good cooperation of the latter two methods) due to the high error rates experienced.

Hoene, Carreras and Wolisz ([23]) also present a scheme which builds on the result presented in section 2. In addition to a similar scheme to ours, which allows the retransmission of higher priority packets at the link layer, they also employ redundant transmissions at the application level as well as a combined solution. In contrast to our architecture, however, where two QoS mappings are performed (Figs. 1 and 5), they propose to directly map the application-level preference to the link level. This avoids any modification (priority marking) to the IP packets (it is a transparent “protocol booster” approach). However, this also limits the applicability of the scheme to the first hop (cf. the discussion in section 4). Hoene et al. furthermore present measurement results using an 802.11 WLAN (in DCF mode - distributed coordination function) in a configuration where losses are introduced by actual channel errors as well as a simulated play-out buffer (and not by contention at the MAC level), also demonstrating a performance gain when employing an SPB-triggered packet protection.

7. Conclusions

We have investigated the impact of packet loss at different positions within a speech signal on the perceived quality for the G.729 codec. It has been shown that the loss of voiced frames after an unvoiced/voiced transition leads to a significant degradation of the speech quality while the loss of other frames is concealed rather well by the decoder’s concealment algorithm.

Based on this result we have developed a selective packet prioritization scheme (SPB-MARK) that protects the packets which are essential to the speech quality by marking them with a higher DiffServ priority, while relying on the decoder’s concealment in case other low priority packets are lost. The priorities are applied at the wireless link layer to identify and retransmit only these essential packets. This architectural separation allows to convey QoS requirements known only at the source to other QoS enforcement entities in the network (particularly a wireless last hop can thus apply similar error control mechanisms as the wireless first hop). Additionally it is possible to map the per-packet QoS requirements to different networking

technologies.

Simulations using a simple network model and subsequent evaluation using objective speech quality measures show that even just a single possible retransmission try enhances significantly the perceptual quality in the presence of bit errors. Particularly for higher bit error rates, the proposed selective ARQ scheme (SPB ARQ) avoids a significant decrease in perceptual quality as well as the explosion in the number of retransmissions (which occurs if every packet is eligible for retransmission). Thus SPB ARQ is offering a reasonable cost versus quality tradeoff.

Future work includes the extension of the link layer error control scheme with Forward Error Correction (hybrid FEC/ARQ). Also, the described marking algorithm could be mapped to prioritized medium access (e.g. by modifying the length of the IEEE 802.11 DIFS inter-frame spacing time period). Future evaluations of speech quality should also take advantage of advanced results in the objective speech quality measurement domain (like the forthcoming ITU-T standard P.862, [24]).

8. Acknowledgments

The valuable comments by Adam Wolisz (TU Berlin, GMD Fokus) and Georg Carle (GMD Fokus) on an earlier version of this work are gratefully acknowledged.

9. References

- [1] International Telecommunications Union, “Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s,” Recommendation G.723.1, ITU-T, March 1996.
- [2] International Telecommunications Union, “Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP),” Recommendation G.729, ITU-T, March 1996.
- [3] W. Mohr and W. Konhäuser, “Access network evolution beyond third generation mobile communications,” *IEEE Communications Magazine*, December 2000.
- [4] R. Braden, D. Clark, and S. Shenker, “Integrated Services in the Internet Architecture: an Overview,” RFC 1633, IETF, 1994, <ftp://ftp.ietf.org/rfc/rfc1633.txt>.
- [5] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Services,” RFC 2475, IETF, December 1998, <ftp://ftp.ietf.org/rfc/rfc2475.txt>.
- [6] Wireless World Research Forum, *Proceedings of the WWRF Kick-Off Meeting*, Alcatel, Ericsson, Nokia, Siemens, Munich, Germany, 2001, <http://www.wireless-world-research.org>.
- [7] Wireless World Research Forum, *Proceedings of the WWRF Second Meeting*, Alcatel, Ericsson, Nokia, Siemens, Helsinki, Finland, 2001, <http://www.wireless-world-research.org>.
- [8] D. Minoli and E. Minoli, *Delivering Voice over IP Networks*, John Wiley and Sons, 1998.
- [9] H. Sanneck and N. Le, “Speech property-based FEC for Internet Telephony applications,” in *Proceedings of the SPIE/ACM SIGMM Multimedia Computing and Networking Conference (MMCN)*, San Jose, CA, January 2000, pp. 38–51, <ftp://ftp.fokus.gmd.de/pub/glone/papers/Sann0001:Speech-FEC.ps.gz>.
- [10] Henning Sanneck, *Packet Loss Recovery and Control for Voice Transmission over the Internet*, Ph.D. thesis, GMD Fokus / Telecommunication Networks Group, Technical University of Berlin, October 2000, <http://sanneck.net/research/publications/thesis/Sann0010.Loss.pdf>.

- [11] L. Sun, G. Wade, B. Lines, and E. Ifeachor, "Impact of packet loss location on perceived speech quality," in *Second IP-Telephony Workshop*, New York, NY, April 2001, pp. 114–122.
- [12] J. Sjoberg et al., "RTP payload format and file storage format for AMR and AMR-WB audio," Internet draft, IETF Audio-Video Transport Working Group, 2001, <ftp://ftp.ietf.org/internet-drafts/drafts-ietf-avt-rtp-amr-08.txt>.
- [13] IEEE, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE Standard 802.11, IEEE, June 1999.
- [14] I. Aad and C. Castelluccia, "Differentiation mechanisms for IEEE 802.11," in *Proceedings IEEE INFOCOM*, Anchorage, Alaska, April 2001.
- [15] M. Barry, A. Campbell, and A. Veres, "Distributed control algorithms for service differentiation in wireless packet networks," in *Proceedings IEEE INFOCOM*, Anchorage, Alaska, April 2001.
- [16] H. Sanneck, N. Le, A. Wolisz, and G. Carle, "Intra-flow loss recovery and control for VoIP," in *Proceedings ACM Multimedia*, Ottawa, Canada, September 2001, http://sanneck.net/research/publications/papers/Sann0109_Intra-Flow-VoIP.pdf.
- [17] W. Yang and R. Yantorno, "Improvement of MBSD by scaling noise masking threshold and correlation analysis with MOS difference instead of MOS," in *Proceedings ICASSP*, Phoenix, AZ, March 1999, pp. 673–676.
- [18] S. Voran, "Objective estimation of perceived speech quality - part i: Development of the measuring normalizing block technique," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 4, pp. 371–382, July 1999.
- [19] International Telecommunications Union, "Objective quality measurement of telephone-band (300-3400 Hz) speech codecs," Recommendation P.861, ITU-T, February 1998.
- [20] S. Voran, "Objective estimation of perceived speech quality - part ii: Evaluation of the measuring normalizing block technique," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 4, pp. 383–390, July 1999.
- [21] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet-loss recovery techniques for streaming audio," *IEEE Network Magazine*, Sept./Oct. 1998.
- [22] M. Veeraraghavan, N. Cocker, and T. Moors, "Support of voice services in IEEE 802.11," in *Proceedings IEEE INFOCOM*, Anchorage, Alaska, April 2001.
- [23] C. Hoene, I. Carreras, and A. Wolisz, "Voice over IP: Improving the quality over wireless LAN by adopting a booster mechanism - an experimental approach," in *work in progress draft*, Berlin, Germany, June 2001.
- [24] International Telecommunications Union, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Draft Recommendation P.862, ITU-T COM 12-5-E, ITU-T, November 2000.