Multiple Descriptions and Path Diversity for Voice Communications over Wireless Mesh Networks

Jagadeesh Balam and Jerry D. Gibson

Department of Electrical and Computer Engineering

University of California, Santa Barbara, California 93106-9560

Email: balam@ece.ucsb.edu and gibson@ece.ucsb.edu

Telephone: (805) 403–7066, Fax: (805) 893–2930

Abstract

A key feature of wireless mesh networks is that multiple independent paths through the network are available. Multiple descriptions coding is often suggested as a source coding scheme to take advantage of this path diversity. We compare multiple description (MD) coding with path diversity (PD) against a full-rate single description (SD) coder without PD, and two simple PD methods of 1) repeating a half-rate SD coder over both paths and 2) repeating the full-rate parent SD coder over the two paths. We first present a theoretical analysis comparing the average distortion per symbol in packetized communication using the above mentioned MD and PD methods to transmit a memoryless Gaussian source over additive white Gaussian noise channels. Next, using two new MD speech coders with balanced side descriptions derived from the AMR-WB and G.729 standards, we evaluate delivered voice quality using PESQ-MOS and compare MD coding against the PD methods for random and bursty packet losses. Both the theoretical analyses and the speech coding experiments show that with packet overheads, the simple PD methods may be preferable to MD coding. A new performance measure that incorporates both quality and bit-rate is shown to account for the tradeoffs more explicitly.

I. Introduction

Wireless mesh networking is a promising technology to provide low cost wireless coverage over wide areas. A wireless mesh network [1] (WMN) consists of fixed nodes (mesh routers) and mobile nodes (mesh clients). The mesh clients can also forward packets of other clients, i.e. the clients also form an ad-hoc network with nearby clients and routers. Because of their ease of deployment, WMNs are projected as a solution for broadband home networking, enterprise networking and emergency situations. With growing usage of Voice over Internet Protocol (VoIP) over 802.11 WLANs, wireless mesh networks are also promising for voice communications. Voice communications over IEEE 802.11 based WMNs is challenging because the 802.11 standard is designed

This research has been supported by the California Micro Program, Applied Signal Technology, Dolby Labs, Inc., Mindspeed, and Qualcomm, Inc., by NSF Grant Nos. CCF-0429884 and CNS-0435527, and by the UC Discovery Grant Program and Nokia, Inc.

primarily for non-real time transfer of data. IEEE 802.11 MAC protocols are designed to minimize collisions and depend on retransmissions to ensure successful transmission, irrespective of the delay incurred by the packet or the number of voice calls supported. Interactive voice communications cannot tolerate large delays (one way end-to-end delay, as specified by the ITU-T recommendation G.114, should be under 200 ms for users to be "very satisfied") and in multi-hop communications, delays occur at each node due to MAC, physical, and network layer protocols. Further, voice quality is affected by packet losses due to bit errors in the wireless channel. The mesh network architecture provides increased robustness by allowing transmission over multiple paths. We look at this key feature of WMNs and investigate multiple description coding as a solution to use multiple paths efficiently.

Using a path diversity scheme not only improves fault tolerance but also reduces overall packet losses and end-to-end delays. However, sending multiple copies of the same packet is inefficient usage of bandwidth. To improve bandwidth efficiency, a source coding diversity method such as multiple description (MD) coding can be used along with path diversity. In MD coding, multiple descriptions/bit-streams of the source are created in such a way that each description can be used to reconstruct the source with acceptable quality and two or more descriptions can be combined to give a better quality reconstruction.

Path diversity has been proposed for increased bandwidth and improved end-to-end connection robustness for video transport over wireless networks in [2], [3]. The authors in [2] compare MD coding and layered coding (LC) when used with path diversity and observe that MD coding is preferable when the paths are symmetric, while in [3] the authors conclude that LC does better only when the base layer can be transmitted error free or with very low error rates. Path diversity has been shown to have significant benefits over conventional single path transmission in terms of reduced packet loss rate and improved video quality for wireless video transmission in [4].

While MD coding has been successful for some video and image applications, voice communications has two distinctive characteristics that distinguish it from images and video. First, since voice communications is usually two-way and latency constrained, the payload sizes are very small compared to the transmission of video and still images. Second, the standardized codecs for many voice communications applications use block-based code excited linear predictive coding (CELP), and it is difficult to obtain balanced side descriptions that perform well at rates near half the full rate with these codecs. The small payload sizes can cause MD coding to lose much of its rate advantage when packet overheads are included. Further, the quality produced by the half rate side descriptions of the standardized CELP codecs is often relatively poor, so having only one description available does not provide acceptable performance.

The performance of the MD coding when the two descriptions are sent over independent symmetric paths is compared against transmitting a full-rate single description (SD) coder without path diversity and two simple path diversity methods of 1) repeating a half-rate SD coder (a coder at a bit rate about the same as each description of the MD coder) and 2) repeating the full-rate parent SD coder over the independent paths. We begin with a theoretical analysis of the problem considering packetized communication of Gaussian sources with source rate dependent packet losses in the channel (to capture the payload length dependence), including the effect of packet overheads on the efficiency of these methods [5]. Then we present experimental results for a similar scenario using MD speech coders. We developed two new MD speech coders - one [6] based on the AMR-WB codec [7], [8] and the other [9] based on G.729 [10]. These coders create two descriptions of equal rate from the bit-stream of a standard single description (SD) coder. The rate of each description is about half the rate of the parent (SD) coder.

The widespread availability of 802.11 based WLANs and the possibility of supporting low-cost wireless voice communications has motivated significant prior work on reliable voice communications over 802.11 based networks. In [11], the authors suggest using multiple interfaces, path diversity, and packet aggregation to increase the number of calls supported by a 802.11 mesh network. Lin et al. [12] suggest using inter-packet redundancy, path diversity and multiple description coding for reduced delay and improved bandwidth efficiency when transmitting speech over an ad-hoc network. The various challenges of voice communications over ad-hoc networks and some possible solutions that include multiple descriptions and path diversity are suggested in [13]. Most efforts have been toward adapting the 802.11 MAC layer for reducing retransmissions and packet losses [14]–[16]. Commercial WLAN phones by *Spectralink* [17] use a priority scheme and zero back-off at the link layer for transmission of speech packets to minimize delay for these packets.

A new standard, IEEE 802.11e, has been approved as an enhancement to the 802.11 MAC for providing different QoS levels for data, voice and video. However, 802.11e does not overcome the degradation in voice quality due to packet losses resulting from noisy communication links or node failures. Cross-layer solutions [18]–[20] that involve interaction between the application layer and the MAC or physical layer have also been suggested.

In this paper, we present a theoretical analysis comparing MD and SD methods with packet overheads taken into consideration. Classical results comparing source coding methods, seldom consider packet overheads. We also describe two new MD speech coders and present experimental results comparing different path diversity methods for voice communications with a new performance measure that combines MOS and effective bit rate. In Section II, we present our theoretical analysis of the problem of communicating memoryless Gaussian sources over two independent channels. We compare MD coding against simple path diversity methods and also show how packet headers result in increased average distortion and reduced efficiency of the MD methods. In the later sections, we see that the experimental results follow the analysis presented in this section. Background on multiple description speech coding is provided in Section III. In Sections IV and V, we describe the two new MD speech coders that we developed. In Section VI, we describe the experiments to find the best path diversity method suitable for communication under different packet loss conditions and discuss the corresponding results. We generate packet loss trace files for different types of packet losses and compare the quality of reconstructed speech when the corresponding frames are dropped for encoded speech using packet loss concealment (PLC).

We do not consider retransmissions due to latency constraints and because we do not wish to overload the access point. As a result, the random packet losses would be due to bit errors in the packet and the packet being discarded due to the failure of the CRC. The bursty errors would be caused by fading or a short term link failure. These simple models of channel error conditions are adequately reflective of more realistic channels for the evaluation of MD coding for path diversity in wireless networks. The different diversity methods considered are compared with respect to the quality of delivered speech in terms of MOS (mean opinion score) calculated using PESQ (Perceptual Evaluation of Subjective Quality) [21]. A new performance measure is introduced in Section VII that captures the quality of the delivered speech and at the same time penalizes the method according to the bandwidth required to support it. Finally, we summarize our contributions and conclusions.

II. Multiple Descriptions and Path Diversity

In this section, we consider packetized communication of *i.i.d.* Gaussian sources over independent, parallel, additive white Gaussian noise channels. Even in the most idealized form, speech is modeled as a first-order autoregressive process but we limit our analysis to memoryless sources, as the rate-distortion region for multiple description coding is readily available from literature for these sources and we hope to be able to make some good guesses for speech communication using these results.

Since the error mechanism on the channels is bit error rate, the probability of packet loss on each channel is proportional to the packet size. In such a scenario that should favor methods employing smaller packets, we illustrate how typical large overheads can result in stealing the potential advantage of MD methods. We compare the following four different methods of communication :

1) **Single description (SD) code of rate** R $(bits/symbol)$ **without path diversity**: For transmitting the single description code, we use only one of the available pair of links (Fig. 2(a)).
2) **Multiple description (MD) coding**: We consider a two-description coder (Fig. 1), where each description is of rate $R/2$ $(bits/symbol)$ and the joint description is of rate R.
3) **Path diversity with rate** $R/2$ $(bits/symbol)$ **SD code**: A single description of the source coded at a rate $R/2$ is duplicated over the two available links.
4) **Path diversity with rate** R $(bits/symbol)$ **SD code**: A single description code of rate R is duplicated over the available pair of links (Fig. 2(b)).

Let the source be *i.i.d.* Gaussian with zero mean and unit variance and the distortion measured by the squared error between the source and the reconstructed sample. The packet loss rate p, when independent bit errors are introduced by the channel, is given by

$$p = 1 - (1 - BER)^L \tag{1}$$

where L is the packet length in bits and BER is the bit error rate. If each packet contains a fixed number N of symbols and each symbol is coded at an average rate of R bits per symbol, the packet length L is now related

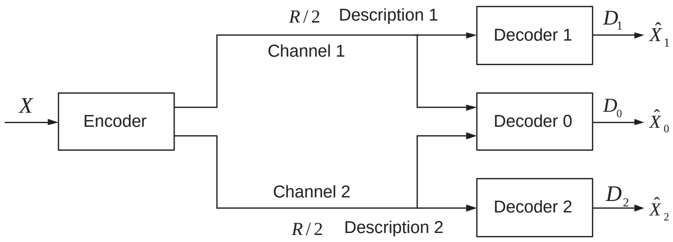


Fig. 1. A two-description coder with each description of rate R/2 sent over two independent channels

to R and N as

$$L\ (bits/packet) = R\ (bits/symbol) \times\ N\ (symbols/packet) \tag{2}$$

A. Multiple Description Coding

The idea of multiple descriptions was posed as a problem of jointly good descriptions by Gersho, Wistenhausen, Wolf, Wyner, Ziv, and Ozarow at the 1979 IEEE Information Theory Workshop. The question posed was that if two individually good descriptions of a stochastic process are sent over an unreliable communication network, then what is the maximum combined information at the receiver when both the descriptions reach the destination [22]. References [22]–[24] give good insights into the achievable rate regions for multiple description codes for Gaussian sources.

Figure 1 is an illustration of a two-description coder where the source is coded into two descriptions of rate $R/2$ each and transmitted separately over two independent links. When only description I (II) is received, the distortion is D_1 (D_2), and when both the descriptions reach the receiver, the central decoder reconstructs the source with a distortion D_0. We consider a symmetric coder where each side description is of the same rate and each gives the same fidelity reconstruction of the source.

1) Two Cases of MD Coding:

a) No Excess Marginal Rate: The individual descriptions of rate $R/2$ are rate-distortion optimal with distortion $D_1 = 2^{-R}$ and the lower bound on the distortion for the joint description (D_0) of rate R for a sequence of i.i.d. Gaussian random variables with unit variance and squared-error distortion measure is given by [23]

$$D_0 \geq \frac{2^{-R}}{2 - 2^{-R}} \tag{3}$$

b) No Excess Joint Rate: In this case, the joint description at rate R is rate-distortion optimal with $D_0 = 2^{-2R}$ and the lower bound on the distortion at the side decoders for i.i.d. Gaussian sources is given by [23]

$$D_1 \geq \frac{1}{2}(1 + 2^{-R}) \tag{4}$$

2) Optimal MD coding: The achievable distortion region for a Gaussian source with unit variance and a fixed rate R ($R/2$ for each description), using MD coding is given by [25]

$$\begin{aligned} D_1 &\geq 2^{-R} && (5)\\ D_0 &\geq 2^{-2R} && (6)\\ (D_0, D_1) &= (a, \frac{1+a}{2} - \frac{1-a}{2}\sqrt{1-\frac{2^{-2R}}{a}}) && (7) \end{aligned}$$

$for\ a\ \in [2^{-2R}, 2^{-R}/(2-2^{-R}]$ where D_0 is the distortion at the central decoder and D_1 is the distortion at the side decoders. For a packet loss rate p, the average distortion achieved at the receiver using a two-description coder is

$$D_{MD} = (1-p)^2 D_0 + 2p(1-p)D_1 + p^2 \tag{8}$$

From Eqs. (7) and (8), we get [25]

$$D_{MD} = (1-p)^2 a + 2p(1-p)(\frac{1+a}{2} - \frac{1-a}{2}\sqrt{1-\frac{2^{-2R}}{a}}) + p^2 \tag{9}$$

For each R and p, we can find the value of 'a' in Eq. (9) that gives the minimum average distortion. This minimum distortion is only achievable when the sender knows *a priori* the packet loss rate and hence can choose the best MD coding method. This gives us a lower bound on the distortion achieved using MD coding but practically achieving this lower bound for changing p's is not possible when information about the channel is not known at the encoder.

For MD coding, we consider three cases: 1) the no-excess joint rate case (MD-NJR), 2) the no-excess marginal rate case (MD-NMR) and 3) the optimal case that gives minimum average distortion for each value of p (MD-OPT).

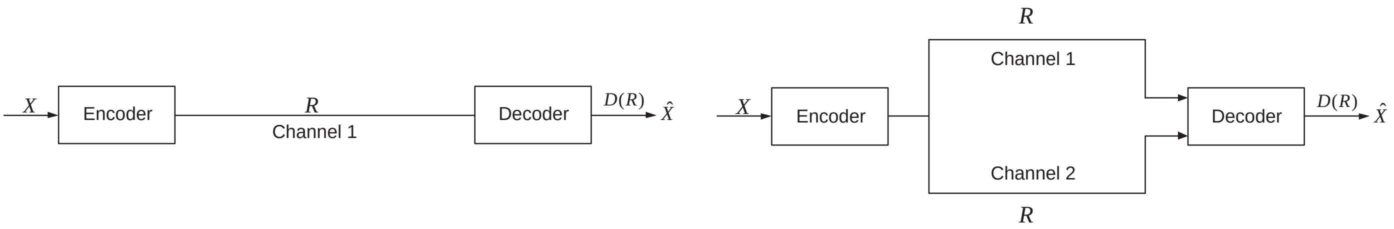
B. Single Description Coding and Path Diversity

The other three methods of communication we consider (methods 1, 3, 4 listed in the previous subsection) involve the use of SD coding. Henceforth, we call an SD coder that operates at rate R with an optimal distortion of $D_{FR} = 2^{-2R}$ as the full-rate (FR) coder and an SD coder that operates at $R/2$ with optimal distortion $D_{HR} = 2^{-R}$ as the half-rate (HR) coder.

The average distortion for each of the communication methods that involve an SD coder, with probability of packet loss p is given as follows:

Single description of rate R without path diversity (SD)

$$D_{SD} = (1-p)2^{-2R} + p \tag{10}$$



(a) A single description coder with the coded stream sent over a single channel

(b) A single description coder of rate R duplicated over the parallel channels

Fig. 2. Single description coding communication methods

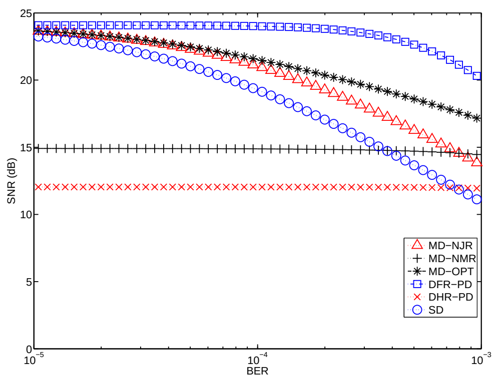


Fig. 3. SNR for the different communication methods considered when source rate is fixed at $R = 4$ and no overheads are added to the packet payload

Half-rate coder with path diversity (DHR-PD)

$$D_{DHR-PD} = (1-p)^2 2^{-R} + 2p(1-p)2^{-R} + p^2 \tag{11}$$

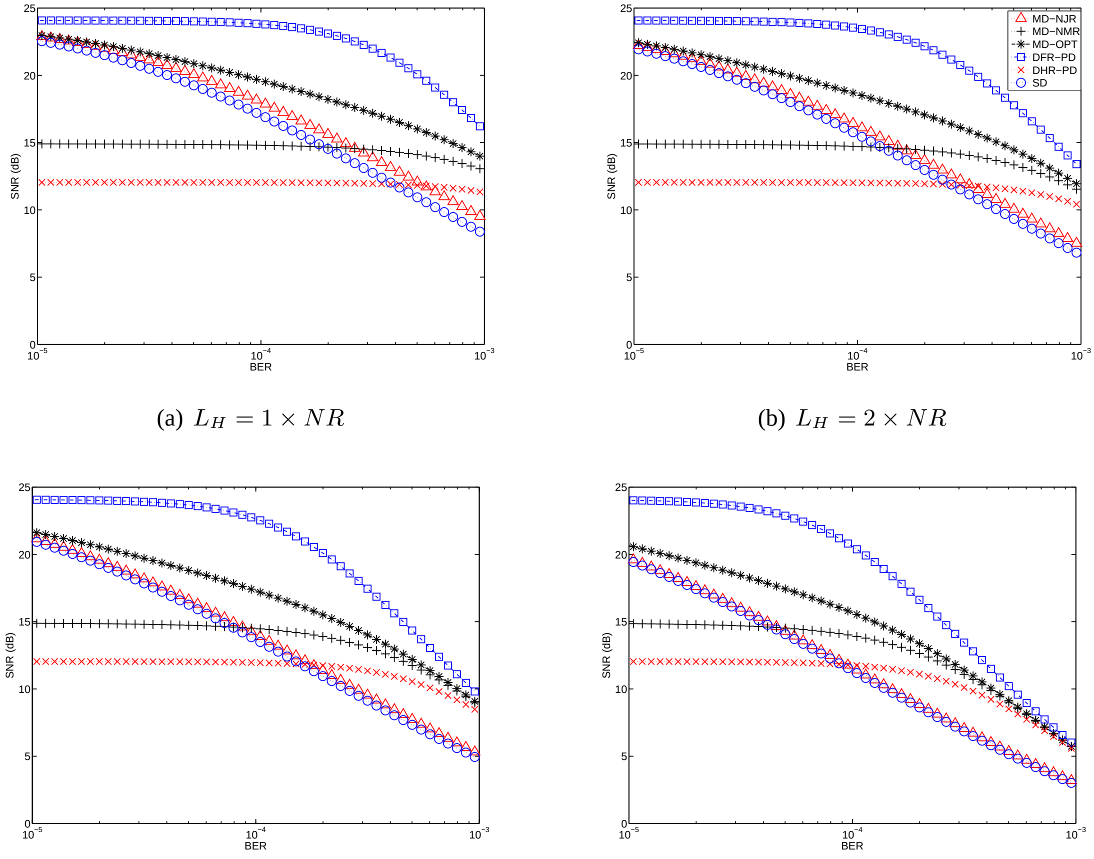
Full-rate coder with path diversity (DFR-PD)

$$D_{DFR-PD} = (1-p)^2 2^{-2R} + 2p(1-p)2^{-2R} + p^2 \tag{12}$$

C. Effect of Packet Losses

We show the effect of packet losses on the performance of each of the communication methods mentioned above through the SNR obtained at the receiver, where SNR for the unit variance Gaussian sources is calculated as $10 * log10(\frac{1}{D_{av}})$ where D_{av} is the average distortion at the receiver. For our illustrations we pick a rate $R = 4 \; bits \; per \; symbol$ and assume that each packet contains 20 symbols, resulting in 80 bits per packet. Such packet lengths are common in packet based voice communications using low bit-rate codecs such as G.729.

In Fig. 3, we compare the performance of each of the methods for different bit error rates in the channel. DHR-PD $(\cdot \cdot \times \cdot \cdot)$ has the worst performance because of the higher distortion at the encoder for the half rate coder and because the SNR for DHR-PD does not increase even if both links successfully deliver packets. SD



(a) $L_H = 1 \times NR$

(b) $L_H = 2 \times NR$

(c) $L_H = 4 \times NR$

(d) $L_H = 8 \times NR$

Fig. 4. SNR for different communication methods for high BERs and different header rates. Full rate, $R = 4$. Source rate for each description of MD coders and HR coders is $R/2 = 2$.

$(\cdot\cdot\bigcirc\cdot\cdot)$, MD-NJR $(\cdot\cdot\triangle\cdot\cdot)$, MD-OPT $(\cdot\cdot\divideontimes\cdot\cdot)$, and DFR-PD $(-\cdot\square-\cdot)$ all give high SNRs at low BERs. This is because when there are no losses in the channel, SD, MD-NJR and DFR-PD should result in essentially the same distortion due to encoding and MD-OPT should coincide with the no excess joint rate MD case, MD-NJR. Observe that DHR-PD and MD-NMR $(\cdot\cdot+\cdot\cdot)$ show very little decrease in SNR at high BERs also. This is because these methods have duplicate $R/2$ bitstreams on both paths and the SNR at the receiver is almost the same even if only one of the links successfully delivers most of the packets. From Equation (3), for MD-NMR, we can see that at high rates the best distortion that can be achieved at the joint decoder is only half the distortion achieved at the side decoders. On the other hand we see a consistent decrease in SNR for SD, because SD completely fails when the single link carrying the source information fails and the distortion at the receiver is maximum. Similarly, for MD-NJR, if one of the links fails, then the SNR decreases considerably because the side descriptions are not optimal. DFR-PD shows only a small deviation in performance as BER increases because of path diversity.

D. Packet headers

In packet based networks, headers are added to each source packet by other protocol layers to facilitate communication over the network. Such overheads affect the probability of packet loss, p, since they increase the length of the packet. If the number of header bits per packet is L_H , then p is given by

$$p = 1 - (1 - BER)^{NR+L_H} \tag{13}$$

We now investigate the behavior of each of the methods for different values of L_H. We consider different payload-header ratios ($NR : L_H$) $1:1$, $1:2$, $1:4$, $1:8$. Such ratios typically occur in voice communications over IEEE 802.11 based WLANs. For example, each packet sent in the transmission of G.729 [10] encoded speech contains 10 bytes (10 ms of speech) or 20 bytes (20 ms of speech) of payload and around 68 bytes of overhead. Speech encoded with AMR-WB [7] at 12.65 kbps contains 32 (20 ms of speech) bytes of payload and 68 bytes of overhead per packet, and a G.711 [26] packet contains 80 (10 ms of speech) or 160 bytes (20 ms of speech) of payload and 68 bytes of overhead.

E. Effect of headers on packet losses and distortion

In Fig. 4 we show the SNR plots for different header sizes. Observe that increasing header sizes worsen the performance of all the methods. MD-NMR ($\cdot\cdot + \cdot\cdot$) outperforms MD-NJR ($\cdot\cdot \triangle \cdot\cdot$) at high BERs as expected, because at high loss rates, only one of the descriptions reaches the receiver for a majority of the time and individual descriptions are optimal in MD-NMR but not in MD-NJR in this case. The performance of MD-OPT ($\cdot\cdot \divideontimes \cdot\cdot$) approaches that of MD-NMR as the BER increases for high values of L_H. This is because, as L_H increases, p increases and only one of the descriptions reaches the destination most of the time. With only one description reaching the receiver, the best distortion that can be achieved at a rate $R/2$ is $D = 2^{-R}$ and this is exactly what each description of MD-NMR achieves.

Although p for DFR-PD ($- \cdot \square - \cdot$) is larger than that of all the other methods except SD ($\cdot\cdot \bigcirc \cdot\cdot$), because of a higher rate on each link ($NR + L_H$ against $NR/2 + L_H$ for other methods), its SNR is higher than any of the other methods. The gain due to a higher source rate and path diversity for DFR-PD is large enough to overcome a higher packet loss rate p. If we compare SD and DFR-PD, there is a difference of about 9.2 dB in the SNRs at BERs around 10^{-4}, for $L_H = 8NR$, and all of this gain for DFR-PD can be attributed to path diversity. The most significant point to note here is that after these large headers are added, the effective rate of DFR-PD ($2 \times (NR + L_H)$) differs from any MD method or DHR-PD ($\cdot\cdot \times \cdot\cdot$) ($NR + 2 \times L_H$) by only R, and so for example, when $L_H = 8NR$, DFR-PD requires less than a 6% increase in bandwidth compared to the MD and DHR-PD methods. For such a small increase in bandwidth, the gain in quality achieved using DFR-PD is quite significant.

Therefore, we observe that when headers dominate the packet size, as can occur for voice communications, the smaller payloads and reduced rate due to MD coding do not provide the expected performance advantage over SD or other path diversity methods. Comparisons between the theoretical results for memoryless Gaussian sources presented in this section and the experimental results for speech presented in Sec. VI are discussed in Sec. VI.G.

III. Multiple Description Speech Coding

A. Background

The initial motivation for multiple description coding of speech was to overcome the problem of link outages and provide uninterrupted telephone service [27]. One of the earliest multiple description coders for speech was introduced by Jayant and Christensen [28] for waveform coders. The interest in MD coding of speech resurfaced with the advent of internet telephony in the 1990s, where MD coding was proposed as an efficient solution to overcome packet losses over the internet. Ingle and Vaishampayan [29] extend the approach of simple odd/even separation of samples for DPCM sytems. Jiang and Ortega [30] suggest constructing two packets by coding odd and even samples separately using PCM or ADPCM and then injecting redundancy into each packet by adding coarse information about the missing samples. In [31] small modifications to existing encoders are suggested that have been shown to work with PCM, ADPCM and LD-CELP coders and the resultant multiple description system is backward interoperable with existing single description decoders. Voran [32] proposes using two different two-dimensional structured vector quantizers to code a pair of G.711 PCM codes. Some MD coders of speech that use transform coding have also been suggested in literature [33], [34].

Many methods have been proposed for creating multiple descriptions using Code Excited Linear Prediction (CELP) based codecs. In [35], two descriptions are created from CELP coded speech by including base or important information that allows an acceptable reproduction of speech in each packet. A subset of enhancement information is added to each packet so that when both packets are received, a finer reproduction is possible. Wah and Dong [36] present a zero redundancy multiple description coding method that uses the correlations in LSPs (Line Spectrum Pairs) of adjacent frames. The excitation is generated for a larger subframe and the same codeword is replicated in all the descriptions. Zhong and Juang [37] propose a novel approach to MD coding by using regular single description coders for the side descriptions and then introducing diversities to generate non-redundant data between the descriptions. In [38], a new MD coder based on the AMR-WB [7] codec is presented that creates two descriptions by dividing the AMR-WB bit-stream into two sub-streams.

B. Our Work

We designed two MD coders, one based on the AMR-WB codec [7] and the other based on the G.729 codec [10]. These coders are extensions of the MD coder introduced in [38]. While the MD coder in [38] was designed as a minimum redundancy coder, our new coders are designed to create balanced descriptions, i.e. each side description is of the same rate, and speech decoded from either description is of similar quality. The descriptions are created in such a way that the missing information when one description is lost can be concealed by simple interpolation of the information in the received description. This MD coder design approach of dividing the bit-stream of the parent coder into balanced sub-streams is similar to the no-excess joint rate case of MD coding, where the individual descriptions can be combined to give an optimal joint description. The difference from the classical case is that, in a no-excess joint rate MD coder, the joint description is optimal at the combined rate of the side descriptions. Here, we take an optimized bit stream and increase the effective joint rate by introducing redundancy in the side descriptions. The distortion at the central decoder is still the same as the full rate SD decoder, but the effective bit-rate is higher due to the redundancy introduced in the side descriptions. Of course, the quality delivered by each side description will be worse than that of an SD codec optimized for the same rate as each individual side description.

Figure 5 shows a block diagram of the proposed MD coding based system. The MD encoder consists of a standard CELP encoder (AMR-WB or G.729) and a bit-stream division block which divides the bit-stream from the CELP encoder into two sub-streams. Each sub-stream (description) is sent over an independent path in the network. The decoder consists of the standard CELP decoder and two more decoder blocks in parallel that are used when only one of the descriptions is received.

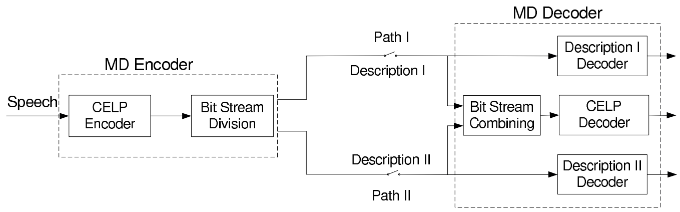


Fig. 5. Block diagram of our MD coder for speech based on CELP codecs

In both the AMR-WB and G.729 standard codecs, each speech frame is divided into sub-frames for estimating various parameters required for CELP coding. Except for the LPC coefficients, all the other parameters are determined on a per sub-frame basis. Therefore, our approach is to use a sub-frame as our basic unit in allocating bits to different descriptions. However, the LPC coefficients are calculated on a frame basis, and they are coded using split-multistage vector quantization in both AMR-WB and G.729. We take advantage of this structure and divide the bits corresponding to the LPC coefficients in such a way that the LPC coefficients can at least be reconstructed coarsely at the side decoders.

In the next two sections we describe in detail the two coders, namely, MD-AMR (MD coder based on the AMR-WB codec) and MD-G.729 (MD coder based on the G.729 codec).

IV. MD-AMR: A Multiple Description Speech Coder based on AMR-WB speech codec

A. AMR-WB

The AMR-WB speech codec was selected in December 2000 for GSM and the third generation WCDMA mobile communication system for providing wideband speech services. It was also selected as recommendation

G.722.2 by the ITU-T. AMR-WB operates on speech covering the band from 50 Hz to 7000 Hz. Traditionally, speech codecs were designed for narrowband telephone bandwidth 200 to 3400 Hz speech, but the evolution of broadband multimedia services has spawned an increased interest in wideband speech. Wideband speech sounds more natural and is more intelligible than the traditional narrowband speech. The AMR-WB speech codec is an ACELP (Algebraic Codebook Excited Linear Prediction) based codec and operates on 20 ms speech frames. The AMR-WB codec operates in nine different modes (0-8) with bit-rates ranging from 6.6 to 23.85 kbps. Mode 2 at 12.65 kbps is the lowest rate at which the AMR-WB codec offers high quality speech [7].

B. MD-AMR: Encoder

The bit splitting for various parameters in the MD-AMR encoder is described below:

a) ISF: The LPC coefficients, computed once per frame, are encoded (after conversion to Immitance Spectral Frequencies (ISFs)) using split-multistage vector quantization. Specifically, the ISFs are coded using a two stage split vector quantizer in AMR-WB mode 2. The 16 bits that are used to code the index of the the code vectors in the first stage are included in both the descriptions, because without this information, LPC parameters cannot be reconstructed at all. The next stage vector is divided into 5 sub-vectors and the five sub-vectors are coded with $6+7+7+5+5=30$ bits. The bits corresponding to the first sub-vector are included in both the descriptions, while the bits corresponding to the second and the fifth sub-vectors are included in Description I only. The remaining bits corresponding to the third and fourth sub-vectors are included in Description II only. This way of splitting was experimentally determined to give the most symmetric quality at the two descriptions.

b) Pitch Delay for Adaptive codebook: In AMR-WB, pitch delay is calculated on a sub-frame basis, but the second sub-frame pitch delay is differentially encoded with respect to the first sub-frame and the fourth sub-frame pitch delay is encoded using the pitch delay for the third sub-frame. Without the first (third) sub-frame pitch delay, the second (fourth) sub-frame bits are useless, so the first and second sub-frame bits are included in Description I and the third and fourth sub-frames are included in Description II.

c) Adaptive and Fixed codebook gains: Gains for the adaptive and fixed codebooks are jointly quantized with seven bits in each sub-frame in the AMR-WB codec. We include bits corresponding to the first and the third sub-frame gains in Description I, and bits corresponding to the second and the fourth sub-frame are included in Description II. This is done because when only one description is received, the missing gain information is concealed using the previous sub-frame information. If only Description I is received, the second sub-frame gains are concealed using the first sub-frame gains and the fourth sub-frame gains are concealed using the third sub-frame information.

d) Fixed codebook Indices: The fixed codebook vector is coded with 36 bits per sub-frame. Here again, first and third sub-frame bits are included in Description I and second and fourth sub-frame bits are included in Description II.

TABLE I

BIT ALLOCATION FOR THE MD CODEC BASED ON AMR-WB

ISP	Stage 1: ***8 8***				I,(II)
	Stage 2: ***6*** 7 (7) (5) 5				34,(34)
	1st sf	2nd sf	3rd sf	4th sf	
VAD					1,(1)
LTP-filtering	1	1	(1)	(1)	2,(2)
Pitch delay	9	6	(9)	(6)	15,(15)
Algebraic Code	36	(36)	36	(36)	72,(72)
Gains	7	(7)	7	(7)	14,(14)
Total					138,(138)

Table I shows the bit allocation for the two descriptions. The table as a whole shows the bit allocation for each parameter in the bit stream of AMR-WB, mode-2 (12.65 kbps). The numbers within the parentheses indicate that the corresponding bits belong only to Description II and the bits corresponding to the emphasized (bold) numbers are replicated in both the descriptions. The remaining bits belong only to Description I. The VAD flag is included in both the descriptions.

The bit-rate for each description is 6.9 kbps and the bit-rate for the combination is 13.8 kbps, of which 1.15 kbps is redundant. The redundant bit-rate is the penalty paid to make the distortion at the side decoders acceptable. However, each description sounds worse than AMR-WB at 6.6 kbps since they are obtained by splitting the rate of a higher rate codec compared to AMR-WB@6.6 kbps which is optimized to give the best quality at that rate.

C. MD-AMR: Decoder

When both the descriptions created using the MD-AMR encoder are delivered at the receiver, the bit streams are combined to form the AMR-WB bit stream and the AMR-WB decoder is used to reconstruct the signal. When both descriptions are lost, AMR-WB packet loss concealment is used to conceal the lost packet. When only one description is received, the missing bits when compared with the AMR-WB bit stream are substituted using information from the most recent frame received.

The decoding process of either description is similar. In the following points, we summarize decoding only using Description I when Description II is lost:

- The sub-vectors corresponding to the missing bits in the ISP indices from Description II are ignored and not added in the second stage of the vector quantizer
- The pitch-lag values of the third and fourth subframes are substituted with the pitch-lag value of the second subframe (available from Description I)
- The LTP-filtering flag of the third and fourth subframe is set to be the same as that of 2nd subframe
- The fixed codebook vector of the second (fourth) subframe is set to be the same as that of the first (third) subframe and the gains for the second (fourth) subframe are set to be first (third) subframe gains attenuated

by 3 dB

Performance results are included in Sec. VI.

V. MD-G.729: A Multiple Description Speech Coder based on G.729

A. G.729

ITU-T G.729 [10] is a conjugate-structure algebraic code-excited linear-predictive (CS-ACELP) coder intended primarily for wireless communications. Today, G.729 is a widely adopted codec in most of the voice over IP and voice over WLAN products. The G.729 codec encodes narrowband speech (speech sampled at 8 kHz) and delivers toll quality speech at the rate of 8 kbps. The encoder operates on 10 ms speech frames and each speech frame is divided into two subframes and all the parameters except the LPC coefficients are determined once per subframe. The LPC coefficients are determined once per frame. This structure is very similar to the AMR-WB codec and we use a similar principle in creating two descriptions from the G.729 bit stream.

B. MD-G.729: Encoder

The MD-G.729 coder creates two descriptions of the same average rate from the bit-stream of the G.729 codec. In order to keep the effective average bit rate of each description the same (4.6 kbps), odd and even numbered frames in each description are coded with a different number of bits. Tables II shows the bit allocations for odd and even frames in each of the descriptions.

The bits corresponding to the pitch delay are included only in alternate frames in each description. The pitch delay for the second subframe in each frame is differentially encoded with respect to the first subframe. Without the first subframe pitch delay, the second subframe pitch delay cannot be decoded. Hence, pitch delay information for both the subframes always has to be included together in one description. For Description I, the 14 bits for adaptive-codebook delay are included in odd-numbered frames and for Description II, these bits are included in even-numbered frames.

Each description has 13 bits allocated to the Line Spectrum Pairs (LSPs). G.729 uses multi-stage split vector quantization to quantize the LSP vector. In the first stage, the vector is not split and 8 bits are used to code the vector. These 8 bits are included in both the descriptions for all frames. This allows for a coarse reconstruction of the 10-dimensional residual vector of LSPs in either description. In the second stage of the vector quantizer, the 10-dimensional residual vector is split into two 5-dimensional sub-vectors and each sub-vector is coded using 5 bits. For odd (even) numbered frames, the codebook index for the first (second) subvector is included only in Description I while the codebook index for the second (first) subvector is included only in Description II. This is done to make the descriptions more symmetric with respect to quality. Experiments revealed that the degradation in the reconstructed speech was greater when the first subvector was removed rather than when the second sub-vector was removed.

TABLE II
BIT ALLOCATION FOR MD-G.729

	Description I					Description II				
	Odd Frame		Even Frame		Sum	Odd Frame		Even Frame		Sum
Frame Indicator	2(00)		2 (01)		4	2(10)		2 (11)		4
LSP	Stage 1: ***8***		Stage 1: ***8***		26	Stage 1: ***8***		Stage 1: ***8***		26
	Stage 2: 5 0		Stage 2: 0 5			Stage 2: 5 0		Stage 2: 0 5		
	sf 1	sf 2	sf 1	sf 2		sf 1	sf 2	sf 1	sf 2	
Pitch delay	9	5	0	0	14	0	0	9	5	14
Fixed Codebook	13	0	13	0	26	0	13	0	13	26
Fixed Codebook Signs	4	0	4	0	8	0	4	0	4	8
Gains	7	0	7	0	14	0	7	0	7	14
Total					92					92

The bits corresponding to the fixed codebook vector and signs of the fixed codebook for the first subframe of all frames are included only in Description I and the fixed codebook information for the second subframe is included only in Description II. The adaptive codebook and the fixed codebook gains for the first (second) subframe are included only in Description I (II). Thus, each odd numbered frame for Description I gets 51 bits from the G.729 bit-stream while Description II gets 37 bits. Similarly, for even numbered frames Description I contains 37 bits and Description II contains 51 bits. Two frame indicator bits are added to indicate the description to which the bit-stream belongs and whether the frame is odd or even numbered. Bit pair '00' indicates that the bit-stream belongs to an odd numbered frame of Description I, '01' indicates Description I and even frame, '10' indicates Description II and odd frame and '11' indicates Description II and even frame.

C. MD-G.729: Decoder

When both the descriptions are received at the decoder, the two descriptions are combined to give the bit-stream of G.729. If both the descriptions are lost, then the frame error concealment algorithm of G.729 [10] is used to conceal the lost frame. If only one of the descriptions is received, then the decoder substitutes the missing information by using the received parameters in the description or information from the most recent correctly received frame as follows. The LSP vectors are constructed from the received first stage vector and one of the received subvectors. The missing second stage subvector is assumed to be zero. The pitch delay in an even (odd) frame in Description I (II) is constructed from the previous received frame's pitch delay increased by 1. This process is same as that used for frame error concealment in the G.729 codec. The missing gain information in the second subframe for Description I and the first subframe for Description II is substituted by an attenuated version of the previous subframe. The memory of the gain predictor is also attenuated in a manner similar to that used in G.729 error concealment.

VI. MD and Path Diversity Peformance Comparisons

The objective of our experiments is to compare the performance of the two MD coders described in the previous section against a single description sent over a single path and duplicates of a single description coder sent over independent paths (path diversity). We call the single description coder upon which each MD coder is designed as the full rate (FR) coder. For the MD-AMR coder, we use mode-2 of the AMR-WB coder (12.65 kbps) as the FR coder and mode-0 of the AMR-WB coder (6.6 kbps) as the half rate (HR) coder. The different cases we compare our new MD-AMR coder against are 1) sending the FR coder over a single path (FR-SD), 2) duplicating the FR coder over two independent paths (DFR-PD) and 3) duplicating the HR coder over two independent paths (DHR-PD). For the MD-G.729 coder we choose G.729 at 8 kbps as the FR coder. We compare MD-G.729 only against FR-SD and DFR-PD based on G.729 since there is no suitable half rate coder available for comparison.

A. Setup

We assume that two independent paths are available for using path diversity. We follow the 802.11 concept wherein a speech packet is dropped if even one of the bits in the packet is in error. Similar to the model suggested in [12], no retransmissions are allowed in the network and the MAC layer does not use an acknowledgement packet to know whether a packet was successfully delivered. For such a scenario with two paths of the same relaibility, we study the quality of speech delivered by the above mentioned communication methods. We consider two kinds of packet losses, 1) random packet losses due to random bit errors in the channel and 2) bursty packet losses due to phenomena like fading or shadowing in the network or other factors like a link failure. We do not explicitly consider losses due to contention and collisions. However, this does not limit the current work since no MD or SD method offers an advantage, and all methods investigated would be equally vulnerable. For random errors, the bit error rate (BER) is assumed to be the same on both paths. We use six different (3 male, 3 female) speech files, each around 8 seconds long, in our experiments. Each speech file consists of two different sentences spoken by the same speaker. The quality of the decoded speech is evaluated using WPESQ [39] for the wideband experiments and PESQ [21] for narrowband G.729 based experiments. The predicted MOS from PESQ and WPESQ is mapped to subjective MOS using different mapping functions. The details of the mapping functions used are given below.

1) PESQ: PESQ (Perceptual Evaluation of Speech Quality) is an ITU standard for objective speech quality measurement of narrowband speech. PESQ compares the degraded signal decoded at the receiver with the reference signal and gives a score between -0.5 and 4.5. PESQ scores have been found to correlate well with subjective MOS scores. PESQ-LQ was then shown to be a good predictor of subjective listening quality in [40]. PESQ-LQ gives a mapping function to map the PESQ scores to an average P.800 MOS scale. The mapping function is given by [40]

$$y = \begin{cases} 1.0, x \leq 1.7 \\ -0.157268x^3 + 1.386609x^2 - 2.504699x + \\ 2.0233454, x > 1.7 \end{cases} \tag{14}$$

where x is the PESQ score and y is the corresponding mapping to LQ MOS.

2) WPESQ: WPESQ (Wideband PESQ) is an extension to ITU-T P.862, proposed in [39], to adapt PESQ for use in measuring wideband speech quality. The difference between WPESQ and PESQ is only the input filter characteristics, since the psychoacoustic model and the error model are the same. We use an implementation based on this proposal to evaluate quality of speech in our wideband experiments. Note that WPESQ is not an approved standard of the ITU as yet. Barriac et al. propose a function for mapping WPESQ scores to subjective MOS values in [41], as

$$y = 1 + \frac{4}{1 + e^{-2x+6}} \tag{15}$$

where x is the WPESQ score and y is the corresponding mapped value.

B. Speech Quality Indicator MOS_{90}

In our experiments, for each packet loss rate, 250 different packet loss patterns are used to drop frames in the speech files. There is a large variation in the MOS values predicted by PESQ for different loss patterns of the same packet loss rate. This is because some frames are perceptually more important and some frames, such as the transition frames, are not concealed as well as the other frames. In such a case, taking an average MOS value calculated by taking the mean of all the realizations for a given PLR does not give a MOS value that is indicative of the user experience. It was observed in [42] that an average MOS value might only be achieved for only 50% of the realizations, i.e. an average MOS value is only an indicator of quality guaranteed for 50% of the time. We need a performance measure that is indicative of the quality for a majority of the conversation time. We choose a performance measure suggested in [42], MOS_{90} , which is the MOS value that can be achieved for at least 90% of the realizations. The MOS_{90} is a better indicator of quality delivered to the user than average MOS.

C. Packetization

We assume that each packet sent over the network contains one coded speech frame. In an 802.11 based network, large delays can occur at each intermediate node because of various factors like contention for the channel or link failure. To allow for the unpredictable delays in the network, we keep the packetization delay at the minimum of one frame. Also, having more frames per packet impairs the performance of the packet loss concealment algorithm since one lost frame is more effectively concealed than two or more successive lost frames. For AMR-WB based experiments, each packet contains data corresponding to 20 ms of speech, and for the G.729 based experiments, each packet contains data corresponding to 10 ms of speech.

D. Random packet losses

We consider packet losses that occur due to random bit errors in the channel. The relationship between packet loss probability and bit error rate (BER) in the channel when independent bit errors are introduced in the channel is given by

$$p = 1 - (1 - BER)^L \tag{16}$$

where L is the packet size in bits. For different BERs, we first find packet loss probabilities (p) and for each p, trace files are created using 250 different seeds of the random number generator. Frames are dropped in the encoded speech files using the trace files created. First, we compare the various communication methods in the classical scenario for comparing source coding methods where only the encoded speech frames are transmitted over the network and no overheads are taken into account in the analysis. Next, we compare the performance after including typical packet overheads such as the transport and MAC layer headers.

E. Random losses and No Packet Headers: Experimental Results

1) MD-AMR Experiments: As stated above, MD-AMR is compared against FR-SD (Full Rate - Single Description), DFR-PD (Duplicate Full Rate with Path Diversity) and DHR-PD (Duplicate Half Rate with Path Diversity). Each description of the MD-AMR coder is sent over the two independent paths available and this method is designated as MD-PD (Multiple Descriptions with Path Diversity)

TABLE III
FRAME SIZES FOR THE CODECS CONSIDERED

Codec	Frame Size (bits)
AMR-WB @ 12.65 kbps (FR)	253
AMR-WB @ 6.6 kbps (HR)	132
MD-AMR	138
G.729	80
MD-G.729	53 or 39

First, we look at a scenario where there are no packet headers added to the speech payloads. The packet sizes are determined by the encoded frame sizes of the speech coder. The packet size for each method is listed in Table III. The channel has random bit errors and the packet loss probability is p given by Eq. (16). FR-SD ($\cdot\cdot\triangle\cdot\cdot$) and DFR-PD ($- - \divideontimes - -$) have the largest p for a given BER, because they use the FR codec with largest encoded frame sizes, while p's for MD-PD (—+—) and DHR-PD ($-\cdot\Box-\cdot$) are almost equal because the difference in their packet sizes is very small. In Fig. 6, we plot the $WPESQ - MOS_{90}$ values for changing BERs. Note that when there are no packet losses, FR-SD, MD-PD and DFR-PD deliver the same information at the receiver, i.e. the bit stream of speech coded with the standard AMR-WB coder, and hence should have the same MOS values. Observe that at the lowest BER of 10^{-5}, there is already a loss of performance in FR-SD, while MD-PD and DFR-PD provide a quality equal to that achieved without any packet losses. At a BER of 10^{-5}, there is no loss in the performance of DFR-PD as the packet loss rate is small and path diversity ensures that at least one path successfully delivers the packet all the time. MD-PD is also not affected at this BER because of the small packet size and a corresponding very small packet loss rate ($\approx .14\%$).

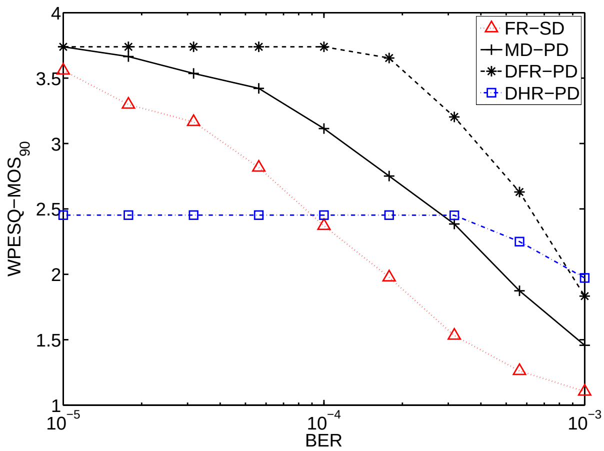


Fig. 6. MD-AMR comparisons: MOS_{90} values for changing BER when NO packet headers are added

The bandwidth required by FR-SD, MD-PD and DHR-PD is almost the same, and among these three methods, MD-PD is a clear winner except at very high BERs, when DHR-PD starts doing better than MD-PD. This is because at such high BERs, only one of the paths successfully delivers for most of the time. In such a case, only one description of MD-AMR is delivered in MD-PD while DHR-PD delivers the single description corresponding to AMR-WB@ 6.6 kbps. As already mentioned in Section III, a single description of the MD-AMR coder sounds worse than speech coded using AMR-WB@6.6 kbps. DFR-PD delivers the best quality of speech but at a penalty of required bandwidth for transmission. Bits required to be sent for DFR-PD are almost double that of either MD-PD or DHR-PD. MD-PD has an advantage over FR-SD for only a small increase in the bits transmitted.

2) MD-G.729 Experiments: The same speech files used for wideband experiments were used for the narrow-band experiments after modified IRS filtering and downsampling. The filter module from ITU-T's software tools library (STL) provided with ITU-T G.191 [43] was used for filtering and downsampling. For each BER, each file was evaluated for 250 different seeds of the random number generator. We compare a single description coder (G.729) with a single path (FR-SD), MD-G.729 with path diversity (MD-PD) and a duplicated full-rate single description coder (G.729) with path diversity (DFR-PD). We did not find a suitable half-rate codec to consider DHR-PD for these comparisons.

As listed in Table III, when no packet headers are considered, the FR-SD and DFR-PD methods transmit 80 bits per packet, since they use the G.729 codec, while the MD-PD method sends either 53 or 39 bits per packet on each path. For a given BER, it is obvious that FR-SD and DFR-PD have a higher packet loss probability than

MD-PD because of the larger packet size.

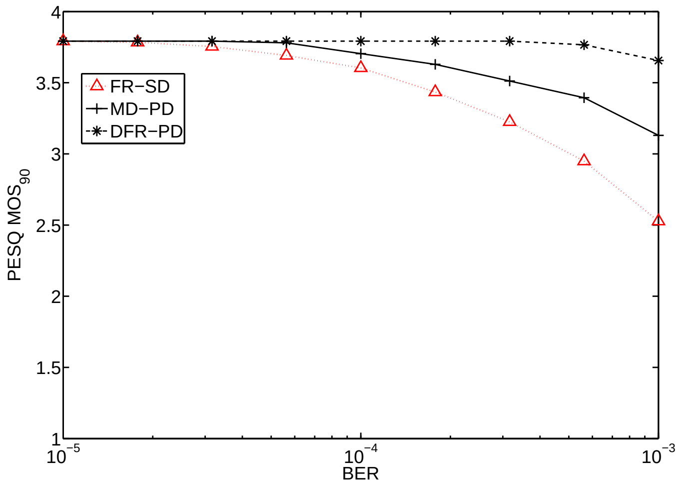


Fig. 7. MD-G.729 comparisons: MOS_{90} values for changing BER when NO packet headers are added

Figure 7 shows the speech quality delivered by each of the methods for increasing BERs. The results are comparable to the AMR-WB results in Fig. 6, where MD-PD (—+—) performs better than FR-SD ($\cdot\cdot\triangle\cdot\cdot$) and DFR-PD ($- - \divideontimes - -$) delivers the best quality, albeit at a penalty of the larger bandwidth required to transmit two full rate streams at 16 kbps compared to 8 kbps required for FR-SD and 9.2 kbps required for MD-PD. The drop in MOS_{90} values for increasing BERs in Fig. 7 is less steep than the decrease noticed in Fig. 6 because the G.729 packets are smaller than the AMR-WB packets.

F. Random losses with Packet Headers: Experimental Results

Now we consider a more realistic scenario where headers are added to the speech packets by the lower protocol layers. In a typical 802.11 based wireless network, headers are added by RTP, UDP, IP and the 802.11 MAC layer protocol. The overheads for each packet add up to 68 bytes (the 802.11 MAC (28 bytes), IP (20 bytes), UDP (8 bytes) and RTP (12 bytes)), significantly larger than the payload, which is a maximum of 32 bytes in our experiments. For path diversity, the overheads are even larger because for each frame, we need to send 68 bytes of packet headers on both the paths. The difference in the payloads of the MD codec and the FR-SD codec becomes insignificant now since the packet length and the effective data rate of MD-PD is almost double that of FR-SD.

1) MD-AMR: Table IV shows the effective packet sizes of each coder for the AMR-WB based experiments after the inclusion of packet headers. Payloads are padded with zeros to form complete octets.

The performance of all the methods with these new packet sizes for changing BERs is shown in Fig. 8. Observe from Figs. 6 and 8 that the overall performance of all the methods drops because of the increased packet loss rates (PLR) resulting from the larger packet sizes. For small BERs, the ordering of the methods with respect to their performance is the same in the two figures. The performance of MD-PD (—+—) drops below that of

TABLE IV
PACKET SIZES WITH HEADERS

Codec	Packet size (bytes)
AMR@12.65kbps (FR)	100(32+68)
AMR@6.6 kbps (HR)	85(17+68)
MD-AMR	86(18+68)

DHR-PD ($- \cdot \Box - \cdot$) at a smaller BER in Fig. 8, because higher PLRs result in only one path delivering the packets for a majority of the time, and in such a scenario, the HR coder performs better than the MD-AMR coder. DFR-PD ($- - \divideontimes - -$) now performs significantly better than MD-PD for a large range of BERs and this improved performance requires only a 16% increase in the bits transmitted per path. Also, compared to the case with no headers (Fig. 6), the advantage of MD-PD over FR-SD ($\cdot \cdot \triangle \cdot \cdot$) narrows considerably.

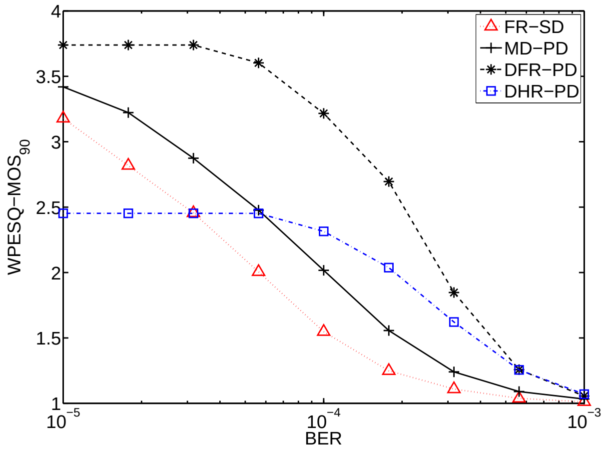


Fig. 8. MD-AMR comparisons: MOS_{90} values for changing BER with packet headers

2) MD-G.729: With the packet headers, the effective packet sizes are 78 bytes for G.729 and 75 or 73 bytes for MD-G.729. Sending duplicate copies of G.729 packets over two independent paths (DFR-PD) would require a bit rate of 124.8 kbps $((78 + 78) \times 8/10)$ while MD-PD needs 118.4 $((75 + 73) \times 8/10)$ kbps. For a small increase in required bandwidth, we can send two copies of G.729 packets instead of sending MD-G.729 packets that have only around half the information as a G.729 packet.

TABLE V
PACKET SIZES WITH HEADERS

Codec	Full headers (bytes)	Compressed Headers
G.729	78	40
MD	75 or 73	37 or 35

From Fig. 9, observe there is a drop in the performance of all the methods compared to the no-header case of Fig. 7, because of larger p's resulting from the larger packet sizes. We see that MD-PD (—+—) performs better than FR-SD ($\cdot\cdot\triangle\cdot\cdot$) but this gain in performance is achieved at a huge penalty (almost 100%) in terms of the bandwidth required for transmission. The packet loss rate experienced by each description of the the MD codec is now almost the same as that of a G.729 packet because the ratio of their packet sizes is close to one. The better performance of MD-PD over FR-SD can be attributed to better error concealment in MD-PD. When a packet is lost in only one of the paths, we need to conceal only about half of the bits in MD-PD, whereas, in the case of FR-SD, no information is received if the single packet is lost in the network. Even after the inclusion of packet headers, DFR-PD ($- - \divideontimes - -$) performs significantly better than MD-PD and this improvement in performance can be achieved at a small percentage increase (about 5.5%) in the number of bits transmitted.

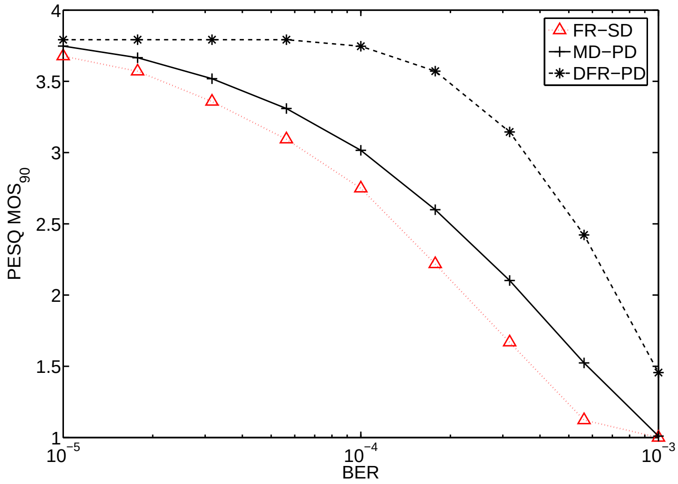


Fig. 9. Comparison of FR-SD, MD-PD, DFR-PD for changing BER (with packet headers)

G. Comparison with Theory

Comparing the theoretical results in Fig. 3 with the experimental results for the AMR-WB codec in Fig. 6 and for the G.729 codec in Fig. 7, we see that even though the theoretical results are for a memoryless Gaussian source, the relative ordering of the coding approaches, namely, DFR-PD best, MD-PD next, and SD the poorest, agree for all three figures. When packet headers are included, the theoretical results in Fig. 4 again compare favorably with the experimental results in Figs. 8 and 9. Even the fact that MD-NJR performance loses much of its advantage over SD is reflected in Fig. 4 as in Figs. 8 and 9, although the theoretical results are more pessimistic than the experimental results. The MD-OPT curve in Fig. 4 is not achievable in practice and there is no counterpart in the experiments. Thus, the simple abstraction of a Gaussian memoryless source transmitted through an additive white Gaussian noise channel subject to the mean squared error distortion measure provides the correct ordering of the methods, and retains the tight connection to the basic information theoretic results as well.

H. Header Compression: A solution?

Using larger payloads by coding longer frame sizes or including more frames per packet might reduce the inefficiency due to the headers, but doing so would result in poorer performance of the concealment algorithm and also increased end-to-end latency, which is a principal concern in conversational voice communications. The best possible solution for the problem of large packet headers today is using a header compression scheme like RoHC (Robust Header Compression). Efforts are underway to make RoHC compatible with IEEE 802.11. Using RoHC, the IP/UDP/RTP headers can be compressed to very small sizes of up to one byte. If we assume an average compressed header size of 2 bytes, the MAC layer header is still of significant size (28 bytes), and the ratio of MD-AMR to FR AMR-WB packet sizes is still around 0.77. Figures 10 and 11 show the performance of all the methods with reduced packet sizes because of compressed headers. DFR-PD still performs significantly better than MD-PD and requires only a 10% increase in the bitrate for G.729.

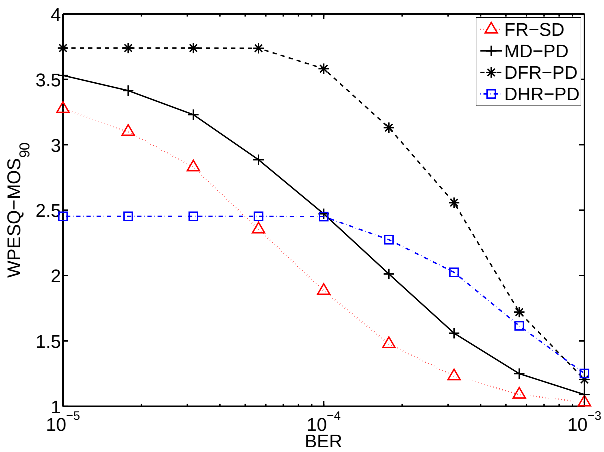


Fig. 10. MD-AMR comparisons: MOS_{90} values for changing BER with compressed headers

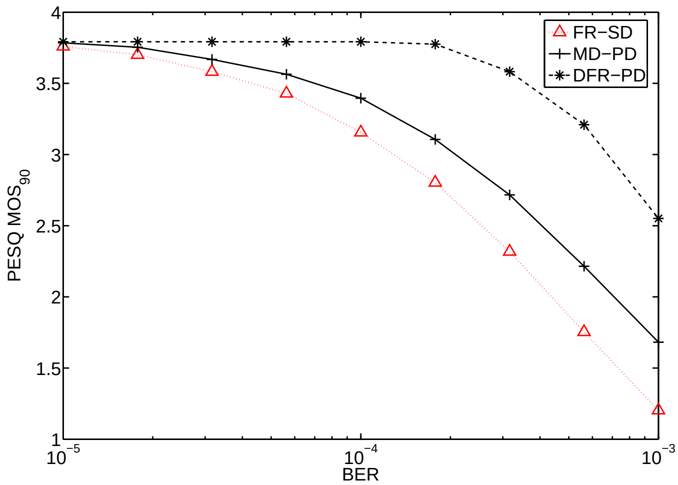


Fig. 11. MD-G.729 comparisons: MOS_{90} values for changing BER with compressed headers

I. Burst Packet Losses

Path diversity is particularly useful in counteracting the burstiness of packet losses. We assume that burst losses are independent of packet size because they are usually caused due to phenomena like fading or shadowing in the network or other factors like a link failure. For low rate speech codecs, the time required for transmission of a single packet is very small and each packet is transmitted at regular intervals of 10 or 20 milliseconds. The difference between the time required for transmitting (say) a half-rate codec packet and a full rate codec packet is less than 1 millisecond at a transmission rate of 2 Mbps. If there is a link failure for t ms, then the number of packets dropped in this time is the same for FR packets and HR packets except in very rare cases where the link failure occurs within the time interval of 1 millisecond when a HR coder would have just finished its transmission but the FR codec needs one millisecond more to finish its transmission. We assume such cases are negligible and the packet loss rate is the same for all the codecs we consider.

We model burst losses using a Gilbert model where the channel is modeled using a two-state Markov chain. The channel exists in either a good state or a bad state. No packets are dropped in a good state and all the packets are dropped when the channel is in a bad state. We assume full headers are added to each packet in these experiments.

1) MD-AMR: The same trace-files were used for MD-PD, DHR-PD and DFR-PD. Figure 12 shows the performance of each of the methods for different percentages of average packet losses and an average burst length of 4 (80 ms) packets. We see that the MD-PD (—+—) method does better than FR-SD (· · △ · ·) but worse than DHR-PD (– · □ – ·). This is because each description of the MD coder is worse than the half-rate coder. When there are burst errors, only one description is received for consecutive packets, which has a quality below the quality of the HR codec. We can see that DFR-PD (– – ⁎ – –) performs much better than any other method with only a small increase (around 16% when packet headers are included) in the rate compared to either MD-PD or DHR-PD.

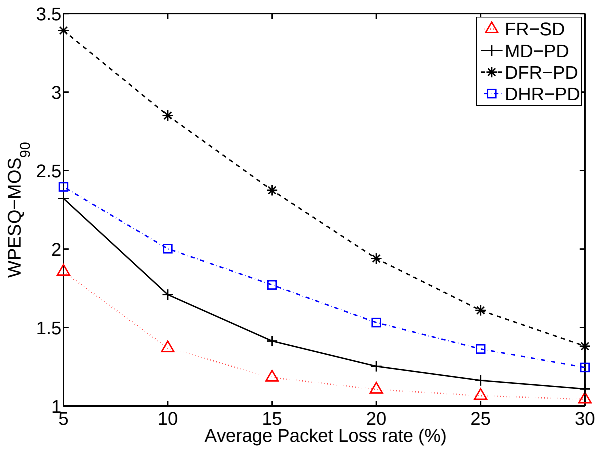


Fig. 12. MD-AMR: Comparison under bursty packet losses only for average burst size = 4

2) MD-G.729: Observe that MD-PD performs significantly better than FR-SD. This is because the packet loss concealment algorithms in CELP codecs are not very effective when successive packets are lost, since the concealment algorithm depends on the last received good frame to conceal the lost frame. Again, the DFR-PD ($--\divideontimes--$) method performs the best and the MOS delivered is significantly better than that of MD-PD (—+—). In a typical network with packet headers, the advantage in performance provided by DFR-PD under burst loss conditions requires only a slight (5.5%) increase in the number of bits transmitted compared to MD-PD.

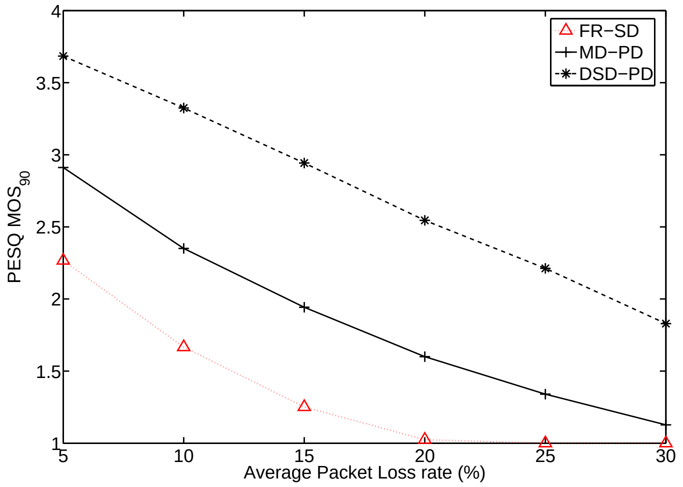


Fig. 13. MD-G.729: Comparison under bursty packet losses only for average burst size = 4

VII. $MOS_{90}/kbps$ AS A NEW PERFORMANCE MEASURE

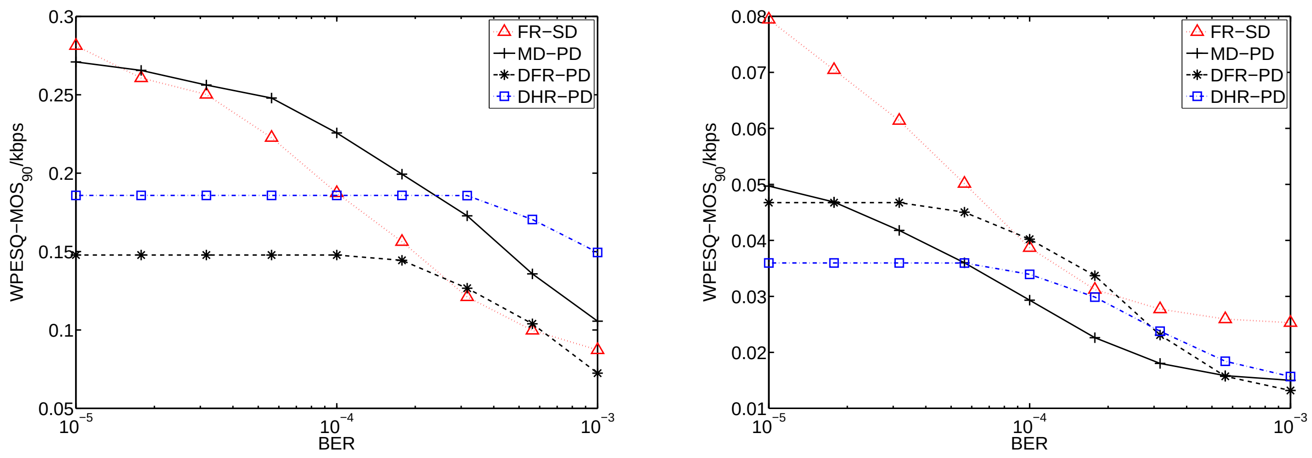
In the last section we showed that although a MD based method seems efficient in the classical scenario where no overheads are considered, the picture completely changes when we consider packet headers added to the speech payloads. In this section we introduce another performance measure to elucidate the same result and clearly show how the bandwidth efficiency of the MD-PD method is mitigated and simple path diversity seems to be a better option.

TABLE VI

EFFECTIVE BIT RATES (KBPS) FOR EACH OF THE AMR-WB BASED METHODS

Method	rate (w/o headers)	rate (with headers)
AMR-WB		
FR-SD	12.65	40
MD-PD	13.8	68.8
DFR-PD	25.30	80
DHR-PD	13.2	68.2
G.729		
Method	rate (w/o headers)	rate (with headers)
FR-SD	8	62.4
MD-PD	9.2	118
DFR-PD	16	124.8

We consider a performance measure that folds in the quality of the delivered speech and also the effective bit-rate needed to achieve that quality. For this, we divide the MOS_{90} values at each BER by the effective bit-rate of each method. Effective bit rates of each communication method in AMR-WB and G.729 based experiments are listed in Table VI with and without the packet headers included. Note that the effective rates of G.729 based methods are larger than the rates for AMR-WB based methods after adding headers because for G.729 we assume that the packets are generated every 10 ms while for AMR-WB the packets are generated every 20 ms. The new measure ($MOS_{90}/kbps$) penalizes a method that needs a larger rate.



(a) Without Headers

(b) with Headers

Fig. 14. Comparison using $MOS_{90}/kbps$ values under different BER conditions

In Fig. 14 we plot the $MOS_{90}/kbps$ values for each of the methods when (a) only the speech information is sent and (b) when packet headers are added to the speech payloads. Figure 14(a) clearly shows the advantage of the MD-PD method in the classical scenario. If we refer to Fig. 6 in the previous section for the same case, we see that the DFR-PD had the best quality of delivered speech but that figure does not capture the inefficiencies of a full rate path diversity method. With our new measure, we see that MD-PD has the best performance because of the smaller redundancy, and at higher BERs, DHR-PD has the better $MOS_{90}/kbps$ values. DFR-PD is now the worst choice because the bit-rate needed for DFR-PD is double the bit-rate required for other methods. In Fig. 14(b), as the headers are added, the inefficiencies in the PD methods increase and FR-SD has the best $MOS_{90}/kbps$ values at most of the BERs. Among the PD methods, DFR-PD has the best performance for most BERs because all the methods are almost equally redundant and from Fig. 8 we know that DFR-PD has the best MOS_{90} values.

For voice communications, we also need to ensure that the speech quality is at least acceptable to the listener. In Fig. 14(a), FR-SD has the best $MOS_{90}/kbps$ values for increasing BER, but in Fig. 8 we see that FR-SD has the worst MOS_{90} values and the quality falls steeply with an increasing BER to very low values. We impose another condition that the MOS_{90} values should be above an acceptable threshold MOS value. We choose the quality of the 6.6 kbps mode of AMR-WB as the minimum quality any method should provide. From Fig. 8, for each method we note the BER above which the quality of the method is not acceptable, and in Fig. 15 we plot the $MOS_{90}/kbps$ but only for the BER range where the speech delivered for each method is acceptable. In Fig. 14(b), FR-SD seems to be the best choice for most of the BERs, but for the same case in Fig. 15(b), we see that FR-SD is not an option above a BER of $10^{-4.5}(\approx 3.16 \times 10^{-5})$. In Fig. 15(a), we notice that MD-PD is the best option for most of the BERs, except at higher BERs where all the methods except DFR-PD fail to deliver at least acceptable MOS_{90} values. However, in Fig. 15(b), when the headers are added to the speech payloads, MD-PD does not figure in the chosen methods at any BER when $MOS_{90}/kbps$ is the criterion.

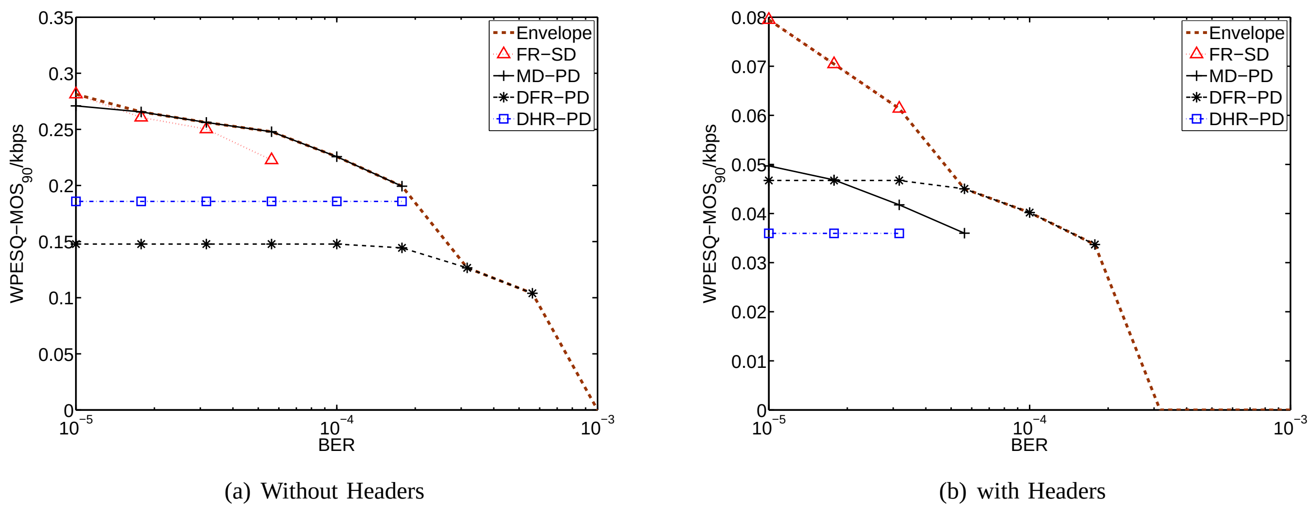


Fig. 15. Comparison using $MOS_{90}/kbps$ values under different BER conditions. The MOS_{90} values here are constrained to be above a minimum value of 2.45 (MOS_{90} for the HR coder without any packet losses)

VIII. Conclusions

We consider the problem of using path diversity to support conversational voice communication over a wireless mesh network . For path diversity, using a multiple description coder is a bandwidth efficient method. We show for memoryless Gaussian sources that in the classical MD scenario, where the coders are compared without considering any non source-coding overheads, the MD coder performs well in terms of bandwidth efficiency and the average distortion at the receiver. When headers are taken into account, the MD method is no longer the best. In fact, all the path diversity methods are now more inefficient compared to just sending a single description without path diversity. The same phenomenon is also observed through the experiments involving the MD speech coders. We designed two MD coders for speech, one based on AMR-WB and the other based on G.729. The MD coders are tested against single description coders that are used with and without path diversity. For smaller BERs, not using path diversity is the best option, while at larger BERs, a simple path diversity method performs better than the multiple descriptions method. For speech experiments, we introduce a new performance measure, $MOS_{90}/kbps$, that incorporates both the end quality and the bit rate required by a communication method. MOS_{90} is a better indicator of quality delivered to the user than average MOS and this quality measure is divided by the total bit-rate that includes the source-coding bits and the packet headers required by a communication method. The new measure is used to account for the various tradeoffs more explicitly. The MD coders do not produce the best performance when packet overheads are considered, but in the future, if packet headers are compressed to a size smaller than the speech packets or better MD speech coders are designed, then using a multiple description codec may be beneficial in terms of bandwidth and quality.

REFERENCES

[1] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, Sept. 2005.

[2] Y. Wang, S. Panwar, S. Lin, and S. Mao, "Wireless video transport using path diversity: multiple description vs layered coding," *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, 2002.

[3] N. Gogate, D. M. Chung, S. S. Panwar, and Y. Wang, "Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, no. 9, pp. 777–792, 2002.

[4] A. Miu, J. G. Apostolopoulos, W. Tan, and M. Trott, "Low-latency wireless video over 802.11 networks using path diversity," *IEEE International Conference on Multimedia and Expo*, July 2003.

[5] J. Balam and J. D. Gibson, "Path diversity and multiple descriptions with rate dependent packet losses," *Information Theory and Applications Workshop, University of California, San Diego, La Jolla, CA*, February 6-10, 2006.

[6] ——, "Multiple descriptions and path diversity using the AMR-WB speech codec for voice communications over MANETs," *Proceedings of IWCMC, Vancouver, Canada*, July 3-6, 2006.

[7] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, Nov. 2002.

[8] *Adaptive Multi-rate Wideband Speech Transcoding*, 3GPP TS 26.190 Std.

[9] J. Balam and J. D. Gibson, "Multiple description coding and path diversity for voice communication over manets," *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, October 30 - November 2, 2005.

[10] *Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)*, ITU-T G.729 Recommendation, Nov. 2000.

[11] D. Niculescu, S. Ganguly, K. Kim, and R. Izmailov, "Performance of VoIP in an 802.11-based wireless mesh network," *IEEE INFOCOM*, 2006.

[12] C.-H. Lin, H. Dong, U. Madhow, and A. Gersho, "Supporting real-time speech on wireless ad hoc networks: inter-packet redundancy, path diversity, and multiple description coding," *WMASH '04: Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, pp. 11–20, 2004.

[13] H. Dong, I. D. Chakeres, C.-H. Lin, A. Gersho, E. Belding-Royer, U. Madhow, and J. D. Gibson, "Speech coding for mobile ad hoc networks," *Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA*, November 2003.

[14] H. Dong, I. D. Chakeres, H. Dong, E. Belding-Royer, A. Gersho, and J. D. Gibson, "Selective bit-error checking at the mac layer for voice over mobile ad hoc networks with IEEE 802.11," *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2004.

[15] A. Servetti and J. C. De Martin, “Link-level unequal error detection for speech transmission over 802.11 networks,” *Proc. Special Workshop in Maui (SWIM)*, Jan. 2004.

[16] ——, “Error tolerant mac extension for speech communications over 802.11 wlans,” *IEEE 61st Vehicular Technology Conference, 2005. VTC 2005-Spring.*, pp. 2330 – 2334, 30 May-1 June 2005.

[17] Spectralink. (2006, Mar.) Spectralink voice priority. White Paper. [Online]. Available: http://www.spectralink.com/files/literature/SVP_white_paper.pdf

[18] M. Petracca, A. Servetti, and J. C. De Martin, “Voice transmission over 802.11 wireless networks using analysis-by-synthesis packet classification,” *First International Symposium on Control, Communications and Signal Processing*, 2004.

[19] C. Hoene, I. Carreras, and A. Wolisz, “Voice over IP: Improving the quality over wireless LAN by adopting a booster mechanism - an experimental approach,” *SPIE 2001 - Voice Over IP (VoIP) Technology*, pp. 157–168, 2001.

[20] A. Servetti and J. C. De Martin, “Adaptive interactive speech transmission over 802.11 wireless LANs,” *Proc. IEEE Int. Workshop on DSP in mobile and Vehicular Systems*, Apr. 2003.

[21] *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, ITU-T P.862 Recommendation.

[22] A. A. El Gamal and T. M. Cover, “Achievable rates for multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

[23] L. Ozarow, “On a source-coding problem with two channels and three receivers,” *Bell Syst. Tech. J.*, vol. 59-10, pp. 1909–1921, 1980.

[24] H. Feng and M. Effros, “On the achievable region for multiple description source codes on gaussian sources,” *IEEE International Symposium on Information Theory*, June 2003.

[25] M. Effros, R. Koetter, A. J. Goldsmith, and M. Medard, “On source and channel codes for multiple inputs and outputs: Does multiple description beat space time?” *Information Theory Workshop 2004*, Oct. 2004.

[26] *Pulse Code Modulation (PCM) for voice frequencies*, ITU-T G.711 Recommendation, 1993.

[27] V. K. Goyal, “Multiple description coding: Compression meets the network,” *IEEE Signal Processing Magazine*, Sept. 2001.

[28] N. S. Jayant and S. W. Christensen, “Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure,” *IEEE Trans. Commun.*, vol. COM-29, pp. 101–109, Feb. 1981.

[29] A. Ingle and V.A. Vaishampayan, “DPCM system design for diversity systems with applications to packetized speech,” *IEEE Trans. on Speech andAudio Processing*, vol. 3, pp. 48–58, Jan. 1995.

[30] W. Jiang and A. Ortega, “Multiple description speech coding for robust communication over lossy packet networks,” *IEEE Int. Conf. Multimedia and Expo*, vol. 1, pp. 444–447, 2000.

[31] C.-C. Lee, “Diversity control among multiple coders: a simple approach to multiple descriptions,” *IEEE Workshop on Speech Coding, 2000*, pp. 69 – 71, 17-20 Sept. 2000.

[32] S. D. Voran, “A multiple-description PCM speech coder using structured dual vector quantizers,” *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005 (ICASSP ’05)*, pp. 129 – 132, March 18-23, 2005.

[33] J. C. Batllo and V.A. Vaishampayan, “Multiple description transform codes with an application to packetized speech,” *IEEE International Symposium on Information Theory, 1994*, 27 June-1 July 1994.

[34] G. Kubin and W. B. Kleijn, “Multiple-description coding (MDC) of speech with an invertible auditory model,” *IEEE Workshop on Speech Coding, 1999*, pp. 81 – 83, 20-23 June 1999.

[35] A. K. Anandkumar, A. V. McCree, and V. Viswanathan, “Efficient CELP-based diversity schemes for VoIP,” *ICASSP*, vol. 6, pp. 3682–3685, 2000.

[36] B. W. Wah and L. Dong, "LSP-based multiple-description coding for real-time low bit-rate voice transmissions," *ICME'02*, vol. 2, pp. 597– 600, 2002.

[37] X. Zhong and B.H. Juang, "Multiple description speech coding with diversities," *ICASSP*, 2002.

[38] H. Dong, A. Gersho, V. Cuperman, and J. D. Gibson, "A multiple description speech coder based on AMR-WB for mobile ad hoc networks," *ICASSP'04*, May 2004.

[39] *Proposed modification to draft P.862 to allow PESQ to be used for quality assessment of wideband speech (BT, United Kingdom, and KPN, The Netherlands)*, ITU-T SG12 COM-D007-E Delayed Contribution, Feb. 2001.

[40] A. Rix, "Comparison between subjective listening quality and P.862 PESQ score," *Proc. of Online Workshop Measurement of Speech and Audio Quality in Networks*, pp. 17–25, May 2003.

[41] V. Barriac, J-Y. Le Saout, and C. Lockwood. (2006, Mar.) Discussion on unified objective methodologies for the comparison of voice quality of narrowband and wideband scenarios. [Online]. Available: http://portal.etsi.org/stq/presentations2004/090-Mainz_FT_objective_methodologies.doc

[42] N. Shetty, S. Choudhury, and J. D. Gibson, "Voice capacity under quality constraints for IEEE 802.11a based WLANs," *Proceedings of IWCMC,Vancouver, Canada*, July 3-6, 2006.

[43] *Software tools for speech and audio coding standardization*, ITU-T G.191 Recommendation, Mar. 1996.