A MULTIPLE DESCRIPTION SPEECH CODER BASED ON AMR-WB FOR MOBILE AD HOC NETWORKS

H. Dong, A. Gersho, J. D. Gibson, V. Cuperman

Dept. of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106
{huidong, gersho, gibson, cuperman}@ece.ucsb.edu

ABSTRACT

To address the challenging task of achieving effective voice communication over mobile ad hoc networks (MANETs), we introduce a new multiple description (MD) speech coder based on the AMR-WB standard. The MD coder splits the bitstream of the AMR-WB coder into two redundant sub-streams by directly selecting overlapping subsets of encoded data generated for each frame. The sub-streams are transmitted on different network paths. When both sub-streams arrive at the decoder, an output identical to that of AMR-WB is recovered. If only one substream arrives at the decoder, degraded but still acceptable speech quality is obtained. The performance of the multiple description coding system is tested for MANETS with a network simulator and an informal listening test. The results demonstrate that this approach makes effective use of the channel capacity and provides reliable end-to-end connections for voice communication in MANETs.

1. INTRODUCTION

The rapid growth and popularity of wireless LANs (WLANs), particularly due to the success of IEEE 802.11b, has stimulated the development of new applications and services. There has been a growing demand for support of voice over WLANs due to the spread of the WLAN environment. A mobile ad hoc network (MANET) is a wireless network of mobile nodes wherein nodes can serve as routers allowing multi-hop routing without relying on any pre-existing infrastructure. Many potential applications of MANETS will require effective real-time voice communication.

The main challenges in the design and operation of MANETs arise from the peer-to-peer network structure, the dynamic network topology, limited wireless bandwidth, and the high bit-error rate of wireless links. To enable speech transmission over MANETs, a flexible speech coding system is desired to adaptively interact with the time-varying environment.

The adaptive multirate wideband (AMR-WB) speech coder is a 3GPP standard for GSM and the third generation cellular WCDMA system. It not only provides superior voice quality over the existing narrowband standards, but also is very robust against transmission errors and takes full advantage of the limited channel capacity.

One powerful approach to achieve reliable transmission of packetized speech over adverse network conditions is the use of multiple description (MD) coding. In MD coders, the source is encoded into two or more descriptions that are then separately transmitted. Each description is individually decodable for a reduced quality reconstruction of the source but if two or more descriptions are available they can be jointly decoded for a better reconstruction. MD coding offers a way to cope with packet loss and transmission erasures in communication networks.

In this paper, we introduce and describe a new MD coder, MD-AMR, based on AMR-WB for voice over MANETs. The MD coder splits the bitstream of the AMR-WB coder into two redundant sub-streams by directly selecting overlapping subsets of encoded data generated for each frame. The sub-streams can then be transmitted in separate packets and on different network paths. When both sub-streams arrive at the decoder, an output identical to that of AMR-WB is recovered. If only one substream arrives at the decoder, degraded but still acceptable speech quality is obtained. The network performance of the MD-AMR coder is tested for MANETS in terms of packet loss rate and transmission delay with a network simulator. The subjective speech quality of the MD-AMR coder is compared with that of AMR-WB coder under different channel conditions.

2. RELATED WORK

2.1. Voice over Wireless LANs

The most efficient way to transmit voice over WLANs is to employ a reservation scheme or high priority scheme that guarantees delay and bandwidth [1, 2]. The IEEE 802.11e standard is under development to support delay-sensitive applications for Quality of Service (QoS) with multiple managed levels of QoS for data, voice, and video applications. An adaptive speech transmission over 802.11 WLANs was proposed in [3] where the bit rate of the encoder is adapted to suit changing channel conditions to address the stringent quality of service requirement.

2.2. Multiple Description Coding

An MD coder produces multiple descriptions, i.e., two or more coded bit streams, from a given source signal as shown in Fig. 1 for two descriptions, so that each bit stream independently represents a “coarse” description of the source (e.g., output 1 or output 2 in Fig. 1), while multiple descriptions jointly convey a “refined” source representation (output 0).

One approach to designing MD coders is to constrain the performance of the joint description and then attempt to minimize the distortion of the individual descriptions for a given constraint on

This research is supported in part by the NSF under grants EIA-9986057, EIA-0080134, and CCR-0243332, the University of California MICRO Program, Dolby Laboratories, Inc., Lucent Technologies, Inc., Microsoft Corporation, and Qualcomm, Inc.

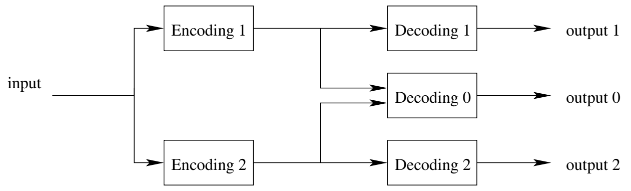

Fig. 1. A multiple description coding scheme.

their bit rates. A key issue in designing MD coders in this way is to effectively minimize redundancy between the two descriptions while trying to minimize distortion in the individual descriptions. While some practical MD coders have been developed for image and video, relatively little attention has been given to MD speech coding. Some notable efforts for MD speech coding are reported in [4, 5, 6, 7, 8].

2.3. Speech Coding

Generally, for wired VoIP applications and telephone bandwidth input speech, the ITU G.711 standard at 64 kbps is used when the relative traffic load is expected to be low. For higher traffic wired VoIP applications, G.729 at 8 kbps is widely used with very good speech quality. There is also an extensive set of speech coding standards for digital cellular applications such as the Adaptive Multi-Rate Narrowband (AMR-NB) speech coder standardized by ETSI in 1998. More recently, the AMR-WB speech coder [9] was selected by the Third Generation Partnership Project (3GPP) for GSM and WCDMA and the same algorithm was adopted by the ITU as Recommendation G.722.2.

AMR-WB, based on ACELP, consists of 9 modes at bit rates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85, and 6.6 kbps and covers the speech bandwidth of 50-7000 Hz with a sampling rate of 16 kHz. AMR-WB provides superior voice quality over the existing narrowband standards and the 12.65 kbps or higher rate modes offer high quality wideband speech. The codec also has optional features that include compression, voice activity detection, forward error correction, and error concealment techniques. Most of these features have specifically been designed for wireless applications. Our study here is based on AMR-WB due to its state-of-the-art performance and its promise for IP telephony and 3G cellular phones.

3. VOICE COMMUNICATION IN MANETS

3.1. Network Model

A typical diagram of a MANET is shown in Fig. 2, where we assume a protocol stack as shown in the oval that is based on 802.11 physical and MAC layers and UDP/IP. We consider a voice communication session between nodes A and B, and assume that A and B cannot hear each other, so that multi-hop routing is needed and multiple paths are generally available. Two of many possible paths for the transmission are shown in the figure. Since nodes are mobile, the network topology is changing and a dynamic routing protocol is needed to establish a correct and efficient route to deliver packets on time.

In such a network, the transmission delay budget is tight due to multi-hop routing and high bit-error rates. Also the packet loss rate can be very high under adverse conditions, such as channel noise or heavy traffic with multiple communications and a high collision rate. Generally, retransmission of packets is not desirable for voice over MANETS since it increases the packet loss rate and adds extra delay. Moreover, the communication might be lost if the packet loss rate is too high or the communication link is broken. It is clearly a challenging task to carry voice over MANETs.

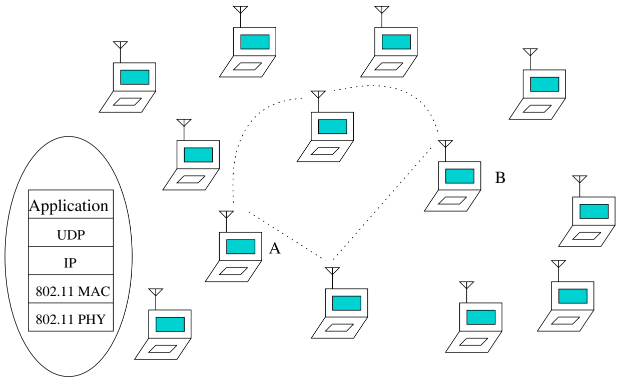

Fig. 2. Example of multi-hop and multi-path communication in a MANET

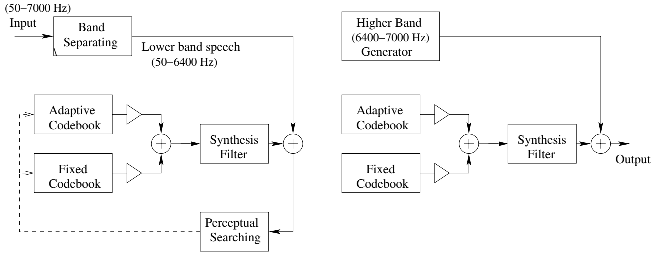

Fig. 3. A simplified block diagram of the AMR-WB coder (excluding 23.85 kbps mode)

This network environment suggests an adaptive coding scheme based on MD speech coding. Under low error rate conditions, conventional AMR-WB could be selected for high quality wideband speech and during severe channel conditions, a suitable control mechanism could switch to an MD-AMR coder. In this paper, however, we limit our focus to the MD coder and its performance when an MD mode is operative in the network.

4. MD-AMR CODER

This multiple description coder is designed to ensure a full quality at 12.65 kbps if two descriptions are received and to have a degraded quality at 6.8 kbps if one description is received. A minimal coding redundancy between two descriptions is used.

The 12.65 kbps mode of AMR-WB processes speech in two frequency bands, 50-6400 Hz and 6400-7000 Hz. The lower frequency band is encoded and decoded by an ACELP algorithm and the higher frequency band is reconstructed in the decoder using the parameters of the lower band and a random excitation. A simple block diagram of the 12.65 kbps mode is shown in Fig. 3. No bits are allocated to the higher band; the bit allocation in lower band is shown in Columns 2 to 6 in Table 1. Linear prediction (LP) analysis is performed once per 20 ms frame. The set of LP parameters is converted to immittance spectrum pairs (ISPs) and vector quantized using split-multistage vector quantization with 46 bits: 16 bits are used at the first stage and 30 bits are used for five subvectors at stage two. The parameters of the adaptive codebook (ACB) and fixed codebook (FCB) are generated for each 5 ms subframe. The 64 positions in a subframe are divided into 4 tracks, where each track contains two pulses (+1 or -1). Each two pulse positions in one track are encoded with 8 bits, and the sign of the first pulse in the track is encoded with 1 bit. This gives a total of 36 bits for the algebraic code. The pitch lag is encoded with 9 bits in odd subframes and relatively encoded with 6 bits in even subframes. One bit per frame is used to determine the low pass filter applied to the past excitation. The pitch and algebraic codebook gains are jointly quantized using 7 bits per subframe.

4.1. Bit Splitting for Two Descriptions

Our MD-AMR at 13.5 kbps forms two descriptions, denoted D1 and D2, for each speech frame by directly selecting overlapping subsets of the encoded data generated by the standard. The most significant ISP bits, which are coded by the first stage VQ, are included in both two descriptions. The ISP bits for the second stage VQ are split into two descriptions: 16 bits of 3 vectors for D1 and 14 bits of the other 2 vectors for D2. The bits for the ACB lags and gains of the first two subframes are placed in D1, and those for the last two subframes are included in D2. The FCB bits for each subframe are split so that tracks 1 and 2 go to D1 and tracks 3 and 4 are assigned to D2. The bit allocations for D1 and D2 of this scheme are shown in Column 7 and 8 in Table 1. The resulting bit rates for D1 and D2 are 6.8 kbps and 6.7 kbps respectively, and the rate for combining two descriptions is 13.5 kbps. The one-bit flag for the voice activity detector (VAD) is included in each description.

Table 1. Bit allocation of 12.65 kbps mode and its two descriptions

	12.65 kbps	sf 0	sf 1	sf 2	sf 3	D1	D2
VAD	1					1	1
LPC	46					32	30
ACB	34	10	7	10	7	17	17
Gains	28	7	7	7	7	14	14
FCB	144	36	36	36	36	72	72
Total	253					136	134

4.2. Multiple Description Decoding

If both descriptions are received, decoder 0 of Fig. 1 performs the standard decoding process of the 12.65 kbps mode after combining the two descriptions into one nonredundant bitstream, and the full quality recovered speech frame at 12.65 kbps is obtained. If one description (D1 or D2) is received, the corresponding decoder of Fig. 1estimates the missing information and then decodes the parameters for a degraded but acceptable quality. The following mechanism for estimation and decoding is adopted.

The missing information of the ISPs at the second stage is reconstructed by choosing a random vector from the stage codebook. The missing pitch lag information, e.g., the lags $t_{n,3}$ and $t_{n,4}$ for the third and fourth subframes of frame n are estimated by a linear interpolation of the pitch lags in the adjacent subframes, according to:

$$\begin{aligned} t_{n,3} &= \alpha_1 t_{n,1} + \alpha_2 t_{n,2} + \alpha_3 t_{n+1,1} + \alpha_4 t_{n+1,2} \\ t_{n,4} &= \alpha_1 t_{n+1,2} + \alpha_2 t_{n+1,1} + \alpha_3 t_{n,2} + \alpha_4 t_{n,1} \end{aligned}$$

This is illustrated in Fig. 4 where missing subframes are marked by dashed lines. We have chosen interpolation coefficients $\alpha_1 = 2/8, \alpha_2 = 3/8, \alpha_3 = 2/8, \alpha_4 = 1/8$. The gains are estimated in a similar way, but in the logarithm domain. When only one description is received, the decoder performs the adaptive anti-sparseness post-processing that is used for the 6.6 kbps mode of AMR-WB to reduce perceptual artifacts due to the sparseness of the codebook vectors.

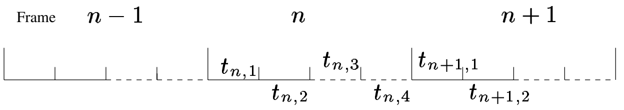

Fig. 4. Missing information of pitch lags and gains in a single description

5. SIMULATION AND RESULTS

We examine speech transmission between two nodes, A and B (Fig. 2), with the NS-2 network simulator [10]. The distances between adjacent nodes are chosen so that the nodes are within "hearing" distance from each other for one link. The data rate of IEEE 802.11 was configured to be 2 Mbps. The destination sequence distance vector (DSDV) routing protocol is used [11].

Network performance in terms of packet loss and average transmission delay is evaluated for AMR-WB coder at 23.85 kbps, 12.65 kbps, 6.6 kbps, and for MD-AMR coder at 13.5 kbps. The transmission delay includes the transport delay in the network and the propagation delay in the air, but does not include the packetization delay or the buffer delay at source and destination. In our simulation, the nodes have no mobility. Also, for simplicity, no packet is dropped due to late arrival and each MD packet is assumed to be sent independently by using suitable multiple path routing and transport protocols that are not studied in this paper. These are of course not realistic conditions; however, they simplify and isolate the analysis of the network performance for assessing packet loss and transmission delay.

In this simulation, the Elliott-Gilbert two state Markov model [12, 13] is used to model the wireless channel. The bit errors generated by this model are introduced to MAC frames. In this model, each state represents a binary symmetric channel. Bit errors occur with a low probability in the "good" state, and with a high probability in the "bad" state. Transition probabilities are specified for switching between good and bad states.

The network simulation results are shown in Fig. 5. If the channel bit error rate is less than 10^{-5}, the packet loss rate is less than 6% and the transmission delay is less than 4 ms for either AMR-WB coding or MD-AMR coding. Hence AMR-WB coder can be used for a good quality and an adaptive source coding is advantageous. If the channel bit rate is 10^{-3} or higher, the packet loss rate ranges from 25% to 40% for AMR-WB coding, and the connection may be lost. By adopting the MD-AMR coder, the packet loss rate is reduced to 8% without introducing an extra delay.

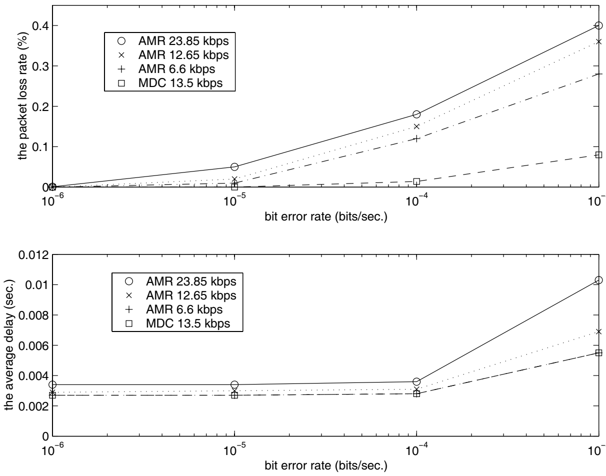

Fig. 5. Packet loss rate and average transmission delay

An informal listening test with 10 untrained listeners was conducted to evaluate the subjective quality of 25 seconds of reconstructed speech after MD-AMR coding and transmission as described in the above simulated conditions. We evaluate the recovered speech from the AMR-WB coder at 12.65 kbps and from the MD-AMR coder at 13.5 kbps under typical channel conditions where the lost packets are concealed by using the standard error concealment algorithm in AMR-WB (Table 2) . When the channel bit error rate is 10^{-5} and the packet loss rate is very low, the coded speech quality from the AMR-WB coder at 12.65 is close to the coded speech quality from the MD-AMR coder at 13.5 kbps. Therefore, the use of MD is not needed in this case and the AMR-WB coder is preferred. When the channel error rate is increased to 10^{-4}, the packet loss rate is 1% with MD-AMR and 15% with the AMR-WB coder and 6 of 10 people prefer the speech from MD-AMR coder to the speech from the AMR-WB coder. When channel bit error rate is increased to 10^{-3}, all listeners prefer the speech from the MD-AMR coder.

Table 2. Subjective Preferences of Coded Speech

P_{ber}	AMR-WB at 12.65 kbps	MD-AMR at 13.5 kbps
10^{-5}	5/10	5/10
10^{-4}	4/10	6/10
10^{-3}	0/10	10/10

6. CONCLUSIONS

This paper proposed the MD-AMR speech coder for voice communication in MANETs. The simulation results show that MD-AMR is indeed preferable to sending the full-rate AMR-WB at 12.65 kbps under high error rate conditions. This suggests that an adaptive coding scheme that switches between AMR-WB and MD-AMR according to channel conditions can be beneficial for MANETS. Such a system offers the promise of making effective use of the channel capacity and providing reliable end-to-end connections for voice communication in MANETs. Further research on multi-path routing protocols, adaptive rate selection algorithms, and further study on effective MD coding schemes are needed to lead to a practical system. Multiple description speech coding can be advantageously combined with many other techniques for voice over MANETs, such as MAC layer modifications, header compression, and silence compression.

7. REFERENCES

[1] M. I. Kazantzidis, L. Wang, and M. Gerla, “On fairness and efficiency of adaptive audio application layers for multihop wireless networks,” in *1999 IEEE International Workshop on Mobile Multimedia Communications*, Nov. 1999, pp. 357–362.

[2] I. Joe and S. G. Batsell, “Reservation CSMA/CA for multimedia traffic over mobile ad hoc networks,” in *IEEE International Conference on Communication*, 2000, vol. 3, pp. 1714–1718.

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

[4] N. S. Jayant, “Subsampling of a DPCM speech channel to provide two ‘self-constrained’ half-rate channels,” in *Bell Syst. Tech. J.*, 1981, vol. 60, pp. 501–509.

[5] Dong Lin and B. W. Wah, “LSP-based multiple-description coding for real-time low bit-rate voice transmissions,” in *IEEE International Conference on Multimedia and Expo*, 2002, vol. 2, pp. 597–600.

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

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

[8] X. Zhong and B.-H. Juang, “Multiple description speech coding with diversity,” in *Proc. of ICASSP*, May 2002, vol. 1, pp. 177–180.

[9] GSM, *3GPP TS 26.171: Speech Codec speech processing functions; AMR wideband Speech codec; General Description*, Mar. 2001.

[10] UCB/LBNL/VINT, *Network Simulator- ns2*, URL: http://www.isi.edu/nsnam/ns, 1997.

[11] C. E. Perkins and P. Bhagwat, “Highly dynamic destination-sequenced distance vector (DSDV) for mobile computers,” *Proc. of the SIGCOMM Conference on Communications Architectures, Protocols and Applications*, vol. 39, pp. 234–244, 1994.

[12] E. N. Gilbert, “Capacity of a burst-noise channel,” *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, 1960.

[13] E. O. Elliot, “Estimates of errors rates for codes on burst-noise channels,” *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, 1963.