

A Transcoding-Free Multiple Description Coder for Voice over Mobile Ad-hoc Networks

Jagadeesh Balam* and Jerry D. Gibson†

* Ditech Networks, Mopuntain View, CA 94043

email: balam@ece.ucsb.edu

† Department of Electrical and Computer Engineering

University of California, Santa Barbara, California 93106-9560

email: gibson@ece.ucsb.edu

Abstract— We propose a new multiple description (MD) coder design based on the Adaptive Multi-Rate Wideband (AMR-WB) coder that can support transcoding-free communication between an ad-hoc network and another network that supports the AMR-WB codec. The encoder of the MD coder consists of the standard AMR-WB coder and a bit-stream splitting block that splits the AMR-WB bit-stream into two balanced descriptions. The decoder consists of a bit-stream substitution block that substitutes the missing bits, when only one description is received, to construct a valid AMR-WB frame that can be decoded using the standard AMR-WB decoder. We show that the performance of the new MD coder is better than a previous non-transcoding-free MD coder based on AMR-WB when transcoding is required and it is significantly better than using a single description of AMR-WB over the ad-hoc network supporting transcoding-free communication.

I. VOICE OVER AD-HOC NETWORKS

Mobile Ad-hoc Networks (MANETs) are formed by mobile wireless hosts without the need of an existing infrastructure, unlike wireless cellular systems which require a centralized control and support system at the base station. Most of the wireless systems deployed today are centralized systems, wherein the nodes connected to the network communicate through an access point or a base station. MANETs do not need such a centralized support and can be deployed any place where nodes want to talk to each other or someone else through their neighbors. Since the nodes now depend on each other for communication, mobility of nodes implies that the routes for data transfer between two nodes are not fixed.

MANETs are seen as future networks for Personal Area Networks (PAN), military environments, emergency operations and in an office or conference environments. Voice communication is essential in all of these scenarios. Real time communication over a MANET is a challenging problem because of the many transient characteristics of the network that arise due to the flexibilities that a MANET offers. The constantly changing routes between the sender and receiver nodes, lack of time synchronization between nodes in the network, broken links and delays involved in establishing

a new link or finding a new route, make interactive voice communication a difficult proposition in MANETs.

Path diversity can be used to improve end-to-end connectivity between the nodes in a MANET. 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 the efficiency of using multiple paths, a source coding diversity method like Multiple Description (MD) coding can be used. 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.

With different communication networks using different speech codecs, merging of the networks requires *cross tandeming / transcoding* of speech, where speech is decoded from the bit-stream of the first network and then re-encoded using the speech coding standard supported by the second network. This method, referred to as conventional cross tandeming in [1] is just referred to as *transcoding* in the rest of this paper.

Transcoding causes degradation in quality and increases delay and complexity of the system [1]. Methods have been proposed to reduce the complexity and delay involved in complete decoding of the signal and re-encoding by only re-encoding the parameters that are coded differently in the two coders, but quality is still degraded by this kind of transcoding of parameters. New efforts are in the direction of avoiding transcoding when the involved networks use the same coder. One example is the Transcoder free Operation effort of 3GPP2 [2]. The AMR-WB coder [3] was adopted as a standard for both wireline (ITU-T) and wireless networks (ETSI/3GPP) facilitating transcoding-free communication over a wide range of networks. To support transcoding free communication using AMR-WB over an ad-hoc network, we present a new MD coder that is designed to be interoperable with an AMR-WB coder used in another network, introducing only nominal complexity into the system and requiring only substitution of parameters instead of complete transcoding.

This work was supported by the California Micro Program, Applied Signal Technology, Dolby Labs, Inc., Marvell Semiconductor, Cisco, and Qualcomm, Inc. and by NSF Grant Nos. CCF-0429884 and CNS-0435527

II. CELP BASED MD CODERS FOR SPEECH

Most of the popular low rate codecs like G.729, Enhanced Full Rate (EFR) and Adaptive Multi-Rate (AMR) for narrow-band speech, and Adaptive Multi-Rate Wideband (AMR-WB), and Variable Multi-Rate (VMR-WB) for wideband speech are based on Code Excited Linear Prediction (CELP). Because of this popularity, quite a few methods have been proposed for creating multiple descriptions using CELP codecs. In [4], 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 [5] present a zero redundancy multiple description coding method that uses the correlations in LSPs (Line Spectrum Pairs) of adjacent frames. LSPs are interleaved and the missing LSPs are substituted with linearly interpolated values. The excitation is generated for a larger subframe and the same codeword is replicated in all the descriptions. Zhong and Juang [6] 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. They also experiment using different diversities with parametric coders like G.729.

In [7], a new MD coder (MD-AMR) based on the AMR-WB [3] codec is presented that creates two descriptions by dividing the AMR-WB bit-stream into two sub-streams. We presented two MD coders, one based on the AMR-WB codec [3] and the other based on the G.729 codec in [8]. These two coders were 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. We extend the idea of MD-AMR [3] to design a transcoding-free MD coder (MD-TrF) based on the AMR-WB codec, that requires bit-stream splitting at the encoder and a bit-stream substitution at the decoder. The reconstructed bit-stream at the decoder is compatible with the standard AMR-WB decoder.

III. TRANSCODING FREE MD CODER : MD-TrF

Our MD-TrF coder supports transcoding-free communication between a mobile ad-hoc network and a network that supports the AMR-WB coder. For example, consider a call that is made by a user within an ad-hoc network to another user that is outside the ad-hoc network, but is connected to the internet through a voice over IP network that supports the AMR-WB coder. The MD-TrF coder at the caller's end, first encodes the speech using the AMR-WB encoder and then splits the encoded bit-stream into two descriptions that are sent over independent paths in the ad-hoc network to reach the network gateway. At the network interface, there is no need to decode and re-encode the signal to create an AMR-WB encoded bit-stream that can be forwarded to the callee's network. When both the descriptions are received, they are combined to form the AMR-WB stream from which the two descriptions were created. When only one of the descriptions is received, the bits corresponding to the missing parameters

TABLE I
BIT ALLOCATION FOR THE MD CODER BASED ON AMR-WB

ISP	Stage 1: 8 8				I,(II)
	Stage 2: 6 7 (7) (5) 5				
	1st sf	2nd sf	3rd sf	4th sf	34,(34)
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)

are added by the MD-TrF decoder to form a valid AMR-WB frame that can be forwarded over the internet to the user at the receiving end. Similarly, when the AMR-WB bit-stream from the callee's end reaches the network interface, the MD-TrF encoder splits the AMR-WB bit-stream into two descriptions that are sent over independent paths in the ad-hoc network to the caller, thus avoiding transcoding at the network interface.

The encoder of our new MD coder is the same as the encoder for the MD-AMR coder presented in [8]. The encoder consists of a bit-stream splitting block after the regular AMR-WB encoder. Each encoded frame, generated every 20 ms, is split into two separate bit-streams of equal rate (6.9 kbps), as shown in Table I. The encoder design already supports transcoding-free communication for an AMR-WB stream coming into the ad-hoc network, because the incoming bit stream is just split into two streams instead of being decoded and re-encoded using an MD coder into two streams. However, the decoder design of MD-AMR [8] does not support transcoding-free communication but the new design of the decoder for MD-TrF makes it a complete transcoding-free MD coder. Our new coder, MD-TrF, consists of a bit-substitution block that substitutes the missing bits in each description and creates a valid AMR-WB bit-stream that can be decoded using a standard AMR-WB decoder. However, the selection of the bits to be substituted must be done carefully.

More specifically, when both the descriptions are received at the decoder, the two descriptions are combined to form the complete AMR-WB encoded frame from which the two descriptions were formed at the encoder. This complete frame can be decoded by a standard AMR-WB decoder and hence can be transmitted as it is to a network that supports the AMR-WB coder. However, when only one of the two descriptions is received, the received frame cannot be fed to a standard AMR-WB coder as it is, because the received frame is not a complete AMR-WB frame.

In the MD-AMR coder [8], when only one of the descriptions is received, the parameters corresponding to the received bits are decoded and the parameters corresponding to the missing bits are concealed using the values of the parameters in the previous subframe. Here, in the MD-TrF coder, the concealment is done at the bit-stream level and the missing bits in each description when compared to a AMR-WB frame

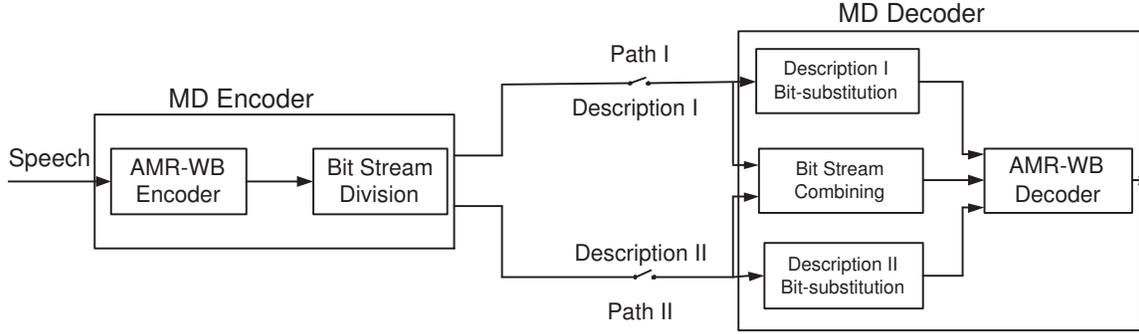


Fig. 1. Block diagram of the transcoding free MD coder based on AMR-WB

are added with bits chosen to conceal the missing information.

When only one of the descriptions is received at the decoder, say description I, all the bits corresponding to description II that are not part of description I need to be added to form a complete AMR-WB frame. In Table I, all the bits which are only present in description II are indicated by the numbers enclosed within brackets. When only description I is received, the following missing bits need to be added

- Third and fourth subvector of the second stage vector quantizer of the ISP vector
- Second VAD bit and the LTP filtering flags for the third and fourth subframes
- Pitch delay of the third and fourth subframes
- Algebraic codevector of the second and fourth subframes
- Jointly quantized adaptive codebook and fixed codebook gains for the second and fourth subframes

The bit substitution for each parameter, when only description I is received, is described next.

From Table I, note that the for ISPs, the bits corresponding to the third and fourth sub-vectors of the second stage of the vector quantizer are missing in description I. Since the second stage vector corresponds to the error between the unquantized vector and the vector chosen in the first stage, when no information is received about the second stage vector, the best strategy is to not add any refinement vector at the second stage, i.e. to add a zero vector. Our best option for the MD-TrF coder would be to substitute the missing bits using the indices that point to a zero vector in the codebook but because there are no zero vectors in any of the codebooks in the AMR-WB coder, we pick the vectors with the smallest 1-norm in each of the codebooks for our substitution. The indices of the smallest norm-1 vector in the codebook of each subvector are listed in Table II.

TABLE II
INDEX OF THE CODEVECTOR WITH THE SMALLEST 1-NORM

subvector	Index
2 (Desc. II)	51
3 (Desc. I)	50
4 (Desc. I)	31
5 (Desc. II)	21

The bits corresponding to the pitch-lag of the third and

fourth subframes are set to be the same as the pitch-lag value of the second subframe (already available in Description I). Also, the LTP filtering flag bit is copied to be the same as that of the second subframe for the third and the fourth subframes. The bits corresponding to the fixed codebook vector of the second (fourth) subframe are set to be the same as that of the first (third) subframe.

Similar copying of bits corresponding to the gains in the second and fourth sub-frame may lead to large distortions in the output speech. To suppress these distortions, bits corresponding to attenuated values of the gains are chosen. The adaptive codebook gain and the fixed codebook (FCB) gain are jointly encoded in AMR-WB using a 7 bit two-dimensional codebook. The missing gain indices in the second and third subframes for description I are substituted with the indices of codevectors closest to the attenuated gains. This requires a small memory in the bit-substitution block to store the gain quantization codebook.

For the search within the codebook, we first get the gains corresponding to the gain index in the previous received sub-frame. The new FCB gain is set as the FCB gain from the previous sub-frame multiplied by an attenuation factor (0.98) and the ACB gain is retained as the same. In the two-dimensional codebook, the magnitude of the ACB gain increases with the value of the index, hence, to keep the ACB gain close to the original we search the codevector within 5 indices above and below the received index that has the FCB gain component closest to the attenuated value. The bits corresponding to the index of the chosen codevector are added in the received description I frame to form a complete AMR-WB frame.

The advantage of the new decoder design is that it can be used for transcoding-free operation, but for a single encoding and decoding without any transcoding, the quality of the MD-TrF output speech will be worse than the MD-AMR coder because the concealment of missing information by MD-TrF is more restricted compared to MD-AMR.

IV. PERFORMANCE

We compare the performance of our MD-TrF coder against the performance of a single description AMR-WB coder at 12.65 kbps and the MD-AMR coder of [8]. We assume that

two independent paths are available for using path diversity. For packet losses, we follow the IEEE 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 [9], no retransmissions are allowed in the network and the MAC layer does not use an acknowledgment packet to indicate whether a packet was successfully delivered. For such a scenario with two paths of the same reliability, 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 significance of 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 [10] for the wideband experiments. WPESQ (Wideband PESQ) is an extension to ITU-T P.862, proposed in [10], 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 (as described in [10]) to evaluate quality of speech in our wideband experiments.

A. 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 [11] that an average MOS value might only be achieved for only 50% of the realizations, i.e. the average MOS value is only an indicator of quality guaranteed for 50% of the users. We need a performance measure that is indicative of the quality for a majority of time the channel is used. We choose a performance measure suggested in [11], MOS_{90} , which is the MOS value that is achieved for at least 90% of the realizations. The MOS_{90} is a better indicator of quality delivered to the user than average MOS.

We assume that each packet sent over the network contains one coded speech frame. To allow for the unpredictable delays in the network, we keep the packetization delay at the minimum of one speech frame of 20 ms. We assume that the

IP/UDP/RTP headers are compressed to an average header size of 2 bytes and the MAC layer header is still 28 bytes.

B. Results

We examine the performance of the MD-TrF coder under different packet loss conditions. First, 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

$$p = 1 - (1 - BER)^L \quad (1)$$

where L is the packet size in bits. For different BERs and corresponding packet loss probabilities (p), we generated 250 trace files using different seeds of the random number generator. The trace files were used to drop frames in the encoded speech files that correspond to the lost packets. For the MD methods, the bit error rates in both the paths are assumed to be the same. Note that the packet size for the MD codecs is smaller than that of the SD AMR-WB coder because of the smaller payload size.

First, we compare the performance of the coders in a scenario where no transcoding is required, such as for example, a call that originates and ends within an ad-hoc network. In such a scenario, the quality is dependent only on a single encoding and decoding on the input speech. In Fig. 2, we plot the MOS_{90} values under different bit error rate conditions for each of the three communication methods. We see that there is a performance loss for the new transcoding-free MD-coder (MD-TrF) compared to the MD-AMR coder. This is because of the suboptimal approach of the MD-TrF coder in concealing the missing bits compared to the MD-AMR coder, but the performance of the MD-TrF coder is still significantly better than that of the SD coder (single description of AMR-WB@ 12.65 kbps) under random packet loss conditions.

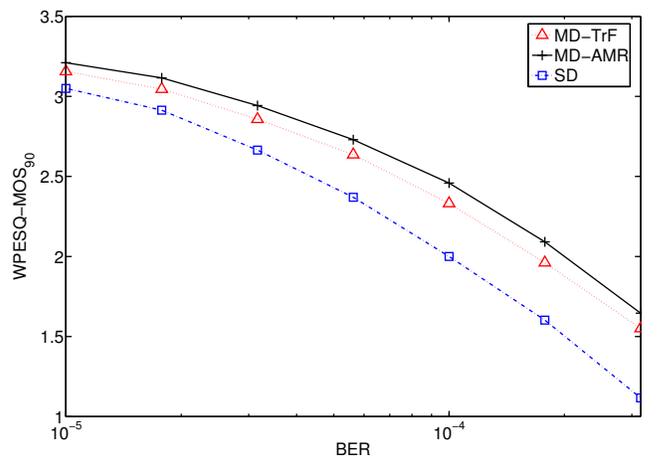


Fig. 2. Voice quality comparisons under random packet losses with no transcoding

Now, we consider the case where calls are made between different networks, such as for example, a user in an ad-hoc network calling another user in a cellular network. If the

cellular network is assumed to be supporting the AMR-WB coder, then the MD-TrF coder and SD coder do not need any transcoding, whereas the MD-AMR coder needs transcoding at the network interface to decode the MD-AMR frame and encode using the AMR-WB coder. For ease of comparison, we assume that there are no losses in the cellular network. Hence, the quality remains the same as in the scenario of Fig. 2 for MD-TrF and SD but the quality of the MD-AMR call drops due to transcoding. This loss in quality for MD-AMR is shown in Fig. 3. The advantage of using the MD-TrF to support transcoding-free communication is clear from Fig. 3 since the quality of delivered speech using MD-TrF is better than the MD-AMR coder at low loss rates and only marginally worse than the MD-AMR coder at large packet losses (above 10% packet loss rate corresponding to a BER of $10^{-3.75}$). Also, note that the PESQ quality measure does not account for loss in quality due to delay in the network. The quality of the MD-AMR call could further be affected by the additional delay introduced due to transcoding, resulting in a significantly better quality of an MD-TrF coder based call.

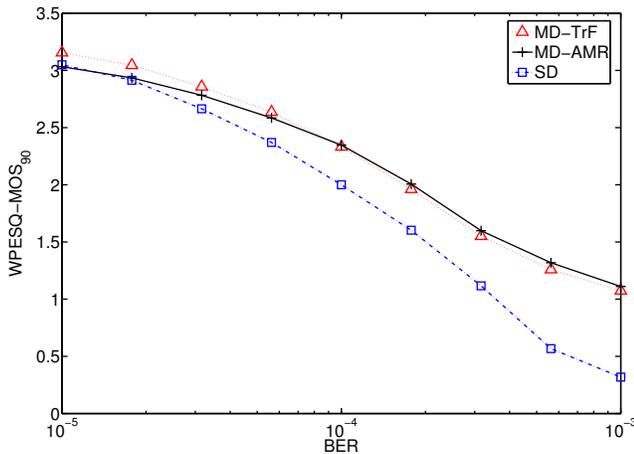


Fig. 3. Voice quality comparisons under random packet losses with no transcoding for MD-AMR

In Fig. 4, we compare the performance of the coders under burst loss conditions. To model burst losses, we use the Gilbert model where the channel is modeled using a two-state Markov chain and we assume that the loss rate is independent of the packet size. No packets are dropped in the good state and all the packets are dropped when the channel is in the bad state. The MD-AMR encoded speech is transcoded with an AMR-WB coder. Observe from Fig. 4 that the WPESQ-MOS scores for MD-TrF are consistently better than that of MD-AMR, particularly in the 5% to 15% packet loss rates MD-TrF does significantly better than MD-AMR, and our new MD method also does significantly better than SD under burst loss conditions.

V. CONCLUSIONS

A new multiple description coder based on the AMR-WB coder is presented to support transcoding-free communication between two networks. The encoder includes the standard

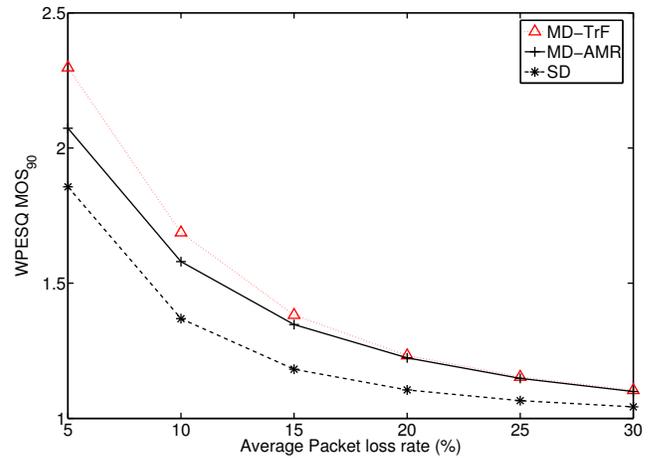


Fig. 4. Burst losses with average burst length = 4

AMR-WB coder and a bit-splitting block that divides the bit-stream into two sub-streams of equal rate that can reproduce speech with similar quality. The decoder consists of a bit-substitution block that substitutes missing bits when only one of the descriptions is received to form a complete AMR-WB encoded frame that can be decoded using the standard AMR-WB decoder. The performance of the new coder is compared against another AMR-WB based MD coder (MD-AMR) and is shown to be better than the MD-AMR under conditions that require transcoding and marginally worse when no transcoding is required. The new coder can be used for reliable voice communication over an ad-hoc network using path diversity with minimal delays introduced at the network interfaces.

REFERENCES

- [1] H.-G. Kang, H. K. Kim, and R. V. Cox, "Improving the transcoding capability of speech coders," *Multimedia, IEEE Transactions on*, vol. 5, no. 1, pp. 24–33, 2003.
- [2] 3GPP2, "Transcoder free operation: Stage 1 requirements." [Online]. Available: www.3gpp2.org/Public_html/specs/S.R0096-0.v1.0_040319.pdf
- [3] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wide-band speech codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, Nov. 2002.
- [4] A. K. Anandkumar, A. V. McCree, and V. Viswanathan, "Efficient CELP-based diversity schemes for VoIP," *ICASSP*, vol. 6, pp. 3682–3685, 2000.
- [5] 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.
- [6] X. Zhong and B.H. Juang, "Multiple description speech coding with diversities," *ICASSP*, 2002.
- [7] 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.
- [8] J. Balam and J. D. Gibson, "Multiple Descriptions and Path Diversity for Voice Communications Over Wireless Mesh Networks," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 9, no. 5, p. 1073, 2007.
- [9] 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.
- [10] *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.
- [11] 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.