Speech Coding for Mobile Ad Hoc Networks *

H. Dong, I. D. Chakares, C.-H. Lin, A. Gersho, E. Belding-Royer†, U. Madhow, J. D. Gibson
Dept. of Electrical and Computer Engineering, Dept. of Computer Science†
University of California, Santa Barbara, CA 93106

Abstract

Achieving effective real-time voice communication over an ad hoc network of mobile wireless nodes is an important new challenge in the wireless arena and opens new directions for research in speech coding. In this paper, we review the problems and issues in supporting speech over mobile ad hoc networks and examine the role of new speech coding techniques and modified network protocols aimed at providing adequate quality of service in this difficult environment.

1. Introduction

A mobile ad hoc network (MANET) is a wireless LAN (WLAN) wherein mobile nodes can communicate with one another and serve as routers for multi-hop connections without relying on any pre-existing infrastructure. MANETs are very attractive for a host of applications and environments, such as conference and convention center communications, as well as emergency response scenarios such as law enforcement and military activities. Real-time voice communication is a critical application for many of these network scenarios.

A MANET is characterized by peer-to-peer network structure, dynamic network topology, limited wireless bandwidth, low-power mobile terminals, high error rates, and the broadcast nature of wireless communications. Along with effective network protocols, suitable speech coding techniques are needed to ensure quality voice communication in MANETs.

In this paper, we focus on speech coding techniques and some modified network protocols for MANETs that take advantage of the character of speech. We start with an overview of the problem of speech communication in MANETs, and introduce the relevant networking issues, challenges, and current research activities in this area. Then we examine some techniques including multiplexing, multiple description speech coding, and scalable speech coding that contribute to effective and reliable voice communications in MANETs.

*This work is supported in part by the NSF under grants EIA-9986057, EIA-0080134, CCR-0243332, and ANI-0220118, ONR grant N00014-03-1-0090, the University of California MICRO Program, Dolby Labs., Lucent Technologies, Microsoft, Qualcomm, and Intel.

2. Voice Communication in MANETs

2.1. Network Protocols

Although many network protocols may be applicable to MANETs, a protocol stack based on the IEEE 802.11 physical (PHY) and medium access control (MAC) layers and UDP/IP for transport and network interface layers, shown in the oval in Fig. 1, is widely used for voice communications. Notably, UDP must replace TCP to allow real-time voice at the price of unreliable packet delivery.

Many networking studies have been conducted to analyze and improve the performance of MANETs. Routing in MANETs is very challenging due to mobility. Dynamic routing protocols are needed to quickly react to changes in network topology and re-establish routes. Several routing solutions have been proposed and some are being studied in the IETF MANET working group [1].

The widely known IEEE 802.11 family of standards [2] was originally designed for data communication in WLANs. The standard includes multiple protocols for the PHY layer and a common MAC protocol to interact with the selected physical layer. The 802.11 MAC layer allows two distinct access methods for transport over a link. Only one of these, the distributed coordination function (DCF), is applicable to MANETs. This method, called carrier sense multiple access with collision avoidance (CSMA/CA), mitigates collisions at the cost of limiting capacity and increasing latency. It also offers a retransmission mechanism for packets that are corrupted or whose receipt is not confirmed.

2.2. Network Characteristics

A typical diagram of a MANET is shown in Fig. 1. The dotted lines indicate wireless connectivity for nodes A, B, C and D. For a voice communication session between nodes A and B, multi-hop routing is needed. Two of many possible routes are shown in the figure; multiple paths are generally available in MANETs.

The loss characteristics of wireless channels have been empirically observed to be bursty due to various fading effects. The Gilbert-Elliott two state Markov model [3, 4] is often used

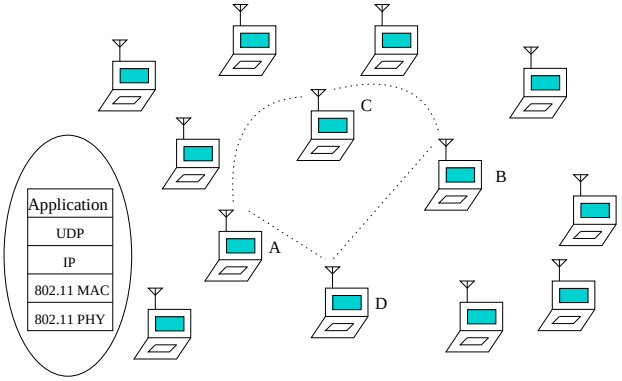

Figure 1. A typical MANET.

to model the bursty error channel. 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. Suitable probabilities are assigned for the transitions between states.

During the communication between nodes A and B, a packet at the MAC layer, which consists of the voice data, UDP header, IP header and MAC header, is lost or discarded in certain situations, including: (i) two or more nodes attempt to send packets in the same time slot and a collision occurs, e.g., nodes A and B are each sending a packet to node C, or nodes A and C attempt to transmit packets; (ii) the media is busy and bandwidth is not available, hence the packets are queued at a router and are dropped when the buffer overflows; (iii) a packet is received with bit errors and the number of retransmissions has reached the retry limit; and (iv) a packet in a jitter buffer at the receiver is discarded when its delay exceeds a threshold set by the application.

Although 802.11 allows for request-to-send (RTS) and clear-to-send (CTS) handshaking before a data packet is sent, for short data packets RTS/CTS is best disabled. Then, in most cases, the likelihood of losing a packet increases as the packet size grows. Thus, smaller packets are preferred for unreliable channels. Since the header size is the same for all data packets and the voice duration in each packet is limited by the application-specific delay requirement for two-way voice communication, the packet size varies with the size of the voice data allocated to a packet. Therefore speech compression is needed to minimize the packet size, bandwidth usage, packet delay, and packet loss. On the other hand, a relatively small payload with large headers implies inefficient use of bandwidth. One way to alleviate this inefficiency is multiplexing, as described later.

A transmitted packet in a MANET has a fairly large header, compared to the size of the speech data. Typically, a speech packet contains about 20 ms of speech data. In such packets, the speech data size ranges from 160 bytes (G.711 at 64 kbps) to 20 bytes (G.729 at 8 kbps). There are at least 58 bytes of data for the UDP, IP and 802.11 MAC headers. Moreover, to transmit a packet, additional overhead is needed for packet-by-packet synchronization, and the Physical Layer Convergence Protocol (PLCP) preamble (144 bits) and header (48 bits) at the PHY layer in 802.11 is transmitted at 1 Mbps regardless of the speed at which the MAC packet is transmitted. Thus, the transmission time for say 20 bytes of speech data is quite short compared to the time consumed for sending the PHY and MAC headers.

2.3. Challenges and Issues

For voice communication in a MANET, 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 high channel noise or heavy traffic with multiple concurrent communication sessions causing a high collision rate. Generally, retransmission of packets is not desirable for voice over MANETs since it adds extra delay and wastes bandwidth. Furthermore, communication might be lost if the packet loss rate is too high or a communication link is considered broken.

Another issue is the need for privacy and security in wireless voice communications. Conventional encryption may impose a nontrivial computational burden for low power handheld terminals.

To meet these challenges, effective network protocols and speech processing techniques are needed. The following network measurements are commonly used to objectively evaluate performance:

Packet loss rate is the fraction of transmitted packets from the source that are not received at the destination.

Jitter is the random variation of the packet inter-arrival time at the destination. Jitter can be many tens of milliseconds in 802.11. A jitter buffer at the receiving end can control a trade-off between speech quality and delay.

End-to-end packet delay is the difference between the packet transmission time at the source and its reception time at the destination. The limit on this delay varies with different applications. The end-to-end packet delay is a major component of the total voice delay between the time a voice sound element is spoken and the time that sound is heard at the receiving end. Other components include encoder and decoder processing delay and jitter buffer delay. The maximum acceptable voice delay for a two-way conversation may range from 80 ms to 400 ms depending on the degree of degradation deemed acceptable.

3. Related Work

3.1. Network Protocols

With the expanding interest in WLANs, there has been a great deal of effort to support voice over these networks. The

most efficient way to transmit voice data is to employ a reservation scheme at the MAC layer that guarantees delay and bandwidth. Many different reservation schemes have been studied to enable speech or multimedia transmission [5–9]. Moreover, a new standard, IEEE 802.11e, 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 [10].

Furthermore, in MANETs, QoS mechanisms have been incorporated by extending the routing protocol to assure certain QoS requirements. Many multiple path routing protocols have been proposed for real-time communications because they decrease the number of route discoveries and eliminate route discovery latency after link breaks by making use of the availability of multiple path in a MANET.

There are many multiple path routing protocols for MANETs, such as Ad-hoc On-demand Multipath Distance Vector (AOMDV), Ad-hoc On-demand Distance Vector Backup Routing (AODV-BR), Dynamic Source Routing (DSR), Split Multipath Routing (SMR), and Zone Routing Protocol (ZRP), etc. [11]. In general, these protocols are designed for different applications in different networks, such as load balancing, or power awareness in heterogeneous or homogeneous wireless networks.

3.2. Speech Coding

Over the years, highly effective speech compression algorithms have been developed with increasing sophistication. A large number of speech coders have been standardized for various applications. 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 networks, G.729 at 8 kbps is widely used with very good speech quality. There is also an extensive set of speech coding standards for cellular networks such as the Adaptive Multi-Rate Narrowband (AMR-NB) speech coder standardized by ETSI in 1998. More recently, the AMR-WB speech coder [12] 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. Most recently for CDMA 2000, the variable multi-rate wideband (VMR-WB) coder was standardized and includes limited interoperability with AMR-WB.

In addition to these standards, recent studies have been directed to multiple description (MD) coding as well as scalable coding (SC) of speech [13]. MD coding offers an interesting way to cope with packet loss and transmission erasures, while SC has potential for effective communications in a heterogeneous multimedia network. Two SC speech standards exist, namely G.727 (ADPCM) and MPEG-4 speech coding tools. However, no MD speech coder has yet been standardized.

4. Speech Coding Techniques for MANETs

A number of techniques are available in cellular and VoIP to provide different QoS levels. For example, multiplexing has been used to improve the link quality of the wireless downlink in a centralized system. Unequal error protection is used in cellular systems for efficient communication by exploiting the speech bit stream structure. Some of these techniques can be applied to MANETs to help achieve stringent QoS specifications.

For MANETs, the choice of speech coding technique depends on network conditions and the quality requirement of the underlying application. At this stage, it is advantageous to make use of existing speech coders. For instance, G.711 is adequate when the network is lightly loaded and it offers very good telephone speech quality. Alternatively, a low bit rate coder such as AMR-NB can be beneficial [14], since its bit rate can adapted to suit channel conditions, and a seamless interconnection between a wireless LAN and a wireless cellular system is possible.

To further address QoS requirements in MANETs, we have been studying both MD and SC speech techniques since both schemes show promise for adaptive and robust real-time multimedia communication over lossy networks. Both these techniques, while distinct from one another, offer the possibility of improving the quality of the received and reconstructed speech by transmitting subsets of the speech data over different paths.

4.1. Selective Error Checking

The MAC layer of 802.11 includes a CRC check on all bits in the MAC frame. However, many of the speech bits in a packet can tolerate errors while other bits are critical for effective recovery of the speech. Hence, retransmissions and dropping of MAC frames can be reduced by limiting the CRC to the headers and the more sensitive part of the speech data. This technique, called selective error checking (SEC), has been simulated at UCSB with the AMR-NB coder. The results have shown that SEC substantially improves the network performance and the speech quality.

4.2. Multiplexing Multiple Voice Channels

As discussed earlier, the high ratio of overhead to payload in a voice packet causes a very inefficient use of bandwidth. In recent work at UCSB, we have studied a way to reduce or avoid this inefficiency. Multi-channel speech data arriving at a node is multiplexed into a single packet with connection ID information in the payload indicating which segment is destined for which neighbor. The mapping between such connection IDs and the neighbor for which the segment is intended can be updated at regular intervals whose lengths depend on the rate of network topology changes. This allows a time segment of multiple voice conversations to be efficiently carried on a

single packet over that part of the route that they have in common. Later, these segments may be split and continue to their destination in separate packets.

Although the packet collision rate increases as a packet size increases due to multiplexing, there is nevertheless a significant reduction in the overall rate of collisions in the network. The relative overhead due to the header is reduced, and each node sends fewer (but larger) packets, thus relieving the burden on the MAC layer. With multiplexing in 802.11, the rate of packet generation per node is reduced due to multiplexing, thereby reducing collisions; On the other hand, the size of each packet is increased by multiplexing, thereby increasing collisions. The former effect is dominant when the payload size is small compared to the header size, since the increase in the overall packet size due to multiplexing is small. We note also that the performance of multiplexing can be improved by use of a repetition code or MD coding.

4.3. Multiple Description Speech Coding

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

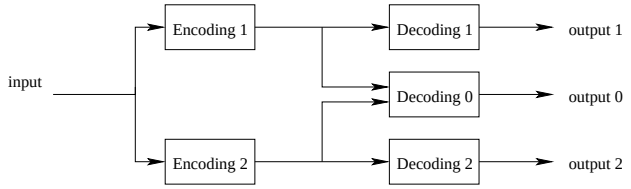

Figure 2. A multiple description coding scheme.

Generally, there are two approaches to designing MD coders. One approach 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 their bit rates. Another approach is to constrain the performance of the individual descriptions and then attempt to minimize the distortion of the joint description for a given constraint on its bit rate. A key issue in designing MD coders is to effectively minimize redundancy between the two descriptions while trying to minimize distortion in the individual descriptions, or to effectively minimize distortion in the individual descriptions while trying to minimize distortion in the joint description.

Both approaches have advantages depending on the transmission conditions. If both descriptions are received most of the time, a high quality from the joint description is of prime importance. If only one description is frequently received, it is more important to obtain good quality from a single description even if the quality of the joint description is compromised. 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 [15–19].

MD speech coding is a promising approach to transmission of packetized speech in MANETS. The availability of multiple paths in MANETs and the severe network operating conditions both suggest that MD coding can be beneficial. MD coding increases the likelihood that at least one description of any particular speech segment reaches its destination while avoiding extra delay due to retransmissions under severe network conditions.

Recently, we developed an MD speech coder based on AMR-WB for MANETs, where the MD speech coder is employed during severe channel conditions. The MD coder splits the bit stream 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 sub-stream arrives at the decoder, degraded but still acceptable speech quality is obtained. Our simulation results showed that the MD coder is beneficial for reliable voice communication under severe channel conditions in MANETs. Therefore, an MD coder with an AMR coder offers the promise of making effective use of the channel capacity and providing reliable end-to-end connections for voice communication in MANETs.

4.4. Scalable Speech Coding

Scalable speech coding consists of a minimum rate bit stream that provides acceptable coded speech quality, along with one or more enhancement bit streams, which when combined with a lower rate coded bit stream, provide improved speech quality. Usually, scalability is implemented in a layered structure, shown as Fig. 3.

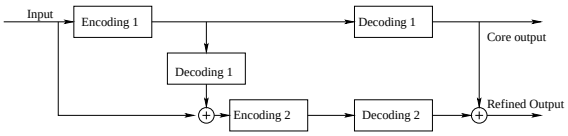

Figure 3. The typical diagram of scalable speech coding system.

Scalable speech coding has a number of advantages. For example, it allows service interworking. Using a scalable approach, users can receive different quality versions of the same source according to available bandwidth. It also offers flexibility for error protection. The high-priority information can be transmitted over a more reliable transmission path or using an unequal error protection of the core and the enhancement

layers. Moreover, it provides an effective encryption option in wireless networks, where only the core bit stream is encrypted.

Scalable coding is attractive for MANETs due to the diversity of the network resource. We suggest applying the layered bit streams of scalable speech coding in MANETs in two ways: either all layers in one packet or each layer for one packet. In either case, the enhancement bit streams can be added on or dropped off according to network conditions such as power, bandwidth, or quality requirement. Therefore scalable coding offers the flexibility to utilize the network resources.

Recently, a new application of scalable coding is discussed to secure voice communication, where only the core bit stream of the data stream is encrypted. The remainder of the data stream is sent in the clear, as shown in Fig 4. It is shown that encryption of the core layer only is sufficient to ensure a high level of protection against eavesdroppers, thus significantly reducing the signal processing power needed for encryption and decryption in comparison to encryption of the full bit stream.

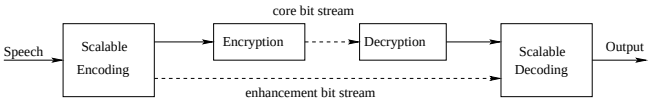

Figure 4. Scalable coding with selective encryption.

5. Summary and Conclusions

Supporting speech over MANETs requires new speech coding techniques as well as modified network protocols that exploit the characteristics of speech and often leverage these techniques. MD and SC are attractive schemes to suit the changing channel conditions on a multi-hop time-varying network. MD speech coding provides a way to achieve reliable transmission without increasing delay. Scalable coding offers flexible use of available network resources and a simple way to enable voice privacy. Selective error checking is a simple MAC layer modification to enhance performance and multiplexing multiple conversations enhances performance by reducing collisions. These techniques as well as other new directions show promise but need much further study to achieve adequate QoS on MANETs.

References

[1] S. Corson and J. Macker, "Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation consideration," *RFC*, Jan. 1999.

[2] ISO/IEC 8802-11, ANSI/IEEE Std.802.11, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,*, 1999.

[3] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, pp. 1253–1265, 1960.

[4] E. O. Elliot, "Estimates of errors rates for codes on burst-noise channels," *Bell System Technical Journal*, vol. 42, pp. 1977–1997, 1963.

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

[6] V. N. Muthiah and W. C. Wong, "A speech-optimised multiple access scheme for a mobile ad hoc network," in *First Annual Workshop on Mobile and Ad Hoc Networking and Computing*, Aug. 2000, pp. 127–128.

[7] H. Wu, C. Hung, M. Gerla, and R. Bagrodia, "Speech support in wireless, multihop networks," in *Third International Symposium on Paprallel Architectures, Algorithms, and Networks Proceedings*, Dec. 1997, pp. 282–288.

[8] 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.

[9] C.-H.R. Lin and M. Gerla, "A distributed control scheme in multi-hop packet radio networks for voice/data traffic support," in *IEEE International Conference on Communication*, 1995, vol. 2, pp. 1238–1242.

[10] *http://www.ieee802.org/11*.

[11] *http://en2.wikipedia.org/wiki/Ad hoc protocol list*.

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

[13] H. Dong, *SNR and bandwidth scalable speech coding*, Ph.D. thesis, Southern Methodist University, Dallas, Texas, December 2002.

[14] 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.

[15] N. S. Jayant, "Subsampling of a DPCM speech channel to provide two 'self-constrained' half-rate channels," in *Bell System Technical Journal*, 1981, vol. 60, pp. 501–509.

[16] 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.

[17] A. K. Anandakumar, A. V. McCree, and V. Viswanathan, "Efficient CELP-based diversity schemes for VoIP," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, June 2000, vol. 6, pp. 3682–3685.

[18] 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.

[19] X. Zhong and B.-H. Juang, "Multiple description speech coding with diversity," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, May 2002, vol. 1, pp. 177–180.