Supporting Real-time Speech on Wireless Ad Hoc Networks: Inter-packet Redundancy, Path Diversity, and Multiple Description Coding

Chi-hsien Lin Hui Dong Upamanyu Madhow Allen Gersho
Department of Electrical and Computer Engineering
University of California – Santa Barbara
Santa Barbara, California
{chlin, huidong, madhow, gersho}@ece.ucsb.edu

ABSTRACT
We consider the problem of supporting real-time traffic over packetized wireless ad hoc networks. Our specific emphasis is on speech, since this is a critical application in many scenarios such as emergency deployment of ad hoc networks. Standard retransmission-based Medium Access Control (MAC) strategies are poorly matched to speech applications, because the payload size for speech as well as for MAC-layer acknowledgements (ACKs) is small compared to the packet header, which contains a large synchronization preamble. In this paper, we show that inter-packet redundancy is significantly more efficient than traditional MAC layer retransmissions, in terms of both network capacity and end-to-end delay. The key observations regarding our design and results are as follows. Because of the small payloads, introducing redundancy across packets only increases the packet transmission time slightly, and hence has negligible impact on the packet collision rate. Thus, we obtain large gains from redundant transmission essentially "for free." Because of the large packet header, elimination of ACKs leads to substantial bandwidth savings. Overall, a combination of inter-packet redundancy (at the MAC layer), path diversity (at the network layer), and multiple description source coding (at the application layer), is shown to provide significant improvements in bandwidth efficiency and delay.

Categories and Subject Descriptors
C.2.1 [**Network Architecture and Design**]: Network communications, Packet-switching networks, Wireless communication

General Terms
Performance, Design

This work was supported by the National Science Foundation under grants ANI-0220118 and EIA-0080134, and by the Office of Naval Research under grant N00014-03-1-0090.

Keywords
Wireless, ad hoc, 802.11, real-time, speech, path diversity

1. INTRODUCTION
Much progress has been made over the past few years on set-up and routing in wireless ad hoc networks [1][2]. Such networks have huge potential for plug-and-play and emergency deployments. A key application in many such scenarios is the support of real-time traffic, especially voice. There are, however, two major technical challenges that must be overcome in order to achieve this. Firstly, ad hoc networks typically have fairly large packet loss rates, due to collisions arising from decentralized transmissions, as well as due to the bit errors inherent to even a collision-free wireless channel. While retransmissions and buffering are effective mechanisms for combating loss for delay-insensitive data applications and non real-time audio, they have limited utility for real-time traffic with strict delay constraints, especially when the traffic must traverse more than one wireless hop. Secondly, the overhead in packetized wireless networks must include not only MAC and network layer information, as in wireline communication, but also a preamble for synchronization, which must typically occur on a packet-by-packet basis. Thus, the efficiency of such networks is poor for small data payloads, as in applications such as voice.

In this paper, we propose and evaluate a number of mechanisms that, compared to simply deploying a speech application over standard ad hoc network infrastructure, provide significant gains in the end-to-end frame loss rate, while simultaneously improving bandwidth efficiency. The main ingredients are as follows:

1) "Free" redundancy and ACK elimination: While the redundancy required for standard forward error control reduces the bandwidth efficiency, we introduce redundancy in a fashion that actually increases bandwidth efficiency, while providing large reductions in end-to-end frame loss rate. This is achieved by exploiting the small payload size of speech, relative to the large headers required in typical packetized wireless networks (e.g., those based on 802.11abg). Specifically, we consider the following simple technique, which can be generalized easily. Instead of sending speech frame d_n, as the payload of the n_{th} packet, we concatenate frames d_{n-1} and d_n as the payload of packet n. Thus, frame n is received successfully if either packet n or packet n+1 is successful. On the other hand, putting two speech frames in the payload only slightly increases the overall packet length (which is dominated by the header), and hence the packet

loss probability p is increased only marginally. In addition, the probability of frame loss is reduced by orders of magnitude, since it now scales as p^2, compared to p in a system that does not employ redundancy or retransmissions. Since we do not employ retransmissions, we can eliminate ACKs. A useful point of comparison is a retransmission-based MAC which allows at most one retransmission. Note that the success probability of a frame again scales as p^2, where p is the packet loss rate, but there is significantly more bandwidth expended than in our inter-packet redundancy method. If no retransmission is needed, then we must expend bandwidth in transmitting an ACK (which has roughly the same transmission time as a speech packet, since both have small payloads compared to the header). Further, if a retransmission is needed, again we expend another packet's worth of overhead.

2) Path diversity: Wireless networks suffer from a number of impairments that can lead to a string of consecutive packet losses; these include fading and changes in connectivity caused by mobility. While our inter-packet redundancy method deals well with random packet loss, path diversity is required for handling correlated losses along a given path.

3) Multiple Description Coding: Multiple Description (MD) coding is a source coding techniques that combines well with path diversity. MD coding splits the information sequence (speech, in our case) into two (or more) equally important streams. The received quality is best if both streams are received, but is acceptable even if only one of the streams is received. Thus, sending MD streams over different paths provides robustness against correlated packet losses. However, applying MD will increase the total number of packets that a source node sends into the network. Thus, an effective way to increase bandwidth efficiency is crucial to help release the extra network load, and the inter-packet redundancy scheme is a good candidate.

Each of the preceding schemes produces gains in performance, and typically operates at a different layer of the OSI hierarchy. Inter-packet redundancy, which provides the biggest gains, is ideally deployed at the link/MAC layer, thereby permitting hop-by-hop packet reconstruction for recovering from isolated losses. (It could also be deployed at the application layer, but then hop-by-hop packet reconstruction would not be possible). Path diversity operates at the network layer, while MD coding is at the application layer. We consider an integrated system combining inter-packet redundancy, path diversity, and MD. The system is strongly robust against random packet loss due to the use of inter-packet redundancy, and is resistant to bursty loss because of path diversity and MD. Our simulations show that the proposed system can indeed provide good delay and loss performance for real-time speech over a difficult operating environment such as a multihop wireless network. Given the widespread availability of 802.11 based Wireless Local Area Networks (WLANs) [3][4], the parameters in our performance evaluation are consistent with those of 802.11b, using the Distributed Coordination Function (DCF) for peer-to-peer communication. There are four allowable rates in 802.11b: 1, 2, 5.5 and 11 Mbps. The physical layer convergence protocol (PLCP) handles synchronization using PLCP preamble and header transmitted at 1 Mbps. Thus, if the MAC Protocol Data Unit (MPDU) is small, as for voice applications, and is transmitted at 11 Mbps, for example, then the time to transmit the PLCP preamble and header becomes significant compared to the transmission of the MPDU, or payload.

The organization of this paper is as follows. Section 2 briefly discusses related work. Section 3 first points out the significant overhead in 802.11, and then describes the proposed inter-packet redundancy scheme and the supplemental packet reconstruction function, then discuss the performance of the proposed scheme. A rough analysis is given, to provide intuition into the expected performance gains (detailed simulations are given in Section 6). An interesting postive feedback in 802.11 MAC is also discussed. Section 4 provides analytical insight into the gains due to path diversity, for a simple Gilbert-Elliott model of correlated packet losses. Section 5 describes the MD coder to be used in our integrated system. Simulation results are provided in Section 6. We first compare the inter-packet redundancy with a conventional retransmission-based MAC. We then demonstrate the additional gain due to path diversity. Finally, the performance of the integrated system is simulated. Section 7 contains concluding remarks.

2. RELATED WORK

Prior attempts to improve the performance of real-time traffic over ad hoc networks include [5-15]. These include ideas such as optimizing packet length, employing forward error control within a packet [5][6], reservation policies [7-10], bandwidth reuse technique [11], retransmission strategy [12], and performance evaluation technique [13]. These papers do not exploit small payloads as we do, and do not employ inter-packet redundancy. Coding across packets has been proposed in the context of wireline networks [14][15]. However, the realization that coding gains can be obtained "for free" for small payloads (relative to the large overhead in packetized wireless communication) appears to be completely novel.

Path diversity has been considered in [16-20], and many multiple path routing protocols based on different routing schemes over wireless ad hoc networks have been proposed [21][22]. However, the focus has mainly been on the support of delay-tolerant data applications, rather than on improving end-to-end performance for real-time traffic. MD coding has a rich history [23-30], and has been discussed for theory [23-26] and image/video applications [27-29]. Applying MD over a wireless environment for both image/video [30][31] and voice [32] is a relatively recent development. To the best of our knowledge, however, there is no prior work that integrates design concepts across multiple layers to provide an integrated system suitable for real-time speech applications over wireless ad hoc networks.

3. INTER-PACKET REDUNDANCY

In the following, we describe our inter-packet redundancy scheme. While the concept is broadly applicable for any packetized system in which the payload is small compared to the overhead, we describe, for concreteness, an example design for supporting voice over an 802.11b based ad hoc network.

3.1 Overhead in a Conventional 802.11 system

First we consider speech transmitted over a conventional system to give a rough idea of the overhead relative to the payload size. Typical speech datagram size ranges from 160 bytes (G.711 at 64 kbps) to 20 bytes (G.729 at 8 kbps) or smaller. These speech datagrams are typically generated every 20 ms. Suppose that neighbors can communicate at the highest rate of 11 Mbps. If we have a 20-byte voice frame to be sent at 11 Mbps, the payload transmission time is

$$8 \times 20 / 11 = 14.55\ (\mu s),$$

The transmission time of headers, including 802.11 MAC, IP and UDP header is

$$8 \times (28 + 20 + 8) / 11 = 40.73\ (\mu s),$$

where 28, 20, 8 bytes are 802.11 MAC header, IP header, and UDP header, respectively. The overhead, in terms of transmission time, due to the header is 280%. According to the 802.11 standard, the synchronization time is

$$(144 + 48) / 1 = 192\ (\mu s),$$

where 144 bits is the length of PLCP preamble, 48 bits is the length of PLCP header, and 1 Mbps is the transmission rate for PLCP preamble and header specified in the 802.11 standard. Thus, the overhead due to the PLCP preamble and header is 1320%. MAC layer ACKs lead to additional overhead, which takes

$$10 + 192 + 8 \times 14 / 11 = 212.18\ (\mu s),$$

to transmit, where 10 μs is the SIFS, 192 μs is the synchronization time mentioned above, 14 bytes is the size of ACK packets, and 11 Mbps is the transmission rate. Thus, the overhead due to ACK is 1458%. Fig. 1 shows actual proportion of the transmission time devoted to synchronization, headers, and payload. Note that the transmission time for the data and ACK packets are almost the same, and that the overhead relative to the payload is huge in both cases. We do not calculate the overhead introduced by RTS/CTS because according to the 802.11 standard, RTS/CTS mechanism is not used if MAC frame size is smaller or equal to RTS threshold (dot11RtsThreshold) and the default RTS threshold is 3000, which is much larger than small packet sizes that we consider.

The preceding overhead computations were for 802.11b. The percentage overhead is a little smaller, but still very large, for newer OFDM-based WLAN standards such as 802.11a/g. In the latter, the PLCP preamble and header takes 20 μs to transmit, and the MAC, IP, TCP header takes three symbol interval ($3 \times 4 = 12$ μs), while the actual payload only takes one symbol interval (4 μs). The overhead due to synchronization and headers is still very significant (800%). The overhead due to ACK (28 μs) is large as well (700%).

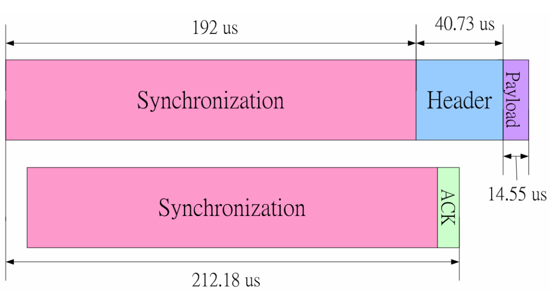


Fig. 1: The overhead of 802.11b

3.2 Adding Redundancy

The inter-packet redundancy scheme that we propose is designed to take advantage of the significant overhead stated above. Consider a voice stream consisting of a sequence of segments $\{d_n\}$. We include a copy of dn in the payload for the packet carrying the next segment d_{n+1}.Thus, d_n is lost only if both the n_{th} and $(n+1)_{\text{th}}$ packets are lost. The redundant segment is equivalent to one MAC retransmission, without incurring the significant overhead associated with retransmissions, including the packet overhead incurred in retransmitting a small payload (if the packet is unsuccessful), and the ACK overhead incurred (for indicating a successful packet for positive acknowledgements, as in 802.11, or for indicating failure, as in a negative acknowledgement based scheme). Note that our method can be generalized to provide functionality equivalent to more than one retransmission, by sending more than two consecutive segments in each packet. Whether or not this is useful depends on the reconstruction delay allowed at the intervening nodes (for a MAC layer implementation) or at the receiver (for an application layer implementation). It also depends on how the increase in payload size impacts the overall packet size, and hence the probability of packet loss due to collisions and bit errors. We confine our simulations to the case of two segments per packet, keeping in mind end-to-end delay constraints. In this case, the increase in packet size, and hence in collision probability, is negligible, thus providing the promised "free" coding gains.

Fig. 2 shows a conventional one-retransmission scheme with our inter-packet redundancy scheme. Note that no ACK is transmitted in inter-packet redundancy scheme. Since actual payload is so small compared to synchronization and header transmission time as stated, the increased transmission time due to attaching dn to the $(n+1)_{\text{th}}$ packet is negligible.

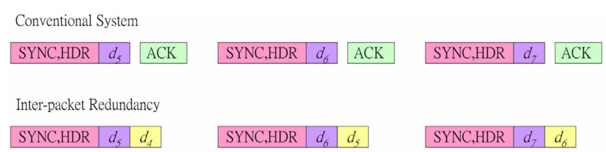


Fig. 2: Conventional system vs. inter-packet redundancy

3.3 Packet Reconstruction

Packet reconstruction is an optional function which can help further reducing the end-to-end segment loss rate. Let (d_n, d_m) denote a packet containing segments d_n and d_m. For simplicity, we assume nodes rearrange the segments in a packet before sending out the packet such that the sequence number of first segment is always larger than the sequence number of the second segment, namely, $n > m$, and $n = m + 1$ if no packet loss has happened.

When a packet (d_n, d_m) arrives at a forwarding node, the node looks into the packet, records segment dn in its memory, and then forwards the packet to its next hop. When another packet from the same stream (d_r, d_s) comes, the node looks into the packet again and checks the sequence number to see if there is any packet got lost. If the difference between larger sequence numbers (the larger sequence number is n for segment (d_n, d_m) and r for segment (d_r, d_s) here) of two successive packets is greater than one, then packet reconstruction is activated. The node picks up d_n from its memory, combines d_n with ds and then sends out a new packet $(d_s,$

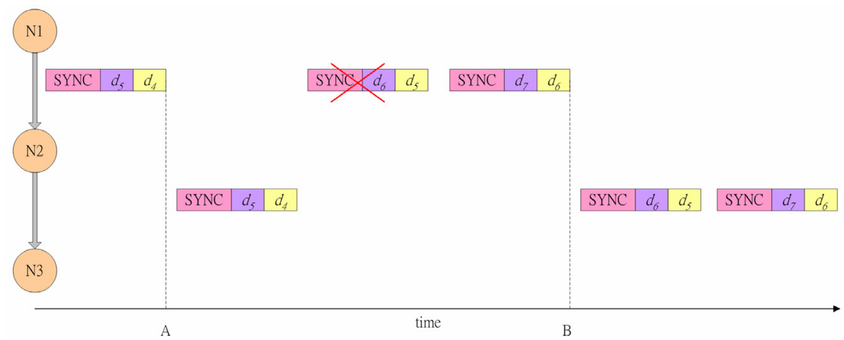


Fig. 3: Packet reconstruction

d_n) before forwarding packet (d_r, d_s). The reconstructed packet is sent only when n is different from s. Note that the d_n in memory will be replaced by dr before (d_r, d_s) is forwarded. The node only keeps d_n in its memory for a predefined period of time, d_n will be deleted from the memory if the node has been waiting too long for the next packet from the same stream.

Fig.3 is an example of packet reconstruction. N1, N2, N3 are forwarding nodes, packets are forwarded form N1 through N2 to N3. At time A, N2 received (d_5, d_4) and put d_5 in its memory. After N2 forwarded (d_5, d_4) to N3, (d_6, d_5) from N1 got lost in the channel. Since N1 did not know that there was a packet loss, it continued sending (d_7, d_6) to N2. At time B, N2 received (d_7, d_6) successfully and detected the loss. It then retrieved d_5 from its memory, reconstructed and sent the lost packet (d_6, d_5) before forwarding (d_7, d_6) to N3. N3 did not even realize that a loss had occurred. Note that the loss recovery occurred without requiring an increase in the number of packets sent, unlike in a conventional retransmission-based scheme.

Packet reconstruction provides a mechanism to recover from isolated packet loss, thus preventing performance degradation when the hop count of a path gets large. With packet reconstruction, only two successive packet losses will cause one segment loss.

3.4 Performance of Inter-packet Redundancy

We want to compare our inter-packet redundancy system with conventional 802.11 MAC. To make the two systems comparable, we focus on an 802.11 system with retransmit limit of 1, namely, the 802.11 MAC has one chance to recover from packet loss, and it requires at least two successive packet losses to incur one segment loss.

In this section, we provide insight into the relative performance of the two methods via rough calculations. Their detailed behavior and performance is investigated using simulations in Section 6.

3.4.1 Packet Error Rate and Collision Probability

First we consider the probability of single transmission failure, namely, packet error rate and collision probability. The impact of retransmissions is discussed later.

We show that the increase in payload size of inter-packet redundancy slightly increases the packet loss rate due to bit errors. However, the decrease in overall traffic due to elimination of ACKs and retransmissions results in a significant decrease in the packet loss rate due to collisions. At the operating Signal-to-Noise Ratios of interest, bit errors are infrequent, so that the second effect is much more significant. Thus, the inter-packet redundancy technique provides significant gains over the one-retransmit scheme.

Packets can be dropped at the receiver end if some bit errors has occurred during the transmission. Here we use packet error rate (PER) to compare the two systems. Denote PER of the original packet as P_p, and the PER of the packet with redundancy as P_p'. Assuming the two systems have the same bit error rate and bit errors are independent, we can estimate the relationship between P_p and P_p':

$$P_p = 1-(1-P_p')^{\frac{b}{a}} = 1-(1-\tfrac{b}{a}P_p' + h.o.t) \approx \tfrac{b}{a}P_p',$$

where a is the number of bits in the packet with redundancy and b is the number of bits in the original packet. Higher order terms are negligible for small P_p'. Note that a, b both include MAC, IP, and UDP headers. For the example in Section 3.1, a = 96 and b = 76, so $\frac{b}{a}$ is 0.79, which is also the PER ratio of inter-packet redundancy system to one-retransmission 802.11 MAC.

We now consider the collision probability. Since RTS/CTS is not employed for small packets that we are considering as stated before, these short packets can only rely on CSMA/CA to avoid collision, and are therefore vulnerable to the hidden terminal problem.

For hidden nodes, the collision probability is roughly proportional to transmission time per packet and the average channel busy time. Note that the transmission time here includes and is dominated by synchronization time. According to parameters in Section 3.1, the packet transmission time ratio of one-retransmission 802.11 MAC to inter-packet redundancy is

$$\frac{192+40.73+14.55}{192+40.73+14.55\times 2} = \frac{247.28}{261.83} = 0.94,$$

and the average channel busy time ratio of one-retransmission 802.11 MAC to inter-packet redundancy is

$$\frac{247.28+212.18}{261.83} = \frac{460.45}{261.83} = 1.75,$$

where 212.18 is the time used by ACK. The use of ACK packets makes the channel busy time much longer, and therefore degrades the performance of the one-retransmission scheme.

For non-hidden nodes, collision probability is not so closely tied to the packet transmission time. Instead, it is determined by the collision avoidance mechanism. From this point of view, having more packets to contend for the channel increases the collision probability. The main difference between the two systems here is that inter-packet redundancy has no ACK, backoff and retransmission. For one-retransmission 802.11 MAC, backoff can reduce the probability of collision, but this only helps the retransmissions. While retransmitted packets are less likely to incur collisions due to its larger contension window, they still increase the number of packets contending for the channel relative to the inter-packet redundancy scheme. Further, while ACKs do not suffer collisions with packets from non-hidden nodes in the 802.11 MAC, the cost in terms of transmission time (and hence overall capacity) is significant.

3.4.2 Behavior under High Network Loads

Now we take the retransmission into account and consider the behavior of the two systems when the network load is heavy.

Fig. 4 shows the transmission cycle of the one-retransmission 802.11 MAC. There is an interesting postive feedback in 802.11 MAC: When the network load is heavy, the collision probability gets high, and thus more losses incur more retransmission. However, this behavior implies that nodes are always injecting more traffic into the network when network condition is bad and making the condition worse. On the other hand, more retransmission implies less channel efficiency, so 802.11 MAC is always spending more transmission chances to transmit a packet when the chances are rare. Another important factor is that, even after the packet is successfully received, retransmission can still happen because there is still a chance that the ACK packet get lost and cannot be received by the sender. Besides, no matter how many retransmissions is allowed, the last ACK is always useless since the sender cannot retransmit due to the limit. These factors make the performance of 802.11 MAC degrade dramatically when traffic load gets high.

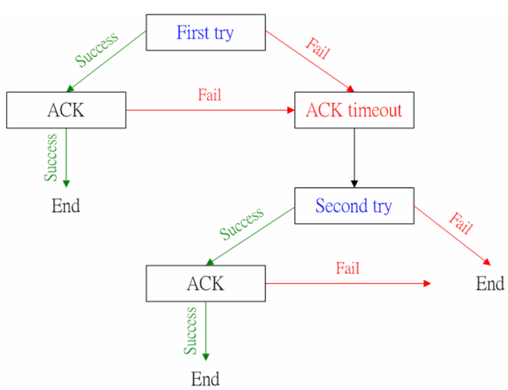


Fig. 4: Transmission cycle of one-retransmission 802.11 MAC

For our inter-packet redundancy technique, things are totally different. Although we also have the ability to recover form single packet loss, we are not introducing any unexpected traffic at any time. This is especially good for real-time applications since we can keep end-to-end delay under control when traffic load gets heavier. Also, since increased traffic does not imply more retransmissions, the loss rate increases gradually with the traffic load rather than the steep increase seen in retransmission-based schemes once the traffic load exceeds a threshold.

4. PATH DIVERSITY

Packet losses in ad hoc networks can be categorized into two kinds: random losses and bursty losses. Bursty losses can be due to routing failure, bursty noise, IP buffer overflow, and mobility. Random losses usually result from collisions or noise. Both types of packet loss are much more prevalent in wireless networks than in wired networks. Bursty losses are a more severe impairment for real-time applications such as speech, since interpolation-based error resilience techniques can be employed at the destination to alleviate degradation due to random loss. While coding and interleaving over a large number of packets is one possible mechanism for dealing with bursty loss, it does not apply to delay-constrained applications such as speech. In this situation, path diversity is a simple but effective method, assuming that we can find multiple paths which are unlikely to incur a burst of packet losses at the same time (e.g, use the Disjoint Path Selection Protocol in [22]). In this section, we provide a quick calculation for estimating performance gains from path diversity, using a Gilbert-Elliott model for bursty packet losses. Detailed simulations are postponed to Section 4.

Suppose there are two paths, path 1 and path 2, modeled by two independent Gilbert-Elliott models, each as shown in Fig. 5. Paths can be in either good or bad state, and can switch their states according to Markov model, depicted in Fig. 5. Here the transition rates from good state to bad state and from bad state to good state are λ and μ, respectively. Packet loss rates of a path are different while in different states, and packet loss rates of different paths in the same state are not necessarily the same. Here we set the packet loss rate in bad state to be 1 for both paths, and set the packet loss rate in good state to be p_1 and p_2 for path 1 and path 2, respectively.

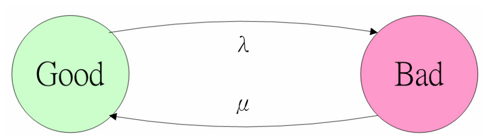


Fig. 5: Two state Gilbert-Elliott model

If only one out of the two paths is randomly chosen to transmit through, the expected loss rate is $\frac{\mu\left(\frac{p_1+p_2}{2}\right)+\lambda}{\lambda+\mu}$, and the expected time in bad state, i.e., the expected burst time, is $\frac{1}{\mu}$.

If we use two paths at the same time, the Gilbert-Elliott model leads to a two dimensional Markov chain as shown in Fig. 6. There are four states, (G, G), (G, B), (B, G), and (B, B), where the first element represents the state of path 1 (G for good, B for Bad) and the second element represents the state of path 2. Since the two paths are assumed independent, the transition rates are still λ and μ as in Fig. 5.

When both path 1 and path 2 are used, the expected loss rate remains the same as randomly choosing a path, which is $\frac{\mu\left(\frac{p_1+p_2}{2}\right)+\lambda}{\lambda+\mu}$. But the average burst time (the average time stay in (B, B) state) is halved, which is $\frac{1}{2\mu}$. Of course, when one path is significantly

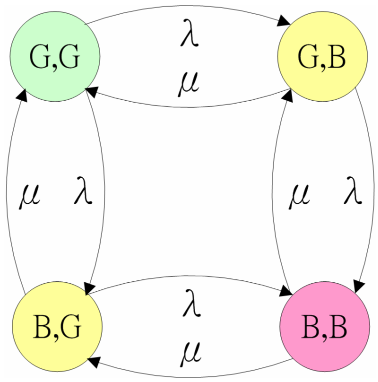


Fig. 6: Gilbert-Elliott model for two paths

superior to another (e.g. in terms of packet loss rate in Good state), and the identity of the better path is known to the source (this information is typically not available from standard ad hoc routing protocols), then the tradeoffs between random and burst loss, from an end-to-end application perspective, must be carefully considered in deciding whether, and in what form, to use path diversity. However, these simple calculations do provide motivation for devising application layer strategies that can exploit path diversity, which leads us into the MD coding techniques discussed next.

5. MULTIPLE DESCRIPTION CODING

An MD coder produces two or more descriptions, or coded bit streams, from a given source signal. Fig. 7 shows the block diagram of the coder and decoder for two descriptions. Bit streams independently represent "coarse" descriptions of the source (output 1 and output 2), while multiple descriptions jointly convey a "refined" source representation (output 0). Different bit streams are transmitted to the receiver separately, usually through different paths to profit from path diversity as showed in last section. If any one of the bit streams is received, the decoder can choose a proper decoding procedure and provide a degraded but acceptable quality of speech. If all the bit streams are successfully received, the decoder will be able to reconstruct better speech. Note that all the descriptions in a MD code are equally important unlike layered coding in which the higher layers are useless if the coarsest layer is not received correctly.

One way to design MD speech coding is to allow some redundancy of two descriptions. So as long as at least one description is received, the decoder will have the basic information and be able to estimate the lost information, which enables the decoder to reconstruct a degraded but acceptable quality.

The MD coder used in this paper is designed to make use of the wideband speech standard, AMR-WB. Each frame of AMR-WB contains 20 ms speech data, and the size of the 12.65 kbps frame is 253 bits. The MD coder splits the bit-streams into two redundant sub-streams, 136 bits and 134 bits, by directly selecting overlapping subsets of encoded data generated for each frame. A full quality at 12.65 kbps is ensured if two descriptions are received and a degraded quality at 6.8 kbps is achieved if only one description is received.

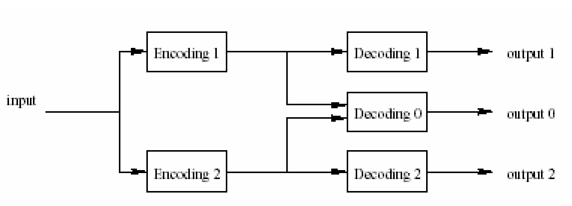


Fig. 7: Block diagram of multiple description coder

A very important thing to be noticed about MD is that, when MD splits the source in to two bit streams, the inter frame interval still remains the same, in other words, the total number of packets that a source node sends into the network is actually doubled. If everyone on the network adopts path diversity and MD coding, the number of connections that a network can support will be reduced given a fixed network capacity. Thus, it is very important to combine MD together with the inter-packet redundancy since it halves the number of packets transmitted into the network and provides better bandwidth efficiency.

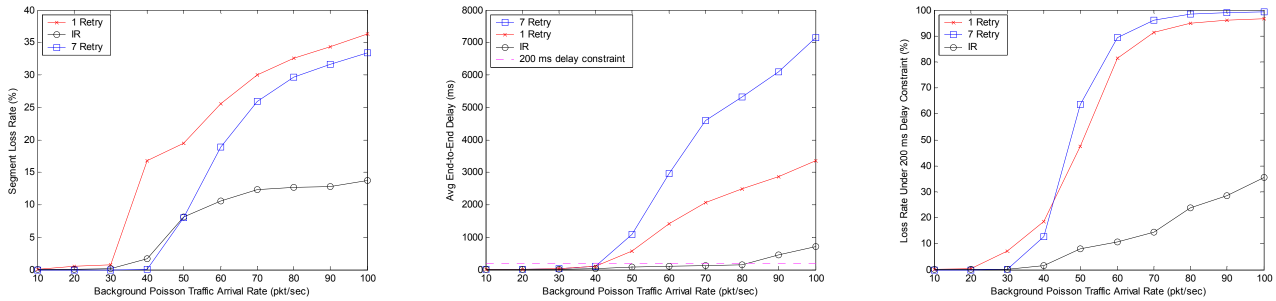
6. PERFORMANCE EVALUATION

Glomosim, a scalable simulation library for wireless network systems developed by UCLA, is used as the simulation tool for evaluating the proposed techniques [33]. For the simulation, 16 nodes are uniformly distributed in an area of 800 × 800 square meters. The distributed coordination function (DCF) of IEEE 802.11 is used as the MAC layer, and a two-ray propagation model is used to model signal propagation. The transmission rate is chosen to be 11Mbps.The traffic pattern of our tagged stream is selected as CBR. At each node, the background traffic is modeled as Poisson, with the destinations chosen among its nearest neighbors for simplifying the simulation model. Each simulation is executed for 5 minutes of simulation time and each speech stream has 13000 frames with a 20 ms inter-frame time. All IR systems in the following simulations include the packet reconstruction function.

In this section, IR is first compared with the retransmission-based scheme (both restricted to one retransmission and the standard). Then, using path diversity, MD coding with two paths and single path transmission without MD coding are compared. Finally, IR, path diversity and MD coding are integrated into a single system for performance evaluation. End-to-end performance measures are considered, including segment loss rate, delay, and burst size for lost segments (number of consecutive segments lost).

6.1 IR vs. Retransmission-based MAC

To compare the proposed IR scheme with a conventional system fairly, 802.11 MAC's retransmission limit is first restricted to 1, and then restored to 7, which is specified in the standard for short packets. In Glomosim, a priority queue with size 100 packets is implemented at the IP layer of each node, providing control packets, real-time data, and non real-time data different priorities to get into MAC layer. Both real-time and non real-time background traffic are simulated to examine the impact of their priorities. The size of non-real-time background traffic packets is 1500 bytes and the size of real-time background traffic packets is 20 bytes. The segment size of the tagged voice stream is 20 bytes, excluding all headers. All short packets, both background and tagged voice packets, in one simulation use the same type of MAC, namely, IR or retransmission-based MAC. Long packets, like the non-real-time background packets, use retransmission-based MAC. For the IR scheme, a segment is considered lost if all redundant copies, both



(a) **(b)** **(c)**

Fig. 8: (a) Segment loss rate, (b)average end-to-end delay, and (c) segment loss rate under 200 ms delay constraint for non real-time background traffic

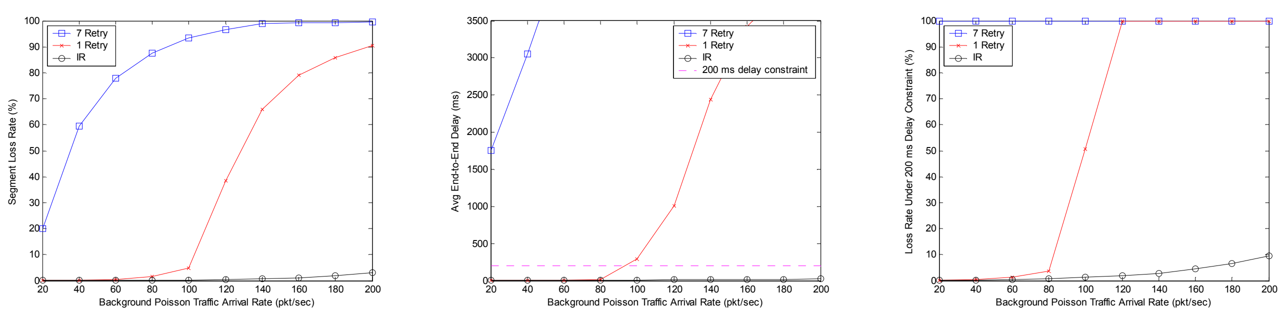
original and reconstructed, of that segment are lost. The end-to-end delay is calculated as the difference between the arrival time of the first successful copy and the transmit time of the first copy. For a desired 200 ms delay constraint, the loss rate is also calculatedfor which all packets with end-to-end delay larger than 200 ms are considered lost.

Fig. 8(a) shows the segment loss rate with non-real-time background traffic. Results of the one-retransmission (1-retry), the IR, and the standard (7-retry) systems are shown. It is found that the segment loss rate of the 1-retry system grows much faster than the IR system, especially when the background traffic arrival rate ranges from 30 to 50 packets per second. This improvement of the IR system over the 1-retry system comes from the different behavior of the two systems under heavy traffic. When transmission opportunities become rarer due to the higher load, the IR system still transmits a new segment together with an old segment at every opportunity. In contrast, the 1-retry scheme usually needs two opportunities to send each segment, increasing the traffic in the network, thus increasing both delay (in waiting for transmission opportunities) and loss rate (due to higher traffic loads). On the other hand, since the 7-retry system has seven chances to recover from transmission failures, the segment loss rate is zero when load is light. But its performance degrades after the background traffic arrival rate exceeds 40 packets per second. At high traffic load, IR performs better than the two retransmission based systems.

Fig. 8(b) shows the average end-to-end delay under non-real-time background traffic. When the background traffic arrival rate is greater than 40 packets per second, the average end-to-end delay of retransmission-based systems grows fast and exceeds the desired 200 ms delay, while the average end-to-end delay of the IR system still remains less than the delay constraint until 80 packets per second. This is not only due to the effective reduction of transmission time in the IR scheme, but also its inherently stable delay characteristic. Note that the average end-to-end delay of the 7-retry system is longer than the 1-retry system when the traffic load is heavy due to too many retransmissions.

Fig. 8(c) displays the segment loss rate for a 200 ms delay constraint. The figure shows the actual loss rate that the receiver end-user sees because voice packets violating the delay constraint are discarded at the receiver. According to Fig. 8(a), the segment loss rate of the 7-retry system is less than 35% when the background traffic arrival rate is 100 packets per second, but it becomes 94.42% when delay constraint is applied. In other words, almost all packets in the 7-retry system are too delayed to be useful when they arrive at the receiver. On the other hand, having a delay constraint does not affect the IR system so much.

Fig. 9(a) shows the segment loss rate for different arrival rates of real-time background traffic. Here too, the performance of the IR scheme has significant gains over the retransmission based systems. In this figure, it is shown that having more retransmissions significantly degrades the system performance. And the difference between the performances here is more significant than Fig.8(a). The key factor of this effect is the priority of background traffic packets. Since the IP queue size is not infinite, packets will be dropped when the queue gets full. And since the voice packets have priority over the non-real-time



(a) **(b)** **(c)**

Fig. 9: (a) Segment loss rate, (b)average end-to-end delay, and (c) segment loss rate under 200 ms delay constraint for real-time background traffic

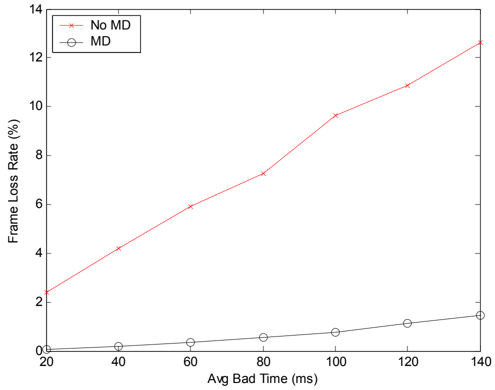


Fig. 10: Frame loss rate for MD and non-MD systems

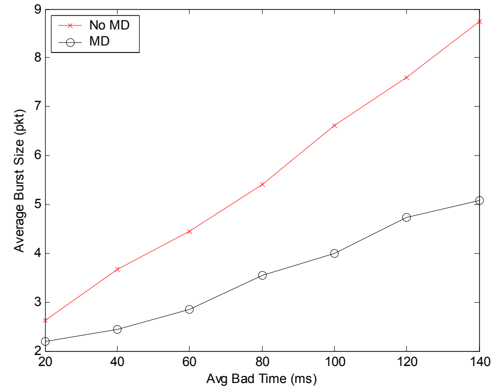


Fig. 11: Average burst size for MD and non-MD systems

background packets in Figs. 8(a), 8(b), and 8(c), the low priority background packets will be dropped first while the voice packets are not likely to be dropped due to IP buffer overflow, even when the packet arrival rate is higher than the system service capability. However, if the background traffic has the same priority as the tagged voice packets, every packet is equally likely to be dropped. Thus, the segment loss rates here are much higher than in Fig. 8(a). Since the IR system has much better bandwidth efficiency, the system can deal with heavier background traffic and does not suffer from IP buffer overflow. The real-time background traffic can be viewed as a coarse model for a system with many real-time connections (not simulated because of the complexity of tracking so many connections), even though only the performance seen by the tagged speech connection is monitored. With this interpretation, Fig. 9(a) also implies that IR can significantly increase the network capacity by allowing more connections in the network for a given segment loss rate.

Fig. 9(b) displays the average end-to-end delay while using short real-time packets as background traffic. While the average end-to-end delay of retransmission-based systems start to increase steeply, the inter-packet redundancy system keeps its end-to-end delay small because it is not affected by IP buffer overflow. Fig. 9(c) shows the segment loss rate after applying the delay constraint. The loss rate of retransmission-based systems becomes extremely large once the postive feedback described in Section 3.4 becomes significant.

6.2 Two Paths with MD vs.Single Path without MD

Here, simulation results are used to show the gains achieved via path diversity. A Gilbert-Elliott model is added into each path in order to model bursty losses, without incurring the complexity of modeling specific causes of burst loss such as fading and mobility. When the state of a path is bad, all the packets are lost; when the state is good, the packets can still be lost due to collision or noise. The average dwell time in the good state is set to 1000 ms. The background traffic is comprised of 20-byte real-time packets with an arrival rate of 30 packets per second. The average time in the bad state ("bad time") is varied from 20 to 140 ms to simulate the impact of channel burstiness. As in Section 5, the frame size for each description in MD coding is 17 bytes and the payload size becomes 23 bytes after adding the sequence number and timestamp to each frame. The payload size for a non-MD coding system is 38 bytes. The two MD bit streams are transmitted through different paths, while the non-MD coded system only uses one path. Both systems adopt the one-retransmission 802.11 MAC.

Fig. 10 shows the frame loss rate of both the MD and non-MD systems. For the MD system, a frame loss occurs only when both descriptions are lost. The frame loss rate of the single path system is 11.08 times greater than the two path MD system when the average bad time is 20 ms, and is 8.57 times greater when the average bad time is 140 ms. This performance measure does not account for the fact that, when one of the two MD streams is received, the speech quality is not as good as when a single description code is reliably received. This is consistent with the objective of providing acceptable quality in a difficult environment.

Fig. 11 shows the average size of bursts of end-to-end lost segments for the two systems. Only successive losses are considered as contributing to a burst, so that the smallest possible burst size is 2. When the channels are not bursty (average bad state dwell time of 20 ms), random loss is the dominant source of loss, hence both systems have average burst sizes slightly larger than 2. But when the channels get bursty, the average burst size of the two path system is approaching half the average burst size of the single path system. The average end-to-end delay is not shown here, since the background traffic load, which is the dominant factor of end-to-end delay, is not changed.

6.3 Integrated System vs. Single Path + 1-Retry

The performance of the integrated system, which combines inter-packet redundancy, path diversity and multiple description coding, is evaluated in this subsection. The integrated system is only compared with the single path system with one-retransmission MAC since the one-retransmission MAC actually performs better than the standard seven-retransmission system as shown in Section 6.1. Here, the average bad time is fixed to 100 ms, and the arrival rate of the real-time background traffic is varied. The packet size of the background traffic is still 20 bytes.

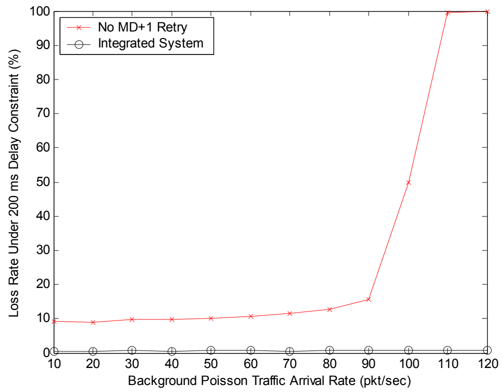


Fig. 12: Segment loss rate for the integrated system and retransmission-based system under 200 ms delay constraint

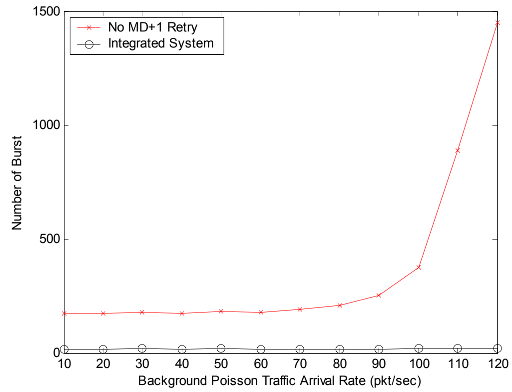


Fig 13: Number of error bursts for the integrated system and retransmission-based system

Fig. 12 shows the frame loss rate after applying the delay constraint. The integrated system has a consistently low frame loss rate (lower than 0.8%). On the other hand, the loss rate for the single path system ranges from 8 ~ 14% as the background traffic arrival rate grows from 10 to 80 packets per second. The loss rate grows dramatically once the background traffic arrival rate becomes greater than 90 packets per second. Fig. 13 shows the number of end-to-end loss bursts during the whole session for these two systems. The burst number of the retransmission-based system remains around 10 times the burst number of the integrated system before the background traffic arrival rate reaches 100 packets per second. Fig. 12 and Fig. 13 show the proposed integrated system employing IR is superior both in terms of loss bursts and loss rates.

7. CONCLUSION

The conventional wisdom regarding packetized systems is that one would not employ inter-packet coding above the physical layer, unless there is a delay constraint. This is because retransmission-based recovery is always considered superior to adding redundancy on an erasures channel. In this paper, we show that this is simply not true for packetized systems in which the payload is small compared to the overhead, as is the case for the real-time speech application considered here. The large overhead makes both retransmissions and ACKs expensive in terms of the additional network load generated. In this case, proactive use of inter-packet redundancy provides large coding gains while incurring negligible bandwidth penalties, thus providing better performance both in terms of capacity and delay compared to retransmission-based error recovery. Our integrated system includes the additional mechanisms of path diversity and MD coding. Path diversity alleviates bursty losses, and is particularly useful for delay-sensitive applications in which data cannot be interleaved with a large enough delay to convert burst losses into random loss. MD source coding is a natural means of exploiting path diversity, by sending equally important streams in parallel over multiple paths. Overall, the simulated performance shows that it is indeed possible to support speech with reasonable end-to-end quality over multihop wireless networks with both random and bursty loss, and with both real-time and non real-time traffic. Note that we do not require complex bandwidth estimation and reservation methods, and require only that individual node schedulers give priority to real-time traffic over non real-time traffic.

An important topic for future work is to demonstrate working systems that incorporate the strategies presented in this paper. It is also important to characterize the practical capacity regions for ad hoc networks with both real-time and non real-time traffic under such optimized strategies. Theoretical rules of thumb would be very useful in this setting, since purely simulation-based approaches suffer from problems of scaling to a large number of connections (since the bandwidth required by an individual speech connection is much smaller than the link bandwidth in a typical 802.11 type network, it is difficult to simulate a large enough number of connections to determine the maximum number the network can support). The effect of non real-time traffic on real-time QoS is particularly important to understand: even when node schedulers give priority to real-time traffic, non real-time traffic from other nodes, including hidden nodes, can cause interference.

8. ACKNOWLEDGMENTS

The authors would like to thank Y. C. Chen for helpful discussions, Dr. E. M. Belding-Royer for her valuable suggestions and Dr. J. D. Gibson for his careful review of the paper.

9. REFERENCES

[1] S. Corson and J. Macker, "Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation consideration," RFC 2501, Jan. 1999.

[2] E.M. Royer, C.-K. Toh,"A review of current routing protocols for ad hoc mobile wireless networks," *IEEE Personal Communications*, vol. 6, pp. 46 -55, April 1999

[3] IEEE standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, ISO/IEC 8802-11:1999(E), Aug. 1999.

[4] http://www.ieee802.org/11.

[5] M.G. Arranz, R. Aguero, L. Murioz, P. Mahonen, "Behavior of UDP-Based Application over IEEE 802.11 Wireless Networks," *12th IEEE International Symposium on Personal,*

Indoor and Mobile Radio Communications, vol. 2 , pp. F-72 -F-77, 30 Sept.-3 Oct. 2001.

[6] L. Munoz, M. Garcia, J. Choque, R. Aguero, P. Mahonen, "Optimizing Internet flows over IEEE 802.11b wireless local area networks: a performance-enhancing proxy based on forward error correction," *IEEE Communications Magazine*, vol. 39, pp. 60 -67, Dec. 2001.

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

[8] V. N. Muthiah and W. C. Wong, "A speech-optimised multiple access scheme for a mobile ad hoc network," *1st Annual Workshop on Mobile and Ad Hoc Networking and Computing*, pp. 127–128, Aug. 2000.

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

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

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

[12] S. Aramvith, Chia-Wen Lin, S. Roy, and Ming-Ting Sun, "Wireless Video Transport Using Conditional Retransmission and Low-Delay Interleaving," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, No. 6, Jun 2002.

[13] H. Wu, C. Hung, M. Gerla, and R. Bagrodia, "Speech support in wireless, multihop networks," *3rd International Symposium on Paprallel Architectures, Algorithms, and Networks Proc.*, pp. 282–288, Dec. 1997.

[14] S.-Y.R. Li, R.W. Yeung, C. Ning, " Linear network coding," *IEEE Trans. on Information Theory*, vol. 49, pp. 371 -381, Feb. 2003.

[15] R. Koetter, M. Medard, " An algebraic approach to network coding,"*IEEE/ACM Trans. on Networking*, vol. 11, Oct. 2003.

[16] N. F. Maxemchuck, "Dispersity routing in store and forward networks," *Ph.D. dissertation, Univ. Pennsylvania, Philadelphia*, May 1975.

[17] T. T. Lee and S. C. Liew, "Parallel communications for ATM network control and management," *Proc. GLOBECOM'93*, Nov. 1993, pp. 442–446.

[18] N. T. Plotkin and P. P. Varaiya, "Performance analysis of parallel atm connections for gigabit speed applications," *Proc. INFOCOM'93*, pp. 1186–1193.

[19] E. Ayanoglu, I. Chih-Lin, R. Gitlin, and J. Mazo, "Diversity coding for self-healing and fault tolerant communication networks," *IEEE Trans. on Communications*, vol. COM-41, pp. 1677–1688, Nov. 1993.

[20] R. Krishnan and J. A. Silvester, "Choice of allocation granularity in multipath source routing schemes," *Proc. INFOCOM'93*, Mar. 1993, pp. 322–329.

[21] Mahesh K. Marina, Samir R. Das, "On-demand multipath distance vector routing in ad hoc networks," *9th International Conference on Network Protocols*, 11-14 Nov. 2001

[22] Panagiotis Papadimitratos, Zygmunt J. Haas, Emin Gun Sirer, "Path set selection in mobile ad hoc networks," *Proc. 3rd ACM international symposium on Mobile ad hoc networking & computing*, Jun. 2002

[23] V. A. Vaishampayan, "Design of multiple description scalar quantizer," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.

[24] Y. Wang, M. Orchard, V. Vaishampayan, and A. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Processing*, vol. 10, pp. 351–366, Mar. 2001.

[25] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2199–2224, Sept. 2001.

[26] H. Jafarkhani and V. Tarokh, "Multiple description trellis coded quantization," *IEEE Trans. Communications*, vol. 47, pp. 799–803, June 1999.

[27] D. Chung and Y.Wang, "Multiple description image coding using signal decomposition and reconstruction based on lapped orthogonal transforms," I*EEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 895–908, Sept. 1999.

[28] S. D. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *Proc. ICIP'98*, pp. 659–663.

[29] J. G. Apostolopoulos, T. Wong, W. Tan, S. Wee, "On Multiple Description Streaming with Content Delivery Networks," *IEEE INFOCOM*, June 2002

[30] Nitin Gogate, Doo-Man Chung, Shivendra S. Panwar, Yao Wang, "Supporting image and video applications in a multihop radio environment using path eiversity and multiple description coding," *IEEE Jounral on Selected Areas in Communications*, vol. 12, pp. 777 -792, Sep 2002.

[31] A. Miu, J. G. Apostolopoulos, W. Tan, M. Trott, "Low-Latency Wireless Video Over 802.11 Networks Using Path Diversity," *Proc. of the IEEE International Conference on Multimedia and Expo(ICME)*, Baltimore, MD, July, 2003

[32] H. Dong, A. Gersho, J. Gibson, and V. Cuperman, "A Multiple Description Speech Coder Based on AMR-WB for Mobile Ad Hoc Networks," *IEEE International Conference of Acoustics, Speech, and Signal Processing, Montreal*, Canada, May 17-21, 2004

[33] X. Zeng, R. Bagrodia, M. Gerla, "GloMoSim: a library for parallel simulation of large-scale wireless networks," *Proc. PADS 98*, 26-29 May 1998, pp. 154–161