Efficient Methods for Broadcasting Multi-Slot Messages with Random Access with Capture

Amanda Peters, Linda Zeger
Harvard University, MIT Lincoln Laboratory

Abstract—**A new technique for best effort delivery of broadcast bursty multi-slot messages is investigated, using random access with capture. It is shown that a combination of packet level coding and random spreading of a message's packets can improve message reception probability, satisfaction of delay constraints, and throughput in many cases. Furthermore, we show how the specific transmission strategy employed should depend on the message length, total offered traffic, and the specific delay constraints. In addition, we show how edge effects, the finite size of the network, and receiver capture significantly impact performance not only at the edges of the network, but also near the center of the network. Finally, we demonstrate how to leverage these edge effects so that the** *introduction of minimal feedback from only one or two nodes* **can substantially improve the probability of message reception for all nodes in the network.**

I. Introduction

The increasing demand for transmission of larger files such as images can result in the need to transmit bursty traffic, which can result in a file or message extending beyond a single physical layer frame or time slot, in wireless networks. Such bursty traffic is not well suited to transmission schemes with fixed channel assignments such as TDMA. Random access protocols are good candidates for bursty traffic when there are many bursty users, each with an average low traffic rate. While protocols that reserve the channel such as CSMA/CA or DBTMA [1] are beneficial for point-to-point transmissions, such protocols were not designed to handle delivery of a message to multiple destinations.

ALOHA type protocols could be used to handle bursty broadcast messages, but they generally suffer from collisions and low throughput, as well as delay at high traffic loads. Furthermore there has been relatively little study on the use of ALOHA for multi-slot messages, particularly when they are broadcast. Transmission of point-to-point multi-slot messages over multiple ALOHA channels was accomplished in [2] and [3] with an erasure code covering the multiple slots. In [4] a protocol for transmission of point-to-point multi-slot messages on a single channel was introduced.

In this work, we handle best effort delivery of *point-to-multipoint* multi-slot messages, and investigate the case of an ALOHA type protocol with capture. These conditions are applicable to Link 16, in which the nearly perfect capture model used here serves as a close approximation to that actually implemented. Furthermore, in Link 16 nodes have knowledge of each others' locations; we leverage this type of knowledge in our efficient feedback protocol at the end of this paper.

The difficulty with multi-slot messages is that for a number of applications, if any single packet or slot is lost, the entire message is considered lost, which proves challenging when ALOHA is used due to the large number of collisions. Therefore, we turn to packet level erasure coding to add redundancy against collisions, as in [4]. Use of coding with random access was also considered in small and moderate sized networks to maximize the throughput in [5] when all nodes are backlogged with packets. In contrast, we consider stochastic arrivals of packets at each node and moderate sized and large networks.

Furthermore, we consider the challenging case of messages that are designed for all nodes in the network. For reliable transmissions, the issue with conventional methods of broadcast messages is that the amount of feedback typically grows with network size. Thus we consider *best effort* delivery, where initially we use no feedback, so that time sensitive newly arriving messages can be delivered without undue delay caused by retransmissions of old messages and there is *no* burden of feedback from the many destination nodes. We lastly show how to leverage the capture effect, along with our coding and access strategies, in order to substantially improve performance for best effort traffic with use of only minimal feedback. Therefore we introduce a very small amount of feedback for *best effort broadcast traffic*, in order to improve reliability by eliminating transmission of coded packets when all nodes have received a message.

If the entire coded message were to be transmitted in consecutive slots, little diversity would be gained by the coding. Therefore, we use packet level erasure coding, along with careful selection of the retransmission probability, to provide favorable tradeoffs between throughput and delay constraints. Our random spreading in time of coded packets introduces a *capture diversity*, that results from different coded packets of a single message receiving very different levels of interference from different interfering nodes. We demonstrate the spatial dependence of performance that results from the capture effect, and we utilize this dependence to further improve performance in the broadcast transmissions through introduction of only a small amount of feedback so as to avoid the conventional "ACK explosion" problem from many recipients.

This work is sponsored by the Department of Defense under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, recommendations, and conclusions are those of the authors and are not necessarily endorsed by the United States Government. Specifically, this work was supported by: 640th Electronics Systems Squadron of the Electronic Systems Center, Air Force Materiel Command, and Information Systems of ASD(R&E).

Amanda Peters is with Harvard University. Linda Zeger is with MIT Lincoln Laboratory.

In Section II we described the network and communication model used. Analysis of this model for no spreading of a message in time and for extreme spreading is presented in Section III. Intermediate levels of spreading are investigated with simulation in Section IV. In Section IV-A we discuss aggregate performance characteristics of the network. In Section IV-B we consider the spatial variance of performance caused by the capture effect, and show the substantial benefit of adding a limited amount of feedback. A summary is provided in Section V.

II. Network and Protocol Model

We investigate the transmission of multi-slot messages when a slotted ALOHA protocol is used with capture. The messages originate at random times from each node within a network and are destined to all other nodes in the network. All nodes have the same probability of message generation, and in any given time slot multiple nodes may be transmitting. In order to protect messages against individual packet collisions, we use erasure coding at the packet level and a random spreading of the k packets, as in [4]. Thus each original k packet message is transmitted by sending n coded packets. We assume the coding is such that if and only if any k or more of the coded packets are received by a node, the message is successfully decoded. We further assume a node is able to decode a message as soon as it receives k of the n coded packets, and that all nodes are half-duplex.

Upon generation at a node, the first coded packet of a message is immediately transmitted, if there are no coded packets from preceding messages in the buffer awaiting transmission. If there are coded packets awaiting transmission in the buffer, the new message is placed at the end of the queue. Each of the remaining coded packets, after the first coded packet is transmitted, is transmitted successively in any given time slot with probability Pt. If $Pt = 1$, then the protocol is similar to unslotted ALOHA with a packet length of n, and the coded packets offer little diversity, since it is likely that multiple coded packets will be lost in any collision. The same is true for large values of Pt close to but less than 1. If Pt is too small, it will take a long time for a node to completely transmit a message.

We consider collisions as the only channel impediment, so that in the absence of concurrent transmission, every node can successfully receive from every other node. We primarily consider an ideal capture model in which the node will receive the packet from its nearest transmitting neighbor. In the event that two transmitting nodes are at equal distances from the receiver and that this distance is the shortest distance between the receiver and any transmitting node, both packets will be lost. If a node is transmitting, it is not able to receive any other messages. Hence packet loss occurs only due to the half-duplex constraint, and when transmitters equally distant from a receiving node transmit simultaneously.

Because in this work we consider broadcast transmissions in a network with a total of N nodes and the capture effect, if T nodes each transmit a packet simultaneously in a single time slot, some of the $N - T$ non-transmitting nodes may capture a packet from one transmitting node, while other non-transmitting nodes may instead capture a different packet from a closer transmitting node in that same time slot.

A line of N equally spaced nodes, as well as a two-dimensional grid of nodes placed on the points of a square lattice, are considered as examples in much of the paper. Since many of the results actually depend only on relative orderings of nodes, rather than on their precise locations, these results also provide insights to many additional node configurations as well. For example, we have considered randomly removing various fractions of nodes from a network that consists of nodes on the 2D grid of a square lattice; for most cases, the performance characteristics were the same as for the fully occupied lattice of nodes. We have considered networks of size ranging from 10 to 400 nodes.

As N is increased from 10 to 20 to 100, significant improvement in throughput is seen at each increase in N, due to the fact that when there are more nodes, the transmit while receive collisions, as well as collisions from equidistant transmitters, become a smaller fraction of the total transmissions. As N is increased beyond a few hundred nodes, little further improvement is seen.

A message buffer of four messages per node is employed. The message buffer queues messages that arrive while another message has already begun transmitting. This buffer size was sufficient to avoid all queue drops for the per node offered traffic levels considered here.

III. Analysis for Extremes in Spreading Packets in Time

In this section we consider the two extreme limiting cases of the range of allocation strategies when no bandwidth is allocated to feedback. First, we consider transmitting coded packets spread enough in time so that loss of one coded packet in a message is independent of loss of the other coded packets of the message. Next, we consider transmission of *uncoded* multi-slot messages with no spreading in time.

A. Independent Coded Packets

If all of the coded packets were independently subject to loss from collisions, then the per node average received throughput from all transmitting nodes would be given by

$$\frac{2}{N}\frac{k}{n}g_c\sum_{f=1}^{g-1}\sum_{g=2}^{N}\sum_{j=k}^{n}\binom{n}{j}\exp(-jN_ig_c)\times(1-\exp(-N_ig_c))^{n-j}, \tag{1}$$

where g_c is the per node offered coded traffic, and the number of interfering nodes N_i that each transmission is subject to is

$$N_i = \min(2\times(g-f), N-f). \tag{2}$$

Equation (1) is obtained from considering all pairs of nodes in the line, and assumes that the traffic from N_i nodes is Poisson distributed. This assumption matches the model used here when Pt is small, as seen in Figure 2 in Section IV, which is precisely when we expect the losses of coded packets from a single message to be the least correlated.

We explored a range of k and n values for large N for the case when coded packets are independently lost, using (1) and (2). For different sized messages of length k slots, Table I summarizes the optimal coded message size n for $N = 200$. In addition, the highest achievable total throughput S is given, along with the corresponding value of total offered traffic G at which this maximum throughput is achieved. When

TABLE I
OPTIMAL CODING AND MAXIMAL THROUGHPUT FOR LARGE NETWORKS WITH INDEPENDENT PACKETS IN EACH MESSAGE.

k	Best n	Highest S	G for Highest S
1	1	.97	5.5
2	3	.54	1.6
3	6	.47	1.4
4	9	.44	1.1

$k = 1$, due to the nearly perfect capture, it is beneficial to have the high values of offered traffic. For larger k values, whereas transmission of more redundant packets (larger n), is beneficial at low values of traffic, when the traffic becomes higher, the additional coded packets cause more collisions, rendering moderate values of n more favorable. The lower values of n do not offer enough protection against loss of the whole message if a few coded packets are lost. Furthermore, the larger the message size, the lower the throughput due to the finite size block length and the loss of use of all coded packets when fewer than k are received. On the other hand, making the block length too long or using a rateless code would mean that most of the receivers, that is those that have already received k coded packets, would be obtaining no useful throughput during many of the redundant coded packet transmissions.

Hereafter, we consider $k = 3$ slot messages, and we use the optimal value of $n = 6$ found for $k = 3$.

B. Transmissions in Bursts

We now consider the case of uncoded packets where each node transmits the entire multi-slot message upon its arrival at the node. In this case, there are no time slots allocated to coding. Hence, with the capture effect for a k slot message, the probability of loss at any given node that has N_i interfering nodes closer to it than the transmitting node is the probability that none of the N_i interfering nodes transmit in any of the $2k-1$ time slots that would cause an overlap of the messages:

$$Ploss \approx 1 - (1 - N_i \times G/N)^{2k-1}. \tag{3}$$

In particular, in Figure 1, we plot the message loss probability for $k = 3$, $N = 21$, when we consider the transmitting node at the edge of a line of nodes. We have also plotted simulation results for comparison. Because (3) calculates the loss resulting from collisions involving only two messages, it slightly underestimates the total probability of loss, as shown in Figure 1. In contrast, the simulated results include collisions of multiple messages.

IV. Simulations

We now investigate the cases intermediate in terms of spreading of packets to those studied in Section III.

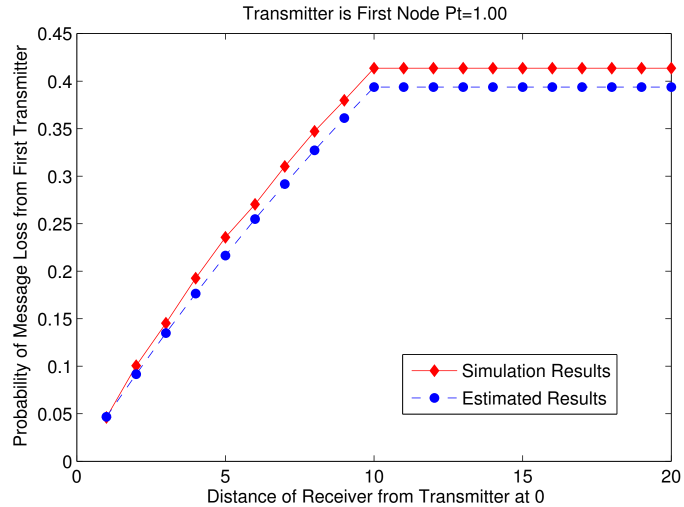

Fig. 1. Probability of loss for uncoded transmissions. Estimated results include collisions between only two nodes at a time, whereas simulated results include collisions between any number of nodes.

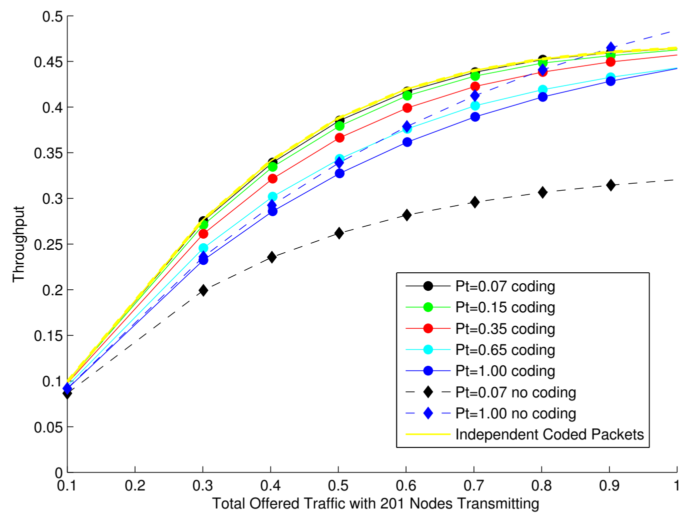

Fig. 2. Throughput received at a node from all other nodes vs. offered information traffic for 201 nodes in a linear network.

A. Aggregate Performance

Figure 2 plots the throughput received at a node as a function of the total offered information traffic in a network with 200 nodes forming a line. The results of Figure 2 are similar to those obtained with 400 nodes on a two-dimensional grid. The per node received throughput is averaged over all N receiving nodes. The total offered information traffic is the product of the average number of coded packets transmitted by all N users in a time slot and k/n. The dashed curves represent no coding, whereas the solid curves represent the use of coding. It is seen that except at the highest offered traffic levels, the largest throughput is achieved by using coding and the lowest values of Pt. The performance of $Pt = .07$ is the same as that of independent coded packets derived in Section III-A and indicated in this figure by the yellow curve; the packets are sufficiently spread in time here that if two coded packets from different messages collide, then the successive packets

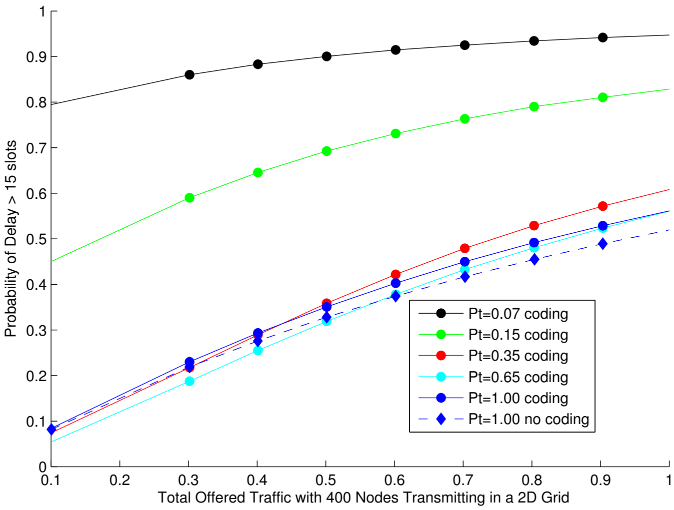

Fig. 3. Probability that the delay exceeds a threshold of 15 time slots, for a range of values of offered traffic.

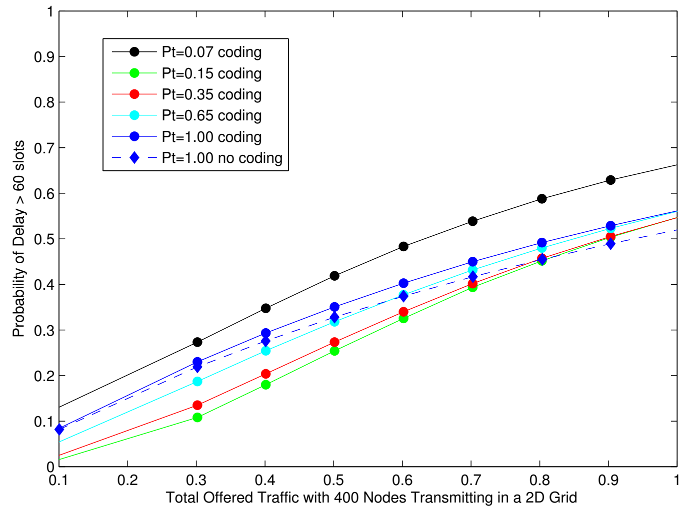

Fig. 4. Probability that the delay exceeds a threshold of 60 time slots, for a range of values of offered traffic.

from each of these messages are no more likely to collide with each other than they are with packets from *different* messages. We dub this effect as "capture diversity" since at any given receiver, the different coded packets of a message will have independent attempts at being captured. When coding is used, as Pt is increased towards 1.0, it is seen in both these figures that the throughput decreases, because if packets from two messages collide, then their successive packets become more likely to also collide, thereby limiting the capture diversity from coding.

On the other hand, it is seen from Figure 2 that the worst performance is obtained from use of $Pt = .07$ and no coding. For uncoded messages, since once a single packet of the message is lost, the entire message is lost, it is more favorable to have multiple packets of two messages collide with each other, rather than having additional coded packets of these already lost messages collide with other additional messages that could potentially still be successfully received.

Finally, at very high traffic levels, it is seen from Figure 2 that the policy of no coding and no spreading ($Pt = 1.0$) slightly outperforms all other coding and spreading strategies. At these high traffic levels, the additional interference caused by coding can not compensate for the capture diversity gain provided. We note that we can decrease the load of extraneous coded packets to further improve the spreading and coding strategies shown here, by adding a small amount of feedback; we illustrate this idea in Section IV-B.

We next consider delay, and in particular we investigate what fraction of the nodes can satisfy specific delay constraints. We define delay as the time from when the first packet of a message is transmitted to the time when a receiving node has received any k of the transmitted packets of that message. First, we discuss a delay constraint threshold of 15 time slots. Figure 3 plots the cumulative distribution function (CDF) at 15 time slots. By comparing the different curves, it is seen that the probability of not satisfying this delay constraint increases as Pt decreases for $Pt \leq .35$. With all of these Pt values, the messages are more likely delayed than lost. For larger values of Pt, further increasing Pt does not necessarily increase the probability of satisfying this delay constraint threshold; while increasing Pt reduces the transmission time between coded packets, at the large Pt values it also increases the message loss probability, and hence increases the probability of infinite delay.

For larger values of a delay constraint, the effect of message loss on delay is even more pronounced relative to the random spreading time between coded packets. In Figure 4, the CDF is plotted at a delay constraint threshold of 60 slots, and it is seen to closely resemble the probability of message loss graph, which is plotted in Figure 5. The probability of the delay exceeding 60 time slots, for all values of Pt explored, except the smallest one of $Pt = .07$, is thus shown to be dominated by lost packets. For Pt values between $.15$ and 1.0 the expected number of slots, in the absence of collisions, to send a full message is well below 60. In these cases, a message not being received after 60 slots is most likely lost. In contrast, for $Pt = 0.07$, a larger number of slots is required transmit the message due to greater spreading. In this case, after 60 slots it is still reasonably likely that the transmitting node has not finished sending all of the packets, causing the packet spreading to be a primary cause of the large delays.

Figures 3 and 4 and the above discussion show that the transmission strategy should depend on the specific delay constraint. Shorter delay constraints can be satisfied by more nodes if a larger Pt is used, whereas longer delay constraints can best be satisfied with a lower value of Pt. Furthermore, in both of these figures, at larger values of traffic, the delay constraints are best satisfied by using no coding and no spreading. While the addition of a small amount of feedback, such as that discussed in Section IV-B, may further increase the traffic level at which no coding and no spreading becomes optimal, if no feedback were used, then the traffic level should also be factored into the choice of transmission strategy.

We give an example comparing our method here to send

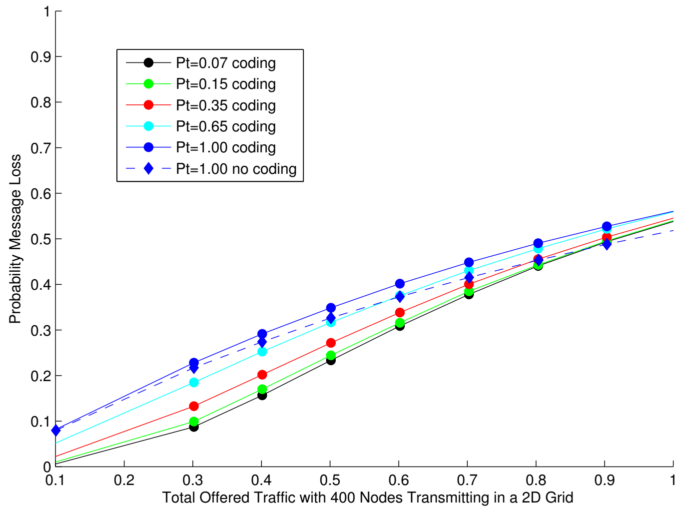

Fig. 5. Probability of message loss vs. offered information traffic.

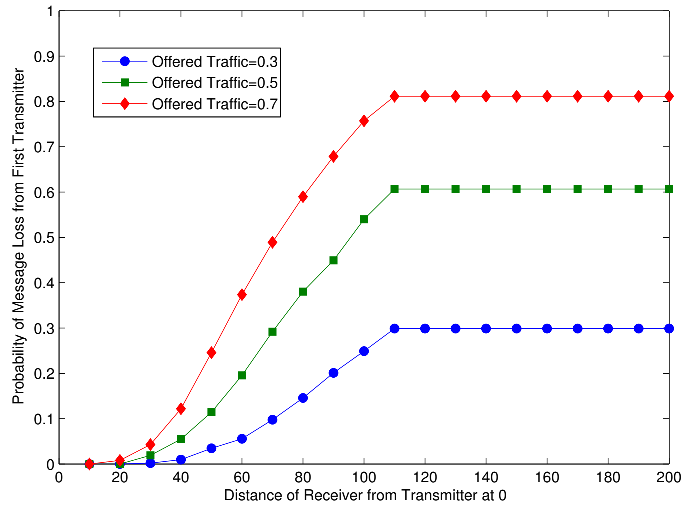

Fig. 6. Message loss vs. distance from the transmitting node for a total offered information traffic of .3.

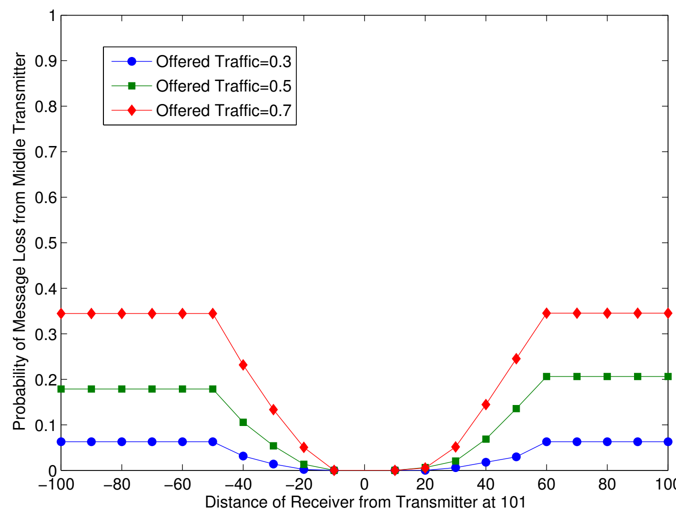

Fig. 7. Percent of Message Loss vs. Distance from Transmitter at total offered information traffic of .3.

multi-slot messages with random access with capture to use of TDMA. For $Pt = .15$, the *mean* number of slots it takes for messages that are not lost to be received ranges from 17 to 23, as the offered information traffic is increased from .1 to .8. In contrast, if instead of random access with our protocol of randomly spacing coded packets, the 400 nodes were to use TDMA, then the *mean* number of slots it would take, using the same definition of delay, which is the number of time slots for successful message reception after the first packet of the message is transmitted, would be 800 slots. While no messages would be lost with TDMA, the delay in many cases would be unacceptable.

As the offered traffic in the system increases, the probability of message loss also increases. Furthermore, larger values of Pt lead to more message loss when coding is used since larger values of Pt mean a higher likelihood of multiple coded packets from the same two messages colliding as they are more clumped in time, thereby negating the diversity benefit of coding. The dashed lines represent simulations in which n=3, which means that there is no coding used.

B. Spatial Dependence of Performance

We now explore the spatial dependence of the probability of message loss. The capture effect makes it more likely that nodes close to a transmitter will receive a message than nodes farther away. In order to illustrate this spatial dependence, we consider a 200 node linear network.

Figure 6 shows the probability that a message transmitted by the left most node node will be lost by each of the receiving nodes for several values of total offered information traffic, when $Pt = .07$. As shown in the figure, there is a saturation point at node 100 beyond which all future nodes are impacted by the same amount of interference. Loss of coded packets occurs when another node closer to the receiver transmits in the same time slot as another node sending a message, so receivers farther away from the transmitter will be subject to more interference, and therefore will experience a higher likelihood of message loss. The saturation occurs due to edge effects, because the nodes in the saturated region are subject to potential interference from every other node in the network.

The spatial dependence of probability of message loss is seen to result from the combination of capture and edge effects. Similar effects shown in [6] in the spatial dependence of offered traffic for single-slot messages, when the offered traffic was chosen to maintain a fixed level throughput.

A similar profile is seen if mean delay is plotted as a function of distance from the transmitting node. More traffic leads to a higher mean delay, as expected. Mean delay increases rapidly with distance from the transmitting node until the plateau point is reached at which point the mean delay stays constant at longer distances due to edge effects.

We now consider packet receptions from transmissions made by the middle node. Figure 7 shows the probability of losing a message that is sent from the central transmitter. This figure shows the expected mirrored loss as the distance from the transmitting node is increased in either direction.

We now propose utilizing the spatial dependence of packet loss that is induced by the combination of the capture effect and edge effects. Since in the model used here, with capture and collisions being the only source of packet loss, it is seen from Figure 6 that if any of the nodes between nodes 100 and 200 received the message, then they all received it. In this case, the reception by the plateau point node, which is node 100 for example when the transmitting node is the left most node, *deterministically* implies that the nodes on the other side of the plateau point have received the packet with certainty.

Therefore, we implement the following feedback protocol when coding and random spacing in time is used: When the receiver at the designated plateau point for each transmitter receives the message (that is any k coded packets), it sends an acknowledgement to the transmitting node. The plateau point identifies the closest receiving node that will be impacted by every other potential interfering node. When this node has received the message, it can be assumed that every other node has as well. By enabling this feedback, we are able to reduce the overall traffic in the system and therefore reduce collisions and loss, as well as delay, of future messages.

We now demonstrate how to determine the plateau point node for an transmitting node. For example, if a transmitting node located at $t < N/2$ nodes from the left most transmitter, where N is the total number of nodes in a linear network, the two potential plateau points would be located at transmitters P_1 and P_2:

$$P_1 = \lceil(\frac{N-t}{2})\rceil + t \quad (4)$$

$$P_2 = \lfloor(t/2)\rfloor \quad (5)$$

In this case N is the number of nodes in the system and P is the plateau point index. This figure also shows that as the offered traffic in the system is increased, the probability of message loss increases. Note: when the middle node is the transmitter there are *two* pivots, defined by each of the equations above, that both must acknowledge message receipt for feedback to be successful.

Figure 8 shows the resulting decrease in message loss from this protocol for $N = 21, k = 3, n = 6$, and Pt = 0.07, for an offered traffic of .3. The upper curve represents no feedback when all n packets are transmitted, whereas the lower curve shows the performance with the feedback protocol. The difference between the two curves shows the significant reduction in loss provided by the feedback protocol. The savings from the feedback is greatest where the loss is greatest, which is beyond the plateau point. Here, the message loss is reduced by more than a factor of 2.

Again, for the half of the nodes beyond the plateau point, the message loss is reduced by a factor greater than 2. The absolute value of the decrease in probability of message loss is only a few percent here, whereas it was 15% when the edge node transmitted, because the absolute value of the original message loss probability is smaller when the middle node transmits. Hence, when averaging over all transmitting and receiving nodes, the absolute value of the probability of message loss decreases a few percent when the feedback protocol is used. The savings from using the feedback protocol are greatest for the receiving nodes most distant from the transmitting node.

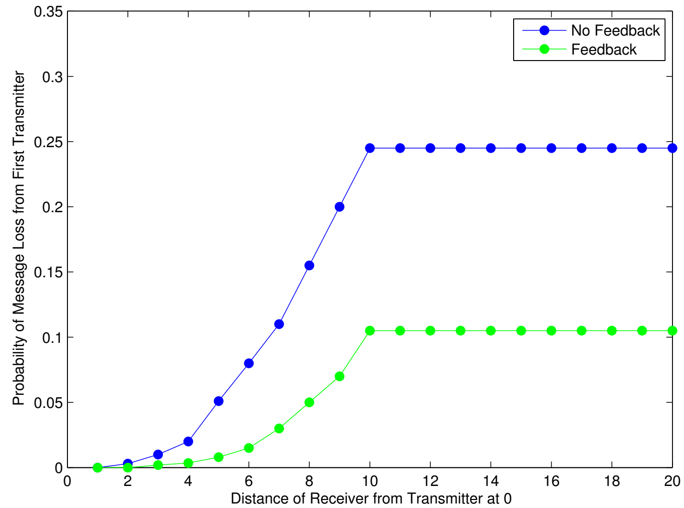

Fig. 8. Feedback to the transmitter at 0 for a 21 node network.

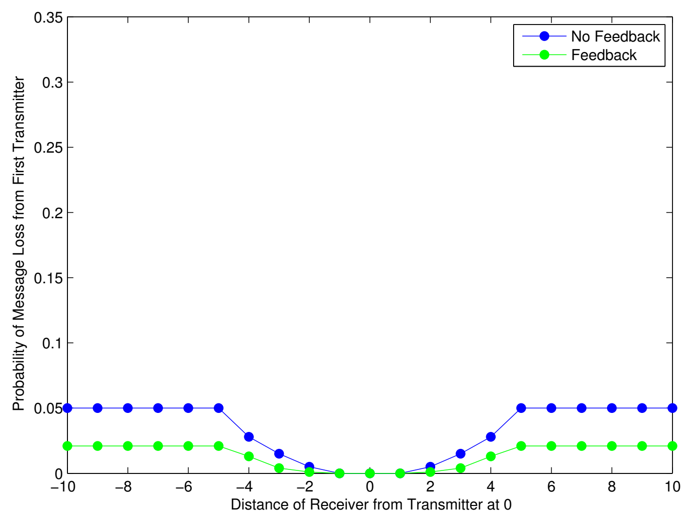

Fig. 9. Feedback to the transmitter in the middle for a 21 node network.

Figure 9 shows the impact of feedback on message loss across the linear network for messages transmitted from the center node, when the same system parameters are used. Feedback is again seen to significantly decrease message loss by a factor greater than 2 relative to loss without feedback, particularly for the half the nodes in the network that are most distant from the center node. We note that similar results for both the edge and center nodes transmitting are obtained for a 200 node linear network.

We note that while some communication systems, such as Link 16, exhibit nearly perfect capture, others may have an imperfect capture. More generally, if there is partial capture, then the plateau points move in towards the transmitter, while the probability of message loss increases at these plateau points. In the limit of no capture, the probability of message loss is greatest, and the nodes adjacent to the transmitter experience the same loss as all the other nodes in the network, and feedback from them would indicate the reception status of all nodes in a pure collision channel. Finally, if in addition to collisions, communication was subject to noise or other interference, then the plateau points could indicate if the remaining nodes definitely need more coded packets. They could also indicate with a probability depending on noise, that the remaining nodes have all received the entire message. Such a feedback mechanism could be extended to reliable communications as well. These topics are areas for future investigation.

We conclude by noting that the cost of feedback is minimal in terms of bandwidth used. In addition, the feedback is not expected to add significant delays, since the messages are delayed from the random spreading of packets.

V. Summary

The use of erasure coding, together with the random spreading in time of individual coded packets of a message to create "capture diversity", enables significant throughput gains in multi-slot messages transmitted with random access by multiple sources. It is shown how the amount of spreading can be selected to satisfy delay constraints, and that when coding is not used, spreading should not be used. Finally, it is shown that edge effects and capture result in an increase in message loss with distance from the transmitting node, until a saturation point is reached. Finally, we propose to use this phenomenon and demonstrate how it can enable consolidation of feedback and elimination of future unnecessary transmissions.

References

[1] Z. Haas and J. Deng, "Dual Busy Tone Multiple Access (DBTMA), A Multiple Access Control Scheme for Ad hoc Networks", IEEE Trans. on Commun., vol. 50, No. 6, 2002.

[2] D. Baron and Y. Birk, "Coding Schemes for Multislot Messages in Multichannel ALOHA with Deadlines", IEEE Trans. on Wireless Comm., Vol. 1, no. 2, April 2002, p. 292-301.

[3] Y. Birk and Y. Keren, "Judicious Use of Redundant Transmissions in Multichannel ALOHA Networks with Deadlines", IEEE JSAC, vol. 17, no. 2, Feb. 1999, p. 257-269.

[4] L. Zeger, "Packet Erasure Coding with Random Access to Reduce Losses of Delay Sensitive Multislot Messages'", MILCOM 2009, Boston, MA, 2009.

[5] M. Riemensberger, M. Heindlmaier, A. Dotzler, D. Traskov, W. Utschick, "Optimal Slotted Random Access in Coded Wireless Packet Networks", RAWNET, Avignon France, June 2010.

[6] L. Zeger, "Spatial Distribution of Traffic Density in ALOHA with Capture", IEEE Trans. Commun., Vol. 46, No. 12, Dec. 1998.