

Network Coding for Efficient Wireless Unicast

Desmond S. Lun and Muriel Médard
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
E-mail: {dslun, medard}@mit.edu

Ralf Koetter
Coordinated Science Laboratory
University of Illinois
Urbana, IL 61801, USA
E-mail: koetter@uiuc.edu

Abstract—We consider the problem of establishing efficient unicast connections over wireless packet networks. We show how network coding, combined with distributed flow optimization, gives a practicable approach that promises to significantly outperform the present approach of end-to-end or link-by-link retransmission combined with route optimization, where performance may be measured in terms of energy consumption, congestion, or any other cost that increases with the number of transmissions made by each node. We present a specific coding scheme and specific distributed flow optimization techniques that may be used to form the basis of a protocol.

I. INTRODUCTION

Most of the work to date on network coding focuses on multicast. Indeed, this is the case with the original work on network coding [1], [2], [3] and, for the main problem considered in the original work, that of establishing a single connection over a lossless wireline packet network, network coding offers no advantage over simple routing in the unicast case. But that is not to say that the utility of network coding is strictly restricted to multicast. It is well-known, for example, that coding among multiple connections offers an advantage, whether they are multicast or unicast. The problem of coding for multiple unicast connections has been studied, and some results (e.g. [4]) are encouraging, while others (e.g. [5]) are not so encouraging. It is, however, generally agreed that the problem is difficult and only really amenable to ad hoc solutions.

It is also known that coding over lossy packet networks offers an advantage, whether it is for a single connection or for multiple connections, and whether they are multicast or unicast. In the case of a single connection, the coding problem is quite benign. The capacity of coding is known [6], and several coding schemes have been proposed [7], [8], [9], [10], [11], both for wireline networks, which consist solely of point-to-point links, and for wireless networks, which may contain broadcast links.

The wireless case is particularly interesting: Here, because the links cause packets to be spread about in probabilistic manner, there is no reason to restrict information flow to a path as in wireline networks. Rather, every node in the network can potentially act as a relay, coding packets it receives and sending out these coded packets. Therefore, the very concept of routing breaks down. Another approach is needed that does not treat wireless packet networks merely as wireline packet networks without the wires, allowing us instead to exploit

natural advantages offered by the wireless medium. In such an approach, the question is not simply what route packets should follow, but rather how many packets nodes should send and what those packets should contain.

This is the question we consider in the present paper. We focus on unicast, though our considerations can be straightforwardly extended to multicast. Thus, we consider the problem of establishing efficient unicast connections over coded wireless packet networks. We give a distributed approach to solving this problem that we believe may be practicable. The approach comes from prior work on efficient operation of coded packet networks (e.g. [12]), which we review.

Before proceeding further, however, it behooves us to ask, how much do we gain by a coding approach? Indeed, is there a compelling reason to change the current approach to wireless unicast?

II. A COMPARISON OF VARIOUS APPROACHES TO WIRELESS UNICAST

To address the issue of how much we gain by a coding approach, we conducted a simple simulation of various approaches to wireless unicast. We simulated wireless networks where nodes were placed randomly according to a uniform distribution over a square region. The size of square was set to achieve unit node density. We considered a network where transmissions were subject to distance attenuation and Rayleigh fading, but not interference (owing to scheduling). So, when node i transmits, the signal-to-noise ratio (SNR) of the signal received at node j is $\gamma d(i, j)^{-\alpha}$, where γ is an exponentially-distributed random variable with unit mean, $d(i, j)$ is the distance between node i and node j , and α is an attenuation parameter that we took to be 2. We assumed that a packet transmitted by node i is successfully received by node j if the received SNR exceeds β , i.e.

$$\gamma d(i, j)^{-\alpha} \geq \beta,$$

where β is a threshold that we took to be $1/4$. If a packet is not successfully received, then it is completely lost.

We considered five different approaches to wireless unicast; approaches (1)–(3) do not use network coding, while approaches (4) and (5) do:

- 1) **End-to-end retransmission:** A path is chosen from source to sink, and packets are acknowledged by the sink, or destination node. If the acknowledgment for

a packet is not received by the source, the packet is retransmitted. This represents the situation where reliability is provided by a retransmission scheme at the transport layer, e.g. by the transport control protocol (TCP), and no mechanism for reliability is present at the link layer.

- 2) **End-to-end coding:** A path is chosen from source to sink, and an end-to-end forward error correction (FEC) code, such as a Reed-Solomon code, an LT code [13], or a Raptor code [14], is used to correct for packets lost between source and sink.
- 3) **Link-by-link retransmission:** A path is chosen from source to sink, and automatic repeat request (ARQ) is used at the link layer to request the retransmission of packets lost on every link in the path. Thus, on every link, packets are acknowledged by the intended receiver and, if the acknowledgment for a packet is not received by the sender, the packet is retransmitted.
- 4) **Path coding:** A path is chosen from source to sink, and every node on the path employs coding to correct for lost packets. The most straightforward way of doing this is for each node to use one of the FEC codes for end-to-end coding, decoding and re-encoding packets it receives. The main drawback of such an approach is delay. Every node on the path codes and decodes packets in a block. A way of overcoming this drawback is to use codes that operate in a more of a “convolutional” manner, sending out coded packets formed from packets received thus far, without decoding. The random linear coding scheme from [7], [8], [9] is such a code. A variation, with lower complexity, is presented in [10].
- 5) **Full coding:** In this case, paths are eschewed altogether. A subgraph that specifies the frequency with which every node transmits packets is chosen, and the random linear coding scheme from [7], [8], [9] is used. This represents the ultimate limit of achievability provided that we are restricted from modifying the design of the physical layer.

In all cases where acknowledgments are sent, acknowledgments are subject to loss in the same way that packets are and follow the same path.

The average number of transmissions required per packet using the various approaches in random networks of varying size is shown in Figure 1. We see that, while end-to-end coding and link-by-link retransmission already represent significant improvements on end-to-end retransmission, the network coding approaches represent more significant improvements still. By a network size of nine nodes, full coding already improves on link-by-link retransmission by a factor of two. Moreover, as the network size grows, the performance of the various schemes diverges. Here, we discuss performance simply in terms of the number of transmissions required per packet; in some cases, e.g. congestion, the performance measure increases super-linearly in this quantity, and the performance improvement is even greater than that depicted in Figure 1. We see, at any rate, that the use of network coding promises sig-

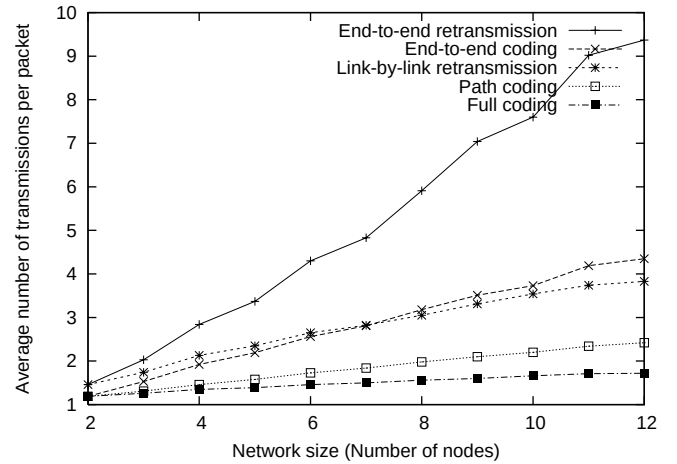


Fig. 1. Average number of transmissions required per packet using various wireless unicast approaches in random networks of varying size. Sources and sinks were chosen randomly according to a uniform distribution. Paths or subgraphs were chosen in each random instance to minimize the total number of transmissions required, except in the cases of end-to-end retransmission and end-to-end coding, where they were chosen to minimize the number of transmissions required by the source node (the optimization to minimize the total number of transmissions in these cases cannot be done straightforwardly by a shortest path algorithm).

nificant improvements, particularly for large networks. Given the potential benefits promised by network coding, we proceed to describe how they may be achieved. We first introduce a model of wireless networks.

III. MODEL

We model the network with a directed hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of hyperarcs. A hypergraph is a generalization of a graph, where, rather than arcs, we have hyperarcs. A hyperarc is a pair (i, J) , where i , the start node, is an element of $\mathcal{N}$ and J , the set of end nodes, is a non-empty subset of $\mathcal{N}$.

Each hyperarc (i, J) represents a wireless broadcast link from node i to nodes in the non-empty set J . Let A_{iJK} be the counting process describing the arrival of packets that are injected on hyperarc (i, J) and received by exactly the set of nodes $K \subset J$, i.e. for $\tau \geq 0$, $A_{iJK}(\tau)$ is the total number of packets that are injected on hyperarc (i, J) and received by all nodes in K (and no nodes in $\mathcal{N} \setminus K$) between time 0 and time τ . For example, suppose that three packets are injected on hyperarc $(1, \{2, 3\})$ between time 0 and time 1 and that, of these three packets, one is received by node 2 only, one is lost entirely, and one is received by both nodes 2 and 3; then we have $A_{1(23)\emptyset}(1) = 1$, $A_{1(23)2}(1) = 1$, $A_{1(23)3}(1) = 0$, and $A_{1(23)(23)}(1) = 1$. We assume that A_{iJK} has an average rate z_{iJK} ; more precisely, we assume that

$$\lim_{\tau \rightarrow \infty} \frac{A_{iJK}(\tau)}{\tau} = z_{iJK}$$

almost surely.

Let $z_{iJ} := \sum_{K \subset J} z_{iJK}$ be the average rate at which packets are injected into hyperarc (i, J) . The rate vector z , consisting

of z_{iJ} , $(i, J) \in \mathcal{A}$, is the coding subgraph for the connection of interest and can be varied within a constraint set Z dictated to us by lower layers (for examples of such constraint sets, see [15], [16], [17], [18], [19]). We reasonably assume that Z is a convex subset of the positive orthant containing the origin. We associate with the network a cost function f that maps feasible coding subgraphs to real numbers and that we seek to minimize. For wireless networks, it is common for the cost function to reflect energy consumption, but it could also represent, for example, average latency, monetary cost, or a combination of these considerations.

IV. DISTRIBUTED RANDOM NETWORK CODING

Following the approach put forward in [20], we separate the problems of determining how many packets nodes should send and what those packet should contain. We start by addressing the second problem. In [7], [8], [9], a coding scheme, which we refer to as distributed random network coding, is described and shown to achieve the capacity of any given subgraph. Therefore, there is loss in throughput in separating the two problems and addressing the second problem by distributed random network coding. We propose doing exactly that and now describe distributed random network coding.

We suppose that, at the source node s , we have K message packets $w_1, w_2, \dots, w_K$, which are vectors of length ρ over the finite field $\mathbb{F}_q$. (If the packet length is b bits, then we take $\rho = \lceil b / \log_2 q \rceil$.) The message packets are initially present in the memory of node s .

The coding operation performed by each node is simple to describe and is the same for every node: Received packets are stored into the node's memory, and packets are formed for injection with random linear combinations of its memory contents whenever a packet injection occurs on an outgoing link. The coefficients of the combination are drawn uniformly from $\mathbb{F}_q$.

Since all coding is linear, we can write any packet x in the network as a linear combination of $w_1, w_2, \dots, w_K$, namely, $x = \sum_{k=1}^K \gamma_k w_k$. We call γ the *global encoding vector* of x , and we assume that it is sent along with x , as side information in its header. The overhead this incurs (namely, $K \log_2 q$ bits) is negligible if packets are sufficiently large.

Nodes are assumed to have unlimited memory. The scheme can be modified so that received packets are stored into memory only if their global encoding vectors are linearly-independent of those already stored. This modification maintains the rate optimality of the scheme while ensuring that nodes never need to store more than K packets.

A sink node collects packets and, if it has K packets with linearly-independent global encoding vectors, it is able to recover the message packets. Decoding can be done by Gaussian elimination, and the scheme can be operated ratelessly, i.e. it can be run indefinitely until successful reception (at which stage that fact is signaled to other nodes).

If a connection is feasible over a given subgraph, then the probability that the sink cannot decode successfully approaches 0 as K approaches infinity. Hence, given a coding

subgraph z , a unicast of rate arbitrarily close to R is achievable with coding from source node s to sink node t if and only if there exists a flow vector x satisfying

$$\sum_{\{J|(i,J) \in \mathcal{A}\}} \sum_{j \in J} x_{iJj} - \sum_{\{j|(j,I) \in \mathcal{A}, i \in I\}} x_{jIi} = \begin{cases} R & \text{if } i = s, \\ -R & \text{if } i = t, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

for all $i \in \mathcal{N}$, and

$$\sum_{j \in K} x_{iJj} \leq \sum_{\{L \subset J | L \cap K \neq \emptyset\}} z_{iJL} \quad (2)$$

for all $(i, J) \in \mathcal{A}$ and $K \subset J$. We define the bounded polyhedron F to be the set of non-negative vectors x satisfying (1) and (2).

V. SUBGRAPH SELECTION

We now move to the problem of determining how many packets nodes should send, or, of finding the coding subgraph of minimum cost that admits the desired connection. We see that it equates to the following optimization problem.

$$\begin{aligned} & \text{minimize } f(z) \\ & \text{subject to } z \in Z, \\ & \sum_{j \in K} x_{iJj} \leq \sum_{\{L \subset J | L \cap K \neq \emptyset\}} z_{iJL}, \quad \forall (i, J) \in \mathcal{A}, K \subset J, \\ & x \in F. \end{aligned} \quad (3)$$

If x conforms to a single path, then path coding results.

Problem (3) is by no means a simple optimization problem, so we make some simplifying assumptions. We assume that f is of the form $f(z) = \sum_{(i,J) \in \mathcal{A}} f_{iJ}(z_{iJ})$, where f_{iJ} is a monotonically increasing, convex function, and that, as z_{iJ} is varied, z_{iJK}/z_{iJ} is constant for all $K \subset J$. Therefore,

$$b_{iJK} := \frac{\sum_{\{L \subset J | L \cap K \neq \emptyset\}} z_{iJL}}{z_{iJ}}$$

is a constant. We now see that problem (3) is a convex optimization problem whose difficulty, and whose ability to be solved in a distributed manner, principally depends on the structure of the constraint set Z . We consider, in this paper, the simplest case where Z can be dropped entirely, which already poses a challenge. We note that, even though we drop Z , separable constraints can still be handled by making f_{iJ} approach infinity as z_{iJ} approaches the upper constraint associated with hyperarc (i, J) .

We arrive at the following convex optimization problem.

$$\begin{aligned} & \text{minimize } \sum_{(i,J) \in \mathcal{A}} f_{iJ}(z_{iJ}) \\ & \text{subject to } \sum_{j \in K} x_{iJj} \leq z_{iJ} b_{iJK}, \quad \forall (i, J) \in \mathcal{A}, K \subset J, \\ & x \in F. \end{aligned} \quad (4)$$

This problem is very similar to the principal optimization problem studied in [12]. The main difference is that it involves more constraints. In fact, the number of constraints grows exponentially in $\max_{(i,J) \in \mathcal{A}} |J|$, the maximum number of nodes reached by a hyperarc. If the connectivity of nodes is kept more or less constant as the network grows, however, then the problem should be computationally tractable. In [12], two distributed approaches to the problem are proposed.

The first applies in the case of linear cost, i.e. in the case where $f_{iJ}(z_{iJ}) = a_{iJ}z_{iJ}$ for all $(i, J) \in \mathcal{A}$, where $a_{iJ} > 0$. We take the Lagrangian dual of the problem:

$$\begin{aligned} & \text{maximize} \quad \sum_{t \in T} q^{(t)}(p^{(t)}) \\ & \text{subject to} \quad \sum_{t \in T} \sum_{K \subset J} p_{iJK}^{(t)} = a_{iJ} \quad \forall (i, J) \in \mathcal{A}, \\ & \quad p_{iJK}^{(t)} \geq 0, \quad \forall (i, J) \in \mathcal{A}, K \subset J, t \in T, \end{aligned}$$

where

$$q^{(t)}(p^{(t)}) := \min_{x^{(t)} \in F^{(t)}} \sum_{(i,J) \in \mathcal{A}} \sum_{j \in J} \left(\sum_{\{K \subset J | K \ni j\}} \frac{p_{iJK}^{(t)}}{b_{iJK}} \right) x_{iJj}. \quad (5)$$

The subproblems given by (5) are simply shortest path problems which can be solved, for example, by the distributed Bellman-Ford algorithm. We then solve the dual problem by approaching the optimum along subgradients and recover an optimal primal solution from the series of iterates. Details of the algorithm can be found in [12].

The second approach is to solve the following approximation of problem (4):

$$\begin{aligned} & \text{minimize} \quad \sum_{(i,J) \in \mathcal{A}} f_{iJ}(z'_{iJ}) \\ & \text{subject to} \quad x \in F, \end{aligned}$$

where

$$z'_{ij} := \left(\sum_{K \subset J, t \in T} \left(\frac{\sum_{j \in K} x_{iJj}^{(t)}}{b_{iJK}} \right)^n \right)^{1/n}.$$

This approximation becomes exact as $n \rightarrow \infty$. We now merely have a convex multicommodity flow problem. There are many algorithms for convex multicommodity flow problems (see [21] for a survey), some of which (e.g. the algorithms of [22], [23]) are well-suited for distributed implementation. In [12], a primal-dual approach is used to solve the convex multicommodity flow problem.

VI. CONCLUSION

We have seen, through simulation, that network coding is a promising way to achieve significant performance improvements for unicast in multi-hop wireless networks, and we have suggested practicable distributed methods to implement coding. But how these methods actually perform in practical situations and how they compare to each other is currently not well understood. We have merely set out a direction, and more

work clearly is required before we can exploit the promised gains.

REFERENCES

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.
- [2] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.
- [3] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.
- [4] S. Katti, D. Katabi, W. Hu, H. Rahul, and M. Médard, "The importance of being opportunistic: Practical network coding for wireless environments," in *Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2005.
- [5] Z. Li and B. Li, "Network coding: The case of multiple unicast sessions," in *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, Sept.–Oct. 2004.
- [6] R. Gowaikar, A. F. Dana, R. Palanki, B. Hassibi, and M. Effros, "On the capacity of wireless erasure networks," in *Proc. 2004 IEEE International Symposium on Information Theory (ISIT 2004)*, Chicago, IL, June–July 2004, p. 401.
- [7] D. S. Lun, M. Médard, R. Koetter, and M. Effros, "On coding for reliable communication over packet networks," submitted to *IEEE Trans. Inform. Theory*.
- [8] —, "Further results on coding for reliable communication over packet networks," in *Proc. 2005 IEEE International Symposium on Information Theory (ISIT 2005)*, Sept. 2005, pp. 1848–1852.
- [9] D. S. Lun, M. Médard, and M. Effros, "On coding for reliable communication over packet networks," in *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, Sept.–Oct. 2004, invited paper.
- [10] P. Pakzad, C. Fragouli, and A. Shokrollahi, "Coding schemes for line networks," in *Proc. 2005 IEEE International Symposium on Information Theory (ISIT 2005)*, Sept. 2005.
- [11] R. Khalili and K. Salamatian, "On the capacity of multiple input erasure relay channels," in *Proc. WINMEE, RAWNET and NETCOD 2005 Workshops*, Apr. 2005.
- [12] D. S. Lun, N. Ratnakar, R. Koetter, M. Médard, E. Ahmed, and H. Lee, "Achieving minimum-cost multicast: A decentralized approach based on network coding," in *Proc. IEEE Infocom 2005*, vol. 3, Miami, FL, Mar. 2005, pp. 1608–1617.
- [13] M. Luby, "LT codes," in *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, Nov. 2002, pp. 271–280.
- [14] A. Shokrollahi, "Raptor codes," Jan. 2004, preprint.
- [15] R. L. Cruz and A. V. Santhanam, "Optimal routing, link scheduling and power control in multi-hop wireless networks," in *Proc. IEEE Infocom 2003*, vol. 1, Mar.–Apr. 2003, pp. 702–711.
- [16] M. Johansson, L. Xiao, and S. Boyd, "Simultaneous routing and power allocation in CDMA wireless data networks," in *Proc. 2003 IEEE International Conference on Communications (ICC 2003)*, vol. 1, May 2003, pp. 51–55.
- [17] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, July 2004.
- [18] M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless mesh networks with orthogonal channels," *IEEE/ACM Trans. Networking*, vol. 13, no. 4, pp. 868–880, Aug. 2005.
- [19] Y. Wu, P. A. Chou, Q. Zhang, K. Jain, W. Zhu, and S.-Y. Kung, "Network planning in wireless ad hoc networks: A cross-layer approach," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 136–150, Jan. 2005.
- [20] D. S. Lun, M. Médard, T. Ho, and R. Koetter, "Network coding with a cost criterion," in *Proc. 2004 International Symposium on Information Theory and its Applications (ISITA 2004)*, Oct. 2004, pp. 1232–1237.
- [21] A. Ouorou, P. Mahey, and J.-P. Vial, "A survey of algorithms for convex multicommodity flow problems," *Manage. Sci.*, vol. 46, no. 1, pp. 126–147, Jan. 2000.
- [22] D. P. Bertsekas, "A class of optimal routing algorithms for communication networks," in *Proc. Fifth International Conference on Computers and Communication (ICCC 80)*, Oct. 1980, pp. 71–76.
- [23] D. P. Bertsekas, E. M. Gafni, and R. G. Gallager, "Second derivative algorithms for minimum delay distributed routing in networks," *IEEE Trans. Commun.*, vol. 32, no. 8, pp. 911–919, Aug. 1984.