On Coding for Reliable Communication over Packet Networks*

Desmond S. Lun Muriel Médard Michelle Effros

Abstract

We present a capacity-approaching coding scheme for unicast or multicast over lossy packet networks. In the scheme, all nodes perform coding, but do not wait for a full block of packets to be received before sending out coded packets. Rather, whenever they have a transmission opportunity, they form coded packets with random linear combinations of previously received packets. All coding and decoding operations in the scheme have polynomial complexity.

Our analysis of the scheme shows that it is not only capacity-approaching, but that the propagation of packets carrying "innovative" information follows that of a queueing network where every node acts as a stable $M/M/1$ queue. We consider networks with both lossy point-to-point and broadcast links, allowing us to model both wireline and wireless packet networks.

1 Introduction

Packet losses in networks result from a variety of causes, which include congestion, buffer overflows, and, in wireless networks, link outage due to fading. Thus a method to ensure reliable communication is necessary, and the prevailing approach is for the receiver to send requests for the retransmission of lost packets over some feedback channel. There are, however, a number of drawbacks to such an approach, which are evident most notably in high-loss environments and for multicast connections. In both instances, many requests for retransmissions are usually required, which place an unnecessary load on the network and which may overwhelm the source. In the latter instance, there is the additional problem that retransmitted packets are often only of use to a subset of the receivers and are therefore redundant to the remainder.

An approach that overcomes these drawbacks is to use erasure-correcting codes. Under such an approach, the original packets are reconstructed from those that are received and little or no feedback is required. This approach has been recently exemplified by digital fountain codes [1, 2, 3], which are fast, near-optimal erasure codes. Such codes can approach the capacity of connections over lossy packet networks, provided that the connection as a whole is viewed as a single channel and coding is performed only at the

*This work was supported by the National Science Foundation under grant nos. CCR-0093349 and CCR-0325496, by Hewlett-Packard under grant no. 008542-008, and by the Air Force Office of Scientific Research under grant no. F49620-01-1-0365. D. S. Lun (dslun@mit.edu) and M. Médard (medard@mit.edu) are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. M. Effros (effros@caltech.edu) is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

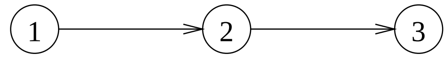

Figure 1: A network consisting of two links in tandem. If both links are lossy, then a greater capacity can be achieved for the connection between node 1 and node 3 by not treating it as a single channel, and allowing for coding at node 2.

source node. But in lossy packet networks where all nodes have the capability for coding, such as overlay networks using UDP and wireless networks, there is no compelling reason to adopt this view, and a greater capacity can in fact be achieved if we do not.

Consider the network consisting of two links in tandem shown in Figure 1. Suppose that both links can transmit one packet per unit time and that packets are lost on links $(1, 2)$ and $(2, 3)$ with probabilities ε_{12} and ε_{23}, respectively. We see that, by using erasure coding (such as digital fountain coding), we are able to transmit packets from node 1 to node 3 at rates not exceeding $(1-\varepsilon_{12})(1-\varepsilon_{23})$ packets per unit time. But if, rather than treating the connection between node 1 and node 3 as a single channel, we recognize that it is composed of two separate channels, then, by allowing node 2 to also code, we are able to transmit packets from node 1 to node 3 at rates not exceeding $\min(1-\varepsilon_{12}, 1-\varepsilon_{23})$ packets per unit time [4], which is, in general, greater than $(1 - \varepsilon_{12})(1 - \varepsilon_{23})$ packets per unit time. An obvious way to approach this capacity is to use an erasure code between nodes 1 and 2 and another between nodes 2 and 3. Such a scheme, however, would introduce significant delay as it would require an entire block of packets to be received at node 2, decoded, then re-encoded, and sent to node 3. Moreover, the severity of this delay problem clearly increases with the size of the network.

In the present paper, we consider an alternative feedback-free operation that is expected to incur considerably less delay: Each node performs coding, but does not wait for a full block of packets to be received before it sends out coded packets. Rather, it has a memory in which it stores received packets, and it forms coded packets with random linear combinations of its memory contents whenever it has a transmission opportunity. The scheme is decentralized, requiring no co-ordination, and can be operated ratelessly, that is, it can be run indefinitely until successful reception (at which stage that fact is signaled to all nodes, requiring a comparatively insignificant amount of feedback), which is a particularly useful property when the loss rates are not known precisely. We analyze the scheme and show that it can approach the capacity of connections over packet networks where all nodes are allowed to perform arbitrary coding. Our work is inspired on the one hand by the proposals of Chou et al. [5] and Ho et al. [6, 7], which are random linear coding schemes for lossless networks, and on the other by LT coding [1] and its extensions, which are random linear coding schemes for lossy channels. Since lossless networks and lossy channels are special cases of lossy networks, we extend and unify previous work.

We commence by considering unicast connections in the following section, then multicast connections in Section 3. In Section 4, we extend our model beyond the conventional model of lossy point-to-point links to consider lossy broadcast links; that is, links modeled by broadcast erasure channels.

2 Unicast connections

We model the network as a directed, acyclic graph $G = (N, A)$, where N is the set of nodes and A is the set of arcs, or links. Each arc (i, j) represents an erasure channel with erasure probability ε_{ij} and maximum input rate r_{ij}. We suppose that we wish to establish a connection of rate arbitrarily close to R packets per unit time from source node s to terminal node t. We intend for G, ε, and r not to represent the full network, but the portion of the network that has been allocated to support the connection from s to t. This allocation is determined, for example, by finding the least costly allocation that allows for the connection to be supported (see [8, 9]). Thus, the assumption that G is acyclic is reasonable.

We assume that link (i, j) allows packets being injected according to a Poisson process with rate not exceeding r_{ij} to be instantaneously received at j, provided that they are not erased. (The assumption that the injected traffic is of a Poisson nature exists solely for ease of analysis, and we expect our results to hold under more general conditions.) We justify this assumption on the basis that the rate r_{ij} represents the portion of the full maximum input rate of link (i, j) that has been allocated to this connection and therefore packets arriving according to a Poisson process with rate not exceeding r_{ij} will presumably not overload the link and eventually be received after some delay if they are not erased. The delay, consisting of queueing, transmission, and propagation delay, will however generally depend on other traffic in the network and so, to abstract away this complication, we assume that packets are received instantaneously if they are not erased. We expect, at any rate, that this delay will not significantly alter our conclusions, just as it does not significantly alter the main conclusions of [10, 11], which deal with linear coding in lossless networks.

Suppose that, at the source node s, we have k message packets $w_1, w_2, \ldots, w_k$, which are vectors of length ρ over the finite field $\mathbb{F}_q$. (If the packet length is b bits, then we take $\rho = \lceil b/\log_2(q)\rceil$.) From $w_1, w_2, \ldots, w_k$, we form n packets $v_1, v_2, \ldots, v_n$ with random linear combinations; so, for all $l = 1, 2, \ldots, n$, we have

$$v_l = \sum_{m=1}^{k} \alpha_{lm} w_m,$$

where each α_{lm} is a random element of $\mathbb{F}_q$ chosen according to a uniform distribution. The packets $v_1, v_2, \ldots, v_n$ are injected at s according to a Poisson distribution with rate R_0, where $R_0 = R(1 - \delta)$ for some $\delta > 0$.

Suppose that

$$R \leq \min_{Q \in \mathcal{Q}(s,t)} \left\{ \sum_{(i,j)\in\Gamma_+(Q)} r_{ij}(1 - \varepsilon_{ij}) \right\},$$

where $\mathcal{Q}(s, t)$ is the set of all cuts between s and t, and $\Gamma_+(Q)$ denotes the set of forward arcs of the cut Q, i.e.

$$\Gamma_+(Q) := \{(i, j) \in A \,|\, i \in Q, j \notin Q\}.$$

Therefore, by the max-flow/min-cut theorem (see, for example, [12, Section 3.1]), there exists a vector f satisfying

$$\sum_{\{j|(i,j)\in A\}} f_{ij} - \sum_{\{j|(j,i)\in A\}} f_{ji} = \begin{cases} R_0 & \text{if } i = s, \\ -R_0 & \text{if } i = t, \\ 0 & \text{otherwise,} \end{cases}$$

for all $i \in N$, and

$$f_{ij} \le r_{ij}(1-\varepsilon_{ij})(1-\delta),$$

for all $(i,j) \in A$.

On each link (i,j), we inject packets, which are formed with random linear combinations of the memory contents of i (namely, all packets received thus far by i), according to a Poisson process with rate $f_{ij}/\{(1-\varepsilon_{ij})(1-\delta)\}$ (which does not exceed r_{ij}). Since all coding performed by all nodes is linear, we can write any packet x in the network as a linear combination of $v_1, v_2, \ldots, v_n$:

$$x = \sum_{l=1}^{n} \beta_l v_l.$$

We call the vector β associated with x the *auxiliary encoding vector* of x. Note also that we have

$$x = \sum_{l=1}^{n} \beta_l \sum_{m=1}^{k} \alpha_{lm} w_j = \sum_{m=1}^{k} \left(\sum_{l=1}^{n} \beta_l \alpha_{lm} \right) w_m = \sum_{m=1}^{k} \gamma_m w_m,$$

where $\gamma_m = \sum_{l=1}^{n} \beta_l \alpha_{lm}$. We call the vector γ associated with x the *global encoding vector* of x, and we assume that it is sent along with x, in its header. We shall see why this is necessary when we later describe the decoding process. Note that the overhead incurred by placing a global encoding vector in every packet header can be made arbitrarily small by making the packets arbitrarily large.

Recall that the packets $v_1, v_2, \ldots, v_n$ are arriving at s according to a Poisson distribution with rate R_0. Now, on each of the outgoing links of s, packets are injected according to a Poisson process with rate $f_{sj}/((1-\varepsilon_{sj})(1-\delta))$; since packets are erased with probability ε_{sj} on link (s,j), they are received according to a Poisson process with rate $f_{sj}/(1-\delta)$. Merging these processes, we see that packets are received by outward neighbors of s according to a Poisson process with rate $\sum_{\{j|(s,j)\in A\}} f_{sj}/(1-\delta) = R_0/(1-\delta)$.

Suppose that, at some time τ, the packet x with auxiliary encoding vector β is received by an outward neighbor of s. We say x is *innovative* if β does not lie in the span of the auxiliary encoding vectors of all innovative packets previously received by outward neighbors of s; that is, we consider a set $V_\tau(\Gamma_+(s))$, which consists of the auxiliary encoding vectors of innovative packets received by outward neighbors of s up to time τ (so $V_0(\Gamma_+(s)) = \emptyset$), and x is considered innovative if $\beta \notin \mathrm{span}(V_\tau(\Gamma_+(s)))$.

Now suppose that packets $v_1, v_2, \ldots, v_m$ have arrived at s and that $|V_\tau(\Gamma_+(s))| < m$. The packet x is formed from a random linear combination of $v_1, v_2, \ldots, v_m$, and it is clear that x is innovative with some non-trivial probability. We can bound this probability using the following lemma from [13]. We quote the lemma without repeating the proof.

Lemma 1. *[13, Lemma 2.1] Let V_1 and V_2 be two collections of vectors from $\mathbb{F}_q^n$, and let β be a random linear combination of the vectors in V_1, with the coefficients of the combination drawn uniformly from $\mathbb{F}_q$. Then*

$$\Pr(\beta \notin \mathrm{span}(V_2) \,|\, \mathrm{span}(V_1) \not\subset \mathrm{span}(V_2)) \ge 1 - \frac{1}{q}.$$

It follows from Lemma 1 that x is innovative with probability not less than $1 - 1/q$. Since we can always discard innovative packets, we assume that x is innovative with probability exactly $1 - 1/q$. Hence, from a point when $v_1, v_2, \ldots, v_m$ are at s and $|V_\tau(\Gamma_+(s))| < m$, the time until an innovative packet is next received by an outward neighbor of s is distributed exponentially with mean $(1-\delta)/(R_0(1-1/q))$. Suppose instead that $|V_\tau(\Gamma_+(s))| = m$. Then we see that x cannot be innovative and this remains true until another arrival occurs at s. Thus, the system set up at s is effectively that of an $M/M/1$ queueing system with arrival rate R_0 and service rate $R_0(1-1/q)/(1-\delta)$.

The queueing system is stable provided that $R_0 < R_0(1-1/q)/(1-\delta)$, i.e. $1/q < \delta$, which can be achieved by making q sufficiently large. Moreover, if the system is run for a sufficiently long period of time to allow it to reach steady-state, then by Burke's theorem (see, for example, [14, Section 2.1]), the departure process is Poisson with rate R_0. And it follows that the arrival process for innovative packets at an outward neighbor j of s is Poisson with rate

$$\frac{f_{sj}/(1-\delta)}{\sum_{\{j|(s,j)\in A\}} f_{sj}/(1-\delta)} \cdot R_0 = \frac{f_{sj}}{R_0} \cdot R_0 = f_{sj}.$$

We consider the remainder of the nodes in the network in a topological ordering. We assume, without loss of generality, that s is the first node in this ordering and t is the last. Thus we can write $N = \{i_1, i_2, \ldots, i_{|N|}\}$, where $i_1 = s$, $i_{|N|} = t$, and $l < m$ for all $(i_l, i_m) \in A$. For $Q = \{i_1, i_2, \ldots, i_l\}$, define

$$U_\tau(\Gamma_+(Q)) := V_\tau(\Gamma_+(i_l)) \cup V_\infty(\Gamma_+(Q) \setminus \Gamma_+(i_l)),$$

where, for a set of arcs B, $V_\tau(B)$ denotes the set of auxiliary encoding vectors of innovative vectors received by end nodes of arcs in B up to time τ. Then we say that a packet x received by an outward neighbor of i_l at time τ is innovative if its auxiliary encoding vector β does not lie in the span of $U_\tau(\Gamma_+(Q))$. Hence for all $\tau \geq 0$ and $l = 1, \ldots, |N|$, the set of vectors $V_\tau(\Gamma_+(Q))$, with $Q = \{i_1, i_2, \ldots, i_l\}$, is linearly-independent.

We claim that, in steady-state, innovative packets arrive at the end node j of all links (i,j) according to a Poisson process with rate f_{ij}. We have already established that this is the case for the outgoing links of s. We step through the nodes in topological order and show inductively that it is true for all the remainder. So consider a node i and let Q be the cut consisting of all nodes that do not exceed i in topological order. By hypothesis, innovative packets arrive at node i according to a Poisson process with rate $\varphi_i = \sum_{\{j|(j,i)\in A\}} f_{ji}$. Moreover, just as at s, packets are received by the outward neighbors of i according to a Poisson process with rate $\sum_{\{j|(i,j)\in A\}} f_{ij}/(1-\delta) = \varphi_i/(1-\delta)$.

Suppose that m innovative packets have arrived at i at time τ and that $|V_\tau(\Gamma_+(i))| < m$. We denote the set of arcs incoming to i (equivalently, the set of backward arcs of the cut $\{i\}$) by $\Gamma_-(i)$. So $\mathrm{span}(V_\tau(\Gamma_-(i))) \not\subset \mathrm{span}(V_\tau(\Gamma_+(i)))$. Simply by considering dimensionality, we see that $\mathrm{span}(V_\tau(\Gamma_-(i)) \cup V_\infty(\Gamma_+(Q) \setminus \Gamma_+(i))) \not\subset \mathrm{span}(U_\tau(\Gamma_+(Q)))$. Moreover, since $V_\infty(\Gamma_+(Q) \setminus \Gamma_+(i))$ is a subset of $U_\tau(\Gamma_+(Q))$, it follows that $\mathrm{span}(V_\tau(\Gamma_-(i))) \not\subset \mathrm{span}(U_\tau(\Gamma_+(Q)))$. The auxiliary encoding vector β for a packet x produced at time τ is a random linear combination of vectors from a set B that contains $V_\tau(\Gamma_-(i))$, so $\mathrm{span}(B) \not\subset \mathrm{span}(U_\tau(\Gamma_+(Q)))$. Hence, by Lemma 1, the packet x is innovative with probability not less than $1 - 1/q$; and, again, we can assume that x is innovative with probability exactly $1-1/q$ by discarding innovative packets on some occasions. If instead $|V_\tau(\Gamma_+(i))| = m$, we again see that x cannot be innovative and this remains true until another arrival occurs at i. We therefore have an $M/M/1$ queueing system with arrival rate φ_i and service rate $\varphi_i(1-1/q)/(1-\delta)$, which is again stable provided that $1/q < \delta$. Applying Burke's theorem completes the demonstration of the claim.

Suppose that we receive a collection of packets $\{x\}$ at t, which have the collection of global encoding vectors $\{\gamma\}$ in their headers. We see that, if there exists a subset of k linearly independent vectors in $\{\gamma\}$, then we are able to recover the message packets $w_1, w_2, \ldots, w_k$. Note that the decoding procedure essentially involves performing Gaussian elimination, which can be done in polynomial time.

If we receive k innovative packets at t, then, because their auxiliary encoding vectors are linearly independent, it follows that the global encoding vectors in their headers collectively form a random $k \times k$ matrix over $\mathbb{F}_q$, with all entries chosen uniformly. Thus, a decoding error occurs only if we receive fewer than k innovative packets at t or if the $k \times k$ matrix obtained from the global encoding vectors in their headers is not invertible. The latter occurs with probability $\prod_{l=1}^{k}(1 - 1/q^l)$, which can be made arbitrarily small by taking k or q arbitrarily large.

To estimate the probability of receiving fewer than k innovative packets at t, we note that every node behaves like a stable $M/M/1$ queueing system in steady-state, so, if the network of queues is run for sufficiently long, then N_i, the number of innovative packets in it, is a time-invariant random variable with finite mean. Now, the number of innovative packets received by t is $n - N_i$, so fewer than k innovative packets are received if and only if $N_i > n - k$. Take $R_c < 1$ and set $n = \lfloor k/R_c \rfloor$. Then, the probability that the number of innovative packets received by t is less than k can be made arbitrarily small by taking k arbitrarily large.

Hence the probability of error of this scheme can be made arbitrarily small. Moreover, the rate achieved by the scheme is k/τ_0, where τ_0 is the time taken for $v_1, v_2, \ldots, v_n$ to all arrive at s. Thus, for n sufficiently large, the rate is, with probability exceeding $1 - \varepsilon$,

$$\frac{k}{\tau_0} > \frac{k}{n(1/R_0 + \varepsilon)} \geq \frac{R_c R_0}{1 + \varepsilon R_0} = \frac{R_c R(1-\delta)}{1 + \varepsilon R(1-\delta)},$$

which can be made arbitrarily close to R.

We summarize our result with the following theorem statement.

Theorem 1. *Consider the network represented by the directed, acyclic graph $G = (N, A)$, the erasure probability vector ε, and the maximum input rate vector r. The random linear coding scheme we describe can achieve, with arbitrarily small error probability, a unicast connection over the network from source node s to terminal node t at rate arbitrarily close to R packets per unit time if*

$$R \leq \min_{Q \in \mathcal{Q}(s,t)} \left\{ \sum_{(i,j) \in \Gamma_+(Q)} r_{ij}(1 - \varepsilon_{ij}) \right\}.$$

The converse to Theorem 1 holds under arbitrary coding [15].

3 Multicast connections

Suppose we wish to establish a connection of rate arbitrarily close to R from source node s to terminal nodes in the set T, where R satisfies

$$R \leq \min_{t \in T} \min_{Q \in \mathcal{Q}(s,t)} \left\{ \sum_{(i,j) \in \Gamma_+(Q)} r_{ij}(1 - \varepsilon_{ij}) \right\}.$$

Let $R_0 = R(1-\delta)$ for some $\delta > 0$. Then, by the max-flow/min-cut theorem, there exists, for all $t \in T$, a vector $f^{(t)}$ satisfying

$$\sum_{\{j|(i,j)\in A\}} f_{ij}^{(t)} - \sum_{\{j|(j,i)\in A\}} f_{ji}^{(t)} = \begin{cases} R_0 & \text{if } i = s, \\ -R_0 & \text{if } i = t, \\ 0 & \text{otherwise,} \end{cases}$$

for all $i \in N$, and

$$f_{ij}^{(t)} \leq r_{ij}(1-\varepsilon_{ij})(1-\delta),$$

for all $(i,j) \in A$. Let $g_{ij} = \max_{t\in T} f_{ij}^{(t)}$. On link (i,j), we inject packets, which are formed with random linear combinations of the memory contents of i, according to a Poisson process with rate $g_{ij}/\{(1-\varepsilon_{ij})(1-\delta)\}$ (which does not exceed r_{ij}).

We track whether packets are innovative with respect to each terminal. Writing the nodes in topological order, $N = \{i_1, i_2, \ldots, i_{|N|}\}$, with $i_1 = s$, we say, for $i_l \in N$, that a packet x received by an outward neighbor of i_l at time τ is innovative with respect to t if its auxiliary encoding vector β satisfies $\beta \notin \text{span}(U_\tau^{(t)}(\Gamma_+(Q)))$, where Q is the cut $\{i_1, i_2, \ldots, i_l\}$, and if the outcome of an independent Bernoulli trial with probability of success $f_{ij}^{(t)}/g_{ij}$ is successful. Then node i is $|T|$ simultaneously-operating $M/M/1$ queueing systems with arrival rate $\varphi_i^{(t)} = \sum_{\{j|(i,j)\in A\}} f_{ij}^{(t)}$ ($\varphi_i^{(t)} = R_0$ for $i = s$) and service rate $\varphi_i^{(t)}(1-|T|/q)/(1-\delta)$ for each $t \in T$, which are all stable provided that $|T|/q < \delta$.

We summarize our result with the following theorem statement.

Theorem 2. *Consider the network represented by the directed, acyclic graph $G = (N, A)$, the erasure probability vector ε, and the maximum input rate vector r. The random linear coding scheme we describe can achieve, with arbitrarily small error probability, a multicast connection over the network from source node s to terminal nodes in the set T at rate arbitrarily close to R packets per unit time if*

$$R \leq \min_{t\in T} \min_{Q\in\mathcal{Q}(s,t)} \left\{ \sum_{(i,j)\in\Gamma_+(Q)} r_{ij}(1-\varepsilon_{ij}) \right\}.$$

The converse to Theorem 2 holds under arbitrary coding [15].

4 Wireless erasure networks

In this section, we consider networks where nodes are connected to each other by broadcast erasure channels. Such networks have recently been dubbed wireless erasure networks [16], as wireless packet links are essentially broadcast links with packet loss rates that vary with the location of the receiver, which are well-modeled by broadcast erasure channels. Indeed, this view of wireless packet links is advocated by Ganesan et al. [17] based on findings from their experimental data.

Under the model of wireless erasure networks, the case for employing a coding scheme such as the one we describe is particularly cogent, and directing packets hop-by-hop along some designated path, which is the standard manner in which wireless networks are operated, is clearly sub-optimal. We proceed to show that our scheme can approach the capacity of connections in such networks.

We presently consider unicast connections, as the extension to multicast follows from the arguments in Section 3. We model the network as a directed, acyclic hypergraph $H = (N, A)$, where N is the set of nodes and A is the set of hyperarcs. A hyperarc is a pair (i, J), where i, the head, is an element of N and J, the tail, is a non-empty subset of N. Each hyperarc (i, J) represents a broadcast erasure channel with erasure distribution ε_{iJ} for each j in J and maximum input rate r_{iJ}. The erasure distribution ε_{iJ} is a mapping from the power set of J to the interval $[0, 1]$ such that $\varepsilon_{iJ}(K)$ is the probability that erasures occur for all $k \in K$. Our model allows for arbitrary erasure distributions; in particular, the erasures for several receivers may be correlated, as we would expect if they are, for example, in close geographic proximity.

We suppose that we wish to establish a connection of rate arbitrarily close to R packets per unit time from source node s to terminal node t, where R satisfies

$$R \leq \min_{t \in T} \min_{Q \in \mathcal{Q}(s,t)} \left\{ \sum_{(i,J) \in \Gamma_+(Q)} r_{iJ} \left(1 - \varepsilon_{iJ}(J \setminus Q)\right) \right\}.$$

Here, $\Gamma_+(Q)$ is the set of forward hyperarcs of the cut, defined as

$$\Gamma_+(Q) := \{(i, J) \in A \,|\, i \in Q, J \setminus Q \neq \emptyset\}.$$

Therefore, there exists a vector f satisfying

$$\sum_{\{J|(i,J) \in A\}} \sum_{j \in J} f_{iJj} - \sum_{\{j|(j,I) \in A, i \in I\}} f_{jIi} = \begin{cases} R_0 & \text{if } i = s, \\ -R_0 & \text{if } i = t, \\ 0 & \text{otherwise,} \end{cases}$$

for all $i \in N$, and

$$\sum_{j \in K} f_{iJj} \leq r_{iJ} \left(1 - \varepsilon_{iJ}(K)\right) (1 - \delta),$$

for all $(i, J) \in A$ and $K \subset J$. On link (i, J), we inject packets, which are formed with random linear combinations of the memory contents of i, according to a Poisson process with rate $\max_{K \subset J} \sum_{j \in K} f_{iJj} / \{(1 - \varepsilon_{iJ}(K))(1 - \delta)\}$ (which does not exceed r_{iJ}). Note that for any $L \subset J$, we have

$$\max_{K \subset J} \frac{\sum_{j \in K} f_{iJj}}{(1 - \varepsilon_{iJ}(K))(1 - \delta)} \cdot (1 - \varepsilon_{iJ}(L)) \geq \frac{\sum_{j \in L} f_{iJj}}{1 - \delta},$$

so the rate at which packets are received at one or more $j \in L$ exceeds $\sum_{j \in L} f_{iJj}/(1 - \delta)$.

When a packet is transmitted on link (i, J) and is received by some $K \subset J$, we randomly choose to discard it (consider it not be innovative) or choose one of the $j \in K$ at which to consider it an innovation if, indeed, it turns out to be innovative. It can be seen that a random rule exists that ensures that the rate at which innovative packets are serviced on link (i, J) for node $j \in J$ is $f_{iJj}/(1 - \delta)$. The remainder of the argument follows from that in Section 2 and can be extended to multicast with the argument in Section 3. Hence we have the following theorem.

Theorem 3. *Consider the network represented by the directed, acyclic hypergraph $H = (N, A)$, the erasure distribution vector ε, and the maximum input rate vector r. The* *random linear coding scheme we describe can achieve, with arbitrarily small error probability, a multicast connection over the network from source node s to terminal nodes in the set T at rate arbitrarily close to R packets per unit time if*

$$R \leq \min_{t \in T} \min_{Q \in \mathcal{Q}(s,t)} \left\{ \sum_{(i,J) \in \Gamma_+(Q)} r_{iJ} \left(1 - \varepsilon_{iJ}(J \setminus Q)\right) \right\}.$$

The converse to Theorem 3 holds under arbitrary coding [16]. Note that the achievability result in [16] requires side information at the decoder in the form of the exact erasure pattern over every link in the network. It our scheme, we have no such side information requirement, but we do require the side information provided by the global encoding vectors in the packet headers. So while, in wireline erasure networks, the overhead of including the global encoding vectors in the headers is a price paid by our scheme that can be avoided with block-by-block coding, it appears that some side information is necessary to approach the capacity of wireless erasure networks, even if block-by-block coding is used. Moreover, in the limit of large networks, the overhead incurred by our scheme is less than that incurred by describing the exact erasure pattern over every link.

5 Conclusion

We have considered a scheme for reliable communication over packet networks where all nodes perform coding, but do not wait for a full block of packets to be received before sending out coded packets, and, rather, form coded packets with random linear combinations of previously received packets whenever they have a transmission opportunity. The scheme is capacity-approaching and potentially practical: The coding is linear and the decoding procedure involves nothing more complicated than Gaussian elimination. Whether the scheme can actually be effectively employed in a practical situation depends on issues such as how long the block length needs to be for acceptable performance and how significant the overhead is relative to the size of the packets. Studying these issues by simulation or analysis are avenues for future work. Other avenues include extending our results to cyclic graphs and to multiple-source multicast, and exploring adding an element of design into the scheme. The last is one that we believe is especially promising and that we are currently pursuing: The linear combinations used for coding do not, of course, need to be chosen completely randomly and improvements can perhaps be gained by not doing so.

Acknowledgments

This work was done in collaboration with Ralf Koetter. The authors would like to thank Negar Kiyavash for helpful discussions and suggestions.

References

[1] M. Luby, "LT codes," in *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, Nov. 2002, pp. 271–280.

[2] A. Shokrollahi, "Raptor codes," Jan. 2004, preprint.

[3] P. Maymounkov, "Online codes," NYU, Technical Report TR2002-833, Nov. 2002.

[4] R. E. Blahut, *Principles and Practice of Information Theory.* Reading, MA: Addison-Wesley, 1987.

[5] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2003.

[6] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "Toward a random operation of networks," submitted to *IEEE Trans. Inform. Theory.*

[7] T. Ho, M. Médard, J. Shi, M. Effros, and D. R. Karger, "On randomized network coding," in *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2003.

[8] D. S. Lun, M. Médard, T. Ho, and R. Koetter, "Network coding with a cost criterion," accepted to *2004 International Symposium on Information Theory and its Applications (ISITA 2004).*

[9] 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," submitted to *IEEE Infocom 2005.*

[10] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[11] 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.

[12] D. P. Bertsekas, *Network Optimization: Continuous and Discrete Models.* Belmont, MA: Athena Scientific, 1998.

[13] S. Deb and M. Médard, "Algebraic gossip: A network coding approach to optimal multiple rumor mongering," submitted to *IEEE Trans. Inform. Theory.*

[14] F. P. Kelly, *Reversibility and Stochastic Networks.* Chichester: John Wiley & Sons, 1979.

[15] S. P. Borade, "Network information flow: Limits and achievability," in *Proc. 2002 IEEE International Symposium on Information Theory (ISIT 2002)*, Lausanne, Switzerland, June–July 2002, p. 139.

[16] 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, USA, June–July 2004, p. 401.

[17] D. Ganesan, B. Krishnamachari, A. Woo, D. Culler, D. Estrin, and S. Wicker, "An empirical study of epidemic algorithms in large scale multihop wireless networks," Intel Research, Technical Report IRB-TR-02-003, Mar. 2002.