

# Toward A Random Operation of Networks

Tracey Ho<sup>†</sup>, Muriel Médard<sup>†</sup>, Ralf Koetter<sup>‡</sup>, David R. Karger<sup>†</sup>, Michelle Effros<sup>\*</sup>,  
Jun Shi<sup>§</sup> and Ben Leong<sup>†</sup>

<sup>†</sup>Massachusetts Institute of Technology, <sup>‡</sup>University of Illinois, Urbana-Champaign,

<sup>\*</sup>California Institute of Technology, <sup>§</sup>University of California, Los Angeles

<sup>†</sup>{trace@, medard@, karger@csail., benleong@}mit.edu, <sup>‡</sup>koetter@uiuc.edu,

<sup>\*</sup>effros@caltech.edu, <sup>§</sup>junshi@ee.ucla.edu

## Abstract

We present a distributed randomized network coding approach for transmission and compression of information in general multi-source multicast networks. Network nodes independently and randomly select linear mappings from inputs onto output links over some field. We show that this achieves optimal capacity with probability rapidly approaching 1 with the code length. We also demonstrate that randomized coding performs compression when necessary in a network, generalizing error exponents for linear Slepian-Wolf coding in a natural way. Benefits of this approach are decentralized operation and robustness to network changes or link failures. We show that this approach can take advantage of redundant network capacity for improved performance and robustness. We illustrate some potential advantages of randomized network coding over routing in two examples of practical scenarios: distributed network operation and online algorithms for networks with dynamically varying connections. Our mathematical development of these results also provides a link between network coding and network flows/bipartite matching, leading to a new bound on required field size for centralized network coding on general multicast networks.

## 1 Introduction

The achievable capacity of multicast networks with network coding was given in [1]. We show how to achieve this capacity in a distributed setting, using an efficient randomized approach.

We consider the most general multicast framework – multi-source multicast, possibly with correlated sources, on general networks. This family of problems includes traditional single-source multicast for content delivery and the reachback problem for sensor networks, in which several, possibly correlated, sources transmit to a single receiver. We use a randomized strategy: all nodes other than the receiver nodes perform random linear mappings from inputs onto outputs over some field. These mappings are selected independently at each node. An illustration is given in Figure 1. The receivers need only know the overall linear combination of source processes in each of their incoming signals. This information can be sent with each signal or packet as a vector of coefficients corresponding to each of the source processes, and updated at each coding node by applying the same linear mappings to the coefficient vectors as to the information signals. The

required overhead of transmitting these coefficients decreases with increasing length of blocks over which the codes and network state are expected to remain constant.

Our primary results show, firstly, that such randomized coding achieves maximum multicast capacity with probability rapidly approaching 1 with the length of code. Secondly, in the context of a distributed source coding problem, we demonstrate that randomized coding also performs compression when necessary in a network, generalizing known error exponents for linear Slepian-Wolf coding [4] in a natural way.

This approach not only recovers the capacity and achievable rates, but also offers a number of advantages. While optimal capacity can be achieved without a randomized approach, this requires, in general, careful and centrally-planned solutions. We consider the case of distributed operation of a network whose conditions may be varying over time. Our work hints at a beguiling possibility for network operation: that a network may be operated in a decentralized manner and still achieve the performance of the optimized solution. The distributed nature of our approach also ties in well with considerations of robustness to changing network conditions. We show that our approach can take advantage of redundant network capacity for improved performance and robustness. Moreover, issues of stability, such as those arising from propagation of routing information, are obviated by the fact that each node selects its code independently from the others.

Our results, more specifically, give a lower bound on the probability of error-free transmission for independent or linearly correlated sources, which, owing to the particular form of transfer matrix determinant polynomials, is tighter than the Schwartz-Zippel bound [18] for general polynomials of the same total degree. This bound, which is exponentially dependent on the code length, holds for any feasible set of multicast connections over any network topology (including networks with cycles and link delays). The result is derived using a relation we establish between multicast network coding and bipartite matching, a useful mathematical tool which leads also to an upper bound on field size required for deterministic centralized network coding over general networks. We further give, for acyclic networks, tighter bounds based on more specific network structure, and show the effects of redundancy and link reliability on success probability. For arbitrarily correlated sources, we give error bounds for minimum entropy and maximum a posteriori probability decoding. In the special case of a Slepian-Wolf source network consisting of a link from each source to the receiver, our error exponents reduce to the corresponding results in [4] for linear Slepian-Wolf coding. The latter scenario may thus be considered a degenerate case of network coding.

We illustrate some possible applications with two examples of practical scenarios – distributed settings and online algorithms for networks with dynamically varying connections – in which randomized network coding shows promise of substantial benefits compared to routing.

This paper is a first exploration on randomized network coding, posing more questions that it answers. We do not consider aspects such as resource and energy allocation, but focus on optimally exploiting a given set of resources. There are also many issues surrounding the adaptation of protocols, which generally assume routing, to random coding approaches. We do not address these here, but rather seek to establish that the potential benefits of randomized network coding justify future consideration of protocol compatibility with or adaptation to network codes.

The basic randomized network coding approach requires no coordination among nodes. If we allow for retries to find successful codes, we in effect trade code length

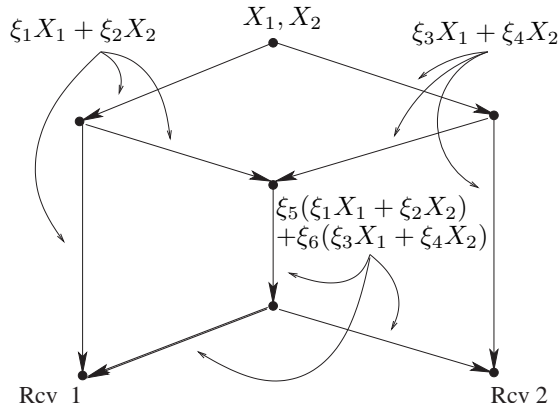


Figure 1: An example of distributed randomized network coding.  $X_1$  and  $X_2$  are the source processes being multicast to the receivers, and the coefficients  $\xi_i$  are randomly chosen elements of a finite field. The label on each link represents the signal being carried on the link.

for some rudimentary coordination. Implementations for various applications may not be completely protocol-free, but the roles and requirements for protocols may be substantially redefined in this new environment.

Portions of this work have appeared in [8], which introduced distributed randomized network coding, [7], which presented connections with bipartite matching/network flows and a new bound on required field size for centralized network coding, [10], which generalized previous results to arbitrary networks and gave tighter bounds for acyclic networks, and [9], on network coding for arbitrarily correlated sources.

## 1.1 Overview

A brief overview of related work is given in Section 1.2. In Section 2, we provide the algebraic model we use in our analyses. Our main results are given in Section 3. These are further developed and discussed in Section 4 linking network coding with bipartite matching and network flows, Section 5 on delay-free networks, Section 6 on general networks with cycles and delay, Section 7 giving tighter bounds for acyclic networks, and Section 8 on arbitrarily correlated sources. We also give examples of practical scenarios in which randomized network coding can be advantageous compared to routing, in Section 9. We present our conclusions and some directions for further work in Section 10.

## 1.2 Related Work

Ahlsvede et al. [1] showed that with network coding, as symbol size approaches infinity, a source can multicast information at a rate approaching the smallest minimum cut between the source and any receiver. Li et al. [16] showed that linear coding with finite symbol size is sufficient for multicast. Koetter and Médard [14] presented an algebraic framework for network coding that extended previous results to arbitrary networks and robust networking, and proved the achievability with time-invariant solutions of the min-cut max-flow bound for networks with delay and cycles. Reference [14] also gave an algebraic characterization of the feasibility of a multicast problem and the validity of a network coding solution in terms of transfer matrices, which we showed in [7] had

equivalent formulations related to bipartite matching and network flows. We used this result in obtaining a tighter upper bound on the required field size than the previous bound of [14], and in our analysis of distributed randomized network coding, introduced in [8]. Concurrent independent work by Sanders et al. [21] and Jaggi et al. [11] considered single-source multicast on acyclic delay-free graphs, showing a similar bound on field size by different means, and giving centralized deterministic and randomized polynomial-time algorithms for finding network coding solutions over a subgraph consisting of flow solutions to each receiver. Lower bounds on coding field size were presented by Rasala Lehman and Lehman [15] and Feder et al. [6]. Reference [6] also gave graph-specific upper bounds based on the number of “clashes” between flows from source to terminals. Dougherty et al. [5] presented results on linear solutions for binary solvable multicast networks, and on non-finite field alphabets. The need for vector coding solutions in some non-multicast problems was considered by Rasala Lehman and Lehman [15], Médard et al. [17] and Riis [20]. Various practical protocols for and experimental demonstrations of randomized network coding [3] and non-randomized network coding [24, 19] have also been presented.

## 2 Model

Our mathematical framework and model is based on that of [14]. A network is represented as a directed graph. Discrete random processes  $X_1, X_2, \dots, X_r$  are observable at one or more source nodes, and there are  $d \geq 1$  receiver nodes. The output processes at a receiver node  $\beta$  are denoted  $Z(\beta, i)$ . The *multicast* connection problem is to transmit all the source processes to each of the receiver nodes.

There are  $\nu$  links in the network. Link  $l$  is an *incident outgoing link* of node  $v$  if  $v = \text{tail}(l)$ , and an *incident incoming link* of  $v$  if  $v = \text{head}(l)$ . We call an incident outgoing link of a source node a *source link* and an incident incoming link of a receiver node a *terminal link*. Edge  $l$  carries the random process  $Y(l)$ . A *path* is a subgraph of the network consisting of a sequence of links  $e_1, \dots, e_k$  such that  $e_i$  is an incident incoming link of  $e_{i+1}$ , and each node is visited at most once.

Edges are assumed to have unit capacity; edges with larger capacities are modeled as parallel edges. Our analysis for the case of independent source processes  $X_i$  assumes that each has unit entropy rate; sources of larger entropy rate are modelled as multiple sources at the same node. For the case of linearly correlated sources, we assume unit source entropy and conditional entropy rates, modeling such sources as pre-specified linear combinations of underlying independent unit entropy rate processes. To simplify the notation in our subsequent development, we denote these underlying independent processes by  $X_1, X_2, \dots, X_r$ . For the case of arbitrarily correlated sources, we assume that the sources have integer bit rates.

The processes  $X_i, Y(l), Z(\beta, i)$  generate binary sequences. We assume that information is transmitted as vectors of bits. The length of the vectors is equal in all transmissions, and all links are assumed to be synchronized with respect to the symbol timing.

We consider linear coding<sup>1</sup>. For independent and linearly correlated sources, operations are carried out on vectors of length  $u$ , viewed as scalar elements of the finite field  $\mathbb{F}_{2^u}$ . For arbitrarily correlated sources, we consider vector operations over  $\mathbb{F}_2$ . This vector coding model can, for given vector lengths, be brought into the scalar algebraic frame-

---

<sup>1</sup>which is sufficient for multicast [16]

work of [14] by conceptually expanding each source into multiple sources and each link into multiple links, such that each new source and link corresponds to one bit in the code vectors. We describe this scalar framework below, and use it to analyze the operation of interior network nodes. Note however that the linear decoding strategies of [14] do not apply when we consider compressible and arbitrarily correlated sources.

In a linear code, the signal  $Y(j)$  on a link  $j$  is a linear combination of processes  $X_i$  generated at node  $v = \text{tail}(j)$  and signals  $Y(l)$  on incident incoming links  $l$ . For the delay-free case, this is represented by the equation

$$Y(j) = \sum_{\{i : X_i \text{ generated at } v\}} a_{i,j} X_i + \sum_{\{l : \text{head}(l) = v\}} f_{l,j} Y(l)$$

and an output process  $Z(\beta, i)$  at receiver node  $\beta$  is a linear combination of signals on its terminal links, represented as

$$Z(\beta, i) = \sum_{\{l : \text{head}(l) = \beta\}} b_{\beta,i,l} Y(l)$$

For multicast on a network with link delays, memory is needed at the receiver nodes, but memoryless operation suffices at all other nodes [14]. We consider unit delay links, modeling links with longer delay as links in series. The corresponding linear coding equations are

$$\begin{aligned} Y_{t+1}(j) &= \sum_{\{i : X_i \text{ generated at } v\}} a_{i,j} X_{it} + \sum_{\{l : \text{head}(l) = v\}} f_{l,j} Y_t(l) \\ Z_{t+1}(\beta, i) &= \sum_{\{l : \text{head}(l) = \beta\}} \sum_{u=0}^{\mu} b_{\beta,i,l_u} Y_{t-u}(l) \end{aligned}$$

where  $\mu$  represents the memory required. These equations, as with the random processes in the network, can be represented algebraically in terms of a delay variable  $D$ :

$$\begin{aligned} Y(j)(D) &= \sum_{\{i : X_i \text{ generated at } v\}} D a_{i,j} X_i(D) + \sum_{\{l : \text{head}(l) = v\}} D f_{l,j} Y(l)(D) \\ Z(\beta, i)(D) &= \sum_{\{l : \text{head}(l) = \beta\}} \left( \sum_{u=0}^{\mu} D^{u+1} b_{\beta,i,l_u} \right) Y(l)(D) \end{aligned}$$

where

$$\begin{aligned} X_i(D) &= \sum_{t=0}^{\infty} X_{i,t} D^t \\ Y(j)(D) &= \sum_{t=0}^{\infty} Y_t(j) D^t, \quad Y_0(j) = 0 \\ Z(\beta, i)(D) &= \sum_{t=0}^{\infty} Z_t(\beta, i) D^t, \quad Z_0(\beta, i) = 0 \end{aligned}$$

The coefficients  $\{a_{i,j}, f_{l,j}, b_{\beta,i,l} \in \mathbb{F}_{2^u}\}$  can be collected into  $r \times \nu$  matrices  $A = (a_{i,j})$  and  $B_\beta = (b_{\beta,i,j})$ , and the  $\nu \times \nu$  matrix  $F = (f_{l,j})$ , whose structure is constrained by the network. For acyclic graphs, we number the links ancestrally, i.e. lower-numbered links

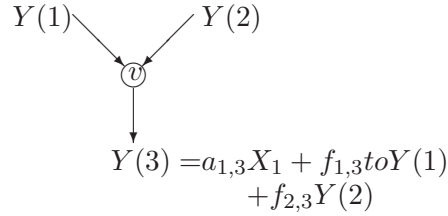


Figure 2: Illustration of linear coding at a node.

upstream of higher-numbered links, so matrix  $F$  is upper triangular with zeros on the diagonal. A triple  $(A, F, B)$ , where

$$B = \begin{bmatrix} B_1 \\ \vdots \\ B_d \end{bmatrix},$$

specifies the behavior of the network, and represents a *linear network code*. We use the following notation:

- $G = \begin{cases} (I - F)^{-1} & \text{in the acyclic delay-free case}^2 \\ (I - DF)^{-1} & \text{in the case with delay}^3 \end{cases}$
- $G_{\mathcal{H}}$  is the submatrix consisting of columns of  $G$  corresponding to links in set  $\mathcal{H}$

Matrix  $AG$  gives the transfer matrix from input processes to signals on each link. For independent or linearly correlated sources, the connection problem is feasible if and only if  $AGB_{\beta}^T$  has full rank for each receiver  $\beta$  [14].

The randomized network coding approach allows for some or all of the coefficients  $\{a_{i,j}, f_{l,j}\}$  to be chosen randomly, as long as the fixed coefficient values preserve feasibility. We denote by  $\eta$  the number of links  $j$  with associated random coefficients  $\{a_{i,j}, f_{l,j}\}$ .

### 3 Main Results

Reference [14] gives the following necessary and sufficient condition for a multicast connection problem with independent or linearly correlated sources to be feasible (or for a particular network code  $(A, F, B)$  to be a valid solution): that for each receiver  $\beta$ , the transfer matrix  $A(I - F)^{-1}B_{\beta}^T$  in the delay-free case, or  $A(I - DF)^{-1}B_{\beta}^T$  in the case with delays, has nonzero determinant. The following result, which we prove in Section 4, is an alternative formulation of this condition that makes a connection with the Edmonds matrix [18] of bipartite matching, and can be used to easily deduce various characteristics of the transfer matrix determinant.

**Theorem 1** (a) *For an acyclic delay-free network, the determinant of the transfer matrix  $M_1 = A(I - F)^{-1}B_{\beta}^T$  for receiver  $\beta$  in a network code  $(A, F, B)$  is equal to*

$$|M_1| = (-1)^{r(\nu+1)} |M_2|$$

where  $M_2 = \begin{bmatrix} A & 0 \\ I - F & B_{\beta}^T \end{bmatrix}$  is the corresponding Edmonds matrix.

<sup>2</sup>The inverse exists since  $F$  is nilpotent.

<sup>3</sup>The inverse exists since the determinant is a nonzero polynomial in  $D$ .

(b) For an arbitrary (possibly cyclic) network with unit delay links, the transfer matrix  $A(I - DF)^{-1}B_\beta^T$  for receiver  $\beta$  in a network code  $(A, F, B)$  is nonsingular if and only if the corresponding Edmonds matrix  $\begin{bmatrix} A & 0 \\ I - DF & B_\beta^T \end{bmatrix}$  is nonsingular.  $\square$

The above result, in illuminating the structure of the transfer matrix determinant, leads to the following two theorems. The first of these is an upper bound on required field size for a feasible network coding problem, which tightens the upper bound of  $q > rd$  given in [14], where  $r$  is the number of processes being transmitted in the network. References [11, 21] independently and concurrently showed the sufficiency of  $\mathbb{F}_q, q \geq d$  for the acyclic delay-free case.

**Theorem 2** *For a feasible multi-source multicast connection problem with independent or linearly correlate sources on an arbitrary network (which may have cycles and delay), there exists a solution in any finite field  $\mathbb{F}_q$  where  $q$  is greater than the number of receivers  $d$ .*  $\square$

Theorem 1 illuminates not only the total degree of the transfer matrix determinant polynomial, but also its particular form, which gives below a bound on randomized coding success rate that is tighter than the Schwartz-Zippel bound of  $1 - d\eta/q$  for general polynomials of the same total degree.

**Theorem 3** *For a feasible multicast connection problem on an arbitrary network with independent or linearly correlated sources, and a network code in which some or all code coefficients are chosen independently and uniformly over all elements of a finite field  $\mathbb{F}_q$  (some coefficients can take fixed values as long as these values preserve feasibility<sup>4</sup>), the probability that all the receivers can decode the source processes is at least  $(1 - d/q)^\eta$  for  $q > d$ , where  $d$  is the number of receivers and  $\eta$  is the number of links with associated randomized coefficients.*  $\square$

The complexity of the code grows as the logarithm of the field size  $q = 2^u$ , since arithmetic operations are performed on codewords of length  $u$ . The error bound is on the order of the inverse of the field size, so the error probability decreases exponentially with the number of codeword bits  $u$ .

The bound of Theorem 3 is very general, applying across all networks with the same number of receivers and the same number of links with independently chosen random linear mappings. Our next goal is to find tighter bounds by taking into account more specific network characteristics. One such bound, for acyclic networks with or without link delays, is based on a connection between randomized coding success probability and network connection feasibility when links are independently deleted with some probability.

**Theorem 4** *For a  $d$ -receiver multicast problem on an acyclic network with independent or linearly correlated sources, the success probability of a random network code in the field of size  $q$  is greater than or equal to the probability that the network connections remain feasible after deleting each link of the original graph with probability  $d/q$ .*

The above bound is useful in cases where analysis of connection feasibility is easier than direct analysis of randomized coding. We apply it to obtain the following result showing how spare network capacity and/or more reliable links allow us to use a smaller field size to surpass a particular success probability.

---

<sup>4</sup>i.e. the result holds for networks where not all nodes perform random coding

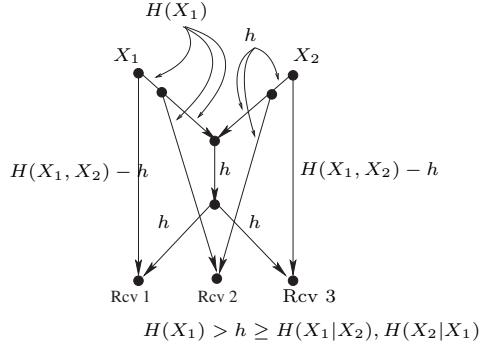


Figure 3: An example network with two correlated sources  $X_1, X_2$  that can be transmitted using distributed randomized network coding. The label on each link represents the capacity of the link.

**Theorem 5** *For a multicast problem on an acyclic network with independent or linearly correlated sources of joint entropy rate  $r$ , and links which fail (are deleted from the network) with probability  $p$ , let  $y$  be the minimum redundancy, i.e. deletion of any  $y$  links in the network preserves feasibility. A random network code transmits all source processes successfully to a particular receiver with probability at least*

$$\sum_{x=r}^{r+y} \binom{r+y}{x} \left(1 - p - \frac{1-p}{q}\right)^{Lx} \left(1 - \left(1 - p - \frac{1-p}{q}\right)^L\right)^{r+y-x}$$

where  $L$  is the longest source-receiver path in the network.

So far we have been considering independent or linearly correlated sources. We next look at the ability of distributed randomized network coding to transmit arbitrarily correlated sources over a network where compression may be required, for example in Figure 3.

The following theorem bounds the probability of successful minimum entropy or maximum a posteriori probability decoding at a receiver, for two sources  $X_1$  and  $X_2$  whose output values in each unit time period are drawn i.i.d. from the same joint distribution  $Q$ . Denote by  $r_i$  the bit rate of source  $X_i$ , and suppose linear coding is done in  $\mathbb{F}_2$  over vectors of  $nr_1$  and  $nr_2$  bits from each source respectively. Let  $m_1$  and  $m_2$  be the minimum cut capacities between the receiver and each of the sources respectively, and let  $m_3$  be the minimum cut capacity between the receiver and both sources. We denote by  $L$  the maximum source-receiver path length.

**Theorem 6** *For distributed randomized network coding of arbitrarily correlated sources  $X_1$  and  $X_2$  over an arbitrary network, the error probability is at most  $\sum_{i=1}^3 p_e^i$ , where*

$$\begin{aligned} p_e^1 &\leq \exp \left\{ -n \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + \left| m_1 \left(1 - \frac{1}{n} \log L\right) - H(X_1 | X_2) \right|^+ \right) + 2^{2r_1 + r_2} \log(n+1) \right\} \\ p_e^2 &\leq \exp \left\{ -n \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + \left| m_2 \left(1 - \frac{1}{n} \log L\right) - H(X_2 | X_1) \right|^+ \right) + 2^{r_1 + 2r_2} \log(n+1) \right\} \\ p_e^3 &\leq \exp \left\{ -n \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + \left| m_3 \left(1 - \frac{1}{n} \log L\right) - H(X_1 X_2) \right|^+ \right) + 2^{2r_1 + 2r_2} \log(n+1) \right\} \end{aligned}$$

The error exponents

$$\begin{aligned}
e^1 &= \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + \left| m_1 \left( 1 - \frac{1}{n} \log L \right) - H(X_1 | X_2) \right|^+ \right) \\
e^2 &= \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + \left| m_2 \left( 1 - \frac{1}{n} \log L \right) - H(X_2 | X_1) \right|^+ \right) \\
e^3 &= \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + \left| m_3 \left( 1 - \frac{1}{n} \log L \right) - H(X_1 X_2) \right|^+ \right)
\end{aligned}$$

generalize the Slepian-Wolf error exponents for linear coding [4]:

$$\begin{aligned}
e^1 &= \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + |R_1 - H(X_1 | X_2)|^+ \right) \\
e^2 &= \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + |R_2 - H(X_2 | X_1)|^+ \right) \\
e^3 &= \min_{X_1, X_2} \left( D(P_{X_1 X_2} \| Q) + |R_1 + R_2 - H(X_1 X_2)|^+ \right)
\end{aligned}$$

where  $R_i$  is the rate of the code for  $X_i$ .

We motivate our interest in randomized network coding with two examples of practical scenarios in which randomized network coding offers useful advantages. The first is in distributed and varying environments where it may be expensive or infeasible to reliably maintain routing state. We consider as a simple example the problem of sending two processes over a grid network from a source node to receiver nodes in various locations unknown to the source and intermediate nodes. We obtain an analytical upper bound on the performance of a distributed randomized routing approach, which is exceeded by our lower bound on the performance of randomized coding for modest code lengths.

Another scenario in which randomized network coding can be advantageous is for online algorithms in dynamic environments. As an illustration, we compare, for dynamically varying connections on randomly generated graphs, distributed randomized coding with an approximate online Steiner tree routing approach from [13] in which, for each transmitter, a tree is selected in a centralized fashion. The high complexity of such a routing scheme requires a simulation-based approach. In practice, an alternative to growing the field size (code length) for improving randomized coding success probability is to allow retrieval of random codes in case of failure. We find that for randomly generated graphs of 8 to 10 nodes, randomized coding with 4-5 bit code lengths and a limit of 3 re-tries per new connection generally performs as well as, and in a non-negligible set of cases, better than the approximate Steiner-tree routing scheme.

These and other results are developed and discussed further in the following sections.

## 4 Connections with Bipartite Matching and Network Flows

Theorem 1 shows the equivalence of the network coding transfer matrix formulation and the Edmonds matrix formulation for checking if a bipartite graph has a perfect matching, which is a classical reduction of the problem of checking the feasibility of an  $s - t$  flow [12]. This latter problem is a special case of network coding, restricted to the binary field and to separate transmission of different signals; it is interesting to find that

the two formulations are equivalent for the more general case of coding in higher order fields.

The combinatorial formulations of Theorem 1 and Theorem 7 below connect network coding with network flows, providing more direct insights into how individual code coefficients affect the overall network code, and making it easier to deduce various characteristics of transfer matrix determinant polynomials without dealing with matrix products and inversions. For instance, Theorem 1 sheds light on the maximum exponent of a variable, the total degree of the polynomial, and its form for networks with linearly correlated sources, leading to Theorems 2 and 3.

*Proof of Theorem 1:*

(a) Note that

$$\begin{bmatrix} I & -A(I-F)^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ I-F & B_\beta^T \end{bmatrix} = \begin{bmatrix} 0 & -A(I-F)^{-1}B_\beta^T \\ I-F & B_\beta^T \end{bmatrix}$$

The first matrix,  $\begin{bmatrix} I & -A(I-F)^{-1} \\ 0 & I \end{bmatrix}$ , has determinant 1. So  $\det\left(\begin{bmatrix} A & 0 \\ I-F & B_\beta^T \end{bmatrix}\right)$  equals  $\det\left(\begin{bmatrix} 0 & -A(I-F)^{-1}B_\beta^T \\ I-F & B_\beta^T \end{bmatrix}\right)$ , which can be expanded as follows:

$$\begin{aligned} & \det\left(\begin{bmatrix} 0 & -A(I-F)^{-1}B_\beta^T \\ I-F & B_\beta^T \end{bmatrix}\right) \\ &= (-1)^{r\nu} \det\left(\begin{bmatrix} -A(I-F)^{-1}B_\beta^T & 0 \\ B_\beta^T & I-F \end{bmatrix}\right) \\ &= (-1)^{r\nu} \det(-A(I-F)^{-1}B_\beta^T) \det(I-F) \\ &= (-1)^{r(\nu+1)} \det(A(I-F)^{-1}B_\beta^T) \det(I-F) \end{aligned}$$

The result follows from observing that  $\det(I-F) = 1$ .

(b) By similar manipulations, we can show that

$$\det\left(\begin{bmatrix} A & 0 \\ I-DF & B_\beta^T \end{bmatrix}\right) = (-1)^{r(\nu+1)} \det(A(I-DF)^{-1}B_\beta^T) \det(I-DF)$$

Since  $\det(I-DF)$  is nonzero, the result follows. ■

*Proof of Theorem 2:* By Theorem 1, the form of the transfer matrix determinant  $|AGB_\beta^T|$  for any receiver  $\beta$  matches the form of the determinant of the Edmonds matrix, in which no variable  $a_{x,j}$ ,  $f_{i,j}$  or  $b_{\beta i,j}$  appears in more than one entry. Thus, no product term in the determinant polynomial contains a variable  $a_{x,j}$ ,  $f_{i,j}$  or  $b_{\beta i,j}$  raised to an exponent greater than 1, and the largest exponent of any variable in the product  $P$  of  $d$  receivers' determinant polynomials is at most  $d$ .

We use an induction argument similar to that in [14] to show that there exists a solution in  $\mathbb{F}_q$ ,  $q > d$ , such that  $P$  is nonzero. Consider one of the variables, denoting it by  $\xi_1$ , and consider  $P$  as a polynomial in the other variables (and  $D$  in the case with delays), with coefficients from  $\mathbb{F}_2[\xi_1]$ . Since these coefficients have maximum degree  $d$ , they are not divisible by  $\xi_1^q - \xi_1$ . Thus,  $\xi_1$  can take some value in  $\mathbb{F}_q$  such that at least one of the coefficients is nonzero. Repeating this procedure for each of the other variables gives the desired result. ■

For acyclic networks, Theorem 7 can be used in place of Theorem 1 to deduce many of the same transfer matrix determinant properties. Furthermore, Theorem 7 allows us to tighten the bound of Theorem 3 for acyclic networks to  $(1 - d/q)^{\eta'}$ , where  $\eta'$  is the maximum number of links with associated random coefficients in any set of links constituting a flow solution from all sources to any receiver. This is used in Section 9 to derive a randomized coding performance bound for grid networks.

**Theorem 7** *A multicast connection problem is feasible (or a particular  $(A, F)$  can be part of a valid solution) if and only if each receiver  $\beta$  has a set  $\mathcal{H}_\beta$  of  $r$  incident incoming links for which*

$$\sum_{\substack{\{\text{disjoint paths } \mathcal{E}_1, \dots, \mathcal{E}_r : \\ \mathcal{E}_i \text{ from outgoing link} \\ l_i \text{ of source } i \text{ to } h_i \in \mathcal{H}_\beta\}}} |A_{\{l_1, \dots, l_r\}}| \prod_{j=1}^r g(\mathcal{E}_j) \neq 0$$

where  $A_{\{l_1, \dots, l_r\}}$  is the submatrix of  $A$  consisting of columns corresponding to links  $\{l_1, \dots, l_r\}$ , and

$$g(\mathcal{E}) = \begin{cases} f_{e_1, e_2} f_{e_2, e_3} \cdots f_{e_{k-1}, e_k} & \text{if } k > 1 \\ 1 & \text{if } k = 1 \end{cases}$$

is the product of gains on the path  $\mathcal{E}$  consisting of links  $e_1 < \dots < e_k$ . The sum is over all flow solutions that transmit all source processes to links in  $\mathcal{H}_\beta$ , each such solution being a set of  $r$  disjoint paths each connecting a different source to a different link in  $\mathcal{H}_\beta$ .  $\square$

*Proof:* Recall that we assume an ancestral numbering for the links of an acyclic graph, i.e. lower-numbered links upstream of higher-numbered links. For  $1 \leq h' \leq h \leq \nu$ , let  $S_{h', h}$  be the set of all sets of integers  $\{e_1, e_2, \dots, e_k\}$  such that  $h' = e_1 < e_2 < \dots < e_k = h$ . Let  $\mathcal{H} = \{h_1, \dots, h_r\}$ , where  $1 \leq h_1 < \dots < h_r \leq \nu$ .

Let  $\underline{a}_j$  and  $\underline{b}_j$  denote column  $j$  of  $A$  and  $AG$  respectively. It follows from the definitions of transfer matrices  $A$  and  $G = I + F + F^2 + \dots$  that  $\underline{c}_h$  can be computed recursively as follows:

$$\underline{c}_1 = \underline{a}_1 \tag{1}$$

$$\underline{c}_h = \sum_{i=1}^{h-1} \underline{c}_i f_{i, h} + \underline{a}_h, \quad h = 2, 3, \dots, \nu \tag{2}$$

to obtain

$$\underline{c}_h = \sum_{i=1}^h \underline{a}_i \sum_{\mathcal{E} \in S_{i, h}} g(\mathcal{E})$$

Using this expression for each column of  $AG_{\mathcal{H}}$  and expanding the determinant linearly in all columns, we obtain

$$\begin{aligned} |AG_{\mathcal{H}}| &= \begin{vmatrix} | & | & | \\ \underline{c}_{h_1} & \cdots & \underline{c}_{h_r} \\ | & | & | \end{vmatrix} \\ &= \sum_{\substack{\{(h'_1, \dots, h'_r) : \\ 1 \leq h'_j \leq h_j \\ h'_i \neq h'_j \forall i \neq j\}}} \begin{vmatrix} | & | & | \\ \underline{a}_{h'_1} & \cdots & \underline{a}_{h'_r} \\ | & | & | \end{vmatrix} \prod_{i=1}^r \sum_{\mathcal{E} \in S_{h'_i, h_i}} g(\mathcal{E}) \end{aligned}$$

$$= \sum_{\substack{\{(h'_1, \dots, h'_r) : \\ 1 \leq h'_j \leq h_j \\ h'_i \neq h'_j \forall i \neq j\}}} \left| \begin{array}{c|c|c} \underline{a}_{h'_1} & \dots & \underline{a}_{h'_r} \\ \hline & & \\ \hline \end{array} \right| \sum_{\substack{\{(\mathcal{E}_1, \dots, \mathcal{E}_r) : \\ \mathcal{E}_j \in S_{h'_j, h_j}\}}} \prod_{j=1}^r g(\mathcal{E}_j).$$

The above expansion does not take into account dependencies among the columns  $\underline{c}_h$ . We can obtain an equivalent expression with fewer terms by using the following alternative sequence of expansions which takes the dependencies into account. We start by expanding the determinant of  $AG_{\mathcal{H}}$  linearly in the  $h_r$ <sup>th</sup> column using (2):

$$\begin{aligned} |AG_{\mathcal{H}}| &= \left| \begin{array}{c|c|c} \underline{c}_{h_1} & \dots & \underline{c}_{h_r} \\ \hline & & \\ \hline \end{array} \right| \\ &= \sum_{\substack{\{i : 1 \leq i < h_r, \\ i \neq h_1, \dots, h_{r-1}\}}} \left| \begin{array}{c|c|c} \underline{c}_{h_1} & \dots & \underline{c}_{h_{r-1}} \\ \hline & & \\ \hline \end{array} \right| f_{i, h_r} + \left| \begin{array}{c|c|c} \underline{c}_{h_1} & \dots & \underline{c}_{h_{r-1}} \\ \hline & & \\ \hline \end{array} \right| \underline{a}_{h_r} \end{aligned}$$

and proceed recursively, expanding each determinant linearly in its column  $\underline{c}_h$  whose index  $h$  is highest, using (2) for  $h > 1$  and (1) for  $h = 1$ . At each expansion stage, the expression for  $AG_{\mathcal{H}}$  is a linear combination of matrix determinants. Each nonzero determinant corresponds to a matrix composed of columns  $\{\underline{a}_{k_1}, \dots, \underline{a}_{k_s}, \underline{c}_{k_{s+1}}, \dots, \underline{c}_{k_r}\}$  such that  $k_i \neq k_j \forall i \neq j$ , and  $\min(k_1, \dots, k_s) > \max(k_{s+1}, \dots, k_r)$ . Its coefficient in the linear combination is a product of terms  $f_{i, h}$  such that  $h > k_{s+1}, \dots, k_r$ , and is of the form  $\prod_{j=1}^r g(\mathcal{E}_j)$  where  $\mathcal{E}_j \in S_{k'_j, h_j}$  and  $\mathcal{E}_i \cap \mathcal{E}_j = \emptyset \forall i \neq j$ . By induction we have that these properties hold for all nonzero determinant terms in the course of the expansion. The expansion terminates when the expression is a linear combination of determinants of the form  $|\underline{a}_{l_1} \dots \underline{a}_{l_r}|$ , at which point we have

$$|AG_{\mathcal{H}}| = \sum_{\substack{\{(h'_1, \dots, h'_r) : \\ 1 \leq h'_j \leq h_j \\ h'_i \neq h'_j \forall i \neq j\}}} \left| \begin{array}{c|c|c} \underline{a}_{h'_1} & \dots & \underline{a}_{h'_r} \\ \hline & & \\ \hline \end{array} \right| \sum_{\substack{\{(\mathcal{E}_1, \dots, \mathcal{E}_r) : \\ \mathcal{E}_j \in S_{h'_j, h_j}, \\ \mathcal{E}_i \cap \mathcal{E}_j = \emptyset \\ \forall i \neq j\}}} \prod_{j=1}^r g(\mathcal{E}_j).$$

The result follows by noting that each set  $\mathcal{E} = \{e_1, e_2, \dots, e_k\}$  such that  $g(\mathcal{E}) \neq 0$  corresponds to a network path consisting of links  $e_1, \dots, e_k$ ; that the condition  $\mathcal{E}_j \cap \mathcal{E}_k = \emptyset$  for all  $j \neq k$ ,  $1 \leq j, k \leq r$  implies that the corresponding paths  $\mathcal{E}_1, \dots, \mathcal{E}_r$  are disjoint; and that  $|\underline{a}_{h'_1} \dots \underline{a}_{h'_r}|$  is nonzero only when links  $h'_j$  are source links carrying  $r$  independent signals.  $\blacksquare$

Having laid some mathematical groundwork by connecting network coding to bipartite matching and network flows, we next consider the performance of randomized network coding.

## 5 Delay-free Networks

We first consider delay-free networks, which must be acyclic in order for all signals to be well-defined. The analysis and results of this section apply also to acyclic networks with delay that are operated in a burst-oriented [16], pipelined [21] or batch-like [3]

fashion, where information may be buffered or delayed so as to be combined with other incoming information from the same batch. A cyclic graph with  $v$  nodes and rate  $r$  may also be converted to an expanded acyclic graph with  $\kappa v$  nodes and rate at least  $(\kappa - v)r$ , communication on which can be emulated over  $\kappa$  time steps on the original cyclic graph [1].

**Lemma 1** *Consider a random network code  $(A, F, B)$  in which  $\eta$  links have associated randomized coefficients. The determinant polynomial of the corresponding Edmonds matrix  $\begin{bmatrix} A & 0 \\ I - F & B_{\beta}^T \end{bmatrix}$  has maximum degree  $\eta$  in variable terms  $\{a_{x,j}, f_{i,j}\}$ , and is linear in each of these variables.*

*Proof:* Each term  $\{a_{x,j}, f_{i,j}, b_{x,j}\}$  appears in only one entry of the Edmonds matrix. Only the  $\eta$  columns corresponding to links carrying random combinations of source processes and/or incoming signals contain variable terms  $\{a_{x,j}, f_{i,j}\}$ .

The determinant can be written as the sum of products of  $r + \nu$  entries, one from each row and column. Each such product is linear in each variable term  $\{a_{x,j}, f_{i,j}\}$ , and has degree at most  $\eta$  in these variables. ■

**Lemma 2** *Let  $P$  be a polynomial in  $\mathbb{F}[\xi_1, \xi_2, \dots]$  of degree less than or equal to  $d\eta$ , in which the largest exponent of any variable  $\xi_i$  is at most  $d$ . Values for  $\xi_1, \xi_2, \dots$  are chosen independently and uniformly at random from  $\mathbb{F}_q \subseteq \mathbb{F}$ . The probability that  $P$  equals zero is at most  $1 - (1 - d/q)^\eta$  for  $d < q$ .*

*Proof:* For any variable  $\xi_1$  in  $P$ , let  $d_1$  be the largest exponent of  $\xi_1$  in  $P$ . Express  $P$  in the form  $P = \xi_1^{d_1} P_1 + R_1$ , where  $P_1$  is a polynomial of degree at most  $d\eta - d_1$  that does not contain variable  $\xi_1$ , and  $R_1$  is a polynomial in which the largest exponent of  $\xi_1$  is less than  $d_1$ . By the Principle of Deferred Decisions, the probability  $\Pr[P = 0]$  is unaffected if we set the value of  $\xi_1$  last after all the other coefficients have been set. If, for some choice of the other coefficients,  $P_1 \neq 0$ , then  $P$  becomes a polynomial in  $\mathbb{F}[\xi_1]$  of degree  $d_1$ . By the Schwartz-Zippel Theorem, this probability  $\Pr[P = 0 | P_1 \neq 0]$  is upper bounded by  $d_1/q$ . So

$$\begin{aligned} \Pr[P = 0] &\leq \Pr[P_1 \neq 0] \frac{d_1}{q} + \Pr[P_1 = 0] \\ &= \Pr[P_1 = 0] \left(1 - \frac{d_1}{q}\right) + \frac{d_1}{q}. \end{aligned} \quad (3)$$

Next we consider  $\Pr[P_1 = 0]$ , choosing any variable  $\xi_2$  in  $P_1$  and letting  $d_2$  be the largest exponent of  $\xi_2$  in  $P_1$ . We express  $P_1$  in the form  $P_1 = \xi_2^{d_2} P_2 + R_2$ , where  $P_2$  is a polynomial of degree at most  $d\eta - d_1 - d_2$  that does not contain variables  $\xi_1$  or  $\xi_2$ , and  $R_2$  is a polynomial in which the largest exponent of  $\xi_2$  is less than  $d_2$ . Proceeding similarly, we assign variables  $\xi_i$  and define  $d_i$  and  $P_i$  for  $i = 3, 4, \dots$  until we reach  $i = k$  where  $P_k$  is a constant and  $\Pr[P_k = 0] = 0$ . Note that  $1 \leq d_i \leq d < q \forall i$  and  $\sum_{i=1}^k d_i \leq d\eta$ , so  $k \leq d\eta$ . Applying Schwartz-Zippel as before, we have for  $k' = 1, 2, \dots, k$

$$\Pr[P_{k'} = 0] \leq \Pr[P_{k'+1} = 0] \left(1 - \frac{d_{k'+1}}{q}\right) + \frac{d_{k'+1}}{q}. \quad (4)$$

Combining all the inequalities recursively, we can show by induction that

$$\Pr[P = 0] \leq \frac{\sum_{i=1}^k d_i}{q} - \frac{\sum_{i \neq j} d_i d_j}{q^2} + \dots + (-1)^{k-1} \frac{\prod_{i=1}^k d_i}{q^k}.$$

Now consider the integer optimization problem

$$\begin{aligned} \text{Maximize } f &= \frac{\sum_{i=1}^{d\eta} d_i}{q} - \frac{\sum_{i \neq j} d_i d_j}{q^2} + \dots + (-1)^{d\eta-1} \frac{\prod_{i=1}^{d\eta} d_i}{q^{d\eta}} \\ \text{subject to } &0 \leq d_i \leq d < q \forall i \in [1, d\eta], \\ &\sum_{i=1}^{d\eta} d_i \leq d\eta, \text{ and } d_i \text{ integer} \end{aligned} \quad (5)$$

whose maximum is an upper bound on  $\Pr[P = 0]$ .

We first consider the problem obtained by relaxing the integer condition on the variables  $d_i$ . Let  $\underline{d}^* = \{d_1^*, \dots, d_{d\eta}^*\}$  be an optimal solution.

For any set  $S_h$  of  $h$  distinct integers from  $[1, d\eta]$ , let  $f_{S_h} = 1 - \frac{\sum_{i \in S_h} d_i}{q} + \frac{\sum_{i, j \in S_h, i \neq j} d_i d_j}{q^2} - \dots + (-1)^h \frac{\prod_{i \in S_h} d_i}{q^h}$ . We can show by induction on  $h$  that  $0 < f_{S_h} < 1$  for any set  $S_h$  of  $h$  distinct integers in  $[1, d\eta]$ .

If  $\sum_{i=1}^{d\eta} d_i^* < d\eta$ , then there is some  $d_i^* < d$ , and there exists a feasible solution  $\underline{d}$  such that  $d_i = d_i^* + \epsilon$ ,  $\epsilon > 0$ , and  $d_h = d_h^*$  for  $h \neq i$ , which satisfies

$$f(\underline{d}) - f(\underline{d}^*) = \frac{\epsilon}{q} \left( 1 - \frac{\sum_{h \neq i} d_h^*}{q} + \dots + (-1)^{d\eta-1} \frac{\prod_{h \neq i} d_h^*}{q^{d\eta-1}} \right).$$

This is positive, contradicting the optimality of  $\underline{d}^*$ .

Next suppose  $0 < d_i^* < d$  for some  $d_i^*$ . Then there exists some  $d_j^*$  such that  $0 < d_j^* < d$ , since if  $d_j^* = 0$  or  $d$  for all other  $j$ , then  $\sum_{i=1}^{d\eta} d_i^* \neq d\eta$ . Assume without loss of generality that  $0 < d_i^* \leq d_j^* < d$ . Then there exists a feasible vector  $\underline{d}$  such that  $d_i = d_i^* - \epsilon$ ,  $d_j = d_j^* + \epsilon$ ,  $\epsilon > 0$ , and  $d_h = d_h^* \forall h \neq i, j$ , which satisfies

$$f(\underline{d}) - f(\underline{d}^*) = - \left( \frac{(d_i^* - d_j^*)\epsilon - \epsilon^2}{q^2} \right) \left( 1 - \frac{\sum_{h \neq i, j} d_h^*}{q} - \dots + (-1)^{d\eta-2} \frac{\prod_{h \neq i, j} d_h^*}{q^{d\eta-2}} \right).$$

This is again positive, contradicting the optimality of  $\underline{d}^*$ .

Thus,  $\sum_{i=1}^{d\eta} d_i^* = d\eta$ , and  $d_i^* = 0$  or  $d$ . So exactly  $\eta$  of the variables  $d_i^*$  are equal to  $d$ . Since the optimal solution is an integer solution, it is also optimal for the integer program (5). The corresponding optimal  $f = \eta \frac{d}{q} - \binom{\eta}{2} \frac{d^2}{q^2} + \dots + (-1)^{\eta-1} \frac{d^\eta}{q^\eta} = 1 - \left(1 - \frac{d}{q}\right)^\eta$ . ■

*Proof of Theorem 3 for delay-free networks:* To check if a network code  $(A, F, B)$  transmits all source processes to receiver  $\beta$ , it suffices to check that the determinant of the corresponding Edmonds matrix is nonzero (Theorem 1). This determinant, which we denote by  $P_\beta$ , is a polynomial linear in each variable  $\{a_{x,j}, f_{i,j}\}$ , with total degree at most  $\eta$  in these variables (Lemma 1). The product  $\prod_\beta P_\beta$  for  $d$  receivers is, accordingly, a polynomial in  $\{a_{x,j}, f_{i,j}\}$  of total degree at most  $d\eta$ , and in which the largest exponent of each of these variables is at most  $d$ .

Recall from the discussion of our model in Section 2 that linearly correlated sources can be viewed as pre-specified linear combinations of underlying independent unit entropy

rate processes. Unlike the independent sources case where each nonzero entry of the  $A$  matrix can be set independently, in this case there are linear dependencies among the entries. The columns of the  $A$  matrix are linear functions  $\sum_k \alpha_j^k \underline{x}_j^k$  of column vectors  $\underline{x}_j^k$  that represent the composition of the source processes at  $\text{tail}(j)$  in terms of underlying independent processes. In distributed randomized coding, the variables  $\alpha_j^k$  are chosen independently and uniformly at random over  $\mathbb{F}_q$ .

It can be seen from Lemma 1 that for any particular  $j$ , each product term in the polynomial  $P_\beta$  for any receiver  $\beta$  contains at most one variable  $a_{i,j} = \sum_k \alpha_j^k v_{i,j}^k$ .  $P_\beta$  is thus linear in the variables  $\alpha_j^k$ , and also in variables  $f_{i,j}$ , which are unaffected by the source correlations. So any variable in the product of  $d$  such polynomials has maximum exponent  $d$ .

Applying Lemma 2 gives us the required bound.

For the single-receiver case, the bound is attained for a network consisting only of links forming a single set of  $r$  disjoint source-receiver paths. ■

These results for acyclic delay-free networks can be generalized to arbitrary networks, as we next show.

## 6 General Networks with Delays

In this section we consider general networks which may have cycles and link delays. As noted earlier, acyclic networks with delay may be operated in a burst-oriented/batch-like fashion which renders the analysis similar to that for acyclic delay-free networks. Here we consider the general cyclic case without buffering, where information is continuously injected into the network. The coefficients of the linear combinations of signals on each link then become polynomials in a delay variable, instead of scalars. The number of terms of these polynomials that must be sent, and the memory required at the receivers, depend on the number of links involved in cycles (memory registers) in the network. For less frequently changing networks, one efficient way to communicate the code coefficients to the receivers following a change is to have a phase in which the sources send a canonical basis through the network.

**Lemma 3** *The determinant polynomial of the Edmonds matrix  $\begin{bmatrix} A & 0 \\ I - DF & B_\beta^T \end{bmatrix}$  associated with a network code  $(A, F, B)$  in a network with delay is a polynomial in delay variable  $D$ , whose coefficients have maximum degree  $\eta$  in variables  $\{a_{x,j}, f_{i,j}\}$ , and are linear in each variable  $\{a_{x,j}, f_{i,j}\}$ .*

*Proof:* The proof is analogous to that of the corresponding result (Lemma 1) for delay-free graphs. ■

*Proof of Theorem 3 for general networks with delay:* To check if a network code  $(A, F, B)$  transmits all source processes to receiver  $\beta$ , it suffices to check that the determinant of the corresponding Edmonds matrix is nonzero (Theorem 1). This determinant is a polynomial in delay variable  $D$ , whose coefficients are linear in each variable  $\{a_{x,j}, f_{i,j}\}$  and have degree at most  $\eta$  in these variables (Lemma 3). The rest of the proof is analogous to that of the corresponding result for acyclic delay-free graphs given in the previous section. ■

While these results hold very generally, they perforce do not take advantage of the particular network structure. However, it is intuitive that redundancy or spare resources

in the network should improve the performance of randomized network coding. The next section presents tighter, more specialized bounds for acyclic networks.

## 7 Connections with Link Reliability and Tighter Bounds for Acyclic Graphs

In this section we prove Theorem 4 relating network coding performance and network connection feasibility in the case of unreliable links, which is used in the proof of Theorem 5 quantifying the benefit of spare capacity and effect of unreliable links. While our results in previous sections have all extended to networks with cycles with the introduction of link delays, our proof of Theorem 4 assumes an acyclic network, with or without delays. We have not proven or disproven whether these results extend to networks with cycles.

**Lemma 4** *Consider any link  $j$ . Let  $\underline{v}_i \in (\mathbb{F}_q[D])^r$  be the vector of source coefficients associated with the  $i^{\text{th}}$  input to link  $j$ , and let  $Y(j) = \sum_i Df_i \underline{v}_i$  be the vector associated with link  $j$ . Consider a number of sets  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_n$  each consisting of  $d'$  arbitrary rank- $(r-1)$  matrices in  $(\mathbb{F}_q[D])^{r \times (r-1)}$ , such that for each matrix in  $\mathcal{S}_k$ ,  $1 \leq k \leq n$ , link  $j$  has among its inputs a signal whose associated vector is not in the column space of the matrix.*

*Denote by  $E_{\mathcal{S}_k, j}$  the event that adding  $Y(j)$  as an additional column to each of the matrices in  $\mathcal{S}_k$  gives a full rank matrix. If coefficients  $f_i$  are chosen uniformly and independently from  $\mathbb{F}_q$ , then  $\Pr(\bigcup_{k=1}^n E_{\mathcal{S}_k, j}) \geq 1 - d'/q$ .*

*Proof:* First consider any one of the sets  $\mathcal{S}_{k'}, 1 \leq k' \leq n$ . Each entry of  $Y(j)$  is a polynomial in  $\mathbb{F}_q[D, f_1, f_2, \dots]$  that is linear in coefficients  $f_i$ . The determinant of an  $r \times r$  matrix which has  $Y(j)$  as one of its columns, and whose  $r-1$  other columns are independent of coefficients  $f_i$ , is thus linear in coefficients  $f_i$ . The product of  $d'$  such determinants has maximum degree  $d'$  in coefficients  $f_i$ .

By the Schwartz-Zippel Theorem, this product is nonzero with probability at least  $1 - d'/q$ . Thus, we have  $\Pr(E_{\mathcal{S}_{k'}, j}) \geq 1 - d'/q$ , which gives  $\Pr(\bigcup_{k=1}^n E_{\mathcal{S}_k, j}) \geq 1 - d'/q$ . ■

*Proof of Theorem 4:* Each receiver receives all processes successfully if the submatrix of  $AG$  corresponding to  $r$  of its incident incoming links, or terminal links, has full rank. The connection problem is feasible if and only if each receiver has a set of  $r$  link-disjoint paths, one from each source.

Let  $j$  be the highest-indexed link in an ancestral ordering, where lower-indexed links feed into higher-indexed links. Consider any given signals on all other links. There are three cases:

Case 1: Regardless of the code coefficients for  $j$ , there cannot exist full rank sets of  $r$  terminal links for each receiver.

Case 2: Regardless of the code coefficients for  $j$ , each receiver has a full rank set of  $r$  terminal links.

Case 3: For some choice of code coefficients for link  $j$ , each receiver has a full rank set of  $r$  terminal links, i.e. link  $j$  has among its inputs signals whose associated vectors are not in the column space of the submatrices of  $AG$  corresponding to the other terminal links of one or more receivers. Applying Lemma 4, we see that such a choice is made with probability at least  $1 - d'/q$ , where  $d'$  is the number of receivers downstream of link  $j$ .

In all three cases, the probability that each receiver has a set of  $r$  terminal links with a full rank set of inputs when code coefficients for link  $j$  are chosen randomly is greater than or equal to that in the case where link  $j$  is deleted with probability  $d/q \geq d'/q$ .

We next consider the problem where link  $j$  is deleted with probability  $d/q$ , and random code coefficients are chosen for all other links. From our earlier arguments, the probability that any set of  $r$  undeleted paths to each receiver has a full rank set of inputs is less than or equal to the probability of success in the original network coding problem.

We continue in this fashion, at each stage considering a new problem in which we delete with probability  $d/q$  the next highest-indexed link as well as each previously considered link. Random code coefficients are chosen for all other links. At each stage, for any choice of surviving links among the set of randomly deleted links, the problem is either infeasible, or there exist one or more sets of random coding links incident to undeleted paths to each receiver which, if full rank, preserve feasibility of the problem. The probability that any set of  $r$  undeleted paths to each receiver has a full rank set of inputs is less than or equal to the probability of success in the original network coding problem.

Note that these arguments hold for the case of independent or linearly correlated sources. ■

*Proof of Theorem 5:* For a given network of non-failed links, we find a lower bound by considering the case of linearly correlated sources, which includes the case of independent sources as a special case, and by analyzing the probability that the connections remain feasible when links fail with probability  $1/q$ , which by Theorem 4 gives us a lower bound on network coding success probability. The success probability for a network whose links fail (i.e. are permanently deleted from the network) with probability  $p$  is thus lower bounded by the probability that the connections remain feasible when links fail with probability  $1 - (1 - p)(1 - 1/q)$ .

We show by induction on  $y$  that a network consisting of  $r + y$  disjoint source-receiver paths of length  $L$ , any  $r$  of which can transmit all processes, has a success probability that is less than or equal to that for any  $y$ -redundant network.

Consider a network  $\mathcal{G}_1$  consisting of  $r + y$  disjoint source-receiver paths of length  $L$ , any  $r$  of which can transmit all the processes. Let  $\mathcal{G}_2$  be any other  $y$ -redundant network with source-receiver paths of length at most  $L$ .

For  $i = 1, 2$ , we consider a set  $\mathcal{P}_i$  of links in graph  $\mathcal{G}_i$  forming  $r$  link-disjoint source-receiver paths sufficient to transmit all processes to the receiver. We distinguish two cases:

Case 1: None of the links in  $\mathcal{P}_i$  fail. In this case the connections are feasible.

Case 2: There exists some link  $j_i \in \mathcal{P}_i$  that fails.

Then we have

$$\begin{aligned} \Pr(\text{success}) &= \Pr(\text{case 1}) + \Pr(\text{case 2}) \Pr(\text{success}|\text{case 2}) \\ &= 1 - \Pr(\text{case 2}) (1 - \Pr(\text{success}|\text{case 2})). \end{aligned}$$

Since  $\mathcal{P}_1$  has at least as many links as  $\mathcal{P}_2$ ,  $\Pr(\text{case 2}, i = 1) \geq \Pr(\text{case 2}, i = 2)$ . Thus, if we can show that  $\Pr(\text{success}|\text{case 2}, i = 1) \leq \Pr(\text{success}|\text{case 2}, i = 2)$ , the induction hypothesis  $\Pr(\text{success}|i = 1) \leq \Pr(\text{success}|i = 2)$  follows.

For  $y = 0$ , the hypothesis is true since  $\Pr(\text{success}|\text{case 2}) = 0$  for  $i = 1, 2$ . For  $y > 0$ , in case 2 we can remove link  $j_i$  leaving a  $(y - 1)$ -redundant graph  $\mathcal{G}'_i$ . By the induction hypothesis, the probability of success for  $\mathcal{G}'_1$  is less than or equal to that for  $\mathcal{G}'_2$ .

Thus,  $\mathcal{G}_1$  gives a lower bound on success probability, which is the probability that all links on at least  $r$  of  $r + y$  length- $L$  paths do not fail. The result follows from observing that each path does not fail with probability  $\left((1 - p)(1 - \frac{1}{q})\right)^L$ . ■

Having seen that our results so far hold for linearly correlated sources naturally leads us to consider the Slepian-Wolf problem, where arbitrarily correlated sources must be independently compressed to be communicated, in the context of randomized network coding.

## 8 Arbitrarily Correlated Sources

We consider transmission of arbitrarily correlated sources in a network by linear network coding, and show error bounds on the probability of successful (non-linear) decoding at a receiver. Analogously to Slepian and Wolf [23], we consider the problem of distributed encoding and joint decoding of two sources whose output symbols in each unit time period are drawn i.i.d. from the same joint distribution  $Q$ . The difference is that in our problem, transmission occurs across a network of intermediate nodes that perform linear transformations from their inputs to their outputs. In the special case of a network consisting of a set of parallel links, this reduces to the original Slepian-Wolf problem.

An  $\alpha$ -decoder (which may be a minimum entropy or maximum  $Q$ -probability decoder) [4] at the receiver maps a block of received signals to the corresponding minimum entropy or maximum  $Q$ -probability inputs. We derive the error probability in terms of the block length when all non-receiver nodes independently and randomly choose linear mappings from inputs to outputs.

*Proof of Theorem 6:* Encoding in the network is represented by the transfer matrix  $AG_{\mathcal{T}}$  specifying the mapping from the vector of source signals  $[X_1 \ X_2] \in \mathbb{F}_2^{n(r_1+r_2)}$  to the vector  $\mathbf{z}$  of signals on the set  $\mathcal{T}$  of terminal links incident to the receiver. Our error analysis, using the method of types, is similar to that in [4]. As there, the type  $P_{\mathbf{x}_i}$  of a vector  $\mathbf{x}_i \in \mathbb{F}_2^{nr_i}$  is the distribution on  $\mathbb{F}_2$  defined by the relative frequencies of the elements of  $\mathbb{F}_2$  in  $\mathbf{x}_i$ , and joint types  $P_{\mathbf{x}_1\mathbf{x}_2}$  are analogously defined.

The  $\alpha$ -decoder maps a vector  $\mathbf{z}$  of received signals onto a vector in  $\mathbb{F}_2^{n(r_1+r_2)}$  minimizing  $\alpha(P_{\mathbf{x}_1\mathbf{x}_2})$  subject to  $[\mathbf{x}_1 \ \mathbf{x}_2]AG_{\mathcal{T}} = \mathbf{z}$ . For a minimum entropy decoder,  $\alpha(P_{\mathbf{x}_1\mathbf{x}_2}) \equiv H(P_{\mathbf{x}_1\mathbf{x}_2})$ , while for a maximum  $Q$ -probability decoder,  $\alpha(P_{\mathbf{x}_1\mathbf{x}_2}) \equiv -\log Q^n(\mathbf{x}_1\mathbf{x}_2)$ . We consider three types of errors: in the first type, the decoder has the correct value for  $X_2$  but outputs the wrong value for  $X_1$ ; in the second, the decoder has the correct value for  $X_1$  but outputs the wrong value for  $X_2$ ; in the third, the decoder outputs wrong values for both  $X_1$  and  $X_2$ . The error probability is upper bounded by the sum of the probabilities of the three types of errors,  $\sum_{i=1}^3 p_e^i$ . Defining the sets of types

$$\mathcal{P}_n^i = \begin{cases} \{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \mid \tilde{X}_1 \neq X_1, \tilde{X}_2 = X_2\} & i = 1 \\ \{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \mid \tilde{X}_1 = X_1, \tilde{X}_2 \neq X_2\} & i = 2 \\ \{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \mid \tilde{X}_1 \neq X_1, \tilde{X}_2 \neq X_2\} & i = 3 \end{cases}$$

where  $\tilde{X}_i \in \mathbb{F}_2^{nr_i}$ , and the sets of sequences

$$\begin{aligned} \mathcal{J}_{X_1X_2} &= \{[\mathbf{x}_1 \ \mathbf{x}_2] \in \mathbb{F}_2^{n(r_1+r_2)} \mid P_{\mathbf{x}_1\mathbf{x}_2} = P_{X_1X_2}\} \\ \mathcal{J}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{x}_1\mathbf{x}_2) &= \{[\tilde{\mathbf{x}}_1 \ \tilde{\mathbf{x}}_2] \in \mathbb{F}_2^{n(r_1+r_2)} \mid P_{\tilde{\mathbf{x}}_1\tilde{\mathbf{x}}_2|\mathbf{x}_1\mathbf{x}_2} = P_{\tilde{X}_1\tilde{X}_2|X_1X_2}\} \end{aligned}$$

we have

$$\begin{aligned}
p_e^1 &\leq \sum_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \\ \in \mathcal{P}_n^1: \\ \alpha(P_{\tilde{X}_1X_2}) \leq \\ \alpha(P_{X_1X_2})}} \sum_{\substack{(\mathbf{x}_1, \mathbf{x}_2) \in \\ \mathcal{J}_{X_1X_2}}} Q^n(\mathbf{x}_1\mathbf{x}_2) \Pr\left(\exists(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) \in \mathcal{J}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{x}_1\mathbf{x}_2) \text{ s.t. } [\mathbf{x}_1 - \tilde{\mathbf{x}}_1 \quad \underline{0}]AG_{\mathcal{T}} = \underline{0}\right) \\
&\leq \sum_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \\ \in \mathcal{P}_n^1: \\ \alpha(P_{\tilde{X}_1X_2}) \leq \\ \alpha(P_{X_1X_2})}} \sum_{\substack{(\mathbf{x}_1, \mathbf{x}_2) \in \\ \mathcal{J}_{X_1X_2}}} Q^n(\mathbf{x}_1\mathbf{x}_2) \min \left\{ \sum_{\substack{(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) \in \\ \mathcal{J}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{x}_1\mathbf{x}_2)}} \Pr([\mathbf{x}_1 - \tilde{\mathbf{x}}_1 \quad \underline{0}]AG_{\mathcal{T}} = \underline{0}), 1 \right\}
\end{aligned}$$

Similarly,

$$\begin{aligned}
p_e^2 &\leq \sum_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \\ \in \mathcal{P}_n^2: \\ \alpha(P_{X_1\tilde{X}_2}) \leq \\ \alpha(P_{X_1X_2})}} \sum_{\substack{(\mathbf{x}_1, \mathbf{x}_2) \in \\ \mathcal{J}_{X_1X_2}}} Q^n(\mathbf{x}_1\mathbf{x}_2) \min \left\{ \sum_{\substack{(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) \in \\ \mathcal{J}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{x}_1\mathbf{x}_2)}} \Pr([\underline{0} \quad \mathbf{x}_2 - \tilde{\mathbf{x}}_2]AG_{\mathcal{T}} = \underline{0}), 1 \right\} \\
p_e^3 &\leq \sum_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \\ \in \mathcal{P}_n^3: \\ \alpha(P_{\tilde{X}_1\tilde{X}_2}) \leq \\ \alpha(P_{X_1X_2})}} \sum_{\substack{(\mathbf{x}_1, \mathbf{x}_2) \in \\ \mathcal{J}_{X_1X_2}}} Q^n(\mathbf{x}_1\mathbf{x}_2) \min \left\{ \sum_{\substack{(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) \in \\ \mathcal{J}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{x}_1\mathbf{x}_2)}} \Pr([\mathbf{x}_1 - \tilde{\mathbf{x}}_1 \quad \mathbf{x}_2 - \tilde{\mathbf{x}}_2]AG_{\mathcal{T}} = \underline{0}), 1 \right\}
\end{aligned}$$

where the probabilities are taken over realizations of the network transfer matrix  $AG_{\mathcal{T}}$  corresponding to the random network code. The probabilities

$$\begin{aligned}
P_1 &= \Pr([\mathbf{x}_1 - \tilde{\mathbf{x}}_1 \quad \underline{0}]AG_{\mathcal{T}} = \underline{0}) \\
P_2 &= \Pr([\underline{0} \quad \mathbf{x}_2 - \tilde{\mathbf{x}}_2]AG_{\mathcal{T}} = \underline{0}) \\
P_3 &= \Pr([\mathbf{x}_1 - \tilde{\mathbf{x}}_1 \quad \mathbf{x}_2 - \tilde{\mathbf{x}}_2]AG_{\mathcal{T}} = \underline{0})
\end{aligned}$$

for nonzero  $\mathbf{x}_1 - \tilde{\mathbf{x}}_1, \mathbf{x}_2 - \tilde{\mathbf{x}}_2$  can be calculated for a given network, or bounded in terms of  $n$  and parameters of the network as we will show later.

As in [4], we can apply some simple cardinality bounds

$$\begin{aligned}
|\mathcal{P}_n^1| &< (n+1)^{2^{2r_1+r_2}} \\
|\mathcal{P}_n^2| &< (n+1)^{2^{r_1+2r_2}} \\
|\mathcal{P}_n^3| &< (n+1)^{2^{2r_1+2r_2}} \\
|\mathcal{J}_{X_1X_2}| &\leq \exp\{nH(X_1X_2)\} \\
|\mathcal{J}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{x}_1\mathbf{x}_2)| &\leq \exp\{nH(\tilde{X}_1\tilde{X}_2|X_1X_2)\}
\end{aligned}$$

and the identity

$$Q^n(\mathbf{x}_1\mathbf{x}_2) = \exp\{-n(D(P_{X_1X_2}||Q) + H(X_1X_2))\}, (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{J}_{X_1X_2} \quad (6)$$

to obtain

$$p_e^1 \leq \exp \left\{ -n \min_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \\ \in \mathcal{P}_n^1: \\ \alpha(P_{\tilde{X}_1X_2}) \leq \\ \alpha(P_{X_1X_2})}} \left( D(P_{X_1X_2}||Q) + \left| -\frac{1}{n} \log P_1 - H(\tilde{X}_1|X_1X_2) \right|^+ \right) + 2^{2r_1+r_2} \log(n+1) \right\}$$

$$p_e^2 \leq \exp \left\{ -n \min_{\substack{P_{X_1 \tilde{X}_1 X_2 \tilde{X}_2} \\ \in \mathcal{P}_n^2; \\ \alpha(P_{X_1 \tilde{X}_2}) \leq \\ \alpha(P_{X_1 X_2})}} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_2 - H(\tilde{X}_2 | X_1 X_2) \right|^+ \right) + 2^{r_1+2r_2} \log(n+1) \right\}$$

$$p_e^3 \leq \exp \left\{ -n \min_{\substack{P_{X_1 \tilde{X}_1 X_2 \tilde{X}_2} \\ \in \mathcal{P}_n^3; \\ \alpha(P_{\tilde{X}_1 \tilde{X}_2}) \leq \\ \alpha(P_{X_1 X_2})}} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(\tilde{X}_1 \tilde{X}_2 | X_1 X_2) \right|^+ \right) + 2^{2r_1+2r_2} \log(n+1) \right\},$$

where the exponents and logs are taken to base 2.

For the minimum entropy decoder, we have

$$\alpha(P_{\tilde{X}_1 \tilde{X}_2}) \leq \alpha(P_{X_1 X_2}) \Rightarrow \begin{cases} H(\tilde{X}_1 | X_1 X_2) \leq H(\tilde{X}_1 | X_2) \leq H(X_1 | X_2) & \text{for } X_2 = \tilde{X}_2 \\ H(\tilde{X}_2 | X_1 X_2) \leq H(\tilde{X}_2 | X_1) \leq H(X_2 | X_1) & \text{for } X_1 = \tilde{X}_1 \\ H(\tilde{X}_1 \tilde{X}_2 | X_1 X_2) \leq H(\tilde{X}_1 \tilde{X}_2) \leq H(X_1 X_2) \end{cases}$$

which gives

$$p_e^1 \leq \exp \left\{ -n \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(X_1 | X_2) \right|^+ \right) + 2^{2r_1+r_2} \log(n+1) \right\} \quad (7)$$

$$p_e^2 \leq \exp \left\{ -n \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_2 - H(X_2 | X_1) \right|^+ \right) + 2^{r_1+2r_2} \log(n+1) \right\} \quad (8)$$

$$p_e^3 \leq \exp \left\{ -n \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(X_1 X_2) \right|^+ \right) + 2^{2r_1+2r_2} \log(n+1) \right\} \quad (9)$$

We next show that these bounds also hold for the maximum  $Q$ -probability decoder, for which, from (6),

$$\alpha(P_{\tilde{X}_1 \tilde{X}_2}) \leq \alpha(P_{X_1 X_2}) \Rightarrow D(P_{\tilde{X}_1 \tilde{X}_2} \| Q) + H(\tilde{X}_1 \tilde{X}_2) \leq D(P_{X_1 X_2} \| Q) + H(X_1 X_2). \quad (10)$$

For  $i = 1$ ,  $\tilde{X}_2 = X_2$ , and (10) gives

$$D(P_{\tilde{X}_1 X_2} \| Q) + H(\tilde{X}_1 | X_2) \leq D(P_{X_1 X_2} \| Q) + H(X_1 | X_2). \quad (11)$$

We show that

$$\begin{aligned} & \min_{\substack{P_{X_1 \tilde{X}_1 X_2 \tilde{X}_2} \in \mathcal{P}_n^1 : \\ \alpha(P_{\tilde{X}_1 \tilde{X}_2}) \leq \alpha(P_{X_1 X_2})}} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(\tilde{X}_1 | X_1 X_2) \right|^+ \right) \\ & \geq \min_{\substack{P_{X_1 \tilde{X}_1 X_2 \tilde{X}_2} \in \mathcal{P}_n^1 : \\ \alpha(P_{\tilde{X}_1 \tilde{X}_2}) \leq \alpha(P_{X_1 X_2})}} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(\tilde{X}_1 | X_2) \right|^+ \right) \\ & \geq \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(X_1 | X_2) \right|^+ \right) \end{aligned}$$

by considering two possible cases for any  $X_1, \tilde{X}_1, X_2$  satisfying (11):

Case 1:  $-\frac{1}{n} \log P_1 - H(X_1|X_2) < 0$ . Then

$$\begin{aligned} & D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(\tilde{X}_1|X_2) \right|^+ \\ & \geq D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(X_1|X_2) \right|^+ \\ & \geq \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(X_1|X_2) \right|^+ \right) \end{aligned}$$

Case 2:  $-\frac{1}{n} \log P_1 - H(X_1|X_2) \geq 0$ . Then

$$\begin{aligned} & D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(\tilde{X}_1|X_2) \right|^+ \\ & \geq D(P_{X_1 X_2} \| Q) + \left( -\frac{1}{n} \log P_1 - H(\tilde{X}_1|X_2) \right) \\ & \geq D(P_{\tilde{X}_1 X_2} \| Q) + \left( -\frac{1}{n} \log P_1 - H(X_1|X_2) \right) \text{ by (11)} \\ & = D(P_{\tilde{X}_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(X_1|X_2) \right|^+ \end{aligned}$$

which gives

$$\begin{aligned} & D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(\tilde{X}_1|X_2) \right|^+ \\ & \geq \frac{1}{2} \left[ D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(\tilde{X}_1|X_2) \right|^+ \right. \\ & \quad \left. + D(P_{\tilde{X}_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(X_1|X_2) \right|^+ \right] \\ & \geq \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_1 - H(X_1|X_2) \right|^+ \right). \end{aligned}$$

A similar proof holds for  $i = 2$ .

For  $i = 3$ , we show that

$$\begin{aligned} & \min_{\substack{P_{X_1 \tilde{X}_1 X_2 \tilde{X}_2} \in \mathcal{P}_n^3 : \\ \alpha(\tilde{P}_{\tilde{X}_1 \tilde{X}_2}) \leq \alpha(P_{X_1 X_2})}} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(\tilde{X}_1 \tilde{X}_2 | X_1 X_2) \right|^+ \right) \\ & \geq \min_{\substack{P_{X_1 \tilde{X}_1 X_2 \tilde{X}_2} \in \mathcal{P}_n^3 : \\ \alpha(\tilde{P}_{\tilde{X}_1 \tilde{X}_2}) \leq \alpha(P_{X_1 X_2})}} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(\tilde{X}_1 \tilde{X}_2) \right|^+ \right) \\ & \geq \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(X_1 X_2) \right|^+ \right) \end{aligned}$$

by considering two possible cases for any  $X_1, \tilde{X}_1, X_2, \tilde{X}_2$  satisfying (10):

Case 1:  $-\frac{1}{n} \log P_3 - H(X_1 X_2) < 0$ . Then

$$\begin{aligned} & D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(\tilde{X}_1 \tilde{X}_2) \right|^+ \\ & \geq D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(X_1 X_2) \right|^+ \\ & \geq \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(X_1 X_2) \right|^+ \right) \end{aligned}$$

Case 2:  $-\frac{1}{n} \log P_3 - H(X_1 X_2) \geq 0$ . Then

$$\begin{aligned} & D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(\tilde{X}_1 \tilde{X}_2) \right|^+ \\ & \geq D(P_{X_1 X_2} \| Q) + \left( -\frac{1}{n} \log P_3 - H(\tilde{X}_1 \tilde{X}_2) \right) \\ & \geq D(P_{\tilde{X}_1 \tilde{X}_2} \| Q) + \left( -\frac{1}{n} \log P_3 - H(X_1 X_2) \right) \text{ by (10)} \\ & = D(P_{\tilde{X}_1 \tilde{X}_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(X_1 X_2) \right|^+ \end{aligned}$$

which gives

$$\begin{aligned} & D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(\tilde{X}_1 \tilde{X}_2) \right|^+ \\ & \geq \frac{1}{2} \left[ D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(\tilde{X}_1 \tilde{X}_2) \right|^+ \right. \\ & \quad \left. + D(P_{\tilde{X}_1 \tilde{X}_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(X_1 X_2) \right|^+ \right] \\ & \geq \min_{X_1 X_2} \left( D(P_{X_1 X_2} \| Q) + \left| -\frac{1}{n} \log P_3 - H(X_1 X_2) \right|^+ \right). \end{aligned}$$

Next we bound the probabilities  $P_i$  in terms of  $n$  and the network parameters  $m_i, i = 1, 2$ , the minimum cut capacity between the receiver and source  $X_i$ ,  $m_3$ , the minimum cut capacity between the receiver and both sources, and  $L$ , the maximum source-receiver path length. Let  $\mathcal{G}_1, \mathcal{G}_2$ , be subgraphs of graph  $\mathcal{G}$  consisting of all links downstream of sources 1 and 2 respectively, and let  $\mathcal{G}_3$  be equal to  $\mathcal{G}$ . Note that in a random linear network code, any link which has at least one nonzero incoming signal carries the zero signal with probability  $\frac{1}{2^{nc}}$ , where  $c$  is the capacity of the link. This is the same as the probability that a pair of distinct values for the link's inputs are mapped to the same output on the link.

For a given pair of distinct source values, let  $E_l$  be the event that the corresponding inputs to link  $l$  are distinct, but the corresponding values on  $l$  are the same. Let  $E(\tilde{\mathcal{G}})$  be the event that  $E_l$  occurs for some link  $l$  on every source-receiver path in graph  $\tilde{\mathcal{G}}$ .  $P_i$  is then equal to the probability of event  $E(\mathcal{G}_i)$ .

Let  $\mathcal{G}'_i, i = 1, 2, 3$  be the graph consisting of  $m_i$  node-disjoint paths, each consisting of  $L$  links each of unit capacity. We show by induction on  $m_i$  that  $P_i$  is upper bounded by the probability of event  $E(\mathcal{G}'_i)$ .

We let  $\tilde{\mathcal{G}}$  be the graphs  $\mathcal{G}_i, \mathcal{G}'_i, i = 1, 2, 3$  in turn, and consider any particular source-receiver path  $\mathcal{P}_{\tilde{\mathcal{G}}}$  in  $\tilde{\mathcal{G}}$ . We distinguish two cases:

Case 1:  $E_l$  does not occur for any of the links  $l$  on the path  $\mathcal{P}_{\tilde{\mathcal{G}}}$ . In this case the event  $E(\tilde{\mathcal{G}})$  occurs with probability 0.

Case 2: There exists some link  $\hat{l}$  on the path  $\mathcal{P}_{\tilde{\mathcal{G}}}$  for which  $E_l$  occurs.

Thus, we have  $\Pr(E(\tilde{\mathcal{G}})) = \Pr(\text{case 2}) \Pr(E(\tilde{\mathcal{G}})|\text{case 2})$ . Since  $\mathcal{P}_{\mathcal{G}'_i}$  has at least as many links as  $\mathcal{P}_{\mathcal{G}_i}$ ,  $\Pr(\text{case 2 for } \mathcal{G}'_i) \geq \Pr(\text{case 2 for } \mathcal{G}_i)$ . Therefore, if we can show that  $\Pr(E(\mathcal{G}'_i)|\text{case 2}) \geq \Pr(E(\mathcal{G}_i)|\text{case 2})$ , the induction hypothesis  $\Pr(E(\mathcal{G}'_i)) \geq \Pr(E(\mathcal{G}_i))$  follows.

For  $m_i = 1$ , the hypothesis is true since  $\Pr(E(\mathcal{G}'_i)|\text{case 2}) = 1$ . For  $m_i > 1$ , in case 2, removing the link  $\hat{l}$  leaves, for  $\mathcal{G}'_i$ , the effective equivalent of a graph consisting of  $m_i - 1$  node-disjoint length- $L$  paths, and, for  $\mathcal{G}_i$ , a graph of minimum cut at least  $m_i - 1$ . The result follows from applying the induction hypothesis to the resulting graphs.

Thus,  $\Pr(E(\mathcal{G}'_i))$  gives an upper bound on probability  $P_i$ :

$$\begin{aligned} P_i &\leq \left(1 - \left(1 - \frac{1}{2^n}\right)^L\right)^{m_i} \\ &\leq \left(\frac{L}{2^n}\right)^{m_i}. \end{aligned}$$

Substituting this into the error bounds (7)-(9) gives the desired result.  $\blacksquare$

Having established the asymptotic optimality of distributed randomized network coding and presented several theoretical bounds on its performance, we next consider types of network problems for which this approach offers promising advantages over routing.

## 9 Benefits of Randomized Coding Over Routing

Network coding, as a superset of routing, has been shown to offer significant capacity gains for specially constructed networks [21]. Apart from such examples, however, the benefits of centralized network coding over centralized optimal routing have not been as clear.

In this section we consider two types of network scenarios in which distributed randomized coding offers other advantages compared to routing based approaches. We give a theoretical comparison of distributed routing and coding for a simple grid topology. In general, however, an exact theoretical analysis of optimal multicast routing is difficult, as it is closely related to the NP-complete Steiner-tree problem. Thus, we use simulations for more complex routing schemes on randomly generated networks.

### 9.1 Distributed Settings

In networks with large numbers of nodes and/or changing topologies, it may be expensive or infeasible to reliably maintain routing state at network nodes. Distributed randomized routing schemes have been proposed [2, 22] which address this kind of issue. However, not allowing different signals to be combined can impose intrinsic penalties in efficiency compared to using network coding.

Consider for example the problem of sending two processes from a source node to receiver nodes in random unknown locations on a rectangular grid network. Transmission to a particular receiver is successful if the receiver gets two different processes instead of

duplicates of the same process. Suppose we wish to use a distributed transmission scheme that does not involve any coordination among nodes or routing state. The network aims to maximize the probability that any node will receive two distinct messages, by flooding in a way that preserves message diversity, for instance using the following scheme RR (ref Figure 4):

- The source node sends one process in both directions on one axis and the other process in both directions along the other axis.
- A node receiving information on one link sends the same information on its three other links (these are nodes along the grid axes passing through the source node).
- A node receiving signals on two links sends one of the incoming signals on one of its two other links with equal probability, and the other signal on the remaining link.

For comparison, we consider the same rectangular grid problem with the following simple random coding scheme RC (ref Figure 4):

- The source node sends one process in both directions on one axis and the other process in both directions along the other axis.
- A node receiving information on one link sends the same information on its three other links.
- A node receiving signals on two links sends a random linear combination of the source signals on each of its two other links.<sup>5</sup>

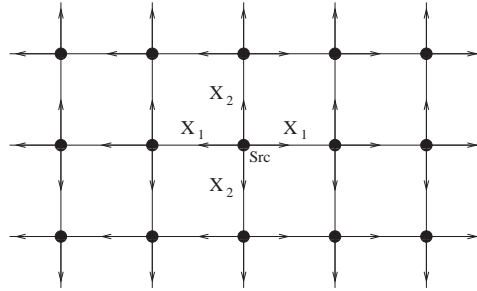


Figure 4: Rectangular grid network with two processes  $X_1$  and  $X_2$  originating at a source node.

**Theorem 8** *For the random routing scheme RR, the probability that a receiver located at grid position  $(x, y)$  relative to the source receives both source processes is at most*

$$\frac{1 + 2^{\lvert x \rvert - \lvert y \rvert + 1} (4^{\min(\lvert x \rvert, \lvert y \rvert) - 1} - 1) / 3}{2^{\lvert x \rvert + \lvert y \rvert - 2}}$$

<sup>5</sup>This simple scheme, unlike the randomized routing scheme RR, leaves out the optimization that each node receiving two linearly independent signals should always send out two linearly independent signals.

Receiver position		(2,2)	(3,3)	(4,4)	(10,10)	(2,3)	(9,10)	(2,4)	(8,10)
RR	actual	0.75	0.672	0.637	-	0.562	-	0.359	-
	upper bound	0.75	0.688	0.672	0.667	0.625	0.667	0.563	0.667
RC	$\mathbb{F}_{2^4}$ lower bound	0.772	0.597	0.461	0.098	0.679	0.111	0.597	0.126
	$\mathbb{F}_{2^6}$ lower bound	0.939	0.881	0.827	0.567	0.910	0.585	0.882	0.604
	$\mathbb{F}_{2^8}$ lower bound	0.984	0.969	0.954	0.868	0.977	0.875	0.969	0.882

Table 1: Success probabilities of randomized routing scheme RR and randomized coding scheme RC. The table gives bounds as well as some actual probability values where exact calculations are tractable.

*Proof:* To simplify notation, we assume without loss of generality that the axes are chosen such that the source is at  $(0, 0)$ , and  $0 < x \leq y$ . Let  $E_{x,y}$  be the event that two different signals are received by a node at grid position  $(x, y)$  relative to the source. The statement of the lemma is then

$$\Pr[E_{x,y}] \leq (1 + 2^{y-x+1}(4^{x-1} - 1)/3) / 2^{y+x-2} \quad (12)$$

which we prove by induction.

Let  $Y_{x,y}^h$  denote the signal carried on the link between  $(x-1, y)$  and  $(x, y)$  and let  $Y_{x,y}^v$  denote the signal carried on the link between  $(x, y-1)$  and  $(x, y)$  (ref Figure 5).

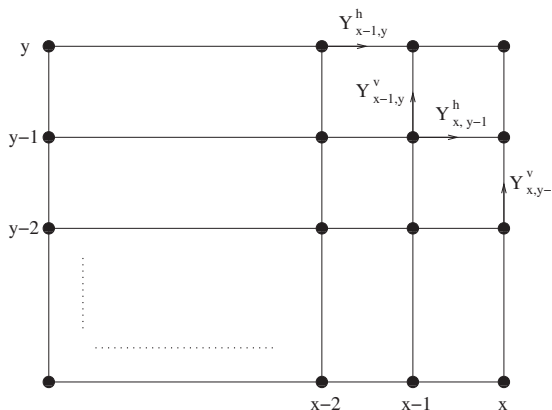


Figure 5: Rectangular grid network.  $Y_{x,y}^h$  denotes the signal carried on the link between  $(x-1, y)$  and  $(x, y)$ , and  $Y_{x,y}^v$  denotes the signal carried on the link between  $(x, y-1)$  and  $(x, y)$ .

Observe that  $\Pr[E_{x,y}|E_{x-1,y}] = 1/2$ , since with probability  $1/2$  node  $(x-1, y)$  transmits to node  $(x, y)$  the signal complementary to whatever signal is being transmitted from node  $(x, y-1)$ . Similarly,  $\Pr[E_{x,y}|E_{x,y-1}] = 1/2$ , so  $\Pr[E_{x,y}|E_{x-1,y} \text{ or } E_{x,y-1}] = 1/2$ .

Case 1:  $E_{x-1,y-1}$

Case 1a:  $Y_{x-1,y}^h \neq Y_{x,y-1}^v$ . With probability  $\frac{1}{2}$ ,  $Y_{x-1,y}^v \neq Y_{x-1,y}^h$ , resulting in  $E_{x,y-1} \cup E_{x-1,y}$ . With probability  $\frac{1}{2}$ ,  $Y_{x,y-1}^v = Y_{x,y-1}^h$ , resulting in  $E_{x,y}$ . So  $\Pr[E_{x,y} | \text{Case 1a}] = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} = \frac{3}{4}$ .

Case 1b:  $Y_{x-1,y}^h = Y_{x,y-1}^v$ . Either  $E_{x,y-1} \cup \overline{E_{x-1,y}}$  or  $\overline{E_{x,y-1}} \cup E_{x-1,y}$ , so  $\Pr[E_{x,y} | \text{Case 1b}] = 1/2$ .

Case 2:  $\overline{E_{x-1,y-1}}$

Case 2a:  $Y_{x-1,y}^h \neq Y_{x,y-1}^v$ . Either  $E_{x,y-1} \cup \bar{E}_{x-1,y}$  or  $\bar{E}_{x,y-1} \cup E_{x-1,y}$ , so  $\Pr[E_{x,y} | \text{Case 2a}] = 1/2$ .

Case 2b:  $Y_{x-1,y}^h = Y_{x,y-1}^v = Y_{x-1,y-1}^h$ . By the assumption of case 2,  $Y_{x,y-1}^v$  is also equal to this same signal, and  $\Pr[E_{x,y} | \text{Case 2b}] = 0$ .

Case 2c:  $Y_{x-1,y}^h = Y_{x,y-1}^v \neq Y_{x-1,y-1}^h$ . Then  $E_{x,y-1}$  and  $E_{x-1,y}$ , so  $\Pr[E_{x,y} | \text{Case 2c}] = 1/2$ .

So

$$\begin{aligned} \Pr[E_{x,y} | E_{x-1,y-1}] &\leq \max(\Pr[E_{x,y} | \text{Case 1a}], \Pr[E_{x,y} | \text{Case 1b}]) \\ &= 3/4 \\ \Pr[E_{x,y} | \bar{E}_{x-1,y-1}] &\leq \max(\Pr[E_{x,y} | \text{Case 2a}], \Pr[E_{x,y} | \text{Case 2b}], \Pr[E_{x,y} | \text{Case 2c}]) \\ &= 1/2 \\ \Pr[E_{x,y}] &\leq \frac{3}{4} \Pr[E_{x-1,y-1}] + \frac{1}{2} \Pr[\bar{E}_{x-1,y-1}] \\ &= \frac{1}{2} + \frac{1}{4} \Pr[E_{x-1,y-1}] \end{aligned}$$

If (12) holds for some  $(x, y)$ , then it also holds for  $(x+1, y+1)$ :

$$\begin{aligned} \Pr[E_{x+1,y+1}] &\leq \frac{1}{2} + \frac{1}{4} \Pr[E_{x,y}] \\ &= \frac{1}{2} + \frac{1}{4} \left( \frac{1 + 2^{y-x+1}(1 + 4 + \dots + 4^{x-2})}{2^{y+x-2}} \right) \\ &= \frac{1 + 2^{y-x+1}(4^x - 1)/3}{2^{y+1+x+1-2}} \end{aligned}$$

Now  $\Pr[E_{1,y'}] = 1/2^{y'-1}$ , since there are  $y' - 1$  nodes,  $(1, 1), \dots, (1, y' - 1)$ , at which one of the signals being transmitted to  $(1, y')$  is eliminated with probability  $1/2$ . Setting  $y' = y - x + 1$  gives the base case which completes the induction. ■

**Theorem 9** *For the random coding scheme RC, the probability that a receiver located at grid position  $(x, y)$  relative to the source can decode both source processes is at least  $(1 - 1/q)^{2(x+y-2)}$ .*

*Proof:* We first establish the degree of the polynomial  $P_\beta$  for a receiver  $\beta$  at  $(x, y)$ , in the indeterminate variables  $f_{i,j}$ . By Theorem 7,  $P_\beta$  is a linear combination of product terms of the form  $a_{1,l_1} a_{2,l_2} f_{i_1,l_3} \dots f_{i_i,l_k}$ , where  $\{l_1, \dots, l_k\}$  is a set of distinct links forming two disjoint paths from the source to the receiver. In the random coding scheme we consider, the only randomized variables are the  $f_{i,j}$  variables at nodes receiving information on two links. The maximum number of such nodes on a source-receiver path is  $x + y - 2$ , so the total degree of  $P_\beta$  is  $2(x + y - 2)$ . Applying the random coding bound of Lemma 2 yields the result. ■

Table 9.1 gives, for various values of  $x$  and  $y$ , the values of the success probability bounds as well as some actual probabilities for routing when  $x$  and  $y$  are small. Note that an increase in grid size from  $3 \times 3$  to  $10 \times 10$  requires only an increase of two in codeword length to obtain success probability lower bounds close to 0.9, which are substantially better than the upper bounds for routing.

## 9.2 Dynamically Varying Connections

Another scenario we considered was multi-source multicast with dynamically varying connections, comparing distributed randomized coding to an approximate online Steiner tree routing approach from [13] in which, for each transmitter, a tree is selected in a centralized fashion. We ran trials each consisting of a number of periods in which any source that was on turned off with probability  $p_o$ , and any source that was off turned on with probability  $p_o$  if it was able to reach all the receivers. A source that was unable to reach all the receivers would be blocked from turning on. For routing, a greedy approximate Steiner tree algorithm from [13] was used. In order for a source to turn on, it would need to find a tree connecting it to all the receivers using spare network capacity unreserved by other sources, and would then reserve capacity corresponding to the tree. A source that turned off freed up its reserved links for new connections. For coding, each network node that tried to turn on would initiate up to three random choices of code coefficients within the network. If the receivers were able to decode the new source in addition to all the sources that were already on, the new source would be allowed to turn on.

The blocking probability and the average throughput were calculated for windows of 250 periods until these measurements reached steady-state. Some results for various randomly generated networks are given in table 2.

These simulations do not attempt to quantify precisely the differences in performance and overhead of randomized coding and online routing. However, they serve as useful illustrations in two ways.

Firstly, they show that the performance of the Steiner tree heuristic is exceeded by randomized coding over a non-negligible set of randomly constructed graphs, indicating that when connections vary dynamically, coding offers advantages that are not circumscribed to a few carefully chosen examples.

Secondly, the simulations illustrate the kinds of field sizes needed in practice for networks with a moderate number of nodes. Field size is important, since it affects memory and complexity requirements. To this end, the simulations use a small field size that still allows randomized coding to generally match the performance of the Steiner heuristic, and to surpass it in networks whose topology makes coding desirable over trees. Our theoretical bounds of previous sections guarantee the optimality of randomized coding for large enough field sizes, but they are tight for worst-case network scenarios. The simulations show the applicability of short network code lengths for moderately-sized networks.

## 10 Conclusion

We have presented a distributed randomized network coding approach which asymptotically achieves optimal capacity in multi-source multicast networks. We have given a general bound on the success probability of such codes for arbitrary networks, showing that error probability decreases exponentially with code length. Our analysis uses connections we make between network coding and network flows/bipartite matching, which also lead to a new bound on required field size for centralized network coding. We have also given tighter bounds for more specific acyclic networks, which show how redundant network capacity and link reliability affect the performance of randomized network coding.

Parameters						Results			
$n$	$r$	$d$	$\rho$	$i$	$p_O$	$b_r$	$t_r$	$b_c$	$t_c$
8	6	1	4	0.5	0.6	1.49	1.55	1.43	1.45
8	6	1	4	0.5	0.6	0.25	2.75	0.24	2.72
9	6	1	3	0.5	0.7	0.75	2.32	0.71	2.29
9	6	1	3	0.5	0.7	2.16	0.86	2.08	0.85
10	6	2	4	0.3	0.6	1.44	0.89	1.34	0.88
10	6	2	4	0.3	0.6	0.22	2.63	0.16	2.77
10	7	3	3	0.5	0.7	1.07	2.39	1.17	2.45
10	7	3	3	0.5	0.7	2.11	1.43	1.80	1.69

Table 2: A sample of results on graphs generated with the following parameters: number of nodes  $n$ , number of sources  $r$ , number of receivers  $d$ , transmission range  $\rho$ , maximum in-degree and out-degree  $i$ .  $b_r$  and  $b_c$  are the rate of blocked connections for routing and coding, respectively, and  $t_r$  and  $t_c$  are the corresponding throughputs. For each trial,  $n$  nodes were scattered uniformly over a unit square. To create an acyclic graph we ordered the nodes by their  $x$ -coordinate and chose the direction of each link to be from the lower numbered to the higher numbered node. Any pair of nodes within a distance  $\rho$  of each other was connected by a link, provided this did not violate the degree constraints. The receiver nodes were chosen as the  $d$  highest numbered nodes, and  $r$  source nodes were chosen randomly (with replacement) from among the lower-numbered half of the nodes. The parameter values for the tests were chosen such that the resulting random graphs would in general be connected and able to support some of the desired connections, while being small enough for the simulations to run efficiently. In 11 of the 135 networks tested, randomized coding with a field size of 17 and up to 3 retrievals performed appreciably better than the Steiner heuristic.

Taking a source coding perspective, we have shown that distributed randomized network coding effectively compresses correlated sources within a network, providing error exponents that generalize corresponding results for linear Slepian-Wolf coding.

Finally, two examples of scenarios in which randomized network coding shows benefits over routing approaches have been presented. These examples suggest that the decentralized nature and robustness of randomized network coding can offer significant advantages in settings that hinder optimal centralized network control.

Further work includes extensions to non-uniform code distributions, possibly chosen adaptively or with some rudimentary coordination, to optimize different performance goals. Another question concerns selective placement of randomized coding nodes. The randomized and distributed nature of the approach also leads us naturally to consider applications in network security. It would also be interesting to consider protocol issues for different communication scenarios, and to compare specific coding and routing protocols over a range of performance metrics.

## References

- [1] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46:1204–1216, 2000.
- [2] K. P. Birman, M. Hayden, O. Ozkasap, Z. Xiao, M. Budiu, and Y. Minsky. Bimodal multicast. *ACM Transactions on Computer Systems*, 17(2):41–88, 1999.
- [3] P. A. Chou, Y. Wu, and K. Jain. Practical network coding. In *Proceedings of 41st Annual Allerton Conference on Communication, Control, and Computing*, October 2003.
- [4] I. Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Transactions on Information Theory*, 28, No.4:585–592, 1982.
- [5] R. Dougherty, C. Freiling, and K. Zeger. Linearity and solvability in multicast networks. *submitted to the IEEE Transactions on Information Theory*, 2003.
- [6] M. Feder, D. Ron, and A. Tavor. Bounds on linear codes for network multicast. *Electronic Colloquium on Computational Complexity*, 10(033), 2003.
- [7] T. Ho, D. R. Karger, M. Médard, and R. Koetter. Network coding from a network flow perspective. In *Proceedings of the IEEE International Symposium on Information Theory*, 2003.
- [8] T. Ho, R. Koetter, M. Médard, D. R. Karger, and M. Effros. The benefits of coding over routing in a randomized setting. In *Proceedings of 2003 IEEE International Symposium on Information Theory*, June 2003.
- [9] T. Ho, M. Médard, M. Effros, R. Koetter, and D. R. Karger. Network coding for correlated sources. In *Proceedings of Conference on Information Sciences and Systems*, 2004. To appear.
- [10] T. Ho, M. Médard, J. Shi, M. Effros, and D. R. Karger. On randomized network coding. In *Proceedings of 41st Annual Allerton Conference on Communication, Control, and Computing*, October 2003.

- [11] S. Jaggi, P. Chou, and K. Jain. Low complexity algebraic network codes. In *Proceedings of the IEEE International Symposium on Information Theory*, 2003.
- [12] R. M. Karp, E. Upfal, and A. Wigderson. Constructing a perfect matching is in random nc. *Combinatorica*, 6 (1):35–48, 1986.
- [13] M. S. Kodialam, T. V. Lakshman, and S. Sengupta. Online multicast routing with bandwidth guarantees: a new approach using multicast network flow. In *Measurement and Modeling of Computer Systems*, pages 296–306, 2000.
- [14] R. Koetter and M. Médard. An algebraic approach to network coding. *IEEE/ACM Transactions on Networking*, to appear.
- [15] A. R. Lehman and E. Lehman. Complexity classification of network information flow problems. In *Symposium on Discrete Algorithms*, 2004.
- [16] S.-Y. R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49:371–381, 2003.
- [17] M. Médard, M. Effros, T. Ho, and D. R. Karger. On coding for non-multicast networks. In *Proceedings of 41st Annual Allerton Conference on Communication, Control, and Computing*, October 2003.
- [18] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.
- [19] T. Noguchi, T. Matsuda, and M. Yamamoto. Performance evaluation of new multicast architecture with network coding. *IEICE Transactions on Communication*, E86-B, No.6, June 2003.
- [20] S. Riis. Linear versus non-linear boolean functions in network flow, preprint, November 2003.
- [21] P. Sanders, S. Egner, and L. Tolhuizen. Polynomial time algorithms for network information flow. In *15th ACM Symposium on Parallel Algorithms and Architectures*.
- [22] S. D. Servetto. Constrained random walks on random graphs: Routing algorithms for large scale wireless sensor networks. In *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, 2002.
- [23] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 25:471 – 480, 1973.
- [24] Y. Zhu, B. Li, and J. Guo. Multicast with network coding in application-layer overlay networks. *IEEE Journal on Selected Areas in Communications*, 22(1), 2004.