The Benefits of Coding over Routing in a Randomized Setting

Tracey Ho, Ralf Koetter, Muriel Médard, David R. Karger and Michelle Effros

***Abstract*— We present a novel randomized coding approach for robust, distributed transmission and compression of information in networks. We give a lower bound on the success probability of a random network code, based on the form of transfer matrix determinant polynomials, that is tighter than the Schwartz-Zippel bound for general polynomials of the same total degree. The corresponding upper bound on failure probability is on the order of the inverse of the size of the finite field, showing that it can be made arbitrarily small by coding in a sufficiently large finite field, and that it decreases exponentially with the number of codeword bits.**

We demonstrate the advantage of randomized coding over routing for distributed transmission in rectangular grid networks by giving, in terms of the relative grid locations of a source-receiver pair, an upper bound on routing success probability that is exceeded by a corresponding lower bound on coding success probability for sufficiently large finite fields.

We also show that our lower bound on the success probability of randomized coding holds for linearly correlated sources. This implies that randomized coding effectively compresses linearly correlated information to the capacity of any network cut in a feasible connection problem.

I. Introduction

In this paper we present a novel randomized coding approach for robust, distributed transmission and compression of information in networks, and demonstrate its advantages over routing-based approaches.

It is known that there exist cases where coding over networks enables certain connections that are not possible with just routing [1]. In this paper we investigate the benefits of coding over routing, not in terms of a taxonomy of network connection problems for which coding is necessary, but in a probabilistic, distributed setting. Distributed randomized routing has previously been considered for achieving robustness and path diversity with minimal state [5].

We give a lower bound on the success probability of a random network code, based on the form of transfer matrix determinant polynomials, that is tighter than the Schwartz-Zippel bound for general polynomials of the same total degree. The corresponding upper bound on failure probability is on the order of the inverse of the size of the finite field, showing that it can be made arbitrarily small by coding in a sufficiently large finite field, and that it decreases exponentially with the number of codeword bits. This suggests that random codes are potentially very useful for networks with unknown or changing topologies.

We demonstrate the advantage of randomized coding over routing for distributed transmission of multiple source processes in the case of rectangular grid networks. We provide an upper bound on the routing success probability for a source-receiver pair in terms of their relative grid locations, which is surpassed by the corresponding lower bound for randomized coding in sufficiently large finite fields.

Randomized coding also has connections with distributed data compression. We show that our lower bound on the success probability of randomized coding applies also for linearly correlated sources, which arise naturally in applications such as networks of sensors measuring the additive effects of multiple phenomena and noise. The effect of randomized coding on such sources can be viewed as distributed compression occuring within the network rather than at the sources. For a feasible multicast connection problem (i.e., one for which there exists some coding solution) and a randomized code of sufficient complexity, with high probability the information flowing across any cut will be sufficient to reconstruct the original source processes. In effect, the source information is being compressed to the capacity of any cut that it passes through. This is achieved without the need for any coordination among the source nodes, which is advantageous in distributed environments where such coordination is impossible or expensive.

Finally, we note that this randomized coding approach achieves robustness in a way quite different from traditional approaches. Traditionally, compression is applied at source nodes so as to minimize required transmission

Tracey Ho and Muriel Médard are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, e-mail: {trace, medard}@mit.edu

Ralf Koetter is with the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801, e-mail: koetter@csl.uiuc.edu

David R. Karger is with the Laboratory for Computer Science, Massachusetts Institute of Technology, MA 02139, e-mail: karger@lcs.mit.edu

Michelle Effros is with the Data Compression Laboratory, California Institute of Technology, Pasadena, CA 91125, e-mail: effros@caltech.edu

rate and leave spare network capacity, and the addition of new sources may require re-routing of existing connections. Our approach fully utilizes available or allocated network capacity for maximal robustness, while retaining full flexibility to accommodate changes in network topology or addition of new sources.

The paper is organized as follows: Section II describes our network model, Section III gives the main results, Section IV gives proofs and ancillary results, and Section V concludes the paper with a summary of the results and a discussion of further work.

II. Model

We adopt the model of [3], which represents a network as a directed graph $\mathcal{G}$. Discrete independent random processes $X_1, \ldots, X_r$ are observable at one or more source nodes, and there are $d \geq 1$ receiver nodes. The output processes at a receiver node β are denoted $Z(\beta, i)$. The *multicast* connection problem is to transmit all the source processes to each of the receiver nodes.

There are ν 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 $Y(l)$.

The time unit is chosen such that the capacity of each link is one bit per unit time, and the random processes X_i have a constant entropy rate of one bit per unit time. Edges with larger capacities are modelled as parallel edges, and sources of larger entropy rate are modelled as multiple sources at the same node.

The processes X_i, $Y(l)$, $Z(\beta, i)$ generate binary sequences. We assume that information is transmitted as vectors of bits which are of equal length u, represented as elements in the finite field $\mathbb{F}_{2^u}$. The length of the vectors is equal in all transmissions and all links are assumed to be synchronized with respect to the symbol timing.

In this paper we consider linear coding[1] on acyclic delay-free networks[2]. 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:

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

[1] which is sufficient for multicast [4]

[2] this algebraic framework can be extended to networks with cycles and delay by working in fields of rational functions in a delay variable [3]

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

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

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. A triple (A, F, B), where

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

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

- $G = (I - F)^{-1}$
- $G_{\mathcal{H}}$ is the submatrix consisting of columns of G corresponding to links in set $\mathcal{H}$
- $\underline{a}_j$, $\underline{c}_j$ and $\underline{b}_j$ denote column j of A, AG and B respectively

III. Main results

Reference [3] gives an algorithm for finding a linear coding solution to a given multicast problem, using knowledge of the entire network topology. In applications where communication is limited or expensive, it may be necessary or useful to determine each node's behavior in a distributed manner. We consider a randomized approach in which network nodes independently and randomly choose code coefficients from some finite field $\mathbb{F}_q$. The only management information needed by the receivers is the overall linear combination of source processes present in each of their incoming signals. This information can be maintained, for each signal in the network, as a vector in ${\mathbb{F}_q}^r$ of the coefficients of each of the source processes, and updated by each coding node applying the same linear combinations to the coefficient vectors as to the data.

Our first result gives a lower bound on the success rate of randomized coding over $\mathbb{F}_q$, in terms of the number of receivers and the number of links in the network. Because of the particular form of the product of transfer matrix determinant polynomials, the bound is tighter than the Schwartz-Zippel bound of of $d\nu/q$ for general polynomials of the same total degree.

Theorem 1: For a feasible multicast connection problem 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[3]), the probability that all the receivers can decode the source processes is at least $(1 - d/q)^\nu$ for $q > d$, where d is the number of receivers and ν is the maximum number of links receiving signals with independent randomized coefficients in any set of links constituting a flow solution from all sources to any receiver. □

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 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. For a fixed success probability, the field size needs to be on the order of the number of links ν multiplied by the number of receivers d.

An implication of this result for linearly correlated sources is that for a feasible multicast connection problem and a randomized code of sufficient complexity, with high probability the information passing through any source-receiver cut in the network contains the source information in a form that is compressed (or expanded) to the capacity of the cut.

Unlike random coding, if we consider only routing solutions (where different signals are not combined), then there are network connection problems for which the success probability of distributed routing is bounded away from 1.

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 communication between nodes or routing state (perhaps because of storage or complexity limitations of network nodes, or frequent shifting of receiver nodes). The best the network can aim for is 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 1):

- 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.

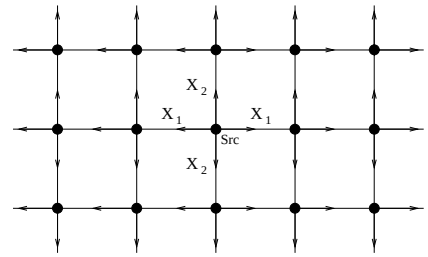

Fig. 1. Rectangular grid network.

Theorem 2: 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

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

□

For comparison, we consider the same rectangular grid problem with the following simple random coding scheme RC (ref Figure 1):

- 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.[4]

Theorem 3: 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 $(1 - 1/q)^{2(x+y-2)}$. □

Table III 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 3×3 to 10×10 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.

IV. Proofs and ancillary results

We make use of the following result from our companion paper [2], which characterizes the feasibility of a multicast connection problem in terms of network flows:

[3] i.e. the result holds for networks where not all nodes perform random coding, or where signals add by superposition on some channels

[4] 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.

TABLE I

SUCCESS PROBABILITIES OF RANDOMIZED ROUTING SCHEME RR AND RANDOMIZED CODING SCHEME RC

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

Theorem 4: A multicast connection problem is feasible (or a particular (A, F) can be part of a valid solution) if and only if each receiver β has a set $\mathcal{H}_\beta$ of r incident incoming links for which

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

where $A_{\{l_1,\ldots,l_r\}}$ is the submatrix of A consisting of columns corresponding to links $\{l_1,\ldots,l_r\}$. The sum is over all flows that transmit all source processes to links in $\mathcal{H}_\beta$, each flow being a set of r disjoint paths each connecting a different source to a different link in $\mathcal{H}_\beta$. □

Corollary 1: The polynomial P_β for each receiver has maximum degree ν and is linear in variables $\{a_{x,j}, f_{i,j}\}$. The product of d such polynomials has maximum degree $d\nu$, and the largest exponent of any variable $\{a_{x,j}, f_{i,j}\}$ is at most d. □

The particular form given in Corollary 1 of the product of determinant polynomials gives rise to a tighter bound on its probability of being zero when its variables take random values from a finite field $\mathbb{F}_q$, as compared to the Schwartz-Zippel bound of $d\nu/q$ for a general $d\nu$-degree multivariate polynomial.

Lemma 1: Let P be a polynomial of degree less than or equal to $d\nu$, in which the largest exponent of any variable is at most d. The probability that P equals zero is at most $1-(1-d/q)^\nu$ for $d < q$.

Proof: For any variable ξ_1 in P, let d_1 be the largest exponent of ξ_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 $d\nu - d_1$ that does not contain variable ξ_1, and R_1 is a polynomial in which the largest exponent of ξ_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 ξ_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 of degree d_1 in ξ_1. By the Schwartz-Zippel Theorem, this probability $\Pr[P = 0 | P_1 \neq 0]$ is upper bounded by d_1/q. So

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

Next we consider $\Pr[P_1 = 0]$, choosing any variable ξ_2 in P_1 and letting d_2 be the largest exponent of ξ_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 $d\nu - d_1 - d_2$ that does not contain variable ξ_2, and R_2 is a polynomial in which the largest exponent of ξ_2 is less than d_2. Proceeding similarly, we assign variables ξ_i and define d_i and P_i for $i = 3, 4, \ldots$ 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 $\sum_{i=1}^{k} d_i \leq d\nu$, so $k \leq d\nu$. Applying Schwartz-Zippel as before, we have for $k' = 1, 2, \ldots, k$

$$Pr[P_{k'} = 0] \leq Pr[P_{k'+1} = 0]\left(1 - \frac{d_{k'+1}}{q}\right) + \frac{d_{k'+1}}{q} \qquad (2)$$

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

$$\begin{aligned}\Pr[P=0] \;\leq\; & \frac{\sum_{i=1}^{k} d_i}{q} - \frac{\sum_{i\neq j} d_i d_j}{q^2} + \ldots \\ & +(-1)^{k-1}\frac{\prod_{i=1}^{k} d_i}{q^k}\end{aligned}$$

where $0 \leq d\nu - \sum_{i=1}^{k} d_i$.

Now consider the integer optimization problem

$$\begin{aligned}\text{Maximize} \quad & f = \frac{\sum_{i=1}^{d\nu} d_i}{q} - \frac{\sum_{i\neq j} d_i d_j}{q^2} + \ldots \\ & +(-1)^{d\nu-1}\frac{\prod_{i=1}^{d\nu} d_i}{q^{d\nu}} \\ \text{subject to} \quad & 0 \leq d_i \leq d < q \,\forall\, i \in [1, d\nu], \\ & \sum_{i=1}^{d\nu} d_i \leq d\nu, \quad \text{and} \quad d_i \text{ integer} \qquad (3)\end{aligned}$$

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

We first consider the non-integer relaxation of the problem. Let $\underline{d}^* = \{d_1^*, \ldots, d_{d\nu}^*\}$ be an optimal solution.

For any set S_h of h distinct integers from $[1, d\nu]$, let $f_{S_h} = 1 - \frac{\sum_{i\in S_h} d_i}{q} + \frac{\sum_{i,j\in S_h, i\neq j} d_i d_j}{q^2} - \ldots +$

$(-1)^h \frac{\prod_{i \in S_h} d_i}{q^h}$. We can show by induction on h that $0 < f_{S_h} < 1$ for any set S_h of h distinct integers in $[1, d\nu]$.

If $\sum_{i=1}^{d\nu} d_i^* < d\nu$, 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

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

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 $\sum_{i=1}^{d\nu} d_i^* \neq d\nu$. 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

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

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

Thus, $\sum_{i=1}^{d\nu} d_i^* = d\nu$, and $d_i^* = 0$ or d. So exactly ν 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 (3). The corresponding optimal $f = \nu\frac{d}{q} - \binom{\nu}{2}\frac{d^2}{q^2} + \ldots + (-1)^{\nu-1}\frac{d^\nu}{q^\nu} = 1 - \left(1 - \frac{d}{q}\right)^\nu$. ■

Proof of Theorem 1: By Corollary 1, the product $\prod_\beta P_\beta$ has degree at most $d\nu$, and the largest exponent of any variable $a_{x,j}$ or $f_{i,j}$ is at most d. These properties still hold if some variables are set to deterministic values which do not make the product identically zero.

Linearly correlated sources can be viewed as prespecified linear combinations of underlying independent 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 $\underline{a}_j$ of the A matrix are linear functions $\underline{a}_j = \sum_k \alpha_j^k \underline{v}_j^k$ of column vectors $\underline{v}_j^k$ that represent the composition of the source processes at tail(j) in terms of the underlying independent processes: Variables α_j^k in column $\underline{a}_j$ can be set independently of variables $\alpha_{j'}^k$ in other columns $\underline{a}_{j'}$. It can be seen from Theorem 4 that for any particular j, each product term in the polynomial P_β for any receiver β contains at most one variable $a_{i,j} = \sum_k \alpha_j^k v_{i,j}^k$. P_β is thus linear in the variables α_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 1 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. ■

Proof of Theorem 2: 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

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

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 2).

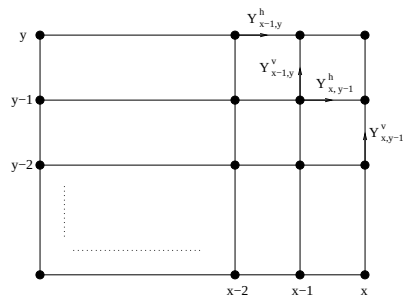

Fig. 2. Rectangular grid network. $Y_{x,y}^h$ denotes the signal carried on the link between $(x-1, y)$ and (x, y), and $Y_{x,y}^v$ denotes the signal carried on the link between $(x, y-1)$ and (x, y).

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: $Y_{x-1,y}^h \neq Y_{x,y-1}^v$. If $Y_{x-1,y}^v \neq Y_{x-1,y}^h$, then $E_{x,y-1} \cup E_{x-1,y}$, and if $Y_{x,y-1}^v = Y_{x,y-1}^h$, then $\overline{E}_{x,y-1} \cup \overline{E}_{x-1,y}$. So $\Pr[E_{x,y}|$ 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 $\Pr[E_{x,y}|$ 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 $E_{x,y-1} \cup \overline{E}_{x-1,y}$ or $\overline{E}_{x,y-1} \cup E_{x-1,y}$, so $\Pr[E_{x,y}|$ 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 $\Pr[E_{x,y}|$ Case 2b$] = 0$.

Case 2c: $Y^h_{x-1,y} = Y^v_{x,y-1} \neq Y^h_{x-1,y-1}$. Then $E_{x,y-1}$ and $E_{x-1,y}$, so $\Pr[E_{x,y}|$ Case 2c$] = 1/2$.

So

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

If Equation 4 holds for some (x, y), then it also holds for $(x + 1, y + 1)$:

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

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

Proof of Theorem 3: We first establish the degree of the transfer matrix determinant polynomial P_β for a receiver β at (x, y), in the indeterminate variables $f_{i,j}$. By Theorem 4, P_β is a linear combination of product terms of the form $a_{1,l_1}a_{2,l_2}f_{i_1,l_3}\ldots f_{i_l,l_k}$, where $\{l_1, \ldots, l_k\}$ 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_β is $2(x+y-2)$. Applying the random coding bound of Lemma 1 yields the result. ■

V. Conclusions and further work

We have presented a novel randomized coding approach for robust, distributed transmission and compression of information in networks, giving an upper bound on failure probability that decreases exponentially with codeword length. We have demonstrated the advantages of randomized coding over randomized routing in rectangular grid networks, by giving an upper bound on the success probability of a randomized routing scheme that is exceeded by the corresponding lower bound for a simple randomized coding scheme in sufficiently large finite fields.

We have also shown that randomized coding has the same success bound for linearly correlated sources, with the implication that randomized coding effectively compresses correlated information to the capacity of any cut that it passes through.

Finally, we note that this randomized coding approach offers a new paradigm for achieving robustness, by spreading information over available network capacity while retaining maximum flexibility to accommodate changes in the network.

Several areas of further research spring from this work. One such area is to study more sophisticated randomized coding schemes on various network topologies, and to compare their performance and management overhead with that of deterministic schemes. Another area would be to extend our results to sources with arbitrary correlations and networks with cycles and delay.

References

[1] R. Ahlswede, N. Cai, S.-Y.R. Li and R.W. Yeung, "Network Information Flow", IEEE-IT, vol. 46, pp. 1204-1216, 2000.

[2] T. Ho, D. R. Karger, M. Médard and R. Koetter, "Network Coding from a Network Flow Perspective ", Submitted to the 2003 IEEE International Symposium on Information Theory.

[3] R. Koetter and M. Médard, "Beyond Routing: An Algebraic Approach to Network Coding", Proceedings of the 2002 IEEE Infocom, 2002.

[4] S.-Y.R. Li and R.W. Yeung, "Linear Network Coding", preprint, 1999.

[5] S. D. Servetto, G. Barrenechea. "Constrained Random Walks on Random Graphs: Routing Algorithms for Large Scale Wireless Sensor Networks", Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, 2002.