Network Coding for Correlated Sources

Tracey Ho, Muriel Médard, Michelle Effros and Ralf Koetter

***Abstract*— We consider the ability of a distributed randomized network coding approach to multicast, to one or more receivers, correlated sources over a network where compression may be required. We give, for two arbitrarily correlated sources in a general network, upper bounds on the probability of decoding error at each receiver, in terms of network parameters. 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 known error exponents for linear Slepian-Wolf coding.**

I. Introduction

The achievable capacity of multicast networks with independent sources employing network coding was given in [1]. Reference [3] showed how to achieve this capacity in a distributed setting using randomized linear codes, and gave error bounds for transmission of independent and linearly correlated sources.

This paper considers linear network coding in the context of a distributed source coding problem, where compression may be required to transmit information from correlated sources over a network to one or more receivers. An example of such a problem is given in Figure 1.

We demonstrate that a distributed randomized network coding approach, an adaptation of that in [3], performs compression when necessary in a multicast network, generalizing known error exponents for linear Slepian-Wolf coding [2] in a natural way. Specifically, for two arbitrarily correlated discrete memoryless sources in a general network, we give error bounds for minimum entropy and maximum a posteriori probability decoding at each receiver. In the special case of a Slepian-Wolf source network consisting of one receiver connected directly by a capacitated link to each source, our error exponents reduce to the corresponding results for linear Slepian-Wolf coding. The latter scenario may thus be considered a degenerate case of network coding.

In this approach, all nodes other than the receiver nodes independently select random linear mappings from vectors of input bits onto vectors of output bits. An illustration is given in Figure 2. The receivers need only know the overall linear combination of source processes in each of their incoming signals. This information can be transmitted to the receivers by sending a canonical basis through the network. The overhead of transmitting these coefficients decreases with increasing length of blocks over which the codes and network state are constant.

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

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

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

This research is supported in part by NSF Grants CCR-0325324 and CCR-0220039, NSF ITR on network coding, University of Illinois subaward #03-25673, Hewlett-Packard 008542-008, and Caltechs Lee Center for Advanced Networking.

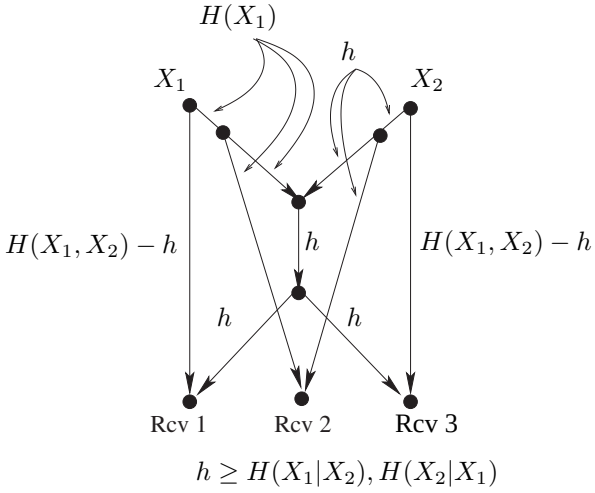

Fig. 1. An example network with two correlated sources X_1, X_2 that can be transmitted using distributed randomized network coding. The label on each link represents the capacity of the link.

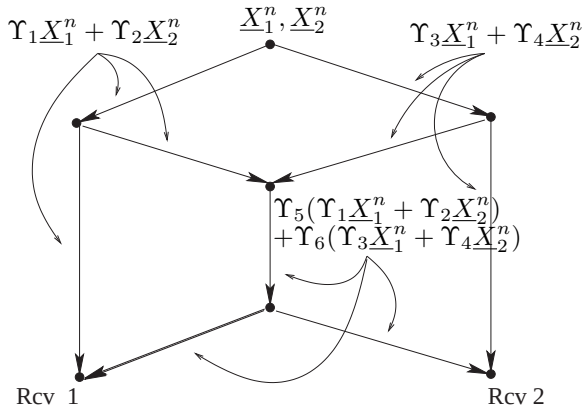

Fig. 2. An example of distributed randomized network coding. $\underline{X}_1^n$ and $\underline{X}_2^n$ are vectors of source bits being multicast to the receivers, and the matrices Υ_i are matrices of random bits. The label on each link represents the signal being carried on the link.

This distributed randomized network coding approach effectively removes or adds redundancy in different parts of the network depending on the available capacity. This is achieved without knowledge of the source entropy rates at interior network nodes. Compression is done simultaneously for multiple receivers in a multicast session.

A. Overview

A brief overview of related work is given in Section I-B. In Section II, we provide the coding and network model we use in our analyses. We present our main result in Section III, and give conclusions and some directions for further work in Section IV.

B. Related Work

Ahlswede et al. [1] showed that with network coding, as symbol size approaches infinity, a source can multicast infor-

mation at a rate approaching the smallest minimum cut between the source and any receiver. Li et al. [7] showed that linear coding with finite symbol size is sufficient for multicast. Koetter and Médard [5] 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. Ho et al. [3] introduced distributed randomized network coding as an efficient, robust way to approach capacity in decentralized settings, giving error bounds for independent and linearly correlated sources. Concurrent independent work by Sanders et al. [10] and Jaggi et al. [4] considered single-source multicast on acyclic delay-free graphs, giving centralized deterministic and randomized polynomial-time algorithms for finding network coding solutions over a subgraph consisting of flow solutions to each receiver. The need for vector coding solutions in some non-multicast problems was considered by Rasala Lehman and Lehman [6], Médard et al. [8] and Riis [9].

II. Model

We represent a network as a directed graph. Discrete memoryless random processes are observable at one or more source nodes, and there are one or more receiver nodes. The *multicast* connection problem is to transmit all the source processes to each of the receiver nodes.

Koetter and Medard [5] give an algebraic framework for network coding which considers unit capacity links and independent unit entropy rate sources. It is assumed 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. Linear coding is used, which is sufficient for multicast [7]. The signal on each link is a scalar linear combination, in a finite field, of incoming links' signals and observable source processes.

We use a slightly different model that lends itself more naturally to consideration of compressible and arbitrarily correlated sources. Our network model allows for links with integer capacities and sources with integer bit rates. Randomized linear network coding is done over vectors of bits in the finite field of size two. This vector coding model can, for given vector lengths, be brought into the scalar algebraic framework of [5] 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 [5] do not apply when we consider compressible and arbitrarily correlated sources.

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 $e_1, \ldots, e_k$ such that e_i is an incident incoming link of e_{i+1}, and each node is visited at most once.

In the scalar linear codes of [5], 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

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

an illustrated in Figure 3. For a network with link delays, each link is assumed to have unit delay; links with longer delay are modeled as links in series. The corresponding linear coding equation is

$$Y_{t+1}(j) = \sum_{\{i\,:\,X_i \text{ generated at } v\}} a_{i,j}X_{it} + \sum_{\{l\ :\ \text{head}(l)\,=\,v\}} f_{l,j}Y_t(l)$$

This equation, as with the random processes in the network, can be represented algebraically in terms of a delay variable D:

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

where

$$X_i(D) = \sum_{t=0}^{\infty} X_{i,t}D^t$$

$$Y(j)(D) = \sum_{t=0}^{\infty} Y_t(j)D^t, \qquad Y_0(j) = 0$$

The coefficients $\{a_{i,j}, f_{l,j} \in \mathbb{F}_{2^u}\}$ can be collected into matrices $A = (a_{i,j})$ and $F = (f_{l,j})$, whose structure is constrained by the network. For acyclic graphs, we number the links ancestrally, i.e. lower-numbered links upstream of higher-numbered links, so matrix F is upper triangular with zeros on the diagonal.

We use the following notation:

- $G = \begin{cases} (I-F)^{-1} \text{ in the acyclic delay-free case}^1 \\ (I-DF)^{-1} \text{ in the case with delay}^2 \end{cases}$
- $G_{\mathcal{H}}$ 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.

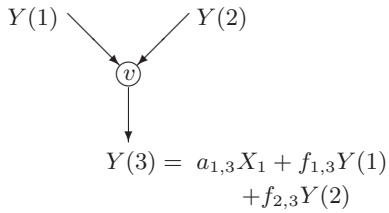

Fig. 3. Illustration of linear coding at a node.

III. Main Result

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 [11], we consider

[2]The inverse exists since F is nilpotent.

[3]The inverse exists since the determinant is a nonzero polynomial in D.

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 α-decoder (which may be a minimum entropy or maximum Q-probability decoder) [2] 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 vector linear mappings from inputs to outputs.

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 1: 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

$$\begin{aligned}
p_e^1 &\leq \exp\Bigg\{ -n \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_1(1-\frac{1}{n}\log L) - H(X_1|X_2) \right|^+ \Bigg) \\
&\quad + 2^{2r_1+r_2}\log(n+1) \Bigg\} \\
p_e^2 &\leq \exp\Bigg\{ -n \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_2(1-\frac{1}{n}\log L) - H(X_2|X_1) \right|^+ \Bigg) \\
&\quad + 2^{r_1+2r_2}\log(n+1) \Bigg\} \\
p_e^3 &\leq \exp\Bigg\{ -n \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_3(1-\frac{1}{n}\log L) - H(X_1X_2) \right|^+ \Bigg) \\
&\quad + 2^{2r_1+2r_2}\log(n+1) \Bigg\}
\end{aligned}$$

The error exponents

$$\begin{aligned}
e^1 &= \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_1(1-\frac{1}{n}\log L) - H(X_1|X_2) \right|^+ \Bigg) \\
e^2 &= \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_2(1-\frac{1}{n}\log L) - H(X_2|X_1) \right|^+ \Bigg) \\
e^3 &= \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_3(1-\frac{1}{n}\log L) - H(X_1X_2) \right|^+ \Bigg)
\end{aligned}$$

generalize the Slepian-Wolf error exponents for linear coding [2]:

$$\begin{aligned}
e^1 &= \min_{X_1,X_2} \left(D(P_{X_1X_2}||Q) + |R_1 - H(X_1|X_2)|^+ \right) \\
e^2 &= \min_{X_1,X_2} \left(D(P_{X_1X_2}||Q) + |R_2 - H(X_2|X_1)|^+ \right) \\
e^3 &= \min_{X_1,X_2} \left(D(P_{X_1X_2}||Q) + |R_1 + R_2 - H(X_1X_2)|^+ \right)
\end{aligned}$$

where R_i is the rate of the code for X_i.

Proof: Encoding in the network is represented by the transfer matrix $AG_{\mathcal{T}}$ specifying the mapping from the vector of source signals $[\ X_1 \quad X_2\ ] \in \mathbb{F}_2^{n(r_1+r_2)}$ 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 [2]. 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 α-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 $[\ \mathbf{x}_1 \quad \mathbf{x}_2\ ]AG_{\mathcal{T}} = \mathbf{z}$. 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

$$\begin{aligned}
\mathcal{T}_{X_1X_2} &= \left\{ [\ \mathbf{x}_1 \quad \mathbf{x}_2\ ] \in \mathbb{F}_2^{n(r_1+r_2)} \;\middle|\; P_{\mathbf{x}_1\mathbf{x}_2} = P_{X_1X_2} \right\} \\
\mathcal{T}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{x}_1\mathbf{x}_2) &= \Big\{ [\ \tilde{\mathbf{x}}_1 \quad \tilde{\mathbf{x}}_2\ ] \in \mathbb{F}_2^{n(r_1+r_2)} \;\Big|\; \\
&\qquad P_{\tilde{\mathbf{x}}_1\tilde{\mathbf{x}}_2\mathbf{x}_1\mathbf{x}_2} = P_{\tilde{X}_1\tilde{X}_2X_1X_2} \Big\}
\end{aligned}$$

we have

$$p_e^1 \leq \sum_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \in \mathcal{P}_n^1 : \\ \alpha(P_{\tilde{X}_1X_2}) \leq \alpha(P_{X_1X_2})}} \sum_{\substack{(\mathbf{x}_1,\mathbf{x}_2) \in \\ \mathcal{T}_{X_1X_2}}} Q^n(\mathbf{x}_1\mathbf{x}_2)$$

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

Similarly,

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

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

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

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 [2], we can apply some simple cardinality bounds

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

and the identity

$$\begin{aligned}
Q^n(\mathbf{x}_1\mathbf{x}_2) &= \exp\{-n(D(P_{X_1X_2}||Q)+H(X_1X_2))\},\\
&\qquad (\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{T}_{X_1X_2} \qquad (1)
\end{aligned}$$

to obtain

$$\begin{aligned}
p_e^1 \le \exp\Bigg\{ &- n \min_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2}\in\mathcal{P}_n^1:\\ \alpha(P_{\tilde{X}_1X_2})\le\alpha(P_{X_1X_2})}} \Bigg(D(P_{X_1X_2}||Q)\\
&+\left|-\frac{1}{n}\log P_1-H(\tilde{X}_1|X_1X_2)\right|^+\Bigg)\\
&+2^{2r_1+r_2}\log(n+1)\Bigg\}\\
p_e^2 \le \exp\Bigg\{ &- n \min_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2}\in\mathcal{P}_n^2:\\ \alpha(P_{X_1\tilde{X}_2})\le\alpha(P_{X_1X_2})}} \Bigg(D(P_{X_1X_2}||Q)\\
&+\left|-\frac{1}{n}\log P_2-H(\tilde{X}_2|X_1X_2)\right|^+\Bigg)\\
&+2^{r_1+2r_2}\log(n+1)\Bigg\}\\
p_e^3 \le \exp\Bigg\{ &- n \min_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2}\in\mathcal{P}_n^3:\\ \alpha(P_{\tilde{X}_1\tilde{X}_2})\le\alpha(P_{X_1X_2})}} \Bigg(D(P_{X_1X_2}||Q)\\
&+\left|-\frac{1}{n}\log P_3-H(\tilde{X}_1\tilde{X}_2|X_1X_2)\right|^+\Bigg)\\
&+2^{2r_1+2r_2}\log(n+1)\Bigg\},
\end{aligned}$$

where the exponents and logs are taken to base 2.

For the minimum entropy decoder, we have

$$\begin{aligned}
&\alpha(P_{\tilde{X}_1\tilde{X}_2})\le\alpha(P_{X_1X_2})\\
&\Rightarrow\begin{cases}
H(\tilde{X}_1|X_1X_2)\le H(\tilde{X}_1|X_2)\le H(X_1|X_2)\\
\qquad\qquad\text{for } X_2=\tilde{X}_2\\
H(\tilde{X}_2|X_1X_2)\le H(\tilde{X}_2|X_1)\le H(X_2|X_1)\\
\qquad\qquad\text{for } X_1=\tilde{X}_1\\
H(\tilde{X}_1\tilde{X}_2|X_1X_2)\le H(\tilde{X}_1\tilde{X}_2)\le H(X_1X_2)
\end{cases}
\end{aligned}$$

which gives

$$\begin{aligned}
p_e^1 \le \exp\Bigg\{ &- n \min_{X_1X_2} \Bigg(D(P_{X_1X_2}||Q)\\
&+\left|-\frac{1}{n}\log P_1-H(X_1|X_2)\right|^+\Bigg)\\
&+2^{2r_1+r_2}\log(n+1)\Bigg\} \qquad (2)\\
p_e^2 \le \exp\Bigg\{ &- n \min_{X_1X_2} \Bigg(D(P_{X_1X_2}||Q)\\
&+\left|-\frac{1}{n}\log P_2-H(X_2|X_1)\right|^+\Bigg)\\
&+2^{r_1+2r_2}\log(n+1)\Bigg\} \qquad (3)\\
p_e^3 \le \exp\Bigg\{ &- n \min_{X_1X_2} \Bigg(D(P_{X_1X_2}||Q)\\
&+\left|-\frac{1}{n}\log P_3-H(X_1X_2)\right|^+\Bigg)\\
&+2^{2r_1+2r_2}\log(n+1)\Bigg\}. \qquad (4)
\end{aligned}$$

We next show that these bounds also hold for the maximum Q-probability decoder, for which, from (1),

$$\begin{aligned}&\alpha(P_{\tilde{X}_1\tilde{X}_2}) \le \alpha(P_{X_1X_2})\\ &\Rightarrow D(P_{\tilde{X}_1\tilde{X}_2}||Q) + H(\tilde{X}_1\tilde{X}_2)\\ &\quad \le D(P_{X_1X_2}||Q) + H(X_1X_2). \end{aligned} \tag{5}$$

For $i=1$, $\tilde{X}_2 = X_2$, and (5) gives

$$D(P_{\tilde{X}_1X_2}||Q) + H(\tilde{X}_1|X_2) \le D(P_{X_1X_2}||Q) + H(X_1|X_2). \tag{6}$$

We show that

$$\begin{aligned}&\min_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2}\in\mathcal{P}_n^1:\\ \alpha(P_{\tilde{X}_1\tilde{X}_2})\le\alpha(P_{X_1X_2})}} \Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_1 - H(\tilde{X}_1|X_1X_2)\right|^+\Big)\\ &\ge \min_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2}\in\mathcal{P}_n^1:\\ \alpha(P_{\tilde{X}_1\tilde{X}_2})\le\alpha(P_{X_1X_2})}} \Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_1 - H(\tilde{X}_1|X_2)\right|^+\Big)\\ &\ge \min_{X_1X_2}\Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_1 - H(X_1|X_2)\right|^+\Big)\end{aligned}$$

by considering two possible cases for any $X_1, \tilde{X}_1, X_2$ satisfying (6):

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

$$\begin{aligned}&D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_1 - H(\tilde{X}_1|X_2)\right|^+\\ &\ge D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_1 - H(X_1|X_2)\right|^+\\ &\ge \min_{X_1X_2}\Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_1 - H(X_1|X_2)\right|^+\Big)\end{aligned}$$

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

$$\begin{aligned}&D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_1 - H(\tilde{X}_1|X_2)\right|^+\\ &\ge\ D(P_{X_1X_2}||Q) - \frac{1}{n}\log P_1 - H(\tilde{X}_1|X_2)\\ &\ge\ D(P_{\tilde{X}_1X_2}||Q) - \frac{1}{n}\log P_1 - H(X_1|X_2) \text{ by (6)}\\ &=\ D(P_{\tilde{X}_1X_2}||Q) + \left|-\frac{1}{n}\log P_1 - H(X_1|X_2)\right|^+\end{aligned}$$

which gives

$$\begin{aligned}&D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_1 - H(\tilde{X}_1|X_2)\right|^+\\ &\ge\ \frac{1}{2}\Big[D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_1 - H(\tilde{X}_1|X_2)\right|^+\\ &\quad + D(P_{\tilde{X}_1X_2}||Q) + \left|-\frac{1}{n}\log P_1 - H(X_1|X_2)\right|^+\Big]\\ &\ge\ \min_{X_1X_2}\Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_1 - H(X_1|X_2)\right|^+\Big).\end{aligned}$$

A similar proof holds for $i=2$.

For $i=3$, we show that

$$\begin{aligned}&\min_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2}\in\mathcal{P}_n^3:\\ \alpha(P_{\tilde{X}_1\tilde{X}_2})\le\alpha(P_{X_1X_2})}} \Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_3 - H(\tilde{X}_1\tilde{X}_2|X_1X_2)\right|^+\Big)\\ &\ge \min_{\substack{P_{X_1\tilde{X}_1X_2\tilde{X}_2}\in\mathcal{P}_n^3:\\ \alpha(P_{\tilde{X}_1\tilde{X}_2})\le\alpha(P_{X_1X_2})}} \Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_3 - H(\tilde{X}_1\tilde{X}_2)\right|^+\Big)\\ &\ge \min_{X_1X_2}\Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_3 - H(X_1X_2)\right|^+\Big)\end{aligned}$$

by considering two possible cases for any $X_1, \tilde{X}_1, X_2, \tilde{X}_2$ satisfying (5):

Case 1: $-\frac{1}{n}\log P_3 - H(X_1X_2) < 0$. Then

$$\begin{aligned}&D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_3 - H(\tilde{X}_1\tilde{X}_2)\right|^+\\ &\ge D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_3 - H(X_1X_2)\right|^+\\ &\ge \min_{X_1X_2}\Big(D(P_{X_1X_2}||Q)\\ &\qquad + \left|-\frac{1}{n}\log P_3 - H(X_1X_2)\right|^+\Big)\end{aligned}$$

Case 2: $-\frac{1}{n}\log P_3 - H(X_1X_2) \ge 0$. Then

$$\begin{aligned}&D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_3 - H(\tilde{X}_1\tilde{X}_2)\right|^+\\ &\ge\ D(P_{X_1X_2}||Q) - \frac{1}{n}\log P_3 - H(\tilde{X}_1\tilde{X}_2)\\ &\ge\ D(P_{\tilde{X}_1\tilde{X}_2}||Q) - \frac{1}{n}\log P_3 - H(X_1X_2) \text{ by (5)}\\ &=\ D(P_{\tilde{X}_1\tilde{X}_2}||Q) + \left|-\frac{1}{n}\log P_3 - H(X_1X_2)\right|^+\end{aligned}$$

which gives

$$D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_3 - H(\tilde{X}_1\tilde{X}_2)\right|^+$$

$$\begin{aligned}
&\geq \frac{1}{2}\Bigg[D(P_{X_1X_2}||Q) + \left|-\frac{1}{n}\log P_3 - H(\tilde{X}_1\tilde{X}_2)\right|^+ \\
&\quad + D(P_{\tilde{X}_1\tilde{X}_2}||Q) + \left|-\frac{1}{n}\log P_3 - H(X_1X_2)\right|^+\Bigg] \\
&\geq \min_{X_1X_2}\Bigg(D(P_{X_1X_2}||Q) \\
&\qquad + \left|-\frac{1}{n}\log P_3 - H(X_1X_2)\right|^+\Bigg).
\end{aligned}$$

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 $\mathcal{G}$. It follows from the algebraic coding model of Section II that in a random linear network code over an arbitrary network, 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:

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

Substituting this into the error bounds (2)-(4) gives the desired results. ■

IV. Conclusion

We have shown that a distributed randomized network coding approach effectively compresses correlated sources within a network, providing error bounds whose exponents generalize corresponding results for linear Slepian-Wolf coding. This randomized network coding approach carries over to any positive number of sources, though we give here a detailed treatment only of the case of two sources.

We give error bounds in terms of minimum cut capacities and maximum source-receiver path length in a network. Bounds in terms of other network parameters, e.g. the number of links upstream of a receiver, or for particular network topologies, can be obtained using similar means.

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.

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] I. Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Transactions on Information Theory*, 28, No.4:585–592, 1982.

[3] 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.

[4] S. Jaggi, P. Chou, and K. Jain. Low complexity algebraic network codes. In *Proceedings of the IEEE International Symposium on Information Theory*, 2003.

[5] R. Koetter and M. Médard. An algebraic approach to network coding. *IEEE/ACM Transactions on Networking*, to appear.

[6] A. R. Lehman and E. Lehman. Complexity classification of network information flow problems. In *Symposium on Discrete Algorithms*, 2004.

[7] S.-Y. R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49:371–381, 2003.

[8] 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.

[9] S. Riis. Linear versus non-linear boolean functions in network flow, preprint, November 2003.

[10] P. Sanders, S. Egner, and L. Tolhuizen. Polynomial time algorithms for network information flow. In *15th ACM Symposium on Parallel Algorithms and Architectures*, pages 286–294, 2003.

[11] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 25:471 – 480, 1973.