

On Network Coding of Independent and Dependent Sources in Line Networks

Mayank Bakshi[†], Michelle Effros[†], WeiHsin Gu[†], Ralf Koetter[‡]

[†] Department of Electrical Engineering
California Institute of Technology
Pasadena, California, CA 91125, USA
Email: {mayank, effros, wgu}@caltech.edu

[‡] Department of Electrical Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
Email: koetter@uiuc.edu

Abstract—We investigate the network coding capacity for line networks. For independent sources and a special class of dependent sources, we fully characterize the capacity region of line networks for all possible demand structures (e.g., multiple unicast, mixtures of unicasts and multicasts, etc.) Our achievability bound is derived by first decomposing a line network into single-demand components and then adding the component rate regions to get rates for the parent network. For general dependent sources, we give an achievability result and provide examples where the result is and is not tight.

I. INTRODUCTION

To date, the field of network coding has focused primarily on finding solutions for families of problems defined by a broad class of networks (e.g., networks representable by directed, acyclic graphs) and a narrow class of demands (e.g., multicast or multiple unicast demands). We here investigate a family of network coding problems defined by a completely general demand structure and a narrow family of networks. Precisely, we give the complete solution to the problem of network coding with independent sources and arbitrary demands on a directed line network. We then generalize that solution to accommodate special cases of dependent sources. Theorem 1 summarizes those results.

Theorem 1: Given an n -node line network \mathcal{N} (shown in Fig. 1) with memoryless sources X_1, \dots, X_n and demands Y_1, \dots, Y_n satisfying $H(Y_i|X_1, \dots, X_i) = 0$ for all $i \in \{1, \dots, n\}$, the rate vector (R_1, \dots, R_{n-1}) is achievable if and only if, for $1 \leq i < n$,

$$R_i \geq \sum_{j=i+1}^n H(Y_j|X_{i+1}, \dots, X_j, Y_{i+1}, \dots, Y_{j-1}), \quad (1)$$

provided one of the following conditions holds:

- Sources X_1, \dots, X_n are independent and each Y_i is a subset of the sources X_1, \dots, X_i .
- Sources X_1, \dots, X_n have arbitrary dependencies and each Y_i is either a constant or the vector (X_1, \dots, X_n) .
- Each source X_i is a (potentially) distinct subset of independent sources W_1, \dots, W_k and each demand Y_i is any subset of those W_1, \dots, W_k that appear in X_1, \dots, X_i .

Lemmas 3, 4, and 5 of Sections III-B, III-C, and III-D give formal statements and proofs of this result under conditions A, B, and C, respectively. Case B is the multicast result of [1]

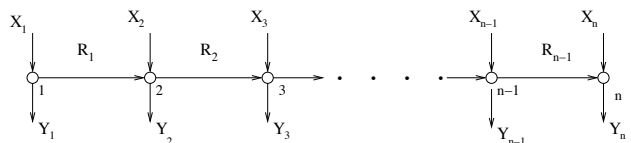


Fig. 1. An n -node line network with sources X_1, \dots, X_n and demands Y_1, \dots, Y_n

generalized to many sources and specialized to line networks. We include a new proof of this result for this special case as it provides an important example in developing our approach.

Central to our discussion is a formal network decomposition described in Section II. The decomposition breaks an arbitrary line network into a family of component line networks. Each component network preserves the original node demands at exactly one node and assumes that all demands at prior nodes in the line network have already been met. (See Fig. 2 for an illustration. Formal definitions follow in Section II.) Sequentially applying the component network solutions in the parent network to meet first the first node's demands, then the second node's demands assuming that the first node's demands are met, and so on, achieves a rate equal to the sum of the rates on the component networks. The given solution always yields an achievability result. The proofs of Lemmas 3, 4, and 5 additionally demonstrate that the given achievability result is tight under each of conditions A, B, and C.

Theorem 2 shows that the achievability result given by our additive solution is tight for an extremely broad class of sources and demands in 3-node line networks. In particular, this result allows arbitrary dependencies between the sources and also allows demands that can be (restricted) functions of those sources (rather than simply the sources themselves).

The form of our solution lends insight into the types of coding sufficient to achieve optimal performance in the given families of problems. Primary among these are entropy codes, including Slepian-Wolf codes for examples with dependent sources. These codes can be implemented, for example, using linear encoders and typical set or minimum entropy decoders [2]. The other feature illustrated by our decomposition is the need to retrieve information from the nearest preceding node where it is available (which may be a sink), thereby

avoiding sending multiple copies of the same information over any link (as can happen in pure routing solutions).

Unfortunately, the given decomposition fails to capture all of the information known to prior nodes in some cases, and thus the achievability result given by the additive construction is not tight in general. Theorem 3 gives a 3-node network where the bound is provably loose. The failure of additivity in this functional source coding example arises since the component network decomposition fails to capture information that intermediate nodes can learn beyond their explicit demands. The same problem can also be replicated in a 4-node network, where all demands also appear as sources in the network. Theorem 4 shows that the gap between the additive solution and the optimal solution can be large.

II. PRELIMINARIES

The following notation is useful in the discussion that follows. For random variables A_1, \dots, A_n , set $S \subseteq \{1, \dots, n\}$, and indices $i, j \in \{1, \dots, n\}$, $A_S = (A_i : i \in S)$ and $A_i^j = (A_i, \dots, A_j)$ denote vectors of A variables and $A_i^j(1 \dots m) = (A_i^j(1), \dots, A_i^j(m))$ denotes consecutive samples of A_i^j .

An n -node line network $(\mathcal{N}, X_1^n, Y_1^n)$ is a directed graph $\mathcal{N} = (V, E)$ with $V = \{1, \dots, n\}$ and $E = \{(1, 2), \dots, (n-1, n)\}$, as shown in Fig. 1. Node i observes source $X_i \in \mathcal{X}_i$ and requires demand $Y_i \in \mathcal{Y}_i$. The random process $(X_1^n(1 \dots \infty), Y_1^n(1 \dots \infty))$ is drawn i.i.d. according to probability mass function $p(x^n, y^n)$. A rate allocation for the n -node line network \mathcal{N} is a vector $(R_i : 1 \leq i < n)$, where R_i is the rate on link $(i, i+1)$. We assume that there are no errors on any of the links. Line networks have been studied earlier in the context of reliable communication (e.g., [3]).

A *simple line network* is a line network with exactly one demand ($Y_i = c$ at all but one node in the network). We next define the *component networks* $\mathcal{N}_1, \dots, \mathcal{N}_n$ for an n -node line network \mathcal{N} with sources X_1^n and demands Y_1^n . (See Fig. 2.) For each $1 \leq i \leq n$, component \mathcal{N}_i is an i -node simple line network. For each $1 \leq j < i$, the source and demand at node j of network \mathcal{N}_j are $X_j^{(i)} = (X_j, Y_i)$ and $Y_j^{(i)} = c$, respectively; the source and demand at node i are $X_i^{(i)} = X_i$ and $Y_i^{(i)} = Y_i$.

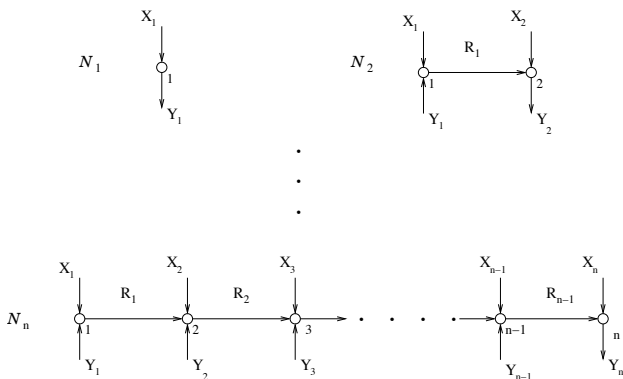


Fig. 2. Component networks

III. RESULTS

A. Cutset bounds and Line Networks

Lemmas 1 and 2 relate the cutset bound to the achievable rate region.

Lemma 1: In an n -node line network $(\mathcal{N}, X_1^n, Y_1^n)$, the cutset bounds are satisfied if and only if

$$R_i \geq \max_{j \geq i+1} H(Y_{i+1}^j | X_{i+1}^j) \quad \forall 1 \leq i < n \quad (2)$$

Proof: The reverse part is immediate since (2) is a subset of the cutset bounds. For the forward part, let $(R_i^* : 1 \leq i < n)$ satisfy (2), and let T be a cut. Each cut is a union of intervals $(T = \bigcup_{k=1}^l T(k))$ with $T(k) = \{m(k), \dots, m(k) + l(k) - 1\} \subseteq \{1, \dots, n\}$ and $m(k-1) + l(k-1) < m(k)$. Then,

$$\begin{aligned} \sum_{k=1}^l R_{m(k)-1}^* &\geq \sum_{k=1}^l \max_{j \geq 0} H(Y_{m(k)}^{m(k)+j} | X_{m(k)}^{m(k)+j}) \\ &\geq \sum_{k=1}^l H(Y_{T(k)} | X_{T(k)}) \\ &\geq H(Y_T | X_T). \end{aligned} \quad (3)$$

Since T is arbitrary, (3) gives the cutset bounds. \blacksquare

Lemma 2: Let $(\mathcal{N}, X_1^3, Y_1^3)$ be a 3-node line network for which the cutset bounds are tight on each of the component networks. Then, the achievable rate region is the set $\mathcal{R} = \{(R_1, R_2) : R_i > R_i^m\}$, where

$$\begin{aligned} R_1^m &= H(Y_2 | X_2) + H(Y_3 | X_2^3, Y_2), \\ R_2^m &= H(Y_3 | X_3). \end{aligned}$$

Proof: Converse: Let C_1 and C_2 be rate R_1 and R_2 m -dimensional codes for links $(1, 2)$ and $(2, 3)$ of \mathcal{N} , and suppose $(1/m)H(Y_i(1 \dots m) | X_i(1 \dots m), C_{i-1}) \leq \epsilon$. Then,

$$\begin{aligned} mR_2 &\geq H(C_2) \geq mH(Y_3 | X_3), \\ mR_1 &\geq H(C_1) \geq H(Y_2(1 \dots m), C_2 | X_2(1 \dots m)) \\ &= mH(Y_2 | X_2) + H(C_2 | X_2(1 \dots m), Y_2(1 \dots m)). \end{aligned}$$

Now,

$$\begin{aligned} &H(C_2 | X_2(1 \dots m), Y_2(1 \dots m)) \\ &\geq I(C_2; Y_3(1 \dots m) | (X_2^3, Y_2)(1 \dots m)) \\ &\geq mH(Y_3 | X_2^3, Y_2) - H(Y_3(1 \dots m) | C_2, X_3(1 \dots m)) \\ &\geq mH(X_2^3, Y_2^3) - m\epsilon. \end{aligned}$$

So $R_1 \geq H(Y_2 | X_2) + H(Y_3 | X_2^3, Y_2) - \epsilon$ and $R_2 \geq H(Y_3 | X_3)$, implying that by picking arbitrarily small ϵ , $(R_1, R_2) \in \mathcal{R}$ and $H(C_2 | C_1) < \epsilon$.

Achievability: Let $R_i = R_i^m + \epsilon$. Since the cutset bound is tight on the components \mathcal{N}_2 and \mathcal{N}_3 for sufficiently large m , there exist m -dimensional codes C_1^2, C_1^3 , and C_2^3 for the links $(1, 2)$ in \mathcal{N}_2 , $(1, 2)$ in \mathcal{N}_3 , and $(2, 3)$ in \mathcal{N}_3 for which

$$\frac{1}{m} H(C_1^2) \leq H(Y_2 | X_2) + \frac{\epsilon}{3}, \quad (4)$$

$$\frac{1}{m} H(C_1^3) \leq H(Y_3 | X_2^3, Y_2) + \frac{\epsilon}{3}, \quad (5)$$

$$\frac{1}{m} H(C_2^3) \leq H(Y_3 | X_3) + \frac{\epsilon}{3}, \quad (6)$$

and $\frac{1}{m}H(Y_i(1 \cdots m)|X_i(1 \cdots m), C_{i-1}^i) \leq \frac{1}{3}$ for $i = 2, 3$. Let $C_1 = C_1^2 C_1^3$ and $C_2 = C_2^3$. Then, (C_1, C_2) is a code for \mathcal{N} satisfying $\frac{1}{m}H(C_i) < R_i$ for $i = 2, 3$, $\frac{1}{m}H(Y_i(1 \cdots m)|X_i(1 \cdots m), C_{i-1}) < \epsilon$ for $i = 1, 2$, and $H(C_2|C_1) < \epsilon$. Thus, (C_1, C_2) achieves rate (R_1, R_2) . ■

B. Independent sources, arbitrary demands

Lemma 3: Consider an n -node line network $(\mathcal{N}, X_1^n, Y_1^n)$ with X_1^n independent and $Y_i = X_{D(i)}$, $D(i) \subseteq \{1, \dots, i\}$ for each $1 \leq i \leq n$. Then, the rate region is fully characterized by the cutset bounds. In particular,

$$R_i \geq H(X_{\cup_{j \geq k+1} D(i) \setminus \{k+1, \dots, n\}}) \quad \forall 1 \leq i < n$$

is achievable.

Proof: Let $Q(i, k) = \cup_{i \leq j \leq k} D(j)$. By Lemma 1, the cut-set bound gives

$$\begin{aligned} R_{i-1} &\geq \max_{k \geq i} H(X_{Q(i, k)} | X_i, X_{i+1}^k) \\ &= \max_{k \geq i} H(X_{Q(i, k) \setminus \{i, \dots, k\}}) \\ &= \max_{k \geq i} \sum_{j \in Q(i, k) \setminus \{i, \dots, k\}} H(X_j). \end{aligned} \quad (7)$$

Since $D(i) \subseteq \{1, \dots, i\}$, $Q(i, k) \subseteq \{1, \dots, k\}$ and $Q(i, k) \setminus \{i, \dots, k\} \subseteq Q(i, k+1) \setminus \{i, \dots, k+1\}$. Thus (7) becomes

$$R_{i-1} \geq \sum_{j \in Q(i, n) \setminus \{i, \dots, n\}} H(X_j).$$

We wish to achieve this bound by coding for the component networks separately. Let $R_i \geq H(X_{Q(i+1, n) \setminus \{i+1, \dots, n\}}) + \epsilon$ for $1 \leq i < n$. Since the demand in network \mathcal{N}_j is $X_{D(j)}$, by calculating the demand across the various links starting from the very last link, the following rate allocation is feasible:

$$\begin{aligned} R_{j-1}^j &= H(X_{D(j) \setminus \{j\}}) + \frac{\epsilon}{n} \\ R_{j-2}^j &= H(X_{(D(j) \setminus \{j\}) \setminus (D(j-1) \cup \{j-1\})}) + \frac{\epsilon}{n} \\ &= H(X_{D(j) \setminus (\{j-1, j\} \cup Q(j-1, j-1))}) + \frac{\epsilon}{n} \\ &\vdots \\ R_1^j &= H(X_{D(j) \setminus (\{2, \dots, j\} \cup Q(2, j-1))}) + \frac{\epsilon}{n}. \end{aligned}$$

Adding the rates from component networks gives

$$\begin{aligned} \sum_{j=i+1}^n R_i^j &= \sum_{j=i+1}^n H(X_{D(j) \setminus (\{i+1, \dots, j\} \cup Q(i+1, j-1))}) + \frac{\epsilon}{n} \\ &= \sum_{j=i+1}^n H(X_{D(j) \setminus (\{i+1, \dots, n\} \cup Q(i+1, j-1))}) + \frac{\epsilon}{n} \\ &= \sum_{j=i+1}^n H(X_{(D(j) \setminus Q(i+1, j-1)) \setminus \{i+1, \dots, n\}}) + \frac{\epsilon}{n} \\ &\leq H(X_{(\cup_{j=i+1}^n D(j)) \setminus \{i+1, \dots, n\}}) + \epsilon \\ &= H(X_{Q(i+1, n) \setminus \{i+1, \dots, n\}}) + \epsilon \\ &\leq R_i \end{aligned}$$

C. Dependent sources, multicast

Lemma 4: Let $(\mathcal{N}, X_1^n, Y_1^n)$ be an n -node line network with X_1^n arbitrarily dependent and Y_1^n feasible multicast demands ($Y_i = c$ or $Y_i = X_1^n$ for each i). Then Theorem 1 holds.

Proof: Let $M = \{1 \leq i \leq n : Y_i = X_1^n\}$ be the multicast receivers. We denote the vertices of M by $\{m_1, \dots, m_k\}$, where $m_i < m_j$ when $i < j$. For each $1 \leq i \leq n$, let $d(i) \triangleq \min\{m \in M : m > i\}$. Consider an achievable rate allocation $(R_i : 1 \leq i < n)$ for \mathcal{N} . For any $1 \leq i < n$, the cutset bound on R_i is tightest possible if we choose the set of vertices for the cutset to be the set $\{i+1, \dots, d(i)\}$. This is true because adding extra vertices to this set adds additional sources to it without increasing the set of demands. Therefore, for all $1 \leq i < n$,

$$R_i \geq H(X_1^n | X_{i+1}^{d(i)}) = H(X_1^{d(i)} | X_{i+1}^{d(i)}) = H(X_1^i | X_{i+1}^{d(i)}) \quad (8)$$

where the first equality follows since $H(X_1^n | X_1^{d(i)}) = 0$ as the demand is feasible for the network. Next, we show that for the component networks $\{\mathcal{N}_j\}_{1 \leq j \leq n}$, there exist rate allocations $\{(R_i^j : 1 \leq i \leq j-1)\}_{j=1}^n$ which come arbitrarily close to satisfying the above bounds with equality. Observe that it suffices to restrict our attention to the networks $\{\mathcal{N}_j\}_{j \in M}$. Fix $\epsilon > 0$ and consider any $m_i \in M$.

When $i = 1$, it suffices to encode X_j at rate $H(X_j | X_{j+1}^{m_1}) + \epsilon/n$ [4]. Summing the over all sources that use link $(r, r+1)$ rate achievable rate $R_r^{m_1}$ for network \mathcal{N}_{m_1} , where

$$\begin{aligned} R_r^{m_1} &= \sum_{j=1}^r H(X_j | X_{j+1}^{m_1}) + \frac{\epsilon}{n} \\ &\leq H(X_1^r | X_{r+1}^{m_1}) + \epsilon \\ &= H(X_1^r | X_{r+1}^{d(r)}) + \epsilon. \end{aligned} \quad (9)$$

When $i > 1$, X_1^n are available at the node m_{i-1} . Hence, the rate required over the link $(r, r+1)$ is zero for all $r < m_{i-1}$. for $m_{i-1} \leq r < m_i$, [4] again gives

$$\begin{aligned} R_r^{m_i} &= H(X_1^n | X_{m_{i-1}+1}^{m_i}) + \sum_{j=m_{i-1}+1}^r H(X_j | X_{j+1}^{m_i}) + \frac{\epsilon}{n} \\ &\leq H(X_1^r | X_{r+1}^{m_i}) + \epsilon \\ &= H(X_1^r | X_{r+1}^{d(r)}) + \epsilon. \end{aligned} \quad (10)$$

Finally, adding the rates over component networks gives

$$\sum_{j=i+1}^n R_i^j = R_i^{d(i)} \leq H(X_1^i | X_{i+1}^{d(i)}) + \epsilon \leq R_i + \epsilon.$$

This shows that the cutset bound in (8) is tight and is achievable by the approach based on component networks. Further, each R_i^j is of the form $H(Y_j | X_{i+1}^j, Y_{i+1}^{j-1})$. ■

D. A class of dependent sources with dependent demands

In this section, we consider sources and demands that are dependent in the following way. We assume the existence of underlying independent sources W_1^k such that the sources are $X_i = W_{S(i)}$ and the demands are $Y_i = W_{D(i)}$ for $1 \leq i \leq n$ for $\{S(i)\}_{i=1}^n$ and $\{D(i)\}_{i=1}^n$ subsets of $\{1, \dots, k\}$. In order

for the demands to be feasible, we require $D(i) \subseteq \cup_{j=1}^i S(j)$ for each $1 \leq i \leq n$. Lemma 5 characterizes the rate region for line networks with the above kind of sources and demands. We need the following notation in order to state the lemma. For $1 \leq j < n$, define $d_i(j)$ and $s_i(j)$ as the first occurrence after the vertex j of W_i in a demand and source respectively.

Lemma 5: Let $(\mathcal{N}, X_1^n, Y_1^n)$ be an n -node line network with $X_i = W_{S(i)}$ and $Y_i = W_{D(i)}$ as defined above. Then the achievable rate region is

$$R_j \geq \sum_{\substack{i=1 \\ i:s_i(j) > d_i(j)}}^k H(W_i) \quad \forall 1 \leq j < n.$$

Proof: We proceed by first decomposing the network \mathcal{N} into k different networks $\{\mathcal{N}^i\}_{i=1}^k$, each corresponding to a different W_i out W_1^k .

For each $1 \leq i \leq k$ and $A \subseteq \{1, \dots, k\}$, let $\tilde{X}_A^i = W_i$ if $i \in A$ and $\tilde{X}_A^i = c$ otherwise. Let $\tilde{\mathcal{N}}_i$ be an n -node line network with sources $\tilde{X}_{S(1)}^i, \dots, \tilde{X}_{S(n)}^i$ and demands $\tilde{X}_{D(1)}^i, \dots, \tilde{X}_{D(n)}^i$. Note that each $\tilde{\mathcal{N}}_i$ is a line network in which both the sources and demands are either W_i or constant. By result of Sec. III-C, it follows that the cutset bound is tight for such network and the rate allocation $(R_{j,i} : 1 \leq j < n)$, defined by

$$R_{j,i} = \begin{cases} H(W_i) & \text{if } s_i(j) > d_i(j) \\ 0 & \text{otherwise} \end{cases}$$

is achievable. Thus, for the parent network \mathcal{N} , the rate allocation $(R_j : j \in \{1, \dots, n\})$ is achievable, where $R_j = \sum_{i=1}^k R_{j,i}$. Further, as all the sources are independent, this approach is optimal. Therefore, the rate region for the network \mathcal{N} is given by all $(R_j : j \in \{1, \dots, n\})$ such that

$$R_j \geq \sum_{\substack{i=1 \\ i:s_i(j) > d_i(j)}}^k H(W_i).$$

Next, we show that the same can be obtained by decomposing \mathcal{N} into simple networks $\mathcal{N}_1, \dots, \mathcal{N}_n$. To this end, we first decompose the network \mathcal{N}^i into simple networks $\mathcal{N}_1^i, \dots, \mathcal{N}_n^i$, noting that the minimum rate $R_{j,i}^l$ required for the link $(j, j+1)$ in \mathcal{N}_l^i is 0 if there is a demand or a source present in one of the nodes $j+1, \dots, l$ and $H(W_i)$ otherwise. Hence, $R_{j,i}^l = H(\tilde{X}_{D(l)}^i | \tilde{X}_{S(j+1)}^i, \dots, \tilde{X}_{S(l)}^i, \tilde{X}_{D(j+1)}^i, \dots, \tilde{X}_{D(l-1)}^i)$. This is an optimal decomposition, since

$$R_{j,i} = \sum_{l=j+1}^n R_{j,i}^l.$$

Adding the rates over components \mathcal{N}^i ,

$$\begin{aligned} R_j^l &= \sum_{i=1}^k R_{j,i}^l \\ &= \sum_{i=1}^k H(\tilde{X}_{D(l)}^i | \tilde{X}_{\cup_{t=j+1}^l S(t)}^i, \tilde{X}_{\cup_{t=j+1}^l D(t)}^i) \\ &= H(X_{D(l)} | X_{\cup_{t=j+1}^l S(t)}, X_{\cup_{t=j+1}^l D(t)}) \end{aligned} \quad (11)$$

is achievable for \mathcal{N}_l . Finally, note that

$$\sum_{l=j+1}^n R_j^l = \sum_{l=j+1}^n \sum_{i=1}^k R_{j,i}^l = \sum_{i=1}^k R_{j,i} = R_j.$$

Thus the rates for \mathcal{N} can also be obtained by summing linkwise the rates for the component networks. By (11), the sum is of the form claimed in Theorem 1-C.

E. 3-node line networks with dependent sources

We now restrict our attention to 3-node line networks of the kind shown in Fig. 3. Sources X_1^3 are arbitrarily dependent and their alphabets are finite; demands Y_2^3 take the form $Y_2 = f(X_1)$ and $Y_3 = g(X_1^3)$ for some $f : \mathcal{X}_1 \rightarrow \mathcal{Y}_1$ and $g : \prod_{i=1}^3 \mathcal{X}_i \rightarrow \mathcal{Y}_2$. The following result shows the tightness of our decomposition approach in this case.

Theorem 2: Given $\epsilon > 0$, for every rate vector (R_1, R_2) achievable for \mathcal{N} , there exist achievable rate allocations R_1^2 and (R_1^3, R_2^3) for component networks \mathcal{N}_2 and \mathcal{N}_3 such that

$$R_1^2 + R_1^3 < R_1 + \epsilon \text{ and } R_2^3 = R_2 + \epsilon.$$

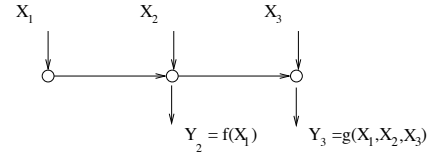


Fig. 3. The three node line network

Proof: Let (R_1, R_2) be achievable for \mathcal{N} . Then, for m large enough, there exist codes (a_m, b_m) for the first and second link, respectively such that $(1/m)H(a_m(X_1(1 \dots m))) < R_1 + \epsilon$, $(1/m)H(b_m(a_m(X_1^2(1 \dots m)))) < R_2 + \epsilon$, and $\Pr(\hat{Y}_2^3(1 \dots m) \neq Y_2^3(1 \dots m)) < \epsilon$.

Let $B_m(k) = b_m(a_m(X_1^2(m(k-1)+1 \dots mk)))$, $F_m(k) = (f(X_1(m(k-1)+1), \dots, f(X_1(mk))), X_{i,m}(k) = X_i(m(k-1)+1 \dots mk)$. Allowing a probability of error ϵ , the problem of coding for the network \mathcal{N} can be reformulated as a functional source coding problem for the network $\mathcal{N}^{(i)}$ with sources $X_{1,i}$ and $X_{2,i}$ and the demand (F_i, B_i) as shown in Fig 4. Prior results on functional coding ([5], [6], [7]) can

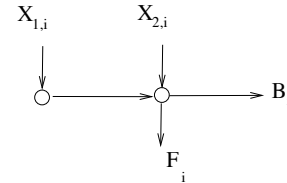


Fig. 4. An equivalent functional coding problem

now be applied. Specifically, by evaluating the functional rate distortion function in [6], [7] at zero distortion, the minimal rate at which $X_{1,i}$ can be coded is given by

$$R^* = \inf_{\hat{X}_{1,i}} I(X_{1,i}; \hat{X}_{1,i} | X_{2,i})$$

where, the infimum is over the set \mathcal{P} consisting of all $\widehat{X}_{1,i}$ for which $\widehat{X}_{1,i} \rightarrow X_{1,i} \rightarrow X_{2,i}$ forms a Markov chain and $H(F_i, B_i | \widehat{X}_{1,i}, X_{2,i}) = 0$.

We show that the above rate can be split into two parts - the rate required to encode $X_{1,i}$ so as to reconstruct F_i with $X_{2,i}$ as the side information, and the rate required to be able to reconstruct B_i with $X_{2,i}$ and F_i as the side information. To this end, let $X_{F,B} \in \mathcal{P}$. Then, the following hold:

$$\begin{aligned}
& I(X_{F,B}; X_{1,i} | X_{2,i}) \\
&= I(X_{F,B}, F_i; X_{1,i} | X_{2,i}) - I(F_i; X_{1,i} | X_{2,i}, X_{F,B}) \\
&= I(X_{F,B}, F_i; X_{1,i} | X_{2,i}) \\
&= H(X_{F,B}, F_i | X_{2,i}) - H(X_{F,B}, F_i | X_{2,i}, X_{1,i}) \\
&= H(X_{F,B}, F_i | X_{2,i}) - H(X_{F,B} | X_{2,i}, X_{1,i}, F_i) \\
&= H(F_i | X_{2,i}) + H(X_{F,B} | F_i, X_{2,i}) \\
&\quad - H(X_{F,B} | X_{2,i}, X_{1,i}, F_i) \\
&= H(F_i | X_{2,i}) + I(X_{F,B}; X_{1,i} | F_i, X_{2,i}). \tag{12}
\end{aligned}$$

Since F_i is a function of $X_{1,i}$, $\frac{1}{i}H(F_i | X_{2,i})$ is an achievable rate for the network \mathcal{N}_1 . Further, since $X_{F,B} \in \mathcal{P}$, it follows that $X_{F,B} \rightarrow X_{1,i} \rightarrow (F_i, X_{2,i})$ is a Markov chain. Combining it with the fact that, $H(B_i | X_{F,B}, X_{1,i}, F_i) = 0$, we note that $I(X_{F,B}; X_{1,i} | F_i, X_{2,i})$ is a sufficient rate for functional source coding with regards to the function B_i given $X_{2,i}$ and F_i as the side information. Therefore, $(\frac{1}{i}I(X_{F,B}; X_{1,i} | F_i, X_{2,i}), \frac{1}{i}H(B_i))$ is an achievable rate for the network \mathcal{N}_3 , hence proving the theorem. ■

F. Networks where additivity does not hold

Theorem 3: There exists a 3-node network $(\mathcal{N}, X_1^3, Y_1^3)$, for which adding the best rate allocations from the component networks does not yield an optimal code.

Proof: Consider the example shown in Fig. 5(a).

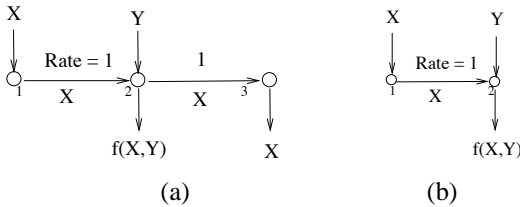


Fig. 5. (a) A set of achievable rates for \mathcal{N} . (b) An optimal rate allocation for component \mathcal{N}_1 .

Let X be distributed uniformly on $\{0, 1\}$ and Y be independently distributed uniformly on $\{0, 1, 2, 3\}$. Let

$$f(x, y) \triangleq \begin{cases} 0 & (x, y) \in \{(0, 0), (0, 2), (1, 1), (1, 2)\} \\ 1 & (x, y) \in \{(0, 1), (0, 3), (1, 0), (1, 3)\} \end{cases}$$

Fig. 5(a) shows an achievable rate allocation that can be achieved by the transmitting X over both links. We next show that best possible rate allocation achievable by optimizing over the achievable rate regions of the component networks is strictly greater than the above rate allocation. To this end, consider \mathcal{N}_1 (see Fig. 5(b)). By [5], the best possible rate on link (1,2) is 1.

To prove our claim, it suffices to show that the rate required over link (1, 2) \mathcal{N}_2 is non-zero. This follows because $H(X|Y, f(X, Y)) > 0$, and the cutset bound requires at least rate $H(X|Y, f(X, Y)) > 0$ across link (1, 2). ■

Theorem 4: For any $n \geq 3$, there exists a $2n$ -node network $(\mathcal{N}, X_1^{2n}, Y_1^{2n})$, such that for any achievable $(R_i^j : 1 \leq i < j)$ for \mathcal{N}_j , there exists an achievable $(R_i : 1 \leq i < n)$ for \mathcal{N} with

$$\sum_{j=i+1}^{2n} R_i^j > R_i + \Omega(n).$$

Proof: Let X and Y be independent sources uniformly distributed over $\{0, 1\}$ and $\{0, \dots, 2n\}$ respectively. Define $f : \{0, 1\} \times \{0, \dots, 2n\} \rightarrow \{0, \dots, 2n\}$ as

$$f(x, y) = \begin{cases} y & \text{if } y \in \{1, \dots, 2n-1\} \\ x & \text{if } y = 2n \end{cases}$$

Consider the $2n$ -node line network $(\mathcal{N}, X_1^{2n}, Y_1^{2n})$, where $X_1 = X, X_2 = X_3 = \dots, X_{2n-1} = Y, X_{2n} = Y_1 = c$, and $Y_i = f(X, Y \oplus (i-1))$ for $2 \leq i \leq 2n$.

By the functional source coding bound, the rate required on each link of the i -th component is at least 1. This rate is achieved by sending X on all the links. Thus, $\sum_{j=i+1}^{2n} R_i^j = 2n - i$. In contrast, rate 1 is sufficient for network \mathcal{N} (sending X over all the links). Therefore, $\sum_{j=i+1}^{2n} R_i^j - R_i \geq 2n - i - 1$. The left side is at most $\mathcal{O}(n)$ for a network with $2n$ components. Hence, $\sum_{j=i+1}^{2n} R_i^j - R_i = \Omega(n)$. ■

ACKNOWLEDGMENT

This material is based upon work partially supported by NSF Grant Nos. CCF-0220039, CCR-0325324 and Caltech's Lee Center for Advanced Networking.

REFERENCES

- [1] T. Ho, R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, October 2006.
- [2] Imre Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Transactions on Information Theory*, IT-28:585–592, July 1982.
- [3] P. Pakzad, C. Fragouli, and A. Shokrollahi. Coding schemes for line networks. In *Proceedings of the IEEE International Symposium on Information Theory*, Adelaide, Australia, September 2005.
- [4] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.
- [5] A. Orlitsky and J. R. Roche. Coding for computing. In *IEEE Symposium on Foundations of Computer Science*, pages 502–511, 1995.
- [6] H. Yamamoto. Wyner-ziv theory for a general function of correlated sources. *IEEE Transactions on Information Theory*, IT-28(5):803–807, September 1982.
- [7] H. Feng, M. Effros, and S. A. Savari. Functional source coding for networks with receiver side information. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2004.