The Multicast Capacity of Acyclic, Deterministic Relay Networks with No Interference

Niranjan Ratnakar and Gerhard Kramer

***Abstract*—The multicast capacity is determined for acyclic networks that have deterministic links with broadcasting at the transmitters and no interference at the receivers. Such networks were studied by M. R. Aref, and are here called Aref networks. The multicast capacity is shown to have a max-flow, min-cut interpretation. This result complements existing theory for networks of directed channels, networks of undirected channels, and packet erasure networks. It is also shown that one cannot always separate channel and network coding in Aref networks.**

I. Introduction

Consider a network represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the respective sets of vertices and directed edges. For example, the graph might represent a communication network where the vertices are terminals and the edges are channels. We study a class of networks known as *deterministic* relay networks with *no interference*. Such networks have one input X_u associated with every vertex u, and one output $Y_{u,v}$ associated with every edge (u, v). By *deterministic*, we mean that $Y_{u,v}$ is some deterministic function of X_u. By *no interference*, we mean that $Y_{u,v}$ is a function of X_u *only*. These restrictions explicitly permit *broadcasting*, since the outgoing edges of a vertex share a common input. We remark that the commonly studied networks that have deterministic point-to-point channels are special types of these networks. To see this, collect the inputs of all the outgoing edges from a vertex u into a vector $\underline{X}_u$, and view $\underline{X}_u$ as being a common input.

The above networks were considered by M. R. Aref in his Ph. D. thesis [1], and we thus call them *Aref networks*. Aref determined the *unicast* capacity of his networks, i.e., the maximum rate for reliable communication of one message from one source vertex to one destination vertex. A *layered* coding scheme turns out to be optimal, i.e., one can separately apply channel codes to a *physical* layer, and routing schemes to a *network* layer. That is, each vertex decodes bit blocks (or packets), reorganizes them into smaller or larger packets, encodes these with a channel code, and sends the encoded symbols out on different edges.

Multicast refers to the scenario when there is one message transmitted from one source vertex to one or more destination vertices. The maximum rate at which one can communicate reliably is called the *multicast capacity*, and we determine this capacity for acyclic Aref networks. We find that, in general, capacity-achieving coding schemes must use *network* coding [2], and not only packet routing, and one cannot separate the physical and network layers.

This document is organized as follows. In Section II, we review problems for which the capacity has a max-flow, min-cut interpretation. In Section III, we define the network model and problem. In Section IV, we derive the multicast capacity of Aref networks. In Section V, we show that one cannot always separate channel and network coding.

Manuscript received March 2005. The material in this paper was presented in part at the First Workshop on Network Coding, Theory and Applications, April 7, 2005, Riva del Garda, Italy. The work of N. Ratnakar was supported by the Vodafone foundation. The work of G. Kramer was partially supported by the Board of Trustees of the University of Illinois Subaward No. 04-217 under NSF Grant No. CCR-0325673.

N. Ratnakar is with the Coordinated Science Laboratory, University of Illinois at Urbana Champaign, 1308 West Main St., Urbana IL 61801 (email: ratnakar@ifp.uiuc.edu).

G. Kramer is with Bell Laboratories, Lucent Technologies, 600 Mountain Ave, Murray Hill NJ 07974 (email: gkr@bell-labs.com).

II. Max-Flow, Min-Cut Results

The unicast capacity of directed or undirected networks was shown to have a max-flow, min-cut interpretation by Ford and Fulkerson, Dantzig and Fulkerson, and Elias, Feinstein and Shannon [3–5]. The multicast capacity of directed networks was similarly shown to have a max-flow, min-cut interpretation by Ahlswede *et al.* [2] (see also [6]), as long as one can combine commodities (or bits) at the vertices. This result was extended to undirected networks in [7, 8]. The important difference between unicasting and multicasting for these problems is that multicasting requires the use of *network coding*, i.e., one must permit the vertices to combine packets before transmitting them.

The above papers label each edge by its *capacity* or *capacity region*. However, to understand how the physical and network layers interact, one needs to specify a channel model for each edge, and not only the capacity region. For instance, for directed graphs one might model the edges as discrete memoryless channels (DMCs) [9], while for undirected graphs one might model the edges as two-way channels (TWCs) [8] (DMCs and TWCs were introduced in [10, Sec. 11] and [11]). The multicast capacity was in both cases[1] shown to have a max-flow, min-cut interpretation, but "min-cut" now refers to an information-theoretic cut-set bound [12, Sec. 14.10].

The physical-layer model studied here might be considered an intermediate step toward modeling wireless networks because it includes broadcasting. Two recent papers take a similar approach and study packet erasure networks [13, 14]. However, these works require the destinations to know where erasures have occurred in the network, i.e., the destinations receive additional "side information" that is assumed to be present in the

[1] The results of [8] require that the TWCs have the property that Shannon's outer bound on the capacity region is the capacity region. This technical condition is sometimes met in practice, e.g., if one uses time or frequency division multiplexing.

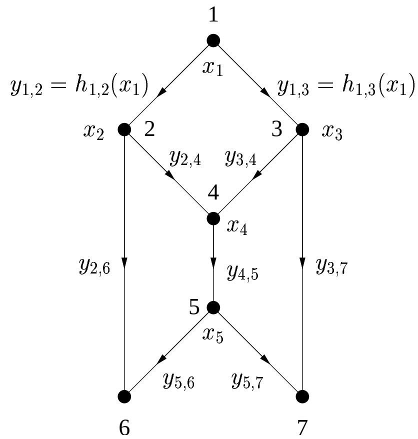

Fig. 1. Example of an Aref network.

packet headers. We will not need to add such an assumption to our problem.

III. Model and Problem

Consider the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ described above. The random variable X_u associated with vertex u has discrete and finite alphabet $\mathcal{X}_u$. We write $p_{X_u}(\cdot)$ for the distribution of X_u, and $p_{X_u}(x_u)$ or $p(x_u)$ for the probability that $X_u = x_u$. Similarly, the random variable $Y_{u,v}$ associated with edge (u, v) has discrete and finite alphabet $\mathcal{Y}_{u,v}$. We write $Y_{u,v} = h_{u,v}(X_u)$, where the function $h_{u,v}(\cdot)$ has domain $\mathcal{X}_u$ and range $\mathcal{Y}_{u,v}$. The network is *clocked*, i.e., all vertices and edges are activated simultaneously N times, and at every time instant vertex u transmits a symbol x_u, $x_u \in \mathcal{X}_u$, and receives symbols $y_{w,u} = h_{w,u}(x_w)$, where w is any vertex for which there is an edge (w, u) in $\mathcal{E}$.

For example, consider the network shown in Fig. 1. Vertex 1 transmits x_1, and vertices 2 and 3 observe $h_{1,2}(x_1)$ and $h_{1,3}(x_1)$, respectively. The models of [2, 6] have this property: vertex 1 transmits a vector $\underline{x}_1 = [x_{1,1}, x_{1,2}]$, and vertices 2 and 3 observe the respective $h_{1,2}(x_{1,2})$ and $h_{1,3}(x_{1,3})$. We can, of course, view this as $y_{12} = h_{1,2}(\underline{x}_1)$ and $y_{13} = h_{1,3}(\underline{x}_1)$.

A. Problem and Coding

The multicasting problem has one source vertex (vertex 1), and several destination vertices that we collect in a set $\mathcal{T}$. The source vertex has a message M that is uniformly distributed over $\{1, 2, \ldots, 2^{NR}\}$, where R is the *rate* and where we assume that NR is an integer for simplicity. A communication strategy consists of encoding functions $f_u^{(i)}(\cdot)$, $u \in \mathcal{V}$, $i = 1, 2, \ldots, N$, and decoding functions $\hat{m}_t(\cdot)$, $t \in \mathcal{T}$. We write $\underline{x}_{\mathcal{S}} = [x_u : u \in \mathcal{S}]$ and $\underline{y}_{\mathcal{S},\mathcal{S}'} = [y_{u,v} : u \in \mathcal{S}, v \in \mathcal{S}']$. We similarly write $\underline{y}_{u,\mathcal{S}'} = [y_{u,v} : v \in \mathcal{S}']$.

- *Encoders*. Suppose $M = m$. At time i, vertex 1 transmits $x_1^{(i)} = f_1^{(i)}(m)$ and every other vertex u receives $\underline{y}_{\mathcal{V},u}^{(i)}$. Vertex u transmits $x_u^{(i)} = f_u^{(i)}(\underline{y}_{\mathcal{V},u}^{(1)}, \underline{y}_{\mathcal{V},u}^{(2)}, \ldots, \underline{y}_{\mathcal{V},u}^{(i-1)})$.
- *Decoders*. After time N, each destination vertex t puts out an estimate $\hat{m}_t(\underline{y}_{\mathcal{V},t}^{(1)}, \underline{y}_{\mathcal{V},t}^{(2)}, \ldots, \underline{y}_{\mathcal{V},t}^{(N)})$.

The error probability is

$$P_e = \Pr\left[\bigcup_{t\in T}\left\{\hat{m}_t\left(\underline{Y}_{\mathcal{V},t}^{(1)}, \underline{Y}_{\mathcal{V},t}^{(2)}, \ldots, \underline{Y}_{\mathcal{V},t}^{(N)}\right) \neq M\right\}\right]. \quad (1)$$

The rate R is said to be *achievable* if, for any $\epsilon > 0$, there exist encoders and decoders that make $P_e \leq \epsilon$ for some N. The *multicast capacity* C is the supremum of the achievable rates.

B. Cuts and Values

Consider a set $\mathcal{S}$ of vertices and let $\overline{\mathcal{S}}$ be its complement in $\mathcal{V}$. $\mathcal{S}$ is called a *cut* if $1 \in \mathcal{S}$ and $\overline{\mathcal{S}}$ contains one or more destination vertices, i.e., if $\overline{\mathcal{S}} \cap \mathcal{T} \neq \emptyset$. We denote the set of all cuts as Λ. The *boundary* of a cut $\mathcal{S}$ is defined as

$$\beta(\mathcal{S}) = \{u : (u, v) \in \mathcal{E}, u \in \mathcal{S}, v \in \overline{\mathcal{S}}\}$$

Let $|\mathcal{V}|$ be the cardinality of $\mathcal{V}$. For fixed input distributions $p_{X_1}(\cdot), p_{X_2}(\cdot), \ldots, p_{X_{|\mathcal{V}|}}(\cdot)$, we define the *value* of a cut $\mathcal{S}$ as

$$\text{Value}(\mathcal{S}) = \sum_{u\in\beta(\mathcal{S})} H(\underline{Y}_{u,\overline{\mathcal{S}}}) \quad (2)$$

where we recall that $\underline{Y}_{u,\overline{\mathcal{S}}} = [Y_{u,v} : v \in \overline{\mathcal{S}}]$. The value of a cut depends on the input distributions, but we do not explicitly include these as arguments in (2).

IV. Main Result

We restrict attention to acyclic Aref networks. We can thus number the vertices so that $(u, v) \in \mathcal{E}$ implies that $u < v$. We can further consider only the subgraph having those vertices and edges on the paths from the source to the destinations. The following is our main result.

Theorem 1: The multicast capacity of an acyclic Aref network is

$$C = \max_{p_{X_1}(\cdot), p_{X_2}(\cdot), \ldots, p_{X_{|\mathcal{V}|}}(\cdot)} \min_{\mathcal{S}\in\Lambda} \text{Value}(\mathcal{S}). \quad (3)$$

This result has a max-flow, min-cut interpretation for fixed input distributions. The minimum value of all cuts, in an information-theoretic sense [12, Sec. 14.10], is

$$\begin{aligned}\min_{\mathcal{S}\in\Lambda} I(\underline{X}_{\mathcal{S}}\,;\, \underline{Y}_{\mathcal{V},\overline{\mathcal{S}}} \mid \underline{X}_{\overline{\mathcal{S}}}) &= \min_{\mathcal{S}\in\Lambda} H(\underline{Y}_{\mathcal{V},\overline{\mathcal{S}}} \mid \underline{X}_{\overline{\mathcal{S}}}) \\ &= \min_{\mathcal{S}\in\Lambda} \text{Value}(\mathcal{S}) \end{aligned} \quad (4)$$

where the first equality follows because the network is deterministic, and the second because the X_u are independent. We discuss the derivation of (4) in more detail in Section IV-B below.

A. Achievability

We use δ-typical sequences for our achievability proof. Let $\mathcal{X}^n$ be the n-fold Cartesian product of $\mathcal{X}$. Let $\nu_{\underline{x}}(a)$ be the number of times the letter a occurs in the sequence $\underline{x}$ of length n. The *empirical frequency* of a in $\underline{x}$ is

$$\pi_{\underline{x}}(a) = \frac{\nu_{\underline{x}}(a)}{n}. \quad (5)$$

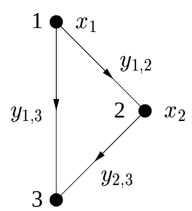

Fig. 2. A second example of an Aref network.

Let $\mathcal{X}$ be the (discrete and finite) alphabet of the entries in $\underline{x}$, and let $\delta > 0$. Let X be a random variable with alphabet $\mathcal{X}$ and distribution $p_X(\cdot)$. The sequence $\underline{x}$ is said to be (robustly) δ-*typical* with respect to X if

$$\left|\pi_{\underline{x}}(a) - p_X(a)\right| \le \delta \cdot p_X(a)$$

for all $a \in \mathcal{X}$ (see [15]). The set of δ-typical sequences with respect to X is denoted $T_\delta(X)$. One can similarly define δ-typical sequences and sets with respect to pairs of random variables. Some of the key properties of these objects are listed in the Appendix.

We code in $L + |\mathcal{V}| - 2$ blocks of length n, i.e., we set $N = (L + |\mathcal{V}| - 2) \cdot n$ for $L \ge 1$. We divide the message m into L parts that each take on values in $\{1, 2, \ldots, 2^{nR}\}$. The l-th part of m is denoted m_l. The overall rate is $R \cdot L/(L + |\mathcal{V}| - 2)$, but since $|\mathcal{V}|$ is finite one can approach R by increasing L.

Let $f_v^n(\cdot) = [f_v^{(i)}(\cdot) : i = 1, 2, \ldots, n]$, and set $f_v^{(i+l\cdot n)}(\cdot) = f_v^{(i)}(\cdot)$ for all i, l, and v, i.e., we use the same encoding function(s) for each block. We associate $\underline{x}_u^n(l)$ and $\underline{y}_{w,u}^n(l)$ (for $(w, u) \in \mathcal{E}$) with the transmissions for the l-th message m_l. That is, we write the $(l + u - 1)$-th transmitted vector at vertex u and $(l + w - 1)$-th received vector on edge (w, u) as the respective $\underline{x}_u^n(l)$ and $\underline{y}_{w,u}^n(l)$ and we write $\underline{y}_{\mathcal{V},u}^n(l) = [\underline{y}_{w,u}^n(l) : w \in \mathcal{V}, (w, u) \in \mathcal{E}]$. We sometimes drop the index l if it does not play an important role. We fix the distributions $p_{X_1}(\cdot), p_{X_2}(\cdot), \ldots, p_{X_{|\mathcal{V}|}}(\cdot)$.

Codebooks. At vertex 1, choose $f_1^n(\cdot)$ to map each of the indices in $\{1, 2, \ldots, 2^{nR}\}$ to a sequence $\underline{x}_1^n$ drawn uniformly from $T_\delta(X_1)$. At vertex u, choose $f_u^n(\cdot)$ to map each sequence in $T_\delta(\underline{Y}_{\mathcal{V},u})$ to a sequence drawn uniformly from $T_\delta(X_u)$. Note that some $\underline{y}_{\mathcal{V},u}^n$ in $T_\delta(\underline{Y}_{\mathcal{V},u})$ might never be used. Note also that we have $\underline{y}_{u,v}^n \in T_\delta(Y_{u,v})$ for all (u, v) because $\underline{x}_u^n \in T_\delta(X_u)$ for all u (see the Appendix, Lemma 4).

Encoding. During the $(l + u - 1)$-th block:

- Vertex $u = 1$ transmits $\underline{x}_1^n(l) = f_1^n(m_l)$ for $l = 1, 2, \ldots, L$ and $\underline{x}_1^n(l) = f_1^n(1)$ otherwise.
- Vertex u, $u \neq 1$, observes $\underline{y}_{\mathcal{V},u}^n(l + 1)$ and transmits

$$\underline{x}_u^n(l) = f_u^n\left(\underline{y}_{\mathcal{V},u}^n(l - 1)\right).$$

For instance, the coding strategy for the network shown in Fig. 2 is given in Table IV-A. Observe how the transmissions are "pipelined".

Decoding. Consider a destination vertex t. Since $\underline{y}_{\mathcal{V},t}^n(l)$ is a function of m_l, we abuse notation and write this sequence of vectors as $\underline{y}_{\mathcal{V},t}^n(m_l)$. Vertex t decodes m_l after block $t + l - 2$ by using the function $\hat{m}_t(\cdot)$, where we again abuse notation by using the same expression as in (1). We further define

$$\hat{m}_t\left(\underline{y}_{\mathcal{V},t}^n(m_l)\right) = \begin{cases} \text{error} & \text{if } \underline{y}_{\mathcal{V},t}^n(m_l') = \underline{y}_{\mathcal{V},t}^n(m_l) \text{ for some } m_l' \neq m_l \\ m_l & \text{otherwise.} \end{cases} \tag{6}$$

Analysis. In the following, we consider only transmissions that pertain to the message m_l. We thus drop the index l for convenience. For example, we write m, $\underline{x}_u^n$, and $\underline{y}_{u,v}^n$ for m_l, $\underline{x}_u^n(l)$ and $\underline{y}_{u,v}^n(l)$, respectively.

Consider a destination vertex t. Let $\overline{P}_e(t, m, m')$ be the average probability that vertex t cannot distinguish between m and m', where the average is over the ensemble of encoding functions. Let $\mathcal{S}(m, m')$ be the set of vertices u for which

$$\underline{y}_{\mathcal{V},u}^n(m) \neq \underline{y}_{\mathcal{V},u}^n(m') \tag{7}$$

i.e., $\mathcal{S}(m, m')$ is the set of the vertices that *can* distinguish between m and m'. We view $\mathcal{S}(m, m')$ as a random variable that is a function of the encoding functions. We clearly have $1 \in \mathcal{S}(m, m')$. Suppose vertex t cannot distinguish between m and m', so that $t \in \overline{\mathcal{S}}(m, m')$ and $\mathcal{S}(m, m')$ is a cut between vertices 1 and t. Let Λ_t be the set of such cuts, i.e., we define $\Lambda_t = \{\mathcal{S} \subset \mathcal{V} : 1 \in \mathcal{S}, t \in \overline{\mathcal{S}}\}$. We can write

$$\begin{aligned} \overline{P}_e(t, m, m') &= \Pr\left[\bigcup_{\mathcal{S} \in \Lambda_t} \{\mathcal{S}(m, m') = \mathcal{S}\}\right] \\ &\le \sum_{\mathcal{S} \in \Lambda_t} \Pr\left[\mathcal{S}(m, m') = \mathcal{S}\right]. \end{aligned} \tag{8}$$

Furthermore, we claim that a necessary condition for the event $\mathcal{S}(m, m') = \mathcal{S}$ is one must have

$$\underline{y}_{u,\overline{\mathcal{S}}}^n(m) = \underline{y}_{u,\overline{\mathcal{S}}}^n(m') \tag{9}$$

for *all* $u \in \mathcal{S}(m, m')$. To see this, note that if (9) was not true for some $u \in \mathcal{S}(m, m')$, then there is a vertex $v \in \overline{\mathcal{S}}$ that can distinguish between m and m', contradicting our original hypothesis. We can thus write

$$\begin{aligned} \Pr\left[\mathcal{S}(m, m') = \mathcal{S}\right] &\le \Pr\left[\bigcap_{u \in \beta(\mathcal{S})} \left\{\underline{Y}_{u,\overline{\mathcal{S}}}^n(m) = \underline{Y}_{u,\overline{\mathcal{S}}}^n(m')\right\}\right] \\ &= \prod_{u \in \beta(\mathcal{S})} \Pr\left[\underline{Y}_{u,\overline{\mathcal{S}}}^n(m) = \underline{Y}_{u,\overline{\mathcal{S}}}^n(m')\right] \end{aligned} \tag{10}$$

where the equality follows because the $\underline{X}_u^n(m)$, $u \in \mathcal{V}$, $m = 1, 2, \ldots, 2^{nR}$, are statistically independent.

We proceed to upper bound the probabilities in the product in (10). We have $(\underline{x}_u^n(m'), \underline{y}_{u,\overline{\mathcal{S}}}^n(m')) \in T_\delta(X_u, Y_{u,\mathcal{S}})$ by Lemma 4 in the Appendix. The event (9) thus implies

$$\left(\underline{x}_u^n(m'), \underline{y}_{u,\overline{\mathcal{S}}}^n(m)\right) \in T_\delta(X_u, Y_{u,\overline{\mathcal{S}}}). \tag{11}$$

TABLE I
CODING STRATEGY FOR THE NETWORK OF FIG. 2

Block	Message	1 Transmits	2 Receives	2 Transmits	3 Receives	Decoder output
1	$m(1)$	$\underline{x}_1^n(1)$	$\underline{y}_{1,2}^n(1)$	$\cdot$	$\underline{y}_{1,3}^n(1)$	$\cdot$
2	$m(2)$	$\underline{x}_1^n(2)$	$\underline{y}_{1,2}^n(2)$	$\underline{x}_2^n(1)$	$\underline{y}_{1,3}^n(2), \underline{y}_{2,3}^n(1)$	$\hat{m}_t(\underline{y}_{1,3}^n(1), \underline{y}_{2,3}^n(1))$
3	$m(3)$	$\underline{x}_1^n(3)$	$\underline{y}_{1,2}^n(3)$	$\underline{x}_2^n(2)$	$\underline{y}_{1,3}^n(3), \underline{y}_{2,3}^n(2)$	$\hat{m}_t(\underline{y}_{1,3}^n(2), \underline{y}_{2,3}^n(2))$
4	$\cdot$	$\cdot$	$\cdot$	$\underline{x}_2^n(3)$	$\underline{y}_{2,3}^n(3)$	$\hat{m}_t(\underline{y}_{1,3}^n(3), \underline{y}_{2,3}^n(3))$

But note that $\underline{X}_u^n(m')$ is independent of $\underline{Y}_{u,\overline{\mathcal{S}}}^n(m)$. The probability of (11) occurring is thus

$$\left|T_\delta(X_u|\, \underline{y}_{u,\overline{\mathcal{S}}}^n(m))\right| \Big/ \left|T_\delta(X_u)\right|. \tag{12}$$

We use Lemmas 2 and 3 in the Appendix to bound

$$|T_\delta(X_u)| \geq (1-\epsilon_\delta(n)) \cdot 2^{n(1-\delta)H(X_u)} \tag{13}$$

$$|T_\delta(X_u|\underline{y}_{u,\overline{\mathcal{S}}}^n(m))| \leq 2^{n(1+\delta_2)H\left(X_u|Y_{u,\overline{\mathcal{S}}}\right)} \tag{14}$$

where $\epsilon_\delta(n) \to 0$ as $n \to \infty$. Inserting (13) and (14) into (12), we have

$$\begin{aligned}\Pr&\left[\underline{y}_{u,\overline{\mathcal{S}}}^n(m) = \underline{y}_{u,\overline{\mathcal{S}}}^n(m')\right] \\ &\leq (1-\epsilon_\delta(n))^{-1} \cdot 2^{n(\delta+\delta_2)} \cdot 2^{-nH\left(Y_{u,\overline{\mathcal{S}}}\right)} \end{aligned} \tag{15}$$

where we have used $H(Y_{u,\overline{\mathcal{S}}}|X_u) = 0$. Inserting (15) into (10), and using $|\beta(\mathcal{S})| \leq |\mathcal{E}|$, we have

$$\begin{aligned}\Pr&\left[\mathcal{S}(m,m') = \mathcal{S}\right] \\ &\leq (1-\epsilon_\delta(n))^{-|\mathcal{E}|} \cdot 2^{n|\mathcal{E}|(\delta+\delta_2)} \cdot 2^{-n\mathrm{Value}(\mathcal{S})}. \end{aligned} \tag{16}$$

Inserting (16) into (8), and using the fact that the number of cuts is less than $2^{|\mathcal{V}|}$, we have

$$\begin{aligned}&\overline{P}_e(t,m,m') \\ &\leq (1-\epsilon_\delta(n))^{-|\mathcal{E}|} \cdot 2^{|\mathcal{V}|+n|\mathcal{E}|(\delta+\delta_2)} \cdot 2^{-n\min_{\mathcal{S}\in\Lambda_t}\mathrm{Value}(\mathcal{S})}. \end{aligned} \tag{17}$$

The above applies to the l-th block of transmission. We now add the index l to m_l. Let $\overline{P}_e(m)$ be the average probability of error when the (overall) message m was transmitted. We use the union bound over all L blocks, all destinations t, and all $m_l' \neq m_l$ to write

$$\begin{aligned}\overline{P}_e(m) &\leq \sum_{l=1}^{L}\sum_{t\in\mathcal{T}}\sum_{m_l'\neq m_l} \overline{P}_e(t,m_l,m_l') \\ &\leq L\cdot|\mathcal{T}|\cdot(2^{nR}-1)\cdot(1-\epsilon_\delta(n))^{-|\mathcal{E}|} \\ &\quad \cdot 2^{|\mathcal{V}|+n|\mathcal{E}|(\delta+\delta_2)} \cdot 2^{-n\min_{\mathcal{S}\in\Lambda_t}\mathrm{Value}(\mathcal{S})}. \end{aligned} \tag{18}$$

We thus find that the average error probability for any message can be made small if n is large and

$$R < -|\mathcal{E}|(\delta+\delta_2) + \min_{\mathcal{S}\in\Lambda_t}\mathrm{Value}(\mathcal{S}). \tag{19}$$

Finally, we optimize over all input distributions, choose δ and δ_2 small, and choose n and L large. The result is that we can make the overall rate approach C in (3) while at the same time ensuring that $P_e \leq \epsilon$ for any positive ϵ.

B. Converse

An Aref network is a special case of the networks described in [12, Sec. 14.10]. We can thus apply [12, Thm. 14.10.1] that we restate here.

Proposition 1: The multicasting capacity is bounded by

$$C \leq \max_{p_{X_1X_2\cdots X_{|\mathcal{V}|}}(\cdot)} \min_{\mathcal{S}\in\Lambda} I(\underline{X}_{\mathcal{S}}\,;\,\underline{Y}_{\mathcal{V},\overline{\mathcal{S}}}\mid\underline{X}_{\overline{\mathcal{S}}}). \tag{20}$$

Proof: See [12, Thm. 14.10.1]. ■

We next optimize the distribution in (20).

Lemma 1: For Aref networks, the bound (20) is optimized by independent inputs.

Proof: For any fixed $p_{X_1X_2\cdots X_{|\mathcal{V}|}}(\cdot)$, we have

$$\begin{aligned} I(\underline{X}_{\mathcal{S}}\,;\,\underline{Y}_{\mathcal{V},\overline{\mathcal{S}}}\mid\underline{X}_{\overline{\mathcal{S}}}) &= H(\underline{Y}_{\mathcal{V},\overline{\mathcal{S}}}\mid\underline{X}_{\overline{\mathcal{S}}}) \\ &\leq H(\underline{Y}_{\mathcal{V},\overline{\mathcal{S}}}) \\ &\leq \sum_{u\in\beta(\mathcal{S})} H(\underline{Y}_{u,\overline{\mathcal{S}}}) \end{aligned} \tag{21}$$

where the first step follows because the network is deterministic, and the other steps because conditioning cannot increase entropy. Furthermore, by replacing the input joint distribution by the product of its marginals, the mutual information in (20) is exactly the sum of the entropies in (21). That is, one can restrict attention to independent inputs. ■

We remark that both Proposition 1 and Lemma 1 apply to cyclic Aref networks as well as acyclic ones. Finally, note that the right-hand side of (21) is Value($\mathcal{S}$). Hence, we have the result that (3) is an upper bound on C.

C. Discussion

For the usual deterministic networks without broadcasting, one can show that the multicasting capacity is

$$C = \min_{t\in T} C_{1,t}$$

where $C_{1,t}$ is the *unicast* capacity from vertex 1 to vertex t. However, for Aref networks such a relationship is not necessarily true. Consider the network shown on the left in Fig. 3. Suppose the channel from x_1 to $y_{1,2}$ and $y_{1,3}$ is the broadcast channel shown on the right in Fig. 3. It is easy to see that $C_{1,2} = C_{1,3} = \log_2(3)$ bits by choosing 3 out of 4 of the input letters to have probability 1/3. The multicast capacity is, however, only $C = 1.5$, and is achieved only if all 4 inputs have probability 1/4.

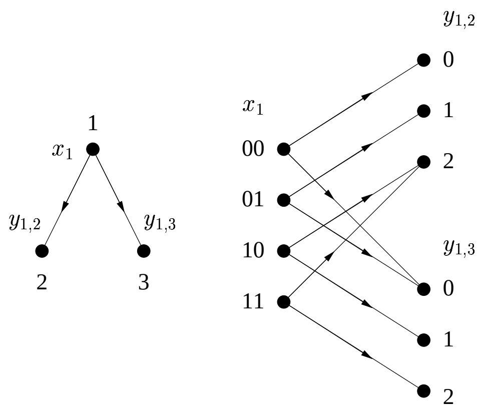

Fig. 3. An Aref network for which $C_{1,T} < \min_{t \in T} C_{1,t}$.

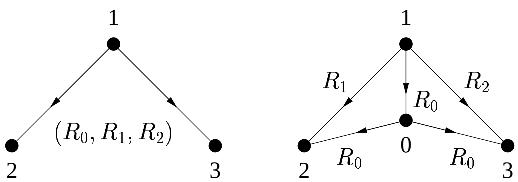

Fig. 4. An Aref network operating at rates (R_0, R_1, R_2) and its PPL network.

V. ON SEPARATING CHANNEL AND NETWORK CODING

A practically interesting question is whether one can *separate* channel and network coding. It is not so obvious how to define such separation in general networks, but for Aref networks a natural definition is as follows. First, on the *physical* layer one performs channel coding for each broadcast channel (BC) to obtain a network comprised of point-to-point links. Second, on the *network* layer one performs network coding for the resulting noise-free network. We consider more details of these layers.

Physical Layer. Suppose that vertex u has m outgoing edges. Each of the $2^m - 1$ non-empty subsets of vertices can be transmitted a different message, and all $2^m - 1$ messages are independent. The BC at vertex u thus has a $2^m - 1$ dimensional capacity region $\mathcal{C}_u$.

For example, consider the BC with $m = 2$ shown on the left in Fig. 4. The capacity region $\mathcal{C}_1$ is a $2^m - 1 = 3$ dimensional region. Suppose that $\underline{R}_1 = (R_0, R_1, R_2)$ is in $\mathcal{C}_1$, where R_0 is the common rate, R_1 the rate to vertex 2, and R_2 the rate to vertex 3. Given $\underline{R}_1$, the BC can effectively be replaced by a network of point-to-point links, PPL($\underline{R}_1$) shown on the right in Fig. 4. More generally, we need to introduce $2^m - 1 - m$ auxiliary vertices to convert a BC to a point-to-point link network.

We apply the above conversion to Aref networks, i.e., at every vertex u we replace the BC by a network PPL($\underline{R}_u$). We call the network thus obtained a Point-to-Point Links (PPL) network, and we denote its graph as $\mathcal{G}'$.

Network Layer. Network coding is performed on $\mathcal{G}'$ where it can achieve the cut-set bound. Code constructions are known, and have been investigated in detail in [16, 17].

A. Properties of Deterministic BCs

We list three properties of deterministic BCs.

Claim 1: The capacity of a discrete, memoryless, deterministic BC with $Y_{1,2} = h_{1,2}(X_1)$ and $Y_{1,3} = h_{1,3}(X_1)$ is the set of non-negative (R_0, R_1, R_2) satisfying

$$R_0 \le \min\{I(T;Y_{1,2}), I(T;Y_{1,3})\} \tag{22}$$
$$R_0 + R_1 \le H(Y_{1,2}) \tag{23}$$
$$R_0 + R_2 \le H(Y_{1,3}) \tag{24}$$
$$R_0 + R_1 + R_2 \le \min\{I(T;Y_{1,2}), I(T;Y_{1,3})\} + H(Y_{1,2}Y_{1,3}|T) \tag{25}$$

where T is arbitrary, but one can restrict attention to T with alphabet $\mathcal{T}$ satisfying $|\mathcal{T}| \le |\mathcal{X}_1| + 2$.

Proof: See Problem 4.11 in [18, p. 391], and [19, 20]. ■

Claim 2: In order to achieve

$$R_0 + R_1 + R_2 = H(Y_{1,2}Y_{1,3}) \tag{26}$$

in (25), the triple $(Y_{1,2}, Y_{1,3}, T)$ must form the following two Markov chains:

$$Y_{1,2} - Y_{1,3} - T \tag{27}$$
$$Y_{1,3} - Y_{1,2} - T. \tag{28}$$

Proof: For (25) and (26) to be the same, we must have

$$I(T;Y_{1,2}) + H(Y_{1,2}Y_{1,3}|T) \ge H(Y_{1,2}Y_{1,3}). \tag{29}$$

Rearranging terms, we obtain

$$I(T;Y_{1,2}) \ge I(T;Y_{1,2}Y_{1,3}). \tag{30}$$

Equivalently, we have

$$I(T;Y_{1,3}|Y_{1,2}) \le 0. \tag{31}$$

The bound (31) implies that (28) forms a Markov chain. Similarly, it can be shown that (27) forms a Markov chain. ■

We continue to write $A - B - C$ to refer to Markov chains.

Claim 3: The triple $(Y_{1,2}, Y_{1,3}, T)$ satisfies the double Markov relations (27) and (28) if and only if there exist functions $f(\cdot)$ of $Y_{1,2}$ and $g(\cdot)$ of $Y_{1,3}$ such that

$$\text{(i)} \quad f(Y_{1,2}) = g(Y_{1,3}) \quad \text{with probability 1} \tag{32}$$
$$\text{(ii)} \quad [Y_{1,2}, Y_{1,3}] - [f(Y_{1,2}), g(Y_{1,3})] - T. \tag{33}$$

Note that if $f(Y_{1,2}) = g(Y_{1,3})$ is a constant, then $[Y_{1,2}, Y_{1,3}]$ is independent of T.

Proof: See Problem 4.25 in [18, p. 402]. ■

B. A Counterexample

We show that layering can be suboptimal. Consider the Aref network in Fig. 5 that has the BC shown in Fig. 6 between vertices 1, 2, and 3. The meaning of the graph in Fig. 6 is that $b_j = h_{1,2}(a_i)$ if there is an edge (a_i, b_j). Similarly, $c_j = h_{1,3}(a_i)$ if there is an edge (a_i, c_j). The edges $(2,4)$, $(2,5)$, $(3,4)$, and $(3,5)$ in Fig. 5 represent point-to-point links with the labeled capacities.

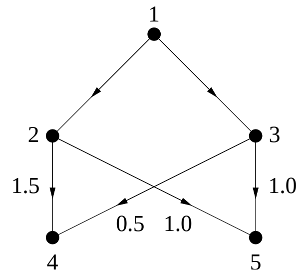

Fig. 5. An Aref network for which separating channel and network coding is suboptimal.

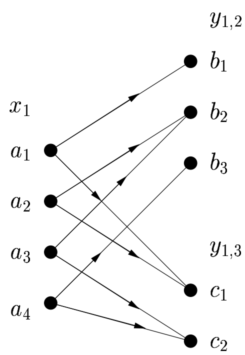

Fig. 6. The broadcast channel between vertices 1,2, and 3 in Fig. 5.

Claim 4: The multicast capacity of the network shown in Fig. 5 is $C = 2$, and one can achieve this rate only if the input X_1 is uniform.

Proof: Suppose X_1 is uniform. We compute $H(Y_{1,2}Y_{1,3}) = 2$, $H(Y_{1,2}) = 1.5$, and $H(Y_{1,3}) = 1$. We further find that $\min_{\mathcal{S}\in\Lambda} \text{Value}(\mathcal{S}) = 2$. However, we also have

$$C \leq \max_{P_{X_1}(\cdot)} H(X_1) = 2. \tag{34}$$

Moreover, if X_1 is not uniform, then $H(X_1) < 2$. ∎

We continue by restricting attention to uniform X_1.

Definition 1: (See Problem 3.12 in [18, p. 350].) The joint distribution the random variables X and Y is *indecomposable* if there are no functions $f(\cdot)$ and $g(\cdot)$ with respective domains $\mathcal{X}$ and $\mathcal{Y}$ so that

- $\Pr\{f(X) = g(Y)\} = 1$ and
- $f(X)$ takes at least two values with non-zero probability.

Claim 5: If the input distribution to the BC in Fig. 6 is uniform, then the joint distribution of $Y_{1,2}$ and $Y_{1,3}$ is indecomposable.

Proof: Suppose there are functions $f(\cdot)$ and $g(\cdot)$ with respective domains $\mathcal{Y}_{1,2}$ and $\mathcal{Y}_{1,3}$ and $\Pr\{f(Y_{1,2}) = g(Y_{1,3})\} = 1$ such that $g(c_1) \neq g(c_2)$. Since $f_{1,2}(a_1) = b_1$ and $f_{1,3}(a_1) = c_1$, it follows that $f(b_1) = g(c_1)$ and $f(b_2) = g(c_2)$. But this would imply $g(c_1) = g(c_2)$, contradicting our hypothesis. ∎

Next, let the point (R_0, R_1, R_2) be a point in the capacity region of the BC in Fig. 6. We perform channel coding as specified above, and obtain the PPL networks shown in Fig. 7.

Claim 6: The PPL networks in Fig. 7 all have multicast capacity less than 2.

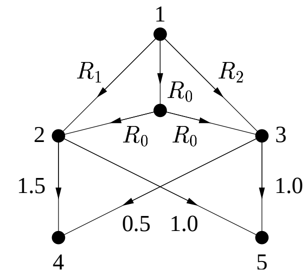

Fig. 7. The PPL network for the network for the Aref network in Fig. 5

Proof: One achieves a multicast rate of 2 only if

$$R_0 + R_1 + R_2 = H(Y_{1,2}Y_{1,3}) = 2.$$

It follows by Claims 2 and 3 that there exist functions $f(\cdot)$ and $g(\cdot)$ satisfying (32) and (33). However, by Claim 5 it follows that $g(\cdot)$ is constant and, hence, that T in Claim 1 is independent of $[Y_{1,2}, Y_{1,3}]$. Using (22), this means that $R_0 = 0$. Applying the cut-set bound for the cuts $\mathcal{S} = \{1, 3, 5\}$ and $\mathcal{S} = \{1, 2, 4\}$ with $R_0 = 0$, we obtain

$$R_1 + 0.5 \geq 2$$
$$R_2 + 1 \geq 2$$

These bounds imply that $R_1 + R_2 \geq 2.5$, which contradicts the condition $R_1 + R_2 \leq H(Y_{1,2}Y_{1,3}) = 2$. ∎

Claim 6 shows that layering cannot achieve a rate of 2.

VI. Conclusions

We have shown that the multicast capacity of acyclic Aref networks has a max-flow, min-cut interpretation. We have also shown that one cannot always separate (or layer) channel and network coding. The sub-optimality of a layering is in contrast to networks of discrete memoryless channels [9], and certain networks of two-way channels [8]. It remains to investigate whether the multicast capacity of Aref Networks with cycles also has a max-flow, min-cut interpretation.

Acknowledgments

The authors wish to thank Tracey Ho and Ralf Koetter for helpful discussions and suggestions.

Appendix A
Robustly Typical Sequences

Let X be a random variable with (discrete and finite) alphabet $\mathcal{X}$ and distribution $p_X(\cdot)$. Recall that the set of δ-typical sequences with respect to X is

$$T_\delta(X) = \big\{\underline{x} \in \mathcal{X}^n : \big|\pi_{\underline{x}}(a) - p_X(a)\big| \leq \delta \cdot p_X(a) \\ \text{for all } a \in \mathcal{X}\big\}.$$

For the technical lemmas below, we will also need to define the *support set* of X to be $\mathcal{S}_X = \{a \in \mathcal{X} : p_X(a) > 0\}$. Also, let μ_X be the smallest nonzero probability of $p_X(\cdot)$.

Similarly, suppose we have two random variables X and Y with respective (discrete and finite) alphabets $\mathcal{X}$ and $\mathcal{Y}$ and joint distribution $p_{XY}(\cdot)$. The set of δ-typical sequences with respect to X and Y is defined as

$$\begin{aligned} T_\delta(X,Y) = \big\{ &(\underline{x},\underline{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \\ &\Big| \pi_{\underline{x},\underline{y}}(a,b) - p_{XY}(a,b) \Big| \leq \delta \cdot p_{XY}(a,b) \\ &\text{for all } (a,b) \in \mathcal{X} \times \mathcal{Y} \big\} \end{aligned}$$

where $\pi_{\underline{x},\underline{y}}(a,b)$ is an empirical frequency defined in the obvious way. We define the support set of X and Y to be $\mathcal{S}_{X,Y} = \{(a,b) \in \mathcal{X} \times \mathcal{Y} : p_{XY}(a,b) > 0\}$. We define μ_{XY} to be the smallest nonzero probability of $p_{XY}(\cdot)$.

We define the set of conditionally δ-typical sequences as

$$T_\delta(Y|\underline{x}) = \{\underline{y} \in \mathcal{Y}^n : (\underline{x},\underline{y}) \in T_\delta(X,Y)\}.$$

The following properties of typical sequences and sets are proved in the appendix of [15].

Lemma 2: Let $0 < \delta \leq 1$ and

$$\epsilon_\delta(n) = 2|\mathcal{S}_X| e^{-\delta^2 \mu_X n/3}. \tag{35}$$

We have

$$(1 - \epsilon_\delta(n)) \cdot 2^{n(1-\delta)H(X)} \leq |T_\delta(X)| \leq 2^{n(1+\delta)H(X)}$$

Lemma 3: Let $0 < \delta_1 < \delta_2 \leq 1$ and

$$\epsilon_{\delta_1,\delta_2}(n) = 2|\mathcal{S}_{X,Y}| e^{-\frac{(\delta_2-\delta_1)^2}{1+\delta_1} \cdot \mu_{X,Y} n/3}.$$

We have

$$\begin{aligned} &(1 - \epsilon_{\delta_1,\delta_2}(n)) \cdot 2^{n(1-\delta_2)H(Y|X)} \leq |T_{\delta_2}(Y|\underline{x})| \\ &|T_{\delta_2}(Y|\underline{x})| \leq 2^{n(1+\delta_2)H(Y|X)} \end{aligned}$$

where the upper bound holds for every $\underline{x} \in \mathcal{X}^n$, and the lower bound holds for every $\underline{x} \in T_{\delta_1}(X)$.

Note that $\epsilon_\delta(n)$ and $\epsilon_{\delta_1,\delta_2}(n)$ approach zero exponentially with n.

Lemma 4: Suppose we have $Y = f(X)$, $\underline{x} \in T_\delta(X)$ and $\underline{y} = (f(x_1), f(x_2), \ldots, f(x_n))$. We then have $\underline{y} \in T_\delta(Y)$ and $(\underline{x},\underline{y}) \in T_\delta(X,Y)$.

Proof: We have $|\pi_{\underline{x}}(a) - p_X(a)| \leq \delta \cdot p_X(a)$ for all $a \in \mathcal{X}$. Consider a pair $(a,b) \in \mathcal{X} \times \mathcal{Y}$. We clearly have $\pi_{\underline{x},\underline{y}}(a,b) = \pi_{\underline{x}}(a)$, $p_{XY}(a,b) = p_X(a)$, and therefore $(\underline{x},\underline{y}) \in T_\delta(X,Y)$.

Next, observe that the following is true if $(\underline{x},\underline{y}) \in T_\delta(X,Y)$:

$$\begin{aligned} |\pi_{\underline{y}}(b) - p_Y(b)| &= \left| \sum_{a \in \mathcal{X}} \pi_{\underline{x},\underline{y}}(a,b) - p_{XY}(a,b) \right| \\ &\leq \sum_{a \in \mathcal{X}} \left| \pi_{\underline{x},\underline{y}}(a,b) - p_{XY}(a,b) \right| \\ &\leq \sum_{a \in \mathcal{X}} \delta \cdot p_{XY}(a,b) \\ &= \delta \cdot p_Y(b). \end{aligned}$$

Thus, we find that $(\underline{x},\underline{y}) \in T_\delta(X,Y)$ implies $\underline{y} \in T_\delta(Y)$. ■

REFERENCES

[1] M. R. Aref, “Information flow in relay networks,” Ph.D. dissertation, Stanford Univeristy, 1981.

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, “Network information flow,” *IEEE Transactions on Information Theory*, vol. IT-46, pp. 1204–1216, 2000.

[3] L. R. Ford and D. R. Fulkerson, “Maximal flow through a network,” *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.

[4] G. B. Dantzig and D. R. Fulkerson, “On the max-flow, min-cut theorem of networks,” *Linear Inequalities, Ann. Math. Studies*, no. 38, pp. 215–221, 1956.

[5] P. Elias, A. Feinstein, and C. E. Shannon, “A note on the maximum flow through a network,” *IRE Transactions on Information Theory*, vol. 2, pp. 117–119, December 1956.

[6] R. Koetter and M. Medard, “An algebraic approach to network coding,” *IEEE/ACM transactions on networking*, vol. 11, no. 5, October 2003.

[7] Z. Li and B. Li, “Network coding in undirected networks,” *Proc. 38th Annual Conf. on Information Sciences and Systems (CISS) (Princeton, NJ)*, pp. 257–262, March 17-19 2004.

[8] G. Kramer and S. Savari, “Cut sets and information flow in networks of two-way channels,” *Proc. ISIT,Chicago*, p. 33, 2004.

[9] S. P. Borade, “Network information flow: limits and achievability,” *Proc. 2002 IEEE Int. Symp. Inform. Theory, Lausanne, Switzerland*, p. 139, June 30 - July 5 2002.

[10] C. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[11] ——, “Two-way communication channels,” *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1, pp. 611–644, 1961.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, New York: John Wiley & Sons, Inc., 1991.

[13] R. Gowaikar, A. F. Dana, R. Palanki, B. Hassibi, and M. Effros, “On the capacity of wireless erasure networks,” *Proc. 2004 IEEE Int. Symp. Inform. Theory, Chicago*, p. 401, June 27 - July 2 2004.

[14] D. S. Lun, M. Mdard, and M. Effros, “On coding for reliable communication over packet networks,” *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, September - October 2004.

[15] A. Orlitsky and J. R. Roche, “Coding for computing,” in *IEEE Transactions on Information Theory*, vol. 47, no. 3, March 2001, pp. 903–917.

[16] S. Jaggi, P. Sanders, P. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhuizen, “Polynomial time algorithms for multicast network code construction,” submitted to IEEE Trans. Inform. Theory, 2003. [Online]. Available: citeseer.ist.psu.edu/jaggi03polynomial.html

[17] T. Ho, M. Medard, J. Shi, M. Effros, and D. Karger, “On randomized network coding,” proceedings of 41st Annual Allerton Conference on Communication, Control, and Computing, October 2003. [Online]. Available: citeseer.ist.psu.edu/ho03randomized.html

[18] I. Csiszar and J. Korner, *Information Theory: Coding Theory for Discrete Memoryless Systems*. Academic Press, NY, 1981.

[19] S. I. Gel’fand and M. S. Pinsker, “Capacity of a broadcast channel with one deterministic component,” *Problemy Peredachi Informatsii*, vol. 16, no. 1, pp. 24–34, Jan. - March 1980.

[20] T. S. Han, “The capacity region for the deterministic broadcast channel with a common message,” *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp. 122–125, January 1980.