Deterministic Relay Networks with State Information

Sung Hoon Lim
School of EECS
KAIST
Daejeon, Korea
Email: sunghlim@kaist.ac.kr

Young-Han Kim
Department of ECE
UCSD
La Jolla, CA 92093, USA
Email: yhk@ucsd.edu

Sae-Young Chung
School of EECS
KAIST
Daejeon, Korea
Email: sychung@ee.kaist.ac.kr

***Abstract*—Motivated by fading channels and erasure channels, the problem of reliable communication over deterministic relay networks is studied, in which relay nodes receive a function of the incoming signals and a random network state. An achievable rate is characterized for the case in which destination nodes have full knowledge of the state information. If the relay nodes receive a *linear* function of the incoming signals and the state in a finite field, then the achievable rate is shown to be optimal, meeting the cut-set upper bound on the capacity. This result generalizes on a unified framework the work of Avestimehr, Diggavi, and Tse on the deterministic networks *with state dependency*, the work of Dana, Gowaikar, Palanki, Hassibi, and Effros on linear erasure networks *with interference*, and the work of Smith and Vishwanath on linear erasure networks *with broadcast*.**

I. Introduction

In their celebrated paper [1] that opened the field of network coding, Ahlswede *et al.* found the multicast capacity of wireline networks. For wireless networks, however, there are some new challenges for reliable communication compared to the wireline network. Among them are *broadcast* and *interference*, and there has been some work that deals with these two features. In [2], the multicast capacity was shown for networks that have deterministic channels with broadcast, but without interference at the receivers. Deterministic networks were further studied in [3] to incorporate interference at the receiving nodes, where the capacity for linear finite field networks was found. These rather simple models were shown to give good insights in solving real-world network problems. For example in [4], Avestimehr *et al.* were able to approximately characterize the capacity of Gaussian relay networks within some constant gap using a similar approach used for deterministic networks. Although previous models consider broadcast and interference, they did not explicitly consider another important feature in wireless communications. The wireless medium in real-world communications suffer *fading*, which in turn cause severe degradation of the transmitted signal. Although the deterministic model can be a good abstraction in understanding broadcast and interference, it does not fully capture the effect of fading in wireless networks. In this sense, the erasure network in which transmitted symbols get erased at random provides a simple model that captures the fading characteristics. In [5], Dana *et al.* considered the erasure networks with broadcast and no interference, where the erasures are at the traversing edges. Smith and Vishwanath [6] considered an erasure network without broadcast, where the interference is modeled as a linear finite field sum of incoming signals that are not erased. In both [5] and [6], if the destination node has perfect knowledge of the state information, they showed that the capacity is given by the cut-set bound.

In this paper, we consider a deterministic network in which the observation at each node is a function of the incoming signals and a random state. The channel state affecting the relay and destination nodes is assumed to be perfectly known at the destinations. We give an achievable rate for this class of networks, and show that the associated coding scheme achieves the capacity for the case in which the relay and destination nodes receive a linear function of the incoming signals and the state over a finite field. This result generalizes the work of Dana *et al.* and the work of Smith and Vishwanath on linear erasure networks to handle both interference and broadcast. As for deterministic networks, our result generalizes the work of Avestimehr, Diggavi, and Tse on the deterministic networks to deterministic state-dependent networks.

II. Problem Statement and Preliminaries

In the following we will give useful definitions for later use. Upper case letters denote random variables (e.g., X, Y, S) and lower case letters represent scalars (e.g., x, y, s). Calligraphic letters (e.g., $\mathcal{A}$) denote sets and the cardinality of the set is denoted by $|\mathcal{A}|$. Subscripts are used to specify node and time indicies. For example, X_u and $X_{u,i}$ denotes the signal sent at node u and the signal sent at node u at time i, respectively. To represent a sequence of random variables we use the notation $X_v^n = X_{v,1}, \ldots, X_{v,n}$. We will frequently use random variables subscripted by sets to denote the set of random variables indexed with elements in the set. For example, $X_{\mathcal{A}} = \{X_a : a \in \mathcal{A}\}$ and $X_{\mathcal{A}}^n = \{X_a^n : a \in \mathcal{A}\}$.

We consider a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ and $\mathcal{E}$ are the set of nodes and directed edges, respectively. Without loss of generality, we let $\mathcal{V} = \{1, \ldots, |\mathcal{V}|\}$ and index the source node with 1. We use $\mathcal{D}$ and $\mathcal{R} = \mathcal{V} - (\{1\} \cup \mathcal{D})$ to denote the set of destination nodes and relay nodes respectively. The network has one channel input $X_u \in \mathcal{X}_u$ associated with each node $u \in \mathcal{V}$, where $\mathcal{X}_u$ is the alphabet of X_u. This incorporates the

978-1-4244-4313-0/09/$25.00 ©2009 IEEE

broadcast nature of the network. Each node $v \in \mathcal{V}$ observes

$$Y_v = f_v\left(X_{\mathcal{N}_v}, S\right), \quad (1)$$

where the input neighbors $\mathcal{N}_v$ of v is defined as $\mathcal{N}_v = \{u : (u,v) \in \mathcal{E}\}$. The random variable S is a random state affecting nodes, which is independent of the source message. The state sequence is memoryless and stationary with $p(s^n) = \prod_{i=1}^n p(s_i)$. We assume that each destination $d \in \mathcal{D}$ has side information of the state sequence. The source node wishes to send a common message $m \in [2^{nR}] \triangleq \{1, \ldots, 2^{nR}\}$ to all destination nodes.

A $(2^{nR}, n)$ code consists of a source encoding function ϕ_1, relay encoding functions $\phi_{v,i}$, $v \in \mathcal{V} - (\{1\} \cup \mathcal{D})$, $i \in \{1, \ldots, n\}$, and decoding functions ψ_d, $d \in \mathcal{D}$, where

$$\phi_1 : [2^{nR}] \rightarrow \mathcal{X}_1^n,$$
$$\phi_{v,i} : \mathcal{Y}_v^{i-1} \rightarrow \mathcal{X}_v, i \in \{1, \ldots, n\}, v \in \mathcal{R},$$
$$\psi_d : \mathcal{Y}_d^n \times \mathcal{S}^n \rightarrow [2^{nR}], \quad d \in \mathcal{D}$$

where M is uniformly distributed over $[2^{nR}]$. The probability of error is defined by

$$P_e^{(n)} = \Pr\{\psi_d(Y_d^n, S^n) \neq M \text{ for some } d \in \mathcal{D}\}.$$

A rate R is said to be *achievable* if there exist a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$.

For each $d \in \mathcal{D}$, a cut $\mathcal{U}_d \subset \mathcal{V}$ is a subset of nodes such that $1 \in \mathcal{U}_d$ and $d \in \mathcal{U}_d^c$. We will omit the destination index when it is clear from the context. We define a boundary of a cut as $\partial(\mathcal{U}) = \{u : (u,v) \in \mathcal{E}, u \in \mathcal{U}, v \in \mathcal{U}^c\}$ and the boundary of a complement of a cut as $\bar{\partial}(\mathcal{U}^c) = \{v : (u,v) \in \mathcal{E}, u \in \mathcal{U}, v \in \mathcal{U}^c\}$.

We say that a node v is in *layer* l if all directed paths from the source to v has l hops. Let L be the longest distance from the source node to any node. We say that a *network is layered* with L layers if every node in $\mathcal{V}$ belong to some layer $l \in \{0, \ldots, L\}$. The set of nodes in layer l is denoted by $\mathcal{V}_l$. Without loss of generality we will assume that $\mathcal{V}_0 = \{1\}$.

For a random variable $X \sim p(x)$, the set $T_\epsilon^{(n)}$ of ϵ-typical n-sequences x^n is defined [7] as

$$T_\epsilon^{(n)} \triangleq \{x^n : |\pi(a|x^n) - p(a)| \leq \delta \cdot p(a), \forall a \in \mathcal{X}\}$$

where $\pi(a|x^n)$ is the relative frequency of the symbol a in the sequence x^n.

III. Main Result

A. General state dependent networks

Given a class of relay networks as defined in (1), the multicast capacity C is upper bounded by

$$C \leq \max_{p(x_\mathcal{V})} \min_{d \in \mathcal{D}} \min_{\mathcal{U}_d} H(Y_{\mathcal{U}_d^c} | X_{\mathcal{U}_d^c}, S). \quad (2)$$

The upper bound is from the cut-set bound [8, Theorem 15.10.1] by treating the state information as additional outputs to the destinations, and using the fact that the state sequences are independent of the message, the memoryless property of the channel, and the deterministic nature of the channel given S.

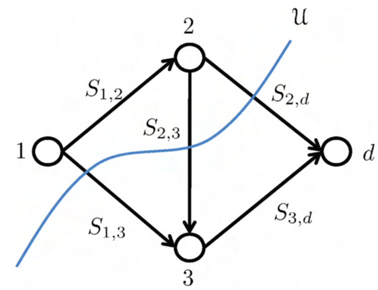

Fig. 1. Example of an erasure network. $S_{u,v}$ are erasures events for links $(u,v) \in \mathcal{E}$.

Remark 1: The cut-set bound is given by (2) whether we assume that the relay nodes have state information or not, as long as the state information at the relays are causal (i.e., $x_{v,i} = \phi_{v,i}(y_v^{i-1}, s^i)$) and destination nodes have the state information.

As our main result we state the following theorem.

Theorem 1: For the multicast relay network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in (1), if all destination nodes in $\mathcal{D}$ have side information of the state, then the capacity C of the network is lower bounded by

$$C \geq \max_{\prod_{i \in \mathcal{V}} p(x_i)} \min_{d \in \mathcal{D}} \min_{\mathcal{U}_d} H(Y_{\mathcal{U}_d^c} | X_{\mathcal{U}_d^c}, S). \quad (3)$$

The proof of this theorem will be given in Sections IV and V.

Remark 2: Theorem 1 includes the special case of unicast networks if $|\mathcal{D}| = 1$.

Example 1 ([5, Theorem 1]): Consider a network with output symbols $Y_v = \{Y_{u,v} : u \in \mathcal{N}_v\}$, where $Y_{u,v}$ is the observation at node v through the edge (u,v). Thus, the receiving nodes receives a separate output for each link connected to the node, i.e., has no interference. Let the output random variables take values from $\mathcal{Y} = \mathcal{X} \cup \{e\}$, where the symbol e is the erasure symbol. Each channel output $Y_{u,v}$ is given by the transmitted signal X_u with probability $1 - \epsilon_{u,v}$ or an erasure symbol e with probability $\epsilon_{u,v}$. Let $S_{u,v,i}$ be a random variable indicating erasure occurrence across channel $(u,v) \in \mathcal{E}$ at time i. If an erasure occurs on link $(u,v) \in \mathcal{E}$ at time i, the value of $S_{u,v,i}$ will be one, otherwise zero. Let $S^n = \{S_{u,v}^n : u \in \mathcal{N}_v\}$. If the destination nodes have the S^n sequence as side information, this channel falls into the channel model described in Section II since the output at each relay is a function of the incoming signals and S^n. It can be shown that the cut-set bound is achieved by the uniform product distribution. Hence, the capacity of this channel is given by (3) with equality.

B. Linear finite field fading networks

Consider a finite field $(GF(q))$ network in which each node $v \in \mathcal{V}$ observes

$$Y_v = \sum_{u \in \mathcal{N}_v} S_{u,v} X_u \quad (4)$$

where Y_v, $X_u, u \in \mathcal{N}_v$, $S_{u,v}, u \in \mathcal{N}_v$, are in $GF(q)$. If we assume that $S_{u,v}, \forall(u,v) \in \mathcal{E}$ is known at the destination nodes, this channel falls into the class of channels in Section II.

Let $\mathbf{Y}_{\mathcal{U}^c}$ and $\mathbf{X}_{\mathcal{U}}$ be vectors of observations in $\bar{\partial}(\mathcal{U}^c)$ and input signals in $\partial(\mathcal{U})$, respectively. These are of observations and input signals of nodes that have an edge passing through the cut. We define a transfer matrix of an arbitrary cut $\mathcal{U}$ as $\mathbf{G}_{\mathcal{U}}$ such that it satisfies $\mathbf{Y}_{\mathcal{U}^c} = \mathbf{G}_{\mathcal{U}}\mathbf{X}_{\mathcal{U}}$.

Thus, the random matrix $\mathbf{G}_{\mathcal{U}}$ consists of zeros when there is no connection between the nodes and $S_{u,v}$ if $u \in \partial(\mathcal{U})$ and $v \in \bar{\partial}(\mathcal{U}^c)$. The column index represents the sending node index in $\partial(\mathcal{U})$ and row index represents the receiving node index in $\bar{\partial}(\mathcal{U}^c)$. For the example in Figure 1, we have the expression

$$\underbrace{\begin{bmatrix} Y_3 \\ Y_d \end{bmatrix}}_{\mathbf{Y}_{\mathcal{U}^c}} = \underbrace{\begin{bmatrix} S_{1,3} & S_{2,3} \\ 0 & S_{2,d} \end{bmatrix}}_{\mathbf{G}_{\mathcal{U}}} \cdot \underbrace{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}}_{\mathbf{X}_{\mathcal{U}}}$$

for the cut $\mathcal{U} = \{1,2\}$.

Theorem 2: The multicast capacity of the linear finite field fading network (4) is

$$C = \min_{d \in \mathcal{D}} \min_{\mathcal{U}_d} E[\text{rank}(G_{\mathcal{U}_d})] \log q.$$

Proof: Proof is omitted due to space limitations. ■

Remark 3: For the special case of $S \in \{0,1\}$, Theorem 2 includes the capacity result for linear finite field erasure networks with broadcast and interference.

IV. Proof of Theorem 1 for Layered networks

We begin by showing the achievabililty of Theorem 1 for layered networks with $\mathcal{D} = \{d\}$. The multicast network is a simple extension of the single destination network and will be treated later.

We use a block Markov encoding scheme in which we divide the message m into K parts $m_k, k \in \{1,\ldots,K\}$. We code in $K+L-1$ blocks of length n. Message m_k takes values from $[2^{nR}]$ for all k and the overall rate is given by $\frac{RK}{(K+L-1)}$ which approaches R as $K \to \infty$.

We will use two types of indexing for the inputs, outputs, and state. We will use $s^n(j)$ to denote the state sequence when message m_j is being sent at the *source* node. For the set of observations and input sequences at layer l carrying message m_j, we will use the notation

$$y^n_{\mathcal{V}_l}(m_j) \triangleq \{y^n_v(x^n_{\mathcal{N}_v}(m_j), s^n(j+l)) : v \in \mathcal{V}_l\} \tag{5}$$

and

$$x^n_{\mathcal{V}_l}(m_j) \triangleq \{x^n_v(y^n_v(m_j)) : v \in \mathcal{V}_l\}, \tag{6}$$

respectively. For example, (5) denotes the set of observation sequences of the nodes in layer l when m_j is received. Due to the layered structure of the network and the coding strategy, which will be explained in the following, the observation sequences corresponding to the jth message at layer l are functions of $s^n(j+l)$. This will be explained in more detail in the following.

Codebook generation: Fix $p(x_u)$ for all $u \in \mathcal{V} - \{d\}$. Randomly and independently generate 2^{nR} sequences $x^n_1(m)$, $m \in [2^{nR}]$, each according to $\prod_{i=1}^n p(x_{1,i})$. For each $u \in \mathcal{V} - \{1\}$, randomly and independently generate $x^n_u(y^n_u)$ sequences for each $y^n_u \in \mathcal{Y}^n_u$, according to $\prod_{i=1}^n p(x_{u,i})$.

Encoding: To send message m_j, $j \in \{1,\ldots,K\}$, the encoder sends $x^n_1(m_j)$, while at each layer l, node $v \in \mathcal{V}_l$ sends $x^n_v(y^n_v(m_{j-l}))$.

Decoding: When the destination receives $y^n_d(m_j)$, it also has $\{s^n(1),\ldots,s^n(j+L)\}$ from previous observations. Assuming the previous blocks were decoded with arbitrarily small error, the receiver declares that a message was sent if it is a unique index $m_j \in [2^{nR}]$ such that

$$\bigcap_{l=0}^{L-1} \left\{ \left(x^n_{\mathcal{V}_l}(m_j), y^n_{\mathcal{V}_{l+1}}(m_j), s^n(j+l)\right) \in T^{(n)}_\epsilon \right\};$$

otherwise an error is declared.

From the encoding we can see that there is a l block delay at layer l, $l = \{1,\ldots,L\}$. When the source sends message m_j, the relays in layer 1 send $x^n_{\mathcal{V}_1}(m_{j-1})$, the relays in layer 2 send $x^n_{\mathcal{V}_2}(m_{j-2})$ and so on. Accordingly, when the source sends the jth block, received observation sequence of node $v \in \mathcal{V}_l$ is a function of $x^n_{\mathcal{N}_v}(m_{j-l-1})$ and $s^n(j)$, which gives (5). Table I shows the coding strategy for a simple diamond network given in Figure 2.

The decoding is a typicality check over an intersection of disjoint sets. Recall that from (5) and (6), as message m_j traverses through the network, the message is being affected by a different state at each layer. Therefore, we require that all inputs and outputs of that layer and a state (corresponding to the specific block time) are uniquely jointly typical.

Before dealing with arbitrarily large networks, we will first give a proof for a simple diamond network. Consider a diamond network depicted in Fig. 2 at the top of the next page. The relay nodes $\{a,b\}$ in layer 1 receives Y_a, Y_b which are deterministic functions of X_1 and S. The destination node in layer 2 observes Y_d, which is a deterministic function of X_a, X_b, and S. Without loss of generality, we will assume that $m_j = 1$ was sent, and show the decoding and probability of error analysis for the jth block. We will omit the message index for simplicity. There are two types of error events:

$$E_0 \triangleq (A^1_1 \cap A^1_2)^c \text{ and } E_1 \triangleq \bigcup_{m \neq 1} (A^m_1 \cap A^m_2)$$

where

$$A^m_1 \triangleq \left\{(X^n_1(m), Y^n_a(m), Y^n_b(m), S^n(j)) \in T^{(n)}_\epsilon\right\},$$

and

$$A^m_2 \triangleq \left\{(X^n_a(m), X^n_b(m), Y^n_d(1), S^n(j+1)) \in T^{(n)}_\epsilon\right\}.$$

For the first error event, we have $P(E_0) \to 0$ as $n \to \infty$ by the law of large numbers. We will decompose E_1 into four

TABLE I
CODING STRATEGY OF THE DETERMINISTIC DIAMOND NETWORK WITH STATE IN FIG. 2

Block index	Layer 0	Layer 1 observes		Layer 1 transmits		Layer 2 observes	State
j	$x_1^n(m_j)$	$y_a^n(m_j, s^n(j))$	$y_b^n(m_j, s^n(j))$	$x_a^n(m_{j-1})$	$x_b^n(m_{j-1})$	$y_d^n(m_{j-1}, s^n(j))$	$s^n(j)$
j+1	$x_1^n(m_{j+1})$	$y_a^n(m_{j+1}, s^n(j+1))$	$y_b^n(m_{j+1}, s^n(j+1))$	$x_a^n(m_j)$	$x_b^n(m_j)$	$y_d^n(m_j, s^n(j+1))$	$s^n(j+1)$
j+2	$x_1^n(m_{j+2})$	$y_a^n(m_{j+2}, s^n(j+2))$	$y_b^n(m_{j+2}, s^n(j+2))$	$x_a^n(m_{j+1})$	$x_b^n(m_{j+1})$	$y_d^n(m_{j+1}, s^n(j+2))$	$s^n(j+2)$

disjoint events. Let

$$B_{\mathcal{Q}}^m \triangleq \{Y_{\mathcal{Q}}^n(m) \neq Y_{\mathcal{Q}}^n(1), Y_{\mathcal{Q}^c}^n(m) = Y_{\mathcal{Q}^c}^n(1)\}$$

where $\mathcal{Q} \subseteq \{a,b\}$ and $\mathcal{Q}^c = \{a,b\} - \mathcal{Q}$. We have four such events since $\{a,b\}$ has four subsets. Then the probability of E_1 is given by

$$\begin{aligned} P(E_1) =& P\left\{\bigcup_{m\neq 1}(A_1^m \cap A_2^m)\right\} \\ \leq& \sum_{m\neq 1} P\{A_1^m \cap A_2^m\} \qquad (7) \\ =& \sum_{m\neq 1}\sum_{\mathcal{Q}\subseteq\{a,b\}} P\{A_1^m \cap A_2^m \cap B_{\mathcal{Q}}^m\} \qquad (8) \end{aligned}$$

where in (7) we have used the union bound and (8) is from the fact that $B_{\mathcal{Q}}^m$ are partitions that cover the whole set. Thus, we have decomposed E_1 into four disjoint events. The event $A_1^m \cap B_{\{a\}}^m$ implies

$$\left\{(X_1^n(m), Y_a^n(m), Y_b^n(1), S^n(j)) \in T_\epsilon^{(n)}\right\} \qquad (9)$$

and $A_2^m \cap B_{\{a\}}^m$ implies

$$\left\{(X_a^n(m), X_b^n(1), Y_d^n(1), S^n(j+1)) \in T_\epsilon^{(n)}\right\} \qquad (10)$$

since $X_b^n(m) = X_b^n(Y_b^n(m))$. Since (9) and (10) are independent events, we have

$$\begin{aligned} &P\{A_1^m \cap A_2^m \cap B_{\{a\}}^m\} \\ &\leq 2^{-n(I(X_1,Y_a;Y_b|S)-3\epsilon)}2^{-n(I(X_a;Y_d|X_b,S)-3\epsilon)} \\ &= 2^{-n(H(Y_b,Y_d|S,X_b)-6\epsilon)} \qquad (11) \end{aligned}$$

where in the last step we have used the Markov structure of the network. Similar to the previous steps, we can bound the other events by

$$P\{A_1^m \cap A_2^m \cap B_{\{b\}}^m\} \leq 2^{-n(H(Y_a,Y_d|S,X_a)-6\epsilon)}, \qquad (12)$$

$$P\{A_1^m \cap A_2^m \cap B_{\phi}^m\} \leq 2^{-n(H(Y_a,Y_b|S,X_a,X_b)-6\epsilon)}, \qquad (13)$$

and

$$P\{A_1^m \cap A_2^m \cap B_{\{a,b\}}^m\} \leq 2^{-n(H(Y_d|S)-3\epsilon)}. \qquad (14)$$

Combining (8), (11), (12), (13), and (14), we get $P(E_1) \to 0$ as $n \to \infty$ if

$$R < \min\left\{\begin{array}{c} H(Y_a,Y_b,Y_d|S,X_a,X_b) - 6\epsilon, \\ H(Y_d|S) - 3\epsilon, \\ H(Y_b,Y_d|S,X_b) - 6\epsilon, \\ H(Y_a,Y_d|S,X_a) - 6\epsilon \end{array}\right\}, \qquad (15)$$

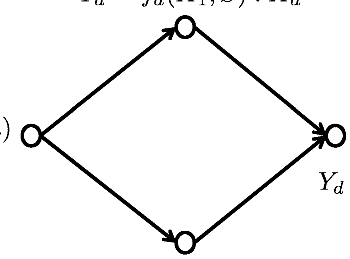

Fig. 2. Deterministic diamond network with state.

which concludes the proof for the diamond network.

We now move on to the proof of Theorem 1 for general layered networks. We will show the proof for decoding message m_j, and define similar events as in the diamond network to lead us through the proof. As before, we omit the message index for simplicity. Let

$$A_l^m \triangleq \left\{\left(X_{\mathcal{V}_l}^n(m), Y_{\mathcal{V}_{l+1}}^n(m), S^n(j+l)\right) \in T_\epsilon^{(n)}\right\}. \qquad (16)$$

Notice that we are abusing notation for the destination observation in (16). For A_{L-1}^m, $Y_{\mathcal{V}_L}^n(m)$ should be Y_d^n, which is the given observation at the destination and is not *tested* for typicality.

Assuming $m=1$ was sent, we have two sources of error:

$$E_0 \triangleq \left(\bigcap_{l=0}^{L-1} A_l^1\right)^c \text{ and } E_1 \triangleq \bigcup_{m\neq 1}\bigcap_{l=0}^{L-1} A_l^m.$$

The error event $P(E_0) \to 0$ as $n \to \infty$. As we did in the diamond network case we will decompose the error event E_1 with each $B_{\mathcal{Q}}^m$, $\mathcal{Q} \subseteq \mathcal{R}$. The probability of E_1 is given by

$$\begin{aligned} P(E_1) =& P\left\{\bigcup_{m\neq 1}\bigcap_{l=0}^{L-1} A_l^m\right\} \\ \leq& \sum_{m\neq 1} P\left\{\bigcap_{l=0}^{L-1} A_l^m\right\} \\ =& \sum_{m\neq 1}\sum_{\mathcal{Q}\subseteq\mathcal{R}} P\left\{\bigcap_{l=0}^{L-1} A_l^m \cap B_{\mathcal{Q}}^m\right\} \qquad (17) \end{aligned}$$

where the inequality is due to the union bound and the last step is due to partitioning the events. The event $A_l^m \cap B_{\mathcal{Q}}^m$

implies

$$\left\{\left(X^n_{\mathcal{Q}_l}(m), X^n_{\mathcal{Q}^c_l}(1), Y^n_{\mathcal{Q}_{l+1}}(m), Y^n_{\mathcal{Q}^c_{l+1}}(1), S^n(j+l)\right) \in T^{(n)}_\epsilon\right\}$$

where $\mathcal{Q}_l = \mathcal{V}_l \cap \mathcal{Q}$ and $\mathcal{Q}^c_l = \mathcal{V}_l - \mathcal{Q}_l$. Then,

$$P\left\{\bigcap_{l=0}^{L-1} A^m_l \cap B^m_{\mathcal{Q}}\right\} \le \prod_{l=0}^{L-1} 2^{-n(I(X_{\mathcal{Q}_l};Y_{\mathcal{Q}^c_{l+1}}|X_{\mathcal{Q}^c_l},S)-3\epsilon)}$$
$$= \prod_{l=0}^{L-1} 2^{-n(H(Y_{\mathcal{Q}^c_{l+1}}|X_{\mathcal{Q}^c_l},S)-3\epsilon)}. \quad (18)$$

From (17) and (18) we get

$$\begin{aligned} P(E_1) &\le \sum_{m\ne 1}\sum_{\mathcal{Q}\subseteq\mathcal{R}}\prod_{l=0}^{L-1} 2^{-n(H(Y_{\mathcal{Q}^c_{l+1}}|X_{\mathcal{Q}^c_l},S)-3\epsilon)} \\ &= \sum_{m\ne 1}\sum_{\mathcal{Q}\subseteq\mathcal{R}} 2^{-n\sum_{l=0}^{L-1}(H(Y_{\mathcal{Q}^c_{l+1}}|X_{\mathcal{Q}^c_l},S)-3\epsilon)} \\ &\le \sum_{\mathcal{Q}\subseteq\mathcal{R}} 2^{nR}2^{-n\sum_{l=0}^{L-1}(H(Y_{\mathcal{Q}^c_{l+1}}|X_{\mathcal{Q}^c_l},S)-3\epsilon)} \\ &= \sum_{\mathcal{Q}\subseteq\mathcal{R}} 2^{nR}2^{-n(H(Y_{\mathcal{U}^c}|X_{\mathcal{U}^c},S)-\epsilon')} \end{aligned}$$

where $\epsilon' = 3L\epsilon$ and $\mathcal{U}^c = \{\mathcal{Q}^c, d\}$ which gives a cut in the network. Thus, $P(E_1) \to 0$ as $n \to \infty$ if

$$R < \min_{\mathcal{U}} H(Y_{\mathcal{U}^c}|X_{\mathcal{U}^c},S) - \epsilon',$$

which proves Theorem 1 for layered networks with a single destination.

Remark 4: Consider a semi-deterministic layered network $\mathcal{G} = (\mathcal{V},\mathcal{E})$ where each node $v \in \mathcal{V} - \{d\}$ observes $Y_v = f_v(X_{\mathcal{N}_v}, Y_d)$ and the final destination gets $Y_d \sim p(y_d|x_{\mathcal{N}_d})$, i.e., a stochastic output. Using the coding scheme above we can show that all rates R that satisfies

$$R < \max_{\prod_{i\in\mathcal{V}} p(x_i)} \min_{\mathcal{U}} I(X_{\mathcal{U}};Y_{\mathcal{U}^c}|X_{\mathcal{U}^c})$$

are achievable for unicast.

For the multicast scenario we declare an error if any of the nodes in $\mathcal{D}$ makes an error. Using the union bound and the same line of proof as in Section IV for each $d \in \mathcal{D}$, we can show that the probability of error is arbitrarily small for sufficiently large n if

$$R < \max_{\prod_{i\in\mathcal{V}} p(x_i)} \min_{d\in\mathcal{D}} \min_{\mathcal{U}_d} H(Y_{\mathcal{U}^c_d}|X_{\mathcal{U}^c_d},S).$$

V. Arbitrary Networks

For extending the layered network result to arbitrary networks we use the same line of proof as done in [3] that unfolds $\mathcal{G}$ into a time-extended network. We will just give an outline of the proof. For more details on unfolding $\mathcal{G}$, we refer to [3] due to space limitations. Given an arbitrary network $\mathcal{G}$, we unfold the original network over T stages to get a layered network $\bar{\mathcal{G}}$. Using the coding scheme for the unfolded layered network, we can achieve

$$R < \frac{1}{T} \max_{\prod_{i\in\mathcal{V}} p(x_i)} \min_{\bar{\mathcal{U}}} H(Y_{\bar{\mathcal{U}}^c}|X_{\bar{\mathcal{U}}^c},S) \quad (19)$$

where $\bar{\mathcal{U}}$ is a cut in the unfolded network. We normalize the right hand side by T since the network gives at most T duplicate paths of the original network. Using Lemma 6.2 in [3] (by including a state random variable in the conditional entropies) we have the relation

$$(T+N-1)\min_{\mathcal{U}} H(Y_{\mathcal{U}^c}|X_{\mathcal{U}^c},S) \le H(Y_{\bar{\mathcal{U}}^c}|X_{\bar{\mathcal{U}}^c},S) \quad (20)$$

where $N = 2^{|\mathcal{V}|-2}$. We also have for any distribution,

$$\min_{\bar{\mathcal{U}}} H(Y_{\bar{\mathcal{U}}^c}|X_{\bar{\mathcal{U}}^c},S) \le T\min_{\mathcal{U}} H(Y_{\mathcal{U}^c}|X_{\mathcal{U}^c},S), \quad (21)$$

since the right hand side corresponds to taking the minimum over only steady cuts (subset of all possible cuts). Combining (20) with (21) we have

$$\lim_{T\to\infty}\frac{1}{T}\max_{\prod_{v\in\mathcal{V}} p(x_v)}\min_{\bar{\mathcal{U}}} H(Y_{\bar{\mathcal{U}}^c}|X_{\bar{\mathcal{U}}^c},S)$$
$$\le \max_{\prod_{v\in\mathcal{V}} p(x_v)}\min_{\mathcal{U}} H(Y_{\mathcal{U}^c}|X_{\mathcal{U}^c},S).$$

Finally, with the relations (19) and (21), we can show that rates arbitrary close to the right hand side of (3) are achievable for sufficiently large T.

Acknowledgment

The work of Sung Hoon Lim and Sae-Young Chung is partially supported by the MKE, Korea, under the ITRC support program supervised by the IITA (IITA-2009-C1090-0902-0005), and the work of Young-Han Kim is partially supported by the National Science Foundation CAREER award CCF-0747111.

References

[1] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, “Network information flow,” *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul 2000.

[2] N. Ratnakar and G. Kramer, “The multicast capacity of deterministic relay networks with no interference,” *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2425–2432, June 2006.

[3] A. Avestimehr, S. N. Diggavi, and D. Tse, “Wireless network information flow,” in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sept. 2007.

[4] ——, “Approximate capacity of Gaussian relay networks,” in *Proc. IEEE Int. Symp. Information Theory*, Toronto, Ontario, Canada, 2008, pp. 474–478.

[5] A. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, “Capacity of wireless erasure networks,” *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789–804, March 2006.

[6] B. Smith and S. Vishwanath, “Unicast transmission over multiple access erasure networks: Capacity and duality,” in *Proc. IEEE Information Theory Workshop*, Tahoe City, California, Sept. 2007, pp. 331–336.

[7] A. Orlitsky and J. Roche, “Coding for computing,” *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar 2001.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.