On the capacity of wireless erasure relay networks*

Radhika Gowaikar, Amir Dana, Ravi Palanki, Babak Hassibi, Michelle Effros

Department of Electrical Engineering,

California Institute of Technology, Pasadena, CA 91125.

Email: {gowaikar,amirf,palanki,hassibi,effros}@caltech.edu

Abstract

In this paper we determine the capacity of a certain class of wireless erasure relay networks. We determine the "cut-capacity" for erasure networks with broadcast at transmission and no interference at reception. We show that with this a max-flow min-cut result holds for the capacity of these networks.

1 Introduction

The capacity for the most general multi-terminal network is an open problem. A max-flow min-cut outer bound for the capacity region is proposed in [1]. This bound is not tight even for some simple channels.

*This work was supported in part by the National Science Foundation under grant no. CCR-0133818, by the office of Naval Research under grant no. N00014-02-1-0578, and by Caltech's Lee Center for Advanced Networking.

However, if we have a wireline network with a single source and a single destination, this outer bound can be achieved. It has been shown that with a wireline network, even for a more general class of network problems, known as multicast problems, the max-flow min-cut outer bound can sometimes be achieved [2], [3], [4]. In the multicast problem the network consists of a source, several destinations, relay nodes and links or channels between some nodes. Each link represents a memoryless channel with some known capacity. All channels are assumed to be independent. The destinations wish to have some subset of the information that the source has. For some special cases of this the capacity region is known and has a max-flow min-cut type interpretation.

In this paper we consider the single source, single destination scenario for a particular class of wireless networks. We look at networks in which each link is modelled as an independent and memoryless erasure channel with no delay. We have broadcast in the network i.e. a node must send out the same message on each outgoing link. This models a wireless system. However, for reception we assume a multiple access model without interference i.e. messages coming in to a node from different incoming links do not act as noise for each other. In general, this is not true for a wireless system. However, this can be realized through some time/frequency/code division multiple access scheme.

Finally, we assume that the side-information regarding erasure locations on each link is available to the destination. If we assume that the erasure network operates on long packets, i.e., packets are either erased or received exactly on each link, then this assumption can be justified as follows. Assume that we are transmitting long packets. Then we can introduce a header for each packet which will have the information about which packets have been erased. This information will be carried through to the destination. By making the packets very long the overhead of transmitting the headers can be made negligible compared to the packet length.

In this paper, however, we work with bits and assume that the side-information regarding erasure locations is available to the destination exactly. Our results go through in either case.

We will show a max-flow min-cut type capacity result for these wireless erasure relay networks under the assumptions mentioned above. Although this paper only concerns itself with the single source, single destination scenario, similar results will go through for some multicast settings also. Many results parallel to those of [4] for this wireless setting can be found in [5].

This paper is organized as follows. We introduce the model and the problem statement in Section 2. Section 4 states the main result and sections 5 and 6 prove it. In section 7 we state without proof some more results related to this problem and conclude in section 8.

2 Model

We assume that the relay network is modeled by a directed graph $G = (V, E)$ where $V = \{v_1, \ldots, v_{|V|}\}$ is the set of vertices and $E \subset V \times V$ is the set of directed edges. We assume that the graph has no cycles. Each edge $(v_i, v_j) \in E$ represents a memoryless erasure channel from v_i to v_j with erasure probability $\epsilon_{i,j}$ associated with it. All channels are assumed independent and operate without delay.

Let $s \in V$ be the source node that wishes to transmit a message to $d(\neq s) \in V$ which is the destination node. Without loss of generality, let $s = v_1$ and $d = v_{|V|}$. (We assume $|V| \geq 2$.) All the nodes in $V - \{s, d\}$ are relay nodes and are to be used to aid communication from s to d.

We define the following sets.

$\Gamma_O(v_i) = \{(v_i, v_j)|(v_i, v_j) \in E\}$, i.e., the set of all edges leaving v_i

$\Gamma_I(v_i) = \{(v_j, v_i)|(v_j, v_i) \in E\}$, i.e., the set of all edges going in to v_i.

We now describe the transmission and reception models. We incorporate broadcast in our network by insisting that vertex v_i transmit the same symbol on all outbound edges i.e. all edges in $\Gamma_O(i)$ carry the same symbol X_i which can be a sequence of bits. On each edge (v_i, v_j), X_i is distorted since some bits may get erased. Let the channel random variable be denoted by $s_{i,j}$. This is a string of bits of the same length as X_i. Each entry is chosed i.i.d. from a Bernoulli distribution with parameter $\epsilon_{i,j}$. The received signal on edge (v_i, v_j) is denoted by $Y_{i,j}$. It consists of bits and erasures and is obtained by replacing with erasures those bits of X_i where $s_{i,j}$ is 1.

For reception, we assume that v_i receives the symbols from each edge in $\Gamma_I(i)$ without interference i.e. for every v_j such that $(v_j, v_i) \in \Gamma_I(i)$, $Y_{j,i}$ is received at v_i. Denote this by Y_i. Thus $Y_i = (Y_{j,i}, (v_j, v_i) \in \Gamma_I(i))$. The relationship between the Y_is and the X_is would define a coding scheme.

Let $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{0, 1, e\}$ where e represents an erasure. These represent the input and output alphabet. If we restrict the X_is to be strings of length n, we have $X_i \in \mathcal{X}^n$, $Y_{i,j} \in \mathcal{Y}^n$ and $Y_i \in \mathcal{Y}_i^n$ where $\mathcal{Y}_i = \mathcal{Y}^{|\Gamma_I(i)|}$.

We will assume that the erasure locations on each link are known to d i.e. it knows the values of the $s_{i,j}$ vectors perfectly. For the erasure channel, this means that perfect channel side-information from across the network is available to d.

3 Formal Definitions

A $(2^{nR}, n)$ code for the single source, single destination network described above consists of the following things.

1. A set of integers $\Omega = \{1, 2, \ldots, \lceil 2^{nR} \rceil\}$ that represent message indices. We assume all messages to be equally likely.

2. An encoding function for the source

$$f_1 : \Omega \rightarrow \mathcal{X}^n$$

   If message k is to be sent, the source sends out the codeword $X_1 = f_1(k)$. The codebook $\mathcal{C}$ is the set of all codewords, viz., $\{f_1(1), f_1(2), \ldots, f_1(\lceil 2^{nR} \rceil\}$.

3. An encoding function f_i for each relay node v_i in the network, i.e., $v_i \in V - \{s, d\}$.

$$f_i : \mathcal{Y}_i^n \rightarrow \mathcal{X}^n$$

   The signal transmitted by the relay node v_i will be $X_i = f_i(Y_i)$ for $i = \{2, 3, \ldots, |V|-1\}$.

4. A decoding function

$$g : \mathcal{Y}_{|V|}^n \rightarrow \Omega$$

If message k is being transmitted, X_i, Y_i and $Y_{i,j}$ will all be functions of k and will be written as $X_i(k)$, $Y_i(k)$ and $Y_{i,j}(k)$ respectively. We will define $Y_1(k) = k$ for uniformity of notation.

If a message $k \in \Omega$ is sent, let

$$\lambda_k = \Pr(g(Y_{|V|}(k)) \neq k)$$

The average probability of error is defined as

$$P_e^{(n)} = \frac{1}{|\Omega|} \sum_{k=1}^{|\Omega|} \lambda_k$$

The rate of the above code is R. A rate R is achievable if there exists a sequence of $(2^{nR}, n)$ codes such that the average probability of error tends to 0 as $n \to \infty$.

4 Main Result

An $s-d$ *cut* is defined as a partition of the vertex set V into two subsets V_s and $V_d = V - V_s$ such that $s \in V_s$ and $d \in V_d$. Clearly, an $s-d$ cut is determined simply by V_s. For the $s-d$ cut given by V_s, let the *cutset* $E(V_s)$ be the set of edges defined below

$$E(V_s) = \{(v_i, v_j) | (v_i, v_j) \in E, v_i \in V_s, v_j \in V_d\}$$

Denote by $W(V_s)$ the *value* of an $s-d$ cut given by V_s. We define $W(V_s)$ below.

$$W(V_s) = \sum_{i:(v_i,v_j)\in E(V_s)} \left(1 - \prod_{j:(v_i,v_j)\in E(V_s)} \epsilon_{i,j}\right) \tag{1}$$

Theorem 1. *The capacity of the erasure relay network described above is given by the value* *of the cut with minimum value.*

$$C = \min_{V_s} W(V_s)$$

where V_s determines an $s-d$ cut.

To put this result in perspective, consider the max-flow min-cut results of [2] in the single source, single destination setting. The above result is of exactly the same flavour. However, it is the definition of the value of the cut or the "cut-capacity" that makes it look different in form. In the wireline case, the value of the cut is simply the sum of the capacities of each edge in $E(V_s)$. Since our system model incorporates broadcast the cut-capacity is the sum of the capacities of each broadcast system that operates across the cut. This gives the outer summation in the definition of $W(V_s)$ above. The capacity of each broadcast system that operates across the cut is obtained by assuming that the receiver nodes within that system get to co-operate and hence have an effective erasure probability ϵ_{eff} equal to the product term in (1). Hence the capacity of each broadcast system is $(1-\epsilon_{\text{eff}})$ which gives the quantity inside the summation.

5 Achievability

In this section we show that all rates $R < C$ are achievable. We obtain the functions f_i as follows

- *Codebook generation* For each argument, $f_1(\cdot)$ is chosen independently and randomly assuming a uniform distribution on $\mathcal{X}^n$. The function f_1, i.e., codebook is known to the destination d.

- *Encoding functions at relay nodes* For each argument, $f_i(\cdot)$ is chosen independently and randomly assuming a uniform distribution on $\mathcal{X}^n$. Each of these functions is known to the destination d.

Now consider the following sequence of events.

1. *Encoding at source* To send message k, the source s transmits $X_1(k) = f_1(k)$.

2. *Encoding at relay nodes* Once the relay node v_i receives $Y_i(k)$, it transmits $X_i(k) = f_i(Y_i(k))$

3. *Decoding by destination* Let the signal received by the destination when message k is being transmitted be denoted by $y(k)$. Recall that the erasure pattern $s_{i,j}$ from each edge (v_i, v_j) is assumed to be known to the destination. Therefore for every $l \in \Omega$ the destination can compute the values of $X_i(l)$, $Y_{i,j}(l)$ and $Y_i(l)$ if l were the message that had been transmitted. If $\exists$ a unique $r \in \Omega$ such that the computed value of $Y_{|V|}(r)$ is equal to the value of the received signal $y(k)$, then r is declared as the decoder output. Otherwise, an error is declared. Denote this decoder by $\mathcal{D}_1$ and the function it represents by $g_1(\cdot)$.

   Note that the computed value $Y_{|V|}(k)$ will always match the value of the received signal $y(k)$. Therefore an error is declared only when $\exists r \neq k$ such that the computed values $Y_{|V|}(r)$ and $Y_{|V|}(k)$ are equal (and equal to $y(k)$), i.e., $y(k) = Y_{|V|}(k) = Y_{|V|}(r)$.

5.1 Probability of Error

To analyze the probability of error, we first describe a slightly different decoder $\mathcal{D}_2$. Consider the following typical sets, defined for each vertex v_i having outgoing edges.

$$A_{\delta_i}^{(n)}(i) = \{(s_{i,j}, (v_i, v_j) \in \Gamma_O(i)) | (s_{i,j}, (v_i, v_j) \in \Gamma_O(i)) \text{ are jointly } \delta_i\text{-typical}\}$$

Decoder $\mathcal{D}_2$ knows all $s_{i,j}$ from the network. If any set of $s_{i,j}$s of the form $(s_{i,j}, (v_i, v_j) \in \Gamma_O(i))$ does not belong to the corresponding typical set $A_{\delta_i}^{(n)}(i)$, an error is declared. Otherwise, a decoding operation identical to that of decoder $\mathcal{D}_1$ is performed. Denote this decoding function by g_2.

Clearly, decoding errors made by $\mathcal{D}_1$ are a subset of those made by $\mathcal{D}_2$. Hence the rate achievable by $\mathcal{D}_2$ is also achievable by $\mathcal{D}_1$.

We now analyze the probability of error with $\mathcal{D}_2$.

Let $P(\mathcal{E})$ be the probability of error averaged over all the functions f_i, i.e., $P_e^{(n)}$ averaged over all codebooks as well as all possible encoding functions being used by the relay nodes.

Because of the symmetry of the code construction, we have

$$P(\mathcal{E}) = P(\mathcal{E}|k = 1)$$

i.e. the average probability of error given that the first codeword was sent. We now calculate this probability.

Let T be the event that for all i such that $\Gamma_I(i) \neq \emptyset$ we have $(s_{i,j}, (v_i, v_j) \in \Gamma_O(i)) \in A_{\delta_i}^{(n)}(i)$. By making n sufficiently large we have $P(A_{\delta_i}^{(n)}(i)) = (1 - \delta_i)$ which can be made arbitrarily close to 1. Now

$$P(T) = \prod_{i:\Gamma_O(i)\neq\emptyset} P(A^{(n)}_{\delta_i}(i))$$

which can also be made arbitrarily close to 1. Let $P(T) = 1 - \Delta$.

If T does not occur decoder $\mathcal{D}_2$ declares an error. If T occurs, we have a decoding error only if $\exists l \neq 1$ such that $Y_{|V|}(1) = Y_{|V|}(l) = y(1)$. Define the event E_l as follows.

$$E_l = \{Y_{|V|}(1) = Y_{|V|}(l)\}$$

i.e. the received messages are identical for messages 1 and l.

$$\begin{aligned}
P(\mathcal{E}) &= P(\mathcal{E}|k=1) \\
&= P(\bigcup_{l=2}^{|\Omega|} E_l) \\
&= P(\bigcup_{l=2}^{|\Omega|} E_l|T^c)P(T^c) + P(\bigcup_{l=2}^{|\Omega|} E_l|T)P(T) \\
&\leq P(T^c) + P(\bigcup_{l=2}^{|\Omega|} E_l|T)P(T) \\
&\leq P(T^c) + P(T)\sum_{l=2}^{|\Omega|} P(E_l|T) \qquad (2)
\end{aligned}$$

We now consider the event $\{E_l|T\} = \{Y_{|V|}(1) = Y_{|V|}(l)|T\}$. For $l \neq 1$, $Y_1(1) \neq Y_1(l)$ holds trivially. Thus event E_l for $l \neq 1$ tells us that the values of $Y_1(1)$ and $Y_1(l)$ are unequal while those of $Y_{|V|}(1)$ and $Y_{|V|}(l)$ are equal. For all the other Y_is the values of $Y_i(1)$ and $Y_i(l)$ may either be equal or unequal.

Recall V_s from the definition of an $s-d$ cut. Consider a particular set V_s. Assume that

$\forall v_i \in V_s$ the values of $Y_i(1)$ and $Y_i(l)$ are unequal and $\forall v_j \in V_d$ the values of $Y_j(1)$ and $Y_j(l)$ are equal. If we now consider all possible sets V_s we will have considered all possible ways in which E_l can occur. Note also that all of these possibilities are mutually disjoint. We abbreviate "such that" with "s.t." below.

$$
\begin{aligned}
& P(E_l|T) \\
= \; & P(Y_{|V|}(1) = Y_{|V|}(l)|T) \\
= \; & P(Y_1(1) \neq Y_1(l), Y_{|V|}(1) = Y_{|V|}(l)|T) \\
= \; & \sum_{V_s : V_s \text{ is a cut}} P(Y_i(1) \neq Y_i(l) \forall i \text{ s.t.} v_i \in V_s, Y_j(1) = Y_j(l) \forall j \text{ s.t.} v_j \in V_d|T) \\
\leq \; & \sum_{V_s : V_s \text{ is a cut}} P((Y_i(1) \neq Y_i(l), Y_j(1) = Y_j(l)) \forall (i,j) \text{ s.t.} (v_i, v_j) \in E(V_s)|T) \\
\leq \; & \sum_{V_s : V_s \text{ is a cut}} P((Y_i(1) \neq Y_i(l), Y_{i,j}(1) = Y_{i,j}(l)) \forall (i,j) \text{ s.t.} (v_i, v_j) \in E(V_s)|T) \\
= \; & \sum_{V_s : V_s \text{ is a cut}} P(Y_{i,j}(1) = Y_{i,j}(l)) \forall (i,j) \text{ s.t.} (v_i, v_j) \in E(V_s)|T, Y_i(1) \neq Y_i(l) \forall i \text{ s.t.} (v_i, v_j) \in E(V_s)) \cdot \\
& \qquad\qquad P(Y_i(1) \neq Y_i(l) \forall i \text{ s.t.} (v_i, v_j) \in E(V_s)) \\
\leq \; & \sum_{V_s : V_s \text{ is a cut}} P(Y_{i,j}(1) = Y_{i,j}(l)) \forall (i,j) \text{ s.t.} (v_i, v_j) \in E(V_s)|T, Y_i(1) \neq Y_i(l) \forall i \text{ s.t.} (v_i, v_j) \in E(V_s)) \\
\leq \; & \sum_{V_s : V_s \text{ is a cut}} \prod_{i:(v_i,v_j) \in E(V_s)} P(Y_{i,j}(1) = Y_{i,j}(l) \forall j \text{ s.t.} (v_i, v_j) \in E(V_s)|T, Y_i(1) \neq Y_i(l))
\end{aligned} \tag{3}
$$

For a fixed i and V_s, consider the event

$$F_i(V_s) = \{Y_{i,j}(1) = Y_{i,j}(l), \forall j \text{ such that } (v_i, v_j) \in E(V_s)|T, Y_i(1) \neq Y_i(l)\}$$

Since $Y_i(1) \neq Y_i(l)$, we know that $X_i(1)$ and $X_i(l)$ will be chosen independently from a uniform distribution on $\{0,1\}^n$. The probability that $X_i(1)$ and $X_i(l)$ differ in at most α places is given by $2^{-(n-\alpha)}$.

For a fixed i we will have $Y_{i,j}(1) = Y_{i,j}(l)$ for every j such that $(v_i, v_j) \in E(V_s)$ only if all the bits where $X_i(1)$ and $X_i(l)$ differ get erased on all these edges. For a fixed i, since T occurs, we know that the number of bits that are erased on every j such that $(v_i, v_j) \in E(V_s)$ is close to its expected value ,i.e., of the order of $n \cdot d(V_s, i)$ where

$$d(V_s, i) = \prod_{(v_i,v_j)\in E(V_s)} \epsilon_{i,j}.$$

Let this range around $d(V_s, i)$ that is required for typicality according to the definition of set $A^{(n)}_{\delta_i}(i)$ be given by $(d(V_s, i) - d_1(V_s, i), d(V_s, i) + d_2(V_s, i))$. Note that $d_1(V_s, i)$ and $d_2(V_s, i)$ depend on δ_i and $\epsilon_{i,j}$s and are independent of n. Therefore the occurence of T ensures that at most $n(d(V_s, i) + d_2(V_s, i))$ bits will be erased on all the edges of interest. Hence $F_i(V_s)$ can occur only if $X_i(1)$ and $X_i(l)$ differ in no more than $n(d(V_s, i) + d_2(V_s, i))$ places. The probability of this is $2^{-(n-n(d(V_s,i)+d_2(V_s,i)))}$. Therefore we can conclude that $P(F_i(V_s)) \leq 2^{-(n-n(d(V_s,i)+d_2(V_s,i)))}$. This argument is formalised below.

Let t_m for $m = 1, 2, \ldots, N$ denote all the distinct values that the vectors $s_{i,j}$ can take such that T occurs. Clearly, the t_ms represent disjoint events and $P(T) = \sum_{m=1}^{N} P(t_m)$. For each of these typical instantiations, t_m, we know that the number of bits erased on all $(v_i, v_j) \in E(V_s)$ emanating from a fixed i cannot exceed $n(d(V_s, i)+d_2(V_s, i))$ The probability of event $F_i(V_s)$ can now be calculated.

$$\begin{aligned}
P(F_i(V_s)) &= P(Y_{i,j}(1) = Y_{i,j}(l) \forall (v_i, v_j) \in E(V_s) | T, Y_i(1) \neq Y_i(l)) \\
&= \frac{(P(Y_{i,j}(1) = Y_{i,j}(l) \forall (v_i, v_j) \in E(V_s)), T, Y_i(1) \neq Y_i(l))}{P(T, Y_i(1) \neq Y_i(l))} \\
&= \frac{\sum_{m=1}^{N} P((Y_{i,j}(1) = Y_{i,j}(l) \forall (v_i, v_j) \in E(V_s)), t_m, Y_i(1) \neq Y_i(l))}{P(T, Y_i(1) \neq Y_i(l))} \\
&= \frac{\sum_{m=1}^{N} P(Y_{i,j}(1) = Y_{i,j}(l) \forall (v_i, v_j) \in E(V_s) | t_m, Y_i(1) \neq Y_i(l)) \cdot P(t_m, Y_i(1) \neq Y_i(l))}{P(T, Y_i(1) \neq Y_i(l))} \\
&\leq 2^{-n(1-d(V_s,i)-d_2(V_s,i))} \frac{\sum_{m=1}^{N} P(t_m, Y_i(1) \neq Y_i(l))}{P(T, Y_i(1) \neq Y_i(l))} \\
&= 2^{-n(1-d(V_s,i)-d_2(V_s,i))}
\end{aligned}$$

Substituting this in the expression in (3) we get

$$\begin{aligned}
P(E_l|T) &\leq \sum_{V_s : V_s \text{ is a cut}} \prod_{i:(v_i,v_j) \in E(V_s)} 2^{-n(1-d(V_s,i)-d_2(V_s,i))} \\
&= \sum_{V_s : V_s \text{ is a cut}} 2^{-n \sum_{i:(v_i,v_j) \in E(V_s)} (1-d(V_s,i)-d_2(V_s,i))} \\
&= \sum_{V_s : V_s \text{ is a cut}} 2^{-n(W(V_s)-d_2(V_s))}
\end{aligned}$$

where

$$d_2(V_s) = \sum_{i:(v_i,v_j) \in E(V_s)} d_2(V_s, i).$$

and $W(V_s)$ is the value of the cut as defined earlier.

Note that this upperbound on $P(E_l|T)$ is independent of l. Substituting this back in (2) we get

$$
\begin{aligned}
P(\mathcal{E}) &\leq P(T^c) + P(T)(2^{nR} - 1) \sum_{V_s : V_s \text{ is a cut}} 2^{-n(W(V_s) - d_2(V_s))} \\
&\leq P(T^c) + P(T) 2^{nR} \sum_{V_s : V_s \text{ is a cut}} 2^{-n(W(V_s) - d_2(V_s))} \\
&= \Delta + (1 - \Delta) \sum_{V_s : V_s \text{ is a cut}} 2^{n(R - W(V_s) + d_2(V_s))}
\end{aligned}
$$

Note that $d_2(V_s, i)$ (and $d_1(V_s, i)$) go to 0 as δ_i goes to 0 provided we let n grow without bound. With this, Δ as well as $d_2(V_s)$ can be made arbitrarily close to 0. Therefore, if $R < W(V_s)$ for every cut V_s in the network, we can make $P(\mathcal{E})$ arbitrarily small by letting n grow without bound. But since $P(\mathcal{E})$ is the value of $P_e^{(n)}$ averaged over all possible codebooks and encoding functions, there must be some specific set of f_is such that for that set $P_e^{(n)}$ goes to zero. Thus the rate

$$R < \min_{V_s : V_s \text{ is a cut}} W(V_s)$$

is achievable.

6 Converse

In this section we show that any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ must have $R \leq C$. Let M denote a random variable drawn assuming a uniform distribution on Ω. For each n bits, let M represent the transmitted message. Since M has a uniform distribution we have

$P_e^{(n)} = \Pr(g(Y_{|V|}(M) \neq M)$. Note also that $M - X_1 - Y_{|V|}$ is a Markov chain. Hence

$$\begin{aligned} nR &= H(W) \\ &= H(W|Y_{|V|}) + I(W;Y_{|V|}) \\ &\leq H(W|Y_{|V|}) + I(X_1;Y_{|V|}) \\ &\leq 1 + P_e^{(n)} nR + I(X_1;Y_{|V|}) \end{aligned} \tag{4}$$

where the last inequality comes from Fano's inequality [1]. Consider an $s-d$ cut given by V_s. Recall $E(V_s)$ defined earlier. We now define quantities $X(V_s)$ and $Y(V_s)$ for this $s-d$ cut.

$$X(V_s) = \{X_i | (v_i, v_j) \in E(V_s)\}$$

$$Y(V_s) = \{Y_{i,j} | (v_i, v_j) \in E(V_s)\}$$

It is easy to see that $P(Y_{|V|}|X_1, X(V_s)) = P(Y_{|V|}|X(V_s))$. Therefore $X_1 - X(V_s) - Y_{|V|}$ is a Markov chain. Similarly, $X(V_s) - Y(V_s) - Y_{|V|}$ is a Markov chain. We now have the following sequence of inequalities from the Data Processing Inequality [1].

$$\begin{aligned} I(X_1;Y_{|V|}) &\leq I(X(V_s);Y_{|V|}) \\ &\leq I(X(V_s);Y(V_s)) \end{aligned} \tag{5}$$

We now evaluate the quantity $I(X(V_s);Y(V_s))$.

$$
\begin{aligned}
& I(X(V_s);Y(V_s)) \\
= \;& H(Y(V_s)) - H(Y(V_s)|X(V_s)) \\
= \;& H(Y(V_s)) - H((s_{i,j} \forall (i,j) \text{ s.t.} (v_i, v_j) \in E(V_s))) \\
= \;& H(Y(V_s)) - \sum_{(i,j):(v_i,v_j)\in E(V_s)} H(s_{i,j}) \\
= \;& H(Y(V_s)) - \sum_{i:(v_i,v_j)\in E(V_s)} \sum_{j:(v_i,v_j)\in E(V_s)} H(s_{i,j}) \\
= \;& H(Y(V_s)) - \sum_{i:(v_i,v_j)\in E(V_s)} H(Y_{i,j} \forall j \text{ s.t.} (v_i, v_j) \in E(V_s)|X_i) \\
\leq \;& \sum_{i:(v_i,v_j)\in E(V_s)} H(Y_{i,j} \forall j \text{ s.t.} (v_i, v_j) \in E(V_s)) - \sum_{i:(v_i,v_j)\in E(V_s)} H(Y_{i,j} \forall j \text{ s.t.} (v_i, v_j) \in E(V_s)|X_i) \\
= \;& \sum_{i:(v_i,v_j)\in E(V_s)} I(Y_{i,j} \forall j \text{ s.t.} (v_i, v_j) \in E(V_s); X_i) \\
= \;& n \sum_{i:(v_i,v_j)\in E(V_s)} (1 - \prod_{j:(v_i,v_j)\in E(V_s)} \epsilon_{i,j}) \\
= \;& nW(V_s) \qquad (6)
\end{aligned}
$$

Putting together 4, 5 and 6, for every $s - d$ cut V_s we have

$$
\begin{aligned}
nR &\leq 1 + P_e^{(n)} nR + nW(V_s) \\
R &\leq \frac{1}{n} + P_e^{(n)} R + W(V_s)
\end{aligned}
$$

Since the first two terms go to 0 as $n \to \infty$ we have

$$
\begin{aligned}
R &\leq W(V_s) \quad \forall V_s \\
\Longrightarrow R &\leq \min_{V_s : V_s \text{ is a cut}} W(V_s) = C
\end{aligned}
$$

7 Comments

The above proof shows that a scheme using random functions to determine the transmitted codewords at each node achieves capacity for the wireless relay networks under consideration. In [5] we show that if the randomly chosen functions are *linear* we can still reach capacity.

Furthermore, several of the multicast results of [4] also go through for the class of wireless erasure relay networks that we have defined. In particular, results regarding the case where several destination nodes wish to have the same information and the case where all destinations wish to have completely disjoint pieces of information go through for the class of wireless erasure networks we have defined [5].

8 Conclusion

We have obtained the capacity for the class of wireless erasure networks with broadcast and no interference at reception. We have generalized some of the capacity results that hold for wireline networks to a certain class of wireless erasure relay networks. We see from the proof that is not necessary for the relay nodes to decode their received message. They simply encode the received signal and broadcast it. (In fact in [6] we show that decoding by the relay nodes can actually reduce the capacity.) Therefore, unlike the wireline scenario where each link is made error free by channel coding and network coding is employed on top of that, our scheme only requires a single encoding function. In this sense, it takes care of channel and network coding in one shot. Only the destination has to decode the received signal.

Our techniques are general enough to apply to a wider range of problems pertaining to this class of networks. It will be interesting to see if similar results can be obtained for other types of networks, such as erasure networks in which interference is incorporated in the reception model, networks involving channels other than the erasure channel etc.

References

[1] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Info. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[3] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Info. Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[4] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking.*

[5] A. F. Dana, R. Gowaikar, B. Hassibi, and M. Effros, "On the capacity of erasure wireless networks," *in preparation for submission.*

[6] A. F. Dana, R. Gowaikar, B. Hassibi, M. Effros, and M. Médard, "Should we break a wireless network into sub-networks?," *Proc. 41st Annual Allerton Conf. on Communication, Control, and Computing*, 2003.