Capacity of Wireless Erasure Networks

Amir F. Dana, *Student Member, IEEE*, Radhika Gowaikar, *Student Member, IEEE*, Ravi Palanki, Babak Hassibi, and Michelle Effros, *Senior Member, IEEE*

***Abstract*—In this paper, a special class of wireless networks, called wireless erasure networks, is considered. In these networks, each node is connected to a set of nodes by possibly correlated erasure channels. The network model incorporates the broadcast nature of the wireless environment by requiring each node to send the same signal on all outgoing channels. However, we assume there is no interference in reception. Such models are therefore appropriate for wireless networks where all information transmission is packetized and where some mechanism for interference avoidance is already built in. This paper looks at multicast problems over these networks. The capacity under the assumption that erasure locations on all the links of the network are provided to the destinations is obtained. It turns out that the capacity region has a nice max-flow min-cut interpretation. The definition of cut-capacity in these networks incorporates the broadcast property of the wireless medium. It is further shown that linear coding at nodes in the network suffices to achieve the capacity region. Finally, the performance of different coding schemes in these networks when no side information is available to the destinations is analyzed.**

***Index Terms*—Wireless erasure networks, multicast problems.**

I. Introduction

DETERMINING the capacity region for general multiterminal networks has been a long-standing open problem. An outer bound for the capacity region is proposed in [1]. This outer bound has a nice min-cut interpretation: The rate of flow of information across any cut (a cut is a partition of the network into two parts) is less than the corresponding cut-capacity. The cut-capacity is defined as the maximum rate that can be achieved if the nodes on each side of the cut can fully cooperate and also use their inputs as side information.

The difficulty in multiterminal information theory is that this outer bound is not necessarily tight. For instance, for the single-relay channels introduced in [2], no known scheme achieves the min-cut outer bound of [1].

However, for a class of network problems called multicast problems in *wireline* networks, it is shown that the max-flow min-cut outer bound can be achieved [3]–[5]. A multicast problem comprises one or more source nodes (at which information is generated), several destinations (that demand all information available at the source nodes), relay nodes, and directed communication channels between some nodes. It is assumed that each channel is statistically independent of all other channels. Also, as the name suggests, the communication between different nodes is done through physically separated channels (wires). This means that the communication between two particular nodes does not affect the communication between others. In this setup, the maximum achievable rate is given by the minimum cut-capacity over all cuts separating the source nodes and a destination node. Because of the special structure of wireline networks, the cut-capacity for any cut is equal to the sum of the capacities of the channels crossing the cut.

This remarkable result for wireline networks is proved by performing separate channel and network coding in the network. First, we perform channel coding on each link of the network, so as to make it operate error free at any rate below its capacity. This way, the problem is transformed into a flow problem in a graph where the capacity of each edge is equal to the information-theoretic capacity of the corresponding channel in the original network. If there is only one destination node, standard routing algorithms for finding the max-flow (min-cut) in graphs [6] achieve the capacity. However, when the number of destinations is more than one, these algorithms can fail. The key idea in [3] is to perform coding at the relay nodes. By [4], [5], linear codes are sufficient to achieve the capacity in multicast problems. These ideas are formulated in an algebraic framework and generalized to some other special network problems in [5]. Since then, there has been a lot of research on the benefits of coding over traditional routing schemes in networks from different viewpoints such as network management, security, etc. [7], [8]. In a wireless setup, however, the problem of finding the capacity region is more complicated. The main reason is that unlike wireline networks, in which communication between different nodes is done using separated media, in a wireless system the communication medium is shared. Hence, all transmissions across a wireless network are *broadcast*. Also any communication between two users can cause *interference* to the communication of other nodes. These two features, broadcast and interference, present new issues and challenges for performance analysis and system design. The capacity regions of many information-theoretic channels that capture these effects are not known. For instance, the capacity region for general broadcast channels is an unsolved problem [9]. The capacity of general relay channels is not known. However, there are some achievable results based on block Markov encoding and random binning [10]. These ideas have been generalized and applied to a multiple relay setup in [11], [12].

Manuscript received July 21, 2004; revised February 23, 2005. This work was supported in part by the National Science Foundation under Grants CCR-0133818 and CCR-0326554, by the David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking.

The authors are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: amirf@caltech.edu; gowaikar@caltech.edu; palanki@caltech.edu.; hassibi@caltech.edu; effros@caltech.edu).

Communicated by M. Médard, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2005.864424

In this paper, we look at a special class of wireless networks which only incorporates the broadcast feature of wireless networks.[1] We model each communication channel in the network as a memoryless erasure channel. We will often assume that the erasure channels are independent; however, we show that the results also hold when the various erasure channels are correlated. We require that each node sends out the same signal on each outgoing link. However, for reception we use a multiple-access model without interference, i.e., messages coming into a node from different incoming links do not interfere. In general, this is not true for a wireless system. However, this can be realized through some time, frequency, or code-division multiple-access scheme. This simplification is important in making solution of the problem tractable.

Finally, we assume that complete side information regarding erasure locations on each link is available to the destination (but not to the relay) nodes. 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 by using headers in the packets to convey erasure locations or by sending a number of extra packets containing this information. By making the packets very long, the overhead of transmitting the erasure locations can be made negligible compared to the packet length. We should remark that provided that the side information is available to the destinations, all the results in this paper hold for any packet length.

We should mention that our model is appropriate for wireless networks where all information transmission is packetized and where some form of interference avoidance is already in place. Channel coding within each packet can be used to make each link behave as a packet erasure channel. Although our model does not incorporate interference (primarily because it is not clear what interference means for erasure channels) one way, perhaps, to account for interference is to allow the erasure channels coming into any particular node to be correlated (something that is permitted in our model).

The main result is that a max-flow min-cut type of result holds for multicast problems in wireless erasure networks under the assumptions mentioned above. The definition of cut-capacity in these networks is such that it incorporates the broadcast nature of the network. We further show that similar to the wireline case, for multicast problems over wireless erasure networks, linear encoding at nodes achieves all the points in the capacity region. Working with linear encoding functions reduces the complexity of encoding and decoding. Building on the results of this paper and using ideas from LT coding [15], it is shown in [16] that it is possible to reduce the delay incurred in the network. In their scheme, instead of using linear *block* codes, which is what we do here, the nodes send random linear combinations of their previously received signals at each time. This way, nodes do not need to wait for receiving a full block before transmitting, which reduces the delay.

We once more need to emphasize the importance of the side information on the erasure locations (or any other mechanism that provides the destination with the mapping from the source nodes to their incoming signals) for our result to hold. Interestingly, all the cut capacities of the network remain unchanged by making the above described side information available to the receiver nodes. Thus, in some sense, what is shown in this paper is that with appropriate side information made available to the receivers, the min-cut upper bound on capacity can be made tight. It would therefore be of further interest to see whether for other classes of networks it is possible to come up with the appropriate side information to make the min-cut bounds tight.

This paper is organized as follows. Section II defines notation used in this paper and reviews some graph-theoretic definitions of importance. We introduce the network model in Section III and the problem setup in Section IV. Section V states the main result for multicast problems over wireless erasure networks with side information available at destinations. Section VI includes proofs of these results. Section VII demonstrates the optimality of linear encoding. Section VIII includes a discussion of our network assumptions. Also the performance of different coding schemes when side information is not available is analyzed and compared. We mention future directions of our work and conclude in Section IX.

II. Preliminaries

A. Notation

Throughout this paper, upper case letters (e.g., X, Y, Z) usually denote random variables and lower case letters (e.g., x, y, z) denote the values they take. Underlined letters (e.g., $\underline{x}$) are used to denote vectors. Sets are denoted by calligraphic alphabet (e.g., $\mathcal{A}, \mathcal{B}, \mathcal{C}$). The complement of a set $\mathcal{A}$ is shown by $\mathcal{A}^c$. The transpose of matrix $\underline{x}$ is shown by $\underline{x}^\dagger$. $\exp(x)$ is used to denote 2^x.

Subscripts specify nodes, edges, inputs, outputs, and time. For instance, v_2 and X_2 could denote node number two and the output of node number two in the network, respectively. Unless otherwise mentioned, commas are used to separate time subscripts from other subscripts. Superscripts are also used to refer to different sources. For example, $w^{(s)}$ could denote the message sent by node s.

We use notation x^n to denote the sequence $x_1, x_2, \ldots, x_n$. We also use notation $(x_i, i \in \mathcal{I})$ to denote the ordered tuple specified by index set $\mathcal{I}$. Finally, $|\mathcal{X}|$ is the cardinality of set $\mathcal{X}$. Table I summarizes our notation.

B. Definitions for Directed Graphs

In this part, we briefly review the concepts and definitions from graph theory used in this paper [17].

A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ has vertex set $\mathcal{V}$ and directed edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. Without loss of generality, let

$$\mathcal{V} = \{1, 2, \ldots, |\mathcal{V}|\}.$$

We assume that the graph is finite, i.e., $|\mathcal{V}| < \infty$. For each node $v \in \mathcal{V}, \mathcal{N}_O(v)$ and $\mathcal{N}_I(v)$ are the set of edges leaving from and the set of edge going into v, respectively. Formally

$$\mathcal{N}_O(v) = \{(v,u) | (v,u) \in \mathcal{E}\}$$
$$\mathcal{N}_I(v) = \{(u',v) | (u',v) \in \mathcal{E}\}.$$

[1]References [13], [14] have considered applications of network coding at the network layer for cost (energy) minimization in lossless wireless *ad hoc* networks. In this paper, we look at wireless features of the network in the physical layer.

TABLE I
SOME IMPORTANT NOTATION IN THIS PAPER

$\mathcal{V}$	node set
$\mathcal{E}$	edge set
$\mathcal{S}$	the set of source nodes
$\mathcal{D}$	the set of destination nodes
$[\mathcal{V}_x, \mathcal{V}_y]$	$x-y$ cut-set described by x-set $\mathcal{V}_x$
X_i	symbol transmitted from node i
X_i^n	a transmitted block of n symbols from node i
Y_{ij}	channel output of edge (i,j)
Y_i	symbols received at node i from all incoming channels
$w^{(s)}$	message transmitted from source s
$\mathcal{W}^{(s)}$	message index set at source node s
$\hat{w}_{d_i}^{(s)}$	estimate at destination d_i of the message transmitted from s
$P_{d_i}^{(n)(s)}$	prob. of error in decoding source s at destination d_i

The *out-degree* $d_O(v)$ and *in-degree* $d_I(v)$ of v are defined as $d_O(v) = |\mathcal{N}_O(v)|$ and $d_I(v) = |\mathcal{N}_I(v)|$. A sequence of nodes $v_0, v_1, \ldots, v_n$ such that $(v_0, v_1), (v_1, v_2), \ldots, (v_n, v_0)$ are all in $\mathcal{E}$ is called a cycle. An acyclic graph is a directed graph with no cycles.

An $x-y$ *cut* for $x, y \in \mathcal{V}$ is a partition of $\mathcal{V}$ into two subsets $\mathcal{V}_x$ and $\mathcal{V}_y = \mathcal{V}_x^c$ such that $x \in \mathcal{V}_x$ and $y \in \mathcal{V}_y$. The x-set, $\mathcal{V}_x$ (or y-set, $\mathcal{V}_y$) determines the cut uniquely. For the $x-y$ cut given by $\mathcal{V}_x$, the *cut-set* $[\mathcal{V}_x, \mathcal{V}_y]$ is the set of edges going from the x-set to y-set, i.e.,

$$[\mathcal{V}_x, \mathcal{V}_y] = \{(u,v) | (u,v) \in \mathcal{E}, u \in \mathcal{V}_x, v \in \mathcal{V}_y\}.$$

We also define $\mathcal{V}_x^*$ as

$$\mathcal{V}_x^* = \{v | \exists u \text{ s. t. } (v,u) \in [\mathcal{V}_x, \mathcal{V}_y]\}.$$

$\mathcal{V}_x^*$ is the set of nodes in the x-set that has at least one of its outgoing edges in the cut-set.

Example 2.1: Consider the acyclic directed graph shown in Fig. 1. $\mathcal{V} = \{1,2,3,4\}$ is the set of nodes and

$$\mathcal{E} = \{(1,2), (3,2), (1,3), (3,4), (2,4)\}$$

is the set of edges. The source and destination nodes are $s = 1$ and $d = 4$, respectively. The *out-degree* of node 3 is 2, i.e., $d_O(3) = 2$. Looking at the $s-d$ cut specified by s-set $\mathcal{V}_s = \{1,3\}$, the cut-set $[\mathcal{V}_s, \mathcal{V}_d]$ is the set $\{(3,4), (3,2), (1,2)\}$ and $\mathcal{V}_s^* = \{1,3\}$.

III. NETWORK MODEL

A. Wireless Packet Erasure Networks

We model the wireless packet[2] erasure network by a directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each edge $(i,j) \in \mathcal{E}$ represents a memoryless packet erasure channel from node i to node j. For most of this paper, we assume that erasure events across different links are independent. However, as described later in the paper, the results go through for correlated erasure events. For independent erasure events, a packet sent across link (i,j) is either erased with probability of erasure ϵ_{ij} or received without error. We denote the input alphabet (the set of possible packets) of the erasure channel by $\mathcal{X}$.[3]

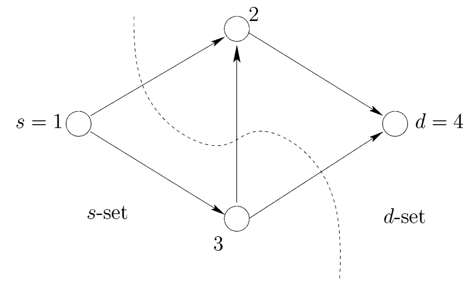

Fig. 1. A directed acyclic graph with four nodes and five edges. The cut-set $\{(3,4), (3,2), (1,2)\}$ is shown by the dashed line.

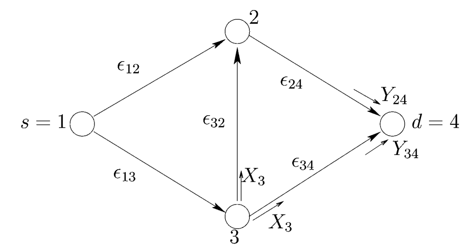

Fig. 2. i) An erasure wireless network with the graph representation of Example 2.1. Probability of erasure on link (i,j) is ϵ_{ij}. Each node (e.g., node 3) transmits the same signal (X_3) across its outgoing channels. Since the network is interference free, node 4 receives both signals Y_{24} and Y_{34} completely. ii) In this network, cut-capacity for s-set $\mathcal{V}_s = \{1,3\}$ is $C(\mathcal{V}_s) = 1 - \epsilon_{12} + 1 - \epsilon_{32}\epsilon_{34}$.

Let $Z_{ij,t}$ be a random variable indicating erasure occurrence across channel (i,j) at time t. For independent erasure events, $Z_{ij,t}$ has a Bernoulli distribution with parameter ϵ_{ij}. If an erasure occurs on link $(i,j) \in \mathcal{E}$ at time t, the value of $Z_{ij,t}$ will be one, otherwise $Z_{ij,t}$ will be zero. Note that the behavior of the network can be fully determined by the values of $Z_{ij,t}$ for all links and all times and the operation performed at each node.

We assume that transmissions on each channel experience one unit of time delay. The input of all the channels originating from node i is denoted by X_i chosen from input alphabet $\mathcal{X}$. Note that with this definition, we have required that each node transmits the same symbol on all its outgoing edges, i.e., all channels corresponding to edges in $\mathcal{N}_O(i)$ have the (same) input X_i (see Fig. 2). This constraint incorporates broadcast in our network model. The output of the communication channel corresponding to edge $(i,j) \in \mathcal{E}$ is denoted by Y_{ij}; Y_{ij} lies in output alphabet $\mathcal{Y} = \mathcal{X} \cup \{e\}$, where e denotes the erasure symbol. We also assume that the outputs of all channels corresponding to edges in $\mathcal{N}_I(i)$ are available at node i. This condition is equivalent to having no interference in receptions in the network. Having this, let $Y_i = (Y_{ji}, (j,i) \in \mathcal{N}_I(i))$ be the symbols that are received at node i from all its incoming channels. We have $Y_i \in \prod_{j:(j,i) \in \mathcal{E}} \mathcal{Y}$. The relation between the Y_i's and X_i's defines a coding scheme for the network.

Based on the properties of the network mentioned above, if we consider the inputs and outputs up to time t, then the conditional probability function of the outputs of all the channels

[2]Throughout this paper, a packet can be of any length. When the length of packets is one, the channel is a binary-erasure channel.

[3]For simplicity and without loss of generality, we consider $\mathcal{X} = \{0,1\}$ in our analysis and proofs. However, we should remark that all the results and analysis hold for input alphabet of arbitrary length.

(edges) up to time t given all the inputs of all the channels up to time t and all the previous outputs, can be written as follows for all t:

$$\mathrm{P}\left((y_{ij,t},(i,j)\in\mathcal{E})\,|\,\left(x_l^t,l\in\mathcal{V}\right),\left(y_{ij}^{t-1},(i,j)\in\mathcal{E}\right)\right) = \mathrm{P}\left((Y_{ij}=y_{ij,t},\,(i,j)\in\mathcal{E})\,|\,(X_l=x_{l,t},\,l\in\mathcal{V})\right).$$

For independent erasure events, we further have

$$\mathrm{P}\left((y_{ij,t},(i,j)\in\mathcal{E})|(x_l^t,l\in\mathcal{V}),(y_{ij}^{t-1},(i,j)\in\mathcal{E})\right) = \prod_{i\in\mathcal{V}}\prod_{j:\,(i,j)\in\mathcal{N}_O(i)}\mathrm{P}(Y_{ij}=y_{ij,t}|X_i=x_{i,t}). \quad (1)$$

B. Multicast Problem

In this paper, we consider a class of network problems called multicast problems. Any network problem is characterized by a collection of information sources, a collection of source nodes at which one or more information sources are available, and a collection of destination nodes. Each destination node demands a subset of information sources. The class of network problems that we consider in this paper is the multiple source/multiple destination multicast, where each of the destinations demands all of the information sources. This problem can be further specified by the following sets.

- $\mathcal{S}=\{s_1,s_2,\ldots,s_{|\mathcal{S}|}\}\subset\mathcal{V}$ denotes the information source nodes. We assume that each of the source nodes generates an information (message) which is modeled by an independent and identically distributed (i.i.d.) uniformly distributed random process. Information sources at different nodes are assumed to be independent.
- $\mathcal{D}=\{d_1,d_2,\ldots,d_{|D|}\}\subset\mathcal{V}$ denotes the set of destination nodes.

Note that $\mathcal{S}\cap\mathcal{D}$ may not be empty, i.e., a node can be a destination node for one information source and a source node for another. Also, destination nodes can act as relay nodes for other destination nodes in the network.

C. Side Information at Destinations

In most parts of the paper we assume that each destination node $d\in\mathcal{D}$ has complete knowledge of the erasure locations on each link of the network that is on a path from the source set to d. In other words, d knows values of the $z_{ij,t}$, for all $(i,j)\in\mathcal{E}$ and all times t, for which (i,j) is on at least one path from one of the sources to d. This serves as channel side information provided to the destinations from across the network. In the case when we consider large packets (alphabet), this side information can be provided using negligible overhead. More discussion of this model appears in Section VIII.

D. Cut-Capacity Definition

Consider an $s-d$ cut given by s-set, $\mathcal{V}_s$ as defined in Section II-B. We define $X(\mathcal{V}_s)$ and $Y(\mathcal{V}_s)$ as

$$\begin{aligned} X(\mathcal{V}_s) &= \{X_i|i\in\mathcal{V}_s^*\} \\ Y(\mathcal{V}_s) &= \{Y_{ij}|(i,j)\in[\mathcal{V}_s,\mathcal{V}_s^c]\}. \end{aligned} \quad (2)$$

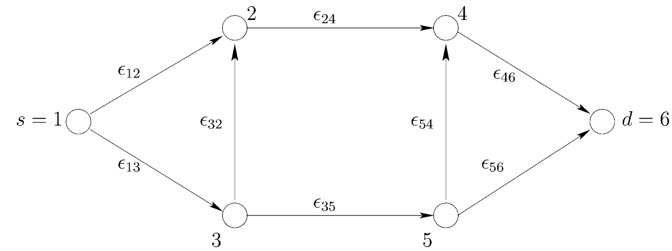

Fig. 3. For the cut-set specified by the s-set $\mathcal{V}_s=\{1,3,4\}$ the cut-capacity is $C(\mathcal{V}_s)=1-\epsilon_{12}+1-\epsilon_{46}+1-\epsilon_{35}\epsilon_{32}$.

At the end of this section, we define the cut-capacity for wireless erasure networks. In wireline networks, the value of the cut-capacity is the sum of the capacities of the edges in the cut-set [5]. Such a definition of cut-capacity in wireline networks makes sense because the nodes can send out different signals across their outgoing edges. However, this is not the case for wireless erasure networks where broadcast transmissions are required. The following definition of cut-capacity is different from that in the wireline network settings, and it incorporates the broadcast nature of transmission in our network.

Definition: Consider an erasure wireless network represented by $\mathcal{G}=(\mathcal{V},\mathcal{E})$ and probabilities of erasure ϵ_{ij} as described in Section III. Let s and d_l be the source and destination nodes, respectively. The cut-capacity corresponding to any $s-d_l$ cut represented by s-set, $\mathcal{V}_s$ is denoted by $C(\mathcal{V}_s)$ and is equal to

$$C(\mathcal{V}_s)=\sum_{i\in\mathcal{V}_s^*}\left(1-\prod_{j:\,(i,j)\in[\mathcal{V}_s,\mathcal{V}_{d_l}]}\epsilon_{ij}\right). \quad (3)$$

Example 3.1: Consider the network represented by the directed graph of Example 2.1 (see Fig. 2). For the $s-d$ cut specified by the s-set, $\mathcal{V}_s=\{1,3\}$, the cut-capacity is

$$C(\mathcal{V}_s)=1-\epsilon_{12}+1-\epsilon_{32}\epsilon_{34}.$$

Looking at this example, we see that all edges in the cut-set that originate from a common node, i.e., edges $(3,2)$ and $(3,4)$, together contribute a value of one minus the product of their erasure probabilities, i.e., $1-\epsilon_{32}\epsilon_{34}$ to the cut-capacity. This observation holds in general for wireless erasure networks.

Example 3.2: As another example, consider the network shown in Fig. 3 with one source $s=1$ and one destination $d=6$. The cut-capacity corresponding to the $s-d$ cut specified by $\mathcal{V}_s=\{1,3,4\}$ is $C(\mathcal{V}_s)=1-\epsilon_{12}+1-\epsilon_{46}+1-\epsilon_{35}\epsilon_{32}$.

IV. Problem Statement

We next define the class of block codes considered in this paper. A $(\lceil 2^{nR_1}\rceil,\ldots,\lceil 2^{nR_{|\mathcal{S}|}}\rceil,n)$ code for the multicast problem in a wireless erasure network described in the previous sections, consists of the following components.

- A set of integers $\mathcal{W}^{(s_i)}=\{1,2,\ldots,\lceil 2^{nR_i}\rceil\}$ for each source node $s_i\in\mathcal{S}$. $\mathcal{W}^{(s_i)}$ represents the set of message indices corresponding to node s_i. $w^{(s)}$ denotes the message of source $s\in\mathcal{S}$. We assume that the messages are equally likely and independent.

- A set of encoding functions $\{f_{i,t}\}_{t=1}^n$ for each node $i \in \mathcal{V}$, where
$$x_{i,t} = f_{i,t}\left(w^{(i)}, y_i^{t-1}\right)$$
is the signal transmitted by node i at time t. Note that $x_{i,t}$ is a function of the message $w^{(i)}$ that node $i \in \mathcal{V}$ wants to transmit in the current block[4] and all symbols received so far by node i from its incoming channels. If i is not a source node, we set $w^{(i)} = 0$ for all blocks and all times.
- A decoding function g_{d_i} at destination node $d_i \in \mathcal{D}$
$$g_{d_i} : \mathcal{W}^{(d_i)} \times \mathcal{Y}_{d_i}^n \times \{0,1\}^{n|\mathcal{E}|} \to \prod_{s\in\mathcal{S}} \mathcal{W}^{(s)}$$
such that
$$\begin{aligned}\underline{\hat{w}}_{d_i} &= \left(\hat{w}_{d_i}^{(s)},\ s\in\mathcal{S}\right)\\ &= g_{d_i}\left(w^{(d_i)}, y_{d_i}^n, (z_{ij,t}, (i,j)\in\mathcal{E},\ 1\le t\le n)\right) \quad (4)\end{aligned}$$
where $\hat{w}_{d_i}^{(s)}$ is the estimate of the message sent from source $s \in \mathcal{S}$ based on received signals at d_i, information source available at d_i,[5] $w^{(d_i)}$, and also the erasure occurrences on all the links of the network in the current block.

Note that X_i, Y_{ij}, and Y_i all depend on the message vector $\underline{w} = (w^{(s)},\ s \in \mathcal{S})$ that is being transmitted. Therefore, we will write them as $X_i(\underline{w}), Y_{ij}(\underline{w})$, and $Y_i(\underline{w})$ to specify what specific set of messages is transmitted.

Associated with every destination node $d \in \mathcal{D}$ and every information source $s \in \mathcal{S}$ is a probability that the message will not be decoded correctly[6]
$$P_d^{(n)(s)} = \Pr\left(\hat{W}_d^{(s)} \neq W^{(s)}\right) \quad (5)$$
where $P_d^{(n)(s)}$ is defined under the assumption that all the messages are independent and uniformly distributed over $\mathcal{W}^{(s)},\ s \in \mathcal{S}$. The set of rates $(R_s,\ s \in \mathcal{S})$ is said to be achievable if there exist a sequence of $(\lceil 2^{nR_1}\rceil, \ldots, \lceil 2^{nR_{|\mathcal{S}|}}\rceil, n)$ codes such that $P_d^{(n)(s)} \to 0$ as $n \to \infty$ for all $s \in \mathcal{S}$ and $d \in \mathcal{D}$. The capacity region is the closure of the set of achievable rates.

V. MAIN RESULTS

In this section we present the main results of this paper.

Theorem 1: Consider a single-source/single-destination wireless erasure network described by the directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and the assumptions of Section III. Let $s \in \mathcal{V}$ and $d \in \mathcal{V}$ denote the network's source and destination, respectively. Then the capacity of the network with side information at the destination is given by the value of the minimum value $s - d$ cut. More precisely, we have
$$C = \min_{\mathcal{V}_s : \mathcal{V}_s \text{an } s-d \text{ cut}} C(\mathcal{V}_s). \quad (6)$$

Remark 1: The results derived in this paper are stated for erasure wireless networks with broadcast property (and no interference). However, based on the results of this paper, it is possible to derive the capacity of multicast problems over *error-free* networks (with the broadcast property and without interference), with or without capacitated links.

Remark 2: Although we have assumed that the erasure events across the network are independent, the capacity results of this paper also hold for the case when the erasure events are correlated, i.e., $Z_{ij},\ (i,j) \in \mathcal{E}$ are dependent on each other. In that case, the definition of the cut capacity should be modified as described in (A3). (See Remark 4 in Appendix A).

Example 5.1: Recall the single-source/single-destination network of Example 3.1 (see Fig. 2). By Theorem 1, the capacity of this network is
$$\begin{aligned}C = \min\{&1-\epsilon_{12}+1-\epsilon_{32}\epsilon_{34}, 1-\epsilon_{34}\\ &+1-\epsilon_{24}, 1-\epsilon_{12}\epsilon_{13}, 1-\epsilon_{13}+1-\epsilon_{24}\}\end{aligned}$$
The following theorems generalize the single-source/single-destination result to general multicast problems.

Theorem 2: Consider a multiple-source/single-destination wireless erasure network described by directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and the assumptions of Section III. Suppose that the destination requests all of the information from all of the sources. Let $\mathcal{S} \subset \mathcal{V}$ and $d \in \mathcal{V}$ denote the set of source nodes and the destination node, respectively. The capacity region of the network with side information provided at the destination is given by
$$C(\mathcal{G}, \mathcal{S}, d) \triangleq \left\{ (R_s,\ s \in \mathcal{S}) \middle| \ 0 \le \sum_{s\in\mathcal{V}'\cap\mathcal{S}} R_s \le C(\mathcal{V}')\ \forall \mathcal{V}' \subset \mathcal{V} - \{d\} \right\}. \quad (7)$$
In other words, the total rate of information transmission to d across any cut $[\mathcal{V}', \mathcal{V}_d]$, should not exceed the cut-capacity of that cut.

Example 5.2: Consider the network shown in Fig. 4 with two sources $\{1, 2\}$ and one destination $\{3\}$. Then according to Theorem 2, the capacity region is
$$\begin{aligned}\{(R_1, R_2) \in \mathbb{R}^+ \times \mathbb{R}^+ | R_1\\ \le 1-\epsilon_{12}\epsilon_{13}, R_2 \le 1-\epsilon_{23}, R_1+R_2 \le 1-\epsilon_{23}+1-\epsilon_{13}\}.\end{aligned}$$

Theorem 3: Consider a multicast problem with multiple sources and multiple destinations. Let $\mathcal{S}, \mathcal{D} \subset \mathcal{V}$ denote the set of source nodes and destination nodes, respectively. The capacity region of the network with side information is given by the intersection of the capacity regions of the multicast problem between the sources and each of the destinations, i.e.,
$$C(\mathcal{G}, \mathcal{S}, \mathcal{D}) = \bigcap_{d\in\mathcal{D}} C(\mathcal{G}, \mathcal{S}, d). \quad (8)$$

Corollary 1: Consider a multicast problem with one source denoted by s and multiple destinations denoted by $d_1, \ldots, d_{|\mathcal{D}|}$. The capacity of the network is given by the minimum value of the cuts between the source node and any of the destinations, i.e.,
$$C = \min_{d_i\in\mathcal{D}} \min_{\mathcal{V}_s : s-d_i \text{cut}} C(\mathcal{V}_s).$$

[4]The value of $w^{(i)}$ does not change in one block.

[5]If $d_i \notin \mathcal{S}$, without loss of generality we set $w^{(d_i)} = 0$ and $\mathcal{W}^{(d_i)} = \{0\}$ for all blocks.

[6]Note that if d is a source node, we assume without loss of generality that $P_d^{(n)(d)} = 0$.

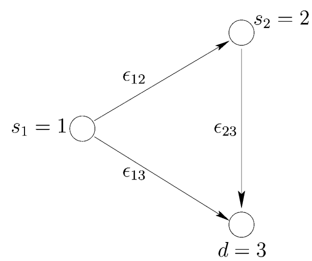

Fig. 4. A wireless erasure network with two sources $\mathcal{S} = \{1, 2\}$ and one destination $\mathcal{D} = \{3\}$.

Example 5.3: Consider the network shown in Fig. 2. Suppose that we are decoding at node 2 and 4, i.e., $\mathcal{D} = \{2, 4\}$. Based on Corollary 1, the capacity of this network is

$$C = \min\{1 - \epsilon_{12} + 1 - \epsilon_{32}, 1 - \epsilon_{34} + 1 - \epsilon_{24}, 1 - \epsilon_{12}\epsilon_{13}, 1 - \epsilon_{13} + 1 - \epsilon_{24}\}.$$

The preceding results show that the capacity region for multicast problems over wireless erasure networks has a max-flow min-cut interpretation. This result is similar to multicast problems in wireline networks [3], however, the definition of the cut-capacity is different. Recall from [3] that in wireline networks, the cut-capacity is the sum of the capacities of the links in the cut-set. Since wireless erasure networks incorporate broadcast, the cut-capacity is the sum of the capacities of each broadcast system that operates across the cut.

The next theorem states that linear network coding is sufficient for achieving the capacity region.

Theorem 4: Consider a multicast problem with multiple sources and multiple destinations. Then any rate vector in the capacity region $C(\mathcal{G}, \mathcal{S}, \mathcal{D})$ of the network defined in Theorem 3 is achievable with linear block coding.

In the next section, we prove Theorems 1–3. In Section VII, we look at the performance of the network using random linear coding and prove Theorem 4.

VI. Proof of Theorems

A. *Proof of Theorems 1 and 2*

In this subsection, we prove the results stated for multiple- source/single-destination network problems. We start by proving the converse.

1) Converse: We prove the converse part by considering perfect cooperation among subsets of nodes. Consider the cut specified by d-set $\mathcal{V}_d$. Let all of the nodes in $\mathcal{V}_d$ and all of the nodes in $\mathcal{V}_d^c$ cooperate perfectly, i.e., each node has access to all of the information known to nodes in its set. In this case, we have a multiple-input, multiple-output point-to-point erasure channel. Consider all source nodes in $\mathcal{V}_d^c$. Then, clearly, the sum-rate of these source nodes must be less than the capacity of the multiple-input multiple-output point-to-point erasure channel. The capacity of this point-to-point communication channel is

$$C_{\rm col} = \max_{P(x_i,\ i\in\mathcal{V}_d^{c*})} I\left((X_i,\ i\in\mathcal{V}_d^{c*});(Y_{ij},\ (i,j)\in[\mathcal{V}_d^c,\mathcal{V}_d])\right).$$

Since the channels are independent and memoryless, the mutual information is maximized when the different X_i's are i.i.d. and uniform on the input alphabet $\mathcal{X}$. In this case, the above mutual information equals the cut-capacity corresponding to the cut-set $[\mathcal{V}_d^c, \mathcal{V}_d]$, i.e.,

$$C_{\rm col} = C\left(\mathcal{V}_d^c\right).$$

Therefore, for any cut-set $[\mathcal{V}_d^c, \mathcal{V}_d]$ the sum-rate of the information sources in set $\mathcal{V}_d^c$ satisfies

$$\sum_{s\in\mathcal{S}\cap\mathcal{V}_d^c} R_s \le C_{\rm col} = C(\mathcal{V}_d^c).$$

The complete analysis appears in Appendix A. The proof follows the same lines as the min-cut upper bound of Cover and Thomas for multiterminal networks [1, Sec. 14.10].

2) Achievability: In this subsection, we prove that all of the rates arbitrarily close to rates in the capacity regions given in Theorems 1 and 2 are achievable for a multiple-sources/single-destination multicast problem. We next use random coding techniques to show this result.

We employ random block codes in the network. Each node transmits the next block of n symbols only after it has received all n symbols corresponding to the present block from each of its incoming channels. Let $L_{\max}$ denote the length of the longest path from a source to the destination in the network. Since each transmission introduces one unit of time delay, the maximal delay between the transmission of a message from one source and its receipt at the destination using block codes of length n is $nL_{\max}$. We do not use any information from previously decoded blocks to decode the current set of messages. Also note that since our model assumes that the reception is interference free, there is no confusion among different blocks at any node. Therefore, if the network operates for nB units of time (i.e., B blocks of length-n symbols) then the destination has received all of the information required for decoding the $B - L_{\max}$ first messages transmitted from each source $s \in \mathcal{S}$, i.e., $w_b^{(s)},\ b = 1, \ldots, B - L_{\max}$. Since the network size is finite, as $B \to \infty$, for fixed n, the rate $R_s \frac{B-L_{\max}}{B}$ approaches R_s.[7] The same codebook and encoding and decoding functions are used for all the blocks. We explain the coding scheme for transmitting one set of messages from the sources to the destination. In the following, we describe the encoding and decoding processes.

- **Codebook Generation and Encoding:** For each node $i \in \mathcal{V}$, the encoding function

$$f_i : \mathcal{W}^{(i)} \times \mathcal{Y}_i^n \to \mathcal{X}^n$$

is generated randomly as follows. For each $y^n \in \mathcal{Y}_i^n$ and for each $w^{(i)} \in \mathcal{W}^{(i)}$ we draw the symbols of $f_i(w^{(i)}, y^n) \in \mathcal{X}^n$ randomly and independently according to a binary Bernoulli distribution with parameter $1/2$. Thus, the channel input at node $i \in \mathcal{V}$ is $X_i^n = f_i(w^{(i)}, y^n)$ when the message at node i is $w^{(i)}$ and the incoming sequence is $y^n \in \mathcal{Y}_i^n$. The destination

[7]We could also consider the case when different sources transmit different numbers of messages in B block uses. In that case, if L_s denotes the longest path from $s \in \mathcal{S}$ to the destination, we could transmit $B - L_s$ messages from information source s to the destination. However, for simplicity of notation and analysis we assume that all of the nodes send the same number of messages in a synchronized fashion.

has perfect knowledge of all the encoding functions $f_i(\cdot),\ i \in \mathcal{V}$ thus generated.[8]

- **Decoding:** The destination "simulates" the network to decode the messages. Suppose that message vector $\underline{w}_0 = (w_0^{(s)},\ s \in \mathcal{S})$ is transmitted and $y_d^n(\underline{w}_0)$ is received at destination d. By assumption, the receiver knows the erasure locations on all the links of the network, i.e., $(z_{ij}^n, (i,j) \in \mathcal{E})$. Having all of the erasure locations and all of the encoding functions applied at different nodes in the network,[9] the destination can compute the values of $X_i^n(\underline{w}), Y_{ij}^n(\underline{w})$, and $Y_i^n(\underline{w})$ for all nodes and edges for any $\underline{w} \in \prod_{s \in \mathcal{S}} \mathcal{W}^{(s)}$. If there exist a unique message vector $\underline{w} \in \prod_{s \in \mathcal{S}} \mathcal{W}^{(s)}$ such that the computed value of $Y_d^n(\underline{w})$ equals the value $y_d^n(\underline{w}_0)$ of the received signal at the destination, then $\underline{w}$ is declared as the decoder output. Otherwise, the decoder declares an error.

  Since the computed value of $Y_d^n(\underline{w}_0)$ for transmitted message $\underline{w}_0$ always matches the received signal at the destination, an error occurs if and only if there is another message vector $\underline{w} \neq \underline{w}_0$ for which $Y_d^n(\underline{w}) = Y_d^n(\underline{w}_0) = y_d^n(\underline{w}_0)$. In the next subsection we compute the probability of this event and show that for large blocks this probability can be made arbitrarily close to zero provided that the rate vector $(R_s,\ s \in \mathcal{S})$ is inside the capacity region described in Theorems 1 and 2.

3) Probability of Error: Let $\Pr(\text{err})$ be the probability of error averaged over all possible functions f_i. In other words, if $P_e^{(n)}$ is the probability that $\hat{w}_0^{(s)}$, the destination's estimate of the transmitted message $\underline{w}_0$ is not equal to $\underline{w}_0$, then $\Pr(\text{err})$ is the expected value of $P_e^{(n)}$ over all possible encoding functions at all nodes.[10] More precisely

$$P_e^{(n)} = \Pr\left(\exists s \in \mathcal{S}\,\text{s.t.}\, \hat{w}_0^{(s)} \neq w_0^{(s)}\right)$$

and $\Pr(\text{err}) = \mathrm{E}\, P_e^{(n)}$. Because of the symmetry of the code construction

$$\Pr(\text{err}) = \Pr(\text{err}|\underline{W} = \underline{w}_0 \,\text{is transmitted}) \tag{9}$$

where $\underline{W} = (W^{(s)},\ s \in \mathcal{S})$. Therefore, we will find the average probability of error when message vector $\underline{w}_0$ is transmitted from the sources. Recall the notation $X_i^n(\underline{w}_0)$ and $Y_i^n(\underline{w}_0)$ and $Y_{ij}^n(\underline{w}_0), (i,j) \in \mathcal{E}$. For each $\underline{w} \in \underline{\mathcal{W}} \triangleq \prod_{s \in \mathcal{S}} \mathcal{W}^{(s)}, \underline{w} \neq \underline{w}_0$, define the following event:

$$E(\underline{w}) = \{Y_d^n(\underline{w}) = Y_d^n(\underline{w}_0)\}. \tag{10}$$

Let $\mathcal{A}_\delta^{(n)}(i)$ be the event that the erasure locations on the channels going out of node i are jointly δ-strongly typical, i.e.,

$$\mathcal{A}_\delta^{(n)}(i) = \left\{\left(z_{ij}^n,\, j : (i,j) \in \mathcal{E}\right) \text{ are jointly } \delta\text{-strongly typical}\right\}$$

[1, Eq. (13.107)] and define

$$\mathcal{A}_\delta^{(n)} = \bigcap_{i=1}^{|\mathcal{V}|} \mathcal{A}_\delta^{(n)}(i).$$

Note that by the weak law of large numbers [1], $\Pr(\mathcal{A}_\delta^{(n)}(i)) \to 1$ as $n \to \infty$, and hence, for all $\delta > 0$

$$\Pr\left(\mathcal{A}_\delta^{(n)}\right) \geq 1 - |\mathcal{V}|\delta, \qquad \text{for } n \text{ sufficiently large.}$$

Using the definition of the above events, $\Pr(\text{err})$ can be written as

$$\begin{aligned}\Pr(\text{err}) &= \Pr(\text{err}|\underline{W} = \underline{w}_0)\\ &= \Pr\left(\bigcup_{\underline{w} \in \underline{\mathcal{W}} - \{\underline{w}_0\}} E(\underline{w})\right)\\ &= \Pr\left(\bigcup_{\underline{w} \in \underline{\mathcal{W}} - \{\underline{w}_0\}} E(\underline{w}) | \mathcal{A}_\delta^{(n)}\right) \Pr(\mathcal{A}_\delta^{(n)})\\ &\quad + \Pr\left(\bigcup_{\underline{w} \in \underline{\mathcal{W}} - \{\underline{w}_0\}} E(\underline{w}) | \mathcal{A}_\delta^{(n)^c}\right) \Pr(\mathcal{A}_\delta^{(n)^c})\\ &\leq \sum_{\underline{w} \in \underline{\mathcal{W}} - \{\underline{w}_0\}} \Pr(E(\underline{w}) | \mathcal{A}_\delta^{(n)}) + |\mathcal{V}|\delta. \end{aligned} \tag{11}$$

Therefore, using strong typicality and the union bound on the probability of events, we only look at network instantiations that are "strongly typical." We next bound the conditional probability of $E(\underline{w})$ given $\mathcal{A}_\delta^{(n)}$.

Corresponding to each cut in the network, represented by d-set $\mathcal{V}_d \ni d$, define the following event:

$$B[\mathcal{V}_d] = \left(\bigcap_{i \in \mathcal{V}_d} \{Y_i^n(\underline{w}) = Y_i^n(\underline{w}_0)\}\right) \bigcap \left(\bigcap_{i \in \mathcal{V}_d^c} \{Y_i^n(\underline{w}) \neq Y_i^n(\underline{w}_0)\}\right). \tag{12}$$

The interpretation of the above event is as follows. By definition of $E(\underline{w})$, we know that the received signal at the destination is the same for $\underline{w}$ and $\underline{w}_0$, but $\underline{w} \neq \underline{w}_0$. Therefore, we can partition the nodes of the network into two sets: the "distinguishable" and the "indistinguishable" set. The "distinguishable" set contains all nodes for which the signal received at those nodes when $\underline{w}$ is transmitted differs from the signal received when $\underline{w}_0$ is transmitted. All the other nodes, for which the received signals for $\underline{w}$ and $\underline{w}_0$ are the same, are in the "indistinguishable" set. Clearly, these two sets define a cut. Event $B[\mathcal{V}_d]$ corresponds to the case when the "indistinguishable" set (containing d) is equal to $\mathcal{V}_d \subset \mathcal{V}$. Note that these events are all disjoint and also

$$E(\underline{w}) = \bigcup_{\mathcal{V}_d:\text{d-set}} B[\mathcal{V}_d].$$

Define

$$\mathcal{M}(\underline{w}) = \left\{s | s \in \mathcal{S}, w_0^{(s)} \neq w^{(s)}\right\} \tag{13}$$

to be the subset of source nodes for which the corresponding messages in $\underline{w}$ and $\underline{w}_0$ are different. Set $\mathcal{M}(\underline{w})$ is not empty since $\underline{w}_0 \neq \underline{w}$ by assumption. In what follows, we bound the

[8]Note that the encoding functions thus constructed satisfy a causality condition that is more strict than what is defined in Section IV. Here each transmitted block is only a function of the immediately preceding block of received symbols. In Section III, each transmitted symbol could be a function of all previous symbols.

[9]We also assume that the destinations knows the topology of the network.

[10]Note that if $P_e^{(n)}$ goes to zero as n grows larger, so will $P_d^{(n)(s)}$ of (5) for every $s \in \mathcal{S}$.

probability of event $B[\mathcal{V}_d]$ by considering the edges in the cut-set $[\mathcal{V}_x, \mathcal{V}_x^c]$ where $\mathcal{V}_x \triangleq \mathcal{V}_d^c \cup \mathcal{M}(\underline{w})$. Note that $\mathcal{V}_x^c$ is a d-set since if the destination is a source of information, it is aware of the message it has transmitted and so $d \notin \mathcal{M}(\underline{w})$.

Consider any edge $(i,j) \in [\mathcal{V}_x, \mathcal{V}_x^c]$. We know that the transmitted signal $X_i^n = f_i(W^{(i)}, Y_i^n)$ from node i is a function of the message it wants to transmit $W^{(i)}$, and the received signal at its incoming edges Y_i^n. For any node i in $\mathcal{V}_d^c \cup \mathcal{M}(\underline{w})$, either the received signal Y_i^n or message $w^{(i)}$ is different for message vectors $\underline{w}$ and $\underline{w}_0$. Thus, for a randomly designed code, the transmitted signal by node i for message vector $\underline{w}$ is independent of the corresponding X_i^n for message vector $\underline{w}_0$. Using this observation, we bound the probability of the event $E(\underline{w})$ conditioned on $\mathcal{A}_\delta^{(n)}$ at the bottom of this page. Here

(a) follows since $\Pr(A,B) \le \Pr(A)$ for any events A and B. Instead of looking at equalities on every edge and every node of the network, we are looking at the nodes having an edge from $[\mathcal{V}_x, \mathcal{V}_x^c]$ connected to them, where $\mathcal{V}_x = \mathcal{V}_d^c \cup \mathcal{M}(\underline{w})$.

(b) is clear from the definition of $\mathcal{V}_x^*$.

(c) follows from the definition of conditional probability.

(d) follows from fact that averaged over all possible functions f_i, the conditional events shown in the equation are independent for different i's in $\mathcal{V}_x^*$.

Now we bound the expression given in (14), as shown at the bottom of the page, for any node $i \in \mathcal{V}_x^*$. Note that since

$$(w^{(i)}, Y_i^n(\underline{w})) \neq (w_0^{(i)}, Y_i^n(\underline{w}_0))$$

at node i, $X_i^n(\underline{w}) = f_i(w^{(i)}, Y_i^n(\underline{w}))$ and

$$X_i^n(\underline{w}_0) = f_i(w_0^{(i)}, Y_i^n(\underline{w}_0))$$

are chosen independently and uniformly from $\{0,1\}^n$. Therefore, the probability that they are the same in at least α_i specific locations is at most $2^{-\alpha_i}$. Looking at a fixed node i, $Y_{ij}^n(\underline{w}) = Y_{ij}^n(\underline{w}_0)$ for all j such that $(i,j) \in [\mathcal{V}_x, \mathcal{V}_x^c]$ only if all the locations that $X_i^n(\underline{w})$ and $X_i^n(\underline{w}_0)$ differ get erased on all these edges. Because of the δ-strong typicality of the erasure locations on edges $(i,j) \in [\mathcal{V}_x, \mathcal{V}_x^c]$, the number of locations at which erasure occurs on all the edges of interest, say $\alpha_i(\mathcal{V}_x)$, satisfies

$$\left|\frac{1}{n}\alpha_i(\mathcal{V}_x) - \Pr\left(Z_{ij}=1,\ j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]\right)\right| \le \frac{\delta}{2^{|\{j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]\}|}} \le \delta.$$

Therefore, $X_i^n(\underline{w})$ and $X_i^n(\underline{w}_0)$ cannot differ in more than

$$n(\Pr(Z_{ij}=1,\ j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]) + \delta)$$

locations and the probability of this event is no more than

$$\exp\left(-n\left(1-\Pr\left(Z_{ij}=1,\ j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]\right)-\delta\right)\right) = \exp\left(-n\left(1-\prod_{j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]}\epsilon_{ij}-\delta\right)\right). \quad (15)$$

$$\begin{aligned}
&\Pr\left(E(\underline{w})|\mathcal{A}_\delta^{(n)}\right)\\
&= \Pr\left(\bigcup_{\mathcal{V}_d:d\text{-set}} B[\mathcal{V}_d]|\mathcal{A}_\delta^{(n)}\right) = \sum_{\mathcal{V}_d:d\text{-set}} \Pr\left(B[\mathcal{V}_d]|\mathcal{A}_\delta^{(n)}\right)\\
&= \sum_{\mathcal{V}_d:d\text{-set}} \Pr\left(\left(\bigcap_{j\in\mathcal{V}_d}\left\{Y_j^n(\underline{w}) = Y_j^n(\underline{w}_0)\right\}\right)\bigcap\left(\bigcap_{i\in\mathcal{V}_d^c}\left\{Y_i^n(\underline{w}) \neq Y_i^n(\underline{w}_0)\right\}\right)|\mathcal{A}_\delta^{(n)}\right)\\
&\overset{(a)}{\le} \sum_{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x} \Pr\left(\bigcap_{i,j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]}\left\{\left(w^{(i)},Y_i^n(\underline{w})\right)\neq\left(w_0^{(i)},Y_i^n(\underline{w}_0)\right),\ Y_{ij}^n(\underline{w}) = Y_{ij}^n(\underline{w}_0)\right\}|\mathcal{A}_\delta^{(n)}\right)\\
&\overset{(b)}{=} \sum_{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x} \Pr\left(\bigcap_{i\in\mathcal{V}_x^*}\ \bigcap_{j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]}\left\{(w^{(i)},Y_i^n(\underline{w})\right)\neq\left(w_0^{(i)},Y_i^n(\underline{w}_0)\right),\ Y_{ij}^n(\underline{w}) = Y_{ij}^n\ (\underline{w}_0)\}|\mathcal{A}_\delta^{(n)}\right)\\
&\overset{(c)}{=} \sum_{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x} \Pr\left(\bigcap_{i\in\mathcal{V}_x^*}\ \bigcap_{j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]}\left\{Y_{ij}^n(\underline{w}) = Y_{ij}^n(\underline{w}_0)\right\}\big|\,\mathcal{A}_\delta^{(n)},\ \bigcap_{i\in\mathcal{V}_x^*}\left\{\left(w^{(i)},Y_i^n(\underline{w})\right)\neq\left(w_0^{(i)},Y_i^n(\underline{w}_0)\right)\right\}\right)\\
&\quad\cdot\Pr\left(\left(w^{(i)},Y_i^n(\underline{w})\right)\neq\left(w_0^{(i)},Y_i^n(\underline{w}_0)\right),\ \forall\, i\in\mathcal{V}_s^*\right)\\
&\le \sum_{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x} \Pr\left(\bigcap_{i\in\mathcal{V}_x^*}\ \bigcap_{j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]}\left\{Y_{ij}^n(\underline{w}) = Y_{ij}^n(\underline{w}_0)\right\}\big|\,\mathcal{A}_\delta^{(n)},\ \bigcap_{i\in\mathcal{V}_x^*}\left\{(w^{(i)},Y_i^n(\underline{w}))\neq(w_0^{(i)},Y_i^n(\underline{w}_0))\right\}\right)\\
&\overset{(d)}{=} \sum_{\substack{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x\\ d\notin\mathcal{V}_x}}\ \prod_{i\in\mathcal{V}_x^*} \Pr\left(\bigcap_{j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]}\left\{Y_{ij}^n(\underline{w}) = Y_{ij}^n(\underline{w}_0)\right\}\big|\,\mathcal{A}_\delta^{(n)},\ \left\{\left(w^{(i)},Y_i^n(\underline{w})\right)\neq(w_0^{(i)},Y_i^n(\underline{w}_0))\right\}\right). \quad (14)
\end{aligned}$$

Combining this with the last equation of (14) gives[11]

$$\begin{aligned}\Pr\left(E(\underline{w})|\mathcal{A}_\delta^{(n)}\right) &\le \sum_{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x}\prod_{i\in\mathcal{V}_x^*} \\ &\quad\times\exp\left(-n\left(1-\prod_{j:\,(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]}\epsilon_{ij}-\delta\right)\right) \\ &= \sum_{\mathcal{V}_x:\mathcal{M}(\underline{w})\in\mathcal{V}_x,d\notin\mathcal{V}_x} 2^{n|\mathcal{V}_x^*|\delta} \\ &\quad\cdot\exp\left(-n\sum_{i\in\mathcal{V}_x^*}\left(1-\prod_{j:\,(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]}\epsilon_{ij}\right)\right) \\ &\le 2^{n|\mathcal{V}|\delta}\sum_{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x,d\notin\mathcal{V}_x} 2^{-nC(\mathcal{V}_x)}. \qquad (16)\end{aligned}$$

Combining (11) and (16) together gives

$$\begin{aligned}\Pr(\mathrm{err}) &\le |\mathcal{V}|\delta + 2^{n|\mathcal{V}|\delta}\sum_{\underline{w}\in\underline{\mathcal{W}}-\{\underline{w}_0\}}\sum_{\substack{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x\\ d\notin\mathcal{V}_x}} 2^{-nC(\mathcal{V}_x)} \\ &= |\mathcal{V}|\delta + 2^{n|\mathcal{V}|\delta} \\ &\quad\times\sum_{\mathcal{M}\subset\mathcal{S}}\sum_{\substack{\underline{w}\in\underline{\mathcal{W}}-\{\underline{w}_0\}\\ \mathcal{M}(\underline{w})=\mathcal{M}}}\sum_{\substack{\mathcal{V}_x:\mathcal{M}(\underline{w})\subset\mathcal{V}_x\\ d\notin\mathcal{V}_x}} 2^{-nC(\mathcal{V}_x)} \\ &= |\mathcal{V}|\delta + 2^{n|\mathcal{V}|\delta} \\ &\quad\times\sum_{\mathcal{M}\subset\mathcal{S}}\sum_{\substack{\mathcal{V}_x:\mathcal{M}\subset\mathcal{V}_x\\ d\notin\mathcal{V}_x}}\sum_{\substack{\underline{w}\in\underline{\mathcal{W}}-\{\underline{w}_0\}\\ \mathcal{M}(\underline{w})=\mathcal{M}}} 2^{-nC(\mathcal{V}_x)} \\ &= |\mathcal{V}|\delta + 2^{n|\mathcal{V}|\delta} \\ &\quad\times\sum_{\mathcal{M}\subset\mathcal{S}}\sum_{\substack{\mathcal{V}_x:\mathcal{M}\subset\mathcal{V}_x\\ d\notin\mathcal{V}_x}}\prod_{s\in\mathcal{M}}(\lceil 2^{nR_s}\rceil-1)2^{-nC(\mathcal{V}_x)} \\ &\overset{(a)}{\le} |\mathcal{V}|\delta + 2^{n|\mathcal{V}|\delta} \\ &\quad\times\sum_{\mathcal{M}\subset\mathcal{S}}\sum_{\substack{\mathcal{V}_x:\mathcal{M}\subset\mathcal{V}_x\\ d\notin\mathcal{V}_x}} 2^{-n(C(\mathcal{V}_x)-\sum_{s\in\mathcal{M}}R_s)} \\ &\overset{(b)}{=} |\mathcal{V}|\delta + 2^{n|\mathcal{V}|\delta} \\ &\quad\times\sum_{\mathcal{V}_x:\mathcal{V}_x\subset\mathcal{V}-\{d\}}\sum_{\mathcal{M}\subset\mathcal{S}\cap\mathcal{V}_x} 2^{-n(C(\mathcal{V}_x)-\sum_{s\in\mathcal{M}}R_s)} \\ &\overset{(c)}{\le} |\mathcal{V}|\delta + 2^{n|\mathcal{V}|\delta} \\ &\quad\times\sum_{\mathcal{V}_x:\mathcal{V}_x\subset\mathcal{V}-\{d\}} 2^{|\mathcal{V}_x\cap\mathcal{S}|}2^{-n(C(\mathcal{V}_x)-\sum_{s\in\mathcal{S}\cap\mathcal{V}_x}R_s)} \qquad (17)\end{aligned}$$

where we have used the inequality $\lceil 2^{nR_s}\rceil - 1 \le 2^{nR_s}$ in (a). Also (b) is derived by changing the order of summation, and (c) follows from bounding $\sum_{s\in\mathcal{M}} R_s$ by $\sum_{s\in\mathcal{V}_x\cap\mathcal{S}} R_s$ in (c). Now by assumption, the rate vector $(R_s,\ \ s\in\mathcal{S})$ is inside the capacity region given in Theorem 2. Therefore, for any partition of the nodes into $\mathcal{V}_x$ and $\mathcal{V}_x^c \ni d$ we have $C(\mathcal{V}_x) - \sum_{s\in\mathcal{S}\cap\mathcal{V}_x} R_s > 0$.

[11]Using (15) it can be easily verified that the arguments that follow will exactly go through for correlated erasure events with cut-capacity $C(\mathcal{V}_x)$, defined as in (A3).

Therefore, the exponent in the last term of the above summation is negative. The above result holds for any $\delta > 0$ and n sufficiently large. By letting $n\to\infty$ and $\delta\to 0$, we can make the upper bound on the probability of error arbitrarily close to zero. Now by standard coding arguments we conclude that there exists some deterministic choice of encoding functions that has arbitrarily small probability of error for the rates in the achievable rate region $C(\mathcal{G},\mathcal{S},d)$.

B. Proof of Theorem 3

In this subsection, we outline the proof of Theorem 3. The analysis is very similar to Theorem 1. The converse part is straightforward. We know that the sources can be recovered at all the destinations, therefore, we have the same argument as the converse part of Theorem 1 for the sources and any of the destinations. In particular, for any destination $d_i,\ \ i\in\mathcal{D}$, we have $(R_s,\ s\in\mathcal{S})\in C(\mathcal{G},\mathcal{S},d_i)$ Therefore, any achievable rate vector should be in the intersection of these capacity regions, i.e.,

$$(R_s, s\in\mathcal{S}) \in \cap_{d_i\in\mathcal{D}} C(\mathcal{G},\mathcal{S},d_i) = C(\mathcal{G},\mathcal{S},\mathcal{D}).$$

Hence, the converse part is done.

In order to prove the achievability of the above rates, we can use random coding argument of Section VI-A2. Note that averaged over all the codebooks and functions, the probability of error for each destination goes to zero. Therefore, using the union bound on probability of events, the probability of having an error in at least one destination (averaged over all the functions and codebooks) goes to zero. Using standard arguments, there exists some deterministic choice of codebooks and functions for which the probability of error in the network become arbitrarily small and that shows the achievability of the rates in $C(\mathcal{G},\mathcal{S},\mathcal{D})$ of Theorem 3 for the multiple destination case.

VII. Linear Encoding

In Section VI-A2, we showed the achievability of the capacity region as defined in Theorem 2 by using general random coding functions at different nodes of the network. In this section, we restrict our attention to linear encoding schemes. The advantage of using a linear encoding scheme is that the decoding process becomes much easier. In this case, the equivalent transfer function of the network from any source to any destination, having the erasure locations at that destination, is linear. Hence, decoding at the destination is simply forming and solving a linear system of equations.

In this section, we show that linear encoders achieve capacity. Let us first define the linear block coding scheme with block length of n:

Recall that $\mathcal{W}^{(s)} = \{1, 2, \ldots, \lceil 2^{nR_s}\rceil\}$ is the message set for information source $s\in\mathcal{S}$. We assume that different messages are equiprobable and independent of each other. For any $w^{(s)}\in\mathcal{W}^{(s)}$, let $\underline{b(w^{(s)})}$ be the length-nR_s binary expansion of $w^{(s)}$–1.

The encoding operation is as follows.

Each node $i\in\mathcal{V}$ transmits n linear combinations of the non-erased symbols received from its incoming edges and the binary representation of the message it wants to transmit across the network. More precisely, node i generates a random binary matrix

B_i of size $n \times n(d_I(i) + R_i)$ where $d_I(i)$ is the in-degree of node i and R_i is the rate of the codebook used at node i (in the case where i is not a source of information $R_i = 0$). Each element of B_i is drawn i.i.d. Bernoulli$(1/2)$. For a given sequence y, let $\widetilde{y}$ be a sequence derived by replacing every e with 0. Note that $\widetilde{y}$ and y have the same length.[12] If node i receives $Y_i^n = y_i^n$ on its incoming edges and wants to transmit message $w^{(i)}$ then it sends out $x_i = B_i \cdot \underline{[b(w^{(i)}), \widetilde{y_i^n}]}^\dagger$. (Since the input–output relation at each node is linear, setting the erased symbols equal to zero is the same as finding linear combinations of only the nonerased bits.) Each destination d knows all the matrices B_i and also the erasure locations Z^n on all the links across the network. Since each received and transmitted symbol at any node is a linear combination of the elements of vector $b(\underline{w}) \triangleq \underline{(b(w^{(s)})}, s \in \mathcal{S})$. Therefore, each destination receives a collection of linear combinations of elements of $b(\underline{w})$. Using $\{B_i\}_{i\in\mathcal{V}}$ and Z^n, destination node d can construct the matrix that corresponds to the linear input–output relation of the network. We denote this matrix by $F(\{B_i\}, Z^n)$, giving

$$\widetilde{Y_d^n}(\underline{w}) = F(\{B_i\}, Z^n) \cdot b(\underline{w})^\dagger.$$

Note that matrix F is a function of different nodes' encoding matrices $\{B_i\}$ and Z^n.

Now, upon receiving $Y_d^n = y \in \{0, 1, e\}^{nd_I(d)}$, the destination node d looks (solves) for the message vector

$$\underline{w} \in \underline{\mathcal{W}} \triangleq \prod_{s\in\mathcal{S}} \mathcal{W}^{(s)}$$

such that $F(\{M_i\}, Z^n) \cdot b(\underline{w})^\dagger = \widetilde{y}$. If there is a unique $\underline{w}$ with this property, node d declares it as the transmitted message vector, otherwise, it declares an error. Note that the actual transmitted message vector, say $\underline{w}_0 \in \underline{\mathcal{W}}$, always satisfies the above property. Therefore, an error occurs only if there is another message vector $\underline{w} \neq \underline{w}_0$ such that $Y_d^n(\underline{w}) = Y_d^n(\underline{w}_0) = y$.

A. Achievability Result for Linear Encoding

Looking at the achievablity proof and probability of error analysis for general random coding in Sections VI-A2 and VI-A3, it can be easily verified that the linear case requires the same error events (10). Since the erasure vector Z^n is available at the destination, there is no difference between $\widetilde{Y}_i$ and Y_i and we can determine one from the other. By expanding the conditional error event $E(\underline{w})$ given $A_\delta^{(n)}$ for different cuts in the network, all of the relations up to step (d) of (14) go through for the linear case. In fact, the relations up to step (d) only require independence of encoding functions for different nodes of the network, which holds for the linear case. Now we look at the following probability in (14):

$$P_i \triangleq \Pr\left(\bigcap_{j:(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]} \{Y_{ij}^n(\underline{w}) = Y_{ij}^n(\underline{w}_0)\} \Big| \times \mathcal{A}_\delta^{(n)}, \left\{ \left(w^{(i)}, Y_i^n(\underline{w})\right) \neq \left(w_0^{(i)}, Y_i^n(\underline{w}_0)\right) \right\} \right). \quad (18)$$

[12]The corresponding mapping from alphabet GF$(q) \cup \{e\}$ to GF(q) again replaces e with 0. This variation is useful for packet erasure networks.

As in the general random coding argument, for a fixed i we have $Y_{ij}^n(\underline{w}) = Y_{ij}^n(\underline{w}_0)$ for all j such that $(i, j) \in [\mathcal{V}_x, \mathcal{V}_x^c]$, only if $X_i^n(\underline{w})$ and $X_i^n(\underline{w}_0)$ differ only in locations where an erasure occurs on all the edges of the interest. Because of strong typicality, the number of these location is at most

$$n\left(\prod_{j:\ (i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]} \epsilon_{ij} + \delta \right).$$

Therefore, $X_i^n(\underline{w})$ and $X_i^n(\underline{w}_0)$ should be the same in at least

$$n\left(1 - \prod_{j:\ (i,j)\in[\mathcal{V}_s,\mathcal{V}_s^c]} \epsilon_{ij} - \delta \right)$$

locations. This means that

$$B_i \cdot \underbrace{([w^{(i)}, Y_i^n(\underline{w})]^\dagger - [w_0^{(i)}, Y_i^n(\underline{w}_0)]^\dagger)}_{\underline{z}}$$

should be zero in at least

$$n\left(1 - \prod_{j:\ (i,j)\in[\mathcal{V}_s,\mathcal{V}_s^c]} \epsilon_{ij} + \delta \right)$$

specific locations. Also note that since $(w^{(i)}, Y_i^n(\underline{w})) \neq (w_0^{(i)}, Y_i^n(\underline{w}_0))$, $\underline{z}$ is a nonzero vector. From the above argument we have

$$P_i \leq \Pr\left(B_i \cdot \underline{z} \text{ be } \mathbf{0} \text{ in at least } n\alpha_i \text{ specific locations} \middle| \ \underline{z} \neq 0\right) \overset{\text{(a)}}{\leq} 2^{-n\alpha_i} = 2^{-n(1-\prod_{j:\,(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]} \epsilon_{ij} - \delta)} \quad (19)$$

where

$$\alpha_i = 1 - \prod_{j:\,(i,j)\in[\mathcal{V}_x,\mathcal{V}_x^c]} \epsilon_{ij} - \delta$$

and (a) follows from the following Lemma and its Corollary. Proof of this lemma is provided in Appendix B.

Lemma 1: Let X be a nonzero vector of size $n \times 1$ from some finite field GF(q). Suppose that A is a random matrix of size $m \times n$ with i.i.d. components distributed uniformly over GF(q). Then the coordinates of $Y = A \cdot X$ are i.i.d. uniform random variables over GF(q).

Corollary 2: The probability that $Y = A \cdot X$ is zero in k specific coordinates equals q^{-k}.

Now note that by replacing P_i in (18) and (14) with its bound from (19) we get the same bound as (16) for random linear codes. Therefore, linear operations are sufficient for achieving the capacity.

VIII. Discussion

A. Packet Size and Cycles

This paper treats binary erasure networks. However, as mentioned earlier, the results presented in this paper hold for any packet length (or more generally for any input alphabet size). The theorems stated in this paper give the maximum achievable rate per packet. Therefore, if one is interested in the maximum achievable rate per bit, assuming that the size of each packet is

L bits (resp., the alphabet size is M) across the network, the capacity will be L (resp., $\log_2 M$) times the capacity as stated in the theorems.

This paper assumes that the graph representation of the wireless erasure network is acyclic. However, the upper bound derived in Appendix A does not rely on this assumption. By an approach similar to [3], [5], [18 (Sec. 11.5.2)], it can be shown that the upper bound is still achievable and therefore the capacity theorems holds. We do not get into this problem in detail here.

B. Side Information at Destination Nodes

The results stated so far are based on perfect knowledge of the erasure locations for each link of the network to be available at destination nodes.

Erasure channels are usually used in modeling networks for which there exists a mechanism by which the receiver (destination) can be informed of a packet dropping. Usually, this side information is provided by using sequencing numbers in the packet header to detect lost packets. However, if we do not provide the destination with this side information, even for the simplest case of point-to-point communication, the capacity is not known. In this case, the communication system is usually modeled by the *deletion channel*. This channel has been studied by some researchers and lower and upper bounds on its capacity are found in [19]–[21].

Looking back at our network, for each block there are $n|\mathcal{E}|$ transmissions of packets across the network. Therefore, the erasure locations on the links of the network can be represented by $n|\mathcal{E}|$ bits. These $n|\mathcal{E}|$ should be provided to the destination through some mechanism. One approach is to use part of each packet as a header to transmit this information.

If the size of each packet is L bits, then based on our result we are able to send nCL bits across the network in a block of length n, where C is the minimum cut-capacity of the network. If the size of the packet, L, is large compared to the size of the network, or if the network is small, i.e., $|\mathcal{E}|$ is small, the amount of side information required is negligible compared to the amount of information sent across the network. We should remark that if one is trying to map a real network to our model, the packet length and the probability of erasure are closely related, and there will be some tradeoff between them.

A number of techniques can be used to reduce the required overhead for providing side information at the destination(s). For instance, consider a wireless erasure network, with one destination d. Let C denote the minimum cut-capacity for this node. Based on Theorem 1, this is the maximum achievable rate at d. Now consider $\mathcal{Q}$ to be the subset of nodes for which the minimum cut-capacity is greater than C. If we decode the messages completely and then re-encode them using random codebooks, we can still achieve the capacity at destination d. However, doing this may reduce the amount of overhead required at the destination. As an example, let us look at a line network (Fig. 5) from this point of view. The source node is the leftmost node and the destination is the rightmost node. The minimum cut-capacity for the destination is less than or equal to the minimum cut-capacity for every other node. Therefore, the intermediate nodes can decode without degrading the performance at the destination d. Further, this approach decreases the amount of overhead required. In fact, for this special case, by decoding at every node, no side information from previous links is required at the destination. Another technique is scheduling among the nodes to minimize the average header size needed for sending across the required side information. In this scheduling, for any link in the network, we determine the nodes that should include the erasure locations on that link as a header in their transmitted packets.

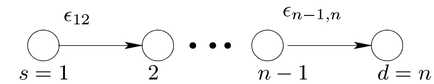

Fig. 5. A wireless erasure line network.

Remark: A closer look at the achievability proof of Section VI-A2 reveals that all that the destination nodes need to know is the mapping from the source nodes to their incoming signals for every instantiation of the network. (In other words, for every instantiation of the network, the destinations should be able to unambiguously compute their output for any given input to the network.) Any mechanism that provides destination nodes with the knowledge of this mapping will work. Providing the erasure locations for each link of the network is one possible mechanism. In the subsequent work of [16], another kind of side information is considered for the linear encoding scenario. There the side information is the global encoding vectors, which also allow for the input output mapping to be determined.

C. Achievable Rates Without Side Information

In this subsection, we look at the achievable rate for single-source/single-destination wireless erasure networks under a number of coding schemes when the side information is not available at the destination.

1) "Forward" and "Decode" Scheme: In this scheme, we analyze the performance of wireless erasure networks when limited operations are allowed at each node. Consider a codebook $\mathcal{C}$ of rate R and block size n. This codebook is available at all the nodes and is used for encoding the information message. The source node uses this codebook to "encode" the information. We assume that all the other nodes are allowed to perform one of the following operations.

- **Forward:** A node operating at this mode forward received strings unchanged.[13]
- **Decode and Re-encode:** In this case, the node first decodes the message transmitted from the source node based on what it has received and the codebook $\mathcal{C}$. Then it sends out the codeword corresponding to that message in $\mathcal{C}$ across its outgoing links. In this way, each relay node acts as a "source" of information for other nodes in the network. The need for successful decoding at intermediate nodes may reduce R the rate of the codebook used.

[13]In this scheme, we consider a modified erasure channel between any two nodes, in which the nodes can forward the erasure symbol without error. In other words, although similar to previous sections the nodes are not allowed to perform coding on the erasure symbol, they can inform their local neighbors of erasure of packets.

The main observation is that since the source message is intended to be decoded only at destination nodes, decoding at one relay node may be suboptimal. In fact, as is observed in [22]–[24], the distinguishing features of wireless media imply that decoding at every relay node and operating below the capacity of each link in the network can result in severe degradation in the achievable rate in the network. For instance, for the erasure wireless networks considered in this paper, the maximum rate when all the nodes are decoding the source message using codebook $\mathcal{C}$ is given by the minimum capacity of the links in the network, i.e.,

$$R_{AD} = 1 - \max_{(i,j)\in\mathcal{E}} \epsilon_{ij} \tag{20}$$

where subscript AD refers to the all-decoding case.

In the "Forward" and "Decode" scheme, instead of requiring that all the nodes decode the source message, we allow for another operation: "forward"ing. The objective is to find the optimal operation at every node so as to maximize the achievable rate, i.e., the rate of the codebook, in this network. In [22], we have proposed an efficient algorithm that finds the optimal rate for this scheme. In this paper, we use R_{FD} to refer to the optimal rate using the "Forward" and "Decode" scheme.

2) Block Markov Superposition Coding: Cover and El Gamal [10] developed a coding strategy for general single-relay channels based on block Markov coding and random partitioning. This strategy is generalized and used in a multiple-relay setup in [11], [12].

Let $\pi(\cdot)$ be a permutation on $\mathcal{V}$ that fixes the source and destination nodes. This permutation determines the order in which the nodes decode and encode the information. Using this coding strategy, it is shown in [11] that we can achieve

$$R_{BM} = \max_{\pi(\cdot)} \min_{1\le t\le|\mathcal{V}|-1} I(X_{\pi(1:t)}; Y_{\pi(t+1)}|X_{\pi(t+1:|\mathcal{V}|-1)})$$

where subscript BM is used to refer to block Markov coding and

$$\pi(i:j) = \{\pi(i), \pi(i+1), \ldots, \pi(j)\}.$$

We should remark that one can choose any distribution on $(X_1, X_2, \ldots, X_{|\mathcal{V}|})$ in the preceding result. Applying this result to our case, we show in Appendix C that the maximum rate is achieved when the X_i's are independent and uniformly distributed. Also, a maximizing permutation of the nodes is one that keeps the partial ordering of the nodes.[14] The maximum achievable rate in this case is

$$R_{BM} = \min_{j\in\mathcal{V}} \sum_{i:\,(i,j)\in\mathcal{E}} (1-\epsilon_{ij}). \tag{21}$$

The above formula suggests that the achievable rate in block Markov scheme is constrained by the minimum of the sum-capacities of incoming edges to any node in the network. This constraint is less severe than the all decoding case in (20). Here, instead of requiring all the links to be error free, we only require that the relay nodes be able to decode the information. However, the achievable rate is still constrained by the capability of each relay node to decode the information. The example that follows demonstrates that block Markov coding is not always efficient in our network.

[14]Since the network is acyclic, we can number the nodes such that if there is a path from node i to j, then the number assigned to node i is less than node j. This defines a partial ordering on the nodes of the graph.

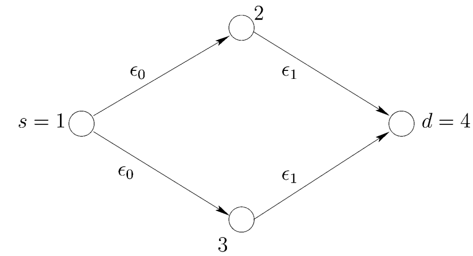

Fig. 6. A simple network.

There are other strategies such as compress forward and also partial decoding at the relay nodes that can be implemented in the network [10, Theorem 6], [25], [12]. Finding the achievable rates for these schemes explicitly and in terms of parameters of the network is usually intractable since the expressions for these schemes are not simple and involve a number of auxiliary random variables. In the rest of this section, we compare the performance of these schemes for a very simple but interesting erasure broadcast network. The capacity of this simple network is not known to our knowledge.

3) A Simple Example: The network under study here has a graph representation shown in Fig. 6. It has two relay nodes and one destination. We assume that the relay nodes are identical, in the sense that their connections to the source (resp., destination) have the same probability of erasure equal to ϵ_0 (resp., ϵ_1). The capacity of the network given side information at the destination is

$$C_{SI} = \min\{1-\epsilon_0^2, 2-2\epsilon_1, 2-\epsilon_1-\epsilon_0\}.$$

Using the "Forward" and "Decode" strategy, the maximum achievable rate is given by [22]

$$R_{FD} = \max\{1-\max\{\epsilon_1,\epsilon_0\}, 1-(1-(1-\epsilon_0)(1-\epsilon_1))^2\}.$$

Using (21), the block Markov coding scheme achieves rates up to

$$R_{BM} = \min\{1-\epsilon_0, 2(1-\epsilon_1)\}.$$

Another strategy that can be used is for the relay nodes to encode and compress their received signals, Y_i, with rate R_i and send it to the destination node reliably. Since the received signals at the relay nodes are correlated, the Slepian–Wolf encoding scheme can be used [26]. However, this scheme works only if the capacity of the channel between relay node i and the destination is larger than R_i. Combining this with the Slepian–Wolf rate region [26], we should have

$$H(Y_2|Y_3) \le 1-\epsilon_1, \qquad H(Y_2,Y_3) \le 2(1-\epsilon_1).$$

If the above conditions are satisfied, the destination will have access to both observations Y_2 and Y_3, and therefore, we can achieve a rate of $I(X_1;Y_2,Y_3)$ in the network.

Now suppose that the distribution on the input signal, X_1, is given by vector $\underline{p}$. It can be verified that

$$H(Y_2|Y_3) = H(\epsilon_0) + \epsilon_0(1-\epsilon_0)H(\underline{p})$$
$$H(Y_3,Y_2) = (1-\epsilon_0^2)H(\underline{p}) + 2H(\epsilon_0)$$

and

$$I(X_1;Y_2,Y_3) = (1-\epsilon_0^2)H(\underline{p}).$$

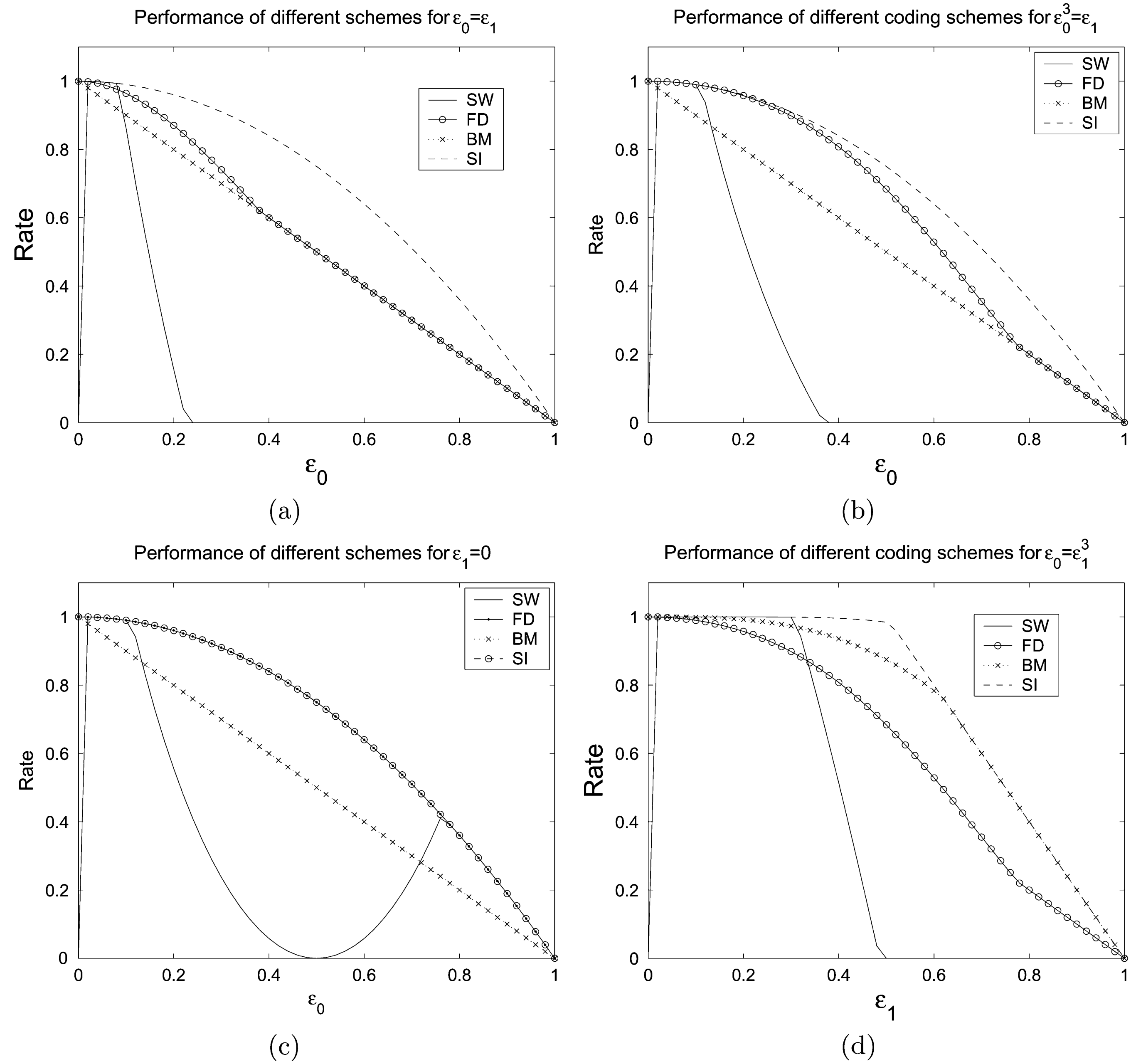

Fig. 7. Performance of different schemes for four scenarios.

By choosing the probability distribution appropriately we can achieve

$$R_{SW} = \min\left\{1-\epsilon_0^2, 2\{1-\epsilon_1-H(\epsilon_0)\}^+\right\}$$

where $\{x\}^+ = \max\{0, x\}$ and the subscript SW is used to refer to the Slepian–Wolf coding scheme. Note that this scheme does not work if the quality of the channels from the relay nodes to the destinations is low, i.e., if ϵ_1 is large. In this case, the second term in the above formula becomes zero and therefore R_{SW} equals zero.

We have plotted the performance of the above-mentioned strategies for four different scenarios in Fig. 7. In Fig. 7(a), we plot the performance for $\epsilon_0 = \epsilon_1 \in [0,1]$. For small values of ϵ_0, the Slepian–Wolf strategy achieves capacity. Unfortunately, this approach performs poorly for large values of ϵ_0 since the quality of the channel between relay nodes and the destination is not good enough to pass the compressed data reliably. Fig. 7(b) shows the results for $\epsilon_1 = \epsilon_0^3$. This choice corresponds to the case when the quality of the channels from the relay nodes to the destination is better than the source-to-relay connections. Note that the performance of the block Markov scheme is not good because of the rate constraint introduced by decoding at the relay nodes. The Slepian–Wolf scheme works for a larger range of ϵ_0 compared to the network considered in Fig. 7(a). This is because the channels between the source and the destination are better than in the former case. In Fig. 7(c), we look at one extreme case when ϵ_1 is zero, giving a perfect channel between the relays and the destination. As the figure shows, the "Forward" and "Decode" scheme achieves the capacity in this cases. The rate of the Slepian–Wolf scheme decreases for intermediate values of ϵ_0 but as we increase ϵ_0 again this scheme achieves capacity. This happens because for values of ϵ around 0.5, the entropy of the received signals at the relay nodes increases. Therefore, the minimum required rate for reliable transmission also increases. However, as we increase the value of ϵ the required rate decreases and the compressed signals can pass across without error. Hence, we achieve capacity. In Fig. 7(d), we look at the case when the quality of channels between the source and relays is better than of those between the relays and destination. For this, we have chosen $\epsilon_0 = \epsilon_1^3$. In this plot, block Markov coding outperforms other schemes and achieves capacity for large values of ϵ_1.

There are other strategies that can be used that are not analyzed in this paper. For instance, the authors of [10] propose a compression-based scheme for general relay channels. The approach is generalized to a multiple-relay setup in [12], [25]. This scheme is based on the Wyner–Ziv compression technique with side information at the receiver [27]. One expects this scheme to be more efficient than the Slepian–Wolf scheme proposed here. However, finding an explicit formula for the achievable rate requires a seemingly intractable optimization over a number of auxiliary random variables. Designing efficient and analytically tractable coding schemes based on these ideas deserves further investigation and can be a subject of further research.

IX. Conclusion and Further Work

We have obtained the capacity for a 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 [3], [5] to these networks. Furthermore, we have shown that linear encoding suffices to achieve the optimal performance. We see from the proof that it is not necessary to perform channel coding and network coding separately from each other. In fact, in [22], [23] we show that decoding at the relay nodes and operating below the capacities of each link can actually significantly reduce the achievable rate. Therefore, unlike the wireline scenario where each link is made error free by channel coding and network coding is then employed on top of that, our scheme only requires a single encoding function. Only the destination has to decode the received signal.

Many problems related to wireless networks remain open. Generalizing the results in this paper for other network problems is one possible extension of this work. As a first step, in broadcast problems over wireless erasure networks are considered. For these problems it can be shown that unlike wireline networks, the capacity region is not given by min-cut bounds. It is shown in [28] that the capacity region of multiple-input erasure broadcast channels is given by time sharing between users at different inputs. This result gives tighter outer bounds on the capacity region of broadcast problems in erasure wireless networks.

It will also be interesting to see if similar results can be obtained for other types of networks, such as erasure wireless networks in which interference is incorporated in the reception model, networks involving channels other than erasure channels, etc.

Appendix

A. Proof of Converse

We have to show that any sequence of

$$(\lceil 2^{nR_1} \rceil, \dots, \lceil 2^{nR_{|\mathcal{S}|}} \rceil, n)$$

codes with $P_{d_1}^{(n)(s)} \to 0$ satisfies the bounds given in Theorem 2 (and Theorem 1). Let $\underline{W} = (W^{(s)}, \quad s \in \mathcal{S})$ be a random vector drawn i.i.dt yo. from a uniform distribution over the set of message indices $\underline{\mathcal{W}}$. Let Z^n be the random vector describing the erasure locations, i.e., $Z^n = (Z_{ij,t}, (i,j) \in \mathcal{E}, t \in \{1, \dots, n\})$. Consider an $s-d$-cut given by s-set $\mathcal{V}_s$. We have (A1) and (A2) at the bottom of the following page, where

(a) follows from Fano's inequality since message $W^{(s)}$ can be decoded at node d from Y_d^n and the erasure locations Z^n across the network.

(b) follows from the properties of the block code defined in Section IV and data processing inequality. The causality of the block code and also the deterministic structure of the relaying functions can be used to inductively show that $(W^{(s)}, \ s \in \mathcal{S} \cap \mathcal{V}_d^c) - Y^n(\mathcal{V}_d^c) - Y_d^n$ forms a Markov chain for any cut. Applying the data processing inequality gives inequality (b).

(c) follows since messages and erasure locations are independent from each other.

(d) follows since the output of every channel is a deterministic function of the erasure locations Z^n and the transmitted messages $(W^{(s)}, \ s \in \mathcal{V}_d^c \cap \mathcal{S})$. Therefore the second conditional entropy is zero.

(e) follows from the fact that conditioning reduces the entropy.

(f) follows from the fact that conditioning reduces the entropy and $H(X_1, \dots, X_m) \leq \sum_{i=1}^m H(X_i)$ for any collection of random variables.

(g) follows from the fact that $(Z_{ij,t}, \ j : (i,j) \in [\mathcal{V}_d^c, \mathcal{V}_d])$ is a deterministic function of

$$(Y_{ij,t}, \ j : (i,j) \in [\mathcal{V}_d^c, \mathcal{V}_d]).$$

(h) follows since

$$H((Y_{ij,t}, \ j : (i,j) \in [\mathcal{V}_d^c, \mathcal{V}_d]) | X_{i,t}) = H(Z_{ij,t}, \ j : (i,j) \in [\mathcal{V}_d^c, \mathcal{V}_d]).$$

(i) follows from the capacity of the memoryless erasure channel. Here the transmitter transmits $X_{l,t}$ and the receiver has access to $Y_{lj,t}, (l,j) \in [\mathcal{V}_d^c, \mathcal{V}_d]$. The receiver experiences an erasure only if all channels $(l,j) \in [\mathcal{V}_d^c, \mathcal{V}_d]$ simultaneously suffer an erasure. Therefore, the equivalent channel's erasure probability is $\Pr(Z_{lj,t} = 1, \ j : (l,j) \in [\mathcal{V}_d^c, \mathcal{V}_d])$.

(j) follows since in the case of independent erasure events $\Pr(Z_{lj,t} = 1, \ j : (l,j) \in [\mathcal{V}_d^c, \mathcal{V}_d])$ is equal to $\prod_{j:\,(l,j) \in [\mathcal{V}_d^c, \mathcal{V}_d]} \epsilon_{lj}$

Remark 3: As we observe from (A1), the upper bound is in terms of the mutual information between the input and outputs of every broadcast system (i.e., a node and its outgoing edges) in the network. We should also mention that the same kind of upper bound of (A1) holds for more general networks (not necessarily erasure) that have interference-free and broadcast property.

Remark 4: From (A2), we see that in the case of correlated erasure events, i.e., when Z_{ij}s are dependent (however, still data independent), we can find an upper bound for the maximum achievable rate for each cut. Furthermore, as mentioned in footnote 11, it can be verified that the probability of error analysis of Section VI-A3 is valid for the correlated erasure events with the following definition of the cut-capacity:

$$C(\mathcal{V}_s) = \sum_{i\in\mathcal{V}_s^*} \left(1 - \Pr\left(Z_{ij} = 1,\ j : (i,j) \in [\mathcal{V}_s, \mathcal{V}_s^c]\right)\right). \tag{A3}$$

Therefore, the capacity results of this paper go through for the correlated erasure events as well.

B. Proof of Lemma 1

First note that if x and y are independent uniform random variables over GF (q) it can be easily verified that $x+y$ is also uniformly distributed over GF (q). By a simple induction it is straightforward that sum of any number of independent uniform random variables is uniformly distributed. By assumption different rows of A are independent from each other. Also, each element of Y is a linear combination of elements of one specific row. Therefore, different elements of Y are independent from each other. Now we show that elements of Y are uniformly distributed. Without loss of generality look at first element, i.e.,

$$y_1 = \sum_{i=1}^{n} a_{1k}x_k.$$

Note that for nonzero x_k's, the $a_{1k}x_k$ are independent uniformly distributed random variables. Hence, y_1 is a sum of a number of independent uniform random variables over GF (q) and based on the above discussion y_1 is uniformly distributed.

Remark: Using Bayes rule, we can easily check that if X is a nonzero uniform random vector over GF (q) and it is independent of A, then $A \cdot X$ is a uniform random vector.

C. Achievable Rates for Block Markov Scheme

As mentioned earlier the achievable rate using block Markov coding is given by

$$R_{BM} = \max_{\pi(.)} \min_{1\le t\le|\mathcal{V}|-1} I\left(X_{\pi(1:t)}; Y_{\pi(t+1)}|X_{\pi(t+1:|\mathcal{V}|-1)}\right)$$

$$\begin{aligned}
n\sum_{s\in\mathcal{V}_d^c\cap\mathcal{S}} R_s &= H\left(W^{(s)},\quad s\in\mathcal{S}\cap\mathcal{V}_d^c\right)\\
&= I\left(\left(W^{(s)},\quad s\in\mathcal{S}\cap\mathcal{V}_d^c\right); Y_d^n, Z^n\right) + H\left(\left(W^{(s)},\quad s\in\mathcal{S}\cap\mathcal{V}_d^c\right)\Big| Y_d^n, Z^n\right)\\
&\overset{(a)}{\le} I\left(\left(W^{(s)},\quad s\in\mathcal{S}\cap\mathcal{V}_d^c\right); Y_d^n, Z^n\right) + n\epsilon_n\\
&\overset{(b)}{\le} I\left(\left(W^{(s)},\quad s\in\mathcal{S}\cap\mathcal{V}_d^c\right); Y^n\left(\mathcal{V}_d^c\right), Z^n\right) + n\epsilon_n\\
&\overset{(c)}{=} I\left(\left(W^{(s)},\quad s\in\mathcal{S}\cap\mathcal{V}_d^c\right); Y^n\left(\mathcal{V}_d^c\right)|Z^n\right) + n\epsilon_n\\
&= H\left(Y^n\left(\mathcal{V}_d^c\right)|\, Z^n\right) - H\left(Y^n\left(\mathcal{V}_d^c\right)|\, Z^n, \left(W^{(s)},\quad s\in\mathcal{S}\cap\mathcal{V}_d^c\right)\right) + n\epsilon_n\\
&\overset{(d)}{=} H\left(Y^n\left(\mathcal{V}_d^c\right)|\, Z^n\right) + n\epsilon_n\\
&\overset{(e)}{\le} H\left(Y^n\left(\mathcal{V}_d^c\right)|\left(Z_{ij}^n, (i,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]\right)\right) + n\epsilon_n\\
&\overset{(f)}{\le} \sum_{t=1}^{n}\sum_{i\in\mathcal{V}_d^{c*}} H\left(Y_{ij,t},\quad j:(i,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]|\, Z_{ij,t},\quad j:(i,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]\right) + n\epsilon_n\\
&\overset{(g)}{\le} \sum_{t=1}^{n}\sum_{i\in\mathcal{V}_d^{c*}} H\left(Y_{ij,t},\quad j:(i,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]\right) - H\left(Z_{ij,t},\quad j:(i,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]\right) + n\epsilon_n\\
&\overset{(h)}{=} \sum_{t=1}^{n}\sum_{i\in\mathcal{V}_d^{c*}} H\left(Y_{ij,t}, j:(i,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]\right) - H\left(Y_{ij,t},\quad j:(i,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]\,|X_{i,t}\right) + n\epsilon_n\\
&= \sum_{t=1}^{n}\sum_{l\in\mathcal{V}_d^{c*}} I\left(X_{l,t}; \left(Y_{lj,t},\quad j:(l,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]\right)\right) + n\epsilon_n \qquad\qquad \text{(A1)}\\
&\overset{(i)}{\le} \sum_{t=1}^{n}\sum_{l\in\mathcal{V}_d^{c*}} \left(1-\Pr\left(Z_{lj}=1, j:(l,j)\in[\mathcal{V}_d^c,\mathcal{V}_d]\right)\right) + n\epsilon_n \qquad\qquad \text{(A2)}\\
&\overset{(j)}{\le} \sum_{t=1}^{n}\sum_{l\in\mathcal{V}_d^{c*}} \left(1-\prod_{j:(l,j)\in\left[\mathcal{V}_d^c,\mathcal{V}_d\right]} \epsilon_{lj}\right) + n\epsilon_n\\
&= nC\left(\mathcal{V}_d^c\right) + n\epsilon_n
\end{aligned}$$

where the maximization is over the joint distribution of the $(X_i, \ \ i \in \mathcal{V})$ and $\pi(\cdot)$, permutations of the nodes that keep source and destination fixed. Now for any mutual information term in the preceding equation we have

$$\begin{aligned} & I\left(X_{\pi(1:t)}; Y_{\pi(t+1)} | X_{\pi(t+1:|\mathcal{V}|-1)}\right) \\ & \quad = H\left(Y_{\pi(t+1)} | X_{\pi(t+1:|\mathcal{V}|-1)}\right) \\ & \quad\quad - H\left(Y_{\pi(t+1)} | X_{\pi(t+1:|\mathcal{V}|-1)}, X_{\pi(1:t)}\right) \\ & \quad \leq H\left(Y_{\pi(t+1)}\right) - H\left(Y_{\pi(t+1)} | X_i, \quad i \in \mathcal{V}\right) \\ & \quad \overset{\text{(a)}}{=} H(Y_{\pi(t+1)}) \\ & \quad\quad - H\left(Y_{\pi(t+1)} | X_i, \quad (i, \pi(t+1)) \in \mathcal{E}\right) \\ & \quad \overset{\text{(b)}}{\leq} \sum_{i:\,(i,\pi(t+1)) \in \mathcal{E}} H\left(Y_{i\pi(t+1)}\right) - H\left(Y_{i\pi(t+1)} | X_i\right) \\ & \quad = \sum_{i:\,(i,\pi(t+1)) \in \mathcal{E}} I\left(Y_{i\pi(t+1)}; X_i\right) \\ & \quad \overset{\text{(c)}}{\leq} \sum_{i:\,(i,\pi(t+1)) \in \mathcal{E}} 1 - \epsilon_{i\pi(t+1)} \end{aligned} \tag{C1}$$

where

- $(\mathrm{a}), (\mathrm{b})$ follow from the fact the network is memoryless, therefore, given $X_i \ \ (i,j) \in \mathcal{E}, Y_{ij}$'s are independent from each other and also output of other nodes.
- (c) follows from the fact that the capacity of an erasure channel with probability of erasure ϵ_{ij} is $1 - \epsilon_{ij}$.

Using (C1), we have

$$R_{\mathrm{BM}} \leq \min_{j \in \mathcal{V}} \sum_{i:\,(i,j) \in \mathcal{E}} 1 - \epsilon_{ij}.$$

Now we can easily verify that by choosing X_i's independent and uniformly distributed and by considering a permutation that is faithful to partial ordering of the nodes we can achieve the right-hand side of the above equation.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[2] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

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

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

[5] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[6] L. R. Ford Jr. and D. R. Fulkerson, *Flows in Networks*. Princeton, NJ: Princeton Univ. Press, 1962.

[7] N. Cai and R. W. Yeung, "Secure network coding," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 323.

[8] T. Ho, M. Médard, and R. Koetter, "An information theoretic view of network management," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1295–1312, Apr. 2005.

[9] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[10] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[11] L.-L. Xie and P. R. Kumar, "An achievable rate for multiple level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.

[12] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[13] Y. Wu, P. A. Chou, and S.-Y. Kung, "Minimum-energy multicast in mobile ad hoc networks using network coding," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1906–1918, Nov. 2005.

[14] D. S. Lun, M. Médard, T. Ho, and R. Koetter, "Network coding with a cost criterion," in *Proc. IEEE Int. Symp. Information Theory and Its Applications (ISITA 2004)*, Parma, Italy, Oct. 2004.

[15] M. Luby, "LT codes," in *Proc. 43rd Annu. IEEE Symp. Foundations of Computer Science*, Vancouver, BC, Canada, Nov. 2002, pp. 271–280.

[16] D. S. Lun, M. Médard, and M. Effros, "On coding for reliable communication over packet networks," in *Proc. 42nd Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2004.

[17] D. B. West, *Introduction to Graph Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[18] R. W. Yeung, *A First Course in Information Theory*. New York: Kluwer Academic/Plenum, 2002.

[19] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertion and reversals," *Sov. Phys.—Dokl.*, vol. 10, pp. 707–710, Feb. 1966.

[20] J. D. Ullman, "On the capabilities of codes to correct synchronization errors," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 95–105, Jan. 1967.

[21] S. N. Diggavi and M. Grossglauser, "On transmission over deletion channels," in *Proc. 39th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, 2001.

[22] R. Gowaikar, A. F. Dana, B. Hassibi, and M. Effros, "Practical schemes for wireless networks operation," *IEEE Trans. Commun.*, submitted for publication.

[23] A. F. Dana, M. Sharif, B. Hassibi, and M. Effros, "Is broadcast plus multi-access optimal for Gaussian wireless networks with fading?," in *Proc. 37th Asilomar Conf. Signals, Systems and Computing*, Pacific Grove, CA, Nov. 2003, pp. 1748–1752.

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

[25] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3061–3080, Dec. 2004.

[26] D. Slepian and J. K. Wolf, "Noiseless coding for correlated information sources," *IEEE Trans. In. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[27] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side-information at the receiver," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[28] A. F. Dana, R. Gowaikar, and B. Hassibi, "On the capacity region of broadcast over wireless erasure networks," in *Proc. 42nd Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2004.

[29] A. F. Dana and B. Hassibi, "The capacity region of multiple input erasure broadcast channel," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 2315–2319.