Classes of Delay-Independent Multimessage Multicast Networks with Zero-Delay Nodes

Silas L. Fong, *Member, IEEE*

Abstract

In a network, a node is said to *incur a delay* if its encoding of each transmitted symbol involves only its received symbols obtained before the time slot in which the transmitted symbol is sent (hence the transmitted symbol sent in a time slot cannot depend on the received symbol obtained in the same time slot). A node is said to *incur no delay* if its received symbol obtained in a time slot is available for encoding its transmitted symbol sent in the same time slot. Under the classical model, every node in the network incurs a delay. In this paper, we investigate the multimessage multicast network (MMN) under a generalized-delay model which allows some nodes to incur no delay. We obtain the capacity regions for three classes of MMNs with zero-delay nodes, namely the deterministic network dominated by product distribution, the MMN consisting of independent DMCs and the wireless erasure network. In addition, we show that for any MMN belonging to one of the above three classes, the set of achievable rate tuples under the generalized-delay model and under the classical model are the same, which implies that the set of achievable rate tuples for the MMN does not depend on the delay amounts incurred by the nodes in the network.

Index Terms

Multimessage multicast network (MMN), zero-delay nodes, capacity region, cut-set bound, delay-independent.

I. INTRODUCTION

In a multimessage multicast network (MMN), each source sends a message and each destination wants to decode all the messages. The set of source nodes and the set of destination nodes may not be disjoint. A node in the network is said to *incur a delay* if its encoding of each transmitted symbol involves only its received symbols obtained before the time slot in which the transmitted symbol is sent. In contrast, a node is said to *incur no delay* if its received symbol obtained in a time slot is available for encoding its transmitted symbol sent in the same time slot. Similarly, the network is said to *contain zero-delay nodes* if there exists a node that incurs zero delay on another node; the network is said to *contain no zero-delay node* if every node incurs a delay on all the nodes. In [1], the *capacity region* of the MMN with zero-delay nodes is defined to be the set of rate tuples achievable by all *feasible* schemes that do not include deadlock loops, and the *positive-delay region* is defined to be the set of rate
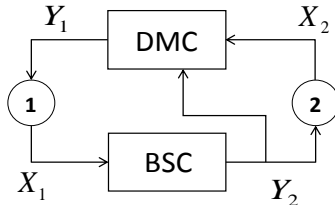
Fig. 1.   BSC-DMCF / BSC-CF

tuples achievable by all classical schemes under the constraint that each node incurs a delay (and hence deadlock loops are automatically excluded). By this definition, the positive-delay region is always a subset of the capacity region.

It is easy to construct a network with zero-delay nodes whose capacity region is strictly larger than the positive-delay region. One such network is the binary symmetric channel with correlated feedback (BSC-CF) considered in [1, Section VII], which will be introduced in the next section.

A. A Motivating Example

Consider a network that consists of two nodes denoted by 1 and 2 respectively. Node 1 and node 2 want to transmit a message to each other. This is a two-way channel [2]. Since we can assume without loss of generality that both nodes want to decode both messages, this network can be regarded as a MMN. In each time slot, node 1 and node 2 transmit X_1 and X_2 respectively, and they receive Y_1 and Y_2 respectively. All the input and output alphabets are binary, and the channel that carries X_1 to node 2 is a binary symmetric channel (BSC) while the channel that carries X_2 to node 1 is a discrete memoryless channel (DMC) whose output may depend on the output of the BSC. In this network, node 1 incurs zero delay on node 2, i.e., node 2 can receive Y_2 before encoding and transmitting X_2. We call this network the *BSC with DMC feedback* (BSC-DMCF), which is illustrated in Figure 1. The BSC-CF is a special case of the BSC-DMCF when $Y_1 = X_2 + Y_2$ [1, Section VII], where $+$ denotes the XOR operation. It is shown in [1, Section IX] that the capacity region of the BSC-CF is strictly larger than the positive-delay region (recall that the positive-delay region is obtained under the assumption that the network contains no zero-delay node while the capacity region of the BSC-CF is achieved when node 1 incurs zero delay on node 2). Other MMNs whose capacity regions are strictly larger than their positive-delay regions include the *relay-without-delay channel* studied by El Gamal *et al*. [3] and the *causal relay network* studied by Baik and Chung [4]. In other words, for some MMNs with zero-delay nodes, their capacity regions can be strictly larger than their positive-delay regions, which motivates us to classify the set of MMNs with zero-delay nodes into the following two categories:

(i)   *Delay-independent MMNs* whose capacity regions coincide with their positive-delay regions.

(ii)   *Delay-dependent MMNs* whose capacity regions are strictly larger than their positive-delay regions.

For each MMN in Category (i), the set of achievable rate tuples does not depend on the delay amounts incurred by the nodes in the network. On the other hand, for each MMN in Category (ii), the set of achievable rate tuples

shrinks if we impose the additional constraint that each node incurs a positive delay. It is important to decide which category a given MMN belongs to because the category of the MMN affects how the delays should be handled and how the transmissions in the network should be synchronized to achieve optimal performance.

B. Main Contribution

The main contributions of this work are identification of three classes of delay-independent MMNs and complete characterizations of their capacity regions. The first class is called the *deterministic MMN dominated by product distribution*. Being a subclass of MMNs consisting of deterministic channels, the deterministic MMN dominated by product distribution is a generalization of the deterministic relay network with no interference in [5] and the finite-field linear deterministic network in [6,7]. The second class is the *MMN consisting of independent DMCs* [8]. The third class is the *wireless erasure network* [9]. We successfully evaluate the capacity regions for the above classes of MMNs with zero-delay nodes and show that their capacity regions coincide with their positive-delay regions, which implies that the above classes of MMNs belong to the category of delay-independent MMNs. A natural consequence of our result is that for any MMN belonging to one of the above three classes, using different methods for handling delays and synchronization in the network does not affect the capacity region.

Given a MMN with zero-delay nodes belonging to one of the above three classes, in order to show its delay-independence, we first evaluate an achievable rate region for the MMN by invoking the noisy network coding (NNC) inner bound [10, Theorem 1] (which was also discovered in [11]). The achievable rate region is contained in the positive-delay region because the NNC inner bound was proved in [10] for classical MMNs. Then, we evaluate an outer bound on the capacity region of the MMN with zero-delay nodes by simplifying the cut-set outer bound in [1, Theorem 1] and show that the cut-set outer bound coincides with the NNC inner bound (which is within the positive-delay region), implying that the MMN is delay-independent.

This work should not be confused with the work by Effros [12], which shows that under the classical model which assumes a positive delay at every node, the set of achievable rate tuples for any MMN does not depend on the amount of positive delay incurred by each node. Here, we prove a different result for the above classes of MMNs with zero-delay nodes that their capacity regions and positive-delay regions are the same. Our result is meaningful given the fact that for some MMNs with zero-delay nodes, their capacity regions are strictly larger than their positive-delay regions (see Section I-A).

C. Paper Outline

This paper is organized as follows. Section II presents the notation used in this paper. Section III presents the formulation of the MMN with zero-delay nodes. Section IV recapitulates the NNC inner bound and the cut-set outer bound for the capacity region of the MMN with zero-delay nodes. In Section V, we use the two bounds obtained in Section IV to identify the three classes of delay-independent MMNs – the deterministic MMN dominated by product distributions, the MMN consisting of independent DMCs and the wireless erasure network, whose

problem formulations and proofs for delay-independence are contained in Section V-A, Section V-B and Section V-C respectively. Concluding remarks are given in Section VI.

II. Notation

We use $\Pr\{\mathcal{E}\}$ to represent the probability of an event $\mathcal{E}$, and use $\mathbf{1}\{\mathcal{E}\}$ to denote the characteristic function of $\mathcal{E}$. We use a capital letter X to denote a random variable with alphabet $\mathcal{X}$, and use the small letter x to denote the realization of X. We use X^n to denote a random tuple $(X_1, X_2, \ldots, X_n)$, where the components X_k have the same alphabet $\mathcal{X}$. We let p_X and $p_{Y|X}$ denote the probability mass distribution of X and the conditional probability mass distribution of Y given X respectively for any discrete random variables X and Y. We let $p_X(x) \triangleq \Pr\{X = x\}$ and $p_{Y|X}(y|x) \triangleq \Pr\{Y = y|X = x\}$ be the evaluations of p_X and $p_{Y|X}$ respectively at $X = x$ and $Y = y$. We let $p_X p_{Y|X}$ denote the joint distribution of (X, Y), i.e., $p_X p_{Y|X}(x, y) = p_X(x) p_{Y|X}(y|x)$ for all x and y. If X and Y are independent, their joint distribution is simply $p_X p_Y$. We will take all logarithms to base 2. For any discrete random variable (X, Y, Z) distributed according to $p_{X,Y,Z}$, we let $H_{p_{X,Z}}(X|Z)$ and $I_{p_{X,Y,Z}}(X; Y|Z)$ be the entropy of X given Z and mutual information between X and Y given Z respectively. For simplicity, we drop the subscript of a notation if there is no ambiguity. If X, Y and Z are distributed according to $p_{X,Y,Z}$ and they form a Markov chain, we write $(X \to Y \to Z)_{p_{X,Y,Z}}$ or more simply, $(X \to Y \to Z)_p$. The sets of natural and real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$ respectively. The closure of a set S is denoted by $\mathrm{closure}(S)$

III. Discrete Memoryless Multimessage Multicast Network with Zero-Delay Nodes

We consider a multimessage multicast network (MMN) that consists of N nodes. Let

$$\mathcal{I} \triangleq \{1, 2, \ldots, N\}$$

be the index set of the nodes, and let $\mathcal{V} \subseteq \mathcal{I}$ and $\mathcal{D} \subseteq \mathcal{I}$ be the sets of sources and destinations respectively. We call $(\mathcal{V}, \mathcal{D})$ the *multicast demand* on the network. The sources in $\mathcal{V}$ transmit information to the destinations in $\mathcal{D}$ in n time slots (channel uses) as follows. Each node $i \in \mathcal{V}$ transmits message

$$W_i \in \{1, 2, \ldots, M_i\}$$

and each node $j \in \mathcal{D}$ wants to decode all the messages $\{W_i : i \in \mathcal{V}\}$. We assume that each message W_i is uniformly distributed over $\{1, 2, \ldots, M_i\}$ and all the messages are independent. For each $k \in \{1, 2, \ldots, n\}$ and each $i \in \mathcal{I}$, node i transmits $X_{i,k} \in \mathcal{X}_i$ and receives $Y_{i,k} \in \mathcal{Y}_i$ in the kth time slot where $\mathcal{X}_i$ and $\mathcal{Y}_i$ are some alphabets that depend on i. After n time slots, node j declares $\hat{W}_{i,j}$ to be the transmitted W_i based on (W_j, Y_j^n) for each $(i, j) \in \mathcal{V} \times \mathcal{D}$.

To simplify notation, we use the following conventions for each $T \subseteq \mathcal{I}$: For any random tuple

$$(X_1, X_2, \ldots, X_N) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_N,$$

we let

$$X_T \triangleq (X_i : i \in T)$$

be a subtuple of $(X_1, X_2, \ldots, X_N)$. Similarly, for any $k \in \{1, 2, \ldots, n\}$ and any random tuple

$$(X_{1,k}, X_{2,k}, \ldots, X_{N,k}) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_N,$$

we let

$$X_{T,k} \triangleq (X_{i,k} : i \in T)$$

be a subtuple of $(X_{1,k}, X_{2,k}, \ldots, X_{N,k})$. For any N^2-dimensional random tuple $(\hat{W}_{1,1}, \hat{W}_{1,2}, \ldots, \hat{W}_{N,N})$, we let

$$\hat{W}_{T \times T^c} \triangleq (\hat{W}_{i,j} : (i, j) \in T \times T^c)$$

be a subtuple of $(\hat{W}_{1,1}, \hat{W}_{1,2}, \ldots, \hat{W}_{N,N})$.

We follow the formulation of the discrete memoryless network with zero-delay nodes in [1], which includes the following six definitions. The definitions are given here for completeness, and the detailed motivations behind them can be found in [1].

Definition 1: An α-dimensional tuple $(\mathcal{S}_1, \mathcal{S}_2, \ldots \mathcal{S}_\alpha)$ consisting of subsets of $\mathcal{I}$ is called an α-*partition of* $\mathcal{I}$ if $\cup_{h=1}^{\alpha} \mathcal{S}_h = \mathcal{I}$ and $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ for all $i \neq j$.

For any $(\mathcal{S}_1, \mathcal{S}_2, \ldots \mathcal{S}_\alpha)$ which is an α-partition of $\mathcal{I}$, we let

$$\mathcal{S}^h \triangleq \cup_{i=1}^{h} \mathcal{S}_i$$

for each $h \in \{1, 2, \ldots, \alpha\}$ to facilitate discussion.

Definition 2: The discrete network consists of N finite input sets $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N$, N finite output sets $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_N$ and α channels characterized by conditional distributions $q_{Y_{\mathcal{G}_1}|X_{\mathcal{S}^1}}^{(1)}$, $q_{Y_{\mathcal{G}_2}|X_{\mathcal{S}^2}, Y_{\mathcal{G}^1}}^{(2)}, \ldots, q_{Y_{\mathcal{G}_\alpha}|X_{\mathcal{S}^\alpha}, Y_{\mathcal{G}^{\alpha-1}}}^{(\alpha)}$, where

$$\boldsymbol{\mathcal{S}} \triangleq (\mathcal{S}_1, \mathcal{S}_2, \ldots \mathcal{S}_\alpha)$$

and

$$\boldsymbol{\mathcal{G}} \triangleq (\mathcal{G}_1, \mathcal{G}_2, \ldots \mathcal{G}_\alpha)$$

are two α-dimensional partitions of $\mathcal{I}$. We call $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{G}}$ the *input partition* and the *output partition* of the network respectively. The discrete network is denoted by $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ where

$$\boldsymbol{q} \triangleq (q^{(1)}, q^{(2)}, \ldots, q^{(\alpha)}).$$

Definition 3: A delay profile is an N-dimensional tuple $(b_1, b_2, \ldots, b_N)$ where $b_i \in \{0, 1\}$ for each $i \in \mathcal{I}$. The delay profile is said to be *positive* if its elements are all 1.

When we formally define a code on the discrete network later, a delay profile $B = (b_1, b_2, \ldots, b_N)$ will be associated with the code and b_i represents the amount of delay incurred by node i for the code. Under the classical model, B can only be positive, meaning that the amount of delay incurred by each node is positive. In contrast, under our generalized-delay model some elements of B can take 0 as long as deadlock loops do not occur. Therefore our model is a generalization of the classical model. The essence of the following definition is to characterize delay profiles which will not cause deadlock loops for the transmissions in the network.

Definition 4: Let $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ be a discrete network. For each $i \in \mathcal{I}$, let h_i and m_i be the two unique integers such that $i \in \mathcal{S}_{h_i}$ and $i \in \mathcal{G}_{m_i}$. Then, a delay profile $(b_1, b_2, \ldots, b_N)$ is said to be *feasible for the network* if the following holds for each $i \in \mathcal{I}$: If $b_i = 0$, then $h_i > m_i$.

Under the classical model, Definition 4 is trivial because any delay profile is positive and hence always feasible for the network. We are ready to define codes that use the network n times as follows.

Definition 5: Let $B \triangleq (b_1, b_2, \ldots, b_N)$ be a delay profile feasible for $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$, and let $(\mathcal{V}, \mathcal{D})$ be the multicast demand on the network. A $(B, n, M_\mathcal{I})$-code, where $M_\mathcal{I} \triangleq (M_1, M_2, \ldots, M_N)$ denotes the tuple of message alphabets, for n uses of the network consists of the following:

1) A message set
$$\mathcal{W}_i \triangleq \{1, 2, \ldots, M_i\}$$
   at node i for each $i \in \mathcal{I}$, where $M_i = 1$ for each $i \in \mathcal{V}^c$. Message W_i is uniformly distributed on $\mathcal{W}_i$.

2) An encoding function $f_{i,k} : \mathcal{W}_i \times \mathcal{Y}_i^{k-b_i} \to \mathcal{X}_i$ for each $i \in \mathcal{I}$ and each $k \in \{1, 2, \ldots, n\}$, where $f_{i,k}$ is the encoding function at node i in the kth time slot such that $X_{i,k} = f_{i,k}(W_i, Y_i^{k-b_i})$.

3) A decoding function $g_{i,j} : \mathcal{W}_j \times \mathcal{Y}_j^n \to \mathcal{W}_i$ for each $(i, j) \in \mathcal{V} \times \mathcal{D}$, where $g_{i,j}$ is the decoding function for W_i at node j such that
$$\hat{W}_{i,j} \triangleq g_{i,j}(W_j, Y_j^n).$$

Given a $(B, n, M_\mathcal{I})$-code, it follows from Definition 5 that for each $i \in \mathcal{I}$, node i incurs a delay if $b_i > 0$, where b_i is the amount of delay incurred by node i. If $b_i = 0$, node i incurs no delay, i.e., for each $k \in \{1, 2, \ldots, n\}$, node i needs to receive $Y_{i,k}$ before encoding $X_{i,k}$. The feasibility condition of B in Definition 4 ensures that the operations of any $(B, n, M_\mathcal{I})$-code are well-defined for the subsequently defined discrete memoryless network; the associated coding scheme is described after the network is defined.

Definition 6: A discrete network $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ with multicast demand $(\mathcal{V}, \mathcal{D})$, when used multiple times, is called a *discrete memoryless multimessage multicast network (DM-MMN)* if the following holds for any $(B, n, M_\mathcal{I})$-code:

Let $U^{k-1} \triangleq (W_\mathcal{I}, X_\mathcal{I}^{k-1}, Y_\mathcal{I}^{k-1})$ be the collection of random variables that are generated before the kth time slot.

Then, for each $k \in \{1, 2, \ldots, n\}$ and each $h \in \{1, 2, \ldots, \alpha\}$,

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^h, k} = x_{\mathcal{S}^h, k}, Y_{\mathcal{G}^h, k} = y_{\mathcal{G}^h, k}\}$$

$$= \Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^h, k} = x_{\mathcal{S}^h, k}, Y_{\mathcal{G}^{h-1}, k} = y_{\mathcal{G}^{h-1}, k}\} q^{(h)}_{Y_{\mathcal{G}_h} | X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}}(y_{\mathcal{G}_h, k} | x_{\mathcal{S}^h, k}, y_{\mathcal{G}^{h-1}, k}) \qquad (1)$$

for all $u^{k-1} \in \mathcal{U}^{k-1}$, $x_{\mathcal{S}^h, k} \in \mathcal{X}_{\mathcal{S}^h}$ and $y_{\mathcal{G}^h, k} \in \mathcal{Y}_{\mathcal{G}^h}$.

Following the notation in Definition 6, consider any $(B, n, M_{\mathcal{I}})$-code on the DM-MMN. In the kth time slot, $X_{\mathcal{I}, k}$ and $Y_{\mathcal{I}, k}$ are generated in the order

$$X_{\mathcal{S}_1, k}, Y_{\mathcal{G}_1, k}, X_{\mathcal{S}_2, k}, Y_{\mathcal{G}_2, k}, \ldots, X_{\mathcal{S}_\alpha, k}, Y_{\mathcal{G}_\alpha, k} \qquad (2)$$

by transmitting on the channels in this order $q^{(1)}, q^{(2)}, \ldots, q^{(\alpha)}$ using the $(B, n, M_{\mathcal{I}})$-code (as prescribed in Definition 5). Specifically, $X_{\mathcal{S}^h, k}$, $Y_{\mathcal{G}^{h-1}, k}$ and channel $q^{(h)}$ together define $Y_{\mathcal{G}_h, k}$ for each $h \in \{1, 2, \ldots, \alpha\}$. It is shown in [1, Section IV] that the encoding of $X_{\mathcal{S}_h, k}$ before the transmission on $q^{(h)}$ and the generation of $Y_{\mathcal{G}_h, k}$ after the transmission on $q^{(h)}$ for each $h \in \{1, 2, \ldots, \alpha\}$ are well-defined.

After defining the DM-MMN with zero-delay nodes in the above six definitions, we are now ready to formally define the *capacity region* and the *positive-delay region* through the following three intuitive definitions.

Definition 7: For a $(B, n, M_{\mathcal{I}})$-code on the DM-MMN, the average probability of decoding error P^n_{err} is defined as

$$P^n_{\text{err}} \triangleq \Pr\left\{ \bigcup_{(i,j) \in \mathcal{V} \times \mathcal{D}} \{\hat{W}_{i,j} \neq W_i\} \right\}.$$

Definition 8: Let B be a feasible delay profile for the network. A rate tuple $(R_1, R_2, \ldots, R_N)$, denoted by $R_{\mathcal{I}}$, is *B-achievable* for the DM-MMN if there exists a sequence of $(B, n, M_{\mathcal{I}})$-codes such that

$$\lim_{n \to \infty} \frac{\log M_i}{n} \geq R_i$$

for each $i \in \mathcal{I}$ and

$$\lim_{n \to \infty} P^n_{\text{err}} = 0.$$

Definition 9: The *B-capacity region*, denoted by $\mathcal{C}_B$, of the DM-MMN is the set consisting of every B-achievable rate tuple $R_{\mathcal{I}}$ with $R_i = 0$ for all $i \in \mathcal{V}^c$. The *capacity region* $\mathcal{C}$ is defined as

$$\mathcal{C} \triangleq \bigcup_{B:B \text{ is feasible}} \mathcal{C}_B$$

and the *positive-delay region* $\mathcal{C}_+$ is defined as

$$\mathcal{C}_+ \triangleq \bigcup_{B:B \text{ is positive}} \mathcal{C}_B.$$

If $\mathcal{C} = \mathcal{C}_+$, the DM-MMN is said to be *delay-independent*. If $\mathcal{C} \supsetneq \mathcal{C}_+$, the DM-MMN is said to be *delay-dependent*.

Roughly speaking, the capacity region is the set of rate tuples which are achievable by codes that do not incur a deadlock loop, and the positive-delay region is the set of rate tuples which are achievable by codes under the constraint that every node incurs a delay. Definitions 3, 4 and 9 imply that $\mathcal{C} \supseteq \mathcal{C}_+$, which implies that each DM-MMN is either delay-independent (i.e., $\mathcal{C} = \mathcal{C}_+$) or delay-dependent (i.e., $\mathcal{C} \supsetneq \mathcal{C}_+$).

IV. INNER AND OUTER BOUNDS FOR THE CAPACITY REGION

We start this section by stating an achievability result for classical DM-MMNs in the following theorem, which is a specialization of the main result of *noisy network coding (NNC)* inner bound by Lim, Kim, El Gamal and Chung [10] (the NNC inner bound was also discovered by Yassaee and Aref [11]).

Theorem 1: Let $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \mathcal{S}, \mathcal{G}, q)$ be a DM-MMN, and let

$$
\mathcal{R}_{\text{in}} \triangleq \bigcup_{\substack{p_{X_\mathcal{I}, Y_\mathcal{I}} : p_{X_\mathcal{I}, Y_\mathcal{I}} = \\ (\prod_{i=1}^N p_{X_i})(\prod_{h=1}^\alpha q_{Y_{\mathcal{G}_h}|X_\mathcal{S}^h, Y_\mathcal{G}^{h-1}}^{(h)})}} \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset}
$$
$$
\left\{ R_\mathcal{I} \,\middle|\, \begin{array}{l} \sum_{i \in T} R_i \leq I_{p_{X_\mathcal{I}, Y_\mathcal{I}}}(X_T; Y_{T^c}|X_{T^c}) - H_{p_{X_\mathcal{I}, Y_\mathcal{I}}}(Y_T|X_\mathcal{I}, Y_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{V}^c \end{array} \right\}. \tag{3}
$$

Then,

$$
\mathcal{R}_{\text{in}} \subseteq \mathcal{C}_+. \tag{4}
$$

Proof: For every (classical) $((1, 1, \ldots, 1), n, M_\mathcal{I})$-code, the MMN $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \mathcal{S}, \mathcal{G}, q)$ is equivalent to the MMN $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, 1, \mathcal{I}, \mathcal{I}, \prod_{h=1}^\alpha q_{Y_{\mathcal{G}_h}|X_\mathcal{S}^h, Y_\mathcal{G}^{h-1}}^{(h)})$ by Theorem 3 in [1]. The intuition behind the above equivalence can be reasoned as follows: If every node incurs a delay, then the outputs of the α channels in q will be independent given their inputs, and hence the relationship between the inputs and outputs of the network can be characterized simply by a product of the α channels, which is $\prod_{h=1}^\alpha q_{Y_{\mathcal{G}_h}|X_\mathcal{S}^h, Y_\mathcal{G}^{h-1}}^{(h)}$.

On the other hand, $\mathcal{R}_{\text{in}}$ is a specialization of the NNC inner bound in [10, Theorem 1] by taking $\hat{Y} = Y$ for the MMN $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, 1, \mathcal{I}, \mathcal{I}, \prod_{h=1}^\alpha q_{Y_{\mathcal{G}_h}|X_\mathcal{S}^h, Y_\mathcal{G}^{h-1}}^{(h)})$. Since the NNC inner bound was developed under the classical model where each node incurs a delay, any rate tuple in $\mathcal{R}_{\text{in}}$ is achievable by some sequence of $((1, 1, \ldots, 1), n, M_\mathcal{I})$-codes, which then implies (4). ■

Following similar procedures for proving Theorem 1 in [1], we can prove an outer bound on $\mathcal{C}$ stated in the following theorem. Since networks with zero-delay nodes can be viewed as networks with in-block memory formulated by Kramer [13, Section VII-D], the following theorem can be seen as the multicast version of Theorem 1 in [13].

Theorem 2: Let $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ be a DM-MMN, and let

$$
\mathcal{R}_{\text{out}} \triangleq \bigcup_{\substack{p_{X_\mathcal{I}, Y_\mathcal{I}}: p_{X_\mathcal{I}, Y_\mathcal{I}} = \\ \prod_{h=1}^{\alpha} (p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}}, Y_{\mathcal{G}^{h-1}}} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}})}} \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_\mathcal{I} \left| \begin{array}{l} \sum_{i \in T} R_i \leq \sum_{h=1}^{\alpha} I_{p_{X_\mathcal{I}, Y_\mathcal{I}}} (X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; Y_{T^c \cap \mathcal{G}_h} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}), \\ R_i = 0 \text{ for all } i \in \mathcal{V}^c \end{array} \right. \right\}. \tag{5}
$$

Then,

$$
\mathcal{C} \subseteq \mathcal{R}_{\text{out}}.
$$

Proof: Let $R_\mathcal{I}$ be an achievable rate tuple for the DM-MMN denoted by $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$. By Definitions 8 and 9, there exists a sequence of $(B, n, M_\mathcal{I})$-codes on the DM-MMN such that

$$
\lim_{n \to \infty} \frac{\log M_i}{n} \geq R_i \tag{6}
$$

for each $i \in \mathcal{I}$ and

$$
\lim_{n \to \infty} P_{\text{err}}^n = 0. \tag{7}
$$

Fix any $T \subseteq \mathcal{I}$ such that $T^c \cap \mathcal{D} \neq \emptyset$, and let d denote a node in $T^c \cap \mathcal{D}$. For each $(B, n, M_\mathcal{I})$-code, since the N messages $W_1, W_2, \ldots, W_N$ are independent, we have

$$
\sum_{i \in T} \log M_i = H(W_T | W_{T^c})
$$

$$
= I(W_T; Y_{T^c}^n | W_{T^c}) + H(W_T | Y_{T^c}^n, W_{T^c})
$$

$$
\leq I(W_T; Y_{T^c}^n | W_{T^c}) + H(W_T | Y_d^n, W_d)
$$

$$
\leq I(W_T; Y_{T^c}^n | W_{T^c}) + 1 + P_{\text{err}}^n \sum_{i \in T} \log M_i, \tag{8}
$$

where the last inequality follows from Fano's inequality (cf. Definition 7). Following similar procedures for proving Theorem 1 in [1], we can show by using (6), (7) and (8) that there exists a joint distribution $p_{X_\mathcal{I}, Y_\mathcal{I}}$ which depends on the sequence of $(B, n, M_\mathcal{I})$-codes but not on T such that

$$
p_{X_\mathcal{I}, Y_\mathcal{I}} = \prod_{h=1}^{\alpha} (p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}}, Y_{\mathcal{G}^{h-1}}} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}})
$$

and

$$
\sum_{i \in T} R_i \leq \sum_{h=1}^{\alpha} I_{p_{X_\mathcal{I}, Y_\mathcal{I}}} (X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; Y_{T^c \cap \mathcal{G}_h} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}). \tag{9}
$$

Since $p_{X_\mathcal{I}, Y_\mathcal{I}}$ depends on only the sequence of $(B, n, M_\mathcal{I})$-codes but not on T, (9) holds for all $T \subseteq \mathcal{I}$ such that $T^c \cap \mathcal{D} \neq \emptyset$. This completes the proof. ∎

V. Classes of Delay-Independent MMNs

In this section, we will use our inner and outer bounds developed in the previous section to calculate the capacity regions for some classes of MMNs with zero-delay nodes and then show that the MMNs are delay-independent, i.e., $\mathcal{C} = \mathcal{C}_+$. In the process of calculating their capacity regions, we will use the following proposition extensively to characterize an important property of Markov chains.

Proposition 1: Suppose there exist two probability distributions $r_{X,Y}$ and $q_{Z|Y}$ such that

$$p_{X,Y,Z} = r_{X,Y} q_{Z|Y}. \tag{10}$$

Then

$$(X \to Y \to Z)_{p_{X,Y,Z}} \tag{11}$$

forms a Markov chain. In addition,

$$p_{Z|Y} = q_{Z|Y}. \tag{12}$$

Proof: The proof of (11) is contained [14, Proposition 2.5]. In addition, (12) follows from (10). ∎

A. Deterministic MMN Dominated by Product Distribution

1) Problem Formulation and Main Result:

Definition 10: A conditional distribution $q_{Y|X}$ is said to be *deterministic* if for each $x^* \in \mathcal{X}$, there exists a $y^* \in \mathcal{Y}$ such that $q_{Y|X}(y^*|x^*) = 1$.

Definition 11: The MMN $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ is said to be *deterministic* if $q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{S}^{h-1}}}$ is deterministic for each $h \in \{1, 2, \ldots, \alpha\}$.

With the help of the following definition, we can completely characterize the capacity region for a class of deterministic MMNs with zero-delay nodes.

Definition 12: The deterministic MMN $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ is said to be *dominated by product distributions* if the following holds for each distribution $p_{X_{\mathcal{I}}}$:

Define s_{X_i} to be the marginal distribution of $p_{X_{\mathcal{I}}}$ for each $i \in \mathcal{I}$, i.e., $s_{X_i}(x_i) = \sum_{x_j : j \in \mathcal{I} \setminus \{i\}} p_{X_{\mathcal{I}}}(x_{\mathcal{I}})$ for all x_i. In addition, define $p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} \triangleq p_{X_{\mathcal{I}}} \prod_{h=1}^{\alpha} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}}$ and $s_{X_{\mathcal{I}}, Y_{\mathcal{I}}} \triangleq (\prod_{i=1}^{N} s_{X_i})(\prod_{h=1}^{\alpha} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}})$. Then for any $T \subseteq \mathcal{I}$, $H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{T^c}|X_{T^c}) \le H_{s_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{T^c}|X_{T^c})$.

The following is our main result in this section.

Theorem 3: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ be a deterministic MMN dominated by product distributions, and let

$$\mathcal{R}_{\text{in}}^{\text{det}} \triangleq \bigcup_{\substack{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} : p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} = \\ (\prod_{i=1}^{N} p_{X_i})(\prod_{h=1}^{\alpha} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}}^h, Y_{\mathcal{G}}^{h-1}})}} \bigcap_{T \subseteq \mathcal{I} : T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \le H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{T^c}|X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{V}^c \end{array} \right. \right\}. \tag{13}$$

Then,

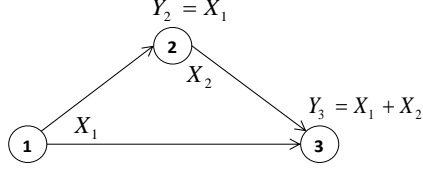$$\mathcal{C} = \mathcal{C}_+ = \mathcal{R}_{\text{in}}^{\text{det}} \tag{14}$$

Fig. 2. A relay channel under the finite-field linear deterministic model.

and hence the network is delay-independent. In particular, (14) holds for the deterministic relay network with no interference in [5] and the finite-field linear deterministic network in [6,7], which implies that they are delay-independent.

Remark 1: It has been shown in [5] that the capacity region of the deterministic relay network with no interference is contained in the classical cut-set bound even though the network contains zero-delay nodes. Therefore, it is intuitive that the capacity region of any deterministic MMN with zero-delay nodes should be contained in the classical cut-set bound. In addition, the cut-set bound can be achieved if the deterministic MMN is dominated by product distributions. Combining the intuition and the fact provided above, it is intuitive that Theorem 3 should hold.

Example 1: Consider a relay channel that consists of three nodes, where node 1 wants to transmit information to node 3 via a relay node 2. In each time slot, node i transmits X_i and receives Y_i for each $i \in \{1, 2, 3\}$. All the alphabets are assumed to be binary, and we assume that $Y_2 = X_1$ and $Y_3 = X_1 + X_2$. This relay channel is illustrated in Figure 2. The relay channel is a finite-field linear deterministic network [6], and it can be formulated as a deterministic MMN with zero-delay nodes by setting $\mathcal{S} \triangleq (\{1\}, \{2, 3\})$ and $\mathcal{G} \triangleq (\{2\}, \{1, 3\})$ and choosing appropriate $q_{Y_2|X_1}^{(1)}$ and $q_{Y_1,Y_3|X_1,X_2,X_3,Y_2}^{(2)}$ such that $Y_2 = X_1$ and $Y_3 = X_1 + X_2$ with probability one. Since node 2 incurs no delay under this formulation, we cannot characterize the capacity region by applying the classical cut-set bound. However, since every finite-field linear deterministic network is dominated by product distributions [10, Section II-A], Theorem 3 implies that this relay channel with a zero-delay node is delay-independent and its capacity region coincides with the classical cut-set bound. ∎

In the following, we provide the proof of Theorem 3. Since the last statement of the theorem follows from the fact that the deterministic relay network with no interference and the finite-field linear deterministic network are dominated by product distributions [10, Section II-A], it suffices to prove (14). To this end, it suffices to prove the achievability statement

$$\mathcal{R}_{\text{in}}^{\text{det}} \subseteq \mathcal{C}_+ \tag{15}$$

and the converse statement

$$\mathcal{C} \subseteq \mathcal{R}_{\text{in}}^{\text{det}}. \tag{16}$$

2) Achievability: In this subsection, we would like to show (15) by using Theorem 1 and Definition 11. Since $\mathcal{R}_{\text{in}} \subseteq \mathcal{C}_+$ by Theorem 1 (cf. (3)), it suffices to prove that

$$\mathcal{R}_{\text{in}} = \mathcal{R}_{\text{in}}^{\text{det}}. \tag{17}$$

Fix any $p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}$ that satisfies

$$p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} = \left(\prod_{i=1}^{N} p_{X_i} \right) \left(\prod_{h=1}^{\alpha} q_{Y_{\mathcal{G}_h} | X_{\mathcal{S}}^h, Y_{\mathcal{G}}^{h-1}}^{(h)} \right). \tag{18}$$

Since

$$p_{Y_{\mathcal{I}} | X_{\mathcal{I}}} = \prod_{h=1}^{\alpha} q_{Y_{\mathcal{G}_h} | X_{\mathcal{S}}^h, Y_{\mathcal{G}}^{h-1}}^{(h)}$$

by (18) and $q_{Y_{\mathcal{G}_h} | X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}}^{(h)}$ is deterministic for each $h \in \{1, 2, \ldots, \alpha\}$, it follows that $p_{Y_{\mathcal{I}} | X_{\mathcal{I}}}$ is deterministic and hence

$$H_{p_{Y_{\mathcal{I}} | X_{\mathcal{I}}}}(Y_{\mathcal{I}} | X_{\mathcal{I}}) = 0,$$

which then implies that

$$I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}) - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_T | X_{\mathcal{I}}, Y_{T^c}) = H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{T^c} | X_{T^c}). \tag{19}$$

Consequently, (17) follows from (3), (13) and (19).

3) Converse: In this subsection, we will show (16). Given a $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code on the deterministic MMN and the messages $W_{\mathcal{I}}$, a careful inspection of Definitions 5, 6 and 11 will reveal that $(X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n)$ is just a function of $W_{\mathcal{I}}$, which is formally stated in the following lemma. Since the proof of the lemma is straightforward, it is relegated to Appendix A.

Lemma 2: Let $(\mathcal{X}_{\mathcal{I}}, \alpha, \mathcal{Y}_{\mathcal{I}}, \mathcal{S}, \mathcal{G}, q)$ be a deterministic MMN. For any $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code on the network,

$$H_{p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n}}(X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n | W_{\mathcal{I}}) = 0, \tag{20}$$

where $p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n}$ is the joint distribution induced by the code according to Definitions 5 and 6.

In order to show that the capacity region of the deterministic MMN with zero-delay nodes lies within the classical cut-set bound, we will prove in Theorem 4, the theorem following the proposition below, that the deterministic MMN with zero-delay nodes is equivalent to some classical deterministic MMN. The following proposition is an important step for proving Theorem 4.

Proposition 3: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \mathcal{S}, \mathcal{G}, q)$ be a deterministic MMN. For any $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code on the network, if some u, $x_{\mathcal{I}}$ and $y_{\mathcal{I}}$ satisfy

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_{\mathcal{I}}\} > 0 \tag{21}$$

and

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_{\mathcal{I}}, Y_{\mathcal{I},k} = y_{\mathcal{I}}\} = 0, \tag{22}$$

then there exists some $h \in \{1, 2, \ldots, \alpha\}$ such that $q^{(h)}(y_{\mathcal{G}_h}|x_{\mathcal{S}^h}, y_{\mathcal{G}^{h-1}}) = 0$ (where $x_{\mathcal{S}^h}$ is a subtuple of $x_{\mathcal{I}}$, and $y_{\mathcal{G}_h}$ is a subtuple of $y_{\mathcal{I}}$).

Proof: Suppose there exist u, $x_{\mathcal{I}}$ and $y_{\mathcal{I}}$ that satisfy (21) and (22). We prove the proposition by assuming the contrary. Assume

$$q^{(h)}(y_{\mathcal{G}_h}|x_{\mathcal{S}^h}, y_{\mathcal{G}^{h-1}}) > 0 \tag{23}$$

for all $h \in \{1, 2, \ldots, \alpha\}$. We now prove by induction on h that

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^h,k} = x_{\mathcal{S}^h}, Y_{\mathcal{G}^h,k} = y_{\mathcal{G}^h}\} > 0 \tag{24}$$

for each $h \in \{1, 2, \ldots, \alpha\}$. For $h = 1$, the LHS of (24) is

$$\begin{aligned}
&\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^1,k} = x_{\mathcal{S}^1}, Y_{\mathcal{G}^1,k} = y_{\mathcal{G}^1}\} \\
&\stackrel{(a)}{=} p_{U^{k-1}, X_{\mathcal{S}^1,k}}(u^{k-1}, x_{\mathcal{S}^1})q^{(1)}(y_{\mathcal{G}_1}|x_{\mathcal{S}^1}) \\
&\stackrel{(b)}{>} 0,
\end{aligned} \tag{25}$$

where

  (a) follows from Definitions 6 and 11.

  (b) follows from (21) and (23).

If (24) holds for $h = m$, i.e.,

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^m,k} = x_{\mathcal{S}^m}, Y_{\mathcal{G}^m,k} = y_{\mathcal{G}^m}\} > 0, \tag{26}$$

then for $h = m + 1$ such that $m + 1 \le \alpha$,

$$\begin{aligned}
&\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^{m+1},k} = x_{\mathcal{S}^{m+1}}, Y_{\mathcal{G}^{m+1},k} = y_{\mathcal{G}^{m+1}}\} \\
&\stackrel{(a)}{=} p_{U^{k-1}, X_{\mathcal{S}^{m+1},k}, Y_{\mathcal{G}^m,k}}(u^{k-1}, x_{\mathcal{S}^{m+1}})q^{(m+1)}(y_{\mathcal{G}_{m+1}}|x_{\mathcal{S}^{m+1}}, y_{\mathcal{G}^m}) \\
&= p_{U^{k-1}, X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k}}(u^{k-1}, x_{\mathcal{S}^m}, y_{\mathcal{G}^m})q^{(m+1)}(y_{\mathcal{G}_{m+1}}|x_{\mathcal{S}^{m+1}}, y_{\mathcal{G}^m}) \\
&\quad p_{X_{\mathcal{S}_{m+1},k}|U^{k-1}, X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k}}(x_{\mathcal{S}_{m+1}}|u^{k-1}, x_{\mathcal{S}^m}, y_{\mathcal{G}^m}) \\
&\stackrel{(b)}{=} p_{U^{k-1}, X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k}}(u^{k-1}, x_{\mathcal{S}^m}, y_{\mathcal{G}^m})q^{(m+1)}(y_{\mathcal{G}_{m+1}}|x_{\mathcal{S}^{m+1}}, y_{\mathcal{G}^m}) \\
&\quad p_{X_{\mathcal{S}_{m+1},k}|U^{k-1}}(x_{\mathcal{S}_{m+1}}|u^{k-1}) \\
&\stackrel{(c)}{>} 0,
\end{aligned} \tag{27}$$

where

  (a) follows from Definitions 6 and 11.

  (b) follows from Lemma 2 that $(X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k})$ is a function of U^{k-1}.

  (c) follows from (26), (23) and (21).

Consequently, it follows from (25), (26) and (27) that (24) holds for $h = \alpha$ by mathematical induction, which then implies that $\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^\alpha,k} = x_{\mathcal{S}^\alpha}, Y_{\mathcal{G}^\alpha,k} = y_{\mathcal{G}^\alpha}\} > 0$, which contradicts (22). ∎

Surprisingly, each deterministic MMN with zero-delay nodes is equivalent to some classical deterministic MMN, which is proved as follows.

Theorem 4: For any $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code, the deterministic MMN specified by $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \mathcal{S}, \mathcal{G}, \boldsymbol{q})$ is equivalent to the deterministic MMN specified by $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, 1, \mathcal{I}, \mathcal{I}, q^{(1)} q^{(2)} \ldots q^{(\alpha)})$.

Proof: Fix a $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code, and let $U^{k-1} \triangleq (W_\mathcal{I}, X_\mathcal{I}^{k-1}, Y_\mathcal{I}^{k-1})$ be the collection of random variables that are generated before the kth time slot. To prove the theorem statement for this code, it suffices to show that the following two statements are equivalent for each $k \in \{1, 2, \ldots, n\}$ (cf. (1)):

Statement 1: For each $h \in \{1, 2, \ldots, \alpha\}$,

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^h,k} = x_{\mathcal{S}^h}, Y_{\mathcal{G}^h,k} = y_{\mathcal{G}^h}\}$$
$$= \Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^h,k} = x_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1},k} = y_{\mathcal{G}^{h-1}}\} q^{(h)}(y_{\mathcal{G}_h} | x_{\mathcal{S}^h}, y_{\mathcal{G}^{h-1}}). \tag{28}$$

Statement 2:

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_\mathcal{I}, Y_{\mathcal{I},k} = y_\mathcal{I}\}$$
$$= \Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_\mathcal{I}\} \prod_{h=1}^{\alpha} q^{(h)}(y_{\mathcal{G}_h} | x_{\mathcal{S}^h}, y_{\mathcal{G}^{h-1}}). \tag{29}$$

Fix a $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code and a $k \in \{1, 2, \ldots, n\}$. We first show that (28) implies (29). Suppose (28) holds for each $h \in \{1, 2, \ldots, \alpha\}$. Consider the following three mutually exclusive cases:

Case $\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_\mathcal{I}\} = 0$**:**

Both the LHS and the RHS of (29) equal zero.

Case $\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_\mathcal{I}\} > 0$ **and**

$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_\mathcal{I}, Y_{\mathcal{I},k} = y_\mathcal{I}\} = 0$**:**

For this case, the LHS of (29) equals zero. By Proposition 3, there exists some $h \in \{1, 2, \ldots, \alpha\}$ such that $q^{(h)}(y_{\mathcal{G}_h} | x_{\mathcal{S}^h}, y_{\mathcal{G}^{h-1}}) = 0$, which implies that the RHS of (29) equals zero.

Case $\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_\mathcal{I}, Y_{\mathcal{I},k} = y_\mathcal{I}\} > 0$**:**

For this case,

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_\mathcal{I}, Y_{\mathcal{I},k} = y_\mathcal{I}\}$$
$$= p_{U^{k-1}, X_{\mathcal{I},k}}(u^{k-1}, x_\mathcal{I}) \prod_{h=1}^{\alpha} p_{Y_{\mathcal{G}_h,k} | U^{k-1}, X_{\mathcal{I},k}, Y_{\mathcal{G}^{h-1},k}}(y_{\mathcal{G}_h} | u^{k-1}, x_\mathcal{I}, y_{\mathcal{G}^{h-1}})$$
$$\stackrel{(a)}{=} p_{U^{k-1}, X_{\mathcal{I},k}}(u^{k-1}, x_\mathcal{I}) \prod_{h=1}^{\alpha} p_{Y_{\mathcal{G}_h,k} | U^{k-1}, X_{\mathcal{S}^h,k}, Y_{\mathcal{G}^{h-1},k}}(y_{\mathcal{G}_h} | u^{k-1}, x_{\mathcal{S}^h}, y_{\mathcal{G}^{h-1}})$$
$$\stackrel{(28)}{=} p_{U^{k-1}, X_{\mathcal{I},k}}(u^{k-1}, x_\mathcal{I}) \prod_{h=1}^{\alpha} q^{(h)}(y_{\mathcal{G}_h} | x_{\mathcal{S}^h}, y_{\mathcal{G}^{h-1}}),$$

where (a) follows from Lemma 2 that for the $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code, $X_{\mathcal{I},k}$ is a function of U^{k-1}. Therefore, the LHS and the RHS of (29) are equal.

Combining the three mutually exclusive cases, we obtain that (28) implies (29). We now show that (29) implies (28). Suppose (29) holds. Then for each $h \in \{1, 2, \ldots, \alpha\}$ and each $m \in \{1, 2, \ldots, h\}$,

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{S}^h,k} = x_{\mathcal{S}^h}, Y_{\mathcal{G}^m,k} = y_{\mathcal{G}^m}\}$$

$$= \sum_{\substack{x_{\mathcal{S}_{h+1}},\ldots,x_{\mathcal{S}_\alpha} \\ y_{\mathcal{G}_{m+1}},\ldots,y_{\mathcal{G}_\alpha}}} p_{U^{k-1},X_{\mathcal{I},k},Y_{\mathcal{I},k}}(u^{k-1}, x_{\mathcal{I}}, y_{\mathcal{I}})$$

$$\overset{(29)}{=} \sum_{\substack{x_{\mathcal{S}_{h+1}},\ldots,x_{\mathcal{S}_\alpha} \\ y_{\mathcal{G}_{m+1}},\ldots,y_{\mathcal{G}_\alpha}}} p_{U^{k-1},X_{\mathcal{I},k}}(u^{k-1}, x_{\mathcal{I}}) \prod_{\ell=1}^{\alpha} q^{(\ell)}(y_{\mathcal{G}_\ell}|x_{\mathcal{S}^\ell}, y_{\mathcal{G}^{\ell-1}})$$

$$= \sum_{x_{\mathcal{S}_{h+1}},\ldots,x_{\mathcal{S}_\alpha}} p_{U^{k-1},X_{\mathcal{I},k}}(u^{k-1}, x_{\mathcal{I}}) \prod_{\ell=1}^{m} q^{(\ell)}(y_{\mathcal{G}_\ell}|x_{\mathcal{S}^\ell}, y_{\mathcal{G}^{\ell-1}})$$

$$\overset{(a)}{=} p_{U^{k-1},X_{\mathcal{S}^h,k}}(u^{k-1}, x_{\mathcal{S}^h}) \prod_{\ell=1}^{m} q^{(\ell)}(y_{\mathcal{G}_\ell}|x_{\mathcal{S}^\ell}, y_{\mathcal{G}^{\ell-1}}), \tag{30}$$

where (a) follow from the fact that $m \le h$. Then, for each $h \in \{1, 2, \ldots, \alpha\}$, the equality in (28) can be verified by substituting (30) into the LHS and the RHS. ∎

The following lemma simplifies the outer bound in Theorem 2 for the deterministic MMN.

Lemma 4: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ be a deterministic MMN. Define

$$\mathcal{R}_{\text{out}}^{\text{det}} \triangleq \bigcup_{\substack{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}:p_{X_{\mathcal{I}},Y_{\mathcal{I}}}= \\ p_{X_{\mathcal{I}}}\prod_{h=1}^{\alpha} q_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h},Y_{\mathcal{G}^{h-1}}}^{(h)}}} \bigcap_{T \subseteq \mathcal{I}:T^c \cap \mathcal{D} \ne \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \le H_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(Y_{T^c}|X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{V}^c \end{array} \right. \right\}. \tag{31}$$

Then,

$$\mathcal{C} \subseteq \mathcal{R}_{\text{out}}^{\text{det}}.$$

Proof: Suppose $R_{\mathcal{I}}$ is an achievable rate tuple for $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$. It follows from Definition 8 that there exists a sequence of $(B, n, M_{\mathcal{I}})$-codes on $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ such that $\lim_{n\to\infty} \frac{\log M_i}{n} \ge R_i$ for each $i \in \mathcal{I}$ and $\lim_{n\to\infty} P_{\text{err}}^n = 0$. Since $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ is equivalent to $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, 1, \mathcal{I}, \mathcal{I}, q^{(1)}q^{(2)} \ldots q^{(\alpha)})$ for any $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code on the deterministic DMN by Theorem 4, it follows from Definition 5 that $R_{\mathcal{I}}$ is achievable for $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, 1, \mathcal{I}, \mathcal{I}, q^{(1)}q^{(2)} \ldots q^{(\alpha)})$, which then implies from Theorem 2 that there exists some $p_{X_{\mathcal{I}},Y_{\mathcal{I}}}^*$ satisfying

$$p_{X_{\mathcal{I}},Y_{\mathcal{I}}}^* = p_{X_{\mathcal{I}}}^* \prod_{h=1}^{\alpha} q_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h},Y_{\mathcal{G}^{h-1}}}^{(h)} \tag{32}$$

such that for any $T \subseteq \mathcal{I}$ such that $T^c \cap \mathcal{D} \ne \emptyset$,

$$\sum_{i \in T} R_i \le I_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}^*}(X_T; Y_{T^c}|X_{T^c})$$

$$= H_{p^*_{X_\mathcal{I},Y_\mathcal{I}}}(Y_{T^c}|X_{T^c}) - H_{p^*_{X_\mathcal{I},Y_\mathcal{I}}}(Y_{T^c}|X_\mathcal{I})$$

$$\stackrel{(a)}{=} H_{p^*_{X_\mathcal{I},Y_\mathcal{I}}}(Y_{T^c}|X_{T^c}) \tag{33}$$

where (a) follows from the fact that

$$p^*_{Y_\mathcal{I}|X_\mathcal{I}} \stackrel{(32)}{=} \prod_{h=1}^{\alpha} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h},Y_{\mathcal{G}^{h-1}}}$$

is deterministic. Consequently, it follows from (31), (32) and (33) that $R_\mathcal{I} \in \mathcal{R}^{\text{det}}_{\text{out}}$. ∎

We are now ready to prove (16) as follows. Using Lemma 4, we obtain $\mathcal{C} \subseteq \mathcal{R}^{\text{det}}_{\text{out}}$ where $\mathcal{R}^{\text{det}}_{\text{out}}$ is defined in (31). In addition, it follows from (13), (31) and Definition 12 that $\mathcal{R}^{\text{det}}_{\text{out}} \subseteq \mathcal{R}^{\text{det}}_{\text{in}}$. Consequently, $\mathcal{C} \subseteq \mathcal{R}^{\text{det}}_{\text{in}}$.

B. MMN Consisting of Independent DMCs

1) Problem Formulation and Main Result: Consider a DM-MMN $(\mathcal{X}_\mathcal{I}, \mathcal{Y}_\mathcal{I}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ defined as follows: The edge set of the network is characterized by

$$\Omega \triangleq \bigcup_{h=1}^{\alpha} \mathcal{S}^h \times \mathcal{G}_h, \tag{34}$$

and a DMC denoted by $q_{Y_{i,j}|X_{i,j}}$ is associated with every edge $(i,j) \in \Omega$, where $\mathcal{X}_{i,j}$ and $\mathcal{Y}_{i,j}$ are the input and output alphabets of the DMC carrying information from node i to node j. The definition of Ω in (34) ensures that $q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h},Y_{\mathcal{G}^{h-1}}}$ can be well-defined for each $h \in \{1, 2, \ldots, \alpha\}$. For each $(i,j) \in \Omega$, the capacity of channel $q_{Y_{i,j}|X_{i,j}}$, denoted by $C_{i,j}$, is attained by some $\bar{p}_{X_{i,j}}$ due to the channel coding theorem, i.e.,

$$C_{i,j} \triangleq \max_{p_{X_{i,j}}} I_{p_{X_{i,j}} q_{Y_{i,j}|X_{i,j}}}(X_{i,j}; Y_{i,j})$$

$$= I_{\bar{p}_{X_{i,j}} q_{Y_{i,j}|X_{i,j}}}(X_{i,j}; Y_{i,j}). \tag{35}$$

For all the other $(\tilde{i}, \tilde{j}) \in \Omega^c$, we assume without loss of generality that

$$\mathcal{X}_{\tilde{i},\tilde{j}} = \mathcal{Y}_{\tilde{i},\tilde{j}} = \{0\} \tag{36}$$

and $C_{\tilde{i},\tilde{j}} = 0$. Then, we define the input and output alphabets for each node i in the following natural way:

$$\mathcal{X}_i \triangleq \mathcal{X}_{i,1} \times \mathcal{X}_{i,2} \times \ldots \times \mathcal{X}_{i,N} \tag{37}$$

and

$$\mathcal{Y}_i \triangleq \mathcal{Y}_{1,i} \times \mathcal{Y}_{2,i} \times \ldots \times \mathcal{Y}_{N,i} \tag{38}$$

for each $i \in \mathcal{I}$. In addition, we define

$$q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h},Y_{\mathcal{G}^{h-1}}} \triangleq \prod_{(i,j)\in\mathcal{S}^h\times\mathcal{G}_h} q_{Y_{i,j}|X_{i,j}} \tag{39}$$
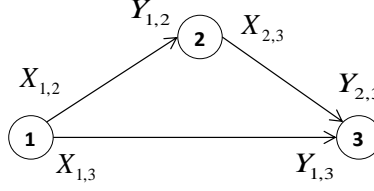
Fig. 3.   A relay channel consisting of independent DMCs.

for each $h \in \{1, 2, \ldots, \alpha\}$, i.e., the random transformations (noises) from $X_{i,j}$ to $Y_{i,j}$ are independent and each channel $q^{(h)}_{Y_{\mathcal{G}_h} | X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}}$ is in a product form. We call the network described above the *DM-MMN consisting of independent DMCs*. The classical MMN consisting of independent DMCs studied in [8] is a special case of this network model when $\alpha = 1$ and $\Omega = \mathcal{I} \times \mathcal{I}$. The following is the main result in this section, and the proof will be presented in the next two subsections.

Theorem 5: For the DM-MMN consisting of independent DMCs, define

$$\mathcal{R}^{\mathrm{DMCs}}_{\mathrm{in}} \triangleq \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq \sum_{(i,j) \in T \times T^c} C_{i,j}, \\ R_i = 0 \text{ for all } i \in \mathcal{V}^c \end{array} \right. \right\}. \tag{40}$$

Then,

$$\mathcal{C} = \mathcal{C}_+ = \mathcal{R}^{\mathrm{DMCs}}_{\mathrm{in}}$$

and hence the network is delay-independent.

Remark 2: In network coding theory, it is well-known that the classical cut-set bound (also called *max-flow bound*) always holds for networks consisting of noiseless bit-pipes with zero-delay nodes [14, Chapter 18]. Therefore, it is intuitive that by replacing the noiseless bit-pipes by independent DMCs, the cut-set bound still serves as an outer bound on the capacity region. On the other hand, the cut-set bound can be achieved for MMNs consisting of independent DMCs. Combining the intuition and the fact provided above, it is intuitive that Theorem 5 should hold.

Example 2: Consider a relay channel that consists of three nodes and three edges connecting the nodes, where node 1 wants to transmit information to node 3 via a relay node 2 through edges $(1, 2)$, $(1, 3)$ and $(2, 3)$. In each time slot, node i transmits $X_{i,j}$ to node j through edge (i, j) and receives $Y_{\ell,i}$ from node ℓ through edge (ℓ, i). Each edge is associated with a DMC. The three DMCs associated with the three edges, denoted by $q_{Y_{1,2}|X_{1,2}}$, $q_{Y_{1,3}|X_{1,3}}$ and $q_{Y_{2,3}|X_{2,3}}$ respectively, are assumed to be independent, i.e.,

$$p_{Y_{1,2}, Y_{1,3}, Y_{2,3} | X_{1,2}, X_{1,3}, X_{2,3}} = q_{Y_{1,2}|X_{1,2}} q_{Y_{1,3}|X_{1,3}} q_{Y_{2,3}|X_{2,3}}$$

regardless of the distribution of $(X_{1,2}, X_{1,3}, X_{2,3})$. This relay channel is illustrated in Figure 3. The relay channel can be formulated as a DM-MMN consisting of independent DMCs by setting $\boldsymbol{\mathcal{S}} \triangleq (\{1\}, \{2, 3\})$, $\boldsymbol{\mathcal{G}} \triangleq (\{2\}, \{1, 3\})$, $\Omega \triangleq (\mathcal{S}_1 \times \mathcal{G}_1) \cup (\mathcal{S}^2 \times \mathcal{G}_2)$, $X_1 \triangleq (X_{1,2}, X_{1,3})$, $X_2 \triangleq X_{2,3}$, $Y_2 \triangleq Y_{1,2}$, $Y_3 \triangleq (Y_{1,3}, Y_{2,3})$, $q^{(1)}_{Y_2|X_1} \triangleq q_{Y_{1,2}|X_{1,2}}$

and $q_{Y_3|X_1,X_2}^{(2)} \triangleq q_{Y_{1,3}|X_{1,3}} q_{Y_{2,3}|X_{2,3}}$. The set of non-trivial edges $\{(1,2),(1,3),(2,3)\}$ is inside Ω by (34). Since node 2 incurs no delay under this formulation, we cannot characterize the capacity region by applying the classical cut-set bound. Surprisingly, Theorem 5 implies that this relay channel with a zero-delay node is delay-independent and its capacity region coincides with the classical cut-set bound. ∎

In the following, we provide the proof of Theorem 5. To this end, it suffices to prove the achievability statement

$$\mathcal{R}_{\text{in}}^{\text{DMCs}} \subseteq \mathcal{C}_+ \tag{41}$$

and the converse statement

$$\mathcal{C} \subseteq \mathcal{R}_{\text{in}}^{\text{DMCs}}. \tag{42}$$

2) Achievability: In this subsection, we would like to prove (41). Since the DMCs $q_{Y_{i,j}|X_{i,j}}$ are all independent and each of the DMC can carry information at a rate arbitrarily close to the capacity, it is intuitive that $\mathcal{R}_{\text{in}}^{\text{DMCs}}$ lies in the positive-delay region of the DM-MMN consisting of independent DMCs, which is proved as follows.

Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ be the DM-MMN consisting of independent DMCs whose positive-delay region is denoted by $\mathcal{C}_+$, and construct a counterpart of the channel $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ as follows: Let $(\bar{\mathcal{X}}_{\mathcal{I}}, \bar{\mathcal{Y}}_{\mathcal{I}}, \alpha, \bar{\boldsymbol{\mathcal{S}}}, \bar{\boldsymbol{\mathcal{G}}}, \bar{\boldsymbol{q}})$ be a noiseless DM-MMN consisting of independent DMCs with multicast demand $(\mathcal{V}, \mathcal{D})$ such that for each $(i,j) \in \mathcal{I} \times \mathcal{I}$, the DMC carrying information from node i to node j is an error-free (noiseless) channel, denoted by $\bar{q}_{\bar{X}_{i,j}|\bar{X}_{i,j}}$, with capacity $C_{i,j}$ (cf. (35)). To be more precise, $\bar{q}_{\bar{X}_{i,j}|\bar{X}_{i,j}}$ can carry $\lfloor nC_{i,j} \rfloor$ error-free bits for each $(i,j) \in \mathcal{I} \times \mathcal{I}$ for n uses of $(\bar{\mathcal{X}}_{\mathcal{I}}, \bar{\mathcal{X}}_{\mathcal{I}}, \bar{q}_{\bar{X}_{\mathcal{I}}|\bar{X}_{\mathcal{I}}})$. Let $\bar{\mathcal{C}}_+$ denote the positive-delay region of $(\bar{\mathcal{X}}_{\mathcal{I}}, \bar{\mathcal{Y}}_{\mathcal{I}}, \alpha, \bar{\boldsymbol{\mathcal{S}}}, \bar{\boldsymbol{\mathcal{G}}}, \bar{\boldsymbol{q}})$. Since the original as well as the counterpart DM-MMNs consist of independent DMCs, it follows from the network equivalence theory [8] that $\mathcal{C}_+ = \bar{\mathcal{C}}_+$. In addition, it has been shown in [10, Section II-A] that $\bar{\mathcal{C}}_+ = \mathcal{R}_{\text{in}}^{\text{DMCs}}$. Consequently, $\mathcal{C}_+ = \bar{\mathcal{C}}_+ = \mathcal{R}_{\text{in}}^{\text{DMCs}}$ and (41) holds.

3) Converse of Theorem 5: In this subsection, we would like to prove (42). Define

$$\mathcal{R}_{\text{out}}^{\text{DMCs}} \triangleq$$

$$\bigcup_{\substack{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}:p_{X_{\mathcal{I}},Y_{\mathcal{I}}}= \\ \prod_{h=1}^{\alpha}(p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}},Y_{\mathcal{G}^{h-1}}} q_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h},Y_{\mathcal{G}^{h-1}}}^{(h)})}} \quad \bigcap_{T \subseteq \mathcal{I}:T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{c} \sum_{i \in T} R_i \leq \sum_{h=1}^{\alpha} I_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; \\ Y_{T^c \cap \mathcal{G}_h}|X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}), \\ R_i = 0 \text{ for all } i \in \mathcal{V}^c \end{array} \right. \right\}. \tag{43}$$

It follows form Theorem 2 that $\mathcal{C} \subseteq \mathcal{R}_{\text{out}}^{\text{DMCs}}$. Therefore, it remains to show that

$$\mathcal{R}_{\text{out}}^{\text{DMCs}} \subseteq \mathcal{R}_{\text{in}}^{\text{DMCs}}. \tag{44}$$

For any $p_{X_{\mathcal{I}},Y_{\mathcal{I}}} = \prod_{h=1}^{\alpha}(p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}},Y_{\mathcal{G}^{h-1}}} q_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h},Y_{\mathcal{G}^{h-1}}}^{(h)})$, it follows from (39) that

$$p_{X_{\mathcal{I}},Y_{\mathcal{I}}} = \prod_{h=1}^{\alpha} \left(p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}},Y_{\mathcal{G}^{h-1}}} \prod_{(i,j) \in \mathcal{S}^h \times \mathcal{G}_h} q_{Y_{i,j}|X_{i,j}} \right). \tag{45}$$

Marginalizing (45), we have

$$p_{X_{\mathcal{S}^m},Y_{\mathcal{G}^m}} = \prod_{h=1}^{m}\left(p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}},Y_{\mathcal{G}^{h-1}}}\prod_{(i,j)\in\mathcal{S}^h\times\mathcal{G}_h}q_{Y_{i,j}|X_{i,j}}\right) \tag{46}$$

for each $m \in \{1,2,\ldots,\alpha\}$. Relabeling symbols in (46) and using (37) and (38), we have

$$p_{X_{\mathcal{S}^h\times\mathcal{I}},Y_{\mathcal{I}\times\mathcal{G}^h}} = \prod_{\ell=1}^{h}\left(p_{X_{\mathcal{S}_\ell\times\mathcal{I}}|X_{\mathcal{S}^{\ell-1}\times\mathcal{I}},Y_{\mathcal{I}\times\mathcal{G}^{\ell-1}}}\prod_{(i,j)\in\mathcal{S}^\ell\times\mathcal{G}_\ell}q_{Y_{i,j}|X_{i,j}}\right)$$

for each $h \in \{1,2,\ldots,\alpha\}$, which implies from Proposition 1 that for each $h \in \{1,2,\ldots,\alpha\}$ and each $(i,j) \in \mathcal{S}^h \times \mathcal{G}_h$ (cf. (34)),

$$(\{(X_{k,\ell},Y_{k,\ell}) : (k,\ell)\in\mathcal{S}^h\times\mathcal{G}^h,(k,\ell)\neq(i,j)\}\to X_{i,j}\to Y_{i,j})_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}} \tag{47}$$

forms a Markov chain. Following (43), we consider the following chain of inequalities for a fixed $p_{X_{\mathcal{I}},Y_{\mathcal{I}}} = \prod_{h=1}^{\alpha}(p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}},Y_{\mathcal{G}^{h-1}}}q_{Y_{\mathcal{G}_h}|X_{\mathcal{S}}^h,Y_{\mathcal{G}}^{h-1}}^{(h)})$ and a fixed $T\subseteq\mathcal{I}$:

$$\sum_{h=1}^{\alpha}I_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h},Y_{T\cap\mathcal{G}^{h-1}};Y_{T^c\cap\mathcal{G}_h}|X_{T^c\cap\mathcal{S}^h},Y_{T^c\cap\mathcal{G}^{h-1}})$$

$$=\sum_{h=1}^{\alpha}\sum_{j\in T^c\cap\mathcal{G}_h}I_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h},Y_{T\cap\mathcal{G}^{h-1}};Y_j|X_{T^c\cap\mathcal{S}^h},Y_{T^c\cap\mathcal{G}^{h-1}},\{Y_\ell\}_{\ell\in T^c\cap\mathcal{G}_h,\ell<j})$$

$$\overset{(38)}{=}\sum_{h=1}^{\alpha}\sum_{i\in\mathcal{I},j\in T^c\cap\mathcal{G}_h}I_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h},Y_{T\cap\mathcal{G}^{h-1}};Y_{i,j}|X_{T^c\cap\mathcal{S}^h},Y_{T^c\cap\mathcal{G}^{h-1}},\{Y_\ell\}_{\ell\in T^c\cap\mathcal{G}_h,\ell<j},\{Y_{m,j}\}_{m\in\mathcal{I},m<i})$$

$$\overset{(a)}{=}\sum_{h=1}^{\alpha}\sum_{i\in\mathcal{S}^h,j\in T^c\cap\mathcal{G}_h}I_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h},Y_{T\cap\mathcal{G}^{h-1}};Y_{i,j}|X_{T^c\cap\mathcal{S}^h},Y_{T^c\cap\mathcal{G}^{h-1}},\{Y_\ell\}_{\ell\in T^c\cap\mathcal{G}_h,\ell<j},\{Y_{m,j}\}_{m\in\mathcal{I},m<i})$$

$$\overset{(b)}{=}\sum_{h=1}^{\alpha}\sum_{i\in T\cap\mathcal{S}^h,j\in T^c\cap\mathcal{G}_h}I_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h},Y_{T\cap\mathcal{G}^{h-1}};Y_{i,j}|X_{T^c\cap\mathcal{S}^h},Y_{T^c\cap\mathcal{G}^{h-1}},\{Y_\ell\}_{\ell\in T^c\cap\mathcal{G}_h,\ell<j},\{Y_{m,j}\}_{m\in\mathcal{I},m<i})$$

$$\overset{(c)}{\leq}\sum_{h=1}^{\alpha}\sum_{i\in T\cap\mathcal{S}^h,j\in T^c\cap\mathcal{G}_h}I_{p_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{i,j};Y_{i,j})$$

$$\overset{(35)}{\leq}\sum_{h=1}^{\alpha}\sum_{i\in T\cap\mathcal{S}^h,j\in T^c\cap\mathcal{G}_h}C_{i,j}$$

$$\leq\sum_{h=1}^{\alpha}\sum_{i\in T,j\in T^c\cap\mathcal{G}_h}C_{i,j}$$

$$=\sum_{(i,j)\in T\times T^c}C_{i,j}, \tag{48}$$

where

(a) follows from the fact that for each $h \in \{1,2,\ldots,\alpha\}$ and each $(i,j) \in (\mathcal{I}\setminus\mathcal{S}^h)\times\mathcal{G}_h$, (i,j) lies in Ω^c (cf. (34))

and hence

$$I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; Y_{i,j} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_\ell\}_{\ell \in T^c \cap \mathcal{G}_h, \ell < j}, \{Y_{m,j}\}_{m \in \mathcal{I}, m < i})$$

$$\leq H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j})$$

$$\overset{(36)}{=} 0.$$

(b) follows from the fact that for each $h \in \{1, 2, \ldots, \alpha\}$ and each $(i,j) \in (T^c \cap \mathcal{S}^h) \times (T^c \cap \mathcal{G}_h) \subseteq \Omega$,

$$I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; Y_{i,j} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_\ell\}_{\ell \in T^c \cap \mathcal{G}_h, \ell < j}, \{Y_{m,j}\}_{m \in \mathcal{I}, m < i})$$

$$= H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_\ell\}_{\ell \in T^c \cap \mathcal{G}_h, \ell < j}, \{Y_{m,j}\}_{m \in \mathcal{I}, m < i})$$

$$\quad - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j} | X_{\mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_\ell\}_{\ell \in T^c \cap \mathcal{G}_h, \ell < j}, \{Y_{m,j}\}_{m \in \mathcal{I}, m < i}, Y_{T \cap \mathcal{G}^{h-1}})$$

$$\overset{(47)}{=} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j} | X_{i,j}) - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j} | X_{i,j})$$

$$= 0.$$

(c) follows from the fact that for each $h \in \{1, 2, \ldots, \alpha\}$ and each $(i,j) \in (T \cap \mathcal{S}^h) \times (T^c \cap \mathcal{G}_h) \subseteq \Omega$,

$$I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; Y_{i,j} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_\ell\}_{\ell \in T^c \cap \mathcal{G}_h, \ell < j}, \{Y_{m,j}\}_{m \in \mathcal{I}, m < i})$$

$$\leq H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j}) - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j} | X_{\mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_\ell\}_{\ell \in T^c \cap \mathcal{G}_h, \ell < j}, \{Y_{m,j}\}_{m \in \mathcal{I}, m < i}, Y_{T \cap \mathcal{G}^{h-1}})$$

$$\overset{(47)}{=} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j}) - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{i,j} | X_{i,j})$$

$$= I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{i,j}; Y_{i,j}).$$

Consequently, it follows from (40), (43) and (48) that (44) holds.

C. Wireless Erasure Network

1) Problem Formulation and Main Result: Consider a DM-MMN $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ defined as follows: Similar to the MMN consisting of independent DMCs discussed in the previous section, we let

$$\Omega \triangleq \bigcup_{h=1}^{\alpha} \mathcal{S}^h \times \mathcal{G}_h \tag{49}$$

characterize the edge set of the network so that $q^{(h)}_{Y_{\mathcal{G}_h} | X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}}$ can be well-defined for each $h \in \{1, 2, \ldots, \alpha\}$. To simulate the broadcast nature of wireless networks, we assume that in every time slot, each node i broadcasts a symbol X_i for each $i \in \mathcal{I}$ and we let $\mathcal{X}_i$ denote the finite alphabet of X_i. For each $(i,j) \in \Omega$, we assume that node j receives X_i with erasure probability $\varepsilon_{i,j} \in [0, 1]$, and we let $Y_{i,j}$ and $\mathcal{Y}_{i,j} \triangleq \mathcal{X}_i \cup \{\varepsilon\}$ denote the received symbol and its alphabet respectively where ε denotes the erasure symbol. For every edge (i', j') that is not in Ω, we set its erasure probability $\varepsilon_{i',j'}$ to be 1 and

$$\mathcal{Y}_{i',j'} \triangleq \{\varepsilon\}, \tag{50}$$

indicating that no information can be transmitted from i' to j'. We let $q_{Y_{i,j}|X_i}$ characterize the channel corresponding to edge (i,j) such that for each $x_i \in \mathcal{X}_i$ and each $y_{i,j} \in \mathcal{Y}_{i,j}$,

$$q_{Y_{i,j}|X_i}(y_{i,j}|x_i) = \begin{cases} 1 - \varepsilon_{i,j} & \text{if } y_{i,j} = x_i, \\ \varepsilon_{i,j} & \text{if } y_{i,j} = \varepsilon. \end{cases} \tag{51}$$

The symbols transmitted on the edges in Ω are assumed to be erased independently, i.e.,

$$\Pr\left\{ \left. \bigcap_{(i,j)\in\Omega} \{Y_{i,j} = y_{i,j}\} \right| X_\mathcal{I} = x_\mathcal{I} \right\} = \prod_{(i,j)\in\Omega} q_{Y_{i,j}|X_i}(y_{i,j}|x_i) \tag{52}$$

for each $x_\mathcal{I} \in \mathcal{X}_\mathcal{I}$ and each $|\Omega|$-dimensional tuple $(y_{i,j} : (i,j) \in \Omega) \in \prod_{(i,j)\in\Omega} \mathcal{Y}_{i,j}$.

For each $(i,j) \in \mathcal{I} \times \mathcal{I}$, let $E_{i,j}$ and $\mathcal{E}_{i,j}$ be the indicator random variable for the erasure occurred on edge (i,j) and its alphabet respectively such that

$$E_{i,j} \triangleq \mathbf{1}\left(\{Y_{i,j} = \varepsilon\}\right) = \begin{cases} 1 & \text{if } Y_{i,j} = \varepsilon, \\ 0 & \text{if } Y_{i,j} \neq \varepsilon. \end{cases} \tag{53}$$

Let

$$E_{\mathcal{I}\times\mathcal{I}} \triangleq (E_{1,1}, E_{1,2}, \ldots, E_{N,N})$$

be the N^2-dimensional random tuple containing all the $E_{i,j}$'s so that $E_{\mathcal{I}\times\mathcal{I}}$ characterizes the *network erasure pattern*, and let $\mathcal{E}_{\mathcal{I}\times\mathcal{I}}$ denote the alphabet of $E_{\mathcal{I}\times\mathcal{I}}$. Recalling that $(\mathcal{V}, \mathcal{D})$ is the multicast demand, we assume that the following two statements hold:

(i) All the destinations are contained in $\mathcal{G}_\alpha$, i.e., $\mathcal{D} \subseteq \mathcal{G}_\alpha$.

(ii) The network erasure pattern $E_{\mathcal{I}\times\mathcal{I}}$ in each time slot is accessible by each destination node in $\mathcal{D}$.

Since $E_{\mathcal{I}\times\mathcal{I}}$ is a function of $Y_\mathcal{I}$ by (53), there exists some conditional distribution $\chi_{E_{\mathcal{I}\times\mathcal{I}}|Y_\mathcal{I}}$ such that

$$\chi_{E_{\mathcal{I}\times\mathcal{I}}|Y_\mathcal{I}}(e_{\mathcal{I}\times\mathcal{I}}|y_\mathcal{I}) = \begin{cases} 1 & \text{if } e_{i,j} = \mathbf{1}(\{y_{i,j} = \varepsilon\}) \text{ for all } (i,j) \in \mathcal{I} \times \mathcal{I}, \\ 0 & \text{otherwise} \end{cases} \tag{54}$$

for all $y_\mathcal{I}$ and $e_{\mathcal{I}\times\mathcal{I}}$. We are now ready to formally define $\mathcal{X}_\mathcal{I}$, $\mathcal{Y}_\mathcal{I}$ and $\boldsymbol{q}$ as follows: For each $i \in \mathcal{I}$, recalling that $\mathcal{X}_i$ is the finite alphabet of the symbol $X_{i,k}$ broadcast by node i, we define

$$\mathcal{X}_\mathcal{I} \triangleq \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_N. \tag{55}$$

Recalling that $\mathcal{Y}_{i,j} = \mathcal{X}_i \cup \{\varepsilon\}$ is the alphabet of the noisy version of $X_{i,k}$ that is received by node j in each time slot for each $(i,j) \in \Omega$, we define for each $h \in \{1, 2, \ldots, \alpha\}$ and each $m \in \mathcal{G}_h$

$$\mathcal{Y}_m \triangleq \begin{cases} \left(\prod_{i\in\mathcal{I}} \mathcal{Y}_{i,m}\right) \times \mathcal{E}_{\mathcal{I}\times\mathcal{I}} & \text{if } m \text{ is an element in } \mathcal{D} \subseteq \mathcal{G}_\alpha, \\ \prod_{i\in\mathcal{S}^h} \mathcal{Y}_{i,m} & \text{otherwise.} \end{cases} \tag{56}$$

The definition of $\mathcal{Y}_m$ in (56) is divided into two cases because we assume according to Statements (i) and (ii) that the destination nodes in $\mathcal{D} \subseteq \mathcal{G}_\alpha$ have access to the network erasure pattern. After defining $\mathcal{Y}_m$ for each $m \in \mathcal{I}$ in (56), we define

$$\mathcal{Y}_\mathcal{I} \triangleq \mathcal{Y}_1 \times \mathcal{Y}_2 \times \ldots \times \mathcal{Y}_N. \tag{57}$$

Based on the definitions of $\mathcal{X}_\mathcal{I}$ and $\mathcal{Y}_\mathcal{I}$ in (55) and (57) respectively and recalling (51) and (54), we define $q_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{S}^{h-1}}}^{(h)}$ for each $h \in \{1, 2, \ldots, \alpha\}$ as

$$q_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{S}^{h-1}}}^{(h)}(y_{\mathcal{G}_h}|x_{\mathcal{S}^h}, y_{\mathcal{S}^{h-1}})$$
$$\triangleq \begin{cases} \left(\prod_{i \in \mathcal{I}} \prod_{m \in \mathcal{G}_h} q_{Y_{i,m}|X_i}(y_{i,m}|x_i)\right) \chi_{E_{\mathcal{I} \times \mathcal{I}}|Y_\mathcal{I}}(e_{\mathcal{I} \times \mathcal{I}}|y_\mathcal{I}) & \text{if } h = \alpha, \\ \prod_{i \in \mathcal{S}^h} \prod_{m \in \mathcal{G}_h} q_{Y_{i,m}|X_i}(y_{i,m}|x_i) & \text{otherwise} \end{cases} \tag{58}$$

for all $x_{\mathcal{S}^h} \in \mathcal{X}_{\mathcal{S}^h}$, $y_{\mathcal{G}^h} \in \mathcal{Y}_{\mathcal{G}^h}$ and $e_{\mathcal{I} \times \mathcal{I}} \in \mathcal{E}_{\mathcal{I} \times \mathcal{I}}$, where for each $m \in \mathcal{G}_h$

$$y_m = \begin{cases} ((y_{i,m} : i \in \mathcal{I}), e_{\mathcal{I} \times \mathcal{I}}) & \text{if } m \text{ is an element in } \mathcal{D} \subseteq \mathcal{G}_\alpha, \\ (y_{i,m} : i \in \mathcal{S}^h) & \text{otherwise.} \end{cases}$$

We call the network described above the *wireless erasure network*. The random variables $X_\mathcal{I}$ and $Y_\mathcal{I}$ in the wireless erasure network are generated according to this order

$$X_{\mathcal{S}_1}, Y_{\mathcal{G}_1}, X_{\mathcal{S}_2}, Y_{\mathcal{G}_2}, \ldots, X_{\mathcal{S}_\alpha}, Y_{\mathcal{G}_\alpha}$$

(cf. (2)), which implies from (56), (57) and (58) that $X_\mathcal{I}$, $\{Y_{i,j}\}_{(i,j) \in \Omega}$ and $E_{\mathcal{I} \times \mathcal{I}}$ are generated according to this order

$$X_{\mathcal{S}_1}, \{Y_{i,j}\}_{(i,j) \in \mathcal{S}^1 \times \mathcal{G}_1}, X_{\mathcal{S}_2}, \{Y_{i,j}\}_{(i,j) \in \mathcal{S}^2 \times \mathcal{G}_2}, \ldots, X_{\mathcal{S}_\alpha}, \{Y_{i,j}\}_{(i,j) \in \mathcal{S}^\alpha \times \mathcal{G}_\alpha}, E_{\mathcal{I} \times \mathcal{I}}. \tag{59}$$

It may not be obvious from (59) that $X_\mathcal{I}$ and $E_{\mathcal{I} \times \mathcal{I}}$ are always independent, but it follows from (51), (52) and (54) that for any $e_{\mathcal{I} \times \mathcal{I}} \in \mathcal{E}_{\mathcal{I} \times \mathcal{I}}$ and $x_\mathcal{I} \in \mathcal{X}_\mathcal{I}$,

$$\Pr\{E_{\mathcal{I} \times \mathcal{I}} = e_{\mathcal{I} \times \mathcal{I}}|X_\mathcal{I} = x_\mathcal{I}\} = \prod_{(i,j) \in \Omega} \left(\varepsilon_{i,j}^{\mathbf{1}\{e_{i,j}=1\}}(1 - \varepsilon_{i,j})^{\mathbf{1}\{e_{i,j}=0\}}\right), \tag{60}$$

which implies the independence between $X_\mathcal{I}$ and $E_{\mathcal{I} \times \mathcal{I}}$, i.e.,

$$\Pr\{E_{\mathcal{I} \times \mathcal{I}} = e_{\mathcal{I} \times \mathcal{I}}|X_\mathcal{I} = x_\mathcal{I}\} = \Pr\{E_{\mathcal{I} \times \mathcal{I}} = e_{\mathcal{I} \times \mathcal{I}}\} \tag{61}$$

for any $e_{\mathcal{I} \times \mathcal{I}} \in \mathcal{E}_{\mathcal{I} \times \mathcal{I}}$ and $x_\mathcal{I} \in \mathcal{X}_\mathcal{I}$. The classical wireless erasure network studied in [9] is a special case of our model when $\alpha = 1$ and $\Omega = \mathcal{I} \times \mathcal{I}$. The following theorem is the main result in this section, and the proof will be provided in the next two subsections.
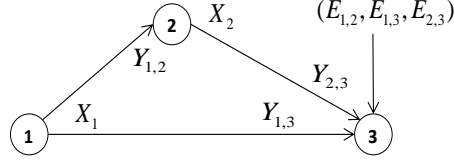
Fig. 4. A three-node wireless erasure network.

Theorem 6: For the wireless erasure network, let

$$\mathcal{R}_{\text{in}}^{\text{WEN}} \triangleq \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq \sum_{i \in T} \left(1 - \prod_{j \in T^c} e_{i,j} \right) |\mathcal{X}_i|, \\ R_i = 0 \text{ for all } i \in \mathcal{V}^c \end{array} \right. \right\}. \tag{62}$$

Then,

$$\mathcal{C} = \mathcal{C}_+ = \mathcal{R}_{\text{in}}^{\text{WEN}}$$

and hence the network is delay-independent.

Remark 3: For the wireless erasure network with zero-delay nodes, due to the independence nature among the erasures, the network can be intuitively viewed as a MMN consisting of independent erasure channels, whose capacity region is contained in the classical cut-set bound by Theorem 5. On the other hand, it has been shown in [9] that the cut-set bound can be achieved for the wireless erasure network. Combining the intuition and the fact provided above, it is intuitive that Theorem 6 should hold.

Example 3: Consider a relay channel that consists of three nodes where node 1 wants to transmit information to node 3 via a relay node 2. In each time slot, node i transmits X_i for each $i \in \{1, 2, 3\}$, while node 2 receives an erased version of X_1 denoted by $Y_{1,2}$ and node 3 receives erased versions of X_1 and X_2 denoted by $Y_{1,3}$ and $Y_{2,3}$ respectively. Let $E_{i,j}$ denote the erasure random variable for (i, j) where

$$E_{i,j} = \begin{cases} 0 & \text{if } X_i \text{ is not erased at node } j, \text{ i.e., } Y_{i,j} = X_i, \\ 1 & \text{otherwise, i.e., } X_i \text{ is erased at node } j. \end{cases} \tag{63}$$

The erasures are assumed to be independent, i.e., $p_{E_{1,2},E_{1,3},E_{2,3}} = p_{E_{1,2}} p_{E_{1,3}} p_{E_{2,3}}$ regardless of the distribution of (X_1, X_2). In addition, node 3 is assumed to have access of the network erasure pattern $(E_{1,2}, E_{1,3}, E_{2,3})$ (note that $E_{1,3}$ and $E_{2,3}$ can be deduced from $Y_{1,3}$ and $Y_{2,3}$ respectively by (63), but $E_{1,2}$ is an extra information provided for node 3 for decoding). This relay channel is illustrated in Figure 4. The relay channel can be formulated as a wireless erasure network by setting $\boldsymbol{\mathcal{S}} \triangleq (\{1\}, \{2,3\})$, $\boldsymbol{\mathcal{G}} \triangleq (\{2\}, \{1,3\})$, $\Omega \triangleq (\mathcal{S}_1 \times \mathcal{G}_1) \cup (\mathcal{S}^2 \times \mathcal{G}_2)$, $Y_2 \triangleq Y_{1,2}$ and $Y_3 \triangleq (Y_{1,3}, Y_{2,3}, E_{1,2}, E_{1,3}, E_{2,3})$. The set of non-trivial edges $\{(1,2), (1,3), (2,3)\}$ is contained in Ω by (49). Since node 2 incurs no delay under this formulation, we cannot characterize the capacity region by applying the classical cut-set bound. Surprisingly, Theorem 6 implies that this three-node wireless erasure network with a zero-delay node is delay-independent. ∎

In the following, we provide the proof of Theorem 6. To this end, it suffices to prove the achievability statement

$$\mathcal{R}_{\text{in}}^{\text{WEN}} \subseteq \mathcal{C}_+ \tag{64}$$

and the converse statement

$$\mathcal{C} \subseteq \mathcal{R}_{\text{in}}^{\text{WEN}}. \tag{65}$$

2) Achievability: In this subsection, we would like to prove (64). Since the achievability statement (64) has been shown in [9] under the classical model which considers no zero-delay nodes, (64) holds naturally under our generalized-delay model. For completeness, the proof of (64) under our generalized-delay model is provided in Appendix B.

3) Converse: In this subsection, we would like to prove (65). We will first prove the following counterpart of Theorem 2 to show an outer bound on $\mathcal{C}$, and then show that the outer bound is contained in $\mathcal{R}_{\text{in}}^{\text{WEN}}$.

Lemma 5: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$ be a wireless erasure network, and let

$$\mathcal{R}_{\text{out}}^{\text{WEN}} \triangleq$$

$$\bigcup_{\substack{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} : p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} = \\ \prod_{h=1}^{\alpha} (p_{X_{\mathcal{S}_h} | X_{\mathcal{S}^{h-1}}, Y_{\mathcal{G}^{h-1}}} q_{Y_{\mathcal{G}_h} | X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}}^{(h)})}} \bigcap_{T \subseteq \mathcal{I} : T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq \sum_{h=1}^{\alpha} I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; \\ \qquad\qquad Y_{T^c \cap \mathcal{G}_h} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, E_{\mathcal{I} \times \mathcal{I}}), \\ R_i = 0 \text{ for all } i \in \mathcal{V}^c \end{array} \right. \right\} \tag{66}$$

where $E_{\mathcal{I} \times \mathcal{I}}$, the network erasure pattern, is a function of $Y_{\mathcal{I}}$ defined by (53). Then,

$$\mathcal{C} \subseteq \mathcal{R}_{\text{out}}^{\text{WEN}}.$$

Proof: Let $R_{\mathcal{I}}$ be an achievable rate tuple for the wireless erasure network denoted by $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, \alpha, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{G}}, \boldsymbol{q})$. Then, there exists a sequence of $(B, n, M_{\mathcal{I}})$-codes on the network such that

$$\lim_{n \to \infty} \frac{\log M_i}{n} \geq R_i \tag{67}$$

and

$$\lim_{n \to \infty} P_{\text{err}}^n = 0 \tag{68}$$

for each $i \in \mathcal{I}$. Fix any $T \subseteq \mathcal{I}$ such that $T^c \cap \mathcal{D} \neq \emptyset$, and let d denote a node in $T^c \cap \mathcal{D}$. Fix a $(B, n, M_{\mathcal{I}})$-code and let $E_{\mathcal{I} \times \mathcal{I}, k}$ denote the network erasure pattern occurred in time slot k for each $k \in \{1, 2, \ldots, n\}$. Then, we consider the following chain of inequalities:

$$\sum_{i \in T} \log M_i \overset{\text{(a)}}{=} H(W_T | W_{T^c})$$

$$\overset{\text{(b)}}{=} H(W_T | W_{T^c}, E_{\mathcal{I} \times \mathcal{I}}^n)$$

$$= I(W_T; Y^n_{T^c}|W_{T^c}, E^n_{\mathcal{I}\times\mathcal{I}}) + H(W_T|Y^n_{T^c}, W_{T^c}, E^n_{\mathcal{I}\times\mathcal{I}})$$

$$\leq I(W_T; Y^n_{T^c}|W_{T^c}, E^n_{\mathcal{I}\times\mathcal{I}}) + H(W_T|Y^n_d, W_d, E^n_{\mathcal{I}\times\mathcal{I}})$$

$$\overset{(c)}{\leq} I(W_T; Y^n_{T^c}|W_{T^c}, E^n_{\mathcal{I}\times\mathcal{I}}) + 1 + P^n_{\text{err}}\sum_{i\in T}\log M_i, \tag{69}$$

where

(a) follows from the fact that the N messages $W_1, W_2, \ldots, W_N$ are independent.

(b) follows from the fact that $W_{\mathcal{I}}$ and $E^n_{\mathcal{I}\times\mathcal{I}}$ are independent.

(c) follows from Fano's inequality.

Following similar procedures for proving Theorem 1 in [1], we can show by using (67), (68) and (69) that there exists a joint distribution $p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}$ which depends on the sequence of $(B, n, M_{\mathcal{I}})$-codes but not on T such that

$$p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} = \prod_{h=1}^{\alpha}(p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}}, Y_{\mathcal{G}^{h-1}}} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}})$$

and

$$\sum_{i\in T} R_i \leq \sum_{h=1}^{\alpha} I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h}, Y_{T\cap\mathcal{G}^{h-1}}; Y_{T^c\cap\mathcal{G}_h}|X_{T^c\cap\mathcal{S}^h}, Y_{T^c\cap\mathcal{G}^{h-1}}, E_{\mathcal{I}\times\mathcal{I}}). \tag{70}$$

Since $p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}$ depends on only the sequence of $(B, n, M_{\mathcal{I}})$-codes but not on T, (70) holds for all $T \subseteq \mathcal{I}$ such that $T^c \cap \mathcal{D} \neq \emptyset$. This completes the proof. ∎

Since

$$\mathcal{C} \subseteq \mathcal{R}^{\text{WEN}}_{\text{out}} \tag{71}$$

by Lemma 5 and our goal is to prove $\mathcal{C} \subseteq \mathcal{R}^{\text{WEN}}_{\text{in}}$, it remains to show that

$$\mathcal{R}^{\text{WEN}}_{\text{out}} \subseteq \mathcal{R}^{\text{WEN}}_{\text{in}}. \tag{72}$$

For any $p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} = \prod_{h=1}^{\alpha}(p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}}, Y_{\mathcal{G}^{h-1}}} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}})$, it follows from (51) and (53) that for each $h \in \{1, 2, \ldots, \alpha\}$ and each $(i, j) \in \mathcal{S}^h \times \mathcal{G}_h$, $Y_{i,j}$ is a function of $(X_i, E_{\mathcal{I}\times\mathcal{I}})$ and hence

$$\left(\{(X_k, Y_{k,\ell}) : (k, \ell) \in (\mathcal{S}^h \setminus \{i\}) \times \mathcal{G}^h\} \to (X_i, E_{\mathcal{I}\times\mathcal{I}}) \to Y_{i,j}\right)_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}} \tag{73}$$

forms a Markov chain. Following (72) and (66), we fix $p_{X_{\mathcal{I}}, Y_{\mathcal{I}}} = \prod_{h=1}^{\alpha}(p_{X_{\mathcal{S}_h}|X_{\mathcal{S}^{h-1}}, Y_{\mathcal{G}^{h-1}}} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}}^h, Y_{\mathcal{G}}^{h-1}})$ and $T \subseteq \mathcal{I}$ such that $T^c \cap \mathcal{D} \neq \emptyset$, and we consider

$$\sum_{h=1}^{\alpha} I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h}, Y_{T\cap\mathcal{G}^{h-1}}; Y_{T^c\cap\mathcal{G}_h}|X_{T^c\cap\mathcal{S}^h}, Y_{T^c\cap\mathcal{G}^{h-1}}, E_{\mathcal{I}\times\mathcal{I}})$$

$$\overset{(56)}{=} \sum_{h=1}^{\alpha}\sum_{i\in\mathcal{S}^h} I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h}, Y_{T\cap\mathcal{G}^{h-1}}; Y_{\{i\}\times(T^c\cap\mathcal{G}_h)}|X_{T^c\cap\mathcal{S}^h}, Y_{T^c\cap\mathcal{G}^{h-1}}, \{Y_{\{\ell\}\times(T^c\cap\mathcal{G}_h)}\}_{\ell<i}, E_{\mathcal{I}\times\mathcal{I}})$$

$$\overset{(a)}{=} \sum_{h=1}^{\alpha}\sum_{i\in T\cap\mathcal{S}^h} I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T\cap\mathcal{S}^h}, Y_{T\cap\mathcal{G}^{h-1}}; Y_{\{i\}\times(T^c\cap\mathcal{G}_h)}|X_{T^c\cap\mathcal{S}^h}, Y_{T^c\cap\mathcal{G}^{h-1}}, \{Y_{\{\ell\}\times(T^c\cap\mathcal{G}_h)}\}_{\ell<i}, E_{\mathcal{I}\times\mathcal{I}})$$

$$\overset{(b)}{\leq} \sum_{h=1}^{\alpha} \sum_{i \in T \cap \mathcal{S}^h} I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_i; Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | Y_{\{i\} \times (T^c \cap \mathcal{G}^{h-1})}, E_{\mathcal{I} \times \mathcal{I}})$$

$$\overset{(c)}{=} \sum_{h=1}^{\alpha} \sum_{i \in T \cap \mathcal{S}^h} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | Y_{\{i\} \times (T^c \cap \mathcal{G}^{h-1})}, E_{\mathcal{I} \times \mathcal{I}})$$

$$= \sum_{i \in T \cap \mathcal{S}^h} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\mathcal{I} \times \mathcal{I}})$$

$$\leq \sum_{i \in T \cap \mathcal{S}^h} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c}) \tag{74}$$

where

(a) follows from the fact that for each $h \in \{1, 2, \ldots, \alpha\}$ and each $i \in T^c \cap \mathcal{S}^h$,

$$I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_{\{\ell\} \times T^c \cap \mathcal{G}_h}\}_{\ell < i}, E_{\mathcal{I} \times \mathcal{I}})$$

$$= H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_{\{\ell\} \times T^c \cap \mathcal{G}_h}\}_{\ell < i}, E_{\mathcal{I} \times \mathcal{I}})$$

$$\quad - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}, \{Y_{\{\ell\} \times T^c \cap \mathcal{G}_h}\}_{\ell < i}, E_{\mathcal{I} \times \mathcal{I}})$$

$$\overset{(73)}{=} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | X_i, E_{\mathcal{I} \times \mathcal{I}}) - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | X_i, E_{\mathcal{I} \times \mathcal{I}})$$

$$= 0.$$

(b) follows from the fact that for each $h \in \{1, 2, \ldots, \alpha\}$ and each $i \in T \cap \mathcal{S}^h$,

$$I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, \{Y_{\{\ell\} \times T^c \cap \mathcal{G}_h}\}_{\ell < i}, E_{\mathcal{I} \times \mathcal{I}})$$

$$\leq H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | Y_{\{i\} \times T^c \cap \mathcal{G}^{h-1}}, E_{\mathcal{I} \times \mathcal{I}})$$

$$\quad - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | X_{\mathcal{S}^h}, Y_{\mathcal{G}^{h-1}}, \{Y_{\{\ell\} \times T^c \cap \mathcal{G}_h}\}_{\ell < i}, E_{\mathcal{I} \times \mathcal{I}})$$

$$\overset{(73)}{=} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | Y_{\{i\} \times T^c \cap \mathcal{G}^{h-1}}, E_{\mathcal{I} \times \mathcal{I}}) - H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | X_i, E_{\mathcal{I} \times \mathcal{I}})$$

$$\leq I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_i; Y_{\{i\} \times (T^c \cap \mathcal{G}_h)} | Y_{\{i\} \times T^c \cap \mathcal{G}^{h-1}}, E_{\mathcal{I} \times \mathcal{I}}).$$

(c) follows from (51) and (53) that $Y_{\{i\} \times (T^c \cap \mathcal{G}_h)}$ is a function of $(X_i, E_{\mathcal{I} \times \mathcal{I}})$.

Following (74) and letting $\mathbf{1}^{T^c}$ denote the $|T^c|$-dimensional all-1 tuple, we consider the following chain of inequalities for each $h \in \{1, 2, \ldots, \alpha\}$ and each $i \in T \cap \mathcal{S}^h$:

$$H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c})$$

$$= \Pr\{E_{\{i\} \times T^c} = \mathbf{1}^{T^c}\} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c} = \mathbf{1}^{T^c})$$

$$\quad + \Pr\{E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c}\} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c})$$

$$\overset{(a)}{=} \Pr\{E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c}\} H_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_i | E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c})$$

$$\leq \Pr\{E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c}\} |\mathcal{X}_i|$$

$$\overset{\text{(b)}}{=} \left(1 - \prod_{j \in T^c} e_{i,j}\right) |\mathcal{X}_i| \tag{75}$$

where

(a) follows from (53) that for each $j \in \mathcal{I}$,

$$Y_{i,j} = \begin{cases} \varepsilon & \text{if } E_{i,j} = 1, \\ X_i & \text{otherwise.} \end{cases}$$

(b) follows from (60) and (61).

Combining (74) and (75), we have

$$\sum_{h=1}^{\alpha} I_{p_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_{T \cap \mathcal{S}^h}, Y_{T \cap \mathcal{G}^{h-1}}; Y_{T^c} | X_{T^c \cap \mathcal{S}^h}, Y_{T^c \cap \mathcal{G}^{h-1}}, E_{\mathcal{I} \times \mathcal{I}})$$

$$\leq \sum_{i \in T \cap \mathcal{S}^h} \left(1 - \prod_{j \in T^c} e_{i,j}\right) |\mathcal{X}_i|$$

$$\leq \sum_{i \in T} \left(1 - \prod_{j \in T^c} e_{i,j}\right) |\mathcal{X}_i|. \tag{76}$$

Consequently, it follows from (62), (66) and (76) that (72) holds, which implies from (71) that $\mathcal{C} \subseteq \mathcal{R}_{\text{in}}^{\text{WEN}}$.

VI. Concluding Remarks

We have investigated under the generalized-delay model three classes of delay-independent multimessage multicast networks (MMNs), namely the deterministic MMN dominated by product distributions, the MMN consisting of independent DMCs and the wireless erasure network respectively. We are able to evaluate the capacity regions for the above classes of MMNs with zero-delay nodes and demonstrate that their capacity regions coincide with the positive-delay regions, which implies that the above classes of MMNs with zero-delay nodes belong to the category of delay-independent MMNs. In other words, for each MMN with zero-delay nodes which belongs to one of the above three classes, the set of achievable rate tuples does not depend on the delay amounts incurred by the nodes in the network. This is in contrast to the fact that for some MMNs with zero-delay nodes, the set of achievable rate tuples shrinks if we impose the additional constraint that each node incurs a positive delay. An important implication of our result is that for each MMN belonging to one of the above three classes, using different methods for handling delay and synchronization does not affect the network capacity.

Future research may continue the theme of this work – to identify other important classes of delay-independent and delay-dependent MMNs under the generalized-delay model. This work is limited to identifying delay-independent MMNs whose capacity regions lie in the corresponding cut-set bounds and at the same time the cut-set bounds can be achieved. The search of delay-independent and delay-dependent MMNs whose capacity regions are strictly smaller than the classical cut-set bounds is an interesting research direction. Another direction is exploring delay-dependent MMNs whose capacity regions are strictly larger than the classical cut-set bounds.

Appendix A

Proof of Lemma 2

Fix a $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code, and let $p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n}$ be the joint distribution induced by the code according to Definitions 5 and 6. For each $k \in \{1, 2, \ldots, n\}$, let $U^{k-1} \triangleq (W_{\mathcal{I}}, X_{\mathcal{I}}^{k-1}, Y_{\mathcal{I}}^{k-1})$ be the collection of random variables that are generated before the kth time slot for the $(B, n, M_{\mathcal{I} \times \mathcal{I}})$-code. In order to prove (20), it suffices to show that

$$H_{p_{U^{k-1}, X_{\mathcal{S}^h,k}, Y_{\mathcal{G}^h,k}}} (X_{\mathcal{S}^h,k}, Y_{\mathcal{G}^h,k} | U^{k-1}) = 0 \tag{77}$$

holds for each $k \in \{1, 2, \ldots, n\}$ and each $h \in \{1, 2, \ldots, \alpha\}$, which will then imply that

$$H_{p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n}} (X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n | W_{\mathcal{I}}) = \sum_{k=1}^{n} H_{p_{U^{k-1}, X_{\mathcal{I},k}, Y_{\mathcal{I},k}}} (X_{\mathcal{I},k}, Y_{\mathcal{I},k} | U^{k-1})$$
$$\overset{(77)}{=} 0.$$

Fix a $k \in \{1, 2, \ldots, n\}$. We prove (77) by induction on h as follows. For $h = 1$, the LHS of (77) is

$$H_{p_{U^{k-1}, X_{\mathcal{S}^1,k}, Y_{\mathcal{G}^1,k}}} (X_{\mathcal{S}^1,k}, Y_{\mathcal{G}^1,k} | U^{k-1})$$
$$= H_{p_{U^{k-1}, X_{\mathcal{S}^1,k}}} (X_{\mathcal{S}^1,k} | U^{k-1}) + H_{p_{U^{k-1}, X_{\mathcal{S}^1,k}, Y_{\mathcal{G}^1,k}}} (Y_{\mathcal{G}^1,k} | U^{k-1}, X_{\mathcal{S}^1,k})$$
$$\overset{(a)}{=} H_{p_{U^{k-1}, X_{\mathcal{S}^1,k}, Y_{\mathcal{G}^1,k}}} (Y_{\mathcal{G}^1,k} | U^{k-1}, X_{\mathcal{S}^1,k})$$
$$\leq H_{p_{X_{\mathcal{S}^1,k}, Y_{\mathcal{G}^1,k}}} (Y_{\mathcal{G}^1,k} | X_{\mathcal{S}^1,k})$$
$$\overset{(1)}{=} H_{p_{X_{\mathcal{S}^1,k}} q^{(1)}_{Y_{\mathcal{G}^1,k} | X_{\mathcal{S}^1,k}}} (Y_{\mathcal{G}^1,k} | X_{\mathcal{S}^1,k})$$
$$\overset{(b)}{=} 0, \tag{78}$$

where

(a) follows from Definitions 4 and 5 that $X_{\mathcal{S}^1,k}$ is a function of U^{k-1} for the code.

(b) follows from the fact that $q^{(1)}$ is deterministic (cf. Definition 11).

If (77) holds for $h = m$, i.e.,

$$H_{p_{U^{k-1}, X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k}}} (X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k} | U^{k-1}) = 0 \tag{79}$$

then for $h = m + 1$ such that $m + 1 \leq \alpha$, the LHS of (77) is

$$H_{p_{U^{k-1}, X_{\mathcal{S}^{m+1},k}, Y_{\mathcal{G}^{m+1},k}}} (X_{\mathcal{S}^{m+1},k}, Y_{\mathcal{G}^{m+1},k} | U^{k-1})$$
$$= H_{p_{U^{k-1}, X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k}}} (X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k} | U^{k-1}) + H_{p_{U^{k-1}, X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k}, X_{\mathcal{S}_{m+1},k}}} (X_{\mathcal{S}_{m+1},k} | U^{k-1}, X_{\mathcal{S}^m,k}, Y_{\mathcal{G}^m,k})$$
$$\quad + H_{p_{U^{k-1}, X_{\mathcal{S}^{m+1},k}, Y_{\mathcal{G}^{m+1},k}}} (Y_{\mathcal{G}_{m+1},k} | U^{k-1}, X_{\mathcal{S}^{m+1},k}, Y_{\mathcal{G}^m,k})$$

$$\stackrel{(79)}{=} H_{p_{U^{k-1},X_{\mathcal{S}^m,k},Y_{\mathcal{G}^m,k},X_{\mathcal{S}_{m+1},k}}}\left(X_{\mathcal{S}_{m+1},k}|U^{k-1},X_{\mathcal{S}^m,k},Y_{\mathcal{G}^m,k}\right)$$

$$+ H_{p_{U^{k-1},X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^{m+1},k}}}\left(Y_{\mathcal{G}_{m+1},k}|U^{k-1},X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^m,k}\right)$$

$$\stackrel{(a)}{=} H_{p_{U^{k-1},X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^{m+1},k}}}\left(Y_{\mathcal{G}_{m+1},k}|U^{k-1},X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^m,k}\right)$$

$$\leq H_{p_{X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^{m+1},k}}}\left(Y_{\mathcal{G}_{m+1},k}|X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^m,k}\right)$$

$$\stackrel{(1)}{=} H_{p_{X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^m,k}}q^{(m+1)}_{Y_{\mathcal{G}_{m+1},k}|X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^m,k}}}\left(Y_{\mathcal{G}_{m+1},k}|X_{\mathcal{S}^{m+1},k},Y_{\mathcal{G}^m,k}\right)$$

$$\stackrel{(b)}{=} 0, \tag{80}$$

where

(a) follows from Definitions 4 and 5 that $X_{\mathcal{S}_{m+1},k}$ is a function of $(U^{k-1},Y_{\mathcal{G}^m,k})$ for the code.

(b) follows from the fact that $q^{(m+1)}$ is deterministic (cf. Definition 11).

For $h=1$, it follows from (78) that (77) holds. For all $1 \leq m \leq \alpha - 1$, it follows from (79) and (80) that if (77) is assumed to be true for $h = m$, then (77) is also true for $h = m+1$. Consequently, it follows by mathematical induction that (77) holds for all $1 \leq h \leq \alpha$.

APPENDIX B
PROOF OF THE ACHIEVABILITY OF THEOREM 6

Our goal is to prove (64). Let u_{X_i} be the uniform distribution on $\mathcal{X}_i$ for each $i \in \mathcal{I}$ and let

$$u_{X_{\mathcal{I}},Y_{\mathcal{I}}} = \left(\prod_{i=1}^{N} u_{X_i}\right)\left(\prod_{h=1}^{\alpha} q^{(h)}_{Y_{\mathcal{G}_h}|X_{\mathcal{S}}^h,Y_{\mathcal{G}}^{h-1}}\right). \tag{81}$$

Fix any $T \subseteq \mathcal{I}$ such that

$$T^c \cap \mathcal{D} \neq \emptyset. \tag{82}$$

In order to apply Theorem 1, we consider

$$H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(Y_T|X_{\mathcal{I}},Y_{T^c})$$
$$\stackrel{(a)}{=} H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(Y_T|X_{\mathcal{I}},Y_{T^c},E_{\mathcal{I}\times\mathcal{I}})$$
$$\stackrel{(b)}{=} 0 \tag{83}$$

and

$$I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_T;Y_{T^c}|X_{T^c})$$
$$\stackrel{(c)}{=} I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_T;Y_{T^c},E_{\mathcal{I}\times\mathcal{I}}|X_{T^c})$$
$$\stackrel{(d)}{=} I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_T;Y_{T^c}|X_{T^c},E_{\mathcal{I}\times\mathcal{I}})$$
$$= H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(Y_{T^c}|X_{T^c},E_{\mathcal{I}\times\mathcal{I}}) - H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(Y_{T^c}|X_{\mathcal{I}},E_{\mathcal{I}\times\mathcal{I}})$$

$$\stackrel{(83)}{=} H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(Y_{T^c}|X_{T^c},E_{\mathcal{I}\times\mathcal{I}})$$

$$\stackrel{(56)}{=} H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(\{Y_{i,j}\}_{(i,j)\in\mathcal{I}\times T^c}|X_{T^c},E_{\mathcal{I}\times\mathcal{I}})$$

$$\stackrel{(e)}{=} H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(\{Y_{i,j}\}_{(i,j)\in T\times T^c}|X_{T^c},E_{\mathcal{I}\times\mathcal{I}}) \tag{84}$$

where

(a) follows from Statements (i) and (ii) in the previous subsection and (82) that Y_{T^c} contains the random variable $E_{\mathcal{I}\times\mathcal{I}}$.

(b) follows from (51) and (53) that Y_T is a function of $(X_{\mathcal{I}}, E_{\mathcal{I}\times\mathcal{I}})$.

(c) follows from Statements (i) and (ii) in the previous subsection and (82) that Y_{T^c} contains the random variable $E_{\mathcal{I}\times\mathcal{I}}$.

(d) follows from (61) that $X_{\mathcal{I}}$ and $E_{\mathcal{I}\times\mathcal{I}}$ are independent, i.e.,

$$I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{\mathcal{I}};E_{\mathcal{I}\times\mathcal{I}}) = 0. \tag{85}$$

(e) follows from (51) and (53) that $\{Y_{i,j}\}_{(i,j)\in T^c\times T^c}$ is a function of $(X_{T^c}, E_{\mathcal{I}\times\mathcal{I}})$.

In order to further simplify (84), consider the following chain of inequalities for any $T_1, T_2 \subseteq \mathcal{I}$ such that $T_1 \cap T_2 = \emptyset$:

$$I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(\{Y_{i,j}\}_{(i,j)\in T_1\times T_2};X_{T_2}|E_{\mathcal{I}\times\mathcal{I}})$$

$$\leq I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T_1},\{Y_{i,j}\}_{(i,j)\in T_1\times T_2};X_{T_2}|E_{\mathcal{I}\times\mathcal{I}})$$

$$\stackrel{(a)}{=} I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T_1};X_{T_2}|E_{\mathcal{I}\times\mathcal{I}})$$

$$\stackrel{(b)}{=} I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T_1};X_{T_2})$$

$$\stackrel{(c)}{=} 0 \tag{86}$$

where

(a) follows from the fact that $\{Y_{i,j}\}_{(i,j)\in T_1\times T_2}$ is a function of $(X_{T_1}, E_{\mathcal{I}\times\mathcal{I}})$.

(b) follows from (61) that $X_{\mathcal{I}}$ and $E_{\mathcal{I}\times\mathcal{I}}$ are independent.

(c) follows from (81) that X_{T_1} and X_{T_2} are independent.

Following (84), consider the following chain of inequalities:

$$H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(\{Y_{i,j}\}_{(i,j)\in T\times T^c}|X_{T^c},E_{\mathcal{I}\times\mathcal{I}})$$

$$\stackrel{(a)}{=} H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(\{Y_{i,j}\}_{(i,j)\in T\times T^c}|X_{T^c},E_{\mathcal{I}\times\mathcal{I}}) + I_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(X_{T^c};\{Y_{i,j}\}_{(i,j)\in T\times T^c}|E_{\mathcal{I}\times\mathcal{I}})$$

$$= H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(\{Y_{i,j}\}_{(i,j)\in T\times T^c}|E_{\mathcal{I}\times\mathcal{I}})$$

$$= \sum_{i\in T} H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(Y_{\{i\}\times T^c}|E_{\mathcal{I}\times\mathcal{I}},\{Y_{m,\ell}\}_{m\in T,m<i,\ell\in T^c})$$

$$\geq \sum_{i\in T} H_{u_{X_{\mathcal{I}},Y_{\mathcal{I}}}}(Y_{\{i\}\times T^c}|E_{\mathcal{I}\times\mathcal{I}},\{Y_{m,\ell}\}_{m\in T,m<i,\ell\in T^c},\{X_m\}_{m\in T,m<i})$$

$$\overset{(b)}{=} \sum_{i \in T} H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\mathcal{I} \times \mathcal{I}}, \{X_m\}_{m \in T, m < i})$$

$$\overset{(86)}{=} \sum_{i \in T} H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\mathcal{I} \times \mathcal{I}})$$

$$\overset{(c)}{\geq} \sum_{i \in T} (H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c}) - I_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(E_{\{i\} \times T^c}; Y_{\{i\} \times T^c} | X_i, E_{\{i\} \times T^c}))$$

$$\overset{(d)}{=} H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c}) \tag{87}$$

where

(a) follows from (86) by letting $T_1 = T$ and $T_2 = T^c$.

(b) follows from (51) and (53) that $\{Y_{m,\ell}\}_{m \in T, m < i, \ell \in T^c}$ is a function of $(\{X_m\}_{m \in T, m < i}, E_{\mathcal{I} \times \mathcal{I}})$.

(c) follows from the fact that

$$I_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(E_{\{i\} \times T^c}; Y_{\{i\} \times T^c} | E_{\{i\} \times T^c})$$

$$\leq I_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(E_{\{i\} \times T^c}; X_i, Y_{\{i\} \times T^c} | E_{\{i\} \times T^c})$$

$$\overset{(85)}{=} I_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(E_{\{i\} \times T^c}; Y_{\{i\} \times T^c} | X_i, E_{\{i\} \times T^c}).$$

(d) follows from (51) and (53) that $Y_{\{i\} \times T^c}$ is a function of $(X_i, E_{\{i\} \times T^c})$.

Following (87) and letting $\mathbf{1}^{T^c}$ denote the $|T^c|$-dimensional all-1 tuple, we consider the following chain of equalities for each $i \in T$:

$$H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c})$$

$$= \Pr\{E_{\{i\} \times T^c} = \mathbf{1}^{T^c}\} H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c} = \mathbf{1}^{T^c})$$

$$+ \Pr\{E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c}\} H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(Y_{\{i\} \times T^c} | E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c})$$

$$\overset{(a)}{=} \Pr\{E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c}\} H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_i | E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c})$$

$$\overset{(b)}{=} \Pr\{E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c}\} H_{u_{X_{\mathcal{I}}, Y_{\mathcal{I}}}}(X_i)$$

$$\overset{(c)}{=} \Pr\{E_{\{i\} \times T^c} \neq \mathbf{1}^{T^c}\} |\mathcal{X}_i|$$

$$\overset{(d)}{=} \left(1 - \prod_{j \in T^c} e_{i,j}\right) |\mathcal{X}_i| \tag{88}$$

where

(a) follows from (53) that for each $j \in \mathcal{I}$,

$$Y_{i,j} = \begin{cases} \varepsilon & \text{if } E_{i,j} = 1, \\ X_i & \text{otherwise.} \end{cases}$$

(b) follows from (61) that $X_{\mathcal{I}}$ and $E_{\mathcal{I} \times \mathcal{I}}$ are independent.

(c) follows from (81) that X_i is uniform on $|\mathcal{X}_i|$.

(d) follows from (60) and (61).

Combining (84), (87) and (88), we have

$$I_{u_{X_\mathcal{I},Y_\mathcal{I}}}(X_T;Y_{T^c}|X_{T^c}) \geq \sum_{i\in T}\left(1 - \prod_{j\in T^c} e_{i,j}\right)|\mathcal{X}_i|. \tag{89}$$

Using Theorem 1, (62), (83) and (89), we have $\mathcal{R}_{\text{in}}^{\text{WEN}} \subseteq \mathcal{C}_+$.

ACKNOWLEDGMENT

REFERENCES

[1] S. L. Fong and R. W. Yeung, "Cut-set bounds for networks with zero-delay nodes," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 3837–3850, 2015.

[2] C. E. Shannon, "Two-way communication channels," in *Proc. the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1961, pp. 611–644.

[3] A. El Gamal, N. Hassanpour and J. Mammen, "Relay networks with delays," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3413–3431, Oct. 2007.

[4] I.-J. Baik and S.-Y. Chung, "Causal relay networks and new cut-set bounds," in *Proc. Allerton Conference on Communication, Control and Computing*, Sep. 2011, pp. 247–252.

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

[6] A. S. Avestimehr, S. N. Diggavi, and D. N. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, 2011.

[7] E. Perron, "Information-theoretic secrecy for wireless networks," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.

[8] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence — Part I: Point-to-point channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, 2011.

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

[10] S. H. Lim, Y.-H. Kim, A. El Gamal and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, 2011.

[11] M. Yassaee and M. R. Aref, "Slepian-Wolf coding over cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3462–3482, 2011.

[12] M. Effros, "On dependence and delay: Capacity bounds for wireless networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, Apr. 2012, pp. 550–554.

[13] G. Kramer, "Information networks with in-block memory," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2105 – 2120, 2014.

[14] R. W. Yeung, *Information Theory and Network Coding*. Springer, 2008.