

# On Equivalence for Networks of Noisy Channels under Byzantine Attacks

Mayank Bakshi      Michelle Effros      Tracey Ho  
California Institute of Technology  
Pasadena, California 91125, USA  
Email: {mayank, effros, tho}@caltech.edu

**Abstract**—We consider the problem of finding network coding capacities of networks of independent point-to-point channels in the presence of a Byzantine adversary. We assume that the adversary knows all messages, and noise values and the code used to communicate across the network. The adversary controls an unknown subset of edges and can replace the channel output vectors from those edges. We show that finding the capacity for the above network is equivalent to finding the capacity of a network that is obtained by replacing each finite input alphabet point-to-point channel by a noiseless link of the noisy channel capacity. Our result shows the asymptotic optimality of separation between channel coding for each link followed by network coding for the resulting network under the corresponding model of adversarial attack.

## I. INTRODUCTION

One common approach for communicating in networks of noisy channels is to separate network coding and channel coding. In this approach, we operate each channel essentially losslessly with the help of a channel code. We then perform network coding on an essentially noise-free network. Indeed, in [1], [2], this approach is shown to be asymptotically optimal when the noise values on the distinct channels of the network are independent of each other. It is also known that when channels corresponding to different links are not independent, operating the channel code for each link independently may be strictly suboptimal (c.f. Example 2, [2]). In these cases, the dependence between the noise values on different links is exploited by first creating an appropriate dependence between the transmitted codewords on these channels and then jointly decoding them at the receiver.

In this work, we consider a network of independent point-to-point channels with the presence of a Byzantine adversary that observes all transmissions, messages, and channel noise values, and can corrupt some of the transmissions by replacing a constrained subset of the received channel outputs. The objective of the adversary is to maximize the probability of decoding error, and the capacity of the network is the set of vectors describing rates at which it is possible to reliably communicate across the network. It is tempting to believe that separation of network coding and channel coding is suboptimal in the case of our adversarial model due to the potential for statistical dependence between the "noise" observed on edges

controlled by the adversary. We show, however, that the capacity of this network equals the adversarial capacity of another network in which each channel is replaced by a noise-free capacitated link of the same capacity. Thus, it is asymptotically optimal to operate the adversarial network code independently of the channel code in this framework. We do not assume any special structure on the topology of the network, e.g., we allow unequal link capacities and networks with cycles. We also allow arbitrary model of adversarial attack, e.g. edge-based or node-based attack. The result immediately extends previous adversarial network coding capacity results from noise-free networks (e.g. [3], [4], [5], [6], [7]) to that of networks of independent point-to-point channels.

The proof follows the strategy introduced in [1], [2]. In Section III, we show that the adversarial capacity of a network is same as that of a *stacked network* comprised of many copies of the same network. In Section IV, we show that replacing one of the channels with a noiseless link of equal capacity does not alter the adversarial network coding capacity of the stacked network. We begin with a formal problem definition in Section II.

## II. PRELIMINARIES

### A. Network Model

We define a network  $\mathcal{N}$  to be a pair  $(\mathcal{G}, \mathcal{C})$ . Here  $\mathcal{G} = (\mathcal{V}, E)$  is a directed graph with vertices  $\{1, \dots, m\}$  and directed edges  $E \subseteq \mathcal{V} \times \mathcal{V}$ . Each edge  $e \in E$  describes the input and output of a point-to-point channel  $\mathcal{C}_e$ . The full collection of channels is given by  $\mathcal{C} = (\mathcal{C}_e : e \in E)$ .

For each  $e \in E$ , channel  $\mathcal{C}_e$  is given by a vector  $(\mathcal{X}^{(e)}, \mathcal{Y}^{(e)}, \mathcal{Z}^{(e)}, P_e, \Upsilon_e)$ , where  $\mathcal{X}^{(e)}$ ,  $\mathcal{Y}^{(e)}$ , and  $\mathcal{Z}^{(e)}$  are, respectively, the input, output, and noise alphabets of the channel,  $P_e$  is the probability distribution of the noise, and  $\Upsilon_e : \mathcal{X}^{(e)} \times \mathcal{Z}^{(e)} \rightarrow \mathcal{Y}^{(e)}$  is the channel map that determines the channel output as a function of the channel input and noise. The noise distribution  $P_e$  and mapping  $\Upsilon_e$  together induces a conditional probability distribution of the channel output given the channel input, here denoted by  $p_e(\cdot|\cdot)$ . Thus, the random variables  $X^{(e)}$ ,  $Y^{(e)}$ , and  $Z^{(e)}$  denoting the input, the output, and the noise value of the channel are related as

$$Y^{(e)} = \Upsilon_e(X^{(e)}, Z^{(e)}),$$

with

$$p_e(y|x) = \int_{\{z: \Upsilon_e(x,z)=y\}} P_e(z) dz.$$

For each  $t \in \mathbb{N}^+$  and  $e \in E$ , let  $X_t^{(e)} \in \mathcal{X}^{(e)}$ ,  $Y_t^{(e)} \in \mathcal{Y}^{(e)}$ , and  $Z_t^{(e)} \in \mathcal{Z}^{(e)}$ , respectively, be the random variables denoting the transmitted, received, and noise values for edge  $e$  at time  $t$ . We assume that each transmission on edge  $e$  involves a delay of unit time and that the noise on all channels is independent and memoryless. Thus,

$$Y_{t+1}^{(e)} = \Upsilon_e(X_t^{(e)}, Z_t^{(e)}) \quad \forall e \in E, t \in \mathbb{N}^+$$

and

$$P_E(Z_\tau^{(e)} : e \in E, \tau = 1, \dots, t) = \prod_{e \in E} \prod_{\tau=1}^t P_e(Z_\tau^{(e)}).$$

Here,  $P_E$  denotes the joint distribution of the noise values.

For notational convenience, we adopt the following convention to represent collections of random vectors. For every collection of random variables  $Q_1, Q_2, \dots$  taking values from a set  $\mathcal{Q}$ , we denote the row vector  $[Q_t, Q_{t+1}, \dots, Q_{t+n-1}] \in \mathcal{Q}^n$  by  $Q_{t:t+n-1}$ . We specify column vectors by underlining them and the element of a given row from the column vector by parenthesis. Thus,  $\underline{Q} \in \mathcal{Q}^N$  represents the column vector  $[\underline{Q}(1), \underline{Q}(2), \dots, \underline{Q}(N)]^T$  with  $\underline{Q}(i) \in \mathcal{Q}$  for all  $i$ .

For each  $v \in \mathcal{V}$  and  $t \in \mathbb{N}^+$ , let  $X_t^{(v,*)} \triangleq (X_t^{(v,w)} : (v,w) \in E)$  and  $X_t^{(*,v)} \triangleq (X_t^{(u,v)} : (u,v) \in E)$  denote the time- $t$  random variables on edges outgoing from  $v$  and incoming to  $v$ , respectively; the alphabets for  $X_t^{(v,*)}$  and  $X_t^{(*,v)}$  are  $\mathcal{X}^{(v,*)} = \prod_{u:(v,u) \in E} \mathcal{X}^{(v,u)}$  and let  $\mathbf{X}_t \triangleq (X_t^{(e)} : e \in E)$  denote all transmitted random variables in the network at time  $t$ . Similarly, define  $Y_t^{(v,*)}$ ,  $Y_t^{(*,v)}$ ,  $Z_t^{(v,*)}$ ,  $Z_t^{(*,v)}$ ,  $\mathbf{Y}_t$ , and  $\mathbf{Z}_t$  for each  $v \in \mathcal{V}$  and  $t \in \mathbb{N}^+$ . Let  $\mathcal{X}$ ,  $\mathcal{X}^{(u,*)}$ , and  $\mathcal{X}^{(*,v)}$  denote the product sets  $\prod_{e \in E} \mathcal{X}^{(e)}$ ,  $\prod_{v:(u,v) \in E} \mathcal{X}^{(u,v)}$ , and  $\prod_{v:(u,v) \in E} \mathcal{X}^{(u,v)}$ . Similarly define  $\mathcal{Y}$ ,  $\mathcal{Y}^{(u,*)}$ ,  $\mathcal{Y}^{(*,v)}$ ,  $\mathcal{Z}$ ,  $\mathcal{Z}^{(u,*)}$ , and  $\mathcal{Z}^{(*,v)}$ .

### B. Network Code

Let  $\mathcal{M} = \{(u, V) : u \in \mathcal{V}, V \subseteq \mathcal{V} \setminus \{u\}\}$  denote the set of possible pairs of source nodes and sink sets. A network coding solution  $\mathcal{S}(\mathcal{N})$  implemented over  $n$  time steps maps is defined by message alphabet  $\mathcal{W} = \prod_{(u,V) \in \mathcal{M}} \mathcal{W}^{(u \rightarrow V)}$ , the collection of encoder maps  $\{f_t^{(u,v)} : (u,v) \in E, t \in \{1, \dots, n\}\}$  with

$$f_t^{(u,v)} : \prod_{V \subseteq \mathcal{V}} \mathcal{W}^{(u \rightarrow V)} \times \prod_{v':(v',u) \in E} (\mathcal{Y}^{(v',u)})^t \rightarrow \mathcal{X}^{(u,v)}$$

that determine the transmitted random variable  $X_t^{(u,v)}$  as a function of the messages  $(W^{(u \rightarrow V)} : V \subseteq \mathcal{V})$  and received vectors  $Y_{1:t}^{(*,u)}$  at node  $u$ , and the decoder maps  $\{g^{(u)} : u \in V\}$  with

$$g^{(u)} : \prod_{V \subseteq \mathcal{V} \setminus \{u\}} \mathcal{W}^{(u \rightarrow V)} \times \prod_{v:(v,u) \in E} (\mathcal{Y}^{(v,u)})^n \rightarrow \prod_{\substack{V \subseteq \mathcal{V}: u \in V \\ v \in \mathcal{V} \setminus \{u\}}} \mathcal{W}^{(v \rightarrow u)}$$

that determine the reconstructed messages  $(\hat{W}^{(v \rightarrow V, u)} : (v, V) \in \mathcal{M}, v \in \mathcal{V}, u \in V)$  as a function of the messages  $(W^{(u \rightarrow V')} : V' \subseteq \mathcal{V} \setminus \{u\})$  and received vectors  $Y_{1:t}^{(*,u)}$  at node  $u$  for all  $t = 1, \dots, n$  and  $u \in \mathcal{V}$ . Let  $\mathbf{R} = (R(u, V) : (u, V) \in \mathcal{M}) \in \mathbb{R}^{|\mathcal{M}|}$ . We say that a solution  $\mathcal{S}(\mathcal{N})$  is a rate  $\mathbf{R}$  solution if  $|\mathcal{W}^{(u \rightarrow V)}| = 2^{nR(u,V)}$  for all  $(u, V) \in \mathcal{M}$ . Without loss of generality, we assume that all messages are either binary vectors or binary matrices of appropriate dimensions.

### C. Adversarial Model

We assume an omniscient Byzantine adversary that observes all messages  $(W^{(u \rightarrow V)} : (u, V) \in \mathcal{M})$ , noise values  $\mathbf{Z}_{1:n}$ , and the network code  $\mathcal{S}$  in operation. Thus, the adversary can deduce all transmitted and received vectors,  $\mathbf{X}_{1:n}$  and  $\mathbf{Y}_{1:n}$ . The adversary picks a subset  $\sigma$  from the set  $\Sigma$  of permissible attack-sets and replaces the vectors  $(Y_t^{(e)} = \Upsilon(X_{t-1}^{(e)}, Z_{t-1}^{(e)})) : e \in \sigma, t = 1, \dots, n$  of channel outputs on these edges with the vector  $\mathbf{A}_{1:n} = (A_{1:n}^{(e)} : e \in \sigma)$  of his own choice. The set  $\Sigma$  is known to the designer of the network code, but the chosen attack set  $\sigma \in \Sigma$  is unknown.

We say that there is a *decoding error* if  $\hat{W}^{(v \rightarrow V, u)} \neq W^{(v \rightarrow V)}$  for some  $v \in \mathcal{V}$ ,  $V \subseteq \mathcal{V} \setminus \{v\}$ , and  $u \in V$ . For a given solution  $\mathcal{S}(\mathcal{N})$  that is implemented over  $n$  time steps, and for each  $(u, V) \in \mathcal{M}$  and  $v \in V$ ,  $\hat{W}^{(u \rightarrow V, v)}$  is a deterministic function of the messages  $W = (W^{(u \rightarrow V)} : (u, V) \in \mathcal{M})$ , the noise values  $\mathbf{Z}_{1:n}$ , the attack-set  $\sigma$ , and the injected vector  $\mathbf{A}_{1:n}$ ; let  $G_S^{(u \rightarrow V, v)} : \mathcal{W} \times \mathcal{Z}^n \times \Sigma \times \mathcal{Y}^n \rightarrow \mathcal{W}$  denote this function. Since the adversary knows  $W$  and  $\mathbf{Z}_{1:n}$ , he can compute the decoded message for every possible choice of  $\sigma$  and  $\mathbf{A}_{1:n}$ . The adversary's goal is to chose  $\sigma$  and  $\mathbf{A}_{1:n}$  to minimize the rate of reliable communication. We define the set  $\mathcal{E}(\mathcal{S}) \subseteq \mathcal{W} \times \mathcal{Z}^n$  as the collection of messages and noise values for which it is possible for the adversary to cause a decoding error for any of the messages, i.e.,

$$\begin{aligned} \mathcal{E}(\mathcal{S}) \triangleq & \{(w, \mathbf{z}_{1:n}) \in \mathcal{W} \times \mathcal{Z}^n : \\ & G_S^{(u \rightarrow V, v)}(w, \mathbf{z}_{1:n}, \mathbf{a}_{1:n}, \sigma) \neq w^{(u, V)} \text{ for some} \\ & (u, V) \in \mathcal{M}, v \in V, \sigma \in \Sigma, \text{ and } \mathbf{a}_{1:n} \in \prod_{e \in \sigma} (\mathcal{Y}^{(e)})^n \}. \end{aligned}$$

The *probability of error* for the solution  $\mathcal{S}(\mathcal{N})$  is

$$\begin{aligned} P_{\mathcal{E}}(\mathcal{S}) & \triangleq \Pr_{W, \mathbf{Z}_{1:n}} ((W, \mathbf{Z}_{1:n}) \in \mathcal{E}(\mathcal{S})) \\ & = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \int_{\{\mathbf{z}_{1:n} : (w, \mathbf{z}_{1:n}) \in \mathcal{E}(\mathcal{S})\}} P_E(\mathbf{z}_{1:n}) d\mathbf{z}_{1:n}. \end{aligned}$$

We say that a solution  $\mathcal{S}(\mathcal{N})$  that is implemented over  $n$  time steps is a  $(\lambda, \mathbf{R})$ -solution if  $|\mathcal{W}^{(u \rightarrow V)}| = 2^{nR(u,V)}$  for every  $(u, V) \in \mathcal{M}$  and  $P_{\mathcal{E}}(\mathcal{S}) < \lambda$ . The *capacity region*  $\mathcal{R}(\mathcal{N})$  of a network  $\mathcal{N}$  is the closure of the set of all rate vectors  $\mathbf{R}$  for which a  $(\lambda, \mathbf{R})$ -solution exists for every  $\lambda > 0$ .

## III. STACKED NETWORK

Following the proof method employed in [2], we define the stacked network as follows. Let  $\mathcal{N} = (\mathcal{G}, \mathcal{C})$  be a network with vertex set  $\mathcal{V} = \{1, \dots, m\}$  and edge set  $E$ .

For each  $e \in E$ , let  $\underline{P}_e$  be a probability distribution on  $(\mathcal{Z}^{(e)})^N$  obtained by forming an  $N$ -fold product of  $P_e$  with itself, i.e.,  $\underline{P}_e(\underline{z}_e) = \prod_{i=1}^N P_e(\underline{z}_e(i))$  for all  $\underline{z}_e \in (\mathcal{Z}^{(e)})^N$ . Next, let  $\underline{\Upsilon}_e : (\mathcal{X}^{(e)})^N \times (\mathcal{Z}^{(e)})^N \rightarrow (\mathcal{Y}^{(e)})^N$  represent a channel that maps pairs  $(\underline{x}_e, \underline{z}_e) \in (\mathcal{X}^{(e)})^N \times (\mathcal{Z}^{(e)})^N$  to  $\underline{\Upsilon}_e(\underline{x}_e, \underline{z}_e) = [\Upsilon_e(\underline{x}_e(1), \underline{z}_e(1)), \dots, \Upsilon_e(\underline{x}_e(N), \underline{z}_e(N))]^T$ . We define the  $N$ -fold stacked network  $\underline{\mathcal{N}}$  derived from  $\mathcal{N} = (\mathcal{G}, \mathcal{C})$  as a pair  $(\underline{\mathcal{G}}, \underline{\mathcal{C}})$ , where,  $\underline{\mathcal{G}} \triangleq \mathcal{G}$  and  $\underline{\mathcal{C}} \triangleq ((\mathcal{X}^{(e)})^N, (\mathcal{Y}^{(e)})^N, (\mathcal{Z}^{(e)})^N, \underline{P}_e, \underline{\Upsilon}_e)$  for all  $e \in E$ .

For the network  $\underline{\mathcal{N}}$ , we denote the messages corresponding to the pair  $(u, V) \in \mathcal{M}$  by matrix  $\underline{W}_{1:nR(u, V)}^{(u \rightarrow V)}$ , and the transmitted, received, and noise values for the edge  $(u, v) \in E$  by matrices  $\underline{X}_{1:n}, \underline{Y}_{1:n}$ , and  $\underline{Z}_{1:n}$  respectively. Let  $\underline{\mathcal{N}}(1), \underline{\mathcal{N}}(2), \dots, \underline{\mathcal{N}}(N)$  be  $N$  copies of the network  $\mathcal{N}$ . For each  $i = 1, \dots, N$ , associate vector  $\underline{W}_{1:nR(u, V)}^{(u \rightarrow V)}(i)$  with the message corresponding to the pair  $(u, V) \in \mathcal{M}$ , and  $\underline{X}_{1:n}(i), \underline{Y}_{1:n}(i)$ , and  $\underline{Z}_{1:n}(i)$ , with the messages, and transmitted, received, and noise values, respectively, for the edge  $(u, v) \in E$  in  $\underline{\mathcal{N}}(i)$ .

We visualize  $\underline{\mathcal{N}}$  as a stack with layers  $\underline{\mathcal{N}}(1), \underline{\mathcal{N}}(2), \dots, \underline{\mathcal{N}}(N)$  and infinite capacity bidirectional edges connecting all  $N$  copies a give vertex  $v \in \mathcal{V}$  to each other. Thus, for each  $v \in \mathcal{V}$  and  $i = 1, \dots, N$ , the transmitted vector  $\underline{X}_{1:n}^{(v, *)}(i)$  may be a function of all messages  $(\underline{W}_{1:nR(u, V)}^{(v \rightarrow U)} : (v, U) \in \mathcal{M})$  and received vectors  $\underline{Y}_{1:n}^{(*, v)}$ .

The capacity region for the stacked network  $\mathcal{R}(\underline{\mathcal{N}})$  is normalized by the number of layers  $N$ . In [2], it is shown that the capacity regions for  $\mathcal{N}$  and  $\underline{\mathcal{N}}$  are equal when none of the edges are corruptible by the adversary. Even though the presence of adversary changes the network capacity, the arguments of Lemma 1 of [2] extend readily to our setup. We state this in the following Lemma without proof.

*Lemma 1:* For any network  $\mathcal{N} = (\mathcal{G}, \mathcal{C})$ ,  $\mathcal{R}(\mathcal{N}) = \mathcal{R}(\underline{\mathcal{N}})$ .

Next, we show that there exists a sequence of solutions to the stacked network such that the error probability decays exponentially with the number of layers. In the non-adversarial case, the mutual independence of  $(\underline{\mathcal{Z}}(i) : i = 1, \dots, N)$  results in independent decoding errors for a solution that operates on each layer independently. Thus, applying a randomly generated error correcting code to all messages  $(\underline{W}_{1:nR(u, V)}^{(u \rightarrow V)} : (u, V) \in \mathcal{M})$  before they are processed by the network code ensures an exponential decay of error probability [8]. However, in the presence of an adversary, decoding errors may no longer be independent across the layers. We overcome this difficulty by first designing a solution to the stacked network for which the error probability is maximum when decoding errors are statistically independent across layers, and then showing that, under this condition, the error probability for this solution decays exponentially in the number of layers.

*Theorem 1:* Given any  $\mathbf{R} \in \text{int}(\mathcal{R}(\mathcal{N}))$ , there exists a  $(2^{-N\delta}, \mathbf{R})$ -solution  $\tilde{\mathcal{S}}(\underline{\mathcal{N}})$  for  $\underline{\mathcal{N}}$  for some  $\delta > 0$  and for all  $N$  large enough.

*Proof:* Let  $\lambda > 0$  and let  $\rho > H(2\lambda)$ , where  $H(\cdot)$  is the binary entropy function. Let  $\mathcal{S}(\mathcal{N})$  be a  $(\lambda, \mathbf{R})$ -solution for  $\mathcal{N}$  with  $\mathcal{W} = \prod_{(u, V) \in \mathcal{M}} \mathcal{W}^{(u \rightarrow V)} = \{0, 1\}^{nR(u, V)}$ . We

design the solution  $\tilde{\mathcal{S}}(\underline{\mathcal{N}})$  as follows. For each  $(u, V) \in \mathcal{M}$ , let  $\underline{w}_{1:nR(u, V)}^{(u \rightarrow V)}$  be a two dimensional binary  $(1 - \rho)N \times nR(u, V)$  matrix. We first encode  $\underline{w}_{1:nR(u, V)}^{(u \rightarrow V)}$  by using a different error correcting code for each column. Next, we transmit each row of the resulting binary matrices  $(\tilde{\underline{w}}_{1:nR(u, V)}^{(u \rightarrow V)} : (u, V) \in \mathcal{M})$  on a different layer using the solution  $\mathcal{S}(\mathcal{N})$ . Finally, we employ nearest-neighbor decoding at each node to reconstruct the messages.

*Code Construction:* Fix a pair  $(u, V) \in \mathcal{M}$  and  $k \in \{1, \dots, nR(u, V)\}$ . Consider a binary symmetric channel  $\tilde{\mathcal{C}}$  with crossover probability  $2\lambda$ . Let  $\Psi_k^{(u \rightarrow V)} : \{0, 1\}^{(1-\rho)N} \rightarrow \{0, 1\}^N$  and  $\Phi_k^{(u \rightarrow V)} : \{0, 1\}^N \rightarrow \{0, 1\}^{(1-\rho)N}$  be the encoder and decoder mappings for an error correcting code for  $\tilde{\mathcal{C}}$  of blocklength  $N$  and rate  $(1 - \rho)$  that is designed randomly as follows.

Select  $\tilde{\mathcal{W}}_k^{(u \rightarrow V)} \subseteq \{0, 1\}^N$  of size  $2^{(1-\rho)N}$  by independently picking each element of  $\tilde{\mathcal{W}}_k^{(u \rightarrow V)}$  from  $\{0, 1\}^N$  using a uniform distribution. The encoder  $\Psi_k^{(u \rightarrow V)}$  maps each message  $\underline{b} \in \{0, 1\}^{(1-\rho)N}$  to a unique codeword  $\hat{\underline{b}} \in \tilde{\mathcal{W}}_k^{(u \rightarrow V)}$ . The decoder  $\Phi_k^{(u \rightarrow V)}$  maps each received vector  $\hat{\underline{b}} \in \{0, 1\}^N$  to the reconstruction  $\underline{\hat{b}} \in \{0, 1\}^{(1-\rho)N}$  that corresponds to the nearest valid codeword. Let where,  $d_H(\cdot, \cdot)$  denote the hamming distance between two binary vectors. In other words,

$$\Phi_k^{(u \rightarrow V)}(\hat{\underline{b}}) = \underset{\hat{\underline{b}} \in \{0, 1\}^{nR(u, V)}}{\text{argmin}} d_H(\Psi_k^{(u \rightarrow V)}(\hat{\underline{b}}), \hat{\underline{b}}),$$

We construct the solution  $\tilde{\mathcal{S}}(\underline{\mathcal{N}})$  as follows. Let the message alphabet be  $\underline{\mathcal{W}} = \prod_{(u, V) \in \mathcal{M}} \mathcal{W}^{(u \rightarrow V)}$ , where  $\underline{\mathcal{W}}^{(u \rightarrow V)} \triangleq \{0, 1\}^{nR(u, V) \times (1-\rho)N}$  is the set of  $nR(u, V) \times (1 - \rho)N$  binary matrices. Let  $\underline{w}_{1:nR(u, V)}^{(u \rightarrow V)} \in \{0, 1\}^{nR(u, V) \times (1-\rho)N}$  be the message intended for the connection  $(u, V) \in \mathcal{M}$ . The solution  $\tilde{\mathcal{S}}(\underline{\mathcal{N}})$  performs the following sequence of operations.

- 1) For each pair of vertices  $(u, V) \in \mathcal{M}$  and message  $\underline{w}_{1:nR(u, V)}^{(u \rightarrow V)} \in \underline{\mathcal{W}}^{(u \rightarrow V)}$ , let  $\tilde{\underline{w}}_{1:nR(u, V)}^{(u \rightarrow V)} \in \{0, 1\}^{nR(u, V) \times N}$  with

$$\tilde{\underline{w}}_k^{(u \rightarrow V)} = \Psi_k^{(u \rightarrow V)}(\underline{w}_k^{(u \rightarrow V)}).$$

- 2) For each  $i = 1, \dots, N$ , communicate messages  $(\tilde{\underline{w}}_{1:nR(u, V)}^{(u \rightarrow V)}(i) : (u, V) \in \mathcal{M})$  using the solution  $\mathcal{S}(\mathcal{N})$  on  $\underline{\mathcal{N}}(i)$ . Let  $(\hat{\underline{w}}_{1:nR(u, V)}^{(u \rightarrow V)}(i) : (u, V) \in \mathcal{M})$  be the reconstructed messages after operating  $\mathcal{S}(\mathcal{N})$  on  $\underline{\mathcal{N}}(i)$ .
- 3) For every  $(u, V) \in \mathcal{M}$ , each vertex  $v \in V$  outputs a reconstruction  $\hat{\underline{w}}_{1:nR(u, V)}^{(u \rightarrow V, v)} \in \{0, 1\}^{nR(u, V) \times (1-\rho)N}$  with

$$\hat{\underline{w}}_k^{(u \rightarrow V, v)} = \Phi_k^{(u \rightarrow V)}(\hat{\underline{w}}_k^{(u \rightarrow V, v)}), \quad k = 1, \dots, nR(u, V).$$

*Analysis of error probability:* Let  $(u, V) \in \mathcal{M}$ ,  $v \in V$  and  $k \in \{1, \dots, nR(u, V)\}$ . Since  $\tilde{\mathcal{C}}$  is symmetrical and the input is uniformly distributed, the decoder  $\Psi_k^{(u \rightarrow V)}$  maps each received vector to the maximum likelihood estimate of the input given the received vector. By previous results on error exponents (c.f.[8]), we know that such a code achieves an error probability of  $2^{-N\delta}$  for some  $\tilde{\delta} = \tilde{\delta}(\rho, \lambda)$  since

the rate  $1 - \rho$  is less than the capacity  $1 - H(2\lambda)$  of the channel  $\tilde{\mathcal{C}}$ . Denote the message and the received vector for the code  $(\Psi_k^{(u \rightarrow V)}, \Phi_k^{(u \rightarrow V)})$  by random variables  $\underline{B}$  and  $\hat{\underline{B}}$  respectively.  $\underline{B}$  is uniformly distributed on  $\{0, 1\}^{(1-\rho)N}$  and  $\hat{\underline{B}}$  is statistically related to  $\Psi_k^{(u \rightarrow V)}(\underline{B})$  via the channel  $\tilde{\mathcal{C}}$ . Therefore,

$$p_{\tilde{\mathcal{C}}}(\underline{B} \neq \Phi_k^{(u \rightarrow V)}(\hat{\underline{B}})) < 2^{-N\delta}. \quad (1)$$

Next, we show that the block error probability for the code  $(\Psi_k^{(u \rightarrow V)}, \Phi_k^{(u \rightarrow V)})$  over the channel  $\tilde{\mathcal{C}}$  is an upper bound for the block error probability  $\Pr(\underline{W}_k^{(u \rightarrow V)} \neq \hat{\underline{W}}_k^{(u \rightarrow V, v)})$ . Note that by the choice of the solution  $\mathcal{S}(\mathcal{N})$ ,  $\Pr(\underline{W}_k^{(u \rightarrow V)}(i) \neq \hat{\underline{W}}_k^{(u \rightarrow V, v)}(i)) < \lambda$  for each  $i = 1, \dots, N$ . Thus,

$$\max_{b \in \{0, 1\}} \Pr(\underline{W}_k^{(u \rightarrow V)}(i) \neq \hat{\underline{W}}_k^{(u \rightarrow V)}(i) | \underline{W}_k^{(u \rightarrow V)}(i) = b) < 2\lambda$$

for each  $i = 1, \dots, N$ . Let  $w_H(\cdot)$  denote the number of 1's in a binary vector. For every  $\tilde{\underline{b}} \in \mathcal{W}_k^{(u \rightarrow V)}$ , let  $\pi_k^{(u \rightarrow V)}(\tilde{\underline{b}}) \subseteq \{0, 1\}^N$  denote the set of all minimal weight error patterns that are decoded incorrectly by  $\Phi_k^{(u \rightarrow V)}$ , i.e.,

$$\begin{aligned} \pi_k^{(u \rightarrow V)}(\tilde{\underline{b}}) &\triangleq \{\underline{e} \in \{0, 1\}^N : \Phi_k^{(u \rightarrow V)}(\tilde{\underline{b}}) \neq \Phi_k^{(u \rightarrow V)}(\tilde{\underline{b}} \oplus \underline{e}) \\ &\text{and } \Phi_k^{(u \rightarrow V)}(\tilde{\underline{b}}) = \Phi_k^{(u \rightarrow V)}(\tilde{\underline{b}} \oplus \underline{e}) \forall \tilde{\underline{e}} \text{ s.t.} \\ &\quad w_H(\tilde{\underline{e}}) = w_H(\underline{e}) - 1 \text{ and } d_H(\tilde{\underline{e}}, \underline{e}) = 1\}. \end{aligned}$$

Let  $\underline{e} \in \{0, 1\}^N$ . Then,

$$\begin{aligned} &\Pr(\underline{W}_k^{(u, V)} \neq \hat{\underline{W}}_k^{(u \rightarrow V, v)} | \underline{W}_k^{(u, V)} = \underline{e}) \\ &= \Pr(\Phi_k^{(u \rightarrow V)}(\tilde{\underline{W}}_k^{(u \rightarrow V)}) \neq \Phi_k^{(u \rightarrow V)}(\hat{\underline{W}}_k^{(u \rightarrow V, v)}) | \underline{W}_k^{(u \rightarrow V)} = \underline{e}) \\ &= \sum_{\underline{e} \in \pi_k^{(u \rightarrow V)}(\underline{e})} \Pr(\bigcap_{i: \underline{e}(i)=1} \{\hat{\underline{W}}_k^{(u \rightarrow V, v)}(i) = \underline{e}(i) \oplus \underline{e}(i)\} \\ &\quad \tilde{\underline{W}}_k^{(u \rightarrow V)} = \underline{e}) \\ &\leq \sum_{\underline{e} \in \pi_k^{(u \rightarrow V)}(\underline{e})} (2\lambda)^{w_H(\underline{e})} \end{aligned} \quad (2)$$

$$\begin{aligned} &= p_{\tilde{\mathcal{C}}}(\Phi_k^{(u \rightarrow V)}(\underline{e}) \neq \Phi_k^{(u \rightarrow V)}(\hat{\underline{W}}_k^{(u \rightarrow V, v)})) \\ &< 2^{-N\delta}. \end{aligned} \quad (3)$$

The bound in (2) is a consequence of the fact that for each  $i = 1, \dots, N$  and  $k = 1, \dots, nR(u, V)$ , the event  $\{\tilde{\underline{W}}_k^{(u \rightarrow V)}(i) \neq \hat{\underline{W}}_k^{(u \rightarrow V)}(i)\}$  depends only on the noise values  $\underline{Z}_{1:n}(i)$  and the messages  $\underline{W}(i)$  and can occur with probability at most  $2\lambda$  under all possible adversarial actions. Step (3) follows from the fact that the transition probability for the channel  $\tilde{\mathcal{C}}$  is  $2\lambda$ . Finally, applying a union bound over all values of  $k$  and  $(u, V)$ , we obtain

$$P_{\mathcal{E}}(\tilde{\mathcal{S}}) \leq n|\mathcal{M}|2^{-N\delta} \max_{(u, V) \in \mathcal{M}} R(u, V) \leq 2^{-N\delta}$$

for every  $\delta < \tilde{\delta}$  for large enough  $N$ . ■

In this section, we prove that finding the capacity region of the network  $\mathcal{N}$  is equivalent to finding the capacity of a network  $\hat{\mathcal{N}}_R$  where one of the links  $\hat{e} = (1, 2)$  is replaced by a noiseless links of capacity  $R = C(\hat{e})$ . Koetter et al [2] prove this by first showing that for every  $R < C(\hat{e})$ , the capacity region of the  $\hat{\mathcal{N}}_R$  is a subset of that of  $\mathcal{N}$ . Next, they show that for every  $R > C(\hat{e})$ , the capacity region of  $\mathcal{N}$  is a subset of that of  $\hat{\mathcal{N}}_R$ . We follow a similar proof outline for our case. The proof of the first part follows the arguments of [2] exactly. We state it without proof in the following Theorem.

*Theorem 2:* For the networks  $\mathcal{N}$  and  $\hat{\mathcal{N}}_R$  defined above,  $\mathcal{R}(\hat{\mathcal{N}}_R) \subseteq \mathcal{R}(\mathcal{N})$  if  $R < C$ .

Next, to prove that  $\mathcal{R}(\hat{\mathcal{N}}_R) \supseteq \mathcal{R}(\mathcal{N})$  if  $R > C$ , Koetter et al argue that a noisy channel may be emulated on the lossless link  $\hat{e}$  in the stacked network  $\underline{\mathcal{N}}_R$  by using a randomly generated source code that operates across the layers. Their proof relies on typicality of the vector  $\underline{X}_t^{(u, v)}$ , which follows from the statistical independence of random variables corresponding to different layers. This assumption is not true in our case because the adversary may introduce dependence between different layers. To accommodate this possibility, we modify the proof of [2] to use a universal source code that first determines the type of the received sequence and then emulates the channel using a source code designed specifically for the observed type. We assume here that the alphabet  $\mathcal{X}^{(\hat{e})}$  is finite.

*Theorem 3:* For the networks  $\mathcal{N}$  and  $\hat{\mathcal{N}}_R$  defined above,  $\mathcal{R}(\hat{\mathcal{N}}_R) \supseteq \mathcal{R}(\mathcal{N})$  if  $R > C$ .

*Proof:* First, we design a sequence of universal source codes  $\{(\alpha_{N,t}, \beta_{N,t})\}_{t=1,2,\dots}$ , each at rate  $\hat{R} \in (C(\hat{e}), R)$ , that operate by describing the type of the input followed by the index of the codeword picked from a codebook designed for that type. Next, we modify a given  $(2^{-N\delta}, \mathbf{R})$ -solution  $\mathcal{S}(\underline{\mathcal{N}})$  for  $\underline{\mathcal{N}}$  to a solution  $\hat{\mathcal{S}}(\underline{\mathcal{N}}_R)$  for  $\underline{\mathcal{N}}_R$  by emulating the channel on the link  $\hat{e}$  by the source codes  $\{(\alpha_{N,t}, \beta_{N,t})\}_{t=1,2,\dots}$ . Finally, we conclude that the error probability for the solution  $\hat{\mathcal{S}}(\underline{\mathcal{N}}_R)$  vanishes as  $N$  grows without bound.

*Construction of  $\hat{\mathcal{S}}(\underline{\mathcal{N}}_R)$ :* Let  $\underline{\mathcal{S}}(\underline{\mathcal{N}})$  be a  $(2^{-N\delta}, \mathbf{R})$ -solution to  $\underline{\mathcal{N}}$  implemented over  $n$  time steps and obtained from the solution  $\mathcal{S}(\mathcal{N})$  by following the construction in Theorem 1. We modify  $\underline{\mathcal{S}}(\underline{\mathcal{N}})$  to obtain a solution  $\hat{\mathcal{S}}(\underline{\mathcal{N}}_R)$  by first designing a sequence of universal source codes  $\{\alpha_{N,t}, \beta_{N,t}\}_{t=1,2,\dots}$  that operate at a rate  $\hat{R} \in (C(\hat{e}), R)$  and next appending the above source code to  $\underline{\mathcal{S}}(\underline{\mathcal{N}})$  to emulate the channel conditional probability  $p_{\hat{e}}$  on the the channel  $\hat{\mathcal{C}}_{\hat{e}}$  in network  $\hat{\mathcal{N}}_R$ .

*Design of  $\{\alpha_{N,t}, \beta_{N,t}\}_{t=1,2,\dots}$ :* Let  $\mathcal{P}_N$  be the set of all strong types of  $N$  length sequences from  $\mathcal{X}^{(\hat{e})}$ . For a vector  $\underline{x} \in (\mathcal{X}^{(\hat{e})})^N$ , let  $\hat{Q}(\underline{x}) \in \mathcal{P}_N$  denote the empirical distribution of  $\underline{x}$ . Let  $Q \in \mathcal{P}_N$ . Let  $(X_Q, Y_Q)$  be random variables jointly distributed on  $\mathcal{X}^{(\hat{e})} \times \mathcal{Y}^{(\hat{e})}$  such that  $\Pr(X_Q = x) = Q(x)$  and the conditional distribution of  $Y_Q$  given  $X_Q$  is  $p_{\hat{e}}$ . For each  $t = 1, \dots, n$ , select  $\mathcal{B}_t^Q \subseteq (\mathcal{Y}^{(\hat{e})})^N$  by choosing  $2^{N\hat{R}}$  elements uniformly at random from  $A_{\epsilon}^{(N)}(Y_Q)$ .

For  $t = 1, \dots, n$ , let the encoder

$$\alpha_{N,t} : (\mathcal{X}^{(\hat{e})})^N \rightarrow \{0, 1\}^{|\mathcal{X}^{(\hat{e})}| \log_2(N+1)} \times \{1, \dots, 2^{N\hat{R}}\}$$

consist of the pair of maps to  $(\alpha_{\mathcal{P}}, \alpha_{\mathcal{B}})$ , where, for each  $\underline{x} \in (\mathcal{X}^{(\hat{e})})^N$ ,  $\alpha_{\mathcal{P}}(\underline{x})$  is the binary description of  $\hat{Q}(\underline{x})$ , and  $\alpha_{\mathcal{B}}(\underline{x})$  is the index of some vector  $\hat{\underline{y}} \in \mathcal{B}_t^{\hat{Q}(\underline{x})}$  such that  $(\underline{x}, \hat{\underline{y}}) \in A_\epsilon^{(N)}(X_{\hat{Q}(\underline{x})}, Y_{\hat{Q}(\underline{x})})$ . If no such vector  $\hat{\underline{y}}$  exists, then  $\alpha_{\mathcal{B}}(\underline{x})$  is set to be 1. Next, let the decoder

$$\beta_{N,t} : \{0, 1\}^{|\mathcal{X}^{(\hat{e})}| \log_2(N+1)} \times \{1, \dots, 2^{N\hat{R}}\} \rightarrow (\mathcal{Y}^{(\hat{e})})^N$$

map pairs  $(x_{\mathcal{P}}, x_{\mathcal{B}})$  to the vector with index  $x_{\mathcal{B}}$  in  $\mathcal{B}_t^{\hat{Q}}$ , where  $\hat{Q}$  is the type described by  $x_{\mathcal{P}}$ .

Appending  $\{\alpha_{N,t}, \beta_{N,t}\}_{t=1,2,\dots}$  to  $\underline{\mathcal{S}}(\underline{N})$ : The solution  $\hat{\mathcal{S}}(\hat{N}_R)$  is identical to  $\underline{\mathcal{S}}(\underline{N})$  except for the maps at nodes 1 and 2. For  $(u, v) \in \mathcal{V} \times \mathcal{V}$ , let  $f^{(u,v)} : \underline{\mathcal{W}}^{(u \rightarrow *)} \times (\mathcal{Y}^{(*,u)})^{nN} \rightarrow (\mathcal{X}^{(u,*)})^{nN}$  denote the encoder that determines the codeword on the edge  $(u, v)$  and let  $g^{(u)} : \underline{\mathcal{W}}^{(u \rightarrow *)} \times (\mathcal{Y}^{(*,u)})^{nN} \rightarrow \prod_{V \subseteq \mathcal{V} \setminus \{u\}} \underline{\mathcal{W}}^{(* \rightarrow \{u\} \cup V)}$  denote the decoder for the messages meant for node  $u$  in the solution  $\underline{\mathcal{S}}(\underline{N})$ .

Let  $\hat{\underline{X}}_{1:n}^{(e)} = \underline{X}_{1:n}^{(e)}$  and  $\hat{\underline{Y}}_{1:n}^{(e)} = \underline{Y}_{1:n}^{(e)}$  for all  $e \neq \hat{e}$ . Let  $\hat{\underline{X}}_{1:n}^{(\hat{e})} = f^{(1,2)}(\underline{W}^{1 \rightarrow *}, \hat{\underline{Y}}_{1:n}^{(*,1)})$  and  $\hat{\underline{Y}}_t^{(\hat{e})} = \beta_{N,t-1}(\alpha_{N,t-1}(\hat{\underline{X}}_{t-1}^{(\hat{e})}))$ . Let  $\hat{\mathcal{S}}(\hat{N}_R)$  be a solution with encoder and decoder mappings  $(\hat{f}^{(u,v)} : (u, v) \in E)$  and  $(\hat{g}^{(u)} : u \in \mathcal{V})$ , where,

$$\hat{f}^{(u,v)}(\underline{W}^{(u \rightarrow *)}, \hat{\underline{Y}}_{1:n}^{(*,u)}) \triangleq \begin{cases} f^{(u,v)}(\underline{W}^{(u \rightarrow *)}, \hat{\underline{Y}}_{1:n}^{(*,u)}) & \text{if } (u, v) \neq \hat{e} \text{ and } u \neq 2 \\ \alpha(f^{(u,v)}(\underline{W}^{(u \rightarrow *)}, \hat{\underline{Y}}_{1:n}^{(*,u)})) & \text{if } (u, v) = \hat{e} \\ f^{(u,v)}(\underline{W}^{(u \rightarrow *)}, \hat{\underline{Y}}_{1:n}^{(*,u)}) & \text{if } u = 2 \end{cases}$$

and

$$\hat{g}^{(u)}(\underline{W}^{(u \rightarrow *)}, \underline{Y}_{1:n}^{(*,u)}) \triangleq \begin{cases} g^{(u)}(\underline{W}^{(u \rightarrow *)}, \underline{Y}_{1:n}^{(*,u)}) & \text{if } u \neq 2 \\ g^{(u)}(\underline{W}^{(u \rightarrow *)}, \hat{\underline{Y}}_{1:n}^{(*,u)}) & \text{if } u = 2. \end{cases}$$

Analysis of error probability: Let  $t = 1, \dots, n$ . Let  $\underline{\mathbf{Z}}_{1:n}^{[\hat{e}]} = (\underline{Z}_{1:n}^{(e)} : e \in E \setminus \{\hat{e}\})$  be the noise values on the edges except the edge  $\hat{e}$  and  $\underline{W} = (\underline{W}^{(u \rightarrow V)} : (u, V) \in \mathcal{M})$  be the messages. Note that for  $t = 1, \dots, n$ ,  $\hat{\underline{X}}_{1:t}^{(\hat{e})}$  is a deterministic function of  $\underline{W}$ ,  $\underline{\mathbf{Z}}_{1:t-1}^{[\hat{e}]}$ , and  $\hat{\underline{Y}}_{1:t-1}^{(\hat{e})}$  while  $\hat{\underline{Y}}_{1:t}^{(\hat{e})}$  is a random variable due to the random design of  $\hat{\mathcal{S}}$ . Let  $\hat{p}_t$  denote the conditional probability distribution of  $\hat{\underline{Y}}_t$  given  $\hat{\underline{X}}_{t-1}$  under a random choice of  $\hat{\mathcal{S}}$  as described above. By Lemma 11, [2],

$$\hat{p}_t(\hat{\underline{y}}_t | \hat{\underline{x}}_{t-1}) \leq \prod_{i=1}^N p_{\hat{e}}(\hat{y}_t(i) | \hat{x}_{t-1}(i)) \cdot 2^{Na(\epsilon, N, t)}.$$

for every  $(\hat{\underline{x}}_t, \hat{\underline{y}}_{t-1})$  such that  $\hat{\underline{x}}_t \in (\mathcal{X}^{(\hat{e})})^N$  and  $(\hat{\underline{x}}_{t-1}, \hat{\underline{y}}_t) \in A_\epsilon^{(N)}(X_{\hat{Q}(\hat{\underline{x}}_{t-1})}, Y_{\hat{Q}(\hat{\underline{x}}_{t-1})})$ . Further, since  $R > C(\hat{e})$ , by standard random coding arguments (e.g. proof of Rate Distortion Theorem, [9]), for large enough  $N$ ,

$$p_t(\{\hat{\underline{y}}_t : (\hat{\underline{y}}_t, \hat{\underline{x}}_{t-1}) \notin A_\epsilon^{(N)}(X_{\hat{Q}(\hat{\underline{x}}_{t-1})}, Y_{\hat{Q}(\hat{\underline{x}}_{t-1})}) | \hat{\underline{x}}_{t-1}\}) \leq \epsilon.$$

Next, note that messages  $\underline{w}$ , noise values  $\underline{\mathbf{z}}^{[\hat{e}]} = (\underline{z}^{(e)} : e \in E \setminus \{\hat{e}\})$ , transmitted vector  $\underline{x}_{1:n}^{(\hat{e})}$ , and received vector  $\underline{y}_{1:n}^{(\hat{e})}$  result in a decoding error under the solution  $\underline{\mathcal{S}}$  if there exists  $\underline{\mathbf{z}}^{(\hat{e})}$  such that  $(\underline{w}, \underline{\mathbf{z}}^{[\hat{e}]}, \underline{\mathbf{z}}^{(\hat{e})}) \in \mathcal{E}(\underline{\mathcal{S}}(\underline{N}))$ , and  $\underline{y}_t^{(\hat{e})} = \Upsilon_{\hat{e}}(\underline{x}_{t-1}^{(\hat{e})}, \underline{z}_{t-1}^{(\hat{e})})$ . Let  $\hat{\mathcal{E}}(\underline{\mathcal{S}}, \underline{w}, \underline{\mathbf{z}}^{[\hat{e}]}) = \{\underline{\mathbf{z}}^{(\hat{e})} : (\underline{w}, \underline{\mathbf{z}}^{[\hat{e}]}, \underline{\mathbf{z}}^{(\hat{e})}) \in \mathcal{E}(\underline{\mathcal{S}}(\underline{N}))\}$ . The expected probability of a decoding error over the choice of  $\hat{\mathcal{S}}(\hat{N}_R)$  for given values of  $\underline{\mathbf{z}}_{1:n}^{[\hat{e}]}$  and  $\underline{w}$  is given by

$$\begin{aligned} & \int_{\hat{\mathcal{E}}(\underline{\mathcal{S}}, \underline{w}, \underline{\mathbf{z}}^{[\hat{e}]})} \left\{ \prod_{t=1}^n \hat{p}_t(\Upsilon_{\hat{e}}(z_{t-1}^{(\hat{e})}) | \hat{\underline{x}}_{t-1}^{(\hat{e})}) \right\} d\underline{\mathbf{z}}_{1:n}^{(\hat{e})} \\ & \leq \int_{\hat{\mathcal{E}}(\underline{\mathcal{S}}, \underline{w}, \underline{\mathbf{z}}^{[\hat{e}]})} \left\{ \prod_{t=1}^n \prod_{i=1}^N p_{\hat{e}}(\Upsilon_{\hat{e}}(z_{t-1}^{(\hat{e})}(i), \hat{x}_{t-1}^{(\hat{e})}(i)) | \hat{x}_{t-1}^{(\hat{e})}(i)) 2^{Na(\epsilon, N, t)} \right\} d\underline{\mathbf{z}}_{1:n}^{(\hat{e})} + \epsilon. \end{aligned}$$

Taking expectation over  $\underline{W}$  and  $\underline{\mathbf{Z}}_{1:n}^{[\hat{e}]}$ ,

$$\begin{aligned} \mathbf{E}_{\hat{\mathcal{S}}, \underline{W}, \underline{\mathbf{Z}}_{1:n}^{[\hat{e}]}} [P_{\mathcal{E}}(\hat{\mathcal{S}})] &= \sum_{\underline{w}} \left[ 2^{-nN \sum_{(u,v) \in \mathcal{M}} R(u,v)} \right. \\ & \left. \int_{\underline{\mathbf{z}} : (\underline{\mathbf{z}}, \underline{w}) \in \mathcal{E}(\underline{\mathcal{S}})} \left\{ \prod_{t=1}^n \prod_{i=1}^N p_{\hat{e}}(\Upsilon_{\hat{e}}(z_{t-1}^{(\hat{e})}(i), \hat{x}_{t-1}^{(\hat{e})}(i)) | \hat{x}_{t-1}^{(\hat{e})}(i)) \right. \right. \\ & \quad \left. \left. 2^{Na(\epsilon, N, t)} P_E(\underline{\mathbf{z}}_{1:n}^{[\hat{e}]}) \right\} d\underline{\mathbf{z}}_{1:n} \right] + \epsilon \\ &= P_{\mathcal{E}}(\underline{\mathcal{S}}) \cdot 2^{nNc(\epsilon, N)} + \epsilon. \end{aligned}$$

Since we assumed that  $\underline{\mathcal{S}}$  is a  $(2^{-N\delta}, \mathbf{R})$  solution, we get

$$\mathbf{E}_{\hat{\mathcal{S}}, \underline{W}, \underline{\mathbf{Z}}_{1:n}^{[\hat{e}]}} [P_{\mathcal{E}}(\hat{\mathcal{S}})] \leq 2^{-N\delta} \cdot 2^{nNa(\epsilon, N, t)} + \epsilon.$$

Finally, for a fixed value of  $n$ , let  $\epsilon < 1/n$ , and choose  $N$  large enough to conclude that  $\mathbf{R} \in \mathcal{R}(\hat{N}_R)$ . ■

## REFERENCES

- [1] R. Koetter, M. Effros, , and M. Médard. On the theory of network equivalence,. In *IEEE Inform. Theory Workshop (ITW)*, 2009.
- [2] M. Effros R. Koetter and M. Médard. A theory of network equivalencepart i: Point-to-point channels. *IEEE Transactions on Information Theory*, 57(2), February 2011.
- [3] S. Kim, T. Ho, M. Effros, and S. Avestimehr. New results on network error correction: Capacities and upper bounds.
- [4] S. Kim, T. Ho, M. Effros, and S. Avestimehr. Network error correction with unequal link capacities. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2009.
- [5] O. Kosut, L. Tong, , and D. Tse. Nonlinear network coding is necessary to combat general byzantine attacks. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2009.

- [6] O. Kosut, L. Tong, , and D. Tse. Polytope codes against adversaries in networks. In *Proceedings of the International Symposium on Information Theory*, Austin, TX, 2010.
- [7] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Medard. Resilient network coding in the presence of byzantine adversaries. In *Proceedings of INFOCOM*, pages 616–624, September 2007.
- [8] R. G. Gallager. *Information Theory and Reliable Communication*. New York: John Wiley, 1968.
- [9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.