Equivalence for Networks with Adversarial State

Oliver Kosut
Department of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ 85287
Email: okosut@asu.edu

Jörg Kliewer
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, NJ 07102
Email: jkliewer@njit.edu

***Abstract*—We address the problem of finding the capacity of networks with independent point-to-point channels where a subset of these channels is replaced either by a compound channel (CC) or an arbitrarily varying channel (AVC). These channels represent a good model for the presence of a Byzantine adversary which controls a subset of links or nodes in the network. We show that equivalence between this network and another network hold in the sense that all links can be replaced by noiseless bit-pipes with the same capacity as the noisy CC or nonsymmetrizable AVC, leading to identical capacity regions for both networks. We then strengthen these results by showing that an additional feedback path between the output and input of a CC or an additional forward path for the AVC extends the equivalent capacity region for both the noisy and the derived noiseless network. This explicitly includes the symmetrizable AVC case.**

I. Introduction

One fundamental problem in wireless and wireline networks is to achieve robustness against active adversaries. A common assumption is to consider Byzantine adversaries who observe all transmissions, messages, and channel noise values and interfere with the transmitted signals, i.e., by replacing a subset of the channel output values or by injecting additional noise to a specific subset of communication channels or nodes (the adversarial set) in the network. For example, for the adversarial noiseless case both in-network error correction approaches and capacity results under network coding have been presented, e.g., in [1]–[4].

The underlying uncertainty in the network due to the action of the adversary leads to channels with varying state in the adversarial set [5]. One possible model is to assume that the corresponding nodes have no knowledge about the exact channel state, but only that the state is selected from a finite set. In the case of a compound channel (CC) [6], [7] the selected state is fixed over the whole transmission of a codeword. In contrast, if the channel state varies from symbol to symbol in an unknown and arbitrary manner we have the case of an arbitrarily varying channel (AVC) [8]–[11].

Note that the AVC either has a (deterministic) capacity with is zero or which equals the random coding capacity [9]. The former case holds for a symmetrizable AVC, since such a channel can mimic a valid input sequence in such a way that it is impossible for the decoder to decide on the correct codeword. Even since in this case transmission is not possible if the AVC is considered in isolation, the situation changes in a network setting, as exemplarily depicted in Fig. 1(a). In this two-node network, source and destination nodes are connected via two parallel channels, a (fixed) channel with capacity C and a symmetrizable AVC. Here, communication over the AVC is possible with a non-zero rate since common randomness with negligible rate $\epsilon > 0$ can be shared between both nodes [9]–[11] via the upper channel in Fig. 1(a). In a more general setup, in Fig. 1 this channel can be replaced with a single-source single-sink network of positive rate C.

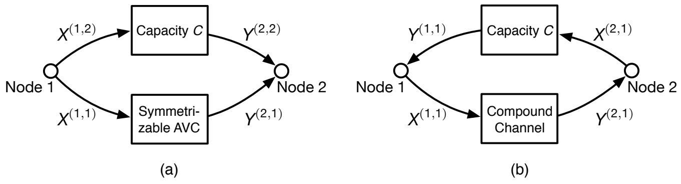

Fig. 1. Two-node networks with a capacity C channel and (a) a symmetrizable AVC, (b) a CC. In general, the upper channel can be replaced with a single-source single-sink network having the same rate.

In the following we consider the problem of reliable communication over a network of independent noisy point-to-point channels in the presence of active adversaries. A subset of the channels either consists of AVCs or CCs. This is in contrast to the model in [12], where the action of the adversary is directly modeled by injecting an arbitrary vector to the network edges in the adversarial set. By building on the results in [13] we identify cases where the adversarial capacity of the network equals the capacity of another network in which each channel is replaced by a noise-free bit-pipe. For a CC, the bit-pipe has capacity equal to the standard CC capacity if there is no feedback path from the output to the input; if there is, then the equivalent bit-pipe has higher capacity, because the state can be estimated at the output and relayed back to the input (see Fig. 1(b)). For an AVC, if it is non-symmetrizable, or there is a forward path as in Fig. 1(a), the equivalent bit-pipe has capacity equal to the random coding capacity. Otherwise, it appears to be difficult to obtain an equivalence result, as the strong converse does not hold for symmetrizable AVCs.

II. Model

Consider a network of nodes $\mathcal{V} := \{1, \ldots, m\}$ with state, given by

$$\mathcal{N} = \left(\prod_{v=1}^{m} \mathcal{X}^{(v)}, \mathcal{S}, p(\mathbf{y}|\mathbf{x}, s), \prod_{v=1}^{m} \mathcal{Y}^{(v)} \right). \tag{1}$$

Herein, $\mathcal{X}^{(v)}$ and $\mathcal{Y}^{(v)}$ denote the input and output alphabets of the node v and $\mathcal{S}$ the set of network states, respectively. This network may represent either a CC or an AVC model.

This work was supported in part by the U.S. National Science Foundation under grants CCF-1439465 and CCF-1440014.

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

These both assume that the state is chosen not randomly but adversarially; in the CC model the adversary chooses a single state $s \in \mathcal{S}$ that remains constant throughout the code block, whereas in the AVC model the adversary chooses an arbitrary state sequence $s^n \in \mathcal{S}^n$. In this paper we are interested in both problems, but only one at a time. Studying networks with both CC-type state and AVC-type state is beyond our scope. For tractability we assume that the overall set of network states decomposes into a product of the set of states for the link being replaced by a bit pipe and the set of states for the rest of the network.

In general, CCs and AVCs can be quite pathological, so we assume that alphabets $\mathcal{X}^{(v)}$, $\mathcal{S}$, and $\mathcal{Y}^{(v)}$ are all finite sets. Most of our results apply for more general alphabets under mild regularity conditions, but to avoid edge cases and complications we restrict ourselves to finite alphabets. We believe that the interesting consequences of the CC and AVC network models are captured with finite alphabets models, and that the complications that arise for general alphabets are unlikely to make a difference in practice.

Notation: Let $[k] = \{1, \ldots, k\}$. A rate vector $\mathcal{R}$ consists of multicast rates $R^{(\{v\} \to U)}$ from each source node v to each destination set $U \subseteq \mathcal{V}$. With a singleton destination set $U = \{u\}$, we sometimes write simply $R^{(v \to u)}$. For each (v, U) pair, there is a message $W^{(\{v\} \to U)} \in \mathcal{W}^{(\{v\} \to U)} = [2^{nR^{(\{v\} \to U)}}]$. Let $W^{(V \to *)}$ denote the vector of all messages originating at nodes $v \in V$, and let $\mathcal{W}^{(V \to *)}$ denote the corresponding message set. Also let W denote the vector of all messages.

A blocklength-n solution $\mathsf{S}(\mathcal{N})$ for network $\mathcal{N}$ consists of a set of causal encoding functions

$$X_t^{(v)} : (\mathcal{Y}^{(v)})^{t-1} \times \mathcal{W}^{(\{v\} \to *)} \to \mathcal{X}^{(v)} \tag{2}$$

for each $v \in \mathcal{V}$ and $t \in [n]$, and decoding functions

$$\widehat{W}^{(\{v\} \to U),u} : (\mathcal{Y}^{(u)})^n \times \mathcal{W}^{(\{u\} \to *)} \to \mathcal{W}^{(\{v\} \to U)} \tag{3}$$

for each (v, U) pair and each $u \in U$. Let $\widehat{W}$ be the complete vector of message estimates, and denote by $\{\widehat{W} \neq W\}$ the event that at least one message is incorrectly decoded. Note that the probability of this event depends on the state sequence S^n.

Definition 1. *The CC-capacity region $\mathscr{R}_{\text{CC}}(\mathcal{N})$ of network $\mathcal{N}$ is given by the closure of the set of rate vectors $\mathcal{R}$ for which there exists a sequence of blocklength-n solutions for which*

$$\max_{s \in \mathcal{S}} \Pr(\widehat{W} \neq W | S^n = (s, s, \ldots, s)) \to 0. \tag{4}$$

Definition 2. *The AVC-capacity region $\mathscr{R}_{\text{AVC}}(\mathcal{N})$ of network $\mathcal{N}$ is given by the closure of the set of rate vectors $\mathcal{R}$ for which there exists a sequence of blocklength-n solutions for which*

$$\max_{s^n \in \mathcal{S}^n} \Pr(\widehat{W} \neq W | S^n = s^n) \to 0. \tag{5}$$

We are especially interested in the case that there is an independent point-to-point channel from node 1 to node 2 with independent state. That is, $\mathcal{X}^{(1)} = \mathcal{X}^{(1,0)} \times \mathcal{X}^{(1,1)}$, $\mathcal{Y}^{(2)} = \mathcal{Y}^{(2,0)} \times \mathcal{Y}^{(2,1)}$, $\mathcal{S} = \mathcal{S}^{(0)} \times \mathcal{S}^{(1)}$, and

$$p(\mathbf{y}|\mathbf{x}, s) = p(\mathbf{y}^{(0)}|\mathbf{x}^{(0)}, s^{(0)}) p(y^{(2,1)}|x^{(1,1)}, s^{(1)}) \tag{6}$$

where $x^{(1,1)} \in \mathcal{X}^{(1,1)}$, $y^{(2,1)} \in \mathcal{Y}^{(2,1)}$, and $s^{(1)} \in \mathcal{S}^{(1)}$ represent the input, output, and state respectively for the point-to-point channel, and $\mathbf{x}^{(0)} \in \mathcal{X}^{(1,0)} \times \prod_{v \neq 1} \mathcal{X}^{(v)}$, $\mathbf{y}^{(0)} \in \mathcal{Y}^{(2,0)} \times \prod_{v \neq 2} \mathcal{Y}^{(v)}$, and $s^{(0)} \in \mathcal{S}^{(0)}$ represent the input, output, and state respectively for the remainder of the network. The point-to-point channel itself is given by

$$\mathcal{C} = (\mathcal{X}^{(1,1)}, \mathcal{S}^{(1)}, p(y^{(2,1)}|x^{(1,1)}, s^{(1)}), \mathcal{Y}^{(2,1)}). \tag{7}$$

We also consider the network $\mathcal{N}^R$ for any $R \geq 0$ in which the noisy point-to-point channel $\mathcal{C}$ is replaced by a rate-R noiseless (and state-less) bit-pipe $\mathcal{C}^R$. By convention, for non-integer R, with n uses $\mathcal{C}^R$ can transmit $\lfloor nR \rfloor$ bits.

Our goal is to prove achievability-type results of the form $\mathscr{R}(\mathcal{N}^R) \subseteq \mathscr{R}(\mathcal{N})$ and converse-type results of the form $\mathscr{R}(\mathcal{N}) \subseteq \mathscr{R}(\mathcal{N}^R)$ for both CC and AVC models.

We adopt the notion from [13] of *stacked networks*, wherein we denote by $\underline{\mathcal{N}}$ a network with N independent copies of the network $\mathcal{N}$. Note that for the CC model the states for the different copies are the same. Underlines denote stacked variables and vectors, and the argument ℓ refers to layer ℓ; e.g., $\underline{X}^{(v)}(\ell)$ is the symbol sent by node v in layer ℓ, and $\underline{X}^{(v)} := (\underline{X}^{(v)}(\ell) : \ell \in [N])$. We state two preliminary lemmas, which are simple extensions of Lemmas 1 and 4 respectively from [13] to include state. The proofs can be found in the extended version of this paper [14].

Lemma 1. *For any network $\mathcal{N}$, $\mathscr{R}_{\text{CC}}(\mathcal{N}) = \mathscr{R}_{\text{CC}}(\underline{\mathcal{N}})$ and $\mathscr{R}_{\text{AVC}}(\mathcal{N}) = \mathscr{R}_{\text{AVC}}(\underline{\mathcal{N}})$.*

Lemma 2. *The capacity regions $\mathscr{R}_{\text{CC}}(\mathcal{N}^R)$ and $\mathscr{R}_{\text{AVC}}(\mathcal{N}^R)$ are continuous in R for all $R > 0$.*

III. Positive Rate Regions

For both CC and AVC models, it will be important to know whether any information at all can be sent between nodes. This positive (but arbitrarily small) rate will be used for feedback in the CC model and for feed-forward in the AVC model (see Fig. 1). Thus in this section we investigate the set of node pairs (u, v) for which positive rate can be sent from u to v. We do this first without state[1], and then extend it for the CC and AVC models. We form a set $\mathcal{P} \subset \mathcal{V} \times \mathcal{V}$ and subsequently show that $\mathcal{P}$ is precisely the set of node pairs that can sustain positive rate. The set $\mathcal{P}$ is formed as follows:

1) Initialize $\mathcal{P}$ as $\{(u, u) : u \in \mathcal{V}\}$.
2) If there is a pair of nodes $(u, v) \notin \mathcal{P}$, node i such that $(u, i) \in \mathcal{P}$, and set $\mathcal{A} \subset \mathcal{V}$ such that $(j, v) \in \mathcal{P}$ for all $j \in \mathcal{A}$, and
$$\max_{p(x^{(i)}), x^{(\{i\}^c)}} I(X^{(i)}; Y^{(\mathcal{A})} | X^{(\{i\}^c)} = x^{(\{i\}^c)}) > 0, \tag{8}$$
then add (u, v) to $\mathcal{P}$.
3) Repeat step 2 until there are no additional such pairs (u, v).

Theorem 3. *If $(u, v) \in \mathcal{P}$, then there exists an $\mathcal{R} \in \mathscr{R}(\mathcal{N})$ with $R^{(u \to v)} > 0$.*

Proof: A node may trivially send arbitrary amounts of information to itself; thus $R^{(u \to u)} > 0$ is achievable for any

[1] That is, $\mathcal{S}$ contains only a single element, in which case $\mathscr{R}_{\text{CC}}(\mathcal{N}) = \mathscr{R}_{\text{AVC}}(\mathcal{N})$, and we denote both by $\mathscr{R}(\mathcal{N})$.

$u \in \mathcal{V}$. We proceed by induction to prove the theorem for pairs $(u,v) \in \mathcal{P}$ with $u \neq v$. Consider the specific step in the construction of $\mathcal{P}$ at which (u,v) is added, and let $i, \mathcal{A}$ satisfy (8). We assume that positive rate can be sent from u to i, and that for all $j \in \mathcal{A}$, positive rate can be sent from j to v. The basic idea to send positive rate from u to v is to employ a point-to-point channel code from $X^{(i)}$ to $Y^{(\mathcal{A})}$. A codeword is chosen at node u, then conveyed to node i using a positive-rate solution. Then the codeword is transmitted by node i and received by nodes in $\mathcal{A}$. Next, the received sequences are transmitted from nodes in $\mathcal{A}$ to node v using positive-rate solutions. Finally, node v decodes the point-to-point code (see [14] for details). ■

The following theorem gives the converse result, stating that if $(u,v) \notin \mathcal{P}$, then values received at node v are conditionally independent of values sent from node u given messages that originate outside node u. This indicates that all information known at node v originates outside of node u; i.e., the input at node u cannot influence the output at node v. This is a much stronger statement than a simple converse, and indeed even stronger than traditional "strong" converses, but it is necessary to prove equivalence results.

Theorem 4. *If $(u,v) \notin \mathcal{P}$, then for any solution $\mathsf{S}(\mathcal{N})$, $X_{1:n}^{(u)} \to W^{(\{u\}^c \to *)} \to Y_{1:n}^{(v)}$ forms a Markov chain.*

Proof: Let $\mathcal{A} := \{i : (i,v) \in \mathcal{P}\}$. By the definition of $\mathcal{P}$, for any $i \notin \mathcal{A}$,

$$\max_{p(x^{(\{i\})}),x^{(\{i\}^c)}} I(X^{(\{i\})};Y^{(\mathcal{A})}|X^{(\{i\}^c)} = x^{(\{i\}^c)}) = 0. \quad (9)$$

As this holds for all $i \notin \mathcal{A}$, we conclude that for any solution $\mathsf{S}(\mathcal{N})$, we have the Markov chain $X_t^{(\mathcal{A}^c)} \to X_t^{(\mathcal{A})} \to Y_t^{(\mathcal{A})}$ for each time t. It can be shown that combining this with the basic dependency requirements in the problem yields the chain $X_{1:n}^{(\mathcal{A}^c)} \to W^{(\mathcal{A}\to *)} \to Y_{1:n}^{(\mathcal{A})}$. This completes the proof since $v \in \mathcal{A}$ and $u \in \mathcal{A}^c$. ■

We now extend the above results for CC- and AVC-type states. For each $s \in \mathcal{S}$, define $\mathcal{P}_s$ as above for $\mathcal{P}$, but with fixed state $S = s$. Let $\mathcal{P}_{\text{CC}} = \bigcap_{s\in\mathcal{S}} \mathcal{P}_s$.

Theorem 5. *Under the CC model, if $(u,v) \in \mathcal{P}_{\text{CC}}$ there exists a rate vector $\mathcal{R} \in \mathscr{R}_{\text{CC}}(\mathcal{N})$ with $R^{(u\to v)} > 0$. Conversely, if $(u,v) \notin \mathcal{P}_{\text{CC}}$, then for any solution $\mathsf{S}(\mathcal{N})$ there exists $s \in \mathcal{S}$ such that with $S^n = (s, s, \ldots, s)$, $X_{1:n}^{(u)} \to W^{(\{u\}^c\to *)} \to Y_{1:n}^{(v)}$ forms a Markov chain.*

Proof: Achievability is proved by constructing a solution with $|\mathcal{S}|$ sessions, one for each $s \in \mathcal{S}$. In the session for s, a code is employed to send positive rate from u to v assuming $S = s$. Such a code exists by Theorem 3 and the fact that if $(u,v) \in \mathcal{P}_{\text{CC}}$, then $(u,v) \in \mathcal{P}_s$. The only difficulty is in node v determining which of the states is the true one. This can be accomplished by, in the inductive construction of the codes for each s, sending training sequences so that the receiving nodes can estimate the state. A similar technique will be used in the proof of Lemma 9. Further details can be found in [14].

To prove the converse, note that if $(u,v) \notin \mathcal{P}_{\text{CC}}$ then $(u,v) \notin \mathcal{P}_s$ for some $s \in \mathcal{S}$. With this fixed state, the proof follows exactly as that of Theorem 4. ■

Define $\mathcal{P}_{\text{AVC}}$ using the same procedure as above for $\mathcal{P}$, but replace (8) with the condition that there exists $x^{(\{i\}^c)} \in \mathcal{X}^{(\{i\}^c)}$ such that the channel from $X^{(i)}$ to $Y^{(\mathcal{A})}$, conditioned on $X^{(\{i\}^c)} = x^{(\{i\}^c)}$, has positive rate which is equivalent to being non-symmetrizable as defined in [10].

Theorem 6. *If $(u,v) \in \mathcal{P}_{\text{AVC}}$, then there exists a rate vector $\mathcal{R} \in \mathscr{R}_{\text{AVC}}(\mathcal{N})$ with $R^{(u\to v)} > 0$.*

Proof: The proof follows from the same argument as for Theorem 3, except that we replace the point-to-point channel code from $X^{(i)}$ to $Y^{(\mathcal{A})}$ with an AVC code. By the assumption that this channel is non-symmetrizable, positive rate can be achieved by the results in [10]. ■

IV. Compound Channel Equivalence

In this section and the next we simplify notation by writing X for $X^{(1,1)}$, Y for $Y^{(2,1)}$, and S for $S^{(1)}$. Since we are primarily interested in the independent channel $\mathcal{C}$, there should be no confusion.

There are two relevant capacities for the compound channel: first, the standard capacity expression for a compound channel

$$\underline{C} = \max_{p(x)} \min_{s\in\mathcal{S}} I(X;Y|S=s), \quad (10)$$

and second, the capacity of a compound channel if the state is known at the encoder and the decoder, wherein the min and max are reversed:

$$\bar{C} = \min_{s\in\mathcal{S}} \max_{p(x)} I(X;Y|S=s). \quad (11)$$

Of course, $\underline{C} \leq \bar{C}$. Let $\mathcal{P}_{\text{CC}}$ be defined as above for $\mathcal{N}$. As stated in the following theorem, the compound channel is equivalent to a bit-pipe with rate either $\underline{C}$ or $\bar{C}$, depending on whether the rest of the network can sustain any positive feedback rate from node 2 to node 1.

Theorem 7.

$$\mathscr{R}_{\text{CC}}(\mathcal{N}) = \begin{cases} \mathscr{R}_{\text{CC}}(\mathcal{N}^{\bar{C}}) & \textit{if } (2,1) \in \mathcal{P}_{\text{CC}} \\ \mathscr{R}_{\text{CC}}(\mathcal{N}^{\underline{C}}) & \textit{if } (2,1) \notin \mathcal{P}_{\text{CC}}. \end{cases} \quad (12)$$

We prove this theorem in several lemmas, which in combination with continuity from Lemma 2 prove the theorem.

Lemma 8. *If $R < \underline{C}$, then $\mathscr{R}_{\text{CC}}(\mathcal{N}^R) \subseteq \mathscr{R}_{\text{CC}}(\mathcal{N})$.*

Proof: The proof follows from the standard achievability argument for the compound channel and Lemma 3 from [15], which is proved for channels without state but applies equally well with state. ■

Lemma 9. *If $R > \bar{C}$, then $\mathscr{R}_{\text{CC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{CC}}(\mathcal{N}^R)$.*

Proof: Let $s^* = \arg\min_s \max_{p(x)} I(X;Y)$. We may use Theorem 4 in [13] to simulate the channel $p(y|x,s^*)$ over the bit-pipe of rate R, since $R > I(X;Y)$ for this channel and any input distribution. ■

Lemma 10. *If $(2,1) \in \mathcal{P}_{\text{CC}}$ and $R < \bar{C}$, then $\mathscr{R}_{\text{CC}}(\mathcal{N}^R) \subseteq \mathscr{R}_{\text{CC}}(\mathcal{N})$.*

Proof: By Theorem 3, since $(2,1) \in \mathcal{P}_{\text{CC}}$, there exists a solution $\mathsf{S}_0(\mathcal{N})$ such that $R^{(2\to 1)} > 0$. Given a solution $\mathsf{S}(\mathcal{N}^R)$, we construct a solution $\mathsf{S}(\mathcal{N})$ with three sessions of blocklength n_1, n_2, n_3 respectively. In the first session node

1 sends an agreed-upon training sequence so that node 2 can estimate the state S of the channel. In the second session, this estimated state is transmitted to node 1 using $\mathsf{S}_0(\mathcal{N})$. In the third session, node 1 uses this estimated state to transmit a message across the compound channel while the rest of $\mathsf{S}(\mathcal{N}^R)$ is conducted. We give more details as follows.

Fix any $\lambda > 0$. Let $\epsilon = \frac{\bar{C}-R}{2}$. Note that $\epsilon > 0$, and that for the true state s

$$R + 2\epsilon \le \max_{p(x)} I(X;Y|S=s). \tag{13}$$

For the first session we employ a random coding argument, wherein we choose a training sequence $\alpha_{1:n_1}$ randomly and uniformly from $\mathcal{X}^{n_1}$. This single sequence forms the "codebook" for the first session, and it is revealed to both nodes 1 and 2. In the first session node 1 transmits $\alpha_{1:n_1}$ into the compound channel while the inputs to all other channels are arbitrary. Let $Y_{1:n_1}$ be the output of the compound channel. Node 2 forms the maximum likelihood estimate for the channel state as

$$\hat{S} := \arg\min_{\hat{s}\in\mathcal{S}} \frac{1}{n}\sum_{t=1}^{n_1} -\log p(Y_t|\alpha_t,\hat{s}). \tag{14}$$

In the second session, solution $\mathsf{S}_0(\mathcal{N})$ is employed with blocklength n_2 to transmit $\hat{S}$ from node 2 to node 1. Let $\check{S}$ be the recovered value at node 1.

Let $\check{p}(x)$ be an optimal input distribution for the channel with state $\check{S}$. In the third session, while the rest of the network conducts $\mathsf{S}(\mathcal{N}^R)$, nodes 1 and 2 employ an n_3-length point-to-point code with input distribution $\check{p}(x)$ and channel transition matrix conditioned on $\check{S}$. Let $W^{(1\to 2)}$ be one of $2^{(n_1+n_2+n_3)R}$ messages, and let $\widehat{W}^{(1\to 2)}$ be the estimated message.

We define the following error events:

$$\mathcal{E}_1 := \left\{p(y|x,s) \neq p(y|x,\hat{S}) \text{ for any } x,y\right\} \tag{15}$$

$$\mathcal{E}_2 := \{\check{S} \neq \hat{S}\}, \quad \mathcal{E}_3 := \{\widehat{W}^{(i\to j)} \neq W^{(i\to j)}\}. \tag{16}$$

Note that $\mathcal{E}_3$ is precisely the overall error event, but we first bound the probability of event $\mathcal{E}_1$. When averaging over the random choice of $\alpha_{1:n_1}$ and the random operation of the code under state s, for any $\hat{s} \in \mathcal{S}$ the quantities $-\log p(Y_t|\alpha_t,\hat{s})$ are i.i.d. with expected value

$$\begin{aligned}&\sum_{x,y} -\frac{1}{|\mathcal{X}|} p(y|x,s)\log p(y|x,\hat{s}) \\ &= H(Y|X,S=s) + \sum_{x} \frac{1}{|\mathcal{X}|} D\big(p(y|x,s)\|p(y|x,\hat{s})\big). \end{aligned} \tag{17}$$

Thus, by the law of large numbers, for sufficiently large n_1, $\Pr(\mathcal{E}_1) \le \lambda$. It follows from the existence of solution $\mathsf{S}_0(\mathcal{N})$ that for sufficiently large n_2, $\Pr(\mathcal{E}_2) \le \lambda$. Now assume that neither $\mathcal{E}_1$ nor $\mathcal{E}_2$ occur, so $\check{S} = \hat{S} = \hat{s}$ such that $p(y|x,s) = p(y|x,\hat{s})$. Hence

$$I_{\check{p}}(X;Y|S=s) = \max_{p(x)} I(X;Y|S=s) \ge R + 2\epsilon. \tag{18}$$

Take n_3 to be sufficiently large so that $(n_1+n_2+n_3)R \le n_3(R+\epsilon)$. By (18), $R + \epsilon < I_{p_{\check{s}}}(X;Y|S=s)$, so for sufficiently large n_3 the point-to-point code used in the third session, having $2^{(n_1+n_2+n_3)R}$ messages, has probability of error at most λ. That is, $\Pr(\mathcal{E}_3|(\mathcal{E}_1\cup\mathcal{E}_2)^c) \le \lambda$. Applying the union bound gives $\Pr(\mathcal{E}_3) \le \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_3|(\mathcal{E}_1\cup\mathcal{E}_2)^c) \le 3\lambda$. ■

Lemma 11. *If $(2,1) \notin \mathcal{P}_{\text{CC}}$ and $R > \underline{C}$, then $\mathscr{R}_{\text{CC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{CC}}(\mathcal{N}^R)$.*

Proof: By Lemma 1 it suffices to show that $\mathscr{R}_{\text{CC}}(\underline{\mathcal{N}}) \subseteq \mathscr{R}_{\text{CC}}(\underline{\mathcal{N}}_R)$. Fix any $\mathcal{R} \in \text{int}(\mathscr{R}_{\text{CC}}(\mathcal{N}))$ and $\lambda > 0$.

Choose code and define distributions: Let $\mathsf{S}(\mathcal{N})$ be a rate-$\mathcal{R}$ solution on network $\mathcal{N}$ for some blocklength n. By Theorem 5, for solution $\mathsf{S}(\mathcal{N})$, $X_{1:n}^{(2)} \to W^{(\{2\}^c\to *)} \to Y_{1:n}^{(1)}$ forms a Markov chain. Moreover, the state S only has direct impact on $Y_{1:n}^{(2)}$, which in turn only has direct impact on $X_{1:n}^{(2)}$. Thus $S \to X_{1:n}^{(2)} \to (W^{(\{2\}^c\to *)}, Y_{1:n}^{(1)})$ forms a Markov chain.[2] Combing these two chains yields

$$S \to X_{1:n}^{(2)} \to W^{(\{2\}^c\to *)} \to Y_{1:n}^{(1)}. \tag{19}$$

Since $W^{(\{2\}^c\to *)}$ is drawn uniformly from $\mathcal{W}^{(\{2\}^c\to *)}$ and independently from S, the distribution of $(W^{(\{2\}^c\to *)}, Y_{1:n}^{(1)})$ does not depend on S. Thus the distribution of $X_{1:n}^{(1)}$ also does not depend on S, as it is a function of $(W^{(\{1\}\to *)}, Y_{1:n}^{(1)})$. Therefore for each time t we may define $p_t(x)$ to be the distribution of $X_t^{(1)}$ independent of S. Let $p(x) = \frac{1}{n}\sum_{t=1}^{n} p_t(x)$ and let

$$s^* = \arg\min_{s\in\mathcal{S}} I(X;Y|S=s) \tag{20}$$

where X is drawn from $p(x)$. Let $p_t(x,y) = p_t(x)p(y|x,s^*)$.

Typical set: Define $\widehat{A}_{\epsilon,t}^{(N)}$ to be the N-length typical set according to distribution $p_t(x,y)$ as in [13].

Design of channel emulators: By concavity of mutual information with respect to the input variable,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} I(X_t;Y_t|S=s^*) &\le I(X;Y|S=s^*) \\ &= \min_{s} I(X;Y|S=s) \le \underline{C} < R. \end{aligned} \tag{21}$$

Let $R_t := I(X_t;Y_t|S=s^*) + \Delta$ where $\Delta > 0$ is chosen so that $\frac{1}{n}\sum_{t=1}^{n} R_t = R$.

Randomly design decoder $\beta_{N,t} : [2^{NR_t}] \to \underline{\mathcal{Y}}$ by drawing codewords $\beta_{N,t}(1), \dots, \beta_{N,t}(2^{NR_t})$ from the i.i.d. distribution with marginal $p_t(\underline{y})$. Define encoder $\alpha_{N,t} : \underline{\mathcal{X}} \to [2^{NR_t}]$ as

$$\alpha_{N,t}(\underline{x}) = \begin{cases} k & \text{if } (\underline{x}, \beta_{N,t}(k)) \in \widehat{A}_{\epsilon,t}^{(N)} \\ 1 & \text{if } \nexists k \text{ s.t. } (\underline{x}, \beta_{N,t}(k)) \in \widehat{A}_{\epsilon,t}^{(N)}. \end{cases} \tag{22}$$

Note that the number of bits required to send $(\alpha_{N,t}(\underline{X}))_{t=1}^{n}$ is $\sum_{t=1}^{n} NR_t = nNR$, so we may send all these encoded functions via a bit-pipe of rate R.

The rest of the proof follows essentially that of Theorem 6 in [13]. This involves creating a stacked solution for $\underline{\mathcal{N}}$ with exponentially decreasing probability of error, and then converting it into a solution for $\underline{\mathcal{N}}^R$ by employing the channel

[2] We have written S as a random variable even though it is arbitrary rather than random. By $S \to A \to B$ we mean that $p(b|a,s) = p(b|a)$.

emulators at nodes 1 and 2 to simulate the noisy channel over the rate-R bit pipe. Finally, the error probability can be bounded provided correct parameters are chosen for the typical set $\widehat{A}_{\epsilon,t}^{(N)}$, which can be done for our problem by virtue of the fact that $R_t - I(X_t; Y_t|S = s^*) = \Delta > 0$. ■

V. Arbitrarily Varying Channel Equivalence

The random coding capacity of a point-to-point AVC is defined as the maximum rate that can be achieved if the encoder and decoder have access to shared randomness (inaccessible to the adversary). It is given by

$$C_r = \max_{p(x)} \min_{p(s)} I(X; Y). \tag{23}$$

Moreover, the max and min may be interchanged without changing the quantity, because of the convexity properties of the mutual information. Without shared randomness, as shown in [10], the capacity of an AVC is 0 if the channel is symmetrizable, and otherwise the capacity is C_r. Thus, in all cases, C_r is an upper bound on the capacity. The following theorem provides the corresponding network-level converse.

Theorem 12. $\mathscr{R}_{\text{AVC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{AVC}}(\mathcal{N}^{C_r})$.

Proof: By the continuity property from Lemma 2, it will be enough to show that $\mathscr{R}_{\text{AVC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{AVC}}(\mathcal{N}^R)$ for all $R > C_r$. Let

$$p^\star(s) := \arg\min_{p(s)} \max_{p(x)} I(X; Y). \tag{24}$$

Let $p^\star(y|x) = \sum_s p^\star(s)p(y|x, s)$. Note that C_r is the capacity of the ordinary channel with transition matrix $p^\star(y|x)$. Since any code on $\mathcal{N}$ must achieve small probability of error for any choice of s^n, it also achieves small probability of error for any random choice of s^n, provided this random choice is independent of the choice of message. In particular, the code works for S^n drawn i.i.d. from $p^\star(s)$. Thus it works if the AVC is replaced by the ordinary channel $p^\star(y|x)$. Now the proof is completed by Theorem 6 of [13]. ■

Theorem 12.11 from [11] states that the capacity of a point-to-point AVC is either 0 or C_r. This is shown by proving that a small header can be transmitted from encoder to decoder that allows the encoder and decoder to simulate common randomness. This small header can be sent using any code that achieves positive rate. The following is an extension of this result to the network setting.

Theorem 13. *If there exists a rate vector $\mathcal{R} \in \mathscr{R}_{\text{AVC}}(\mathcal{N})$ with $R^{(1\to 2)} > 0$, then $\mathscr{R}_{\text{AVC}}(\mathcal{N}) = \mathscr{R}_{\text{AVC}}(\mathcal{N}^{C_r})$.*

Before proving the theorem, we state the following lemma, asserting that C_r can be achieved with a random code that requires a relatively small amount of shared randomness between encoder and decoder. This lemma is a simple combination of Lemmas 12.8 and 12.10 from [11].

Lemma 14. *Given an AVC $p(y|x, s)$, for any $R < C_r$ and $\epsilon > 0$, for any integer K satisfying*

$$K \geq \frac{2n}{\epsilon}(R + \log|\mathcal{S}|). \tag{25}$$

there exist $(n, 2^{nR})$ channel codes (f_ℓ, ϕ_ℓ) for $\ell = 1, \dots, K$ consisting of functions $f_\ell : [2^{nR}] \to \mathcal{X}^n$ and $\phi_\ell : \mathcal{Y}^n \to [2^{nR}]$ such that

$$\max_{m\in[2^{nR}]} \max_{s^n\in\mathcal{S}^n} \frac{1}{K} \sum_{\ell=1}^{K} p^n(\phi_\ell^{-1}(m)^c | f_\ell(m), s^n) \leq \epsilon. \tag{26}$$

Proof of Theorem 13: In light of Theorem 12 and Lemma 2, we have only to prove that $\mathscr{R}(\mathcal{N}^R) \subseteq \mathscr{R}(\mathcal{N})$ for all $R < C_r$. The proof of this follows essentially from the same argument as the proof of Theorem 12.11 from [11]. Fix $\epsilon > 0$ and $R < C_r$. Choose integer K to satisfy (25). By Lemma 14 there exists K channel codes (f_ℓ, ϕ_ℓ) satisfying (26).

Let $\mathsf{S}_0(\mathcal{N})$ be a solution with $R^{(1\to 2)} > 0$. Coding proceeds in two sessions. In the first session node 1 chooses one of the K channel codes to use, and transmits it to node 2 using $\mathsf{S}_0(\mathcal{N})$. Since in this solution $R^{(1\to 2)} > 0$, for sufficiently large blocklength the probability of error for this session can be made arbitrarily small. In the second session nodes 1 and 2 employ whichever channel code was selected in the first session. Since the selection of the channel code is random, by (26), on average the probability of error for the second session is bounded by ϵ. ■

The following corollary provides a sufficient condition for equivalence for the AVC. It follows immediately from Theorem 6 and Theorem 13.

Corollary 15. *If $(1, 2) \in \mathcal{P}_{\text{AVC}}$, then $\mathscr{R}_{\text{AVC}}(\mathcal{N}) = \mathscr{R}_{\text{AVC}}(\mathcal{N}^{C_r})$.*

References

[1] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Medard, and M. Effros, "Resilient network coding in the presence of Byzantine adversaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, Jun. 2008.

[2] S. Kim, T. Ho, M. Effros, and S. Avestimehr, "Network error correction with unequal link capacities," in *Proc. 47th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2009, pp. 1387–1394.

[3] O. Kosut, L. Tong, and D. Tse, "Nonlinear network coding is necessary to combat general Byzantine attacks," in *Proc. 47th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2009, pp. 593–599.

[4] ——, "Polytope codes against adversaries in networks," in *Proc. IEEE Int. Sympos. on Inform. Theory*, Austin, TX, Jun. 2010, pp. 2423–2427.

[5] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[6] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[7] J. Wolfowitz, *Coding Theorems of Information Theory*. Springer Verlag, 1978.

[8] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, 1960.

[9] R. Ahlswede, "Elemination of correlation in random codes for arbitrarily varing channels," *Probability Theory and Related Fields*, vol. 44, no. 2, pp. 159–175, 1978.

[10] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: positivity, constraints," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, 1988.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[12] M. Bakshi, M. Effros, and T. Ho, "On equivalence for networks of noisy channels under Byzantine attacks," in *Proc. IEEE Int. Sympos. on Inform. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 973–977.

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

[14] O. Kosut and J. Kliewer, "Equivalence for networks with adversarial state," Apr. 2014, arXiv.org: 1404.6701 [cs.IT].

[15] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence— Part II: Multiterminal channels," 2010, arXiv.org:1007.1033.