

A Generalized Cut-Set Bound

Amin Aminzadeh Gohari
EECS Department
University of California
Berkeley, CA 94720, USA
Email: aminzade@eecs.berkeley.edu

Venkat Anantharam
EECS Department
University of California
Berkeley, CA 94720, USA
Email: ananth@eecs.berkeley.edu

Abstract—In this paper, we generalize the well known cut-set bound (see for example [1, p. 444]) to the problem of lossy transmission of functions of arbitrarily correlated sources over a discrete memoryless multiterminal network.

I. INTRODUCTION

A general multiterminal network is a model for reliable communication of sets of messages among the nodes of a network, and has been extensively used in modeling of wireless systems. It is known that unlike the point-to-point scenario, in a network scenario the separation of the source and channel codings is not necessarily optimal [4]. In this paper we study the limitations of joint source-channel coding strategies for lossy transmission across multiterminal networks.

A discrete memoryless general multiterminal network (GMN) is characterized by the conditional distribution

$$q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)}),$$

where $X^{(i)}$ and $Y^{(i)}$ ($1 \leq i \leq m$) are respectively the input and the output of the channel at the i^{th} party. In a general multiterminal channel with correlated sources, the m nodes are observing i.i.d. repetitions of m , possibly correlated, random variables $W^{(i)}$ for $1 \leq i \leq m$. The i^{th} party ($1 \leq i \leq m$) has access to the i.i.d. repetitions of $W^{(i)}$, and wants to reconstruct, within a given distortion, the i.i.d. repetitions of a function of all the observations, i.e. $f^{(i)}(W^{(1)}, W^{(2)}, \dots, W^{(m)})$ for some function $f^{(i)}(\cdot)$. If this is asymptotically possible within a given distortion (see section II for a formal definition), we call the source $(W^{(1)}, W^{(2)}, \dots, W^{(m)})$ admissible. In some applications, each party may be interested in recovering i.i.d. repetitions of functions of the observations made at different nodes. In this case the function $f^{(i)}(W^{(1)}, W^{(2)}, \dots, W^{(m)})$ takes the special form of $(f^{(i,1)}(W^{(1)}), f^{(i,2)}(W^{(2)}), \dots, f^{(i,m)}(W^{(m)}))$ for some functions $f^{(i,j)}(\cdot)$.

The admissible source region of a general multiterminal network is not known when the sources are independent except in certain special cases; less is known when the sources are allowed to be arbitrarily correlated. It is known that the source–channel separation theorem in a network scenario breaks down [4]. In this paper, we prove a new outer bound on the admissible source region of GMNs that as a special case reduces to the well known cut-set bound. Other extensions of cut-set bound can be found in [2] and [5]. Furthermore some

existing works show the possibility and benefit of function computation during the communication (see for instance [3], [6], [7]).

A main contribution of this paper is its proof technique which is based on the “potential function method” introduced in [8] and [9]. Instead of taking an arbitrary network and proving the desired outer bound while keeping the network fixed throughout, we consider a function from the set of all m -input/ m -output discrete memoryless networks to subsets of \mathbb{R}_+^c , where \mathbb{R}_+^c is the set of all c -tuples of non-negative reals. We then identify properties of such a function which would need to be satisfied in one step of the communication for it to give rise to an outer bound. The generalized cut-set bound is then proved by a verification argument. Properties that such a function would need to satisfy are identified, intuitively speaking, as follows: take an arbitrary code of length say n over a multiterminal network. During the simulation of the code, the information of the parties begins from the i^{th} party having the i.i.d. repetitions of the random variable $W^{(i)}$; gradually evolves over time with the usage of the network; and eventually after n stages of communication reaches its final state where the parties know enough to estimate their objectives within the desired average distortion. The idea is to quantify this gradual evolution of information; *bound the derivative of the information growth at each stage* from above by showing that one step of communication can buy us at most a certain amount; and conclude that at the final stage, i.e. the n^{th} stage, the system can not reach an information state better than n times the outer bound on the derivative of information growth. An implementation of this idea requires quantification of the information of the m parties at a given stage of the process. To that end, we evaluate the function we started with at a *virtual channel* whose inputs and outputs represent, roughly speaking, the initial and the gained knowledge of the parties at the given stage of the communication. See Lemma 1 of section III and the proof of Theorem 1 of section IV for a formal formulation.

The outline of this paper is as follows. In section II, we introduce the basic notations and definitions used in this paper. Section III contains the main results of this paper followed by section IV that sketches the proofs. For the full proofs, see [10].

II. DEFINITIONS AND NOTATION

Throughout this paper we assume that each random variable takes values in a finite set. \mathbb{R} denotes the set of real numbers and \mathbb{R}_+ denotes the set of non-negative reals. For any natural number k , let $[k] = \{1, 2, 3, \dots, k\}$. For a set $S \subset [k]$, let S^c denote its compliment, that is $[k] - S$. The context will make the ambient space of S clear.

We represent a GMN by the conditional distribution

$$q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$$

meaning that the input by the i^{th} party is $X^{(i)}$ and the output at the i^{th} party is $Y^{(i)}$. We assume that the i^{th} party ($1 \leq i \leq m$) has access to i.i.d. repetitions of $W^{(i)}$. The message that needs to be delivered (in a possibly lossy manner) to the i^{th} party is taken to be $M^{(i)} = f^{(i)}(W^{(1)}, W^{(2)}, \dots, W^{(m)})$ for some function $f^{(i)}(\cdot)$. We assume that for any $i \in [m]$, random variables $X^{(i)}$, $Y^{(i)}$, $W^{(i)}$ and $M^{(i)}$ take values from discrete sets $\mathcal{X}^{(i)}$, $\mathcal{Y}^{(i)}$, $\mathcal{W}^{(i)}$ and $\mathcal{M}^{(i)}$ respectively. For any natural number n , let $(\mathcal{X}^{(i)})^n$, $(\mathcal{Y}^{(i)})^n$, $(\mathcal{W}^{(i)})^n$ and $(\mathcal{M}^{(i)})^n$ denote the n -th product sets of $\mathcal{X}^{(i)}$, $\mathcal{Y}^{(i)}$, $\mathcal{W}^{(i)}$ and $\mathcal{M}^{(i)}$. We use $Y_{1:k}^{(i)}$ to denote $(Y_1^{(i)}, Y_2^{(i)}, \dots, Y_k^{(i)})$.

For any $i \in [m]$, let the distortion function $\Delta^{(i)}$ be a function $\Delta^{(i)} : \mathcal{M}^{(i)} \times \mathcal{M}^{(i)} \rightarrow [0, \infty)$ satisfying $\Delta^{(i)}(m^{(i)}, m^{(i)}) = 0$ for all $m^{(i)} \in \mathcal{M}^{(i)}$. For any natural number n and vectors $(m_1^{(i)}, m_2^{(i)}, \dots, m_n^{(i)})$ and $(m_1'^{(i)}, m_2'^{(i)}, \dots, m_n'^{(i)})$ from $(\mathcal{M}^{(i)})^n$, let

$$\Delta_n^{(i)}(m_{1:n}^{(i)}, m_{1:n}'^{(i)}) = \frac{1}{n} \sum_{k=1}^n \Delta^{(i)}(m_k^{(i)}, m_k'^{(i)}).$$

Roughly speaking, we require the i.i.d. repetitions of random variable $M^{(i)}$ to be reconstructed, by the i^{th} party, within the average distortion of $D^{(i)}$.

Definition 1: Given natural number n , an (n) -code is the following set of mappings: for all $i \in [m]$, and $k \in [n] - \{1\}$

$$\begin{aligned} \zeta_1^{(i)} &: (\mathcal{W}^{(i)})^n \rightarrow \mathcal{X}^{(i)}; \\ \zeta_k^{(i)} &: (\mathcal{W}^{(i)})^n \times (\mathcal{Y}^{(i)})^{k-1} \rightarrow \mathcal{X}^{(i)}; \\ \vartheta^{(i)} &: (\mathcal{W}^{(i)})^n \times (\mathcal{Y}^{(i)})^n \rightarrow (\mathcal{M}^{(i)})^n. \end{aligned}$$

Intuitively speaking $\zeta_k^{(i)}$ is the encoding function of the i^{th} party at the k^{th} time instance, and $\vartheta^{(i)}$ is the decoding function of the i^{th} party.

Given positive reals ϵ and $D^{(i)}$ ($1 \leq i \leq m$), and a source marginal distribution $p(w^{(1)}, w^{(2)}, \dots, w^{(m)})$, an (n) -code is said to satisfy the average distortion interval $D^{(i)}$ (for all $i \in [m]$) over the channel $q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$ if the following ‘‘average distortion’’ condition is satisfied:

Assume that random variables $W_{1:n}^{(i)}$ for $i \in [m]$ are n i.i.d. repetition of random variables $(W^{(1)}, W^{(2)}, \dots, W^{(m)})$ with joint distribution $p(w^{(1)}, w^{(2)}, \dots, w^{(m)})$. Random variables $X_k^{(i)}$ and $Y_k^{(i)}$ ($k \in [n]$, $i \in [m]$) are defined according to the following constraints:

$$\begin{aligned} p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, \dots, w_{1:n}^{(m)}, x_{1:n}^{(1)}, \dots, x_{1:n}^{(m)}, y_{1:n}^{(1)}, y_{1:n}^{(2)}, \dots, y_{1:n}^{(m)}) = \\ \prod_{k=1}^n p(w_k^{(1)}, w_k^{(2)}, \dots, w_k^{(m)}) \times \\ \prod_{k=1}^n q(y_k^{(1)}, y_k^{(2)}, \dots, y_k^{(m)} | x_k^{(1)}, x_k^{(2)}, \dots, x_k^{(m)}) \times \\ \prod_{k=1}^n \prod_{i=1}^m p(x_k^{(i)} | w_{1:n}^{(i)}, y_{1:k-1}^{(i)}); \end{aligned}$$

and that $X_1^{(i)} = \zeta_1^{(i)}(W_{1:n}^{(i)})$, and for any $2 \leq k \leq n$, $X_k^{(i)} = \zeta_k^{(i)}(W_{1:n}^{(i)}, Y_{1:k-1}^{(i)})$. Random variables $X_k^{(i)}$ and $Y_k^{(i)}$ are representing the input and outputs of the i^{th} party at the k^{th} time instance and satisfy the following Markov chains:

$$\begin{aligned} W_{1:n}^{(1)} \dots W_{1:n}^{(m)} Y_{1:k-1}^{(1)} \dots Y_{1:k-1}^{(m)} - W_{1:n}^{(i)} Y_{1:k-1}^{(i)} - X_k^{(i)}, \\ W_{1:n}^{(1)} \dots W_{1:n}^{(m)} Y_{1:k-1}^{(1)} \dots Y_{1:k-1}^{(m)} - X_k^{(1)} \dots X_k^{(m)} - Y_k^{(1)} \dots Y_k^{(m)}. \end{aligned}$$

We then have the following constraint for any $i \in [m]$:

$$\mathbb{E} \left[\Delta_n^{(i)} \left(\vartheta^{(i)}(W_{1:n}^{(i)}, Y_{1:n}^{(i)}), M_{1:n}^{(i)} \right) \right] \leq D^{(i)} + \epsilon,$$

where $M_k^{(i)} = f^{(i)}(W_k^{(1)}, W_k^{(2)}, \dots, W_k^{(m)})$.

Definition 2: Given positive reals $D^{(i)}$, a source marginal distribution $p(w^{(1)}, w^{(2)}, \dots, w^{(m)})$ is called an *admissible source* over the channel

$$q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$$

if for every positive ϵ and sufficiently large n , an (n) -code satisfying the average distortion $D^{(i)}$, exists.

The ‘‘independent messages zero distortion capacity region’’ of the GMN is a subset of m^2 -tuples of non-negative numbers $R^{(i,j)}$ for $i, j \in [m]$ defined as follows: consider the set of all sets $\mathcal{W}^{(1)}, \mathcal{W}^{(2)}, \dots, \mathcal{W}^{(m)}$, functions $f^{(i)}(W^{(1)}, W^{(2)}, \dots, W^{(m)})$ ($1 \leq i \leq m$) having the special form of

$$(f^{(i,1)}(W^{(1)}), f^{(i,2)}(W^{(2)}), \dots, f^{(i,m)}(W^{(m)})),$$

the distortion functions $\Delta^{(i)}(m^{(i)}, m'^{(i)})$ (for $1 \leq i \leq m$) being equal to the indicator function $\mathbf{1}[m^{(i)} \neq m'^{(i)}]$, $D^{(i)}$ being set to be zero for all $1 \leq i \leq m$ and admissible sources $p(w^{(1)}, w^{(2)}, \dots, w^{(m)})$ for which $f^{(i,j)}(W^{(j)})$'s are mutually independent of each other. The capacity region is then taken to be the set of all achievable $R^{(i,j)} = H(f^{(j,i)}(W^{(i)}))$ (for $i, j \in [m]$) given the above constraints. Intuitively speaking, $R^{(i,j)}$ is the communication rate from i^{th} party to the j^{th} party.

Definition 3: For any natural number c and any two sets of points K and L in \mathbb{R}_+^c , let $K \oplus L$ refer to their Minkowski sum: $K \oplus L = \{v_1 + v_2 : v_1 \in K, v_2 \in L\}$. For any real number r , let $r \times K = \{r \cdot v_1 : v_1 \in K\}$. We also define $\frac{K}{r}$ as the set formed by shrinking K through scaling each point of it by a factor $\frac{1}{r}$.

Definition 4: For any two points \vec{v}_1 and \vec{v}_2 in \mathbb{R}_+^c , we say $\vec{v}_1 \geq \vec{v}_2$ if and only if each coordinate of \vec{v}_1 is greater than or equal to the corresponding coordinate of \vec{v}_2 . For any two sets of points A and B in \mathbb{R}_+^c , we say $A \leq B$ if and only if for any point $\vec{a} \in A$, there exists a point $\vec{b} \in B$ such that $\vec{a} \leq \vec{b}$. For a set $A \in \mathbb{R}_+^c$, the down-set $\Pi(A)$ is defined as: $\Pi(A) = \{\vec{v} \in \mathbb{R}_+^c : \vec{v} \leq \vec{w} \text{ for some } \vec{w} \in A\}$.

Definition 5: Given a specific network architecture $q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$, and the source marginal distribution $p(w^{(1)}, w^{(2)}, \dots, w^{(m)})$, it may be possible to find properties that the inputs to the multiterminal network throughout the communication satisfy. For instance in an interference channel or a multiple access channel

with no output feedback, if the transmitters observe independent messages, the random variables representing their information stay independent of each other throughout the communication. This is because the transmitters neither interact nor receive any feedback from the outputs. Other constraints on the inputs to the network might come from practical requirements such as a maximum instantaneous power used up by one or a group of nodes in each stage of the communication. Given a multiterminal network $q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$ and assuming that $\mathcal{X}^{(i)}$ ($i \in [m]$) is the set $X^{(i)}$ is taking value from, let Ψ be a set of joint distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \times \dots \times \mathcal{X}^{(m)}$ for which the following guarantee exists: for any communication protocol, the inputs to the multiterminal network at each time stage have a joint distribution belonging to the set Ψ . Such a set will be called a *permissible set* of input distributions. Some of the results below will be stated in terms of this nebulously defined region Ψ . To get explicit results, simply replace Ψ by the set of all probability distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \times \dots \times \mathcal{X}^{(m)}$.

III. STATEMENT OF THE RESULTS

Theorem 1: Given any GMN

$$q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)}),$$

a sequence of non-negative real numbers $D^{(i)}$ ($i \in [m]$), an arbitrary admissible source $W^{(i)}$ ($i \in [m]$), and a permissible set of input distributions of the network Ψ , there exists

- joint distribution $q(x^{(1)}, x^{(2)}, \dots, x^{(m)}, z)$ where size of the alphabet set of Z is $2^m - 1$ and furthermore $q(x^{(1)}, x^{(2)}, \dots, x^{(m)} | z)$ belongs to Ψ for any value z that the random variable Z might take;
- joint distribution $p(\widehat{m}^{(1)}, \widehat{m}^{(2)}, \dots, \widehat{m}^{(m)}, w^{(1)}, \dots, w^{(m)})$ where the average distortion between

$$M^{(i)} = f^{(i)}(W^{(1)}, W^{(2)}, \dots, W^{(m)})$$

and $\widehat{M}^{(i)}$ is less than or equal to $D^{(i)}$, i.e. $\Delta^{(i)}(M^{(i)}, \widehat{M}^{(i)}) \leq D^{(i)}$,

such that for any arbitrary $T \subset [m]$ the following inequality holds:

$$\begin{aligned} I(W^{(i)} : i \in T ; \widehat{M}^{(j)} : j \in T^c | W^{(j)} : j \in T^c) \leq \\ I(X^{(i)} : i \in T ; Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c, Z), \end{aligned}$$

where $Y^{(1)}, Y^{(2)}, \dots, Y^{(m)}, X^{(1)}, X^{(2)}, \dots, X^{(m)}$ and Z are jointly distributed according to

$$q(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}) \cdot q(x^{(1)}, \dots, x^{(m)}, z).$$

Discussion 1: The fact that the expressions on both sides of the above inequality are of the same form is suggestive. To any given channel $q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$ and input distribution $q(x^{(1)}, x^{(2)}, \dots, x^{(m)})$, assign the down-set of a vector in $\mathbb{R}_+^{2^m}$ whose k^{th} coordinate is defined as

$$I(X^{(i)} : i \in T_k ; Y^{(j)} : j \in T_k^c | X^{(j)} : j \in T_k^c),$$

where T_k is defined as follows: there are 2^m subsets of $[m]$; take an arbitrary ordering of these sets and take T_k to be the k^{th} subset in that ordering. Next, to any channel $q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$ and a set of permissible input distributions, we assign a region by taking the convex hull of the union over all permissible input distributions, of the region associated to the channel and the varying input distribution. A channel is said to be weaker than another channel if the region associated to the first channel is contained in the region associated to the second channel.

Intuitively speaking, given a communication task one can consider a virtual channel whose inputs and outputs represent, roughly speaking, the raw and acceptable information objectives at the m parties. Furthermore, let the only permissible input distribution for this virtual channel to be one given by the statistical description of the raw information of the parties. More specifically, given any $p(\widehat{m}^{(1)}, \widehat{m}^{(2)}, \dots, \widehat{m}^{(m)}, w^{(1)}, w^{(2)}, \dots, w^{(m)})$ such that $\Delta^{(i)}(M^{(i)}, \widehat{M}^{(i)}) \leq D^{(i)}$ holds, consider the virtual channel $p(\widehat{m}^{(1)}, \widehat{m}^{(2)}, \dots, \widehat{m}^{(m)} | w^{(1)}, w^{(2)}, \dots, w^{(m)})$ and the input distribution $p(w^{(1)}, w^{(2)}, \dots, w^{(m)})$. The inputs of this virtual channel, i.e. $W^{(1)}, W^{(2)}, \dots, W^{(m)}$, and its outputs, i.e. $\widehat{M}^{(1)}, \widehat{M}^{(2)}, \dots, \widehat{M}^{(m)}$, can be understood as the raw information and acceptable information objectives at the m parties. The region associated to the virtual channel $p(\widehat{m}^{(1)}, \widehat{m}^{(2)}, \dots, \widehat{m}^{(m)} | w^{(1)}, w^{(2)}, \dots, w^{(m)})$ and the input distribution $p(w^{(1)}, w^{(2)}, \dots, w^{(m)})$ would be the down-set of a vector in $\mathbb{R}_+^{2^m}$ whose k^{th} coordinate is defined as

$$I(W^{(i)} : i \in T ; \widehat{M}^{(j)} : j \in T^c | W^{(j)} : j \in T^c).$$

Theorem 1 is basically saying that this region associated to this virtual channel and the corresponding input distribution should be included inside the region associated to the channel $q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$. Here the complexity of transmission of functions of correlated messages is effectively translated into the performance region of a virtual channel at a given input distribution. This virtual channel at the given input distribution must be, in the above mentioned sense, weaker than any physical channel fit for the communication problem.

Remark 1: This bound reduces to the traditional cut-set bound for the “independent messages zero distortion capacity region,” (see Definition 2) when Ψ is taken to be the set of all input distributions, and $I(X^{(i)} : i \in T ; Y^{(i)} : i \in T^c | X^{(i)} : i \in T^c, Z)$ is bounded from above by

$$I(X^{(i)} : i \in T ; Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c).$$

This bound is sometimes tight; for instance it is tight for a multiple access channel with independent source messages when Ψ is taken to be the set of all mutually independent input distributions.

A. The Main Lemma

During the simulation of the code, the information of the parties begins from the i^{th} party having $W_{1:n}^{(i)}$ and gradually

evolves over time with the usage of the network. At the j^{th} stage, the i^{th} party has $W_{1:n}^{(i)} Y_{1:j}^{(i)}$. We represent the information state of the whole system at the j^{th} stage by the virtual channel $p(w_{1:n}^{(1)} y_{1:j}^{(1)}, \dots, w_{1:n}^{(m)} y_{1:j}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})$ and the input distribution $p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})$. In order to quantify the information state, we map the information state to a subset of \mathbb{R}_+^c (c is a natural number) using a function $\phi(\cdot)$. A formal definition of ϕ and the properties we require it to satisfy are as follows:

Let $\phi(p(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \Psi)$ be a function that takes as input an arbitrary m -input/ m -output GMN and a subset of probability distributions on the inputs of this network and returns a subset of \mathbb{R}_+^c where c is a natural number. $\phi(\cdot)$ is thus a function from the set of all conditional probability distributions defined on finite sets and a corresponding set of input distributions, to subsets of \mathbb{R}_+^c .

Assume that the function $\phi(\cdot)$ satisfies the following three properties. The intuitive description of the properties is provided after their formal statement. Please see Definitions 3 and 4 for the notations used.

- 1) Assume that the conditional distribution $p(y^{(1)} y'^{(1)}, y^{(2)} y'^{(2)}, \dots, y^{(m)} y'^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$ satisfies the following

$$\begin{aligned} & p(y^{(1)} y'^{(1)}, y^{(2)} y'^{(2)}, \dots, y^{(m)} y'^{(m)} | x^{(1)}, \dots, x^{(m)}) \\ &= p(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)}) \cdot \\ & p(y'^{(1)}, y'^{(2)}, \dots, y'^{(m)} | x'^{(1)}, x'^{(2)}, \dots, x'^{(m)}), \end{aligned}$$

where $X'^{(i)}$ is a deterministic function of $Y^{(i)}$ (i.e. $H(X'^{(i)} | Y^{(i)}) = 0$ ($i \in [m]$)). Random variable $X'^{(i)}$ (for $i \in [m]$) is assumed to take value from set $\mathcal{X}'^{(i)}$. Take an arbitrary input distribution $q(x_1, x_2, \dots, x_m)$. This input distribution, together with the conditional distribution $p(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$, impose a joint distribution $q(x'^{(1)}, x'^{(2)}, \dots, x'^{(m)})$ on $(X'^{(1)}, X'^{(2)}, \dots, X'^{(m)})$. Then the following constraint needs to be satisfied for any arbitrary set Ψ of joint distributions on $\mathcal{X}'^{(1)} \times \mathcal{X}'^{(2)} \times \dots \times \mathcal{X}'^{(m)}$ that contains $q(x'^{(1)}, x'^{(2)}, \dots, x'^{(m)})$:

$$\begin{aligned} & \phi \left(p(y^{(1)} y'^{(1)}, \dots, y^{(m)} y'^{(m)} | x^{(1)}, \dots, x^{(m)}) \right. \\ & \quad \left. , \{q(x_1, \dots, x_m)\} \right) \subseteq \\ & \phi(p(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \{q(x_1, \dots, x_m)\}) \\ & \oplus \phi(p(y'^{(1)}, y'^{(2)}, \dots, y'^{(m)} | x'^{(1)}, \dots, x'^{(m)}), \Psi). \end{aligned}$$

- 2) Assume that

$$p(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}) = \prod_{i=1}^m \mathbf{1}[y^{(i)} = x^{(i)}].$$

Then we require that for any input distribution $q(x_1, x_2, \dots, x_m)$, the set

$$\phi(p(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \{q(x_1, \dots, x_m)\})$$

contains only the origin in \mathbb{R}^c .

- 3) Assume that

$$\begin{aligned} & p(z^{(1)}, \dots, z^{(m)}, y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}) = \\ & p(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}) \prod_{i=1}^m p(z_i | y_i). \end{aligned}$$

Then we require that for any input distribution $q(x_1, x_2, \dots, x_m)$,

$$\begin{aligned} & \phi(p(z^{(1)}, \dots, z^{(m)} | x^{(1)}, \dots, x^{(m)}), \{q(x_1, \dots, x_m)\}) \subseteq \\ & \phi(p(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \{q(x_1, \dots, x_m)\}). \end{aligned}$$

The first condition is intuitively saying that additional use of the channel

$$p(y'^{(1)}, y'^{(2)}, \dots, y'^{(m)} | x'^{(1)}, x'^{(2)}, \dots, x'^{(m)})$$

can expand $\phi(\cdot)$ by at most

$$\phi(p(y'^{(1)}, y'^{(2)}, \dots, y'^{(m)} | x'^{(1)}, x'^{(2)}, \dots, x'^{(m)}), \Psi).$$

The second condition is intuitively saying that $\phi(\cdot)$ vanishes if the parties are unable to communicate, that is each party receives exactly what it puts at the input of the channel. The third condition is basically saying that making a channel weaker at each party can not cause $\phi(\cdot)$ expand.

Lemma 1: For any function $\phi(\cdot)$ satisfying the above three properties, and for any multiterminal network

$$q(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)}),$$

distortions $D^{(i)}$ and arbitrary admissible source $W^{(i)}$ ($i \in [m]$), positive ϵ and (n) -code satisfying the distortion constraints and a permissible set Ψ of input distributions, we have (for the definition of multiplication of a set by a real number see Definition 3):

$$\begin{aligned} & \phi(p(\widehat{m}_{1:n}^{(1)}, \dots, \widehat{m}_{1:n}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})\}) \subseteq \\ & n \times \text{Convex Hull} \{ \phi(q(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \Psi) \}, \end{aligned}$$

where $W_{1:n}^{(i)}$ ($i \in [m]$) are the messages observed at the nodes; $\widehat{M}_{1:n}^{(i)}$ ($i \in [m]$) are the reconstructions by the parties at the end of the communication satisfying

$$\mathbb{E} \left[\Delta_n^{(i)} \left((\widehat{m}_{1:n}^{(i)}, m_{1:n}^{(i)}) \right) \right] \leq D^{(i)} + \epsilon,$$

for any $i \in [m]$.

IV. PROOFS

Proof of Lemma 1: Let random variables $X_k^{(i)}$ and $Y_k^{(i)}$ ($k \in [n]$, $i \in [m]$) respectively represent the inputs to the multiterminal network and the outputs at the nodes of the network. We have:

$$\phi(p(\widehat{m}_{1:n}^{(1)}, \dots, \widehat{m}_{1:n}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})\}) \subseteq (1)$$

$$\begin{aligned} & \phi(p(w_{1:n}^{(1)} y_{1:n}^{(1)}, \dots, w_{1:n}^{(m)} y_{1:n}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})\}) \\ & \subseteq (2) \end{aligned}$$

$$\phi(p(w_{1:n}^{(1)} y_{1:n-1}^{(1)}, \dots, w_{1:n}^{(m)} y_{1:n-1}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})\})$$

$$\oplus \phi(q(y_n^{(1)}, \dots, y_n^{(m)} | x_n^{(1)}, \dots, x_n^{(m)}), \Psi) \subseteq$$

$$\dots \subseteq$$

$$\phi(p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})\})$$

$$\begin{aligned}
& \oplus \phi(q(y_1^{(1)}, \dots, y_1^{(m)} | x_1^{(1)}, \dots, x_1^{(m)}), \Psi) \\
& \oplus \phi(q(y_2^{(1)}, \dots, y_2^{(m)} | x_2^{(1)}, \dots, x_2^{(m)}), \Psi) \oplus \dots \\
& \oplus \phi(q(y_n^{(1)}, \dots, y_n^{(m)} | x_n^{(1)}, \dots, x_n^{(m)}), \Psi) \subseteq \\
& \phi(q(y_1^{(1)}, \dots, y_1^{(m)} | x_1^{(1)}, \dots, x_1^{(m)}), \Psi) \\
& \oplus \phi(q(y_2^{(1)}, \dots, y_2^{(m)} | x_2^{(1)}, \dots, x_2^{(m)}), \Psi) \oplus \dots \\
& \oplus \phi(q(y_n^{(1)}, \dots, y_n^{(m)} | x_n^{(1)}, \dots, x_n^{(m)}), \Psi) \subseteq \\
& n \times \text{Convex Hull}\{\phi(q(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \Psi)\},
\end{aligned} \tag{3}$$

where in equation 1 we have used property (3); in equation 2 we have used property (1) because

$$\begin{aligned}
& p(w_{1:n}^{(1)} y_{1:n}^{(1)}, \dots, w_{1:n}^{(m)} y_{1:n}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}) = \\
& p(w_{1:n}^{(1)} y_{1:n-1}^{(1)}, \dots, w_{1:n}^{(m)} y_{1:n-1}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}) \cdot \\
& p(y_n^{(1)}, \dots, y_n^{(m)} | x_n^{(1)}, \dots, x_n^{(m)})
\end{aligned}$$

and furthermore $H(X_n^{(i)} | W_{1:n}^{(i)} Y_{1:n-1}^{(i)}) = 0$ for all $i \in [m]$, and that

$$p(y_n^{(1)}, \dots, y_n^{(m)} | x_n^{(1)}, \dots, x_n^{(m)}) = q(y_n^{(1)}, \dots, y_n^{(m)} | x_n^{(1)}, \dots, x_n^{(m)}).$$

The definition of permissible sets implies that the joint distribution $p(x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(m)})$ is in Ψ ; in equation 3 we have used property (2). In equation 4, we first note that the conditional distributions $q(y_i^{(1)}, \dots, y_i^{(m)} | x_i^{(1)}, \dots, x_i^{(m)})$ for $i = 1, 2, \dots, n$ are all the same. We then observe that whenever $\vec{v}_i \in \phi(q(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \Psi)$ for $i \in [n]$, their average, $\frac{1}{n} \sum_{i=1}^n \vec{v}_i$ falls in the convex hull of $\phi(q(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \Psi)$. ■

Proof of Theorem 1: Full proof of this Theorem is provided in [10]. The sketch of the proof is as follows: We define a function $\phi(\cdot)$ as follows: for any conditional distribution $p(y^{(1)}, y^{(2)}, \dots, y^{(m)} | x^{(1)}, x^{(2)}, \dots, x^{(m)})$ and an arbitrary set Ψ of distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \dots \times \mathcal{X}^{(m)}$, let

$$\begin{aligned}
& \phi(p(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \Psi) = \\
& \bigcup_{p(x^{(1)}, \dots, x^{(m)}) \in \Psi} \varphi(p(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}) p(x^{(1)}, \dots, x^{(m)})),
\end{aligned}$$

where $\varphi(p(y^{(1)}, y^{(2)}, \dots, y^{(m)}, x^{(1)}, x^{(2)}, \dots, x^{(m)}))$ is defined as the down-set (see Definition 4) of a vector of size $c = 2^m - 2$ whose k^{th} coordinate equals $I(X^{(i)} : i \in T_k ; Y^{(j)} : j \in (T_k)^c | X^{(j)} : j \in (T_k)^c)$ where T_k is defined as follows: there are $2^m - 2$ subsets of $[m]$ that are neither empty nor equal to $[m]$. Take an arbitrary ordering of these sets and take T_k to be the k^{th} subset in that ordering. We then verify that $\phi(\cdot)$ satisfies the three properties of Lemma 1 for the choice of $c = 2^m - 2$. Lemma 1 thus implies that:

$$\begin{aligned}
& \phi(p(\hat{m}_{1:n}^{(1)}, \dots, \hat{m}_{1:n}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})\}) = \\
& \varphi(p(\hat{m}_{1:n}^{(1)}, \dots, \hat{m}_{1:n}^{(m)} | w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)}) p(w_{1:n}^{(1)}, \dots, w_{1:n}^{(m)})) \subseteq \\
& n \times \text{Convex Hull}\{\phi(q(y^{(1)}, \dots, y^{(m)} | x^{(1)}, \dots, x^{(m)}), \Psi)\}.
\end{aligned} \tag{5}$$

For the right hand side we use Carathéodory theorem to express points in the convex hull as convex combination of $c + 1 = 2^m - 1$ points in the set. Random variable Z in the theorem is related to this convexification. The set on the left

hand side is by definition the down-set of a vector of length $2^m - 2$ whose k^{th} coordinate is equal to

$$I(W_{1:n}^{(i)} : i \in T_k ; \widetilde{M}_{1:n}^{(j)} : j \in (T_k)^c | W_{1:n}^{(j)} : j \in (T_k)^c).$$

The next step is to show that this vector is coordinate-wise greater than or equal to the vector whose k^{th} element equals

$$n \cdot I(\widetilde{W}^{(i)} : i \in T_k ; \widetilde{M}^{(j)} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c),$$

for some $\widetilde{W}^{(i)}$ and $\widetilde{M}^{(i)}$ ($i \in [m]$) such that the joint distribution of $\widetilde{W}^{(i)}$ ($i \in [m]$) is the same as that of $W^{(i)}$ ($i \in [m]$), and that the average distortion between $\widetilde{M}^{(i)} = f^{(i)}(\widetilde{W}^{(1)}, \widetilde{W}^{(2)}, \dots, \widetilde{W}^{(m)})$ and $\widetilde{M}^{(i)}$ is less than or equal to $D^{(i)} + \epsilon$. In the next step of the proof, we perturb random variables $\widetilde{M}^{(i)}$ (for $i \in [m]$) and define random variables $\widetilde{M}'^{(i)}$ (for $i \in [m]$) such that for every $i \in [m]$, the average distortion between $\widetilde{M}'^{(i)}$ and $\widetilde{M}^{(i)}$ is less than or equal to $D^{(i)}$ (rather than $D^{(i)} + \epsilon$ as in the case of $\widetilde{M}^{(i)}$) and also for every k ,

$$\begin{aligned}
& I(\widetilde{W}^{(i)} : i \in T_k ; \widetilde{M}'^{(j)} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c) - O(\tau(\epsilon)) \\
& \leq I(\widetilde{W}^{(i)} : i \in T_k ; \widetilde{M}^{(j)} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c),
\end{aligned}$$

where $\tau(\cdot)$ is a real-valued function that satisfies the property that $\tau(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. The proof continues by recalling the definition of the sets involved in equation 5 and letting ϵ converge zero. ■

ACKNOWLEDGEMENT

The authors would like to thank the anonymous referees for their comments. The authors also would like TRUST (The Team for Research in Ubiquitous Secure Technology), which receives support from the National Science Foundation (NSF award number CCF-0424422) and the following organizations: Cisco, ESCHER, HP, IBM, Intel, Microsoft, ORNL, Pirelli, Qualcomm, Sun, Symantec, Telecom Italia and United Technologies, for their support of this work. The research was also partially supported by NSF grants CCF-0500023, CCF-0635372, and CNS-0627161.

REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.
- [2] M. Gastpar, "Cut-set Arguments For Source-Channel Networks," Proc IEEE Int Symp Info Theory, 34, 2004.
- [3] B. Nazer and M. Gastpar, "Computation over multiple-access channels," IEEE Trans. IT, 53(10): 3498-3516, 2007.
- [4] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," IEEE Trans. IT, 26 (6): 648-657, (1980).
- [5] G. Kramer and S. A. Savari, "Cut sets and information flow in networks of two-way channels," Proc IEEE Int Symp IT, 33, 2004.
- [6] A. Orlitsky and J. R. Roche, "Coding for computing," IEEE Trans. IT, 47 (3): 903917 (2001).
- [7] H. Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources," IEEE Trans. IT, 28 (5): 803807 (1982).
- [8] A. A. Gohari and V. Anantharam, "Information-Theoretic Key Agreement of Multiple Terminals – Part I: Source Model," *Preprint*, Dec. 2007. Available at <http://www.eecs.berkeley.edu/~aminzade/SourceModel.pdf>
- [9] A. A. Gohari and V. Anantharam, "Information-Theoretic Key Agreement of Multiple Terminals – Part II: Channel Model," *Preprint*, Dec. 2007. Available at <http://www.eecs.berkeley.edu/~aminzade/ChannelModel.pdf>
- [10] A. A. Gohari and V. Anantharam, "A Generalized Cut-Set Bound," Available at <http://arxiv.org/abs/0904.4542>