Abstract—Cut-set bounds are not, in general, tight for all classes of network communication problems. In this paper, we introduce a new technique for proving converses for the problem of transmission of correlated sources in networks, which results in bounds that are tighter than the corresponding cut-set bounds. We also define the concept of “uncertainty region” which might be of independent interest. We provide a full characterization of this region for the case of two correlated random variables. The bounding technique works as follows: on one hand, we show that if the communication problem is solvable, the uncertainty of certain random variables in the network with respect to imaginary parties that have partial knowledge of the sources must satisfy some constraints that depend on the network architecture. On the other hand, the same uncertainties have to satisfy constraints that only depend on the joint distribution of the sources. Matching these two leads to restrictions on the statistical joint distribution of the sources in communication problems that are solvable over a given network architecture. Our technique also provides nontrivial outer bounds for communication problems with secrecy constraints.

Index Terms—Correlated sources, cut-set bound, edge cut, Gray–Wyner problem, network coding, uncertainty region.

I. INTRODUCTION

Consider a directed network with a source $s$ and two sinks $t_1$ and $t_2$. Suppose that the source observes i.i.d. copies of random variables $X$, $Y$ jointly distributed according to $p(x, y)$. Sink $t_1$ is interested in the i.i.d. copies of $X$, while sink $t_2$ is interested in the i.i.d. copies of $Y$. We consider the problem of reliable transmission to fulfill the demands of both sink nodes with probability converging to one as the number of i.i.d. observations of $X$, $Y$ grows without bound.

The cut-set bound (see [2, Sec. 15.10]) says that if the demands of both sinks can be fulfilled, each of the cuts that separate $s$ from $t_1$ must have capacity at least $H(X)$, each of the cuts that separate $s$ from $t_2$ must have capacity at least $H(Y)$, and each of the cuts that separate $s$ from $(t_1, t_2)$ must have capacity at least $H(X, Y)$. The cut-set bound is known to be tight when $X = (M_0, M_1)$ and $Y = (M_2, M_3)$ for some mutually independent random variables $M_0, M_1, M_2$ [3], [4]. Another case is when $X$ and $Y$ are “linearly correlated” in the sense that one can express $X$ and $Y$ as $X = AU^m$ and $Y = BU^m$ for some random vector $U^m$, and matrices $A$ and $B$ all taking values in a given field. Without loss of generality, one can assume that the rows of $A$ and $B$ are linearly independent. By applying suitably chosen invertible linear transformations $T_1$ and $T_2$, we can write

$$T_1 X = \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} U^m,$$

$$T_2 Y = \begin{bmatrix} A_0 \\ B_1 \end{bmatrix} U^m,$$

where the rows of $A_0$, $A_1$, and $B_1$ are linearly independent. Because the linear transformations $T_1$ and $T_2$ are invertible, the communication task is to transmit the common message $A_0 U^m$ to both the sinks, and the private messages $A_1 U^m$ and $B_1 U^m$ to the two sinks. Clearly, this problem reduces to the one mentioned above if $A_0 U^m$, $A_1 U^m$, and $B_1 U^m$ are mutually independent. Therefore, the cut-set bound is also tight in such cases.

However, in general, when the joint distribution of $X$ and $Y$ is arbitrary, the cut-set bound is not always tight. To go beyond the cut-set bound, we devise a new technique for proving converses for the problem of transmission of correlated sources over networks. We provide an example for which the cut-set bound is not tight, but the new converse is tight. Nonetheless, the problem of finding joint distribution of the sources in communication problems that are solvable over a given network remains an open problem. One can refer to the several papers written on this topic for treatments of special cases of this problem (see, for instance, [5]–[14]). Some of these works discuss different settings in which separated source coding and network coding become either optimal or suboptimal.

In particular, in [5], [6], [8], and [9], the authors examine the problem of multicasting for correlated sources—in these scenarios, network coding is known to achieve the cut-set bound (in...
some scenarios even linear codes suffice). The work of [7] first demonstrated that for general distributed source coding problems, cut-set bounds do not always match what can be achieved by separation of source and network coding. The techniques in [10]–[14] result in achievable schemes for either lossless or lossy distributed source coding over networks—in general, the performance of these schemes is not known to match that of any outer bound. The work of [15]–[17] provides nontrivial outer bounds on the capacity of network coding problems with independent sources.

At the heart of our technique lies the concept of “uncertainty region” and how we relate it to networks. We define the uncertainty region as the set of all possible uncertainty vectors where each of these vectors is trying to capture the uncertainty of a given random variable from the perspective of different observers who have access to distinct but dependent sources. More precisely, given an arbitrary random variable $K$, a vector formed by listing the uncertainty left in $K$ when conditioned on different subsets of i.i.d. copies of, i.e.,

$$
\frac{1}{n} H(K), \frac{1}{n} H(K|X^n), \frac{1}{n} H(K|Y^n), \frac{1}{n} H(K;X^n, Y^n)
$$

is called an uncertainty vector. Since the statistical dependence between the sources affects the uncertainty region in a crucial way, our discussion of correlated sources here is not a straightforward extension of the case of independent sources.

Our technique also differs from those developed by Kramer and Savari [15], Harvey, et al. [16], and Thakor et al. [17], most of which were developed for outer bounds on the communication of independent sources over networks. Many of these works (see, for instance, [17]) build on the technique of [18], of constraining the set of entropic vectors satisfying the network communication problem via an LP—as such these bounds also apply to correlated sources, but can be viewed in this setting as a generalization of the cut-set bound.

The work of [19] (done independently and parallel to this paper) is probably closest to that of this work—it notes (as we do) via an example that cut-set bounds are not, in general, tight when correlated sources need to be transmitted over networks. The authors then propose adding auxiliary random variables that constrain the set of feasible solutions and thus may provide tighter bounds than cut-set bounds—this too is similar to the approach in this study. The primary difficulty in using the techniques of [19] is that they do not have single-letter characterizations, and hence cannot be “effectively computed.” A primary contribution of our work is to provide nontrivial single-letter outer bounds that often improve on the cut-set bounds.

The work of [20] considers, among other scenarios, the problem of transmitting correlated sources over noisy networks and, in some special cases, is able to provide achievable schemes that match cut-set bounds.

In Section VI, we introduce the concept of “edge cut” and use it to prove better outer bounds for the problem of transmission of correlated sources over a network. (Our concept of edge-cuts should not be confused with that of [15], [21], or [22].) Our discussion on edge-cuts reveals a difference between transmission of independent sources and that of correlated sources. We consider the problem of transmission of two sources over the simple network of Fig. 5. The cut-set bound is tight for this network if the sources are independent. In other words, it is a collection of the edges on a cut (rather than the individual edges) that affect the plausibility of transmission of the two sources. Not so for correlated sources. If we slightly perturb the joint distribution of the sources to make them correlated, individual edges (rather than their collective on a cut) can be the bottleneck. Our discussion of “edge-cut” is meant to highlight this fact.

We also note that the techniques developed in this paper have implications for communications with “secrecy constraints” [23]. This is because a primary technique in this study is to characterize uncertainty regions, an understanding of which seems critical for characterizing secrecy capacity of communication problems. We demonstrate novel outer bounds obtained via our technique for an example network (see Example III in Section V).

The rest of this paper is organized as follows. In Section II, we motivates our new technique via an example. Sections III
and IV contain the main results of this paper, respectively, a complete characterization of the uncertainty region, and a discussion on how to use this to write better converses. Section V illustrates the use of the tools developed in Sections III and IV via some examples. Section VI demonstrates directions in which our new converse technique can be improved upon even further, via “edge-cuts.” Section VII contains the proofs of our theorems.

II. MOTIVATING EXAMPLE

This section uses an example to motivate our technique which is based on “uncertainty computations.” For the ease of exposition and to convey the main ideas, the discussion of this example is quite intuitive and not rigorous. Examples with precise calculations are in Section V, following the main results and tools developed in Sections III and IV.

We consider the well-known butterfly network shown in Fig. 1. Assume that the source is observing \( n \) i.i.d. repetitions of the correlated binary sources \((X, Y)\). Thus, the source has a length-\( n \) vector \( X^n \) and the length-\( n \) vector \( Y^n \). The first sink is interested in recovering the \( n \) i.i.d. repetitions of \( X \), whereas the second sink is interested in recovering the \( n \) i.i.d. repetitions of \( Y \). Probabilities of error at both sinks are required to converge to zero as the number of i.i.d. observations of \( X \), \( Y \) grows without bound. A cut from the source node to a sink node is a set of edges such that the source node and the sink node are disjointed if the edges in this set are removed.

For the sake of simplicity, we restrict ourselves to networks such that the cut toward the first receiver across edges 4 and 6, and the cut toward the second receiver across edges 5 and 6, are tight, that is, \( C_4 + C_6 = H(X) \) and \( C_5 + C_6 = H(Y) \). Let \( K \) denote the random variable that is put on edge 6 as shown in Fig. 1. Using the source coding theorem and the fact that \( C_4 + C_6 = H(X) \), one can conclude that \( H(K, X^n) \) ought to be negligible if the demand of the first sink is to be fulfilled. Similarly, \( H(K, Y^n) \) ought to be negligible. Therefore, \( K \) corresponds to common randomness between \( X^n \) and \( Y^n \) in the sense of Gács–Körner [24]. This common information is equal to \( \max T \), where \( T \) is both a function of \( X \) and \( Y \). For binary sources, this common information is nonzero if and only if \( X = Y \) or \( X = Y^c \). Thus, in the general case, the Gács–Körner common information for binary random variables is zero, implying that \( \frac{1}{n} H(K) \) should be “almost zero”. This effectively implies that we are not using edge 6 in communication at all. But the cuts at the two sinks are tight (by assumption), implying that \( C_4 < H(X) \) and \( C_5 < H(Y) \). There is not enough capacity to communicate \( X^n \) and \( Y^n \) through these links. This implies that the required communication demands cannot be simultaneously satisfied. Note that because even a small perturbation in the joint distribution can destroy the Gács–Körner common information between two random variables, a given network that supports transmission of certain correlated sources may not support transmission of correlated sources in its immediate vicinity. Hence, this is a “discontinuity-type” phenomenon, asymptotically in the block-length \( n \).

A more systematic way of thinking about this example comes from thinking about the “uncertainty” vector

\[
\left[\frac{1}{n} H(K), \frac{1}{n} H(K|X^n), \frac{1}{n} H(K|Y^n), \frac{1}{n} H(K, X^n, Y^n)\right],
\]

i.e., the vector formed by listing the uncertainty left in \( K \) conditioning on different subsets of \( \{X^n, Y^n\} \). Each of \( X^n \) and \( Y^n \) is “almost sufficient” to determine \( K \). Thus, the last three coordinates of the uncertainty vector are “almost zero”. Thus, the Gács–Körner common information can be reinterpreted as providing an upper bound for the first coordinate of the uncertainty vector when all the other coordinates are essentially zero.

III. UNCERTAINTY REGION

The above section motivates the definition of the uncertainty region. In this section, we formally define this region and then provide a complete characterization of it. In the next section, we discuss the use of the uncertainty region in proving converses. Given joint distribution \( p(x, y) \); on discrete random variables \( X \) and \( Y \), let us define a 4-D region uncertainty region, \( U(p) \), as the closure of the set of nonnegative 4-tuples \((u_1, u_2, u_3, u_4)\) such that for some \( n \) and \( p(k|x^n, y^n) \) we have

\[ u_1 = \frac{1}{n} H(K), \quad u_2 = \frac{1}{n} H(K|X^n), \quad u_3 = \frac{1}{n} H(K|Y^n), \quad u_4 = \frac{1}{n} H(K, X^n, Y^n). \]

Intuitively speaking, the coordinates of this vector are the uncertainties of \( K \) when i.i.d. copies of a subset of variables \( X \) and \( Y \) are available. We are interested in the set of all plausible uncertainty vectors. Note that we define the uncertainty region in terms of \( p(x, y) \) alone, irrespective of the network architecture.
We now fully characterize the uncertainty region.

**Theorem 1:** The region $U(p)$ is equal to the convex envelope of the union of the following four sets of points. The first set is the union over all $c \geq 0$ and $p(e, x, y)$ of nonnegative 4-tuples $(u_1, u_2, u_3, u_4)$ where

$$
\begin{align*}
    u_1 &= c + I(E; X, Y), \\
    u_2 &= c + I(E; Y | X), \\
    u_3 &= c + I(E; X | Y), \\
    u_4 &= c.
\end{align*}
$$

The second set of points is the union over all $c \geq 0$ of 4-tuples $(u_1, u_2, u_3, u_4)$, where

$$
\begin{align*}
    u_1 &= c + H(Y | X), \\
    u_2 &= c + H(Y | X), \\
    u_3 &= c, \\
    u_4 &= c.
\end{align*}
$$

The third set of points is the union over all $c \geq 0$ of 4-tuples $(u_1, u_2, u_3, u_4)$, where

$$
\begin{align*}
    u_1 &= c + H(X | Y), \\
    u_2 &= c, \\
    u_3 &= c + H(X | Y), \\
    u_4 &= c.
\end{align*}
$$

The fourth set of points is the union over all $c \geq 0$, $0 \leq f \leq \max(H(X | Y), H(Y | X))$ of nonnegative 4-tuples $(u_1, u_2, u_3, u_4)$, where

$$
\begin{align*}
    u_1 &= c + f, \\
    u_2 &= c + \min(f, H(Y | X)), \\
    u_3 &= c + \min(f, H(X | Y)), \\
    u_4 &= c.
\end{align*}
$$

**Remark 1:** One can use the strengthened Carathéodory theorem of Fenchel [25] to prove a cardinality bound of $X[2] + 2$ on the auxiliary random variable $E$ in the first set of points.

Although the above theorem characterizes the region, the following outer bound is useful in some instances. The extreme points of this outer bound belong to the first set of points of the above theorem.

**Theorem 2:** The uncertainty region is a subset of the union over all $c, g, h \geq 0$ and $p(e, x, y)$ of 4-tuples $(u_1, u_2, u_3, u_4)$ where

$$
\begin{align*}
    u_1 &= c + I(E; X, Y), \\
    u_2 &= c + I(E; Y | X) + g, \\
    u_3 &= c + I(E; X | Y) + h, \\
    u_4 &= c.
\end{align*}
$$

### A. Uncertainty Region for Independent Random Variables

Assume that $X$ and $Y$ are independent. In this case, we claim that the uncertainty region reduces to the union of all $(u_1, u_2, u_3, u_4)$ such that

$$
\begin{align*}
    u_1 &= c + d, \\
    u_2 &= c + e, \\
    u_3 &= c + f, \\
    u_4 &= c
\end{align*}
$$

for some $c \geq 0, 0 \leq e \leq H(Y), 0 \leq f \leq H(X), \max(e, f) \leq d \leq c + f$.

**Proof of the Claim:** We first prove that every point in the uncertainty region belongs to the above set. For any arbitrary $n, K, X^n, Y^n$, set $c = \frac{1}{n} H(K | X^n Y^n)$. First, $u_4 = c \geq 0$. Let $e = \frac{1}{n} I(K; Y^n X^n)$. We have $e \leq H(Y)$, and

$$
\begin{align*}
    u_2 &= c - \frac{1}{n} I(K; X^n Y^n), \\
    &= \frac{1}{n} I(K; X^n Y^n), \\
    &= e.
\end{align*}
$$

Similarly, let $f = \frac{1}{n} I(K; Y^n | X^n)$. We have $f \leq H(X)$ and $u_3 - c = f$. Next, let $d = \frac{1}{n} I(K; X^n Y^n)$. We can similarly verify that $u_1 - c = d$. Further, we have

$$
\begin{align*}
    d &= \frac{1}{n} I(K; X^n Y^n) \\
    &\geq \max\{\frac{1}{n} I(K; X^n Y^n), \frac{1}{n} I(K; Y^n X^n)\} \\
    &= \max\{e, f\}.
\end{align*}
$$

and using the fact that $I(X^n; Y^n) = 0$,

$$
\begin{align*}
    d &= \frac{1}{n} I(K; X^n Y^n) \\
    &\leq \frac{1}{n} I(K; X^n Y^n) + \frac{1}{n} I(K; Y^n | X^n) \\
    &= e + f.
\end{align*}
$$

Thus, every point in the uncertainty region belongs to the above set.

We now prove that every point in the above set is in the uncertainty region. We verify the inclusion for $c = 0$. Using the convexity of the uncertainty region, it suffices to prove the inclusion for $d, e$, and $f$, as a three-tuple $(d, e, f)$, in the set

$$
\begin{align*}
    \begin{cases}
        \{0, 0, 0\}, \\
        \{H(Y), H(Y), 0\}, \\
        \{H(X), 0, H(X)\}, \\
        \{H(X), H(Y), H(X)\}, \\
        \{\max\{H(X), H(Y)\}, H(Y), H(X)\}, \\
        \{\max\{H(X), H(Y)\}, H(X), H(X)\}).
    \end{cases}
\end{align*}
$$

Note that $H(Y X) = H(Y)$ and $H(X Y) = H(X)$ when $X$ and $Y$ are independent. The case of $(d, e, f) = (0, 0, 0)$ follows by the first set of points in Theorem 1 by letting $E$ be constant. The case of $(d, e, f) = (H(Y), H(Y), 0)$ follows by the second set of points in Theorem 1. The case of $(d, e, f) = (H(X), 0, H(X))$ follows by the third set of points in Theorem 1. The case of $(d, e, f) = (\max\{H(X), H(Y)\}, H(Y), H(X))$ follows by the fourth set of points in Theorem 1 by letting $f = \max\{H(X), H(Y)\}$. 

The case of $(d, e, f) = (H(X) + X(Y), H(Y), H(X))$ follows by using the first set of points in the statement of the main theorem and by letting $E = (X, Y)$. We are done.

IV. WRITING CONVERSES USING THE UNCERTAINTY REGION

Take an arbitrary directed network $\mathcal{N}$ with a source $s$ and two sinks $t_1$ and $t_2$. Suppose that the source observes i.i.d. copies of random variables $X$, $Y$ jointly distributed according to $p(x, y)$. Sink $t_1$ is interested in the i.i.d. copies of $X$, while sink $t_2$ is interested in the i.i.d. copies of $Y$. The capacity of an edge $v$ is denoted by $C_v$. Our discussion in this section can be extended naturally to more general networks, e.g., a network with $X$ and $Y$ on different network nodes and a network with a sink interested in both $X$ and $Y$. To make the main idea clear, we focus on the problem discuss above and leave the discussion of some extensions in the following sections.

An $(n, \epsilon)$ code for this network consists of a set of encoding functions at the intermediate nodes such that $X^n$ and $Y^n$ can be recovered at the first and second sinks, respectively, with probabilities of error less than or equal to $\epsilon$, and furthermore, the number of bits passed on a given edge $e$ is at most $n(C_e + \epsilon)$. Note that the above description of an $(n, \epsilon)$ is sufficient for our purpose. Readers are referred to [26, Sec. 21.3] for a detailed definition of codes for acyclic networks. (Similar codes for cyclic networks can be defined as well with respect to certain coding order.)

In order to write a converse for $\mathcal{N}$, we take the edges one by one and write a converse for that particular edge. At the end, we intersect all such converses.

Take an $(n, \epsilon)$ code. Choose a particular edge $e$ and let $K$ denote the random variable associated with the symbols transmitted on the edge $e$. The idea is to find as many constraints as possible on the uncertainty vector associated to $K$, i.e.,

$$
\frac{1}{n} H(K) - \frac{1}{n} H(K|X^n) - \frac{1}{n} H(K|Y^n) - \frac{1}{n} H(K|X^n, Y^n).
$$

Let us denote the first coordinate $\frac{1}{n} H(K)$ by $d_e$, defined as the entropy rate of the random variable on edge $e$. This $d_e$ is required to satisfy $0 \leq d_e \leq C_e + \epsilon$. Every cut that has the edge $e$ and separates the source from the first sink imposes a constraint on $\frac{1}{n} H(K|X^n)$ as follows. For a set of edges $E$, let $C_E$ denote the sum of the capacities of the edges in $E$.

**Lemma 1**: Let $E$ be an arbitrary cut from $s$ to both sinks containing an edge $e$. Then, $\frac{1}{n} H(K|X^n)$ must satisfy the following inequality:

$$
C_e + \frac{1}{n} H(K|X^n) \leq C_E + k(\epsilon),
$$

or equivalently

$$
\frac{1}{n} H(K|X^n) \leq C_E - C_e + d_e - H(X) + k(\epsilon),
$$

for some function $k(\epsilon)$ that converges to zero as $\epsilon$ goes to zero.

**Proof**: Noting that by definition $d_e = \frac{1}{n} H(K)$ and that the components of $X^n$ are i.i.d, (2) is equivalent to (3), which is proved by the following analysis.

Let $Q$ denote the collection of random variables passing over the edges in $E \setminus e$. We have $\frac{1}{n} H(Q) \leq C_K - C_e + m\epsilon$, where $m$ is the number of edges in the graph. Since $(Q, K)$ is the collection of the random variables passing the edges in the cut $E$ from $s$ to $t_1$, $X^n$ should be recoverable from $(Q, K)$ with probability of error less than or equal to $\epsilon$. Thus, by Fano’s inequality, $\frac{1}{n} H(X^n|Q, K) \leq k_1(\epsilon)$ for some function $k_1(\epsilon)$ that converges to zero as $\epsilon$ converges to zero. Thus

$$
\frac{1}{n} H(K|X^n) \leq \frac{1}{n} H(Q, K|X^n) - \frac{1}{n} H(X^n|Q, K) - \frac{1}{n} H(K) + \frac{1}{n} H(X^n|K) - \frac{1}{n} H(K|X^n, K) - \frac{1}{n} H(K|X^n, Y^n) \leq C_E - C_e + m\epsilon + d_e - H(X) + k_1(\epsilon).
$$

We get the desired result by setting $k(\epsilon) = k_1(\epsilon) + m\epsilon$.

One can use similar ideas to impose constraints on $\frac{1}{n} H(K|Y^n)$. Without loss of generality, we assume that $K$ is a function of $(X^n, Y^n)$ as randomized coding would only reduce the throughput. Thus, the last coordinate $\frac{1}{n} H(K|X^n, Y^n)$ in (1) will be zero. The following lemma (whose proof is similar to that of Lemma 1, and hence is omitted) is also useful.

**Lemma 2**: Let $E$ be an arbitrary cut from $s$ to both sinks containing an edge $e$. Then, $\frac{1}{n} H(K|X^n, Y^n)$ must satisfy the following inequality:

$$
C_e + \frac{1}{n} H(X^n, Y^n|K) \leq C_E + k(\epsilon),
$$

or equivalently

$$
\frac{1}{n} H(K|X^n, Y^n) \leq C_E - C_e + d_e - H(X, Y) + k(\epsilon),
$$

for some function $k(\epsilon)$ that converges to zero as $\epsilon$ converges to zero.

Thus, for every $(n, \epsilon)$ code, we write all such constraints on the coordinates in (1). Finally, we look at these constraints over a sequence of codes $(n_i, \epsilon_i)$ where $\epsilon_i \to 0$ as $i \to \infty$. Let $\text{Mincut}^\epsilon_x$ be the smallest cut that has the edge $e$ and separates the source from the first sink. $\text{Mincut}^\epsilon_y$ and $\text{Mincut}^\epsilon_{xy}$ are defined similarly. For the code $(n_i, \epsilon_i)$, we have

$$
\frac{1}{n_i} H(K|X^n) = d_{ei},
$$

$$
\frac{1}{n_i} H(K|X^n) \leq \text{Mincut}^\epsilon_x - C_e + d_{ei} - H(X) + k(\epsilon),
$$

$$
\frac{1}{n_i} H(K|X^n) \leq \text{Mincut}^\epsilon_y - C_e + d_{ei} - H(Y) + k(\epsilon),
$$

$$
\frac{1}{n_i} H(K|X^n, Y^n) = 0 \leq \text{Mincut}^\epsilon_{xy} - C_e + d_{ei} - H(X, Y) + k(\epsilon).
$$
Sine there is a convergent subsequence $d_{e_i}$ converging to some $d^*_e \leq C_e$, the region $U(p)$ contains a point $[u_1, u_2, u_3, u_4]$ such that

\begin{align}
  u_1 &= d^*_e, \\
  u_2 &\leq \text{MinCut}_{x} - C_e + d^*_e - H(X), \\
  u_3 &\leq \text{MinCut}_{y} - C_e + d^*_e - H(Y), \\
  u_4 &= 0 \leq \text{MinCut}_{x,y} - C_e + d^*_e - H(X,Y).
\end{align}

From Theorem 2, we know that there exist $c, g, h \geq 0$ and $p(e|x,y)$ such that

\begin{align}
  u_1 &= c + I(E; X, Y), \\
  u_2 &= c + I(E; Y|X) + g, \\
  u_3 &= c + I(E; X|Y) + h, \\
  u_4 &= c.
\end{align}

Thus, there exists a $p(e|x,y)$ such that

\begin{align}
  d^*_e &= I(E; X, Y) \leq C_e, \\
  \text{MinCut}_{x} - C_e + d^*_e - H(X) &\geq I(E; Y|X), \\
  \text{MinCut}_{y} - C_e + d^*_e - H(Y) &\geq I(E; X|Y), \\
  \text{MinCut}_{x,y} - C_e + d^*_e - H(X,Y) &\geq 0
\end{align}

imply that $\text{MinCut}_{x} - H(X) > 0$, $\text{MinCut}_{y} - H(Y) > 0$ and $\text{MinCut}_{x,y} - H(X,Y) > 0$. Since edge $e$ was arbitrary, one can see that this converse is no worse than the cut-set bound.

When $X$ and $Y$ are independent, by the discussion in Section III-A, there exist $c \geq 0$, $0 \leq g \leq H(Y)$, $0 \leq f \leq H(X)$, $\max(g, f) \leq d$ such that

\begin{align}
  u_1 &= c + d, \\
  u_2 &= c + g, \\
  u_3 &= c + f, \\
  u_4 &= c.
\end{align}

Thus, (4)–(7) imply

\begin{align}
  d &\leq C_e \\
  H(X) &\leq \text{MinCut}_{x} - C_e + d - g \\
  H(Y) &\leq \text{MinCut}_{y} - C_e + d - f \\
  H(X) + H(Y) &\leq \text{MinCut}_{x,y} - C_e + d.
\end{align}

But due to the constrains on $d, f, g$, the above converse for independent $X$ and $Y$ is again no worse than the cut-set bound.

**Remark 2**: We note that our techniques may also be helpful for communication problems wherein an auxiliary goal is to ensure secrecy against a “passive wiretapper” [23]. For a variety of such problems, the *secrecy capacity* is open (for instance, see [27]). In such scenarios, other restrictions on $\frac{1}{n}H(K|X^n)$ may come from secrecy constraints. For instance, if $K$ is observed by an eavesdropper and there is an equivocation rate constraint on how much the eavesdropper can learn about $X^n$, say $\frac{1}{n}H(K|X^n) \leq R$, we can conclude that $\frac{1}{n}H(K) - R = d_e - R$. We demonstrate an example of non-trivial outer bounds on the secrecy capacity obtained by our technique in Section V-C, but in the main leave aside the question of providing outer bounds for the secrecy capacity for future work.

**Remark 3**: A concrete algorithm for writing an outer bound for a network coding problem with two correlated sources can be summarized as follows.

1. Compute $U(p)$ as in Theorem 1 (or Theorem 2, for an easier bound). Note that Remark 1 provides a bound on the alphabet size of $E$.
2. For each edge, use Lemmas 1 and 2 as in (4)–(7), secrecy constraints or “edge-cuts” defined later, to find constraints on the uncertainty vector corresponding to the random variable of the edge. This uncertainty vector must belong to $U(p)$ and leads to an outer bound corresponding to this edge. Take the intersection of all such bounds for individual edges.

**V. EXAMPLES ILLUSTRATING OUR TECHNIQUE**

We demonstrate our technique discussed in the last two sections and some extensions via several examples.

**A. Example I**

Let us consider the network given in Fig. 2. Assume that $C_3 = C_4 = C_5$. This network is known as the Gray–Wyner system [28]. Let us write the converse for the edge number 3. The converse says that there exists a $p(e|x,y)$ such that

\begin{align}
  d^*_3 &= I(E; X, Y) \leq C_3, \\
  \text{MinCut}_{x} - C_3 + d^*_3 - H(X) &\geq I(E; Y|X), \\
  \text{MinCut}_{y} - C_3 + d^*_3 - H(Y) &\geq I(E; X|Y), \\
  \text{MinCut}_{x,y} - C_3 + d^*_3 - H(X,Y) &\geq 0
\end{align}

Note that $\text{MinCut}_{x} = C_4 + C_1 = C_3 + C_1$, $\text{MinCut}_{y} = C_5 + C_2 - C_3 + C_2$ and $\text{MinCut}_{x,y} = C_1 + C_2 + C_4 - C_3 + d^*_3 - H(X,Y) \geq 0$. After simplification and substituting the value of $d^*_3 = I(E; X, Y)$ from the first equation into the other equations, we get that

\begin{align}
  C_3 &\geq I(E; X, Y), \\
  C_1 + C_2 &\geq I(E; Y|X) - I(E; X, Y) + H(X) = H(X|E), \\
  C_3 &\geq I(E; X|Y) - I(E; X, Y) + H(Y) = H(Y|E), \\
  C_1 + C_2 &\geq H(X,Y) - I(E; X, Y) = H(X,Y|E).
\end{align}
Since the last equation above is redundant, we get
\[ C_3 \geq I(E;X,Y), C_1 \geq H(X,E), C_2 \geq H(Y,E) \]
for some \( p(e|x,y) \). But this is exactly the solution to the Gray–Wyner system [28]. Therefore, the new converse is tight. On the other hand, the cut-set bound is not tight for this network. Let us consider the minimum of \( C_3 \) such that \( C_1 + C_2 + C_3 = H(X,Y) \) over the actual region and the cut-set bound. It is known that in the Gray–Wyner system this minimum is equal to the Wyner’s common information [29]. However, in the cut-set bound, this minimum is \( I(X;Y) \) which can be strictly less than the Wyner’s common information.
Therefore, the new converse represents a strict improvement over the cut-set bound.

### B. Example II

Consider the network in Fig. 3. Further assume that \( C_2 + C_3 = H(X,Y) \), i.e., all the cuts to the middle bottom node are tight. We discuss this network in general, and in the special case where \( C_4 + C_5 = H(Y) \), i.e., when the cut to the right bottom node is tight, \( C_2 > H(X,Y) \) and \( C_1 < H(X) \).

Take an \( (n,\epsilon) \) code for this problem. Let \( K_i \) be the collection of all random variables transmitted on the edge with capacity \( C_i \). Let \( d_i = \frac{1}{n} H(K_i) \). From the definition of the code, we have
\[ d_i \leq C_i + \epsilon. \tag{11} \]

Since we should be able to recover \( (X^n,Y^n) \) from \( (K_2, K_3) \), we can use Lemma 2 to conclude that
\[ \frac{1}{n} H(K_2|X^n,Y^n) \leq C_2 + C_3 - C_2 + d_2 - H(X,Y) + k_1(\epsilon) \]
\[ = d_2 - C_2 + k_1(\epsilon) \]
\[ \leq k_1(\epsilon) + \epsilon \tag{12} \]
\[ \leq k_1(\epsilon) + \epsilon \tag{13} \]
where (12) follows from the hypothesis \( C_2 + C_3 = H(X,Y) \), and the last inequality follows from (11). Thus, \( K_2 \), and similarly \( K_3 \), are almost functions of \( X^n,Y^n \). By (12) and \( \frac{1}{n} H(K_2|X^n,Y^n) \geq 0 \), we further have
\[ d_2 \geq C_2 - k_1(\epsilon). \tag{14} \]

Using Lemma 1 on the cut on edges 1 and 2, we have
\[ \frac{1}{n} H(K_2|X^n) \leq C_2 + C_3 - C_2 + d_2 - H(X) + k_2(\epsilon) \]
\[ \leq C_2 + C_3 - H(X) + k_2(\epsilon) + \epsilon \tag{15} \]
where the last equality follows from (12). Similarly, using Lemma 1 on the cut on edges 3 and 4, we have
\[ \frac{1}{n} H(K_3|Y^n) \leq C_4 + C_3 - H(Y) + k_4(\epsilon) + \epsilon \tag{16} \]

Next, since \( \hat{X}^n \) and \( \hat{Y}^n \) are functions of \( K_2 \) and \( K_3 \), we can write
\[ \frac{1}{n} H(K_2, K_3|Y^n) > \frac{1}{n} H(\hat{X}^n, \hat{Y}^n|Y^n) > H(X,Y) - k_3(\epsilon) \tag{17} \]
where the last inequality follows from
\[ \frac{1}{n} H(\hat{X}^n, \hat{Y}^n Y^n) \geq \frac{1}{n} H(X^n, Y^n|\hat{X}^n, \hat{Y}^n, Y^n) \]
\[ \geq \frac{1}{n} H(X^n, Y^n Y^n) - k_3(\epsilon) \]
where we have used \( H(A|B) \geq H(C|B) - H(C|A) \) for any arbitrary \( A, B, C \) in the first step, and the Fano inequality in the second step. Thus, by (16) and (17),
\[ \frac{1}{n} H(K_2 Y^n) \]
\[ = \frac{1}{n} H(K_2, K_3|Y^n) - \frac{1}{n} H(K_3|K_2, Y^n) \]
\[ \geq \frac{1}{n} H(K_2, K_3|Y^n) - \frac{1}{n} H(K_3|Y^n) \]
\[ \geq H(X,Y) - k_3(\epsilon) - (C_4 + C_3 - H(Y) + k_4(\epsilon) + \epsilon) \]
\[ = H(X,Y) - C_4 - C_3 - k_3(\epsilon). \tag{18} \]
where \( k_3(\epsilon) \) is the minimum of \( C_4 + C_3 \).

Therefore, to sum this up, we have by (13)–(15) and (18)
\[ \frac{1}{n} H(K_2|Y^n) \leq C_2 - k_1(\epsilon), \]
\[ \frac{1}{n} H(K_2|Y^n) \leq C_2 + C_1 - H(X) + k_2(\epsilon) + \epsilon, \]
\[ \frac{1}{n} H(K_2|Y^n) \leq H(X,Y) - C_4 - C_3 - k_3(\epsilon), \]
\[ \frac{1}{n} H(K_2|X^n, Y^n) \leq k_1(\epsilon) + \epsilon. \]

We further have \( \frac{1}{n} H(K_2|X^n) \leq C_2 + \epsilon \) and
\[ \frac{1}{n} H(K_2|X^n, Y^n) \leq \frac{1}{n} H(K_2|X^n) + \frac{1}{n} I(K_2;X^n|Y^n) \]
\[ \leq \frac{1}{n} H(K_2|X^n, Y^n) + H(X,Y) \]
\[ \leq k_2(\epsilon) + H(X,Y) \]
\[ \leq k_2(\epsilon) + H(X,Y). \]

Since \( k_1(\epsilon) \to 0 \) as \( \epsilon \to 0 \), the above inequalities imply that there should exist a point \( (u_1, u_2, u_3, u_4) \) in the uncertainty region satisfying:
\[ u_1 = C_2, \]
\[ u_2 \leq C_2 + C_1 - H(X), \]
\[ u_3 \in [H(X,Y) - C_4 - C_3, H(X|Y)], \]
\[ u_4 = 0. \]

For the case of \( C_4 + C_3 > H(Y) \), we can use the inclusion of the above point to write a converse. However, we can say more explicit statements for the special case of \( C_4 + C_3 = H(Y) \), i.e., when the cut to the right bottom node is tight, \( C_2 > H(X|Y) \) and \( C_1 < H(X) \). In this case, \( H(X,Y) - C_4 - C_3 = H(X|Y) \), meaning that the following point has to be in the uncertainty region:
\[ u_1 = C_2, \]
\[ u_2 \leq C_2 + C_1 - H(X), \]
\[ u_3 = H(X|Y), \]
\[ u_4 = 0. \]
Since we assume that $C_2 > H(X|Y)$ and $C_1 < H(X)$, we have $w_1 > \max\{w_2, w_3\}$ and the only case where this point could belong to the uncertainty region is that it belongs to the first set of points. In other words, there has to exist some $E$ such that
\[
I(E; XY) - C_2,
\]
\[
I(E; Y|X) \leq C_2 + C_1 - H(X),
\]
\[
I(E; X Y) - H(X|Y).
\]
The last constraint implies that $H(X|EY) = 0$. So for every given $C_2 > H(X|Y)$, the minimum value of $C_1$ can be computed as follows:
\[
C_1 \geq \min_{E: H(X|EY) = 0} I(E; XY) - C_2 + H(X).
\]

Comparing the Above Bound With the Cut-Set Bound: The only constraint that we can get on $C_1$ using the cut-set bound is that $C_1 > H(X)$. Or in other words, $C_1 > H(X) - C_2$. Comparing this bound with the above lower bound on $C_1$, we see the extra term $I(E; Y|X)$. We claim that our bound is strictly better than the cut-set bound when $C_1$ is arbitrary where $C_1 > H(X)$. Assume otherwise. Our bound reduces to the cut-set bound only when there is an auxiliary random variable $E$ such that
\[
H(X|EY) = 0, \quad I(E; XY) = C_2, \quad I(E; Y|X) = 0.
\]
The last equality implies the Markov chain $E \rightarrow X \rightarrow Y$. We claim that $H(X|E) = 0$. If this is not the case, one can find $x_1$, $x_2$ and $\epsilon$ such that $p(\epsilon) > 0$, $p(x_1 \epsilon) > 0$ and $p(x_2 \epsilon) > 0$. Observe that for any arbitrary $y$, $p(x_1, y) = p(\epsilon)p(x_1 \epsilon)p(y|x_1) > 0$ and $p(x_2, y) = p(\epsilon)p(x_2 \epsilon)p(y|x_2) > 0$. Therefore, $p(x_1|y, \epsilon) > 0$ and $p(x_2|y, \epsilon) > 0$. But this implies that $X$ is not a function of $(E, Y)$ which is a contradiction with $H(X|EY) = 0$. Therefore, we must have $H(X|E) = 0$. Therefore, if our bound reduces to the cut-set bound, there is an auxiliary random variable $E$ such that
\[
H(X|E) = 0, \quad I(E; XY) = C_2, \quad E \rightarrow X \rightarrow Y.
\]
Now, observe that these equations imply that
\[
C_2 = I(E; XY) = I(E; X) + I(E; Y|X)
\]
\[= I(E; X) = H(X) - H(X|E) = H(X).
\]
But we had assumed that $C_2 < H(X)$. This is a contradiction.

C. Example III

In this section, we provide an example showing an outer bound on the secrecy capacity of networks. This example is again based on the butterfly network of Fig. 4. A passive eavesdropper is assumed to be on one of the nodes as shown in the figure. The eavesdropper can observe random variable $K$ (shown on the figure) but cannot tamper with any of the messages. The goal of the code is to keep the eavesdropper “almost ignorant” of the message of the first sink. That is, we would like to restrict our attention to those codes in which $K$ is almost independent of $X^n$, i.e., an $(n, \epsilon)$ could have have $\frac{1}{n}I(K; X^n) \leq \epsilon$. Further, assume that the cut at the second sink is tight, i.e., $C_5 + C_6 = H(Y)$. We claim that one must then have $C_6 \leq H(Y|X)$, $C_5 \geq I(X; Y)$, Otherwise, the sources are not transmittable.

Take an $(n, \epsilon)$ code. Consider the “uncertainty” vector $[\frac{1}{n}H(K), \frac{1}{n}H(K|X^n), \frac{1}{n}H(K|Y^n), \frac{1}{n}H(K|X^n, Y^n)]$, i.e., the vector formed by listing the uncertainty left in $K$ conditioning on different subsets of $\{X^n, Y^n\}$. The secrecy constraint $\frac{1}{n}I(K; X^n) \leq \epsilon$ implies that the first and the second coordinate of the uncertainty vector can differ by at most $\epsilon$. Let $R$ denote the message that is put on the edge with capacity $C_5$. Since the cut at the second sink is tight, both $K$ and $R$ must essentially be functions of $Y^n$. More precisely from $C_5 + C_6 = H(Y)$, we have $\frac{1}{n}H(K) + \frac{1}{n}H(R) < H(Y) + 2\epsilon$. Since $Y^n$ should be almost recoverable from $R$ and $K$, we have $\frac{1}{n}H(K^n|KR) - k_1(\epsilon)$ for some function $k_1(\epsilon)$ converging to zeros as $\epsilon$ goes to zero. These two equations imply that
\[
\frac{1}{n}H(K^n|KR) \leq \frac{1}{n}H(K^n|KR|Y^n)
\]
\[= \frac{1}{n}H(K^n|KR) - \frac{1}{n}I(K^n; Y^n)
\]
\[\leq \frac{1}{n}H(K) + \frac{1}{n}H(R) - \frac{1}{n}I(K^n; Y^n)
\]
\[\leq H(Y) + 2\epsilon - \frac{1}{n}H(K^n|KR)
\]
\[= H(Y) + 2\epsilon \frac{1}{n}H(Y^n) + \frac{1}{n}H(Y^n|KR)
\]
\[= 2\epsilon + k_1(\epsilon).
\]
Thus, the third and the fourth coordinate are almost zero (they tend to zero as $\epsilon$ tends to zero). Letting $\epsilon$ converge to zero, we get an uncertainty vector of the form $[a, a, 0, 0]$. The constraint $C_6 \leq H(Y|X)$ comes from the fact that the maximum value of $a$ such that the uncertainty vector $[a, a, 0, 0]$ is in $U(p)$, is $a = H(Y|X)$. Using $C_5 + C_6 = H(Y)$ one can show that $C_5 \geq I(X; Y)$.

VI. USING “EDGE-CUTS” TO WRITE BETTER CONVERSES

The new converse as expressed in the previous sections is not also tight in general. In the above discussion, we observed that every cut that has the edge $\epsilon$ and separates the source from the first sink imposes a constraint on $I(\epsilon)$. However, it turns out that one can write strictly better converses by looking at what might be termed “edge-cuts” (certain cuts in certain subgraphs of the original graph) if there are multiple source nodes in the network. Our concept of edge-cuts should not be confused with that of [15].

In order to construct an explicit example for multisource problems that shows the benefit of using edge-cuts, we consider a directed network with two sources $s_1$ and $s_2$ and two sinks $t_1$ and $t_2$ of Fig. 5 under the assumption that $C_4 = C_7 = C_8$.

Suppose that the source $s_1$ observes i.i.d. copies of the random variable $X$, and source $s_2$ observes i.i.d. copies of $Y$. Note that, in general, allowing for stochastic encoding/randomness in coding operations inside the network can sometimes help in the case of secrecy problems (for instance see [30]). In this case, if there is residual capacity in a cut (the cut has a strictly larger value than the entropy of the source that one desires to reconstruct), the random variables in the cut may not be functions of the source. However, in this example, we focus exactly on a tight cut.
the random variable $Y$. As before, random variables $X$ and $Y$ are jointly distributed according to $p(x, y)$, and sink $s_1$ is interested in the i.i.d. copies of $X$, while sink $s_2$ is interested in the i.i.d. copies of $Y$. We consider the problem of reliable transmission to fulfill the demands of both sink nodes, with probability of decoding error converging to zero as the number of i.i.d. observations of $X$, $Y$ grows without bound.

One can verify that the cut-set bound is tight for the network of Fig. 5 if the sources are independent. Our discussion below indicates that if we slightly perturb the joint distribution of $p(x, y)$ to make $X$ and $Y$ correlated in such a way that the Wyner common information of $X$ and $Y$ is strictly larger than $I(X; Y)$, the cut-set bound and its extension discussed in the previous section will be both loose. The reason is that the individual edges (rather than their collective on a cut) will become the bottleneck.

### A. Edge-Cuts

Take an arbitrary edge $e$ in a directed graph from a vertex $v_1$ to a vertex $v_2$. Consider the subgraph formed by including all the directed paths from the two sources to $v_2$ that pass through edge $e$. We can think of $v_2$ as an imaginary sink in this subgraph. Let $K$ denote the random variable carried on the edge $e$. We can consider three types of cuts between the two sources and the imaginary sink in this subgraph: 1) cuts that separate the first source from node $v_2$ but do not separate the second source from node $v_2$; 2) cuts that separate the second source from $v_2$ but do not separate the first source from node $v_2$; and 3) cuts that separate both sources from node $v_2$. Let $Cut_{x,y,v_2}$ denote the sum-capacity of an arbitrary cut that separates both sources from node $v_2$ in the subgraph. By the cut-set bound, we have

$$Cut_{x,y,v_2} \geq \frac{1}{n} I(K; X^n, Y^n).$$

Let $Cut_{x,v_2}$ denote the sum-capacity of an arbitrary cut that separates the first source from node $v_2$ in the subgraph. By the cut-set bound, we have

$$Cut_{x,v_2} \geq \frac{1}{n} I(K; X^n | Y^n).$$

If $X$ and $Y$ are independent, the cut-set bound implies that

$$H(X) \leq C_4 + \min\{C_2, C_5, C_6, C_7\},$$

$$H(Y) \leq C_5 + \min\{C_3, C_4, C_6\},$$

$$H(X^n) + H(Y^n) \leq C_4 + C_5 + C_6.$$

This bound can be achieved as follows. If $H(X) < C_4$, we can achieve the above bound by using only edge 4 to transmit $X$ and using other part of the network to transmit $Y$. Similarly, if $H(Y) \leq C_5$, the above bound is achievable. Finally, if $H(X) > C_4$ and $H(Y) > C_5$, we split $s_1, s_2$ into two parts: the first part has capacity $C_4$ ($C_5$) and is transmitted through the edge 4 (5), while the second part is transmitted through the path formed by edges 2, 6, 7 (3, 6, 8). Since we can rewrite the above outer bound as

$$H(X) - C_4 \leq \min\{C_2, C_5, C_6, C_7\},$$

$$H(Y) - C_5 \leq \min\{C_3, C_4, C_6\},$$

$$H(X^n) + H(Y^n) - C_4 - C_5 \leq C_6,$$

we see that the second part of both sources can be achieved by splitting the capacity of edge 6 into two parts.

Similarly, let $Cut_{y,v_2}$ denote the sum-capacity of an arbitrary cut that separates the second source from node $v_2$ in the subgraph. We have

$$Cut_{y,v_2} \geq \frac{1}{n} I(K; Y^n | X^n).$$

These inequalities have consequences for the uncertainty vector $[1/n H(K); 1/n H(K | X^n); 1/n H(K | Y^n); 1/n H(K | X^n, Y^n)].$

Consider the edge 6 in Fig. 5. The resulting subgraph formed by including all the directed paths from the two sources to the end point of this edge is shown in Fig. 6. Let $K_6$ denote the random variable carried on this edge. Observe that edge 6 is a cut that separates the first source only from the imaginary sink. Therefore, we can write $1/n I(K_6; X^n Y^n) \leq C_2$ by (19). Since $H(K_6 | X^n, Y^n) = 0$, we conclude that $1/n H(K_6 | Y^n) \leq C_2$. It is not possible to get this constraint on the uncertainty of $K_6$ given $Y^n$ by looking at the cuts between the sources and the sinks in the original graph. To see this, note that if we use (8)–(10) for all the cuts that have the edge 6, we get the following set of equations:

$$d_6 = I(E_6; XY) \leq C_6$$

$$C_4 + C_5 - C_6 + d_6 - H(Y) \geq I(E_6; X | Y)$$

$$C_5 + C_6 - C_6 - H(Y) \geq I(E_6; X | Y)$$

for some $p(e_6, x, y)$, where the second inequality follows that $\{4, 6\}$ is a cut from $s_1, s_2$ to $t_1$ in the original graph, and the third inequality follows that $\{5, 6\}$ is a cut from $s_1, s_2$ to $t_2$ in the original graph.

The next step is to incorporate the inequality $1/n H(K_6 | Y^n) \leq C_2$ with the above set of inequalities. By Lemma 1,

$$C_5 + d_6 - H(Y) \geq \frac{1}{n} H(K_6 | Y^n),$$

where the LHS is the same to the LHS in the third inequality above. By Theorem 2,

$$I(E_6; X | Y) \leq \frac{1}{n} H(K_6 | Y^n).$$

Now, using the inequality $1/n H(K_6 | Y^n) \leq C_2$, we can conclude that

$$\min\{C_2, C_5 + d_6 - H(Y)\} \text{ is an upper bound on } 1/n H(K_6 | Y^n),$$

and hence, an upper bound on $I(E_6; X | Y)$.

Thus, we can write

$$d_6 = I(E_6; XY) \leq C_6$$

$$C_4 + d_6 - H(Y) \geq I(E_6; X | Y)$$

$$\min\{C_2, C_5 + d_6 - H(Y)\} \geq I(E_6; X | Y)$$

for some $p(e_6, x, y)$, where the second inequality follows that $\{4, 6\}$ is a cut from $s_1, s_2$ to $t_1$ in the original graph, and the third inequality follows that $\{5, 6\}$ is a cut from $s_1, s_2$ to $t_2$ in the original graph.
for some $p(e|z,y)$. This set of equations can be simplified in the following form:

\begin{align}
C_6 &\geq I(F_6;XY) \\
C_4 &\geq H(X|E_4) \\
C_5 &\geq H(Y|F_6) \\
C_2 &\geq I(E_6;X|Y)
\end{align}

(20) (21) (22) (23)

for some $p(e|z,y)$. Note that the first three inequalities are similar to those of the solution to the Gray–Wyner system [28]. However, there is no parallel for the last inequality.

**B. Comparison of Two Converses**

We now compare the converse derived by looking at all the cuts from the sources to the sinks (no edge-cuts) with the converse given by (20)–(23), which takes edge-cuts into consideration. The former converse is derived in the Appendix and given by (28)–(47).

The former converse enforces a lower bound on the value of $C_6$. More precisely, if we denote this lower bound on $C_6$ by $T$, this converse implies that when $C_6 < T$ the transmission of sources is impossible. We claim that $T \leq I(X;Y)$ if we restrict ourselves to networks where $C_2 + C_4 = H(X|Y)$; thus, this converse does not allow us to rule out the case $C_6 = I(X;Y)$. This claim is shown at the end of the Appendix. On the other hand, the converse written using edge-cuts and given by (20)–(23) allows us to show that for most joint distributions $p(x,y)$, transmission of sources is impossible when $C_6 = I(X;Y)$. Therefore, the edge-cut converse is strictly better.

To show this, we prove that the edge-cut converse implies that $C_6$ is greater than or equal to $\min_{X \rightarrow E_6 \rightarrow Y} I(E_6;X|Y)$, i.e., Wyner’s common information of $X$ and $Y$. From (21) and (23), we have

\begin{align*}
C_2 + C_4 &\geq H(X|E_6) + I(E_6;X|Y) \\
&= H(X|E_6) + I(X;Y) - H(X|E_6,Y) \\
&= H(X|Y) + I(X;Y,E_6).
\end{align*}

Since we restrict ourselves to networks where $C_2 + C_4 = H(X|Y)$, it must be the case that $I(X;Y|E_6) = 0$, i.e., random variables $X \rightarrow E_6 \rightarrow Y$ form a Markov chain. Therefore, the minimum of $C_6$ in the outer region of the converse is $\min_{X \rightarrow E_6 \rightarrow Y} I(E_6;X,Y)$, which is equal to Wyner’s common information. Noting that Wyner’s common information is, in general, larger than $I(X;Y)$, we conclude that the latter converse is strictly better than the former converse.

**VII. PROOFS**

**Lemma 3:** Given any three random variables $X,Y,K$ where $K$ is a function of $(X,Y)$, we have

\begin{align*}
I(K;X) &\geq [H(K) - H(Y;X)]_+ \\
I(K;Y) &\geq [H(K) - H(X;Y)]_+
\end{align*}

where $|x|_+$ is 0 when $x$ is negative and $x$ when it is nonnegative.

**Proof:** We prove the first equation. The proof for the second one is similar. It suffices to show that $I(K;X) \geq H(K) - H(Y;X)$, which is equivalent with $H(Y;X) \geq H(K,X)$ and obviously true.

**Proof of Theorem 1:**

**Achievability:** We begin by showing that each of the four set of points is a subset of $U(p)$. This would complete the proof noting that $U(p)$ is a convex set in $R^4$ as it implies that the convex envelope of the union of the four sets of points is also a subset of $U(p)$.

\[
\lim_{n \to \infty} \frac{1}{n} H(K_n) = I(E;X,Y),
\]

\[
\lim_{n \to \infty} \frac{1}{n} H(K_n|X^n) = I(E;X|Y),
\]

\[
\lim_{n \to \infty} \frac{1}{n} H(K_n|Y^n) = I(E;Y|X),
\]

\[
\lim_{n \to \infty} \frac{1}{n} H(K_n|X^n,Y^n) = 0.
\]

\text{We use part 1 of Theorem 5 of [31] which says that one can find a sequence of $p(k_n,x^n,y^n)$ such that}

\[
\lim_{n \to \infty} \frac{1}{n} I(X^n,Y^n;K_n) = I(X;Y,E),
\]

\[
\lim_{n \to \infty} \frac{1}{n} H(K_n|X^n) = I(E;X|Y),
\]

\[
\lim_{n \to \infty} \frac{1}{n} H(K_n|Y^n) = I(E;Y|X),
\]

\[
\lim_{n \to \infty} \frac{1}{n} H(K_n|X^n,Y^n) = 0.
\]

The difference between these set of equations and the ones we would like to have is the first one. However, these four set of equations are indeed equivalent. Note that

\[
H(K_n) = H(K_n|X^n) + H(K_n|Y^n) - H(K_n|X^n,Y^n) = I(X^n;Y^n) - I(X^n;Y^n;K_n).
\]

Thus,

\[
\lim_{n \to \infty} \frac{1}{n} H(K_n) = \lim_{n \to \infty} \frac{1}{n} H(K_n|X^n) + \lim_{n \to \infty} \frac{1}{n} H(K_n|Y^n)
\]

\[
- \lim_{n \to \infty} \frac{1}{n} I(X^n;Y^n|K_n)
\]

\[
= I(E;X|Y) + I(E;Y|X)
\]

\[
+ I(X;Y) - I(X;Y|E)
\]

\[
= I(E;X,Y).
\]

\text{To see that $U(p)$ is a convex set in $R^4$, it suffices to show that $U(p)$ contains the convex hull of point in $U(p)$ that arise from joint distributions $(k_n,x^n,y^n)$ for some $p(k_n,x^n,y^n)$. Take two arbitrary points in $U(p)$ that arise from joint distributions $(k_n,x^n,y^n)$ and $(k'_n,x'^n,y'^n)$. Construct a sequence $(k_m,x''_m,y''_m)$ for $m = 1,2,3,\ldots$ as follows: take $n_m$ i.i.d. copies of $X''_m$, $Y''_m$ and some binary random variable $Q$ independent of that. If $Q = 0$ let $K_m = k_m$ be $n_m$ i.i.d. copies of $K$. If $Q = 1$ let $K_m = k'_m$ be $n_m$ i.i.d. copies of $K$. The sequence of points corresponding to the sequence $(k_m,x''_m,y''_m)$ converges, as $n_m$ converges to infinity, to the middle point of the two original points in $U(p)$ that we started with. Thus, $U(p)$ is a convex set in $R^4$.}
We now prove that the second and the third set of points are in $U(p)$. Slepian–Wolf tell us that for any $\epsilon$, one can find $N$ such that for any $n > N$, there are functions $M_{n_1}: X^n \to [1 : 2^n[\mathbb{H}(X|Y)+\epsilon)]$ and $M_{n_2}: Y^n \to [1 : 2^n[\mathbb{H}(Y|X)+\epsilon)]$ such that $X^n$ can be recovered from $(M_{n_1}(X^n), Y^n)$, and $Y^n$ can be recovered from $(M_{n_2}(X^n), Y^n)$ with probability $1-\epsilon$. One can prove that for some functions $r_1$ such that $r_1(\epsilon)$ converges to zero as $\epsilon$ converges to zero. Setting $K_n = M_{n_1}(Y^n)$ would give us the second set of points as $n \to \infty$. To see this, note that $\lim_{n \to \infty} \frac{1}{n} I(K_n|X^n) = I(Y|X)$ because of (27) and the fact that $M_{n_1}$ is taking value in $[1 : 2^n[\mathbb{H}(Y|X)+\epsilon)]$. Furthermore, one can show that $\lim_{n \to \infty} \frac{1}{n} I(K_n, X^n) = I(Y|X)$ using (25). Similarly, setting $K_n = M_{n_2}(X^n)$ asymptotically gives us the third set of points.

We now prove that the fourth set of points is in $U(p)$. In order to define $K_n$ appropriately to get this set of points, we are going to use random variables $M_{n_1}$ and $M_{n_2}$ defined above. For every $n \in \mathbb{N}$, we can find some $\epsilon_n$ such that (24)–(27) hold, and that $\epsilon_n$ converges to zero as $n$ goes to infinity. Next, take some arbitrary $0 \leq \epsilon \leq \max I(Y|X), I(X|Y)$. We would like to find a sequence of pairs $p(k_n, x^n, y^n)$ such that

$$\lim_{n \to \infty} \frac{1}{n} H(K_n) = f,$$

$$\lim_{n \to \infty} \frac{1}{n} H(K_n|X^n) = \min(f, H(Y|X)),$$

$$\lim_{n \to \infty} \frac{1}{n} H(K_n|Y^n) = \min(f, H(X|Y)),$$

$$\lim_{n \to \infty} \frac{1}{n} H(K_n, X^n, Y^n) = 0.$$

Let us define the functions $M_{n_1} \in [1 : 2^n[\mathbb{H}(X|Y)+\epsilon_n]$ and $M_{n_2} \in [1 : 2^n[\mathbb{H}(Y|X)+\epsilon_n]$ as above. We can think of $M_{n_1}(X^n)$ and $M_{n_2}(Y^n)$ as two random binary sequences of length $\lceil n[H(Y|X) + \epsilon_n] \rceil$ and $\lceil n[H(Y|X) + \epsilon_n] \rceil$, respectively. Let us use the notation $M_{n_2}(X^n)$ to denote the set of the $i^{th}$ to $j^{th}$ bits of $M_{n_2}(Y^n)$. We use a similar notation for $M_{n_1}(X^n)$.

Without loss of generality, let us assume that $H(X|Y) \geq H(Y|X)$. Consider the following two cases:

**Case 1.** $f < H(Y|X)$: In this case, we let $K_n$ be equal to the bitwise XOR of the first $\lceil n f \rceil$ bits of $M_{n_1}(X^n)$ and $M_{n_2}(Y^n)$, i.e., the bitwise XOR of $M_{n_1}^{\lceil n f \rceil}(X^n)$ and $M_{n_2}^{\lceil n f \rceil}(Y^n)$. Clearly, $\frac{1}{n} H(K_n|X^n, Y^n) = 0$. We would like to show that

$$\lim_{n \to \infty} \frac{1}{n} H(K_n) = f,$$

$$\lim_{n \to \infty} \frac{1}{n} H(K_n|X^n) = f,$$

$$\lim_{n \to \infty} \frac{1}{n} H(K_n|Y^n) = f.$$

It suffices to prove the last two inequalities since $H(K_n|X^n) \leq H(K_n) \leq \log |K_n| \leq nf$. We prove the second one, and the proof for the third is similar. Note that

$$H(K_n|X^n) = H(K_n, X^n, M_{n_1}^{\lceil n f \rceil}(X^n)) = H(M_{n_1}^{\lceil n f \rceil}(Y^n), X^n).$$

Equation (25) implies

$$\frac{1}{n} I(M_{n_1}^{\lceil n f \rceil}(Y^n); X^n) \leq r_2(\epsilon_n).$$

Thus,

$$\lim_{n \to \infty} \frac{1}{n} H(K_n|X^n) = \lim_{n \to \infty} \frac{1}{n} H(M_{n_1}^{\lceil n f \rceil}(Y^n)).$$

Clearly $\lim_{n \to \infty} \frac{1}{n} H(M_{n_1}^{\lceil n f \rceil}(Y^n)) \leq f$. If $\lim_{n \to \infty} \frac{1}{n} H(M_{n_1}^{\lceil n f \rceil}(Y^n)) < f$, then additionally considering the $\lceil nf \rceil + 1$ to $\lfloor nH(Y|X) + nf \rfloor$ bits of $M_{n_1}$ can at most increase the asymptotic entropy rate by $H(Y|X) - f$ bits. On the other hand, (27) implies that $\lim_{n \to \infty} \frac{1}{n} H(M_{n_2}(Y^n)) = H(Y|X)$. This is a contradiction because using the fact that the joint entropy is less than or equal to the individual entropies, one can write

$$\lim_{n \to \infty} \frac{1}{n} H(M_{n_1}(Y^n)) \leq \lim_{n \to \infty} \frac{1}{n} H(M_{n_1}^{\lceil n f \rceil}(Y^n)) + \lim_{n \to \infty} \frac{1}{n} H(M_{n_2}^{\lceil n f \rceil+1: \lceil nH(Y|X)+nf \rceil}(Y^n)) \leq f + n - f = n.$$

**Case 2.** $H(Y|X) \leq f \leq H(Y|X)$: In this case, let $K_n$ be equal to the bitwise XOR of $M_{n_1}^{\lceil nH(Y|X) \rceil}(X^n)$ and $M_{n_2}^{\lceil nH(Y|X) \rceil}(Y^n)$, together with $M_{n_2}^{\lceil nH(Y|X) \rceil+1: \lceil nH(Y|X) \rceil}(X^n)$. In this case, one needs to show that

$$\lim_{n \to \infty} \frac{1}{n} H(K_n) = f,$$

$$\lim_{n \to \infty} \frac{1}{n} H(K_n|X^n) = nH(Y|X),$$

$$\lim_{n \to \infty} \frac{1}{n} H(K_n|Y^n) = f.$$

As in case 1, the third equation implies the first. The proof for the last two limits is similar to the one discussed above in case 1.

**Converse:** Since $U(p)$ is convex, to show that the region $U(p)$ is equal to the convex envelope of the given set of points, it suffices to show that for any real $\lambda_1, \ldots, \lambda_d$, the maximum of $\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4$ over $U(p)$ is achieved at one of
the given points. We show this by a case by case analysis. First assume that \( \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 > 0 \). In this case, maximum will be infinity and is achieved at the point \( |c, c, c, c| \) when \( c \to \infty \). If \( \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \leq 0 \), we can write the maximum of \( \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4 \) over \( U(p) \) as

\[
\limsup_{n \to \infty} \frac{1}{n} \left( \lambda_1 I(K; X^n Y^n) + \lambda_2 I(K; Y^n | X^n) + \lambda_3 I(K; X^n Y^n) + (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)H(K | X^n, Y^n) \right).
\]

The last term \((\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)H(K | X^n, Y^n)\) is less than or equal to zero. Given any \((K, X^n, Y^n)\), we can always use part 1 of Theorem 5 of [31] as in the achievability to find \((K', X'^{mn}, Y'^{mn})\) for some \( m \) such that \( K' \) is a function of \((X'^{mn}, Y'^{mn})\) and sum of the first three terms is asymptotically unchanged. \( K' \) being a function of \((X'^{mn}, Y'^{mn})\) implies that \((\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)H(K' | X'^{mn}, Y'^{mn})\) is zero. To sum up, without loss of generality, we can consider only random variables \( K \) that are deterministic functions of \((X^n, Y^n)\), and furthermore, we only need to compute the following expression over such random variables:

\[
\limsup_{n \to \infty} \frac{1}{n} \left( \lambda_1 I(K; X^n Y^n) + \lambda_2 I(K; Y^n | X^n) + \lambda_3 I(K; X^n Y^n) \right).
\]

We now continue by a case-by-case analysis:

1) \( \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0 \): Note that if we replace \( K \) with \((K, X^n, Y^n)\), the expression will not decrease. Since \( K \) is a function of \((X^n, Y^n)\), we conclude that \( K = X^n Y^n \) is the optimal choice in this instance. In the case, the maximum of \( \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4 \) over \( U(p) \) will be equal to the maximum of the same expression over the first set of points with the choice of \( K = X^n Y^n \). To see this, write \( \lambda_2 I(K; Y^n | X^n) \) as \( \lambda_2 I(K; X^n, Y^n) - \lambda_2 I(K; X^n) \) and note that the expression is maximized when \( K = X^n Y^n \).

If \( \lambda_1 + \lambda_2 \leq 0 \) first note that if we replace \( K \) with \((K, X^n, Y^n)\), the expression will not decrease. In this case, the expression \( \lambda_1 I(K, X^n; X^n Y^n) + \lambda_2 I(K, X^n; Y^n X^n) + \lambda_3 I(K, X^n; X^n Y^n) \) will be equal to \( \lambda_1 H(X^n) + \lambda_3 I(K; X^n Y^n) + (\lambda_1 + \lambda_2)I(K; Y^n | X^n) \). Since \( \lambda_1 + \lambda_2 < 0 \), we have \( \lambda_1 + \lambda_2 I(K; Y^n | X^n) < 0 \). Thus, the maximum of \( \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4 \) over \( U(p) \) will be less than or equal to \( \lambda_1 H(X^n) + \lambda_3 I(K; X^n Y^n) \), which is equal to the maximum of the same expression over the first set of points with the choice of \( K = X^n Y^n \).

3) \( \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \leq 0 \): This case is similar to case 2 by symmetry.

4) \( \lambda_1 > 0, \lambda_2 < 0, \lambda_3 < 0 \): Take some arbitrary \( n \) and \( K = f(X^n, Y^n) \). Let the random index \( J \) be uniformly distributed on \( \{1, 2, 3, \ldots, n\} \) and independent of \((K, X^n, Y^n)\). Define the auxiliary random variables

\[
E = (K, X_{j-1}, Y_{j-1}, J), X = X_j, Y = Y_j.
\]

Note that

\[
I(K; X^n, Y^n) = \sum_{j=1}^{n} I(K; X_j, Y_j | X_{1:j-1}, Y_{1:j-1})
\]

\[
= \sum_{j=1}^{n} I(K, X_{1:j-1}, Y_{1:j-1}; X_j, Y_j)
\]

\[
= n I(E; X, Y),
\]

\[
I(K; Y^n | X^n) = \sum_{j=1}^{n} I(K, Y_{j+1}; X_{j+1} - Y_{j+1}; X_j, Y_j)
\]

\[
\geq \sum_{j=1}^{n} I(K, X_{1:j-1}; Y_j, X_j)
\]

\[
= n I(E; Y X),
\]

and similarly

\[
I(K; X^n Y^n) \geq n I(E; X Y).
\]

Since \( \lambda_2 \leq 0, \lambda_3 \leq 0 \), we have \( \lambda_2 \frac{1}{n} I(K; Y^n | X^n) \leq I(E; Y | X) \) and \( \lambda_3 \frac{1}{n} I(K; X^n Y^n) \leq I(E; X | Y) \). Therefore, the maximum of \( \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4 \) over \( U(p) \) will be less than or equal to the maximum of the same expression over the first set of points.

5) \( \lambda_1 \leq 0, \lambda_2 \geq 0, \lambda_3 \leq 0 \): If \( \lambda_1 + \lambda_2 \geq 0 \), we can write

\[
\lambda_1 I(K; X^n Y^n) + \lambda_2 I(K; Y^n | X^n) + \lambda_3 I(K; X^n Y^n)
\]

\[
\leq \lambda_1 I(K; X^n Y^n) + \lambda_2 I(K; Y^n | X^n)
\]

\[
= \lambda_1 I(K; X^n Y^n) + \lambda_1 \lambda_2 I(K; Y^n | X^n)
\]

\[
\leq (\lambda_1 + \lambda_2) I(K; X^n Y^n)
\]

\[
\leq (\lambda_1 + \lambda_2) H(Y^n | X^n).
\]

Thus, the maximum of \( \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4 \) over \( U(p) \) will be less than or equal to the maximum of \( (\lambda_1 + \lambda_2) H(Y^n | X^n) \), which is equal to the maximum of the same expression over the second set of points.

6) \( \lambda_1 \leq 0, \lambda_2 \leq 0, \lambda_3 \geq 0 \): This is similar to case 5.

7) \( \lambda_1 \leq 0, \lambda_2 \leq 0, \lambda_3 \leq 0 \): This is similar to case 4.

8) \( \lambda_1 < 0, \lambda_2 > 0, \lambda_3 > 0 \): If \( \lambda_1 + \lambda_2 + \lambda_3 < 0 \),

\[
\lambda_1 I(K; X^n Y^n) + \lambda_2 I(K; Y^n | X^n) + \lambda_3 I(K; X^n Y^n)
\]

\[
= (\lambda_1 + \lambda_2 + \lambda_3) I(K; X^n Y^n) - \lambda_2 I(K; X^n)
\]

\[
- \lambda_3 I(K; Y^n)
\]

\[
\leq 0.
\]
Thus, $K$ constant works here. If $\lambda_1 + \lambda_2 + \lambda_3 \geq 0$, using Lemma 3, we have 

$$\lambda_1 I(K; X^n Y^n) + \lambda_2 I(K; Y^n X^n) + \lambda_3 I(K; X^n Y^n)$$

$$= (\lambda_1 + \lambda_2 + \lambda_3) I(K; X^n Y^n) - \lambda_2 I(K; X^n) - \lambda_3 I(K; Y^n)$$

$$\leq (\lambda_1 + \lambda_2 + \lambda_3) I(K; X^n Y^n)$$

$$- \lambda_2 |I(K; X^n Y^n) - H(Y^n | X^n)| +$$

$$- \lambda_3 |I(K; X^n Y^n) - H(X^n Y^n)| +$$

$$= n \left( \lambda_1 \frac{I(K; X^n Y^n)}{n} + \lambda_2 \min \left( \frac{I(K; X^n Y^n)}{n}, H(Y | X) \right) + \lambda_3 \min \left( \frac{I(K; X^n Y^n)}{n}, H(X | Y) \right) \right).$$

Thus, the maximum of the original expression is less than or equal to

$$\max_{0 \leq t \leq H(Y | X)} \left( \lambda_1 t + \lambda_2 \min \left( t, H(Y | X) \right) \right)$$

$$+ \lambda_3 \min \left( t, H(Y | X) \right)$$

$$\leq \max_{0 \leq t \leq \max_{H(Y \mid X)} H(Y | X)} \left( \lambda_1 t + \lambda_2 \min \left( t, H(Y | X) \right) \right)$$

$$+ \lambda_3 \min \left( t, H(Y | X) \right).$$

Thus, the maximum of $\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4$ over $U[p]$ will be less than or equal to the maximum of the same expression over the fourth set of points.

**Proof of Theorem 2:** Take some $n$ and $p(k|x^n, y^n)$ and consider the 4-tuples $(u_1, u_2, u_3, u_4)$

$$u_1 = \frac{1}{n} H(K),$$

$$u_2 = \frac{1}{n} H(K X^n),$$

$$u_3 = \frac{1}{n} H(K Y^n),$$

$$u_4 = \frac{1}{n} H(K X^n Y^n).$$

Let $e = \frac{1}{n} H(K X^n, Y^n)$. Let the random index $J$ be uniformly distributed on $\{1, 2, 3, \ldots, n\}$ and independent of $(K, X^n, Y^n)$. Define the auxiliary random variables $E = (K, X, Y, J), X = X_j, Y = Y_j$. One can then verify that

$$I(K; X^n, Y^n) = n I(E; X, Y),$$

$$I(K; Y^n | X^n) \geq n I(E; Y | X),$$

$$I(K; X^n Y^n) \geq n I(E; X | Y).$$

Thus, $u_1 = e + I(E; X, Y), u_2 \geq e + I(E; Y, X)$ and $u_3 \geq e + I(E; Y, X)$ for some $p(e|x, y)$.

**APPENDIX**

**USING THE CUTS TO WRITE A CONVERSE**

In this appendix, we use cuts between sources and sinks to write a converse for the network of Fig. 5. Since there are two sources and two sinks in this network there are more types of cuts to consider. Every cut divides the nodes of the network into two sets $A$ and $A^c$. We use the notation cut$(s_i)$ for sources in $A$; sinks in $A^c$ to denote the edges of such a cut. For instance, in Fig. 5, $\{4, 2\}$ is cut$(s_1, s_2, t_1, t_2)$ meaning that edges 4 and 2 are the edges of a cut that has $s_1$ in $A$, $s_2$ in $A^c$ and sinks $t_1, t_2$ in $A^c$. Suppose we want to write the converse for an edge $e$ in cut$(s_i)$ for sources in $A$, sinks in $A^c$.

**Lemma 1 (Revisited):** Take an arbitrary cut containing $e$ from the first source to the first sink, and let $Cut_e$ denote the sum of the capacities of the edges on this cut. Further assume that $s_2$ is in $A^c$. Then, $\frac{1}{n} H(K, X^n)$ must satisfy the following inequalities:

$$- \frac{1}{n} H(K | X^n) \leq Cut_e - C_e + d_e + H(Y) - H(X) + k(e)$$

$$\frac{1}{n} H(K, X^n, Y^n) \leq Cut_e - C_e + d_e + H(Y) - H(X, Y) + k(e)$$

for some functions $k(e)$ that converges to zero as $e$ converges to zero.

**Proof:** Let $Q$ denote the collection of random variables passing over the edges of the cut (except $e$). Clearly $\frac{1}{n} H(Q) \leq Cut_e - C_e + m\epsilon$, where $m$ is the number of edges in the graph. Since $(Q, K)$ is the collection of the random variables passing the edges of the cut, $X^n$ should be recoverable from $(Q, K, Y^n)$ with probability of error less than or equal to $\epsilon$. Thus, by Fano’s inequality, $\frac{1}{n} H(E | Q, K, Y^n) \leq k_1(\epsilon)$ for some function $k_1(\epsilon)$ that converges to zero as $\epsilon$ converges to zero.

We have

$$- \frac{1}{n} H(K | X^n) \leq - \frac{1}{n} H(Q, K, Y^n | X^n)$$

$$= - \frac{1}{n} H(Q, K, Y^n, X^n) - \frac{1}{n} H(X^n)$$

$$\leq - \frac{1}{n} H(Q) + \frac{1}{n} H(K) + H(Y)$$

$$+ \frac{1}{n} H(X^n | Q, K, Y^n) - H(X, Y)$$

$$\leq Cut_e - C_e + H(Y)$$

$$+ m\epsilon + d_e - H(X, Y) + k_1(\epsilon).$$

We get the first inequality by setting $k(\epsilon) = k_1(\epsilon) + m\epsilon$. For the second inequality, note that

$$- \frac{1}{n} H(K | X^n, Y^n) \leq - \frac{1}{n} H(Q, K | X^n, Y^n)$$

$$= - \frac{1}{n} H(Q, K, Y^n, X^n) - \frac{1}{n} H(X^n, Y^n)$$

$$\leq - \frac{1}{n} H(Q) + \frac{1}{n} H(K) + H(Y)$$

$$+ \frac{1}{n} H(X^n | Q, K, Y^n) - H(X, Y, Y)$$

$$\leq Cut_e - C_e + H(Y)$$

$$+ m\epsilon + d_e - H(X, Y) + k_1(\epsilon).$$
We can now write down the converse using the edge-cuts. We proceed in a similar fashion that we did in deriving (8)--(10) using Lemma 1 (revisited) and Theorem 2. Lemma 1 (revisited) gives us upper bounds on the elements of the uncertainty vector, whereas Theorem 2 gives us lower bounds on these elements.

Cuts that have edge 2:
\[ d_2 = I(E_2; XY) \leq C_2 \]
\[ C_2 + C_6 - C_5 + d_2 + H(X) - H(Y) \geq I(E_2; Y|X) \]
\[ C_2 + C_4 - C_2 + d_2 + H(Y) - H(X,Y) \geq 0 \]
\[ \text{because } \{2, 4\} \text{ is cut}(s_1; s_2; t_1, t_2) \]
\[ C_2 + C_4 + C_5 - C_2 + d_2 - H(X,Y) \geq 0 \]
\[ \text{because } \{2, 3, 4; 5\} \text{ is cut}(s_1, s_2; \emptyset; t_1, t_2) \]

for some \( p(e_2, x, y) \).

Cuts that have edge 3:
\[ d_3 = I(E_3; XY) \leq C_3 \]
\[ C_3 + C_6 - C_5 + d_3 + H(X) - H(Y) \geq I(E_3; X|Y) \]
\[ C_3 + C_6 - C_5 + d_3 + H(X) - H(X,Y) \geq 0 \]
\[ \text{because } \{3, 5\} \text{ is cut}(s_2; s_1; t_1, t_2) \]
\[ C_2 + C_3 + C_4 + C_5 - C_3 + d_3 - H(X,Y) \geq 0 \]
\[ \text{because } \{2, 3, 4, 5\} \text{ is cut}(s_1, s_2; \emptyset; t_1, t_2) \]

for some \( p(e_3, x, y) \).

Cuts that have edge 4:
\[ d_4 = I(E_4; XY) \leq C_4 \]
\[ C_4 + C_6 - C_5 + d_4 + H(X) - H(Y) \geq I(E_4; Y|X) \]
\[ C_2 + C_4 - C_4 + d_4 + H(Y) - H(X,Y) \geq 0 \]
\[ \text{because } \{2, 4\} \text{ is cut}(s_1; s_2; t_1, t_2) \]
\[ C_2 + C_4 + C_5 - C_2 + d_4 - H(X,Y) \geq 0 \]
\[ \text{because } \{2, 3, 4, 5\} \text{ is cut}(s_1, s_2; \emptyset; t_1, t_2) \]
\[ C_4 + C_5 + C_6 - C_4 + d_4 - H(X,Y) \geq 0 \]
\[ \text{because } \{4, 5, 6\} \text{ is cut}(s_1, s_2; \emptyset; t_1, t_2) \]
\[ C_4 + C_7 - C_4 + d_4 - H(X) \geq I(E_4; Y|X) \]
\[ \text{because } \{4, 7\} \text{ is cut}(s_1, s_2; \emptyset; t_1) \]

for some \( p(e_4, x, y) \).

Cuts that have edge 5:
\[ d_5 = I(E_5; XY) \leq C_5 \]
\[ C_5 + C_6 - C_5 + d_5 + H(X) - H(Y) \geq I(E_5; X|Y) \]
\[ C_5 + C_4 - C_6 + d_5 + H(X) - H(X,Y) \geq 0 \]
\[ \text{because } \{3, 5\} \text{ is cut}(s_2; s_1; t_1, t_2) \]
\[ C_2 + C_3 + C_4 + C_5 - C_5 + d_5 - H(X,Y) \geq 0 \]
\[ \text{because } \{2, 3, 4, 5\} \text{ is cut}(s_1, s_2; \emptyset; t_1, t_2) \]
\[ C_4 + C_5 + C_6 - C_5 + d_5 - H(X,Y) \geq 0 \]
\[ \text{because } \{4, 5, 6\} \text{ is cut}(s_1, s_2; \emptyset; t_1, t_2) \]
\[ C_5 + C_7 - C_5 + d_5 - H(Y) \geq I(E_5; X|Y) \]
\[ \text{because } \{5, 8\} \text{ is cut}(s_1, s_2; \emptyset; t_2) \]

for some \( p(e_5, x, y) \).

Since the capacities of edges 6, 7, and 8 are all the same, we can assume that they are all carrying the same message. Therefore, we can compute the uncertainty of the message on edge 6 by looking at cuts that include edge 7 or 8:
\[ d_6 = I(E_6; XY) \leq C_6 \]
\[ C_4 + C_5 + C_6 - C_6 + d_6 - H(X,Y) \geq 0 \]
\[ \text{because } \{4, 5, 6\} \text{ is cut}(s_1, s_2; \emptyset; t_1, t_2) \]
\[ C_4 + C_6 - C_6 + d_6 + H(Y) - H(X,Y) \geq I(E_6; Y|X) \]
\[ C_4 + C_6 - C_6 + d_6 + H(Y) - H(X,Y) \geq 0 \]
\[ \text{because } \{4, 6\} \text{ is cut}(s_1; s_2; t_1, t_2) \]
\[ C_5 + C_6 - C_6 + d_6 + H(X) - H(Y) \geq I(E_6; X|Y) \]
\[ C_5 + C_6 - C_6 + d_6 + H(X) - H(X,Y) \geq 0 \]
\[ \text{because } \{5, 6\} \text{ is cut}(s_2; s_1; t_1, t_2) \]
\[ C_4 + C_6 - C_6 + d_6 - H(X) \geq I(E_6; Y|X) \]
\[ \text{because } \{4, 7\} \text{ is cut}(s_1; s_2; \emptyset; t_1) \]
\[ C_5 + C_6 - C_6 + d_6 - H(X) \geq 0 \]
\[ \text{because } \{5, 7, 8\} \text{ is cut}(s_1; s_2; \emptyset; t_1, t_2) \]

for some \( p(e_6, x, y) \). After simplification and removal of redundant equations and noting that \( C_6 = C_7 = C_8 \), these inequalities can be written as follows:

\[ I(E_2; X|Y) \leq C_2 \] (28)
\[ C_4 \geq H(X|Y|E_2) - H(Y) \] (29)
\[ C_3 + C_4 + C_6 \geq H(X, Y|E_2) \] (30)

\[ \text{From equations for edge 2:} \]
\[ I(E_3; X|Y) \leq C_3 \] (31)
\[ C_5 \geq H(X|Y|E_3) - H(X) \] (32)
\[ C_2 + C_4 + C_6 \geq H(X, Y|E_3) \] (33)

\[ \text{From equations for edge 3:} \]
\[ I(E_4; X|Y) \leq C_4 \] (34)
\[ C_2 \geq H(X|Y|E_4) - H(Y) \] (35)
\[ C_2 + C_5 + C_6 \geq H(X, Y|E_4) \] (36)
\[ C_5 + C_6 \geq H(X|Y|E_4) \] (37)
\[ C_5 \geq H(X|E_4) \] (38)

\[ \text{From equations for edge 4:} \]
\[ I(E_5; X|Y) \leq C_5 \] (39)
\[ C_3 \geq H(X|Y|E_5) - H(X) \] (40)
\[ C_2 + C_5 + C_6 \geq H(X, Y|E_5) \] (41)
\[ C_4 + C_6 \geq H(X, Y|E_5) \] (42)
\[ C_6 \geq H(X|E_5) \] (43)

\[ \text{From equations for edge 5:} \]
\[ I(E_6; X|Y) \leq C_6 \] (44)
\[ C_4 + C_5 \geq H(X, Y|E_6) \] (45)
\[ C_4 \geq H(X|E_6) \] (46)
\[ C_5 \geq H(Y|E_6) \] (47)

\[ \text{From equations for edge 6:} \]

for some \( p(e_2, e_3, e_4, e_5, e_6|x, y) \).
This converse enforces a lower bound on the value of $C_{B}$. More precisely, if we denote this lower bound on $C_{B}$ by $T$, this converse implies that when $C_{B} < T$, the transmission of sources is impossible. We claim that $T \leq I(X;Y)$ if we restrict ourselves to networks where $C_{B} + C_{A} = H(X,Y)$; thus, this converse does not allow us to rule out the case $C_{B} = I(X;Y)$. This is because the choice of $C_{B} = 0$, $C_{A} = H(Y)$, $C_{A} = H(X,Y)$, $C_{B} = H(X,Y)$ and $C_{B} = I(X;Y)$ is a valid point in this converse region. To see this, take $E_0$ in a way that $E_0 \rightarrow X \rightarrow Y$ forms a Markov chain, and furthermore, $p(e_0, x) \sim p(y | x)$. Take $E_1$ in a way that $E_1 \rightarrow X \rightarrow Y$ forms a Markov chain, and furthermore, $I(E_1, X) = H(X,Y)$. Take $E_2 = (X,Y)$, $E_3 = Y$ and $E_2$ = constant. To verify these equations, it is useful to note that since $C_{B} = H(X,Y)$ those equations involving $C_{B}$ will be automatically satisfied. Because $E_0 \rightarrow X \rightarrow Y$ forms a Markov chain and $p(e_0, x) \sim p(y | x)$, we have $I(E_0; X,Y) = I(E_1; X) = I(Y; X)$.

ACKNOWLEDGMENT

The authors would like to thank R. Yeung for insightful discussions, his comments on the manuscript and continued support throughout the research. The authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] A. A. Gohari, S. Yang, and S. Jaggi, “Beyond the cut-set bound: Uncertainty computations in network coding with correlated sources,” in Proc. Int. Symp. Inf. Theory, 2011, pp. 598–602.

[2] T. M. Cover, Elements of Information Theory, 2nd ed. New York, NY: Wiley, 2005.

[3] C. K. Ngai and R. W. Yeung, “Multisource network coding with two sinks,” in Proc. Int. Conf. Commun., Circuits Syst., 2004, vol. 1, pp. 34–37.

[4] E. Erez and M. Feder, “Capacity region and network codes for two receivers multicast with private and common data,” presented at the Workshop Coding, Cryptog. Comb., 2003.

[5] T. Ho, M. Medard, M.Effros, and R. Koetter, “Network coding for correlated sources,” presented at the Conf. Inf. Sci. Syst., 2004.

[6] J. Barros and S. D. Servetto, “Network information flow with correlated sources,” IEEE Trans. Inf. Theory, vol. 52, no. 1, pp. 155–170, Jan. 2006.

[7] A. Ramamoorthy, K. Jain, P. A. Chou, and M. Effros, “Separating distributed source coding from network coding,” IEEE Trans. Inf. Theory, vol. 52, no. 6, pp. 2785–2795, Jun. 2006.

[8] M. Effros, M. Médard, T. Ho, S. Ray, D. Karger, and R. Koetter, “Linear network codes: A unified framework for source, channel and network coding,” presented at the DIMACS Workshop Network Inf. Theory, Picataway, NJ, 2003.

[9] T. P. Coleman, E. Martinian, and E. Ordentlich, “Joint source-channel coding for transmitting correlated sources over broadcast networks,” IEEE Trans. Inf. Theory, vol. 55, no. 8, pp. 3864–3868, Aug. 2009.

[10] H. Park, N. Thomas, and P. Frossard, “Transmission of correlated information sources with network coding,” presented at the 18th Eur. Signal Process. Conf., Aalborg, Denmark, Aug. 2010.

[11] A. Lee, M. Médard, K. Z. Haigh, S. Gowan, and P. Rubel, “Minimum-cost subgraphs for joint distributed source and network coding,” presented at the Inf. Theory Appl. Workshop, San Diego, CA, 2007.

[12] Y. Wu, V. Stankovic, Z. Xiong, and S.-Y. Kung, “On practical design for joint distributed source and network coding,” presented at the Int. Symp. Netw. Cod., 2005.

[13] T. Cui, L. Chen, and T. Ho, “Distributed distortion optimization for correlated sources with network coding,” IEEE Trans. Commun., vol. 60, no. 5, pp. 1336–1344, May 2012.

[14] T. Shimizu, T. Matsumoto, H. Murata, K. Yamamoto, and S. Yoshida, “Code design for joint decoding of correlated sources using algebraic network coding over awgn channels,” in Proc. 7th Int. Conf. Inf. Signal Process., Macau, Dec. 2009, pp. 1–5.

[15] G. Kramer and S. Savari, “Edge-cut bounds on network coding rates,” J. Netw. Syst. Manag., vol. 14, pp. 49–67, 2006.

[16] N. J. A. Harvey, R. Kleinberg, and A. R. Lehman, “On the capacity of information networks,” IEEE Trans. Inf. Theory, vol. 52, no. 6, pp. 2345–2364, Jun. 2006.

[17] S. Thakor, A. Grant, and T. Chan, “Network coding capacity: A functional dependence bound,” in Proc. Int. Symp. Inf. Theory, 2009, pp. 263–267.

[18] R. Yeung, A First Course in Information Theory. New York, NY: Plenum, 2002.

[19] S. Thakor, T. Chan, and A. Grant, “Bounds for network information flow with correlated sources,” in Proc. Australian Commun. Theory Workshop, 2011, pp. 43–48.

[20] M. Gastpar, “On source-channel communication in networks,” Dimacs Ser. Discr. Math. Theor. Comput. Sci., vol. 66, pp. 217–238, 2004.

[21] S. Kamath and P. Viswanath, “An information-theoretic meta-theorem on edge-cut bounds,” in Proc. Int. Symp. Inf. Theory, Cambridge, MA, Jul. 2012, pp. 1657–1661.

[22] S. Kamath, D. Tse, and V. Anantharam, “Generalized network sharing outer bound and the two-unicast problem,” in Proc. Int. Symp. Netw. Cod., Beijing, China, Jul. 2011, pp. 1–6.

[23] S. Jaggi and M. Langberg, “Secure network coding: Bounds and algorithms for secret and reliable communication,” in Network Coding: Fundamentals and Applications. New York, NY: Elsevier, 2012, pp. 183–216.

[24] P. Gács and J. Körner, “Common information is far less than mutual information,” Probl. Contr. Inf. Theory, vol. 2, pp. 149–162, 1973.

[25] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” IEEE Trans. Inf. Theory, vol. IT-24, no. 3, pp. 339–348, May 1978.

[26] R. W. Yeung, Information Theory and Network Coding. New York, NY: Springer, 2008.

[27] W. Huang, T. Ho, M. Langberg, and J. Kliewer, CoRR “On secure network coding with uniform wiretap sets,” 2012, vol. abs/1208.4390.

[28] R. M. Gray and A. D. Wyner, “Source coding for a simple network,” Bell Syst. Tech. J., vol. 53, no. 9, pp. 1681–1721, Nov. 1974.

[29] A. D. Wyner, “On source coding with side information at the decoder,” IEEE Trans. Inf. Theory, vol. 21, no. 3, pp. 294–300, May 1975.

[30] T. Cui, T. Ho, and J. Kliewer, “On secure network coding with unequal link capacities and restricted wiretapping sets,” presented at the IEEE Inf. Theory Workshop, 2010.

[31] Z. Zhang, “On a new non-Shannon-type information inequality,” Commun. Inf. Syst., vol. 3, no. 1, pp. 47–60, 2003.

Amin Aminzadegh Gohari (S’10–M’11) is an Assistant Professor at Sharif University of Technology, Tehran, Iran. Dr. Gohari received his M.Sc. and Ph.D. degree in electrical engineering in 2010 from the University of California, Berkeley, and his B.Sc. degree in 2004 from Sharif University of Technology, Iran. He received the 2010 Eli Jury Award from UC Berkeley, Department of Electrical Engineering, for “outstanding achievement in the area of communication networks,” and the 2009–2010 Bernard Friedman Memorial Prize in Applied Mathematics from UC Berkeley, Department of Mathematics, for “demonstrated ability to do research in applied mathematics.” He also received the Gold Medal from the 41st International Mathematical Olympiad (IMO 2000) and the First Prize from the 9th International Mathematical Competition for University Students (IMC 2002).

Shenghao Yang is an Assistant Professor with the Institute for Interdisciplinary Information Sciences, Tsinghua University. He received the B.S. degree from Nankai University in 2001, M.S degree from Peking University in 2004, and Ph.D. degree in Information Engineering from The Chinese University of Hong Kong in 2008. He was a Postdoctoral Fellow in the University of Waterloo from 2008 to 2009, and a Postdoctoral Fellow/Research Associate in the Institute of Network Coding, The Chinese University of Hong Kong from 2010 to 2012.

Sidharth Jaggi (SM’00–M’07) is an Assistant Professor in Information Engineering at the Chinese University of Hong Kong. He received a Ph.D. (2006) and M.S. (2001) from the California Institute of Technology, and a B.Tech. degree (2000) in Electrical Engineering from the Indian Institute of Technology, Bombay. His primary research interests are in network coding, coding theory and information theory.