ISIT'2013 1569720191

General Index Coding with Side Information: Three Decoder Case

Sinem Unal
School of Electrical & Computer Eng.
Cornell University
Ithaca, NY USA
Email: su62@cornell.edu

Aaron B. Wagner
School of Electrical & Computer Eng.
Cornell University
Ithaca, NY USA
Email: wagner@ece.cornell.edu

***Abstract*—The problem of general index coding with side information is investigated for the case of three receivers. We view the problem as a special case of lossy source coding with side information at the decoders, and we find the optimal rate.**

I. Introduction

Index coding with side information, first proposed by Bar-Yossef *et al.* has recently received considerable attention due to its connection with network coding [2], [3] and its potential use in wireless systems. The problem is defined in the following way [1]. There is a sender (or encoder), which has a vector of bits, $x_1, \ldots, x_m$, and m receivers (decoders) $\mathrm{Rec}_1, \ldots, \mathrm{Rec}_m$. Rec_i wishes to learn the bit x_i and has side information consisting of an arbitrary subset of the other bits. The sender must broadcast a common message, of minimum rate, to the receivers that simultaneously meets all of their demands. The minimum rate has been found for several special cases [3], [4], and the minimum rate achieved by linear codes has been characterized [1]. It is also known that nonlinear codes can beat linear ones [5], and that coding over several realizations of the source (i.e., block coding) can also yield improved rates [6]. As such, we shall consider block codes that allow for nonlinear operations. We shall also consider the more general problem in which the number of bits and the number of receivers may be different; that is, some bits may be demanded by multiple users and some may be demanded by none. It is then convenient to lump together all receivers with the same side information into one receiver that may demand multiple source bits.

For this problem of *general index coding with side information*, we exactly characterize the minimum rate achieved by nonlinear, block codes for the case of three decoders. This result contrasts with the existing literature, for which conclusive results are available for a potentially large number of users but only if the side information and demands have a particular structure. Our approach is motivated by the following observation. The techniques developed for index coding can be deployed only in situations where several receivers sharing a common transmitter (such as a cellular base station or WiFi access point) have overlapping demands for content and are willing to coordinate their actions to meet these demands. This is more likely to occur with few receivers than with many.

Whereas most prior work on index coding used graph-theoretic techniques or tools from network coding, we use techniques from network information theory. Specifically, we approach the problem as an instance of lossy source coding with side information at the receivers and show that standard upper and lower bounds on the optimal rate coincide if the auxiliary random variables in the former are chosen in the right way.

II. Main Result

Let $\mathbf{S} = (S_1, \ldots, S_k)$ be a source vector of i.i.d. uniform binary random variables. Let $\mathbf{X} = (X_1, \ldots, X_{k_1})$, $\mathbf{Y} = (Y_1, \ldots, Y_{k_2})$, $\mathbf{Z} = (Z_1, \ldots, Z_{k_3})$ denote the side information at decoder 1, decoder 2, and decoder 3 respectively, where for each i, $X_i = S_j$ for some $j \in \{1, \ldots, k\}$ and likewise for Y_i and Z_i. Let $f_1(\mathbf{S}) = (F_{11}, \ldots, F_{1l_1})$, $f_2(\mathbf{S}) = (F_{21}, \ldots, F_{2l_2})$, $f_3(\mathbf{S}) = (F_{31}, \ldots, F_{3l_3})$ denote the demands at decoder 1, decoder 2 and decoder 3 respectively, where for each i, $F_{1i} = S_j$ for some $j \in \{1, \ldots, k\}$ and likewise F_{2i} and F_{3i}. It is common to view the side information and demands at the decoders as a subset of source variables (see e.g. [7]); our achievable scheme uses a bit alignment procedure that makes it convenient to view them as vectors, however.

Definition 1: An (n, M) code is a collection of mappings

$$\begin{aligned} &f : \{0,1\}^{kn} \to \{1, \ldots, M\} \\ &g_1 : \{1, \ldots, M\} \times \{0,1\}^{k_1 n} \to \{0,1\}^{l_1 n} \\ &g_2 : \{1, \ldots, M\} \times \{0,1\}^{k_2 n} \to \{0,1\}^{l_2 n} \\ &g_3 : \{1, \ldots, M\} \times \{0,1\}^{k_3 n} \to \{0,1\}^{l_3 n}. \end{aligned}$$

We call f the encoding function at the encoder and g_i the decoding function at decoder i, where $i \in \{1, 2, 3\}$.

Definition 2: The rate R is *achievable* if there exists a sequence of (n, M) codes with rate $n^{-1} \log M \le R$ such that the probability of error, P_e, tends to zero as n tends to infinity, where

$$\begin{aligned} P_e = P\{ &g_1(f(\mathbf{S^n}), \mathbf{X^n}) \neq f_1^n(\mathbf{S^n}) \cup g_2(f(\mathbf{S^n}), \mathbf{Y^n}) \neq f_2^n(\mathbf{S^n}) \\ &\cup g_3(f(\mathbf{S^n}), \mathbf{Z^n}) \neq f_3^n(\mathbf{S^n}) \}. \end{aligned}$$

The optimal rate is defined as

$$R_{opt} = \inf\{R : R \text{ is achievable}\}.$$

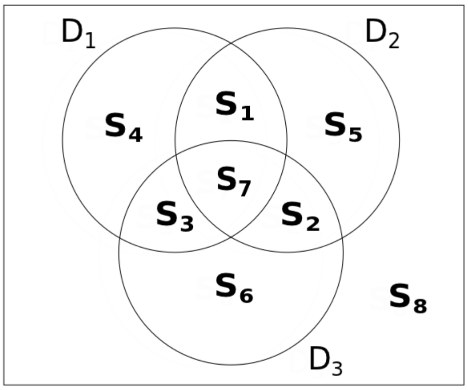

Fig. 1. Relation of Side Information at Decoders

Our goal is to characterize this quantity.

Without loss of generality, one can assume that there are no repetitions in the components of the side information and the demands at the decoders and that the components of $\mathbf{S}$ have been ordered so that the side information at the decoders is as shown in Fig. 1, where the circle D_i denotes the side information at decoder i, $\mathbf{S_1}$ denotes the variables in $\mathbf{S}$ that appear in $\mathbf{X}$ and $\mathbf{Y}$ but not $\mathbf{Z}$, $\mathbf{S_4}$ denotes the set of variables in $\mathbf{S}$ that appear in only $\mathbf{X}$, and so on. Grouping the source variables by this relation, we can write $\mathbf{S} = (\mathbf{S_1}, \mathbf{S_2}, \mathbf{S_3}, \mathbf{S_4}, \mathbf{S_5}, \mathbf{S_6}, \mathbf{S_7}, \mathbf{S_8})$ where $\mathbf{S_8}$ is the vector of source variables that are not contained in any of the three side information vectors. Note that each $\mathbf{S}_i$ may be empty, a single bit, or an vector of i.i.d. bits.

For each decoder i, the components of the demand f_i can be grouped according to which S_j they appear in. The general problem can thus be reduced to the one appearing in Fig. 2, where we use f_{ij} to denote $f_{ij}(\mathbf{S})$, the components of $f_i(\mathbf{S})$ that are in $\mathbf{S_j}$ at decoder i. Hence, we view f_{ij} as a random vector. Note that we have ignored components of f_1 that appear in $\mathbf{X}$, since these demands can always be met, regardless of what the encoder sends, and likewise for f_2 and f_3. Also note that although each decoder's demand formally has four components, some of these may be empty. In that case, one can interpret the corresponding f_{ij} as a constant function of the source. Finally, note that the components of the demand for a decoder may form a strict subset of the set of source variables that it does not have as side information.

We can now state our main result.

Theorem: The optimal rate is

$$\begin{aligned} R_{opt} = \max\{&H(f_{18}, f_{28}, f_{38}, f_{12}, f_{15}, f_{16}, f_{26}|\mathbf{X}), \\ &H(f_{18}, f_{28}, f_{38}, f_{12}, f_{15}, f_{16}, f_{35}|\mathbf{X}), \\ &H(f_{18}, f_{28}, f_{38}, f_{23}, f_{24}, f_{26}, f_{16}|\mathbf{Y}), \\ &H(f_{18}, f_{28}, f_{38}, f_{23}, f_{24}, f_{26}, f_{34}|\mathbf{Y}), \\ &H(f_{18}, f_{28}, f_{38}, f_{31}, f_{34}, f_{35}, f_{15}|\mathbf{Z}), \\ &H(f_{18}, f_{28}, f_{38}, f_{31}, f_{34}, f_{35}, f_{24}|\mathbf{Z})\}. \end{aligned}$$

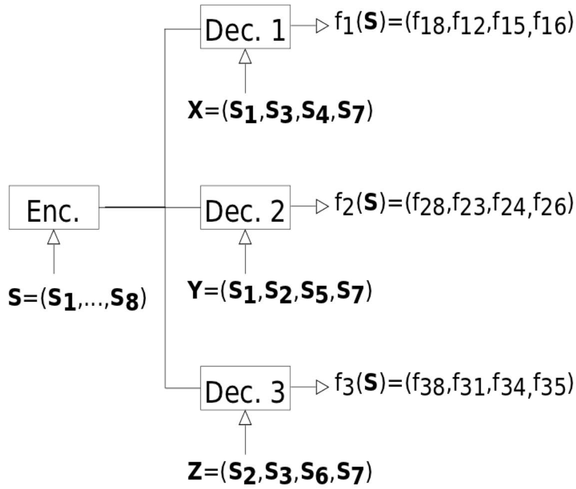

Fig. 2. Problem Setup

III. Achievability

For our upper bound on R_{opt}, we use a random coding technique from network information theory rather than graph theoretic tools from network coding. Specifically, we use a weakening of a result of Timo *et al.* [8]. This result applies to a rate-distortion model, but it can be easily modified to handle the block error probability formulation used here. The Timo *et al.* result corrects an earlier result in the literature [9], although this correction is immaterial when one considers the weakened form used here.

Proposition:

$$R_{opt} \leq \min_{W} [\max\{I(\mathbf{S}; W|\mathbf{X}), I(\mathbf{S}; W|\mathbf{Y}), I(\mathbf{S}; W|\mathbf{Z})\}]$$

where the minimization is over the set of all random variables $(W) \in W$ jointly distributed with $(\mathbf{S})$ such that there exist functions $g_1(W, \mathbf{X})$ and $g_2(W, \mathbf{Y})$ and $g_3(W, \mathbf{Z})$ such that

$$g_1(W, \mathbf{X}) = f_1(\mathbf{S}), \quad g_2(W, \mathbf{Y}) = f_2(\mathbf{S}), \quad g_3(W, \mathbf{Z}) = f_3(\mathbf{S}).$$

It is worth noting that our scheme sends only a common message (W); Timo *et al.* also allow for the transmission of "private" messages to specific subsets of receivers. However, we do not make use of this capability since sending only a common message is sufficient to reach the optimal rate for our problem.

We turn to the choice of W. A straightforward choice is $W = (f_1(\mathbf{S}), f_2(\mathbf{S}), f_3(\mathbf{S}))$, i.e., all of the demands. This gives the rate

$$\begin{aligned} &\max\{H(f_{18}, f_{28}, f_{38}, f_{12}, f_{15}, f_{16}, f_{26}, f_{35}|\mathbf{X}), \\ &H(f_{18}, f_{28}, f_{38}, f_{23}, f_{24}, f_{26}, f_{16}, f_{34}|\mathbf{Y}), \\ &H(f_{18}, f_{28}, f_{38}, f_{31}, f_{34}, f_{35}, f_{15}, f_{24}|\mathbf{Z})\}. \end{aligned} \tag{1}$$

This choice fails to capitalize on all of the coding opportunities available, however. Using $\oplus$ to denote bit-wise exclusive-or, it is sufficient to send $f_{26} \oplus f_{35}$ (with the shorter of the two strings zero padded as necessary) in place of the pair (f_{26}, f_{35}). This is because only the second decoder demands f_{26} and only the third decoder demands f_{35}, and the second decoder has $\mathbf{S}_5$ as side information while the third decoder has $\mathbf{S}_6$ as side information. Likewise, we can send $f_{16} \oplus f_{34}$ and $f_{15} \oplus f_{24}$ in place of $(f_{16}, f_{34}, f_{15}, f_{24})$. This yields the choice $W = (f_{18}, f_{28}, f_{38}, f_{12}, f_{31}, f_{23}, f_{15} \oplus f_{24}, f_{16} \oplus f_{34}, f_{26} \oplus f_{35})$ and the rate

$$\begin{aligned} &\max\{H(f_{18}, f_{28}, f_{38}, f_{12}, f_{15}, f_{16}, f_{26} \oplus f_{35}|\mathbf{X}), \\ &H(f_{18}, f_{28}, f_{38}, f_{23}, f_{24}, f_{26}, f_{16} \oplus f_{34}|\mathbf{Y}), \\ &H(f_{18}, f_{28}, f_{38}, f_{31}, f_{34}, f_{35}, f_{15} \oplus f_{24}|\mathbf{Z})\}. \end{aligned} \tag{2}$$

This is evidently better than the rate in (1). This this rate can be improved further via a careful alignment of the vectors that are being summed. To see this, consider the entropy $H(f_{15}, f_{16}, f_{35} \oplus f_{26})$ as an example, where $f_{15} = (s_1)$, $f_{35} = (s_1, s_2, s_3, s_4)$, $f_{16} = (s_5, s_6)$, $f_{26} = (s_5, s_6, s_7, s_8)$, and each s_i is independent uniform binary random variable. If we reorder f_{35} and f_{26} so that $f_{35} \oplus f_{26} = (s_1 \oplus s_8, s_2 \oplus s_7, s_3 \oplus s_6, s_4 \oplus s_5)$ then $H(f_{15}, f_{16}, f_{35} \oplus f_{26}) = 7$ bits.

However, if we reorder f_{35} and f_{26} so that $f_{35} \oplus f_{26} = (s_2 \oplus s_7, s_3 \oplus s_8, s_4 \oplus s_6, s_1 \oplus s_5)$, then $H(f_{15}, f_{16}, f_{35} \oplus f_{26}) = 6$ bits. It turns out that this same rate can be achieved by summing only a subset of the components: if we interpret "$f_{35} \oplus f_{26}$" as $(s_2 \oplus s_7, s_3 \oplus s_8, s_4, s_6, s_1, s_5)$, then $H(f_{15}, f_{16}, f_{35} \oplus f_{26}) = 6$ bits. Either way, we must pay attention to which bits in f_{35} and f_{26} are being summed. The choice given by the next definition turns out to be optimal.

Definition 3: Let f_{ij}, f_{mn} be two functions from Fig. 2, with $i, m \in \{1, 2, 3\}$, $j, n \in \{4, 5, 6\}$ and $j \neq n$. We define the new function $f_{ij} \oplus^* f_{mn}$ as follows. Observe that there is a unique $k \in \{1, 2, 3\}$ so that f_{kj} appears in the figure. Let $\tilde{f}_{ij}$ denote f_{kj} and define $\tilde{f}_{mn}$ similarly. Let $(U_1, U_2, \ldots, U_{a_1})$ be a vector containing those variables that are in f_{ij} but not $\tilde{f}_{ij}$. Let $(V_1, V_2, \ldots, V_{a_2})$ be a vector containing those variables that are in f_{mn} but not $\tilde{f}_{mn}$. Let $(P_1, P_2, \ldots, P_{b_1})$ be a vector containing those variables that are in both vectors f_{ij} and $\tilde{f}_{ij}$ and let $(R_1, R_2, \ldots, R_{b_2})$ be a vector containing those variables that are in both f_{mn} and $\tilde{f}_{mn}$. We assume that $a_1 \leq a_2$ without loss of generality. Then we define

$$f_{ij} \oplus^* f_{mn} = (A_1, A_2),$$

where

$$\begin{aligned} A_1 &= (U_1 \oplus V_1, \ldots, U_{a_1} \oplus V_{a_1}) \\ A_2 &= (V_{a_1+1}, \ldots, V_{a_2}, P_1, \ldots, P_{b1}, R_1, \ldots, R_{b_2}). \end{aligned}$$

The $\oplus^*$ operation the definition is illustrated in Fig. 3, where each square corresponds to a bit.

The following lemma is proven in the Appendix.

Lemma: Let $f_{ij}, \tilde{f}_{ij}, f_{mn}, \tilde{f}_{mn}$ be as defined in Definition 3.

$$\begin{aligned} \text{Then, } &H(f_{ij} \oplus^* f_{mn}|\tilde{f}_{ij}, \tilde{f}_{mn}) \\ &= \max\{H(f_{ij}|\tilde{f}_{ij}, \tilde{f}_{mn}), H(f_{mn}|\tilde{f}_{ij}, \tilde{f}_{mn})\}. \end{aligned}$$

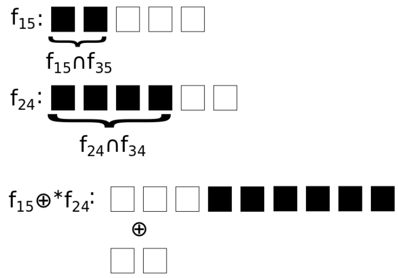

Fig. 3. Example for $\oplus^*$ Operation

Our final choice shall be $W = (f_{18}, f_{28}, f_{38}, f_{12}, f_{31}, f_{23}, f_{15} \oplus^* f_{24}, f_{16} \oplus^* f_{34}, f_{26} \oplus^* f_{35})$. By using the lemma, one can verify that this yields the following rate

$$R_{ach} = R_{opt}.$$

IV. Converse

We next show that for any achievable rate, R, $R_{ach} \leq R$. Suppose R is achievable, and let J be the output of the encoder of an (n, M) code such that $(1/n) \log M \leq R$ and the error probability is upper bounded by δ. Then we have

$$\begin{aligned} nR &\geq H(J) \\ &\geq I(f_1^n, f_2^n, f_3^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}; J) \\ &\overset{(a)}{=} I(\mathbf{X^n}; J) + I(f_1^n; J|\mathbf{X^n}) \\ &\quad + I(\mathbf{Y^n}; J|f_1^n, \mathbf{X^n}) + I(f_2^n; J|f_1^n, \mathbf{X^n}, \mathbf{Y^n}) \\ &\quad + I(\mathbf{Z^n}; J|f_1^n, f_2^n, \mathbf{X^n}, \mathbf{Y^n}) \\ &\quad + I(f_3^n; J|f_1^n, f_2^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}) \\ &\overset{(b)}{\geq} H(f_1^n|\mathbf{X^n}) + H(f_2^n|f_1^n, \mathbf{X^n}, \mathbf{Y^n}) \\ &\quad + H(f_3^n|f_1^n, f_2^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}) \\ &\quad - H(f_1^n|J, \mathbf{X^n}) - H(f_2^n|J, f_1^n, \mathbf{X^n}, \mathbf{Y^n}) \\ &\quad - H(f_3^n|J, f_1^n, f_2^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}) \\ &\overset{(c)}{\geq} n\big[H(f_1|\mathbf{X}) + H(f_2|f_1, \mathbf{X}, \mathbf{Y}) \\ &\quad + H(f_3|f_1, f_2, \mathbf{X}, \mathbf{Y}, \mathbf{Z})\big] - 3n\epsilon \end{aligned}$$

where (a) follows by the chain rule and (b) follows by dropping the terms

$$I(\mathbf{X^n}; J), I(\mathbf{Y^n}; J|f_1^n, \mathbf{X^n}), I(\mathbf{Z^n}; J|f_1^n, f_2^n, \mathbf{X^n}, \mathbf{Y^n})$$

and expanding the remaining mutual information, and (c) follows by Fano's inequality (assuming that δ is sufficiently small compared with ϵ) and the fact that the source is i.i.d.

Since ϵ was arbitrary, this implies

$$\begin{aligned} R &\geq \big[H(f_1|\mathbf{X}) + H(f_2|f_1, \mathbf{X}, \mathbf{Y}) + H(f_3|f_1, f_2, \mathbf{X}, \mathbf{Y}, \mathbf{Z})\big] \\ &= \big[H(f_{18}, f_{12}, f_{15}, f_{16}|\mathbf{S_1}, \mathbf{S_3}, \mathbf{S_4}, \mathbf{S_7}) \\ &\quad + H(f_{28}, f_{23}, f_{24}, f_{26}|f_{18}, f_{16}, \mathbf{S_7}, \mathbf{S_1}, \mathbf{S_2}, \mathbf{S_3}, \mathbf{S_4}, \mathbf{S_5}) \\ &\quad + H(f_{38}, f_{31}, f_{34}, f_{35}|f_{18}, f_{28}, \mathbf{S_1}, \ldots, \mathbf{S_7})\big] \end{aligned}$$

If we expand and rearrange the terms at the right hand side we obtain

$$\begin{aligned}
R &\geq H(f_{18}) + H(f_{12}) + H(f_{15}) + H(f_{16}) \\
&\quad + H(f_{28}|f_{18}) + H(f_{26}|f_{16}) + H(f_{38}|f_{18}, f_{28}) \\
&= H(f_{18}, f_{28}, f_{38}) + H(f_{12}) + H(f_{15}) + H(f_{16}, f_{26}) \\
&= H(f_{18}, f_{28}, f_{38}, f_{12}, f_{16}, f_{26}, f_{15}) \\
&= H(f_{18}, f_{28}, f_{38}, f_{12}, f_{16}, f_{26}, f_{15}|\mathbf{S_1}, \mathbf{S_3}, \mathbf{S_4}, \mathbf{S_7}) \\
&= H(f_{18}, f_{28}, f_{38}, f_{12}, f_{16}, f_{26}, f_{15}|\mathbf{X}). \qquad (3)
\end{aligned}$$

Similarly we find that

$$\begin{aligned}
nR &\geq H(J) \\
&\geq I(f_1^n, f_2^n, f_3^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}; J) \\
&\overset{(a)}{=} I(\mathbf{X^n}; J) + I(f_1^n; J|\mathbf{X^n}) \\
&\quad + I(\mathbf{Z^n}; J|f_1^n, \mathbf{X^n}) + I(f_3^n; J|f_1^n, \mathbf{X^n}, \mathbf{Z^n}) \\
&\quad + I(\mathbf{Y^n}; J|f_1^n, f_3^n, \mathbf{X^n}, \mathbf{Z^n}) \\
&\quad + I(f_2^n; J|f_1^n, f_3^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}) \\
&\overset{(b)}{\geq} H(f_1^n|\mathbf{X^n}) + H(f_3^n|f_1^n, \mathbf{X^n}, \mathbf{Z^n}) \\
&\quad + H(f_2^n|f_1^n, f_3^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}) \\
&\quad - H(f_1^n|J, \mathbf{X^n}) - H(f_3^n|J, f_1^n, \mathbf{X^n}, \mathbf{Z^n}) \\
&\quad - H(f_2^n|J, f_1^n, f_3^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}) \\
&\overset{(c)}{\geq} n\big[H(f_1|\mathbf{X}) + H(f_3|f_1, \mathbf{X}, \mathbf{Z}) \\
&\quad + H(f_2|f_1, f_3, \mathbf{X}, \mathbf{Y}, \mathbf{Z})\big] - 3n\epsilon
\end{aligned}$$

where (a) follows by the chain rule and (b) follows by dropping the terms

$$I(\mathbf{X^n}; J), I(\mathbf{Z^n}; J|f_1^n, \mathbf{X^n}), I(\mathbf{Y^n}; J|f_1^n, f_3^n, \mathbf{X^n}, \mathbf{Z^n})$$

and expanding the remaining mutual information terms, and (c) follows by Fano's inequality (assuming that δ is sufficiently small compared with ϵ) and the fact that the source is i.i.d. Then when $\epsilon \to 0$, we get

$$\begin{aligned}
R &\geq \big[H(f_1|\mathbf{X}) + H(f_3|f_1, \mathbf{X}, \mathbf{Z}) + H(f_2|f_1, f_3, \mathbf{X}, \mathbf{Y}, \mathbf{Z})\big] \\
&= \big[H(f_{18}, f_{12}, f_{15}, f_{16}|\mathbf{S_1}, \mathbf{S_3}, \mathbf{S_4}, \mathbf{S_7}) \\
&\quad + H(f_{38}, f_{31}, f_{34}, f_{35}|f_{18}, f_{15}, \mathbf{S_7}, \mathbf{S_1}, \mathbf{S_2}, \mathbf{S_3}, \mathbf{S_4}, \mathbf{S_6}) \\
&\quad + H(f_{28}, f_{23}, f_{24}, f_{26}|f_{18}, f_{38}, \mathbf{S_1}, \ldots, \mathbf{S_7})\big].
\end{aligned}$$

If we expand and rearrange the terms at the right hand side

$$\begin{aligned}
R &\geq H(f_{18}) + H(f_{12}) + H(f_{15}) + H(f_{16}) \\
&\quad + H(f_{38}|f_{18}) + H(f_{35}|f_{15}) + H(f_{28}|f_{18}, f_{38}) \\
&= H(f_{18}, f_{28}, f_{38}) + H(f_{12}) + H(f_{15}, f_{35}) + H(f_{16}) \\
&= H(f_{18}, f_{28}, f_{38}, f_{12}, f_{16}, f_{15}, f_{35}) \\
&= H(f_{18}, f_{28}, f_{38}, f_{12}, f_{16}, f_{15}, f_{35}|\mathbf{S_1}, \mathbf{S_3}, \mathbf{S_4}, \mathbf{S_7}) \\
&= H(f_{18}, f_{28}, f_{38}, f_{12}, f_{16}, f_{15}, f_{35}|\mathbf{X}). \qquad (4)
\end{aligned}$$

Combining (3) and (4) gives

$$\begin{aligned}
R \geq \max\{&H(f_{18}, f_{28}, f_{38}, f_{12}, f_{16}, f_{15}, f_{26}|\mathbf{X}), \\
&H(f_{18}, f_{28}, f_{38}, f_{12}, f_{16}, f_{15}, f_{35}|\mathbf{X})\}.
\end{aligned}$$

We can summarize our procedure as follows. Note that we have

$$\begin{aligned}
nR &\geq H(J) \\
&\geq I(f_1^n, f_2^n, f_3^n, \mathbf{X^n}, \mathbf{Y^n}, \mathbf{Z^n}; J).
\end{aligned}$$

Observe that while applying the chain rule to the above expression, we first extracted $\mathbf{X}$, followed by its corresponding demand, followed by $\mathbf{Y}$ and its corresponding demand, and finally $\mathbf{Y}$ and its corresponding demand. This led to the bound in (3). We then reversed the order of $\mathbf{Y}$ and $\mathbf{Z}$ and the corresponding demands to obtain (4). There are four other possible orderings, each of which yields a lower bound that equals one of the entropy expressions in the Theorem. Combining all of these bounds yields

$$\begin{aligned}
R \geq= \max\{&H(f_{18}, f_{28}, f_{38}, f_{12}, f_{15}, f_{16}, f_{26}|\mathbf{X}), \\
&H(f_{18}, f_{28}, f_{38}, f_{12}, f_{15}, f_{16}, f_{35}|\mathbf{X}), \\
&H(f_{18}, f_{28}, f_{38}, f_{23}, f_{24}, f_{26}, f_{16}|\mathbf{Y}), \\
&H(f_{18}, f_{28}, f_{38}, f_{23}, f_{24}, f_{26}, f_{34}|\mathbf{Y}), \\
&H(f_{18}, f_{28}, f_{38}, f_{31}, f_{34}, f_{35}, f_{15}|\mathbf{Z}), \\
&H(f_{18}, f_{28}, f_{38}, f_{31}, f_{34}, f_{35}, f_{24}|\mathbf{Z})\}.
\end{aligned}$$

This completes the proof of the converse.

V. Discussion

A. Related Work

Consider the problem setup in which we provide the side information at the first decoder to the second and third decoders and the side information at the second decoder to the third decoder. Since this new setup consists of decoders with degraded side information, we know the optimal rate for it by [9]. Also note that the optimal rate for this new setup is lower than equal to the that of original problem. Therefore, the solution to this new problem provides a lower bound to the original. Since we have three decoders, this yields six different lower bounds, the maximum of which is of course also a lower bound. This is essentially what we do in the converse proof.

Maleki *et al.* [7] apply interference alignment techniques to the index coding problem. The bit alignment procedure given in Definition 3 is different from what is proposed there, however, as Maleki *et al.* only consider the problem in which each source component is demanded by exactly one decoder. The procedure in Definition 3, on the other hand, is nondegenerate only when multiple decoders demand the same source components. Also, at this conference Arbabjolfaei *et al.* [10] consider index coding as a network information theory problem and find the optimal rate for small networks, as we do here. As with Maleki *et al.* they do not allow for a source component to be demanded by multiple decoders, however, and their optimality proof is numerical, unlike the analytical proof given here.

B. An Example

Lastly, we give a numerical example for our optimal scheme. Let $\mathbf{S} = (S_1, \ldots, S_{17})$, $\mathbf{S_1} = (S_1, S_2)$,

$\mathbf{S_2} = (S_3, S_4, S_5)$, $\mathbf{S_3} = (S_6)$, $\mathbf{S_4} = (S_7, S_8)$, $\mathbf{S_5} = (S_9, S_{10}, S_{11}, S_{12})$, $\mathbf{S_6} = (S_{13}, S_{14}, S_{15}, S_{16})$, $\mathbf{S_7} = \emptyset$, $\mathbf{S_8} = (S_{17})$ and let the demands be

$$\begin{aligned}
&f_{31} = (S_2), f_{12} = (S_3, S_4), f_{23} = \emptyset,\\
&f_{24} = (S_7, S_8), f_{34} = (S_7),\\
&f_{15} = (S_9), f_{35} = (S_9, S_{10}, S_{11}),\\
&f_{16} = (S_{13}, S_{14}), f_{26} = (S_{13}, S_{14}, S_{15}, S_{16}),\\
&f_{18} = f_{28} = (S_{17}), f_{38} = \emptyset,
\end{aligned}$$

where each S_i is independent uniform binary random variable. Then optimal rate can be calculated as

$$\begin{aligned}
R_{opt} = \max\{&H(f_{18}, f_{28}, f_{38}, f_{12}, f_{15}, f_{16}, f_{26}|\mathbf{X}) = 8 \text{ bits},\\
&H(f_{18}, f_{28}, f_{38}, f_{12}, f_{15}, f_{16}, f_{35}|\mathbf{X}) = 8 \text{ bits},\\
&H(f_{18}, f_{28}, f_{38}, f_{23}, f_{24}, f_{26}, f_{16}|\mathbf{Y}) = 7 \text{ bits},\\
&H(f_{18}, f_{28}, f_{38}, f_{23}, f_{24}, f_{26}, f_{34}|\mathbf{Y}) = 7 \text{ bits},\\
&H(f_{18}, f_{28}, f_{38}, f_{31}, f_{34}, f_{35}, f_{15}|\mathbf{Z}) = 6 \text{ bits},\\
&H(f_{18}, f_{28}, f_{38}, f_{31}, f_{34}, f_{35}, f_{24}|\mathbf{Z}) = 7 \text{ bits}\}\\
= 8 \text{ bits}.
\end{aligned}$$

If we calculate the rate in (1) obtained by selecting $W = (f_1(\mathbf{S}), f_2(\mathbf{S}), f_3(\mathbf{S}))$, we get 10 bits. Also, the rate in (2) obtained by choosing $W = (f_{18}, f_{28}, f_{38}, f_{12}, f_{31}, f_{23}, f_{15} \oplus f_{24}, f_{16} \oplus f_{34}, f_{26} \oplus f_{35})$ is equal to 9 bits, where

$$\begin{aligned}
&f_{15} \oplus f_{24} = (S_9 \oplus S_7, S_8),\\
&f_{16} \oplus f_{34} = (S_{13} \oplus S_7, S_{14}),\\
&f_{26} \oplus f_{35} = (S_{13}, S_{14} \oplus S_{11}, S_{15} \oplus S_{10}, S_{16} \oplus S_9).
\end{aligned}$$

VI. ACKNOWLEDGEMENT

This work was supported by Intel, Cisco, and Verizon under the Video-Aware Wireless Networks (VAWN) program.

APPENDIX

Proof of the Lemma:

Let $(U_1, U_2, \ldots, U_{a_1}), (V_1, V_2, \ldots, V_{a_2}), (P_1, P_2, \ldots, P_{b_1}), (R_1, R_2, \ldots, R_{b_2}), f_{ij}, \tilde{f}_{ij}, f_{mn}, \tilde{f}_{mn}$ be as in Definition 3.

Also, let $a_1 \leq a_2$ without loss of generality and $f_{ij} \oplus^* f_{mn} = (A_1, A_2)$, where A_1,A_2 are defined as in Definition 3.

$$\begin{aligned}
\text{Then, } &H(f_{ij} \oplus^* f_{mn}|\tilde{f}_{ij}, \tilde{f}_{mn})\\
&= H(A_1, A_2|P_1, \ldots, P_{b_1}, R_1, \ldots, R_{b_2}).
\end{aligned}$$

Note that

$$\begin{aligned}
&H(A_1|\tilde{f}_{ij}, \tilde{f}_{mn})\\
&= H(U_1 \oplus V_1, \ldots, U_{a_1} \oplus V_{a_1}|P_1, \ldots, P_{b_1}, R_1, \ldots, R_{b_2})\\
&= H(U_1 \oplus V_1, \ldots, U_{a_1} \oplus V_{a_1})\\
&= H(A_1),\ a_1 \text{ bits}.
\end{aligned}$$

Then,

$$\begin{aligned}
&H(f_{ij} \oplus^* f_{mn}|\tilde{f}_{ij}, \tilde{f}_{mn})\\
&= H(A_1, A_2|\tilde{f}_{ij}, \tilde{f}_{mn})\\
&= H(A_1) + H(A_2|\tilde{f}_{ij}, \tilde{f}_{mn}, A_1)\\
&= H(A_1) + H(V_{a_1+1}, \ldots, V_{a_2}, P_1, \ldots, P_{b1},\\
&\quad R_1, \ldots, R_{b_2}|P_1, \ldots, P_{b_1}, R_1, \ldots, R_{b_2}, A_1)\\
&= H(A_1) + H(V_{a_1+1}, \ldots, V_{a_2}),\ a_2 \text{ bits}
\end{aligned}$$

Therefore, we can conclude that

$$\begin{aligned}
&H(f_{ij} \oplus^* f_{mn}|\tilde{f}_{ij}, \tilde{f}_{mn}) = a_2 = \max(a_1, a_2)\\
&= \max\{H(f_{ij}|\tilde{f}_{ij}, \tilde{f}_{mn}), H(f_{mn}|\tilde{f}_{ij}, \tilde{f}_{mn})\}.
\end{aligned}$$

■

REFERENCES

[1] Z. Bar-Yossef, Y. Birk, T. S. Jayram, T. Kol "Index Coding with Side Information" *47th Annu. IEEE Symp. Foundations of Computer Science (FOCS)*, pp. 197–206, Oct. 2006.

[2] S. El Rouayheb, A. Sprintson, C. Georghiades, "On the relation between the Index Coding and the Network Coding problems" *IEEE Int. Symp. on Info. Theory (ISIT)*, pp. 1823–1827, July 2008.

[3] A. Blasiak, R. D. Kleinberg, E. Lubetzky, "Index coding via linear programming" *ArXiv preprint,* vol. 1004.1379, 2011.

[4] L. Ong, Chin Keong Ho, "Optimal index codes for a class of multicast networks with receiver side information" *IEEE Int. Conference on Communications (ICC)*, pp. 2213–2218, June 2012.

[5] E. Lubetzky, U. Stav, "Nonlinear Index Coding Outperforming the Linear Optimum" *IEEE Trans. on Info. Theory,* vol. 55, no. 8, Aug. 2009.

[6] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, A. Hassidim, "Broadcasting with Side Information" *49th Annu. IEEE Symp. Foundations of Computer Science (FOCS)*, pp. 823–832, Oct. 2008.

[7] H. Maleki, V. Cadambe, S. Jafar "Index Coding: An Interference Alignment Perspective" *Int. Symp. on Info. Theory (ISIT)*, pp. 2236–2240, July 2012.

[8] R. Timo, T. Chan, A. Grant, "Rate Distortion With Side-Information at Many Decoders" *IEEE Trans. on Info. Theory*, vol. 57, no. 8, pp. 5240–5257, Aug. 2011.

[9] C. Heegard, T. Berger, "Rate distortion when side information may be absent" *IEEE Trans. on Info. Theory*, vol. 31, no. 6, Nov. 1985.

[10] F. Arbabjolfaei, B. Bandemer, Y. H. Kim, E. Sasoglu, L. Wang, "On the Capacity Region for Index Coding" *ArXiv preprint,* vol. 1302.1601, 2013.