Distributed Index Coding

Parastoo Sadeghi[†], Fatemeh Arbabjolfaei[*], and Young-Han Kim[*]
[†]Research School of Engineering, Australian National University, Canberra, ACT, 2601, Australia
[*]Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093, USA
Emails: parastoo.sadeghi@anu.edu.au, {farbabjo, yhk}@ucsd.edu

***Abstract*—In this paper, we study the capacity region of the general distributed index coding. In contrast to the traditional centralized index coding where a single server contains all n messages requested by the receivers, in the distributed index coding there are $2^n - 1$ servers, each containing a unique non-empty subset J of the messages and each is connected to all receivers via a noiseless independent broadcast link with an arbitrary capacity $C_J \geq 0$. First, we generalize the existing outer bound on the capacity region of the centralized problem to the distributed case. Next, building upon the existing centralized composite coding scheme, we propose three distributed composite coding schemes and derive the corresponding inner bounds on the capacity region. We present a number of interesting numerical examples, which highlight the subtleties and challenges of dealing with the distributed index coding, even for very small problem sizes of $n = 3$ and $n = 4$.**

I. Introduction

The index coding problem is a canonical problem in network information theory and has been studied over the past two decades using tools from various disciplines such as combinatorics, algebra, and information theory [1], [2]. In the traditional index coding problem, it is assumed that one server has all messages requested by the receivers. However, in many practical circumstances this assumption might not be true and the messages might be distributed over multiple servers. We refer to this more general version as the *distributed index coding problem* compared to the traditional *centralized* index coding problem. The distributed index coding problem was first studied by Ong, Ho, and Lim [3], where lower and upper bounds on the optimal codelength were derived in the special case in which each message is only known to one receiver and each receiver only knows one message a priori and it is shown that the bounds match if no two servers have any messages in common.

In this paper, we study the distributed index coding problem in its general form, which to the best of our knowledge has not been investigated before. First, we generalize the outer bound on the capacity region of the centralized problem [4] to the distributed case. Next, building upon the existing centralized composite coding scheme [4], we propose three distributed composite coding schemes, derive the corresponding inner bounds on the capacity region and show their use via examples. Although the outer bound is tight for all centralized index coding problems with up to $n = 5$ messages [4], there exist instances of the distributed problem with $n = 3$ messages for which the outer bound is not tight. Nevertheless, using customized Shannon-type inequalities, we show that the proposed distributed composite coding scheme achieves the capacity region for all problems with $n = 3$ messages.

In this paper, $[n]$ denotes the set $\{1, 2, \ldots, n\}$ and the set of all nonempty subsets of $[n]$ is $N \doteq \{J \subseteq [n] : J \neq \emptyset\}$.

II. System Model and Problem Setup

Consider the following index coding problem. There are n messages in the system, $x_1, x_2, \cdots, x_n$, where $x_j \in \{0,1\}^{t_j}$ for $j \in [n]$ and some t_j. There are n receivers, where receiver j wants to obtain message x_j and knows a subset of the messages a priori, denoted by $x(A_j)$ for some $A_j \subseteq [n] \setminus \{j\}$. For simplicity of notation throughout the paper and where the context is clear, we will refer to j as the wanted message and to A_j as the side information of receiver j, respectively. Any instance of this problem can be specified by a side information graph G with n nodes, in which a directed edge $i \to j$ represents that receiver j has message i as side information ($i \in A_j$). For instance, Fig. 1 shows the directed graph representing the index coding problem with $A_1 = \emptyset$, $A_2 = \{3\}$, and $A_3 = \{2\}$.

The main difference in the system model compared to traditional (centralized) index coding problem is in the server setup. Instead of a single server which contains all messages, there are $2^n - 1$ servers. For each $J \in N$, there is a server that contains all messages $j \in J$ and the capacity of the broadcast link connecting server J to all receivers is denoted by C_J. Hence, we assume that there are $2^n - 1$ ideal bit pipes to the receivers with arbitrary link capacities. This is a fairly general model that allows for all possible message availabilities on different servers. If $C_J = 1$ only for $J = [n]$ and is zero otherwise, we recover the centralized index coding problem. A special normalized symmetric case is where $C_J = 1$ for all $J \in N$. Server J sends sequence $y_J \in \{0,1\}^{s_J}$ for some s_J to all receivers. A $((t_j, j \in [n]), (s_J, J \in N))$ *distributed index code* is defined by

- $2^n - 1$ encoders, one for each server $J \in N$, such that $\phi_J : \prod_{i \in J} \{0,1\}^{t_i} \to \{0,1\}^{s_J}$ maps the messages in server J, $(x_i, i \in J)$, to an s_J-bit sequence y_J, and
- n decoders $\psi_j : \prod_{J \in N} \{0,1\}^{s_J} \times \prod_{k \in A_j} \{0,1\}^{t_k} \to \{0,1\}^{t_j}$ that maps the received sequences $\phi_J(x_i, i \in J)$ and the side information $x(A_j)$ back to x_j for $j \in [n]$.

Thus, for every $x^n \in \prod_{i=1}^n \{0,1\}^{t_i}$,

$$\psi_j((\phi_J, J \in N), x(A_j)) = x_j, \quad j \in [n].$$

We say that rate-capacity tuple $(\mathbf{R}, \mathbf{C}) = ((R_j, j \in [n]), (C_J, J \in N))$ is achievable if there exists a $((t_j, j \in [n]), (s_J, J \in N))$ distributed index code and r such that

$$R_j \leq \frac{t_j}{r}, \quad C_J \geq \frac{s_J}{r}, \quad \forall j \in [n], \quad \forall J \in N. \qquad (1)$$

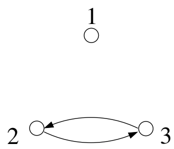

Fig. 1. The side information graph for $A_1 = \emptyset, A_2 = \{3\}, A_3 = \{2\}$.

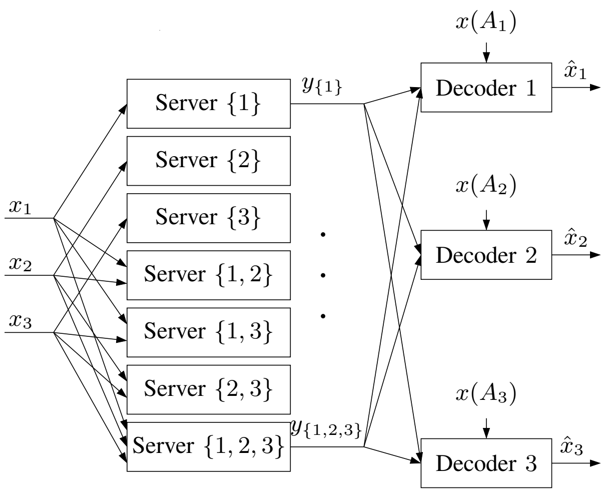

Fig. 2. The distributed index coding problem with $n = 3$.

For a given $\mathbf{C}$, the capacity region $\mathscr{C}$ of this index coding problem is the closure of the set of achievable rate tuples $\mathbf{R} = (R_1, \cdots, R_n)$. Throughout the paper, we will compactly represent the distributed index coding problem (for a given $\mathbf{C}$) as sets of $(j|i \in A_j)$. For example, for $A_1 = \emptyset$, $A_2 = \{3\}$, and $A_3 = \{2\}$, we write $(1); (2|3); (3|2)$.

III. Outer Bounds

In this section, we present an outer bound on the capacity region of the distributed index coding problem by considering all non-empty subsets of messages and adapting the outer bound for the centralized problem (see, for example, [4, Theorem 1]) to take into account the sum-capacity constraints on the corresponding servers.

Theorem 1: Let B_j be the set of interfering messages at receiver j, i.e., $B_j = [n] \setminus (A_j \cup \{j\})$. If $(\mathbf{R}, \mathbf{C})$ is achievable, then for every $T \in N$,

$$R_j \le f_T(B_j \cup \{j\}) - f_T(B_j), \quad j \in T, \tag{2}$$

for some $f_T(S)$, $S \subseteq T$, such that

1) $f_T(\emptyset) = 0$,
2) $f_T(T) = \sum_{J: J \cap T \neq \emptyset} C_J$,
3) $f_T(A) \le f_T(B)$ for all $A \subseteq B \subseteq T$, and
4) $f_T(A \cup B) + f_T(A \cap B) \le f_T(A) + f_T(B), \forall A, B \subseteq T$.

The proof is provided in Appendix A. A relaxed version of this bound, which is a generalized version of the maximal acyclic induced subgraph (MAIS) bound, is handy and also useful for some problems.

Corollary 2: If $(\mathbf{R}, \mathbf{C})$ is achievable for an index coding problem represented by the directed graph G, then for every $T \in N$ it must satisfy

$$\sum_{j \in S} R_j \le \sum_{J: J \cap T \neq \emptyset} C_J, \tag{3}$$

for all $S \subseteq T$ for which the subgraph of G induced by S does not contain a directed cycle.

Example 3: Consider the index coding problem shown in Fig. 1. The following is the generalized MAIS outer bound for this problem (inactive inequalities are not shown).

$$\begin{aligned}
R_1 &\le C_{\{1\}} + C_{\{1,2\}} + C_{\{1,3\}} + C_{\{1,2,3\}}, \\
R_2 &\le C_{\{2\}} + C_{\{1,2\}} + C_{\{2,3\}} + C_{\{1,2,3\}}, \\
R_3 &\le C_{\{3\}} + C_{\{1,3\}} + C_{\{2,3\}} + C_{\{1,2,3\}}, \\
R_1 + R_2 &\le \sum_{J = N \setminus \{3\}} C_J, \\
R_1 + R_3 &\le \sum_{J = N \setminus \{2\}} C_J.
\end{aligned}$$

IV. Inner Bounds

In this section, we present a series of inner bounds on the capacity region of the distributed index coding problem. The first inner bound is a simple extension of [4], in which we solve the composite coding problem separately for each server. We show the shortcoming of this method through an example and then present a new composite coding scheme that solves the problem collectively across all servers. The limitation of this scheme, which is assigning the same decoding sets across all servers, is demonstrated through an example. This leads us to the last scheme as a generalization of the two previous methods, where we allow a general grouping of servers for solving distributed index coding.

A. Distributed Composite Coding For Individual Servers

The idea is to separately apply the composite coding scheme of [4] to each server and then appropriately combine the achievable rates.

For each non-empty subset $K \subseteq J$ at server $J \in N$, there is a virtual encoder at that server with associated composite coding rate $C_{K,J}$. In the first step of composite coding, virtual encoder K at server J maps the messages indexed by K, $(x_j, j \in K)$, into a single composite index $W_{K,J}$, which is generated randomly and independently as a Bern(1/2) sequence of length $2^{s_J C_{K,J}}$ bits. In the second step, server J uses flat coding to encode the composite indices $(W_{K,J}, K \subseteq J)$ into $y_J \in \{0,1\}^{s_J}$. As with the encoding, decoding also takes place in two steps. Each receiver first recovers all composite indices $(W_{K,J}, K \subseteq J)$ for each server $J \in N$. This can be achieved without an error if

$$\sum_{K: K \not\subseteq A_{j,J}} C_{K,J} \le C_J, \quad \forall j \in J, \forall J. \tag{4}$$

is satisfied, where $A_{j,J} = A_j \cap J$ is the common side information of receiver $j \in J$ with server J. As the second step of decoding, each receiver recovers the desired message from the composite indices. Let $D_{j,J}$ be the set of messages that receiver $j \in J$ decodes from server J (note, index coding requires that $j \in D_{j,J}$). Then the probability that message x_j can be recovered correctly at rate $R_{j,J}$ goes to

TABLE I

CAPACITY REGION FOR ALL NON-ISOMORPHIC PROBLEMS OF SIZE $n = 3$.

Index Coding Problem	R_1	R_2	R_3	$R_1 + R_2$	$R_1 + R_3$	$R_2 + R_3$	$R_1 + R_2 + R_3$	An Optimal Decoding Set
$(1); (2); (3)$ $(1\|2); (2); (3)$ $(1\|2,3); (2); (3)$ $(1); (2\|3); (3\|1)$ $(1); (2\|1); (3\|1)$ $(1); (2\|1); (3\|1,2)$	≤ 4	≤ 4	≤ 4	≤ 6	≤ 6	≤ 6	≤ 7	$D_j = [n] \setminus A_j$
$(1\|3); (2\|1); (3\|2)$	≤ 4	≤ 4	≤ 4	≤ 6	≤ 6	≤ 6	≤ 9	$D_j = [n] \setminus A_j$
$(1); (2\|3); (3\|2)$ $(1); (2\|1,3); (3\|2)$ $(1); (2\|1,3); (3\|1,2)$	≤ 4	≤ 4	≤ 4	≤ 6	≤ 6	≤ 8	≤ 9	If $A_j = \emptyset$ then $D_j = \{j\}$, otherwise $D_j = [n] \setminus A_j$
$(1\|3); (2\|3); (3\|2)$ $(1\|3); (2\|1,3); (3\|2)$ $(1\|2,3); (2\|3); (3\|2)$	≤ 4	≤ 4	≤ 4	≤ 6	≤ 6	≤ 8	≤ 10	$D_j = [n] \setminus A_j$
$(1\|3); (2\|3); (3\|1,2)$ $(1\|2,3); (2\|3); (3\|1,2)$	≤ 4	≤ 4	≤ 4	≤ 6	≤ 8	≤ 8	≤ 10	$D_j = [n] \setminus A_j$
$(1\|2,3); (2\|1,3); (3\|1,2)$	≤ 4	≤ 4	≤ 4	≤ 8	≤ 8	≤ 8	≤ 12	$D_j = [n] \setminus A_j$

1 as $s_J \to \infty$, if the rates of the composite messages belong to the polymatroidal rate region $\mathcal{R}(D_{j,J}|A_{j,J})$ defined by

$$\sum_{j \in L_J} R_{j,J} < \sum_{K \subseteq D_{j,J} \cup A_{j,J} :\, K \cap L_J \neq \emptyset} C_{K,J}, \tag{5}$$

for all $L_J \subseteq D_{j,J} \setminus A_{j,J}$. Then the achievable rate region for server J is given by

$$\mathbf{R}_J \in \bigcap_{j \in J} \quad \bigcup_{D_{j,J} \subseteq J : j \in D_{j,J}} \mathcal{R}(D_{j,J}|A_{j,J}), \tag{6}$$

After finding the composite coding achievable rate regions $\mathbf{R}_J$ for all servers, we obtain a combined achievable rate region by applying the following constraints.

$$R_j < \sum_{J : j \in J} R_{j,J}, \quad \forall j \in [n] \tag{7}$$

and eliminating $(R_{j,J}, j \in J, J \in N)$ through Fourier-Motzkin elimination [5, Appendix D]. This yields an inner bound on the rate-capacity tuple $(\mathbf{R}, \mathbf{C})$. We can view this scheme as a rate splitting method across the servers in N.

We applied this technique to the problem $(1|3); (2|1); (3|2)$ when $C_J = 1$ for all $J \in N$. Focusing on the sum rate only due to space limitations, we observed that $R_1 + R_2 + R_3 < 7.5$. However, the outer bound of Theorem 1 gives a sum rate of $R_1 + R_2 + R_3 < 9$. Next, we will see that this outer bound can indeed be achieved by treating servers collectively.

B. Distributed Composite Coding Across All Servers

Here, we solve a single composite coding problem across all servers. Intuitively, the advantage of this scheme is that receivers can collectively use servers to decode messages, even though some of those messages do not exist on some servers. Hence, unlike the previous scheme, this method does not use rate splitting across the servers in N.

Let D_j be the index of messages receiver j wishes to decode. The rates of the composite messages belong to the polymatroidal rate region $\mathcal{R}(D_j|A_j)$ defined by

$$\sum_{j \in L} R_{j,N} < \sum_{K \subseteq D_j \cup A_j : K \cap L \neq \emptyset} \quad \sum_{J : K \subseteq J} C_{K,J}, \tag{8}$$

for all $L \subseteq D_j \setminus A_j$. The subscript N in $R_{j,N}$ signifies that the problem is solved across all servers, N. The second sum on the RHS is because all servers that contain K contribute to the composite rate. Similar as before, we have

$$\mathbf{R}_N \in \bigcap_{j \in [n]} \quad \bigcup_{D_j \subseteq [n] : j \in D_j} \mathcal{R}(D_j|A_j), \tag{9}$$

and the constraints on the composite rates are

$$\sum_{K : K \nsubseteq A_j} C_{K,J} \leq C_J, \quad \forall j \in [n], \forall J \in N. \tag{10}$$

Table I shows the composite coding inner bound along with an optimal decoding set for all non-isomorphic cases with $n = 3$ and when $C_J = 1$ for all $J \in N$. Revisiting the problem $(1|3); (2|1); (3|2)$, the inner bound matches the MAIS outer bound, hence establishing the capacity region. In fact, for all non-isomorphic cases shown in Table I, the composite coding inner bound matches the MAIS outer bound, except for the cases that are in the same group as the problem $(1); (2|3); (3|2)$. For these instances, the inequality

$$R_1 + R_2 + R_3 \leq 9 \tag{11}$$

is not deducible by Theorem 1. However, using customized Shannon-type inequalities, we show that this inequality is also an outer bound inequality (see Appendix B). Therefore, for all non-isomorphic index coding problems with $n = 3$, the rates shown in Table I are indeed capacity regions. It can be easily verified that the capacity regions of all problems with $n = 3$ are achieved by this composite coding scheme for arbitrary values of C_J, $J \in N$. Furthermore, in Table I we can see that the capacity regions of the two problems $(1); (2|3); (3|2)$ and $(1|3); (2|3); (3|2)$ are not the same. This proves the important point that, as opposed to the structural property of the centralized index coding [6], [7], [8], removing an edge that does not belong to any directed cycle may change the capacity region of the distributed index coding problem.

To show the limitation of this scheme, consider the problem $(1|4); (2|3,4); (3|1,2); (4|2,3)$ with $C_J = 1, \forall J \in N$. Focusing on the sum rate due to space limitations, we observe that the sum rate is limited to $R_1 + R_2 + R_3 + R_4 < 23$. However, the outer bound of Theorem 1 gives $R_1 + R_2 + R_3 + R_4 \leq 24$. In the next subsection, we will show that this sum rate can indeed be achieved by appropriately grouping servers and assigning suitable different decoding sets to each group.

TABLE II

A POSSIBLE SERVER GROUPING AND DECODING SETS FOR THE PROBLEM $(1|4); (2|3,4); (3|1,2); (4|2,3)$ FOR THE SCHEME OF SECTION IV-C. NOTE, ONLY ACTIVE INEQUALITIES FOR NON-ZERO RATES ARE WRITTEN. DUE TO SPACE LIMITATIONS, $R_{j,P}$ IN THE COLUMN HEADINGS IS WRITTEN AS R_j

Server Groupings	R_1	R_2	R_3	R_4	R_1+R_2	R_1+R_3	R_1+R_4	R_3+R_4	$R_1+R_2+R_3$	$R_1+R_2+R_4$	**Decoding Sets**
P = $\{\{1\},\{2\},\{3\},\{4\}\}$	≤ 1	≤ 1	≤ 1	≤ 1							$D_{j,P}=[p]\setminus A_{j,P}$
P = $\{\{1,2\}\}$					≤ 1						$D_{j,P}=[p]\setminus A_{j,P}$
$P=\{\{1,3\},\{3,4\}\}$	≤ 1			≤ 1		≤ 2		≤ 2			$D_{j,P}=[p]\setminus A_{j,P}$
$P=\{\{2,3\},\{2,4\}\}$		≤ 2	≤ 1	≤ 1							$D_{j,P}=[p]\setminus A_{j,P}$
$P=$ $\{\{1,4\},\{1,2,3,4\}$ $\{1,2,3\},\{1,2,4\},$ $\{1,3,4\},\{2,3,4\}\}$	≤ 5	≤ 4	≤ 4	≤ 5	≤ 6	≤ 6	≤ 6	≤ 6	≤ 9	≤ 10	$D_{1,P}=$ $\{1\}$, $D_{j,P}=$ $[p]\setminus A_{j,P}$, $j=2,3,4$

C. Partitioned Distributed Composite Coding

This scheme can is a generalization of the previous two schemes. In this new scheme, we jointly optimize grouping of servers and decoding sets across server groupings.

Consider a partition of N, denoted by Π, such that for all $P \in \Pi$ we have $P \subseteq N, P \neq \emptyset, \bigcup_{P\in\Pi} P = N$, and if $P, Q \in \Pi$ and $P \neq Q$, then $P \cap Q = \emptyset$.

For each element of the partition $P \in \Pi$, let $[p] = \{j|\exists J \in P : j \in J\}$ be the union of all messages held by the servers in P. For every $j \in [p]$, let $D_{j,P}$ be the index of messages that receiver j wishes to decode from servers in P and $A_{j,P} = A_j \cap [p]$. The rates of the composite messages belong to the polymatroidal region $\mathcal{R}(D_{j,P}|A_{j,P})$:

$$\sum_{j\in L_P} R_{j,P} < \sum_{K\subseteq D_{j,P}\cup A_{j,P}:\, K\cap L_P\neq\emptyset} \; \sum_{J:J\in P, K\subseteq J} C_{K,J},$$

for all $L_P \subseteq D_{j,P} \setminus A_{j,P}$. Then the achievable rate region for servers in P is given by

$$\mathbf{R}_P \in \bigcap_{j\in[p]} \; \bigcup_{D_{j,P}\subseteq[p]:j\in D_{j,P}} \mathcal{R}(D_{j,P}|A_{j,P}), \tag{12}$$

with the following constraint on composite rates

$$\sum_{K:K\not\subseteq A_{j,P}} C_{K,J} \leq C_J, \quad \forall j \in [p], \forall J \in P. \tag{13}$$

After finding the composite coding rate regions $\mathbf{R}_P$ for all elements of the partition Π, we obtain a combined achievable rate region by applying the following constraints.

$$R_j < \sum_{P:j\in[p]} R_{j,P}, \quad \forall j \in [n] \tag{14}$$

and eliminating $(R_{j,P}, j \in [p], P \in \Pi)$ through Fourier-Motzkin elimination. This yields an inner bound on the rate-capacity tuple $(\mathbf{R}, \mathbf{C})$. We can view this scheme as a rate splitting method across the elements of Π, whenever $|\Pi| \geq 2$.

Note that the previous two composite coding schemes are special cases of this method. In Section IV-A, there are $2^n - 1$ server groups of the form $\Pi = \{P : P = J, \forall J \in N\}$. In Section IV-B, there is a single server group, $\Pi = N$.

The probelm $(1|4); (2|3,4); (3|1,2); (4|2,3)$ is revisited in Table II. Each row represents a server grouping and the corresponding decoding sets are shown in the last column. After applying (14), we find the overall inner bound to be:

$$R_j \leq 8, \qquad j \in [n], \tag{15}$$

and

$$R_1 + R_2 \leq 12, \qquad R_1 + R_3 \leq 12, \tag{16}$$

$$R_1 + R_4 \leq 12, \qquad R_3 + R_4 \leq 12, \tag{17}$$

$$R_1 + R_2 + R_3 \leq 18, \qquad R_1 + R_3 + R_4 \leq 19. \tag{18}$$

There indeed exists an explicit coding scheme with $R_1 = R_2 = R_3 = R_4 = 6$ which matches the maximum sum rate given by Theorem 1. However, there is still a gap between the inner and outer bounds, which is being investigated.

V. CONCLUSION

We studied a more general class of index coding problems than those considered in the literature, where for a given problem of size n we allowed all possible distributed servers with arbitrary broadcast link capacities. We derived a more general outer bound on the capacity region, compared to the existing centralized (single-server) outer bound. We showed the necessity of using customized Shannon-type inequalities to establish the capacity region of some problems with $n = 3$ messages. Moreover, we showed instances of the problems of size $n = 3$, whose graph structural properties are fundamentally different from their centralized index coding counterparts. Our proposed inner bound was based on generally partitioning servers into groups, solving the composite coding problem for each group, and then combining the achievable rates. We showed the usefulness of this scheme to obtain tight sum rates through a numerical example for $n = 4$.

We are currently working on more general composite coding schemes. One direction for future research is to derive tighter outer bounds on the capacity region. Another general direction is to study the graph structural properties of the problem and other graph-based inner and outer bounds.

APPENDIX A
PROOF OF THEOREM 1

Consider a distributed index code for the problem $(j|A_j)$, $j \in [n]$. Let X_j be the uniform random variable over $\{0,1\}^{t_j}$ corresponding to message $j \in [n]$ and Y_J be the uniform random variable over $\{0,1\}^{s_J}$ corresponding to the output of server $J \in N$. For each $T \in N$ we set $x_j = \emptyset$, $j \notin T$. Therefore, $Y_J = \emptyset$, if $J \cap T = \emptyset$. Let $Y \doteq (Y_J, J \cap T \neq \emptyset)$. For $j \in T$

$$\begin{aligned} t_j = H(X_j) &= I(X_j; Y | X(A_j \cap T)) \\ &= H(Y|X(A_j \cap T)) - H(Y|X_j, X(A_j \cap T)). \end{aligned}$$

Define

$$f_T(S) = \sum_{J:J\cap T\neq\emptyset} C_J \quad \frac{H(Y|X(S^c\cap T))}{H(Y)}. \tag{19}$$

Then the set function f_T satisfies $f_T(\emptyset)=0$, and $f_T(T)=\sum_{J:J\cap T\neq\emptyset} C_J$. Let $A\subseteq B\subseteq T$ ($B^c\subseteq A^c$), then $f_T(A)\leq f_T(B)$ since conditioning reduces entropy. We show that the set function f_T is submodular. Consider $A,B\subseteq T$, then

$$f_T(A\cup B)-f_T(A)=\frac{\sum_{J:J\cap T\neq\emptyset} C_J}{H(Y)}\times$$
$$(H(Y|X(A^c\cap B^c\cap T))-H(Y|X(A^c\cap T))), \tag{20}$$
$$f_T(B)-f_T(A\cap B)=\frac{\sum_{J:J\cap T\neq\emptyset} C_J}{H(Y)}\times$$
$$(H(Y|X(B^c\cap T))-H(Y|X((A^c\cup B^c)\cap T))). \tag{21}$$

We have

$$\begin{aligned}&H(Y|X(A^c\cap B^c\cap T))-H(Y|X(A^c\cap T))\\ &=I(Y;X(A^c\cap B\cap T)|X(A^c\cap B^c\cap T))\\ &=H(X(A^c\cap B\cap T))\\ &-H(X(A^c\cap B\cap T)|Y,X(A^c\cap B^c\cap T)),\end{aligned} \tag{22}$$

and similarly

$$\begin{aligned}&H(Y|X(B^c\cap T))-H(Y|X((A^c\cup B^c)\cap T))\\ &=H(X(A^c\cap B\cap T))\\ &-H(X(A^c\cap B\cap T)|Y,X(B^c\cap T)).\end{aligned} \tag{23}$$

Since conditioning reduces entropy, (20)-(23) imply

$$f_T(A\cup B)-f_T(A)\leq f_T(B)-f_T(A\cap B),$$

which completes the proof of the submodularity of the set function f_T defined in (19). Let $(\mathbf{R},\mathbf{C})$ be achievable. Then for every $T\in N$ and every $j\in T$ we have

$$\begin{aligned}R_j\leq\frac{t_j}{r}&=(f_T(B_j\cup\{j\})-f_T(B_j))\,\frac{H(Y)}{r\sum_{J:J\cap T\neq\emptyset} C_J}\\ &\leq(f_T(B_j\cup\{j\})-f_T(B_j))\,\frac{\sum_{J:J\cap T\neq\emptyset} s_J}{r\sum_{J:J\cap T\neq\emptyset} C_J}\\ &\leq f_T(B_j\cup\{j\})-f_T(B_j).\end{aligned}$$

Appendix B
Proof of inequality (11)

Consider a distributed index code for the problem $(1);(2|3);(3|2)$. Let X_j be the uniform random variable over $\{0,1\}^{t_j}$ corresponding to message $j\in\{1,2,3\}$ and Y_J be the uniform random variable over $\{0,1\}^{s_J}$ corresponding to the output of server $J\in N_3=\{J\subseteq\{1,2,3\},J\neq\emptyset\}$. Since the output of server J is a function of the messages available at that server,

$$H(Y_J|X_i,i\in J)=0,\quad J\in N_3, \tag{24}$$

and due to the exact recovery condition at each receiver

$$H(X_1|Y_J,J\in N_3)=0, \tag{25}$$
$$H(X_2|(Y_J,J\in N_3),X_3)=0, \tag{26}$$
$$H(X_3|(Y_J,J\in N_3),X_2)=0. \tag{27}$$

The independence of the messages and (24)-(27) leads to

$$\begin{aligned}&H(X_2\,|\,X_1)=H(X_2\,|\,X_1,X_3)\\ &=H(X_2\,|\,X_1,X_3,Y_{\{1\}},Y_{\{3\}},Y_{\{1,3\}})\\ &=H(X_2,Y_{\{2\}},Y_{\{1,2\}},Y_{\{2,3\}},Y_{\{1,2,3\}}\,|\,X_1,X_3,Y_{\{1\}},Y_{\{3\}},Y_{\{1,3\}})\\ &=H(Y_{\{2\}},Y_{\{1,2\}},Y_{\{2,3\}},Y_{\{1,2,3\}}\,|\,X_1,X_3,Y_{\{1\}},Y_{\{3\}},Y_{\{1,3\}})\\ &\leq H(Y_{\{2\}},Y_{\{1,2\}},Y_{\{2,3\}},Y_{\{1,2,3\}}\,|\,X_1).\end{aligned} \tag{28}$$

Next consider

$$\begin{aligned}&I(X_2;Y_{\{2\}},Y_{\{1,2\}},Y_{\{2,3\}},Y_{\{1,2,3\}}\,|\,X_1) &(29)\\ &=H(X_2\,|\,X_1)-H(X_2\,|\,Y_{\{2\}},Y_{\{1,2\}},Y_{\{2,3\}},Y_{\{1,2,3\}},X_1)\\ &=H(Y_{\{2\}},Y_{\{1,2\}},Y_{\{2,3\}},Y_{\{1,2,3\}}\,|\,X_1)\\ &\quad-H(Y_{\{2,3\}},Y_{\{1,2,3\}}\,|\,X_1,X_2). &(30)\end{aligned}$$

Comparing (28) and (29) we get

$$\begin{aligned}&H(X_2\,|\,Y_{\{2\}},Y_{\{1,2\}},Y_{\{2,3\}},Y_{\{1,2,3\}},X_1)\\ &\leq H(Y_{\{2,3\}},Y_{\{1,2,3\}}\,|\,X_1,X_2).\end{aligned} \tag{31}$$

We have

$$\begin{aligned}&H(X_1,X_2,X_3\,|\,Y_J,J\in N_3)=H(X_2\,|\,(Y_J,J\in N_3),X_1)\\ &\leq H(Y_{\{2,3\}},Y_{\{1,2,3\}}\,|\,X_1,X_2) &(32)\\ &\leq H(Y_{\{2,3\}},Y_{\{1,2,3\}}), &(33)\end{aligned}$$

where (32) follows from (31).

$$\begin{aligned}&t_1+t_2+t_3=H(X_1,X_2,X_3)\\ &=H(X_1,X_2,X_3)+H(Y_J,J\in N_3\,|\,X_1,X_2,X_3)\\ &=H(Y_J,J\in N_3)+H(X_1,X_2,X_3\,|\,Y_J,J\in N_3)\\ &\leq H(Y_J,J\in N_3)+H(Y_{\{2,3\}},Y_{\{1,2,3\}}) &(34)\\ &\leq\sum_{J\in N_3}s_J+s_{\{2,3\}}+s_{\{1,2,3\}}\\ &\leq r\left(\sum_{J\in N_3}C_J+C_{\{2,3\}}+C_{\{1,2,3\}}\right)=9r, &(35)\end{aligned}$$

where (34) follows from (33) and (35) follows from the assumption of $C_J=1$, $J\in N_3$. Therefore, $R_1+R_2+R_3\leq 9$.

References

[1] Y. Birk and T. Kol, "Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2825–2830, Jun. 2006.

[2] M. Effros, S. El Rouayheb, and M. Langberg, "An equivalence between network coding and index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 967–971.

[3] L. Ong, C. K. Ho, and F. Lim, "The single-uniprior index-coding problem: The single-sender case and the multi-sender extension," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3165–3182, Jun. 2016.

[4] F. Arbabjolfaei, B. Bandemer, Y.-H. Kim, E. Sasoglu, and L. Wang, "On the capacity region for index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 962–966.

[5] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[6] M. Gadouleau and S. Riis, "Graph-theoretical constructions for graph entropy and network coding based communications," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6703–6717, Oct. 2011.

[7] M. Tahmasbi, A. Shahrasbi, and A. Gohari, "Critical graphs in index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2014, pp. 281–285.

[8] F. Arbabjolfaei and Y.-H. Kim, "Structural properties of index coding capacity using fractional graph theory," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 1034–1038.