On Independent Distributed Source Coding Problems with Exact Repair

Congduan Li*, Fangwei Ye†, Xuan Guang†, Zhiheng Zhou†, Chee Wei Tan*, Raymond W. Yeung†

*Department of Computer Science, City University of Hong Kong,

† Institute of Network Coding, The Chinese University of Hong Kong

***Abstract*—In conventional distributed storage exact repair problems, all sources are reconstructed when the decoder has access to a certain number of encoders (disks). So, the underlying reconstruction network is equivalent to a single-source problem. This paper considers a variant of the exact repair problem, where the underlying reconstruction network is the independent distributed source coding problem, a type of multi-source problem. As the first non-trivial case with two sources and three encoders, the storage-repair tradeoff regions are proved for all the 33 instances, and it is shown that binary codes are optimal.**

***Index Terms*—Independent distributed source coding, exact repair, distributed storage**

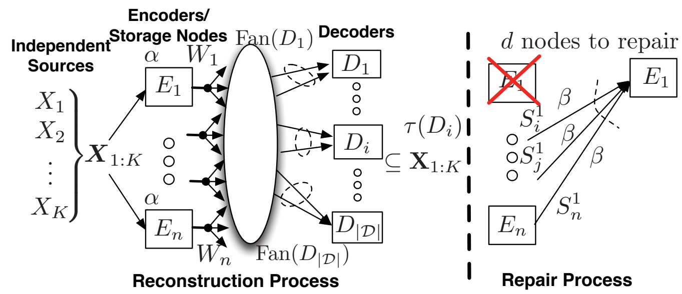


Figure 1. Diagram of an IDSC with exact repair: the reconstruction part is an IDSC problem while the exact repair constraints are added.

I. Introduction

The distributed storage problem has drawn a lot of attention in recent years. In [1], they consider a system that stores files coded in a distributed manner on n disks such that any k disks can reconstruct all the files and a failed disk can be repaired by any other d disks. If the repaired disk only needs to work functionally the same as the failed disk, it is called *functional repair*. If the repaired disk needs to store the same content as the failed disk, it is called *exact repair*. The measure of interest is the tradeoff between the disk capacity α and the repair bandwidth β that limits the information transmitted from each helper in the repair process. We can assume that the transmission is error-free.

In the conventional setup, the reconstruction part is actually a single source problem, since all decoders require all the source files. In this paper, we are interested in a variant on the reconstruction part, where decoders may decode different subsets of the files. We formulate the underlying reconstruction part as a multi-source problem. Specifically, we regard the reconstruction part as the independent distributed source coding (IDSC) problem [2], [3], which is a class of multi-source multicast network problem [4]. Thus, we denote our problem model as IDSC with exact repair (IDSCER). We are interested in the tradeoff between storage and repair bandwidth in terms of the source entropies. One special case was studied in [5] and the inner bound achieved by binary codes was given.

The main contributions of this paper are: i) the tradeoff region of general IDSCER is formulated (§II); ii) all 33 two-source three-encoder IDSCER instances are solved and it is shown that binary linear codes are optimal (§III); iii) the results show that the inner bound provided in [5] is actually the exact region of the problem considered in it (§III).

II. Problem Formulation

We start with an introduction to the model of IDSC with exact repair, and then define its tradeoff region.

A. System Model

The IDSCER model considered in this paper can be viewed as an IDSC problem [2] with exact repair constraints on the encoders, or equivalently, an exact repair problem in distributed storage [1] with the reconstruction part being an IDSC problem. Specifically, in an IDSCER (Fig. 1), there are K independent sources $X_1, \dots, X_K$, which are coded in a distributed manner by n encoders $E_1, \dots, E_n$ with same capacities α. In the reconstruction process, a decoder D_i has access to a subset $\mathrm{Fan}(D_i)$ of the encoders and needs to recover a subset $\tau(D_i)$ of the sources, where $\tau(D_i)$ can be different for different D_i. Denote the set of all decoders as $\mathcal{D}$. In the repair process, when an encoder fails, every size-d subset of other encoders can exactly repair this failed node, with the repair bandwidth β at each helper node. We denote the coded (or stored) messages on encoders as $W_1, \dots, W_n$ and the transmitted message from the helper E_i to repair a failed encoder, say E_j, as S_i^j.

For a fixed (K, n) tuple, there exist many valid IDSC instances. Details about the constraints on possible reconstructions can be found in [3]. The tuple (K, n, d) is used to denote all IDSCER instances with K sources, n encoders, and d helpers in the exact repair process when an encoder fails.

B. Tradeoff Region Formulation

For an (K, n, d) IDSCER instance, we define an $(N, \boldsymbol{\omega}, \alpha, \beta)$ block code, with $\boldsymbol{\omega} = [H(X_1), \dots, H(X_K)]$.

978-1-5090-3097-2/17/$31.00 ©2017 IEEE

(i) The K mutually independent block sources $X_i^{(N)}, i \in \{1,\ldots,K\}$ are uniformly distributed in $\mathcal{X} = \{1,\ldots,\lceil 2^{NH(X_i)} \rceil\}$.

(ii) The block encoders, one for each encoder i, are functions that map a block of N source observations, denoted as $m = [m_1,\ldots,m_K]$, from all sources to one of $\lceil 2^{N\alpha} \rceil$ different descriptions in $\mathcal{W}_i = \{0,1,\ldots,\lceil 2^{N\alpha} \rceil - 1\}$,

$$\phi_i^{(N)} : \prod_{j\in\{1,\ldots,K\}} \mathcal{X}_j \to \mathcal{W}_i, i \in \{1,\ldots,n\}. \quad (1)$$

(iii) The exact repair constraints indicate that when an encoder E_k fails, any d helpers $\mathcal{E}_d^k$ can exactly repair E_k with transmitted messages $S_i^k, E_i \in \mathcal{E}_d$ from the helper nodes. That is, for any $\mathcal{E}_d^k \subseteq \{E_1,\ldots,E_n\} \setminus E_k, |\mathcal{E}_d^k| = d, k \in \{1,\ldots,n\}$,

$$\psi_k^{(N)} : \prod_{E_i\in\mathcal{E}_d^k} \mathcal{S}_i^k \to \mathcal{W}_k, \quad (2)$$

where each transmitted message S_i^k for repair is a mapping from the corresponding stored message,

$$f_i^k : \mathcal{W}_i \to \mathcal{S}_i^k, \mathcal{S}_i^k = \{0,1,\ldots,\lceil 2^{N\beta} \rceil - 1\}. \quad (3)$$

In addition, it is required that the repaired content should be exactly the same as stored in E_k, that is,

$$\psi_k^{(N)}(\prod_{E_i\in\mathcal{E}_d^k} S_i^k) = \phi_k^{(N)}(m). \quad (4)$$

(iv) The block decoders are functions that map observations of $\mathrm{Fan(D_i)}$ to the sources $\tau(D_i)$,

$$\mu_i^{(N)} : \prod_{j\in\mathrm{Fan(D_i)}} \mathcal{W}_j \to \prod_{k\in\tau(D_i)} \mathcal{X}_k, \quad (5)$$

with the reconstructed messages being the same as observations $\{m_k : k \in \tau(D_i)\}$.

After defining a code, we now define the storage-repair tradeoff region.

Definition 1. *The storage-repair tradeoff region $\mathcal{R}_{K,n,d}$ of an IDSCER is the closure of the set of all achievable vectors $(\boldsymbol{\omega},\alpha,\beta)$, where a vector is achievable if there exist a sufficiently large N, a sequence of encoders $\{\phi^{(N)} = [\phi_i^{(N)}|i = 1,\ldots,n]\}$ satisfying the storage and repair bandwidth limits, a sequence of repair functions $\psi^N = [\psi_k^N|k \in \{1,\ldots,n\}]$, and decoders $\{\mu^{(N)} = [\mu_i^{(N)}|i = 1,\ldots,|\mathcal{D}|\}$ satisfying the exact repair and reconstruction constraints, with a varnishing error $\epsilon \to 0$, as $N \to \infty$.*

It is not difficult to see that the relations between the entropies of the variables in the network are as follows.

$$H(\mathbf{X}_{1:K}^N) = \sum_{i=1}^K NH(X_i), \quad (6)$$

$$H(W_i|\mathbf{X}_{1:K}^N) = 0, i = 1,\ldots,n, \quad (7)$$

$$H(S_i^k|W_i) = 0, i,k = 1,\ldots,n, i \neq k \quad (8)$$

$$H(W_k^N|S_i^k, i : E_i \in \mathcal{E}_d^k) = 0, k = 1,\ldots,n \quad (9)$$

$$H(\tau(D_i)^N|\mathbf{W}_{\mathrm{Fan(D_i)}}) = 0, i = 1,\ldots,|\mathcal{D}|, \quad (10)$$

$$H(W_i) \leq \alpha, i \in \{1,\ldots,n\} \quad (11)$$

$$H(S_i^k) \leq \beta, i,k \in \{1,\ldots,n\}, i \neq k \quad (12)$$

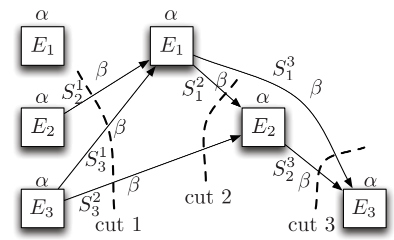


Figure 2. Some cuts in the repair process used in the converse proof. It is assumed that E_1, E_2, E_3 are repaired in a sequence.

where (6) represents the independence of sources, (7) represents the encoding constraints for storage, (8) represents the encoding constraints for generating repair message at each encoder, (9) represents the repair constraints, (10) represents the decoding/ reconstruction constraints, (11) represents the storage capacity constraints, and (12) represents the repair bandwidth constraints.

III. 2-Source 3-Encoder IDSC with Exact Repair

In this section, we fully characterize the storage-repair tradeoff region of the 2-source 3-encoder IDSCER with $(K,n,d) = (2,3,2)$, which is the first non-trivial tuple. According to [3], there are 33 non-isomorphic IDSC instances in total. For each of them, we will characterize its exact tradeoff region. Note that, though the reconstruction of the instances are different, the tradeoff regions can be exactly the same or are permutations of each other, since the key constraints determining the characterization can be the same. We will classify the 33 instances into 7 groups and select one representative for each group, since the proofs of the others follow a similar argument. We will use the indices in [6] for the instances proved in this section.

For each group of the network instances, we will use a distinct theorem to present their storage-repair tradeoff regions. Fig. 2 may be helpful in understanding the converse proofs by illustrating the repair processes and some cuts. The topologies of the representatives and their achievability proofs can be found in Table I, where the constructions of codes to achieve each extreme ray in the tradeoff regions are provided. Hence, in the following, we will only show the converse proof following each theorem, where we assume that a point $\boldsymbol{r}$ is selected in the tradeoff region and there exists an $(N,\boldsymbol{r})$ block code to achieve it.

Theorem 1. *The tradeoff region of cases $1,4,6,7,17$ of the $(2,3,2)$ IDSCER problems contains all rate tuples characterized by the following inequalities:*

$$\alpha \geq H(X) + H(Y) \quad (13)$$

$$2\beta \geq H(X) + H(Y) \quad (14)$$

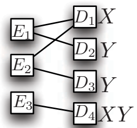
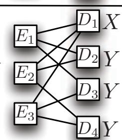
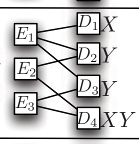
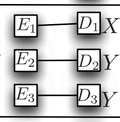
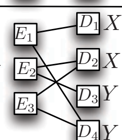
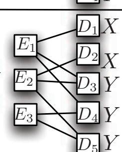
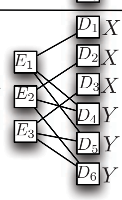
Proof. We will show the converse proof for case 17 here.

For (13), we have

$$N\alpha \geq H(W_3) \geq H(X^N Y^N) = N(H(X) + H(Y)) \quad (15)$$

which follows from the capacity constraint on E_3, the decoding constraint $H(XY|W_3) = 0$ at D_4, and the source independence.

Table I
ALL REPRESENTATIVE CASES OF THE 2-SOURCE 3-ENCODER IDSC WITH EXACT REPAIR AND THE CORRESPONDING ACHIEVABILITY PROOFS.

Cases	Topology	Extreme rays $(\alpha, \beta, H(X), H(Y))$	Achieving Codes (W_1, W_2, W_3)
case 17 Equivalent: (1, 4, 6, 7)	XY 	2 1 2 0 2 1 0 2	(X_1X_2, X_1X_2, X_1X_2) (Y_1Y_2, Y_1Y_2, Y_1Y_2)
case 25 Equivalent: (2, 3, 15, 16, 20, 23)	XY 	2 1 3 0 2 1 0 3 1 1 2 0 1 1 0 2	(X_1X_2, X_2X_3, X_3X_1) (Y_1Y_2, Y_2Y_3, Y_3Y_1) (X_1, X_2, X_1+X_2) (Y_1, Y_2, Y_1+Y_2)
case 12 Equivalent: (8, 11, 13, 14, 19, 31)	XY 	1 1 0 2 2 1 0 3 2 1 2 0 2 1 2 1 1 1 1 1	(Y_1, Y_2, Y_1+Y_2) (Y_1Y_2, Y_2Y_3, Y_1Y_3) (X_1X_2, X_1Y, X_2Y) (X_1X_2, X_1Y, X_2Y) $(X, Y, X+Y)$
case 9 Equivalent: (26)	XY 	2 1 0 2 2 1 2 1 2 1 2 0	(Y_1Y_2, Y_1Y_2, Y_1Y_2) (X_1X_2, X_1Y, X_2Y) (X_1X_2, X_1Y, X_2Y)
case 30 Equivalent: (10)	XY 	2 1 0 2 2 1 1 2 2 1 2 0 2 1 2 1 1 1 1 1	(XY_1, Y_1Y_2, XY_2) (XY_1, Y_1Y_2, XY_2) (X_1X_2, X_1Y, X_2Y) (X_1X_2, X_1Y, X_2Y) $(X, Y, X+Y)$
case 29 Equivalent: (18, 22, 27, 28)	XY 	2 1 0 3 2 1 1 2 2 1 2 0 1 1 0 2	(Y_1Y_2, Y_2Y_3, Y_3Y_1) (XY_1, XY_2, Y_1Y_2) (X_1X_2, X_1X_2, X_1X_2) (Y_1, Y_2, Y_1+Y_2)
case 33 Equivalent: (21, 24, 32)	XY 	2 1 0 3 2 1 2 0 1 1 0 2	(Y_1Y_2, Y_2Y_3, Y_3Y_1) (X_1X_2, X_1X_2, X_1X_2) (Y_1, Y_2, Y_1+Y_2)

For (14), we consider the cut 3 in Fig. 2 in repairing E_3. Then, we have

$$2N\beta \geq H(S_1^3) + H(S_2^3) \geq H(W_3). \tag{16}$$

Here, (16) follows from the repair bandwidth constraints and the repair constraints $H(W_3|S_1^3S_2^3) = 0$, and the proof follows from the second inequality in (15). □

Theorem 2. *The tradeoff region of cases* $2, 3, 15, 16, 20, 23, 25$ *of the* $(2, 3, 2)$ *IDSCER problems contains all rate tuples characterized by the following inequalities:*

$$2\alpha \geq H(X) + H(Y) \tag{17}$$

$$3\beta \geq H(X) + H(Y) \tag{18}$$

$$\alpha + \beta \geq H(X) + H(Y) \tag{19}$$

Proof. We will show the converse proof for case 25 here.

For (17), we have

$$2N\alpha \geq H(W_1) + H(W_2) \geq H(W_1W_2) = H(W_1W_2W_3) \geq H(X^NY^N) = N(H(X) + H(Y)) \tag{20}$$

where (20) follows from the storage capacity constraints, the fundamental property of joint entropy, and the fact that W_3 can be repaired by W_1 and W_2 together, i.e., $H(W_3|W_1W_2) = 0$. Then, the rest follows from the decoding constraints at D_1, and the source independence.

For (18), we consider the cut 1 in Fig. 2 in repairing E_1 and E_2. Then, we have

$$3N\beta \geq H(S_3^1) + H(S_3^2) + H(S_2^1) \geq H(W_1W_2). \tag{21}$$

Here, (21) follows from the repair bandwidth constraints and the repair constraint $H(W_3|S_1^3S_2^3) = 0$, and the rest follows from (20)..

For (19), we have

$$N(\alpha + \beta) \geq H(W_3) + H(S_2^1) \geq H(S_3^1S_3^2) + H(S_2^1) \geq H(S_3^1S_3^2S_2^1) \geq H(W_1S_1^2S_3^2) \geq H(W_1W_2). \tag{22}$$

Here, (22) follows from the storage and repair bandwidth constraints, repair coding constraints at E_3, fundamental property of joint entropy, and the repair processes of W_1 and W_2, i.e.,

$H(W_1|S_2^1S_3^1)=0$ and $H(W_2|S_1^2S_3^2)=0$, and the rest follows from (20). . □

Theorem 3. *The tradeoff region of cases* $8,11,12,13,14,19,31$ *of the* $(2,3,2)$ *IDSCER problems contains all rate tuples characterized by* (17)*,* (18)*,*(19)*, and the following inequalities:*

$$\alpha \geq H(X) \tag{23}$$
$$2\beta \geq H(X) \tag{24}$$

Proof. We will show the converse proof for the new inequalities (23)–(24) for case 12 here, since the others can be proved similarly as (18)–(19).

(23) can be derived by the storage capacity constraint on E_1 and then the decoding constraint at D_1. (24) can be derived by considering repairing E_1, that is

$$2N\beta \geq H(S_2^1)+H(S_3^1) \geq H(S_2^1,S_3^1) \geq H(U_1) \geq NH(X). \tag{25}$$

□

Theorem 4. *The tradeoff region of cases* 9 *of the* $(2,3,2)$ *IDSCER problems contains all rate tuples characterized by* (23)*,*(24)*, and the following inequalities:*

$$2\alpha \geq H(X)+2H(Y) \tag{26}$$
$$4\beta \geq H(X)+2H(Y) \tag{27}$$

Proof. Similarly, we will only show the proofs for (26) and (27).

For (26), we have

$$\begin{aligned} 2N\alpha &\geq H(W_2)+H(W_3)=H(W_2Y^N)+H(W_3Y^N)\\ &\geq H(W_2W_3Y^N)+H(Y^N)\\ &\geq H(W_1W_2W_3Y^N)+H(Y^N)\\ &\geq NH(XY)+NH(Y)=NH(X)+2NH(Y), \end{aligned} \tag{28}$$

which follows from the storage capacity constraints on E_2,E_3, decoding abilities of D_2,D_3, non-negativity of conditional mutual information, the fact that W_1 can be repaired by W_2,W_3, decoding abilities of all decoders, and then the source independence.

For (27), we consider cut 2 and cut 3 in Fig. 2. Then, we will have

$$\begin{aligned} 4N\beta &\geq H(S_3^2)+H(S_1^2)+H(S_1^3)+H(S_2^3)\\ &\geq H(W_2)+H(W_3). \end{aligned} \tag{29}$$

Then, the rest follows from (28). Here, (29) simply follows from the repair bandwidth constraints and repair processes of W_2,W_3. □

Theorem 5. *The tradeoff region of cases* $10,30$ *of the* $(2,3,2)$ *IDSCER problems contains all rate tuples characterized by* (23)*,*(24)*,*(18)*,*(19)*, and the following inequalities:*

$$\alpha \geq H(Y) \tag{30}$$
$$2\beta \geq H(Y) \tag{31}$$

Proof. We will show the proofs for (30) and (31) for case 30. (30) can be derived by considering the storage capacity constraint on E_2 and the decoding ability at D_3. Similarly, (31) can be derived by considering the repair of W_2 and then the decoding ability at D_3. □

Theorem 6. *The tradeoff region of cases* $18,22,27,28,29$ *of the* $(2,3,2)$ *IDSCER problems contains all rate tuples characterized by* (26)*,*(27)*,*(18)*, and* (19)*.*

Proof. Though we select case 29 for the achievability proof in Table I, the converse proof is omitted here, since it can be derived from similar arguments above, with a permutation of variables X,Y. □

Theorem 7. *The tradeoff region of cases* $21,24,32,33$ *of the* $(2,3,2)$ *IDSCER problems contains all rate tuples characterized by the following inequalities:*

$$2\alpha \geq 2H(X)+H(Y) \tag{32}$$
$$6\beta \geq 3H(X)+2H(Y) \tag{33}$$
$$2\alpha+2\beta \geq 3H(X)+2H(Y) \tag{34}$$

Proof. We will show the proofs for case 33.

(32) can be derived similarly as (26) with permuting X,Y.

For (33), we will consider cut 1, cut 2, and cut 3 in Fig. 2. Then, we have

$$\begin{aligned} 12N\beta \geq\, &2(H(S_2^1)+H(S_3^1)+H(S_1^2)\\ &+H(S_3^2)+H(S_1^3)+H(S_2^3))\\ \geq\, &2(H(S_2^1S_3^1W_1X^N)+H(S_1^2S_3^2W_2X^N)\\ &+H(S_1^3S_2^3W_3X^N)) \qquad (35)\\ =\, &2(H(S_2^1S_3^1S_1^2S_1^3W_1X^N)+H(S_1^2S_3^2S_2^1S_2^3W_2X^N)\\ &+H(S_1^3S_2^3S_3^1S_3^2W_3X^N)), \qquad (36) \end{aligned}$$

where (35) follows from the repair processes of E_1,E_2,E_3 and the decoding abilities at D_1,D_2,D_3, and (36) follows from the fact that all transmitted messages for repair are functions of corresponding stored (coded) messages at the encoders.

Next, we apply submodularity property (or equivalently, the non-negativity of conditional mutual information) to the three pairs of the terms in (36). Then, we have

$$\begin{aligned} 12N\beta \geq\, &H(W_1W_2X^N)+H(W_1W_3X^N)+H(W_2W_3X^N)\\ &+H(S_1^2S_2^1X^N)+H(S_1^3S_3^1X^N)+H(S_2^3S_3^2X^N). \end{aligned} \tag{37}$$

From the decoding abilities, we can see that

$$\begin{aligned} &H(W_1W_2X^N)+H(W_1W_3X^N)+H(W_2W_3X^N)\\ &\geq 3H(X^NY^N)=3NH(X)+3NH(Y). \end{aligned} \tag{38}$$

For the other three terms in (37), we will apply the submodularity property (or equivalently, the non-negativity of conditional mutual information) step to step and the decoding abilities. Then, we have

$$\begin{aligned} &H(S_1^2S_2^1X^N)+H(S_1^3S_3^1X^N)+H(S_2^3S_3^2X^N)\\ &\geq H(S_1^2S_2^1S_1^3S_3^1X^N)+H(X)+H(S_2^3S_3^2X^N) \qquad (39)\\ &\geq H(S_1^2S_2^1S_1^3S_3^1S_2^3S_3^2X^N)+2H(X^N) \qquad (40)\\ &\geq H(X^NY^N)+2H(X^N) \qquad (41)\\ &\geq 3NH(X)+NH(Y). \qquad (42) \end{aligned}$$

Then, from (37), (38), and (42), we have

$$12\beta \geq 6H(X) + 4H(Y), \tag{43}$$

which is equivalent to (33).

For (34), we have

$$\begin{aligned}
&6N\alpha + 6N\beta \geq 2(H(W_1) + H(W_2) + H(W_3)) + H(S_2^1) \\
&+H(S_3^1) + H(S_1^2) + H(S_3^2) + H(S_1^3) + H(S_2^3) \quad (44)\\
&\geq 2H(W_1X^NS_1^2S_1^3) + 2H(W_2X^NS_2^1S_2^3) \\
&+2H(W_3X^NS_3^1S_3^2) + H(S_2^1S_3^1) \\
&+H(S_1^2S_3^2) + H(S_1^3S_2^3) \quad (45)\\
&\geq 2H(W_1X^NS_1^2S_1^3) + 2H(W_2X^NS_2^1S_2^3) \\
&+2H(W_3X^NS_3^1S_3^2) + H(S_2^1S_3^1W_1X^N) \\
&+H(S_1^2S_3^2W_2X^N) + H(S_1^3S_2^3W_3X^N) \quad (46)
\end{aligned}$$

where (45) follows from the encoding constraints for generating repair messages at each encoder and the fundamental property of joint entropies, while (46) follows from (35).

Note that, by applying the submodularity property (or equivalently, the non-negativity of conditional mutual information) twice, we will have

$$\begin{aligned}
&\frac{1}{2}H(W_2X^NS_2^1S_2^3) + \frac{1}{2}H(W_3X^NS_3^1S_3^2) + H(S_2^1S_3^1W_1X^N) \\
&\geq \frac{1}{2}H(W_1W_2S_2^1S_3^1S_2^3) + \frac{1}{2}H(W_1W_3S_2^1S_3^1S_3^2) \\
&+\frac{1}{2}H(X^NS_2^1) + \frac{1}{2}H(X^NS_3^1) \quad (47)\\
&\geq H(X^NY^N) + \frac{1}{2}H(X^NS_2^1) + \frac{1}{2}H(X^NS_3^1). \quad (48)
\end{aligned}$$

Similarly, we can have

$$\begin{aligned}
&\frac{1}{2}H(W_1X^NS_1^2S_1^3) + \frac{1}{2}H(W_3X^NS_3^1S_3^2) + H(S_1^2S_3^2W_2X^N) \\
&\geq H(X^NY^N) + \frac{1}{2}H(X^NS_1^2) + \frac{1}{2}H(X^NS_3^2), \quad (49)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{2}H(W_1X^NS_1^2S_1^3) + \frac{1}{2}H(W_2X^NS_2^1S_2^3) + H(S_1^3S_2^3W_3X^N) \\
&\geq H(X^NY^N) + \frac{1}{2}H(X^NS_1^3) + \frac{1}{2}H(X^NS_2^3). \quad (50)
\end{aligned}$$

Further, by applying submodularity (or equivalently, the non-negativity of conditional mutual information) twice, we will have

$$\begin{aligned}
&H(W_1X^NS_1^2S_1^3) + \frac{1}{2}H(X^NS_3^2) + \frac{1}{2}H(X^NS_2^3) \\
&\geq \frac{1}{2}H(W_1X^NS_1^2S_1^3S_3^2) + \frac{1}{2}H(W_1X^NS_1^2S_1^3S_2^3) + H(X^N) \\
&\geq \frac{1}{2}H(W_1W_2) + \frac{1}{2}H(W_1W_3) + H(X^N) \quad (51)\\
&\geq H(X^NY^N) + H(X^N) \quad (52)\\
&\geq 2H(X^N) + H(Y^N) \quad (53)
\end{aligned}$$

Similarly, we will have

$$\begin{aligned}
&H(W_2X^NS_2^1S_2^3) + \frac{1}{2}H(X^NS_3^1) + \frac{1}{2}H(X^NS_1^3) \\
&\geq 2H(X^N) + H(Y_N), \quad (54)
\end{aligned}$$

and

$$\begin{aligned}
&H(W_3X^NS_3^1S_3^2) + \frac{1}{2}H(X^NS_2^1) + \frac{1}{2}H(X^NS_1^2) \\
&\geq 2H(X^N) + H(Y^N). \quad (55)
\end{aligned}$$

Applying (48)–(55) to (46), we will have

$$6N\alpha + 6N\beta \geq 9H(X^N) + 6H(X^N) = 9NH(X) + 6NH(Y), \tag{56}$$

that is,

$$2\alpha + 2\beta \geq 3H(X) + 2H(Y). \tag{57}$$

□

From the achievability proofs in Table I, we have the following corollary.

Corollary 1. *Binary linear codes suffice for the 2-source 3-encoder IDSC problems with exact repair.*

Note that, the problem considered in [5] actually is the case 16 in [3] with exact repair. Then, from Theorem 2, we will have the following corollary.

Corollary 2. *In [5], the inner bound provided in Theorem 6 is the exact storage-repair tradeoff region of the problem considered therein.*

IV. Conclusion

This paper investigates the independent distributed source coding with exact repair constraints, which modifies the reconstruction part of the traditional distributed storage exact repair problem. As the first step towards solving the general case, all 33 non-isomorphic IDSCER instances with two sources and three encoders are solved and binary codes are shown to be optimal.

Acknowledgement

This work was in part supported by the Research Grants Council of Hong Kong RGC 11207615, in part by the Hong Kong Innovation and Technology Fund (ITF) Project (ITS/180/16) and in part by the Science, Technology and Innovation Commission of Shenzhen Municipality, under Project 20170221180035973 and JCYJ20170307090810981.

References

[1] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4539–4551, Sept 2010.

[2] R. W. Yeung and Z. Zhang, "Distributed source coding for satellite communications," *IEEE Trans. on Information Theory*, vol. 45, no. 4, pp. 1111–1120, 1999.

[3] C. Li, S. Weber, and J. M. Walsh, "Computer aided proofs for rate regions of independent distributed source coding problems," in *2015 International Symposium on Network Coding (NetCod)*, June 2015, pp. 81–85.

[4] X. Yan, R. Yeung, and Z. Zhang, "An implicit characterization of the achievable rate region for acyclic multisource multisink network coding," *IEEE Transactions on Information Theory*, vol. 58, no. 9, pp. 5625–5639, Sept 2012.

[5] J. Apte, C. Li, S. Weber, and J. Walsh, "Exact repair with multiple sources," in *Conference on Information Sciences and Systems (CISS)*, March 2014.

[6] C. Li, J. M. Walsh, and S. Weber, "Exact Rate Regions and Codes for all (K, E) IDSCs with $K, E \in \{2, 3\}$." available at http://goo.gl/WP80CK.