Cooperative Data Exchange with Weighted Cost based on d-Basis Construction

Su Li, Abhin Shah and Michael Gastpar
EPFL

Abstract— We consider the cooperative data exchange problem, in which nodes are fully connected with each other. Each node initially only has a subset of the K packets making up a file and wants to recover the whole file. Node i can make a broadcast transmission, which incurs cost w_i and is received by all other nodes. The goal is to minimize the total cost of transmissions that all nodes have to send, which is also called weighted cost. Following the same idea of our previous work [1] which provided a method based on d-Basis construction to solve cooperative data exchange problem without weighted cost, we present a modified method to solve cooperative data exchange problem with weighted cost. We present a polynomial-time deterministic algorithm to compute the *minimum weighted cost* and determine the rate vector and the packets that should be used to generate each transmission. By leveraging the connection to Maximum Distance Separable codes, the coefficients of linear combinations of the optimal coding scheme can be efficiently generated. Our algorithm has significantly lower complexity than the state of the art. In particular, we prove that the minimum weighted cost function is a *convex* function of the total number of transmissions for integer rate cases.

I. INTRODUCTION

Cooperative communication is widely used to reduce the traffic from the base station to users, by enabling the users to communicate with each other. Consider a group of users that want to download the same file from the base station. It is common that some packets are not successfully delivered from the base station to users due to the change of the channel quality. Each user is likely to get a subset of the desired file but not the complete file. Instead of waiting for the retransmissions from the base station, users can communicate with each other to recover the file, provided that collectively they have received all packets. By exploiting the packets that have already been successfully received by each node, coded packets can be transmitted and reduce the total number of transmissions as compared to using pure packets as transmissions. Additionally, the links between spatially close users are more reliable than the links between the base station and users. We want to minimize the overall cost of cooperative communications between users. To solve this problem, we need to answer the following two questions: (1) What is the minimum overall cost of communications? (2) What should each node transmit individually? This problem is referred to as *Cooperative Data Exchange* (CDE) or *Communication for Omniscience* with weighted cost.

A. Related Work

Conventionally, in the cooperative data exchange problem, the nodes are assumed to be fully connected and initially have a subset of packets. The optimization goal is to minimize the total number of transmissions [1], [2], [3], [4]. It has been shown that, for the fully connected network cases, the CDE problem can be formulated as an Integer Linear Program with the Slepian-Wolf constraints on all proper subsets of the packet distribution information of nodes. Although the number of constraints grows exponentially with the number of nodes, this integer linear program can be solved by polynomial-time algorithms due to the submodularity of the constraints. A randomized algorithm was proposed to estimate the minimum number of required transmissions [5]. Deterministic algorithms based on Dilworth Truncation optimization were proposed to compute the exact minimum number of required transmissions [3], [4], [6]. It has been recently shown that a deterministic algorithm based on conditional basis construction [1] can solve this optimization problem with lower complexity. The CDE problem has also been extended in heterogeneous directions, including CDE for arbitrarily connected network [7], [8], successive omniscience and priority [9], [10], and optimization on weighted cost or fairness rate allocations [6], [11], [12].

Constructing the linear coding scheme which enables all nodes to recover all packets requires the rate vector that indicates the number of transmissions that should be generated and sent by each node. In the previously mentioned deterministic algorithms, Dilworth Truncation based algorithms [3], [4], [6] can output the rate vector which can be used to generate linear coding schemes by multicast network code construction algorithm [13] or randomized linear coding generation algorithm [14]. However, conditional basis construction based algorithm [1] can output not only the rate vector but also the packet index vectors, which indicate the packets that should be used to generate linear combinations for transmissions. Hence, an efficient way to generate the linear coding scheme was proposed in [1] by leveraging the connection to MDS codes.

B. Contributions

In this paper, we focus on the CDE problem with weighted cost. This is a more general problem since the basic CDE problem is a special case where all nodes have the same weighted cost for making each transmission. Our contributions can be summarized as follows:

(1) Intuitively, solving a more general problem requires equal or higher complexity. We actually show that solving the CDE problem with weighted cost is as complex

as solving the basic CDE problem. In particular, we propose a polynomial time deterministic algorithm based on the conditional basis construction to compute the optimal (i.e. minimum) weighted cost, the corresponding rate vector, and packet index vectors. The complexity of our algorithm is bounded by $\mathcal{O}(N^3K^3\log(K))$[1], which is the same as the complexity of the best-known algorithm for basic CDE problem and significantly lower than the algorithms proposed in [11], [6], [12].

(2) Explicitly, we prove that the minimum weighted cost function is a convex function of the sum rate when the rate allocation vector has integer value entries and can achieve universal recovery. This convexity property of minimum weighted cost function for a fixed sum rate has only been proved for relaxed conditions in [6], where the entries of rate vector can be non-integer.

C. Organization

The rest of this paper is organized as follows. Section II formally introduces the system model and defines the main problem studied in this paper. Section III presents our main results. Section IV proposes our algorithms to compute the optimal weighted cost, rate vector, and packet index vectors.

II. SYSTEM MODEL AND PROBLEM DEFINITIONS

Before we formally introduce the problem, we define some notations. Let $[n]$ denote the set $\{1,\ldots,n\}$. Let $w_H(u)$ denote the number of non-zero entries in vector u. For any set of vectors $\mathbf{U} = \{u_1, u_2, \ldots\}$ and a subset of vector $\mathbf{S} \subseteq \mathbf{U}$: $u_{\mathbf{S}}$ denote the bitwise **OR** of the components of all vectors in set $\mathbf{S}$.

A. System Model

We consider a fully connected network which has N nodes and the desired file is composed of K packets. Let $\mathbf{N} = [N]$ and $\mathbf{P} = \{P_i, i \in [K]\}$ denote the set of nodes and the set of packets, respectively. Each $P_i \in \mathbb{F}$, where $\mathbb{F}$ is some finite field. Without loss of generality, we assume that every packet is initially possessed by at least two nodes and at most $N-1$ nodes[2]. Each node initially only has a subset of packets, denoted by $\mathbf{X}_i$ ($\mathbf{X}_i \subseteq \mathbf{P}, \forall i \in [N]$). The union set of the packets initially available at a subset of nodes $\mathbf{I} \subseteq \mathbf{N}$ is denoted as $\mathbf{X}_{\mathbf{I}} = \bigcup_{i \in \mathbf{I}} \mathbf{X}_i$. We assume that all the nodes collectively have all packets, which means $\mathbf{X}_{\mathbf{N}} = \mathbf{P}$. The notation $\mathbf{X}_{\mathbf{I}}^c = \mathbf{P} \setminus \mathbf{X}_{\mathbf{I}}$ denotes the jointly missing packets at nodes in set $\mathbf{I}$. Let $\mathcal{M} = \min_{i \in \mathbf{N}} |\mathbf{X}_i|$ be the minimal number of initially available packets at any single node.

Definition 1: Define packet distribution matrix E as the matrix with the ith row j^{th} column entry

$$E_{ij} = \begin{cases} 1, & P_j \in \mathbf{X}_i \\ 0, & otherwise \end{cases} \quad (1)$$

Let K-dimensional binary vector e_i, the ith row of E, be the Packet Distribution Vector (PDV) of node i.

Let $\mathbf{T} = \{T_1, \ldots, T_R\}$ denote a linear coding scheme with R transmissions[3]. Each transmission T_i is a linear combination of packets available at the sender node. Let $\mathbf{r} = [r_1, \ldots, r_N]^{\mathsf{T}}$ denote the rate vector where each r_i is the number of transmissions made by node i and $\mathcal{S}(\mathbf{r}) = \sum_{i=1}^{N} r_i$ denote the sum rate of $\mathbf{r}$. In CDE problem, both the sum rate and individual rate of each node should be integers. Let $\mathbf{w} = [w_1, \ldots, w_N]^{\mathsf{T}}$ denote the weight vector where each w_i is the cost of one transmission sent by node i. Then, the cost of a coding scheme with rate vector $\mathbf{r}$ is denoted by $\mathcal{C}(\mathbf{r}) = \mathbf{w}^{\mathsf{T}} \cdot \mathbf{r} = \sum_{i=1}^{N} w_i r_i$.

B. Problem Definition

For the CDE problem with weighted cost, we have to find a coding scheme with rate vector $\mathbf{r} = [r_1, \ldots, r_N]^{\mathsf{T}}$ which has the minimum weighted cost $\mathcal{C}(\mathbf{r})$. After we find the optimal rate vector $\mathbf{r}$, we still need to generate the linear coding scheme which achieves universal recovery. Since generating a linear coding scheme which achieves universal recovery for a given rate vector is exactly the same way as what we can do in the basic CDE problem, in this paper we ignore this part and focus on finding the optimal rate vector.

It has been shown that any rate vector which achieves universal recovery if and only if the following Slepian-Wolf constraints are satisfied [3]:

$$\sum_{i \in \mathbf{N} \setminus \mathbf{I}} r_i \geq |\mathbf{X}_{\mathbf{I}}^c|, \forall \mathbf{I} \subsetneq \mathbf{N} \quad (2)$$

Let $\Omega = \{\mathbf{r} = [r_1, \ldots, r_N]^{\mathsf{T}} : \sum_{i \in \mathbf{N} \setminus \mathbf{I}} r_i \geq |\mathbf{X}_{\mathbf{I}}^c|, \forall \mathbf{I} \subsetneq \mathbf{N}\}$ denote the set of all rate vectors which satisfy the Slepian-Wolf constraints in (2). Then, the minimum weighted cost can be computed as

$$\mathcal{C}^* = \min_{\mathbf{r} \in \Omega} \mathcal{C}(\mathbf{r}) = \min_{\mathbf{r} \in \Omega} \sum_{i=1}^{N} w_i r_i \quad (3)$$

Although the optimization should be over all vectors in Ω, we can actually decompose this optimization problem into two sub-optimization problems. We first find the optimal rate vector under the condition that the sum rate is fixed. Then, the further optimization should only be taken over the optimal rate vectors for all fixed sum rates.

Definition 2: Let $\mathcal{K}(R)$ denote the minimum weighted cost of all rate vectors that can achieve universal recovery and has sum-rate equal to R.

$$\mathcal{K}(R) = \min_{\mathbf{r} \in \Omega, \mathcal{S}(\mathbf{r}) = R} \mathcal{C}(\mathbf{r}) = \min_{\mathbf{r} \in \Omega, \mathcal{S}(\mathbf{r}) = R} \sum_{i=1}^{N} w_i r_i \quad (4)$$

Let R_{min} denote the minimum sum rate such that there exists a rate vector that can achieve universal recovery. Only rate vectors with sum rate between R_{min} and K should

---

be considered. The minimum weighted cost can also be computed as

$$\mathcal{C}^* = \min_{R \in \{R_{min},\ldots,K\}} \mathcal{K}(R)$$

$$= \min_{R \in \{R_{min},\ldots,K\}} \min_{\mathbf{r} \in \Omega, \mathcal{S}(\mathbf{r})=R} \sum_{i=1}^{N} w_i r_i \quad (5)$$

Example 1: Consider a CDE problem for the fully connected network with 5 nodes and 9 packets with the goal of minimizing the weighted cost of transmissions. The packet distribution matrix is as follows:

$$E = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

The weights of nodes are as follows:

Node(i)	1	2	3	4	5
w_i	2	3	6	8	10

By using the methods proposed in [6], [11], we can find that the optimal rate vector is $\mathbf{r}^* = [3,3,1,0,0]^\mathsf{T}$ and the minimum weighted cost is 21. However, for basic CDE problem (unweighted case) with the same packet distribution matrix, the optimal rate vector is $\mathbf{r} = [1,1,1,1,1]^\mathsf{T}$.

Remark 1: By using algorithms in [1], [6], [7], we can show that the minimum sum rate R_{min} for Example 1 is 5. But for CDE problem with weighted cost, the optimal rate vector has sum rate 7, which is larger than the minimum required sum rate. Thus, only finding the rate vector with sum rate R_{min} is not enough, we have to optimize $\mathcal{K}(R)$ over all $R \in \{R_{min},\ldots,K\}$. However, we show that it is not necessary to compute $\mathcal{K}(R)$ for all $R \in \{R_{min},\ldots,K\}$. By exploiting the convexity of the function $\mathcal{K}(R)$, we can search the optimal R and rate vector by using the binary searching style method.

C. Background

In our previous work [1], we presented a method based on constructing d-Basis (see Definition 3) to solve basic CDE problem. For each CDE problem with specific packet distribution matrix E, we first proposed Algorithm 1 [1, Section IV] which searches the existence of any d-Basis that can be generated by packet distribution vectors. If there exists any d-Basis, Algorithm 1 in [1] will output the d-Basis and corresponding rate vector. We proved that the existence of d-Basis is a sufficient and necessary condition for generating linear coding schemes which achieve universal recovery. The total number of transmissions of such linear coding scheme is $K - d$ and each transmission is a linear combination of $d+1$ packets, for any $0 \leq d \leq K - R_{min}$. In particular, packets that are used to generate transmissions are indexed by the basis vectors. Hence, we also call the basis vectors packet index vectors. Additionally, a binary search algorithm is proposed in [1] to find the rate vector with the

minimum sum rate and the corresponding d-Basis which achieve universal recovery. By leveraging the connection between linear coding schemes based on d-Basis and MDS codes, we showed that the coefficient matrix of the optimal coding scheme could be efficiently generated by performing elementary row operations on Vandermonde matrices.

For the CDE problem with weighted cost, we propose Algorithm 1 which is a modified version of Algorithm 1 in [1]. We show that our Algorithm 1 can solve the optimization problem (4) and output the optimal rate vector and corresponding packet index vectors for fixed sum rate. Further, we propose Algorithm 2, a binary search style method, to find the globally optimal rate vector with minimum weighted cost and the corresponding d-Basis vectors. Based on the rate vector and d-Basis vectors, a linear coding scheme can be generated to achieve universal recovery by using the same way as we presented in [1]. In this paper, we focus on finding the optimal rate vector and d-Basis. Now we introduce some necessary definitions.

Definition 3 (d-Basis): A set of K-dimensional binary linearly independent vectors ($\mathbf{V} = \{v_i : i \in [K - d]\}, 0 \leq d \leq K - 1$) is called a d-Basis if

$$w_H(v_i) = d + 1, \qquad \forall i \in [K-d] \quad (6)$$
$$w_H(v_\mathbf{S}) \geq |\mathbf{S}| + d, \qquad \forall \emptyset \neq \mathbf{S} \subseteq \mathbf{V} \quad (7)$$

Definition 4: A binary vector u can generate another binary vector v if u and v have the same dimensions and

$$\{m : v_m = 1\} \subseteq \{n : u_n = 1\}. \quad (8)$$

Moreover, let $\mathcal{G}(u)$ denote the set of all binary vectors that can be generated by u. Define $\mathcal{G}(\mathbf{S}) = \cup_{u \in \mathbf{S}} \mathcal{G}(u)$ and $\mathcal{G}(u, d) = \{v : v \in \mathcal{G}(u), w_H(v) = d + 1\}$.

A set of K-dimensional binary vectors $\mathbf{U} = \{u_1,\ldots,u_L\}$ is able to generate a d-Basis $\{v_i : i \in [K-d]\}$ if $\forall i \in [K-d]$, $v_i \in \mathcal{G}(\mathbf{U}, d)$. Let d^*-Basis denote the d-Basis with largest d that can be generated by given vectors.

Definition 5: For binary vector u and $w_H(u) > d$, define $\mathcal{B}(u, d) = \{b_i \in \mathcal{G}(u, d) : i \in [w_H(u) - d], w_H(b_\mathbf{S}) \geq |\mathbf{S}| + d, \forall \mathbf{S} \subseteq [w_H(u) - d]\}$. For binary vector u and $w_H(u) \leq d$, define $\mathcal{B}(u, d) = \emptyset$.

Set $\mathcal{B}(u, d)$ is a set of binary vectors which are generated by u. Each of the vector has weight $d + 1$ and they satisfy the Constraint (7) of d-Basis. Therefore, they are basis vector candidates for d-Basis.

III. MAIN RESULTS

We state the main result of this paper in the following theorems.

Theorem 1: For any $R \in \{R_{min},\ldots,K\}$ and $d = K-R$, let $\mathbf{r} = [r_1,\ldots,r_N]^\mathsf{T}$ be the output rate vector of Algorithm 1 with input E and d, then $\mathcal{K}(R) = \sum_{i=1}^{N} w_i r_i$.

The proof detail is in Appendix IV-A. The proof outline for Theorem 1 is that we show that for any other rate vector which has the same sum rate as the rate vector $\mathbf{r}$ output by Algorithm 1: (1) If it can achieve universal recovery, it has equal or larger weighted cost than $\mathbf{r}$. (2) It cannot achieve universal recovery, hence it should not be considered.

Remark 2: In words, Theorem 1 says that the output rate vector of Algorithm 1 is the optimal rate vector among all the rate vectors which have sum rate R and can achieve universal recovery.

Now we have a method to get the optimal solution to the sub-optimization problem (4). To get the globally optimal solution to the optimization problem (5), it is sufficient to only consider the rate vectors that are output by Algorithm 1 with different values of input parameter d ($d = K - R$). However, it is not necessary to run Algorithm 1 with all possible $R \in \{R_{min}, \ldots, K\}$, leveraging convexity of the function $\mathcal{K}(R)$ which is stated by the following Theorem. The optimal weighted cost and rate vector can be found by using binary search style method.

Theorem 2: For $R_{min} \leq R \leq K$, the function defined by (4): $\mathcal{K}(R) = \min_{\mathbf{r} \in \Omega, \mathcal{S}(\mathbf{r}) = R} \sum_{i=1}^{N} w_i r_i$ is convex.

The proof detail is in Appendix IV-B. To prove Theorem 2, it is sufficient only to consider the coding schemes with rate vector output by Algorithm 1, since they are the conditional optimal solution for fixed sum rate R. In particular, we exploit some properties of rate vector output by Algorithm 1 to show that the second order difference of $\mathcal{K}(R)$ is non-negative, i.e. $\mathcal{K}(R + 2) + \mathcal{K}(R) - 2\mathcal{K}(R + 1) \geq 0$. By induction, we prove that $\mathcal{K}(R)$ is a convex function of R.

Remark 3: In [6], it has been proved that the function $\mathcal{K}(R)$ defined in (4) is convex for $R_{min} \leq R \leq K$ for a relaxed condition where each entry of $\mathbf{r} = [r_1, \ldots, r_N]^\mathsf{T}$ can be non-integer rate vector. However, the rate vector should always be an integer for the cooperative data exchange problem. The improvement of our theorem is we prove that for integer rate vector, the function $\mathcal{K}(R)$ defined in (4) is still convex for $R_{min} \leq R \leq K$.

Since the function $\mathcal{K}(R)$ is a convex function, it is not necessary to search all possible R to get the optimal solution to the optimization problem (5). We propose Algorithm 2 to compute the minimum weighted cost by using binary searching style method.

IV. ALGORITHMS

We propose an efficient deterministic algorithm based on d-Basis to solve the optimization problem (4). For a given fixed number of transmissions R, Algorithm 1 searches the existence of corresponding d-Basis where $d = K - R$.

Comparing to the Algorithm in [1] which checks the existence of d-Basis for basic CDE problem and output the corresponding d-Basis vectors if they exist, Algorithm 1 requires that the input PDVs should be ordered according to their weights. The nodes with smaller weights have smaller indexes. The node with the smallest weight would be selected to generate as many d-Basis vectors as it can. Then, the nodes with larger weights would be selected to generate d-Basis vectors that can not be generated by previous nodes. And we show that by just ordering the input PDVs according to the ascending order of their weights, Algorithm 1 can find the optimal rate vector and corresponding d-Basis vectors which can achieve universal recovery by using $K - d$ transmissions and has minimum overall weighted cost. Since the algorithm

---

Algorithm 1 Search d-Basis (SdB)

1: **Input:** $E = [e_1, \ldots, e_N]^\mathsf{T}$ ($w_i \leq w_j \ \forall i \leq j$) and d.
2: **Output: True**, **r**, **V** or **False**.
3: **Initialization**: $\mathbf{Q} = \emptyset$, $\mathbf{V} = \emptyset$, $\mathbf{r} = [r_1, \ldots, r_N]^\mathsf{T} = \mathbf{0}_{N \times 1}$.
4: **for** $i : i \in \{1, \ldots, N\}$ **do**
5:     **for** $b \in \mathcal{B}(e_i, d)$ **do**
6:         **if** $b \notin \mathcal{G}(\mathbf{Q}, d)$ **then**
7:             $r_i = r_i + 1$
8:             $\mathbf{V} = \mathbf{V} \cup \{b\}$
9:             **while** $\exists \mathbf{S} \subseteq \mathbf{Q}, |\mathbf{S}| \leq 2 : (9)$ *holds* **do**

$$w_H(q_\mathbf{S} \vee b) \leq \sum_{q_i \in \mathbf{S}} w_H(q_i) + w_H(b) - |\mathbf{S}|d \quad (9)$$

10:                $b = b \vee q_\mathbf{S}$, $\mathbf{Q} = \mathbf{Q} \setminus \mathbf{S}$
11:             $\mathbf{Q} = \mathbf{Q} \cup \{b\}$
12:         **if** $|\mathbf{V}| = K - d$ **then**
13:             return **True**, **r** and **V**
14: return **False**

---

just requires the ordering of PDVs according to their weights, the optimal rate vector would remain the same for a fixed number of transmissions irrespective of the actual individual weights of the PDVs. Since the ordering of the PDVs according to their weights can be done before the start of Algorithm 1 and only requires complexity $\mathcal{O}(N \log(N))$, as compared to the complexity of searching the existence of a d-Basis, which is $\mathcal{O}(N^3 K^3)$ [1], it can be ignored.

Example revisited 1: On applying Algorithm 1 on Example 1 for $d = \{0, 1, 2, 3, 4\}$, we can get the results as shown in Table I. As can be seen from the table, the

TABLE I
SUM RATE, OPTIMAL WEIGHTED COST AND RATE VECTOR

d	R=K-d	$\mathcal{K}(R)$	r_1	r_2	r_3	r_4	r_5
4	5	29	1	1	1	1	1
3	6	22	2	2	2	0	0
2	7	21	3	3	1	0	0
1	8	23	4	3	1	0	0
0	9	25	5	3	1	0	0

minimum cost is achieved by a coding scheme that uses 7 transmissions, which is larger than the minimum number of required transmissions ($R_{min} = 5$) for achieving universal recovery. Additionally, we plot the function $\mathcal{K}(R)$ vs R for example 1 in Fig. 1. As can be seen from the figure, the function $\mathcal{K}(R)$ is convex for $R \in \{5, \ldots, 9\}$. For $R < 5$, there does not exist any coding schemes with sum rate R that can achieve universal recovery. For $R \geq 9$, the number of transmissions is equal to or larger than the number of packets. Hence, the optimal coding scheme in such cases is sending each pure packet without coding.

To solve the optimization problem (5), we propose Algorithm 2, which is a binary search style algorithm to find the minimum of $\mathcal{K}(R)$, i.e. $\mathcal{C}^*$, for $R \in \{R_{min}, \ldots, K\}$.
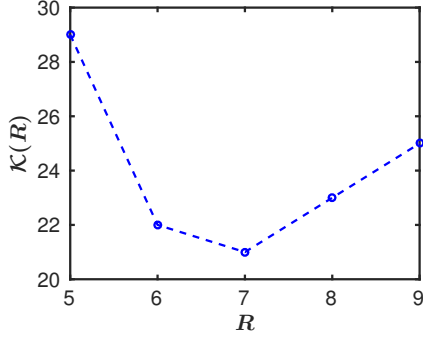
The complexity of binary search algorithm 2 is approx-

Fig. 1. Optimal weighted cost ($\mathcal{K}(R)$) vs Sum rate (R) for Example 1.

Algorithm 2 Finding $\mathbf{r}^*$ and $\mathcal{C}^*$ using Binary Search Algorithm

1: **Input:** $E = [e_1, \ldots, e_N]^\mathsf{T}$, K and $\mathbf{w} = [w_1, \ldots, w_N]^\mathsf{T}$ such that $(w_i \leq w_j \; \forall i \leq j)$
2: **Output:** $\mathbf{r}^*$ and $\mathcal{C}^*$
3: **Initialization:** $d_{start} = 0$, $d_{end} = \mathcal{M}$
4: **while** $d_{start} < d_{end}$ **do**
5:      $d = \max\{\lfloor \frac{d_{start}+d_{end}}{2} \rfloor, d_{start}+1\}$
6:      $(\mathbf{F}, \mathbf{r}, \mathbf{V}) = SdB(E, d)$
7:      **if** $\neg \mathbf{F}$ **then**
8:          $d_{end} = d$
9:      **else**
10:          $\hat{d} = d - 1$
11:          $(\hat{\mathbf{F}}, \hat{\mathbf{r}}, \hat{\mathbf{V}}) = SdB(E, \hat{d})$
12:          **if** $\mathbf{w}^\mathsf{T} \cdot \mathbf{r} > \mathbf{w}^\mathsf{T} \cdot \hat{\mathbf{r}}$ **then**
13:              $d_{end} = \hat{d}$, $\mathbf{r}^* = \hat{\mathbf{r}}$
14:          **else**
15:              $d_{start} = d$, $\mathbf{r}^* = \mathbf{r}$
16: $R^* = K - d$, $\mathcal{C}^* = \mathbf{w}^\mathsf{T} \cdot \mathbf{r}$

---

imately $\mathcal{O}(\log(K))$. Hence, the overall complexity of our two algorithms is $\mathcal{O}(N^3 K^3 \log(K))$ which is the same as complexity as the algorithm of state of the art for basic CDE problem [1].

APPENDIX

A. Proof of Theorem 1

To prove Theorem 1, we first prove two useful Lemmas.

Lemma 1: Let $\mathbf{r}^* = [r_1^*, r_2^*, \ldots, r_N^*]^\mathsf{T}$ denote the rate vector output by Algorithm 1. For any rate vector $\mathbf{r} = [r_1, \ldots, r_N]^\mathsf{T}$ such that $\mathbf{r} \in \Omega$ and $\mathcal{S}(\mathbf{r}^*) = \mathcal{S}(\mathbf{r})$, there does not exist any node pair (i, j) such that $i < j$, $r_i > r_i^*$ and $r_j < r_j^*$.

Proof: If the coding scheme with rate vector $\mathbf{r}$ can achieve universal recovery and uses the same total number of transmissions, then the coding scheme can be implemented as a d-Basis based coding scheme which has the same d value as the coding scheme with rate vector $\mathbf{r}^*$. As Algorithm 1 guarantees that $\forall i \in [N]$, if $r_i^* > 0$, then there must exist as many as $\sum_{n=i}^N r_n^*$ d-Basis vectors that cannot be generated by nodes in set $\{1, 2, \ldots, i-1\}$. If $\exists i < j$ such that, $r_i >$

r_i^* and $r_j < r_j^*$, then $\sum_{j=i}^N r_j < \sum_{j=i}^N r_j^*$ which is not possible as such vectors can only be generated by nodes in set $\{i, i+1 \ldots, N\}$. Hence, there does not exist node pair (i, j) such that $i < j$, $r_i > r_i^*$ and $r_j < r_j^*$. ∎

Lemma 2: Let $\mathbf{r}^* = [r_1^*, r_2^*, \ldots, r_N^*]^\mathsf{T}$ denote the rate vector output by Algorithm 1. If there exists a coding scheme with rate vector $\mathbf{r} = [r_1, \ldots, r_N]^\mathsf{T}$ such that $\mathbf{r} \in \Omega$, $\mathcal{S}(\mathbf{r}^*) = \mathcal{S}(\mathbf{r})$ and there exists node pair (i,j) such that $i < j$, $r_i < r_i^*$ and $r_j > r_j^*$, then $\mathcal{C}(\mathbf{r}) \geq \mathcal{C}(\mathbf{r}^*)$.

Proof: Let $\mathbf{S}_1 = \{i : r_i < r_i^*\}$, $\mathbf{S}_2 = \{j : r_j > r_j^*\}$ and $\mathbf{S}_3 = \{k : r_k = r_k^*\}$. Since $\mathcal{S}(\mathbf{r}^*) = \mathcal{S}(\mathbf{r})$, we have

$$0 = \sum_{n=0}^N (r_n - r_n^*) \tag{10}$$

$$= \sum_{i \in \mathbf{S}_1} (r_i - r_i^*) + \sum_{j \in \mathbf{S}_2} (r_j - r_j^*) + \sum_{k \in \mathbf{S}_3} (r_k - r_k^*) \tag{11}$$

$$= \sum_{i \in \mathbf{S}_1} (r_i - r_i^*) + \sum_{j \in \mathbf{S}_2} (r_j - r_j^*) \tag{12}$$

Hence, for each $i \in \mathbf{S}_1$ such that $r_i < r_i^*$ which sends one less transmission, there must exist one corresponding $j \in \mathbf{S}_2$ which sends one more transmission. According to Lemma 1, if there exists such pair of (i, j), it must satisfy $i < j$ and $w_i \leq w_j$. Let $P = \sum_{i \in \mathbf{S}_1} (r_i^* - r_i) = \sum_{j \in \mathbf{S}_2} (r_j - r_j^*)$ denote the total number of such pairs and $\mathcal{P}$ denote the partition of such pairs. Therefore,

$$\mathcal{C}(\mathbf{r}) - \mathcal{C}(\mathbf{r}^*) = \sum_{i=0}^N w_i r_i - \sum_{i=0}^N w_i r_i^* \tag{13}$$

$$= \sum_{i \in \mathbf{S}_1} w_i (r_i - r_i^*) + \sum_{j \in \mathbf{S}_2} w_j (r_j - r_j^*) \tag{14}$$

$$= \sum_{(i,j) \in \mathcal{P}} (w_j - w_i) \tag{15}$$

$$\geq 0 \tag{16}$$

∎

Now, we are ready to prove Theorem 1.

Proof: [**Theorem 1**] If there exists any linear coding scheme that achieves universal recovery by using $K - d$ transmissions with rate vector $\mathbf{r} = [r_1, \ldots, r_N]^\mathsf{T}$ ($\mathcal{S}(\mathbf{r}) = K - d$), it is always possible to generate a corresponding linear coding scheme based on d-Basis that have the same rate vectors [1]. Hence, they have the same weighted cost and we can only consider the coding schemes based on d-Basis. Let $r^* = [r_1^*, \ldots, r_N^*]^\mathsf{T}$ denote the rate vector output by Algorithm 1. According to Lemma 1, there does not exist any node pair (i, j) such that $i < j$, $r_i > r_i^*$ and $r_j < r_j^*$. Additionally, since $\mathcal{S}(\mathbf{r}^*) = \mathcal{S}(\mathbf{r})$, if rate vector $\mathbf{r}$ is different from $\mathbf{r}^*$, the change can only be $\exists i < j : r_i < r_i^*$ and $r_j > r_j^*$. According to Lemma 2, $\mathcal{C}(\mathbf{r}) \geq \mathcal{C}(\mathbf{r}^*)$. Therefore, the rate vector output by Algorithm 1 has minimum weighted cost in all coding schemes which use $K - d$ transmissions and achieve universal recovery. ∎

B. Proof of Theorem 2

To prove Theorem 2, we first prove two useful Lemmas. Let $\mathbf{r}(l)$ be the rate vector output by Algorithm 1 for input

E and $d = K - l$. Thus, $\mathbf{r}(l)$ is the optimal rate vector with minimum weighted cost among all the rate vectors with $\mathcal{S}(\mathbf{r}) = l$.

Lemma 3: For the coding schemes with rate vectors $\mathbf{r}(l) = [r_{(l,1)}, \ldots, r_{(l,N)}]^{\mathsf{T}}$ with $l \in \{R_{min}, \ldots, K - 1\}$ yielded by Algorithm 1, we have

(1) $r_{(l+1,1)} = r_{(l,1)} + 1$.
(2) $r_{(l+1,m)} \le r_{(l,m)} + 1, \forall 2 \le m \le N$.
(3) If $r_{(l+1,m)} < r_{(l,m)}$, then $r_{(l+2,m)} \le r_{(l+1,m)}$.

Proof: (1) In Algorithm 1, we always start the generation of basis vectors from the PDV of node 1. There is no previously generated basis vector. Thus the number of basis vectors that should be generated by node 1 is

$$r_{(l,1)} = (w_H(e_1) - d)^+ = (w_H(e_1) - K + l)^+ \quad (17)$$

According to Theorem 3 in [1], the minimum number of transmissions required to achieve universal recovery satisfies $R_{min} = K - \min\{\mathcal{M}, d^*\}$. Hence, we have

$$w_h(e_1) \ge \mathcal{M} \ge \min\{\mathcal{M}, d^*\} = K - R_{min} \quad (18)$$

Since $R_{min} \le l \le K$, we have $w_H(e_1) \ge K - l$, which implies that the number of transmissions made by node 1 is always positive. Therefore, for any feasible l, we have $r_{(l+1,1)} = r_{(l,1)} + 1$. This means the first node generates 1 more vector when the total number of transmissions increases by 1. When $r_{(l,1)} = |X_1|$, each transmissions is just a pure packet. In such cases, we have $d = 0$ and $l = K$. Universal recovery can always be achieved when all packets have been sent individually. No coding scheme with more than K transmissions need to be considered.

(2) Similarly, for any $2 \le m \le N$, the total number of feasible basis vectors that can be generated by node m is $w_H(e_m) - K + l$. However, some of them may not be compatible with basis vectors that have been generated by previous nodes. Hence we have

$$r_{(l,m)} = w_H(e_m) - K + l - h_l \quad (19)$$

where h_l is the number of $(K - l)$-Basis vectors that can be generated by node m but not compatible with $(K - l)$-Basis vectors generated by previous node. When the sum rate increases from l to $l + 1$, we need a $(K - l - 1)$-Basis. Thus, we have

$$r_{(l+1,m)} = w_H(e_m) - K + l + 1 - h_{l+1} \quad (20)$$
$$= r_{(l,m)} + 1 + h_l - h_{l+1} \quad (21)$$
$$\le r_{(l,m)} + 1 \quad (22)$$

The last inequality is due to $h_l \le h_{l+1}$. In words, node m can generate at most 1 more basis vector when the total number of transmissions increases by 1.

(3) As the total number of transmissions (sum rate) goes from l to $l+1$, the corresponding basis changes from $(K-l)$-Basis to $(K-l-1)$-Basis. Therefore, the number of packets that are used to generate each transmission decreases by 1. Note that $w_H(e_m) \ge \mathcal{M} \ge K - R_{min}$, $\forall m \in [N]$. When $l = R_{min}$, nodes m with $w_H(e_m) = K - R_{min}$ are not considered to generate any basis vector, since every basis

vector needs $K - R_{min} + 1$ ones. But when $l > R_{min}$, every node is considered to generate basis vectors. If node i is not used to generate any basis vector, that means all basis vectors that can be generated by node i are not compatible with the basis vectors generated by previous nodes. If $r_{(l+1,m)} < r_{(l,m)}$, that means besides the first node, there exists at least one node with lower weight than node m that generates more basis vector(s), i.e. $\exists n$ s.t. $n < m$ and $r_{(l+1,n)} > r_{(l,n)}$. The set of basis vectors that are generated to form $(K - l - 1)$-Basis by node m is a subset of $\mathcal{B}(e_m, K-l-1)$. Let $\mathcal{D}(m, l+1)$ denote vectors in $\mathcal{B}(e_m, K - l - 1)$ but not selected to form $(K - l - 1)$-Basis. Then every vector in $\mathcal{D}(m, l+1)$ is not compatible with $(K - l - 1)$-Basis vectors generated by previous nodes. Any vector in $\mathcal{B}(e_m, K - l - 2)$ which can be generated by vectors in $\mathcal{D}(m, l+1)$ is also not compatible with $(K - l - 2)$-Basis vectors generated by previous nodes. The number of vectors that are in $\mathcal{B}(e_m, K - l - 2)$ and can be generated by vectors in $\mathcal{D}(m, l + 1)$ is $|\mathcal{D}(m, l + 1) + 1|$. Although the total number of basis vectors that can possibly be generated by node m increases by 1 when sum rate increases by 1, the number of incompatible vectors also increases by at least 1. Hence, the maximum number of basis vectors that can be generated by node m for next round is upper bounded by $r_{(l+1,m)}$. Therefore, If $r_{(l+1,m)} < r_{(l,m)}$, then $r_{(l+2,m)} \le r_{(l+1,m)}$, $\forall 2 \le m \le N$. ∎

Definition 6: Let $\mathbf{S}_{(l,\uparrow)}$ denote the set of nodes which generate more number of transmissions when the sum rate increases from l to $l + 1$. Let $\mathbf{S}_{(l,0)}$ denote the set of nodes which generate the same number of transmissions when the sum rate increases from l to $l + 1$. Let $\mathbf{S}_{(l,\downarrow)}$ denote the multiset of nodes which generate fewer transmissions when the sum rate increases from l to $l + 1$. The multiplicity of node i in $\mathbf{S}_{(l,\downarrow)}$ equals $r_{(l,i)} - r_{(l+1,i)}$.

Lemma 4: For $\forall R_{min} \le l \le K - 1$, we have (1) $\mathbf{S}_{(l+1,\uparrow)} \subseteq \mathbf{S}_{(l,\uparrow)}$ and (2) $\max_{i \in \mathbf{S}_{(l+1,\downarrow)}} w_i \le \max_{j \in \mathbf{S}_{(l,\downarrow)}} w_j$.

Proof: Let $\mathbf{r}(l) = [r_{(l,1)}, \ldots, r_{(l,N)}]^{\mathsf{T}}$ and $\mathbf{r}(l + 1) = [r_{(l,1)}, \ldots, r_{(l+1,N)}]^{\mathsf{T}}$ denote the rate vectors output by Algorithm 1 for $d = K - l$ and $d = K - l - 1$, respectively. According to Theorem 1, $\mathbf{r}(l)$ and $\mathbf{r}(l)$ are optimal rate vectors for fixed sum rate l and $l + 1$, respectively.

(1) Assuming the contraction that $\mathbf{S}_{(l+1,\uparrow)} \not\subseteq \mathbf{S}_{(l,\downarrow)}$, then there must exist at least one node k, such that $k \in \mathbf{S}_{(l+1,\uparrow)}$ and $k \notin \mathbf{S}_{(l,\uparrow)}$. Hence, k must be in $\mathbf{S}_{(l,0)}$ or $\mathbf{S}_{(l,\downarrow)}$. It is apparent that $k \ne 1$, since the first node always increases the rate by 1 when the total sum-rate increases by 1. For $k \in \mathbf{S}_{(l+1,\uparrow)} \setminus \{1\}$, there must always exist a corresponding node $m \in \mathbf{S}_{(l+1,\downarrow)}$ such that $w_k < w_m$.

(i) If $k \in \mathbf{S}_{(l,0)}$, we know that $r_{(l+1,k)} = r_{(l,k)}$. Coding scheme with rate vector $\hat{\mathbf{r}}(l) = [\hat{r}_{(l,1)}, \ldots, \hat{r}_{(l,N)}]^{\mathsf{T}}$ such that

$$\hat{r}_{(l,k)} = r_{(l+1,k)} = r_{(l,k)} + 1 \quad (23)$$
$$\hat{r}_{(l,m)} = r_{(l+1,m)} = r_{(l,m)} - 1 \quad (24)$$
$$\hat{r}_{(l,i)} = r_{(l,i)}, \forall i \in [N] \setminus \{k, m\} \quad (25)$$

can also achieve universal recovery. Moreover, coding

scheme with rate vector $\hat{\mathbf{r}}(l)$ has lower cost than coding scheme with rate vector $\mathbf{r}(l)$. This contradicts that coding scheme with rate vector $\mathbf{r}(l)$ is optimal for all rate vector with sum rate l.

(ii) If $k \in \mathbf{S}_{(l,\downarrow)}$, we know that $r_{(l+1,k)} < r_{(l,k)}$. According to Lemma 3, $r_{(l+1,k)} \leq r_{(l,k)}$. This contradicts our assumption that $k \in \mathbf{S}_{(l+1,\uparrow)}$.

Thus we have $\mathbf{S}_{(l+1,\uparrow)} \subseteq \mathbf{S}_{(l,\uparrow)}$.

(2) Let $w_m = \max_{i \in \mathbf{S}_{(l+1,\downarrow)}} w_i$ and $w_n = \max_{j \in \mathbf{S}_{(l,\downarrow)}} w_j$. Assuming the contraction that $w_m > w_n$, then $m \notin \mathbf{S}_{(l,\downarrow)}$. Since $\mathbf{S}_{(l+1,\uparrow)} \subseteq \mathbf{S}_{(l,\uparrow)}$, coding scheme with rate vector $\hat{\mathbf{r}}(l) = [\hat{r}_{(l,1)}, \ldots, \hat{r}_{(l,N)}]^{\mathsf{T}}$ which satisfies

$$\hat{r}_{(l,m)} = r_{(l,m)} - 1 \tag{26}$$
$$\hat{r}_{(l,n)} = r_{(l,n)} + 1 \tag{27}$$
$$\hat{r}_{(l,i)} = r_{(l,i)}, \forall i \in [N] \setminus \{m, n\} \tag{28}$$

can also achieve universal recovery with the same sum-rate and has lower weighted cost. This contradicts that coding scheme with rate vector $\mathbf{r}(l) = [r_{(l,1)}, \ldots, r_{(l,N)}]^{\mathsf{T}}$ is optimal for all rate vector with sum rate l. Thus, we have $\max_{i \in \mathbf{S}_{(l+1,\downarrow)}} w_i \leq \max_{j \in \mathbf{S}_{(l,\downarrow)}} w_j$. ∎

Now we are ready to prove Theorem 2.

Proof: [**Theorem 2**] For any $R_{min} \leq l \leq K - 2$, we show that the second order difference of $\mathcal{K}(l)$ is non-negative, i.e. $\mathcal{F}(l+1) - \mathcal{F}(l) \geq 0$, where $\mathcal{F}(l) = \mathcal{K}(l+1) - \mathcal{K}(l)$. We compute the difference of the weighted cost of two coding schemes when sum-rate increases by 1.

$$\mathcal{F}(l+1) = \mathcal{K}(l+2) - \mathcal{K}(l+1) \tag{29}$$
$$= \sum_{i \in \mathcal{S}_{(l+1,\uparrow)}} w_i - \sum_{i \in \mathcal{S}_{(l+1,\downarrow)}} w_i \tag{30}$$
$$= w_1 + \sum_{i \in \mathcal{S}_{(l+1,\uparrow)} \setminus \{1\}} w_i - \sum_{i \in \mathcal{S}_{(l+1,\downarrow)}} w_i \tag{31}$$

According to Lemma 3, node 1 always generates 1 more transmission when the total number of transmissions increases by 1. And for other nodes, if their rate increases, the increment is 1, whereas if their rate decreases, the decrement can be more than 1. And the number of multiplications of the nodes in $\mathcal{S}_{(l+1,\downarrow)}$ is equal to the decrease in rate. Similarly, for sum-rate change from l to $l+1$, we have

$$\mathcal{F}(l) = \mathcal{K}(l+1) - \mathcal{K}(l) = w_1 + \sum_{i \in \mathcal{S}_{(l,\uparrow)} \setminus \{1\}} w_i - \sum_{i \in \mathcal{S}_{(l,\downarrow)}} w_i \tag{32}$$

The reason why node 1 is separated from other nodes is that the total number of transmissions only increases by 1, which implies that the total number of transmissions sent by other nodes, except node 1, remains the same. Hence

$$|\mathcal{S}_{(l,\uparrow)} \setminus \{1\}| = |\mathcal{S}_{(l,\downarrow)}| \tag{33}$$
$$|\mathcal{S}_{(l+1,\uparrow)} \setminus \{1\}| = |\mathcal{S}_{(l+1,\downarrow)}| \tag{34}$$

Therefore, $\forall i \in \mathcal{S}_{(l,\uparrow)} \setminus \{1\}$, $\exists j \in \mathcal{S}_{(l,\downarrow)}$ such that $w_i < w_j$. We can construct a partition of node pairs (i, j), where $i \in$ $\mathcal{S}_{(l,\uparrow)} \setminus \{1\}$ and $j \in \mathcal{S}_{(l,\downarrow)}$ as follows

$$\mathcal{P}(l) = \{(i, j) : i \in \mathcal{S}_{(l,\uparrow)} \setminus \{1\}, j \in \mathcal{S}_{(l,\downarrow)}, i < j\} \tag{35}$$

Note that the number of node pairs in $\mathcal{P}(l)$ is equal to $|\mathcal{S}_{(l,\uparrow)} \setminus \{1\}|$. Then we have

$$\mathcal{F}(l) = w_1 + \sum_{(i,j) \in \mathcal{P}(l)} (w_i - w_j) \tag{36}$$

where every term of the summation $(w_i - w_j)$ is negative.

We show that for each pair $(i, j) \in \mathcal{P}(l+1)$, there always exists a pair $(\hat{i}, \hat{j}) \in \mathcal{P}(l)$ such that

$$w_i - w_j - (w_{\hat{i}} - w_{\hat{j}}) \geq 0 \tag{37}$$

Assuming that there exists a node pair $(i, j) \in \mathcal{P}(l+1)$ such that for all possible pairs $(\hat{i}, \hat{j}) \in \mathcal{P}(l)$:

$$w_i - w_j - (w_{\hat{i}} - w_{\hat{j}}) < 0 \tag{38}$$

Equivalently, we have

$$w_i - w_j < \max_{\hat{i} \in \mathcal{S}_{(l,\uparrow)}, \hat{j} \in \mathcal{S}_{(l,\downarrow)}} (w_{\hat{i}} - w_{\hat{j}}) \tag{39}$$

If $i \in \mathcal{S}_{(l,\uparrow)}$, then $w_j > \max_{\hat{j} \in \mathcal{S}_{(l,\downarrow)}} w_{\hat{j}}$, which contradicts Lemma 4. If $i \notin \mathcal{S}_{(l,\uparrow)}$, consider another coding scheme with rate vector $\mathbf{r} = [r_1, r_2, \ldots, r_N]^{\mathsf{T}}$ such that

$$r_i = r_{(l+1,i)} + 1, r_j = r_{(l+1,j)} - 1 \tag{40}$$
$$r_{\hat{i}} = r_{(l+1,\hat{i})} - 1, r_{\hat{j}} = r_{(l+1,\hat{j})} + 1 \tag{41}$$
$$r_m = r_{(l+1,m)}, \forall m \notin \{i, j, \hat{i}, \hat{j}\} \tag{42}$$

It can be verified that this coding scheme can also achieve universal recovery with total $l+1$ transmissions. It has lower weighted cost than the coding scheme with rate vector $[r_{(l+1,1)}, \ldots, r_{(l+1,N)}]^{\mathsf{T}}$, which contradicts that coding scheme with rate vector $[r_{(l+1,1)}, \ldots, r_{(l+1,N)}]^{\mathsf{T}}$ has the minimum weighted cost over all coding schemes that achieve universal recovery with $l+1$ transmissions. Therefore, we can always find $(\hat{i}, \hat{j}) \in \mathcal{P}(l)$ such that Eqn (37) is satisfied for every pair $(i, j) \in \mathcal{P}(l+1)$. Hence, we have

$$\mathcal{F}(l+1) - \mathcal{F}(l)$$
$$= \sum_{(i,j) \in \mathcal{P}(l+1)} (w_i - w_j) - \sum_{(m,n) \in \mathcal{P}(l)} (w_m - w_n) \tag{43}$$
$$= \sum_{(i,j) \in \mathcal{P}(l+1), (\hat{i},\hat{j}) \in \mathcal{P}(l)} [(w_i - w_j) - (w_{\hat{i}} - w_{\hat{j}})]$$
$$- \sum_{(m,n) \in \mathcal{P}(l)\{(\hat{i},\hat{j})\}} (w_m - w_n) \tag{44}$$
$$\geq 0 \tag{45}$$

where every $(w_i - w_j) - (w_{\hat{i}} - w_{\hat{j}}) \geq 0$ and every $w_m - w_n < 0$. Hence, the function $\mathcal{K}(l) = \min_{\mathbf{r} \in \Omega, \mathcal{S}(\mathbf{r}) = l} \sum_{i=1}^{N} w_i r_i$ is convex. ∎

ACKNOWLEDGMENT

REFERENCES

[1] S. Li and M. Gastpar, "Cooperative data exchange based on MDS codes," in *2017 IEEE International Symposium on Information Theory (ISIT) (ISIT'2017)*, Aachen, Germany, Jun. 2017, pp. 1411–1415.

[2] S. El Rouayheb, A. Sprintson, and P. Sadeghi, "On coding for cooperative data exchange," in *Information Theory (ITW 2010, Cairo), 2010 IEEE Information Theory Workshop on*. IEEE, 2010, pp. 1–5.

[3] T. A. Courtade, B. Xie, and R. D. Wesel, "Optimal exchange of packets for universal recovery in broadcast networks," in *Military Communications Conference, 2010-Milcom 2010*. IEEE, 2010, pp. 2250–2255.

[4] N. Ding, C. Chan, Q. Zhou, R. A. Kennedy, and P. Sadeghi, "Communication for Omniscience," *ArXiv e-prints*, Nov. 2016.

[5] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb, "A randomized algorithm and performance bounds for coded cooperative data exchange," in *2010 IEEE International Symposium on Information Theory*. IEEE, 2010, pp. 1888–1892.

[6] N. Milosavljevic, S. Pawar, S. El Rouayheb, M. Gastpar, and K. Ramchandran, "Efficient algorithms for the data exchange problem," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1878–1896, 2016.

[7] T. A. Courtade and R. D. Wesel, "Coded cooperative data exchange in multihop networks," *Information Theory, IEEE Transactions on*, vol. 60, no. 2, pp. 1136–1158, 2014.

[8] M. Gonen and M. Langberg, "Coded cooperative data exchange problem for general topologies," *Information Theory, IEEE Transactions on*, vol. 61, no. 10, pp. 5656–5669, 2015.

[9] A. Heidarzadeh, M. Yan, and A. Sprintson, "Cooperative data exchange with priority classes," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2324–2328.

[10] A. Heidarzadeh and A. Sprintson, "Successive local and successive global omniscience," *arXiv preprint arXiv:1702.01773*, 2017.

[11] D. Ozgul and A. Sprintson, "An algorithm for cooperative data exchange with cost criterion," in *2011 Information Theory and Applications Workshop*, Feb 2011, pp. 1–4.

[12] S. E. Tajbakhsh, P. Sadeghi, and R. Shams, "A generalized model for cost and fairness analysis in coded cooperative data exchange," in *Network Coding (NetCod), 2011 International Symposium on*. IEEE, 2011, pp. 1–6.

[13] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1973–1982, 2005.

[14] M. Yan and A. Sprintson, "Algorithms for weakly secure data exchange," in *2013 International Symposium on Network Coding (NetCod)*. IEEE, 2013, pp. 1–6.