

DIRECT INFORMATION EXCHANGE IN WIRELESS NETWORKS:  
A CODING PERSPECTIVE

A Thesis

by

DAMLA OZGUL

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2010

Major Subject: Electrical Engineering

DIRECT INFORMATION EXCHANGE IN WIRELESS NETWORKS:  
A CODING PERSPECTIVE

A Thesis

by

DAMLA OZGUL

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of  
MASTER OF SCIENCE

Approved by:

Chair of Committee,	Alex Sprintson
Committee Members,	Henry D. Pfister
	Tie Liu
	Donald K. Friesen
Head of Department,	Costas N. Georghiades

August 2010

Major Subject: Electrical Engineering

## ABSTRACT

Direct Information Exchange In Wireless Networks:

A Coding Perspective. (August 2010)

Damla Ozgul, B.S., METU, Ankara, Turkey

Chair of Advisory Committee: Dr. Alex Sprintson

The rise in the popularity of smartphones such as Blackberry and iPhone creates a strain on the world's mobile networks. The extensive use of these mobile devices leads to increasing congestion and higher rate of node failures. This increasing demand of mobile wireless clients forces network providers to upgrade their wireless networks with more efficient and more reliable services to meet the demands of their customers. Therefore, there is a growing interest in strategies to resolve the problem and reduce the stress on the wireless networks.

One strategy to reduce the strain on the wireless networks is to utilize cooperative communication. The purpose of this thesis is to provide more efficient and reliable solutions for direct information exchange problems. First, algorithms are presented to increase the efficiency of cooperative communication in a network where the clients can communicate with each other through a broadcast channel. These algorithms are designed to minimize the total transmission cost so that the communication will be less expensive and more efficient. Second, we consider a setting in which several clients exchange data through a relay. Our algorithms have provable performance guarantees. We also verify the performance of the algorithms in practical settings through extensive simulations.

To my family

## ACKNOWLEDGMENTS

Among all people, I would like to express my deepest gratitude and appreciation to my advisor, Dr. Alex Sprintson. His contribution to my education and professional development is invaluable. I am so grateful, and I consider myself lucky to have had such an advisor.

I wish to express my appreciation and thanks to my committee members, Dr. Pfister, Dr. Liu and Dr. Friesen, for their kindness and supportiveness.

Finally, I would like to thank my family, especially my sister and my friends, for supporting and encouraging me to pursue this degree.

## TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION . . . . .	1
II	BASIC MODEL . . . . .	5
	A. Fundamental Model . . . . .	5
	B. Direct Information Exchange Problems . . . . .	6
	C. Index Coding . . . . .	9
III	DATA EXCHANGE WITH COSTS (DEWC) PROBLEM . . . . .	13
	A. Background . . . . .	13
	B. Model . . . . .	16
	C. Randomized Algorithm . . . . .	17
	1. Algorithm . . . . .	17
	2. Analysis . . . . .	19
	3. Numerical Results . . . . .	22
	D. Deterministic Algorithm . . . . .	25
	1. Algorithm . . . . .	25
	2. Analysis . . . . .	27
	3. Numerical Results . . . . .	29
	E. Randomized Algorithm with Restrictions on Number of Transmissions . . . . .	30
	1. Algorithm . . . . .	30
	2. Analysis . . . . .	35
	3. Numerical Results . . . . .	37
IV	DIRECT INFORMATION EXCHANGE THROUGH RE- LAY PROBLEM . . . . .	39
	A. Data Exchange Through Relay (DETR) Problem . . . . .	39
	1. Background . . . . .	39
	2. Model . . . . .	41
	3. Main Results . . . . .	43
V	CONCLUSIONS . . . . .	47
	A. Summary of the Thesis . . . . .	47

CHAPTER	Page
B. Future Works . . . . .	48
REFERENCES . . . . .	49
VITA . . . . .	53

## LIST OF FIGURES

FIGURE		Page
1	Wireless Network Coding. . . . .	2
2	Index Coding Problem . . . . .	6
3	Cooperative Data Exchange Problem . . . . .	7
4	Data Exchange with Costs Problem . . . . .	8
5	Data Exchange through Relay Problem . . . . .	9
6	General Properties of Direct Wireless Exchange Problems . . . . .	10
7	The Index Coding Problem. . . . .	11
8	Possible Transmission Schemes. . . . .	14
9	Optimal Transmission Scheme. . . . .	15
10	Performance Results for Randomized Algorithm with $k = n$ . . . . .	23
11	Performance Results for Randomized Algorithm with $k = 5$ . . . . .	24
12	Performance Comparison for DEWC Problem. . . . .	30
13	DEWC Problem without any Restrictions. . . . .	31
14	DEWC Problem with Restrictions on the Number of Transmissions. . . . .	32
15	Performance Comparison for DEWC with Restrictions. . . . .	37
16	The CDE Problem. . . . .	40
17	Data Exchange Through Relay (DETR) Problem. . . . .	41
18	Uplink (a) and Downlink (b) Transmissions for Problem DETR . . . . .	42



FIGURE		Page
19	The Graph Representation of Problem CDE. . . . .	43
20	The Graph Representation of Problem DETR. . . . .	45

## CHAPTER I

## INTRODUCTION

Wireless technology has become an increasingly popular way to gain network access. Wireless networks are expected to provide an efficient and reliable service in many applications, such as multimedia streaming and video conferencing. The novel technique of network coding has a significant potential to improve the throughput, reliability and efficiency of wireless networks by taking advantage of the broadcast nature of the wireless medium. Network coding is a generalization of network operation beyond traditional routing or store-and-forward approaches. In the traditional approach, coding is employed at source nodes to compress redundant information or to protect the network against losses. It can also be employed at the link level to protect against random errors or erasures on individual links. On the other hand, the usual task of networks is to transport information supplied by source nodes without any modification. Network coding, in contrast, allows interior network nodes to combine or mix information from different sources.

The first study highlighting the utility of network coding was performed by Ahlswede et al.[1] where the term *network coding* was coined. In this work, the advantage of network coding over routing, the traditional way of operating a network, was pointed out for the first time by means of a very simple example known as the butterfly network. Since their work, the topic of network coding has been undergoing an active development in the research community [2, 3, 4, 5].

The technique of *network coding* has a significant potential for improving the performance of wireless networks by exploiting their broadcast nature. For example,

---

The journal model is *IEEE Transactions on Automatic Control*.

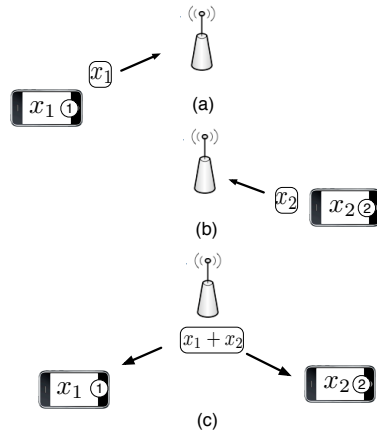


Fig. 1. Wireless Network Coding.

consider the wireless network coding scheme depicted in Figure 1. In this example, two wireless clients need to exchange packets  $x_1$  and  $x_2$  through a relay node (clients cannot communicate directly, e.g., due to power constraints). The network coding approach requires three transmissions: two from the clients to the relay node as shown in Figure 1(a) and Figure 1(b) and one from the relay node to the clients as shown in Figure 1(c). In contrast, the traditional approach would require four transmissions.

Wireless ad-hoc networks are one of the main areas where the network coding technique is expected to be most beneficial. In particular, the network coding technique allows to exploit the broadcast nature of the wireless medium. The underlying principle of wireless network coding architectures is *opportunistic listening* [6, 7]. With this approach each network node is snooping on all communications over the wireless medium. The overheard packets are stored for a limited period of time. The key idea is to take advantage of the overheard packets, also referred to as *side information*, to achieve a higher rate of the information exchange. For instance, *Index Codes* introduced in [8] utilize from prior side

information to satisfy more efficient and reliable information exchange.

Recently, increasing demand of mobile wireless clients has resulted in a challenging problem of achieving efficient and reliable information exchange on mobile networks. There is a considerable stress on cellular networks in terms of bandwidth provision and network cost. This creates a growing interest in cooperative wireless communication [9] which is a promising strategy to resolve the problem providing energy efficiency [10], increased coverage [11] and enhanced data rates [12].

In this thesis, we have considered several direct information exchange problems that are mostly utilizing the side information to satisfy efficient information exchange via cooperative wireless communication. In each scenario, the main goal is to improve efficiency and reliability of the transmissions to exchange information. Indeed, these problems can be classified in two major groups according to the decision whether clients are allowed to communicate with each other or not.

One scenario could be the case when clients are allowed to directly communicate with each other. In this problem, some wireless clients are interested in the same large file (such as an audio file). Initially, a base station broadcasts the file to the mobile clients. The long-range link between the base station and the mobile clients can be expensive, slow and unreliable, which causes some clients to receive only some parts of the whole file. Indeed, partial reception results from channel fading or shadowing, connection loss or network saturation and congestion such as in peer-to-peer systems. Although the clients have received some portion of the file, if the whole file is collectively known by the mobile clients, they can cooperate with each other to obtain the whole file using direct communication which is more reliable and faster. However, each client could be associated with different transmission costs. The costs can capture different factors, such as the quality of the communication channel, residual battery charge, etc. For example, a client with a low residual charge

will have a higher transmission cost, whereas a client with a higher charge will have a lower cost. In fact, if a client with low battery is requested to have a large number of transmissions, it will find itself with a empty battery. Since our algorithm minimizes the total transmission cost, it would give a preference for clients with higher residual charge in the transmission order.

On the other hand, there could be some networks where the clients cannot directly communicate with each other. Suppose again that the clients collectively hold a large file and each client would like to obtain the whole file. The clients exchange data using a relay node to obtain the missing parts of the large file. Since the clients collectively have the file, then it is possible that all clients will obtain the whole file.

This thesis is organized as follows. In Chapter II, we describe the basic algebraic model and introduce different scenarios that are mainly taking the advantage of side information of the clients, i.e., the opportunistic listening technique. In addition, Index Codes are explained in detail as a leading study utilizing opportunistic listening in networks. In Chapter III, we examine Data Exchange with Costs (DEWC) problem in detail and establish randomized and deterministic algorithms to satisfy efficient information exchange. We also examine the performance of the randomized algorithm when there is a restriction on the number of transmissions. Furthermore, simulation results are presented to corroborate the analytical results. In Chapter IV, we study Data Exchange through Relay (DETR) problem, in which the clients cannot communicate with each other, in detail and state our results. Finally, we present our conclusions and some directions for further work in Chapter V.

## CHAPTER II

## BASIC MODEL

## A. Fundamental Model

A general instance of the basic model of the direct information exchange problems includes a set of information messages (data packets)  $P = \{p_1, p_2, \dots, p_m\}$ , a set of receivers (clients)  $C = \{c_1, \dots, c_n\}$ . Each one of the clients,  $c_i$ , has a side information set,  $H(c_i) \subseteq P$ , represented by a subset of  $P$ , and a demand set,  $W(c_i) \subseteq P$ , which is another subset of  $P$ . After the completion of the transmissions, the clients will obtain the packets in their demand set. In addition, the demand set could be the set which is the complement of side information set, i.e.,  $W(c_i) = \overline{H}(c_i)$  and  $W(c_i) \cup H(c_i) = P$  as well as it could be some other set whose size is smaller than the complement of side information set, i.e., the client could be interested in some specific packets in  $P$ .

We have considered two settings in this thesis. In the first one, there is no server/base station; that is, the clients can communicate among themselves to complete the communication and ensure that all the clients eventually obtain their demand packets. On the other hand, in the second setting there is a server through which the communication between the clients takes place. The server and the clients can broadcast encodings of messages in  $P$  over a noiseless broadcast channel. In other words, we assume that transmissions of the packets between the clients and/or between a client and the server occur without an error. The objective of all scenarios is to ensure that all of the clients eventually possess the packets in their demand sets in an efficient and inexpensive way.

The generic model is the underlying part of all the scenarios studied in this thesis. Although the basic properties are the same for all the problems, they differ in

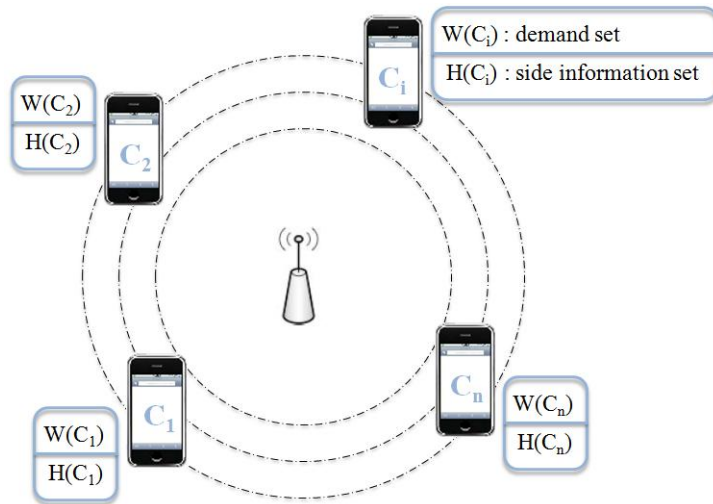


Fig. 2. Index Coding Problem

some aspects. Therefore, each scenario is studied separately. In the remainder of this chapter, each problem is introduced and explained briefly. Then, Index Coding problem is examined in detail since it considers a basic setting that utilize opportunistic listening and all other scenarios studied in this thesis are built upon this problem.

### B. Direct Information Exchange Problems

*Index Codes* utilize prior side information [8]. An instance of the Index Coding problem includes a set of wireless clients and a set of packets that need to be delivered to clients, and a server node, base station, that holds these information packets. Each client has a *demand* set which is the set of packets required by the client and a *side information* set which is the set of packets available at the client, as shown in Figure 2. In each round of communication, base station can transmit a single packet or a combination of the packets. We assume that all packets transmitted by the server are received by all clients without an error. The goal is to allow each client to decode the packets it requested while minimizing the number of transmissions.

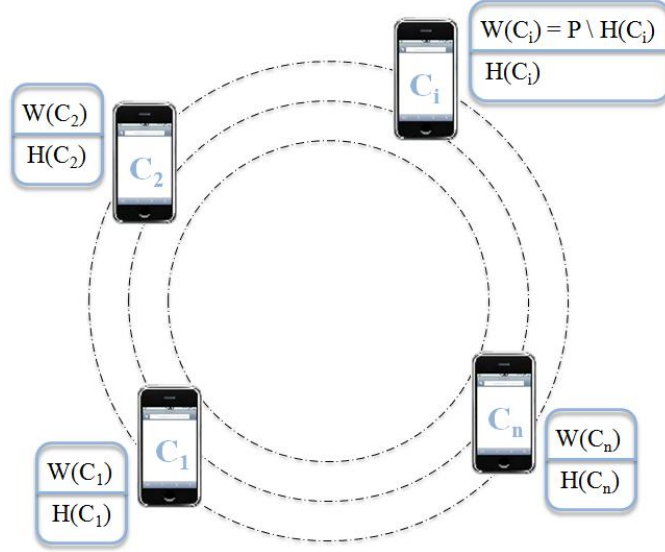


Fig. 3. Cooperative Data Exchange Problem

*Cooperative Data Exchange (CDE)* Problem is another scenario utilizing the overheard packets which was introduced by El Rouayheb et al.[13]. An instance of the CDE Problem includes a set of wireless clients and a set of packets that need to be delivered to all clients. This problem differs from the Index Coding problem in the following way; there is no server in this scenario, rather clients try to communicate with each other. In other words, they exchange the packets in a cooperative manner as shown in Figure 3. In addition, unlike the Index Coding problem, in this scenario all clients require all the packets; not some specific packets. In each round of communication, a client can transmit a single packet or a combination of the packets. Assuming that the transmitted packets are received by all clients without an error, the aim is to guarantee that all the clients eventually obtain all the packets with minimum the number of transmissions.

*Data Exchange with Costs (DEWC)* Problem is similar to the CDE Problem except that in this problem each client is associated with a cost value as shown in Figure 4. An instance of the DEWC Problem includes a set of wireless clients and



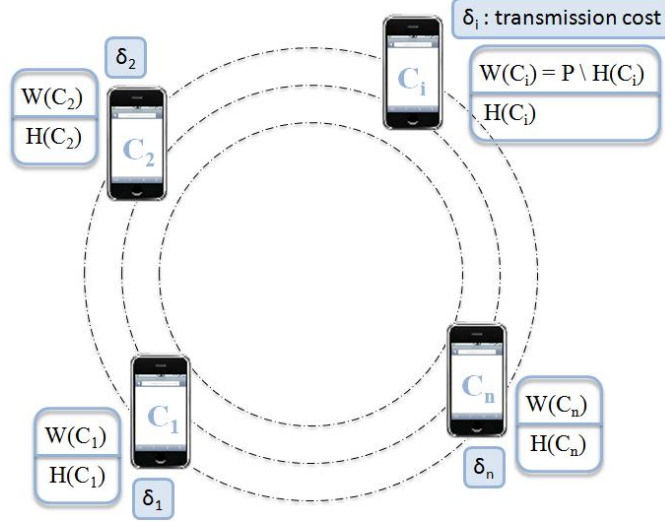


Fig. 4. Data Exchange with Costs Problem

a set of packets that need to be delivered to all clients as in the case of CDE. The major difference between these problems is that in the CDE Problem the goal is to guarantee that at the end all clients will get all the packets with the minimum number of transmissions whereas in the DEWC Problem the goal is to minimize the total transmission cost.

A closely-related problem is that of *Data Exchange Through Relay (DETR)* Problem. An instance of this problem includes a set of wireless clients and a set of packets that need to be delivered to clients, and a server node as shown Figure 5. At the end, all clients require all the packets as in the CDE Problem. In each round of communication, base station or a client can transmit a single packet or a combination of the packets. Again, we assume that the transmitted packets are received without an error. The goal is again to ensure that all the clients eventually get all the packets with minimum the number of uplink and downlink transmissions through the server node.

Figure 6 shows general properties of all the problems considered in this thesis.

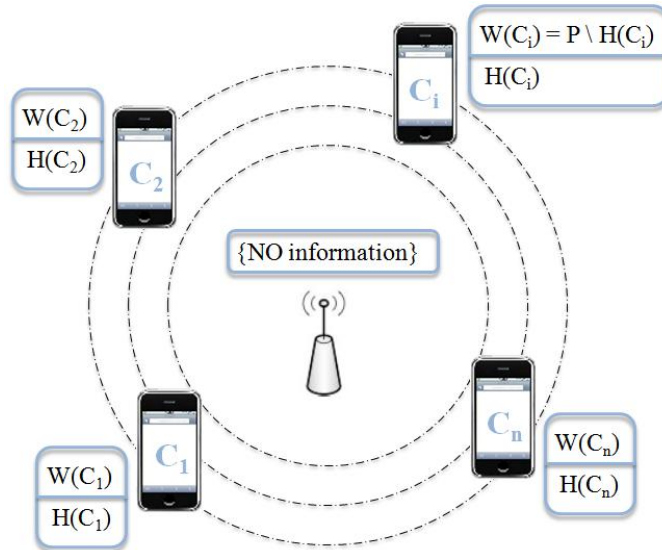


Fig. 5. Data Exchange through Relay Problem

Although these direct exchange problems are similar, they differ in some properties as shown in the Figure 6.

### C. Index Coding

Index Coding problem has been recently introduced in [8] and it is one of the fundamental problems in wireless network coding. Index Coding problem has been studied in several recent works [8, 14, 15, 16, 17, 18, 19, 20, 21] and it has several applications in wireless networking and distributed computing. In addition, these codes are instrumental in satellite communication networks in which the clients have limited storage and save only some part of the received information [18]. An instance of the Index Coding problem includes a server/base station that holds a set of information packets  $P$  and a set of receivers (clients)  $C$ , each one of them has some side information represented by a subset of  $P$ , known to the server, and demands another subset of  $P$ . The server can broadcast encodings of messages in  $P$  over a noiseless channel. The objective is to identify an encoding scheme that satisfies the demands

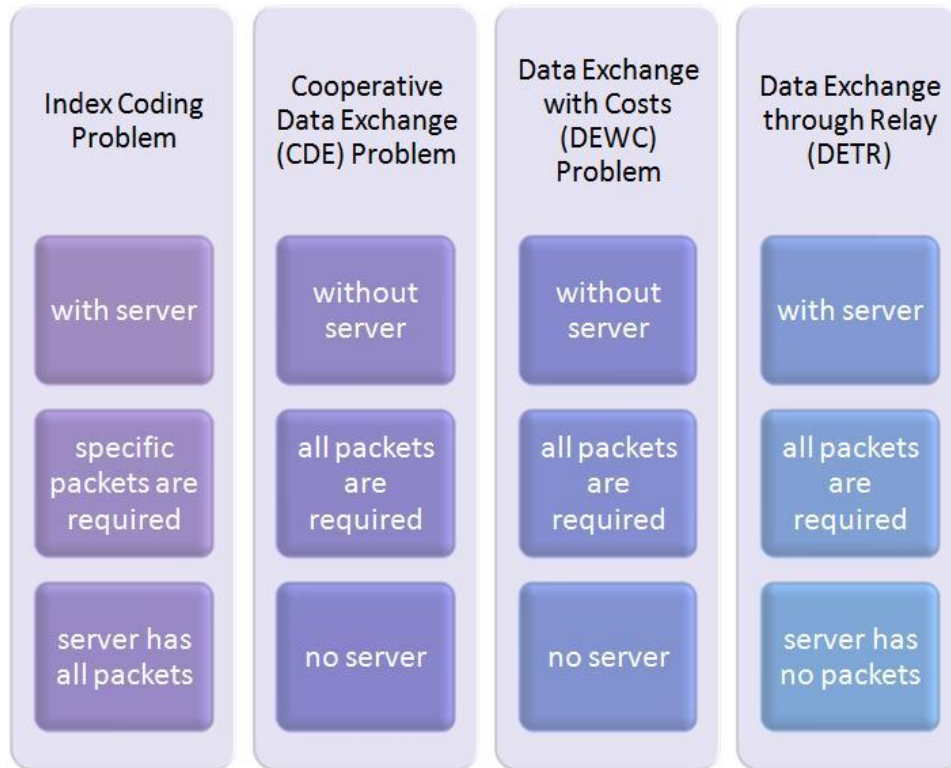


Fig. 6. General Properties of Direct Wireless Exchange Problems

of all clients with the minimum number of transmissions.

Figure 7 demonstrates an instance of the Index Coding problem in which the overheard packets can be used to minimize the number of transmissions. In this example, the sender node needs to deliver four packets  $p_1, \dots, p_4$  to its neighbors. The figure shows, for each neighbor, the required packets, as well as the packets available through opportunistic listening. The standard technique would require four packets to be transmitted by the central node, while the network coding approach requires only two transmissions  $p_1 + p_2 + p_3$  and  $p_1 + p_4$ , all operations are performed over  $GF(2^n)$ . In other words, the encoding technique decreases the total number of transmissions by 50%, resulting in a coding gain of 2. The problem of minimizing the

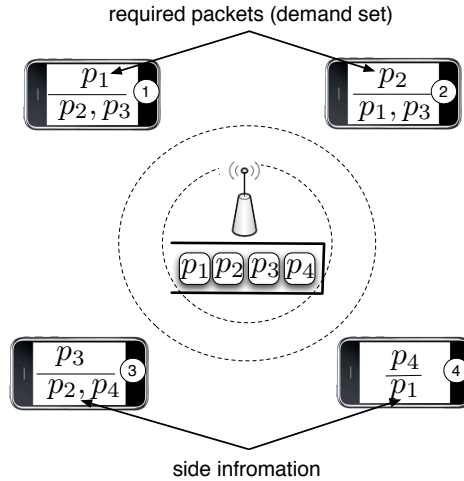


Fig. 7. The Index Coding Problem.

number of required transmissions is referred to as the *Index Coding* problem.

More formally, an instance of the Index Coding includes a relay node  $r$ , a set  $C = \{c_1, \dots, c_n\}$  of wireless clients and a set  $P = \{p_1, p_2, \dots, p_m\}$  of packets that need to be delivered to clients in  $C$ . Each client  $c_i \in C$  is associated with two sets:

- *demand* set  $W(c_i) \subseteq P$  - the set of packets required by  $c_i$
- *side information* set  $H(c_i) \subseteq P$  - the set of packets available at  $c_i$

In each *round* of communication the relay can transmit a single packet. We assume that all packets transmitted by the relay are received by all clients without an error. The  $j$ 'th round of communication is specified by an encoding function  $g_j : \Sigma^m \rightarrow \Sigma$ . The objective is to find the set of encoding functions  $\Phi = \{g_i\}_{i=1}^\ell$  that will allow each client to decode the packets it requested while minimizing the number of transmissions  $\ell = |\Phi|$ . Client  $c_i \in C$  can decode packets in  $W(c_i)$  if there exists a decoding function  $\gamma_i : \Sigma^\ell \times (\Sigma)^{|H(c_i)|} \rightarrow (\Sigma)^{|W(c_i)|}$  that allows  $c_i$  to obtain all packets

in  $W(c_i)$ .

The study of Index Codes by Bar-Yossef, Birk, Jayram and Kol, [8], has attracted a significant amount of attention in the research community. The problem also motivated us to investigate various problems related to Index Coding problem and to design effective algorithms satisfying efficient information exchange to solve the problems.

## CHAPTER III

## DATA EXCHANGE WITH COSTS (DEWC) PROBLEM

## A. Background

In this chapter, we consider the problem of cooperative data exchange between clients each having an associated transmission cost value. As a motivation of the problem, consider some wireless clients interested in the same large file. Initially, a base station broadcasts the file to the mobile clients. The long-range link between the base station and the mobile clients can be expensive, slow and unreliable, which causes some clients to receive only some parts of the whole file. Indeed, partial reception results from channel fading or shadowing, connection loss or network congestion such as in peer-to-peer systems. Although the clients have received some portion of the file, if the whole file is collectively known by the mobile clients, they can communicate among each other to obtain the whole file using direct communication which is more reliable and faster. However, each client could be associated with different transmission costs. The costs can capture different factors, such as residual battery charge. For example, a client with a low residual charge will have a higher transmission cost, whereas a client with a higher charge will have a lower cost. In fact, if a client with low battery is requested to have a large number of transmissions, it will find itself with a empty battery. Since our algorithm minimizes the total transmission cost, it would give a preference for clients with higher residual charge in the transmission order.

To illustrate the problem further, consider four wireless clients that are interested in obtaining three packets of  $m$  bits each,  $p_1, p_2$  and  $p_3 \in GF(2^m)$ . The second, third and fourth clients have received packets  $\{p_1, p_3\}$ ,  $\{p_2, p_3\}$  and  $\{p_1, p_2\}$ , respectively,

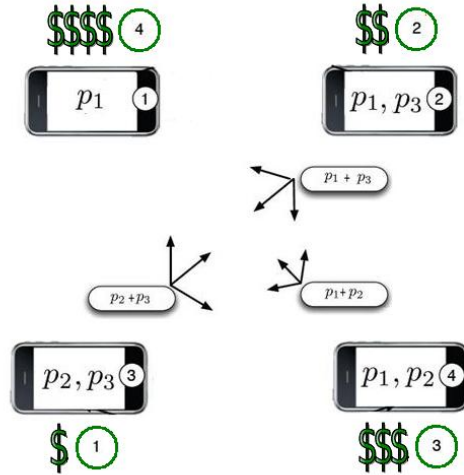


Fig. 8. Possible Transmission Schemes.

i.e., each of these clients misses one packet. However, the first client has only  $\{p_1\}$  and misses two packets due to channel imperfections. Each client is associated with a cost value; for example the transmission cost when the first client,  $c_1$  sends its packet  $\{p_1\}$  is 4. The transmission costs associated with  $c_2$ ,  $c_3$  and  $c_4$  are 2, 1, and 3, respectively. Assume that each client can broadcast the packets via a noiseless broadcast channel and is fully aware of the packets available to other nodes. Now, the clients can try to complete the communication among themselves. Since they have collectively known all the packets, it is possible to choose a transmission scheme which ensures that all the clients eventually obtain all the packets with a total minimum cost value.

A simple cooperation scheme would consist of three uncoded transmissions with a minimum cost of four; the third client first sends  $p_2$  and then  $p_3$  separately costing  $1 + 1 = 2$  and the second client sends  $p_1$  letting the total cost as  $2 + 2 = 4$ . However, this is not an optimal solution not only for minimum number of transmissions but also for the total minimum transmission cost. Since the clients can send coded packets

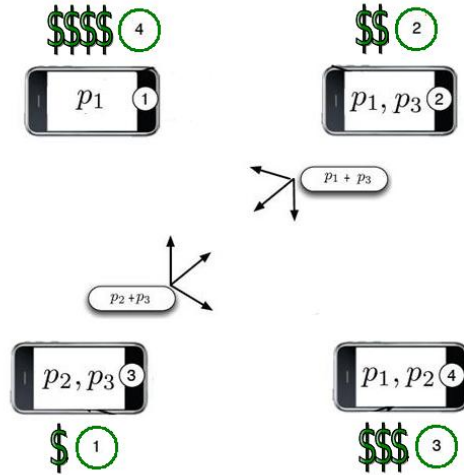


Fig. 9. Optimal Transmission Scheme.

and help multiple clients with a single transmission, the number of transmissions for this problem can be decreased to two letting only two clients among second, third and fourth clients send packets as shown in Figure 8. Indeed, an optimal solution can be obtained if  $(c_2, c_3)$  or  $(c_2, c_4)$  or  $(c_3, c_4)$  are the clients that transmit packets. However, since we are interested in the total minimum cost, we need to choose  $(c_2, c_3)$  as transmitters as in Figure 9. Because, if the third client sends  $p_2 + p_3$  while the second client sends  $p_1$ , then the total cost will be  $1 + 2 = 3$  which is the minimum cost that could be obtained. It can be verified that all four mobile clients can then decode all the packets.

A closely-related problem is that of Cooperative Data Exchange problem which was introduced by El Rouayheb et al.[13] in which the main goal is to minimize the number of transmissions rather than to minimize the total cost.

In the following, we present efficient algorithms to find optimum transmission schemes for cooperative data exchange with costs and verify the effectiveness of our



algorithms through simulation studies. We first define the problem and present the model. Then, we present and analyze the algorithms. We also show the performance results of our algorithms.

## B. Model

Consider a set  $C = \{c_1, \dots, c_k\}$  of  $k$  wireless clients. Each client  $c_i$  is associated with a transmission cost  $\delta(c_i)$ . A set  $P = \{p_1, \dots, p_n\}$  of  $n$  packets needs to be delivered to  $k$  clients in  $C$ . The packets are elements of a finite field,  $F$ . Initially, each client  $c_i \in C$  is associated with *side information* set  $H(c_i) \subseteq P$  while the clients collectively know all packets in  $P$ , i.e.,  $\cup_{c_i \in C} H(c_i) = P$ . The *demand* set of the client  $c_i$ , which is the set of packets required, is denoted as  $W(c_i) = H(\bar{c}_i) = P \setminus H(c_i) \subseteq P$ . We assume that each client knows the indices of packets that are available to other clients.

To ensure all clients will obtain all packets in  $P$ , the clients exchange packets over a lossless broadcast channel. The information packets are transferred in communication rounds, such that at round  $j$  one of the clients,  $c_{i_j}$ , broadcasts a packet,  $p^j \in F$ , to other clients in  $C$  with a transmission cost of  $\delta(c_{i_j})$ . Packet  $p^j$  could be one of the packets in  $H(c_{i_j})$ , or it could be a combination of packets in  $H(c_{i_j})$  and the packets that are previously transmitted over the channel. This is only restricting  $c_{i_j}$  to use the information it knows at the start of round  $j$ . The aim is to find a scheme that satisfies each client  $c_i$  obtains all packets in  $W(c_i)$  while minimizing the total cost of transmissions. The scheme uses linear coding over the field  $F$ . In other words, each packet is an element of a finite field  $F$  and all coding operations are also linear over the field,  $F$ .

The number of packets initially possessed by the client,  $c_i$ , is denoted by  $n_i = |P_i|$ ,

while the number of unknown packets to the client,  $c_i$ , is therefore,  $\bar{n}_i = |H(\bar{c}_i)| = n - n_i$ . We denote by  $n_{min} = \min_{1 \leq i \leq k} n_i$  the minimum number of packets known to a client.

A client  $c_i$  is said to have a unique packet  $p_j$  if  $p_j \in H(c_i)$  but  $p_j \notin H(c_l)$  for all  $l \neq i$ . A client who owns a unique packet can broadcast the packet in an uncoded fashion at any stage without any penalty in terms of optimality. Therefore, without loss of generality, it can be assumed that there are no unique packets in the system.

### C. Randomized Algorithm

#### 1. Algorithm

In this section, we present an efficient randomized algorithm for the data exchange with costs problem. Our algorithm assumes a large finite field  $F$  of size  $q$  and provides an optimal solution with high probability. The algorithm uses linear coding in which any packet,  $p^j$ , transmitted by the algorithm is a linear combination of the original packets in  $P$ ;

$$p^j = \sum_{p_i \in P} \alpha_i^j p_i \quad (3.1)$$

where  $\alpha_i^j \in F$  are encoding coefficients and  $\alpha^j = \{\alpha_1^j, \dots, \alpha_n^j\}$  is the encoding vector of  $p^j$ . We denote  $u_i = \{u_i^1, \dots, u_i^n\}$  as the unit encoding vector corresponding to the original packet  $p_i$  where  $u_i^i = 1$  and  $u_i^j = 0$  for  $i \neq j$ . Moreover, we denote by  $U(c_i)$  the set of unit vectors that corresponds to packets in  $H(c_i)$ . In the analysis of the algorithm which runs in rounds, instead of expressing the original packets we use encoding vectors. In other words, rather than saying that a packet  $p^j = \sum_{p_i \in P} \alpha_i^j p_i$  has been transmitted by a client  $c_{i_j}$  at round  $j$ , we state that the client transmits the encoding vector  $\alpha^j = \{\alpha_1^j, \dots, \alpha_n^j\}$ . The transmitted vector  $\alpha^j$  is a random linear combination of the unit vectors in  $U(c_{i_j})$ , i.e.,  $\alpha_i^j = 0$  for  $p_i \notin H(c_{i_j})$ . Other

elements of  $\alpha^j$  are selected at random from the field  $F$ . Then, the set of encoding vectors that have been transmitted up to and including round  $j$  can be expressed as  $A_j = \{\alpha^1, \dots, \alpha^j\}$ .

Let  $T$  is the number of transmissions with which minimum total cost for the overall transmissions can be obtained, that is, it is the number of the transmissions for the optimal solution of the problem. Assume that we have made a good guess on the value of  $T$ . Indeed, it is not hard to find the exact value of  $T$ . Since  $n - n_{min} \leq T \leq n$ , i.e., the value of  $T$  is upper bounded by  $n$ , we can implement exhaustive search to find its value. So, if the randomized algorithm assuming that the value  $T$  is known is a low complexity algorithm, the overall algorithm with the exhaustive search will not be an algorithm with high complexity.

At each iteration  $j$  of the algorithm, we denote by  $n_i^j$  the number degrees of freedom available for client  $c_i$ . More specifically,  $n_i^j$  is defined as follows:

$$n_i^j = \text{rank}U(c_i) \cup A_{j-1}, \quad (3.2)$$

where  $A_{j-1} = \{\alpha^1, \dots, \alpha^{j-1}\}$  is the set containing the packets that have been transmitted so far. Note that  $n - n_i^j$  is a minimum number of packets that needs to be received by client  $c_i$  to satisfy its demands.

At the iteration  $j$  of the algorithm we divide the clients in  $C$  into two groups  $C_1^j$  and  $C_2^j$ .

- Set  $C_1^j = \{c_i \in C \mid n - n_i^j = T - (j - 1)\}$  contains clients that require  $T - (j - 1)$  packets at iteration  $j$ ,
- Set  $C_2^j = \{c_i \in C \mid n - n_i^j < T - (j - 1)\}$  contains clients that require less than  $T - (j - 1)$  packets at iteration  $j$ .

Since the clients in  $C_1^j$  need at least  $T - (j - 1)$  transmissions to decode the

required packets, they can not transmit at the current iteration. Therefore, at each round  $j$ , our algorithm selects a client with lowest cost in  $C_2^j$  as the transmitter, i.e,

$$c_{i_j} = \arg \min_{c_i \in C_2^j} \delta(c_i).$$

The steps performed by the algorithm can be summarized as follows:

#### RANDOMIZED ALGORITHM

- 1 **for**  $T \leftarrow n - n_{min}$  **to**  $n$
- 2     **for**  $j \leftarrow 1$  **to**  $T$
- 3         Determine sets  $C_1^i$  and  $C_2^j$  as defined above,
- 4         Select a client  $c_{i_j} \in C_2^j$  with minimum  
transmission cost,
- 5         Create an encoding vector  $\alpha^j$  by randomly  
combining unit vectors in  $U(c_{i_j})$ ,
- 6         Transmit the packet  $p^j = \sum_{p_i \in P} \alpha_i^j p_i$ .
- 7     Calculate the total transmission cost for chosen  $T$ ,  
i.e.,  $\Delta_T = \sum_{i=1}^T \delta(c_{i_j})$ .
- 8 **return** the total minimum cost among all  $T$  values,  
i.e.,  $\Delta = \arg \min_{T \in [n - n_{min}, n]} (\Delta_T)$ .

## 2. Analysis

We proceed to analyze the correctness and optimality of the algorithm. Consider an iteration  $j$  of the algorithm. We denote  $\text{OPT}_j$  as the minimum total transmission cost of completing the information transfer after round  $i$ , provided that at least  $T - j$  transmissions are allowed after round  $j$ . In other words, in addition to the first  $j$  transmissions, at most  $T - j$  transmissions of total cost  $\text{OPT}_j$  are needed to satisfy

the demands of all clients.

**Lemma 1** *With a probability at least  $1 - \frac{k}{q}$ ,  $OPT_j = OPT_{j-1} - \delta(c_{i_j})$ .*

**Proof:** We denote  $\Omega_{j-1}$  as an optimal set of encoding vectors which are necessary to complete data transfer. In other words,  $\Omega_{j-1}$  has  $T - (j - 1)$  encoding vectors such that:

- each vector is a linear combination of  $U(c_i)$  for some  $c_i \in C$
- for each client  $c_i \in C$  it holds that the set  $A_{j-1} \cup \Omega_{j-1} \cup U(c_i)$  is of rank  $n$  where  $A_{j-1} = \{\alpha^1, \dots, \alpha^{j-1}\}$  is the set containing the packets that have been transmitted so far.

Let  $\mu = \text{rank}(A_{j-1})$  be the rank of the set of packets transmitted so far and  $\mu^j = \text{rank}(U(c_{i_j}) \cup A_{j-1})$  be the rank of the set of encoding vectors which a client  $c_{i_j}$  own at the beginning of iteration  $j$ . Assume that  $c_{i_j}$  is the client which has a minimum cost among the set  $C_2^j$ . Observe that  $T - (j - 1)$  is at least  $T - \mu^j$  which is strictly larger than 0. This follows from the fact that  $c_{i_j}$  is in the set  $C_2^j$  meaning that it has a demand set strictly smaller than the solution set. So,  $T - (j - 1)$  could be at least 1. Therefore, we can remove at least one packet,  $v$ , from  $\Omega_{j-1}$  so that  $\tilde{\Omega}_{j-1} = \Omega_{j-1} \setminus \{v\}$  such that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_{i_j})$  still remains of rank  $n$ .

Assume that  $c_i$  is a client in  $C \setminus \{c_{i_j}\}$ . We prove that with probability at least  $1 - \frac{k}{q}$  it holds that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i) \cup \{\alpha^j\}$  is of rank  $n$ . Note that the rank of vector set  $S_i = A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i)$  is at least  $n - 1$ . Indeed, if  $c_i$  is a client in the set  $C_1^j$ , then since all of the transmissions,  $T$ , are necessary to get all the packets; when we remove a packet  $v$  from  $\Omega_{j-1}$  we decrease its rank to  $n - 1$ . But if  $c_i$  is a client in  $C_2^j \setminus \{c_{i_j}\}$ , then the rank of the set  $S_i = A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i)$  could still be  $n$ . Because, the vector,  $v$ , which is removed from  $\Omega_{j-1}$  may not be necessary for the client  $c_i$ . If

it is useful for the client  $c_i \in C_2^j \setminus \{c_{i_j}\}$ , then the situation will be same as previous case, a client which is in the set  $C_1^j$ . Therefore, the rank of vector set  $S_i$  is at least  $n - 1$  and we only need to consider the case in which  $S_i$  is of rank  $n - 1$  since a client with  $S_i$  of rank  $n$  will not be affected by the removal of packet  $v$ .

We denote  $\gamma_i$  as the normal vector to the span of  $S_i$  which can be written as  $\gamma_i = \sum_{u_g \in U(c_{i_j})} \beta_g u_g + \sum_{u_g \in \bar{U}(c_{i_j})} \beta_g u_g$ , where  $\bar{U}(c_{i_j})$  is the set of unit encoding vectors that correspond to  $W(c_{i_j}) = P \setminus H(c_{i_j})$ . In fact, If we show that  $\gamma_i$  and  $\alpha^j$  are not orthogonal with high probability, then we prove the claim that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i) \cup \{\alpha^j\}$  is of rank  $n$  with high probability at most  $1 - \frac{k}{q}$ . In other words, to prove the claim it will suffice to show that the inner product  $\langle \gamma_i, \alpha^j \rangle$  between  $\gamma_i$  and  $\alpha^j$  is not equal to zero with probability at least  $1 - \frac{k}{q}$ .

We can show that there exists  $u_g \in U(c_{i_j})$  such that  $\beta_g \neq 0$  by contradiction. If that is not the case, then we can write  $\gamma_i$  as  $\gamma_i = \sum_{u_g \in \bar{U}(c_{i_j})} \beta_g u_g$ . For each  $u_g \in \bar{U}(c_{i_j})$ , the span of  $\alpha^{j-1} \cup \tilde{\Omega}_{j-1}$  must include a vector  $v_g = u_g + \sum_{u_t \in U(c_{i_j})} \alpha_t$  which is orthogonal to  $\gamma_i$ . However, since this means that  $\beta_g$  is equal to zero for each  $u_g \in \bar{U}(c_{i_j})$ , we will have a contradiction with the fact that  $\gamma_i$  is not identical to zero.

We can write the inner product  $\langle \gamma_i, \alpha^j \rangle$  as  $\langle \gamma_i, \alpha^j \rangle = \sum_{u_g \in U(c_{i_j})} \alpha_g^j \beta_g$  since  $\alpha^j$  is a random linear combination of vectors in  $U(c_{i_j})$ , i.e.,  $A_j = \sum_{u_g \in U(c_{i_j})} \alpha_g^j u_g$  where  $\alpha_g^j$  are random coefficients over a field  $F$ . Let  $\hat{U}$  be a subset of  $U(c_{i_j})$  such that for each  $u_g \in \hat{U}$  it holds that  $\beta_g \neq 0$ . Using the statement we have showed above, we can be sure that the set  $\hat{U}$  is not empty and so,  $\langle \gamma_i, \alpha^j \rangle = \sum_{u_g \in \hat{U}} \alpha_g^j \beta_g$ . Since for each  $u_g \in \hat{U}$ ,  $\alpha_g^j$  is a random variable chosen independently of  $\{\beta_g u_g \in \hat{U}\}$  the probability that  $\langle \gamma_i, \alpha^j \rangle$  is equal to zero is at most  $\frac{1}{q}$ .

Finally, we can prove that the probability that  $\langle \gamma_i, \alpha^j \rangle = 0$  for some client  $c_i \in C$  is bounded by  $\frac{k}{q}$  by utilizing the union bound. Therefore, for each client  $c_i \in C$ , it is true that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i) \cup \{\alpha^j\}$  is of rank  $n$ , with probability at least  $1 - \frac{k}{q}$ .

This means that after iteration  $j$  of the algorithm, the data transfer can be completed within  $T - (j - 1) - 1 = T - j$  transmissions by using vectors in  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i)$ .

Note that at the iteration  $j$ , only the clients in the set  $C_2^j$  are allowed to transmit. Thus, both client  $c_{i_j}$  and the client that transmits vector  $v$  belong  $\in C_2^j$ . Since  $c_{i_j}$  has the lowest transmission cost among all clients in  $C_2^j$ , this implies that the cost of  $\tilde{\Omega}_{j-1} \cup \{\alpha^j\}$  is equal to  $OPT_{j-1}$ . Note that after iteration  $j$  the information exchange can be completed by broadcastings vectors in  $\tilde{\Omega}_{j-1}$ , hence the cost of  $\tilde{\Omega}_{j-1}$  is equal to  $OPT_j$ . This, in turn, implies that  $OPT_j = OPT_{j-1} - \delta(c_{i_j})$  and the lemma follows.

More formally, Lemma 1 states that after each iteration, the optimality of the DEWC problem holds with a probability at least  $1 - \frac{k}{q}$ . Since the information transfer will be completed after  $T$  iterations, the algorithm computes an optimal solution for Data Exchange with Costs problem with probability at least

$$\left(1 - \frac{k}{q}\right)^T \geq \left(1 - \frac{k}{q}\right)^n \geq 1 - \frac{nk}{q} \quad (3.3)$$

provided that  $q > n$ . Indeed, probability of success can be increased by performing multiple iterations and by averaging the total minimum cost values among all the iterations.

### 3. Numerical Results

In this section, we verify the performance of the algorithm presented in the previous section. We have compared the minimum total cost value without coding and the minimum total cost value obtained by the algorithm. In the analysis, total minimum cost values for both traditional approach and network coding are calculated. In traditional approach, for each packet,  $p_i$ , the client,  $c_j$ , who owns that specific packet and has the smallest associated cost value,  $\delta(c_j)$ , is found and let that client transmit the packet,  $p_i$ . This value is compared with the total minimum cost found

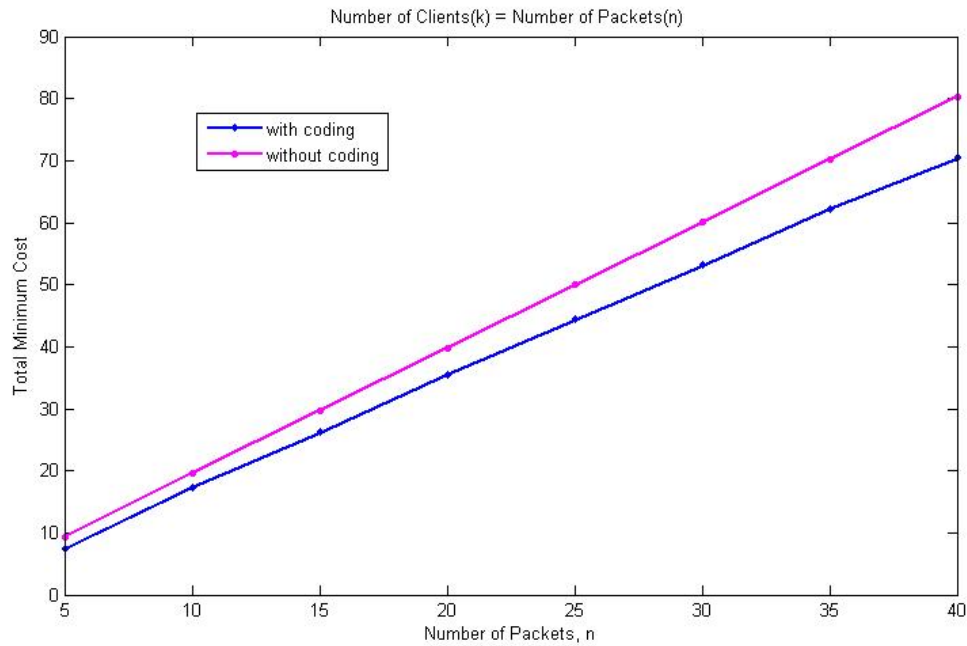


Fig. 10. Performance Results for Randomized Algorithm with  $k = n$ .

by the randomized algorithm.

Figure 10 and Figure 11 show numerical results of minimum total cost values for  $n = (5, 10, \dots, 40)$  packets. In our numerical analyses, each client has different cost values starting from 1 to  $k$  to make the analyses more convenient. In addition, each cost value is calculated by averaging over 100 random initializations of the problem.

In Figure 10 the upper line represents total minimum cost values for traditional approach; on the other hand the lower line shows the total minimum cost values when network coding technique is utilized. In this analysis, the number of clients are chosen to be equal to the number of packets. Remarkably, the algorithm performs better than traditional approach for all the cases.

We have observed similar trends when we fix the number of clients,  $k$ , and



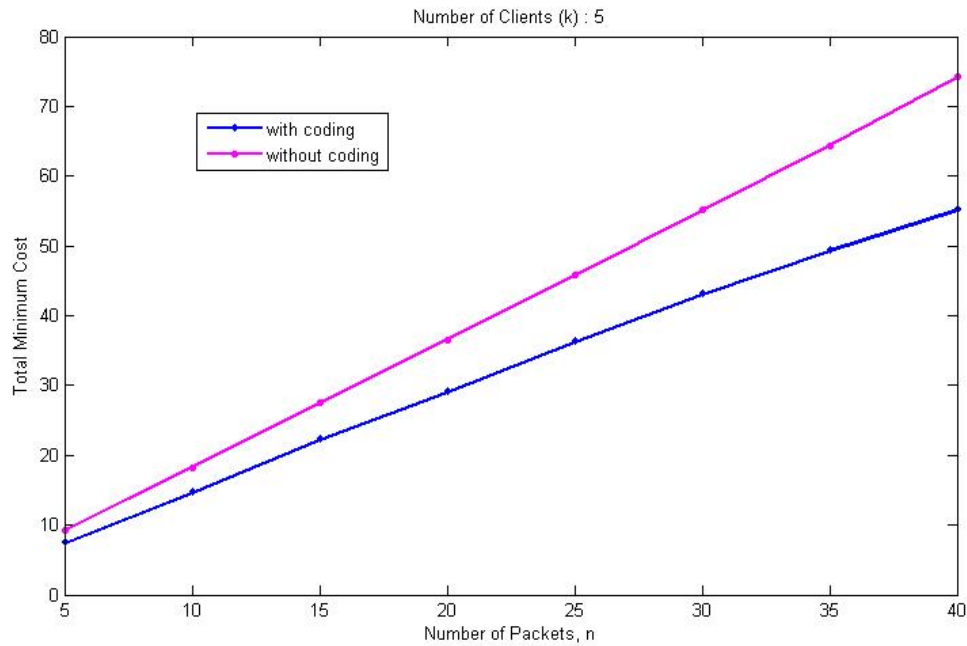


Fig. 11. Performance Results for Randomized Algorithm with  $k = 5$ .

changed the number of packets,  $n$ . In particular, Figure 11 shows the performance comparison between traditional approach and the randomized algorithm for 5 clients. As in Figure 10 the upper line represents total minimum cost values for traditional approach while the lower line shows the total minimum cost when network coding technique is considered. The algorithm again performs better than traditional approach for all the cases. Indeed, in Figure 11 the gap between the lines is wider than the one in Figure 10 meaning that coding assures higher gain for the case of fixed client number. In other words, when we hold the number of clients stable, then the performance gain of network coding can be better especially for large number of packets.

## D. Deterministic Algorithm

### 1. Algorithm

The randomized algorithm explained in the previous part gives an optimal solution for Data Exchange with Costs Problem with a high probability when the field size is large. However, this randomness creates ambiguity in the solution even if the probability of failure is so small for a very large field. Therefore, it is important to get rid of this indeterminacy and modify the algorithm such that the problem can be solved in a deterministic way. In this section, we present a deterministic algorithm for Data Exchange with Costs Problem.

We again describe and analyze the algorithm in terms of encoding vectors, rather than the original packets, i.e., rather than saying that a packet  $p^j = \sum_{p_j \in P} \alpha_i^j p_j$  has been transmitted by a client  $c_{i_j}$  at round  $j$ , we state that the client transmits the encoding vector  $\alpha^j = \{\alpha_1^j, \dots, \alpha_n^j\}$  as in the previous part.

The algorithm runs over a finite field  $F_q$  where the size  $q$  must be larger than  $2k$  for  $k$  is the number of clients. Let  $U(c_{i_j})$  be the set of encoding vectors available to client  $c_{i_j}$ . For a client  $c_i \in C$  we define  $A(c_i)$  as the set of all possible encoding vectors in  $F_q^n$  that can be generated by client  $c_i$ . In other words,  $A(c_i) = \text{span}(U(c_i))$  such that each vector  $\alpha^j \in A(c_i)$  can be expressed as  $\alpha^j = \sum_{u_g \in U(c_j)} \alpha_g^j \cdot u_g$ .

As in the randomized algorithm the clients are divided into two subsets  $\{C_1^j, C_2^j\}$  where only the clients in  $C_2^j$  are allowed to transmit. The deterministic algorithm also executes in iterations like the randomized algorithm. In each iteration a client in the set  $C_2^j$  that will be transmitting at that round is determined but not the coefficients of the encoding vector. The encoding coefficients of each transmitted packets will be determined at the last stage of the algorithm.

More formally, for each client  $c_i \in C$  a counter  $b_i$  is saved to specify the number of

the packets that will be transmitted by that client. In the beginning of the algorithm  $b_i = 0$  for all  $c_i \in C$ . Indeed, since the a client  $c_l \in C_1^j$  is not allowed to transmit any packet,  $b_l$  value for a client  $c_l \in C_1^j$  will always be zero. Once it is determined that a client  $c_{i_j} \in C_2^j$  is transmitting a packet in  $j^{th}$  iteration, we increment the corresponding counter  $b_{c_{i_j}}$ . Assume  $B_j$  is the vector that specifies the number of transmissions made by each client  $c_i \in C$  at round  $j$ .

Assume for a client  $c_i \in C$  we define  $A(c_i) = \text{span}(U(c_i))$ , i.e.,  $A(c_i)$  is the set of all possible encoding vectors in  $F_q^n$  that can be generated by client  $c_i$ . Also, we say that the set  $A$  fits  $B_j$  if it is a union of  $b_1$  vectors from  $A(c_1)$ ,  $b_2$  vectors from  $A(c_2)$ , ... and  $b_k$  vectors from  $A(c_k)$ .

Deterministic algorithm utilizes the max-rank concept. Let  $M(B_j)$  as the collection of all sets of encoding vectors that fit  $B_j$ . Then, given  $B_j$  and  $U(c_i)$  and  $M(B_j)$ ,  $\text{Maxrank}(B_j; U(c_i))$  can be defined as  $\text{Maxrank}(B_j; U(c_i)) = \max_{A \in M(B_j)} \text{rank}(A \cup U(c_i))$ . For given  $B_j$  and  $U(c_i)$ , it is possible to compute efficiently the value of  $\text{Maxrank}(B_j; U(c_i))$  in polynomial time by constructing a bipartite graph  $G = (V_1; V_2; E)$  where the set of nodes in  $V_1$  correspond to packets in  $P$ . For each client  $c_i \in C$ , the nodes in  $V_2$  are composed of  $b_i$  nodes which are connected to all nodes in  $V_1$  that correspond to packets in  $H(c_i)$  and  $|H(c_i)| = |U_i|$  nodes each of which is connected to a corresponding packet in  $V_1$ . Indeed, each node in  $V_2$  corresponds to a linear combination of packets in  $P$ . It is in fact easy to verify that the value of  $\text{Maxrank}(B_j; U(c_i))$  is equal to the maximum size of a matching in  $G = (V_1; V_2; E)$ .

Like in the randomized algorithm, let  $T$  is the number of transmissions with which minimum total cost for the overall transmissions can be obtained, that is, it is the number of the transmissions for the optimal solution of the problem. Assume that we have made a good guess on the value of  $T$ . Indeed, it is not hard to find the exact value of  $T$ . Since  $n - n_{min} \leq T \leq n$ , i.e., the value of  $T$  is upper bounded by

$n$ , we can implement exhaustive search to find its value.

The steps performed by the algorithm can be summarized as follows:

**DETERMINISTIC ALGORITHM**

- 1 **for**  $T \leftarrow n - n_{min}$  **to**  $n$
- 2      $B_0 = (b_1, \dots, b_k)$  where all of them are zero.
- 3     **for**  $i \leftarrow 1$  **to**  $T$
- 4         Choose the client  $c_{i_j} \in C_2$  with min. transmission cost.
- 5         Let  $b_{c_{i_j}} \leftarrow b_{c_{i_j}} + 1$ ,  
 $B_j = (b_1, \dots, b_k)$ .
- 6         Update the sets  $C_1^j$  and  $C_2^j$ .
- 7         Find a vector set  $\hat{A} \in M(B_T)$  such that  
 $\text{rank}(\hat{A} \cup U(c_i)) = n$  for all  $c_i \in C$ .
- 8         Calculate the total minimum cost for chosen  $T$  value, i.e.,  
 $\Delta_T = \sum_{i=1}^T \delta(c_{i_j})$ .
- 9 **return** the total minimum cost among all  $T$  values, i.e.,  
 $\Delta = \arg \min_{T \in [n - n_{min}, n]} (\Delta_T)$ .

## 2. Analysis

We proceed to analyze the correctness and optimality of the algorithm. Consider an iteration  $j$  of the algorithm. We denote  $\text{OPT}_j$  as the minimum total transmission cost of completing the information transfer after round  $j$ , provided that at least  $T - j$  transmissions are allowed after round  $j$ . Recall that the vector  $B_j = (b_1, b_2, \dots, b_k)$  specifies the number transmissions made by each client  $c_i$  during iterations  $1, \dots, j$ . Also, assume  $L_j = (l_1, l_2, \dots, l_k)$  is the number of additional transmissions that each client needs to do to achieve optimum  $T - j$ , i.e.,  $T - j = \sum_{i=1}^k l_i$ . As in the randomized

algorithm we try to show that for each iteration  $j$  it holds  $\text{OPT}_j = \text{OPT}_{j-1} - \delta(c_{i_j})$  in order to prove the optimality of the algorithm.

**Lemma 2** *For each iteration  $i$ ,  $\text{OPT}_j = \text{OPT}_{j-1} - \delta(c_{i_j})$ .*

**Proof:** Consider iteration  $j$  of the algorithm. Let  $c_{i_j}$  be the client selected at that iteration. Consider first the case in which  $l_{c_{i_j}} > 0$ . Let  $B_j$  be a vector formed from  $B_{j-1}$  be incrementing  $b_{c_{i_j}}$  by one and  $L_j$  be a vector formed from  $L_{j-1}$  be decrementing  $b_{c_{i_j}}$  by one. Since  $B_j + L_j = B_{j-1} + L_{j-1}$ , after iteration  $j$  we need  $T - (j - 1) - 1$  transmissions to satisfy  $\text{Maxrank}(B_T; U(c_i)) = n$  for each client  $c_i \in C$ . Since  $\delta(c_{i_j})$  is the client having minimum cost value among from all the clients allowed to transmit, it holds that  $\text{OPT}_j = \text{OPT}_{j-1} - \delta(c_{i_j})$ . Consider now the case in which  $l_{c_{i_j}} = 0$ . Note that for each client  $c_i \in C$  it holds that  $\text{Maxrank}(B_{j-1} + L_{j-1}; U(c_j)) = n$ . Then, there exist vector sets  $A_{j-1} \in M(B_{j-1})$  and  $\Omega_{j-1} \in M(L_{j-1})$  satisfying the conditions;  $\text{rank}(A_{j-1} \cup U(c_i)) = \text{Maxrank}(B_{j-1}; U(c_i))$  and  $\text{rank}(A_{j-1} \cup \Omega_{j-1} \cup U(c_i)) = n$ . Since we choose the client  $c_{i_j}$  with minimum cost among from the set  $C_2^j$ , it is true that we can remove at least one packet,  $v$ , from  $\Omega_{j-1}$  so that  $\tilde{\Omega}_{j-1} = \Omega_{j-1} \setminus \{v\}$  such that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_{i_j})$  still remains of rank  $n$ . Indeed, the proof of this statement is shown in the previous section.

Let  $c_i$  be a client in  $C \setminus c_{i_j}$  and denote  $\gamma_i$  as the normal vector to the span of  $S_i$  which can be written as  $\gamma_i = \sum_{u_g \in U(c_{i_j})} \beta_g u_g + \sum_{u_g \in \bar{U}(c_{i_j})} \beta_g u_g$ , where  $\bar{U}(c_{i_j})$  is the set of unit encoding vectors that correspond to  $W(c_{i_j}) = P \setminus H(c_{i_j})$ . For  $\bar{\gamma}_i$  is a projection of  $\gamma_i$  to span  $U(c_{i_j})$  and  $w \in \text{span}(U(c_{i_j}))$  is a vector for which  $\langle w, \bar{\gamma}_j \rangle \neq 0$  for each  $c_i \in C$ , it is true that  $A_j \in M(B_j)$  and  $\Omega_j \in M(L_j)$  where  $A_j = A_{j-1} \cup w$  and  $\Omega_j = \Omega_{j-1} \setminus w$ . Note that for each client  $c_i \in C$  it holds that  $\text{Maxrank}(B_j + L_j; U(c_i)) = n$ . Thus, after iteration  $j$  we need  $T - (j - 1) - 1 = T - j$  transmissions to satisfy  $\text{Maxrank}(B_T; U(c_i)) = n$  for each client  $c_i \in C$ . Since at the iteration  $j$ , only the

clients in the set  $C_2^j$  are allowed to transmit, both client  $c_{i_j}$  and the client that transmits vector  $v$  belong  $\in C_2^j$ . Since  $c_{i_j}$  has the lowest transmission cost among all clients in  $C_2^j$ , this implies that the cost of  $\tilde{\Omega}_{j-1} \cup \{\alpha^j\}$  is equal to  $OPT_{j-1}$ . Note that after iteration  $j$  the information exchange can be completed by broadcastings vectors in  $\tilde{\Omega}_{j-1}$ , hence the cost of  $\tilde{\Omega}_{j-1}$  is equal to  $OPT_j$ . This, in turn, implies that  $OPT_j = OPT_{j-1} - \delta(c_{i_j})$ .

### 3. Numerical Results

In this section, we verify the performance of the deterministic algorithm together with the randomized algorithm. We have compared the minimum total cost value without coding and the minimum total cost value obtained by both the randomized and deterministic algorithms. In the analysis, total minimum cost values for both traditional approach and network coding are calculated. In traditional approach, for each packet,  $p_i$ , the client,  $c_j$ , who owns that specific packet and has the smallest associated cost value,  $\delta(c_j)$ , is found and let that client transmit the packet,  $p_i$ . This value is compared with the total minimum cost found by the randomized and deterministic algorithms.

Figure 12 shows numerical results of minimum total cost values for a total number of 5 clients and  $n = (5, 10, \dots, 40)$  packets. In the analysis, each client has different cost values starting from 1 to  $k$  to make the analyses more convenient. In addition, each cost value is calculated by averaging over 100 random initializations of the problem.

In Figure 12 the upper line represents total minimum cost values for traditional approach; on the other hand the lower lines show the total minimum cost values when network coding technique is utilized. Remarkably, the deterministic algorithm

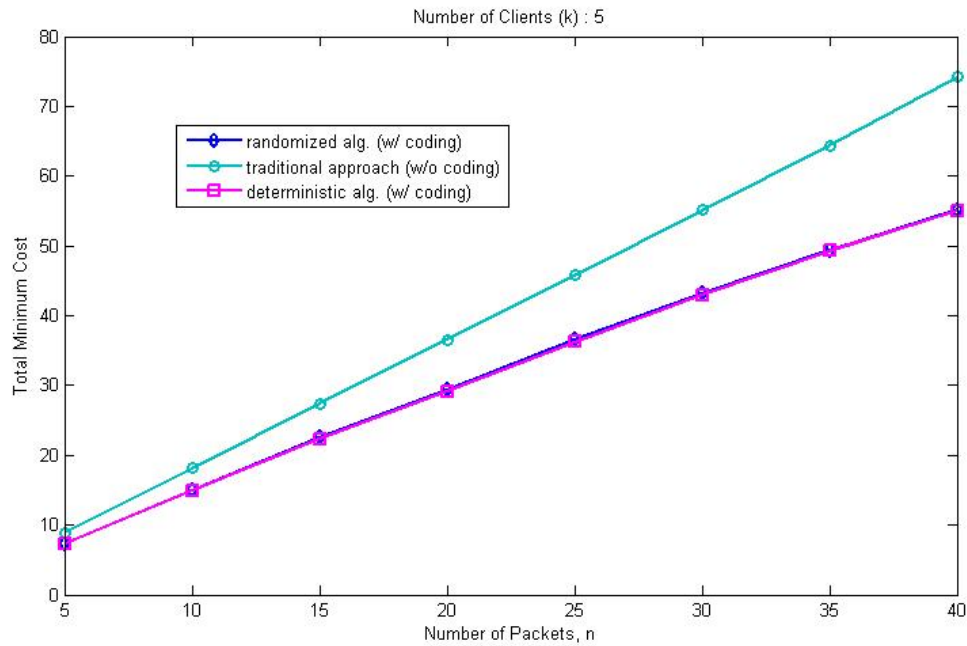


Fig. 12. Performance Comparison for DEWC Problem.

performs nearly same as the randomized algorithm. Indeed, there is a really small difference for some cases which is caused by the fact that the randomized algorithm gives optimal results that are really close to 1 but not exactly, however deterministic algorithm always gives optimal total transmission cost value.

#### E. Randomized Algorithm with Restrictions on Number of Transmissions

##### 1. Algorithm

The randomized algorithm explained in the first part gives an optimal solution for Data Exchange with Costs Problem with a high probability if the field size is large enough where there exists no limit in the number of transmissions each client can make. However, each client can have different battery lives, i.e., the number of

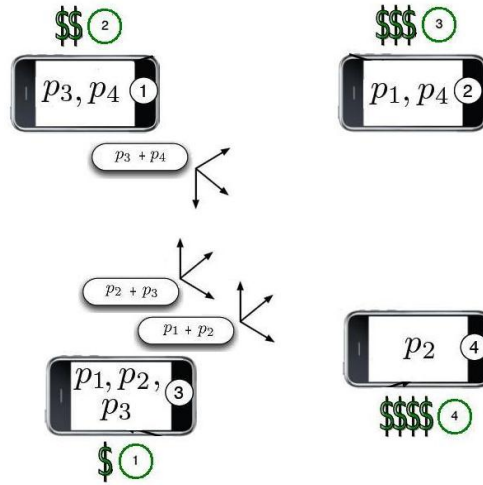


Fig. 13. DEWC Problem without any Restrictions.

transmissions each client is allowed to make could be restricted. In this part of the thesis, we improved the randomized algorithm for Data Exchange with Costs Problem when there exists a restriction on the number of transmissions of the clients. This algorithm also assumes a large finite field  $F$  of size  $q$  and provides an optimal solution with high probability.

Figure 13 is an example with four wireless clients that are interested in obtaining four packets of  $m$  bits each,  $p_1, p_2, p_3$  and  $p_4 \in GF(2^m)$ . Each client lacks some packets, however they collectively know all the packets. Also, each client is associated with a transmission cost. Assume again that each client can broadcast the packets via a noiseless broadcast channel and is fully aware of the packets available to other nodes.

Figure 13 shows the solution which is composed of three coded transmission two of which is from the client,  $c_3$  and one of which is from the client,  $c_1$ . The randomized algorithm first chooses the number of transmission as three, then in the



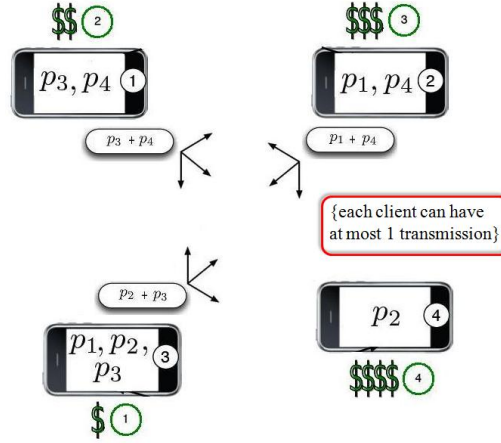


Fig. 14. DEWC Problem with Restrictions on the Number of Transmissions.

first and second rounds it will chose the client,  $c_3$  as transmitter since it has the minimum cost. However, in the last round the client,  $c_3$  will be in the set  $C_1$ , so it can not broadcast a packet anymore. Therefore, the algorithm will chose the client,  $c_1$  which is the one with minimum cost in the set  $C_2$ . Therefore, total minimum cost obtained from the algorithm will be  $1 + 1 + 2 = 4$ .

In this scenario the client,  $c_3$  is responsible for two transmission. If we restrict all the clients to make at most 1 transmissions, then the solution of the problem should be changed. Figure 14 gives the optimal solution for the problem when the number of transmissions of each client is restricted to at most 1. Now, in the first round  $c_3$  will transmit a packet, then in the second round since it cannot transmit anymore,  $c_1$  will broadcast a packet and finally  $c_2$  will be responsible for the final round. Therefore, total minimum cost obtained will increase to  $1 + 2 + 3 = 6$ . Even though the total cost is increased, we can still find the optimal solution for Data Exchange with Costs problem even there exists restrictions on the number of transmissions allowed.

In the construction of the problem DEWC with restriction, we assume that  $\tau(c_i)$

is the value specifying the number of transmissions that a client  $c_i$  is allowed to make. Therefore, a client,  $c_{i_j}$  with the minimum cost can transmit information to the other clients at round  $j$  if  $\tau c_{i_j} \geq 1$ . In the beginning, for each client,  $c_i$ ,  $\tau(c_i)$  value is chosen randomly. However, we put a lower and an upper bound on the number of total transmissions that all the clients can make such that  $n - n_{min} \leq \sum_{i=1}^k \tau(c_i) \leq 2n$ . It is worth to state that this problem is not feasible for all various scenarios, so we only consider and analyse the scenarios that have feasible solutions.

The algorithm works in the same way as the randomized algorithm. We can partition the clients as  $C_1^j$  and  $C_2^j$  such that only the clients in the set,  $C_2^j$  can transmit. Then, at each iteration  $j$  a client,  $c_{i_j} \in C_2$ , with a minimum cost value is chosen. However, the client,  $c_{i_j}$  can transmit data only if it is allowed to transmit, i.e.,  $\tau(c_{i_j}) \geq 1$ ; otherwise the algorithm finds another client  $c_l \in C_2^j$  which has the second minimum cost among the clients that are allowed to transmit information, i.e.,  $\tau(c_l) \geq 1$ . Therefore, we have updated the allocation of the clients for  $j^{th}$  round as;

- a set  $C_1^j = \{c_i \in C \mid n - n_i^j = T - (j - 1)\}$  of clients whose demand set has a size of  $T - (j - 1)$ ;
- a set  $C_2^j = \{c_i \in C \mid n - n_i^j < T - (j - 1)\}$  of clients whose demand set has a size strictly smaller than  $T - (j - 1)$ .
  - a subset  $C_{2_1}^j = \{\tau(c_i) < 1 \mid c_i \in C_2^j\}$  of clients that are not allowed to transmit
  - a subset  $C_{2_2}^j = \{\tau(c_i) \geq 1 \mid c_i \in C_2^j\}$  of clients that are allowed to transmit

At each round of the algorithm, a client in  $C_{2_2}^j$  with the minimum cost value will be responsible for the transmission. Since this client has the minimum cost among the

clients that are allowed to transmit, at the end the algorithm gives optimal solution for the problem if the problem has a feasible solution.

If an instance of the problem has a feasible solution, when the algorithm executes we will get the total minimum cost as the sum of the transmission costs of each of  $T$  transmissions.

The steps performed by the algorithm can be summarized as follows:

**RAND. ALG. WITH RESTRICTIONS**

- 1 **for**  $T = n - n_{min}$  **to**  $n$
- 2     **for**  $i = 1$  **to**  $T$
- 3         Select a client  $c_{i_j} \in C_{2_2}^j$  with min. trans. cost value.
- 4         Create a new encoding vector  $\alpha^j$ , with random coefficients.
- 5         Transmit the packet  $p^j$ , let
 
$$\tau(c_{i_j}) \leftarrow \tau(c_{i_j}) - 1,$$
- 6         Update the sets  $C_{1_1}^j$ ,  $C_{2_1}^j$  and  $C_{2_2}^j$ .
- 7         Calculate the total minimum cost for chosen  $T$  value, i.e.,
 
$$\Delta_T = \sum_{i=1}^T \delta(c_{i_j}).$$
- 8     Check if the problem has a feasible solution;
- 9     If for all  $T$  values there is no solution, the problem is unfeasible.
- 10    Otherwise, **return** the total minimum cost among all  $T$  values, i.e.,
 
$$\Delta = \arg \min_{T \in [n - n_{min}, n]} (\Delta_T).$$

This algorithm is also constructed on linear coding in which  $\alpha^j = \{\alpha_1^j, \dots, \alpha_n^j\}$  is the encoding vector of any packet,  $p^j$ . In the analysis of the algorithm which runs in rounds, instead of expressing the original packets we again use encoding vectors.

## 2. Analysis

We now analyze the correctness and optimality of the algorithm. As in the randomized algorithms, let  $\text{OPT}_j$  as the minimum total transmission cost of completing the information transfer after round  $j$ , provided that at least  $T - j$  transmissions are allowed after round  $j$ .

**Lemma 3** *With a probability at least  $1 - \frac{k}{q}$ ,  $\text{OPT}_j = \text{OPT}_{j-1} - \delta(c_{i_j})$ .*

**Proof:** We again denote  $\Omega_{j-1}$  as an optimal set of encoding vectors necessary for the data transfer.  $\mu = \text{rank}(A_{j-1})$  is the rank of the set of packets transmitted so far and  $\mu^j = \text{rank}(U(c_{i_j}) \cup A_{j-1})$  is the rank of the set of encoding vectors which client  $c_{i_j} \in C_{22}^j$  own at the beginning of iteration  $j$ , Assume that  $c_{i_j}$  is the client which is allowed to transmit,  $\tau(c_{i_j}) \geq 1$ , and has a minimum cost among the set  $C_2^j$ . Note that if we could not find such a client, then either our guess on  $T$  is wrong or the problem does not have a feasible solution. Otherwise,  $T - (j - 1)$  is at least  $T - \mu^j$  which means that it is at least 1. So, we can remove at least one packet,  $v$ , from  $\Omega_{j-1}$  so that  $\tilde{\Omega}_{j-1} = \Omega_{j-1} \setminus \{v\}$  such that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_{i_j})$  still remains of rank  $n$ .

We prove that for another client  $c_i$  in  $C \setminus \{c_{i_j}\}$ , it holds that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i) \cup \{\alpha^j\}$  is of rank  $n$  with probability at least  $1 - \frac{k}{q}$ . Note that, for the client  $c_i$  the rank of the set  $S_i = A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i)$  is at least  $n - 1$  as explained in the previous part. Therefore, again we only need to consider the case in which  $S_i$  is of rank  $n - 1$  since a client with  $S_i$  of rank  $n$  will not be affected by the removal of packet  $v$ .

To prove the claim that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i) \cup \{\alpha^j\}$  is of rank  $n$  with high probability at most  $1 - \frac{k}{q}$ , we show that  $\gamma_i$  and  $\alpha^j$  are not orthogonal with high probability where  $\gamma_i = \sum_{u_g \in U(c_{i_j})} \beta_g u_g + \sum_{u_g \in \bar{U}(c_{i_j})} \beta_g u_g$  is the normal vector to the span of  $S_i$ . In other words, we will show that the inner product  $\langle \gamma_i, \alpha^j \rangle$  between  $\gamma_i$  and  $\alpha^j$  is not equal to zero with probability at least  $1 - \frac{k}{q}$ .

We can show that there exists  $u_g \in U(c_{i_j})$  such that  $\beta_g \neq 0$  by contradiction. If that is not the case, then we can write  $\gamma_i$  as  $\gamma_i = \sum_{u_g \in \bar{U}(c_{i_j})} \beta_g u_g$ . For each  $u_g \in \bar{U}(c_{i_j})$ , the span of  $\alpha^{j-1} \cup \tilde{\Omega}_{j-1}$  must include a vector  $v_g = u_g + \sum_{u_t \in U(c_{i_j})} \alpha_t$  which is orthogonal to  $\gamma_i$ . However, since this means that  $\beta_g$  is equal to zero for each  $u_g \in \bar{U}(c_{i_j})$ , we will have a contradiction with the fact that  $\gamma_i$  is not identical to zero.

We can again write the inner product  $\langle \gamma_i, \alpha^j \rangle$  as  $\langle \gamma_i, \alpha^j \rangle = \sum_{u_g \in U(c_{i_j})} \alpha_g^j \beta_g$  since  $\alpha^j$  is a random linear combination of vectors in  $U(c_{i_j})$ , i.e.,  $A_j = \sum_{u_g \in U(c_{i_j})} \alpha_g^j u_g$  where  $\alpha_g^j$  are random coefficients over a field  $F$ . Let  $\hat{U}$  be a subset of  $U(c_{i_j})$  such that for each  $u_g \in \hat{U}$  it holds that  $\beta_g \neq 0$ . Using the statement we have showed above, we can be sure that the set  $\hat{U}$  is not empty and so,  $\langle \gamma_i, \alpha^j \rangle = \sum_{u_g \in \hat{U}} \alpha_g^j \beta_g$ . Since for each  $u_g \in \hat{U}$ ,  $\alpha_g^j$  is a random variable chosen independently of  $\{\beta_g u_g \in \hat{U}\}$  the probability that  $\langle \gamma_i, \alpha^j \rangle$  is equal to zero is at most  $\frac{1}{q}$ .

Finally, we can prove that the probability that  $\langle \gamma_i, \alpha^j \rangle = 0$  for some client  $c_i \in C$  is bounded by  $\frac{k}{q}$  by utilizing the union bound. Therefore, for each client  $c_i \in C$ , it is true that  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i) \cup \{\alpha^j\}$  is of rank  $n$ , with probability at least  $1 - \frac{k}{q}$ . This means that after iteration  $j$  of the algorithm, the data transfer can be completed within  $T - (j - 1) - 1 = T - j$  transmissions by using vectors in  $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_i)$ .

Note that at the iteration  $j$ , only the clients in the set  $C_{22}^j$  are allowed to transmit. Thus, both client  $c_{i_j}$  and the client that transmits vector  $v$  belong  $\in C_{22}^j$ . Since  $c_{i_j}$  has the lowest transmission cost among all clients in  $C_{22}^j$ , this implies that the cost of  $\tilde{\Omega}_{j-1} \cup \{\alpha^j\}$  is equal to  $OPT_{j-1}$ . Note that after iteration  $j$  the information exchange can be completed by broadcastings vectors in  $\tilde{\Omega}_{j-1}$ , hence the cost of  $\tilde{\Omega}_{j-1}$  is equal to  $OPT_j$ . This, in turn, implies that  $OPT_j = OPT_{j-1} - \delta(c_{i_j})$ . Since, the algorithm has a total of  $T$  iterations, then the optimal solution for the DEWC with restrictions problem can be obtained by the modified randomized algorithm with a probability at least  $1 - \frac{nk}{q}$  for all feasible scenarios.

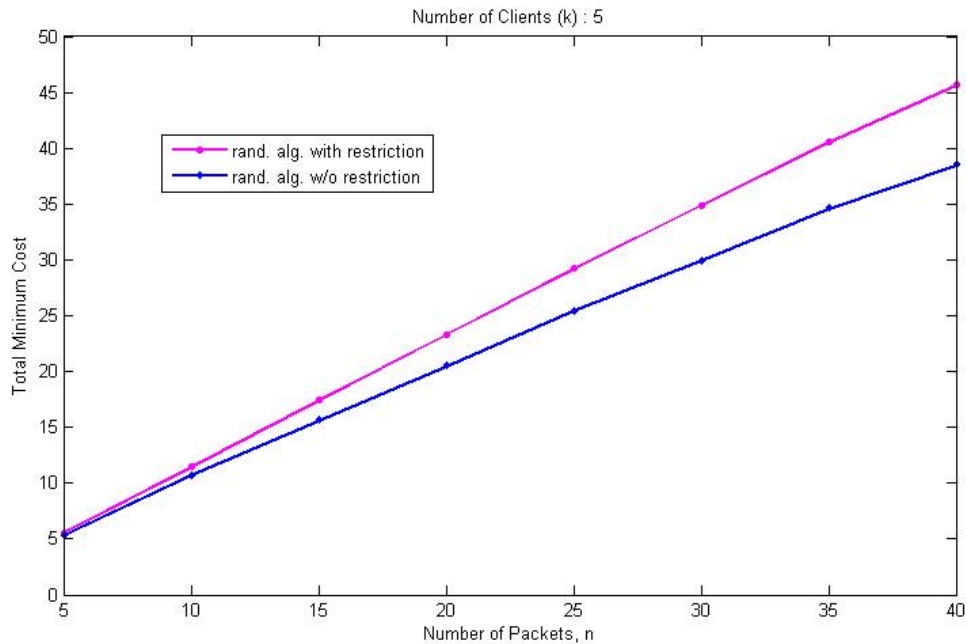


Fig. 15. Performance Comparison for DEWC with Restrictions.

### 3. Numerical Results

In this section, we compared the performance of the algorithms for Data Exchange with Costs Problem with and without restrictions on the number of transmissions. Since there now exists unfeasible cases, we will ignore these cases and just consider the scenarios in which all clients will eventually obtain all the packets.

Figure 10 shows numerical results of minimum total cost values for  $n = (5, 10, \dots, 40)$  packets with  $k = 5$  clients. In addition, each cost value is calculated by averaging over 100 random initializations of the problem. It is worth to note that, in the analysis, we have ignored the scenarios that are not feasible for the Data Exchange with Costs Problem with restrictions on the number of transmissions.

In Figure 15 the upper line represents total minimum cost values calculated by

the randomized algorithm when the number of transmissions each clients can make is restricted. On the other hand, the lower line shows the total minimum cost values calculated by the randomized algorithm without any restrictions on the number of transmissions. Remarkably, the restriction decreases the performance of the randomized algorithm, but if the problem is feasible we could still get the optimal solution for Data Exchange with Costs Problem.

## CHAPTER IV

## DIRECT INFORMATION EXCHANGE THROUGH RELAY PROBLEM

## A. Data Exchange Through Relay (DETR) Problem

## 1. Background

In this chapter, we consider Data Exchange through Relay (DETR) Problem. In this problem, clients can not communicate among each other and communication takes place through a server. Initially, the server has no information about the clients. First, the clients send packets or combinations thereof to the server (uplink transmissions). Then, the server broadcasts linear combinations of the packets to clients (downlink transmissions). Each client needs all the lacking information packets. In other words; at the end all clients would obtain all the packets in the packet set. Our goal is again to minimize the number of uplink and downlink transmissions.

This problem is very similar to Cooperative Data Exchange (CDE) Problem studied in [13] which is introduced in Data Exchange with Costs (DEWC) Problem. In this part of the thesis, we consider Cooperative Data Exchange (CDE) Problem in a more detailed manner.

Figure 16 demonstrates the problem studied in [13]. There are four wireless clients who had requested  $m = 4$  packets,  $p_1, \dots, p_4 \in GF(2^m)$ , from the base station. However, due to channel imperfections, the third client could have received packets  $\{p_1, p_2, p_3\}$ , while first, second, and fourth clients received packets  $\{p_3, p_4\}$ ,  $\{p_1, p_4\}$ , and  $\{p_1, p_2\}$ , respectively. Since they have collectively received all the packets, they can now try to communicate among themselves to complete the communication and ensure that all the clients eventually possess all the packets. Here, it is assumed that:

1. Each mobile client can broadcast data to all other clients at a rate of one packet



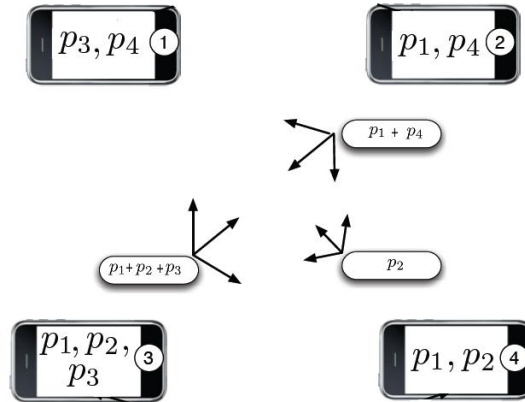


Fig. 16. The CDE Problem.

per transmission and every one receives this transmission error-free;

2. Each client knows which packets were received by others.

An optimal solution to this problem would require three transmissions as shown in Figure 16 where the second, and third clients send the coded packets  $p_1 + p_4$ ,  $p_1 + p_2 + p_3$ , respectively and the fourth client sends  $p_2$ . It can be verified that all the mobile clients can then decode all the packets.

On the other hand, for the DETR case, clients cannot communicate with each other; rather there is a relay node to provide information transfer among the clients. The relay node acts a server such that it will collect all the necessary packets with the minimum number of uplink transmissions and after encoding the packets properly, it will broadcast to the clients. Figure 17 demonstrates the same problem in Figure 16 but with a relay node. Indeed, an optimal solution to this problem could be constructed using the solution of the previous problem; first the second, and third clients send the coded packets  $p_1 + p_4$ ,  $p_1 + p_2 + p_3$ , respectively and the fourth client

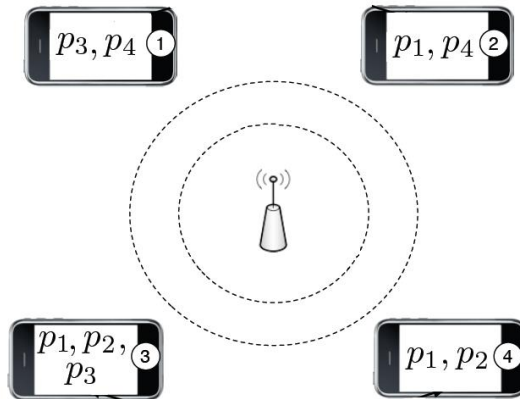
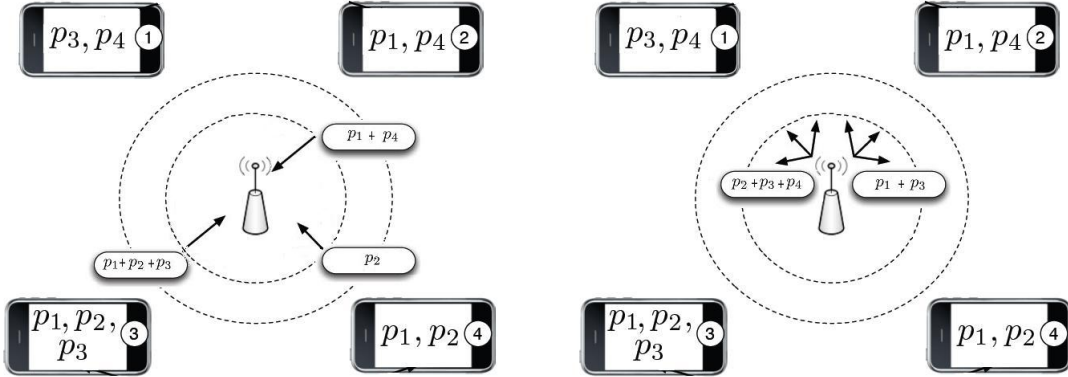


Fig. 17. Data Exchange Through Relay (DETR) Problem.

sends  $p_2$  to the relay node as uplink transmissions. For the downlink solution, two transmissions will be enough; the relay node first creates a linear combination of first and second coded packets;  $p_1 + p_4$  and  $p_1 + p_2 + p_3$  and broadcasts  $p_2 + p_3 + p_4$ . Then, it creates a linear combination of second coded packet and the packet sent by the fourth client;  $p_1 + p_2 + p_3$  and  $p_2$  to broadcast  $p_1 + p_3$  as shown in Figure 18. It can again be verified that all the mobile clients can then decode all the packets. That is, it is possible to create an optimal solution for the DETR Problem using the solution of the CDE problem studied in [13] as uplink transmission solution and the solution of the Index Coding Problem as downlink solution.

## 2. Model

More generally, an instance of the DETR problem includes a relay node  $r$ , a set  $C = \{c_1, \dots, c_n\}$  of wireless clients and a set  $P = \{p_1, p_2, \dots, p_m\}$  of packets that need to be delivered to clients in  $C$ . Each client  $c_i \in C$  is associated with only one set: *side information* set  $H(c_i) \subseteq P$  - the set of packets available at  $c_i$ . Indeed, the



(a) Uplink solution for problem DETR (b) Downlink solution for problem DETR

Fig. 18. Uplink (a) and Downlink (b) Transmissions for Problem DETR

*demand* set will simply be  $W(c_i) = P \setminus H(c_i) \subseteq P$ . In each *round* of communication the relay or a client can transmit a single symbol of  $\Sigma$  (*i.e.*, a single packet). We assume that transmissions of the packets between the clients and the relay occur without an error. We also assume that *Index Coding* solutions are known. In other words, letting the  $j$ 'th downlink round of communication be specified by an encoding function  $g_j^d : \Sigma^m \rightarrow \Sigma$ , the set of encoding functions  $\Phi^d = \{g_i^d\}_{i=1}^{\ell^d}$  that will allow each client to decode the packets it requested while minimizing the number of downlink transmissions  $\ell^d = |\Phi^d|$  are well known. On the other hand, the  $j$ 'th uplink round of communication is specified by an encoding function  $g_j^u : \Sigma^m \rightarrow \Sigma$ . Now, the objective of DETR problem is to find the set of encoding functions  $\Phi^u = \{g_i^u\}_{i=1}^{\ell^u}$  that will allow the relay to construct the set of encoding functions  $\Phi^d = \{g_i^d\}_{i=1}^{\ell^d}$  which are the required *Index Coding* solutions while minimizing the number of uplink transmissions  $\ell^u = |\Phi^u|$ .

### 3. Main Results

In this part of our work, we have showed that we can find an optimal solution for DETR Problem if we consider the problem as the combination of *Cooperative Data Exchange* and *Index Coding* Problems. We have illustrated the equivalence of the DETR Problem with these problems using *Network Information Flow* concept which is the main topic of [1].

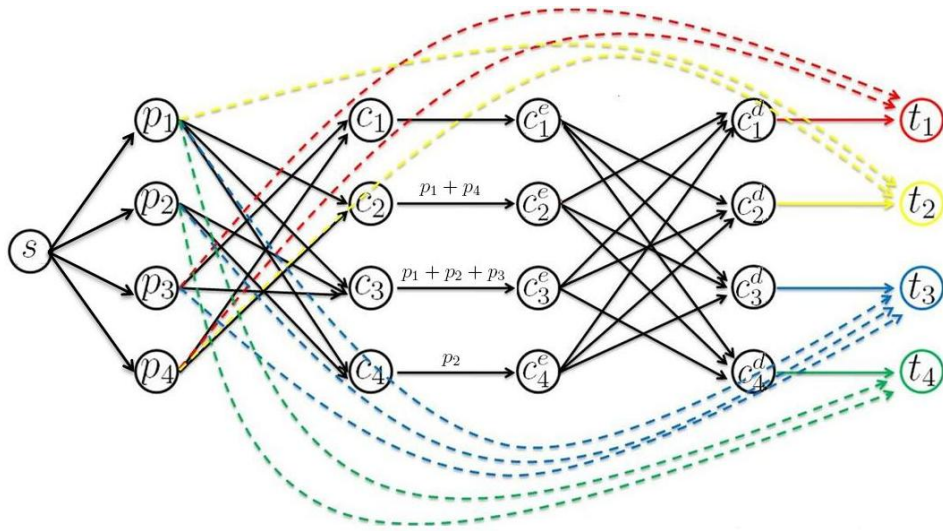


Fig. 19. The Graph Representation of Problem CDE.

Figure 19 demonstrates the graph representation of the example shown in the Figure 16. In this graph  $G = (V, E)$ , there are four sources which are the packets in the set  $P = \{p_1, p_2, p_3, p_4\}$ . These packets are connected to client nodes according to the set of packets available at each client; if there is an edge between  $p_i$  and  $c_j$  then  $p_i \in H(c_j)$ . For example, first client  $c_1$  has the packets  $\{p_3, p_4\}$  so there should be edges going from  $p_3$  and  $p_4$  to  $c_1$ . In the figure, it is easy to see that this construction is also true for other clients. After the encoding in the clients, the coded packets will be

forwarded to the other clients. In the graph  $\{c_1^e, c_2^e, c_3^e, c_4^e\}$  are the nodes representing the transmitter part of the clients in which the encoded packets are broadcasted to the other clients. On the other hand,  $\{c_1^d, c_2^d, c_3^d, c_4^d\}$  are the nodes representing the receiver part of the clients such that the transmitted packets are received for decoding. Since each client  $c_i$  broadcasts its coded(or uncoded) packet(s); there are edges between  $c_i^e$  and  $c_j^d$  where  $i \neq j$ . For the decoding part, we have again for sinks  $\{t_1, t_2, t_3, t_4\}$  each will be able to obtain all the packets except for the ones it has using the received coded packets and the packets they have. Therefore; as shown in the figure for a client  $c_i$  there is a sink  $t_i$  which is connected to  $c_i^d$  and some source nodes in  $\{p_1, p_2, p_3, p_4\}$  specified by its *side information* set  $H(c_i) \subseteq P$ . In the graph, the flow of the solution of the problem is also shown. If each sink is examined, it is easy to see that incoming flows of each sink is equal or greater than the total number of packets, which is four for this specific example, meaning that each client could recover all the packets. Indeed, in [22], it is shown that this solution is optimal for  $n$  clients and  $m$  packets case meaning that this flow is the max-flow with a value of  $|P|$  for each client.

On the other hand, Figure 20 demonstrates the graph representation of the example shown in the Figure 17. In this graph  $G' = (V', E')$ , again there are four sources and sinks in the graph and up to encoding part of the packets in the clients will be exactly the same. However, now since there is a server which is responsible for the communication between clients, the encoded packets will be picked up by the server node  $\tau^e$  and linear combinations of the packet combinations will be broadcasted from the node  $\tau^d$  to the receiver nodes of the clients;  $\{c_1^d, c_2^d, c_3^d, c_4^d\}$ . Then, the decoding part will again be same as the graph  $G$ . In fact, if the graph is examined, it can be verified that with the flow values given we have the optimal solution for each sink which has max-flow value of four which is the total number of packets available. Now,

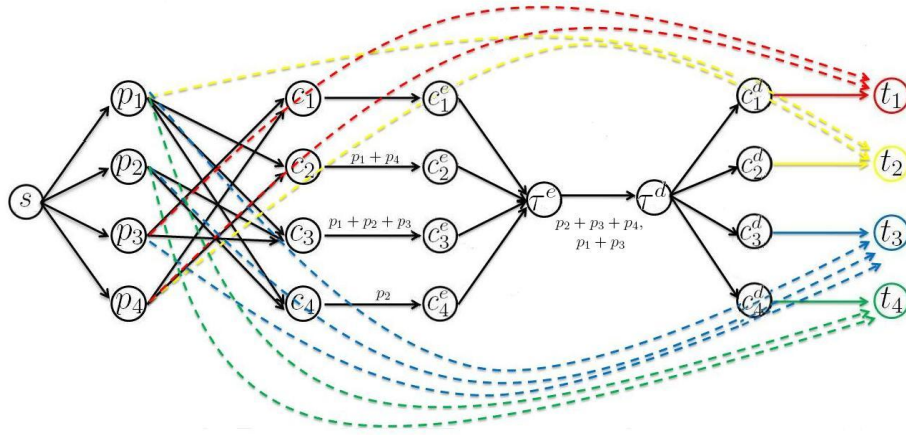


Fig. 20. The Graph Representation of Problem DETR.

our goal is to prove that this can be generalized for  $n$  clients and  $m$  packets case such that the graph  $G'$ , which is the graph representation of Problem DETR, has also a max-flow value of  $|P| = m$  for each client so that the solution is optimal using the equivalence of the graphs  $G$  and  $G'$ .

In general we can construct the multicast graph of the CDE problem with edge capacities  $\{b_1, \dots, b_n\}$  for the transmissions of the clients. Each edge capacity  $b_j$  represents the number of transmissions the client  $c_j$  makes which is found by the algorithm explained in [22]. Finally, all clients will get a total number of  $b = \sum_{j=1}^n b_j$  transmissions letting each edge between  $c_j^d$  to  $t_j$  as  $b$ . Indeed, a client  $c_j$  is interested in all  $m$  packets but it has some packets as side information, which can be represented as direct edges from the packet nodes. In [22] it is shown that it is possible to find a network coding solution to the problem with a field size  $|F_q| \geq n$  for this standard multicast problem using the results in [23]. In other words, it is possible to find  $m$  disjoint paths for each  $t_j$ . Since each  $t_j$  has  $m_j$  packets as a side information which are already represented by disjoint paths, then for the client  $c_j$  it will be enough to

get only  $m - m_j$  packets. Since the minimum number of packets among all the clients have is  $m_{min}$ , then  $b$  should be greater than or equal to  $m - m_{min}$ . Indeed, for DETR problem, rather than a total of  $b$  transmissions the base station will make  $m - m_{min}$  transmissions which is a linear combination of the transmission vectors from all the clients determined the algorithm in [13]. Indeed, this will be just to decrease the number of downlink transmissions without disturbing the multicast problem.

More formally, let  $Q$  is the solution matrix for the CDE problem obtained from the randomized algorithm in [22], i.e.,  $Q = \sum_{i=1}^b \sum_{j=1}^m q_{ij}$  where  $q_{ij}$  is 1 if in the  $i^{th}$  transmission the packet  $m_j$  is transmitted, 0 otherwise. Then, it is possible to obtain a new matrix  $Z$  which will be the solution matrix for the downlink transmissions of the DETR problem in which the number of total transmissions is decreased to  $m - m_{min}$ . In other words,  $Z = \sum_{i=1}^{m-m_{min}} \sum_{j=1}^m q_{ij}$  where  $q_{ij}$  is 1 if in the  $i^{th}$  downlink transmission the packet  $m_j$  is transmitted, 0 otherwise. Indeed, the matrix  $Z$  can be obtained from the row vectors of  $Q_1, \dots, Q_b$  using a proper conversion matrix  $\xi$ , i.e.,  $Z^{(m-m_{min}) \times m} = \xi^{(m-m_{min}) \times b} \cdot Q^{b \times m}$ .

In the example shown in Figure 17, the matrix for the downlink transmissions of DETR problem can be obtained from the solution matrix of CDE as;

$$Z = \xi \cdot Q = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad (4.1)$$

## CHAPTER V

### CONCLUSIONS

In this chapter, we summarize the work discussed in the thesis and suggest possible future research topics.

#### A. Summary of the Thesis

To summarize, this thesis has considered direct information exchange problems in which the clients can communicate in a cooperative manner to satisfy efficient and reliable information exchange utilizing opportunistic listening technique.

First, the Index Coding problem has been examined to understand the fundamental model of direct information exchange problems with opportunistic listening. Then, the efficiency of information exchange has been considered in networks where clients are allowed to communicate directly with each other. Cooperative Data Exchange (CDE) and Data Exchange with Costs (DEWC) Problems have been introduced. Randomized and deterministic algorithms have been presented to ensure efficient information exchange between the clients in Data Exchange with Costs (DEWC) Problem. The performance analyses of the algorithms have also corroborated with simulation results.

Next, the efficiency and reliability of information exchange have been considered in networks where clients are not allowed to communicate directly with each other. Data exchange through relay (DETR) has been modeled. Data Exchange through Relay (DETR) Problem is studied and utilizing the network information flow concept the equivalence of the problem with the combination of Cooperative Data Exchange (CDE) and Index Coding problems has been shown.



## B. Future Works

There are numerous directions for future research work. One direction is to explore the Data Exchange with Costs (DEWC) problem when the clients require some selective packets rather than all the packets. Another direction is to consider direct wireless exchange problems with lossy channels. All the problems discussed in this thesis assume that the transmissions are error-free. However, in general, the channels could be lossy. Therefore, the extension of these problems to a lossy environment can be considered as an interesting future work.

## REFERENCES

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, “Network information flow,” *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.
- [2] R. Yeung, SY Li, and N. Cai, *Network coding theory*, Boston: Now Publishers, 2006.
- [3] T. Ho and D. Lun, *Network coding: an introduction*, Cambridge: Cambridge University Press, 2008.
- [4] C. Fragouli and E. Soljanin, *Network coding fundamentals*, Hanover: Now Publishers Inc, 2007.
- [5] C. Fragouli and E. Soljanin, *Network coding applications*, Hanover: Now Publishers Inc, 2008.
- [6] S. Katti, D. Katabi, W. Hu, H. S. Rahul, and M. Médard, “The importance of being opportunistic: Practical network coding for wireless environments,” in *Proc. 43rd. Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, September 2005.
- [7] S. Katti, H. Rahul, D. Katabi, W. Hu M. Médard, and J. Crowcroft, “XORs in the air: Practical wireless network coding,” in *ACM SIGCOMM*, Pisa, Italy, September 2006.
- [8] Ziv Bar-Yossef, Yitzhak Birk, T. S. Jayram, and Tomer Kol, “Index coding with side information,” in *Proc. of 47th. Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, Berkeley, CA, USA, October 2006, pp. 197–206, IEEE Computer Society.

- [9] A. El Gamal H. El Gamal M. Franceschetti M. Gastpar G. Kramer, R. Berry and J. N. Laneman, “Introduction to the special issue on models, theory, and codes for relaying and cooperation in communication networks [Guest editorial],” *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3297–3301, October 2007.
- [10] Y.W. Hong, W.J. Huang, F.H. Chiu, and C.C. J. Kuo, “Cooperative communications in resource-constrained wireless networks,” *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 47–57, May 2007.
- [11] E. Erkip, A. Sendonaris, A. Stefanov, and B. Aazhang, “Cooperative Communications in Wireless Systems,” in *Advances in network information theory: DIMACS Workshop Network Information Theory, March 17-19, 2003, Piscataway, New Jersey*. Amer Mathematical Society, 2004, pp. 303–320, AMS DIMACS Series.
- [12] A. Sendonaris, E. Erkip, and B. Aazhang, “User cooperation diversity–Part I: System description,” *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.
- [13] S. El Rouayheb, A. Sprintson, and P. Sadeghi, “On coding for cooperative data exchange,” in *Proc. IEEE Information Theory Workshop (ITW) (to appear)*, Cairo, Egypt, Feb. 2010.
- [14] E. Lubetzky and U. Stav, “Non-linear index coding outperforming the linear optimum,” Los Alamitos, CA, USA, 2007, pp. 161–168, IEEE Computer Society.
- [15] M.A.R. Chaudhry and A. Sprintson, “Efficient algorithms for index coding,” in *Proc. of Student Workshop, IEEE Conference on Computer Communications (INFOCOM)*, Phoenix, AZ, USA, April 2008, pp. 1–4.

- [16] M. Langberg and A. Sprintson, “On the hardness of approximating the network coding capacity,” in *IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, July 2008, pp. 315–319.
- [17] S.Y. El Rouayheb, M.A.R. Chaudhry, and A. Sprintson, “On the minimum number of transmissions in single-hop wireless coding networks,” in *IEEE Information Theory Workshop (ITW)*, Lake Tahoe, 2007, pp. 120–125.
- [18] Y. Birk and T. Kol, “Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients,” *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2825–2830, 2006.
- [19] S. El Rouayheb, A. Sprintson, and C. Georghiades, “On the relation between the index coding and the network coding problems,” in *IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, July 2008, pp. 1823–1827.
- [20] S. El Rouayheb, A. Sprintson, and C. N. Georghiades, “A new construction method for networks from matroids,” in *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009, pp. 2872–2876.
- [21] S.Y. El Rouayheb, A. Sprintson, and C.N. Georghiades, “On the Index Coding Problem and Its Relation to Network Coding and Matroid Theory,” *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3187–3195, 2010.
- [22] A. Sprintson, P. Sadeghi, S. El Rouayheb, and G. Booker, “Cooperation with side information: The data exchange problem,” *IEEE Trans. Inform. Theory* (under review), April 2010.

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

## VITA

Damla Ozgul was born in Tekirdag, Turkey. After obtaining a B.S. degree in Electrical and Electronics Engineering from Middle East Technical University, Ankara, Turkey, she came to Texas A&M University to pursue a M.Sc. degree in Electrical Engineering.

Ms. Ozgul can be reached at 232 Zachry Engineering Center, TAMU, College Station, TX 77843-3128. Her email address is [damlaozgul@gmail.com](mailto:damlaozgul@gmail.com).

The typist for this thesis was Damla Ozgul.