An Algorithm for Cooperative Data Exchange with Cost Criterion

Damla Ozgul and Alex Sprintson
Department of Electrical and Computer Engineering
Texas A&M University, College Station, TX 77843
{damlaozgul, spalex}@tamu.edu

***Abstract*—We consider the problem of minimizing the cost of cooperative data exchange between a group of wireless clients. In this problem, a group of clients needs to exchange a set of packets over a shared lossless broadcast channel. Each client initially holds a subset of packets and needs to obtain the packets held by other clients. At each round, one of the clients can broadcast its packets or a combination thereof over the channel. Each client is associated with a certain transmission cost that captures its ability to transmit packets. Such cost can depend on various factors, e.g., available battery life. In this paper, we present an efficient randomized algorithm that ensures that all clients receive all packets and minimizes the total transmission cost. We prove the optimality of the algorithm and perform simulation studies to estimate the advantage achievable by using the network coding technique.**

I. Introduction

Recently, cellular networks are experiencing a growing demand for bandwidth intensive services, resulting in large volumes of voice and data traffic. As a result, there is a growing interest in cooperative communication and direct information exchange between mobile clients (see e.g., [1]–[5] and references therein).

Recently, Sprintson et al. [6] have introduced the problem of direct information exchange between a group of wireless clients. In this problem, a group of clients would like to exchange a set of packets over a shared lossless broadcast channel. Each client initially holds a subset of packets and needs to obtain all the packets held by other clients. At each round, one of the clients can broadcast its packets or a combination thereof over a noiseless broadcast channel of capacity one packet per channel use. Assuming that clients can cooperate with each other and know which packets are available to other nodes, the aim is to minimize the total number of transmissions needed to satisfy the demands of all clients.

In this paper, we consider a more general version of the problem that minimizes the overall transmission cost. In this version, each client is associated with a cost of transmitting a single packet. The transmission costs allow us to differentiate between different clients. For example, a client which has small residual battery level can be assigned a high cost, whereas a client with a fully charged battery will be given a low cost.

The work of Alex Sprintson was supported by NSF grant CNS-0954153 and by Qatar Telecom (Qtel), Doha, Qatar.

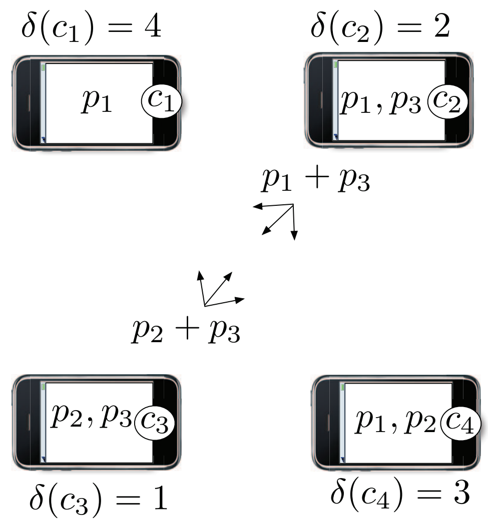

Fig. 1. Optimal Transmission Scheme.

Figure 1 illustrates the problem considered in this paper. In this picture we have four wireless clients $c_1, \ldots, c_4$ that would like to obtain three packets $p_1, \ldots, p_3$. Each client initially has a subset of packets. Specifically, client c_1 has p_1, client c_2 has p_1 and p_3, client c_3 has p_2 and p_3, and client c_4 has p_1 and p_2. Each client c_i is associated with a transmission cost $\delta(c_i)$, as depicted in the figure. Without network coding, each packet p_1, p_2, and p_3 needs to be transmitted, resulted in a total transmission cost of at least four (achieved if c_3 broadcasts p_2 and p_3 and c_2 broadcasts p_1). By using network coding, the total transmission cost can be reduced to 3. Indeed, consider a solution in which client c_3 broadcasts a bitwise "exclusive-OR" of packets p_2 and p_3 and client c_2 broadcasts a bitwise "exclusive-OR" of packets p_1 and p_3. It can be easily verified that after these two transmissions all four clients will be able to decode all three packets.

In this paper, we present an efficient algorithm for the cooperative data exchange problem with costs and prove its optimality. We also verify the effectiveness of the network coding approach through extensive simulations.

Related work. A closely-related problem is Index Coding Problem [7]–[11] in which the clients receive transmissions from a server node and try to obtain the required packets.

References [6] and [12] present two algorithms for the direct information exchange problem (without costs). The first algorithm, presented in [6], is based on a randomization technique and the second algorithm, presented in [12], uses a deterministic approach.

The rest of the paper is organized as follows. In Section II, we present the formal definition of the problem. Then, in Section III we present our randomized algorithm. The performance of the algorithm is analyzed in Section IV. The results of our simulation study are presented in Section V. Finally, conclusions and directions for future work are presented in Section VI.

II. Model

Our problem setup includes a set $C = \{c_1, \ldots, c_k\}$ of k wireless clients that need to obtain a set $P = \{p_1, \ldots, p_n\}$ of n packets. We assume that each packet is an element of a finite field $\mathbf{F}$ of size q. Initially, each client has an access to a subset $H(c_i) \subseteq P$ of packets. We refer to set $H(c_i)$ as a "side information" or a "has" set of client c_i. The clients collectively know packets in P, i.e., $\cup_{c_i \in C} H(c_i) = P$. The set of required packets, or a *demand* set of the client c_i is denoted as $W(c_i) = \bar{H(c_i)} = P \setminus H(c_i) \subseteq P$. We assume that each client knows the indices of packets that are available to other clients.

The clients use a lossless broadcast channel to obtain all packets in P. Each client $c_i \in C$ is associated with a transmission cost $\delta(c_i)$. The information packets are transmitted in communication rounds, such that at round j one of the clients, denoted by c_{i_j}, broadcasts a packet, $p^j \in \mathbf{F}$, to other clients in C. Packet p^j can be one of the packets in $H(c_{i_j})$, or a combination of the packets in $H(c_{i_j})$. We use linear coding over the field $\mathbf{F}$, i.e., each packet is an element of a finite field $\mathbf{F}$ and all coding operations are also linear over $\mathbf{F}$.

Our goal is to find a transmission scheme that satisfies the following two conditions:

1) Each client $c_i \in C$ is able to decode all packets in P from the packets in its has set $H(c_i)$ and the packets transmitted over the channel $p^1, \ldots, p^T$;
2) The total transmission cost $\sum_{j=1}^{T} \delta(c_{i_j})$ is minimal among all the schemes that satisfy the first condition.

Here, T is the total number of transmission required by the scheme.

We denote by $n_i = |H(c_i)|$ the number of packets originally available at client c_i. We also denote by $n_{min} = \min_{1 \le i \le k} n_i$ the minimum number of packets available to a client.

We say that a client c_i has a unique packet p_j if $p_j \in H(c_i)$ but $p_j \notin H(c_l)$ for any other client c_l. It can be verified that a unique packet can be broadcasted by the client that holds it without mixing with other packets without any penalty in terms of optimality. Therefore, without loss of generality, we can assume that are no unique packets in the system.

In general, c_{i_j} can also use packets in its "has" set together with the packets $p^1, \ldots, p^{j-1}$ previously transmitted over the channel. However, it is easy to verify that this will not help to minimize the total transmission cost.

III. Randomized Algorithm

In this section, we present an efficient randomized algorithm for the problem at hand. The algorithm provides a minimum cost solution with high probability, assuming that the underlying finite field $\mathbf{F}$ is sufficiently large.

Since our algorithm uses linear coding over field $\mathbf{F}$, every packet p^j transmitted by the algorithm is a linear combination of the original packets in P:

$$p^j = \sum_{p_i \in P} \alpha_i^j p_i, \tag{1}$$

where $\alpha_i^j \in \mathbf{F}$ are encoding coefficients.

In the description and analysis of our algorithm we refer to packets by their corresponding 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 say that the client transmits the encoding vector of the packet p^j, $\alpha^j = \left\{\alpha_1^j, \ldots, \alpha_n^j\right\}$.

Let $u_i = \left\{u_i^1, \ldots, u_i^n\right\}$ be the encoding vector that corresponds to a packet $p_i \in P$, where $u_i^i = 1$ and $u_i^j = 0$ for $i \neq j$. Also, let $U(c_i)$ the set of unit vectors that correspond to packets in $H(c_i)$. Our algorithm employs *random linear coding*. That is, each transmitted vector α^j is a random linear combination of the unit vectors in $U(c_{i_j})$. This implies that $\alpha_i^j = 0$ if $p_i \notin H(c_{i_j})$. Other elements of α^j are selected at random from the field $\mathbf{F}$. Then, the set of encoding vectors that have been transmitted up to and including round j can be expressed as $A_j = \left\{\alpha^1, \ldots, \alpha^j\right\}$.

The key decision that our algorithm needs to make is to determine which client transmits a packet at each communication round. Since our goal is to minimize the total transmission cost, we would like to select clients with low transmission costs. However, we also need to make sure that the selected clients have sufficient side information to satisfy other clients.

The key idea of our algorithm is to determine number of transmissions T required to complete the information transfer at minimum total cost. Note that this number might be greater than the minimum number of transmissions. Our algorithm determines the value of T through exhaustive search. That is, the algorithm tries all possible value of T in the range $[n - n_{min}, n]$ and selects the one that yields minimum overall cost. Indeed, the number of transmission made by any feasible algorithm is lower bounded by $n - n_{min}$ and upper bounded by n.

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:

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

where $A_{j-1} = \left\{\alpha^1, \ldots, \alpha^{i-1}\right\}$ is the set containing the packets that have been transmitted so far. Note that $n - n_i^j$ is the 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 contains clients that require $T-(j-1)$ packets at iteration j, i.e.,

$$C_1^j = \left\{ c_i \in C |\; n - n_i^j = T-(j-1) \right\};$$

- Set C_2^j contains clients that require less than $T-(j-1)$ packets at iteration j, i.e.,

$$C_2^j = \left\{ c_i \in C |\; n - n_i^j < T-(j-1) \right\}.$$

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

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

IV. Correctness Analysis

We proceed to analyze the correctness and optimality of the algorithm. Consider an iteration j of the algorithm. We denote 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. In other words, in addition to the first j transmissions, at most $T-j$ transmissions of total cost OPT_j are needed to satisfy the demands of all clients.

Lemma 1: With a probability at least $1-\frac{k}{q}$, where q is the size of the finite field $\mathbf{F}$.

$$\text{OPT}_j = \text{OPT}_{j-1} - \delta(c_{i_j}). \tag{2}$$

Proof: Let Ω_{j-1} be an optimal set of encoding vectors which are necessary to complete data transfer. In other words, Ω_{j-1} has $T-(j-1)$ encoding vectors such that:

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

Assume that c_{i_j} is the client which has a minimum cost among the set C_2^j. Let $\mu^j = \text{rank}\left(U(c_{i_j}) \cup A_{j-1}\right)$ be the rank of the set of encoding vectors which a client c_{i_j} has at the beginning of iteration j. Note that $T-(j-1) > T - \mu^j$. This follows from the fact that c_{i_j} belongs to the set C_2^j. Therefore, we can remove at least one packet, say v, from Ω_{j-1} so that $\tilde{\Omega}_{j-1} = \Omega_{j-1} \setminus \{v\}$ satisfies $A_{j-1} \cup \tilde{\Omega}_{j-1} \cup U(c_{i_j})$ is of rank n.

Let c_i be an arbitrary a client in $C \setminus \left\{c_{i_j}\right\}$. 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 $n-1$. Note also that it is sufficient to focus on the case in which the rank is equal to $n-1$.

We denote γ_i as the normal vector to the span of S_i which can be written as

$$\gamma_i = \sum_{u_g \in U(c_{i_j})} \beta_g u_g + \sum_{u_g \in \bar{U}(c_{i_j})} \beta_g u_g, \tag{3}$$

where $\bar{U}(c_{i_j})$ is the set of unit encoding vectors that correspond to $W(c_{i_j}) = P \setminus H(c_{i_j})$. If we show that γ_i and α^j are not orthogonal with high probability, then we prove the claim that $A_{i-1} \cup \tilde{\Omega}_{i-1} \cup U(c_i) \cup \{\alpha^j\}$ is of rank n with high probability. In other words, to prove the claim it will suffice to show that the inner product $\left\langle \gamma_i, \alpha^j \right\rangle$ between γ_i and α^j is not equal to zero with probability at least $1-\frac{k}{q}$.

First, we show that there exists $u_g \in U(c_{i_j})$ such that $\beta_g \neq 0$. Indeed, if this is not the case, then we can write γ_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 γ_i. However, since this means that β_g is equal to zero for each $u_g \in \bar{U}(c_{i_j})$, we will have a contradiction with the fact that γ_i is not identical to zero.

We can write the inner product $\left\langle \gamma_i, \alpha^j \right\rangle$ as

$$\left\langle \gamma_i, \alpha^j \right\rangle = \sum_{u_g \in U(c_{i_j})} \alpha_g^j \beta_g, \tag{4}$$

since α^j is a random linear combination of vectors in $U(c_{i_j})$, i.e., $\alpha^j = \sum_{u_g \in U(c_{i_j})} \alpha_g^j u_g$ where α_g^j are random coefficients over a field $\mathbf{F}$. Let $\hat{U}$ be a subset of $U(c_{i_j})$ such that for each $u_g \in \hat{U}$ it holds that $\beta_g \neq 0$. Since there exists $u_g \in U(c_{i_j})$ such that $\beta_g \neq 0$ the set $\hat{U}$ is not empty and so, $\left\langle \gamma_i, \alpha^j \right\rangle = \sum_{u_g \in \hat{U}} \alpha_g^j \beta_g$. Since for each $u_g \in \hat{U}$, α_g^j is a random variable chosen independently of $\left\{\beta_g u_g \in \hat{U}\right\}$ the probability that $\left\langle \gamma_i, \alpha^j \right\rangle$ is equal to zero is at most $\frac{1}{q}$.

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

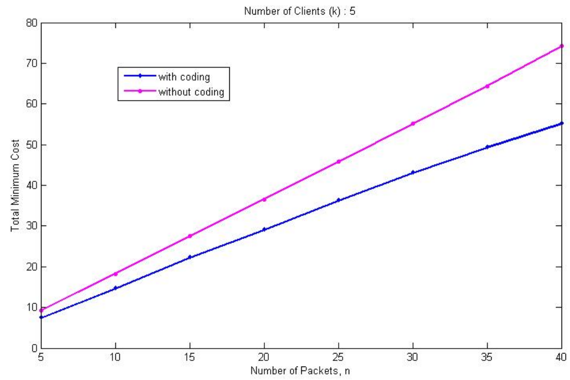

Fig. 2. Performance Comparison for Data Exchange with Costs Problem.

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 $\text{OPT}_j = \text{OPT}_{j-1} - \delta(c_{i_j})$ and the lemma follows.

■

Lemma 1 implies that at each interruption j it holds that $\text{OPT}_j = \text{OPT}_{j-1} - \delta(c_{i_j})$ with probability at least at least $1 - \frac{k}{q}$. The condition will hold for all T iterations with probability at least

$$\left(1 - \frac{k}{q}\right)^T \geq \left(1 - \frac{k}{q}\right)^n \geq 1 - \frac{nk}{q} \qquad (5)$$

provided that the size q of the underlying field $\mathbf{F}$ is larger than n.

The probability of success can be increased by performing multiple iterations and by averaging the total minimum cost values among all the iterations.

V. Numerical Results

In this section, we verify the performance of the algorithm presented in Section III. We have compared the minimum total cost value without coding and the minimum total cost value obtained by the algorithm. Figure 2 shows numerical results of minimum total cost values for $k = 5$ clients and $n = (5, 10, \ldots, 40)$ packets. In our experiments all clients have different cost values starting from 1 to k.

In our simulation study, total minimum cost values for both traditional approach and network coding are calculated by averaging over 100 random initializations of the problem. In traditional approach, for each packet, $p_j \in P$ we select a client c_{i_j} that has p_j and whose transmission cost is minimal among all clients that have p_j. The total cost of transmissions with the traditional approach is compared with the total minimum transmission cost of the randomized algorithm. The performance comparison is shown in Figure 2.

In Figure 2 the upper line represent total minimum cost values for traditional approach; on the other hand the lower line shows the total minimum cost values when the network coding technique is utilized. Remarkably, the coding algorithm performs better than traditional approach for all the cases.

VI. Conclusions

In this paper, we considered the problem of cooperative data exchange between wireless clients each having an associated cost value and presented a randomized algorithm which minimizes the total transmission cost with high probability. We also analyzed the performance of the algorithm. Empirical results show that the randomized algorithm performs better than traditional approach that do not utilize network coding. As a future work we plan to work on a deterministic algorithm for this work. In addition, there are many interesting problems for future research. For example, we can consider additional constraints, such as the upper bounds on the number of transmissions that can be made by each client.

References

[1] T. Korakis, M. Knox, E. Erkip, and S. Panwar. Cooperative network implementation using open-source platforms. *IEEE Communications Magazine*, 47(2):134 –141, February 2009.

[2] A. Scaglione, D. L. Goeckel, and J. N. Laneman. Cooperative communications in mobile ad hoc networks. *IEEE Signal Processing Magazine*, 23(5):18–29, September 2006.

[3] 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*, 24(3):47–57, May 2007.

[4] A. Stefanov E. Erkip, A. Sendonaris and B. Aazhang. *Cooperative Communication in Wireless Systems*, pages 303–320. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 2004.

[5] A. Sendonaris, E. Erkip, and B. Aazhang. User Cooperation Diversity–Part I: System Description. *IEEE Transactions on Communications*, 51(11):1927–1938, November 2003.

[6] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb. A Randomized Algorithm and Performance Bounds for Coded Cooperative Data Exchange. *Proc. of IEEE International Symposium on Information Theory (ISIT'10)*, 2010.

[7] Y. Birk and T. Kol. Coding-on-demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients. *IEEE Transactions on Infromation Theory*, 52(6):2825–2830, June 2006.

[8] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Ko. Index coding with side information. In *Proc. of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 197–206, 2006.

[9] S. El Rouayheb, A. Sprintson, and C. N. Georghiades. On the relation between the index coding and the network coding problems. *Proc. of IEEE International Symposium on Information Theory (ISIT08)*, 2008.

[10] S. El Rouayheb, A. Sprintson, and C. N. Georghiades. A new construction method for networks from matroids. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009.

[11] S. El Rouayheb, A. Sprintson, and C. N. Georghiades. On the index coding problem and its relation to network coding and matroid theory. submitted to IEEE Transactions on Information Theory, 2009.

[12] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb. Deterministic algorithm for coded cooperative data exchange. In *Proceedings of the 7th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine 2010)*, Houston, Texas, November 2010.