# Minimum-Cost Subgraphs for Joint Distributed Source and Network Coding

Anna Lee and Muriel Médard
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: {annalee, medard}@mit.edu

Karen Zita Haigh, Sharon Gowan and Paul Rubel
BBN Technologies
10 Moulton St.
Cambridge, MA 02138
email: {khaigh, sgowan, prubel}@bbn.com

***Abstract*— We consider multicast of correlated sources over a network. Assuming the use of random network coding, we provide a linear optimization formulation for allocation of link rates in the network, also known as subgraph construction. Such an approach requires joint distributed source and network coding, which often has a lower cost than of that required by separated source and network coding. We support this result with simulations on randomly generated networks and on network data collected from a Future Combat Systems (FCS) exercise at Lakehurst, NJ.**

## I. Introduction

Network coding, the notion that packets traversing a network can be combined and mixed rather than merely forwarded, has garnered much interest since its inception by Ahlswede et al. [1]. In particular, network coding is optimal for multicast. The use of network coding for multicast can be decoupled into two independent components, minimum-cost subgraph construction [9] and random linear network coding [6]. Subgraph construction is the selection of network resources by choosing links and corresponding flow rates to support a multicast connection, preferably by minimizing some cost such as energy or latency. Once the subgraph is established, nodes produce coded packets at their assigned rates by transmitting random combinations of their incoming packets. In this paper we examine the first problem, minimum-cost subgraph construction for multicasting of correlated sources.

### A. Overview and Related Work

Ho et al. [7] showed that random network coding can be used to multicast correlated sources, and moreover generalized the error exponents for linear Slepian-Wolf coding [4]. In this setting, compression may occur within the network, resulting in joint distributed source and network coding. We complement this approach with a corresponding subgraph construction algorithm by adapting the approach presented in Lun et al. [9], described in section II.

Complexity concerns over the joint coding approach of Ho et al. [7] motivated Ramamoorthy et al. [11] to study the separation of source and network coding. They defined a ”price of separation” to quantify the gap between joint coding and separate source and network coding, and showed that the two-source, two receiver connection is always separable. Their experimental results focused on networks with capacity constraints, and showed that separation held in all their test cases. Our simulations focus on cost constraints, rather than capacity constraints. In this situation, a separable solution can always be found, but in general has a higher cost than a joint coding solution. In section III-A we present results for randomly generated networks that highlight that difference in cost. We also show that the benefit from joint coding is substantial for existing networks, using data from Future Combat Systems (FCS), in section III-B.

## II. Problem Formulation

In this section we present the linear optimization formulation for calculating the minimum-cost subgraph for two sources. We assume that each link has a cost linearly proportional to the rate. Suppose we are given the following inputs:

- a graph $G = (N, A)$
- edge weights $w_{ij} : (i, j) \in A \rightarrow R^+$
- edge capacities $c_{ij} : (i, j) \in A \rightarrow R^+$
- set of two source nodes $S$
- sources $X_i$ generated at $s_i$ with rate $H(X_i)$
- joint rate $H(X_1, X_2)$
- set of receiver nodes $T$.

The goal is to transmit at the joint rate $H(X_1, X_2)$ to the all the receivers in $T$. We augment the graph by adding a virtual source $s^*$ and a virtual edge from $s^*$ to each real source. The capacities of the virtual edges are set to the marginal entropy of the corresponding source. The overall rate from the virtual source to each of the receivers is set to the joint entropy of the sources. These conditions ensure that the flow rates from the sources to each receiver meet the Slepian-Wolf constraints for distributed source coding. Figure 1 illustrates how the graph is augmented.

We give the formal definition of the problem.
Let

$$\begin{aligned} G^* &= (N^*, A^*) \\ N^* &= N \cup s^* \\ A^* &= A \cup E, \end{aligned}$$

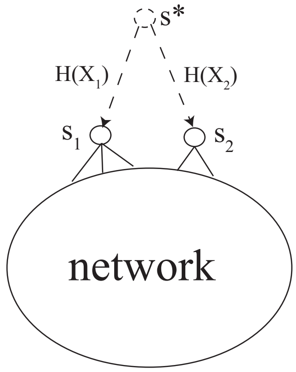



Fig. 1. Problem Formulation

where

$$\begin{aligned} E &= \{(s^*, j) | j \in S\} \\ c_{s^*j} &= H(X_j) \\ w_{s^*j} &= 0. \end{aligned}$$

Also, let $R = H(X_1, X_2)$.

The desired minimum-cost subgraph is found using the following linear optimization problem.
Minimize

$$\sum_{(i,j)\in A^*} w_{ij} z_{ij},$$

subject to

$$c_{ij} \geq z_{ij}, \qquad \forall (i,j) \in A, \tag{1}$$

$$c_{ij} \geq x_{ij}, \qquad \forall (i,j) \in E, \tag{2}$$

$$z_{ij} \geq x_{ij}^{(t)} \geq 0, \qquad \forall (i,j) \in A^*, t \in T, \tag{3}$$

$$\sum_{\{j|(i,j)\in A^*\}} x_{ij}^{(t)} - \sum_{\{j|(j,i)\in A^*\}} x_{ji}^{(t)} = \sigma_i^{(t)}, \quad \forall i \in N^*,\ t \in T, \tag{4}$$

where

$$\sigma_i^{(t)} = \begin{cases} R & i = s^*, \\ -R & i = t, \\ 0 & \text{otherwise.} \end{cases}$$

The desired subgraph is

$$\{z_{ij} | (i,j) \in A\}$$

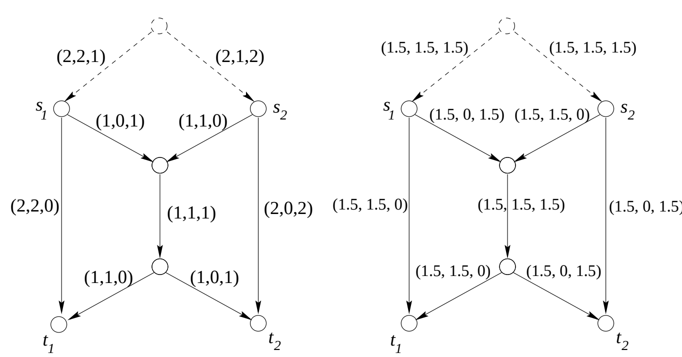
and the subgraph cost is

$$\sum_{(i,j)\in A} w_{ij} z_{ij}.$$

$\{z_{ij} | (i,j) \in A\}$ are the actual rates that are to be used by the random code in the network. (1) upper bounds $z_{ij}$ by the capacities $c_{ij}$ (it is not necessary to consider the virtual links). $\{z_{ij}\}$ must support a network flow of rate $R$ from the virtual source to each receiver. $\{x_{ij}^{(t)}\}$ represents the actual flow from the virtual source to receiver $t \in T$. (4) creates the flow constraints $R$ for each receiver. (2) ensures that the flow supplied by each real source does not exceed the corresponding marginal entropy. (3) upper bounds $\{x_{ij}\}$ for each receiver by $\{z_{ij}\}$, to ensure that the $\{z_{ij}\}$ supports a flow of rate $R$ for each receiver.

The rates $z_{s^*s_i}$ on the two virtual links are the rates to which each source individually compresses outside the network. The sum of the source rates may equal or exceed the joint entropy $R$. Recall that the rate at each receiver is also $R$. Therefore if the sum of the source rates exceeds $R$, then the random network code also performs compression, and thus joint source-network decoding is necessary (minimum entropy decoding [5], for example). If the sum of the source rates is equal to $R$, then source coding does not occur within the network. Once the rate links are established by construction of the subgraph, one network code is used for the entire multicast connection. Our notion of separability concerns the decoding stage. In a separable coding situation, all redundancy is removed at the sources, only network coding occurs within the network, and decoding is straightforward. In a joint coding situation, some redundancy is kept, subsequently removed within the network, and joint source-network decoding is required.

Note that the weights for the virtual edges are 0. This encourages the optimization problem to find a solution that results in joint source-network coding. We can force a separable solution by setting the virtual edge weights arbitrarily high (the sum of the edge weights of the original graph, for example). Furthermore, we can easily determine whether a solution is a separable solution or not. If the sum of the flow rates $z_{ij}$ for $z \in E$ (the virtual edges) equals the joint entropy, then the solution is a separable one. Figure 2 gives an example of the subgraph for a two-source, two-receiver network. The marginal entropies of the two sources is 2 and the joint entropy is 3. The links have unit weight and no capacity constraints. Each link is labelled with the triplet $(z_{ij}, x_{ij}^{(1)}, x_{ij}^{(2)})$. The joint solution has a cost of 9, and the separable solution has a cost of 10.5.




(a) Joint Solution (cost = 9) (b) Separable Solution (cost = 10.5)

Fig. 2. Example of Minimum Cost Subgraphs. Each link is labelled with the triplet $(z_{ij}, x_{ij}^{(1)}, x_{ij}^{(2)})$.

This linear optimization problem is equivalent in form to the linear optimization problem in Lun et al. [9]. Thus, the same decentralized techniques described in Lun et al. [9] can also be used for our formulation. It may appear that our formulation does not allow a decentralized approach because of the virtual links that indirectly connect the real sources. But centralized coordination is only needed for the allocation of rates among the sources, which would be necessary in any Slepian-Wolf setting. The actual source-network coding remains distributed.

## III. Simulations

### A. Randomly Generated Networks

We present simulation results for randomly generated networks. The networks were generated in the following manner. $n$ nodes were placed in a box of size $w$ by $h$. 2 source nodes were placed at the top two corners of the box, and 2 receiver nodes were placed at the bottom two corners of the box. The remaining $n-4$ nodes were randomly placed in the box. Nodes within a distance of $r$ were connected by an edge, with edges pointing downwards. All edges had unit weight and infinite capacity. Figure 3 is an example of such a network.

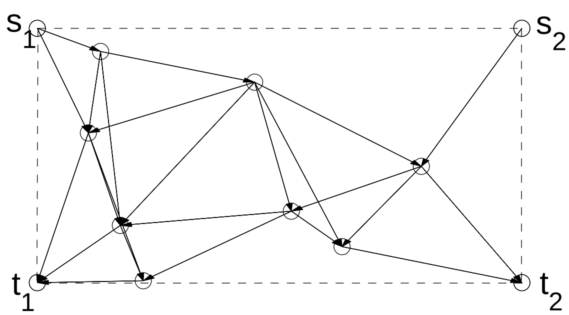



Fig. 3. Example of a Randomly Generated Network

For each network configuration, 1000 feasible graphs were generated, where feasibility meant all sources were connected to all receivers. For each graph, the minimum-cost subgraph was computed. The minimum-cost separable subgraph was also computed for comparison. Table I shows the results for some sample configurations. The two sources had marginal entropies equal to 2. For each configuration, we give the average subgraph cost, the percentage of graphs with a lower joint coding solution subgraph cost than a separable solution subgraph cost, and of those cases, the percent cost saving of the joint coding solution over the separable solution. The overall saving is the product of the first two percentages.

As the network becomes wider, such that the sources (and receivers) are farther apart, a joint coding solution has a significantly lower cost than a separable solution. If the network size is fixed but the correlation between the sources is increased (resulting in a lower joint entropy, and thus, lower rate $R$), the benefit from a joint coding solution also increases. These results validate our intuition. As the network becomes wider, there is more benefit in retaining some redundancy, as it becomes cheaper for each receiver to get its information from the source closer to it. If all redundancy is removed, as in the case for a separable solution, then each receiver must get some information from the farther source, which

TABLE I

Results for Randomly Generated Networks

| h x w | n | r | R | avg. cost | % joint $<$ sep. | % cost diff. | % overall saving |
|---|---|---|---|---|---|---|---|
| 1 x 1 | 12 | 0.6 | 3 | 14.29 | 6.5 | 4.28 | 0.27 |
| | | | 2.5 | 11.66 | 6.6 | 9.8 | 0.65 |
| | | | 2 | 9.15 | 8.9 | 11.88 | 1.06 |
| 1 x 2 | 18 | 0.8 | 3 | 15.06 | 79.5 | 12.52 | 9.95 |
| | | | 2.5 | 11.79 | 79.9 | 12.52 | 10.0 |
| | | | 2 | 8.56 | 78.6 | 20.98 | 16.5 |
| 1 x 3 | 28 | 0.9 | 3 | 16.53 | 99.2 | 10.19 | 10.1 |
| | | | 2.5 | 12.32 | 99.6 | 18.64 | 18.6 |
| | | | 2 | 8.10 | 99.4 | 31.96 | 31.8 |

was originally available at the closer source. It follows that this benefit also increases as the correlation of the sources increases, as more shared information is available at the closer source.

In a two-source, two-receiver network, a separable solution always exists [11]. But the subgraph cost for a separable solution can be no better than that for a joint coding solution. Depending on the network topology, the joint coding subgraph will almost always have a lower cost than the separable coding subgraph.

### B. Future Combat Systems Data

We also collected network management data from a 10-node Future Combat Systems (FCS) exercise at Lakehurst, NJ. The Lakehurst terrain combined with the mobility of the nodes causes connectivity among nodes to change frequently and generally be very unreliable. Figure 4 illustrates the connectivity of the network over the sequence of the first 5000 graphs, where an edge denoted by a solid line indicates that routes were found between that pair of nodes in most of the graphs ($> 70\%$), and dotted lines indicates that routes were found less frequently ($30\% - 55\%$). Routes found less than $< 30\%$ of the time are not shown.

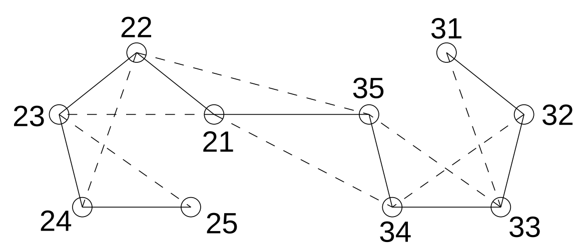



Fig. 4. FCS Network Topology

We calculated the correlation coefficients for each variable in the Management Information Base (MIB), for each pair of nodes in the network. Table II shows the node-to-node correlation coefficients for one particular variable.

To test the effectiveness of network coding, we performed three experiments using the correlations shown in Table II. For each experiment, two sources and two receivers were randomly chosen and the following correlation model was constructed.

| | 21 | 22 | 23 | 24 | 25 | 31 | 32 | 33 | 34 | 35 |
|---|---|---|---|---|---|---|---|---|---|---|
| 21 | 1.00 | 0.95 | 0.95 | 0.95 | 0.33 | 0.80 | 0.85 | 0.80 | 0.80 | 0.95 |
| 22 | 0.95 | 1.00 | 1.00 | 1.00 | 0.45 | 0.69 | 0.94 | 0.69 | 0.68 | 1.00 |
| 23 | 0.95 | 1.00 | 1.00 | 1.00 | 0.45 | 0.70 | 0.94 | 0.69 | 0.69 | 1.00 |
| 24 | 0.95 | 1.00 | 1.00 | 1.00 | 0.44 | 0.69 | 0.94 | 0.69 | 0.68 | 1.00 |
| 25 | 0.33 | 0.45 | 0.45 | 0.44 | 1.00 | 0.20 | 0.42 | 0.20 | 0.20 | 0.45 |
| 31 | 0.80 | 0.69 | 0.70 | 0.69 | 0.20 | 1.00 | 0.60 | 1.00 | 1.00 | 0.69 |
| 32 | 0.85 | 0.94 | 0.94 | 0.94 | 0.42 | 0.60 | 1.00 | 0.60 | 0.60 | 0.94 |
| 33 | 0.80 | 0.69 | 0.69 | 0.69 | 0.20 | 1.00 | 0.60 | 1.00 | 1.00 | 0.68 |
| 34 | 0.80 | 0.68 | 0.69 | 0.68 | 0.20 | 1.00 | 0.60 | 1.00 | 1.00 | 0.68 |
| 35 | 0.95 | 1.00 | 1.00 | 1.00 | 0.45 | 0.69 | 0.94 | 0.68 | 0.68 | 1.00 |

TABLE II

CORRELATION COEFFICIENTS FOR THE VARIABLE MI_F_THF_FFFF__0_ IN THE FCS MIB.

The marginal entropies were set to 1, and the joint entropy was set to $1 + H(\frac{\rho+1}{2})$, where $\rho$ is the correlation coefficient for the two sources. All edges were given unit weight and infinite capacity. The minimum-cost subgraph was computed for each node adjacency graph, and the separable minimum-cost subgraph was also computed for comparison. Table III contains the results for the three sample experiments. For each configuration, we give the percentage of feasible connections, average subgraph cost, the percentage of feasible graphs with a lower joint coding solution subgraph cost than a separable solution subgraph cost, and of those cases, the percent cost saving of the joint coding solution over the separable solution. The overall cost saving over all 5000 graphs is the product of the three percentages.

TABLE III

RESULTS FOR FCS DATA

| sources | 24, 32 | 23, 32 | 22, 32 |
|---|---|---|---|
| receivers | 23, 31 | 25, 31 | 24, 34 |
| $\rho$ | 0.94 | 0.94 | 0.94 |
| feasible connection (%) | 80.50 | 65.76 | 81.72 |
| average cost | 3.30 | 3.05 | 3.90 |
| joint $<$ separable (%) | 58.51 | 30.99 | 52.67 |
| saving of joint over separable (%) | 38.42 | 40.33 | 28.66 |
| overall saving (%) | 18.09 | 8.22 | 12.34 |

## IV. CONCLUSION

We have presented a linear optimization formulation for calculating a minimum-cost subgraph for network coding of correlated sources. This approach results in a joint distributed source and network coding solution. We have shown through simulations that a joint coding approach has a lower subgraph cost than of a separate source and network coding approach.

We propose some directions for further research. One extension of this work is to generalize the problem formulation to handle more than two sources. Another avenue for further research is to design practical codes (codes with low decoding complexity) for joint source network coding. Wu et al. [12] is an example of work in this direction, but for a specific correlation structure. Or perhaps we can extend existing low-complexity schemes for Slepian-Wolf coding [10], [3] to incorporate network coding. We have presented the analysis for a multicast connection with correlated sources. Another interesting problem is, given a network and a set of receivers, determine the best placement of source nodes and correlation structure to obtain the minimum cost subgraph. This can be seen as designing generalized mirror sites, where sources are not restricted to replication of data and transmission by routing. The use of more varied correlation structures and network coding for transmission allows for a richer set of possible solutions. In work related to this problem, Jiang [8] incorporates network coding for jointly minimizing storage and transmission cost, and [2] compares network coding with routing for P2P networks.


## ACKNOWLEDGMENT

This work was supported by the National Reconnaissance Office and DARPA ATO N00014-05-1-0197.


## REFERENCES


[1] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung. “Network Information Flow,” *IEEE Transactions on Information Theory*, 46(4):1204-1216, 2000.

[2] D. M. Chiu, R. W. Yeung, J. Huang and B. Fan. “Can Network Coding Help in P2P Networks?”, *International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, April 2006.

[3] T. P. Coleman, A. H. Lee, M. Médard, M. Effros,“Low-Complexity Approaches to Slepian-Wolf Near-Lossless Distributed Data Compression”, *IEEE Transactions on Information Theory*, 52(8):3546-3561, 2006.

[4] I. Csiszar, “Linear Codes for Sources and Source Networks: Error Exponents, Universal Coding”, *IEEE Transactions on Information Theory*, 28(4):585-592, 1982.

[5] I. Csiszar and J. G. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Inc., Orlando, FL, 1982.

[6] T. Ho, R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger, “Toward a Random Operation of Networks”, *IEEE Transactions on Information Theory*, submitted (2004). http://web.mit.edu/trace/www/itrandom5.pdf

[7] T. Ho, M. Médard, M. Effros and R. Koetter, “Network Coding for Correlated Sources”, Invited Paper, *Conference on Information Sciences and Systems (CISS)*, 2004.

[8] A. Jiang, “Network Coding for Joint Storage and Transmission with Minimum Cost”, *IEEE International Symposium on Information Theory*, Seattle, Washington, July 2006.

[9] D. S. Lun, N. Ratnakar, M. Médard, R. Koetter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, “Minimum-cost Multicast Over Coded Packet Networks”, *IEEE Transactions on Information Theory*, 52(6):2608-2623, 2006.

[10] S. S. Pradhan and K. Ramchandran, ”Distributed Source Coding Using Syndromes (DISCUS): Design and Construction”, *IEEE International Symposium on INformation Theory*, 49(3):626-643, 2003.

[11] A. Ramamoorthy, K. Jain, P. A. Chou, and M. Effros, “Separating Distributed Source Coding from Network Coding”, *IEEE Transactions on Information Theory*, 52(6):2785-2795, 2006.

[12] Y. Wu, V. Stankovic, Z. Xiong, and S.-Y. Kung, “On practical design for joint distributed source and network coding,” *1st Workshop on Network Coding, Theory, and Applications*, Apr. 2005.