On Networks with Side Information

A. Cohen, S. Avestimehr and M. Effros
California Institute of Technology

Motivation
Side information is data available within the network, which is correlated with the main source we wish to transmit. Side information can significantly reduce the rates required throughout the network.

The model is interesting in cases where:
- Not enough capacity from the source.
- Side information is "close" or "less expensive".
- Collaborative networks (sensor, organizational, army).

Problem Statement
- Network is modeled as a directed acyclic graph.
- One source node, one side information node, possibly several terminals requiring the source (with arbitrarily small error).
- Data processing is allowed at all nodes.

How much data do we have to send?
- Goal: compute the amount of data required - the Rate Region

For a network with nodes \( V \), edges \( E \), source at \( s \) \( \in V \), side information at \( v \) \( \neq V \) and terminals \( T \subset V \), characterize \( R(\{V, E, s, T\}) \subseteq R^{|T|} \): the set of achievable rates.

Converse:
- What are the minimal link capacities required?
- Achievability:
  - Which \( X \) descriptions to create?
  - Which \( Y \) descriptions to create?
  - How to decode \( X \)?

Complexity:
- What are the required operations at the nodes?

Cut sets
The majority of network coding related work is on cases where the cut-set bounds are tight.

Generalizing Point-to-Point Networks and Beyond
- Shannon 48:
  \( \text{Theorem} \)
  Assume both \( X \) and \( Y \) are binary symmetric. The outer bound is tight for the "Y" network, and \( \text{Theorem} \)
  \( \sum_{e \in C} H(e) \geq H(X|U) \)
  \( \sum_{e \in C} H(e) \geq H(Y|U) \)

Ford-Fulkerson 56:
Let \( A = B = \emptyset \) be the set of edges \( A \neq B \). Let \( C \) be the set of edges \( e \in E \) for which \( d(e) \in A \) and \( A(e) \in B \).

The cut-set bound is:
\( \sum_{e \in C} H(e) \geq H(X|U) \)
\( \sum_{e \in C} H(e) \geq H(Y|U) \)

Examples:
- Example 1: one terminal
  - But we know how to solve more than point-to-point networks!
- Example 2: more than one terminal
  - \( (v, e, x, a, z, T) \) if \( c(e) = x \in E \in R(\{V, E, x, a, z, T\}) \), then for each \( e \in E \) and each cut \( Y \), there exists a random variable \( U \), such that \( Y = X \), \( |U| \leq |Y| \), and
  \( \sum_{e \in C} H(e) \geq H(X|U) \)

Multi-resolution codes are Necessary
- A source may have a high rate to one terminal and a low rate to the other.
- As a result, two descriptions may be required.

Multi-resolution codes:
- Each point represents a sequence in \( \{0,1\}^n \).
- \( R = \log \left(\frac{\text{num regions}}{\text{bin size}}\right) \).
- Ideally, the high rate description refines the coarse one.

K Sinks with Direct Links from Y - Result

Multi-resolution codes:
- Each link of the source sequence in \( \{0,1\}^n \).
- \( R = \log \left(\frac{\text{num regions}}{\text{bin size}}\right) \).
- Ideally, the high rate description refines the coarse one.

Proof Sketch - random binning and independent equations

The figure below represents binning to 8 bins and its binary representation:

The main idea: randomly bin the source sequence:
- Each node sends as much independent equations on the bin index as it can.

Routing is sub-optimal
- To send the two descriptions (at the lowest possible rates) we must use multicast network coding.

Conclusion
- We derived inner and outer bounds on the rate region, and identified scenarios in which they are tight.
- The outer bound is at least as tight as the cut-set, and tighter in most non-trivial cases.
- While the cut-set bounds are loose, we extended the range of scenarios for which cut analysis describes the rate region.
- Our methods inherit the desirable properties of the building blocks.
- We identified an interesting connection between coding for networks with side information and successive refinement.
- In fact, this method might yield interesting results for other source coding problems as well. In a sense, this is a way to rigorously formulate the intuition from canonical problems to larger networks.