On Average Throughput Benefits in Network Coding

Emina Soljanin

Mathematical Sciences Research, Bell Labs

Collaborations with C. Chekuri, C. Fragouli, S. El Rouayheb, and Y. Shi

University of Southern California

8/28/2006

On Average Throughput Benefits in Network Coding

Emina Soljanin

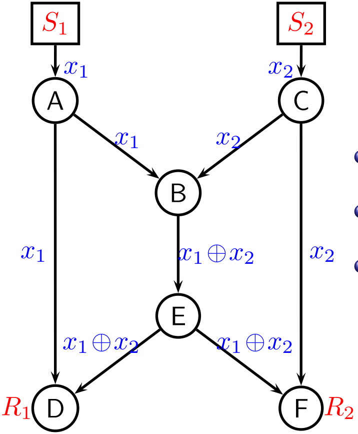
Mathematical Sciences Research, Bell Labs

Collaborations with C. Chekuri, C. Fragouli, S. El Rouayheb, and Y. Shi

University of Southern California

8/28/2006

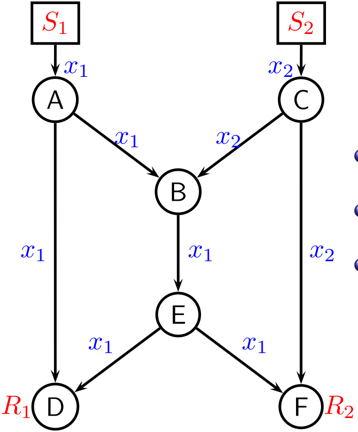
NETWORK CODING – Example 1



- Sources S_1 and S_2 produce bits x_1 and x_2.
- Each receiver needs bits from both sources.
- The edges have unit capacity.

Each receiver gets 2 bits. The average throughput with coding is 2.

NETWORK MULTICAST – The Butterfly



- Sources S_1 and S_2 produce bits x_1 and x_2.
- Each receiver needs bits from both sources.
- The edges have unit capacity.

The average throughput with routing is $\frac{3}{2} = \frac{3}{4} \cdot 2$.

OUTLINE OF THE TALK

- network coding basics
- examples of networks with small coding benefits
- examples of networks with large coding benefits
- some general results on coding benefits
- on network code alphabet size
- work in progres

COMMUNICATION NETWORK – A Mathematical Model

- A network is represented as a directed acyclic graph $G = (V, E)$.
- G has unit-capacity edges and parallel edges are allowed.
- There are h unit-rate information sources $S_1, \ldots, S_h$.
- There are N receivers $R_1, \ldots, R_N$ located at N distinct nodes.
- Can all sources simultaneously transmit information to all receivers? (We are interested in the transmission at full rate.)

NETWORK MULTICAST – Throughput

- Can all sources simultaneously transmit information to receiver j?
  Yes, if between the sources and the j-th receiver node
  - the number of edges in the min-cut is h (or equivalently)
  - there are h edge-disjoint paths (S_i, R_j) for $1 \leq i \leq h$.

  [Ford, Fulkerson], [Elias, Feinstein, Shannon] $\sim$ 50s
- Can all sources simultaneously transmit information to all receivers?
  Yes, if in addition each node of G can re-encode information.
  [Alshwede, Cai, Li, Yeung] $\sim$ 2000

NETWORK MULTICAST – Linear Combining

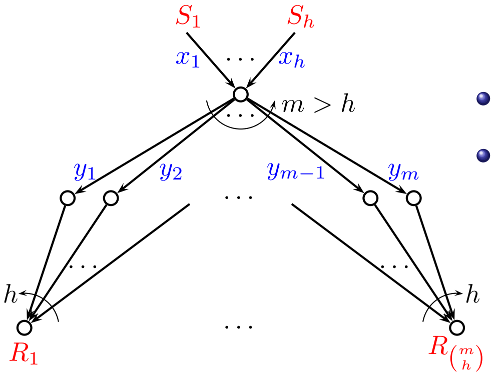
- Source S_i emits σ_i which is an element of some finite field $\mathbb{F}_q$.
- Each edge carries a linear combination of its parent node inputs.
- Consequently,

  each edge carries a linear combination of source symbols.
- The h edges a receiver observes

  should carry independent linear combination of source symbols.
- Network code design problem:

  How should nodes combine their inputs to ensure this independence?

NETWORK MULTICAST – Combination Network



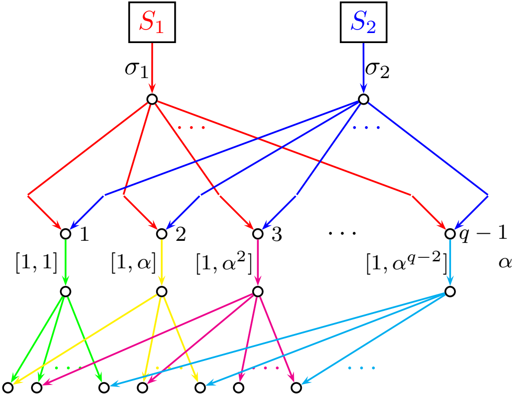
- Assume $x_j \in \mathbb{C}$, $j = 1, \dots, h$.
- Apply an overcomplete FT to $\{x_i\}$:

$$y_k = \frac{1}{\sqrt{h}} \sum_{j=1}^{h} x_j W_m^{(j-1)(k-1)}$$

where $k = 1, \dots, m$ and $W_m = e^{\imath \frac{2\pi}{m}}$.

Each receiver has access to h of the m numbers $\{y_k\}$ and can recover $\{x_j\}$.

NETWORK MULTICAST – An Example



α is a primitive element of $\mathbb{F}_q$.

$\longleftarrow \binom{q-1}{2}$ receivers.

A receiver observes $\sigma_1 + \alpha^i \sigma_2$ and $\sigma_1 + \alpha^j \sigma_2$.

AVERAGE THROUGHPUT WITH ROUTING

- The average throughput with coding is $T^c = 2$.
- Route σ_1 through one half of the $q-1$ intermediate nodes, and σ_2 through the other half.
- The average throughput with routing (for even $q-1$):

$$T^r = \frac{1}{\binom{q-1}{2}} \left[\frac{q-1}{2} \left(\frac{q-1}{2} - 1 \right) \cdot 1 + \left(\frac{q-1}{2} \right)^2 \cdot 2 \right] > \frac{3}{4} \cdot 2$$

CODING v.s. ROUTING – Throughput Measures

- $T_i^c = h$ is the throughput to receiver i when coding is used.
- T_i^r is the throughput to receiver i when only routing is used.
- Compare the average routing throughput with coding throughput:

$$T^r = \frac{1}{N}\sum_{i=1}^{N} T_i^r \quad \text{vs} \quad T^c = \frac{1}{N}\sum_{i=1}^{N} T_i^c = h.$$

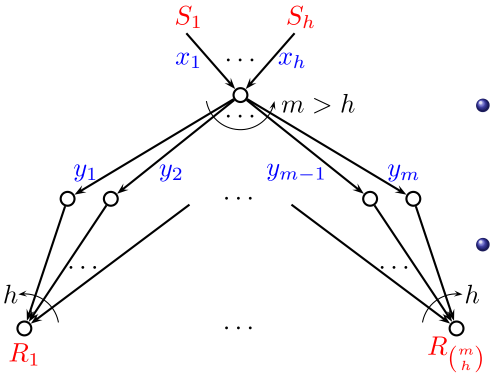
(We are not interested in the maximal routing throughput.)

- Coding benefits T^c/T^r depend on the throughput measure.

CODING v.s. ROUTING – Measures and Scenarios

- Individual receiver requirements: symmetric or sum.
- Type of routing: integral or fractional.
- Communications schemes: time sharing or channel coding.
- Type of network: directed or undirected.
- General results for symmetric throughput and fractional routing:
  1. In undirected networks, coding can at most double the throughput. (Li&Li '04)
  2. In both undirected and directed networks, coding benefits are equal to the integrality gap of the standard Steiner tree problem. (Agrawal&Charikar '04).

NETWORK MULTICAST – Combination Network



- Coding solution:
  Encode the h source symbols by an (m, h) RS code into m coding symbols.
- Routing solution:
  For each of the m edges, chose one of the h sources uniformly at random.

The routing scheme can be analyzed by an occupancy model.

AVERAGE THROUGHPUT WITH ROUTING

- There are h urns (sources) and each receiver has h balls.
- $\mu_0(h)$ is the number of unobserved sources (empty urns):

$$\Pr\{\mu_0(h) = k\} = \binom{h}{k}\left(1 - \frac{k}{h}\right)^h \Pr\{\mu_0(h-k) = 0\}$$

$$\Pr\{\mu_0(n) = 0\} = \sum_{l=0}^{n} \binom{n}{l} (-1)^l \left(1 - \frac{l}{h}\right)^h$$

- A receiver observes

$$h[1 - (1 - 1/h)^h] \gtrsim h(1 - e^{-1}) > h/2$$

sources on the average.

SYMMETRIC THROUGHPUT – A Coding Theorem

- The probability that a receiver does not observe source S_i is

$$\epsilon = (1 - 1/h)^h.$$

- The expected value of the throughput is

$$h[1 - (1 - 1/h)^h] = h(1 - \epsilon).$$

- Theorem: There exist a sequence of channel codes of rates

$$k/n \to 1 - \epsilon$$

  and a routing strategy such that the symmetric throughput

$$T_i(n) \to hk/n \to h(1 - \epsilon) \text{ as } n \to \infty.$$

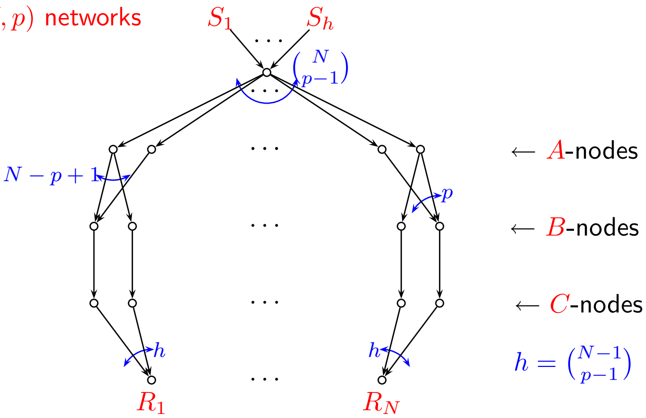
AVERAGE THROUGHPUT BENEFITS

- In our two-source network example, $T^r \geq 3T^c/4$.
- For general networks with two sources, $T^r \geq T^c/2 + 1/N$.
- For networks with two receivers, $T^r \geq T^c/2$.
- For several other classes of networks, $T^r \geq T^c/2$.
- There are networks for which

$$T^r \simeq \frac{1}{\sqrt{N}} \cdot T^c.$$

AVERAGE THROUGHPUT WITH ROUTING

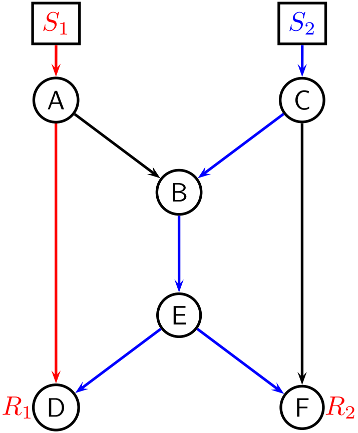
ZK(N, p) networks



Theorem: We have $\frac{T^r}{T^c} = \frac{p-1}{N-p+1} + \frac{1}{p}$.

For $p = \frac{N+1}{\sqrt{N}+1}$, we get $\frac{T^r}{T^c} = 2\,\frac{\sqrt{N}}{1+N} \lesssim \frac{2}{\sqrt{N}}$.

(PARTIAL) STEINER TREES



STEINER TREE PROBLEMS – Instance $\{G, S, \mathcal{R}, w, c\}$

Optimal packing $T^r(G, S, \mathcal{R}, c)$:

$$\max \sum_{t \in \tau} \frac{n_t}{N} y_t$$

$$\sum_{t \in \tau : e \in t} y_t \leq c_e, \quad \forall \ e \in E$$

$$y_t \geq 0, \quad \forall \ t \in \tau$$

$$\beta(G, S, \mathcal{R}) = \max_c \frac{T^c(G, c, S, \mathcal{R})}{T^r(G, c, S, \mathcal{R})}$$

Minimum weight $\text{OPT}(G, w, S, \mathcal{R})$:

$$\min \sum_{e \in E} w_e x_e$$

$$\sum_{e \in \delta(\mathcal{D})} x_e \geq 1, \quad \forall \ \mathcal{D}\text{: } \mathcal{D} \text{ is separating}$$

$$x_e \in \{0, 1\}, \quad \forall \ e \in E$$

$$\alpha(G, S, \mathcal{R}) = \max_w \frac{\text{OPT}(G, w, S, \mathcal{R})}{\text{LP}(G, w, S, \mathcal{R})} \quad \boxed{w^*, x^*_{\text{LP}}}$$

Theorem: $\max_{\mathcal{R}' \subseteq \mathcal{R}} \beta(G, S, \mathcal{R}') \geq \max\{1, \alpha(G, S, \mathcal{R}) / \sum_{j=1}^{N} 1/j\}$

Note: $\beta(G, S, \mathcal{R}) \geq \frac{T^c(G, c = x^*, S, \mathcal{R})}{T^r(G, c = x^*, S, \mathcal{R})} \geq \frac{1}{T^r(G, c = x^*, S, \mathcal{R})} = \frac{1}{\sum \frac{n_t}{N} y_t^*}$

STEINER TREE PROBLEMS – Instance $\{G, S, \mathcal{R}, w, c\}$

- There exists a partial tree t_1 of weight $w_{t_1} = \sum_{e \in t_1} w_e^* x_e^*$ such that

$$w_{t_1} \leq \frac{1}{\sum_{t \in \tau} \frac{n_t}{N} y_t^*} \cdot \frac{n_{t_1}}{N} \sum_{t \in \tau} w_t y_t^* \leq \underbrace{\boxed{\frac{1}{\sum_{t \in \tau} \frac{n_t}{N} y_t^*}}}_{\leq \beta} \cdot \frac{n_{t_1}}{N} \sum_{e \in E} w_e^* x_e^*$$

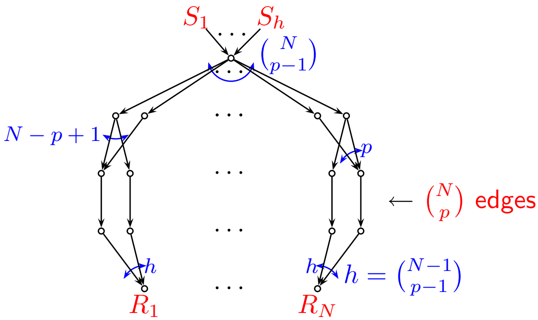
- Therefore,

$$w_{t_1} \leq \beta(G, S, \mathcal{R}) \cdot \frac{n_{t_1}}{N} \sum_{e \in E} w_e^* x_e^*.$$

- If $n_1 = N$, we have

$$\beta(G, S, \mathcal{R}) \geq \frac{w_{t_1}}{\sum_{e \in E} w_e^* x_e^*} \geq \alpha(G, S, \mathcal{R}).$$

RANDOM CODING FOR THE ZK NETWORKS



- Edges carry random linear combination of their parent nodes' inputs.
- The probability that all N receivers can decode all h sources is at least

$$\left(1 - \frac{N}{q}\right)^{\binom{N}{p}} \approxeq e^{-\frac{N\binom{N}{p}}{q}}.$$

CODE-ALPHABET SIZE – Deterministic Coding

- For networks with h sources and N receivers, the field of size

$$q = N$$

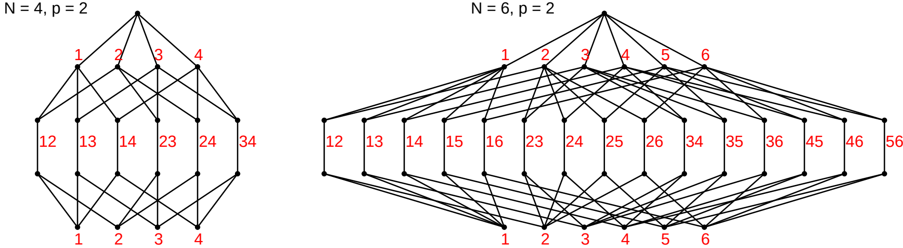
is sufficient. There exist configurations for which the field of size $q = O(\sqrt{2N})$ is necessary. [Jaggi et al. '03].

- For networks with two sources and N receivers, the field of size

$$q = \lfloor \sqrt{2N - 7/4} + 1/2 \rfloor$$

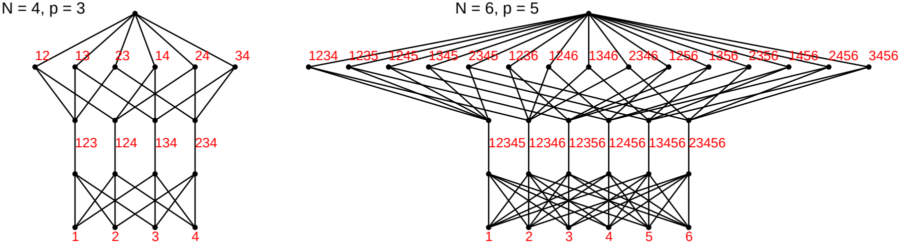
is sufficient. There exist configurations for which it is necessary. [Fragouli& Soljanin '04]

DETERMINISTIC CODING FOR ZK$(N, 2)$



- The number of information sources is $h = N - 1$.
- Remove the A-node with label N. B-nodes add their inputs.
- Receiver N gets σ_i, $i, j = 1, \ldots, N - 1$.
- Receiver i gets σ_i and $\sigma_j + \sigma_i$ for $i, j = 1, \ldots, N - 1$ and $j \neq i$.

DETERMINISTIC CODING FOR ZK$(N, N-1)$



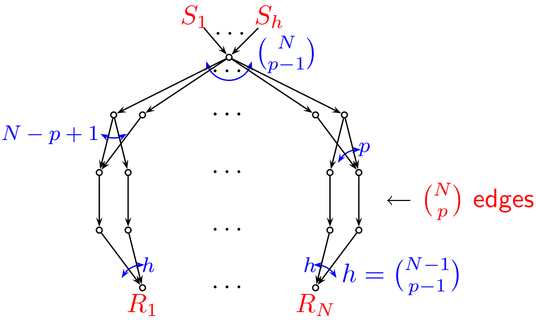
- The number of information sources is $h = N - 1$.
- Remove the A-node whose label contains N.
- The $B - C$ edges carry the following: $\sum_j \sigma_j$, σ_i for $i = 1, \ldots, N-1$.

CODE-ALPHABET SIZE – ZK Networks

- The code design generalizes to arbitrary p.
- The binary alphabet is sufficient.
- What about coding randomly over the alphabet $\mathbb{F}_q$?
- A bound on the probability that all receivers can decode

$$\left(1 - \frac{N}{q}\right)^{\binom{N}{p}} \approx e^{-\frac{N\binom{N}{p}}{q}}.$$

ZK NETWORKS WITH LIMITED RESOURCES



- Coding is done at $\binom{N-1}{p}$ nodes. Suppose that only k nodes can code.
- Theorem: This hybrid scheme achieves the average throughput of

$$T^c \cdot \frac{1}{N}\Big(p + \frac{N-p}{p} + k\frac{p-1}{h}\Big).$$

CONCLUSIONS

- Network coding benefits
  - are related to certain combinatorial optimization problems
  - depend on the throughput measure
- Symmetric throughput
  - benefits from network coding more than the average, but
  - can approach the average in symmetric networks by channel coding and routing.
- Average throughput
  - gets at most doubled by coding for large classes of networks.

WORK IN PROGRESS

- Networks with nonuniform demands:
  - In undirected networks, coding can at most double the throughput.
  - To code or not to code?
  - Neglected users or unfulfilled demands?
- Multicast in quantum networks:
  - No information without representation and no cloning!
  - Compression into the symmetric subspace.
  - Combining information from different sources?
- Wiretapped networks:
  - Ozarow&Wyner Wiretap II generalization.
  - Alphabet size issues.
  - Code design.