An Overview of Network Information Flow Problem

Yun Sun
sunyun@dcs.qmul.ac.uk

February 4, 2005

Abstract

The effect of network coding began widely realized when a network with its optimized throughput can be achieved only if linear coding was introduced to its intermediate nodes was presented by R.Ahlswede et al. in [1]. A series of advantages of network coding over routing have been discovered during afterwards research by R.Koetter, M.Médard et al. This talk tries to provide an overview of network information flow problem.

1 Basics of network coding

Network coding has been proved to be an effective technology in solving *network information flow* problem, which is derived from traditional multicommodity flow problems and have recently absorbed some ideas from information theory and coding theory.

Our information network are directed acyclic graphs $G = (V, E)$. The vertex V contains three disjoint subsets: source nodes S, target nodes T and intermediate nodes I.

For each **source node** s, the messages it transmits through intermediate nodes to target nodes through edge set E are drawn from a fixed finite alphabet A with size $|A| > 1$. For each **target node** t, the messages it requires is a subset of messages from source nodes. The **intermediate nodes** can not only duplicate and forward messages they receive from in-edges, but also use mathematical functions to compute these messages before forwarding them(the power of node's computability is unlimited).

If we can find a set of functions which help satisfy all target nodes, then we say this network is **solvable** and we have found a solution for it. If messages transmitted in edges are scalar quantities, we call this solution scalar solutions; if these messages are made up of multiple scalar quantities, we call this solution vector solution. If output message of each intermediate node is one of its incoming messages, we call this solution routing solution, similarly, we also have linear solution and non-linear solution.

1.1 Examples of network coding

A classical example of network coding (Figure 1) was presented by R.Ahslwede et al. in [1]. Each edge in this network has capacity 1. Our task is to transfer two messages x and y concurrently from source node $1(s_1)$ and $2(s_2)$ to both target nodes $5(t_1)$ and $6(t_2)$.

The maximum throughput of this network is 2 which is the largest possible capacity for each source-target pair. But this throughput is not achievable if we don't use network coding because two messages cannot be sent at the same time through the middle edge ($3 \rightarrow 4$) whose capacity is only 1.

On the contrary, if we use XOR(exclusive or) as encoding function, we can solve this bottleneck. Thus, we reconstruct x and y from $y \bigoplus (x \bigoplus y)$ and $x \bigoplus (x \bigoplus y)$ respectively.

2 Multicast networks

2.1 The definition of multicast

Multicast information flow problems have received the most study till date. In multicast network there is only one source node and multiple target nodes, and all messages are available at the source, while they are demanded by each target nodes.

2.2 Max-flow Min-cut theorem

T.Ho in [9] presented this theorem in this way: Coding within a network allows sources to multicast information at a rate approaching the smallest minimum cut between the source and any receiver, as the coding symbol size approaches infinity.

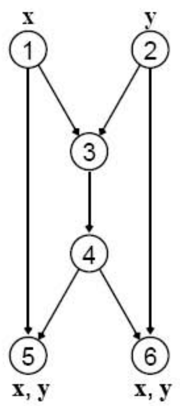

Figure 1: Network coding example:Routing Capacity = 1/2

2.3 The upper and lower bound of the scale of multicast alphabet set

Theorem Every solvable multicast network has a scalar linear solution over a sufficiently large finite field alphabet.

Actually a finite field is enough to promise a linear solution for a given multicast network. However, determining the minimum alphabet size for a specific multicast problem is NP-hard [6].

For multicast session with a throughput r and the cardinality of target set k, there exist a solution based on a finite field $GF(2^m)$, where $m \leq \log_2(r+1)$; and meanwhile, the lower bound of coding field size is to be at least $2\sqrt{k}$.

Notice **Theorem** A multicast network that has a solution for a given alphabet might not have a solution for all larger alphabets.

An example Figure 2 has been give in [8]. This multicast network is solvable if and only of the alphabet size is neither 2 nor 6. Take $|A| = 3$ for instance, let $\mathcal{S}$ and $\mathcal{T}$ are two orthogonal latin squares: $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 3 & 2 \\ 3 & 2 & 1 \\ 2 & 1 & 3 \end{pmatrix}$.
Then, $\lambda_1 = x, \lambda_2 = y, \lambda_3 = \mathcal{S}(x,y), \lambda_4 = \mathcal{T}(x,y)$ is a solution of Figure 2.

The $|A|$ cannot be 2 or 6 because pairs of orthogonal latin squares exist if and only if the order of the matrices is neither 2 nor 6.

2.4 Network coding vs. routing

Network coding makes it possible to achieve maximum throughput given by the max-flow min-cut theorem, which might not be achieved if only routing is allowed. Figure 3 presents an ideal example.

R.Yeung in [2] provide two measures of network coding's advantage over trivial routing: Coding gain C and Bandwidth saving S.
The definition of *Coding gain* and *Bandwidth saving* are as follows:

$$C = \frac{\textit{network coding capacity}}{\textit{network routing capacity}} - 1 \quad (1)$$

$$S = 1 - \frac{\textit{total bandwidth required with network coding}}{\textit{total bandwidth required with routing only}} \quad (2)$$

Bandwidth saving S is well-defined for multi-source problems, while Coding gain C is not.

When calculating Coding gain C, network coding capacity is its max-flow bound, and network routing capacity is the supremum of all possible fractional message throughputs achievable by routing [3].

Example 1: Figure 3
For node 6 and 7, each should receive at least $2k$ in order to reconstruct x, y; on the other hand, total capacity for edge $2 \rightarrow 6, 4 \rightarrow 5, 3 \rightarrow 7$ is $3n$, thus $2 * (2k) \leq 3n$, routing capacity $k/n \leq 3/4$ and $3/4$ is achievable.

Example 2: Figure 1
Edge $3 \rightarrow 4$ must be shared by x, y, thus $2k \leq n$, routing capacity $k/n \leq 1/2$ and $1/2$ is achievable.

Example 3: Figure 4
For each element in x, y, it should be presented $(N + 1)$ times among node $2N + 4 \ldots 4n + 3$, otherwise at least one of target nodes cannot reconstruct it successfully. Then, we have $(2k) * (N + 1) \leq (2N) * n$, routing capacity $k/n \leq N/(N + 1)$ and $N/(N + 1)$ is achievable.
This network can be generalized to S.Riis's solvable network with routing capacity equals $1/n$.

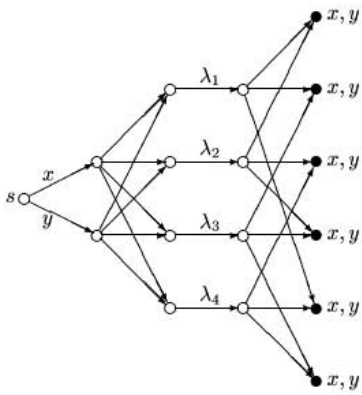

Figure 2: A multicast network solvable when q is neither 2 nor 6

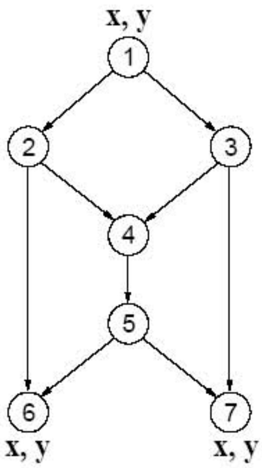

Figure 3: Routing Capacity = 3/4

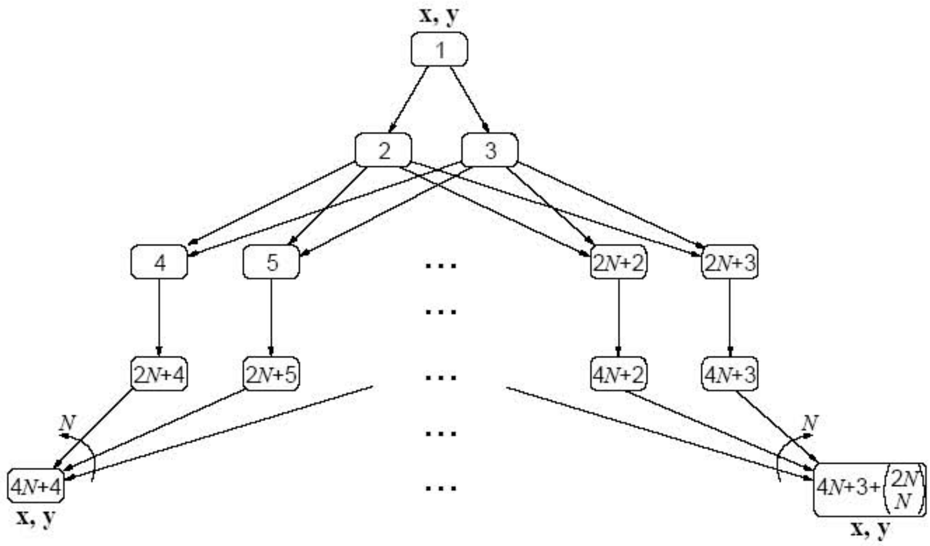

Figure 4: Routing Capacity = N/(N+1)

Example 4: Figure 5
For (km) elements in $x^{(i)}$, it should be presented $(n-I+1)$ times, thus we have $(mk)*(N-I+1) \leq N*n$, that is routing capacity $k/n \leq N/(m(N-I+1))$ and is achievable.

Example 5: Figure 6
For edge $1 \to 3, 1 \to 4$, $2k \leq 2n$, thus routing capacity $k/n \leq 1$ and is achievable. A solution can be presented when $k = n = 2$.

Example 6: Figure 7
$3k \leq 4n$, the routing capacity of this network is 4/3.

3 Non multicast networks

We have some interesting results for non multicast networks. According to [5], for each vector size k, we can find a network which allows no vector-linear solution; but this network has vector linear solution for a larger vector size.
Example 6: Figure 8 is a solvable non multicast network with no vector linear solution [10].
Figure 9 demonstrate one solution over an alphabet of cardinality 4. The

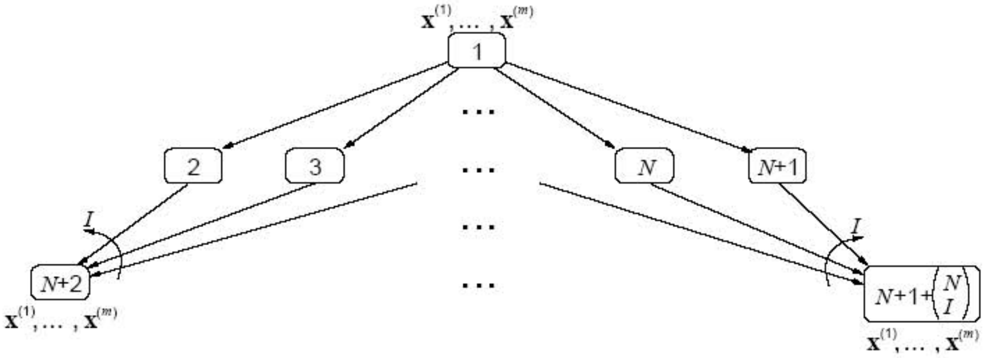

Figure 5: Routing Capacity = N/(m(N-I+1))

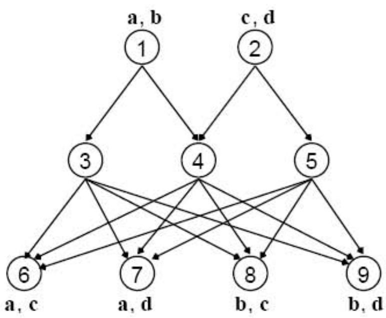

Figure 6: Routing Capacity = 1

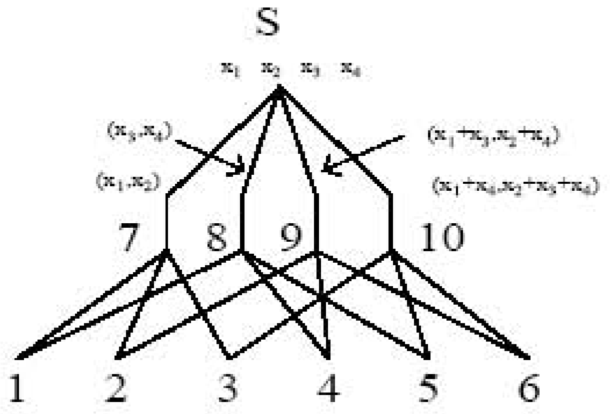

Figure 7: flow problem $N_{2,4,2}$

function $\bigoplus$ indicates binary XOR and $t(x)$ indicates exchanging the order of the bits in a 2-bit binary word x. Its solution is shown in Figure 9.

4 Network Coding and Error Correcting Codes

S.Riis in [4] mentioned that we can build a bridge from network's linear boolean solution with maximum distance separable code (MDS code), In [7], S.Riis established an one-to-one correspondence between solutions of network flow problem and error correcting codes.

Theorem The flow problem $N_{k,r,s}$ has a solution if and only if there exists an $(r, |A|^k, r-s+1)$ $|A|$-ary error correcting code.
In which a flow problem $N_{k,r,s}$ means a multicast network with source node transmits k messages, $|I| = r, |T| = \binom{r}{s}$; a (n, k, d) linear code is a linear code with length n, dimension k and minimum weight d; a q-ary linear code C means a linear subspace of F_q^n. For instance, Figure 7 show that there is (4,16,3) 4-ary MDS code.

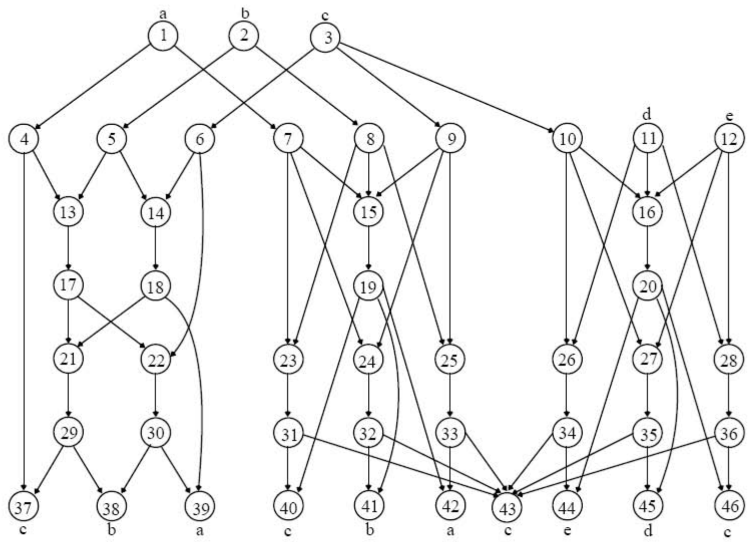

Figure 8: N1:A solvable network without vector solution

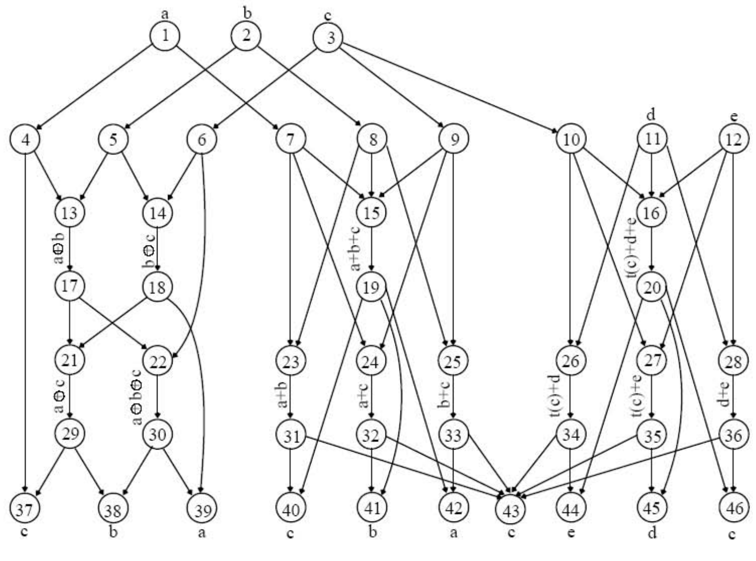

Figure 9: The solution of network N1

References

[1] R.Ahlswede, N.Cai, S.-Y.R.Li, and R.W.Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. IT-46, no.4, pp. 1204-1216, July 2000.

[2] R. W. Yeung, "Two Approaches to Quantifying the Bandwidth Advantage of Network Coding," presented at 2004 IEEE Information Theory Workshop, San Antonio, Oct 25-29, 2004.

[3] J. Cannons, R. Dougherty, C. Freiling, and K. Zeger, "Network Routing Capacity", IEEE/ACM TRANSACTIONS ON NETWORKING (submitted October 16, 2004).

[4] S. Riis, Linear versus non-linear Boolean functions in Network Flow, preprint, November, 2003

[5] A. Lehman and E. Lehman "Network Coding: Does the Model Need Tuning?" Proc. of 16th Symposium on Discrete Algorithms, January 2005.

[6] A. Rasala-Lehman and E. Lehman, "Complexity Classification of Network Information Flow Problems", 41st Annual Allerton Conference on Communication Control and Computing, Oct. 2003.

[7] S.Riis, Problems in Network coding and error correcting codes, preprint, December, 2004

[8] R. Dougherty, C. Freiling, and K. Zeger, "Linearity and Solvability in Multicast Networks", submitted to the *IEEE Transactions on Information Theory*

[9] Tracey Ho, "Networking from a network coding perspective", PhD thesis for EECS, MIT, May, 2004

[10] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of Linear Coding in Network Information Flow", preprint, February 2004