Mathematical Modelling of Computer Networks: Part II

Module 1: Network Coding

Lecture 1: Introduction to Network Coding
29th October 2013

Laila Daniel and Krishnan Narayanan
Dept. of Computer Science, University of Helsinki, Finland

Network Coding

- Network Coding began with the paper Ashwede-Cai-Li-Yeung Network Information Flow, IEEE-IT 2000 (ACLY00)
- Network Coding has made rapid strides in several applications such as content distribution networks, disruption/delay tolerant networks, wireless and sensor networks.
- Network Coding can provide the advantages of efficient operation, improved security and robustness and QoS support in wired and wireless networks.
- The mathematics used here is based on Finite Fields

Contents of the Network Coding Module

- We provide a concise introduction to a major result in the field of network coding - **the network multicast problem**
- Specifically, we discuss the following results
  - Multicast capacity in directed networks
  - Show that linear network coding can achieve multicast capacity in directed networks
  - Show that linear network coding can be implemented by an efficient algorithm
  - Describe Random linear network coding as an attractive distributed solution to the multicast problem
- As applications of Network coding we describe in detail
  - How Network Coding can be combined with TCP
  - How Network Coding can improve throughput and reliability in wireless networks

Today's Lecture: Overview

- What is network coding ?
- The butterfly network example
- Some applications of network coding
- Single source multicasting problem and network coding solutions
- Derivation of the multicast capacity
- Linear network coding formulation
- Summary

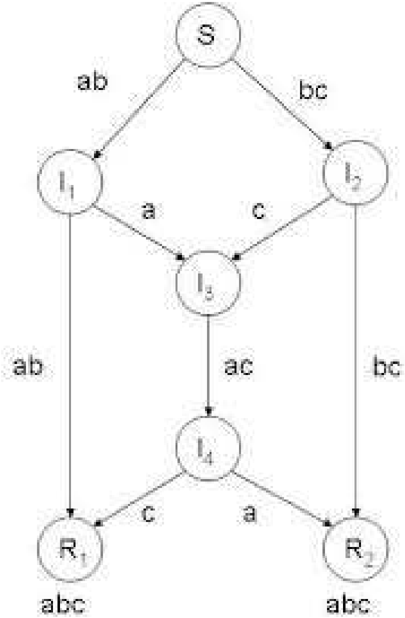
What is Network Coding?

- Traditionally in a packet network, the routers forward the packets on their incoming links to their outgoing links
- In the Network Coding approach, the intermediate nodes in the network can mix or combine the packets on their incoming links and create new packets that are sent on their outgoing links
- Formally, the intermediate nodes compute functions of the packets they receive and send them on their outgoing links

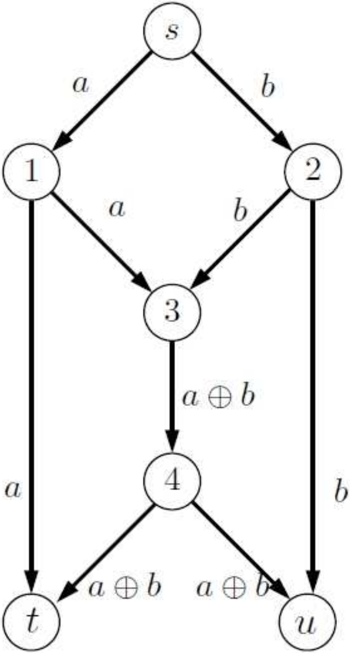
Network Coding : A generalization of packet routing

- Network Coding is a generalization of packet routing and is essential to achieve network capacity (e.g., multicast)
- Network Coding can simplify the design of distributed algorithms for networks
- For example, the multicast problem is known to be NP-hard and hard to approximate for general network. However, with Network Coding approach a polynomial time algorithm exists for the optimal solution of the multicast problem
- Packet information flow in networks not quite the same as commodity flow or traffic flow in networks
- This simple idea of Network Coding turns out to be very versatile in several packet networking applications

Butterfly network using routing



Butterfly network using Network Coding



Butterfly network

- The butterfly network example to illustrate Network Coding introduced in first paper on (ACLY00)
- The butterfly network is to Network Coding what the "Hello World" program is to C programming
- It is a simple example, yet reveals many insights that arise in this approach.
- The routing has maximal efficiency of 1.5 packets per channel use in this network whereas Network Coding can achieve an efficiency of 2 packets per channel use

Network Coding can be viewed as a generalization of packet routing

Butterfly Network - Some comments

- The specific routing scheme shown for the butterfly network achieves a throughput of 1.5 packets/channel use.
- Is this the best possible for **ANY** routing scheme in this network?
- The Network Coding scheme achieves throughput of 2 packets/channel use
- Is this the best possible Network Coding scheme for this network? (hint: use max-flow min-cut theorem)

Network Coding gain

- The increased throughput of network coding has been attained at the cost of encoding at an intermediate node and decoding at the receiver nodes

$$\textit{coding gain} = \frac{\textit{throughput under network coding}}{\textit{throughput under routing}}$$

- The coding gain for the butterfly network is $\frac{4}{3}$
- How large can such coding gains be in arbitrary networks?
- In general coding gain can be as large as $\Omega(\frac{log(|V|)}{log\ log(|V|)})$ and $\Omega(|T|)$ where $|V|$ is the number of vertices V in the directed network and $|T|$ is the number of multicast receivers.

General Network Information Flow problem

- Let $S = \{s_1, s_2, \ldots, s_{|S|}\}$ be the set of sources
- Assumption: All sources have infinitely many packets to send
- Objective: Each sink vertex $t \in T$ wants to reconstruct all packets sent by a subset $D_t \subset S$ of sources
- Source s_i has packets $p_1, p_2, p_3, \ldots$ to send.
- Suppose over n channel uses k of these packets can be reconstructed by t, then $r_i = \frac{k}{n}$ is the rate at which sink t gets packets from source s_i (number of packets per channel use)
- Each sink requiring a particular source s_i is to be served at the same rate r_i
- A vector of rates $R = (r_1, r_2, \ldots, r_n)$ is achievable if for some (possibly time-varying) local encoding operations at nodes and for some n, each sink can reconstruct the source it demands at the specified rate

Multicast Problem (1/2)

- The multicast problem is an important special case of network information flow.
- Here every sink t demands each of the sources i.e, $D_t = S$ for all $t \in T$
- Single source multicast is the case where a single source is demanded by all the receivers in the network
- Basic result: Linear network coding (LNC) can achieve multicast capacity (maximum possible multicast rate)
- LNC means that the node computes a linear function of the incoming packets to form new packets that are forwarded to the outgoing links
- In the Butterfly network example, the exclusive-or ($\oplus$) operation performed at the node is a linear operation as it corresponds to addition in the finite field $\mathbb{F}_2$ consisting of the two elements 0 and 1

Multicast Problem (2/2)

- For general flow problems, Linear Network Coding is not in general sufficient to achieve capacity.
- Unicast problem (single source and single sink) and multiple unicast problem (multiple sources and equally as many sinks each receiving a single source) are examples of general flow problems
- Next we formulate the notion of multicast capacity after a brief review of max-flow min-cut theorem and data processing inequality

Recap: Max-flow min-cut theorem (Ford-Fulkerson)

- The maximum flow from a source s to a sink t is equal to the capacity of the minimal cut separating the source and the sink.
- An s-t cut is defined to be the set of edges that meet every path from s to t i.e, every path from s to t includes some edge in the given s-t cut.
- The capacity of a cut is the sum of the weights of the edges that belong to it.
- If the edge weights are chosen to be unity, then the number of pairwise edge disjoint paths is equal to the capacity of the minimum cut (Menger's theorem).
- The Ford-Fulkerson algorithm can be used to efficiently find these paths.

Recap: Data Processing Inequality

- Let $X \to Y \to Z$ be a Markov chain.
- $I(X;Y)$ denotes the mutual information between random variables X and Y
- The mutual information between X and Y is the reduction in uncertainty about X when Y is known
- $I(X;Y) = H(X) - H(X|Y)$
- H(X) is the entropy of the random variable X (which captures the uncertainty in random variable X)
- The conditional entropy $H(X|Y)$ is the uncertainty about X when Y is known
- $I(X;Z) \leq I(X;Y)$ is called Data processing Inequality (for proof see Cover-Thomas book on Information Theory)

Multicast Network Capacity -(1/3)

- Consider a multicast network $N = (V; E)$ with E a multiset of directed edges of unit capacity with k 'parallel' edges between a pair of nodes corresponding to the capacity k of the link between the nodes.
- X_B : collection of packets transmitted on a subset B of edges
- Let s be the source node, t the sink node and C any (s, t)-cut
- Let $O(s)$ denote the outgoing links from node s
- Let $I(t)$ be the incoming links into node t
- The packets from the outgoing links in the source node s have to pass through cut C before reaching the terminal node t.
- This implies that $X_{O(s)} \to X_C \to X_{I(t)}$ is a Markov chain

Multicast Network Capacity . (2/3)

- $X_{O(s)} \to X_C \to X_{I(t)}$ is a Markov chain
- So, using Data Processing Inequality,
  $I(X_{O(s)}; X_{I(t)}) \leq I(X_{O(s)}; X_C)$
- The quantity $I(X_{O(s)}; X_C)$ has a trivial upper bound of $|C|$, the number of packets that go across the cut C for each channel use
- So $I(X_{O(s)}; X_{I(t))} \leq |C|$ where C is any (s, t) -cut
- Minimizing over all such cuts, yields the upper bound:
- $I(X_{O(s)}; X_{I(t)}) \leq mincut\,(s, t)$

Multicast Network Capacity . (3/3)

- Let $R(s,t)$ denote the communication rate between the nodes s and t
- It follows from the derivation just concluded that $R(s,t) \leq \ mincut\,(s,t)$
- To achieve the upper bound it is necessary that for each mincut C, $I(X_{O(s)}; X_C) = |C|$
- This implies that the packets sent through the mincut should be independently and uniformly distributed over the packet alphabet
- We can extend this result easily to the case of multiple sinks T
- The multicast rate from s to T, $R(s,T) \leq \ min_{t \in T} mincut\,(s,t)$ as the multicast rate from s to T cannot exceed the transmission rate from s to any of the receivers in T

Main Theorem of Network Multicasting

- The multicast rate from s to T, $R(s,T) \leq min_{t \in T} mincut\,(s,t)$ as the multicast rate from s to T cannot exceed the transmission rate from s to any of the receivers in T
- The Network Multicasting theorem states that the upper bound in the above inequality can be attained (with equality) using Network Coding and indeed using Linear Network Coding over a sufficiently large finite field

Summary

- Network Coding generalizes packet routing
- Butterfly network
- Network Coding mixes packets in the network
- The Multicast capacity is the maximum rate $R(s, T) = min_{t \in T} mincut\,(s, t)$
- The Multicast Network Coding Theorem states that Network Coding can achieve the Multicast Capacity
- Even Linear Network Coding can achieve the Multicast Capacity