LECTURE 38
Research Topics: Computational Science & Computer Networking

Definition 38.1. Computational science is the field of study concerned with the solution of scientific problems by constructing mathematical models and quantitative analysis techniques.

The three modes of science:

- theory
- experimentation/observation
- theory

Related fields/areas of specialization:

- Bioinformatics
- Cheminformatics
- Chemometrics
- Computational biology
- Computational chemistry
- Computational economics
- Computational electromagnetics
- Computational engineering
- Computational finance
- Computational fluid dynamics
- Computational forensics
- Computational geophysics
- Computational linguistics
- Computational mathematics
- Computational mechanics
- Computational neuroscience
- Computational particle physics
- Computational physics
- Computational statistics
- Computer algebra
- Environmental simulation

- Financial modeling
- Geographic information system (GIS)
- High performance computing
- Machine learning
- Network analysis
- Numerical weather prediction
- Pattern recognition

38.1. Network Coding

Network coding is a concept from information theory and coding theory that can be used to attain the maximum information flow in a network. In networks using network coding, instead of relaying received packets, nodes take several packets and combine together into a packet for transmission.

So, some preliminaries:

Definition 38.2. A network can be represented by a directed graph $G = (V, E, C)$. V is the set of nodes in the network, E is the directed links (edges of the graph), and C the capacity of each link in E.

38.1.0.1. *Linear network coding.* In the Linear Network Coding problem, a group of nodes P are involved in moving data from S source nodes to K sink nodes. Each node generates a new packet that is a linear combination of earlier received packets by coefficients in a Galois (finite) field $GF(2^S)$.

Each message X_k is related by the received messages M_i by

$$X_k = \sum_{i=1}^{S} g_k^i \dot{M}_i \tag{38.1}$$

Each node forwards the computed value X_k along with all the coefficients used in the k^{th} level g_k^i. The result is a linear system $X = GM$ where with the knowledge of X, G, we need to compute M. Each receiver in K tries to solve this linear equation, and for which you have received at least $N \leq S$ packets.

So, from a mathematical standpoint, network coding is performing operations on strings of s bits (working in a Galois field $GF(2^s)$). Addition is the standard bitwise `xor`. Multiplication is a product of two polynomials following by computing the remainder modulo of a polynomial irreducible in $GF(2^s)$. The result is that you can implement both addition and multiplication as combination of s shifts and additions.

Each node continually performs Gaussian elimination to reduce G to row-echelon form. The result is then used to back substitute to get the message M.

38.1.0.2. *Disadvantages of network coding.* It is important to note that loss of one packet could affect many other packets. This means that a one bit loss in the network results in several bits losses for the receiving nodes.

Synchronization is critical as there is more than one incoming data stream at the input of an intermediate node. This is not a problem for nonreal-time applications but can be serious problem for real-time applications like voice and video transmission.

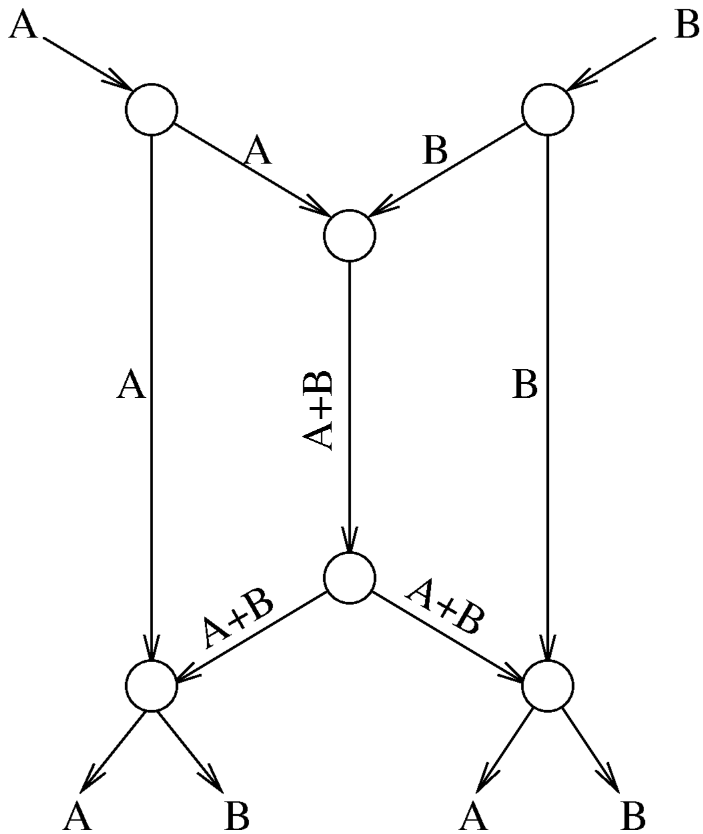

Figure 1. Network Coding: Butterfly Example

38.1.0.3. *Research problems.*

Performance: Gaussian elimination is $O(n^3)$. Can we do better?

Multisource network coding: What happens when more than one node attempts to multicast?

Cyclic Networks: How deal with loops in your network, esp. when applying to routing applications?

File distribution: Use network coding to avoid swarming in P-2-P networks. Microsoft Research is actively looking at this with the Avalanche project.