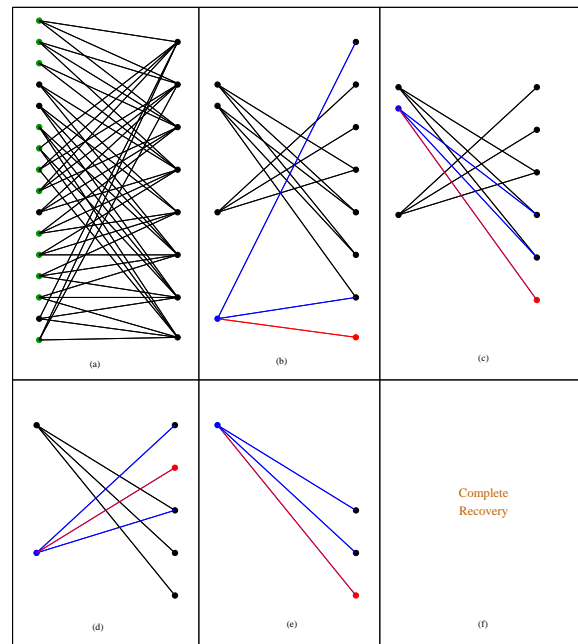


Low-Density Codes and the Erasure Channel



M. Amin Shokrollahi

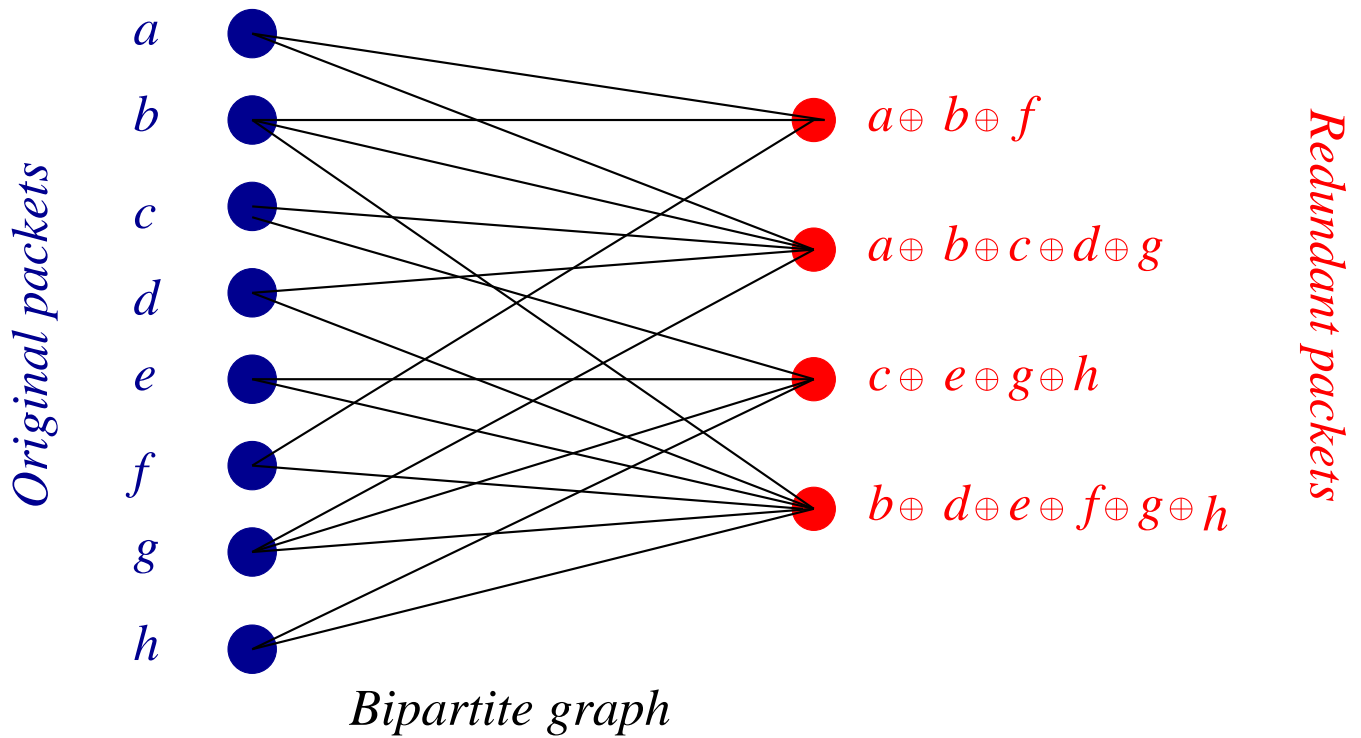
Bell Laboratories



Lucent Technologies
Bell Labs Innovations

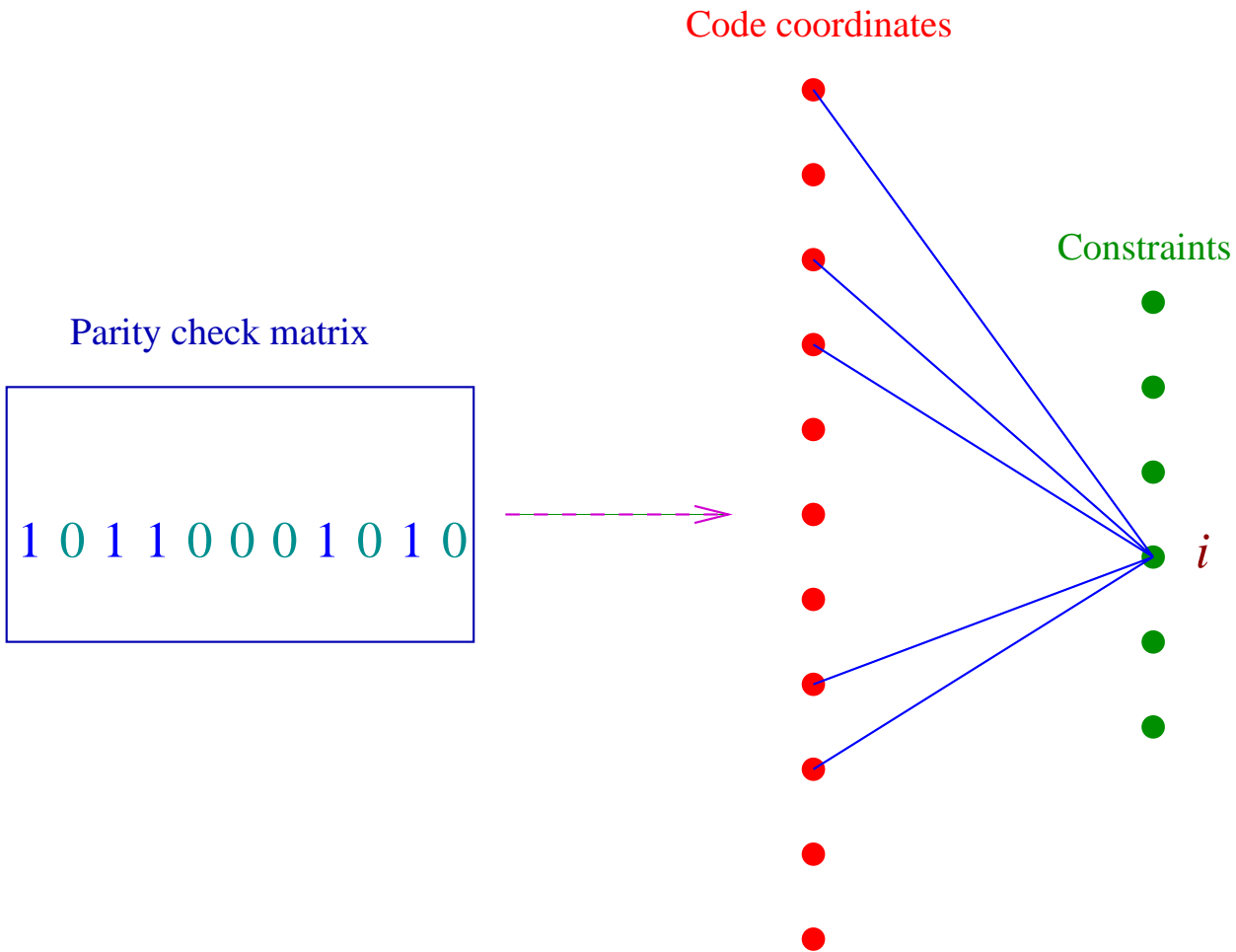
Encoding with Bipartite Graphs

Packets = Bits

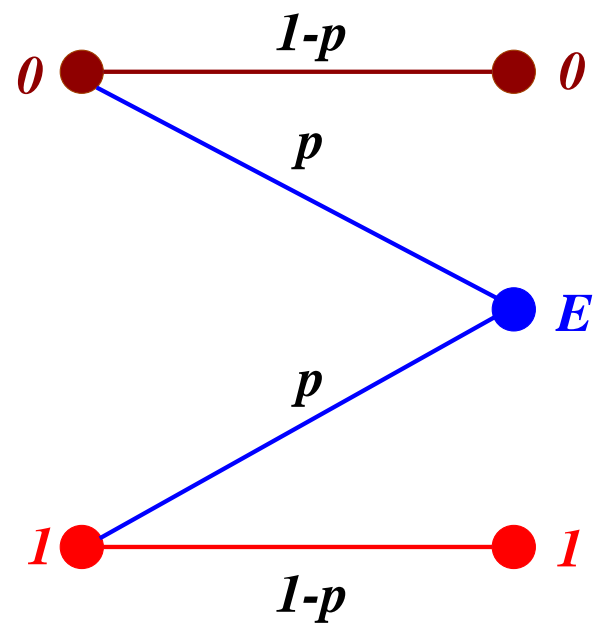


Encoding time is proportional to number of edges in graph.

Different Perspective

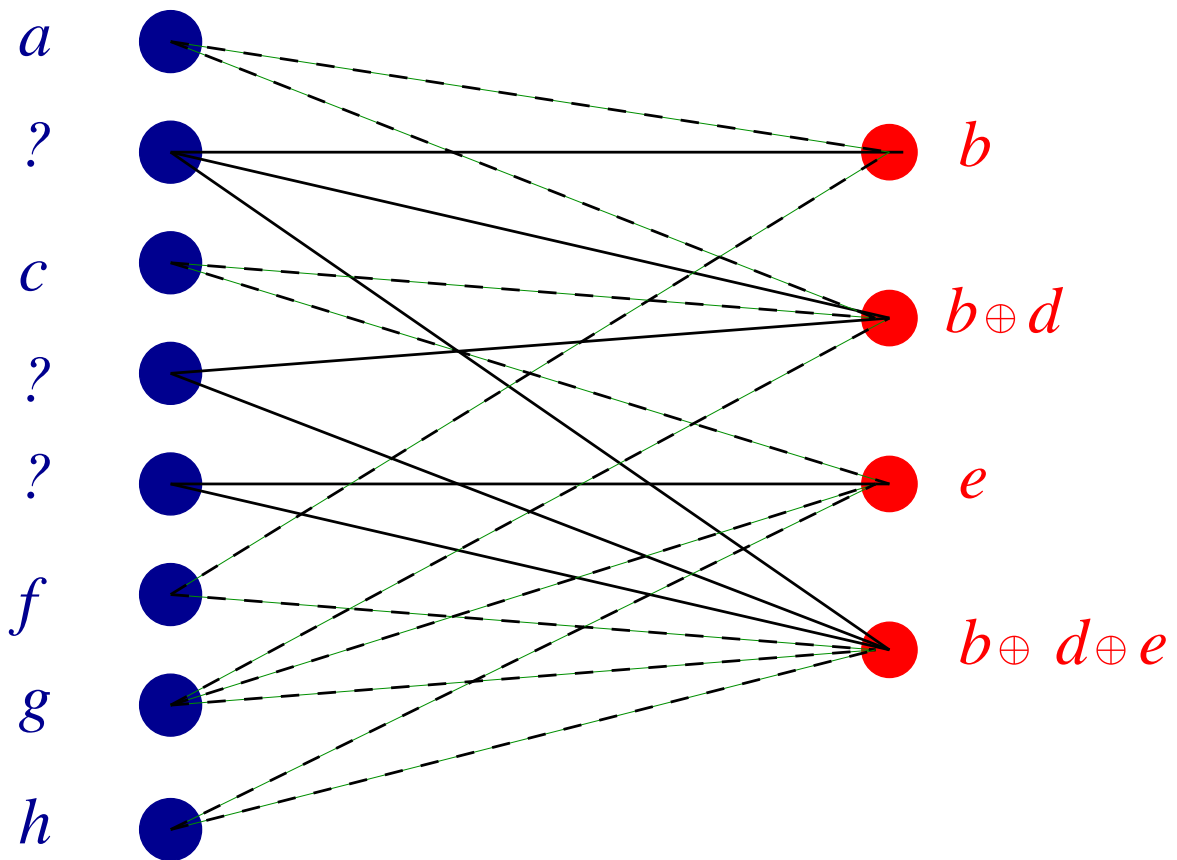


Erasure Channel



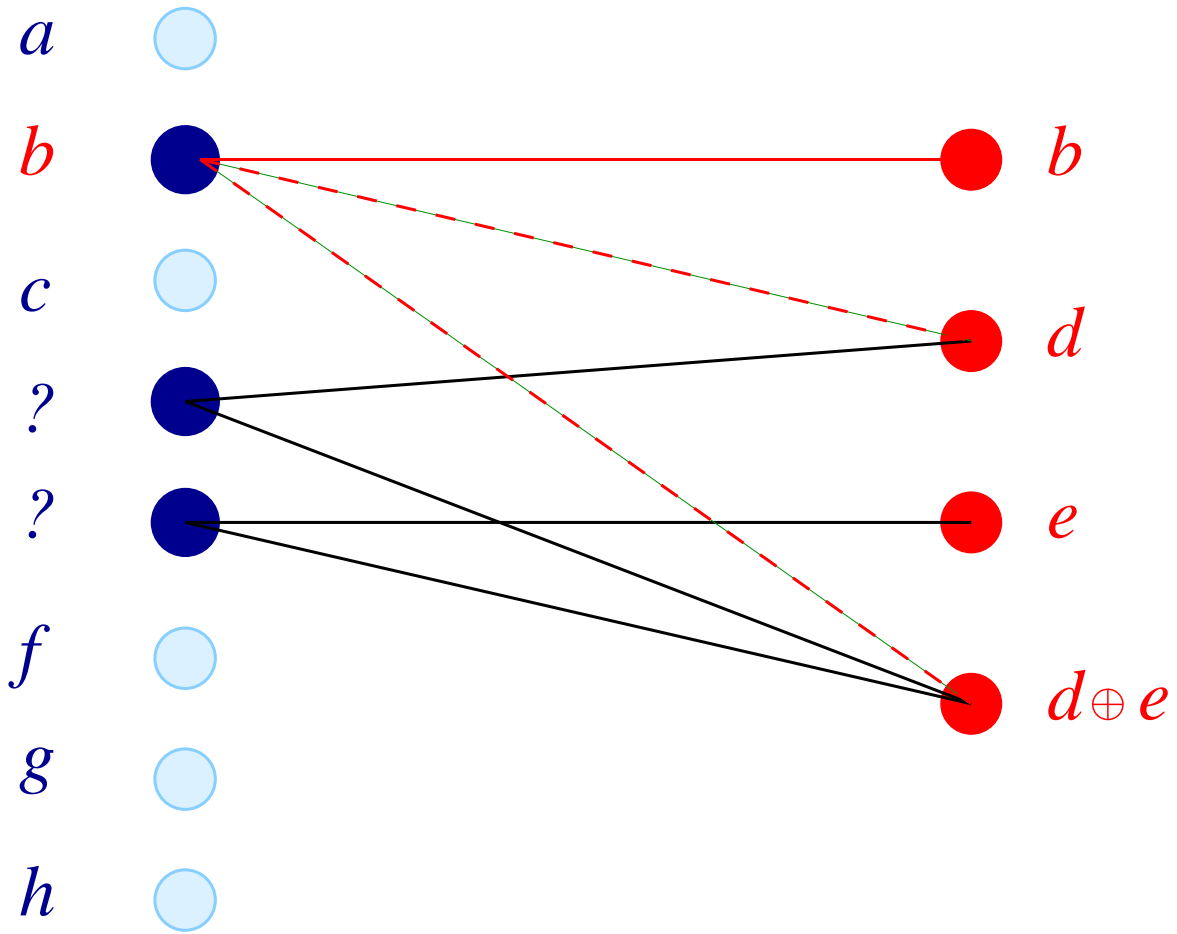
Decoding

Stage 1: Direct Recovery



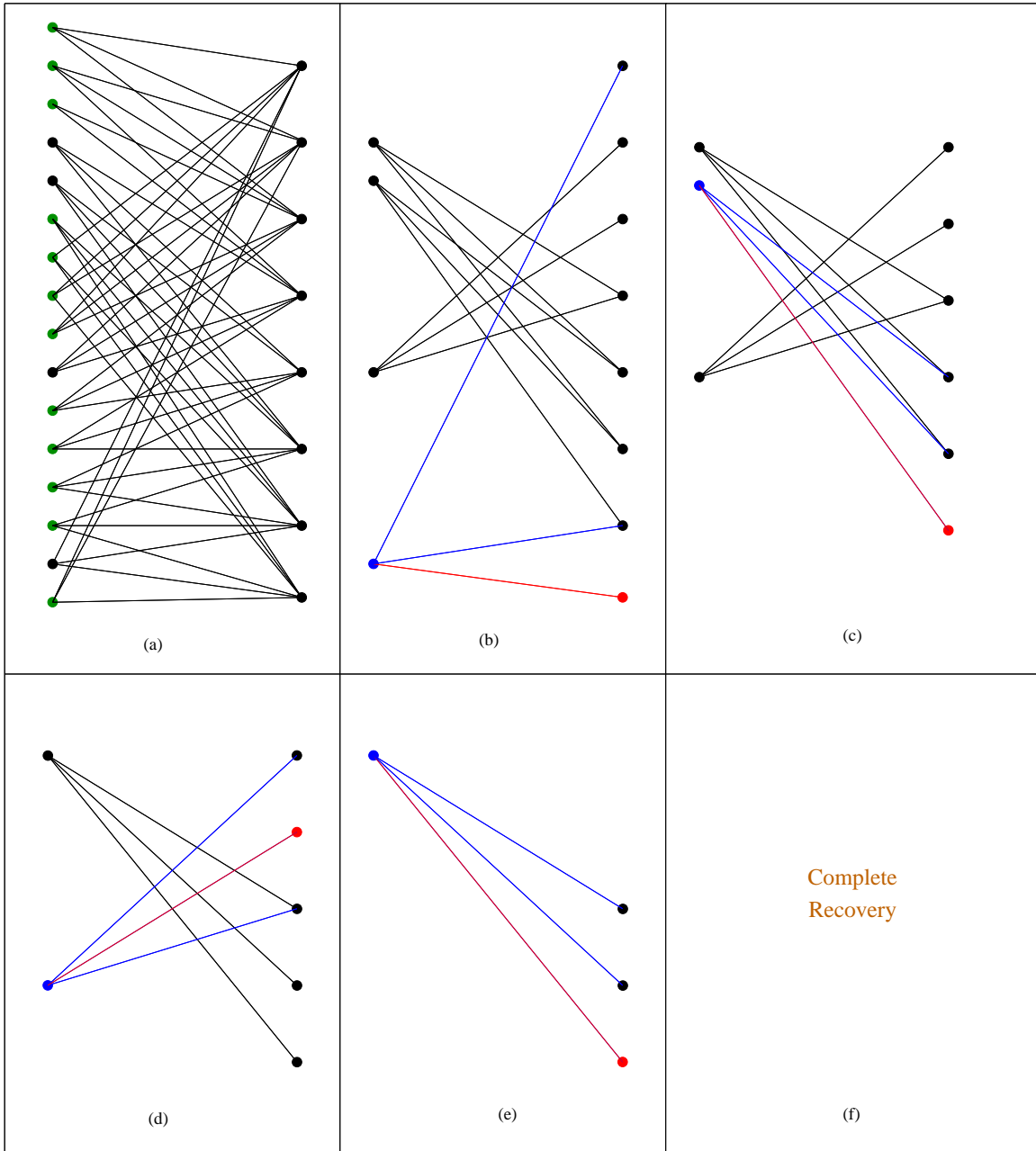
Decoding

Stage 2: Substitution Recovery



Decoding time is proportional to number of edges in graph.

Example



The Problem

Have **fast encoding** and **decoding** algorithm, if graph is **sparse**.

Want to **design codes (=graphs)** that perform good with respect to these algorithms.

How?

Experiments

Choose **regular graphs**.

A **(3, 6)**-graph recovers from **42.9%** erasures.

A **(4, 8)**-graph recovers from **38.3%** erasures.

A **(5, 10)**-graph recovers from **34.1%** erasures.

What are these numbers?

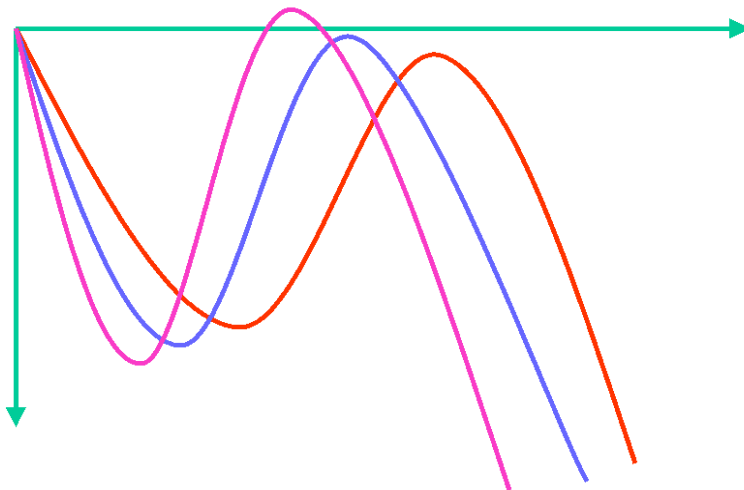
Revelation

Theorem. A random (k, d) -graph recovers from a p -fraction of erasures with high probability iff

$$p(1 - (1 - x)^{d-1})^{k-1} < x \quad \text{for } x \in (0, p).$$

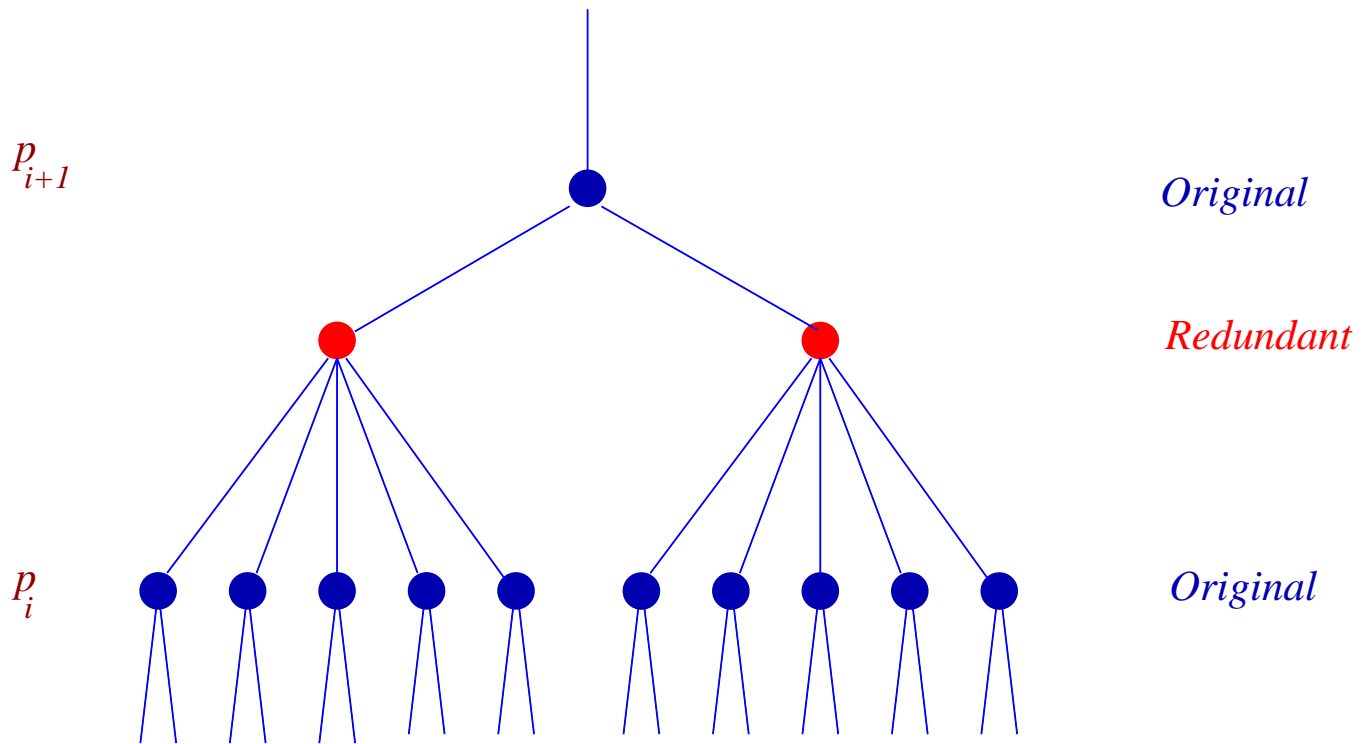
(Luby, Mitzenmacher, Shokrollahi, Spielman, Stemann)

Proof: uses probability theory (martingales, tail inequalities, large deviation results).



p=0.433
p=0.425
p=0.4

Example: (3, 6)-regular graph



$$p_{i+1} = p_0(1 - (1 - p_i)^5)^2, \text{ so}$$

$$p_0(1 - (1 - p_i)^5)^2 < p_i$$

guarantees successful decoding.

General Case

λ_i, ρ_i fraction of edges of degree i on the left and the right hand side.

$$\lambda(x) := \sum_i \lambda_i x^{i-1} \text{ and } \rho(x) := \sum_i \rho_i x^{i-1}.$$

Successful decoding for erasure probability p_0 if and only if

$$p_0 \lambda(1 - \rho(1 - x)) < x$$

for all $x \in (0, p_0)$.

Design of Graphs: Linear Programming

Fix right hand side $\rho(x)$, and find best left hand side $\lambda(x)$ using the condition

$$p_0 \lambda (1 - \rho(1 - x)) < x$$

on $(0, 1)$ using linear programming.

Once best left hand side found, fix left hand side and use dual condition

$$\rho (1 - p_0 \lambda (1 - x)) > x$$

on $(0, 1)$ with linear programming to find best right hand side.

Iterate!

Capacity Achieving Codes

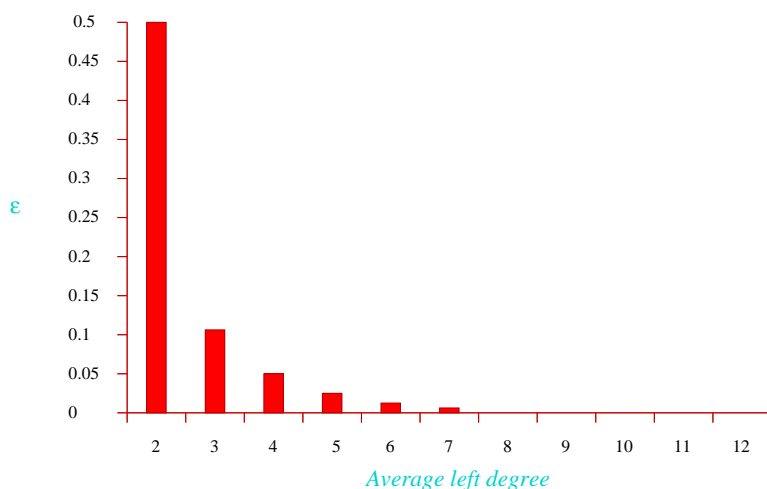
Highly irregular graphs give for any **rate R** sequences of codes that **achieve capacity** of erasure channel asymptotically.

Degree structure? Fix **design parameter D** .

$$\lambda(x) := \frac{1}{H(D)} \left(x + \frac{x^2}{2} + \dots + \frac{x^D}{D} \right)$$

$$\rho(x) := \exp(\mu(x-1))$$

$H(D)$ is **harmonic sum** $1 + 1/2 + \dots + 1/D$ and $\mu = H(D) / (1 - 1/(D+1))$.



Right Regular Sequences

Fix $a \geq 2$, $n \geq 2$, $\alpha = 1/(a - 1)$.

$$\lambda_{a,n}(x) := \frac{\sum_{k=1}^{n-1} \binom{\alpha}{k} (-1)^{k+1} x^k}{1 - n \binom{\alpha}{n} (-1)^{n+1}},$$
$$\rho_a(x) := x^{a-1}.$$

give codes of rate

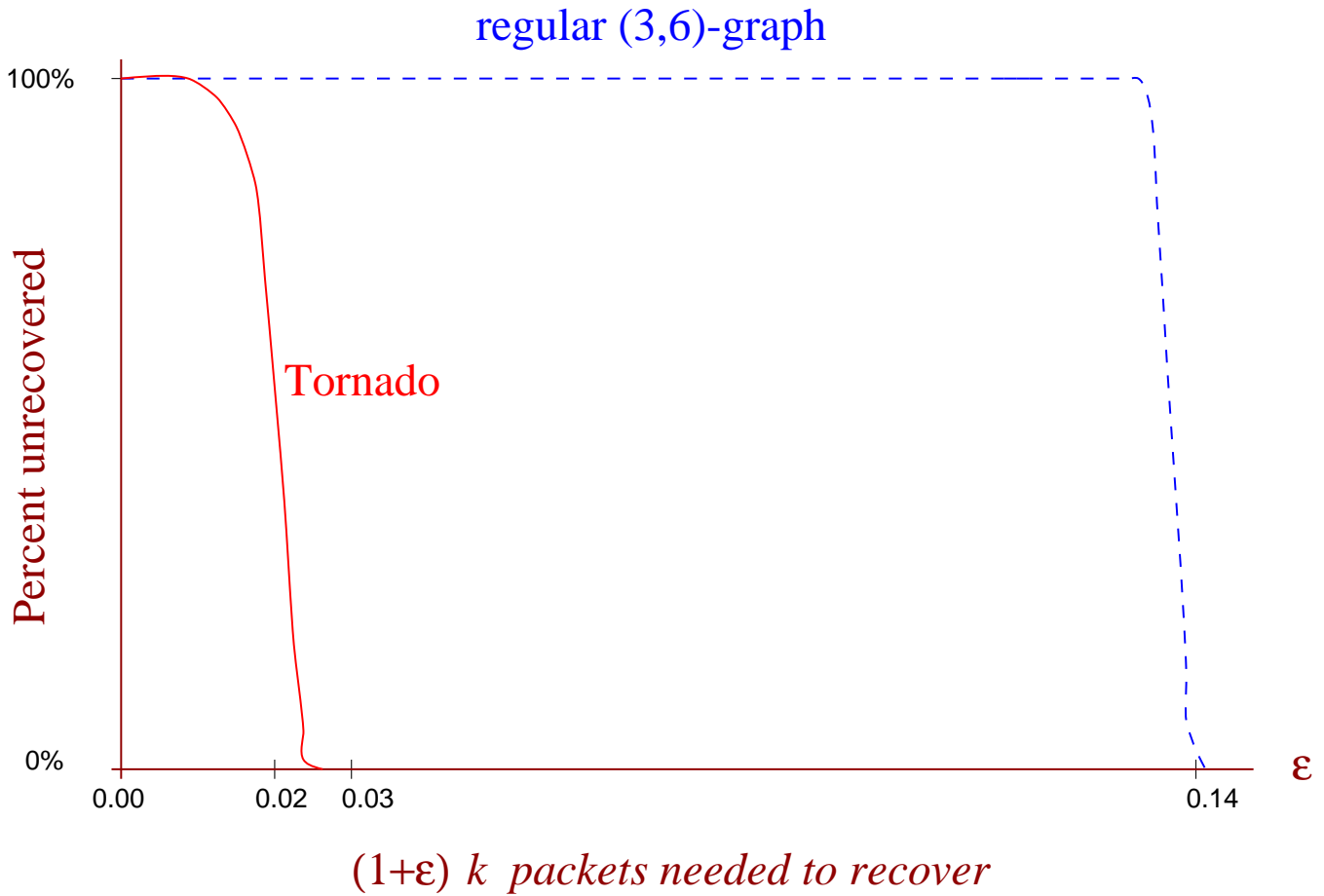
$$1 - \frac{\alpha - \binom{\alpha}{n} (-1)^{n+1}}{\alpha - n \binom{\alpha}{n} (-1)^{n+1}}.$$

In both cases: ε close to channel capacity needs graphs of average degree $\log(1/\varepsilon)$.

Optimal!

(Shokrollahi, 1999)

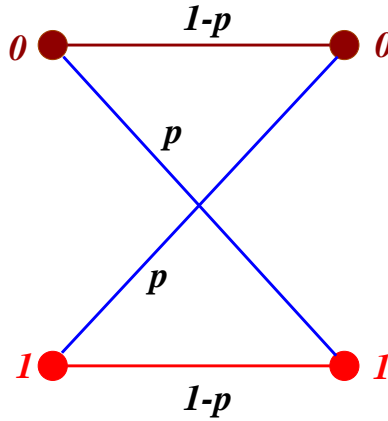
Experiments



640,000 message packets were encoded into 1,280,000 packets using a code with average left degree 8.

Extensions

Binary symmetric channel with hard-decision decoding

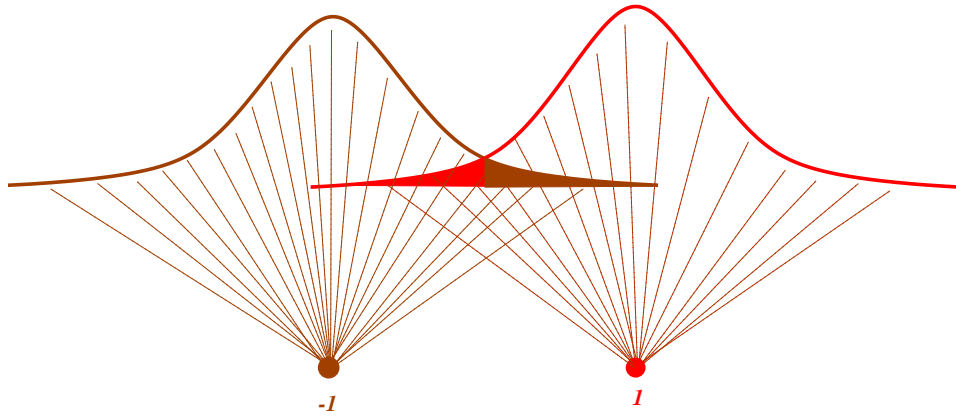


Luby, Mitzenmacher, Shokrollahi, Spielman, STOC 1998

Proof methods were vastly generalized by Richardson & Urbanke.

Extensions

AWGN channel: empirical results

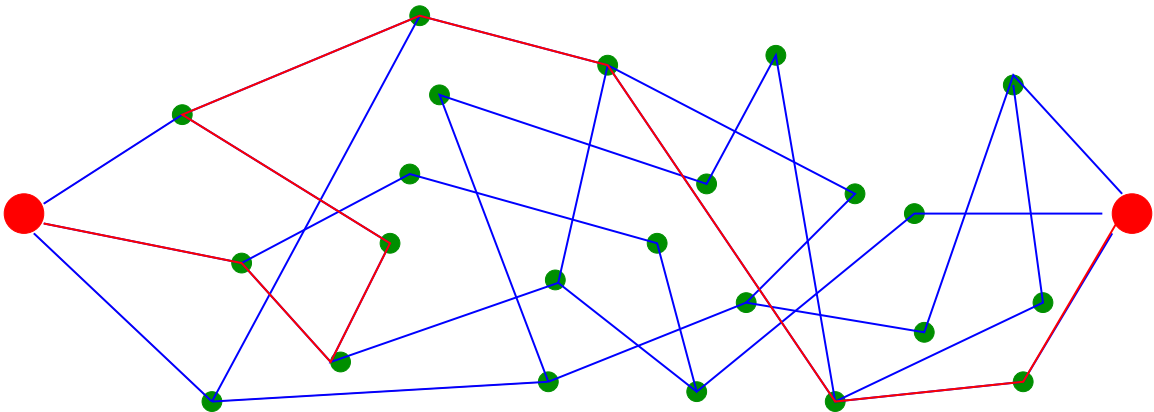


Luby, Mitzenmacher, Shokrollahi, Spielman, ISIT 1998

Applications

Packets in Networks

Data sent in a network is divided into **packets** which are routed through the network from a sender to a recipient.



Packet Loss

Each packet has an **identifier**.

Packets can get **lost** or **corrupted**.

Corruption is checked via **checksums**.

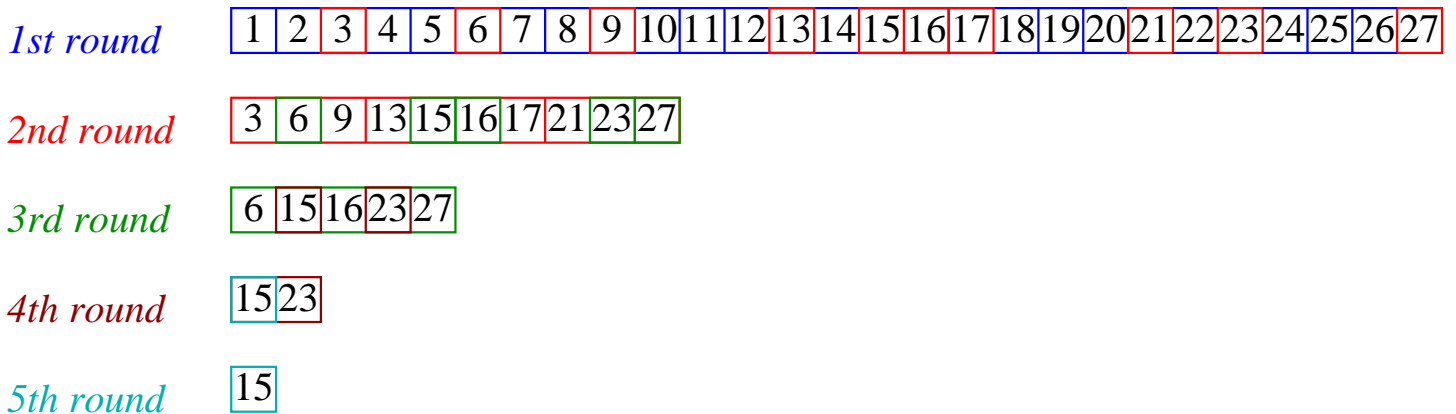
Corrupted packets are regarded as **lost**.

May without loss of generality only concentrate on **losses**.

Retransmission

In many communication protocols lost packets are re-transmitted.

Process is repeated until all packets are received.

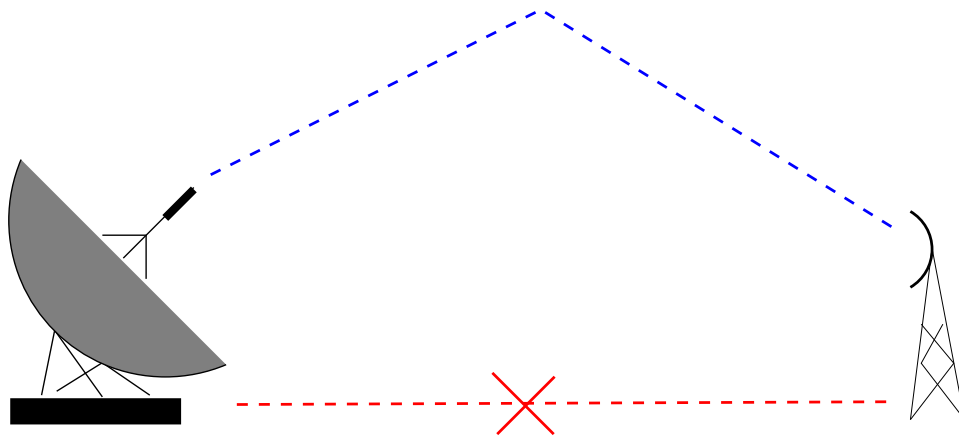


Retransmission Protocols

Requires existence of **feedback channel**.

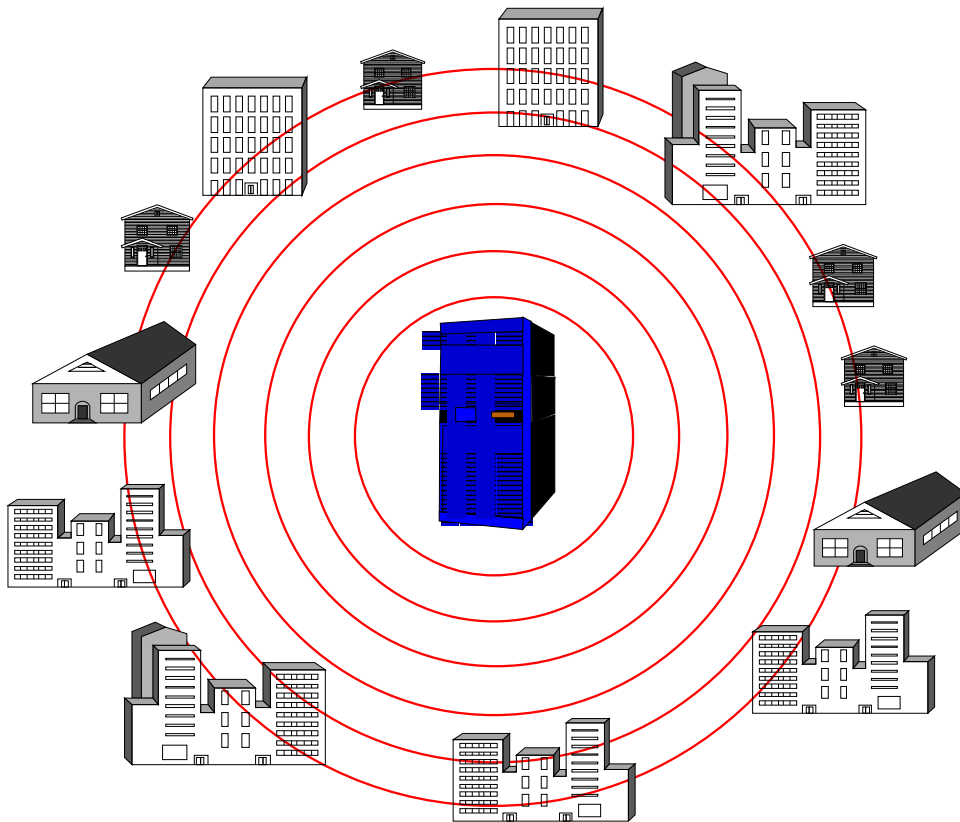
May **not exist**, or maybe **too expensive**.

Example: **satellite links**.



Retransmission Protocols

Not good enough in **broadcast** application: **one server**, **many clients**. Request for retransmission leads to huge **server load**.



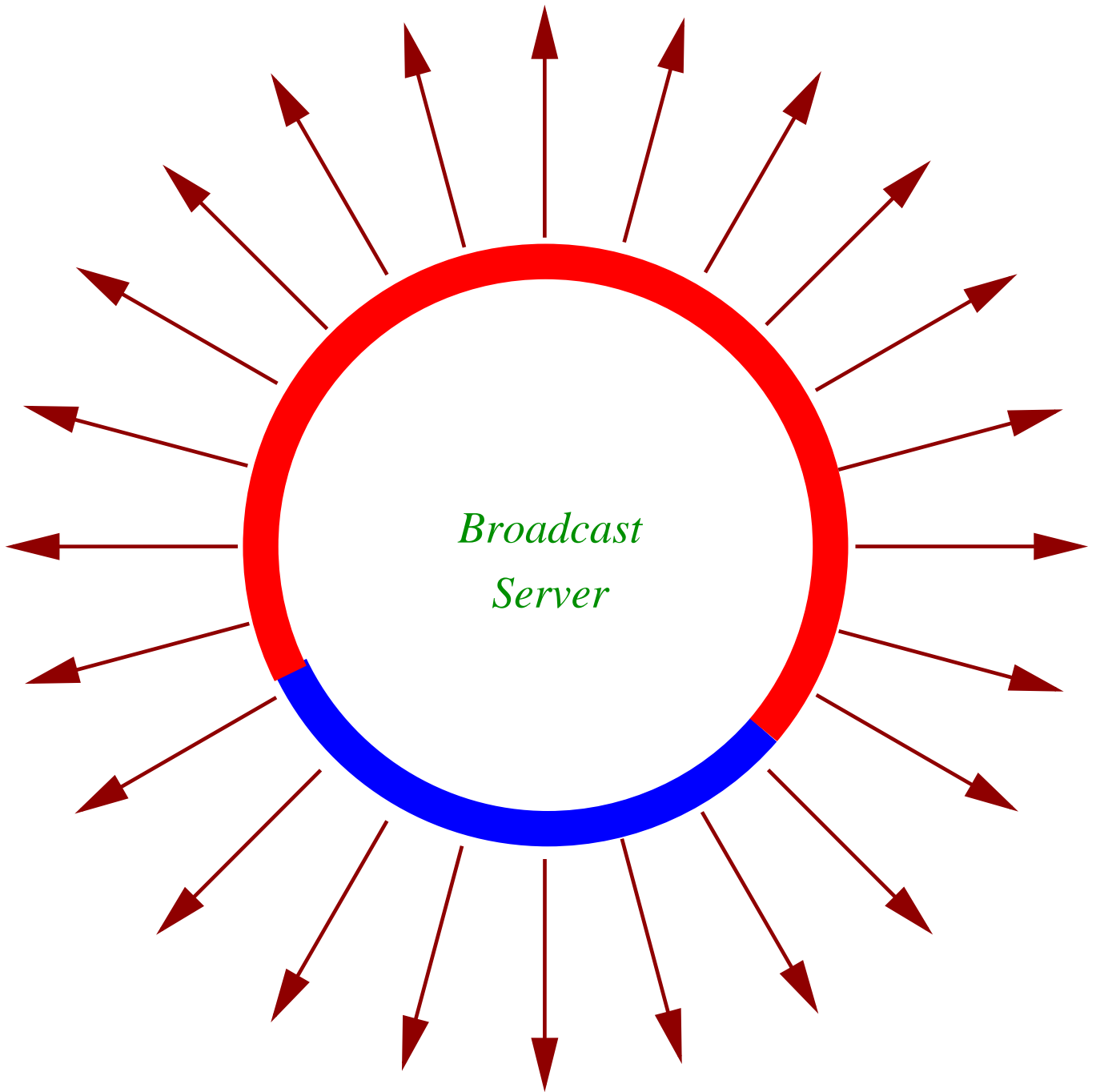
Bulk Data Distribution

Distribution of bulk data to a **large** number of clients.

Want

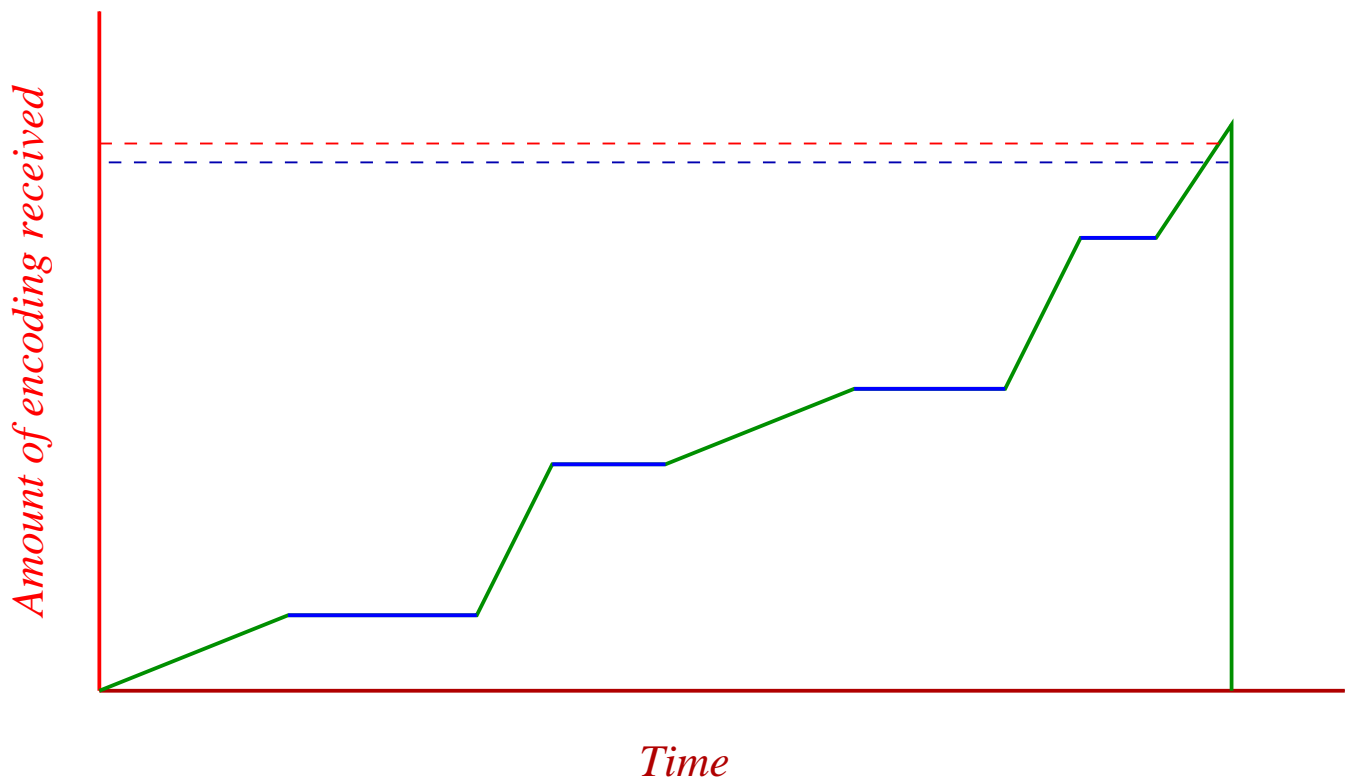
- **fully reliable,**
- **low network overhead,**
- **support vast number of receivers with heterogeneous characteristics**
- **users want to access data at times of their choosing and these access times overlap.**

Our solution



Our solution

Client joins multicast group until **enough** of the encoding has been received, and then decodes to obtain original data.



\$\$ Applications \$\$

- video and financial information broadcast
- database replication
- popular web site access

Done by

- Tornado codes
- Digital Fountain
- <http://www.dfountain.com>