

In-Network Coding for Resilient Sensor Data Storage and Efficient Data Mule Collection

Michele Albano

**Instituto de telecomunicacoes,
Aveiro, Portugal**

Jie Gao

**Stony Brook University, Stony
Brook, USA**

Data collection

- Gather sensor data to a base station
- Traditional approach:
 - Build aggregation tree rooted at the sink
 - E.g., TinyDB family.
- Problem:
 - Sensors near sink are overloaded.
 - Sink disconnected from the network prematurely.
- Our approach: **mobile sink, or, data mule.**

Data collection using mules

- Network of n nodes, with k of them have data.
- Data mules tour around to pick up data.
- **Challenge #1:** path planning
 - TSP or multi-TSP problem, NP-hard.
 - Random walk.
 - Coupon collection
 - $O(n^2)$ hops to cover all nodes.

Data collection using mules

- Network of n nodes, with k of them have data.
- Data mules tour around to pick up data.
- **Challenge #2:** information brokerage
 - Mule is not aware of the data nodes.
 - In any predetermined scheme, mule may visit many nodes without data.
 - Need data processing, i.e., data nodes initiate certain actions

Our approach: in-network coding

- Sensor data are stored in encoded format in the network.
- Original data: **symbols** s_1, s_2, \dots, s_k .
- Coded data: **codewords** w_1, w_2, \dots, w_n .
- We use **random linear coding**:
 - Codeword = random linear combination of symbols, $w_j = \sum_{i=1}^k s_i \lambda_{ij}$.
 - Every node keeps a different codeword.

Data mule collection and decoding

- Data mule visit **any** k nodes and collect k codewords
- If the coefficient matrix is **full rank**, the symbols can be recovered.

$$S \begin{bmatrix} \lambda_{ij} \end{bmatrix} = W$$

- Main focus: how to build codewords in **distributed** and **communication efficient** manner?

Gossip algorithms

- In a round, each node:
 - Selects **another node** randomly
 - **Exchanges** information via multi-hop routing
 - Repeats every round
- Simple
- Distributed
- Robust to link dynamics, transmission errors

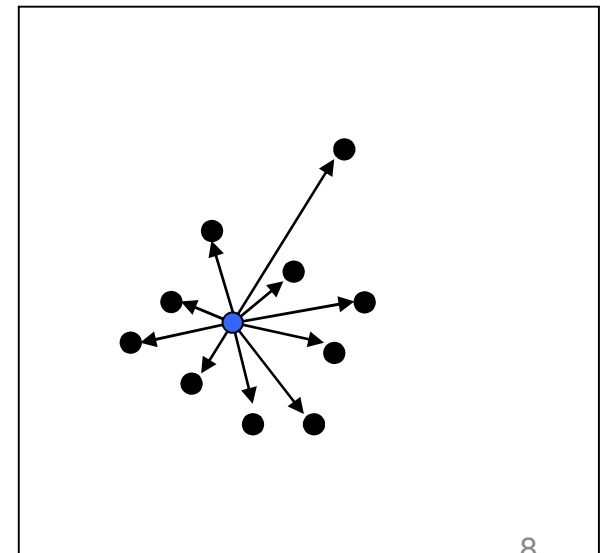
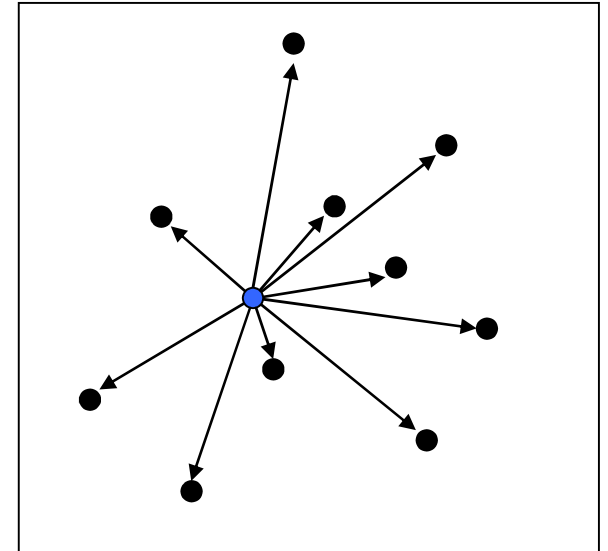
Types of gossip

- Uniform/Geographic gossip
 - Select a node q **uniformly randomly** and gossip

[Dimakis, Sarwate, Wainwright, IPSN 06]

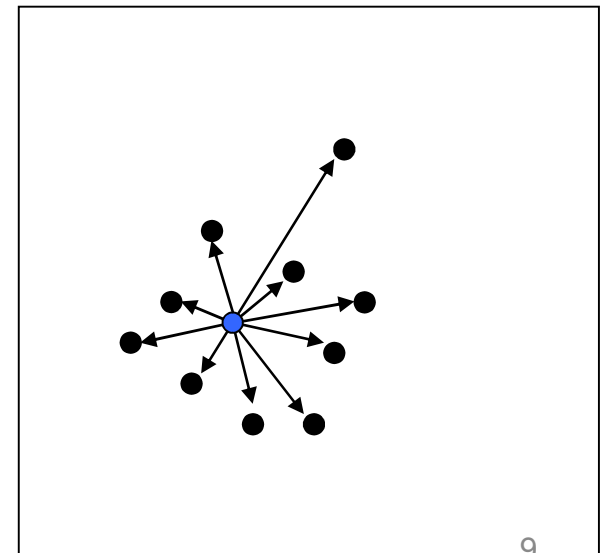
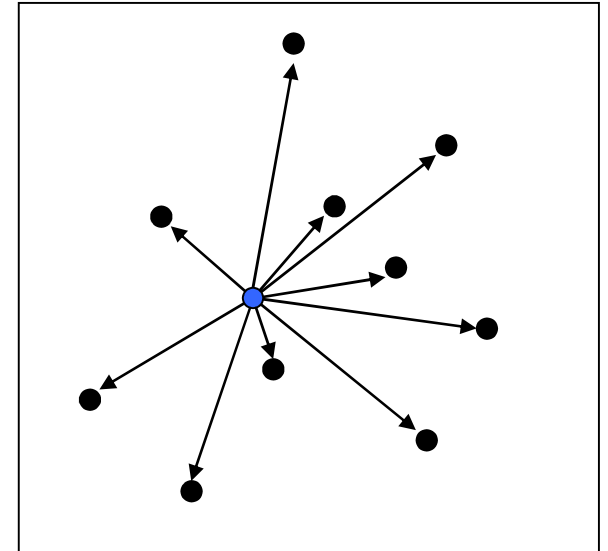
- Spatial Gossip
 - Select a node q at distance r with probability $1/r^\alpha$.

[Kempe, Kleinberg, Demers, STOC 01]



Communication cost

- Uniform/Geographic gossip
 - Cost per step $\sim O(n\sqrt{n})$
 - # rounds for a message to reach everyone $\sim O(\log n)$
- Spatial Gossip
 - Prob= $1/r^2$, cost per step $\sim O(n\sqrt{n})$
 - Prob= $1/r^3$, cost per step $\sim O(n \log n)$
 - # rounds for a message to reach everyone $\sim O(\log n)$



Spatial gossip for in-network coding

- Nodes proceed in synchronous rounds
- Each node p :
 1. Multiply its current data by a **random coefficient**;
 2. Send it to node q chosen by **spatial** distribution;
 3. Store **linear combination** of all data received.
- Use total $O(\log^{3.4}n)$ rounds.
- Total communication cost = $O(n \log^{4.4}n)$

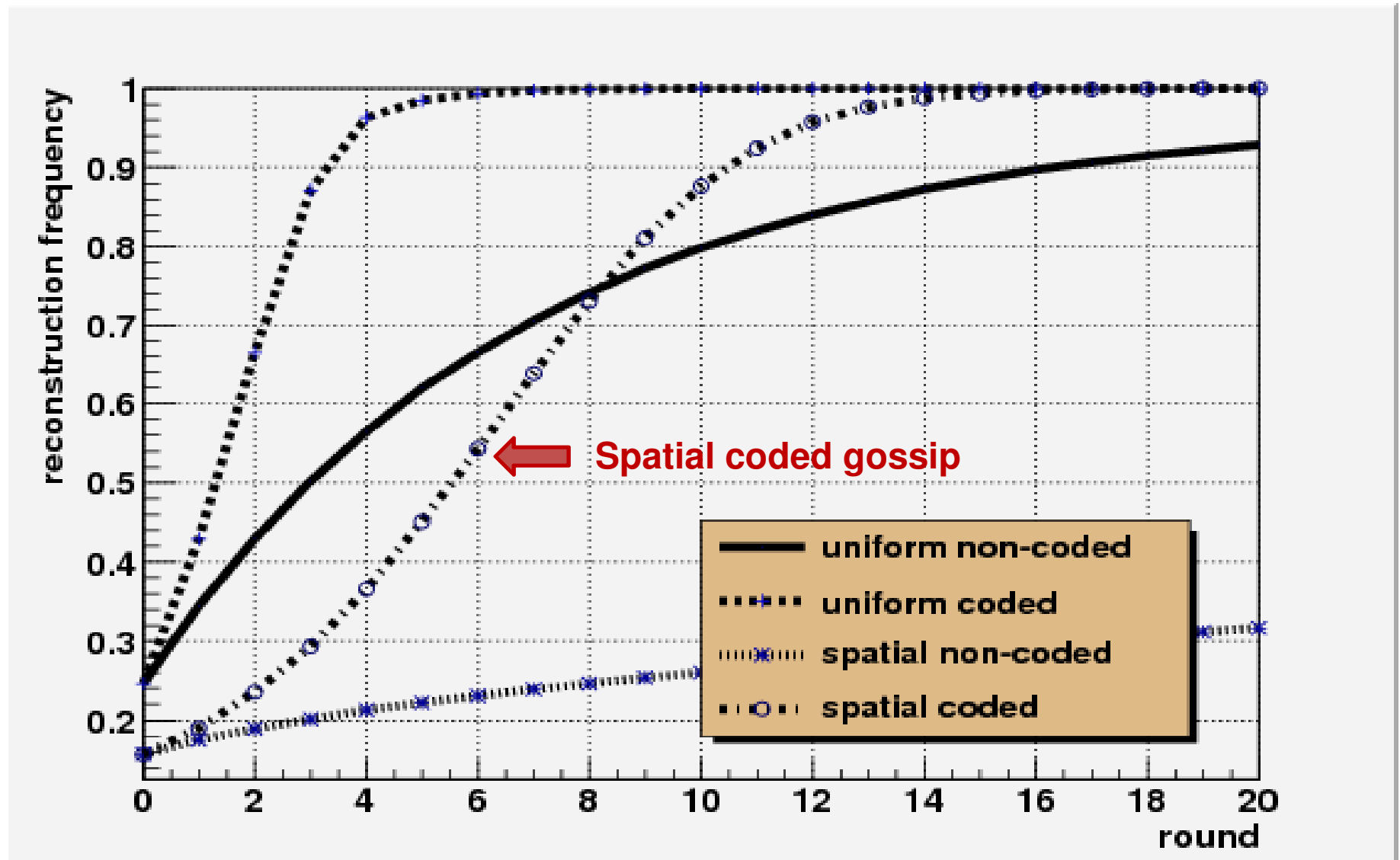
Theorems

- Theorem: The codeword at each node has a non-zero coefficient for any symbol w.h.p.
- Theorem: Any k codewords can decode for the original symbols with prob $\rightarrow 1$ if $n \rightarrow \infty$.
- Mule can successfully decode by picking up any k codewords!

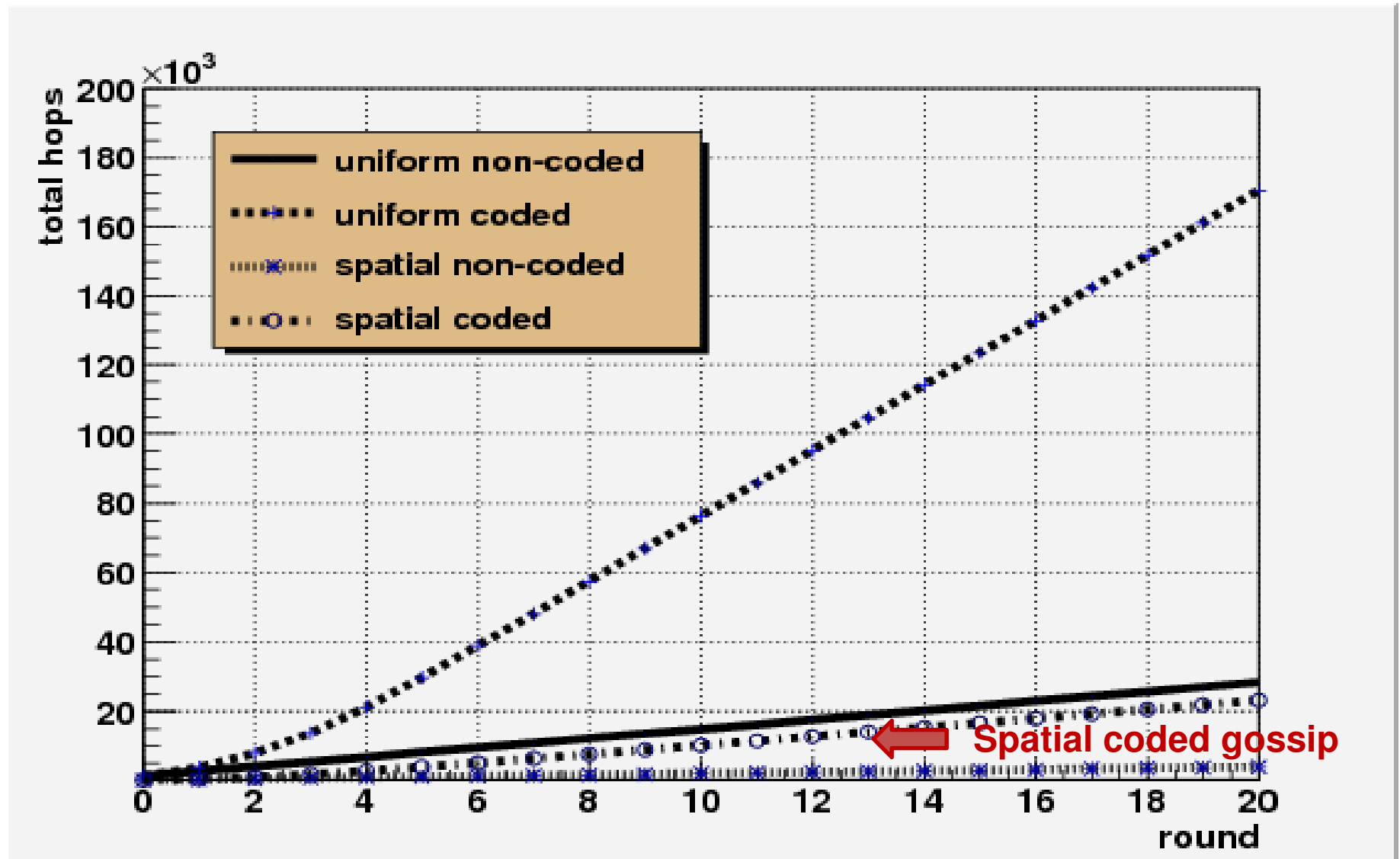
Simulations

- 700 nodes in a square region
- Compare 4 schemes:
 - Uniform gossip v.s. spatial gossip
 - Disseminate codewords v.s. symbols.
- Major metrics to evaluate:
 - Decoding success rate
 - Communication cost

Frequency of correct reconstruction



Routing cost in hops



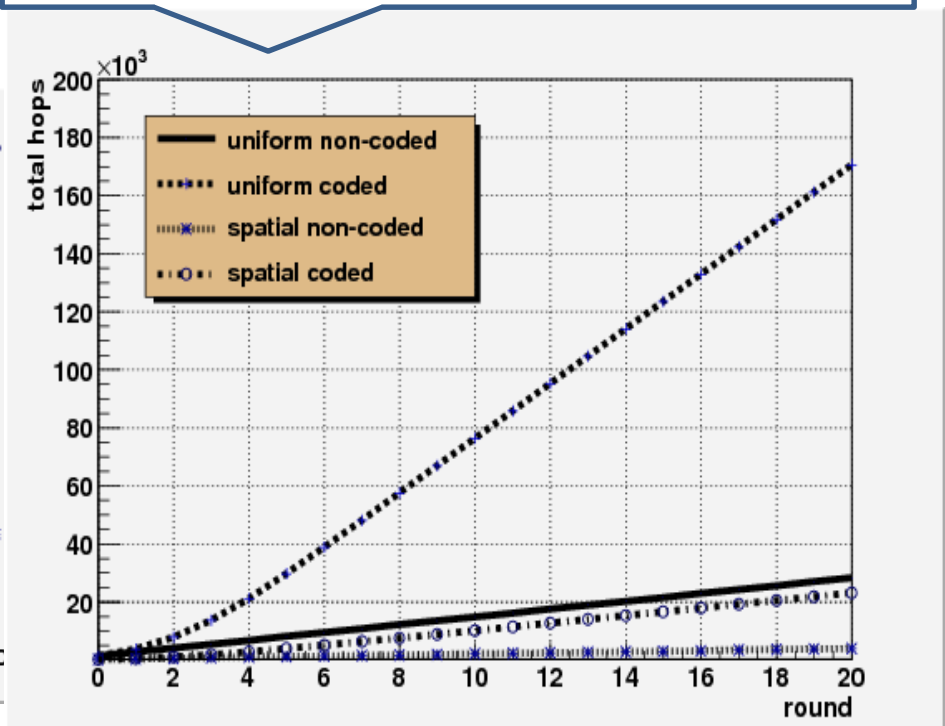
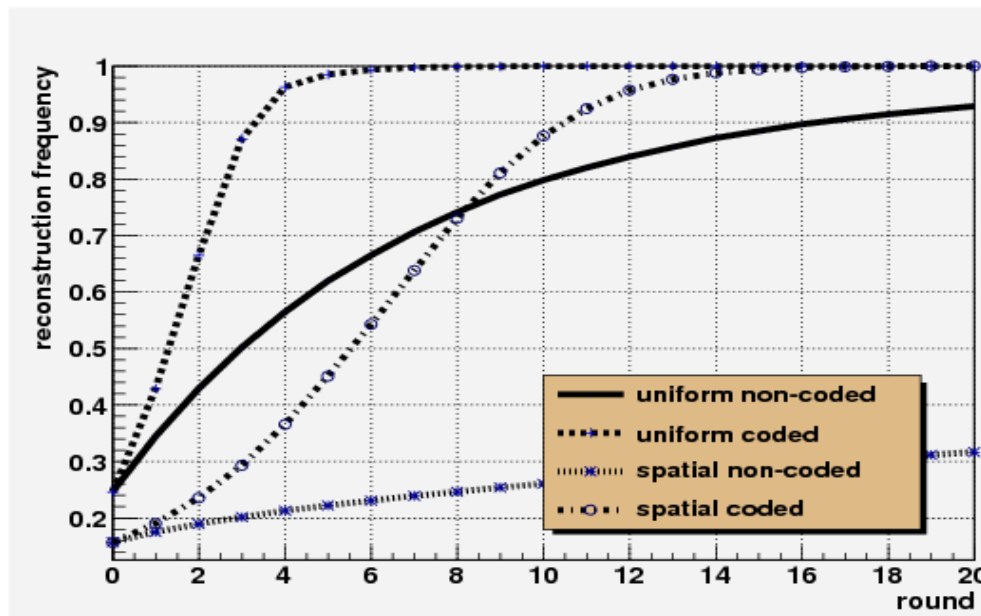
Frequency of correct reconstruction

gossip rounds for correct reconstruction:

- Uniform non-coded – 230 rounds
- Uniform coded – 10 rounds
- Spatial coded – 20 rounds
- Spatial non-coded – Too many rounds

Total routing cost for correct reconstruction:

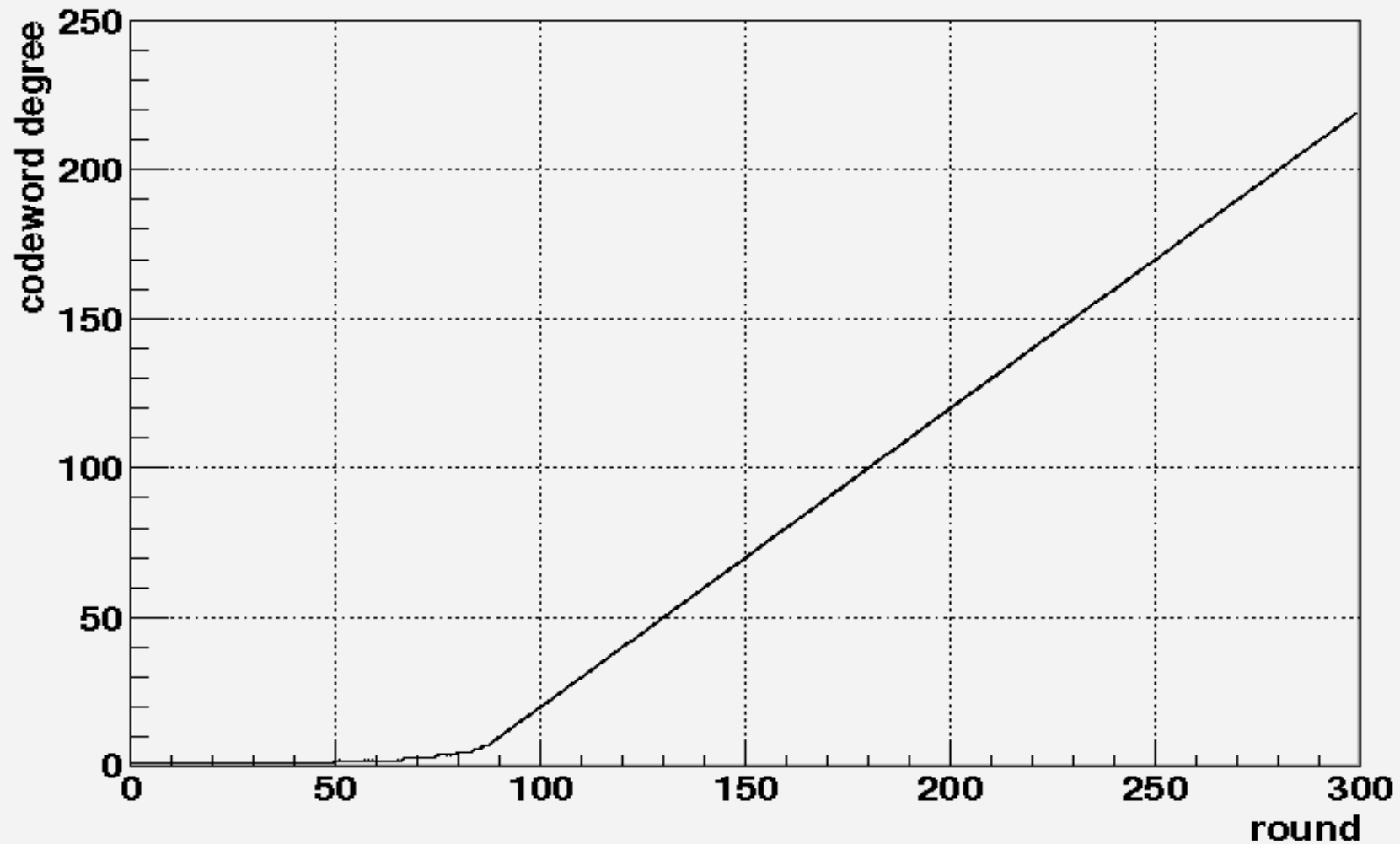
- Uniform non-coded – 353,000 hops
- Uniform coded – 90,000 hops
- Spatial coded – 30,000 hops
- Spatial non-coded – too high



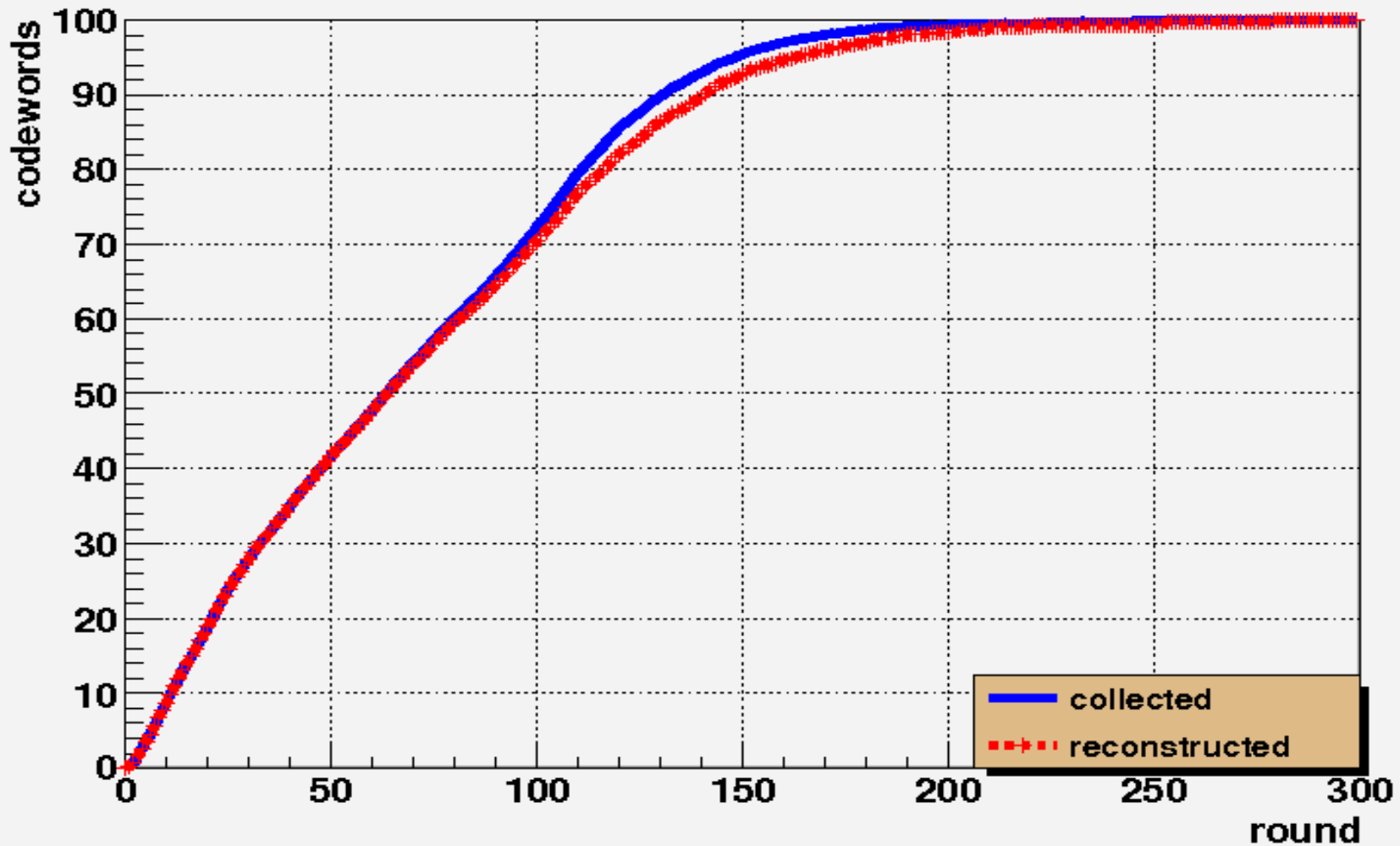
Online reconstruction

- Decode symbols as soon as possible.
- Each round is composed of
 - ONE gossip round
 - ONE data collection step of the mule
- Degree of a codeword: # symbols with non-zero coefficients.
 - grows exponentially in spatial gossip.
 - For online construction, degree should grow much slower.

Our heuristic for codeword degree



Collected codewords vs reconstructed symbols



Conclusions & Future Work

- Combining spatial gossip with coding results in an efficient data collection mechanism
- It is possible to implement online data reconstruction
- What is the best threshold for code degree for online collection?

Thank you!

- Questions and comments?