Reconstruction of Network Coded Sources From Incomplete Datasets

Nikolaos Thomos
in collaboration with
E. Bourtsoulatze and P. Frossard
Efficient data delivery with network coding

SERVER 1

SERVER 2

NETWORK NODE

CLIENT 1

CLIENT 2

CLIENT 3

processing
Efficient data delivery with network coding

- High throughput
- Reduced delay
- Efficiency

- Robustness to losses
- Deployment in distributed systems

University of Bern – Communication and Distributed Systems (CDS)
http://cds.unibe.ch
Decoding from incomplete network coded data

- No guarantee that a sufficient number of packets reach the clients
  - exact source recovery is not feasible
  - approximate reconstruction may be meaningful

University of Bern – Communication and Distributed Systems (CDS)
http://cds.unibe.ch
Outline

• Problem formulation
• Theoretical analysis
• Practical decoder
• Simulation results
• Conclusions
Decoding from incomplete NC data
Decoding from incomplete NC data

- $\text{rank}(A) = N$: exact reconstruction
Decoding from incomplete NC data

- \(\text{rank}(A) = N\): exact reconstruction
- \(\text{rank}(A) < N\): approximate reconstruction using source priors
Decoding from incomplete NC data

- Goal:
  - analyze the performance
  - design a practical decoding algorithm
MAP decoding

- Maximum a posteriori decoding rule

\[
\hat{x}^* = \arg\max_{\hat{x} \in \hat{X}^N} \mathbb{1}\{A, \hat{y}\}(\hat{x}) \hat{f}(\hat{x})
\]

- \(\hat{x}^*\): transmitted sequence
- \(\hat{x}\): decoded sequence
- \(f(\hat{x}^*) \geq f(\hat{x})\)
MAP decoding

- Maximum a posteriori decoding rule

\[ \hat{x}^* = \arg \max_{\hat{x} \in \hat{X}^N} \mathbb{1}\{A, \hat{y}\}(\hat{x}) \hat{f}(\hat{x}) \]

- Upper bound on the block error probability

\[ P_e \leq \min_{0 \leq \rho \leq 1} q^{-\rho L} \left[ \sum_{\mathbf{x} \in \hat{X}^N} f(\mathbf{x})^{\frac{1}{1+\rho}} \right]^{1+\rho} \]

\( \hat{x} : \) transmitted sequence
\( \hat{x}^* : \) decoded sequence
\( \hat{f}(\hat{x}^*) \geq \hat{f}(\hat{x}) \)
Illustration of the bound

- Let \( f(x) = \prod_{n=1}^{N} f(x_i|x_{i-1}) \)
  \[
f(x_1) = \frac{1}{q}
\]
  \[
f(x_i|x_{i-1}) = \frac{1}{K} \frac{1 - p}{1 + p} p^{|x_i-x_{i-1}|}, \quad p \in (0, 1)
\]

- Decoding performance vs Correlation

- Decoding performance vs Finite field size
Correlation model

- Linearly correlated sources
- Correlation coefficient
  \[ \rho_{ij} \geq 0 \]
- Correlation expressed as correlation noise

\[ W_m = X_i - X_j \]
\[ g_m(w) : \mathcal{W} \to [0, 1] \]
Factor graph

\[
\begin{align*}
\mathbb{1}\{A_1, \hat{y}_1\}(\hat{x}) & \quad \mathbb{1}\{A_l, \hat{y}_l\}(\hat{x}) & \quad \mathbb{1}\{A_L, \hat{y}_L\}(\hat{x}) \\
\hat{x}_1 & \quad \cdots & \quad \hat{x}_n & \quad \cdots & \quad \hat{x}_N
\end{align*}
\]

Variable nodes  $\leftrightarrow$  Source symbols
Check nodes  $\leftrightarrow$  Network coded symbols
Edges  $\leftrightarrow$  Coding coefficients
Initialization

- Input
  - coding matrix $A$
  - network coded symbols $\hat{y}$
  - adjacency matrix $C_G$
  - pmf $g_m(w)$ and $\hat{f}_n(\hat{x})$

• Messages from variable nodes
  $$q_{nl}(a) = \hat{f}_n(a) \text{ or } q_{nl}(a) = \frac{1}{q}$$

• Messages from check nodes
  $$r_{ln}(a) = 1$$

University of Bern – Communication and Distributed Systems (CDS)
http://cds.unibe.ch
Message passing

• Messages from variable nodes to check nodes

\[ q_{nl}(a) = \alpha_{nl} \prod_{l' \in \mathcal{L}(n) \setminus l} r_{l' n}(a) \]
Message passing

\[ r_{ln}(a) = \sum_{\hat{x} : \hat{x}_n = a} \mathbb{1}_{\{A_l, \hat{y}_l\}}(\hat{x}) \prod_{n' \in \mathcal{N}(l) \setminus n} \mu_{n'l}(\hat{x}_{n'}) \]

\[ \mu_{n'l}(\hat{x}_{n'}) = \begin{cases} g_{n'n}(\hat{x}_{n'} - a)q_{n'l}(\hat{x}_{n'}), & \text{if } \rho_{nn'} \neq 0 \\ q_{n'l}(\hat{x}_{n'}), & \text{if } \rho_{nn'} = 0 \end{cases} \]
Reconstruction

If \( \hat{y} = A\hat{x}^* \), valid solution

Otherwise, repeat until \( k > k_{max} \) and \( \hat{x}_n^* = E[\hat{X}_n] \)
Complexity

- Computational complexity per node and per iteration:
  - variable nodes: $\mathcal{O}(d_v q)$
  - check nodes: $\mathcal{O}(d_c q^2)$

- Reduced complexity with operations in transform domain: $\mathcal{O}(q^2) \rightarrow \mathcal{O}(q \log_2 q)$

- Overall, the complexity depends on:
  - finite field size: $q$
  - number of sources: $N$
  - number of NC symbols: $L$
  - density of the coding matrix: $A'$
Synthetic signals

- \( N = 20 \) sensors uniformly distributed over a unit square

- \( N \)-dimensional random vector
  \[
  s = (s_1, s_2, \ldots, s_N)^T \sim \mathcal{N}(0, \Sigma)
  \]

- Covariance matrix
  \[
  \Sigma = \begin{bmatrix}
  1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1N} \\
  \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2N} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  \rho_{N1} & \rho_{N2} & \rho_{N3} & \cdots & 1
  \end{bmatrix}
  \]

- Correlation noise \( W_m \) between sensors \( S_i \) and \( S_j \)
  \[
  g_m(w) = \sum_{x_i, x_j : x_i - x_j = w} \int \int p(s_i, s_j) ds_i ds_j
  \]

  \[
  \rho_{ij} = e^{-\beta d_{ij}}, \quad \beta > 0
  \]
Decoding performance

- The influence of correlation on the decoding performance

3-bit uniform quantizer \((q = 8)\)  
4-bit uniform quantizer \((q = 16)\)
Decoding performance

- The influence of field size on the decoding performance

| $q$ | $\beta = 0.01$ | $\beta = 0.05$ |
|-----|----------------|----------------|
| 8   | 3-bit uniform quantizer | 3-bit uniform quantizer |
| 16  |                   |               |
| 32  |                   |               |
| 64  |                   |               |

University of Bern – Communication and Distributed Systems (CDS)
http://cds.unibe.ch
Correlated images

- $N = 15$ consecutive images from video sequences *Silent* and *Foreman* in QCIF format

- original images quantized to $n$ bits per pixel
  $x_i \in [0, 2^n - 1]$

- Laplacian correlation noise $W_m$ between images $I_i$ and $I_j$

\[
g_m(w) = \frac{1 - p_m}{1 + p_m} p_m |w|, \quad p_m \in (0, 1)
\]
Decoding performance is worse for sequences with higher motion due to lower correlation.
Visual quality - Silent

• Images reconstructed from $L = 13$ out of $N = 15$ network coded symbols

  original sequence  reconstructed sequence  error sequence

• Decoding errors occur in regions with higher motion, e.g., around the edges
Visual quality - Silent

- Images reconstructed from $L = 13$ out of $N = 15$ network coded symbols

original sequence  reconstructed sequence  error sequence

- Decoding errors occur in regions with higher motion, e.g., around the edges
Visual quality - Silent

• Images reconstructed from $L = 13$ out of $N = 15$ network coded symbols

original sequence  reconstructed sequence  error sequence

• Decoding errors occur in regions with higher motion, e.g., around the edges
Visual quality - Foreman

• Images reconstructed from $L = 13$ out of $N = 15$ network coded symbols

original sequence  reconstructed sequence  error sequence

• Decoding errors occur in regions with higher motion, e.g., around the edges
Visual quality - Foreman

- Images reconstructed from $L = 13$ out of $N = 15$ network coded symbols

- Decoding errors occur in regions with higher motion, e.g., around the edges
Visual quality - Foreman

• Images reconstructed from $L = 13$ out of $N = 15$ network coded symbols

original sequence  reconstructed sequence  error sequence

• Decoding errors occur in regions with higher motion, e.g., around the edges
Conclusions

- Study the problem of decoding of network coded data from incomplete data sets
- Derive the bound of the MAP decoder used to approximate the source data
- Propose a practical belief propagation decoder
- Jointly consider network and source constraints
- Demonstrate the performance for synthetic signal and image sequences
Thank you!