Binary Locally Repairable Codes from Complete Multipartite Graphs

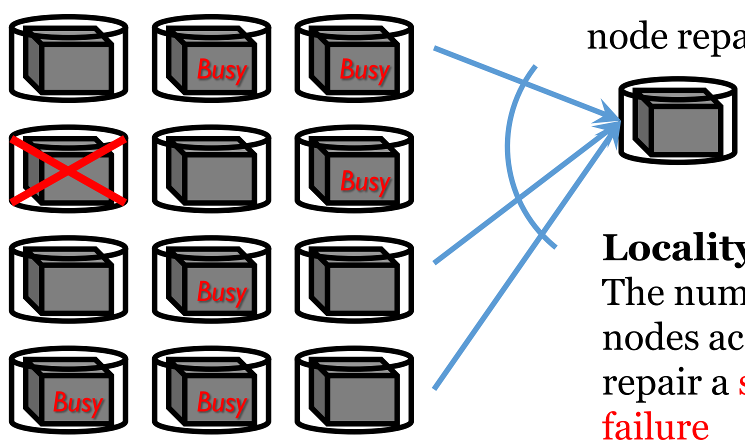
Jung-Hyun Kim, Mi-Young Nam, and Hong-Yeop Song

International Conference on ICT Convergence
Oct. 28-30, 2015, Jeju, Korea

▪ Distributed Storage System (DSS)

Frequent node failure

Node failure :
A disk fault or a node in use



Locality :
The number of nodes accessed to repair a single node failure

P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," IEEE Trans. Inform. Theory, vol. 58, no. 11, pp. 6925-6934, Nov. 2012.

▪ Locally repairable code (LRC)

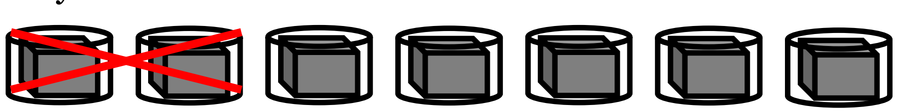
- Codes with good (small) locality

Locality (Generalized definition)
- **ℓ-locality (r_ℓ)** : locality for ℓ symbols repair

* 1-locality (r_1) is the same with "locality" in the previous definition

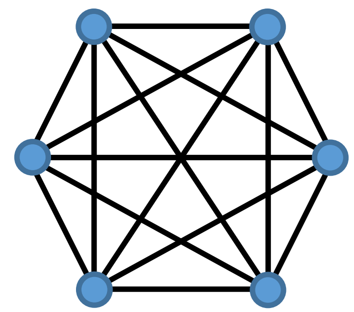
Symbol set

Codeword :



A. S. Rawat, A. Mazumdar, and S. Vishwanath, "Cooperative local repair in distributed storage," arXiv Preprint arXiv:1409.3900, 2014.

▪ Complete Graph (CG) codes



K_6 complete graph ($k = 6$)

$$\left(\begin{array}{cccccc|ccccccccccccccc}
1&0&0&0&0&0&1&1&0&1&0&0&1&0&0&0&1&0&0&0&0\\
0&1&0&0&0&0&1&0&1&0&1&0&0&1&0&0&0&1&0&0&0\\
0&0&1&0&0&0&0&1&1&0&0&1&0&0&1&0&0&0&1&0&0\\
0&0&0&1&0&0&0&0&0&1&1&1&0&0&0&1&0&0&0&1&0\\
0&0&0&0&1&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&1\\
0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1
\end{array}\right)$$

Generator matrix of $[21, 6, 6]_2$ code

Theorem: It has $d_{min} = k$ and
1) $r_1 = 2$ for $k = d_{min} \geq 2$.
2) $r_2 = 3$ for $k = d_{min} \geq 3$.
3) $r_\ell \leq \min(2\ell, k)$ for $d_{min} \geq 2$ and $\ell = 1, 2, 3, \ldots$

▪ Complete Multipartite Graph (CMG) codes
(p-partite)

Note that p can be any (positive) divisor of k.

- $p = k$ (CG codes) $\leftrightarrow$ **lowest code rate**
- p is the smallest non-trivial prime factor of k. (CMG codes) $\leftrightarrow$ **highest code rate**
- $p = 1 \leftrightarrow G = I$ and $d_{min} = 1$ (trivial)

- **Joint locality**
  - A set of numbers of symbols for repairing various erasure patterns of symbols

Q1. Can we design a Binary LRC with joint locality (2, 3) or (2, 4)?

One choice would be binary simplex codes with the parameter $n = 2^k - 1, k, d_S = 2^{k-1}$

$$G_{S4} = \begin{pmatrix} 1000 & 111000 & 1110 & 1 \\ 0100 & 100110 & 1101 & 1 \\ 0010 & 010101 & 1011 & 1 \\ 0001 & 001011 & 0111 & 1 \end{pmatrix} \text{ for } k = 4$$

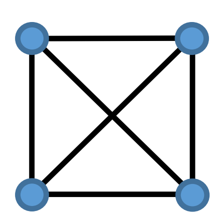
$r_\ell \leq \ell + 1$ (Rawat-14)

- It has $(r_1, r_2) = (2, 3)$. – proof is straightforward

- **Simplex codes**
  - Code rate : $R_S = \frac{k}{2^k - 1}$ (VERY LOW)

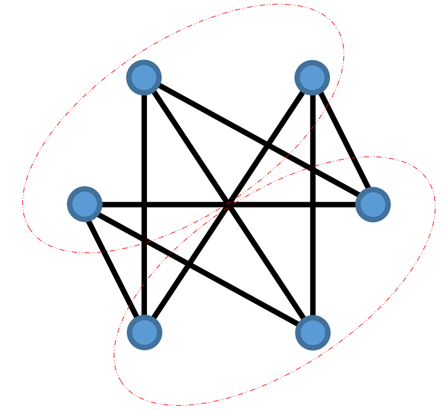
Q2. Can we improve the rate maintaining joint locality (2, 3) or (2, 4)?

$$G = \begin{pmatrix} 1000 & 111000 & 1110 & 1 \\ 0100 & 100110 & 1101 & 1 \\ 0010 & 010101 & 1011 & 1 \\ 0001 & 001011 & 0111 & 1 \end{pmatrix} \text{ for } k = 4$$



- This code STILL has $(r_1, r_2) = (2, 3)$.
- How to describe the code ?

✓ Its generator matrix has all the columns of weight 1 and weight 2 ONLY.



Complete 2-partite graph ($p = 2, \ k = 6$)

$$\begin{pmatrix} 1&0&0&0&0&0&1&1&1&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&1&1&1&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0&0&0&1&1&1 \\ 0&0&0&1&0&0&1&0&0&1&0&0&1&0&0 \\ 0&0&0&0&1&0&0&1&0&0&1&0&0&1&0 \\ 0&0&0&0&0&1&0&0&1&0&0&1&0&0&1 \end{pmatrix}$$

Generator matrix of $[15, 6, 4]_2$ code

- **Code rate :** $R = \frac{2}{k - \frac{k}{p} + 2} \geq R_S = \frac{k}{2^k - 1}$
- Minimum distance : $d = k - \frac{k}{p} + 1 \leq d_S = 2^{k-1}$

Theorem: It has $d_{min} = k - \frac{k}{p} + 1$ and

1) $r_1 = 2$ for $d_{min} \geq 2$.

2) $r_2 = \begin{cases} 3, & \text{for } p = k \text{ and } d_{min} \geq 3, \\ 4, & \text{for } p < k \text{ and } d_{min} \geq 3. \end{cases}$

3) $r_\ell \leq \min(2\ell, k)$ for $d_{min} \geq 2$ and $\ell = 1, 2, 3, \ldots$

- **Concluding Remarks**
  - The rate of CG/CMG codes gives a **global lower bound**. How good is it?
  - LRC construction **not based on** simple graph
  - Binary LRC with joint locality (r_1, r_2, r_3, r_4)
  - **Non-binary LRC** construction with the same G for either CG or CMG code

Communication Signal Design Lab