Reliability Comparison of Various Regenerating Codes for Cloud Services

Yonsei Univ. Seoul, KOREA

Jung-Hyun Kim, Jin Soo Park, ***Ki-Hyeon Park***,

Inseon Kim, Mi-Young Nam, and Hong-Yeop Song

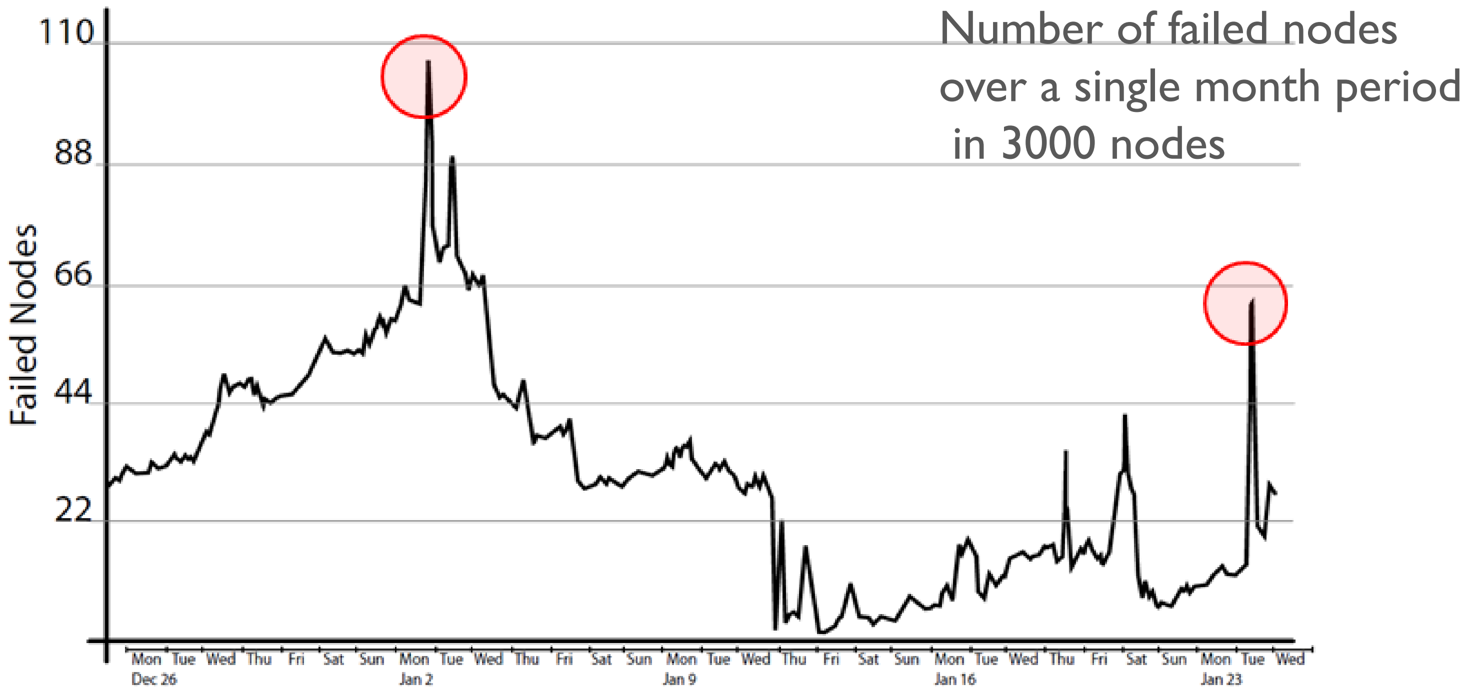
ICTC'13, Oct. 14 - 16, 2013

Contents

- Introduction
  - Why we need regenerating codes for clouds?
  - How to regenerating failed nodes?
- Background of Regenerating Codes
  - Regenerating code framework
  - Tradeoff between storage size and repair bandwidth
- Various Regenerating Codes
  - Minimum Storage Regenerating (MSR) codes
  - Minimum Bandwidth Regenerating (MBR) codes
  - Local Reconstruction Codes (LRC)
  - LT Regenerating codes
- Simulation Results
- Conclusion

Introduction

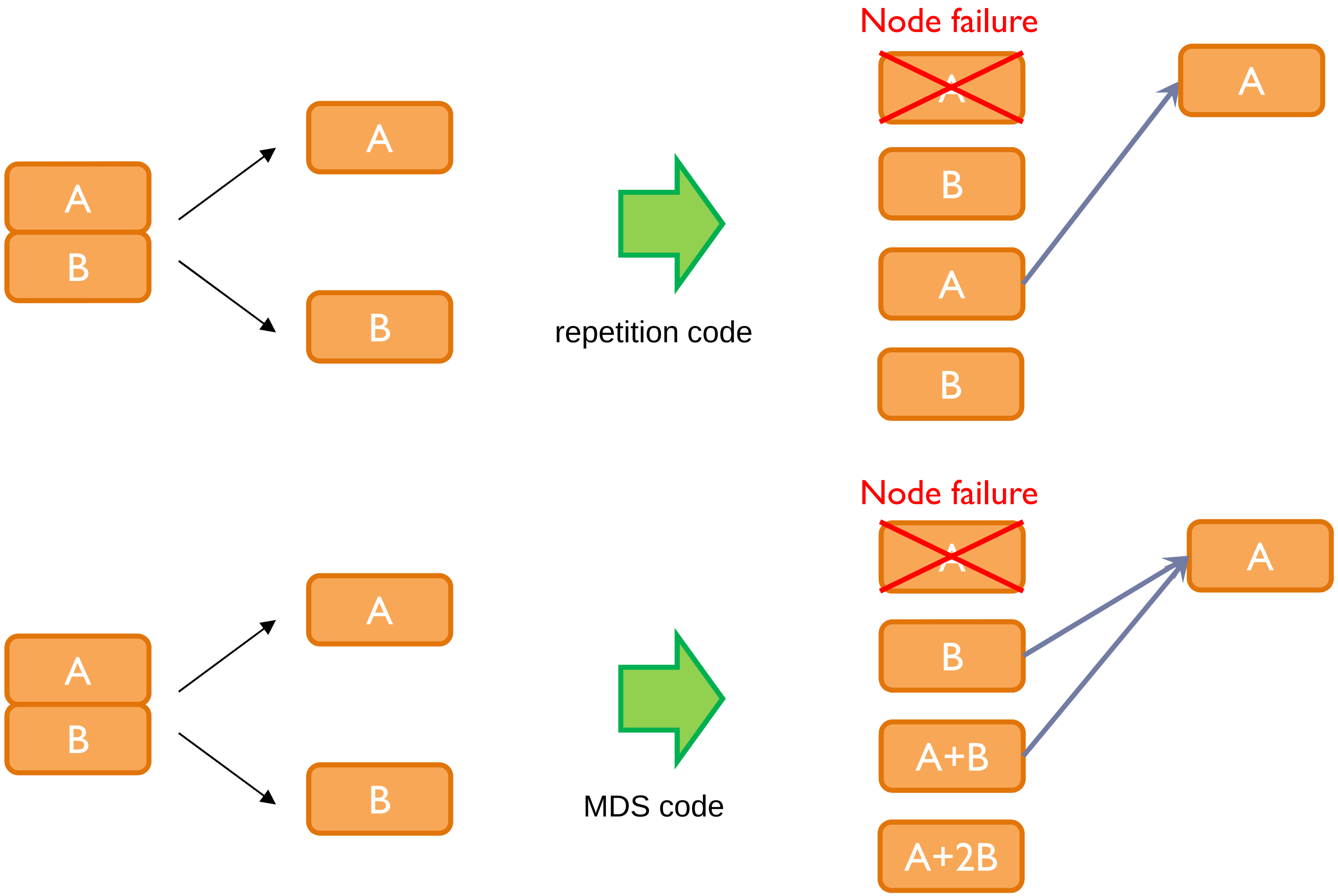
- Why we need regenerating codes for clouds?
  - To repair ***node failure***
    - At ***Facebook***, it is quite typical to have 20 or more node failures per day.



M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, D. Borthakur, "XORing Elephants: Novel Erasure Codes for Big Data," in Proc. of the 39th International Conf. on Very Large Data Bases, 2013.

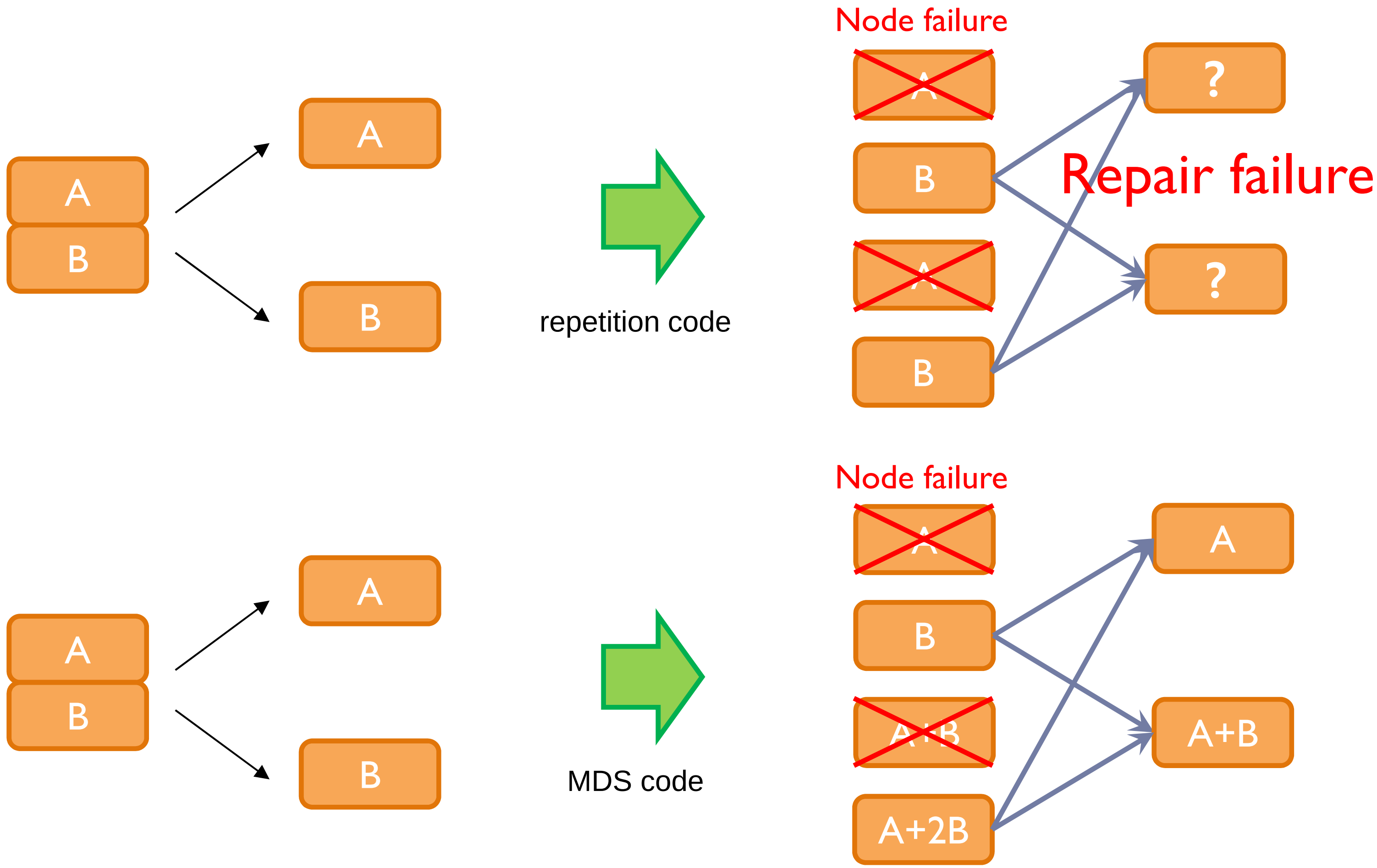
Introduction

- How to regenerate failed nodes?
  - Node repair using codes for erasure channel



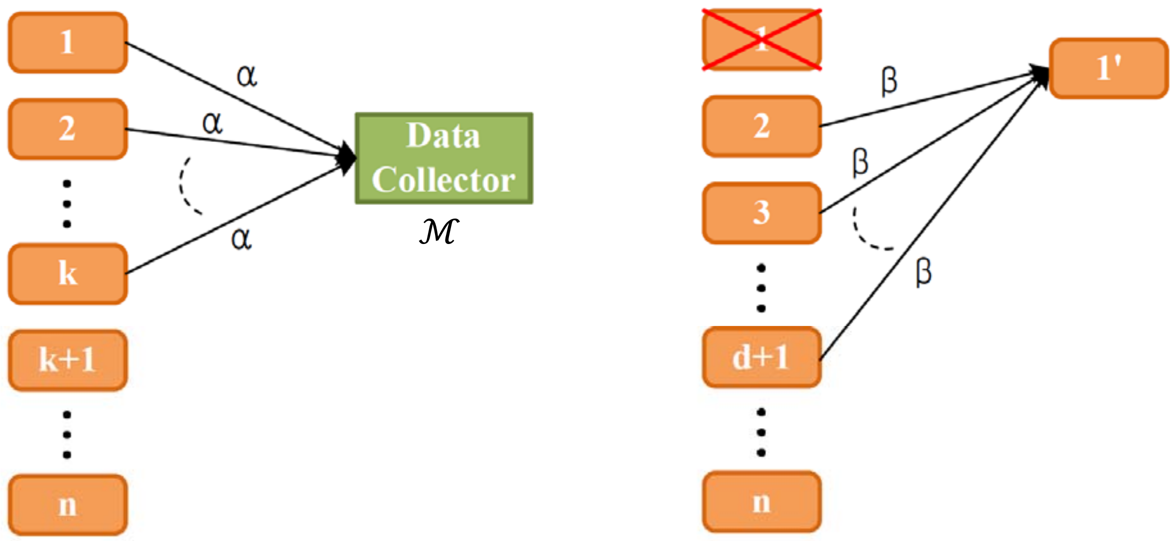
Introduction

- How to regenerate failed nodes?
  - MDS codes have higher reliability than repetition codes



Background of Regenerating codes

- Regenerating Codes Framework



(a) Data collection (b) Node repair

n : # of storage nodes

k : # of storage nodes for data collection

α : storage size

$\mathcal{M}$: data size

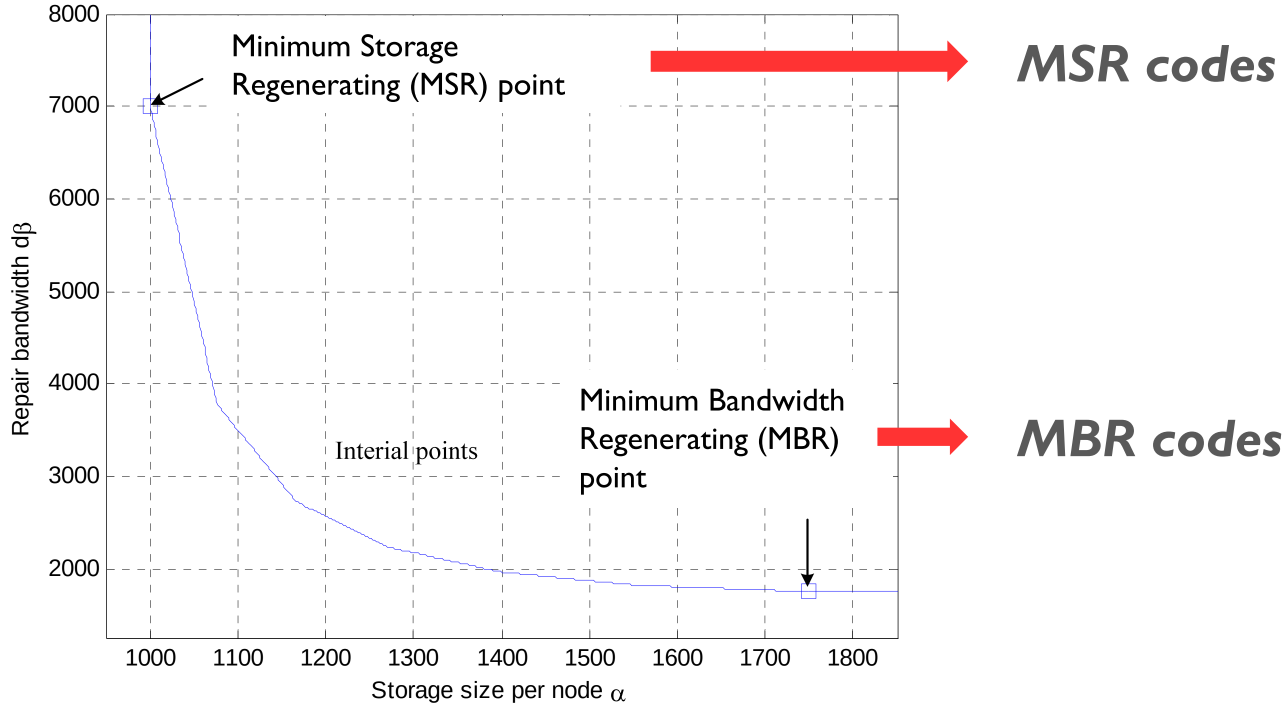
d : # of storage nodes for node repair (read cost)

β : download size

$d\beta$: repair bandwidth

Background of Regenerating codes

- Reducing ***storage size and repair bandwidth***
  - Based on the min-cut bound : $\sum_{i=0}^{k-1} \min\{(d-i)\beta, \alpha\} \geq \mathcal{M}$

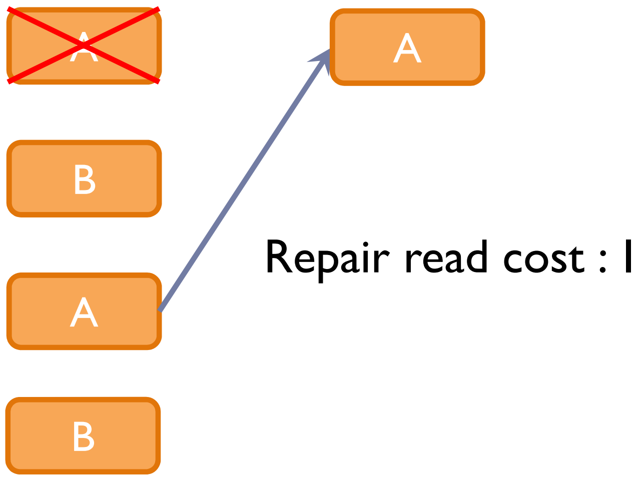


< Tradeoff between storage size and repair bandwidth ($\mathcal{M}$=7000, n=15, k=7, d=7) >

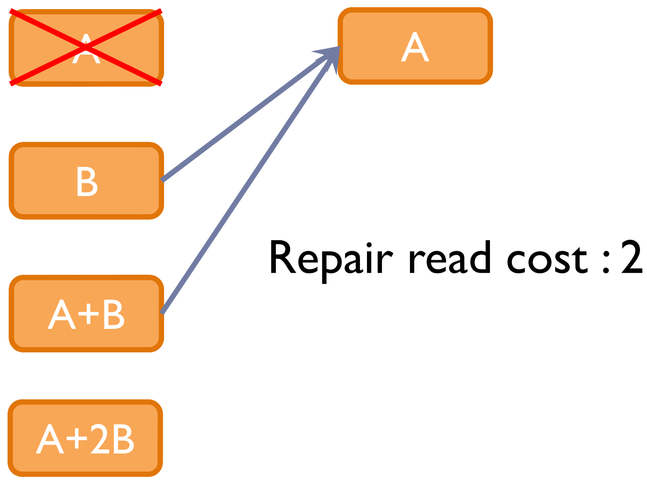
N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, “Distributed Storage Codes With Repair-by-Transfer and Nonachievability of Interior Points on the Storage-Bandwidth Tradeoff,” *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1837–1852, Mar. 2012.

Background of Regenerating codes

- Reducing ***repair read cost***
  - Repair read cost : the minimum number of nodes for repair



repetition code



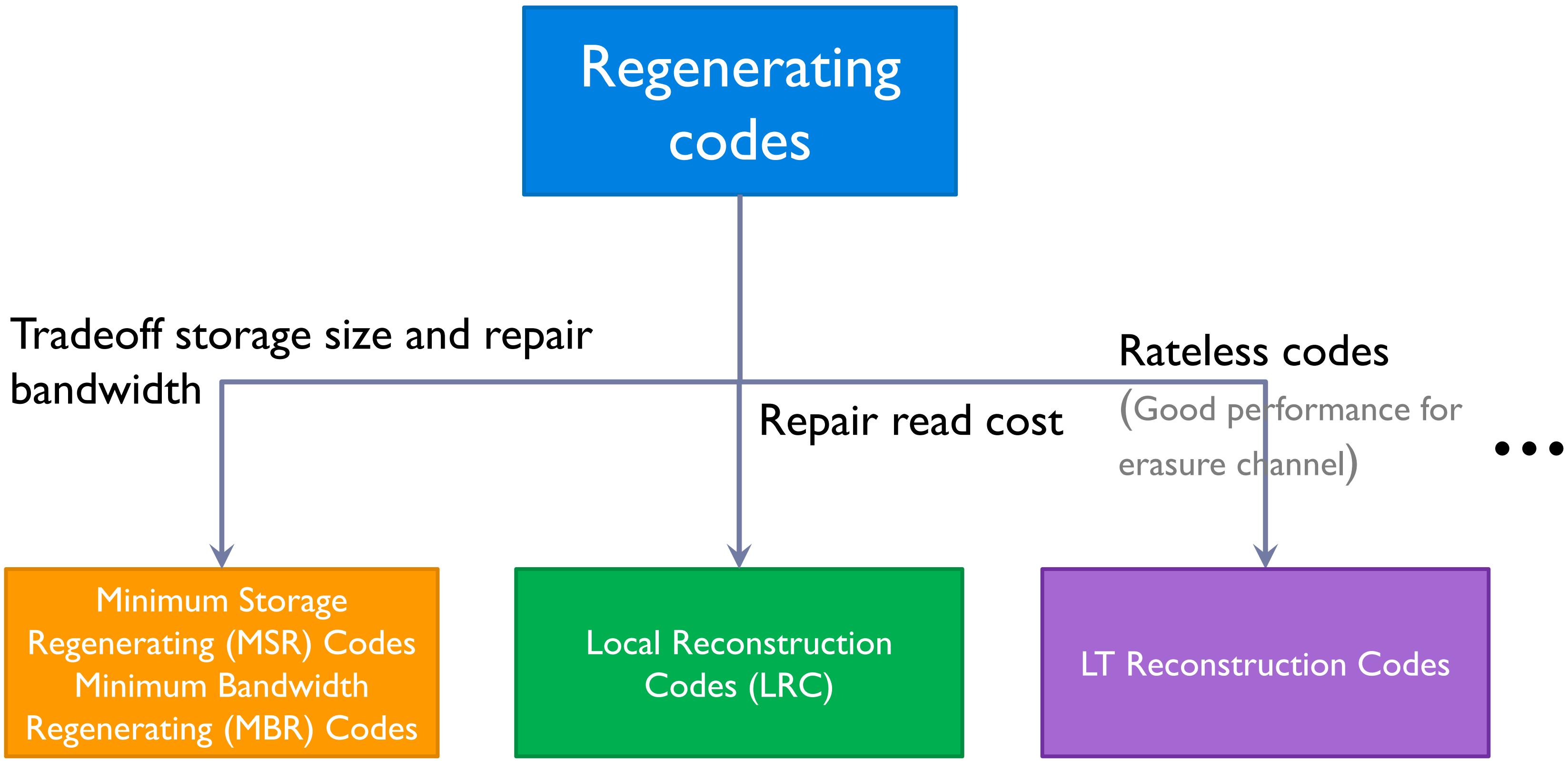
MDS code

Efficient Repair
Low Reliability

High Reliability
Inefficient Repair

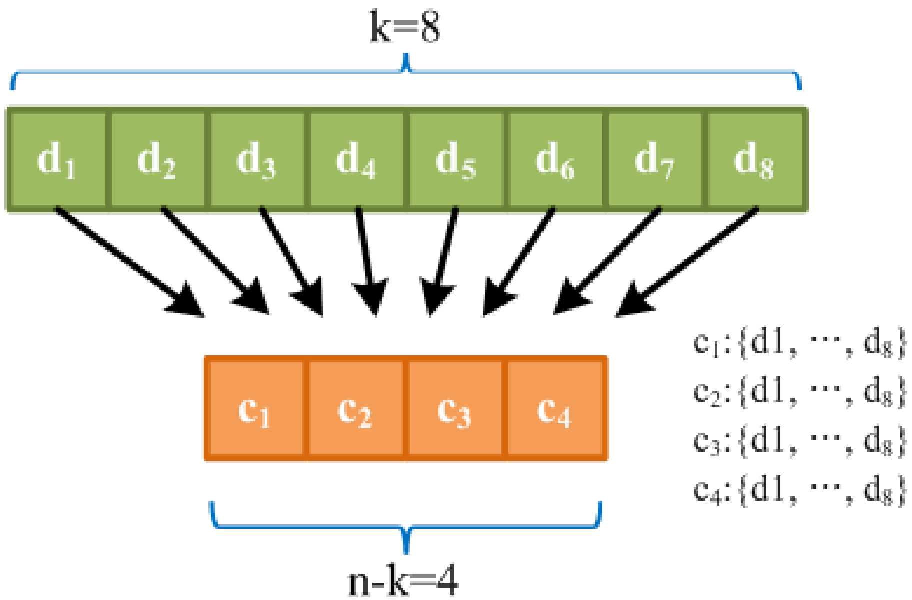
Q. Tradeoff between "repair read cost" and "reliability"?

Various Regenerating codes



Various Regenerating codes

- Minimum Storage Regenerating (MSR) codes
  - Using a Maximum Distance Separable (MDS) code



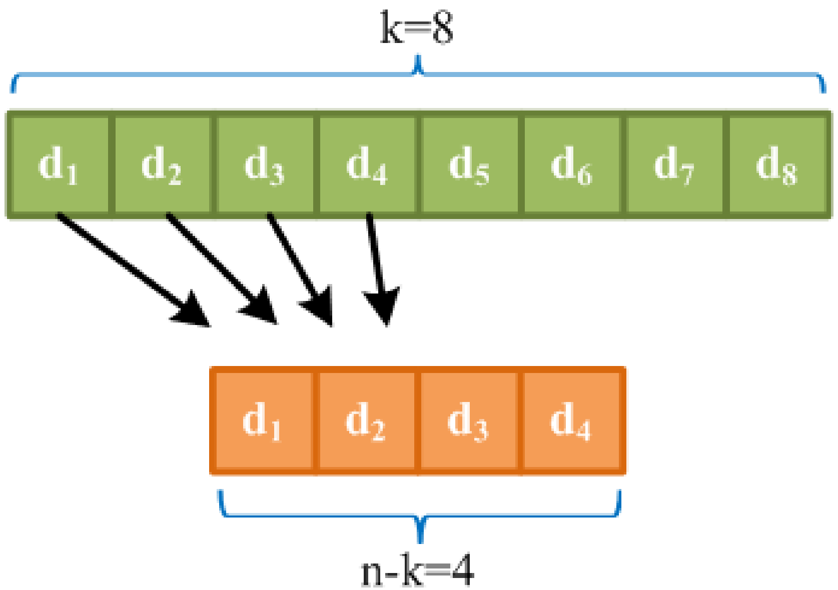
< (8, 4) MDS code >

- Code construction methods
  - Interference Alignment method, Product-Matrix method, etc.

N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Interference Alignment in Regenerating Codes for Distributed Storage: Necessity and Code Constructions," IEEE Trans. Inf. Theory, vol. 58, no. 4, pp. 2134–2158, April 2012.
K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for the MSR and MBR points via a product-matrix construction," IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 5227-5239, Aug. 2011.

Various Regenerating codes

- Minimum Bandwidth Regenerating (MBR) codes
  - Using a Fractional Repetition (FR) code



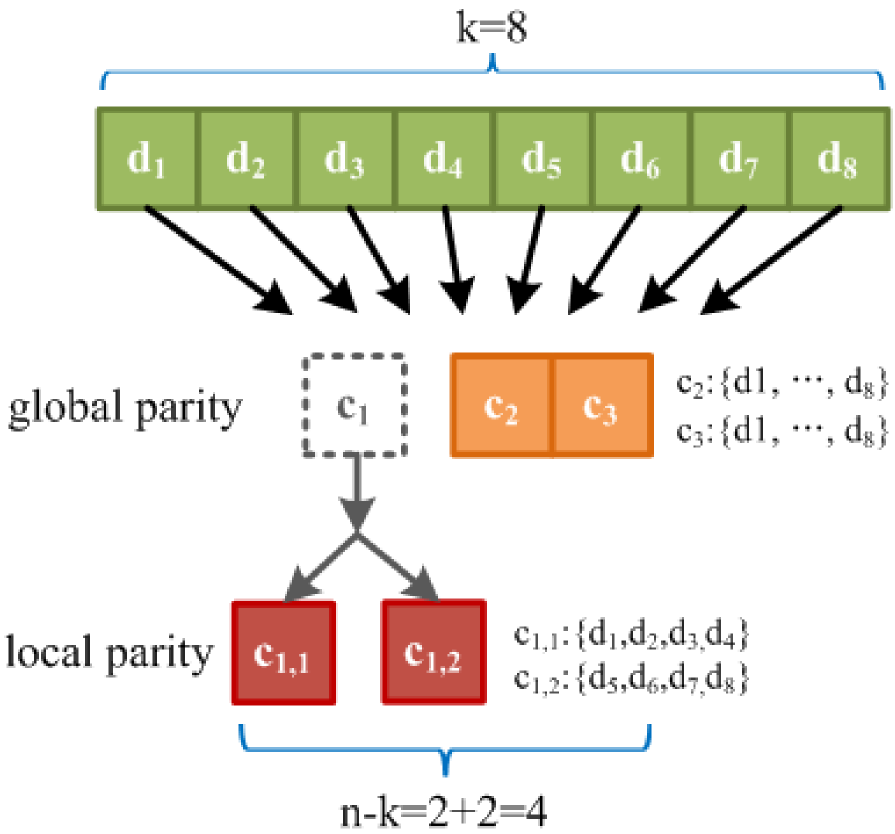
< (8, 4) FR code >

- Code construction methods
  - Repair-by-product method, Product-Matrix method, etc.

K. W. Shum, and Y. Hu, "Functional-Repair-by-Transfer Regenerating Codes," in Proc. of 2012 IEEE International Symposium on Information Theory, Cambridge, MA, July 2012.
K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for the MSR and MBR points via a product-matrix construction," IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 5227-5239, Aug. 2011.

Various Regenerating codes

- Local Reconstruction Codes (LRC)
  - Extending an MDS code

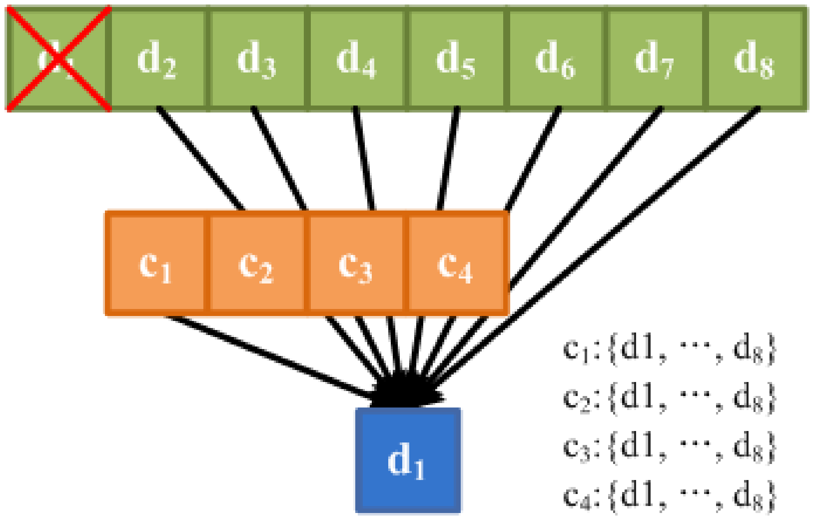


< (8, 2, 2) LRC >

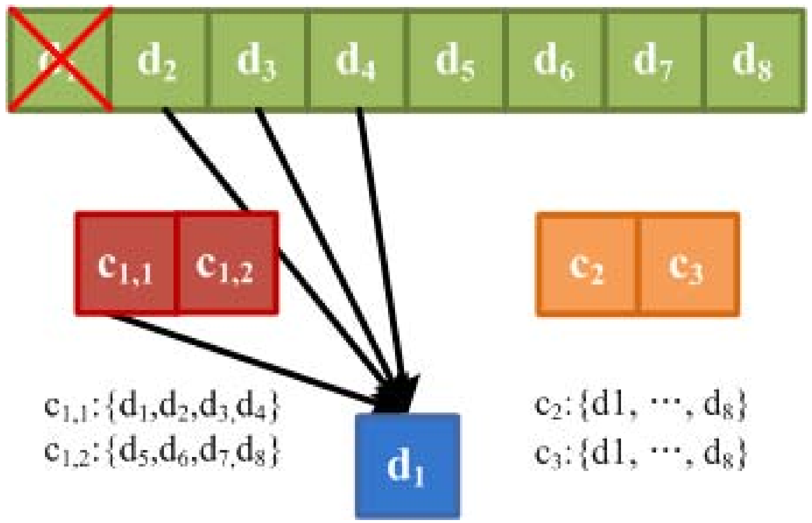
Cheng Huang, Minghua Chen, and Jin Li. “Pyramid codes: flexible schemes to trade space for access efficiency in reliable data storage systems,” In Sixth IEEE International Symposium on Network Computing and Applications (NCA 2007), pp. 79-86, 2007.

Various Regenerating codes

- Local Reconstruction Codes (LRC)
  - Repair read cost comparison between MSR code and LRC



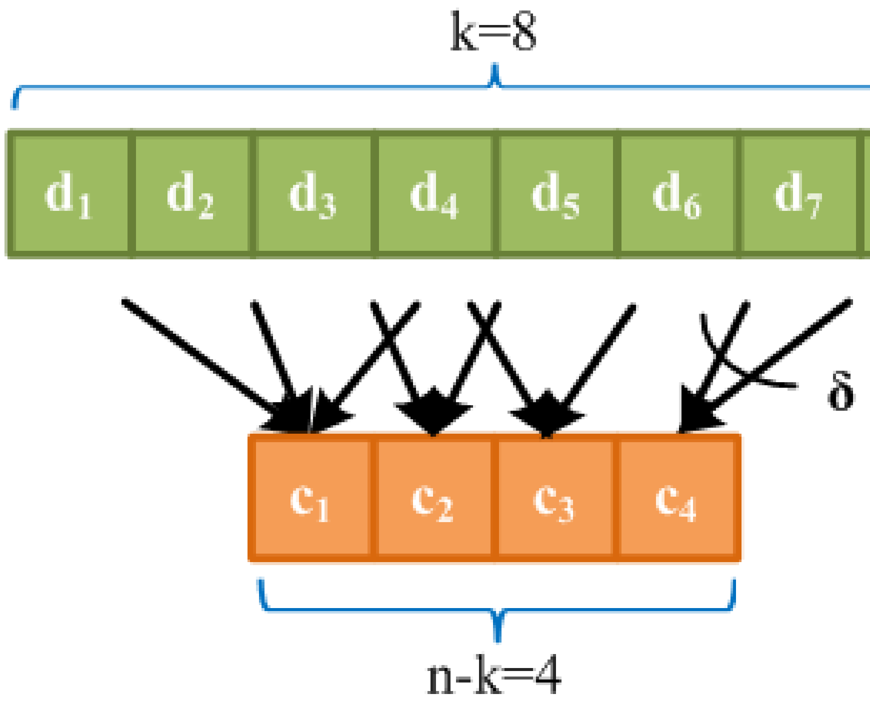
MSR code
repair read cost = 8



LRC
repair read cost = 4

Various Regenerating codes

- LT Regenerating Codes
  - Using the ideal/robust soliton distribution



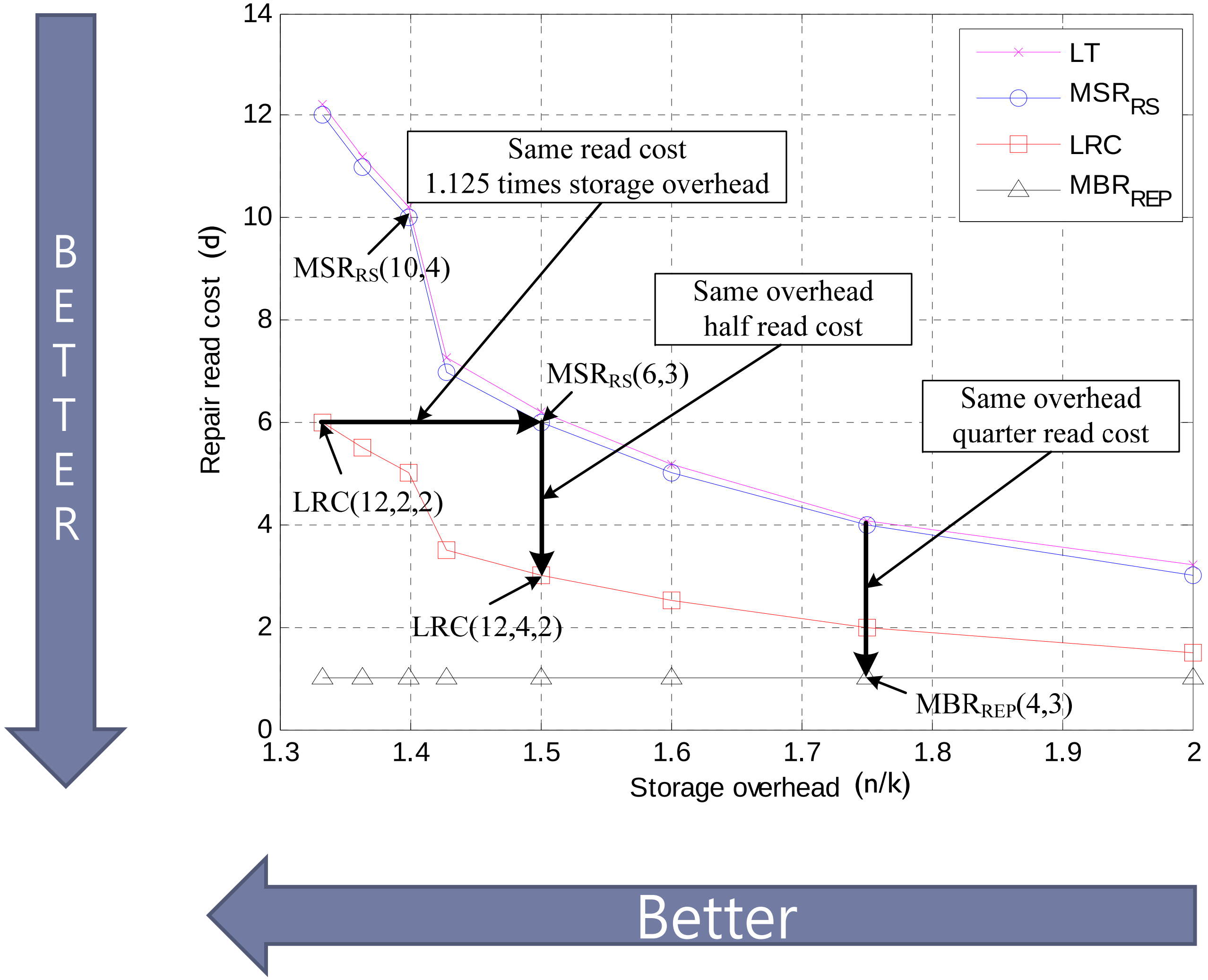
< (8, 4) LT regenerating code >

: degree distribution

M. Asteris and A. G. Dimakis, "Repairable fountain codes," in Proc. of 2012 IEEE International Symposium on Information Theory, Cambridge, MA, July 2012.

Simulation Results

- Better cost and overhead trade-off

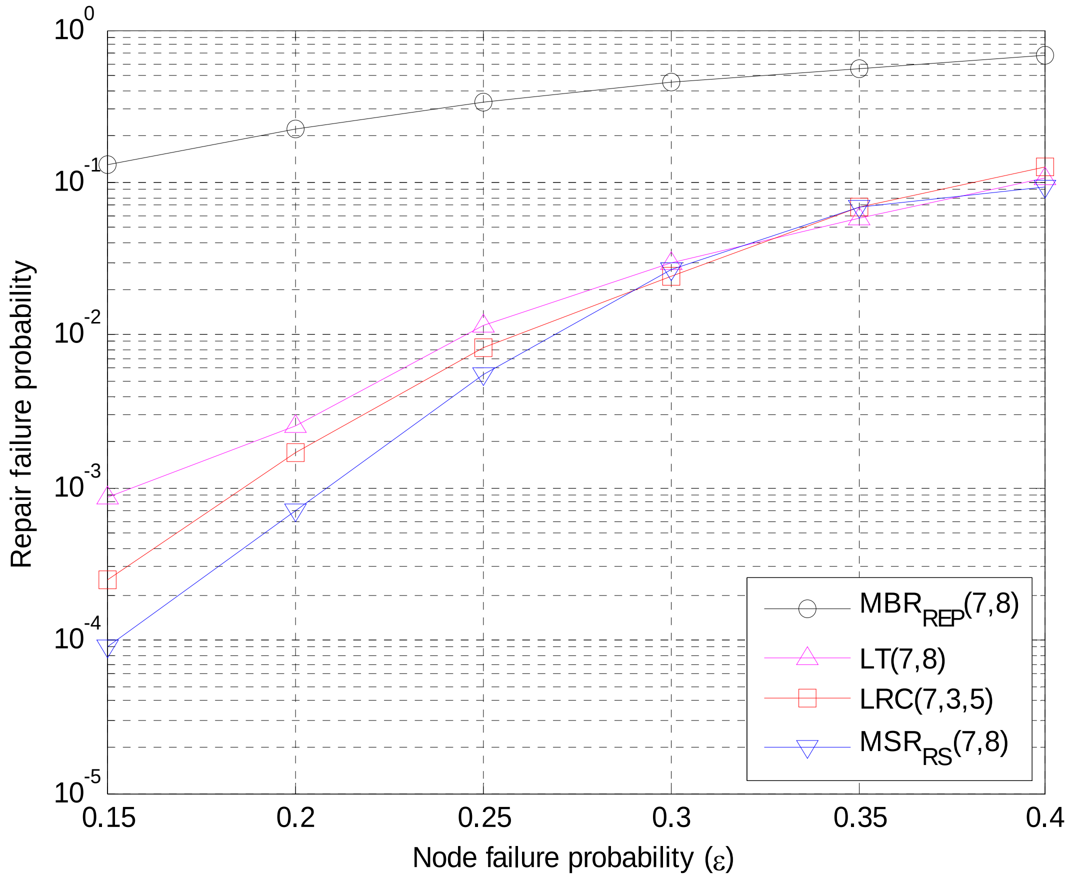


Storage overhead : the ratio of all storage nodes, n, to storage nodes for data collection, k

Repair read cost : the number of storage nodes for node repair

Simulation Results

- Repair failure probability for different node failure probability



Node failure prob. :
the probability that a node is unavailable

Repair failure prob. :
the probability that any newcomer nodes can not repair the original data symbol from coded data symbols of surviving storage nodes

Conclusion

- Through the trade-off between repair read cost and storage overhead, we can expect that *the optimal coding scheme might be different according to system requirements*.

- Although LRC is not an MDS code, it achieves both low repair read cost and low storage overhead by relaxing MDS property. Hence *LRC can be a good candidate for practical systems* and it should be studied more as a future coding scheme for cloud services.