# CODED DISTRIBUTED STORAGE FOR CLOUD ENVIRONMENTS

Maheswaran Sathiamoorthy

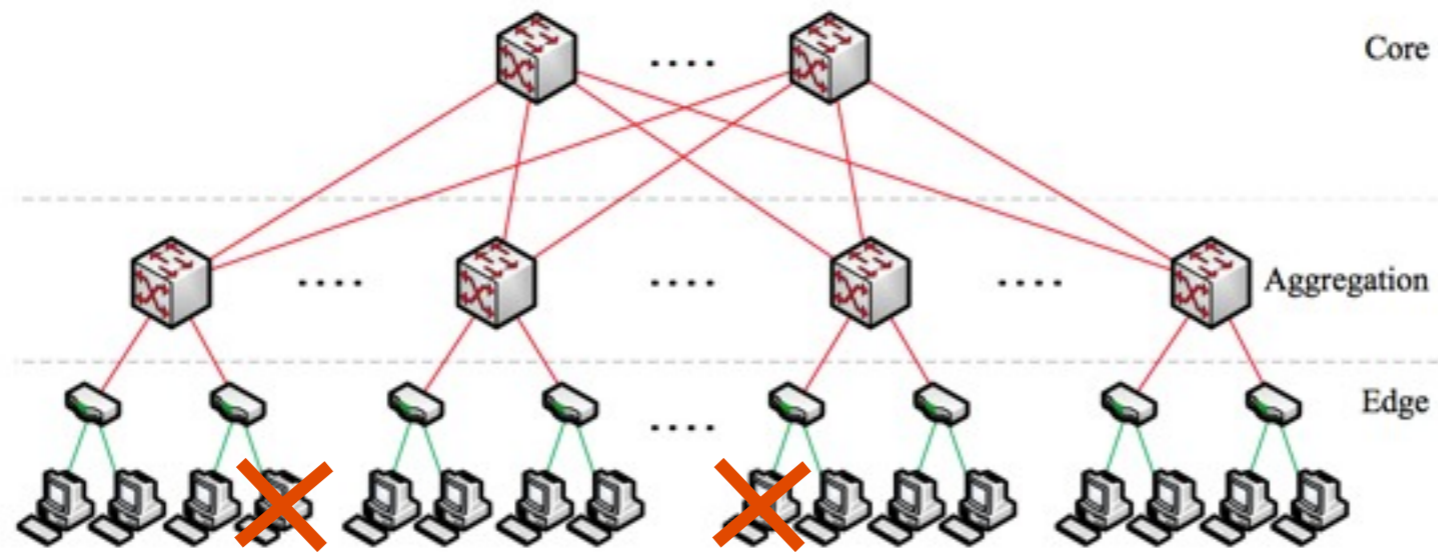USC Viterbi
School of Engineering

ANRG

# Data Centers

- Thousands of nodes under one roof.



- Failure is the norm rather than the exception.
- Need high reliability. Use some form of *redundancy*
  - Most data centers rely on 3x replication
    - Storage-redundancy = 3
  - But the amount of useful (logical) data that can be stored is **only 33%**

# Goal

Given a fixed amount of storage, store as much logical data as possible, while maintaining high data reliability.

Can we do better than 33%?

# Erasure Coding

- Erasure codes offer better storage redundancy for similar or higher reliability, but they generate a lot of repair traffic.
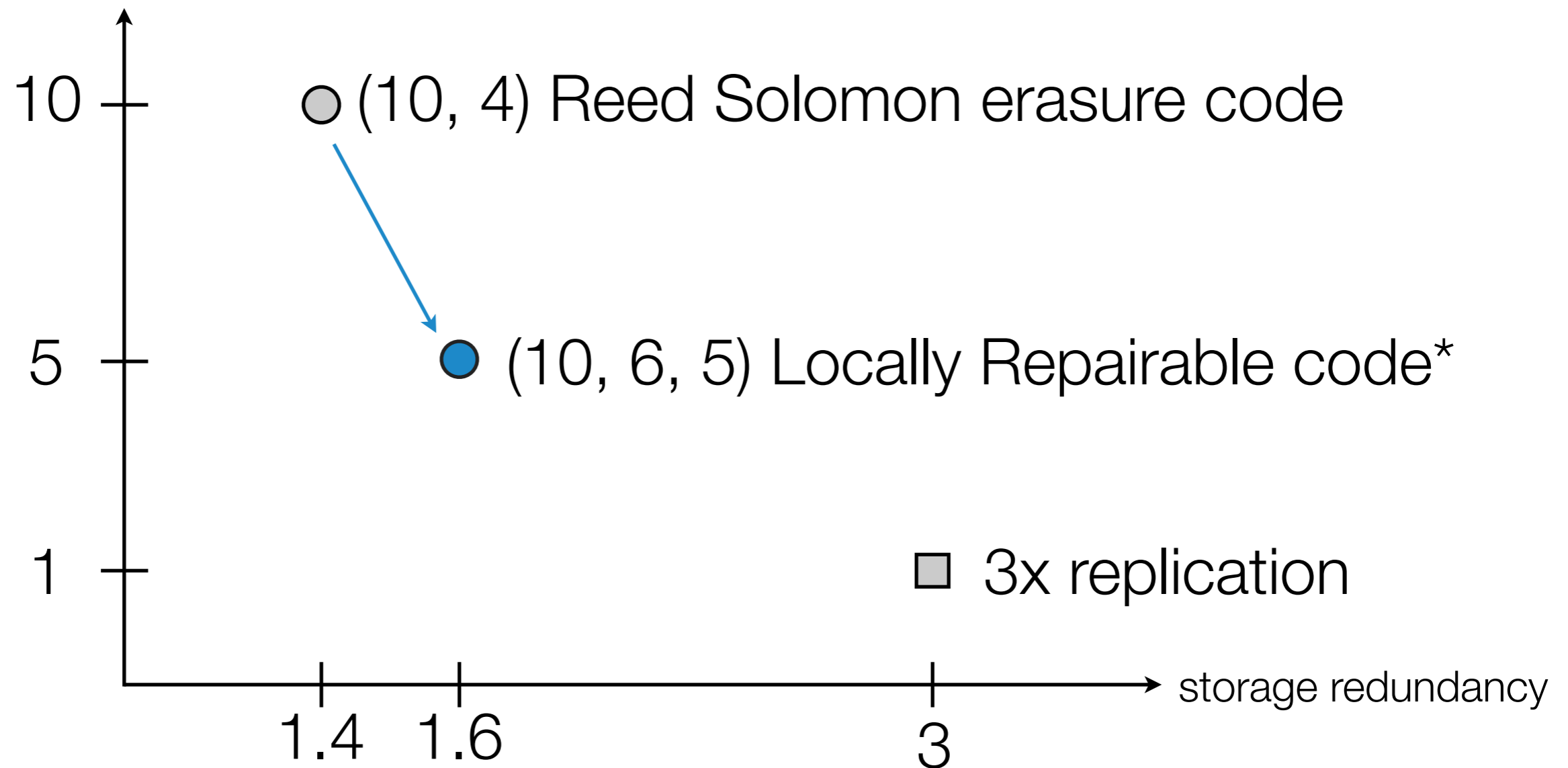
repair traffic required per block loss

10 — ⬤ (10, 4) Reed Solomon erasure code

5 —

1 — ◻ 3x replication

storage redundancy

1.4                3
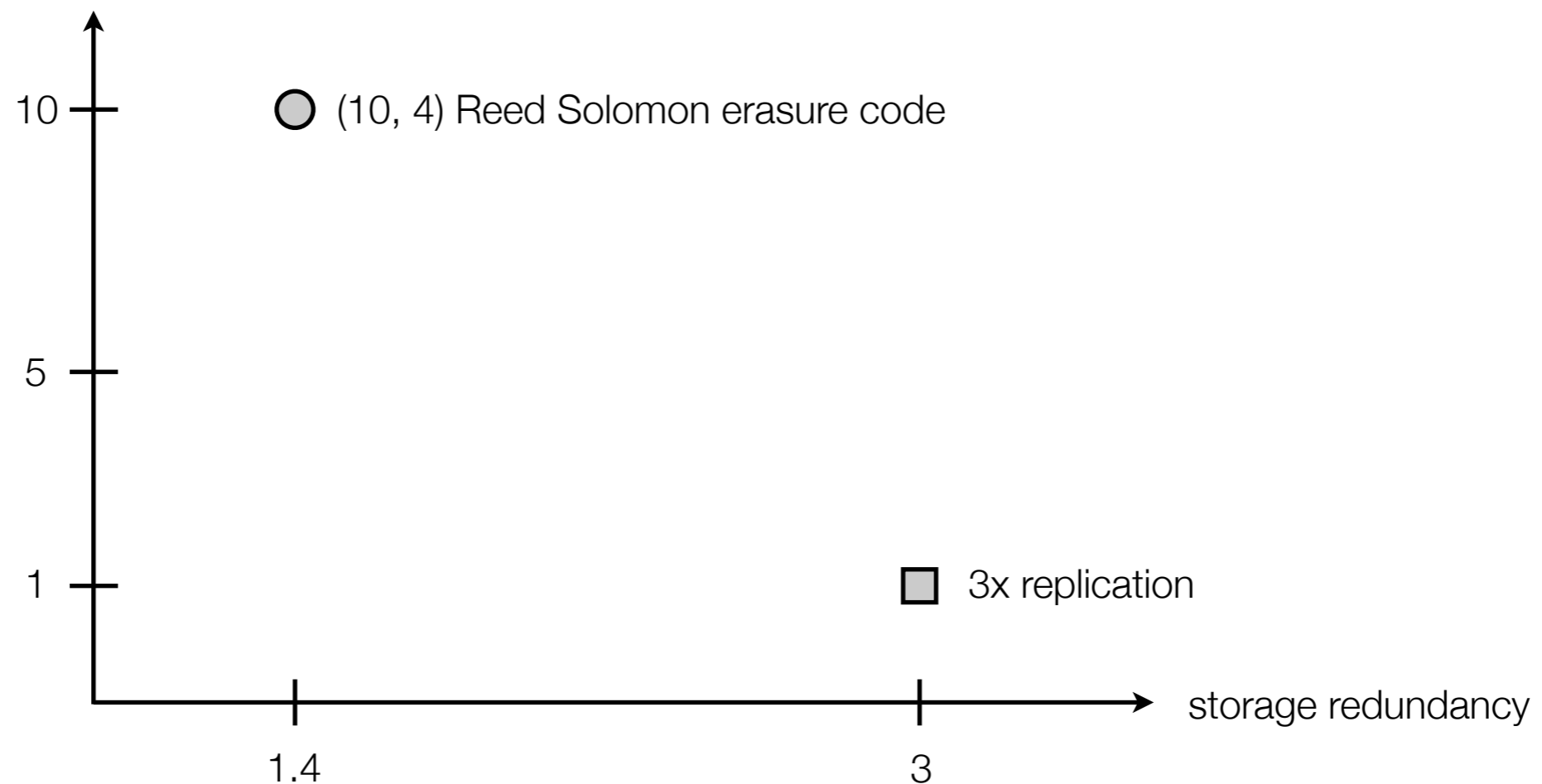
- Facebook encodes **only 8%** of data using (10,4) RS codes due to repair bottleneck.
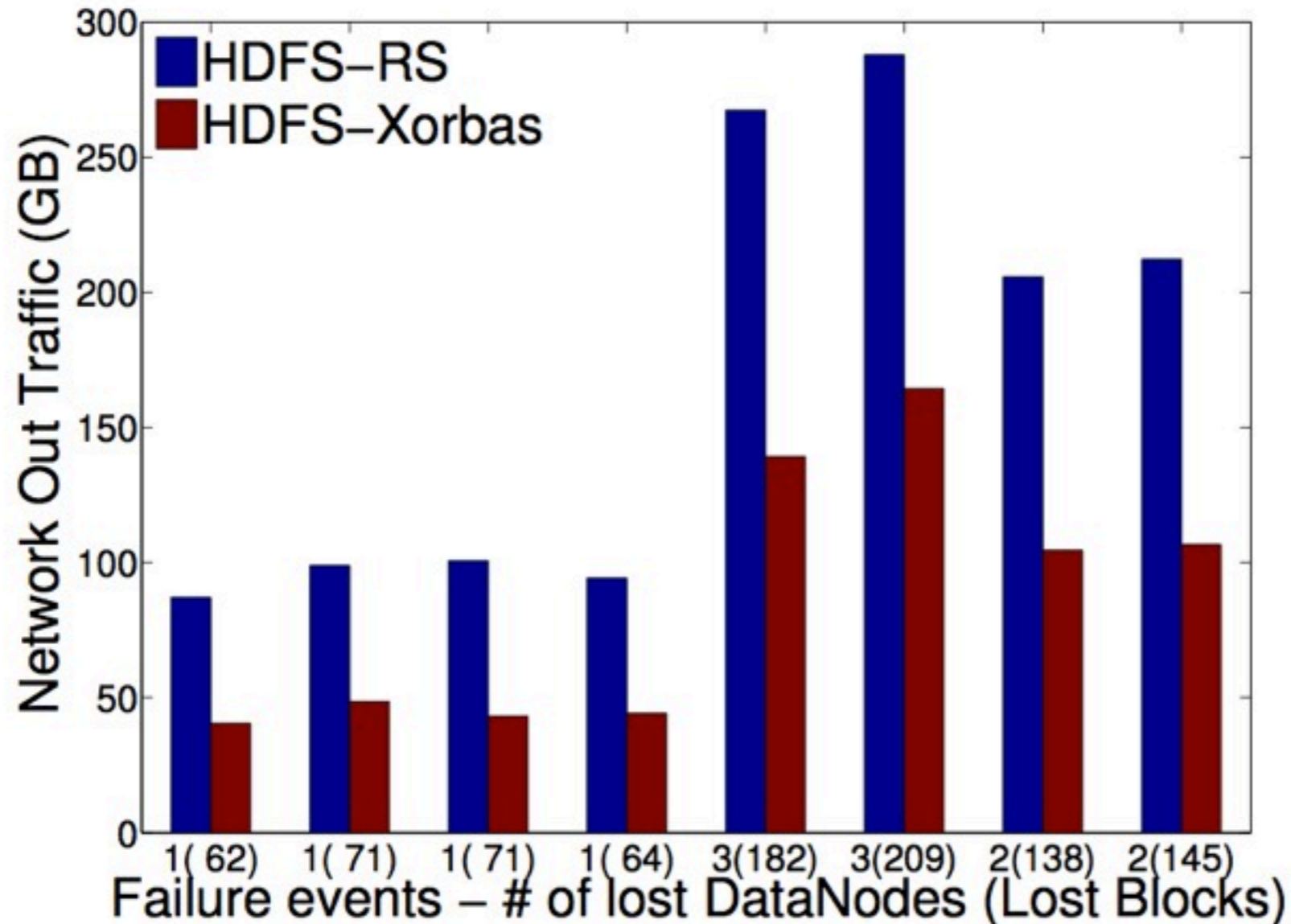- Reduce repair traffic ➡ more data erasure coded.

# Storage/Network tradeoff

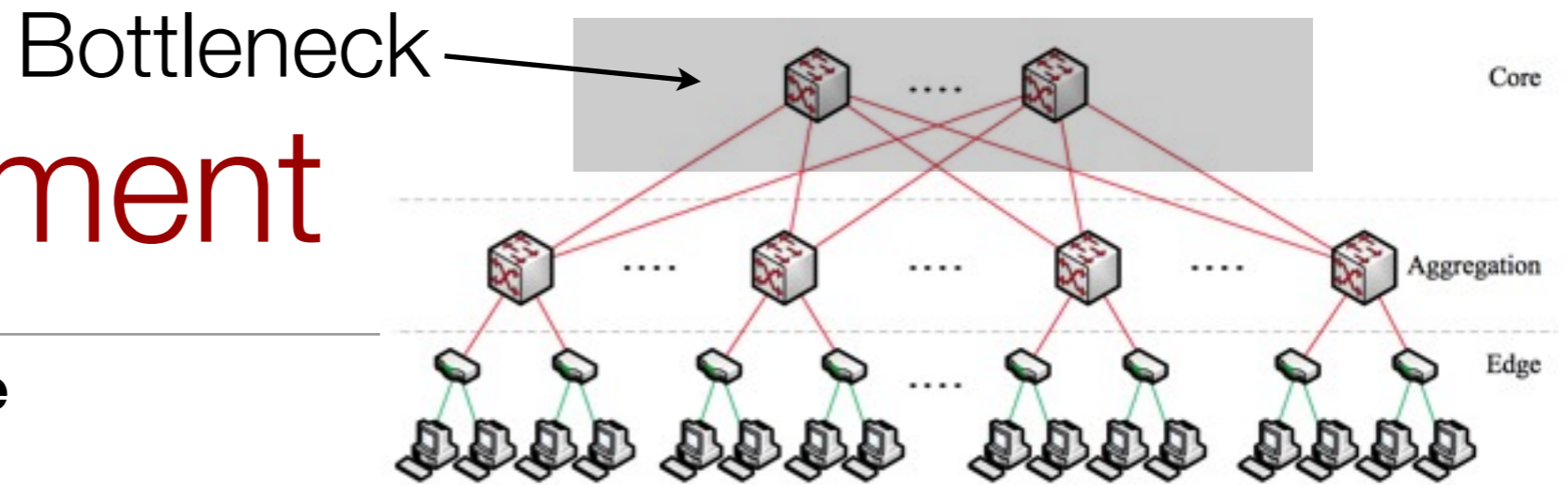repair traffic required per block loss



* Designed by Papailiopoulos et al. 5
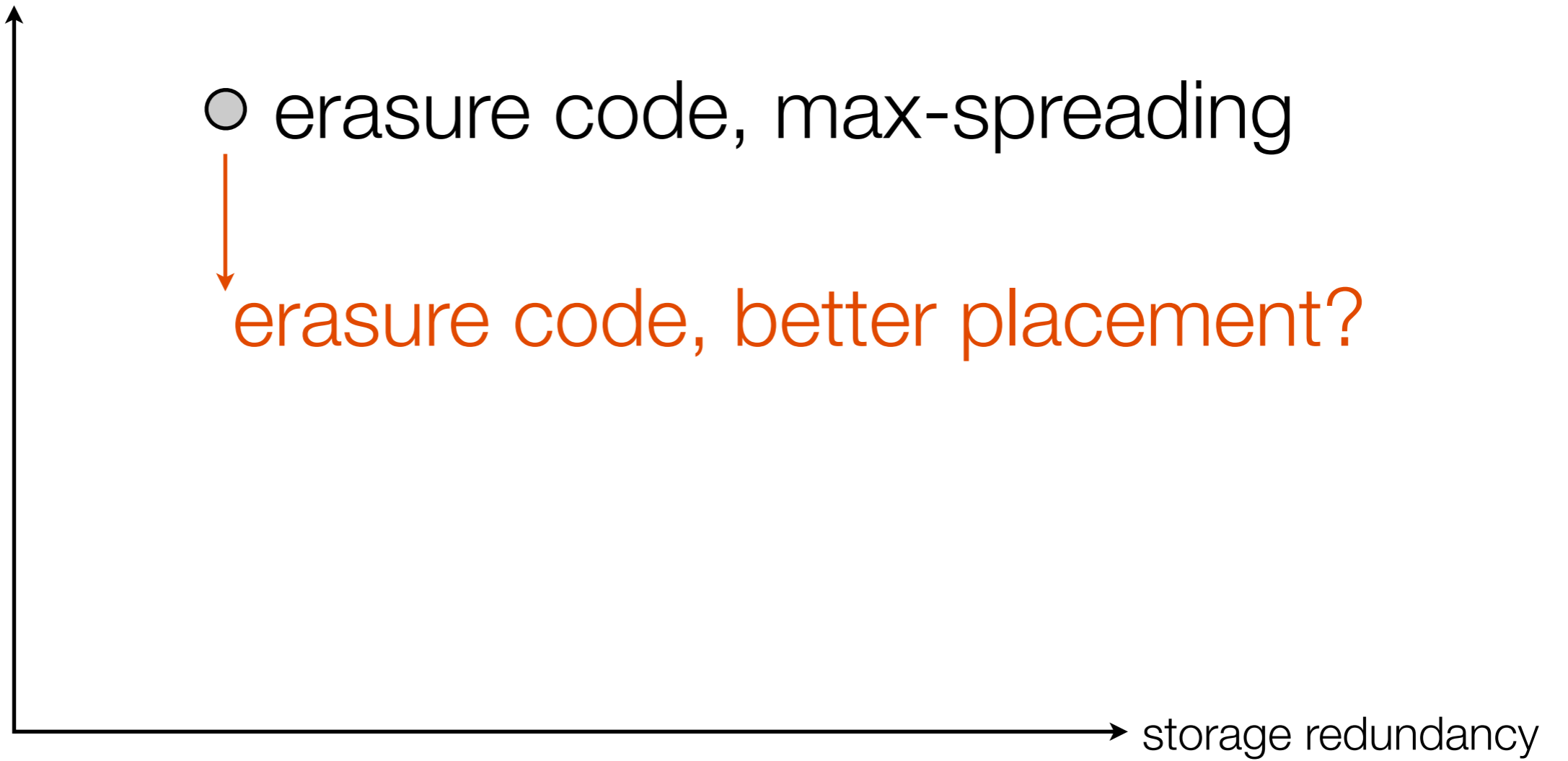
# HDFS-Xorbas (Hadoop implementation)



Experiments run over Amazon EC2 involving 100 nodes.
Overall 2x reduction in Network and Disk I/O.

# Block Placement

Bottleneck

Core

Aggregation

Edge

repair traffic required **at the core** per block loss

○ erasure code, max-spreading

erasure code, better placement?

storage redundancy

Use Markov chains to model reliability for various placements to find the best. Ongoing work.

# Conclusion

- To increase fraction of logical (useful) data stored:

  - Use erasure codes designed for Data Centers, such as Locally Repairable Codes.

  - Use better placement schemes to reduce repair traffic at the core.

- Savings can run into petabytes and thus tens of thousands of dollars.

Thank You
[msathiam@usc.edu](mailto:msathiam@usc.edu)