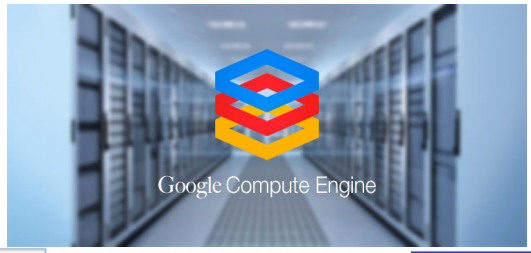
Coded Distributed Computing

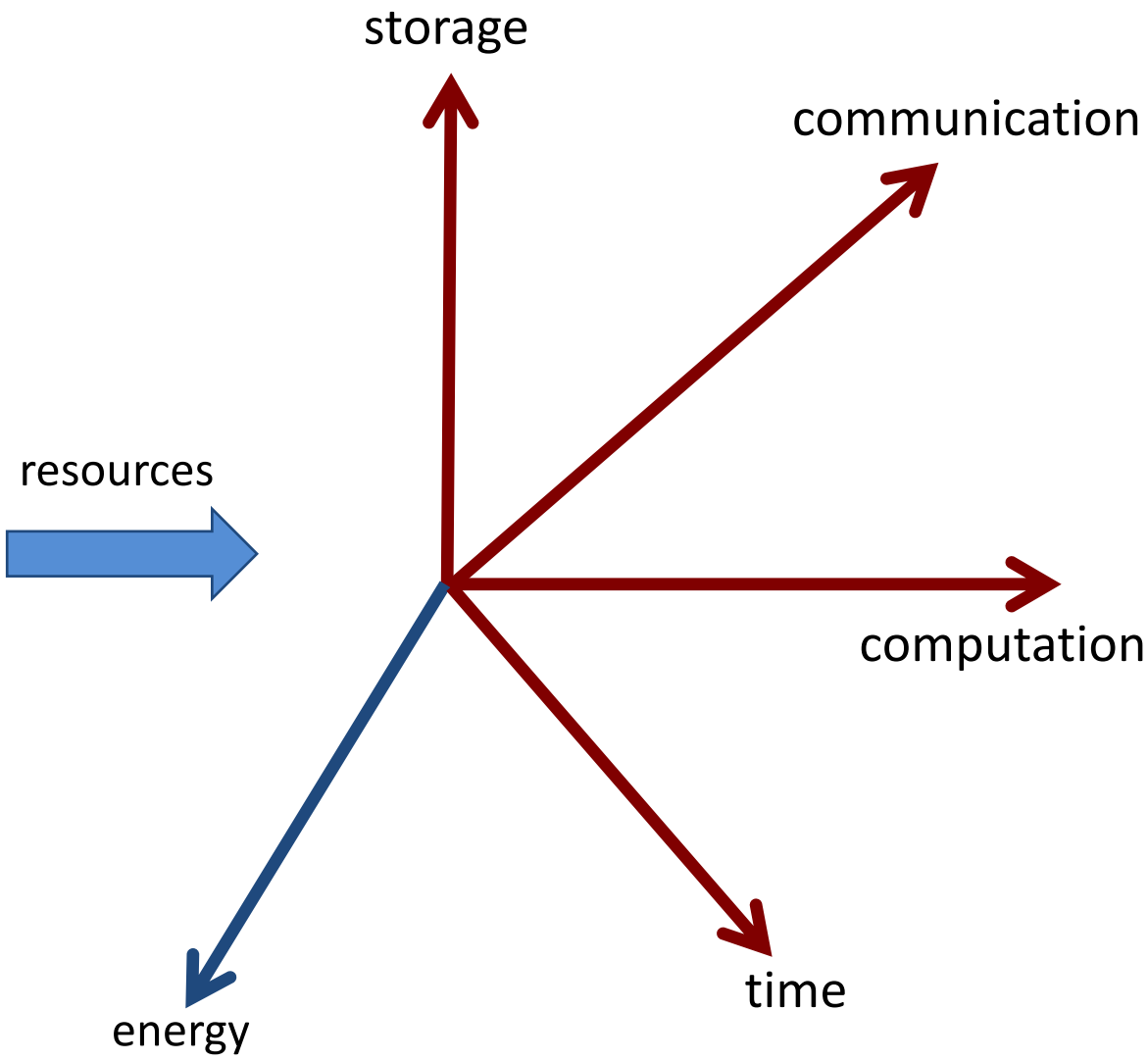
Salman Avestimehr

Computing Infrastructure of CPS

amazon web services

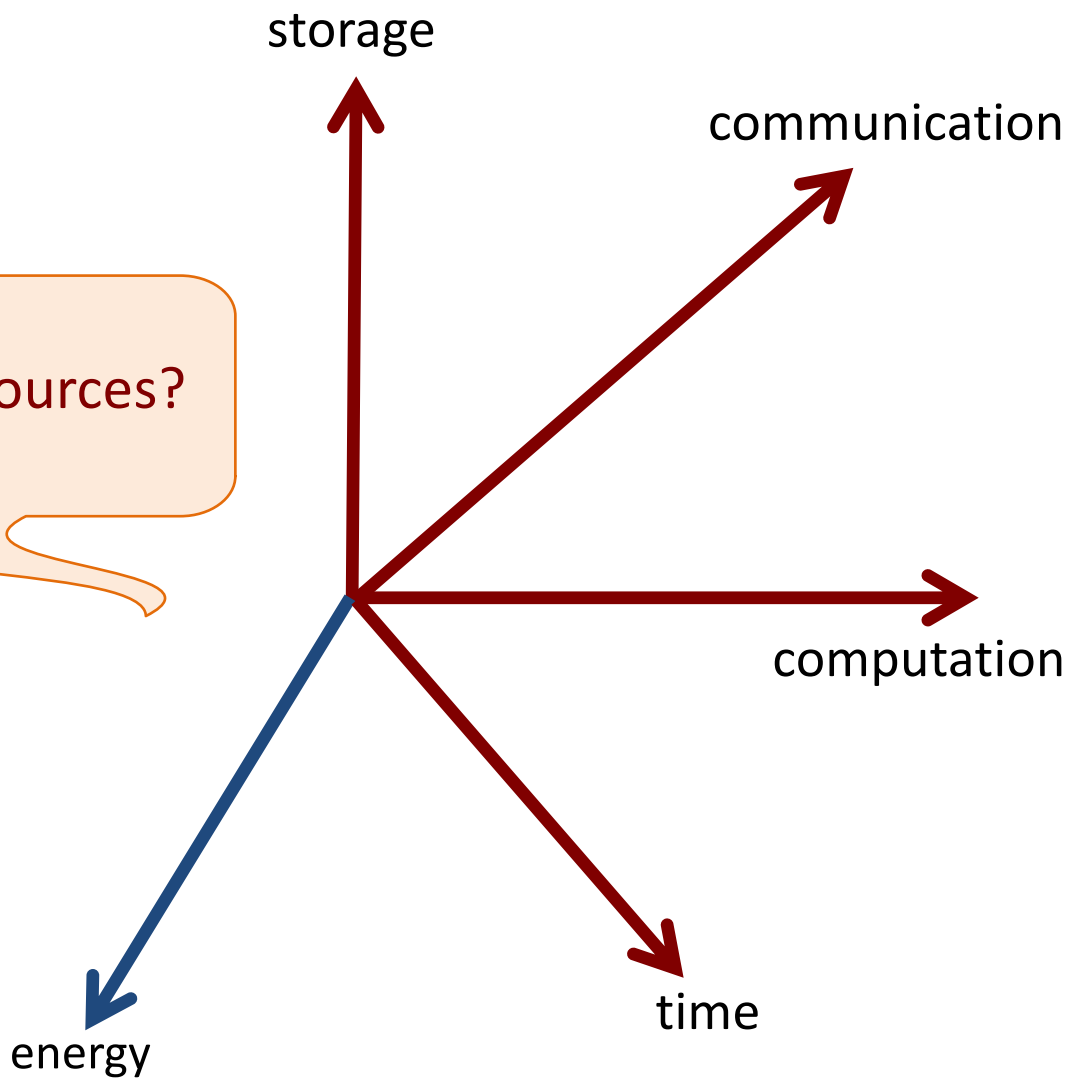


How to optimally utilize resources?

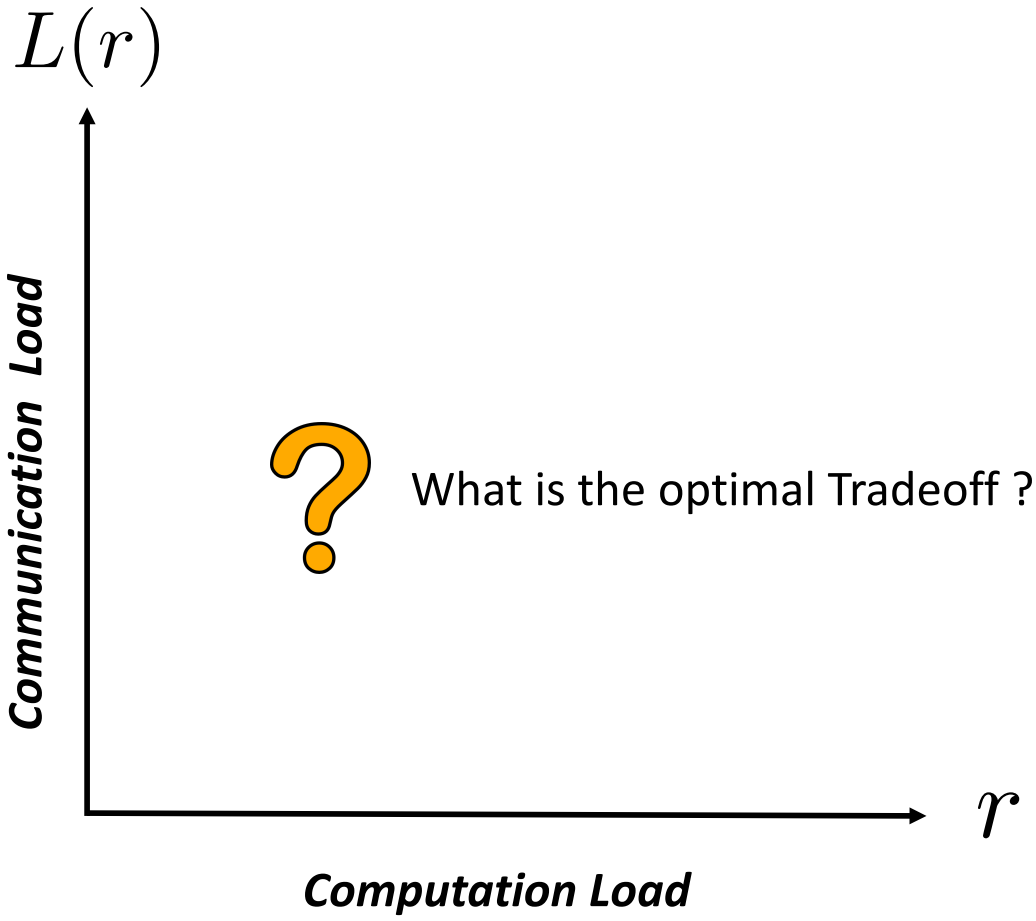


Fundamental Tradeoffs between Resources

how to optimally **trade** network resources?

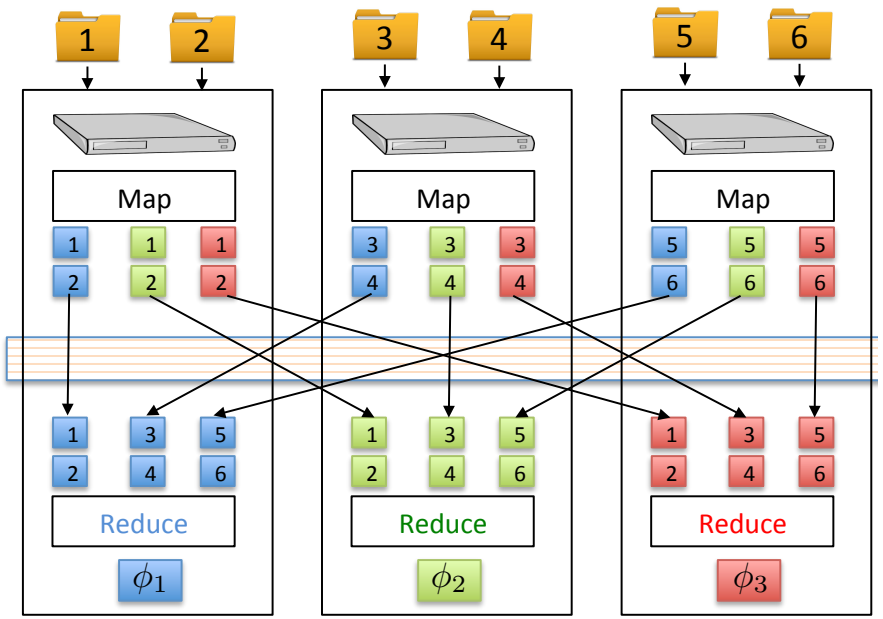


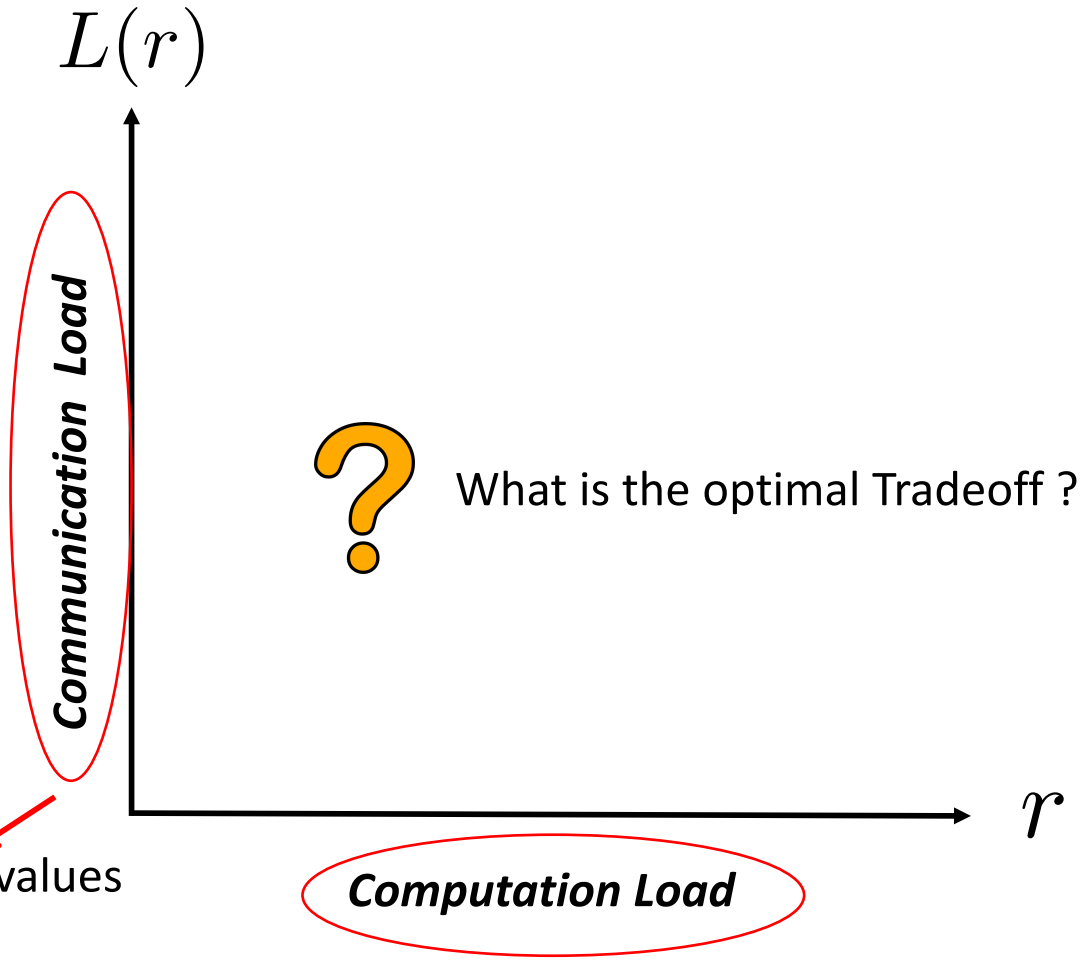
Computation-Communication Tradeoff



Computation-Communication Tradeoff

MapReduce-Type Frameworks



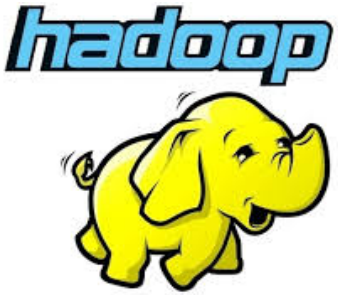


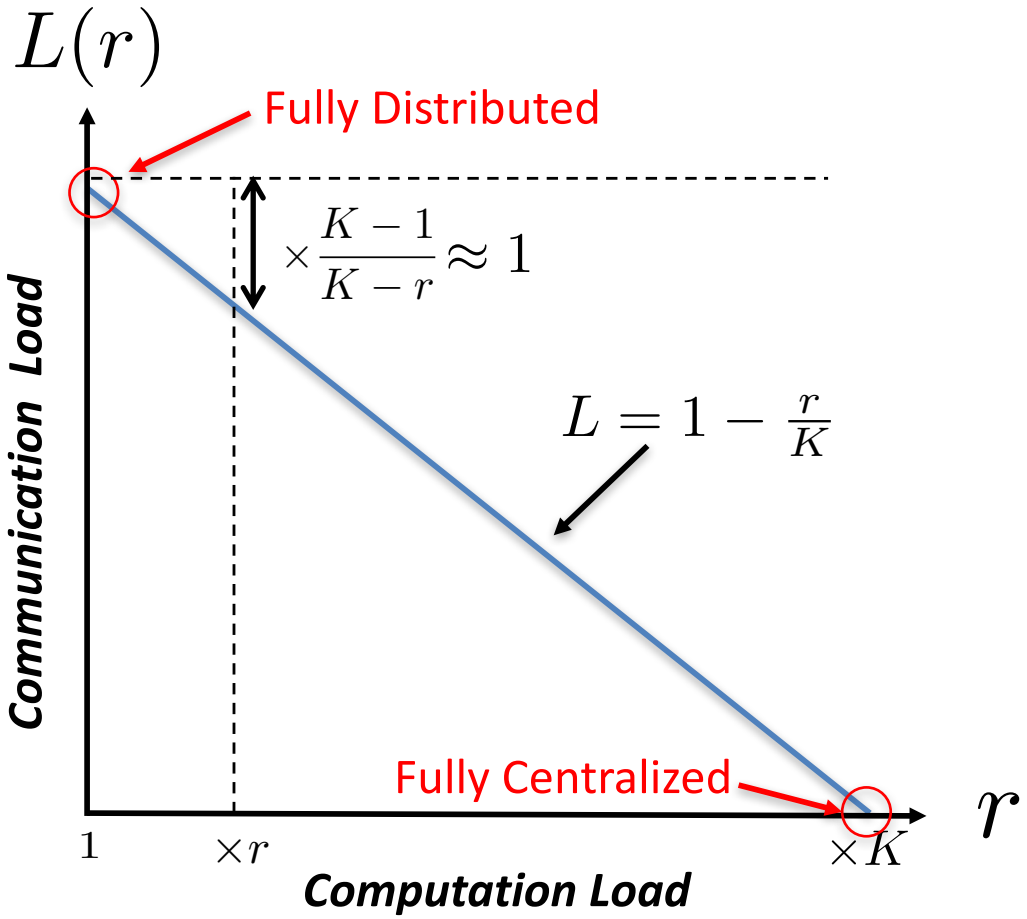
What is the optimal Tradeoff ?

# of exchanged intermediate values

average # times each file is mapped

Today's Design

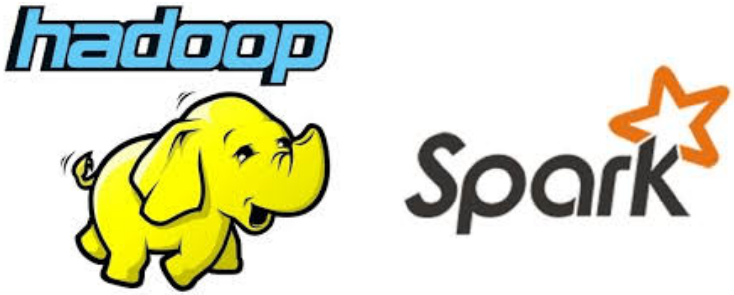


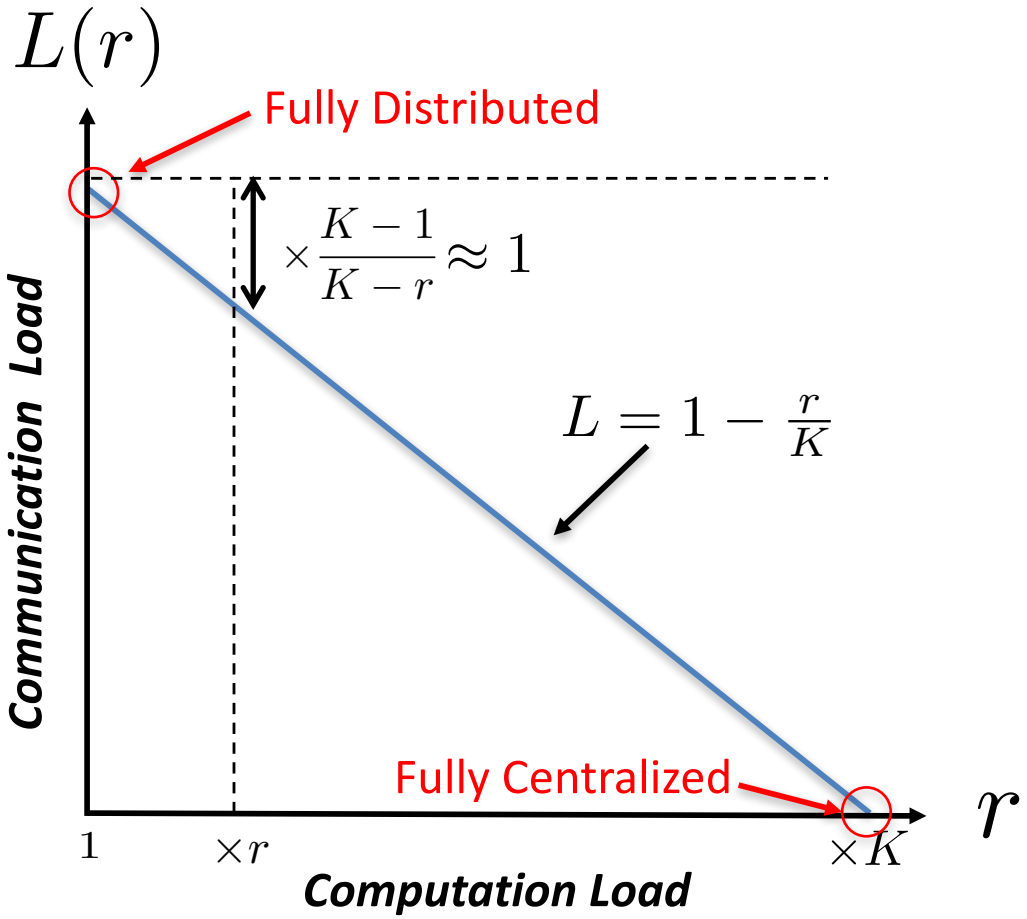


K: number of participating servers

For K=100, doubling computation only reduces the communication by 1% !

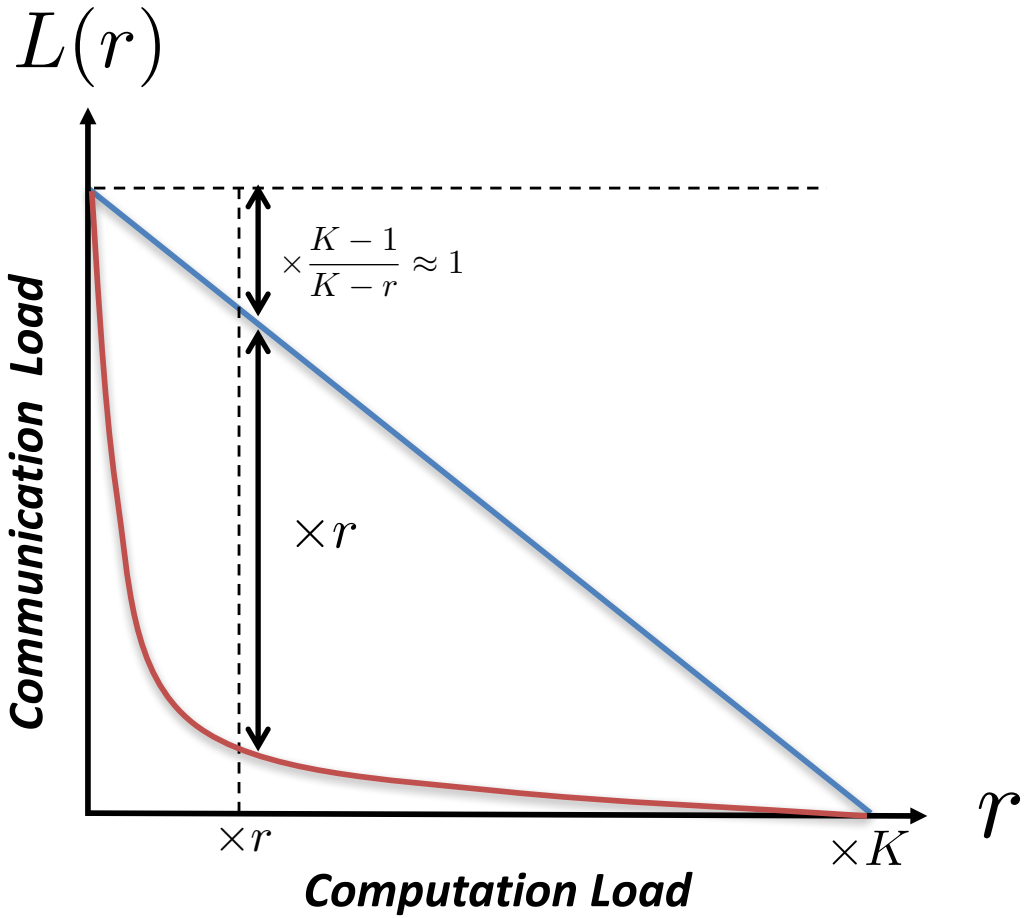
Today's Design





Is this the best tradeoff?

Coded Distributed Computing

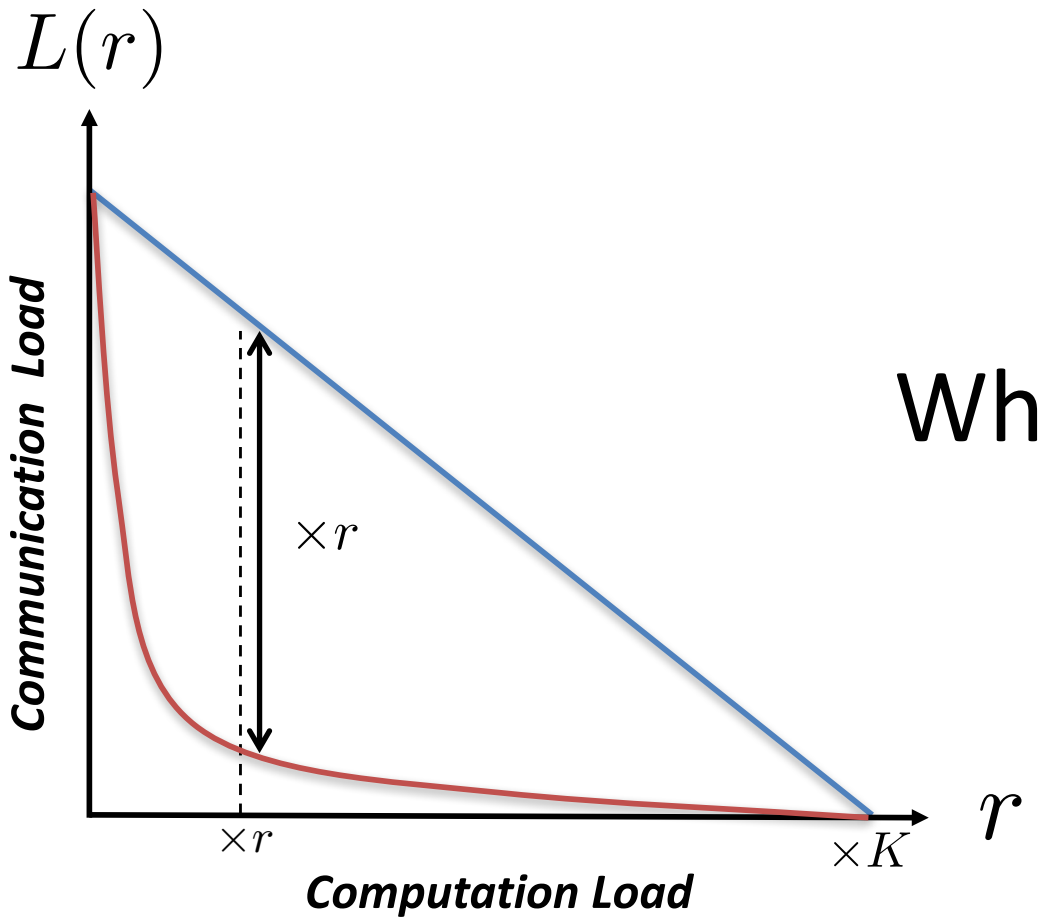


Comm. Load (Coded)

$$L_{\text{coded}} = (1 - \frac{r}{K})\frac{1}{r}$$
$$= \frac{L_{\text{uncoded}}}{r}$$

For K=100, doubling computation reduces the communication by 50% !

$$\text{communication load} \approx \frac{1}{\text{computation load}}$$



What is the Impact?

(1) Speeding Up Distributed Computing

- We can reduce the total computation time by trading Map time with Shuffle time

$$T_{\text{total}} = \mathbb{E}[T_{\text{Map}} + T_{\text{Shuffle}} + T_{\text{Reduce}}]$$

$$T_{\text{total, CDC}} = \min_{r} \quad \mathbb{E}[rT_{\text{Map}} + \frac{T_{\text{Shuffle}}}{r} + T_{\text{Reduce}}]$$

- For example, consider distributed sorting using Terasort algorithm

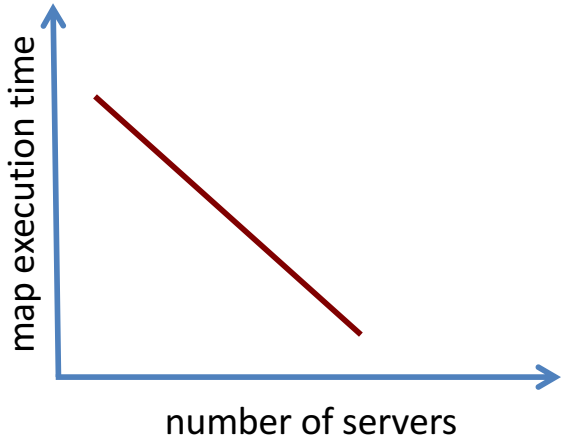
SORTING 12 GB DATA WITH $K = 16$ NODES AND 100 MBPS NETWORK SPEED

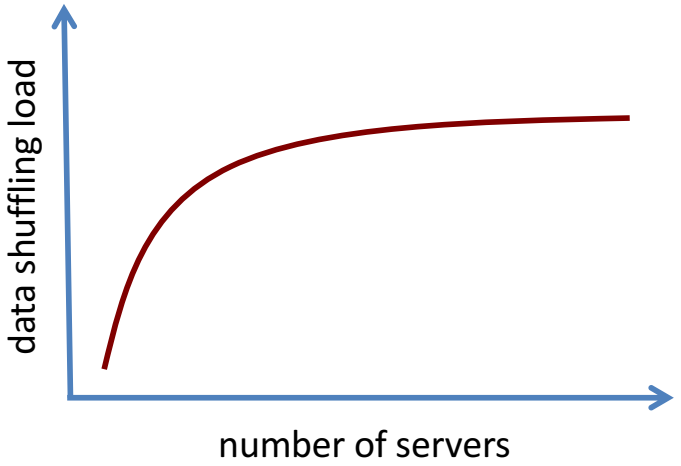
	CodeGen (sec.)	Map (sec.)	Pack/Encode (sec.)	Shuffle (sec.)	Unpack/Decode (sec.)	Reduce (sec.)	Total Time (sec.)	Speedup
TeraSort:	–	1.86	2.35	945.72	0.85	10.47	961.25	
CodedTeraSort: $r = 3$	6.06	6.03	5.79	412.22	2.41	13.05	445.56	2.16×
CodedTeraSort: $r = 5$	23.47	10.84	8.10	222.83	3.69	14.40	283.33	3.39×

CDC provides 50% - 70% speed up

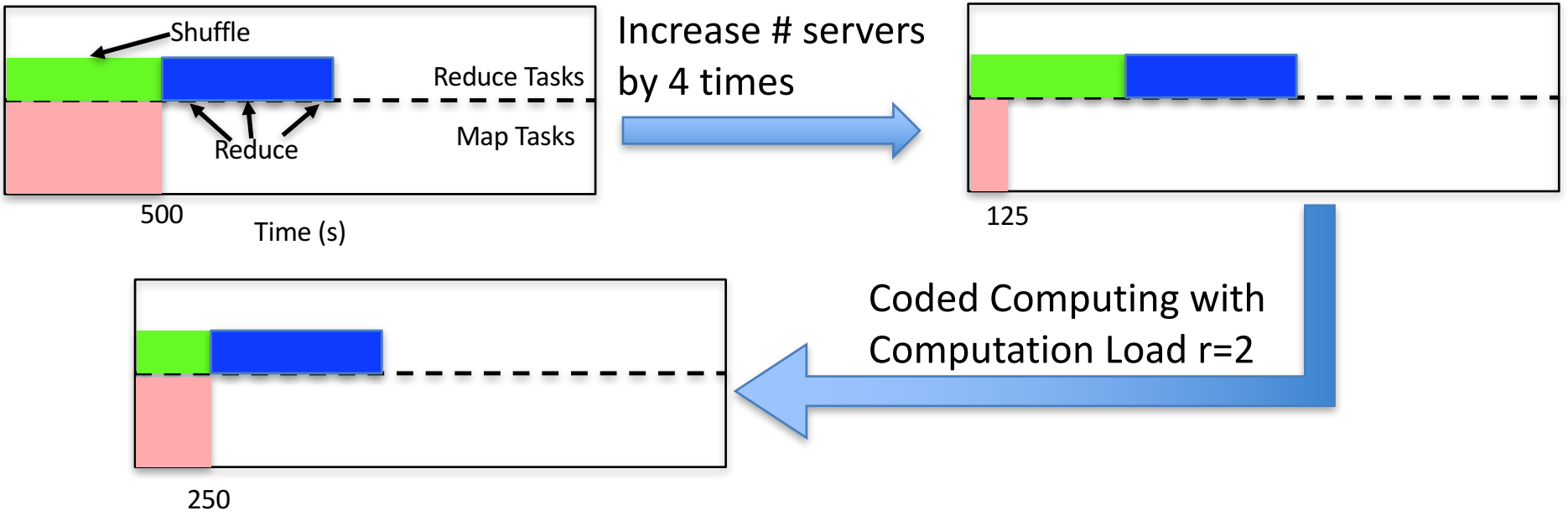
(2) Breaking the Parallelization Limit

- Current view:





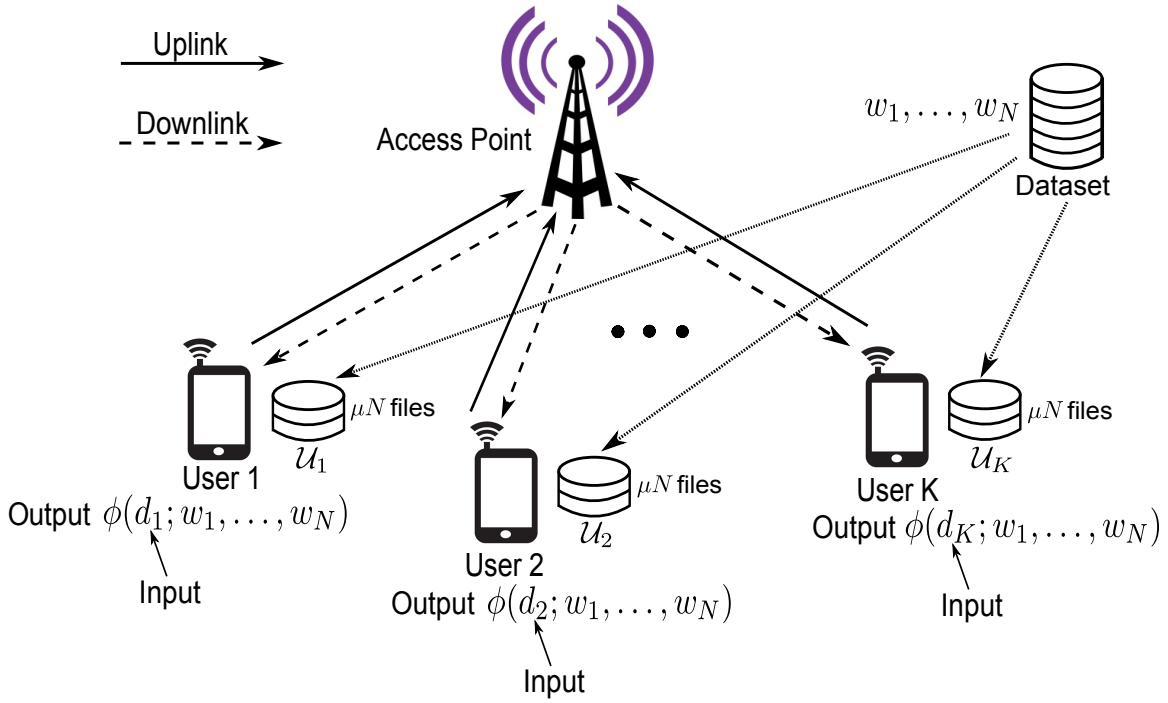
➢ Spread just enough to make Map execution time ≈ data shuffle time
  - e.g., iShuffle [Guo, et al.' 13]



(3) Scalable Wireless Distributed Computing

$$L_u^* = L_d^* = L_{coded} = \frac{K(1-\mu)}{K\mu} = \frac{1}{\mu} - 1$$

can accommodate any number of users without increasing the communication load



Conclusions and Research Directions

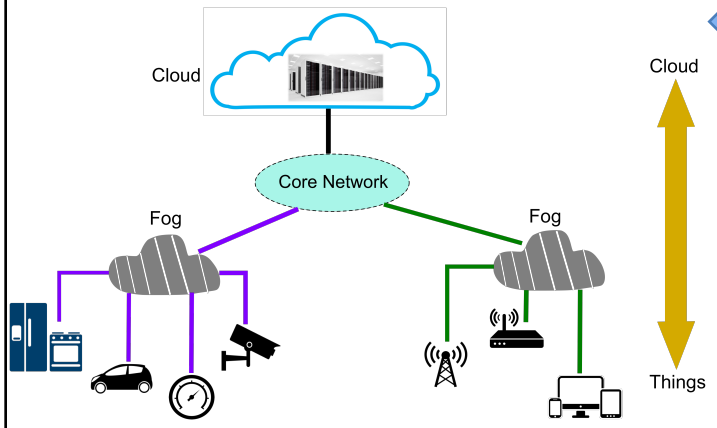
- Coding plays a fundamental role in distributed computing by enabling optimal tradeoffs between resources
- Many interesting research directions

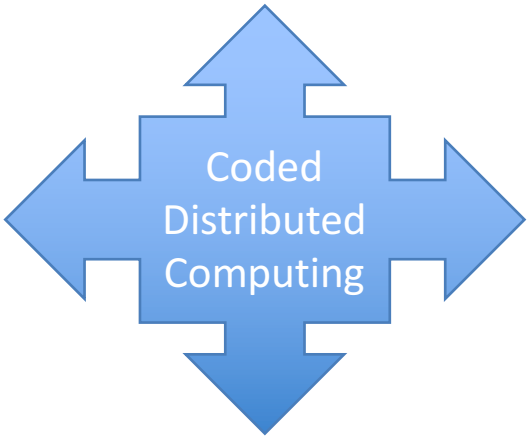
Scaling/Speeding Machine Learning and Graph Processing Algorithms

e.g., Coded Terasort, Gradient Coding, Coded Clustering, etc

Edge and Fog Computing

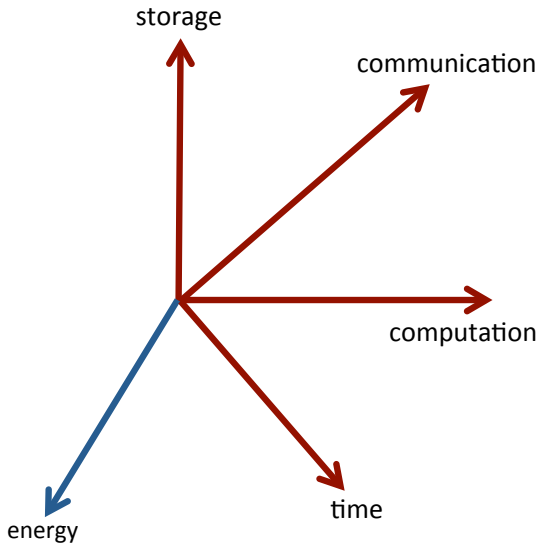
e.g., PHY-aware computing





Coding for Stragglers and Failures

Optimal Tradeoff Between Resources



Some References

- "A Fundamental Tradeoff between Computation and Communication in Distributed Computing," S. Li, M. Maddah-Ali, Q. Yu, and A. S. Avestimehr, http://arxiv.org/abs/1604.07086.

- "Coded Terasort,",S. Li, M. Maddah-Ali, and A. S. Avestimehr, 2017 International Workshop on Parallel and Distributed Computing for Large Scale Machine Learning and Big Data Analytics. https://arxiv.org/abs/1702.04850.

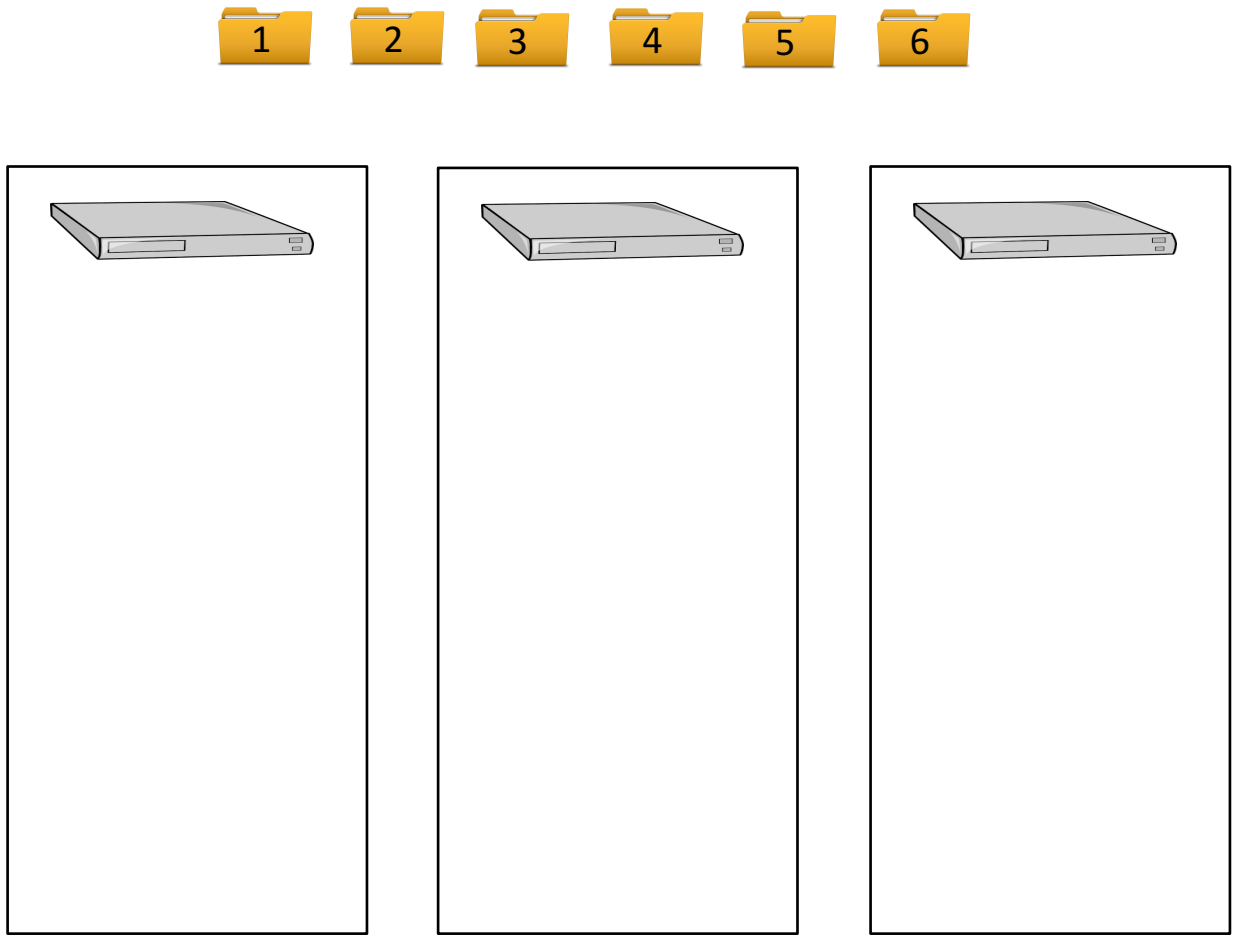
- ``Coding for Distributed Fog Computing", S. Li, M. A. Maddah-Ali and A. S. Avestimehr, to appear in IEEE Communications Magazine issue for Fog Computing and Networking, April 2017. Available online at https://arxiv.org/abs/1702.06082.

- "A Unified Coding Framework for Distributed Computing with Straggling Servers," S. Li, M. Maddah-Ali, and A. S. Avestimehr, http://arxiv.org/abs/1609.01690.

- "A Scalable Framework for Wireless Distributed Computing," S. Li, Q. Yu, M. Maddah-Ali, and A. S. Avestimehr, http://arxiv.org/abs/1608.05743.

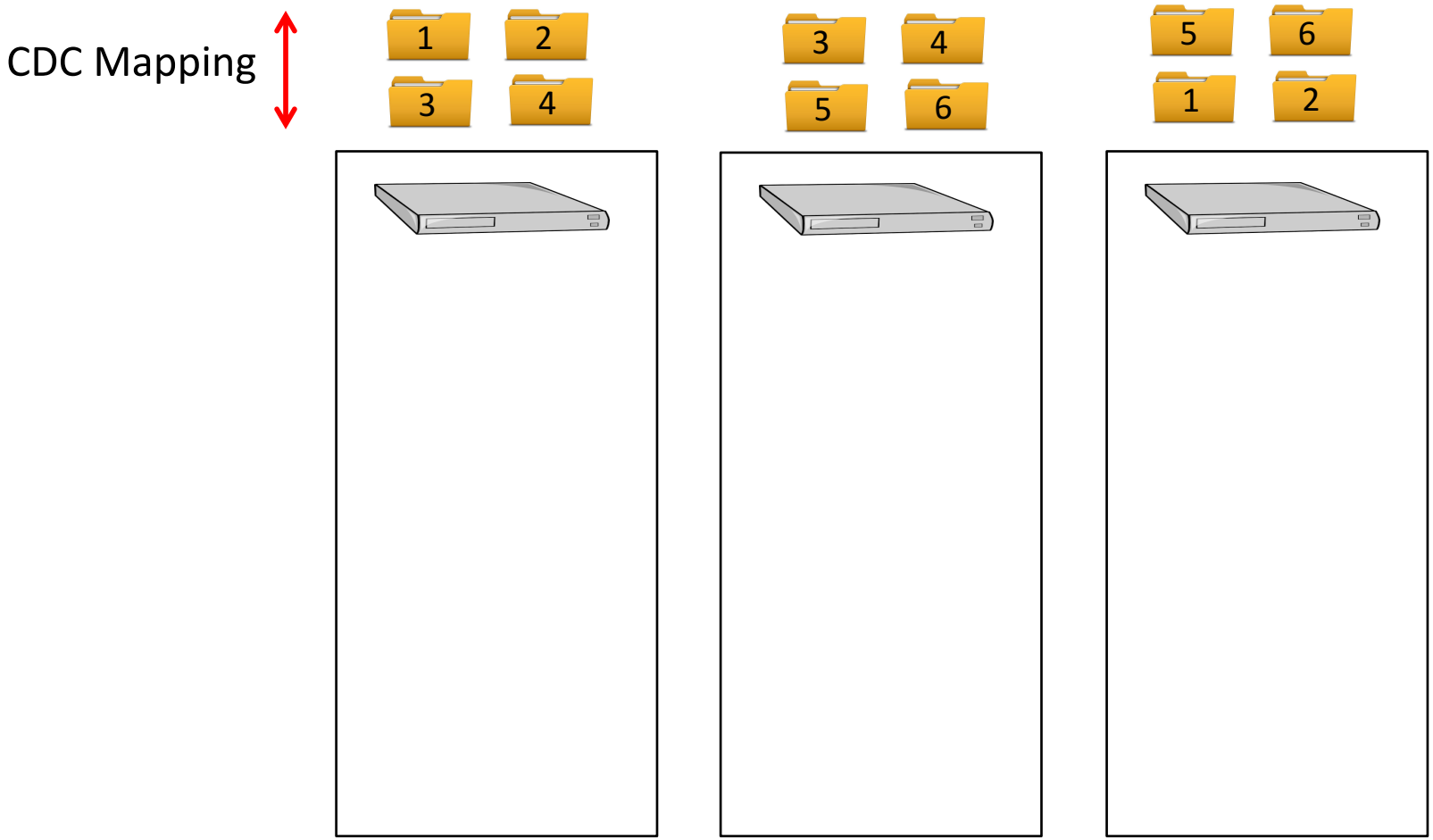
A Toy Example

- **Key Idea:** Careful assignment of tasks to servers, such that multicast coding opportunity of size r arises in the data shuffling phase
- **Example:** 6 inputs, 3 servers, 3 functions, computation load of r=2



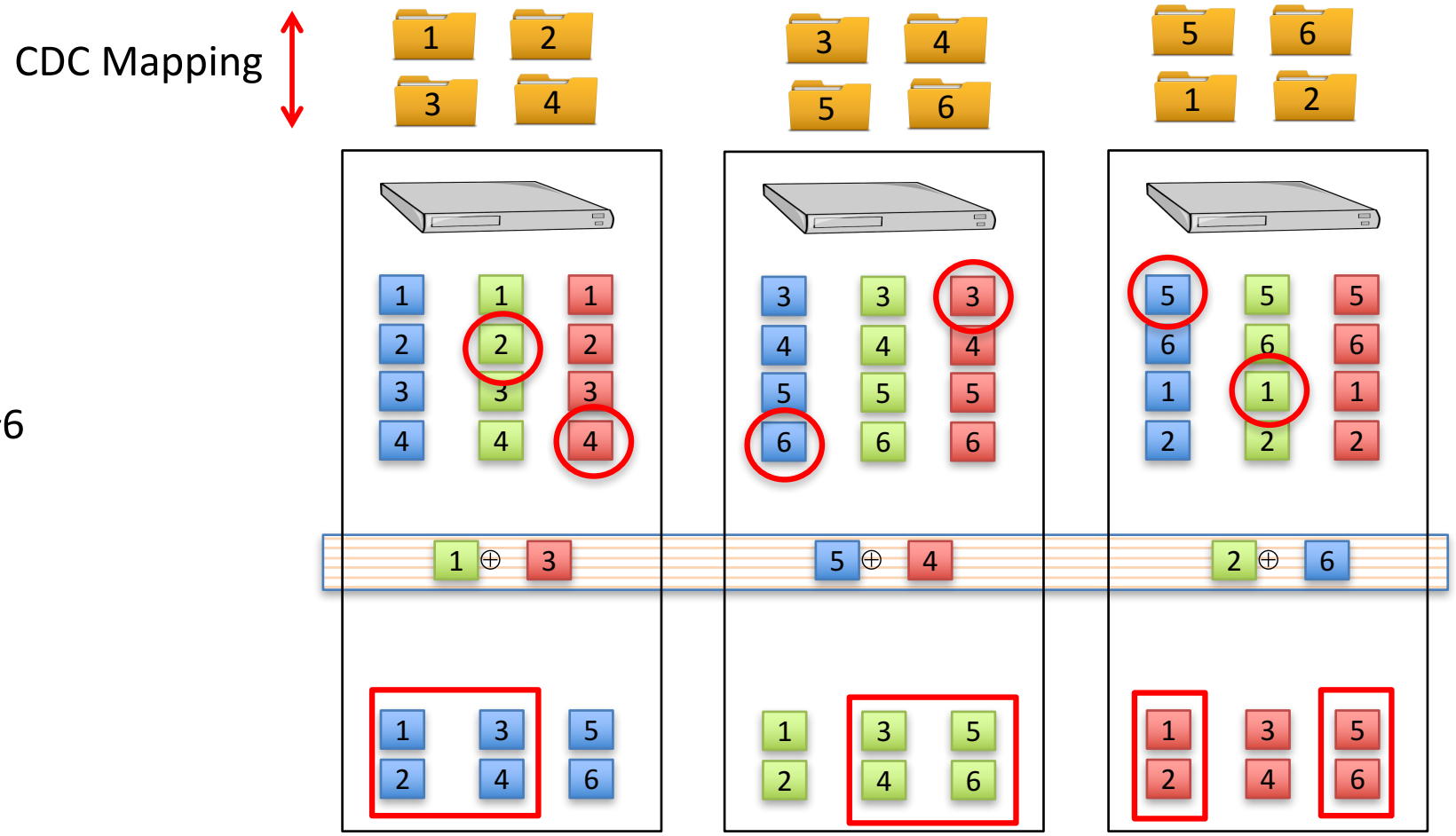
A Toy Example

- **Key Idea:** Careful assignment of tasks to servers, such that multicast coding opportunity of size r arises in the data shuffling phase
- **Example:** 6 inputs, 3 servers, 3 functions, computation load of r=2



A Toy Example

- **Key Idea:** Careful assignment of tasks to servers, such that multicast coding opportunity of size r arises in the data shuffling phase
- **Example:** 6 inputs, 3 servers, 3 functions, computation load of r=2



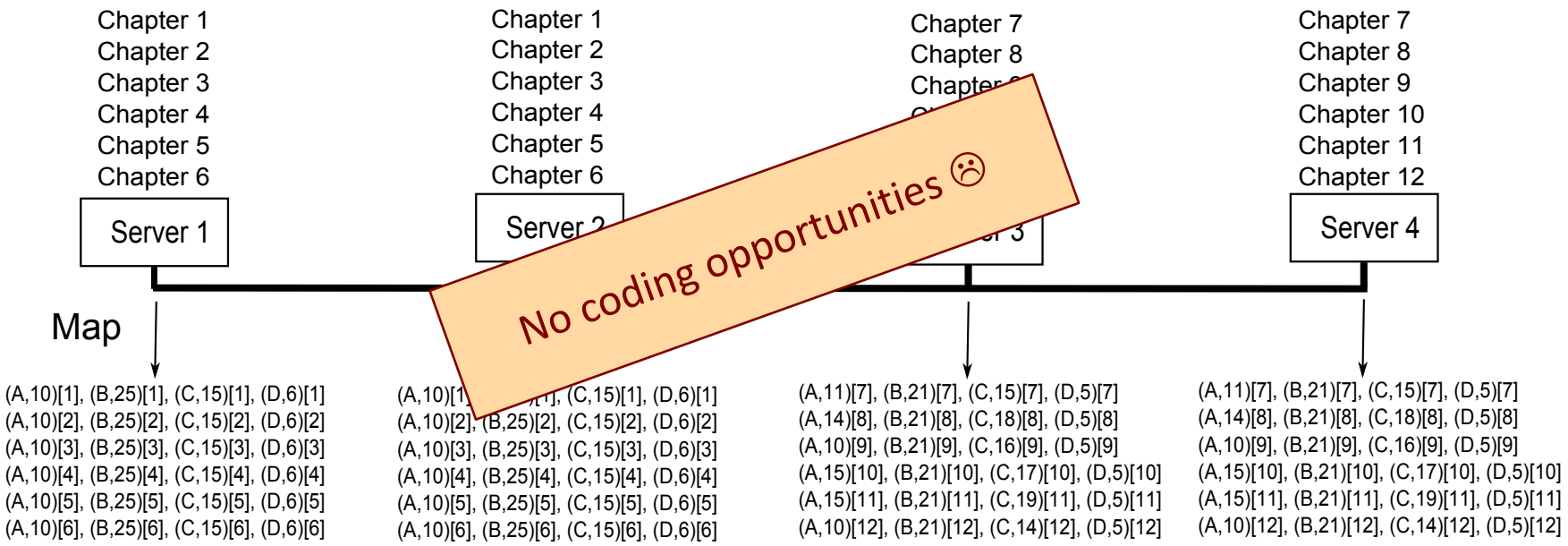
$L_{uncoded}$=6

L_{coded}=3

Each coded packet is useful for two servers

Key Challenge

- Careful assignment of MapTasks to servers, such that multicast coding opportunity of size r arises in the shuffling phase
- Example: K=Q=4, N=12, r=2



Key Challenge

- Careful assignment of MapTasks to servers, such that multicast coding opportunity of size r arises in the shuffling phase
- Example: K=Q=4, N=12, r=2