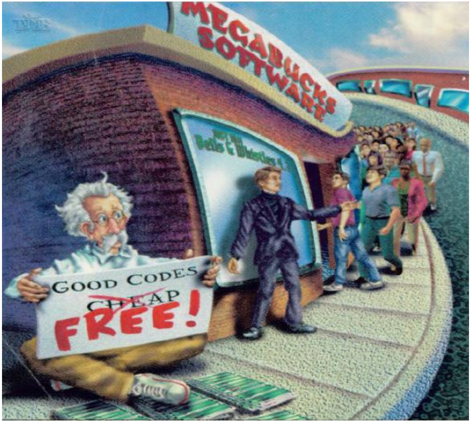
Service Rates of Codes and Vertex Covers of Graphs

Emina Soljanin, Rutgers



Joint work with

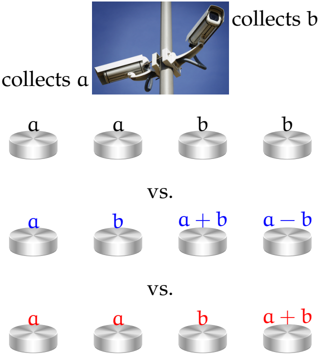
M. Aktaş, S. Andersen, A. Johnston, G. Joshi,
S. Kadhe, G. Matthews, C. Mayer, G. Yadgar

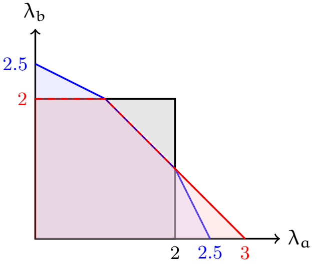
Demand for Low Latency

- Webpage download
  Amazon: 100ms $\sim$ costs 1% sales, Google: 1s $\sim$ page view drops 11%
- Interactive Tasks: 100ms - 150ms
- Online Gaming: 30ms
- Augmented Reality: 7ms - 20ms
- 5G, The Tactile Internet: 1ms

Service Rates of Codes

New applications create new performance metrics for codes.





Service Capacity Problem Formulation

System Model:

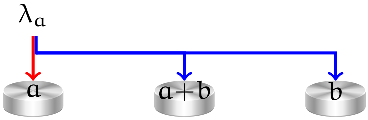
- k files $F_1, \ldots, F_k$ are stored redundantly across n servers.
- The size of each file and the size of data on each server are equal.
- Time to download a file from a server is exponential with rate 1.
- Requests to download F_i arrive at rate λ_i.

THE OBJECTIVE:

1. Determine the set of rates $(\lambda_1, \ldots, \lambda_k)$ that can be supported by the system implementing some common redundancy schemes.
2. Provide guidance on how to choose a redundancy scheme in order to maximize and/or shape the of region of supported arrival rates.

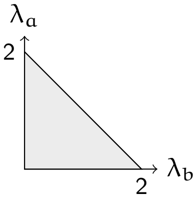
[3, 2] Simplex Code: $(a, b) \to (a, b, a+b)$

How can requests for file a be served?

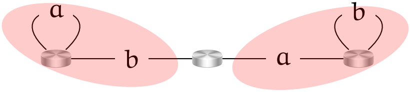


$\Rightarrow \lambda_a \leqslant 2$

Service Capacity Bound:

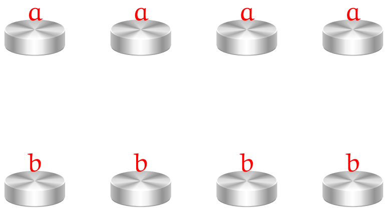


Relating to a graph covering problem:

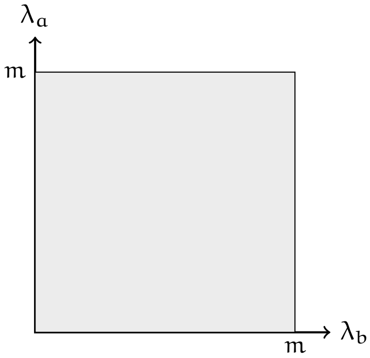


$\Rightarrow \lambda_a + \lambda_b \leqslant 2$

$[2m, 2]$ Repetition Code



Service Capacity Region:

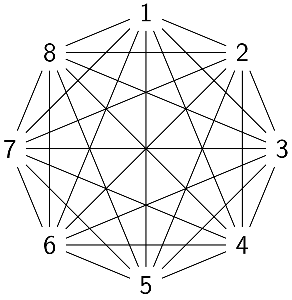


$[2m, 2]$ RS Code with no Systematic Nodes

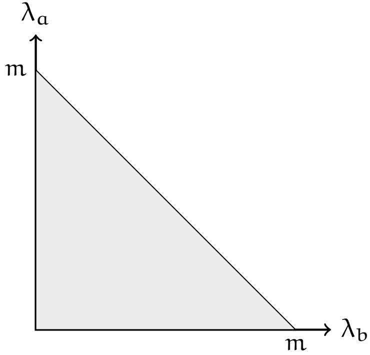
$a+b$ $a+\alpha b$ $a+\alpha^2 b$ $a+\alpha^3 b$ $a+\alpha^4 b$ $a+\alpha^5 b$ $a+\alpha^6 b$ $a+\alpha^7 b$

α is primitive in some F_q, $q \geqslant 8$

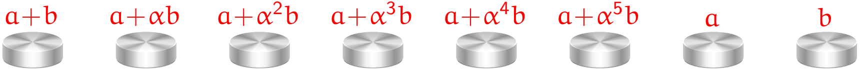
A graph to cover:



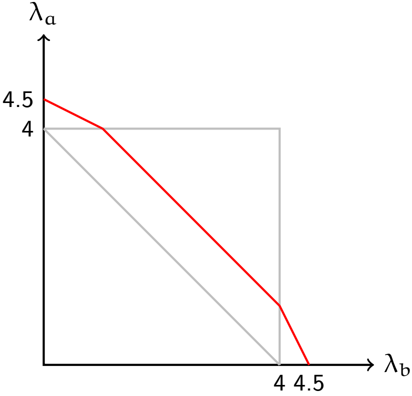
Service Capacity Region:



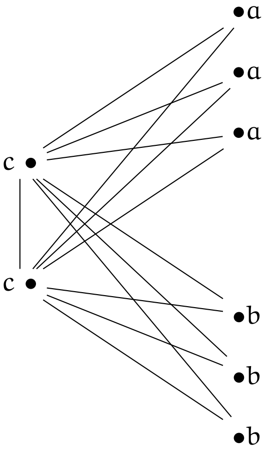
A Systematic [8, 2] RS Code



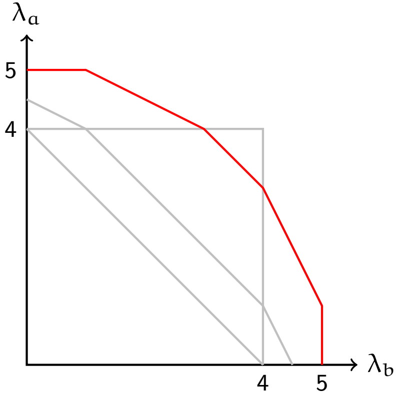
Service capacity region:



How About Both Coding and Replicating

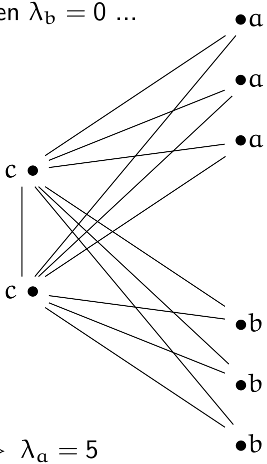


Service capacity region:



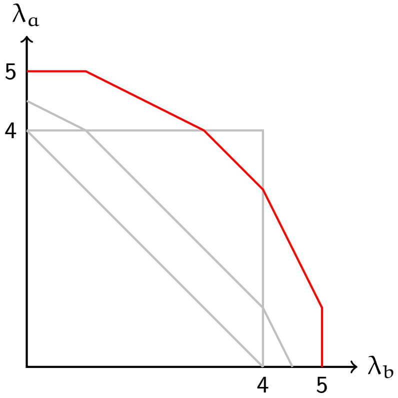
How About Both Coding and Replicating

When $\lambda_b = 0$ ...



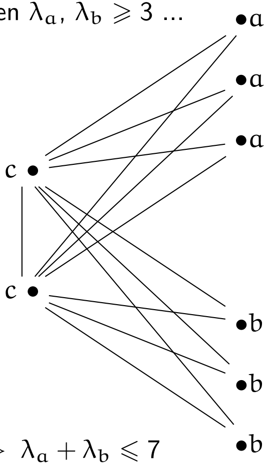
$\implies \lambda_a = 5$

Service capacity region:



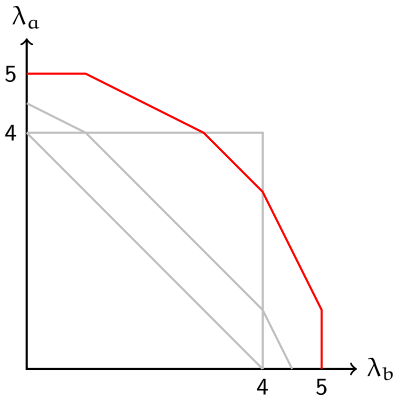
How About Both Coding and Replicating

When $\lambda_a, \lambda_b \geqslant 3$ ...



$\implies \lambda_a + \lambda_b \leqslant 7$

Service capacity region:



A General Theorem

Let there A systematic nodes for file a, B for file b, & C coded nodes, and assume that any 2 coded nodes or a coded and a systematic node can recover both a and b.

Then the service rate region is bounded by

$\lambda_a = 0$, $\lambda_b = 0$, $\lambda_a = \min\{(A+C)\mu,\ (A+\frac{B}{2}+\frac{C}{2})\mu\}$, and

$$L(\lambda_a) = \begin{cases} (B+C)\mu & \text{if } A > C \text{ and } 0 \leqslant \lambda_a \leqslant (A-C)\mu \\ -\frac{1}{2}\lambda_a + (\frac{A}{2}+B+\frac{C}{2})\mu & \text{if } A > C \text{ and } (A-C)\mu < \lambda_a \leqslant A\mu \\ -\frac{1}{2}\lambda_a + (\frac{A}{2}+B+\frac{C}{2})\mu & \text{if } A \leqslant C \text{ and } 0 \leqslant \lambda_a \leqslant A\mu \\ -\lambda_a + (A+B+\frac{C}{2})\mu & \text{if } A\mu < \lambda_a \leqslant (A+\frac{C}{2})\mu \\ -2\lambda_a + (2A+B+C)\mu & \text{if } B > C \text{ and } (A+\frac{C}{2})\mu < \lambda_a \leqslant A+C \\ -2\lambda_a + (2A+B+C)\mu & \text{if } B \leqslant C \text{ and } (A+\frac{C}{2})\mu < \lambda_a \leqslant (A+\frac{B}{2}+\frac{C}{2})\mu. \end{cases}$$

Dagstuhl development:

Gven a service region, find A, B, C s.t. $A+B+C$ is minimal.

Locally Repairable Codes (LRC) with Availability

A code has (r, t) availability if

- there are t disjoint repair groups for each data symbol &
- each repair group has at most r symbols.

A $(2, 3)$-availability code:

$$\{a, b, c\} \longrightarrow \{\boxed{a}, b, c, a+b, b+c, a+c, a+b+c\}$$

The minimum distance penalty for an $[n, k]$ code with (r, t) availability is

$$d_{\text{min}} \leqslant n - k + \left\lceil \frac{t(k-1)+1}{t(r-1)+1} \right\rceil + 2$$

There are code alphabet dependent bounds.

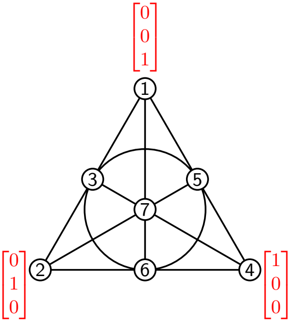
A Simplex Code Example

$$\{a, b, c\} \longrightarrow \{\boxed{a}, b, c, a+b, b+c, a+c, a+b+c\}$$

The $[7, 3, 4]$ Simplex code is a $(2, 3)$-availability code:

$$G_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Repair groups for position 1 are $\{2\&3, 4\&5, 6\&7\}$
(any two columns of G that sum to 1)



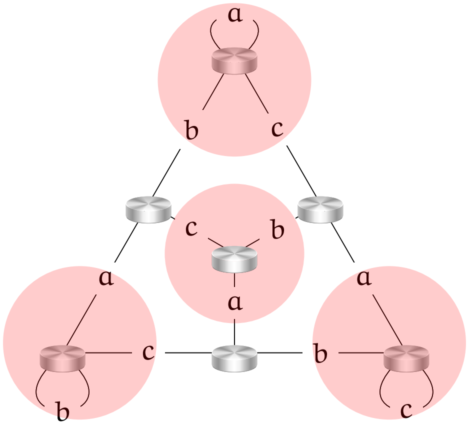
The Binary Simplex Code $\mathcal{S}_m$, $m \geqslant 2$

- G_m consist of all distinct nonzero vectors of $\mathbb{F}_2^m$
- $\mathcal{S}_m$ is a $[2^m - 1, m, 2^{m-1}]$ binary code.
- $\mathcal{S}_m$ has locality $r = 2$ and availability $t = 2^{m-1} - 1$.

Why do we care about these codes?

Considering rate, alphabet size, minimum distance, locality, availability, the binary simplex codes are in certain sense optimal.

[7, 3] Simplex Code



$[2^m - 1, m]$ Simplex Code

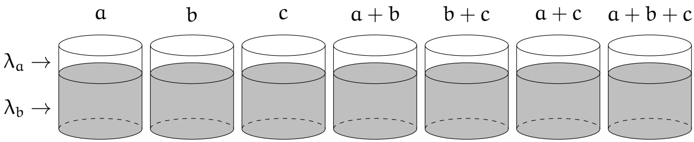
Theorem:

The service capacity region of the $[2^m - 1, m]$ Simplex coded system consists of all $\lambda_1, \ldots, \lambda_m$ s.t. $\lambda_1 + \cdots + \lambda_m \leqslant 2^{m-1}\mu$.

Proof Sketch for the Achevability:

Note that $\lambda_i \leqslant 2^{m-1}\mu$. Each server dedicates the fraction $\lambda_i/(2^{m-1}\mu)$ of its capacity solely to serving requests for file F_i.

$[7, 3]$ simplex code:



$[2^m - 1, m]$ Simplex Code

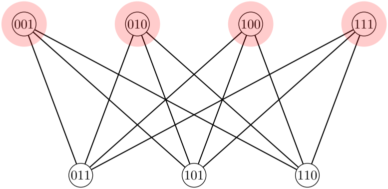
Theorem:

The service capacity region of the $[2^m - 1, m]$ Simplex coded system consists of all $\lambda_1, \ldots, \lambda_m$ s.t. $\lambda_1 + \cdots + \lambda_m \leqslant 2^{m-1}\mu$.

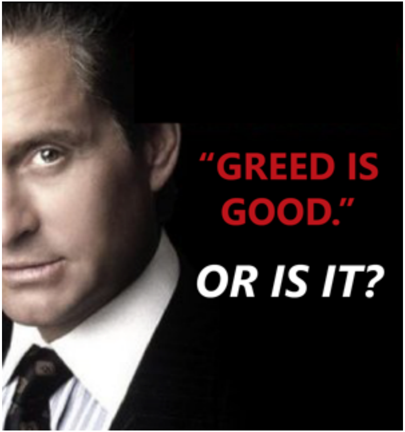
Proof Sketch for the Converse:

We consider graph Γ_m with $2^m - 1$ vertices labeled by all non-zero vectors of $\mathbb{F}_2^m$. Two vertices are connected iff their labels differ by 1. $\Longrightarrow$ Γ_m is a complete bipartite graph. The 2^{m-1} vertices with odd number of 1s cover each edge (recovery group) exactly once.

$[7, 3]$ simplex code:

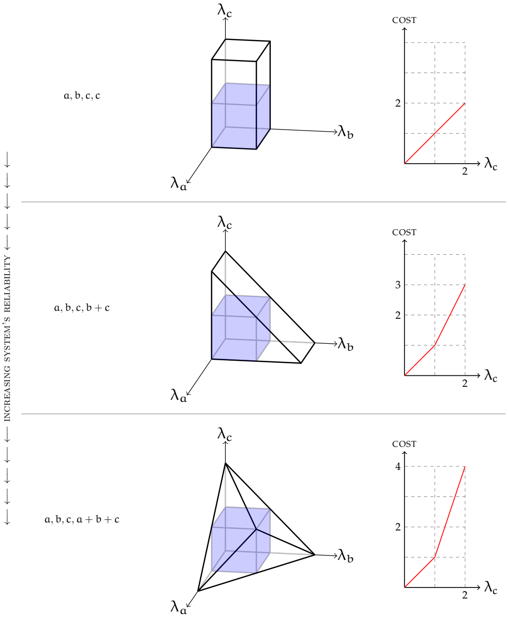


What about Systematic $[n, k]$ MDS Codes?



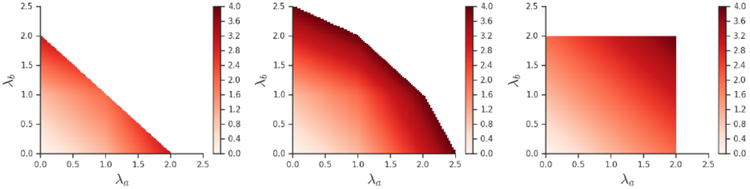
What about 1) water filling or 2) sharing?

Systems Issues



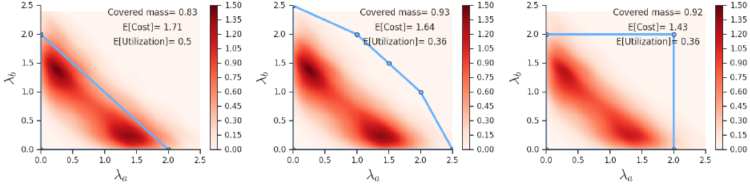
Systems Issues

Cost:



(a) $S(a, b, a+b)$ **(b)** $S(a, b, a+b, a+2b)$ **(c)** $S(a, a, b, b)$

Coverage and Utilization:



(a) $S(a, b, a+b)$ **(b)** $S(a, b, a+b, a+2b)$ **(c)** $S(a, a, b, b)$

and Σ robustness (related to Batch codes).