Coded Caching with Error Correction

B. Sundar Rajan

Nujoom Sageer Karat

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore

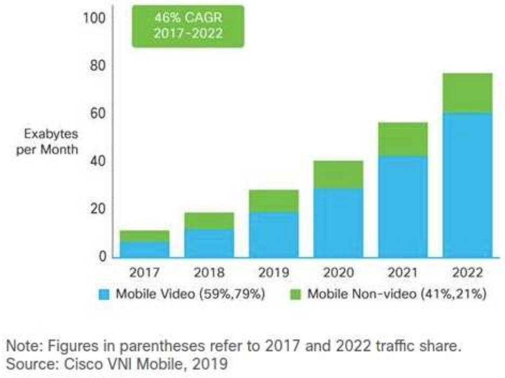
NCC 2020
IIT Kharagpur

February 2020

Outline

- Coded Caching Problem
- Index Coding Problem
- Error Correction for Index Coding Problems
- Error Correction (Optimal) for Coded Caching:
    - Uncoded Prefetching Scheme: Symmetric Batch Prefetching.
    - Coded Prefetching Scheme: CFL Scheme
    - Shared Cache scheme
    - Ali-Niesen Decentralized Scheme
    - Least Recently Sent (LRS) Online Coded Caching Scheme

Motivation



- Video on demand is increasing in popularity
- Mobile video will generate nearly four-fifths of mobile data traffic by 2022.
- Coded caching can be employed to reduce the peak hour traffic.

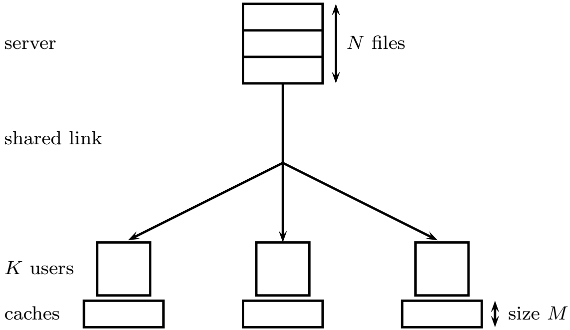
Motivation (continuation)

- Video Service providers
  - YouTube
  - Netflix streaming service
  - Amazon Instant Video
  - TikTok

- Service provider has to have large bandwidth

- Caches enable the server to reduce the transmission rates
  - High temporal traffic variability.
  - During low traffic period caches are filled with portions of content.
  - During peak traffic period server makes use of the content in cache to deliver the demands at reduced transmission rate.

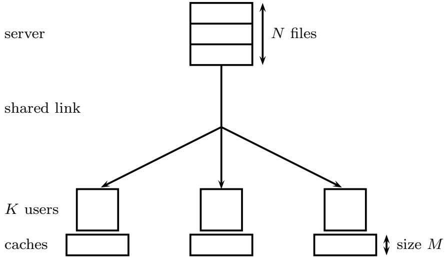
Problem Setting



- Server has N files each of size F bits.
- K users each having a cache of size MF bits

M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

Placement and Delivery Phases (Joint Design)



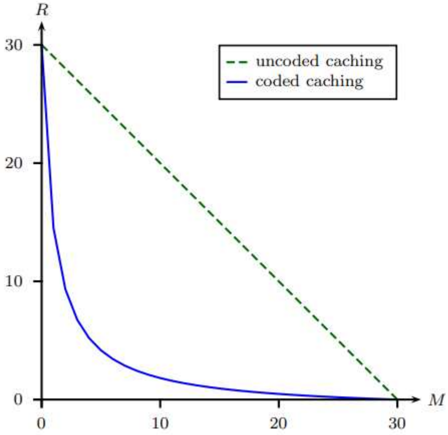
- **Placement Phase**:
  - The network not congested.
  - The main limitation is the size of the cache memories.
  - Demands not revealed
- **Delivery Phase**:
  - Demands are revealed.
  - The network is congested.
  - The main limitation is the rate.
- Finding the optimal caching scheme for a given N, K and M.

Uncoded Caching Vs Coded Caching

- No Caching
  - Rate $R = K$
- Uncoded Caching
  - $\frac{M}{N}$ fraction of each file stored in the cache.
  - Server to send to each user $(1 - \frac{M}{N})$ fraction of the demanded file.
  - Rate $R_U = K(1 - \frac{M}{N})$.
- Coded Caching
  - Apart from the local caching gain, global caching gain is also achieved.
  - Rate $R_C = K \underbrace{\left(1 - \frac{M}{N}\right)}_{\text{Local Caching Gain}} \underbrace{\left(\frac{1}{1 + \frac{KM}{N}}\right)}_{\text{Global caching gain}}$
  - Global caching gain is proportional to global cache size.

Uncoded Caching Vs Coded Caching

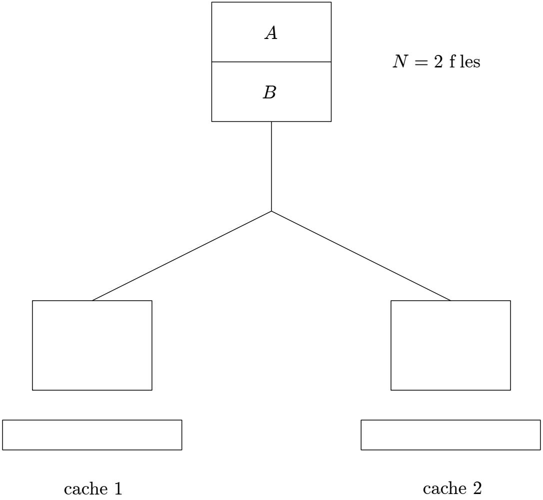
- Rate-Memory trade-off for $N = K = 30$.



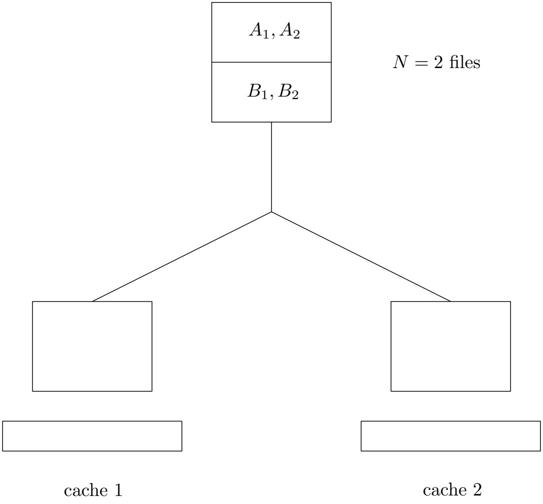
M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

Example: Maddah-Ali and Niesen Scheme

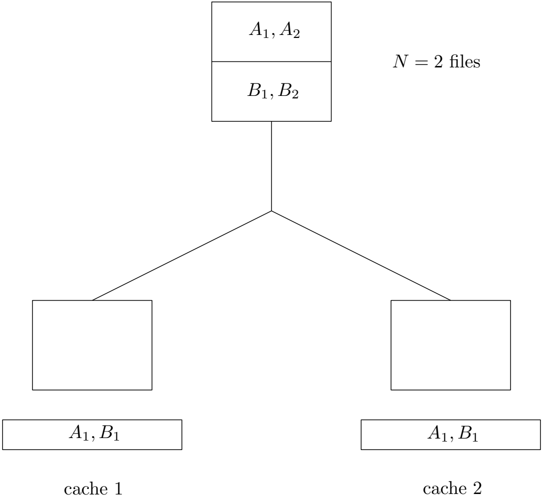
- $N = 2$ files, $K = 2$ users, cache size $M = 1$



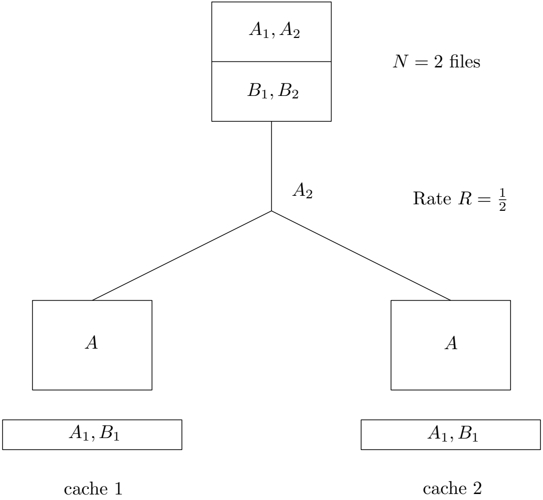
Example: $N = 2$ files, $K = 2$ users, cache size $M = 1$



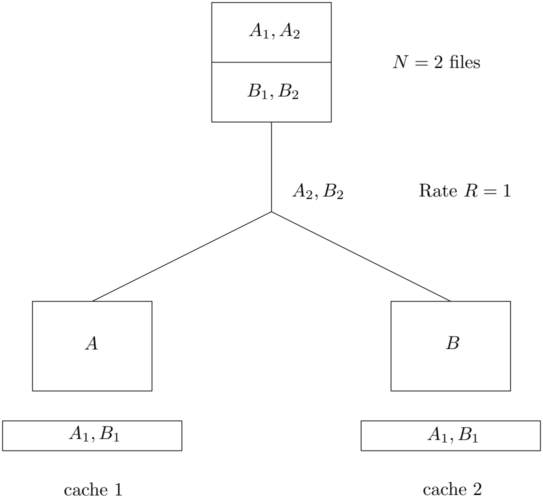
Example: $N = 2$ files, $K = 2$ users, cache size $M = 1$



Example: $N = 2$ files, $K = 2$ users, cache size $M = 1$

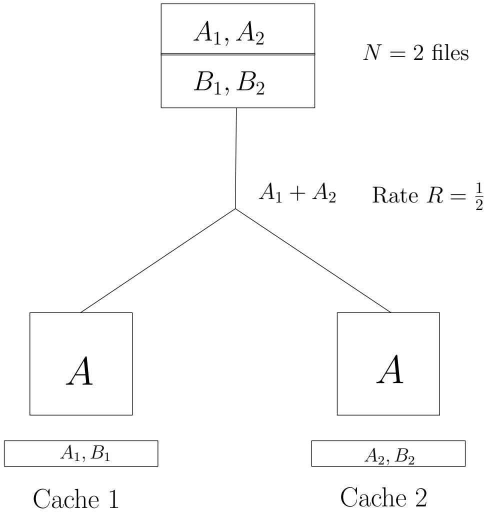


Example: $N = 2$ files, $K = 2$ users, cache size $M = 1$

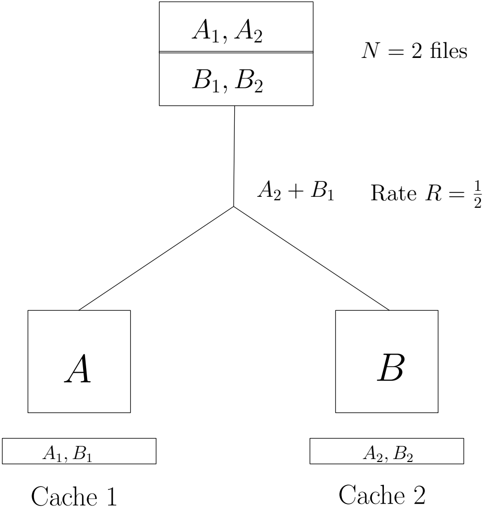


- Change in demand changes the number of transmissions required.

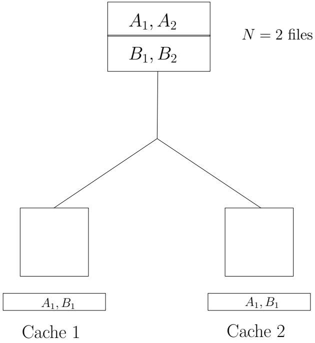
Example: $N = 2$ files, $K = 2$ users, cache size $M = 1$



Example: $N = 2$ files, $K = 2$ users, cache size $M = 1$



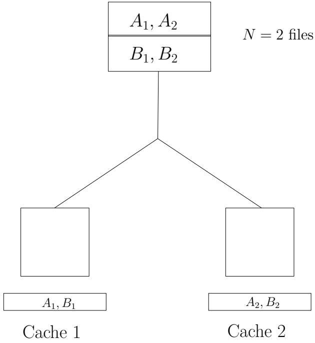
$N = 2,\ K = 2,\ M = 1$ (Coded Multicasting)



Demands	Transmissions
A, A	A_2
A, B	A_2, B_2
B, A	B_2, A_2
B, B	B_2

Worst case rate $R = 1$

Uncoded transmissions



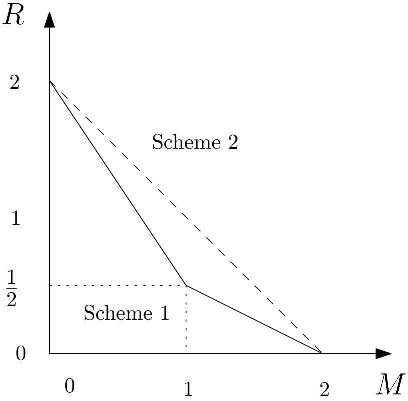
Demands	Transmissions
A, A	$A_2 + A_1$
A, B	$A_2 + B_1$
B, A	$B_2 + A_1$
B, B	$B_1 + B_2$

Worst case rate $R = 1/2$

Coded transmissions

Rate-Memory trade-off

- As memory increases, demands can be delivered at a smaller rate.
- $M = 0$: All files have to be delivered when demanded, $R = N$
- $M = N$: All files can be stored in cache, $R = 0$.
- Scheme 1 (Coded Caching) and Scheme 2 (Uncoded Caching) for $N = K = 2$



General Setup

- N files $X_1, X_2, \ldots, X_N$.
- Demand $\mathbf{d} = (d_1, d_2, \ldots, d_K)$.
  - d_i is the index of the file demanded by user i.
- Cache contents $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_K)$
  - $\mathcal{M}_i$ is indices of the bits stored in the cache of user i.
- Rate R is achievable for a demand $\mathbf{d}$ and a prefetching $\mathcal{M}$ if there exists a message of length RF such that every user is able to decode X_{d_k}.
- $R^*(\mathbf{d}, \mathcal{M})$: Minimum achievable rate for given $\mathbf{d}$ and $\mathcal{M}$.

- $R^*(\mathcal{M}) = \mathbb{E}_{\mathbf{d}}[R^*(\mathbf{d}, \mathcal{M})]$.
- $R^* = \min_{\mathcal{M}} R^*(\mathcal{M})$

- $R^*_{\text{peak}}(\mathcal{M}) = \min \max_{\mathbf{d}} R^*(\mathbf{d}, \mathcal{M})$
- $R^*_{\text{peak}} = \min_{\mathcal{M}} R^*_{\text{peak}}(\mathcal{M})$

Symmetric batch prefetching

- For $r = 1, 2, \cdots, K$, each file X_i is partitioned into $\binom{K}{r}$ non-overlapping subfiles.
- A subfile of file X_i is $X_{i,\mathcal{A}}$, where $\mathcal{A} \subseteq \{1, ..., K\}, |\mathcal{A}| = r$.
- Each user $k \in \{1, 2, \ldots, K\}$ caches subfiles $X_{i,\mathcal{A}}$ for all $i = 1, 2, \cdots, N$.
- Each user caches $\binom{K-1}{r-1} N$ subfiles each of size $F / \binom{K}{r}$ bits.
- The number of bits cached at each user is $\binom{K-1}{r-1} NF / \binom{K}{r} = \frac{rN}{K} F$ bits
- This correesponds to $M = \frac{rN}{K}$.

M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

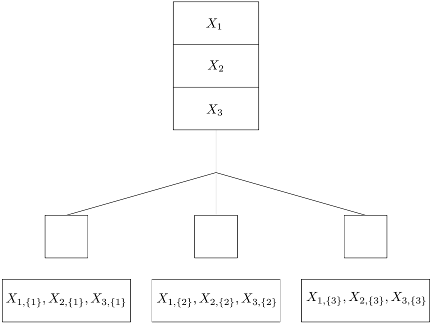
Delivery Scheme

- Delivery scheme for distinct demands.
  - For every subset $\mathcal{B}$ of $r+1$ users, $Y_{\mathcal{B}} = \bigoplus_{k\in\mathcal{B}} X_{d_k,\mathcal{B}\setminus\{k\}}$
  - The expression for peak rate: $R^*_{\text{peak}} = K\left(1-\frac{M}{N}\right)\left(\frac{1}{1+\frac{KM}{N}}\right)$
- Considering all the demand cases (including the non-distinct demand case):
  - The expression for average rate: $R^* = \mathbb{E}_{\mathbf{d}}\left[\frac{\binom{K}{r+1}-\binom{K-N_e(\mathbf{d})}{r+1}}{\binom{K}{r}}\right]$
  - $N_e(\mathbf{d})$ is the number of distinct files demanded.
- YMA scheme is optimal for symmetric batch prefetching.

Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory trade-off for caching with uncoded prefetching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017.

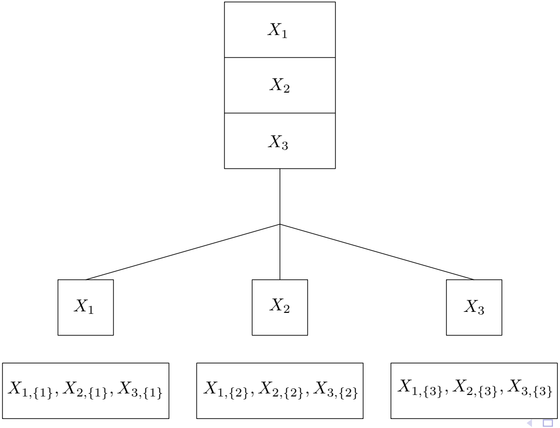
Example: Symmetric batch prefetching

- $N = 3$ files : X_1, X_2 and X_3.
- $K = 3$ users; cache size $M = 1$
- $r = \frac{KM}{N} = 1$
- Each file is partitioned into $\binom{K}{r} = 3$ subfiles
  - $X_1 : X_{1,\{1\}}, X_{1,\{2\}}, X_{1,\{3\}}$
  - $X_2 : X_{2,\{1\}}, X_{2,\{2\}}, X_{2,\{3\}}$
  - $X_3 : X_{3,\{1\}}, X_{3,\{2\}}, X_{3,\{3\}}$
- User i caches $\mathcal{M}_i = \{X_{1,\{i\}}, X_{2,\{i\}}, X_{3,\{i\}}\}$



Example: Delivery Scheme

- $\mathbf{d} = (1, 2, 3)$
- For every subset $\mathcal{A}$ of $r+1$ users $Y_{\mathcal{A}} = \bigoplus_{k \in \mathcal{A}} X_{d_k, \mathcal{A} \setminus \{k\}}$



- $\mathcal{A} = \{1, 2\}$

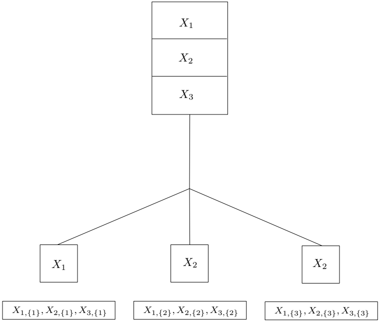
  $X_{1,\{2\}} \oplus X_{2,\{1\}}$
- $\mathcal{A} = \{1, 3\}$

  $X_{1,\{3\}} \oplus X_{3,\{1\}}$
- $\mathcal{A} = \{2, 3\}$

  $X_{2,\{3\}} \oplus X_{3,\{2\}}$

Example: Delivery Scheme

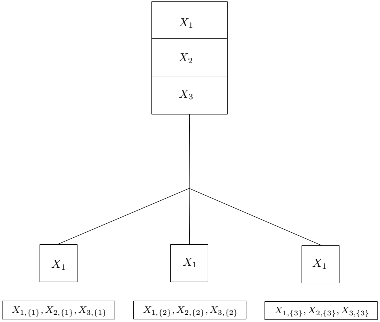
- $\mathbf{d} = (1, 2, 2)$
- Leader set of users ($\mathcal{U}$): set of users demanding distinct files.
- Here $\mathcal{U} = \{1, 2\}$.
- For every subset $\mathcal{A}$ of $r+1$ users such that $\mathcal{A} \cap \mathcal{U} \neq \emptyset$ transmit $Y_{\mathcal{A}} = \bigoplus_{k \in \mathcal{A}} X_{d_k, \mathcal{A} \setminus \{k\}}$



- $\mathcal{A} = \{1, 2\}$: $X_{1,\{2\}} \bigoplus X_{2,\{1\}}$
- $\mathcal{A} = \{1, 3\}$: $X_{1,\{3\}} \bigoplus X_{2,\{1\}}$
- $\mathcal{A} = \{2, 3\}$: $X_{2,\{3\}} \bigoplus X_{2,\{2\}}$

Example: Delivery Scheme

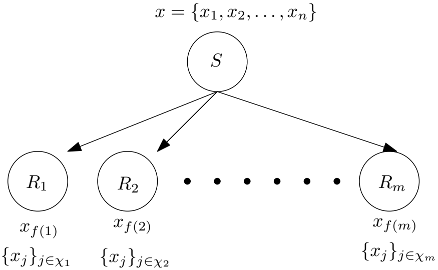
- $\mathbf{d} = (1,1,1)$
- Leader set of users ($\mathcal{U}$): set of users demanding distinct files.
- Here $\mathcal{U} = \{1\}$.
- For every subset $\mathcal{A}$ of $r+1$ users such that $\mathcal{A} \cap \mathcal{U} \neq \emptyset$ transmit $Y_{\mathcal{A}} = \bigoplus_{k \in \mathcal{A}} X_{d_k, \mathcal{A}\setminus\{k\}}$



- $\mathcal{A} = \{1,2\}$: $X_{1,\{2\}} \oplus X_{1,\{1\}}$
- $\mathcal{A} = \{1,3\}$: $X_{1,\{3\}} \oplus X_{1,\{1\}}$

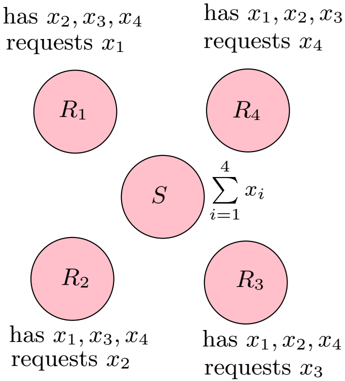
Index Coding Problem

- Sender S having n independent messages $x = \{x_1, x_2, \ldots, x_n\}$.
- m receivers $R_1, R_2, \ldots, R_m$.
- Each receiver R_i knows a subset of messages $\mathcal{K}_i \subset x$ (or equivalently $\{x_j\}_{j \in \mathcal{X}_i}$).
- Each receiver R_i demands a subset of messages $\mathcal{W}_i \subseteq x$ (or equivalently and w.l.o.g. a single message $x_{f(i)}$).



[1] Z. Bar-Yossef, Z. Birk, T. S. Jayram and T. Kol, "Index coding with side information," in *Proc. 47th Annu. IEEE Symp. Found. Comput. Sci.*, 2006, pp. 197-206.

A simple example



- Naive approach - transmit four messages one after another.
- Index coding - Only one transmission (sum of all the messages).

IC and δ-ECIC: Definitions

- An *index code* over $\mathbb{F}_q$ is an encoding function $\mathfrak{C} : \mathbb{F}_q^n \to \mathbb{F}_q^N$ such that for each receiver R_i, $i \in [m]$, there exists a decoding function $\mathfrak{D}_i : \mathbb{F}_q^N \times \mathbb{F}_q^{|\chi_i|} \to \mathbb{F}_q$ satisfying

$$\textit{for} \quad IC : \quad \mathfrak{D}_i(\mathfrak{C}(x), x_{\chi_i}) = x_{f(i)} \quad \forall\, x \in \mathbb{F}_q^n$$

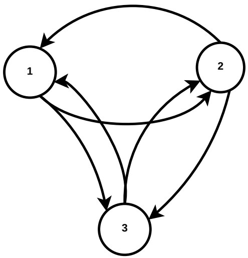
$$\textit{for} \quad ECIC : \quad \mathfrak{D}_i(\mathfrak{C}(x) + \epsilon_i, x_{\chi_i}) = x_{f(i)} \quad \forall x \in \mathbb{F}_q^n, \epsilon_i \in \mathbb{F}_q^N, wt(\epsilon_i) \leq \delta.$$

- The parameter N is the *length* of the index code.
- An index code is linear if $\mathfrak{C}$ is a linear transformation.
- For linear index codes $\mathfrak{C}(x) = xL$ where L is an $n \times N$ matrix.

Son Hoang Dau, V. Skachek,and Yeow Meng Chee, "Error Correction for Index Coding With Side Information", *IEEE Trans.Inf.Theory*, vol. 59, Issue: 3, pp. 1517-1531, 2013.

Graphical Representation of Index Coding problems

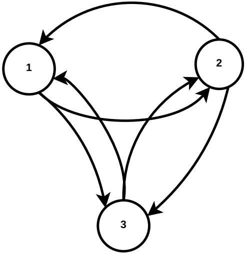
- Any Index Coding problem can be represented using side-information (directed) hypergraph [1] with vertex set $\mathcal{V}=[n]$ hyper-edge set $\mathcal{E}_{\mathcal{H}}$, where $\mathcal{E}_{\mathcal{H}} = \{(f(i), \mathcal{X}_i) \ : \ i \in [m]\}$.
- Vertex v_i represents message x_i and each hyper-edge a receiver.
- Example: $n = 3, m = 4$: $f(1) = 1$, $f(2) = 2$, $f(3) = 3$, $f(4) = 2$, $\mathcal{X}_1 = \{2, 3\}$, $\mathcal{X}_2 = \{1\}$, $\mathcal{X}_3 = \{1, 2\}$, $\mathcal{X}_4 = \{3\}$, the $\mathcal{H}$ is



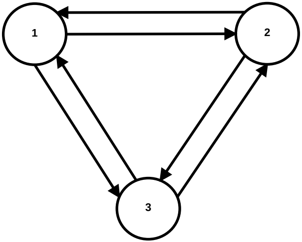
[1] N. Alon, A. Hassidim, E. Lubetzky, U. Stav, and A. Weinstein, "Broadcasting with side information", in *Proc. 49th Annu. IEEE symp. Found. Comput. Sci*, pp.823-832, 2008.

Graphical Representation of Index Coding problems

- For general index coding problems, the side information hypergraph $\mathcal{H}$ cannot be replaced by corresponding directed graph in Figure (b).



(a)



(b)

(a) Side-information hypergraph

(b) Directed graph

Min-rank $\kappa_q(\mathcal{H})$ of a hypergraph $\mathcal{H}$

- The min-rank of the hypergraph $\mathcal{H}$ over the finite field with q elements $\mathbb{F}_q$ is defined in [1] as,

$$\kappa_q(\mathcal{H}) \triangleq min\{rank_q(\{\mathbf{v}_i + \mathbf{e}_{f(i)}\}_{i\in[m]}) : \mathbf{v}_i \in \mathbb{F}_q^n, \mathbf{v}_i \lhd \mathcal{X}_i\},$$

- $\mathbf{v}_i \lhd \mathcal{X}_i$ denotes that the support of $\mathbf{v}_i$ is a subset of $\mathcal{X}_i$.
- $\mathbf{e_i} = (\underbrace{0,...,0}_{i-1}, 1, \underbrace{0,...,0}_{n-i}) \in \mathbb{F}_q^n$.

---

[1] Z. Bar-Yossef, Z. Birk, T.S. Jayaram, and T. Kol, "Index coding with side information", in *Proc. 47th Annu. IEEE symp. Found. Comput. Sci*, pp.197-206, Oct. 2006.

Example: Min-rank Calculation

- $n = 3$ messages, $m = 4$ receivers.
- Demands: $f(1) = 1$, $f(2) = 2$, $f(3) = 3$ and $f(4) = 2$.
- Side-information: $\mathcal{X}_1 = \{2, 3\}$, $\mathcal{X}_2 = \{1\}$, $\mathcal{X}_3 = \{1, 2\}$, and $\mathcal{X}_4 = \{3\}$.
- $v_1 + e_{f(1)} = [1\ x\ x]$, $v_2 + e_{f(2)} = [x\ 1\ 0]$, $v_3 + e_{f(3)} = [x\ x\ 1]$ and $v_4 + e_{f(4)} = [0\ 1\ x]$, where x can be either 0 or 1.
- The min-rank, $\kappa_2(\mathcal{H})$ is the minimum rank of the $n \times m$ matrix

$$\begin{bmatrix} 1 & x & x & 0 \\ x & 1 & x & 1 \\ x & 0 & 1 & x \end{bmatrix}.$$

- For this example, the minimum rank is 2, corresponding realization is $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$.
- Index coded transmissions: $x_1 + x_3$ and x_2.

Generalized Independence Number $\alpha(\mathcal{H})$

- Consider the case where there are n messages and m receivers and receiver R_i demanding the message $x_{f(i)}$ and having side information $\{x_j\}_{j \in \mathcal{X}_i}$. For each receiver R_i, the set $\mathcal{J}(\mathcal{H})$ is defined as,

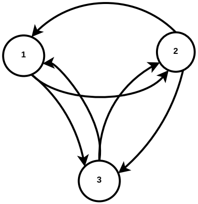
$$\mathcal{J}(\mathcal{H}) \triangleq \cup_{i \in [m]} \{\{f(i)\} \cup Y_i : Y_i \subseteq [n] \setminus \mathcal{X}_i\}.$$

- A subset H of $[n]$ is called a generalized independent set in $\mathcal{H}$, if every non-empty subset of H belongs to $\mathcal{J}(\mathcal{H})$.
- The size of the largest generalized independent set in $\mathcal{H}$ is called the generalized independence number and is denoted by $\alpha(\mathcal{H})$.

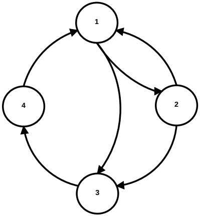
Generalized Independence Number $\alpha(\mathcal{H})$

A set of vertices H of size K forms a generalized independent set if,

- Every vertex in H has at least one hyperedge originating from it.
- For every k-subset of H and for all $k \in \{2, 3, ..., K\}$, there should be at least one hyperedge originating from one vertex and not ending at any of the remaining $k - 1$ vertices.

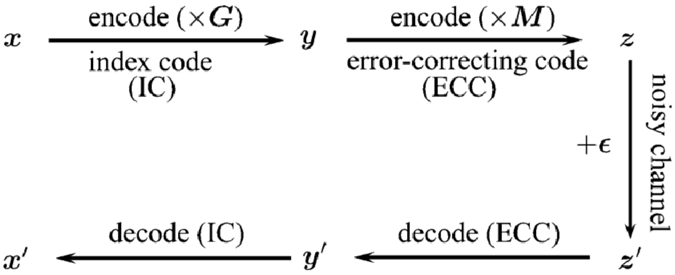


(a) $\alpha(\mathcal{H}) = 2$ and maximum independent set is H=$\{1, 2\}$



(b) $\alpha(\mathcal{H}) = 3$ and maximum independent set is H=$\{1, 2, 3\}$

Insufficiency of concatenation



Mere application of an ECC on top of an IC may not be optimal.

S.H.Dau, V. Skachek and Y.M. Chee, "Error Correction for Index Coding with Side Information," IEEE Trans. Information Theory, Vol.59, No.3, pp. 1517-1531, March 2013,

Bounds for ECIC

- $\alpha(\mathcal{H}) \leq \kappa_q(\mathcal{H})$
- The optimal length $\mathcal{N}_q[\mathcal{H}, \delta]$ of linear ECIC is bounded by

$$\underbrace{N_q[\alpha(\mathcal{H}), 2\delta+1]}_{\alpha-bound} \quad \leq \quad \mathcal{N}_q[\mathcal{H}, \delta] \quad \leq \quad \underbrace{N_q[\kappa_q(\mathcal{H}), 2\delta+1]}_{\kappa-bound}.$$

- The upper bound is met by concatenating an optimal index code with an optimal error correcting code.
- When the bounds are met with equality, concatenation of an optimal error correcting code with an optimal index code gives an optimal error correcting index code.

S.H.Dau, V. Skachek and Y.M. Chee, "Error Correction for Index Coding with Side Information,"
IEEE Trans. Information Theory, Vol.59, No.3, pp. 1517-1531, March 2013.

Insufficiency of concatenation (Example)

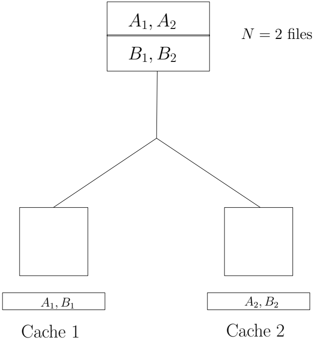
- Let $n = m = 5$ with $\mathcal{W}_i = i, \quad i = 1, 2, 3, 4, 5$ and
- $\mathcal{K}_1 = \{2, 5\}$, $\mathcal{K}_2 = \{1, 3\}$, $\mathcal{K}_3 = \{2, 4\}$, $\mathcal{K}_4 = \{3, 5\}$, $\mathcal{K}_5 = \{1, 4\}$.
- $\alpha(\mathcal{H}) = 2$ and $\kappa_2(\mathcal{H}) = 3$
  - α-bound =8; κ-bound =10 for $\delta = 2$.
- An optimal 2-ECIC of length 9 is

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

S.H.Dau, V. Skachek and Y.M. Chee, "Error Correction for Index Coding with Side Information,"
IEEE Trans. Information Theory, Vol.59, No.3, pp. 1517-1531, March 2013,

Coded Caching problem and Index Coding

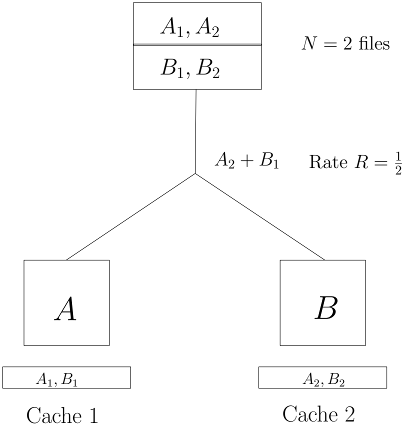
- For a given demand, delivery phase of caching problem is an index coding problem.



- Corresponds to an index coding problem with $n = 4$ messages and $K = m = 2$ receivers.

Coded caching problem and index coding

- Receiver 1
  - Demands A_2
  - Side-information A_1 and B_1.
- Receiver 2
  - Demands B_1
  - Side-information A_2 and B_2.
- Min-rank $= 1$



- As demand changes the corresponding index coding problem changes.
- Delivery phase of a (N, K, M) coded caching problem consists of N^K index coding problems.

Symmetric batch prefetching and index coding

- $\mathcal{I}(\mathcal{M}_{SB}, \mathbf{d})$
  - $\mathcal{M}_{SB}$: Symmetric batch prefetching
  - $\mathbf{d}$: demand vector
- $\alpha(\mathcal{M}_{SB}, \mathbf{d})$: Generalized independence number of $\mathcal{I}(\mathcal{M}_{SB}, \mathbf{d})$
- $\kappa(\mathcal{M}_{SB}, \mathbf{d})$: Min-rank of $\mathcal{I}(\mathcal{M}_{SB}, \mathbf{d})$
- $\alpha(\mathcal{M}_{SB}, \mathbf{d}) \leq \kappa(\mathcal{M}_{SB}, \mathbf{d})$

N. S. Karat, A. Thomas and B. S. Rajan, "Error Correction in Coded Caching With Symmetric Batch Prefetching," in *IEEE Transactions on Communications*, vol. 67, no. 8, Aug. 2019.

Example : Symmetric batch prefetching for $N = K = 3$, $M = 1$

- Subfile division:

$$X_1 : X_{1,\{1\}}, X_{1,\{2\}}, X_{1,\{3\}}$$

$$X_2 : X_{2,\{1\}}, X_{2,\{2\}}, X_{2,\{3\}}$$

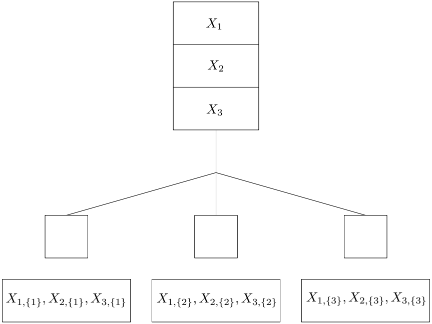
$$X_3 : X_{3,\{1\}}, X_{3,\{2\}}, X_{3,\{3\}}$$

- The cache contents:

$$\mathcal{M}_1 : X_{1,\{1\}}, X_{2,\{1\}}, X_{3,\{1\}}$$

$$\mathcal{M}_2 : X_{1,\{2\}}, X_{2,\{2\}}, X_{3,\{2\}}$$

$$\mathcal{M}_3 : X_{1,\{3\}}, X_{2,\{3\}}, X_{3,\{3\}}$$



- For $\mathbf{d} = (1, 2, 3)$, the corresponding index coding problem has $n = 6$ messages and $m = 3$ receivers.

Example: Distinct Demands

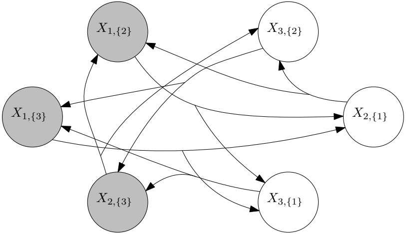
- The demand vector is $\mathbf{d} = (1, 2, 3)$.
- Generalized independent set B for this case:

$$B = \{X_{1,\{2\}}, X_{1,\{3\}}, X_{2,\{3\}}\}.$$

- $\alpha(\mathcal{M}_{SB}, \mathbf{d}) \geq 3$.

- $n = 6$ and the message set is

$$x = \{X_{1,\{2\}}, X_{1,\{3\}}, X_{2,\{1\}}, X_{2,\{3\}}, X_{3,\{1\}}, X_{3,\{2\}}\}$$



- Delivery phase transmissions

$$X_{1,\{2\}} \oplus X_{2,\{1\}} \quad X_{2,\{3\}} \oplus X_{3,\{2\}} \quad X_{1,\{3\}} \oplus X_{3,\{1\}}$$

- $\kappa(\mathcal{M}_{SB}, \mathbf{d}) \leq 3$
- $\alpha(\mathcal{M}_{SB}, \mathbf{d}) = \kappa(\mathcal{M}_{SB}, \mathbf{d}) = 3$.

N. S. Karat, A. Thomas and B. S. Rajan, "Error Correction in Coded Caching With Symmetric Batch Prefetching," in *IEEE Transactions on Communications*, vol. 67, no. 8, Aug. 2019.

Example: Non-distinct demand case

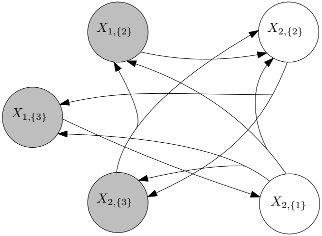
- The demand vector is $\mathbf{d} = (1, 2, 2)$.
- Generalized independent set B for this case:

$$B = \{X_{1,\{2\}}, X_{1,\{3\}}, X_{2,\{3\}}\}.$$

- $\alpha(\mathcal{M}_{SB}, \mathbf{d}) \geq 3$.

- $n = 5$ and the message set is

$$x = \{X_{1,\{2\}}, X_{1,\{3\}}, X_{2,\{1\}}, X_{2,\{2\}}, X_{2,\{3\}}\}$$



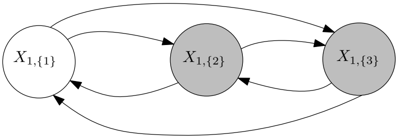
- Delivery phase transmissions

$$X_{1,\{2\}} \oplus X_{2,\{1\}} \quad X_{2,\{3\}} \oplus X_{2,\{2\}} \quad X_{1,\{3\}} \oplus X_{2,\{1\}}$$

- $\kappa(\mathcal{M}_{SB}, \mathbf{d}) \leq 3$
- $\alpha(\mathcal{M}_{SB}, \mathbf{d}) = \kappa(\mathcal{M}_{SB}, \mathbf{d}) = 3$.

Example: Non-distinct demand case

- All users demand the same file X_1.
  - $\mathbf{d} = (1, 1, 1)$
- $n = 3$ and the message set is $x = \{X_{1,\{1\}}, X_{1,\{2\}}, X_{1,\{3\}}\}$



- Generalized independent set B for this case:

$$B = \{X_{1,\{2\}}, X_{1,\{3\}}\}.$$

- $\alpha(\mathcal{M}_{SB}, \mathbf{d}) \geq 2.$

- Delivery phase transmissions

$$X_{1,\{1\}} \oplus X_{1,\{2\}}$$

$$X_{1,\{1\}} \oplus X_{1,\{3\}}$$

- $\kappa(\mathcal{M}_{SB}, \mathbf{d}) \leq 2$
- $\alpha(\mathcal{M}_{SB}, \mathbf{d}) = \kappa(\mathcal{M}_{SB}, \mathbf{d}) = 2.$

List of α and κ values for all the possible demands

Demand	α	κ	Demand	α	κ	Demand	α	κ
(X_1, X_2, X_3)	3	3	(X_1, X_3, X_3)	3	3	(X_2, X_2, X_3)	3	3
(X_1, X_3, X_2)	3	3	(X_3, X_1, X_3)	3	3	(X_2, X_3, X_2)	3	3
(X_2, X_1, X_3)	3	3	(X_3, X_3, X_1)	3	3	(X_3, X_2, X_2)	3	3
(X_2, X_3, X_1)	3	3	(X_1, X_1, X_2)	3	3	(X_2, X_3, X_3)	3	3
(X_3, X_1, X_2)	3	3	(X_1, X_2, X_1)	3	3	(X_3, X_2, X_3)	3	3
(X_3, X_2, X_1)	3	3	(X_2, X_1, X_1)	3	3	(X_3, X_3, X_2)	3	3
(X_1, X_2, X_2)	3	3	(X_1, X_1, X_3)	3	3	(X_1, X_1, X_1)	2	2
(X_2, X_1, X_2)	3	3	(X_1, X_3, X_1)	3	3	(X_2, X_2, X_2)	2	2
(X_2, X_2, X_1)	3	3	(X_3, X_1, X_1)	3	3	(X_3, X_3, X_3)	2	2

- For any demand $\mathbf{d}$,

$$\alpha(\mathcal{M}_{SB}, \mathbf{d}) = \kappa(\mathcal{M}_{SB}, \mathbf{d}).$$

Number of messages in the corresponding index coding problem

- N files, K users.
- $\mathbf{d}$: Demand vector
- $N_e(\mathbf{d})$: Number of distinct files demanded
- WLOG let the demanded files be $X_1, X_2, \ldots, X_{N_e(\mathbf{d})}$
- f_i: The number of users demanding file i, $i \in \{1, 2, \ldots, N_e(\mathbf{d})\}$.
- The corresponding index coding problem has n messages, where

$$n = N_e(\mathbf{d})\binom{K}{r} - \sum_{i=1}^{N_e(\mathbf{d})} \binom{K - f_i}{r - f_i}$$

Generalization

- Construction of a generalized independent set

$$B(\mathbf{d}) = \bigcup_{i \in [N_e(\mathbf{d})]} \{X_{d_i,\{a_1,\ldots,a_r\}} : a_1, \ldots, a_r \neq 1, 2, \ldots, i\}.$$

- $N_e(\mathbf{d})$: Number of distinct files in $\mathbf{d}$.
- $\alpha(\mathcal{M}_{SB}, \mathbf{d}) \geq |B(\mathbf{d})|$.
- We have shown that

$$|B(\mathbf{d})| = \binom{K}{r+1} - \binom{K - N_e(\mathbf{d})}{r+1}.$$

- $|B(\mathbf{d})|$ is the number of transmissions in YMA delivery scheme.
- $\kappa(\mathcal{M}_{SB}, \mathbf{d}) \leq |B(\mathbf{d})|$
- YMA scheme is optimal for symmetric batch prefetching.

N. S. Karat, A. Thomas and B. S. Rajan, "Error Correction in Coded Caching With Symmetric Batch Prefetching," in *IEEE Transactions on Communications*, vol. 67, no. 8, Aug. 2019.

Optimal error correcting delivery scheme for $\mathcal{M}_{SB}$

- Delivery phase of a (K, N, M) caching scheme corresponds to N^K index coding problems.
- For every index coding problem $\mathcal{I}(\mathcal{M}_{SB}, \mathbf{d})$, we have

$$\alpha(\mathcal{M}_{SB}, \mathbf{d}) = \kappa(\mathcal{M}_{SB}, \mathbf{d}).$$

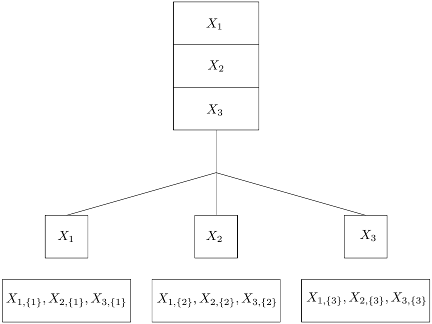
- α-bound and κ-bound satisfy

$$N_q[\alpha(\mathcal{H}), 2\delta + 1] = \mathcal{N}_q[\mathcal{H}, \delta] = N_q[\kappa_q(\mathcal{H}), 2\delta + 1].$$

- Optimal error correcting delivery scheme: Concatenation of YMA delivery scheme with optimal classical error correcting code.

N. S. Karat, A. Thomas and B. S. Rajan, "Error Correction in Coded Caching With Symmetric Batch Prefetching," in *IEEE Transactions on Communications*, vol. 67, no. 8, Aug. 2019.

Example: Error correcting delivery scheme



- $\mathbf{d} = (1, 2, 3)$.
- Corresponding index coding problem has: $n = 6$ and $m = 3$.
- Three YMA scheme transmissions: $X_{1,\{2\}} \oplus X_{2,\{1\}}$, $X_{2,\{3\}} \oplus X_{3,\{2\}}$, $X_{1,\{3\}} \oplus X_{3,\{1\}}$

- The corresponding index coding matrix:

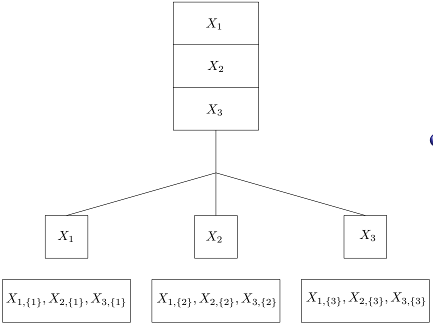
$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- $\kappa(\mathcal{M}_{SB}, \mathbf{d}) = 3$.
- To correct $\delta = 1$ error: concatenate with an optimal error correcting code of dimension 3.

Example: Error correcting delivery scheme

- Generator matrix

$$M = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$



- $\mathcal{N}_2[\kappa(\mathcal{M}_{SB}, \mathbf{d}), 3] = 6$
- Error correcting delivery matrix:

$$GM = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

- Delivery scheme
  - $Y_1 : X_{1,\{2\}} \oplus X_{2,\{1\}}$
  - $Y_2 : X_{2,\{3\}} \oplus X_{3,\{2\}}$
  - $Y_3 : X_{1,\{3\}} \oplus X_{3,\{1\}}$
  - $Y_4 : X_{1,\{2\}} \oplus X_{2,\{1\}} \oplus X_{2,\{3\}} \oplus X_{3,\{2\}}$
  - $Y_5 : X_{1,\{2\}} \oplus X_{2,\{1\}} \oplus X_{1,\{3\}} \oplus X_{3,\{1\}}$
  - $Y_6 : X_{2,\{3\}} \oplus X_{3,\{2\}} \oplus X_{1,\{3\}} \oplus X_{3,\{1\}}$

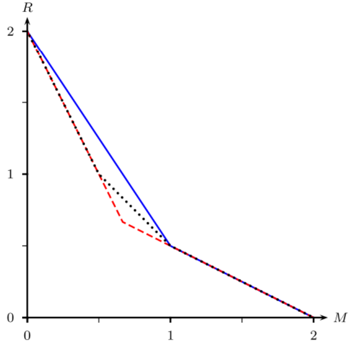
Uncoded and Coded Prefetching

- Coded caching with uncoded prefetching
  - Parts of files are placed as such without coding.
- Coded caching with coded prefetching
  - Parts of files are coded and placed.
  - Performs better than the uncoded prefetching case.
- The fundamental scheme by Maddah-Ali and Niesen: Centralized scheme with uncoded prefetching.

Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: improved bounds for users with small buffers," *IET Communications*, vol. 10, no. 17, pp. 2315-2318, Nov. 2016.

J. Gómez-Vilardebó, "Fundamental Limits of Caching: Improved Rate-Memory Tradeoff With Coded Prefetching," in *IEEE Transactions on Communications*, vol. 66, no. 10, Oct. 2018.

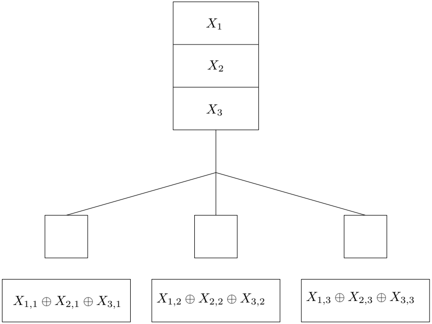
Coded and Uncoded Prefetching: $N = K = 2$



- The solid blue curve indicates Maddah-Ali and Niesen Scheme-uncoded prefetching.
- **The dotted black curve** indicates the improvement using **coded prefetching**.
- The dashed red curve indicates the cut-set based lower bound.

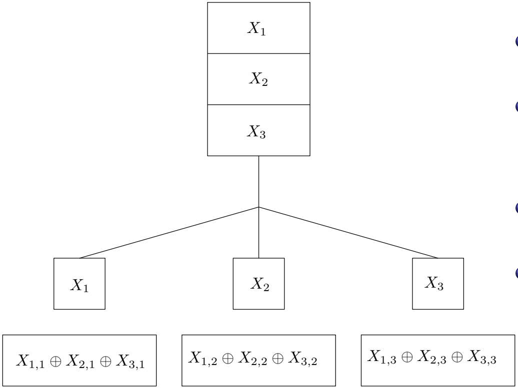
Coded prefetching

- Example $N = K = 3$ and $M = 1/3$.
- Each file is split into three subfiles and XORed messages are cached.



Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: improved bounds for users with small buffers," *IET Communications*, vol. 10, no. 17, pp. 2315-2318, Nov. 2016.

CFL Scheme



- $\mathbf{d} = (1, 2, 3)$.
- CFL transmissions: $X_{2,1}, X_{3,1}, X_{1,2}, X_{3,2}, X_{1,3}, X_{2,3}$.
- Rate, $R_{\text{CFL}} = 6/3 = 2$.
- Rate achieved when YMA scheme is used: $R_{\text{YMA}} = 2.33$.

- Coded prefetching performs better than uncoded prefetching.

CFL scheme : K users, N files and $M = 1/K$

- Prefetching scheme depends on the relation between N and K.
- For $N = K$
  - Each file split into N subfiles: $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,N})$.
  - Cache of user i: $C_i = X_{1,i} \oplus X_{2,i} \oplus \ldots X_{N,i}$.
- For $K > N$
  - Each file split into NK subfiles: $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,NK})$.
  - Cache of user i: $C_i = X_{1,N(i-1)+j} \oplus \ldots \oplus X_{N,N(i-1)+j}$, for $j = 1, 2, \ldots, N$.
- Rate achieved in CFL scheme

$$R = \begin{cases} N_e(\mathbf{d}) & \text{if } N_e(\mathbf{d}) \leq N-1 \\ N - \frac{N}{K} & \text{if } N_e(\mathbf{d}) = N \end{cases}.$$

Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: improved bounds for users with small buffers," *IET Communications*, vol. 10, no. 17, pp. 2315-2318, Nov. 2016.

CFL Scheme and Generalized Index Coding Problem

- For CFL prefetching scheme, coded packets are placed in cache of each user.
- For fixed demand, the delivery phase is a generalized index coding problem $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$.
- Sender S having N independent messages $X = \{x_1, x_2, \ldots, x_N\}$.
- i^{th} Receiver demands some linear combination $D_i X$
- i^{th} Receiver knows s_i independent linear combinations $V^{(i)}X, \ where \ V^{(i)} \in \mathbb{F}_q^{s_i \times N}$.
- The row space of $V^{(i)}$ is $\mathcal{X}^{(i)}$
- D, $m \times N$ matrix with ith row as D_i, represents demand of all m users.

K.W. Shum, D. Mingjun, and C.W. Sung, "Broadcasting with coded side information," in *IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2012.

Min-rank and GIN of Generalized Index Coding Problem

- The min-rank of the instance $\mathcal{I}$ of the GICP over $\mathbb{F}_q$ is defined as

$$\kappa(\mathcal{I}) = \min\{\text{rank}(A + D) : A \in \mathbb{F}_q^{m\times N}, A_i \in \mathcal{X}^{(i)}, i \in [m]\}.$$

- $\mathcal{Z}^{(i)}$ is defined as

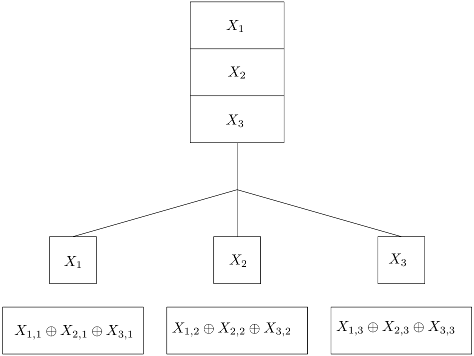
$$\mathcal{Z}^{(i)} \triangleq \{Z \in \mathbb{F}_q^{N\times 1} : V^{(i)}Z = 0, D_i Z \neq 0\}.$$

- $\mathcal{Z} = \cup_{i\in[m]}\mathcal{Z}^{(i)} \cup \{0\}$.
- Generalized independence number $\alpha(\mathcal{I})$ is the maximum dimension of any subspace of $\mathbb{F}_q^N$ in $\mathcal{Z}$.

Eimear Byrne and Marco Calderini, "Error correction for index coding with coded side information," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3712-3728, Jun. 2017.

Example: $\alpha(\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d}))$

- $N = K = 3$ and $M = 1/3$.



- $\mathbf{d} = (1, 2, 3)$.
- Set of equations
  - $e_1 : X_{1,1} \oplus X_{2,1} \oplus X_{3,1} = 0$
  - $e_2 : X_{1,2} \oplus X_{2,2} \oplus X_{3,2} = 0$
  - $e_3 : X_{1,3} \oplus X_{2,3} \oplus X_{3,3} = 0$
- $\mathcal{S}$: set of all vectors which satisfy these equations.
- $\mathcal{S}$ is a subspace of $\mathbb{F}_q^9$

Example Contd...

- $\dim(\mathcal{S}) \geq 6$
- $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ has 9 messages and 9 receivers.
- $\mathcal{Z}^{(d_i,j)} \triangleq \{Z \in \mathbb{F}_q^9 : e_i, X_{d_i,j} \neq 0\}$.
  - All vectors in $\mathcal{S}$, with $X_{d_i,j} \neq 0$ belongs to $\mathcal{Z}^{(d_i,j)}$.
- All vectors in $\mathcal{S}$ lie in $\mathcal{Z} = \cup_{i,j \in [3]} \mathcal{Z}^{(d_i,j)} \cup \{\mathbf{0}\}$.
- $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 6$.
- From CFL delivery scheme : $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq 6$.
- $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = \alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = 6$.

Example: Non-distinct demands

- Let $\mathbf{d} = (1, 1, 1), N_e(\mathbf{d}) = 1$
- In addition to e_1, e_2 and e_3 consider

$$e_4 : X_{2,1} = 0$$

$$e_5 : X_{2,2} = 0$$

$$e_6 : X_{2,3} = 0$$

- Let $\mathcal{S}$ be the solution subspace.
- $\dim(\mathcal{S}) \geq 9 - 6 = 3$.
- All vectors in $\mathcal{S}$ are present in $\mathcal{Z}$.
- CFL delivery scheme: $\kappa(\mathcal{M}_{\mathrm{CFL}}, \mathbf{d}) \leq 3$
- $\alpha(\mathcal{M}_{\mathrm{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\mathrm{CFL}}, \mathbf{d}) = 3$.

Generalization for $N = K$ and $N < K$

Theorem

For $N = K$ *and* $M = 1/N$,

$$\alpha(\mathcal{M}_{CFL}, \mathbf{d}) = \kappa(\mathcal{M}_{CFL}, \mathbf{d}) = \begin{cases} N(N_e(\mathbf{d})) & \text{if } N_e(\mathbf{d}) \leq N-1 \\ N(N-1) & \text{if } N_e(\mathbf{d}) = N \end{cases},$$

where $N_e(\mathbf{d})$ *is the number of distinct demands.*

Theorem

For $N < K$ *and* $M = 1/K$,

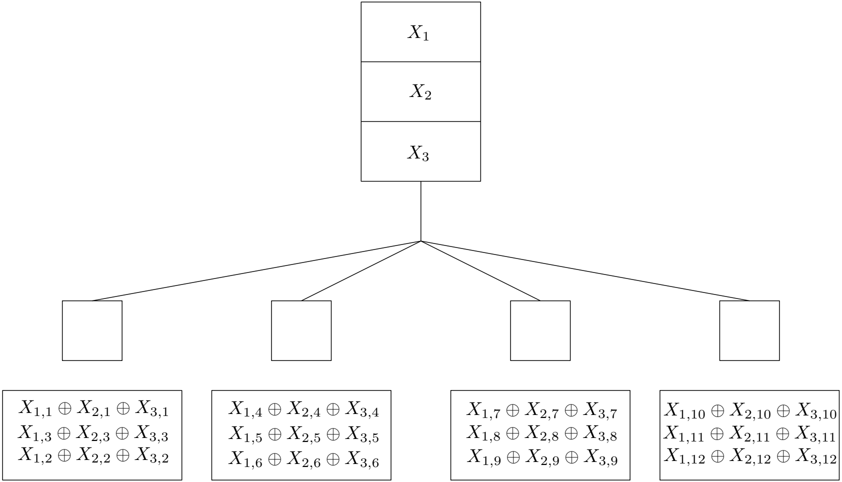
$$\alpha(\mathcal{M}_{CFL}, \mathbf{d}) = \kappa(\mathcal{M}_{CFL}, \mathbf{d}) = \begin{cases} NK(N_e(\mathbf{d})) & \text{if } N_e(\mathbf{d}) \leq N-1 \\ N^2(K-1) & \text{if } N_e(\mathbf{d}) = N, \end{cases}$$

where $N_e(\mathbf{d})$ *is the number of distinct demands.*

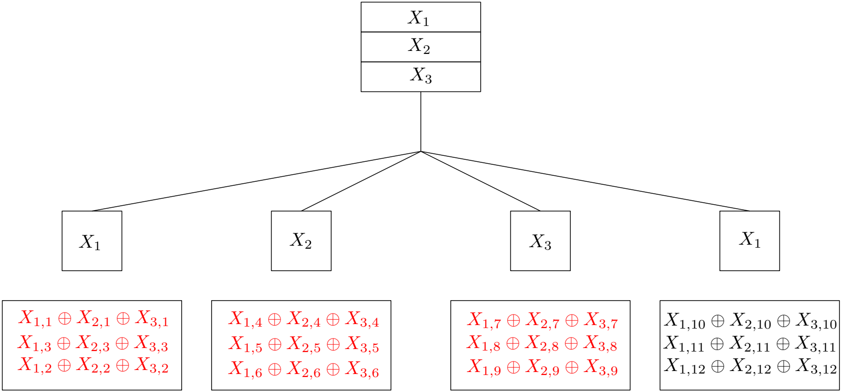
N. S. Karat, A. Thomas and B. S. Rajan, "Optimal Error Correcting Delivery Scheme for an Optimal Coded Caching Scheme with Small Buffers," IEEE ISIT 2018, Vail.

Example: $K > N$, $M = 1/K$ (N=3, K =4, M=1/4).

- Each file is split into $NK = 12$ subfiles.



- Let $\mathbf{d} = (1, 2, 3, 1)$.
- $N_e(\mathbf{d}) = N = 3$



$e_{i,j} : (X_{1,3(i-1)+j} \oplus X_{2,3(i-1+j)} \oplus X_{3,3(i-1)+j}) = 0$ for $i = 1, 2, 3$ and $j = 1, 2, 3$.

Example

- Consider the nine equations

$$e_{i,j} : (X_{1,3(i-1)+j} \oplus X_{2,3(i-1+j)} \oplus X_{3,3(i-1)+j}) = 0$$

  for $i = 1, 2, 3$ and $j = 1, 2, 3$.
- $\mathcal{S}$ be the set of vectors satisfying these equations.
- $\dim(\mathcal{S}) \geq 36 - 9 = 27$.
- $\mathcal{Z}^{(i,j)} \triangleq \{Z \in \mathbb{F}_q^{36} : e_{i,1}, e_{i,2}, e_{i,3}, X_{d_i,j} \neq 0\}$ for $i \in [4]$
- $\mathcal{Z} = \cup_{i\in[4],j\in[12]} \mathcal{Z}^{(i,j)} \cup \{\mathbf{0}\}$
- Every vector in $\mathcal{S}$ belongs to $\mathcal{Z}$.
- Hence $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 27$.
- CFL delivery scheme: $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq 27$.
- $\alpha(\mathcal{M}_{\text{CFL}}) = \kappa(\mathcal{M}_{\text{CFL}}) = 27$.

Example: Optimal error correcting delivery scheme

- $N = K = 3,\ M = 1/3$.
- Let $N_e(\mathbf{d}) = 3$ and $\mathbf{d} = (1, 2, 3)$.
- $\alpha(\mathcal{M}_{\mathrm{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\mathrm{CFL}}, \mathbf{d}) = 6$.
- CFL transmissions: $Y_1 : X_{2,1},\ Y_2 : X_{3,1},\ Y_3 : X_{1,2},\ Y_4 : X_{3,2},$ $Y_5 : X_{1,3}$ and $Y_6 : X_{2,3}$. The corresponding index coding matrix G is given by

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

Example

- $\delta = 1$.
- Generator matrix corresponding to $[10, 6, 3]_2$ code is

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

- Concatenation gives rise to 10 transmissions.

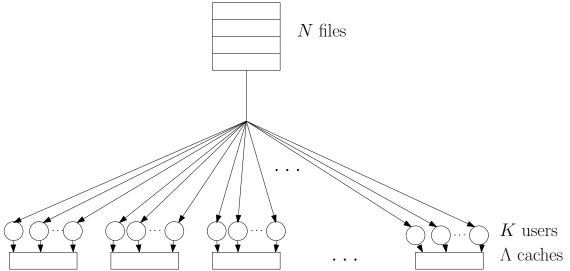
- $Y_1 : X_{2,1}$
- $Y_2 : X_{3,1}$
- $Y_3 : X_{1,2}$
- $Y_4 : X_{3,2}$
- $Y_5 : X_{1,3}$
- $Y_6 : X_{2,3}$
- $Y_7 : X_{2,1} \oplus X_{3,1} \oplus X_{1,2}$
- $Y_8 : X_{2,1} \oplus X_{3,2} \oplus X_{1,3}$
- $Y_9 : X_{3,1} \oplus X_{3,2} \oplus X_{2,3}$
- $Y_{10} : X_{1,2} \oplus X_{1,3} \oplus X_{2,3}$

- Error correcting delivery matrix:

$$GM = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Shared Cache Scheme (SCS)

- N files, K users and Λ helper nodes.



- Each user has access to a helper cache.
- Helper cache accessed by multiple users.
- Shared Cache Scheme involves three steps:
  - SCS Placement: Uncoded placement of files in helper cache nodes.
  - User-to-cache association: Each user assigned to exactly one helper node. Each cache is assigned to a set of users $\mathcal{U}_\lambda$ for $\lambda = 1, 2, \ldots, \Lambda$.
  - SCS Delivery: Once the demands are revealed, server transmits.

E. Parrinello, A. Unsal, P. Elia, " Fundamental Limits of Caching in Heterogeneous Networks with Uncoded Prefetching," Available on arXiv:1811.06247 [cs.IT], Nov. 2018.

Error Correction for Shared Cache Scheme

- We proved that $\alpha(\mathcal{M}_{\text{SC}}) = \kappa(\mathcal{M}_{\text{SC}}) = \sum_{i=1}^{\Lambda-\Lambda\gamma} \mathcal{L}_i \binom{\Lambda-i}{\Lambda\gamma}$.
- Hence the optimal error correcting delivery is by concatenation.

N. S. Karat, S. Dey, A. Thomas and B. S. Rajan, "An Optimal Linear Error Correcting Delivery Scheme for Coded Caching with Shared Caches," *2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, 2019.

Centralized and Decentralized Coded Caching

- Centralized coded caching [1]
  - Prefetching is carried out with a central coordination.
  - Number of users (K) is a parameter during placement.
- Decentralized coded caching [2]
  - No central coordination during placement.
  - Number of users (K) is not considered as a parameter during placement.
  - Rate expression [2]:

$$R^*_{\text{peak}} = K\left(1-\frac{M}{N}\right)\left\{\frac{N}{KM}\left(1-\left(1-\frac{M}{N}\right)^K\right)\right\}.$$

[1] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[2] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," IEEE/ACM Trans. Networking, vol. 23, no. 4, pp. 1029-1040, Aug. 2015.

Ali-Niesen Decentralized Coded Caching

- In centralized coded caching scheme, there may be considerable time gap between placement and delivery phase because of which number of users may vary.
- So, the number of users may not be known in the placement phase.
- No centralized coordination is assumed during placement.
- Placement phase : Each user independently caches a subset of $\frac{MF}{N}$ bits of each file, chosen uniformly at random.

M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," in *IEEE/ACM Transactions on Networking*, vol. 23, no. 4, pp. 1029–1040, Aug. 2015.

Error correcting delivery for Ali-Niesen decentralized scheme

- From the Ali-Niesen decentralized transmissions,

$$\kappa(\mathcal{M}_D, \mathbf{d}) \leq (1 - M/N).\frac{N}{M}(1 - (1 - M/N)^{N_e(\mathbf{d})})F$$

Theorem

For the index coding problem $\mathcal{I}(\mathcal{M}_D, \mathbf{d})$,

$$\alpha(\mathcal{M}_D, \mathbf{d}) \geq (1 - M/N).\frac{N}{M}(1 - (1 - M/N)^{N_e(\mathbf{d})})F.$$

- So, $\alpha(\mathcal{M}_D, \mathbf{d}) = \kappa(\mathcal{M}_D, \mathbf{d})$. Thus concatenation is optimal.

N. S. Karat and B. S. Rajan, "On the Optimality of Ali-Niesen Decentralized Coded Caching Scheme With and Without Error Correction," available on arXiv:1903.02408v1 [cs.IT], 5 Mar 2019.

LRS Online Coded Caching (Least Recently Sent)

- There are multiple time slots.
- A set of popular files for every time slot, denoted by $\mathcal{N}_t$.
- $\mathcal{N}_t$ evolves from $\mathcal{N}_{t-1}$ with at most one change.
- In each time slot, there is delivery phase and a cache update phase.
- Two phases:
  - Delivery phase : Same as Ali-Niesen Decentralized scheme.
  - Cache update phase:
    - $N = |\mathcal{N}_t|$, cardinality of the set of popular files.
    - $Q = \beta N$ files are partially cached for some $\beta \geq 1$
    - If all the demanded files are partially cached: No cache update.
    - If one demanded file is not partially cached: Cache is updated by replacing the new file with the least recently sent file.

R. Pedarsani, M. A. Maddah-Ali, and U. Niesen, "Online coded caching," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 836–845, Apr. 2016.

LRS Online coded caching (Cache update phase)

- There are two cases of situations that can arise during the cache update phase.
  - *Case I: All the requested files are partially cached*. Same as Ali-Niesen Decentralized scheme. The prefetching scheme here is denoted as $\mathcal{M}_{\text{LRS1}}$. Rate, $R(\mathcal{M}_{\text{LRS1}}, 0) = (1 - M/Q)(Q/M)\big(1 - (1 - M/Q)^K\big)$.
  - *Case II: One of the requested files is not partially cached.* $(K - 1)$ users demand files which are partially cached, and one user demands a file which is not cached. $K - 1$ users use Ali-Niesen decentralized scheme and the demanded file of one user is transmitted as such uncoded. Rate, $R(\mathcal{M}_{\text{LRS2}}, 0) = (1 - M/Q)(Q/M)\big(1 - (1 - M/Q)^{K-1}\big) + 1$.
- For both the cases of LRS scheme, we proved $\alpha(\mathcal{M}_{LRS1}, \mathbf{d}) = R(\mathcal{M}_{\text{LRS1}}, 0)F$ and $\alpha(\mathcal{M}_{LRS2}, \mathbf{d}) = R(\mathcal{M}_{\text{LRS2}}, 0)F$
- Since for all cases, $\alpha(\mathcal{M}_{LRS}, \mathbf{d}) = \kappa(\mathcal{M}_{LRS}, \mathbf{d})$, the concatenation scheme is optimal.

Our Publications

- N. S. Karat, S. Samuel and B. S. Rajan, "Optimal Error Correcting Index Codes for Some Generalized Index Coding Problems," in *IEEE TCOM*, vol. 67, no. 2, pp. 929-942, Feb. 2019.
- N. S. Karat, A. Thomas and B. S. Rajan, "Error Correction in Coded Caching With Symmetric Batch Prefetching," in *IEEE TCOM*, vol. 67, no. 8, pp. 5264-5274, Aug. 2019.
- N. S. Karat, A. Thomas and B. S. Rajan, "Optimal Error Correcting Delivery Scheme for an Optimal Coded Caching Scheme with Small Buffers," *ISIT 2018*, Vail, CO, 2018, pp. 1710-1714.
- N. S. Karat, A. Thomas and B. S. Rajan, "Optimal Error Correcting Delivery Scheme for Coded Caching with Symmetric Batch Prefetching," *ISIT 2018*, Vail, CO, 2018, pp. 2092-2096.
- N. S. Karat, S. Dey, A. Thomas and B. S. Rajan, "An Optimal Linear Error Correcting Delivery Scheme for Coded Caching with Shared Caches," *ISIT 2019*, Paris, France, 2019, pp. 1217-1221.
- N. S. Karat and B. S. Rajan, "On the Optimality of Ali-Niesen Decentralized Coded Caching Scheme With and Without Error Correction," available on arXiv:1903.02408v1 [cs.IT], 5 Mar 2019.

Thank you