Coded Caching: Discussing Promises and Challenges

Eleftherios Lampiris
lampiris@eurecom.fr

EURECOM

September 2017

Outlook

Promises

- Coded Caching
- MIMO Coded Caching
- D2D Coded Caching
- Decentralised Approach
- Caching under Popularity Distributions

Outlook

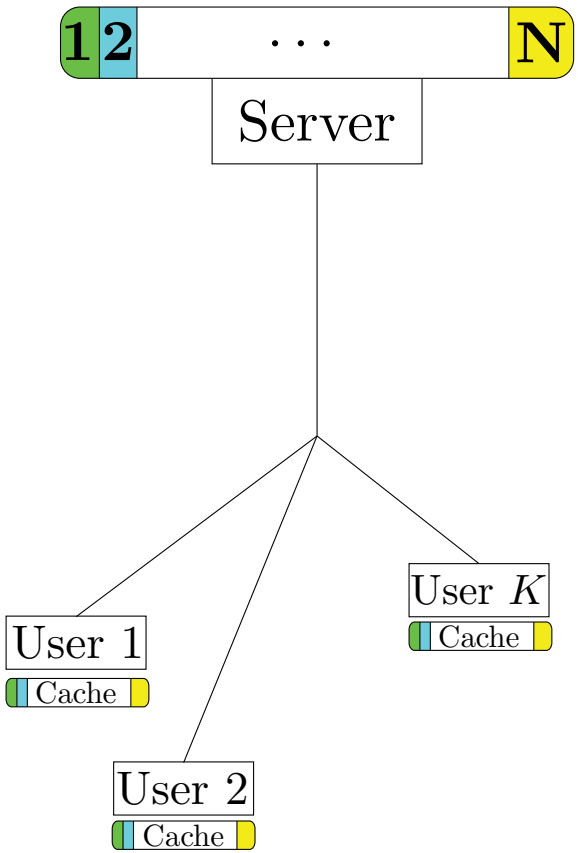
Promises

- Coded Caching
- MIMO Coded Caching
- D2D Coded Caching
- Decentralised Approach
- Caching under Popularity Distributions

Challenges

- Uneven Channel Strengths
- CSI Requirements
- Subpacketization issues

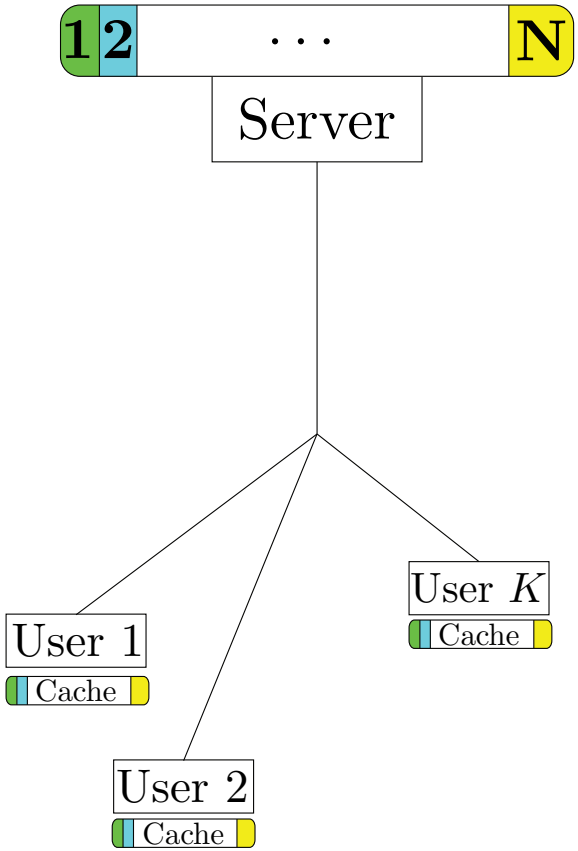
Coded Caching



[Maddah-Ali, Niesen '12]

- N Files
- F bits per File
- K Users
- $M \cdot F$ User Memory
- $\gamma = \dfrac{M}{N}$ Normalized Cache

Coded Caching



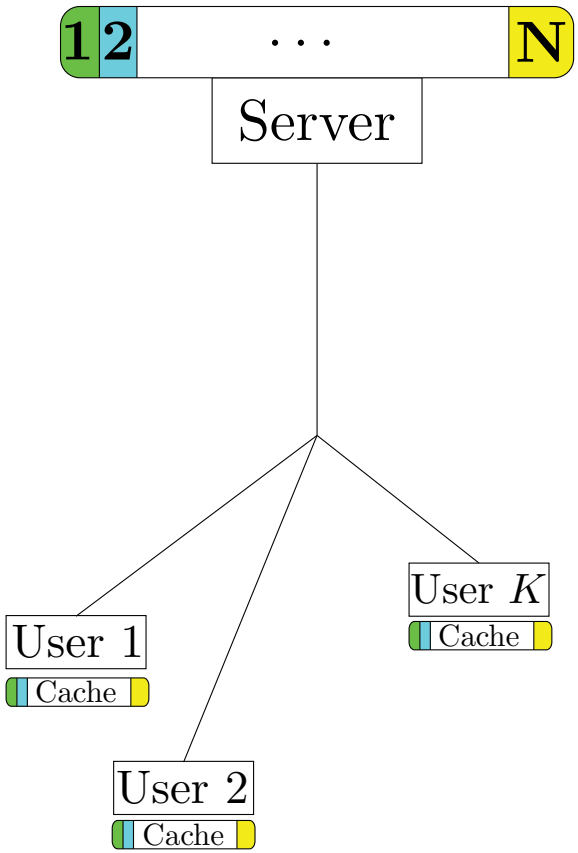
[Maddah-Ali, Niesen '12]

- N Files
- F bits per File
- K Users
- $M \cdot F$ User Memory
- $\gamma = \dfrac{M}{N}$ Normalized Cache

Users Served Simultaneously

$$d_\Sigma = K\gamma + 1$$

Coded Caching



[Maddah-Ali, Niesen '12]

- N Files
- F bits per File
- K Users
- $M \cdot F$ User Memory
- $\gamma = \dfrac{M}{N}$ Normalized Cache

Users Served Simultaneously

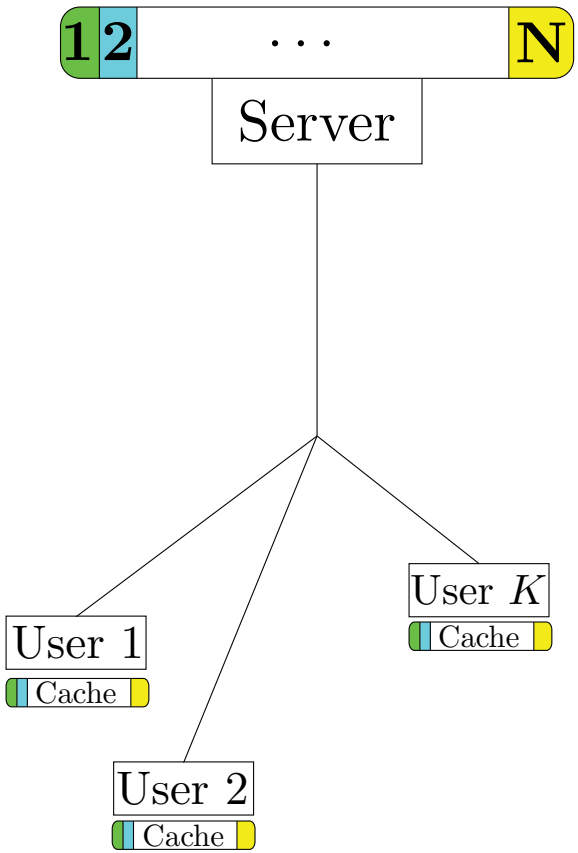
$$d_{\Sigma} = K\gamma + 1$$

Optimality

Optimal within a multiplicative factor of 2.008.

[Qian, Maddah-Ali, Avestimehr '17]

Effect of Coded Caching

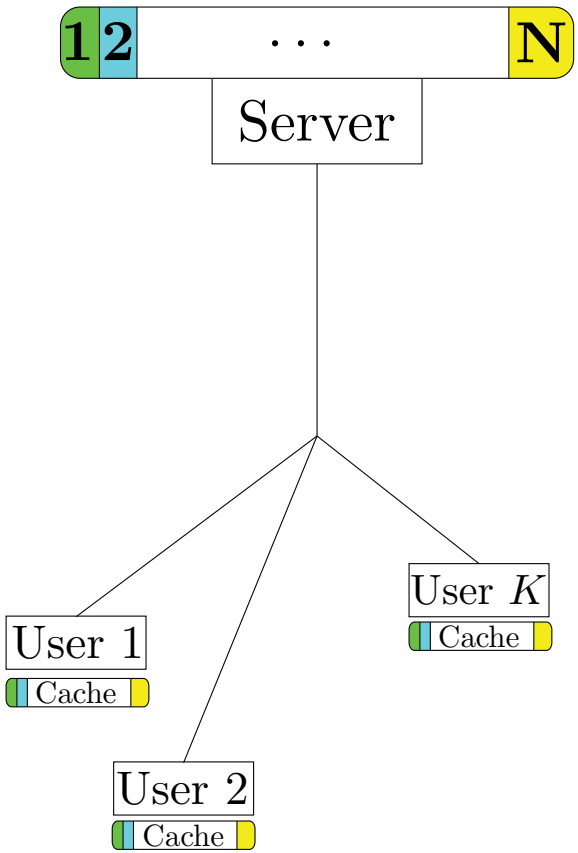


[Maddah-Ali, Niesen '12]

Time under Uniform Demand

$$T = K(1-\gamma)\frac{1}{K\gamma+1} \approx \frac{1-\gamma}{\gamma}$$

Effect of Coded Caching



[Maddah-Ali, Niesen '12]

Time under Uniform Demand

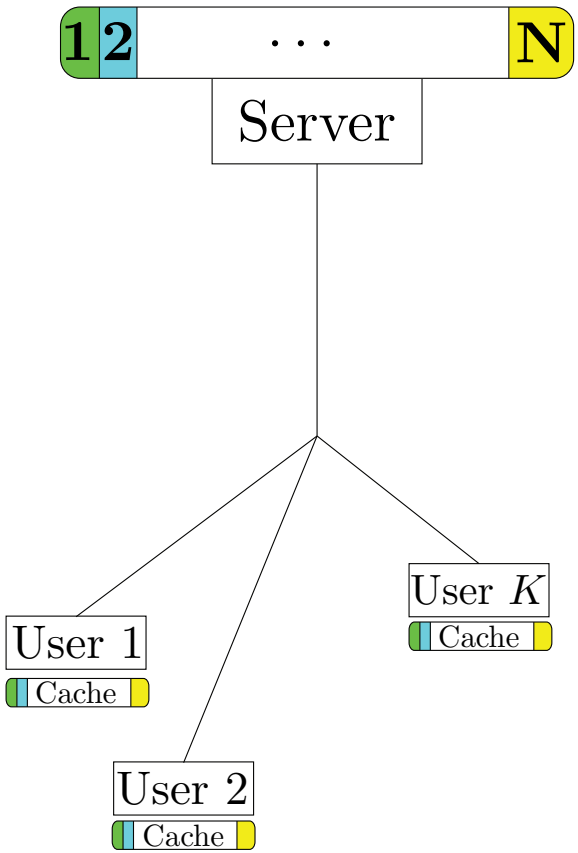
$$T = K(1-\gamma)\frac{1}{K\gamma+1} \approx \frac{1-\gamma}{\gamma}$$

Per User Rate

$$R_u = \left(\gamma + \frac{1}{K}\right) \cdot R_{\text{tot}} \approx \gamma \cdot R_{\text{tot}}$$

Each gets a γ piece of the pie!!

Effect of Coded Caching



[Maddah-Ali, Niesen '12]

Time under Uniform Demand

$$T = K(1-\gamma)\frac{1}{K\gamma+1} \approx \frac{1-\gamma}{\gamma}$$

Per User Rate

$$R_u = \left(\gamma + \frac{1}{K}\right) \cdot R_{\text{tot}} \approx \gamma \cdot R_{\text{tot}}$$

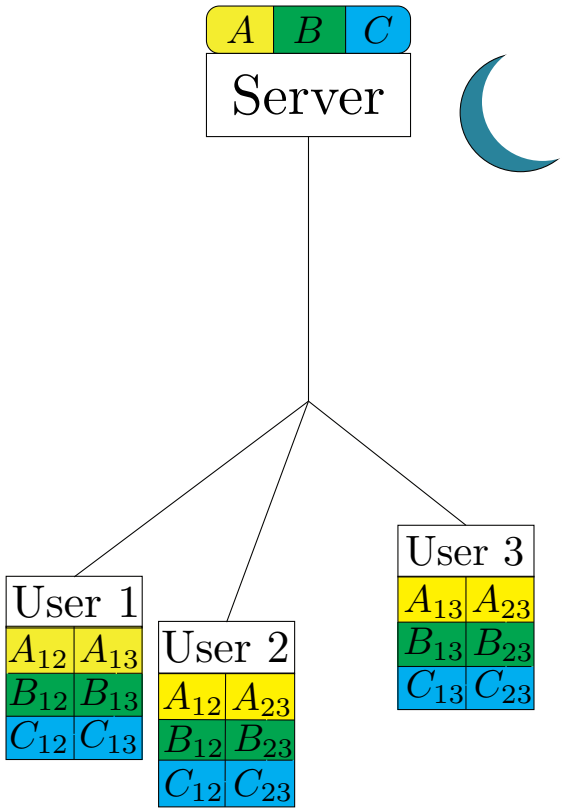
Each gets a γ piece of the pie!!

Intuition

Even unwanted packets can help reduce interference

Example 1

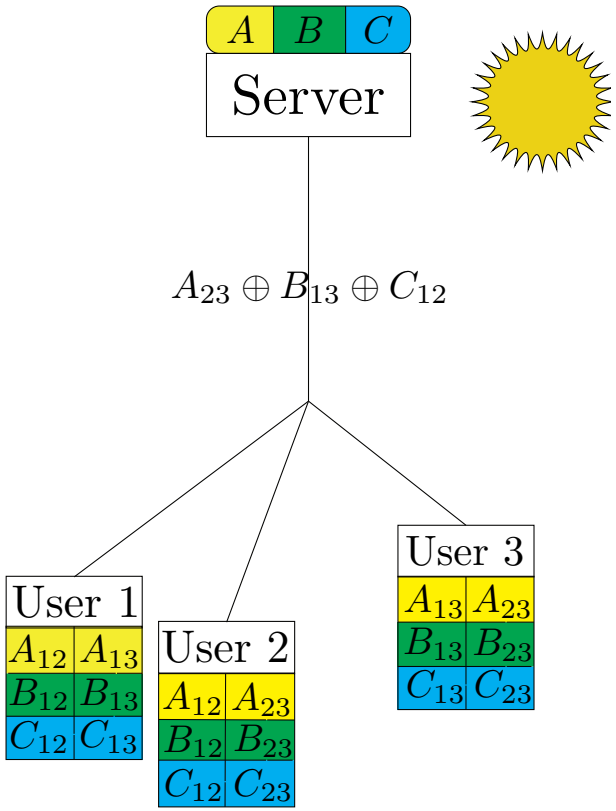
3 Users - 3 Files - $\gamma = \frac{2}{3}$



- $M = 2$
- Files are divided into 3 parts $\{12, 13, 23\}$
- Cache i is filled-up with all parts indexed with i

Example 1

3 Users - 3 Files - $\gamma = \frac{2}{3}$

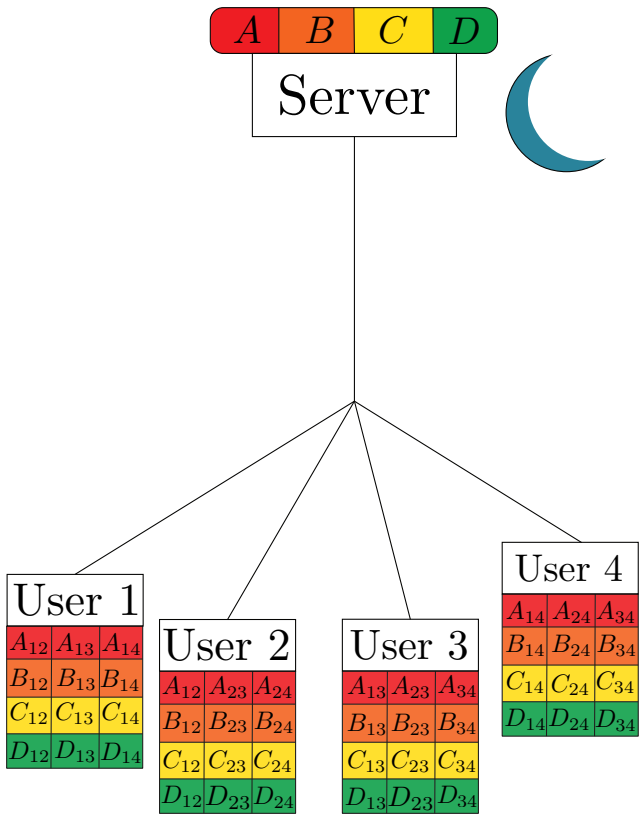


$A_{23} \oplus B_{13} \oplus C_{12}$

Single Message serves all users at the same time

Example 2

4 Users - 4 Files - $\gamma = \frac{2}{4}$

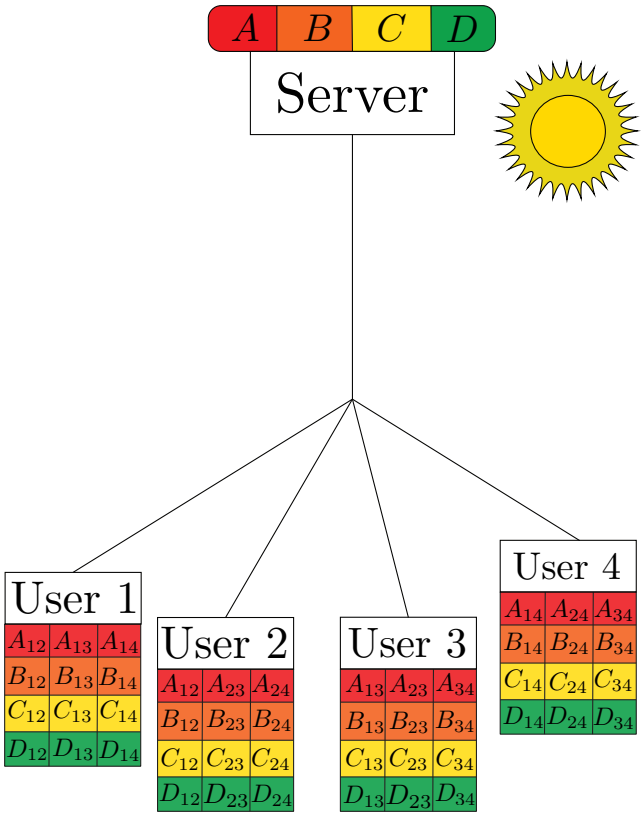


Placement Phase

- Files are divided into 6 parts $\{12, 13, 14, 23, 24, 34\}$
- Cache i is filled-up with all parts indexed with i

Example 2

4 Users - 4 Files - $\gamma = \frac{2}{4}$

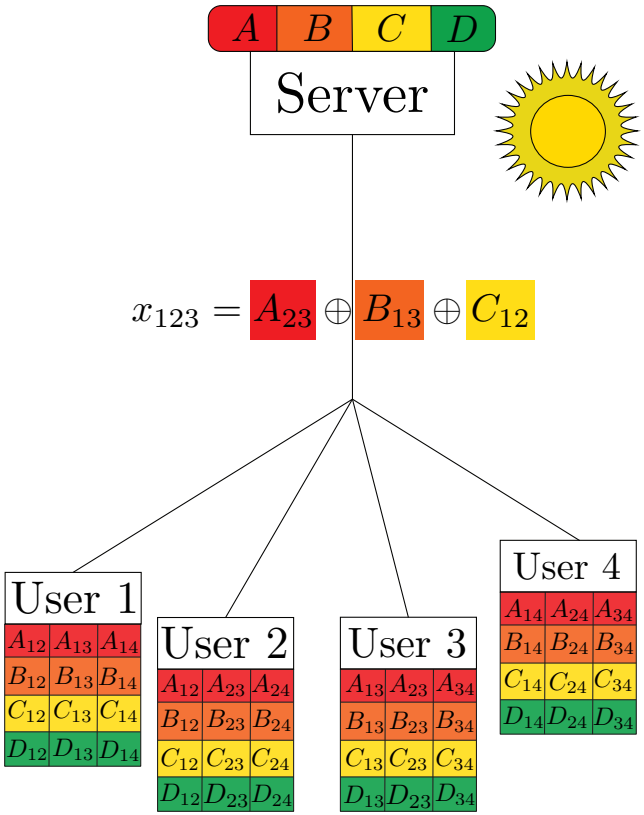


Delivery Phase

Every Message serves 3 users at the same time

Example 2

4 Users - 4 Files - $\gamma = \frac{2}{4}$



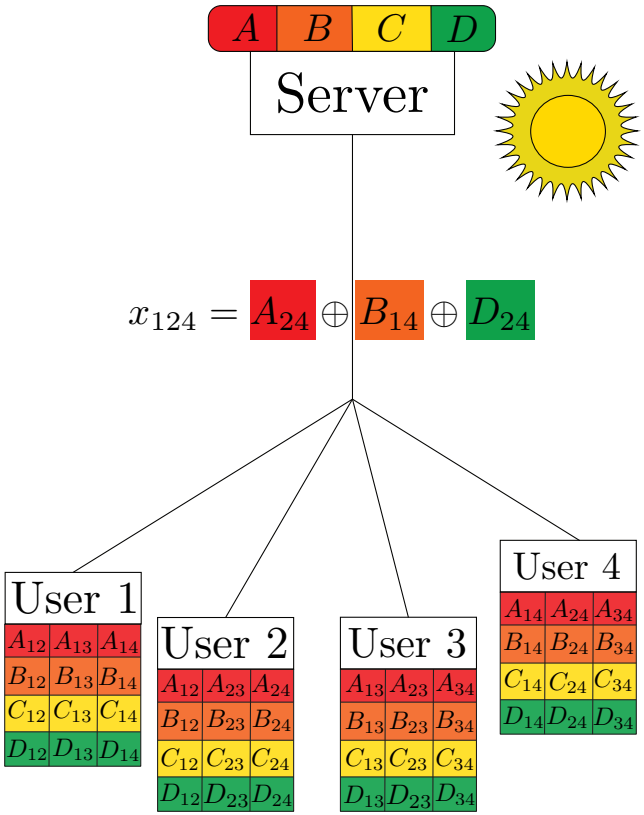
Delivery Phase

Every Message serves 3 users at the same time

- $x_{123} = A_{23} \oplus B_{13} \oplus C_{12}$

Example 2

4 Users - 4 Files - $\gamma = \frac{2}{4}$



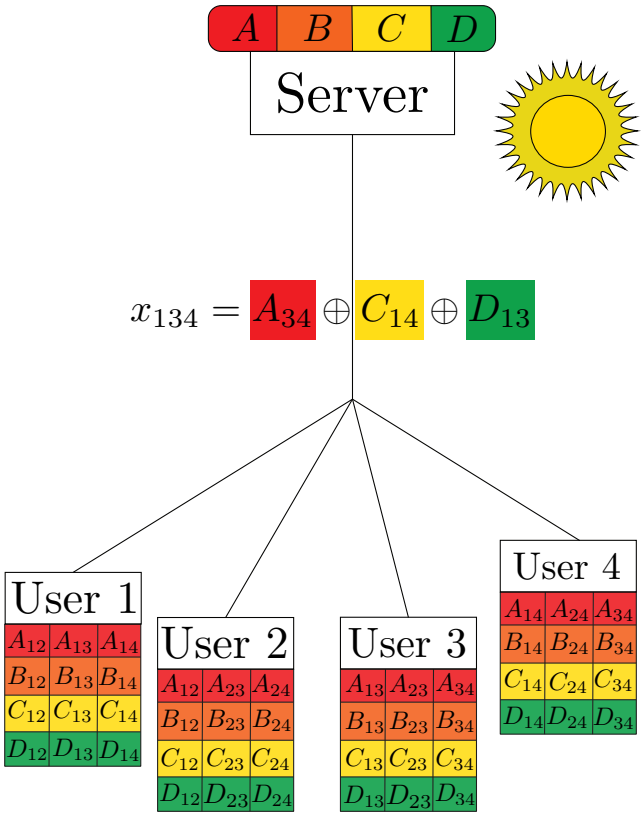
Delivery Phase

Every Message serves 3 users at the same time

- $x_{123} = A_{23} \oplus B_{13} \oplus C_{12}$
- $x_{124} = A_{24} \oplus B_{14} \oplus D_{24}$

Example 2

4 Users - 4 Files - $\gamma = \frac{2}{4}$



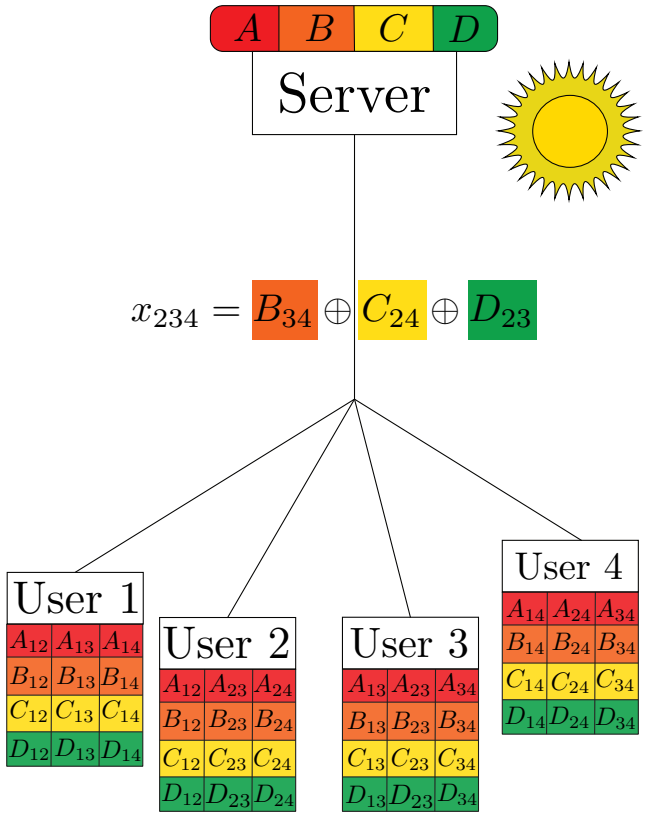
Delivery Phase

Every Message serves 3 users at the same time

- $x_{123} = A_{23} \oplus B_{13} \oplus C_{12}$
- $x_{124} = A_{24} \oplus B_{14} \oplus D_{24}$
- $x_{134} = A_{34} \oplus C_{14} \oplus D_{13}$

Example 2

4 Users - 4 Files - $\gamma = \frac{2}{4}$

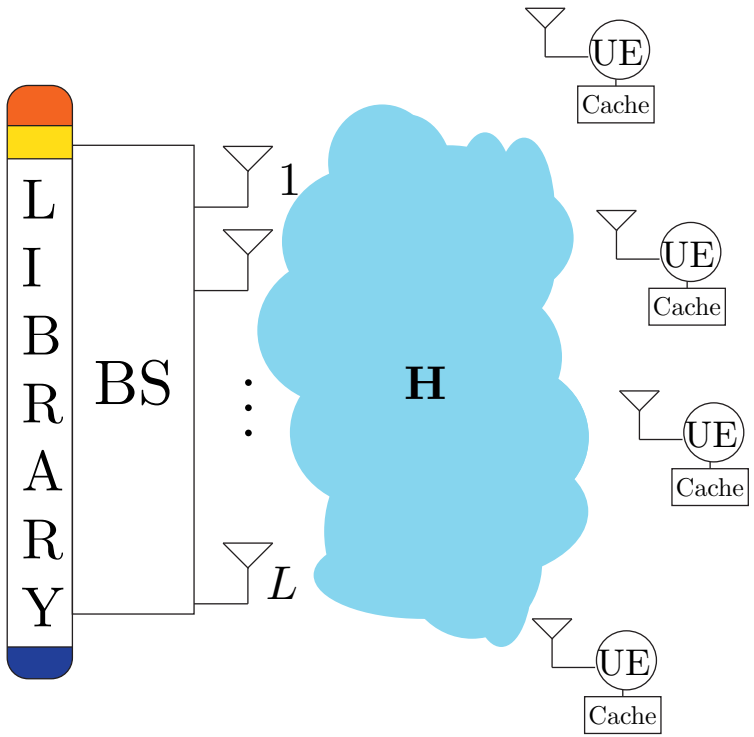


Delivery Phase

Every Message serves 3 users at the same time

- $x_{123} = A_{23} \oplus B_{13} \oplus C_{12}$
- $x_{124} = A_{24} \oplus B_{14} \oplus D_{24}$
- $x_{134} = A_{34} \oplus C_{14} \oplus D_{13}$
- $x_{234} = B_{34} \oplus C_{24} \oplus D_{23}$

MIMO Coded Caching



[Shariatpanahi, Motahari, Khalaj '15 &
Naderializadeh, Maddah-Ali, Avestimehr '16]

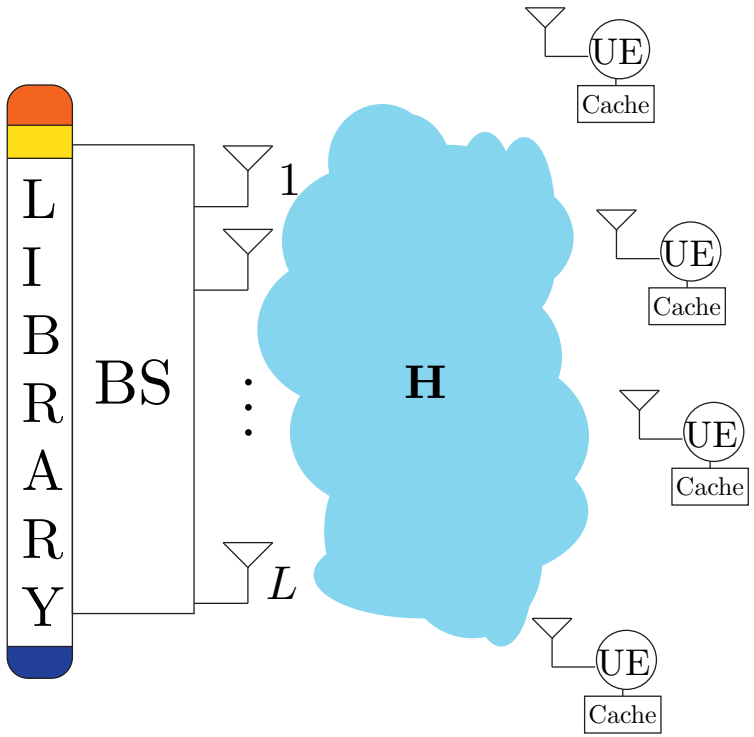
- N files
- K users
- L antennas
- γ fractional cache

Users Served

$$d_\Sigma = K\gamma + L \ ^\star$$

$^\star d_\Sigma = \min\{K, K\gamma + L\}$
Order Optimal Gap$= 2$

MIMO Coded Caching



[Shariatpanahi, Motahari, Khalaj '15 & Naderializadeh, Maddah-Ali, Avestimehr '16]

- N files
- K users
- L antennas
- γ fractional cache

Users Served

$$d_\Sigma = K\gamma + L \ ^\star$$

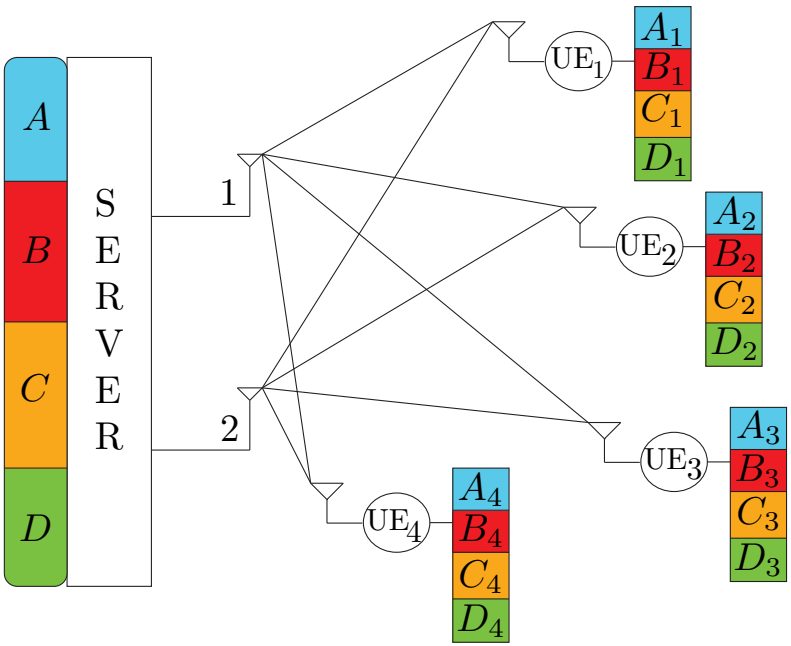
Intuition

Send a vector of linear Combinations such that some packets are cached-out and some packets are nulled-out.

$^\star d_\Sigma = \min\{K, K\gamma + L\}$
Order Optimal Gap$= 2$

MIMO CC Example

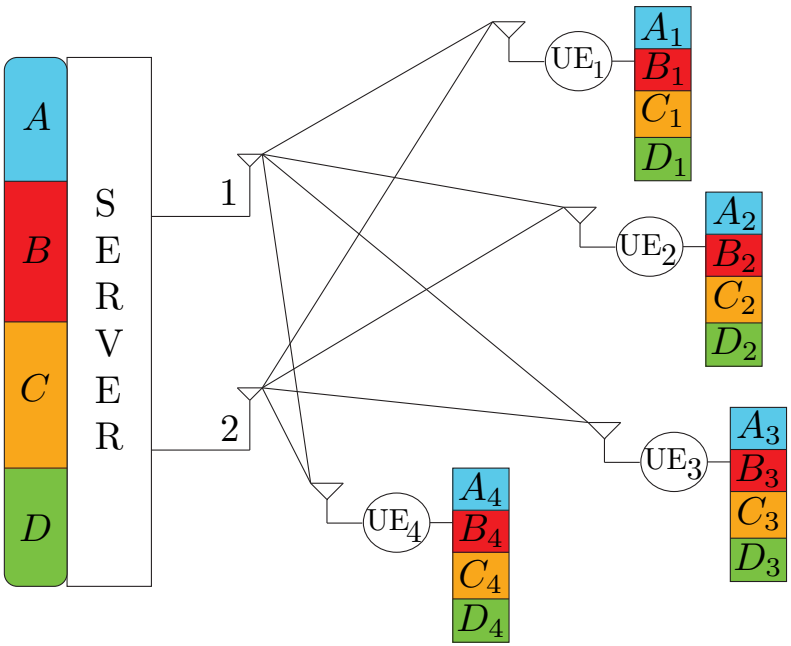
$K = 4,\ N = 4,\ L = 2,\ \gamma = \frac{1}{4}$



h_{ij} : Antenna i to User j Channel

MIMO CC Example

$K = 4,\ N = 4,\ L = 2,\ \gamma = \frac{1}{4}$

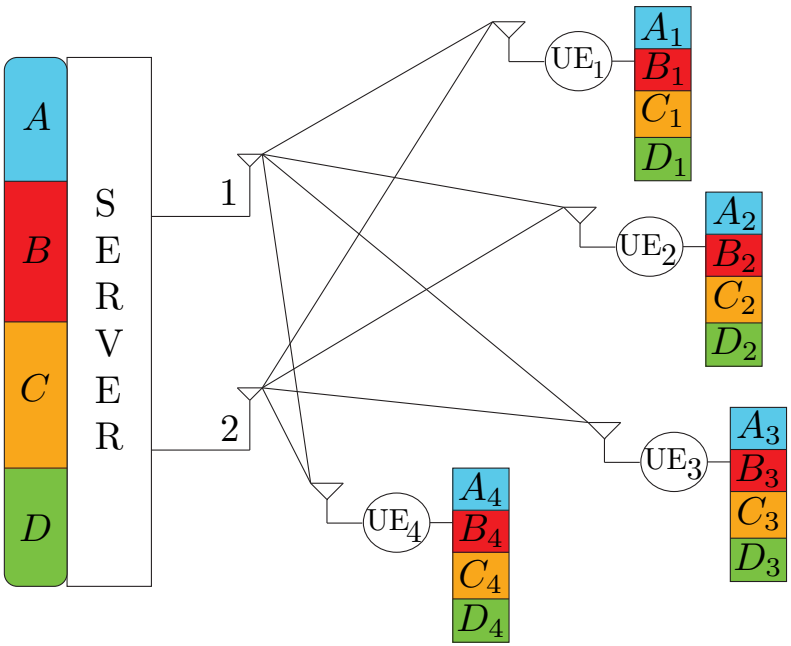


$$\underline{x}_{123} = \begin{bmatrix} h_{13}^{-1} A_2 + h_{11}^{-1} B_3 + h_{12}^{-1} C_1 \\ -h_{23}^{-1} A_2 - h_{21}^{-1} B_3 - h_{22}^{-1} C_1 \end{bmatrix}$$

h_{ij} : Antenna i to User j Channel

MIMO CC Example

$K = 4,\ N = 4,\ L = 2,\ \gamma = \frac{1}{4}$



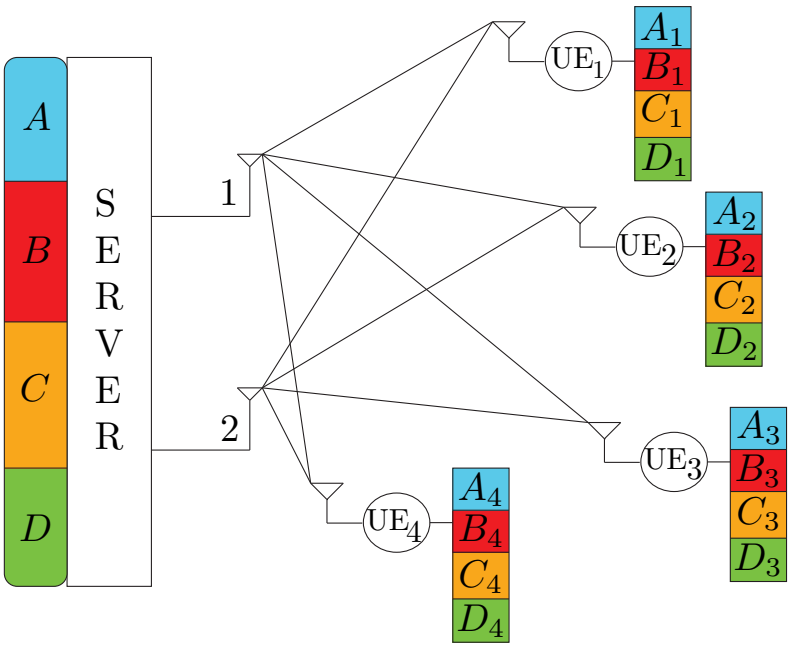
$$\underline{x}_{123} = \begin{bmatrix} h_{13}^{-1}A_2 + h_{11}^{-1}B_3 + h_{12}^{-1}C_1 \\ -h_{23}^{-1}A_2 - h_{21}^{-1}B_3 - h_{22}^{-1}C_1 \end{bmatrix}$$

$$\begin{aligned} y_1 &= A_2 + C_1(h_{11}h_{12}^{-1} - h_{21}h_{22}^{-1}) \\ y_2 &= B_3 + A_2(h_{12}h_{13}^{-1} - h_{22}h_{23}^{-1}) \\ y_3 &= C_1 + B_3(h_{13}h_{11}^{-1} - h_{23}h_{21}^{-1}) \end{aligned}$$

h_{ij} : Antenna i to User j Channel

MIMO CC Example

$K = 4,\ N = 4,\ L = 2,\ \gamma = \frac{1}{4}$



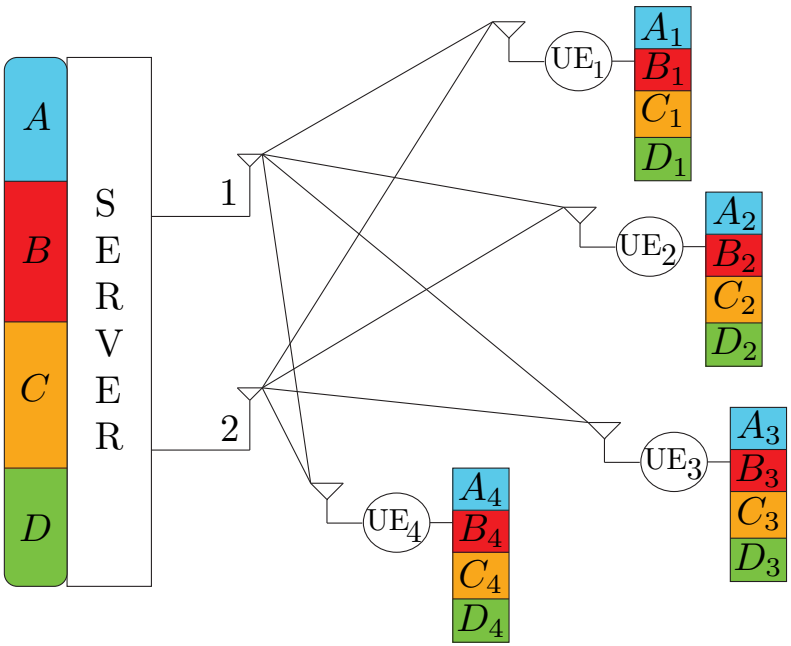
$$\underline{x}_{123} = \begin{bmatrix} h_{13}^{-1} A_2 + h_{11}^{-1} B_3 + h_{12}^{-1} C_1 \\ -h_{23}^{-1} A_2 - h_{21}^{-1} B_3 - h_{22}^{-1} C_1 \end{bmatrix}$$

$$\underline{x}_{124} = \begin{bmatrix} h_{12}^{-1} A_4 + h_{14}^{-1} B_1 + h_{11}^{-1} D_2 \\ -h_{22}^{-1} A_4 - h_{24}^{-1} B_1 - h_{21}^{-1} D_2 \end{bmatrix}$$

h_{ij} : Antenna i to User j Channel

MIMO CC Example

$K = 4,\ N = 4,\ L = 2,\ \gamma = \frac{1}{4}$



h_{ij} : Antenna i to User j Channel

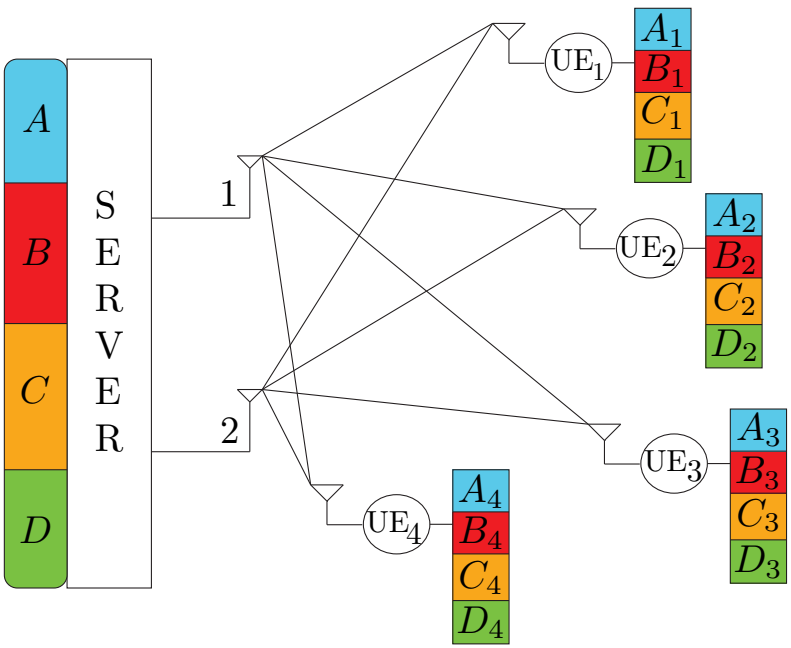
$$\underline{x}_{123} = \begin{bmatrix} h_{13}^{-1}A_2 + h_{11}^{-1}B_3 + h_{12}^{-1}C_1 \\ -h_{23}^{-1}A_2 - h_{21}^{-1}B_3 - h_{22}^{-1}C_1 \end{bmatrix}$$

$$\underline{x}_{124} = \begin{bmatrix} h_{12}^{-1}A_4 + h_{14}^{-1}B_1 + h_{11}^{-1}D_2 \\ -h_{22}^{-1}A_4 - h_{24}^{-1}B_1 - h_{21}^{-1}D_2 \end{bmatrix}$$

$$\underline{x}_{134} = \begin{bmatrix} h_{14}^{-1}A_3 + h_{11}^{-1}C_4 + h_{13}^{-1}D_1 \\ -h_{24}^{-1}A_3 - h_{21}^{-1}C_4 - h_{23}^{-1}D_1 \end{bmatrix}$$

MIMO CC Example

$K = 4,\ N = 4,\ L = 2,\ \gamma = \frac{1}{4}$



h_{ij} : Antenna i to User j Channel

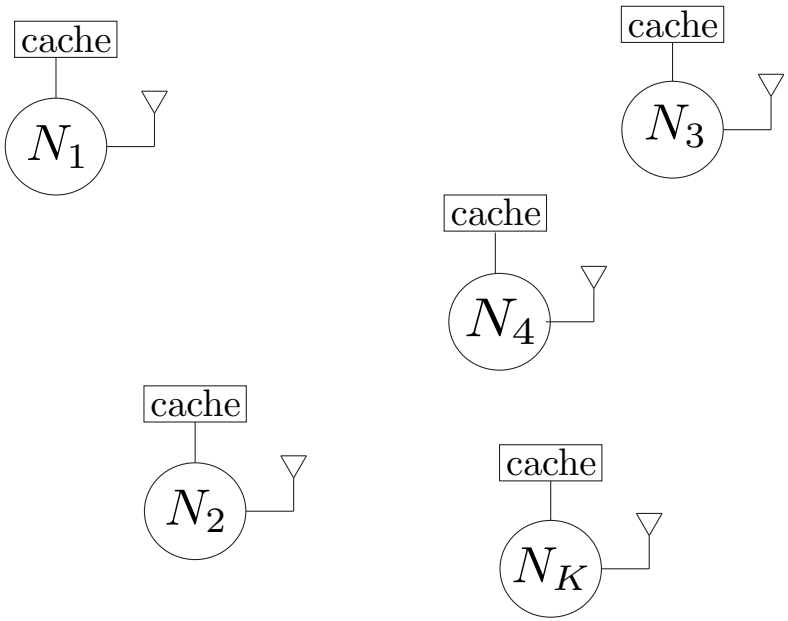
$$\underline{x}_{123} = \begin{bmatrix} h_{13}^{-1}A_2 + h_{11}^{-1}B_3 + h_{12}^{-1}C_1 \\ -h_{23}^{-1}A_2 - h_{21}^{-1}B_3 - h_{22}^{-1}C_1 \end{bmatrix}$$

$$\underline{x}_{124} = \begin{bmatrix} h_{12}^{-1}A_4 + h_{14}^{-1}B_1 + h_{11}^{-1}D_2 \\ -h_{22}^{-1}A_4 - h_{24}^{-1}B_1 - h_{21}^{-1}D_2 \end{bmatrix}$$

$$\underline{x}_{134} = \begin{bmatrix} h_{14}^{-1}A_3 + h_{11}^{-1}C_4 + h_{13}^{-1}D_1 \\ -h_{24}^{-1}A_3 - h_{21}^{-1}C_4 - h_{23}^{-1}D_1 \end{bmatrix}$$

$$\underline{x}_{234} = \begin{bmatrix} h_{13}^{-1}B_4 + h_{14}^{-1}C_2 + h_{12}^{-1}D_3 \\ -h_{23}^{-1}B_4 - h_{24}^{-1}C_2 - h_{22}^{-1}D_3 \end{bmatrix}$$

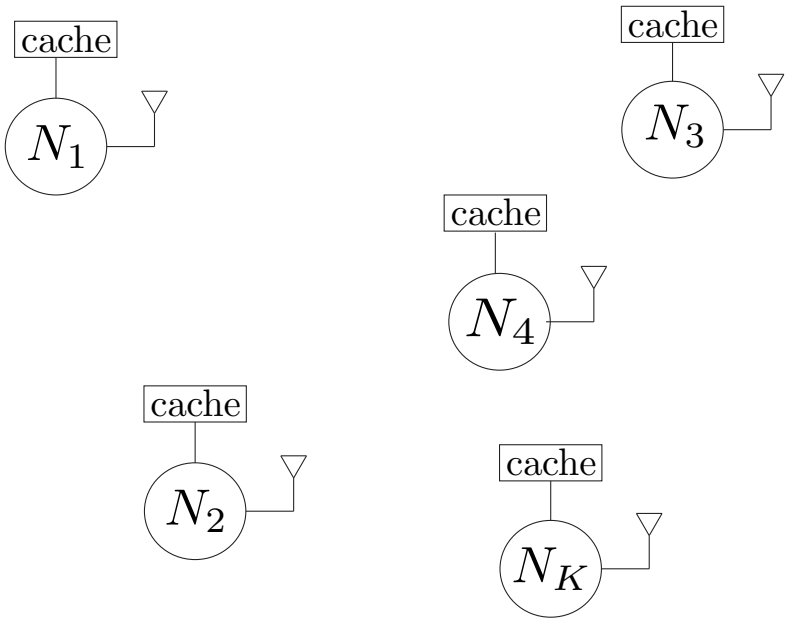
D2D Coded Caching



[Ji, Caire, Molisch '14]

- K users
- N files
- M files cache capacity

D2D Coded Caching



[Ji, Caire, Molisch '14]

- K users
- N files
- M files cache capacity

Intuition

Each user assumes the role of the transmitter as in original MN

D2D Coded Caching

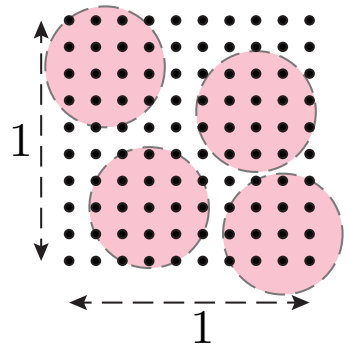
In a nutshell

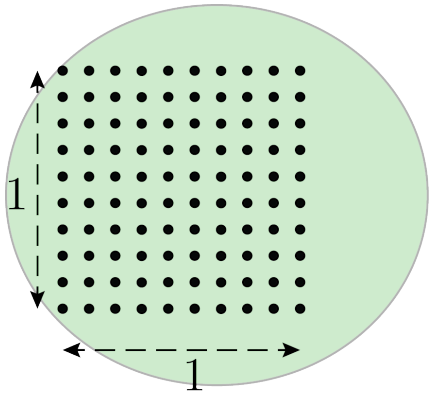
Results

- $K\gamma$ users served at a time
- $T = \dfrac{1-\gamma}{\gamma}$

Surprising Result

D2D Gains are the same with or without spatial reuse





Decentralized Coded Caching

- Number of users showing up is unknown
- Cache uniformly at random

Result

$$T = \frac{(1-\gamma)}{\gamma}(1-(1-\gamma)^K) \quad \star$$

$\star$ Order Optimal

[Maddah-Ali, Niesen '13]

Decentralized Coded Caching

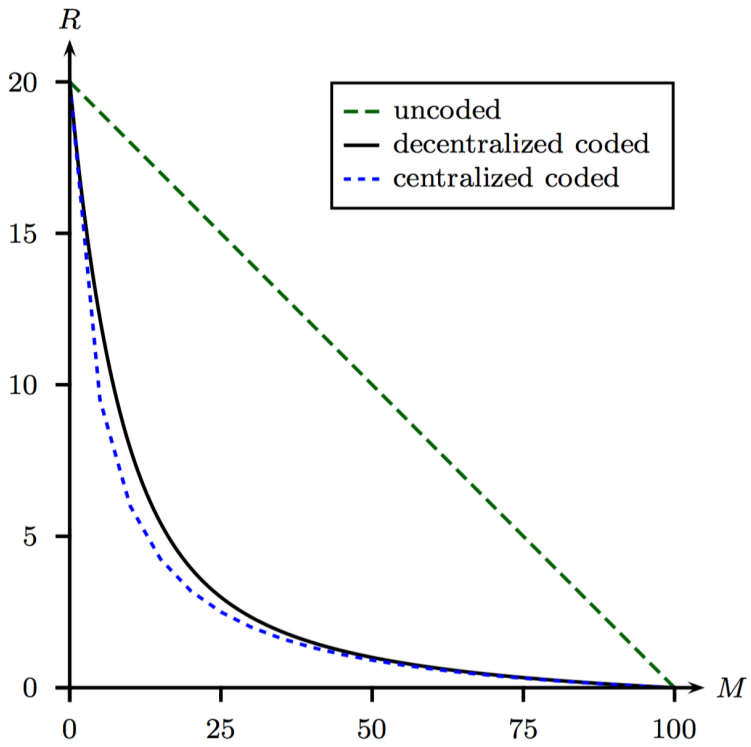
- Number of users showing up is unknown
- Cache uniformly at random

Result

$$T = \frac{(1-\gamma)}{\gamma}(1-(1-\gamma)^K) \quad \star$$

$\star$ Order Optimal

[Maddah-Ali, Niesen '13]

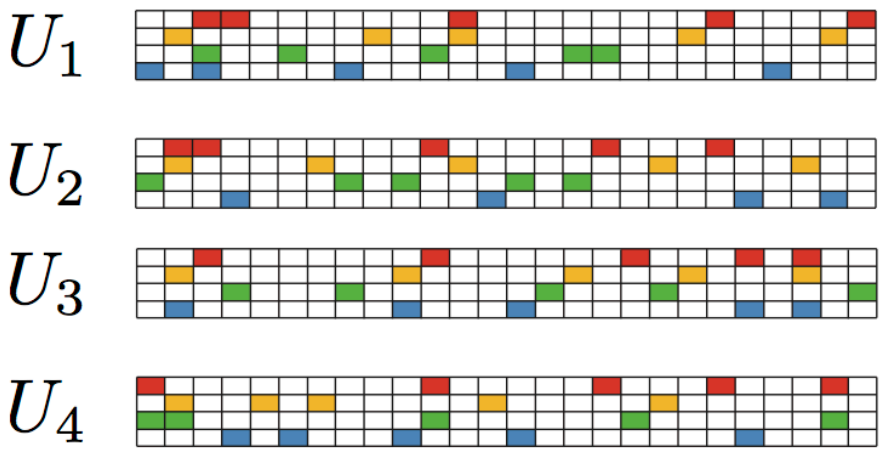


[source: Maddah-Ali, Niesen '13]

Decentralized Coded Caching

Example

Placement



Algorithm

- Start from higher order cliques and move to lower order
- Greedy approach

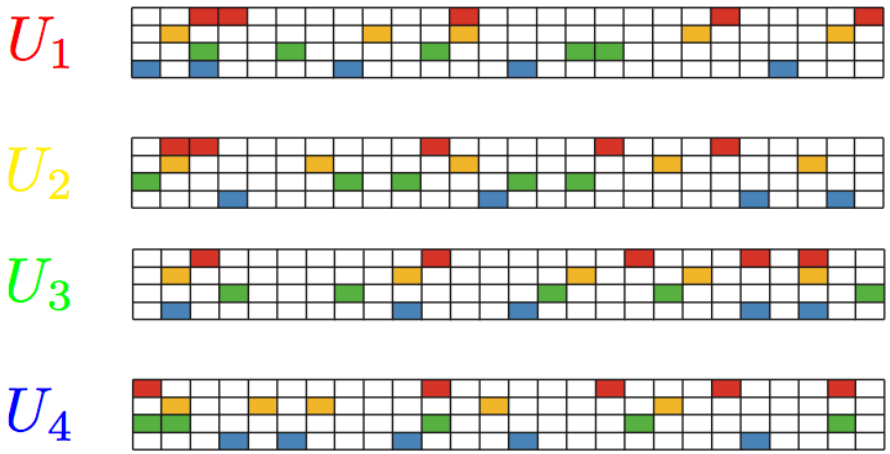
Decentralized Coded Caching

Example

Algorithm

- Start from higher order cliques and move to lower order
- Greedy approach

Requests



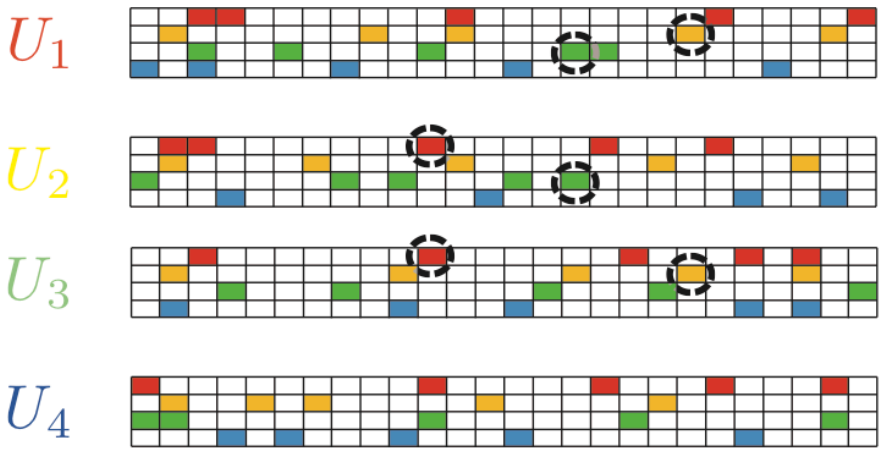
Decentralized Coded Caching

Example

Algorithm

- Start from higher order cliques and move to lower order
- Greedy approach

Clique of 3



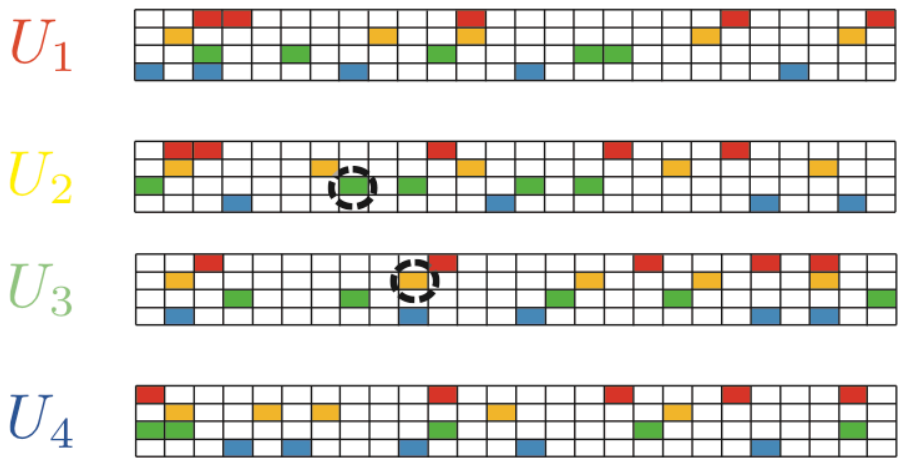
Decentralized Coded Caching

Example

Algorithm

- Start from higher order cliques and move to lower order
- Greedy approach

Clique of 2



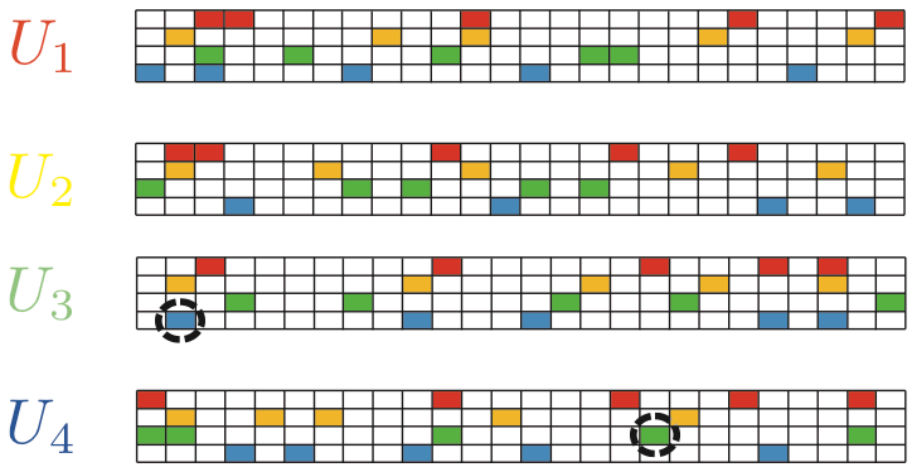
Decentralized Coded Caching

Example

Algorithm

- Start from higher order cliques and move to lower order
- Greedy approach

Clique of 2

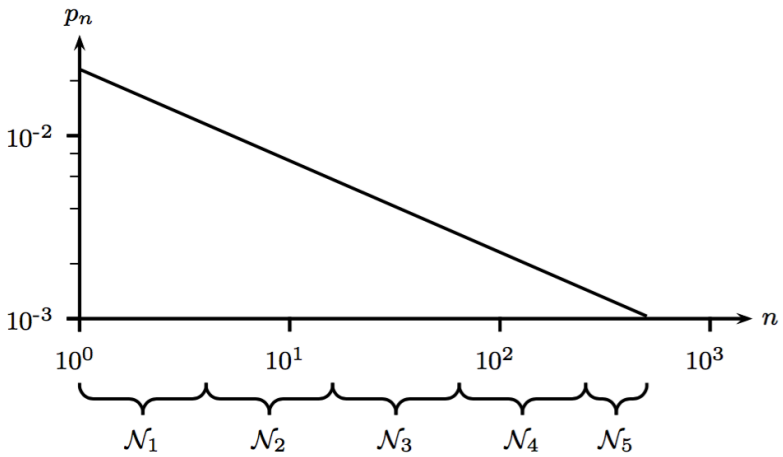


Requests under Popularity Distributions

Algorithm 1

- Proved Sub-optimality of HPF
- Group files according to popularity

[Maddah-Ali, Niesen '13]



[source: Maddah-Ali, Niesen '13]

Requests under Popularity Distributions

Algorithm 1

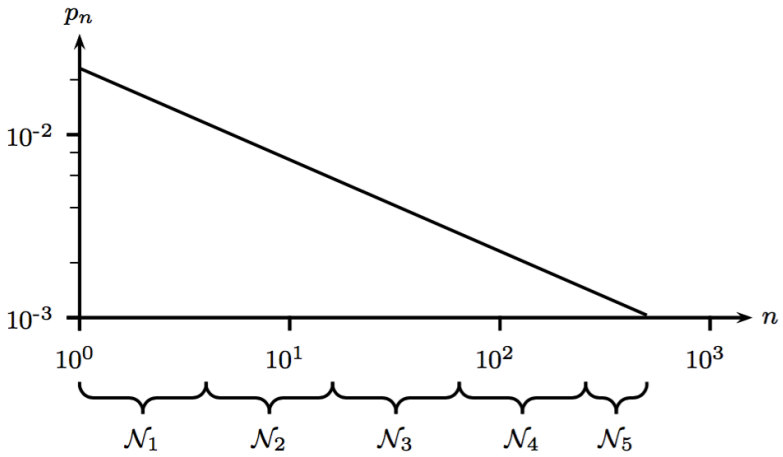
- Proved Sub-optimality of HPF
- Group files according to popularity

[Maddah-Ali, Niesen '13]

Algorithm 2

- Cache file if popularity $\geq \frac{1}{KM}$
- Increases γ by reducing N
- Order optimal with Gap $= 87$

[Zhang, Lin, Wang '15]



[source: Maddah-Ali, Niesen '13]

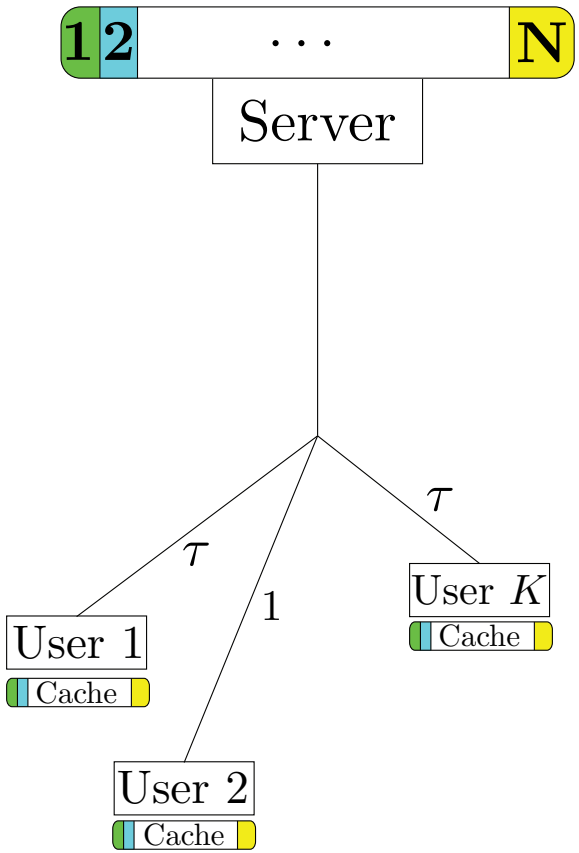
Requests under Popularity Distributions

Key Questions

- Algorithms to achieve near optimal solutions?
- How better can these algorithms perform?
- Does knowledge of the popularity significantly improve original MN?

Implementation Challenges

1) Uneven Channel Strengths

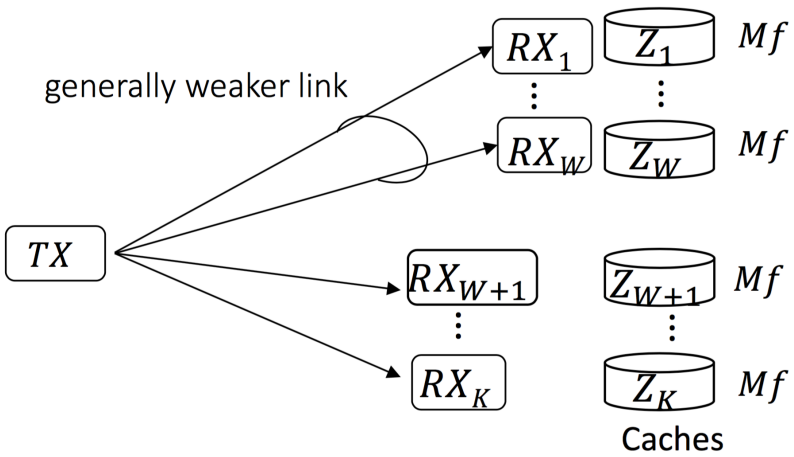


$\tau \leq 1$

- Worst-user effect
- Common reality in Wireless

Implementation Challenges

1) Uneven Channel Strengths: SISO with Topology

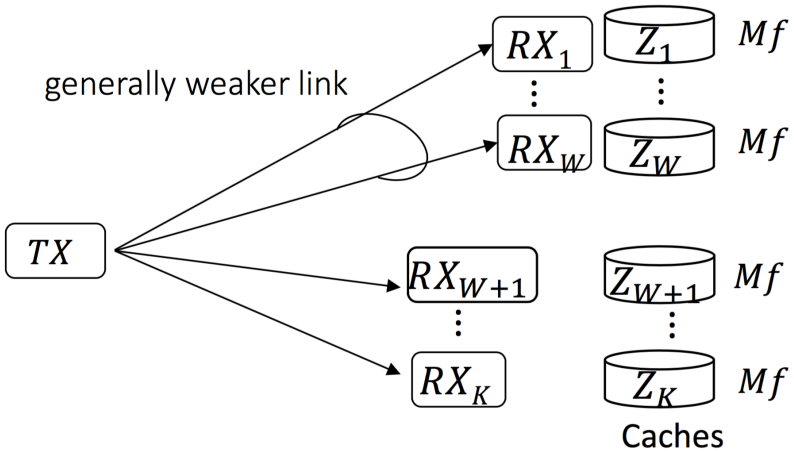


[Zhang, Elia '16
source: Zhang, Elia '16]

- W users have link capacity $\tau \leq 1$
- $K - W$ users have link capacity 1

Implementation Challenges

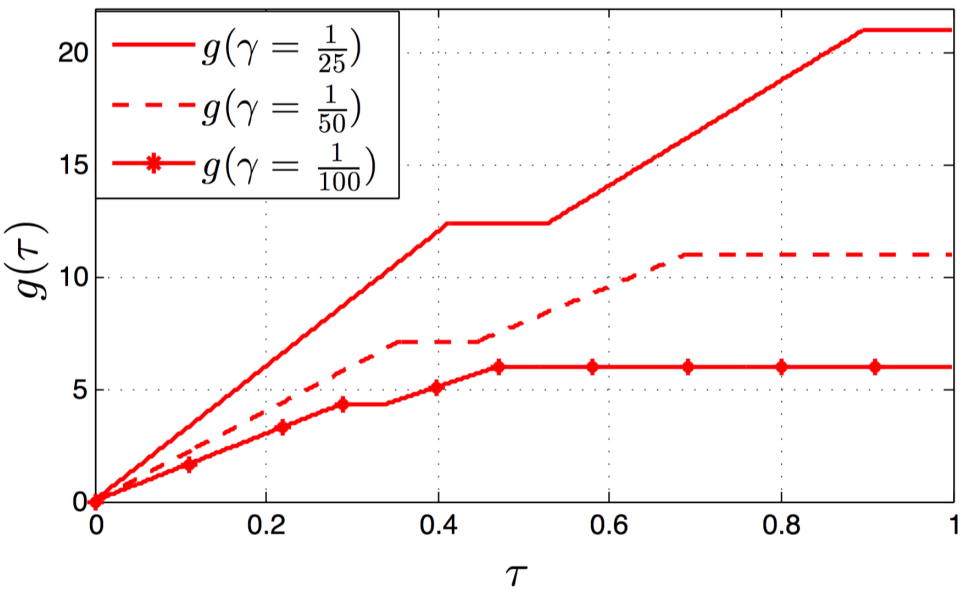
1) Uneven Channel Strengths: SISO with Topology



[Zhang, Elia '16
source: Zhang, Elia '16]

- W users have link capacity $\tau \leq 1$
- $K - W$ users have link capacity 1

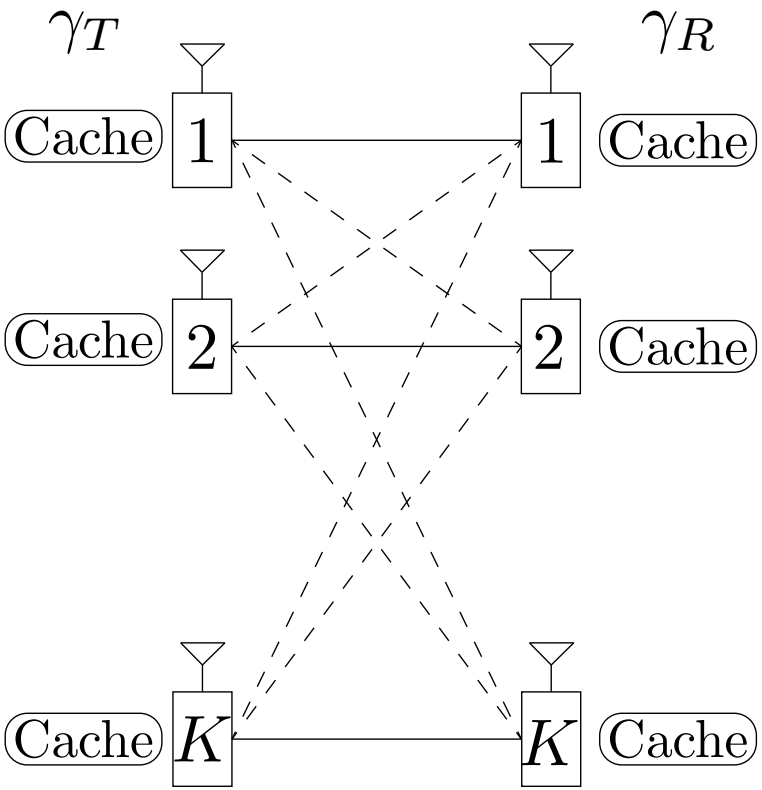
$$K = 500$$



[source: Zhang, Elia '16]

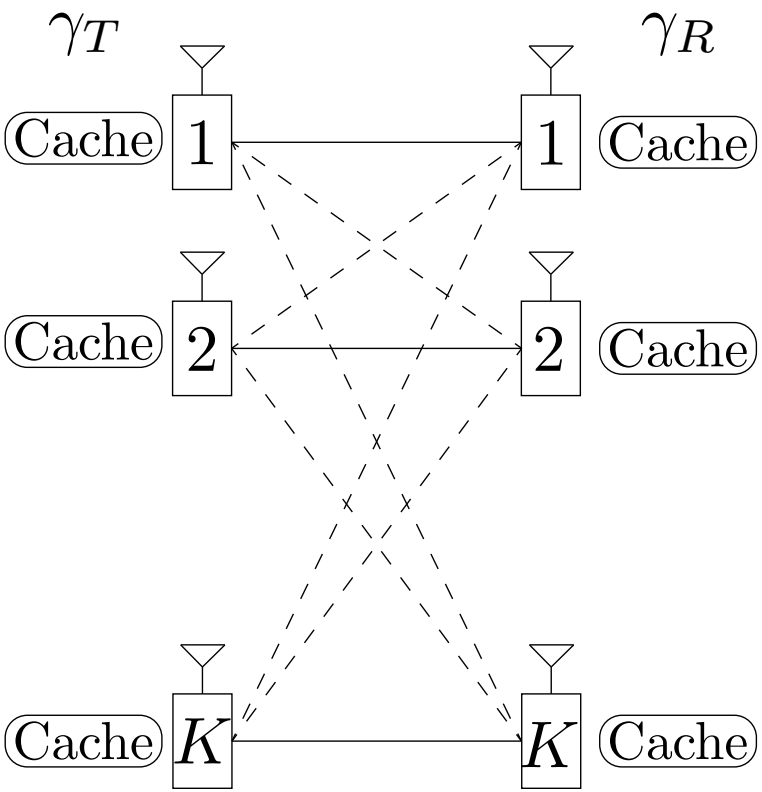
Implementation Challenges

1) Uneven Channel Strengths: Cooperation without CSIT



Implementation Challenges

1) Uneven Channel Strengths: Cooperation without CSIT



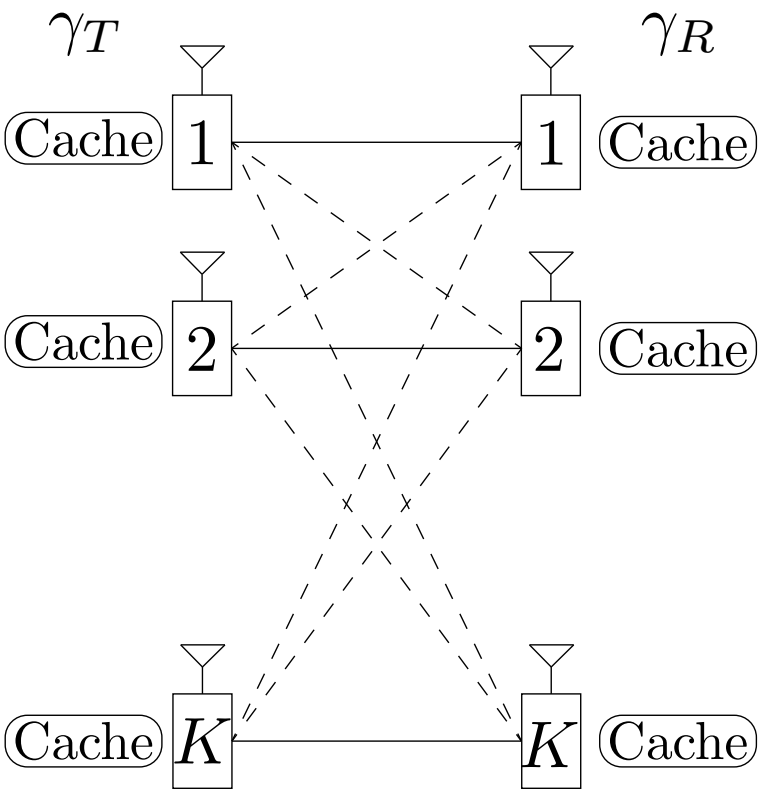
[Lampiris, Zhang, Elia '17]

Assumptions

- No CSIT
- Different Link Capacities
- Allows Cooperative Transmissions
- Caches at both ends

Implementation Challenges

1) Uneven Channel Strengths: Cooperation without CSIT



[Lampiris, Zhang, Elia '17]

Assumptions

- No CSIT
- Different Link Capacities
- Allows Cooperative Transmissions
- Caches at both ends

Achievable Cache-Aided GDoF of MISO BC ($\gamma_T = 1$)

$$d_\Sigma = K(1-\tau) + (K\gamma_R + 1)\tau$$

Topology and Caching complement each other

Implementation Challenges

Achievable Cache-Aided GDoF of Interference Channel

$$\mathcal{T}(\gamma_T,\gamma_R)=\frac{K\gamma_T(1-\gamma_R)}{K(1-\tau)+(K\gamma_R+1)\left(\min\{\frac{2\tau-1}{1-\gamma_T},\tau\}\right)^+}+\frac{K(1-\gamma_T)(1-\gamma_R)}{K_g\min\left\{\frac{\tau}{1-\gamma_T},1\right\}}x_s$$

$$K_g=\frac{K(1-\gamma_T)\binom{K-1}{K\gamma_R}}{\binom{K}{K\gamma_R+1}-\gamma_T\binom{K-1}{K\gamma_R}-\binom{K\gamma_T}{K\gamma_R+1}\frac{K-K\gamma_R-1}{K}}\approx K\gamma_R+1$$

$$x_s=\frac{K(1-\tau)}{K(1-\tau)+\left((K\gamma_R+1)\min\left\{\frac{2\tau-1}{1-\gamma_T},\tau\right\}\right)^+}$$

Implementation Challenges

Achievable Cache-Aided GDoF of Interference Channel

$$\mathcal{T}(\gamma_T,\gamma_R)=\frac{K\,\gamma_T\,(1-\gamma_R)}{K(1-\tau)+(K\gamma_R+1)\left(\min\{\frac{2\tau-1}{1-\gamma_T},\tau\}\right)^+}+\frac{K(1-\gamma_T)(1-\gamma_R)}{K_g\,\min\left\{\frac{\tau}{1-\gamma_T},1\right\}}x_s$$

- Transmitter Side Caching

Implementation Challenges

Achievable Cache-Aided GDoF of Interference Channel

$$\mathcal{T}(\gamma_T,\gamma_R)=\frac{K\,\gamma_T\,(1-\gamma_R)}{K(1-\tau)+(K\gamma_R+1)\left(\min\{\frac{2\tau-1}{1-\gamma_T},\tau\}\right)^+}+\frac{K(1-\gamma_T)(1-\gamma_R)}{K_g\,\min\left\{\frac{\tau}{1-\gamma_T},1\right\}}x_s$$

- Transmitter Side Caching
- Interference Enhancement

Implementation Challenges

Achievable Cache-Aided GDoF of Interference Channel

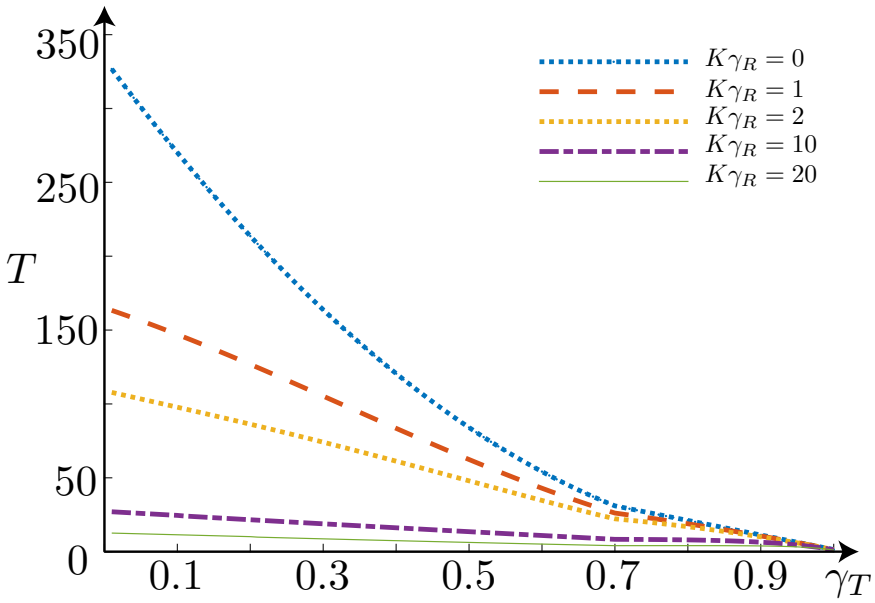
$$\mathcal{T}(\gamma_T,\gamma_R)=\frac{K\,\gamma_T\,(1-\gamma_R)}{K(1-\tau)+(K\gamma_R+1)\left(\min\{\frac{2\tau-1}{1-\gamma_T},\tau\}\right)^+}+\frac{K(1-\gamma_T)(1-\gamma_R)}{K_g\,\min\left\{\frac{\tau}{1-\gamma_T},1\right\}}x_s$$

- Transmitter Side Caching
- Interference Enhancement
- Coded Caching

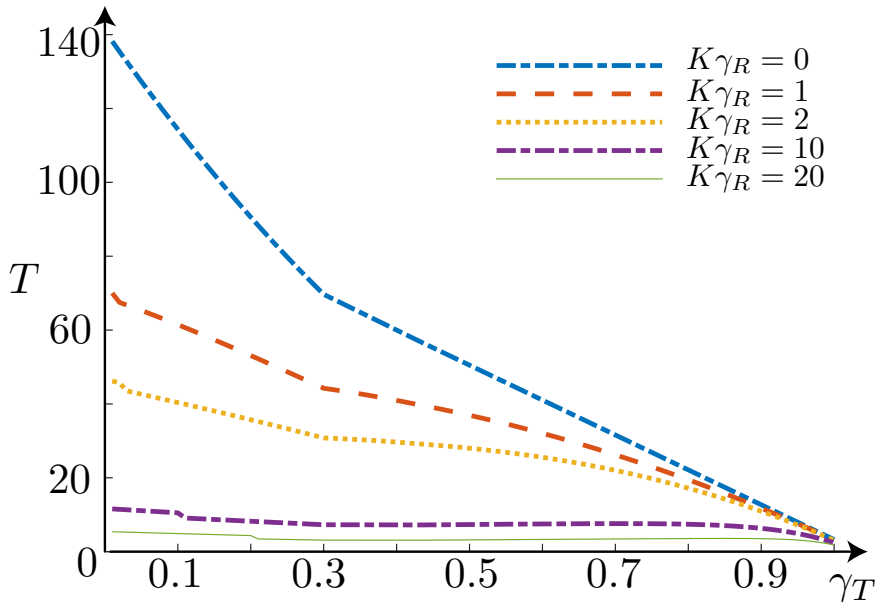
Implementation Challenges

Impact of Transmitter Side Cache

- $\tau = 0.3,\ K = 100$



- $\tau = 0.7,\ K = 100$

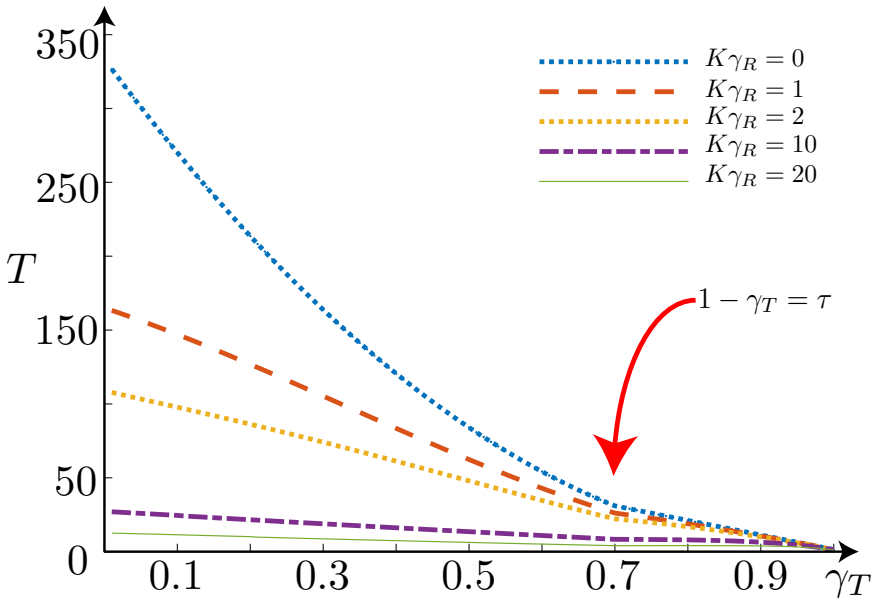


$$\mathcal{T}(\gamma_T,\gamma_R) = \frac{K\gamma_T(1-\gamma_R)}{K(1-\tau)+(K\gamma_R+1)\left(\min\{\frac{2\tau-1}{1-\gamma_T},\tau\}\right)^+} + \frac{K(1-\gamma_T)(1-\gamma_R)}{K_g \min\left\{\frac{\tau}{1-\gamma_T},1\right\}} x_s$$

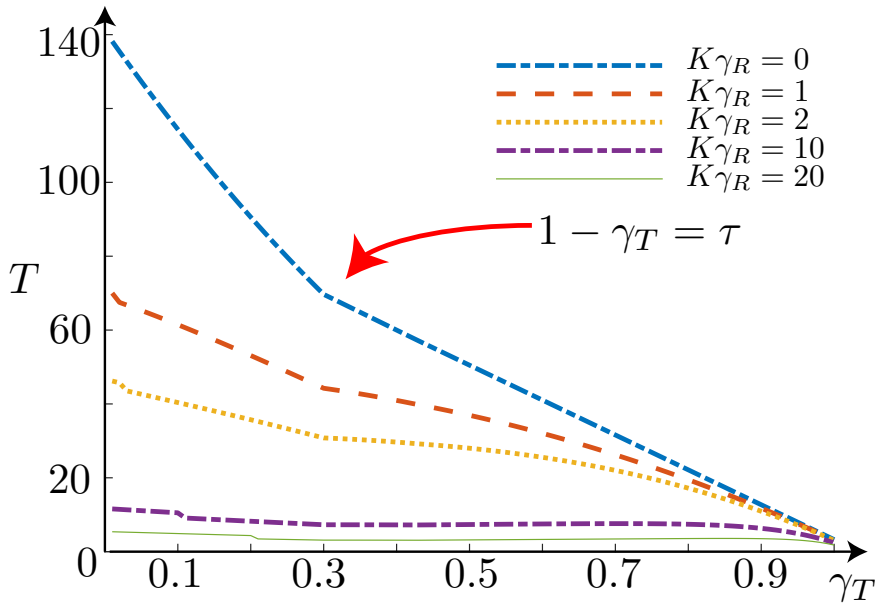
Implementation Challenges

Impact of Transmitter Side Cache

- $\tau = 0.3,\ K = 100$



- $\tau = 0.7,\ K = 100$

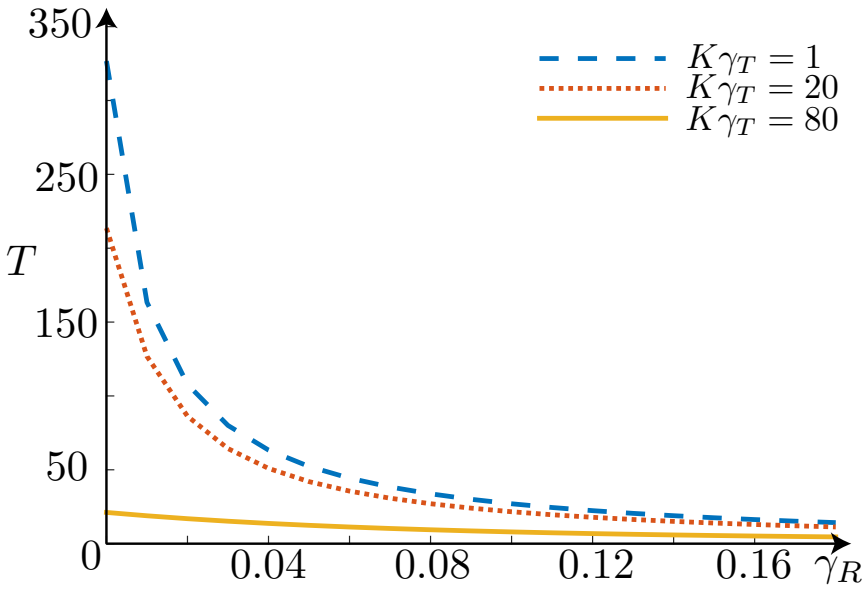


$$\mathcal{T}(\gamma_T,\gamma_R) = \frac{K\gamma_T(1-\gamma_R)}{K(1-\tau)+(K\gamma_R+1)\left(\min\{\frac{2\tau-1}{1-\gamma_T},\tau\}\right)^+} + \frac{K(1-\gamma_T)(1-\gamma_R)}{K_g \min\left\{\frac{\tau}{1-\gamma_T},1\right\}} x_s$$

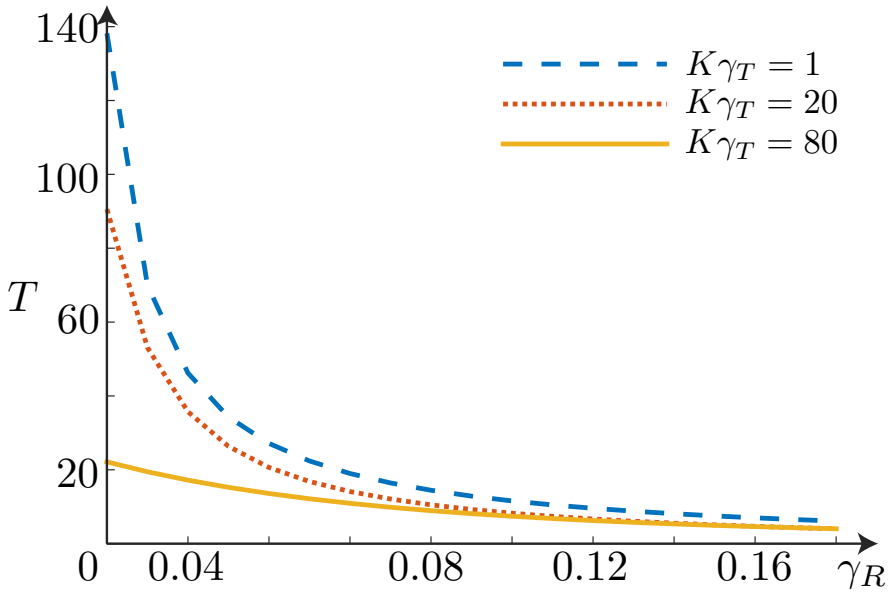
Implementation Challenges

Impact of Receiver Side Cache

- $\tau = 0.3,\ K = 100$



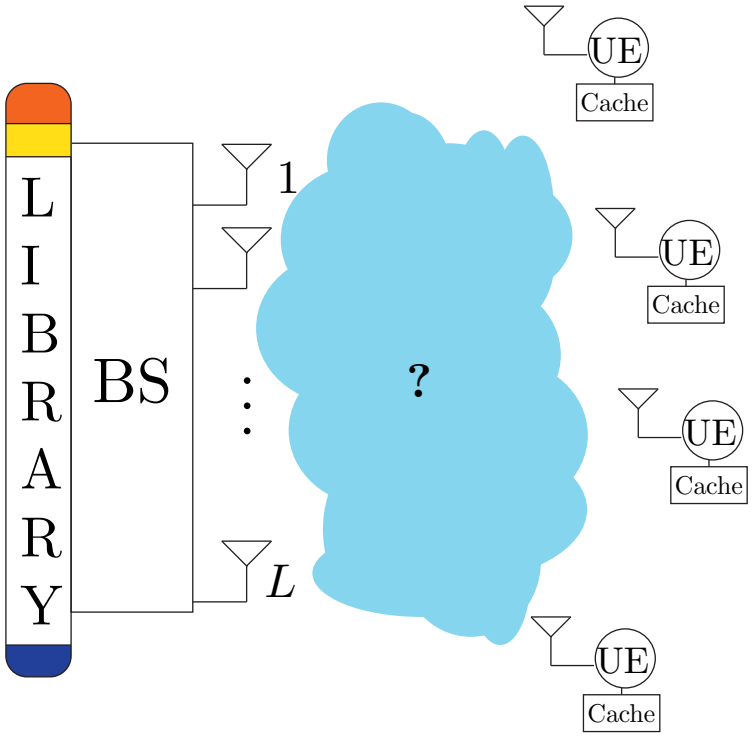
- $\tau = 0.7,\ K = 100$



$$\mathcal{T}(\gamma_T,\gamma_R) = \frac{K\gamma_T(1-\gamma_R)}{K(1-\tau)+(K\gamma_R+1)\left(\min\{\frac{2\tau-1}{1-\gamma_T},\tau\}\right)^+} + \frac{K(1-\gamma_T)(1-\gamma_R)}{K_g \min\left\{\frac{\tau}{1-\gamma_T},1\right\}} x_s$$

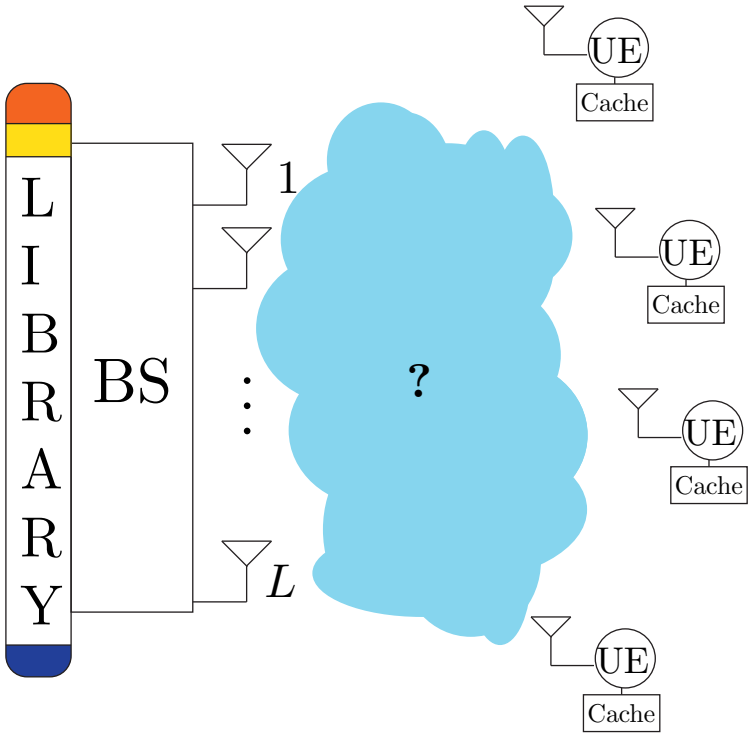
Implementation Challenges

2) Too much CSI needed for MIMO Coded Caching



Implementation Challenges

2) Too much CSI needed for MIMO Coded Caching



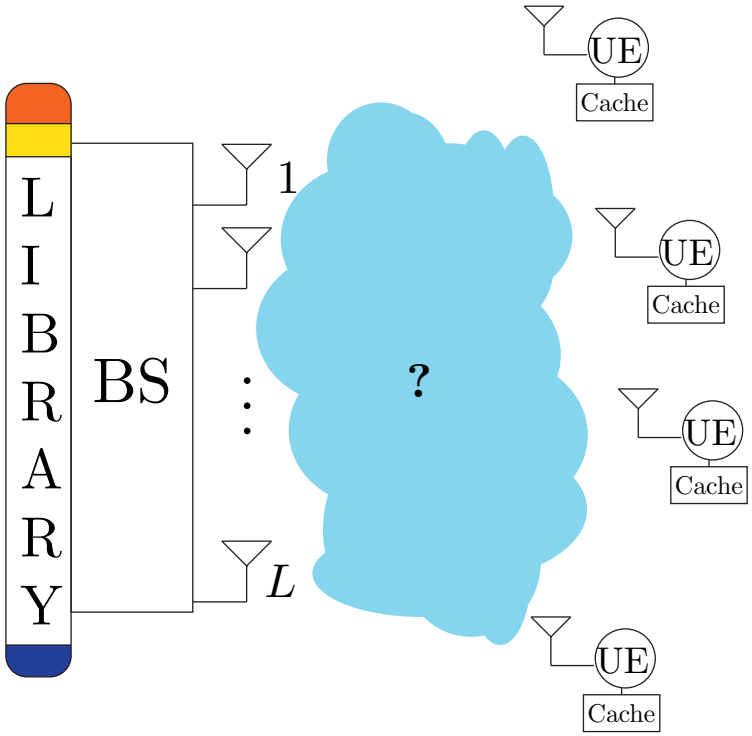
RECALL

$$y_1 = A_2 + C_1(h_{11}h_{12}^{-1} - h_{21}h_{22}^{-1})$$

Requires CSIR

Implementation Challenges

2) Too much CSI needed for MIMO Coded Caching

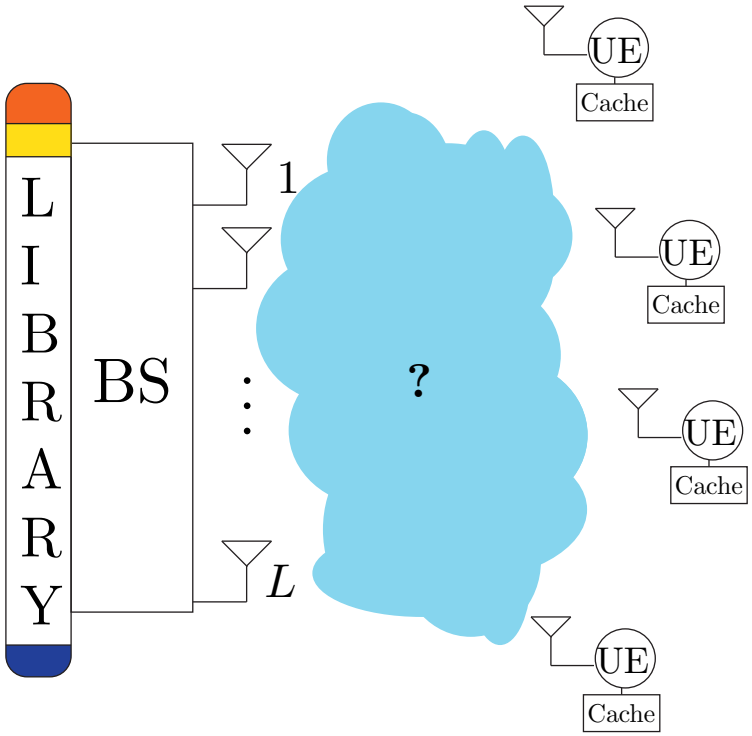


CSIT/CSIR required :
$(K\gamma + L) \cdot L$ variables

MIMO CC requires a lot of training for high gains

Implementation Challenges

2) Too much CSI needed for MIMO Coded Caching



CSIT/CSIR required :
$(K\gamma + L) \cdot L$ variables

MIMO CC requires a lot of training for high gains

More Coming Soon

Stay tuned!

Implementation Challenges

3) Subpacketization: Finite Length Analysis

First result discussing the astronomical file size requirement

Results

- Gain≈ 2 for Decentralized MN if $|F| \leq (\frac{e}{\gamma})^{K\gamma}$
- For a gain of g under any decentralized scheme $|F| = \mathcal{O}(\left(\frac{1}{\gamma}\right)^{g})$

[Shanmugam, Ji, Tulino, Llorca, Dimakis '15]

Implementation Challenges

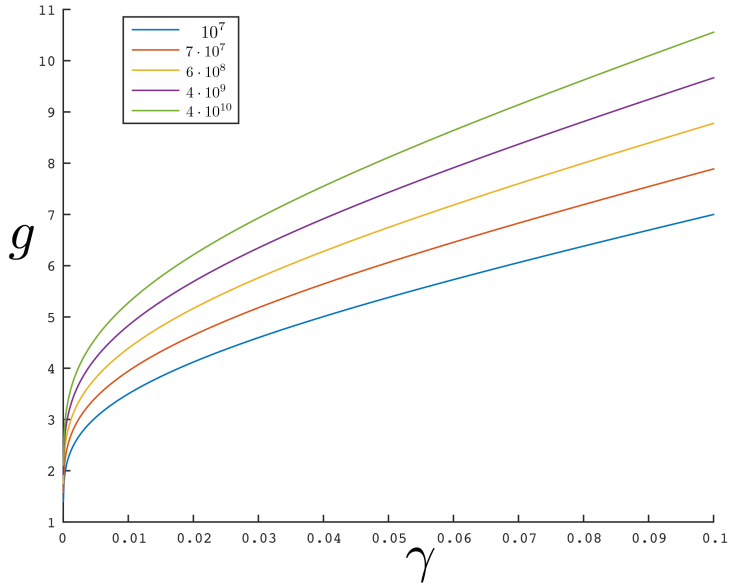
3) Subpacketization: Finite Length Analysis

First result discussing the astronomical file size requirement

Results

- Gain≈ 2 for Decentralized MN if $|F| \leq (\frac{e}{\gamma})^{K\gamma}$
- For a gain of g under any decentralized scheme $|F| = \mathcal{O}(\left(\frac{1}{\gamma}\right)^{g})$

[Shanmugam, Ji, Tulino, Llorca, Dimakis '15]



Take home message

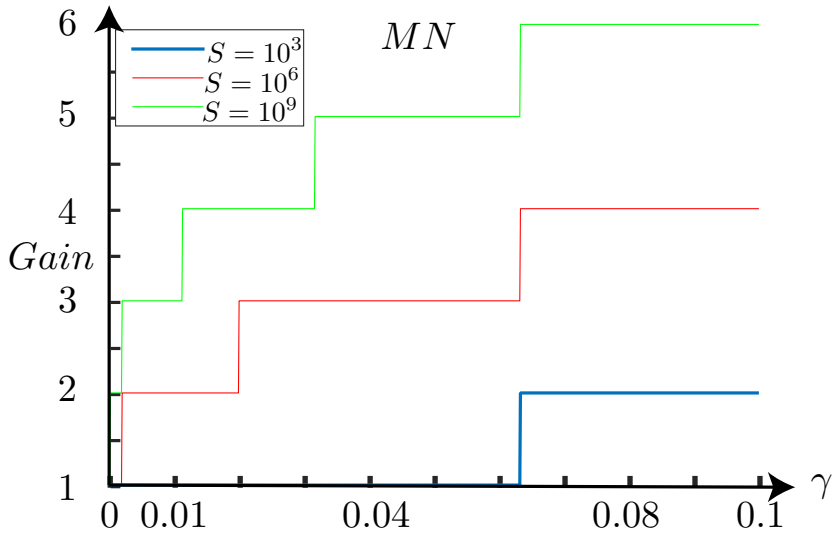
Decentralized CC is limited by a gain of at most 7

Implementation Challenges

3) Subpacketization

Original Subpacketization

$$S_{MN} = \binom{K}{K\gamma}$$



Implementation Challenges

3) Subpacketization

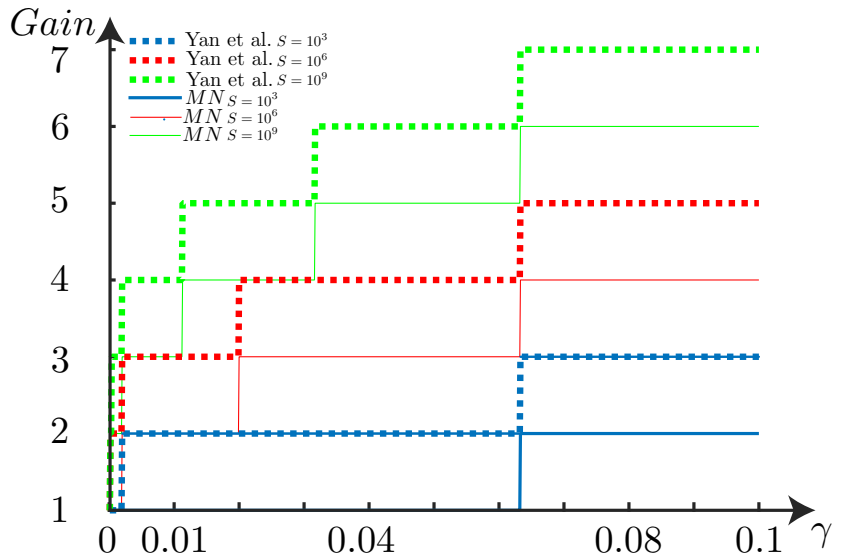
Original Subpacketization

$$S_{MN} = \binom{K}{K\gamma}$$

Reduced Subpacketization

$$S_2 = \left(\frac{1}{\gamma}\right)^{K\gamma-1}$$

[Yan, Cheng, Tang, Chen '15 & Tang, Ramamoorthy '16]



Implementation Challenges

3) Subpacketization: Other Schemes

Linear Subpacketization is Possible

- There exists a caching scheme with linear subpacketization and polynomial delivery time, $|F| = \mathcal{O}(K)$.
- In reality, $K \to \infty \Rightarrow F \to \infty$

[Shanmugam, Tulino, Dimakis '17]

Centralized CC: A hypergraph approach

- There exist no constant rate caching schemes with $|F| = \mathcal{O}(K)$
- Tradeoff between performance and subpacketization
- Schemes require $K > \frac{4}{\gamma^2}$ for $g \geq 2$

[Shangguan, Yiwei Zhang, Gennian Ge '16]

Implementation Challenges

3) Subpacketization: Other Schemes

More Coming Soon

Stay tuned!

[Lampiris, Elia '17]

General Conclusions

Promises

- Coded Caching in theory can severely reduce traffic
- Merging CC with PHY leads to rich and powerful solutions
- MIMO CC can pave the way to supplement MIMO gains with caching.

General Conclusions

Promises

- Coded Caching in theory can severely reduce traffic
- Merging CC with PHY leads to rich and powerful solutions
- MIMO CC can pave the way to supplement MIMO gains with caching.

Hot Questions

- How can subpacketization be reduced?
  - NO Current Algorithm can go beyond a gain of 7 for reasonable γ and subpacketization
- Can feedback in MIMO CC be detangled from Caching Gains?

Thanks!