Caching in wireless communications: feedback and topological considerations

(some salient features of caching for wireless)

Petros Elia
Eurecom
(Joint work with Jingjing Zhang and Eleftherios Lampiris)

Tyrrhenian Workshop - 2016

Main parameters

- K receivers: Each receiver's cache is equipped with a cache of size Mf bits
- Library of size Nf bits (N files – each of size f)

$$\gamma \overset{\text{def}}{=} \frac{M}{N} = \frac{individual\ cache\ size}{library\ size}$$

- Worst-case duration $T(\gamma)$ (rate required)
  - High SNR setting, with $f \propto \log SNR$

- Per-user DoF

$$d(\gamma) = \frac{1-\gamma}{T} \in [0,1]$$

1. Feedback

Feature: Synergy and interplay between memory and feedback

Background

- In most cases, DoF impact of coded caching:

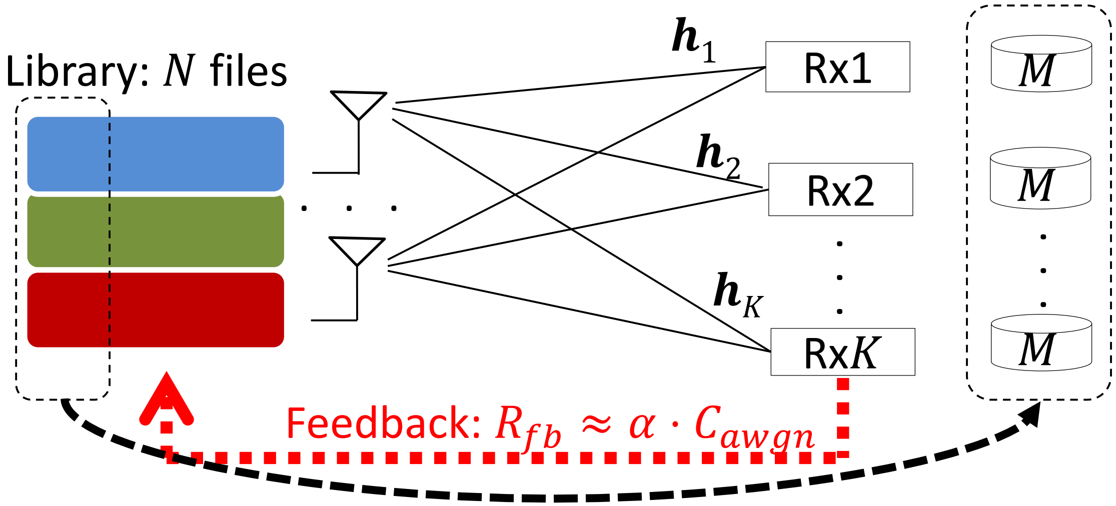
$$d(\gamma) - d(\gamma = 0) = \gamma$$

  - Even in settings with perfect feedback and many antennas

Gains due to cachiing are $\approx \gamma \approx 10^{-3} \rightarrow 10^{-2}$ (Roberts et al.)

- Are there settings for which the impact of caching is substantially larger?

Cache-aided K-user BC with mixed CSIT



- Delayed CSIT + imperfect-quality current CSIT
- High-SNR <u>current-CSIT</u> quality exponent

$$\alpha = -\lim_{P \to \infty} \frac{\log \mathbb{E}\left[||\boldsymbol{h}_k - \widehat{\boldsymbol{h}}_k||^2 \right]}{\log P}, \qquad k \in \{1, \dots, K\}$$

➢ $\alpha = 0$ means ≈no current feedback, and $\alpha = 1$ means perfect CSIT

CSIT/Caching Interplay: MISO BC

Corollary (Zhang-Elia):

$$T(\gamma,\alpha) = \frac{(1-\gamma)\cdot\log(\frac{1}{\gamma})}{\alpha\cdot\log(1/\gamma) + (1-\alpha)(1-\gamma)}$$

Per-user DoF

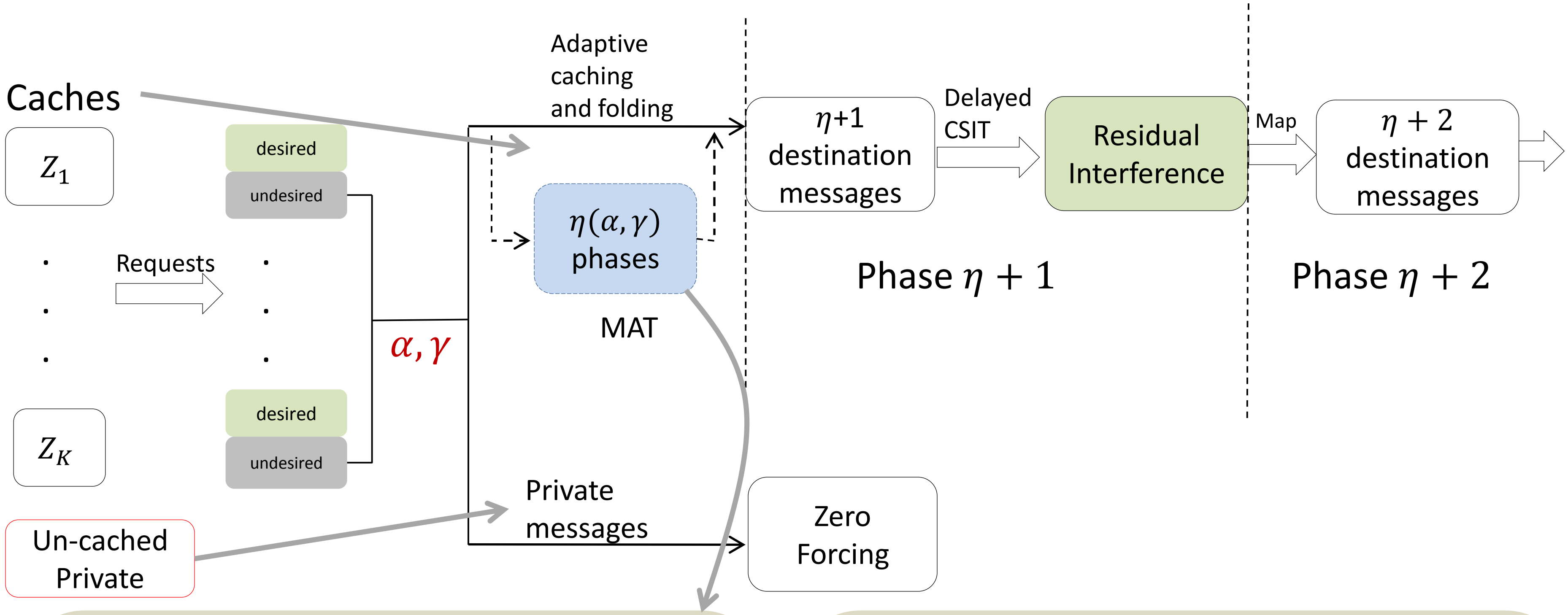
$$d(\gamma,\alpha) = \alpha \; + \; (1-\alpha) \; \frac{1-\gamma}{\log\left(\frac{1}{\gamma}\right)}$$

Features:

- additive combination of resources
- Initial offset due to FB (larger K), and then substantial additional boost due to memory

Under the logarithmic approximation $H_n \approx \log(n)$ (Exact for large K)

Cache-aided Prospective-hindsight Scheme



Feature:

- With delayed CSIT, multicasting is much faster than broadcasting
- Memory boosts broadcasting

Feature: current CSIT increases caching redundancy

$\alpha \uparrow$ ⇒ can have more private data

⇒ Less to be cached

⇒ Caching can be more redundant

⇒ XORS have higher order

⇒ multicast to more users at a time

⇒ Much much faster

Using Coded Caching to `Buffer' CSI

Feature: Caching allows for CSIT reductions (and `buffering')

$\boldsymbol{\gamma'_\alpha = e^{-\frac{1}{\alpha}}}$ can achieve – without current CSIT – the optimal DoF $d^*(\gamma = 0, \alpha)$ associated to a system with delayed CSIT and α-quality current CSIT.

Example (large $\boldsymbol{K}$)

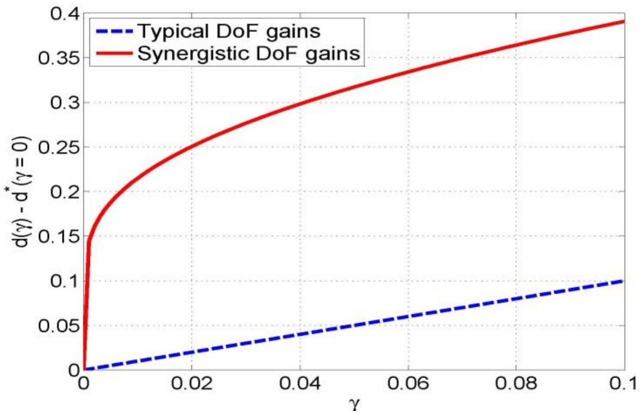
- Assume D-CSIT and α= 1/5. Then

$$\gamma'_{\alpha=\frac{1}{5}} = e^{-5} = 0.0067 \approx \frac{1}{150}$$

$$d^*(\gamma = 0.0067, \alpha = 0) \geq d^*(\gamma = 0, \alpha = 1/5)$$

- The optimal $d^*(\gamma = 0, \alpha = 1/5)$, can be achieved by substituting all current CSIT with DCSIT and coded caching employing $\gamma \approx 1/150$.

Synergistic DoF Gains ($\alpha = 0$)



$$\frac{\delta}{\delta\gamma}\left(d(\gamma) - d(\gamma=0) \right)\Big|_{\gamma=\frac{1}{K}} \approx \frac{K}{\log^2 K} \quad \text{(for all } K\text{)}$$

vs.

$$\frac{\delta}{\delta\gamma}\left(d(\gamma) - d(\gamma=0) \right)\Big|_{\gamma=\frac{1}{K}} = \frac{\delta}{\delta\gamma}(\gamma) = 1$$

- **Feature:** CSIT allows for boost from small (reasonable) amounts of caching
- Synergy because PHY and CC exceed sum of two individual compontents

$$\mathrm{d}(\gamma) > \mathrm{d}_{\mathrm{SS}}(\gamma) + \mathrm{d}_{\mathrm{PHY}}(\gamma = 0)$$

- `Exponential' effect of coded caching (for sufficiently large K)
  - ➢A very small $\gamma = e^{-G}$ can offer a very satisfactory

$$d(\gamma = e^{-G}) - d(\gamma = 0) \rightarrow \frac{1}{G}$$

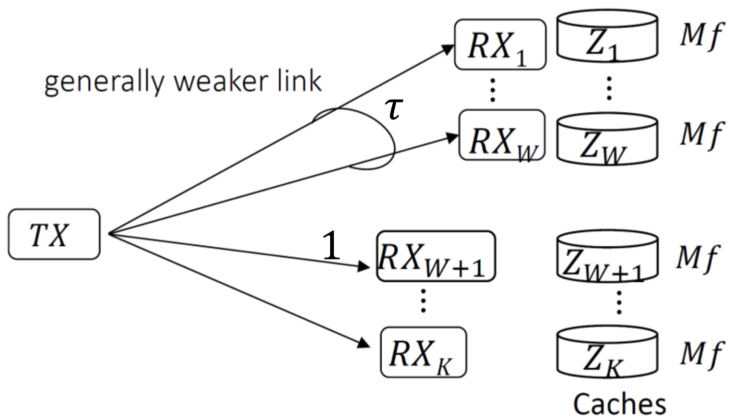
2. Topology (no FB)

Wireless Coded Caching: A Topological Perspective

Features/Opportunities:

- Topological `holes' to attenuate interference
- XORING on the air
- XORs need not be common
- Interesting relationship between coding gain and local caching gain

2A. SISO BC

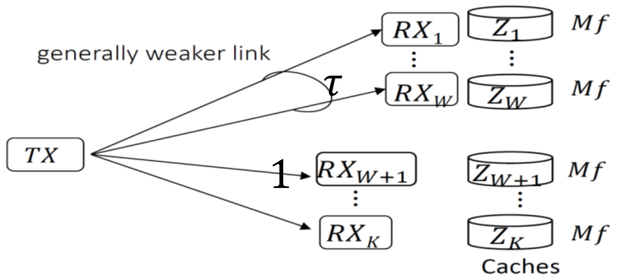


Topologically-uneven wireless <u>SISO</u> K-user BC:

- W weak users with normalized capacity $\tau < 1$
- $K - W$ strong users with normalized capacity $= 1$

- Same cache size per user (M)

System Model



- When $\tau = 1$ (M&N)

$$T(K) = \frac{K(1-\gamma)}{1+K\gamma}, \qquad \gamma = \frac{M}{N}$$

which gives a caching gain

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

- Problem: multicasting can suffer from “worst-user” effect

$$d(\gamma) \to \tau \cdot d(\gamma)$$

- **Opportunity:** Topological `holes’ to attenuate interference
- Question: how is the performance affected as τ decreases?

Main Result

Theorem:

In the K-user topological cache-aided SISO BC with W weak users,

$$T(\tau)=\begin{cases}\dfrac{T(W)}{\tau}, & 0\le \tau\le \bar{\tau}_{thr}\\ \min\left\{T(K-W)+T(W),\dfrac{\tau_{thr}T(K)}{\tau}\right\}, & 0\le \tau\le \tau_{thr}\\ T(K), & \tau_{thr}\le \tau\le 1\end{cases}$$

is achievable and has a gap-to-optimal

$$\frac{T}{T^*}<8$$

that is always less than 8.

$$T(N)=\frac{N(1-\gamma)}{1+N\gamma},\bar{\tau}_{thr}=\frac{T(W)}{T(W)+T(K-W)},\tau_{thr}=\begin{cases}1-\dfrac{\binom{K-W}{K\gamma+1}}{\binom{K}{K\gamma+1}}, & for\ W<K(1-\gamma)\\ 1, & otherwise\end{cases}$$

Topology Threshold

Corollary :

There is a threshold

$$\tau_{thr} \approx 1 - \left(1 - \frac{W}{K}\right)^{g_{max}}$$

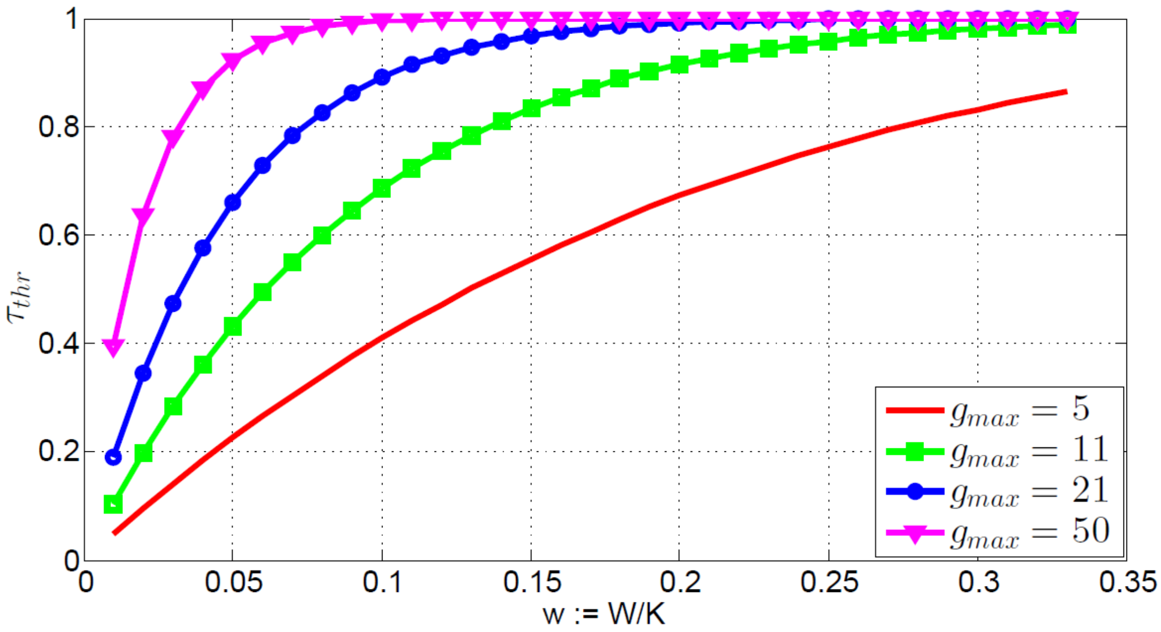
which guarantees full-capacity performance

$$T(\tau \geq \tau_{thr}) = T(K)$$

Recall $g_{max} \stackrel{\text{def}}{=} \mathrm{K}\gamma + 1, w \stackrel{\text{def}}{=} \frac{W}{K}$

$$\tau_{thr} \in \left[1 - (1 - w)^{g_{max}}, 1 - \left(1 - w - \frac{w\gamma}{1-\gamma}\right)^{g_{max}}\right]$$

Topology Threshold

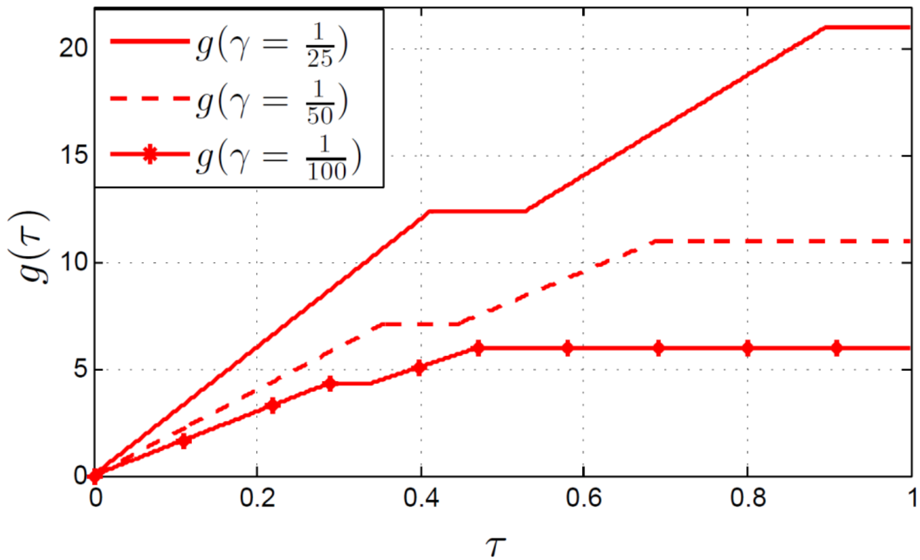


- τ_{thr} corresponding to distinct values for gains g_{max}
- E.g., for $g_{max} = 5$ and $w = 0.1$, then $\tau_{thr} \approx 0.4$

Coded-caching Gain

- Coded-caching gain under topology setting

$$g(\tau) \triangleq \frac{K(1-\gamma)}{T} \in [0, g_{max}]$$

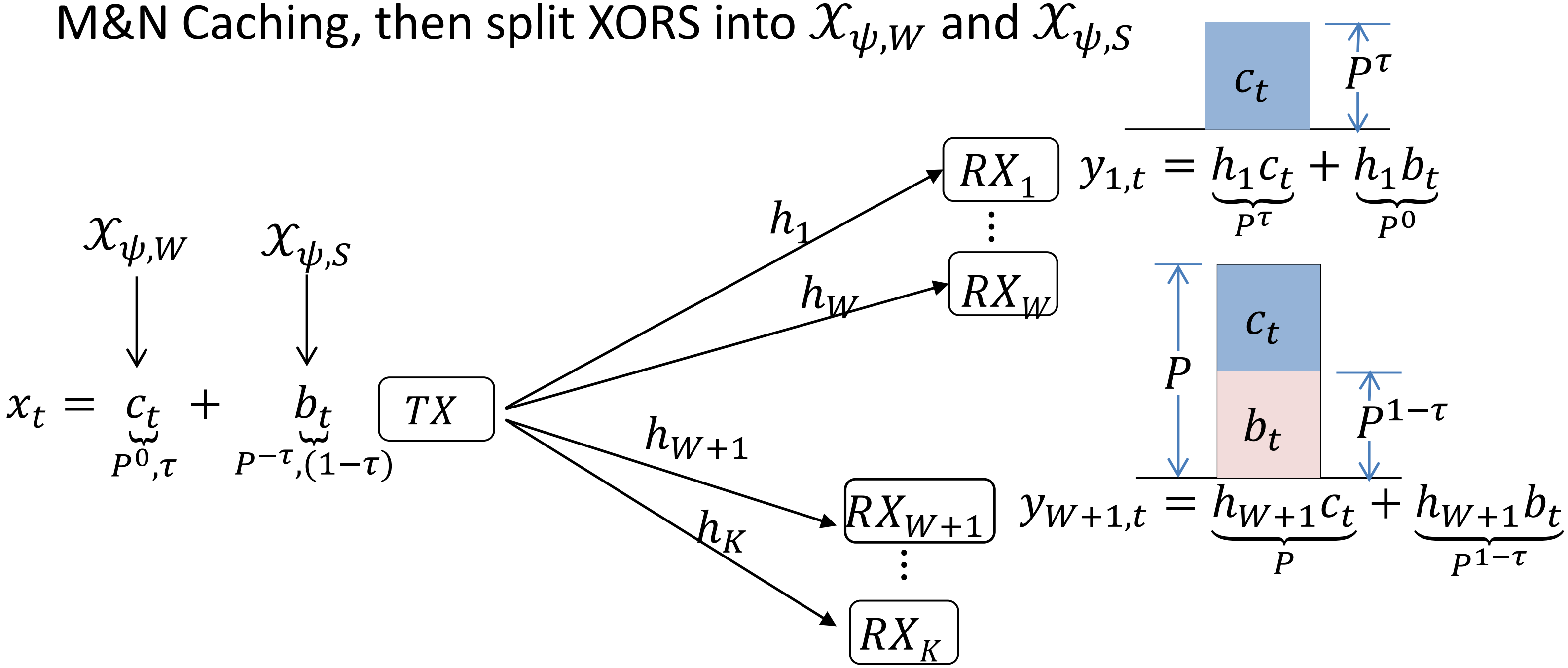


The caching gain for $K = 500, W = 50$

- The horizontal lines denote the maximum gain g_{max} corresponding to $\tau = 1$
- Demonstrate how these can be achieved even with lesser link capacities.

Intuition of the schemes

- M&N Caching, then split XORS into $\mathcal{X}_{\psi,W}$ and $\mathcal{X}_{\psi,S}$



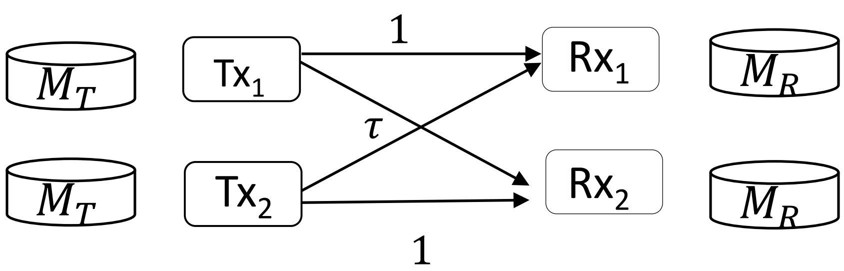
- Interference $\mathcal{X}_{\psi,S}$ hidden from weak users due to topology
  - Treat strong users ($\mathcal{X}_{\psi,S}$) while slowly serving weak ($\mathcal{X}_{\psi,W}$)
  - Transmission rate can be kept (in some cases) at 1 (as if all strong)
  - This ameliorates the negative effects of uneven topology

(Insight)

- For large K (actually for large envisioned gains),… we are in trouble

- Else, `worst-user’ effect can be ameliorated
  - **Feature: Sometimes strong users can lift the performance of the weak users** without any penalties on the overall (worst-case) T

2.B. Cache-aided user cooperation in the topological IC (No CSIT)



2×2 interference channel with caches at the rxs and the txs

- Strong direct link; weak indirect link with strength τ
- The caches at the transmitters hold the entire library, i.e., $M_T \geq \frac{N}{2}$
  - ➢ No need for backhaul the next day
- Difficult setting: no `cooperation’, no CSIT, no backhaul (even for your own data), weak links (dreaded reverse IC)
  - Can `play’ 4 `tricks’

Main Result

Theorem:

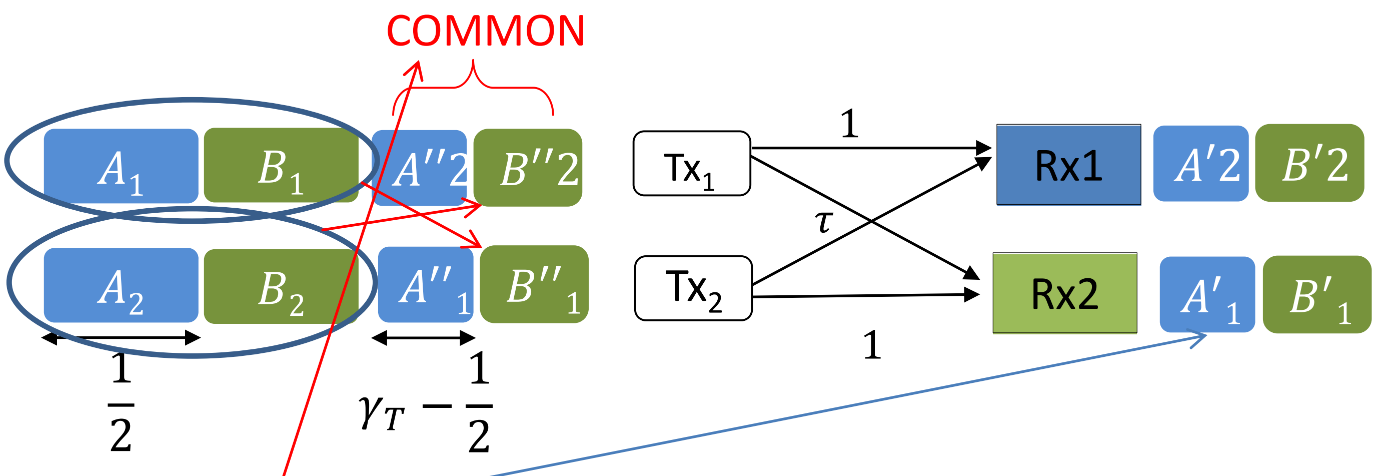
In the 2×2 topological interference channel with $\tau \in \left[0, \frac{1}{2}\right]$, and $\gamma_R = \frac{M_R}{N} \leq \frac{\tau}{2}$, we have

$$d(\tau) = \begin{cases} \dfrac{\tau(1-\gamma_T)(2-\tau)}{2-2\gamma_T-2\gamma_R-\tau+3\tau\gamma_T}, & \gamma_T \geq \dfrac{2-\tau-\gamma_R}{4-3\tau} \\ \dfrac{\tau(1-\tau)(1-\gamma_T)}{1-\gamma_T-\gamma_R}, & \gamma_T \leq \dfrac{2-\tau-\gamma_R}{4-3\tau} \end{cases}$$

is achievable.

$\gamma_R \triangleq \frac{M_R}{N}, \gamma_T \triangleq \frac{M_T}{N}$

Example: $N = 2, \gamma_T > \frac{1}{2}$

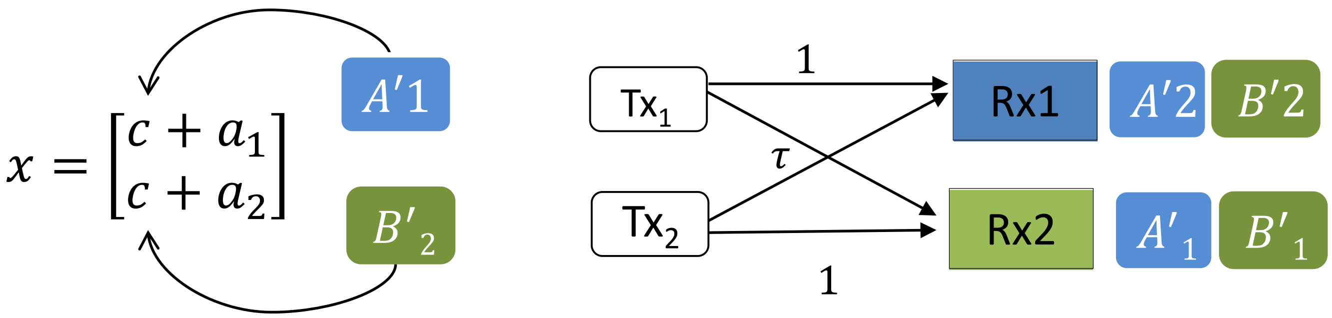


- $A_1 = A''_1 \cup A'_1 \cup A_P$
- Transmission form

$$x = \begin{bmatrix} \underbrace{c}_{P^0,\tau} + \underbrace{a_1}_{P^{-\tau},(1-\tau)} \\ \underbrace{c}_{P^0,\tau} + \underbrace{a_2}_{P^{-\tau},(1-\tau)} \end{bmatrix} \rightarrow \begin{matrix} y_1 = \underbrace{h_1 c}_{P^1,\tau} + \underbrace{h_1 a_1}_{P^{1-\tau},\tau} + \underbrace{h_1 c}_{P^\tau,\tau} \\ y_2 = \underbrace{h_2 c}_{P^1,\tau} + \underbrace{h_2 a_2}_{P^{1-\tau},\tau} + \underbrace{h_2 c}_{P^\tau,\tau} \end{matrix}$$

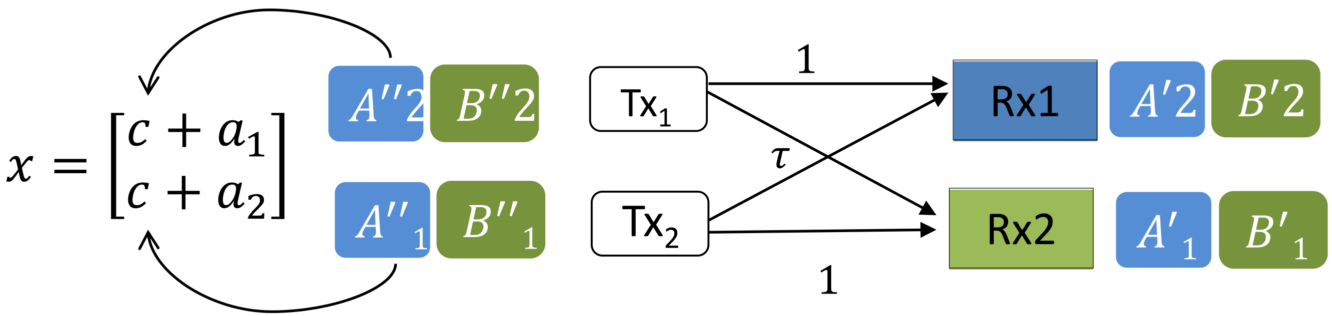
➢ Each user can get its own private data and common data

Example: $N = 2, \gamma_T > \frac{1}{2}$



- **Feature/Trick** 1: XORING in the air
  - ➢ Do not waste the common information
- Trick 2: Embed private information in topological ‘holes’
  - ➢ a_1 and a_2 carry private information (uncached and unshared)

Example: $N = 2, \gamma_T > \frac{1}{2}$



- Feature/Trick 1: XORING in the air
  - Do not waste the common information
- Trick 2: Embed private information due to topology
  - a_1 and a_2 carry private information (uncached and unshared)
- Trick 3: Caching assisted cooperation
  - No need to have full cooperation to trigger full BC benefits
  - Eg. $\tau = \frac{1}{2}$ requires common size = $\frac{1}{3}$ to achieve $d_{\Sigma - BC} = 2 - \tau = \frac{3}{2}$
- Feature/Trick 4: Local caching gain
  - Helps to reduce the data sent from weak links , e.g., for user 1, there is no need to send A'_2 and B'_2, which are only placed in Tx2.

Conclusions

- Several salient features when caching is for wireless
- Certain non-separability between caching and PHY
- Feedback and topology are unexplored frontiers in caching for wireless.
  - Among interesting differentiating ingredients (wireless vs wired)

- Interesting tradeoffs, synergies, and opportunities
- More in tutorials for ICC-16 and Sigmetrics-16

GRAZIE