Multi-user XR

Users immersed into the interactive virtual world via 3D interface

Highly reliable high-throughput wireless connections with low latency are required.
Location-Dependent Multimedia Content

Large part of static (infrastructure) and dynamic (avatars, shapes) content can be stored in advance at user devices.

Coded Caching (CC)

Traditional Caching

<table>
<thead>
<tr>
<th>Request 1</th>
<th>Request 2</th>
<th>Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>Average load</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Coded Caching

<table>
<thead>
<tr>
<th>Request 1</th>
<th>Request 2</th>
<th>Codeword</th>
<th>Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>A₂ ∘ B₁</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>B₂ ∘ A₁</td>
<td>0.5</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A₂ ∘ A₁</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>B₂ ∘ B₁</td>
<td>0.5</td>
</tr>
<tr>
<td>Average load</td>
<td></td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

CC gain is proportional to the total cache size in the network.
Multi-Antenna Coded Caching

CC gain is additive with the spatial multiplexing gain.

---


Location-dependent (Coded) Caching [5, 6]

Location dependency of the content to alleviate wireless bottlenecks:
- Use a priori knowledge of the expected throughput per location;
- Allocate higher cache share for content relevant to locations with poor connectivity;

Can we enable a CC gain for such a non-uniform memory allocation?


Dynamic environment with $K$ users.

Coverage area divided into $|S|$ single transmission units (STU).

Memory allocation per STU based on predicted/expected rate, e.g.:

$$r(s) = \frac{\Omega}{F} C_p \mathbb{E} \left[ \log \left(1 + \frac{P_T \| h_{k,s} \|^2}{N_0} \right) \right] \quad \text{[files/second]} \quad (1)$$
Memory Allocation

- Placement phase: 1) memory allocation and 2) cache placement.
- Worst case total delivery time is first approximated based on $r(s), \forall s \in S$:

$$\hat{T}_T = \frac{K}{K\bar{m} + L} \max_{s \in S} \frac{1 - m(s)}{r(s)} \text{ [seconds]},$$

(2)

where $\bar{m} = \min_{s \in S} m(s)$.

- Memory allocation $m(s) \forall s$ is performed based on minimizing (2):

$$\min_{m(s), \gamma \geq 0, \bar{m} \geq 0} \frac{\gamma}{\bar{m} + \frac{L}{K}}$$

s.t. $$\frac{1 - m(s)}{r(s)} \leq \gamma, \forall s \in S,$$

$$\bar{m} \leq m(s), \forall s \in S, \sum_{s \in S} m(s) \leq M.$$
Memory Allocation – Example

- The term $\frac{L}{K}$ in (3) can be substituted with a general parameter $\phi$.
- A $30 \times 30 [m^2]$ room with $K = 6$ users, $L = 2$ transmit antennas, and $\frac{M}{S} = 0.33$ relative memory size.

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>Avg. DoF</th>
<th>Max. DoF</th>
<th>Min. DoF</th>
<th>DoF std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gg \frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
</tr>
<tr>
<td>$\approx \frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
</tr>
<tr>
<td>$\ll \frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
<td>$\frac{L}{K}$</td>
</tr>
</tbody>
</table>

© Antti Tölli, Centre for Wireless Communications
Cache placement

- Cache placement can be based, e.g., on multi-server scheme [5] or PDA-based scheme [6].
- Non-uniform placement of STU-specific file-fragments.
- A toy example with 5 STUs [5]:

<table>
<thead>
<tr>
<th></th>
<th>$s=1$</th>
<th>$s=2$</th>
<th>$s=3$</th>
<th>$s=4$</th>
<th>$s=5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r(s)$</td>
<td>$3 \times 10^3$</td>
<td>$2 \times 10^3$</td>
<td>$1 \times 10^3$</td>
<td>$2 \times 10^3$</td>
<td>$3 \times 10^3$</td>
</tr>
<tr>
<td>$m(s)$</td>
<td>0.25</td>
<td>0.5</td>
<td>0.75</td>
<td>0.5</td>
<td>0.25</td>
</tr>
</tbody>
</table>

![Toy Example Diagram]


Content Delivery

Delivery phase: users request location-dependent content.

The BS transmits precoded messages $x_U$ to users $U \subseteq K$ over several transmit intervals.
Different Delivery Methods

- NCM-based scheme [5] enables parallel multicast transmissions to user groups with size $|\mathcal{T}|$: $x_{\mathcal{U}} = \sum_{\mathcal{T} \subseteq \mathcal{U}} v_{\mathcal{T}} x_{\mathcal{T}}$.
- The PDA-based scheme [6] enables unicast transmission by sending $x_{\mathcal{U}} = \sum_{k \in \mathcal{U}} v_k x_k$.
- Achievable degrees of freedom (DoF) for both schemes is $\min_{k \in \mathcal{U}} K m(s_k) + L$.

$$(A_2 \ast B_1) V_{12} \rightarrow A_2 V_1 + B_1 V_2$$
Different Transmission/Reception Signals (PDA - based)

- Consider the following transmitted signal using PDA,

\[ x_{\text{PDA}}(1) = A_4 v_1 + B_1 v_2 + C_1 v_3 \]

- Corresponding received signal at the users,

\[ y_1 = A_4 h_1^H v_1 + B_1 h_1^H v_2 + C_1 h_1^H v_3 + w_1, \]
\[ y_2 = A_4 h_2^H v_1 + B_1 h_2^H v_2 + C_1 h_2^H v_3 + w_2, \]
\[ y_3 = A_4 h_3^H v_1 + B_1 h_3^H v_2 + C_1 h_3^H v_3 + w_3 \]

- Underlined terms are regenerated and removed using cache and the double-underlined terms suppressed by precoder \( v_k \).

- No need for SIC at the receiver.
Weighed Maxmin (WMM) Beamforming

- Different fractions of files missing given the UE locations $s_k, k \in \mathcal{U}$.
- The WMM multicast beamformer design [5] (l.h.s) and the unicast beamformer design [6] (r.h.s):

\[
\max \left\{ v_T, \gamma^k_T, R^Q_{\text{sum}} \right\} \min_{k \in \mathcal{U}} \min_{Q \subseteq \mathcal{D}_k} \frac{1}{c_k |Q|} R^Q_{\text{sum}} \\
\text{s.t.} \quad R^Q_{\text{sum}} \leq \log \left( 1 + \sum_{T \in Q} \gamma^k_T \right),
\]

\[
\gamma^k_T \leq \frac{|\mathbf{h}_k^H v_T|^2}{\sum_{V \in I_k} |\mathbf{h}_k^H v_V|^2 + \sigma^2},
\]

\[
\sum_{\mathcal{T} \subseteq \mathcal{U}, |\mathcal{T}| = \hat{t} + 1} \|v_T\|^2 \leq P_T,
\]

\[
\max \left\{ v_k, \gamma_k \right\} \min_{k \in \mathcal{U}} \frac{1}{c_k} \log(1 + \gamma_k) \\
\text{s.t.} \quad \gamma_k \leq \frac{|\mathbf{h}_k^H v_k|^2}{\sum_{i \in I_k} |\mathbf{h}_k^H v_i|^2 + N_0},
\]

\[
\sum_{k \in \mathcal{U}} \|v_k\|^2 \leq P_T,
\]

Comparison with uniform placement: lower DoF \( \frac{(K \min_k (m_k) + L)}{(K \frac{M}{S} + L)} < 1 \) is compensated by improved transmission rates \( \frac{R_w}{R_u} > 1 \).
Numerical Examples [6]

Large number of users ($K = 12 - 60$). Parameter $\sigma$ models the impact of random placement of obstacles.

Significant performance improvement due to optimal memory allocation → further gain by leveraging coded caching transmission.
Simulation Results [6]

- Higher gains at lower available transmit powers.
- At very high SNR, non-uniform placement provides less gains.
- Largest gains available when pathloss variation ($\sigma$) is high.

Simulation Results [6]

- Higher achievable coded caching gain $K \min m(s_k)$ in larger networks $\rightarrow$ higher performance gain.

- When the number of spatial multiplexing gain $L$ is relatively larger than CC gain, unicast transmission achieves almost the same performance as CC-based schemes.
Conclusion

- Location-aware CC scheme is proposed for location-dependent data delivery to avoid excessive transmission time.

- A trade-off between local and global caching gain is achieved through novel memory allocation.

- Using the CC scheme a global caching gain additive to the multiplexing gain is achieved.

- Most effective in scenarios with large expected throughput variation.
Nested Code Modulation

- Multi-rate transmission, based on side information at the RX
- An example of nesting QPSK constellations within 16 QAM.
Different Transmission/Reception Signals (NCM - based)

Consider the following transmitted signal using NCM,

$$x_{\text{NCM}}(1) = [A_2 \ast B_1]v_{12} + [A_3 \ast C_1]v_{13} + [B_3 \ast C_2]v_{23}$$

Corresponding received signal at the users,

$$y_1 = [A_2 \ast B_1]h_1^H v_{12} + [A_3 \ast C_1]h_1^H v_{13} + [B_3 \ast C_2]h_1^H v_{23} + w_1,$$

$$y_2 = [A_2 \ast B_1]h_2^H v_{12} + [A_3 \ast C_1]h_2^H v_{13} + [B_3 \ast C_2]h_2^H v_{23} + w_2,$$

$$y_3 = [A_2 \ast B_1]h_3^H v_{12} + [A_3 \ast C_1]h_3^H v_{13} + [B_3 \ast C_2]h_3^H v_{23} + w_3$$

Underlined terms are removed using cache and the double-underlined terms are suppressed using precoder $v_{ij}$.

Superposition of multiple messages at each user $\rightarrow$ successive interference cancellation (SIC) at UEs.