Understanding Optimal Caching and Opportunistic Caching at “The Edge” of Information Centric Networks

Ali Dabirmoghaddam¹
Maziar Mirzazad-Barljough¹
J.J. Garcia-Luna-Aceves¹,²

¹UC Santa Cruz
²Palo Alto Research Center
jj@soe.ucsc.edu
Outline

How important is edge caching for ICNs?

- Compare optimum on-path caching with opportunistic caching at the edge [near the consumers of content] using a hierarchical caching model (approximation).
- Model the spatio-temporal locality of reference in content requests. [just a start!]
- Compare on-path caching with caching at the edge in random networks (ndnSim simulations).
- Modeling and design implications.
Related Work

- Models rely on strong assumptions to simplify problem:
  - Equal size objects (unit size)
  - Independent reference model (IRM): requests for information objects arrive according to i.i.d. processes

- Caching used in most ICN approaches is on-path caching
  - All routers participate in content caching.
  - Caching done along path to origin sites.
  - Inconclusive results on impact of caching at the core of ICNs.

- Not enough comparison of on-path caching with edge caching.
Hierarchical Caching Model

- Simple hierarchy of LRU caches in $L+2$ levels.
- Consumers are at level 0.
- Content source at level $L+1$.
- $L$ levels of caching.
- Each tree node has $k$ children.
- On-path caching:
  - Forward request from consumer to root until cache hit occurs.
  - Cache content along entire reverse path.
- Assume IRM.
- How much should each layer of the caching hierarchy store to get the most from a constrained caching budget?

Of course, Level 1 should have a much larger degree, but that would help make our case stronger.
Hierarchical Caching Model

- Caching tree structure of \( L+2 \) levels, assume IRM.
- \( \tau(n) \): Expected time to access content (ETTA):
  - Measured in terms of the number of hops between consumer and nearest copy of content.
  - \( m_j(n) \): Miss probability of a given cache for object \( n \) at level \( i \) of the caching system.
  - \( h_j(n) = 1 - m_j(n) \)

\[
\tau(n) = 1 + \sum_{i=1}^{L} \prod_{j=1}^{i} m_j(n)
\]
Hierarchical Caching Model

- $\tau_i(n)$: Expected time to visit state $H$ from state $i$.
- $\tau_0(n)$ can be expressed recursively:
  \[ \tau_0(n) = 1 + \tau_1(n) = 1 + 1 + m_1(n) \tau_2(n) + h_1(n) \tau_H(n) = 1 + 1 + m_1(n) \tau_2(n) + [1 - m_1(n)] \tau_H(n) \]
- By induction:
  \[ \tau_0(n) = 2 + \sum_{i=1}^{L} \prod_{j=1}^{i} m_j(n) \]
- Given that state $H$ has the content,
  \[ \tau(n) = \tau_0(n) - 1 \] and the result follows.
Hierarchical Caching Model

Optimal cache allocation problem:

- Given a total cache budget $C$ and a caching tree of $L$ levels with each node having $k$ children
- **Find the optimum breakdown of $C$ across all levels that minimizes ETTA.**

\[
c^* = \text{argmin}_{c} \sum_{n=1}^{N} q(n) \tau(n; c)
\]

\[
\text{s.t.} \sum_{\ell=1}^{L} c(\ell) k^{(L-\ell)} = C, \quad \text{and}
\]

\[
c(\ell) \geq 0 \text{ and integer } \forall \ell \in \{1, \ldots, L\}
\]

- $q(n) = \text{popularity of content } n.$
- $c(\ell) = \text{capacity of a cache at level } \ell.$
- $c^* = \text{vector of optimal cache sizes.}$
Optimal Breakdown of Caching Budget

- Consumers at level 0, four caching levels, content at level six.
- 1M objects, IRM.
- Identical objects have same popularity among all users.
- Proportion of caching for increasing degree $k$ and budget $C$.

Edge caching becomes more important as $C$ increases!
ETTA for Optimal Caching and Edge Caching

- Same assumptions as before for optimal caching.
- **Edge caching**: All caching budget $C$ only at Level 1.
- Results based on model and ndnSim simulations.

Edge caching is within 10% of the optimum!
Capturing Spatio-Temporal Reference Locality

- **Spatial locality of reference**: Impact of user geographical diversity on content requests.
- **Temporal locality of reference**: Temporal evolution of content popularity.

Generate “center points” of a Poisson process that generate off-springs and so on based on number of objects, object popularity [Zipf distribution], and localization factor.

**Reference locality** \((\beta) = \text{ave. } \# \text{ off-springs for each center point } (0 \leq \beta < 1);\)

\[
q_\ell(n) = \begin{cases} 
q(n) & \ell = 0, \\
kq_{\ell-1}(n)m_{\ell-1}(n) & 0 < \ell \leq L,
\end{cases}
\]

\[
q_u(n, t) = \begin{cases} 
\lim_{\Delta t \to 0} \frac{\mathbb{E}[N_u(n, t + \Delta t)]}{\Delta t} & u \in \{\ell_1\}, \\
\sum_{c \in C_u} q_c(n, t)m_c(n, t) & u \notin \{\ell_1\}.
\end{cases}
\]
Algorithm for Generating Object References with Localization in a d-dimensional Space

**Generate_Trace(α, β)**

\[ X \leftarrow \emptyset \]

for every object \( n \) {

\[ q_n \propto n^{-\alpha} \]

\( \Pi_n \leftarrow \text{Hawkes}(q_n, \beta) \)

\( X \leftarrow X \cup \Pi_n \)

} return \( X \)

**Hawkes(\( \rho, \beta \))**

\[ n_t \leftarrow \text{Poisson}(\rho) \]

for \( i \leftarrow 1 \) to \( n_t \) {

\[ \Pi_i \leftarrow \text{Uniform}(0,1) \]

\( \text{idx} \leftarrow 1 \)

end \( \leftarrow n_t \)

While \( \text{idx} < n_t \) {

\( n_c \leftarrow \text{Poisson}(\beta) \)

for \( j \leftarrow 1 \) to \( n_c \) {

\( \Pi \leftarrow \Pi \cup (\Pi_{\text{idx}} + \mathcal{N}(0,1)) \)

\( n_t \leftarrow n_t + n_c \)

} return \( \Pi \)
Optimal Breakdown of Caching Budget

- Consumers at level 0, four caching levels, content at level six.
- Reference locality ($\beta$) varied from 0 (IRM) to 1.
- Proportion of caching for increasing $\beta$ and budget $C$.

Edge caching dominates as $\beta$ and $C$ increase!
ETTA for Optimal Caching and Edge Caching

- Same assumptions as before for optimal caching.
- **Edge caching**: All caching budget $C$ only at Level 1.
- Results based on model and ndnSim simulations.

Edge caching is within 8% of the optimum as $C$ and $\beta$ increase.
Caching On Random Networks

- Model random networks with random geometric graphs.
- Voronoi cells and local caches in each cell.
- On-path caching: Caching along entire path.
- Edge caching: Caching only within the cell in which request originated.
- Compare the two using ndnSim simulations:
  - 200 nodes; 8.9 average node degree
  - 1,000 objects with Zipf popularity distribution, uniformly distributed among nodes.
Edge vs. On-Path Caching in Random Networks

Edge caching outperforms on-path caching for all values of $\beta$!
Edge vs. On-Path Caching in Random Networks

Edge caching outperforms on-path caching for all cache sizes.
Modeling framework for hierarchical caching:
- Optimal on-path caching provides only marginal benefits over edge caching.

Tool introduced to synthesize spatial and temporal locality in traces of object requests
- Optimal caching tends towards the edge for both larger locality of reference and caching budget.

Compared edge and on-path caching in random networks:
- Edge caching outperforms on-path caching.

Edge caching provides “less pain, most of the gain” in ICNs.
Next Steps and Implications

- More realistic topologies.
- Verify synthetic traces for locality of reference model with real traffic traces.
- Develop model for random networks.
- Since edge caching provides most of the caching gains:
  - Only edge routers in ICNs need large caches.
  - New approaches to integrate routing with edge caching.
THANK YOU!

ANY QUESTIONS?