Content, Topology and Cooperation in In-Network Caching

Liang Wang

Email: liang.wang@cs.helsinki.fi

Department of Computer Science

University of Helsinki, Finland

Why We Need Content Networking

- Content distribution is the primary task for today's Internet. E.g., the estimated video traffic will reach 79% of the Internet traffic by 2018.
- Traditional paradigm of communication network is Point-to-Point.
- Point-to-Point paradigm has many drawbacks when dealing with large-scale content distribution - efficiency, security and privacy.

Content consumer only cares what it is instead of where it is from.

Information-Centric Network Architecture

This thesis focuses on Information-Centric Networking (ICN).

ICN is a clean-slate redesign of the current Internet infrastructure,

- Content is accessed by name.
- Caching is universal in the network.

ICN tries to solve the problems confronting the current Internet, e.g., content distribution efficiency, security, network congestion and etc.

Meanwhile, ICN also poses new challenges on cache management, content addressing, routing and etc.

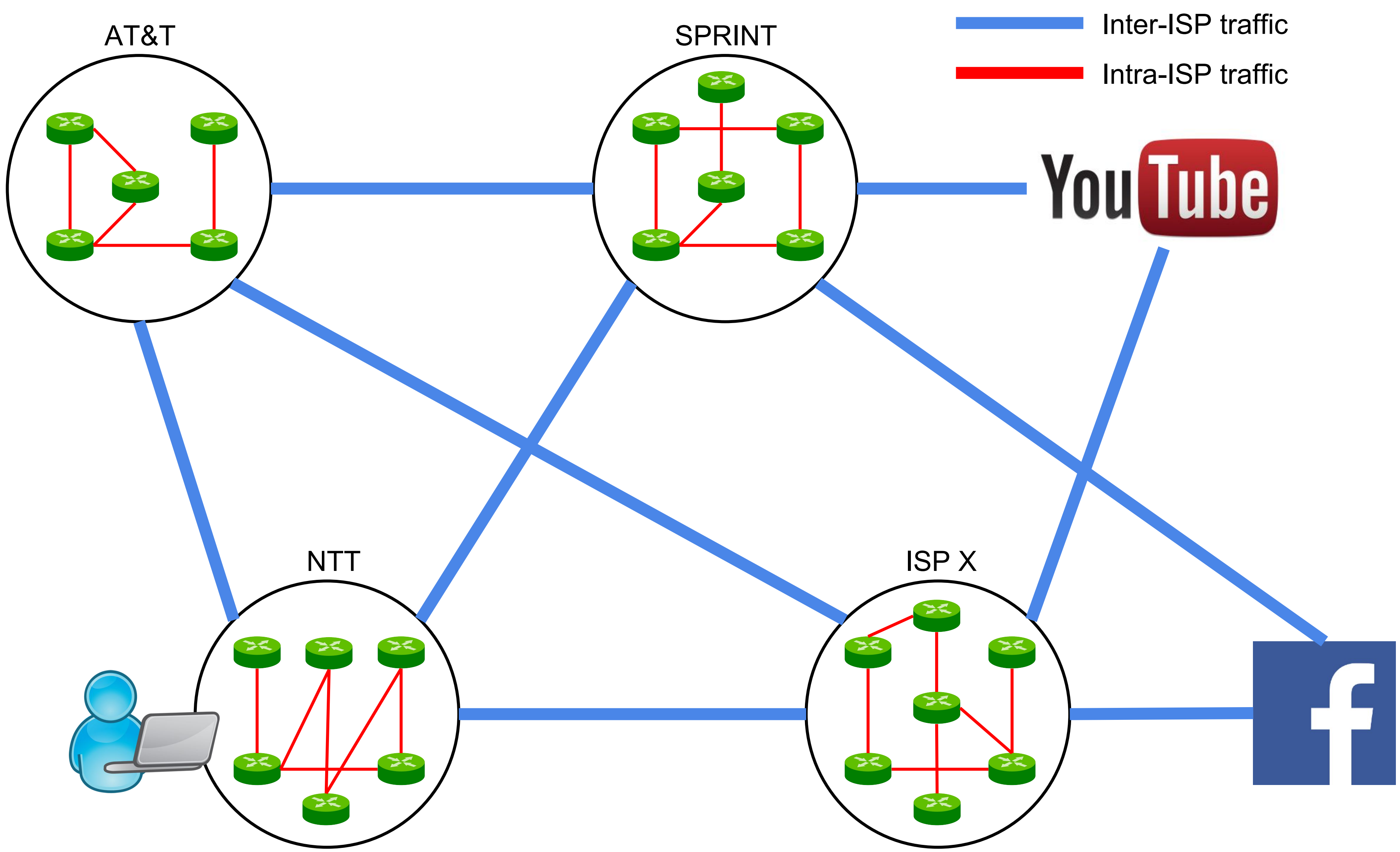
Thesis Contributions

This thesis studied the following topics in ICN context.

1. Evaluation methodology and metrics
2. Model of cache networks and collaboration
3. Fair collaborative caching game
4. Collaboration cost on general topologies
5. Compact routing in ICN
6. Content chunking analysis

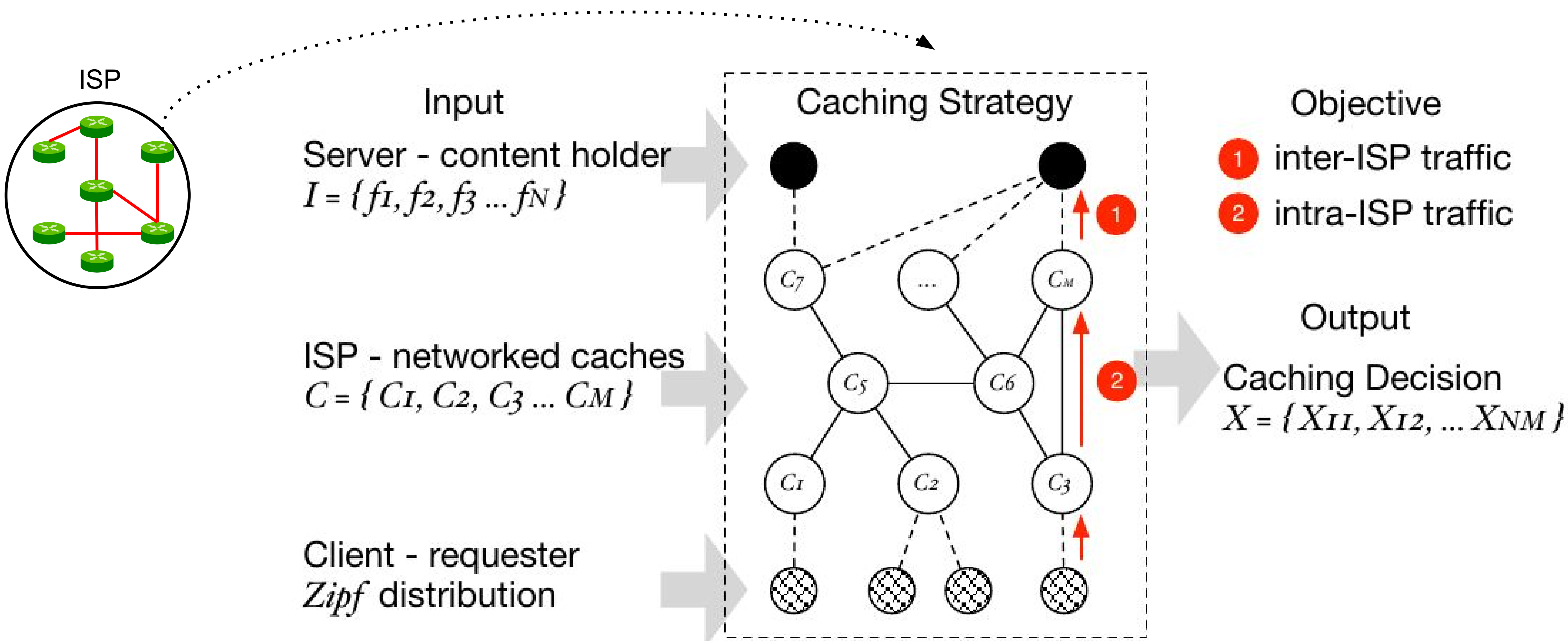
The Big Picture of Today's Internet

A very high-level abstraction of current Internet: ISPs are interconnected with each other, along with big service providers. End-users are attached to various ISP networks.



Model of Cache Network

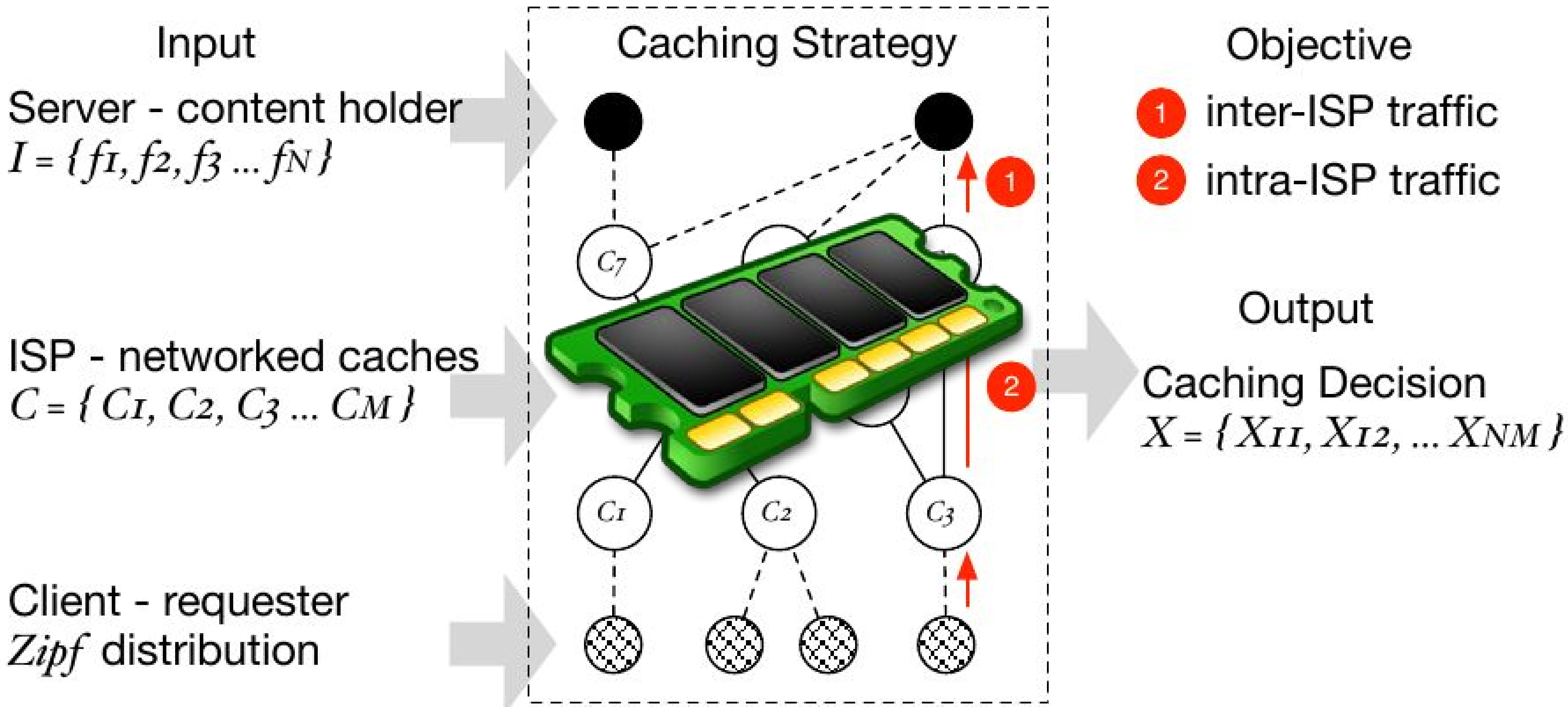
Given a group of networked caches, how to utilize them smartly and efficiently in order to push the system to its optimal state?



Essentially, ICN manages a group of networked caches.

Model of Cache Network

Given a group of networked caches, how to utilize them smartly and efficiently in order to push the system to its optimal state?



We want to use them as a single big cache.

Simple Question, Hard Challenges

As mentioned, the goal is to achieve "the optimal state of the system",

However,

- How do you define the optimum for an ICN system?
- What metrics do you use to quantify the performance?
- What is the model for collaboration?

We need enough metrics to build up a holistic view of ICN systems.

Single Cache vs. Cache Network

The fundamental difference between a single cache and a cache network:

The topological structure becomes a system parameter in ICN designs.

- Content caching ≠ Content addressing
- Effective capacity ≠ Aggregated cache size
- Local optimum ≠ Global optimum

The whole system should not be treated as a simple "entity", we need examine the internal topological structures of a cache network.

The Metrics

In-network caching concerns traffic localization

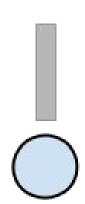
→ i.e. how much saving on inter-ISP and intra-ISP traffic?

→ i.e. how many cache hits and where they occur in the network?

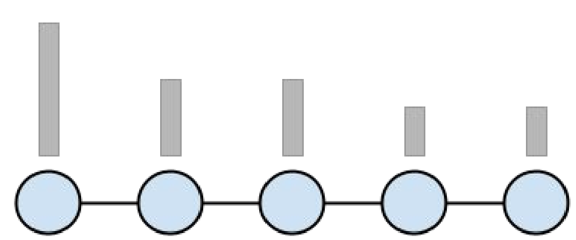
Byte hit rate (BHR) - the saving on inter-ISP traffic.

Footprint Reduction (FPR) - the saving on intra-ISP traffic.

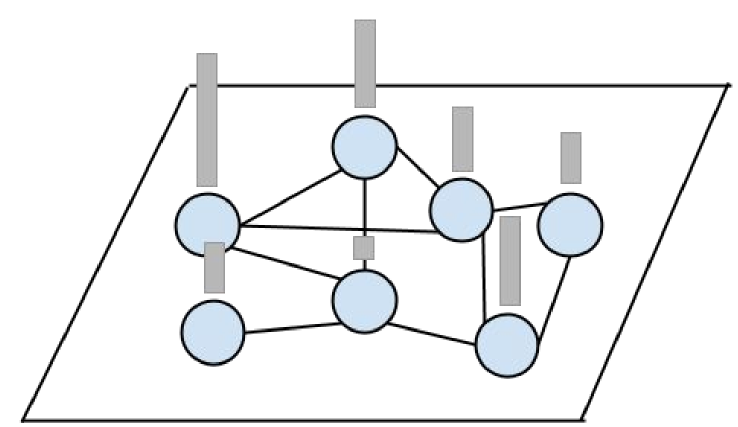
Coupling factor (CPF) - the distribution of the saving within a network.



BHR



FPR



CPF

The Role of Collaboration

In the conventional single cache context, admission control and replacement policy answers WHAT question.

- What content to admit into a cache?
- What content to evict from a cache?

In the cache network context, collaboration answers WHERE question.

- Where to cache the popular content in a cache network?
- Where to fetch the popular content in a cache network?

W. Wong, L. Wang, and J. Kangasharju, ”Neighborhood Search and Admission Control in Cooperative Caching Networks,” in the Proceedings of IEEE Globecom. IEEE, December 3-7 2012.

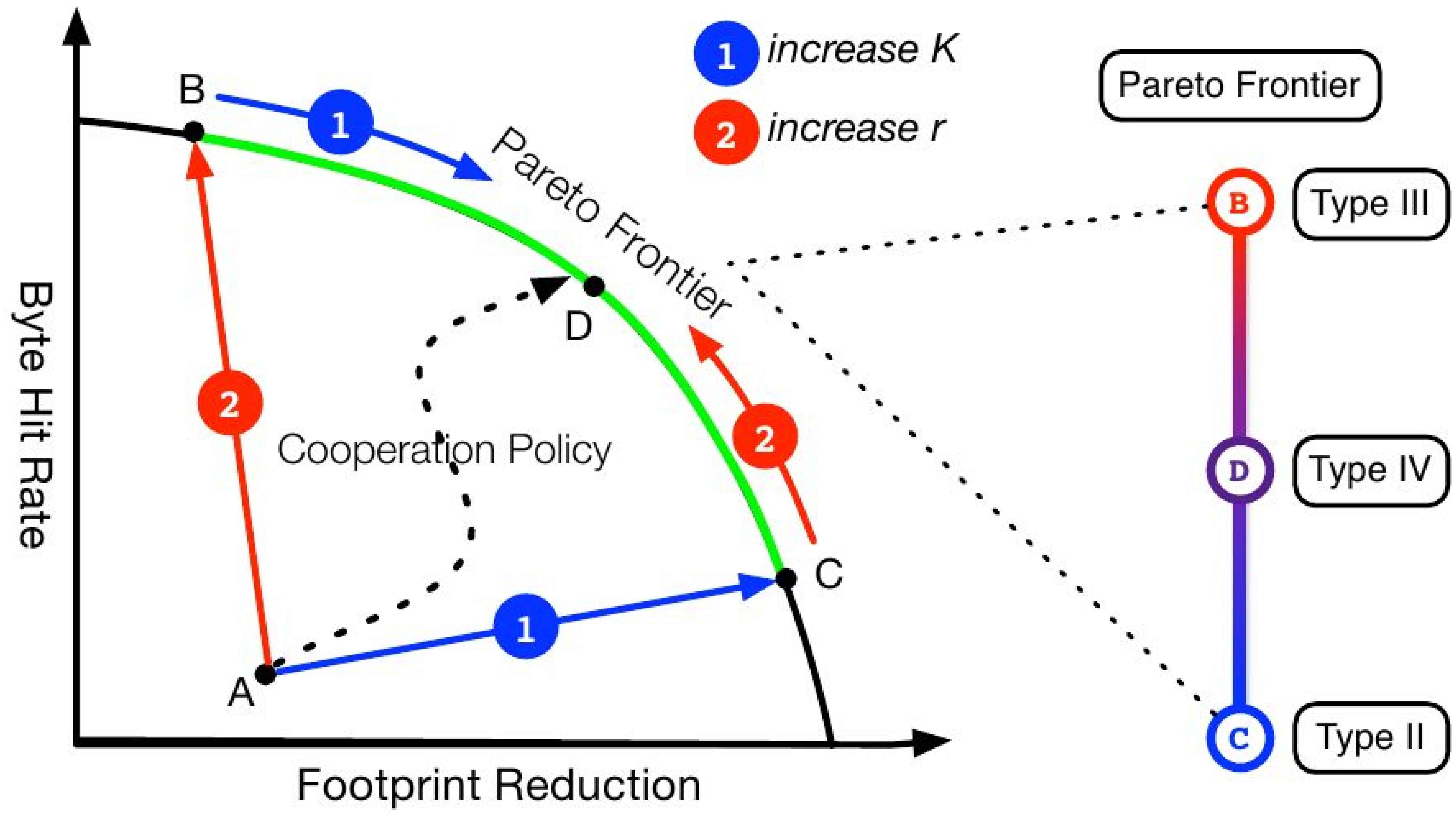
Model of Collaboration

(K, r)-Collaboration Model

- r is the maximum search radius of a given node, it uniquely defines a neighborhood of collaboration. I.e., the range of collaboration.
- K is the maximum number of content replicas in the neighborhood defined by the search radius r. I.e., the tolerance on duplicates.

L. Wang, S. Bayhan, and J. Kangasharju, "Effects of Cooperation Policy and Network Topology on Performance of In-network Caching," IEEE Communication Letters. IEEE, Vol.18, No.4, April 2014.

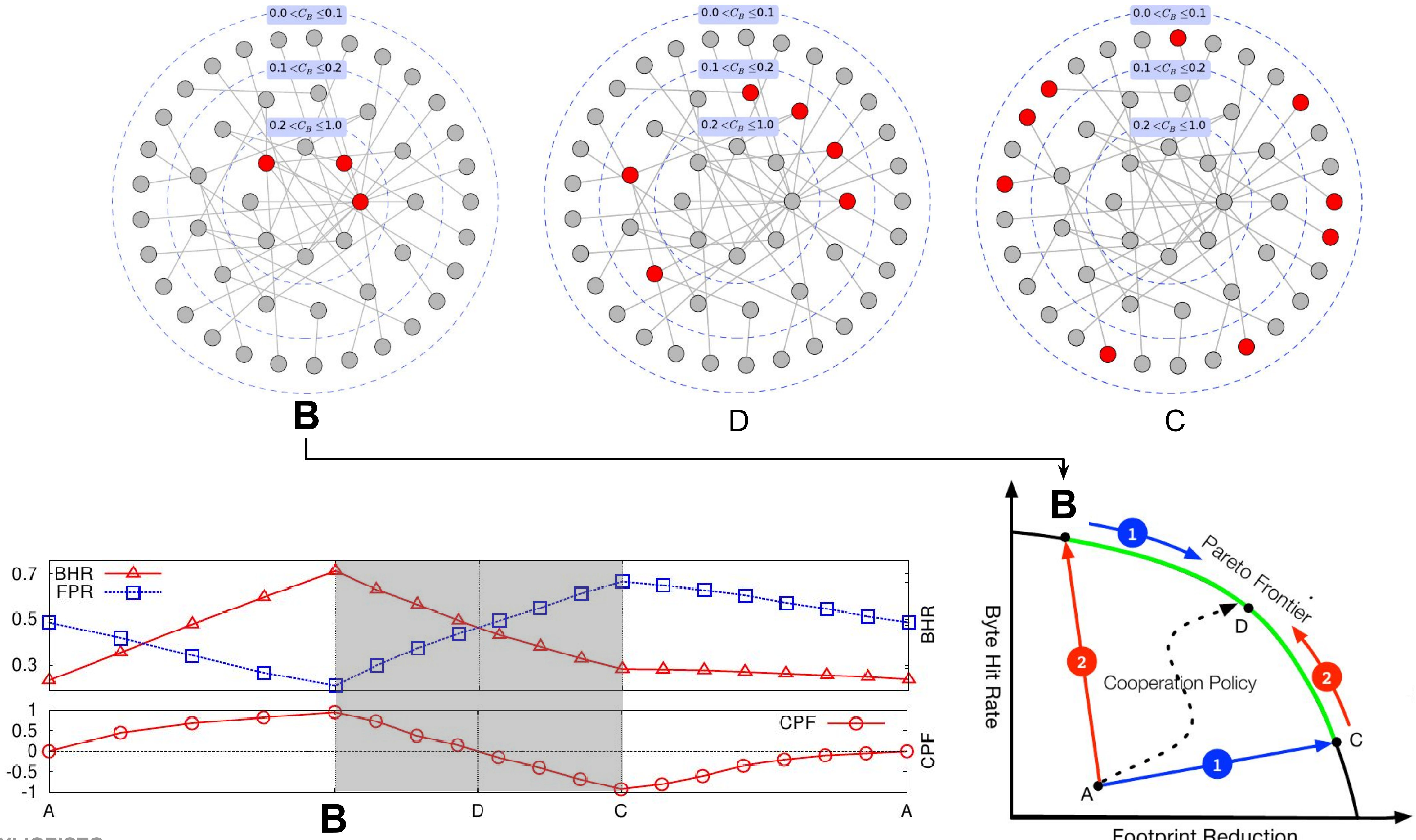
Interplay of Content, Topology and Collaboration



The figure summarizes the cache system behavior by using the metrics and collaboration model we presented in the previous slides.

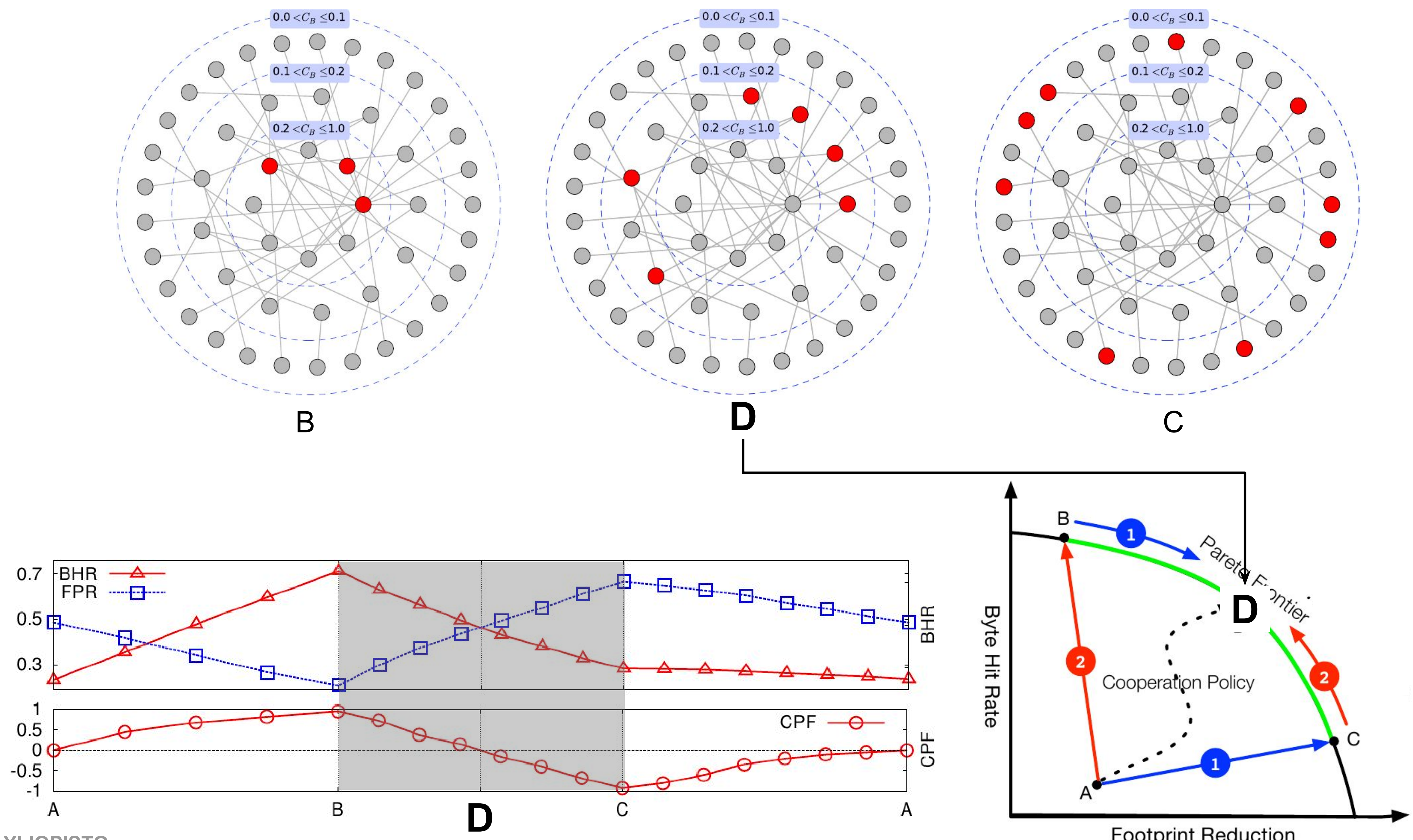
Interplay of Content, Topology and Collaboration

Move along the Pareto frontier (B→D→C) where CPF varies in (-1, +1).



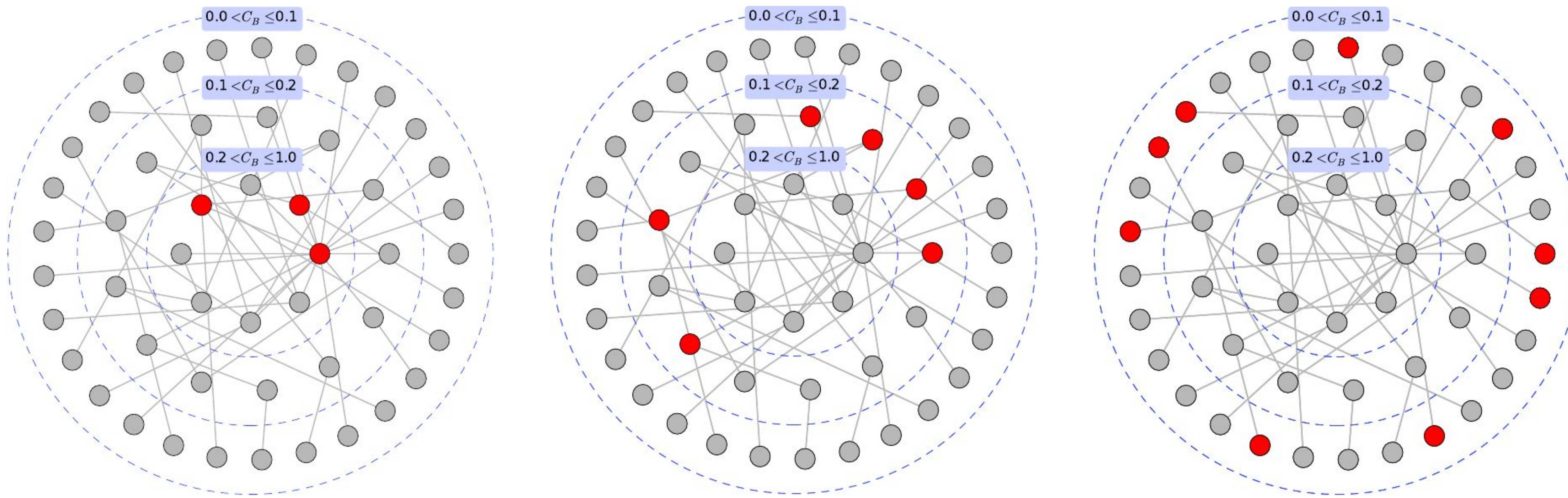
Interplay of Content, Topology and Collaboration

Move along the Pareto frontier (B→D→C) where CPF varies in (-1, +1).

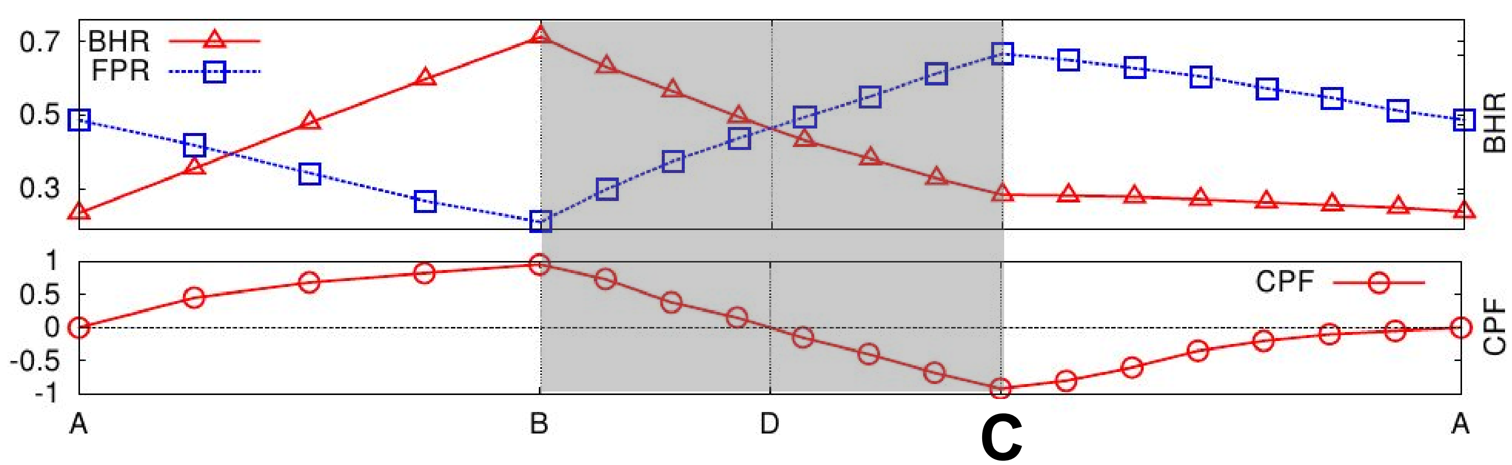


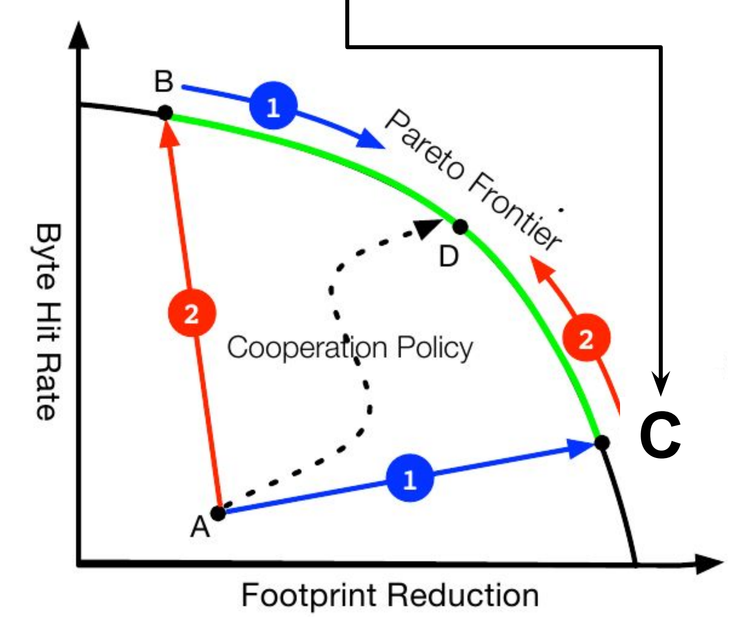
Interplay of Content, Topology and Collaboration

Move along the Pareto frontier (B→D→C) where CPF varies in (-1, +1).



B D C





Basis of Collaboration

- There are potentially infinite Pareto optimal solutions on the frontier, how are we going to find the optimal point "**D**" for collaboration?

- What is the fundamental basis of being collaborative?
  - Incentive of collaboration. We assumed nodes are altruistic.
  - What if selfishness is an inherent and intrinsic characteristic?
  - It is hard to justify that a node would like to sacrifice for others.
  - Fairness is important!

L. Wang and J. Kangasharju, "A Fair Collaborative Game on Cache Networks," in submission.

Fair Game and Nash Bargaining

Core idea: "*A node maximizes the extra benefit from collaboration without degrading its own performance.*"

- Fair caching game is formulated in Nash Bargaining framework.
- Axiomatic game theory, agnostic about negotiation mechanisms.
- Three well-defined fairness: Egalitarian, Max-min and Proportional.

Definition 3. *An in-network caching game is a tuple* (Ω, u^0), *where* $\Omega \subset \mathbb{R}^{|V|}$ *contains all the utility values obtainable via collaboration,* $u^0 \subset \mathbb{R}^{|V|}$ *contains all the disagreement values leading to a negotiation breakdown.*

Definition 4. *A fair collaborative game is a game* (Ω, u^0) *with Nash bargaining solution, namely a function* $f : \Omega^e \rightarrow \Psi$ *such that* $f(\Omega, u^0) = (\mathbf{x}, \mathbf{y})$ *uniquely maximizes* $\prod_{v_i \in V}(U_i - u_i^0)$.

Analysis of Collaboration Cost

Collaboration cost on general topologies

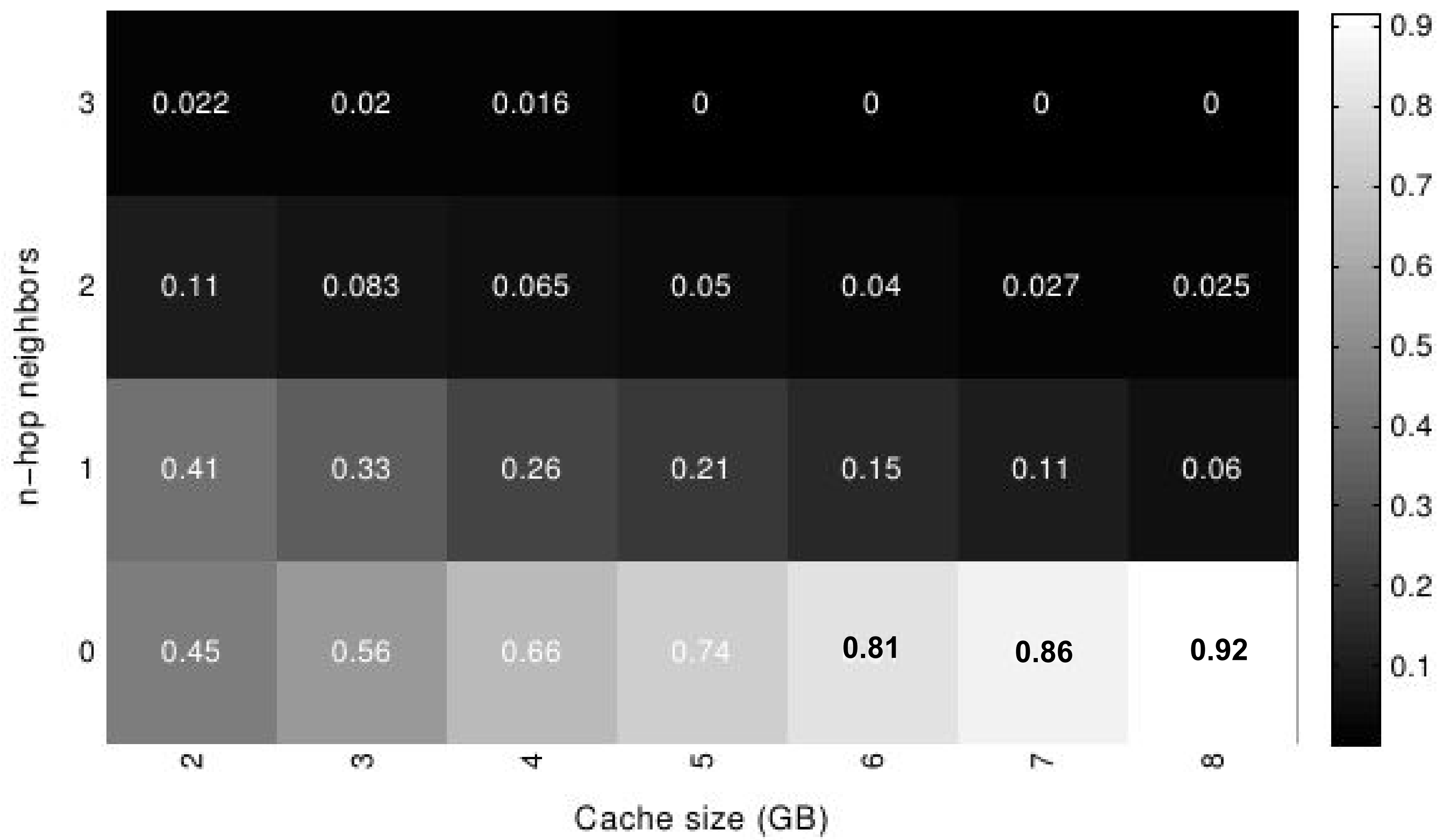
- The cost is measured in terms of number of exchanged messages.
- The cost grows exponentially when the search radius increases.
- Collaboration has to be restricted within a very small neighborhood to keep the cost reasonable.

Theorem 3. *In a random network $G = (V, p)$ where nodes have average search radius r, the induced system overhead $\Delta_r^{r+1}\Phi$ by increasing the average search radius by 1 equals*

$$\Delta_r^{r+1}\Phi = \theta \times |V| \times \left[\frac{z_2}{z_1}\right]^r \times z_1 \qquad (16)$$

Collaboration Localization

- Is collaboration doomed due to its cost? Fortunately, collaboration localization.
- Further investigation strongly indicates the collaboration is highly localized in a small neighborhood due to the highly skewed content popularity distribution.



Summary

The key findings of this thesis can be summarized as below:

- A cache network is fundamentally different from a single cache. To gain a holistic view, measurement metrics must be carefully designed.
- Collaboration in ICN can be modeled with (K,r)-Collaboration Model.
- There are potentially infinite Pareto optimal solutions on a non-trivial topology, with different balance on the intra- and inter-ISP traffic.
- Fairness is the basis of collaboration since nodes can be selfish. This thesis proposed and analyzed the fair in-network caching solution.
- The collaboration on general topologies is costly, but most of the gain can be obtained from a small neighborhood due to localization.

Thank You!

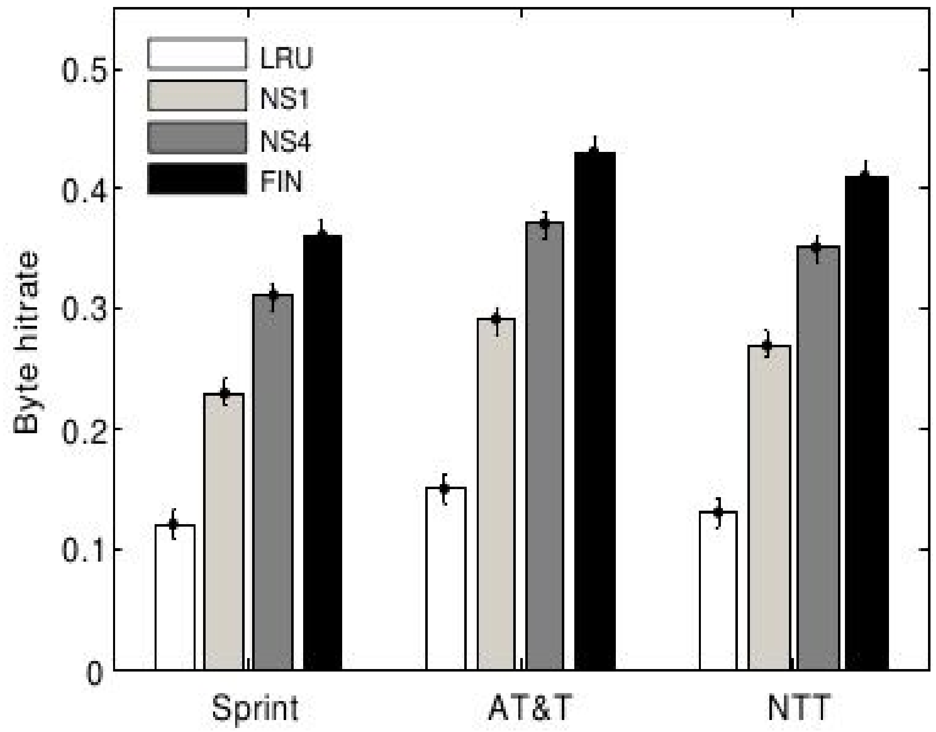
Rationale Behind K and r

(K, r)-Collaboration Model is simple yet expressive.

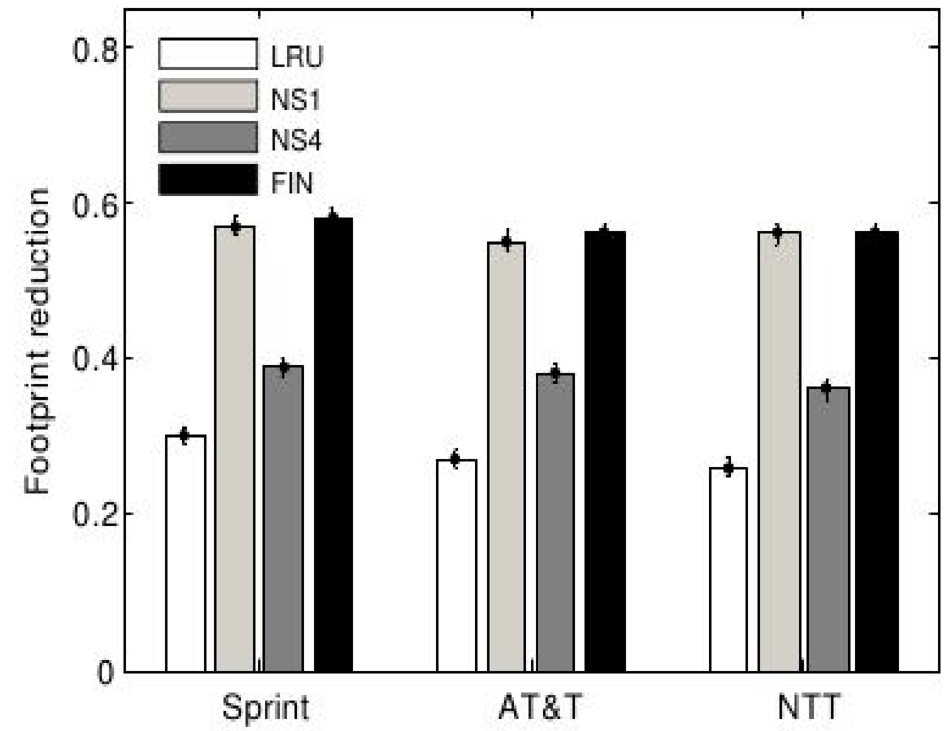
The rationale behind K and r

- We can always characterize collaboration by its search strength (r) and capability of reducing duplicates (K).
- In other words, the capability of discovering content (r) and the capability of utilizing cache efficiently (K).
- K and r represent the tradeoff between BHR and FPR.
- Collaboration model has direct impacts the content distribution in the network.

Performance of Fair Caching Algorithm



Byte hit rate comparison, 4GB cache.



Footprint reduction comparison, 4GB cache.

Content Discovery and Delivery

Namely, how the content is addressed and how the query is routed.

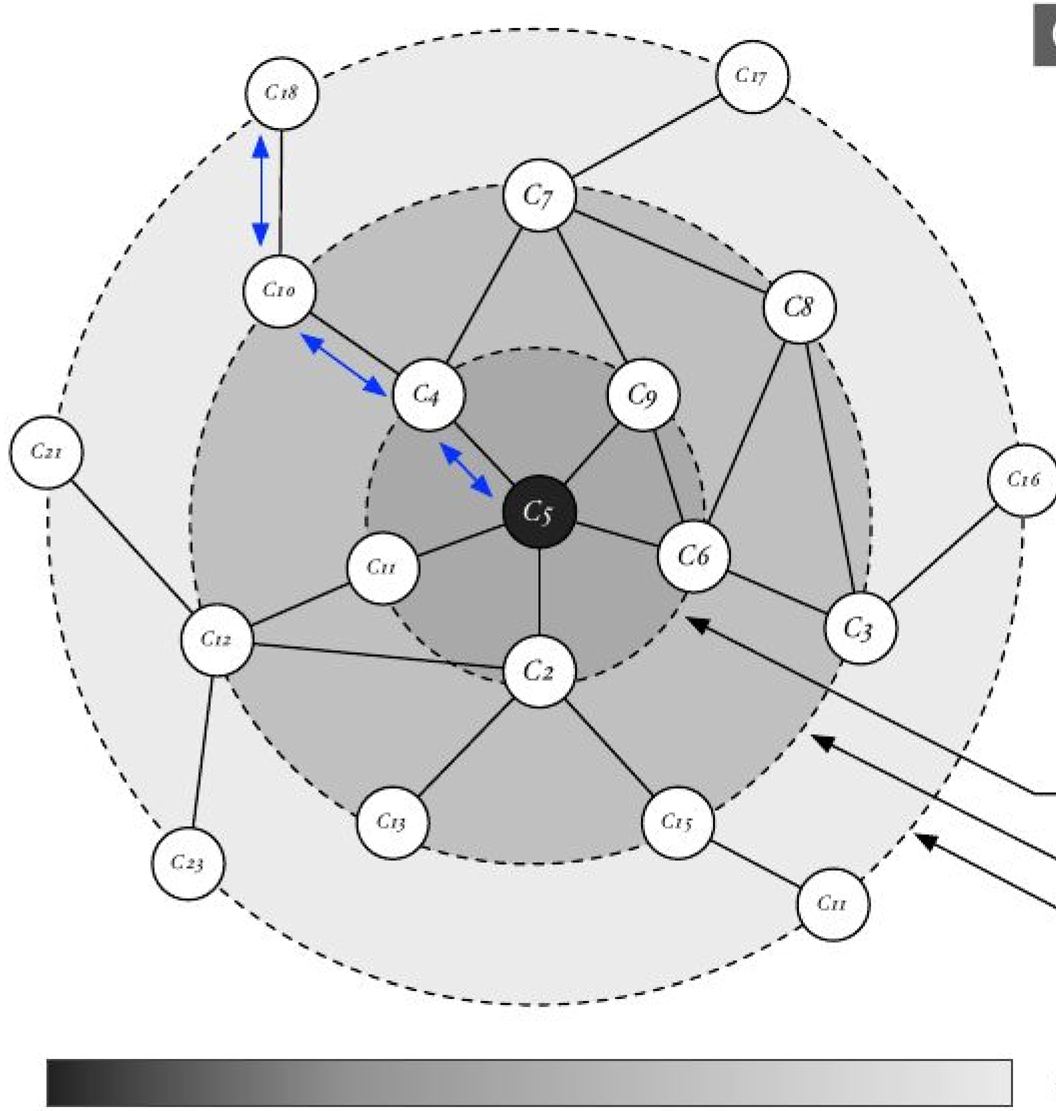
- Probabilistic solution.
- Deterministic solution.

Some examples:

- En-route discovery, simply forward the request to the next hop along the path. Lowest complexity, but no guarantees on finding the content.
- Explicitly exchanging information, the communication cost can be reduced by using compression like Bloom Filter and etc.
- Distributed Hash Table, global name resolution service and etc.

L. Wang, O. Waltari, and J. Kangasharju, "MobiCCN: Mobility Support with Greedy Routing in Content-Centric Networks," in the Proceedings of IEEE Globecom. IEEE, December 9-13 2013.

A Graphical Illustration of the Model



Output Caching strategy (x, y) consists of two parts

1. x : caching decision, what to cache
2. y : retrieving decision, where to retrieve

E.g. $x_{18,3} = 1, y_{5,18,3} = 1$ indicates C_{18} caches O_3, and C_5 retrieves O_3 from O_{18} ⟷

Input Topology G, nodes C, content O

N_i^+ defines who has C_i in its neighborhood

$C_5 \in N_6^+, \quad C_5 \in N_9^+, \quad C_5 \in N_{10}^+, \quad C_5 \in N_{11}^+ \quad ......$

N_i defines C_i has whom in its neighborhood

C_5 1-hop neighbors z_1
C_5 2-hop neighbors z_2
C_5 3-hop neighbors z_3
......

} C_5 neighborhood N_5

Content localization: fraction of retrieved content