Cooperation Policies for Efficient In-Network Caching

Liang Wang
University of Helsinki
liang.wang@cs.helsinki.fi

Suzan Bayhan
HIIT, Aalto University
bayhan@hiit.fi

Jussi Kangasharju
University of Helsinki
jakangas@cs.helsinki.fi

ABSTRACT

Caching is a key component of information-centric networking, but most of the work in the area focuses on simple en-route caching with limited cooperation between the caches. In this paper we model cache cooperation under a game theoretical framework and show how cache cooperation policy can allow the system to converge to a Pareto optimal configuration. Our work shows how cooperation impacts network caching performance and how it takes advantage of the structural properties of the underlying network.

Categories and Subject Descriptors

C.2.2 [**Communication Networks**]: Network protocols

Keywords

In-network caching, Game theory, Cooperative caching

1. MOTIVATION

In-network caching is an indispensable component of information-centric networking (ICN). However, managing a group of networked caches is a big challenge. Most previous work on caching in ICN focuses on simple en-route caching [1,3,5], with limited cooperation among the routers. Lack of cooperation engenders redundancy and degrades efficiency of network storage use. In [5] we showed that cooperation efficiently reduces duplicates and boosts performance, however at the cost of (some) increase in network traffic.

Game theory has been used to analyze distributed caching systems [2]. In [4] caching is modelled as a replication game on graphs, showing the existence of Nash equilibrium on arbitrary topologies. However, the impact of network structure on cooperation has been overlooked. Our work focuses on the relationship between cooperation, topological properties, and caching performance. With an optimization model, we find the Pareto frontier of network caching performance and show the area which a cooperation policy can explore.

2. SYSTEM MODEL

We use two caching performance metrics: Byte Hit Rate (BHR) and Footprint Reduction (FPR) [5]. BHR measures the reduction in traffic going outside an ISP and FPR measures the reduction in traffic within an ISP.

2.1 Caching as a Game

Let $\mathcal{G} = (V, E)$ represent the ISP network where $V = \{v_i\}$ is the set of routers with cache capacities C_i, and E is the links between the routers. We have a content set $\mathcal{O} = \{o\}$, and a demand matrix $W = \{w_i^o\}$, w_i^o indicates the demand for object o on v_i. Since a router makes caching decisions locally and they are also affected by the content population in other routers, we model the caching decision as a non-cooperative game. In our model, a router v_i serves a request for a content item either from local storage, or by retrieving it from another router v_j within its r-hop neighborhood if v_j stores it, or by retrieving it from the content provider (CP). Let *searchable set* $\mathcal{N}_i^r$ be the set of routers within the r-hop neighborhood of v_i. We refer to r as the *search radius*.

Searchable set depends on $\mathcal{G}$'s topological properties so we allow each v_i to select its own r. $\mathcal{L}_i$ is v_i's cache configuration representing the content set stored locally at v_i. In the game theory context, $\mathcal{L}_i$ is the action of v_i, and is *feasible* if both cache capacity and maximum replica constraints are met. The former ensures that a cache cannot store more objects than it has storage capacity and the latter aims to decrease redundancy and favor more efficient cache space usage. We let each item to be stored in at most N_{copy} routers in the network. However, as v_i is aware of the contents only in $\mathcal{N}_i^r$, we modify the maximum replica constraint for o as follows: $\sum_{v_j \in \mathcal{N}_i^r} 1_{[o \in \mathcal{L}_j]} \leqslant \left\lceil \frac{|\mathcal{N}_i^r|}{|V|} N_{copy} \right\rceil$ where $1_{[x]}$ is the indicator function taking value 1 if x is true.

The cache profile of the whole network is the aggregation of all individual cache configurations, $\mathcal{L} = (\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2...)$. $\mathcal{L}_{-i}$ is defined as all other routers except v_i sticking to configurations in $\mathcal{L}$. Let $\mathcal{H}_i = \bigcup_{\forall v_j \in \mathcal{N}_i^r} \mathcal{L}_j$ be the content set that v_i can retrieve from its neighborhood, and $\mathcal{R}_i = \mathcal{O} \setminus (\mathcal{L}_i \bigcup \mathcal{H}_i)$ be the set of contents that are not stored at v_i or in $\mathcal{N}_i^r$. We assume caching an object locally induces no cost, but retrieving it remotely does and denote the cost by α_i^o. We define α_i^o as the length of the shortest path between v_i and the nearest copy of o it can find in $\mathcal{N}_i^r$. Total cost on v_i under cache configuration $(\mathcal{L}_i, \mathcal{L}_{-i})$ is:

$$cost_i(\mathcal{L}_i, \mathcal{L}_{-i}) = \sum_{\forall o \in \{\mathcal{H}_i \setminus \mathcal{L}_i\}} w_i^o \alpha_i^o + \sum_{\forall o \in \mathcal{R}_i} w_i^o \beta_i^o \quad (1)$$
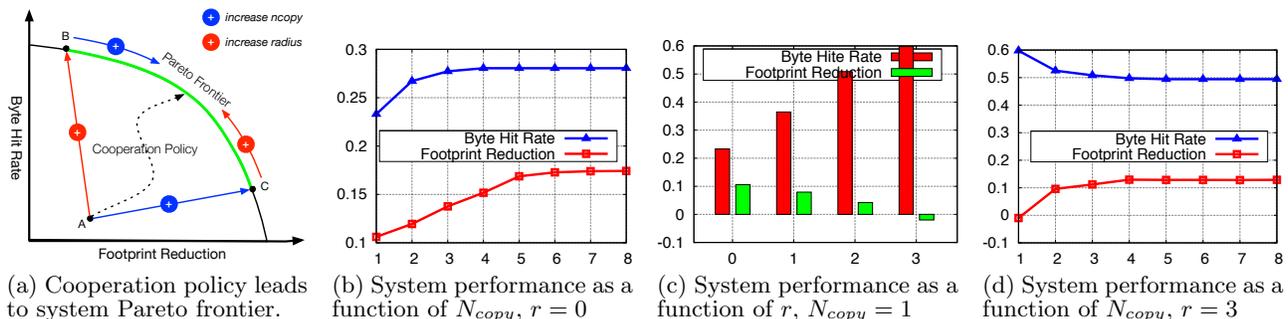
(a) Cooperation policy leads to system Pareto frontier.

(b) System performance as a function of N_{copy}, $r = 0$

(c) System performance as a function of r, $N_{copy} = 1$

(d) System performance as a function of N_{copy}, $r = 3$

Figure 1: Performance evaluation on a 4-level 2-tree topology. Clients are attached to the leaves.

where β_i^o is the length of the shortest path between v_i and the CP that stores o. A router gets benefit by serving requests both from its directly attached clients and requests it receives from its neighbor routers. Let us define $\widetilde{\mathcal{N}_i} = \{v_j | \forall v_j \in V, v_i \in \mathcal{N}_j^r\}$ as the set of routers who have v_i in their searchable sets. The benefit of v_i is as follows:

$$benefit_i(\mathcal{L}_i, \mathcal{L}_{-i}) = \sum_{\forall o \in \mathcal{L}_i} (w_i^o + \sum_{\forall v_j \in \widetilde{\mathcal{N}_i}} w_j^o) \qquad (2)$$

Given $\mathcal{L}_{-i}$, v_i tries to maximize its own payoff which is defined as the difference between benefit and cost, and calculated as follows:

$$u_i(\mathcal{L}_i, \mathcal{L}_{-i}) = benefit_i(\mathcal{L}_i, \mathcal{L}_{-i}) - cost_i(\mathcal{L}_i, \mathcal{L}_{-i}) \qquad (3)$$

subject to the aforementioned feasibility constraints.

A router's best response is given by $\max_{L_i} u_i(\mathcal{L}_i, \mathcal{L}_{-i})$. However, due to other routers making caching decisions simultaneously, the network state, i.e., cache configurations at each v_i, may not be stable. Therefore, we need to investigate if the system has a stable point (Nash equilibrium) in which no router has incentive to change its strategy given others stick with $\mathcal{L}_{-i}$. Also, convergence time is of interest.

2.2 Social Optimal Cache Configuration

To quantify the limits of the sub-optimality due to routers' limited knowledge, we compute the optimal performance by designing the *social optimal* cache configuration. We define the *social optimal configuration* $\mathcal{L}^{opt}$ as the set of all cache configurations that lead to the network-wide maximum net benefit, and derive it by solving the following optimization problem: $\max \sum_{\forall v_i \in V} u_i(\mathcal{L}_i, \mathcal{L}_{-i})$. Using this upper bound, we can calculate the Price of Anarchy (PoA) which shows the performance loss in case of sub-optimal cache configurations.

3. PRELIMINARY EVALUATION

Figure 1 shows how N_{copy} and r impact the performance of caching. Even though the optimization model can calculate the optimal cache profile for given N_{copy} and r (e.g., point A in Figure 1a), it does not indicate what are the best values for these two parameters. In other words, we can still achieve better performance by tuning N_{copy} and r, because system may be underutilized. However, the optimization model can be used to find Pareto frontier of the performance (green arc BC in Figure 1). When we reach the Pareto frontier, we cannot improve either of BHR or FPR without hurting the other. The fan-shaped area defined by ABC is the area which a cooperation policy can explore to find the best trade-off between the N_{copy} and r. Note that

the lines AB and AC are not parallel to the x- and y-axis in Figure 1a, since changing either of r or N_{copy} affects both BHR and FPR, as we show below.

Figures 1b, 1c and 1d show how performance changes along the segments AB, AC and arc BC by increasing r and N_{copy} respectively. In Figure 1b, we increase N_{copy} by fixing $r = 0$. We can see both BHR and FPR improving along the line AC in Figure 1a. Since we set $r = 0$, this is equivalent to en-route caching, where allowing some duplicates on disjoint paths will clearly improve performance. In Figure 1c, we increase r by fixing $N_{copy} = 1$. As we move from A towards B along the line AB by increasing r, we see that BHR increases and FPR drops, even to negative at the extreme, indicating that the search traffic overhead is larger than the traffic reduction gains of caching. The reason is that as search radius increases, a node's neighborhood grows which will increase BHR significantly, but since we allow only one copy in the network, the request needs to be redirected to other nodes which degrades FPR. Figure 1d shows how performance changes along the Pareto frontier after we reach point B.

4. FUTURE DIRECTIONS

In the future work, we first explore the possibility of discovering a Nash equilibrium of our game model, and find out how close the performance in the steady state is when compared to the Pareto frontier by calculating the Price of Anarchy. We also plan to study how the topological structures impact the cooperation policy and performance. (e.g. how hetero-radius balances between performance gain and search cost based on the router's centrality, degree and etc.) These results can be used to design better distributed heuristics for making caching decisions, while taking fully into account the topological properties of the underlying network.

5. REFERENCES

[1] K. Cho, M. Lee, K. Park, T. Kwon, Y. Choi, and S. Pack. Wave: Popularity-based and collaborative in-network caching for content-oriented networks. In *Infocom workshop on Name-Oriented Networking*, pages 316–321, 2012.

[2] B.-G. Chun, K. Chaudhuri, H. Wee, M. Barreno, C. H. Papadimitriou, and J. Kubiatowicz. Selfish caching in distributed systems: a game-theoretic analysis. In *Proceedings of ACM PODC*, 2004.

[3] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard. Networking named content. In *Proceedings of ACM CoNEXT*, 2009.

[4] V. Pacifici and G. Dán. Selfish content replication on graphs. In *Proceedings of International Teletraffic Congress*, 2011.

[5] W. Wong, L. Wang, and J. Kangasharju. Neighborhood search and admission control in cooperative caching networks. In *Proceedings of IEEE Globecom*, 2012.