Computer Networks 103 (2016) 129–142

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

On fair network cache allocation to content providers

Sahar Hoteit[a,*], Mahmoud El Chamie[b], Damien Saucez[c], Stefano Secci[d]

[a] INRIA-Saclay, Palaiseau 91192, France
[b] Department of Aeronautics and Astronautics, University of Washington, Seattle, WA 98195, USA
[c] Inria Sophia Antipolis, 2004 Route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex, France
[d] Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005 Paris, France

ARTICLE INFO

Article history:
Received 2 November 2015
Revised 22 March 2016
Accepted 7 April 2016
Available online 12 April 2016

Keywords:
In-network caching
Information Centric Networking
Mechanism design
Game theory
Cache allocation

ABSTRACT

In-network caching is an important solution for content offloading from content service providers. However despite a rather high maturation in the definition of caching techniques, minor attention has been given to the strategic interaction among the multiple content providers. Situations involving multiple content providers (CPs) and one Internet Service Provider (ISP) having to give them access to its caches are prone to high cache contention, in particular at the appealing topology cross-points. While available cache contention situations from the literature were solved by considering each storage as one autonomous and self managed cache, we propose in this paper to address this contention situation by segmenting the storage on a per-content provider basis (e.g., each CP receives a portion of the storage space depending on its storage demand). We propose a resource allocation and pricing framework to support the network cache provider in the cache allocation to multiple CPs, for situations where CPs have heterogeneous sets of files and untruthful demands need to be avoided. As cache imputations to CPs need to be fair and robust against overclaiming, we evaluate common proportional and max–min fairness (PF, MMF) allocation rules, as well as two coalitional game rules, the Nucleolus and the Shapley value. When comparing our cache allocation algorithm for the different allocation rules with the naive least-recently-used-based cache allocation approach, we find that the latter provides proportional fairness. Moreover, the game-theoretic rules outperform in terms of content access latency the naive cache allocation approach as well as PF and MMF approaches, while sitting in between PF and MMF in terms of fairness. Furthermore, we show that our pricing scheme encourages the CPs to declare their truthful demands by maximizing their utilities for real declarations.

1. Introduction

With the advent of broadband and social networks, the Internet became a worldwide content delivery platform [1,2], with high bandwidth and low latency requirements. To meet the always increasing demand, contents are pushed as close as possible to their consumers and *content providers* (CP) install dedicated storage servers directly in the core of Internet Service Provider (ISP) networks [3]. However, the TCP/IP protocol suite uses a conversational mode of communication between hosts that can be considered not appropriate for content delivery [2]. Therefore, a complex machinery is developed (around the Domain Name System, DNS, protocol and the HyperText Transfer Protocol, HTTP) to compensate the limitations of the TCP/IP protocol suite. Conscious of the mismatch between the network usage and its conception, the research community recently proposed the concept of in-network caching (e.g., Information Centric Networking (ICN) [2,4]). For instance, in ICN, content objects can be accessed and delivered natively by the network according to their name rather than relying on IP addresses [2]. Hence, this technology removes the concept of location or topology from communication primitives and uses the notion of contents and their name instead. These contents can therefore be found potentially anywhere in the network, moved or replicated at different locations [4–6].

ISP networks then become native distributed storage systems, i.e., network cache providers that can directly sell caching capabilities to content providers instead of hosting their servers. However, it is most probable that the storage demand exceeds the total ISP storage offer, at least for the content caching locations the closest to the users. So far, the contention is solved by considering each storage as one autonomous and self managed cache (e.g., using a LRU, least-recently-used, mechanism), as depicted in the rightmost part of Fig. 1. With this approach CPs are unable to provision their own infrastructure accurately as they cannot predict what contents

* Corresponding author.
E-mail addresses: sahar.hoteit@inria.fr (S. Hoteit), melchami@uw.edu (M. El Chamie), damien.saucez@inria.fr (D. Saucez), stefano.secci@upmc.fr (S. Secci).

http://dx.doi.org/10.1016/j.comnet.2016.04.006

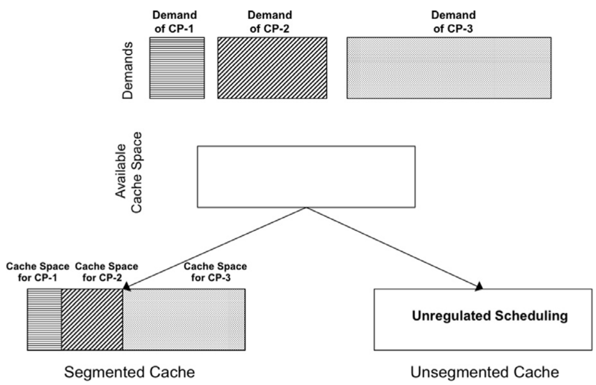

Fig. 1. Representation of segmented and unsegmented caches with many content providers (CPs).

will be cached by the ISP as it depends on the workload of the other CPs using the ISP infrastructure.

In this paper, we propose to address this contention situation by segmenting the storage on a per-content provider basis, as depicted in the leftmost part of Fig. 1. It is worth mentioning that, to the best of our knowledge, this is the first work in the literature to propose such a partitioning of the ISP cache to CPs. Thus the viability of the model is not yet investigated. However, we believe it would be rather straightforward to deploy, at least from a technical standpoint. Maybe it would be less straightforward from an ICN protocol design standpoint, if for some reasons it can be useful to compute and disseminate the content cache segmentation to other ICN nodes in a distributed fashion. We believe these details are implementation-specific and do not warn against the validity of our work.

Hence in our proposition, each content provider receives a portion of the storage space depending on its storage demand. For this, based on application of results in economics and game theory to the target problem, we propose a *2-step mechanism design* [7,8] that computes a fair and rational sharing of resources between CPs. The first step relies on a content cache allocation algorithm where, as a function of content cache demands coming from CPs, the network cache provider decides the *imputation* of cache spaces to CPs. The second step uses a predefined *payment rule* by auctions to decide the selling price of the storage unit in the network; its purpose is to prevent content providers from lying about their true demands. The paper is organized as follows. Section 2 presents an overview of related works. In Section 3, we analytically introduce the context of our work: Section 3.2 presents the resource allocation problem by modeling it as a cooperative game, and Section 3.4 develops our pricing scheme based on mechanism design theory. Section 4 presents the implementation of our proposed pricing scheme for the different cache imputations. Section 5 compares the proposed cache allocation rules with other schemes. Finally, Section 6 concludes the paper.

2. Background

Several researches have recently proposed various cache allocation solutions. Rossi and Rossini compare the in-network caching performance in homogeneous (i.e., where the routers have the same overall cache size) and heterogeneous cache deployments (i.e., where the routers have not the same cache size) [9]. In the latter case, they propose to allocate cache capacity proportionally to the router centrality metric measured according to different criteria: degree, stress, betweenness, closeness, graph, and eccentricity centrality. Authors of both [9] and [10] show that allocating cache capacity across the network in a heterogeneous manner slightly improves network performance compared to the homogeneous manner; however, the benefits of heterogeneous deployments become apparent with larger networks (e.g., more than 100 nodes). Moreover, Wang et al. study the influence of content popularity distribution on network performance showing that (i) for uniformly distributed content demands (e.g., catch-up TV), pushing caches into the core yields better performance while (ii) highly skewed popularity request patterns (e.g., YouTube, mobile VoD system or Vimeo) are better served by edge caching [10]. This latter point is confirmed by Fayazbakhsh et al. [11].

Recently, there has been significant interest in applying game theory to the analysis of communication networks, with the aim to identify rational strategic solutions for multiple decision-maker situations. Indeed, as opposed to mono-decision maker problems, game-theoretic approaches adopt a multi-agent perspective to account for different objective functions and counter objections to rationally non-justified solutions [12].

Thus far, many papers from the literature have tackled game-theoretic approaches for cache allocation using non-cooperative game theory. These papers consider servers or routers or networks as selfish entities seeking to maximize their own profit at the expense of globally optimum behavior. For example Pacifici and Dan study a non-cooperative game to characterize the problem of replication of contents by a set of selfish routers aiming to minimize their own costs [13]. In the same context, Chun et al. characterize the caching problem among selfish servers using a non-cooperative game [14]. For each content in the network, selfish servers have two possible actions: either caching the content if all its replicas are located too far away or not caching it if one of its replicas is located at a nearby node. As in [13], they show the existence of pure strategy Nash equilibrium of the caching game.

Motivated by the intuition that forms of collaboration between different network cache providers could yield an enhancement in network performance, some of the papers have tackled cache allocation problem using cooperative game theory. For instance, authors in [15] propose a cooperative game whereby the routers behave as rational agents that seek to minimize their aggregate content access cost. Going beyond routers, Saucez et al. [16] describe how content providers could shape their content access prices and discounts to favor the emergence of cache space distribution overlays across independent networks, toward the formation of incentive-prone overlay equilibria.

Under the similar rationale of collaboration between different content providers, yet a broader context, in this paper we focus on cooperative game theory. We investigate how the network cache provider, modeling CPs as players of a cooperative game, can design a cache allocation framework so that cache imputations to CPs are strategically fair and robust against cache space over-claiming, while outperforming legacy approaches in terms of content access latency. Up to our knowledge, there are no other works precisely addressing this problem, despite the above-cited works do share similar concerns in cache allocation and component sharing but do no tackle the over-claiming cache space issue. As detailed in the following, we propose various cache allocation rules, including coalitional game theory rules for *bankruptcy* situations [17] to solve the atomic cache contention problem, motivated by the fact that a similar algorithmic approach has shown high performances in strategic shared spectrum allocation problems [18,19].

3. Cache allocation framework and rules

In the context of a network cache provider, the cache capacity is used to host content files in order to enhance users' quality of experience by decreasing content access latency. Assuming contents are owned by external CPs, the network cache provider