

Impact of Revenue-Driven CDN on the Competition among Network Operators

Patrick Maillé
Telecom Bretagne, France
patrick.maill@telecom-bretagne.eu

Gwendal Simon
Telecom Bretagne, France
gwendal.simon@telecom-bretagne.eu

Bruno Tuffin
Inria, France
bruno.tuffin@inria.fr

Abstract—We investigate the impact of decisions made by a CDN willing to maximize its revenue through the management of cache servers. Based on a model with two network providers, we highlight that revenue-oriented management policies can affect the user-perceived quality of experience, impacting the competition among network access providers in favor of the largest one. Since this contradicts the principle underpinning network neutrality—although not with the technical net neutrality rules—we discuss the necessity to regulate CDN activity.

I. INTRODUCTION

Content Delivery Network (CDN) providers have become key actors in the value chain of Internet delivery: both network operators and content providers benefit from CDNs to improve the satisfaction of their customers. Indeed, the presence of CDN *edge servers*, which store content and serve end-users at the edges of the network, is expected to improve the Quality of Experience (QoE), and consequently the engagement, of the end-users [6]. An abundant literature is related to the performances of CDNs; a less-studied implication of the presence of CDN edge servers deep in the network of Internet Service Providers (ISPs) is that ISPs and content providers can leverage the CDN to differentiate their services from their competitors. The CDN providers being revenue-driven actors, it is indeed possible to pay to get a better service or to degrade the Quality of Service (QoS) of competitors. In [8], we showed that a dominant Content Provider (CP) can pay to harm a smaller competitor. In the current paper, we focus on the impact that CDNs have on the competition between network operators.

We study the CDN behavior from an economic standpoint. We first describe a topology that we will use throughout the paper, with two ISPs and a CDN which can place its servers close to users in the ISPs' networks or at an intermediate level before reaching the ISPs. We analyze the best (revenue-maximizing) caching strategies and show that depending on popularities, the quality perceived by users of an ISP can be affected by those caching strategies. Then based on VoD traces, we illustrate the practical relevance of the analysis and show that an economic differentiation can (again) be at the expense of some users through degraded QoE, affecting the competition among ISPs. Finally we highlight the need to include CDNs in the network neutrality debate.

This work has been partially funded by Orange, however the ideas developed here solely reflect the authors' vision.

II. MODEL

We consider a CDN, which delivers content on behalf of a set of CPs to a population of end-users spread into k distinct networks (owned by distinct ISPs). For the sake of simplicity, we consider $k = 2$ in the following (and index ISPs with A and B). The model can be extended to more ISPs.

The CDN includes one *intermediate server* (noted as I), which is shared by both network operators, and optionally an *edge server* in each ISP (noted as A and B , for ISP_A and ISP_B respectively). Various forms of agreement between CDNs and ISPs exist in reality for the management of an edge server. For example the CDN sometimes pays a fee to the ISP for the rental of a rack in an ISP's colocation center [3]. Another typical agreement is based on resource exchanges in a federation of ISP-CDNs [10]. We consider here a situation close to the latter example where the CDN pays the ISP for the rental of storage resources in privileged network locations.

When an end-user, regardless of the network she belongs to, requests some content, she can be served either by the edge server of her ISP (if any), or by the shared intermediate server, or, if none of them store the content, by the origin server. In the latter case, the CDN *cache miss* results in a drop of the QoE. Note that an end-user from a given ISP cannot be served by the edge server of another ISP. See Figure 1 for an illustration.

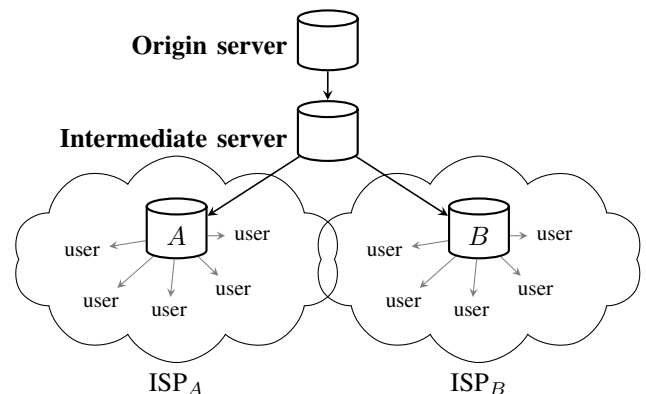


Fig. 1. CDN Topology with two ISPs, an intermediate server, and two edge servers A and B .

A. Terminology and Notations

We use the same model as in [8]. Details are below.

Content popularities. Each piece of content has two popularity values, x_A and x_B , where x_j represents the expected number of downloads from population $j \in \{A, B\}$ per time unit for that specific content.

Service prices. The CDN charges the CPs at higher rates when content is accessed from a CDN server closer to users. We consider fixed prices per unit of content (and per download), and without loss of generality we assume that the service is free for content accessed from the origin server. This leaves three prices p_A , p_B , and p_I , which are assumed given since we focus on the cache management.

Transit costs. We denote by q_I the transit cost per content unit from the origin server to the intermediate server, and by q_j the transit cost per content unit from the origin server to the edge server j ($j = A, B$). Those transit costs are borne by the CDN, hence an incentive to cache popular content to avoid undergoing those costs for each download.

Revenue per download. Based on the service price and the transit cost, we denote by r_j , ($j = I, A, B$) the extra revenue for the CDN per content served from cache j , as a comparison to a CDN cache miss (service from the origin server). In addition to the received payment p_j , the CDN also saves q_j in transit costs, hence $r_j = p_j + q_j$.

Storage costs or limits. We consider three alternatives to set the capacities of CDN servers. In each case, we define three non-negative values $\lambda_I, \lambda_A, \lambda_B$ to represent the marginal prices (per time unit) for storage: those values appear when maximizing the CDN revenue under the capacity constraints. To model the capacities and set the storage cost parameters, we encompass several cases, including:

- Fixed, given, capacities for each server, as in [8]. The three parameters are then Lagrangian multipliers (shadow price) for each capacity constraint.
- Capacities agreed with a (convex) cost function for each server. The three parameters are the marginal capacity costs at each CDN location.
- Capacities agreed with a linear function (a particular case of the one above). The parameters are unit capacity prices at each CDN location.

B. Where to Store the Content (if at all)?

As in [8], the objective of the CDN provider is to choose caching strategies maximizing its revenue per time unit. Mathematically, let us denote by

- $S_i(C_i)$ the cost of storage capacity C_i ($i = A, B, I$): if capacity is fixed beforehand to \tilde{C}_i then we have $S_i(x) = \begin{cases} 0 & \text{if } x \leq \tilde{C}_i \\ +\infty & \text{otherwise;} \end{cases}$
- \mathbb{M} a measure representing the distribution of popularities, more precisely for a Lebesgue-measurable subset \mathcal{S} of \mathbb{R}_+^2 , $\mathbb{M}(\mathcal{S})$ is the total volume of content whose popularities belong to \mathcal{S} .

Then, for each piece of content with popularities (x_A, x_B) , the CDN (assumed to know those values) selects cache location(s) $L(x_A, x_B) \subset \{I, A, B\}$ to store it. Also including capacities

in the decisions, the revenue maximization problem consists in finding capacities C_A, C_B, C_I and a correspondence L maximizing gains under the capacity constraints, i.e., in solving

$$\begin{aligned} \max_{L, C_A, C_B, C_I} \int_{x_A, x_B} & x_A r_A \mathbb{1}_{\{A \in L(x_A, x_B)\}} + x_B r_B \mathbb{1}_{\{B \in L(x_A, x_B)\}} \\ & + x_A r_I \mathbb{1}_{\{(I \in L(x_A, x_B)) \cap (A \notin L(x_A, x_B))\}} \\ & + x_B r_I \mathbb{1}_{\{(I \in L(x_A, x_B)) \cap (B \notin L(x_A, x_B))\}} \\ & - (x_A q_A + x_B q_B) d\mathbb{M}(x_A, x_B) \\ & - S_A(C_A) - S_B(C_B) - S_I(C_I) \\ \text{s.t. } \int_{x_A, x_B} & \mathbb{1}_{\{i \in L(x_A, x_B)\}} d\mathbb{M}(x_A, x_B) \leq C_i \quad i \in \{A, B, I\}. \end{aligned}$$

The three constraints in the optimization problem reflect storage capacity limits, and the components in the objective respectively express the gains from serving content from its servers (A, B, I), the transit costs, and the capacity costs. Note that gains and costs are weighted by popularities (since they occur for each request) while capacity constraints only involve volumes.

To solve that optimization problem, one can form the Lagrangian, denoting by λ_A, λ_B , and λ_I the respective multipliers for the constraints. We study the local optimality conditions for L , reasoning on the marginal revenue changes resulting from changes in L around popularity values (x_A, x_B) . There are eight different possibilities for the CDN to handle that piece of content, but the option of storing the content in I, A , and B is never optimal (the content would never be retrieved from I). The remaining seven options are listed in the first column of Table I, where we give the marginal differences in the Lagrangian of the optimization problem that stem from a different handling of that specific content, taking as a reference the decision of not caching anywhere in the CDN.

Where to store	Marginal objective variation
nowhere	0
in I only	$(x_A + x_B)r_I - \lambda_I$
in A only	$x_A r_A - \lambda_A$
in B only	$x_B r_B - \lambda_B$
in I and A	$x_A r_A + x_B r_I - \lambda_I - \lambda_A$
in I and B	$x_A r_I + x_B r_B - \lambda_I - \lambda_B$
in A and B	$x_A r_A + x_B r_B - \lambda_A - \lambda_B$

TABLE I. MARGINAL DIFFERENCES IN THE OBJECTIVE FUNCTION (LAGRANGIAN) DEPENDING ON THE DECISIONS FOR A PIECE OF CONTENT WITH POPULARITY $x = (x_A, x_B)$. (Reference case: the content is not stored close to end-users.)

Note that in Table I we treat λ_I, λ_A and λ_B as fixed: this is true for the linear-cost model for capacity, while for the other models those multipliers are determined by the caching decisions and the capacity constraints. Nevertheless, the reasoning above remains valid since we only consider marginal (infinitesimal) variations.

When $\lambda_I, \lambda_A, \lambda_B$ are known, Table I immediately provides us with the optimal decision for the CDN manager: for each piece of content, pick the option with the largest objective variation. For example, the two first lines in the table indicate that it is preferable to cache some content in I rather than not cache it at all if $(x_A + x_B)r_I - \lambda_I > 0$, or equivalently if $x_A + x_B > \frac{\lambda_I}{r_I}$: indeed, the extra unit revenue for allocating

some storage to that content would be $(x_A + x_B)r_I$ (from payments and economies in transit traffic for each download, hence dependent on popularities) while the unit perceived cost for that space would be λ_I . Figure 2 shows the best decision for any content with popularity (x_A, x_B) , when $\lambda_I = 2, \lambda_A = 4, \lambda_B = 5, r_I = 2, r_A = 2.5, r_B = 3$.

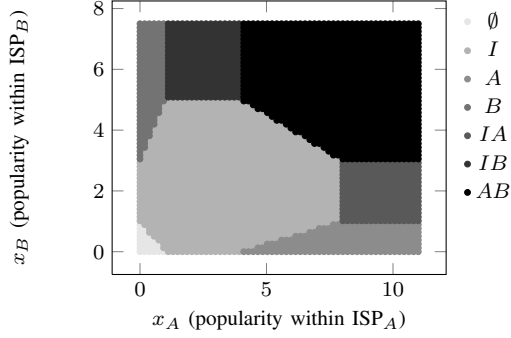


Fig. 2. Optimal caching for three caches

III. AN ASYMMETRIC SCENARIO

We consider now an “asymmetric” scenario where the CDN agrees with only one ISP, say ISP_A , on the rental of storage resources for an edge server A . Since the CDN has no agreement with ISP_B , the customers of ISP_B have only one option to get content from a nearby server (the server I) while the customers of ISP_A have two options (the server I and the server A). The goal of ISP_A is to both improve the QoE for its customers and to generate some revenue from the exploitation of a privileged location in the network.

The CDN has four ways to store a given content: in none of the edge servers (*i.e.*, in the origin server), in I only, in A only, or in both I and A . A rational CDN again decides with regard to economic objectives, so both servers I and A are filled so that the net revenue is maximum. Figure 3 summarizes the one-to-one comparisons of Table I, and illustrates the best caching decisions as a function of the content popularity (x_A, x_B) , for given values of λ_I and λ_A .

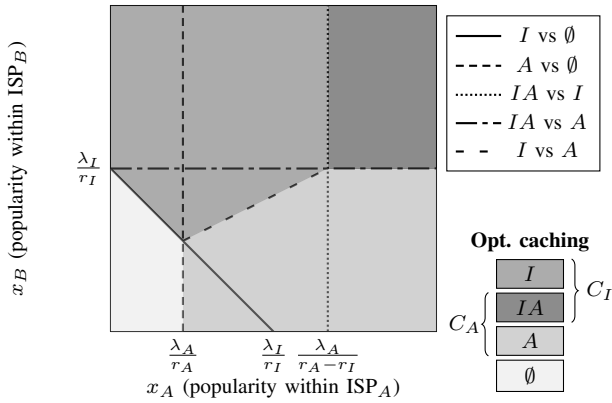


Fig. 3. Optimal caching decisions when there is no cache in B ($\lambda_I = 1.5, \lambda_A = 2, r_I = 1, r_A = 2$).

The figure highlights the impact that the edge server A can have on the QoE of customers of ISP_B . Consider some

content with a given popularity x_B , and assume that $x_B \in \left(\frac{\lambda_I}{r_I} - \frac{\lambda_A}{r_A}, \frac{\lambda_I}{r_I}\right)$. Then the quality perceived by population B for that content depends on the content popularity among population A :

- if $x_A < \frac{\lambda_I}{r_I} - x_B$, the content is not cached at all and users in population B get basic quality;
- if $x_A \in \left[\frac{\lambda_I}{r_I} - x_B, \tilde{x}_A\right)$ with $\tilde{x}_A = \frac{\lambda_A}{r_I} + (x_B - \frac{\lambda_I}{r_I})(\frac{r_A}{r_I} - 1)$, the content is stored in the intermediate cache I , therefore users in both populations get better (say, medium) quality;
- if $x_A > \tilde{x}_A$, the content is kept in cache A , hence users in population A get top quality while those in population B only get basic quality.

Similarly, the quality perceived by population A for some content depends on the popularity of that content in population from ISP_B . In particular, some content that is popular in ISP_B can be cached in I instead of A , which can thus slightly degrade the QoE of users in ISP_A .

A. Capacity of CDN Servers

The amount of data managed in each possible way (not stored, stored in I and/or A) corresponds to the mass of content in each zone as displayed in Figure 3, computed using the joint distribution of (x_A, x_B) . This amount is a function of the service and storage prices $(r_j, j = I, A)$, and the values of the unit (or shadow) prices $(\lambda_j, j = I, A)$. Since prices are fixed, let us denote the amount of stored content in I and A by $C_I(\lambda_I, \lambda_A)$ and $C_A(\lambda_I, \lambda_A)$, respectively.

(i) In the case of linear capacity costs (*i.e.*, fixed λ_A and λ_I), we immediately derive the storage capacity;

(ii) In the fixed-capacity case, λ_A and λ_I are solutions of

$$\begin{cases} C_I(\lambda_I, \lambda_A) = \tilde{C}_I \\ C_A(\lambda_I, \lambda_A) = \tilde{C}_A, \end{cases}$$

with \tilde{C}_I and \tilde{C}_A the fixed capacities of I and A , respectively;

(iii) Finally in the case of non-linear (convex) capacity costs, λ_A and λ_I are solutions of the first-order optimality conditions stating that the marginal revenue from capacity in I (*resp.*, A) equals the marginal cost $C'_I(C_I(\lambda_I, \lambda_A))$ (*resp.*, $C'_A(C_I(\lambda_I, \lambda_A))$) of that capacity, where C_I and C_A are the capacity cost functions.

B. Analysis with VoD Traces

We have previously shown that the presence of edge servers in the network of a given ISP can impact other ISPs. In the following we quantify this impact on a concrete example.

Let us consider an initial configuration where the CDN has only one intermediate server. Then, ISP_A decides to rent an edge server to the CDN in its network. The presence of the edge server A becomes a source of conflicts, when content previously stored in I gets now only stored in A (and no longer in I). The QoE of end-users from ISP_B is degraded because the content is now accessed from the origin server.

We study the problem by confronting our model to some real traces from a French Video on Demand (VoD) service. All

the requests for movies have been recorded during two years from 2010 to 2012. A key point of this dataset is that it includes the network origin of the requesters, which is here the Point of Presence from which each request is emitted. There are twelve “sub-networks” in the dataset, corresponding to twelve different populations.

We use the dataset to build a realistic environment as follows. We associate the sub-network population that generated the highest number of requests with ISP_A (the network with an edge server in our model). ISP_A is thus the *dominant* network in the system. The ISP_B is the average of the other sub-networks in the dataset. That is, for each movie, the popularity x_B is the average popularity over the eleven sub-networks. Finally, we build a standard CDN with one intermediate server for the whole country, and an edge server close to the end-users in the regional sub-network ISP_A .

Figure 4 depicts the joint popularity distribution of movies using the same representation as in Figure 3. As can be seen, the density of movies that are candidate for storage in CDN (popular-enough movies) is mainly on south east, which corresponds to storing content in A only.

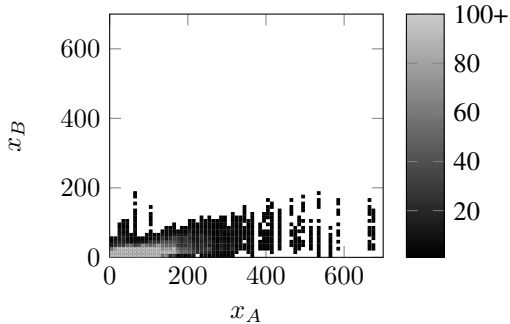


Fig. 4. Correlation between popularities from ISP_A and ISP_B . The gray scale represents the number of movies that have a given popularity (x_A, x_B) .

The results for this dataset are summarized in Figure 5, where we show the impact of edge server A on the QoE of ISP_B users with respect to the initial configuration without A . The QoE for content accessed from the origin server is half the QoE from intermediate and edge servers. Our model for capacity is the fixed-cost one: λ_I and λ_A are unit prices for storage, and the CDN chooses the capacities such that the marginal revenues equal those prices. On the y -axis, we compute the ratio between the satisfaction with and without the edge server A . On the x -axis, we set different storage costs for A , from an aggressive policy (cheaper storage than in I) to a benefit-oriented one (A more expensive than I).

As Figure 5 illustrates, the edge server A has *both* a positive impact on ISP_A user satisfaction (which is expected), and a negative impact on ISP_B user satisfaction (which is not suitable). Observe that the negative impact on ISP_B exceeds the positive impact on ISP_A for most of the storage cost prices on A . Hence a powerful network operator does not need to implement any aggressive pricing policy to degrade the performances of competitors: an efficient strategy can be to set a high rental price for the storage in A to both get revenues from the CDN and acquire customers due to churn from ISP_B .

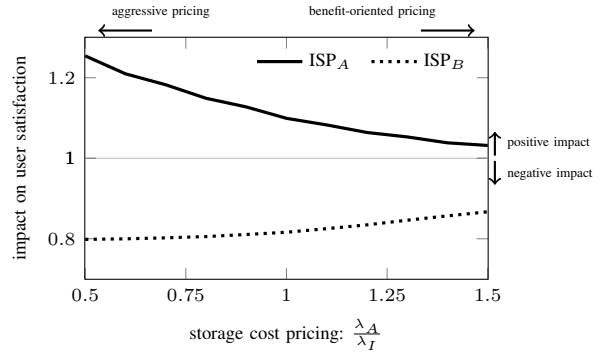


Fig. 5. Impact of the cache within ISP_A on user perceived quality in both ISPs.

IV. SHOULD THE CDNS BE REGULATED?

In February 2015, the US regulator reinforced the principle of *net neutrality*, classifying broadband Internet service as a public utility. A key rule of net neutrality is the non-differentiation of packets in the network, regardless of the application. A reasoning behind it is to ensure that new entrants can enter the market without massively investing in the infrastructure. This rule mainly targets CPs, and also applies to ISPs. A “neutral” network should allow a new ISP to start providing access service and to benefit from the services of a CDN despite a smaller population than its competitors.

In this paper, we have shown that edge servers located deep in the network can violate this principle because the CDN favors incumbent ISPs and prevent smaller network operators from benefiting from a good quality of service. On the other hand, one can argue that, in our scenario, the ISP_A is rewarded from investing in its infrastructure (here installing an edge server in its network). In our opinion, the positive impact on the satisfaction of customers of ISP_A is legitimate but the negative impact on ISP_B is not.

We call for a scientific debate on the definition and regulation of a neutral CDN. In the context of the competition among content providers, we sketched some proposals in [9], in the vein of other past papers related to net neutrality [1], [4] and CDN management [2], [5].

This paper is also another approach for studying networks of caches. In the recent years, the scientific community has paid a lot of attention on Content Centric Network (CCN) proposals, which are based on in-network caching and decoupling between content and destination host. Many papers have dealt with the management of in-network caches [7], [11] but to our knowledge the economic aspects behind network management have been ignored so far.

Our future works include the merge of both models from the present paper and the one in [8], i.e., to consider content popularities distributions differing among ISPs and CPs. Our motivation is to understand all (sometimes conflicting) interests that a CDN provider has to take into account when managing its CDN, to identify the scenarios that may require the attention of regulators regarding neutrality, and to provide solutions that preserve not only the profits of CDNs but also a fair competition among all other involved actors.

REFERENCES

- [1] E. Altman, A. Legout, and Y. Xu. Network non-neutrality debate: An economic analysis. In *Proc. of IFIP Networking*, 2011.
- [2] R. Buyya, A.-M. Pathan, J. Broberg, and Z. Tari. A case for peering of content delivery networks. *Distributed Systems Online, IEEE*, 7(10), 2006.
- [3] Cisco. Wholesale Content Delivery Networks: Unlocking New Revenue Streams and Content Relationships. Technical Report 710667, Cisco, 2011.
- [4] P. Coucheney, P. Maillé, and B. Tuffin. Impact of Reputation-Sensitive Users and Competition Between ISPs on the Net Neutrality Debate. *IEEE Trans. on Network and Service Management*, 4(10), 2013.
- [5] K. Hosanagar, J. Chuang, R. Krishnan, and M. Smith. Service adoption and pricing of content delivery network (CDN) services. *Management Science*, 54(9):1579–1593, 2008.
- [6] S. S. Krishnan and R. K. Sitaraman. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. *IEEE/ACM Trans. on Networking*, 21(6):2001–2014, 2013.
- [7] Z. Li, G. Simon, and A. Gravey. Caching policies for in-network caching. In *Proc. of IEEE ICCCN*, 2012.
- [8] P. Maillé, K. Pires, G. Simon, and B. Tuffin. How neutral is a CDN? an economic approach. In *Proc. of CNSM*, 2014.
- [9] P. Maillé, G. Simon, and B. Tuffin. Toward a Net Neutrality Debate that Conforms to the 2010s. <https://hal.inria.fr/hal-01127958>, Mar. 2015.
- [10] S. Puopolo, M. Latouche, F. L. Faucheur, , and J. Defour. CDN Federations: How SPs Can Win the Battle for Content-Hungry Consumers. Technical report, Cisco, 2011. "Point of View" White Paper.
- [11] E. J. Rosensweig, J. Kurose, and D. Towsley. Approximate models for general cache networks. In *Proc. of IEEE INFOCOM*, 2010.