



**HAL**  
open science

## The Economics of CDNs and Their Impact on on Service Fairness

Eric Gourdin, Patrick Maillé, Gwendal Simon, Bruno Tuffin

► **To cite this version:**

Eric Gourdin, Patrick Maillé, Gwendal Simon, Bruno Tuffin. The Economics of CDNs and Their Impact on on Service Fairness. *IEEE Transactions on Network and Service Management*, 2017, 14 (1), pp.22-33. 10.1109/TNSM.2017.2649045 . hal-01398923v2

**HAL Id: hal-01398923**

**<https://inria.hal.science/hal-01398923v2>**

Submitted on 14 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Economics of CDNs and Their Impact on Service Fairness

Eric Gourdin, Patrick Maillé, Gwendal Simon, Bruno Tuffin

**Abstract**—Content Delivery Networks (CDN) have become key telecommunication actors. They contribute to improve significantly the quality of services delivering content to end users. However, their impact on the ecosystem raises concerns about their “fairness”, and therefore the question of their inclusion in the neutrality debates becomes relevant. This paper aims at analyzing the impact of a revenue-maximizing CDN on some other major actors, namely, the end-users, the network operators and the content providers, at comparing the outcome with that of a fair behavior, and at providing tools to investigate whether some regulation should be introduced. We present a mathematical model and show that there exists a unique optimal revenue-maximizing policy for a CDN actor, in terms of dimensioning and allocation of its storage capacity, and depending on parameters such as prices for service/transport/storage. Numerical experiments are then performed with both synthetic data and real traces obtained from a major Video-on-Demand provider. In addition, using the real traces, we compare the revenue-based policy with policies based on several fairness criteria.

**Index Terms**—Content Delivery Networks, Fairness, Revenue maximization

## I. INTRODUCTION

In less than fifteen years, Content Delivery Networks (CDNs) have become key actors of the value and supply chain of the Internet. Besides, the CDN activity impacts the other economic actors in several ways. First, the major Content Providers (CPs) have to develop partnerships with CDNs because they need to rely on a wide, distributed infrastructure to deliver their data. The size of CDN infrastructures (more than 200,000 servers for Akamai<sup>1</sup>) and their deployment close to end-users should ultimately improve the user quality of experience (QoE), especially for the services that are sensitive to delays and bandwidth stability (for instance, video streaming or online gaming). To differentiate from their competitors and guarantee a good QoE, CPs have to either pay for a CDN service, or invest in their own infrastructure (sometimes with the support of a CDN). Second, the dramatic growth of the traffic managed by CDNs has impacted the traffic exchanges between network providers, and consequently their economic relationships [1]–[4]. Typically, several CDNs are now among the ten biggest contributors in terms of traffic while this ranking included only Internet Service Providers (ISPs) less than ten years ago [4]. Third, various traditional actors, which pay attention to the revenue generated by the main CDN providers (around \$2 billion dollars for Akamai), have started developing their own CDN activity. As the CDN development reduces the demand volumes for transit services, some major transit network operators such as Level 3 shifted a fraction

of their activities to CDN. Some ISPs, CPs and equipment vendors have also a CDN service [5]–[7]. This shift, which is referred to as *vertical integration* of CDN activity, illustrates the major importance of this type of business.

Surprisingly, although CDNs have undoubtedly become key actors in the telecommunication world, their *economic* aspirations have been barely analyzed using rigorous modeling and performance tools. Among the few exceptions, we can mention [8]–[10] where the author(s) consider a (single) CDN and a time period of interest; the number of requests to the CP during the period is unknown and treated as a random variable, and the pricing strategy is optimized, showing that the common practice, known as *economy of scale*, and consisting in having a decreasing price per unit as the total traffic increases, is indeed optimal. But the complex relationships between actors (and their consequences on fairness and social welfare) are neither included nor investigated in those works. Faratin [11] describes and explains the wholesale market failure due to the lack of end-to-end coordination between ISPs and CPs—hence the appearance of intermediaries like CDNs—and analyzes the best strategies of intermediaries. This last work uses game theory as a modeling tool. To the best of our knowledge, game theory has otherwise only been used to analyze the interactions between several CDNs [12]–[14], but not to study the interactions between the different types of actors. Technical performances of CDNs have been extensively studied, but without including in the picture the interactions with other economic actors or profit-driven strategies.

In this paper, we propose a model capturing the main elements that a CDN should take into account for a profit-driven management of its infrastructure. We focus on the management problems faced by a CDN having to dimension and optimally use its infrastructure, sharing it among its clients (content/service providers). The model that is presented here extends the work that was introduced in [15] by addressing the problem of infrastructure dimensioning, by extending the analysis to more than two CPs and general popularity distributions, and by providing new specific results for the case when popularities for the content from each CP follow power laws. Also, with the objective of investigating whether regulation is needed for CDNs, we discuss several definitions of fairness and analyze their impact. Finally, with respect to [15] the numerical experimentation is based on VoD traces instead of live streaming channels, a framework that is today the main usage of CDNs [16]. Based on our previous model, which considers several content providers in competition for the CDN service, we analyze the behavior of a profit-maximizing CDN, determine the optimal CDN policy, and assess its impact on the quality perceived by users and on the fairness among content

<sup>1</sup><https://www.akamai.com/us/en/about/facts-figures.jsp> (February 2016)

providers. We illustrate our theoretical results with an analysis based on real data from two major content providers (obtained through real-world measurements), which we assume compete for the CDN resources. In particular, we investigate the impact of the CDN economic behavior on the quality perceived by users and on the fairness among content providers when the CDN charges them different prices. We show that a CDN implementing a revenue-maximizing policy tends to favor the content providers that apply aggressive financial policies (paying more for content), and at an even bigger extent when the aggressive content provider is the biggest player of the competition. Since this inclination toward the financially attractive provider results in a gap in the QoE of end-users of both service providers, it goes against fairness principles that regulators generally try to enforce. We also study the behavior of other CDN management strategies, which can be thought of as more fair, but we show that these policies are far from being perfect, either for the treatment of heterogeneous service providers, or to ensure revenues for the CDN, or both. This analysis calls for future works, which we believe has not received enough attention, especially when compared to its importance in public debates.

Note that some of our previous works intend to address other fairness-related aspects of the CDN activity: in [17] we analyze its impact on the competition among ISPs, stressing the fact that larger ISPs would be favored by the presence of CDNs; in [18] we also focus on the competition among ISPs, but with a different model for user demand and covering peering/transit agreements among ISPs; finally [19] focuses on innovation, i.e., the amount of content offered by CPs (that amount is assumed fixed in the present submission). Covering all those aspects at the same time makes them difficult to understand, hence we separate the problems into several models.

The remainder of the paper is organized as follows. In Section II, we present the main elements of the problem, including our model based on revenues and cost for the CDN, and the impact on the QoE for the end-users. In Section III, we present our two main results, which are the optimization of the storage sharing policy for a given server capacity, and a joint optimal server capacity dimensioning and storage sharing. In Section IV, we highlight some results that illustrate the fairness problem due to revenue-based policy for caches. We first provide an analytical evaluation based on synthetic configurations for CPs whose content popularities follow a power law. We then provide an example based on real traces, also used to compare the revenue-based policy with policies based on various fairness notions. Finally, Section V highlights the links we think our model has with the current debates on neutrality [35]–[38], and Section VI presents our conclusions.

## II. MODEL

### A. Economic relationships between actors

A CDN manages a set of storage servers (called *edge servers* or CDN caches), which are located “close” (in a network sense) to users, sometimes within the network of the ISP. The idea is to store (or cache) in these edge servers the

content that is most likely to be downloaded by users. The edge server hence partially replaces the (often distant) server that stores the original content (called *origin server* or *entry point*). Big CDNs have multiple origin servers, so that each service provider can have an easy entry access to the CDN infrastructure.

The use of CDNs offer several advantages. They improve the QoE for the end-users because they alleviate the load on the origin servers, and reduce the response time by shortening the path between the end-user and the content. CDNs also reduce the traffic in the backbone network because a given content is downloaded from the origin server only once, and each subsequent request for this content is fulfilled from the edge server on a much shorter path, which limits transit through intermediary transit networks.

A CDN is a *multi-tenant* infrastructure: its resources are shared among multiple CPs. This is illustrated in Figure 1, where  $n$  is the number of CPs. Only one origin server per CP is depicted, although in the reality origin servers are shared among multiple CPs from the same region. We also display only one ISP with end-users, but the same figure could be replicated with several ISPs: the CDN would then own another set of storage servers located in those other ISPs, each one being studied independently.

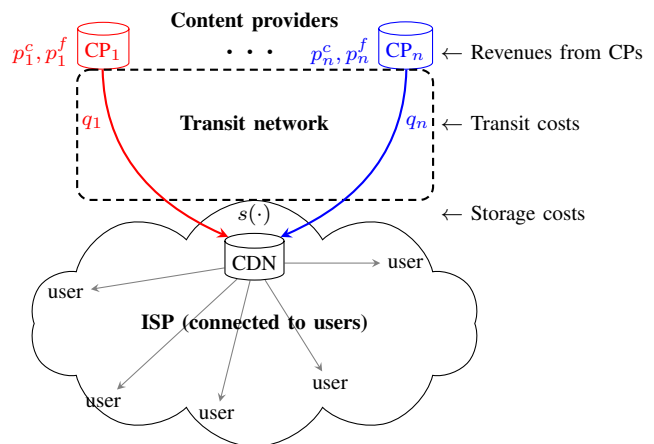


Fig. 1. Costs and revenues for a CDN located within an ISP's network.

We define a cost and pricing model, which captures a typical agreement between a CDN, an ISP, and two CPs. Figure 1 shows the main economic flows exchanged by the CDN within this model. We acknowledge the fact that our model does not cover all possible situations. There exists multiple forms of agreements between these actors, according to their relative weights. For example, an ISP does not negotiate in the same way with a powerful CDN provider like Akamai or with a small niche CDN. As a matter of fact, the CDN sometimes pays a fee to the ISP for the rental of a rack in an ISP's colocation center [20], while the agreement is sometimes based on resource exchanges in a federation of ISP-CDNs [21]. Nevertheless, we believe that our model addresses the most common configuration. In particular, for the CDN provider, our model allows the distinction between:

- **Revenues:** Each CP subscribes to the CDN service to reach its customers. To simplify the typical Service Level Agreement (SLA) between CDNs and CPs, we say that the CDN charges CPs a different price per unit of data volume delivered to users, according to whether users are served from the edge server (unit price  $p_i^c$  for  $CP_i$ ,  $i = 1, \dots, n$ ) or from the CP origin server (unit price  $p_i^f \leq p_i^c$  for  $CP_i$ ), in this case, with lower QoE.
- **Costs:** The CDN has two types of costs. First the storage in the CDN cache incurs a cost represented by a function  $s(\cdot)$  of the storage capacity, aggregating hardware, estate, and energy costs. Second, the CDN pays transit fees (to transit networks) to get the content from the origin servers of the CPs and make it available to the end-users. This cost is a function of the required volume of data to be delivered. When the data are taken from the origin server, the cost is the transit cost  $q_i$  per unit of volume (which can be low if the CDN owns the transit network, but large otherwise). Remark that those transit costs can differ among CPs (e.g.,  $q_i \neq q_j$ ) because CPs may have different origin servers to provide data, so the path between the origin server and the ISP of interest may differ. When the data is taken from the edge servers, the cost is significantly smaller. Here, we opted for not including it in our study because (i) the cost is the same for both CPs, so this part of traffic does not generate any inclination toward a CP over another, and (ii) with the CDN redirection, every request travels through the edge server before being routed to the origins servers, so this cost has no impact on the decision of whether the content is cached or not.

### B. Monopolistic or competing CDNs?

We refer to a unique CDN throughout the paper, but it does not mean that we are assuming a monopolistic CDN, which can fix prices so as to maximize revenues. Indeed, our analysis takes place *after* prices (especially  $(p_i^f, p_i^c)$  pairs for  $i = 1, \dots, n$ ) are fixed. How those prices are decided is out of the scope of this paper. These prices are mostly a consequence of the CDN market structure. For instance, a monopolistic CDN would set prices considering the elasticity of demand, to extract as much as possible from the willingness-to-pay of CPs for the CDN delivery service, while in the other extreme a perfect competition (with a large number of competing CDNs) would lead to unit prices close to marginal costs, hence null revenues for CDN actors. In the middle, oligopolistic situations should be studied using game-theoretic models [22], [23].

In this paper, we set the CDN prices as an input of the problem, these prices being the result of the interactions among CDNs. The study we propose here focuses on what the CDN (or any one of them, if there are several) do with the content it manages once those prices are fixed, and the consequences on QoE and possibly fairness among CPs.

Note also prices and cache management are not the only decisions a CDN actor has to make: at a larger time scale, the decisions on the number of cache servers to install, their locations, and the ISPs with which to peer are also crucial [7],

[24]. In this paper we assume those are fixed, and we focus on the shortest time scale, at which cache management policies can be modified.

### C. Decision variables of the CDN

We assume that the above costs and prices are fixed (from long-term contracts, and/or from competition interactions among CDNs). We thus focus on two major aspects of the CDN management strategies, namely:

- The **dimensioning** of the storage servers (in Figure 1 we represent only the server farm located in the considered ISP). The decision variable is the storage capacity of that farm, which we denote by  $C$ ;
- The **storage policy** implemented, i.e., the management of the storage space in the CDN cache (storage) servers. The decision variables are the (fixed) storage capacities  $(C_i)_{i=1, \dots, n}$  allocated to each CP.

Note that, in this paper, we chose not to deal at the same time with the aspects related to the dynamics of cache management. We hence treat the content popularity values as *constant and known to the CDN operator* (this also amounts to assuming the estimation schemes to be efficient enough to adapt rapidly to popularity changes). Thus our approach differs from most of the CDN-related studies, where the goal is to estimate the likelihood that each piece of content will be requested in the future so as to keep the most popular content in the cache. We can mention, among the economy-oriented approaches, the use of different types of contracts to maintain content in the CDN cache servers [25], which converges to our popularity-based approach when popularities are static. For completeness, we can recall that caching strategies—the practical implementations of cache management—involve two main algorithms: one—called *admission policy*—to decide what new content to insert, and another—called *replacement policy*—to decide which content to remove from the cache when a new piece of content has to be stored. There are several replacement policies, among which the most often used are LRU (*least recently used*) and LFU (*least frequently used*, where the frequency is estimated by measurements, and corrected with a limited memory effect), see for example [12] and references therein. We ignore this level of treatment here, focusing on the economic tradeoffs that already occur when popularities are perfectly known and static. In that case, the LFU policy would correspond to a CDN keeping in the cache only the most popular content. This policy can be interpreted as being fair while some economic factors may incentivize the CDN to deviate from it.

To represent the popularity of the content at each CP, we denote by  $F_i(x)$ ,  $i = 1, \dots, n$  the request frequency (number of requests per time unit) for the  $x^{\text{th}}$  most popular unit of content of  $CP_i$ . We assume that  $F_i$  is continuous and strictly decreasing, and that some content has infinitely small popularity, i.e.,  $F_i(V_i) = 0$  with  $V_i$  the total volume of content proposed by  $CP_i$ ,  $i = 1, \dots, n$ .

### D. Revenue of the CDN

Let us now express the incomes that the CDN gets when deploying a storage capacity of  $C$ , and allocating  $C_i$  of that

capacity for the most popular data from  $CP_i$ . The gains obtained from CPs are then

$$\sum_{i=1}^n p_i^c \int_{x=0}^{C_i} F_i(x) dx + \sum_{i=1}^n p_i^f \int_{x=C_i}^{V_i} F_i(x) dx,$$

The expression of the gains can be simplified by introducing the functions

$$G_i(y) := \int_{x=0}^y F_i(x) dx$$

and the values

$$\bar{G}_i := G_i(V_i) \text{ and } \bar{G} := \sum_{i=1}^n \bar{G}_i$$

for  $i = 1, \dots, n$ .  $G_i(y)$  is the user throughput generated when requests from the  $y$  most popular units of content are downloaded from  $CP_i$ , and  $\bar{G}_i$  (resp.  $\bar{G}$ ) is the total user download throughput of  $CP_i$  (resp. of all CPs) content. With these new notations, the (per time unit) incomes of the CDN can be rewritten as

$$\sum_{i=1}^n p_i^f \bar{G}_i + (p_i^c - p_i^f) G_i(C_i).$$

Without loss of generality, we ignore contents with no demand, so that we can assume  $F_i(x) > 0$  for all  $x < V_i$  ( $i = 1, \dots, n$ ).

Let us now focus on CDN costs, which come from storage and transit. The storage costs equal  $s(C)$ , where  $s(\cdot)$  is a strictly increasing function which we assume to be convex.<sup>2</sup> In the following, we will treat the linear case as a special example. For the transit costs, we neglect the one-shot costs for the first time contents are stored in the CDN cache: therefore transit costs only correspond to the content that is not in the cache, and for each CP they are proportional to the aggregated download rate for that content since they are incurred each time the content is requested by a user. The total transit costs can therefore be expressed as follows

$$\sum_{i=1}^n q_i (\bar{G}_i - G_i(C_i)).$$

Overall, the net revenue of the CDN operator (still, per time unit) is then

$$R = \sum_{i=1}^n r_i^f \bar{G}_i + (p_i^c - r_i^f) G_i(C_i) - s(C), \quad (1)$$

with

$$r_i^f := p_i^f - q_i \quad \text{for } i = 1, \dots, n.$$

### E. Average quality experienced by users

User QoE is the core concern of CPs and the main reason why they pay a CDN to deliver their content. Indeed, it has been shown [26]–[28] that user engagement strongly depends on QoE, and the higher/lower the QoE, the stronger/weaker the engagement of a user with a specific service. For HTTP video streaming content (which is by far the dominant content

nowadays), studies [26]–[30] have shown that the prevailing QoE criteria include the occurrence of re-buffering events (i.e., stops in the video playback) and the time it takes to start the video, two criteria that are expected to improve when the end-user is served from a nearby cache instead of a distant origin server. Therefore, we consider two QoE metrics, according to whether the end-user is served from the cache or not:  $Q_c$  is the quality experienced by users for the cached content, while  $Q_{f,i}$  is for content retrieved from the origin server of  $CP_i$ . The higher the value, the higher the quality, and hence  $Q_{f,i} < Q_c$  for all  $i$ . The average quality experienced per download is then

$$Q^{\text{avg}} = \frac{1}{\bar{G}} \left( Q_c \sum_{i=1}^n G_i(C_i) + \sum_{i=1}^n Q_{f,i} (\bar{G}_i - G_i(C_i)) \right).$$

This measure will be used, later in this paper, to investigate the impact of the CDN strategies on user experience quality.

Note that we consider a better QoE when the content is served from the edge server rather than from the origin server ( $Q_{f,i} < Q_c$ ) in the model because the closest edge server to a user is closer than the origin server. However, in practice CDNs may fail to find the closest caches as shown in [31]; in this paper we neglect that effect of wrong CDN re-directions.

### F. Fairness among service providers

To complete the analysis of the CDN impact on the ecosystem, we also consider the standpoint of CPs, and we focus on potential unfair behaviors, which could ultimately limit innovation and the development of new services. The fairness among CPs is a key concern in the ongoing debates on neutrality (of networks and/or search engines). However translating fairness principles into the CDN framework is not trivial.

A revenue-maximizing CDN chooses values of  $C_i$  leading to minimum frequencies of cached content  $F_i(C_i)$  which can be very different between CPs because of different prices and/or costs. This decision may be considered an unfair treatment of CPs by the CDN. Regulators can force CDNs to adopt other strategies, e.g., to adopt the following ‘‘fair’’ behaviors.

- The CDN treats content without regard to the origin, caching the most popular content whatever the benefits and costs. This implies choosing  $C_i$  such that  $F_i(C_i) = F_j(C_j)$  for all  $1 \leq i, j \leq n$ .
- The CDN allocates the same amount  $C/n$  to all CPs, i.e., treating them in an equal way without regard to the popularity of their content. Observe that this solution is essentially a case study. Indeed, it would first require the implementation of an administrative authority to *register* the CPs. Second, it is hardly justifiable since it gives the same cache space to a small newcomer CP as to a large incumbent CP. Moreover it is far from optimal not only in terms of CDN revenue but also for the overall user perceived quality. Finally, a CP would be incentivized to artificially split into two or more CPs in order to aggregate a larger share of the storage space.

<sup>2</sup>More data stored may involve more expenses in rack space, which can become scarce, and more air-conditioning costs; marginal costs can increase in both cases.

- The CDN provides the same average user QoE to all CPs, i.e., choosing  $C_i$  such that for all  $i, j$ ,

$$\begin{aligned} Q_c \frac{G_i(C_i)}{\bar{G}_i} + Q_{f,i} \left(1 - \frac{G_i(C_i)}{\bar{G}_i}\right) \\ = Q_c \frac{G_j(C_j)}{\bar{G}_j} + Q_{f,j} \left(1 - \frac{G_j(C_j)}{\bar{G}_j}\right). \end{aligned} \quad (2)$$

- The CDN ensures the same *hit-ratio* to all CPs, i.e., choosing  $C_i$  such that

$$\frac{G_i(C_i)}{\bar{G}_i} = \frac{G_j(C_j)}{\bar{G}_j}.$$

Note that this is a special case of (2), obtained when  $Q_{f,i} = Q_{f,j}$  for all  $i, j$ .

### III. MAXIMIZATION OF THE CDN REVENUE

The goal of this section is to determine the optimal decisions of a revenue-maximizing CDN regarding the total storage capacity  $C$  and the sharing of the storage among CPs  $C_i, 1 \leq i \leq n$ . We consider two scenarios:

- The storage capacity  $C$  of the CDN is fixed. For instance, one can consider that the CDN rents a fixed storage  $C$  and that this rental comes from agreements between the CDN and the ISP. The goal for the CDN is then to determine the values  $C_i, 1 \leq i \leq n$  that maximize the net revenue in (1) under the constraint  $\sum_{i=1}^n C_i \leq C$ . We detail this scenario in Section III-A.
- The CDN can freely choose the total storage capacity  $C$  in a consumption-based pricing model (or on-demand storage). The goal is then to *jointly* determine the optimal storage  $C$  and its allocation among CPs. We address this problem in Section III-B.

#### A. Optimal Storage Sharing, Fixed Capacity

##### 1) Main Result:

*Proposition 1:* assume the total storage capacity  $C$  is fixed. There exists a unique allocation of storage capacities  $(C_i^{\text{opt}})_{1 \leq i \leq n}$  maximizing the net revenue of the CDN. This allocation satisfies

$$\begin{aligned} \forall i, j \quad C_i^{\text{opt}} > 0 \Rightarrow (p_i^c - r_i^f) F_i(C_i^{\text{opt}}) \geq (p_j^c - r_j^f) F_j(C_j^{\text{opt}}) \quad (3) \\ \sum_{i=1}^n C_i^{\text{opt}} = C. \end{aligned} \quad (4)$$

In particular, if two CPs  $i$  and  $j$  store some content then  $\frac{F_i(C_i^{\text{opt}})}{F_j(C_j^{\text{opt}})} = \frac{p_j^c - r_j^f}{p_i^c - r_i^f}$ .

**Proof.** From the expression of the net revenue in (1) we can formulate the CDN revenue maximization problem as:

$$\begin{aligned} \max_{C_1, \dots, C_n} \quad & \sum_{i=1}^n r_i^f \bar{G}_i + (p_i^c - r_i^f) G_i(C_i) - s(C) \quad (5) \\ \text{s.t.} \quad & \begin{cases} C_i \geq 0 & \forall i \\ \sum_{i=1}^n C_i \leq C. \end{cases} \end{aligned}$$

Note that the first term in the objective function is a constant that can be removed when focusing on the part to be optimized. Each function  $G_i$  ( $i = 1, \dots, n$ ) is continuously differentiable (with derivative  $F_i$ ), strictly increasing and strictly concave

on  $[0, V_i]$ , and the constraints form a convex and compact set, hence the maximum exists and is unique. Since the revenue is strictly increasing in each  $C_i$ , all of the capacity is used, giving (4). Denoting by  $\lambda_i$  the Lagrange multipliers for the constraint  $C_i \geq 0$ , and by  $\mu$  the multiplier for the constraint  $\sum_{i=1}^n C_i = C$ , the unique solution  $(C_i^{\text{opt}})_{i=1, \dots, n}$  is characterized by the set of Karush-Kuhn-Tucker conditions [32]:

$$\begin{aligned} (p_i^c - r_i^f) \underbrace{G'_i(C_i^{\text{opt}})}_{F_i(C_i^{\text{opt}})} &= -\lambda_i + \mu \quad i = 1, \dots, n \quad (\text{stationarity}) \\ \sum_{i=1}^n C_i^{\text{opt}} &= C \\ \text{and } C_i^{\text{opt}} &\geq 0 \quad i = 1, \dots, n \quad (\text{primal feasibility}) \\ \lambda_i &\geq 0 \quad i = 1, \dots, n \quad (\text{dual feasibility}) \\ \lambda_i C_i^{\text{opt}} &= 0 \quad i = 1, \dots, n \quad (\text{complementary slackness}). \end{aligned}$$

From the last two conditions, we remark that if  $C_i^{\text{opt}} > 0$  then  $\lambda_i = 0$ , and the first condition yields  $(p_i^c - r_i^f) G'_i(C_i^{\text{opt}}) = \mu$ . Then since  $\lambda_j \geq 0$  for all  $j$ , the first condition implies  $(p_j^c - r_j^f) G'_j(C_j^{\text{opt}}) \leq \mu = (p_i^c - r_i^f) G'_i(C_i^{\text{opt}})$ . ■

2) *Discussion:* We would like to emphasize two observations.

First, for two CPs  $i$  and  $j$ , we can quantify the “unfairness” in their treatment through the ratio  $\frac{F_i(C_i^{\text{opt}})}{F_j(C_j^{\text{opt}})}$ : a value strictly above 1 indicates that CP  $i$  is disfavored with respect to  $j$ , having uncached content more popular than some cached content of CP  $j$ . When both CPs are allocated storage space in the cache, that unfairness is exactly the ratio of the unit benefits for their content (*prices minus costs*), that is  $\frac{p_j^c - r_j^f}{p_i^c - r_i^f}$ .

Second, for each  $i, 1 \leq i \leq n$ , the value of  $C_i^{\text{opt}}$  increases with  $C$ , strictly if  $C_i^{\text{opt}} > 0$ . Indeed, if  $\tilde{C} > C$ , there is at least one CP  $j$  for which  $C_j^{\text{opt}}$  strictly increases since the whole capacity is always used. Denoting the new optimal values by  $\tilde{C}_i^{\text{opt}}$  and  $\tilde{C}_j^{\text{opt}}$ , we have by applying (3) twice, for  $\tilde{C}_j^{\text{opt}} > 0$  and then for  $C_i^{\text{opt}} > 0$ ,

$$F_j(\tilde{C}_j^{\text{opt}}) \geq \frac{p_i^c - r_i^f}{p_j^c - r_j^f} F_i(\tilde{C}_i^{\text{opt}}) \geq \frac{F_i(\tilde{C}_i^{\text{opt}})}{F_i(C_i^{\text{opt}})} F_j(C_j^{\text{opt}}).$$

Therefore  $F_i(\tilde{C}_i^{\text{opt}}) \leq F_i(C_i^{\text{opt}}) \frac{F_j(\tilde{C}_j^{\text{opt}})}{F_j(C_j^{\text{opt}})}$ , and due to the strict decreasingness of  $F_j$  and  $F_i$ ,  $\tilde{C}_j^{\text{opt}} > C_j^{\text{opt}}$  implies that  $\tilde{C}_i^{\text{opt}} > C_i^{\text{opt}}$ .

3) *Application to CPs with power-law popularity distributions:* Assume two CPs with power-law content popularity distributions, i.e., the probability density that a piece of content (assumed infinitesimal) is requested  $y$  times is proportional to  $y^{-\tau}$  for  $y$  above a threshold  $y_{\min}$  (most studies have shown that  $\tau$  is greater than two for media services [33]). We derive in Appendix A that the request rate  $F_i(x)$  for the  $x^{\text{th}}$  most popular unit of content of CP  $i$  equals

$$F_i(x) = A_i x^{-\alpha}$$

where  $\alpha = \frac{1}{\tau-1}$  (so for real cases,  $0.5 < \alpha < 1$ ), as long as  $A_i x^{-\alpha} \geq y_{\min}$ , i.e.,  $x \leq \left(\frac{A_i}{y_{\min}}\right)^{1/\alpha} := C_{i,\max}$ , which

indicates the domain of validity of the power law. If the values  $C_{i,\max}$  are large enough,

$$\frac{F_2(C_{2,\max})}{F_1(C - C_{2,\max})} < \frac{p_1^c - r_1^f}{p_2^c - r_2^f} < \frac{F_2(C - C_{1,\max})}{F_1(C_{1,\max})},$$

and thus from (3) the optimal value of  $C_1$  is in  $(C - C_{2,\max}, C_{1,\max})$ , where we used the fact that  $C_1 + C_2 = C$  at the optimum. Equation (3) gives

$$\frac{C_1^{\text{opt}}}{C - C_1^{\text{opt}}} = \left( \frac{A_1 p_1^c - r_1^f}{A_2 p_2^c - r_2^f} \right)^{1/\alpha},$$

that is,

$$C_1^{\text{opt}} = \frac{C}{1 + \left( \frac{A_2 p_2^c - r_2^f}{A_1 p_1^c - r_1^f} \right)^{1/\alpha}}. \quad (6)$$

## B. Joint optimal storage capacity and capacity sharing

### 1) Main Result:

**Proposition 2:** There exists a unique cache capacity  $C$  maximizing revenue. It is strictly positive if and only if storage costs are such that  $s'(0) < \max_{i=1,\dots,n} (p_i^c - r_i^f) F_i(0)$ .

Moreover, the cache capacity that the CDN should allocate to each CP  $i$  is also unique, and satisfies

$$C_i^{\text{opt}} = F_i^{(-1)} \left( \frac{s' \left( \sum_{j=1}^n C_j^{\text{opt}} \right)}{p_i^c - r_i^f} \right), \quad (7)$$

with  $F_i^{(-1)}$  the inverse of  $F_i$  and  $s'$  the derivative of  $s$ .

In particular, if the unit storage cost is a constant  $q_s > 0$ , then

$$C_i^{\text{opt}} = F_i^{(-1)} \left( \frac{q_s}{p_i^c - r_i^f} \right).$$

With respect to Proposition 1, our optimization model here jointly solves the problem of sharing among CPs and the problem of determining the optimal  $C$ .

**Proof.** The problem faced by the CDN actor is still as in (5), but where now the total capacity  $C$  is part of the decision variables. As before, the whole storage capacity is used, so  $C = \sum_j C_j$ . We rewrite the optimization problem (without the constant part) as

$$\begin{aligned} \max_{C_1, \dots, C_n} \quad & \sum_{i=1}^n (p_i^c - r_i^f) G_i(C_i) - s \left( \sum_{i=1}^n C_i \right) \quad (8) \\ \text{s.t.} \quad & C_i \geq 0 \quad 1 \leq i \leq n. \end{aligned}$$

The set of constraints again defines a convex set, which we can limit to a bounded subset since  $C_i \leq V_i$  (allocating to a CP more space than its total data volume brings only extra costs and no benefits). Consequently, since the objective function is strictly concave on the compact set  $\Pi_i[0, V_i]$ , the problem has a unique solution. For each  $i$ , the Karush-Kuhn-Tucker conditions imply that this solution  $(C_i^{\text{opt}})_{1 \leq i \leq n}$  satisfies

$$\lambda_i = \underbrace{(p_i^c - r_i^f) G'_i(C_i^{\text{opt}})}_{=F_i(C_i^{\text{opt}})} - s' \left( \sum_{j=1}^n C_j^{\text{opt}} \right) \quad (9)$$

$$\lambda_i \geq 0 \quad (10)$$

$$\lambda_i C_i = 0, \quad (11)$$

where for  $1 \leq i \leq n$ ,  $\lambda_i$  is the Lagrange multiplier corresponding to the constraint  $C_i \geq 0$ . We then immediately get from the complementary slackness condition (11) that  $C_i^{\text{opt}} > 0$  implies  $\lambda_i = 0$  and thus from (9) that

$$(p_i^c - r_i^f) F_i(C_i^{\text{opt}}) = s' \left( \sum_{j=1}^n C_j^{\text{opt}} \right),$$

which is (7). If  $C_i^{\text{opt}} = 0$  then we have from (10) and (9) that

$$F_i(0) \leq \frac{s' \left( \sum_{i=1}^n C_i^{\text{opt}} \right)}{p_i^c - r_i^f},$$

which is also consistent with (7).

In particular, if  $C_i^{\text{opt}} = 0$  for all  $i$  then this means that  $s'(0) \geq \max_i F_i(0) p_i^c - r_i^f$ . Reciprocally, assume  $s'(0) \geq F_i(0)(p_i^c - r_i^f)$  for all  $i$ : then since  $s'$  is assumed increasing, (7) implies that for all  $i$ ,  $C_i^{\text{opt}} = 0$ . ■

2) *Application to CPs with power-law popularities:* Considering again the case of power-law distributions and constant unit storage cost  $q_s$ , the optimal dimensioning of the storage space devoted to CP  $i$  content is simply

$$C_i = \left( \frac{(p_i^c - r_i^f) A_i}{q_s} \right)^{1/\alpha}. \quad (12)$$

We notice from (12) that a CP that is  $k$  times less popular than another (say,  $A_1 = A_2/k$ ) can still get the same space in the CDN cache if the benefit to the CDN is  $k$  times larger (i.e.,  $p_1^c - r_1^f = k(p_2^c - r_2^f)$ ).

## IV. NUMERICAL ANALYSIS

In this Section, we use our model to study how a CDN impacts the competition among CPs in various configurations. For simplicity and clarity sakes, we limit our analysis to two CPs. Our analysis is in two steps:

- First we use synthetic configurations where both CPs have a content popularity that follows a power-law. Since we aim at studying possible competition distortion, we consider two CPs having similar popularity settings. Our goal is to show here the differences of treatment that a revenue-driven CDN can generate.
- Second we use real traces for two different video providers: one VoD provider with a large catalog of long movies, and a free video provider, which offers a smaller catalog of short videos. Here, our goal is to study the impact of the “fair” policies that the regulators may enforce on the CDN, based on what is described in Section II-F.

### A. Analysis Based on Synthetic Configurations

We first consider the case of two identical CPs in the sense that the request rates among pieces of content follow the same distribution:

$$F_1(x) = F_2(x) = Ax^{-\alpha}.$$

By default, the *peak popularity*  $A$  is set to 10,000 and the *skewed parameter*  $\alpha$  to 0.8. The default values that we set in

the remainder of this numerical analysis are summarized in Table I. Note that we have repeated the analysis with other realistic parameters and obtained similar qualitative results.

We consider the case where one of the origin servers, here the origin server of CP<sub>2</sub>, is not as well connected to this ISP as the other origin server, here the one of CP<sub>1</sub>. The consequences are that the transit cost to reach CP<sub>2</sub>'s origin server is larger than to reach CP<sub>1</sub>'s origin server ( $q_2 > q_1$ ). In this configuration, from (7), a revenue-maximizing CDN tends to favor CP<sub>2</sub> in order to reduce the number of requests treated from CP<sub>2</sub>'s origin server. Indeed, the benefit for serving a request from CP<sub>1</sub> origin server is greater than the benefit from serving a request from CP<sub>2</sub> origin server:  $r_1^f = p_1^f - q_1 > r_2^f = p_2^f - q_2$ . Hence the CDN is incentivized to store in priority the content from CP<sub>2</sub> in its edge server.

QoE for cache $Q^c$	5	
QoE for origin server $Q^f$	3.5	
storage cost $s(x)$	1000 $\times$ $x$	
size of the cache	100	
skewed parameter $\alpha$	0.8	
	CP <sub>1</sub>	CP <sub>2</sub>
transit cost $q_i$	0.2	0.4
hit revenues $p_i^c$	1	1
miss revenues $p_i^f$	0.8	0.8
volume of content $V_i$	1000	1000

TABLE I  
DEFAULT PARAMETERS FOR OUR NUMERICAL ANALYSIS

We show in Figure 2 the ratio  $C_1/(C_1 + C_2)$  of content from CP<sub>1</sub> in the cache for different values of the skewed parameter  $\alpha$ , and for three settings. The reference case in plain line is the revenue-maximizing implementation of a CDN. The two other settings correspond to configurations where one of the CPs implements an aggressive pricing policy, here by paying more for accesses to their content from the edge servers,  $p_i^c = 1.3$  instead of  $p_i^c = 1$ . The skewed parameter ranges from 0.4 (which corresponds to a rather homogeneous popularity distribution) to 1 (which is a highly-skewed popularity distribution).

We highlight two observations from this figure. First, the revenue-maximizing CDN increases the cache hit of CP<sub>2</sub> in an unfair manner with regards to CP<sub>1</sub> by allocating more cache space to CP<sub>2</sub> than if ignoring monetary aspects, a stance we call *inclination* toward CP<sub>2</sub>. This inclination is especially strong for low skewed parameter: almost three quarters of the edge server are filled with CP<sub>2</sub> content for  $\alpha = 0.4$ . For skewer distributions of popularity, the CDN has more benefits storing the most popular content from CP<sub>1</sub> rather than some less popular content from CP<sub>2</sub> despite the inclination toward CP<sub>2</sub>. So the ratio of content of CP<sub>1</sub> in the edge server increases, but it is still less than 0.4 for highly skewed popularity distribution. The slight monetary differences between CPs (higher transit cost) make the CDN being in favor of CP<sub>2</sub> over CP<sub>1</sub> although both CPs are identical regarding both the popularity of content and their fees.

Our second observation is that this inclination toward one CP can be either amplified or reduced/inverted by more aggressive policies regarding the fees from CP<sub>2</sub> or CP<sub>1</sub> respectively. To represent it, we keep one of the CPs with the default

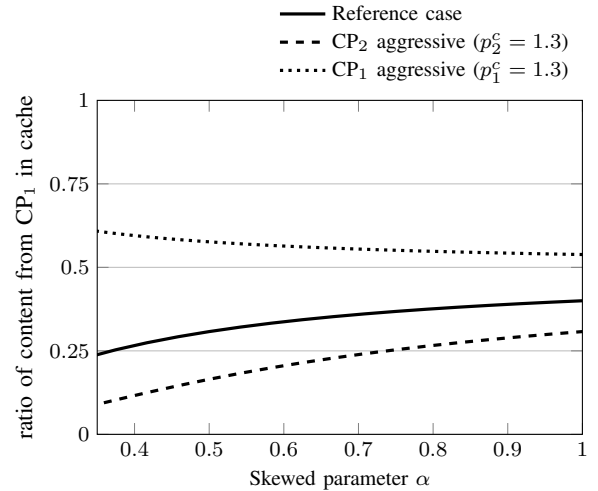


Fig. 2. Ratio of content from CP<sub>1</sub> (formally  $C_1^{\text{opt}}/C$ ) for various values of the skewed parameter  $\alpha$ . The reference case corresponds to a revenue-maximizing policy with  $p_1^c = p_2^c (= 1)$ . The “aggressive” lines correspond to one CP paying more for caching its content.

parameters while we change the price setting for the other CP to  $p_i^c = 1.3$ . We say that the CP is financially “aggressive”: it pays more for cache-served content, putting more pressure on the CDN to store its content in the edge server. In Figure 2, we show that when CP<sub>2</sub> is aggressive, the inclination toward CP<sub>2</sub> stays high (more than 60% of cached content) even for heterogeneous popularities ( $\alpha > 0.8$ ). CP<sub>1</sub> can pay to get more cached content, especially for homogeneous popularity ( $\alpha$  small), but to a lesser extent than what CP<sub>2</sub> can obtain by being aggressive.

We study now the impact of this inclination on the QoE of the respective end-users. The difference between both CPs is directly related to the concerns about equal treatment among CPs. Figure 3 displays the average QoE for each CP when one of the CPs, here CP<sub>2</sub>, pays for hits (*i.e.*, for requests that are treated by the cache) different cache-hit prices  $p_2^c$ . On the left part of the thin vertical line, the CP<sub>2</sub> is less aggressive than CP<sub>1</sub> ( $p_2^c < p_1^c$ ) while CP<sub>2</sub> is more aggressive on the right part of the figure.

As could be expected, the larger  $p_2^c$ , the more cache space devoted to CP<sub>2</sub> content, and naturally the better QoE for CP<sub>2</sub>'s end-users. Also, since CP<sub>2</sub> is already privileged because of higher transit costs from CP<sub>2</sub>'s origin server, CP<sub>2</sub> does not need to pay as much as CP<sub>1</sub> to provide a similar QoE as CP<sub>1</sub>: both QoE intersect when  $p_2^c$  is lower than  $p_1^c$ .

Another observation from Figure 3 is about the impact of CP<sub>2</sub> aggressive policy on CP<sub>1</sub> QoE. When the price paid by CP<sub>2</sub> is reasonable (typically after the crossing point, *i.e.*, for  $0.9 < p_2^c < 2.5$ ), the hit-ratio for CP<sub>2</sub> is almost constant, while that for CP<sub>1</sub> significantly decreases as CP<sub>2</sub> pays more. Even if not directly improving its own QoE significantly, paying the (revenue-maximizing) CDN a high price allows a CP to harm its competitor.

Finally, let us now consider the case where one CP is a bigger player in terms of global audience. Here the overall audience is given by the parameter  $A_i$ . The peak popularity



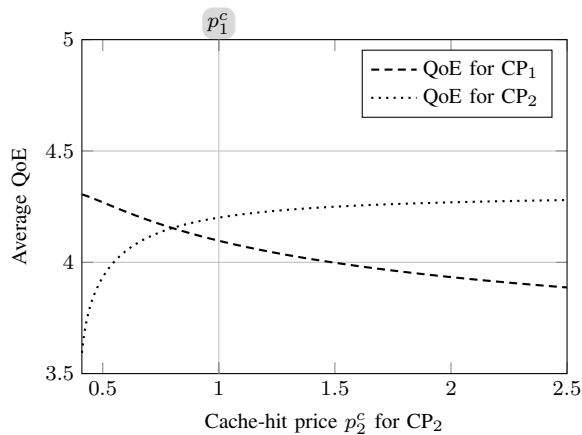


Fig. 3. Average Quality of Experience (QoE) vs. cache-hit price  $p_2^c$  paid by  $CP_2$ . Both CPs have identical popularity distributions. The transit cost of  $CP_2$  is higher than that of  $CP_1$ .

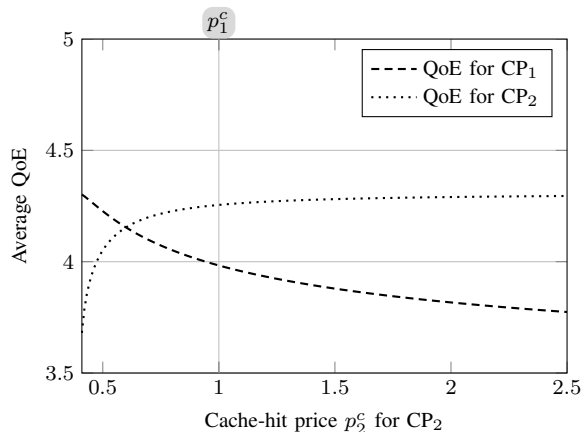


Fig. 4. Average Quality of Experience (QoE) vs. cache-hit price  $p_2^c$  paid by  $CP_2$ .  $CP_2$  is twice larger than  $CP_1$  ( $A_2 = 2A_1$ ). The transit cost of  $CP_2$  is higher than the one of  $CP_1$ .

$A_1$  of  $CP_1$  is again set to 10,000, while that of  $CP_2$  is twice larger,  $A_2 = 20,000$ . In Figure 4, we show the same figure as in Figure 3 where the cache-hit price of  $CP_2$  varies. When comparing both figures, we observe that the lines of QoE intersect for lower values of  $p_2^c$  when  $CP_2$  is bigger (around 0.6, compared with 0.8 in Figure 3). This can be interpreted as the CDN having an even bigger inclination toward  $CP_2$  when this provider is bigger. We also observe that, by increasing the cache-hit price  $p_2^c$ ,  $CP_2$  does not increase significantly the QoE for its end-users, but the impact it has on the QoE of  $CP_1$  end-users is much bigger than when both CPs are identical in size. Here, the decreasing of  $CP_1$ 's QoE is sharper than in Figure 3 and the overall QoE of  $CP_1$  reaches 3.75 for high values of  $p_2^c$ .

### B. Analysis Based on Real Traces

We now use real data traces from Orange, an ISP but also a large VoD provider, collected over two years (2012-2014) for a total of about 1.8 million views over 44,000 video files, from which we derive the empirical distributions of popularities from the number of requests for each file. Some

of the requested videos are *movies* (with duration around 100 minutes), which required the end-users to pay for watching, while some others are movie *trailers* (with duration 5 minutes), which anybody can watch for free. In the following, we consider that movies and trailers are proposed by two different CPs:  $CP_1$  provides movies, and  $CP_2$  provides trailers.

To have content elements of the same size, and also include the fact that users rarely watch a movie entirely, we separate each movie into 20 *chunks* of the same size as a trailer. The number of views for the  $i^{\text{th}}$  chunk ( $i = 1, \dots, 20$ ) is then generated by multiplying the number of requests for the movie with the probability of watching more than a proportion  $i/20$  of the video, as observed in [34]. This allows us to build two distributions,  $F_1(\cdot)$  (for movies) and  $F_2(\cdot)$  (for trailers).

The scenario we consider in the following emphasizes some of the issues that the CDN has to consider when dealing with heterogeneous CPs. The main parameters, which are typical for such configuration, are summarized in Table II.  $CP_1$  (movies) has a catalog of videos that is two orders of magnitude bigger than the catalog of  $CP_2$ . Nevertheless, the popularity of the most popular content of both CPs is in the same range. They are competitors although the volume of traffic generated by  $CP_1$  is bigger than the one generated by  $CP_2$ . The consequence is that the negotiation between CPs and the CDN does not result in the same arrangement, here we assume that  $CP_2$  pays ten times more than  $CP_1$ . Since we focus on the sharing of the edge servers, we put low revenues for videos that are missed by the cache. Finally, to highlight some critical trends of such a configuration, we consider that  $CP_2$  is more concerned by the QoE of the users. We reflect it by setting different values for the QoE for both CPs.

	$CP_1$ (movies)	$CP_2$ (trailers)
Number of items	743340	7337
Number of views for the most popular item	2386	2590
QoE for content accessed from cache $Q_i^c$	4.5	5
QoE for content accessed from origin server $Q_i^o$	3.5	3
hit revenues $p_i^c$	1	10
miss revenues $p_i^o$	0.01	0.01
transit cost $q_i$	0.01	0.01

TABLE II

DEFAULT PARAMETERS FOR THE NUMERICAL ANALYSIS

Our goal is then to study the behavior of the revenue-based CDN against other policies that are not driven by the revenue. Based on the details given in Section II-F, we study the following policies:

- **Popularity-driven:** only the popularity of content matters
- **Equal-sharing:** both CPs have the same amount of content in the cache
- **Equal-QoE:** the average QoE is the same for the users of both CPs
- **Equal-hit-ratio:** the hit-ratio is the same for requests to both CPs

and we compare all those policies with the revenue-maximizing one. In each case, we compute numerically the allocation corresponding to each criterion, using the empirical distributions  $F_1(\cdot)$  and  $F_2(\cdot)$ . For example we find the revenue-maximizing allocation by solving (3)-(4) by dichotomy on  $C_1$ .

We first illustrate the revenues generated by the different policies in Figure 5. We compute the maximum revenues, which is achieved by the revenue-driven policy, and we normalize the revenues for all the other policies. As seen in Figure 5, the other policies, while they make sense in terms of guaranteeing some form of fairness, can severely impact the revenue of the CDN. Typically, for this configuration, the revenues for the CDN are four times less when imposing the *Equal Hit-Ratio* policy as compared with the revenue-based one. Let us also highlight the performances of the *Popularity-based* policy, which is often depicted as a reasonable option since it is the most efficient regarding the traffic reduction (transit traffic from/to the edge-server) and it does not consider monetary exchanges. Unfortunately, we show in Figure 5 that the implementation of this policy corresponds to a drop of revenues ranging from half to one quarter for the CDN. Note that the drop can be less in other price configurations or with more homogeneous CPs.

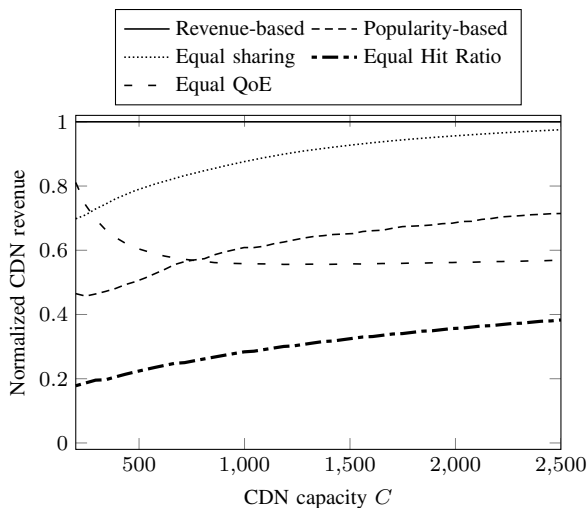


Fig. 5. Normalized CDN revenue when the cache capacity varies, relative to the revenue-maximizing policy.

To understand the main reasons behind the revenue differences, we present in Figure 6 the ratio of content from CP<sub>1</sub> in the cache. The different policies result in a different sharing of the cache storage among CPs. Of course, the *Equal sharing* policy divides the storage into two similar subsets. On the contrary, the *Equal hit-ratio* reserves almost all the cache to CP<sub>1</sub> because it has to deal with the volume of requests for CP<sub>1</sub>, which is one hundred times larger than for CP<sub>2</sub>, although the most popular videos have equivalent popularities. Finally, the other extremal policy is the *Revenue-driven*, which considers that the videos from CP<sub>2</sub> generate larger revenues, so, since the popularity of content is similar, the trailers are given the highest caching priority. The ratio of content from CP<sub>1</sub> increases when all the popular trailers are cached (CDN capacity greater than 1,000 here).

Finally, we set the CDN cache capacity to 1,000, and we make the hit-price of CP<sub>2</sub>,  $p_2^c$ , vary from 1 (which is equal to  $p_1^c$ ) to 20. We show in Figure 7 the average QoE for the end-users of both CPs. In that case, all the policies except the *Revenue-based* one have a constant average QoE since they

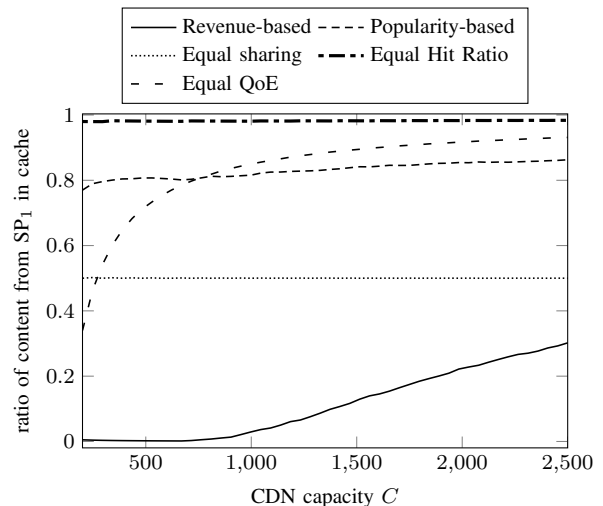


Fig. 6. Proportion of cache capacity for CP<sub>1</sub> when the cache capacity varies

do not depend on the revenues. In that case, the price paid by CP<sub>2</sub> is only based on negotiation with the CDN without any impact on the overall QoE for its end-users. We make three observations. First, in our configuration (please recall that the dataset is extracted from a real VoD provider), the large volume of requests for movies and the low popularity of the most popular movies with respect to the volume of requests make that the cache of the CDN has not a big impact on the average QoE of the end-users of CP<sub>1</sub>. The average QoE is slightly greater than 3.5 regardless of the CDN policy. However, and this is our second observation, the policy has a significant impact on CP<sub>2</sub>. The implementation of the *Equal hit-ratio*, which favors the high-volume CP<sub>1</sub>, results in a overall disastrous average QoE of 3.1 although the *Equal sharing* policy achieves a QoE of 4.1 in average. Finally our third observation is about the average QoE of the *Revenue-based* policy, which is equal to the *Popularity-based* policy when  $p_2^c = p_1^c$  but quickly increases to reach a QoE of 4.5.

To summarize our findings in this evaluation, the other options for a more fair CDN management policy should be carefully studied since their impact on the cache management can introduce major bias among competitors. The "fair" policy that is the most reasonable is the one based on content popularity, because it is both insensitive to monetary pressures and it is the most efficient regarding traffic reduction. However this policy can result in a severe drop of revenues for the CDN, and it does not allow one of the competitors to invest in the CDN infrastructure to offer a better QoE for its end-users.

## V. LINK WITH THE NETWORK NEUTRALITY DEBATE

As our results show, economic considerations can lead CDNs to treat differently some comparable CPs. Indeed, a CP paying more can benefit from preferential storage in the cache, hence from a better quality, the equivalent of a "fast-lane" with respect to the other CPs. And even for two CPs paying the same prices for the service, an unequal treatment can result from differences in transit costs. In our opinion, those phenomena are precisely what regulators want to avoid in

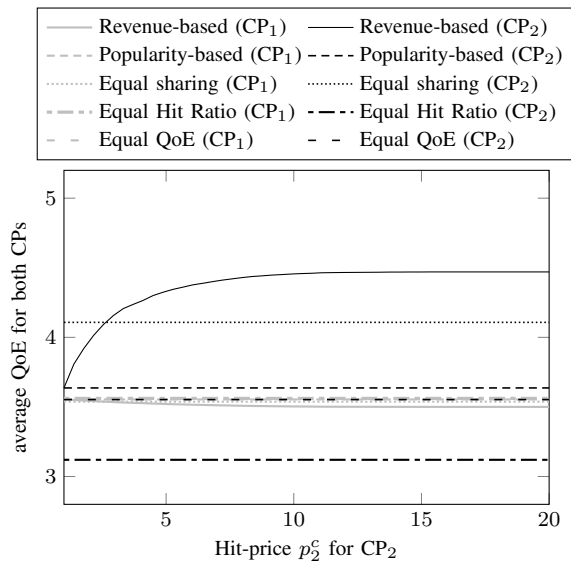


Fig. 7. Quality of Experience, when the hit-revenue of  $CP_2$ ,  $p_2^c$ , varies. Note that the hit-revenue  $p_1^c$  is equal to one.

the current *network neutrality* [35]–[38] and *search neutrality* debates [39], [40], hence we think it is worth considering those actors in the discussions. To deal with net neutrality, regulators all over the world have launched public consultations and issued recommendations. Yet, these recommendations barely mentioned the place of CDNs: for example, CDNs are absent in the French set of recommendations [41]. To our knowledge, the only mention of CDNs in the responses to national consultations is in the Norwegian one [42], where it is stated that “the ordinary use of CDN servers is not a breach of net neutrality. However the Norwegian Post and Telecommunications Authority<sup>3</sup> has not expressed a *carte blanche* that any CDN implementation is net neutral.” This lack of recommendations regarding CDNs is surprising given that the CDN weight in the content delivery business is increasing with respect to other actors such as transit providers, and their payment can be seen as a substitute of the disparaged side payment asked by ISPs. In line with our previous remark, many recent works have dealt with modeling and analyzing the interactions between CPs and ISPs to address the network neutrality debate, and sometimes propose regulation remedies [43]–[47]; but none incorporates CDNs into the picture yet.

The question of whether a CDN may breach neutrality in the CP competition shares some similarities with the case of ISPs. A CDN is a shared infrastructure, which has become necessary to deliver content at a worldwide scale. Since the delivery capacity of CDN is physically limited, the implementation of privileged partnerships with some CPs (similar as *fast lanes* by ISP) is detrimental to standard other CPs. Finally, given the nature of the considered CDN-CP partnerships, the lack of regulation may favor incumbent CPs over newcomers, an argument often raised to justify net neutrality regulation for ISPs. Our goal in this paper is to formally study the interplay of CDNs with regards of CP competition.

<sup>3</sup>now Norwegian Communications Authority

In the previous sections, we have highlighted situations where a CP obtains a preferential treatment because of economic considerations, which is against the principle of neutrality, although not against the technical rules that regulators may enforce [48]. Hence we think those situations are worth considering within debates connected to neutrality.

The theoretical contribution presented in this paper can be further used to address the questions of whether CDNs represent a breach of neutrality and whether a regulation should apply to CDNs.

## VI. CONCLUSION

CDN providers have become key players of the Internet delivery chain of value, but, despite this prominent position, CDNs are rarely identified as potential breaches of neutrality for packet treatment. We define in this paper a model, which integrates the main elements of the value chain for the content providers and the CDN. We show that a CDN whose purpose is to maximize its profit implements infrastructure management policies that can be regarded as unfair. In particular, our model shows that the inclination of the CDN toward the most paying content provider has an impact on the Quality of Experience of the end-users of competing content providers. Hence we think the role of CDNs regarding the net neutrality debate is ambiguous, and we suggest not to ignore those actors in the discussions.

In the last part of this paper, we use a real data set and compare the performance of the revenue-maximizing policy with various fairness-based policies. Those policies may induce large revenue reductions and may not be implemented by the CDN unless enforced by a regulator.

This paper opens some significant perspectives regarding the treatment of neutrality problems. We showed that, despite the most common belief, many actors should be concerned by neutrality issues, and especially the CDNs since a rational profit-driven management policy can result in unequal treatments among identical content providers.

In future works, we will continue our investigation toward a better understanding of the net neutrality debate. We will in particular study the problem caused by vertical integration of CDN into ISPs. Also, an important extension of the present work would be to consider how the prices for the CDN service are set: depending on whether we have a monopoly, an oligopoly, or perfect competition among CDN actors, different prices would be attained. We would like to investigate that extra step of price setting, and the impact of the market structure on the resulting service (or price) differentiation.

## ACKNOWLEDGEMENT

A preliminary version of this paper was published in the Proceedings of the 10th International Conference on Network and Service Management (CNSM), IEEE, Rio de Janeiro, Brazil, Nov 17-21, 2014.

The authors acknowledge the support of Orange. Note however that the content presented in this paper is solely the responsibility of the authors.

APPENDIX A  
CONTENT POPULARITY FROM A POWER-LAW  
DISTRIBUTION

Let us assume a CP  $i$  whose content popularities follow a power law, and express the expected number of downloads per time unit  $F_i(x)$  for its  $x^{\text{th}}$  most popular unit of content. By definition of the power-law, there exist  $\tau$ ,  $K_i$  and  $y_{\min}$  such that for a given piece of content, the density probability for the number of downloads (per time unit) at the value  $y$  equals

$$p(y) = K_i y^{-\tau} \quad \text{if } y \geq y_{\min}.$$

For  $y < y_{\min}$  we are out of the validity region of the power law.

Therefore, the probability that a given piece of content is requested at least  $y \geq y_{\min}$  times equals

$$\int_{t=y}^{\infty} p(t) dt = K_i \int_{t=y}^{\infty} t^{-\tau} dt = K_i \frac{y^{1-\tau}}{\tau + 1}.$$

Recall that  $V_i$  is the total volume of data of CP  $i$ . Picking randomly and uniformly a piece of content (assumed infinitesimal) in that amount, for any  $x \in [0, V_i]$  the probability that this piece has at least  $F_i(x)$  downloads is the probability that it is among the  $x$  most popular content from CP  $i$ . Hence it is just  $x/V_i$ .

Take  $y \geq y_{\min}$ , such that there is  $x \in (0, V_i]$  such that  $y = F_i(x)$  (note that such an  $x$  is unique because of the strict decreasingness of  $F_i$ ). From our computation above we have  $x/V_i = K_i \frac{F_i(x)^{1-\tau}}{\tau+1}$ , or equivalently

$$F_i(x) = \left( \frac{x(\tau + 1)}{V_i K_i} \right)^{\frac{1}{1-\tau}}.$$

Summarizing,  $F_i$  is of the form  $F_i(x) = A_i x^{-1/(\tau-1)}$  when that expression is above  $y_{\min}$ . We can then go back in the proof to check that with that expression, for any  $y \geq y_{\min}$  there exists  $x$  such that  $y = F_i(x)$ .

REFERENCES

- [1] DrPeering, "A business case for peering in 2010," 2010. [Online]. Available: <http://drpeering.net/white-papers/A-Business-Case-For-Peering.php>
- [2] —, "Internet transit prices - historical and projected," 2010. [Online]. Available: <http://drpeering.net/white-papers/Internet-Transit-Pricing-Historical-And-Projected.php>
- [3] A. Kovaks, "Internet peering and transit," Technology Policy Institute, Tech. Rep., 2012. [Online]. Available: <http://www.techpolicyinstitute.org/files/amkinternetpeeringandtransit.pdf>
- [4] H. Saunders, "IP interconnection: trends and emerging issues," 2012. [Online]. Available: [http://berec.europa.eu/files/news/ofcom\\_ipic.pdf](http://berec.europa.eu/files/news/ofcom_ipic.pdf)
- [5] V. Bonneau, "Evolution of the CDN market," in *CDN World Summit*, 2010.
- [6] B. Schwarz, "Content delivery networks 3.0," CTOiC White paper, Tech. Rep., 2013. [Online]. Available: [http://www.broadpeak.tv/upload/documentation/fichier/73-264-live\\_webtv\\_white\\_paper\\_b\\_schwarz\\_201212.pdf](http://www.broadpeak.tv/upload/documentation/fichier/73-264-live_webtv_white_paper_b_schwarz_201212.pdf)
- [7] M. Yu, W. Jiang, H. Li, and I. Stoica, "Tradeoffs in CDN designs for throughput oriented traffic," in *Proc. of ACM CoNEXT*, 2012.
- [8] K. Hosanagar, R. Krishnan, M. Smith, and J. Chuang, "Optimal pricing of content delivery network (CDN) services," in *Proc. of the Annual Hawaii Int. Conf. on System Sciences*, 2004.
- [9] K. Hosanagar, J. Chuang, R. Krishnan, and M. Smith, "Service adoption and pricing of content delivery network (CDN) services," *Management Science*, vol. 54, no. 9, pp. 1579–1593, 2008.

- [10] K. Hosanagar, "CDN pricing," in *Content Delivery Networks*, ser. Lecture Notes Electrical Engineering, R. Buyya, M. Pathan, and A. Vakali, Eds. Springer, 2008, vol. 9.
- [11] P. Faratin, "Economics of overlay networks: An industrial organization perspective on network economics," in *Proc. of ACM EC*, 2007.
- [12] H. Atzmon, R. Friedman, and R. Vitenberg, "Replacement policies for a distributed object caching service," in *Proc. of CoopIS*, 2002, pp. 661–674.
- [13] N. Christin, J. Chuang, and J. Grossklags, "Economics-informed design of CDNs," in *Content Delivery Networks*, ser. Lecture Notes Electrical Engineering, R. Buyya, M. Pathan, and A. Vakali, Eds. Springer Berlin Heidelberg, 2008, vol. 9, pp. 183–210. [Online]. Available: [http://dx.doi.org/10.1007/978-3-540-77887-5\\_7](http://dx.doi.org/10.1007/978-3-540-77887-5_7)
- [14] 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 the Twenty-third Annual ACM Symposium on Principles of Distributed Computing*, ser. PODC '04, New York, NY, USA: ACM, 2004, pp. 21–30. [Online]. Available: <http://doi.acm.org/10.1145/1011767.1011771>
- [15] P. Maillé, K. Pires, G. Simon, and B. Tuffin, "How neutral is a CDN? an economic approach," in *Proc. of CNSM*, 2014.
- [16] M. Fluczynski, "CDN Use Case Revisited," in *Proc. of OPNFV Summit*, 2015.
- [17] P. Maillé, G. Simon, and B. Tuffin, "Impact of revenue-driven CDN on the competition among network operators," in *Proc. of 11th International Conference on Network and Service Management (CNSM)*, Barcelona, Spain, 2015.
- [18] P. Maillé and B. Tuffin, "How do content delivery networks affect the economy of the internet and the network neutrality debate?" in *Proc. of 11th International Conference on Economics of Grids, Clouds, Systems and Services (GECON)*, Cardiff, UK, 2014.
- [19] —, "Impact of Content Delivery Networks on service and content innovation," in *Proc. of NetEcon*, Portland, OR, USA, 2015.
- [20] Cisco, "Wholesale Content Delivery Networks: Unlocking New Revenue Streams and Content Relationships," Cisco, Tech. Rep. 710667, 2011.
- [21] S. Puopolo, M. Latouche, F. L. Faucheur, , and J. Defour, "CDN Federations: How SPs Can Win the Battle for Content-Hungry Consumers," Cisco, Tech. Rep., 2011, "Point of View" White Paper.
- [22] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. MIT Press, 1994.
- [23] P. Maillé and B. Tuffin, *Telecommunication Network Economics: From Theory to Applications*. Cambridge University Press, 2014.
- [24] S. Hasan, S. Gorinsky, C. Dovrolis, and R. K. Sitaraman, "Trade-offs in optimizing the cache deployments of CDNs," in *Proc. of IEEE INFOCOM*, 2014.
- [25] R. T. B. Ma and D. Towsley, "Cashing in on caching: On-demand contract design with linear pricing," in *Proc. of CoNEXT*, 2015.
- [26] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs," *IEEE/ACM Trans. on Networking*, vol. 21, no. 6, pp. 2001–2014, 2013.
- [27] P. Casas, A. Sackl, S. Egger, and R. Schatz, "YouTube & Facebook Quality of Experience in mobile broadband networks," in *Proc. of IEEE Globecom Workshops*, 2012.
- [28] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of Quality of Experience for internet video," in *Proc. of ACM SIGCOMM*, 2013.
- [29] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, "Quantification of youtube qoe via crowdsourcing," in *Proc. of IEEE International Symposium on Multimedia, ISM*, 2011.
- [30] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the Quality of Experience of HTTP video streaming," in *Proc. of IFIP/IEEE IM*, 2011.
- [31] P. Casas, A. D'Alconzo, P. Fiadino, A. Bär, A. Finamore, and T. Zseby, "When youtube does not work - analysis of qoe-relevant degradation in google CDN traffic," *IEEE Trans. Network and Service Management*, vol. 11, no. 4, pp. 441–457, 2014.
- [32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [33] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. B. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Trans. on Networking*, vol. 17, no. 5, pp. 1357–1370, 2009.
- [34] A. Finamore, M. Mellia, M. M. Munafò, R. Torres, and S. G. Rao, "Youtube everywhere: Impact of device and infrastructure synergies on user experience," in *Proc. of Internet Measurement Conference (IMC)*, Berlin, Germany, 2011, pp. 345–360.

- [35] P. Crocioni, "Net neutrality in Europe: Desperately seeking a market failure," *Telecommunications Policy*, vol. 35, no. 1, pp. 1 – 11, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0308596110001461>
- [36] P. Maillé, P. Reichl, and B. Tuffin, "Internet governance and economics of network neutrality," in *Telecom. Eco. - Selected Results of the COST Action IS605 Econtel*, A. Hadjiantonis and B. Stiller, Eds. Springer Verlag, 2012, pp. 108–116.
- [37] T. Lenard and R. May (Eds.), *Net Neutrality or Net Neutering: Should Broadband Internet Services be Regulated*. Springer, 2006.
- [38] T. Wu, "Network neutrality, broadband discrimination," *Journal of Telecom. and High Tech.*, 2003.
- [39] J. Crowcroft, "Net neutrality: The technical side of the debate: A white paper," *ACM SIGCOMM Computer Communication Review*, vol. 7, no. 1, January 2007.
- [40] P. L'Ecuyer, P. Maillé, N. Stier-Moses, and B. Tuffin, "Revenue-maximizing rankings for online platforms with quality-sensitive consumers," *Operations Research*, 2017, in press.
- [41] ARCEP, "Les cahiers de l'ARCEP: Net neutralité: les 10 propositions de l'ARCEP," [http://www.arcep.fr/uploads/tx\\_gspublication/Cahiers\\_ARCEP\\_03\\_light.pdf](http://www.arcep.fr/uploads/tx_gspublication/Cahiers_ARCEP_03_light.pdf), 2010.
- [42] F. Sorensen, "The norwegian model for net neutrality," <http://is.gd/HtQuuJ>.
- [43] E. Altman, A. Legout, and Y. Xu, "Network non-neutrality debate: An economic analysis," in *Proc. of IFIP Networking*, 2011.
- [44] P. Coucheney, P. Maillé, and B. Tuffin, "Network Neutrality Debate and ISP Inter-Relations: Traffic Exchange, Revenue Sharing, and Disconnection Threat," Inria, Tech. Rep., 2012.
- [45] —, "Impact of Reputation-Sensitive Users and Competition Between ISPs on the Net Neutrality Debate," *IEEE Trans. on Network and Service Management*, vol. 4, no. 10, 2013.
- [46] R. Frieden, "Internet Packet Sniffing and its Impact on the Network Neutrality Debate and the Balance of Power between Intellectual Property Creators and Consumers," SSRN, Tech. Rep., 2007, <http://is.gd/FG978B>.
- [47] P. Njoroge, A. Ozdaglar, N. Stier-Moses, and G. Weintraub, "Investment in two sided markets and the net neutrality debate," Columbia Univ., Decision, Risk and Operations Working Papers Series, Tech. Rep. DRO-2010-05, 2010.
- [48] P. Maillé, G. Simon, and B. Tuffin, "Toward a net neutrality debate that conforms to the 2010s," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 94–99, March 2016.



**Eric Gourdin** obtained his Ph.D. in 1994 from Ecole Polytechnique de Montréal. He worked on a TMR Grant from 1996 to 1998 at ULB (Université Libre de Bruxelles) on traffic management problems. He joined Orange (former France Telecom) in 1998 where he has been working on various network optimization problems, with a special focus on IP networks. He has been in charge of OR (Operation Research) for several years and has lead various optimization oriented projects at Orange Labs. He has co-authored

many scientific papers and contributed to several books. His latest interest lies in routing problems, Content Delivery architectures and virtual networks. Since nov 2016, he is in charge of the department TRM (Traffic Resource Management).



**Patrick Maillé** graduated from Ecole polytechnique and Telecom ParisTech (France) in 2000 and 2002, respectively. He has been with the Networks, Security, Multimedia department of Telecom Bretagne (Institut Mines-Telecom) since 2002, where he obtained his PhD in applied mathematics in 2005 and is now a professor. His research interests are on economics and game theory concepts applied to networked ecosystems: resource pricing, routing, consequences of user selfishness on global performance, and regulation issues. He has co-authored more than 80 research papers on those topics, as well as the book "Telecommunication Network Economics" published by Cambridge University Press in 2014. He has held visiting scholar appointments at Columbia University (2006) and UC Berkeley (2014-2015). His email address is [patrick.maille@imt-atlantique.fr](mailto:patrick.maille@imt-atlantique.fr).



**Gwendal Simon** received his PhD in computer science from University Rennes 1 (France) in 2004. During his PhD, he worked at Orange Research Labs. From 2004 to 2006 he was a researcher at Orange Labs. Since 2006, he has been Associate Professor at Telecom Bretagne, a graduate engineering school within the Institut Mines-Telecom. He was a visiting researcher at University of Waterloo from September 2011 to September 2012. His research interests include multimedia delivery systems (video and gaming) and network management. He has published more than 70 papers in international conferences and journals.



**Bruno Tuffin** received his PhD degree in applied mathematics from the University of Rennes 1 (France) in 1997. Since then, he has been with Inria in Rennes. His research interests are twofolds: the development Monte Carlo simulation techniques and quasi-Monte Carlo simulation techniques for the performance evaluation of telecommunication systems, and network economics analysis. He is currently Simulation Area Editor for *INFORMS Journal on Computing* and Associate Editor for *ACM Transactions on Modeling and Computer Simulation*. He has written or co-written three books: *Rare event simulation using Monte Carlo methods* published by John Wiley & Sons in 2009, *La simulation de Monte Carlo* (in French), published by Hermes Editions in 2010, and *Telecommunication Network Economics: From Theory to Applications*, published by Cambridge University Press in 2014. His email address is [bruno.tuffin@inria.fr](mailto:bruno.tuffin@inria.fr).