Network Friendly Video Distribution

Zhe Li[‡], Mohamed Karim Sbai[‡], Yassine Hadjadj-Aoul[†], Annie Gravey[‡],
Damien Alliez[*], Jeremie Garnier[*], Gwendal Simon[‡], Kamal Singh[‡]
[‡]Institut Mines Télécom - Télécom Bretagne, France, {firstname.lastname}@telecom-bretagne.eu
[†]Inria Bretagne-Atlantique, France, yassine.hadjadj-aoul@irisa.fr
[*]NDS Limited, France, dalliez@nds.com, jGarnier@nds.com

***Abstract*—The rise of popularity of video streaming services has resulted in increased volumes of network traffic, which in turn have created Internet bottlenecks leading to perceived quality degradations. In this paper, we argue that a good way to tackle this type of congestion is to make the contents available inside ISPs' networks. We thus propose a network-friendly content delivery architecture that considers the complex video distribution chain and its associated business models. This comprehensive architecture allows a network operator to fully engineer video traffic distribution in order to both alleviate peering links' workload and improve delivered QoS. This proposal is fully compatible with Adaptive Bitrate Streaming (ABS) architectures, which are currently used to distribute video in the Internet. Extensive simulations show that the proposed architecture optimizes resources' use while clearly improving the quality of experience by reducing peering links' load and by shortening significantly the waiting time of clients. From an ISP's point of view, such an architecture promises dramatic savings on the costs associated with video delivery. The feasibility of the architecture is proved by a prototype implementation based on DASH ("Dynamic Adaptative Streaming over HTTP"), which is the MPEG version of ABS. This architecture actually presents an evolutionary and pragmatic method to efficiently deploy an Information Centric Networking (ICN) architecture based on the collaboration between service providers or traditional CDNs and peer-assisted CDNs operated by ISPs.**

I. Introduction

More than one billion of users access daily multimedia content. The sharp growth of content distribution has opened new business opportunities in particular for *Content Delivery Network* - CDN - providers. Acting as intermediaries between content providers and consumers, CDNs have proliferated rapidly with the explosive increment of Internet services. While CDNs clearly benefit from increasing business, they need to devise new methods for handling the drastically augmenting Internet traffic. In particular, recent studies show that congested peering links are nowadays the real bottleneck of the Internet [1], [2]. Moreover, the forecasted dramatic growth of both the number of connected devices (24 billions of connected devices in 2020 [3]) and the volume of video traffic (90 per cent of global consumer Internet traffic in 2015 [4]) calls for shifting some of the mechanisms that are currently in use in the Internet. Some networking experts propose to redesign the Internet through some clean-slate approaches as *Information Centric Network* (ICN). However, a naive, purely academic, vision of the evolution of Internet, that does not consider the main actors and their business models, may not be realistic.

In this paper, we show that an information-centric evolution of Internet should take account of the economic relationships between the content distribution Internet actors. We then argue in favour of a collaboration between the CDN providers and the Internet Service Providers (ISPs). Our analysis endorses some recent works about the convergence between ISPs and CDNs [5]. *Our first claim is that the most appealing location for the deployment of information-centric architecture is indeed the access networks (at the regional scale) under the joint management of both CDNs and ISPs.*

We also highlight the profound evolution of video delivery systems in the last couple of years. Indeed, it is now frequent that a video, referenced by a name, has multiple encoded forms. The client requests one of these forms according to both network conditions and client device characteristics. This new video delivery paradigm finds its roots in two recent business-driven developments. First, the broadening diversity of devices has forced the development of transcoding systems for matching the content with devices' characteristics (especially tablets' and smartphones'). Second, most major video providers have invested into *Adaptive Bitrate Streaming* (ABS) architectures where servers store multiple encodings of the same movies. The traffic related to ABS already represents about 50% of overall Internet traffic at peak time [7]. Many proprietary ABS architectures are currently used, while standard bodies (3GPP and MPEG) specify a standard ABS architecture, namely *Dynamic Adaptive Streaming over HTTP* (DASH) systems [6] on which we build our proposal.

In a new information-centric Internet, name resolution does not any longer aim at determining the location of one equipment (server or proxy) storing the content. It aims instead at determining one copy of the requested content. As highlighted in [8], the HTTP protocol provides the flexibility required by the service providers (and their allies the CDN providers) to transform a human-readable name into the location of a server storing the most adequate content. This leads to *our second claim, namely that name resolution in an ISP level Information-centric network should be done at the application layer by an actor who has a comprehensive view of the candidate alternatives.*

Based on these two claims, we present a **Network Friendly DASH** (NF-DASH) architecture, which aims at building up an application-layer ICN for video traffic engineering in ISPs' networks. NF-DASH leverages on a peer-assisted system or "distributed CDN" (hereafter referred to as *dCDN*), which is an overlay controlled by the network operator. The objective of a dCDN is to facilitate the content dissemination within a regional network (typically an ISP's network). This objective leads to a neat delineation between the role of a CDN and that of a dCDN: a CDN is in charge of disseminating content at a wide scale (say, at the international level) while a dCDN aims at delivering content within a medium size network (say, regional). The ISP is thus able to engineer video traffic over network and to implement DASH in a network friendly fashion.

The remainder of this paper is organized as follows. Section II provides the economic and technological context and outlines the motivations behind the NF-DASH architecture. Section III portrays the components of the envisioned dCDN and illustrates the function of the whole system. In Section IV, we discern the feasibility of NF-DASH through a short introduction of implementation and evaluate its performance by simulation. Finally, the paper concludes in Section V with a summary recapping the main advantages of the proposed architecture.

II. Context and Motivations

A. Business Consideration

1) Recent market evolutions: As the development of Internet goes on, four major market evolutions have recently changed the business of CDN providers:

- The ISP market matured and reached denser and more complex levels of dependencies in peering with other networks. In the meantime, transit costs drastically decayed, especially through the development of direct peering between ISPs. Consequently, the CDN "market need" from the ISP side has at best stabilized.
- The amount of traffic from CDN servers increased, bottlenecks appeared in the peering links toward CDN. To overcome this problem, ISPs started to restrict CDN generated traffic, by raising (previously overlooked) variable costs. Especially, ISPs recently charged upstream bandwidth used from CDN infrastructure to ISP network and reduced the share of cache allocated to CDN business.
- Facing direct peering between ISPs, Internet Transit Providers (ITPs) developed their own caching capacities, drastically lowered CDN prices to gain market shares. ITPs also lowered the global market revenue potentials of CDN market at large.
- As the nature of Internet traffic evolved from web and asynchronous applications to video, low latency and real time services, the CDN business has evolved which challenges the traditional CDN mechanisms that were optimized for serving static HTML pages.

Confronting revenue cuts and changes in cost structures, the CDN business needs to reinvent itself. The original paradigm of a single third party between ISPs and service providers is also challenged: as bottlenecks moved over peering links, CDNs are now considering servicing their technological expertise instead of running CDN services themselves. The future of CDN businesses is likely to lie deeper within the ISP networks, and to be more integrated into and interleaved with ISP infrastructures.

A deeper integration of CDN into the regional network of ISPs raises numerous new issues. In particular, the share of investment into the key components of a distributed hosting infrastructure has to be re-considered. The traffic management (QoS at large) into an access network has little in common with what CDN providers are used to dealing with. From a more practical point of view, the serviceability of the key components (localization, maintenance, support) should be questioned.

Lastly, legal constraints appear when examining the value chain. Most prominent are the conditions of network and caching operation and thereby the players potentially involved in the funding of distributed caching services. The current debate about net neutrality (or quasi-neutrality) interferes with the CDN providers' business goals, potentially affecting then global value chain by applying stringent legal constraints.

2) Main Actors' Christmas Lists: In the absence of regulation, the positions of players in the value chain are clear at this stage:

Service providers want to serve any end-user without having to engage specific deals with ISPs (excluding potential exclusivity or syndication deals to be considered as particular use cases). They also want to maintain exclusive relationships with their clients. Since personalization is considered as a promising way to monetize services, the providers want to be notified of each and every action of their clients. Therefore traffic interception by un-authorized third-parties is not acceptable. Lastly, service providers wish to drastically lower their cost structure. Two approaches emerge. In the positive one, service providers assume a share of the required investment if this leads to build non-variable costs structures. They also respect distribution formats to match the preferred content delivery mode of network operators (as attested by the current development of ABS). In the negative one, service providers pressure ISPs to deliver revenues back in proportion to the popularity of the provided services.

CDN providers want to remain in the game by returning to more healthy levels of profitability, by getting rid of variable costs and by finding added value in content handling. CDN providers also want to participate to global infrastructure investments, in order to better balance and value their core assets (technologies, knowledge and current infrastructures).

ISPs want to maintain cost structures under control in order to remain competitive on the Internet access market (i.e. maintaining low pricing/level of margin). They also want to control the global QoS offered to certain services and to maintain a certain level of differentiation.

End-users want to access any service, at any time and from anywhere in the world. As most services are either free or cheap, end-users tend to confound the actors in the value chain, which leads to requests for one overall monthly bill in order to access any service on any device through any network.

This analysis highlights that many market players objectives converge (in spite of some disparities). The global interest of the value chain players is converging toward a better collaboration in order to match Internet evolution at large and its impact on a wider scope of industries such as the content industry. The present paper explores paths of potential collaborations to enable the next generation of distributed services.

B. Technological Consideration

After the assessment of latest business model evolutions, we now provide a more technology-oriented analysis of the recent development in the video delivery techniques.

1) HTTP as a Support for Video Services: HTTP/TCP tends to replace RTP/UDP as the main protocol for the delivery of video. In comparison to UDP flows, an HTTP encapsulation guarantees the seamless traversal of NAT and firewalls and offers data centric properties via human-readable widely-adopted named resources and transparent redirections. Multi-layered video coding over UDP has been proposed as an elegant way to design adaptive video distribution solutions, but the main video provider actors have opted for HTTP-based solutions. More generally, the HTTP protocol is replacing IP as the narrow waist of the Internet [8].

For video transmission, the first attempt to use HTTP at large scale was the *progressive download* architecture, which consists in downloading a video as fast as possible, while starting the playout before the complete content is received. This technique is commonly used to deliver short video documents (e.g. YouTube clips). However, progressive download is not adaptive since the encoding cannot be changed once the download has started. Users may experience degradations when link capacity does not support the selected encoding.

ABS architectures, including DASH, tackle the limitations of HTTP progressive downloading [6]. Most players have adopted and implemented ABS as their main content delivery protocol (Microsoft, Adobe, Apple and Netflix, which is today the dominant video provider [7]). Facing this plebiscite, the MPEG consortium has launched the process of standardizing DASH into MPEG [9].

In short, ABS requires the segmentation of a media file into *chunks* of same duration. Chunks are proposed in several formats, corresponding e.g. to different coding rates, and are encapsulated in HTTP (see Figure 1). A video server publishes documents (*manifests* and *playlists*) where each version of each chunk is identified by an URL or Byte-Range URL. The player requests successive chunks from the server; it estimates the bottleneck size between the server and itself and thus adapts its requests to available bandwidth, by requesting the appropriate coding for each chunk. This enables seamless switching from one coding rate to another when network conditions change.

Since video providers have recently massively invested in ABS architectures, any ICN proposal providing alternatives to HTTP-based solutions has to prove a significantly better efficiency than ABS. Moreover, a progressive deployment of any new ICN oriented architecture mandates the support of ABS architectures.

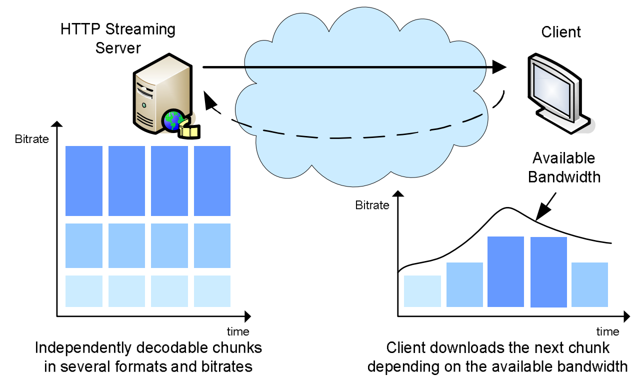

Fig. 1. Adaptive Bitrate Streaming Principle

2) Traffic Engineering for Adaptive Bitrate Streaming: The development of ABS forces ISPs to address new challenges. Because the quality of experience of popular video services is a key demand from their customers, ISPs need to ensure a good QoS for the traffic from delay-sensitive ABS. However, all HTTP traffic is currently considered by ISPs as Internet traffic and cannot be selectively recognized as video traffic to be prioritized.

Another problem is the location of bottleneck links. Intuitively, ABS adaptive algorithms work well when the bottleneck is the last-mile. However, when several concurrent ABS connections share the same bottleneck, clients incessantly adapt the flow bit-rate, and ABS adaptive algorithms may interfere with TCP adaptive congestion control.

The only practical solution for an ISP is thus to blindly increase both peering and access bandwidth in order to accomodate the ever-increasing HTTP traffic, which leads to constantly increasing investment and operational costs.

3) The Case for Application-Layer ICN: ABS (and DASH) seem to be rather incompatible with ICN's fundamental principles. For instance, an ABS connection does not support swarming (one client fetching a large video content from multiple servers) which is natural in currently considered ICN proposals. Indeed, when successive chunks come from different servers, congestion cannot be accurately measured and the adaptive algorithm may adopt an inconsistent behavior. Furthermore, if a single "name" is associated to many diversely encoded content, ICN's name resolution cannot apply to the complete video but to the video chunks. However the decision of which encoding to choose depends on the performances of the link from the client to the equipment that stores the chunk, so it depends on the location of the equipment, which contradicts the *a posteriori* content discovery process typical to ICN. Furthermore, caches' performance can be severely degraded if every content is composed of multiple independently cacheable files. Lastly, name resolution in ICN does not ensure that every request is received by the CDN and the service providers, whereas the latter actors consider this feature as mandatory for commercial reasons.

From the ISP's point of view, ICN clean slate approaches have yet to demonstrate traffic engineering capabilities. Some proposed ICN architectures require intrusive modifications to the communication stack or even the entire network architecture. Moreover, addressing, naming and routing issues are not yet solved to scale and need more research. In comparison, ABS architectures over the existing Internet infrastructure are currently deployed over the Internet.

In order to allow a smooth transition of video content delivery services, with incremental deployments, toward a ICN-like adaptive architecture, we suggest that an overlay architecture implementing an application layer design based on ABS principles is preferable to a clean slate ICN approach. The next section sketches out how NF-DASH integrates ICN principles with DASH.

III. Proposed Intra-Domain Content Delivery Service

We assume that there is a single ISP entity, which provides Internet access to end-users and controls the access network, from peering points to the last mile. Actually this is a common situation, in particular in Europe.

A. dCDN building blocks

ISPs can leverage on the next-generation switches and routers, which are equipped with caching capabilities. These network nodes are deployed and managed by the ISP, which thus has the opportunity to dive into the market of content delivery. Home gateways and set-top-boxes can also become content servers. The ISP owns at least one equipment in every home of its customer (the home gateway), and possibly more (the home gateway and one or several set-top boxes). In Europe, home gateways are powerful ISP-controlled equipments with large storage areas. Finally, ISP can directly manage servers or data-centers.

NF-DASH uses these reliable storage equipments which are referred as *dServers*. We are not the only one to notice the huge storage capacity that the ISP is now, or will soon, administer close to the end users. As illustrated in [5], other concurrent projects also consider building a CDN into the infrastructure of an ISP, that is, into a regional access network.

Most ISPs have also already deployed hardware and software tools that are in charge of monitoring the traffic, and available storage spaces. Access networks are operated, which means that the network topology and the operating equipments allow a centralized supervision of any large-scale service. The development of Software Defined Networking "SDN" technologies is yet another step toward a centralized traffic management. In other words, the ISP has already the capacity to "engineer" a vast overlay of storage equipments in the access network.

B. CDN interconnection

Another argument in favor of the management of a so-called dCDN inside the ISP infrastructure is the recent works toward a standardization of the inter-connection among CDN [10]. The collaboration with existing traditional CDNs is a critical point for new actors in the content delivery chain. In the VIPEER project[1], we call for reinforcing the collaboration between CDN providers and ISP operators.

- The role of the ISP is to bring its expertise in the management of access networks and to provide new storage spaces with guaranteed links to the customers,
- The role of the CDNs is to get and distribute contents from the service providers either by its own servers or by delegating this distribution to the dCDN.

This collaboration preserves the essence of the Internet, which is to let every network operator manage its own network. Actually, a network operator manages its own dCDN according to the characteristics of its network. For instance, an ISP currently investing in a new line of routers will be interested in massively using *in-network caching* strategies that leverage on the features of new routers, whereas an ISP presenting a large base of customers equipped with caching enabled home gateways will focus on mechanisms that exploit these resources. The aggregation of dCDNs cooperating with traditional CDNs form a information-centric network, which aligns with domain specific policies. Let us now highlight some features of such architecture.

C. dCDN Functional elements

The dCDN is composed by nodes that are interconnected by (virtual) links. The dCDN nodes form a virtual network which overlays the ISP's network, and which is fed by the client CDNs. The dCDN is composed of two types of entities: the *dServers* and the *dTracker* (see Figure 2).

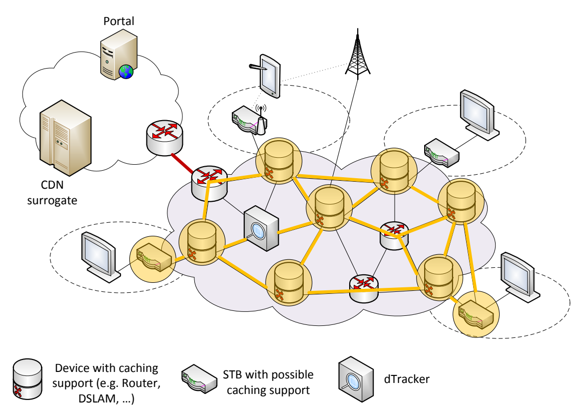

Fig. 2. Overview of the proposed architecture

A *dTracker* interfaces with the CDN. New content is injected in the dCDN via a well identified dTracker that

[1]The work described in this paper was carried out with the support of the French Agence Nationale de la Recherche in the framework of the VIPEER project (grant number ANR-09-VERS-014).

represents the interface between the CDN and the dCDN. When a CDN uploads content in the dCDN, it may provide popularity and identification of the previous requests regarding this content. The dTracker is responsible for controlling the chunks' placement on the dCDN nodes (i.e. the dServers). It maintains a list of those chunks that have been injected into the dCDN and the list of chunks stored by the different dServers. This knowledge is not shared with the CDN who has indeed delegated the content's distribution to the dCDN.

D. NF-DASH operation

Consider that a CDN and dCDN agree that a given content, say "*Matrix*", should now be served by the dCDN. Figure 3 illustrates how this is achieved. The CDN notifies the dCDN with the manifest so that the dCDN can prefetch the various chunks and store them on its own dServers. The dCDN should now generate a new manifest, which associates chunks (with a given video quality) with the network addresses of the dServers where these chunks are stored. The dCDN could also transcode the video in order to increase the number of video quality levels. The result of this first step is that a name, here "Matrix", is now associated with multiple IP-routable addresses of servers that collectively store multiple versions of the given Matrix movie.

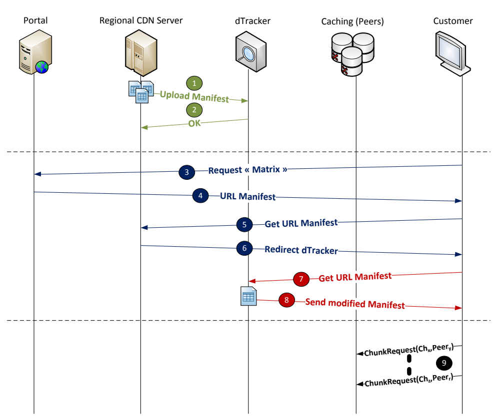

Fig. 3. Adaptive Streaming in a dCDN/CDN world

The remaining steps to achieve video distribution rely on classical request redirection processes : a request for "Matrix" from a customer is first sent to the video service provider, which redirects the customer's request to the CDN provider. The CDN provider in turn redirects the customer to the dTracker. The dTracker sets up an adaptive streaming session with the customer, through a modified manifest that takes into account the current performance of the network and the availability of the content in the dServers.

Such a scenario presents all the desirable properties of an ICN ; it is also respectful of all actors (video provider, CDN provider and ISP) and is deployable according to the agenda of each actor. It leverages on a better network resource utilization and the advances provided by adaptive streaming. It does not need any TCP/IP protocol revamping and offers the ISP the opportunity to rely on NF-DASH to implement lightweight traffic engineering policies. Finally, contrarily to [5], which is based on traffic interception, our solution does not bypass the service provider and the CDNs: they still received the requests from clients, so they are able to monitor and personalize the services.

IV. Evaluation and Implementation

This section presents simulation results that demonstrate the benefits of operating a dCDN, both from the users and the operators' point of view. It also describes a prototype implementation of NF-DASH.

A. Simulation Setup

In order to assess the impact of dCDN operation both on network performance and on delivered QoS, we developped an ad-hoc simulation program (in C) that tested several configurations for a video distribution service.

We used a one-week content download trace of Orange VoD service for the period June 12, 12 o'clock, to June 19, 12 o'clock, 2011. The structure of each record in the trace is illustrated in figure 4. The first field identifies a user and the second identifies the video or clip that is downloaded by the user. *Region* indicates the user's geographical location and the last field is the time when the video was requested. Since the duration of a download session is not given in the trace, we assumed that it followed a uniform distribution from 60 minutes to 120 minutes. The average session's duration is thus 90 minutes. The trace contained 364,663 download requests for 17,223 distinct videos. These requests come from users distributed in 13 geographical areas in France, which are considered to set up the dCDN topology.

user ID	*item ID*	*region*	*timestamp*

Fig. 4. Structure of trace record

We assume that each of the 13 regions contains a dServer. The dServers are identical with storage capacity of 5,000 videos and serving capacity of 500 users. The single CDN server is much more powerful than the dServers as it can store all required videos and can serve up to 5,000 users.

The dCDN operation is as follows : a user's request is directed by the dCDN to a dServer (according to one of the simulated policies specified later). If this dServer is saturated (i.e. already serves 500 requests), the request is forwarded to another server that stores a replica of the video (either another dServer or the CDN server). If the CDN server is saturated, the request is queued in order to postpone the video delivery. Only the CDN server can queue requests.

Network's topology is shown in figure 5. The red circle represents the CDN server located at the Point of Presence (PoP) in Paris. Each dServer is directly connected to the CDN server. These connections represent *external* links (i.e. toward CDN provider or transit networks). Traffics transfered on red links will be charged (either by the CDN provider or by a

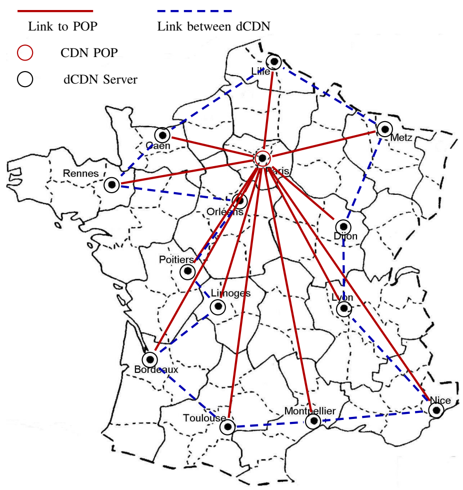

Fig. 5. Simulated dCDN topology; the cost of using an external (red line) link is twice the cost of using a peering (blue dashed line) link.

transit network provider), and thus generate external costs for the ISP.

In both placement and redirection policies, we may assume that dServers are organized in *cooperation groups*, where dServers in a given cooperation group are geographically close. For example, the cooperation group for the dServer located in Rennes includes the dServers in Caen and Orleans, while the cooperation group for the dServer located in Nice includes the dServers in Toulouse, Montpellier, Lyon and Dijon.

Each pair of geographically adjacent dServers is connected by an *internal* link shown by a dashed blue line in figure 5. Internal links are intra-domain links managed by the ISP, therefore cheaper to use than *external* links. Specifically, we assume that the cost for transferring a unit of data on a blue link is a half of the cost for transferring a unit of data on a red one.

B. Measurements

In order to highlight the benefits yielded by NF-DASH, we assess the impact of both placement and redirection policies on the cost and performance of the video retrieval service, by comparing it to a benchmark centralized solution where all requests are directly treated by CDN server. The considered NF-DASH configurations are the following combinations of placement and redirection settings:

- *Random (video) placement*, where a video is replicated on randomly selected dServers.
- *Optimal placement*, where each video follows an optimal placement strategy described in [11].
- *Random (request) redirection*, where each request is initially randomly assigned to a dServer. If either the dServer does not store the video, or if it is saturated, the request is redirected to the CDN server.
- *Optimal redirection*, where a user's request is first assigned to its closest dServer. If the request cannot be satisfied by this dServer, it will be redirected successively to the dServers in the cooperative group of the initially selected dServer. If none can deliver the requested video, the request is finally redirected to the CDN server.

We thus consider four implementations of NF-DASH differing on the placement and redirection policies: *Rand-Rand* and *Opt-Rand* respectively represent random and optimal placement with random redirection whereas *Rand-Opt* and *Opt-Opt* respectively represent random and optimal placement with optimal redirection. Note that the CDN is responsible neither for placing the videos in the dServers nor for redirecting the requests to appropriate dServers.

The metric used to assess the cost of serving a single video is the geographical distance between servers multiplied by the bandwidth consumed by video streaming. Since we assume here that all videos have the same playback bit-rate, the cost is proportional to the distance between the user and its server. The performance delivered to users (QoE) is assessed by the delay observed in serving requests (neglecting redirection times). We evaluate the number of requests that are postponed due to the saturation of the CDN server, and the average response time to postponed requests. The performance of the considered policies is also assessed by comparing the numbers of requests served by the dServers and by the CDN server.

C. Simulation Results

All simulation results are obtained by running each simulation 100 times, and then averaging the 100 results.

First, we show the ISP operation cost produced by different configurations in figure 6 which represents each cost as a fraction of the cost for the centralized CDN scheme.

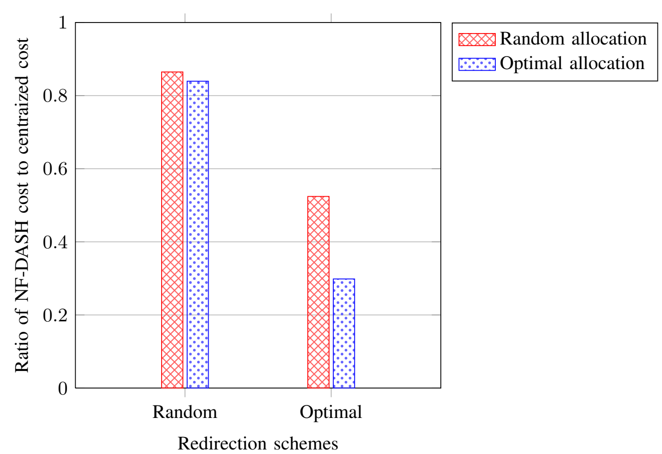

Fig. 6. Comparison of ISP operation costs

Figure 6 shows that all configurations clearly outperform the centralized solution. Even a simplistic NF-DASH implementation where contents are randomly placed in dServers, and requests are randomly directed to dServers (the *Rand-Rand* configuration) allows to reduce the cost by about 15%. The redirection strategy is shown to have the major influence on operation cost: for random redirection, otimizing placement

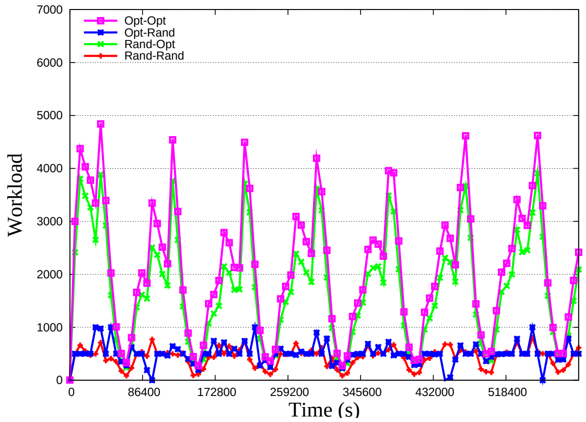

Fig. 7. Workload on dServers

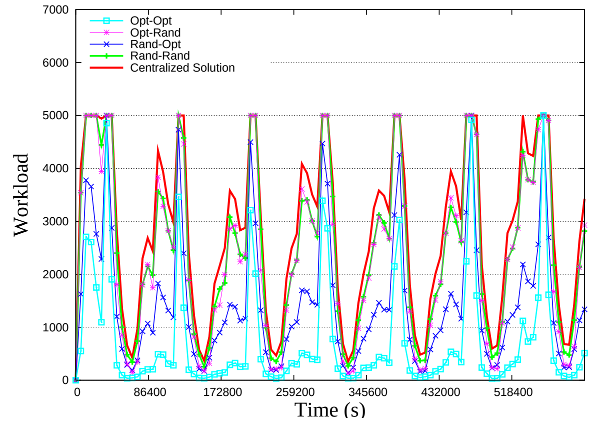

Fig. 8. Workload on CDN server

Measurements / configurations	Access to dCDN	Access to CDN	Average delay
Centralized	0	100%	$4639s$
Rand-Rand	13.5%	86.5%	$1716s$
Rand-Opt	39.2%	60.8%	$1220s$
Opt-Rand	16.1%	83.9%	$1705s$
Opt-Opt	76.9%	23.1%	$886s$

TABLE I
DISTRIBUTION OF ACCESS AND AVERAGE WAITING TIME

(the *Opt-Rand* configuration) presents a limited gain compared to random placement (the *Rand-Rand* configuration). On the other hand, in case of optimal redirection, the operation cost is drastically lowered. In the random placement case (the *Rand-Opt* configuration), the cost is roughly half of the cost of the centralized CDN, whereas optimizing also the placement (the *Opt-Opt* configuration) allows to reduce the operation cost to only 30% of benchmark's operation cost.

In figure 7, 8 and 9, each curve represents one configuration of NF-DASH or the centralized solution.

Figure 7 demonstrates the overall workload undertaken by the dCDN system. Keep in mind that the task of the dCDN is to quickly reply to video requests in order to avoid delay. So a large workload on the dServers translates into a lower workload on the CDN server and less postponed requests (i.e. improved QoE). First of all, we see that the workload's variance shows clearly a daily periodicity. Two daily peaks happen at around 2:00 P.M. and 11:00 P.M, and the workload is minimum at around 4:00 A.M. As the trace starts at Sunday noon, we observe the difference of the workload on week days and weekends. The peak of the workload decreases from Sunday to Thursday, and then increases again on Friday and Saturday nights. These characteristics matches perfectly the user behavior in real world. Thus, the results obtained from our trace are valuables references for future systems designs. Obviously, the *Opt-Opt* configuration treats the largest number of requests. At the "rush hour", the number of requests served by dCDN servers is close to 5,000. This number indicates that almost all service capacity offered by the NF-DASH is occupied. The *Opt-Opt* configuration outperforms the *Rand-Opt* by about 25%, but both perform well. On the contrary, the performances of *Rand-Rand* and *Opt-Rand* configurations are not ideal. The utilization ratio of dCDN bandwidth is only 10% even in the peak period. This proves once again that the redirection scheme is a critical issue concerning the dCDN performance.

Corresponding to figure 7, figure 8 gives the workload dealt with by the CDN server in various NF-DASH configurations and the centralized CDN configuration. The advantage of *Opt-Opt* configuration stands out prominently in the figure. When *Opt-Opt* is implemented, the peak workload at 2:00 P.M. almost disappears at the CDN side. During the night peak on weekdays, only around 70% of the CDN bandwidth is occupied in *Opt-Opt*, while the download link is always saturated in the centralized solution. At weekends, the heaviest workload in all configurations reaches the bandwidth capacity of the CDN server. However, the *Opt-Opt* configuration yields the shortest duration of the saturated state. The *Rand-Opt* configuration also alleviate the CDN workload visibly, and the reduction is about 10%. The other two configurations only slightly diminish the bandwidth utilization of the CDN server.

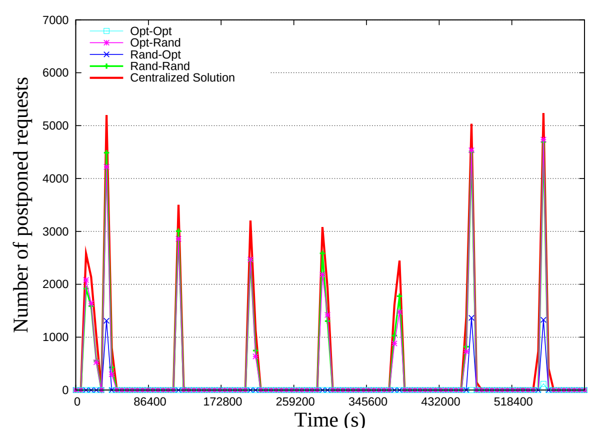

Fig. 9. Delayed requests at CDN server

Figure 9 counts the number of requests postponed by the

CDN server. Same as it is shown in previous figures, the *Opt-Rand* and *Rand-Rand* configurations do not lessen much the number comparing with the centralized solution. The *Rand-Opt* configuration clears up all delayed requests during week days. However, the number at weekends still exceeds $1,000$, which is not acceptable for VoD service. Fortunately, the *Opt-Opt* configuration performs satisfactorily. There is almost no delayed request all the week, except at Saturday night. And the number is less than 100.

In table I, we show the percentage of requests that are respectively satisfied by dServers and by the CDN server, together with the average waiting time of delayed requests. All NF-DASH configurations largely reduce the waiting time. Among all configurations, *Opt-Opt* works extremely well : close to 80% of the requests are served by the dCDN and the waiting time for postponed requests is dramatically shortened.

D. Prototyping NF-DASH

The objective of designing a prototype is to show the feasibility of the NF-DASH architecture and to measure some basic performance metrics like the gain in inter-operator links load and the enhancement in the quality of experience of clients.

The implementation of a dCDN must be as simple as possible and must be compatible with existing technologies. We propose to use the current specification and the current implementation of DASH VLC clients and to push the intelligence of the system into the operator's network. We considered videos structured into equal duration chunks, and coded into several formats. In order to distribute the work between different servers, we partitioned the network into geographic areas which contain their own sets of dServers.

When connecting to a classical CDN server, a client requests an MPD file containing URLs of the video chunks. In our prototype, the original CDN server redirects the client to an *MPD generator server* located in the dTracker of the client's operator network. An MPD generator Servlet creates an MPD file which is client specific. The URLs provided in this file are not directly leading to chunks on dServers, as they are *r*equests to a redirection Servlet located in the client's geographic area ; this allows managing sets of dServers per geographical area. Receiving such a request, the redirection Servlet can in turn redirect the client to the appropriate (e.g. the nearest available server) dServer hosting the requested chunk. In this way, the operator can decide from where the client will retrieve the chunks and thus optimize chunks' location depending on dServers and network conditions.

To test the deployment of our prototype architecture, we have implemented all the components (dTracker and dServers) and deployed them on a VIPEER inter-parnter platform. Each partner hosts a dServer and several streaming clients. A single dTracker is located in TELECOM Bretagne, Brest.

V. Conclusion

The present paper proposes to enhance the current combination of CDNs using DASH to distribute video content. These methods, although allowing more than one billion of Internet users to daily access multimedia contents, do not lend themselves easily to the control of QoS in the ISP network.

We have first shown that all actors in the video distribution value chain (Service Providers, ISPs and CDN operators) could benefit from a closer collaboration between ISPs and CDN operators.

Our proposed architecture, NF-DASH, empowers the ISPs with more control on their resources in the framework of formal agreements between the ISPs and external CDNs: CDNs who are in charge of dissiminating content at a wide scale delegate the distribution at regional levels to ISPs. NF-DASH is fully compatible with the current Internet architecture as it only relies on HTTP redirections and standard HTTP adaptive streaming.

NF-DASH can be considered as a basic realization of information centric networking, which is easily deployable according to the agenda of each actor (Service Provider, ISPs and CDN operators). It is also very flexible, depending on the access network architecture and available storage capacities.

A first prototype implementation of NF-DASH proves the feasibility of our approach, while simulation results highlight that the NF-DASH not only improves the users' QoE but also yields significant cost savings for ISPs.

Future works include content placement optimization, comprehensive bandwidth and storage capacity monitoring and proactive cache control at very large scale. It is also interesting to assess whether NF-DASH can be extended to support all ABS architectures, including those which are proprietary.

References

[1] S. Secci, J.-L. Rougier, A. Pattavina, F. Patrone, and G. Maier, "ClbuMED: Coordinated Multi-Exit Discriminator Strategies for Peering Carriers," in *Next Generation Internet Networks (NGI'09)*, 2009.

[2] T. Leighton, "Improving Performance on the Internet," *Communications of the ACM*, February 2009.

[3] GSMA, "M2M Global Forecast and Analysis," GSMA, Tech. Rep., October 2011.

[4] Cisco, "Visual Networking Index Forecast (2010-2015) White Paper," Cisco, Tech. Rep., June 2011.

[5] K. Cho, H. Jung, M. Lee, D. Ko, T. Kwon, and Y. Choi, "How can an ISP merge with a CDN?" *IEEE Communications Magazine*, vol. 49, no. 10, pp. 156 –162, oct. 2011.

[6] T. Stockhammer, "Dynamic adaptive streaming over HTTP: standards and design principles," in *Proceedings of ACM conference on Multimedia systems (MMSys)*, 2011.

[7] Sandvine, "Global internet phenomena report: Fall 2011," White Paper, October 2010. [Online]. Available: http://www.sandvine.com/news/global_broadband_trends.asp

[8] L. Popa, A. Ghodsi, and I. Stoica, "HTTP as the Narrow Waist of the Future Internet," in *Proceedings of Hotnet-IX*, 2010.

[9] N. Zong, "Survey and gap analysis for HTTP streaming standards and implementations," Internet-Draft (work in progress), draft-zong-httpstreaming-gap-analysis-01, October 2010. [Online]. Available: http://tools.ietf.org/html/draft-zong-httpstreaming-gap-analysis-01

[10] T. Zahariadis, Y. L. Louedec, C. Timmerer, S. Spirou, and D. Griffin, "Catalogue of Advanced Use Cases for Content Delivery Network Interconnection," Internet Draft (work in progress), draft-fmn-cdni-advanced-use-cases-00, Octobre 2011. [Online]. Available: http://tools.ietf.org/html/draft-fmn-cdni-advanced-use-cases-00

[11] C. Bothorel, Z. Li, P. Michiardi, Y. Hadjadj-Aoul, and J. Garnie, "Report on cdn/dcdn design," Tech. Rep., Feb 2012.