Demonstration of Performance Gains Using Information-Aware Content Delivery Mechanisms

Walid Benchaita
UPMC Sorbonne Universités
walid.benchaita@lip6.fr

Gioacchino Tangari
University College London
uceegta@ucl.ac.uk

Samir Ghamri-Doudane
Nokia Bell Labs France
samir.ghamri-doudane@nokia.com

Sébastien Tixeuil
UPMC Sorbonne Universités
sebastien.tixeuil@lip6.fr

***Abstract*—We propose an ICN-inspired request routing framework to improve CDN performance with a minimal impact on currently deployed architectures. In this demonstration, we show that our framework is capable of optimizing delivery delays while enabling the implementation of cost efficient strategies for request routing.**

I. Introduction

Although content delivery solutions are progressively integrated with the network infrastructure, they still exhibit technological limitation when dealing with the increasing amount of bandwidth hungry applications. Such inefficiencies mostly derive from the mismatch between the content oriented service model of CDN solutions and the underlying host oriented networking principles. As an example, mapping a content name to a server location is a required operation that typically generates multiple DNS resolutions, introducing additional latency. Moreover, this mechanism comes short when it comes to realizing a dynamic load balancing of requests over multiple servers, based on varying network conditions. Besides, the connection-based transport model of HTTP/TCP is not adequate for mobile, point-to-multipoint communications. Then, it is essential to develop innovative technical solutions to meet the ever-increasing expectations of performance and functionality from content delivery infrastructures.

Recent solutions have tackled these problems by improving the control on video traffic, *e.g.* through centralized optimization [3], or by addressing the inaccuracy of end-user mapping [2]. However, a common bottleneck for all these solutions lies in the limitations of DNS-based request-routing: *(i)* lack of reactivity (due to DNS architecture and TTLs), *(ii)* additional delays (name resolution and multiple redirections), and *(iii)* low flexibility, since new policies for resolving the client requests need to be reflected to third-party servers. This issue has been recently challenged by the ICN (Information Centric Networking) paradigm [6], however a clean-slate deployment would require severe changes to the network infrastructure [4].

We propose to enrich content delivery primitives with information awareness in order to overcome the mismatch between the content-oriented service model and the underlying delivery infrastructure. For this purpose, we designed information-aware mechanisms that enable clear differentiation with currently deployed CDN solutions, and aim at optimized name-based request routing, connectionless multipath transport, and

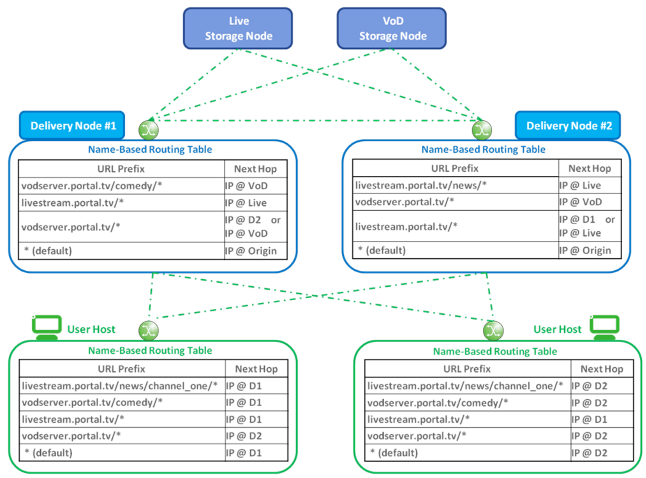

Fig. 1: Name-Based Request Routing

effective cache management at delivery nodes. These mechanisms are easily integrated within the CDN infrastructure with no modifications to the underlying support network, which enables deployment with minimal impact.

In the demonstration, we show that this solution reduces CDN serving costs thanks to more accurate request resolutions (up to the granularity of a single content name) and a more effective connectionless data transport mechanism. Furthermore, it provides improved flexibility and reactivity, as it allows the CDN operator to disregard third-party services (*e.g.* DNS) when enforcing new request routing strategies. For a detailed description of the solution design, please refer to the related published work [1].

II. Name-Based Request Routing

We propose a solution that incorporates information-aware delivery mechanisms into the CDN infrastructure. Differently from the related clean-slate ICN approaches, we eliminate the need to any modifications of the network physical infrastructure or on the Internet protocol suite. Fig 1 presents the main idea of name-based request routing, which is at the cornerstone of our proposed framework. The CDN clients and delivery nodes maintain name-based routing tables integrated in their server software suite. In these individual tables, entries correspond to a specific channel or category name (e.g livestream.portal.tv/news/channel1/*). Similarly to

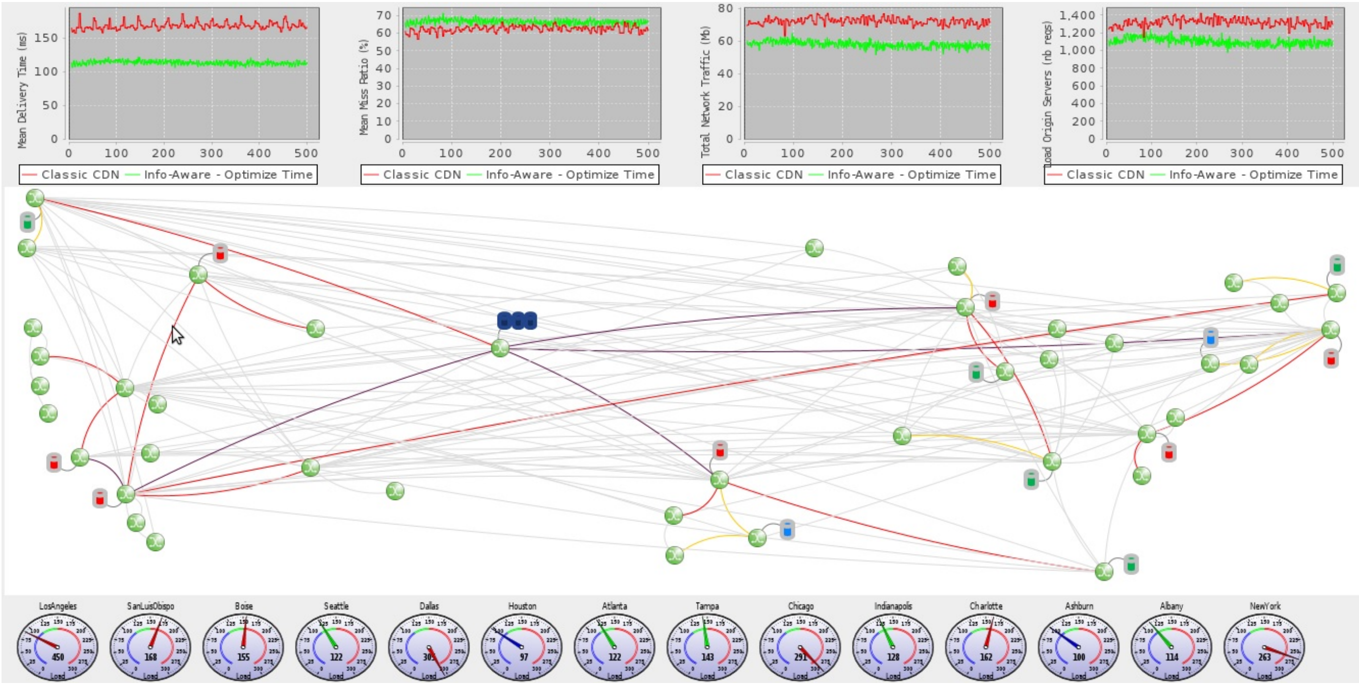

Fig. 2: A screenshot of the demo dashboard with several live statistics on CDN performance

traditional routing, when the table is checked against a requested content name, the matching entry is produced by longest-prefix matching. Such entry maps a single or a pool of possible next hops. Furthermore, the selected next-hop can be a storage node (including the origin server) or a delivery node empowered with caching capabilities. When a copy of the requested content is found, the content is returned to the requester by following a data path that is symmetric to the request one and by using a connectionless transport protocol.

The name-based routing tables are produced by one or more CDN service nodes and distributed to all delivery nodes and client hosts (or proxies) in their scope. New tables are generated in response to changing profiles of user demands, server loads, network traffic, cache setups, or provider policies. For instance, as shown in the demonstration, different routing table configurations can represent specific optimization goals (e.g. minimize delivery costs, minimize delivery delays, balance server charge, etc.) or a tradeoff between different strategies. The ability for frequent updates, with fast implementation, of the routing configuration allows the CDN administrators to deal with the unpredictable nature of request patterns and traffic conditions (e.g. episodes like Flash Crowds).

III. Demonstration Details

In this demonstration, we present how our model can be integrated in present CDN architectures at a minimal cost of few software updates. Then, we show the framework capabilities against an actual operational use case composed of a real video workload from a North American CDN operator. Real CDN topology and cache configurations from the same operator are also used. Besides, to accurately quantify performance gains, we have used the actual request routing policy and settings adopted by the CDN operator.

To implement the proposed architecture and perform an evaluation of the system, a scalable simulation platform, realistically emulating the CDN environment and including all the Information-aware mechanisms introduced earlier, has been developed. One of the main features included in the model, besides name-based request routing, is the connectionless transport, which brings a considerable gain in terms of reduced latencies and is more suitable in a context of dynamic network behaviour and client mobility.

To this end, an Omnet++/INET [5] platform was used as fundamental tools to create a trace-driven, event-based simulation environment, with full support of transport, network and data link layer protocols.

Due to large execution times, the demonstration is mainly based on the replay of various experimentation scenarios. Also, a dashboard will be used as control station by the CDN operator. The dashboard, as depicted in Figure 2, includes a monitor for visualizing live network statistics (*e.g.* delivery delay, hit ratio, cache load, network load) as well as a control panel enabling the enforcement of operator's objectives.

In order to demonstrate the flexibility of our design, we implement and show different strategies that depict potential operator objectives:

(i) optimize delivery delays — in this strategy, client requests are directed to the nearest cache server, and in case of a cache miss, requests are directed to the next-hop cache on the path toward the origin server;

(ii) optimize cost efficiency — each cache server is assigned a subset of the content catalog. The requests are routed according to these assignment in order to maximize content availability (presence at caches);

(iii) trade-off strategy — a hybrid configuration in which the content catalog and the cache servers are splitted into two classes. Then, client requests are directed to the nearest cache server of the same class.

The ultimate goal of the demonstration is to showcase the following benefits and gains: reducing transport and processing costs, improving delivery times, enabling flexibility of the CDN infrastructure, as well as providing fast and dynamic adaptability.

IV. Conclusion

In this demonstration, we present a novel approach to content delivery aiming at reducing the drawbacks of the currently used schemes (DNS-based request routing and static connection bindings). We show that by introducing the information-centric networking principles within the content delivery mechanisms of an operational CDN, it is possible to deploy a faster yet accurate request routing and to achieve considerable improvements in terms of reactivity and flexibility.

References

[1] W. Benchaita, G. Tangari, S. Ghamri-Doudane, and S. Tixeuil, "Enhancing content distribution through information-aware mechanism," in *accepted for publication in the 5th IEEE International Conference on Cloud Networking (CloudNet 2016)*.

[2] F. Chen, R. K. Sitaraman, and M. Torres, "End-user mapping: Next generation request routing for content delivery," in *the 2015 ACM SIGCOMM Conference*. ACM, 2015, pp. 167–181.

[3] M. K. Mukerjee, D. Naylor, J. Jiang, D. Han, S. Seshan, and H. Zhang, "Practical, real-time centralized control for cdn-based live video delivery," in *the 2015 ACM SIGCOMM Conference*. ACM, 2015, pp. 311–324.

[4] D. Perino and M. Varvello, "A reality check for content centric networking," in *the 2011 ACM SIGCOMM Workshop on Information-Centric Networking (ICN 2011)*, 2011, pp. 44–49.

[5] A. Varga and R. Hornig, "An overview of the omnet++ simulation environment," in *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, p. 60.

[6] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. V. Katsaros, and G. C. Polyzos, "A survey of information-centric networking research," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 1024–1049.