Chapter 7
Optimizing Content Delivery in QoS-Aware Multi-CDNs

Nadia Ranaldo
University of Sannio, Italy

Eugenio Zimeo
University of Sannio, Italy

ABSTRACT

Broadband network technologies have improved the bandwidth of the edge of the Internet, but its core is still a bottleneck for large file transfers. Content Delivery Networks (CDNs), built at the edge of the Internet, are able to reduce the workload of network backbones, but their scalability and network reach is often limited, especially in case of QoS-bound delivery services. By using the emerging CDN internetworking, a CDN can dynamically exploit resources of other cooperating CDNs to face peak loads and temporary malfunctions without violating QoS levels negotiated with content providers. In this chapter, after a wide discussion of the problem, the authors propose an architectural schema and an algorithm, based on the divisible load theory, which optimizes delivery of large data files by satisfying an SLA, agreed with a content provider, while respecting the maximum budget that the delivering CDN can pay to peer CDNs to ensure its revenue.

INTRODUCTION

The increasing desire of using rich media services is often not satisfied due to bandwidth limitations, high latency and low quality of service. In fact, even if broadband technologies are significantly improving the bandwidth of the edge of the Internet, backbones remain limited and congested because of the huge traffic generated by service providers for replying to user requests. Therefore, smarter infrastructures are necessary to improve quality, bandwidth availability and profitability by using edge technologies, such as Content Delivery Networks (CDNs) (Hofmann & Beaumont, 2005; Pallis & Vakali, 2006).

DOI: 10.4018/978-1-4666-1794-0.ch007

A CDN is a collection of cooperating servers, distributed across multiple locations, optimized to deliver content, such as static Web pages, applications files, stored multimedia objects and real-time audio/video. Content providers, such as large enterprises, Web service providers, media companies and news broadcasters, exploit CDN services to deliver content and applications with high performance and reliability to a wide and geographically distributed audience.

The basic idea of a CDN is to replicate contents on servers (called replica servers) and deliver them to end-users from "the best" server with the aim of minimizing user-perceived response time and of maximizing overall CDN performance with respect to some metrics, such as bandwidth consumption, server utilization and reliability. The redirection of the user's request to the best replica server is performed using a request routing mechanism, such as the Global Server Load Balancing (GSLB), DNS-based request routing, dynamic metafile generation and HTML rewriting (Pathan & Buyya, 2008a).

From a business point of view, deploying a large-scale CDN for Internet-wide available content is complex and requires high operation costs. It, in fact, involves the availability and management of distributed resources with large storage capacity connected among them by means of high throughput networks. Against sustained costs, CDNs obtain their revenues by service contracts stipulated with content providers, which typically define fees related to the amount of stored and of transferred content. Moreover, Service Level Agreements (SLAs) (Bouman, Trieneken, & van der Zwan, 1999), used to formally define a service contract, agreed between a customer and a service provider, are becoming more and more popular in the context of content delivery services. Commercial CDNs, in fact, are recently starting to adopt SLAs to agree a desired degree of quality of service (QoS) against a penalty to pay to the customer if an SLA is not satisfied; however, at our knowledge, only service availability is taken into consideration by existing providers (Amazon Web Services, 2010).

CDNs could negotiate further QoS parameters for becoming more attractive to content providers, such as maximum response time for dynamic Web transactions or maximum download time for large multimedia data. However, satisfying SLAs for content delivery services introduces an additional complexity in CDN infrastructures management. This leads to further increase the economic investments against low revenue margins.

An approach that can be exploited to foster QoS-aware CDNs is *CDN internetworking*, also called *content peering* (Day, Cain, Tomlinson, & Rzewski, 2003). This emerging approach, based on interoperability and resource sharing mechanisms, allows a CDN to optimize its content delivery services by dynamically exploiting capabilities of other CDNs participating to the same virtual organization, called *multi-CDN*. Since content peering allows a CDN to expand beyond its borders and to deliver large-scale (eventually Internet-wide) services without huge economical investments, it is gaining popularity for improving scalability and network reach of CDN providers and to tear down the initial investment of new ones.

Content peering allows a CDN to avoid the violation of negotiated QoS levels with content providers by dynamically addressing unforeseen situations caused by malfunctions of CDN infrastructures or peak loads typically generated by global and limited-time interest events, such as worldwide sport competitions and historical significance actuality and political occurrences. As an example, in conjunction with the Soccer World Cup 2010, during which Akamai delivered live and on-demand Web content for 24 global broadcasters in 65 countries for one month, a peak of more than 20 million visitors per minute was registered on June 24th, over twice higher than the World Cup 2006 (Young & Faris, 2010).

In the context of a business multi-CDN, CDNs offer their content delivery services with fees to content providers negotiating SLAs with them. If

Related Content

A Review on the Impact of Hybrid Renewable Energy Systems for Sustainability and Energy Management
Swetha Shekarappa G., Manjulata Badi, Saurav Raj, Sheila Mahapatraand M. Senbagavalli (2023). *Handbook of Research on Network-Enabled IoT Applications for Smart City Services (pp. 263-281).*
www.irma-international.org/chapter/a-review-on-the-impact-of-hybrid-renewable-energy-systems-for-sustainability-and-energy-management/331337

Secured Optimal Cost Approach for Bimodal Deep Face Recognition in IoT and Its Applications
Madhavi Gudavalli, Vidaysree P, S Viswanadha Rajuand Surekha Borra (2018). *Big Data Management and the Internet of Things for Improved Health Systems (pp. 163-175).*
www.irma-international.org/chapter/secured-optimal-cost-approach-for-bimodal-deep-face-recognition-in-iot-and-its-applications/196045

Content Delivery Networks: On the Path Towards Secure Cloud-Native Platforms at the Edge
Yannick Le Louédec, Gaëlle Yven, Valéry Bastide, Yiping Chen, Gwenaëlle Delsart, Mateusz Dzida, Frédéric Fieau, Patrick Fleming, Ivan Froger, Lahcene Haddak, Nathalie Labidurie Omnesand Vincent Thiebaut (2021). *Design Innovation and Network Architecture for the Future Internet (pp. 66-95).*
www.irma-international.org/chapter/content-delivery-networks/276695

Design and Analysis of Active Hypertext Views on Databases
Gilles Falquet, Jacques Guyot, Luka Nerimaand Seongbin Park (2003). *Information Modeling for Internet Applications (pp. 40-58).*
www.irma-international.org/chapter/design-analysis-active-hypertext-views/22967

Applications of Big Data and Green IoT-Enabling Technologies for Smart Cities
Onur Doganand Omer Faruk Gurcan (2019). *Handbook of Research on Big Data and the IoT (pp. 22-41).*
www.irma-international.org/chapter/applications-of-big-data-and-green-iot-enabling-technologies-for-smart-cities/224261