Developing Content Delivery Networks

Ioannis Chochliouros
OTE S.A., General Directorate for Technology, Greece

Anastasia S. Spiliopoulou
OTE S.A., General Directorate for Regulatory Affairs, Greece

Stergios P. Chochliouros
Independent Consultant, Greece

INTRODUCTION

Over the past decades, the expansion of the converged Web-based facilities/infrastructures, together with new business perspectives, have created new needs for all (potential) categories of end-users. Although various effects were significant in most sectors (European Commission, 2005) the fast progress has, however, promoted more complex issues, especially for the delivery of multimedia-based applications.

It is now a common view that there is a growing need for delivering high-quality services in the scope of liberalized and competitive markets, where multiple factors of different origin (i.e., technological, business, economic, regulatory, social, etc.) can drastically affect further deployment, establishment or upgrading of existing infrastructures and of any possible (innovative) services offered through them, especially if considering the continuous expansion of the broadband perspective (Chochliouros, & Spiliopoulou, 2005). Furthermore, multimedia applications are bandwidth consuming and new applications for absorbing the available assets appear. As the "converged" sector of information technologies, communication, and media industries is currently on the "edge" of a crucial phase of growth, several challenges appear in the global scene: Appropriate infrastructures for delivering mails, exchanging data files (of various forms of content) and simple Web browsing are now required to be adopted and used, to support the streaming of multimedia content and, *simultaneously*, to "compose" a reliable means of transmitting information between several entities (physical and legal persons) using digital facilities.

Although technological advances have enhanced the deployment of faster (lesser latency) and greater (more bandwidth) "network lines" possessing significant advantages, the demands of the extravagant use of Internet from users worldwide (Dilley, Maggs, Parikh, Prokop, Sitaraman, & Weihl, 2002; Shoniregun, Chochliouros, Laperche, Logvynovskiy, & Spiliopoulou-Chochliourou, 2004), together with an extensive variance of services offered, were primary motives for researchers to develop a specific category of modern infrastructures, known as content distribution (or delivery) networks (CDNs) (Hull, 2002; Verma, 2002).

The development of suitable content delivery networking comprises one of the most important challenges in the global networking area, together with the expansion of various IP trends. Content networks influence high-layer network intelligence to efficiently manage the delivery of various forms of data (which is becoming progressively more multimedia in nature). At an initial stage, they were built upon the structure of the public Internet (Saroiu, Gummadi, Dunn, Gribble, & Levy, 2002), to accelerate Web site performance (Johnson, Carr, Day, & Frans Kaashoek, 2000). This option has been fulfilled in numerous cases, and such intelligent network tools can be applied in other beneficial and profitable ways.

According to the present market experience, several definitions may appear to depict both the specific nature and the usage of a CDN. Although some people think of it as the means for "delivery" of streaming video or television over the Internet (or over private networks), others consider it as Web switching or content-switching. An alternative approach suggests that it may be considered as a "way" to improve Web site performance. All possible approaches are real, to a certain extent, according to the specific application. In fact, a CDN is a network optimized to deliver specific

content, such as static Web pages, transaction-based Web sites, streaming media, or even real-time video or audio, especially to enable the distribution and the delivery of rich media over wide area networks, such as the Internet or corporate WANs (Tiscali, 2005).

BACKGROUND

Approaches for the Building of CDNs

A CDN is a network of servers that cache or store Web content (i.e., Web pages and embedded objects) and intelligently deliver it to users based on their geographic location. CDN servers are typically collocated with Internet service providers (ISPs) with which the CDN has alliances. When users request content, the request is redirected to the nearest CDN server, where *nearness* is based on expected latency, which is in turn determined by geographical proximity, server load, and network conditions. By delivering content from the edge of the Internet, CDNs speed content delivery, circumvent bottlenecks, and provide protection from sudden traffic surges that can bring down servers, rendering Web sites unreachable (Nottingham, 2000).

Thus, CDNs are an important element of the digital supply-chain for the delivery of information goods. The supply-chain consists of content providers (CPs) that create the content, backbone, and access networks that help transport the content, and CDNs that store and deliver the content to the end-users. Thus, CDNs function as content storage and distribution centres, performing similar functions to those accomplished by distributors/retailer warehouses in traditional supply-chains.

There are two general approaches to build CDNs. The first one is the so-called "overlay model," which replicates content to thousands of servers worldwide. Application specific servers or caches, at various points in the entire network infrastructure, handle the distribution of specific content types (such as Web graphics or streaming video). The core network infrastructure, including two routers and switches, plays no part in content delivery, short of providing basic connectivity or perhaps guaranteed quality of service (QoS) for specific types of traffic. (Good examples are those CDNs deployed by certain world-wide known companies such as Akamai (www.akamai.com), Digital Island (www.digitalisland.net), Speedera (www.speedera.com) and many more].

In various existing cases, most of which are found in business scenarios, a CDN can be quite well considered as an overlay network that is built to deliver content to a distributed audience, which is constructed as a layer on top of the network infrastructure.

The second approach is the so-called "network model." This sets up code to routers and switches so that they can recognize and distinguish specific application types and make forward decisions on the basis of exact predefined policies. Examples of this approach include devices that redirect content requests to local caches or switch traffic coming into data centres to specific servers optimized to serve certain content types. Some content delivery designs use both the network and overlay approaches. IP Multicast is a good example of an early network-based approach to optimizing the delivery of specific content types (Tiscali, 2005).

OPTION FOR THE DEVELOPMENT OF CDNS

Basic CDN Features

A content delivery network mainly intends to fulfill three critical needs (Vakali & Pallis, 2003), and is able to have a strong effect on network or service performances (Dilley et al., 2002). These are listed as follows:

- Decrease of the network traffic, minimization of the bandwidth consumed, and prevention of network congestion effects.
- Minimization of the external latency encountered by the end-users involved in network usage.
- Provision of greater reliability for more effective (and reasonable) availability of the desired content.

To conform to the above targeted requirements, a specific technique is both suggested and implemented in multiple practical cases; in fact, this technique also constitutes a basic "feature" of the corresponding CDN (Rabinovitch & Spastscheck, 2002). This mainly implicates the deployment of surrogate-caching servers, which are able to host the desired content. In such a case, when a client submits a request for specific con-