Chapter 2
On the Performance of Content Delivery Clouds

Mukaddim Pathan
Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia

James Broberg
University of Melbourne, Australia

Rajkumar Buyya
University of Melbourne, Australia

ABSTRACT

Extending the traditional Content Delivery Network (CDN) model to use Cloud Computing is highly appealing. It allows developing a truly on-demand CDN architecture based upon standards designed to ease interoperability, scalability, performance, and flexibility. To better understand the system model, necessity, and perceived advantages of Cloud-based CDNs, this chapter provides an extensive coverage and comparative analysis of the state of the art. It also provides a case study on the MetaCDN Content Delivery Cloud, along with highlights of empirical performance observations from its world-wide distributed platform.

INTRODUCTION

Content Delivery Networks (CDNs) (Buyya, et al., 2008; Pallis & Vakali, 2006) are designed to improve Web access performance, in terms of *response time* and *system throughput*, while delivering content to Internet end-users through multiple, geographically distributed replica servers. The CDN industry, i.e. content delivery, consumption and monetization, has been undergoing rapid changes. The multi-dimensional surge in content delivery from end-users has lead to an explosion of new content, formats as well as an exponential increase in the size and complexity of the digital content supply chain. These changes have been accelerated by economic downturn in that the content providers are under increasing pressure to reduce costs while increasing revenue.

With the traditional model of content delivery, a content provider is locked-in for a particular period of time under specific Service Level Agreements (SLAs) with a high monthly/yearly fees and excess

DOI: 10.4018/978-1-4666-1794-0.ch002

data charges (Hosanagar, et al., 2008). Thus, far from democratizing content delivery, most CDN services are often priced out of reach for all but large enterprise customers (Rayburn, 2009). On the other hand, a commercial CDN provider realizes high operational cost and even monetary penalization if it fails to meet the SLA-bound commitments to provide high quality of service to end-users. Thus, it suffers from—spiraling ownership costs; resource wastage for maintaining infrastructure; inability to grow or to profit from economics of scale; inability to fully monetize new or long tail content—to leave lucrative business deals on the table and forfeit profits.

Furthermore, the main value proposition for CDN services has shifted over time. Initially, the focus was on improving end-user perceived experience by decreasing response time, especially when the customer Web site experiences unexpected traffic surges. Nowadays, CDN services are treated by content providers as a way to use a shared infrastructure to handle their peak capacity requirements, thus allowing reduced investment cost in their own Web site infrastructure. Moreover, recent trends in CDNs indicate a large paradigm shift towards a utility computing model (Canali, et al., 2004), which allows customers to exploit advanced content delivery services without having to build a dedicated infrastructure (Gayek, et al., 2004; Subramanya & Yi, 2005). To break through these barriers, a more efficient content delivery solution is required—a truly on-demand architecture based upon standards designed to ease interoperability, scalability, performance, and flexibility.

One approach to address these issues is to exploit the recent emergence of "Cloud Computing" (Buyya, et al., 2009), a recent technology trend that moves computing and data away from desktop and portable PCs into computational resources such as large Data Centers ("Computing") and make them accessible as scalable, on-demand services over a network (the "Cloud"). The main technical underpinnings of Cloud Computing infrastructures and services include virtualization, service-orientation, elasticity, multi-tenancy, power efficiency, and economics of scale. The perceived advantages for Cloud-service clients include the ability to add more capacity at peak demand, reduce cost, experiment with new services, and to remove unneeded capacity.

Extending the traditional CDN model to use clouds for content delivery, i.e. a Content Delivery Cloud (Cohen, 2008), is highly appealing as cloud providers, e.g. Amazon Simple Storage Service (S3), Mosso Cloud Files, and Nirvanix Storage Delivery Network (SDN), charge customers for their utilization of storage and transfer of content (*pay-as-you-go*), typically in order of cents per gigabyte. Cloud providers, on the face value, offer SLA-backed performance and uptime guarantees for their services. Moreover, they can rapidly and cheaply scale-out during flash crowds (Arlitt & Jin, 2000) and anticipated increases in demand. By exploiting the power of Cloud computing, CDN providers endeavor to improve cost efficiency, accelerate innovations, attain faster time-to-market, and achieve application scalability (Leighton, 2009). There are a number of major players in this domain that are providing cloud-based content delivery services on a commercial basis, either by themselves or by partnering with an existing CDN, such as Amazon CloudFront, VoxCAST CDN, and Akamai Cloud Optimizer.

An example research initiative in this context is MetaCDN (Broberg, et al., 2009; Pathan, et al., 2009), an integrated overlay network that leverages resources from existing storage clouds to provide content delivery services. The main goals of the MetaCDN system is to provide economics of scale and high content delivery performance through its simple yet general purpose, reusable, and reliable geographically distributed framework. MetaCDN delivers high performance content delivery via an on-demand cloud service, eliminating costly capital expenditures or infrastructure upgrades. MetaCDN can be deployed as a fully outsourced, end-to-end services platform or as a complement