White Paper

Virtual CDNs: Maximizing Performance While Minimizing Cost

Prepared by

Roz Roseboro
Senior Analyst, Heavy Reading
www.heavyreading.com

on behalf of

www.6wind.com

www.hp.com

November 2014

Introduction

Content delivery network (CDN) operators are coping with ever-increasing demands for online content, while at the same time being challenged to manage costs. Their customers want to access content on a plethora of devices, and want that content to be available to them anywhere, at any time. This results in a need for content to be distributed closer to the edge.

As a result, CDN operators need to consider alternative architectural approaches and business models. Whereas today they often manage thousands of servers and procure transport from telecom operators, they could soon be able to rent the needed resources from telecom operators offering hosted virtualized services, which would dramatically reduce their operational expenditure (opex).

Companies such as 6WIND and HP have developed technologies that will provide platforms to enable this hosted model. Their combined solution leverages 6WIND's packet-processing acceleration and HP's telecom-grade servers to cost-effectively achieve the scale required for virtual CDNs.

This white paper is structured as follows:

- **Section II** discusses the trends in content delivery that are driving the need for greater capacity in CDNs. It presents current CDN architecture and discusses existing CDN business models.
- **Section III** discusses the impact of network function virtualization (NFV) and how it can enable virtual CDN. It highlights how emerging technologies can provide platforms with the requisite performance for content delivery.
- **Section IV** shows how CDN operators, telecom operators and application developers can benefit from new architectures and business models enabled through innovations from 6WIND and HP.

To Meet Demand, CDN Operators Must Expand Capacity While Managing Costs

Greater Numbers & Variety of Devices Consuming Content

Consumers have no end of choices for consuming content. Smartphone penetration continues to grow, and tablets have ramped up more quickly than nearly any other consumer electronics device in history. Video consumption per user has grown accordingly, and shows no signs of abating.

Consumers are also using devices such as Roku, Google Chromecast and Amazon Fire to watch Internet content on their TV screens. Much of the viewing that had historically been done via broadcast and cable is now being done online.

More Content Moving Online

Content owners are making ever more movie and TV series content available for streaming on services such as Netflix and Amazon Prime. Meanwhile, services such as WatchESPN and HBO Go allow consumers to view live cable programs over the

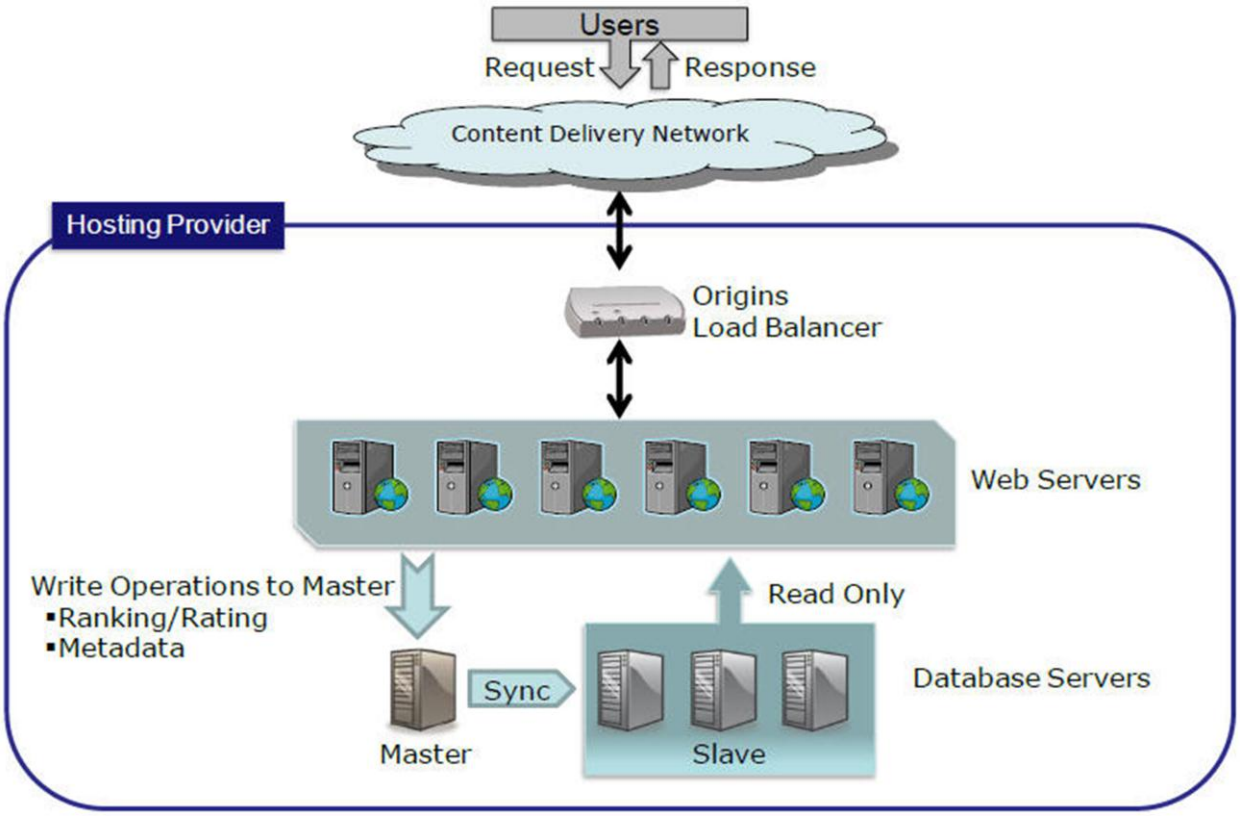
Internet. Especially among younger consumers, YouTube is seen as a viable place to find and watch content that is increasingly more professionally created and curated. Noting the movement to online content, Netflix and Amazon have begun sponsoring the creation of new content.

Greater Mobility Leads to the Need for More Distributed Delivery

As mentioned above, consumers are accessing online content from a wide array of devices, some of which are highly mobile. This drives CDN operators to provide more and more distributed delivery. They need to build more data centers, holding more servers, load balancers (and other) equipment, in order to support constantly increasing amounts of content. It also means that CDN operators need to procure ever-larger quantities of transport capacity from telecom operators.

Figure 1 shows a conceptual view of a CDN. The master server contains the original source content that is then fed to the Web servers. Rather than a single server, this is more likely to be a cluster of servers. Load balancers are used in order to maximize utilization of server resources.

Figure 1: Conceptual Architecture Overview of a CDN



Source: semanticommunity.info

Currently, CDN operators usually run in one of two business models. In one, a CDN operator runs its data centers separately from those run by telecom operators, and

connects to telecom networks via peering arrangements. In this scenario, the CDN operator has two sets of costs to manage: one for the data center, and one for operator peering (comprising connection capacity, fiber rental and other fees).

In the second business model, a CDN operator rents space in a telecom operator's data center. In this scenario, the CDN operator only needs to manage one set of costs from the telecom operator – data center rentals, connection capacity and other feeds. In each case, however, the CDN provider has significant opex to operator physical servers.

More Video

While CDNs initially delivered primarily static content, today they deliver much more dynamic content, including video. Video is more sensitive to latency, and because the impact on the end user is so visible, CDN operators are challenged to maximize and optimize network performance. Because they face numerous competitive challenges, they must do so while also minimizing their costs.

Video also requires significantly more bandwidth than static Web pages. This means that the amount of networking capacity they need from telecom operators is also increasing rapidly.

NFV Will Change How Services Are Delivered

Moving Away From Proprietary Hardware to Standardized Servers

Many telecom operators are currently in the process of trialing network functions virtualization (NFV). They are looking to move away from dedicated, proprietary hardware devices to leverage standardized servers. Besides lowering capex, they expect to reduce opex by automating processes. Of course, they are also looking at new ways to create and deliver services once functions have been virtualized.

However, telecom operators want at least the same level of performance as on specialized platforms. They need to provide the same service quality to their customers as before. They also want to be able to support multiple virtualized network functions (VNFs) on common infrastructure. That will enable them to support a hosted service delivery model, as well as maximize hardware utilization. For CDN operators, this means they will be able move away from owning their own servers, and instead rent computing and storage resources from a telecom operator. In this model, the opex cost structure for the CDN operator changes dramatically.

Enabling Technologies Provide Requisite Performance

Companies looking to capitalize on the move toward NFV are focusing on utilizing IT industry technologies that will provide the platform performance and system management capabilities that telcos demand. 6WIND and HP are two such companies, and have partnered to provide a high-volume, low-cost platform that combines 6WIND's TCP stack and Virtual Accelerator and HP's industry-standard servers.

6WIND's accelerated TCP networking stack processes packets outside of the Linux kernel, thereby reducing Linux bottlenecks and increasing performance. Its Virtual Accelerator, which combines accelerated virtual switching with virtual networking features, provides the basis for the NFV infrastructure, on top of which run the CDN

VNFs (which themselves may leverage 6WIND's TCP stack). The Virtual Accelerator allows the operator to either increase VNF density or increase performance for a smaller number of virtual machines. In addition, 6WIND provides the lower level software tools to tune and optimize software on HP servers.

HP is working closely with industry-leading packet-processing software providers, such as 6WIND, to ensure that its high-volume, industry-standard servers are optimally configured and tested for the highest possible packet-processing capabilities. HP offers a complete portfolio of blade, rackmount and workload-optimized servers that have been designed to support telecom applications. Its servers support configurations that are NEBS- and ETSI-compliant, and support a full range of operating systems and hypervisors, including carrier-grade Linux.

This integrated solution allows VNFs to run on standard HP servers, rather than dedicated, proprietary devices. Processes now run as virtual machines. In the context of CDNs, this means all the functions that currently run on specialized servers – such as processing user requests, caching, customizing and serving content, and disseminating status and control information – can now be run on a standard HP server. Indeed, multiple functions can even run on a single server, due to the session and throughput scalability of the 6WIND and HP solution. Higher density means more customers for the operator and a better balanced price point for resource renting for customers.

6WIND recently tested its TCP Termination module, which accelerates networking applications by terminating TCP connections in the 6WINDGate Fast Path, rather than the Linux kernel. It found that on an HP DL580 server, it was able to respond to 6.5 million HTTP requests of 4KB pages per second, and sustain creation of 5 million new sessions per second, to reach 107 million open connections. Because it enables parallel operations for both session creation and data path, linear performance scalability is achieved. Previously, the multi-core architecture was masked to programs, but 6WIND has made it visible.

Players Across the CDN Ecosystem Can Benefit

CDN Operators Can Maximize Footprint & Scale Cost-Effectively

For CDN operators, this solution greatly reduces the number of servers and other devices, such as load balancers, required to support a given number of users for video, or a given amount of traffic for static content. **Figures 2** and **3** show the configuration differences between a best-of-breed web server and that of the 6WIND/HP solution for a 160Gbit/s web server with 600GB of in-memory 40KB images.

Figure 2: Best-of-Breed Web Server Hardware

ELEMENT	QTY	TOTAL U'S	TOTAL WATTS
DL380 784GBRAM 2*10Gbit/s	8	16	4,000
Load Balancer 2*40Gbit/s + 8*10Gbit/s	1	16	2,600
Total		**32**	**8,600**

Source: 6WIND

Figure 3: Hardware With 6WINDGate TCP Stack

ELEMENT	QTY	TOTAL U'S	TOTAL WATTS
DL380 784GB 4*40Gbit/s	1	2	500
Total		**2**	**500**

Source: 6WIND

For video delivery, the number of subscribers (i.e., the number of concurrent connections) is more important than the actual throughput. Rather than 100 million connections on a single server, what is needed is linear behavior from 1 to 10 million concurrent connections. This is what the 6WIND/HP solution can provide. Also, for small file servers, 6WIND/HP can support 70,000 file requests per second – which is significantly more efficient than existing Web servers.

Because 100 million TCP connections on a single server is more than is needed to support the number of subscribers accessing the network, there is plenty of capacity to manage DDoS attacks when they occur. With 240 Gbit/s of throughput, the CDN will no longer be the bottleneck in service delivery. Also, the need for load balancing is dramatically reduced at this scale. Opex should also be reduced, as the constant firmware updates needed as applications change will now be avoided.

In addition to fewer servers being needed, lower-cost IT industry servers can now be deployed, resulting in substantial hardware and system management cost savings. This will make it more cost-effective for CDN operators to distribute content even further out into the network, and to scale the network to support increasing levels of traffic. This will also make it more cost-effective to provide high availability, as the cost of providing backup servers for redundancy is dramatically lowered.

Hosted Model to Serve CDN Marketplace

Telcos can also benefit from this new platform. In terms of NFV, they can optimize their hardware spending by running multiple VNFs on a single server. They can also directly leverage 6WIND's Virtual Accelerator to accelerate the CDN virtual machines on a server. Because the platform scales to 240 Gbit/s and more than 100 million TCP connections, telcos can support multiple CDN operators from a common cluster of servers by providing CDN functionality in a hosted model. Having the hosted component could provide an edge when competing for connectivity services and help reduce churn, since content is a stickier service than connectivity alone. The hosted architecture will look like a set of 40Gbit/s optimized VNFs on top of a Virtual Accelerator-enabled NFV infrastructure, probably leveraging a form of software load balancing that has also been accelerated by 6WIND TCP.

Platform for Application Developers

CDN technology providers such as network equipment providers and independent software vendors can leverage 6WIND's TCP stack combined with its Virtual Accelerator to enhance their products, so that CDN operators can replace their servers with HP servers + 6WIND software with no degradation in performance. Developers can take advantage of new multi-core CPU architectures by allocating additional processor cores to their applications.

Conclusion

Given that demand for online content will continue to grow at a rapid pace, CDN operators should strongly consider new architectural approaches and business models in order to support demand in a cost-effective manner. Telecom operators will soon be virtualizing many network functions – including those supporting content delivery – so CDN operators will be able to leverage a hosted model and dramatically reduce their opex.

Telecom operators can take advantage of emerging technologies from companies such as 6WIND and HP that will enable them to provide the performance they need to support CDN customers. Application developers will also benefit from having a platform they can design upon with confidence.