



the
diffusion
group

Insights in Front of the Curve™

The Cloud Effect:

Can Cloud Resources Be Used to Deliver
Quality, Scalable and Affordable OTT Video?

Presented by



Colin Dixon,
Senior Partner, Advisory

Table of Contents

1.0 Introduction	3
2.0 Streaming Video is Hard to Do Well	4
3.0 CDNs Fall Short of the Challenge	5
4.0 Leveraging the Cloud for a Scalable Approach	7
5.0 How Cloudmass Works	8
6.0 Can the Cloudmass Approach Contain Streaming Costs	10
7.0 Summarizing the Cloudmass Approach	11
8.0 Conclusion	12
TDG Transparency Statement	12

List of Figures

Fig. 1: Video Bandwidth Required to the Home through 2014	4
Fig. 2: The Mesh Network of Streaming Resources	9

1.0 Introduction

Consumers love web video. By every measure, consumers are looking to the Internet for their entertainment needs more and more. The average Internet video user watched 15.1 hours in October 2010,¹ a 40% increase over the previous year.² Increasingly, consumers are turning to the web first to find an episode of a favorite television show they missed, to get up-to-the-minute news updates, and to watch sports broadcasts that aren't carried on their local cable networks.

As television joins the ranks of connected devices, consumers' expectations — and demand — for a quality experience are rising. Indeed, the grainy postage stamp-sized videos of the past have given way to full HD sports and movies, yet delivering a stellar video experience from the web is far from a simple task. For example, MLB.TV's promise to bring live HD quality games and multi-screen viewing to the baseball-loving masses has disappointed many subscribers. One frustrated web viewer put it this way:

"MLB.TV has made beautiful promises, has taken my money, and has yet to deliver. Wouldn't you feel robbed, too?"³

For content providers, failing to offer viewers the high-quality experience they expect means watching customers turn to other sites or mediums for content. And lost viewers means lost revenue.

An additional problem for content producers and aggregators is the cost of delivery. For the largest providers of video on the Internet, delivery costs have declined dramatically — in some cases as low as a penny per gigabyte.⁴ But costs have remained stubbornly high for smaller providers — in the 40 to 50 cent per gigabyte range for the last two years.⁵ This means the cost of delivering a one-hour HD

show can be as much as \$1. When the going rate for TV shows at Amazon Video On Demand is \$1.99,⁶ it's difficult to see how the small provider can make any money at all!

This paper will look at the environment for streaming video on the web today. We will examine how video is currently scaled and streamed from the web by existing Content Delivery Networks (CDN), as well as how these approaches impact video delivery costs and quality. In addition, we will examine how Octoshape's Cloudmass solution attempts to address the short-comings of the CDN Content Delivery Network (CDN) approach by creating a mesh of computing resources to help contain costs and improve streaming quality.

¹ comScore Press Release, Hulu Delivers 1 Billion Video Ads as Fall Season Viewership Increases (comScore Mediametrix, November 15, 2010)

² comScore Press Release, Hulu Delivers Record 856 Million U.S. Video Views in October During Height of Fall Season (comScore Mediametrix, November 25, 2009)

³ Matt O'Donnell, Problems Abound for MLB.TV, <http://www.fenwaywest.com/2010-archives/april/problems-abound-for-mlbtv.html>, (accessed on December 17 2010)

⁴ Tony Greenberg, CDN Pricing a Penny A Gig Ramp Rate, (February 18, 2010)

⁵ Various postings by Dan Rayburn at www.cdnpricing.com

⁶ Amazon Video On Demand, <http://www.amazon.com/b/?&node=16261631>, (accessed on January 18, 2011)

2.0 Streaming Video Is Hard to Do Well

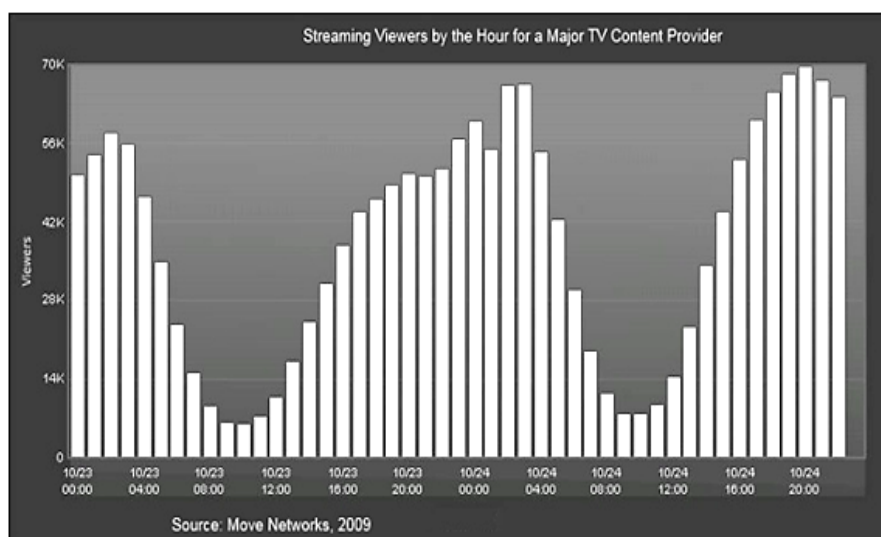
For CDNs – companies in the business of delivering streaming video to consumers – delivering a high-quality video experience is no easy task. Consider the unpredictable demand for a single TV show in a single day (see Figure 1.) Viewers varied from 5,000 to 70,000. This unpredictability means the CDN needs to have streaming capacity to cover the peak of 70,000 viewers, even though that full capacity will only be used for a short peak period during each day.

On a daily, weekly and yearly cycle it is possible to anticipate the size and approximate times of usage peaks for mainstream on-demand streaming capacity. However, huge amounts of this capacity will be idle for much of the time. This reality dramatically increases the cost of delivering on-demand shows.

Unfortunately, it's almost impossible to predict the size of demand for live events. An unexpected occurrence can cause a huge spike in demand for streamed video. For example, the death of Michael Jackson caused such an overwhelming spike in

demand that even Google couldn't keep up with all the search queries.⁷ Although CDNs have huge amounts of capacity allocated to cope with on-demand video delivery peaks, this alone is simply not enough to accommodate some live events. To understand why we must first look at how the CDNs have been built to deliver streamed video.

Figure 1:
Streaming Viewers by the Hour for a Major TV Content Provider



⁷ Vassiliss Manoussos, Michael Jackson's Death and the World Wide Web: The First 48 Hours (WebUpOn, June 30, 2009)

3.0 CDNs Fall Short of the Challenge

In order to understand the limitations of the approach employed by companies such as Akamai and Limelight Networks, we must first explore the fundamentals that drive costs in the CDN business model.

Established streaming technologies, and even newer technologies such as adaptive bitrate streaming, are bound by one core principle: the further the consumer is from the streaming server, the poorer the video quality the consumer experiences. CDNs actually use this distance/quality constraint to create differentiators and barriers to entry in the market.

For example, Akamai addresses quality by creating a huge distributed network of servers that push resources as close to the edge of the network as possible. When a client requests a video, the content is streamed from the closest available streaming server. This minimizes the number of network segments the video must traverse before it reaches the client and, it is hoped, minimizes the chances of delays in video delivery.

Limelight addresses the issue with a more centralized set of server farms streaming directly to the clients. In effect, Limelight has created a separate network parallel to the Internet. The Limelight network interconnects directly with ISP networks. So the majority of video Limelight delivers goes directly from the CDN to the ISP without traversing the Internet at all. As Michael Gordon, Chief Strategy Officer of Limelight, says:

“The Internet cloud is a part of the story. But in our case... an increasingly shrinking part of the story as our network grows.”⁸

Limelight’s approach, then, is to remove the Internet from the delivery network and, therefore, attempt to reduce the number of routers and amount of latency between the consumer and the server.

In both approaches, huge networks of video servers have been dedicated to handling the highly variable load of Internet video. This can yield strong streaming results, but at significant cost. Consider the cost factors associated with the approaches of Akamai and Limelight:

- Both CDNs must maintain huge overcapacity to be able to cope with expected and unexpected spikes in demand.
- Both approaches do nothing to optimize “last mile” delivery, the connection from the ISP to the consumer’s house. Unfortunately, this is where many of the delays occur.

⁸<http://uk.limelightnetworks.com/network.htm>, (accessed on November the 12, 2010)

- Both CDNs have a finite amount of streaming capacity in any given area. If there is a spike in demand in an area that exceeds the capacity to deliver, video quality will plummet.
- Both use traditional streaming techniques that create a connection between client and server. If the link between the two degrades, video quality will suffer.

The existing CDN approaches have massive infrastructure and recurring operational costs that are directly proportional to the quality of the video delivered. This massive investment cost is the “ante” required to play in the big CDN game — and the reason why there are very few CDNs at the top. It also means there is definitive floor in the cost to achieve an acceptable quality level. Once that cost floor is reached, the only differentiator left is a better quality experience. For today’s CDNs, that translates into which company has the most hosts closest to the target users.

4.0 Leveraging the Cloud for a Scalable Approach

There is a new approach to handling variable demand called cloud computing. Companies such as Google, Amazon, and Microsoft are building computing centers attached to the Internet that are available on-demand for a variety of computing purposes.

Customers can easily leverage this capacity “as needed” without having to pay for resources they are not using. Could a streaming video solution leverage cloud resources to handle the peaks in demand without having to maintain this peak capacity itself?

Unfortunately, cloud computing resources are not particularly well suited to handling the rigorous demands of delivering streamed video. First and foremost, cloud resources are centralized and do not fit the distance/quality-driven architecture of the CDNs. Cloud resources are also shared and are only available on a best-effort basis. This works against the demands of traditional streaming technologies with regard to quality. As well, streaming from the cloud can be more expensive than the CDN approach because the resources are not optimized for streaming and need to be initialized each time they are used.

In order to leverage cloud resources in a video streaming network, a video-sensitive approach is needed; one that anticipates imminent need and ensures that resources are ready to handle that need. This approach must overlay cloud (and other resources) to create a cooperative mesh that meets streaming video demand. Octoshape seeks to implement such an approach with their Cloudmass solution. So, let us examine how Cloudmass works.

5.0 How Cloudmass Works

As we have seen, traditional CDNs are currently constrained by the purported relationship between distance and quality. The Octoshape transport was not built bound by this constraint. Rather, Cloudmass can employ any centralized cloud resource to deliver consistent high quality streaming video either ondemand or for live events.

The Octoshape Cloudmass solution unifies four types of streaming resources into a single network:

1. **Origin Servers:** These servers store all videos available in the network.
2. **Streaming Servers:** These are designated resources to stream videos to clients.
3. **Cloud Streamers:** These streaming resources can be called on when needed.
4. **Multicast Resources:** Routers, relays, and servers that enable multicast video streaming.

Each of these resources is in constant communication with the other resources, forming a mesh network that can flexibly deploy resources as needed.

A video in this network begins its lifecycle on the Origin server. The Origin server can automatically distribute video to the streaming servers, or the video can be delivered to them on-demand. As requests for the video come in from clients, the streaming servers deliver the videos to them.

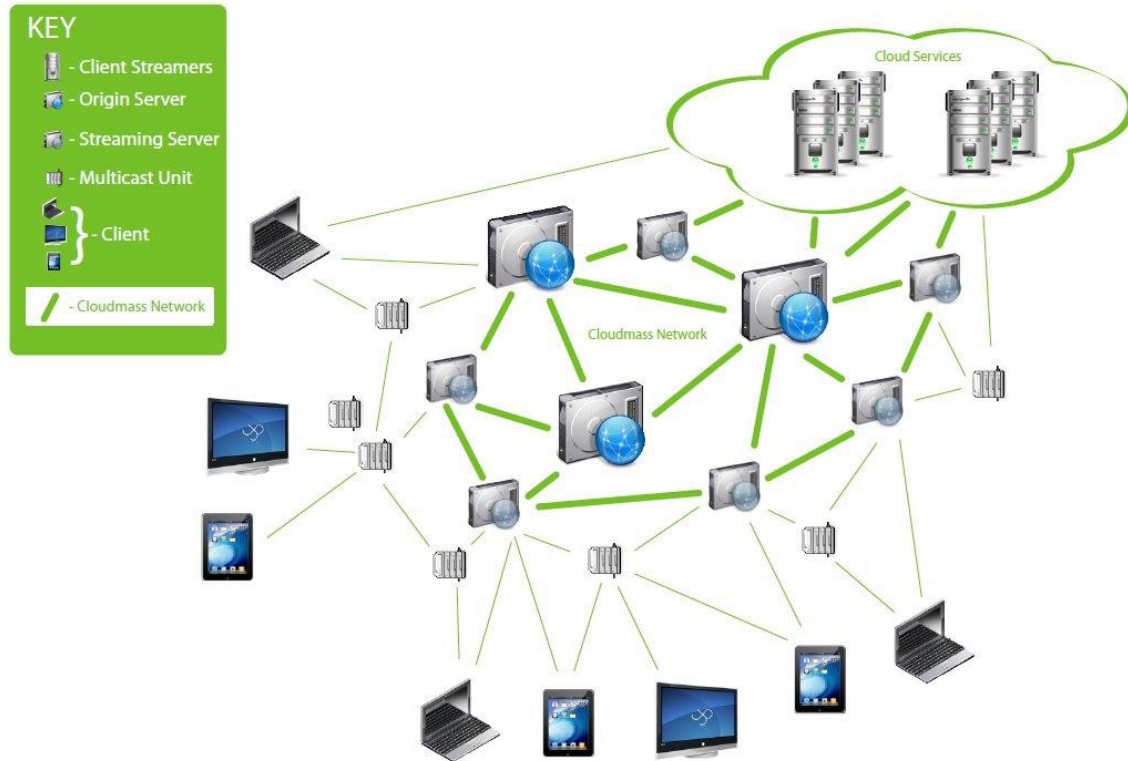
Multicast Streaming –

Normally, video is streamed from one server directly to one client, so called unicast streaming. In multicast streaming, one server streams to many clients at the same time.

Multicast streaming uses much less bandwidth than unicast as many clients share a single video stream rather than each client having its own stream.

Figure 2:

The Mesh Network of Streaming Resources



By contrast, in traditional video streaming, a client has a direct relationship with the server delivering the video. A client requests a video from a specific server and the server and client create a connection between themselves. The server sends some of the video to the client, the client tells the server the video has been received, and the client requests more.

Unfortunately, if the server is overloaded or the link between them becomes congested, the client may not get the next part of the video in time and the viewer may see the picture freeze. With newer streaming technologies, such as adaptive bit rate, at minimum the video quality will be reduced significantly.

With the Cloudmass approach, there isn't a specific connection created between an individual server and client. Instead, the client sends the request for a video to the

network and any resource in the Cloudmass mesh that has the video can start sending small components of it. This means the requesting client will receive video from the fastest responding resources first.

Also, the client may receive parts of the video from multiple resources that are part of the Cloudmass mesh at the same time. It reassembles the video from all the replies it receives, and plays the video back. Since there is no connection between the client and a particular server, the client isn't affected if one of the servers is overloaded or if a link becomes congested. And since distance from the server is no longer tied to quality, the client always has a cloud of resources across multiple diverse paths and multiple global cloud infrastructures from which to pull the data.

6.0 Can the Cloudmass Approach Contain Streaming Costs?

For live events, Cloudmass has the potential to reduce costs by using multicast streaming. In unicast streaming, every client that wants to watch a live event gets their own individual stream. So if one person is watching an HD video of their local college basketball game on their connected TV, it would consume 3 mbps of bandwidth between the TV and the server. If 10 of his neighbors are watching the same game at the same time, that would use 30 mbps. Even though all the people watching the game are in the same neighborhood, the server must send each person their own individual stream. This also means that streaming charges will be 10 times higher than streaming for one person.

With multicast, the first person to watch the basketball game still uses 3 mbps. But if the second person watching the game is in the same neighborhood, they simply receive the same stream that is already being sent to the first person. And if 10 people in the same neighborhood start watching the game, they all watch the same stream. No matter how many people watch the local college basketball game, the bandwidth used between the server and the clients is just 3 mbps. That's a 90% saving in streaming costs over unicast.

The Cloudmass approach may also contain streaming costs in another way. With on-demand and live streaming, audience levels can be very unpredictable. The resources in the mesh are constantly monitoring streaming load and quality. If an onset of a spike in demand is detected, the mesh requests additional cloud-based resources. These new resources are readied and allocated by the cloud services in real time to handle the spike in demand and once the spike is passed, the resources are released back to the cloud service. This is a very cost-effective method since charges for the cloud resources are only incurred while they are being used.

7.0 Summarizing the Cloudmass Approach

The Cloudmass mesh approach has some distinct differences from traditional CDN delivery:

- Cloudmass does not require the maintenance of huge over-capacity to cope with expected and unexpected spikes in demand. This, potentially, removes large amounts of capital and operational costs from the equation.
- Utilizing multicast offers drastic scale increases in the last mile and the potential for large reduction in streaming costs.
- The mesh network can flexibly recruit additional resources to manage variability in video demand. Using this strategy allows the streaming network to scale to handle any size live or On Demand broadcast event.
- The mesh is resilient, meaning it can withstand server and network failures by seamlessly switching load to other resources.

By recruiting a diverse set of computing resources to cooperate in the delivery of streaming video, Octoshape's Cloudmass may be very helpful in solving the problems of scale and cost for content providers. In breaking the relationship between distance and quality the Octoshape solution takes direct aim at the CDN "ante" requiring massive capital investment. The pooling of multiple clouds together at once means it is possible to grow to any amount of streaming capacity without investing in permanent resources to handle unexpected demand.

8.0 Conclusion

While traditional CDNs have built massive networks to handle web streaming that is subject to the constrictions of the distance/quality equation, Octoshape has taken a different approach. Octoshape's Cloudmass creates a mesh network of servers, cloud resources, and multicast devices that cooperate to deliver scalable, quality streaming video. This allows Cloudmass to grow and shrink, recruit resources across multiple cloud infrastructures, stream from multiple locations, and do so transparently to the end user.

By leveraging all of these resources simultaneously in the delivery of streaming video, Cloudmass clearly has the potential to contain the cost of video delivery while at the same time scaling to any number of web viewers.

Presented by  **octoshape**

TDG Transparency Statement

As a leader in the analysis of the emerging digital delivery market, we are often asked to help companies explain new technologies as well as market challenges and opportunities. These research pieces are released as TDG Private Papers and we've created this disclaimer to maintain full transparency surrounding this research. We maintain high objectivity standards and have created these simple rules:

1. TDG's time has been compensated, but our opinion is not for sale.
2. TDG retains final editorial control.
3. TDG Private Papers do not include rankings or potentially subjective rating systems.
4. TDG Private Papers are not released on our website.

Copyright Notice:

This document was published by
The Diffusion Group, Inc.
Copyright 2011.

Reproduction is forbidden unless authorized.
All rights reserved.

About TDG

TDG provides actionable intelligence on the quantum shifts impacting consumer technology and media behaviors. Since 2004, our market research and advisory services have helped technology vendors, media companies, and service providers understand how consumers access, navigate, distribute, and consume broadband media - whenever and wherever they may be.

For more information, visit our website at <http://www.tdgresearch.com>.