Successful Caching for Internet Media

Executive Summary

Internet traffic, particularly from media usage, is skyrocketing. This is both good news and bad news for Internet service providers (ISPs): while customer demand for the Internet rises, ISPs contend with new business and technical challenges. The cost of extra bandwidth cuts into profits, while users expect expanded service without paying more. Surges in usage as customers seek a popular video threaten to slow response times, lowering the customers' Quality of Experience (QoE).

Web caching, where a proxy server saves copies of objects as they're first downloaded from the Internet and then uses the local copies for subsequent requests for the same objects, can cut down bandwidth usage and deliver objects faster. Traditional caching techniques, which have been deployed at most ISPs, tend to miss media objects, rendering these techniques less effective as customers seek more and more video and audio online. New intelligent caching can be designed to catch media objects and significantly save bandwidth while improving download times and customers' QoE.

Internet Consumers Want More and More

It's no surprise that Internet usage continues to climb, maintaining a trend that's been steady since the inception of the Internet. In recent years, global Internet traffic has been rising at staggering rates, as consumers, feeding their desire for online media, demand more and more bandwidth. ISPs who can deliver fast, reliable access to video and music sites stand to gain customer satisfaction and loyalty, but this growth also brings business and technical challenges.

Traffic, Particularly from Media, is on the Rise

Studies show that the demand for online media is stronger than ever and not likely to slow anytime soon. The number of users seeking online video has grown 339 percent in the last six years. More tellingly, the time those users spent watching video climbed almost 2000 percent. In the last year, the number of unique viewers grew by 10 percent, and their time watching online went up 71 percent.[1]

Industry analysts predict a fivefold rise in global Internet traffic in the next five years,

[1] Report, "Social Media And Video Site Engagement Reshapes The Web," Nielsen.com, April 22, 2009

due largely to the rise in media downloads. 90 percent of the predicted 667 exabytes of global traffic is expected to be from video downloads and sharing.[2]

New Challenges for ISPs

While these trends imply an increased demand for Internet services, ISPs face a new set of technical and logistical challenges that come with exponential growth. Higher customer demand doesn't necessarily translate to increased revenue for an ISP, as it often requires costly investments in network infrastructure and bandwidth. As the industry becomes more and more competitive, profit margins shrink.

At the same time, customer expectations of reliability are higher than ever. As the Internet replaces broadcast radio stations and televisions, customers expect broadcast-quality playback for audio and video on the web. Consumers are turning to their computers for coverage of breaking news stories, and they rely more on immediate access to online video. An important event now creates a large demand for online video coverage, which can cripple networks across a country.[3] Slow response times, stuttering media, and outages are not acceptable to customers, and poor Quality of Experience (QoE) leads to subscriber turnover. Even customers who haven't experienced QoE issues may make decisions based on reports of these issues from other, vocal customers. Since the rise in social media, a single, vocal customer's dissatisfaction can become a very public issue if the customer vents his or her frustrations on blogs and review websites.

Regulating Bandwidth: A Risky Solution

As the demand for Internet usage goes up and up, some ISPs resort to imposing bandwidth caps or changing the caps they already have in place. Others limit the amount of data allowed to pass through in a specified period of time, called throttling. Changes in bandwidth allowances and throttling can help ISPs get through peak hours, but they risk lowering customer satisfaction and generating negative publicity.[4] Users don't want to feel needlessly restricted by their ISP, and this perception can affect their satisfaction with a provider. With more mainstream media coverage of technical issues, customers who in the past would never have known nor cared about practices like bandwidth caps and throttling now take them into account when choosing an ISP.

[2] Report, "New Cisco Visual Networking Index Forecasts Global IP Traffic to Increase Fivefold by 2013," Cisco.com, June 9, 2009

[3] Article, "Online Video of Inauguration Sets Records," *New York Times*, January 20, 2009

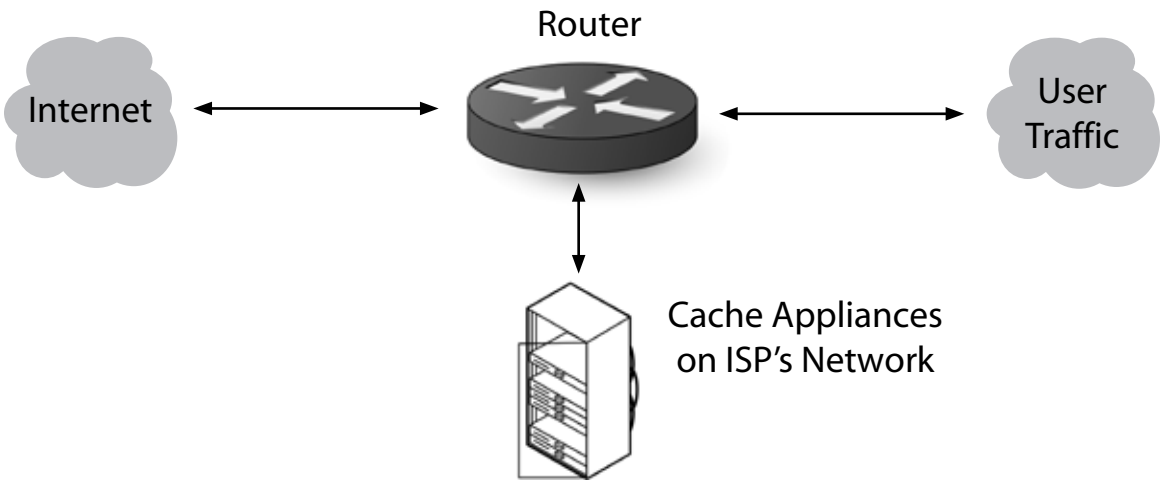
[4] Article, "BT Seeks to End 'Free Ride' by Video Websites," *Financial Times*, June 11, 2009

Web Caching: An Essential for Survival in Today's Market

To reduce bandwidth usage and improve speed, ISPs regularly deploy web caching, storing copies of web documents such as HTML pages or media files on their own network and then using the local copy for subsequent user requests for the same content. Caching proxies are everywhere—most systems that connect users to the Internet use one.

Web caching works by intercepting the client's request for an object and checking to see if an up-to-date copy is available. If so, it answers the client's request with the copy. If not, it fetches a copy from the source out on the Internet, saves it for future requests, and returns it to the client. How does the caching proxy know when a file is up-to-date? RFC 2616, which defines the current version of HTTP, establishes one method for validating cached objects without re-checking with the Internet source. A *freshness lifetime* sets an expiration time for each object. If the age of the cached object has exceeded the freshness lifetime, the object is said to be *stale*. Otherwise it is considered *fresh*. Stale objects must be re-checked before they can be returned to clients.

ISPs deploy web caching with *caching servers* and *cache appliances*. A cache server per-



forms web caching with software installed on the ISP's servers. Most ISPs have deployed caching proxy servers, like the free software Squid, to speed up their users' access to HTTP, FTP, and gopher data objects. A cache appliance combines hardware and software to deliver more powerful and efficient caching. Cache appliances dedicate memory to caching, avoid the need to configure other hardware in the system for caching, and are optimized for the fastest return times.

Once a caching proxy decides to keep a local copy of an object, it needs a place to store it. Many software-only cache servers store objects on general-purpose file systems, which can introduce their own latency issues because they aren't designed for the rigorous disk access required for caching. Most servers can recover from inefficient disk reads faster than they can retrieve a large object from a remote server, so caching with software-only cache servers shows an advantage over not caching at all. Some of the advantage gained can be lost to inefficiencies on the local system. Cache appliances, which have been designed to handle the demands of writing and reading cached objects, avoid many of these problems and further speed up client access to cached files.

The connection to a caching proxy can be explicitly set up in a browser, or it can involve a little sleight-of-hand between the browser and the websites the user visits on the Internet. Often, when customers use an ISP, they configure their browsers to specify a proxy server to use. Routers can also be set up for *transparent caching*, where all traffic on port 80 (the HTTP port) is diverted to a proxy cache. This requires no extra configuration in the user's browser. The process is often called *semi-transparent caching*, because the remote server will see the IP address of the proxy server and not that of the end user. Occasionally, semi-transparent caching causes problems with websites, which tips off users that a proxy cache is operating between their browser and the Internet at large. A *fully transparent* proxy cache, however, uses IP spoofing to transmit packets that look like they came from the end-user. Full transparency requires careful network routing, but when operating properly, both the client and the server work as though they are not aware of the proxy.

Caching for Media Objects

Media downloads present special opportunities and challenges for web caching. Because media files can be large, effective caching offers vast potential for bandwidth savings. The "viral" nature of popular clips only magnifies this effect: when a significant percentage of an ISP's clients want to download the same clip, that ISP can gain tremendous savings and greatly improve download times by caching it.

Unfortunately, media files and the sites that serve them tend to resist traditional caching methods. The most popular media sites use complicated, dynamic URLs to deliver video and audio files. A website generates *dynamic URLs* when it stores its contents in a database and builds pages on demand for users. (These differ from *static* URLs, which have been around since the beginning of the Internet and do not change depending on how

they're accessed.) Dynamic URLs are often easy to spot, since they tend to include question marks and other characters.

A relatively simple dynamic URL looks like this:

http://www.youtube.com/results?search_query=crocodile+lion+buffalo&search_type=&aq=2&oq=crocodile+lion

URLs of embedded video and audio objects, which do not appear on a web browser's screen but are essential for delivering media, are often more complicated. Some media sites deliberately use one-time or time-limited URLs as a security measure, and others include server details or even the client's IP address in the URL. Often, no two views of the same file, even from the same user, use the same URL.

Traditional caching solutions rely heavily on an object's URL, and many don't cache dynamic URLs at all. These solutions therefore effectively cache only the most static objects on the web, like HTML pages and some images. Caching these static objects does generate some bandwidth savings, but this method misses the media files that now make up the bulk of Internet traffic.

Intelligent caching appliances designed explicitly for media objects can break through these limitations by analyzing the contents of media objects or by addressing known URL patterns for popular media servers, like YouTube, Google Video, and the like.
A caching proxy that has been designed for video and audio will have other advantages over traditional caching methods. Determining the freshness of a clip requires careful attention, and validation methods for media content must keep current to be sure only valid objects are cached.

Web Caching with CacheApp XXX

CacheApp XXX, Company Y's high-performance Internet cache appliance helps ISPs manage their customers' increasing demand for media. Designed for optimal caching of media files, the CacheApp XXX can deliver hit ratios unheard of with most cache servers.

- By serving media files from within the ISP's network, the CacheApp XXX reduces the backhaul bandwidth consumed by repeated downloads of frequently requested media.
- CacheApp XXX supports both Web and video caching in a single solution. It can be deployed side-by-side with an ISP's existing web caching solutions, so the ISP can address new traffic growth and new types of traffic without the need to alter its existing infrastructure.
- Since fully transparent and semi–transparent deployments are both supported, the CacheApp XXX facilitates different network configurations.
- As it improves network efficiency for video and audio delivery, the CacheApp XXX supports subscriber growth without massive investments in additional network infrastructure and bandwidth. ISPs can deliver more content without oversubscription of costly transit and peering links.
- ISPs deploy the CacheApp XXX to reduce latency, serving the data as close to the subscriber as possible. Serving cached files reduces delivery times considerably and reduces buffer waiting periods. Users see a near instantaneous rendering of the video they requested.
- With these faster response times, ISPs give users a better video viewing experience and greater customer satisfaction. With accelerated media delivery, ISPs can promote premium price points for top-tier broadband packages.
- CacheApp XXX caching validates media content and provides expiration functions that determine which streams are cacheable. Careful monitoring verifies that cached content was not changed at the source, ensuring that only fresh content is returned to users.
- ISPs can manage and monitor the CacheApp XXX remotely. CacheApp's management interface facilitates all configuration via an easy-to-use web based GUI.