Powering your Web Strategy with CDN Services

April 2007

Table of Contents

Executive Summary

The Web has become an integral component in the business activities of many organizations. As a result, an Internet infrastructure has become a keystone in many corporate technology blueprints. The planning goal is to leverage an infrastructure that includes hardware and software to enable Internet operations to effectively support the corporate Web application strategy.

Growing demand has overwhelmed many Internet infrastructures. Wave after wave of new users, or traffic spikes, have flooded Web servers with requests resulting in demand for more bandwidth and processing power. In addition, the advent of bandwidth-intensive content, like video and other rich media, has further burdened the Web infrastructure. To address growing demands and to ensure the best corporate face for Web visitors, many corporate planners have resorted to costly capital investments to expand the organization's infrastructure, and purchasing additional and larger Internet connections to support their bandwidth requirements. Other organizations have outsourced their entire Web infrastructure, or individual components, to avoid large capital expenditures and operating costs. This tactic enables the company to focus on their core business. There are traditional outsourcing options, such as centralized hosting and site mirroring, as well as distributed Internet infrastructure services such as Content Delivery Networks (CDN).

Centralized Hosting still requires ongoing investment in equipment and maintenance. Like an in-house solution, scaling requires the addition of more hardware, which requires long-term planning and can not respond to rapidly growing demand as well as event-driven or seasonal traffic spikes without over-provisioning. With hosting's single point of failure, the availability of your site is also susceptible to technical failure, natural disaster, and other threats. If the site is not available, your business suffers.

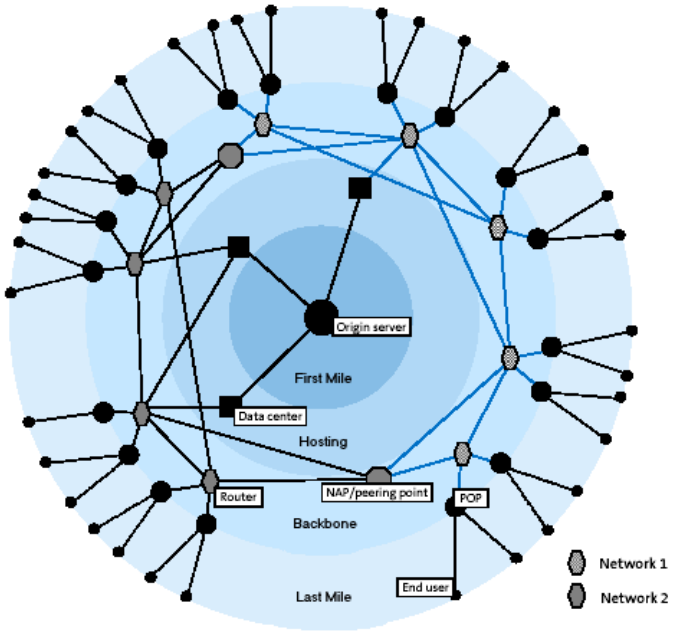
Site mirroring attempts to address some of the reliability issues associated with the Centralized Hosting model. However, mirroring requires a substantial investment in hardware and operating costs and it does not address many of the performance problems inherent to the Internet. In addition, site replication introduces data synchronization challenges, which increases the complexity and cost of maintaining your site.

Alternatively, an organization can purchase managed Internet services to address specific business needs such as scalability, reliability, performance and security through a CDN provider. The Internet infrastructure of a CDN furnishes a set of set of servers distributed across the Internet. The servers provide services and a network that ensures high-speed global performance, on-demand scalability, uninterrupted availability, and security. The services enable an organization to reduce their infrastructure investment and still meet the demands of Internet traffic and strategy.

The Internet Challenge

The Internet has proven to be a valuable resource for business, but it is also highly problematic. Every user has experienced the frustration of slow responses and site outages. From its inception, the Internet was never centrally planned. It grew organically with various networks agreeing on transmission standards and linking to one another. Today's Internet consists of a large number of private networks that interconnect to each other, forming one global inter-network. It is a vast heterogeneous system of networks, routers, and other devices that can transport virtually any type of data to anywhere in the world. However, the Internet was never designed to handle the diversity and volume of traffic that it now supports. Its performance shortcomings reflect an architecture that is performing well beyond its original intent.

The Internet can be divided into four segments (see below). The connection between the origin server and the Internet is referred to as the First Mile. The Hosting segment refers to Internet data centers and the network infrastructures that exist within data centers. The Backbone segment includes the fiber connections that link data centers to points of presence, as well as the public and private peering points. Finally, the Last Mile represents the connection between the end-user and the Internet.



Poor Web site performance can often be attributable to the First Mile. This is especially true when the content providers choose to manage their own Internet infrastructure. Limited bandwidth between the origin server and the Internet can be the source of delays and outages. Web site data served from data centers and managed hosting service providers can bypass the bandwidth issues of the First Mile. However, this solution cannot solve many of the other Internet network performance challenges. Data centers can offer scalable Internet infrastructures and on demand capacity to meet traffic surges and typically perform better than data that must traverse the 'first mile,' but the data center solution can not address the key issues that cause inherent Internet latency and interruptions.

Most Internet latency and reliability problems can be traced to the Internet Backbone. The Internet's architecture is extremely complex and its data pathways can be intricate, illogical, and occasionally sub-optimal. As data packets travel through the Internet they are routed through many points. Each point-to-point link in a data pathway is referred to as a "hop." Any given transmission can consist of many hops. In general, a larger number of hops results in greater total latency or time delays, which in turn reduces perceived overall performance. The data does not travel through a pre-defined pathway, but rather it is routed at each point in the network. Each hop requires router processing to determine the appropriate path. This, in turn, reduces overall transmission speed. Occasionally, some routers are overwhelmed with traffic, which can impair performance and lead to packet loss. While the router processing considers network conditions, its scope is often limited to conditions in the immediate network vicinity and it does not include the overall state of the network. As a result, packets can be routed in ways that appear optimal from a local perspective, but are sub-optimal overall.

Finally, the Last Mile is often the source of significant Internet performance and reliability problems. Since this is due to the nature of the connection between the end-user and the Internet, content providers have little control over last mile performance. Further, it's a common misconception that the foreseeable displacement 56K dial-up connections with broadband will solve Internet performance problems: it will only highlight the limitations of the other three Internet segments. In addition, the growth of broadband access has been accompanied by richer, more bandwidth-intensive content, like video, that causes further Internet congestion.

The CDN Solution

Internet service and technology providers have responded to the market demand for greater Web site performance with a myriad of solutions. Each technology and service attempts to provide faster content delivery to end-users. CDN vendors have developed network architectures and software solutions that address the inherent Internet performance problems.

A Content Delivery Network delivers Web content and applications from a set of strategically located and load-balanced servers. The distributed infrastructure provides services to ensure a Web site supports the needs of

the business without costly capital expenditures and operating costs. It is a global network of servers that provide load balancing, object caching, streaming, authorization, secure access, usage analysis, network monitoring, replicated storage, and other services. CDN services reduce the infrastructure investment required for a Web site to deliver fast and reliable performance for normal and peak demand. An enterprise can easily enable the CDN solution and, with little or no changes to their own systems, quickly offload a significant portion of its processing and bandwidth needs.

CDN services overcome many of the performance issues associated with the Internet by deploying caching services located on the Internet near the end-user. The infrastructure provides a distributed set of caches, with a routing layer that determines which cache is closest to any given Web site visitor. Commonly requested pages are served from the cache location closest to the user, circumventing Internet hops and minimizing access to the content provider's Web site infrastructure. With caches close to end users, a CDN can eliminate many content delivery performance obstacles, including those associated with the Internet backbone, without costly hardware and software investments in the content provider's web site infrastructure. Content providers can continue to control their web sites while leveraging the technology and expertise provided by the CDN vendor. Analysts estimate that customers can achieve savings between 12% and 85% by utilizing a distributed content delivery infrastructure along with other managed web services.

The CDN architecture has two direct benefits for a content provider. First, from the perspective of the web site visitor, caching reduces the time it takes to retrieve objects, thereby improving the visitor's Internet experience. Second, from the perspective of the web site, distributed caching reduces the number of requests for a particular object. As a result, the web site can effectively support a greater number of visitors with a less costly hardware configuration.

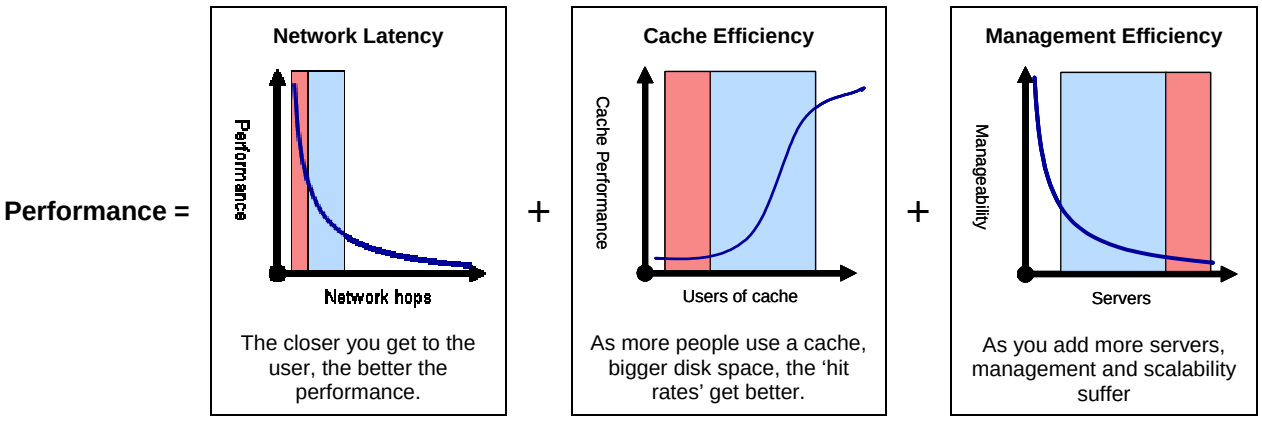
Not Every CDN is Alike

Some CDN vendors have architected their network with thousands of locations placed very close to end users. Other CDN networks are designed with fewer and larger servers strategically situated between end users and the Internet backbone. These locations optimize proximity to the end users, capacity, and peering (the interconnection between networks within the Internet backbone).

The Gilder Technology Group describes the CDN architecture with many small locations as "the 7-Elevens of the content distribution space. … Sure it's convenient to stop by and pick up the morning paper and a Big Gulp® on the way to work, but notice the manager's face when you ask him for directions to the shoe department." In this "convenience store" model, the cache servers are smaller and dispersed close to the end-users, but they have limited content. Constrained by space limitations, the distributed convenience store architecture must prioritize cache content – not all "items" are available at every store. As traffic increases, the routes mapped by their intelligent networks across the Internet to find requested content become more convoluted. The average data packet must contend with many more routers and resulting network hops. Furthermore, data traffic jams at one or more routers are not uncommon.

With fewer and larger servers, the distributed "superstore" architecture is not limited by cache size. The superstore architecture with fewer and larger servers provides more processing power and storage than the smaller-scale characteristic of the convenience store server. The superstore infrastructure is not forced to prioritize content. In addition, the "superstore" architecture can ensure the integrity and security of web content by deploying servers exclusively in class A co-location sites. CDN architectures with 1000's of small locations cannot rely solely on such a limited number of global locations. For example, "convenience store" CDN vendors deploy servers at private data centers and academic institutions. Unfortunately, these sites can not always provide robust security and control characteristic of class A co-location sites to protect the systems and your data.

The ability of the CDN to improve performance depends on several variables including network latency, cache efficiency, and management efficiency. Network latency is attributable to Internet traffic, as well as the number of Internet nodes the content must cross to reach its destination. Locating the content closer to the end user can reduce latency, however it has implications on cache efficiency and management efficiency.



Conceptually, you could replicate your entire web site near every user, but this is not practical. Content Delivery Networks distribute content close to users by storing commonly requested objects in cache. Performance is improved if the object is in cache when a user requests it. The object is in cache if a prior visitor previously requested it. The probability that a particular page would have already been requested from a particular server is much higher in the superstore infrastructure, so any particular object is more likely in cache when it is requested, avoiding additional request routing and subsequent network hops.

The more concentrated superstore model also adds greater content management efficiency. Configuration or content changes can be deployed faster. For example, when an object is modified, it would take much longer for the change to be applied across the vast, intricate set of servers in the convenience store network. As a result, it's more likely that your visitors would view outdated content.

How to Select a CDN Vendor

Your web infrastructure must support the business needs of your organization. It must provide a foundation that enables you to cost-effectively fulfill your Internet strategy. A CDN solution can help you meet this goal, but not every CDN solution is the same. It's important to evaluate your options based on a sound set of criteria including global performance, on-demand scalability, uninterrupted availability, content security, and implementation time and costs. Most important is the ability to minimize your service costs and understand how variable traffic patterns and surges will impact your budget.

Unpredictable Traffic and Costs

Traffic can surge to more than ten times average and exceed site capacity, degrading performance or even resulting in outages. Traditionally, sites have addressed the problem by over-provisioning hardware and bandwidth. This is not cost effective since most of the investment would remain unused outside of peak traffic periods.

CDN services can help you avoid costly investments to overbuild your web server infrastructure in order to handle peak loads or intermittent traffic spikes (flash crowds). That means your server infrastructure can be much slimmer, which reduces your capital costs for things like servers and switches. CDN support can enable you to accommodate web site traffic spikes without costly infrastructure investments. A related benefit is that an outsourced service model helps your company stay focused on what you do best, which isn't necessarily building and operating web site infrastructure.

From a business and planning perspective, it is also important to be able to forecast your costs. Unfortunately, many CDN price models amplify costs resulting from variable traffic patterns and impose additional "bursting fees" when your traffic exceeds committed levels. With the CDN, your web site could handle unusual peak traffic resulting from a marketing program or some unforeseen event, but the resultant unpredictable fees could undermine your budget.

A Dependable Budget

Your CDN solution should protect your web site from traffic surges and ensure predictable performance. The solution should also protect you from excessive cost variability. Unforeseen service charge spikes can make it

difficult to manage your budget and your business. Unlike other vendors, Mirror Image offers pricing options that allow you to minimize your costs and provide predictability in spite of variable web site traffic patterns.

Typically, CDN service fees are based on peak bandwidth measured in megabits per second (Mbps). These models are characteristic of traditional pricing techniques for bandwidth providers and offer a relatively effective metric for traffic patterns with comparatively constant traffic levels. However, these models tend to amplify the costs resulting from traffic spikes. The Mirror Image pricing options include the traditional bandwidth model (Mbps) as well as pricing based on the quantity of data delivered (GBytes served), a more predictable metric. For instance, an executable file delivered via the Internet at noon on a business day (a period of peak demand) costs exactly the same as the executable delivered at midnight on Sunday morning (a period of low usage).

Since every site experiences different traffic patterns, no one pricing model is best for all. Mirror Image offers both models so you can select the option that best fits your needs to minimize costs and provide a predictable expense. In addition, there are no overage penalties or bursting fees, and Mirror Image allows you to rollover unused CDN and streaming dollars.

Many CDN vendors charge their customers the maximum contractual monthly fee even if the customer has not consumed the purchased capacity for that period. Customers can not carryover unused services from one month to the next. In contrast, Mirror Image allows customers to rollover their unused contractual monthly services with the Total Value Plan. Mirror Image gives you more control over your costs by matching your service charges to your usage patterns, with flexible terms, more predictable costs and no unanticipated fees.

Things Are Not Always How They Appear - Measuring Performance

Object caching is a core technology of a Content Delivery Network. Through caching, the distributed infrastructure can serve commonly requested objects from the cache location closest to the user without accessing your web site origin servers. The result is improved performance for your site visitor and reduced processing for your web server. The key metric for evaluating the effectiveness of the caching process is the "cache hit ratio." This measures the portion of requests that are served from the cache. The higher the ratio, the more often the cache is able to directly deliver the object without retrieving content from your origin.

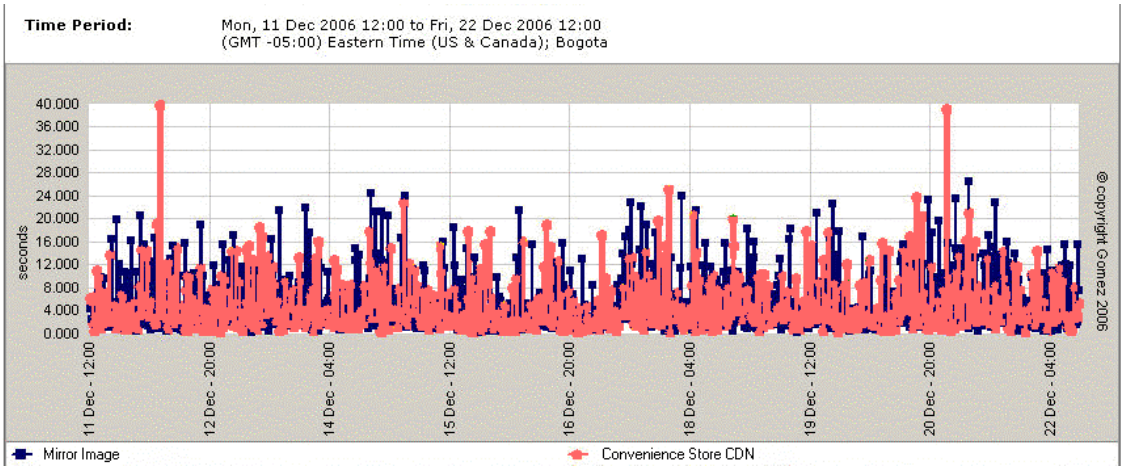
The cache hit ratio not only measures the ability of the delivery network to improve performance, it also reveals how effectively the distributed infrastructure will reduce your web site costs. The more often the CDN can serve the object from cache, the slimmer your server infrastructure can be, while still delivering the required level of performance. When your web page object is not in cache, then latency is added because the delivery network must retrieve the object from other caches or, worst, from your web server, increasing processing demands on your infrastructure. A CDN with a higher cache hit ratio can reduce network latency and origin server workload, while accelerating the user experience.

Superior Performance and Infrastructure Savings

Mirror Image brings the Internet closer to users, and reduces content provider costs, by storing, processing and delivering Internet applications and web site content through its content delivery services layered over its global Content Access Point® (CAP) network. The CAP network is a specialized content delivery system with built-in fail-over that plugs into web applications easily and rapidly, adding virtually unlimited scalability, reliability, efficiency, stability, and consistency to the Internet operations of Mirror Image customers.

Mirror Image has built and deployed high-capacity CAPs at optimal Internet exchange locations within major Internet usage and population areas. These Internet access points have become more concentrated and the Mirror Image CAPs are situated at the natural convergence of Internet networks in their respective regions and their public and private peering points. The proximity of the CAPs to these Internet exchange locations enables the Mirror Image network to effectively overcome Internet latency performance problems. The growing concentration of regional access in specific Internet exchange locations has virtually eliminated any performance gains attributed to caching content closer to the edge beyond these strategic exchange points. In addition, the greater capacity of each CAP enables Mirror Image to deliver more performance, processing power, and storage than the small-scale server characteristic of the convenience store model.

CDN Performance



The chart above illustrates a comparison between the Mirror Image network and a leading CDN designed with thousands of cache servers dispersed close to end-users. Performance was measured for more than 10 days at 15-minute intervals based on the delivery of an object of similar size (about 11 Kbytes) through each network and monitored by globally distributed agents located on end-user desktops. Average performance for the testing period for each network (i.e., response time) was within 0.6 seconds.

The Mirror Image delivery network achieves greater cache effectiveness with a 98% cache hit ratio compared with a 70% to 80% cache hit ratio generally available through convenience store infrastructures. Mirror Image attains its superior cache hit ratio, and the resulting faster response and reduced traffic at the origin server, by storing the majority of a web site's content in each CAP. In contrast, the limited storage characteristic of the distributed convenience store model restricts the volume of content each cache can maintain so the system must resort to complex routing systems to find the content when it's requested.

Mirror Image Solution

Mirror Image's content delivery infrastructure provides a scalable platform that reliably delivers web content to users around the world. The Mirror Image content delivery solution is based on the patented Content Access Point (CAP) network. This global network of strategically located CAP facilities works to automatically identify and place your web content closer to millions of users worldwide.

Mirror Image has recognized that the key to optimizing content delivery goes well beyond simply bringing content physically closer to the end user. Unlike other solutions that solely rely on a highly distributed network of small server appliances, Mirror Image's global CAP network combines an optimal mix of proximity, connectivity, processing power, storage and software. This innovative approach effectively offloads origin servers, protects existing infrastructures from surges in demand, and bypasses Internet congestion to ensure the fast and reliable delivery of content, regardless of user location or demand. It also increases your web site capabilities and significantly reduces bandwidth requirements, capital expenditure, and management costs while expanding your global market reach. The Mirror Image solution allows you to optimize the performance you provide your web site visitors, regardless of traffic surges, control your web site costs, and manage your budget.

The Mirror Image distributed infrastructure improves your web site visitor's experience with an advanced network design, and performance that is backed by quantifiable metrics. With its unsurpassed cache hit ratio, the Mirror Image CAP network effectively overcomes Internet congestion and latency, and trims your web server processing requirements. The greater capacity of each CAP enables Mirror Image to deliver more effective caching, and better performance, than the small-scale characteristic of the "convenience store" infrastructures.

Mirror Image also helps ensures the predictability of your costs to help you manage your budget. An object returned to a web site visitor at noon on a business day (a period of peak demand) should cost you exactly the same as that object returned at midnight on Sunday morning (a period of low usage). Mirror Image's "quantity delivered" pricing approach (Gbytes) can mitigate unforeseen cost spikes. With the option of either the "quantity delivered" pricing model or the traditional "peak Mbps" pricing model, along with no overage penalties or bursting fees and no administration traffic charges, Mirror Image can help you control your costs and eliminate budget risk.

The Mirror Image Internet infrastructure leverages the power of optimally distributed servers to deliver high-speed global performance, on demand scalability, uninterrupted availability, security, and rapid implementation. The Mirror Image services can enable you to reduce your infrastructure costs and still meet the demands of your Internet traffic and strategy. With Mirror Image, you can ensure reliable performance, while your control infrastructure investment and managing your budget.