Building a CDN with F5

Content delivery networks are useful services for transferring data over long distances, delivering an optimized user experience, and securing data in an otherwise public environment. However, concerns about cost and control of critical data drive many organizations to seek an alternative delivery mechanism, one that can be achieved with F5 products.

White Paper

by Don MacVittie

Introduction

Content delivery networks (CDNs) offer a large, geographically dispersed network of servers and optimization tools that enables customers to rapidly deploy and propagate data across multiple regions. Many organizations use CDNs to share data across geographic boundaries, optimize delivery of complex content, and secure the transfer of data to affiliates in remote locations.

Although CDNs have been available for many years, recent developments in the cloud space—which is similar enough to CDNs to confuse many IT professionals—has driven increased interest in CDN solutions.

While the functions that CDNs deliver are beneficial, the way that they are implemented can vary widely. When all options are considered, implementing an internal system is often a more cost-effective choice for organizations with offices in each of the geographies where data and applications are needed.

There are two primary reasons that organizations choose to avoid a CDN deployment. The first is the ongoing monthly cost of the service and any additional services—such as application acceleration—that may be needed. When the amount of data being transferred, the number of users online, and the complexity of the networking systems that serve up web applications and back-end replication increase, the monthly cost of a CDN also increases. The second reason for avoiding a CDN is control of data and the desire to manage optimizations within the IT department rather than at an outside provider. The control of data is a significant issue, with numerous security concerns driving some organizations to look for a viable alternative. The most prevalent security concern is a requirement that if the connections and data are to be secured, the CDN vendor will need copies of relevant certificates, which most companies prefer to keep in-house.

For organizations that have offices in several different geographic regions, however, there is an alternative. The F5 BIG-IP platform with BIG-IP Application Acceleration Manager (AAM) and BIG-IP Global Traffic Manager (GTM) can deliver an internal CDN managed by the IT department with no monthly fees and with bandwidth limited only by connection, not cost per megabyte.

CDN Use Cases

There are a few specific reasons that an organization contracts for a CDN. These use cases involve offering something that, at the time the contract was signed, the enterprise could not or was not willing to deliver from within the IT department.

A CDN as an application optimization engine. When applications are being delivered to low bandwidth or high latency clients (such as in certain geographic areas or via cellular network), application optimization can reduce the amount of data being sent while increasing the utilization of available client bandwidth. CDNs connect the client to the geographically closest copy of the application, which has the effect of reducing the latency introduced by the network.

A CDN for file replication. When large data files are required at multiple geographies, timely replication of data from a corporate headquarters in one region to one or more regional headquarters can be enabled with a CDN deployment.

A CDN as an unlimited bandwidth source. When an application has a very bursty usage pattern or is expected to experience rapid but ill-defined growth, a CDN is often used to ensure that the application or service remains responsive no matter how high the request volume becomes.

Each of these use cases utilizes the CDN to address a perceived problem with the organization's existing infrastructure.

How Much Benefit?

In one deployment, an enterprise CDN cost 16 percent of the cost of contracting for three years of comparable services from a CDN vendor.

The CDN as an Application Optimization Engine

Application optimization is a chore frequently and readily handled within an organization's normal application infrastructure without resorting to the use of a CDN. Still, some organizations need to deliver a high-performance application to a variety of geographic locales and several different types of client systems. CDNs make this delivery easy by placing the application in question close to several geographic locales and directing traffic to the one closest to the client, thus reducing latency and speeding performance. Some CDNs also offer separate application protocol acceleration, such as optimizations to TCP that will enhance the performance of applications when communicating with low-bandwidth clients. Finally, some CDN providers have allowed for application optimizations such as selective compression and image resizing to match the client's needs and reduce the number of bytes sent over the wire. The combination of all of these optimizations can drastically improve application performance—a common rationale for CDN usage. Currently, however, CDNs do not have specific optimizations that account for—and take advantage of—the unique constraints of mobile clients.

Many CDN providers offer these optimization services, but most of them charge extra fees for each of them, increasing the monthly bill. While many organizations find the extra expense worthwhile for targeted applications, the fees can be a barrier to those wishing to introduce new applications into this environment.

The CDN as a Replication Engine

For truly dispersed organizations, synchronizing or backing up data across a large geographical area can be problematic. While there is almost always a “database of record” for structured data, making certain that all other data centers are in sync with this database can be a painful process. Additionally, sharing unstructured data across large geographic distances can be problematic from the perspectives of network bandwidth and backup considerations such as recovery time objectives (RTOs) and recovery point objectives (RPOs).

CDNs are a common mechanism for mitigating these issues. The primary source updates the CDN, and then all other locations pull from the CDN source, not tying up the bandwidth of the primary source. Since CDNs generally have large connections to client sites, the speed of delivery is greatly enhanced and the amount of time the primary data center spends updating is greatly reduced.

Although not specifically replication, other common uses of a CDN include sharing files on a regular basis with regional offices. The performance improvements offered by CDNs provide clients with a more timely update process and allow for tighter scheduling of updates.

The CDN as an Unlimited Bandwidth Source

For a wide selection of web applications, forecasting the number of connections per second is problematic. While everyone wants their applications to be found useful by the target market, too much access is a curse unto itself when legitimate overuse lengthens application response times. There are also cases where the nature of the application or target market makes connections come in bursts that spike, after which usage levels off or even drops to zero. CDNs are often used in these scenarios to mitigate the effect of high-volume traffic and keep the application performing at an acceptable level.

The monthly cost of a CDN in this instance is acceptable based on the protection of the organization’s corporate identity and the assurance that the application will be up during high traffic times. But the cost of maintaining such an application on a CDN is directly related to the usage rates the application sees. While most CDN providers charge a standard monthly fee, they also charge for bandwidth over a threshold set by the provider or contract. These extra fees can mount, so for organizations with offices in the correct locations, it is worth considering replacing a CDN with another viable option.

Each of these cases has a strong business driver, but each also has associated expenses that can be rather high—and that represent a lot of funding per year that could be invested in alternative solutions. In addition, a vendor contract invokes administrative and managerial complexities, from control of optimization strategies to data security. Not least of these concerns is the necessity of providing the organization's security certificates to the vendor. This loss of control is not acceptable for every organization.

In addition, some organizations have infrastructures nearly sufficient to answer the issues within their own network and thus save monthly CDN payments. One alternative is a private, enterprise CDN implemented by the organization's IT department, with the costs incurred just once, as a capital expenditure, and with the organization retaining all control. The ROI period for deploying a private, enterprise CDN can be one to three years, depending on traffic growth and associated rising costs. The decision to go with a predictable CapEx versus variable OpEx depends upon each organization's business requirements and approach to application delivery.

F5 Solutions for CDN Use Cases

If an organization has offices in all of the major geographies where it provides service and also has an application infrastructure capable of serving applications at the rate they will be requested, the only missing piece of the application delivery optimization puzzle is network infrastructure that can handle the demands of each of the CDN use cases.

With an Application Delivery Controller (ADC) infrastructure sitting at critical points of control in the network, it is possible to meet all of these requirements. An F5 ADC, in particular, meets these requirements and allows the organization to place as many applications as desired into the resulting enterprise CDN without increasing per-month fees.

Specifically, F5 products work together to address each of the purposes of a CDN contract:

1. Application optimization is implemented in BIG-IP AAM. BIG-IP AAM handles optimizations at all networking levels, accelerating applications protocols, network protocols, and application delivery objects to speed delivery of applications across all types of networks and for all types of clients. Utilizing BIG-IP GTM with BIG-IP AAM improves performance by directing users to the nearest data center.
2. Wide area replication and file distribution is handled by BIG-IP AAM with symmetric adaptive compression, symmetric data deduplication, TCP optimizations, and secure point-to-point tunneling.

3. Bandwidth optimization is covered by BIG-IP AAM, which correctly sizes and reduces image files, caches frequently used objects, minimizes the data being sent from the application to the client, and applies protocol optimizations. Adding BIG-IP GTM to BIG-IP AAM directs traffic to the closest data center, allowing for regional customization of content and reduction of distance-based latency.

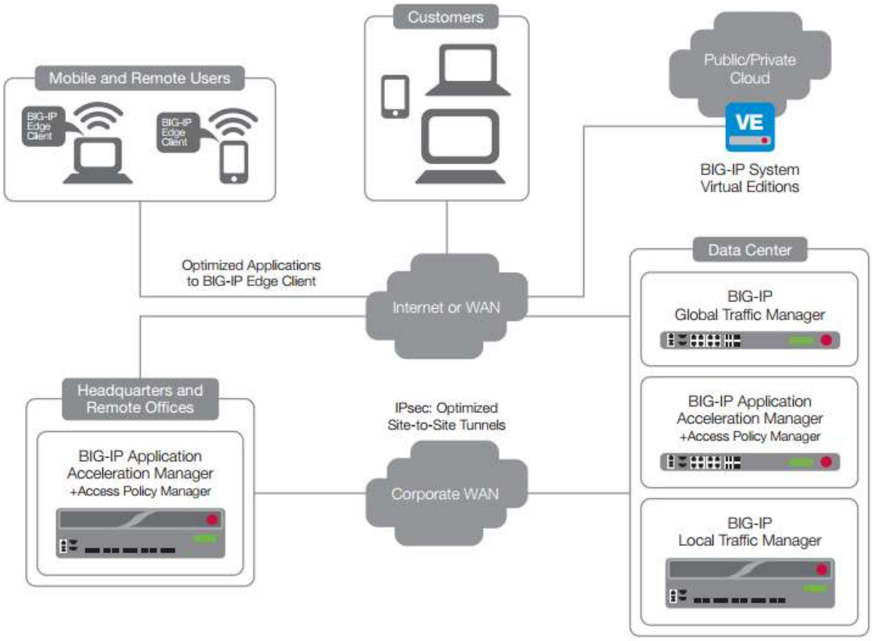

Mobile and Remote Users
BIG-IP Edge Client
BIG-IP Edge Client
Customers
Public/Private Cloud
VE
BIG-IP System Virtual Editions
Optimized Applications to BIG-IP Edge Client
Internet or WAN
Data Center
BIG-IP Global Traffic Manager
BIG-IP Application Acceleration Manager +Access Policy Manager
BIG-IP Local Traffic Manager
Headquarters and Remote Offices
BIG-IP Application Acceleration Manager +Access Policy Manager
IPsec: Optimized Site-to-Site Tunnels
Corporate WAN

Figure 1: A typical architecture for CDN replacement in a large enterprise makes global application delivery secure, fast, and available.

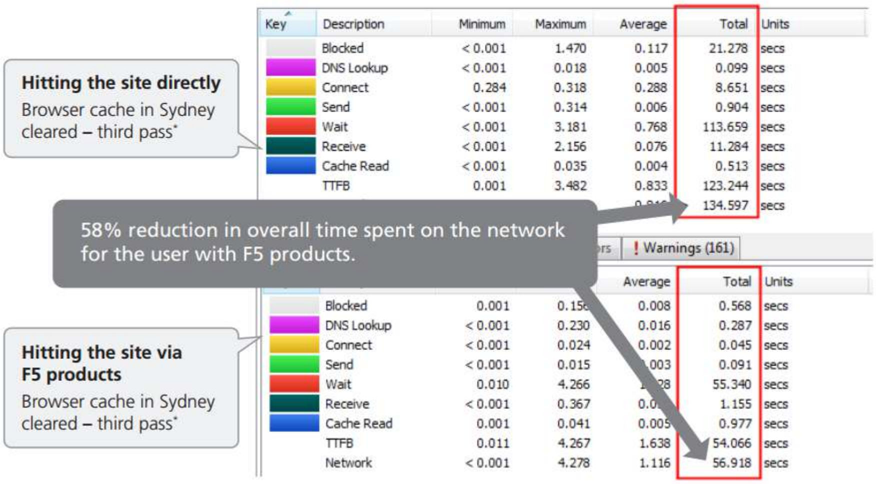
In a typical CDN replacement scenario, BIG-IP AAM and BIG-IP GTM can be deployed on the client side of the network to resolve application delivery optimization and bandwidth optimization issues, or BIG-IP AAM can be deployed between data centers to address wide area replication and file distribution issues. BIG-IP AAM can also optimize content for specific client attributes, such as performing more optimization for mobile clients with smaller screens, slower bandwidth, and less memory. With BIG-IP GTM directing traffic to the nearest data center, users experience more consistent application performance, whether they are in Shanghai or Los Angeles. In addition, when a data center is down for any reason, the connections that would normally be directed to that data center can easily be redirected to a different data center.

With BIG-IP AAM deployed on both sides of the data center–to–data center connection, high-performance replication and file distribution can take place at a fraction of the time it would take on a bare network. Additionally, if the organization utilizes one of several replication tools such as Oracle GoldenGate to keep databases in synch, BIG-IP AAM can provide an even greater level of performance over the Internet, making increased use of available bandwidth. When necessary, the optimized, secure tunnel created by BIG-IP AAM can also be used to send user requests to a remote data center.

BIG-IP AAM sits between application servers and clients, directing clients to the correct application server for their needs, optimizing the content being returned to the client, and managing connections between clients and servers. Additionally, if the application in question requires encryption, that encryption can be offloaded to BIG-IP AAM, keeping application server CPUs available for application processing.

The combination of BIG-IP AAM and BIG-IP GTM also offers load balancing at the local and global level. This load balancing, when combined with encryption offloading, improves the ability of the network to handle a spike in traffic and mitigates the impact of latency by reducing the content sent to each client and improving TCP performance.

BIG-IP Application Security Manager (ASM) and BIG-IP Access Policy Manager (APM) can be added to this solution set to integrate the functionality of centralized authentication, authorization, and accounting (AAA) and the unified security of a web application firewall.



Hitting the site directly
Browser cache in Sydney cleared – third pass*

Key	Description	Minimum	Maximum	Average	Total	Units
⬜	Blocked	< 0.001	1.470	0.117	21.278	secs
🟪	DNS Lookup	< 0.001	0.018	0.005	0.099	secs
🟨	Connect	0.284	0.318	0.288	8.651	secs
🟩	Send	< 0.001	0.314	0.006	0.904	secs
🟥	Wait	< 0.001	3.181	0.768	113.659	secs
⬛	Receive	< 0.001	2.156	0.076	11.284	secs
🟦	Cache Read	< 0.001	0.035	0.004	0.513	secs
	TTFB	0.001	3.482	0.833	123.244	secs
	[illegible]	[illegible]	[illegible]	[illegible]	134.597	secs

58% reduction in overall time spent on the network for the user with F5 products.

Hitting the site via F5 products
Browser cache in Sydney cleared – third pass*

				Average	Total	Units
⬜	Blocked	0.001	0.15[illegible]	0.008	0.568	secs
🟪	DNS Lookup	< 0.001	0.230	0.016	0.287	secs
🟨	Connect	< 0.001	0.024	0.002	0.045	secs
🟩	Send	< 0.001	0.015	[illegible].003	0.091	secs
🟥	Wait	0.010	4.266	[illegible]28	55.340	secs
⬛	Receive	< 0.001	0.367	0.0[illegible]	1.155	secs
🟦	Cache Read	0.001	0.041	0.005	0.977	secs
	TTFB	0.011	4.267	1.638	54.066	secs
	Network	< 0.001	4.278	1.116	56.918	secs

Figure 2: Sample benefits of BIG-IP AAM deployed symmetrically as a CDN replacement in a Sydney-to-London scenario.

Customers who already use F5 products have the advantage of familiarity with related policies and tools, including F5 iRules and F5 iApps Templates, which will already have been developed and put into place to support existing infrastructure. When additional F5 products are deployed to create a private, enterprise CDN, the same policies and iApps Templates can be extended to applications hosted in the CDN replacement, saving time and eliminating any need to implement separate management tools and policies for the applications residing on the CDN.

Conclusion

CDNs offer solutions to some of the most traffic-intense problems of a growing enterprise, and they offer a pay-as-you-grow model to accommodate the budgets of growing businesses. For many organizations, the monthly bills are acceptable based upon the benefits CDNs offer. For others, the overhead of a monthly bill drives interest in alternative solutions to the problems CDNs resolve.

For those organizations seeking to reduce or eliminate the monthly outflow of money, there are alternative solutions. F5 products can solve three of the most prevalent issues that drive CDN adoption, without the monthly fees. The architecture required to implement a CDN replacement with F5 products is relatively simple, and once deployed, these devices—because they sit within the data center at strategic points of control—can also benefit applications that are not currently part of a CDN deployment, without extra fees. Finally, implementing an enterprise CDN means that an organization's security certificates are retained within the walls of the corporation, not given out to a vendor.

Customers who already use F5 products can attain even greater benefits, applying existing policies, iRules, and iApps Templates to minimize complexity and streamline management of the CDN as a component of the organization's infrastructure, instead of as a separate, contracted service.

F5 Networks, Inc.
401 Elliott Avenue West, Seattle, WA 98119
888-882-4447 www.f5.com

Americas
info@f5.com

Asia-Pacific
apacinfo@f5.com

Europe/Middle-East/Africa
emeainfo@f5.com

Japan
f5j-info@f5.com