Think You Have An Application Delivery Strategy?

FEBRUARY 2009

The Role of Internet Application Delivery Services Within The Enterprise

By Jim Metzler, Cofounder
Webtorials Editorial/Analyst Division

Produced by Webtorials

Sponsored by Akamai

Think You Have an Application Delivery Strategy?

By Jim Metzler, Cofounder, Webtorials Editorial/Analyst Division

Introduction

Starting a few years ago, IT organizations began to focus on ensuring acceptable application delivery to their remote employees. The motivation for this new focus stemmed in large part from the fact that up until a few years ago the vast majority of employees worked in a headquarters facility. Today the vast majority of employees work in wide variety of non-headquarters facilities, including regional offices, branch offices, home offices, hotels, coffee shops and airport lounges. Unlike employees who work in a headquarters facility, remote employees typically access applications over a Wide Area Network (WAN). Unfortunately, the introduction of the WAN into the application delivery environment introduces a number of performance challenges.

The specific performance challenges that are introduced by accessing applications over the WAN depend on type of WAN service being utilized. For example, one type of WAN service is based on technologies such as private lines, Frame Relay, ATM and MPLS. This type of service is typically used to connect headquarters sites with each other and with branch and regional offices. Throughout this white paper, this class of WAN services will be referred to as private WAN services. When IT organizations began to be concerned with application delivery a few years ago, the majority of their focus was on one or two initiatives – either making some improvements within the data center or on improving the performance of applications that employees accessed over private WAN services.

While these two initiatives are important, they are only a part of the application delivery challenge as enterprise applications are evolving and the boundaries of the typical enterprise are beginning to blur. For example, until recently enterprise applications were accessed by just a single constituency - the employees of the enterprise. Today the typical enterprise has multiple constituencies, including employees, business partners, suppliers and customers. An example of this is a company whose supply chain management application extends beyond its employee base and reaches out to its customers, suppliers and distributors.

In virtually all cases today, at least some of the constituencies that an IT organization has to support don’t access enterprise applications by using private WAN services. They access enterprise applications by using the Internet. In addition, the boundaries of the typical enterprise will continue to become blurred due to a more diverse user community, as well as the adoption of new distributed application architectures (e.g., Web-enabled applications and business processes, SOA/Web Services, SaaS, and Cloud Computing) that often traverse multiple enterprises. As the boundaries of the enterprise become increasingly blurred, an increasing percentage of enterprise WAN traffic will be carried on the Internet.

As noted, over the last few years that IT organizations have focused on application delivery, the vast majority of that focus has been on either making some improvements within the data center or on improving the performance of applications that are delivered to branch office employees over private WAN services. What has been missing is a comprehensive approach to making application delivery over the Internet be of similar quality to application delivery over a private WAN service. Since it is usually not feasible to provide every Internet user with their own hardware and/or software, optimizing the delivery of applications that transit the Internet requires that application flows be optimized within the Internet itself. As will be shown in this white paper, the type of optimization techniques that get applied at layers 4 through 7 of the OSI stack are only effective if the underlying network is performing well. As will also be shown in this white paper, optimizing application traffic flows within the Internet can only be achieved by subscribing to an Application Delivery Service (ADS) offered by a managed service provider.

To provide additional insight into the challenges and opportunities relative to optimizing the performance of Internet traffic, two IT professionals were interviewed. One interviewee is the lead analyst for application technology for a Fortune 500 company in the food and beverage industry. This interviewee cannot be quoted by name or company and so will be referred to in this white paper as The Lead Analyst. The other interviewee is Dan Larsen, senior IT architect at Qualcomm.

The Limitations of the Internet

When comparing the Internet with private WAN services, the primary advantages of the private WAN services are better control over latency and packet loss, as well as better isolation of the enterprise traffic and of the enterprise internal network from security threats. As will be discussed in this section, the limitations of the Internet result in performance problems. These performance problems impact all applications, including bulk file transfer applications as well as delay sensitive applications such as Voice over IP (VoIP), video conferencing and telepresence.

The primary reason for the limitation of the Internet is that as pointed out by Wikipedia[1], the Internet "Is a 'network of networks' that consists of millions of private and public, academic, business, and government networks of local to global scope." In the case of the Internet, the only service providers that get paid to carry Internet traffic are the providers of the first and last mile services. All of the service providers that carry traffic between the first and last mile do so without compensation. One of the affects of this business model is that there tend to be availability and performance bottlenecks at the peering points. Another affect is that since there is not a single, end-to-end provider, service level agreements (SLAs) for the availability and performance of the Internet are not available.

As noted, the primary source of packet loss within the Internet occurs at the peering points. Packet loss also occurs when router ports become congested. In either case, when a packet is dropped, TCP-based applications (including most critical enterprise data applications) behave as good network citizens, reacting to a lost packet by reducing the offered load through halving the transmission window size and then following a slow start procedure of gradually increasing the window size in a linear fashion until the maximum window size is reached or another packet is dropped and the window is halved again.

With UDP-based applications, such as VoIP, Videoconferencing, and streaming video, there is no congestion control mechanism triggered by packet loss. As a result, the end systems continue to transmit at the same rate regardless of the number of lost packets. In the Internet, the enterprise subscriber has no control of the amount of UDP-based traffic flowing over links that are also

[1] http://en.wikipedia.org/wiki/Internet

carrying critical TCP application traffic. As a result, the enterprise subscriber cannot avoid circumstances where the aggregate traffic consumes excessive bandwidth which increases the latency and packet loss for TCP applications.

Another aspect of the Internet that can contribute to increased latency and packet loss is the use of the BGP routing protocol for routing traffic among Autonomous Domains (ADs). When choosing a route, BGP strives to minimize the number of hops between the origin and the destination networks. Unfortunately, BGP does not strive to choose a route with the optimal performance characteristics; i.e., the lowest delay or lowest packet loss. Given the dynamic nature of the Internet, a particular network link or peering point router can go through periods exhibiting severe delay and/or packet loss. As a result, the route that has the fewest hops is not necessarily the route that has the best performance.

Virtually all IT organizations have concerns regarding security intrusions via the Internet and hence have decided to protect enterprise private networks and data centers with firewalls and other devices that that can detect and isolate spurious traffic. At the application level, extra security is provided by securing application sessions and transactions using SSL authentication and encryption. Processing of SSL session traffic, however, is very compute-intensive and this has the affect of reducing the number of sessions that a given server can terminate. SSL processing can also add to the session latency even when appliances that can provide hardware-acceleration of SSL are deployed.

Increased latency and packet loss within the Internet cloud can exacerbate problems with protocols whose performance is highly sensitive to delay and loss, such as HTTP (Hypertext Transfer Protocol) and TCP. HTTP and other chatty protocols require hundreds or even thousands of round trips or applications turns to complete a single transaction. For example, if the round-trip delay over the WAN is 100 ms and a particular transaction over HTTP requires two hundred round trips then the user's application response time will exceed twenty seconds.

TCP has a number of other characteristics that can cause the protocol to perform poorly when run over a lossy, high latency network. One of these characteristics is TCP's retransmission timeout. This parameter controls how long the transmitting device waits for an acknowledgement from the receiving device before assuming that the packets were lost and need to be retransmitted. If this parameter is set too high, it introduces needless delay as the transmitting device sits idle waiting for the timeout to occur. Conversely, if the parameter is set too low, it can increase the congestion that was the likely cause of the timeout occurring.

Another important TCP parameter is the TCP slow start algorithm, mentioned earlier in this section. The slow start algorithm is part of the TCP congestion control strategy and it calls for the initial data transfer between two communicating devices to be severely constrained. The algorithm calls for the data transfer rate to increase linearly if there are no problems with the communications. When a packet is lost, however, the transmission rate is cut in half each time a packet loss is encountered.

The affect of packet loss on TCP has been widely analyzed [2]. Mathis, et.al., provide a simple formula that provides insight into the maximum TCP throughput on a single session when there is packet loss. That formula is:

Throughput <= (MSS/RTT)*(1 / sqrt{p})

where:

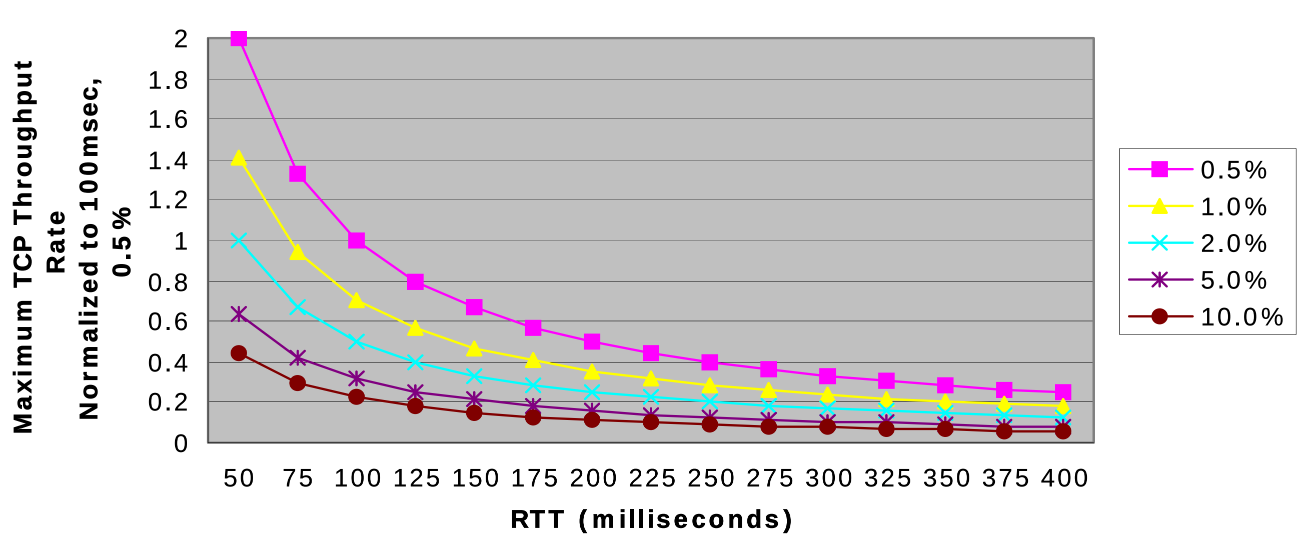
MSS: maximum segment size
RTT: round trip time
p: packet loss rate.

One obvious conclusion that can be drawn from the preceding equation is that the throughput decreases as either RTT or p increases. To exemplify the impact of packet loss and round trip time on TCP throughput, assume that MSS is 1,460 bytes, RTT is 100 ms., and p is 0.5%. In this case, the maximum TCP throughput is 1.65 Mbps independent of the size of the WAN link.

Figure 1 demonstrates the impact of delay and packet loss on TCP throughput given an MSS of 1,460 bytes. The data in Figure 1 is normalized relative to an RTT of 100 ms. and packet loss of 0.5%. These values for RTT and p will be referred to as The Normalized Parameters. The statement that the data in Figure 1 is normalized means that the maximum TCP throughput with The Normalized Parameters is 1.0. It also means that the maximum TCP throughput for other values of RTT and p are depicted in the graph relative to the maximum TCP throughput for The Normalized Parameters.

For example, if the packet loss increases from 0.5% to 1.0%, then the normalized TCP throughput drops to approximately 0.7. This means that the maximum TCP throughput is

Figure 1: Impact of Delay and Packet Loss on TCP Throughput



[2] The macroscopic behavior of the TCP congestion avoidance algorithm by Mathis, Semke, Mahdavi & Ott in Computer Communication Review, 27(3), July 1997

reduced by 30%. Since the maximum TCP throughput with The Normalized Parameters is 1.65 Mbps, this results in a maximum TCP throughput of 1.15 Mbps.

If the packet loss were to increase to 2.0%, the maximum TCP throughput is reduced by approximately 50%. Analogously, if the packet loss stays fixed at 0.5%, but the RTT increases to 200 ms. then the maximum TCP throughput is also reduced by approximately 50%. In both cases, the maximum TCP throughput is roughly 0.83 Mbps independent of the size of the WAN link.

An Application Delivery Strategy

The goal of an application delivery strategy is to ensure an acceptable quality of experience whenever and wherever an organization's employees, customers or partners access an application. This means that the company's applications must be highly available, exhibit suitable levels of performance, implement appropriate levels of security and be cost effective independent of the WAN service that is used to access the applications.

As noted, the vast majority of the attention that IT organizations have paid to application delivery has targeted one or two initiatives. Those initiatives are focused on either improving the performance of the data center or improving the performance of applications that are delivered to branch office employees over private WAN services.

Data Center Initiatives

The primary goal of these initiatives is to optimize the performance of the servers in the data center. One of the principal ways that IT organizations achieve that goal is by implementing devices that are typically referred to as an Application Delivery Controller (ADC). Unlike a WAN optimization controller (see subsequent paragraph), an ADC is an *asymmetric solution*. By asymmetric solution is meant that there is an ADC in the data center, but there is not a corresponding device at the user's premise.

In order to maximize server performance, ADCs implement a range of functionality, including:

- Server Load Balancing (SLB) to maximize the scalability and the availability of an application
- Layer 4 - Layer 7 switching to direct application queries to the most appropriate server
- Firewall functionality to ensure the integrity of the company's data and to provide application-specific security
- Off-loading computationally intensive tasks, such as the processing of SSL traffic
- Switch, accelerate and secure XML applications and web services

An ADC often contains functionality referred to as a *reverse cache*. One of the primary roles of a reverse cache is to store frequently accessed content. Since users retrieve this content from the reverse cache and not from the servers, the performance of the servers improves. However, this information still needs to transit the WAN in order to be delivered to the end user. As a result, the value of a reverse cache reflects the overall value of an ADC. That value being that an ADC alleviates some of the performance bottlenecks in the data center. However, because it is an asymmetric solution, an ADC cannot alleviate the performance bottlenecks that are associated with delivering applications over a WAN.

Private WAN Optimization Initiatives

The goal of these initiatives is to improve the performance of applications that are delivered to branch office employees over private WAN services. This goal is achieved by overcoming the characteristics of private WAN services that cause application performance to degrade. These characteristics include insufficient bandwidth, high latency, packet loss and network contention. For example, insufficient WAN bandwidth makes it a challenge to transmit large files, such as the files that are associated with database replication. High latency reduces the quality of delay sensitive applications such as voice and video.

In order to overcome the impact of these detrimental WAN characteristics, many IT organizations have implemented a product that is often referred to as a WAN optimization controller (WOC). WOCs are also referred to as *symmetric solutions* because they typically require an appliance in both the data center as well as the branch office.

In addition to appliance-based WOCs, some vendors have implemented solutions that call for an appliance in the data center, but instead of requiring an appliance in the branch office the solution only requires software. Typically this software resides on the computer of a particular employee and hence services just that employee. In some cases the software resides on an office computer and services multiple employees in the same office. This class of solution is often referred to as a software-only solution, or as a soft WOC. Since an IT organization can rarely control the software that gets deployed by its customers, partners and suppliers, the primary application of a soft WOC is to support the employees of a given organization.

The Lead Analyst said that his company has a low capacity MPLS network and that for cost reasons they are using the Internet for an increasing percentage of their WAN traffic. He added that they have deployed ADCs, primarily for offloading SSL processing from their servers. They are in the process of deploying WOCs in their restaurants to minimize the amount of traffic that traverses their MPLS network.

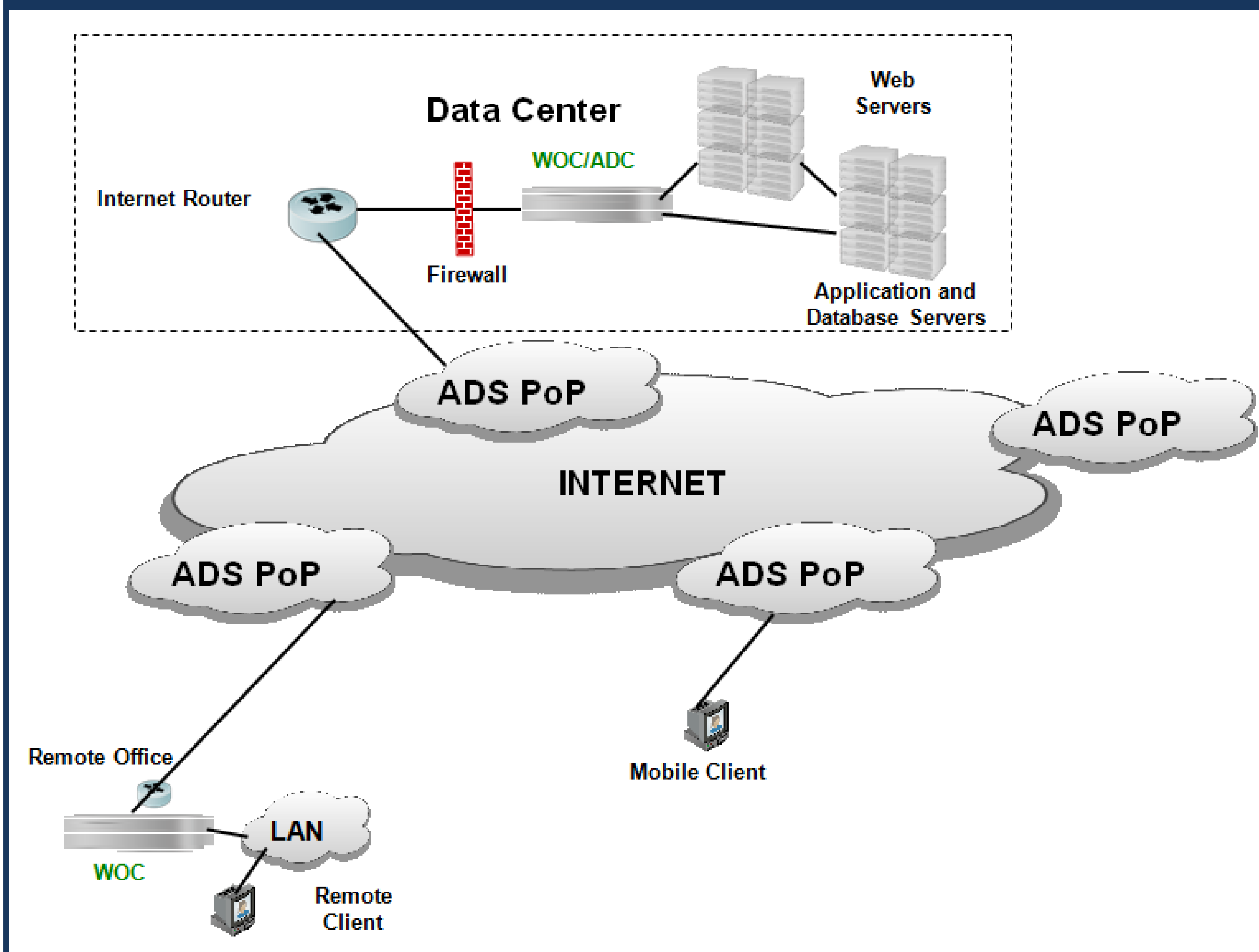
Larsen stated that his organization has deployed WOCs on their private WAN services and ADCs in their data centers. He highlighted the fact that because of the length of their WAN circuits (and hence the amount of delay and packet loss), that they were particularly pleased with the performance improvements that resulted from deploying WOCs.

However, for enterprises that use the Internet as a complement to private WAN services a complete application delivery strategy must address the performance limitations of the Internet that were described earlier. This can only be accomplished by subscribing to Application Delivery Services, such as those described in the next section of this white paper.

Internet-Based Application Delivery Services

As described in the preceding section, the traditional classes of application delivery solutions (ADC, WOC, soft WOC) were designed to address application performance issues at both the client and server endpoints. These solutions make the assumption that performance characteristics within the WAN itself are not optimizable because they are determined by the relatively static service parameters controlled by the WAN service provider. This assumption is reasonable in the case of private WAN services. However, this assumption does not apply to enterprise application traffic that transits the Internet because there are significant opportunities to optimize performance within the Internet itself based on Application Delivery Services (ADSs). An ADS leverages service provider resources that are distributed throughout the Internet in order to optimize the performance, security, reliability, and visibility of the enterprise's Internet traffic. As shown in Figure 2, all client requests to the application's origin server in the data center are redirected via DNS to an ADS server in a nearby point of presence (PoP) close to application users, typically within a single network hop. This edge server then optimizes the traffic flow to the ADS server closest to the data center's origin server.

Figure 2: The Internet Infrastructure for an ADS



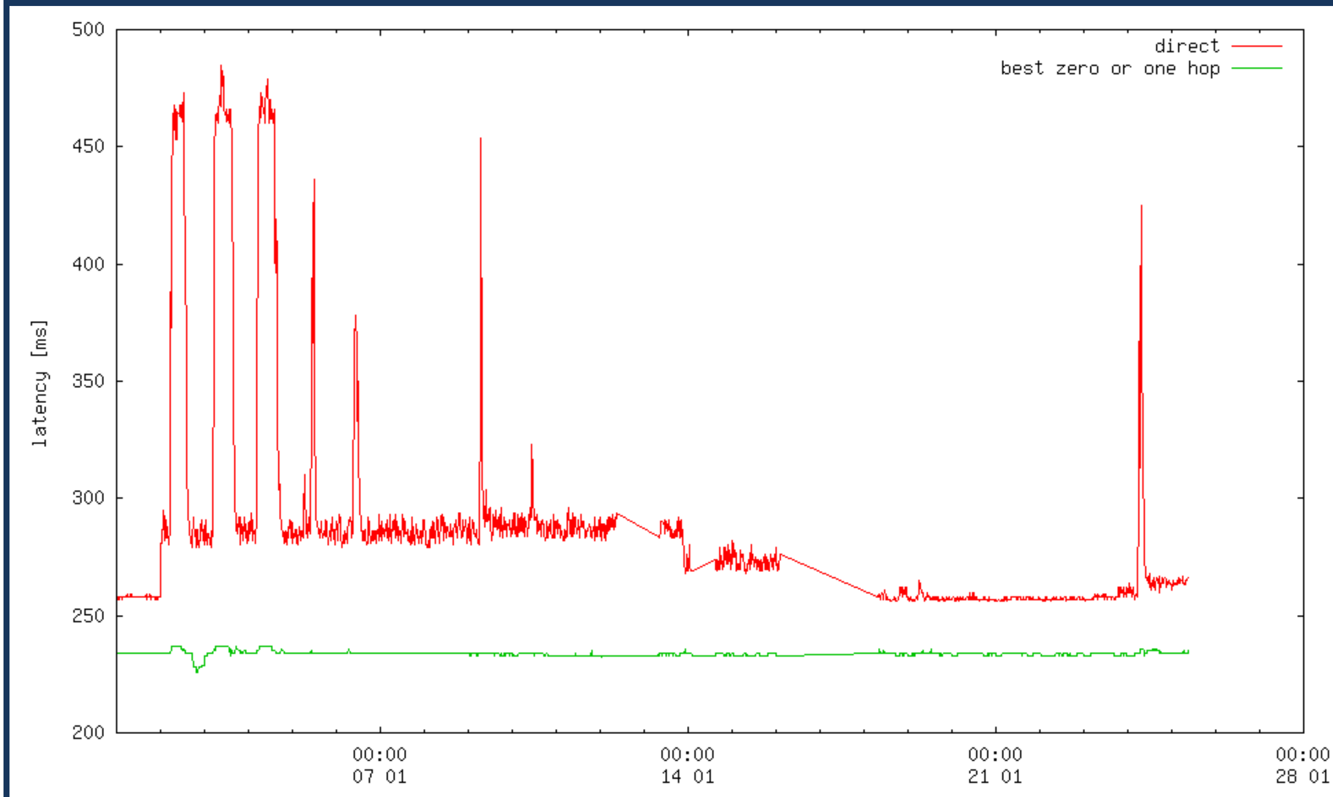
The servers at the ADS provider's PoPs perform a variety of optimization functions that generally complement the traditional application delivery solutions rather than overlap or compete with them. Some of the ADS functions include:

Route Optimization

Route optimization is a technique for circumventing the limitations of BGP by dynamically optimizing the round trip time between each end user and the application server. A route optimization solution leverages the intelligence of the ADS servers throughout the PoPs to measure the performance of multiple paths through the Internet and chooses the optimum path from origin to destination. The selected route factors in the degree of congestion, traffic load, and availability on each potential path to provide the lowest possible latency and packet loss for each user session.

As shown in Figure 3, the impact of route optimization can be dramatic. The data in Figure 3 depicts the round trip latency between Los Angeles, CA and Bangalore, India. The red graph reflects the round trip latency over the Internet and the green graph represents the round trip latency that results from using an ADS. Not only is the round trip latency significantly reduced by using an ADS, but the spikes in latency are removed.

Figure 3: Internet Round Trip Latency With and Without Dynamic Route Optimization



Transport Optimization

TCP performance can be optimized by setting retransmission timeout and slow start parameters dynamically based on the characteristics of the network such as the speed of the links and the distance between the transmitting and receiving devices. TCP optimization can be implemented either asymmetrically (typically by an application delivery controller) or symmetrically over a private WAN service between two WOCs, or within the Internet cloud by a pair of ADS servers in the ingress and egress PoPs. The edge ADS servers can also apply asymmetrical TCP optimization to the transport between the subscriber sites and the ADS PoPs. It should be noted that because of its ability to optimize based on real time network parameters, symmetrical optimization is considerably more effective than is asymmetrical optimization.

Another approach to transport optimization is to replace TCP with a higher performing transport protocol for the traffic flowing over the Internet between the ingress and egress ADS servers. By controlling both ends of the long-haul Internet connection with symmetric ADS servers, a high performance transport protocol can eliminate most of the inefficiencies associated with TCP, including the three-way handshake for connection setup and teardown, the slow start algorithm, and re-transmission timer issues. For subscriber traffic flowing between ADS servers, additional techniques are available to reduce packet loss, including forward error correction and packet replication.

There is a strong synergy between route optimization and transport optimization because either an optimized version of TCP or a higher performance transport protocols will operate more efficiently over route-optimized paths that exhibit lower latency and packet loss. Unfortunately, as shown in Table 1, conventional Internet routes often exhibit high levels of latency and packet loss[3]. What is even worse for some applications is that as shown in Table 1, the latency over the Internet can vary widely.

Table 1: Internet Performance with Chicago as the Origin

Destination	Median RTT	Maximum RTT	Peak Packet Loss
NYC, New York	22 ms	185 ms	28%
Dallas, TX	22 ms	65 ms	8%
Paris, France	94 ms	148 ms	14%
Beijing, China	280 ms	510 ms	54%
Seoul, South Korea	275 ms	580 ms	15%
Tel Aviv, Israel	193 ms	295 ms	16%

[3] The data in Table 1 was compiled over two weeks in March 2008.

HTTP Protocol Optimization

HTTP inefficiencies can be eliminated by techniques such as compression and caching at the edge ADS server with the cache performing intelligent pre-fetching from the origin. With pre-fetching, the ADS edge server parses HTML pages and brings dynamic content into the cache. When there is a cache hit on pre-fetched content, response time can be nearly instantaneous. With the caches located in nearby ADS PoPs, multiple users can leverage the same frequently accessed information.

Content Offload

Static content can be offloaded out of the data-center to caches in ADS servers and through persistent, replicated in-cloud storage facilities. Offloading content and storage to the Internet cloud reduces both server utilization and the bandwidth utilization of data center access links, significantly enhancing the scalability of the data center without requiring more servers, storage, and network bandwidth. The use of persistent connections also reduces SSL processing requirements by the server. ADS content offload complements ADC functionality to further enhance scalability of the data center.

Security

The ADS servers can also be used to move the outer limits of the enterprise security perimeter from the data center into the cloud. Security services in the cloud can provide firewall-like traffic screening with Level 3-7 intelligence for access control, filtering, and validity checking that can keep malicious traffic outside of the data-center. The extra layer of security can also isolate the data center from large scale DDoS attacks.

Availability

Dynamic route optimization technology can improve the effective availability of the Internet itself by ensuring that viable routes are found to circumvent outages, peering issues or congestion. For users accessing applications over the Internet, availability of the cloud is just as important as the availability of data center resources to maximize application uptime from an end-user perspective.

Visibility

Intelligence within the ADS servers can also be leveraged to provide extensive monitoring, configuration control and SLA monitoring of a subscriber's application with performance metrics, analysis, and alerts made visible to the subscriber via a Web portal.

The Lead Analyst said that his organization recognizes the need to have their application delivery strategy cover both their MPLS network as well as their use of the Internet. He added that one of the challenges that they face is running SAP over the Internet as SAP may require as many as 300 gets to load a single page. This challenge is exacerbated by the fact that the company headquarters is in the central portion of the US and they need to communicate between their headquarters and their restaurants, as well as with a variety of third parties around the world. The combination of the chatty nature of SAP combined with excessive WAN delay resulted in unacceptable application performance which has improved significantly by using an ADS. The

Lead Analyst pointed out that one of the features of the ADS that they use that is 'a big win' is route optimization. Other beneficial features include TCP optimization and pre-fetching.

Larsen stated that while Qualcomm communicates with some of their larger customers over private WAN services, that Qualcomm uses the Internet for the vast majority of their communications. Because of Qualcomm's desire to provide a consistent experience to their customers, combined with the performance issues that they faced when using the Internet for that communications, Qualcomm has adopted an ADS. Some of the ADS functionality that has been of the most benefit to Qualcomm includes:

- Route optimization

  By using route optimization through the Internet, Qualcomm has seen performance increase by at least a factor of 5, and by as much as a factor of twenty for sites such as Shanghai. In addition, when the Taiwan earthquake occurred in 2006 it severed Internet routes from Singapore and Taiwan. Because of route optimization, Qualcomm was not impacted.

- Edge Caching

  According to Larsen, people expect rich Web pages but that "If you have 75 images on your page, it can take a long time to render. Serving up the data locally makes a huge difference." He added that almost of the information on a Web page is static and can be compressed.

- Pre-Fetching

  Larsen stated that a lot of Web pages have between 50 and 80 sub elements and that the ADS that they use can "fire hose the elements across the Internet and then spoon feed them to the user's browser." Using pre-fetching the browser only has to go back to the origin for a small number of dynamic elements which according to Larsen "might me only one of the 50 elements on the page."

Summary

Enterprises are extending their business critical applications to an increasingly mobile user community comprised of employees, customers, and business partners. Many of these users must rely on the Internet for WAN connectivity to centralized enterprise applications and data resources. As enterprises increase their reliance on delivering high-value business applications across the Internet, they will need to develop an application delivery strategy that incorporates in-cloud Internet application optimizations as well as the more traditional WOC and ADC optimizations.

A comprehensive application delivery strategy must include Application Delivery Services such as the ones described in this white paper. The benefits of these services include complete transparency to both the application infrastructure and the end users - namely no new hardware, software or application redesign is required. This transparency ensures the compatibility of the ADS with complementary application acceleration technologies provided by WOCs or ADCs deployed in the data center or at remote sites. ADS services are available for optimizing all IP-based applications as well as Web applications and these ADS services provide the visibility into the Internet traffic that is comparable to the visibility most IT organizations have relative to the traffic that transits private WAN services. As a result, these ADS services transform the Internet into a business-grade WAN service that delivers the performance, security, scalability, and availability that is comparable to that of private WAN optimization services deployed over private WAN connections

About the Webtorials® Editorial/Analyst Division

The Webtorials® Editorial/Analyst Division, a joint venture of industry veterans Steven Taylor and Jim Metzler, is devoted to performing in-depth analysis and research in focused areas such as Metro Ethernet and MPLS, as well as in areas that cross the traditional functional boundaries of IT, such as Unified Communications and Application Delivery. The Editorial/Analyst Division's focus is on providing actionable insight through custom research with a forward looking viewpoint. Through reports that examine industry dynamics from both a demand and a supply perspective, the firm educates the marketplace both on emerging trends and the role that IT products, services and processes play in responding to those trends.

For more information and for additional Webtorials® Editorial/Analyst Division products, please contact Jim Metzler at jim@webtorials.com or Steven Taylor at taylor@webtorials.com.

Webtorials Briefs
Vol 3, Number 1

Published by Webtorials Editorial/Analyst Division
www.Webtorials.com

Division Cofounders:
Jim Metzler
jim@webtorials.com
Steven Taylor
taylor@webtorials.com

Transforming the Internet into a Business-Ready

Application Delivery Platform

Ensuring application performance supports your business goals

As organizations expand globally, they need to make a variety of business-critical applications – including extranet portals, sales order processing, supply chain management, product lifecycle management, customer relationship management, financial management, VPNs, and voice over IP – available to employees, business partners and customers throughout the world. These organizations must also be sensitive to the economic pressures driving IT consolidation and centralization initiatives.

Though global delivery of enterprise applications provides remote users with essential business capabilities, poor application performance can quickly turn the user experience into a costly, productivity-sapping exercise. Business applications must perform quickly, securely, and reliably at all times. If they don't, application use and adoption will suffer, threatening not just the benefits linked to the applications, but the overall success of the business itself.

Key Challenges in Delivering Applications

When delivering applications via the Internet to their global user communities, businesses face significant challenges, such as poor performance due to high latency, spotty application availability caused by unplanned internet disruptions, and inadequate application scalability to deal with growing user bases and spiky peak usage. Each of these problems severely undermines the application's effectiveness and the company's return on investment.

The performance issues associated with the Internet are not new. They are, however, having more of an impact because of business trends, such as globalization. Because of its lower cost, quick time to deploy, and expansive reach, IT organizations are increasingly turning to the Internet to support their globalization efforts. At the same time, chatty protocols like HTTP and XML, are introducing additional performance issues.

Akamai's Application Performance Solutions

Today, more than 2,700 businesses trust Akamai to distribute and accelerate their content, applications, and business processes. Akamai's Application Performance Solutions (APS) are a portfolio of fully managed services that are designed to accelerate performance and improve reliability of any application delivered over the Internet – with no significant IT infrastructure investment.

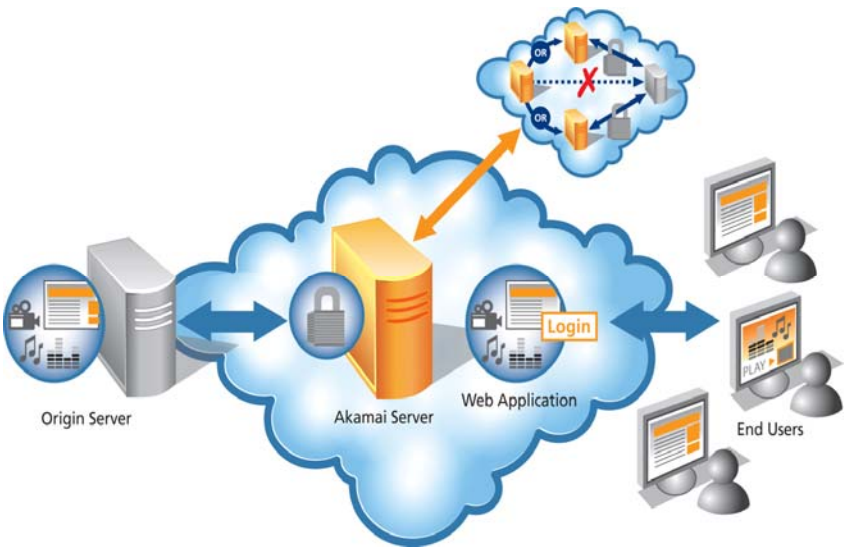
Akamai leverages a highly distributed global footprint of over 36,000 servers, ensuring that users are always in close proximity to the Akamai network. Application performance improvements are gained through several Akamai technologies, including SureRoute route optimization, which ensures that traffic is always sent over the fastest path; the Akamai Protocol, a high performance transport protocol that reduces the number of round trips over the optimized path; caching and compression techniques; and packet loss reduction to maintain high reliability and availability.

APS comprises two solutions – Web Application Accelerator and IP Application Accelerator. Web Application Accelerator accelerates dynamic, highly-interactive Web applications securely, resulting in greater adoption through improved performance, higher availability, and an enhanced user experience. It ensures consistent application performance, regardless of where users are located, and delivers capacity on demand, where and when it's needed.

IP Application Accelerator, like Web Application Accelerator, is built on an optimized architecture for delivering all classes of applications to the extended enterprise, ensuring increased application performance and availability for remote wireline and wireless users. Applications delivered by any protocol running over IP, such as SSL, IPSec, UDP and Citrix ICA will benefit from IP Application Accelerator.

Examples of applications delivered by APS include Web-based enterprise applications, Software as a Service (SaaS), Web services, client/server or virtualized versions of enterprise business processes, Voice over IP, live chat, productivity, and administration functions, such as secure file transfers. Akamai APS also addresses performance problems associated with the delivery of applications to wireless handheld devices, such as PDAs and smart phones.

How it works



Akamai Managed Services for Application Delivery

1. Akamai's dynamic mapping system directs user requests for secure application content to an optimal Akamai server.

2. Route optimization technology identifies the fastest and most reliable path back to the origin infrastructure to retrieve dynamic application content.

3. A high-performance transport protocol transparently optimizes communications between the Akamai server and the origin, improving performance and reliability.

4. The Akamai server retrieves the requested application content and returns it to the user over secure optimized connections.

Customer Benefits

Akamai's Application Performance Solutions offer a number of performance and business benefits:

• **Superior Application Performance –** Akamai provides unsurpassed application performance by accelerating both cacheable and dynamic content.

• **Superior Application Availability -** Akamai provides unique protections to ensure that poor Internet reliability never gets in the way of end user access to your application. Users are dynamically mapped to edge servers, in real-time, based on considerations including Internet conditions, server proximity, content availability and server load, ensuring that users are served successfully, with minimal latency.

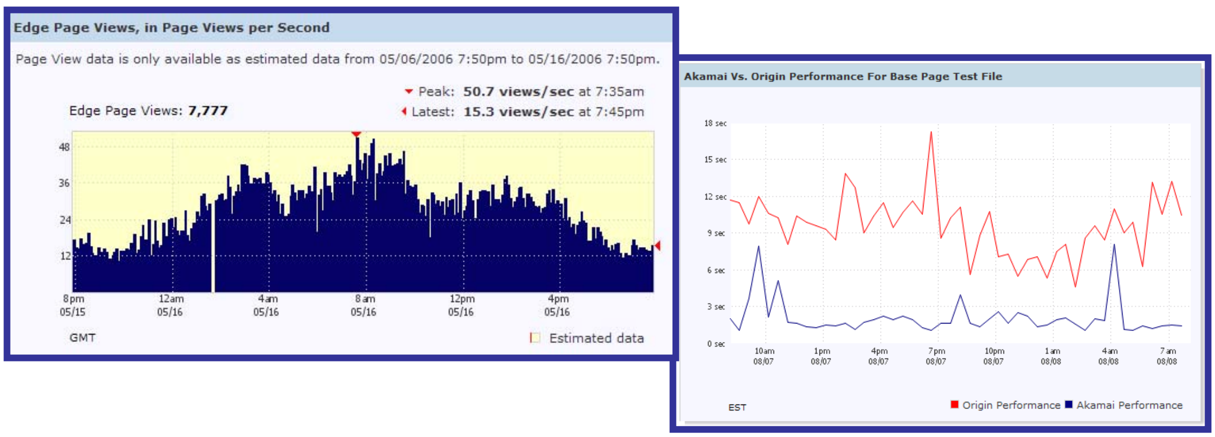
• **Rigorous Application Security –** Akamai allows you to eliminate the public entry points to corporate infrastructure by taking the initial hit at the Akamai edge, outside of your data-center.

• **Complete Visibility and Flexible Control -** Akamai's EdgeControl Management Center provides clear and effective tools that allow IT to manage and optimize their extended application infrastructure. In addition to sophisticated historical reporting and real-time monitoring functionality, Akamai also provides automated alert notifications when origin site problems are detected or user performance degrades. In addition, a sophisticated secure Network Operations Command Center continually monitors Akamai's global distributed network.

About Akamai

Akamai® provides market-leading managed services for powering rich media, dynamic transactions, and enterprise applications online. Having pioneered the content delivery market one decade ago, Akamai's services have been adopted by the world's most recognized brands across diverse industries. The alternative to centralized Web infrastructure, Akamai's global network of tens of thousands of distributed servers provides the scale, reliability, insight and performance for businesses to succeed online. Akamai has transformed the Internet into a more viable place to inform, entertain, interact, and collaborate. To experience The Akamai Difference, visit www.akamai.com.

For helpful insights and thought leading whitepapers on how Akamai Performance Solutions enable your key IT initiatives, visit www.akamai.com/aps.



The Akamai Portal provides valuable, real-time performance and usage information