

WAN Optimization:

How to rev up
sluggish applications



The WAN optimization market continues to expand, and enterprise IT buyers are the beneficiaries. Vendors competing for enterprise customers are adding all sorts of extras to their application acceleration and traffic optimization products, including built-in security features, performance management capabilities and add-ons for mobile users.



Silver Peak

Sponsored by Silver Peak

www.silver-peak.com

Table of Contents



Profile: Silver Peak	3
Introduction	5
WAN Optimization	5
News and trends	6
The next three big WAN optimization targets	6
Vista over the WAN: good but not great	10
Carriers pitch a WAN nirvana	12
Balance sought for SSL WAN optimization	14
Analysts' take on the market	16
What makes a WAN optimization controller?	16
Application acceleration: Divide and conquer	17
How do your WAN optimization efforts measure up?	18
Real-world tests and shopping tips for buyers	20
WAN acceleration offers huge payoff	20
How we tested application-acceleration devices	27
A buyer's checklist for application acceleration	28
Application acceleration: Making sense of a crowded technology market.....	30
Speed safely: application acceleration best practices	33
Lessons from the field	35
User plugs in an appliance, reduces WAN traffic by 60%	35
Gitmo gets high-bandwidth makeover	37
Transportation authority speeds up WAN with Silver Peak	39
The human side of net optimization	40
Learning from experience	41
Silver Peak White Paper	42

Profile: Silver Peak

How does Silver Peak define WAN optimization?

The WAN is undergoing a transformation as a result of many key enterprise initiatives. These include business continuity and disaster recovery (both of which drive the need for network backup and data replication), server and storage consolidation, and the move to MPLS and Internet VPNs. WAN optimization appliances incorporate a collection of techniques that transparently and securely optimize WAN traffic to mitigate the impact of limited bandwidth, latency and packet loss for any collection of IP-based applications. The more advanced techniques are 'data reduction', latency mitigation and error correction, which recovers from dropped packets in the network. Other techniques include compression and traffic management with advanced quality of service (QoS).

Unlike previous approaches like file caching (WAFS) and HTTP caching that are application-specific optimizations, WAN optimization is application-agnostic and does not alter client/server communications or configurations.

How does Silver Peak enable better application performance across a distributed enterprise?

WAN optimization has transitioned from a 'tactical' technology to a strategic element of distributed enterprise networks. This changes WAN optimization product requirements. When

an enterprise is planning a large-scale rollout across its entire network, functionality such as security, capacity and being able to accelerate all application types all become critical WAN optimization requirements. Architecturally, Silver Peak's appliances operate at the IP layer of the network stack so they can perform advanced data reduction across all application types, including realtime and interactive applications. This unique architecture also results in higher system capacity — Silver Peak appliances deliver up to 500Mbps of optimized WAN-side throughput across hundreds of thousands of flows. Silver Peak appliances also fully encrypt content on disk and over the network to ensure data privacy. Silver Peak's disk-based data reduction technology, called Network Memory™, uses up to seven terabytes of local storage to 'remember' the traffic that has traversed each WAN link. When duplicate information is detected, instructions are sent to the remote appliance allowing it to deliver the information from its local 'Network Memory'. If data is changed, only modifications traverse the WAN. Network Memory can eliminate 99 percent of WAN traffic, resulting in better application performance.

What are the core differentiators/unique attributes of Silver Peak's approach vs. others in the industry?

Silver Peak combines innovative new technologies, such as disk-based Network Memory, latency mitigation and loss mitigation,

Profile: Silver Peak (Cont'd.)

with traditional techniques, such as QOS, TCP acceleration and compression, to deliver a high-capacity solution that works across all enterprise applications. Differences that set Silver Peak apart include:

- **All applications** — Silver Peak performs all of its optimizations, including data reduction on all IP traffic. This provides benefits across ALL enterprise applications, including data replication, network backup and recovery, SQL, VOIP, SQL, Citrix, Remote Desktop, video and other UDP traffic, file, email and Web traffic. TCP-proxy based schemes that use slow session buffers are not compatible with most of these applications.
- **Network capacity** — Silver Peak NX appliances can process nearly 10 times more application sessions and bandwidth than alternative appliances. This makes Silver Peak ideally suited for high-capacity data centers often associated with enterprise-wide WAN optimization implementations.
- **Security** — Silver Peak's 'Secure Content Architecture'™ ensures all data, including data on local disk drives, is encrypted in hardware using advanced AES 128 bit encryption. Deep packet inspection enables fine-grained security policies, such as barring unwanted traffic from traversing the WAN.

How does Silver Peak address the particular management needs of the enterprise?

Silver Peak's Global Management System (GMS) provides a centralized, automated, point-and-click interface for managing, controlling and

configuring an entire network of NX appliances. For added network visibility, GMS provides centralized reporting and real-time monitoring of network conditions like packet loss, latency and bandwidth across each link. 'Netflow' records are exported to any external device that accepts and analyzes Netflow records for full application-level traffic visibility.

Which applications does your solution address and how do you see that list evolving over the next 18 months?

The typical enterprise with over 1,000 employees runs 100 applications across the WAN, and these applications are constantly evolving. No one can predict how these applications will change, which makes it virtually impossible to effectively keep application-specific optimization techniques up to date. By avoiding application-specific "tweaks", Silver Peak's Network Memory works consistently across all IP traffic types. This ensures that Silver Peak's products will continue to deliver industry-leading data reduction and performance for all enterprise applications, ranging from transactional applications like SQL to data replication and streaming video as these applications evolve. Another trend we see is that more and more application traffic is encrypted. As a result, Silver Peak developed a Secure Content Architecture that delivers a hardware and software foundation to accelerate even encrypted traffic at line rate performance. This is particularly important as security-conscious enterprises turn to WAN optimization as a key enabler of their business initiatives.

Introduction

WAN Optimization

WAN optimization technology is maturing and buyers are more comfortable than ever with tools that actively manipulate WAN data loads, but instead of resting on their laurels, vendors continue to roll out impressive new features and functions.

Application acceleration and traffic optimization products traditionally were designed to speed up TCP and other communications protocols, such as the Common Internet File System. The chattiness and inefficiency of these protocols took a toll on performance as enterprises started running applications designed for LAN environments over WAN links, so gear-makers responded with tools that expedite data handling using everything from compression to caching, load balancing and wide-area file services.

These methods can work to great effect. *Network World* tests show leading WAN acceleration devices can sharply reduce response time and bandwidth for Windows applications running across wide-area links, cutting WAN consumption by as much as 80 times and speeding file transfers by as much as 45 times.

Yet even as enterprise adoption increases, players in the WAN optimization market continue to innovate. For example, product sets today can include application-specific acceleration features for popular applications such as Microsoft SharePoint or SAP. Some vendors offer content delivery network features that pre-position content in a cache to speed the delivery of live and on-demand streaming media content. Others feature the ability to monitor and enforce application-specific service-level agreements (SLA).

Optimization products also are getting security add-ons, such as features for optimizing SSL traffic, tools to build encrypted tunnels between devices, and encrypted disks. For enterprises seeking an all-in-one branch-office appliance, some vendors are bundling in DHCP, DNS, Active Directory and print server features. Others are building routing features

into their WAN optimization equipment, such as policy-based routing and route selection.

Some vendors also are beginning to address the acceleration needs of mobile users. Blue Coat Systems, Expand Networks, Packeteer and Stampede Technologies are among vendors that offer client software designed to speed application delivery to remote PCs.

Zeus Kerravala, a research vice president at Yankee Group, likes the idea of client-side optimization. "For mobile employees, the big unknown is bandwidth. This client software will deliver performance without employees having to figure out the best way to work with their applications while on the road," he says.

Making a selection

With all these options to choose from, the trick for enterprises is finding the right fit.

Before choosing a vendor, it's important to understand the applications and services running on the network and the protocols they use, Gartner suggests in a recent report. "Also conduct a detailed analysis of your network traffic to identify specific problems — for example, excessive latency, bandwidth oversubscription or lack of prioritization for certain types of traffic," the research firm says. "Finally, insist on a real-life trial before committing to any purchase."

That's what Sound Transit in Seattle did — and it discovered Silver Peak Systems' gear worked so well that users were unhappy when the trial period ended.

Slow WAN links had hindered railway-improvement projects at Sound Transit, so Gary Nayar installed test boxes at headquarters and at a remote site that had the most complaints, then ran a test for two weeks.

"When we took it down the remote site

started to send complaints to their senior management saying we need that thing back on because we were getting a lot more work done," says Nayar, senior integration engineer for the transit authority.

WAN services revival

Another significant WAN trend that's boosting enterprises' performance options is carriers' adoption of technologies such as Transport Multi-protocol Label Switching and the new Provider Backbone Transport that make it easier for carriers to manage their IP networks. With easier management comes the ability to concentrate on speeding connections, improving customer service options and supporting higher-level emerging IP services.

"If carriers have a simpler, more flexible and more manageable network infrastructure, then enterprises get a better quality of service," says Steve Taylor, president of Distributed Networking Associates.

While converging their voice, data, video and wireless infrastructures, carriers hope to provide enterprises with the same level of service and quality of experience, no matter where they are located or what device they are using. So, if users are in an office in New York, at home in Vermont or on the road in Madrid, they can access the same applications, with the same security levels and functionality.

For enterprises, what's most exciting is that improving application performance over the WAN has become an industry goal of gear vendors and service providers alike. This Executive Guide to WAN Optimization reviews the latest trends and technologies designed to boost application performance, no matter where users log in.

WAN OPTIMIZATION

Section 1

News and trends

The next three big WAN optimization targets

Performance management, client-side acceleration and network security top priority lists

■ *By Denise Dubie*

WAN-optimization appliances from such companies as Packeteer, Riverbed Technology and Silver Peak Systems have been instant industry darlings, surging to the top of enterprise priority lists with their speedy, performance-improving results and quantifiable ROI. By combining such technologies as compression, caching, wide-area file services, TCP optimization and SSL acceleration, these devices have made doing business over the WAN not only tolerable but in many cases preferable.

Now vendors are hoping for follow-on hits with product improvements in client-side acceleration, network security and performance management.

Managing while optimizing

Performance-management technologies are becoming the add-on of choice for many optimization vendors looking to maximize the functions of their appliances. Coupling management capabilities and optimization lets network administrators use the reams of traffic and application data collected by acceleration tools to troubleshoot problems and identify areas in



With an appliance that combines WAN optimization and performance management, the network team at Rawlings Sporting Goods not only accelerates applications but also finds trouble spots, say Jack Matthews (l), and Richard Truex, network managers.

Sponsored by Silver Peak

www.silver-peak.com

Section 1: News and trends ●●●

need of a performance overhaul.

For instance, when the network team at Rawlings Sporting Goods in St. Louis needed to figure out why the performance of a Web-based customer-order application was suffering, it turned to a WAN-optimization appliance from Packeteer rather than rely on traditional network-monitoring software. Packeteer's PacketShaper used a variety of techniques to speed traffic across the WAN and provided insight into performance problems, say Jack Matthews, director of MIS, and Richard Truex, manager of network operations.

PacketShaper technology analyzed traffic, found bandwidth hogs and helped the Rawlings network team manage bandwidth consumption. With Packeteer the team determined what type of traffic could traverse the net, and created policies to throttle down application packets that weren't mission-critical in favor of those deemed essential to the business.

"Rather than buying network taps and sniffer probes, and having to take all that data and do the analysis ourselves, we bought PacketShaper," Truex says. Using WAN optimization "immediately puts you in a proactive mode, because once the issue is identified, you can go into the appliance and create a policy to prevent the problem from happening again," he adds.

Rawlings also uses PacketShaper to make sure it doesn't exceed its allocated share of its service provider's DS-3 line and to monitor QoS on VoIP calls. "PacketShaper is a management tool, a network analyzer tool and an optimization tool. It allows us to improve the quality of our network applications without spending more money than we had planned," Matthews says.

Industry watchers say companies should expect other vendors to provide systems management capabilities in concert with optimization tools. For instance, Citrix Systems acquired application-acceleration vendor NetScaler and Reflectent Software, a maker of client-side performance-management software. Combined, the products could provide performance improvements and real-time measurement, analysts say. Citrix already has integrated Reflectent with its Presentation

Server product and plans to continue working toward delivering optimized applications to user machines.

Examples of vendors planning to pin together management and optimization include Radware, through its acquisition of Covelight Systems. Covelight's Inflight product captures transactional data related to business events in real time, including user identity and session information. Radware says the merging of its technology with Covelight's could help network managers respond to performance and security issues more quickly. For instance, if a Web transaction deviated from normal in a way that indicated fraud was taking place, the network could be instructed to sever that transaction.

F5 Networks is planning to license Microsoft System Operation center in its F5 ControlPoint Management appliance. The F5 product, the two vendors say, will collect, analyze and report data from F5's portfolio of application optimization, availability and security devices using Microsoft management technology.

"Once you distribute the [WAN optimization] technology, there is no reason why you shouldn't be able to use it as probes

to report back to you in real time about end-to-end application performance," says Robert Whiteley, a senior analyst with Forrester Research.

Considering the client

Next up for companies whose appliances already sit in data centers and branch offices will be optimization software for installation on client machines. These include Blue Coat Systems, Expand Networks, Packeteer and Stampede Technologies. "We are going to see a lot of activity around client-based code in the second half of this year," says Joe Skorupa, a research director at Gartner. "Nomadic employees need the acceleration just as much as those in a remote office, so they need that code on their machines."

This client software would be a peer to a branch-office appliance and communicate with the data-center appliance serving as the central management device. A client machine in a branch office would rely on the appliance there to perform optimization tasks. If the client machine is in a home or at a hotel kiosk, the optimization software would kick in and apply the technologies needed to ramp



Silver Peak

The Leader in Scalable WAN Acceleration

100x Improvement in WAN Performance

"Best Current Product Offering" – Forrester

The Best Performance on Large Networks

- > High capacity WAN links
- > Many distributed offices
- > Diverse mix of applications

www.silver-peak.com

Sponsored by Silver Peak
www.silver-peak.com

Section 1: News and trends ●●●

up performance for the remote user.

In some cases, customers can choose to add the software only to certain client machines – those of the traveling sales staff, for example – and not use an optimization appliance at a remote sales office. This option could appeal to IT administrators who are overseeing small offices with a handful of staff and can't always justify the \$3,000 to \$5,000 for an appliance.

In other cases, the optimization software would not need to be preloaded on client machines. For instance, Blue Coat plans to offer acceleration software on demand, via a small code download, to remote users when they log on to the corporate network through a VPN. And Citrix announced a software-based acceleration client expected to be available this summer. The company claims it will deliver the first interoperable WAN-

optimization and SSL VPN product set with its Citrix Access Gateway integration.

Zeus Kerravala, a research vice president at Yankee Group, says he likes the idea of client-side optimization. "For mobile employees, the big unknown is bandwidth. This client software will deliver performance without employees having to figure out the best way to work with their applications while on the road," he says.

WAN optimization turf wars

Does WAN optimization belong in the network or on the edge?
Depends on whom you're asking **BY DENISE DUBIE**

For some industry watchers, the integration of WAN optimization and acceleration technologies into routers and switches is inevitable.

"The technology is all about improving the performance of applications for clients through manipulation of the network," says Zeus Kerravala, a research vice president at Yankee Group. "It will be a bit of time before it is fully integrated, but unless you are going to teach every software developer about how to make applications work on networks, you are going to have to embed the intelligence and capabilities into the network."

For others, the application-delivery technology isn't a logical extension of a packet-delivery network.

"It doesn't make sense for the router to swallow up WAN optimization," says Joe Skorupa, a research director at Gartner. "Networks and applications are still two very different skill sets. It's one thing to build a router with an integrated firewall, but it's an entirely different beast to build [application intelligence] into network gear."

And although many users like the idea of reducing the number of appliances, they must distribute to remote locations, they understand that eliminating appliances altogether isn't always feasible. Such is the case at Nanometrics, a semiconductor vendor in Milpitas, Calif.

"We have a lot of offices around the world, and I need to get our CRM application out to remote sales and service staff. When I did the math on how to support all these places, well, spending money on a DS-3 was just not cost effective," says David Kizer, IT director at Nanometrics.

Kizer supports 22 facilities using either a Cisco Wide-Area

Application Services appliance or a WAAS router module. For small offices, Kizer slips the WAAS module into existing routers. For larger offices or those that require more performance or scalability, he uses a WAAS appliance.

But ultimately, he says, he'd like to deliver applications worldwide without having to install hardware.

"I am on a mission to consolidate everything and support just the bare minimum at remote sites," Kizer says. "Having one device reduces support costs and simplifies troubleshooting, so the more features I can get on one box, the better."

Gartner's Skorupa points out that even though the WAAS module sits inside Cisco gear, it isn't integrated fully into the router. The technology shares the same physical location but remains a separate function, at least for the moment.

"This is not a router function, but more a services-overlay model. It may live inside the router's box, but that would be like saying that a firewall that lives in an Ethernet switch is an Ethernet switch," he says. "Maybe in 10 years there will be enough processing power within the router to do this, but for now, they are separate functions."

"I am on a mission to consolidate everything and support just the bare minimum at remote sites."

DAVID KIZER, IT director at Nanometrics.

Section 1: News and trends ●●●

Protecting secure packets

Looking ahead, security also will be a big focus of WAN-optimization vendors. As encrypted traffic increasingly becomes the norm, enterprises will need to apply optimization technologies to that traffic or the purpose of their WAN-optimization appliances will be lost. If the appliance can optimize only a small percentage of the traffic, companies won't continue to see the remarkable performance gains they have seen so far.

"A company today might only have 10% to 15% of its traffic encrypted, but in some industries, such as financial, it is as high as 60%, and overall that number is just going to keep growing," Forrester's Whiteley says. "It will be imperative for vendors to be able to optimize this traffic."

In the past, WAN-optimization vendors maintained that accelerating encrypted traffic was not feasible, because their tools would have to decrypt the packets to get at the content to determine whether compression was possible – thus defeating the purpose of encryption.

Some vendors have begun to figure out workarounds. Riverbed, for example, has upgraded the operating-system software used on its Steelhead appliances to be able to terminate SSL sessions, optimize the traffic, then re-encrypt it as SSL.

To avoid putting encrypted data at risk, SSL keys and certificates that reside within servers don't leave the data center, Riverbed says. Instead, the keys and certificates get transferred to a Steelhead appliance, which intercepts SSL requests headed for the servers. Then it establishes SSL sessions with another Steelhead appliance across the WAN using temporary session keys. These two devices talk to each other while the Steelhead device on the data center side talks to the servers, and the remote appliance talks to client machines trying to access the servers.

Besides Riverbed, Blue Coat (which has roots in security) and Certeon have begun to offer acceleration of SSL traffic across the WAN.

Keeping applications in mind

Although WAN-optimization vendors are broadening their product lines to include client operations, security and performance management, some also are ramping up optimization for one or more particular application types. By gaining deeper insight into these, vendors can learn – then apply – the best ways to improve performance over the WAN.

F5, for example, recently unveiled an application-delivery network specifically for Microsoft applications. This is part of the company's newly announced Application Ready Network program, intended to simplify the design, deployment and management of an integrated network from a single vendor. And Certeon, with its S-Series appliances, uses application blueprints to understand the user requests that will require specific application-data objects to pass over the WAN. It has S-Series Application Acceleration Blueprints for Oracle and SAP applications, as well as for the 2007 Microsoft Office, SharePoint and Exchange platforms.

"More vendors are looking to go beyond generic acceleration. They are being very specific in how they add optimization value," Whiteley says. "Customers with big investments in certain applications will find the targeted knowledge appealing."

Indeed, network managers interested in WAN optimization

"A company today might only have 10% to 15% of its traffic encrypted, but in some industries, such as financial, it is as high as 60%, and overall that number is just going to keep growing. It will be imperative for vendors to be able to optimize this traffic."

ROBERT WHITELEY, a senior analyst with Forrester Research

need to discuss application road maps with their prospective vendors, says Yankee Group's Kerravala. They need to ask vendors about the applications they plan on supporting, as well as whether they'll handle optimization at the client and on mobile devices, he says.

In addition, network managers shouldn't overlook optimization for homegrown applications.

Homegrown happiness

At Userplane, an AOL company that offers a communications platform for online communities, for example, WAN optimization of a custom application made a key function possible, says Michael Jones, founder and CEO of the Santa Monica, Calif., company. One of its mission-critical applications drops a small script on Web pages and calls back to Userplane servers every few seconds to report on visitors to the site. "It's a very low-transfer, high-volume application, transferring small bits of data every 5 seconds for millions of connections," Jones explains. "We could not take it live without having load-balancing technology in place."

That capability comes from Citrix's NetScaler load balancing and TCP offloading technology. The TCP offloading removed the overhead of establishing and closing connections to the Web server. NetScaler made more Web servers unnecessary by optimizing the application traffic coming into the data center, and made it possible to roll out the application in line with business plans.

Today's WAN optimization tools won't be around for long in their current incarnations. The products that worked their way into enterprise networks promising to speed applications will deliver those capabilities and then some. Enterprise network managers can be certain their investments in WAN optimization today will pay off in more advanced capabilities tomorrow.

"Customers can go to these vendors now and get a clear picture of what their products will look like in a couple of years, and really make the case for getting the features they want in their WAN optimization gear," Yankee Group's Kerravala says. "It's definitely a buyer's market."

Section 1: News and trends ●●●

Vista over the WAN: good but not great

Here's the skinny on how Microsoft's new client operating system handles WAN optimization

■ By Ann Bednarz

Windows Vista has users buzzing about network performance improvements, but industry watchers caution that the new client operating system is no WAN-optimization cure-all.

In certain scenarios, the improvements implemented in Vista and slated for Windows Server 2008 (Longhorn) will alleviate the problems of latency and inefficient protocols that have hampered application performance over the WAN. But the improvements will have limited or no effect on performance until PC hardware is upgraded, applications are reworked and Microsoft Vista or Longhorn machines sit at both ends of a WAN link.

From a communications perspective, Longhorn and Vista are works in progress, says Joe Skorupa, a Gartner research vice president. "There's a lot to like, but don't dash to deploy Vista just to get these features. As well tested as it has been, you'll want to make sure you take a measured approach to deployment."

Although the vast majority of enterprises plan to adopt Vista, most are in waiting mode, according to a 2007 survey conducted by Walker Information in collaboration with CDW. Among 753 IT buyers, 1% have deployed Vista, 25% are implementing it and 61% plan to do so within the next 12 months or more.

Respondents acknowledged Vista's potential performance improvement; 56% cited that as a key perceived benefit.

Protocol improvements

For the most part, Vista's performance gains are directed at file-sharing over WANs, which has become an important New Data Center issue for enterprises as the number

of employees working in remote offices climbs. Most significant, Vista features a rewritten TCP/IP stack and improvements to Common Internet File System (CIFS), the native Windows access protocol that enables file- and print-sharing among devices.

For the new TCP/IP stack, Microsoft changed its congestion-control algorithms so more data is sent at higher speeds. "It does some autotuning things, and it takes advantage of large window sizes – stuff the research has been pointing to for a while that will make TCP run a lot better in high-bandwidth, high-latency environments," Skorupa says.

Microsoft based its Vista implementation of CIFS on a new version of the Server Message Block (SMB) protocol that lets multiple data blocks be sent over the WAN simultaneously instead of requiring blocks be sent individually and waiting for acknowledgment that each has been received.

These CIFS and TCP improvements will have the most positive impact on long file-transfers, such as large media files that have been compressed as much as they can be, says Eric Siegel, a senior analyst at Burton Group. But don't count on Vista to do data-reduction-based compression as dedicated WAN-optimization appliances can, he says. Nor will Vista's performance enhancements have much effect on Web-based applications, Skorupa adds.

"They're not going to take a badly structured, browser-based app that does 70 round trips on the network to paint a screen down to five round trips," he says. "Vista also doesn't do anything to provide optimization specifically for HTTPS-encrypted traffic."

Another consideration is that companies won't benefit fully from the TCP and CIFS enhancements unless they have Longhorn or Vista machines on both ends of a link.

Further complicating the CIFS issue is that a number of third-party and home-grown applications, because they were developed specifically for the LAN, never took advantage of some of the advanced features in SMB 1.0. Getting these applications to take advantage of the improvements in SMB 2.0, on which the Vista CIFS implementation is based, is not an attractive prospect for many IT departments, Skorupa says. Deploying WAN-optimization gear is much quicker and far less painful than restructuring an application, he says.

Not fully baked yet

Another performance feature Microsoft has focused on in Vista is QoS, incorporating management tools that let administrators link application policies to user profiles via Active Directory. Using Active Directory information about employees' and groups' access rights as the base, administrators potentially could make decisions as fine-grained as giving a particular finance employee high-priority access to SAP applications during the last two weeks of a quarter.

"If you can do all that through Active Directory and put it out through group policy, and lock it down so that the person on the end station can't change it ... that becomes a really powerful system to have," Skorupa says. To date, he adds, the policy-based QoS features in Vista don't allow that level of detail. "You can't say, 'Do this for this application running over Port 80.' It's not enough to have a huge impact yet. But it's a start."

On the whole, Microsoft has made great strides to improve application performance over the WAN, but Vista's enhancements are not going to wipe out the market for specialized WAN-optimization gear any time soon. "There's a lot of cleanup here, and it

Section 1: News and trends ●●●

took Microsoft a long time to do it. But that said, it's good stuff," Skorupa says.

For enterprises, taking advantage of all the good stuff will require hardware upgrades. Most existing desktop machines lack the processing power, memory and graphics capabilities to take advantage of Vista, Skorupa says. "This is not a no-cost upgrade."

Significant software testing is required, too. "Enterprises have hundreds, and in some cases thousands, of in-house applications that they have to test," Skorupa says.

That's a serious consideration for Hyundai Information Services North America. David Jung, who is technical lead for infrastructure engineering security at the automaker's Irvine, Calif., IT unit, is running Vista on his work PC, but says Hyundai has no immediate plans to implement Vista companywide. The appeal of performance enhancements is outweighed by the daunting reality of how a client operating-system upgrade could break dozens of existing applications, Jung says.

"One of the big issues is that we need to verify all the applications will keep working on Vista. It will take some time," he says. In addition, he says, the CPU and memory required to run Vista is a factor for waiting.

Vista's performance-boosting trio

With these three targets, Microsoft aims to boost application performance over the WAN:

Updated TCP/IP stack

IMPACT: Enhancements include autotuning of the receive window's size to optimize throughput on long-distance links, a new Compound TCP (CTCP) algorithm for moderating TCP flow control from the transmitting end, and the ability to detect default-router failures and improperly configured maximum-transmission unit settings.

LIMITATIONS: CTCP is best used for high-speed Longhorn server-to-server data replications.

New Common Internet File System (CIFS) implementation

IMPACT: Support for Microsoft's Server Message Block (SMB) 2.0 protocol in computers running Vista or Longhorn enables the sending of multiple commands within a packet, reducing the number of round trips between a client and server. SMB 2.0 supports larger buffer sizes than SMB 1.0 does, and can withstand short interruptions in network availability.

LIMITATIONS: SMB 2.0 will improve performance when file servers are accessed across a WAN, but only between Vista clients and Longhorn servers.

QoS enhancements

IMPACT: Tools to set application- and transaction-level QoS via Active Directory policies could simplify configuration and management for users.

LIMITATIONS: QoS controls aren't detailed enough yet, and asymmetrical QoS requires a homogenous Vista and Longhorn environment.

SOURCE: GARTNER

Section 1: News and trends ●●●

Carriers pitch a WAN nirvana

Next-generation transport technologies could lead to cool new IP services

■ By Joanne Cummings

The public network is getting interesting again. In the early '90s, the buzz was about the Internet's rise and the advent of frame relay. Today, it's about PBT, T-MPLS and next-generation networks, such as Verizon's FiOS and BT's 21CN.

Wade through all the acronyms, and it becomes clear that carriers and network equipment providers are focusing on injecting more intelligence into public networks. That is making them more reliable and responsive, as well as capable of handling a variety of business-critical applications. On tap are managed services for security, storage, collaboration, disaster recovery and other business essentials.

It comes down to this: "If carriers have a simpler, more flexible and more manageable network infrastructure, then enterprises get a better quality of service because it is deliverable," says Steve Taylor, president of Distributed Networking Associates. "And that goes all the way from the hardware, up through things you would never think of, like billing systems. The entire infrastructure of the carrier is being overhauled and simplified, and that bodes well for the enterprise."

Debating MPLS vs. PBT

This carrier overhaul is a direct result of the recent focus on IP. That WAN technology is far more unwieldy than older ones, such as ATM or frame relay, which were more point-to-point in nature and easier to manage at the carrier level. "If you don't know where the traffic is, you can't guarantee it, and therein lies the problem with IP," says Thomas Nolle, president of technology assessment firm CIMI. "The adaptive behavior of IP tends to destroy control over just where the traffic is going in the network."

This uncertainty means carriers have been struggling to provide the requisite QoS and

uptime assurances for a range of emerging IP business applications. In a frame relay environment, carriers may have been able to guarantee five-nines of uptime per 24 hours, but they cannot get close to that level for today's VPNs. Sometimes those service-level agreements (SLA) guarantee five-nines over a month instead of 24 hours. "An IP service just doesn't have the meticulously orderly behavior that a frame relay service does," Nolle says.

That's why the carriers are pursuing more deterministic IP-transport mechanisms, such as Transport Multi-protocol Label Switching (TMPLS) and the new Provider Backbone Transport (PBT), also called Provider Backbone Bridging-Traffic Engineering (PBB-TE).

Both technologies support multiprotocol packet transport and scale globally, key requirements for next-generation networks, experts say.

The question of which technology – TMPLS or PBT – is better is akin to the Betamax vs. VHS debate, which started almost 30 years ago. Some say TMPLS wins because of its superior traffic-engineering capabilities; others – most notably Nortel – hold that PBT's Ethernet-like behavior and lower costs make it the technology of choice.

For now, carriers and equipment makers are hedging their bets. BT, for example, has been fairly public in its support for Nortel's PBT gear, announcing it is purchasing the vendor's 8600 Metro Ethernet Switch and 1850



Network switches should be more highly intelligent and application-aware, says Tony Bishop, vice president, Wachovia Bank. But for now he cares less about what the carriers are doing in that regard and more about how he can boost application performance across the bank's own network.

BRIAN GOMSAK

Sponsored by Silver Peak
www.silver-peak.com

Section 1: News and trends ●●●

Metro Ethernet Services Unit for use in its next-generation 21CN (21st-Century Network) program. However, the carrier also is highly supportive of MPLS as a managed-services-enabler in the WAN.

"We're looking at PBT, but there's nothing that we see in there that we cannot deliver through our MPLS network today," says Stefan Van Overtveldt, vice president of IT transformation at BT.

The TMPLS-or-PBT decision clearly means more to the carriers than their customers. "The argument behind PBT is it gives carriers a lower cost of ownership and management, so in theory, they can pass along that savings to customers. Obviously, there's an advantage to that, but enterprises can get the same services no matter which technology is used," says Johna Till Johnson, president of Nemertes Research.

David Creamer, director of network systems and TV operations at St. Petersburg College in Florida, agrees. "I'm out of the loop on whether they do it with MPLS or they do it with whatever. As long as they give me the service and SLAs I'm looking for, I don't care," he says.

Future applications

The commonality between these competing technologies is that they make it easier for carriers to manage their next-generation networks. And with easier management comes the ability to concentrate on supporting higher-level emerging IP services. While converging their voice, data, video and wireless infrastructures, carriers hope to provide enterprise users with the same level of service and quality of experience, no matter where they are located or what device they are using. So, if users are in an office in New York, at home in Vermont or on the road in Madrid, they can access their same applications, with the same security levels and functionality.

"Service providers will be able to personalize and manage individual sessions and applications differently, while at the same time guaranteeing SLAs to the enterprise," says Arpit Joshipura, a vice president at Redback Networks, which provides multiprotocol routers for several next-generation carrier networks.

He uses the example of an ad hoc videoconference: "A year from now, when these networks migrate, we will be able to videoconference through our PCs with built-in cameras and microphones via a secure IPVPN on the fly. The carriers will give us a secure videoconference with high-definition-equivalent

bandwidth on the click of a button. That's what things are moving toward."

Another example is BT's Corporate Fusion offering, available in Europe now and in the United States by yearend. Companies with the service issue each employee a dual-mode cell or VoIP phone that works over regular cell and Wi-Fi networks. Corporate Fusion determines which type of service is available and the least expensive at the time the employee makes a call.

"If you're home, you make the call over your broadband DSL or cable modem line, and if you're in the office, you call over the Wi-Fi that's connected into your PBX that then goes out into the public telephone system," BT's Van Overtveldt says. "It's a much cheaper way of communicating, and it comes with the additional plus of only one contact list, so there's no need to look up a number on a cell phone and dial it from a regular phone, etc."

Future carrier networks supposedly also will handle many applications that today are managed at the enterprise level, such as security, single contact and calendar, storage and disaster recovery.

Some services are done best at the carrier level, St. Petersburg's Creamer says. "We use a managed security service from our ISP for handling [distributed denial-of-service] attacks," he says. "It's not real complicated, but ... if we stop attacks at the carrier's core, then we're not dealing with the issue. That fits very well for any enterprise."

Nemertes' Johnson agrees that security services play well in carrier networks. "Managed security services for things like DDoS, [viruses and spam] make sense, and another example is storage applications," she says. "I don't see any good way you can get very high-speed, very distributed storage without involving the network, so the carriers have a play there, too. I can see enterprises migrating that way."

Next-generation carrier networks likely also will handle disaster recovery and server virtualization services, Creamer says. "For example, our grant programs are temporary in nature... It's important to do the work or whatever the grant entails, but you don't want to hire a bunch of IT people or bring a lot of infrastructure in. In a situation like that, a carrier server virtualization service that was personalized, secure and could be brought up and taken down quickly would be advantageous."

Show me

Just because carriers will be able to handle

cool new services on their next-generation networks doesn't mean enterprises will let them.

"Say I buy these managed services, and our students and customers become dependent on them, and they become an important part of my business," Creamer says. "If I don't have the infrastructure in place or the staff to support them in-house, then the carriers' pricing models are bound to change. I'll be stuck. I'll have to pay what they charge."

Experts agree and say that for this new model to work, competition has to be robust. "We're almost in some ways back to 1984 again, in terms of the number of major players in the carrier space," Distributed Networking Associates' Taylor says. "I think we have to be very careful that we make sure, as enterprises, we continue to push for a competitive environment."

Others say they are concentrating on their own internal infrastructures, making their enterprise networks more intelligent, and taking more of a wait-and-see attitude toward carrier services.

"We're interested in Cisco's [Application Oriented Networking] initiative and making our internal switches more intelligent and application aware," says Tony Bishop, a vice president at Wachovia Bank. "But a number of regulatory and security concerns will have to be addressed before we move more of our [processing control] into the public cloud."

More public than private

Experts say, however, that as these next-generation carrier networks mature, enterprise networks of the future increasingly will rely on key services provided by the carriers and their public infrastructures.

"It's really a continuum of what functions are delivered by the carrier network and what functions are delivered at the enterprise - where is that demarcation point?" Taylor says. "We started building private data networks 20 years ago, not because it was absolutely the best idea in the world but because it was the only way you could get the services you needed. The service providers were dragging their feet too much."

"Now, with these next-gen networks, the service providers should be able to lead the way," he says. "The opportunity is theirs to lose."

Cummings is a freelance writer in North Andover, Mass. She can be reached at jocummings@comcast.net.

Section 1: News and trends ●●●

Balance sought for SSL WAN optimization

Businesses need to balance WAN performance with data security

■ By Tim Greene

With more and more WAN optimization vendors extending their capabilities to include encrypted traffic, corporate IT executives have a decision to make: Should they trust the security these devices provide?

Rather than passing through SSL sessions between clients and servers located in remote data centers, some WAN optimization gear can terminate the SSL sessions, shrink the traffic and re-encrypt it for the next leg of the trip. These chains of encrypted sessions introduce potential vulnerabilities that different vendors address in different ways.

SSL traffic represents a growing percentage of total traffic on WAN links, according to Forrester Research. So SSL support in WAN optimization appliances will become more important to businesses that want to keep traffic secure while minimizing the size of their WAN links.

In a recent survey of 1,300 IT executives by WAN-optimization vendor Blue Coat Systems, one-third of respondents said that 25% of their WAN traffic is SSL. And of those surveyed, 45% plan to roll out more SSL applications this year.

About a third of all WAN traffic at Richardson Partners Financial Ltd. in Toronto is SSL, says Andrew McKinney, director of technical services for the firm. But if only the urgent business traffic is considered, the percentage is much higher. "For critical business traffic, it's all encrypted," he says. So he uses Blue Coat Systems gear to secure traffic and optimize it for good performance.

But first he got the devices in and grilled the vendor about the security at each point of the proxy chain until he was satisfied it would keep the firm's data safe. "Our big concern was that we would have control of what was being

cached," he says. He didn't want sensitive data left on the disk of the Blue Coat appliance.

"We wanted to be sure the data could be flushed as we required but also that it was securely being stored. In the end we were satisfied," McKinney says.

So far, just three vendors - Blue Coat, Certeon and Riverbed Technology - offer SSL acceleration on their appliances, and the capability is also on the road map for Juniper Networks and Silver Peak.

Such devices sit at both ends of WAN links and perform a number of functions that serve to speed up transaction times. These include optimizing TCP sessions, enforcing QoS, byte-level pattern matching and protocol optimization.

Without SSL support, when SSL traffic hits these boxes they are limited to using TCP optimization and QoS.

SSL support relies on terminating SSL sessions, decrypting the traffic storing segments of the data for future reference and re-encrypting. Later traffic through the devices is compared with these segments. When data being sent matches a segment, the devices send an abbreviated reference rather than the longer complete segment, thereby reducing the amount of traffic that has to cross the wire.

These segments are analogous to paperwork put through a shredder, says Mark Day, chief scientist at Riverbed. "The individual pieces are still intelligible, but total documents are somewhere between difficult and impossible to reconstitute," he says. Customers concerned that these data segments represent a threat can turn off SSL optimization and pass the traffic through, he says.

Blue Coat doesn't encrypt the stored segments either, arguing that the difficulty



Silver Peak

The Leader in Scalable WAN Acceleration

100x Improvement in WAN Performance

"Best Current Product Offering" – Forrester

The Best Performance on Large Networks

- > High capacity WAN links
- > Many distributed offices
- > Diverse mix of applications

www.silver-peak.com

Sponsored by Silver Peak

www.silver-peak.com

Section 1: News and trends ●●●

of using the data makes it practically inaccessible. "It's not encrypted but you're going to need NSA-level talent to get that information out of the disk," says Chris King, director of strategic marketing for Blue Coat.

Certeon does encrypt on the disk, storing the keys needed to decrypt it on the appliance that sits across from it on the WAN. That way if one box is stolen, the data on the disk is secure.

In any SSL proxying architecture, the server must give up its certificates to at least one other device. In the cases of Blue Coat, Certeon and Riverbed, that device is a WAN optimizer that sits inside the data center where it is as physically secure as the

server itself. How those certificates are handled varies from vendor to vendor.

Blue Coat creates a session key for its client-side device using the public key of the server that is held by the serverside Blue Coat device. The serverside device uses the private key of the server to decrypt the session key that is used over the WAN.

Riverbed also generates a session key for the remote SSL session and migrates it to its client-side machine in two steps. Certeon proxies between the server and its serverside appliance, then transmits traffic across the WAN via an IPsec connection. The client side Certeon

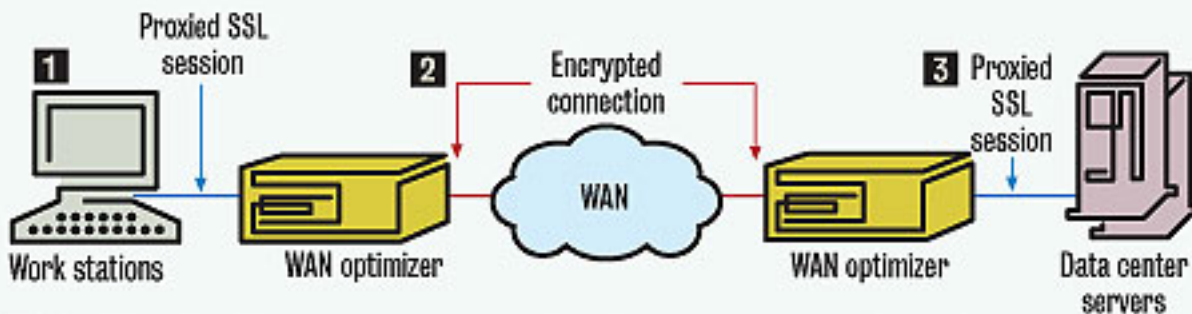
box creates a separate session to the client machine.

The fact that SSL sessions are being proxied shouldn't scare anyone away, says Joe Skorupa, an analyst with Gartner. Akamai, for instance, proxies SSL transactions – some of them involving e-commerce – for its customers, he says.

"There's precedent for doing it and doing it where significant amounts of money are at risk," Skorupa says, but each potential user has to balance the need for speed and the need for security. "Customers will have different levels of sensitivity about how it's being done."

Speeding SSL traffic

Blue Coat Networks, Certeon and Riverbed offer WAN-optimization gear that can look inside SSL traffic, reduce it so it takes up less space on the WAN, then re-encrypt it to cross the wire. Here is how the gear works and some of its differences.



- 1** WAN-optimizing gear appears to the client machine as if it is the server and establishes an SSL connection with the client. It terminates the SSL session, optimizes the traffic, then re-encrypts it. Part of optimizing calls for storing data segments on the disk of the appliance. Only Certeon encrypts the data on the disk.
- 2** The WAN optimizers set up a secure connection between themselves. In the case of Blue Coat and Riverbed, this is an SSL session; in the case of Certeon, it is an IPsec connection.
- 3** The data-center side optimizer decrypts the traffic from the WAN, rebuilds it to its unoptimized form, then re-encrypts it into an SSL session with a server. All three vendors copy server SSL certificates to their appliances, but do not copy the certificates to branch machines.

WAN OPTIMIZATION

Section 2

Analysts' take on the market

What makes a WAN optimization controller?

Need to increase application performance over the WAN? Choosing a vendor is a key first step

■ *By Ann Bednarz*

Choosing a vendor can be the most challenging part of a WAN optimization project, given the wide range of vendors, features and technology options available today. Enterprises in the market for WAN optimization controllers (WOC) may find a new report from Gartner can help distinguish among the offerings available from leading players.

Gartner's magic quadrant for WAN optimization controllers was written by analysts Andy Rolfe and Joe Skorupa. In it the analysts rate vendors according to the completeness of their technology vision and their ability to execute on that vision, reserving the "leaders" quadrant of the matrix for those WOC vendors that excel in both categories.

WOCs typically are dedicated devices, deployed on either end of a WAN link to improve application response times, according to the research firm. WOC technologies generally work to prevent network latency using protocol optimization techniques; reduce data traveling over the WAN via compression or caching; and prioritize traffic streams according to business needs.

But the similarities end there. In a market as innovative and dynamic as the WAN optimization arena, players have chosen to focus on

specific areas to distinguish their offerings. According to Gartner, some of the features buyers can find in WOC product sets today include:

- Application-specific acceleration features for TCP/HTTP and Microsoft Common Internet File System (CIFS) file access.
- Enterprise content delivery network (CDN) features that pre-position content in the cache to speed the delivery of live and on-demand streaming media content.
- Client software designed to speed application delivery to remote PCs.
- Bundled DHCP, DNS, Active Directory and print server features to appeal to enterprises seeking an "all-in-one" branch-office appliance.
- Ability to monitor and enforce application-specific service-level agreements.
- Security capabilities such as SSL optimization and encrypted tunnels between WOCs and encrypted discs.
- Integration of router features into WOC equipment, such as policy-based routing and route selection.

The variety of features available can make vendor selection difficult, the research firm warns. Before choosing a vendor, "ensure you understand the applications and services running on your network, and the protocols they use," the Gartner report reads. "Also conduct a detailed analysis of your network traffic to identify specific problems - for example, exces-

sive latency, bandwidth oversubscription or lack of prioritization for certain types of traffic. Finally, insist on a real-life trial before committing to any purchase."

Also, don't think too long term, the firm warns. Looking ahead, Gartner expects that the growing popularity of managed WAN optimization services and the trend to bundle WOC features into other network gear, such as routers, may change the market:

"Our advice on stand-alone WOC equipment selection should therefore be considered in light of these anticipated changes in the market, and we recommend that selection is made on the basis of relatively short payback times (typically less than three years) and on current and near-term product capabilities," the report reads.

Among the WOC vendors Gartner studied, four placed in its leaders quadrant: Riverbed Technology, Juniper Networks, Blue Coat Systems and Expand Networks. Here is just a small sample of the positive things Gartner analysts had to say about the vendors in the leaders quadrant:

Riverbed's Steelhead family "is the broadest lineup of appliances on the market, ranging from a low-cost entry model to a high-end cluster," Gartner reports. Expand stands out for offering "the best mix of QoS, reporting and acceleration features," while Blue Coat "has strong market understanding, demonstrated

Sponsored by Silver Peak

www.silver-peak.com

Section 2: Analysts' take on the market ●●●

through its broad WOC range and feature set." Gartner calls Juniper an "early compression innovator," and says its WX and WXC WOC products "have a broad feature set, including User Datagram Protocol (UDP) and MAPI acceleration."

Those aren't the only standout products, however. "Challengers" Cisco and Packeteer, as well as "visionaries" F5 Networks and Citrix Systems also excel in key areas, as do "niche players" like Silver Peak Systems, Ipanema Technologies, Exinda Networks, Certeon, Intelligent Compression Technologies and Stampede Technologies.

Certeon offers encrypted disks in its WOC appliance, for example, and Exinda "offers the lowest-entry-cost disk-based appliances on the market," Gartner reports. Silver Peak "has shown particularly good understanding of the acceleration needs of data replication and backup," and Packeteer "has excellent QoS, reporting and application visibility."

Still, none are perfect for every enterprise scenario - even the leaders. Juniper lacks a SoftWOC, for example, and Riverbed has "less-capable QoS and reporting features than some leading vendors," Gartner points out. Blue Coat's Mach5 WOC "has weaker reporting and monitoring capabilities than some other leading vendors' products," Gartner warns, and Expand "has limited WAN capacity at high end."

For each of the 14 vendors in the matrix, Gartner spells out strengths and weaknesses related to variables such as company viability, sales execution, market understanding, business model, innovation and, of course, the actual products being sold.

For enterprises, the trick is finding the vendor with the set of features that's the best fit for their network setting. To do so, the strengths and weaknesses of each player have to be weighed against the needs of the enterprise.

To be included in the magic quadrant, vendors' products had to include capabilities in at least three of the four broad categories of WAN acceleration techniques Gartner identified: traffic management capabilities such as WAN QoS classification and enforcement or traffic shaping; compression, caching and/or data replication or reduction capabilities; generic protocol acceleration such as for TCP or HTTP; and application-specific optimization features, such as acceleration of the CIFS file-sharing protocol.

Application acceleration: Divide and conquer



■ By Kevin Tolly

More than perhaps any other area except virtualization, the term "application acceleration" can be defined and characterized in almost as many ways as there are products on the market. Still, if you take the right approach, it is relatively easy to build a taxonomy that will help you choose what is right for your needs.

At The Tolly Group, we've been testing application acceleration products since before the technology picked up that appellation. Over time we've had the opportunity to benchmark and dissect products from, among others, Allot Communications, Citrix Systems, Expand Networks, Ipanema Technologies, Packeteer and Riverbed Technology (and more in the works)

In the beginning, acceleration typically was implemented using compression to make the oversubscribed WAN pipe appear larger. In a way, that was virtualization as well - because a T1 could function as if it virtually was multiple T1 links bonded together.

To this usually was added a QoS function that made sure packets belonging to latency-sensitive applications wouldn't sit behind a time-insensitive file transfer that might be shoveling thousands of large data packets across the link.

Over the years, however, the techniques employed to deliver acceleration increased dramatically. For example, vendors dissected elements of the data stream and found, for example, that TCP acknowledgement packets flowing back and forth on the link could slow down user performance. Or they found information was sent repeatedly across the link - again, a cost to performance that offered little benefit. As a result, application acceleration products began to tinker with the underlying protocols - reaching ever higher to the ultimate target of acceleration: the application.

That brings us to the "divide and conquer" of the title. Somewhere along the way, some vendors realized specific applications often had unique characteristics that, once understood, could be optimized.

Thus, we see the split between vendors offering general-purpose acceleration vs. those citing very specific applications that their products optimize. Make sure you are aware of the difference, because some vendors might not make it crystal clear. If you note that a product's literature says, "improves Microsoft SharePoint" or "better PeopleSoft response time," it may well be that the acceleration is tuned to those applications. Although the product might improve general applications also, realize that your mileage may vary: You might not get the dramatic benefits that will accrue to users of the applications the product specifically optimizes.

In addition, with all these products it is critically important the scenarios benchmarked are relevant to you. Labeled "real world" or not, the only important real world is your world. Because any application-acceleration benchmark, practically speaking, looks at only a very small slice of all network and application scenarios, prudent network architects must diligently attempt to understand the specifics of the optimization being illustrated and ultimately, its relevance to the corporate environment.

And just to make things more interesting, our friends at Microsoft now are making an impact on the world of application acceleration. As noted earlier, many application-acceleration products "fix" parts of protocols, such as TCP and Server Message Block (SMB), to improve WAN performance.

With the introduction of Vista and Longhorn, Microsoft also introduces the first major rewrite of the TCP stack in years: It implements RFCs that improve WAN performance. And SMB 1.0, which has been around forever - since about 1980, at least - finally is being replaced by SMB 2.0.

Will this stir the pot in application acceleration? Yes, but that's another story.

Sponsored by Silver Peak
www.silver-peak.com

Section 2: Analysts' take on the market ●●●

How do your WAN optimization efforts measure up?

Aberdeen Group studies how companies improve application response times and combat WAN latency

■ By Ann Bednarz

How does a thousand-fold improvement in application response time sound? Or an 87% improvement in bandwidth utilization? It may sound too good to be true, but one research firm says companies with the right policies and technologies in place are seeing gains like these.

Aberdeen Group surveyed 235 companies for its new report, "Optimizing WAN for Application Acceleration." On the one hand, the responding companies have a lot in common. Nine out of 10 say they need more visibility into and control over the performance of their networks. In addition, 87% agree that managing their networks has become more complex over the last two years.

Another commonality is an increasingly distributed workforce. Among the respondents, 45% have 20 or more remote network locations, and 49% have 100 or more.

But when it comes to dealing with these performance concerns and management challenges, the respondents reported varying degrees of success.

To gauge that success, Aberdeen divided the respondents into three camps: Best-in-class, which includes the top 20% of respondents as measured by the research firm's performance criteria; industry average, which represents the middle 50%; and laggard, which makes up the bottom 30%.

Of course it's the best-in-class companies that have yielded the most gains, including 87% average improvement

in bandwidth utilization, 1,158% average improvement in application response times, and 100% decreased WAN latency. The laggards, by comparison, reported 16% average improvement in bandwidth utilization, 10% average improvement in application response times, and 0% decrease in WAN latency.

The industry average fell in between the two extremes, logging 40% average improvement in bandwidth utilization, 99% average improvement in application response times, and 41% decreased WAN latency.

But just what are the best-in-class companies doing that the laggards aren't?

For starters, they are twice as likely to have policies in place for prioritizing WAN traffic (73% compared to 35%) and twice as likely to have defined traffic thresholds for optimal application performance (41% compared to 19%).

Manpower is also a distinguishing factor. Thirty-five percent of best-in-class companies have a dedicated staff for application performance management, compared to just 5% of laggards.

Not surprisingly, the WAN optimization and application acceleration technologies used by best-in-class companies far exceed those used by laggards. For example, 42%



Silver Peak

The Leader in Scalable WAN Acceleration

100x Improvement in WAN Performance

"Best Current Product Offering" – Forrester

The Best Performance on Large Networks

- > High capacity WAN links
- > Many distributed offices
- > Diverse mix of applications

www.silver-peak.com

Sponsored by Silver Peak
www.silver-peak.com

Section 2: Analysts' take on the market ● ● ●

of best-in-class companies use byte caching, compared to 2% of laggards. Application-specific compression tools are used by 40% of best-in-class companies, compared to 15% of laggards. TCP acceleration tools? About 32% of best-in-class have them, compared to just 7% of laggards.

Lastly, 65% of best-in-class organizations can measure application and response times in real time, or near real time, compared to 22% of laggards.

So what do you think? Is your company best-in-class when it comes to WAN optimization? A laggard? Or somewhere in between? If there's room for improvement in your outfit, consider some of Aberdeen's suggestions for getting better mileage out of WAN resources:

- Develop policies and deploy tools for prioritizing WAN traffic. "Organizations need to make decisions about how they want their WAN bandwidth to be used and deploy technology solutions to enable network managers to reinforce those decisions," the report reads.
- Define traffic thresholds and plan WAN capacity based on historic data. Companies that can view bandwidth consumption per network location, and per application, can better plan for changes in capacity and preserve optimal application performance, Aberdeen says.
- Implement auto alerts for issues with application performance. The quicker network managers become aware that traffic has reached a defined threshold, the faster any performance problems can be repaired.
- Centralize management of WAN optimization appliances. The ability to manage remote WAN appliances from a central location is critical if a company wants to get to best-in-class performance, Aberdeen says.

What doesn't work is simply throwing bandwidth at the problem. Most companies have tried adding bandwidth capacity, but that alone isn't going to bolster application response times as much as end users would like, Aberdeen reports. While 93% of respondents said they increased their bandwidth over the last two years, only 52% saw an improvement in application response times and only 36% saw WAN latency decrease.

For lots more tips and information, check out the full report, "Optimizing WAN for Application Acceleration," sponsored by Virtela Communications, Shunra, Mazu Networks and Exinda Networks. A free copy is available on Aberdeen's site.

WAN OPTIMIZATION

Section 3

Real-world tests and shopping tips for buyers

WAN acceleration offers huge payoff

Riverbed edges Cisco, Silver Peak and Blue Coat in test of devices that speed traffic, cut bandwidth and save money

■ *By David Newman*

Imagine walking into the CIO's office tomorrow and saying, "I can cut our WAN consumption by as much as 80 times, speed file transfers as much as 45 times and make our Windows users a whole lot happier." Think you'd get the CIO's attention?

Those aren't just idle claims. Seven months of rigorous testing showed us why application acceleration is such a hot area: These devices really work.

We tested boxes from Blue Coat Systems, Cisco, Riverbed Technology and Silver Peak Systems in a true enterprise context, with a massive test bed pushing data over multiple T3 and T1 links. After pounding the systems with the most popular enterprise applications, we're inclined to believe the hype.

Even if average speedups are "only" around 5 to 10 times, that's still a big improvement. With 31% of IT budgets eaten up by recurring monthly WAN costs, according to a recent Nemertes Research study, application acceleration promises potentially huge cost savings.

Riverbed's Steelhead appliances outperformed the field in most tests, and won our Clear Choice award.

But all these devices deserve serious consideration: Blue Coat's SG appliances for solid HTTP optimization; Cisco's Wide Area Application System (WAAS) for excellent compression, traffic transparency and interoperability with other devices; and Silver Peak's NX appliances for strong scalability and intuitive traffic reporting tools.

Why is Windows so bad?

The problem statement for application acceleration is simple: Windows performance in the WAN is lousy. To begin with, Windows' two workhorse protocols – TCP and NetBIOS – were never intended for use in low-bandwidth or high-delay networks. Windows XP Service Pack 2 compounds these problems with some spectacularly

suboptimal configuration defaults. (Windows Vista is better, but it isn't widely implemented yet.)

By default, XP's TCP stack advertises a receive window – the maximum amount of data allowed in flight without acknowledgment – of 64KB. That's fine as far as it goes, but XP isn't very responsive about resizing that window in response to loss or delay. A large, static receive window contributes to retransmissions, possible packet loss and poor response time.

To make matters worse, XP doesn't use a common TCP option called window scaling that can expand a 64KB receive window by a factor of four or more. Even when network conditions let XP go much faster, it won't. (There is a registry hack to enable window-scaling, but even then, it isn't used by the Windows file-handling protocol.)

WAN performance is always limited by the so-called bandwidth-delay product, but the constraints with Windows clients are especially severe. For example, if a link between Boston and Los Angeles has a 100-msec round-trip delay and the Windows TCP receive window is 64KB, the highest transmission rate possible is only around 5.6Mbit/s, regardless of link speed. Ordering up a T3 or OC-3 won't help, at least not for any given Windows TCP connection; 5.6Mbit/s is as good as it gets.

WAN-acceleration devices compensate for these shortcomings with a variety of tricks, including block caching, compression, connection multiplexing and application-layer optimization. While not all devices implement every method, all sharply reduce response time and bandwidth for Windows applications across the WAN.

Faster file service

As part of our research for this test, we asked vendors and several corporate IT shops to name their top five candidates for application



Sponsored by Silver Peak

www.silver-peak.com

Section 3: Real-world tests and shopping tips for buyers ● ● ●

acceleration, and every respondent named Common Internet File System (CIFS) as its top pick. This is understandable, given that Microsoft's notoriously chatty file-handling protocol originally was intended for LAN-only operations. Given its popularity and performance issues, we made CIFS the highlight of our performance testing.

We tested application acceleration the way enterprises use it – with multiple WAN links and round-trip times. Our test bed modeled a hub-and-spoke WAN linking with a headquarters office plus four remote sites, two apiece on T-1 and T-3 links. The remote sites represented every permutation of high and low bandwidth and delay.

At each of the remote sites, we configured XP clients to upload and download directories containing Word documents from a Windows Server 2003 machine at headquarters.

To measure the effects of block and/or file caching, we ran the CIFS tests three times. First was a "cold run" with all caches empty. Second was a "warm run" that repeated the same transfer as the cold run, this time with the files already in cache. Finally, we changed the contents of 10%

of the files; this "10% run" forced devices to serve some but not all content from the origin server.

The two most important application-acceleration metrics are bandwidth reduction and response-time improvement. While we measured both in this test, our results show there's not necessarily a strong correlation between the two. A device with a powerful compression engine might do well at reducing bandwidth consumption, but the time spent putting the squeeze on data might increase response time or, at best, yield only modest improvements. Conversely, some devices might willingly trade off a bit more bandwidth consumption if the net result is faster overall data delivery.

Looking first at bandwidth-reduction

results, all products substantially lightened the WAN load, but big differences exist across devices depending on cache contents.

For example, in the cold run (caches empty), Cisco's Wide Area Engine (WAE) appliances were by far the most effective at compression, using nearly 28 times less bandwidth than was used in our baseline, no-device test. In contrast, the bandwidth savings for other devices seeing data for the first time was usually less than a two-times reduction in bandwidth, according to measurements taken by a ClearSight Networks Network Analyzer.

Note that we're presenting all results in terms of relative improvement rather than absolute numbers. For example, in the

SCORECARD

Category	Weight	Riverbed	Cisco	Silver Peak	Blue Coat
Performance	45%	5	4.5	4.25	4.5
Functionality	20%	4.5	4.5	4.5	4.5
Manageability	20%	4.5	4.5	4.5	2.75
Usability	15%	4	4	4	2.5
Total score		4.65	4.43	4.31	3.85

NETRESULTS

Product	Steelhead	Wide Area Application System (WAAS)	NX appliances	SG8100, SG200, Director
Vendor	Riverbed Technology www.riverbed.com	Cisco Systems www.cisco.com	Silver Peak Systems www.silver-peak.com	Blue Coat Systems www.bluecoat.com
Version tested	4.0.0a	4.0.12.b33	2.0.0.0_15619	5.1.4.21.29594
Price	Steelhead 5520, \$70,000; Steelhead 3520, \$45,000; Steelhead 1020, \$12,500; Central Management Console 8000, \$5,000. Total as tested, \$190,000.	WAE-7371, \$120,000; WAE-612, \$22,000; WAE-512, \$14,800; Central Manager, \$11,800. Total as tested, \$303,400.	NX-7500, \$65,000; NX-5500, \$45,000; NX-2500, \$10,000; GX-1000, \$10,000. Total as tested, \$185,000.	SG8100, \$51,200; SG200, \$5,500; Director, \$18,000. Total as tested, \$182,700.
Pros	Top performer; excellent reporting; easy-to-use interface.	Strong in CIFS performance, compression, concurrent connections; good reporting via new dashboard.	Long list of predefined application types; good reporting; easy-to-use interface.	Long list of optimized applications; strong SSL performer.
Cons	External marking required for QoS tests; SSL optimization not supported yet.	Sped up HTTP less than some other devices; SSL optimization not supported yet.	Performance in some tests didn't match vendor's results.	Some software issues in testing; central management not as unified as in others.
Score	4.65	4.43	4.31	3.85

Sponsored by Silver Peak
www.silver-peak.com

Section 3: Real-world tests and shopping tips for buyers ● ● ●

CIFS cold run, Cisco's devices consumed 130MB of WAN bandwidth, compared with 3.6GB with no acceleration device inline, which translates into using 27.82 times less bandwidth.

Given that enterprise data patterns are repetitive and subject to change, bandwidth reduction in the warm and 10% test cases can be more meaningful – and this is where these devices really shine.

Riverbed's Steelhead appliances topped these tests, reducing bandwidth by a factor of 84 in the warm run and a factor of 32 in the 10% run. While the other devices reduced bandwidth by a lesser degree, the improvements were still dramatic. Any device that reduces bandwidth use by 20 or 30 times must be considered a boon to IT budgets.

We also used the ClearSight analyzer to measure LAN bandwidth consumption (see graphic "CIFS LAN bandwidth reduction").

LAN differences among products were not as dramatic as WAN differences. The Blue Coat and Cisco devices reduced LAN bandwidth consumption by factors of 1.5 to 2 in our warm run and 10% run, because these vendors' headquarters devices served objects out of cache instead of from the origin servers. In contrast, the Riverbed and Silver Peak devices increased LAN use by 2% to 10%, probably because of appliance-control traffic. Changes in bandwidth use don't always correlate with changes in response time, however.

Measuring CIFS response time

We used a common enterprise task to gauge CIFS response time, measuring how long it took for a client to upload or download a set of Word files to or from a server. We measured transfer times at each of our four remote sites – each representing a different permutation of high and low bandwidth and delay. We're presenting the results for each site because users' requirements differ depending on where they work. As our results suggest, some appliances do a better job at accelerating CIFS in low-bandwidth settings; others are better for high-delay settings.

Arguably, the most important results for enterprises are from the 10% runs, where we

offered 10% new content and 90% existing content to each set of appliances. This represents an enterprise where many users might see the same documents repeatedly but where there also would be some new documents added to the mix.

In the download tests, low-bandwidth sites tended to see the biggest improvements in response time, regardless of the amount of delay present.

Riverbed's Steelhead appliances sped up file transfers 45 times to a low-bandwidth, low-delay site and 34 times to a low-bandwidth, high-delay site. The Steelhead appliances were also tops for the high-bandwidth sites, but to a lesser degree, with speed increases of four to seven times.

The Silver Peak NX appliances were next most efficient overall, with speedups of three to 16 times (again, with the most

improvement shown for low-bandwidth sites), followed by the Cisco and Blue Coat appliances.

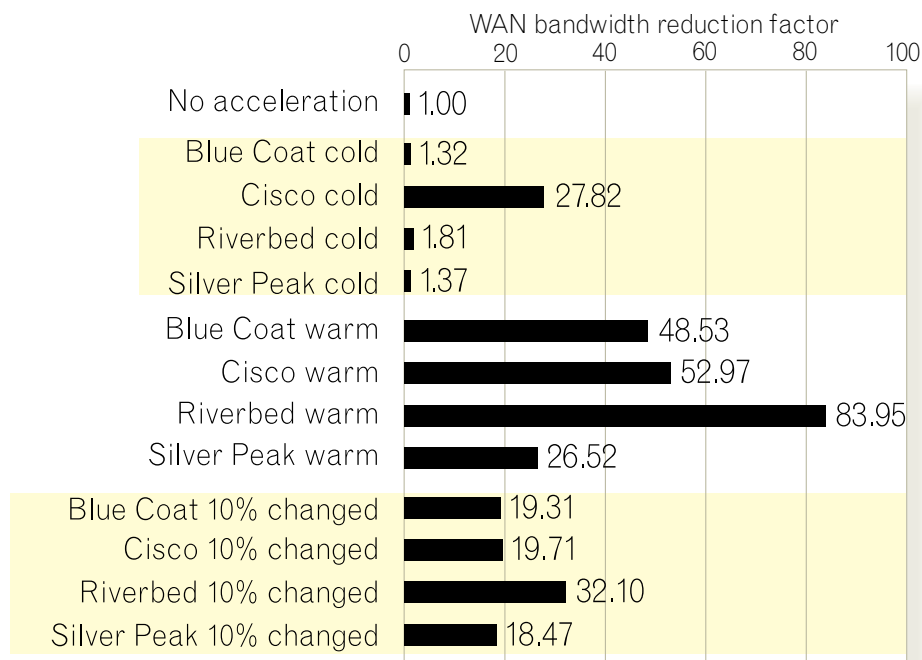
File uploads generally don't benefit from application acceleration as much as downloads do. When handling client downloads, acceleration devices either serve content from a client-side cache, pipeline data using read-ahead operations or employ some combination of the two approaches. That's not possible with write operations, because an acceleration device can't predict in advance what data the client will send.

Even so, big improvements in upload performance are still possible.

Riverbed's Steelhead appliance again led the pack, with speedups of three to 34 times compared with no acceleration. Accelerations from the Silver Peak, Cisco and Blue Coat devices were less dramatic but

CIFS WAN bandwidth reduction [FIGURE 1

Application-acceleration devices reduce WAN bandwidth consumption the most when they handle data they've previously seen. We tested each device three times: "cold run" with caches empty, a "warm run" with caches populated and a "10% run" with 10% of files changed. In the warm run, Riverbed's Steelhead used nearly 84 times less bandwidth than a baseline test with no device.



Section 3: Real-world tests and shopping tips for buyers ● ● ●

still significant, moving traffic 1.3 to 16 times faster than our baseline test. Most devices sped up data the most from low-bandwidth sites. Blue Coat's SG was an exception; it delivered the greatest upload benefit to the high-bandwidth, high-delay site.

Note that response-time improvements do not track linearly with bandwidth-reduction results. For example, Cisco's devices were more efficient, relative to their competitors, at reducing WAN bandwidth consumption than at speeding CIFS transfer times.

In reviewing the CIFS results, Riverbed commented that it achieved even greater improvement over no-acceleration baselines by using many small files. Our tests used a mix of random file sizes of 25KB to 1MB. Both approaches have their merits: Riverbed's short-file methodology is more stressful on devices' CIFS processing engines (stress is a good thing in device benchmarking), while a mix of larger files may offer a more meaningful prediction of device performance in production settings.

Mail call

After CIFS, the next most popular candidate for acceleration is Messaging API (MAPI) traffic. MAPI is the e-mail protocol used by the Microsoft Exchange server and Outlook clients. All devices tested can speed up MAPI traffic, but in our tests the improvements were far less significant than in the CIFS tests.

In our MAPI tests, all clients sent messages – some with Word attachments, some without – to all other clients through an Exchange 2003 server. As with the CIFS tests, the number of messages was proportional to each site's link speed – fewer messages for clients at T-1 sites, more for those at T-3 sites.

There was significantly less differentiation among products when accelerating MAPI traffic, compared to CIFS traffic.

All products sped mail delivery, but only by factors of 1.24 to 2.39 compared with a no-device baseline. Averaging results across all sites, the Blue Coat devices provided the biggest boost for mail traffic, but by a relatively small margin over the Riverbed, Silver Peak and Cisco devices.

Doubling e-mail performance is nothing

to sneeze at, but we also wanted to understand why MAPI performance didn't match CIFS performance. A few minutes with the ClearSight analyzer gave us the answer: The Outlook 2007 clients we used in this test encrypt e-mail traffic by default.

To the acceleration appliances, most of the MAPI data structures weren't visible to be optimized. Some acceleration was still possible, through TCP optimizations or because some MAPI traffic was visible. After reviewing the results, Riverbed said it encourages Outlook 2007 users to disable encryption for highest performance. That said, network managers using the new version of Outlook should consider whether the security/performance tradeoff is worthwhile.

A faster Web

We measured acceleration of HTTP traffic in two tests, one with 248 and one with 2,480 concurrent users. The results were a bit surprising: While the products delivered Web traffic as much as seven times faster than a baseline test without acceleration,

performance didn't necessarily improve as we added more users.

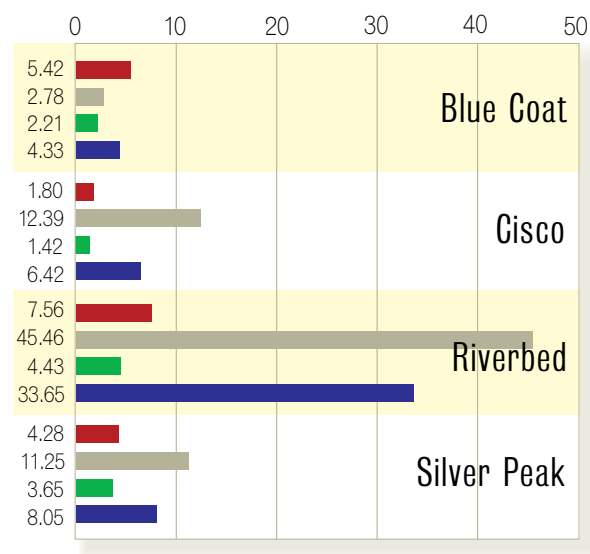
To avoid overloading the sites on slower links, we put proportionately fewer users at the T-1 sites than at the T-3 sites. For example, our 2,480-user test involved 1,200 clients at each of two sites on a T-3, and 40 clients at each of two sites on a T-1. We used Spirent Communications' Avalanche/Reflector tool to emulate Web clients and servers. Because previous studies of Web objects place the average size at 8K to 13KB, we configured the clients to request an 11KB object from the servers.

As in the CIFS and MAPI tests, the Riverbed Steelhead appliances delivered Web traffic the fastest (see graphic Web acceleration). In all three ways we measured – transactions per second, traffic rates and response time – the Steelhead appliances delivered Web traffic seven times faster than tests with no device inline. We observed the same seven-times improvement with 248 and 2,480 users; because LAN and WAN bandwidth use was almost identical in each

Downloading CIFS [FIGURE 2

File downloads go much faster with application acceleration. In situations where 10% of user data changed, downloads moved as much as 45 times faster with acceleration than without it. The biggest boosts generally came on lower-speed T-1 links, regardless of delay.

RESPONSE-TIME IMPROVEMENT FACTOR (NO DEVICE = 1.00)



■ HIGH BANDWIDTH, HIGH DELAY ■ HIGH BANDWIDTH, LOW DELAY
 ■ LOW BANDWIDTH, LOW DELAY ■ LOW BANDWIDTH, HIGH DELAY

Section 3: Real-world tests and shopping tips for buyers ●●●

test, it's likely that WAN bandwidth was the bottleneck.

Blue Coat's SG appliances were second fastest, but that result must be stated with a caveat: The Blue Coat boxes worked better with fewer Web users, not more. Compared with no acceleration, the Blue Coat appliances boosted Web performance by around seven times for 248 users, but by around six times for 2,480 users (and that's just for transactions per second and data rate; the response time improved by only a factor of three).

We noticed some erratic Address Resolution Protocol (ARP) behavior in tests involving 2,480 users when Blue Coat forwarded either Web or SSL traffic (see "What about SSL?"). Although Blue Coat replicated our issue in-house and produced a software fix (now available to customers), we still observed sluggish behavior in the 2,480-user tests after applying the update.

Silver Peak's NX appliances were third-fastest, tripling transaction and data rates and reducing response time by around 2.5 times when handling 248 users. With 2,480 users, performance dipped slightly (by about the same margin as Blue Coat's appliances), though traffic still moved substantially faster than in our no-device baseline test. Silver Peak says these results are roughly in line with its in-house testing.

Cisco's WAE appliances better than doubled performance with 248 users, and more than tripled performance with 2,480 users. Cisco's WAE devices don't proxy Web traffic as they do with CIFS, so the performance improvements here are largely attributable to TCP optimizations.

QoS testing

QoS testing revealed some of the most interesting – and in some ways most problematic – results of all our performance testing. While three of four products did a virtually perfect job of prioritizing traffic, the path there was anything but straightforward, involving much tuning – and in some cases external devices to protect key flows during congestion.

To measure QoS capabilities, we offered a small amount of high-priority traffic – in this

case, a single VoIP call, which is sensitive to delay and jitter – while walloping the WAN with huge amounts of background traffic. We used User Datagram Protocol (UDP) for both high- and low-priority flows; VoIP uses UDP by default, and TCP was not suitable as background traffic, because of its built-in congestion control.

We also determined whether devices could "re-mark" Diff-Serv code points (DSCP), a good practice in guarding against rogue users or applications marking their flows with an inappropriate priority.

Blue Coat's SG appliances couldn't participate in this test, because they don't optimize UDP traffic. The other vendors turned in excellent results but used different paths to get there.

Cisco recommends using WAN routers (in this case, the Cisco 3845 and ISR 2800 Series devices it supplied) rather than application accelerators for shaping traffic. Cisco's

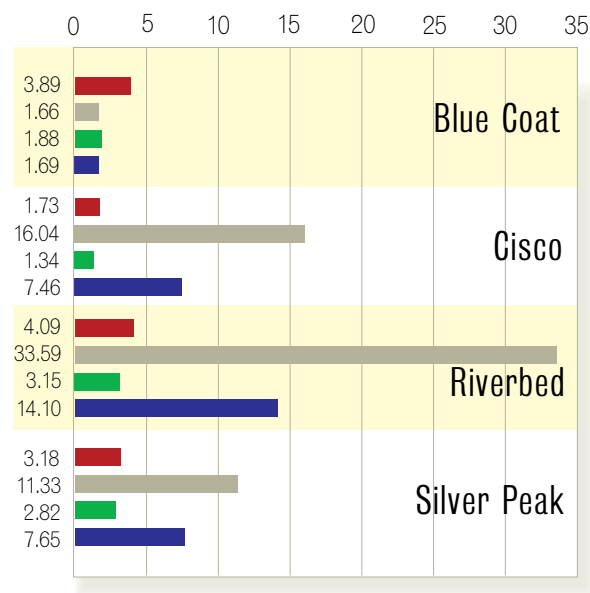
WAAS-acceleration devices and routers work together using network-based application recognition (NBAR). We verified in testing that flows the acceleration devices classified using NBAR will be prioritized by the routers during congestion. The routers turned in great results; the ClearSight analyzer measured R-value, an audio-quality metric, as 92.03 out of a possible 93, and they correctly re-marked DSCPs.

Note that ultimately Cisco's entry performed prioritization on its routers, not on the application-acceleration devices, though the latter did play a role in classifying traffic. This differs from the Riverbed and Silver Peak devices, which performed prioritization on board. Many network managers already run QoS on WAN routers, and for them handing off this function to a router isn't a big deal. For users just getting started with QoS, it may be simpler to set it up on application-acceleration devices, and leave

Uploading CIFS [FIGURE 3

While file uploads don't benefit from acceleration as much as downloads, the benefits still can be substantial. Uploads from clients to servers moved as much as 33 times faster with acceleration than without it. As with downloads, the greatest benefits came on lower-speed T-1 links, although all the speedups were significant.

RESPONSE-TIME IMPROVEMENT FACTOR (NO DEVICE = 1.00)



■ HIGH BANDWIDTH, HIGH DELAY ■ HIGH BANDWIDTH, LOW DELAY
 ■ LOW BANDWIDTH, LOW DELAY ■ LOW BANDWIDTH, HIGH DELAY

Section 3: Real-world tests and shopping tips for buyers ●●●

routers alone, at least for now.

The Riverbed and Silver Peak appliances also protected voice traffic, with R-value scores of 91.80 and 90.07, respectively, and both correctly re-marked DSCPs.

Of the two, the Silver Peak NX appliances were easier to configure. They correctly classified VoIP streams and shaped traffic according to the parameters we defined. Riverbed's Steelhead appliances don't classify real-time protocol streams automatically, and a bug in the software version we tested wouldn't let us manually define port ranges. Instead, we used other criteria, such as source address, to classify VoIP streams.

Concurrent connections

Our final performance test determined the maximum number of TCP connections each system could optimize. This is an important metric for enterprises with many remote offices and hub-and-spoke network designs, where connection counts for data-center devices can run into the tens of thousands. All the devices we tested get into that tens-of-thousands range, but there was more than a fourfold difference between the highest and lowest capacities.

To measure connection concurrency, we configured Spirent's Avalanche to issue a Web request once a minute, letting us establish many connections and keep them alive. We kept adding connections until transactions began to fail or the devices stopped optimizing new connections.

Cisco's new WAE-7371 came out tops in this test, accelerating more than 50,000 TCP connections (see graphic Maximum accelerated TCP connections). Silver Peak's NX appliances were next, optimizing 43,306 concurrent connections. This is well short of the NX 7500's rated capacity of 128,000 optimized connections, a level that Silver Peak achieved in internal testing. We were unable to reproduce that result in our lab, and, despite extensive troubleshooting, neither we nor Silver Peak's engineers were able to explain the difference. The Blue Coat SG appliances were next, handling about 19,500 optimized connections.

Riverbed's Steelhead 5520 optimized more than 12,200 connections, but that result

reflects the limits of the two Steelhead 3520 units through which we set up connections. Riverbed says the higher-end 5520 model can optimize 15,000 connections. We were unable to confirm that result, but our tests did show that each 3520 slightly outperformed its rated limit of 6,000 connections to get to the 12,200 total mentioned previously.

Features and functions

Most testing focused on performance, but we also assessed devices for functionality, manageability and usability. Each of these areas turned up at least as many differences as the performance tests did.

All acceleration devices reduce the number of bits sent across the WAN, but they do this in very different ways. The Blue Coat and Cisco devices act as proxies, terminating connections between clients and servers and setting up new sessions on their behalf. Riverbed's devices can proxy traffic, though the vendor did not enable that feature for this test. Silver Peak's NX appliances don't

proxy traffic.

Transparency is another architectural difference. Blue Coat and Silver Peak engineers respectively configured SSL or generic routing-encapsulated tunnels between appliances, and Riverbed can use SSL tunneling. Tunneling may pose a problem if other inline devices, such as firewalls or bandwidth managers, need to inspect traffic.

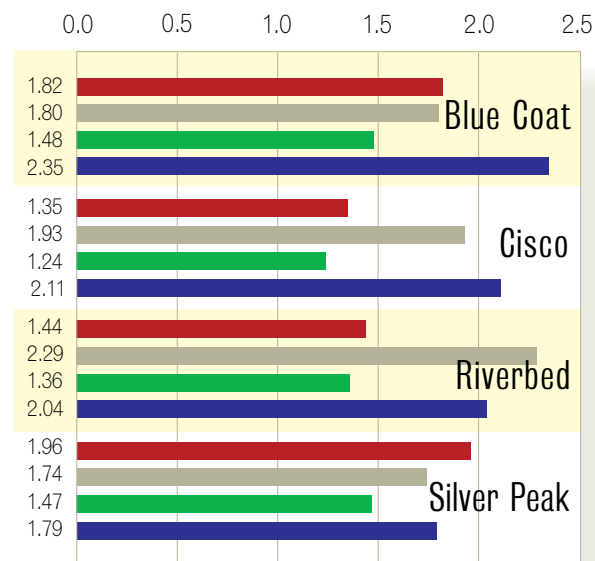
Cisco claims this is a major differentiator for its WAAS offering, which doesn't hide traffic from other devices and automatically learns about new traffic types from other Cisco devices using NBAR. A powerful classification engine, NBAR in our tests classified even applications using ephemeral port numbers, such as for such as those used for H.323 and Session Initiation Protocol. Silver Peak's appliances also classified such traffic. Then again, transparency isn't an issue for users who don't need application visibility among acceleration devices.

Application support also varies, but it's less important a differentiator than

MAPI acceleration [FIGURE 4

The MAPI e-mail protocol showed lower but still significant improvement in our tests compared with CIFS file traffic. Differences between products also were much smaller. This may be tied to our use of Outlook 2007 clients, which by default enable encryption of e-mail contents, giving appliances fewer opportunities for optimization.

RESPONSE-TIME IMPROVEMENT FACTOR (NO DEVICE = 1.00)



■ HIGH BANDWIDTH, HIGH DELAY ■ HIGH BANDWIDTH, LOW DELAY
 ■ LOW BANDWIDTH, LOW DELAY ■ LOW BANDWIDTH, HIGH DELAY

Sponsored by Silver Peak
www.silver-peak.com

Section 3: Real-world tests and shopping tips for buyers ● ● ●

performance, manageability and usability. It's tempting – but also a bit misleading – to compare the number of predefined application types each vendor claims to optimize. First, the applications involved are important only if they're running in your enterprise. Second, acceleration devices still may boost performance even if a given application isn't predefined, thanks to compression and TCP optimization. Finally, all devices we tested allow manual definition of new application classes based on addresses and port numbers (though these may not be subject to the same speedups as some predefined types).

Manageability

To look after all the devices in our test bed's enterprise, we asked each vendor to supply a central management system.

We assessed centralized management in terms of functions and reporting features. On the functions side, all vendors but Blue Coat offer a centralized method of pushing out configuration changes or software upgrades to all appliances. Blue Coat indeed can push changes and upgrades but only by manually defining a job to push out the change. All vendors allow appliances to be defined into groups (though Blue Coat's Director appliance requires a manually defined job to perform an action on a given group).

All devices use a dashboard display to show application distribution and volume during predefined periods. These displays can be enormously helpful in managing application traffic even before acceleration is enabled. It's pretty common to find during

installation that enterprises are running applications they didn't know about.

Once acceleration is enabled, these devices use pie charts and bar graphs to report on compression, percentage of optimized vs. pass-through traffic and data reduction.

The Cisco, Riverbed and Silver Peak appliances aggregate displays across multiple devices, a useful feature for capacity planning. There were differences in terms of the application data and time periods supported; for example, Silver Peak's display was useful in troubleshooting because – uniquely among the products tested – it reported on packet loss and did so in per-minute intervals.

Usability

There are significant usability differences among the accelerators, but we'll be the first to admit this is a highly subjective area. If we had to rank the systems in terms of ease of use, the lineup would be Riverbed, Silver Peak, Cisco and Blue Coat.

Riverbed's Steelhead appliances came closest to the goal of "just working." Setup took less than half a day. Once we were up and running, we found the user interface to be simple and well designed. It was easy to make changes and view reports, even without delving into the company's well-written documentation.

Silver Peak's NX appliances also feature a simple user interface with excellent reporting on current and historical statistics. The central management display wasn't as polished or fully featured as Riverbed's,

although unlike Riverbed's, it includes a topology map of all appliances.

Cisco's display bristles with features and commands – perhaps too many. Cisco's redesigned dashboard offers whizzy graphics, useful pie charts on CIFS application performance and (like Riverbed and Silver Peak devices) real-time connection monitoring and per-device reporting on connection statistics. Getting to specific commands or opening logs often took more steps than with other devices, however; further, not all the commands available from the device command line were available from the GUI, and vice versa.

Blue Coat's management software, while powerful, was the most difficult to use. Individual appliances used a Web-based Java application that was sluggish; further, it worked with Internet Explorer but not Firefox. And some predefined tasks in other vendors' devices, such as updating configuration or images, required manual definition in the Blue Coat devices, or touching each appliance individually.

Newman is president of Network Test, an independent test lab in Westlake Village, Calif. He can be reached at dnewman@networktest.com. Newman is also a member of the Network World Lab Alliance, a cooperative of the premier reviewers in the network industry, each bringing to bear years of practical experience on every review. For more Lab Alliance information, including what it takes to become a member, go to www.networkworld.com/alliance.

Section 3: Real-world tests and shopping tips for buyers ● ● ●

How we tested application-acceleration devices

■ By David Newman

We tested application-acceleration devices for performance, features, manageability and usability.

To assess performance, we constructed a test bed modeling an enterprise hub-and-spoke network with five sites: a headquarters in Boston and branch offices in Portsmouth, N.H.; Newton, Mass.; El Segundo, Calif.; and San Francisco.

We used a Spirent Converged Network Impairment Emulator to emulate WAN rates and delays. The Newton and San Francisco remote links ran at T1 (1.5-Mbit/s) rates, while the other two ran at T3 (45-Mbit/s) rates.

The Newton and Portsmouth links used 16-millisecond round-trip delays, while the other two used 100-millisecond round-trip delays, representing all permutations of low and high bandwidth and delay.

We measured application acceleration performance with CIFS/SMB Windows file transfers, Outlook/Exchange, HTTP and SSL traffic. In a separate test, we assessed devices' QOS capabilities by generating VoIP traffic while simultaneously loading the network with HTTP traffic. We developed custom software for this project to generate CIFS/SMB and Outlook/Exchange traffic.

To measure bandwidth reduction, we used a ClearSight hardware-based analyzer with taps in both the Boston LAN and WAN sides of the test bed. To measure application-response time, our custom software measured CIFS and MAPI transfers.

For the CIFS file transfers, two clients at each remote site simultaneously sent and received Microsoft Word documents from the Boston site. Clients on T3 links transferred 750 files in each direction, while clients on T1 links transferred 25 files each way. We ran each CIFS test three times: a "cold" run with empty device data stores, a "warm" run once the data store had been populated and a "modified" run in which we altered the contents of 10% of the files.

To measure MAPI/Exchange perfor-

mance, Outlook 2007 on each client created 1,000 or 34 messages for T3 or T1 circuits, respectively. Some of the messages had Microsoft Word files as attachments, and some were simple text. Each client sent messages to all other clients, but always through an Exchange 2003 server at the Boston site.

To measure HTTP performance, we configured the Spirent Avalanche and Reflector 2500 appliances to emulate Web clients and servers, respectively. As many as 2,048 clients at remote sites requested 11KB objects from servers at the Boston site. We measured HTTP response time and transfer rates. We repeated these tests twice, once with 256 clients across all remote sites, and again with 2,048 clients.

To measure SSL performance, we repeated the HTTP tests using HTTPS, loading server certificates on the acceleration devices where they support SSL proxying.

To assess devices' QOS capabilities, we simultaneously offered small amounts of VoIP and large amounts of HTTP traffic. To generate and measure VoIP traffic, we used GL Communications' PacketGen and VQT products to set up and measure SIP/RTP calls.

We again used Spirent Avalanche/Reflector for HTTP traffic. In these tests, we compared VoIP audio-quality measurements with and without HTTP present. As an added test of QOS functionality, we also checked whether devices could classify and remark the Diff-Serv code points (DSCP) for voice and Web traffic.

We also measured the maximum TCP connection capacity of the Boston device. In this test, the Avalanche appliance emulated a Web client requesting a 1KB object every 60 seconds. Because Web clients use HTTP 1.1, this test establishes a large number of TCP connections over time. We attempted to measure a maximum connection count supported by the Boston appliance to the nearest 1,000 connections.



Much of our assessment for functionality, manageability and usability occurred in the course of running the performance tests. For example, most vendors in this

test issued one or more software upgrades in the course of this project. We compared vendors' management tools in terms of ease of use for pushing out a new software image to multiple remote sites.

In the area of device functionality, we attempted to describe device taxonomy. The devices we tested work in different ways: The network design and functional differences included whether devices sit inline, whether they tunnel traffic or send traffic transparently, whether devices compress traffic in-flight or support multiple connections (and possibly load-sharing) between sites and offer high-availability features.

To assess device manageability, we performed common administration tasks and rated each system on its ability to perform these tasks. In addition to the aforementioned software upgrade, we attempted to make configuration changes from a central management platform and then push the change out to all appliances with a single operation.

We subjectively assessed each vendor's ability to present a coherent view of all devices from a central location. We looked for the ability to define multiple administrative roles, where different classes of managers had different privilege levels. We compared devices' ability to present aggregated and individual device logs on a central system. We looked for support for management via command-line interface as well as through a centralized Web or proprietary user interface.

Our usability assessment was mainly a set of subjective judgments on the ease of accomplishing all the foregoing tasks. We also looked for a few features that make products easier to use, such as autodiscovery of devices and application types.

Sponsored by Silver Peak

www.silver-peak.com

Section 3: Real-world tests and shopping tips for buyers ● ● ●

A buyer's checklist for application acceleration

6 tips on how to pick a WAN-optimization winner

■ *By David Newman*

Faced with big bandwidth bills every month, it's tempting simply to buy the application accelerator with the best performance. Tempting, but not necessarily correct.



Performance matters, but it's far from the only consideration. Numerous other issues should factor into any buying decision, including functionality, network design, security and application support. What follows are six key questions buyers should take into account while considering which application-acceleration system will best suit their own environment.

1. What are my goals for application acceleration? All accelerators reduce the number of bits on the wire, but they do so with different goals.

Most devices focus on WAN bandwidth reduction. That's a worthy goal when links are overloaded and the cost of adding more WAN capacity is an issue. But reducing bandwidth isn't the only thing application-acceleration devices do.

In other situations, enterprises may need to speed bulk data transfers or improve response times for interactive applications. Examples of the former include backups and disaster-recovery processes, both of which require moving a lot of data in a hurry. (Silver Peak, in particular, focuses on speeding high-bandwidth applications.) Examples of the latter include databases and other transaction-processing applications where there's revenue tied to every transaction.

And organizations may have yet other needs for application acceleration beyond bandwidth reduction or faster transfer times. For example, a company that routinely distributes large videos or databases might want to locate data closer to customers using "prepopulation" or "prepositioning" capabilities, intelligent forms of caching that places

frequently requested data on remote-site appliances.

Our advice: Make sure vendors understand your main goal for application acceleration – bandwidth reduction, faster bulk transfers or response-time improvement – and let them pinpoint which of their systems come closest to achieving that goal.

2. What's the difference between caching and application acceleration? Caching – getting data close to the user – is the oldest trick in performance tuning, and it's still a great idea. Application-acceleration devices use caching, but do so in fundamentally different ways than conventional Web caches and their optimization toolkits extend well beyond caching.

Conventional caches work at the file level. That's fine for static content, but it's no help when something changes. Consider a manufacturing company that routinely distributes a 10GB parts database to multiple sites. If just one record changes, caches would need to retrieve the whole database again.

Application-acceleration devices work smarter: They retrieve only the changes. As user data flows through a pair of devices, each one catalogs the blocks of data it sees

and makes an index of those blocks. Note that a "block" is not the same as a file; it's just a fixed amount of data.

The next time users request data, the devices compare their indexes. If nothing changed, the device closest to the user serves up the data. If something's new, the remote device retrieves the changed data, and both devices put new blocks and new indexes into their data stores. Over time, application-acceleration devices build up "dictionaries" that are hundreds of gigabytes or terabytes in size.

Dictionaries have three advantages over conventional caching. First, they require transmission only of changes to an object, not the entire object. Second, they still save bandwidth if an object is changed and then later changed back, because the original data still exists in the dictionaries. Finally, dictionaries are application-agnostic. In contrast, caches typically work only with a single application. All the devices we tested use dictionaries. Blue Coat's devices are also Web caches, while the Cisco devices are CIFS caches.

Acceleration devices perform many other optimizations as well. All compress blocks of data flowing between pairs of devices, with big bandwidth savings shown in our tests. But compression won't help with near-random data patterns, such as video or encrypted data (an exception is when clients repeatedly request the same near-random data).

These devices also tweak TCP and CIFS parameters to speed data transfer. At the TCP level, these devices make use of many high-performance options missing from Microsoft's stack. Some devices do inverse multiplexing of client-side connections, reducing connection setup overhead. The devices we tested also optimize CIFS, Microsoft's infamously chatty file-transfer pro-

Section 3: Real-world tests and shopping tips for buyers ● ● ●

to col. For sites looking to optimize Windows traffic, CIFS-optimization efficiency is a top concern.

Our advice: Make certain vendors are not pushing the notion that application-acceleration devices are “just” file caches; they’re smarter about storing data and employ other optimizations to boot.

3. How do application-acceleration devices operate with the rest of my network? Imagine the effect on the network if an intermediate device were to terminate TCP connections, alter IP addresses and port numbers, and possibly scramble packet payloads. That’s one way of describing exactly what many acceleration devices do. While these effects aren’t always harmful and may be desirable, network transparency may be a concern.

Altering or hiding packet contents can cripple devices that need to see those contents, such as firewalls, bandwidth managers, and QoS-enabled routers. All the devices we tested optionally can be configured to run in a transparent mode, but might lose optimization efficiency in doing so. Of course, if other devices don’t examine traffic contents, this isn’t an issue.

Another design concern is whether devices operate inline or in so-called off-path mode. Cisco and Riverbed devices can be configured for off-path operation, meaning traffic passes up through a separate software module, while the device simply bridges nonoptimized traffic.

All devices tested fall back to passthrough mode if acceleration is disabled, a useful feature in maintaining availability, and all offer failover capabilities. To further enhance availability, the Blue Coat

and Cisco devices also support clustering of multiple application-acceleration devices.

Our advice: Grill vendors on whether or not their product will “blind” other devices, such as firewalls or bandwidth managers, that need to see packet contents.

4. What are the security implications for application acceleration? On the plus side, acceleration devices can improve data privacy by setting up encrypted tunnels between sites. Because these tunnels carry all data (or some user-defined portion of data that’s sensitive), there’s no need to set up authentication and encryption on a per-application basis.

But these devices also keep copies of all user data, creating disclosure concerns and possible compliance issues for industries that require end-to-end encryption. Network managers will need to revise security policies to cover data stored on acceleration devices, not only while it’s in use but when it’s retired (to ensure its disks are wiped clean before disposal or recycling). The Cisco and Silver Peak devices have an elegant solution: They encrypt data on disk, rendering it useless to an attacker.

Our advice: Push potential vendors to explain how you could revise security policies as appropriate to deal with use and disposal of sensitive data stored on their application-acceleration devices.

5. What’s my application mix? Acceleration-device vendors differ in terms of the number and type of applications they can optimize.

Not all application-acceleration devices optimize UDP-based traffic, including the Blue Coat appliances in our tests. Given that

voice, video, file sharing and some backup traffic may use UDP, support for UDP-based applications will likely become more important over time.

For many enterprises, the mission-critical, revenue-bearing application is something developed in-house. Even the longest list of supported standard applications won’t help here, but even so the application may still be a good candidate for TCP or other optimizations. Testing support for custom applications is critical in such situations.

Our advice: Force any potential vendor to address how its product will directly address the prospect of speeding up your organization’s application mix.

6. Where are my users? Most acceleration today is done between sites, with a symmetrical pair of devices at either end of a WAN link. However, some client software is beginning to become available for road warriors and telecommuters.

Blue Coat recently released client software that performs most, but not all, the same functions of its hardware appliances. The client code does caching, compression, L4/L7 optimization, but it doesn’t perform WAN-data reduction. Riverbed also has announced an acceleration client, and other vendors are likely to follow.

Our advice: If you need to speed traffic going out to a mobile workforce, press vendors about their plans to provide application-acceleration clients as well as site-to-site appliances.

David Newman is president of Network Test, an independent test lab in Westlake Village, Calif. He can be reached at dnewman@networktest.com.

Section 3: Real-world tests and shopping tips for buyers ● ● ●

Application acceleration: Making sense of a crowded technology market

■ *By David Newman*

Confused about application acceleration? You've got company.

Dozens of vendors have entered this hot area, using another dozen or so techniques to reduce response time, cut bandwidth consumption, or both. As with any market where multiple sellers all speak at once, it's easy to get lost amid the claims and counterclaims. It's harder still when the wares for sale are new and unfamiliar to many buyers.

As always, education is key. This article describes the major types of acceleration devices; introduces the players; explains the workings of acceleration mechanisms; and looks into what the future holds for this technology.

Application acceleration products generally fall into one of two groups: Data center devices and symmetrical appliances that sit on either end of a WAN link. A third category, acceleration client software, is emerging, but it is in relatively early stages.

Application acceleration may be a relatively new market niche, but the technology behind it has been around for some time. For close to a decade, companies such as Allot Communications and Packeteer have sold bandwidth optimization appliances that prioritize key applications and optimize TCP performance (Packeteer also offers a symmetrical WAN device.) Other acceleration technologies such as caches, compression devices, and server load balancers have been around even longer. For the most part, though, the application acceleration market today is split between data-center and WAN-based devices.

The two device types differ not just in their location in the network but also in the problems they address and the mechanisms they use to solve these problems.

Into the data center

Data centers have high-speed pipes and numerous servers. Some also have multi-tiered designs, with Web servers arrayed in front of

application and database servers. In this context, improving performance means reducing WAN bandwidth usage for out going and incoming traffic and offloading TCP overhead and/or SSL overhead or eliminating servers.

Prominent vendors of data-center acceleration devices include Array Networks, Cisco Systems, Citrix Systems, Coyote Point Systems, Crescendo Networks, F5 Networks, Foundry Networks, and Juniper Networks.

Data-center acceleration devices use a variety of mechanisms to achieve these ends. Weapons in their acceleration arsenal include TCP connection multiplexing, HTTP compression, caching, content load balancing, and SSL offload. Though more of a security measure than a performance feature, some data-center accelerators also rewrite content on the fly.

Of these mechanisms, connection multiplexing and HTTP compression do the most to reduce WAN bandwidth usage. Connection multiplexing is helpful when server farms field requests from large numbers of users. Even with load balancers in place, TCP connection overhead can be very significant. Acceleration devices lighten the load by multiplexing a large number of client-side connections onto a much smaller number of serverside connections. Previous test results show reductions of 50:1 or higher are possible.

Note that 50:1 multiplexing doesn't translate into a 50-fold reduction in servers. Other factors such as server CPU and memory utilization come into play. Still, multiplexing can lower overhead and speed content delivery.

As its name suggests, HTTP compression puts the squeeze on Web payloads. Most Web browsers can decompress content; usually the stumbling block is on the server side, where compression is often disabled to reduce delay and save CPU cycles. By offloading this function off the servers and onto the acceleration devices make it feasible to do compression.

Obviously, results vary depending on the compressibility of content. Since most sites serve up a mix of compressible text and

uncompressible images, HTTP compression offers at least some bandwidth reduction, and may even be able to reduce the number of Web servers needed. One caveat: Compression won't help at all with seemingly random data streams, such as encrypted SSL traffic, and could even hurt performance.

The remaining data-center application acceleration mechanisms help lighten the load on servers. Caching is one of the oldest tricks in the book. The acceleration device acts as a "reverse proxy," caching oft-requested objects and eliminating the need to retrieve them from origin servers every time. Caching can deliver very real performance gains, but use it with care: Real-time content such as stock quotes must never be cached. Object caching also won't help when a small part of a large object changes, for example when a single byte in a large document is deleted.

Content load-balancing is conceptually similar to previous generations of layer-4 load balancing, but in this case the decision about where to send each request is based on layer-7 criteria. For example, devices run SQL queries and other "health checks" on back-end databases to decide which server will provide the lowest response time.

SSL offload also helps speed delivery of secure communications. In some cases, acceleration devices act as SSL proxies; the encrypted tunnel ends on the acceleration appliance, with cleartext traffic flowing between it and the origin servers. This frees up the server from computationally expensive SSL encryption, and in many cases it can dramatically reduce server count in the data center. It's also possible to achieve end-to-end encryption through proxying; the acceleration device terminates a client's SSL session and then begins a new session with the server. Some performance gain is still possible through TCP multiplexing.

Because data-center acceleration devices are application-aware, they have the added capability of being able to rewrite URLs or

Section 3: Real-world tests and shopping tips for buyers ● ● ●

even traffic contents on the fly. Citrix recently announced the ability to replace credit card numbers in data streams with Xs instead, preventing theft by interception. Similarly, it's possible to rewrite URLs, either to make them shorter or more recognizable or to hide possible security vulnerabilities. On this latter point, an attacker may be less likely to probe for Microsoft Active Server Page vulnerabilities if a URL ending in ".asp" gets rewritten to end with ".html".

Going Symmetrical

For many enterprises, the biggest bang for the acceleration buck comes not in the data center, but on the dozens or hundreds of WAN circuits linking remote sites to data centers. A recent Nemertes Research survey found that monthly WAN fees alone account, on average, for 31% of total enterprise IT spending. In that context, even a small performance improvement can mean big savings.

That's not to suggest that symmetrical WAN devices provide small improvements. Recent Network World test results show WAN bandwidth reduction of up to 80 times (not 80 percent) and 20- to 40-times improvements in file-transfer rates. Considering the huge bite of the IT budget that WAN circuits take every month, symmetrical WAN acceleration devices are very much worth considering.

The technology certainly has gotten vendors' attention, with numerous companies offering this type of acceleration device. Players in this crowded field include Blue Coat Systems, Cisco Systems, Citrix Systems, Exinda Networks, Juniper Networks, Riverbed Technology, Silver Peak Systems, and Streamcore.

All these vendors offer appliances and/or acceleration modules large and small, with size depending on WAN link capacity and the number of connected sites and users. Devices generally include disks for caching (though caching may have a different meaning than the caching capability of data-center devices; more on that later). All seek to address the number one bottleneck in enterprise WAN traffic: the sluggish performance of the Microsoft Windows TCP/IP stack across the WAN.

Beyond those common capabilities, these devices may offer at least some of the following mechanisms to reduce WAN bandwidth usage or to speed data transfer: application- and transport-layer optimizations;

pre-positioning (a method of storing content closer to users); data compression; read-ahead/write-behind methods; and protocol prioritization.

Application-layer awareness is the most potent WAN acceleration technique. All vendors in this area can optimize the two most common application-layer protocols in enterprise networks - CIFS (common Internet file system), used in Windows file transfers, and MAPI (messaging application program interface), used by Exchange email servers and Outlook clients.

Because CIFS is notoriously chatty, it's a terrible performer in the WAN. Even a simple operation like opening a directory and listing files can involve the transfer of hundreds or even thousands of CIFS messages, each one adding delay. Acceleration devices streamline and reduce CIFS chatter using a variety of proprietary techniques. The results are impressive: CIFS performance in Network World tests of four offerings in this space was 30 to 40 times faster than a baseline test without acceleration.

All vendors can optimize CIFS, MAPI, and other popular applications such as HTTP, but there's a considerable amount of specsmanship about how many applications are supported beyond the basics. Some vendors' data sheets claim to optimize more than 100 different applications, but often this means simply classifying traffic by TCP or UDP port number, and not necessarily doing anything specific with application-layer headers or payloads. Network managers are well advised to quiz prospective vendors on what specific optimizations acceleration devices offer for their organization's particular application mix.

Pre-positioning, another big bandwidth saver, is essentially an automated form of caching. Say a large electronics distributor regularly distributes a 75-Mbyte parts catalog to all 15,000 of its employees. Rather than have employees retrieve the catalog from headquarters over and over again, a better option is to load the presentation locally at each remote site's acceleration device, and then distribute it locally. Most caches can do that, but pre-positioning goes further by automating the catalog's distribution to all acceleration devices at remote sites. Especially for organizations with many large sites, the bandwidth savings can be very substantial.

Caching can take two forms: object caching,

as discussed previously in our data center discussion and byte caching (called "network memory" by some vendors). With byte caching, each appliance inspects and caches the stream of data going by, and creates an index for each block of data it sees. The index may contain some form of hash uniquely identifying that block. The first time a device forwards data, the byte cache will be empty. On each successive transfer, the pair of devices won't transfer the data again; instead, it just sends the indexes, in effect saying "just send block X that you already have stored in your cache."

Byte caching has two benefits. First, like object caching, it greatly reduces the amount of data traversing the WAN. Second, unlike object caching, it chops the byte stream into relatively small blocks rather than dealing with potentially huge objects. If only a small part of a very large file changes, the acceleration device just sends the updated data, not the whole object. Some devices, such as those from Blue Coat and Cisco, employ both forms of caching (in Cisco's case, for Windows file traffic only. Others such as those from Riverbed and Silver Peak rely on byte caching alone.

WAN acceleration devices also use data compression to reduce WAN bandwidth usage. This isn't just the HTTP compression seen in data-center devices; instead, symmetrical WAN devices compress entire payloads of all packets, regardless of application. Compression works best for data streams comprised mainly of text or other repetitive data; for near-random byte patterns (such as images or encrypted data), it's not much help.

Cisco's WAAS acceleration devices use "read-ahead/write-behind" techniques to speed up file transfers. While these techniques aren't new (server and PC designers have employed them for years), they can speed file transfers. Both techniques take advantage of the fact that enterprise data tends to be repetitive. Over time, devices can predict that if a user requests block A of data, then a request for blocks B and C are likely to follow. With that knowledge, the device can line up the next blocks and serve them out of memory instead of a much slower retrieval from disk. And speaking of disk operations, it takes a relatively long time to write data to a disk. Write-behind operation defers write requests until several have accumulated and then does them all at once. From the user's

perspective, read-ahead and write-behind both translate into faster response times.

Many acceleration devices (with the notable exception of Cisco's) also use various QoS mechanisms to prioritize key applications or flows during periods of congestion. Cisco also has a prioritization story, but it involves communication with routers, which then perform the actual queuing. For enterprises that already have enabled QoS features on their routers, this is a useful approach; for others just getting started with QoS it may make sense to consider using the acceleration device for queuing. As with application support, there is considerable variation among products as to which types of traffic acceleration devices can prioritize.

Future trends

Client software, security, and device consolidation are likely to be the next major trends in application acceleration. Acceleration client software already is available from Blue Coat and others have clients in development. These software packages give PC-toting road warriors and telecommuters some if not all the techniques used in acceleration appliances.

Security is another hot-button issue, with acceleration vendors adding support for SSL optimization (supported by Blue Coat and Riverbed in the recent Network World test, with plans announced by Cisco and Silver Peak). Cisco and Silver Peak devices also encrypt all user data stored on appliances, a key consideration for regulatory compliance in some industries.

If past history is any guide, it's also likely switch and router vendors will fold at least some acceleration features into their devices. However, the market for standalone devices is highly unlikely to disappear anytime soon. Most switches and routers aren't TCP-aware today, let alone application-aware, and getting there will take time. Moreover, the form factors and component costs of acceleration devices (many have beefy CPUs, memory, and disks) argue against rapid consolidation, especially into low-end branch office switches and routers. For fully featured acceleration, standalone pairs of devices are likely to be the platform of choice for at least a few more years.

Section 3: Real-world tests and shopping tips for buyers ● ● ●

Speed safely: application acceleration best practices

■ *By David Newman*

Five times, 10 times, 20 times or more: The performance benefits from application acceleration are real, provided you understand what the technology can and can't do for your network. What follows are selected best practices for deploying application acceleration devices in enterprise networks.

Define your goals.

Application acceleration takes many different forms. There's no one definition for "making an application go faster."

For some users, reducing WAN bandwidth consumption and cutting monthly circuit costs may be the key goals. For others, it's speeding bulk data transfer, such as in backup, replication, or disaster recovery scenarios. For yet others, improving response times for interactive applications is most important, especially if those transaction-based applications carry an organization's revenue.

Deciding where to deploy application acceleration is also a consideration. Different types of acceleration devices work in the data center; in pairs with devices deployed on either end of a WAN link; and, increasingly, as client software installed on telecommuters' or road warriors' machines. Identifying the biggest bottlenecks in your network will help you decide which parts of your network can benefit most from application acceleration.

It's also worth considering whether application acceleration can complement other enterprise IT initiatives. For example, many organizations already have server consolidation plans under way, moving many remote servers into centralized data centers. Symmetrical WAN-link application acceleration

devices can help here by reducing response time and WAN bandwidth usage, and giving remote users LAN-like performance. In a similar vein, application acceleration may help enterprise VoIP or video rollouts by prioritizing key flows and keeping latency and jitter low.

Classify before you accelerate.

Many acceleration vendors recommend initially deploying their products in "pass-through" mode, meaning devices can see and classify traffic but they don't accelerate it. This can be an eye-opening experience for network managers.

The adage "you can't manage what you can't see" definitely applies here. It's fairly common for enterprises to deploy acceleration devices with the goal of improving performance of two to three key protocols - only to discover your network actually carries five or six other types of traffic that would also benefit from acceleration. On the downside, it's unfortunately also all too common to find applications you didn't realize existed on your network.

The reporting tools of acceleration devices can help here. Most devices show which applications are most common in the LAN and WAN, and many present the data in pie charts or graphs that easily can be understood by non-technical management. Many devices also report on LAN and WAN bandwidth consumption per application, and in some cases per flow.

Understanding existing traffic patterns is critical before enabling acceleration. Obtaining a baseline is a mandatory first step in measuring performance improvements from application acceleration.

For products that do some form of caching, a corollary to classification is understanding the size of the data set. Many acceleration devices have object or byte caches, or both, often with terabytes of

storage capacity. Caching can deliver huge performance benefits, provided data actually gets served from a cache. If you regularly move, say, 3 Tbytes of repetitive data between sites and but your acceleration devices have only 1 Tbyte of cache capacity, then obviously caching is of only limited benefit. Here again, measuring traffic before enabling acceleration is key.

Even without acceleration devices deployed, it's still possible (and highly recommended) to measure application performance. Tools such as Cisco NetFlow or the IETF's open sFlow standard are widely implemented on routers, switches, and firewalls; many network management systems also classify application types.

Choose between in-line or off-path.

If forced to choose between high availability and high performance (even really high performance), network architects inevitably opt for better availability. This is understandable - networks don't go very fast when they're down - and it has implications when deciding which acceleration device type to select.

WAN acceleration devices use one of two designs: in-line and off-path. An in-line device forwards traffic between interfaces, same as a switch or router would, optimizing traffic before forwarding it. An off-path device may also forward traffic between interfaces or it may simply receive traffic from some other device like a router, but in either case it sends traffic through a separate module for optimization. Because this module does not sit in the network path, it can be taken in and out of service without disrupting traffic flow.

There's no one right answer to which design is better. For sites that put a premium on the highest possible uptime, off-path operation is preferable. On the other hand,

Section 3: Real-world tests and shopping tips for buyers ● ● ●

there may be a higher delay introduced by passing traffic to and from an off-path module. The extra delay may or may not be significant, depending on the application. If minimal delay is a key requirement, in-line operation is preferable.

Some devices combine both modes; for example, Cisco's WAAS appliances perform off-path optimization of Windows file traffic but use in-line mode to speed up other applications.

Note that "pass-through" operation is different than in-line or off-path mode. In case of power loss, virtually all acceleration devices will go into pass-through mode and simply bridge traffic between interfaces. Devices in pass-through mode won't optimize traffic, but then again they won't cause network downtime either.

Choose between transparent and tunneled traffic.

One of the most contentious debates in WAN application acceleration is whether to set up encrypted tunnels between pairs of devices or whether traffic should remain visible to all other devices along the WAN path. The answer depends upon what other network devices, if any, need to inspect traffic between pairs of WAN acceleration boxes.

Some vendors claim tunneling as a security benefit because traffic can be authenticated, encrypted, and protected from alteration in flight. That's true as far as it goes, but encrypted traffic can't be inspected - and that could be a problem for any firewalls, bandwidth managers, QoS-enabled routers or other devices that sit between pairs of acceleration devices. If traffic transparency is an issue, then acceleration without tunneling is the way to go.

On the other hand, transparency is a requirement only if traffic actually requires inspection between

pairs of WAN acceleration devices. If you don't have firewalls or other content-inspecting devices sitting in the acceleration path, this is a nonissue.

Know your limits.

Application acceleration is a worthy addition to the networking arsenal, but it's not a silver bullet. It's important to distinguish between problems that acceleration can and can't solve.

For example, acceleration won't help WAN circuits already suffering from high packet loss. While the technology certainly can help in keeping congested WAN circuits from becoming even more overloaded, a far better approach here would be to address the root causes of packet loss before rolling out acceleration devices.

Further, not all protocols are good candidates for acceleration. Some devices don't accelerate UDP-based traffic such as NFS (network file system) or multimedia. And even devices that do optimize UDP may not handle VoIP based on SIP (session initiation protocol) due to that protocol's use of ephemeral port numbers (this problem isn't limited to acceleration devices; some firewalls also don't deal with SIP). SSL is another protocol with limited support; in a recent Network World test only two of four vendors' products sped up SSL traffic.

Despite these limitations, application acceleration is still a technology very much worth considering. The performance benefits and cost savings can be significant, even taking into account the few caveats given here. Properly implemented, application acceleration can cut big bandwidth bills while simultaneously improving application performance.

WAN OPTIMIZATION

Section 4 Lessons from the field

User plugs in an appliance, reduces WAN traffic by 60%

Faster remote file-access and swifter Web surfing are two big payoffs

■ *By Beth Schultz*

Sometimes the easiest projects are among the most gratifying – in the case of Thomas Weisel Partners and its WAN-optimization technology deployment, instantly so. With little effort and minimal cost, the network engineering team boosted WAN performance and turned user complaints into compliments, within a matter of minutes.

With WAN-optimization appliances from Riverbed Technology, Thomas Weisel Partners has reduced the amount of data traversing the network by about 60%, says Kevin Fiore, director of engineering services at the San Francisco-based investment bank. Of the 23TB of data requested each month by remote users, about 9TB has to be delivered over the WAN, he says. As a result, remote file-sharing has improved, backup times have decreased and the network supports an increasing amount of real-time data, such as VoIP and video, without pause.

Thomas Weisel Partners taps into numerous capabilities of the Riverbed Steelhead appliances and the Riverbed Optimization System software to achieve that 60% optimization rate. These include packet compression, TCP optimization, application acceleration and data reduction. For the latter, the operating-system software analyzes and stores large traffic patterns locally, eliminating trips across the WAN. Any type of application can tap into the data.

With the appliances, “WAN utilization is more efficient, redundant WAN data is no longer sent, and more data is being sent per payload on each TCP round trip,” Fiore says.



GARY LAUGHMAN

“The benefit of WAN optimization has been nothing short of remarkable — honestly.”

KEVIAN FIORE, director of engineering services, Thomas Weisel Partners

Sponsored by Silver Peak
www.silver-peak.com

Section 4: Lessons from the field ● ● ●

The CIFS nightmare

At Thomas Weisel Partners, remote Windows Common Internet File System (CIFS)-based file-sharing accounts for a good chunk – 40% – of the WAN traffic benefiting from the appliances, Fiore says. The optimization of CIFS has been a godsend to users in New York, who had expressed extreme dissatisfaction with how much time it took to work with data sitting in San Francisco. Fiore says he discovered the appliances from Riverbed as he searched for ways to address the slowness associated with the company's file-sharing environment.

"Associates would be trying to open up financial models, presentations or other relatively large files, and it would take a while to do so over the WAN ... because of latency," Fiore says. (CIFS, because it relies on a large number of back-and-forth transactions to complete a request, is known for performing poorly over high-latency WAN links.)

Even worse for users, Fiore says, was the prolonged amount of time it took Windows Explorer to display the directory structure. "One of the main complaints was that they spent more time trying to find a file than they did actually opening the file," he says. "Mentally, they could handle waiting 15 seconds or however long for the file to open, because they knew it was opening. But not even being able to start that process while the directory structure opened – that annoyed them."

With productivity at stake, finding a fix

was a high priority. Fiore's team members attacked the problem in many ways. They tried applying search technology to circumvent using the directory, for instance. And they worked with Microsoft to tweak operating system and server configurations. "Nothing we tried gave us enough satisfaction," he says.

Optimization appliances were a logical next step, but at the time (late 2005), Fiore says he found many of the options too limiting. He wanted to be able to optimize other data types, such as HTTP, Microsoft Exchange Messaging API and Oracle database logs; and the products were too focused on optimizing the Microsoft CIFS process alone, he says.

Riverbed turned out to be the exception.

Fiore's team began evaluating the Steelhead appliances one night, installing one at each end of the WAN link between New York and San Francisco. Within about 20 minutes of uncrating the devices, they were up and running. The next morning, New Yorkers noticed the improvement at once. "By the time I got in my office at 8 o'clock Pacific, e-mails were circulating from key business users thanking IT for solving their problem," Fiore recalls. He cut a purchase order for the appliances that same morning.

Since then, Fiore has added Steelhead appliances to optimize traffic between San Francisco and the company's other offices, in Boston; Portland, Ore.; and Mumbai, India. In total, WAN optimization has cost about \$200,000, including maintenance.

Across all those WAN links, Fiore reports reducing the amount of CIFS traffic for April by about 59%.

Optimization by association

True to Fiore's expectations, CIFS is far from the only data type benefiting from the appliances. April statistics show significant optimization percentages for a wide variety of other data, he says. For example, Oracle data (log shipping for databases) is 88% optimized; HTTP data (for CRM and other internal Web applications), 85%; EMC Centerra-based e-mail archiving repository data, 84%; and replicated storage-area network data, 84%.

There have been unexpected benefits, too – "little weird things we've caught over time," Fiore says. Web surfing is an example. At Thomas Weisel Partners, a Web-monitoring application sits in San Francisco. That means that the local router – for instance, in New York – must communicate back and forth over the WAN to request permission to allow access to a particular site. With optimization, that yes/no process is much faster. "It looks like we improved the Internet connection, but we didn't. We just improved the lookup for approvals," Fiore says.

"The reality is, Riverbed is doing a lot more for us than solving the problem we were trying to solve," he says. "That's pretty cool, given that a more typical IT scenario is buying custom solutions that only solve specific problems," he adds. "So this has been nothing short of remarkable – honestly."

Section 4: Lessons from the field ●●●

Gitmo gets high-bandwidth makeover

U.S. Naval base at Guantanamo Bay reworks its network

■ By Tim Greene

The U.S. Naval base at Guantanamo Bay, Cuba, notorious as the prison for enemy combatants, poses a lot of other thorny problems, including a copper communications infrastructure extremely unfriendly to high-speed networking.

As a result, Kenn Devine, manager of BRSC Information Systems, which is part of Burns and Roe Services, has struggled to create an effective WAN to tie its 17 Guantanamo sites to a local data center and link that back to the company's facility in Virginia Beach, Va.

BRSC has held contracts to supply a range of services to the military at Guantanamo Bay since the 1950s, including providing drinking water and electricity, shuttling barges across the bay, running the airport, making concrete and providing housing.

The firm's offices - scattered around the 90-square-mile base, which spans both sides of the bay - share business process applications supported at the data center that each project office ties into via copper phone lines. The data center is tied back to the mainland, and carries mainly HTTP, HTTPS, Citrix and some Microsoft CIFS traffic.

Before Devine could work on the WAN problems, he had to upgrade the network infrastructure - a series of cascading hubs - to a 100Mbps HP ProCurve switched infrastructure supported by Cisco routers and VPNs. To squeeze the most out of the links between the sites and the data center and between Guantanamo Bay and Virginia Beach, Devine installed nine Expand Accelerator WAN-optimization appliances.

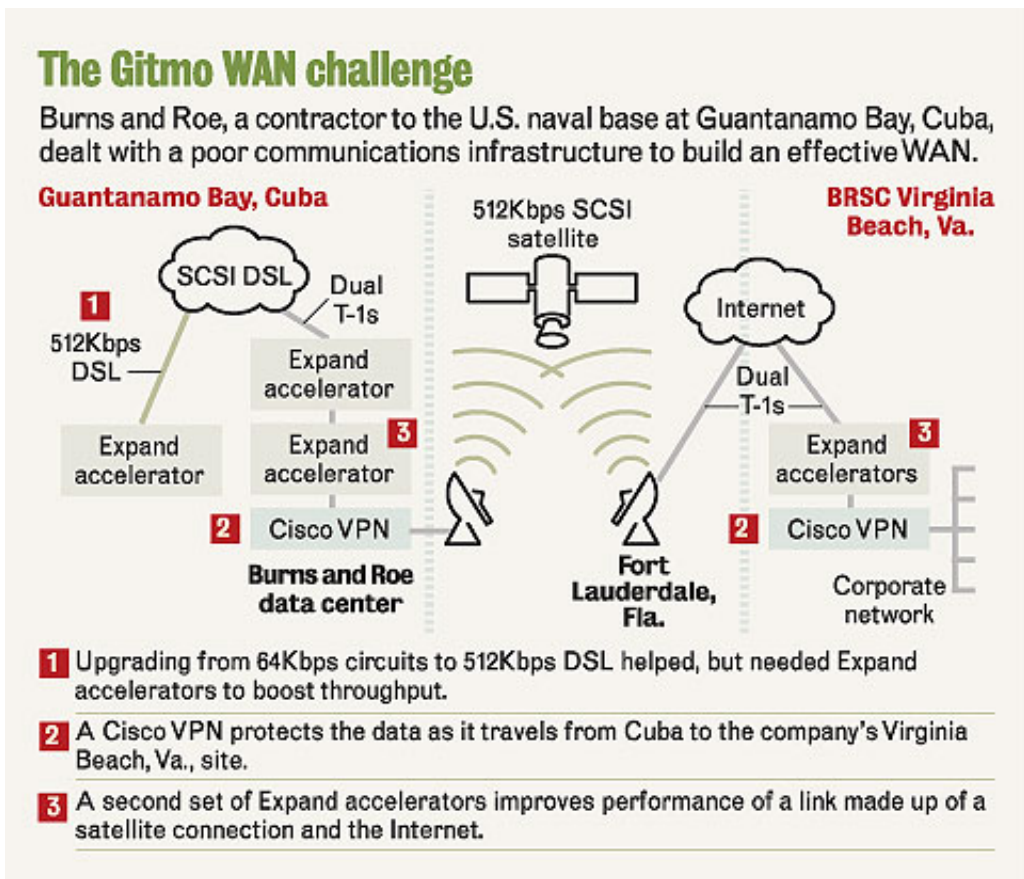
He says the network now works efficiently, and he is in preparing to swap out most of the antiquated workstations for locked-down Citrix thin-client machines to give him cen-

tralized control. That step is on hold until he can upgrade application servers in the data center, probably early next year.

The project isn't saving any money, but it is giving the firm a faster, more reliable and more efficient network that over time will bring down management and maintenance costs, Devine says.

After replacing the hubs, the first move was upgrading the 64Kbps connections between the remote sites around the base and the data center. The base provides two options for connecting the remote sites to the network operations center: Phoenix Cable, the base cable TV network, or the phone network run by SCSI. Wireless is out because the military bans use of unclassified bandwidth on the base.

Cable would be the less expensive option, but reliability is poor,



Section 4: Lessons from the field ● ● ●

Devine says. "It's not as fast as you get with your cable connection at home," he says.

So he chose DSL, which gave him a leap to 512Kbps links between the sites and SCSI. Each link costs \$250 per month, plus \$550 per month for dual T-1 trunks that feed traffic from the SCSI network to the data center.

The only option for connecting to the mainland is satellite service also operated by SCSI - \$10,000 per month for a 512Kbps hop to Ft. Lauderdale, Fla. Traffic is routed from there over the Internet to the BRSC Virginia Beach facility that taps into the Internet via dual T-1s provided by Cox Communications, Devine says.

Because of the limited copper infrastructure at the base and because the copper is in poor shape, 512Kbps is the top speed BRSC can afford. But six of the sites have significant amounts of traffic that warrant larger connections, he says. So he looked around for WAN-acceleration gear that could make the 512Kbps connections behave as if they were larger.

He considered equipment from Peribit, now part of Juniper Networks, because it supported caching to speed up transfers. By the time the network was upgraded, though, Expand also had caching, and it made appliances for smaller connections for about 30% less than Juniper equipment would cost, he says. "Expand was designed more for a trickle," he says.

The Expand boxes make the 512Kbps links seem like 2Mbps pipes, with some spikes in throughput where they behave like 11Mbps links, he says.

BRSC had trouble getting the data center Expand box to take in all the branch traffic and feed it directly onto the satellite, so it installed a second Expand box there. Traffic flows from one Expand Accelerator to the other via a Layer 3 ProCurve gigabit switch, into an IPSec VPN tunnel created by Cisco ASA5510 router and over the satellite link.

Both the VPN tunnel and the Expand optimization tunnel are terminated in Virginia Beach, where traffic bound for the Internet is decrypted and dropped onto dual T-1 Internet access lines.

Devine says that with the company's WAN and LAN settled, he is focusing on better management of endpoints by locking them down and tightening up on who has administrative access. Using Citrix thin clients will allow him to centralize most applications at the Guantanamo Bay data center, making maintenance simpler.

Section 4: Lessons from the field ●●●

Transportation authority speeds up WAN with Silver Peak

WAN optimizing gear drops transaction times and makes applications respond faster.

■ *By Tim Greene*

Sound Transit in Seattle unclogged slow WAN links that hindered railway-improvement projects by using WAN optimization gear that made traffic run faster and that cost much less over time than other options.

Now applications on the network perform well, reducing transaction time for a test set of traffic by as much as 69%. And, as an unexpected bonus, the transit authority has also dramatically cut back bandwidth that was unnecessarily dedicated to VoIP thanks to data gleaned from the Silver Peak WAN optimization gear it uses, says Garv Nayyar, senior integration engineer for the transit authority.

The Sound Transit WAN is arranged in a hub-and-spoke configuration, with a headquarters data center that feeds seven other sites via point-to-point T-1 connections.

Workers at the remote sites access a variety of applications including Microsoft SharePoint for file sharing, Opentext Livelink EMC for document management, Primavera Expedition for project management, and Microsoft Exchange for e-mail.

Part of the project calls for workers to send construction photos and CAD drawings over the WAN, transfers that slowed down everything else on the links, Nayyar says. "The problem was if somebody was pulling a 5MB file it just took the site down or it would slow them down for a few seconds, making it so other folks couldn't work effectively," he says.

End users complained about the delays and demanded that something be done. "They wanted us to put either servers at all the local sites or add another T-1 line or possibly up to three T-1 lines to each site," he says.

In addition to those possibilities, he looked into WAN acceleration gear from Cisco, Riverbed and Silver Peak. Cisco kept pushing back the deadline for when its gear would be ready to test and Sound Transit couldn't wait, Nayyar says. He could not get Riverbed gear to work and was told he needed an upgrade to his Cisco switch infrastructure, so he rejected them as well.

He says he got the Silver Peak gear up and running relatively quickly without any network upgrades.

Price was also a factor in his decision. Setting up separate servers in each branch so client-server traffic could be kept local would have cost about \$15,000 per site, he says.

Adding T-1s would have cost \$350 per month per T-1, for a total between \$2,450 and \$7,350 per month extra but the extra bandwidth would be overkill for the average amount of traffic sent over the lines.

The Silver Peak gear cost about \$90,000, he says, making it a cheaper option than placing local servers in all the branches, and less expensive than extra T-1s after three years.

Nayyar says he installed test Silver Peak boxes in about two hours, one at headquarters and one at the remote site that had the loudest complainers. He ran it as a test for two weeks and it improved application response times so much that end users complained when the test was over. "When we took it down the remote site started to send complaints to their senior management saying we need that thing back on because we were getting a lot more work done," he says.

He ran a baseline test of sample traffic over each WAN connection twice before turning up the Silver Peak equipment, and ran the same tests again after it was working. The smallest reduction in time to complete the test was 32% and the largest reduction was 69%, according to the numbers he gathered.

In setting aside 1Mbps of the 1.5Mbps available on the T-1s for data, Sound Transit found that the VoIP traffic on the network suffered. This was because the VoIP vendor had already dedicated 50% of the bandwidth to VoIP, so the voice and data were contending for the same bandwidth.

When the Silver Peak gear finally helped sort the problem out, the bandwidth dedicated to VoIP was cut to 110Kbps, and the voice works fine, Nayyar says.

Using Silver Peak statistical data, Sound Transit has better network monitoring than it had before to determine whether performance on the network meets the promises IT has given to workers about service levels, he says.

Section 4: Lessons from the field ●●●

The human side of net optimization

Making the most of new traffic management gear spurred Ceridian to revamp its IT structure

■ *By Ann Bednarz*

With a new technology implementation, it's easy to focus on the products rather than the staffing and operational resources that go into a project. In reality, it's the proper combination of technology and manpower that makes a rollout successful.

At Ceridian, coordination among different IT groups and changes to the reporting structure enabled the human resources specialist to adopt a new business model. The Minneapolis-based software maker over the last several years has shifted from solely selling shrink-wrapped versions of its HR applications to offering hosted services for applications such as payroll, time and attendance, and labor management.

The need to create network-ready applications that are delivered over the Internet, rather than deployed at customer sites, raised new IT considerations for Ceridian. When customers were responsible for the software, they dealt with issues such as network capacity, performance and data management. When Ceridian adopted a hosted model, those issues became more critical internally. "We went from that being on the customer side of the equation to it being on our side of the equation," says Rodney Bowers, director of data center services at Ceridian.

To ensure solid application performance, Ceridian implemented 3400 series Big-IP appliances from F5 Networks. The F5 gear today handles load balancing among more than 1,900 servers at Ceridian's data centers as well as authentication of SSL traffic.

One of the advantages of using the Big-IP appliances is that it freed development staff – who were under pressure to prepare the hosted offerings for launching – from having to address traffic management and SSL authentication on their own, Bowers says.

"We could take and aggregate TCP/IP sessions, also do the SSL acceleration, and offload that from the application," he says. "To be able to use hardware to offload those two important components was a very beneficial find for us."

At the same time it was deploying the F5 gear, Ceridian tweaked its IT structure to better address the challenges of delivering applications to customers over the Internet. Shifting to an application service provider (ASP) mode was an effort that crossed network, application development, storage, and customer support disciplines, and getting it right required better coordination among these distinct departments, Bowers says.

"The groups were truly siloed. The network guys didn't talk to the storage guys, the storage guys didn't talk to the application guys," he recalls. "We had to get people understanding the other groups' problems."

For example, it was important for network staff to work closely with field service staff who were charged with helping Ceridian customers

prepare their own internal networks for the hosted applications. Likewise, application development teams paired with data center specialists to design distinct Web and database application tiers to better isolate and monitor application performance issues.

Ceridian also needed to address how the new delivery model would impact storage requirements, which meant getting the network and storage teams to work together. "We tuned our storage requirements, we tuned our clustering configuration working with our SAN infrastructure, and then we worked on replication between the two data centers," Bowers says.

In breaking down the organizational barriers that existed between the IT factions, Bowers wound up taking on greater responsibility. At the beginning of the project, he had responsibility for the network teams at Ceridian. As the makeover transpired, it became clear that individual IT groups needed to work more closely with one another, and Bowers assumed responsibility for the company's storage and data center teams as well.

The application development team does not report to Bower, but it's a group his team members work closely with. "We provide those guys with lots of data, and now they're learning more about how to put their applications together for delivery to an external customer," says Bowers.

Bowers worked closely with Ceridian's director of infrastructure and data architecture, Craig Henne, to get the all the different groups talking regularly. "We would get people in the room and facilitate the discussions," he says. "That was a key thing that we established early on – we need everybody's help to make this better."

In addition to resolving technical issues more quickly, planning inclusive meetings gives talented staff a chance to shine, Bowers says.

"My storage guys, for example, know about more than just storage. They're sharp guys in a lot of other ways," Bowers says. "So they get to use their talents for some other problem-solving mechanisms. They get to contribute to the overall goals, and they get to see the bigger picture. When you start to allow people to do that, you'd be amazed at how receptive they are to coming to meetings and talking about opportunities to improve."

Cross-training also is a key to success, according to Bowers. "We've trained all of our primary WAN engineers on all aspects of F5. Day to day, some of them don't even touch the boxes. But I want them to be aware of the functionality and capabilities of it," he says.

Cross-training is important not only for tactical reasons, but also for strategic planning. "There's the support component, but we also want them to be thinking outside the box when we present them with an architectural problem or an infrastructure problem," Bowers says.

Sponsored by Silver Peak

www.silver-peak.com

Section 4: Lessons from the field ●●●

Learning from experience

Liz Claiborne is a veteran when it comes to WAN optimization

■ *By Ann Bednarz*

It's not hard to find companies that are jumping into WAN optimization. Every month vendors announce new customers that have signed on to purchase their gear. What's more rare is finding an old pro like apparel company Liz Claiborne, which has been using traffic-shaping gear from Packeteer for six years.

Joe Yankauskas, IT director with Liz Claiborne, says the company deployed Packeteer's PacketShaper appliances to prioritize and deliver its business-critical applications, including its ERP and retail management applications.

Six years ago, HTTP, CIFS and MAPI traffic was consuming too much bandwidth, imposing performance penalties on essential network traffic, Yankauskas says. "The original impetus was simply to make sure that our mission-critical business applications were delivered to the sites and didn't have conflict with non-mission-critical applications like e-mail and regular HTTP traffic," he recalls.

Prioritizing voice traffic is another key use of the traffic-shapers. Liz Claiborne started deploying VoIP technology three years ago and today it has about 5,000 VoIP phones in use. "Voice requires the highest priority traffic on the network because you can't retransmit voice packets. If you drop a packet, you get the jitter and scratchiness on the phone," Yankauskas says.

Liz Claiborne so far has deployed the PacketShaper appliances at its corporate data center and roughly 40 satellite offices, which are networked via a variety of T1 and DS-3 circuits. The PacketShaper appliances have "now become staple as we roll out new sites, new locations," Yankauskas says.

But while the company's deployment dates back six years, that doesn't mean it's not dynamic.

One thing that is new is Liz Claiborne's in-progress rollout of Packeteer's Report Center software, which will enable the IT department to centrally monitor traffic and perform trend analyses. The software creates a dashboard that consolidates metrics to make it easier for administrators to spot throughput issues and track network behavior, for example.

"At a glance, they can see how each circuit is performing and determine whether or not they have to fine-tune any of the parameters and shaping to give more availability to a particular application," Yankauskas says.

Without Report Center, reviewing metrics requires individually pulling data from each device. "You have to attach to each PacketShaper at each site and pull down the reports," he says.

Yankauskas is hopeful the Report Center software, once fully operational, will help Liz Claiborne more accurately evaluate and predict bandwidth requirements. With Report Center, the IT team will be able to "do careful analysis to determine exactly when we need an upgrade and when we don't need an upgrade," he says.

Silver Peak White Paper





Top 10 Considerations for Scaling a WAN Acceleration Solution

Taking WAN Performance to New Heights

Security and compliance concerns are creating increased pressure to remove data out of branch offices and to implement robust business continuity plans. At the same time, new WAN acceleration technologies have emerged that deliver order of magnitude performance improvements across a wide breadth of applications. These two factors are creating a "perfect storm" in the IT industry, which is elevating WAN acceleration from a tactical fix to a strategic enabler of key business initiatives.

As WAN acceleration becomes more critical to business operations, the average size and scope of deployment is steadily increasing. This creates a variety of questions pertaining to product scalability. What happens when new applications are deployed? How easy and cost effective is it to add new offices? How is performance affected by network load? Do higher capacity WAN links behave the same as lower capacity links?

There are numerous factors that affect the scalability of a WAN acceleration solution, from hardware architecture to ease of use. This paper identifies the top 10, and provides guidelines for overcoming scalability challenges in real-world enterprise environments.

I. WAN Throughput.

Approximately one third of all enterprises have data centers with 45 Mbps or higher WAN connections, fifteen percent of which are over 100 Mbps. Large WAN links are required to connect many remote offices to centralized resources (i.e. to support a large "hub and spoke" architecture), and to support data center-to-data center applications like replication, backup, and disaster recovery.

To accommodate large offices, WAN acceleration appliances require adequate WAN throughput capabilities. A scalable solution can achieve several hundred Mbps of throughput in a single appliance. This is much more cost effective and easier to manage than solutions that require multiple appliances to be clustered within a data center to deliver adequate WAN capacity. In addition, clustering often leads to reduced performance in large networks. When only one appliance is deployed, it is guaranteed that successive requests for data will always go through that device, which makes it easy to detect and eliminate duplicate data across the WAN. When multiple appliances are deployed, successive requests for data



are sometimes handled by different devices, limiting the effectiveness of data reduction and compression technologies.

WAN throughput can vary with acceleration techniques. For example, when only latency mitigation is employed, such as TCP window sizing and selective acknowledgement, it is fairly easy to exceed 600 Mbps of throughput in a single appliance. It is more challenging, however, to achieve high throughput when more advanced techniques are employed, such as data reduction. Enterprises should be aware of this distinction, making their vendors highlight the maximum throughput that can be achieved (with a subset of features) separately from the throughput that is expected when all features are enabled.

It is often useful to perform a “sanity check” on vendor’s throughput claims by dividing WAN capacity by LAN capacity. If, for example, an appliance supports only 1 GB of LAN traffic and the vendor is claiming 600 Mbps of WAN throughput, then that product is delivering less than 2x improvement when fully loaded. That is fine when latency mitigation is all that is required, as discussed above. But, that ratio is not enough to support the massive performance gains that are achieved when doing data reduction across the WAN, which can easily average in the 10-20x range.

2. Flow limits

WAN acceleration appliances may never reach their stated WAN capacity if the maximum number of TCP sessions supported is reached well before the WAN pipe is filled.

For example, a mid sized enterprise with 45 offices and 50 users in each office would generate approximately 22,500 TCP flows (assuming the average user has 10 TCP flows open at any given time.) If all of these offices are connected to a main data center via a 45 Mbps link, one might assume that a single WAN acceleration appliance with 45 Mbps of WAN capacity could be deployed at the head-end to support this entire network. However, this is not always the case. If the appliance has a low TCP flow limit, then WAN capacity is irrelevant when it comes to sizing this network. For example, if only 4,500 TCP sessions are supported simultaneously per appliance, then 5 appliances would actually be required in this data center instead of the 1 previously assumed. That is a significant difference in terms of cost, management, and support – which becomes even more significant as the network grows even larger.

It is important to determine the TCP flow limits of a WAN acceleration solution prior to sizing a deployment. In addition, one must determine what scenarios may affect this number. For example, do flow limits actually decrease when specific features are enabled, such as disconnected file services? In addition, what happens when TCP flow limits are exceeded - is WAN traffic blocked or sent un-accelerated across the WAN? This last question is particularly important when dealing with applications that have long-lived connections, such as CIFS and email. If a user logs on during a busy period and their connection is sent across the WAN un-accelerated, will the session remain un-accelerated for the lifetime of the connection – even if TCP flow counts drop to acceptable levels?

Not surprisingly, a product may demo great in a lab trial where flows are kept to a minimum, but fall over when deployed in a large network. To ensure that there are no surprises during full-scale deployment, one must take measures to test TCP flow limits from the onset of a WAN acceleration project.

3. Performance under load

Performance can vary based upon the number of users and/or applications deployed in a network. If a WAN acceleration appliance does not have enough memory or disk space to support large environments, performance will start to degrade as the network grows. This often starts to occur well before stated throughput and flow limits are reached, and worsens as the network continues to increase in size.

The right hardware platform can help to ensure better performance under load. A 64 bit architecture, for example, will have more memory and disk space than older platforms, making it better suited for large environment. By using multi processor, multi-core processors, appliances can achieve higher performance with increased memory addressability. In addition, while native 64-bit architectures enable physical and virtual memory to be directly addressed, 32 bit architectures can only address 4 GB of physical memory directly. This requires more complex memory de-referencing, which is substantially slower under heavy load.

4. Hardware acceleration for encryption

The last thing that enterprises want to do is sacrifice security for the sake of application performance when they replace branch office servers with new WAN acceleration

appliances. To overcome this issue, many vendors will employ encryption on their WAN acceleration devices. Encryption of local drives can protect data at rest, while technologies like IPsec can use encryption to protect data sent across the WAN. To perform data reduction and compression techniques on SSL traffic, the WAN appliance must securely become part of the trusted security domain, decrypt the SSL streams, optimize the traffic, then re-encrypt the traffic. This requires fast authentication, and high speed encryption and decryption. Often these techniques are used together. For instance it is very dangerous to store decrypted SSL content in-the-clear on a disk. So encrypting data at rest is a co-requisite for secure SSL optimization.

Encryption is very computationally intensive, which can have an adverse impact on the performance and scalability of a WAN acceleration appliance. Even if the WAN link is only a few Mbps (and within reach of software encryption), peak LAN bandwidth can be greater than 100 Mbps. In addition, WAN acceleration appliances often require several hundred Mbps of bandwidth to perform disk read and writes. As all of this traffic needs to be encrypted and decrypted, dedicated, multi-Gbps security hardware co-processors are required to offload encryption functionality.

This is the only way to ensure maximum throughput when encrypting network traffic, enabling a WAN acceleration solution to grow in a secure and reliable fashion. Even small branch office appliances can benefit from hardware acceleration for encryption to handle SSL acceleration and encryption of local data stores.

It is generally good practice to see how performance changes when all encryption services are enabled. This is the only way to ensure that there is adequate processing capacity with a WAN acceleration solution to perform security with sacrificing scalability and performance.

5. Breadth of Application Supported

The business case for WAN acceleration is often predicated on the number of applications being accelerated. When more applications are supported by a WAN acceleration solution, enterprises experience larger productivity gains. This makes it easier to justify investments in new WAN acceleration hardware and software.

The most scalable WAN acceleration solutions are designed to support new applications and future revisions of existing applications with minimal configuration.

To achieve this, the following is required:

- **Transparency.** When modifications are required to clients, servers, routers, or the application itself, it becomes increasingly harder to support future revisions or new types of traffic. The most scalable WAN acceleration solutions are completely transparent to existing infrastructure.
- **Multiprotocol support.** A WAN acceleration solution that supports only TCP-based applications will have less applicability across the enterprise than a solution that supports both TCP and UDP. The latter, for example, can be used to improve the performance of VoIP, video, disaster recovery, and other applications.
- **Low latency.** If a WAN acceleration solution imposes high latency on WAN traffic, it cannot be used for time sensitive traffic. This limits its focus to bulk applications, such as file transfers and email.

While initial deployments sometimes focus on a few key applications, such as file and email, most enterprises ultimately want to scale their WAN acceleration investments to support all business critical WAN traffic, including Voice over IP (VoIP), video, ERP, CRM, and custom applications. To satisfy this scalability requirement, application breadth is an important consideration when evaluating WAN acceleration solutions.

6. Quality of Service (QoS)

QoS is a key feature for scalability as it enables more applications to co-exist with one another on a single network. In addition, it ensures that bandwidth is adequately distributed as more applications are vying for a shared limited resource.

Policing and traffic shaping are an indispensable part of any complete WAN acceleration solution. Even with drastic reductions in aggregate bandwidth consumption, there are still bursts of uncompressible data, so it is critical to guarantee bandwidth for enterprise critical applications. In a well-designed network, QoS will be managed at every potential bottleneck point. It is particularly important to implement QoS in the WAN acceleration appliance as it is the only element that has both a pre- and post-optimization view of the traffic.

7. RAM vs. Disk based Data Reduction

The amount of storage that is used for data reduction techniques will impact the overall performance in larger networks. That is because the efficiency of a compression or data reduction technique is dependent on the likelihood that data streams that have traversed the network in the past can be recognized on subsequent passes and eliminated across the WAN.

The most effective data reduction solutions leverage hard drives on appliances to store several weeks or months' worth of traffic patterns, a concept called "disk based data reduction". This efficiently eliminates the transfer of duplicate data by leveraging information collected over extended periods of time. In contrast, some products perform data reduction in RAM (as opposed to using disk drives). While the two solutions may provide similar benchmark results for short data runs, (e.g., in simple lab tests), significant differences will be seen over time and under increased load. Better "long term" memory and higher storage capacity will perform better in real world environments as resources fill up and more information is learned about the particular network environment.

A fully utilized 155 Mbps WAN link, for example, will send 60 Gigabytes (GB) of data per hour in a single direction (120 GB per hour if operating in full duplex.) At 12% utilization, that same link will send nearly a Terabyte (TB) of traffic over a 24 hour period. A WAN acceleration appliance that has a large, rapidly referenceable effective data store of multiple TBs will be extremely efficient in this type of environment.



In contrast, an appliance that operates only out of physical RAM will be limited to 4GB or 8 GB of active data store, resulting in less than 10 minutes of traffic retention. While the differences may not be noticeable when transferring small file sets (e.g., a few GBs), they are apparent when large volumes of information are sent over extended periods of time, as is the case with data center replication and disaster recovery.

8. Effective storage capacity vs. stated capacity

When performing disk based data reduction, vendors use different methods to store information. Some techniques are more efficient than others, resulting in better usage of available storage space, which leads to improved scalability.

A scalable WAN acceleration appliance will store a single local instance of information that is independent of all peer appliances. The alternative – storing a separate instance for each individual WAN link (i.e. peer appliance) – consumes significantly more disk space, particularly when many offices are involved. In addition, detecting and storing repetitive patterns in a bidirectional fashion can dramatically improve scalability by eliminating the need to store a separate instance of information for each direction of a WAN link. For example, if a 10 MB file is sent back and forth between 100 offices, a

head-end appliance that works bi-directionally and stores a single instance of information would use up approximately 10 MB of local data store. If the same appliance stored a separate instance for each direction of each WAN link, it would require 2 GB of local hard drive capacity to support this scenario. The first solution is getting 200x better effective storage capacity than the second solution!

9. Management

Remote provisioning, management, and monitoring are important to scaling WAN operations. This is particularly true in large networks, especially where servers are removed from remote offices and there is limited IT staff onsite to handle day-to-day tasks. Features such as zero-touch provisioning, which allow a device to be sent to a remote office and then automatically provisioned from a central location, are critical for growing a large deployment. The ability to centrally generate network traffic reports and export them to commonly used formats, such as Cisco Netflow, are also important for scalability. Other capabilities that can help ease the burden of managing a large WAN acceleration deployment include centralized alarm aggregation, scheduled tasks, auditing, trending reports, and centralized software updates.

10. High availability and clustering

As WAN acceleration deployments become larger, they often become more strategic to the enterprise. As a result, availability often goes hand in hand with scalability. This is especially true when WAN acceleration is used for data center-to-data center applications, such as disaster recovery, where downtime can mean a significant loss of revenue.

To address the need for high availability, most vendors will support a wide range of network configurations. These often use common techniques to ensure redundancy, such as Virtual Router Redundancy Protocol (VRRP) and Web Cache Coordination Protocol (WCCP). These same protocols can also be used to cluster multiple appliances together for better scalability. This enables traffic to be balanced across multiple devices for better performance under heavy load.

Resiliency can also be built into the WAN acceleration appliance itself. Common features include redundant load-sharing power supplies, fail-to-wire network interfaces, ECC memory and hot-plug RAID drive configurations.



Silver Peak

Silver Peak Systems, Inc.
471 El Camino Real
Santa Clara, CA 95050
408.935.1800
www.silver-peak.com