

Managed Service Provider's Guide to Application Performance Optimization Services

By Rebecca Wetzel and Peter Sevcik

May 2009

The time for service offerings that measure, accelerate, and control application performance over wide area networks (WANs) has come. Products that optimize WAN application performance are now main stream, and enterprises want to purchase them as managed services. So if you already deploy and manage network-based services and equipment, adding application performance optimization to your service portfolio makes business sense. This report describes the customer need, what is in it for you, the anatomy of such services, and how to sell them to your customers.

The Need

The combined forces of globalization, server centralization, inter-office collaboration, and application “webification” conspire to slow application performance. Why? Because these forces increase application response time by: extending distances data must travel, increasing the amount of information transferred, and increasing application “chattiness”.

These forces affect all forms of network traffic from the smallest user inquiry to the largest file transfer, and they affect responsiveness as seen by end users as well as machine-to-machine communication for functions like data replication.

To understand how these forces impair application performance, it is important to understand key factors influencing application response time (R). These include: application turns (round trips), application payload, network bandwidth, network round-trip time (RTT), server compute time (Cs), and client compute time (Cc).

The following formula shows how these factors combine to determine response time for a given task.

$$R \approx \frac{\text{Payload}}{\text{Bandwidth}} + \text{AppTurns (RTT)} + Cs + Cc$$

Two important trends affect the formula to significantly increase response time. Payload size is relentlessly increasing as more data is needed to accomplish application tasks or record business events—and application turn counts are burgeoning.

Long response times due to payload size can be offset by adding bandwidth, but additional bandwidth has limited utility in many scenarios¹, making payload reduction an important means to improve response times.

High turn counts combined with long round-trip times can cripple application performance, thus reducing application turn counts can be critically important in shortening response times. Our research shows that application turn counts for typical business-to-business transactions have increased about 12 percent per year for the past 15 years, while business trends like globalization and data center consolidation have increased data travel distances. The confluence of these trends dramatically lengthens round trip times.

**NetForecast Report
NFR5099**

**©2009
NetForecast, Inc.**

To make matters even more challenging, IT groups are under pressure to do more with less. They are expected to keep application users productive and satisfied even as IT staffs dwindle and budgets shrink.

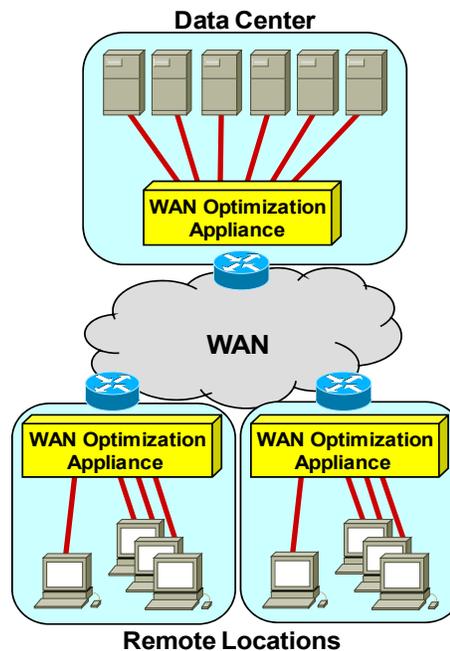
In response to these challenges, enterprises need service providers to offer managed solutions that alleviate application performance problems more effectively and inexpensively than they can do it in house.

Solution Components

A complete application performance optimization solution includes three components—measurement, acceleration, and control. *Measurement* identifies the applications running over the WAN, reports on how well they are performing, and provides information to identify and fix problems. *Acceleration* changes how the application delivery system operates to streamline the process and improve performance for all users all the time. *Control* ensures that performance is as good as it can be during periods of insufficient bandwidth.

Acceleration and control capabilities are delivered by WAN optimization products with solution elements located at remote sites as well as in central locations as Figure 1 shows.

Figure 1. WAN Optimization Product Architecture



WAN optimization products can employ a variety of the acceleration and control techniques shown in Figure 2. Acceleration techniques like TCP optimization, dynamic compression, dictionary compression, and turn reduction actually speed up application performance for all users all of the time by overcoming technical limitations to streamline application delivery. Control techniques like traffic marking, bandwidth allocation, and priority handling keep performance from degrading during times of system stress.

Figure 2 shows the complete set of techniques implemented by WAN optimization vendors. Leading vendors implement a comprehensive mix of these techniques to provide maximum performance improvement for a wide range of applications.

Figure 2. WAN Optimization Product Technologies²

Accelerate	Control (QoS)
TCP Optimization	Application Traffic Discovery
Packet Packing	Traffic Marking
Header Compression	Bandwidth Allocation
Forward Error Control	Priority Handling
Dynamic Compression	TCP Rate Control
Dictionary Compression	DoS Attack Prevention
Remote File Caching	
Application Turn Reduction	

Delivering Performance Optimization as a Service

Innovative service providers are developing application performance optimization services by wrapping service offerings around commercial off-the-shelf WAN optimization products. Since many service providers already manage customer premises-based network infrastructure, it is a natural progression for them to manage the additional infrastructure needed to measure, accelerate, and control application performance. Many customers want and are requesting such services.

Many service providers NetForecast has interviewed describe these offerings as natural enhancements to IP VPN and network management services that can be sold as upgrades to their base services. It is common, for example, for service providers to position application performance optimization as an upgrade to an MPLS-based VPN service offering, and to charge a 15 to 20 percent uplift over the monthly base service price.

Application performance optimization services are commonly sold for a 15 to 20 percent uplift over monthly VPN service prices.

Who Buys Application Performance Optimization Services?

Application performance optimization service buyers fall into two main camps: enterprises (often but not always large) with many distributed locations, and small-to-medium sized businesses with modest numbers of locations. The common purchase motivation is that outsourcing performance optimization capabilities is easier and cheaper than the do-it-yourself alternative. Many of these enterprises already outsource services like router and/or firewall management, and trust their service providers with the additional responsibility.

Highly distributed enterprise buyers must deliver excellent performance to far-flung users, but are plagued by distance-induced latency, low bandwidth links and congestion, all of which slow response times. They also must ensure timely and efficient computer-to-computer communications. To date the market sweet spot for application performance optimization services has been multinational companies that must deliver acceptable performance to corporate users in many global nooks and

crannies. But national enterprises face many of the same performance challenges as their international counterparts, and their requirements are growing.

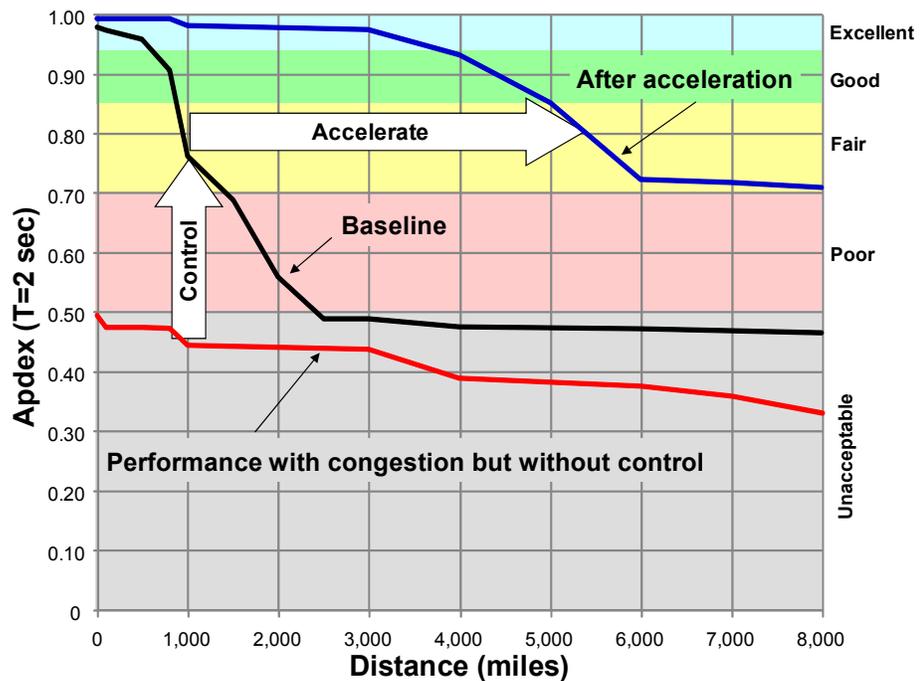
Small-to-medium sized business buyers also have high performance requirements, but they often cannot afford to invest in application management and delivery systems—and they lack internal resources needed to effectively manage the solution. Economic pressures deter them from large capital outlays needed for a do-it-yourself solution, and attract them to the lower operating costs of a service alternative.

Service providers often can do things enterprises aren't prepared to do or don't have time to do. Many enterprises, for example, find it difficult to hire and train employees who can delve into performance details, interpret them, and act upon that information. Such talent is hard to find and expensive to train and retain. It is often easier for service providers to hire and train such professionals and deliver their knowledge in the form of professional services, than it is for an enterprise to develop such capabilities internally. Service providers also manage to service level agreements (SLAs), which can result in a higher level of service performance than an in-house alternative.

Performance Optimization Services “Shrink” the World

Figure 4 shows the effects of server-to-user distance on the performance of Web-based SAP. With congestion and without control, performance is poor even near the server—and turning on control restores the user experience to or near baseline levels. Acceleration has a much more dramatic effect on performance. With acceleration turned on, users located thousands of miles from the server experience the same performance as those who are much closer—thus profoundly “shrinking” the distance from the user’s viewpoint.

Figure 4. Effects of Control and Acceleration on SAP Performance



Apdex is an open standard developed by the Apdex Alliance. It is a numerical measure of user satisfaction with performance on a zero-to-one scale, with one being excellent and values below 0.5 deemed unacceptable. Visit www.apdex.org to learn more.

The Market Potential

WAN optimization products are now thoroughly main stream and customers are looking to service providers to buy the capabilities as services. In fact, sixty percent of US enterprises surveyed by Ovum indicated that they would consider buying or they already buy application performance optimization services.³ NetForecast expects managed application performance optimization service adoption to follow the trajectory of other services such as managed router and managed firewall services, which are now commonplace.

As of this writing more than 20 service providers worldwide offer application performance optimization services—up from eight in April of 2008. The current list includes: Alestra, AT&T, BT Global Services, Cable & Wireless, EDS (HP), Global Crossing, NTT, Orange Business Services, Pacnet, SingTel, Sprint, Swisscom, Telefonica Brazil, Telindus, Telstra, TELUS, TMC Communications, T-Systems, Vanco, Verizon Business, Virtela, and XO Communications.

What's In It for Service Providers?

Application performance optimization services enable you to migrate your service portfolio up the protocol stack into the application layer—and improve your competitiveness, customer retention, cross-selling capabilities and profitability in the process.

Competitiveness: Network transport is quite simply a commodity. Buyers of “dumb pipes” routinely switch service providers based on price, and it is hard to differentiate connectivity services sufficiently to keep old customers and win new ones. When you provide insight into your customers’ WAN traffic makeup and behavior, and add the ability to control and accelerate application performance, you improve your chances to win and keep business for your entire service line.

Customer Retention: Application performance optimization services play an important role in the application layer where your enterprise customers’ businesses run. This makes your customer relationship more strategic because it enables you to ensure that application performance matches your customers’ business needs. This stronger relationship increases customer loyalty and prevents churn.

Cross-Selling Opportunities: Monitoring and reporting on application performance provides insight into customers’ needs. This information can help you identify other solutions in your service portfolio that can meet those needs. For example, if reports consistently indicate that a customer’s email application performance is poor, you may be able to help the customer by migrating them to a hosted email server. This type of cross selling opportunity can generate additional revenue.

Profitability: NetForecast’s research indicates that on average service providers charge a 15-20 percent premium for services that optimize application performance. This additional revenue offsets price erosion for commoditized services, thus improving profitability.

Anatomy of a Performance Optimization Service Portfolio

An application performance optimization service portfolio can include an array of service modules. You will need to craft unique offerings that capitalize upon your organization’s strengths and your customers’ needs, however, in this section we outline a prototypical service portfolio that can serve as a useful starting point.

NetForecast's research shows that although the services described here can stand independently, managed service providers routinely layer them atop MPLS-based IP VPN services and/or existing network management services.

The suggested NetForecast "starter" service portfolio includes application performance:

- Assessment
- Monitoring
- Acceleration
- Control
- Service level assurance

I. Application Performance Assessment

Customer Challenge: Most enterprises lack insight into the applications running over their WAN, and they have little information about each application's performance throughout the network.

Solution Overview: The application performance assessment service identifies the applications traversing the corporate WAN and provides performance information by application. The service provides reports to IT as well as business managers. These reports contain information which helps the organization understand and benchmark current performance and set goals that will ensure the performance for each application supports business requirements.

II. Application Performance Monitoring

Customer Challenge: Enterprises find it difficult to identify and correct application performance quality problems and ensure acceptable ongoing performance for business-critical applications—however, failure to do so hampers an enterprise's ability to meet business goals.

Solution Overview: The application performance monitoring service provides end-to-end visibility into an enterprises' network infrastructure and the applications traversing it including the user experience. The service provides around-the-clock monitoring of all networked applications and it shows server and network contributions to application performance separately. The monitoring information is summarized daily, monthly, and quarterly in reports about application bandwidth usage and application performance. This information is also available in real time via an online portal.

III. Application Performance Acceleration

Customer Challenge: Factors such as long distances, high application turn counts, and large payloads slow application performance—and adding bandwidth cannot correct the problem.

Solution Overview: The application performance acceleration service improves performance for all data applications at all times by changing how the delivery system operates to overcome inherent technical limitations and streamline application delivery. Not only does this improve the end-user experience and speed up machine-to-machine functions like data replication, it also maximizes infrastructure efficiency so enterprises realize the best return on their infrastructure investment.

IV. Application Performance Control

Customer Challenge: Performance for all applications suffers when the network is under stress. Enterprises need to ensure satisfactory performance for critical applications when the network is subject to performance-degrading factors such as congestion, resource loss, or traffic conflicts.

Solution Overview: The application performance control service protects application performance during periods of system stress by adjusting server or network throughput to favor critical applications or users.

V. Application Performance Service Level Assurance

Customer Challenge: Networked applications are increasingly important to businesses, and enterprises need consistent and acceptable end-to-end performance for critical applications over the corporate WAN.

Solution Overview: Application performance service level assurance differentiates performance contributions by the network and the data center, and provides automated system measurements and reports on actual end-user (or machine-to-machine) responsiveness. Response time information is subdivided to document performance at critical grades of the delivery system reflecting the portion attributable to the service provider as a part of the total.

The service provider works with the customer to define business requirements for the end-user experience. The service provider can then offer additional assurance including some or all of the following elements based on ITIL service level management guidelines:

- Service Level Requirement (definition of the business need)
- Service Level Objective (business need defined in IT terms)
- Service Level Target (goal in measurements of the IT terms)
- Service Level Agreement (terms and conditions added)

Note: Each element has its own key performance indicator (KPI) and reports tailored to customer needs and the technical aspects of the service offering. Customers can migrate from service level requirement up the service stack to service level agreement. A graduated fee structure should reflect the degree of service level assurance.

VI. Other Potential Service Offerings

In addition to ongoing measurement, control and acceleration services, many service providers add professional services to the mix. Solution assessments and proof of concept trials are increasingly common, along with consulting and ongoing reporting and recommendations. Some service providers also offer application performance troubleshooting services.

How to Sell to Enterprise Customers

A primary selling point for a comprehensive application performance optimization service portfolio is that it helps your enterprise customers get more from their network infrastructure, especially during lean economic times. The services improve end-user experience quality, provide visibility into application and network infrastructure performance, and ensure consistent application performance during periods of system stress. They also free customers from day-to-day application performance operations to

focus on core business activities, and help them make the most efficient use of existing hardware, software and bandwidth resources.

Well implemented application performance optimization services:

- Enable data center centralization
- Enable server consolidation
- Improve employee productivity
- Increase IT staff productivity
- Improve customer and business partner satisfaction
- Improve disaster preparedness through faster data replication
- Increase the return on current infrastructure and bandwidth investment

Conclusions

The time has come for managed service providers to add application performance services to their service portfolios. Products to optimize application performance over the WAN are now main stream. Enterprises are familiar with managed WAN optimization solutions from major vendors, and their penchant to purchase these solutions as managed services is increasing.

Service providers are stepping up to the plate with services to measure, accelerate and control application performance. You should consider joining their ranks if you already offer network services, and if your customer base includes large distributed enterprises that must deliver excellent performance to far-flung users or small-to-medium sized enterprises that need good application performance to multiple locations.

If you meet those criteria, NetForecast recommends that you start developing a set of application performance optimization service offerings. Executed well, such a service portfolio will create more strategic, long-term customer relationships that will improve customer retention, improve your competitiveness, generate additional revenue, help you cross sell other services, and offset price erosion.

In short, we see application performance optimization services as a must-have for forward-thinking managed service providers.

About the Authors

Rebecca Wetzel is a principal of NetForecast and a veteran of the data networking industry with unparalleled inside knowledge of the Internet service and product markets. She works with network product vendors and service providers to develop and implement product strategies. She can be reached at rebecca@netforecast.com.

Peter Sevcik is a principal of NetForecast and is a leading authority on measuring, assessing, and improving the performance of networked applications. Peter has contributed to the design of more than 100 enterprise systems and pioneered many performance management techniques including Apdex. He can be reached at peter@netforecast.com.

References

- 1 - *Improving Effective WAN Throughput for Large Data Flows*, NetForecast Report 5095, November 2008 (available at <http://www.netforecast.com/ReportsFrameset.htm>)
- 2 - *Field Guide to Application Delivery Systems*, NetForecast Report 5085, September 2006 (available at <http://www.netforecast.com/ReportsFrameset.htm>)
- 3 - *The Adoption of WAN Optimisation as a Managed Service*, Ovum Report, July, 2008

NetForecast helps enterprises and vendors understand and improve the performance of networked applications.

Additional information is available at:
www.netforecast.com

NetForecast and the curve-on-grid logo are registered trademarks of NetForecast, Inc.