Buyer’s Guide:

Application Delivery Solutions

Ashton, Metzler & Associates

Leverage Technology & Talent for Success

February 2006

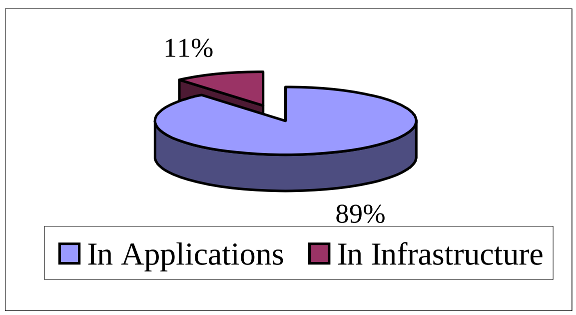
Table of Contents

1.0 Introduction

In both an article in the Harvard Business Review [1] as well as a subsequent book, Nicolas Carr made the assertion that "IT has become a commodity. Affordable and accessible to everyone, it no longer offers strategic value to anyone."

It is true that some components of IT, such as Ethernet switches, have become commodities. However, Carr's assertion dramatically underestimates the strategic value of IT. That is because over the last few years the IT value proposition has shifted away from a focus on provisioning the infrastructure and towards a focus on ensuring the effective delivery of business applications.

This trend was confirmed in a recent study performed by Ashton, Metzler & Associates. Approximately 200 IT professionals were asked to identify where their company's business and functional managers saw value from the IT function. Was it in the infrastructure? Or was it in the applications? As shown in Figure 1, the response to that question was overwhelming. In particular, the vast majority of enterprises recognize that the value of IT is predicated on its ability to successfully deliver business critical applications to employees.



Source: Ashton, Metzler & Associates

Relative Value of IT

Figure 1

Combining the salient aspects of Carr's argument with the data in Figure 1 indicates that the real mandate facing IT organizations is to simultaneously reduce both the cost and the risk associated with the IT infrastructure while also ensuring optimum application performance. Unfortunately, these objectives are often at odds with one another.

An example of when these objectives can be at odds with each other is server consolidation. Enterprises want to consolidate servers out of branch offices and into centralized data centers to reduce the cost associated with the hardware, software licenses, real estate, and IT resources required to support and manage enterprise applications. Centralizing servers also allows an IT organization to demonstrate that it is managing risk because it allows IT organizations to respond to the ongoing enactment of government and industry regulations, such as the Sarbanes-Oxley

[1] "IT Doesn't Matter", Nicholas G. Carr, Harvard Business Review, May 2003

Act (SOX). However, server consolidation often creates a variety of application performance problems due to protocol limitations and the inherent challenges associated with communicating across a Wide Area Network (WAN). For example, while opening a file might take milliseconds when the client and the server are in the same branch office, it can take several seconds when the client is in a branch office and the server is hundreds or thousands of miles away in a data center.

A new generation of Application Delivery Solutions has recently emerged. This generation of solutions is designed to enable IT organizations to be able to centralize servers and still ensure optimum application performance.

The goal of this paper is to create a set of criteria that helps the reader evaluate these solutions and determine which Application Delivery Solution is best for them. By choosing the best solution, IT can achieve its mandate of reduced cost, tighter security, and better compliance, while also playing a strategic role in achieving business objectives by ensuring optimal performance for all key applications.

2.0 The Application Bottleneck

Throughout this white paper the performance issues encountered by users in a branch office attempting to access centralized applications will be referred to as The Application Bottleneck. There are several factors that cause The Application Bottleneck, ranging from application specific issues to limitations in the network infrastructure.

2.1 Wide Area Networking Limitations

For a number of reasons, running an application over a Wide Area Network (WAN) is fundamentally different than running the same application over a Local Area Network (LAN). For example, unlike a LAN, there are monthly recurring charges that are associated with a WAN circuit and these charges increase as the speed of the circuit increases. So, while companies run their LANs at speeds ranging from 10 Mbps up to 10 Gbps, virtually no company can afford to implement WAN circuits that run at these speeds.

Another key distinction between a WAN and a LAN is latency. Given the limited geographic extent of a LAN, latency within the typical LAN is negligible. In contrast, latency within the WAN can have a significant impact on application performance, particularly on delay sensitive applications such as Citrix and Voice over IP (VoIP). WAN latency can also have a significant impact if the application uses a chatty protocol such as CIFS (Common Internet File System) or NFS (Network File System). Chatty protocols, which were designed to run over a LAN, exchange tens or even hundreds of messages between sender and receiver for each transaction.

Another factor that introduces WAN performance issues is the TCP windowing algorithm. The TCP window indicates how much data can be outstanding without an acknowledgement. Neither the business criticality nor the delay sensitivity of the application that is run on top of TCP has any influence on the TCP window size. What does impact the TCP window size is successful transmission. In particular, as the application is successful in sending packets across the network, the size of the TCP window is steadily increased, allowing these applications to have greater throughput. Conversely, if there is packet loss, the size of the TCP window size is reduced and this causes the application throughput to also be reduced.

2.2 Challenges with Existing Applications

There are many application-specific factors that also contribute to The Application Bottleneck. For example, one well-known problem that is associated with server consolidation is that Microsoft file services rely on the CIFS protocol. As previously mentioned, CIFS is a chatty protocol that was designed to run over a LAN. When run over a WAN, CIFS performs very badly and can cause highly disgruntled users.

One solution to this problem that has garnered a lot of attention in the last year or so is Wide-Area File Systems (WAFS). WAFS are focused on reducing the latency that is associated with taking applications that are based on specific LAN protocols such as CIFS or NFS, and running these applications over a WAN.

For many companies, WAFS solves an important problem in the short term. However, one of the important considerations relative to providing branch office workers with a solution that allows them access to applications is that the key characteristics of these applications change over time. For example, Microsoft recently announced R2, the new Windows Server release. One of the goals of R2 is to eliminate the aspects of CIFS that lead to CIFS based applications running badly over the WAN. If R2 is successful, the need for a WAFS solution will diminish.

Another well-known problem that is associated with server consolidation is that Microsoft Exchange makes use of the MAPI (Message Application Programming Interface) protocol. MAPI, like CIFS, was designed to run over a LAN and can behave very poorly when run over a WAN. As a result, point products appeared in the market with the sole purpose of eliminating the latency that is associated with centralized deployments of Microsoft Exchange. However, the enhancements that Microsoft made to Exchange 2003 reduced many of the performance problems associated with running MAPI over a WAN.

A final well-know problem is that many applications are very sensitive to key network parameters such as delay, jitter and packet loss. An example of such an application is Citrix, which has been deployed by the majority of large companies. One of the key advantages that Citrix provides to a user in a branch office accessing an inquiry-response application is that only the changes (i.e., the keystrokes, mouse clicks and screen refreshes) are transmitted over the network. Since entire screens of data and graphics are seldom transmitted, this approach requires relatively little bandwidth. Citrix is, however, sensitive to delay and packet loss.

Another such application is VoIP. The ITU (International Telecommunication Union) recommends that the end-to-end delay associated with a voice call not exceed 150 milliseconds. Experience has shown that it is possible to exceed that goal by a small amount. However, if the delay becomes too large, the quality of the voice call degrades noticeably.

Jitter is a measure of how packet delay changes over time. The vast majority of data applications are not impacted by jitter. Again, voice is different. Using the RFC 1889 definition of jitter, jitter should not exceed 30 ms or else the voice quality degrades noticeably.

3.0 Deployment Considerations

3.1 Performance

As previously noted, many companies that have consolidated servers into one or more centralized data centers have encountered significant application performance problems. If the IT organization cannot resolve these application performance problems, the server consolidation initiative is doomed to failure.

The important role that performance plays in a server consolidation initiative is reflected in terms of how IT organizations justify to management the investment in solutions to The Application Bottleneck. The conventional wisdom is that justifying these solutions requires a thorough ROI analysis. A recent market study [2] indicated that it might be time to change that conventional wisdom. In particular, that market study analyzed how companies have justified investments in these solutions. The study reported that only one quarter of IT organizations actually completed a detailed Return on Investment analysis and presented the results to management for their approval. The study also reported that well over half of IT organizations have used improving user productivity through better performance as a justification for investing in solutions to The Application Bottleneck.

3.2 Need for an Integrated Solution

The previous section of this white paper described how point products such as WAFS and Microsoft Exchange accelerators might not be long-term solutions in part because the problems that they were designed to solve are either going away or changing significantly. Another reason why point products may not be a good long-term solution for many companies is because they require IT organizations to implement numerous appliances in each branch office. In many cases these highly specialized appliances sit in branch offices next to numerous other appliances, each of which is performing some other highly specialized task in order to make individual applications run better over the WAN. Few companies want a deployment strategy that results in removing servers from branch offices and replacing them with multiple appliances.

3.3 Supporting UDP

There is no doubt that TCP is the most commonly used transport protocol in the Internet Protocol (IP) suite. However, a significant amount of network traffic does not use TCP. In particular recent market research[3] indicates that on average 27% of network traffic is not TCP.

A lot of this traffic is UDP (User Datagram Protocol). UDP is used by numerous protocols, including:

---

[2] The Impact of Data Center Consolidation on Branch Office Performance, COMPUTERWORLD, September 2005
[3] Ibid

- DHCP (Dynamic Host Configuration Protocol)
- TFTP (Trivial File Transfer Protocol)
- RIP-1, RIP-2 (Routing Information Protocol)
- SNMP (Simple Network Management Protocol)

In addition, DNS (Domain Name Server) uses UDP for most message exchanges and early versions of NFS (Network File System) use UDP.

Many custom developed applications also use UDP. For example, UDP is used in applications that require multicast or broadcast transmissions, as these functions are not supported in TCP. UDP is also used when an application developer decides to use a simple protocol to provide some of the functionality found in TCP and hence avoid the added overhead that is associated with a full TCP implementation. Perhaps most important is that UDP is used when an application values the timely delivery of packets more than the reliable delivery of packets.

A very important example of an application in which the timely delivery of packets is more meaningful than the reliable delivery of those packets is VoIP. The human ear is not very sensitive to a small amount of packet loss. As a result, when transmitting voice over an IP network it is more important that the stream of packets keeps flowing than it is to retransmit a packet that will arrive out of context and hence be of no value.

The reason that VoIP is such an important example of the use of UDP is because virtually every study of the topic [4] indicates that the percentage of companies that have deployed VoIP has steadily increased to the point where the majority of companies have already begun to deploy it. Market research also indicates that most company's initial deployment of VoIP is relatively limited. However, over time most companies expect to increase their deployment of VoIP.

The increasing percentage of companies that are deploying VoIP, coupled with the increasing proliferation of voice in environments already using VoIP, leads to the conclusion that the amount of network traffic that is based on UDP will steadily increase over time.

[4] The 2004 VoIP State of the Market Report, Steven Taylor, Distributed Networking Associates, Inc.

4.0 Evolution of Application Delivery Solutions

4.1 First Generation Solutions

Over the last few years there has been significant innovation relative to enhancing application delivery. This spate of innovation has lead to multiple generations of Application Delivery Solutions, each with a different emphasis. For example, the first generation of Application Delivery Solutions emphasized WAN optimization and featured technologies such as compression and QoS.

The value of compression is that it reduces the size of a file of data to be transmitted over a WAN. As such, compression can often be used to cause the WAN to perform as if it had received a one-time bandwidth upgrade.

QoS refers to the ability of the network to provide preferential treatment to certain classes of traffic, such as voice traffic. QoS is required in those situations in which bandwidth is scarce and there are one or more delay sensitive, business critical applications.

In a typical scenario, deploying a first generation Application Delivery Solution results in performance that is two to three times better than it had been. However, these solutions do not address some of the latency specific issues associated with server consolidation; i.e., running CIFS over a WAN. In addition, these solutions have not been widely deployed as they are very tactical and address only short-term performance problems on individual WAN links.

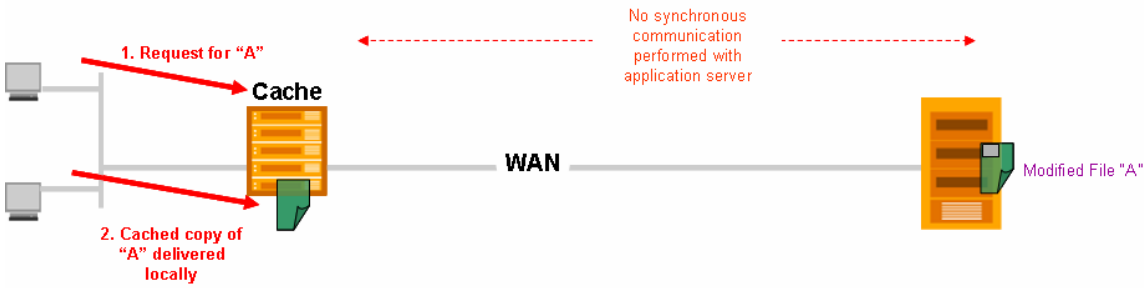
4.2 Second Generation Solutions

The second generation of Application Delivery Solutions resolved some of the latency specific issues that were not addressed in the first generation of Application Delivery Solutions by emphasizing application acceleration and by focusing on technologies such as caching and application proxies. Caching refers to keeping a local copy of information with the goal of either avoiding or minimizing the number of times that the information must be accessed from a remote location. For example, most Web browsers contain a cache that stores the most recent Web files that have been downloaded.

Application proxies are typically deployed in a branch office and simulate the functionality of a remote application server. This allows certain content to be delivered to the branch office users with LAN-like performance. The previously mentioned WAFS solution is an example of using a proxy server to accelerate some specific applications.

In many cases, application caches and proxies offer significant value. However, these solutions also have significant limitations. For example, as Figure 2 demonstrates, in a caching environment the client does not interface directly with the server, it interfaces directly with the cache. As a result, caching changes the overall communications flow between the client and the server

.



Second Generation Caching
Figure 2

For example, when a client requests File A the cache intercepts the request. If a copy of File A has previously been stored in the cache, the cache returns the file locally, eliminating the backend request to the server and the transfer of the file over the WAN. If a copy of File A is not already in the cache, the cache proxies a request for File A on behalf of the client, and File A is transmitted over the WAN to the branch office.

Because caches do not refresh local content on a real-time basis, they suffer from data coherency issues. As a result, caches in remote offices can be out of synch with each other and with the data stored at the central site. Referring to Figure 2, this means that Copy A may not be the same as File A or the same as the copy of File A that exists in other branch offices.

In addition to data coherency issues that arise from altering client-server communications, second generation caching solutions have other significant limitations, which include:

- Caches are Application Specific

  This means that IT organizations have to implement and manage separate caches for each application.

- Limited Applicability

  Second generation caching only works when the identification of the object that is being requested (i.e., URL, file name) matches the identification of data that is stored locally. Again referring to Figure 2, this means that if the client requested File B that was a "virtual" match with Copy A (i.e. same content, but using a different filename), File B would still be transmitted in its entirety over the WAN.

- Management and Security Complexity

  Because a cache interfaces directly with clients, it is responsible for functionality such as handling authentication and authorization as well as file locking. Inserting this layer of intelligence into a network adds management and security complexity.

- Re-Configuration Requirements

  In many cases, it is necessary to re-configure the client so that it points to the cache proxy server and not the original application server. This adds additional overhead and complexity.

5.0 Buying Guide for a Third Generation Solution

A third generation of Application Delivery Solutions has emerged. This generation of solution is designed to enable IT organizations to successfully consolidate servers while overcoming the application delivery and deployment issues discussed in sections 2 and 3. This section of the white paper will begin by presenting a characterization of this new generation of Application Delivery Solutions. This section will conclude by presenting a set of selection criteria that enterprise IT organizations should use when evaluating third generation Application Delivery Solutions. The first appendix to this document contains a checklist that IY organizations can use to evaluate Application Delivery Solutions.

In order to provide context for what IT organizations are looking for in a third generation Application Delivery Solution, Ashton, Metzler & Associates interviewed two IT professionals. One of the interviewees was a senior network administrator and the other was a CIO. They will be referred to in this white paper as The Administrator and The CIO respectively.

5.1 Characterization of a Third Generation Solution

The primary business goal addressed by the third generation of Application Delivery Solutions is to be able to have centralized servers and yet still provide LAN like application performance to users in branch offices. Achieving this goal allows IT organizations the flexibility to consolidate existing servers out of branch offices and into centralized data centers. Achieving this goal also allows IT organizations to minimize server sprawl as new applications are deployed.

The CIO stated that stopping server sprawl was a high priority of his. He further stated that at the time that his organization decided to deploy Microsoft Exchange, they already had three or four servers in each of their remote offices and he wanted to "put the brakes on deploying additional servers". One of the reasons his organization deployed an Application Delivery Solution was to avoid having to put a Microsoft Exchange server in each branch office.

It is important to note that point products (i.e., WAFS and Microsoft Exchange accelerators) will continue to provide value to IT organizations for at least the near term. However, a third generation Application Delivery Solution is not designed to be a point product, but rather an integrated solution to a broad range of factors that impact The Application Bottleneck. In particular, a third generation Application Delivery Solution avoids littering branch offices with multiple appliances by providing the same functionality as was provided by both of the previous generations of solutions. That means that a third generation Application Delivery Solution simultaneously optimizes WAN Bandwidth and improves Application Performance.

A third generation Application Delivery Solution must also overcome the limitations of previous generations of solutions. In particular, a third generation Application Delivery Solution must be transparent to as many aspects of the network environment as possible; i.e., this generation of solution must not be application specific. In addition, a third generation Application Delivery Solution must provide at least an order of magnitude performance increase to overcome the bandwidth and latency limitations present in most WAN environments.

5.2 Selection Criterion

The previous sub-section of this white paper characterized what is meant by a third generation Application Delivery Solution. In order to assist IT organizations to evaluate these solutions, this sub-section will discuss specific functionality that should be present in a third generation Application Delivery Solution.

5.2.1 Performance

To the CIO, performance is "the main feature" of an Application Delivery Solution. He stated that, "this is the metric that you compare first and foremost." The Administrator agreed. He stated that performance is "a key consideration" and that the solution "needs to flat out perform".

There are many ways to improve application performance over a WAN. A third generation Application Delivery Solution will address a wide breadth of performance attributes through techniques such as the following:

<u>Improve WAN bandwidth utilization</u>

This includes techniques such as payload and header compression that reduce the size of packets traversing the WAN. Coalescing, which refers to combining multiple packet fragments into a single packet in order to use WAN bandwidth more efficiently, can also provide significant benefits. More advanced solutions will also implement data reduction (described below), which can provide enormous improvements in the utilization of WAN bandwidth.

<u>Minimize the effects of latency</u>

This includes TCP acceleration techniques such as window scaling (RFC 1323) which increases throughput by enabling a larger TCP window size. It also includes techniques such as CIFS Read-Ahead and Write-Behind. These refer to situations in which an application is performing sequential file reads and writes. In these situations, CIFS can anticipate the user's requirements and read/write portions of the file before the user has requested those portions.

Third generation solutions also leverage various techniques to deliver information locally whenever possible. Since these techniques minimize the amount of data that must traverse the WAN, they also minimize the latency associated with transmitting data over a WAN.

<u>Mitigate packet loss</u>

This includes techniques that overcome performance challenges in WAN environments that display high loss characteristics. A primary tool to achieve this is called Forward Error Correction (FEC). FEC has long been used at the physical level to ensure error free transmission with a minimum of re-transmissions. FEC can also be applied at the network layer whereby an additional error recovery packet is transmitted for every 'n'packets that are sent. Network layer FEC enhances performance if the network is congested and hence dropping a significant number of packets. A third generation Application Delivery Solution monitors the network for packet loss and employs network layer FEC only if the amount of packet loss is sufficiently high to justify it. More advanced FEC implementations will be adaptive, adjusting the amount of FEC packets in accordance with levels of packet loss. This helps to minimize overhead while maximizing performance.

5.2.2 QoS

As described in section 4, QoS refers to the ability of the network to provide preferential treatment to certain classes of traffic, such as voice traffic. Accelerating the performance of the network and of applications does not mitigate the fact that QoS is required in those situations in which bandwidth is scarce and there are one or more delay sensitive, business critical applications.

QoS involves two separate, but related functions. One function is packet classification based on factors such as the source address, the destination address, or the application itself. The second function is servicing packets differently based on their classification.

An IT organization may choose to implement QoS classification in a variety of locations, including the LAN switch, WAN router, or in the application acceleration appliance itself. It is important to note, however, that most application acceleration appliances modify IP addresses and port numbers, compress headers and payload content, and encrypt the data. Therefore, it is not always feasible to leave packet classification to a downstream device, such as the WAN router.

A third generation Application Delivery Solution must provide the IT organizations with as much flexibility as possible in order to support varying network environments. That means that a third generation Application Delivery Solution must be able to honor any QoS classification set by an upstream device. It also means that a third generation Application Delivery Solution must also be able to provide the QoS functionality itself including sophisticated classification logic, a number of packet marking techniques, as well as queuing and traffic shaping.

The Administrator commented that within his company's network, "the Application Delivery Solution is the place to implement QoS in order to free up the router to focus on routing packets." The Administrator also thought that QoS would become increasingly important to his company as they began to implement VoIP.

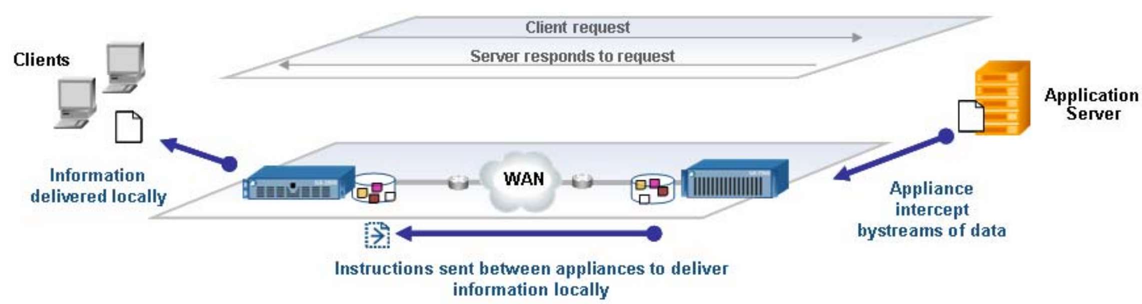
The CIO agreed on the importance of QoS. He commented that, "all of the company's revenues come in over their ERP system. It is very important to use to be able to give a high priority to ERP traffic."

QoS will prove to be even more important as more real-time applications, like VoIP, are deployed throughout the enterprise. These types of applications require guaranteed bandwidth and/or specific delivery characteristics that distinguish them from other enterprise applications.

5.2.3 Data Reduction

For the sake of this white paper, data reduction will refer to techniques that reduce the amount of data that is transmitted from a central site to a remote site over a WAN. Data reduction differs from compression in that data reduction causes large amounts of traffic destined for remote sites to be delivered locally whenever possible. Since the traffic does not traverse the WAN this functionality results in maximizing WAN bandwidth efficiency while avoiding latency and other performance limitations associated with communicating across a WAN.

As shown in Figure 3, a data reduction solution works by inspecting all inbound and outbound WAN traffic at the byte level in real-time and storing a single instance of the data at each location within the enterprise. Prior to transmitting traffic streams to the remote site, a third generation Application Delivery Solution at the central site compares the traffic streams to patterns already present in the Application Delivery Solution at the remote site. If there is a match, a small reference pointer is sent to the remote site instructing the Application Delivery Solution at the remote site to deliver the traffic stream from its local memory. In this way, data that is already stored at the remote site is not transmitted over the WAN.



A Data Reduction Solution
Figure 3

If, however, the traffic stream contains some modifications to what is stored in the remote site, a third generation Application Delivery Solution at the central site will identify those modifications at the byte level. Just those modifications are sent to the Application Delivery Solution at the remote site, where the modifications are combined with the original content to create and deliver a current data stream.

Whether or not the traffic stream contains modification, a key characteristic of a third generation Application Delivery Solution is the ability to ensure 100% concurrency of the data.

5.2.4 Scalability

The Administrator emphasized the importance of a scalable solution when he stated that he didn't want a solution "that I have to rip out in six months". Many factors influence the scalability of an Application Delivery Solution. For example, the more information that an Application Delivery Solution can remember, the better its performance will scale over extended periods of time.

Given that requirement, an Application Delivery Solutions should utilize a disk-based storage system that can store weeks or possibly months worth of data in order to facilitate extensive pattern matching and data reduction. The size of the data store used in each appliance will vary based upon the requirements of the individual offices as well as their expected growth. For example, a small branch office might use 250 GB of local store, while a large data center might require 2 TB of hard drive space.

Application Delivery Solutions use different techniques to store data within their memory, which can also affect scalability. To exemplify this, consider a network comprised of 100 branches, each of which has 20 Gb of data. In some solutions, the central site has to keep a copy of the data for every branch office. In this example, this approach would require the central site to store 2 terabytes of data. A more efficient Application Delivery Solution would only store one instance of each unique piece of information at the central site, requiring only 20 GB of storage for the same network. The second solution has much larger effective storage capacity than the first solution, despite the fact that both solutions might claim similar physical storage capacities.

Several other factors can also influence the scalability of an Application Delivery Solution. One such factor is the number of simultaneous users that the solution can support. Another factor is the number of remote sites that can be connected to a central site. This is often limited by TCP flow counts. As a result, it is advisable to look for an Application Delivery Solution that does not originate and terminate TCP connections, as these devices often deliver greater overall scalability.

5.2.5 Transparency

A fundamental requirement of a third generation Application Delivery Solution is transparency. In particular, second generation Application Delivery Solutions are highly specialized solutions that continually require modifications as applications get modified. The goal of transparency is to install a solution and have it work over as broad a range of network conditions as possible and to require as little tweaking as possible as those conditions change.

Transparency is particularly important to The CIO. He stated that, “I want to solve this problem (application performance) once and be done with it. If we later deploy another application, the solution needs to also work with that application”.

Transparency was also very important to The Administrator. He stated that “I want a solution that I can deploy and have it work from day one. With all the changes in protocols and applications, I don’t want to have to make constant changes to my WAN optimization solution.”

There are many facets of transparency that must be in a third generation Application Delivery Solution. This includes:

<u>Application Transparency</u>

A third generation Application Delivery Solution must support all applications without requiring any customization of the solution. For example, a third generation Application Delivery Solution must enable an IT organization to implement the type of server centralization that was described in section 1 without having to re-configure clients, servers, or the application itself. It must also preserve client/server communications in order to eliminate any potential coherency issues.

In addition, there can be no adverse impact on application performance. For example, some solutions might increase the performance of file services and email, but add latency to time sensitive applications, like VoIP or Citrix. For a solution to be strategic, it must provide improvements across all applications.

Transport Protocol Transparency

As described in section 2, over a quarter of network traffic does not use TCP. As a result, in order to be truly transparent, a third generation Application Delivery Solution must support all transport protocols, not just TCP. This requirement means that the solution must operate at the network layer of the OSI (Open Systems Interconnection) model.

Communications Transparency

A third generation Application Delivery Solution will require communications between devices on both sides of a WAN. As a result, it must support common techniques for encoding/compressing data on one end of the WAN and directing it to the device on the other end of the WAN for decoding/decompression.

The two primary ways of doing this are via tunneling and NAT. Tunnels function by inserting a packet of data inside of an IP header. This header contains the address of the destination device and is used to route the packet between the originating and destination device. NAT is a technique that remaps IP address and ports and is often used to hide IP addresses that are internal to an organization.

In order to be truly transparent, a third generation Application Delivery Solution must work in either or both of these environments.

5.2.6 Security and Compliance

IT organizations have the mandate to continually implement more effective security. Part of this mandate comes from an organization's internally generated desire to protect its assets. However, the mandate also comes from some of the recent regulations that are very focused on security. For example, the Federal Information Security Management Act requires that each federal agency must develop, document and implement an agency-wide program to provide information security.

Other key regulations that are driving an increased emphasis on security include:

The Sarbanes-Oxley Act

This act requires management to make a written assertion stating their responsibility for establishing and maintaining an adequate control structure and procedures for financial reporting.

HIPAA

HIPAA (the Health Insurance Portability and Accountability Act) requires companies in the health care industry to provide administrative simplification, security and privacy.

The Gramm-Leach-Bliley Act

This act requires companies to give consumers privacy notices that explain the institution's information-sharing practices and to give consumers the right to limit some of the sharing of its information.

A third generation Application Delivery Solution must enable IT organizations to respond to the mandate for more effective security. This starts by enabling branch office server centralization because as mentioned in section 1, it is very difficult to ensure the security of data when there a multiple copies of that data located on servers in branch offices.

Beyond server centralization, there are specific security measures that must be taken within an Application Delivery Solution. This includes using sophisticated encryption algorithms such as the 128 bit Advanced Encryption Standard (AES) to encrypt both the data that is stored in the branch offices as well as the data as it transmits the WAN.

This requirement was strongly supported by The Administrator who stated that having security on an Application Delivery Solution "is a must". He further stated that his company's policy is to encrypt data everywhere, so he can only deploy an Application Delivery Solution that encrypts data as it resides on the local disks and as it transits the WAN.

A third generation Application Delivery Solution must support additional security functionality such as secure login via protocols such as HTTPS and SSH, as well as user authentication using protocols such as RADIUS and TACACS+.

6.0 Summary

In order to be perceived by the company's business managers as providing business value, IT professionals need to be able to show that they can enable the optimum performance of the company's key applications.

Over the last few years there has been significant innovation that was focused on allowing the IT organization to enhance application performance. The first generation of Application Delivery Solutions emphasized WAN optimization and featured technologies such as compression, and QoS, whereas the second generation of Application Delivery Solutions emphasized application acceleration and focused on technologies such as caching and application proxies.

Each of these generations of Application Delivery Solutions provides value, but also has significant limitations. IT organizations that are taking a forward look at optimizing application performance should analyze the third generation of Application Delivery Solutions. These solutions are designed to eliminate the limitations of the first two generations of Application Delivery Solutions while simultaneously providing performance improvements that are an order of magnitude better than those solutions.

Some of the key characteristics of a third generation Application Delivery Solution are:

<u>Performance</u>

These solutions use a combination of compression and acceleration techniques, coupled with newer technologies that increase data reduction and enable local information delivery. As a result, a third generation Application Delivery Solution can improve application response time by a factor of at least fifty. In contrast, previous generations of Application Delivery Solutions typically improved application response time by a factor of two or three.

<u>QoS</u>

These solutions are able to honor any QoS classification that is set by an upstream device and they also are able to provide sophisticated QoS functionality themselves. QoS becomes especially important when enterprises deploy a mix of applications to branch offices, including real-time traffic such as VoIP, and interactive applications such as Citrix.

<u>Data Reduction</u>

These solutions transmit the absolute minimum amount of information over the WAN while maintaining data concurrency throughout the enterprise. To accomplish this, the devices on each end of the WAN circuit exchange pointers, which are used to cause information to be delivered locally whenever possible. This results in delivering LAN-like performance in many scenarios.

<u>Scalability</u>

These solutions are able to scale to meet present and future application delivery requirements, and are not unduly impacted by the number of simultaneous users, the number of remote sites that are attached to the central site, or the number of TCP flows that are traversing the WAN.

Transparency

These solutions optimize performance across all applications, independent of transport protocol and communications methods

Security

These solutions are able to implement powerful encryption on both the data that is stored in the branch offices, as well as the data as it transits the WAN.

The latest generation of Application Delivery Solutions enables IT organizations to centralize servers and yet still provide LAN like application performance to users in branch offices. Achieving this goal allows IT organizations the flexibility to consolidate existing servers out of branch offices and into centralized data centers. Achieving this goal also allows IT organizations to minimize server sprawl as new applications are deployed.

Appendix A: Product Evaluation Scorecard

Vendor Name:

Product Name:

Breadth of Applications Supported

- File Systems (CIFS, NFS)
- Email
- Web
- VoIP
- Citrix
- SQL
- Other

Core Features

- Data Reduction
- Local Information Delivery
- Compression
- TCP acceleration
- Loss mitigation
- Quality of Service

Scalability

- Long term memory
- WAN capacity per appliance
- Optimized flows per appliance

Security

- Encrypted data store
- Encrypted communications

Appendix B: Terminology

AES	Advanced Encryption Standard
CIFS	Common Internet File System
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name Server
FEC	Forward Error Correction
IP	Internet Protocol
IT	Information Technology
ITU	International Telecommunication Union
LAN	Local Area Network
MAPI	Message Application Programming Interface
NAT	Network Address Translation
NFS	Network File System
OSI	Open Systems Interconnection
QoS	Quality of Service
RIP	Routing Information Protocol
SNMP	Simple Network Management Protocol
SOX	The Sarbanes-Oxley Act
TCP	Transmission Control Protocol
TFTP	Trivial File Transfer Protocol
UDP	User Datagram Protocol
VoIP	Voice over IP
WAN	Wide Area Network
WAFS	Wide-Area File Systems