

The
ESSENTIAL GUIDE
for
Upgrading
your **Network**

.....
1 NETWORKING EVOLUTION
AND ROADMAP

.....
2 MOVING TOWARD THE
APPLICATION-CENTRIC NETWORK

.....
> 3 **UPGRADING
DISTRIBUTED NETWORKS**

.....
4 SECURING THE NEW
NETWORK ARCHITECTURE

.....
5 CASE STUDY:
TOMORROW'S NETWORK—TODAY

CHAPTER 3

Upgrading Distributed Networks

By Robin Layland

At a major enterprise, executive IT management has already made the decision to consolidate local file servers in branch offices to the data center. Consolidation makes sense—it will save money, provide a more secure environment and make maintenance easier. The problems, however, begin to rear their heads early in the project. The branch offices consolidated first begin complaining because the time it takes to download files is affecting their productivity. The consolidation project is saved by implementing a new branch office technology: application acceleration. With an application acceleration appliance, download times are reduced to acceptable levels and the project is back on track.

Branch offices are important to the business. They need the same quality IT services as corporate headquarters and can't be treated as second-class locations. When branch offices lose connectivity, there is an immediate

impact on the overall business. The fear is that providing excellent service to branch offices can also create a money pit. It is not just server consolidation that has made managing distributed networks more difficult. The move to Web browser-based applications has increased the size of each transaction. If that were not enough, the HTTP protocol has some inefficiency built into it that can make it slower than older client/server applications.

Server consolidation

The growth of server-based applications has done wonders for productivity and provides important functionality to the people in distributed locations. File servers allow users to quickly retrieve important business data. Email servers such as Microsoft's Exchange provide fast and efficient email service. Having these and other servers in branch offices has made good response time, and thus productivity, the norm.

But the growth of servers and applications in the branch office has a dark side. Maintenance and problem resolution are expensive. It takes the IT staff extra time and expensive tools to remotely diagnose problems. The remoteness leads to frustration for both the branch office workers and the IT

staff. Remote servers waste resources if they are running at low utilization, which is a common occurrence.

Backup and recovery takes longer and uses expensive WAN resources when the server is remote. The security of the server is also challenging and makes it harder to meet many of the regulatory requirements of providing data protection.

All these reasons have led to the desire to move remote servers to the data center. The IT staff is located there and can react quickly when a problem arises. Backing up or restoring a server is faster when it is at the data center. It is easier to apply best practices and ensure that the data on the server is secure in the data center. With all servers at the data center, the IT staff can take advantage of server virtualization technologies, such as VMWare, to combine several servers into one.

It doesn't matter whether it is called server consolidation or data center consolidation—the concept solves many problems. However, moving servers to the data center is not the perfect solution. The workers in branch offices frequently see poor response time that negatively affects productivity and morale. Consolidation can also affect the budget, because servers residing in the data

center and transmitting all data to branch offices require significant WAN resources.

The question is: How does a good infrastructure design become a better infrastructure design? The answer is incorporation of new technologies that allow an enterprise to capture all the benefits of server consolidation while solving its problems. In this case, the solution at hand goes by two names: WAN optimization and application acceleration. Both names refer to virtually identical technological solutions. If a vendor wants to emphasize WAN bandwidth savings and resulting cost savings, it focuses on WAN optimization. If the vendor wants to highlight improvement in response time and productivity, it focuses on application acceleration.

Application acceleration and WAN optimization

Acceleration applies many techniques to solve the twin problems of poor branch office response time and the extra bandwidth required. The techniques can be grouped into two general categories. The first are “generic” techniques. Generic techniques apply to all the data going to the branch office, no matter the protocol. The benefit is that the technique helps CAD/CAM, file and Web traffic equally.

The primary generic techniques include TCP/IP protocol optimization, bandwidth management and shaping, quality of service (QoS), and compression. Some of these optimization techniques have been around for a while; the biggest recent improvements are in the area of compression.

Older compression techniques generally reduced the amount of data sent by two to three times, while newer techniques, called dictionary compression or de-duping, can reduce bandwidth requirements by 10 to 50 times. Applying these newer compression algorithms means that

How Dictionary Compression Works

IN DICTIONARY COMPRESSION, also known as de-duping, the accelerator learns patterns from the data flowing through it and stores them in a large cache, located both in memory and on a disk drive. The patterns are generally 100 characters long. Accelerators are located at the data center and the branch office, and they both learn the same patterns from the data. The first time the data passes the accelerator, it can only apply older compression techniques; the real advantage comes when it sees the pattern the second time. When the pattern shows up again—this can be in any data, including data totally unrelated to the first instance—the accelerator substitutes the entire pattern with a reference number. The reference number refers to the pattern it has stored, and because the accelerator on the other end has learned the same pattern, it can easily rebuild the message.

For example, if a PowerPoint presentation is attached to an email, the first time it is sent to the branch office there is some data reduction. When the file is sent back to the data center with a few changes, the accelerator can use its pattern database to remove all the parts that haven't changed and send only reference numbers to those parts along with the changes. The result is that a file that was 5 MB can be reduced to a few kilobytes.

despite the large bandwidth requirements for server consolidation, the overall utilization of a WAN link could be less than it was before con-

Accelerators can do wonders, but they don't help with certain types of traffic.

solidation. Response time is also improved because the overall amount of data that needs to be sent is decreased and because the smaller compressed packets are automatically combined into larger packets, reducing the number of packets sent.

The second group of techniques that accelerators apply to improve response time and solve the problems of consolidation are “protocol specific.” Many protocols, including Microsoft's Common Internet File System (CIFS) and HTTP, are not very efficient. This inefficiency is unnoticed in a LAN environment because of the speed of the LAN and the short distances traveled. Over the slower WAN, however, protocol inefficiencies can affect response time. Accelerators understand the protocols and apply techniques that overcome their shortcomings. For example, Microsoft file servers can ex-

perience close to LAN-like service with the combination of generic techniques and CIFS-specific acceleration.

Accelerators can do wonders, but they don't help with certain types of traffic. Video, such as training films, is not helped much by accelerators because video is already highly compressed. Voice traffic can actually suffer because there is little an accelerator can do, and trying to accelerate it can actually slow it down. It is best to have the accelerator recognize voice traffic and pass it directly through at a high priority.

Acceleration architectural challenges

Acceleration can do wonders for response time and significantly reduce bandwidth requirements, but several issues must be addressed for a successful implementation. The first is a good understanding of which applications are using the WAN.

Gone are the days when identifying traffic by port number was enough. Knowing that Web applications are using port 80, for example, tells you little. Web-based applications using the same port number can include those that are mission critical along with those that are time wasters. Accelerating all Web applications may mean that music-sharing applications

run faster. Before applying acceleration, the network group needs to implement application monitoring tools that report on the applications that are using the network, not just the ports that are being used. This information will allow the accelerator to accelerate business applications before non-critical applications. It is also important because, in many cases, network managers will not be aware of all the applications using the network.

The next issue is what to do about encrypted traffic. The movement of

applications to Web interfaces has made it easier to encrypt the traffic using SSL. There are many good reasons to use encryption, but it is impossible to apply many of the acceleration and optimization techniques, such as dictionary compression, to encrypted traffic. If business critical traffic, or a significant amount of overall traffic, is encrypted, then an accelerator that can de-encrypt traffic, accelerate it and then re-encrypt is needed. There are accelerators that can perform this function, but not all do it equally well.

Keeping the Branch Connected

FAST APPLICATION RESPONSE time is meaningless if branch connectivity to the data center is lost in the event of an outage or disaster. Providing backup connectivity has always been difficult and expensive. Even if two service providers are in the area, their actual infrastructure often follows the same route out of the building and may be subject to the same backhoe accidents or other disasters.

A new alternative is wireless connectivity from cellular vendors. The connectivity many mobile workers use to get a broadband connection can also be used to connect a branch office. Branch office routers are available that integrate this option directly into the router. The biggest advantage of this option is that the cellular last-mile infrastructure is completely separate from landline facilities. The speeds are not as high as normal landline connection, but wireless can provide significant bandwidth that allows the office to continue working.

Securing the accelerator is also necessary because of the new capabilities of dictionary compression techniques and file caching. The dictionary compression file has a copy of all the patterns that have passed through the appliance. With file caching, a copy of the file is stored on the appliance. If someone hacks into the accelerator or runs out the door with it, then it is possible that the sensitive data could be compromised, and traffic could even be recreated from the stored patterns. This is not likely, because the patterns are short and nothing in the accelerator relates one pattern to another, making it very difficult to reconstruct a file. The solution is to encrypt both the cache and the compression files. This feature is available from many acceleration vendors, but not all of them.

Another architectural issue is transparency. This issue has two layers. The first is how the traffic is packaged when it is sent between the two accelerators. The most common way is to create a tunnel between the two accelerators with all the accelerated traffic having a new TCP/IP header added to the packet. The transparency issue is that any monitoring or security device between the two accelerators will no longer see the traffic as coming from the client or server. This

loss of visibility causes problems for monitoring and security equipment. The solution is to move all the monitoring or security devices before the

When any monitoring or security device that performs deep packet inspection looks into the packet, what it will see is nothing like what the client or server sent.

accelerator. Some of the acceleration vendors do not create tunnels between the accelerators and thus do not have this problem.

A larger transparency issue is created by the accelerators. Accelerators significantly change the traffic by compressing it and combining multiple packets into one larger packet. When any monitoring or security device that performs deep packet inspection looks into the packet, what it will see is nothing like what the client or server sent. Because compressing packets is inherent in the acceleration and optimization process, the only solution is to place all security and monitoring devices before the accelerator. ■

Robin Layland is president of Layland Consulting. As an industry analyst and consultant, Robin has covered all aspects of networking from both the business and technical sides and has published more than 100 articles in leading trade journals, including *Network World*, *Business Communication Review*, *Network Magazine* and *Data Communications*. Prior to his current role, Robin spent a combined 15 years at American Express and Travelers Insurance in a wide range of jobs, including network architect, technical support, management, programming, performance analysis and capacity planning.