SPECIAL REPORT

WAN:
The value of visibility

By Jim Metzler

Sponsored Exclusively by:
FLUKE networks

Produced by:
NETWORKWORLD®
CUSTOM MEDIA SOLUTIONS
www.networkworld.com

Table of Contents

The value of visibility

During the dotcom boom, IT organizations were under much pressure to show the business value that they provided. During that period, it was implicit in the thinking of most companies that the IT function provided immeasurable value because IT would allow them to quickly and easily enable fundamental business transformations.

Now that the dotcom boom is over, it is incumbent on all IT organizations to continually make the company's senior management aware of the general value that IT provides, as well as the particular value provided by a company's Wide Area Network (WAN). Making the latter task a bit easier is the fact that the business value of a company's WAN has been steadily growing because of a number of significant trends that have been building over the last decade. Those trends include:

- The mandate to establish linkages with partners, suppliers and distributors.

  As recently as the mid to late 1990s a typical corporate WAN connected a company headquarters' sites to each other and to a company's branch offices. The WAN of that era was used primarily for e-mail and some online applications, such as sales order entry.

  Today, in addition to connecting headquarters' sites and to branch offices, the typical corporate WAN also connects a company to its key partners, suppliers and distributors. The contemporary WAN is used for a variety of critical business purposes, including supporting sophisticated supply chain management processes.

- The movement of employees out of headquarters' sites.

  As recently as 10 years ago, the vast majority of employees worked in a headquarters' sites. Today, the vast majority of employees work in branch offices. A company's WAN enables these branch-office employees to have access to the same applications that are available to employees who work at a headquarters' sites.

No organization can function in today's business environment without a WAN. However, it is important to realize that in addition to being a business critical resource, managing a WAN involves a number of complex issues. For example, a WAN is expensive, performs in ways that are notably different than the way a Local Area Network (LAN) performs and introduces security vulnerabilities.

Given both the complex issues associated with managing a WAN, as well as its business criticality, all IT organizations should have extensive visibility into the performance of the WAN. This special report will outline some of the key issues associated with running applications over a WAN. It will also highlight the value that results from having visibility into the performance of the WAN and will describe some of the most popular sources of management data.

Classes of WANs

There are three general classes of WANs. They are:

- **Private**
  In a private WAN, the IT organization acquires its component pieces (that is, transmission links, routers, frame relay access devices) from various suppliers and is responsible for planning, designing, implementing and managing a WAN.

- **Public**
  In a public WAN, the majority of the planning, designing and implementing is the responsibility of the service provider. In most cases, ongoing WAN management is a joint responsibility of the IT organization and the service provider. It has become common to refer to the current generation of public network services as Virtual Private Networks (VPNs).

- **Hybrid**
  A hybrid WAN combines a private and public WAN. In most instances of a hybrid WAN, the IT organization uses a private WAN to connect its headquarters' sites and relies on a public WAN to connect to some, or possibly all, of its remote and branch offices.

Whether a WAN is private, public or hybrid, it is necessary for the IT organization to have visibility into its performance. This topic will be expanded upon in the WAN visibility section of this report.

Key WAN characteristics

For a number of reasons, running an application over a WAN is fundamentally different than running the same application over a LAN. For example, unlike a LAN, recurring monthly charges are associated with a WAN circuit. In virtually all cases, the recurring monthly charges associated with a WAN increase as the capacity of the circuit increases. In the case of some technologies (that is, a private line) the cost of the WAN also increases as the distance between locations increases. So, while companies typically run their LANs at speeds of 10Mbps and higher, few companies can afford to implement WAN circuits that run at these speeds, other than within metropolitan areas.

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, significant latency within the WAN is commonplace. The processing that occurs within the service provider's equipment as the data travels from origin to destination causes some of this latency. However, the delay in the speed of light as a signal travels through a combination of copper and fiber optic facilities causes the majority of the latency.

WAN latency typically has a negligible effect on such applications as e-mail and bulkfile transfer. As will be described in the next section, WAN latency can have a significant impact on VoIP, as well as Enterprise Resource Planning (ERP). WAN latency also can 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, designed to run over a LAN, exchange tens or even hundreds of messages between sender and receiver in each transaction.

Other factors, such as packet loss or packets errors, also have an impact on WAN performance. For example, assume that two devices are exchanging packets over an ATM network and that the size of each packet is 1,000

bytes. Since an ATM cell contains 48 bytes of data, it takes 21 ATM cells to encode just one of these packets. If even one of these ATM cells is either lost or received in error, then the entire packet has to be retransmitted.

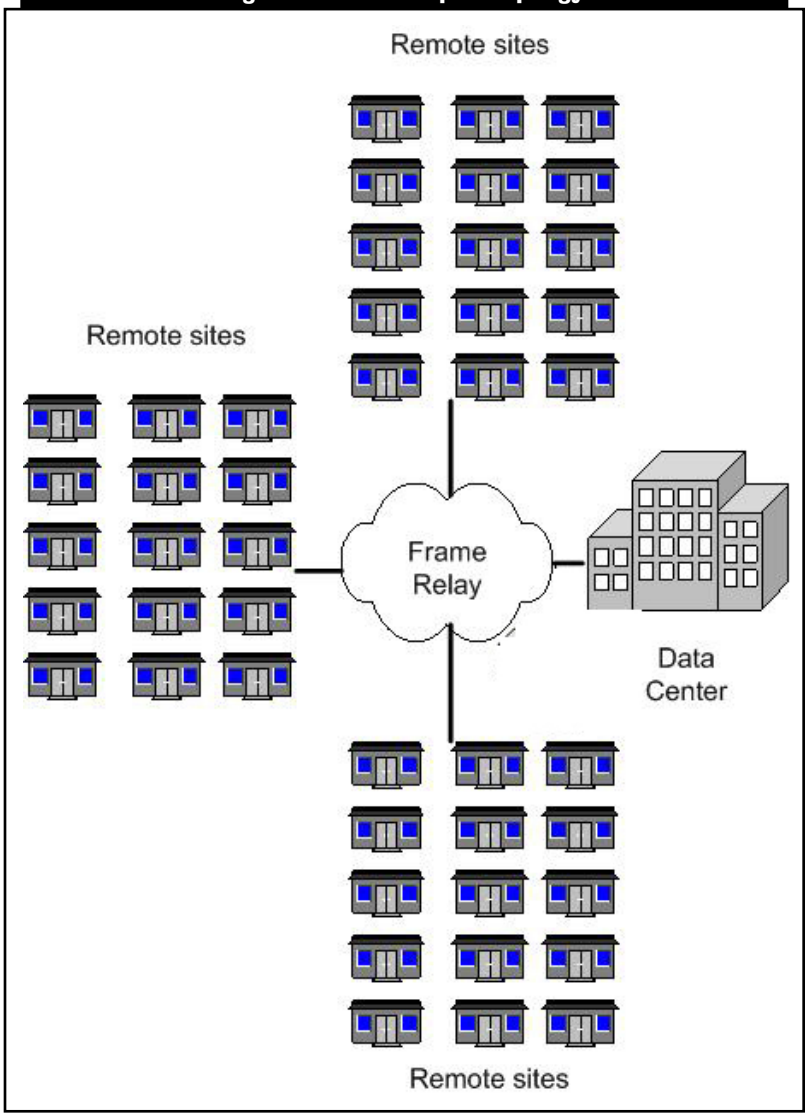
Having to retransmit a packet has a significant impact on application performance because of the way that the TCP windowing algorithm functions. The TCP window indicates how much data can be outstanding without an acknowledgment. Neither the business criticality nor the delay sensitivity of the application that is run on top of TCP has any influence on the TCP window size. What affects a 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 steadily increases, allowing these applications to have greater throughput. Conversely, if a packet has to be retransmitted, the size of the TCP window size is reduced, and this cuts application throughput.

Evolving technologies and topologies

For most of the 1990s, the most common type of branch office WAN was a hub and spoke network based on frame relay or ATM (Figure 1). A hub and spoke network is one in which there is a Permanent Virtual Circuit (PVC) that connects each branch office to a central site, typically a company's data center. Note that in a hub and spoke network, any communications between branch offices has to transit through a central site.

A hub and spoke topology is appropriate when the bulk of communications is between a branch office and a headquarters' site. However, a number of trends are shifting that type of traffic pattern in a fundamental way. Those trends include the need for employees at the branch offices to have access to applications that reside in multiple data centers, as well as to a disaster recovery site.

Figure 1 : Hub and spoke topology



Another trend driving this shift in traffic patterns is the deployment of VoIP, which shifts traffic patterns, because it is an any-to-any type of application. That is, employees at a branch office do not talk just to other employees at headquarters, but to a broad range of people. As a result, voice traffic flows from each branch office to a wide range of locations. While it would be possible to have this voice traffic move through a central site, that is very inefficient if there is a significant traffic load.

The applications challenge

This section of the report will discuss a number of the challenges associated with supporting applications on a WAN.

VoIP

Over the last few years, there has been a strong movement on the part of IT organizations to deploy VoIP. In addition to shifting the traditional traffic patterns, as previously described, VoIP deployments will also place stringent demands on the IT infrastructure.

For example, a traditional data application, such as e-mail, can accept a small amount of downtime. That is not true with voice. The vast majority of people expect that their voice call will go through every time they pick up the phone and dial a number.

In addition to higher availability, another requirement that distinguishes VoIP from a more typical data application is the rigorous demands that voice places on the underlying IP network. These rigorous demands include the requirement to have exceedingly low levels of delay, jitter and packet loss.

For example, for many data applications, end-to-end delay is not a critical issue. That is not the case with voice. The International Telecommunication Union recommends that the end-to-end delay associated with a voice call not exceed 150 millisecs. 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 affected by jitter. Again, voice is different. Using the RFC 1889 definition, jitter should not exceed 30 millisecs, or voice quality degrades noticeably.

Like virtually all data applications, VoIP traffic is sensitive to packet loss. However, unlike most data applications, lost voice packets should not be retransmitted. This follows in large part because modern voicecoding algorithms (called CODECs) are explicitly designed to deal with the occasional loss of a packet. As such, it is better to let the CODEC do what it was designed for, than to retransmit a voice packet and introduce significant delay and jitter. To avoid retransmitting packets, VoIP does not use TCP, but uses User Datagram Protocol.

Enterprise applications

Many enterprise applications, such as ERP, are sensitive to some key network parameters. Table 1 depicts the effect that network latency, measured in millisecs has on two modules of an ERP application. Note that each module has the goal of a 5-second response time.

Table 1: Application performance

Network Latency	Module #1 Measured response time	Module #2 Measured response time
0 millisecs	2 seconds	2 seconds
100 millisecs	2 seconds	3 seconds
150 millisecs	2 seconds	4 seconds
250 millisecs	2 seconds	7 seconds
350 millisecs	4 seconds	18 seconds
450 millisecs	4 seconds	34 seconds
550 millisecs	12 seconds	57 seconds

Source: Ashton, Metzler & Associates

As shown in Table 1, if there is no network latency, each module has a 2-second response time, which is well within the target response time. As network latency is increased up to 450 millisecs, there is little impact on module No. 1's response time. If network latency is increased above 450 millisecs, module No. 1's response time increases rapidly and is soon well above the target response time.

Module No. 2 behaves somewhat differently. As latency is increased, its performance begins immediately to gracefully degrade. Once the network latency has reached 250 millisecs, the module's response time exceeds the target and begins to degrade further quite rapidly.

Chatty protocol-based applications

As mentioned, chatty protocol refers to those that exchange tens or even hundreds of messages between sender and receiver for each transaction. Chatty protocols

were designed to run on a LAN, and they perform well in that environment. However, applications that use a chatty protocol often have significant performance issues when run over a WAN.

Over the last couple of years many companies have encountered the difficulty of running chatty protocols over the WAN after consolidating servers from branch offices and into centralized data centers. These initiatives are widespread because of the benefits that result from server consolidation.

One of the benefits that result from server consolidation is that an IT organization gets more control over the company's data resources. Historically, many companies have kept key company data on servers in branch offices. Having multiple copies of a company's data on servers in branch offices makes it difficult for companies to exert the type of control required by recent governmental regulations, such as the Sarbanes-Oxley Act. These regulations require that companies put a greater emphasis on assuring accuracy, security and confidentiality of data. Consolidating servers in a centralized data center makes it easier for the IT organization to exert this control.

Consolidating servers in a centralized data center reduces cost by requiring fewer servers. Fewer servers reduce the cost of the relevant software licenses, support and real estate.

However, one well-known problem associated with server consolidation is that Microsoft file services rely on the CIFS protocol. CIFS works by sending packets from the client to the server in order to request some kind of service, such as opening, closing or reading a file. The server processes these packets and checks to see whether the client has the appropriate file permissions. If the client has the proper permissions, the server executes the request and sends one or more packets back to the client.

When run over a LAN, these service request packets introduce negligible latency. However, companies that are consolidating their servers into a centralized data center end up running CIFS over the WAN. When run over the WAN, these service request packets add latency that is potentially noticeable to the user.

However, a more important factor that influences the user's experience is that CIFS decomposes all files into smaller blocks before transmitting them. For example, assume that a client was attempting to open up a 5MB file on a remote server. CIFS would decompose that file into tens or possibly hundreds of small data blocks. The server sends each of these blocks to the client, where it is verified, and an acknowledgment is sent back to the server. The server must wait for an acknowledgement before sending the next data block. As a result, it can take several seconds for the user to be able to open up the file.

Network misuse

One factor that contributes to the difficulty that an IT organization has in supporting applications over the WAN is network misuse. The phrase refers to instances when the WAN is used to carry unauthorized applications.

Controlling network misuse has always been important. However, the issue of network misuse is growing in importance because of the creative and taxing forms of misuse possible. For example, a company recently discovered that one of its employees is a big Jimmy Buffet fan, so

much so that the user was continually streaming music from a Jimmy Buffet Web site. Many companies have discovered that their employees have used Kazaa or a similar program to share music with their friends.

But it is not just music that IT organizations have to worry about. Sports fans can keep up with their favorite teams, as it is now possible to watch major league baseball games over the Internet. It is also possible for employees to stream Sirius radio to their PC and listen to their favorite shock jock or perhaps to their favorite opera.

If network misuse like the ones mentioned above goes undetected, it can consume sizable portions of a company's WAN.

The cost of the WAN

The costs associated with an investment in any component of IT can be assigned into three categories: facilities, capital and personnel.

Facilities costs include the expenses associated with heating and air conditioning, cabling, floor space and rack space. Facilities costs include the expenses associated with the WAN transmission links.

Capital costs include the expenses associated with the initial acquisition and implementation of the hardware and software used for tasks such as switching and routing data, as well as for managing and monitoring the network.

Personnel costs include the salaries of the internal personnel, as well as third parties, that are associated with the ongoing support of the IT infrastructure.

To do a true total cost of ownership analysis of the WAN, it would be necessary to quantify all of the facilities, capital and personnel costs associated with the WAN over some appropriate time frame. For the sake of simplicity, the analysis in this section will look at just the expenses associated with the WAN transmission links over three years.

To analyze these expenses, this section will use a hypothetical company referred to as Acme. The Acme WAN comprises a single hub and 50 branch offices and is based on the use of public frame relay services.

There are three primary cost components associated with Acme's frame relay network. They are the costs associated with:

Access circuits

This includes the circuits that connect Acme's branch offices to the frame relay network, as well as the circuit that connects Acme's headquarters to the frame relay network.

PVC speed

This refers to the committed transfer rate between one of Acme's branch offices and the Acme headquarters' site.

Port speed

This defines the maximum rate of data transfer that is possible.

To perform this cost analysis, a number of assumptions have to be made. For example, it is assumed that each of Acme's branch offices uses a T1 access circuit and has a 256Kbps frame relay PVC and a 512Kbps frame relay port. It is also assumed that that a T3 circuit and frame relay port is used to connect the Acme headquarters' site to the frame relay network. Table 1 contains some assumptions about the pricing of the service Acme uses.

Table 1: Frame relay pricing

Cost component	Monthly recurring cost
Access circuits	
T1	$200
T3	$1,600
PVC speeds	
256 Kbps	$45
384 Kbps	$75
Port speeds	
512 Kbps	$425
768 Kbps	$550
T3	$5,200

Based on the assumptions listed above, the monthly recurring charges associated with the Acme WAN total $40,300. Those charges over a three-year period would total $1,450,800.

Now assume that the traffic on Acme's network is consuming all of the available bandwidth and that users are starting to complain about application performance. Further assume that Acme is contemplating "throwing bandwidth" at the performance problems. The phrase means that Acme does not have a good understanding of either how its WAN is used or what is causing the applications to behave badly. Acme hopes that if it adds bandwidth to the WAN, the problems will go away.

The WAN upgrade that Acme is considering is to increase the frame relay PVCs from 256Kbps to 384Kbps and to increase the frame relay ports from 512Kbps to 768Kbps. It is important to note that this is a very minor WAN upgrade. Not only is the increase in the speed of the frame relay PVCs and ports relatively small, there is no increase to any other part of the Acme WAN; that is, the access circuits.

While the upgrade to Acme's WAN is minor, the cost impact is major. Over three years, this upgrade would cost Acme an additional $279,000.

WAN visibility

The preceding sections discussed some of the challenges associated with running applications over a WAN, as well as the significant cost associated with operating a such a network. This section will discuss how visibility into WAN performance can enable network organizations to improve the performance of applications while reducing the WAN's cost. This section also will discuss some of the primary sources of management data that are used to get this visibility.

Turning visibility into value

To improve application performance and reduce the WAN's cost, many parameters must be monitored. A minimum set includes the:

- Type of WAN circuit; for example, frame relay, ATM.
- Throughput capacity of the circuit.
- Availability of the circuit.
- Utilization of the circuit.
- Set of devices (for example, switches, routers) that are using the circuit.
- Problem log that shows any alarms or alerts that have been generated.

To identify the source of degraded application performance, a number of additional parameters need to be monitored. This includes the time it takes:

- To establish a TCP connection between the client and the server.
- For the server to begin to respond to a request.
- From when the server begins to respond to a request until it is finished.
- For a packet to traverse the network.

It is also important to know what applications are running on the network. This knowledge allows a network

organization to shut down access to unauthorized applications, such as Kazaa or Internet radio, and hence free up considerable bandwidth. This information also can be used to identify applications that are running during the peak time of the day. If some of these applications can be shifted to non-peak hours, this also frees up considerable bandwidth. In both examples, knowledge of the application enables the network organization to avoid throwing bandwidth at a performance problem.

From a security perspective, knowledge of the applications helps a network organization identify suspicious traffic that might be associated with someone trying to hack into the network. It also helps the IT organization to eliminate viruses and worms.

Knowing the applications that run on the network enables an IT organization to improve the performance of those applications. For example, if an IT organization realizes that voice traffic is running on the network, it will know that it has to closely monitor delay and packet loss, as voice quality is very sensitive to these parameters.

The IT organization may also use its knowledge about which applications are running on the network to determine whether implementing functionality such as Quality of Service would enhance the performance of those applications. As is the case with implementing any significant change to the network, having visibility into the performance of the WAN allows an IT organization to understand whether their implementation of QoS had the desired effect or needs to be modified.

Sources of management data

NetFlow

NetFlow is a Cisco IOS software feature that is a readily available, often untapped resource, which offers significant value as a source of management data. Within NetFlow, a network flow is defined as a unidirectional sequence of packets between a given source and destination. The branch office router outputs a flow record after it determines that the flow is finished. This record contains such information as time stamps for the flow start and finish time, the volume of traffic in the flow, its source and destination IP addresses and source and destination port numbers.

NetFlow data can be used for a variety of purposes, including:

- **Network monitoring:** NetFlow data can be used to visualize traffic patterns either network-wide or associated with a given switch or router.
- **User monitoring and profiling:** NetFlow data can be used to identify how individual users are using the network and applications.
- **Security analysis:** NetFlow data can be used to identify and classify distributed denial of service attacks, as well as viruses and worms in real time.
- **Accounting and billing:** NetFlow data can be used as input to systems that produce detailed information about resource utilization and also can be used to enable usage sensitive chargeback.

IPFIX

The Internet Engineering Task Force used Cisco's NetFlow Version 9 as the basis of IP Flow Information Export (IPFIX). The goal of IPFIX is to allow for a standards-based approach that developers of network management applications can use to gain access to management data. IPFIX achieves this by standardizing the format used for exporting router-based information about network traffic flows to data collection devices and network management systems.

Probes

Probes attach to the network via passive taps to collect management data on applications, hosts and networks. Whereas NetFlow and IPFIX provide broad visibility across the entire enterprise over time, probes enable the IT organization to have rapid, deep visibility into critical links or trouble spots. Probes enable this visibility by providing the most sophisticated and complete class of management data across all seven layers of the OSI model.

In many cases, the data collected from probes is used to augment the data provided by NetFlow by offering

analysis of the physical links, error detection and notification. These probes provide remote-monitoring statistics that feed various management systems, and they can also provide packet capture when that is beneficial.

The management data generated by probes can be used for varied purposes, including application and network monitoring, capacity planning, troubleshooting, fault prevention, service level management, modeling and billing. In particular, the data collected from probes can be use for:

- Device discovery and analysis.
- Rapid troubleshooting of a particular site, segment or circuit.
- Monitoring a service provider's SLA.

While probes provide significantly more management data than is provided by NetFlow, they also cost more. As such, IT organizations must make a trade-off between the level of visibility they need and the cost of getting that visibility. One common approach is to install probes in various facilities at a company's headquarters and to use NetFlow in its small branch offices. That still leaves unanswered the question of whether a company should use probes or NetFlow in its other facilities. There is no right or wrong answer to that question. What some companies do is to use NetFlow in these other facilities but have some extra probes that can be used in these facilities when the need arises.

Summary

All IT organizations are under constant pressure to demonstrate the business value that they provide. One powerful approach that an infrastructure or management organization can use to show such value is to enable better application performance while simultaneously reducing cost.

To accomplish these goals, IT organizations need extensive visibility into the performance of their WANs. This visibility allows them to identify the applications that are running on the network and to ensure that those applications are getting the type of WAN performance that they require. This visibility also allows IT organizations to reduce cost through removing unauthorized traffic and shifting traffic that is not time critical to off-peak hours.

The management data necessary to provide this visibility can come from a variety of sources, including NetFlow and probes. When choosing between NetFlow and probes, IT organizations have to make a classic trade-off between cost and functionality. As ever, there is no right or wrong way to make that trade-off. One common approach is to use probes in headquarters' sites, NetFlow in small offices and have some extra probes that can be used in the remaining facilities when the need arises.

© 2006 Network World, Inc. All rights reserved.

To request reprints of this special report contact networkworld@reprintbuyer.com