White Paper

WAN Optimization – A Clear Cut Comparison of Leading Technologies

Table of Contents

WAN Optimization – The Need

In today's global business arena, enterprises are increasingly dependent on their networks for real-time access to critical company resources. Enterprise IT organizations are challenged to provide more applications, services and performance without exceeding the limitations of costly international access or investing in infrastructure upgrades. From VoIP to accelerating the performance of mission-critical ERP and CRM applications, the performance of an enterprise's WAN is becoming increasingly vital to ensuring the ongoing success of its business. What's more, in light of the ongoing shift towards server consolidation, a rapidly growing number of end-users, often in multiple branch locations, are accessing data via the WAN and both performance and security are being compromised as a result.

IT managers look to WAN optimization (also referred to as traffic management) solutions to mitigate the negative effects of the WAN, such as low bandwidth, high latency, packet loss, or network congestion. When network capacity constraints put operations at risk, there are a range of technologies available to address these issues. This has opened up the opportunity for multiple WAN optimization techniques, based on various technologies, such as traffic shaping, compression and acceleration techniques, to make their way into the market.

This paper offers an overview and comparison of an array of predominant WAN optimization technologies available in the market today:

1. Compression
2. Protocol Acceleration
3. CIFS Acceleration
4. Traffic Shaping

WAN Optimization – Technologies

1) Compression

Compression solutions are marketed as a method for expanding network capacity by shrinking the volume of data to transport and improving WAN application performance. Compression is the equivalent of zipping and unzipping every packet on your WAN. The technology reduces the number of bytes that need to be transferred, and hence the bandwidth needed to carry out the transaction. This allows more data to flow through WAN or Internet links, freeing bandwidth to enhance the performance of applications traversing the network.

In essence, the sender's end shrinks data by substituting a type of shorthand for common sequences, and the receiver's end interprets the shorthand to restore the original content. A compression algorithm is the particular method used to shrink the size of transferred traffic. A compression dictionary is the location where an algorithm stores its data sequences, shortened substitutions, and/or any other data it needs to do its job. For data transmission, compression can be performed on just the data content or on the entire transmission unit (including header data).

Deployment

Compression devices operate in pairs - one to compress traffic at the source, the other to decompress it at the destination. Both the sending (compressing) and receiving (decompressing) units speak the same language - they use the same compression algorithm for the same traffic and keep each other posted on the contents of associated compression dictionaries.

Benefits

Compression enables more data to flow through WAN or Internet links, freeing bandwidth to enhance the performance of applications that are most critical to business. This enables organizations to postpone or avoid bandwidth upgrades and reduce both operational and capital costs.

Drawbacks

Despite the obvious advantages of introducing a compression technology in a specific network environment, the technique also carries a number of inherent faults. Although applying compression may increase the "virtual size" of a WAN connection, it does not differentiate between applications and does not contribute to the overall improvement of WAN performance when applied as a sole optimization solution.

- **Devices are non-discriminatory in terms of application**: Compression performs the same action on all defined traffic, whether it is a P2P session or an HTTP transaction. This limits network managers' understanding of the types of traffic crossing the network and whether performance problems remain.
- **High Total Cost of Ownership:** Two compression devices are required per connection point.
- **Not suitable for high-end networks:** In high-end environments with numerous connections and/ or a large number of packets, compression is limited due to its resource-consuming activity and is therefore rapidly overloaded and cannot perform sufficiently.
- **Stop-gap solution**: Because compression does not increase bandwidth or optimize its utilization, the underlying causes for congestion are not being solved and will only recur as the existing bandwidth is consumed by more and more traffic.

- **Effectiveness is dependant on the file type**: Voice-over-IP (VoIP), Citrix, certain multimedia applications and many image and video files are pre-compressed and cannot be further reduced. In other cases, compression devices simply cannot compress data from applications such as Adobe Acrobat, Microsoft Corp.'s SQL Server or ZIP software.
- **Encrypted packets cannot be compressed:** As traffic encryption becomes increasingly prevalent compression leaves larger amounts of uncompressed data on the network.
- **Deployment adds to network complexity and increases the number of points-of-failure**: This directly counters industry trends and expert recommendations that users consolidate equipment and network functionality wherever possible.
- **Adds latency**: Standard compression adds latency and variability.

2) Protocol Acceleration

Protocol acceleration stems from the fact that bandwidth applications are often unable to utilize available bandwidth due to inherent protocol limitations. The two major protocols which are today candidates for acceleration are TCP and HTTP. As these are the two predominant protocols used for network communications, protocol acceleration technologies are emerging to address a range of limitations they propose such as distance delay, latency, packet loss and unresponsive HTTP sessions. The aim of this technology is to reduce the effects of these limitations and enable the utilization of wasted bandwidth.

Accelerating TCP

TCP utilizes a sliding window mechanism to limit the amount of data in flight at any time. When the window becomes full, the sender stops transmitting until it receives new acknowledgements (ACKs). While waiting for an acknowledgment the transmission it said to have "windowed out" Over long distance networks, where acknowledgements are slow to return, the TCP window size often sets a hard limit on the maximum throughput rate. Each operating system has a predefined window size. Depending on the distance of the link, it can take up to 600 ms to get the ACK.

TCP acceleration devices create a tunnel between them, through which all TCP traffic is directed. The devices intercept all TCP traffic and "cheat" each session by sending an acknowledgement regardless of the predefined window size. This saves a great deal of wasted time and drastically reduces latency involved in transmitting files over a great distance.

Accelerating HTTP

HTTP is the most common application used for web browsing and downloads, and as a transport layer for many business applications. When a user attempts to download a webpage the browser contacts the server and requests that the server deliver the document to it. The request is in essence a series of HTTP GET commands that are immediately sent to the site's server. Although HTTP defines what the browser and web server are transmitting, the actual communication protocol they use is TCP. This is where the same system of transmissions and acknowledgements, is optimized using acceleration technologies as clarified above. If there are references to images, Java applets, sound clips etc, these will be separate requests, and add additional load to the server and network. When the user follows another link the whole sequence starts anew.

Deployment

Much like with compression technologies, protocol acceleration involves two acceleration devices – one at each endpoint location, designed to accelerate the entire transmission process. In most cases, IT managers can decided whether to deploy a certain type of acceleration on a given connection in line with network requirements.

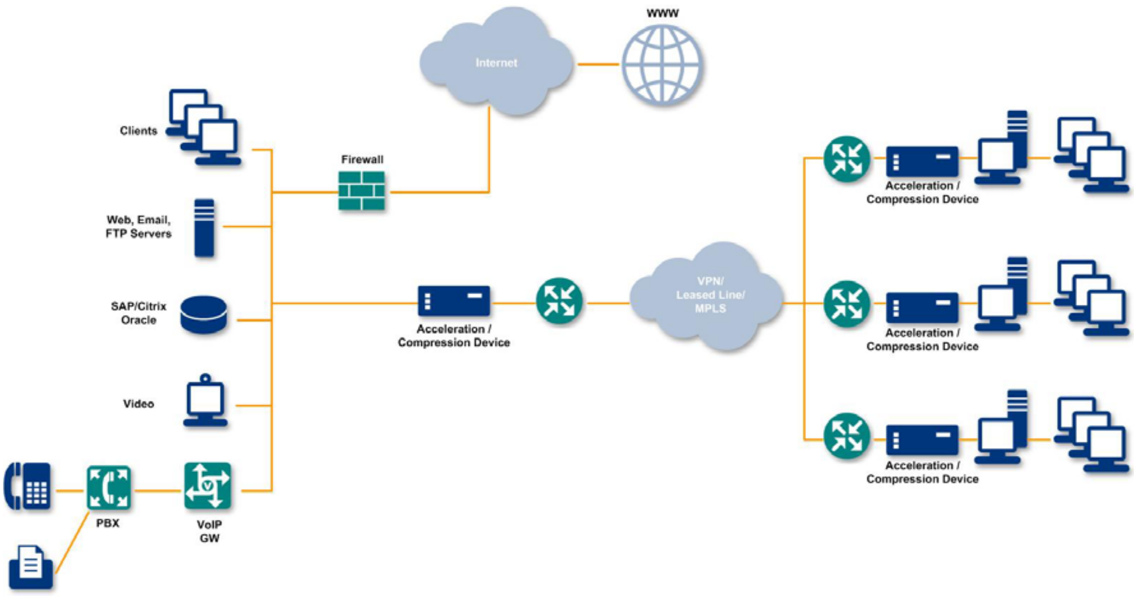

Diagram: Typical deployment of an acceleration or compression device on an enterprise WAN.

Benefits

TCP and HTTP acceleration technologies are effective in improving the performance mainly in terms of delay of certain applications over high-latency links. It successfully accelerates file transfers and can be useful in speeding up backups and synchronizations over the WAN.

Drawbacks

Opting with acceleration solutions to optimize an enterprise WAN is suitable for an enterprise environment which needs to prioritize the use of either TCP/ HTTP based applications only. However, as a great deal of recreational traffic also falls into this category, to be effective acceleration should be integrated with an additional optimization technique such as compression.

- **Partial solution**: Because acceleration technologies must be applied per protocol they are therefore usually integrated with other optimization tools.
- **Devices are non-discriminatory in terms of application**: Channeling an entire "protocol" of traffic into a tunnel that is collectively accelerated means that a great deal of recreational or uncritical traffic such as P2P, that uses the TCP protocol for communication will too be accelerated. What's more, a different acceleration technology must be applied for each protocol, which not only adds to the complexity of this process.
- **High TCO**: Two acceleration devices are required per connection point.
- **Low performance**: Performance of acceleration devices is also an issue due to a built-in limitation in cache memory used for the process.
- **Involves complex calculations**: In order to utilize acceleration technologies effectively, IT managers must calculate parameters such as the operating system's window size, latency, link size and more, for each link.

- **Stop-gap solution**: As applications are extremely dynamic and becoming adjustable to Internet and network environments, protocol acceleration presents at best a stop-gap measure.

3) CIFS Acceleration

CIFS acceleration was developed explicitly to overcome the limitations of the Common Internet File System (CIFS) file sharing protocol developed by Microsoft. The basis for the implementation of CIFS was developed during the 1980's (then known as Server Message Block) and has since become a standard that comes pre-bundled with all Microsoft-based client and server platforms. In addition to file sharing, CIFS is also used as a transport protocol for various higher level Microsoft communications protocols, as well as for network printing, resource location services, remote management/administration, network authentication (secure establishment services) and RPC (Remote Procedure Calls). This protocol is considered extremely "chatty" and inefficient.

CIFS acceleration technologies attempt to solve this problem by pipelining the data blocks, sending as many in quick succession as needed to fill the available WAN capacity. By transmitting more data across the WAN simultaneously without having to wait for individual acknowledgements, CIFS acceleration technologies reduce the amount of time required to complete a transaction, resulting in a considerable improvement in application performance.

Deployment

CIFS acceleration devices are deployed in pairs – one at the server side and one at the client end. Each device acts on behalf of a CIFS client (e.g. Microsoft XP computer) and server (e.g. Windows Server 2003 computer) to make the interaction between the two much more efficient.

Benefits

CIFS acceleration devices can be greatly beneficial in reducing the latency experienced by the CIFS client. They provide an excellent solution for two main CIFS-related interactions:

- File Access - file download, upload and remote access
- Directory Browsing - Moving around in a directory on a remote server

Drawbacks

- **Limited capabilities**: CIFS acceleration is designed expressly for Microsoft applications.
- **High Total Cost of Ownership**: Implementation of this technology requires deployment and operation per connection point.
- **Devices are non-discriminatory in terms of application**: Much like in the case of other acceleration technologies, CIFS acceleration does not differentiate between critical and non-critical applications and relates to a bulk of traffic traversing the network.

4) Traffic Shaping

Traffic shaping is founded on a proactive, policy-powered approach to WAN optimization. Traffic shaping devices control the traffic running at the bottlenecked links of a network—such as backbone link between the head office and a branch office — thereby optimizing the utilization of a network's infrastructure, guaranteeing and prioritizing bandwidth for business-critical applications and limiting bandwidth for less important ones. This ensures that mission critical traffic moves through the network causing business applications to respond more quickly. It also guarantees that the maximum amount of traffic is transmitted over a WAN's Internet connection in the most efficient manner possible—so that packets are not dropped or re-transmitted.

Most traffic shaping devices rely on deep packet inspection (DPI) technology to deliver granular visibility of the network. DPI enables real-time visibility of all traffic traversing the network and can therefore be used to monitor and analyze traffic flowing across the WAN. DPI also enables active per-user, per-application control of the traffic to correspond with an enterprise's business objectives. Enterprises can therefore, limit or even block the use of recreational applications such as P2P, at certain times or locations to ensure the optimal performance of mission critical traffic at all times.

Quality of Service (QoS) policies defines how bandwidth optimization is to be achieved. Each policy defines both the priority conditions for matching the traffic with policies and the network actions that need to be applied to the traffic when the conditions are met. Assume for a moment that increasing revenues is a corporation's top goal, and that implementing a new e-commerce capability is the most important relevant departmental initiative. Using a traffic management device, an organization's IT department can create a policy instructing the network that e-commerce users, applications and transactions are to be treated as a higher priority than any other activities.

In addition to prioritizing traffic, today's traffic shaping devices should also provide the following capabilities:

- Setting a minimum amount of bandwidth for an application/user ("guaranteeing")
- Setting a maximum amount of bandwidth for an application/user ("limiting")
- Enforcing a specific CBR (Constant Bit Rate) level for specific connections
- Allowing bursts of traffic on certain connections that exceed maximum defined limits
- Enabling hierarchical policies that ease policy creation and maintenance

Network managers can maximize bandwidth efficiency by predefining application priorities and bandwidth allocations. For example, they can establish FTP bandwidth limits during periods of congestion, or prioritize CRM traffic ahead of email.

Deployment

Various traffic shaping devices are deployed at different locations on the network. However, as oppose to other WAN optimization devices, in most cases, only one traffic shaping device is required per WAN link. Centralized deployment at the headquarters office or at the enterprise data center is a huge benefit in today's increasingly consolidated network environment.

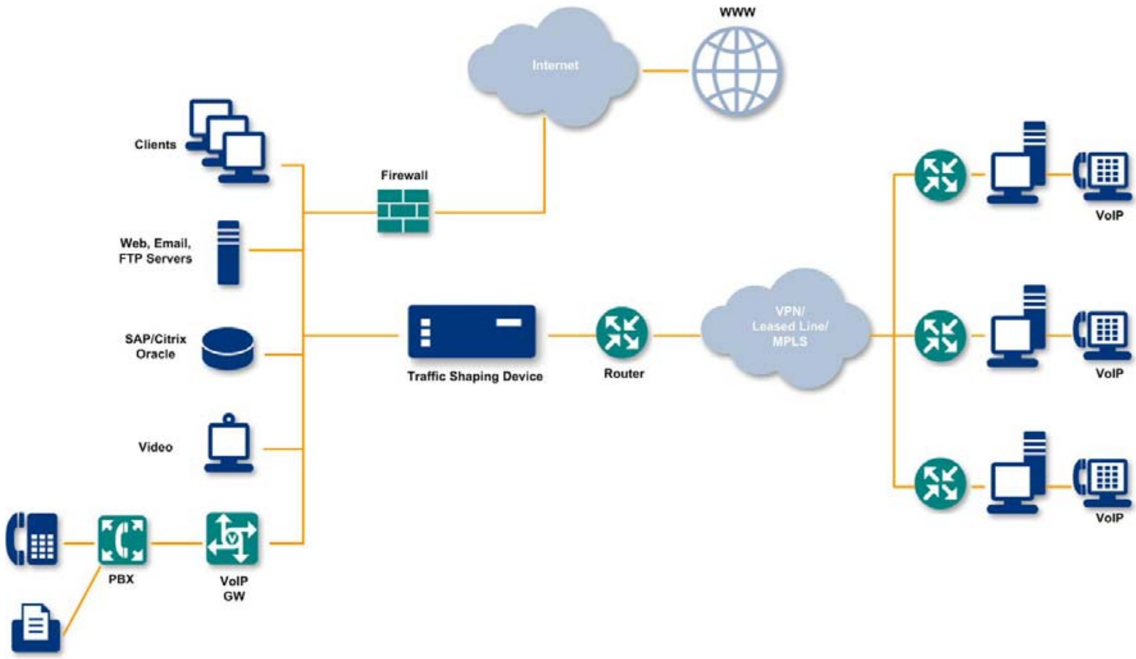

Diagram: Typical deployment of a traffic shaping device on a centralized connection point of the enterprise WAN.

Benefits

Utilizing QoS policies created by system administrators, traffic shaping devices enforce each network session to the appropriate policy, effectively instructing the device on what action should be taken. In this way, traffic management devices are able to fundamentally alter the network's natural first-in-first-out design so that an organization's most important activities are handled before lower priority transactions.

Equipped with DPI technology, traffic shaping devices deliver an unsurpassed visibility of an enterprises network. Furthermore, because deployment of a traffic shaping device only requires one traffic shaping device per WAN link location, these activities can be performed from a centralized location, allowing for one team to manage the traffic of an entire multi-branch network. Using centrally located devices coincides with the migration towards consolidated data centers witnessed in most enterprise environments today.

The real-time visibility enabled by a device's DPI technology greatly improves an IT staff's operational capabilities, allowing them to monitor, optimize and troubleshoot the entire bulk of network traffic from a centralized location. This saves significant investments in both network equipment and the resources required to operate them. Based on a highly-granular visibility of the WAN, IT managers can troubleshoot congestion and take immediate action to protect mission-critical traffic. Using traffic shaping devices makes it possible to define policies that guarantee performance during all usage periods and limit less delay-sensitive traffic like file transfers and curb or block recreational traffic.

Because traffic shaping devices can differentiate between the various applications and even users on the network, they can be used to control bandwidth-consuming and suspicious traffic on the network. They can be configured to set off alarms or isolate infected workstations in cases of connections-based (Dos/ DDos) attacks, allowing IT staff time to mitigate the potential damage that can be caused.

Drawbacks

- **Suitable for multitasking networks:** On dedicated links used for only a limited variation of applications/ protocols or mission-critical traffic only, traffic shaping could perhaps prove to be most effective when integrated with additional optimization technologies.
- **Requires ongoing configuration:** Traffic shaping is by nature a dynamic technique which requires ongoing configuration and management for its effective implementation. Optimal configuration reflects business objectives and must therefore be based on the accurate and granular visibility of a network's activity.

ROI: Performance Gain vs. Solution Cost

In determining the optimal WAN optimization solution that will deliver the greatest business value for a specific network environment, there are numerous factors such as IT budget, target applications and services for optimization, network topology and more. However, the proliferation of a variety of WAN optimization technologies is making it increasingly hard to evaluate the benefits of implementing a specific optimization technique on an enterprise's WAN. Whereas some techniques do appear to hold technical advantages, when applied to certain network environments they can often be extremely expensive to deploy and operate. What's more, with enterprise networks taking on more and more functionality, such techniques will usually materialize as only partial solutions to an ever-growing predicament. IT professionals are therefore challenged to choose a technique that will reap the most performance gain both immediately and in the long run, in proportion to the cost of implementation.

Subject	Compression	Acceleration	CIFS	Traffic Shaping
TCO	High	High	High	Low-Average
Application differentiation	No	No	No	Yes
Traffic volume	Low	Low	Low	High
Monitoring	No	No	No	Yes
Dos/DDos protection	No	No	No	Yes

ROI Compression

Investing in compression entails significant capital and operating expenditures. In a fairly typical network environment with a corporate headquarters and 15 branch offices, the average price of a high-capacity central compression device is $24,000, with branch office devices averaging $6,500. In this scenario, first-year capital outlays will total $121,500. In addition, ongoing operating expenses will average 15% of capital expenditure, or $18,225 per year for maintenance and support. Over 5 years, this deployment will incur total direct costs of nearly $213,000.

ROI Acceleration

The cost of implementing and operating an acceleration technology is approximately the same as with compression. With a 15-branch enterprise, at least 16 devices must be deployed and operated at each office independently. A five-year deployment of such a scenario will breed similar results in terms of both operational and capital expenditure. Due to the limitations of the technology in optimizing the entire traffic of an organization's WAN effectively, acceleration will in most cases be deployed in concurrence with additional optimization techniques, namely compression. This will also severely augment the original investment estimation.

ROI Traffic Shaping

Investing in Traffic shaping is much cheaper in the matter of TCO since instead of investing in many low end units the customer invest in one high end unit in the main branch. So in the same example as above with 15 branches, the customer will buy only one traffic shaping unit.

1G traffic shaping device will cost approximately $60,000. In this scenario, first-year capital outlays will total will be about $70,000 that will include about 15% maintenance costs. In addition, ongoing operating expenses will average 15% of capital expenditure, or about $9,000 per year for maintenance and support. Over 5 years, this deployment will incur total direct costs of nearly $105,000.

Conclusion

When taking into consideration the many variables related to a network's ideal WAN optimization solution, traffic shaping stands out as the most cost-effective. To begin with, deployment of a traffic shaping device does not require the installation of a branch office unit, regardless of the network's throughput. This significantly reduces both the Total Cost of Ownership and operational expenditures involved in what can often prove to be a highly complex installation project that consumes a great deal of man power and time. In other words, you need only one device, to control all applications and usage on the network, with greater efficiency and less effort. Furthermore, traffic shaping devices impart a new level of intelligence to the network infrastructure, which enables network operators to limit the impact of low-priority network activities and maximize the performance of high-priority transactions. They do not apply to any specific protocol or application and can be controlled on a per-application, per-user basis. More importantly, they are unique in providing a direct linkage between board-level strategic priorities, and the priorities of the IT infrastructure.