Important Considerations When Choosing A Wan Optimization Solution

Organizations are becoming increasingly dependent on effective delivery of IT applications, services and data over their wide area network (WAN). As a result, technologies that optimize the performance of applications over the WAN have become increasingly strategic. Early deployments of WAN optimization have been tactical in nature, addressing bandwidth or latency challenges across a selected number of WAN links. In these narrowly focused scenarios, the main selection criteria have been around the core features of traffic compression, and/or response time improvements. A WAN Optimization solution is an integral part of an application delivery network and to make the best choice customers should expand the selection criteria beyond core features and take a look at the entire application delivery infrastructure when considering a strategic network-wide deployment of WAN optimization. This document outlines the key considerations for selecting a WAN optimization solution for a network integrated deployment.

Consideration #1—Network Transparency

Network transparency is an important consideration for a WAN optimization solution. Transparent WAN optimization architecture applies optimizations to the traffic "payload", while preserving the original and critically-important IP and TCP headers. Non-transparent architectures on the other hand either encapsulate optimized traffic in "tunnels" or otherwise replace this critical information with something else, and therefore obstruct network services that depend on TCP headers.

A WAN optimization solution with a transparent architecture fully integrates into existing networks while preserving existing network services. As a result, network transparency protects the organization's investment in networks, and reduces ongoing operational expenses.

Network services which rely on such transparency include the following:

- **Access Control Lists**—ACL examine source or destination IP and other TCP information. They cannot operate if traffic is encapsulated into a transport tunnel
- **Firewall policies**—most firewall policies rely on examining source/destination IP addresses and TCP information. Transparent WAN optimization solutions support IP stateful inspection of optimized traffic and offer full firewall policies compliance. The use of traffic tunnels renders firewall policies ineffective. In effect, network administrators must 'punch a hole' in the firewall to allow optimization traffic through, and all optimized traffic is obfuscated, thereby defeating any existing firewall policies. Furthermore, any stateful inspection performed on this traffic is done against the tunnel packets and not against the original flow.
- **NetFlow statistics**—Most networking devices (e.g. routers, switches) can export traffic stats using the NetFlow interface. Analysis of NetFlow information is critical to monitoring and troubleshooting any network. Transparent WAN optimization solutions preserve the packet headers and therefore maintain the validity of NetFlow stats from any device along the path. Non-transparent solutions obstruct the NetFlow exports from any device along the path, as the router/switch only sees un-optimized flows and tunneled traffic.
- **Route selection**—Mechanisms such as Policy Based Routing (PBR), Performance Routing (PFR), and Dynamic Multipoint VPN (DMVPN) rely on packet header information and classification to determine route selection. Lack of transparency causes route selection to fail, as the network can only see un-optimized flows and tunnels between WAN optimization devices.

- **Quality of Service**—Quality of service is a means of differentiating traffic in the network with the intent of treating one type of traffic differently than another. Quality of Service Classification is performed on any aspect of a flow, including payload data, IP or TCP header information, VLAN tag, DSCP, or others. The network element must be able to see the original characteristics of the data, otherwise classification, and everything that follows, is broken. Devices that obfuscate data from being properly classified, due to lack of transparency, make end-to-end Quality of Service impossible.

A Transparent WAN optimization solution should be available without additional complexity and should be integrated with all other WAN optimization capabilities

Consideration #2—Interoperability with Application Performance Monitoring Tools

Measurement and monitoring of application performance is an essential part of daily network operations. Application performance monitoring tools highlight deviations from service level agreements (SLAs), and help troubleshoot problems related to application performance across the network.

Most application performance monitoring tools rely on the following data sources:

- **NetFlow statistics:** used to examine the traffic and application mix across multiple nodes in the network. NetFlow services capitalize on the flow nature of network traffic to provide detailed traffic accounting information with minimal impact on router or switch performance. NetFlow technology watches the traffic flows through network routers or switches and builds accounting data that characterizes the IP traffic being forwarded
- **TCP time-stamps:** through correlation of time stamps on TCP flows between client and servers, a monitoring tool calculates the overall transaction time, and the respective contribution of client, server and network delays.

For example, the NetQoS Reporter/Analyzer presents NetFlow statistics, while the NetQoS SuperAgent presents application response time, using TCP time stamps.

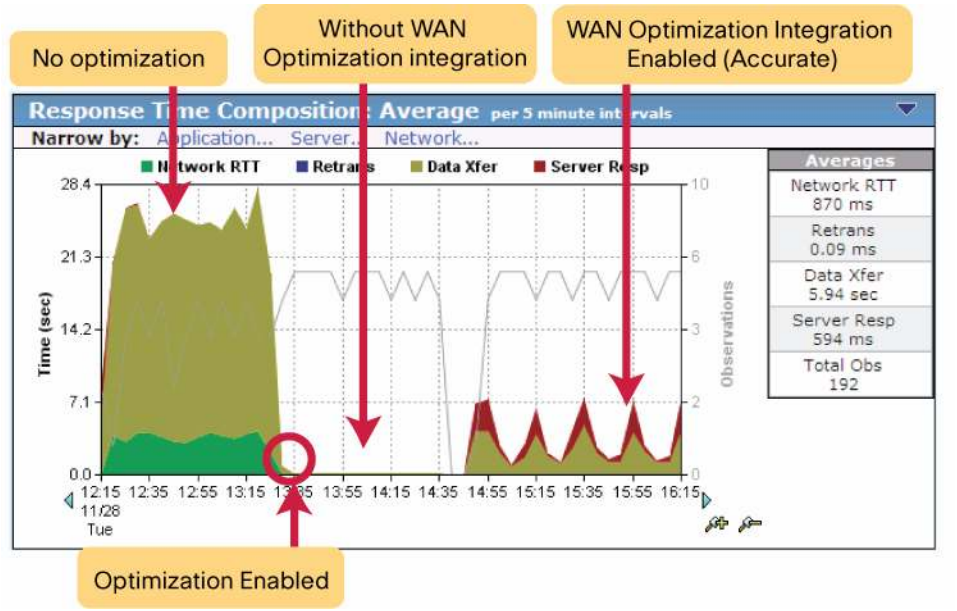
The interoperability between a WAN optimization solution and Application Performance Monitoring tools is an important consideration. NetFlow based interoperability is achieved through a transparent architecture as discussed above. The interoperability between WAN optimization and TCP response time tools requires additional interfaces. Fundamentally, WAN optimization divides the TCP flow between each client and server into three segments:

1. Between the client and the ‘branch’ WAN optimization device
2. Between the ‘branch’ and the ‘datacenter’ WAN optimization devices (across the WAN)
3. Between the ‘datacenter’ WAN optimization device and the application server

The division of the TCP flow into 3 segments presents a challenge to Application Performance Monitoring tools. Without proper interoperability, the presence of a WAN optimization solution would distort the results presented by an Application performance monitoring tool.

A WAN optimization solution should be integrated with performance management tools to measure end-to-end response time, from the client to the server over an optimized link. This requires the ability to export the necessary TCP/IP information to the performance monitoring tool. This ensures that end-to-end application response time reporting is accurate even in the presence of WAN acceleration devices that implement TCP proxy architecture. See Figure 1 for details.

Figure 1. Application Response Time Accuracy



Consideration #3—Partnership with Application Vendors

WAN optimization solutions accelerate critical business application traffic across the WAN. It is important that optimizations applied to application traffic are “safe”, and avoid risks of data corruption or changes to application behavior. To achieve this goal, WAN optimization vendors should establish partnerships with application vendors, license the respective protocols, and establish a framework for joint development, cooperative validation of solutions, and support for customers using the technology. WAN optimization vendors should avoid any risky “reverse engineering” of application protocols which may result in suboptimal performance, unexpected application errors, and limited support or optimization for less common application functions.

The combination of joint interoperability testing, proper licensing agreements and having escalation support in place with application vendors like Oracle, Microsoft, SAP and others minimizes the risk of application corruption and allows networking and application groups within IT organizations to confidently accelerate applications. Co-branded whitepapers and designs documents offer best practices and significantly reduce the risk of deploying WAN optimization to accelerate these applications.

Consideration #4—Interoperability with Network-Based QoS

Efficient delivery of latency and bandwidth sensitive applications over the network requires the use of quality of service (QoS) techniques. Those involve identification of specific application flows and prioritizing their respective traffic. Most routers and switches have built-in QoS support, thus enabling organizations to apply a network-based QoS policy. The successful application of network-based QoS involves the following considerations:

- Network devices along the path can fully “sense” and react to traffic congestion. Congestion avoidance mechanisms such as Weighted Random Early Detection (WRED) take advantage of TCP's congestion control mechanism. Therefore, QoS policies are best applied throughout the network, rather than in LAN-attached devices which can not see congestion points.
- QoS is applied as an end-to-end function—on LAN devices and WAN devices. It relies on metrics and functions being met and applied at each hop within the network. QoS should not be approached as a point-to-point function.

- Traffic is generally ‘classified’ at the edge of the network; however specific “prioritization” actions occur at each hop within the network based on ‘marking’, typically applied at the edge.

WAN optimization solutions should fully interoperate with network-based QoS. Here again, transparent architecture for WAN optimization is key for such interoperability. Any tunnel based implementations typically obstruct the operation of a network-based QoS. Furthermore, attempting to apply QoS policies within standalone WAN optimization devices may yield inefficient and unexpected results, as those standalone devices have little or no visibility into true network congestion points.

Consideration #5—Interoperability with Voice over IP (VoIP)

The transport of voice over IP data networks has been growing rapidly. The economic advantages of using a single network for application data and VoIP are extremely compelling. Since voice traffic is latency sensitive, successful implementation of VoIP requires extensive use of QoS policies throughout the network, Transparent WAN optimization solutions seamlessly integrate with network-based QoS, and therefore are fully interoperable with VoIP.

Transparent WAN optimization solutions integrate with the Cisco IOS-based QoS architecture to ensure reliable prioritization, bandwidth allocation, protection, and control of latency sensitive voice traffic. This results in a single, unified QoS framework, and eliminates the need for managing and correlating between two sets of independent policies.

Such interoperability allows the router to tag the VoIP traffic once correctly, rather than creating two sets of conflicting QoS policies. Time-sensitive VoIP traffic receives prioritization through existing router QoS policies and avoids any processing delay caused by WAN optimization. The result is better VoIP quality and performance.

Interoperability with Performance Routing (PfR) enables optimized route selection for application and voice traffic based on latency, packet loss, monetary cost, capacity and jitter of the various network links between the branch and the data center. This ensures that latency-sensitive voice traffic continues gets the best route when application acceleration is deployed.

Consideration #6—Secure and Trustworthy Handling of SSL Optimization

Secure Sockets Layer Version 3 (SSLv3), also known as Transport Layer Security Version 1 (TLSv1), is one of the most common protocols used to encrypt content transported over IP networks. The significant growth in use of SSL/TLS-secured applications, including both web-based and non-web-based applications, suggests the need to apply policy-based WAN optimization to the secured traffic. It is required to have SSL optimization capabilities that integrate fully with existing data center key management and trust. In such a solution private keys and certificates should be stored in a secure vault in a central location. The private keys and certificates should be distributed in a secure manner to the WAN optimization devices in the data center and stored in a secure vault, maintaining the trust boundaries of server private keys. SSL optimization should be fully transparent to end users and servers and requires no changes to the network environment.

Here are some important capabilities which should be available in SSL Optimization:

- **Simple, easy to deploy architecture**—The solution should allow creation of aggregated services with additional support for wildcard certificates and IP addresses.
- **Preservation of the trust boundary**—The solution should not distribute private keys beyond the secure data center WAN optimization devices.

- **Scalable secure storage of keys**—all certificates and private keys should be stored securely on a Central Manager and only distributed to the WAN optimization devices in the data center. The private keys should not be distributed to remote WAN optimization devices. A Central Manager should provide management of encryption services for all WAN optimization devices in the network, including the secure vault for encryption key-pairs and keys necessary for WAN optimization device disk encryption. All sensitive data used or generated by the WAN optimization devices should be stored—and transmitted—in a secure manner.
- **Disk encryption**—The solution should include the ability to enable selectively or globally disk encryption with keys managed by the Central Manager. This ensures that data written to the WAN optimization device disks is completely unusable should a WAN optimization device be compromised.
- **Interoperability with existing proxy infrastructure**—the SSL Optimization solution should provide full support for automatic identification, interception, optimization, and acceleration of SSL traffic even in environments where web proxies have already been deployed or where clients are configured to use explicit proxies.
- **Online Certificate Status Protocol (OCSP) support**—by providing support for OCSP a real-time security check of certificates can be used. The solution should support OCSP for real time revocation status of a certificate in compliance with the DoD Class 3 PKI definition. OCSP is also very useful where client certificates are used in the SSL handshake for Client Authentication.
- **Client authentication support**—The solution should provide full support for client certificate-based authentication during initial session establishment. By supporting this capability a verification check on the client can be performed before allowing the SSL session to proceed with the server. Client certificate authentication is commonly deployed in highly secure environments where message layer authentication mechanisms using userid and password, or a token, are not considered sufficient from a security standpoint.
- **Role Based Access Control (RBAC)**—The Central Manager RBAC framework should allow for controlled access to SSL configuration and monitoring.

Consideration #7—Intelligent Delivery of Live and On-demand Video

Many organizations leverage the power of video-based information delivery to employees over the network. The quality of the video being delivered has a significant effect on the ability to engage the viewer and impart the message. Previous efforts to deliver high-quality video messaging, either live or recorded on video tape or DVD, have proven expensive, time consuming, difficult to control, as well as difficult to maintain.

Network-based delivery of live video and video on demand (VoD) must be performed in a cost-effective way with limited impact on bandwidth. Video delivery should be done over the existing IP network, avoiding the use of dedicated content distribution networks.

The WAN optimization solution should provide an integrated and automated Video delivery solution for both Live and Video-on-demand content. This will enable efficient delivery of video to remote sites with minimal impact on bandwidth. A video delivery solution integrated with the WAN optimization solution can also save considerable bandwidth by allowing pre-positioning of high quality Video on demand (VoD) files in off-peak hours and by providing local availability of this content. Such a solution should provide interoperability with digital signage solutions to eliminate the need for additional devices to cache signage content.

Consideration #8—System Scalability and High Availability

With regards to wan optimization, response time must be the ultimate goal and shouldn't be degraded when the deployment size increases and scalability is required. A scalable WAN Optimization solution should provide the following capabilities:

- High Single Device System Capacity

  - WAN throughput of the Data Center Accelerator should support high speed links (such us multiple OC-3 links or OC-12) in order to minimize the number of devices needed to fully utilize the existing WAN infrastructure
  - Support for maximum optimized TCP sessions without compromising system stability
- Support for WCCP—which is a highly scalable (clusters/load balances up to 32 WAN optimization devices) and compatible with most routers and switches. Support for hardware acceleration of WCCP redirection from routers and switches is required.
- Integration with off the shelf load Balancers for high availability and large scalability
- Comprehensive Central management tool to manage and monitor high capacity of devices
- A tightly synchronized cache to ensure that response times are maintained even when data might be evicted from either the edge or core WAN optimization devices.

Consideration #9—Hybrid Caching Architecture

All WAN optimization solutions utilize cache to perform their acceleration function. There are two primary cache architectures utilized for WAN optimization:

- **TCP segment cache:** used for storing common patterns of TCP packet payload. The WAN optimization device continuously builds a history of such patterns. When new packets arrive, the WAN optimization device can recognize an already 'cached' byte pattern. When such a 'cache hit' occurs, the device transmits a small signature across the WAN that represents the known pattern. Its peer WAN optimization device will recognize the signature and recreate the traffic pattern using its own segment cache.
- **Object cache:** used for storing logical, large data objects—most notably files. An object, or file cache processes high level file operations such as read, write, copy, rename, etc. By caching file data at a remote site, the WAN optimization device can handle file-based operations without having to traverse the WAN. Note that file based protocols (e.g. CIFS) have built in mechanism to ensure file cache coherency, thus assuring the user always has access to the most recent copy of the data. Object caches can be preloaded (or pre-positioned) with file data based on administrative distribution policies. This enables the delivery of large files to remote sites at off-peak hours.

Hybrid cache architectures combine the benefit of both TCP segment cache and an object cache. The benefits of the Hybrid Cache Model include:

- A consistent application response time and consistent remote user experience
- High service quality levels for common Enterprise applications
- Enterprise-class scalability and deployment flexibility-from two sites to thousands of sites
- Reduced head-end solution footprint
- Offloading of the WAN head-end devices and servers to support server consolidation initiatives

Consideration #10—Acceleration of Data Center to Data Center Applications

Optimization of data protection applications over the Wide Area Network (WAN) between data centers—including replication—brings challenges and requirements that are different from the optimization of application traffic for remote users. While remote users care about reducing the amount of time it takes to access a specific file or application over limited-bandwidth WAN connections, storage administrators responsible for Disaster Recovery operations are mostly concerned about reducing the time required for backup/replication windows, and the inability of storage array systems to fill a high-throughput WAN connection due to latency, as well as the high cost of bandwidth for data replication.

A datacenter-to-datacenter (DC-DC) optimization solution should be part of any WAN optimization solution. It should be managed and monitored by the same central management tool as the branch to Data Center WAN optimization devices in order to have a single view of the overall network optimization.

Such DC-DC solution should be tested and certified by the main storage vendors such as EMC and NetApp, These suite of tests ensures that IT organizations can safely consolidate distributed servers into the data center while maintaining performance expectations without compromise to data integrity or security.

Consideration #11—Router-integrated Solution

A WAN optimization solution allows IT departments to centralize applications and storage in the data center, maintain LAN-like application performance and reduce the overall floor space consumed by branch-office devices.

A router integrated solution reduces capital expenditures (CapEx), operating expenses (OpEx), and support costs compared to a stand alone appliance. By integrating the WAN optimization solution into an existing network device, such as the Cisco Integrated Service Router (ISR) organizations can reduce the device footprint in the branch office even further as well as reduce the total cost of ownership for the WAN optimization solution.

Consideration #12—WAN Optimization for the Mobile Worker

Acceleration of mobile VPN connections over the public Internet brings different technical challenges compared to corporate WAN and branch-office optimization such us:

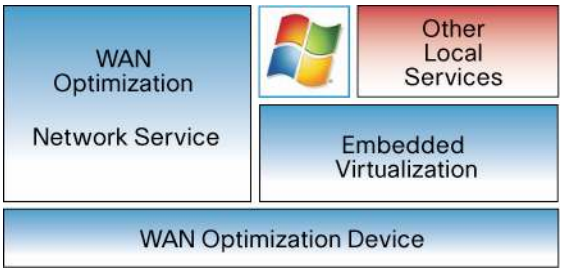
- **Handling lower-quality network connections**—Mobile users use Internet connections such as DSL, Wi-FI, Satellite, dial up, cable and cellular which have lower bandwidth, high packet loss, latency and additional challenges such as time-slicing delay in cellular environments
- **Integration with Desktop Windows environments**—WAN optimization for mobile users must share laptop or PC computing resources with many other PC applications while branch office users can rely on a dedicated WAN optimization device. The open environment of a Windows PC, in contrast to the controlled environment of an appliance, has a very different class of interoperability requirements, with numerous applications running on the PC: a variety of operating systems, browser versions, end-point security applications, and VPN client software and a wide range of business applications.
- **Integration with Windows management tools**—WAN optimization for Windows clients should be managed in a similar way to any other desktop utility. Integration with the Windows management platform is important in order to lower support costs and streamline ongoing operations

A comprehensive WAN optimization solution should include an optimization solution for the mobile user. A Mobile optimization solution should provide acceleration in challenging network connectivity conditions as described here. It should have a small footprint on the PC and must be easy to deploy. A mobile optimization solution should be compatible and validated with all VPN client options including IPSec and SSL based VPN's, and finally it should natively integrate into a Windows management environment to deliver the best performance and compatibility.

Consideration #13—Network-embedded Virtualization

Remote sites often include small Windows servers that host local applications and services. WAN optimization can enable the migration (or consolidation) of many of those services into a datacenter, thus lowering costs and improving manageability. There are however scenarios (e.g. Print) where services must be available locally, causing one or more dedicated servers to reside at the remote site.

Figure 2. Network-embedded Virtualization



Virtualization technology has rapidly gained adoption within IT organizations helping reduce costs and increase services agility. Running applications and services on virtual machines has become the norm within the datacenter. A similar virtualization approach can be applied to branch offices and remote sites. A solution that integrates virtualization with WAN optimization offers the benefits of application acceleration, improved network utilization and local services hosting.

Unlike general business applications, WAN optimization is a real-time network service designed to run natively on the underlying device hardware. General purpose server virtualization solutions do not support this requirement. Network-embedded virtualization architecture supports the native aspects of WAN optimization, while adding virtual machine support for locally hosted services.

Summary

WAN optimization has evolved from a tactical deployment into a strategic one. Improving network utilization and increasing application performance remain important considerations, but not the only ones. The additional considerations described above must be taken into account when planning a strategic, network-wide deployment of WAN optimization.

Integration with the existing network platform and interoperability with other services and applications are paramount to any strategic deployment scenario. The selected solution should be based on an architecture that can scale to meet future network expansions. The selected solution should address the needs of remote and mobile users, and integrate with data center applications and services. When strategically deployed, a WAN optimization solution enables IT organizations to reduce costs, improve productivity and consolidate remote site infrastructure.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-557825-00 08/09