

Distributed systems optimization: A technology primer white paper



Introduction – distributed systems optimization (DSO).....	3
Overview of WAN environments.....	3
Limited bandwidth.....	3
High latency.....	4
Transport protocols – chattiness and fixed window size.....	4
Chatty application protocols.....	4
Why is DSO important?.....	5
Options for DSO.....	5
Deploy local infrastructure.....	6
What is it?.....	6
Benefits.....	6
Drawbacks.....	7
Network compression.....	7
What is it?.....	7
Benefits.....	7
Drawbacks.....	7
Caching (web, file and email).....	8
What is it?.....	8
Benefits.....	8
Drawbacks.....	8
Transport protocol optimization.....	8
What is it?.....	8
Benefits.....	9
Drawbacks.....	9
QoS (quality of service) or rate-shaping.....	9
What is it?.....	9
Benefits.....	9
Drawbacks.....	9
Thin client.....	10
What is it?.....	10

Benefits.....	10
Drawbacks.....	10
HP technology's "transaction acceleration"	10
What is it?	10
Benefits.....	11
Summary	11
For more information.....	12

Introduction

Enterprises rely on distributed systems for a wide array of critical applications and business functionality. From file sharing to email, from web-based applications to storage, distributed systems and the wide area networks (WANs) they run on are a fundamental part of any business today.

Ever since WANs were first deployed though, whether they are based on private networks, virtual private networks (VPNs), or the public Internet, they have been subject to the fundamental challenges of moving data quickly over long distances. Simply put, WANs pose a challenging environment for applications for a variety of reasons:

- **Limited bandwidth**—WANs interconnecting remote branch offices (RBOs) typically have well under 1% of the bandwidth available on local area networks (LANs). Therefore, traffic that flows nicely on LANs can be badly congested on WAN links.
- **High latency**—The latency, or round trip time, on a WAN is often hundreds of times higher than on LANs where it is nearly zero. Applications that generate thousands of client-server interactions, also known as “round trips,” can run fine on LANs. However, they can break down or stop completely on a WAN, even if there is plenty of bandwidth.
- **Chatty transport protocols**—TCP, the universal transport protocol used by most enterprises for most applications, was not developed with high latency, low bandwidth in mind. TCP fundamentally limits the throughput of WAN traffic.
- **Even chattier application protocols**—On top of TCP, many of the most common applications rely on their own protocols (for example, Microsoft® file sharing uses the Common Internet File System protocol [CIFS]).

Over time, a host of technologies has been developed to try to overcome these issues and improve the throughput (also known as response time or performance) of distributed systems.

Some address limited bandwidth, some address file sharing, some address web performance, and others optimize specific protocols. This white paper outlines the primary alternatives available today. Although they all work in different ways, because all of these techniques are intended to improve the performance of distributed systems, it is helpful to compare and contrast the different approaches.

At the end of this paper, the approach to Distributed Systems Optimization (DSO) taken by HP and how it compares to the other alternatives is explained.

Overview of WAN environments

Before discussing the specific optimizations offered by different technologies, the effect of each characteristic is described in detail.

Limited bandwidth

Bandwidth is the easiest to understand, so easy in fact that many enterprise IT professionals believe that most or all of their application problems stem from not having “enough bandwidth.”

Despite the conventional wisdom that “bandwidth is free,” it is not. The aggregate cost of bandwidth to remote offices adds up quickly, even for U.S.-based companies. If an enterprise has overseas offices, those links can easily be \$10,000 per month or more for private T1s. Even when companies switch to VPNs, when they operate hundreds of remote offices, the total cost for bandwidth is significant, and upgrading all links can be prohibitively expensive.

The fact is that most enterprises rely on relatively low-bandwidth links to their remote offices, often in the range of 128 kbps to 1.54 Mbps (T1), running on a variety of transports (frame relay, VPN, ATM, satellite, and so forth). Compared to LAN bandwidth of 100 Mbps to 1,000 Mbps, WANs are stuck with 0.1% to 1.5% of the bandwidth available on a LAN. Obviously, sending large files through thin WAN pipes is a problem.

High latency

The issue of high latency on WANs is not as well understood. On a LAN, where the latency is essentially zero, data can flow almost unimpeded, and users perceive almost no throughput delay. On a WAN though, instead of latency in the range of 0.1 milliseconds, it can easily be 50, 100, or 200 ms. On a satellite link, it may even be a full second (1,000 ms).

Simply put, high latency limits how much data can be transmitted through a given link, no matter how much bandwidth is available. Even with a “Long Fat Network” or LFN, many customers find they cannot fill the pipe. This issue is fundamentally interrelated to the chattiness of both transport and application protocols, which are discussed in the following sections.

Transport protocols—Chattiness and fixed window size

Communication protocols, both transport and application, typically have a fixed or maximum window size. TCP for example has a maximum window size of 64 KB. That means that on each round trip, TCP allows a maximum of 64 KB to be sent. If you are sending a 64-MB CAD file, for example, TCP might require 1,000 round trips to complete that transaction.

Each TCP packet or frame of data cannot be sent until the client sends an “ack” (an acknowledgment) that it received the last one successfully. That means that there would be 1,000 serial trips, one after the other. If the round trip time between New York and San Francisco is 100 ms, then that transaction is going to take 100 seconds to complete, no matter how much bandwidth is available.

Chatty application protocols

In addition to the transport protocol that underlies enterprise applications, many applications have their own protocols that ride on top of TCP. Windows® relies on the CIFS protocol, for example, and Exchange relies on the Messaging Application Programming Interface (MAPI) protocol. In many cases, these protocols have two problems:

Application protocols are often more chatty and inefficient than TCP. There are often hundreds of round trips generated by the application for reasons that no longer make sense, but that are deeply embedded into the protocols for historical reasons. For example, in a simple drag-and-drop from a remote Windows file share, CIFS can generate literally thousands of round trips, no matter what TCP does. Application protocols often have even smaller window sizes than TCP, sometimes as small as 8 or 16 KB compared to the 64-KB maximum window size for TCP. That means a given transaction will require more round trips because the application can finish a smaller fraction of the job with each trip.

That means that even if you got rid of the inefficiency of TCP, many of your most common applications would still be constrained by these application protocols.

Why is distributed systems optimization (DSO) important?

There are three primary drivers behind DSO:

- **Higher productivity**—Particularly in data-intensive sectors like engineering, manufacturing, architecture, software development, and semiconductor design, sending substantial amounts of data around the world is mission critical. Sending a 100-MB CAD file from Detroit to Munich can literally take hours on a typical enterprise WAN link. Obviously, sending the same data faster directly benefits users through higher productivity, less rework and “wrong-work,” and better collaboration. Some, but not all, DSO products can improve the throughput of applications on WANs.

- **Site consolidation**—In today’s business climate, most large enterprises are realizing that deploying, managing, and maintaining hundreds or even thousands of servers, tape backup systems, and the associated IT staff and other expenses associated with those distributed systems means very high costs and risks. There is a strong desire to physically consolidate IT infrastructure to datacenters, but users are then subjected to the poor performance previously discussed because of the WAN between their client and the newly centralized servers.

To enable site consolidation, you have to improve WAN throughput until it feels “LAN-like” for all the applications you want consolidated.

- **Bandwidth optimization**—Since WAN links are expensive, particularly overseas, and since in many places you cannot get more bandwidth even if you are willing to pay for it, most CIOs are anxious to do more with what they have. Most DSO products can help squeeze more bandwidth out of your links, but many of them will not necessarily improve throughput much.

Remember that optimizing or expanding bandwidth may not help throughput, especially on high-latency links or where chatty applications are being used (like Windows).

Options for DSO

These are not new problems. They have existed from the first time someone separated clients from servers. The problems are caused by the facts that underlie distributed systems. But, because the problems have been around for awhile, there are a host of technologies available to improve the performance of distributed systems, all of which fall into the category of DSO.

Some of the solutions address bandwidth, some address transport protocols, some attack the problem through caching, and others by prioritizing traffic. It is important to understand how they all fit together, what each technology can do, and most importantly, what each cannot do for your business.

As an introduction, there is a limited set of techniques available to improve the performance of applications over WANs. You can do one or more of the following:

- Have distributed infrastructure (deploy servers and backup in remote offices)
- Make the pipes bigger (buy more bandwidth)
- Send less data (deploy WAN optimization or network compression)
- Send important data first (implement QoS technology)
- Store local copies of data (deploy file caching, web caching)
- Use file systems, Wide Area File Systems (WAFS), or email caching
- Make TCP more efficient (modify the transport protocol, for example, tweak TCP to use big windows)
- Use thin clients (deploy a thin client solution instead of running client-server architecture across the WAN)

The following table summarizes the various technologies available and shows what techniques each of them use.

Table 1. Technologies and techniques

		Options available to optimize distributed systems					
		Add or expand bandwidth	Reduce WAN traffic	Store data locality	Prioritize WAN traffic	Optimize transport (TCP)	Optimize app. protocol
Available products/technologies	HP	X	X	X		X	X
	Distributed infrastructure		X	X			
	- File servers						
	- Network attached storage (NAS)						
	- Mail servers						
	- Tape auto-loaders						
	Higher bandwidth	X					
	Network compression	X	X	X	X	X	
	Caching		X	X			
Distributed servers		X					
QoS	X			X			
Modified protocols					X		
Thin client		X					

Deploy local infrastructure

What is it?

This is the existing path that most companies follow. A typical remote office will have its own file server and tape autoloader. It may have NAS or a mail/messaging server (like Notes or Exchange). Over time, as servers became cheaper IT managers deployed servers in remote offices because it appeared to be the least expensive way to deploy services and applications.

Benefits

The primary benefit of having local IT infrastructure is that users get high performance, meaning quick response time when opening, sharing, saving files, or when using locally hosted applications like email. Since local file servers are on the LAN, typically users are connected with 100-Mbps or 1000-Mbps Ethernet with nearly zero latency.

Drawbacks

The key drawbacks of deploying local infrastructure have to do with the cost of maintaining and managing the servers and associated software and backup systems. Software vendors release a steady flow of patches that need to be applied to all distributed servers, so the more you have, the more patching is required. Backing up remote file servers typically involve tape autoloaders, off-site tape storage, and potentially online backup services.

Many times, the job of managing the process of tape backup is delegated to local employees who are not fully trained, and who have other responsibilities. The success rate for restores from local tape backup systems is typically less than 50%. Lastly, the simple job of keeping the servers running is often far more expensive than the original cost of the servers.

Another huge drawback to maintaining local file servers in a collaborative environment has to do with managing data, particularly large files that need to be accessed by colleagues working from different offices (for example, CAD/CAM files, software development source trees, semiconductor designs, and so on). The risk of rework or “wrong-work” because of mix-ups about which file is the most recent one can cost companies significant time and lost productivity.

Network compression

What is it?

There are a variety of network compression appliances available. The basic idea behind them is a form of “dictionary”- or “code book”-based compression that looks for repeated patterns in data flowing across WAN links. By recognizing repeated patterns and substituting “codes” that are smaller, the devices can effectively squeeze WAN traffic to consume less bandwidth.

Some network compression appliances use disk in addition to memory to store more of those repeated data patterns, which lets them “look back in time” to find patterns that occurred days, weeks, or months earlier. The key for this part of the technology is to measure how much data can be represented by the “address” or “reference.” If that ratio is limited to 10:1 or 20:1, then there is a ceiling on the effective compression ratio that can be delivered.

Benefits

Network compression will save WAN bandwidth and may help defer a WAN upgrade. The compression ratios achieved by network compression products vary widely among manufacturers and the workload specific to your company, but generally they will reduce traffic by something like 20% to 90% (or conversely, they will improve or expand bandwidth by 5 to 10 times in the best case. Expanding bandwidth should **not** be confused with improving throughput.

As network compression devices add more memory and disk, their functionality begins to approach that of a cache. Over time, the distinction between compression and caching will become blurrier.

Drawbacks

Network compression has little effect on application performance beyond the increase generated by freeing up some bandwidth. If WAN link is completely full, then congestion may be the primary issue affecting throughput, and relieving it will have a material impact. However, when there is sufficient bandwidth to relieve the congestion-induced slowdowns, then application throughput is often governed by the behavior of the transport protocol (TCP) or the application layer protocol (for example, CIFS, MAPI) in high-latency environments.

To address that issue, some network compression vendors have added TCP optimization between their appliances. As discussed in the TCP optimization section, TCP is the limiting factor in application performance, but will not offer much help if there is a higher level application in use that is chattier (for example, Windows or Exchange).

Caching (web, file, and email)

What is it?

The idea behind caching is to store data in a remote site so that subsequent requests for the same data do not have to be fulfilled by the origin server, which is primarily back in the datacenter. There are several flavors of caches: Web caches, file caches, and email caches. File caches are referred to by some analysts as Wide Area File Systems (WAFS) since they enable wide area access to centralized files through the caching mechanism.

Distributed File Systems (DFS) can also be implemented, though most companies find replacing their standard file system to be unattractive.

Benefits

Caches reduce bandwidth consumption and improve response time for applications on WANs, but only within a narrow set of data types. For example, a file cache will improve response time for data stored in file systems (Windows, for example), but will have no impact whatsoever on the performance of web-based applications or email. Conversely, a web cache will improve the performance of web-based content, but will not improve the performance of file systems.

Drawbacks

Caching is not a general solution for application performance, or for site consolidation. It will not help with a general re-architecture of a distributed system at an enterprise because of its application specificity.

Caches also only help with identical content requests. For example, with a file cache, the response time will be improved when the client requests a file that has already been across the WAN before and is therefore stored in the local cache. A request for a file with a different name, even if the content is identical, will result in a cache “miss,” and the entire file will be sent across the WAN again.

Lastly, caches are fairly complex to deploy and administer because they are proxy servers, acting on behalf the actual or “origin” servers. Proxy servers require that IT configure every client so that they know about the proxy and are redirected there first.

Transport protocol optimization

What is it?

Most enterprise applications ride on top of the widely deployed transport protocol known as TCP. The exceptions are primarily Voice over IP (VOIP) and traffic like video, which is typically sent on UDP.

While TCP is a highly reliable and robust protocol, it has some drawbacks when it comes to WANs. Two issues that affect maximum throughput are the “maximum window size” for TCP (normally 64 KB) and the fact that TCP requires an “ack” from a remote host before sending the next packet in a sequence.

These issues are well known and a host of companies have been formed with various optimizations for TCP, either introducing larger window sizes and/or spoofing acks to enable subsequent packets to be sent earlier than would otherwise be possible. Some companies also terminate TCP to substitute a different transport layer protocol, usually UDP.

Benefits

TCP protocol optimization can have a big impact on throughput of applications that are bound by TCP chattiness and window size because they can enable you to “fill the pipe,” especially on an LFN. This is most useful when the content is huge and non-repetitive, and when the bandwidth is large.

Drawbacks

Optimizing TCP or using a non-TCP protocol will not affect the throughput of an application whose “application protocol” is less efficient than TCP. For example, Windows relies on CIFS, which is far chattier than TCP. Therefore, even if you were to optimize TCP perfectly, the throughput of any application that is dependent on moving data with the Windows file system will be bound not by TCP, but the behavior of CIFS over the WAN. Relying on optimized TCP is not going to be as effective under the following conditions:

- The content is iterative (so removing redundant traffic is more advantageous than sending the same traffic quickly a second time)
- There is an application protocol involved (for example, Windows) that gates throughput independent of the TCP protocol optimization
- The bandwidth is limited, so “filling the pipe” does not help (because even a 100% full thin pipe is not going to be very fast)

In those circumstances it is preferable to avoid sending data in the first place rather than to “fill the pipe.”

Quality of service (QoS)

What is it?

The basic idea behind QoS is to detect packet characteristics and assign priority based on application or protocol type, reserving the highest priority for the data belonging to the most critical applications.

Benefits

If WAN links are shared by a range of applications, some of which are mission critical like SAP or Oracle®, and others that are perhaps less critical like web traffic or .mp3 file sharing, then you can give higher priority to that essential traffic, thereby improving the response time for users of those applications.

Drawbacks

QoS does not materially improve the throughput of applications above what they could achieve if bandwidth was 100% reserved for them. In other words, adding QoS may avoid disaster by de-prioritizing non-essential traffic, or it may improve the throughput of essential apps if they are already adversely affected by that traffic. But, QoS in and of itself will make little difference to the raw throughput of an application like Windows or Exchange over WANs because they are still fundamentally gated by the bandwidth available and the latency of the link. Or, to sum up, QoS only helps when there is congestion.

Thin client

What is it?

Many companies have decided that centralizing application servers is an economical way to deploy their IT architecture. With thin client technology, an application like Microsoft Word may be deployed on a central server, and the client accesses the application over the WAN, simplifying the IT administration of the application.

Benefits

The benefits are primarily lower cost and complexity for IT. By not having applications loaded on every client PC, the job of maintaining, updating, patching, and so forth is much simpler.

Drawbacks

Thin clients are most appropriate when the application is not particularly interactive (like basic Office applications, data entry, or simple client-server apps). If the application is highly interactive (like CAD/CAM applications), or if the latency is too high (usually over 30 ms or so), or if the workflow requires that the users save, print, or share files across the WAN, then thin client is not the best solution.

Other drawbacks are that thin clients maintain a fixed amount of bandwidth per client, regardless of the actual work going on, usually about 18 kbps per client. So, if you have 10 people in an office using a thin client, you need 180 kbps reserved for them all the time. Thin clients also typically limit the color depth information that is sent across the WAN to conserve bandwidth, which is another reason they are not normally appropriate for design or engineering applications.

HP “transaction acceleration”

The following sections describe the HP approach and how it differs.

What is it?

HP transaction acceleration is a revolutionary new approach to the issue of application throughput on WANs. Transaction acceleration is deployed on HP StorageWorks Enterprise File Services (EFS) WAN Accelerators.

Transaction acceleration combines several patent-pending innovations into one solution:

- Remove all redundant WAN traffic. By storing all WAN traffic on disk using HP Scalable Data Referencing (SDR) algorithms, virtually all redundancy is removed from your WAN traffic. Unlike caching products, SDR optimizes all TCP-based WAN traffic.
- Minimize turns generated by TCP. Using Virtual TCP Window Expansion (VWE), HP can reduce the number of round trips, or turns, generated by TCP, which in turn increases the throughput of TCP-bound applications like FTP. Unlike modifications to TCP, VWE involves no programming or changes to your underlying network configuration.
- Minimize turns generated by applications. Separate from VWE, HP optimizes specific applications even more through Transaction Prediction, which minimizes the number of round trips across the WAN generated by applications like Windows and Exchange.

Benefits

The HP approach delivers a range of benefits in one product.

- **Higher throughput for applications on WANs**—By combining the three techniques previously outlined (SDR, VWE, and Transaction Prediction), EFS WAN Accelerators can improve throughput by up to 100 times.
- **Multi-layer latency optimization**—Only HP implements latency optimizations both for TCP and application layer transports like CIFS and MAPI.
- **Unparalleled bandwidth reduction**—With bandwidth reductions up to 400:1 on single transactions, EFS WAN Accelerators can reduce bandwidth consumed by applications on WANs far more effectively than any other solution. Unlike other hardware or disk-based compression appliances, the HP hierarchical approach, SDR, means much higher bandwidth reductions than normally available.
- **Flexibility**—With WAN throughput that approximates LANs, IT architects have the option to design distributed systems in the way that makes the most sense. For example, it may be desirable to centralize specific servers like Microsoft Exchange servers, but without LAN-like WAN throughput, the impact on users is unacceptable.
- **Transparency**—Unlike other proxy-server based technologies (such as caches), EFS WAN Accelerators are transparent, requiring no configuration at the client or the server.

Summary

Distributed systems are a fundamental piece of enterprise infrastructure. Despite years of innovations in various areas of WAN performance, the throughput of applications on WANs has historically been quite poor.

While compression, caching, QoS, and other technologies have their place, until now no solution had solved the problem of WAN latency for the general case. HP not only delivers best-in-class WAN optimization, but also optimizes the behavior of both transport and application-level protocols on high-latency wide area links.

HP EFS WAN Accelerators deliver LAN-like throughput to a wide range of applications, including web-based applications, Exchange, Notes, file sharing, backup, and FTP. The products can be used to speed up the performance of applications or to enable site consolidation of such distributed systems as file servers, Exchange servers, Lotus Notes servers, tape backup systems, and NAS.

For more information

For more information on HP StorageWorks EFS WAN Accelerators, visit:

<http://www.hp.com/go/efs>

For additional HP StorageWorks EFS Accelerator white papers, visit:

<http://h18006.www1.hp.com/storage/efswhitepapers.html>

© 2005 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. Oracle is a registered trademark of Oracle Corporation and/or its affiliates.

5983-2462EN, 05/2005

