IT site consolidation white paper

How HP StorageWorks Enterprise File Services (EFS) WAN Accelerators enable consolidation of remote office Microsoft Exchange servers

Introduction

At many large enterprises, CIOs are making the strategic choice to consolidate as much of their IT infrastructure as possible. One of the key pieces of infrastructure targeted for consolidation is Microsoft® Exchange servers. Exchange often performs poorly over wide area networks (WANs) due to the way the application protocol works (the Messaging Application Programming Interface [MAPI]).

Despite the strong desire of CIOs to centralize Exchange servers, doing so can lead to poor performance for end users in remote sites. This white paper explains how HP StorageWorks Enterprise File Services (EFS) WAN Accelerators can enable IT architects to centralize Exchange servers without giving up the high performance their users expect and need.

The drivers for site consolidation of all remote office IT infrastructure, including Exchange servers, include reducing cost, simplifying IT, and improving compliance and security. It is just much easier to manage servers, storage, and backup if they are deployed in a few highly secure places instead of hundreds or thousands of sites all over the globe.

Historically, the downside of "site consolidation" has always been that end users are severely impacted—in short, removing local infrastructure usually means that the performance for remote users becomes completely unacceptable. HP recently introduced its line of IT appliances to improve the performance of virtually any client-server application on WANs. The increase in throughput can often be up to 100 times, or roughly equivalent to local area network (LAN) throughput. EFS WAN Accelerators improve throughput for applications as diverse as Microsoft Windows® file systems, Lotus Notes, Microsoft Exchange, network attached storage (NAS), backup applications, virtually any HTTP (web)-based enterprise application, as well as TCP-based applications like FTP.

With that kind of "LAN-like" throughput, moving infrastructure from remote sites to datacenters can be done without disrupting end users, which means IT architects can achieve their key goals:

- Reducing cost and complexity
- Improving compliance
- Improving security
- Improving resource utilization

HP can help IT architects centralize some or the entire following IT infrastructure from remote sites:

- Microsoft Exchange servers
- File servers
- NAS
- Lotus Notes servers
- Tape backup systems (autoloaders)
- Associated IT services (that is, off-site backup, outsourced server management, and so on)

Can you consolidate Exchange servers?

While many companies are trying to consolidate infrastructure as much as possible, many are surprised at how difficult it is to actually complete a successful site consolidation project, including Microsoft Exchange messaging servers. The reason is that when WANs are involved, applications that worked fantastically on LANs break down and work poorly, or not at all. Many IT architects characterize the issue as a "bandwidth problem," when in fact it is often much more than just bandwidth.

Over time, CIOs have gone to a distributed systems approach, which means giving each remote site their own local servers, storage, and backup systems to work around the WAN performance problem. At many companies, the threshold for deploying a dedicated local Exchange server is 20–30 employees. Since Exchange servers are supposed to have a capacity of several thousand users, deploying a dedicated server for a few dozen means you are vastly over-provisioning infrastructure. On top of that initial investment, all those servers have to be managed, backed up, repaired, and patched. Life would be much simpler if those servers were not needed at all, yet IT could just deliver services with a centralized architecture.

So, why can't you just take out those Exchange servers and move them to larger, more highly utilized servers to your datacenter? The reason is that in addition to the issue of WAN bandwidth, which is almost never high enough at remote offices, all applications rely on underlying communications protocols. In most cases, enterprise applications run over TCP.[1]

On top of TCP, applications have their own communications protocols. For example, Microsoft Windows uses the Common Internet File System (CIFS). Microsoft Exchange uses MAPI. Web-based applications rely on HTTP, and so forth.

Most protocols (application or transport) are extremely "chatty," which means they generate hundreds or even thousands of round trips from client to server, even to accomplish seemingly simple tasks. For example, dragging and dropping a 1-MB file in Windows can trigger over 4,000 WAN round trips. On a LAN, when the latency between client and server is often less than a tenth of a millisecond, those thousands of round trips are completed virtually instantaneously. When the same operation is done on a WAN though, the latency is usually in the range of 50 ms to 250 ms, or even more when satellites are involved. As a simple example, take a hypothetical transaction:

Table 1. LAN versus WAN time to complete—1-MB file drag and drop in Windows

	LAN	WAN
Latency (ms)	0.10	100.00
Number of round trips	4,000	4,000
Time to complete (ms)	400	400,000
Seconds	0.4	400
Minutes	0.01	6.67

That difference, 0.4 seconds to complete versus almost 7 minutes, is why "just moving the servers" does not work—the users notice. Application protocols also have a limited amount of data they can carry on each round trip. So the problem of many round trips is worse for large files. If the application protocol has a "max window size" of 16 KB, then a 16-MB file will require 1,000 trips, just to deliver the data, plus lots of additional round trips generated by the application to manage the data transfer, file system operations, or whatever other operations are required. The same chattiness issue applies to TCP, which affects like web-based business apps, as well as applications like Notes, FTP, and other mission-critical applications.

[1] HP products only address TCP-based applications. All UDP and VOIP traffic is passed through transparently.

Adding bandwidth—That will not work either

On top of the latency issue, the bandwidth available on WANs is almost always a tiny fraction of the LAN. A typical remote office has between 64 kbps and T1 bandwidth (1.544 Mbps) or E1 in Europe. Compared to modern LANs that have 100 Mbps to 1,000 Mbps, a remote site typically relies on less than 1% of the bandwidth.

Lots of companies figure they will just buy more bandwidth or add compression appliances in advance of consolidating remote infrastructure, to compensate for the additional traffic that is generated by centralizing servers. More bandwidth is always helpful, but it is insufficient to solve the problem. Unfortunately, adding bandwidth or implementing compression technology will not improve application performance much, if at all, which means that successfully completing site consolidation projects is still a problem.

Why won't adding more WAN bandwidth help? Because on a WAN, adding bandwidth does not help alleviate the chattiness of the application, which means that all those round trips still have to take place. No matter how much bandwidth you buy, when the initial congestion has been alleviated, the application performance will not be materially affected.

Caching solutions are not sufficient

Some companies have investigated caching appliances as a way to enable site consolidation. That approach can work for single data types, but will not provide a general solution. For Exchange, there are special purpose mail caching appliances available, but they are not a general purpose solution to the problem.

The reason is that caching is an application-specific technology: File caching works for file systems, web caching works for web pages, mail caching works for email, and so on. So while adding an Exchange mail cache will help by storing attachments locally, it adds complexity and only affects the perceived performance of Exchange. Compared against file caching, HP offers several key advantages:

- **Broader applicability**—EFS WAN Accelerators optimize all TCP traffic, not just one type. Whether a company is focused on centralizing Exchange servers, Notes servers, file servers, NAS, tape backup, or some combination of these, EFS WAN Accelerators can help.
- **Much higher "hit rate"**—Because EFS WAN Accelerators do not store any application-specific information (like copies of files, web pages, or emails), even on brand new data you will see a performance improvement. File caches only give you a "hit" when a user requests an identical file to one that has been requested before. EFS WAN Accelerators will deliver improvement even on partial hits, like new versions, different file names, different applications, and so on.
- **Easier deployment**—Because caches are proxy servers, end-user machines have to be configured to know about the proxies, which means touching and changing every client. EFS WAN Accelerators require no end-user configuration, which means the rollout is much simpler and quicker.

The effect of high latency on performance

This effect is easy to understand intuitively: the rate of work that can be performed by a client-server application that executes serialized steps to accomplish its tasks is inversely proportional to the round trip time (RTT) between the client and the server. If the client-server application is bottlenecked in a serialized computation (that is, it is "chatty"), then increasing the round-trip time by a factor of two causes the throughput to decrease by a factor of two—it takes twice as long to perform each step (while the client waits for the server and vice versa).

More generally, the throughput of client-server applications that are not chatty but run over a window-based protocol (like TCP) can also suffer a similar fate. This can be modeled with a simple equation that accounts for the RTT and the protocol window (W). The window is how much the sender can transmit before receiving acknowledgment from the receiver. When a window's worth of data is sent, the sender must wait until it hears from the receiver. Since it takes an RTT to receive the acknowledgment from the receiver, the rate at which data can be sent is the window size divided by the RTT:

$$T = W / RTT$$

A protocol like TCP obviously must figure out how to set its window and as you could imagine the optimal choice depends on a number of factors. To perform well across a range of network conditions, TCP attempts to adapt its window to the underlying capacity of the network. So, if the underlying bottleneck bandwidth (or the TCP sender's share of the bandwidth) is roughly B bits per second, then TCP will set the sender's window to B x RTT:

$$T = (B \times RTT) / RTT = B$$

That is, the throughput is equal to the available rate. Unfortunately, reality is not this simple. Many protocols, like TCP and CIFS, have an upper bound on the window size that is built into the protocol. For example, the maximum request size in CIFS is 64 KB. And, in the original TCP protocol, the maximum window size is limited by the fact that the advertised window field in the protocol header is 16 bits, limiting the window also to 64 KB. While modern TCP stacks implement the window scaling method in RFC 1323 to overcome this problem, there are still many legacy TCP implementations that do not negotiate scaled windows, and there are more protocols like CIFS that have application-level limits on top of the TCP window limit. So, in practice, the throughput is actually limited by the maximum window size (MWS):

$$T = \min(B \times RTT, MWS) / RTT \leq B$$

Even worse, there is an additional constraint on throughput that is fundamental to the congestion control algorithm designed into TCP. This flaw turns out to be non-negligible in WANs where bandwidth is above a few megabits and is probably the key reason why enterprises often fail to see marked performance improvements of individual applications after substantial bandwidth upgrades.

Essentially, this problem stems from conflicting goals of the TCP congestion control algorithm that are exacerbated in a high-delay environment. Namely, upon detecting packet loss TCP reacts quickly and significantly to err on the side of safety (that is, to prevent a set of TCP connections from overloading and congesting the network). Yet, to probe for available bandwidth, TCP dynamically adjusts its sending rate and continually pushes the network into momentary periods of congestion that cause packet loss. In short, TCP continually sends the network into congestion, and then aggressively backs off, and in a high-latency environment, the slow reaction time results in throughput limitations.

Only in recent years have network researchers begun to understand this problem. In fact, an equation was derived in the late 1990s that models the behavior as a function of the packet loss rate that TCP induces:

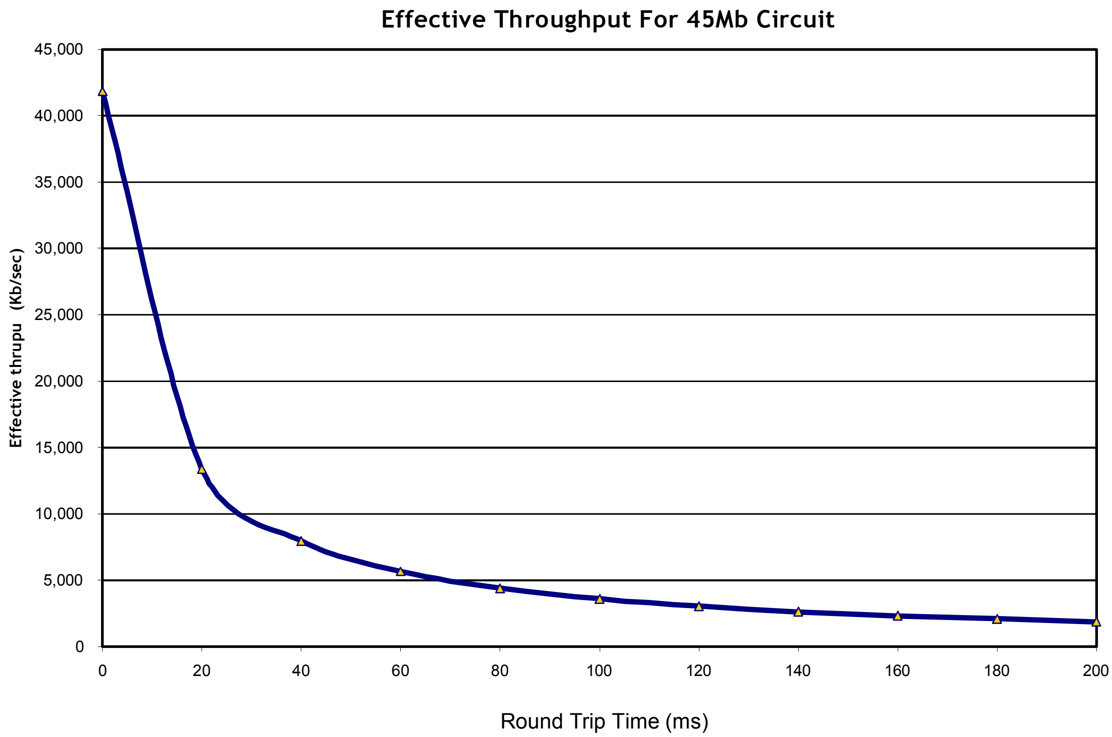
$$CWS = 1.2 \times S / \mathrm{sqrt}(p)$$

This equation says the average congestion window size (CWS) is roughly determined by the packet size (S) and the loss rate (p). So taking this into account, the actual throughput of a client-server application running over TCP is:

$$T = W/RTT = \min(MWS, CWS, B \times RTT) / RTT$$

To see how severe the impact is, see Figure 1.

Figure 1. Effective throughput for 45-Mb circuit



Accelerating applications by modifying TCP

Many IT professionals are aware that TCP has a fixed MWS of 64 KB (the amount of data that can be carried in each trip), but that it can be modified with some work. Actually, in most enterprises, the TCP window size is set to 16 kB or 32 kB, which makes the problem even worse. Even companies that elect to go the route of modifying TCP find that fixing or improving TCP does not help application performance if the application protocol is less efficient than TCP.

In the case of applications like Windows or Exchange, the application protocol (CIFS and MAPI, respectively) are much chattier and less efficient than the underlying network protocol. Therefore, in many cases, making TCP more efficient is necessary, but insufficient because of the added inefficiency of application protocols like MAPI or CIFS. HP EFS WAN Accelerators optimize both application and transport protocol chattiness, and offer unprecedented bandwidth optimization, which delivers the highest performance across the widest array of applications.

The HP solution—HP StorageWorks Enterprise File Services (EFS) WAN Accelerators

HP has introduced its line of EFS WAN Accelerators to boost the performance of all applications running on WANs over TCP. The improvements in performance can be up to 100 times. With this kind of “LAN-like” performance, site consolidation projects can proceed without impacting end users.

Latency optimization—Transaction prediction

To address the high-latency environments inherent in WANs, HP has developed a set of algorithms known as Transaction Prediction, which minimize the number of round trips taken across the WAN by application protocols like MAPI, without interfering with the client-server semantic. These modules are available for specific applications (Windows and Exchange in version 1 of the appliance). Over time, HP will add other latency optimizations.

With specific knowledge of applications protocols like CIFS and MAPI, EFS WAN Accelerators are able to predict upcoming client requests in advance, inject requests on behalf of the client, and then "bundle" many transactions into a few. Each round trip avoided saves a discrete amount of time, independent of how much bandwidth is available. When thousands of round trips are avoided, the time saved can be measured in minutes or even hours, depending on the workload.

Virtual Window Expansion

To address the latency effect on other applications that run on TCP (like any web-based application, FTP, and so forth), the other HP latency optimization, Virtual Window Expansion (VWE), minimizes the number of round trips. EFS WAN Accelerators terminate TCP, and are able to repack TCP payloads, often substituting references to arbitrarily large amounts of data (see Scalable Data Referencing).

This technique "virtually" expands TCP windows because the amount of data that is represented by a reference can be 1 MB, 10 MB, or much more. By virtually expanding the TCP window size, the number of round trips is reduced, which in turn increases throughput. All of this is done without changing the underlying TCP protocol.

Bandwidth optimization—Scalable Data Referencing

HP Scalable Data Referencing (SDR) replicates data within and across the network in a new and unique protocol-independent format to reduce subsequent transmissions of the same data. Rather than attempt to replicate data blocks from a disk volume, files from a file system, email messages, or Web content from application servers, EFS WAN Accelerators represent and store data in a protocol and application-independent format.

A working set of data and references is maintained in persistent storage (on the "Data Store") within each EFS WAN Accelerator appliance. The elegance of the approach is that quite surprisingly there are no consistency issues to be tackled even in the presence of replicated data. This remarkable property is explained in further detail in the following section.

Core mechanisms of SDR

As data is produced at the server site, the SDR system transforms the data payloads into a sequence of data and "references" to that data. As data and their accompanying references are created at the server site in this fashion, they are also created on the Data Store situated at the client site. Conversely, as references are created at the client site, they are also created in the Data Store situated at the server site.

In addition, HP has also developed a proprietary technique to represent an arbitrary amount of repeatedly accessed data with a single (or small number) of references. This approach has the added benefit that it automatically and dynamically adjusts its granularity to the manner in which changes are made to the underlying data. This allows EFS WAN Accelerators to use a relatively small mean data size (that is, to capture fine-grained changes to data) without sacrificing the efficiency of a large data size (that is, to represent a very large amount of with a small reference).

The references and data can either be pulled on demand by client activity or they can be pushed from server site to the client sites using intelligence external to an EFS WAN Accelerator that invokes a service interface in the device to proactively move references and data across the network.

Using the persistent Data Store, a client-server transaction is transformed by a pair of EFS WAN Accelerators into a thin "transaction envelope" that traverses the network. When this envelope reaches the client-side HP EFS WAN Accelerator, it is expanded back to its original form to be processed by the rest of the system and ultimately by the client.

For example, in Windows file sharing (CIFS), when a client reads a file, each block of the file is sent from the server, across the network, and delivered to the client, resulting in a potentially large amount of network traffic. With SDR in place, each read is effectively delivered by sending a reference to the set of data that is stored at the client-side HP EFS WAN Accelerator.

Although the client-server transaction flows across the network (and thus the protocol semantics are fully preserved), no actual data is transferred since it already exists in the form of references and data at the client site. Moreover, the algorithms are entirely symmetric. Consequently, a client dialogue that produces large amounts of data written to the server can also be mapped onto a thin envelope.

Because EFS WAN Accelerators manipulate data in the mission-critical path between the server and the client, it is absolutely imperative that the integrity of the underlying data is maintained as it is transported through the system. To this end, EFS WAN Accelerators employ end-to-end integrity checks to blocks of data as they enter and leave this system (for example, using cyclic-redundancy checks). If the check fails, the system is taken offline until an operator intervenes to diagnose the problem. While this is a drastic measure, it will never occur when the system is operating correctly. Having these checks in place gives the customer assurance that any flaws in the system will be caught and the device disabled rather than allow enterprise data to be silently corrupted.

Exchange 2003

Microsoft has recognized the problems surrounding MAPI, especially surrounding the performance of Exchange 5.5 and 2000 when deployed over WANs. With the introduction of Exchange 2003 and Outlook 2003, they have taken steps to improve at least the perceived performance of Exchange over WANs. So, what have they done?

There are two primary improvements:

- Client-side mail caching
- MAPI protocol improvements

Client-side caching essentially hides the delay in getting email from servers to clients by not displaying any new email headers until the entire email and any attachments are fully delivered. By the time a user is notified that a new message has arrived, the entire message (and any attachments) have already been cached in the client, so the perceived performance is much better. The actual time to deliver the message from the moment a sender clicked "send" may not be much better, but from the recipient's point of view, the performance will seem much faster.

The second set of optimization has been to enlarge the maximum window size of the MAPI protocol from 8 or 16 KB to 64 KB, which means a reduction in the number of round trips generated by the application, which does improve the throughput.

HP offers several areas of significant value, on top of Microsoft's improvements to their product:

- **Higher performance even for older versions of Exchange or partial 2003 rollouts**—To get the full benefits of these new features, you have to deploy both the server (Exchange 2003) and the client (Outlook 2003). If you do not use the Outlook 2003 client (either an earlier version of Outlook, or a different client altogether), then you will not get the benefit of the new features.

  Even if you have not upgraded Exchange and Outlook to the 2003 version, HP can deliver much higher performance to version 5.5 and version 2000 implementations. If you deploy Exchange 2003, but not Outlook 2003 clients, HP can also help improve the throughput of your implementation.

- **Cross-application optimization**—Exchange 2003 + Outlook 2003 will help the real and perceived throughput of email for end users; however, it will do nothing for other enterprise applications being used in remote sites.

  For example, users trying to access centralized Windows servers will still have poor performance over WANs. Also, HP EFS WAN Accelerators will provide significant optimization for content that is sent by email sometimes and by other methods also (like FTP, file sharing, or HTTP, for example). A mail-only solution will not help with those other protocols.

Conclusion

Site consolidation has a tremendous return on investment (ROI), as long as user application performance can be preserved. Most enterprises have a range of remote office IT infrastructure: File servers, Exchange servers, web applications, app servers, Notes servers, NAS, tape backup, and so forth. The more infrastructure that can be centralized or consolidated to the datacenter, the higher the ROI for the IT department.

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

5983-2467EN, 05/2005

invent