It is not about bandwidth white paper

Why adding bandwidth or compression does not solve the problem of application performance on wide area networks

The HP vision

For geographically distributed customers who are looking to consolidate their server environments, the HP vision is to enable the "serverless remote office," where IT infrastructure like servers, storage, and the associated backup equipment can be centralized to the datacenter, leaving minimal equipment in the remote office. With HP StorageWorks Enterprise File Services (EFS) WAN Accelerators deployed, large enterprises can centralize their Microsoft® Exchange servers, storage or file servers, and/or the associated tape backup systems to cut capital and management costs, yet simultaneously preserve high-application performance for their distributed users.

Enterprise customers who cannot or do not want to centralize IT infrastructure can focus on increasing the performance of their existing distributed systems to improve the productivity of their users around the world. Either way, the HP solution can support their goals.

Over the last 30 years networks have gone from a highly centralized mainframe-terminal model, to a distributed client-server model. Now, as the costs of distributed networks continue to mount, there is a reversal underway at many large enterprises looking for ways to "re-centralize" their infrastructure to cut costs.

Simultaneously, globalization is an unstoppable force. Organizationally, corporations are more and more decentralized. Unlike generations past, the key managers, decision makers, engineers, and other professionals in any large company are just as likely to be working from remote branch offices (RBOs) as they are to be working in the home office.

So, IT architects are faced with a paradox: they feel the need to centralize infrastructure to cut costs, but that is counterbalanced by an imperative to deliver high performance to users all over the world. HP foresees a world where distributed enterprise IT infrastructure is designed and built much differently than today; where long distance is no longer an impediment that forces large enterprises to spend millions of dollars on redundant servers or excess bandwidth to compensate for slow wide area networks (WANs).

Executive summary

Wide-area client-server applications are a critical part of almost every large enterprise. However, access to widely used and critical infrastructure across WANs (like file servers, mail servers, and networked storage) is handicapped by very poor throughput and response time when compared to the performance across local area networks (LAN).

Whether an enterprise is taking a centralized approach to IT or needs to maintain a distributed infrastructure, high-performance communication across the WAN is essential to minimize costs and maximize productivity.

Enterprise IT managers typically take one of two approaches to compensate for the performance challenges inherent in WANs:

- If their IT infrastructure is decentralized and they intend to keep it that way for operational reasons, corporate network and server managers typically have to deploy local file servers, data storage, mail servers, and backup systems in each remote office to ensure fast and reliable access to critical data and applications. Often, they also maintain over-provisioned WAN links to enable reasonable levels of performance for file transfers and similar data-intensive tasks.
- If the IT infrastructure is already centralized or the company is moving strategically to a centralized or semi-centralized architecture, then they are faced with a constant demand for "more bandwidth" to remote sites in an effort to improve the performance for distributed users.

The cost of bandwidth and distributed servers can be very high. Many CIOs have decided that a better approach would be to centralize servers, if the performance obstacles of WANs could be overcome or neutralized. On top of the higher costs of decentralized architectures, there are also higher risks of non-compliance with government or industry data retention regulations, lower physical security, and more complex IT management.

The HP advantage

HP develops infrastructure appliances to improve the performance of client-server interactions over WANs without breaking the semantics of the protocols, file systems, or applications. Whether it is copying a file from a distant file server, getting mail from a remote Exchange server, backing up remote file servers to a main datacenter, or sending a very large CAD file to a colleague across the Pacific, slow WANs cost money. The costs are borne in redundant infrastructure, over-provisioned bandwidth, and lost productivity.

The HP approach can improve the performance, or throughput, of client-server interactions over WANs by up to 100 times or more, giving the illusion that the server is local rather than remote. That degree of improvement can enable enterprise customers to centralize currently distributed resources like storage, mail servers, and file servers and deliver new WAN-based IT services that have not been possible before.

The root causes of slow throughput

The primary causes of the slow throughput on WANs are well known: high delay (round trip time [RTT] or latency), limited bandwidth, and chatty application protocols. Since virtually all large enterprises rely on WANs for mission-critical services, and since no one has solved these problems, most companies today compensate for slow throughput by building a highly redundant IT infrastructure.

Large enterprises spend a significant fraction of their IT budgets on storage and networks, much of it spent to compensate for slow throughput by deploying redundant servers and storage and the required backup equipment. HP solutions enable enterprise customers to consolidate and centralize key IT resources to save money, reduce capital expenditures, simplify key business processes, and improve productivity.

Introduction: The problem with WAN performance

Users at remote sites typically rely on connections that have 1/100th the bandwidth of a LAN (or even 1/1,000th) and over 100 to 1,000 times higher latency. Performance of client-server applications over WANs can be measured by the net throughput. In many cases, this throughput is too low to meet critical enterprise needs. Despite all the advances in information technology over the last several decades, two things that affect WAN performance have not changed: the speed of light and the relatively high cost of bandwidth over long-distance network connections.

Problems caused by latency

The speed of light, a fundamental and fixed constant, implies that information transmitted across a network always incurs some non-zero latency as it travels from the source to the destination. In practical terms, this means that sending a packet from Silicon Valley to New York and back would never occur faster than about 30 ms, since this is the amount of time required for light (in a vacuum) to travel in a direct path cross country. In reality, this cross-country RTT is closer to 100 ms as signals travel more slowly in fiber and copper compared to light in a vacuum and packets incur processing delays through each switch and router. This amount of latency is quite significant as it is at least two orders of magnitude higher than typical sub-millisecond LAN latencies.

Round trips from the West Coast of the United States to Europe can be in the range of 100–200 ms, and some links using geo-stationary satellites into remote sites can have latencies in the 500–1,500-ms range. Latency that is higher than 50 ms means that client-server protocols (like file systems or other applications that expect very low latency) that work very well on LANs function very poorly on WANs, or not at all.

Problems caused by limited bandwidth

The relatively limited bandwidth interconnecting remote offices and branches has not changed much either. While many employees routinely depend upon Fast Ethernet (100 Mbps) or Gigabit Ethernet (1 Gbps) within most corporate sites and headquarters facilities, the bandwidth interconnecting many corporate and industrial sites in the world is much lower.

For example, in the United States and Europe, sites like regional offices, retail stores, banks, supermarkets, gas stations, sales offices, and others are still connected with quite low bandwidth. Even though many of them now use DSL, Frame Relay, or other broadband technologies where those services are available, those are quite slow compared with LAN bandwidth. For example, 1-Mbps DSL service offers only 1/100th the bandwidth of Fast Ethernet, and a tiny fraction of (1/1,000th) of what is available using Gigabit Ethernet.

Even in countries like South Korea and Japan that have invested in high-bandwidth backbone networks like Metro Ethernet, the latency and bandwidth issues persist whenever data needs to travel outside the country. For example, a Japanese manufacturer with plants in Japan and the United States constantly needs to send data like CAD/CAM files between plants, or between engineering and manufacturing. The latency from Japan to the East Coast of the United States may be as high as 200 ms, and the trans-Pacific bandwidth is still quite expensive and limited for most companies.

Effects of limited WAN bandwidth

Limited bandwidth is a familiar problem to anyone who has tried to send large files through a constrained link. While email and other relatively low bandwidth, asynchronous traffic can perform adequately under these conditions, when the files are very large, performance breaks down. For files like CAD/CAM files, 3D renderings, or large software builds, the transfer time can often be measured in hours. Delay of this magnitude impacts product development cycles and other critical processes. Even relatively small files like PowerPoint or Adobe Acrobat files can take a long time to open from distant file servers.

Several approaches are available to address different aspects of this problem.

Buy more bandwidth

Of course this seems like the most direct solution, but it is not always the most effective approach. As explained later, adding bandwidth does not usually improve throughput much on WAN links. Furthermore, adding bandwidth has other issues:

- Contrary to popular belief, bandwidth is not free, especially when you manage hundreds of bank branches, sales offices, or retail stores. The costs add up quickly. For international links, the monthly costs of a T1 can easily be $10,000 or more.
- As discussed later in this paper, adding bandwidth does not necessarily improve application throughput.
- In some places adding more bandwidth is not possible or too expensive especially outside the United States, Europe, and certain Asian countries, or in remote areas of the world.

Compress WAN traffic

Another common approach is to deploy commercially available network compression products to send less data over the WAN. This is a reasonable thing to do but the effectiveness is limited. Current methods of compression typically yield roughly 3:1 to 10:1 in bandwidth savings, averaged over time. While that is helpful, it is not enough to dramatically change things if the amount of data being sent is large. Furthermore, if the application throughput limit is due mostly to network latency, then compressing the underlying data will have little or no impact on throughput.

Use software that handles incremental changes

Some applications have developed mechanisms to recognize only when incremental changes have been made to a file, and then send only those incremental changes. What they do not handle well are scenarios where the filename changes or when the data is sent across the WAN using different methods (that is, a file transfer using "Windows Copy" followed by an email with an attachment of the same file). Furthermore, the ability to handle incremental data is usually application specific and does not handle the general case.

Maintain multiple copies of important files

This approach generally involves storing copies of popular or recently accessed content in local servers for quick access. There are several methods available to enterprises to store redundant copies of data in replicated file systems, redundant or local storage servers, or by using any number of new distributed file systems or file caching technologies. There are several key challenges with a file caching approach.

- The implementation of caching solutions is quite complex because every client machine has to be configured to talk to the proxy (file cache).
- Caching is application specific: you need a file cache to store files, a web cache to store web pages, and so on.
- There is a fundamental problem of managing the ever-exploding amount of data, which requires scaling up storage, application and file servers in many places, and trying to ensure that the files people need are available where and when they are needed. This is a daunting and complex challenge.
- Some file caching systems are "read only," and therefore useful for static data that does not need to be edited. This can add to management costs.

Effects of high WAN latency

Network latency, or the RTT of a packet of data, has a direct effect on the performance or throughput of a window-based protocol like TCP or any request-response protocol like Common Internet File System [CIFS], used by Microsoft Windows®) or Messaging Application Programming Interface (MAPI), used by Microsoft Exchange. High RTTs particularly slow down "chatty" applications, even if the actual amounts of data transmitted in each transaction are not large.

Adding bandwidth (or compressing data) does not improve throughput when the RTT exceeds the critical point where throughput is bounded (see Figure 2), or when the problem is primarily latency and not bandwidth related. When the latency exceeds the critical point, throughput decays quickly.

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 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 RTT 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 $B \times RTT$:

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

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*):

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

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:

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

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:

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

Figure 1. Throughput versus latency (T3)

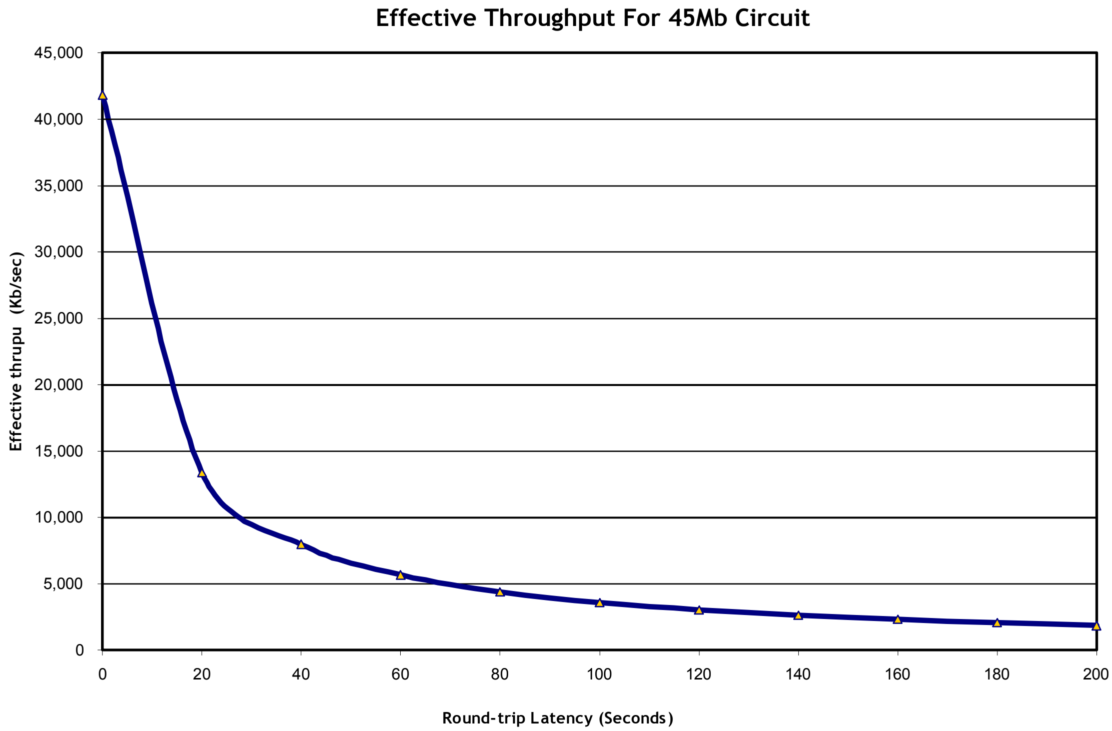
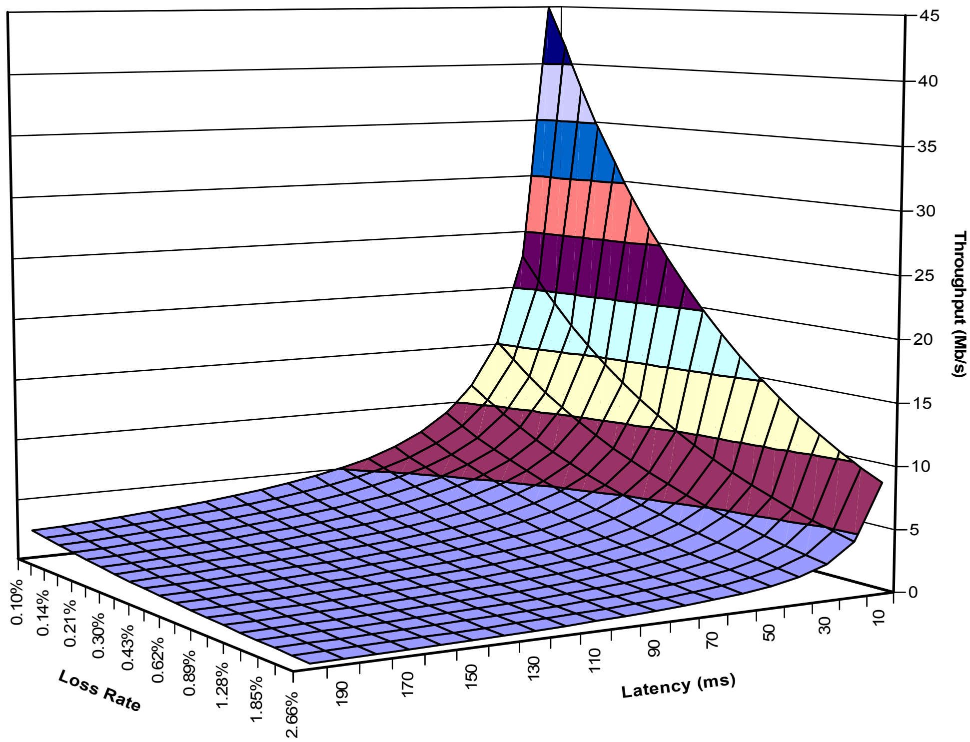


Figure 1 illustrates this problem from a very practical perspective. This graph shows the performance of a TCP data transfer when the network is experiencing a low degree of network loss (less than 1/10 of 1 percent) for increasing amounts of latency. The curve illustrates the performance impact of the protocol window at higher bandwidths.

With a T3 line, the TCP throughput starts out at the available line rate (45 Mbps) at low latencies, but at higher latencies the throughput begins to decay rapidly (in fact, hyperbolically). This effect is so dramatic that at 100 ms of delay (that is, a typical cross country link), TCP is able to use only 4.5 Mbps of the 45-Mbps link. No wonder people are often surprised that their application performance does not increase proportionately when they throw bandwidth at the problem.

As one can see in the preceding figure, if the RTT is greater than a critical point (just 15 ms or so in this example) then increasing the bandwidth of the link will only marginally improve throughput at higher latency. And at even higher latencies, throughput is not increased at all with increases in bandwidth.

Figure 2. Throughput as a function of RTT and loss



To summarize, Figure 2 graphs a surface of modeled TCP throughput over a 45 Mbps T3 link, as a function of both RTT and loss rates. This graph shows that both loss and latency impact performance. While latency has the more dramatic impact, they combine to severely impact performance. In environments with relatively low loss rates and normal WAN latencies, throughput can be dramatically limited.

Financial implications of slow WAN throughput

Network throughput directly affects the architecture of IT infrastructure. Since WANs exhibit very different characteristics than LANs, the services and applications they can support are more limited. To overcome the throughput limits, enterprise IT managers have to compensate by adding equipment in remote locations and by altering or limiting the kinds of business processes in a particular place, based on the available performance.

For example, a global automobile manufacturer might replicate CAD file storage in remote design centers to ensure that the engineers who work at that location have ready access to their work. If the company did not replicate the files locally somehow, then the productivity of those engineers might be severely limited or they may not be able to work out of that location at all. Those CAD files need to then be backed up in each location daily, and proper off-site storage policies and procedures need to be followed to ensure recoverability.

Another scenario is an insurance company with hundreds of offices that deploys file servers or storage servers at every remote office to ensure high performance. If it relied on a centralized storage architecture and WAN access for distributed insurance agents and adjusters, the access to centralized data for those employees would be terrible because of the limited throughput available on the WAN. This would be true even if every office had relatively high bandwidth available, like a T-1 line.

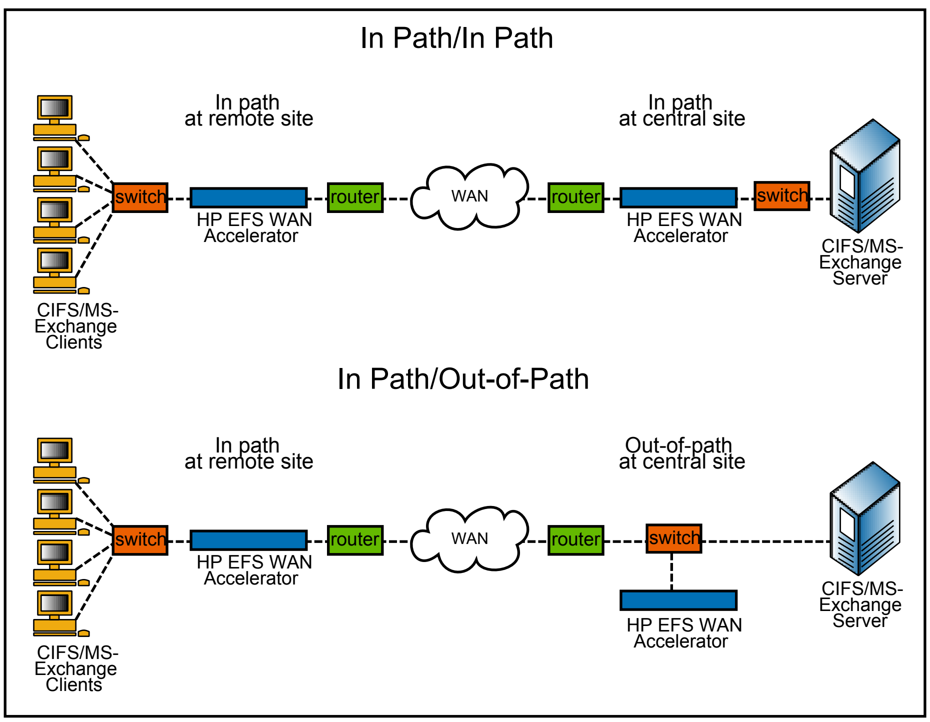
In both these examples, and in many others, slow WAN throughput directly impacts the bottom line. Improving the throughput of a WAN by an order of magnitude or more not only helps productivity but can actually save an enterprise hundreds of thousands, if not millions of dollars in direct and indirect IT costs every year.

There are both security and legal risks also for distributed infrastructure. With the strict compliance regulations in place for many industries and public companies, the consequences of non-compliance are severe. Distributed infrastructure is also much more difficult to patch and upgrade. With the constant threat of viruses and hackers, particularly with Windows-based systems, it is critical that IT managers keep up on a daily basis with the latest patches available for their servers. If the servers can be kept in the datacenter, these risks are substantially lower.

The HP solution for high latency and limited bandwidth

HP has developed an appliance solution that addresses both the bandwidth problem as well as the latency problem. The product is called the HP StorageWorks EFS WAN Accelerator. It is an HP ProLiant-based appliance deployed in the network infrastructure at both client side (typically remote offices) and server side (typically a datacenter or other major facility).

Figure 3. HP EFS WAN Accelerator deployment scenarios



EFS WAN Accelerators are transport-level TCP proxies that conceptually operate in pair-wise configurations, with an HP EFS WAN Accelerator situated near one or more servers (the "server site") and another situated near clients (the "client site"). Though EFS WAN Accelerators communicate with one another in a paired fashion, the architecture allows them to be clustered and meshed across the

WAN, as any EFS WAN Accelerator can communicate directly with any other, so that a server-side EFS WAN Accelerator can communicate with "n" client-side EFS WAN Accelerators.

EFS WAN Accelerators intercept client-server connections without interfering in the normal client-server interactions, file semantics, or protocols. All client requests are passed through to the server normally, yet relevant traffic is optimized to improve performance. Using this approach, EFS WAN Accelerators can be easily introduced into an enterprise environment without requiring any significant changes to the network or architecture.

Transaction acceleration

HP Transaction Acceleration technology is composed of two primary optimizations—a connection bandwidth-reducing technique called Scalable Data Referencing (SDR) and a latency reduction and avoidance mechanism called Transaction Prediction.

The two optimizations can work independently, or in conjunction with one another, depending on the characteristics and workload of the data being sent across the network. The results of the optimization will vary, again depending on these factors, but typically it will result in throughput improvements in the range of 10x to 100x over un-accelerated links.

SDR

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. 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 later in this paper.

Core mechanisms for 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 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 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 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.

Algorithm summary

In summary, the following series of steps outlines the algorithm:

- All TCP requests over supported protocols are intercepted by the EFS WAN Accelerator.
- Data is scanned using HP SDR algorithm to create references to data stored in the "Data Store."
- For each input reference, an index is consulted to determine if the data already exists in the Data Store and, if so, its globally unique reference is retrieved from the index; if not, a new reference is assigned to the data.
- The data is stored locally on disk in the EFS WAN Accelerator, in the Data Store.
- A single reference can represent an arbitrarily large amount of data, thus minimizing the WAN traffic.
- All references are stored in the Data Store until purged according to freshness policies that are configured into the device.

It is easy to see that this approach has a variety of attractive performance features. For example:

When files, web pages, emails, or other data are sent for the first time across a network (no commonality with any file ever sent before), all data and references are "new" and therefore sent to the EFS WAN Accelerator on the far side of the network. This new data and the accompanying references are compressed using conventional algorithms, when, and if, it improves performance.

When a document, file, or email is changed in any way, new data and references are created. In this manner changes are easily incorporated into the Data Store.

Thereafter, whenever new requests are sent across the network, the references created will be compared with those that already exist in the local Data Store. Any data that an EFS WAN Accelerator determines already exists on the far end of the network are not sent; only the references are sent across the network.

As files are copied, edited, renamed, and otherwise changed or moved, the EFS WAN Accelerator continually builds out the Data Store to include more and more data and references.

References may be shared by different files and even by files in different applications if the underlying bits are common.

Virtual TCP window expansion

Another technical feature of the HP approach to bandwidth optimization is something known as "Virtual TCP Window Expansion" or VWE. VWE allows EFS WAN Accelerators to repack TCP payloads with references that represent arbitrary amounts of data. This is possible because unlike other compression products, EFS WAN Accelerators operate at the application layer and terminate TCP, which gives them much more flexibility in the way they can optimize WAN traffic.

The effect of this is as if the TCP payload was increased from its normal 64 KB to some arbitrarily large amount. Because of that increased payload, a given application that relies on TCP performance (HTTP or FTP, for example) can take far fewer trips across the WAN than they used to accomplish the same task. The net-net is that often many applications will run 10 to 15 times faster when EFS WAN Accelerators are deployed.

Transaction prediction

Even if the SDR system previously described eliminated virtually every byte of data that traversed the network, there would still be a network bottleneck tied to the inherent latency imposed by the speed of light. And, as described earlier, this latency bottleneck can have a dramatic impact on overall client-server throughput and performance.

Many client-server applications and protocols were designed for use on local networks where round trips are not costly in terms of performance. Because low latency was presumed, they were not optimized for high-delay environments and the original designers did not consider the effects of many successive round trips on throughput and application performance; the round trips on a LAN do not cost much in terms of time or performance.

On the WAN, however, each round trip can take 50 to 250 ms, depending on the distance involved, the hops required, and the processing delays encountered throughout the path. This level of delay, as much as 1,000 times higher than typical LANs, plays havoc with the performance of file systems and protocols like CIFS, TCP, MAPI, NFS, and FTP, rendering them either very slow or, in some cases, unusable.

To deal with this aspect of the network performance problem, EFS WAN Accelerators rely on Transaction Prediction to mask the effects of wide-area network latency.

Essentially, EFS WAN Accelerators attempt to anticipate client behaviors before they occur and execute predicted transactions ahead of client activity. This technique is known as Transaction Prediction. Then, when the client actually issues the predicted transaction, the answer can be immediately produced without incurring a wide-area round trip.

This approach is quite different from how conventional caching systems function. If the EFS WAN Accelerator used caching, it would have to maintain a store of data that represented the objects to be served locally (like files, file blocks, HTTP objects, email messages, and so forth). While building a store of such objects is straightforward, keeping the store coherent with the original copies in the midst of multiple clients accessing and modifying the objects, with security protocols, locking mechanisms, and so forth all create a complex, difficult-to-manage architecture.

Transaction prediction, on the other hand, avoids the complexities of coherency by logically sending all transactions back to the server and relying on SDR to avoid clogging the network with unnecessary data transfers.

This new approach to overcoming the client-server performance problems over WANs has several attractive advantages over caching. Unlike a cache, which must implement a complete server-side protocol (to talk to the client) and complete client-side protocol (to talk the server), an EFS WAN Accelerator only needs partial protocol knowledge to know when and if it is safe to perform various optimizations. There is no server or client mechanism embedded in the system.

Likewise, the time scales over which the prediction scheme operates are very short, on the order of several round trips (that is, maybe a second or so). In contrast, caches attempt to store data for arbitrary amounts of time in the hope that clients will attempt to access the same data repeatedly. Keeping this data current and consistent with the master copy is a very difficult and complex problem. It is far easier to ensure that transactions executed a few round trips ahead of time are "safe." Yet this much simpler approach gives all the benefits (and more) of a caching system.

Figure 4. Effect on latency on client server transactions

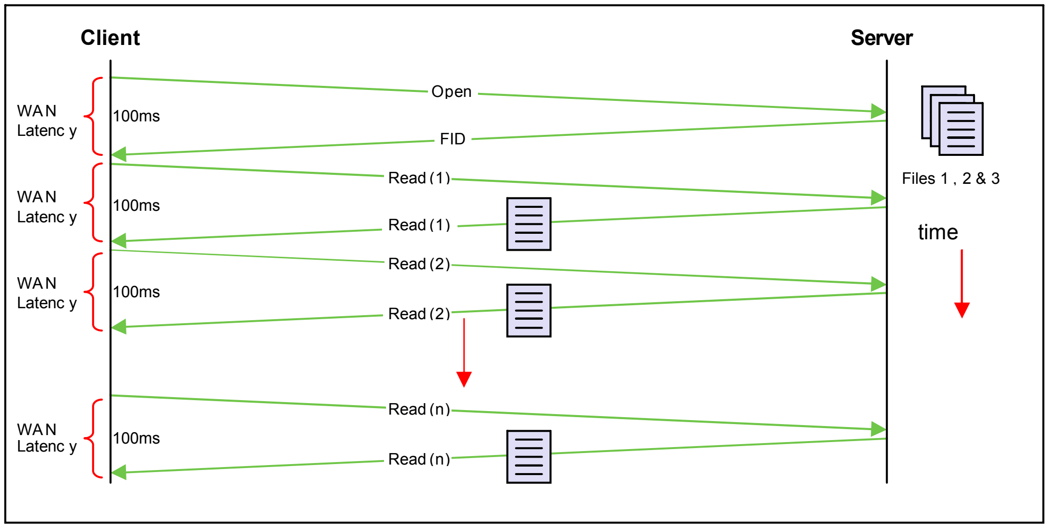


Figure 5 shows how latency affects the performance of client server interactions. In this example, each block is read serially as each round trip is completed. Latency does not have a big effect on performance when the RTT is below 1 ms, like on a LAN. However, when latency is high, taking many round trips degrades performance markedly. In Figure 8, the benefit of Transaction Prediction in reducing round trips is illustrated.

Figure 5. Transaction prediction

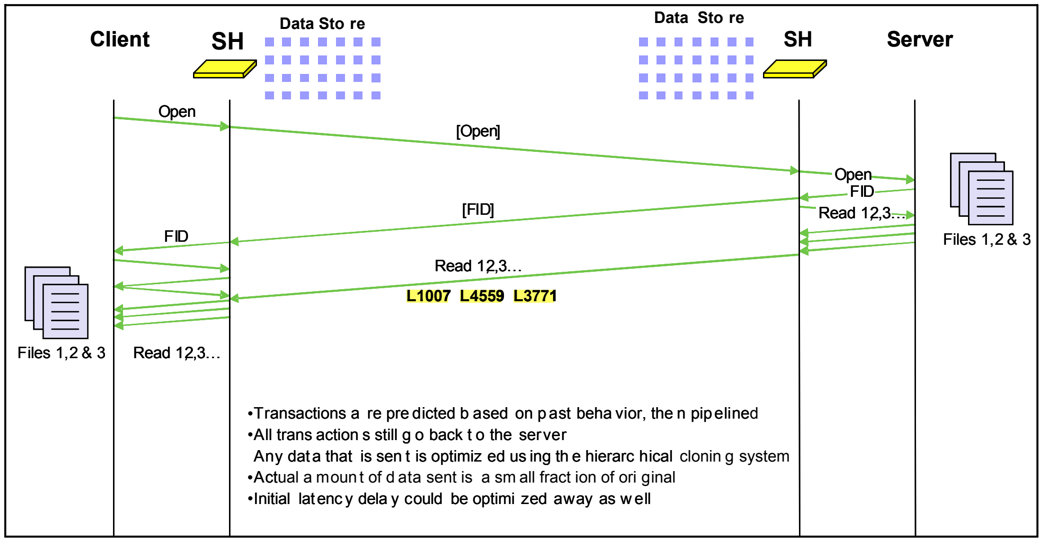


Figure 5 illustrates how the HP Transaction Prediction system works. This diagram shows the simple case of a client that opens a file and sequentially reads all the blocks of a file. In this example, the open request flows across the system through the EFS WAN Accelerators and ultimately to the origin file server. Upon receiving the open response from the server, the server-side EFS WAN Accelerator is in a position to consult its database of past client behaviors and decide, for example, that this file (perhaps in the context defined by earlier transactions) is always sequentially read and closed.

Thus, the server-side EFS WAN Accelerator can inject synthetically produced read requests into the client's session. It might further note that when the blocks have been retrieved and passed through the SDR subsystem, 1 MB of read-ahead data fits into a 20-byte envelope. Consequently, the EFS WAN Accelerator can transmit an envelope representing the 1 MB of read-ahead with virtually no impact on the network.

In a sense, Transaction Prediction is a generalization of this read-ahead concept, designed to work with virtually any client-server based software operating over a WAN. For example, in Figure 5 even though the read-ahead is very effective at eliminating wide-area round trips, there is still a round trip associated with the initial open. In fact, this initial round trip can also be eliminated by the EFS WAN Accelerator when an access to one particular file can predict an access to another file. For example, there may be some earlier transaction that causes the EFS WAN Accelerator to send a synthetic open for the file in question (perhaps doing the read-ahead as well) so that all round trips are completely eliminated. This results in LAN-like performance even though every transaction is served remotely.

As transactions are executed between the client and server, the EFS WAN Accelerators intercept each transaction, compare it to the database of past transactions, and make decisions about the probability of future events. Over time, as more transactions are observed, transition probabilities are improved and the confidence levels increase. For transactions that are not amenable to prediction, confidence levels never increase, which informs the transaction predictor to be less aggressive in such cases.

In addition, different types of information can be attached to each state (for example, to remember that in a certain state a file is read sequentially). By computing the maximum likelihood path through given the current state, fairly deep predictions can be made about the set of future transactions that are likely to occur, allowing the EFS WAN Accelerator to anticipate client activity several transactions before it actually occurs.

Based on this model, if an EFS WAN Accelerator determines that there is a very high likelihood of a future transaction occurring, it might decide to go ahead and perform that transaction, rather than wait for the response from the server to propagate back to the client, and then back to the server. The performance improvement in this scenario comes from the time saved by not waiting for each serial transaction to arrive before making the next request. Instead, the transactions are pipelined one right after the other.

Of course, transactions can only be executed by EFS WAN Accelerators ahead of the client when it is safe to do so. To this end, EFS WAN Accelerators are designed with enough knowledge of the underlying protocols (for example, CIFS oplocks, and so forth) to know precisely when and if it is safe to do so. In cases where such predictions are unsafe, the transactions are relayed back to the origin server and the benefit of Transaction Prediction is lost in these rare cases (but the bandwidth benefits of SDR are still reaped). Fortunately, a wide range of important applications turns out to have very predictable behaviors and, as a consequence, Transaction Prediction can enhance performance significantly. When combined with SDR, the overall performance improvement can be up to 100 times.

Synergy between the two optimizations

The SDR and Transaction Prediction algorithms have synergy that is important to the overall efficacy of the EFS WAN Accelerator. By exploiting information known to the SDR stage, the Transaction Prediction logic can modulate its behavior and adjust its aggressiveness to limit its overall impact on the network.

For example, if the server-side EFS WAN Accelerator decides that it should perform a certain set of predictions on behalf of the client, it first passes the predicted results through the SDR algorithm. If these results are represented in a sufficiently small envelope of information (perhaps guided by the bandwidth policies described later), they can be shipped across the network to the client side to potentially short-circuit the predicted client activity. If the results are too large, the prediction can be aborted, or the depth or scope of prediction can be reduced to adhere to prescribed bandwidth policies.

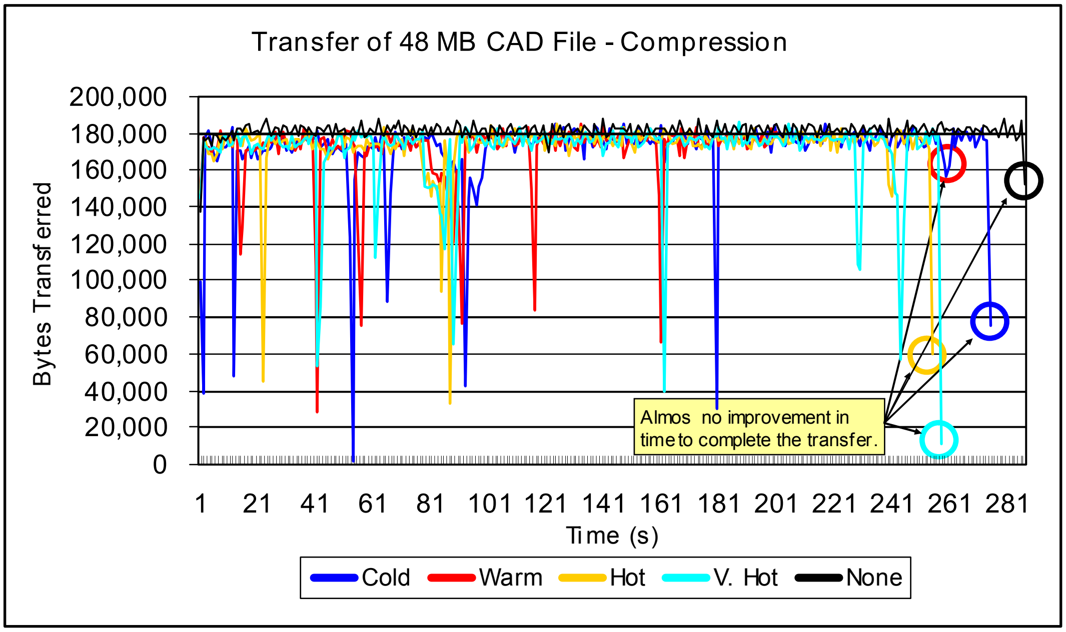
HP versus compression

HP technology, like compression, optimizes WAN traffic by removing redundant bits from the network. Unlike compression technologies, EFS WAN Accelerators perform optimizations for application-level protocols also. The comparison between the two can be dramatic.

Figure 6 shows that using a WAN optimization appliance (a compression appliance) has little to no effect on the transfer of this 48-MB AutoCAD file as it is sent repeatedly over the WAN link. The time for it to complete the transaction remains approximately the same, even as the same data is passed through the compression appliance several times.

The circles show the end of each pass of this file through the compression appliance. The base case (black) took about 290 seconds, which was reduced to 260 seconds after the same data had passed through several times. In addition, the bytes sent were not materially reduced either; the net compression ratio was about 20% (4/5 of the data was sent as compared to the base case). This is because compression systems are not able to look at large data sets, nor do they account for the effects of high latency on throughput.

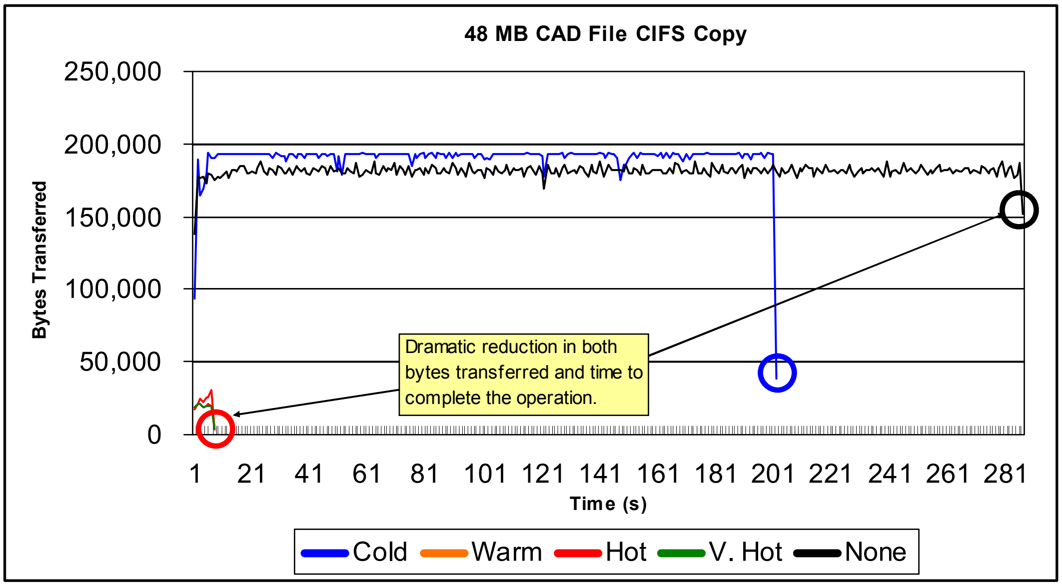
Figure 6. Conventional approach to compression



When compared to the results of the same test done through HP appliances, the effect is quite different, as is shown in this chart.

With HP appliances deployed, the transfer was completed in about 7 seconds (versus 290) and the effective compression ratio was between 318:1 and 382:1, depending on which trip was measured (compared to 1.2:1 for compression alone).

Figure 7. HP approach to compression



These two charts show decisively the difference between the HP approach and that of a more conventional packet-level compression product.

Technical benefits of the HP approach

The overall approach of combining SDR to reduce bandwidth use with Transaction Prediction to mask the effects of high wide-area latencies admits an effective system approach to WAN client-server communication, characterized by several important and attractive properties:

- **Transparency**—EFS WAN Accelerators pass all transaction requests through to the server, just as if it were not deployed at all. This means that applications and protocols work normally with no modification—they just work faster than before.
- **Efficiency**—HP SDR is much more efficient than a file-based or even block-based approach. Even if a change is made in a file that affects every block of data in the file, the SDR system recognizes commonality across references and optimizes for them. A caching algorithm based on blocks or files does not retain its optimization under these conditions.
- **Robustness**—HP architecture is very robust in the face of changes to the underlying data. As files are edited and written to disk, new references are created and incorporated into the existing tree without disrupting the existing data structure.
- **Simplicity**—HP architecture is application-independent and simple in concept. There is no complex cache coherency protocol, and the product does not have to keep track of every file in every location on the network. Furthermore, because of the shared nature of the data in the Data Store, the optimizations are more broadly realized than alternate approaches.
- **Scalability**—The Data Store on EFS WAN Accelerators is disk-based, not RAM-based, so it is inexpensive to scale-up the capacity to size the solution to the data in question. Furthermore, EFS WAN Accelerators can be clustered to provide both scalability and reliability.

- **Shared references**—References can be shared by many files, which means that redundancies in data stored in enterprise networks are automatically be factored out. Shared references result in a dramatic reduction in the amount of storage space required for a given set of files. The degree of reduction varies, depending on the level of redundancy in the data.
- **Long time scales**—Since the data and references are stored in disk and not in RAM, they can be preserved indefinitely. They are purged according to a freshness algorithm, with stale data being deleted over time. However, over the relevant time scales and for "popular" references, they will remain in disk and therefore be available for bandwidth optimization over long time periods.
- **Application independent**—Since the references are created with no knowledge of the application in question, the process is independent of the files and applications being used. If there are references that are common across different filenames, or across different applications, then the optimizations will still be useful.

Key business and IT benefits

Following is a brief overview of the key customer benefits.

Enable the consolidation of distributed infrastructure

At a high level, the HP solution allows IT architects to centralize key assets that until now had to be distributed to deliver the high performance required. Storage servers are an example, where the deficiencies of the underlying protocols (CIFS or NFS) meant that storage was a "local" resource, and that accessing storage over a WAN was not practical. Because of this, enterprise customers had to deploy redundant storage to solve the distributed performance challenge, a costly approach. HP EFS WAN Accelerators give enterprise customers the flexibility to choose the lower cost centralized approach to storage, if that makes sense. The benefits of consolidation are well known, but to summarize, they are:

- **Lower IT costs**—Higher machine utilization, fewer upgrades, less travel.
- **Easier IT administration and lower complexity**—Patching, upgrading, repairing, and so forth are much easier when the equipment is in the datacenter. Whether your company is performing this work in house or outsourcing to external vendors, the costs are much lower than dealing with distributed equipment.
- **Higher compliance**—Data retention and recovery is easier when resource are centralized. Regulations like Sarbanes Oxley can be complied with much more readily, thus reducing the legal exposure for company officers.
- **More reliable backup**—Removing tape drives from remote offices means that the whole unreliable process of backing up to tape, storing tapes off site, and then trying to do restores from tape can be avoided. Instead, all data is kept in the datacenter where trained, knowledgeable staff can perform largely automated and highly reliable backups.
- **Higher physical security**—Keeping servers with sensitive or proprietary data in remote offices invites attack and abuse. With centralized servers, the IT department has much more control over the physical resources they are responsible for.

Improve distributed client-server performance

For companies with distributed workforces, a high-performance WAN means less time waiting for the latest files, emails, and web pages. When the time it takes to receive large files is measured in hours, reducing that time by 90% or more has a direct and significant impact on the bottom line.

Accelerate the performance of web-based applications in RBOs

Many enterprise applications are moving to a web-based implementation. Like client-server–based applications, performance for remote users can be quite poor, affecting their productivity and the effectiveness of the application.

Since EFS WAN Accelerators support HTTP, any web-based enterprise application (such as CRM or SFA) that is accessed by users in RBOs will be automatically accelerated, in many cases quite dramatically. Any application that generates similar web pages, like customer record lookups, parts lists, inventory checks, financial results, and so forth will benefit from the deployment of EFS WAN Accelerators.

Enable new services

There are some things you just do not do over a WAN today. For example, backing up RBOs over your WAN will kill it. So, instead most companies deploy local tape backup systems to avoid overloading their WANs with backup traffic. The problem with that approach is now your backed-up data is scattered over all your remote offices, with little hope of your ever getting control of it. With HP and off-the-shelf third-party network back-up software, you can use your WAN to bring all your backup data from branch offices back to the datacenter for more efficient and cost-effective backups.

Minimize WAN traffic

With the combination of SDR and compression, the amount of data that needs to actually be sent over a WAN link is a tiny fraction of the actual data requested. In some cases, it can be as low as 1/400th of the original data. This can be directly translated into cost savings if bandwidth upgrades are looming, or to reclaim bandwidth for other uses or new applications.

Transparent, incremental deployment

The HP solution can dramatically improve your WAN, and it fits into your IT architecture with little to no changes. EFS WAN Accelerators can be deployed incrementally on one or more WAN links, with minimal configuration and with no changes to the underlying file systems or semantics.

Extensibility

The HP architecture is not tied to any particular protocol or application. Rather, it is a general solution to client-server performance over WANs. The initial product supports a limited set of commonly used protocols (that is, MAPI, CIFS, HTTP, FTP, WebDAV, Remote Backup, and so on), but may be extended over time as customers dictate. For example, it might be desirable to support additional protocols such as NFS, IMAP, or ODBC, or to extend support to particular applications such as Lotus Notes.

Persistent storage of data

Because EFS WAN Accelerators store data and references over long time periods, the HP bandwidth optimizations are longer lived than solutions that rely on RAM-based memory. When two files with common elements are sent across the WAN days or weeks apart, the persistence of the references over time means the optimization can still be done. With a RAM based solution, the common data patterns can be optimized over seconds or minutes, but not over long time periods.

Application acceleration

HP optimizations directly benefit HTTP-based applications and enterprise portals. These applications typically retrieve and display a great deal of common data that can be very effectively stored in the EFS WAN Accelerator's Data Store.

Speed up application development and rollout

One key part of developing and rolling out new applications is optimizing them for slow links. Given the deployment of EFS WAN Accelerator (simple to do by using HTTP) this step can be dramatically shortened or eliminated. This can result in a substantial cost savings or enable deadlines to be met that would otherwise be missed.

Application independence

Since data and references are created without consideration of the source application or objects within the file, they are highly flexible in the way they are used for optimizations independent of the application from which they originated. For example, an engineer might open a 100-MB CAD file from a remote directory at headquarters, work on it, and save the changes as a new filename in a local directory. Then, a week later the engineer emails the revised CAD file with a new filename as an email attachment to a colleague back at headquarters. Even though the filename is different and the application is different, the fact that the references comprising the file are largely the same means that the HP bandwidth optimization techniques can be invoked, resulting in a big performance gain.

Summary

HP has developed a solution to one of the most pervasive problems facing network architects: client-server application performance over WANs. The HP EFS WAN Accelerator family of appliances provides a way for IT managers in large enterprises to simultaneously save money by consolidating and centralizing key infrastructure like storage, without giving up the performance of their distributed systems.

The flexible and seamless architecture from HP means that enterprises that adopt the technology do not have to alter their current IT approach or modify their file systems in any way. EFS WAN Accelerators were designed from the ground up to interoperate in the most demanding WANs.

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

5983-2465EN, 05/2005

invent