 Improving Web Application Response Time for Remote and Mobile Users

network acceleration
Improving Web Application Response Time with Network Acceleration Technologies

Application performance is affected by many factors associated with both network and application logic, which must be addressed in order to achieve satisfactory application performance results. At the network level, application performance is limited by high latency (the effect of physical distance), jitter, packet loss, and congestion.

This guide provides an overview on the network issues, and what to look for in solving these challenges to poor application performance for remote and mobile users.

**Bandwidth and The Network**

Bandwidth limitations continue to play a significant role in application performance. Most networks fall into one of two categories. In the first category are low speed networks such as dial up and frame relay. These networks often range in speeds from 56 Kbps to 2 Mbps and have many endpoints. While not suitable for large offices, these data rates are common among small branch offices. The second class of wide area networks is high speed networks typically used for data replication and communication between large offices. Unlike small office networks, these networks often range in speed from 45 Mbps to 622 Mbps.

Although networks have continued to improve over time, application traffic has increased at an alarming rate. Bandwidth-efficient client server applications have been replaced with bandwidth-demanding web applications. Where previous generation client server transactions involved tens of kilobytes of data, rich web based portal applications can transfer hundreds of kilobytes per transaction. Files attached to email and accessed across remote file shares have increased in size. Even data replication environments with dedicated high speed links have encountered bandwidth challenges due to increases in the amount of data requiring replication.

**Add More Bandwidth?**

For both low speed and high speed networks, provisioning additional bandwidth to meet the increased demand is often prohibitively expensive. Bandwidth prices have not declined as rapidly as expected and networks have been unable to keep up with application demands.

Previous attempts to apply compression at the network level have been relatively unsuccessful. Routers have touted compression capabilities for years, yet very few organizations enable this capability. The reason for this is that compression typically adds overhead both in terms of additional load placed on the routers themselves and in terms of additional latency due to the time it takes for the router to compress each packet.

Implementing an optimal compression scheme requires not only an algorithm, but also an architecture designed to maximize efficiency and performance. A key factor in compression efficacy is how the data is presented to the compression routine. All compression routines achieve greater levels of compression when operating on
homogenous data. When presented with heterogeneous data, such as a collection of packets from multiple different protocols, compression ratios fall dramatically.

**Packet Based Compression**
The primary problem with packet based compression is that it mixes multiple different data types together when compressing. Packet based compression systems have additional problems. When compressing packets, these systems must choose between writing small packets to the network and performing additional work to aggregate and encapsulate multiple packets. Neither option produces optimal results. Writing small packets to the network increases TCP/IP header overhead. In addition, aggregating and encapsulating packets adds encapsulation headers to the stream.

F5’s web acceleration products utilize a technology called Transparent Data Reduction (TDR) to address these bandwidth challenges. Unlike previous forms of compression, TDR utilizes a two stage compression process (TDR-1 and TDR-2) to maximize bandwidth savings while minimizing processing latency. The first step of the process examines the transmitted data to determine if any part of it has been previously sent. If so, the previously transmitted regions are replaced with references. The second step further compresses the data through the use of dictionary based compression and advanced encoding schemes.

**TDR-2**
The data reduction routines utilized by TDR-2 are designed to identify and remove all repetitive data patterns on the WAN. As data flows through two web acceleration appliances, the products record the byte patterns and build synchronized dictionaries. Should an identical pattern of bytes traverse the WAN a second time, the device near the sender replaces the byte pattern with a reference to its copy in the dictionary. When this reference reaches the remote device, the reference is replaced with the original data in the dictionary and the resulting stream is identical to the original one sent.

Because TDR-2 locates repetitive patterns in data, it is ideal for improving file transfers such as CIFS, email attachments, and FTP as well as application protocols that repeatedly send identical data. Unlike caching technologies, TDR-2 ensures that all transactions are received and fully processed by the server. This leaves existing security checks in place and allows clients and servers to operate as normal.

TDR-2 differs from caching technologies in another significant fashion. With TDR-2 there are no stale data problems. With traditional time based caching, objects are stored for a predetermined amount of time. If the object changes on the server before this time has elapsed, the cache will serve stale data. With TDR-2, all transactions are fulfilled by the server and the stored data is only used to reduce the number of bytes transferred across the WAN. Should the data change on the server, the new data will be recorded in the dictionary on its way to the client.

**TDR-1**
During periods of high congestion, TDR-1 increases compression levels to reduce congestion and network queuing delay. During periods of low congestion, TDR-1 reduces compression levels to minimize compression induced latency. The adaptive nature of TDR-1 ensures that the optimal compression strategy is applied and enables network administrators to deploy compression without fear of degrading application performance.

In addition to improving application performance, TDR-1 simplifies configuration as well. Unlike other systems which require an in depth understanding of the compression throughput and latency characteristics, TDR-1 automatically selects the strategy appropriate for the network conditions and updates its strategy in
F5 web acceleration technology enables F5 devices to adapt, in real time, to the latency, packet loss, and congestion characteristics of WAN links, and accelerate virtually all application traffic.

Real-Time. This intelligence allows TDR-1 to scale from low speed 64 Kbps frame relay networks to 155 Mbps OC3 networks while optimizing a wide range of protocols.

**Improves Windows File Sharing**
Common Internet File System (CIFS) is a remote file access protocol that is the basis for Windows file sharing. CIFS performs very poorly across a high-latency WAN because it requires a large number of back and forth transactions. WANJet accelerates CIFS performance by predicting client requests and pre-staging data on the WANJet local to the client. This greatly improves performance for Windows file transfers, directory browsing, and remote access to Microsoft applications such as MS Word, Powerpoint and Excel.

**Eliminates Negative Effects of Distance**
When application performance suffers, IT managers often assume that adding bandwidth will solve the problem. Unfortunately, because of the way the TCP protocol works, adding bandwidth is often ineffective. TCP throughput degrades significantly on the WAN, particularly on high-latency, intercontinental links.

To overcome these inherent protocol limitations, F5 web acceleration technology employs adaptive TCP optimization (which combines session-level application awareness, persistent tunnels, selective acknowledgements, error correction, and optimized TCP windows) to fully utilize available bandwidth. This enables F5 devices to adapt, in real time, to the latency, packet loss, and congestion characteristics of WAN links, and accelerate virtually all application traffic.

**Makes the WAN Faster and Safer**
With the product’s site-to-site SSL encryption, IT managers no longer need to choose between fast and safe. Applications communicate at high speed, over any WAN, in full privacy. Users experience transparent operation and high speed, IT managers rest easy knowing that corporate data is protected, and organizations ensure compliance with regulations such as HIPAA and Sarbanes Oxley.

**Ensures Predictable Performance**
F5’s web acceleration products also include a QoS feature that lets IT managers tune their global networks for round-the-clock application performance. IT resources are logically grouped and managed according to business rules. As a result, bandwidth is preserved for real-time applications like VoIP. This translates into a predictable user experience for application performance, regardless of WAN conditions, time of day, or geographical location.