



F5 White Paper

# Optimize and Accelerate Applications across the WAN

Bandwidth-efficient client server applications have been replaced with bandwidth-demanding web applications, which can put a strain on your infrastructure. Maximize your network and application delivery with WAN optimization

**by Saxon Amdahl**

Architect

**with contributions from Peter Silva**

Technical Marketing Manager



# Contents

<b>Introduction</b>	<b>3</b>
<hr/>	
<b>Deployment Benefits</b>	<b>4</b>
F5 Value for Users	4
<hr/>	
<b>Conclusion</b>	<b>7</b>



## Introduction

Bandwidth limitations continue to play a significant role in application performance. Most WANs fall into one of two main categories. The first category includes lower-speed networks such as dial up, frame relay, and many mobile data networks. These networks often range in speed 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 category is high-speed networks that range in speed from 45 Mbps to 622 Mbps; these are typically used for data replication and communication between large offices. A third, less common category includes Wide Area Ethernets (sometimes referred to as Carrier Ethernets), which are typically used for extending Local Area Networks (LANs) over fiber and are often built with point-to-point circuits.

Although networks have improved 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—and with the explosion of social media and video, megabytes per transaction is not uncommon. Files attached to email and accessed across remote file shares have also 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. This bandwidth-hungry society, in which people often watch videos right on their mobile devices, can have both a financial and technical effect on the infrastructure needed to deliver such large pieces of content.

For both low- 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.

## Network Challenges

### Network Compression

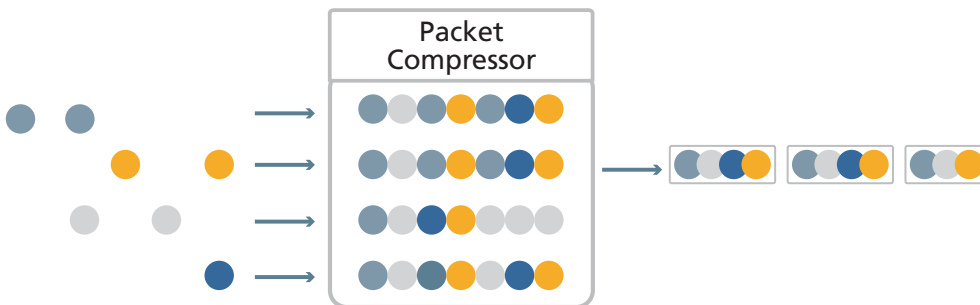
Attempts to apply compression at the network level have been relatively unsuccessful. Routers have touted compression capabilities for years, yet very few organizations use this capability because it typically adds overhead both in terms of



increased load on the routers themselves and additional latency due to the time it takes for the router to compress each packet.

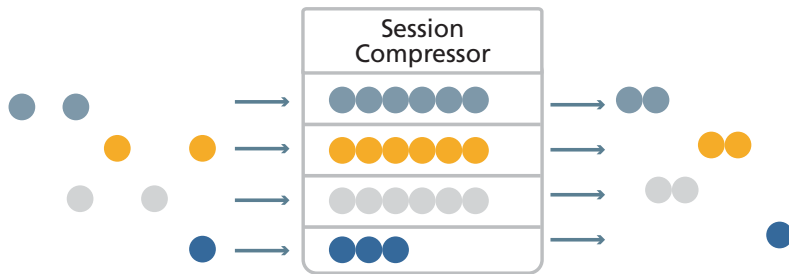
Implementing an optimal compression scheme requires not only an algorithm, but also an architecture designed to maximize efficiency and performance. How the data is presented to the compression routine is a key factor in compression efficacy. All compression routines achieve greater levels of compression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

Packet-based compression systems come with multiple problems, the biggest of which is that they mix multiple data types together when compressing. Another problem is that 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, and, aggregating and encapsulating packets adds encapsulation headers to the stream.



**Figure 1: Packet compressor mixes data types and must aggregate and encapsulate multiple packets.**

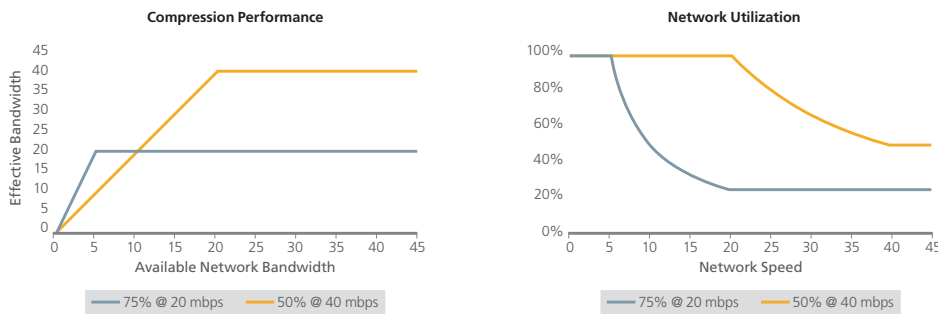
Unlike previous compression solutions, the F5® BIG-IP® WAN Optimization Module™ (WOM) operates at the session layer. This allows it to apply compression across a completely homogenous data set while addressing all application types. This results in higher compression ratios than what comparable packet-based systems can produce.



**Figure 2: Session compressor offers better compression ratios along with improved performance.**

Furthermore, operating at the session layer eliminates packet boundary and re-packetization. This enables BIG-IP WOM to easily find matches in data streams which at layer 3 may be many bytes apart, but at layer 5 are contiguous. Compressing at the session layer also increases system throughput by eliminating the encapsulation stage.

Achieving a high compression ratio is only part of the performance puzzle. To improve performance, the compressor must actually increase network throughput, so it must be able to achieve greater than line speed throughput. Consider the following two theoretical compression devices. Compressor A (Blue Line) achieves a 75 percent reduction in data and can read in and compress data at 20 Mbps. Compressor B (Red Line) compresses twice as fast, but is only able to achieve a 50 percent reduction in data.



At first glance Compressor A appears to be the more effective compression device; but Compressor B actually results in better network performance for link speeds greater than 10 Mbps. This is because as network speeds increase, Compressor A is unable to fully utilize the available bandwidth. For optimal performance—and to use all of the bandwidth you’ve paid for, rather than leaving a half-empty pipe—you want to apply the best compression ratio for the bandwidth available.



# Reduce Data Transparently

The BIG-IP WOM uses a two-stage compression process to maximize bandwidth savings while minimizing processing latency. The first step of the process, symmetric data deduplication (SDD), examines the transmitted data to determine whether any part of it has been previously sent. If so, the previously transmitted regions are replaced with references. The second step, symmetric adaptive compression (SAC), further compresses the data by using dictionary-based compression and advanced encoding schemes.

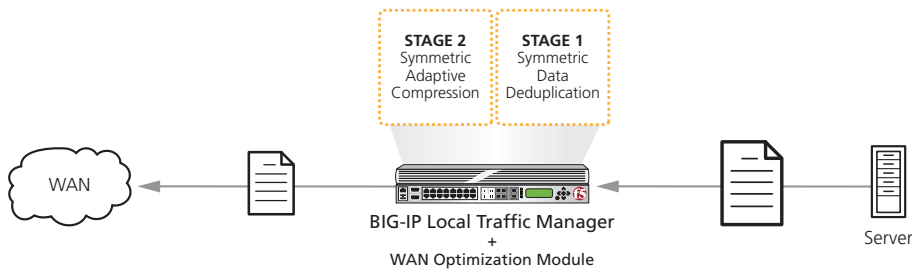


Figure 5: F5 WAN optimization uses a two-stage compression process to maximize bandwidth savings while minimizing processing latency.

## Symmetric Data Deduplication

The data reduction routines used by symmetric data deduplication (SDD) are designed to identify and remove all repetitive data patterns on the WAN. As data flows through two BIG-IP WOM appliances, the two BIG-IP WOM devices records the byte patterns and builds synchronized dictionaries. Should an identical pattern of bytes traverse the WAN a second time, the BIG-IP WOM near the sender replaces the byte pattern with a reference to its copy in the dictionary. When this reference reaches the remote BIG-IP WOM, it is replaced with the original data in the dictionary and the resulting stream is identical to the original one sent.

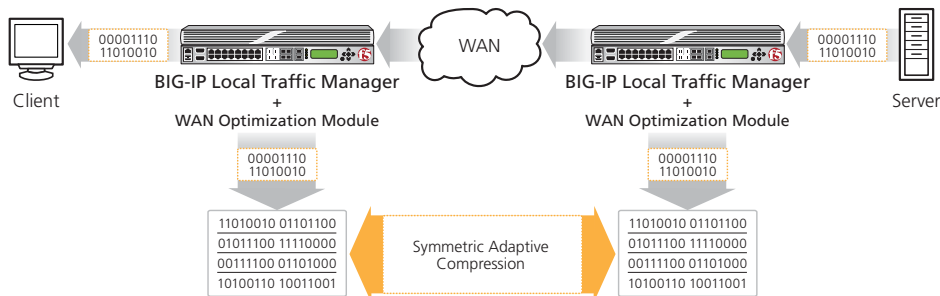


Figure 6: Symmetric data deduplication



SDD dictionary data is not dependant on application protocol or direction of transfer. This means that, should a pattern of bytes first appear in one protocol, and then in another, the second transfer will fully benefit from the first. This holds true even if the first transfer is a download while the second transfer is an upload (such as when a user downloads an email attachment and then uploads it to a Windows file share). Because SDD recognizes byte patterns and is unaffected by protocol differences and direction of transfer, it can leverage the initial email (MAPI) transfer to improve the subsequent Windows file share (CIFS) transfer.

Because SDD 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, SDD ensures that the server receives and fully processes all transactions. This leaves existing security checks in place and allows clients and servers to operate normally.

SDD differs from caching technologies in another significant way: there are no stale data problems. With traditional time-based caching, objects are stored for a predetermined amount of time. If the object is changed on the server before that time has elapsed, the cache will serve stale data. With SDD, all transactions are fulfilled by the server and the stored data is used only 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.

File Caching	SDD
Stale data possible	Accurate data guaranteed
Access control/security compromised	Access control/security intact
Protocol-specific (for example, HTTP)	Protocol-agnostic
File name dependent	File name independent
File edits transfer the entire file	File edits transfer only the new data

This approach has the added benefit of providing partial acceleration for files that have undergone revision. With classic caching technologies, any update to a file requires downloading the new version in its entirety. Because SDD is byte-based rather than file-based, only novel portions of the file are transferred.

However, achieving a high compression ratio alone is not enough to improve application performance. The compression system must also have high throughput, and BIG-IP WOM ensures that data reduction translates into actual performance gains and not simply an empty WAN.



In cloud deployments, SDD can have a significant impact, particularly when deploying virtual machine images or moving applications from the local data center to cloud peering-points. While virtual machine images can be quite large (in the tens of gigabytes), they often comprise of a significant amount of redundant data—like the underlying OS. Virtual machines can be optimal candidates for SDD processing. Because much of the magic happens within a secure, optimized iSessions tunnel between two BIG-IP devices, you can control the availability, scalability, and security of your data.

## Symmetric Adaptive Compression

After SDD has removed all previously transferred byte patterns, BIG-IP WOM applies a second class of data reduction routines called symmetric adaptive compression (SAC). While SDD is optimized to enhance repeat transfer performance, SAC is designed to improve first transfer performance with advanced encoding techniques and dictionaries optimized for very small repetitive patterns.

Compression as implemented in routers and other network devices operates in an all-on or all-off mode. For protocols such as HTTP, adapting compression behavior to congestion levels is critical. Unlike most protocols, HTTP exhibits both interactive and bulk data transfer characteristics. When a user interacts with a web application, the majority of HTTP exchanges involve a large number of small transfers. However, during file download, only a single transfer occurs and the amount of data transferred is often 5 MB or more.

This change in network behavior presents an interesting challenge. In the case of web browsing, it is critical that compression routines minimize the amount of additional latency. Even a few milliseconds added to each transaction can substantially degrade the performance of some applications. At the same time, larger transfers via HTTP, such as file download, typically benefit substantially from compression because the improvements in transfer time overshadow any additional processing latency.

SAC solves this problem by constantly adapting to changing network conditions and application requirements mid-stream. During periods of high congestion, SAC increases compression levels to reduce congestion and network queuing delay. During periods of low congestion, SAC reduces compression levels to minimize compression-induced latency. By examining every packet and adjusting the codec based on the flow, the adaptive nature of SAC ensures that the optimal compression strategy is applied and enables network administrators to deploy compression

## White Paper

Optimize and Accelerate Applications across the WAN

without fear of degrading application performance. When there is little to no benefit to performance, often when a file is encrypted, BIG-IP WOM will leave it alone to save valuable CPU cycles.

In addition to improving application performance, SAC simplifies configuration. Unlike other systems that require an in-depth understanding of the compression throughput and latency characteristics, SAC automatically selects the strategy appropriate for the network conditions, and then updates its strategy in real time. This intelligence allows SAC to scale from low-speed 64 Kbps frame relay networks to 1244 Mbps (OC24) networks while optimizing a wide range of protocols.

Like SDD, SAC can benefit cloud deployments as well. The elasticity of the cloud requires that the infrastructure be just as dynamic. By responding to the ever-changing network conditions of both the cloud and end user, SAC can make certain that you are using your bandwidth efficiently and quickly delivering content to any user around the globe.

## Conclusion

Today's networks are challenged to keep up with the rising demand for bandwidth. TDR as implemented in BIG-IP WOM combines adaptive compression and data reduction technologies with a purpose-built architecture designed to maximize application performance. BIG-IP WOM provides this optimization in a two-step process. First, redundant patterns are identified and replaced by small references to the remote BIG-IP WOM's data store. Second, the remaining network data is compressed through intelligent, network- and application-aware routines that optimally encode the remaining data in as few bytes as possible.

All of this ultimately results in a solution that delivers LAN-like application performance over the WAN, accelerating disaster recovery, client server applications, file transfers, email, data replication, and more while delivering predictable, fast performance for all WAN users.

**F5 Networks, Inc.** 401 Elliott Avenue West, Seattle, WA 98119 888-882-4447 [www.f5.com](http://www.f5.com)

F5 Networks, Inc.  
Corporate Headquarters  
[info@f5.com](mailto:info@f5.com)

F5 Networks  
Asia-Pacific  
[apacinfo@f5.com](mailto:apacinfo@f5.com)

F5 Networks Ltd.  
Europe/Middle-East/Africa  
[emeainfo@f5.com](mailto:emeainfo@f5.com)

F5 Networks  
Japan K.K.  
[f5j-info@f5.com](mailto:f5j-info@f5.com)

