

Fast, Secure, and Reliable Replication for Midrange Storage Using the Brocade[®] 7810 Extension Switch

Challenges

Today's IT organizations exist in a world where extreme weather-related and natural events—hurricanes, floods, wildfires, and earthquakes—are becoming common occurrences. More than ever, organizations must protect their data for disaster recovery. To ensure proper protection, data must be moved to a location sufficiently far away from any disaster affecting the primary data center. Meeting this distance requirement, however, presents several challenges:

- Achieving sufficient replication performance over distance
- Containing costs for the equipment and long-distance connections that support replication
- Securing data-in-flight between data centers
- Keeping data flowing over unreliable WAN connections

With increasing amounts of data being stored on low-priced, midrange arrays, it is essential to find an affordable replication connectivity solution that does not compromise on performance and enterprise-class features. Traditionally, most replication solutions have focused on large data center environments and high-end storage arrays, but these solutions can be prohibitively expensive for smaller regional sites with just one or two midrange arrays. This paper explores ways that organizations with midrange storage can overcome these challenges and effectively protect their data for disaster recovery.

RPO and RTO

An organization's requirements for replication should be driven by two factors: the Recovery Point Objective (RPO) and the Recovery Time Objective (RTO). These objectives directly impact the replication strategy, the type of equipment, and the type of connections that storage administrators may need to deploy to meet their replication requirements.

The RPO determines how much data organizations can afford to lose in the event of a disaster, expressed in time. For example, at the instant a disaster occurs, how many of the most recent writes of data can an organization stand to lose? The last five minutes? The last few hours? An entire day's worth?

No two organizations are alike, and the specific requirements of the organization will determine the right RPO. For one organization, it may be okay to lose a day's worth of data. If the content from that last day is easy to re-create or does not represent a significant amount of revenue, a day's worth of lost data may be viewed as nothing more than an inconvenience and a temporary loss of productivity. However, for another organization, any lost data may pose a severe problem and result in lost revenue, damaged brand reputation, and potential lawsuits if the loss impacts customers and shareholders. In such cases, a disaster event is not merely a nuisance or a source of frustration—it can be devastating, with the potential to destroy a business.

The RTO determines how quickly an organization must get its data back online after a catastrophic event. In simple terms, once the disaster occurs and access to the primary data center is lost, how long is it okay to be down until access to the data can be restored from a secondary site? A few hours? A few days? A few weeks?

Again, the requirements will vary considerably among organizations, and the RTO is often different than the RPO. For some organizations, the RPO is absolutely critical—as long as they do not lose any data, it is acceptable to wait for access to be restored. For others, it may be acceptable to lose some data, but the RTO is absolutely critical, and systems need to be up and running again as soon as possible.

Figure 1: Time Is Money—Data Protection across Data Centers Matters

\$740K+

Average cost of a data center outage in 2016

38%

Percentage of data center outages is up from 2010

\$8,851

Per-minute loss of revenue and unproductive employees

Source: Ponemon Institute, 2016 Cost of Data Center Outages Report

The RPO and RTO ultimately drive the replication requirements. If the RPO is zero, then synchronous replication is required, and often a three-site replication scheme is deployed (synchronous replication to a nearby site and then asynchronous replication to a faraway site). With a nonzero RPO, asynchronous replication is often the obvious choice. However, even with asynchronous replication, organizations must take care to ensure that the replication scheme can accommodate the I/O traffic throughout the workday/workweek without falling too far behind (and outside the bounds of the RPO).

For the RTO, organizations need sufficient bandwidth and network performance to facilitate access to the data after a disaster has occurred. If that involves replication of data from the disaster recovery (DR) site to a new data center, then there must be sufficient bandwidth to move the data to the new site within the RTO. If the data is to be accessed directly from the DR site, then there must be sufficient network performance to access the data from the remote site.

Achieving Good Replication Performance over Distance

Growing numbers of midrange storage arrays now offer built-in IP replication capabilities that conceptually make replication simple. Storage administrators can just plug the replication ports into an IP router, configure the IP addresses, and go, right?

Unfortunately, it is not that simple. Although establishing a basic connection can be fairly straightforward, it is very difficult to achieve good replication performance over distance through a real-life WAN connection, given its inherent challenges of latency and packet loss.

Latency and packet loss are often called “the mortal enemies of replication performance.” However, all WAN connections experience some degree of latency and packet loss, and nearly all storage arrays suffer from significant droop in replication performance due to latency and packet loss. Some arrays are more sensitive to packet loss, whereas other arrays are sensitive to both packet loss and latency. There is always a performance impact when one or both are present.

Latency affects the time it takes to acknowledge data transmitted across the connection. This means that the longer the distance, the higher the latency. Every replication protocol has some degree of “chattiness”—that is, the replication protocol always involves transmitting a little bit of data, waiting for an acknowledgment, transmitting a little bit more data, and so on. This does not matter when there is no latency because thousands of these acknowledgments would not appreciably affect the replication performance. However, once there is meaningful latency (usually at 10 ms to 20 ms or more of round-trip

time), every acknowledgment now takes a considerable amount of time, and many replication products become “latency-bound.” In other words, the storage array spends most of its time waiting for acknowledgments and is not transmitting new data as quickly across the connection. Once that happens, adding bandwidth does not help. If the amount of latency remains the same, so does the throughput, and the organization cannot take advantage of the added bandwidth.

Packet loss is even more insidious. Packet loss affects replication at the TCP/IP level. When a packet is lost, the usual assumption is that it was due to congestion, and the response at a protocol level is “exponential back-off.” That is a very important principle in a LAN. However, over a WAN connection, this behavior becomes very problematic. Some level of packet loss always occurs over a WAN connection. Usually, the loss is not very significant—perhaps only 0.1%. But if each time the protocol response is to stop transmitting for a while, then even a very small amount of packet loss can have a profound impact on throughput. In fact, performance tests show that just 0.1% packet loss can reduce replication throughput by 90% or more (see Figure 2).

The Brocade® 7810 Extension Switch provides IP extension technology to overcome the effects of latency and packet loss over a WAN connection. The Brocade 7810 Extension Switch features a WAN-optimized TCP stack that is specifically optimized to achieve high replication performance in the presence of latency and packet loss. Optimization at the TCP layer (Layer 4) speeds datagrams to the remote site, even when WAN connections have poor I/O characteristics.

Figure 2: Storage Replication with and without Extension

Test Results in MB/s vs. ms RTT (Latency)

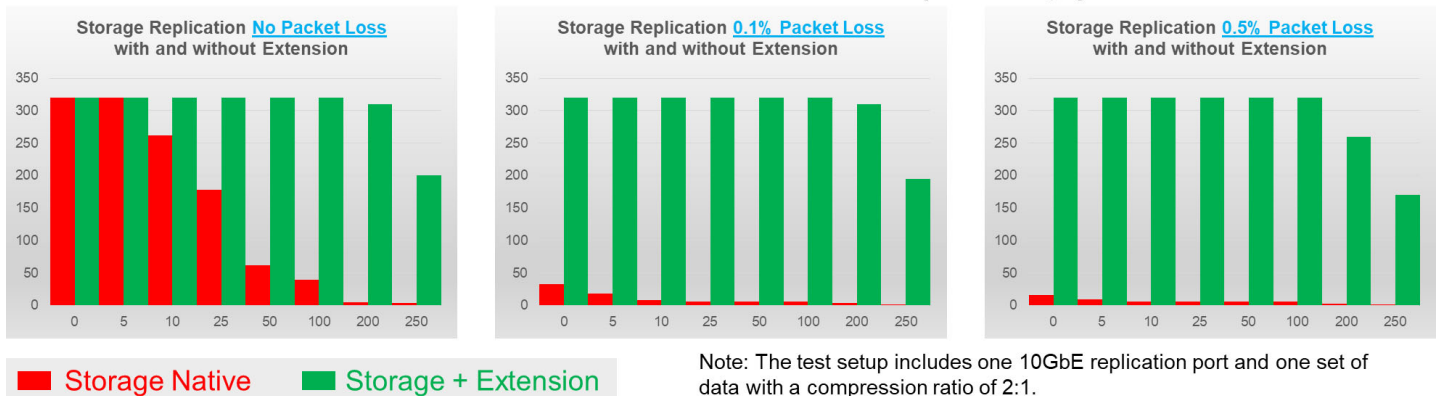


Figure 2 shows the resulting replication throughput when using IP extension technology in the Brocade 7810 Extension Switch. The red line represents midrange storage “native” replication throughput across a WAN connection. The green line represents the replication throughput achieved by running the same replication traffic through a Brocade 7810 Extension Switch, with IP extension technology marshaling the traffic through the WAN and to the storage array at the remote site.

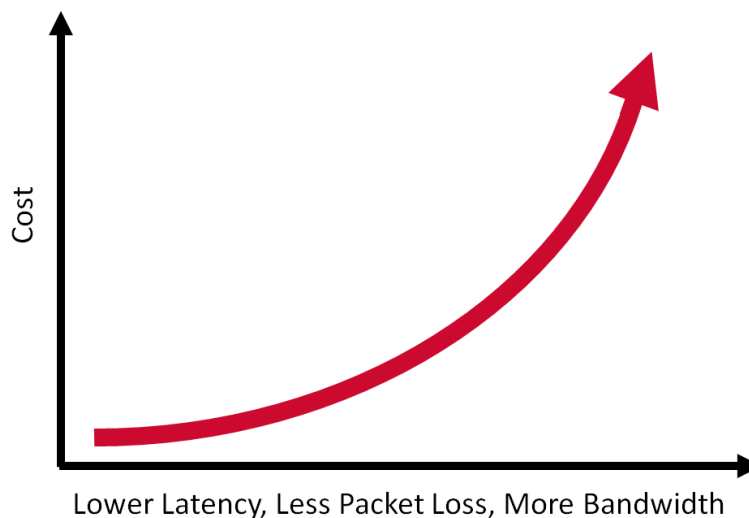
Based on these findings, the Brocade 7810 Extension Switch provides much higher replication throughput than a storage array can achieve on its own when experiencing latency and packet loss. At 25 ms of latency and 0.1% packet loss, the Brocade 7810 Extension Switch can provide 50 times more throughput, and it can deliver that same level of throughput over WAN connections with much higher levels of latency and packet loss.

Containing Costs

Finding a cost-effective way to replicate data over distance to another site can be challenging. Long-distance WAN connections have become less expensive over the years, but pricing varies dramatically (by several orders of magnitude) depending on the specific location that the data needs to be replicated to and from. Pricing is driven by factors such as the amount of fiber in the ground, the available bandwidth, the demand for bandwidth over those connections, and the amount of competition among service providers.

A major factor in connection costs is connection quality. A connection with low latency and low packet loss may be an order of magnitude more expensive than a lower-quality connection. Advanced hardware options can help organizations mitigate these costs.

Figure 3: Costs Increase with More Bandwidth, Lower Latency, and Less Packet Loss



With the Brocade 7810 Extension Switch, organizations can reduce connection costs in two ways:

- Deploying a smaller connection with less bandwidth
- Deploying a connection with inferior I/O characteristics (more latency and packet loss)

The Brocade 7810 Extension Switch has built-in technology that enables organizations to overcome the challenges of latency and packet loss and to dramatically speed up replication performance on lower-cost connections. Performance gains vary among storage arrays, but Broadcom® has yet to see a storage array where IP extension does not provide a dramatic gain in performance. Often these performance gains range from 10 times to 50 times—and occasionally up to 300 times—the throughput of what the array can do on its own.

Not only can the Brocade 7810 Extension Switch provide faster replication than what the array can do on its own over a given WAN connection, it can also provide much faster replication over lower-quality WAN connections. As a result, organizations can potentially reduce costs by deploying a less expensive WAN connection that perhaps takes an awkwardly routed path and therefore has far more latency and packet loss than a more direct connection, and still achieve high replication performance.

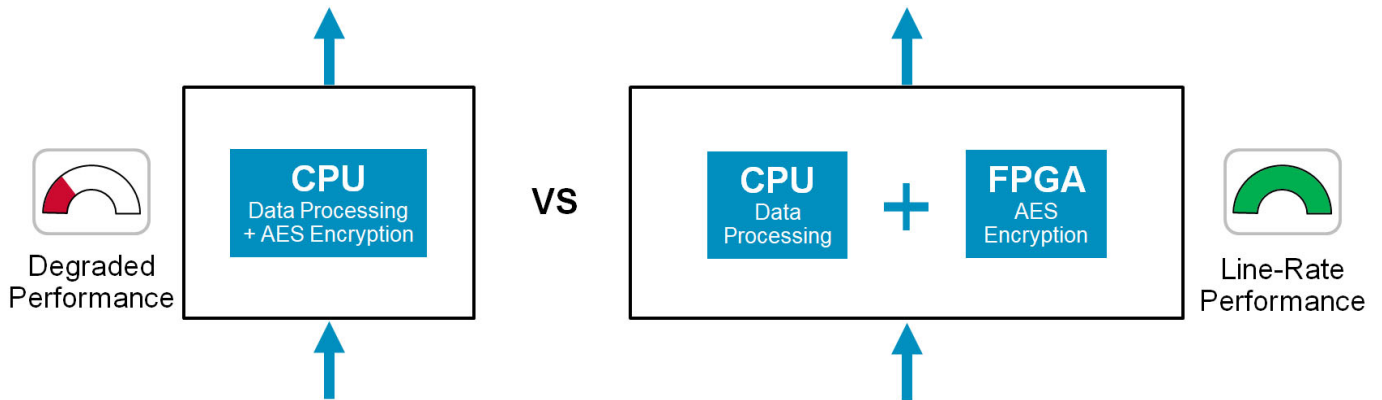
Furthermore, the Brocade 7810 Extension Switch has advanced compression technology that typically provides a compression ratio of 4:1. (The compression is data dependent: The actual compression ratio will vary depending on whether the data is easily compressible or already compressed.) This allows organizations to drive more throughput through limited bandwidth or shared connections.

Securing Data-In-Flight

When organizations replicate data over long distances, they have to protect that data. Unless they own the fiber in the ground and have physical control over the entire length of the connection (the most expensive type of connection), the long-distance connection will be susceptible to tapping somewhere along the way. A typical, cost-effective connection over distance uses shared infrastructure from a service provider, which means that there are many points along the way where data can be intercepted.

If an unauthorized entity gains access to the replication stream, and the stream is not encrypted, the attacker can reassemble the transmitted data and eventually rebuild a full copy of the replicated data. Strong encryption is therefore a fundamental requirement. However, encrypting the data is not simple. Most midrange arrays do not offer encryption of data-in-flight with their replication engines, and for the few arrays that do, there is a severe trade-off in terms of replication performance. When organizations turn on encryption, they often lose 40% to 70% of the replication throughput.

Figure 4: Dedicated Hardware Allows Encryption at Full Line Rate with No Performance Impact and Leaves CPU for Data Processing



The Brocade 7810 Extension Switch can help organizations achieve the desired level of security without compromising on replication performance. It features built-in, dedicated hardware for strong, full-line-rate, 256-bit AES encryption. Since encryption is performed by hardware, throughput is not diminished. Organizations can choose to turn encryption on or off, but there is less than 1% overhead when it is enabled.

For maximum security, organizations should ensure that all data is always protected when replicating between sites. For this reason, Broadcom provides built-in hardware encryption in even the base model of the Brocade 7810.

Keeping Data Flowing over Unreliable WAN Connections

In addition to performance, security, and affordability challenges, organizations must also contend with the poor reliability of WAN connections. The average WAN circuit has more than one disruption event per day. Some of those disruption events may be short—perhaps the WAN circuit is flapping on and off, dropping for a few seconds before the connection is restored again. However, other disruption events may be much longer, lasting hours or more. It is common for multiple service providers to be involved with different segments of a WAN connection. When an outage occurs, it can be difficult to pinpoint the source of the problem quickly and get the connection restored.

Network disruptions can also have a profound impact on an organization’s ability to meet an RPO or RTO. For asynchronous replication, organizations rely on being able to continuously replicate data to keep up with the incoming I/O to the storage array. That incoming I/O continues locally even when the WAN goes down. During the outage, a significant amount of unreplicated and unprotected writes build up. Once the WAN connection is restored, it may take a long time to get the buffer of writes flushed to the other side. When a WAN outage occurs, organizations may quickly find themselves beyond their RPO threshold, and even after the WAN is restored, it can take a long time for them to catch up and meet their RPO.

An even worse scenario is a very long WAN outage in which an organization exceeds the array’s allotted space for unreplicated writes. In such cases, the storage array cannot keep track of all the ordered I/Os and resorts to merely copying changed blocks once the connection is back online. In other words, the array may have to inspect a bitmap of changed blocks to see if they are in sync and replicate all changed blocks without regard for I/O order and consistency. This process can take many hours—or even days—to complete, and during that window of time, the data is not in a protected and safely recoverable state.

The Brocade 7810 Extension Switch can help organizations overcome the reliability issues inherent in most WAN connections. It features technologies that keep data flowing even when problems arise on the WAN.

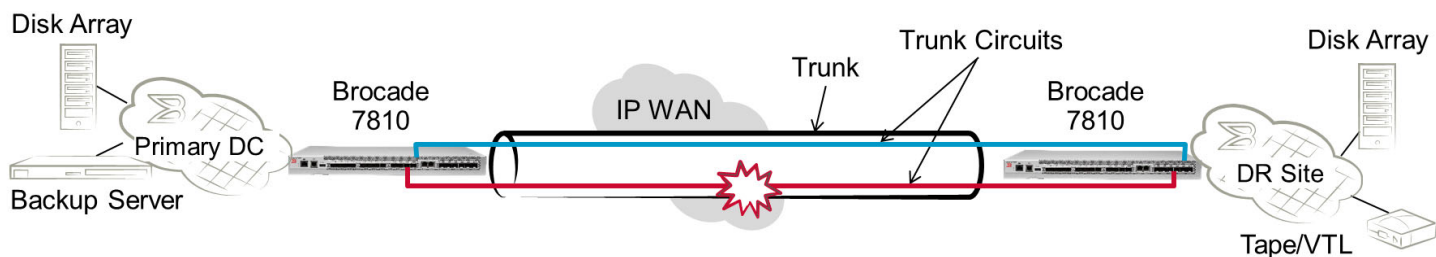
The most critical technology is Extension Trunking. Available in the fully upgraded version of the Brocade 7810 Extension Switch, Extension Trunking allows organizations to aggregate multiple WAN connections, and it performs full active/active load balancing and failover through the various WAN connections.

It is important to note that these WAN connections do not need to have similar characteristics. If multiple WAN links are procured from the same service provider, all of them most likely go through many of the same interconnections and much of the same equipment along the way. As a result, when a backhoe, for example, accidentally digs through and breaks a cable somewhere along the path, or a tree falls on an overhead line, all of these WAN connections go down at the same time.

The most effective strategy to guard against such catastrophic, widespread failure is to procure WAN connections from different service providers and ensure that the connections are running over physically different routes. Not only can this approach protect organizations from WAN disruption events, it may also give them the flexibility to take advantage of competitive service provider pricing.

When organizations procure WAN connections from different service providers, the WAN connections invariably have different I/O characteristics. Perhaps one WAN link has 30 ms of RTT, and the other has 40 ms. One WAN link might have 0.1% packet loss, and the other 0.15%. Or, possibly, one WAN link has 1Gb/s of throughput, and the other 622Mb/s.

Figure 5: Available Bandwidth Is Aggregated, and Traffic Is Load-Balanced across Dissimilar Paths



As shown in [Figure 5](#), the Brocade 7810 Extension Switch can handle such scenarios easily. It aggregates and manages all available bandwidth from each of the WAN links. The switch actively load-balances all traffic flows (FCIP + IP extension) across these dissimilar paths. When a WAN disruption event occurs, the Brocade 7810 Extension Switch retries any lost packets over a surviving WAN link and ensures in-order packet delivery. The Brocade 7810 Extension Switch also provides technologies, such as Adaptive Rate Limiting, that automatically adjust traffic flows to always deliver the maximum available replication bandwidth, especially when something goes wrong along one of the paths or when links are shared with nonstorage flows.

By leveraging technologies such as Extension Trunking and Adaptive Rate Limiting, the Brocade 7810 Extension Switch protects organizations from common disruption events, while always ensuring the maximum available replication throughput.

Summary

Achieving fast, secure, and reliable data replication is critical for meeting RPOs and RTOs. Traditional replication solutions are expensive and out of reach for many organizations with smaller sites that have only one or two storage arrays. The Brocade 7810 Extension Switch provides the technology needed for fast, secure, and reliable replication, in an easy-to-deploy form factor, and at a price point that works for midrange storage arrays.

Copyright © 2018 Broadcom. All Rights Reserved. Broadcom, the pulse logo, Brocade, and the stylized B logo are among the trademarks of Broadcom in the United States, the EU, and/or other countries. The term “Broadcom” refers to Broadcom Inc. and/or its subsidiaries.

Broadcom reserves the right to make changes without further notice to any products or data herein to improve reliability, function, or design. Information furnished by Broadcom is believed to be accurate and reliable. However, Broadcom does not assume any liability arising out of the application or use of this information, nor the application or use of any product or circuit described herein, neither does it convey any license under its patent rights nor the rights of others.

The product described by this document may contain open source software covered by the GNU General Public License or other open source license agreements. To find out which open source software is included in Broadcom products, view the licensing terms applicable to the open source software, and obtain a copy of the programming source code, please visit <https://www.broadcom.com/>.