



DATA CENTER

Brocade Adaptive Rate Limiting

Adaptive Rate Limiting on the Brocade® 7800 Extension Switch and Brocade FX8-24 Extension Blade for the Brocade DCX® 8510 and DCX Backbones enables efficient and cost-effective sharing of WAN bandwidth for FCIP storage and non-storage applications. This paper details the operation and advantages of Brocade Adaptive Rate Limiting in shared WAN bandwidth environments.

BROCADE

CONTENTS

Introduction3

Shared Bandwidth3

Brocade ARL Defined3

The Need for Adaptive Rate Limiting5

The WAN Optimization Myth7

Summary8

INTRODUCTION

Common to all Remote Data Replication (RDR) systems is some type of WAN connection. WAN characteristics include propagation delay, often referred to as Round Trip Time (RTT), bandwidth, and the type of connection such as Multiprotocol Label Switching (MPLS), Dense Wavelength-Division Multiplexing (DWDM), and Carrier Ethernet. The use of WAN optimization, which is discussed at the end of this paper, may also be characteristic. However, there is yet another characteristic that is of considerable importance in RDR network design, an attribute that causes many difficulties in operations. This attribute is whether the WAN is dedicated or shared among multiple storage applications or shared with other non-storage traffic.

More and more, enterprises are opting to use a shared WAN of considerable bandwidth for RDR, rather than a dedicated connection. There are many economic reasons for this. This paper discusses how Brocade Adaptive Rate Limiting (ARL) is used to facilitate the proper operation of RDR across a shared WAN connection.

Implementing synchronous RDR (RDR/S) over any type of shared bandwidth infrastructure is not considered best practice and is not discussed at length in this paper. Tape and asynchronous RDR (RDR/A) can be implemented over a shared bandwidth infrastructure, which is the focus of this paper.

SHARED BANDWIDTH

Shared bandwidth is now common in networks used by RDR and BURA (Backup, Recovery, and Archive); however, sharing bandwidth with these applications often presents difficult problems. Here are examples of shared bandwidth:

- Bandwidth shared between multiple Fibre Channel over IP (FCIP) flows originating from another interface on the same or different Brocade 7800/FX8-24s
- Bandwidth shared with non-storage-related applications such as user traffic (web pages, email, enterprise-specific applications) and Voice over IP (VoIP)

Sharing bandwidth poses a problem for storage traffic. FCIP traffic is, in essence, Inter-Switch Link (ISL) traffic, and FC ISL traffic characteristically makes certain demands on a network. For instance, storage traffic typically consumes a considerable amount of bandwidth. This depends on the number of volumes being replicated and the amount of data being written to the volumes within a given period of time.

In general, shared bandwidth is neither best practice nor recommended for storage; however, considerations of practicality and economics might make it necessary for organizations to share the same link with other user applications. Experience has demonstrated that allowing storage traffic over shared bandwidth—without planning and without implementing mechanisms that facilitate storage transmission—either fails in practice or is plagued with problems. Mechanisms to facilitate storage transmission include: Adaptive Rate Limiting (ARL), massive overprovisioning of bandwidth, Quality of Service (QoS), and Committed Access Rate (CAR). Each of these can be used alone or in combination with the others. This paper focuses on Brocade ARL.

BROCADE ARL DEFINED

Here are some characteristics of Brocade ARL:

- It is used primarily when a link has to share its bandwidth among multiple FCIP interfaces or other user traffic; in other words, a single interface does not have its own dedicated WAN link.
- It is a rate limiting technique exclusive to Brocade 7800 and FX8-24 extension platforms.
- It is not part of TCP (Transmission Control Protocol), nor does it work by adjusting TCP windows or TCP functionality.
- It uses TCP to obtain current traffic conditions over the network path.

- Configurable by the user, Brocade ARL limits traffic to a maximum bandwidth up to full interface speed. The aggregate of the maximum bandwidth values across multiple circuits assigned to an Ethernet interface can exceed the bandwidth of the interface. Brocade ARL permits circuits that are active during one time of day to use all the bandwidth on the same interface that another circuit uses during a different time of day.
- Configurable by the user, Brocade ARL reserves a minimum amount of bandwidth (as low as 10 Mbps) on an interface. The aggregate of the minimum bandwidth values for multiple circuits assigned to an Ethernet interface cannot exceed the bandwidth of the interface. The minimum values cannot oversubscribe the interface, which always permits at least that much bandwidth to exist for that circuit.

Tape and replication traffic are sensitive to latency; even asynchronous traffic can be negatively affected, primarily in terms of reduced throughput. There are Small Computer Systems Interface (SCSI) and Fibre Connection (FICON) protocol optimization methods such as FastWrite and FICON emulation, which mitigate the effects of latency; however, optical propagation delay is not the only source of latency. Latency is also caused by buffering delays during periods of congestion, as well as retransmission due to packet loss. Rate limiting in general must be used to counter these problems, otherwise network performance degrades.

Proper rate limiting of FCIP traffic into an IP network prevents packet loss. Feeding unlimited FCIP into an IP network that has less capacity results in congestion and packet loss, both of which can cause retransmission and added utilization of the WAN. To make the situation worse, when congestion events occur, TCP limits transmission as a flow control mechanism in an effort to dissipate congestion and prevent it from immediately happening again. Using TCP as a flow control mechanism is inefficient and results in poor overall performance. Simply said, you cannot feed more data into a network than it is capable of transmitting; otherwise, data is going to get lost in transmission and performance is going to suffer. It is essential, therefore, to provide the network with a comfortable data rate that avoids such problems. Since the inception of FCIP, this has been accomplished through simple static rate limiting. This approach, however, leaves a lot to be desired.

There are three common scenarios in which Brocade ARL is effective:

1. Storage and non-storage traffic over a single link
2. More than one FCIP interface feeding a single link that has been dedicated solely to storage
3. A combination of these two scenarios, which in essence works out to scenario 2. In all three situations, the link may be oversubscribed and rate limiting required to prevent congestion and packet loss.

Sharing a link with storage and non-storage traffic is the most problematic. User data tends to be unpredictable and very bursty, causing bouts of congestion. QoS can be used to give storage a higher priority. CAR can be used to logically partition the bandwidth over the same WAN connection. Brocade has developed SO-TCP (Storage Optimized TCP), which is an aggressive TCP stack that facilitates and continues storage transmission—even after experiencing congestion events—by maintaining throughput in adverse conditions. This means that other TCP flows on the shared link back off much more than SO-TCP does, which helps free bandwidth for storage. SO-TCP cannot overcome UDP-based traffic, as UDP has no flow control mechanisms; therefore, it is best to prevent collision with such traffic and to use Brocade ARL as the mediating mechanism.

Brocade ARL is easily configured by setting a ceiling (-B) and a floor (-b) bandwidth level per circuit. Brocade ARL always attempts to maintain the floor level, and it does not attempt to exceed the ceiling level. Brocade ARL tries to increase the rate limit up to the ceiling until it detects that no more bandwidth is available. If it detects that no more bandwidth is available, and it is not at the ceiling level, it continues to periodically test for more bandwidth. If it determines that more bandwidth is available, it continues to climb up to the ceiling level. On the other hand, if congestion events are encountered, Brocade ARL drops down to the floor level and attempts to climb up again.

Brocade ARL plays an important role in shared links with user (non-FCIP) traffic. Brocade ARL backs off the current rate limit and permits user traffic to pass, if congestion events are detected. Typically, within an enterprise an agreement is made determining the minimum bandwidth that storage receives across a shared link. However, if user traffic is not using its portion of the bandwidth, storage traffic uses that portion. When storage gets to use extra

bandwidth from time to time, RDR/A applications then have a chance to catch up, if they are behind due to interims of high load. Brocade ARL facilitates these types of SLAs (Service Level Agreements).

In the case of multiple FCIP interfaces feeding a single link, the sources can be from a variety of storage devices, such as tape and arrays. Rarely does a host communicate with storage across a WAN due to poor response times and reliability. Alternatively, the source may be a single array with “A” and “B” controllers connected to respective “A” and “B” Brocade 7800/FX8-24s in a high-availability (HA) architecture, but utilizing only a single WAN connection. In any case, the link is oversubscribed and congestion occurs.

THE NEED FOR ADAPTIVE RATE LIMITING

Static rate limiting does not provide optimal operation in all situations. There are three primary areas of interest:

- Oversubscribed during normal operation
- Undersubscribed during maintenance/failure
- Changing application demands

“Oversubscribed during normal operation” means that when all devices are online, and there are no failures in the network, the available bandwidth is oversubscribed by some amount. As shown in Figure 1, the interfaces on the Brocade 7800/FX8-24 have been configured to statically rate limit for the bandwidth of the link. This example uses an OC-12. If multiple interfaces are rate limited to 622 Mbps and feed a single OC-12, it oversubscribes the link, and the system suffers poor performance. This is the worst case scenario. The motivation for this configuration is to maintain full utilization of the link during periods when a path is offline due to a channel extender, IP network device, or array controller outage, which segues to the next scenario.

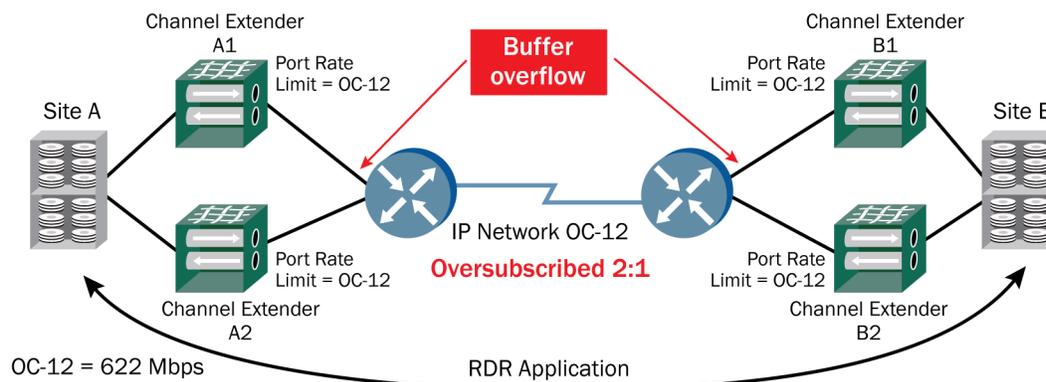


Figure 1. Oversubscribed WAN link scenario

In the event that a Brocade 7800/FX8-24, LAN switch/router, or array controller is taken offline for maintenance, technology refresh, or due to failure, a path becomes unavailable. In the previous scenario this is not a problem, because the link is no longer oversubscribed. Normal operation, however, does not work properly due to oversubscription.

In the case of being “undersubscribed during maintenance or failure,” during normal operation there is no oversubscription, which permits proper operation. During times of failure or maintenance there is undersubscription of the WAN, which presents a problem of a different kind. This may be fine for many organizations, but for some this is just not acceptable. It is imperative that during periods of maintenance and failures there is no increased risk of data loss. Exactly that situation occurs when half the total available bandwidth is no longer available, as shown in Figure 2.

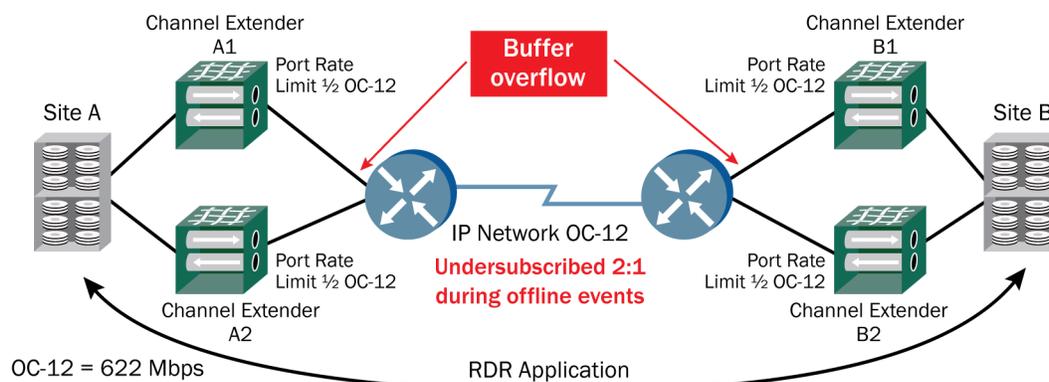


Figure 2. Undersubscribed WAN link scenario

If suddenly only half the bandwidth is available, and assuming that more than half the bandwidth is used on a regular basis, this creates a situation in which an RDR/A application has to start journaling data because of insufficient bandwidth. This data is stored on disk journals or cache side files, which have limited space. Depending on the load during the outage, these journals can fill quickly, as they are not meant for storage beyond perhaps a couple of minutes or hours. The journal has to be large enough to hold all the writes that are replicated during the outage. Once it fills to maximum capacity, the RDR/A application has no choice but to sever the local and remote volume pair relationship. When connectivity is restored, an initial resync of the volume pair is required. From the time the journals start to fill, through losing and regaining pair status, and up to completing initial resynchronization of the pairs, data is not protected by RDR. This can be a lengthy period of time, especially if the volumes are very large. During all this time, data is not protected, which exposes the enterprise to a significant liability. This is why many companies are interested in maintaining available bandwidth at all times and why it is risky not to do so.

The problem described in the previous paragraph can be solved by simply purchasing a link with twice the bandwidth than is actually needed or will be used, and by setting traditional rate limiting to half that bandwidth. During times of failure, upgrades or maintenance—when one path is offline—only half the bandwidth can be used, due to static rate limiting. This is not a problem with an overprovisioned link, because there is twice the needed bandwidth, so half the bandwidth works well. Obviously, this is not an optimal or cost-effective solution, because it means that during normal operation half the bandwidth goes to waste.

Brocade ARL permits more efficient use of that bandwidth, reducing Operational Expenses (OpEx) and delivering a higher Return on Investment (ROI). With Brocade ARL, a proper size link can be provisioned, with the ceiling set to the link bandwidth and (assuming two FCIP interfaces) the floor set to half of the link bandwidth. If there are four FCIP interfaces, the floor is set to $\frac{1}{4}$ of the link bandwidth. Of course, manual intervention by resetting the rate limit settings can help here as well; however, the downside is that this requires more management, more chance of error, more change control, middle of the night and holiday problems, and increased operational costs.

Consider buying and using only the bandwidth that is actually needed. If the remaining online interfaces can automatically adjust their rate limit to compensate for any temporary offline interfaces, then all the bandwidth will be consumed all the time. This is the problem that Brocade ARL addresses—by automatically adjusting the rate limiting on FCIP interfaces. It maintains bandwidth during outages and maintenance windows, which does several things:

1. It reduces the likelihood that an RDR/A application will have to journal data.
2. It reduces the risk of data loss.
3. It facilitates provisioning the proper amount of bandwidth versus excessive unused amounts.

Brocade ARL can perform another important task; it allows multiple storage applications, such as tape and RDR, to access bandwidth fairly. As an application's demand diminishes, the other application(s) can utilize the unused bandwidth. If two storage applications that are assigned to two different FCIP interfaces and attached to the same WAN link are both demanding bandwidth, they adaptively rate limit to an equalization point at which they both get

equal bandwidth. No application can preempt the bandwidth used by another application beyond the equilibrium point. However, if the demand of one application tapers off, the other application can then utilize any unused portion of the bandwidth. In Figure 3, you can see that as the blue flow tapers off, the red flow starts to increase its consumption of bandwidth. Conversely, if the blue application later needs more bandwidth again, it preempts the red application's use of the bandwidth up to the equilibrium point, but not beyond it. In this way, bandwidth is fully utilized without jeopardizing either of the storage applications.

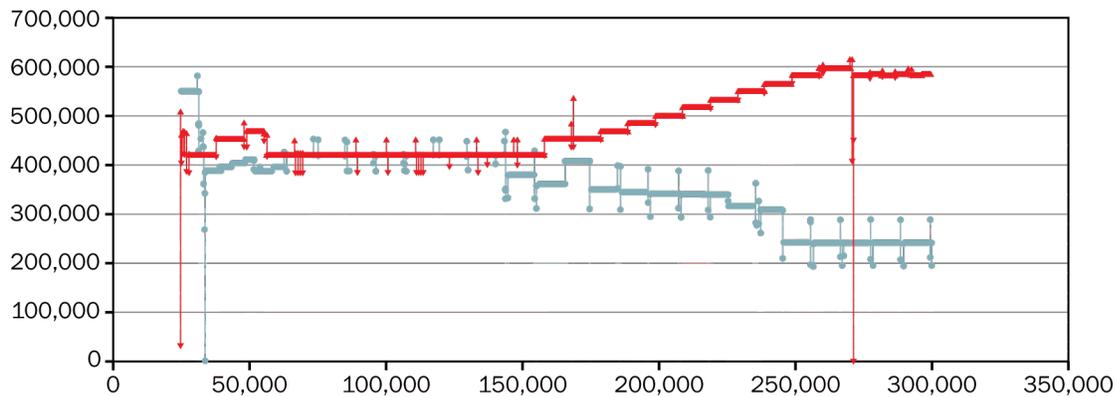


Figure 3. Application demand sharing a single WAN link

THE WAN OPTIMIZATION MYTH

Clearly, WAN optimization is *not* required for high-performance FCIP extension. In fact, third-party WAN optimization products often provide no benefit over the extensive, sophisticated, and very advanced optimizations already engineered into the Brocade 7800/FX8-24 extension platforms. Brocade FCIP extension devices are designed to provide high compression rates and ratios. The compression and deduplication of data by third-party WAN optimizer solutions tend to be less effective than what is provided as standard in the Brocade extension platforms. There are a variety of reasons for this.

One reason is that a better TCP stack than the Brocade SO-TCP does not exist for storage applications. The Brocade SO-TCP is an optimized, high-performance TCP stack capable of large Bandwidth Delay Products (BDPs) and very fast recovery. BDP is the product of the bandwidth x RTT (Round Trip Time), which is the amount of data in flight that TCP must be capable of managing. The Brocade FX8-24 blade can manage 200 MB of data in flight, with 1 percent packet loss on each 10 GbE interface, and can continue to maintain exceedingly good throughput. This means that an OC-192 link has to span more than 16,000 km (approximately 10,000 miles) before exceeding the capabilities of the Brocade SO-TCP. The Brocade 7800/FX8-24 GbE interfaces maintain maximum throughput over very high latency links, up to 200 ms RTT as well.

Another reason is that SCSI, Fibre Channel Protocol (FCP), and FICON protocol optimization are provided on the Brocade 7800/FX8-24 (and are not provided by any of the WAN optimization products on the market), and Brocade protocol optimizations deliver a great deal of acceleration. With the latest generation of Brocade FCIP extension products, the Brocade 7800 and FX8-24, there is little to no value proposition for adding third-party WAN optimizers. Brocade ARL is also used to manage multiple flows and optimize bandwidth utilization.

Also, WAN optimizers have issues with the amount of data that can be optimized. WAN optimizers have been primarily designed to relocate remote office equipment back into the data center, which constitutes considerably less data to learn. For example, can a WAN optimizer optimize the number of terabytes being replicated? The data store on the WAN optimizer would have to be considerably large; otherwise, it is going to cycle its data store, which causes a continuous “cold” state, meaning that data has to be constantly and repeatedly learned.

In addition to the data store size problem, RDR—which in large part consists of database disk data—tends not to contain frequently repeated patterns. New data patterns are referred to as “cold,” meaning that a smaller size token

has not been established at each end; therefore, no token can be sent the first time. WAN optimizers are strongest with “warm” data, which is consistent with Network File System (NFS), Common Internet File System (CIFS), and other remote office user data. WAN optimizers are *not* strong with RDR data, due to the large quantity of “cold” data. For “cold” data, WAN optimizer devices must process the data using traditional compression mechanisms, which are the same as those found on the Brocade 7800/FX8-24. In fact, because of the storage-specific optimizations Brocade has developed for its FCIP compression algorithms, the Brocade 7800/FX8-24 achieves better compression as compared to WAN optimization devices that encounter “cold” data, and a large quantity of RDR data is “cold.”

Finally, WAN optimizers do not have the bandwidth capacity that the Brocade 7800/FX8-24 has. High-bandwidth applications above 1 Gbps (on the WAN side)—which are often associated with, but not limited to, tape—means the WAN optimizer devices become a bottleneck.

WAN optimizers should not be used to optimize FCIP traffic; however, they can certainly be used across a shared link. WAN-optimized user traffic can coexist with Brocade FCIP traffic. A shared link uses WAN optimization for user applications only. In parallel, FCIP data connects downstream closer to the WAN, or not optimized as it passes through the WAN optimization device, as shown in Figure 4. WCCP (Web Cache Control Protocol) or PBR (Policy-Based Routing) on a router redirects specific traffic to the WAN optimizer and should be configured to exclude the storage FCIP traffic. This is a common design and a well-known architecture. Brocade ARL facilitates coexistence and high performance in such environments, as described in the section on shared links with storage and non-storage traffic.

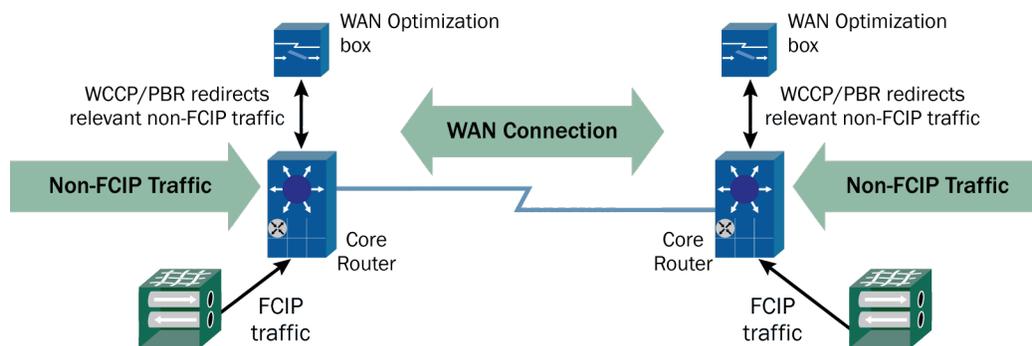


Figure 4. FCIP and user traffic WAN optimization can coexist.

SUMMARY

For many growing enterprises, dedicating WAN bandwidth to individual applications or data flows is increasingly inefficient and cost-prohibitive. But to avoid overprovisioning and effectively share IP WAN bandwidth across multiple FCIP storage applications, such as RDR/A and tape, or across storage and non-storage applications, requires advanced extension capabilities and dynamic rate limiting. Brocade ARL, available on the industry-leading Brocade 7800 and Brocade FX8-24 extension platforms, provide the performance and advanced capabilities you need to safely share bandwidth and ensure that all available bandwidth is fully utilized—all the time.

© 2012 Brocade Communications Systems, Inc. All Rights Reserved. 01/12 GA-TB-419-00

Brocade, Brocade Assurance, the B-wing symbol, DCX, Fabric OS, MLX, SAN Health, VCS, and VDX are registered trademarks, and AnyIO, Brocade One, CloudPlex, Effortless Networking, ICX, NET Health, OpenScript, and The Effortless Network are trademarks of Brocade Communications Systems, Inc., in the United States and/or in other countries. Other brands, products, or service names mentioned may be trademarks of their respective owners.

Notice: This document is for informational purposes only and does not set forth any warranty, expressed or implied, concerning any equipment, equipment feature, or service offered or to be offered by Brocade. Brocade reserves the right to make changes to this document at any time, without notice, and assumes no responsibility for its use. This informational document describes features that may not be currently available. Contact a Brocade sales office for information on feature and product availability. Export of technical data contained in this document may require an export license from the United States government.