

Centralized Data Backup

It Doesn't Have to Cripple Your WAN

CENTRALIZED DATA BACKUP

Introduction

The need to be close to customers, manufacturing facilities and specialized labor have required organizations to extend the traditional concept of “headquarters” to offices and factories hundreds or even thousands of miles away. However, along with the opportunities that come with workforce globalization, come the realities of dealing with data that sprawls across the organization. Whether the data is at the Munich branch or at HQ in New York, it is equally susceptible to loss, requiring that data recovery and security plans apply to all parts of the organization, regardless of location.

To protect company data and ensure its availability to users, IT organizations have been conflicted between two backup approaches. The first approach, local tape backup, requires that tape libraries be present wherever there are servers in racks. Local area network (LAN) access to the servers gives administrators fast data backup and recovery. The newer approach, centralized backup, puts high-density tape libraries in one location to which data from servers around the world is backed up. While centralized backup requires less hardware, reduces administration time, and solves the security problem associated with loose tape media, it can introduce greater bandwidth consumption and longer backup/restore windows. Because of these issues, centralized backup has been a leap some managers have not been willing to make.

With the right wide-area data services (WDS) technology, a more scalable and secure data protection model can be implemented without the expense of an expanded wide area network (WAN). WDS is a superset of several network acceleration categories, including data reduction and compression as well as protocol and application optimization. WDS can eliminate the bandwidth and time constraints that stall many centralized backup deployments and are the primary enabler of many technology consolidation projects.

The pioneer in WDS, Riverbed® offers a family of appliances designed to work with your existing data and network infrastructure. Riverbed Steelhead® appliances can accelerate your centralized backup and recovery processes, reduce bandwidth usage, and enable data and IT consolidation. The Steelhead appliance frees business from the constraints of latency and bandwidth, allowing your company to save millions of dollars by utilizing centralized technology solutions with unlimited scalability.

This whitepaper will help you understand the tremendous shift occurring in data protection. Leveraging the capabilities of WDS, storage specialists are replacing local tape libraries with more scalable and secure network-based systems. Security analysts are no longer drafting extensive procedural documents to manage a growing swarm of loose tape media. CIOs are revising disaster recovery SLAs to reflect new levels of reliability, and CFOs are planning budgets on more practical technology growth. The Steelhead appliance will not only enable IT organizations to efficiently manage the increasingly critical technology infrastructure of today's business, but will also facilitate compliance with new government regulations mandating the security of sensitive company data.

The Evolution from Local to Central Tape Backup

Though a very small bandwidth consumer, the local backup model is rapidly being supplanted by more manageable architectures. IT departments are finding the headaches and costs associated with maintaining an increasingly complex matrix of distributed tape technologies can be alleviated by performing backup over the WAN. But the move toward replacing the local backup model is also driven by the security implications of keeping sensitive data on thousands of tape cartridges at remote offices across the world. Tapes can be lost or stolen, compromising not only disaster recovery efforts but the protection of company data from competitors, thieves or others intending harm to the company or its customers.

Providing better control over loose media and reducing administrative overhead with standardized tape media, centralized backup provides a more secure and scalable approach to data protection. Centralized backup can be successfully deployed over metropolitan-area networks (MANs), where low-latency, high-bandwidth branch inter-connectivity has provided a feasible, low-cost medium for bulk data transport. Extending this model beyond the MAN, however, introduces high connectivity costs and latency issues that cannot be overcome with more bandwidth alone. IT departments have had to look to smarter technology to help scale data protection architectures beyond a few city blocks to anywhere in the world.

The Emergence of Disk-Based Backup

For faster and more reliable access to protected data, backup technology has moved beyond tape cartridges to spinning disk. Using disk-to-disk and more scalable volume management systems, technologies like Network Appliance's NearStore VTL allow low-latency data to be backed up or restored in a fraction of the time required with tape. And with specialized support for transactional databases that allows for differential point-in-time snapshots, software like Network Appliance's SnapManager or

EMC's Replication Manager further reduces data backup and recovery windows. These innovations open the way for IT organizations to build security and scalability into their data protection architecture. Yet without a WDS strategy to extend these benefits to distant remote branches, centralized backup cannot operate beyond the MAN without also creating congested pipes and soaring bandwidth costs.

Solution	Benefit	Cost
Local Backup	<ul style="list-style-type: none"> • Shorter backup times • Reduced bandwidth 	<ul style="list-style-type: none"> • More hardware and staff • Security risks
Central Backup	<ul style="list-style-type: none"> • Less hardware and staff 	<ul style="list-style-type: none"> • Increased bandwidth costs • Increased backup times
Central Backup with the Steelhead appliance	<ul style="list-style-type: none"> • Shorter backup times • Reduced bandwidth • Less hardware and staff 	<ul style="list-style-type: none"> • One-time technology investment

How Wide-Area Data Services Accelerate WAN Backup

Whether protecting remote data onto centralized tape or disk backup systems using NetBackup, or replicating company assets between redundant data centers with SnapMirror, WDS solutions enable organizations to move data between sites without the constraints of distance and throughput. The Riverbed Optimization System (RiOS™), software that powers Riverbed's Steelhead product family, accelerates applications typically by five to 50 times and in some cases up to 100 times faster than conventional transport mechanisms with up to a 95% reduction in WAN bandwidth utilization. This technology delivers immediate benefits to remote data protection processes, transporting backup payloads that don't congest existing circuits and providing remote data recovery at rates fast enough to match local tape metrics.

RiOS consists of four tiers, each streamlining the inherent inefficiencies of applications across the WAN:

- Data streamlining ensures the same data never crosses the WAN twice;
- Transport streamlining eliminates dead time on high-latency circuits;
- Application streamlining accelerates popular business applications for remote users to speeds available at the central office;
- Management streamlining facilitates deployment and administration of WDS technology.

A key value of RiOS is that it is designed to accelerate all traffic that runs over TCP. While this paper is designed to focus on the backup and recovery process, it is important to note that this same technology can be used to accelerate file sharing, email, document management, and all other applications that are important to your organization. For greater detail on RiOS, please refer to the [RiOS Technical Overview](http://www.riverbed.com) whitepaper available on www.riverbed.com.

Data Streamlining: Eliminates redundant bytes from transfers

For all differential data that is read by a backup process, more than 95% of it hasn't changed. Steelhead appliances are aware of all the data that travels over the WAN. Data streamlining reads into every file and recognizes repeated data patterns. Whether that data is present in one or one hundred files, Data Streamlining ensures it never travels across the WAN twice. Where NetBackup and SnapMirror reduce the raw redundancies of daily backups, the Steelhead appliance reduces the bandwidth resources these technologies need by 95%, altogether eliminating backup-induced congestion on the WAN.

Transport Streamlining: Eliminates transport protocol inefficiencies

Backup processes are big network consumers. Using only a single connection to a remote agent, a backup server will pull anywhere from 5% to 100% of the contents of a file server. Efficient data transport is critical to completing the backup process within an allotted window. Inherent limitations to the TCP protocol introduce idle time and retransmissions on data streams, which lengthen overall data transport times by as much as 50x beyond a circuit's optimal throughput rate. Transport streamlining incorporates both patented and industry-accepted features to fully maximize the power of TCP. These include virtual window expansion, window scaling, explicit congestion notification, limited and fast retransmits, adaptive initial congestion windows and bandwidth delay control.

A Technical Examination of the TCP Protocol as a WAN Bottleneck¹

Assuming an 8k TCP window and 75ms latency over a T3 between a file server in Los Angeles and a tape library in New York, effective bandwidth can be calculated by dividing the TCP window by latency:

$$(8\text{kbyte}/75\text{ms}) * 8 = 853\text{kpbs}$$

At 853kpbs, a 500Mbyte payload would require 62,500 round trips and approximately 80 minutes to complete. With Transport Streamlining, Steelheads will learn the latency between two sites and dynamically adjust the window to maximize WAN bandwidth:

$$(420\text{kbyte}/75\text{ms}) * 8 = 45\text{Mbps}$$

A Steelhead-enabled network would allow a 500Mbyte backup to complete in 1200 round trips and in less than 90 seconds.

Application Streamlining: Optimizes application WAN performance

Data protection processes either employ proprietary agent-based protocols or rely on the underlying application protocol for data packaging and re-assembly. Where well-known file system protocols like CIFS are used, RiOS employs protocol-specific application streamlining to minimize the total number of round trips required to move large numbers of files over the wide-area. By proxying client-server communications on both ends of the pipe and executing proprietary transaction prediction algorithms, Steelhead appliances ensure data protection processes have maximum access to bandwidth when they need it.

Management Streamlining: Enables transparent deployment and centralized management

Network infrastructures have grown more complex than just routers and switches. Today's network administrators are responsible for a host of specialized appliances, including firewalls, IPS devices, authorization gateways, load balancers, and network monitors. Distributed computing has also seen the evolution of single servers into large Web farms, and tape drives into self-maintaining robotic libraries. The success of any additional technology will hinge on its ease of deployment. Steelhead appliances require no configuration changes on any other devices and will accelerate any TCP traffic transparently with little configuration out-of-the-box. Steelhead appliances can be administered through an intuitive CLI or a feature-rich Web interface that also provides powerful insight into traffic flows on the WAN.

A Technical Examination of the CIFS Protocol as a WAN Bottleneck²

Assume a 1Gbyte payload split over 1200 files to be copied between a Windows file server in Los Angeles and its replication partner in New York with the following network parameters:

- Streamlined TCP optimization
- 75ms latency cross-country
- 45Mbps throughput

The total number of round trips can be calculated by adding the sum of all protocol dialog to the sum of data windows dispatched over the WAN. Protocol dialog for each file copied over CIFS can require approximately six round trips for setup and one round trip per data window acknowledgment. The CIFS protocol requires a four to 32kbyte data window to be received before an acknowledgment can be transmitted. High-latency connections typically keep the data window near 4kbytes, but for this exercise a faster 32kbyte window is used.

The typical number of round trips required to process all setup operations for 1200 files is $6 \times 1200 = 7200$.

The total number of round trips required to acknowledge the full payload is $1\text{Gbyte} / 32\text{kbytes} = 31,250$.

The total number of round trips required to transfer the data itself is equivalently $1\text{Gbyte} / 32\text{kbytes} = 31,250$.

The sum total of round trips is $7200 + 2 * 31,250 = 69,700$. At 75ms latency, the total time to complete the transfer is $69,700 \times 75\text{ms} = 87\text{ min}$.

With Application Streamlining, protocol dialog is virtually eliminated and data is dynamically packaged to fill the pipe. For a streamlined transfer, the total number of round trips becomes a function of the pipe size itself. A 1Gbyte transfer would require $1\text{Gbyte}/45\text{Mbps} = 3\text{ min}$ to complete, or only about 2400 round trips.

Accelerate your Backup Processes Immediately

With Steelhead appliances powering your data center and remote offices, data protection processes will benefit from gains in performance by five to 50 times and in some cases up to 100 times, and bandwidth reduction by up to 95%. In addition, as Steelhead appliances “learn” data that is traveling between sites, data transfer times continue to decrease.

Consider the scenarios where RiOS technology can improve your data protection operations:

1. A desktop drive at a remote sales office fails and its user files need to be restored to a rebuilt PC. Restoring 2GB of documents from the central office used to take four hours where now it takes just under five minutes.
2. Backing up the remote office's file server used to take eight hours with a 25% chance of completing successfully. Now backups take 25 minutes and succeed 100% of the time.
3. Running a backup or restore always meant access to headquarters was slow. Now, data protection processes can be run during business hours with little to no impact to branch users.
4. Pushing out an operating system patch to a remote office required either a remote file server as a distribution point or an all-night push from the central office. Now a 10Mbyte patch can be installed onto 40 remote machines in under two minutes.

Case Study: Backup windows reduced from a day to a few hours

Mercury Interactive (www.mercury.com), the global leader in business technology optimization (BTO), is one of the largest enterprise software companies in the world. Nearly 90 percent of Fortune 100 companies and dozens of government agencies globally rely on Mercury to fulfill their BTO strategies.

Mercury has four key development locations around the world. Kirkland, Washington, Boulder, Colorado, and Yehud, Israel are all connected to the Sunnyvale, California headquarters via an MPLS network. Throughput ranges anywhere from 3 to 5 megabits per second on each WAN connection, and latency ranges from 50 to 200 milliseconds.

“The challenge for the Mercury IT team,” notes their Director of IT, “is to develop an infrastructure that can effectively support our development process around the world.”

“Given our long distance WAN connections and the latency on those connections, throughput has always slowed down application performance,” he continued. “It was common for us to see throughput of approximately .5 Gigabytes of data per hour, or about 10 Gigabytes per day when backing up data. Even with SnapMirror's differencing algorithms, our backups could take up to 24 hours. If backups ever exceed 24 hours, we're in danger of not having up-to-date data in the event of a mishap.”

The IT Director remarked, “The Riverbed Steelhead appliances have dramatically shortened our backup windows. Backup of remote Filers now takes roughly 6 to 8 hours, instead of 24 hours.”

Because of the Steelhead appliance's transparent deployment, accelerating SnapMirror takes no extra configuration or integration. He recalled, “We were quite surprised when one day our backup process took 25% of its usual time, because we were busy analyzing the effects of Riverbed Steelhead appliances on our FTP transfers. We made no changes to our software – the Steelhead appliances just started working right away.”

Additionally, because of the protocol-independent nature of the Steelhead appliance architecture, Mercury accelerates all other TCP applications that run over the corporate network.

Conclusion

To reduce costs and meet today's stringent security standards, IT organizations are centralizing data protection architectures. WDS technology accelerates application performance over the wide-area and provides networks with the appropriate interface for data transport between geographically-disparate sites. Riverbed's Steelhead appliances extend the life of existing network infrastructure by minimizing the bandwidth required by distributed backup processes. Steelhead appliances also can provide datacenters with low-latency data concurrency for unmitigated disaster preparedness.

Application and protocol independent, Steelhead appliances will optimize all TCP traffic and can easily serve future initiatives to further centralize technology infrastructure up to and including remote file servers themselves. With Steelhead appliances, it is easy to roll out a low-cost, high-speed centralized tape backup solution that will stay off the network and eliminate data protection headaches, beginning today.

About Riverbed

Riverbed Technology is the performance leader in wide-area data services (WDS) solutions. By enabling application performance over the wide area network (WAN) that is orders of magnitude faster than what users experience today, Riverbed is changing the way people work, and enabling a distributed workforce that can collaborate as if they were local.

Riverbed's Steelhead appliance has been named *InfoWorld's* "Technology of the Year" in both 2005 and 2006 as the "Best WAN Accelerator". In addition, Riverbed was named the winner of *The Wall Street Journal's* 2005 Technology Innovation Award in the Network/ Broadband/Internet category, *Network Computing's* 2006 Well-Connected Award for Remote Office Network Infrastructure, and *eWeek's* 2006 Excellence Award for Networking Infrastructure. Riverbed's award-winning solutions are available worldwide from resellers who are members of the Riverbed Partner Network, from Riverbed OEM partners, or directly from Riverbed.

¹This example is for illustrative purposes only. An examination of the TCP protocol's complex approach to congestion and loss would be beyond the scope of this paper.

²This example is for illustrative purposes only. An examination of all of the variables involved in client-server communications over CIFS and the underlying protocols on which it is dependent would be beyond the scope of this paper.

Riverbed Technology, Inc.
501 Second Street, Suite 410
San Francisco, CA 94107
Tel: (415) 247-8800
www.riverbed.com

Riverbed Technology Ltd. UK
200 Brook Drive, Green Park
Reading RG2 6UB
United Kingdom
Tel: +44 118 949 7002

Riverbed Technology Pte. Ltd.
350 Orchard Road #21-01/03
Shaw House
Singapore 238868
Tel: +65 68328082

Riverbed Technology K.K.
Shibuya Mark City W-22F 1-12-1
Dogenzaka, Shibuya-ku Tokyo
Japan 150-0043
Tel: +81 3 4360 5357

© 2006 Riverbed Technology, Inc. All rights reserved. Riverbed Technology, Riverbed, Steelhead and the Riverbed logo are trademarks or registered trademarks of Riverbed Technology, Inc. Portions of Riverbed's products are protected under Riverbed patents, as well as patents pending. WP-CB063006