Data Storage Anywhere, Any Time

Metro and Wide Area Storage Networking
with Nishan Systems IP Storage Switches

The Need for Distant Storage — Remote Data Replication

Impressive forces of nature continually shape our planet. Over the ages, earthquakes have formed mountain ranges and floods have carved huge canyons that today attract tourists and sports enthusiasts. However, when people and buildings — and data centers — are in their path, these forces can be devastating.

In the early nineties, hurricane Andrew flooded more than three dozen major data centers. Since that time, we have become even more dependent on quick access to stored data. Most IT managers now have remote data replication systems for disaster recovery, but those systems can be costly and cumbersome to manage.

Remote data replication consists of archiving or mirroring. Archiving is a commonly used process in which data is written to portable media, such as optical disks or magnetic tapes. For disaster recovery purposes, the disks or tapes are physically transported to an offsite location and stored. Depending on the regulatory or self-imposed replication requirements, the data may be transported across town or hundreds of miles away.

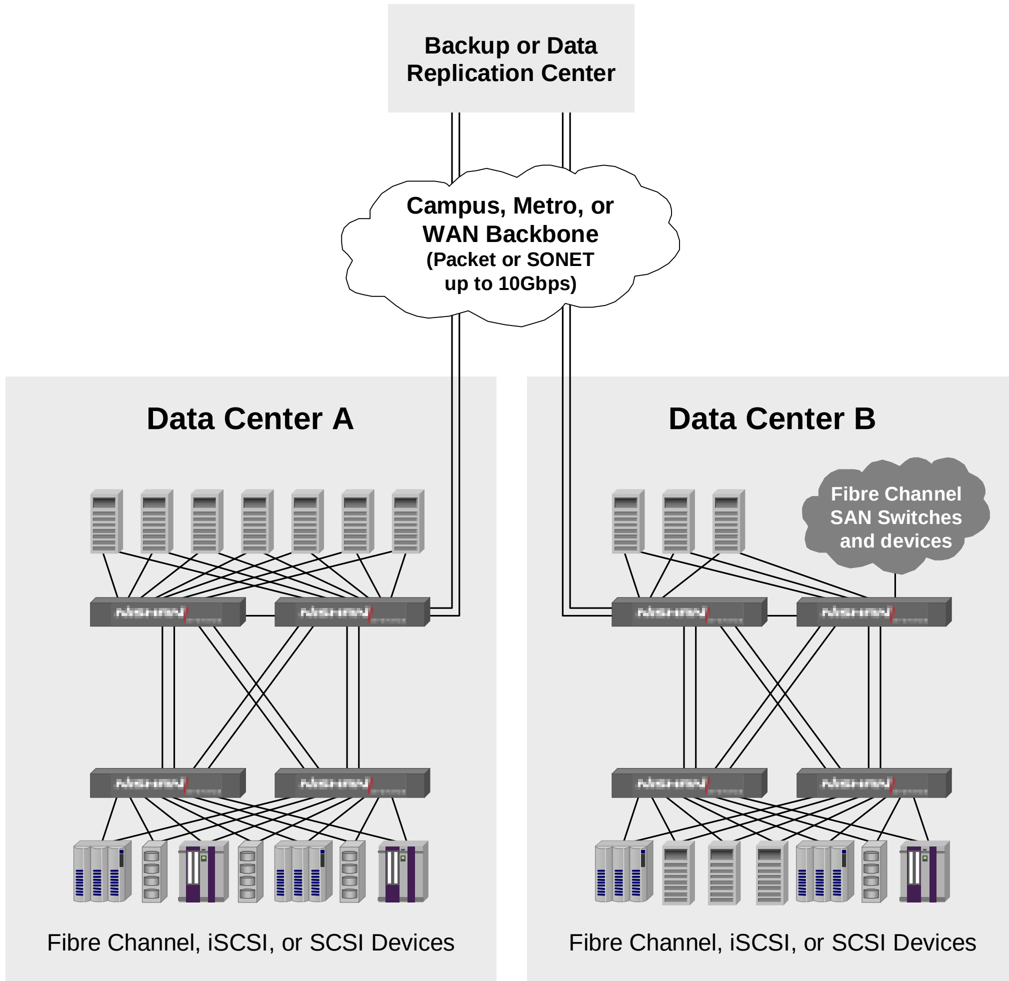
Data mirroring is more bandwidth intensive, as it requires either continuous replication using synchronous mirroring of disk arrays, or batch processing for asynchronous mirroring. Mirroring enables two sites to share the same set of data, so that either can continue to operate on that data in the case of a site outage. Due to the distance and performance limitations of the conventional Fibre Channel switch technology, mirroring has, until now, been used mainly for critical applications and has been limited to relatively short distances.

But a new type of storage networking product has arrived and it is breaking down those distance and performance barriers — Nishan's family of IP Storage switches. These new switches employ the latest high-speed processing technology, and they also use the proven, decades-old protocol, IP — the Internet Protocol. Until the advent of IP Storage switches, IT professionals could not take advantage of the existing worldwide public and private data highways for transferring block-based mass storage data.

The powerful Nishan switches not only enable the construction of large-scale storage networks for data centers, but they also enable storage networks to span campuses, metropolitan areas, and beyond, using existing networks. Like Fibre Channel switches, IP Storage switches can be linked by fixed-bandwidth, dedicated dark fiber or optical multiplexer connections, but IP Storage switches also can take advantage of the more economical and accessible public or private IP-based networks. These networks can be either dedicated or shared, and have a worldwide reach with backbone links up to 10Gbps.

Data centers can back up each other, or a centralized backup data center can be established and shared among the primary data centers, as shown in Figure 1. The Nishan multiprotocol SoIP framework supports iSCSI, iFCP, Fibre Channel, SCSI, and all types of Network-Attached Storage (NAS) interfaces, so virtually all types of storage devices are supported. Conventional Fibre Channel SANs also are supported, as shown.

Figure 1: Remote Data Replication with Nishan IP Storage Switches



Commonly used backup software applications from Legato and VERITAS have been certified for use with Nishan IP Storage switches. Because the Nishan IP Storage networking solutions enable the storage data to be moved faster and further over economical high speed networks, it now is possible to back up more applications, at greater distances from the areas that might be impacted by an earthquake or hurricane.

And during off-peak periods, the storage network can be used to transfer data to archival sites, where it can be copied to optical disks or tapes. Compared to physically transporting the media, storage networking offers a less labor-intensive, more secure, and more reliable approach. Centralizing the media transfer also may be more economical, as many sites can share the backup equipment and staff.

The Technology Behind IP Storage Data Replication

Transmission Latency — Physics and ASICs

Today, virtually all high-speed networks use fiber optic cables to transmit digital voice and data. While the speed of light in a vacuum is about 186,000 miles per second (300,000 Km/s), the index of refraction of the typical single-mode fiber optic cable reduces that to about 100,000 miles per second, or 100 miles per millisecond.

In other words, if the distance between the primary and backup data centers within a metropolitan area is 50 miles, then the round-trip delay will be about a millisecond. Conventional Fibre Channel switches can tolerate a millisecond of delay without much performance degradation, but that may not be sufficiently far from the primary data center to escape the same disaster, defeating the purpose of the backup data center. Or the backup center also may be serving as a centralized data replication site for data centers in other cities that are much farther away.

For those reasons, many IT professionals want to transfer block-mode storage data beyond the metropolitan area, to other cities. Taken to the extreme, that could be hundreds or thousands of miles and tens of milliseconds of round-trip latency. For example, it would take about 50ms for data to be transmitted over fiber optic cable on a 5000-mile round trip across the US or Europe.

Added to that delay, is the latency of the switches, routers, or multiplexers that may be in the transmission path. Today's optical multiplexers add virtually no latency, and Gigabit Ethernet switches also add only tens of microseconds (a small fraction of a millisecond), but some routers can add a few milliseconds of latency to the data path. Modern ASIC-based routers, however, are much faster than their earlier counterparts. For example, in a recent test conducted by the web site, *Light Reading*, the maximum latency of a Juniper M160 router with OC-48c links, was about 170 microseconds. In a typical network, data would pass through two or three of these routers in each direction, adding only about a millisecond to the round trip.

IP Storage Tolerates Greater Latencies

Conventional Fibre Channel switches were designed for use in a single data center, so little consideration was given to long distance transfers of data between Fibre Channel Storage Area Networks (SANs). Consequently, SANs built with Fibre Channel switches must be no more than about 50 miles apart to maintain adequate levels of performance.

In contrast, IP was designed to tolerate the latencies of congested global networks, which can be hundreds of milliseconds. It also withstands packet loss with no loss of data. And with the advent of the iFCP and iSCSI protocols, which adapt block-mode storage data to the standard TCP/IP stack, mass storage data now can be transmitted over hundreds or thousands of miles.

This new IP Storage capability lets IT professionals make their data more accessible in three important ways:

1. Mirroring, rather than just archiving data, for instant access to the backup data
2. Using new high speed IP networks with backbone links up to 10Gbps, for faster transfers
3. Moving the backup data center beyond the metro area, for greater survivability

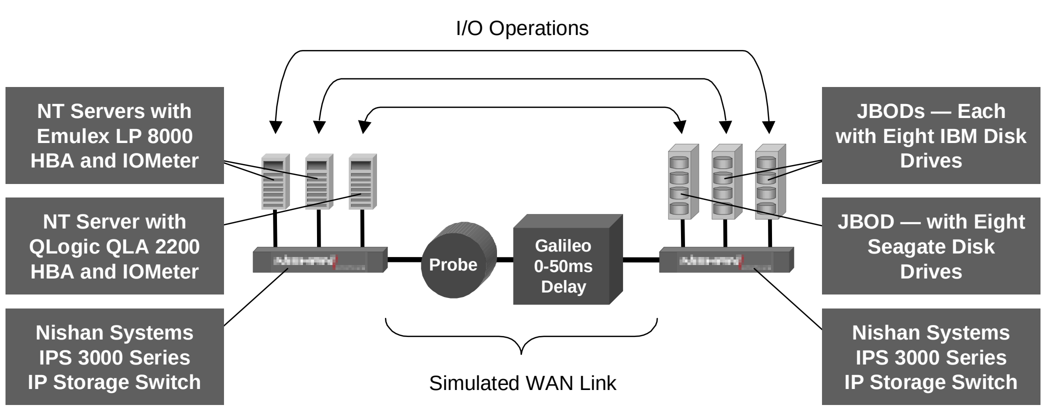
Verification Tests

The IP Storage protocols, iFCP and iSCSI, each have an important role in IP Storage networks. For native IP Storage end systems, such as servers, disk arrays, and tape libraries, iSCSI is used, and for connecting Fibre Channel end systems or Fibre Channel SANs, iFCP is used. Since no wire-speed iSCSI end systems are yet available, the tests were run with Fibre Channel end systems with wire speed iFCP over the TCP/IP connections. Nishan Systems IPS 3000 Series switches were used to convert Fibre Channel to iFCP at wire speed, then back to Fibre Channel at the other end of the TCP/IP connection. Galileo, a Nishan-provided distance simulator was used to inject round-trip delays of up to 50ms, thereby simulating transcontinental data transfers.

The first test simulated a long-distance connection of two data centers over a high-speed link. It was designed to stress the IPS 3000 Series switches by simulating enterprise data centers that have many storage devices connected through a single, long-distance high-speed data link. Each Nishan switch supported a total of 24 separate high-speed TCP connections, and was responsible for multiplexing up to 13,000 I/O's per second into these connections. In this test, three servers in one data center were configured to access data stored on disks in the remote data center via the shared, long-distance link. The servers performed both read and write operations with their remote storage facilities.

The test configuration is shown in Figure 2. Three NT-based Windows servers, each with a Fibre Channel HBA (Host Bus Adapter), were each configured to communicate with eight Fibre Channel targets. A total of 24 disk drives were housed in JBODs (Just a Bunch of Drives) that were connected to an IPS 3000 Series switch. The HBAs were connected directly to another IPS 3000, and the two switches were connected through the Galileo long-distance link simulator, which injected varying amounts of delay into the traffic streams.

Figure 2: Long Distance IP Storage Test Configuration

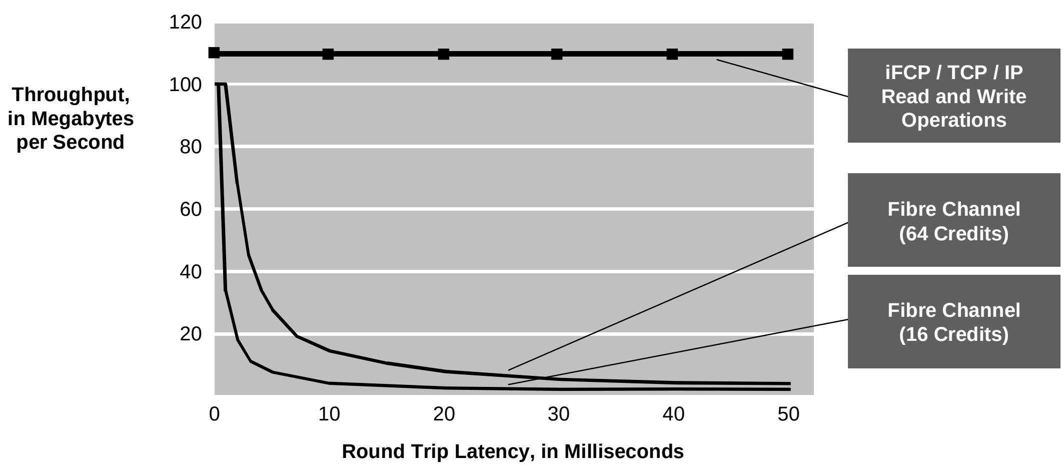


Emulex and QLogic HBAs were used in the servers, along with Seagate and IBM disk drives. The network was configured to allow each HBA access to a total of eight drives housed in the JBOD enclosures, as shown in the diagram. Using Intel's Iometer simulation software, sequential read and write operations were tested and measured separately.

The Galileo distance simulator was used to inject between 0 and 50 milliseconds of round-trip latency in the data path between the Nishan switches. Performance was measured for each increment of latency.

The results of this test are summarized in Figure 3. TCP/IP was demonstrated to support wire-speed gigabit-per-second transfer rates for storage data, with latencies that would be expected in links spanning the continent. Fixed bandwidth DWDM and SONET networks or low-latency IP and Gigabit Ethernet networks could be used to carry the data.

Figure 3: Long Distance IP Storage Performance Results



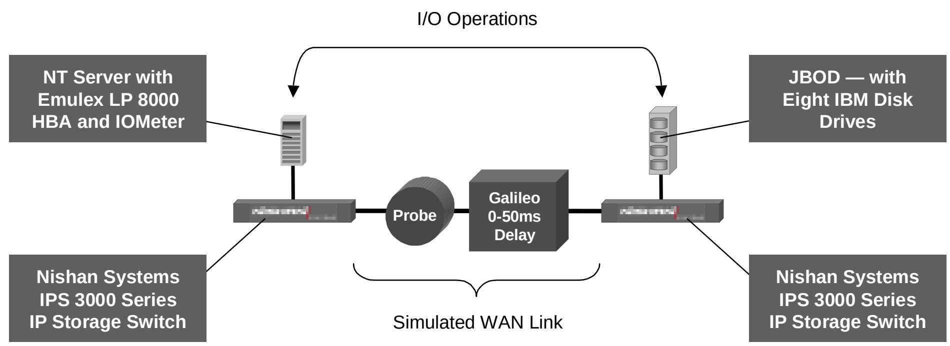
For comparison, Figure 3 also includes graphs of the theoretical maximum performance of Fibre Channel switches, given the corresponding latency of the long distance link. It was assumed that the Fibre Channel frames were fully loaded — a payload of 2112 bytes — and that there were no link errors, so this is the best performance possible. With a round-trip delay of 10ms, for example, the maximum performance is reduced to only 13.5MBps for 64-credit switches (2112B x 64 / 0.010s), and 3.4MBps for 16-credit switches. Compared to TCP/IP, Fibre Channel switches perform very poorly over long distances, as shown. Whether it allows 16 or 64 credits, Fibre Channel's credit-based flow control mechanism severely constrains throughput over distances that introduce more than about one millisecond of latency.

In addition to operating at wire speed, the Nishan switches were able to balance the traffic load evenly among all pairs of Fibre Channel devices sharing the long distance link. Although only a total of 1Gbps was available over the long distance link, the Nishan switches allocated this bandwidth evenly among the Fibre Channel HBAs, with each HBA receiving approximately 36MB/s of throughput.

Performance Testing with a Single HBA

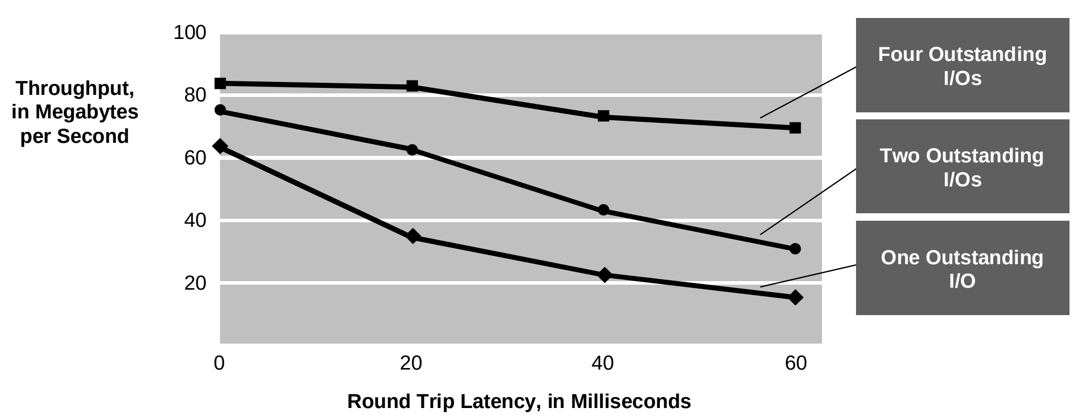
The first test had multiple devices sharing a long distance data link. A second test analyzed the behavior of an individual HBA accessing remote storage devices. The test topology is shown in Figure 4.

Figure 4: Long Distance IP Storage Test Configuration with a Single HBA



To further analyze the effects of latency, performance was measured with the HBA configured to issue one, two, and four I/O operations. Each setting corresponds to the maximum number of I/O's that could be outstanding at any given moment. As the number of I/O operations is increased, more storage data can be "in-the-pipe" at any moment, thereby boosting performance. The results from this second test are shown in Figure 5.

Figure 5: Long Distance IP Storage Performance Results with a Single HBA



The results in Figure 5 clearly indicate that a greater number of outstanding I/O operations leads to improved performance over longer distances. Each pending I/O increases the utilization of the long distance link. Further tests revealed that extending the number of I/O operations beyond four yielded limited additional performance gains, as the server generating the data traffic did not generate a larger number of I/O requests when the number of outstanding I/Os was increased.

More powerful systems that have a greater capacity to issue multiple I/O operations should produce additional performance gains. Faster, high-end platforms should be able to process a greater number of I/O operations simultaneously, thus individually achieving greater throughputs over extended distances.

Final Points

A new type of storage networking product has arrived and it is breaking down distance and performance barriers — Nishan's family of IP Storage switches. These new switches employ the latest high-speed processing technology, but they also use the proven, decades-old protocol, IP. Until the advent of IP Storage switches, IT professionals could not take advantage of the existing worldwide public and private data highways for transferring block-based mass storage data.

The powerful Nishan switches not only enable the construction of large-scale storage networks for data centers, but they also enable storage networks to span campuses, metropolitan areas, and beyond, using existing networks. Like Fibre Channel switches, IP Storage switches can be linked by fixed-bandwidth, dedicated dark fiber or optical multiplexer connections, but IP Storage switches also can take advantage of the more economical and accessible public or private IP-based networks. These networks can be either dedicated or shared, and have a worldwide reach with backbone links up to 10Gbps.

Storage networks built with Nishan IP Storage switches deliver large-scale, wire-speed, mass storage connections that can span great distances. And because the products rely on standard, proven IP and Ethernet technology, IT professionals can immediately leverage their existing knowledge and diagnostics equipment to cope with the challenges of large-scale storage networking.