Evaluation of Multi-protocol Storage Latency in a Multi-switch Environment

Evaluation report prepared under contract with Cisco Systems, Inc.

Executive Summary

In today's complex storage networking environments and with the advent of converged, or unified networking, it is possible to run multiple storage protocols in the same network. Current datacenter networks deploy a variety of network topologies, from simple one-switch networks found in small and medium business, to complex aggregate/core and access multi-switch topologies found in large datacenters.

There are often questions regarding latencies for storage systems when deployed in various network topologies. Cisco Systems® commissioned Demartek to evaluate the effects of various networking topologies in a server and storage environment. Specifically, three storage protocols were tested in the same network simultaneously and with different network topologies. These protocols are iSCSI, FCoE and native Fibre Channel.

We found that the average latency was consistent for each protocol, regardless of the network topology. From the simple network topology found in small and medium-sized businesses to the full core and edge networks found in large datacenters, the various network topologies were very similar in their latency measurements for each storage traffic type.

1 – Storage Protocol Background

Storage area network (SAN) architecture provides a way to use block access methods over a network such as Ethernet or Fibre Channel to provide storage for host computers. The storage in a SAN is not owned by one server but is accessible by all of the servers on the storage network. This SAN storage can be carved into logical storage pools or volumes that can be assigned to particular host servers. These logical volumes are independent of the geometries or components of the storage hardware and are independent of the network topology used to connect the storage to the host servers. The storage appears to host servers and applications in the same way that local storage appears, but because SAN storage uses a network, storage can be a long distance away from the host servers.

SAN architectures use block Small Computer System Interface (SCSI) protocol for sending and receiving storage data over their respective networks. Fibre Channel (FC) SANs implement the SCSI protocol within the FC frames on a native Fibre Channel network. Internet SCSI (iSCSI) SANs implement the same SCSI protocol within TCP/IP packets in an Ethernet network. Fibre Channel over Ethernet (FCoE) is a newer interface that encapsulates the Fibre Channel protocol within Layer 2 Ethernet packets using a relatively new technology called Data Center Bridging (DCB). DCB is a set of IEEE 802.1Q enhancements to traditional Ethernet that allows the LAN behavior of dropping packets upon congestion to co-exist with the SAN requirement of no loss of frames, and is currently implemented with some 10GbE infrastructure, including the switches tested for this report. Because each of these technologies allow applications to access storage using the same underlying SCSI command protocol, it is possible to use all of these technologies in the same enterprise, or to move from one to the other. Generally speaking, applications running on a host server cannot tell the difference between Fibre Channel SAN storage, FCoE SAN storage, and iSCSI SAN storage.

The networking equipment required to carry these types of storage traffic includes traditional Ethernet switches for iSCSI traffic, DCB Ethernet switches that support FCoE and native Fibre Channel switches. Some of the switch technologies, including the switches tested for this report, can support all three types of traffic from within the same switch.

2 – Network Topologies

Modern datacenters contain a variety of compute, storage and networking elements and deploy these elements in various topologies to fit the needs of the business. In this report, we examine four topologies that are representative of many of today's datacenters, with an emphasis on the server and storage elements. In networking parlance, we are focused in this report on the "East to West" traffic that flows between servers and storage.

The four topologies tested for this report are:

- SMB topology (single-switch)
- Multi-hop core/edge with FC, FCoE and iSCSI
- Multi-hop converged core/edge
- Multi-hop core with FC, FCoE and iSCSI

Multiple Protocol Capabilities

For those environments where multiple storage protocols are used, such as Fibre Channel, FCoE and iSCSI, these network topologies show the flexibility available to support these storage protocols. The topologies tested for this report include converged and non-converged traffic flow. The Nexus switches used in these topologies support all three protocols, FC, FCoE and iSCSI. The MDS switch used in these topologies supports FC and FCoE protocols.

SMB Topology

The SMB topology represents a typical network topology found in a small or medium-sized business that depends on a single network switch for all of its networking needs. In this environment, there is only one network layer (see "access" description below), and all the servers and storage units are connected to a single switch. Because the Cisco Nexus 5500UP supports universal ports, the ports can be configured as Ethernet ports or Fibre Channel ports, and one switch can satisfy both types of networks.

Multi-hop Core/Edge

These two topologies are similar to each other in that there is a Cisco Nexus 5548 acting as an access layer ("edge" switch) for the servers and the larger switches are acting as the aggregate or core of the network. The larger switches, the Cisco Nexus 7010 and Cisco MDS 9506, are connected to the storage systems. In the multi-hop core/edge with FC, FCoE and iSCSI topology, the Nexus 7010 handles the Ethernet traffic and the MDS 9506 handles the Fibre Channel traffic within the core and coming from the access layer. In the converged topology, the Nexus 7010 handles all the traffic as a core switch/director and passes the FC and FCoE traffic to the MDS 9506 to be sent on to the storage systems.

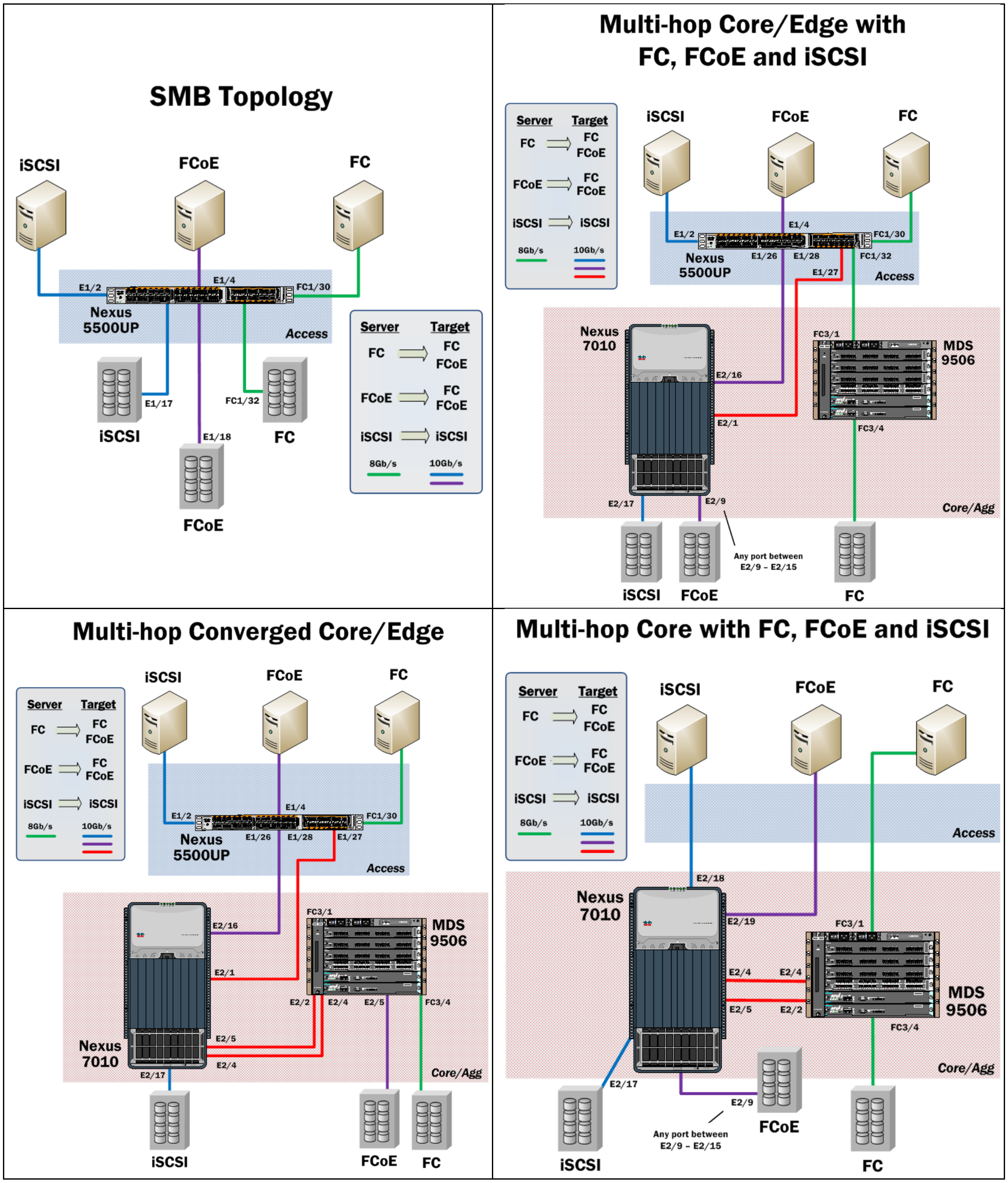
Multi-hop Core

In the multi-hop core with FC, FCoE and iSCSI topology, the access layer, or edge switch (Cisco Nexus 5548) has been removed and the aggregate or core switches handle all the traffic coming from the servers.

Access and Aggregate

The terms "access" and "aggregate" are used to describe the relative positions of the switches in a two-layer networking topology. Access switches are typically logically close to the servers and/or storage and are sometimes known as edge switches. The aggregate switches are in the core of the network and provide connectivity to other access switches in different parts of the network. The switches tested for this report support all of the storage protocols, which allows for some flexibility in designing and deploying the network topologies.

Topology Diagrams



3 – Performance Tests

In these tests, we placed workloads from different servers onto a storage system using different storage protocols and measured the latencies for each of the networking topologies. The workloads consisted of Windows servers running SQLIO for the iSCSI, FC and FCoE workloads. SQLIO is a utility application provided by Microsoft that sends database I/O workloads to a storage system for the purpose of stress testing a storage system.

All of the servers were connected to a NetApp FAS3270 or a NetApp FAS3240 storage system via the switches shown in the topology diagrams. The NetApp FAS3270 and FAS3240 storage systems support 10 Gbps Ethernet including iSCSI and FCoE and supports native 8 Gbps Fibre Channel host interfaces.

The latencies shown in the table below are from the perspective of the SQLIO application running on the host servers. The average latency was consistent for each protocol, regardless of the network topology. Generally speaking, the latency for 10GbE switches is measured in microseconds, while storage latencies are generally measured in milliseconds.

SQLIO Paramters

```
sqlio.exe -kR -s3600 -fsequential -o1 -b1024 -LS -Fparam.txt
sqlio v1.5.SG
using system counter for latency timings, 1562587 counts per second
parameter file used: param.txt
   file M:\sqlio_test_001.dat with 16 threads (0-15) using mask 0x0 (0)
16 threads reading for 3600 secs from file M:\sqlio_test_001.dat
   using 1024KB sequential IOs
   enabling multiple I/Os per thread with 1 outstanding
using specified size: 2048 MB for file: M:\sqlio_test_001.dat
```

Latency Results

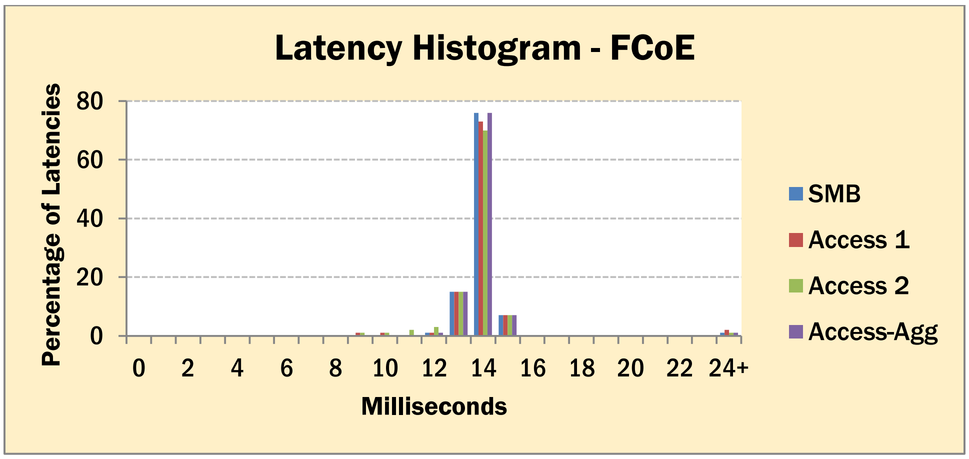
SQLIO records several metrics in its raw output, including minimum, average and maximum latency. It also provides a latency histogram showing the percentage of latencies within each one millisecond range.

The latency results for each network topology are shown below, grouped by protocol. Some of the test results had the occasional "outlier" data point for latency greater than 24 milliseconds, but generally speaking, the latencies were fairly consistent. For the FCoE tests, 95% of the latencies were within a 3 millisecond range. This latency is affected somewhat by the network but is affected primarily by the storage system.

For this network topology and set of equipment, the FCoE traffic had the lowest average latency, followed by the Fibre Channel traffic and then the iSCSI traffic. The Fibre Channel traffic had the lowest minimum latency, and the iSCSI traffic had the highest minimum latency.

FCoE Results

Topology	IOPS	MBPS	Minimum Latency (ms)	Average Latency (ms)	Maximum Latency (ms)
SMB Topology	1108.75	1108.75	4	13	45
Access Topology 1	1104.72	1104.72	4	13	9952
Access Topology 2	1105.50	1105.50	4	13	59990
Access – Aggregate Topology	1108.81	1108.81	4	13	45



FC Results

Topology	IOPS	MBPS	Minimum Latency (ms)	Average Latency (ms)	Maximum Latency (ms)
SMB Topology	689.82	689.82	1	22	85
Access Topology 1	689.02	689.02	1	22	78
Access Topology 2	702.91	702.91	1	22	205
Access – Aggregate Topology	689.02	689.02	1	22	78

iSCSI Results

Topology	IOPS	MBPS	Minimum Latency (ms)	Average Latency (ms)	Maximum Latency (ms)
SMB Topology	593.69	593.69	6	26	47
Access Topology 1	592.94	592.94	6	26	41
Access Topology 2	592.62	592.62	6	26	82
Access – Aggregate Topology	592.99	592.99	6	26	39

Conclusion

We found that the various network topologies had little effect on latencies between the servers and the storage. The latencies were more influenced by the traffic type and the storage system performance than they were by the particular network topology.

The four network topologies were consistent in their latency for each traffic type. We noted that the average latency was the lowest for FCoE with these network topologies and storage configurations.

With the topologies and switches that were tested, we found complete flexibility in the ability to design the network to handle native Fibre Channel, FCoE and iSCSI traffic. We were able to handle all three storage protocols in "converged" or "non-converged" modes. We were able to run in a single-switch environment, core/edge environments and in multi-hop configurations.

I.T. administrators should take comfort knowing that there is good opportunity for flexibility in network topology design, especially for servers and storage, while maintaining consistent latency.

Appendix – Test Environment

Host Server for FCoE

- Intel Xeon E5240, 3.00 GHz, 2 cores, 2 logical processors
- 20 GB RAM
- Internal 15K RPM disk drives
- Emulex OCe10102 10GbE CNA
- Windows Server 2008 R2

Host Server for iSCSI

- 2x Intel Xeon E5345, 2.33 GHz, 8 total cores, 8 logical processors
- 48 GB RAM
- Internal SAS disk array, 15K RPM disk drives
- Emulex OCe11102 10GbE CNA
- Windows Server 2008 R2

Host Server for FC

- 2x Intel Xeon E5345, 2.33 GHz, 8 total cores, 8 logical processors
- 48 GB RAM
- Internal SAS disk array, 15K RPM disk drives
- Emulex LPe11002 8Gb FC HBA
- Windows Server 2008 R2

Storage System for FCoE

- NetApp FAS3270
- 48x 15K RPM disk drives
- ONTAP 8.0.1 RC

Storage System for iSCSI and FC

- NetApp FAS3240
- 24x 10K RPM disk drives
- ONTAP 8.1 RC

Switches

- Cisco Nexus 5548UP, operating system: 5.1(3)N2(1)
- Cisco Nexus 7010, operating system: 6.1(1)
- Cisco MDS 9506, operating system 5.2(2a)

The original version of this report is available at
http://www.demartek.com/Demartek_Cisco_Multi-protocol_Multi-topology_Latency_Evaluation_2011-07.html