Cisco Nexus 5020 Switch Performance in Market-Data and Back-Office Data Delivery Environments

Overview

In the capital markets industry today, significant growth in the levels of data traffic for pre- and post-trade information processing has become the norm. Addressing this continued and accelerating growth is a major challenge and concern for all involved. Those involved are looking at many technologies to stay on top of this accelerating growth: higher-speed processors, more memory, larger-scale environments, and more recently, high-bandwidth and low-latency 10 Gigabit Ethernet switching technologies.

The move to these higher-bandwidth technologies is not without its own challenges. In-host protocol processing continues to be a significant challenge and an impediment to increased performance as it accounts for anywhere between 70 and 90 percent of the end-to-end application latency. Fortunately this problem is being addressed through the offerings of offload technologies such as TCP/IP offload engines, user-space communications libraries, and Remote Direct Memory Access (RDMA) capabilities in a number of 10 Gigabit Ethernet host adapters.

As the host-side performance bottlenecks are addressed, the capabilities and effects of the underlying network fabric become far more evident. Forwarding methodologies, buffering architectures, ingress and egress queuing processes, transceiver-to-transceiver throughput and latencies levels, and jitter become significantly more apparent.

With the recent announcement of the Cisco Nexus™ 5000 Series 10 Gigabit Ethernet switches, a new generation of reliable, lossless, and deterministic performance capabilities can now provide the next level of performance in highly demanding market-data environments.

This document describes the test results from a proof-of-concept evaluation of the Cisco® Nexus 5020 Switch at a world-leading capital market company. The goals of the testing were to detect:

- Increases in throughput and reductions in data load and unload times or data latency
- Increases in the rate of update messages per second in a multicast market-data environment
- Throughput, latency, and loss levels across the fabric
- Reductions in data load and unload times

The findings were as follows:

- Data delivery time for single-threaded data transfers of a 5-GB data set between a single host and a NetApp OnTap GX system decreased by 88 percent. Times for multithreaded and multipathed data transfers increased by a factor of 10.
- Wombat Latency Busters Messaging (LBM) multicast-based market data reached 1 million update messages per second at the subscriber for a two-publisher-to-one-subscriber test scenario in comparison to fewer than 200,000 messages per second over Gigabit Ethernet.

In-host microbenchmarks for 10 Gigabit Ethernet were as follows:

- Line rate was achieved at all packet sizes for both unidirectional and bidirectional traffic flows.
- Transceiver-to-transceiver latencies maintained a constant rate of 3.2 microseconds for all packet sizes tested.

Background

In market-data environments, application performance is one of the most significant factors in ensuring maximum opportunities for return in response to new information. Variables that affect performance, either negatively or positively, include node configuration, interconnects, switching fabric composition and associated configuration, protocol offload technologies, and application communication characteristics.

Latency is often discussed as the primary performance metric in determining whether application run time is optimal. In today's competitive marketplace, as little as 1 millisecond less latency can result in significant profit and strong competitive advantage.

The effectiveness with which the network communication components (switches, adapters, host drivers, and offload technologies) interoperate continues to be assessed to determine potential areas for performance improvement. Gigabit Ethernet has been the standard implementation, but with the rapid growth in message rates and the problem of microburst traffic due to increases in spot trading increases, Gigabit Ethernet has reached its limits in many cases.

The desire to achieve competitive advantage and the increasing volume of market-data traffic are making the increased throughput levels and low latencies of 10 Gigabit Ethernet switches increasingly necessary. The higher throughput rates of 10 Gigabit Ethernet easily absorb the traffic increase associated with traffic bursts and provide the capability for growth at lower switching latencies and with less jitter.

The Cisco Nexus 5000 Series of 10 Gigabit Ethernet, low-latency, Layer 2 cut-through switches provide deterministic and lossless delivery at line rate, which easily meets the high demands of the market-data application environment. The Cisco Nexus 5000 Series Switches supports up to 52 ports of nonblocking, line rate 10 Gigabit Ethernet and Fibre Channel over Ethernet (FCoE) per port. These characteristics are ideal for market-data applications, and can also support high levels of I/O traffic for back-office and post-trading environments.

Description of Tests

Methodology

The tests employed followed the procedures set forth by RFC 2544 and 2899 for device benchmarking in the Ixia test suites, Wombat LBM Options Price Reporting Authority (OPRA) and Middleware Agnostic Messaging API (MAMA) benchmarks, and TCP and User Datagram Protocol (UDP) synthetic benchmarks.

Losslessness and Throughput Tests

The tests measured the 10 Gigabit Ethernet throughput levels of the Cisco Nexus 5020 cut-through switch. Multiple packet sizes were tested for throughput and loss. An Ixia traffic generation device generated the traffic. Traffic was both unidirectional and bidirectional for unicast and multicast.

The objective of the tests was to confirm line rate throughput with zero loss at all frame sizes.

Losslessness and Throughput Test Results

Unicast unidirectional and bidirectional traffic achieved line rate for all packet sizes, with zero loss for packet sizes of up 9216-byte jumbo frames.

Multicast tests were run using datagram sizes of 64 to 1518 bytes using the Protocol Independent Multicast Sparse Mode (PIM-SM) tests provided by the Ixia test suite. Multicast performance at all datagram sizes achieved zero loss for all multicast groups.

Figure 1. Packet Loss Test for Unicast Unidirectional 64- to 1528-Byte Packets

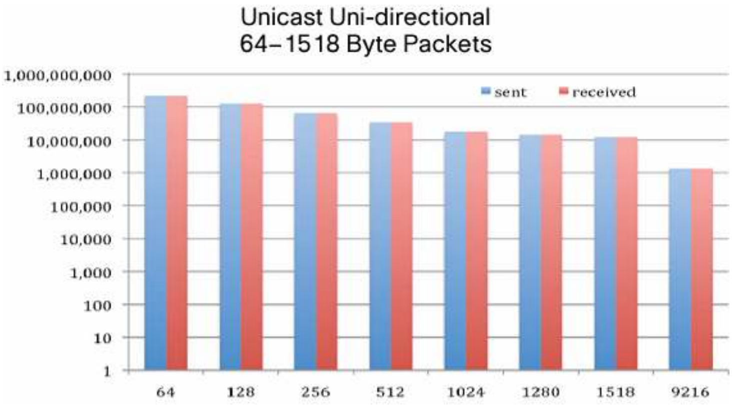
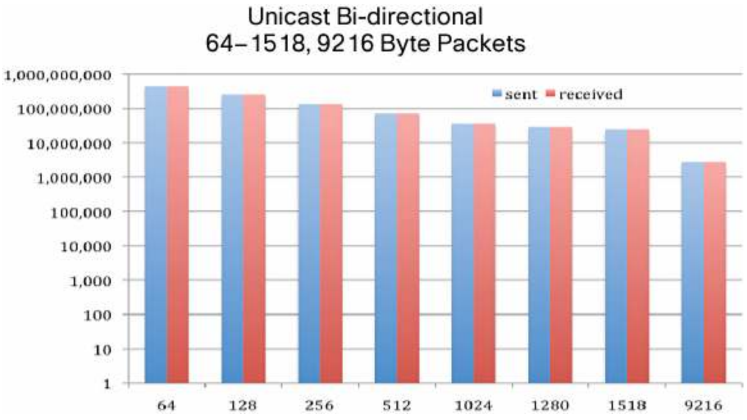


Figure 2. Packet Loss test for Unicast Bidirectional 64- to 1518- and 9216-Byte Packets



Latency Tests

Two sets of latency tests were performed. The first set of tests was based on the latency tests found in the Ixia test suites. For these tests, latency of the Ixia tester was determined through back-to-back testing using two test ports that were directly connected. After the test device latency was determined, the latency of the Cisco Nexus 5020 was tested.

The second test was based on TCP synthetic benchmarks to reflect real-world performance for TCP-based applications. These tests used message sizes of from 1 byte to 16 kilobytes with the host maximum transmission unit (MTU) set at 1500 and 9000 to support 1518- and 9216-byte packets. Each message size was transmitted 1000 times.

This test also provided mixed packet sizes, with the larger messages sizes distributed over n payloads: for example, a 16,384-byte message would be divided into ten 1500-byte payloads and one 1384-byte payload. To test the effect of the scheduler, buffering mechanisms, and queuing and congestion management services, background traffic of from 10 to 40 percent was pushed though the switch.

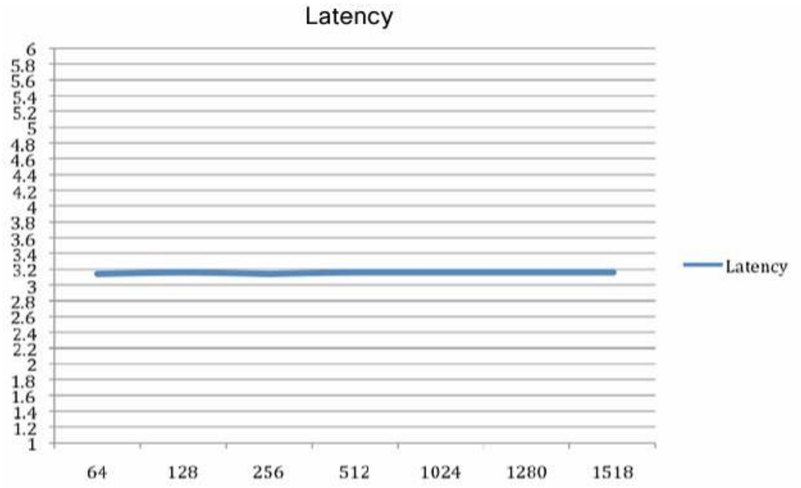
Traffic for this test was generated by two Dell PowerEdge 2950 rack servers running Red Hat Enterprise Linux Version 4 with Update 4 (2.6.9-42). An Intel Oplin 10 Gigabit Ethernet Small Form-Factor Pluggable Plus (SFP+) adapter provided the network connectivity from the hosts to the Cisco Nexus 5020 over fiber. The SFP+ SR optical transceiver latency was 1 microsecond and was included in the total switch latency.

As in the Ixia latency tests, the host latency was calculated using back-to-back connections. Upon completion of the host latency test, the same tests were run with the hosts connected to the Cisco Nexus 5020. The transceiver-to-transceiver latencies of the switch were calculated by subtracting the back-to-back latencies from the total end-to-end latencies.

Latency Test Results

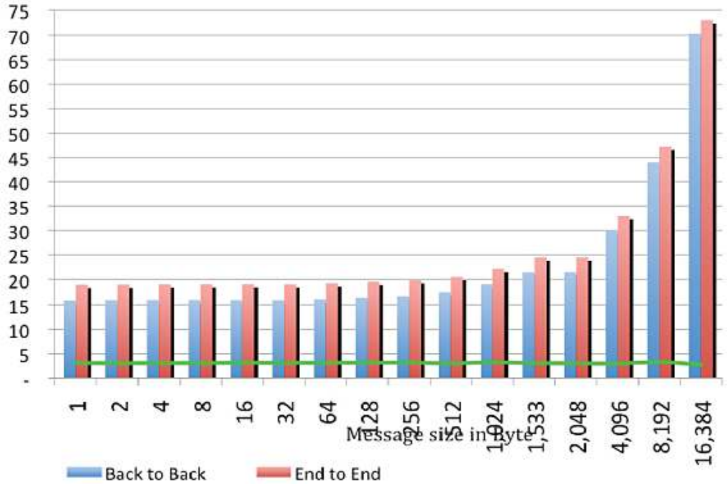
For each packet size tested, from 64 to 1518 bytes, latency did not exceed 3.2 microseconds when using the Ixia test suite (Figure 3).

Figure 3. Switch Latency by Message Size



TCP latencies for all messages sizes did not exceed 3.2 microseconds for the TCP-based tests at any background traffic level (from no background traffic up to 40 percent background traffic). There was zero loss during these tests. Figure 4 shows the back-to-back latencies (blue bar), end-to-end latencies (red bar), and switch latency (green line).

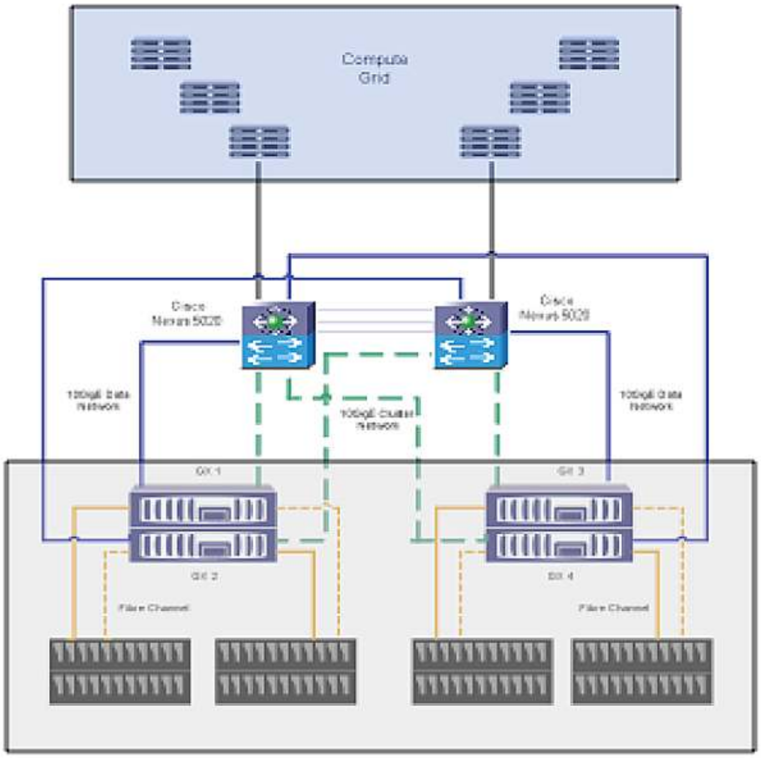
Figure 4. Latency Curve by Message Size



Data Delivery Tests

The design for the data delivery and the Wombat LBM application tests required a low-latency, multicast, 10 Gigabit Ethernet switching platform provided by a pair of Cisco Nexus 5020 Switches. The Cisco Nexus 5020 Switches provided two different networks: a 10 Gigabit Ethernet intercluster communication network that provides connectivity between the four NetApp filer nodes, and a 10 Gigabit Ethernet network for data sent to servers (Figure 5). Through the use of VLANs, the networks can be virtualized on same hardware.

Figure 5. Cluster Connectivity



Data Delivery Test Results

A 5-GigaByte data set was transferred from the NetApp filers connected by 10 Gigabit Ethernet to multiple computing nodes connected by Gigabit Ethernet. The baseline performance for noncached read operations was 5 minutes total transfer time from NetApp filers connected by Gigabit Ethernet to computing nodes connected by Gigabit Ethernet. The data delivery time for the same noncached read operations with the NetApp filer connected to the Cisco Nexus 5020 by 10 Gigabit Ethernet per filer head completed in 36 seconds, or 8.3 times faster.

Multicast Wombat LBM Tests

The tests required a multicast-capable network with servers connected as trunk ports for multi-VLAN access. Servers were required to process one million messages per second with 100 multicast groups

Multicast Wombat LBM Test Results

Two publishers were used to send multicast data to a single subscriber. Test results with data feeds using 500-byte packets achieved one million messages per second at the subscriber. Levels above one million messages were tested but were unachievable due to both hardware limitations and the lack of stateful offload within the Myricom Myri10GE adapter at the subscriber. The Cisco Nexus 5020 easily handled the multicast levels with zero datagram loss.

Conclusion

Much has been done to improve application performance with higher speed systems, but as message rates and data sizes have increased, Gigabit Ethernet has become an inhibitor to achieving optimal performance. For market data, Gigabit Ethernet impedes performance during increased pricing update spikes due to the bandwidth limitations, switching and buffering architectures, and host side ability to drain data as fast as it is received.

For data delivery, time to deliver data to the processors is becoming more and more critical, and in many application environments it is the largest inhibitor to reducing wall clock or run times of the application. The options to change this are to either to linearly scale the data delivery mechanisms to increase concurrent session levels, but this does not necessarily reduce the time to deliver data, as the host connectivity using Gigabit Ethernet will still be a limiting factor.

10 Gigabit Ethernet will provide a delivery mechanism to benefit multicast and unicast traffic with a higher level of available bandwidth, which eliminates the hard performance walls of Gigabit Ethernet and through a reduction in latency for both message passing and data delivery.

The Nexus 5000 series of switches, with up to 52 ports of non-blocking, line rate 10 Gigabit Ethernet, is the optimal choice for these environments as it provides a high performance, low latency switching technology for host and storage connectivity. In addition to the low latency performance comes deterministic performance for port-to-port latency levels of no more than 3.2 microseconds and message-to-message (packet-to-packet) latency variations, or jitter, of less than 100 nanoseconds.

For More Information

http://www.cisco.com/go/nexus5000

Appendix

Cisco Nexus 5020 Switch Configuration

Figure 6. Data load unload testing topology

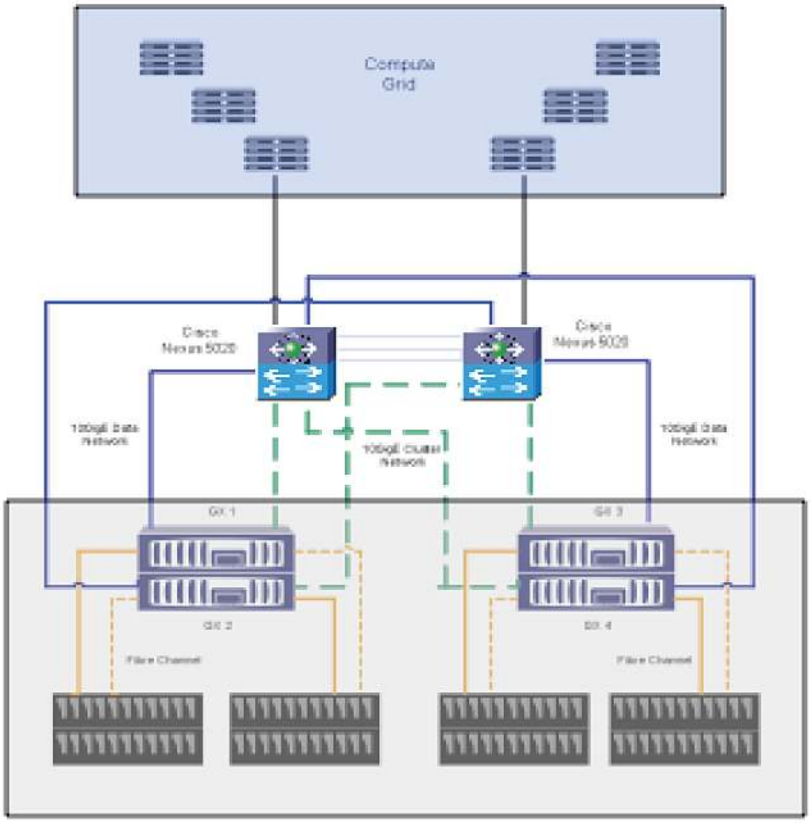
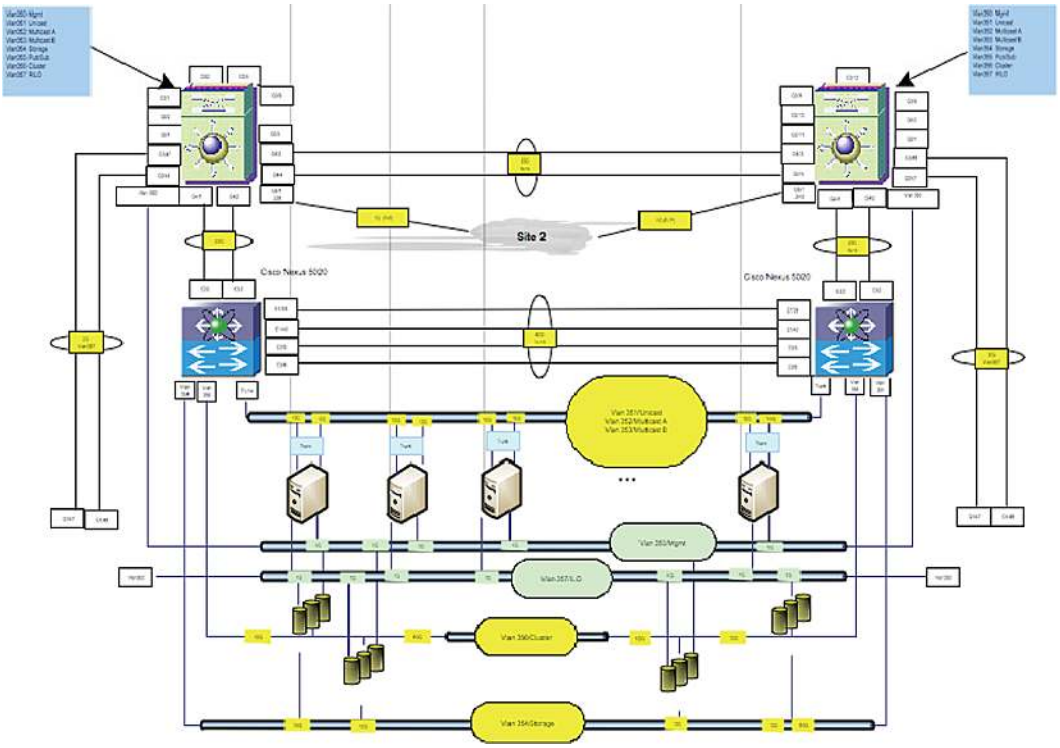


Figure 7. Market Data and Multicast Test Topology



10 Gigabit Ethernet Servers

Data delivery and Wombat LBM multicast test:

- DL 580G5: 16-core 4RU with 32 GB of memory (2 GB per core)
- DL 380G5: 8-core 2RU with 16 GB of memory (2 GB per core)
- 64-bit Linux: RH4 (latest update)

TCP latency tests:

- Dell 2950: 4-core 2RU with 8 GB of memory (2 GB per core)
- RHEL4u4: 2.6.9-42 kernel

Basic Test Information

Ixia Information:

- IxOS Version 5.10.350.24 EASP2
- IxNetwork 5.30.40 EASP1Patch1
- IxAutomate 6.30.32.13 EASP2
- IxLoad 3.40.49.49 EA

Cisco Catalyst® 6509 Switch:

- 2 Supervisor Engine 720 devices: 1 per Cisco Catalyst 6509 Switch
- 2 Cisco 6704 10 Gigabit Ethernet line cards
- Cisco Catalyst 6509-E with Supervisor Engine 720
- Cisco IOS® Software Version 12.1(26)SXF6

Wombat Womark Test Environment:

- 4 HP DL-380 G5–RH4 REL5
- Wombat/29West/Womark test scripts
- Mircom 10G-PCIE-8A-R+E
- Intel 10 Gigabit SFP (XFP) based 10 Gigabit Ethernet network interface card (NIC)

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-492751-01 05/09