Data Center 40GE Switch Study

Cisco Nexus 9516
DR 140127E

Contents

1 Executive Summary

Miercom is pleased to publish these results, of recent testing we conducted on one of the highest-density and highest-throughput data center switches we have ever seen – the powerful 16-slot Cisco Nexus 9516.

Conducted in January 2014, Miercom engineers were invited to give the impressive new switch a work-out. It was configured to scale – that is, packed with 16 of the vendor's latest and highest density line cards – the *36p 40G QSFP+ ACI spine line card*, each offering 36 x 40GE ports. The system we tested, Nexus 9516 with 16 line cards, supported a staggering 576 x 40GE (QSFP+ fiber) ports.

The system incorporates a novel chassis design. Rather than classical multi-slot switches, where a backplane or mid-plane typically connects line cards and fabric modules, Nexus 9500 is the first in the industry to employ a mid-plane free design. Its fabric modules are directly attached to all the line cards through connecting pins, eliminating the airflow obstruction caused by chassis mid-plane. This midplane-free design greatly increases cooling efficiency and easies upgrade for higher speed in the future. The Nexus 9516 switch supports up to six fabric modules, each yielding 10.24 Tbps of bidirectional connectivity throughput.

In addition to validating the vendor's published throughput figures, our testing examined aspects of the switch as far as replicating packets for multicast traffic handling, and the delay, or latency, that the switch applies to packets traversing the system. In each test scenario we were impressed.

Key findings for the Cisco Systems Nexus 9516:

- The switch delivered full line-rate throughput on every 40GB port for all packet sizes, with zero packet loss, based on full-mesh, 576-port x 40GE, high-density throughput testing. Using IP packets and hundreds of flows with different IP source and destination addresses, each 36p 40G line card can transmit 1.44 Tbps of traffic.
- Traffic traversing the switch incurs consistent, low latency, even under heavy load, regardless of packet size, based on the fully populated configuration tested.
- The switch applies very low jitter – latency variance – regardless of frame size, again based on our fully populated test configuration.

All of the performance testing was conducted in accordance with widely accepted IETF standards, including Requests for Comments (RFCs) 2544, 2889, 3393 and 3918.

Miercom has independently observed the performance of the Cisco Systems Nexus 9516 and awards the ***Miercom Performance Verified Certification*** in recognition of the product's proven performance in the Miercom ongoing Data Center Class 40GE Switch Testing Study.

2 Introduction

The powerful, impressive switch tested for this report, the Cisco Nexus 9516, is the latest entry in the Cisco Nexus 9500 Series. It doubles the 40GE port counts of the 8-slot Nexus 9508 from 288 to 576. A compact 4-slot version is also reportedly in the works.

Like the other models in the family, the 9516 is modular and employs the same innovative "mid-plane-free" architecture. A chassis mid-plane or backplane – essentially a high-speed bus – has long been the traditional means in modular switches for connecting line cards into and across the switching fabric. In the Nexus 9500 Series, line cards and fabric modules directly attach to each other via connecting pins and a precise alignment mechanism. Each of up to six fabric modules connects directly to all line cards.

To date, this is the highest density and highest throughput switch we have tested and confirmed "wire speed" performance on all ports concurrently. In other words, the internal architecture is capable of switching as much data as 576 ports of 40GE can deliver to the system. That is a bit more than 46 Tbps of bidirectional traffic with all 40GE ports loaded.

The test traffic was not all random bits. It was legitimate IP packets, and the Nexus 9516 had to examine each, check its forwarding table, and route each one appropriately. The testing is discussed more in the following sections.

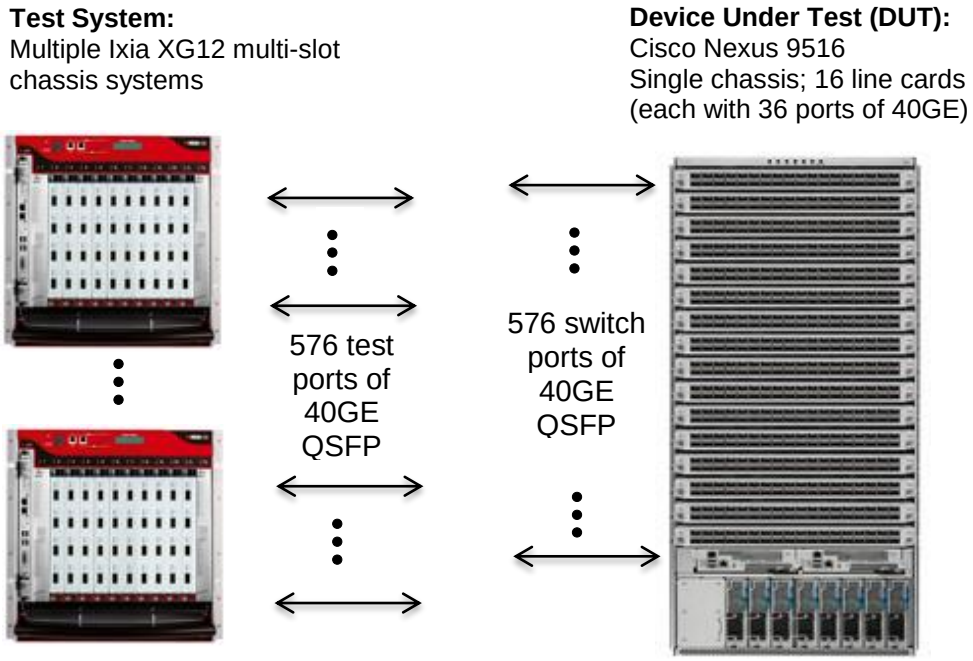
We created this load with a battery of Ixia test systems, which collectively delivered 40 GE of traffic on each of the Nexus 9516's 576 ports, and then checked the outputs on all of the ports to make sure it was all there and all handled correctly. The results proved that all of the IP packets were processed and forwarded by the Nexus 9516 properly.

Since the test traffic was all IP packets, all of the regular IP traffic rules had to be obeyed – interframe gaps, packet-handling overhead and so on. This means that traffic applied never actually filled the full 40GE clock rate of the optical link. This is true of any link carrying IP traffic for that matter. Payload gets closer to 100 percent of link speed as larger frame sizes are used (1,518-byte, and "Jumbo" Ethernet frames, 9,216 bytes). Conversely, frame ***rate*** diminishes as the frame size grows. The entire spectrum of frame sizes was applied in testing the Nexus 9516, and in ***all*** cases the maximum load was accepted and forwarded – with no frame loss.

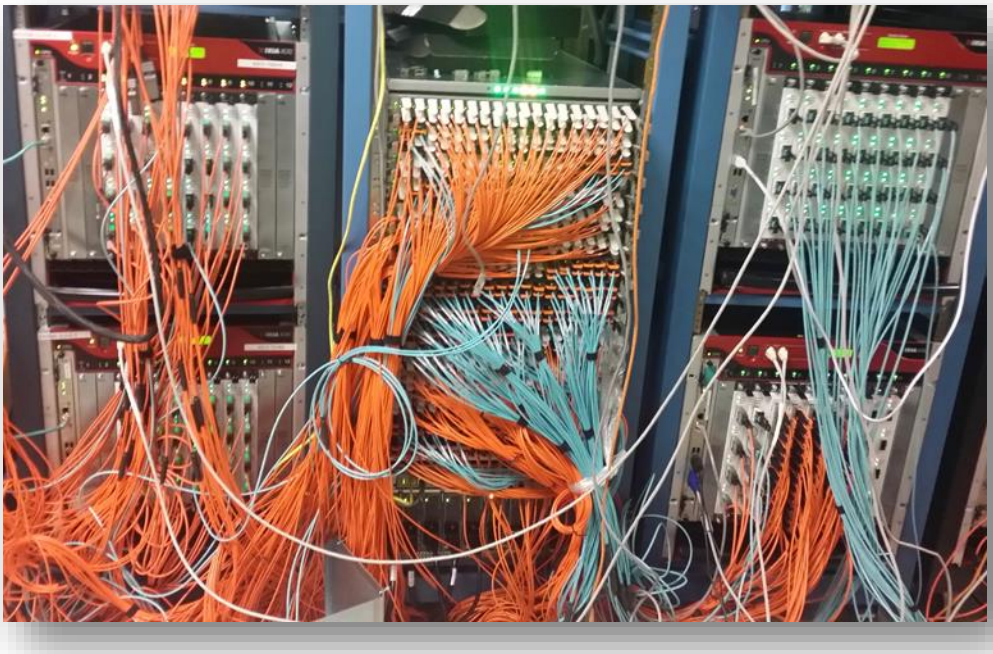
A number of vendors are preparing or already shipping switches designed to function at the heart of today's data centers. Miercom is slated to test several of these in the months ahead. But the current leader, based on all the test metrics we applied here – latency, throughput – is the Cisco Nexus 9516. And the bar for the others has been set very high.

3 Test Bed Setup and How We Did It

All testing of the Cisco Nexus 9516 employed the same test system – multiple Ixia XG12 multi-slot chassis test systems, which are centrally controlled (see details of the Ixia test system below). Sufficient XG12 systems were aggregated to yield 576 Ixia bidirectional 40GE test ports, one for each corresponding port on the Cisco Nexus 9516.



Actual Test Site: Shown below is the switch under test, surrounded by Ixia test systems.



Once the array of Ixia XG12 tests systems were connected to the Nexus 9500, one Ixia system test port connected to each 40GE port on the switch, and connectivity confirmed on each port, we did not have to re-cable for the duration of the testing. The multiple chassis of the Ixia system are centrally controlled.

The battery of tests we applied is discussed individually in this report. They examine different aspects of the device under test, or DUT, in this case the Nexus 9516. The procedures for testing Layer-2 and Layer-3 switches and routers have been more or less standardized in recent years, and this testing employed a number of those procedures. Of course the sheer number and speed of switch ports, and the mind-boggling volume of traffic applied to fill them, are orders of magnitude more today for data center switches than just a few years ago. Even so, the same procedures still apply.

Four of the standards we used in this testing, and which the Ixia test system incorporates, are publicly available as Internet Requests for Comments, or RFCs. The ones applicable here include RFCs 2544, 2889, 3393 and 3918 – for throughput and latency measurements of IPv4 (Layer-3) bidirectional traffic.

RFC 2544, issued in 1999, describes how to conduct basic benchmark tests for latency and throughput measurement. Bidirectional Layer-3 (IP) traffic is applied on port pairs on the device under test (DUT) so that test traffic is processed across the switch fabric.

RFC 2889, issued in 2000, is a reference for conducting more stressful full-mesh tests for throughput and latency measurement. The Ixia test system provides a fully meshed bidirectional traffic flow for these measurements and fully stresses the switch fabric.

RFC 3393, a 2002 document that defines the metric for variation in delay, or jitter, of IP packets passing through a system.

RFC 3918, a 2004 specification, addresses throughput and latency measurement for IP multicast traffic. Based on RFC 3918, the Ixia system supports a combination of traffic profiles with an adjustable number of transmit and receive ports for multicast traffic flows.

Tests were conducted for unicast and multicast traffic throughput, latency, and delay variation. The maximum throughput achievable through the switch was verified, as was proper multicast packet replication.

The multiple Ixia XG12 chassis were driven by the vendor's IxNetwork application, featuring an extensive library of test methodologies and supported scenarios. This was the primary traffic generator that delivered test network traffic through the Nexus 9516 switch.

The tests in this report are intended to be reproducible for users who want to recreate them, with the appropriate test and measurement equipment. Those interested in repeating these tests may contact Miercom at reviews@miercom.com for more details on the configurations applied in this testing. A Miercom professional services sales representative can provide assistance.

4 Throughput and Latency Performance Test

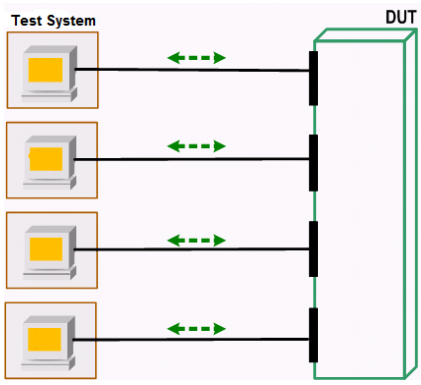
This throughput test determines the maximum rate at which the Nexus 9516 switch receives and forwards traffic without frame loss. Frames are sent at a specified rate, and then the rate is continually stepped up, using a binary search algorithm, to determine the maximum rate at which the switch does not lose frames. Frames can be MAC only, IPv4, IPv6 (with or without Extension Headers) or an IPv4/IPv6 mixture. IPv4 frames, of a complete spectrum of frame sizes, were used for testing the Nexus 9516.

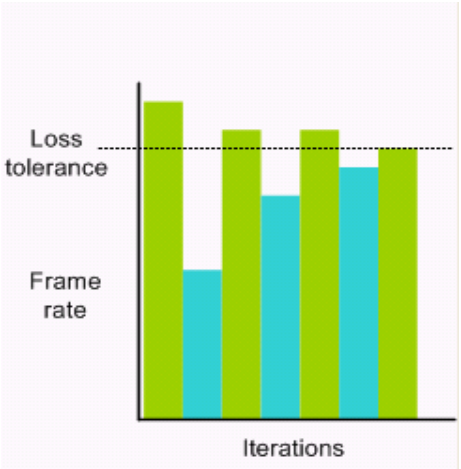
Once the maximum traffic rate is established for a particular frame size, latency through the switch is then calculated – by subtracting the transmit time stamp from the receive time stamp. Based on traffic tests that usually run for a minute or two, the minimum, maximum and average latencies are reported.

The test system's load generator was configured to forward and receive traffic to and from each directly connected port on the switch. As a rule, frames are initially sent at the maximum theoretical rate based on the speed of the port. This test is configured with a one-to-one traffic mapping. The results below show the maximum throughput of the switch without frame loss. When a switch accepts and successfully processes and forwards all traffic at the maximum theoretical rate based on the speed of the port, the switch is said to perform at "wire speed" or "full line rate" for the particular packet size.

The Nexus 9516 was configured for Layer-3 switching (IP routing). Port-pair combinations were assigned within the test system so that bidirectional traffic was transmitted between line-card ports across the fabric modules, in accordance with RFC 2544. All 576 of the Nexus 9516's 40GE ports were connected to the Ixia load-generation system for these tests. The test system established that traffic delivered to and received from each port on the Nexus 9516 could be sent at "wire speed," at all tested packet sizes, without any data loss. For latency measurements testing was conducted using FIFO (First In, First Out) mode.

4.1 RFC 2544 Throughput Test System and DUT Configuration

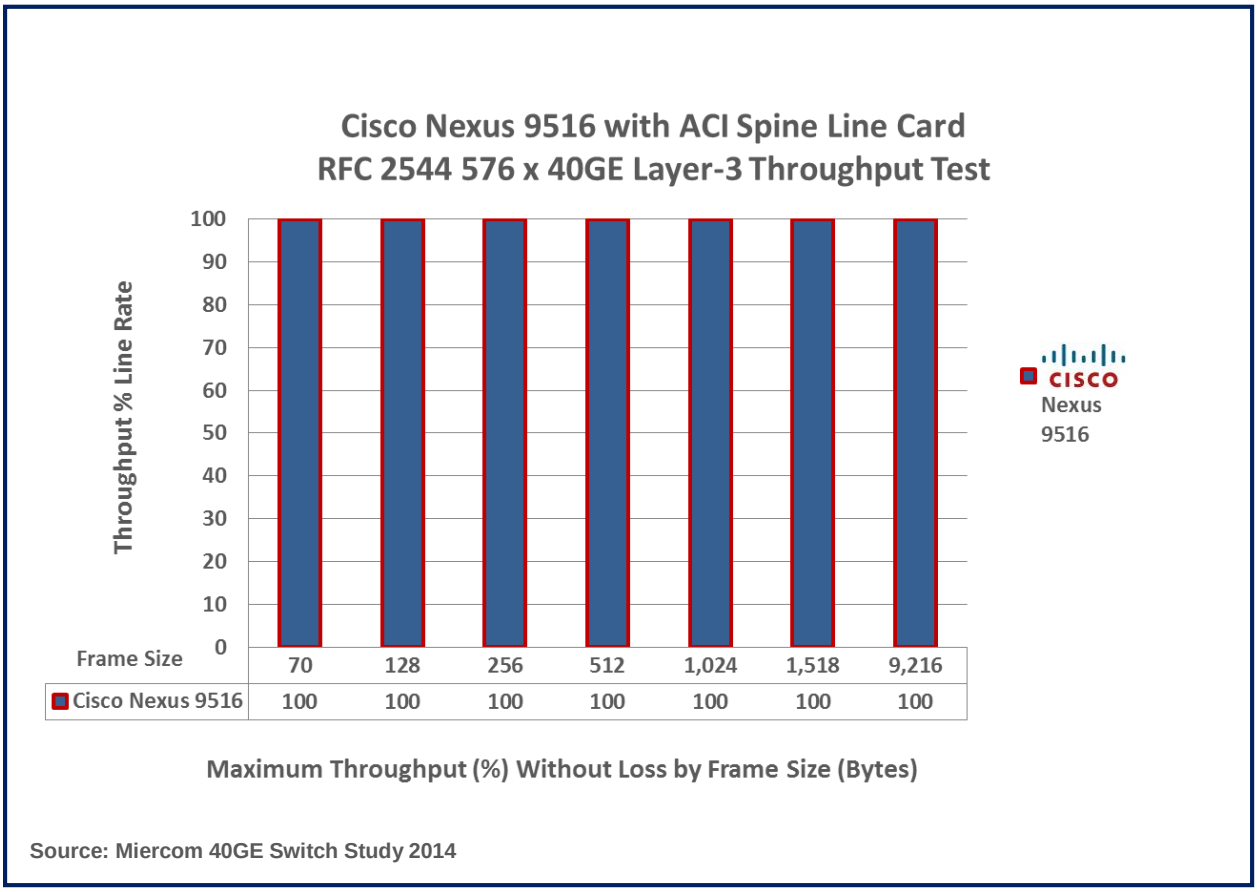




Source Ixia, IxNetwork

4.2 RFC 2544-based 40GE Layer-3 Throughput

"The Cisco Nexus 9516 with sixteen 36-port 40GE ACI spine line cards delivered line-rate performance for all packet sizes during RFC 2544-based Layer-3 Throughput Tests."

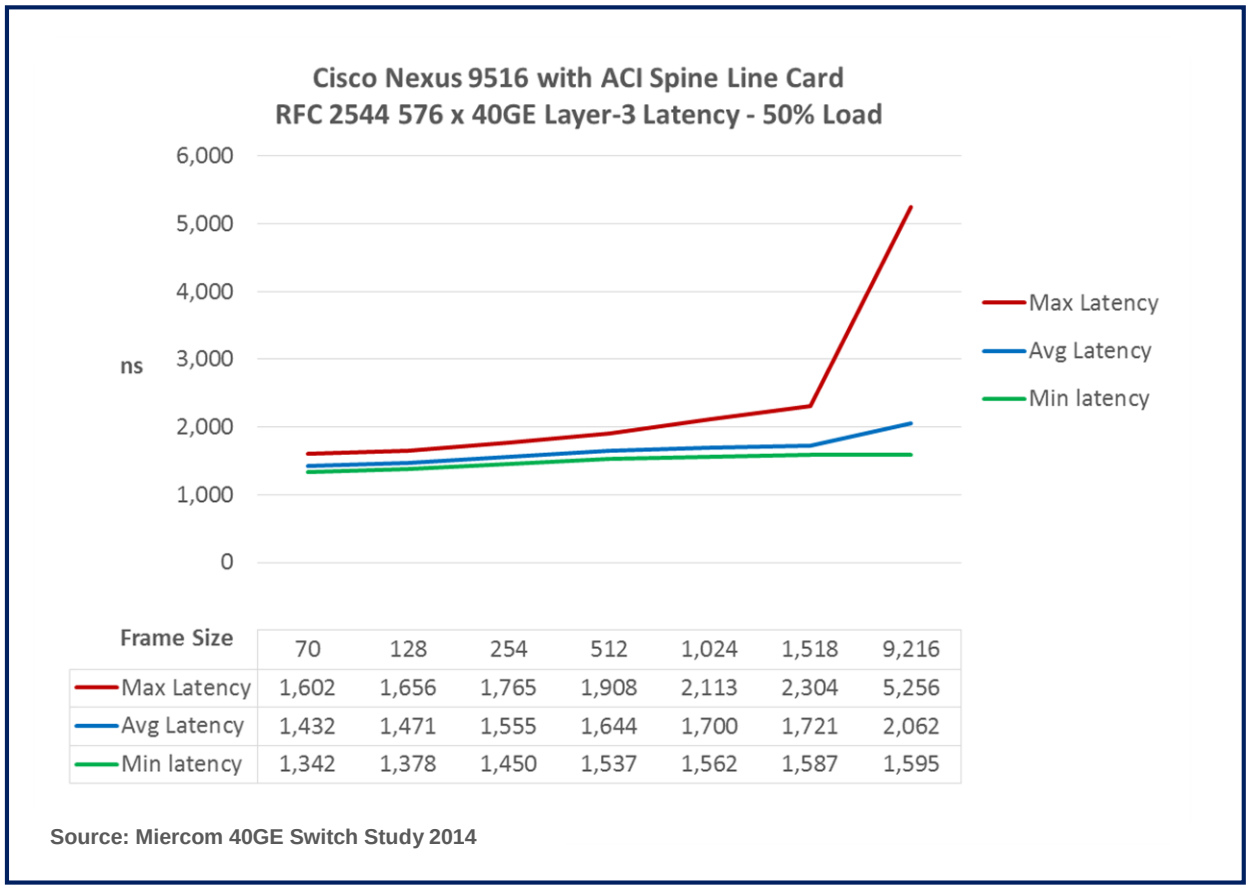


The Cisco Nexus 9516 with sixteen 36p 40G QSFP+ ACI spine line cards exhibits full line-rate forwarding performance with all 576 of its 40GE ports loaded. Test results shown are for 40GE port-pair testing, cross-fabric configuration. The switch was configured with 576 ports under maximum sustainable load without loss for Layer-3 (IP) unicast traffic for the particular frame size. Testing was conducted in accordance with RFC 2544.

Observations - Throughput was tested for different packet sizes in the range of 70 Bytes to 9,216 Bytes. The Cisco Nexus 9516 could handle full-line-rate traffic on all ports, at all frame sizes, without incurring any loss.

4.3 RFC 2544 40GE Layer-3 Latency - 50% Load

"The Cisco Nexus 9516 with sixteen ACI spine line cards exhibits consistent low latency for all packet sizes tested during RFC 2544-based Latency Tests with 50% load."



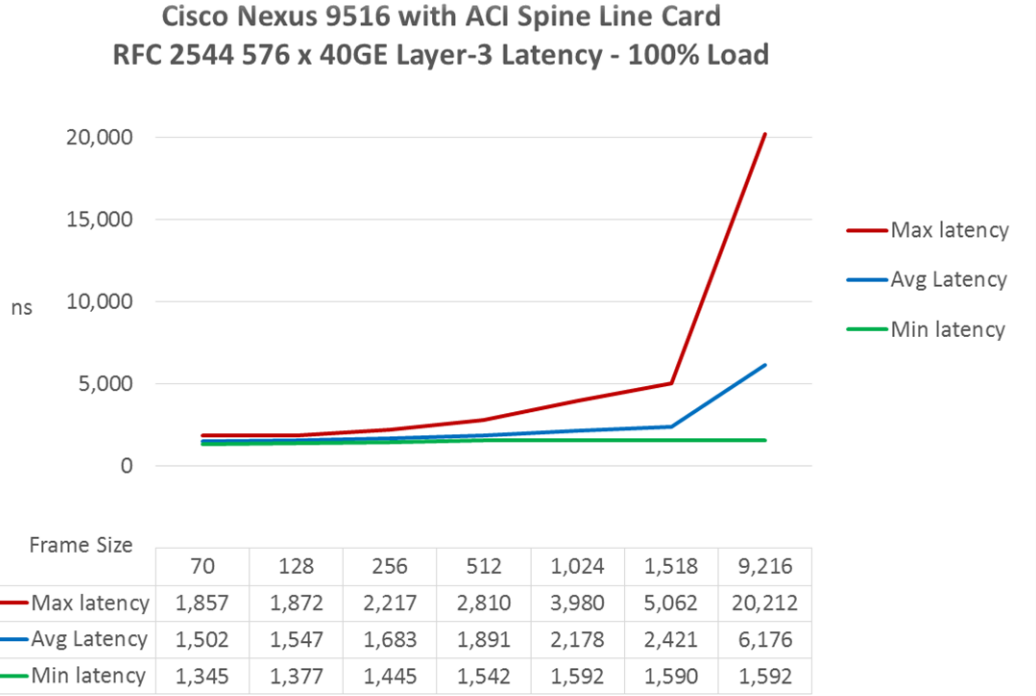
Frame Size	70	128	254	512	1,024	1,518	9,216
Max Latency	1,602	1,656	1,765	1,908	2,113	2,304	5,256
Avg Latency	1,432	1,471	1,555	1,644	1,700	1,721	2,062
Min latency	1,342	1,378	1,450	1,537	1,562	1,587	1,595

Source: Miercom 40GE Switch Study 2014

The Cisco Nexus 9516 with sixteen 36p 40G QSFP+ ACI spine line cards exhibited the latency results shown above during 40GE data center switch testing. The switch was subjected to a 50 percent traffic load of Layer-3 (IP) unicast traffic on all of its 576 ports, for the specified frame size, and exhibited low and consistent latency. From the smallest packet size of 70-Bytes to the largest 9216-Bytes, the average latency ranges from 1.4 to 2.1 microseconds. Tests were conducted in accordance with RFC 2544.

4.4 RFC 2544 40GE Layer-3 Latency - 100% Load

"The Cisco Nexus 9516 with sixteen ACI spine line cards exhibits consistent low latency for all packet sizes tested during RFC 2544-based Latency Tests even with the switch fully loaded 100% on all ports."



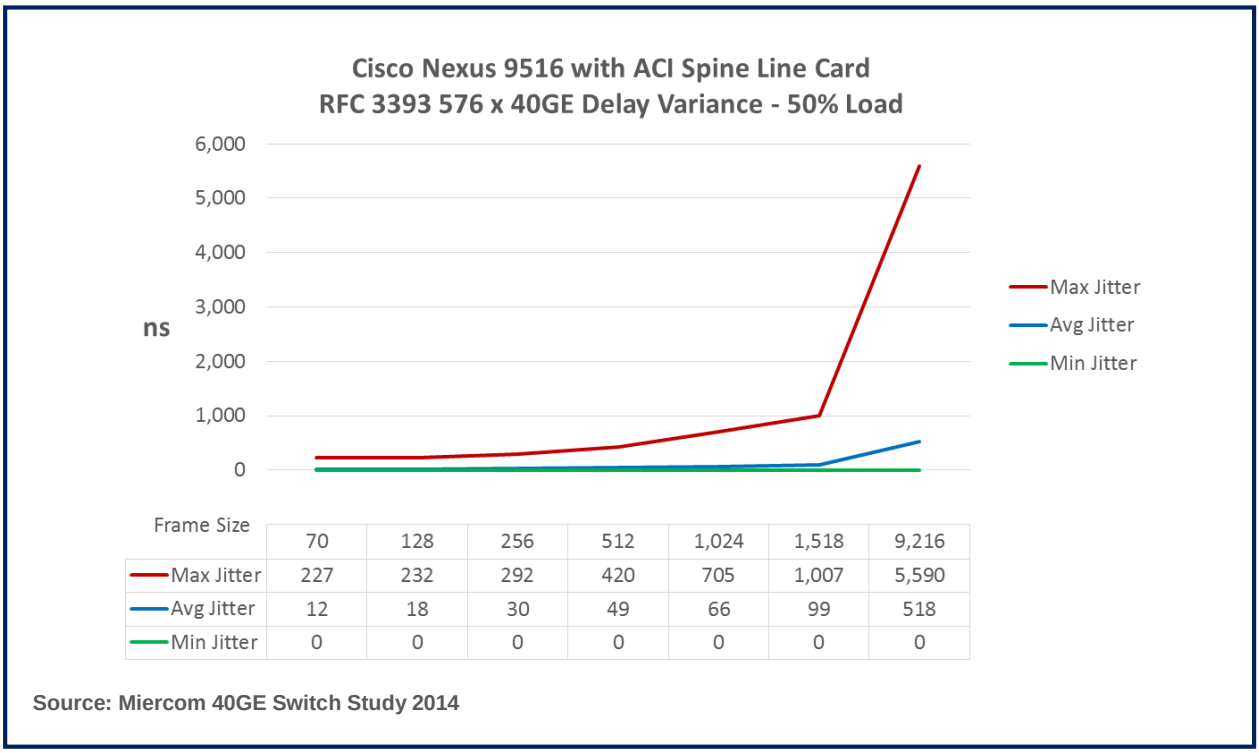
Frame Size	70	128	256	512	1,024	1,518	9,216
Max latency	1,857	1,872	2,217	2,810	3,980	5,062	20,212
Avg Latency	1,502	1,547	1,683	1,891	2,178	2,421	6,176
Min latency	1,345	1,377	1,445	1,542	1,592	1,590	1,592

Source: Miercom 40GE Switch Study 2014

The Cisco Nexus 9516 with 16 x 36p 40G QSFP+ ACI spine line cards exhibited the latency results shown above. The switch was subjected to a 100 percent traffic load of Layer-3 (IP) unicast traffic on all of its 576 ports, for the specified frame size, and exhibited low and consistent latency. From the smallest packet size of 70-Bytes to the largest 9216-Bytes, the average latency ranges from 1.5 to 6.2 microseconds (uSec). Tests were conducted in accordance with RFC 2544.

4.5 RFC 3393-based 40GE Layer-3 Delay Variance- 50% Load

"The Cisco Nexus 9516 with sixteen ACI spine line cards exhibits very little variance and consistent latency for all packet sizes tested during RFC 3393-based Latency Tests with 50% load."



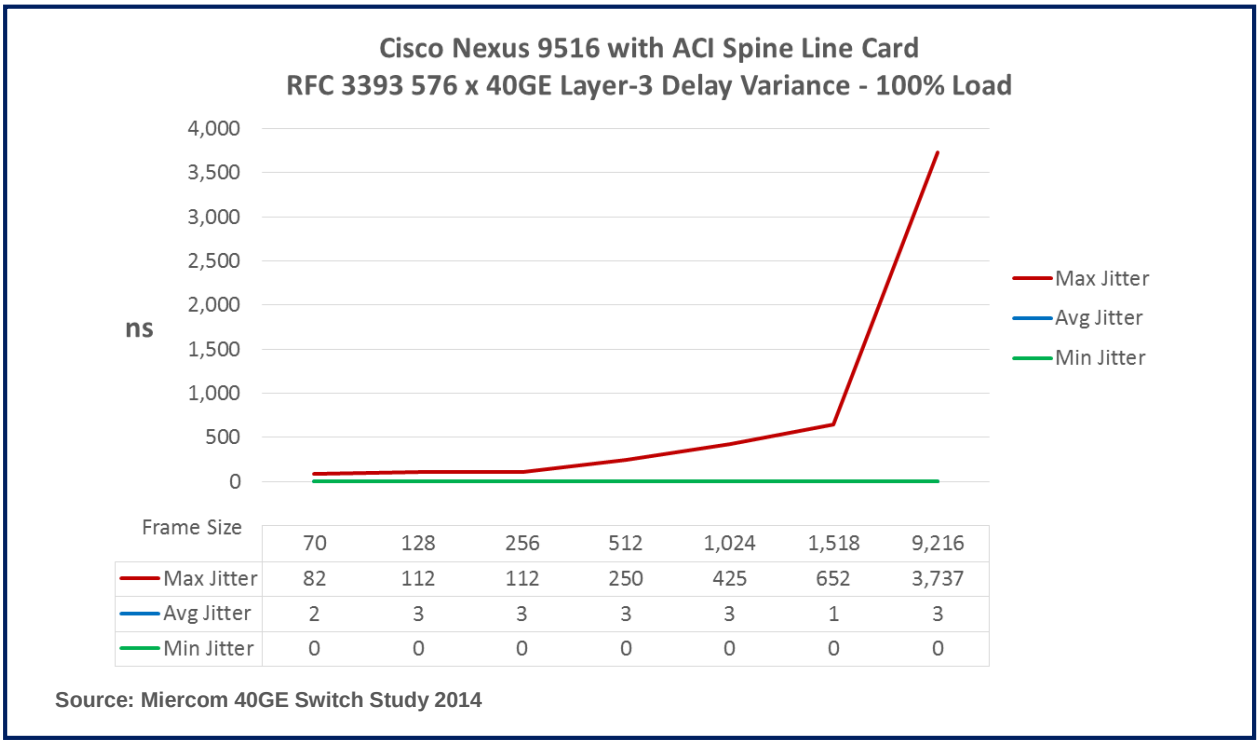
Frame Size	70	128	256	512	1,024	1,518	9,216
Max Jitter	227	232	292	420	705	1,007	5,590
Avg Jitter	12	18	30	49	66	99	518
Min Jitter	0	0	0	0	0	0	0

Source: Miercom 40GE Switch Study 2014

The Cisco Nexus 9516 exhibited very little variance in latency, also called "jitter" – less than 1 microsecond for all packet sizes up to 9,216-byte. The Cisco switch configured with 576 ports was subjected to a 50 percent traffic load. Layer-3 IP unicast traffic was used for the specified frame size. Tests were conducted in accordance with RFC 3393.

4.6 RFC 3393-based 40GE Layer-3 Delay Variance- 100% Load

"The Cisco Nexus 9516 with sixteen ACI spine line cards exhibits very little variance and consistent latency for all packet sizes tested during RFC 3393-based Latency Tests even at 100% load."



Frame Size	70	128	256	512	1,024	1,518	9,216
Max Jitter	82	112	112	250	425	652	3,737
Avg Jitter	2	3	3	3	3	1	3
Min Jitter	0	0	0	0	0	0	0

Source: Miercom 40GE Switch Study 2014

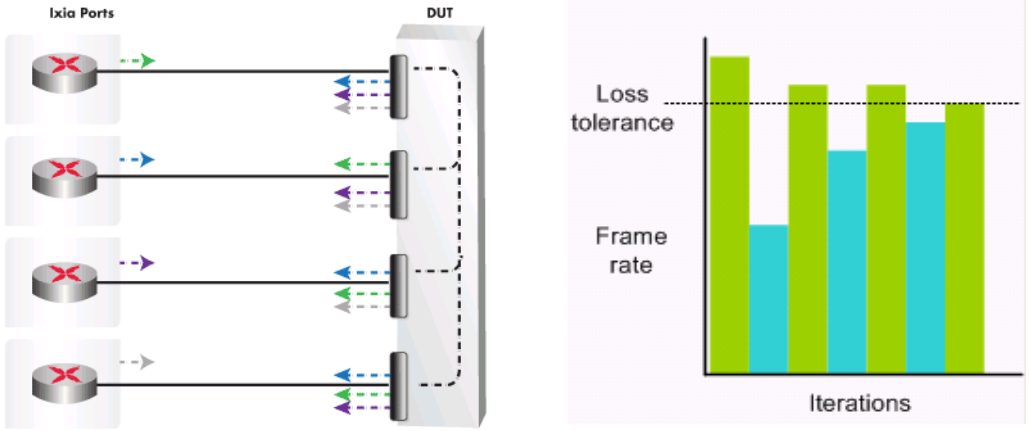
The Cisco Nexus 9516 exhibited very little variance in latency, also called "jitter" – less than 10 nanoseconds or less, on average, for all packet sizes up to 9,216-byte. The Cisco switch configured with 576 ports was subjected to a 100 percent traffic load. Layer-3 IP unicast traffic was used for the specified frame size. Tests were conducted in accordance with RFC 3393.

5 Fully Meshed Throughput and Latency Test

Test Description - The Fully Meshed throughput performance test, as described in RFC 2889, determines the total number of frames that the device under test (DUT) can handle when receiving frames on all ports. The test results show the total number of frames transmitted from, and the total number of frames received on, all ports. In addition, the percentage loss of frames for each frame size is also determined.

Procedure and Configuration - In accordance with RFC 2889 and best known networking practices, all ports of the Ixia load-generation tool are connected, and traffic flows of fixed packet sizes are sent in a mesh-distribution fashion. The device under test is configured for Layer-3 switching (IP routing). The test inherently stresses the switch by sending a "mesh traffic" load distribution, with traffic traversing both the local line card and other line cards, therefore forcing traffic across all the fabric modules. The total number of frames obtained for each frame size for the fully populated switch is recorded. A bidirectional traffic load is used for this test (each port is sending and receiving traffic simultaneously). Testing was conducted using FIFO (First In, First Out) mode.

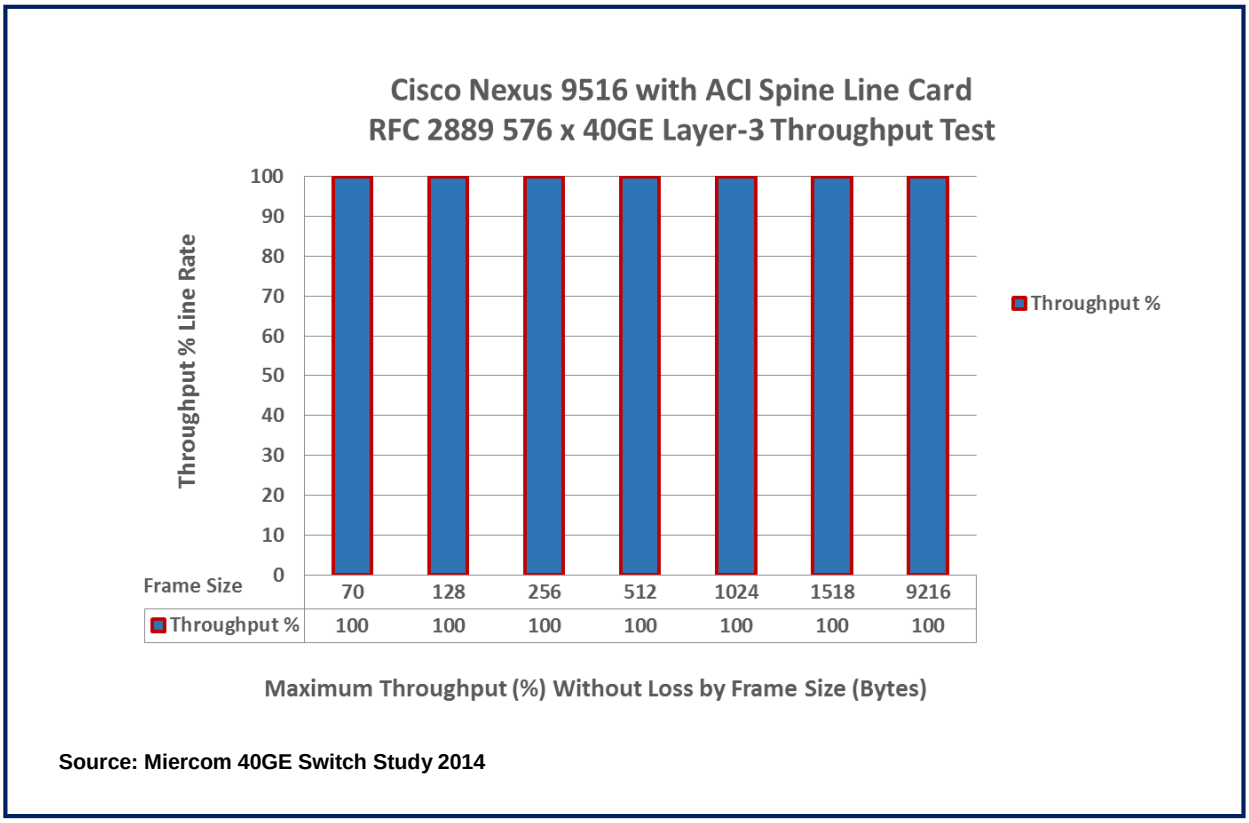
5.1 RFC 2889-based Fully Meshed Test System and DUT Configuration



Source Ixia, IxNetwork

5.2 RFC 2889-based 40GE Layer-3 Full Mesh Throughput

"The Cisco Nexus 9516 with sixteen ACI spine line cards proved full-line-rate mesh throughput performance for all packet sizes tested during RFC 2889-based Throughput Tests without a single packet dropped."

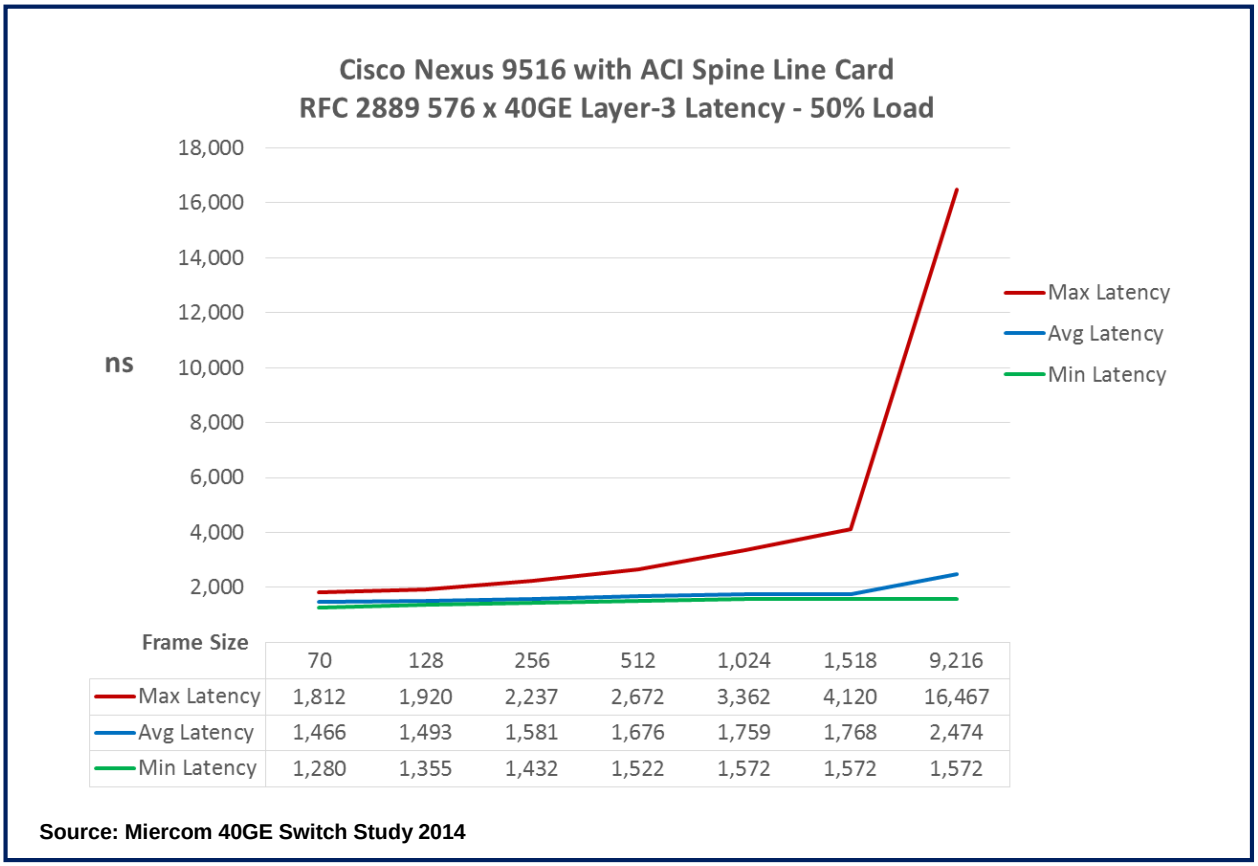


Source: Miercom 40GE Switch Study 2014

The Cisco Nexus 9516 exhibits full-line-rate traffic handling across 576 x 40GE ports loaded. These results are for testing of 40GE ports in a cross-fabric configuration, with full-mesh traffic loads. Tests subjected the switch to the maximum sustainable load that it could maintain without loss of Layer-3 unicast traffic for the specified frame size. Tests were conducted in accordance with RFC 2889.

5.3 RFC 2889-based Full Mesh Layer-3 Latency - 50% Load

"The Cisco Nexus 9516 with sixteen ACI spine line cards exhibits consistent low latency for all packet sizes tested during RFC 2889-based Full Mesh Latency Tests with 50% load."



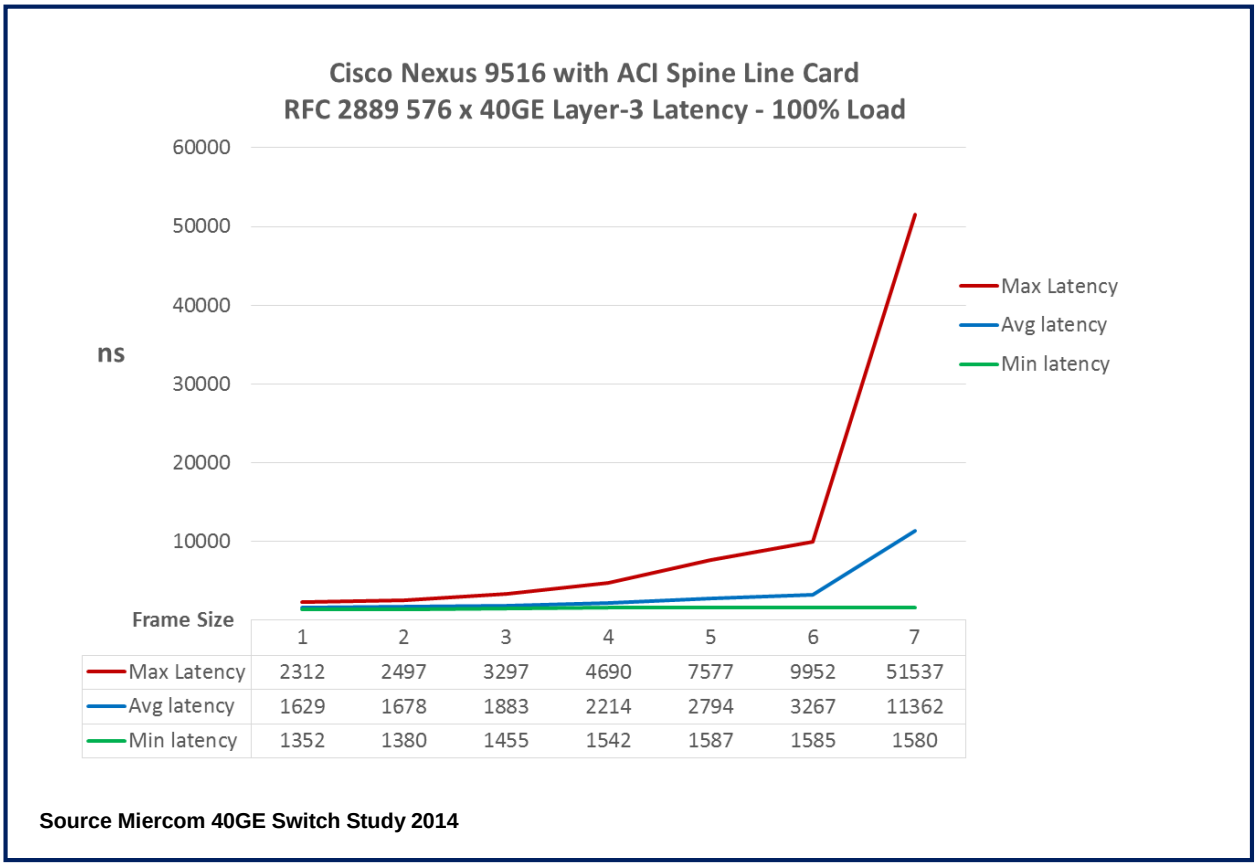
Frame Size	70	128	256	512	1,024	1,518	9,216
Max Latency	1,812	1,920	2,237	2,672	3,362	4,120	16,467
Avg Latency	1,466	1,493	1,581	1,676	1,759	1,768	2,474
Min Latency	1,280	1,355	1,432	1,522	1,572	1,572	1,572

Source: Miercom 40GE Switch Study 2014

The Cisco Nexus 9516 exhibits consistent low latency for the full range of packet sizes tested across all 576 of its 40GE ports with 50 percent load on all ports. From the smallest packet size of 70-Bytes to the largest 9216-Bytes, the average latency ranges from 1.5 to 2.5 microseconds. Results are shown for 576 x 40GE full-mesh testing at 50 percent traffic load. Tests were conducted in accordance with RFC 2889.

5.4 RFC 2889-based Full Mesh Layer-3 Latency - 100% Load

"The Cisco Nexus 9516 with sixteen ACI spine line cards exhibits consistent low latency for all packet sizes tested during RFC 2889 Full Mesh Latency Tests even with the switch loaded 100 percent on all ports."



Frame Size	1	2	3	4	5	6	7
Max Latency	2312	2497	3297	4690	7577	9952	51537
Avg latency	1629	1678	1883	2214	2794	3267	11362
Min latency	1352	1380	1455	1542	1587	1585	1580

Source Miercom 40GE Switch Study 2014

The Cisco Nexus 9516 exhibits consistent low latency for the full range of packet sizes tested across all 576 of its 40GE ports with 100 percent load on all ports. From the smallest packet size of 70-Bytes to the largest 9216-Bytes, the average latency ranges from 1.6 to 11.4 microseconds. Results are shown for 576 x 40GE full-mesh testing at 100 percent traffic load. Tests were conducted in accordance with RFC 2889.

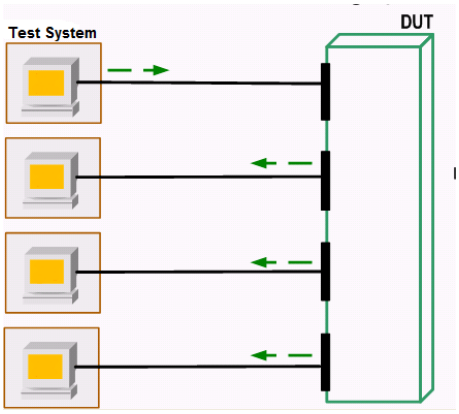
6 IP Multicast Throughput and Latency

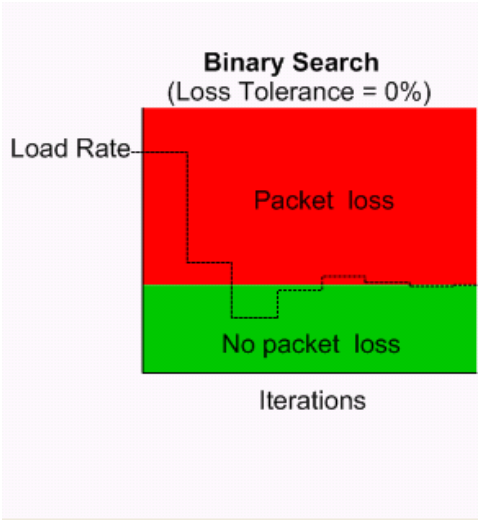
Description - RFC 3918 describes Multicast Forwarding and Latency test, for measuring the forwarding throughput and the minimum, maximum and average latency of multicast frames. These are frames generated and received by the test and measurement equipment. The multicast frames are sent to "clients" on multiple subnets (ports), which are also configured via the test and measurement system.

The test reveals how much processing overhead the device under test (DUT) requires, on average, to forward multicast frames. In a typical scenario the tester defines the multicast protocol (our testing used IGMPv1 2 3; PIM-SM; SSM), and the number of multicast groups to be sent. Traffic streams are automatically built by the Ixia test system. A combination of throughput, latency, group capacity, frame loss, join delay, and leave delay can be calculated from the results.

This test is used to measure the IP multicast forwarding throughput of the DUT and then calculate the latency of traffic. This is done by examining the "sent" timestamp that the test system places within test frames. When frames are received from the ports, the test system compares this timestamp with the current time and calculates the difference, which is the latency. The test system records the average, minimum and maximum latencies for each multicast group, in microseconds (μs).

6.1 RFC 3918 IP Multicast Test System and DUT Configuration





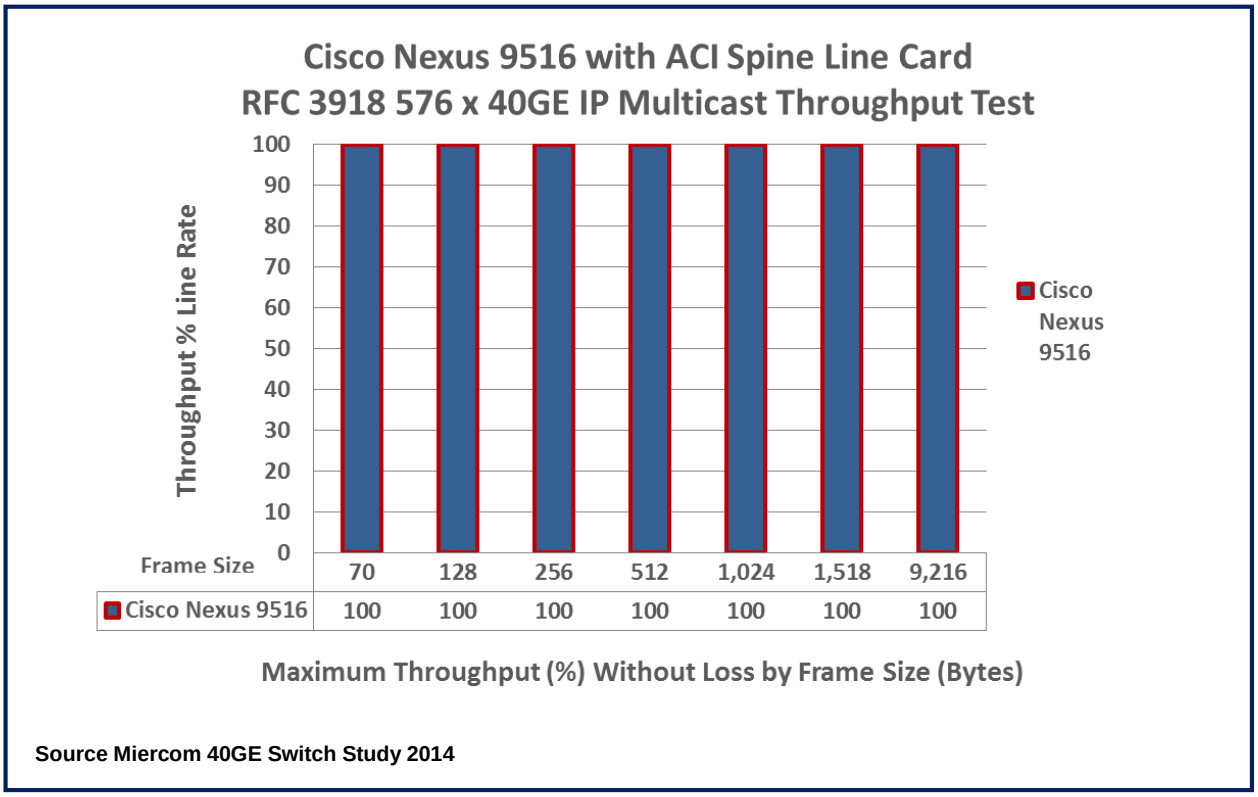
Source Ixia, IxNetwork

6.2 RFC 3918 40GE IP Multicast Throughput

IP multicast is a process and protocol for sending the same IP packet stream to a group of interested receivers. It is an efficient way to achieve one-to-many or many-to-many communication between an affiliated group of IP devices. IP multicast is heavily used by financial trading applications, where performance is critical to those applications.

To understand the IP-multicast-handling capability of the Nexus 9516 we performed RFC 3918-based IP multicast throughput tests, where IP multicast traffic was sent from one input port to all the other ports on the switch. So there were 575 receivers on the Nexus 9516 (traffic sent in on one port and delivered via IP multicast to all the other ports). The graph below shows the throughput testing results.

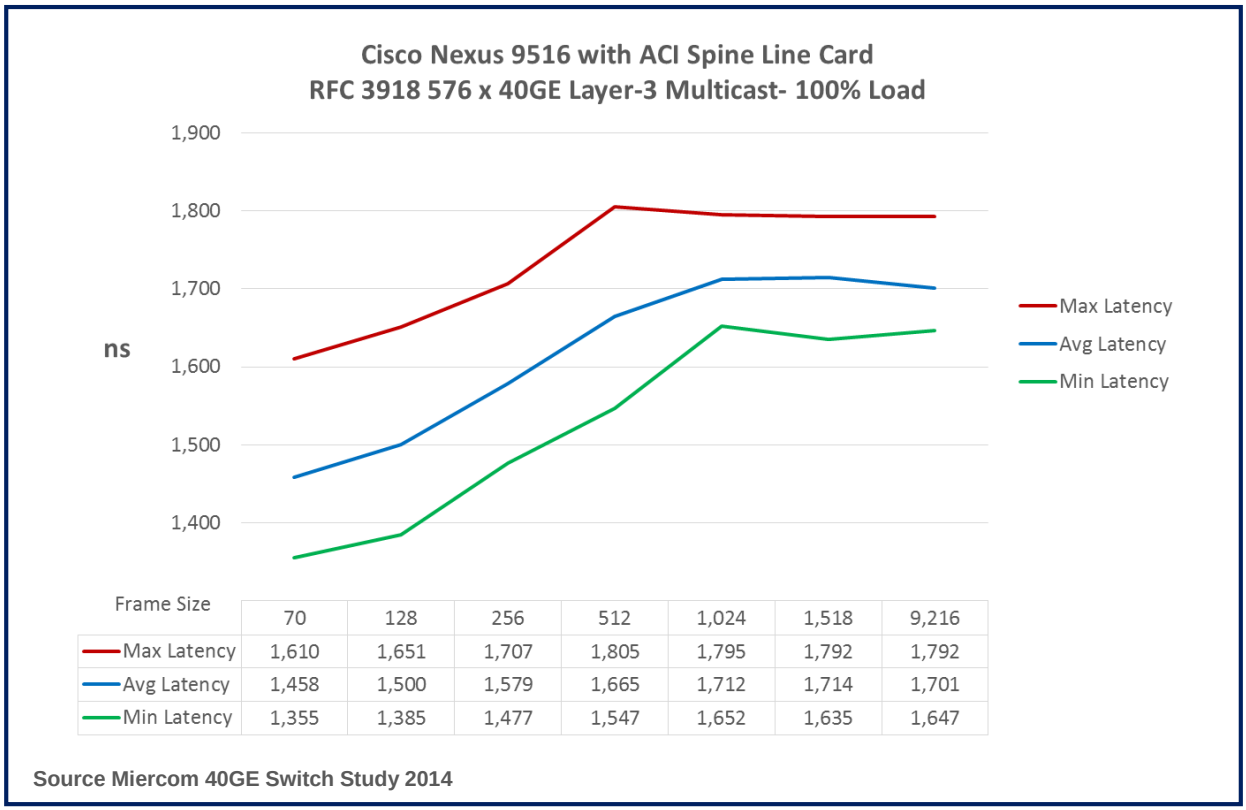
"The Cisco Nexus 9516 with sixteen ACI spine line cards delivered line-rate performance for all packet sizes during RFC 3918-based IP-Multicast Tests."



Source Miercom 40GE Switch Study 2014

The Cisco Nexus 9516 with 16 x 36p 40G QSFP+ ACI spine line cards exhibits full-line-rate multicast handling across all 576 of its 40GE ports, from a fully loaded input traffic stream. Tests were conducted in accordance with RFC 3918.

6.3 RFC 3918-based 40GE IP-Multicast Latency



Frame Size	70	128	256	512	1,024	1,518	9,216
Max Latency	1,610	1,651	1,707	1,805	1,795	1,792	1,792
Avg Latency	1,458	1,500	1,579	1,665	1,712	1,714	1,701
Min Latency	1,355	1,385	1,477	1,547	1,652	1,635	1,647

Source Miercom 40GE Switch Study 2014

The Cisco Nexus 9516 exhibits low latency for IP-Multicast traffic across all 576 of its 40GE ports, under 100 percent traffic load. From the smallest packet size of 70-Bytes to the largest 9216-Bytes, the average latency ranges from 1.5 to 1.7 microseconds. The test results shown are for 576 x 40GE IP Multicast testing. The Cisco switch was configured with 576 ports under 100 percent load without loss for Layer-3 IP-Multicast traffic for the specified frame size. Tests were conducted in accordance with RFC 3918.

7 Miercom's 40GE Switch Industry Study

This report was sponsored by Cisco Systems, Inc. The data was obtained completely and independently as part of the Miercom 40GE Switch Industry Study, an ongoing endeavor in which all vendors have equal opportunity to participate and contribute to the test methodology.

8 About Miercom

Miercom has published hundreds of network-product-comparison analyses in leading trade periodicals including ***Network World, Business Communications Review - NoJitter, Communications News, xchange, Internet Telephony*** and other leading publications. Miercom's reputation as the leading, independent product test center is undisputed.

Miercom's private test services include competitive product analyses, as well as individual product evaluations. Miercom features comprehensive certification and test programs including: Certified Interoperable, Certified Reliable, Certified Secure and Certified Green. Products may also be evaluated under the Performance Verified program, the industry's most thorough and trusted assessment for product usability and performance.