networktest

Cisco Nexus 9516: Scaling 100G Performance to New Heights

June 2015

Executive Summary

There's a capacity boom in the data center, with servers and access switches putting pressure on the network backbone like never before. How to scale up to handle today's load, and provide a path for future expansion?

Cisco Systems commissioned Network Test, an independent third-party test lab, to help answer that question. Network Test assessed the Cisco Nexus 9516 core switch fully loaded with 128 100G interfaces, making this the largest Cisco Nexus 9000 Series evaluation ever conducted.

Seven sets of rigorous benchmark test results demonstrated high performance and versatility in every case:

- Zero frame loss in all tests, covering unicast, multicast, Layer-2, Layer-3, and routing across all 128 100G Ethernet ports
- Low latency and jitter across test cases
- Loss-free performance when forwarding to BGP and BGP-MP routes using IPv4 and IPv6 traffic
- Loss-free performance when forwarding to more than 520,000 IP multicast routes
- Support for a variety of 100G Ethernet transceiver types

This report is organized as follows. This section provides an overview of the test results. The "About This Test" section explains the importance of each metric used and briefly describes the issues common to all test cases. The "Performance Test Results" section provides full results from individual test cases. The "Test Methodology" section describes test procedures in detail, allowing interested parties to reproduce these results if desired. Appendix A presents maximum jitter measurements from all tests. Appendix B provides software versions used in testing.

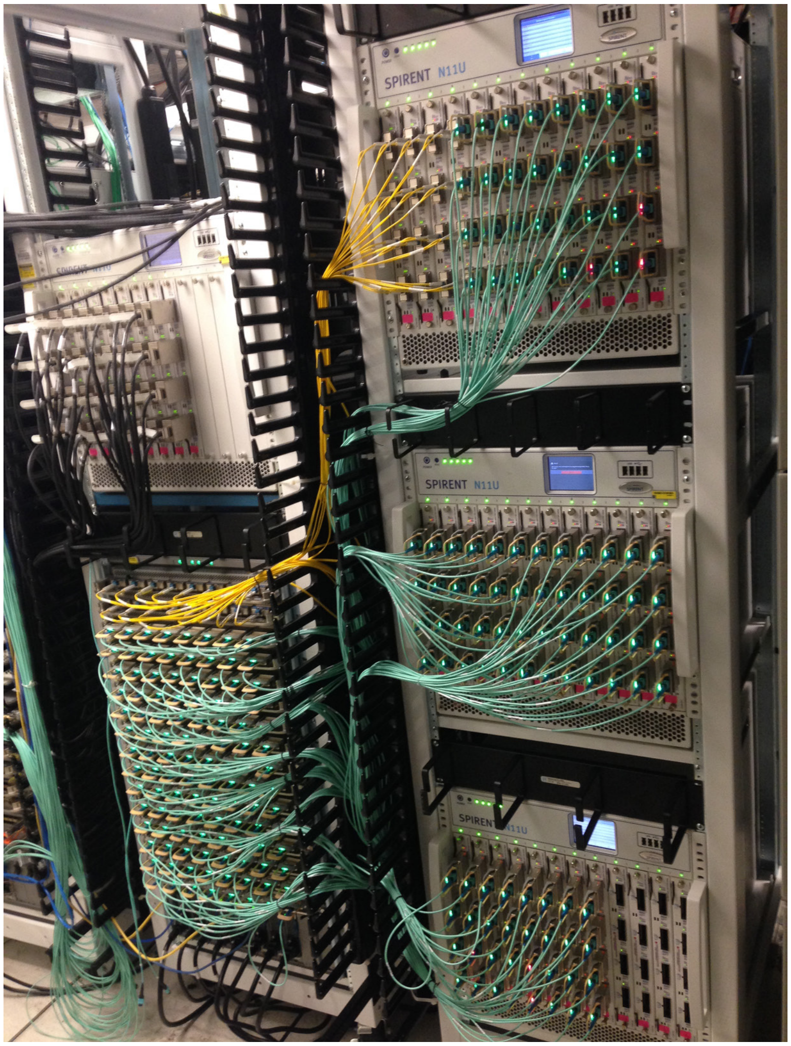


Figure 1: The Cisco Nexus 9516 Test Bed: 128 100G Ethernet Ports

About This Test

This project characterized Cisco Nexus 9516 performance with the switch in seven configuration modes, all involving 128 100G Ethernet interfaces:

- RFC 2889 Ethernet unicast performance
- RFC 2544 IPv4 unicast performance
- RFC 2544 IPv4 unicast performance with BGP routing
- RFC 5180 IPv6 unicast performance
- RFC 5180 IPv6 unicast performance with BGP-MP routing
- RFC 3918 Ethernet multicast performance
- RFC 3918 IPv4 multicast performance

Figure 1 shows the Cisco Nexus 9516 on the test bed, fully loaded with dual Supervisor A engines and N9K-X9408PC-CFP2 modules, along with the Spirent TestCenter traffic generator/analyzer with dX2-100G-P4 modules. Spirent TestCenter is capable of offering traffic at wire speed on all ports with transmit timestamp resolution of 2.5 nanoseconds.

In these relatively early days of 100G Ethernet deployment, transceiver availability can be an issue. To accommodate the maximum number of choices, both the Cisco Nexus 9516 and Spirent instrument use CFP2 form factor, allowing the use of a variety of optics. Figure 2 shows a matrix with the different SR10 and LR4 transceiver types used on the test bed.

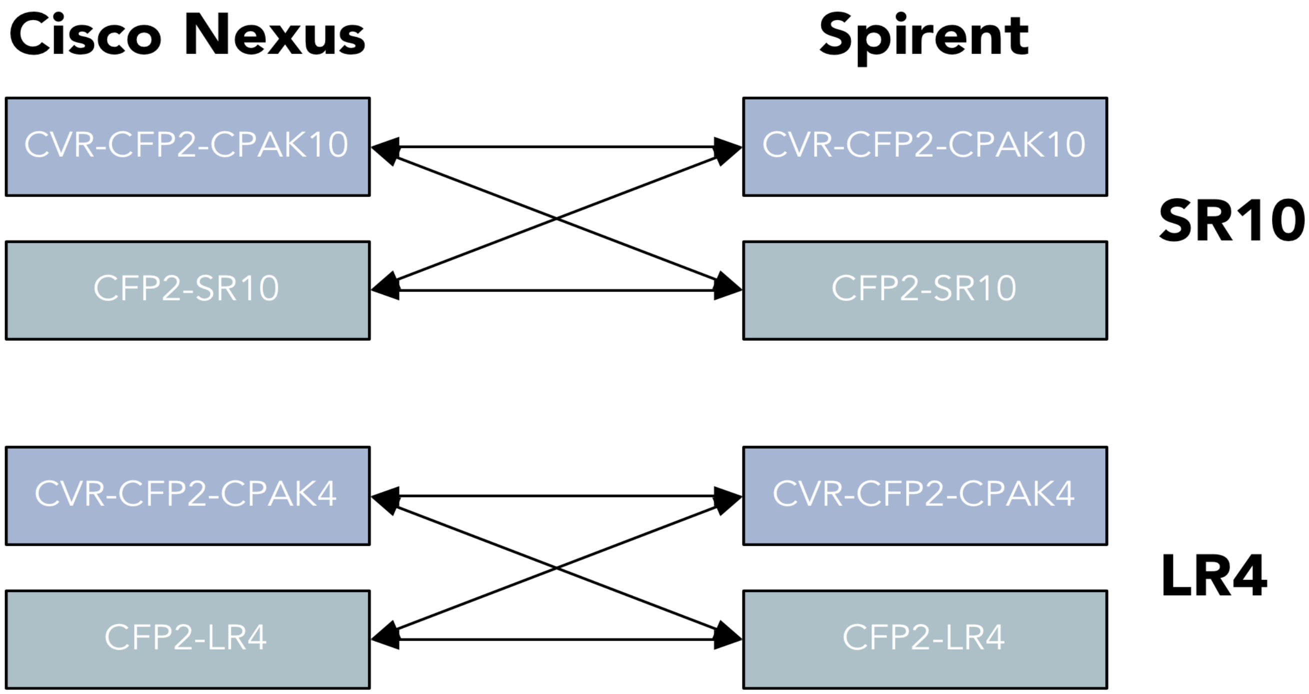


Figure 2: 100G Ethernet Transceiver Matrix

For all tests, the primary metrics were throughput, latency, and jitter.

RFC 2544, the industry-standard methodology for network device testing, determines throughput as the limit of system performance. In the context of lab benchmarking, throughput describes the maximum rate at which a device forwards all traffic with zero frame loss. Describing "real-world" performance is explicitly a non-goal of RFC 2544 throughput testing. Indeed, production networks load are typically far lower than the throughput rate.

Latency and jitter respectively describe the delay and delay variation introduced by a switch. Both are vital, and arguably even more important than throughput, especially for delay-sensitive applications such as video, voice, and some financial trading applications.

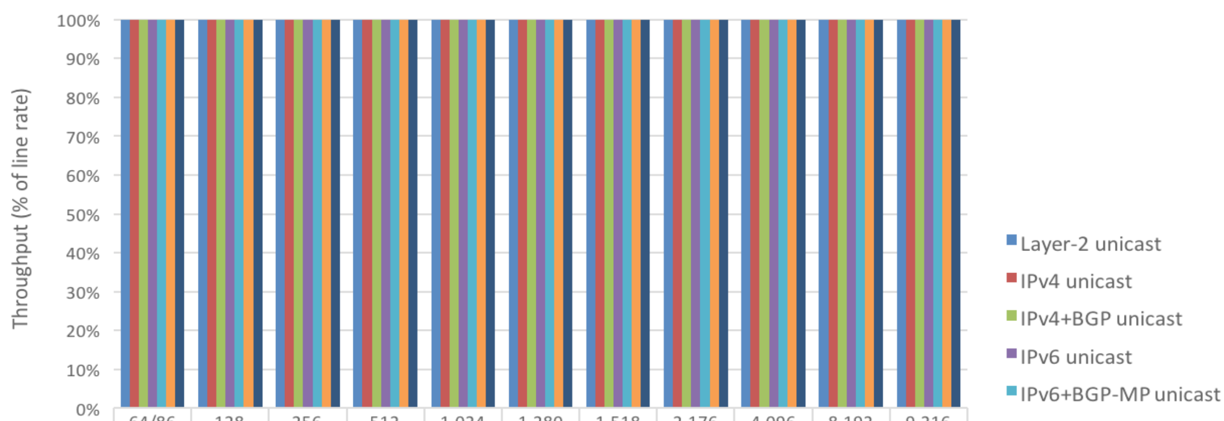
RFC 2544 requires latency be measured at, and only at, the throughput rate. Since average utilization in production networks is typically far lower than line rate, it can also be useful to characterize delay for traffic at lower rates.

All tests described here present latency and jitter not only at the throughput rate, but also at 10, 50, and 90 percent of line rate. These results should help network professionals understand Cisco Nexus 9516 in *their* networks, modeling *their* network utilizations.

Performance Test Results

This section describes results for each configuration mode. See the "Test Methodology" section for details on test procedures.

Before delving into each test scenario in detail, Figure 3 presents throughput results across all cases. The Cisco Nexus 9516 delivered virtual line rate performance in every test case, for every frame size.



	64/86	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
Layer-2 unicast	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%
IPv4 unicast	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%
IPv4+BGP unicast	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%
IPv6 unicast	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%
IPv6+BGP-MP unicast	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%
Layer-2 multicast	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%
Layer-3 multicast	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%	99.999%

Frame size (bytes)

Figure 3: Cisco Nexus 9516 throughput across test scenarios

RFC 2889 Ethernet Unicast Performance

For Layer-2 testing, engineers configured all 128 100G Ethernet ports of the Cisco switch to be access-mode members of the same VLAN. The Spirent TestCenter traffic generator/analyzer offered test traffic to the Cisco Nexus 9516 in a "fully meshed" pattern, meaning traffic offered to each port was destined to all 127 other ports. The Spirent test instrument emulated one host attached to each port in the VLAN.

The test instrument offered traffic at the throughput rate, and also measured latency and jitter at that rate for a variety of frame sizes. Frame sizes range from the Ethernet minimum of 64 bytes to the maximum of 1,518, and beyond to jumbo frame sizes of up to 9,216 bytes. The 2,176-byte frame size is common for data centers using Fibre Channel over Ethernet (FCoE) encapsulation.

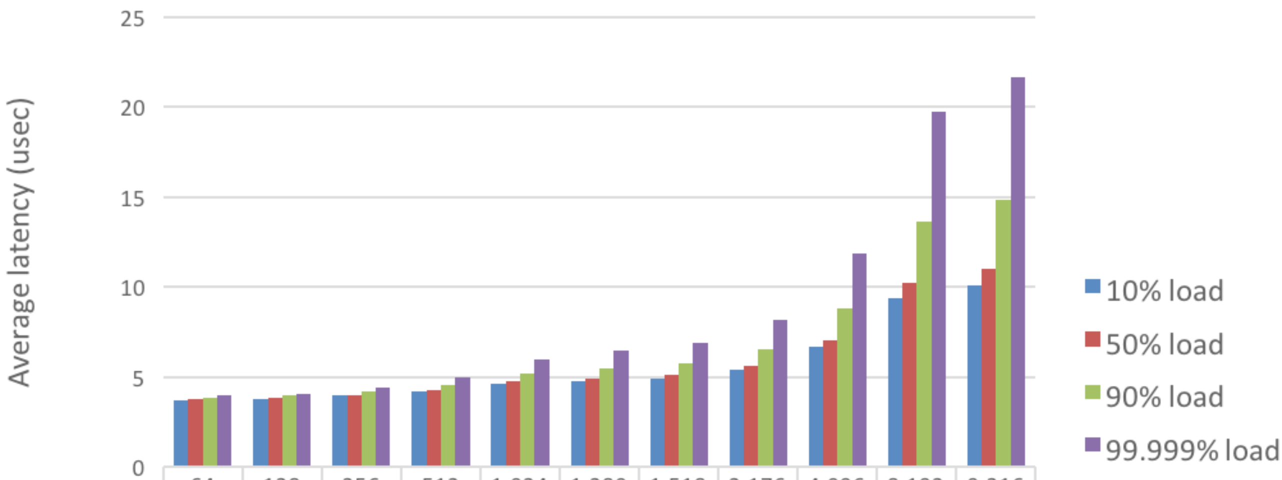
For all tests, the "throughput rate" was 99.999 percent of line rate. Test engineers opted to use 99.999 percent of line rate, which is 10 parts per million (10 ppm) slower than nominal line rate, to avoid clocking differences between the traffic generator and the switch under test. The IEEE 802.3 Ethernet specification requires interfaces to tolerate clocking differences of up to +/- 100 ppm.

Table 1 presents throughput, latency, and jitter results from the Layer-2 unicast tests.

Figures 4 and 5 compare average and maximum delay measurements, respectively, with offered loads of 10, 50, 90, and 99.999 percent of line rate.

	Throughput			**Latency @ throughput rate**			
Frame size (bytes)	**Frames/s**	**Tbit/s**	**% line rate**	**Minimum (usec)**	**Average (usec)**	**Maximum (usec)**	**Maximum jitter (usec)**
64	19,047,430,585	9.752	99.999%	2.22	3.95	6.41	2.20
128	10,810,703,846	11.070	99.999%	2.26	4.08	6.77	2.42
256	5,797,044,091	11.872	99.999%	2.38	4.41	7.53	2.89
512	3,007,489,040	12.319	99.999%	2.50	4.94	8.15	3.37
1,024	1,532,551,887	12.555	99.999%	2.72	5.99	10.28	4.52
1,280	1,230,757,053	12.603	99.999%	2.80	6.47	11.05	5.21
1,518	1,040,301,801	12.633	99.999%	2.84	6.90	12.48	6.04
2,176	728,590,241	12.683	99.999%	3.03	8.16	16.11	8.63
4,096	388,723,073	12.738	99.999%	3.57	11.84	25.37	15.83
8,192	194,834,897	12.769	99.999%	4.72	19.71	46.67	32.21
9,216	173,233,453	12.772	99.999%	5.01	21.68	52.17	36.23

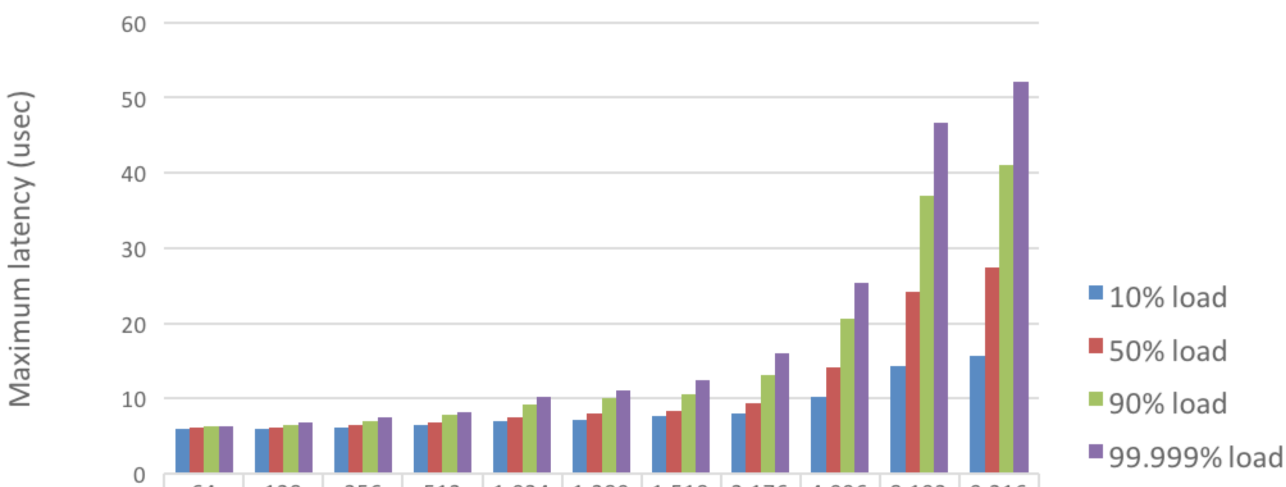
Table 1: Layer-2 unicast performance results



	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	3.72	3.79	3.96	4.19	4.62	4.79	4.93	5.38	6.66	9.40	10.08
50% load	3.76	3.82	4.00	4.25	4.73	4.93	5.08	5.60	7.07	10.21	11.00
90% load	3.86	3.95	4.19	4.53	5.20	5.49	5.74	6.52	8.78	13.61	14.82
99.999% load	3.95	4.08	4.41	4.94	5.99	6.47	6.90	8.16	11.84	19.71	21.68

Frame size (bytes)

Figure 4: Layer-2 unicast average latency vs. load



	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	6.02	6.07	6.24	6.58	6.98	7.20	7.73	8.10	10.29	14.28	15.69
50% load	6.12	6.22	6.46	6.77	7.53	8.09	8.45	9.39	14.17	24.26	27.36
90% load	6.34	6.58	7.05	7.85	9.14	10.15	10.57	13.09	20.60	36.91	41.03
99.999% load	6.41	6.77	7.53	8.15	10.28	11.05	12.48	16.11	25.37	46.67	52.17

Frame size (bytes)

Figure 5: Layer-2 unicast maximum latency vs. load

RFC 2544 IPv4 Unicast Performance

As with the layer-2 tests, IPv4 performance tests involved a "fully meshed" pattern, with traffic offered to each port destined to all other ports. In this case, however, the Cisco Nexus 9516 acted as a router. Engineers configured each interface to be a member of a different IP subnet, requiring the device to route traffic between subnet. The Spirent test instrument emulated one host per subnet.

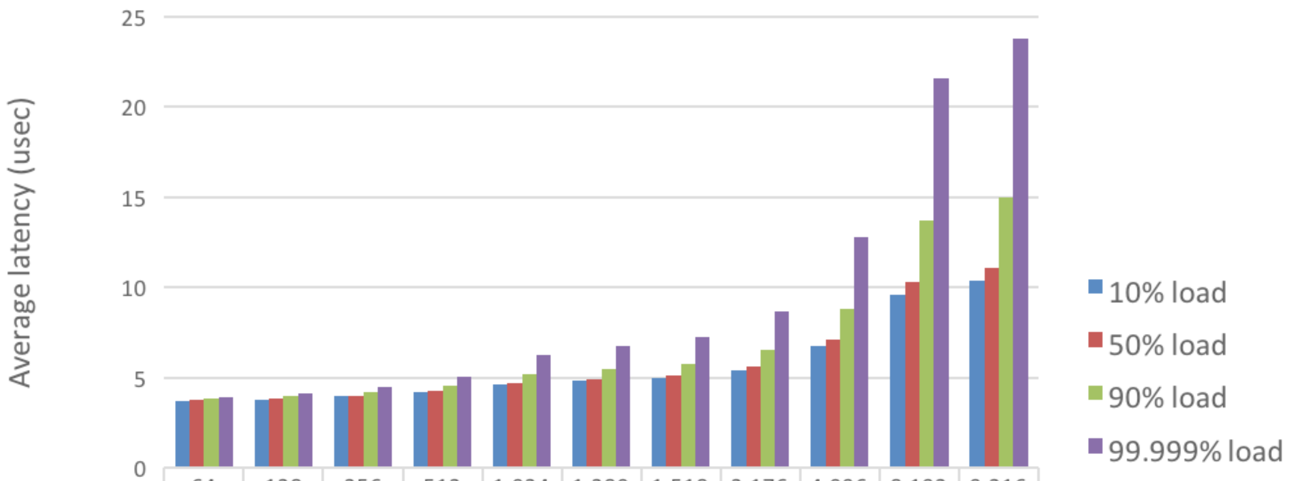
Test traffic carried not only IPv4 but also UDP headers, in this case using a range of 8,000 unique source and destination ports. By default, the Cisco Nexus 9516 uses a hashing algorithm that uses Layer-2/3/4 criteria to distribute flows across the switch fabric. The range of UDP source and destination ports helps ensure a uniform distribution of flows across the fabric.

Throughput again was equivalent to 99.999 percent of line rate, regardless of frame size. Table 2 presents throughput, latency, and jitter results for all frame sizes.

Figures 5 and 6 compare average and maximum delay measurements, respectively, with offered loads of 10, 50, 90, and 99.999 percent of line rate.

	Throughput			Latency @ throughput rate			
Frame size (bytes)	**Frames/s**	**Tbit/s**	**% line rate**	**Minimum (usec)**	**Average (usec)**	**Maximum (usec)**	**Maximum jitter (usec)**
64	19,047,430,585	9.752	99.999%	2.23	3.96	6.42	2.25
128	10,810,703,845	11.070	99.999%	2.25	4.11	6.79	2.54
256	5,797,044,091	11.872	99.999%	2.37	4.48	7.38	2.84
512	3,007,489,040	12.319	99.999%	2.51	5.08	8.75	3.58
1,024	1,532,551,886	12.555	99.999%	2.71	6.24	10.95	4.93
1,280	1,230,757,053	12.603	99.999%	2.77	6.78	11.93	5.73
1,518	1,040,301,801	12.633	99.999%	2.85	7.27	13.75	6.52
2,176	728,590,241	12.683	99.999%	3.03	8.67	16.94	9.15
4,096	388,723,073	12.738	99.999%	3.56	12.78	27.63	17.19
8,192	194,834,897	12.769	99.999%	4.69	21.58	50.60	33.67
9,216	173,233,453	12.772	99.999%	4.99	23.79	56.38	37.74

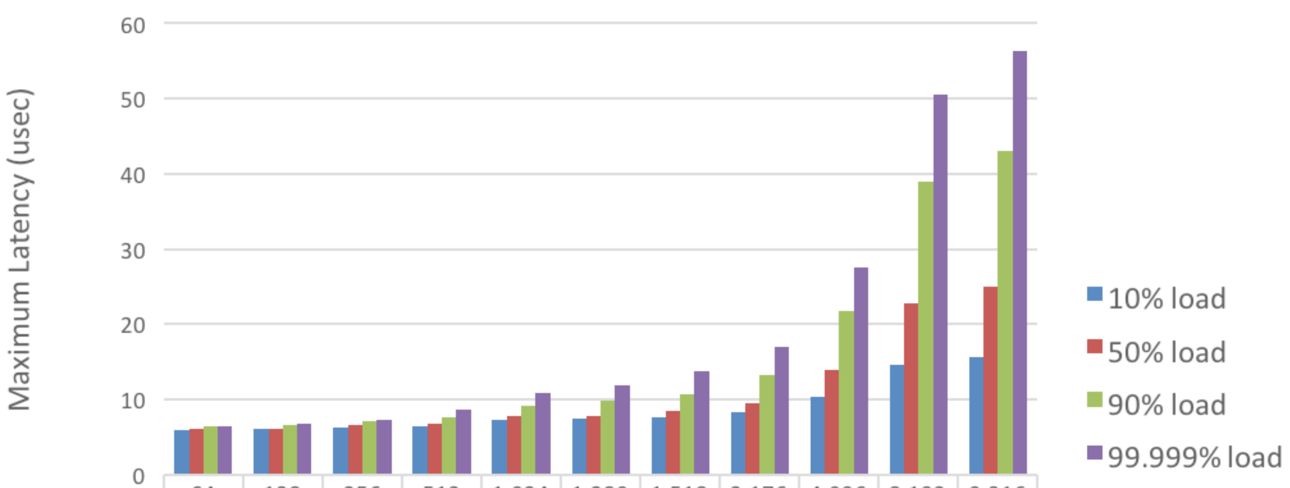
Table 2: IPv4 unicast performance results



	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	3.72	3.79	3.96	4.21	4.65	4.82	4.97	5.44	6.77	9.63	10.34
50% load	3.76	3.82	4.00	4.26	4.74	4.94	5.10	5.62	7.10	10.29	11.09
90% load	3.86	3.96	4.20	4.55	5.22	5.52	5.77	6.56	8.85	13.75	14.97
99.999% load	3.96	4.11	4.48	5.08	6.24	6.78	7.27	8.67	12.78	21.58	23.79

Frame size (bytes)

Figure 6: IPv4 unicast average latency vs. load



	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	5.99	6.06	6.28	6.53	7.22	7.39	7.57	8.31	10.29	14.60	15.69
50% load	6.10	6.16	6.58	6.85	7.73	7.80	8.51	9.59	13.89	22.76	25.09
90% load	6.39	6.60	7.08	7.68	9.21	9.78	10.65	13.31	21.85	38.90	42.99
99.999% load	6.42	6.79	7.38	8.75	10.95	11.93	13.75	16.94	27.63	50.60	56.38

Frame size (bytes)

Figure 7: IPv4 unicast maximum latency vs. load

RFC 2544 IPv4 Unicast Performance With BGP Routing

This tests adds Border Gateway Protocol (BGP) routing to the previous IPv4 test, which involved routing only between subnets locally configured on the switch. Here, the Cisco Nexus 9516 routed traffic among 4,096 unique networks learned via BGP.

In this test case, test engineers configured Spirent TestCenter to emulate 128 BGP routers, each using a unique Autonomous System Number (ASN). Each Spirent BGP router brought up a peering session with the Cisco Nexus 9516, then advertised a total of 4,096 unique routes. The Spirent instrument then offered fully meshed traffic between all networks learned using BGP.

Concurrently with BGP, the Cisco Nexus 9516 also ran Bidirectional Forwarding Detection (BFD) for rapid fault detection.

It's important to note that the number of routes used in this test is due to a limit in the number of trackable streams supported by the Spirent DX2 test modules. The Cisco Nexus 9516 supports up to 128,000 longest prefix match (LPM) routes, according to the Cisco data sheet, but Network Test did not verify this. Other Spirent test modules also support higher trackable stream counts. With the DX2 module, a higher route count also would have been possible using fewer than 128 ports.

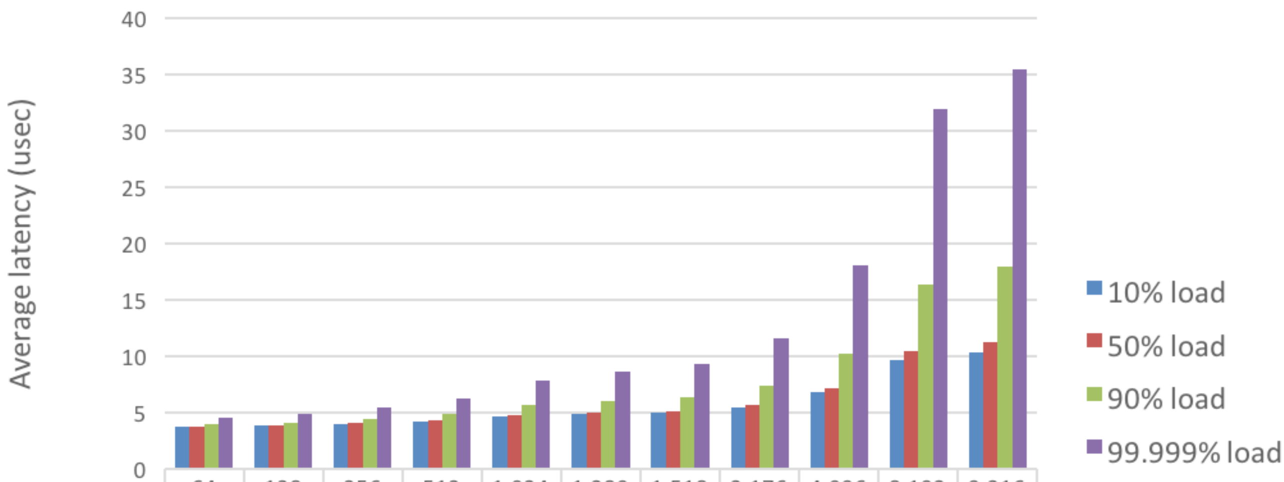
Throughput again was equivalent to 99.999 percent of line rate, regardless of frame size, despite the higher amount of overall traffic (due to a small number of BGP routing messages).

Table 3 presents throughput, latency, and jitter results for all frame sizes.

Figures 8 and 9 compare average and maximum delay measurements, respectively, with offered loads of 10, 50, 90, and 99.999 percent of line rate.

	Throughput			**Latency @ throughput rate**			
Frame size (bytes)	**Frames/s**	**Tbit/s**	**% line rate**	**Minimum (usec)**	**Average (usec)**	**Maximum (usec)**	**Maximum jitter (usec)**
64	19,047,427,571	9.752	99.999%	3.27	4.52	8.01	3.64
128	10,810,702,141	11.070	99.999%	3.32	4.84	8.55	3.54
256	5,797,043,179	11.872	99.999%	3.51	5.40	10.05	4.23
512	3,007,488,568	12.319	99.999%	3.80	6.19	12.26	4.57
1,024	1,532,551,646	12.555	99.999%	4.22	7.78	20.09	6.59
1,280	1,230,756,860	12.603	99.999%	4.40	8.59	23.78	8.21
1,518	1,040,301,637	12.633	99.999%	4.52	9.33	27.27	10.10
2,176	728,590,127	12.683	99.999%	4.97	11.53	37.28	14.00
4,096	388,723,012	12.738	99.999%	6.26	18.03	66.85	26.09
8,192	194,834,866	12.769	99.999%	8.99	31.94	127.78	51.40
9,216	173,233,426	12.772	99.999%	9.68	35.41	143.27	58.17

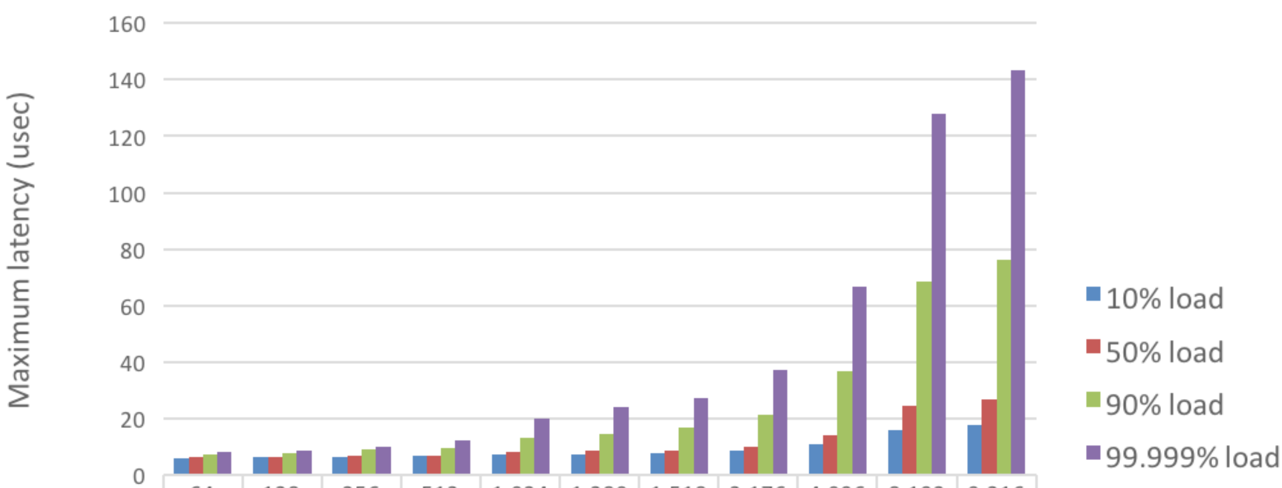
Table 3: IPv4/BGP unicast performance results



	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	3.74	3.81	3.98	4.22	4.66	4.84	4.98	5.44	6.76	9.60	10.31
50% load	3.77	3.84	4.03	4.29	4.78	4.98	5.14	5.66	7.17	10.39	11.20
90% load	3.96	4.11	4.43	4.89	5.69	6.05	6.37	7.34	10.19	16.35	17.90
99.999% load	4.52	4.84	5.40	6.19	7.78	8.59	9.33	11.53	18.03	31.94	35.41

Frame size (bytes)

Figure 8: IPv4/BGP unicast average latency vs. load



	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	6.02	6.06	6.25	6.58	7.10	7.37	7.58	8.64	10.65	15.93	17.40
50% load	6.12	6.28	6.53	6.92	8.30	8.60	8.73	9.73	14.13	24.25	26.78
90% load	7.17	7.60	8.85	9.53	13.20	14.67	16.50	21.39	36.89	68.34	76.39
99.999% load	8.01	8.55	10.05	12.26	20.09	23.78	27.27	37.28	66.85	127.78	143.27

Frame size (bytes)

Figure 9: IPv4/BGP unicast maximum latency vs. load

RFC 5180 IPv6 Unicast Performance

IPv6 performance tests were conceptually similar to the IPv4 tests, again using a fully meshed traffic pattern, and with the Cisco Nexus 9516 acting as a router. Engineers configured each interface to be a member of a different IPv6 subnet, requiring the device to route traffic between subnet. The Spirent test instrument emulated one IPv6 host per subnet.

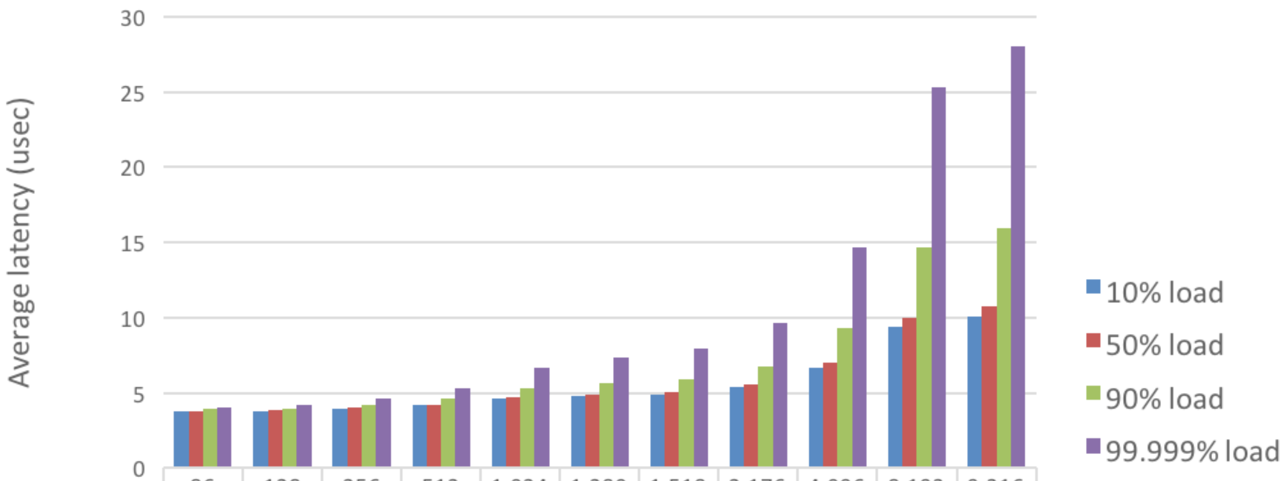
In all IPv6 tests, test engineers configured a minimum frame size of 86 bytes rather than 64 bytes to accommodate the 20-byte "signature field" added by the Spirent test instrument and an 8-byte UDP header (see the Test Methodology section for more details on 86-byte frames). Engineers configured the UDP header with a range of 8,000 source and destination port numbers to ensure more uniform distribution of flows across the switch fabric.

Throughput again was equivalent to 99.999 percent of line rate, regardless of frame size. Table 4 presents throughput, latency, and jitter results for all frame sizes.

Figures 10 and 11 compare average and maximum delay measurements, respectively, with offered loads of 10, 50, 90, and 99.999 percent of line rate.

	Throughput			Latency @ throughput rate			
Frame size (bytes)	**Frames/s**	**Tbit/s**	**% line rate**	**Minimum (usec)**	**Average (usec)**	**Maximum (usec)**	**Maximum jitter (usec)**
86	15,094,190,275	10.385	99.999%	2.24	4.06	6.71	2.30
128	10,810,703,846	11.070	99.999%	2.27	4.17	6.87	2.54
256	5,797,044,091	11.872	99.999%	2.39	4.59	7.72	3.07
512	3,007,489,040	12.319	99.999%	2.49	5.29	9.08	3.74
1,024	1,532,551,886	12.555	99.999%	2.73	6.69	12.15	6.35
1,280	1,230,757,053	12.603	99.999%	2.78	7.35	14.11	7.27
1,518	1,040,301,801	12.633	99.999%	2.84	7.94	15.71	8.87
2,176	728,590,241	12.683	99.999%	3.03	9.66	20.51	12.67
4,096	388,723,073	12.738	99.999%	3.56	14.66	33.03	23.36
8,192	194,834,897	12.769	99.999%	4.71	25.35	61.98	46.92
9,216	173,233,453	12.772	99.999%	4.99	28.03	69.24	52.77

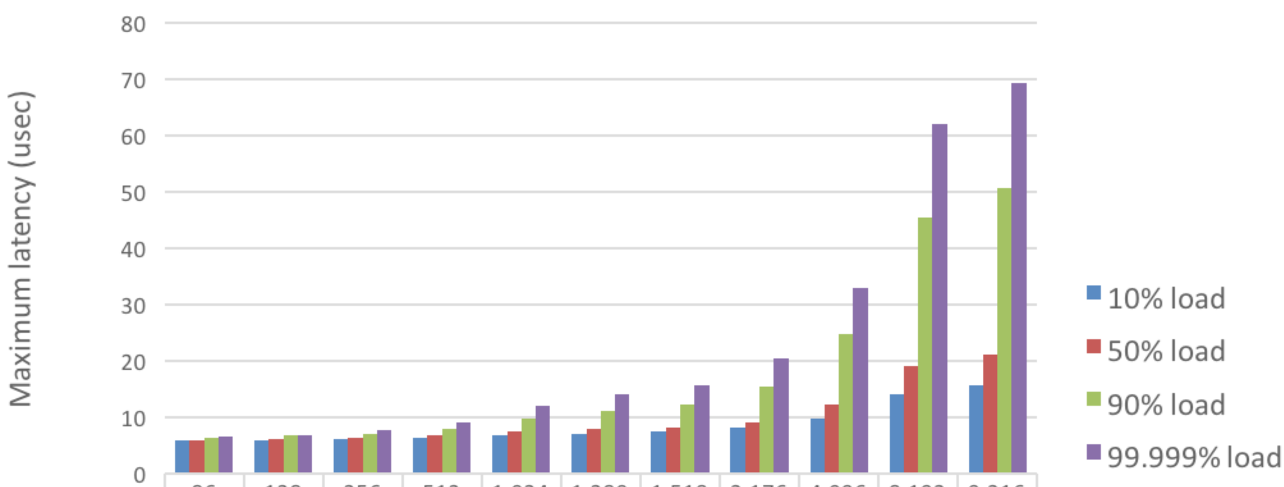
Table 4: IPv6 unicast performance results



	86	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	3.76	3.79	3.96	4.19	4.62	4.79	4.92	5.37	6.64	9.37	10.06
50% load	3.80	3.82	4.00	4.25	4.71	4.90	5.06	5.55	6.98	10.04	10.80
90% load	3.92	3.98	4.23	4.60	5.33	5.65	5.93	6.79	9.29	14.64	15.98
99.999% load	4.06	4.17	4.59	5.29	6.69	7.35	7.94	9.66	14.66	25.35	28.03

Frame size (bytes)

Figure 10: IPv6 unicast average latency vs. load



	86	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	6.02	6.06	6.25	6.46	6.98	7.17	7.59	8.29	9.93	14.19	15.66
50% load	6.09	6.20	6.42	6.85	7.50	8.01	8.25	9.25	12.31	19.10	21.22
90% load	6.49	6.80	7.09	8.10	9.78	11.12	12.24	15.50	24.73	45.51	50.73
99.999% load	6.71	6.87	7.72	9.08	12.15	14.11	15.71	20.51	33.03	61.98	69.24

Frame size (bytes)

Figure 11: IPv6 unicast maximum latency vs. load

RFC 5180 IPv6 Unicast Performance With BGP-MP Routing

This tests adds BGP-Multiprotocol (BGP-MP) routing to the previous IPv6 test, which involved routing only between subnets locally configured on the switch. Here, the Cisco Nexus 9516 routed traffic among 2,048 unique IPv6 networks learned via BGP-MP.

RFC 4760 describes BGP-MP, a set of multiprotocol extensions to BGP for carrying topology information about multiple network-layer protocols, including IPv6. The Spirent test instrument brought up BGP-MP peering sessions, advertised 2,048 unique IPv6 routes, and then offered fully meshed traffic destined to all routes.

Concurrently with BGP-MP, the Cisco Nexus 9516 also ran Bidirectional Forwarding Detection (BFD) for rapid fault detection.

It's important to note that the 2,048 IPv6 routes used in this test is due to a limit in the number of trackable streams supported by the Spirent DX2 test modules. The Cisco Nexus 9516 supports up to 80,000 IPv6 routes, according to Cisco, but Network Test did not verify this. Other Spirent test modules also support higher track-able stream counts. With the DX2 module, a higher route count also would have been possible using fewer than 128 ports.

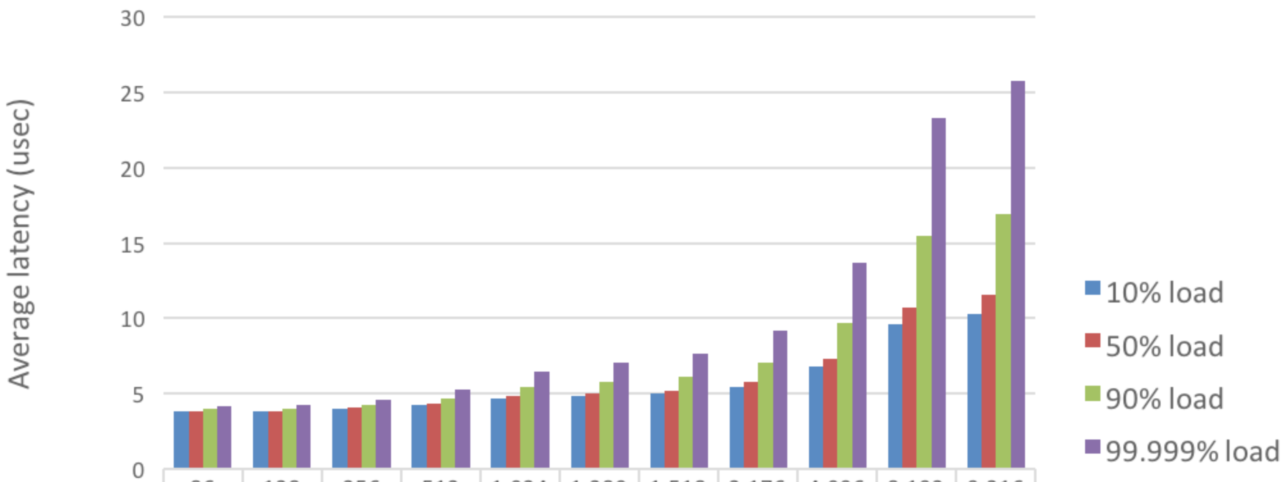
Also, the minimum frame size in these test was 86 bytes, for the same reasons discussed in the "IPv6 Unicast Performance" section.

Throughput again was equivalent to 99.999 percent of line rate, regardless of frame size, despite the higher amount of overall traffic (including the small amount of BGP control-plane traffic).

Table 5 presents throughput, latency, and jitter results for all frame sizes. Figures 12 and 13 compare average and maximum delay measurements, respectively, with offered loads of 10, 50, 90, and 99.999 percent of line rate.

	Throughput			**Latency @ throughput rate**			
Frame size (bytes)	**Frames/s**	**Tbit/s**	**% line rate**	**Minimum (usec)**	**Average (usec)**	**Maximum (usec)**	**Maximum jitter (usec)**
86	15,094,187,638	10.385	99.999%	3.26	4.15	7.38	2.92
128	10,810,701,934	11.070	99.999%	3.29	4.23	7.75	3.11
256	5,797,043,068	11.872	99.999%	3.54	4.60	8.28	3.32
512	3,007,488,510	12.319	99.999%	3.80	5.23	10.32	4.16
1,024	1,532,551,617	12.555	99.999%	4.21	6.49	15.40	4.64
1,280	1,230,756,837	12.603	99.999%	4.39	7.08	18.15	6.04
1,518	1,040,301,618	12.633	99.999%	4.53	7.60	20.66	6.92
2,176	728,590,113	12.683	99.999%	4.98	9.15	27.98	9.38
4,096	388,723,005	12.738	99.999%	6.24	13.68	49.18	18.66
8,192	194,834,862	12.769	99.999%	8.97	23.34	94.74	38.68
9,216	173,233,423	12.772	99.999%	9.65	25.76	106.01	43.53

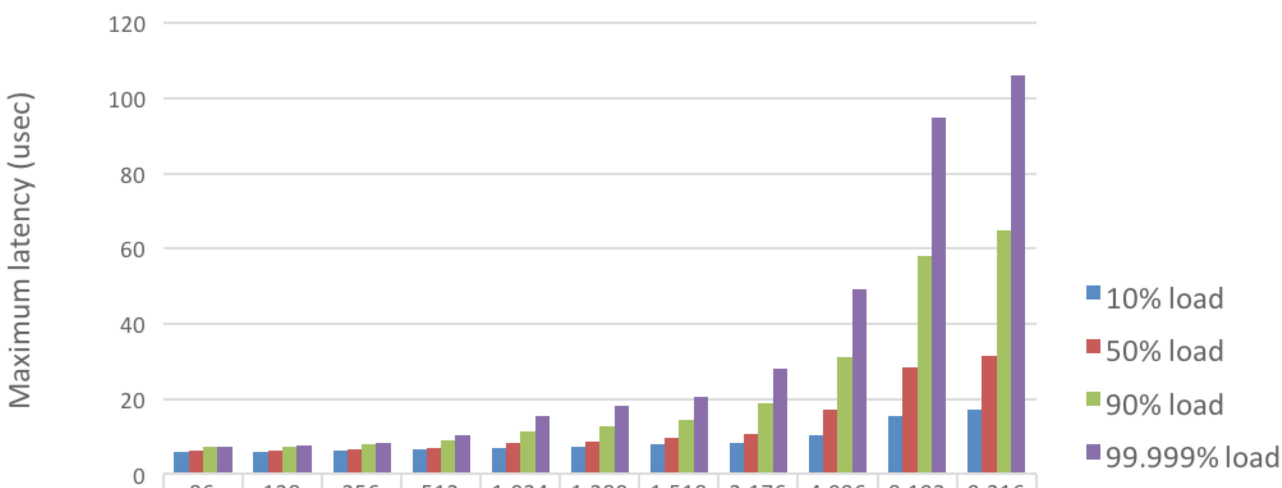
Table 5: IPv6/BGP-MP unicast performance results



	86	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	3.78	3.81	3.98	4.22	4.67	4.84	4.98	5.45	6.77	9.61	10.32
50% load	3.80	3.84	4.03	4.30	4.81	5.02	5.19	5.74	7.32	10.70	11.54
90% load	3.95	3.99	4.24	4.64	5.43	5.79	6.09	7.03	9.72	15.49	16.94
99.999% load	4.15	4.23	4.60	5.23	6.49	7.08	7.60	9.15	13.68	23.34	25.76

Frame size (bytes)

Figure 12: IPv6/BGP-MP unicast average latency vs. load



	86	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	6.05	6.06	6.23	6.54	7.07	7.35	7.86	8.45	10.32	15.61	17.33
50% load	6.23	6.28	6.58	6.96	8.22	8.83	9.83	10.83	17.24	28.51	31.58
90% load	7.24	7.30	8.01	9.04	11.30	12.78	14.32	18.80	31.09	58.14	64.83
99.999% load	7.38	7.75	8.28	10.32	15.40	18.15	20.66	27.98	49.18	94.74	106.01

Frame size (bytes)

Figure 13: IPv6/BGP-MP unicast maximum latency vs. load

RFC 3918 Ethernet Multicast Performance

IP multicast tests stressed both the control and data planes of the Cisco Nexus 9516, scaling up both planes to maximum performance levels.

On the control plane, the Spirent test instrument emulated hosts on 127 subscriber ports, each joining 4,095 IP multicast groups, requiring the Cisco device to maintain an Internet Group Management Protocol (IGMP) snooping table with more than unique 520,000 entries. On the data plane, the Spirent instrument offered traffic in a way that required massive replication on the part of the Cisco switch. The Spirent generator offered traffic destined to all 4,095 IP multicast group addresses on all 127 receiver interfaces. The use of 4,095 multicast groups was due to a stream count limit in the Spirent test modules, and is not a limit of the Cisco Nexus 9516.

In this Layer-2 test, all interfaces on the Cisco switch were members of a single VLAN. The switch used IGMP version 2 (IGMPv2) to build snooping tables.

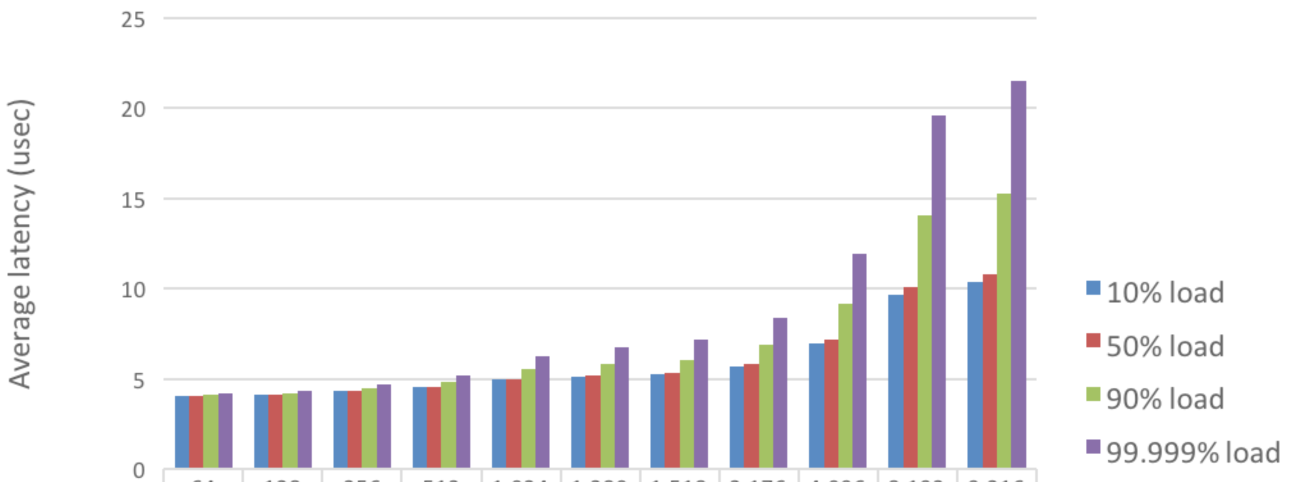
As in all unicast tests, the Cisco Nexus 9516 delivered all traffic with zero frame loss with throughput equivalent to 99.999 percent of line rate.

Table 6 presents throughput, latency, and jitter results for all frame sizes.

Figures 14 and 15 compare average and maximum delay measurements, respectively, with offered loads of 10, 50, 90, and 99.999 percent of line rate.

	Throughput			Latency @ throughput rate			
Frame size (bytes)	**Frames/s**	**Tbit/s**	**% line rate**	**Minimum (usec)**	**Average (usec)**	**Maximum (usec)**	**Maximum jitter (usec)**
64	18,898,622,534	9.676	99.999%	2.81	4.20	6.53	1.91
128	10,726,245,222	10.984	99.999%	2.88	4.35	6.76	2.01
256	5,751,754,684	11.780	99.999%	3.01	4.67	7.21	2.28
512	2,983,993,032	12.222	99.999%	3.15	5.20	8.12	2.68
1,024	1,520,578,825	12.457	99.999%	3.45	6.24	9.56	3.10
1,280	1,221,141,764	12.504	99.999%	3.56	6.72	10.46	3.42
1,518	1,032,174,443	12.535	99.999%	3.66	7.14	11.15	3.71
2,176	722,898,130	12.584	99.999%	3.99	8.37	13.60	4.23
4,096	385,686,174	12.638	99.999%	4.91	11.94	20.84	5.77
8,192	193,312,749	12.669	99.999%	6.87	19.57	35.79	11.50
9,216	171,880,066	12.672	99.999%	7.35	21.48	39.67	12.94

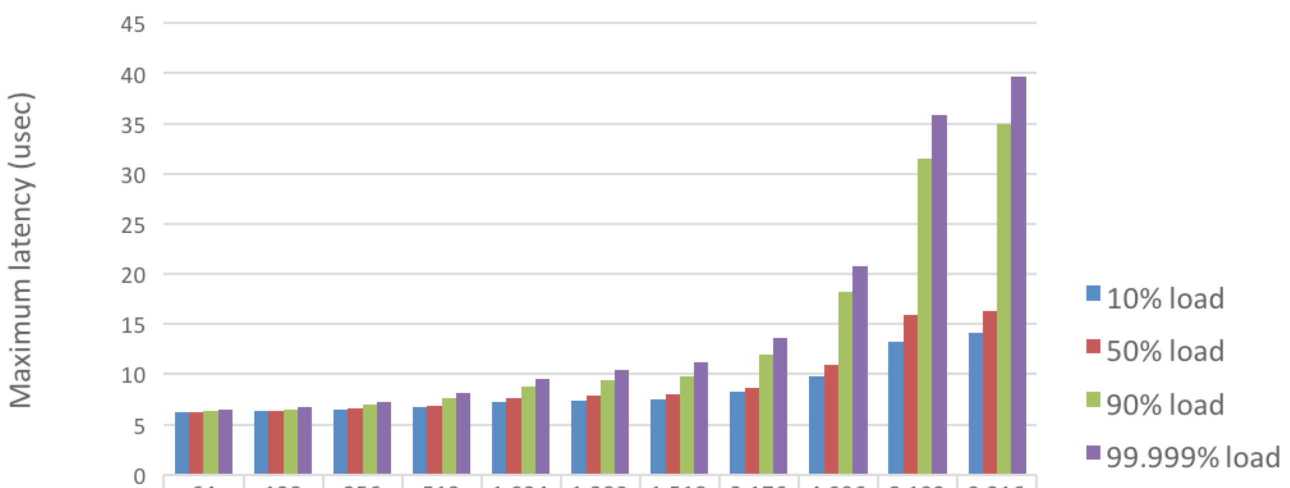
Table 6: Layer-2 multicast performance results



Frame size (bytes)	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	4.05	4.12	4.29	4.51	4.94	5.11	5.25	5.69	6.96	9.68	10.36
50% load	4.04	4.12	4.30	4.54	5.00	5.18	5.32	5.80	7.16	10.08	10.81
90% load	4.08	4.19	4.44	4.79	5.50	5.80	6.06	6.86	9.14	14.04	15.26
99.999% load	4.20	4.35	4.67	5.20	6.24	6.72	7.14	8.37	11.94	19.57	21.48

Frame size (bytes)

Figure 14: Layer-2 multicast average latency vs. load



Frame size (bytes)	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	6.25	6.30	6.50	6.72	7.22	7.42	7.56	8.24	9.82	13.29	14.20
50% load	6.27	6.35	6.55	6.88	7.58	7.82	8.04	8.68	10.93	15.94	16.29
90% load	6.39	6.53	6.96	7.68	8.74	9.39	9.81	11.94	18.19	31.53	34.89
99.999% load	6.53	6.76	7.21	8.12	9.56	10.46	11.15	13.60	20.84	35.79	39.67

Frame size (bytes)

Figure 15: Layer-2 multicast maximum latency vs. load

RFC 3918 IPv4 Multicast Performance

The Layer-3 IP multicast tests again stressed both the control and data planes of the Cisco Nexus 9516, this time moving traffic across subnet boundaries in a test involving more than 520,000 unique multicast routes (mroutes).

In this Layer-3 test, engineers configured each interface on the Cisco switch to reside in a different IPv4 subnet. For Layer-3 IP multicast support, the switch ran the Protocol Independent Multicast-Sparse Mode (PIM) routing protocol as well as IGMPv2 to maintain client snooping tables.

On the control plane, 127 subscriber ports each joined 4,095 IP multicast groups, requiring the Cisco device to maintain a routing table with more than 520,000 unique mroutes. On the data plane, the Spirent traffic generator offered traffic to all IP multicast groups on all 127 receiver interfaces. The use of 4,095 multicast groups was due to a stream count limit in the Spirent test modules, and is not a limit of the Cisco Nexus 9516.

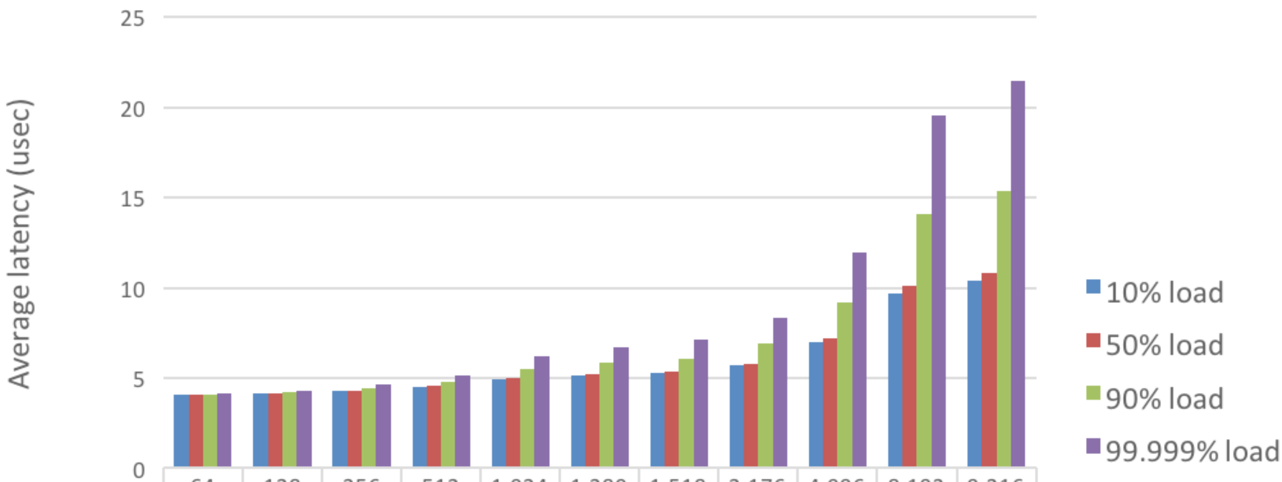
The Cisco Nexus 9516 again delivered all traffic with zero frame loss with throughput equivalent to 99.999 percent of line rate.

Table 7 presents throughput, latency, and jitter results for all frame sizes.

Figures 16 and 17 compare average and maximum delay measurements, respectively, with offered loads of 10, 50, 90, and 99.999 percent of line rate.

	Throughput			Latency @ throughput rate			
Frame size (bytes)	**Frames/s**	**Tbit/s**	**% line rate**	**Minimum (usec)**	**Average (usec)**	**Maximum (usec)**	**Maximum jitter (usec)**
64	18,898,622,534	9.676	99.999%	2.83	4.17	6.43	1.92
128	10,726,245,222	10.984	99.999%	2.88	4.31	6.68	1.99
256	5,751,754,684	11.780	99.999%	3.03	4.65	7.19	2.37
512	2,983,993,032	12.222	99.999%	3.15	5.18	7.99	2.69
1,024	1,520,578,825	12.457	99.999%	3.45	6.22	9.84	3.43
1,280	1,221,141,764	12.504	99.999%	3.56	6.70	10.57	3.64
1,518	1,032,174,443	12.535	99.999%	3.66	7.12	11.01	4.19
2,176	722,898,130	12.584	99.999%	3.99	8.35	13.24	4.76
4,096	385,686,174	12.638	99.999%	4.90	11.92	19.94	6.16
8,192	193,312,749	12.669	99.999%	6.86	19.55	35.08	12.20
9,216	171,880,066	12.672	99.999%	7.34	21.46	38.88	13.73

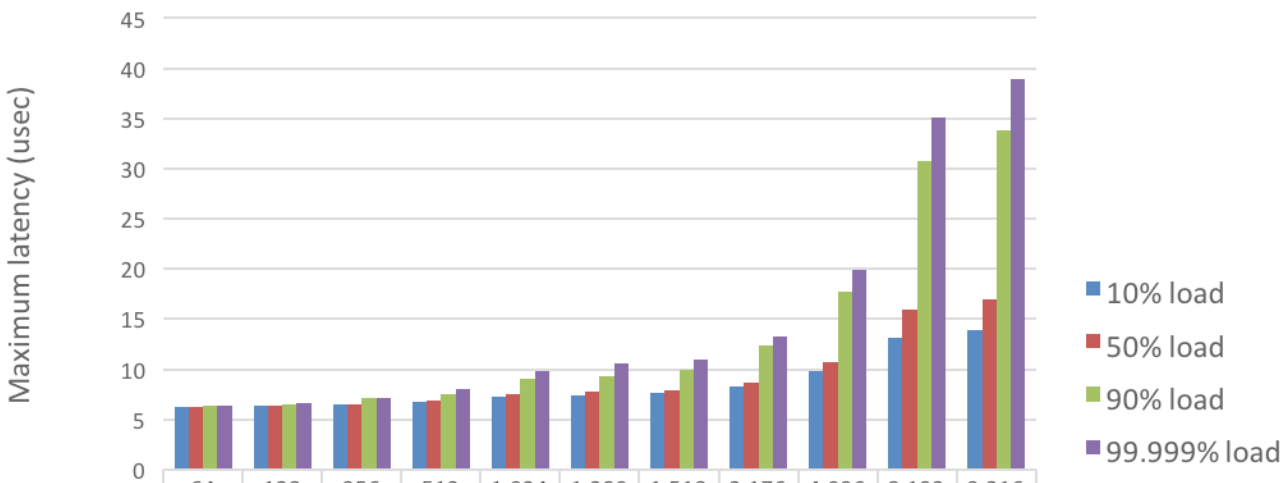
Table 7: Layer-3 multicast performance results



	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	4.06	4.13	4.30	4.52	4.95	5.12	5.26	5.70	6.97	9.69	10.37
50% load	4.05	4.13	4.31	4.55	5.00	5.19	5.33	5.81	7.17	10.09	10.82
90% load	4.09	4.20	4.45	4.81	5.51	5.82	6.08	6.90	9.18	14.10	15.39
99.999% load	4.17	4.31	4.65	5.18	6.22	6.70	7.12	8.35	11.92	19.55	21.46

Frame size (bytes)

Figure 16: Layer-3 multicast average latency vs. load



	64	128	256	512	1,024	1,280	1,518	2,176	4,096	8,192	9,216
10% load	6.28	6.33	6.50	6.73	7.23	7.40	7.62	8.25	9.81	13.13	13.96
50% load	6.30	6.38	6.55	6.87	7.55	7.75	7.93	8.69	10.76	15.96	16.98
90% load	6.43	6.55	7.11	7.51	8.99	9.33	9.96	12.37	17.72	30.75	33.79
99.999% load	6.43	6.68	7.19	7.99	9.84	10.57	11.01	13.24	19.94	35.08	38.88

Frame size (bytes)

Figure 17: Layer-3 multicast maximum latency vs. load

Test Methodology

The principle objective of this test was to characterize the performance of the Cisco Nexus 9516 equipped with 128 100-gigabit Ethernet interfaces in various Layer-2 and Layer-3 configurations. Network Test evaluated the Cisco Nexus 9516 in seven scenarios:

- RFC 2889 Ethernet unicast performance
- RFC 2544 IPv4 unicast performance
- RFC 2544 IPv4 unicast performance with BGP routing
- RFC 5180 IPv6 unicast performance
- RFC 5180 IPv6 unicast performance with BGP-MP routing
- RFC 3918 Ethernet multicast performance
- RFC 3918 IPv4 multicast performance

For all configurations, the performance metrics consisted of RFC 2544 throughput; minimum, average, and maximum latency; and maximum jitter.

The principle test instrument for this project was the Spirent TestCenter traffic generator/analyzer equipped with dX2-100G-P4 CFP2 modules. For unicast tests, the Spirent instrument offered traffic to all 128 ports in a fully meshed pattern, meaning all traffic was destined for all other ports. For multicast tests, the Spirent instrument used the IGMPv2 protocol to subscribe to 4,095 IP multicast group addresses on 127 receiver ports. A single Spirent transmitter port then offered traffic to all IP multicast group addresses.

The Spirent test instrument offered traffic at 99.999 percent of line rate for a duration of 300 seconds, and measured the latency of every frame received. Test engineers used 99.999 percent of line rate, which is 10 parts per million (10 ppm) slower than nominal line rate, to avoid any clocking differences between the traffic generator and the switch under test. The IEEE 802.3 Ethernet specification requires interfaces to tolerate clocking differences of up to +/- 100 ppm, so a 10-ppm difference is well within that specification.

Test engineers repeated this test with 11 frame sizes: 64-, 128-, 256-, 512-, 1,024-, 1,280, 1,518-, 2,176, 4,096, 8,192, and 9,216-byte Ethernet frames. The first seven sizes are recommended in RFC 2544. The 2,176-byte size is common in data centers that use Fibre Channel over Ethernet (FCoE) encapsulation. The larger sizes, while not formally part of the Ethernet specification, are recommended in RFC 5180 section 5.1.1 for jumbo frame tests.

Engineers configured the Spirent instrument to measure latency using the first-in, first-out (FIFO) measurement method described in RFC 1242. FIFO latency measurement is appropriate for cut-through devices such as the Cisco Nexus 9516 that can begin forwarding a frame before the entire frame has been received. (The Cisco device also can operate in store-and-forward mode, but engineers used the default cut-through setting in these tests.)

Engineers also measured switch delay for three loads lower than the throughput rate – at 10, 50, and 90 percent of line rate. RFC 2544 requires latency to measured at, and only at, the throughput rate. Since production networks typically see far lower average utilization, Cisco requested additional tests to be run to characterize delay at lower offered loads.

IPv6 tests used 86-byte instead of 64-byte frames as the minimum frame length. This is due to two requirements. First, the Spirent test instrument embeds a 20-byte “signature field” in every test frame. Second, test engineers configured traffic to use an 8-byte UDP header to ensure optimal distribution of flows across the internal switch fabric. Adding up all the field lengths (18 bytes for Ethernet header and CRC; 40 bytes for IPv6 header; 8 bytes for UDP header; and 20 bytes for Spirent signature field) yields a minimum frame size of 86 bytes.

Some IPv4 and IPv6 unicast tests involved direct routes (1 per port), while others used Border Gateway Protocol (BGP). In the IPv4 tests, each Spirent test interface represented one BGP router advertising reachability to 32 networks, for a total of 4,096 unique IPv4 networks. In the IPv6 tests, each Spirent test interface used BGP-Multiprotocol Extensions (BGP-MP) to advertise 16 networks, for a total of 2,048 unique IPv6 networks. It's important to note that both IPv4 and IPv6 network counts represent limits of the Spirent test interfaces and not of the Cisco Nexus 9516.

In IPv4 BGP and IPv6 BGP-MP tests, the Cisco Nexus 9516 also ran Bidirectional Forwarding Detection (BFD). As described in RFC 5880, BFD detects link or interface faults much faster than the "hello" mechanisms used in many routing protocols. Engineers configured the Cisco Nexus 9516 with desired minimum transmit and receive intervals of 150 milliseconds and a detect multiplier of 3.

The RFC 3918 Ethernet multicast traffic tests involved a traffic pattern with one transmitter port and 127 receiver (subscriber) ports. Here, all 127 receiver ports on the Spirent TestCenter instrument joined the same 4,095 multicast groups using IGMPv2. After the switch's IGMP snooping table was fully populated, the test instrument then offered traffic to the single transmit port, with destination addresses of all 4,095 multicast groups. As in the unicast tests, the instrument measured throughput and latency for 11 frame sizes.

The layer-3 multicast tests used the same traffic pattern as the layer-2 tests, with one transmitter port and 127 receiver (subscriber) ports. In this case, however, all switch ports also ran the protocol independent multicast-sparse mode (PIM-SM) routing protocol. All switch ports used PIM-SM to learn multicast routes. Then, all 127 receiver ports on the Spirent TestCenter instrument joined the same 4,095 multicast groups using IGMPv2. The instrument measured throughput and latency for the same 11 frame sizes as in the other performance tests.

The duration for all tests was 300 seconds. This is more stressful than conventional switch tests that use 30- or 60-second durations, in that switch buffers tend to fill over time. If anything, switch latency and jitter will be lower with shorter tests.

Notably, test engineers did not configure the Spirent test instrument with latency compensation or parts-per-million (PPM) clocking adjustments. These adjustments exist in test instruments to compensate for very specific use cases, but also can be abused. The misadjustment of time measurements in a test instrument for purposes of "improving" test results is generally considered to be an unscrupulous practice.

For reproducibility of these results, it's important to note the contents of test traffic, especially with regard to MAC and IP addresses and UDP port numbers. In the Layer-2 unicast tests, all Spirent emulated hosts used pseudorandom MAC addresses as described in RFC 4814. The Spirent IP addresses were 10.x.x.2/8, where *x* was a value between 1 and 128, incrementing by port number. All Cisco Nexus 9516 interfaces were members of the same VLAN, which was bound to an IPv4 address of 10.1.1.1/8 (though this was not used in this Layer-2 test). The UDP headers used 8,000 unique source and destination ports, each beginning at 4001 and incrementing by 1 up to 12,000. In tests involving BGP and BGP-MP, UDP headers used random source and destination port numbers. The Layer-2 multicast tests used Spirent default MAC addresses and IPv4 addressing of x.1.1.2/8, where *x* was a value between 1 and 128, incrementing by port number.

In the Layer-3 tests, Spirent emulated hosts used Spirent TestCenter default MAC addresses. The IPv4 addresses were x.1.1.2/24, where *x* was a value between 1 and 128, incrementing by port number. In IPv6 tests, the Spirent addresses were 200x::2/64, where x was a hexadecimal value between 0x1 and 0x80. The Cisco addresses were 200x::1/64, where x was a hexadecimal value between 0x1 and 0x80. In tests involving direct IPv4 or IPv6 unicast routes, the UDP headers used 8,000 unique source and destination ports, each beginning at 4001 and incrementing by 1 up to 12,000. In tests involving BGP or BGP-MP routing, UDP headers used random source and destintation port numbers.

Conclusion

"Ultra high performance" was the watchword in this largest-ever test of a Cisco Nexus 9000 Series switch. The Cisco Nexus 9516 never dropped a single frame in rigorous benchmarks covering unicast, multicast, Ethernet, IPv4, IPv6, and BGP traffic, all with traffic flowing at virtual line rate across its 128 100G Ethernet ports.

Latency and jitter also remained low and constant across test cases, a critical finding for time-sensitive applications. Average and maximum delay is lower still in test cases involving traffic at 10, 50, and 90 percent of wire speed, providing a complete picture of how the switch is likely to perform in production settings.

Significantly, the Cisco Nexus 9516 handled all traffic using a combination of CFP2 form-factor transceivers, using both LR10 and SR4 optics.

These results should help assure network professionals that the Cisco Nexus 9516 is a highly capable performer, ready for service at the core of even the very largest data centers.

Appendix A: Jitter Measurements

This section presents maximum jitter measurements across varying offered loads, ranging from 10 percent to 99.999 percent of line rate.

Jitter, or delay variation, is a critical metric for any application sensitive to delay. High maximum jitter can severely degrade the performance of video and video applications, as well as any other application that requires real-time delivery of messages.

Table 8 presents maximum jitter measurements from the RFC 2889 Ethernet unicast performance tests.

	Maximum jitter (usec)			
Frame size (bytes)	**10% load**	**50% load**	**90% load**	**99.999% load**
64	1.90	1.98	2.12	2.20
128	1.86	2.11	2.32	2.42
256	1.87	2.15	2.58	2.89
512	1.94	2.21	3.09	3.37
1,024	2.07	2.51	3.74	4.52
1,280	2.34	2.73	4.74	5.21
1,518	2.49	3.05	5.23	6.04
2,176	2.50	3.57	7.01	8.63
4,096	3.24	5.23	13.15	15.83
8,192	4.71	10.46	26.61	32.21
9,216	5.20	11.75	29.96	36.23

Table 8: Ethernet unicast maximum jitter

Table 9 presents maximum jitter measurements from the RFC 2544 IPv4 unicast performance tests.

	Maximum jitter (usec)			
Frame size (bytes)	**10% load**	**50% load**	**90% load**	**99.999% load**
64	1.89	2.00	2.08	2.25
128	1.88	1.97	2.25	2.54
256	1.94	2.21	2.65	2.84
512	2.01	2.34	3.38	3.58
1,024	2.21	2.69	3.70	4.93
1,280	2.19	2.68	4.51	5.73
1,518	2.28	2.97	5.25	6.52
2,176	2.54	4.04	7.30	9.15
4,096	3.26	7.19	13.58	17.19
8,192	4.81	13.35	26.63	33.67
9,216	5.22	14.98	29.89	37.74

Table 9: IPv4 unicast maximum jitter

Table 10 presents maximum jitter measurements from the RFC 2544 IPv4 unicast performance tests with BGP routing enabled.

	Maximum jitter (usec)			
Frame size (bytes)	**10% load**	**50% load**	**90% load**	**99.999% load**
64	1.87	1.99	2.80	3.64
128	1.85	2.05	3.32	3.54
256	1.87	2.12	3.57	4.23
512	1.92	2.23	4.66	4.57
1,024	2.08	2.45	4.75	6.59
1,280	2.33	3.05	4.98	8.21
1,518	2.24	2.75	6.56	10.10
2,176	2.23	3.28	7.11	14.00
4,096	3.24	4.57	13.31	26.09
8,192	6.19	9.10	26.55	51.40
9,216	6.95	10.23	29.82	58.17

Table 10: IPv4 unicast maximum jitter with BGP routing

Table 11 presents maximum jitter measurements from the RFC 5180 IPv6 unicast performance tests.

	Maximum jitter (usec)			
Frame size (bytes)	**10% load**	**50% load**	**90% load**	**99.999% load**
86	1.91	1.97	2.39	2.30
128	1.86	2.00	2.60	2.54
256	1.88	2.10	2.85	3.07
512	1.90	2.21	3.35	3.74
1,024	2.07	2.53	4.80	6.35
1,280	2.31	2.75	5.89	7.27
1,518	2.25	3.01	7.04	8.87
2,176	2.47	3.26	9.11	12.67
4,096	2.47	5.28	17.13	23.36
8,192	3.40	9.54	32.28	46.92
9,216	3.52	10.75	35.56	52.77

Table 11: IPv6 unicast maximum jitter

Table 12 presents maximum jitter measurements from the RFC 5180 IPv6 unicast performance tests with BGP-MP routing enabled.

	Maximum jitter (usec)			
Frame size (bytes)	**10% load**	**50% load**	**90% load**	**99.999% load**
86	1.90	2.05	2.82	2.92
128	1.85	2.14	2.95	3.11
256	1.87	2.11	2.96	3.32
512	1.92	2.22	3.75	4.16
1,024	2.07	2.67	4.00	4.64
1,280	2.17	2.83	4.46	6.04
1,518	2.25	2.90	4.67	6.92
2,176	2.51	3.16	5.87	9.38
4,096	3.24	5.06	11.00	18.66
8,192	6.18	10.09	21.54	38.68
9,216	6.96	11.33	24.68	43.53

Table 12: IPv6 unicast maximum jitter with BGP-MP routing

Table 13 presents maximum jitter measurements from the RFC 3918 Ethernet multicast tests.

	Maximum jitter (usec)			
Frame size (bytes)	**10% load**	**50% load**	**90% load**	**99.999% load**
64	1.95	1.97	2.04	1.91
128	1.91	1.96	2.02	2.01
256	1.94	2.00	2.27	2.28
512	1.94	2.06	2.42	2.68
1,024	1.99	2.22	3.18	3.10
1,280	2.06	2.39	3.59	3.42
1,518	2.10	2.48	3.88	3.71
2,176	2.23	2.81	4.33	4.23
4,096	2.55	3.19	4.79	5.77
8,192	3.30	5.59	7.59	11.50
9,216	3.50	5.20	8.49	12.94

Table 13: Ethernet multicast maximum jitter

Table 14 presents maximum jitter measurements from the RFC 3918 IPv4 multicast tests.

	Maximum jitter (usec)			
Frame size (bytes)	**10% load**	**50% load**	**90% load**	**99.999% load**
64	1.94	1.97	2.03	1.92
128	1.92	1.93	2.13	1.99
256	1.92	2.03	2.23	2.37
512	1.95	2.07	2.46	2.69
1,024	1.97	2.28	3.57	3.43
1,280	2.04	2.48	3.93	3.64
1,518	2.07	2.54	3.85	4.19
2,176	2.23	2.80	4.51	4.76
4,096	2.59	3.47	6.42	6.16
8,192	3.35	5.61	9.39	12.20
9,216	3.56	6.29	8.70	13.73

Table 14: IPv4 multicast maximum jitter

Appendix B: Software Releases Tested

This appendix describes the software versions used on the test bed. Network Test conducted all benchmarks in May and June 2015 in a Cisco engineering lab in San Jose, California, USA.

Component	Version
Cisco NX-OS	7.0(3)I1(2)
Spirent TestCenter	4.50.5906

About Network Test

Network Test is an independent third-party test lab and engineering services consultancy. Our core competencies are performance, security, and conformance assessment of networking equipment and live networks. Our clients include equipment manufacturers, large enterprises, service providers, industry consortia, and trade publications.

Disclaimer

Network Test Inc. has made every attempt to ensure that all test procedures were conducted with the utmost precision and accuracy, but acknowledges that errors do occur. Network Test Inc. shall not be held liable for damages which may result for the use of information contained in this document. All trademarks mentioned in this document are property of their respective owners.

networktest

Network Test Inc.

31324 Via Colinas, Suite 113 • Westlake Village, CA 91362-6761 • USA
+1-818-889-0011 • http://networktest.com • info@networktest.com