
networktest

Five Pillars:
Assessing the
Cisco Catalyst 4948E
for Data Center Service

August 2010

Contents

Executive Summary.....	3
Features and Manageability	3
Fast Convergence With Flex Link	4
Control Plane Policing	5
SPAN Performance and Capacity	6
Performance and Scalability	8
Throughput and Latency	9
Head-of-Line Blocking	13
Conclusion.....	14
Appendix A: Software Versions Tested.....	15
Appendix B: Disclaimer	15

Executive Summary

Switches suitable for top-of-rack service in today's data centers must move traffic fast, but that's only the beginning. High performance and scalability is only one of five key pillars in the data center. Top-of-rack switches also need strong support in terms of features, resiliency, manageability, and environmental factors.

Cisco Systems commissioned Network Test to assess its new Cisco Catalyst 4948E top-of-rack switch in each of these areas. Although this document devotes the greatest attention to performance and scalability, Network Test also found strong support for each of the five areas considered.

Among the test highlights:

- ✓ Line-rate throughput of up to 131 million frames/second in layer-2 and layer-3 unicast tests, for both IPv4 and IPv6 traffic, across 48 gigabit Ethernet and four 10 gigabit Ethernet ports
- ✓ Average latency as low as 4.68 microseconds, with line-rate traffic across all ports
- ✓ MAC address table supports up to 55,000 dynamically learned addresses
- ✓ OSPF scales to support more than 50,000 routes, with 52 concurrent adjacencies
- ✓ IGMPv3 snooping scales to support 32,767 multicast groups
- ✓ PIM scales to support at least 28,000 multicast routes (mroutes)
- ✓ 48 10/100/1000 ports with up to four 10 gigabit Ethernet uplink ports
- ✓ Fully redundant, hot-swappable components
- ✓ Convergence times of 6.6 milliseconds after link failure using Cisco Flex Link
- ✓ Front-to-back airflow with no blocking of airways
- ✓ Control plane policing successfully protects switch CPU
- ✓ Support for eight concurrent line-rate SPAN sessions

Features and Manageability

The Catalyst 4948E offers a full set of data center top-of-rack switching features in a 1 rack unit (1.75-inch) form factor. The switch offers up to 52 ports, with 48 copper gigabit Ethernet interfaces and four uplink interfaces that accept either gigabit or 10 gigabit Ethernet SFP transceivers. For most measurements described here, Network Test used a 48+4 configuration, with 48 downlink ports and

four 10 gigabit Ethernet uplinks equipped with 10GBase-SR transceivers. The switch is around 30 inches deep, allowing it to fit easily inside most four-post cabinets and racks.

Cisco took the term “front-to-back airflow” literally in designing the Catalyst 4948E. The airflow – a critical consideration in designing data centers with hot and cold aisles for maximum cooling efficiency – makes use of the 4948E’s perfectly rectangular shape. Since the only air intake is on the front panel of the switch, it cannot be obstructed by placing another switch or other device directly on top of the Catalyst 4948E, where top vents could be blocked.

The Catalyst 4948E offers redundant, hot-swappable components such as power supplies and fan trays (something Network Test verified by pulling each component during performance tests). In addition, the Catalyst 4948E supports major loop prevention and failover protocols such as IEEE 802.1D spanning tree (STP); IEEE 802.1w rapid spanning tree (RSTP); and virtually all IP-based routing protocols for layer-3 configurations, both for IPv4 and IPv6.

Fast Convergence With Flex Link

Although spanning tree is widely used to protect against loops and network failures, it carries a performance penalty: Convergence times following a failure can last up to 45-60 seconds with standard spanning tree, or typically 1-3 seconds with rapid spanning tree. Given that the threshold where application performance can suffer is often measured in *milliseconds*, these convergence times may be too high to help avoid degraded performance.

Cisco’s Flex Link technology aims to provide link redundancy with much faster convergence times than either STP or RSTP. As an alternative to spanning tree, Flex Link works at layer 2, with one switch port acting as backup for another.

To verify Flex Link functionality and measure convergence time, Network Test and Cisco engineers constructed a test bed with four Catalyst 4948E switches. As shown in Figure 1 below, the switches used Flex Link instead of STP across redundant paths. Engineers then configured a Spirent TestCenter traffic generator/analyzer to offer traffic at a rate of 1 million frames per second between two emulated hosts; thus, each dropped frame would correlate to 1 microsecond of convergence time.

Initially, test traffic flowed across link 1 as shown in the figure. Engineers then administratively shut down one link on the test bed, forcing Flex Link to redirect traffic over the backup link, labeled link 2 in the figure. Finally, Network Test determined convergence time by measuring frame loss.

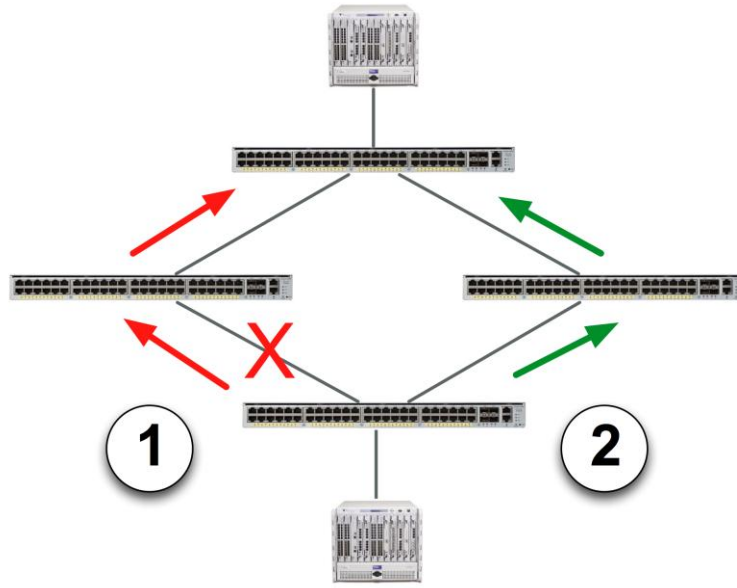


Figure 1: Cisco Catalyst 4948E Flex Link Test Bed

In five trials, Catalyst 4948E switches using Flex Link converged in an average of 6.6 milliseconds after a link failure. That is approximately 150 times faster than a best-case scenario with rapid spanning tree, and more than 500,000 times faster than a worst-case scenario using spanning tree. Clearly, Flex Link offers superior convergence times compared with STP and RSTP.

Control Plane Policing

For a switch to remain in service, its control-plane CPU must always have enough processing cycles available to handle incoming requests. A switch CPU faces any number of risks: An attacker can spray a device with malformed packets. A failure elsewhere in the network can cause a flood of MAC addresses, requiring the device to repopulate its address table. A newly attached subnet or service can bring a sudden influx of IGMP join messages. In all these scenarios, the switch CPU is potentially vulnerable; utilization can rise to near 100 percent, leaving the switch unable to handle any new requests and potentially leading to a loss of connectivity.

The **control plane policing** feature of the Catalyst 4948E offers a safeguard against CPU overload. By configuring a maximum rate at which the control plane will accept traffic, network managers can ensure the CPU in each Catalyst 4948E will remain available to service new and existing flows.

To validate the effectiveness of control plane policing, Network Test used “before” and “after” tests involving a mix of benign OSPF and unauthorized multicast traffic. In the “before” test, control plane policing was disabled on the Catalyst 4948E. Test engineers then established 52 OSPF adjacencies using Spirent TestCenter, one per switch port, and then configured the test tool to offer multicast traffic at

line rate to the reserved all-hosts address (224.0.0.1). Since no receivers previously had subscribed to any multicast group, the switch forwarded all multicast packets to the CPU, in turn causing the loss of all 52 OSPF adjacencies. The IOS `show process cpu` command indicated the switch’s CPU was 99 percent utilized.

The “after” test involved the same routing and traffic parameters, but with control plane policing enabled. This time, there was no change in routing state; all 52 adjacencies remained fully formed. The IOS command line reported switch CPU utilization at just 8 percent, compared with 99 percent without control plane policing.

Table 1 below summarizes results from the control plane policing test.

Test case	Surviving OSPF adjacencies	CPU utilization
Control plane policing disabled	0	99%
Control plane policing enabled	52	8%

Table 1: Cisco Catalyst 4948E Control Plane Policing

SPAN Performance and Capacity

Mirroring is a key capability when it comes to switch management. Copying all traffic to a destination switch port for analysis can be invaluable in troubleshooting and capacity planning, but mirroring has a couple of caveats.

First, a switch’s SPAN (switched port analyzer) performance must be characterized to determine whether it can forward all frames when mirroring traffic. A switch without line-rate mirroring capability may drop frames, leaving network engineers without key information needed to solve a given problem.

Second, in both campus and data center contexts it is often desirable to configure multiple SPAN instances, for example when multiple teams work on separate issues. Here, the number of concurrent SPAN sessions supported becomes a significant question.

The Catalyst 4948E supports up to eight concurrent SPAN sessions using any combination of gigabit Ethernet and 10 gigabit Ethernet ports. To validate SPAN performance and capacity, test engineers configured eight concurrent SPAN instances using the IOS `monitor session` command. One of these sessions mirrored traffic offered to a 10 gigabit Ethernet port; the remaining sessions monitored traffic on gigabit Ethernet ports.

Engineers then configured the Spirent TestCenter instrument to offer a known quantity 64-byte frames at line rate to each monitored switch port, and to capture traffic on each SPAN port¹. The Catalyst 4948E mirrored traffic successfully to all eight SPAN ports, with zero frame loss.

Table 2 below summarizes results from the SPAN performance and capacity tests.

SPAN port instance	Offered load (fps)	Frames dropped
1	14,880,952	0
2	1,488,095	0
3	1,488,095	0
4	1,488,095	0
5	1,488,095	0
6	1,488,095	0
7	1,488,095	0
8	1,488,095	0

Table 2: Cisco Catalyst 4948E SPAN Performance and Capacity

¹ The use of a hardware-based capture tool is significant here. Software-based analyzers cannot capture all frames at gigabit Ethernet line rates, let alone at 10 gigabit Ethernet rates. A hardware-based capture capability is a must when analyzing line-rate traffic on high-speed networks.

Performance and Scalability

Mention “switch performance” to many network professionals, and the image that inevitably comes to mind is frames moving fast through a device. Certainly that’s important, but high throughput is far from the only metric that matters in data center switching. For some applications, latency – the time it takes each frame to cross the switch – is even more important than throughput. And all switches need to scale performance on the control plane as well as the data plane.

Control-plane scalability is a key aspect of performance in the data center. To that end, our performance tests, while primarily focused on the data plane, also loaded up the control plane in a number of interesting ways:

- **Access control lists (ACLs)** have long been considered a security best practice. ACLs require the switch to inspect every frame before deciding whether to forward it. Unless otherwise noted, **the Catalyst 4948E used a 32,292-line ACL in all tests** (621 ACL conditions on each of 52 ports). The ACL consisted of 620 “deny” rules followed by a single “permit” rule.
- Dynamic routing protocols such as EIGRP and OSPF require substantial **routing table capacity**. In OSPF routing tests, **the Catalyst 4948E routed traffic between hosts on 52,000 inter-area networks learned using 52 concurrent OSPF adjacencies**. Even the world’s largest service providers typically advertise far less than 50,000 inter-area routes on their internal networks.
- In multicast scenarios, **IGMP snooping** is needed to switch traffic, and a **large multicast route (mroute) capacity** may be needed to route traffic. In multicast performance tests, **the Catalyst 4948E performed IGMPv3 snooping on 28,000 multicast groups**. In layer-3 tests, **the Catalyst 4948E forwarded traffic to 28,000 multicast routes (mroutes)**.
- **IGMPv3 snooping capacity** can scale even higher when external devices function as Protocol Independent Multicast (PIM) routers, as they typically would in a data center deployment. In a test of multicast group capacity, **the Catalyst 4948E successfully forwarded traffic to 32,767 IGMPv3 groups**.
- Virtualization in data centers requires enormous **MAC address capacity**, with broadcast domains stretching into the thousands or tens of thousands of addresses. In a test of address caching capacity, **the Catalyst 4948E forwarded traffic to 55,000 MAC addresses without flooding**.

Test engineers devised these highly scaled control-plane parameters not to represent some “real-world” network condition, but rather to demonstrate high performance even at the limits of switch performance. As the following discussion of data-plane performance will show, the switch exhibited line-rate throughput and low latency across all tests. Thus, **there is no performance cost to scaling control-plane parameters to arbitrarily high levels**.

Throughput and Latency

As defined in [RFC 1242](#), **throughput** describes the highest rate at which a device can forward traffic with zero frame loss. The zero-loss criterion is especially important in high-speed data centers, where even a single dropped frame can have an adverse impact on application performance and business operations.

The same RFC also defines **latency**, the time it takes the switch to forward each frame. For delay-sensitive applications such as video and voice, latency is an even more important metric than throughput. Latency is also a key concern in some vertical industries, such as financial services. Since latency is cumulative across switches, the cost of even a little added delay on every switch and router can have a direct impact on a company's revenue-bearing application traffic.

Significantly, [RFC 2544](#) (the methodology companion to RFC 1242) requires latency to be measured *at the throughput rate*. Although virtually all switches exhibit lower delay given less stressful offered loads, testing in that manner does not comply either with the letter or the spirit of the industry-standard IETF benchmarks. The tests described here measure latency at, and only at, the throughput rate.

Also in keeping with the IETF benchmarks' notion of testing under the most stressful possible conditions, all unicast tests described here use "fully meshed" traffic patterns, where traffic offered to each switch port is destined to all other switch ports. A switch may exhibit lower latency and/or higher throughput using a less stressful pattern, such as running traffic between pairs of ports, but this is somewhat analogous to driving a car but never turning left or right.

Because a switch buyer has a reasonable expectation of sending traffic to and from any arbitrary set of ports, fully meshed patterns are used here. Again, while it's possible to craft a less stressful test that will give a given switch a "better" result, only testing under the most stressful condition will adequately describe switch limits for all users under any condition.

The actual patterns in unicast testing involved two full meshes, one apiece for the gigabit Ethernet and 10 gigabit Ethernet ports, creating the heaviest possible load on the switch. For multicast testing, test engineers also divided gigabit Ethernet and 10 gigabit Ethernet ports into separate sets, each with one transmit port and all remaining ports subscribing to all multicast groups. Here again, this creates the most stressful load on the switch.

Moreover, all tests were run for a 300-second duration rather than the 30- or 60-second trials often used in switch testing; this too is more stressful and better reflects the long-lived flows found in some data-center applications.

Network Test measured unicast throughput and latency for five different configurations:

- Layer-2 switching without ACLs
- Layer-2 switching with a 32,292-line ACL (to determine what cost, if any, the ACL would have)
- Layer-3 IPv4 static routing with a 32,292-line ACL
- Layer-3 IPv4 OSPF (52,000 routes, 52 adjacencies) with a 32,292-line ACL
- Layer-3 IPv6 static routing with a 32,292-line ACL

All five test cases involved traffic offered at line rate from a Spirent TestCenter traffic generator/analyzer. **In all cases, the Catalyst 4948E forwarded traffic at line rate.**

Figure 2 below summarizes results from unicast throughput tests, including a comparison with the theoretical maximum rate for each frame length.

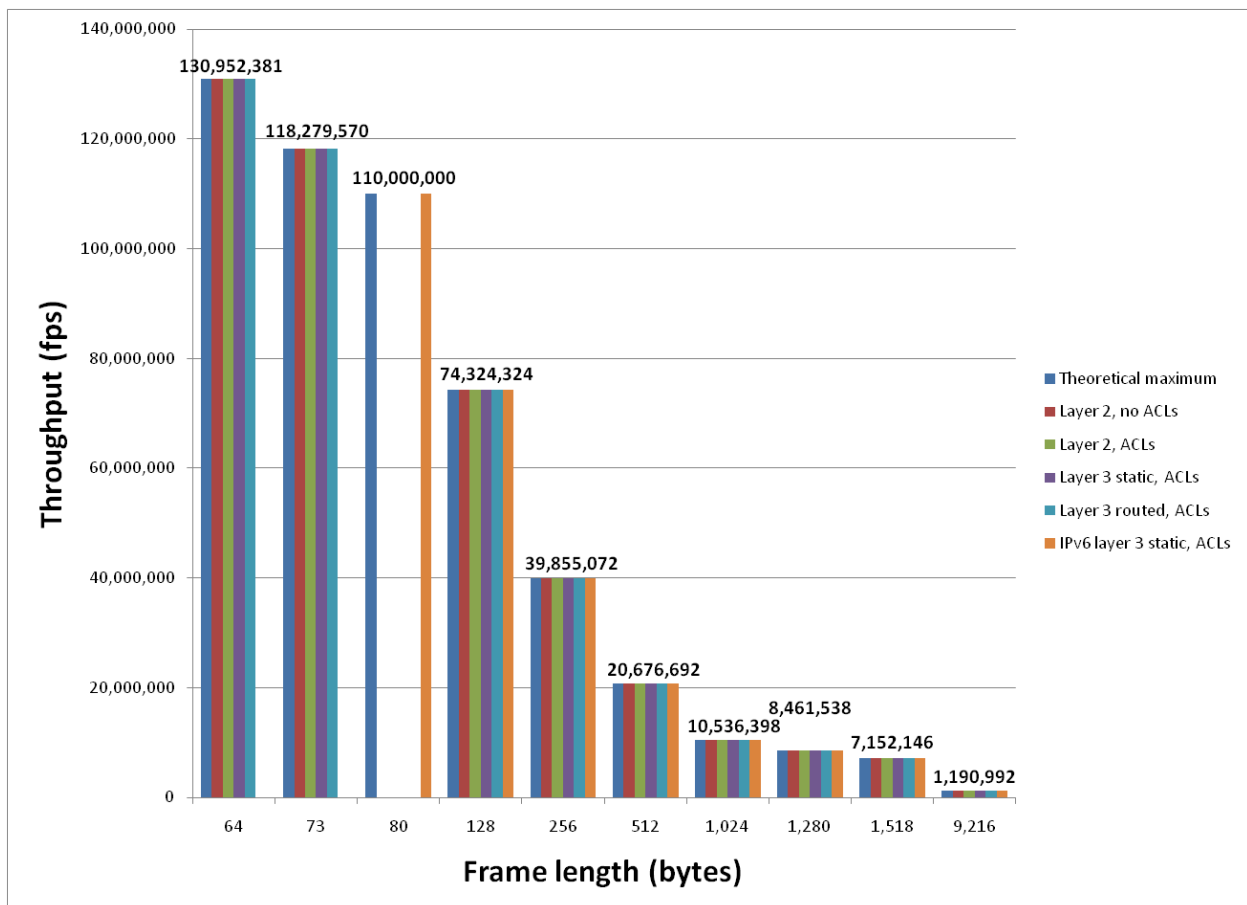


Figure 2: Cisco Catalyst 4948E Throughput

The throughput and latency testing includes a few frame sizes not required by the IETF benchmarking specifications. Test engineers used 73-byte frames to show the switch handling an odd frame length; 80-

byte frames to show IPv6 throughput (see below); and 9,216-byte frames to show jumbo frame handling, a non-standard but nonetheless important consideration in the data center.

Throughput was at line rate in all cases shown, meaning the switch never dropped a frame. There was one exception, not shown, in IPv6 testing. With 78-byte frames (the minimum length supported for IPv6 testing), the Catalyst 4948E drops 0.11 percent of offered traffic. With 80-byte frames, as shown here, the Catalyst 4948E routes traffic at 110 million frames per second with zero loss.

As a rule, average latency scaled linearly with frame length. Linear scaling of latency is an important attribute in a store-and-forward device such as the Catalyst 4948E, since it shows switch buffers do not become backed up regardless of the number of bytes the switch handles.

Table 3 below summarizes unicast average latency measurements.

Frame length (bytes)	Layer 2, no ACLs	Layer 2, ACLs	Layer 3 static, ACLs	Layer 3 routed, ACLs	Layer 3 IPv6 static, ACLs
64	4.68	4.68	4.67	4.88	NA
73	5.48	5.48	5.26	5.94	NA
80	NA	NA	NA	NA	4.92
128	5.54	5.54	5.51	5.55	5.52
256	6.64	6.64	6.57	7.10	6.59
512	9.63	9.63	9.49	9.34	9.52
1,024	15.55	15.55	15.28	15.10	15.35
1,280	18.49	18.49	18.19	35.56	18.25
1,518	21.29	21.29	21.00	20.84	21.10
9,216	117.64	117.64	118.15	118.40	117.28

Table 3: Cisco Catalyst 4948E Unicast Average Latency

Line-rate throughput was also the rule in multicast testing. To assess the Catalyst 4948E's multicast forwarding capabilities, Network Test used three configurations:

- Layer-2 switching without ACLs
- Layer-2 switching with a 32,292-line ACL (to determine what cost, if any, the ACL would have)
- Layer-3 IPv4 PIM routing with a 32,292-line ACL

As noted, all three test cases involved one gigabit Ethernet and one 10 gigabit Ethernet transmitter port, with all remaining ports using IGMPv3 messages to join 28,000 multicast groups. In both layer-2 and layer-3 tests, the Catalyst 4948E acted as a PIM router. When an external router is present, the switch can scale multicast snooping capacity still higher, to 32,767 groups.

Figure 3 below summarizes throughput results. As in unicast testing, the switch handles multicast traffic at line rate in all tests with zero frame loss. In all cases, measured throughput is identical to the theoretical maximum.

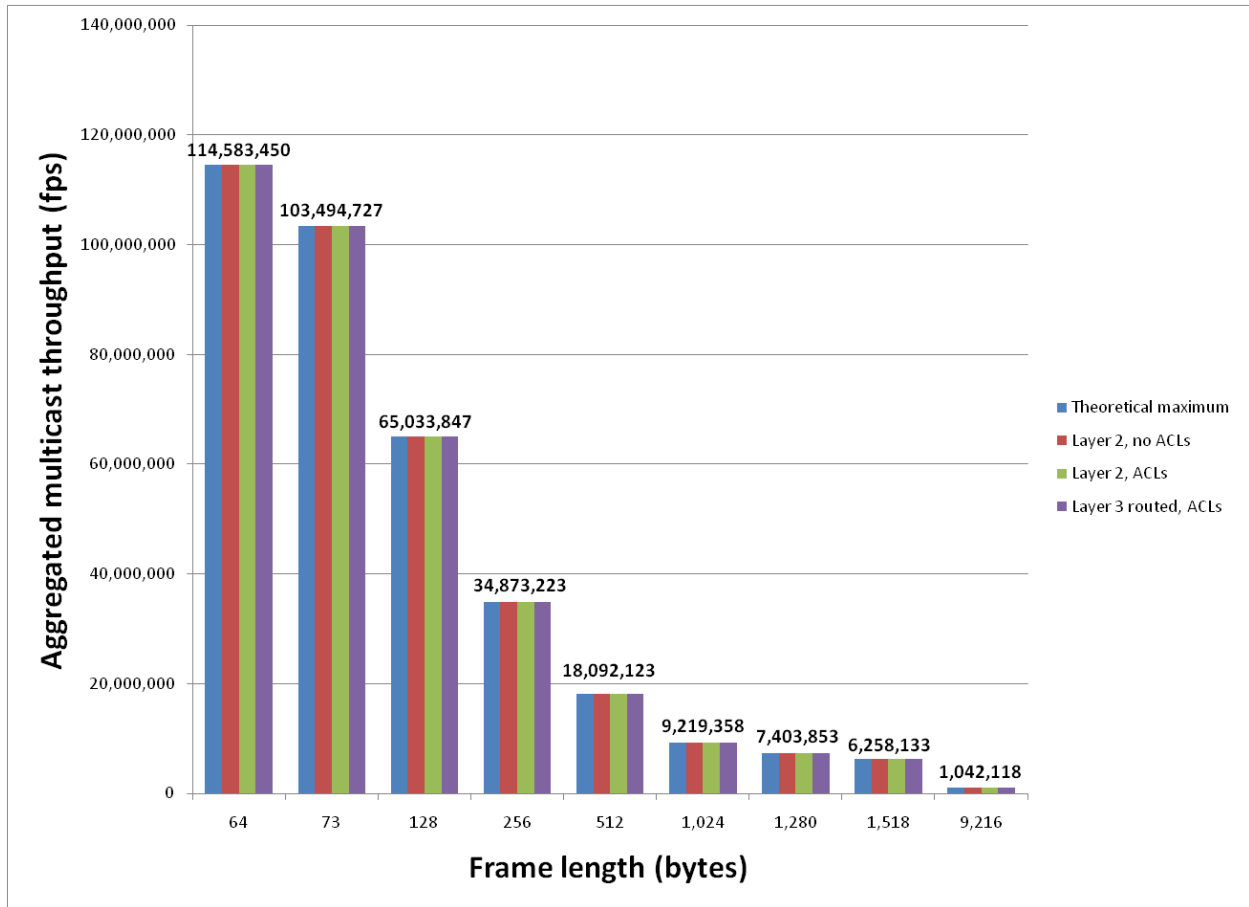


Figure 3: Cisco Catalyst 4948E Multicast Throughput

Average latency scales linearly with frame length for multicast traffic, just as with unicast traffic. Indeed in some cases, average latency is lower for multicast traffic than for unicast, even though the switch has to replicate each incoming frame up to 47 times.

Table 4 below summarizes multicast average latency measurements.

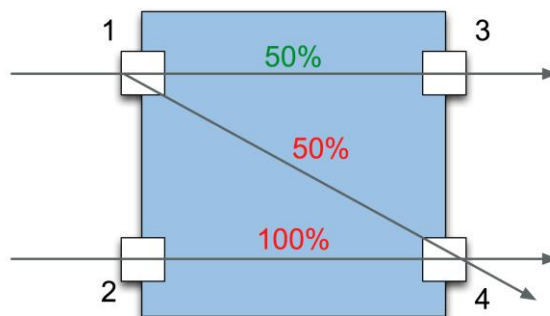
Frame length (bytes)	Layer 2, no ACLs	Layer 2, ACLs	Layer 3 routed, ACLs
64	6.26	6.26	6.23
73	6.39	6.39	6.35
128	6.90	6.90	6.85
256	7.69	7.69	7.57
512	9.90	9.90	9.69
1,024	14.31	14.31	13.93
1,280	16.54	16.54	16.07
1,518	18.59	18.59	18.05
9,216	85.35	85.35	82.59

Table 4: Cisco Catalyst 4948E Multicast Average Latency

Head-of-Line Blocking

A head-of-line blocking (HOLB) problem exists when congestion on one switch port leads to frame loss on another, uncongested port. Thus far, all the testing we've discussed has involved traffic patterns carefully constructed so as not to congest the switch. In contrast, HOLB testing deliberately creates congestion, as often occurs in production networks when multiple frames arrive at a single egress port at the same instant.

As described in [RFC 2889](#) on Ethernet switch testing, the HOLB pattern involves four switch ports:



Here, the traffic generator offers frames at line rate to ports 1 and 2. Since port 4 receives half the frames offered to port 1 and all the frames offered to port 2, a 150 percent overload exists on that port. The central question in HOLB testing is whether congestion on port 4 leads to frame loss on the uncongested port 3.

The Catalyst 4948E architecture uses different application-specific integrated circuits (ASICs) and different ASIC “stubs” for different sets of gigabit and 10 gigabit Ethernet ports. To check for HOLB conditions in all possible cases, Network Test ran the HOLB test using four configurations:

THE CISCO CATALYST 4948E: FIVE PILLARS

- Across all four 10 gigabit Ethernet ports
- Across gigabit Ethernet ports on the same ASIC stub
- Across gigabit Ethernet ports with transmit and receive ports on different ASIC stubs
- Across gigabit Ethernet ports with all ports on different ASIC stubs

For each configuration, Network Test checked for HOLB congestion using 64-, 1,518-, and 9,216-byte frames. **In all configurations with all frame lengths, the Catalyst 4948E did not exhibit head-of-line blocking behavior.**

This test required a configuration change when using jumbo frames. By default, the Catalyst 4948E allocates a relatively large (17-Mbyte) shared memory buffer across egress ports. This default setting can lead to HOLB conditions with jumbo frames. By using a policy-map to limit egress queuing to 16 frames, no HOLB condition exists with jumbo frames. No HOLB condition exists with standard lengths between 64 and 1,518 bytes, even with the default memory allocation.

Conclusion

Cisco commissioned Network Test to validate the performance, scalability, and features functionality of its Cisco Catalyst 4948E. Network Test found an extensive feature set, both for layer-2 switching and layer-3 routing (and then both for IPv4 and IPv6); a solid set of “environmentals” such as front-to-back airflow and configuration options for 10/100/1000/10000 connectivity; and strong performance across a rigorous battery of IPv4 and IPv6 unicast and multicast performance tests. Moreover, most performance tests also involved very high levels of control-plane scalability. By building on five pillars— high performance and scalability, features, resiliency, manageability, and environmental factors – the Cisco Catalyst 4948E offers a strong, credible choice for top-of-rack service in the data center.

Appendix A: Software Versions Tested

This appendix lists the software versions used on Cisco and Spirent equipment on the test bed.

Cisco Catalyst 4948E: IOS 12.2(54)SG

Spirent TestCenter: 3.47.0346

Appendix B: Disclaimer

Version 2010080500. 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.



Network Test Inc.
31324 Via Colinas, Suite 113
Westlake Village, CA 91362-6761
USA
+1-818-889-0011
<http://networktest.com>