Denser, Smaller, Faster:

Assessing Cisco's Catalyst 4948-10GE for High-Performance Data Centers

prepared for Cisco Systems

April 2005

C O N T E N T S

I L L U S T R A T I O N S

Executive Summary

In data center network design, ensuring high throughput and low latency are *the* key requirements. As organizations deploy servers with low-cost gigabit Ethernet adapters and high-performance computing applications such as parallel processing and clustering, the performance demands on network infrastructure have never been greater.

Cisco Systems is meeting these demands with the Catalyst 4948-10GE, the highest-performing 1-rack-unit switch/router the company has ever shipped. This aggregation switch is well-suited to data center service with its 136-Gbit/s switch fabric, two 10-gigabit Ethernet uplink ports, and 48 10/100/1000-Mbit/s copper edge ports.

Cisco commissioned Opus One, an independent networking consultancy, to assess the Catalyst 4948-10GE in terms of performance and features. We validated Cisco's performance claims in several areas, including overall system throughput and latency; layer-2 and layer-3 backbone performance; OSPF performance; and IP multicast handling.

Among the key findings of the Opus One tests:

- Overall system throughput is nearly 101 million packets per second, by far the highest level ever measured by Opus One in a 1U Cisco device
- Average latency for 10-gigabit Ethernet interfaces is 3.8 microseconds, at least one order of magnitude below the point where even the most time-sensitive application would be affected
- Line-rate throughput when moving traffic in a layer-2 backbone topology from 10-gigabit uplinks to gigabit Ethernet edge interfaces
- Line-rate throughput when moving traffic in a layer-3 backbone topology from 10-gigabit uplinks to gigabit Ethernet edge interfaces
- Line-rate throughput with OSPF when routing traffic to more than 2 million hosts
- Identical latency in tests lasting 60 seconds or 600 seconds
- Identical latency when routing traffic to 8,192 hosts or more than 2 million hosts
- Line-rate throughput when moving IP multicast traffic in a backbone topology from 10-gigabit uplinks to gigabit Ethernet edge interfaces over 1,000 mroutes

In addition to these performance highlights, the Catalyst 4948-10GE also demonstrated a number of features aimed at maximizing uptime and reducing cost of ownership. These are described in the next section, "Introducing the Catalyst 4948-10GE."

Introducing the Catalyst 4948-10GE

The Cisco Catalyst 4948-10GE is a 1-rack-unit (1.75-inch) switch/router equipped with two 10-gigabit Ethernet ports and 48 10/100/1000-Mbit/s Ethernet ports. Built around a 136-Gbit/s switch fabric embedded in a single ASIC, the Catalyst 4948-10GE represents the densest, highest-performing 1U switch Cisco has ever produced.

While high performance is a key feature in the Catalyst 4948-10GE, the device also offers numerous feature innovations in the areas of power consumption, device resiliency, and even transceiver type.

A Green Switch

Although the Catalyst 4948-10GE is equipped with dual 300-watt power supplies, it actually consumes far less power, even under heavy load.

Using the "show power" command, the Catalyst 4948-10GE reports power consumption of only 200 watts, even when forwarding traffic on all ports. If anything, Cisco's engineers say this reading is rounded up, and that actual power draw is 170 watts at most times. Opus One did not attempt to validate this, although we did validate the 200-watt reading from the switch:

```
4948-10GE#sh power
Power                                            Fan      Inline
Supply  Model No          Type       Status       Sensor   Status
------  ----------------  ---------  -----------  -------  -------
PS1     PWR-C49-300AC     AC 300W    good         good     n.a.
PS2     PWR-C49-300AC     AC 300W    good         good     n.a.

Power supplies needed by system    : 1
Power supplies currently available : 2

Power Summary                    Maximum
 (in Watts)              Used     Available
----------------------   ----     ---------
System Power (12V)        200         300
Inline Power (-50V)         0           0
Backplane Power (3.3V)      0           0
----------------------   ----     ---------
Total                     200         300
```

The Catalyst 4948-10GE's power supplies are hot-swappable with 1+1 power redundancy. We disconnected one power supply during testing with no impact on device performance. The switch will issue a warning upon loss of a power supply. We also disconnected the fan tray. If two or more fans fail, users have approximately 5 minutes to restore the fan tray before the system automatically shuts down.

Downtime Avoidance

Just as important as performance is maintaining uptime and network resiliency. Given the high value produced by networks, preventing downtime is an absolute must.

The Catalyst 4948-10GE helps maintain maximum uptime by ensuring the switch will always be able to reboot, even if all system images are accidentally deleted or corrupted. Cisco's TACs (technical assistance centers) have fielded calls from customers who have loaded an incorrect system image in memory, or deleted all images. One remedy in such situations is to reload a system image via a switch's console port, but this can take hours.

In contrast, the Catalyst 4948-10GE can rapidly recover from such situations using a special management port that becomes active when no boot image is present. An IP address can be assigned to this port and the switch may then be booted with an image stored on a TFTP server.

Cisco asked Opus One to verify the Catalyst 4948-10G's rapid recovery capability. An example is given below.

```
rommon 1 >b

 boot: can not determine first file name on device "bootflash:"

rommon 2 >set interface fa1 172.20.80.87 255.255.255.0

rommon 3 >set ip route default 172.20.80.1

rommon 4 >ping 171.69.1.129

 Host 171.69.1.129  is alive

rommon 5 >boot tftp://171.69.1.129/rupa/cat4000-i5s-mz.122_25_EWA_146
```

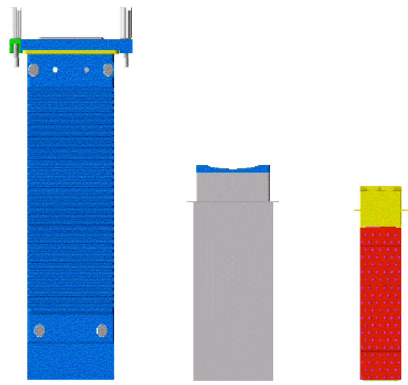
In this example, the switch did not respond to the boot command "b" (prompt 1) because we had deleted all system images. At prompts 2 and 3, we configured the management port with the necessary IP address, netmask, and gateway. At prompt 4, we verified reachability of a TFTP server. Finally, at prompt 5, we booted the system using an image downloaded from the TFTP server.

This downtime avoidance feature not only ensures maximum resiliency for the switch, but also greatly reduces recovery time. The other method of restoring a system image to the switch, transferring the image via console port at 9600 bits per second, literally can take hours to complete.

In contrast, the Catalyst 4948-10GE's management port is a gigabit Ethernet port. We completed the transfer above in just a few seconds.

X2 Transceivers for Investment Protection

The Catalyst 4948-10GE is the among the first 10-gigabit Ethernet devices to make use of new, smaller X2 transceivers. As shown below, an X2 transceiver is only slightly larger than the GBIC transceiver used in gigabit Ethernet applications.



X2 transceivers perform all the same functions as previous-generation XENPAK optics, a point that has important implications for investment protection. Both XENPAK and X2 transceivers incorporate a serializer/deserialzer (SerDes) function that converts between serial and parallel data streams. If the SerDes function were to fail, only the transceiver would need to be replaced.

Another new 10-gigabit Ethernet transceiver design called XFP is smaller than X2. However, the XFP design achieves its small size by offloading the SerDes function to a separate chipset located on the switch. This means that if the SerDes component were to fail, the *entire switch* would need to be replaced. Given the large difference in cost between a single transceiver and an entire switch, X2 has a clear advantage over XFP in terms of investment protection.

We should note one caveat regarding the use of X2 transceivers in the Catalyst 4948-10GE. To save space, Cisco inverted the transceiver bay on the second 10-gigabit Ethernet interface (TenGigabitEthernet1/50). Thus, users must insert the X2 transceiver into this bay upside down from the direction used for the first 10-gigabit Ethernet transceiver bay. Users should never attempt to force transceivers; they can be inserted in only one direction.

Testing With Released Software

Opus One conducted all performance tests in this report on a Catalyst 4948-10GE running IOS Version 12.2(25)EWA. This is released software; the same image we used in testing is available to customers.

The following commands illustrate the output of the "show version" and "show bootflash" commands. The first command shows the version number of the image we used in testing, and the second shows the CRC (cyclic redundancy check) and file size for the image.

```
4948-10GE#sh ver
Cisco IOS Software, Catalyst 4000 L3 Switch Software (cat4000-I5K91S-M),
Version 12.2(25)EWA, RELEASE SOFTWARE (fc1)


4948-10GE#sh bootflash:
-#- ED ----type---- --crc--- -seek-- nlen -length- ---------date/time---------
name
1   .. image        3FD60081  CA7DEC   28 13008236 Mar 10 2005 08:03:35 -08:00
cat4000-i5k91s-mz.122-25.EWA
```

Performance Tests

Overall System Performance

Performance is the name of the game when assessing devices for use in data-center aggregation applications. Data-center aggregation switches must be able to process packets at very high rates and introduce only minimal latency while doing so.

Cisco claims aggregate system throughput for the Catalyst 4948-10GE in excess of 100 million packets per second. To validate this claim, Cisco asked Opus One to conduct a series of performance tests aimed at measuring two key metrics – throughput and latency.

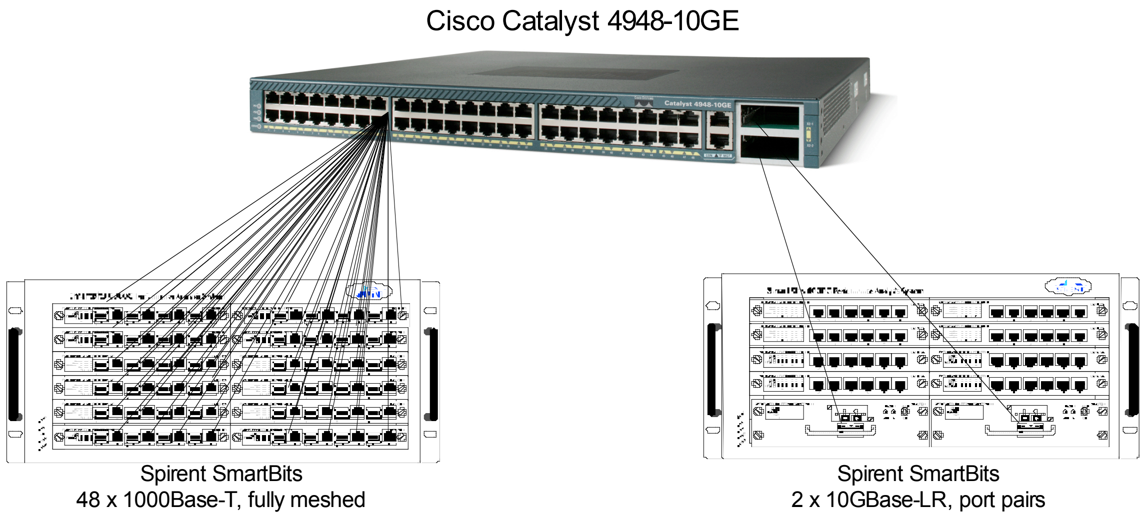
As defined in RFCs 1242 and 2544, throughput measures the maximum rate at which a device can forward packets with zero loss. The same RFCs define latency as the time required to forward a single packet.

To assess overall system throughput and latency, Opus One used the SmartBits traffic generator/analyzer from Spirent Communications. We configured the SmartBits to offer 64-byte frames in a fully meshed pattern to all 48 gigabit Ethernet interfaces of the Catalyst 4948-10GE, meaning traffic offered to all ports was destined for all other ports. At the same time, we set the SmartBits to offer bidirectional streams of 64-byte frames between the 10-gigabit Ethernet interfaces of the Catalyst 4948-10GE.

This pattern of minimum-sized frames offered at very high rates represents the most stressful load a switch will have to contend with.

Figure 1 below illustrates the test bed topology for the overall system performance tests.

Figure 1: Overall System Performance Test Bed



In assessing test results, Opus One recorded overall system throughput of 100,892,855 frames per second when the Catalyst 4948-10GE forwarded 64-byte frames for a 60-second duration.

Average latency during the same test was 3.8 microseconds for 64-byte frames on the 10-gigabit Ethernet interfaces. The latency introduced by the Catalyst 4948-10GE is nowhere near the point where it would degrade performance of even the most time-sensitive applications such as VoIP and video. Indeed, that threshold is usually well into the millisecond range, while the Catalyst 4948-10GE's numbers flirt with the nanosecond range.

Table 1 summarizes results from the overall system performance test.

Table 1: Overall System Performance

Aggregate system throughput (64-byte frames per second)	100,892,855
10-gigabit Ethernet average latency (microseconds)	3.8

Layer-2 Backbone Performance

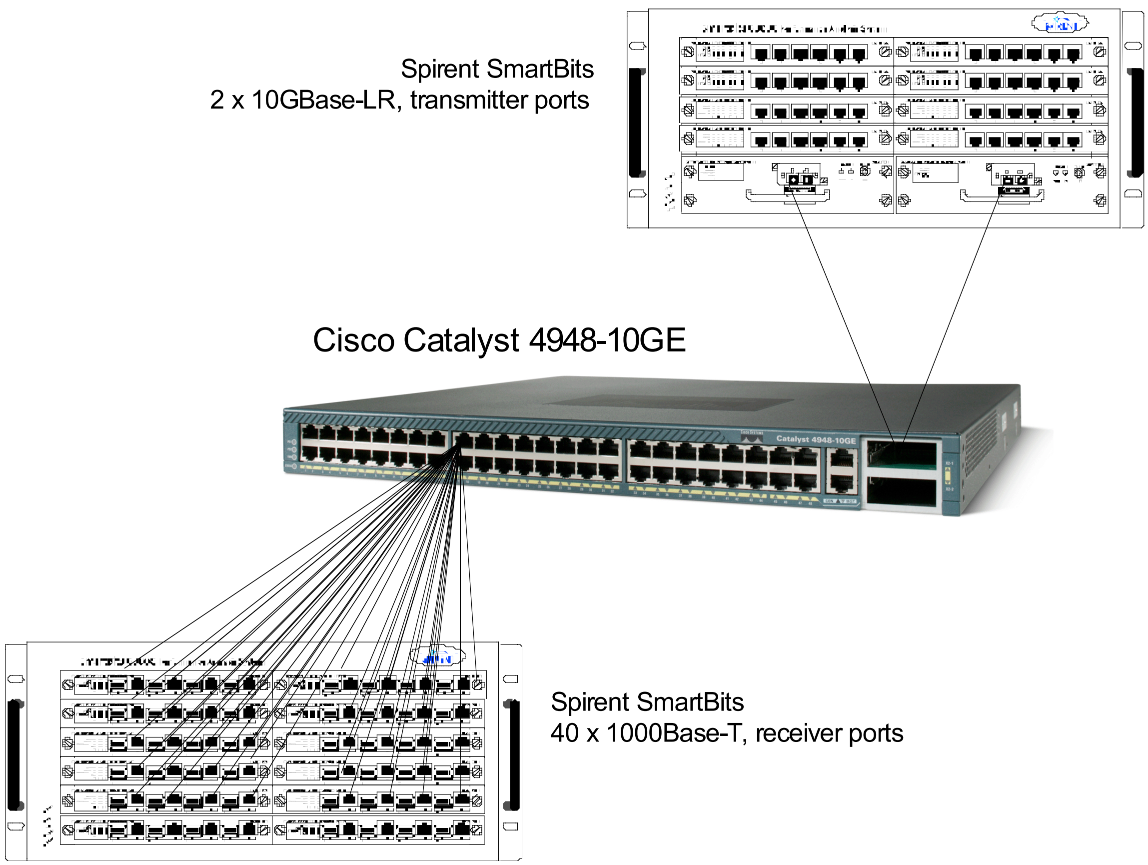
In data-center network designs, a common topology is for an aggregation switch to move traffic from one or more high-speed uplinks to multiple lower-speed "downlink" ports. The Catalyst 4948-10GE is likely to be deployed in this manner, with traffic from its dual

10-gigabit Ethernet uplinks distributed to numerous servers or other devices attached to the Catalyst's gigabit Ethernet edge ports.

Cisco asked Opus One to measure the Catalyst 4948-10GE's throughput and latency using this topology. We began with a simple layer-2 configuration in which all ports – gigabit and 10-gigabit Ethernet alike – shared the same broadcast domain. A key goal of this simple L2 configuration was to establish a baseline for comparison with later tests involving more complex L3 requirements.

Figure 2 below illustrates the test bed topology used in tests of L2 backbone performance. Other tests covered in this report – including measurements of L3 backbone performance, IP routing, and IP multicast handling – used similar topologies.

Figure 2: The Backbone Performance Test Bed



In these tests, we configured the SmartBits to offer traffic at line rate to each of the Catalyst 4948-10GE's 10-gigabit Ethernet interfaces. These streams of traffic were destined for 40 of the Catalyst 4948-10GE's gigabit Ethernet ports. This backbone traffic pattern carried unidirectional streams from the 10-gigabit Ethernet interfaces across the switch fabric and on to the 10/100/1000 edge interfaces.

As in the overall system performance tests, we used 64-byte frames, the minimum legal length in Ethernet and also the most stressful on the Catalyst's switching ASIC.

Recognizing that no production network exclusively carries 64-byte frames, we repeated the backbone test with a range of frame lengths. We used the standard lengths recommended in RFC 2544, including 64-, 128-, 256-, 512-, 1024-, 1280, and 1518-byte frames. In addition, we measured throughput for 9216-byte jumbo frames. Although the IEEE has not officially standardized jumbo frames, they have become a de facto standard in data-center applications, especially for applications requiring bulk data transfer.

In all cases, we offered traffic at line rate for 60 seconds, and measured throughput and latency.

Figure 3 below shows results from the L2 backbone throughput tests, along with the theoretical maximum numbers. The Catalyst 4948-10GE delivered line-rate throughput for all frame sizes in this test.

Figure 3: L2 Backbone Throughput

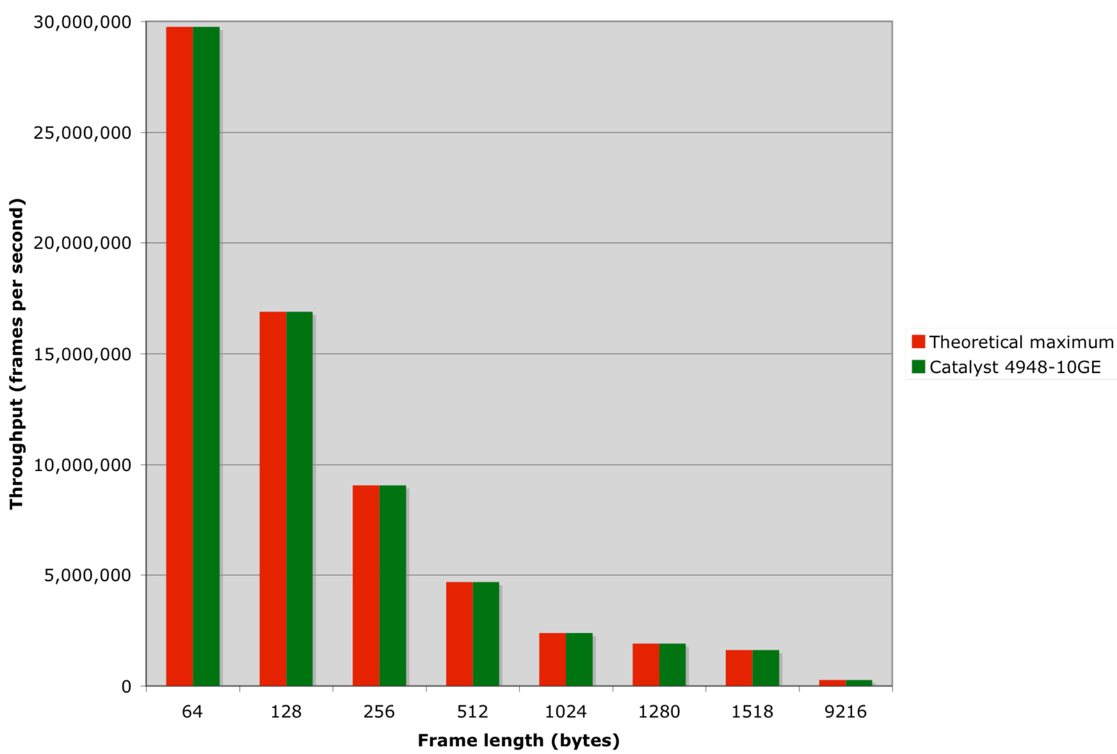
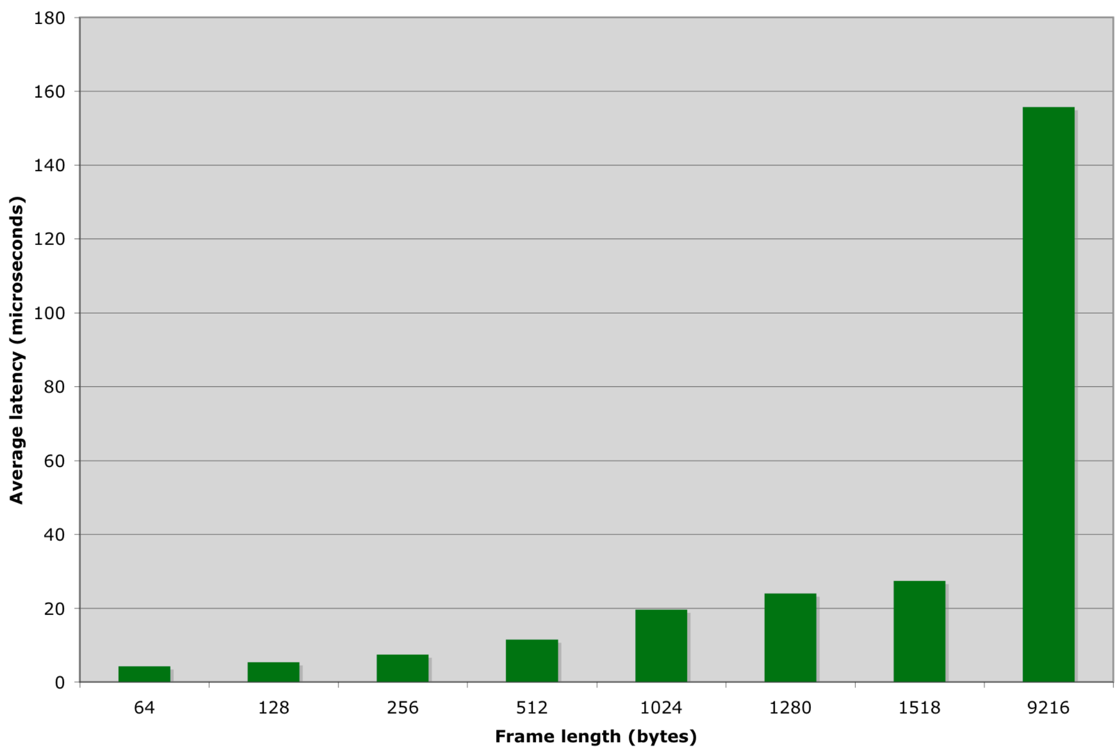


Figure 4 below presents average latency measurements from the L2 backbone tests. As noted previously, application performance begins to suffer when latency climbs into the milliseconds. None of the latencies introduced by the Catalyst 4948-10GE come close to that threshold. Indeed, for the two most common frame lengths – 64 and 1518 bytes – latency is between 4 and 27 microseconds.

At first glance, latency for jumbo frames appears abnormally large, but it is actually well in line with that of shorter frame lengths. Latency naturally increases with frame length; after all, the longer the frame, the longer it takes for any device to forward it.

Calculated proportionately, latency for jumbo frames is actually slightly *lower* for jumbos than for 1,518 byte frames. If the 27-microsecond latency for 1518-byte frames were extended proportionately to cover the 9216-byte length of a jumbo frame, latency would be 166 microseconds. In fact, jumbo latency is slightly lower, at 156 microseconds.

Figure 4: L2 Backbone Latency



Layer-3 Backbone Performance

Many organizations prefer IP subnetting over flat L2 networks in data center network designs, both to contain broadcasts and add security. Cisco claims the Catalyst 4948-10GE performs equally well in L2 and L3 settings.

To evaluate that claim, Opus One reran the backbone tests using IP subnets in place of a single L2 network. The test bed topology was identical to that used in the L2 tests (see Figure 2, again). The only difference was that we assigned one IP subnet on each port of the Catalyst 4948-10GE. In this test, the Catalyst had to route traffic among 42 different

IP subnets (one on each of the gigabit Ethernet interfaces, and one on each of the 10-gigabit Ethernet interfaces).

Figure 5 below presents results from the L3 backbone throughput tests. We have included theoretical maximum rates as well as results from the L2 backbone tests. The Catalyst 4948-10GE delivered line-rate throughput for all frame sizes in this test. These results validate Cisco's claim that throughput is identical in L2 switching and L3 IP forwarding scenarios.

Figure 5: L3 Backbone Throughput

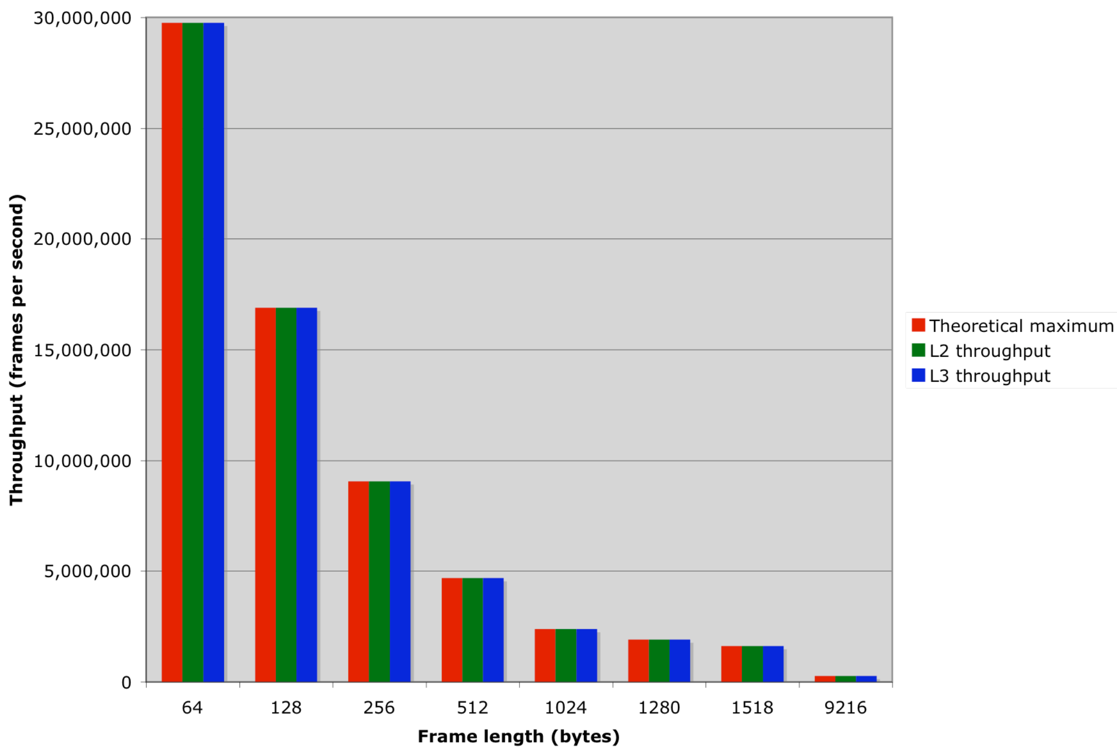
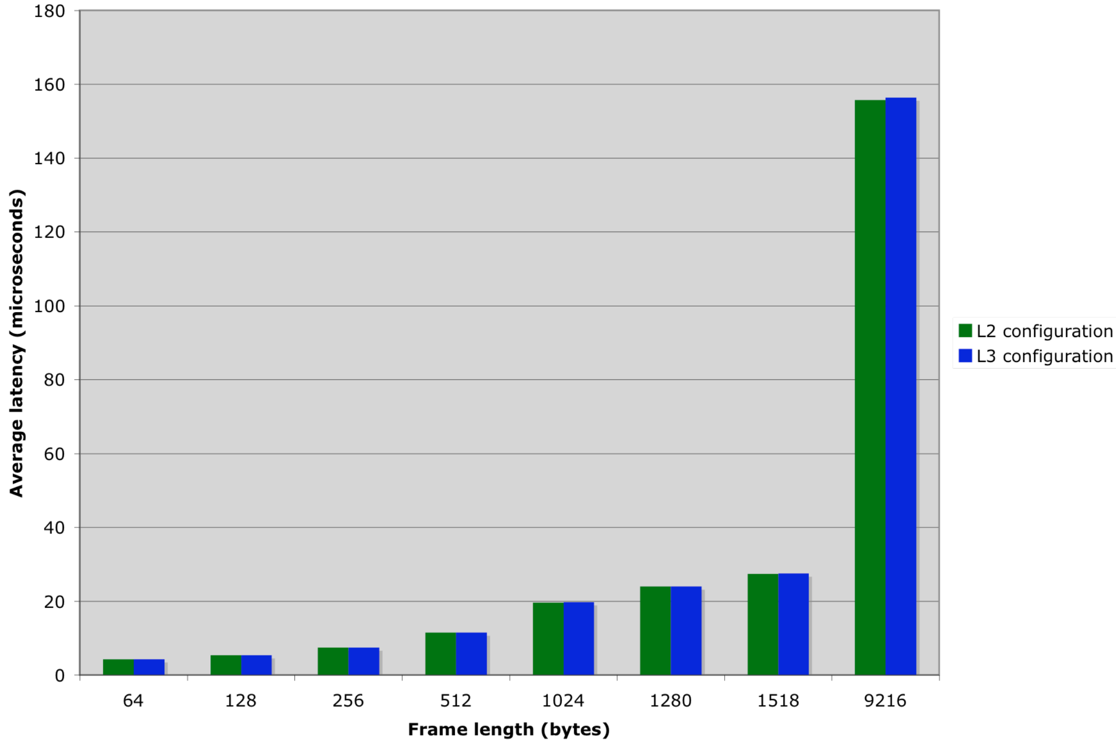


Figure 6 below presents results from tests of L3 backbone average latency. We also present L2 average latency measurements for comparison. As with throughput, note that latency is virtually identical in the L2 and L3 test cases – and once again, latency is well below the point where application performance would be affected. Based on these results, we would conclude that L2 and L3 forwarding and delay are identical with the Catalyst 4948-10GE.

Also, as in the L2 case, note that jumbo latency increases proportionately with frame length. Latency for L3 jumbo frames is actually slightly lower than that for 1518-byte frames when latency is calculated proportional to frame length.

Figure 6: L3 Backbone Latency



Latency Over Extended Durations

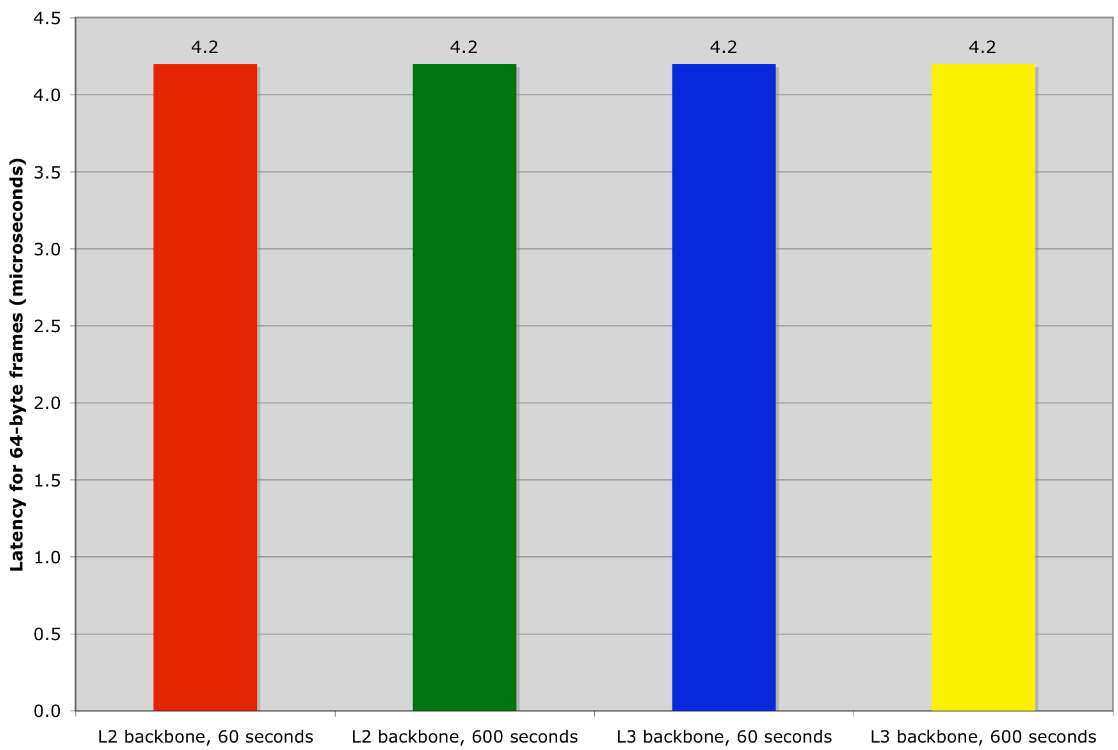
Simplicity is a hallmark of good system design, and that is especially true when it comes to keeping latency low. A key challenge is reducing the amount of processing inside the switch. Switch designs that use multiple ASICs require each frame to be buffered multiple times, increasing latency. Over time, latency can rise to unacceptably high levels.

In contrast, the Catalyst 4948-10GE is built around a single switching ASIC. This design ensures latency remains low and constant over time.

To measure latency introduced by the Catalyst 4948-10GE over extended durations, Opus One reran the L2 and L3 backbone tests, this time with test durations of 600 seconds instead of 60 seconds.

Figure 7 below presents results from these tests, along with comparative numbers from the 60-second tests. Note that latency is identical across the board. Because of the simplicity of the Catalyst 4948-10GE's design, latency remains low and constant over time.

Figure 7: Latency Over Extended Durations



OSPF Routing Performance

In data-center network designs, it is likely – indeed, probable – that IP routing protocols will be in use. Enabling routing protocols places extra processing burdens on aggregation switches such as the Catalyst 4948-10GE.

On the data plane, switches must continue to deliver high throughput and low latency, as in our previous tests involving no routing or static routing. Dynamic routing protocols such as OSPF add control-plane functions as well: Now, switches must keep track of topologies involving thousands of networks, and potentially millions of hosts.

To determine the Catalyst 4948-10GE's capabilities as an IP router, we used two test scenarios. First, we used the same topology as in the L2 and L3 backbone tests (see Figure 2, again). However, for this test case we also enabled OSPF and advertised 8,192 networks behind the gigabit Ethernet ports. The total number of unique flows was 8,192, with one host emulated on each destination network.

We then repeated the test with 250 hosts emulated on each of the 8,192 networks, making for a total of 2,048,000 unique flows. A key goal of this second test was to determine if throughput and latency would be degraded due to the large number of flows.

Some switches use a flow-based design in which each unique source-destination address pair requires its own entry in memory. These flow-based designs do not scale well, with degraded performance as flow count increases. In true router-based designs such as that used in the Catalyst 4948-10GE, flow count should not affect performance. We created this test to determine whether throughput and latency would differ with 8,192 flows or more than 2 million flows.

Figure 8 below summarizes results from the OSPF throughput tests. For comparison, we present the theoretical maximum rates and L3 backbone throughput alongside results from our OSPF tests. In all test cases, the results are identical: The Catalyst 4948-10GE delivers line-rate throughput regardless of the number of flows involved, even when routing traffic to more than 2 million hosts.

Figure 8: OSPF Throughput

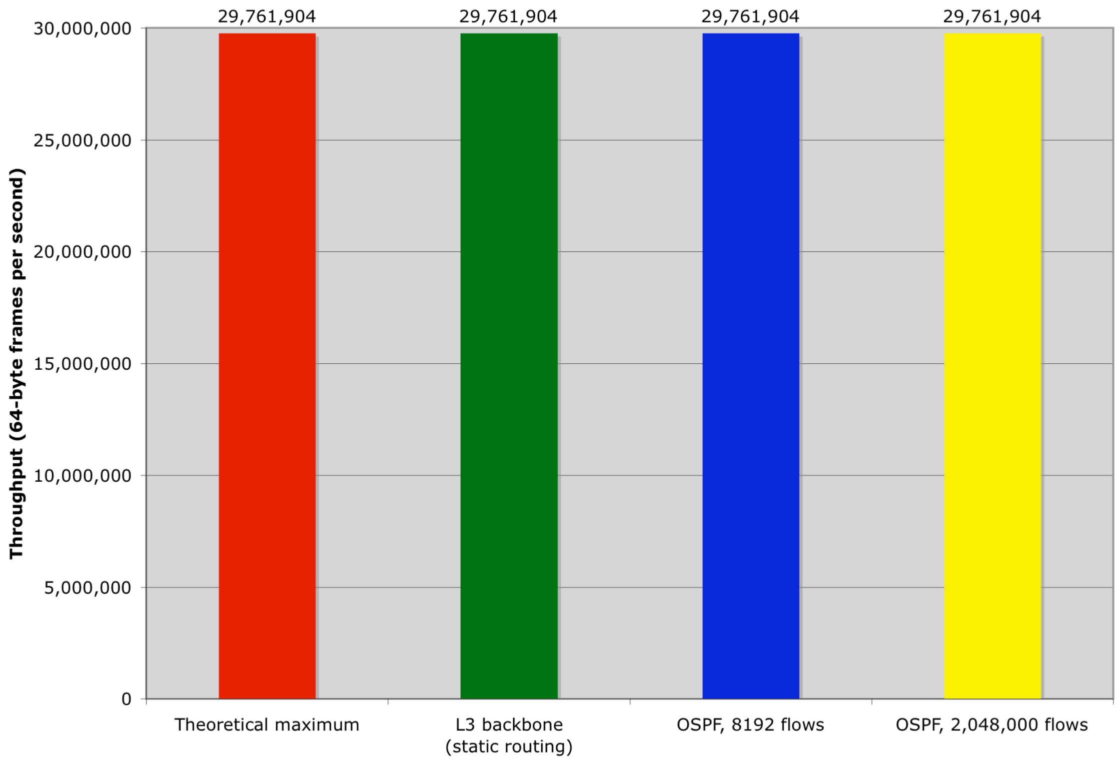
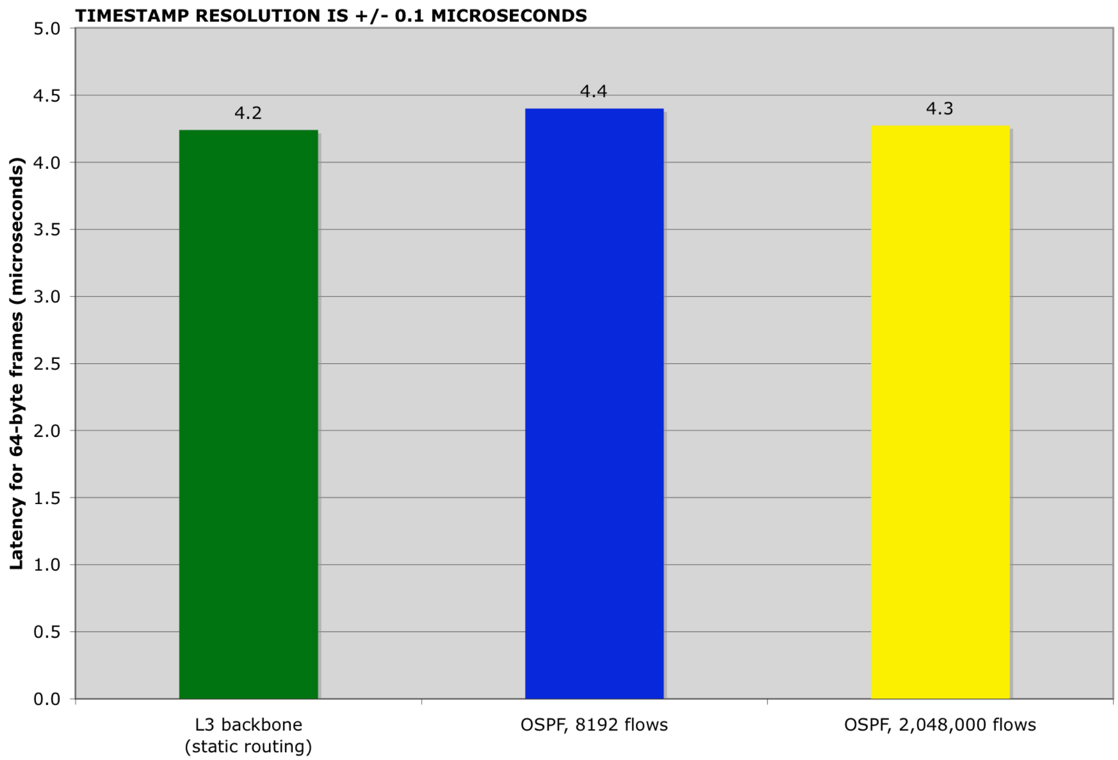


Figure 9 below summarizes results from tests of OSPF latency. The results are virtually identical, with measured differences of at most 0.2 microseconds across test cases. The actual differences may be even smaller. The timestamp resolution of SmartBits latency measurements is +/- 0.1 microseconds. Thus, a measurement of 4.3 microseconds may be as small as 4.2 microseconds or as large as 4.4 microseconds. In any event, these distinctions are academic: None of the numbers here are anywhere close to the point where they would degrade performance of even the most time-sensitive applications.

Figure 9: OSPF Latency



IP Multicast Performance

When it comes to handling streaming video and other multimedia applications, IP multicast support is a must. Switches and routers handling IP multicast have multiple requirements: First, as networks grow they must support ever-larger numbers of IP multicast groups. Second, they must deliver the same high throughput for IP multicast traffic as for unicast traffic. Finally, they must ensure that only members of given multicast groups – and *only* members of those groups – receive intended traffic streams.

To measure the Catalyst 4948-10GE's IP multicast capabilities, we designed a test that assesses all these requirements. We configured the SmartBits ports to join a total of 500 multicast groups (making for 1,000 mroutes). These were divided up so that SmartBits receivers on two gigabit Ethernet ports each joined 25 different multicast groups; with 20 different sets of 2-port receivers, this made 500 groups total. Having two receiver ports in each group exercised the Catalyst 4948-10GE's ability to replicate multicast traffic.

To assess multicast throughput, we then configured the SmartBits to offer traffic to receivers in all 500 groups via one 10-gigabit Ethernet port on the Catalyst 4948-10GE. In all cases, the SmartBits offered multicast traffic to the Catalyst 4948-10GE at line rate.

For this test, we offered multicast traffic in three frame sizes:

- 64-byte frames are the minimum length allowed in Ethernet, and thus the most stressful on a switch in terms of packet processing;
- 256-byte frames are close to the average packet length of around 300-400 bytes, according to studies of Internet traffic;
- 1518-byte frames are the maximum allowed in Ethernet, and are commonly used in streaming video and other high-bandwidth applications.

As a check against "leakage" of multicast traffic onto ports not subscribed to specific groups, we also configured the SmartBits to offer unicast traffic between two additional gigabit Ethernet ports. After the test was run, we verified that these ports forwarded only unicast traffic. Figure 10 below illustrates the test bed topology for the multicast tests.

Figure 10: IP Multicast Test Bed Topology

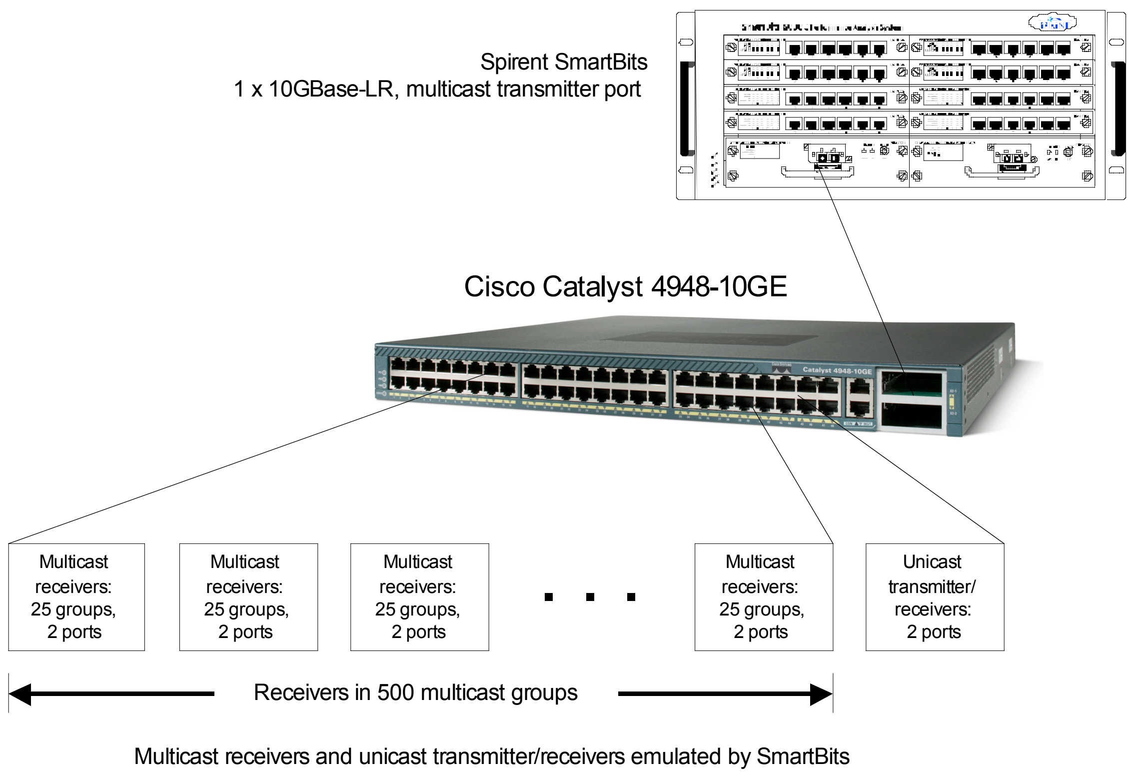
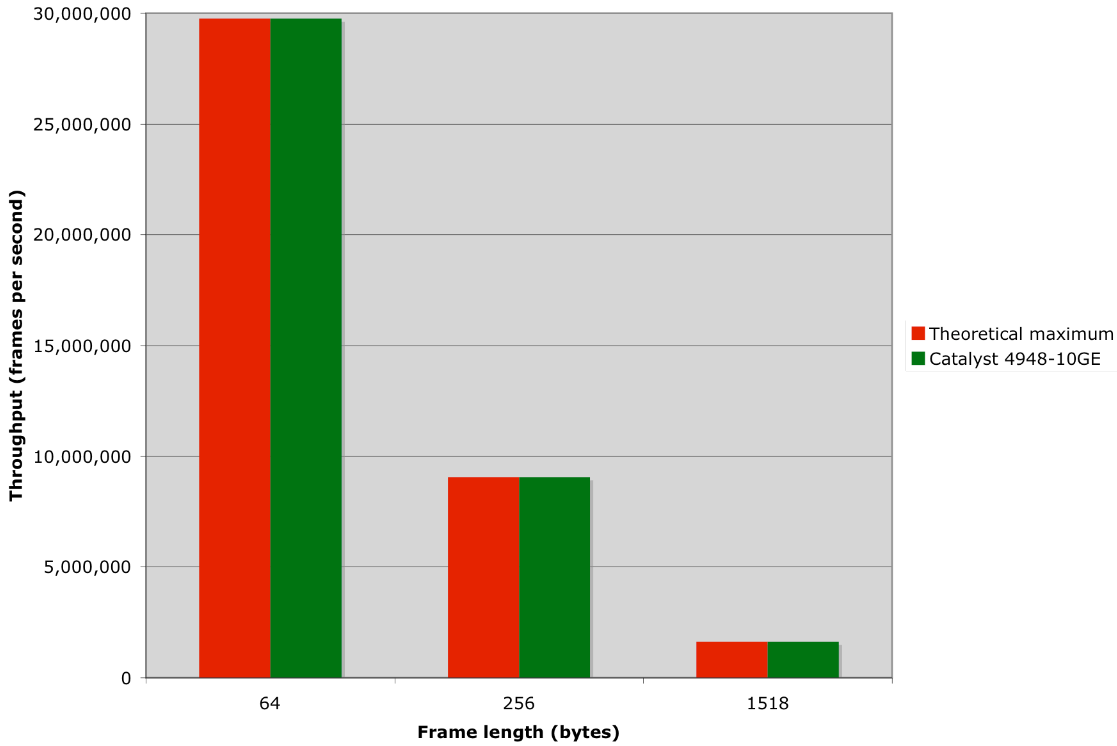


Figure 11 below presents results from the IP multicast throughput tests, as well as theoretical maximum rates for each frame length. Note that the Catalyst 4948-10GE's results match the theoretical numbers in all cases. When we offered a line-rate stream of multicast traffic to a single 10-gigabit Ethernet port, we expected traffic to be replicated at 50 percent of line rate to each of 40 edge ports – and that is exactly what we observed.

Figure 11: IP Multicast Throughput



As noted, we also offered bidirectional streams of unicast traffic on two additional gigabit Ethernet ports to verify the switch did not flood multicast traffic. The Catalyst 4948-10GE passed this test; only unicast test traffic appeared on these additional ports.

Conclusion

With the introduction of the Catalyst 4948-10GE, Cisco Systems raises the bar for aggregation switch performance. The switch offers the highest throughput in the smallest form factor of any device yet introduced by Cisco. It offers latencies that are remarkably low across all tests, ensuring safe delivery of delay-sensitive voice and video traffic. And performance is virtually the same regardless of whether the Catalyst 4948-10GE is configured in L2, L3, or dynamic routing configurations. Further, the Catalyst 4948-10GE delivers the same high throughput for IP multicast traffic as it does for unicast.

The Catalyst 4948-10GE's features also help ensure investment protection. The switch consumes only 200 watts of power despite delivering high performance. It uses new X2 transceivers that protect switch longevity. And it rapidly recovers from potentially disastrous loss of system images thanks to its ability to boot from system ROM. All these features will help Catalyst 4948-10GE deliver high performance in data centers for many years to come.

Acknowledgements

Opus One gratefully acknowledges the support of Spirent Communications, which supplied test equipment and engineering support for this project. Spirent assembled a tiger team especially for this project whose members included Charles Seifert, Ken Dayton, Tim Jefferson, Martha Lawrence, Steve Pagliarulo, Marc Pelletier, Waddell Pierce, Timmons Player, Rick Rabinovich, Peter Reddy, John Reidy, Gregg Rudshagen, Alex Su, Hue Tranh, Raymond Trinh, and Brett Wolmarans. Testing was conducted in March 2005 at Spirent's headquarters in Calabasas, CA.

About Opus One®

Opus One® is a consulting and information technology firm based in Tucson, AZ. Founded in 1989, Opus One's corporate goal is to help our clients make the best use of information technology. We focus on efficient and effective solutions in the areas of data networking, electronic mail, and security. For more information, see http://opus1.com or contact us at:

Opus One
1404 East Lind Road
Tucson, AZ 85719
+1-520-324-0494