Dell EMC Networking S4200-ON Data Center Switch

Performance, Buffer and Route Capacity Testing

DR180316O

June 2018

Miercom.com

www.miercom.com

Contents

1.0 Executive Summary

Data center switches manage communications that rely heavily on aggregating and routing technology to ensure intelligent forwarding. The port density is ever increasing to meet data center needs, and performance in realistic deployment scenarios should be evaluated to validate data sheet specifications.

Dell EMC engaged Miercom to validate aggregate throughput capacity, latency, buffer and route capacity of the Dell EMC S4248FBL-ON 100/40/10 Gigabit Ethernet (GbE) Switch. Using an Ixia IxNetwork Test System, we simulated high-volume traffic of different frame sizes, ranging from 64 to 9216 bytes. Tests were conducted in accordance with RFC 2544 for both Layer 2 and Layer 3 traffic. Percentage of line rate, forwarding rate, aggregate throughput and latency were recorded. Deep buffer and BGP route convergence capacities were also verified.

Key Findings

- **Verified rated 800 Gbps line rate performance**
- **Achieved 100% line rate for Layer 2 and Layer 3 traffic**
- **Observed 720 Mpps frame-handling ability**
- **Average store-and-forward latency as low as 1.9 µs**
- **5.9 GB buffer capacity for traffic overload from 2x10GbE ports to 1x10GbE port**
- **Buffered 100 times more packets than the shallow buffer switch S4148F-ON to accommodate large packet bursts with no frame loss**
- **BGP route convergence for 100K routes in 27.9 seconds using Ixia IxNetwork test**
- **Observed maximum BGP route convergence of 2 million routes in 402 seconds**
- **Achieved BGP convergence failover time of 535 seconds for 2 million routes**

Based on our testing observations, we find the Dell EMC S4248FBL-ON 100/40/10 Gigabit Ethernet Switch is well suited for data center environments as a deep buffer switch and WAN Edge for Internet peering. We proudly award the Dell EMC Networking S4248FBL-ON Switch the ***Miercom Performance Verified*** certification.

Robert Smithers

CEO

Miercom

2.0 Product Tested

The Dell EMC Networking S4200-ON switch is Dell EMC's latest disaggregated hardware and software data center networking solution, providing a broad range of functionality to meet the growing demands of today's data center environment.

The S4200 is a performance-optimized open networking switch that provides deep buffer capability and enhanced hardware table capacities for Internet routing or high-density flow tables for software defined networking applications.

Key Features

- 1RU high-density 10/40/100 GbE Top-of-Rack switch with forty ports of 10 GbE (SFP+), two ports of 40 GbE (QSFP+), and six ports of 100 GbE (QSFP28)
- There are two variants of S4200-ON:
  - S4248FB-ON: with deep-buffer only
  - S4248FBL-ON: with deep-buffer and additional TCAM for expanded FIB and ACL tables and flows
- Multi-rate 100GbE ports support 10/25/40/50 GbE
- 800 Gbps (half-duplex)/1.6 Tbps (full-duplex) non-blocking, cut-through switching fabric delivers line-rate performance under full load
- Deep buffer capability of up to 6 GB
- Supports virtual output queue (VoQ) architecture to eliminate head of line blocking
- Supports up to 2 million IPv4 route entries
- High-performance SDN/OpenFlow 1.3 enabled with ability to interoperate with industry standard OpenFlow controllers
- IO panel to PSU airflow or PSU to IO panel airflow
- Supports the open source Open Network Install Environment (ONIE) for zero touch installation of alternate network operating systems
- Redundant, hot-swappable power supplies and fans
- IEEE 1588v2 capable hardware

Dell EMC Networking S4248FBL-ON Switch

OS Version: 10.4.OE (R2)

Build Version: 10.4.OE (R2.27)

Source: Dell EMC

3.0 How We Did It

Miercom features hands-on testing designed to simulate real-world deployments to provide a robust, realistic assessment of products' capabilities and their effectiveness. Tested solutions are given the chance to analyze the product's performance before, during and after testing. Firmware versions should remain constant throughout testing to ensure a fair and accurate assessment.

The fundamental aspect of these test criteria is to provide a means to validate product performance and functionality in a simulated data center environment.

Test Tools

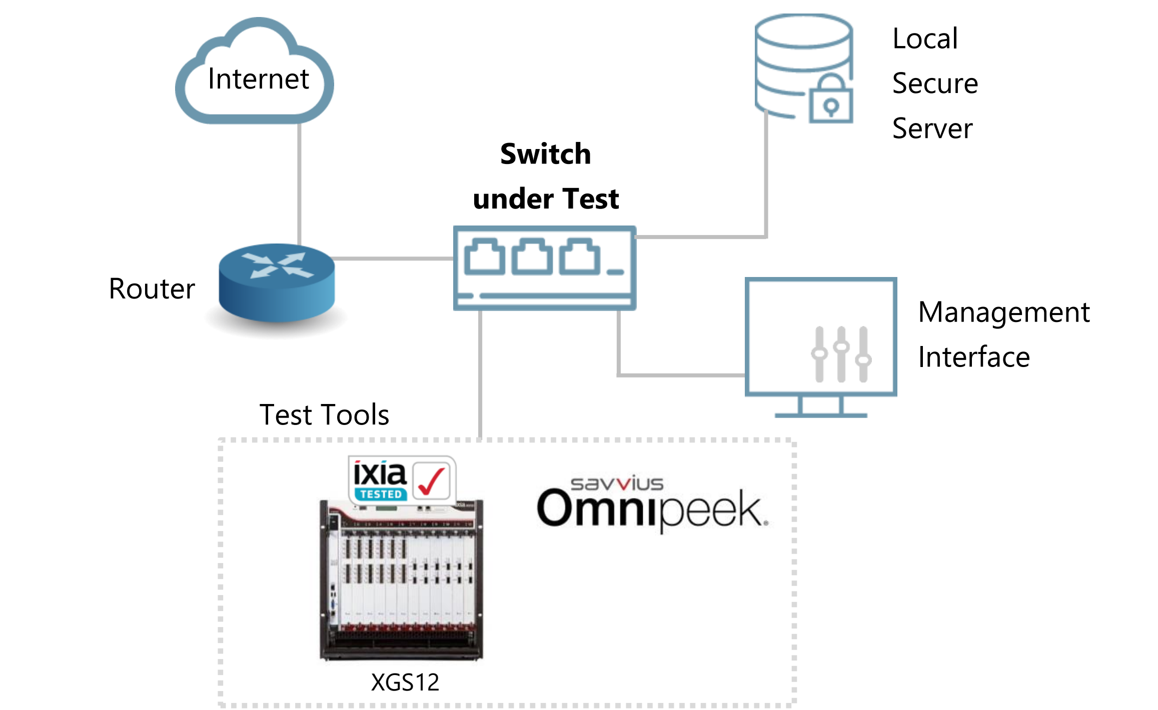
Source: Miercom

The test tools featured above are used for traffic and threat generation, real-time monitoring and capturing of network activity.

Ixia XGS12 is an end-to-end testing solution which has long been a pioneer in developing test software and hardware. This system is designed to fully and accurately assess device performance for the latest core-switch port data rates of up to 100 Gbps.

Savvius Omnipeek captures network traffic and creates packet files for replay. Statistics can help monitor changes in real-time. By baselining normal activity, changes can be observed to analyze problem areas in the network.

Test Bed Diagram



Source: Miercom

The Device Under Test (DUT) is deployed between the untrusted network from the Internet/switch and the local private network. The industry test tools are connected to the switch using 10 GbE and 100 GbE port pairs. The test tools generate traffic with different frame sizes, time intervals and device addresses to simulate a real-world data center environment.

For throughput and latency tests, a 40x10 GbE/4x100 GbE 44-port configuration was used on Layer 2 (Data Link Layer, switched frames) and Layer 3 (Network Layer, IP routed) traffic. Traffic flows consisted of 64, 128, 256, 512, 1024, 1280, 1518 and 9216-byte (B) frame sizes.

The following results show the maximum forwarding throughput the switch achieves without frame loss. When a switch accepts and successfully processes and forwards all traffic at the maximum theoretical rate of the port, the switch is said to perform at wire speed or full line rate for the particular packet size.

4.0 RFC 2544 Port-Pair Unicast Throughput and Latency

Description

RFC 2544 is a standard test case for verifying throughput and latency for unicast traffic. As a rule, in the beginning frames are sent at a specified 100 percent line rate, which is continually stepped down/up in subsequent iterations, using a binary search algorithm, until frame loss no longer occurs.

Ixia test ports were connected to every port on the Dell EMC S4248FBL-ON switch. RFC 2544 describes how to conduct basic benchmark tests for throughput and latency. Bi-directional Layer 2/3 (Ethernet/IP) traffic is applied on port pairs of the switch to be processed across the switch fabric. The 44-port configuration of 40x10 GbE, 4x100 GbE was used.

The maximum throughput achievable through the switch was measured and latency was observed for different frame sizes. Frames can be MAC-only Layer 2, or Layer 3 IPv4, with or without extension headers. This test used 64, 128, 256, 512, 1024, 1280, 1518 and 9216-byte traffic.

Theoretical maximum frame rates (frames/second) differ based on Ethernet transmission speeds and frame size. The 44-port configuration has the following attributes which contribute to the measured percent line rate received on the egress ports:

The effective theoretical forwarding rate of each frame size can be calculated using the adjoining table. For example, the 64-byte frames have a maximum aggregate forwarding rate of 1.19 million packets per second (Mpps). We observed 720 Mpps, and therefore achieved 60.5 percent line rate performance. Line rate is relative to the maximum achievable for each frame size.

Frame Size (Bytes)	Maximum Frames/sec	
64	10GbE	14,880,952
	100GbE	148,809,520
128	10GbE	8,445,946
	100GbE	84,459,460
256	10GbE	4,528,986
	100GbE	45,289,860
512	10GbE	2,349,625
	100GbE	23,496,250
1024	10GbE	1,197,318
	100GbE	11,973,180
1280	10GbE	961,539
	100GbE	9,615,390
1518	10GbE	812,744
	100GbE	8,127,440
9216	10GbE	135,340
	100GbE	1,353,400

Test Setup

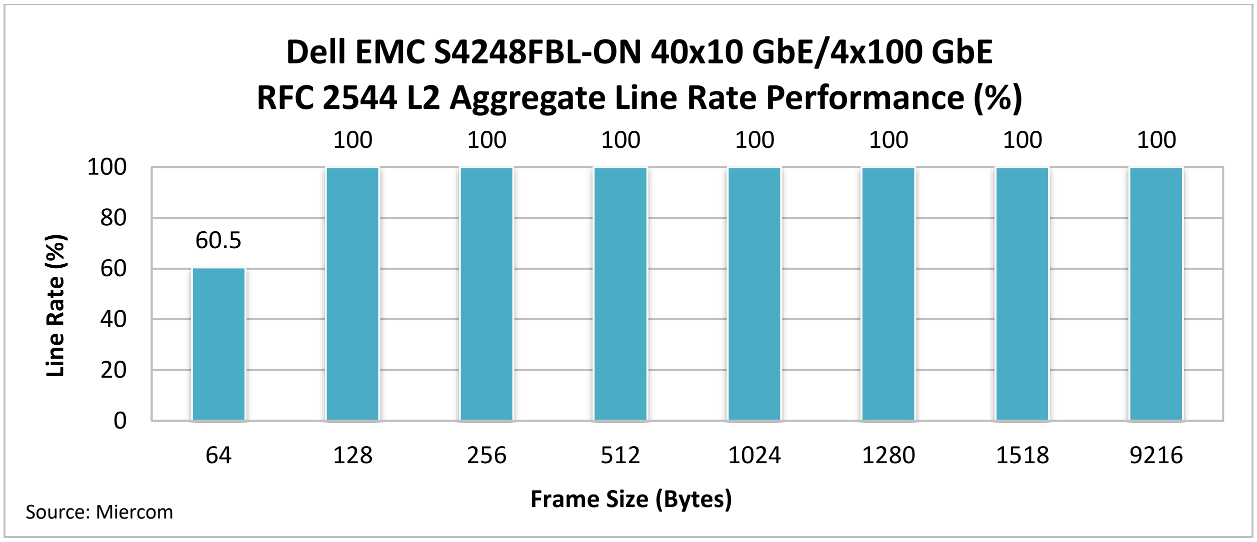
The Ixia load generator was configured to forward traffic to, and receive traffic from, each port of the switch. This test was configured with a one-to-one, port-to-port traffic mapping between port pairs; all traffic arriving on one port in the port pair is delivered by the switch to the other port in the port pair and vice versa.

The Dell EMC S4248FBL-ON switch was configured for Layer 2 switching, as well as Layer 3 forwarding for IP routing. Port pair combinations were assigned within the test system so that bi-directional traffic was transmitted between ports across the fabric, in accordance with RFC 2544. All ports of the Dell EMC S4248FBL-ON switch were connected to the Ixia load generation system for bi-directional traffic to and from each port.

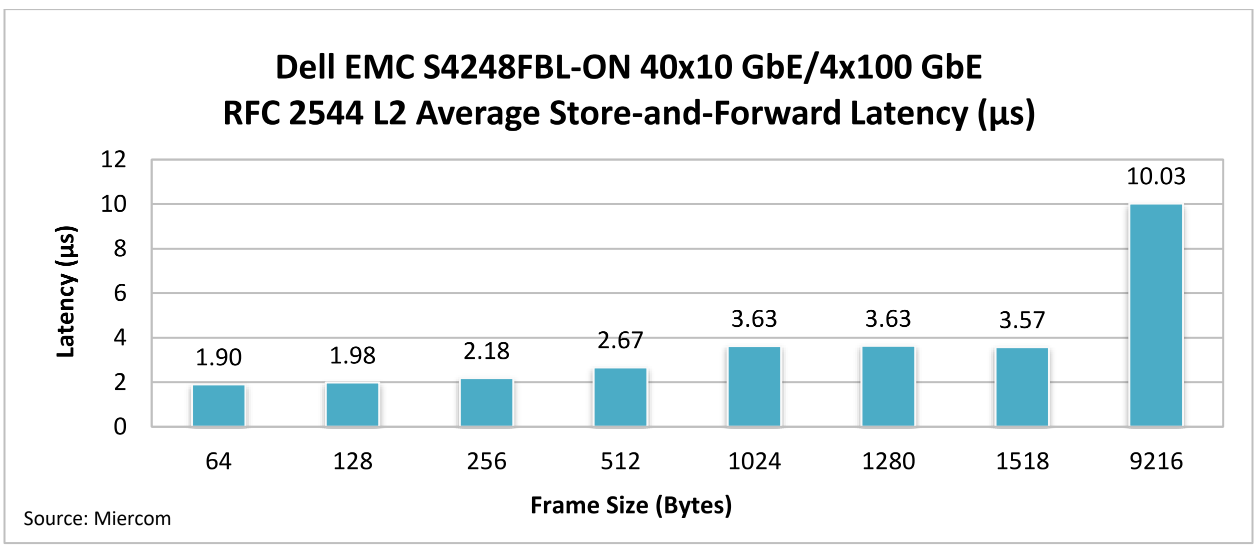
Once the maximum traffic rate without loss is established for a frame size, latency through the switch is calculated. The store-and-forward latency is calculated by subtracting the transmit timestamp from the receive timestamp. The minimum, maximum and average latencies for all packets sent and received are calculated and reported. This test is run for one to two minutes.

Results

The performance of the Dell EMC S4248FBL-ON switch was tested over Layer 2 (L2) using 44 ports. This configuration used 40x10 GbE and 4x100 GbE ports. The Dell EMC S4248FBL-ON switch achieved 100 percent line rate without any loss for all frames above 64 bytes and experienced latency as low as 1.9 microseconds. The charts on the following page depict throughput and latency for each frame size.

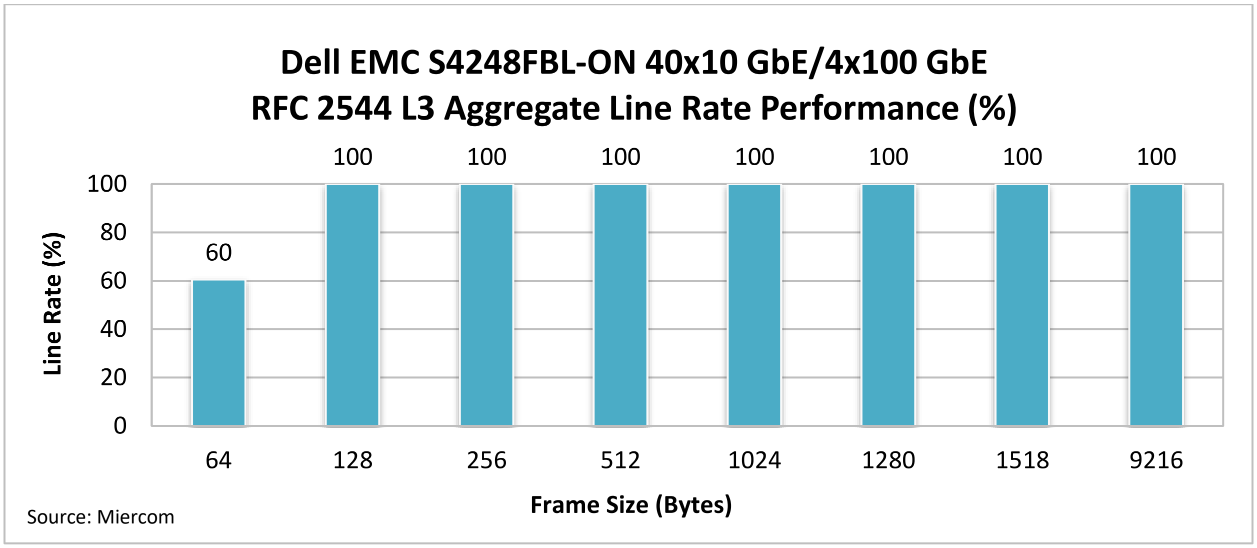


The test-load generator forwarded and received 800 Gigabits per second (Gbps) of traffic through each directly connected 10 and 100 GbE port pairs. The test was conducted for a range of frame sizes including 64, 128, 256, 512, 1024, 1280, 1518 and 9216 bytes. The Dell EMC S4248FBL-ON switch forwards Layer 2 frames at 100 percent line rate for all frame sizes above 64 bytes, achieving a maximum forwarding rate of 720 Mpps. The switch can forward frames at wire-speed without any loss.

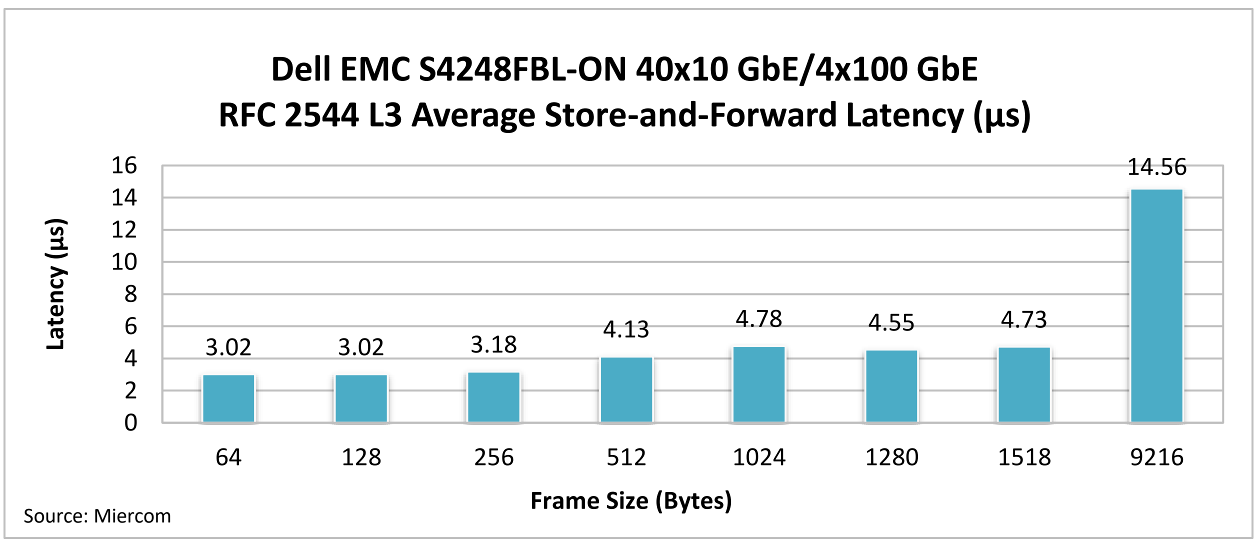


Average latency of the Dell EMC S4248FBL-ON switch was as low as 1.9 microseconds for 64-byte frames. As expected, larger frames typically require more processing, resulting in more latency but remained under 10.03 microseconds. The maximum latency by packet size rose from 1.9 to 10.03 microseconds. For these latency measurements, the switch was subjected to a 100-percent traffic load of Layer 3 IPv4 unicast traffic on all 44 ports. Latency measurements were done in accordance with RFC 2544.

Aggregate throughput performance and store-and-forward latency was tested over Layer 3 (L3) using 44 ports. The Dell EMC S4248FBL-ON switch achieved 100 percent line rate without any loss for all frame sizes above 64 bytes and latency as low as 3 microseconds.



After Layer 2 traffic testing, throughput was measured for Layer 3 traffic. The test-load generator forwarded and received 800 Gbps of traffic through each directly connected 10/100 GbE port using 64, 128, 256, 512, 1024, 1280, 1518 and 9216-byte frame sizes. The Dell EMC S4248FBL-ON switch forwards Layer 3 packets at 100 percent line rate for all frame sizes above 64 bytes, achieving a maximum forwarding rate of 720 Mpps. The switch can forward frames at wire-speed without any loss.



Lowest latency was observed for 64 and 128-byte frames, at 3.02 microseconds. Latency for all frame sizes remained under 14.56 microseconds. As expected, smaller frames were easier to process and had lower latency. The Dell EMC S4248FBL-ON switch was subjected to a 100-percent traffic load of Layer 3 unicast traffic on all 44 ports. Latency measurements were calculated in accordance with RFC 2544.

5.0 Deep Buffer

Description

Network traffic on average, observed over a time scale of seconds, seems steady and consistent. But when analyzing the same traffic at micro- or nanosecond intervals, it appears to be full of bursts of data. These microbursts, along with port speed differences and many-to-one traffic, require large buffer memory to avoid dropped packets.

Deep buffer switches are better able to handle large bursts of data, whereas shallow buffer switches tend to drop packets from such large bursts. They provide predictable performance, fewer drops and shorter flow completion times to a shallow buffer switch.

Deployment of deep buffer switches is best suited for distributed applications, storage interconnects and video applications that generate large bursts of data and are sensitive to packet loss.

Switch buffer size may impact cloud computing in a spine-and-leaf architecture where network performance can be a bottleneck at the shared uplink or downlink port. Two types of switch chips commonly used in data center switches are:

- **On-chip SRAM buffer:** provides less packet buffer; this pool of memory is allocated to ports in the event of congestion
- **External DRAM buffer:** provides on the order of 500 times more packet buffer than on-chip SRAM

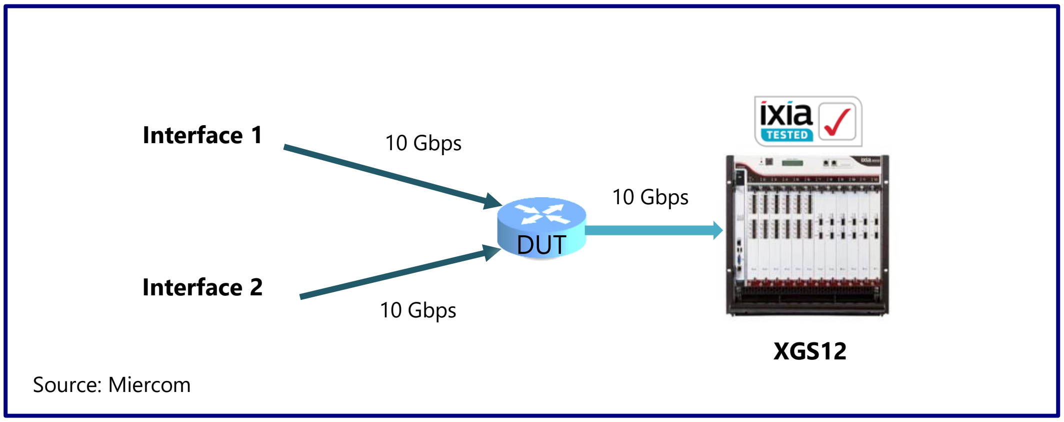
To observe the advantages of deep buffer, the buffer capacity of a 10/40/100 GbE switch with external DRAM buffer is compared to a 10 GbE switch with on-chip SRAM buffer, which isn't equipped to handle bursts of high-volume traffic. The Ixia IxNetwork test tool generates Layer 2 traffic in a controlled overload condition and is then sent through the switch.

There were two buffer tests carried out on the S4248FBL-ON:

1. Deep buffer stress test
2. Performance comparison between the deep buffer switch (S4248FBL-ON) versus a shallow buffer switch (S4148F-ON)

Testing was conducted at Miercom and Miercom-verified partner facilities.

Buffer Test Topology



The test topology for both tests was setup as shown above. Buffer was tested by sending traffic from two 10 GbE ports to one 10 GbE port with dynamic threshold and packet size of 64 to 9000 bytes.

Test Setup

The Ixia IxNetwork 7.3 and IxExplorer 6.90 tools generated and applied simulated Layer 2 traffic overload conditions to the Dell EMC S4248FBL-ON deep buffer and Dell EMC S4148F-ON shallow buffer switches.

Test Setup for deep buffer stress test:

1. Line rate traffic was sent from one 10 GbE interface for 9 seconds.
2. Line rate burst traffic was sent from a second 10 GbE interface. A burst size of 3 GB was used.
3. The total number of packets that arrived at egress 10 GbE interface was captured using Ixia.
4. The total number of buffered packets and percentage of frame loss was derived from the data captured in Step 3.

The above tests were repeated with different packet sizes, ranging from 64 to 9000 bytes on the S4248FBL-ON.

Test Setup for comparison test:

1. Line rate traffic was sent from one 10 GbE interface for 9 seconds.
2. Line rate burst traffic was sent from a second 10 GbE interface. Two separate tests were run with burst sizes of 5 MB and 1 GB, respectively.
3. The total number of packets that arrived at egress 10 GbE interface was captured using Ixia.
4. The total number of buffered packets and percentage of frame loss was derived from the data captured in Step 3.

The above tests were repeated with different packet sizes, ranging from 64 to 9000 bytes on the S4248FBL-ON and S4148F-ON switches, separately.

Results

Test 1: Deep Buffer Switch S4248FBL-ON Buffer Size

Deep Buffer Switch Dell EMC S4248FBL-ON Buffer Capacity 2x10 GbE ingress ports to 1x10 GbE egress port			
Packet Size (B)	**Number of Line Rate Frames in Port 1 (for 9 sec)**	**Number of Frames per Burst (total 3 GB) in Port 2**	**Physical Buffer Utilization (GB)**
64	133928572	46875000	5.74
512	21146617	5859375	5.74
1500	7401316	2000000	4.80
2000	5569307	1500000	3.61
6000	1868771	500000	3.64
9000	1247229	333334	4.05

Buffer capacity was measured during a traffic overload scenario using two 10 GbE ingress ports with line rate traffic on each port. One port sent a continuous stream of data, and the other sent 3 GB bursts of data at line rate. Both links sent incoming data to a single 10 GbE egress port. Total physical buffer utilization was calculated based on the number of packets that arrived at the egress port. A total number of 5.879 GB buffer memory was observed for the deep buffer switch. It was further observed that the S4248FBL-ON deep buffer switch had a maximum buffer utilization efficiency with 2000 B packets.

Test 2: Comparison of Deep Buffer Switch S4248FBL-ON vs Shallow Buffer Switch S4148F-ON

	Buffer Performance: Dell EMC S4248FBL-ON vs S4148F-ON 2x10 GbE ingress ports to 1x10 GbE egress port			
	S4248FBL-ON Deep Buffer		**S4148F-ON Shallow Buffer**	
Burst Size (5MB)	**Number of Buffered Packets**	**Lost Burst Frames (%)**	**Number of Buffered Packets**	**Lost Burst Frames (%)**
Packet Size (1500 B)	3340	0	3340	0
Packet Size (9000 B)	557	0	557	0
Burst Size (1 GB)	**Number of Buffered Packets**	**Lost Burst Frames (%)**	**Number of Buffered Packets**	**Lost Burst Frames (%)**
Packet Size (1500 B)	666674	0	5454	99.13
Packet Size (9000 B)	111113	0	990	99.11

The deep buffer switch buffers 100 times more packets than the shallow buffer switch, thereby accommodating large bursts of data without packet loss. This is important when considering distributed workloads, like BigData, video and other applications – which generate large bursts of data with sensitivity to packet loss.

6.0 Deep Routing

Border Gateway Protocol (BGP) is used for facilitating communications between routers in different autonomous systems – a network under the control of a single administrative entity.

Equipment vendors, carriers, service providers and enterprise customers depend on the interoperability, scalability, and performance of their network equipment to perform multiple services. These services are critical to their communications and core infrastructures.

The challenge for routing tables is to deliver data to its desired destination within an expected time frame. Routing tables are limited by the allotted storage space and, if overrun, can result in persistent routing loops or network instability as the device attempts to remove itself from the topology.

To test deep routing capacity, a scalability scenario is used to determine the maximum number of routes that a BGP-enabled switch can sustain at a single time. This helps test engineers evaluate network devices and understand network limitations before actual implementation in a live network.

Tests were run using Ixia IxNetwork, which contains a rich set of routing and switching protocol emulations – IGP speaking routers which support IS-IS, OSPF, EIGRP, RIP and BGP – to provide large quantities of routes to stress-test the switch using real Internet routes to exercise the routing table.

6.1 BGP Route Capacity

Description

Routing entries change frequently in a dynamic network. In a high-capacity network, these changes may cause delays or packet loss. Despite these changes, a data center switch is expected to handle millions of routes while providing high performance and scalability.

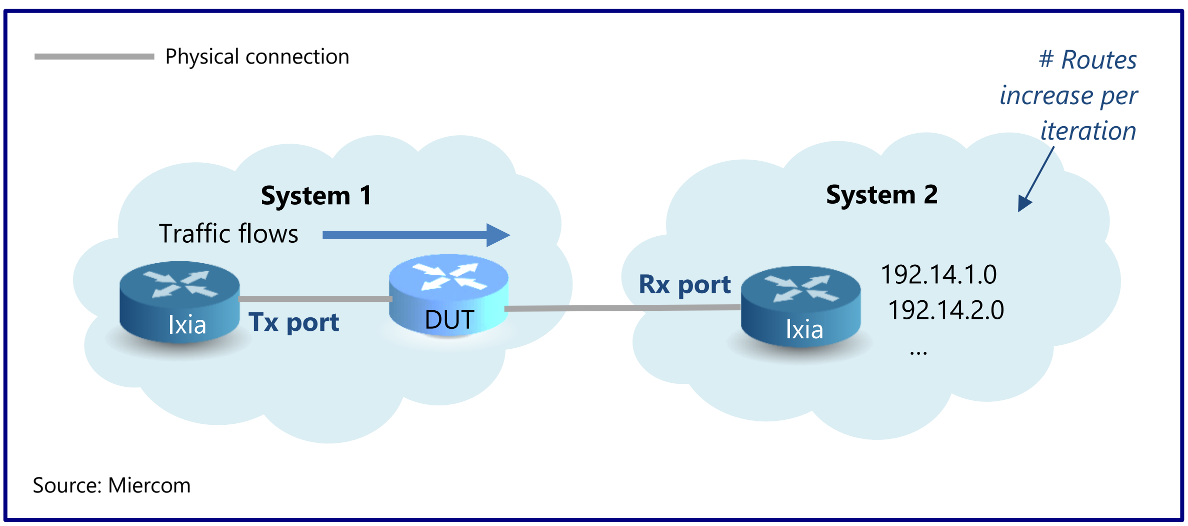
The deep routing test determines the number of routes that a BGP-enabled DUT can sustain at a single time. This scalability test reveals BGP route capacity and network limitations.

Test Setup

The test requires two test ports; one to transmit and one to receive. The transmit direction of traffic is unidirectional. Test port 2 advertises the eBGP routes, and test port 1 sends traffic to verify advertised prefixes. During the test, test port 2 increases the number of advertised routes until the maximum route capacity is sustained by the switch. The Ixia IxNetwork 5.5 application was used to configure, control, and execute this test.

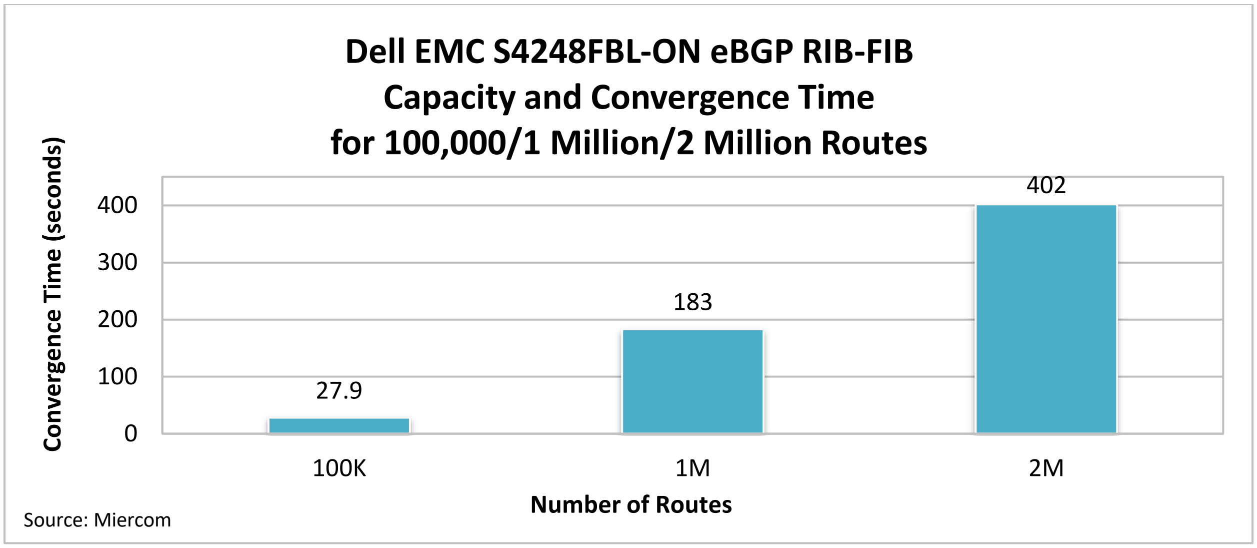
Route capacity was calculated when the DUT was able to send 100 percent ingress traffic, which was being sent to every route advertised, to verify a fully updated Routing Information Base (RIB) and Forwarding Information Base (FIB) information.

Deep Routing: BGP Route Capacity Test Topology



Results

First the Dell EMC S4248FBL-ON switch was tested for handling of 100,000 eBGP routes. All routes were installed in the RIB and FIB in 27.87 seconds. The switch was tested up to its rated capacity of 2 million routes. The switch was able to accept 2 million routes and forward traffic to all 2 million routes.



The Dell EMC S4248FBL-ON switch eBGP convergence with 100,000 (100K) routes took 27.9 seconds to converge. The switch was then tested for the scalability of up to 2 million (2M) routes. These routes were successfully processed with forwarding data in 402 seconds.

6.2 BGP Failover

Description

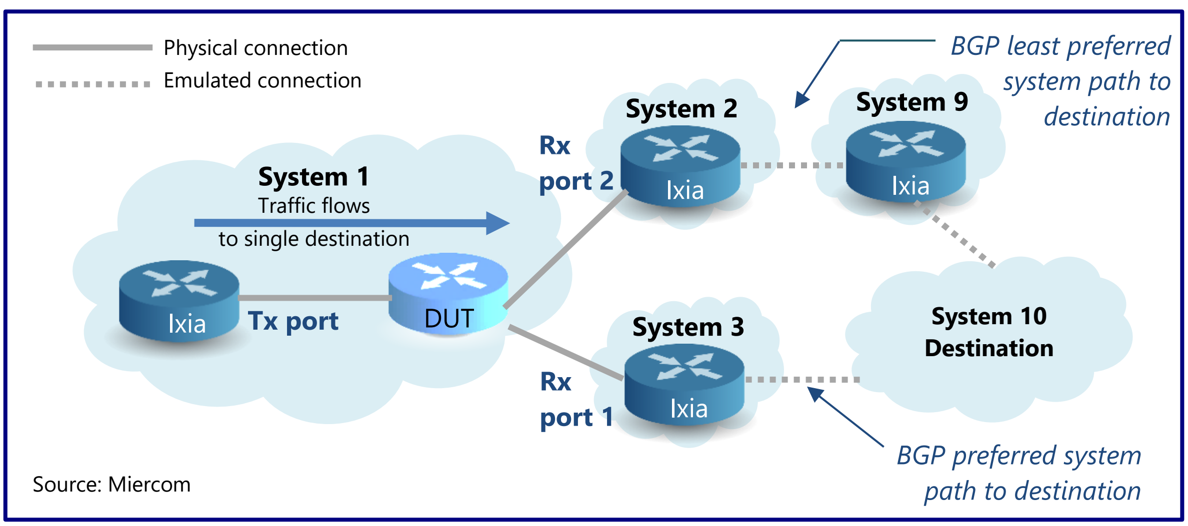
In today's multiplay networks, it's essential that we consider service interruption time with respect to traffic type. Layer 3 routing protocols such as RIP, OSPF, ISIS and BGP include the ability to re-route IP traffic if a link or network fails.

Test Setup

The test requires three test ports; one to transmit and two to receive. Both receiving ports emulate BGP networks and are eBGP-enabled. The transmit direction of traffic is unidirectional. All three ports were configured for IP and have unique subnets in which to communicate with the tester ports. Receiving ports were configured with two eBGP autonomous system numbers having unequal cost path. Both systems were advertising the same routes to the switch.

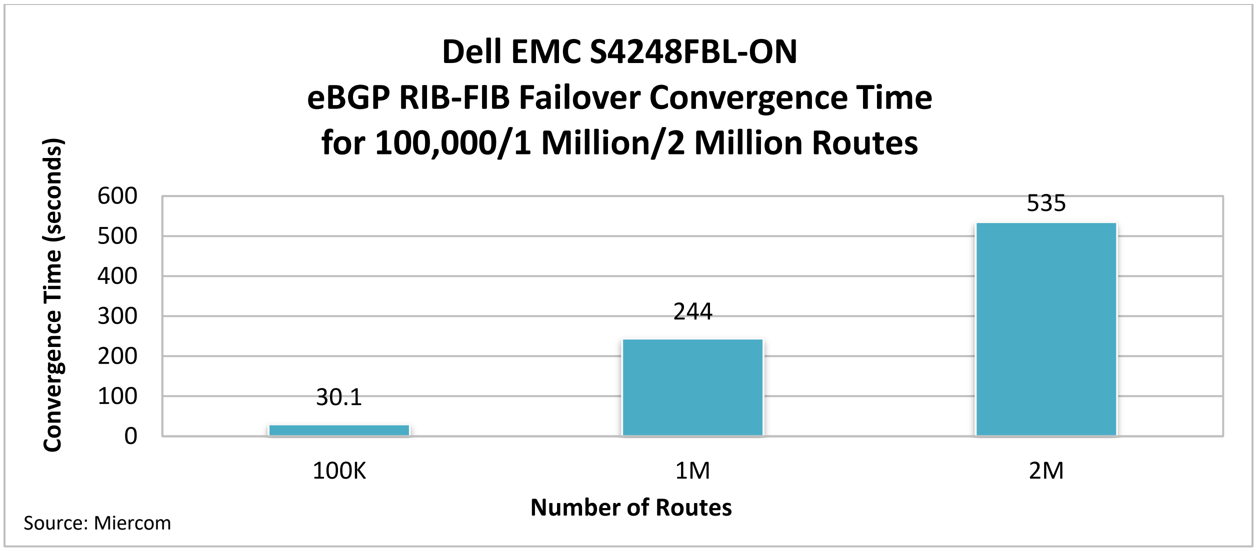
The link to the lower cost path was shut down at the switch interface using the shutdown command. Now all traffic takes a high cost BGP path, and the switch should update its forwarding information base (FIB). A convergence time was recorded when full traffic connectivity was established with full FIB convergence.

Deep Routing: BGP Failover Test Topology



Results

First the Dell EMC S248FBL-ON switch was tested for the failover of 100,000 eBGP routes. Failover convergence occurred after 30.12 seconds. The switch was tested up to its rated capacity of 2 million routes. The failover convergence of 2 million routes took 535 seconds to complete.



The Dell EMC S4248FBL-ON switch eBGP failover convergence of 100,000 (100K) routes took 30.1 seconds to complete. The switch was then tested for the scalability of up to 2 million (2M) routes. Failover handling of these routes required 535 seconds.

About Miercom

Miercom has published hundreds of network product analyses in leading trade periodicals and other publications. Miercom's reputation as the leading, independent product test center is undisputed. Private test services available from Miercom include competitive product analyses, as well as individual product evaluations. Miercom features comprehensive certification and test programs including: Certified Interoperable, Certified Reliable, Certified Secure and Certified Green. Products may also be evaluated under the Performance Verified program, the industry's most thorough and trusted assessment for product usability and performance.

Customer Use and Evaluation

We encourage customers to do their own product trials, as tests are based on the average environment and do not reflect every possible deployment scenario. We offer consulting services and engineering assistance for any customer who wishes to perform an on-site evaluation.

Use of This Report

Every effort was made to ensure the accuracy of the data contained in this report, but errors and/or oversights can occur. The information documented in this report may also rely on various test tools, the accuracy of which is beyond our control. Furthermore, the document relies on certain representations by the vendors that were reasonably verified by Miercom but beyond our control to verify to 100 percent certainty.

This document is provided "as is," by Miercom and gives no warranty, representation or undertaking, whether express or implied, and accepts no legal responsibility, whether direct or indirect, for the accuracy, completeness, usefulness or suitability of any information contained in this report.

All trademarks used in the document are owned by their respective owners. You agree not to use any trademark in or as the whole or part of your own trademarks in connection with any activities, products or services which are not ours, or in a manner which may be confusing, misleading or deceptive or in a manner that disparages us or our information, projects or developments.