TEST REPORT
Tolly.

#216112
March 2016
Commissioned by
Mellanox Technologies, Ltd.

Mellanox Spectrum vs. Broadcom StrataXGS Tomahawk

25GbE & 100GbE Performance Evaluation

Evaluating Consistency & Predictability

EXECUTIVE SUMMARY

One of the fundamental premises for building a data center, whether for Cloud or for traditional Enterprises, is that network infrastructure needs to be predictable in the way it performs. Predictability can be measured in the consistency of throughput regardless of the packet size or the type of applications the network is carrying. However, another aspect of predictability is for performance to stay consistent regardless of which ports are plugged in. A key aspect of the predictability of the network is how fairly traffic is divided when it is needed. Multiple applications and clients share the same infrastructure and when there is contention for example, when a microburst or incast (many-to-one) event occurs, the network needs to fairly divide the resources, buffers and bandwidth, in a predictable way. One application or client cannot be accidentally allowed to starve the other applications of network capacity. Unfortunately, not all switches divide traffic in a fair way.

Mellanox commissioned Tolly to benchmark the performance and predictability of the Mellanox Spectrum-based 100 Gigabit Ethernet switch and compare that to the performance and predictability of switches built by a leading network vendor with Broadcom's StrataXGS Tomahawk ASIC. The Mellanox solution delivered wire speed layer 2/3 performance with zero packet loss in tests up to 32x 100GbE ports and fairly allocated resources in incast congestion and microburst scenarios, where the Tomahawk switch failed in both cases. The Mellanox solution was able to divide bandwidth fairly. See Figure 1.

THE BOTTOM LINE

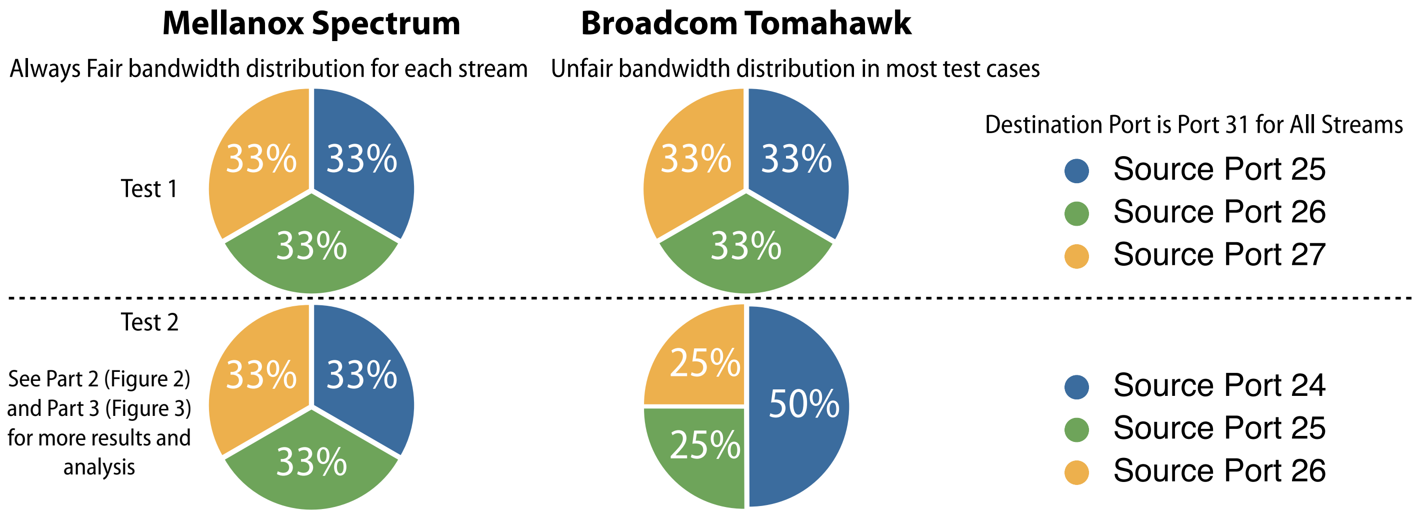
The Mellanox Spectrum ASIC delivers:

1. Predictable performance, fairly dividing traffic in all scenarios
2. Zero packet loss, wire rate performance at all packet sizes and port combinations, compared to 30% loss for Broadcom
3. Better buffering: predictable buffer allocation to any port & packet size vs. Broadcom's variance spreading by ~600%
4. Low latency, up to 90% lower latency in a typical top of rack deployment scenario

Fairness for Port Results: Bandwidth Distribution for Each Stream in Congestion

Part 1: Three 100% Line-rate Streams from Three 100GbE Ports to One 100GbE Ports

(as reported by Ixia IxNetwork 7.50.1009.20EA)



Source: Tolly, February 2016

Figure 1

Tests focused on establishing essential performance characteristics of the ASICs as implemented by a leading network vendor. Tests included fairness (in congestion), L2/L3 throughput/frame loss, microburst absorption and latency.

Tests showed significant strengths in the Mellanox Spectrum ASIC and highlighted several performance deficiencies in the Broadcom Tomahawk ASIC.

Fairness - Mellanox Spectrum distributed available bandwidth and buffers equally to all input streams (with the same input rate) in all scenarios where Broadcom Tomahawk demonstrated unfair and inconsistent results. In some cases, Broadcom Tomahawk provided 50% of the bandwidth to a single input while providing only 3% of the available bandwidth to each of the 15 remaining streams with the same input rate. In a cloud environment this behavior may lead to poor performance for tenants who lose the ability to forecast and control traffic behavior.

Frame Loss - In L2 and L3 throughput tests of 32 100GbE ports, Mellanox Spectrum delivered 100% line rate throughput with zero frame loss in at all frame sizes from 64-byte to 9216-byte jumbo frames in port-pair and full mesh scenarios.

The Broadcom Tomahawk suffered significant frame loss with frame sizes of 218-bytes and smaller. Frame loss of this nature is avoidable because it is not the result of a sustained oversubscribed scenario. 64-byte frames, Tomahawk lost 29.56% of the frames. Even with 200-byte frames, loss was 17.97%. Even when the traffic load was reduced to just six 100GbE ports in three port pairs, the Tomahawk lost packets at packet sizes of 64- and 146-bytes.

Microburst Absorption - Tests illustrated the Mellanox Spectrum buffered more than 10x as many frames in a microburst as the Broadcom Tomahawk. Furthermore, the Broadcom microburst absorption demonstrated inconsistent behavior with a buffering capacity that varied up to 6 fold in different scenarios. This behavior leads to difficulties when attempting to configure and tune the buffers and congestion control in the network, as configuration does not always affect the actual micro-burst absorption in the network using Broadcom Tomahawk.

Latency - In tests of 32 100GbE ports, Mellanox Spectrum demonstrated cut-through (first-in-first-out) L2 latency of ~300 nanoseconds (with zero frame loss) at all frame sizes tested from 64- to 9216-bytes both with and without the Mellanox forward error correction (FEC) feature operational.

By contrast, Broadcom Tomahawk latency was always at least 600 nanoseconds in the 100GbE tests. Furthermore, in tests of 25GbE, Broadcom Tomahawk reverted to store-and-forward mode resulting in significant increases in latency results to as high as more than 3 microseconds.

Test Results

Fairness

Oversubscription of an output port is inevitable at some point in all core networks. When a congestion situation occurs, such as in incast scenarios, it is important that the ASIC, in the absence of higher level quality of service (QoS) mechanisms, allocate bandwidth equally among streams thus providing "fairness".

Tolly engineers ran simple and straightforward oversubscription scenarios using real-world iMIX traffic and involving three, six and then sixteen 100GbE input ports with traffic destined for a single, oversubscribed 100GbE output port on a 32-port switch. Different switch source port combinations were used with 10 different port scenarios in all. Additional details of this and all tests can be found in the Test Methodology section of this report. Results are summarized in Figures 1-3.

Mellanox Technologies, Ltd.

Spectrum ASIC

100 Gigabit Ethernet Performance

Tolly. Certified

Tested February 2016

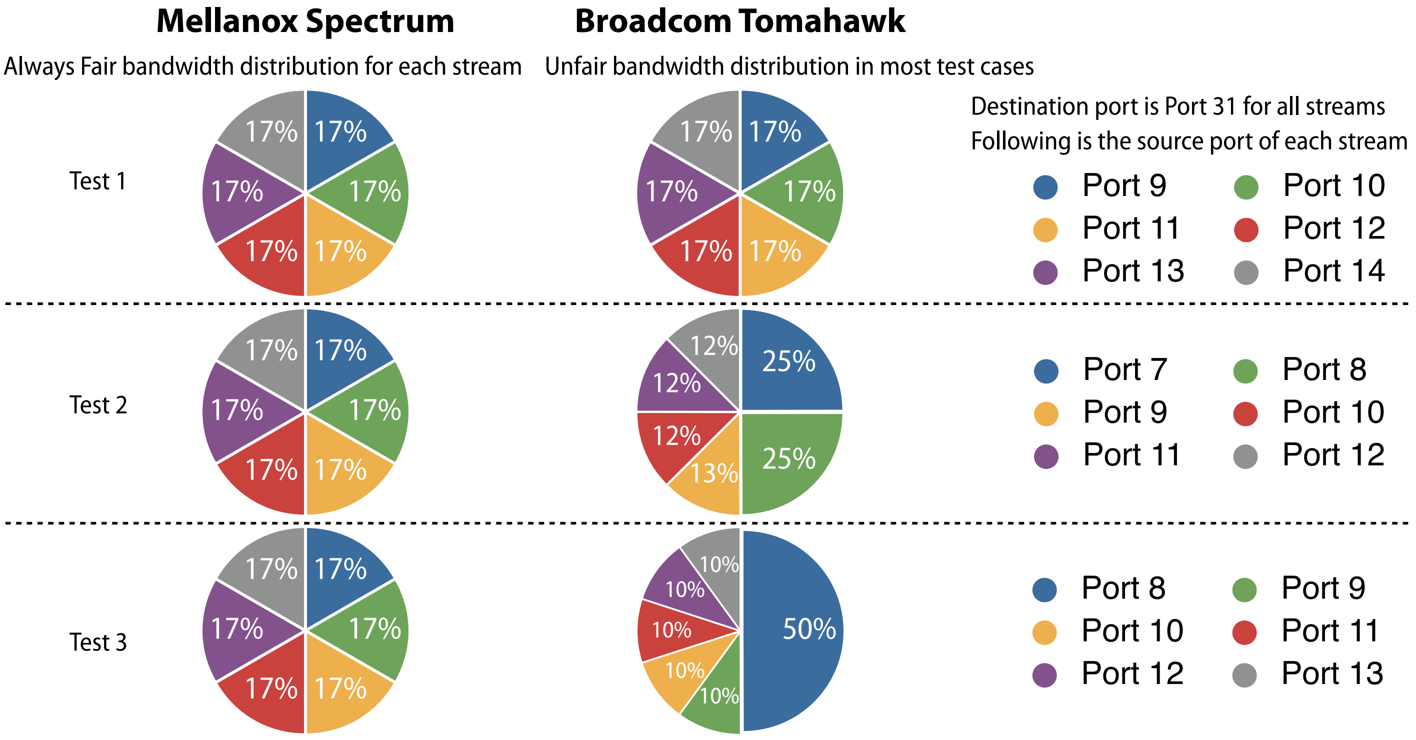
In every one of the ten different scenarios, Mellanox Spectrum distributed bandwidth equally among all input ports. Each and every input port received its fair share of bandwidth.

By contrast, Broadcom Tomahawk did not demonstrate fairness, and was unpredictable in its results. Results varied almost randomly, depending purely on which ports were sending at the same time. The results of these tests illustrate that Broadcom Tomahawk cannot be expected to deliver port-level fairness regardless of input port.

Figure 2 illustrates this point clearly. when traffic ingressed ports 9-14, the traffic was divided fairly, but when the traffic was sent through ports 7-12, two of the ports consumed twice bandwidth as the others. Still worse, when ports 8-13 were used, a single port consumed 50% of the available bandwidth.

Fairness for Port Results: Bandwidth Distribution for Each Stream in Congestion

Part 2: Six 100% Line-rate Streams from Six 100GbE Ports to One 100GbE Ports
(as reported by Ixia IxNetwork 7.50.1009.20EA)



Analysis: For Mellanox, without QoS, each stream with the same transmitting rate shares the bandwidth equally in congestion.

Note: Tolly iMIX traffic profile (Frame Size: Weight - 64:55, 78:5, 576:17, 1518:23) in IxNetwork was used in the test. Default configuration was used.

Source: Tolly, February 2016

Figure 2

While IXIA measures both bandwidth and packet rate per ingress port, our engineers expected to see the same amount of share for these two metrics: packet rate and bandwidth. While this was the result for Mellanox Spectrum, Broadcom Tomahawk showed a higher share of % packet rate than % bandwidth. At the 16−ports scenario a ~6% share bandwidth and a ~9% pps were measured on the Tomahawk switch. These findings mean that the packet loss is not fairly distributed between large and small packets, and Tomahawk unfairly drops larger packets to a greater degree. This could have negative ramifications for applications, like databases, that utilize many full size frames are sharing the network with other applications that send many small packets.

Frame Loss

Forwarding all traffic without loss is the fundamental task of any switch. Lost frames can cause unpredictable results for applications and, at a minimum, can result in delays while higher level protocols detect the loss and retransmit data.

Tolly engineers ran a series of standard RFC2544 L2 and L3, and RFC2889 full mesh test benchmarks on the systems under test using all 32 100GbE ports.

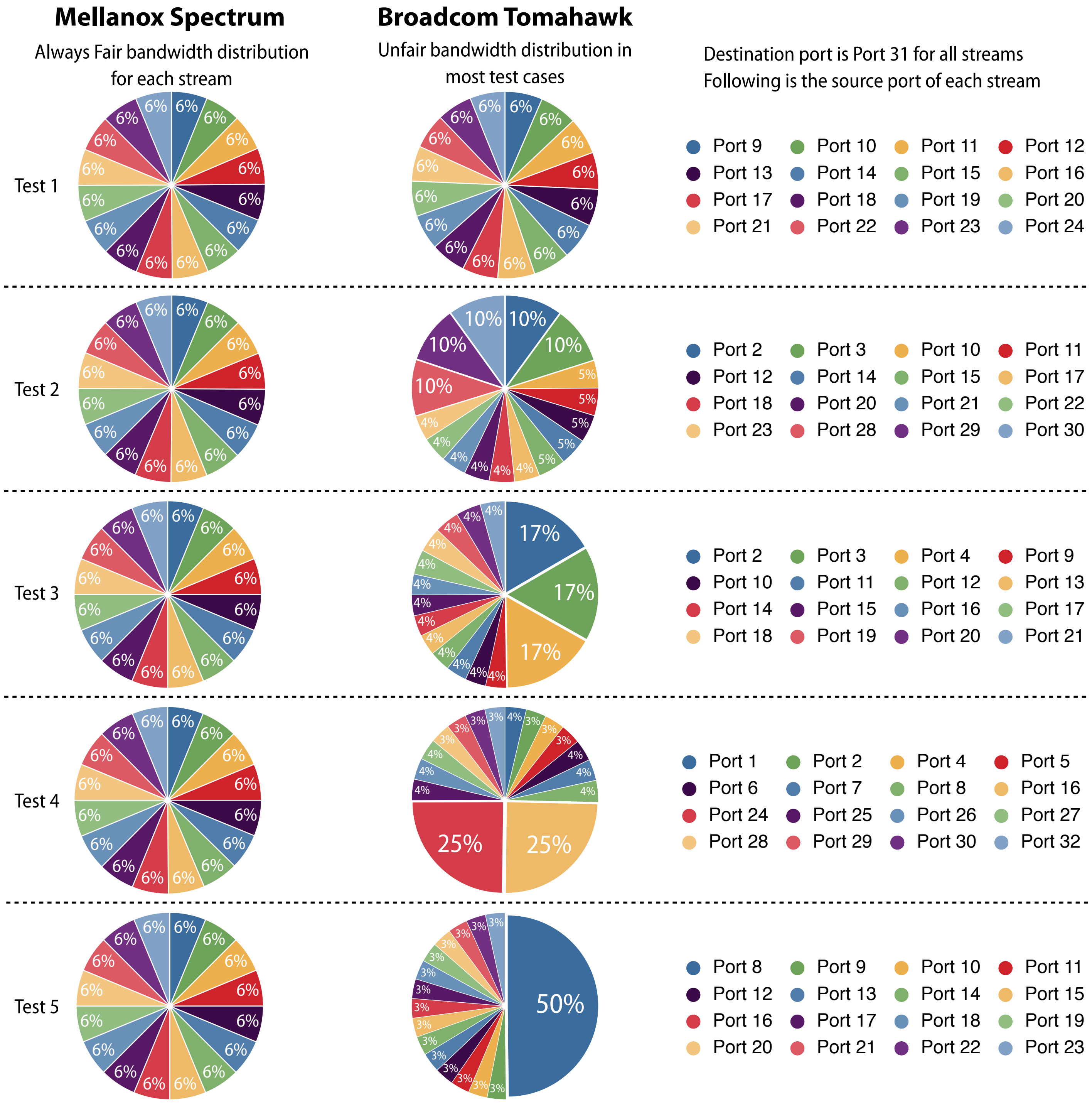
Mellanox Spectrum demonstrated zero loss in all scenarios and frame sizes from 64- through 9216-byte jumbo frames.

By contrast, Broadcom Tomahawk demonstrated 29.56% loss at 64-bytes as well as loss at various sizes up to 218-bytes including 17.97% loss at 200-bytes. See Table 1 and Figure 8.

A subset of these tests were run again using six 100GbE running RFC2544 L2 in port pairs. Mellanox demonstrated zero frame loss. Broadcom still lost 6.7% of the frames at 64-bytes and 7.52% at 146-bytes. See Table 2.

Fairness for Port Results: Bandwidth Distribution for Each Stream in Congestion

Part 3: Sixteen 100% Line-rate Streams from Sixteen 100GbE Ports to One 100GbE Ports (as reported by Ixia IxNetwork 7.50.1009.20EA)



Mellanox Spectrum

Always Fair bandwidth distribution for each stream

Broadcom Tomahawk

Unfair bandwidth distribution in most test cases

Destination port is Port 31 for all streams
Following is the source port of each stream

Note: Tolly iMIX traffic profile (Frame Size: Weight - 64:55, 78:5, 576:17, 1518:23) in IxNetwork was used in the test. Default configuration was used.

Source: Tolly, February 2016

Figure 3

Microburst Absorption

There are times when contention for an output port is momentary, for example when an incast event occurs which is common in Hadoop, CEPH, Spark, and MapReduce deployments. Microburst absorption tests measure how many frames a switch can hold in its buffer while waiting for the output port to become available. The greater the size of this buffer, the less traffic is dropped thus avoiding possible degradation of applications.

Tests showed that the microburst buffer capacity for Mellanox Spectrum was dramatically greater at all frame sizes from 64- to 9216-bytes. Tests were run on two different port configurations to determine if the results would be consistent, regardless of which ports where chosen. Unfortunately, with Broadcom the microburst capacity fluctuated depending on which ports were tested.

Mellanox results remained identical in both test configurations. With 64-byte frames, the Mellanox Spectrum demonstrated the ability to absorb 7.5x more frames than the better of the two Broadcom results. With 9216-byte jumbo frames, Mellanox delivered 4.5x that of Broadcom Tomahawk's best result. See Figure 4.

Where Mellanox provided a minimum of ~5MB of capacity for small frames, best case capacity for Broadcom Tomahawk was 0.65MB for 64-byte frames and hovered in the range of ~1MB through 200-byte frames. For larger frames of 512-bytes, Mellanox provided over 8MB of capacity compared to only 1.71MB for Broadcom Tomahawk. Across every scenario Broadcom could absorb less than half of the packets that Spectrum could absorb. See Table 4.

Latency

A switch has but one job - to move every frame across its ASIC as rapidly as possible. Dropping frames and/or excess latency (delay) can only have a negative impact on the applications that are communicating

Frame Loss Results: Mellanox Spectrum vs. Broadcom Tomahawk
32*100GbE Ports, RFC2544 and RFC2889, Layer 2/3 100% Line-rate Frame Loss Test
(as reported by Ixia IxNetwork 7.50.1009.20EA)

Frame Size (Bytes)	64	82	100	118	128	146	164	182	200	218	236	256	512	1024	1280	1518	2176	4096	9216
Mellanox Frame Loss	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Broadcom Frame Loss	29.56%	14.46%	0	0	0	30.64%	23.12%	15.59%	17.97%	0.33%	0	0	0	0	0	0	0	0	0

Notes: Transmitting rate: 100% line-rate. Three tests were run with the same results - i) RFC2544, 32*100GbE Ports in Port Pairs, Layer 2; ii) RFC2544, 32*100GbE Ports in Port Pairs, Layer 3; iii) RFC2889, 32*100GbE Ports in Full-mesh, Layer 2.

Source: Tolly, February 2016

Table 1

Frame Loss Results: Mellanox Spectrum vs. Broadcom Tomahawk
6*100GbE Ports, RFC2544, Layer 2, 100% Line-rate Frame Loss Test
(as reported by Ixia IxNetwork 7.50.1009.20EA)

Frame Size (Bytes)	64	82	100	118	128	146	164	182	200	218	236	256	512	1024	1280	1518	2176	4096	9216
Mellanox Frame Loss	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Broadcom Frame Loss	6.07%	0	0	0	0	7.52%	0	0	0	0	0	0	0	0	0	0	0	0	0

Notes: Transmitting rate: 100% line-rate. Port 1, 2, 3, 4, 5 and 6 on each switch were used in port pairs.

Source: Tolly, February 2016

Table 2

across the switch. For years, switches running port speeds even as high as 10GbE could forward even the smallest 64-byte frames without loss. As this report shows, that isn't necessarily the case with all 100GbE switch ASICs.

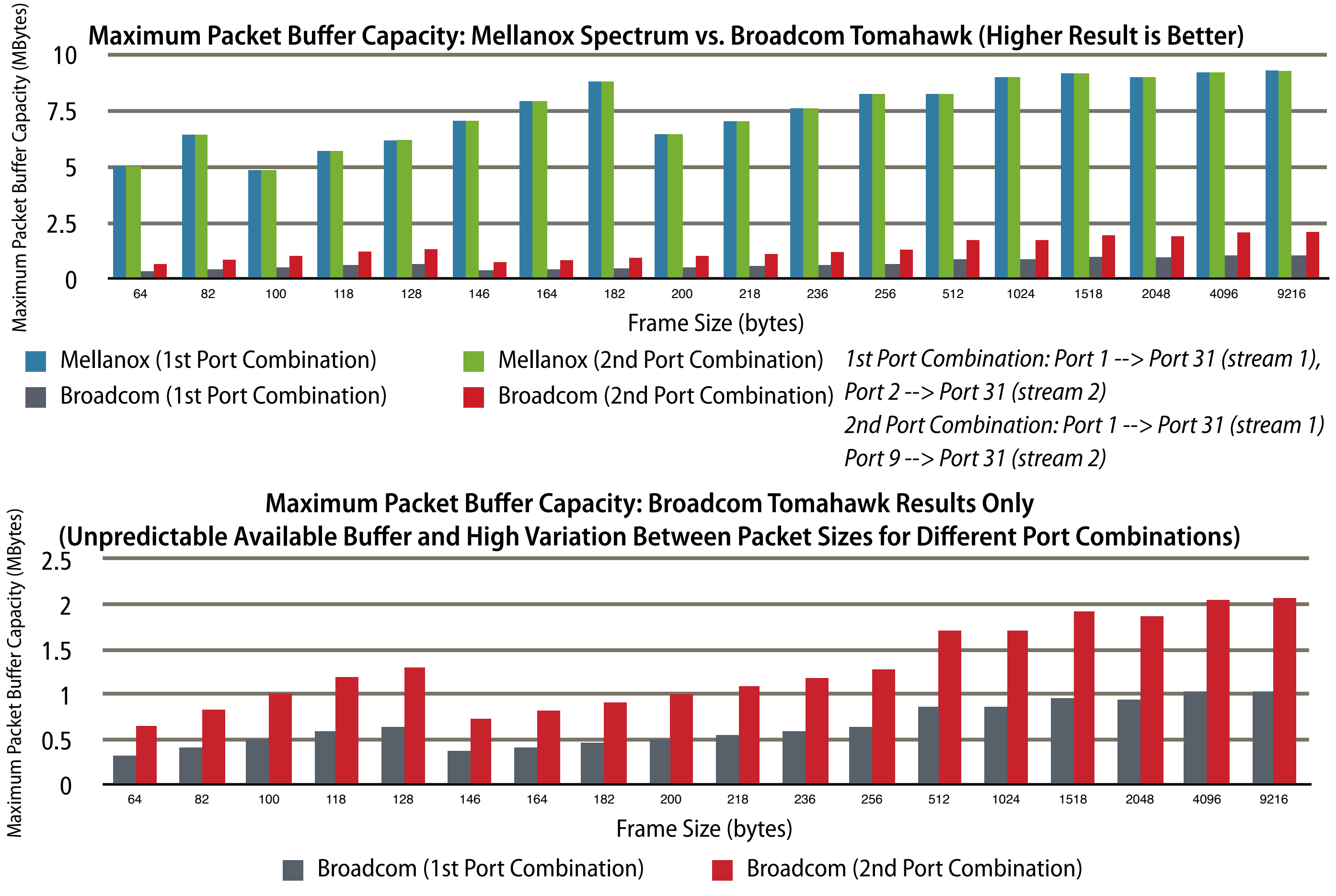
Across all test scenarios, the cut-through latency of Mellanox Spectrum is better than that of the Broadcom solution. See Table 5.

Additional testing benchmarked the performance between two 25GbE ports in a typical scenario for top-of-rack (ToR) server environments where east/west traffic between servers is common.

Testers found that the Broadcom-based solution functioned in store-and-forward for this scenario rather than in cut-through mode, despite the fact it was configured to work in cut-through mode. This resulted in dramatically higher latency for the Broadcom solution compared to the Mellanox solution that continued to operate as a cut-through switch. In the worst case of 9216-byte jumbo frames, the Broadcom solution delivered average 25GbE-to-25GbE latency of 3,334 nanoseconds compared to 336 for Mellanox. See Figures 5 and 6 and Table 5.

Microburst Absorption Capacity Results

Two 100% Line-rate bursts: from Two 100GbE Ports to One 100GbE Ports in Congestion (as reported by Ixia IxNetwork 7.50.1009.20EA)



Maximum Packet Buffer Capacity: Mellanox Spectrum vs. Broadcom Tomahawk (Higher Result is Better)

1st Port Combination: Port 1 --> Port 31 (stream 1), Port 2 --> Port 31 (stream 2)

2nd Port Combination: Port 1 --> Port 31 (stream 1) Port 9 --> Port 31 (stream 2)

Maximum Packet Buffer Capacity: Broadcom Tomahawk Results Only (Unpredictable Available Buffer and High Variation Between Packet Sizes for Different Port Combinations)

Maximum Packet Buffer Capacity (MBytes)
2.5
2
1.5
1
0.5
0
64
82
100
118
128
146
164
182
200
218
236
256
512
1024
1518
2048
4096
9216
Frame Size (bytes)
⬛ Broadcom (1st Port Combination)
🟥 Broadcom (2nd Port Combination)

Note: Default configuration was used for both switches under test. The maximum tested available buffer was 9.25MBytes for Mellanox and 2.07MBytes for Broadcom.

Source: Tolly, February 2016

Figure 4

Test Setup & Methodology

Systems Under Test

For Mellanox, the MSN2700-CS2F switch was tested. This switch had 32 ports of 100GbE and is based on the Mellanox Spectrum ASIC.

The other switch under test had 32 ports of 100GbE and is based on the Broadcom StrataXGS Tomahawk ASIC from a market leading switch vendor.

Traffic Generation

All test traffic was generated and all measurements were made using Ixia benchmarking equipment consisting of 100GbE test ports in an Ixia XG12 chassis and Ixia IxNetwork 7.50.1009.20EA.

Fairness Test

This test evaluates a scenario that multiple source ports send 100% 100Gbps line-rate traffic to one 100GbE port deliberately to create congestion. So there was one the same type of stream from each source port to the destination port. The Tolly iMIX profile in Ixia IxNetwork (Frame Size:Weight as 64:55, 78:5, 576:17, 1518:23) was used for each stream. Each stream used 100 MAC addresses, but Tolly engineers found the same result when tests were run with just a single MAC address per stream.

The default configuration of each switch was used. So engineers would expect the DUT to treat the streams fairly as the only difference for the streams is the source port and MAC address.

Engineers tried different combination of source ports to generate the traffic streams. The destination port was port 31 of the DUT for all streams.

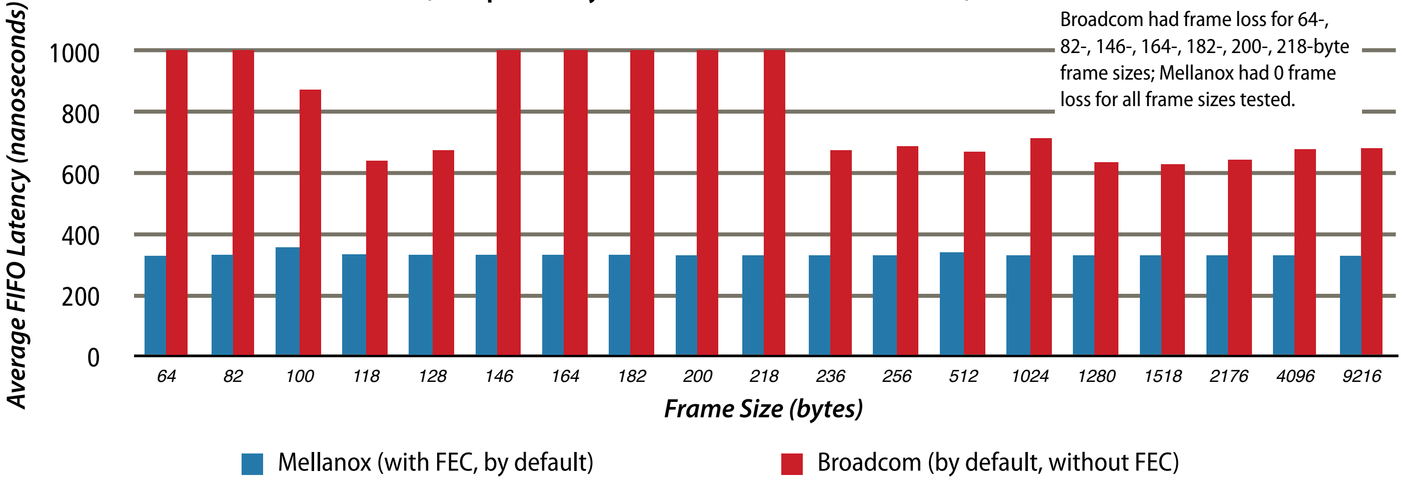
The throughput for each stream was recorded in Gbps in the Layer 2 test. The detailed results are in Table 3. The throughput in Gbps is used to analyze the fairness for source ports.

Frame Loss Test

This test evaluates the forwarding performance of the DUT.

There were 32 100GbE ports on each DUT. Engineers first evaluated the performance with all 32 ports. Three tests were run: i) RFC2544, 32*100GbE Ports in Port Pairs, Layer 2; ii) RFC2544, 32*100GbE Ports in Port Pairs, Layer 3; iii) Full mesh RFC2889, 32*100GbE Ports in Layer 2. For a line-rate

RFC2544 Cut-through Latency Results: Mellanox Spectrum vs. Broadcom Tomahawk
Layer 2 32*100GbE Ports 100% Line-rate Test (Lower Result is Better)
(as reported by Ixia IxNetwork 7.50.1009.20EA)



Notes: 1. Both switches were in cut-through mode by default. Mellanox's latency was less than Broadcom's in all tests.
2. Results reported here are the Ixia IxNetwork reported results minus 20ns due to the 4 meters cable length. FIFO latency was measured.
3. The results that reach 1,000ns are actually higher and are result of drops. Transmitting rate: 100% line-rate.
4. The Mellanox solutioin can reduce latency an additional 50ns by disabling forward error correction (FEC).

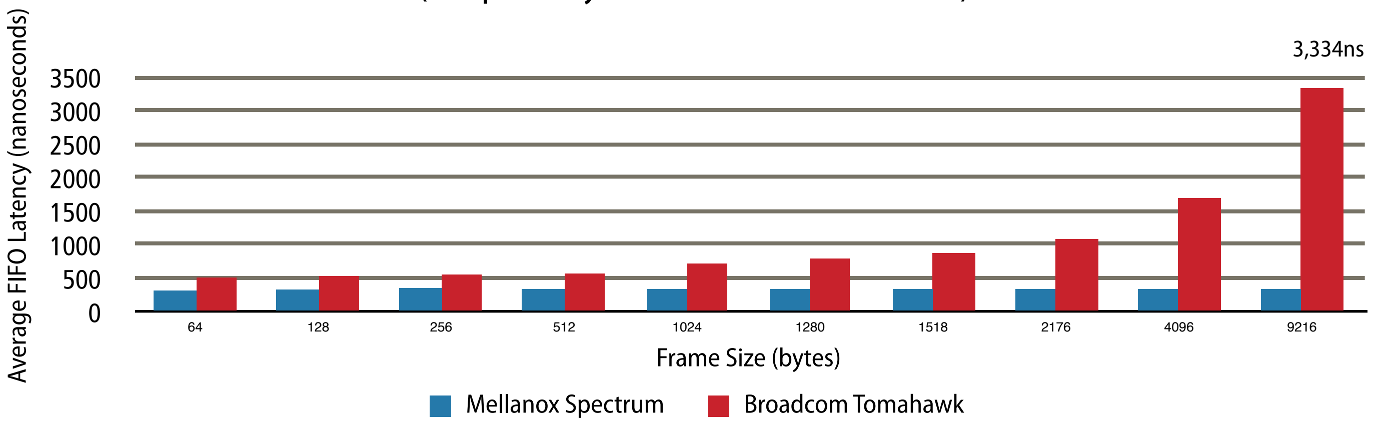
Source: Tolly, February 2016

Figure 5

25GbE-25GbE ToR Latency Results: Mellanox Spectrum vs. Broadcom Tomahawk

RFC2544, Layer 2, Lower Result is Better

(as reported by Ixia IxNetwork 7.50.1009.20EA)



Notes: 1. Both switches were configured to work in cut-through mode by default. Mellanox Spectrum was actually running cut-through while Broadcom Tomahawk actually performed store-and-forward switching. Broadcom Tomahawk supports cut-through mode, however, it appears that the Broadcom Tomahawk ASIC can only run cut-through mode when all ports are running in the same speed. So when administrators are using mixed speeds, which happens in a typical ToR design, the switch can only perform store-and-forward even between ports running the same speed.

2. Neither Mellanox nor Broadcom experienced frame loss in these tests. 25GbE ports had 100% line-rate traffic. Bidirectional traffic was used in the test. The 25GbE ports under test were split from the 100GbE ports on the switches.

3. Results reported here are the Ixia IxNetwork reported results minus 20ns due to the 4 meters cable length. FIFO latency was measured.

Source: Tolly, February 2016

Figure 6

forwarding switch, there should be no frame loss in any of these tests. 32 ports frame loss results are reported in Table 1.

Engineers then evaluated the performance with just 6 ports (the first 6 ports on each DUT). RFC2544, Layer 2, port pairs topology were used to run the test. 6 ports frame loss results are reported in Table 2.

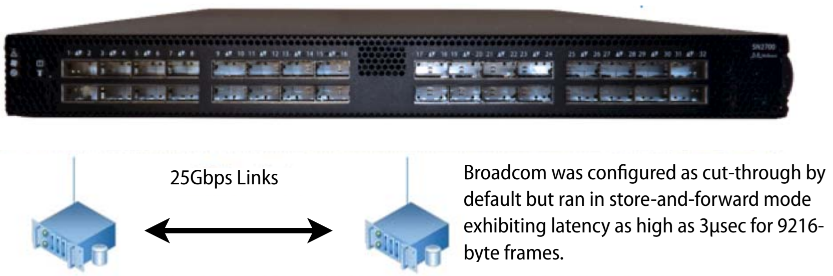
Microburst Absorption Capacity Test

This test evaluates the buffer on each DUT. Two port combinations were used to evaluate whether the available buffer is fair for streams coming from different source ports.

In each combination, there are two source ports and one destination port. Engineers sent a burst from each source port to the

ToR Latency Test Bed

Typical Data Center ToR Switch User Scenario



Broadcom was configured as cut-through by default but ran in store-and-forward mode exhibiting latency as high as 3µsec for 9216-byte frames.

Servers (simulated with the Ixia IP Performance Tester to evaluate the latency with 25Gbps line-rate traffic between them)

Note: One 100GbE port on Ixia was split into four 25GbE ports for the test. The same for the DUT.

Source: Tolly, February 2016

Figure 7

All Detailed Fairness Results - Throughput of Each Stream (Gbps)

Three/Six/Sixteen 100% Line-rate Streams from Three/Six/Sixteen 100GbE Ports to One 100GbE Ports in Congestion
Mellanox Spectrum vs. Broadcom Tomahawk (as reported by Ixia IxNetwork 7.50.1009.20EA)

Three Source Ports Test

Stream (Source Port)	*Port 25*	*Port 26*	*Port 27*
Mellanox	33.3	33.3	33.3
Broadcom	33.4	33.3	33.3
Stream (Source Port)	*Port 24*	*Port 25*	*Port 26*
Mellanox	33.3	33.3	33.3
Broadcom	50.0	25.1	25.0

Six Source Ports Test

Stream (Source Port)	*Port 9*	*Port 10*	*Port 11*	*Port 12*	*Port 13*	*Port 14*
Mellanox	16.7	16.7	16.7	16.7	16.7	16.7
Broadcom	16.7	16.7	16.7	16.7	16.7	16.7
Stream (Source Port)	*Port 7*	*Port 8*	*Port 9*	*Port 10*	*Port 11*	*Port 12*
Mellanox	16.7	16.7	16.7	16.7	16.7	16.7
Broadcom	25.0	25.0	12.6	12.5	12.5	12.5
Stream (Source Port)	*Port 8*	*Port 9*	*Port 10*	*Port 11*	*Port 12*	*Port 13*
Mellanox	16.7	16.7	16.7	16.7	16.7	16.7
Broadcom	50.0	10.1	10.0	10.0	10.0	10.0

Sixteen Source Ports Test

Stream (Source Port)	*Port 9*	*Port 10*	*Port 11*	*Port 12*	*Port 13*	*Port 14*	*Port 15*	*Port 16*	*Port 17*	*Port 18*	*Port 19*	*Port 20*	*Port 21*	*Port 22*	*Port 23*	*Port 24*
Mellanox	6.2	6.2	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3
Broadcom	6.5	6.5	6.4	6.4	6.4	6.4	6.4	6.4	6.2	6.1	6.1	6.1	6.1	6.1	6.1	6.1
Stream (Source Port)	*Port 8*	*Port 9*	*Port 10*	*Port 11*	*Port 12*	*Port 13*	*Port 14*	*Port 15*	*Port 16*	*Port 17*	*Port 18*	*Port 19*	*Port 20*	*Port 21*	*Port 22*	*Port 23*
Mellanox	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3
Broadcom	50.0	3.3	3.3	3.3	3.3	3.4	3.3	3.3	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Stream (Source Port)	*Port 2*	*Port 3*	*Port 4*	*Port 9*	*Port 10*	*Port 11*	*Port 12*	*Port 13*	*Port 14*	*Port 15*	*Port 16*	*Port 17*	*Port 18*	*Port 19*	*Port 20*	*Port 21*
Mellanox	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3
Broadcom	16.6	16.6	16.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	4.3	4.3	4.3	4.3	4.3
Stream (Source Port)	*Port 2*	*Port 3*	*Port 10*	*Port 11*	*Port 12*	*Port 14*	*Port 15*	*Port 17*	*Port 18*	*Port 20*	*Port 21*	*Port 22*	*Port 23*	*Port 28*	*Port 29*	*Port 30*
Mellanox	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.2	6.2	6.3	6.3	6.2	6.3	6.3
Broadcom	10.1	10.1	4.8	4.8	4.8	4.8	4.8	4.4	4.4	4.4	4.4	4.4	4.4	9.9	10.0	10.0
Stream (Source Port)	*Port 1*	*Port 2*	*Port 4*	*Port 5*	*Port 6*	*Port 7*	*Port 8*	*Port 16*	*Port 24*	*Port 25*	*Port 26*	*Port 27*	*Port 28*	*Port 29*	*Port 30*	*Port 32*
Mellanox	6.2	6.3	6.3	6.3	6.2	6.2	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.2
Broadcom	3.7	3.5	3.5	3.5	3.7	3.7	3.8	24.9	24.8	3.7	3.7	3.7	3.5	3.5	3.5	3.5

Note: Destination is Port 31 for all streams in all tests to generate congestion.

Source: Tolly, February 2016

Table 3

All Detailed Microburst Absorption Capacity Results - Buffer in Use (MBytes)

Mellanox Spectrum vs. Broadcom Tomahawk (as reported by Ixia IxNetwork 7.50.1009.20EA)

Frame Size (Bytes)		64	82	100	118	128	146	164	182	200	218	236	256	512	1024	1518	2048	4096	9216
Mellanox Spectrum	Test 1	4.99	6.39	4.81	5.67	6.15	7.02	7.89	8.76	6.41	6.99	7.57	8.21	8.21	8.96	9.13	8.95	9.16	9.25
	Test 2	4.99	6.39	4.81	5.68	6.16	7.02	7.89	8.76	6.41	6.99	7.57	8.21	8.21	8.96	9.12	8.96	9.16	9.23
Broadcom Tomahawk	Test 1	0.32	0.41	0.50	0.59	0.64	0.37	0.41	0.46	0.50	0.55	0.59	0.64	0.86	0.86	0.96	0.94	1.03	1.03
	Test 2	0.65	0.83	1.01	1.19	1.30	0.73	0.82	0.91	1.00	1.09	1.18	1.28	1.71	1.71	1.92	1.87	2.05	2.07

Note: Test 1 is with bursts from Port 1 --> Port 31 and Port 2 --> Port 31. Test 2 is with bursts from Port 1 --> Port 31 and Port 9 --> Port 31.

Source: Tolly, February 2016 Table 4

All Detailed Latency Results - Latency (nanoseconds)

Mellanox Spectrum vs. Broadcom Tomahawk (as reported by Ixia IxNetwork 7.50.1009.20EA)

32*100GbE Ports Test

Frame Size (Bytes)	64	82	100	118	128	146	164	182	200	218	236	256	512	1024	1280	1518	2048	4096	9216
Mellanox without Correction	284	285	311	285	285	284	285	284	283	283	283	283	292	283	283	282	283	282	281
Mellanox with Correction (default)	328	332	356	333	332	332	332	332	330	330	331	331	340	331	330	330	330	330	328
Broadcom without Correction (default)	20,399	20,400	875	641	676	5,703	5,703	5,705	4,777	4,763	676	689	670	716	637	629	645	679	682

One 25GbE Port to One 25GbE Port Test

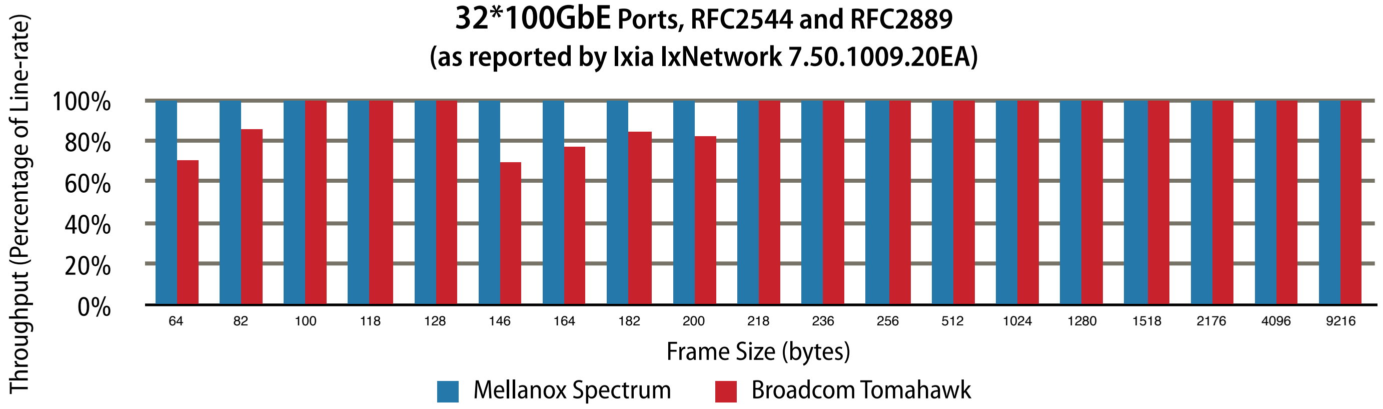
Frame Size (Bytes)	64	128	256	512	1024	1280	1518	2176	4096	9216
Mellanox Spectrum	313	333	354	337	337	337	336	336	336	336
Broadcom Tomahawk	511	528	556	567	717	793	872	1082	1694	3334

Note: Results reported here are the Ixia IxNetwork reported results minus 20ns due to the 4 meters cable length. FIFO latency was measured. Correction is FEC.

Source: Tolly, February 2016 Table 5

Throughput Results: Mellanox Spectrum vs. Broadcom Tomahawk

32*100GbE Ports, RFC2544 and RFC2889
(as reported by Ixia IxNetwork 7.50.1009.20EA)



Notes: 100% Line-rate without frame loss for all Mellanox results. Broadcom was not able to support 100% line-rate without frame loss for 64-, 82-, 146-, 164-, 182-, 200- and 218-byte frame sizes. Three tests were run with the same results - i) RFC2544, 32*100GbE Ports in Port Pairs, Layer 2; ii) RFC2544, 32*100GbE Ports in Port Pairs, Layer 3; iii) RFC2889, 32*100GbE Ports in Full-mesh, Layer 2.

Source: Tolly, February 2016 Figure 8

destination port in line-rate.

Take the Port 1 --> Port 31 and Port 2 --> Port 31 port combination for the Mellanox Spectrum based switch for example. For the 64-byte frame size, engineers sent 82,000 frames from port 1 --> port 31 and 82,000 frames from port 2 --> port 31. Without using buffer, the switch should be able to pass 82,000 frames. While using buffer, the switch passed more. In the test, the switch forwarded 81,863 + 81,845 = 163,708 frames. So the buffered frames are 163,708 - 82,000 = 81,708 frames. The Maximum Packet Buffer Capacity = 81,708 * 64 / 1024 / 1024 = 4.99MBytes.

Latency Test

This test has two parts. First, when both DUT work in 100GbE mode for all ports, engineers evaluated the cut-through latency of each DUT and compare. Second, when both DUT have 100GbE ports split into 25GbE ports as ToR switches, engineers evaluated the latency and analyze whether the switch still worked in cut-through mode or changed to store-and-forward mode.

All latency results used the latency reported by Ixia IxNetwork minus 20ns to compensate for the inherent latency of the 2x2 meter copper cables (5ns per meter).

See Figure 7 for a diagram of the latency test bed.

Devices Under Test

Mellanox MSN2700-CS2F Chassis	MLNX-OS 3.5.0530-29 Mellanox Spectrum ASIC
Broadcom Tomahawk-based Switch from a market-leading vendor	Broadcom Tomahawk ASIC

Source: Tolly, February 2016 Table 6

Test Equipment Summary

The Tolly Group gratefully acknowledges the providers of test equipment/software used in this project.

Vendor	Product	Web
Ixia	**Optixia XG12 Chassis** **8 x Xcellon-Multis QSFP28 Enhanced 100/50/25GbE Load Modules** **Software: IxNetwork 7.50.1009.20 EA**	ixia TESTED ✓ http://www.ixiacom.com

About Tolly

The Tolly Group companies have been delivering world-class IT services for more than 25 years. Tolly is a leading global provider of third-party validation services for vendors of IT products, components and services.

You can reach the company by E-mail at sales@tolly.com, or by telephone at
+1 561.391.5610.

Visit Tolly on the Internet at:
http://www.tolly.com

Interaction with Competitors

In accordance with Tolly's Fair Testing Charter, Tolly personnel invited representatives from Broadcom to review the test plan and its products results. Tolly did not receive a response to this invitation.

For more information on the
Tolly Fair Testing Charter, visit:

http://www.tolly.com/FTC.aspx

Fair Testing Charter

Terms of Usage

This document is provided, free-of-charge, to help you understand whether a given product, technology or service merits additional investigation for your particular needs. Any decision to purchase a product must be based on your own assessment of suitability based on your needs. The document should never be used as a substitute for advice from a qualified IT or business professional. This evaluation was focused on illustrating specific features and/or performance of the product(s) and was conducted under controlled, laboratory conditions. Certain tests may have been tailored to reflect performance under ideal conditions; performance may vary under real-world conditions. Users should run tests based on their own real-world scenarios to validate performance for their own networks.

Reasonable efforts were made to ensure the accuracy of the data contained herein but errors and/or oversights can occur. The test/ audit documented herein may also rely on various test tools the accuracy of which is beyond our control. Furthermore, the document relies on certain representations by the sponsor that are beyond our control to verify. Among these is that the software/ hardware tested is production or production track and is, or will be, available in equivalent or better form to commercial customers. Accordingly, this document is provided "as is," and Tolly Enterprises, LLC (Tolly) 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 herein. By reviewing this document, you agree that your use of any information contained herein is at your own risk, and you accept all risks and responsibility for losses, damages, costs and other consequences resulting directly or indirectly from any information or material available on it. Tolly is not responsible for, and you agree to hold Tolly and its related affiliates harmless from any loss, harm, injury or damage resulting from or arising out of your use of or reliance on any of the information provided herein.

Tolly makes no claim as to whether any product or company described herein is suitable for investment. You should obtain your own independent professional advice, whether legal, accounting or otherwise, before proceeding with any investment or project related to any information, products or companies described herein. When foreign translations exist, the English document is considered authoritative. To assure accuracy, only use documents downloaded directly from Tolly.com. No part of any document may be reproduced, in whole or in part, without the specific written permission of Tolly. 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.