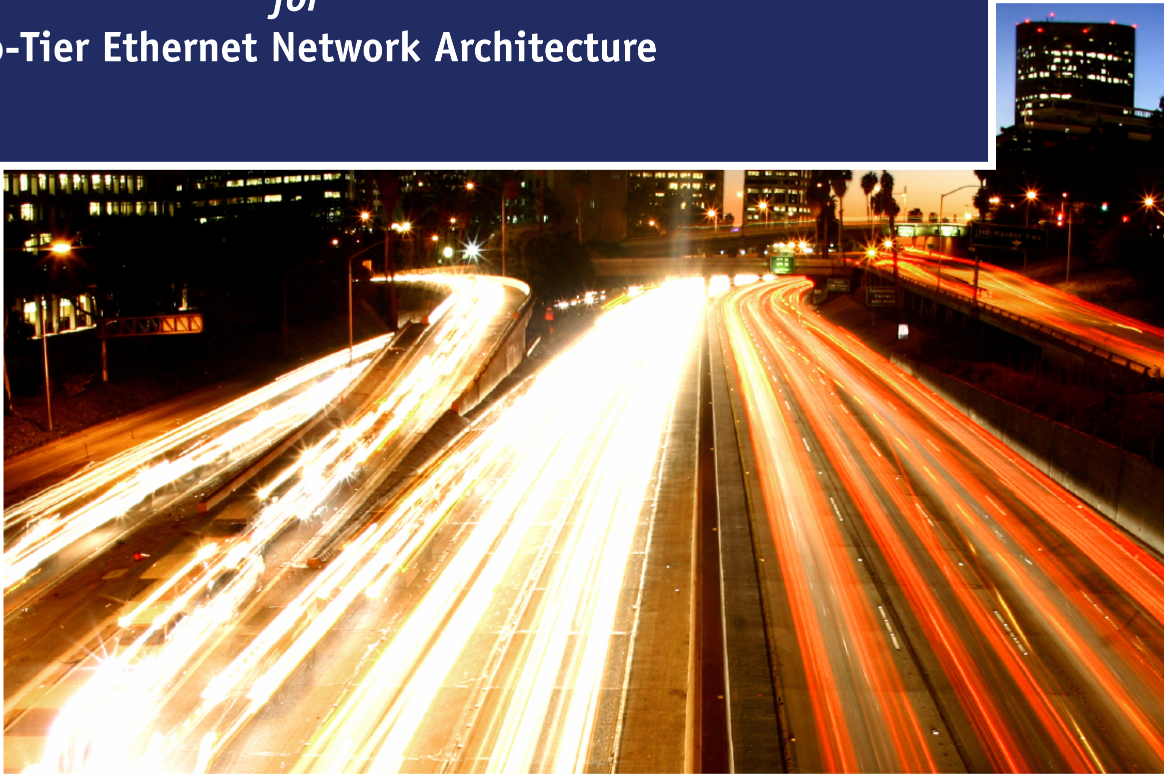
Lippis Report

Lippis Open Industry Active-Active Cloud Network Fabric Test *for* Two-Tier Ethernet Network Architecture



A Report on the:

Arista Software-Defined Cloud Network

April, 2013

Note that, currently Ixia's statistics do not support the combination of Multicast traffic running over ports within a LAG. Therefore, packet loss for this scenario was not accurately calculated and is therefore not valid.

Acknowledgements

We thank the following people for their help and support in making possible this first industry evaluation of active-active Ethernet Fabrics:

Victor Alston, CEO of Ixia—for his continuing support of these industry initiatives throughout its execution since 2009.

Leviton—for the use of fiber optic cables equipped with optical SFP+ connectors to link Ixia test equipment to 10GbE switches under test.

Siemon—for the use of copper and fiber optic cables equipped with QSFP+ connectors to link Ixia test equipment to 40GbE switches under test.

Michael Githens, Lab Program Manager at Ixia—for his technical competence, extra effort and dedication to fairness as he executed test week and worked with participating vendors to answer their many questions.

Henry He, Technical Product Manager at Ixia—for his insights into active-active networking protocols and multiple test design interactions with the vendor community.

Jim Smith, VP of Marketing at Ixia—for his support and contribution to creating a successful industry event.

Mike Elias, Photography and Videography at Ixia—for video podcast editing.

All participating vendors and their technical plus marketing teams—for their support of not only the test event and multiple test configuration file iterations but to each other, as many provided helping hands on many levels to competitors.

Bill Nicholson for his graphic artist skills that make this report look amazing.

Jeannette Tibbetts for her editing that makes this report read as smooth as a test report can.

Table of Contents

Executive Summary

To assist IT business leaders with the design and procurement of their private or public data center cloud fabrics, the Lippis Report and Ixia have conducted an open industry evaluation of Active-Active Ethernet Fabrics consisting of 10GbE (Gigabit Ethernet) and 40GbE data center switches. In this report, IT architects are provided the first comparative Ethernet Fabric performance and reliability information to assist them in purchase decisions and product differentiation.

The Lippis test report based on independent validation at Ixia's iSimCity laboratory communicates credibility, competence, openness and trust to potential buyers of Ethernet Fabrics based upon active-active protocols, such as TRILL, or Transparent Interconnection of Lots of Links, and SPBM, or Shortest Path Bridging MAC mode, and configured with 10GbE and 40GbE data center switching equipment. Most suppliers utilized MLAG, or Multi-System Link Aggregation, or some version of it, to create a two-tier fabric without an IS-IS (Intermediate System to Intermediate System) protocol between switches. The Lippis/Ixia tests are open to all suppliers and are fair, thanks to well-vetted custom Ethernet Fabric tests scripts that are repeatable. The Lippis/Ixia Active-Active Ethernet Fabric Test was free for vendors to participate and open to all industry suppliers of 10GbE, 40GbE and 100GbE switching equipment, both modular and fixed configurations.

This report communicates test results that took place during the late autumn and early winter of 2012/2013 in the modern Ixia test lab, iSimCity, located in Santa Clara, CA. Ixia supplied all test equipment needed to conduct the tests while Leviton provided optical SPF+ connectors and optical cabling. Siemon provided copper and optical cables equipped with QSFP+ connectors for 40GbE connections. Each Ethernet Fabric supplier was allocated lab time to run the test with the assistance of an Ixia engineer. Each switch vendor configured its equipment while Ixia engineers ran the test and logged resulting data.

The tests conducted were an industry first set of Ethernet Fabric Test scripts that were vetted over a six-month period with participating vendors, Ixia and Lippis Enterprises. We call this test suite the Lippis Fabric Benchmark, and it consisted of a single- and dual-homed fabric configuration which three traffic types, including multicast, many-to-many or unicast mesh and unicast from multicast returns. The Lippis Fabric Benchmark test suite measured Ethernet Fabric latency in both packet size iterations and Lippis Cloud Simulation modes. Reliability or packet loss and packet loss duration was also measured at various points in the Fabric. The new Lippis Cloud Simulation measured latency of the fabric as traffic load increased from 50% to 100% consisting of north-to-south plus east-to-west traffic flows.

Ethernet Fabrics evaluated were:

Arista Software-Defined Cloud Network (SDCN) consisting of its 7050S-64 10/40G Data Center Switch ToR with 7508 Core Switches

Avaya Virtual Enterprise Network Architecture or VENA Fabric Connect consisting of its VSP 7000

Brocade Virtual Cluster Switching or VCS consisting of its VDX™ 6720 ToR and VDX™ 8770 Core Switch

Extreme Networks Open Fabric consisting of its Summit® X670V ToR and BlackDiamond X8 Core Switch

The following lists our top ten findings:

1. **New Fabric Latency Metrics:** The industry is familiar with switch latency metrics measured via cut-through or store and forward. The industry is not familiar with fabric latency. It's anticipated that fabric latency metrics reported here will take some time for the industry to digest. The more fabrics that are tested, the greater utility of this metric.

2. **Consistent Fabric Performance:** We found, as expected, that fabric latency for each vendor was consistent, meaning that as packet sizes and loads increased, so did required processing and thus latency. Also, we found that non-blocking and fully mesh configurations offered zero fabric packet loss providing consistency of operations.

3. **No Dual-Homing Performance Penalty:** In fully meshed, non-blocking fabric designs, we found no material fabric latency difference as servers were dual homed to different Top-of-Racks (ToRs). Fabric latency measurement in dual homed were the same as single-homed configuration (as expected) even though increased reliability or availability was introduced to the design via dual-homing server ports to two ToRs plus adding MLAGs between ToRs. This was true in both Arista and Extreme's test results.

4. **CLI-Less Provisioning:** Brocade's VCS, which is TRILL based, offered several unique attributes, such as adding a switch to its Ethernet Fabric plus bandwidth between switches without CLI (Command Line Interface) provisioning. Fabric configuration was surprisingly simple, thanks to its ISL, or Inter-Switch Link Trunking.

5. **Different Fabric Approaches:** In this Lippis/Ixia Active-Active Ethernet Fabric Test, different approaches to fabrics were encountered. Avaya tested its Distributed ToR, or dToR, as part of its Fabric Connect offering, which stacks of ToRs horizontally and can offer advantages for smaller data centers with dominate east-west flows. Extreme Network's Open Fabric utilizes its high performance and port dense BlackDiamond X8 switch, which enables a fabric to be built with just a few devices.

6. **ECMP n-way Scales:** MLAG at Layer 2 and Equal Cost Multi-Path, or ECMP, at Layer 3 are dominant approaches to increasing bandwidth between switches. Arista demonstrated that an ECMP-based fabric scale to 32 with bandwidth consistency among links that is bandwidth is evenly distributed between the 32 10GbE.

7. **Balanced Hashing:** We found that vendors utilized slightly different hashing algorithms, yet we found no difference in hash results. That is, we found evenly distributed traffic load between links within a LAG (Link Aggregation) during different traffic load scenarios.

8. **vCenter Integration:** All fabric vendors offer vCenter integration so that virtualization and network operations teams can view each other's administrative domains to address vMotion within L2 confines.

9. **MLAG the Path to Two Tier Data Center Networks:** MLAG provides a path to both two-tier data center networking plus TRILL and/or SPB (Shortest Path Bridging) in the future. MLAG takes traditional link aggregation and extends it by allowing one device to essentially dual home into two different devices, thus adding limited multipath capability to traditional LAG.

10. **More TRILL and SPB:** It's disappointing that there are not more vendors prepared to publicly test TRILL and SPB implementations as Avaya and Brocade demonstrated their ease of deployment and multipathing value in this Lippis/Ixia Active-Active Ethernet Fabric Test.

Market Background

Data center network design has been undergoing rapid changes in a few short years after VMware launched VM (Virtual Machine) Virtual Center in 2003. Server virtualization enabled not only efficiency of compute, but a new IT delivery model through private and public cloud computing plus new business models to emerge. Fundamental to modern data center networking is that traffic patterns have shifted from once dominant north-south or client-to-server to now a combination of north-south plus east-west or server-server and server-storage. In many public and private cloud facilities, east-west traffic dominates flows. There are many drivers contributing to this change, in addition to server virtualization, such as increased compute density scale, hyperlinked servers, mobile computing, cloud economics, etc. This simple fact of traffic pattern change has given rise to the need for few network switches or tiers, lower latency, higher performance, higher reliability and lower power consumption in the design of data center networks.

In addition to traffic shifts and changes, service providers and enterprise IT organizations have been under increasing pressure to reduce network operational cost and enable self-service so that customers and business units may provision IT needed. At the February 13, 2013, Open Networking User Group in Boston at Fidelity Investments, hosted by the Lippis Report, Fidelity showed the beginning of exponential growth in VM Creation/Deletion by business unit managers since August 2012. Reduced OpEx and self-service are driving a fundamental need for networking to be included in application, VM, storage, compute and workload auto provisioning.

To address these industry realities, various non-profit foundations have formed, including the Open Compute Project Foundation, The OpenStack Foundation, The Open Networking Foundation, etc. These foundations seek to open up IT markets to lower acquisition cost or CapEx and inject innovation, especially around auto provisioning to lower OpEx. While the foundations are developing open source software and standards, the vendor community has been innovating through traditional mechanism, including product development and standards organizations, such as the IETF, IEEE and others.

Data center networks have been ramping up to build private and public cloud infrastructure with 10/40GbE and soon 100GbE data center switches with 400GbE on the horizon. At the center of next generation data center/cloud networking design are active-active protocols to increase application performance, thanks to its lower latency and increased reliability via dual-homed, full-meshed non-blocking network fabric plus CLI-less bandwidth provisioning.

To deliver a two-tier network, spanning tree protocol (STP) is eliminated and replaced with active-active multipath links between servers and ToR switches, ToR Core switches and between, Core switches. The industry is offering multiple active-active protocol options, such as Brocade's VCS Fabric, Cisco's FabricPath, Juniper's QFabric, TRILL, SPBM and LAG Protocol. MLAG and ECMP are design approaches to limited active-active multipathing; they are widely used and central to many vendors' STP alternative strategies, but they lack CLI-less provisioning.

Ethernet fabrics are promoted to be the optimal platform to address a range of data center design requirements, including converged storage/network, network virtualization, Open Networking, including Software-Defined Networking, or SDN, and simply the way to keep up with ever-increasing application and traffic load.

In this **industry first** *Lippis/Ixia Active-Active Ethernet Fabric Test*, we provide IT architects with comparative active-active protocol performance information to assist them in purchase decisions and product differentiation. New data center Ethernet fabric design requires automated configuration to support VM moves across L3 boundaries, low latency, high performance and resiliency under north-south plus east-west flows, and minimum number of network tiers.

The goal of the evaluation is to provide the industry with comparative performance and reliability test data across all active-active protocols. Both modular switching (Core plus End-of-Row, or EoR) products and fixed ToR configuration switches were welcome.

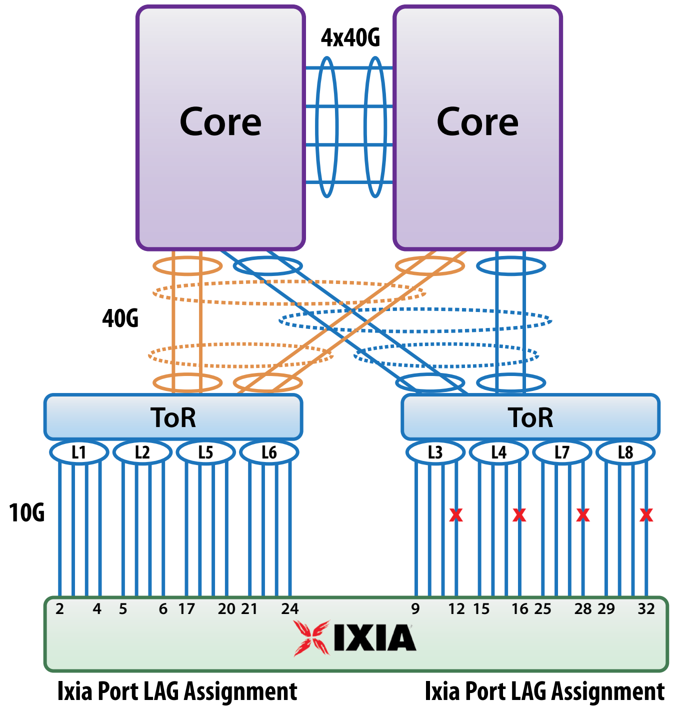
Lippis Active-Active Fabric Test Methodology

There are two active-active test configurations—single and dual homed—used in the Lippis/Ixia Active-Active Ethernet Fabric Test. These configurations consisted of two or four ToRs and two Core switches to test Ethernet fabric latency, throughput and reliability. For those vendors that offer TRILL, SPBM or FabricPath Core switches were not needed as Ixia's IxNetwork simulated these active-active protocols where latency, throughput and reliability can be measured. Most companies ran both simulated and non-simulated test runs for IS-IS based active-active and MLAG, respectively. The single- and dual-homed configurations and traffic profiles are detailed below. These configurations were used for those vendors wishing to test MLAG and/or TRILL, SPBM and FabricPath without the use of Ixia's simulation.

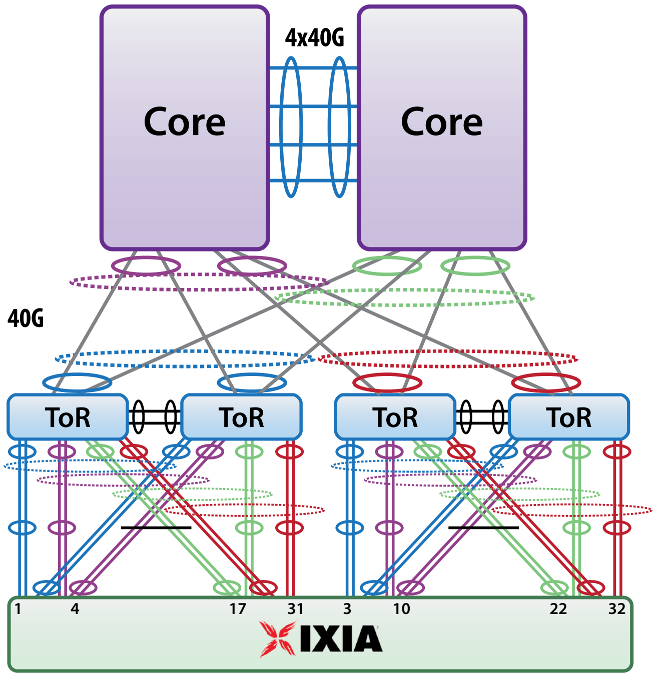
Single-Homed Configuration: In the single-homed configuration, two ToR and two Core switches made up an Ethernet fabric. Thirty-two-10GbE links connected Ixia test equipment to two ToR, which are divided into eight, four-port LAGs. Each ToR connected to two Core switches with 16-10GbE or four-40GbE links. Therefore, the load placed on this Ethernet fabric was 32-10GbE ports, or 320Gbs. A mix of unicast, multicast and mesh or any-to-any flows to represent the Brownian motion typical in modern data center networks was placed upon this fabric while latency and throughput were measured from ingress to egress; that is, from ToR-to-Core-ToR, representing fabric latency and throughput.

Dual-Homed Configuration: In the dual-homed configuration, four ToRs and two Core switches created the Ethernet fabric. Each Ixia port, acting as a virtual server, was dual homed to separate ToR switches, which is a best practice in high availability data centers and cloud computing facilities. The load placed on this Ethernet fabric was the same 32 10GbE, or 320Gbs, as in the single-homed configuration, with a mix of unicast, multicast and mesh or any-to-any flows placed upon the fabric. Each ToR was configured with eight-10GbE server ports. Eight-10GbE or two-40GbE lagged ports connect ToRs. Finally the ToRs were connected to Core switches via eight-10GbE or dual-40GbE port MLAGs. Latency and throughput were measured from ingress to egress; that is, from ToR-to-Core-ToR, representing fabric latency and throughput rather than device.

Single-Homed Topology

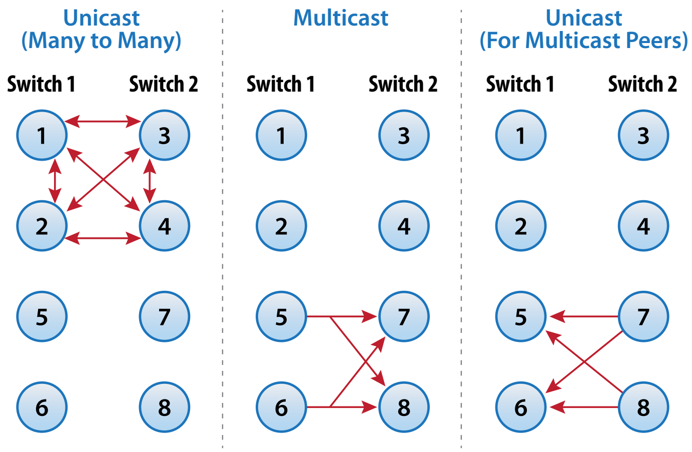


Dual-Homed Topology



Test Traffic Profiles: The logical network was an iMix of unicast traffic in a many-to-many or mesh configuration, multicast traffic and unicast return for multicast peers where LAGs are segmented into traffic types. LAGs 1, 2, 3 and 4 were used for unicast traffic. LAGs 5 and 6 were multicast sources distributing to multicast groups in LAGs 7 and 8. LAGs 7 and 8 were unicast returns for multicast peers within LAGs 5 and 6.

Traffic Profiles



Lippis Cloud Performance Test

In addition to testing the fabric with unicast, multicast and many-to-many traffic flows at varying packet sizes, the Lippis Cloud Performance Test iMix was used to generate traffic and measure system latency and throughput from ingress to egress. To understand the performance of the Ethernet fabric under load, six iterations of the Lippis Cloud Performance Test at traffic loads of 50%, 60%, 70%, 80%, 90% and 100% were performed, measuring latency and throughput on the ToR switch. The ToR was connected to Ixia test gear via 28-10GbE links.

The Lippis Cloud Performance Test iMix consisted of east-west database, iSCSI (Internet Small Computer System Interface) and Microsoft Exchange traffic, plus north-south HTTP (Hypertext Transfer Protocol) and YouTube traffic. Each traffic type is explained below:

East-West database traffic was set up as a request/response. A single 64-byte request was sent out and three different-sized responses were returned (64, 1518 and 9216 bytes). A total of eight ports were used for east-west traffic. Four ports were set as east and four ports were set as west. These eight ports were not used in any other part of the test. The transmit rate was a total 70% of line rate in each direction. The response traffic was further broken down with weights of 1/2/1 for 64/1518/9216 byte frames for the three response sizes. That is, the weight specifies what proportion from the rate set per direction will be applied to the corresponding Tx ports from the traffic profile.

East-West iSCSI traffic was set up as a request/response with four east and west ports used in each direction. Each direction was sending at 70% of line rate. The request was 64 bytes and the response was 9216 bytes.

East-West Microsoft Exchange traffic was set up on two east and two west ports. The request and response were both 1518 and set at 70% of line rate.

The following summarizes the east-west flows:

Database: 4 East (requestors) to 4 West (responders)

iSCSI: 4 East (requestors) to 4 West (responders)

MS Exchange: 2 East (requestors) to 2 West (responders)

Database/iSCSI/MS Exchange Weights: 1/2/1, i.e., 25%/50%/25% of rate set per direction and applicable on selected ports. East rate: 70% = West rate: 70%.

North-South HTTP traffic was set up on four north and four south ports. The request was 83 bytes and the response was 305 bytes. The line rate on these ports was 46.667% line rate in each direction.

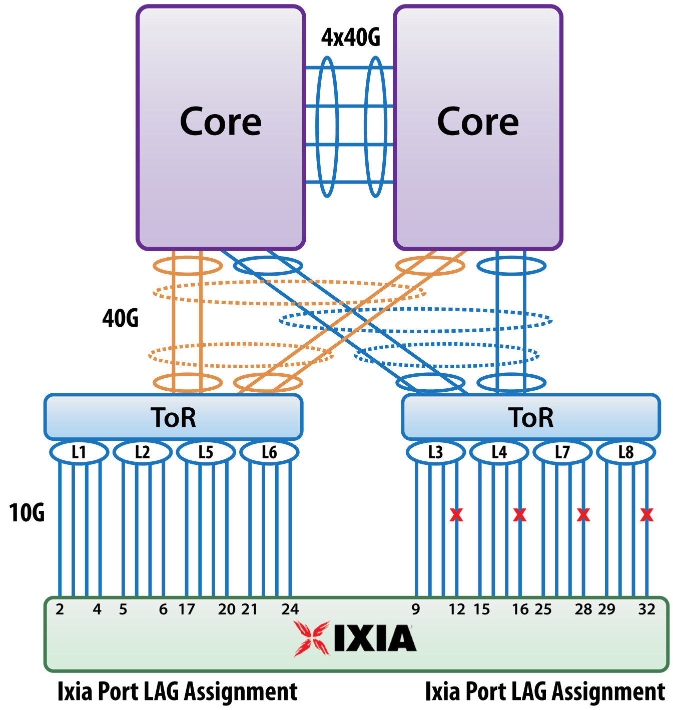
North-South YouTube traffic was using the same four north and south ports as the HTTP traffic. The request was 500 bytes at line rate of 23.333%. There were three responses totaling 23.333% in a 5/2/1 percentage breakdown of 1518, 512 and 64 bytes.

Reliability

Fabric reliability was tested in three critical areas: 1) between Ixia test gear and the ToR, 2) between ToR and core and 3) with the loss of an entire core. The system under test (SUT) was configured in the “single-homed” test design. Only the Ixia-to-ToR reliability test was required, all other reliability tests were optional.

Server to ToR Reliability Test: A stream of unicast many-to-many flows at 128-byte size packets was sent to the network fabric. While the fabric was processing this load, a 10GbE link was disconnected in LAGs 3, 4, 7 and 8 with packet loss and packet loss duration being measured and reported. Note that packet loss duration can vary as the link failure detection is based on a polled cycle. Repeated tests may show results in the nanosecond range or slightly higher numbers in the millisecond range. The poll interval for link detection is not configurable.

Single-Homed Topology



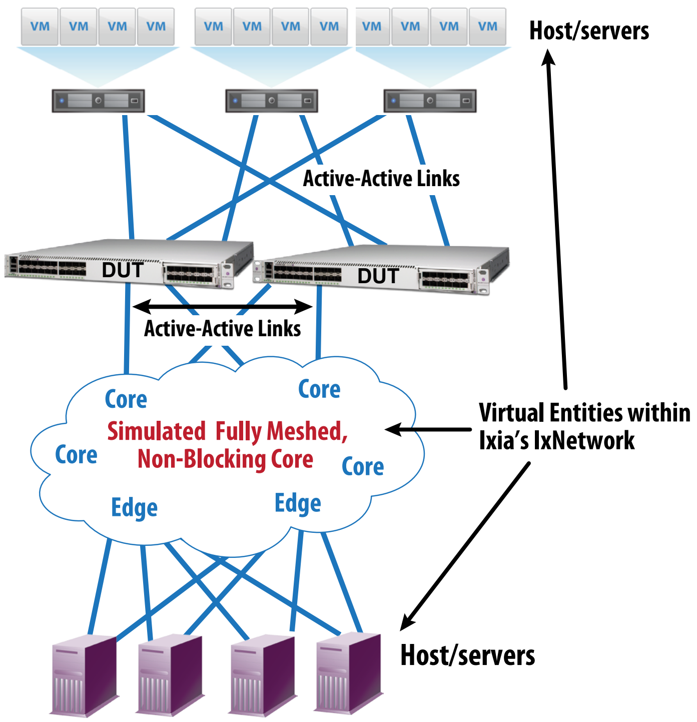
ToR to Core Reliability Test: There were two parts to this Lippis/Ixia Reliability Test. First, a link that connected ToR switches to Core switches while unicast many-to-many traffic flows were being processed was pulled with the resulting packet loss plus packet loss duration recorded by Ixia test equipment. Then the link was restored, and the resulting packet loss plus packet loss duration was recorded by Ixia test equipment. A stream of unicast many-to-many flows at 128-byte size packets was sent to the Ethernet fabric. While the fabric was processing this load, a link was disconnected, and packet loss plus packet loss duration was measured. When the link was restored, the fabric reconfigured itself while packet loss and packet loss duration was measured. Again note that packet loss duration can vary as the link failure detection is based on a polled cycle. Repeated tests may show results in the nanosecond range or slightly higher numbers in the millisecond range. The poll interval for link detection is not configurable.

Core Switch Shut Down Reliability Test: As above, packet loss and packet loss duration was measured when the fabric was forced to re-configure due to a link being shut down and restored while a 128-byte size packet of many-to-many unicast traffic flowed through the fabric. This Reliability Test measured the result of an entire Core switch being shut down and restored. Again note that packet loss duration can vary as the link failure detection is based on a polled cycle. Repeated tests may show results in the nanosecond range or slightly higher numbers in the millisecond range. The poll interval for link detection is not configurable.

Active-Active Simulation Mode Test

For those wishing to test their fabric with TRILL, SPBM and FabricPath, the Lippis/Ixia Active-Active Test offered a simulated core option. The simulated core option eliminated the need for Core switches in the configuration, requiring only ToRs for the single- and dual-homed configurations.

IxNetwork IS-IS simulated a fully meshed, non-blocking core. Two and then four ToR switches connected to the simulated core. Note that an equal number of server and core facing links were required to achieve non-blocking. ToR switches were connected in an “n-way” active-active configuration between ToR and Ixia test equipment with Ixia gear configured for the specific fabric protocol. N is a maximum of 32. The DUT was connected to Ixia test equipment with



enough ports to drive traffic equal to the "n-way" active-active links. The active-active links were connected to Ixia test equipment in core simulation mode where throughput, packet loss and latency for unicast from multicast returns, multicast and unicast many-to-many traffic were measured. The same LAG and traffic profiles as detailed above were applied to the simulation mode configuration while latency and throughput were measured for TRILL, or SPBM and/or FabricPath.

Optional Fabric Tests

The following were optional tests, which were designed to demonstrate areas where fabric engineering investment and differentiation has been made. The acquisition of the 2013 Active-Active Fabric Test report distribution license was required to participate in these optional tests; please contact nick@lippis.com to request a copy.

Most of these optional tests, if not all, were short 10-minute video podcast demonstrations as they were focused upon optional cost reduction via either ease of use or automation of network configuration.

VM Fabric Join/Remove and Live Migration Demonstration: One of the most pressing issues for IT operations is for an Ethernet fabric to support VMs. By support, we mean the level of difficulty of a VM to join or be removed from the fabric. In addition to live VM joins and removes, the ability for the fabric to support the live migration of VMs across L3 boundaries without the need for network re-configuration is a fundamental requirement.

Therefore, the objective for this test was to observe and measure how the network fabric responded during VM join/remove plus live migration. In this optional test, the vendor demonstrated the process in which a VM joins and is removed from the fabric. In addition, a VM was migrated live while we observed needed, if any, network fabric configuration changes. This demonstration was captured on video and edited into a (maximum) 10-minute video podcast. A link to the video podcast is included in the final test report. Vendors may use SPB, TRILL, FabricPath, VXLAN (Virtual Extensible Local Area Network) over ECMP, etc. Two servers, each with 30 VMs, were available for this test. Vendors were responsible for NIC (Network Interface Controller) cards plus other equipment necessary to perform this test.

East-West Traffic Flow Performance Test: Networking ToR switches so that east-west traffic flow does not need to traverse a Core switch is being proposed by various vendors as part of their network fabric. As such, an optional test ran RFC 2544 across three interconnected ToR switches with bidirectional L2/L3 traffic ingress at ToR switch 1 and egress at ToR 3. Throughput, latency and packet loss was measured with horizontal ToR latency compared to traditional ToR-Core-ToR.

Video feature: Click to view a discussion on the Lippis Report Test Methodology

Arista Networks Software-Defined Cloud Network

Arista Networks 7050S and 7508 Active-Active Test Configuration

	Hardware	Configuration/Ports	Test Scripts	Software Version
System under test	http://www.aristanetworks.com			
Single Homed	2x 7050S-64 ToR	36 10GbE/16 Each ToR		
	2x 7508 Core	32 10GbE connect ToRs		EOS 4.11
		8-10GbE MLAG between Cores		EOS 4.11
Dual Homed	4x 7050S-64 ToR	16 10GbE Each ToR: 2-10GbE MLAG between ToRs		
	2x 7508 Core	64 10GbE connect ToRs via 2-32 port LAGs		EOS 4.11
		8-10GbE MLAG between Cores		EOS 4.11
Test Equipment	Ixia XG12 High Performance Chassis		Single Home Active-Active	IxOS 6.30 EA SP2
			Dual Home Active-Active	IxNetwork 6.10 EA
			Cloud Performance Test	IxNetwork 7.0 EA & IxOS 6.40 EA
Ixia Line Cards	Xcellon Flex AP10G16S (16 port 10G module)			
	Xcellon Flex Combo 10/40GE AP (16 port 10G / 4 port 40G)			
	Xcellon Flex 4x40GEQSFP+ (4 port 40G)			
	www.ixiacom.com			
Cabling	Optical SFP+ connectors. Laser optimized duplex lc-lc 50 micron mm fiber, 850nm SPF+ transceivers			
	www.leviton.com			
	Siemon QSFP+ Passive Copper Cable 40 GbE 3 meter copper QSFP-FA-010			
	http://www.siemon.com/sis/store/cca_qsfp+passive-copper-assemblies.asp			
	Siemon Moray Low Power Active Optical Cable Assemblies Single Mode, QSFP+ 40GbE optical cable QSFP30-03 7 meters			
	http://www.siemon.com/sis/store/cca_qsfp+passive-copper-assemblies.asp			

For the Lippis/Ixia Active-Active Ethernet Fabric Test, Arista Networks built a fault-tolerant, two-tier Ethernet data center Fabric with its Software-Defined Cloud Network (SDCN) solution, consisting of 10/40GbE switches including the Arista 7050S-64 ToR switches and high performance 10 terabit per second (Tbs) capacity Arista 7508 Core Switch.

This data center Fabric design is unique on multiple levels. First, the Arista 7508 Core Switch is one of the fastest, high capacity Core switches Lippis/Ixia has tested to date with high 10GbE port density, wire speed 10GbE forwarding and just 10W per 10GbE port of power, measured in a previous Lippis/Ixia test. The Arista 7500 modular switch is available in either four- or eight-slot chassis configurations, supporting 192 or 384 wire speed 10GbE ports. Congestion points are avoided, thanks to its generous 2.3 GB packet buffer per line card module coupled with its Virtual Output Queue (VOQ) buffer-less architecture where packets are stored once on ingress.

Based upon previous Lippis/Ixia test at iSimCity, the 7504 store-and-forward latencies were measured at less than 10 microseconds for a 64B packet at L3. As a point of differentiation, the high non-blocking port density of the Arista 7508 switch enables a large data center Fabric to be built with just a few devices.

The combination of the Arista 7050S-64 ToRs and Arista 7508s form the basis of Arista's SDCN architecture. This architecture is utilized in high performance environments, such as financial Electronic Communications Networks (ECN), Financial Trading, Big Data, High Performance Computing (HPC) clusters, cloud networks and data centers.

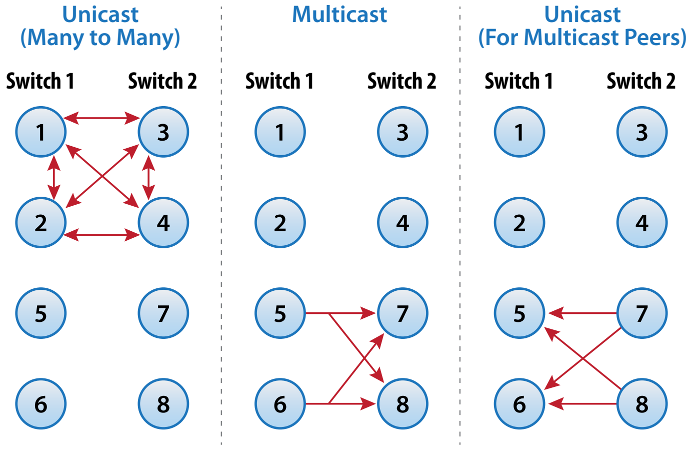
The Arista 7508 and 7050S-64 switches run Arista's Extensible Operating System (EOS), which is a single binary image for the entire Arista family of switches. With EOS, the 7508 and 7050S-64 deliver in-service-software-upgrade (ISSU), self-healing stateful fault repair (SFR), APIs to customize the software and advanced monitoring and automation capabilities for data monitoring and precise control in mission critical environments.

In terms of advanced monitoring, Arista's Latency Analyzer (LANZ) is an enhanced capability to monitor and analyze network performance, generate early warning of pending congestion events, and track sources of bottlenecks. In addition, a 50GB Solid State Disk (SSD) is available as a factory-option and can be used to store packet captures on the switch, LANZ generated historical data, and take full advantage of the Linux-based Arista EOS. Other EOS automation and monitoring modules include ZeroTouch Provisioning (ZTP) and VMTracer, which ease set-up and overall manageability of networks in virtualized data center environments.

For this Lippis/Ixia Active-Active Ethernet Fabric Test, the tested configuration consisted of Arista 7050S-64 ToR switches connecting Ixia test gear to simulate server facing devices. These, in turn, connected to the Arista 7508 switches in the Core. The Arista 7050S-64 ToRs connected to Ixia test gear with 16-10GbE links arranged as four 4x10GbE LAGs. Each Arista 7050S-64 ToR switch connected to the Arista 7508s using 16-10GbE port LAG (eight-10GbE links between each 7508 and 7050S-64) in a full-mesh configuration. The Arista 7508s were inter-connected via eight-10GbE trunks leveraging MLAG.

The logical network was a mix of unicast traffic in a many-to-many or mesh configuration, multicast traffic and unicast return for multicast peers where the LAGs were segmented into traffic types. For example, LAGs 1, 2, 3 and 4 were used for unicast traffic. LAGs 5 and 6 were multicast sources distributing to multicast groups in LAGs 7 and 8. LAGs 7 and 8 were unicast returns for multicast peers within LAGs 5 and 6.

Traffic Profiles



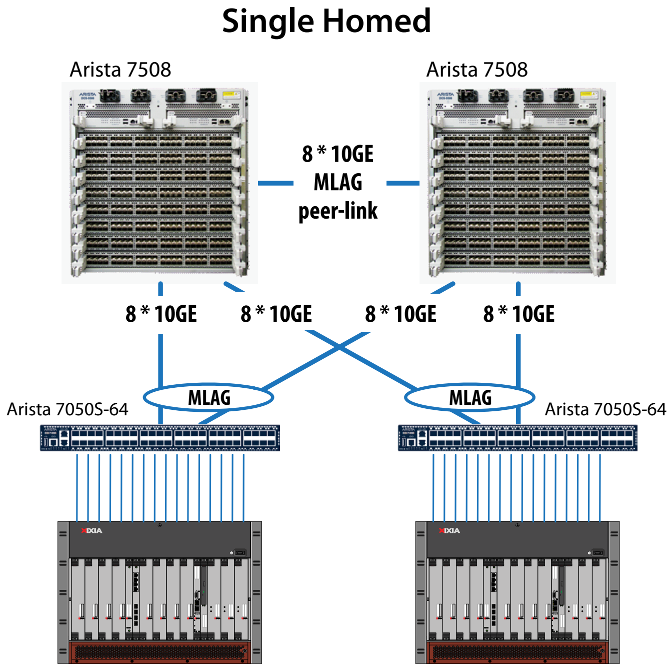
We tested this SDCN Fabric in both single-homed and dual-homed configurations and measured its overall system latency and throughput. We also tested for reliability, which is paramount, since Arista's SDCN architecture provides the option to place much of the packet processing on a few high-density Arista 7500 Core Switches instead of dozens of lower density switches. Another configuration utilized in data centers or cloud computing facilities where the numbers of servers are in the ones, tens and hundreds of thousands is to take advantage of Arista's 7508 port density and connect servers directly into a network of Arista 7508 switches.

Arista Networks demonstrated its Fabric with MLAG as the active-active protocol and thus eliminating the slower, more rudimentary STP, creating a highly efficient two-tier leaf-spine network design.

Single Homed

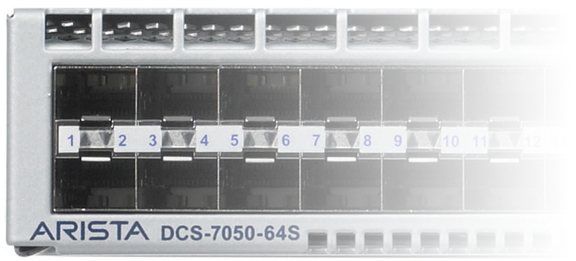
For the Single-Homed Server Lippis/Ixia Test, Arista configured two Arista 7050S-64s and two Arista 7508s for its SDCN Fabric. Thirty-two-10GbE links connected Ixia test equipment to two Arista 7050S-64s, which were divided into eight, four-port LAGs. Each Arista 7050S-64 connected to two Arista 7508s with a 16-port LAG where an eight-10GbE link connected each ToR to each core. Therefore, the load placed on this Ethernet Fabric is 32-10GbE ports, or 320Gbs, with a mix of unicast, multicast and mesh or any-to-any flows to represent the Brownian motion typical in modern data center networks.

Latency and throughput were measured from ingress to egress from Arista 7050S-64 to Arista 7508 to Arista 7050S-64, representing Fabric latency and throughput rather than a single device.



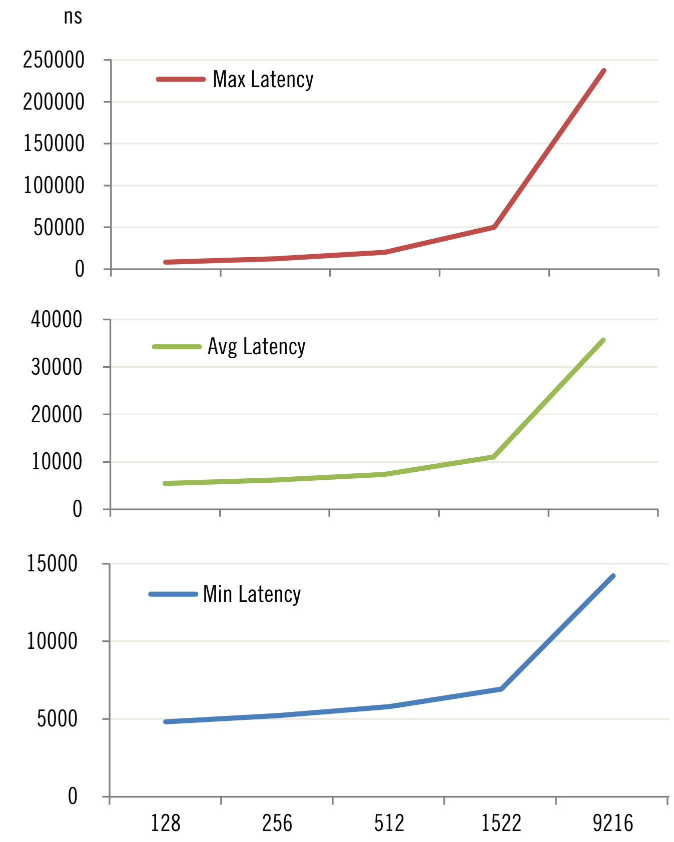
Unicast Results

Arista Networks SDCN system latency for unicast traffic varied from a high of 230 microseconds to a low of 4.8 microseconds. As expected, latency increased with packet size where 9216 size packets experienced the largest system delay. Zero packet loss was observed across all packet sizes of unicast traffic.



Arista 2-7050S-64 ToR 2-7508 Core Single Homed Test

Unicast Return From Multicast Flows (min, max, avg latency)



	128	256	512	1522	9216
Max Latency	8520	12360	20140	50000	237260
Avg Latency	5437	6183	7358	11059	35693
Min Latency	4820	5220	5800	6920	14220

Multicast Results

The Arista SDCN system latency reported by the test for multicast traffic varied from a high of 1 millisecond to a low of 5 microseconds. The latency measurement at the high end was unexpectedly high and upon further investigation it's highly probable that it was the result of buffering.

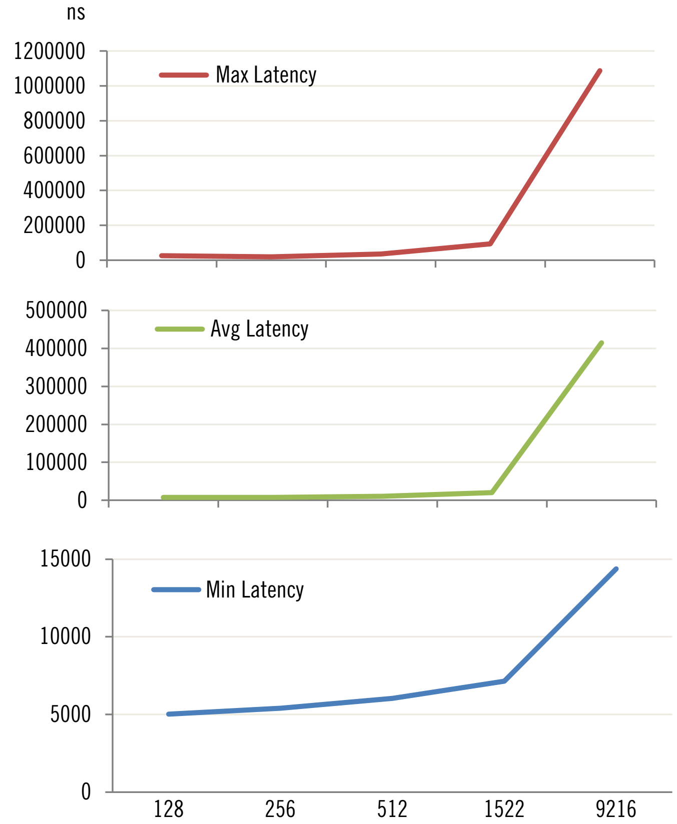
The reason for this increase is Arista 7500 has a unique architecture with a credit-based Fabric scheduler. This design allows for fairness across flows throughout the system and allows for efficient utilization of the Fabric bandwidth. Unicast traffic is buffered on ingress using VOQs. There are over 100,000 VOQs in the system divided into eight classes of traffic. By design multicast traffic cannot be buffered on ingress as multiple ports could be members of a multicast group, and the packets must be transmitted to all destination ports without dependencies on each other.

The Arista 7500 provides a very large amount of multicast bandwidth by replicating on the ingress silicon, at the Fabric and also at the egress. This three-stage replication tree allows the platform to deliver wire speed multicast to all 384 ports simultaneously. If there is congestion on the egress port, multicast packets destined to that port are buffered at egress. Other ports are not affected by this congestion and head of line blocking is avoided. When there are multiple multicast sources, and 9K packets are used, traffic is burstier and it's possible to overflow the egress buffers. Such a burst could result in dropping a small percentage of the overall traffic thus increasing measured latency. This separation of buffering in the ingress for unicast traffic and egress for multicast traffic allows the 7500 to perform well under real-world scenarios with mixed unicast and multicast traffic patterns. To obtain an accurate fabric latency measurement, Arista recommends that multicast traffic is run at a no-drop rate across all nodes.

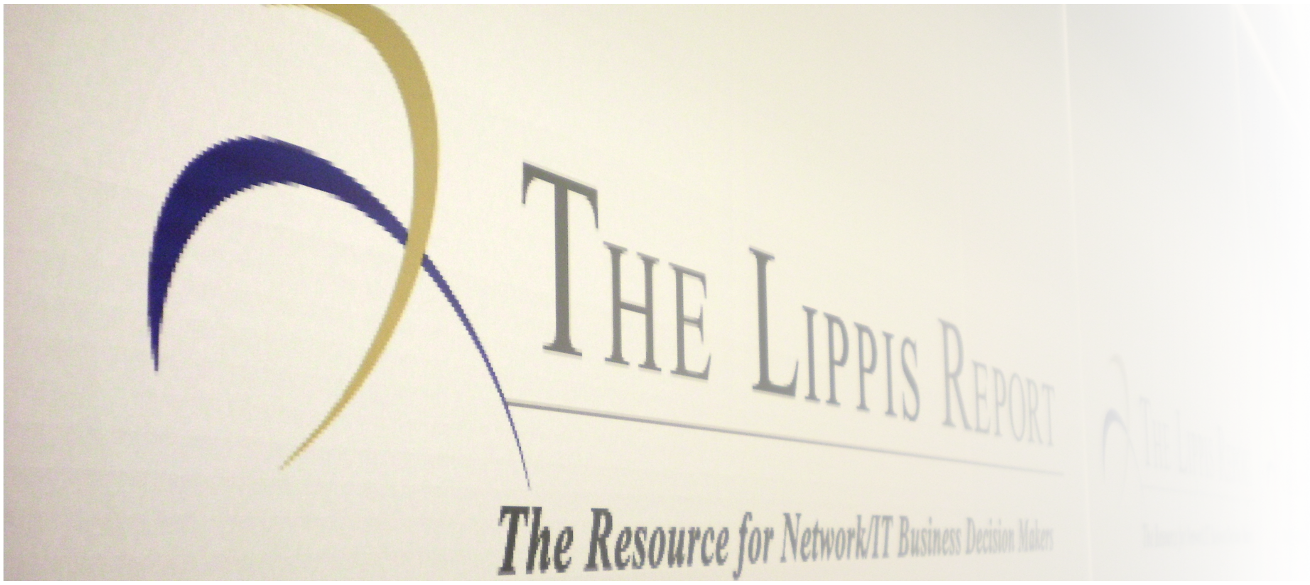
As expected, latency increased with packet size where 9216 size packets experienced the largest system delay. Note that currently Ixia's statistics do not support the combination of multicast traffic running over ports within a LAG. Therefore, packet loss for this scenario was not accurately calculated and is, therefore, not valid.

Arista 2-7050S-64 ToR 2-7508 Core Single Homed Test

Multicast Traffic (min, max, avg latency)



	128	256	512	1522	9216
Max Latency	25280	19680	35200	92980	1088200
Avg Latency	7693	7792	10302	19983	415230
Min Latency	5020	5400	6040	7140	14380

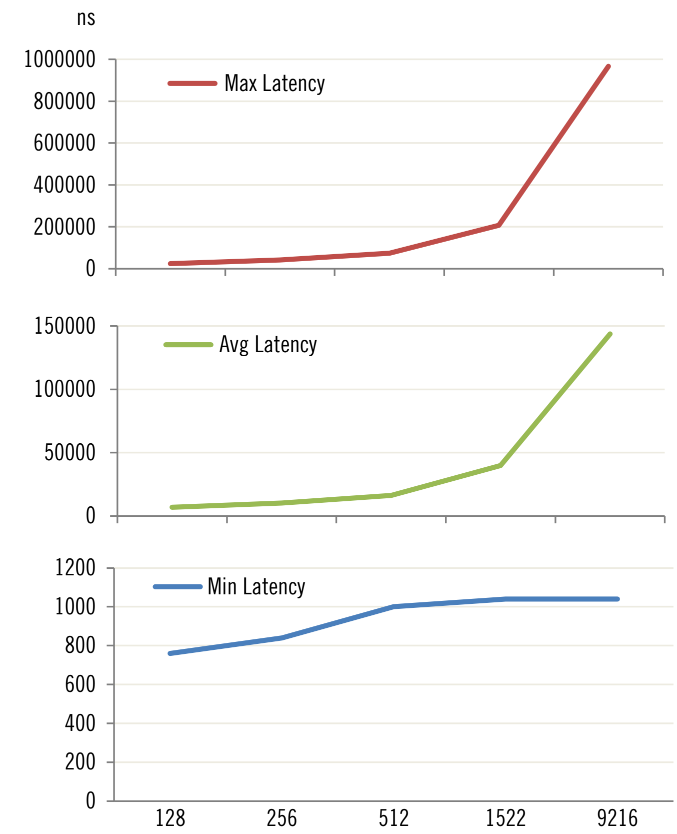


Many-to-Many Results

The Arista SDCN system latency for many-to-many unicast traffic in a mesh configuration varied from a high of 966 microseconds to a low of 760 nanoseconds. As expected, latency increased with packet size where 9216 size packets experienced the largest system delay. Zero packet loss was observed across all packet sizes of unicast traffic.

Arista 2-7050S-64 ToR 2-7508 Core Single Homed Test

Many-to-Many Full Mesh Flows (min, max, avg latency)

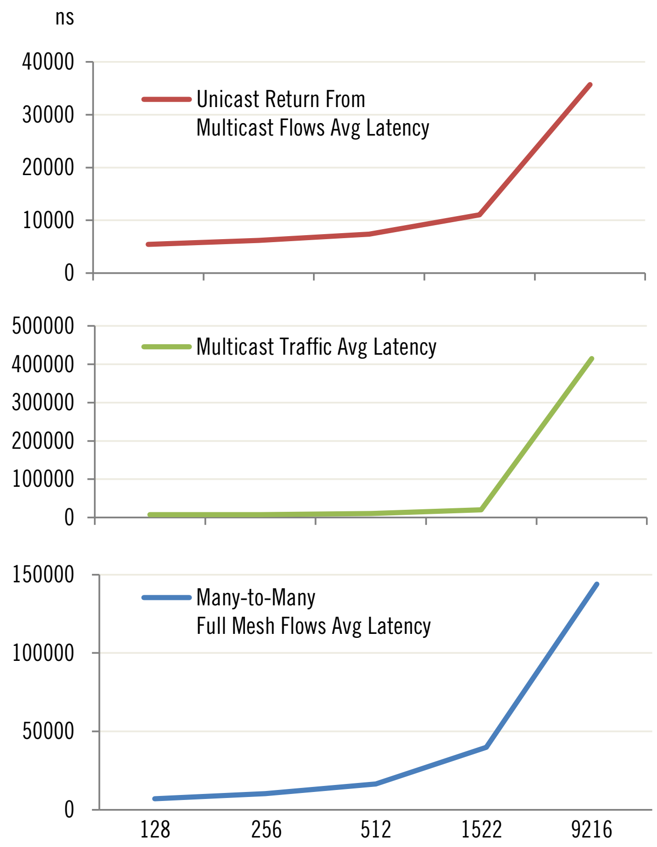


	128	256	512	1522	9216
Max Latency	24080	41200	74500	207140	966400
Avg Latency	6993	10189	16375	39780	143946
Min Latency	760	840	1000	1040	1040

The next table illustrates the average system latency across packet sizes from 128 to 9216 for unicast, multicast and many-to-many traffic flows through the Arista Networks single-homed configuration. Flows with 128-byte to 512-byte packet sizes performed in a tight latency range between 2 and 6 microseconds. The higher packet sizes of 1522 and 9216 showed wider range of latency of around a factor of six or higher. This is mostly due to the fact that the 9216-packet size was six times larger than the previous 1522 size and required substantially more time to pass through the Fabric.

Arista 2-7050S-64 ToR 2-7508 Core Single Homed Test

Unicast Return From Multicast Flows, Multicast Traffic, Many-to-Many Full Mesh Flows (avg latency)

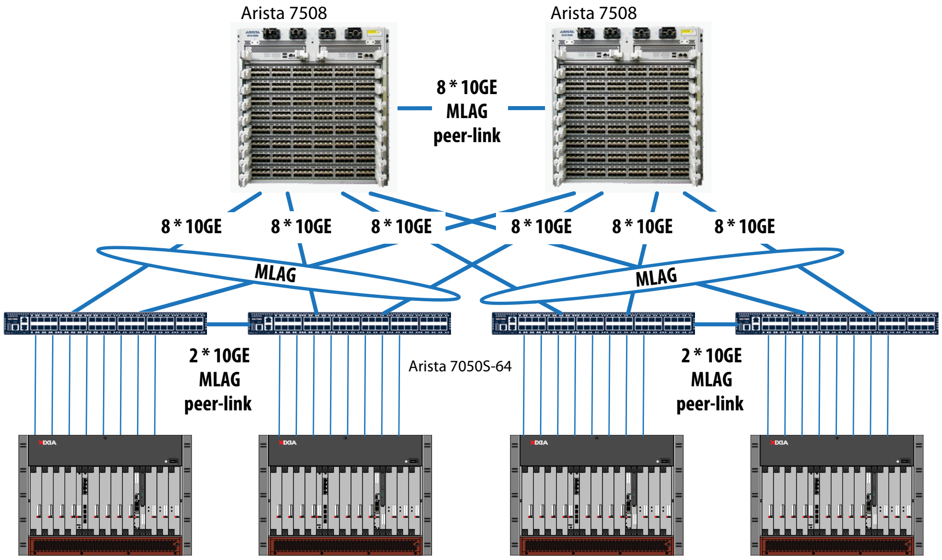


	128	256	512	1522	9216
Unicast	5437	6183	7358	11059	35693
Multicast	7693	7792	10302	19983	415230
Many-to-Many	6993	10189	16375	39780	143946

Dual Homed

For the Dual-Homed Server Lippis/Ixia Test, Arista configured four Arista 7050S-64s and two Arista 7508s for its SDCN Fabric. Ixia ports simulated virtual servers dual homed to separate Arista 7050S-64s. This active-active configuration is a best practice in high availability data centers and cloud computing facilities. The load placed on this Ethernet Fabric was the same 32 10GbE, or 320 Gbs, with a mix of unicast, multicast and mesh or any-to-any flows. Each Arista 7050S-64 was configured with eight-10GbE Ixia server ports. Two-10GbE LAG ports interconnect the Arista 7050S-64s. Finally each Arista 7050S-64 was connected to each Arista 7500 Core Switch via an eight-10GbE port MLAG. Latency and throughput were measured from ingress to egress from Arista 7050S-64 to Arista 7508 to Arista 7050S-64, representing Fabric latency and throughput rather than a single device.

Dual Homed

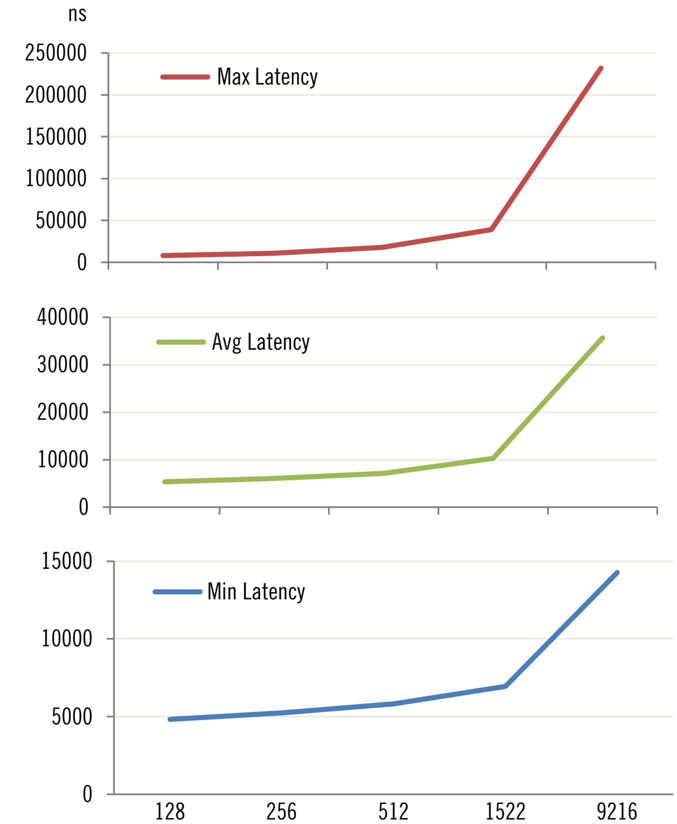


Unicast Results

The Arista SDCN system latency for unicast traffic varied from a high of 230 microseconds to a low of 4.8 microseconds. The result was the same as the single-homed configuration (as expected) even though increased reliability was part of the design at the ToR level with dual homing server ports to ToRs and MLAG connections between ToRs. In line with expectations, latency increased with packet size where 9216 size packets experienced the largest system delay. Zero packet loss was observed across all packet sizes of unicast traffic.

Arista 2-7050S-64 ToR 2-7508 Core Dual Homed Test

Unicast Return From Multicast Flows (min, max, avg latency)



	128	256	512	1522	9216
Max Latency	7980	10880	17680	39100	231780
Avg Latency	5346	6062	7111	10292	35661
Min Latency	4820	5240	5820	6940	14260

Multicast Results

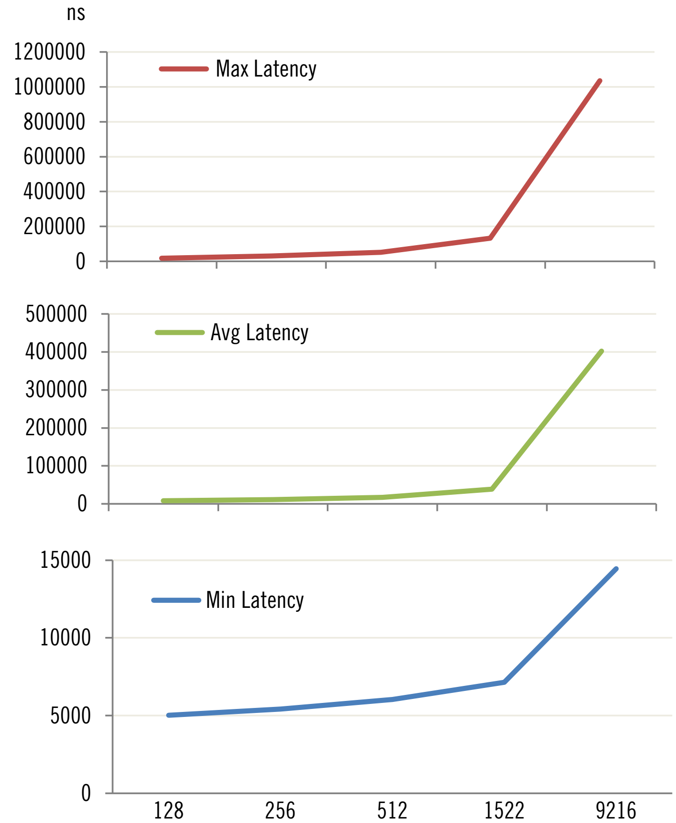
The Arista SDCN system latency for multicast traffic varied from a high of 1 millisecond to a low of 5 microseconds.

This result at the high end was unexpectedly high and upon further investigation it's highly probable that it was the result of buffering or more specifically, due to the buffer depth. Please see the single-home test results for a full explanation.

As expected, the results were the same as the single-homed configuration, even though increased reliability was introduced to the design at the ToR level. As expected, latency increased with packet size where 9216 size packets experienced the largest system delay. Note that currently Ixia's statistics do not support the combination of multicast traffic running over ports within a LAG. Therefore, packet loss for this scenario was not accurately calculated and is, therefore, not valid.

Arista 2-7050S-64 ToR 2-7508 Core Dual Homed Test

Multicast Traffic (min, max, avg latency)



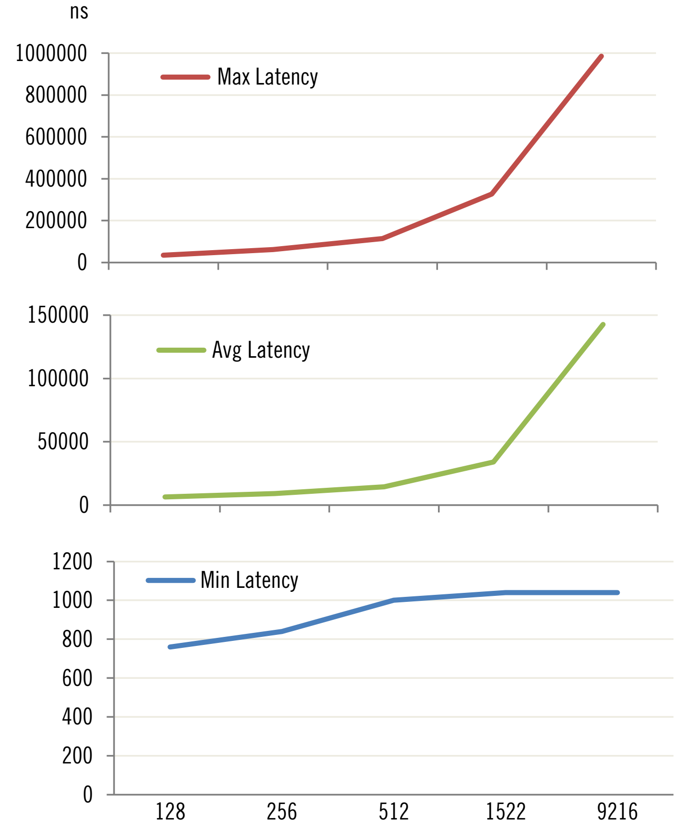
	128	256	512	1522	9216
Max Latency	18320	30380	52000	132480	1036100
Avg Latency	8104	11329	16931	38407	402035
Min Latency	5020	5420	6040	7140	14440

Many-to-Many Results

The Arista SDCN system latency for many-to-many unicast traffic in a mesh configuration varied from a high of 986 microseconds to a low of 760 nanoseconds. Again, the result was approximately the same as the single-homed configuration even though increased reliability was introduced to the design at the ToR level. As expected, latency increased with packet size where 9216 size packets experienced the largest system delay. Zero packet loss was observed across all packet sizes of unicast traffic.

Arista 2-7050S-64 ToR 2-7508 Core Dual Homed Test

Many-to-Many Full Mesh Flows (min, max, avg latency)

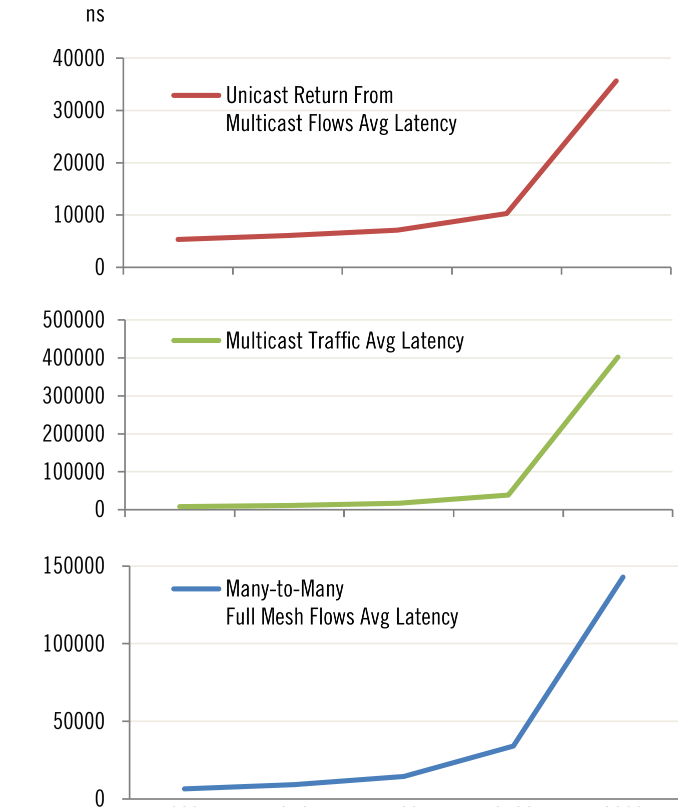


	128	256	512	1522	9216
Max Latency	35520	61400	114720	326600	986040
Avg Latency	6446	9167	14375	34007	142800
Min Latency	760	840	1000	1040	1040

The table below illustrates the average system latency across packet sizes 128-byte to 9216-byte packet sizes for unicast, multicast and many-to-many traffic flows through the Arista Networks dual-homed configuration. As expected, these results were the same as the single-homed configuration, proving that there is no latency or loss penalty for adding redundancy at the ToR level, assuming a corresponding increase in connectivity between ToR and Core is provided.

Arista 2-7050S-64 ToR 2-7508 Core Dual Homed Test

Unicast Return From Multicast Flows, Multicast Traffic, Many-to-Many Full Mesh Flows (avg latency)



	128	256	512	1522	9216
Unicast	5346	6062	7111	10292	35661
Multicast	8104	11329	16931	38407	402035
Many-to-Many	6446	9167	14375	34007	142800

Cloud Performance Test

In addition to testing the Arista SDCN with unicast, multicast and many-to-many traffic flows at varying packet sizes, the Lippis Cloud Performance Test iMix was also used to generate traffic and measure system latency and throughput from ingress to egress. The Lippis Cloud Performance Test iMix consisted of east-west database traffic, iSCSI and Microsoft Exchange traffic, plus north-south HTTP and YouTube traffic. Each traffic type is explained above in the methodology section.

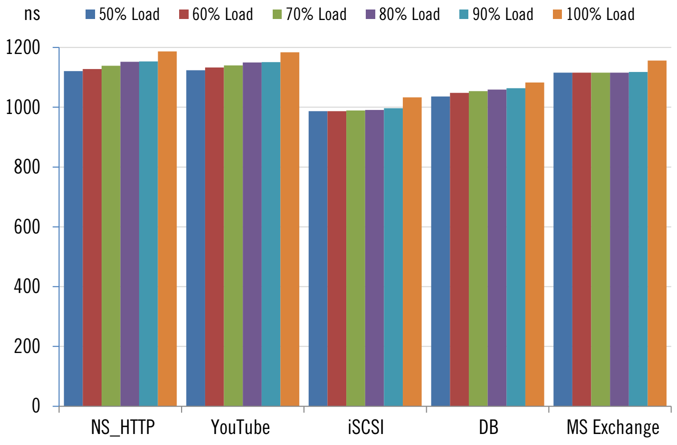
To understand the performance of Arista's SDCN system under load, we ran six iterations of the Lippis Cloud Performance Test at traffic loads of 50%, 60%, 70%, 80%, 90% and 100% measuring latency and throughput on the Arista 7050S-64 ToR switch. The Arista 7050S-64 ToR was connected to Ixia test gear via 28-10GbE links.

Lippis Cloud Performance Test Results

The Arista 7050S-64 performed flawlessly over the six Lippis Cloud Performance iterations. Not a single packet was dropped as the mix of east-west and north-south traffic increased in load from 50% to 100% of link capacity. The average latency was stubbornly consistent as aggregate traffic load was increased across all traffic types. HTTP, Microsoft Exchange, YouTube and database traffic were the longest to process with approximately 100 nanoseconds more latency than iSCSI flows. The difference in latency measurements between 50% and 100% of load across protocols was 66ns, 59ns, 47ns, 47ns and 40ns, respectively, for HTTP, YouTube, iSCSI and Database and Microsoft Exchange traffic. This was a very tight range with very impressive results as it signified the ability of the Fabric to deliver consistent performance under varying load.

Arista 2-7050S-64 ToR IxCloud Performance Test

28 ports of 10GbE (avg latency)



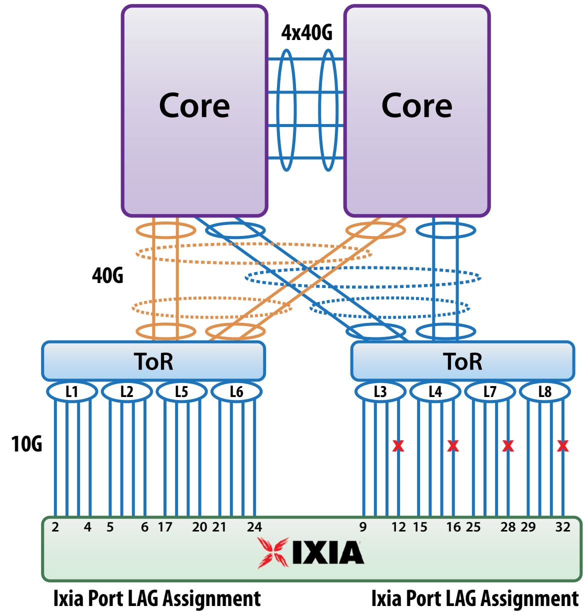
Reliability

To demonstrate and test the reliability of a data center Fabric built with Arista 7050S-64 and Arista 7508, we tested for packet loss and packet loss duration between Ixia test equipment and the Arista 7050S-64 while in the single-homed configuration.

Server to Arista 7050S-64 Reliability Test

A stream of unicast many-to-many flows at 128-byte size packets was sent to the Arista SDCN. While the fabric is processing this load, a 10GbE link was disconnected in LAGs 3, 4, 7 and 8. For many-to-many unicast traffic, packet loss was observed over 70.1 milliseconds, which, in this test, represented 0.118% of the total packets transmitted.

Single-Homed Topology



Cloud Scale and Performance Tests

Arista went above and beyond the Lippis/Ixia Active-Active Fabric Test to test the scalability of ECMP. We ran three scale tests. The first test was to demonstrate how Arista 7508s scales to support up to 15K VMs. The second test was to configure the Arista 7508 in 16-way ECMP and test its hashing algorithm to assure that traffic is well balanced across all 16-10GbE links. In the third test, we configured the Arista 7050S-64 in a 32-way ECMP design, again to test its hashing algorithm to assure that traffic is well balanced and distributed across all 32-10GbE links.

This level of ECMP scale is important as enterprises and cloud providers embrace Cloud Networking Designs. There are various choices available for active-active connectivity using standard L2 protocols (LACP) or using L3 protocols, such as OSPF (Open Shortest Path First) or BGP (Border Gateway Protocol) with ECMP.

Layer 2 designs using MLAG with up to 16 members per LAG (32 per MLAG) are supported through all Arista 7000 family products. An MLAG design provides scale by utilizing the 10+Tbps Fabric on each 7508 to provide 20 Tbps+ of active-active bandwidth at L2 for a cluster of compute and storage nodes. This can scale to 16,000 hosts in a single cluster with just one MLAG pair. Additional scale can be achieved via L3 ECMP designs.

Layer 3 ECMP is a common approach to scale and increased resiliency with just two tiers of L2/L3 switching. This approach provides non-blocking east-west and north-south bandwidth, allowing the cluster to be utilized for all applications from simple web hosting to the most advanced Big Data distributed compute jobs. The net result is improving application response times at lower cost.

Many small and large data center needs are met with both two-way and four-way multipathing, but knowing that both fixed and modular systems are designed and capable of supporting advanced scalability ensures long-term growth can be met.

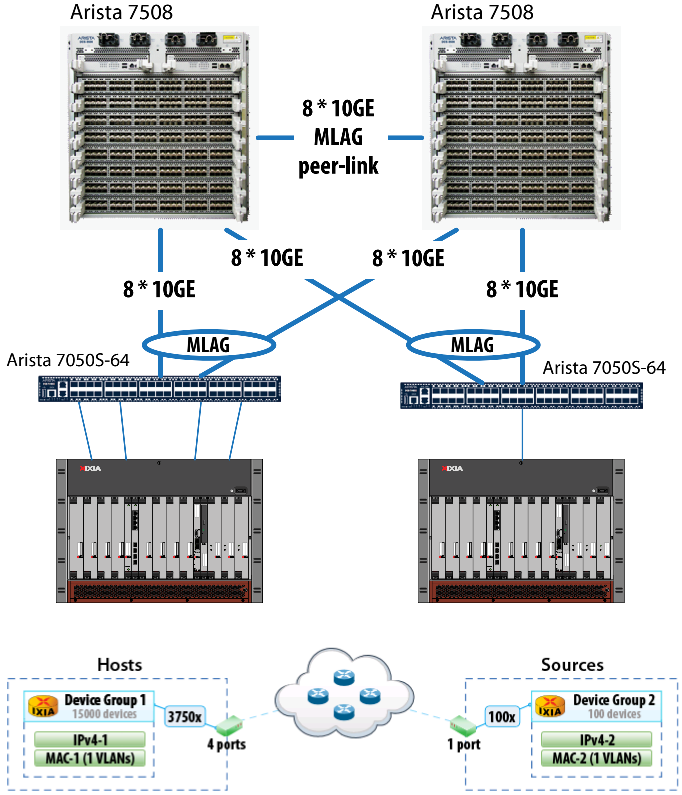
To demonstrate Arista's SDCN Fabric scales to support 15K VMs and ECMP, we configured and tested a unique VM stress test plus a 16-way and 32-way ECMP test. The results follow.

Arista Virtualization Scalability Test Results

In this configuration, Arista 7508's ability to scale VMs was tested. The Arista 7508s were configured as MLAG peers. Two 7050S-64s were connected to two 7508s using eight 10GE ports each, configured as a 16-port MLAGs. One Ixia 10GE port was connected to one of the 7050S-64s with 100 emulated hosts in VLAN (Virtual Local Area Network) 11. Four more Ixia 10GE ports were connected to the second 7050S-64 with 3,750 hosts each in VLANs 12, 13, 14 and 15, respectively. The Arista 7508s were configured with Arista Virtual ARP (Address Resolution Protocol) feature on all five VLANs to provide inter VLAN routing. The MLAGs from the 7050S-64 to the 7508s as well as the ports connected to Ixia were configured as trunks, carrying all five

VLANs to detect flooding. Bidirectional traffic flow was configured between the 100 hosts in VLAN 11 and the other 15,000 hosts. Arista 7508s successfully learned all 15,100 ARP and MAC (Media Access Control) entries and routed traffic as desired without any drops. Further, no flooding was detected. Traffic was hashing evenly across all MLAG ports in both directions.

Arista 75008 VM Scalability Test

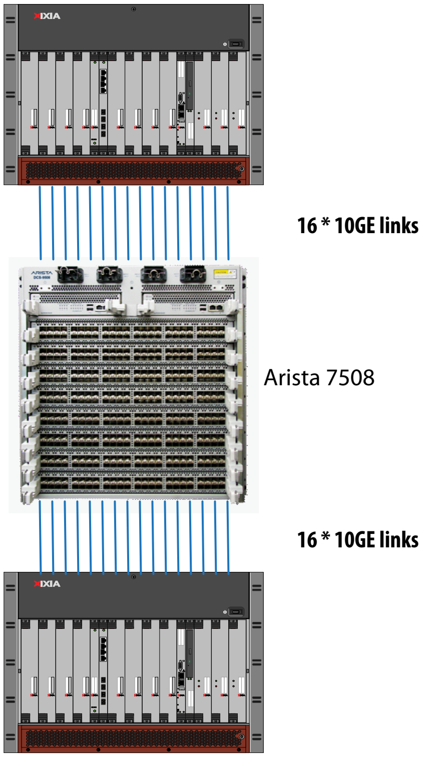


Arista 7508 16-way ECMP Test Results

To test Arista's 7508's ability to support 16-way ECMP, an Arista 7508 was connected to 32-10GE Ixia ports. Sixteen of the Ixia ports emulated 200 hosts in a single VLAN providing the traffic source. The other 16 Ixia ports emulated an OSPF router each, advertised the same IP subnet and emulated 200 hosts in the advertised route and acted as traffic destination. The Arista 7508 was configured for OSPF ECMP. The route advertised by the 16 emulated OSPF routers was correctly learned as 16-way ECMP. Each of the 16 source ports was configured to send traffic at 98.5% line rate. Arista 7508 L3 switched traffic and hashed evenly across the 16 equal cost paths with zero loss.

The theoretical expected loss per 10GbE port was 93.75%, if the hash is distributing traffic evenly across the entire 16-port ECMP. Assuming 100% of bandwidth over 16 ports calculates to 6.25 bandwidth/port or 93.75% loss per port. The table below illustrates theoretical expected loss, measured hash that, as expected, was evenly distributed.

Arista 7508 16 way ECMP test



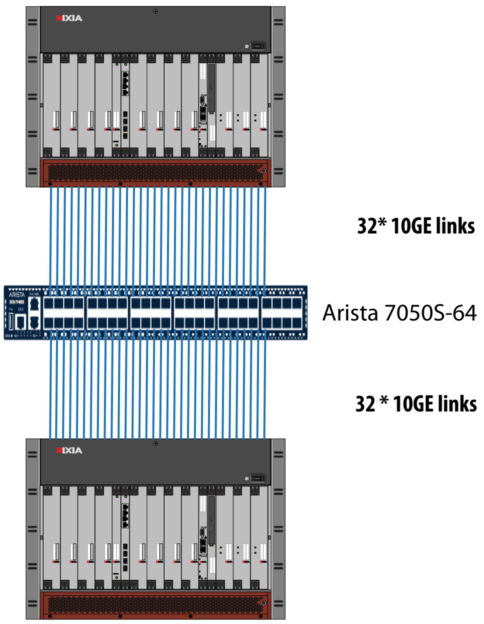
7508 16-way ECMP MLAG Test

ECMP Configuration	Theoretical Expected Hash	Measured Hash
yo409 - e3/31	93.75	93.806
yo409 - e3/39	93.75	93.788
yo409 - e4/31	93.75	93.681
yo409 - e4/39	93.75	93.681
yo409 - e5/31	93.75	93.69
yo409 - e5/39	93.75	93.74
yo409 - e6/31	93.75	93.777
yo409 - e6/39	93.75	93.775
yo409 - e7/31	93.75	93.779
yo409 - e7/39	93.75	93.806
yo409 - e8/31	93.75	93.665
yo409 - e8/39	93.75	93.798
yo409 - e9/31	93.75	93.765
yo409 - e9/39	93.75	93.794
yo409 - e10/31	93.75	93.704
yo409 - e10/39	93.75	93.752

Arista 7050S-64 32-Way ECMP Test Results

To test Arista's 7050S-64's ability to support 32-way ECMP, an Arista 7050S-64 was connected to 64 10GE Ixia ports. Thirty-two of the Ixia ports emulated 200 hosts in a single VLAN providing source traffic. The other 32 Ixia ports emulated an OSPF router each, advertised four IP subnets and emulated 200 hosts in each advertised route and acted as traffic destination. The Arista 7050S-64 was configured for OSPF ECMP. The four routes advertised by the 32 emulated OSPF routers were correctly learned as 32-way ECMP. Each of the 32 source ports was configured to send traffic at 98% line rate. Arista 7050S-64 L3 switched the traffic and hashed evenly across the 32 equal cost paths with zero loss.

Arista 7050S-64 32 way ECMP test



The theoretical expected loss per 10GbE port was 96.875%, if the hash was distributing traffic evenly across the entire 32 ECMP. Assuming 100% of bandwidth over 32 ports calculates to 3.125 bandwidth/port or 96.875% loss per port. The table below illustrates theoretical expected loss, measured hash that, as expected, was evenly distributed.

7050S-64 32-way MLAG ECMP Test

ECMP Configuration	Theoretical Expected Hash	Measured Hash
sq370 - e17	96.875	96.857
sq370 - e18	96.875	96.95
sq370 - e19	96.875	96.914
sq370 - e20	96.875	96.862
sq370 - e21	96.875	96.848
sq370 - e22	96.875	96.877
sq370 - e23	96.875	96.862
sq370 - e24	96.875	96.901
sq370 - e25	96.875	96.892
sq370 - e26	96.875	96.848
sq370 - e27	96.875	96.869
sq370 - e28	96.875	96.822
sq370 - e29	96.875	96.947
sq370 - e30	96.875	96.897
sq370 - e31	96.875	96.87
sq370 - e32	96.875	96.828
sq370 - e33	96.875	96.872
sq370 - e34	96.875	96.854
sq370 - e35	96.875	96.848
sq370 - e36	96.875	96.934
sq370 - e37	96.875	96.916
sq370 - e38	96.875	96.865
sq370 - e39	96.875	96.903
sq370 - e40	96.875	96.885
sq370 - e41	96.875	96.818
sq370 - e42	96.875	96.872
sq370 - e43	96.875	96.838
sq370 - e44	96.875	96.914
sq370 - e45	96.875	96.87
sq370 - e46	96.875	96.904
sq370 - e47	96.875	96.825
sq370 - e48	96.875	96.839

Discussion

The Arista Networks SDCN, built with its Arista 7050S-64 ToR switches and Arista 7508 Core Switches, has undergone the most comprehensive public test for data center network Fabrics and has achieved outstanding results in each of the key aspects of networking. While the SDCN architecture supports multiple, high availability network designs and configurations, during the Lippis/Ixia Active-Active Ethernet Fabric Test, a two-tier network was implemented.

We found that its system latency was very consistent under different packet sizes, payloads and traffic types. It performed to expectations; that is, large-size packet streams required more time to serialize and resulted in greater latency as they required more time to pass through the SDCN. There was no packet loss in either single- or dual-homed configurations, while it was supplied 320Gbs of unicast, multicast and many-to-many traffic types to process. In addition to processing a mix of different traffic types, Arista Network's SDCN performed outstandingly during the Lippis Cloud Performance Test processing a mix of HTTP, YouTube, iSCSI, Database and Microsoft Exchange traffic that increased in load from 50% to 100% of capacity. Here, too, its latency was consistently low with zero packet loss and 100% throughput achieved.

Arista used MLAG to implement its active-active protocol, which performed flawlessly proving that a two-tier data center network architecture built with its Arista 7050S-64 ToR and Arista 7508 Core Switches will scale with performance.

One of the surprise and delights of Arista's set of Active-Active Test results concerned its ECMP and VM scale results. The VM stress test showed that the Arista SDCN scale up to 15,000 VMs—a first in these Lippis/Ixia test at iSimCity. Not only does the Arista SDCN scale in terms of VM support, but in terms of active-active links, too. A 16- and 32-way ECMP network was tested with expected results of balanced traffic distribution between links. Both 16- and 32-way ECMP was the largest we tested at iSimCity in these Lippis/Ixia tests. These results provide comfort in the fact that Arista's SDCN can scale in to the 20K- to 100K-plus-size server data centers using ECMP and a two-tier leaf-spine network architecture.

Data center networking is moving in multiple directions of efficiency. Converged I/O hopes to reduce cabling and storage switch cost by combining both storage and data traffic over one Ethernet Fabric. The Open Networking standards approach to networking looks to reduce operational spend by centralizing network control where northbound APIs abstract network services so that applications and data center orchestration systems can automate network configuration. As

Video feature: Click to view Lippis/Francini Arista Software Defined Cloud Network VM Migration Video Podcast

In this video podcast, we demonstrate automated provisioning of network devices as new VMs are added, moved and deleted, showing synergy between virtualized compute and Arista's SDCN. We used the Arista VMTracer, which is a management tool that integrates into VMware's vCenter to give an end-end view of the virtualized data center, including VMs, physical hosts and network devices across both physical and virtual environments.

these trends develop and grow, a stable, high performing data center network infrastructure that scales becomes ever more important. Arista Networks SDCN demonstrated high performance, low latency and high scale under various loads and conditions during this Lippis/Ixia Active-Active Fabric Test.

Arista's SDCN road map includes 100GbE, OpenFlow, OpenStack, Open vSwitch, VXLAN support, multivendor API support, multiple hypervisor virtualization orchestration, a set of northbound APIs, plus cloud control including its AEM (Advanced Event-Driven Management), ZTP/ZTR (Zero Touch Provisioning/Zero Touch Replacement), LANZ and new DANZ (Data Analyzer) features. We find that Arista Networks SDCN is an excellent choice to consider for modern data center networking requirements.

Active-Active Fabric Cross-Vendor Analysis

To deliver the industry's first test suite of Ethernet fabrics Ixia, Lippis Enterprises and all vendors provided engineering resources to assure that the configuration files are suitable for MLAG, TRILL and SPB configurations. In addition, the results are repeatable, a fundamental principal in testing. This test was more challenging thanks to the device under test being a fabric or "system" versus a single product. These participating vendors are:

Arista Software-Defined Cloud Network

Avaya Virtual Enterprise Network Architecture Fabric Connect

Brocade Virtual Cluster Switching

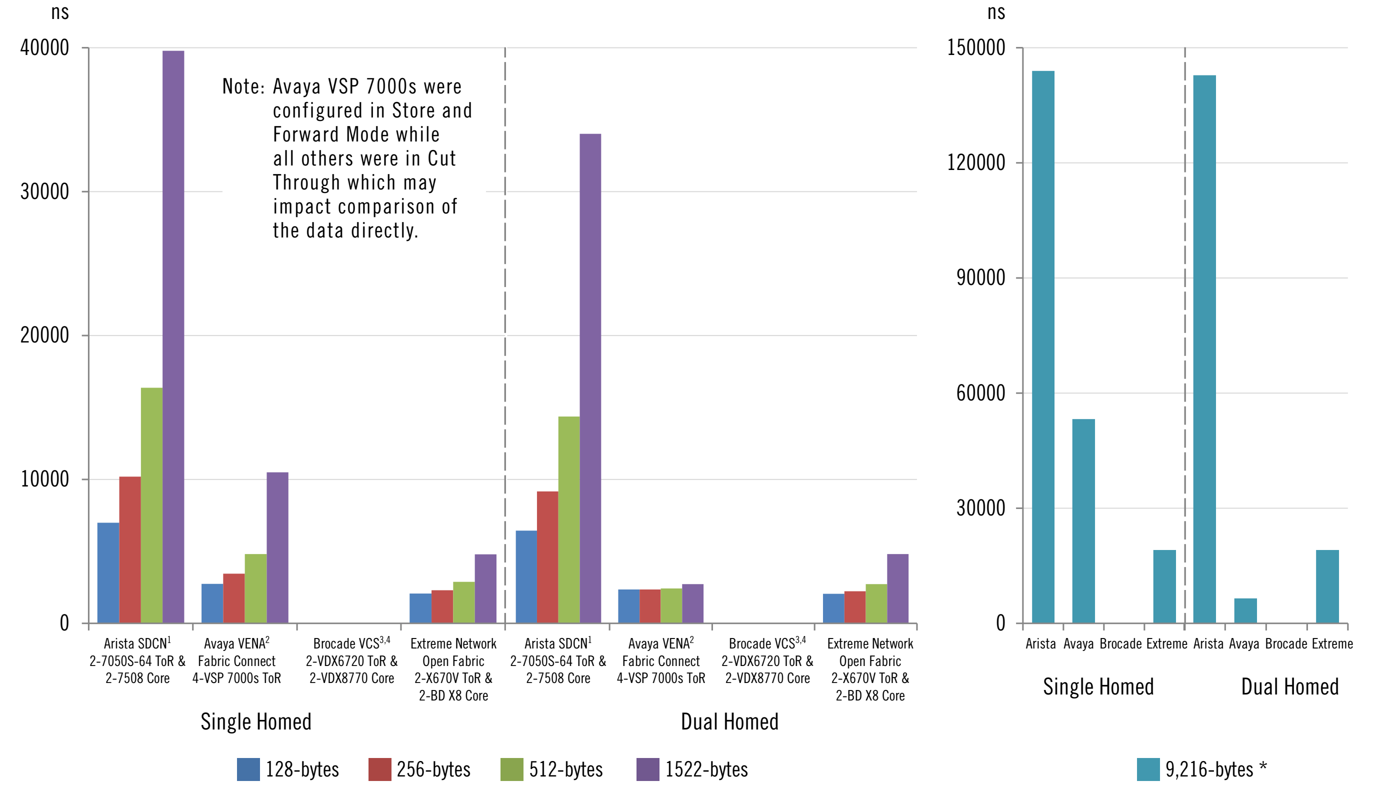
Extreme Networks Open Fabric

Brocade was the first company to be tested. The test suite evolved after its first test, thus we do not include Brocade's single and dual homed results in the cross-vendor section due to a different traffic mix utilized for Brocade's VCS. Further, each vendor was offered optional testing opportunities which some accepted and some declined. For the cross-vendor analysis we report on only required aspects of the Lippis/Ixia Active-Active Ethernet Fabric Test.

These fabric configurations represent the state-of-the-art in two-tier networking. Pricing per fabric varies from a low of $75K to a high of $150K. ToR and Core switch port density impact pricing as does 10GbE vs 40GbE. Price points on a 10GbE per port basis are a low of $351 to a high of $670. 40GbE ToR switch price per port is as low as $625 to $2,250 per port. In the Core 10GbE price per port is as low as $1,200 while 40GbE ports are as low as $6,000 per port.

We compared each of the above firms' fabrics in terms of their ability to forward packets: quickly (i.e., latency), without loss of their throughput at full line rate for three types of traffic, unicast mesh in a many-to-many configuration, multicast and unicast returns from multicast peers. We compare Server-ToR reliability and how each ToR performs during the Lippis Cloud simulation test.

Fabric Test, Many-to-Many Full Mesh Unicast - Average Fabric Latency
All Switches Performed at Zero Frame Loss



[1]Software-Defined Cloud Network
[2]Virtual Enterprise Network Architecture
[3]Virtual Cluster Switching
[4]Brocade's test was based on slightly different traffic profile and thus is not included here

*The latency measurement was unexpectedly high and highly probable that it was the result of buffering and not a true measure of fabric latency.

	Single Homed				Double Homed			
Framesize (bytes)	Arista SDCN[1] 2-7050S-64 ToR & 2-7508 Core	Avaya VENA[2] Fabric Connect 4-VSP 7000s ToR	Brocade VCS[3,4] 2-VDX & 6720 ToR 2-VDX8770 Core	Extreme Networks Open Fabric 2-X670V ToR & 2-BD X8 Core	Arista SDCN[1] 2-7050S-64 ToR & 2-7508 Core	Avaya VENA[2] Fabric Connect 4-VSP 7000s ToR	Brocade VCS[3,4] 2-VDX & 6720 ToR 2-VDX8770 Core	Extreme Networks Open Fabric 2-X670V ToR & 2-BD X8 Core
128	6993	2741	Brocade's test was based on a slighty different traffic profile and thus is not included here.	2065	6446	2356	Brocade's test was based on a slighty different traffic profile and thus is not included here.	2045
256	10189	3447		2297	9167	2356		2223
512	16375	4813		2887	14375	2428		2730
1,522	39780	10491		4789	34007	2724		4882
9,216	143946	53222		19115	142800	6258		19121

[1]Software-Defined Cloud Network
[2]Virtual Enterprise Network Architecture
[3]Virtual Cluster Switching
[4]Brocade's test was based on slightly different traffic profile and thus is not included here

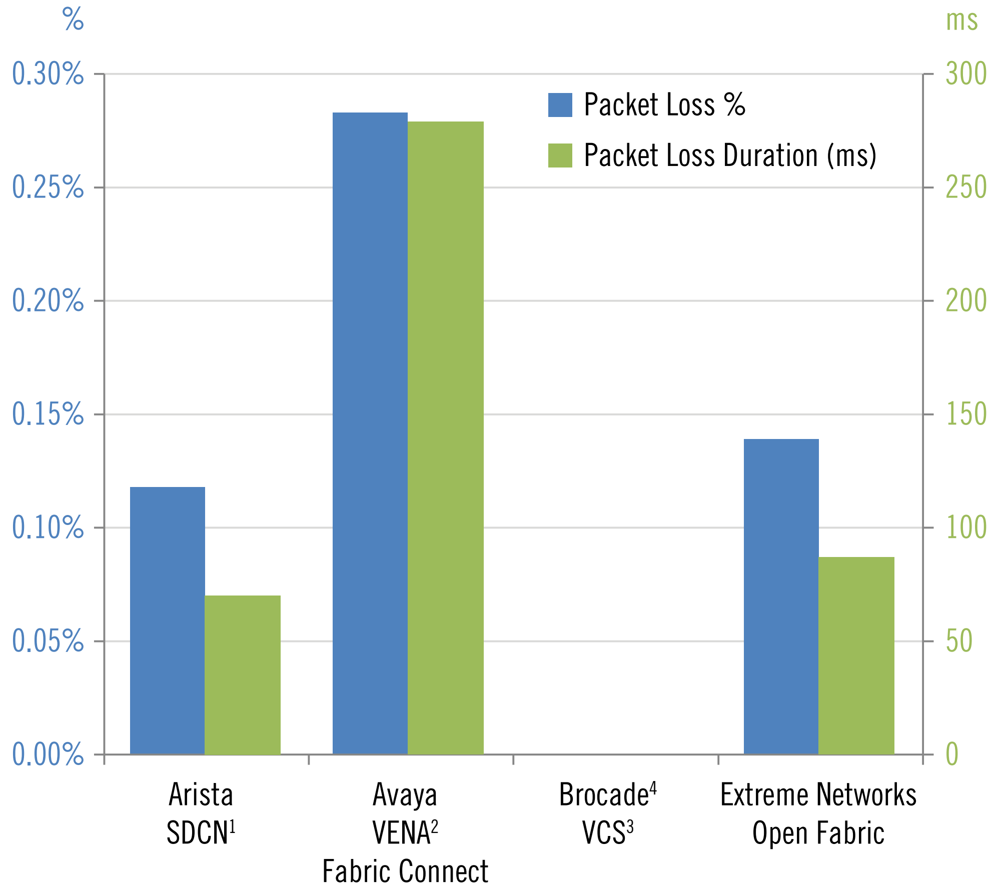
Jumbo frame 9216 size packet size traffic requires significantly more time to pass through the fabric, thanks to serialization, therefore, its plotted on a separate graphic so that smaller size packet traffic can be more easily viewed.

Extreme Networks Open Fabric delivered the lowest latency for singled homed fully meshed unicast traffic of packet sizes 128-1522 followed by Avaya Fabric Connect and Arista's SDCN. Avaya's Fabric Connect delivered the lowest latency for dual homed fully meshed unicast traffic of packet sizes 128-1522. The Extreme Networks Open Fabric and Avaya's Fabric Connect Dual homed results for fully meshed unicast traffic of packet sizes 128-1522 were nearly identical. Note that Avaya's Fabric Connect was configured with ToR switches while Arista and Extreme provided ToR and Core Switches, therefore, there is significantly more network capacity with these configurations.

Both Arista and Extreme dual homed results are the same as single homed, as expected. Avaya's dual homed result is lower than its single homed and is due to increased bandwidth between the ToRs. Further, Core switches do take longer to process packets, thanks to their higher port densities and intermodule system fabrics. Avaya's lack of a Core switch provided it an advantage in this test.

We don't believe that IxNetwork latency measurement of a fabric in Cut-through (CT) or Store and Forward (SF) is material. As the SF RFC 1242 latency measurement method is the time interval starting when the last bit of the input frame reaches the input port and ending when the first bit of the output frame is seen on the output port (LIFO) while CT RFC 1242 latency measurement method is the time interval starting when the end of the first bit of the input frame reaches the input port and ending when the start of the first bit of the output frame is see on the output port (FIFO). The measurement difference between CT vs. SF in a fabric under test is the size of one packet from the starting point; in essence it's the serialization delay of one packet size. CT vs. SF on device latency measurement are material. Given the above and the fact that Avaya's VSP 7000s were configured in SF while all other switches were configured for CT, we cannot rule out an anomaly that may impact Avaya's fabric latency measurement during this industry test.

Server-ToR Reliability Test
128 Byte Size Packet In Many-to-Many Full Mesh Flow Through Fabric
One 10GbE Cable in LAGs 3, 4, 7, and 8 Between Ixia-to-ToR is Pulled



[1] Software-Defined Cloud Network
[2] Virtual Enterprise Network Architecture
[3] Virtual Cluster Switching
[4] Brocade did not test for reliability

Server-ToR Reliability Test
128 Byte Size Packet In Many-to-Many Full Mesh Flow Through Fabric
One 10GbE Cable in LAGs 3, 4, 7, and 8 Between Ixia-to-ToR is Pulled

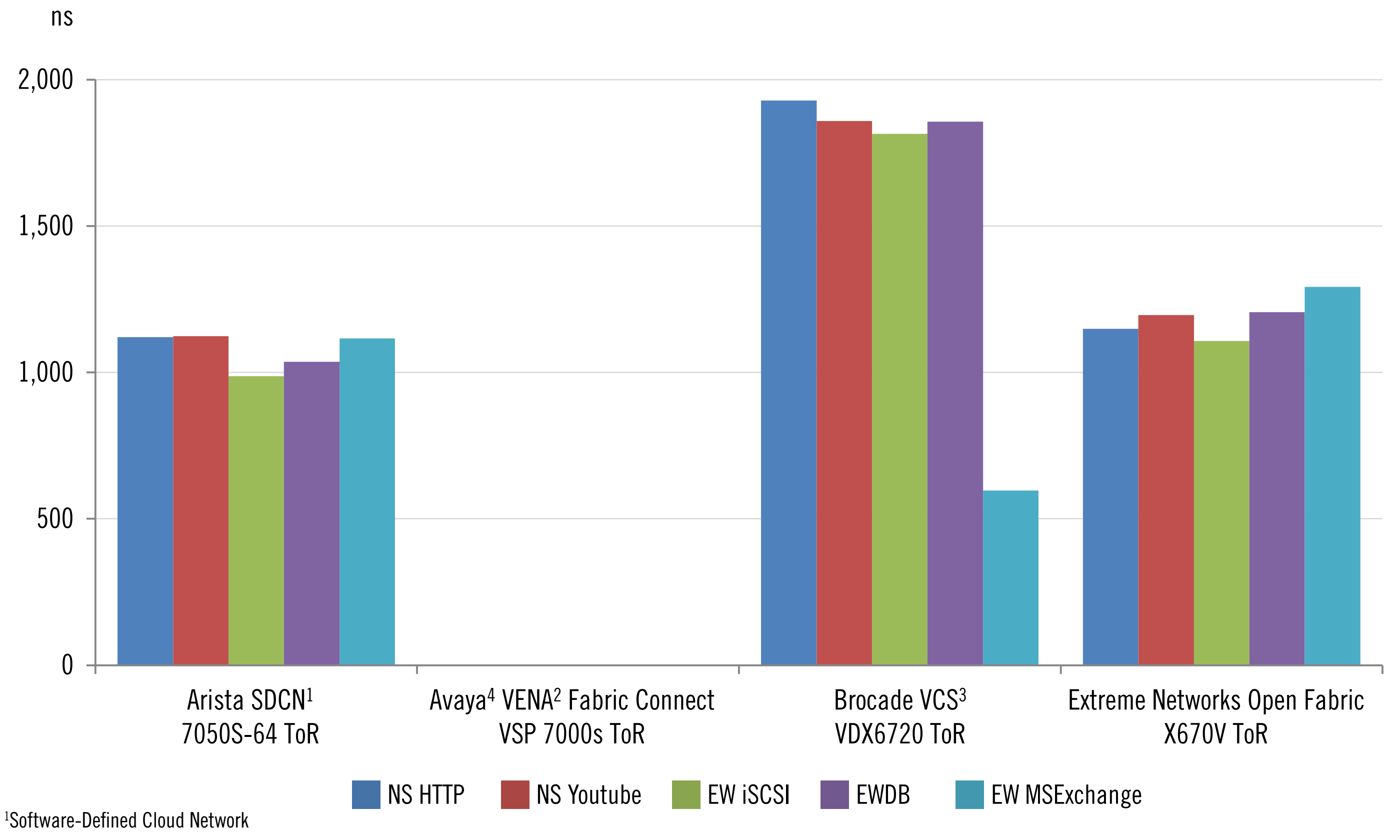
Company	Fabric Name	Fabric Products	Packet Loss %	Packet Loss Duration (ms)
Arista	SDCN[1]	2-7050S-64 ToR & 2-7508 Core	0.118%	70.179
Avaya	VENA[2] Fabric Connect	4-VSP 7000s ToR	0.283%	278.827
Brocade[4]	VCS[3]	2-VDX6720 ToR & 2-VDX8770 Core		
Extreme Networks	Open Fabric	2-X670V ToR & 2-BD X8 Core	0.139%	87.018

[1] Software-Defined Cloud Network
[2] Virtual Enterprise Network Architecture
[3] Virtual Cluster Switching
[4] Brocade did not test for reliability

Arista's SDCN delivered the lowest packet loss and shortest packet loss duration in the server-ToR reliability test followed by Extreme Network's Open Fabric, then followed by Avaya's Fabric Connect. The difference between Arista and Extreme's results for this reliability test is 17 milliseconds of packet loss duration and .024% packet loss; a narrow difference. Avaya's Server-ToR packet loss duration is approximately four times that of Arista.

Cloud Simulation ToR Switches At 50% Aggregate Traffic Load
Zero Packet Loss: Latency Measured in ns

28-10GbE Configuration Between Ixia-ToR Switch
Tested While In Single Homed Configuration



[1] Software-Defined Cloud Network
[2] Virtual Enterprise Network Architecture
[3] Virtual Cluster Switching
[4] Avaya did not test for Cloud Simulation

Cloud Simulation ToR Switches At 50% Aggregate Traffic Load
Zero Packet Loss: Latency Measured in ns

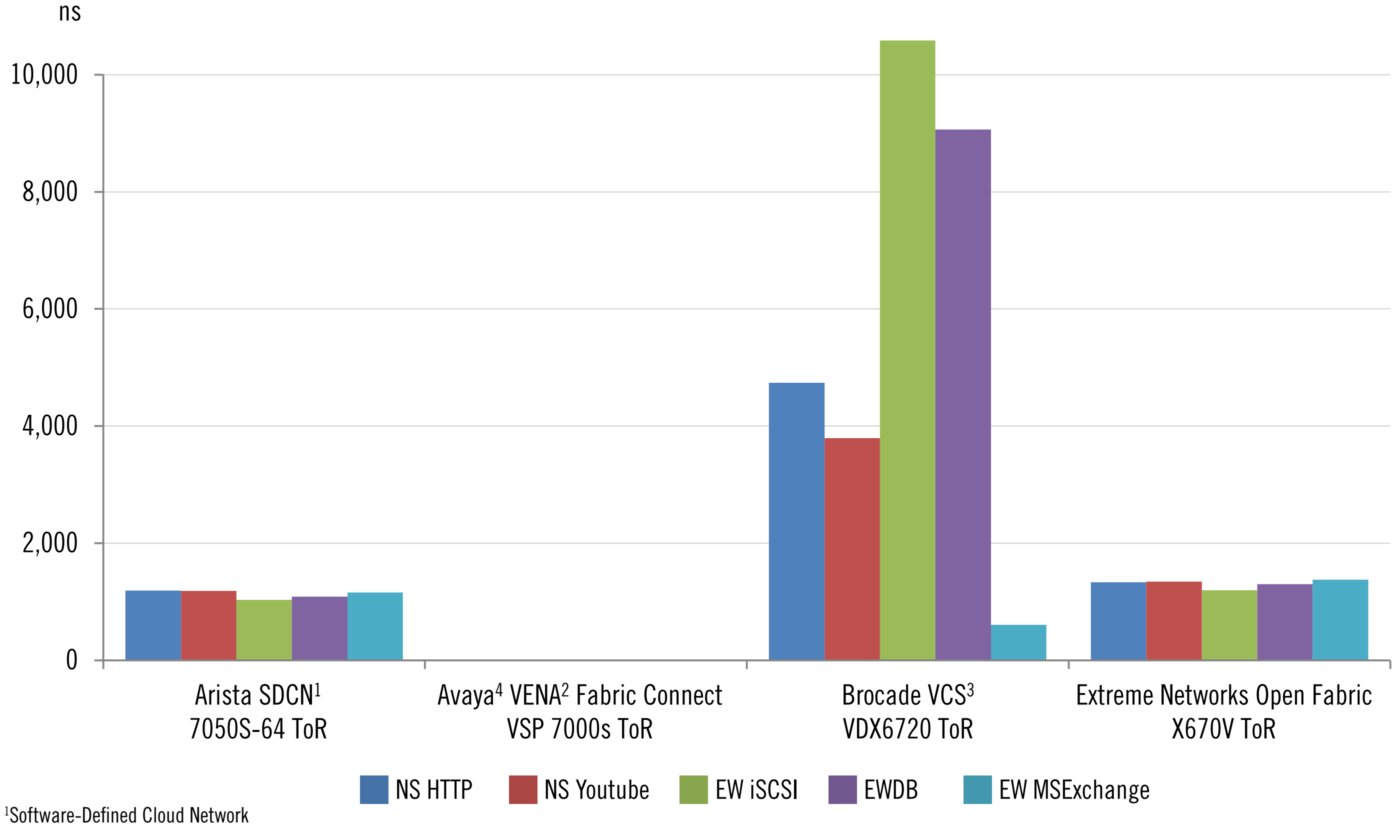
28-10GbE Configuration Between Ixia-ToR Switch
Tested While In Single Homed Configuration

Company	Fabric Name	Fabric Products	NS HTTP	NS Youtube	EW iSCSI	EW DB	EW MSExchange
Arista	SDCN[1]	7050S-64 ToR	1121	1124	987	1036	1116
Avaya[4]	VENA[2] Fabric Connect	VSP 7000s ToR					
Brocade	VCS[3]	VDX6720 ToR	1929	1859	1815	1857	597
Extreme Networks	Open Fabric	X670V ToR	1149	1196	1108	1206	1292

[1] Software-Defined Cloud Network
[2] Virtual Enterprise Network Architecture
[3] Virtual Cluster Switching
[4] Avaya did not test for Cloud Simulation

Cloud Simulation ToR Switches At 100% Aggregate Traffic Load Zero Packet Loss: Latency Measured in ns

28-10GbE Configuration Between Ixia-ToR Switch
Tested While In Single Homed Configuration



[1] Software-Defined Cloud Network
[2] Virtual Enterprise Network Architecture
[3] Virtual Cluster Switching
[4] Avaya did not test for Cloud Simulation

Cloud Simulation ToR Switches At 100% Aggregate Traffic Load Zero Packet Loss: Latency Measured in ns

28-10GbE Configuration Between Ixia-ToR Switch
Tested While In Single Homed Configuration

Company	Fabric Name	Fabric Products	NS HTTP	NS Youtube	EW iSCSI	EW DB	EW MSExchange
Arista	SDCN[1]	7050S-64 ToR	1187	1184	1033	1083	1156
Avaya[4]	VENA[2] Fabric Connect	VSP 7000s ToR					
Brocade	VCS[3]	VDX6720 ToR	4740	3793	10590	9065	602
Extreme Networks	Open Fabric	X670V ToR	1330	1342	1195	1300	1376

[1] Software-Defined Cloud Network
[2] Virtual Enterprise Network Architecture
[3] Virtual Cluster Switching
[4] Avaya did not test for Cloud Simulation

Arista Network's 7050S-64 delivered the lowest latency measurement for the Lippis Cloud simulation test at 50% and 100% load, followed by Extreme Networks' X670V and Brocade's VCS VDX 6720. Both the Arista Networks' 7050S-64 and Extreme Networks' X670V delivered nearly consistent performance under 50% and 100% load with variation of a few hundred nanoseconds per protocol, meaning that there is plenty of internal processing and bandwidth capacity to support this traffic load. The Brocade's VCS VDX 6720 was slightly more variable. All products delivered 100% throughput, meaning that not a single packet was dropped as load varied.

Ethernet Fabric Industry Recommendations

The following provides a set of recommendations to IT business leaders and network architects for their consideration as they seek to design and build their private/public data center cloud network fabric. Most of the recommendations are based upon our observations and analysis of the test data. For a few recommendations, we extrapolate from this baseline of test data to incorporate key trends and how these Ethernet Fabrics may be used in their support for corporate advantage.

Consider Full Mesh Non Blocking: Most of the fabric configurations were fully meshed and non-blocking which provided a highly reliable and stable infrastructure. This architecture scales, thanks to active-active protocols, and enables two-tier design, which lowers equipment cost plus latency. In addition to being highly reliable, it also enables dual-homed server support at no performance cost.

Consider Two-Tier Network Fabric: To reduce equipment cost, support a smaller number of network devices and increase application performance, it's recommended to implement a two-tier leaf-spine Ethernet Fabric. This Lippis/Ixia Active-Active Ethernet Fabric Test demonstrated that two-tier networks are not only ready for prime-time deployment, they are the preferred architecture.

MLAG and ECMP Proven and Scales But: It was proven that a two-tier network can scale, thanks to ECMP up to 32-way links. In addition, ECMP offers multipathing, too, at scale. What's missing from ECMP is auto provisioning of links between switches. In short, ECMP requires manual configuration.

Consider Utilizing TRILL and/or SPB: Over time, most vendors will support TRILL and SPB in addition to MLAG and ECMP. Both TRILL and SPB offer unique auto-provisioning features that simplify network design. It's recommended that network architects experiment with both active-active protocols to best understand its utility within your data center network environment.

Strong Underlay for a Dynamic Overlay: The combination of fully meshed, non-blocking two-tier network build with standard active-active protocols constructs a strong underlay to support a highly dynamic overlay. With the velocity of change in highly virtualized data centers ushering in virtualized networks or overlays, a stable and scalable underlay is the best solution to support the rapid build-up of tunneled traffic running through Ethernet Fabrics. This huge demand in overlay traffic is yet another good reason to consider a two-tier active-active Ethernet Fabric for data center and cloud networking.

Be Open to Different Fabric Architectures: Not all data centers support 10,000 or 100,000 servers and require enormous scale. There are different approaches to building Ethernet Fabrics that are focused on converged I/O or simplicity of deployment, auto provisioning, keeping east-west traffic at the ToR tier, etc. Many vendors offering Ethernet Fabrics offer product strategies to scale up as requirements demand.

Get Ready for Open Networking: In this Lippis/Ixia Test, we focused on the active-active protocols for all the reasons previously mentioned. When considering an Ethernet Fabric, it's important to focus on open networking, such as the integration of the network operating system with OpenStack, or how ToR and Cores support various SDN controllers, do the switches support OpenFlow or have a road map for its support. There are three types of networks in data centers today, L2/L3, Network Virtualization overlays that tunnel traffic through L2/L3 and soon OpenFlow flows. Consider those vendors that support all types of networking as this is a fast-moving target. Auto provisioning of networking with compute and storage is increasingly important; therefore, look for networking vendors that support network configuration via SDN controllers plus virtualization and cloud orchestration systems.

Terms of Use

This document is provided 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 vendors 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, avail¬able in equivalent or better form to commercial customers. Accordingly, this document is provided "as is," and Lippis Enterprises, Inc. (Lippis), 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. Lippis is not responsible for, and you agree to hold Lippis 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.

Lippis makes no claim as to whether any product or company described herein is suitable for in¬vestment. 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 www.lippisreport.com .

No part of any document may be reproduced, in whole or in part, without the specific written permission of Lippis. 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.

About Nick Lippis

Nicholas J. Lippis III is a world-renowned authority on advanced IP networks, communications and their benefits to business objectives. He is the publisher of the Lippis Report, a resource for network and IT business decision makers to which over 35,000 executive IT business leaders subscribe. Its Lippis Report podcasts have been downloaded over 200,000 times; ITunes reports that listeners also download the *Wall Street Journal's* Money Matters, *Business Week's* Climbing the Ladder, *The Economist* and The *Harvard Business Review's* IdeaCast. He is also the co-founder and conference chair of the Open Networking User Group, which sponsors a bi-annual meeting of over 200 IT business leaders of large enterprises. Mr. Lippis is currently working with clients to design their private and public virtualized data center cloud computing network architectures with open networking technologies to reap maximum business value and outcome.

He has advised numerous Global 2000 firms on network architecture, design, implementation, vendor selection and budgeting, with clients including Barclays Bank, Eastman Kodak Company, Federal Deposit Insurance Corporation (FDIC), Hughes Aerospace, Liberty Mutual, Schering-Plough, Camp Dresser McKee, the state of Alaska, Microsoft, Kaiser Permanente, Sprint, Worldcom, Cisco Systems, Hewlett Packet, IBM, Avaya and many others. He works exclusively with CIOs and their direct reports. Mr. Lippis possesses a unique perspective of market forces and trends occurring within the computer networking industry derived from his experience with both supply- and demand-side clients.

Mr. Lippis received the prestigious Boston University College of Engineering Alumni award for advancing the profession. He has been named one of the top 40 most powerful and influential people in the networking industry by *Network World*. *TechTarget*, an industry on-line publication, has named him a network design guru while *Network Computing Magazine* has called him a star IT guru.

Mr. Lippis founded Strategic Networks Consulting, Inc., a well-respected and influential computer networking industry-consulting concern, which was purchased by Softbank/Ziff-Davis in 1996. He is a frequent keynote speaker at industry events and is widely quoted in the business and industry press. He serves on the Dean of Boston University's College of Engineering Board of Advisors as well as many start-up venture firms' advisory boards. He delivered the commencement speech to Boston University College of Engineering graduates in 2007. Mr. Lippis received his Bachelor of Science in Electrical Engineering and his Master of Science in Systems Engineering from Boston University. His Masters' thesis work included selected technical courses and advisors from Massachusetts Institute of Technology on optical communications and computing.