Open Industry Network Performance & Power Test *for* Cloud Networks

Evaluating 10/40 GbE Switches

Fall 2011 Edition

A Report on the
Extreme Networks
BlackDiamond® X8
Data Center Switch

December, 2011

Forward by Scott Bradner

It is now almost twenty years since the Internet Engineering Task Force (IETF) initially chartered the Benchmarking Methodology Working Group (bmwg) with me as chair. The aim of the working group was to develop standardized terminology and testing methodology for various performance tests on network devices such as routers and switches.

At the time that the bmwg was formed, it was almost impossible to compare products from different vendors without doing testing yourself because each vendor did its own testing and, too often, designed the tests to paint their products in the best light. The RFCs produced by the bmwg provided a set of standards that network equipment vendors and testing equipment vendors could use so that tests by different vendors or different test labs could be compared.

Since its creation, the bmwg has produced 23 IETF RFCs that define performance testing terminology or methodology for specific protocols or situations and are currently working on a half dozen more. The bmwg has also had three different chairs since I resigned in 1993 to join the IETF's steering group.

The performance tests in this report are the latest in a long series of similar tests produced by a number of testing labs. The testing methodology follows the same standards as I was using in my own test lab at Harvard in the early 1990s thus the results are comparable. The comparisons would not be all that useful since I was dealing with far slower speed networks, but a latency measurement I did in 1993 used the same standard methodology as do the latency tests in this report.

Considering the limits on what I was able to test way back then, the Ixia test setup used in these tests is very impressive indeed. It almost makes me want to get into the testing business again.

Scott is the University Technology Security Officer at Harvard University. He writes a weekly column for Network World, and serves as the Secretary to the Board of Trustees of the Internet Society (ISOC). In addition, he is a trustee of the American Registry of Internet Numbers (ARIN) and author of many RFC network performance standards used in this industry evaluation report.

Acknowledgements

We thank the following people for their help and support in making possible this first industry evaluation of 10GbE private/public data center cloud Ethernet switch fabrics.

Atul Bhatnagar, CEO of Ixia, for that first discussion of this project back at Interop in May 2009 and his continuing support throughout its execution.

Scott O. Bradner of Harvard University for his work at the IETF on network performance benchmarks, used in this project, and the Forward of this report.

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 cooper 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, Kelly Maloit, Director of Public Relations, Tara Van Unen, Director of Market Development, Jim Smith, VP of Marketing, all from Ixia, for their support and contributions to creating a successful industry event.

Bob Anderson, 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 event but to each other, as many provided helping hands on many levels to competitors.

Tiffany O'Loughlin for her amazing eye in capturing test week with still and video photography.

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 5

Market Background 7

Participating Supplier

Extreme Networks BlackDiamond® X8 12

Extreme Networks BlackDiamond® X8 192-40GbE Test Results 20

Cross Vendor Analysis

Core Switches 27

Latency 28

Throughput 30

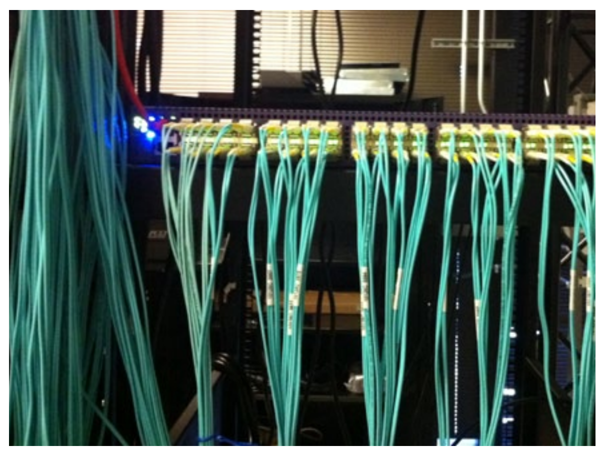
Power 36

Core Switches Industry Recommendations 38

Test Methodology 40

Terms of Use 45

About Nick Lippis 46



Executive Summary

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

The resources available for this test at Ixia's iSimCity are out of reach for nearly all corporate IT departments with test equipment on the order of $9.5M, devices under test on the order of $2M, plus costs associated with housing, power and cooling the lab, and lastly nearly 22 engineers from around the industry working to deliver test results. It's our hope that this report will remove performance, power consumption, virtualualization scale and latency concern from the purchase decision, allowing IT architects and IT business leaders to focus on other vendor selection criteria, such as post sales support, platform investment, vision, company financials, etc.

The Lippis test reports based on independent validation at Ixia's iSim City, communicates credibility, competence, openness and trust to potential buyers of 10GbE and 40GbE data center switching equipment as the tests are open to all suppliers and are fair, thanks to RFC and custom-based tests that are repeatable. The private/public data center cloud 10GbE and 40GbE fabric test was free for vendors to participate and open to all industry suppliers of 10GbE and 40GbE switching equipment, both modular and fixed configurations.

This report communicates the results of three set of test which took place during the week of December 6 to 10, 2010, April 11 to 15, 2011 and October 3 to 14, 2011 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 equippped with QSFP+ connectors for 40GbE connections. Each 10GbE and 40GbE supplier was allocated lab time to run the test with the assistance of an Ixia engineer. Each switch vendor configured their equipment while Ixia engineers ran the test and logged the resulting data.

The tests conducted were IETF RFC-based performance and latency test, power consumption measurements and a custom cloud-computing simulation of large north-south plus east-west traffic flows. Virtualization scale or how was calcluated and reported. Core switches were evaluated.

Core switches evaluated were:

Alcatel-Lucent OmniSwitch 10K

Arista 7504 Series Data Center Switch

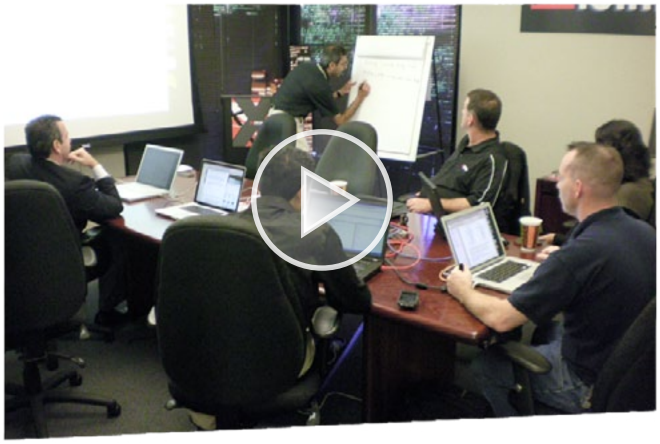
Extreme Networks BlackDiamond® X8 Data Center Switch

Juniper Network EX Series EX8200 Ethernet Switch

From the above list the following products were tested in October, 2011 and added to the products evaluated in December 2010 and April 2011.

Extreme Networks BlackDiamond® X8

What We Have Learned from 18 Months of Testing:



Video feature: Click to view "What We Have Learned" video podcast

The following list our top ten findings:

1. 10GbE ToR and Core switches are ready for mass deployment in private/public data center cloud computing facilities, as the technology is mature, and products are both stable and deliver stated goals of high performance, low latency and low power consumption.

2. There are differences between suppliers, and it's recommended to review each supplier's results along with other important information in making purchase decisions.

3. The Lippis/Ixia industry evaluations are the "litmus test" for products to enter the market, as only those firms who are confident in their products and ready to go to market participate. For each of the three industry test, all ToR and Core switches evaluated in the Lippis/Ixia test were recently introduced to market where the Lippis/Ixia test was their first open public evaluation.

4. Most ToR switches are based upon a new generation of merchant silicon that provides a single chip-forwarding engine of n-10GbE and n-40GbE ports that delivers consistent performance and low power consumption.

5. All ToR and Core switches offer low power consumption with energy cost over a three-year period estimated between 1.3% and 4% of acquisition cost. We measure WATTS per 10GbE port via ATIS methods and TEER values for all switches evaluated. It's clear that these products represent the state-of-the-art in terms of energy conservation. ToR switches consume between 1.5 and 5.5 WATTS per 10GbE port while Core switches consume between 5.62 and 21.68 WATTS per 10GbE port.

6. Most switches design airflow from front-to-back or back-to-front rather than side-to-side to align with data center hot/cold aisle cooling design.

7. There are differences between suppliers in terms of the network services they offer, such as virtualization support, quality of service, etc., as well as how their Core/Spine switches connect to ToR/Leaf switches to create a data center fabric. It's recommended that IT business leaders evaluate Core switches with ToR switches and vice versa to assure that the network services and fabric attributes sought after are realized throughout the data center.

8. Most Core and ToR switches demonstrated the performance and latency required to support storage enablement or converged I/O. In fact, all suppliers have invested in storage enablement, such as support for Converged Enhanced Ethernet (CEE), Fibre Channel over Ethernet (FCoE), Internet Small Computer System Interface (iSCSI), Network-attached Storage (NAS), etc., and while these features were not tested in the Lippis/Ixia evaluation, most of these switches demonstrated that the raw capacity is built into the switches for its support.

9. Three ToR switches support 40GbE while all Core switches possess the backplane speed capacity for some combination of 40 and 100GbE. Core switches tested here will easily scale to support 40GbE and 100GbE while ToR suppliers offer uplinks at these speeds. Of special note is the Extreme Networks BlackDiamond® X8 which was the first core switch tested with 24-40GbE ports, marking a new generation of data center core products entering the market. The Extreme Networks BlackDiamond® X8 switch is capable of up to 192 40GbE ports.

10. The Lippis/Ixia test demonstrated the performance and power consumption advantages of 10GbE networking, which can be put to work and exploited for corporate advantage. For new server deployments in private/public data center cloud networks, 10GbE is recommended as the primary network connectivity service as a network fabric exists to take full advantage of server I/O at 10GbE bandwidth and latency levels. 40GbE connections are expected to ramp up during 2012 for ToR to Core connections with server connections in select high performance environments.

Market Background

The following market background section provides perspective and context to the major changes occurring in the IT industry, and its fundamental importance to the network switches that were evaluated in the Lippis/Ixia test.

The IT industry is in the midst of a fundamental change toward centralization of application delivery via concentrated and dense private and public data center cloud computing sites. Corporations are concentrating IT spending in data centers and scaling them up via server virtualization to reduce complexity and associated operational cost while enjoying capital cost savings advantage. In fact, there is a new cloud-o-nomics that is driven by the cost advantage of multi-core computing coupled with virtualization that has enabled a scale of computing density not seen before. Couple this with a new tier of computing, Apple's iPad, Android tablets plus smartphones that rely upon applications served from private and public cloud computing facilities, and you have the making of a systemic and fundamental change in data center design plus IT delivery.

The world has witnessed the power of this technology and how it has changed the human condition and literally the world. On May 6, 2010, the U.S. stock market experienced a flash crash that in 13 short seconds, 27,000 contracts traded consisting of 49% of trading volume. Further, in 15 minutes $1 trillion of market value disappeared. Clusters of computers connected at 10 and 40GbE programmed to perform high frequency trades based upon proprietary algorithms did all this. The making of the movie *Avatar* was produced at WETA Digital with 90% of the movie developed via giant computer clusters connected at 10GbE to process and render animation. As big science cost has soared, scientists have moved to high performance computing clusters connected at 10 and 40 GbE to simulate rather than build scientific experiments. Biotech engineers are analyzing protein-folding simulations while military scientists simulate nuclear explosions rather than detonate these massive bombs. Netflix has seen its stock price increase and then crash as it distributes movies and TV programs over the Internet and now threatens Comcast and others on-demand business thanks to its clusters of computers connected at 10GbE. At the same time, Blockbuster Video filed for bankruptcy.

A shift in IT purchasing is occurring too. During the market crash of 2008, demand for information increased exponentially while IT business decision makers were forced to scale back budgets. These diverging trends of information demand and IT budget contraction have created a data center execution gap where IT leaders are now forced to upgrade their infrastructure to meet information and application demand. To close the gap, IT business decision makers are investing in 10 and 40GbE switching. Over the past several quarters, sales of 10GbE fixed and modular switches have grown in the 60% to 70% range and now represent 25% of all ethernet shipments. Note that data center etherent ports represent just 10% of the market, but nearly 40% of revenue. 40GbE shippments are expected to ramp during 2012.

At the center of this new age in private and public data center cloud computing is networking, in particular Ethernet networking that links servers to storage and the internet providing high-speed transport for application traffic flow. As the data center network connects all devices, it is the single most important IT asset to assure end users receive an excellent experience and service is continually available. But in the virtualization/cloud era of computing, networking too is fundamentally changing as demand for low latency, high performance and lower power consumption switches increase to address radically different traffic patterns. In addition, added network services, such as storage enablement, are forcing Ethernet to become not just a connectivity service in the data center but a fabric that connects computing and storage; transporting data packets plus blocks of storage traffic. In short, IT leaders are looking for Ethernet to provide a single data center fabric that connects all IT assets with the promise of less equipment to install, simpler management and greater speed to IT service delivery.

The following characterize today's modern private and public data center cloud network fabric.

Virtualization: The problems with networking in virtualized infrastructure are well documented. Move a virtual machine (VM) from one physical server to another and network port profile, Virtual Local Area Networking (VLAN), security settings, etc., have to be reconfigured, adding cost, delay and rigidity to what should be an adaptive infrastructure. Many networking companies are addressing this issue by making their switches virtualization aware, auto reconfiguring and/or recommending a large flat layer 2 domain. In addition, networking companies are offering network management views that cross administrative domains of networking and virtualization in an effort to provide increased visibility and management of physical and virtual infrastructure. The fabric also needs to eliminate the boundary between physical and virtual infrastructure, both in terms of management and visibility plus automated network changes as VMs move.

Traffic Patterns Shift with Exponential Volume Increase: Data centers are experiencing a new type of traffic pattern that is fundamentally changing network design plus latency and performance attributes. Not only are traditional north-south or client-server flows growing but east-west or server-server and storage-server flows now dominate most private and public data center cloud facilities. A dominate driver of east-west flows is the fact that the days of static web sites are over while dynamic sites take over where one page request can spawn 50 to 100 connections across multiple servers.

Large contributors to east-west flows are mobile devices. Mobile application use is expanding exponentially, thanks to the popularity of the iPhone, iPad and increasingly Android smartphones plus tablets. As of this writing, there are some 700,000 plus smartphone applications. The traffic load these applications are placing on data center Ethernet fabrics is paradoxically immense. The vast majority of mobile applications are hosted in data centers and/or public cloud facilities. The application model of mobile devices is not to load them with thick applications like Microsoft Word, PowerPoint, Excel, etc, but to load them with thin clients that access applications and data in private and/or public data centers, cloud facilities.

A relatively small query to a Facebook page, *New York Times* article, Amazon purchase, etc., sends a flurry of east-west traffic to respond to the query. Today's dynamic web spreads a web page over multiple servers and when called upon, drives a tsunami of traffic, which must be delivered well before the person requesting the data via mobile device or desktop loses interest and terminates the request.

East-west traffic will only increase as this new tier of tablet computing is expected to ship 171 million units in 2014 up from 57 million this year, according toIsupplie. Add tens of millions of smartphones to the mix, all relying on private/public data center cloud facilities for their applications, and you have the ingredients for a sustained trend placing increased pressure for east-west traffic flows upon Ethernet data center fabrics.

Low Latency: Data experiences approximately 10 ns of latency as it traverses across a computer bus passing between memory and CPU. Could networking become so fast that it will emulate a computer bus so that a data center operates like one giant computer? The answer is no, but the industry is getting closer. Today's 10GbE switches produce 400 to 700 ns of latency. By 2014, it's anticipated that 100GbE switching will reduce latency to nearly 100 ns.

Special applications, especially in high frequency trading and other financial markets, measure latency in terms of millions of dollars in revenue lost or gained, placing enormous pressure on networking gear to be as fast as engineers can design. As such, low latency is an extremely important consideration in the financial services industry. For example, aggregating all 300 plus stock market feeds would generate approximately 3.6Gbs of traffic, but IT architects in this market choose 10GbE server connections versus 1GbE for the sole purpose of lower latency even though the bandwidth is nearly three times greater than what theoretically would be required. They will move to 40GbE too as 40GbE is priced at three to four times that of 10GbE for four times the bandwidth.

More generally, latency is fundamental to assuring a user's excellent experience. With east-west traffic flows making up as much as 80% of data center traffic, low latency requirements are paramount as every trip a packet makes between servers during the processes of responding to an end user's query adds delay to the response and reduces the user experience or revenue potential.

Storage Traffic over Ethernet Fabric: Converged I/O or unified networking where storage and network traffic flow over a single Ethernet network will increasingly be adopted in 2012. Multiple storage enablement options exist, such as iSCSI over Ethernet, ATA over Ethernet (AoE), FCoE, etc. For IP-based storage approaches, such as iSCSI over Ethernet and AoE, all that is needed is a 10GbE network interface card (NIC) in a server to support both storage and data traffic, but the Ethernet fabric requires lossless Ethernet to assure integrity of storage transport over the fabric as blocks of data transmit between server and storage farms.

For FCoE, a single converged network adaptor or CNA plugged into a server provides the conduit for storage and application traffic flows to traverse over an Ethernet fabric. The number of suppliers offering CNAs has grown significantly, including Intel, HP, Emulex, IBM, ServerEngines, QLogic, Cisco, Brocade, etc. In addition, the IEEE opened up the door for mass deployment as it has ratified the key Ethernet standards for lossless Ethernet. What will drive converged I/O is the reduced cost of cabling, NIC and switching hardware.

Low Power Consumption: With energy costs increasing plus green corporate initiatives as well as government mandates pervasive, all in an effort to reduce carbon emissions, IT leaders have been demanding lower energy consumption of all IT devices, including data center network equipment. The industry has responded with significant improvements in energy efficiency and overall lower wattage consumption per port while 10 and 40 GbE switches are operating at full line rate.

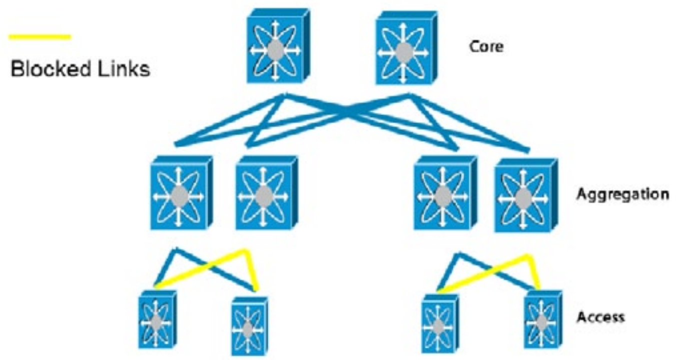
Virtualized Desktops: 2012 promises to be a year of increased virtualized desktop adoption. Frustrated with enterprise desktop application licensing, plus desktop support model, IT business leaders will turn toward virtualizing desktops at increasing numbers. The application model of virtualized desktops is to deliver a wide range of corporate applications hosted in private/public data center clouds over the enterprise network. While there are no estimates to the traffic load this will place on campus and data center Ethernet networking, one can only assume it will add north-south traffic volume to an already increasing and dominating east-west flows.

Fewer Network Tiers: To deliver low latency and high throughput performance to support increasing and changing traffic profiles mentioned above, plus ease VM moves without network configuration changes, a new two-tier private/public data center cloud network design has emerged. A three-tier network architecture is the dominant structure in data centers today and will likely continue as the optimal design for many smaller data center networks. For most network architects and administrators, this type of design provides the best balance of asset utilization, layer 3 routing for segmentation, scaling and services plus efficient physical design for cabling and fiber runs.

By three tiers, we mean access switches/ToR switches, or modular/End-of-Row switches that connect to servers and IP-based storage. These access switches are connected via Ethernet to aggregation switches. The aggregation switches are connected into a set of Core switches or routers that forward traffic flows from servers to an intranet and internet, and between the aggregation switches. It is common in this structure to oversubscribe bandwidth in the access tier, and to a lesser degree, in the aggregation tier, which can increase latency and reduce performance. Spanning Tree Protocol or STP between access and aggregation plus aggregation and core further drive oversubscription. Inherent in this structure is the placement of layer 2 versus layer 3 forwarding that is VLANs and IP routing. It is common that VLANs are constructed within access and aggregation switches, while layer 3 capabilities in the aggregation or Core switches route between them.

Within the private and public data center cloud network market, where the number of servers are in the thousands to tens of thousands plus, east-west bandwidth is significant and applications require a single layer 2 domain, the existing Ethernet or layer 2 capabilities within this tiered architecture do not meet emerging demands.

Traditional three-tier Network Fabric Architecture



An increasingly common design to scale a data center fabric is often called a "fat-tree" or "leaf-spine" and consists of two kinds of switches; one that connects servers and the second that connects switches creating a non-blocking, low-latency fabric. We use the terms "leaf" or "Top of Rack (ToR)" switch to denote server connecting switches and "spine" or "core" to denote switches that connect leaf/ToR switches. Together, a leaf/ToR and spine/core architecture create a scalable data center fabric. Key to this design is the elimination of STP with some number of multi-links between leaf and spine that eliminate oversubscription and enable a non-blocking fabric, assuming the switches are designed with enough backplane capacity to support packet forwarding equal to the sum of leaf ingress bandwidth. There are multiple approaches for connecting leaf and spine switches at high bandwidth, which fall under the category of Multi-Chassis Link Aggregation Group or MC-LAG covered in project IEEE 802.3ad.

Paramount in the two-tier leaf-spine architecture is high spine switch performance, which collapses the aggregation layer in the traditional three-tier network. Another design is to connect every switch together in a full mesh, via MC-LAG connections with every server being one hop away from each other. In addition, protocols such as TRILL or Transparent Interconnection of Lots of Links, SPB or Shortest Path Bridging, et al have emerged to connect core switches together at high speed in an active-active mode.

Two-Tier Leaf-Spine Network Fabric Architecture

The above captures the major trends and demands IT business leaders are requiring from the networking industry. The underpinnings of private and public data center cloud network fabric are 10GbE switching with 40GbE and 100GbE ports/modules. 40GbE and 100GbE are in limited availability now but will be increasingly offered and adopted during 2012. Network performance including throughput performance and latency are fundamental switch attributes to understand and review across suppliers. Because if the 10/40GbE switches an IT leader selects cannot scale performance to support increasing traffic volume plus shifts in traffic profile, not only will the network fail to be a fabric unable to support converge storage traffic, but business processes, application performance and user experience will suffer too.

During 2012, an increasing number of servers will be equipped with 10GbE LAN on Motherboard (LOM) driving 10GbE network requirements. In addition, with nearly 80% of IT spend being consumed in data center infrastructure with all IT assets eventually running over 10/40GbE switching, the stakes could not be higher to select the right product upon which to build this fundamental corporate asset. Further, data center network equipment has the longest life span of all IT equipment; therefore, networking is a long-term investment and vendor commitment.

Participating Suppliers

All Core switches including Juniper's EX8216, Arista Networks' 7504, and Alcatel-Lucent's OmniSwitch 10K were tested for the first time in an open and public forum too. These "spine" switches make up the core of cloud networking. A group of spine switches connected in partial or full mesh creates an Ethernet fabric connecting ToR or leaf switches and associated data plus storage traffic. These Core switches range in list price from $250K to $800K and price per 10GbE port of $1.2K to $6K depending upon port density and software license arrangement. The Core/Spine switches tested were:

Alcatel-Lucent OmniSwitch 10K

Arista 7504 Series Data Center Switch

Extreme Networks BlackDiamond® X8

Juniper Network EX Series EX8200 Ethernet Switch

We present the results of the Extreme Networks BlackDiamond® X8 configured with 256-10GbE and 24-40GbE ports. We then present test results for a special case of the Extreme Networks BlackDiamond® X8 configured with 192-40GbE port. A cross-vendor test result analysis follows.

Video feature: **Click to view Extreme Networks video podcast**

Extreme Networks BlackDiamond® X8 Core Switch

Extreme Networks launched its new entry into the enterprise data center and cloud computing markets with its new line of BlackDiamond® X series switches in May of 2011. This line of Top of Rack and Core switches pushes the envelope on performance, power consumption, density and data center network design.

Extreme Networks submitted their new BlackDiamond® X8 (BDX8) core switch into the Fall Lippis/Ixia industry test. The BDX8 is the highest density 10GbE and 40GbE core switch on the market. It's capable of supporting 768 10GbE and 192 40GbE ports in the smallest packaging footprint of one third of a rack or only 14.5RU. This high port density core switch in a small footprint offers new flexibility in building enterprise core networks and cloud data centers. For example, a single BDX8 could support 768 servers connected directly at 10GbE creating a non-blocking flat layer 2 network. But this is but one option in many designs afforded by the BDX8's attributes.

From an internal switching capacity point of view, the BDX8 offers 1.28 Tbps slot capacity and more than 20 Tbps switching fabric capacity; enough to support the 768 ports of 10 GbE or 192 ports of 40 GbE per chassis. In addition to raw port density and footprint, the BDX8 offers many software features. Some notable software features include automated configuration and provisioning, and XNV™(ExtremeXOS Network Virtualization), which provides automation of VM (Virtual Machine) tracking, reporting and migration of VPPs (Virtual Port Profile). These are but a few software features enabled through ExtremeXOS® network operating system. The Extreme Networks BlackDiamond® X8 is also key to Extreme Network's OpenFabric architecture that enables best-in-class choice of switching, compute, storage and virtualization solutions.

Extreme Networks BlackDiamond® X8 Core Switch Test Configuration*

	Hardware	Software Version	Port Density
Device under test	BSX8 http://www.extremenetworks.com	15.1	352
Test Equipment	Ixia XM12 High Performance Chassis	IxOS 6.10 EA IxNetwork 6.10 EA	
Ixia Line Cards	Xcellon Flex AP10G16S 16 port 10G module		
	Xcellon Flex Combo 10/40GE AP 16 port 10G / 4 port 40G		
	http://www.ixiacom.com/		
Cabling	10GbE Optical SFP+ connectors. Laser optimized duplex lc-lc 50 micron mm fiber, 850mm SPF+ transceivers		
	www.leviton.com		
	Siemon QSFP+ Passive Copper Cable 40 GbE 3 meter copper QSFP-FA-010		
	Siemon Moray Low Power Active Optical Cable Assemblies Single Mode QSFP+ 40GbE optical cable QSFP30-03 7 meters		
	www.siemon.com		

* Extreme BDX8 was configured in 256 10GbE ports and 24-40GbE ports or an equivalent 352-10GbE ports.

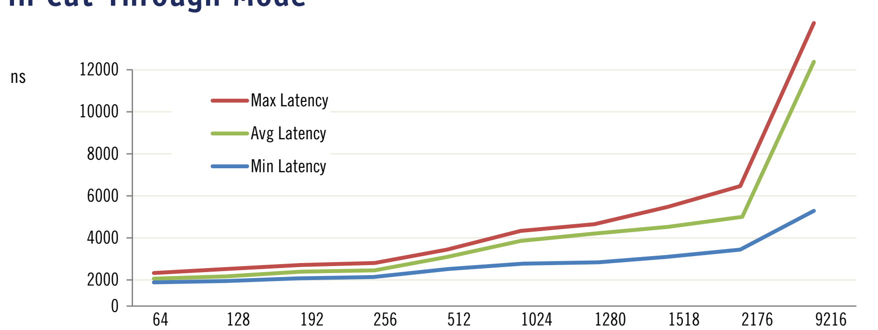
Extreme Networks BlackDiamond® X8 Core Switch

The BDX8 breaks all of our previous records in core switch testing from performance, latency, power consumption, port density and packaging design. In fact, the Extreme Networks BDX8 was between 2 and 9 times faster at forwarding packets over a wide range of frame sizes than previous core switch measurements conducted during Lippis/Ixia iSimCity test of core switches from Alcatel-Lucent, Arista and Juniper. In other words, the BDX8 is the fastest core switch we have tested with the lowest latency measurements. The BDX8 is based upon the latest Broadcom merchant silicon chip set.

For the fall Lippis/Ixia test we populated the Extreme Networks BlackDiamond® X8 with 256 10GbE ports and 24 40GbE ports, thirty three percent of its capacity. This was the highest capacity switch tested during the entire series of Lippis/Ixia cloud network test at iSimCity to date.

We tested and measured the BDX8 in both cut through and store and forward modes in an effort to understand the difference these latency measurements offer. Further, latest merchant silicon forwards packets in store and forward for smaller packets while larger packets are forwarded in cut-through making this new generation of switches hybrid cut-through/store and forward devices. The Extreme Networks BlackDiamond® X8 is unique in that its switching architecture forward packets in both cut-through and store-and-forward modes. As such, we measure the Extreme Networks BlackDiamond® X8 latency with both measurement methods. The purpose is not to compare store-and-forward vs cut through, but to demonstrate the low latency results obtained independent upon forwarding method. During store-and-forward testing, the latency of packet serialization delay (based on packet size) is removed from the reported latency number by test equipment. Therefore, to compare cutthrough to store-and-forward latency measurement, packet serialization delay needs to be added to store-and-forward latency number. For example, in a store-and-forward latency number of 800ns for a 1,518 byte size packet, the additional latency of 1240ns (serialization delay of a 1518 byte packet at 10Gbps) is required to be added to the store-and-forward measurement. This difference can be significant. Note that other potential device specific factors can impact latency too. This makes comparisons between two testing methodologies difficult.

Extreme Networks BDX8 RFC 2544 Layer 2 Latency Test in Cut-Through Mode



	64	128	192	256	512	1024	1280	1518	2176	9216
Max Latency	2510	2657	2862	2997	3542	4642	4877	5277	6402	18262
Avg Latency	2324.921	2442.854	2579.339	2673.136	3137.229	3919.9	4169.825	4396.996	5087.025	12755.718
Min Latency	2100	2162	2222	2277	2522	2897	2977	3062	3317	5982
Avg. delay Variation	8.314	5.403	9.179	5.386	7.636	7.146	5.632	9.314	4.986	9.09

The Extreme Networks BDX8 was tested across all 256 ports of 10GbE and 24 ports of 40GbE. Its average cut through latency ranged from a low of 2324 ns or 2.3 µs to a high of 12,755 ns or 12.7 µs at jumbo size 9216 Byte size frames for layer 2 traffic. Its average delay variation ranged between 5 and 9 ns, providing consistent latency across all packet sizes at full line rate; a welcome measurement for converged I/O implementations

Extreme Networks BlackDiamond® X8 Core Switch

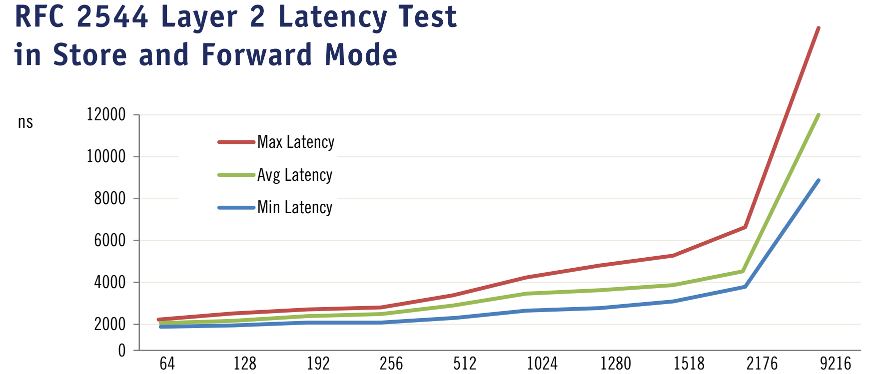
The Extreme Networks BDX8 was tested across all 256 ports of 10GbE and 24 ports of 40GbE. Its average store and forward latency ranged from a low of 2273 ns or 2.2 µs to a high of 12,009 ns or 12 µs at jumbo size 9216 Byte size frames for layer 2 traffic. Its average delay variation ranged between 5 and 9 ns, providing consistent latency across all packet sizes at full line rate.

For layer 3 traffic, Extreme Networks BDX8's measured average cut through latency ranged from a low of 2,321 ns at 64Bytes to a high of 12,719 ns or 12.7 µs at jumbo size 9216 Byte size frames. Its average delay variation for layer 3 traffic ranged between 4 and 9 ns, providing consistent latency across all packet sizes at full line rate; again a welcome measurement for converged I/O implementations.

For layer 3 traffic, Extreme Networks BDX8's measured average store and forward latency ranged from a low of 2,273 ns at 64Bytes to a high of 11,985 ns or 11.9 µs at jumbo size 9216 Byte size frames. Its average delay variation for layer 3 traffic ranged between 4 and 9 ns, providing consistent latency across all packet sizes at full line rate; again a welcome measurement for converged I/O implementations.

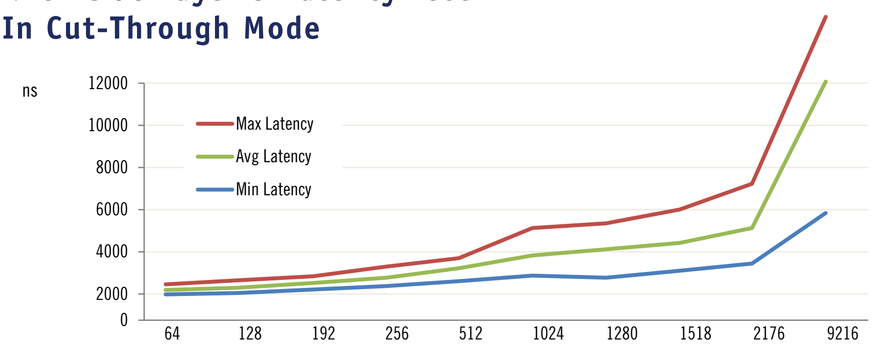
The above latency measures are the lowest we have measured for core switches by a large degree. The BDX8 forwards packets ten to six times faster than other core switches we've tested.

Extreme Networks BDX8
RFC 2544 Layer 2 Latency Test in Store and Forward Mode



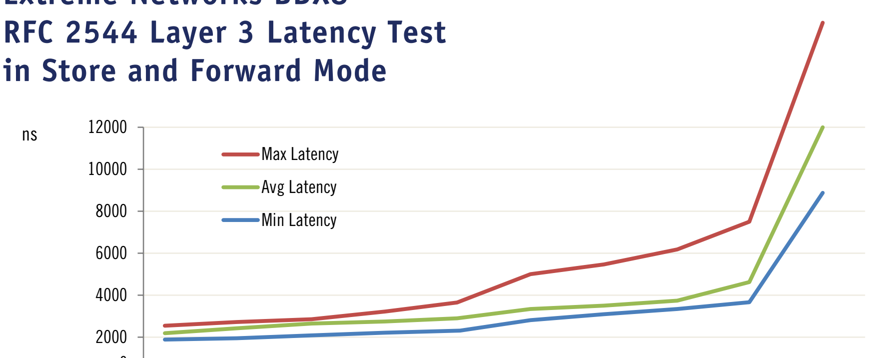
	64	128	192	256	512	1024	1280	1518	2176	9216
Max Latency	2482	2642	2797	2910	3442	4547	4990	5430	6762	21317
Avg Latency	2273.521	2351.750	2436.686	2493.575	2767.593	3331.511	3588.750	3830.261	4542.814	12095.05
Min Latency	2040	2100	2160	2180	2417	2780	2980	3140	3660	8960
Avg. delay Variation	8.314	5.403	9.179	5.386	7.636	7.146	5.632	9.314	4.986	9.079

Extreme Networks BDX8
RFC 2544 Layer 3 Latency Test In Cut-Through Mode



	64	128	192	256	512	1024	1280	1518	2176	9216
Max Latency	2550	2722	2902	3162	3770	5102	5430	6010	7290	21970
Avg Latency	2321.164	2440.482	2577.779	2668.411	3132.839	3898.596	4162.443	4406.857	5126.400	12719.646
Min Latency	2100	2162	2230	2302	2547	2870	2987	3070	3302	5997
Avg. delay Variation	9.982	5.454	9.2	5.411	7.593	7.164	5.529	9.321	4.982	9.118

Extreme Networks BDX8
RFC 2544 Layer 3 Latency Test in Store and Forward Mode



	64	128	192	256	512	1024	1280	1518	2176	9216
Max Latency	2542	2717	2937	3117	3657	5062	5537	6130	7677	25070
Avg Latency	2273.346	2353.996	2432.400	2492.879	2766.289	3324.279	3581.018	3826.468	4504.139	11985.946
Min Latency	2040	2100	2140	2220	2400	2780	3000	3160	3640	8960
Avg. delay Variation	9.982	5.454	9.2	5.411	7.593	7.164	5.529	9.321	4.982	9.118

Extreme Networks BlackDiamond® X8 Core Switch

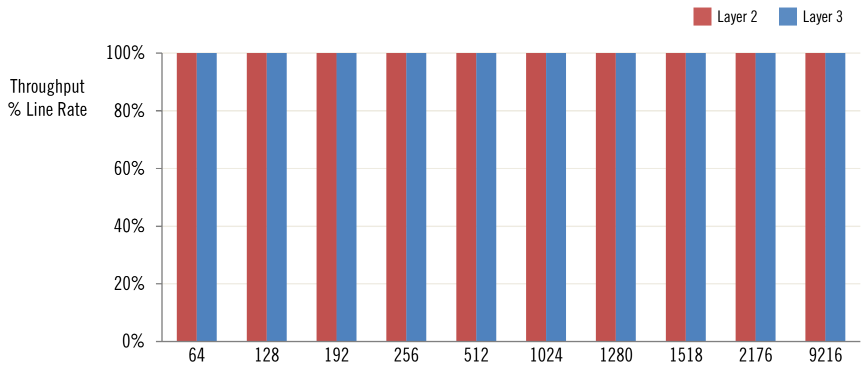
The Extreme Networks BDX8 demonstrated 100% throughput as a percentage of line rate across all 256-10GbE and 24-40GbE ports. In other words, not a single packet was dropped while the Extreme Networks BDX8 was presented with enough traffic to populate its 256-10GbE and 24-40GbE ports at line rate simultaneously for both L2 and L3 traffic flows measured in both cut-through and store and forward; a first in these Lippis/Ixia test.

The Extreme Networks BDX8 demonstrated 100% of aggregated forwarding rate as percentage of line rate during congestion conditions. We tested congestion in both 10GbE and 40GbE configurations. A single 10GbE port was flooded at 150% of line rate. In addition, a single 40GbE port was flooded at 150% of line rate; another first in these Lippis/Ixia test.

The Extreme Networks BDX8 did not use HOL blocking which means that as the 10GbE and 40GbE ports on the BDX8 became congested, it did not impact the performance of other ports. Back pressure was detected. The BDX8 did send flow control frames to the Ixia test gear signaling it to slow down the rate of incoming traffic flow.

Priority Flow Control (PFC) was enabled on the BDX8 during congestion test, as this is Extreme Network's direction to run Data Center Bridging protocols on the BDX8 in support of data center fabric requirements. We expect more core switches will enable and support PFC as this enables Ethernet fabrics to support storage and IP traffic flows under congestion conditions without packet loss, thus maintaining 100% throughput.

Extreme Networks BDX8
RFC 2544 L2 & L3 Throughput Test



	64	128	192	256	512	1024	1280	1518	2176	9216
Layer 2	100	100	100	100	100	100	100	100	100	100
Layer 3	100	100	100	100	100	100	100	100	100	100

Extreme Networks BDX8 RFC 2889 Congestion Test

150% of Line Rate into a single 10GbE

Frame (Bytes)	Agg Forwarding Rate (% Line Rate)	Head of Line Blocking	Back Pressure	Agg Flow Control Frames	
				LAYER 2	LAYER 3
64	100	no	yes	8641332	9016074
128	100	no	yes	5460684	5459398
192	100	no	yes	6726918	6399598
256	100	no	yes	7294218	5798720
512	100	no	yes	8700844	5081322
1024	100	no	yes	5887130	5342640
1280	100	no	yes	5076130	5076140
1518	100	no	yes	4841474	4840594
2176	100	no	yes	5268752	5268552
9216	100	no	yes	5311742	5219964

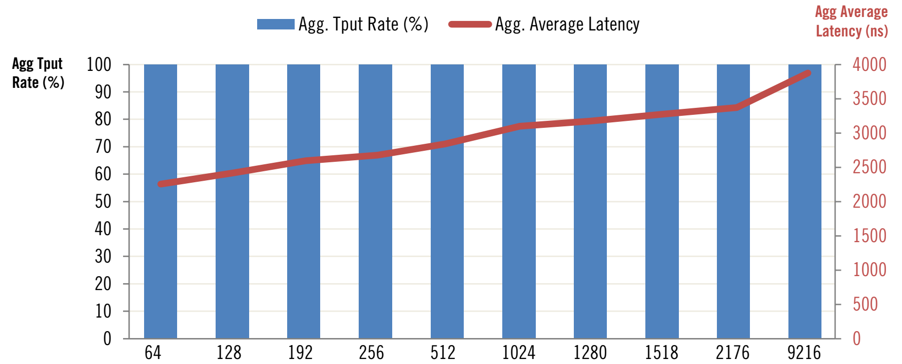
150% of Line Rate into a single 40GbE

Frame (Bytes)	Agg Forwarding Rate (% Line Rate)	Head of Line Blocking	Back Pressure	Agg Flow Control Frames LAYER 2	Agg Flow Control Frames LAYER 3
64	100	no	yes	11305858	11303930
128	100	no	yes	9679780	9682464
192	100	no	yes	9470606	9466194
256	100	no	yes	29566584	10193602
512	100	no	yes	23892956	8909568
1024	100	no	yes	22794016	8915370
1280	100	no	yes	24439074	7944288
1518	100	no	yes	21749118	9093160
2176	100	no	yes	20897472	9101282
9216	100	no	yes	12938948	8977442

Extreme Networks BlackDiamond® X8 Core Switch

The Extreme Networks BDX8 demonstrated 100% aggregated throughput for IP multicast traffic measured in cut through mode with latencies ranging from a 2,382 ns at 64Byte size packet to 3,824 ns at 9216Byte size packets.

Extreme Networks BDX8
RFC 3918 IP Multicast Test Cut Through Mode



	64	128	192	256	512	1024	1280	1518	2176	9216
Agg Tput Rate (%)	100	100	100	100	100	100	100	100	100	100
Agg. Aver. Latency	2382.767	2448.613	2519.792	2578.477	2835.100	3117.849	3143.043	3161.283	3217.358	3824.054

Extreme Networks BlackDiamond® X8 Core Switch

The Extreme Networks BDX8 performed very well under cloud simulation conditions by delivering 100% aggregated throughput while processing a large combination of east-west and north-south traffic flows. Zero packet loss was observed as its latency stayed under 4.6 µs and 4.8 µs measured in cut through and store and forward modes respectively. This measurement too breaks all previous records as the BDX8 is between 2 and 10 times faster in forwarding cloud based protocols under load.

The Extreme Networks BDX8 represents a new breed of cloud network core switches with power efficiency being a core value. Its WattsATIS/port is 8.1 and TEER value is 117. To put power measurement into perspective, other core switches are between 30% and 2.6 times less power efficient. In other words, the BDX8 is the most power efficient core switch tested to date in the Lippis/Ixia iSimCity lab.

Note that the total number of BDX8 10GbE ports to calculate Watts/10GbE were 352, as 24-40Gbs ports or 96 10GbE ports were added to the 256-10GbE ports for Watts/ATIS/port calculation. The 40GbE ports are essentially four 10GbE ports from a power consumption point of view.

While these are the lowest Watts/10GbE port and highest TEER values observed for core switches, the Extreme Networks BDX8's actual Watts/10GbE port is actually lower; we estimate approximately 5 Watts/10GbE port when fully populated with 756 10GbE or 192 40GbE ports. During the Lippis/Ixia test, the BDX8 was only populated to a third of its port capacity but equipped with power supplies, fans, management and switch fabric modules for full port density population. Therefore, when this full power capacity is divided across a fully populated BDX8, its WattsATIS per 10GbE Port will be lower than the measurement observed.

The Extreme Networks BDX8 power cost per 10GbE is calculated at $9.87 per year. The three year cost to power the Extreme Networks BDX8 is estimated at $10,424.05 and represents less than 1.7% of its list price. Keeping with data center best practices, its cooling fans flow air front to back.

Extreme Networks BDX8 Cloud Simulation Test

Traffic Direction	Traffic Type	Cut-Through Avg Latency (ns)	Store and Forward Avg Latency (ns)
East-West	Database_to_Server	2574	3773
East-West	Server_to_Database	1413	1339
East-West	HTTP	4662	4887
East-West	iSCSI-Server_to_Storage	1945	2097
East-West	iSCSI-Storage_to_Server	3112	3135
North-South	Client_to_Server	3426	3368
North-South	Server_to_Client	3576	3808

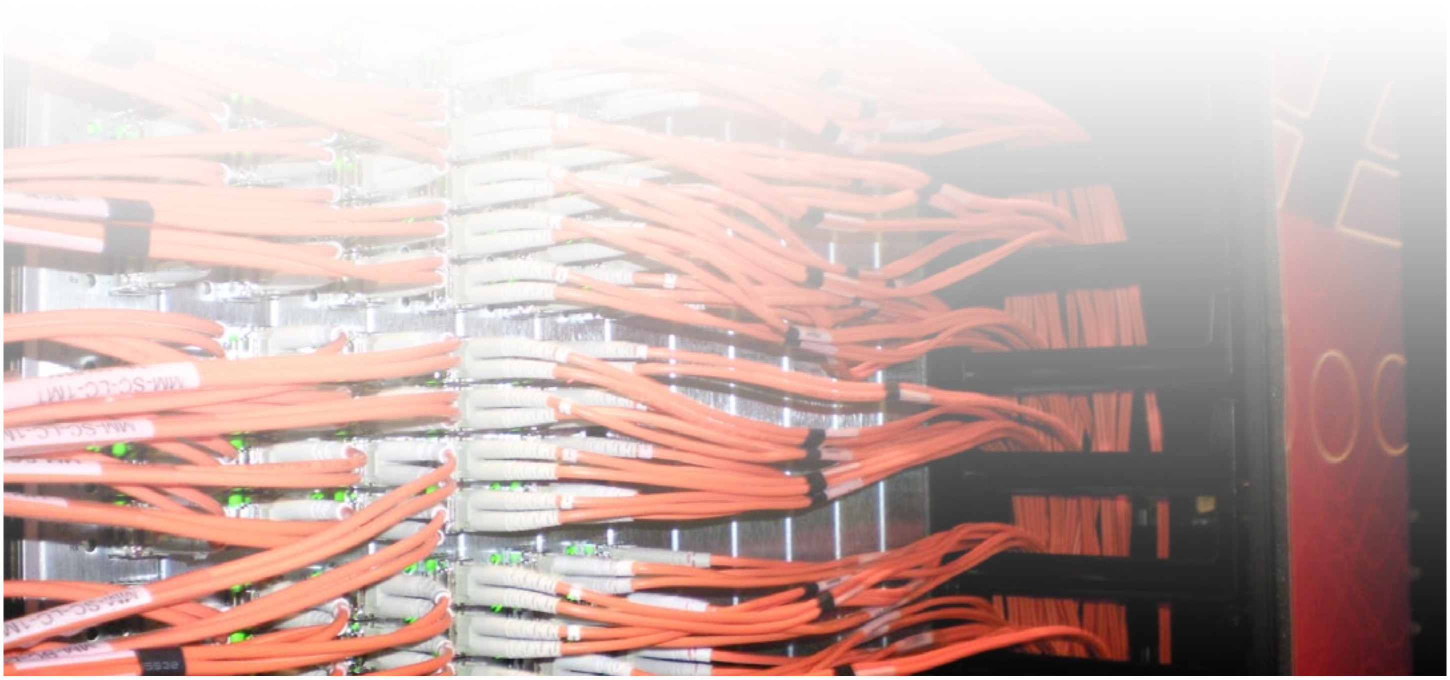
Extreme Networks BDX8 Power Consumption Test

$\text{Watts}_{\text{ATIS}}$/10GbE port	8.1
3-Year Cost/$\text{Watts}_{\text{ATIS}}$/10GbE	$29.61
Total power cost/3-Year	$10,424.05
3 yr energy cost as a % of list price	1.79%
TEER Value	117
Cooling	Front to Back, Reversible

Virtualization Scale Calculation

Within one layer 2 domain, the Extreme Networks BlackDiamond® X8 can support approximately 900 physical servers and 27,125 VMs, thanks to its support of 28K IP host route table size and 128K MAC address table size. This assumes 30 VMs per physical server with each physical server requiring two-host IP/ MAC/ARP entries; one for management, one for IP storage, etc into a data center switch. Each VM will require from a data center switch one MAC entry, one 32/IP host route and one ARP entry. In fact, its IP host route and MAC address table sizes are larger than the number of physical ports that it can support at this VM to physical server ratio.

From a deployment mode very large deployments are enabled. For example, assume 40 VMs/server and with a 28K host route table, 700 servers can be connected into a single BDX8 in an End of Row (EOR) configuration. Here the BD-X8 can serve as the routed boundary for the entire row accommodating 28,000 VMs and approximately 700 physical servers. Scaling further, multiple BDX8s each placed in EOR could be connected to a core routed BDX8 that serves as the Layer 3 interconnect for the EOR BDX8s. Since each EOR BDX8 is a Layer 3 domain, the core BDX8 does not see the individual IP host table entries from each VM (which are hidden behind the EOR BDX8). A very large 2-tier network with BDX8s can be created with the first tier being the EOR BDX8s which is routing, while the second tier are core BDX8(s) which is also routing, thus accommodate hundreds of thousands of VMs. The BDX8 can scale very large in its support of virtualized data center environments.



Discussion:

The Extreme Networks BDX8 seeks to offer high performance, low latency, high port density of 10 and 40GbE and low power consumption for private and public cloud networks. This architecture proved its value as it delivered the lowest latency measurements to date for core switches while populated with the equivalent of 352 10GbE ports running traffic at line rate. Not a single packet was dropped offering 100% throughput a line rate for the equivalent of 352 10GbE ports. Even at 8 Watts/10GbE port, the Extreme Networks BDX8 consumes the least amount of power of core switches measured to date at the Lippis/Ixia industry test at iSimCity. Observations of the BDX8's IP Multicast and congestion test performance were the best measured to date in terms of latency and throughput under congestion conditions in both 10GbE and 40GbE scenarios.

The Extreme Networks BDX8 was found to have low power consumption, front-to-back cooling, front-accessible components and an ultra compact form factor. The Extreme Networks BDX8 is designed to meet the requirements for private and public cloud networks. Based upon these Lippis/Ixia test, it achieves its design goals.

Extreme Networks BlackDiamond® X8 Core Switch 192-40GbE Test Results

In addition to the Lippis Report test above, we stressed the Extreme Networks BlackDiamond® X8 (BDX8) in a fully populated configuration of 192-40GbE ports, which is equivalent to 768-10GbE ports. This is the first time, in the industry, that such a test of this scale and magnitude has been conducted. This configuration allowed us to measure the (BDX8) coreswitch at full capacity for throughput, latency, jitter, IP Multicast, power consumption and cloud simulation. What follows are the results of these test which set a new bar for performance and power consumption.

Extreme Networks BlackDiamond® X8 Core Switch Test Configuration*

	Hardware	Software Version	Port Density
Device under test	BSX8 40G24X Modules http://www.extremenetworks.com	15.1.0.14	192-40GbE
Test Equipment	Ixia XM12 High Performance Chassis	IxOS 6.10 EA SP2 IxNetwork 6.10 EA	
Ixia Line Cards	Xcellon Flex 4x40GEQSFP+ (4 port 40G)		
	Xcellon Flex Combo 10/40GE AP (16 port 10G / 4 port 40G)		
	http://www.ixiacom.com/		
Cabling	Siemon QSFP+ Passive Copper Cable 40 GbE 3 meter copper QSFP-FA-010		
	www.siemon.com		

* Extreme Networks BDx8 was configured with 192-40GbE ports or an equivalent 768-10GbE ports.

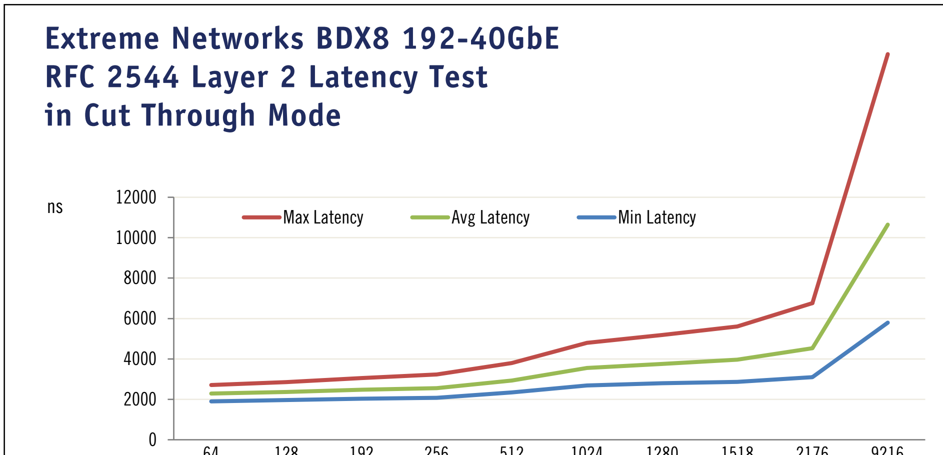
Just during the Lippis/Ixia test conducted at iSimCity, the Extreme Networks BDX8 was tested in both cut through and store and forward modes in an effort to understand the difference these latency measurements offer. Further, latest merchant silicon forwards packets in store and forward for smaller packets while larger packets are forwarded in cut-through, making this new generation of switches hybrid cut-through/store and forward devices. The BDX8 is unique in that its switching architecture forward packets in both cut-through and store-and-forward modes. As such, we measure the BDX8 latency with both methods.

The purpose is not to compare store-and-forward vs. cut-through but to demonstrate the low latency results obtained independent upon forwarding method. Note that other potential device-specific factors can impact latency too.

Extreme Networks BlackDiamond® X8 Core Switch 192-40GbE Test Results

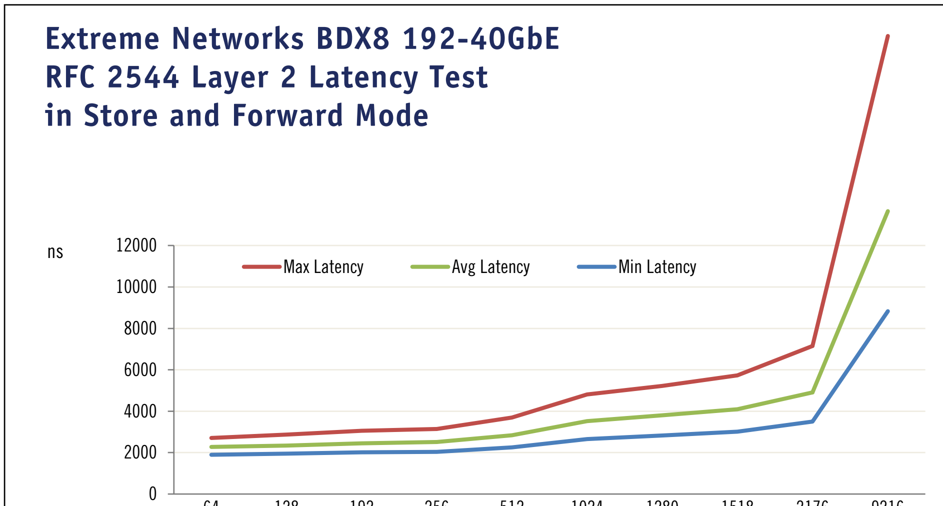
The Extreme Networks BDX8 was tested across all 192 ports of 40GbE. Its average cut-through latency ranged from a low of 2285 ns or 2.2 µs to a high of 10,645 ns or 10.6 µs at jumbo size 9216 Byte size frames for layer 2 traffic. Its average delay variation ranged between 5 and 9 ns, providing consistent latency across all packet sizes at full line rate--a welcome measurement for converged I/O implementations. Note that these jumbo frame latency results are better than the Lippis/Ixia jumbo latency test results by over 2µs!

The Extreme Networks BDX8 was tested across all 192 ports of 40GbE. Its average store-and-forward latency ranged from a low of 2271 ns or 2.2 µs to a high of 13,673 ns or 13.6µs at jumbo size 9216 Byte size frames for layer 2 traffic. Its average delay variation ranged between 4 and 9 ns, providing consistent latency across all packet sizes at full line rate.



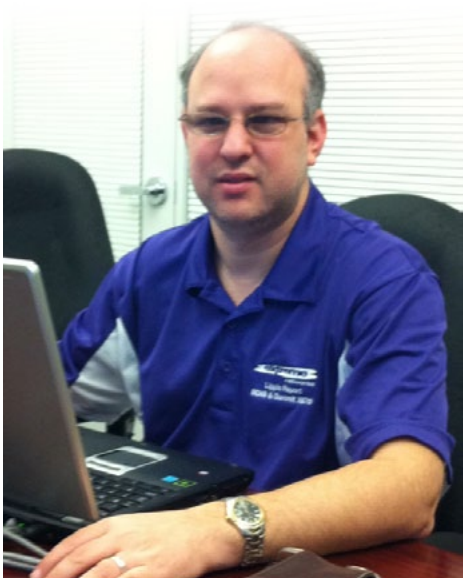
	64	128	192	256	512	1024	1280	1518	2176	9216
Max Latency	2710	2850	3050	3237	3797	4802	5190	5602	6762	19097
Avg Latency	2285	2361	2474	2560	2931	3551	3754	3965	4531	10645
Min Latency	1902	1970	2030	2077	2342	2690	2797	2870	3102	5797
Avg. Delay Variation	4.9	9.4	5.9	6.8	8.1	9.1	6.3	9.0	4.9	8.9

Test was done between far ports to test across chip performance/latency.



	64	128	192	256	512	1024	1280	1518	2176	9216
Max Latency	2702	2870	3057	3137	3690	4810	5222	5730	7150	22137
Avg Latency	2271	2341	2441	2512	2834	3515	3799	4091	4908	13673
Min Latency	1897	1950	2010	2030	2250	2657	2822	3010	3502	8827
Avg. Delay Variation	5.0	9.4	5.9	6.8	8.0	9.2	3.9	9.0	4.9	8.9

Test was done between far ports to test across chip performance/latency.



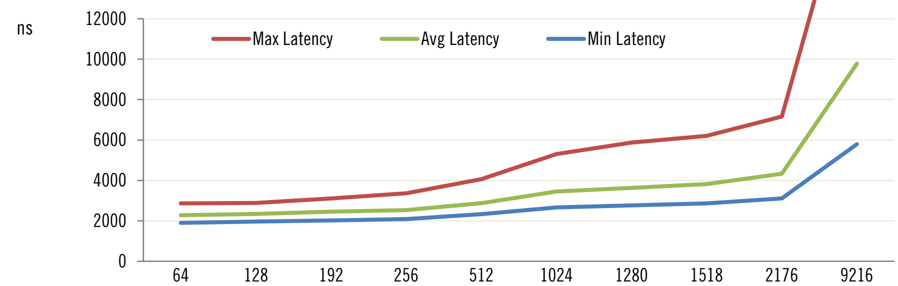
Extreme Networks BlackDiamond® X8 Core Switch 192-40GbE Test Results

For layer 3 traffic, Extreme Networks BDX8's measured average cut- through latency ranged from a low of 2,276 ns at 64 Bytes to a high of 9,780 ns or 9.7 µs at jumbo size 9216 Byte size frames. Its average delay variation for layer 3 traffic ranged between 4 and 9ns, providing consistent latency across all packet sizes at full line rate--again a welcome measurement for converged I/O implementations. Again, these jumbo frame latency results are better than the Lippis/Ixia jumbo latency test results by nearly 3 µs, providing lower latency or faster forwarding when fully populated with the equivalent of 768-10GbE ports.

For layer 3 traffic, Extreme Networks BDX8's measured average store and forward latency ranged from a low of 2,263 ns at 64 Bytes to a high of 12,805 ns or 12.8 µs at jumbo size 9216 Byte size frames. Its average delay variation for layer 3 traffic ranged between 4 and 9 ns, providing consistent latency across all packet sizes at full line rate—again, a welcome measurement for converged I/O implementations.

The above latency measures are the lowest we have measured for core switches by a large degree while populated with three times the number of ports of previous Lippis/Ixia test. The BDX8 forwards packets 10 to six times faster than other core switches we've tested while being three times greater in port capacity/load.

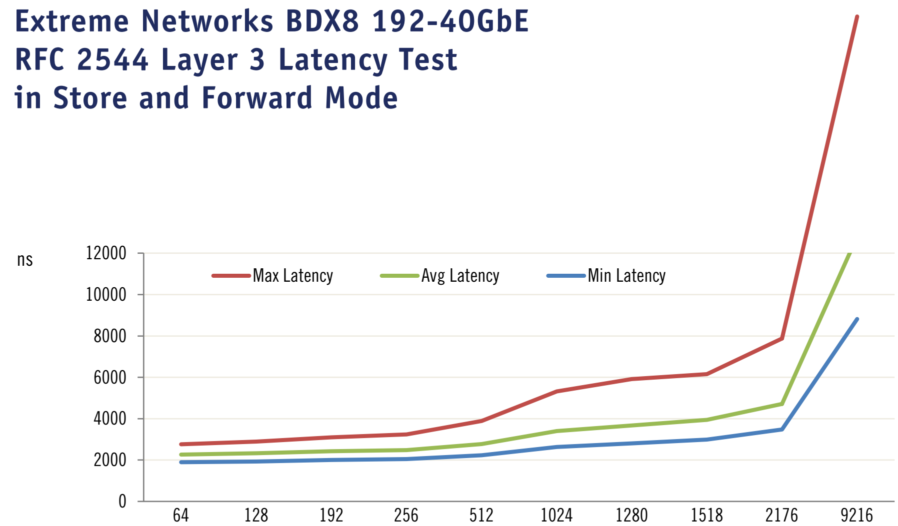
Extreme Networks BDX8 192-40GbE RFC 2544 Layer 3 Latency Test in Cut Through Mode



	64	128	192	256	512	1024	1280	1518	2176	9216
Max Latency	2862	2890	3107	3362	4062	5307	5877	6202	7162	20390
Avg Latency	2276	2349	2457	2530	2877	3448	3634	3822	4330	9780
Min Latency	1897	1962	2022	2090	2330	2670	2770	2862	3110	5797
Avg. Delay Variation	4	9	6	7	8	9	5	9	5	9

Test was done between far ports to test across chip performance/latency.

Extreme Networks BDX8 192-40GbE RFC 2544 Layer 3 Latency Test in Store and Forward Mode



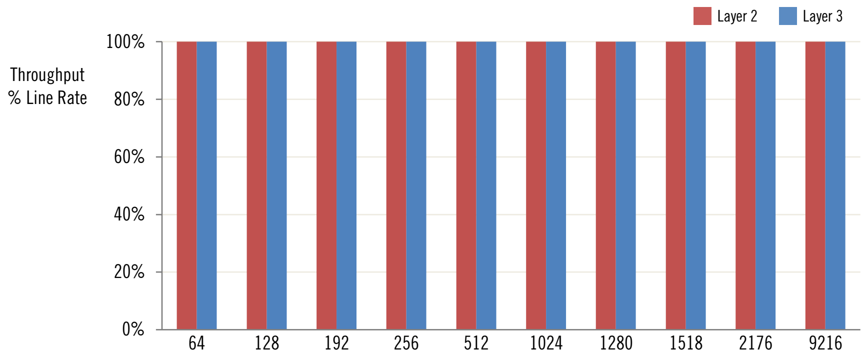
	64	128	192	256	512	1024	1280	1518	2176	9216
Max Latency	2757	2890	3102	3242	3890	5322	5917	6150	7877	23462
Avg Latency	2263	2328	2424	2483	2776	3401	3668	3946	4709	12805
Min Latency	1890	1930	1997	2042	2230	2630	2802	2990	3477	8822
Avg. Delay Variation	4	9	6	7	8	9	4	9	5	9

Test was done between far ports to test across chip performance/latency.

Extreme Networks BlackDiamond® X8 Core Switch 192-40GbE Test Results

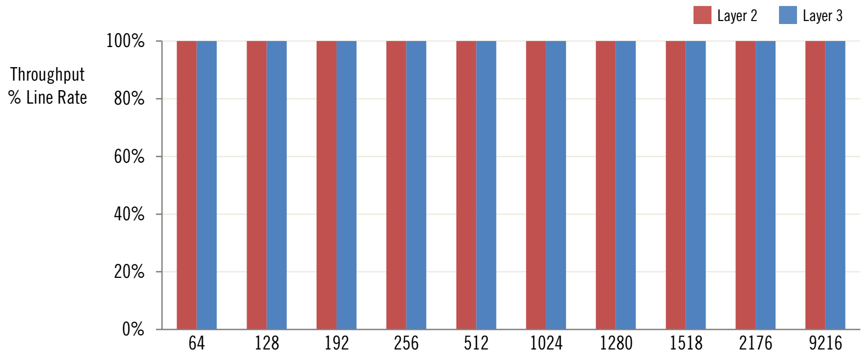
The Extreme Networks BDX8 demonstrated100% throughput as a percentage of line rate across all 192 40GbE ports. In other words, not a single packet was dropped while the Extreme Networks BDX8 was presented with enough traffic to populate its fully configured 192-40GbE ports at line rate for both L2 and L3 traffic flows measured in both cut-through and store-and-forward—a first in these Lippis/Ixia test and the industry.

Extreme Networks BDX8 192-40GbE
RFC 2544 L2 & L3 Throughput Test - Cut Through Mode



	64	128	192	256	512	1024	1280	1518	2176	9216
Layer 2	100	100	100	100	100	100	100	100	100	100
Layer 3	100	100	100	100	100	100	100	100	100	100

Extreme Networks BDX8 192-40GbE
RFC 2544 L2 & L3 Throughput Test - Store and Forward Mode



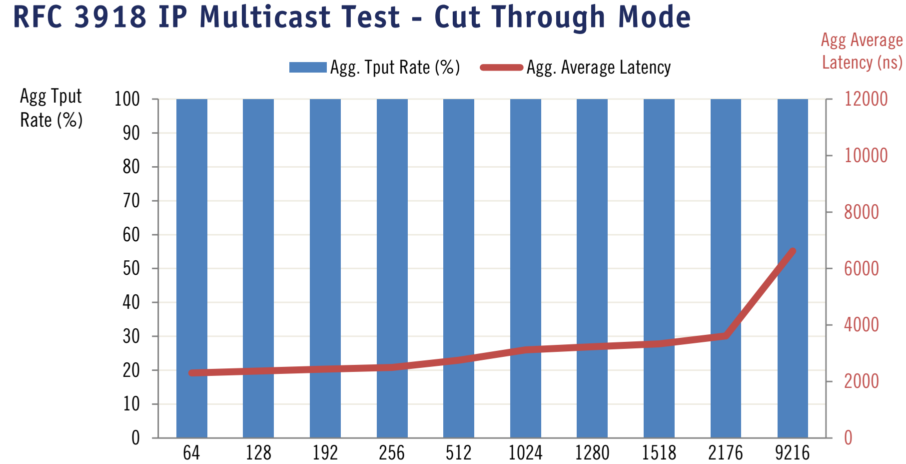
	64	128	192	256	512	1024	1280	1518	2176	9216
Layer 2	100	100	100	100	100	100	100	100	100	100
Layer 3	100	100	100	100	100	100	100	100	100	100

Extreme Networks BlackDiamond® X8 Core Switch 192-40GbE Test Results

The Extreme Networks BDX8 demonstrated 100% aggregated throughput for IP multicast traffic measured in cut-through and store-and-forward mode. IP Multicast latency measured in cut-through mode ranged from a 2,302 ns at 64 Byte size packet to 6,625 ns at 9216 Byte size packets. IP Multicast latency measured in store-and-forward mode ranged from a 2,303 ns at 64Byte size packet to 11,409 ns at 9216 Byte size packets.

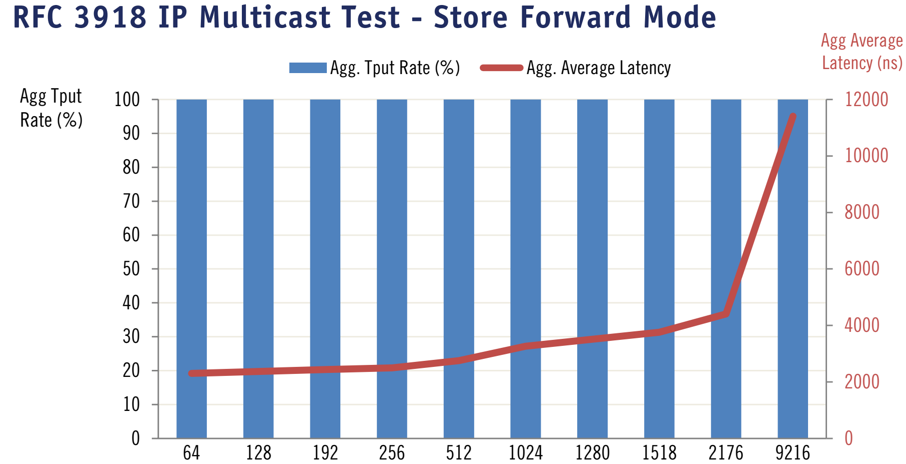
The BDX8 delivered the fastest IP Multicast performance measured to date, being able to forward IP Multicast packets between 3 and 11 times faster than previous core switch measures. Note that the only difference is that the BDX8 was configured with three times the number of equivalent 10GbE ports, making its performance even more impressive.

Extreme Networks BDX8 192-40GbE
RFC 3918 IP Multicast Test - Cut Through Mode



	64	128	192	256	512	1024	1280	1518	2176	9216
Agg Tput Rate (%)	100	100	100	100	100	100	100	100	100	100
Agg. Average Latency	2302	2367	2438	2498	2752	3117	3227	3332	3612	6625

Extreme Networks BDX8 192-40GbE
RFC 3918 IP Multicast Test - Store Forward Mode



	64	128	192	256	512	1024	1280	1518	2176	9216
Agg Tput Rate (%)	100	100	100	100	100	100	100	100	100	100
Agg. Average Latency	2303	2367	2438	2497	2752	3261	3514	3756	4410	11409

Extreme Networks BlackDiamond® X8 Core Switch 192-40GbE Test Results

The Extreme Networks BDX8 performed very well under cloud simulation conditions by delivering 100% aggregated throughput while processing a large combination of east-west and north-south traffic flows. Zero packet loss was observed as its latency stayed under 3 µs and 3.2µs measured in cut-through and store-and-forward modes, respectively. Note that these cloud simulation test results are better than the Lippis/Ixia industry test, thanks to the use of 40GbE. The use of 40GbE reduced latency by as much as 2.6 µs and 3 µs for select cloud protocols measured in cut-through and store-and-forward modes, respectively. The maximum latency observed during the cloud simulation test in the Lippis/Ixia industry test was reduced by 1.6 µs for both cut-through and store-and-forward measurements during the 192-40GbE custom tests. This measurement breaks all previous records as the BDX8 is between 3 and 17 times faster in forwarding cloud-based protocols under load.

The Extreme Networks BDX8 represents a new breed of cloud network core switches with power efficiency being a core value. Its WattsATIS/port is a low 5.62 and TEER value is 168. To put power measurement into perspective, other core switches consume between 1.8 to 3.8 times more power per port. In other words, the BDX8 is the most power efficient core switch tested to date by Lippis/Ixia, consuming nearly 50% to 75% less power on a per 10GbE port basis.The BDX8's WattsATIS/port of 5.62 is slightly higher than a typical holiday Christmas bulb.

Note that the total number of BDX8 10GbE ports to calculate Watts/ATIS per 10GbE was 768…as 192 40Gbs ports equal 768 10GbE ports. The 40GbE ports are essentially four 10GbE ports from a power consumption point of view.

The Extreme Networks BDX8 power cost per 10GbE is calculated at $6.85 per year. The three-year cost to power the Extreme Networks BDX8 fully populated with 768-10GbE is estimated at $15,780. Keeping with data center best practices, its cooling fans flow air front to back.

Extreme Networks BDX8 192-40GbE Cloud Simulation Test

Traffic Direction	Traffic Type	Cut-Through Avg Latency (ns)	Store and Forward Avg Latency (ns)
East-West	Database_to_Server	2948	3244
East-West	Server_to_Database	860	838
East-West	HTTP	2623	2860
East-West	iSCSI-Server_to_Storage	1047	970
East-West	iSCSI-Storage_to_Server	3034	3118
North-South	Client_to_Server	1265	1234
North-South	Server_to_Client	917	788

Extreme Networks BDX8 192-40GbE Power Consumption Test

$\text{Watts}_{\text{ATIS}}$/10GbE port	5.62
3-Year Cost/$\text{Watts}_{\text{ATIS}}$/10GbE	$20.55
1-Year Cost/$\text{Watts}_{\text{ATIS}}$/10GbE	$6.85
Total power cost/3-Year	$15,779.97
TEER Value	168
Cooling	Front to Back, Reversible

* The Extreme Networks BDX8 was fully populated with 192-40GbE ports
FORMULA USED=(WATIS/1000)*(3*365*24)*(0.1046)*(1.33), where WATIS= ATIS weighted average power in Watts.
3*365*24 = 3 years @ 365 days/yr @ 24 hrs/day.
0.1046 = U.S. Average retail cost (in US$) of commercial grade power as of June 2010 as per Dept. of Energy Electric Power Monthly.
(http://www.eia.doe.gov/cneaf/electricity/epm/table5_6_a.html) 1.33 = Factor to account for power costs plus cooling costs @ 33% of power costs.
Factor to account for power costs plus cooling costs @ 33% of power costs. IMIX used for traffic generation during power consumption test.

Discussion - 192-40GbE Results:

The Extreme Networks BDX8 offers the highest port density of 192-40GbE and 768-10GbE in the industry. Only one other vendor has announced a core switch with 768 10GbE, but it has not been tested in an open industry Lippis/Ixia test. The Extreme Networks BDX8 offers high performance, low latency, high-port density of 10GbE and 40GbE plus low power consumption for private and public cloud networks. This architecture proved its value as it delivered the lowest latency measurements to date for core switches while populated with the 192-40GbE ports or the equivalent of 768-10GbE ports running traffic at line rate. Not a single packet was dropped, offering 100% throughput, at line rate for the equivalent of 768 10GbE ports in both cut-through and store-and-forward measurement modes. At 5.6 Watts/10GbE port, the Extreme Networks BDX8 consumes the least amount of power of core switches measured to date in the Lippis/Ixia industry set of test. Observations of the BDX8's IP Multicast performance were the best measured to date, too, in terms of latency and throughput. Based upon our previous Lippis/Ixia tests and to the best of our knowledge the Extreme Networks BDX8 is the fastest Ethernet core switch in the world thanks to its ability to forward packets on a single to low double digit microsecond scale over a wide range of packet sizes.

The Extreme Networks BDX8 was found to have very low power consumption, front-back cooling, front-accessible components and an ultracompact form factor. The Extreme Networks BDX8 is designed to meet the requirements for private and public cloud networks. Based upon these Lippis/Ixia test, it achieves its design goals.

Core/Spine Cross-Vendor Analysis

There were four Core/Spine Switches evaluated for performance and power consumption in the Lippis/Ixia test. These participating vendors were:

Alcatel-Lucent OmniSwitch 10K

Arista 7504 Series Data Center Switch

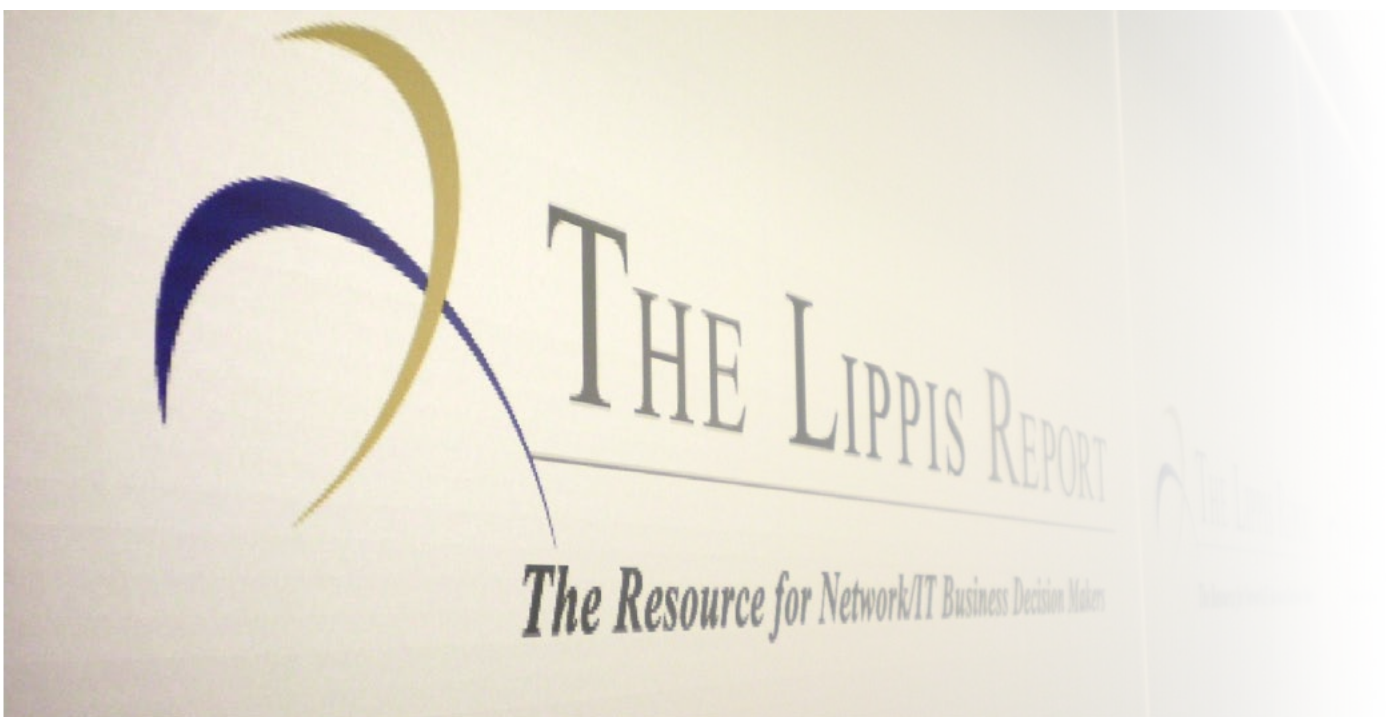
Extreme Networks BlackDiamond® X8

Juniper Network EX Series EX8200 Ethernet Switch

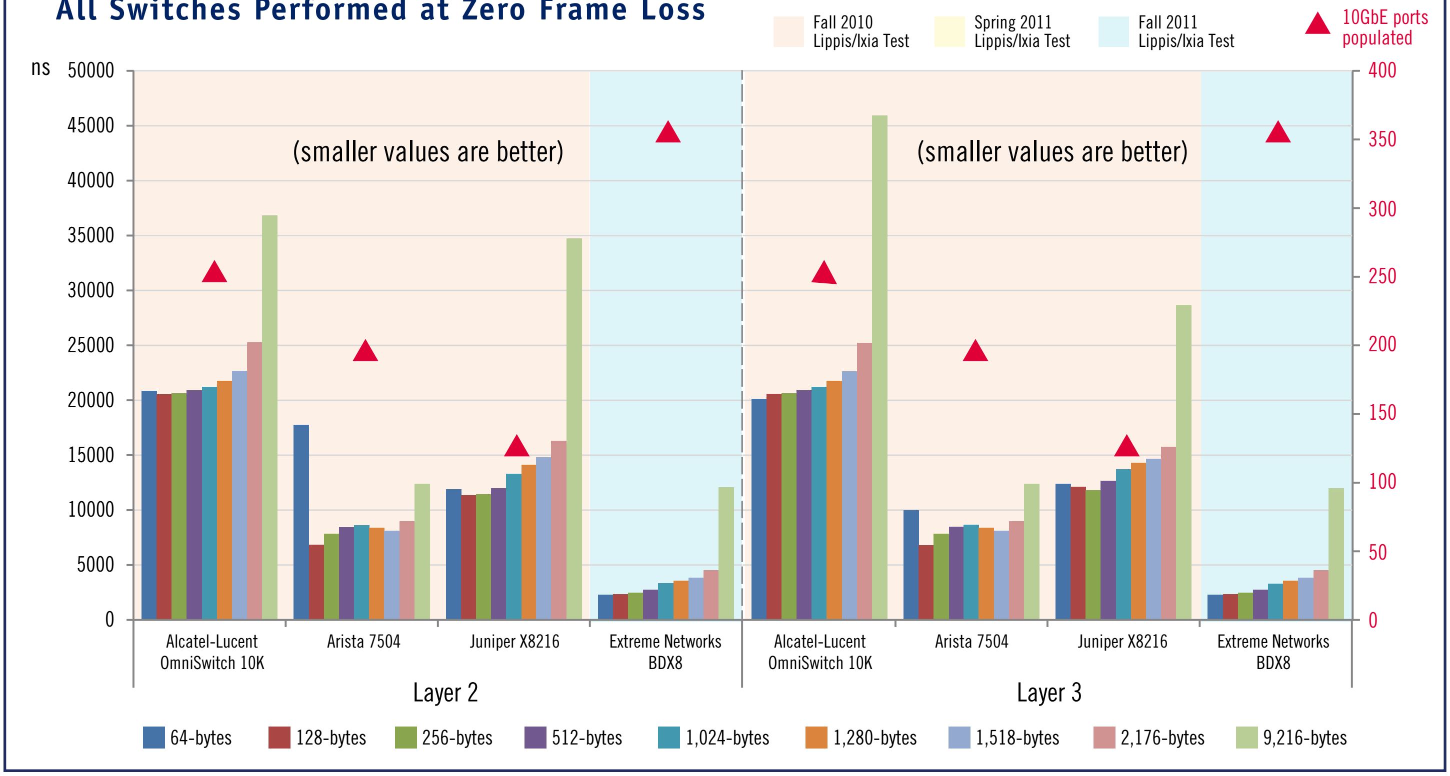
These switches represent the state-of-the-art of computer network hardware and software engineering, and are central to private/public data center cloud computing infrastructure. If not for this category of Ethernet switching, cloud computing would not exist. The Lippis/Ixia public test was the first evaluation for every Core switch tested. Each supplier's Core switch was evaluated for its fundamental performance and power consumption features. The Lippis/Ixia test results demonstrate that these new Core switches provide state-of-the-art performance at efficient power consumption levels not seen before. The port density tested for these Core switches ranged from 128 10GbE ports to a high of 256 10GbE and for the first time 24-40GbE ports for the Extreme Networks BlackDiamond® X8

IT business leaders are responding favorably to Core switches equipped with a value proposition of high performance, high port density, competitive acquisition cost, virtualization aware services, high reliability and low power consumption. These Core switches currently are in high demand with quarterly revenues for mid-size firms in the $20 to $40M plus range. The combined market run rate for both ToR and Core 10GbE switching is measured in the multi-billion dollar range. Further, Core switch price points on a 10GbE per port basis are a low of $1,200 to a high of $6,093. 40GbE core ports are priced between 3 and 4 times that of 10GbE core ports. Their list price varies from $230K to $780K with an average order usually being in the million plus dollar range. While there is a large difference in list price as well as price per port between vendors, the reason is found in the number of network services supported by the various suppliers and 10GbE/40GbE port density.

We compare each of the above firms in terms of their ability to forward packets: quickly (i.e., latency), without loss or their throughput at full line rate, when ports are oversubscribed with network traffic by 150%, in IP multicast mode and in cloud simulation. We also measure their power consumption as described in the Lippis Test Methodology section.



Core Switches RFC 2544 Latency Test - Average Latency
All Switches Performed at Zero Frame Loss



	Layer 2				Layer 3			
Frames/ Packet Size (bytes)	Alcatel-Lucent* OmniSwitch 10K	Arista 7504	Juniper EX8126	Extreme Networks BDX8	Alcatel-Lucent* OmniSwitch 10K	Arista 7504	Juniper EX8126	Extreme Networks BDX8
64	20864.00	17791.20	11891.60	2273.50	20128.00	9994.88	12390.38	2273.30
128	20561.00	6831.50	11366.20	2351.80	20567.00	6821.32	12112.50	2354.00
256	20631.00	7850.70	11439.60	2493.60	20638.00	7858.69	11814.05	2492.90
512	20936.00	8452.00	11966.80	2767.60	20931.00	8466.53	12685.41	2766.30
1,024	21216.00	8635.50	13322.90	3331.50	21211.00	8647.96	13713.25	3324.30
1,280	21787.00	8387.00	14131.80	3588.80	21793.00	8387.26	14318.48	3581.00
1,518	22668.00	8122.10	14789.60	3830.30	22658.00	8112.95	14664.84	3826.50
2,176	25255.00	8996.00	16302.70	4542.80	25242.00	8996.55	15767.44	4504.10
9,216	36823.00	12392.80	34722.70	12095.10	45933.00	12387.99	28674.08	11985.90

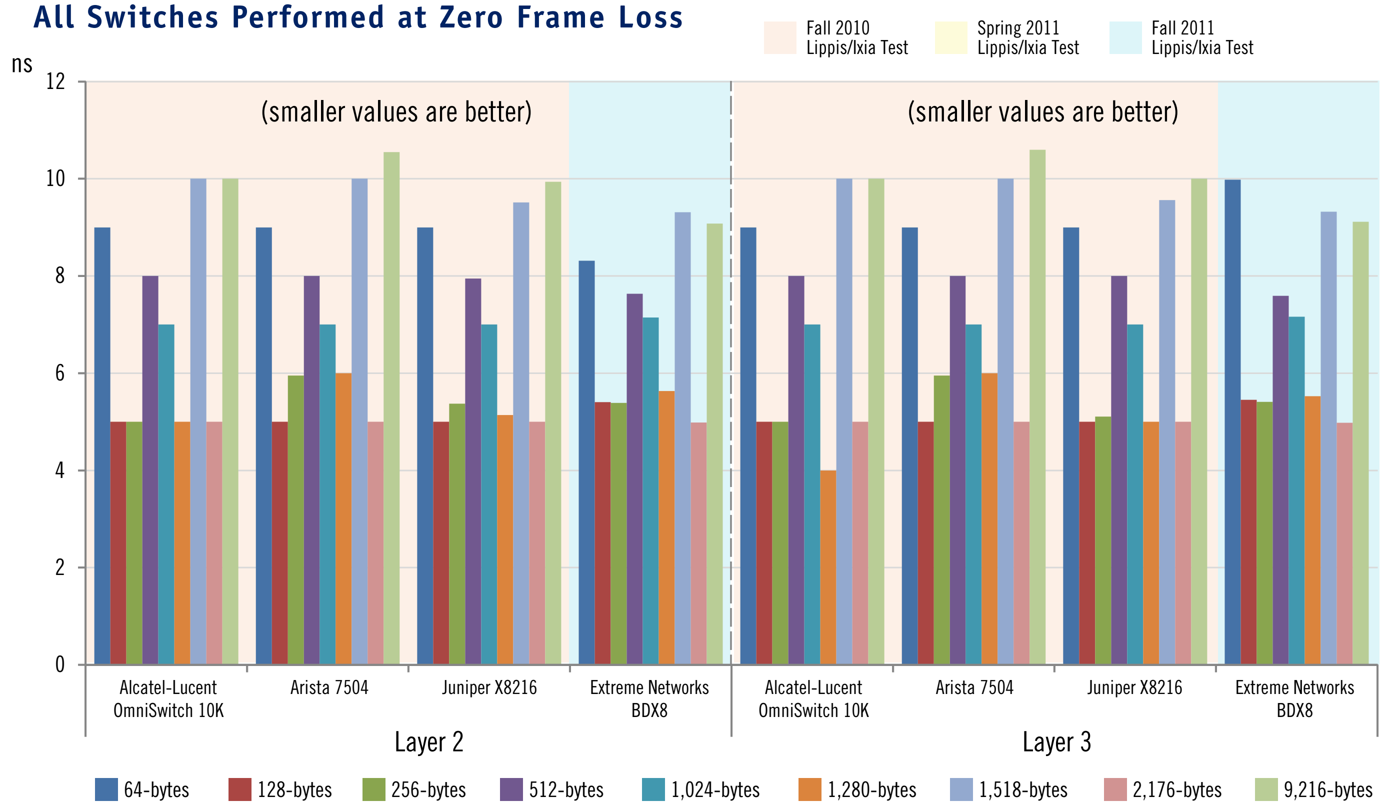
* RFC2544 was conducted on the OmniSwitch 10K via default configuration, however Alcatel-Lucent has an optimized configuration for low latency networks which it says improves these results.

We show average latency and average delay variation across all packet sizes for layer 2 and 3 forwarding. Measurements were taken from ports that were far away from each other to demonstrate the consistency of performance between modules and across their backplanes.

The clear standout here is the Extreme Networks BlackDiamond® X8 latency measurements which are 2 to 9 times faster at forwarding packets than all others.

Another standout here is the Arista 7504‘s large latency at 64Byte size packets. According to Arista, the Arista 7500 series performance and design are fine tuned for applications its customers use. These applications use mixed packet sizes, rather than an artificial stream of wire speed 64 byte packets.

Core Switches RFC 2544 Latency Test - Average Delay Variation
All Switches Performed at Zero Frame Loss

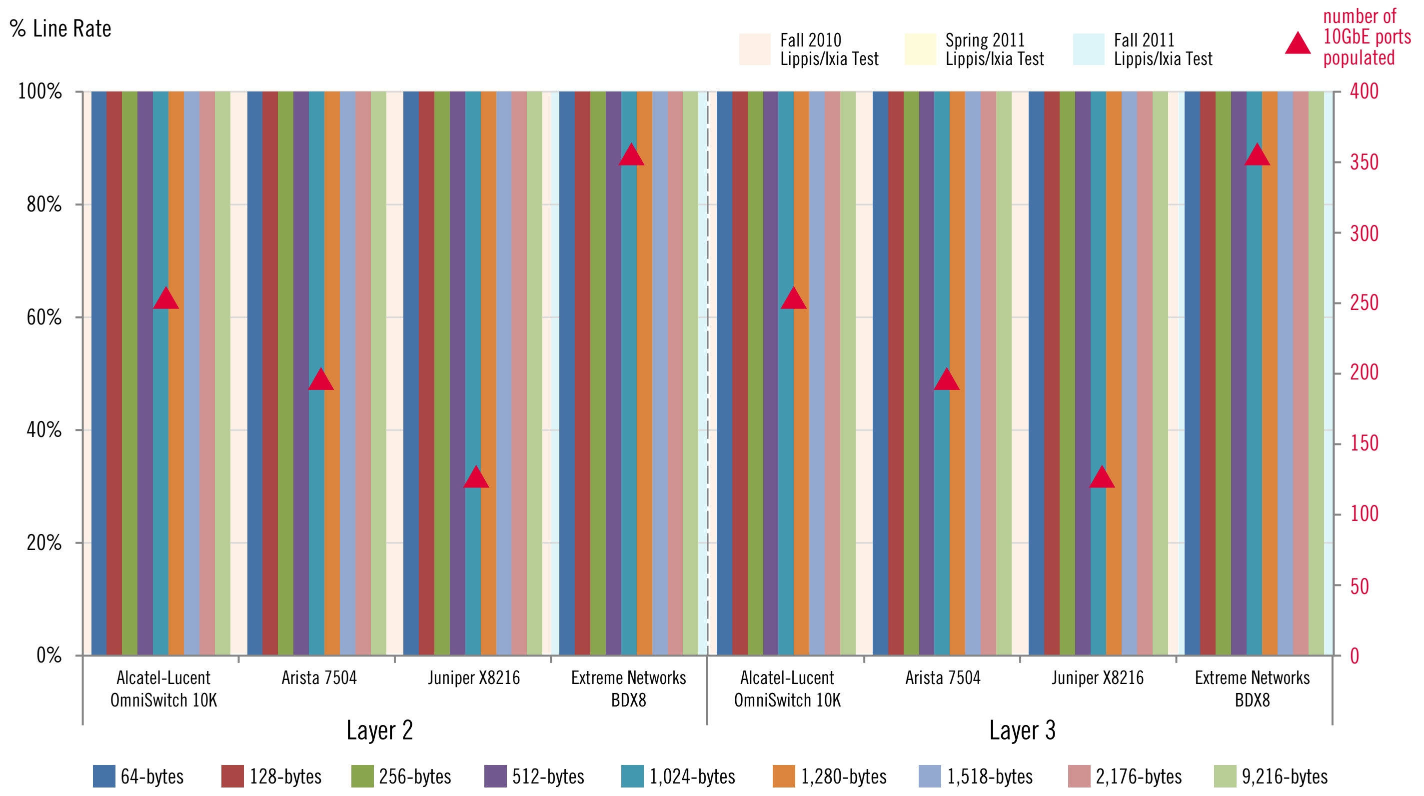


	Layer 2				Layer 3			
Frames/ Packet Size (bytes)	Alcatel-Lucent* OmniSwitch 10K	Arista 7504	Juniper EX8126	Extreme Networks BDX8	Alcatel-Lucent* OmniSwitch 10K	Arista 7504	Juniper EX8126	Extreme Networks BDX8
64	9.00	9.00	9.00	8.30	9.00	9.00	9.00	9.98
128	5.00	5.00	5.00	5.40	5.00	5.00	5.00	5.50
256	5.00	6.00	5.40	5.40	5.00	5.95	5.11	5.40
512	8.00	8.00	7.90	7.60	8.00	8.00	8.00	7.60
1,024	7.00	7.00	7.00	7.10	7.00	7.00	7.00	7.20
1,280	5.00	6.00	5.10	5.60	4.00	6.00	5.00	5.50
1,518	10.00	10.00	9.50	9.30	10.00	10.00	9.56	9.30
2,176	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
9,216	10.00	10.50	9.90	9.10	10.00	10.60	10.00	9.10

* RFC2544 was conducted on the OmniSwitch 10K via default configuration, however Alcatel-Lucent has an optimized configuration for low latency networks which it says improves these results.

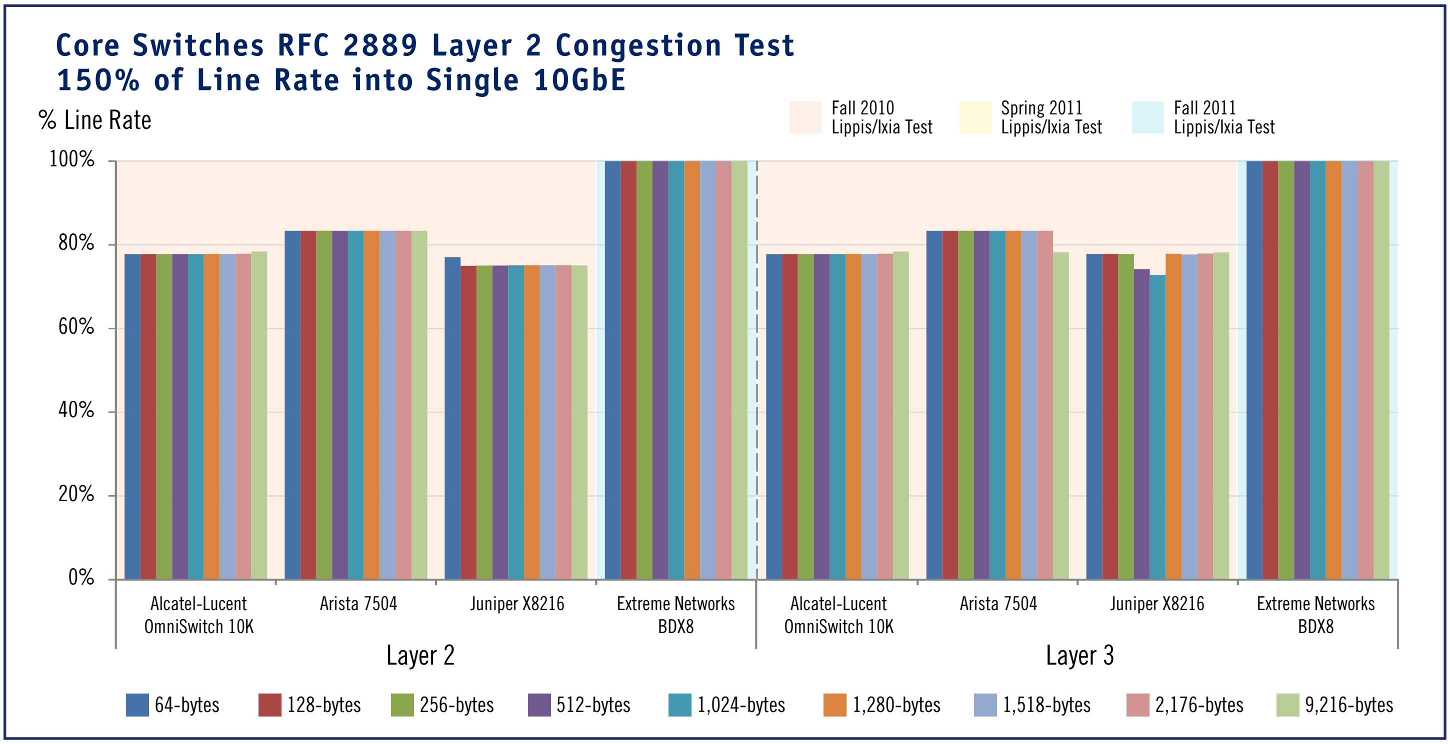
Notice there is little difference in average delay variation across all vendors proving consistent latency under heavy load at zero packet loss for L2 and L3 forwarding. Average delay variation is in the 5 to 10 ns range. Difference does reside within average latency between suppliers, thanks to different approaches to buffer management and port densities.

Core Switches RFC 2544 Throughput Test



	Layer 2				Layer 3			
Frames/ Packet Size (bytes)	Alcatel-Lucent OmniSwitch 10K	Arista 7504	Juniper EX8126	Extreme Networks BDX8	Alcatel-Lucent OmniSwitch 10K	Arista 7504	Juniper EX8126	Extreme Networks BDX8
64	100%	100%	100%	100%	100%	100%	100%	100%
128	100%	100%	100%	100%	100%	100%	100%	100%
256	100%	100%	100%	100%	100%	100%	100%	100%
512	100%	100%	100%	100%	100%	100%	100%	100%
1,024	100%	100%	100%	100%	100%	100%	100%	100%
1,280	100%	100%	100%	100%	100%	100%	100%	100%
1,518	100%	100%	100%	100%	100%	100%	100%	100%
2,176	100%	100%	100%	100%	100%	100%	100%	100%
9,216	100%	100%	100%	100%	100%	100%	100%	100%

As expected, all switches are able to forward L3 packets at all sizes at 100% of line rate with zero packet loss, proving that these switches are high performance devices. Note that while all deliver 100% throughput the Extreme Networks BlackDiamond® X8 was the only device configured with 24-40GbE ports plus 256-10GbE ports.



	Layer 2												Layer 3											
Frames/ Packet Size (bytes)	Alcatel-Lucent OmniSwitch 10K			Arista 7504			Juniper EX8126			Extreme Networks BDX8			Alcatel-Lucent OmniSwitch 10K			Arista 7504			Juniper EX8126			Extreme Networks BDX8		
	AFR%	HOL	BP	AFR%	HOL	BP*	AFR%	HOL	BP	AFR%	HOL	BP	AFR%	HOL	BP	AFR%	HOL	BP*	AFR%	HOL	BP	AFR%	HOL	BP
64	77.80%	N	N	83.34%	N	Y	77.01%	N	N	100%	N	Y	77.80%	N	N	84.34%	N	Y	77.81%	N	N	100%	N	Y
128	77.80%	N	N	83.34%	N	Y	75.01%	N	N	100%	N	Y	77.80%	N	N	84.34%	N	Y	77.84%	N	N	100%	N	Y
256	77.80%	N	N	83.35%	N	Y	75.02%	N	N	100%	N	Y	77.80%	N	N	84.35%	N	Y	77.83%	N	N	100%	N	Y
512	77.80%	N	N	83.35%	N	Y	75.03%	N	N	100%	N	Y	77.80%	N	N	84.35%	N	Y	74.20%	N	N	100%	N	Y
1,024	77.80%	N	N	83.36%	N	Y	76.06%	N	N	100%	N	Y	77.80%	N	N	84.36%	N	Y	72.78%	N	N	100%	N	Y
1,280	77.80%	N	N	83.36%	N	Y	76.06%	N	N	100%	N	Y	77.80%	N	N	84.36%	N	Y	77.87%	N	N	100%	N	Y
1,518	77.80%	N	N	83.36%	N	Y	75.12%	N	N	100%	N	Y	77.80%	N	N	84.36%	N	Y	77.69%	N	N	100%	N	Y
2,176	77.90%	N	N	83.36%	N	Y	76.06%	N	N	100%	N	Y	77.90%	N	N	84.36%	N	Y	77.91%	N	N	100%	N	Y
9,216	78.40%	N	N	83.36%	N	Y	75.06%	N	N	100%	N	Y	78.40%	N	N	78.21%	Y	N	78.16%	N	N	100%	N	Y

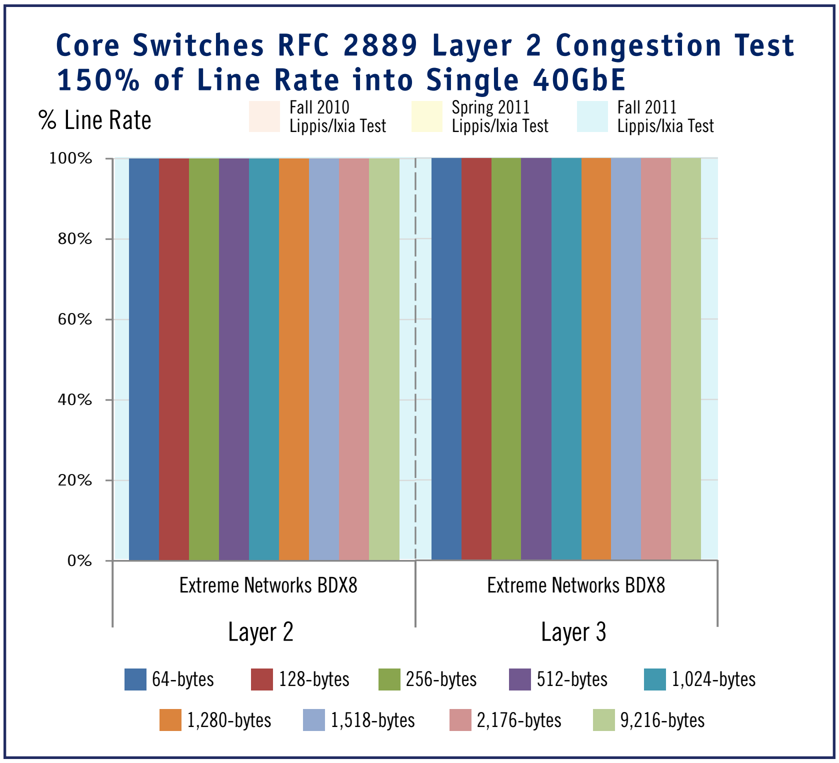
AFR% = Agg Forwarding Rate (% Line Rate) HOL = Head of Line Blocking BP = Back Pressure

* Note that while the Arista's 7504 shows back pressure, in fact there is none. All test equipment including IXIA calculates back pressure per RFC 2889 paragraph 5.5.5.2. which states that if the total number of received frames on the congestion port surpasses the number of transmitted frames at MOL (Maximum Offered Load) rate then back pressure is present. Thanks to the 7504's 2.3GB of packet buffer memory it can overload ports with more packets than the MOL, therefore, IXIA or any test equipment "calculates/sees" back pressure, but in reality this is an anomaly of the RFC testing method and not the 7504. The Arista 7504 can buffer up 40ms of traffic per port at 10GbE speeds which is 400K bits or 5,425 packets of 9216 bytes.

All Core switches performed extremely well under congestion conditions with nearly no HOL blocking and between 77 to 100% aggregated forwarding as a percentage of line rate. It's expected that at 150% of offered load to a port, that that Core switch's port would show 33% loss if it's receiving at 150% line rate and not performing back pressure or flow control.

A few standouts arose in this test. First, Arista's 7504 delivered nearly 84% of aggregated forwarding rate as percentage of line rate during congestion conditions for both L2 and L3 traffic flows, the highest for all suppliers without utilizing flow control. This is due to its generous 2.3GB of buffer memory design as well as its VOQ architecture. Note also that Arista is the only Core switch that showed HOL blocking for 9216 byte size packets at L3. Note the 7504 was in beta testing at the time of the Lippis/IXIA test and there was a "*corner case*" at 9216 bytes at L3 that needed further tuning. Arista states that its production code provides wire-speed L2/L3 performance at all packet sizes without any head of line blocking

Second, Alcatel-Lucent and Juniper delivered consistent and perfect performance with no HOL blocking or back pressure for all packet sizes during both L2 and L3 forwarding.



	Layer 2			Layer 3		
Frames/ Packet Size (bytes)	Extreme Networks BDX8			Extreme Networks BDX8		
	AFR%	HOL	BP	AFR%	HOL	BP
64	100%	N	Y	100%	N	Y
128	100%	N	Y	100%	N	Y
256	100%	N	Y	100%	N	Y
512	100%	N	Y	100%	N	Y
1,024	100%	N	Y	100%	N	Y
1,280	100%	N	Y	100%	N	Y
1,518	100%	N	Y	100%	N	Y
2,176	100%	N	Y	100%	N	Y
9,216	100%	N	Y	100%	N	Y

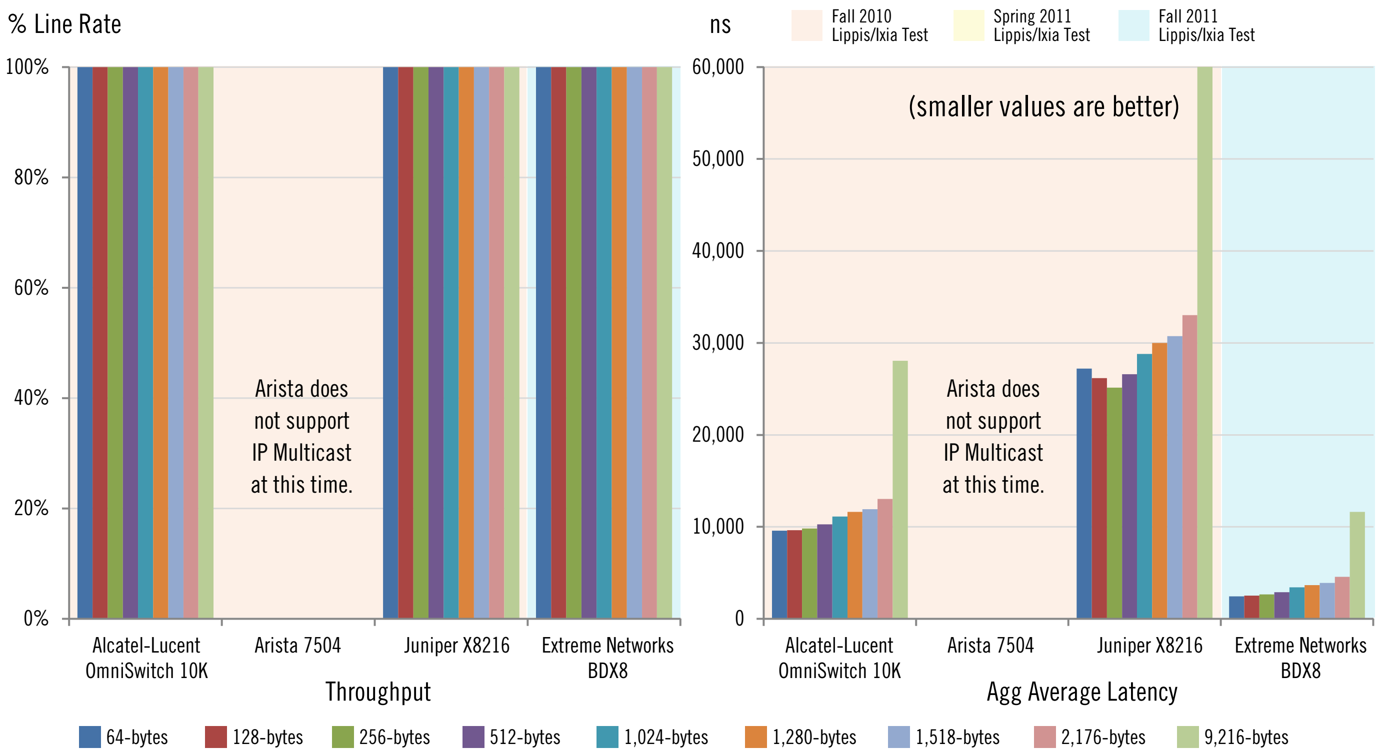
AFR% = Agg Forwarding Rate (% Line Rate) HOL = Head of Line Blocking BP = Back Pressure

* Extreme Networks' BDX8 was the only core switch to utilize back pressure to slow down the rate of traffic flow so as to maintain 100% throughput

The Extreme Networks BDX8 demonstrated 100% of aggregated forwarding rate as percentage of line rate during congestion conditions. We tested congestion in both 10GbE and 40GbE configurations, a first for these test.

The Extreme Networks BDX8 sent flow control frames to the Ixia test gear signaling it to slow down the rate of incoming traffic flow. We anticpate that this will become a new feature that most core switch vendors will adopt as it provides reliable network forwarding when the network is the data center fabric supporting both storagte and data gram flows.

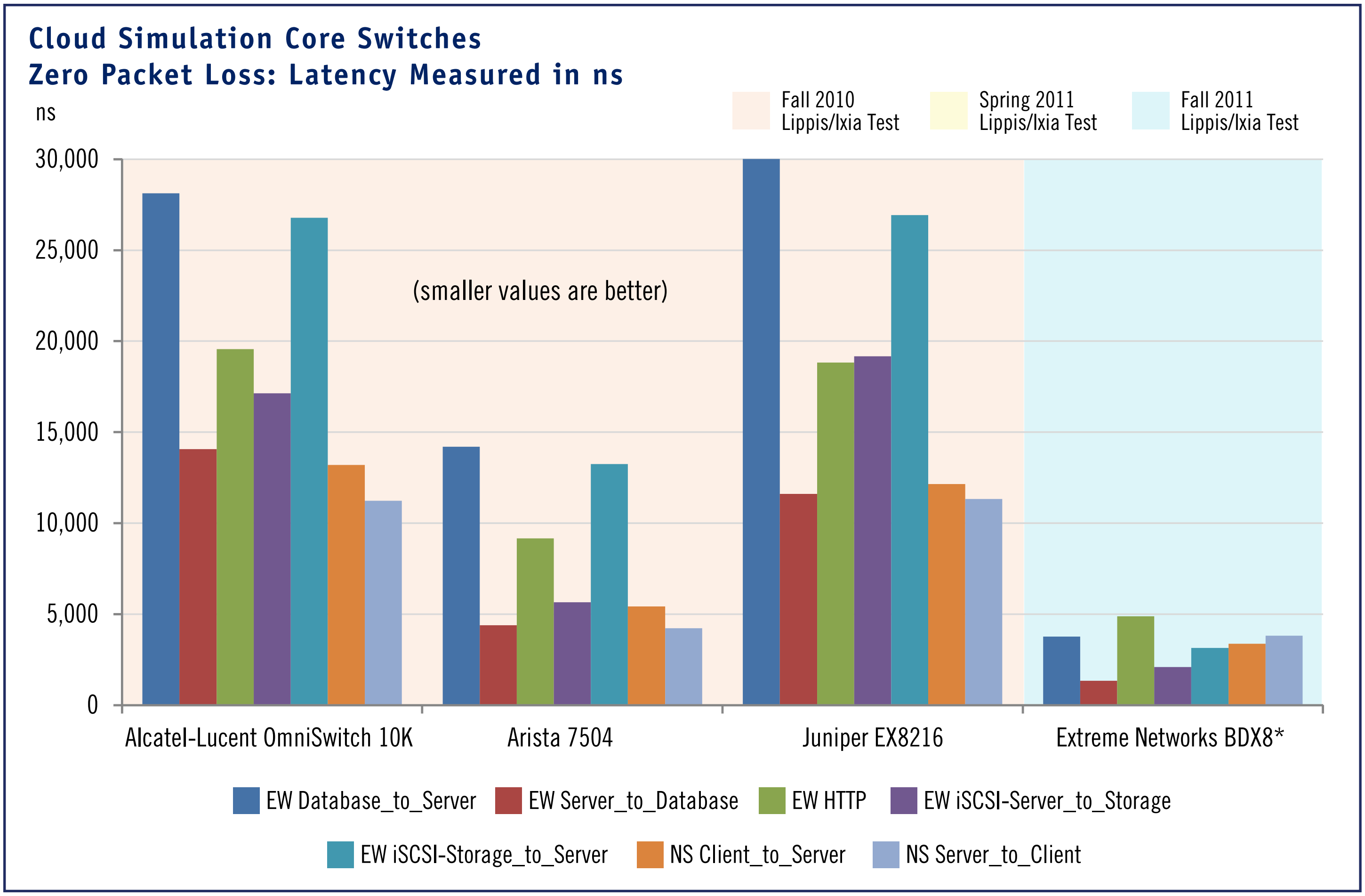
Core Switches RFC 3918 IP Multicast Test



* Arista does not suport IP Multicast at this time.

	Throughput (% Line Rate)				Agg Average Latency (ns)			
Frames/ Packet Size (bytes)	Alcatel-Lucent OmniSwitch 10K	Arista 7504*	Juniper EX8126	Extreme Networks BDX8	Alcatel-Lucent OmniSwitch 10K	Arista 7504*	Juniper EX8126	Extreme Networks BDX8
64	100%	Arista does not support IP Multicast at this time.	100%	100%	9596	Arista does not support IP Multicast at this time.	27210	2452
128	100%		100%	100%	9646		26172	2518
256	100%		100%	100%	9829		25135	2648
512	100%		100%	100%	10264		26600	2904
1,024	100%		100%	100%	11114		28802	3417
1,280	100%		100%	100%	11625		29986	3668
1,518	100%		100%	100%	11921		30746	3915
2,176	100%		100%	100%	13023		33027	4574
9,216	100%		100%	100%	28059		60055	11622

Arista does not support IP Multicast at this time and thus is excluded from this test. The Extreme Networks BDX8 offers the lowest IP Multicast latency at the highest port density tested to date in these industry test. The Extreme Networks BDX8 is the fastest at forwarding IP Multicast thanks to its replication chip design delivering forwarding speeds that are 3-10 times faster than all others. Alcatel-Lucent's OmniSwitch 10K demonstrated 100% throughput at line rate and the shortest aggregated average latency. Juniper's EX8216 demonstrated 100% throughput at line rate and the IP multicast aggregated average latency twice that of Alcatel-Lucent's OmniSwitch.



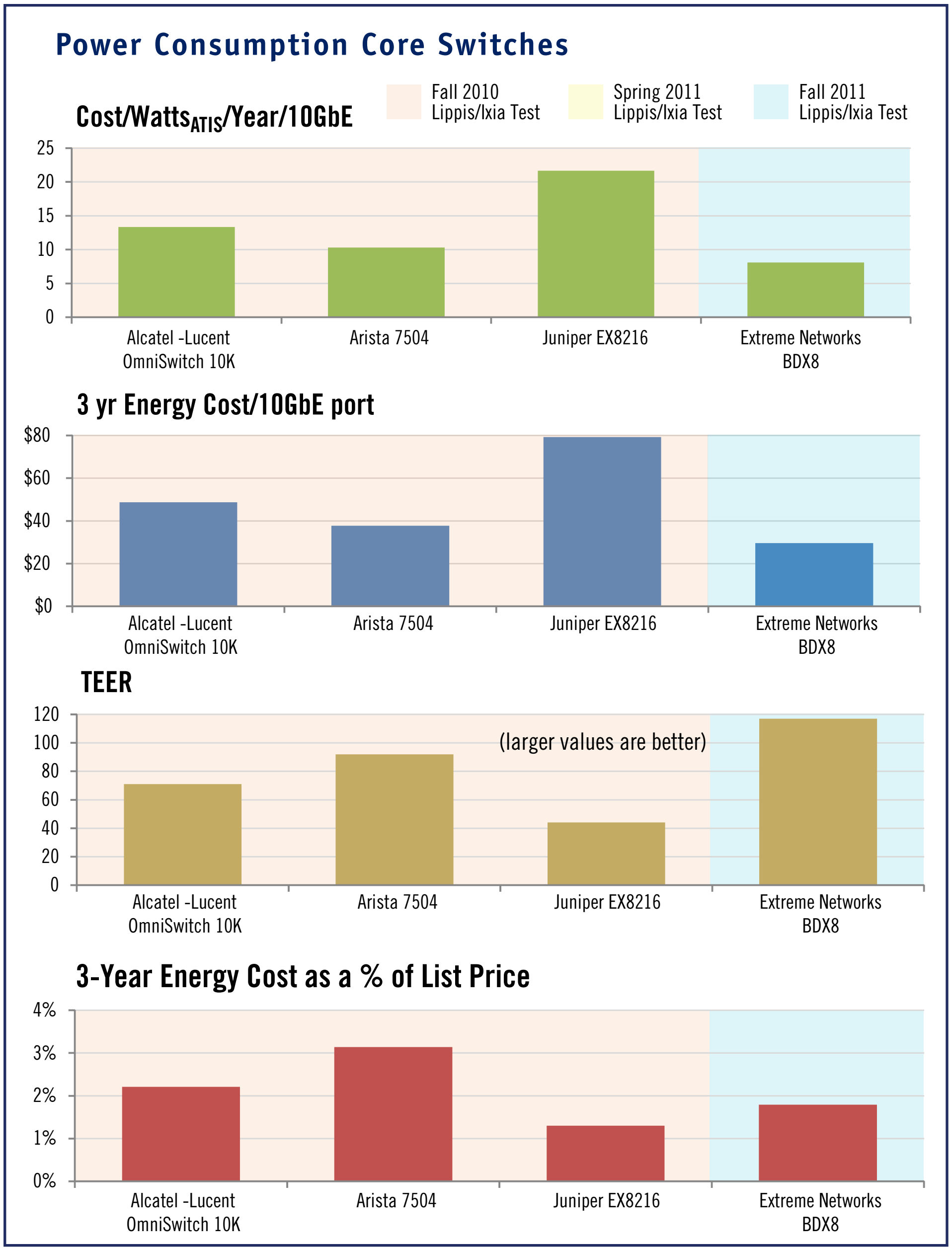
* Extreme Networks BDX8 was measured in Store & Forward plus Cut-Through Mode. For cut-through cloud simulation latency results see the X8 write-up.

Cloud Simulation Core Switches
Zero Packet Loss: Latency Measured in ns

Company Product	EW Database_ to_Server	EW Server_to_ Database	EW HTTP	EW iSCSI-Server_to_ Storage	EW iSCSI-Storage_to_ Server	NS Client_ to_Server	NS Server_ to_Client
Alcatel-Lucent OmniSwitch 10K	28125	14063	19564	17140	26780	13194	11225
Arista 7504	14208	4394	9171	5656	13245	5428	4225
Juniper EX8216	30263	11614	18833	19177	26927	12145	11322
Extreme Networks* BDX8	3773	1339	4887	2097	3135	3368	3808

* Extreme Networks BDX8 was measured in Store & Forward plus Cut-Through Mode. For cut-through cloud simulation latency results see the X8 write-up.

All Core switches delivered 100% throughput with zero packet loss during the cloud simulation test. Alcatel-Lucent's and Juniper's average latencies for various traffic flows were consistent with each other while Arista's latency was measured, at specific traffic types, significantly lower. But the clear stand-out here is the Extreme Networks BDX8 with forwards a range of cloud protocols 4 to 10 times faster than others tested.



There are differences in the power efficiency across these Core switches. The Alcatel-Lucent OmniSwitch 10K, Arista 7540, Extreme X8 and Juniper EX8216 represent a breakthrough in power efficiency for Core switches with previous generation of switches consuming as much as 70 WATTS$_{ATIS}$. These switches consume between 8 and 21 WATTS$_{ATIS}$ with TEER values between 117 and 71; remember the higher the TEER value the better.

The standout here is Extreme Networks BDX8 with the lowest WATTS$_{ATIS}$ and highest TEER values measured thus far for Core swithes in these Lippis/Ixia test. The Extreme Networks BDX8's actual Watts/10GbE port is actually lower; when fully populated with 756 10GbE or 192 40GbE ports. During the Lippis/Ixia test, the BDX8 was only populated to a third of its port capacity but equipped with power supplies, fans, management and switch fabric modules for full port density population. Therefore, when this full power capacity is divided across a fully populated BDX8, its WattsATIS per 10GbE Port will be lower than the measurement observed.

Company Product	Watts$_{ATIS}$/ port	3-Yr Energy Cost per 10GbE Port	TEER	3 yr Energy Cost as a % of List Price	Cooling
Alcatel-Lucent OmniSwitch 10K	13.3	$48.77	71	2.21%	Front-to-Back
Arista 7504	10.3	$37.69	92	3.14%	Front-to-Back
Juniper EX8216	21.7	$79.26	44	1.30%	Side-to-Side*
Extreme Networks BDX8	8.1	$29.60	117	1.79%	Front-to-Back

* 3rd party cabinet options from vendors such as Chatsworth support hot-aisle cold-aisle deployments of up to two EX8216 chassis in a single cabinet.

The three year cost per 10GbE port shows the difference between the three core switches with 3-yr estimated cost per 10GbE ranging from a low of $37.69 to a high of $79.26. In terms of cooling, Juniper's EX8216 is the only Core switch that does not support front-to-back airflow. However, third party cabinet options from vendors such as Chatsworth support hot-aisle cold-aisle deployments of up to two EX8216 chassis in a single cabinet.

Core Switch 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 Core switches may be used in their support for corporate advantage.

10GbE Core Switches Ready for Deployment: 10GbE Core switches are ready for prime time, delivering full line rate throughput at zero packet loss and nanosecond latencies plus single to double-digit delay variation. In addition, these Core switches offer low power consumption with energy cost over a three-year period estimated between 1.3% and 3.14% of acquisition cost.

Evaluate Each Core Switch Separately: While this set of Core switch category tested very well, there are differences between vendors especially in the RFC 2889 congestion, RFC 3918 IP Multicast, power consumption and cloud simulation test. Therefore, it is recommended to review each supplier's results and make purchase decisions accordingly.

Deploy Core Switches that Demonstrate Efficient Power and Cooling: In addition to solid TEER values and low WATTSATIS , all Core switches, except the EX8216, tested support front-to-back or rear-to-front cooling in support of data center hot/cold aisle airflow designs. Juniper does support front to back airflow via the purchase of a 3rd party adapter. Therefore, it is recommended that these Core switches can be used as part of an overall green data center initiative.

Ready for Storage Enablement: Most of the Core switches demonstrated the performance and latency required to support storage enablement or converged I/O. In fact, all suppliers have invested in storage enablement such as support for CEE, FCoE, iSCSI, NAS, etc., and while these features were not tested in the Lippis/Ixia evaluation, these switches demonstrated that the raw capacity is built into the switches for its support. Look for more vendors to support flow control in their core switches.

Evaluate Core Switches for Network Services and Fabric Differences: There are differences between suppliers in terms of the network services they offer as well as how their Core switches connect to ToR/Leaf switches to create a data center fabric. It is recommended that IT business leaders evaluate Core switches with ToR switches to assure that the network services and fabric attributes sought after are realized.

Connect Servers at 10GbE: The Lippis/Ixia test demonstrated the performance and power consumption advantages of 10GbE networking, which can be put to work and exploited for corporate advantage. For new server deployments in private/public data center cloud networks, 10GbE is recommended as the primary network connectivity service as a network fabric exists to take full advantage of server I/O at 10GbE bandwidth and latency levels.

40GbE Uplink Termination Options Available: The Extreme Networks X8 supports 192 40GbE ports, 24 were tested here. The OmniSwitch 10K, Juniper EX8216 and Arista 7504 possess the hardware architecture and scale to support 40GbE and 100GbE modules. While only Extreme Networks was tested for 40GbE, we expect that during the Spring 2012 Lippis Report test at iSimCity more will be aviable. Further, we expect 40GbE adoption will start in earnest during 2012 thanks to 40GbE cost being only 3-4 times that of 10GbE.

Consider a Two-Tier Network Architecture: Core switch port desnsity is on the rise with 10GbE port densities in the 128 to 756 port range plus 40GbE in 192 port density range. We tested Core switches with 256-10GbE and 24-40GbE and found that these products forward L2 and L3 packets at wire speed at nano and μs latency while offering excellent congestion management and are ready to be deployed in a two-tier network architecture made up of Core and ToR switches. This design is preferable for private/public data center cloud computing scale deployments. The OmniSwitch 10K, Arista 7504, Extreme Networks X8 and Juniper EX8216 possess the performance and power efficiency required to be deploy in a two-tier network architecture, reducing application latency plus capital and operational cost from these budgets.

Evaluate Link Aggregation Technologies: Not tested in the Lippis/Ixia evaluation was a Core switch's ability to support MC-LAG (Multi-Chassis Link Aggregation Group), ECMP (Equal-Cost Multi-Path routing , TRILL (Transparent Interconnection of Lots of Links) or SPB (802.1aq (Shortest Path Bridging), Cisco's FastPath, Juniper's QFabric and Brocade's VCS. This is a critical test to determine Ethernet fabric scale. It is recommended that IT business leaders and their network architects evaluate each supplier's MC-LAG, ECMP TRILL, SPB, FabricPath, QFabric, VCS approach, performance, congestion and latency characteristics.

The Lippis Report Test Methodology

To test products, each supplier brought its engineers to configure its equipment for test. An Ixia test engineer was available to assist each supplier through test methodologies and review test data. After testing was concluded, each supplier's engineer signed off on the resulting test data. We call the following set of testing conducted "The Lippis Test." The test methodologies included:

Throughput Performance: Throughput, packet loss and delay for layer-2 (L2) unicast, layer-3 (L3) unicast and layer-3 multicast traffic was measured for packet sizes of 64, 128, 256, 512, 1024, 1280, 1518, 2176, 9216 bytes. In addition, a special cloud computing simulation throughput test consisting of a mix of north-south plus east-west traffic was conducted. Ixia's IxNetwork RFC 2544 Throughput/Latency quick test was used to perform all but the multicast tests. Ixia's IxAutomate RFC 3918 Throughput No Drop Rate test was used for the multicast test.

Latency: Latency was measured for all the above packet sizes plus the special mix of north-south and east-west traffic blend. Two latency tests were conducted: 1) latency was measured as packets flow between two ports on different modules for modular switches, and 2) between far away ports (port pairing) for ToR switches to demonstrate latency consistency across the forwarding engine chip. Latency test port configuration was via port pairing across the entire device versus side-by-side. This meant that a switch with N ports, port 1 was paired with port (N/2)+1, port 2 with port (N/2)+2, etc. Ixia's IxNetwork RFC 2544 Throughput / Latency quick test was used for validation.

Jitter: Jitter statistics was measured during the above throughput and latency test using Ixia's IxNetwork RFC 2544 Throughput/Latency quick test.

Congestion Control Test: Ixia's IxNetwork RFC 2889 Congestion test was used to test both L2 and L3 packets. The objective of the Congestion Control Test is to determine how a Device Under Test (DUT) handles congestion. Does the device implement congestion control and does congestion on one port affect an uncongested port? This procedure determines if Head-of-Line (HOL) blocking and/or if back pressure are present. If there is frame loss at the uncongested port, HOL blocking is present. Therefore, the DUT cannot forward the amount of traffic to the congested port, and as a result, it is also losing frames destined to the uncongested port. If there is no frame loss on the congested port and the port receives more packets than the maximum offered load of 100%, then Back Pressure is present.

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

RFC 2544 Throughput/Latency Test

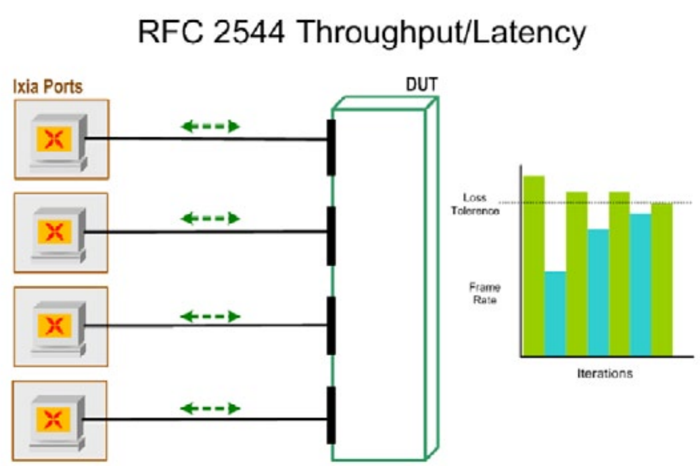
Test Objective: This test determines the processing overhead of the DUT required to forward frames and the maximum rate of receiving and forwarding frames without frame loss.

Test Methodology: The test starts by sending frames at a specified rate, usually the maximum theoretical rate of the port while frame loss is monitored. Frames are sent from and received at all ports on the DUT, and the transmission and reception rates are recorded. A binary, step or combo search algorithm is used to identify the maximum rate at which no frame loss is experienced.

To determine latency, frames are transmitted for a fixed duration. Frames are tagged once in each second and during half of the transmission duration, then tagged frames are transmitted. The receiving and transmitting timestamp on the tagged frames are compared. The difference between the two timestamps is the latency. The test uses a one-to-one traffic mapping. For store and forward DUT switches latency is defined in RFC 1242 as 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. Thus latency is not dependent on link speed only, but processing time too.

Results: This test captures the following data: total number of frames transmitted from all ports, total number of frames received on all ports, percentage of lost frames for each frame size plus latency, jitter, sequence errors and data integrity error.

The following graphic depicts the RFC 2554 throughput performance and latency test conducted at the iSimCity lab for each product.



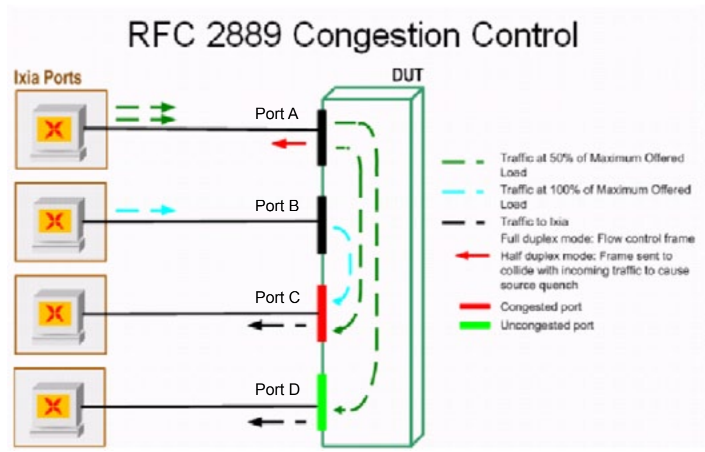
RFC 2889 Congestion Control Test

Test Objective: The objective of the Congestion Control Test is to determine how a DUT handles congestion. Does the device implement congestion control and does congestion on one port affect an uncongested port? This procedure determines if HOL blocking and/or if back pressure are present. If there is frame loss at the uncongested port, HOL blocking is present. If the DUT cannot forward the amount of traffic to the congested port, and as a result, it is also losing frames destined to the uncongested port. If there is no frame loss on the congested port and the port receives more packets than the maximum offered load of 100%, then Back Pressure is present.

Test Methodology: If the ports are set to half duplex, collisions should be detected on the transmitting interfaces. If the ports are set to full duplex and flow control is enabled, flow control frames should be detected. This test consists of a multiple of four ports with the same MOL (Maximum Offered Load). The custom port group mapping is formed of two ports, A and B, transmitting to a third port C (the congested interface), while port A also transmits to port D (uncongested interface).

Test Results: This test captures the following data: intended load, offered load, number of transmitted frames, number of received frames, frame loss, number of collisions and number of flow control frames obtained for each frame size of each trial are captured and calculated.

The following graphic depicts the RFC 2889 Congestion Control test as conducted at the iSimCity lab for each product.



RFC 3918 IP Multicast Throughput No Drop Rate Test

Test Objective: This test determines the maximum throughput the DUT can support while receiving and transmitting multicast traffic. The input includes protocol parameters (IGMP, PIM), receiver parameters (group addressing), source parameters (emulated PIM routers), frame sizes, initial line rate and search type.

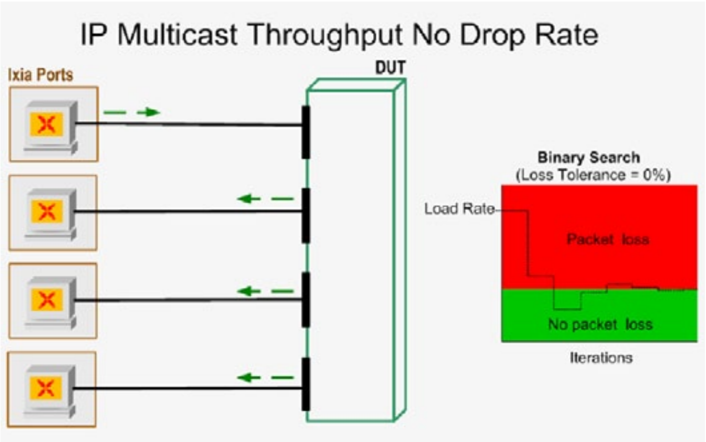
Test Methodology: This test calculates the maximum DUT throughput for IP Multicast traffic using either a binary or a linear search, and to collect Latency and Data

Integrity statistics. The test is patterned after the ATSS Throughput test; however this test uses multicast traffic. A one-to-many traffic mapping is used, with a minimum of two ports required.

If choosing OSPF or ISIS as IGP protocol routing, the transmit port first establishes an IGP routing protocol session and PIM session with the DUT. IGMP joins are then established for each group, on each receive port. Once protocol sessions are established, traffic begins to transmit into the DUT and a binary or linear search for maximum throughput begins.

If choosing "none" as IGP protocol routing, the transmit port does not emulate routers and does not export routes to virtual sources. The source addresses are the IP addresses configured on the Tx ports in data frame. Once the routes are configured, traffic begins to transmit into the DUT and a binary or linear search for maximum throughput begins.

Test Results: This test captures the following data: maximum throughput per port, frame loss per multicast group, minimum/maximum/average latency per multicast group and data errors per port.The following graphic depicts the RFC 3918 IP Multicast Throughput No Drop Rate test as conducted at the iSimCity lab for each product.



Power Consumption Test

Port Power Consumption: Ixia's IxGreen within the IxAutomate test suite was used to test power consumption at the port level under various loads or line rates.

Test Objective: This test determines the Energy Consumption Ratio (ECR), the ATIS (Alliance for Telecommunications Industry Solutions) Telecommunications Energy Efficiency Ratio (TEER) during a L2/L3 forwarding performance. TEER is a measure of network-element efficiency quantifying a network component's ratio of "work performed" to energy consumed.

Test Methodology: This test performs a calibration test to determine the no loss throughput of the DUT. Once the maximum throughput is determined, the test runs in automatic or manual mode to determine the L2/L3 forwarding performance while concurrently making power, current and voltage readings from the power device. Upon completion of the test, the data plane performance and Green (ECR and TEER) measurements are calculated. Engineers followed the methodology prescribed by two ATIS standards documents:

ATIS-0600015.03.2009: Energy Efficiency for Telecommunication Equipment: Methodology for Measuring and Reporting for Router and Ethernet Switch Products, and

ATIS-0600015.2009: Energy Efficiency for Telecommunication Equipment: Methodology for Measuring and Reporting - General Requirements

The power consumption of each product was measured at various load points: idle 0%, 30% and 100%. The final power consumption was reported as a weighted average calculated using the formula:

WATIS = 0.1*(Power draw at 0% load) + 0.8*(Power draw at 30% load) + 0.1*(Power draw at 100% load).

All measurements were taken over a period of 60 seconds at each load level, and repeated three times to ensure result repeatability. The final WATIS results were reported as a weighted average divided by the total number of ports per switch to derive at a WATTS per port measured per ATIS methodology and labeled here as $\text{WATTS}_{\text{ATIS.}}$

Test Results: The L2/L3 performance results include a measurement of WATIS and the DUT TEER value. Note that a larger TEER value is better as it represents more work done at less energy consumption. In the graphics throughout this report, we use $WATTS_{ATIS}$ to identify ATIS power consumption measurement on a per port basis.

With the $WATTS_{ATIS}$ we calculate a three-year energy cost based upon the following formula.

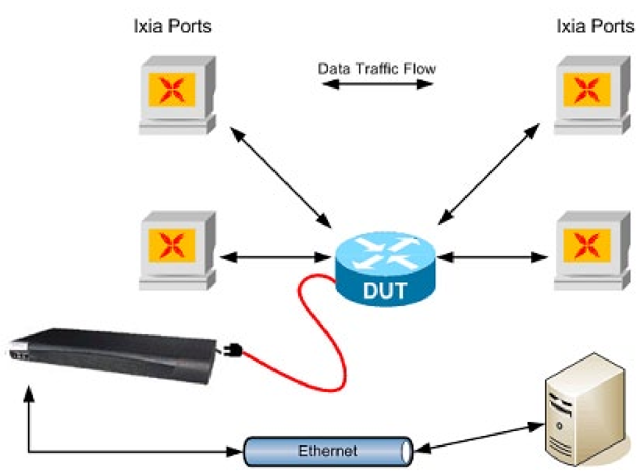
$Cost/Watts_{ATIS}/3\text{-}Year =$
$(WATTS_{ATIS}/1000)*(3*365*24)*(0.1046)*(1.33)$,
where $WATTS_{ATIS}$ = ATIS weighted average power in Watts

3*365*24 = 3 years @ 365 days/yr @ 24 hrs/day

0.1046 = U.S. average retail cost (in US$) of commercial grade power as of June 2010 as per Dept. of Energy Electric Power Monthly

(http://www.eia.doe.gov/cneaf/electricity/epm/table5_6_a.html)
1.33 = Factor to account for power costs plus cooling costs @ 33% of power costs.

The following graphic depicts the per port power consumption test as conducted at the iSimCity lab for each product.



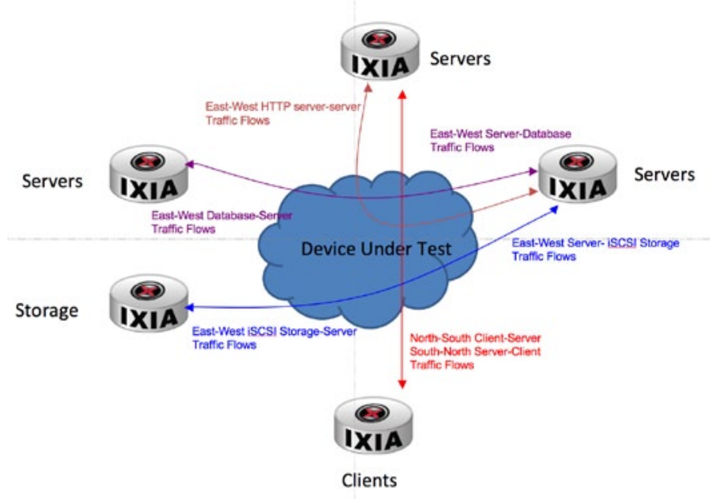
Public Cloud Simulation Test

Test Objective: This test determines the traffic delivery performance of the DUT in forwarding a variety of north-south and east-west traffic in cloud computing applications. The input parameters include traffic types, traffic rate, frame sizes, offered traffic behavior and traffic mesh.

Test Methodology: This test measures the throughput, latency, jitter and loss on a per application traffic type basis across M sets of 8-port topologies. M is an integer and is proportional to the number of ports the DUT is populated with. This test includes a mix of north-south traffic and east-west traffic, and each traffic type is configured for the following parameters: frame rate, frame size distribution, offered traffic load and traffic mesh. The following traffic types are used: web (HTTP), database-server, server-database, iSCSI storage-server, iSCSI server-storage, client-server plus server-client. The north-south client-server traffic simulates Internet browsing, the database traffic simulates server-server lookup and data retrieval, while the storage traffic simulates IP-based storage requests and retrieval. When all traffic is transmitted, the throughput, latency, jitter and loss performance are measured on a per traffic type basis.

Test Results: This test captures the following data: maximum throughput per traffic type, frame loss per traffic type, minimum/maximum/average latency per traffic type, minimum/maximum/average jitter per traffic type, data integrity errors per port and CRC errors per port. For this report we show average latency on a per traffic basis at zero frame loss.

The following graphic depicts the Cloud Simulation test as conducted at the iSimCity lab for each product.



40GbE Testing: For the test plan above, 24-40GbE ports were available for those DUT that support 40GbE uplinks or modules. During this Lippis/Ixia test 40GbE testing included latency, throughput congestion, IP Multicast, cloud simulation and power consumption test. ToR switches with 4-40GbE uplinks are supported in this test.

Virtualization Scale Calculation

We observe technical specifications of MAC address size, /32 IP host route table size and ARP entry size to calculate the largest number of VMs supported in a layer 2 domain. Each VM will consume a switch's logical resources of, especially a modular/core switch, (1) MAC entry, (1) /32 IP host route, and (1) ARP entry. Further, each physical server may use (2) MAC/IP/ARP entries, one for management, one for IP storage and/or other uses. Therefore, the lowest common denominator of the three (MAC/IP/ARP) entries will determine the total number of VM and physical machines that can reside within a layer 2 domain.

If a data center switch has a 128K MAC table size, 32K IP host routes, and 8K ARP table, then it could support 8,000 VM (Virtual Machines) + Physical (Phy) servers. From here we utilize a VM:Phy consolidation ratio to determine the approximate maximum # of physical servers. 30:1 is a typical consolidation ratio today but more dense (12) core processors entering the market may increase this ratio to 60:1. With 8K total IP endpoints at a 30:1 VM:Phy ratio, this calculates to an approximate 250 Phy servers, and 7,750 VMs.

250 physical servers can be networked into approximately 12 48-port ToR switches, assuming each server connects to (2) ToR swiches for redundancy. If each ToR has 8x10G links to each core switch, that's 96 10GbE ports consumed on the core switch. So we can begin to see how the table size scalability of the switch determines how many of it's 10GbE ports, for example, can be use and therefore how large a cluster can be built. This calculation assumes the core is the Layer 2/3 boundary, providing a layer 2 infrastructure for the 7,750 VMs.

More generally, for a Layer 2 switch positioned at ToR for example, the specification that matters for virtualization scale is the MAC table size. So if an IT leader purchases a layer 2 ToR switch, say switch A, and its MAC table size is 32K, then switch A can be positioned into a layer 2 domain supporting up to ~32K VMs. For a Layer 3 core switch (aggregating ToRs), the specifications that matter are MAC table size, ARP table size, IP host route table size. The smaller of the three tables is the important and limiting virtualization scale number. If an IT leader purchases a layer 3 core switch, say switch B and its IP host route table size is 8K, and MAC table size is 128K, then this switch can support a VM scale of ~8K. If an IT architect deploys the Layer 2 ToR switch A supporting 32K MAC addresses and connects it to a core switch B, then the entire layer 2 virtualization domain scales to the lowest common denominator in the network, the 8K from switch B.

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, available 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 Enterprises, Inc. is not responsible for, and you agree to hold Lippis Enterprises, Inc. 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 Enterprises, Inc. 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 www.lippisreport.com .

No part of any document may be reproduced, in whole or in part, without the specific written permission of Lippis Enterprises, Inc. 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 160,000 times; i-Tunes 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. Mr. Lippis is currently working with clients to design their private and public virtualized data center cloud computing network architectures 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, Cigitel, 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 firm's 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.