Lippis Report

Research Note

Lippis Report 208:

A Robust Underlay to Support Hyper Growth and Dynamic Overlay

By Nicholas John Lippis III
President, Lippis Consulting

June, 2013

At the Open Networking Users Group or ONUG this past February in Boston at Fidelity Investments, Chief Network Architects (CNAs) from the largest corporations discussed their Software-Defined Networking or SDN plans, obstacles, requirements and experiences. Two key systemic messages from this meeting were that network opex has to fundamentally change and that business unit managers are demanding self-service IT provisioning. These two requirements provide CNAs permission to break away from 25 years of networking to implement a new architecture that will meet these needs. One large bank showed a graphic plotting the rate of adoption of business unit self-service virtual machine adds, pauses and deletions. The curve was exponential and so, too, was the rise of its overlay virtualized network. And it's just the beginning as virtualized network overlay adoption is in its infancy, considering only 48% of servers are virtualized and a much smaller number of virtual machines are networked. The questions many IT business leaders are asking are: "Can my underlay or physical network infrastructure support the huge growth of my overlay?" "What changes are required for my data center network infrastructure?" "What capabilities should I be looking for?" In this Lippis Report Research Note, we explore the role of network underlays in their support of virtualized network overlays.

Overlays: Layer 2 and 3

With the advent of Open Networking and the large deployment of virtualized servers, virtualized networking has taken center stage as the newest overlay. Just to review, there are two basic overlays: layer 2 or tunneling through a layer 2/3 underlay, and simply a layer 3 overlay. In layer 2 overlays, packets are encapsulated and tunneled through the underlay. For virtualized networks, the tunnel can be VxLAN, NVGRE, STT, GRE, etc. These overlays emulate a LAN segment and provide mobility of VMs within a single subnet. Layer 2 overlays are useful as they are independent of network state and can emulate various physical topologies. On the down side, they don't scale well as it's difficult to penetrate the encapsulation mechanism to manage/view flows plus they are exposed to layer 2 flooding. Layer 3 overlays, however, transport IP packets directly without encapsulation, thus extend mobility beyond layer 3 boundaries and are useful in abstracting connectivity and policy. However, the fact is layer 2 overlays dominate by far, thanks to automated provisioning.

This exponential growth in virtualized networks creates traffic patterns that are unpredictable, and worse of all, there is a lack of visibility, thus adding network management complexity, thanks to tunneled traffic. If one could see overlay traffic patterns, it would most likely look like spaghetti. So, what is the best approach to support this unstoppable growth of nearly invisible traffic that is as predictable as Brownian motion? A robust underlay!

Underlay Progress

In advanced data center networks, underlays or physical networks have been increasingly upgraded to support larger and larger traffic volume. Thus data center, and especially cloud infrastructure, has transitioned from 1GbE to 10GbE at volume with 40 and 100GbE starting their own adoption curves. All of this is occurring as 400GbE is now under discussion in the IEEE. In addition to higher speed, increased reliability or availability has been engineered into data center network design, thanks to active-active multi-pathing protocols such as M-LAG, Virtual Port-Channel, Cisco's FabricPath and emerging TRILL as well as SPB. These protocols replace the slower, inefficient and redundant Spanning Tree Protocol, allowing two-tier networking in the data center to occur, which reduces latency and speeds up application performance. To reduce cabling cost, streamline the number of devices needed in the data center plus reduce management points, the technologies to deploy a unified or converged data center infrastructure have emerged. Technologies such as Fiber Channel over Ethernet or FCoE, ATA over Ethernet, iSCSI over Ethernet, as well as Converged Network Adaptors are enabling a single physical Ethernet fabric that supports datagram and storage traffic.

Modular Switching

The two main networking devices that make up a modern data center network are Top of Rack or ToR and Core switches. ToRs are fixed configuration switches while the Core is modular switching. Modular Core switching is the staple for connecting ToRs, Internet and Campus LANs and will be for some time to come. To see how modular switching will keep pace with overlays, we analyze the Cisco Nexus 7000 as it represents the lion's share of the data center modular switch market being in the market for over five years.

Key Benefits of Modular Switching

High Availability via Dual Supervisors: Network high availability or HA is critical for continuous data center operation. A scalable network infrastructure for virtualization built with overlay networking relies upon a HA capable underlying network to provide connectivity and reachability. Cisco's Nexus 7000 goes beyond traditional HA approaches such as bandwidth increases, active-active protocol support, nx1 power supplies, etc. In fact, Cisco has invested in a range of HA features to assure a stable and reliable underlay. For example, dual supervisor engines or the ability for a Nexus 7000 to support active and standby supervisors simultaneously mitigate packet loss or service disruption in the case of a supervisor failure. Dual supervisors also enable in-service upgrades providing non-stop network operations. This level of HA, reliability and serviceability offers business value in the form of reduced downtime and ease of new technology adoption, which is so important with the large volume of overlay traffic flows.

Fundamental Network Design Technologies: As data center network requirements have evolved, so, too, have modular switches. The modular nature of chassis-based systems enables vendors to introduce new modules and software to support a changing data center landscape. For example, the Cisco Nexus 7000 addressed the requirement of geographically distributed data centers at layer 2 by introducing OTV (or Overlay Transport Virtualization). To enable workload mobility between layer 3 subnets, Cisco introduced the LISP overlay or Locator/ID Separation Protocol. To create a flatter, two-tier, non-blocking, fully meshed data center network, Cisco introduced multipath connectivity options with MLAG and FabricPath. To connect distributed data centers over the wide area MPLS and VPLS, functionality was added to the Nexus 7000. Many of these technologies are, in fact, network-based overlays, as is the emergence of virtual switching/hypervisor-based overlay networks. Therefore, based upon prior innovations, it's pretty clear that physical switches will evolve to provide intelligence and visibility for tunneled overlay traffic. IT executives deploying modular switches will expect and demand that new hardware modules support emerging overlay technologies without the need to do a rip and replace.

Data Center Fabric designs are also evolving toward a "scale-out" spine-leaf design architecture also known as a Clos network, named after Charles Clos who formalized the approach in 1953. These spine-leaf network designs are especially prevalent in large data centers. Spine-leaf architecture enables large cloud network scale plus flexibility to place workloads and services anywhere in the network. System HA is a fundamental critical attribute of the overall design process. A highly available redundant switch can add significant value to the design by eliminating reliance on protocol level convergence, for example. Switches with dual supervisors and ISSU or In-Service Software Upgrade capabilities provide a scalable spine network, which forwards traffic even during maintenance windows.

Lots of Head Room: Application trends, such as Big Data analytics, VDI deployments, BYOD, increased use of video, etc., are driving bandwidth growth. As bandwidth demand grows, so, too, does the need for workload mobility as virtualized workloads and application mobility requirements skyrocket. User experience needs to be consistent, irrespective of workload physical location. Therefore, ample east-west bandwidth is required to support workload mobility. However, increasing data center bandwidth should be simple and seamless.

Anticipating increased data center switching capacity, Cisco's Nexus 7000 engineers developed a second-generation Nexus 7000 fabric module, the Supervisor 2 and 2E introduced in 2012, which increased bandwidth per slot from 230Gbs to 550Gbs. One terabit per second (1Tbps) per slot should not be too far off into the future as the fabric architecture scales to 17.6 Tbps plus. Both 40 and 100GbE were made available in 2012 with the introduction of the M2-Series modules. These new modules can be deployed while the Nexus 7000 is in-service, minimizing service disruption and easing system upgrades.

Open Networking: SDN support is a growing industry requirement to support programmability of application infrastructure. While many vendors have introduced new hardware to support SDN functionality, Cisco will add SDN support to the Nexus 7000 as part of the Cisco ONE initiative. The Nexus 7000 will expose network state to programmers via Cisco's onePK API, and its forwarding tables will be accessible via OpenFlow support, thanks to Cisco ONE. Further, to support virtualized overlay tunneled traffic in a modular system, IT executives expect hardware support for VxLAN to be available and to provide VLAN-to-VxLAN gateways plus multi-tenancy.

Unified Networking: Network convergence of storage and LANs is occurring in highly virtualized environments with Cisco's Unified Computing System. The same converged I/O functionality can be extended in the Core today with the Nexus 7000. Cisco has introduced modules, software and design guides to support director-class FCoE, allowing the Nexus 7000 series to connect Core Fiber Channel capacity to the FCoE access layer. This places the Nexus 7000 at the Core of the network for both storage transport and IP traffic flows.

Innovation with Investment Protection: Investment protection is a key attribute of modular system design, and as such, modular switches are engineered for a long lifecycle. IT executives deploy Core-switching infrastructure under the assumption that it will be operational for seven to 10 years, thanks to incremental hardware and software upgrades. These upgrades tend to add new innovations to support business requirements. The Nexus 7000 has been an innovation engine for Cisco. OTV, LISP, FabricPath, FCoE, et al., first appeared on the Nexus 7000. Also, consider that the Nexus 7000 is available in four chassis sizes, including 4, 9, 10 and 18 slots, providing IT business leaders network design options based on port density that span from small- to massively scaled data centers. The current Nexus 7000 series supports 550Gbps of bandwidth per slot equipped with the 2nd generation fabric module; this is more than twice the slot bandwidth of the first generation fabric of 230Gbps per slot. That is, the same chassis, same backplane wiring, same power supply, same footprint, etc., now scales to 550Gbps per slot. Forward and backward module compatibility is very important to maintain a long product lifecycle, which Cisco has incorporated into its Nexus and Catalyst designs. Future bandwidth growth in the Nexus 7000 is possible with new fabrics and modules.

Considering that supervisor engines enjoy a 10-year-or-so life span and the Supervisor 2 and 2E was just released, it's highly probable that this supervisor will be in production into 2022. During that time, it's also highly probable that a range of supervisors will be available that are forward and backward compatible within the same chassis as well as modules such as the F1, F2, M2 and future additions. Based upon this, the Nexus 7000 should enjoy a life span well in to the late 2020s, if not the early 2030s. That's investment projection!

Looking Forward

If the Nexus 7000 is to be in projection for that stretch of time, Cisco will need to continue its innovation investment cycle, which it seems willing and ready to do. The upgrade cycles have been similar to that of the Catalyst 6500 in that engineers tap customers regularly to understand what features need to be added to address major changes in networking. The best practice of backward- and forward-compatible upgrades, be it modules, service engines, service modules or NXOS software, is part of the Nexus 7000's engineering team's organizing principle. All of this is possible, thanks to the Nexus 7000 being a modular switch.

Modular switches such as the Nexus 7000 have the advantage of large system bandwidth that affords different paces for module, supervisor and service module development. Also, as the NXOS is a modular and compartmentalized operating system, it can be modified for Nexus products. For example, as new software features that are common across all Nexus products are available, one single NXOS image upgraded to all Nexus switches will thus share a common feature implementation.

So what's in store for the Nexus 7000 over this next business cycle that will enable it to support hyper growth and a dynamic overlay? To provide visibility of overlay traffic, flexible NetFlow and Network Analysis Model or NAM service module provides deep visibility into application and user traffic. Both flexible NetFlow and NAM increase troubleshooting coverage to support tunneled overlay traffic. It's inevitable that an increase in bandwidth per slot will be made available to support increased flow of traffic plus higher speed 40 and 100GbE ports. Cisco contributed to the development of VxLAN, MPLS/VPLS, TRILL and FCoE as well as developed OTV and LISP. Expect modules that support these network design options at speed. It would not be a shock either to see the Nexus 7000 engineers develop and offer service modules for security, NAM and L4-7 services.

Nexus 7000: The Robust Underlay for Burgeoning Overlays

The Nexus 7000 was launched in 2008 as a 10GbE Core modular switch. It has evolved from a 10GbE optimized switching platform to a flexible 40/100GbE data center Core routing and switching platform. This platform's features and capabilities have grown as market demands changed to now being the Modular Core Switching platform for not only HA forwarding of layer 2 and 3 traffic, but the enabler of converged I/O and Cisco's Open Networking Environment. Much like Cisco's Catalyst switch, Cisco understands how to manage a platform by innovating and investing into new features that add value, be it cost reduction or service enablement. As virtualized networks and other overlays increase in deployment, the Nexus 7000 is now being focused on providing a stable, reliable, highly available and robust network infrastructure underlay to meet the dynamic market needs now and for many years to come.

About Nick Lippis

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

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

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

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