# What's the Best Alternative to the Spanning Tree Protocol?

**Jim Metzler, Ashton, Metzler & Associates**

- **The Data Center LAN Evolution Series**
- **A Webtorials Thought Leadership Discussion**
- **Dr. Jim Metzler, Moderator**
- **Featuring Arista, Avaya, Brocade, Cisco Systems, Extreme Networks and HP**

This is a summary of an online discussion that took place during August 2012. The discussion was about the best alternative to the spanning tree protocol (STP) and involved Avaya, Brocade, Cisco, Extreme and HP. A full copy of the discussion can be found at http://www.webtorials.com/content/tls.html

One of the most obvious conclusions that can be drawn from the discussion is that the vendors that took part in the discussion have widely varying views relative to the best alternative to STP. Extreme, for example, was a solid proponent of Multi System Link Aggregation (MLAG). Part of their argument is that MLAG interoperates with the existing infrastructure and preserves the existing investment that IT organizations have made in tools for fault management and troubleshooting. Extreme also stated that MLAG is simpler and better understood than alternatives such as Shortest Path Bridging (SPB) and TRILL (Transparent Interconnection of Lots of Links).

HP advocated what HP refers to as IRF (Intelligent Resilient Framework), which is conceptually similar to MLAG. HP added that they are supporting both SPB and TRILL and expect to leverage IRF to extend the scalability and reliability of TRILL and SPB.

Avaya advocated what they refer to as switch clustering (SC), which has some similarities to MLAG. They also advocated SPB and stated that SPB provides a highly reliable, highly scalable multi-path network where services are provisioned only at the edge. Avaya also stated that SPB can be used in conjunction with SC to provide dual/multi-homing at the edge of the SPB fabric.

Cisco was the only participant in the discussion that advocated the possibility of keeping STP in place. What they stated was that some companies might want to use Virtual Port Channels (vPC), which keep STP in place but eliminate its shortcomings. Cisco also recommended that customers who have more demanding requirements, such as the need for a flattened L2 architecture, should consider either TRILL or FabricPath, a Cisco technology that brings routing concepts to Layer 2. Cisco pointed out that it believes that the goal of having SPB be backward hardware compatible will limit the evolution of that protocol.

Similar to Cisco, Brocade stated that the Brocade Ethernet Fabric using VCS (Virtual Cluster Switching) provides Layer 3-type intelligence at Layer 2 and eliminates the need for STP. Brocade also advocated TRILL as a replacement for STP because of the ability of TRILL to support multi-path networking, the resiliency it provides and its ability to reduce the complexity that is associated with subnets. Brocade also discussed technologies such as Data Center Bridging (DCB) in part for the ability of these technologies to provide a foundation for myriad types of storage traffic.

**Jim Metzler, Ashton, Metzler & Associates**

Enterprise applications are increasingly driving east-west traffic within the data center and many IT organizations are concerned with the limitations of the spanning tree protocol. ***What is the best technology or technologies, that are either currently available, or are likely to be available within the next 18 months, to replace the spanning tree protocol?***

The best technology to replace spanning tree in the data center is Multi System Link Aggregation (MLAG). MLAG works by extending the link level redundancy and load sharing mechanism of Link Aggregation (LAG), to support device and network level redundancy, active-active load sharing for full utilization of network bandwidth, and fast convergence. (See **whitepaper here**.)

Devices on the other end of the MLAG use traditional Link Aggregation to talk to the MLAG peers and as such MLAG offers interoperability with existing servers, network switches, blade switches, and other network devices such as firewalls, routers, IPS/IDS, all of which are part of the data center network ecosystem. MLAG builds on the concepts of traditional link aggregation to achieve this, without requiring any new hardware encapsulation or infrastructure refresh, unlike newer protocols such as TRILL. This allows MLAG to provides an easy and cost-effective migration path, while also preserving investment in existing tools for fault management and troubleshooting.

All of this makes MLAG a simple, cost-effective, and scalable technology to replace spanning tree in the data center.

**Jim Metzler, Ashton, Metzler & Associates**

On a going forward basis, do you see a role for TRILL and/or SPB? If so, which of them do you think is the better technology? Why is that?

It depends on the problem you are trying to solve. If what you want is active-active redundancy and fast failover in the data center, MLAG provides a much simpler, well understood and widely available solution, today, as compared to either TRILL or SPB.

Comparing TRILL and SPB, SPB has more of a service provider lineage and hence has more of a “provisioning” approach for example to multipath forwarding, multicast forwarding, etc. Additionally SPB uses MAC-in-MAC encapsulation again more reminiscent of service providers, which works with existing OAM. TRILL on the other hand has more of an enterprise type evolution path. For example it provides a hop by hop approach to multipath forwarding. TRILL also has a TTL (Time To Live)

counter which provides a measure of comfort for those who have lived through buggy routing implementations in the past, which SPB lacks. On the other hand TRILL introduces a new packet header for which existing OAM does not work. And it inherits the 4k VLAN limitation which SPB allows you to overcome. So really both TRILL and SPB have their own strengths and weaknesses reflecting their evolution lineage and target market. For the data center network, however, if one were to extract the goodness from both TRILL and SPB, I think you would come down quickly to the realization that MLAG does just fine for most data center network implementations.

The technologies used to create L2 domains and control traffic at L2 are well established and have been around for many years. Spanning Tree Protocol (STP) and its subsequent enhancements and extensions have met the majority of our networking needs, but have significant drawbacks when applied to the modern virtualized data center.

IRF was the first network virtualization technology in the industry and enabled customers to remove STP and other cumbersome and error prone loop avoidance technologies (VRRP) at every layer of the network. Its ability to deliver multipathing and rapid 50ms convergence enables customers to massively simplify their network topologies and eliminate significant downtime associated with STP config errors and slow reconvergence times. In addition IRF is complimentary to new Open Standard initiatives such as TRILL and will provide additional scalability and functionality.

Building on **HP's current platform/network virtualization capabilities**, HP is taking active roles in both the IEEE and the IETF efforts to standardize new Layer 2 intra and inter-data center connectivity technologies, respectively 802.1aq SPB and TRILL. These technologies will allow customers to build even larger-scale Layer 2 networks and enable multi-site extension using industry standards as the foundation.

**Jim Metzler, Ashton, Metzler & Associates**

Which to you think is a better technology TRILL or SPB and why is that? If the answer is that it depends on something (e.g., current environment, long term goals) kindly elaborate.

- HP is committed to supporting standards and will adopt those as they make the most sense for our customers.
- HP is actively working on TRILL based solutions for the traditional Enterprise customer segment, including open standards participation with the IETF. HP will start to release TRILL compliant products in 2H 2011 as part of its Comware OS.

- PBB is included in HP's Comware OS today with hardware support of the PBB packet format in both 12500 and 9500 DC switches. HP is continuing its investment in PBB by evolving to current standards such as SPB that collectively provides scalability between edge and core, multi-tenancy services, resiliency, and multi-pathing. While these standards started life as solutions for Service provider and carrier customers, they are also becoming more relevant to large scale data center enterprise environments.
- In addition, HP is leveraging its unique network virtualization and clustering technology, IRF (Intelligent Resilient Framework), to enable Enterprises to extend the scalability and reliability of TRILL and SPB without affecting the interoperability desired from standards based solutions. By virtualizing multiple physical IS-IS nodes into a single logical node, IRF enables a Data Center fabric to scale more significantly while keep the hop count low for greater efficiency and faster convergence.

AVAYA The best technologies to avoid drawbacks and pitfalls associated with STP are Avaya's Switch Clustering (SC) and the IEEE's **Shortest Path Bridging (SPB)**.

SC is a mature and proven technology, pioneered by Avaya, which virtualizes the network core, offering consistency, sub-second recovery from failures, together with utilization of all links and resources.

SPB provides an open and standards-based solution for larger deployment scenarios, providing a highly reliable, highly scalable multi-path network where services are provisioned only at the edge. A dynamic link state protocol provides the optimal path to any destination, creating a fully distributed and dynamically maintained fault-tolerant topology. SPB is based on a standardized data plane, offers comprehensive OA&M, and supports efficient Multicast distribution. It also enables secure traffic separation through the creation of Virtual Service Networks (VSNs), empowering long-distance workload mobility, and applications such as storage, voice, and video to be managed individually. Through comprehensive L3 extensions, enterprise-friendly functionality is also offered. Crucially, SPB can be used in combination with SC, providing dual/multi-homing at the edge of the SPB fabric.

Avaya's innovative Fabric Interconnect Stack delivers a virtual backplane of multiple Terabits capacity, combined with ultra-low latency, specifically addressing the east/west traffic demand in Top-of-Rack deployment scenarios.

 **Jim Metzler, Ashton, Metzler & Associates**

Will you also support TRILL once it has been standardized? What do you see as the major weaknesses of TRILL?

Put simply, TRILL offers auto-topology with loop-free multi-pathing; great. However, it doesn't offer any service abstraction or orchestration; undoubtedly better than STP, but TRILL only addresses part of the problem. On the other hand, Shortest Path Bridging delivers both crucial elements and therefore it's the genuine solution for next-generation private cloud infrastructures. Combining the promise of standardized OA&M with proven interoperability – and multi-vendor testing has already been successfully undertaken – Shortest Path Bridging simplifies the network to empower consolidation and virtualization.

Having said that, Avaya is in the unique position of having the agility to implement complex technology without impacting our customers with an expensive fork-lift upgrade or the inefficiency of a performance hit. Our modular products feature innovative re-programmable network processors that enable us to evolve with technology trends, delivering continued high-performance support for the likes of IPv6, MPLS, and most recently, Shortest Path Bridging. Avaya's premium fixed-format products feature the most sophisticated merchant silicon available, specifically chosen for its unmatched flexibility in handling complex data formats. These product lines definitely could support TRILL, or another technology for that matter, should the mainstream demand make that good business sense.

The answer to this question is really rooted in what customers are ultimately trying to achieve in terms of new functionality or mitigation of pain points. We see that many of our customers want to simply eliminate the blocked port that Spanning Tree Protocol (STP) introduces and gain access to that lost bandwidth, although the faster convergence and better L2 scaling also comes into play. In this case, the simplest approach is using Virtual Port Channels (vPC), which keep STP in place, but eliminate its shortcomings. The advantage of this approach is that vPCs are an NX-OS feature already built into their Nexus switches. For customers with more demanding requirements such as the need for more bi-sectional bandwidth, larger scale or a flattened L2 architecture, FabricPath or TRILL offers another alternative. One of the advantages of this approach is that its standards-based interoperability allows customers to maintain flexibility and choice with their architecture. As always, we will continue to provide robust STP support for stable legacy environments. The net result is that these technologies can be deployed in a granular manner where they are needed without needing to replace or disrupt infrastructure that is otherwise working fine.

Here are some resources for the readers. The first document is an **overview of FabricPath**, the second is a **whitepaper on eliminating STP shortcomings with vPC**, and the final link is a **whitepaper on FabricPath**.

**Jim Metzler, Ashton, Metzler & Associates**

What is the relationship between FabricPath and TRILL? Will you also support SPB? What do you see as the major weaknesses of SPB?

With regards to Shortest Path Bridging (SPB), or IEEE 802.1AQ, while Cisco believes the approach has merit, we believe that TRILL and FabricPath offer a better long term solution. We believe one of the underlying tenets of SPB, backward hardware compatibility, will become a hindrance to evolution of the protocol--for example, IEEE is revising the frame format to support functionality like TTL, which validates some of the initial choices made by IETF. From a practical perspective, since moves to flattened network architectures are often done in concert with a switching hardware refresh, the hardware compatibility tends to not be a consequential issue. Finally, when looking at the dynamic and oft changing traffic patterns of the next generation data center, driven by requirements like VM mobility, scale-out app architectures, and cloud deployment models, FabricPath/TRILL provides a more operationally efficient approach.

Using Brocade products, customers today are leveraging the following technologies and solutions to replace STP and improve East-West traffic:

1. Flatter, faster network architectures using a distributed control plane and based on cut-through architectures with wire-speed performance and hardware-based load balancing
2. Based on the emerging TRILL (Transparent Interconnection of Lots of Links) standard to provide the following:
    - Eliminates the need for Spanning Tree Protocol (STP)
    - Delivers an advanced multi-path network using link state routing.
    - Traffic is automatically distributed across equal-cost paths ensuring it takes the shortest path for minimum latency without manual configuration.
    - Events such as added, removed, or failed links are not disruptive and traffic is automatically rerouted in less than a second
    - Reduces dependency on and complexity with subnets to dramatically increase VM sphere of mobility
3. Data Center Bridging (DCB), Priority-based Flow Control (802.1Qbb) and Enhanced Transmission Selection (802.1Qaz) capabilities ensure traffic is lossless and priorities are properly set and maintained. These technologies are ideal for improving east-west traffic and provide a great foundation for transporting Fibre Channel over Ethernet (FCoE), iSCSI and other storage traffic while enabling LAN and SAN convergence for Tier 2 and 3 applications.

For more information on how Brocade is helping customers improve East-West traffic communication, use **this link**.

**Jim Metzler, Ashton, Metzler & Associates**

As you note, TRILL is an emerging standard. What, if anything, are you recommending today as an alternative to spanning tree? Will you also support SPB? What do you see as the major weaknesses of SPB?

BROCADE

Customers have been deploying Brocade Ethernet Fabric using VCS for nearly a year primarily because it allows them to eliminate the need for Spanning Tree. VCS is a viable alternative because it provides L3-type intelligence at L2 and completely eliminates STP enabling customers to build flatter, faster networks that are twice as resilient, self-aggregating trunks to reduce OpEx 15-20% (compared to 3-tier architectures), and hardware-based load balancing for achieving a 2x improvement in performance.

While VCS is based on TRILL in the data path, VCS leverages Brocade's 15 years of leadership and experience in storage area networking for delivering ultra reliability and scale as a more efficient, and scalable alternative to Shortest Path Bridging for Enterprise data center LANs.

As with any industry standard, each vendor will interpret and leverage parameters differently and while Brocade continues to collaborate with industry vendors such as Cisco to drive the TRILL standard to approval, the TRILL specification has matured to a point where customers can and are realizing significant value and we will continue to help develop and leverage industry standards when they solve problems and create opportunities for our customers.

**Jim Metzler, Ashton, Metzler & Associates**

In order to clarify your response to my question about the best alternative to the spanning tree protocol, ***kindly provide a simple network diagram (as a PDF)*** to show how your STP-free solutions would look in the following scenario:

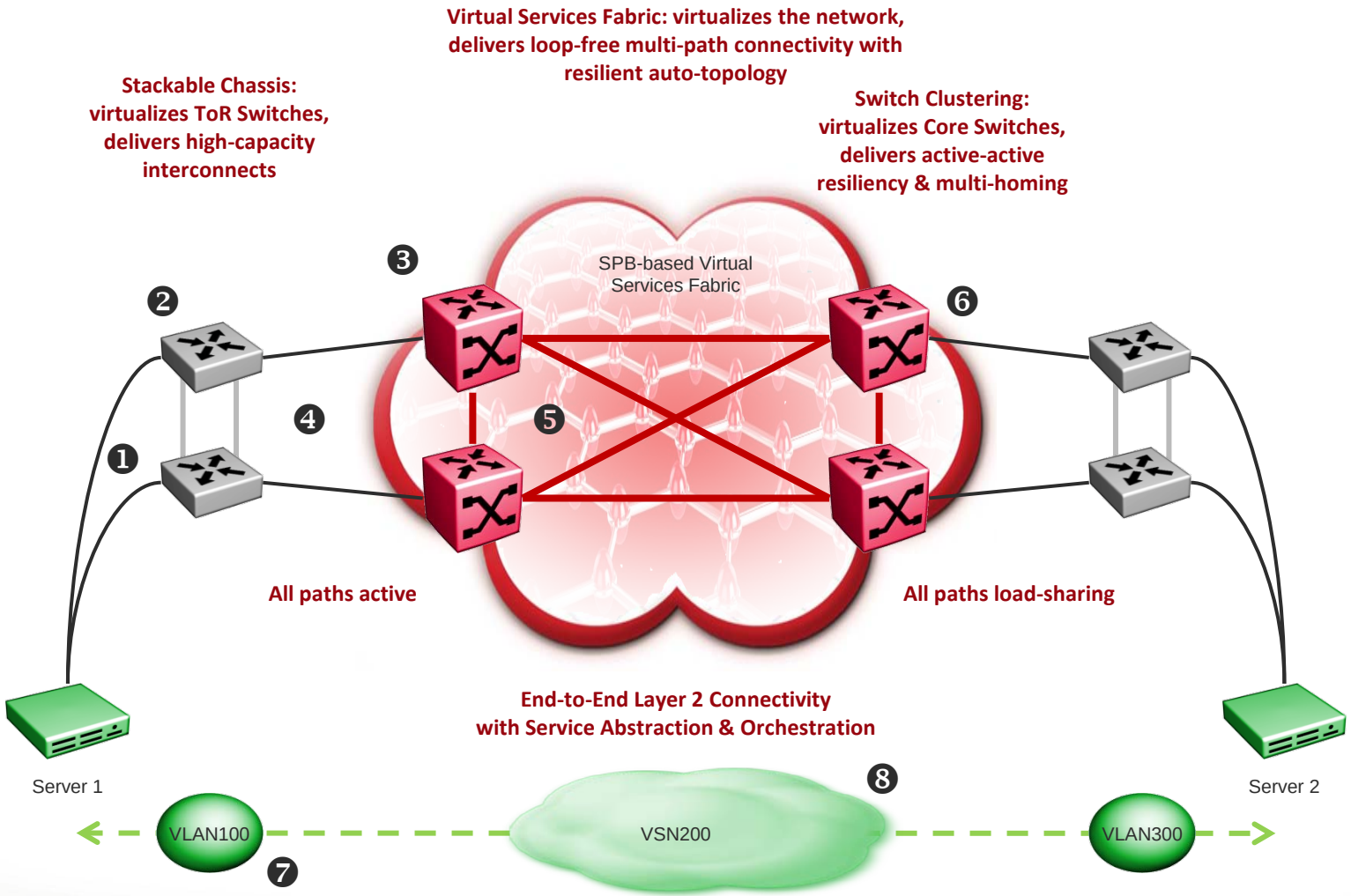
It is desired to have full network redundancy with dual path load sharing between a pair of servers located in PODs at opposite ends of the data center. Each server is dual attached, with one 10 GbE link to each of two access switches. Kindly show how end-to-end dual load sharing paths can be extended through an additional tier of switching that provides connections between the two PODs. Indicate the relationship between the switches in the same tier (e.g, a virtualized pair of switches) and switches in separate tiers (e.g., access/aggregation or leaf/spine). Kindly also state whether the end-to-end path is entirely Layer 2 or if a Layer 3 boundary is crossed.

AVAYA

Avaya's solution for an STP-free Data Center can be demonstrated using the following **network diagram**. (Click **here** for diagram).

# STP-free Data Center

## Featuring: Virtual Services Fabric, Switching Clustering, & Stackable Chassis



Servers are dual-attached via 10GbE links to separate Top-of-Rack Switches, providing active-active resiliency and capacity

1. The ToR Switches are deployed as a virtual “Stackable Chassis”, avoiding any single point-of-failure and providing high-capacity, low-latency bandwidth for east-west traffic

2. The “Switch Cluster” configuration provides a virtualized Core, appearing as a single network entity, simplifying deployments and ensuring full utilization of all links and resources

3. ToR Switches are multi-homed into the Core, with all links supporting active-active load-sharing

4. The Virtual Services Fabric delivers a loop-free multi-pathing topology that is automatically built and maintained; all links are available for the load-sharing of traffic flows

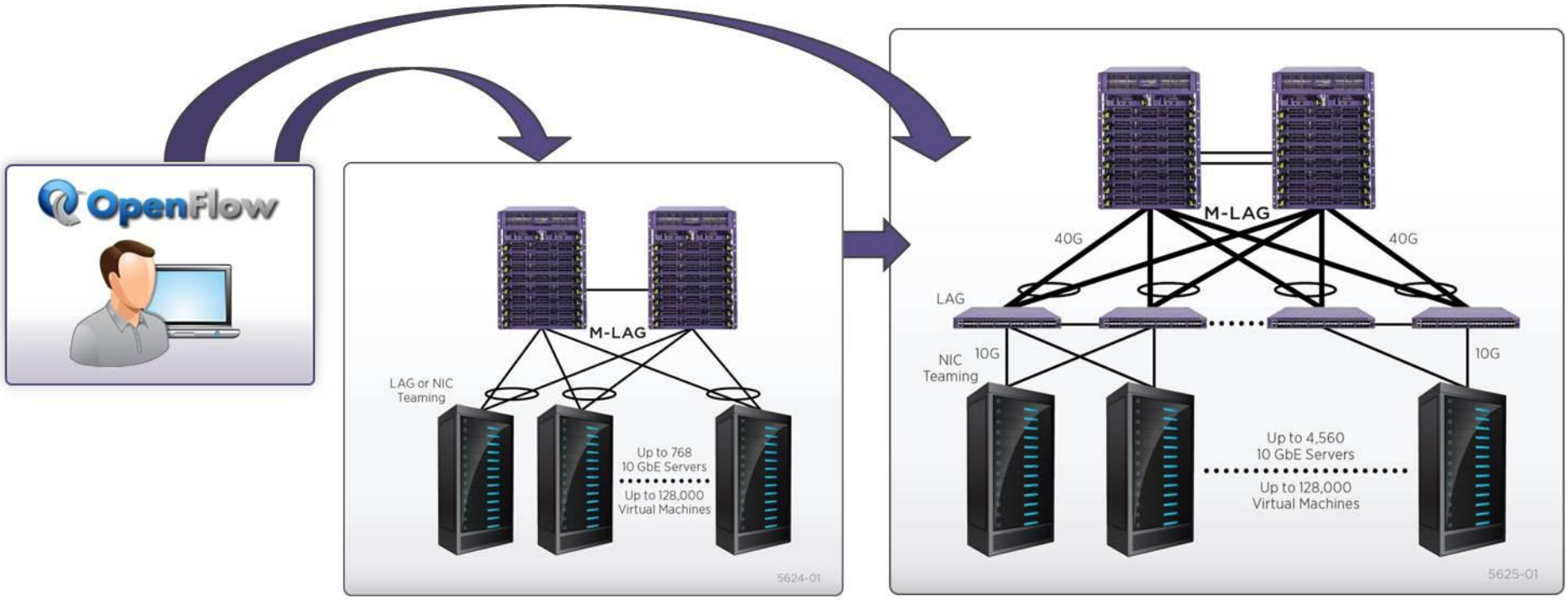
5. SPB’s Service Abstraction capability allows any-to-any connectivity to be dynamically provisioned at the edge and instantaneously advertised throughout the Fabric

6. Mapping of Services is locally-significant, which enables total provisioning flexibility; there’s no limitations around maintaining or extending IDs

7. Services, such as VM migration, can be orchestrated in real-time with automated tools that significantly reduce time-to-service and errors

The solution delivered is Layer 2 end-to-end and any-to-any; support is also provided for extending VRFs and other L3 scenarios.

A Layer 2 network is shown in this **network diagram**, where M-LAG can be implemented from servers into the first layer of switches. If additional network tiers are required, they also can support M-LAG, as depicted in the **diagram**. (Click **here** for diagram.)




Data Center Bridging | M-LAG | Direct Attach™ / VEPA | XNV™ | OpenFlow*