# Avaya VENA Fabric Connect

Ashton, Metzler & Associates

Leverage Technology & Talent for Success

## Executive Summary

The Avaya VENA Fabric Connect solution is based on the IEEE 802.1aq Shortest Path Bridging (SPB) protocol in conjunction with Avaya extensions that add Layer 3 routing capabilities to SPB functionality. The Fabric Connect solution has a range of features that can greatly simplify network operations and at the same time enhance the network's ability to provide the appropriate level of service for business critical applications. These features include:

- Edge-only provisioning
- Layer 2 network virtualization supporting traffic isolation and automated network configuration for VM mobility
- Layer 3 network virtualization for enhanced traffic isolation
- Equal cost Layer 2 multi-pathing for optimized North-South and East-West traffic flows together with active-active connectivity for dual-attached servers
- Protocol simplification via the use of IS-IS for both Layer 2 and Layer 3 control planes

With Avaya enhancements, an SPB-based network offers a fully field tested and proven alternative to other proposed alternatives for virtualization and simplification of data center networks, such as those based on Transparent Interconnection of Lots of Links (TRILL), MAC-in-UDP network virtualization overlays and Software-Defined Networks (SDNs).

## Introduction

The Avaya VENA Fabric Connect solution is based on the MAC-in-MAC variant of the IEEE 802.1aq Shortest Path Bridging (SPB) protocol in conjunction with Avaya extensions that add Layer 3 routing capabilities to the base SPB protocol. MAC-in-MAC SPB uses IEEE 802.1ah encapsulation and the IS-IS routing protocol to provide Layer 2 network virtualization in addition to loop-free equal cost multi-path Layer 2 forwarding. SPB multi-path forwarding eliminates the long convergence times and blocked link inefficiencies that are characteristics of LANs based on the Spanning Tree Protocol (STP). Layer 2 network virtualization and VLAN extension are enabled by Service IDs (I-SIDs) that are part of the outer MAC encapsulation; Avaya refers to these private networks as 'Virtual Service Networks (VSNs)'. Since the I-SID header field length is 24 bits, the theoretical limit is for 16 million VSNs, and current hardware implementations typically have the memory capacity to support 4,000 I-SIDs per switch. Given that each node will typically only directly support a fraction of the I-SIDs in the network, the practical limit on VSNs far exceeds the network-wide limit of 4,094 for traditional 802.1Q VLAN-based networks.

One of the main advantages of the SPB protocol is the fact that provisioning of network virtualization is achieved by configuring only the Edge SPB Switches, which may be either Access or Aggregation Switches. When new VMs are created and need VLAN connectivity, the virtual service networks that provide VLAN extension are provisioned at the network edge only and don't need to be configured throughout the rest of the network infrastructure as would be the case for the IEEE 802.1Q trunks that are traditionally used for VLAN extension within the data center. This edge-only provisioning model provides a significantly faster time-to-service for network configuration, which allows the provisioning of the network to be as dynamic as the provisioning of new application instances on virtualized servers.

In Avaya's Fabric Connect, basic SPB has been extended to also support Layer 3 forwarding and Layer 3 network virtualization. A common requirement in the data center is the ability to route traffic between IEEE 802.1Q VLANs. The capability to route between Layer 2 VSNs is provided by an Inter-VSN Routing function that is encompassed in the IP/SPB extensions to SPB implemented by Avaya and described in an IETF draft. Inter-VSN Routing allows Fabric Connect nodes to act as default gateways/routers. Therefore, in the scenario where traffic needs to move between VSNs, within a virtualized server or within a Pod, it can be efficiently routed by the Fabric Connect Access Switch, rather than by upstream Layer 3 Aggregation or Core Switches.

IP/SPB also provides Layer 3 VSNs by extending Virtual Routing and Forwarding (VRF) instances at the edge of the network across the Fabric Connect network without requiring that the core switches also support VRF instances. VLAN-extension VSNs and VRF-extension VSNs can run in parallel on the same Fabric Connect network to provide isolation of both Layer 2 and Layer 3 traffic for multi-tenant environments. With the combination of Layer 2 and Layer 3 VSNs, Fabric Connect can be used as a unified backbone technology not only in the data center, but also in the campus network, the metropolitan area network, and even a private Ethernet WAN.

The remainder of this white paper provides further details on the Avaya VENA Fabric Connect implementation of SPB and IP/SPB, as well as a more detailed discussion of the benefits that it provides.

## SPB Layer 2 Fabrics

In a three-tier data center LAN, the SPB Fabric encompasses the Aggregation and Core Switches, and possibly the Access Switches as well. As part of the transition to an SPB-based Fabric it's only the Core and any Aggregation Switches that must be SPB-capable. Whether or not the existing Access Switches are replaced with SPB-capable Access Switches, there is not a requirement to reconfigure the physical or virtual servers.

The boundary between the MAC-in-MAC SPB domain and the traditional IEEE 802.1Q domain is handled by the SPB Edge Switches, defined as Backbone Edge Bridges (BEBs) within the standard. At the BEBs, VLANs are mapped into Layer 2 Virtual Service Networks, identified by I-SIDs, based on local service provisioning and the MAC-in-MAC encapsulation shown in Figure 1 is performed. After IS-IS has determined the topology, the BEBs use link metrics to calculate at least one shortest path tree to reach every destination BEB MAC Address (B-MAC). When a packet arrives at an ingress

BEB, the BEB looks up the egress BEB's B-MAC Address in its forwarding database and forwards the encapsulated packet over the shortest path tree. Each BEB learns end system Addresses, or Customer MAC Addresses (C-MACs), within the Layer 2 VSN and maintains a forwarding database that maps these C-MACs to their corresponding B-MACs. As end stations are provisioned or migrated, the IS-IS protocol will advertise the changes across the SPB network.


| ◄ Edge Address ► | | | | ◄ Backbone Address ► | | | |
|---|---|---|---|---|---|---|---|
| Payload | VLAN ID | Source C-MAC | Dest C-MAC | I-SID | B VLAN ID | Source B-MAC | Dest B-MAC |

**Figure 1: Layer 2 VSN with MAC-in-MAC Encapsulation**

Up to sixteen Equal Cost Multi-Tree (ECMT) paths are initially provided for in an IEEE 802.1aq network, however, many more are possible. The choice as to which ECMT path that a specific flow will be assigned to depends on the path/selection/load-balancing algorithm the operator has chosen. The algorithms that SPB support include deterministic and symmetric multiple equal cost routes via hashing across different end-to-end routes from the head-end using either the normal IP n-tuple, or other micro flow order-preserving hash mechanisms.

In contrast to STP's single path forwarding and blocking of parallel paths, SPB uses Reverse Path Forwarding Check (RPFC) for loop mitigation. RPFC leverages SPB's symmetry between forward and reverse paths and discards frames that do not arrive on the shortest path from the source BEB.

Figure 2 shows how the Fabric Connect network supports server connectivity in a network where the Server's Access Switches are SPB-capable. Redundant connectivity between the VLAN domain and the SPB infrastructure is achieved by operating two SPB-capable Switches in Switch Clustering mode (using Avaya's Split Multi-Link Trunk technology). This allows the active-active dual-homing of any traditional standard IEEE 802.3ad LAG-capable device into Fabric Connect network. This means that from the perspective of the Server that the Switch Cluster pair appears to be an IEEE 802.3ad LAG end-point.

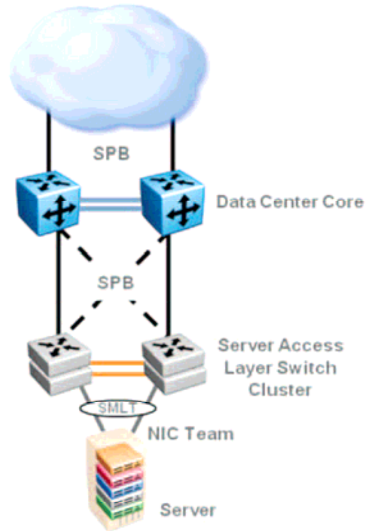


**Figure 2: Fabric Connect Server Connectivity**

Figure 3 shows how a Fabric Connect core can be deployed without requiring that the Server's Access Switches support SPB. In this case, Switch Clustering is also used for the link between the Access Switches and the Core Switches. This configuration could be used in an initial phase of transition of the Core to SPB; e.g., as a precursor to the configuration of Figure 1. In each of these migration phases end devices with dual-homed Switch Cluster attachments do not require any configuration changes. It is important to note that an infrastructure operating the SPB protocol can coexist with the other existing protocols. This allows for a smooth migration whereby User/Server VLANs can be moved one at the time from the traditional transport to the SPB-based transport,

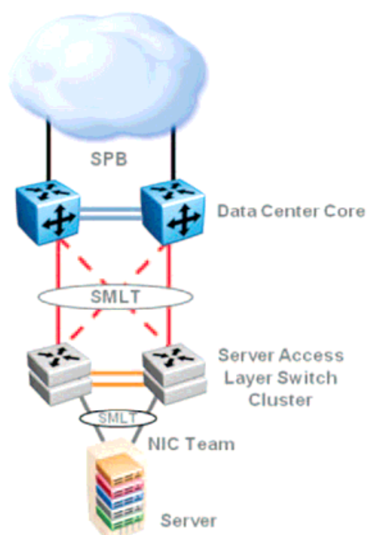


**Figure 3: Traditional Access Switches Connecting to the Fabric Connect Core**

## Fabric Connect Layer 3 Extensions

Avaya's Fabric Connect Layer 3 extensions to SPB are based on the previously mentioned IP/SPB IETF draft. The intent of IP/SPB is to extend SPB with Layer 3 functionality by leveraging the flexibility of the IS-IS Routing protocol. With IP/SPB, the Edge BEB provides a standard IP interface to attached IP devices. The Ingress BEB performs a route lookup on the destination IP Address, which will resolve to the B-MAC of the remote BEB to which the destination VRF or IP end system is attached. The BEB encapsulates the IP packet in an SPB Backbone Ethernet header as shown in Figure 4. It should be noted that with IP/SPB the IP packet is forwarded directly at Layer 2 to the Egress BEB, eliminating multiple IP lookup operations on the intermediate nodes on the end-to-end path. IP/SPB supports several types of Layer 3 functionality including:

**Layer 3 VSNs/VPNs:** As shown in the top half of Figure 4, IP/SPB supports Layer 3 virtualization by extending VRF instances at the edge of the network across the Fabric Connect network without requiring that the Core Switches also support VRF instances. In Figure 4, Enclave A denotes a network domain that could be an enterprise network, a departmental network, or a tenant's virtual network in a public cloud data center. Layer 3 VSNs use I-SIDs to identify and segregate different Layer 3 VSN connectivity services and different Enclaves within the network. As shown in the bottom half of Figure 4, in a multi-tenant network, each tenant can be assigned a set of unique services instances (I-SIDs) that identify that tenant's Layer 2 and Layer 3 VSNs. Layer 2 and Layer 3 VSNs can run in parallel on the same Fabric Connect network to provide complete traffic isolation for multi-tenant environments.

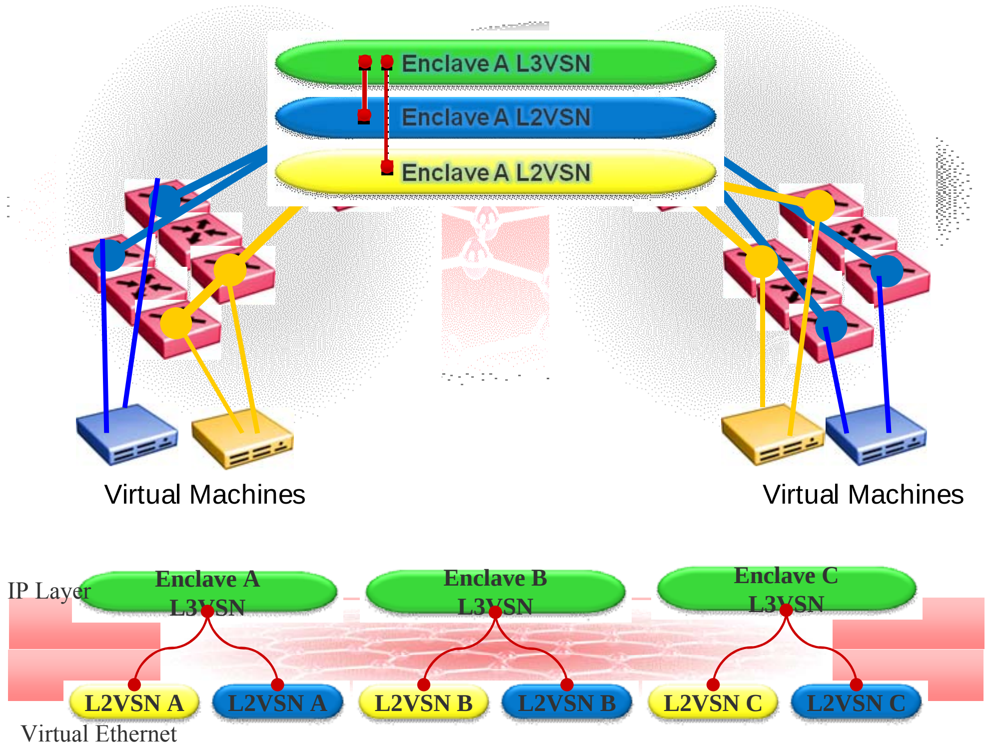


**Figure 4: Layer 3 VSNs**

With IP/SPB, the Edge BEB provides a standard IP interface to attached IP devices. IP reachability information is carried across the network using normal or extended IS-IS type-length-values (TLVs) and then mapped to the remote B-MAC Destination Address. The Ingress BEB performs a route lookup on the Destination IP Address, which will resolve to the B-MAC of the remote BEB to which the destination VRF or IP end system is attached. For IP VPNs, a VRF is associated with an I-SID value, and the BEB encapsulates the IP Packet in an SPB Backbone Ethernet Header as shown in Figure 5.

| Received at Ingress BEB | | |
|---|---|---|
| IP Packet | Source C-MAC | Ingress BEB C-MAC |

| In Transit Through the SPB Network | | | |
|---|---|---|---|
| IP Packet | IPVPN I-SID | Ingress BEB B-MAC | Egress BEB B-MAC |

| Received at Destination | | |
|---|---|---|
| IP Packet | Egress BEB C-MAC | Dest C-MAC |

**Figure 5: IPVPN-BMAC Encapsulation**

**Inter-VSN Routing:** As shown in Figure 6, SPB BEBs can route traffic directly between Layer 2 VSNs without involving external routing devices or VRFs. Encapsulation is similar to that shown in Figure 5, with the exception that the I-SID field is not used.

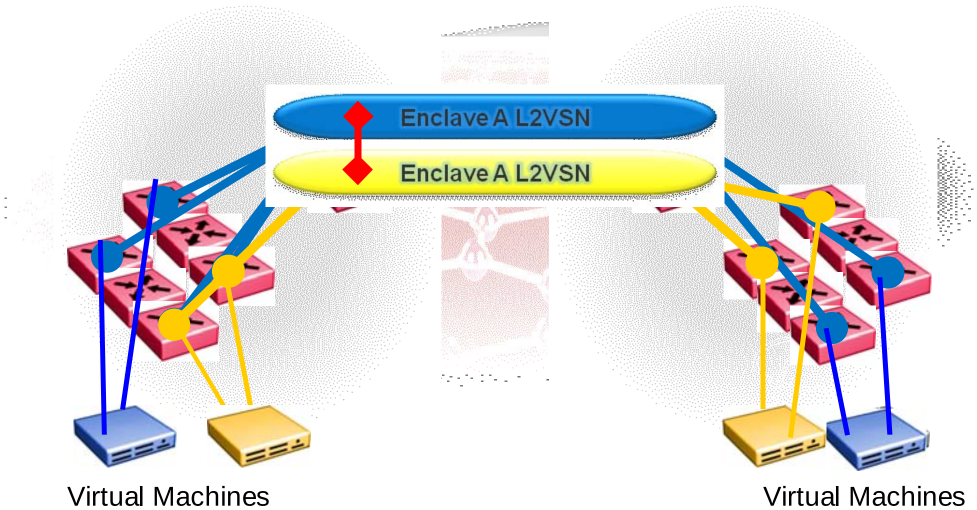


**Figure 6: Inter-Layer 2 VSN Routing**

**IP Shortcut Routing:** As shown in Figure 7, SPB BEBs can also route IP traffic natively between VLANs/Subnets without requiring that they be mapped to VSNs/I-SIDs.

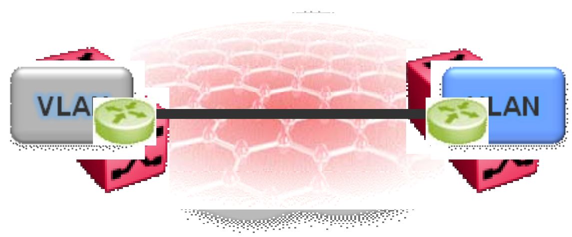


**Figure 7: Shortcut Routing**

**VXLAN Extension:** As shown in Figure 8, the Avaya VENA Fabric Connect solution can also provide VXLAN extension for VMware environments. Avaya BEBs can support VXLAN virtual networks by providing Layer 3 services for forwarding VXLAN packets. In the control plane, Fabric Connect can use SPB IP Multicast as an alterative to Protocol Independent Multicast (PIM) to improve scalability and to reduce overall complexity. PIM was not intended for a many-to-many Multicast environment, which has proven to be a problem in environments with large numbers of VXLAN Tunnel End Points (VTEPs.) In the future, Fabric Connect BEBs will be able to map the VXLAN virtual network segment IDs into VSN/I-SIDs and bridge VXLAN UDP/IP packets across the Layer 2 fabric. The latter capability will allow a single Fabric Connect network to support both VXLAN and VLAN endpoints, eliminating the need for VXLAN/VLAN gateway functionality in virtual appliances of other network devices.


**Figure 8: VXLAN Extension**

## Unique Benefits of the Avaya VENA Fabric Connect solution

The Avaya VENA Fabric Connect implementation, with its foundation on and extensions to SPB, provide a wide variety of unique benefits compared to other less unified approaches to edge virtualization, network virtualization and multi-path fabrics. Some of these benefits include:

**Support for a Wide Range of Network Topologies.** The network topology can be selected from a variety of options with the goal of optimizing bandwidth, hop count and/or scalability of the network. In addition to more conventional 2-tier and 3-tier hierarchical topologies, possible topologies include fat trees and dragonfly networks. While fat trees are generally preferred for High Performance Computing (HPC) applications, dragonfly networks offer a combination of very high scalability, minimal hop count between nodes, and high east-west bandwidth that is proving desirable both in hyper-scale data centers and for latency-sensitive enterprise applications. Dragonfly networks achieve high scalability by using fully meshed groups or clusters of Switches to create high port-count virtualized Switches at the access layer. East-west traffic flows are optimized by intra-group meshing as well as direct attachments among virtual Switch groups, which occurs without the involvement of the aggregation layer Switches. The Avaya VENA Distributed Top-of-Rack configuration shown in the data center that is depicted in Figure 9 is an example of a hybrid dragonfly/2-tier network topology that can be enabled by the Avaya VSP 7000 and VSP 9000 Switches.

In addition to the functionality described above, an Avaya Fabric Connect network can be extended to cover not only the data center but also the campus network and the entire enterprise network as is also shown conceptually in Figure 9.

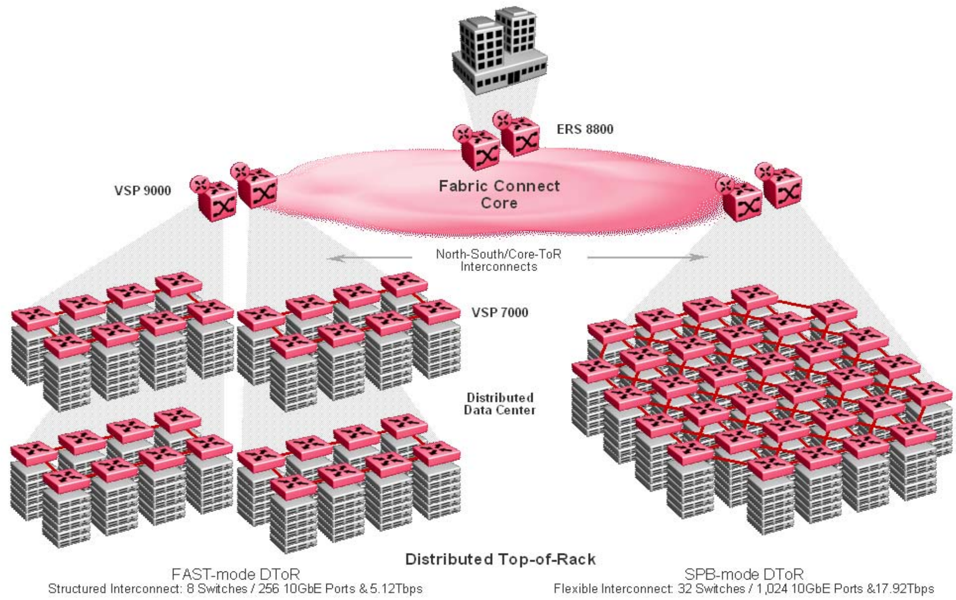


**Figure 9: Avaya Fabric Connect Spanning the Enterprise**

**Unified/Simplified Control/Forwarding Plane.** The extended SPB control and forwarding planes effectively decouples the network services layer from the underlying infrastructure, while supporting the full range of possible network services, including both Layer 2 and Layer 3 virtualization. This occurs without the need for additional protocols to deal with special situations; e.g., Layer 2 virtualization over the WAN. The Fabric Connect implementation of SPB provides essentially all of the service richness of MPLS without any of the complexity that has resulted in very few IT organizations implementing MPLS themselves.

**Simplified Edge Provisioning.** New physical servers, virtual servers and/or end systems can be added to the Fabric Connect network with a single command that links that entity to its VSN at the edge. This avoids error-prone configuration changes to multiple devices within the core of the network. The Avaya Virtualization Provisioning Service delivers automation and orchestration for VM mobility.

**Optimized Traffic Flows.** The Avaya Fabric Connect network provides symmetrical Layer 2 multi-pathing that optimizes the utilization of bandwidth and enables re-convergence in under 50 milliseconds after a change in the network topology. While Fabric Connect supports virtually any meshed LAN or network topology, the Avaya VENA Distributed Top-of-Rack access configuration optimizes East-West traffic flows by bypassing Aggregation Switches via direct ToR-to-ToR connectivity as shown in Figure 9. In multi-tenant environments, Fabric Connect supports the option of Layer 3 traffic isolation in addition to Layer 2 traffic isolation that is provided by other network virtualization solutions.

**Investment Protection at the Network Edge.** A Fabric Connect implementation beginning in the core of network does not impact the edge of the network where existing physical access switches, hypervisor vSwitches, and server configurations can remain unchanged. In addition, Fabric Connect is fully compatible with, and improves the control plane scalability of, existing VXLAN Layer 2 virtualization deployments

**Application-Driven Networking.** The scalability and flexibility of VSNs allows Fabric Connect to support network-wide *application virtualization* in the sense that all entities participating in a networked application can be assigned to an application-specific VSN. Additionally, access to a VSN can be controlled by the Avaya Identity Engines solution with the authenticated user accessing ports being automatically configured for authorized VSNs and for appropriate levels of QoS.

## Summary and Call to Action

Avaya VENA Fabric Connect supports both a wide range of network functionality, including Layer 2 and Layer 3 network virtualization and multi-path forwarding, and it can serve as a unified backbone technology not only in the data center, but also in the campus network, the metropolitan area network, and even a private Ethernet WAN. The network simplification created by the SPB technology that underpins Fabric Connect is typified by the fact that provisioning of network virtualization is achieved by configuring only at Edge Switches. This edge-only provisioning model provides a significantly faster time-to-service for network configuration, which allows the provisioning of the network to be as dynamic as the provisioning of new application instances on virtualized servers.

The traditional approach generally taken by the IT industry to deliver additional network functionality is to design and implement a new standalone, mission-specific protocol; the goal of each protocol being to provide a solution for a specific problem. This has led to protocol proliferation and an ever-increasing level of complexity, especially as network technology evolves to fully support initiatives such as server virtualization and multi-tenancy. The situation is further exacerbated in the current environment in which multiple competing protocols have been proposed as possible solutions to some of the particular problems that are associated with initiatives such as virtualization and multi-tenancy.

IT organizations are in a position to make a fundamental choice relative to how they approach network design on a going forward basis. They can either continue with the current approach of continually adding additional protocols to an already complex network or they can choose an approach that dramatically simplifies network design. SPB, together with extensions proposed in IP/SPB and implemented in the Avaya VENA Fabric Connect solution, offers network designers an opportunity to satisfy emerging requirements for next-generation network functionality. Additionally, they dramatically reduce the complexity of the network, both in terms of the number of control and data plane protocols that must be supported in order to deliver network services, and in terms of the operational effort required to provision and operate increasingly dynamic IT infrastructures.

## Acronyms

| | |
|---|---|
| B-MAC | BEB MAC |
| BEB | Backbone Edge Bridges |
| C-MAC | Customer MAC |
| ECMT | Equal Cost Multi-Tree |
| HPC | High Performance Computing |
| IEEE | Institute of Electrical and Electronics Engineers |
| IETF | Internet Engineering Task Force |
| IP | Internet Protocol |
| I-SID | Instance Service IDs |
| IS-IS | Intermediate System To Intermediate System |
| LAG | Link Access Group |
| MAC | Media Access Control |
| MPLS | Multiprotocol Label Switching |
| PIM | Protocol Independent Multicast |
| QoS | Quality of Service |
| RPFC | Reverse Path Forwarding Check |
| SDN | Software Defined Network |
| SMLT | Split Multi-Link Trunk |
| SPB | Shortest Path Bridging |
| SPBM | Shortest Path Bridging MAC-in-MAC |

| | |
|---|---|
| STP | Spanning Tree Protocol |
| TLV | Type-Length-Value |
| ToR | Top-of-Rack |
| TRILL | Transparent Interconnection of Lots of Links |
| UDP | User Datagram Protocol |
| VLAN | Virtual LAN |
| VM | Virtual Machine |
| VPN | Virtual Private Network |
| VRF | Virtual Routing and Forwarding |
| VSF | Virtual Services Fabric |
| VSN | Virtual Service Network |
| VTEP | VXLAN Tunnel End Point |
| VXLAN | Virtual eXtensible LAN |
| WAN | Wide Area Network |