avaya.com

Network Virtualization using Shortest Path Bridging and IP/SPB

Abstract

This White Paper discusses the benefits and applicability of the IEEE 802.1aq Shortest Path Bridging (SPB) protocol which is augmented with sophisticated Layer 3 routing capabilities. The use of SPB and the value to solve virtualization of today's network connectivity in the enterprise campus as well as the data center are covered.

This document is intended for any technically savvy network manager as well as network architect who are faced with:

- Reducing time to service requirements
- Less tolerance for network down time
- Network Virtualization requirements for Layer 2 (VLAN-extensions) and Layer 3 (VRF-extensions)
- Server Virtualization needs in data center deployments requiring a large set of Layer 2 connections (VLANs)
- Traffic separation requirements in campus deployments for security purposes as well as robustness considerations (i.e. contractors for maintenance reasons needing access to their equipment or guest access needs)
- Multi-tenant applications such as airports, governments or any other network with multiple discrete (legal) entities that require traffic separation

Table of Contents

1. Introduction

The evolution of Ethernet technologies continues with the IEEE 802.1aq standard of Shortest Path Bridging. This next generation virtualization technology will revolutionize the design, deployment and operations of the enterprise campus core networks along with the enterprise data center. The benefits of the technology will be clearly evident in its ability to provide massive scalability while at the same time reducing the complexity of the network. This will make network virtualization a much easier paradigm to deploy within the enterprise environment.

Shortest Path Bridging with its extensions eliminates the need for multiple overlay protocols in the core of the network by reducing the core to a single Ethernet based link state protocol which is providing all virtualization services (virtualization of bridging, routing, multicast) in an integrated model. In addition, by relying on end-point-service provisioning only, the idea of "build it once and don't have to touch it again" becomes a true reality. This simplicity also aids in greatly reducing time to service for new applications and network functionality.

The design of networks has evolved throughout the years with the advent of new technologies and new design concepts. Customer requirements drive this evolution and the adoption of any new technology is primarily based on the benefit it provides versus the cost of implementation. The cost in this sense is not only cost of physical hardware and software, but also in the complexity of implementation and on-going management. New technologies that are too "costly" may never gain traction in the market even though they provide a theoretical benefit.

In order to change the way networks are designed, the new technologies and design criteria must be easy to understand and easy to implement. When Ethernet evolved from a simple shared media with huge broadcast domains to a switched media with segregated broadcast domains, there was a shift in design. The ease of creating a VLAN and assigning users to that VLAN made it commonplace and a function that went without much added work or worry. In the same sense, if Shortest Path Bridging is to be successful, then the implementation of network virtualization must become as common and easy as creating a VLAN is today.

The key value propositions for SPB include:

- Standards-based
- IEEE 802.1aq standard – no lock in technology
- Resiliency
- Single robust protocol with sub-second failover
- Optimal network bandwidth utilization
- Simplicity
- One protocol for all network services
- Plug & Play deployment reduces time to service
- Scalability
- Evolved from carrier with enterprise-friendly features
- Separates infrastructure from connectivity services
- Flexibility
- No constraints on network topology
- Easy to implement virtualization

2. Benefits of Shortest Path Bridging

As with any new technology, it is important to understand the benefits that can be expected from its use. It is critical to weigh these benefits against the cost in order to truly arrive at the realistic value proposition. Understanding the values also sets the proper expectations up front and will be a key factor in how and where the technology is deployed. Several of these benefits are provided here as proof points for the use of Shortest Path Bridging*.

2.1 Network Service Enablement

2.1.1 Data Center Bridging

Data Center Bridging (DCB) is gaining attention by many enterprises. The ability to support storage traffic over Ethernet has many significant benefits. The most compelling being the cost savings by converging the data center on one infrastructure. Significant savings can be realized for both capital expenditures (CAPEX) as well as operational expenditures (OPEX). Reduction in the amount of hardware (network interfaces, host bus adapters, storage switches) contributes to these large and attainable savings.

iSCSI and NAS are storage technologies which are based on TCP/IP, thus operate on today's Ethernet networks without any additional functionality. Fibre Channel by nature runs on a separate infrastructure from traditional Ethernet. Recent enhancements to the T11 standard have introduced Fibre Channel over Ethernet (FCoE) that provides capabilities of running Fibre Channel over an Ethernet infrastructure. In order for Fibre Channel storage traffic to be converged on Ethernet, several enhancements must be made. Ethernet by design is a transport that can lose packets and simply retransmit. In a storage network, this is not acceptable; therefore the Ethernet infrastructure must provide a lossless functionality. This new functionality is part of the 802.1 DCB standardization projects.

Another major requirement for storage transport based on FCoE is transparent Layer 2 connectivity. There is no concept of Layer 3 routing domains for Storage Area Networks which makes the extension of SANs between geographically dispersed data centers that much more of a challenge. Root Bridge based Spanning Tree Layer 2 topologies using VLANs are not seen as robust enough for storage transport. In addition, these networks do not support any form of shortest path switching, which is required to provide minimal latency for storage traffic.

With the convergence of the SAN and traditional LAN within the data center, SPB provides a unique value proposition to seamlessly extend the Layer 2 SAN domains within and across data centers. A shortest path with minimal latency will automatically be created and if there is a failure of a link or switch, the failover time will be less than sub-second. SPB also removes the complexity of manual VLAN extensions and eliminates the cumbersome Spanning Tree protocol from the design. SPB's capability of using multiple parallel paths (ECT's) is another major advantage in providing a truly superior transport solution for Storage Area Networks.

2.1.2 Server Virtualization

The expansion of the data center, a result of both scaled up server architectures and traditional "one application, one server" sprawl, has created problems in housing, powering, and cooling large numbers of underutilized servers. In addition, IT organizations continue to deal with the traditional cost and operational challenges of matching server resources to organizational needs that seem fickle and ever changing. These are two leading factors that have led to the mass adoption of server virtualization. The use of virtualization and specifically virtual machines is profoundly changing data center dynamics.

Virtual machines can significantly mitigate many of these challenges by enabling multiple application and operating system environments to be hosted on a single physical server while maintaining complete isolation

* For the purposes of simplicity, this paper with focus on the MAC-in-MAC version - otherwise known as SPBM - of the IEEE 802.1aq standard.

between the guest operating systems and their respective applications. Hence, server virtualization facilitates server consolidation by enabling organizations to exchange a number of underutilized servers for a single highly utilized server running multiple virtual machines.

These new server virtualization technologies allow the dynamic placement of the applications on any virtualized server infrastructure. With this data center(s) are becoming a "cloud", services can be placed wherever needed or where most resources are available. To enable this transparency, Layer 2 VLAN extensions within the data center as well as across the backbone infrastructure between the data centers are required to provide a robust (transparent) connectivity service.

In today's traditional LAN/WAN design, the extension of numerous VLANs and their propagation within data centers can prove challenging. Ensuring that all redundant links are properly configured as well as all switches can be a time-consuming operation, and can introduce significant risk due to the need to regularly administer the configurations of crucial core devices. This is especially true in data center environments that are continually shifting to match application and business requirements.

SPB removes the complexity by eliminating the need to configure multiple points throughout the network. The simple end-point provisioning is done where the application meets the network, while all points in between are automatically provisioned through SPB's robust link state protocol. The ability to transparently extend Layer 2 and/or Layer 3 domains across a virtual backbone with virtually no effort and no risk enforces SPB's unique value proposition.

2.1.3 Multi-Tenant Applications

As large enterprises continue to evolve, many have become very similar to network service providers/carriers. The enterprise IT organization is the "service provider" for its internal customers. With this comes a new and evolving set of requirements that traditional providers have been accustomed to for many years. The new network requirements are instantiating enhanced design methodologies in order to create complete traffic separation between the customer domains, provide uninterrupted service for business applications, significantly reduce the time to service from weeks/months to hours/days and accommodate flexible network deployments.

With the need to support these complex multi-tenant environments comes the added cost and complexity of operating a "carrier-class" network. In most cases, enterprise network operations teams have a relatively small staff and budget. Carrier technologies, which have been built to scale to thousands of customers, have an inherent complexity, which is in many cases too expensive to operate for enterprise customers. A simpler solution which provides the same or even more functionality can help reduce network operation costs significantly.

SPB is the technology that will help satisfy all aspects of the multi-tenant customer. The technology evolved from similar protocols used by carriers and service providers. SPB has been enhanced to add "enterprise friendly" features to give it the best of both worlds, carrier robustness / scalability and applicability with enterprise-class features and interoperability. The simplicity of the technology doesn't require an entire team with specialized training or knowledge and therefore makes it very appealing. Existing staff will quickly understand the simple end-point provisioning and the ease of troubleshooting a much less complex network that inherently supports Layer 2 and Layer 3 virtualization. SPB provides all the benefits of a carrier-class network without all the overhead, complexity, or cost, it's simple and scalable.

2.2 Time to Service Improvements

With server virtualization come the feature / requirement to move server instances from one physical device to another. This flexibility now allows the server instance to move within a data center or between data centers.

The easiness of moving a server instance from one physical server to another physical server puts additional requirements on the network infrastructure. The move of a server instance will be transparent to the rest of the network. The physical addressing is kept intact and moved to the new location. For the network piece, it's all about how quickly the same IP subnet can be extended and made available on a different location in the enterprise. Traditionally extending VLANs and IP subnets across a network infrastructure required careful planning and was not an instantaneous job.

When there are only a few VLANs requiring this functionality, it may not be such a daunting task, however, as the number of VLANs grow, the number of services grow, and as traffic separation through network virtualization becomes commonplace, the task suddenly is not so simple and straightforward and definitely requires more work and more attention.

SPB helps to reduce the time to service by as much as 90% for the network connectivity that is supporting the application virtualization. The VLAN and VRF extension capabilities and its end-point-provisioning improve time-to-service drastically compared to legacy technologies. The built-in features of network virtualization also reduce the time to service for creating the virtualized domains needed to maintain the traffic separation between different enterprise functions and/or organizations.

2.3 Robustness

As more services are converged onto the enterprise network, the mission critical nature grows exponentially. Many enterprises are global in nature and therefore access to networks, applications and services are truly required 24x365. It is a common assumption that the network will always be there and available. It has become yet another utility that users take for granted. With this being today's reality, it is imperative to maximum network availability. Network down time almost always results in lost revenue. The network design and the underlying technologies must ensure uninterrupted access to business critical services.

New technologies must provide enhanced capabilities in order to become accepted and utilized. One critical capability of SPB can be seen with an increase in the robustness of the network. In essence, SPB can add another 9 to the enterprise availability. Striving for five 9's availability (which equates to less than five minutes of unplanned downtime per year) has been the goal of every network design and implementation. The deployment of SPB with its robust link state protocol, its sub-second end-to-end network restoration and its end-user MAC encapsulation provides significant network availability improvements and can add another '9' to over availability. By moving to a single protocol and not using legacy technologies such as spanning tree based Layer 2 VLAN transport solutions, SPB simplifies and adds another level of resiliency.

2.4 Predictable Network Behavior

Today's layered approach for network protocols inherently creates dependencies of upper layer protocols on lower layer protocols. In some cases, protocols rely on each other for proper operation. A multicast routing protocol relies on the underlying unicast routing protocol for route and path information. In other cases, the protocols operate independently between systems on their layer, but are reliant on the availability of the lower layers. In a Spanning Tree network, a higher layer unicast routing protocol only re-establishes communication after the lower layer (Spanning Tree) has converged.

In all scenarios, the convergence time of all the protocols on the network will vary. Unicast and Multicast protocols have different convergence times. Spanning Tree convergence times vary depending on what fails and where it fails in the network. This makes it very difficult, if not impossible, to have any type of predictability in the network when changes or failures occur. The more protocols running, the more unpredictability exists.

SPB with IP/SPB provides an integrated model where Layer 2 as well as Layer 3 functionality is provided by one protocol, thus network behavior is very predictable. SPB has eliminated the need to run any form of Spanning

Tree, Layer 3 Unicast or Layer 3 Multicast routing in the core of the network, thus increasing efficiency, reducing complexity, and providing predictability.

2.5 Reduce Operational Expenses

Traditional network technologies intertwine the provisioning of connectivity services with the infrastructure that has been put in place to provide the service. This deep interlocking of service with infrastructure as well as the multilayer approach causes complexity to manage today's networks.

In order to scale networks and move to a model that is more like the service provider / carrier, the network infrastructure must be decoupled from the connectivity services. The network must also be able to provide new functionality such as traffic separation, extension of Layer 2 and Layer 3 capabilities, and still be easy to deploy and manage. SPB provides clear separation of the connectivity services layer and the infrastructure. It also uses only one control plane protocol for Layer 2 VLAN and Layer 3 VRF extension services and therefore provides a significantly simplified operational model. This translates directly into a significant reduction in operational expenses.

3. Networking Issues with Today's Technologies

Many issues are commonly seen in today's networks when examining the existing technologies and their challenges to fit the requirements of the virtualized environment. Each of these issues must be reviewed and understood as the enterprise moves forward to tie the application virtualization with the impending network virtualization. The ease of mapping these two together will result in not only the overall success of the enterprise network, but also in the long term scalability and total cost of ownership.

3.1 Service and Infrastructure Separation

A quandary of sorts exists in today's traditional network infrastructure. The difficulty lies in the balance of two seemingly opposing forces; building a network as simple as possible to keep operations cost down while implementing the many different connectivity requirements on that same network. With the existing technologies, more connectivity requirements equal more complexity in design, deployment, and operations.

The major problem that complicates network operations significantly is the limited abstraction of a "network connectivity service" from the infrastructure. The connectivity services and infrastructure configurations are tightly coupled and cannot easily be separated from each other. To illustrate this, a common requirement in the enterprise network is to extend a Layer 2 domain between different end-points. This could be between floors in a building or between buildings on a campus or between data centers across the country. For example, if a VLAN (100) is required to connect two service access points (the point in the network where a server, PC, or other end user device is plugged in), the complete path between these access ports throughout the network, including a redundant path(s), needs to be provisioned with VLAN (100). Even though there are only two service access points (SAP), there will likely be numerous ports, switches, routers throughout the path of the infrastructure that needs to be properly provisioned to accommodate connectivity between those two SAPs.

This simple example shows the difficulty and complexity in design and provisioning when the service and infrastructure are tightly coupled. Many different touch points exist in the network to make this simple Layer 2 extension happen. Ideally, only each service access point should need to be provisioned, thus reducing the configuration down to two places in this example, leaving the entire intermediate touch points to automatically create the shortest path between the SAPs.

The success of the IP protocol can be attributed to this fact; a new IP subnet has to be added only at the service access point, the end-to-end connectivity is established "automatically" by the IP routing protocol. SPB with IP/SPB achieves a similar experience by providing end-point-only provisioning.

3.2 Network Virtualization

3.2.1 Large Layer 2 Domains

Since the mid 1990s (IEEE standardized 802.1Q in 1998) VLAN tagging is the predominant way of virtualizing enterprise networks. With IP being the transport protocol, a one-to-one mapping between VLAN and IP subnet has been established as the defacto standard design. In some cases exceptions exist where business requirements or migrations require features such as IP multi-netting, where multiple IP subnets exist on a single VLAN.

Most applications have used the VLAN concept only to segment into IP subnets to reduce the broadcast domains, improve performance, and ease troubleshooting. Some network applications are looking for true segmentation to restrict access between communication domains. In small topologies, VLANs can fulfill this segmentation requirement nicely, but in larger domains, true segmentation is achieved by virtualizing the Layer 3 domain as well.

Layer 3 device virtualization is achieved by using multiple routing instances known as Virtual Route Forwarder (VRF) technology. With this technology, VLANs are mapped to different VRF instances which in turn create traffic separation. The next step is to achieve network virtualization by building an IP VPN and extending VRFs across the network.

3.2.2 VRF Extensions

To extend VRFs across a multi-hop network, VRF-extension technology is required. In smaller scenarios this is achieved by running a separate routing protocol for each routing instance (i.e. VRF). This may work for a few network links and a few VRF instances, but the larger the topology the more complex it gets to manage such a network since all switches and all links will have multiple routing instances running in parallel. Presently, carriers have deployed BGP and MPLS-based IP VPN services to accommodate the needs of their enterprise customers. Extending these technologies into the campus and data center can be complex and requires additional skill sets not present in many enterprise network operations teams.

3.2.3 Complexity of MPLS and MP-BGP (RFC 4364)

In networks where hundreds to thousands of customer instances are required, BGP has proven to be a robust protocol to achieve the scalability that is required. BGP is used to carry all virtualized routing tables across TCP sessions between virtualized Provider Edge routers (PEs).

MPLS virtualizes the fast path and thus separates each user domain, by identifying it with a different (service) label. MPLS tunnel labels are swapped at each hop throughout the network and forwarding decisions are made based on local forwarding tables. This makes the management and troubleshooting of the network more tenuous, as each hop must be looked at to figure out the entire path through the network.

This approach to network design works very well for the carrier / service provider as it is very scalable and provides the required functionality. For the typical enterprise customer, the many layers of protocols make the solution quite complex to design, deploy, and operate.

The key point is to provide traffic separation, required scalability, and the ability to troubleshoot without multiple layers of protocols or specialized network expertise.

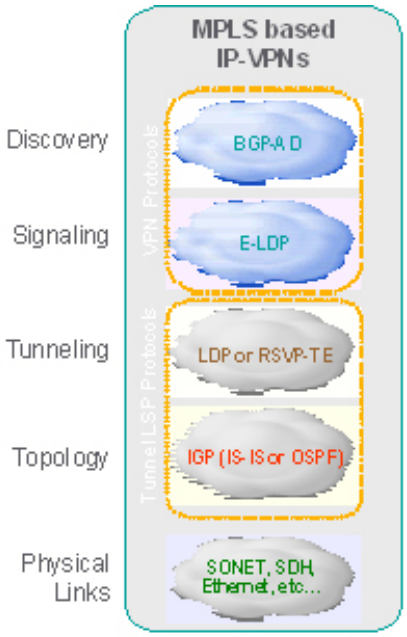


Figure 1 - MPLS Protocol Layers

3.3 Bridged Domain Issues

As discussed earlier, server virtualization technologies have increased the need for Layer 2 VLAN extensions between SAPs where servers are physically connected. These extensions either within a data center or between data centers have become table stakes requirements.

Whenever VLANs are extended across the core of a network, questions arise in regard to the impact this can have on the stability of the network and especially the core. Typical questions include:

- What is the impact of MAC learning on the core?
- Is there the possibility of loops negatively affecting the network?
- If spanning tree is used for Layer 2 redundancy across the core, what about all these blocked and unused links?

3.3.1 MAC Explosion

MAC table size exhaustion in enterprise networks is usually not an issue. The core systems normally scale well beyond the typical number of MACs within an enterprise. What can become an issue in large spanning tree deployments are excessive topology change notification (TCN) messages. TCNs are sent whenever there is a spanning tree port state change. Each TCN causes all bridges in a STG network to reduce their MAC aging timers, thus traffic is being flooded and paths are re-learnt each time a TCN is sent in its Spanning Tree Group.

The other issue that may arise in large scaled networks is the time it takes a core system to re-learn all the MACs during a recovery. When a failed system is restored to service, the MACs must all be learned, in some cases this can negatively impact recovery time.

Reducing TCN generation and MAC table sizes are good practices when designing bridged topologies.

3.3.2 Loop Sensitivity

VLAN-based Layer 2 networks are sensitive to loops. Network loops can occur due to many different reasons, but the effect is always the same. Broadcast and Multicast traffic looping in the network will quickly exhaust network bandwidth in a matter of seconds. Even worse, the looping traffic triggers the bridges to re-learn the end station MAC addresses from different ports (normal path and looped path). This in turn triggers all devices to be unreachable and at the same time the bridge control planes to constantly update the systems bridge forwarding / filtering tables.

Shielding the network core from the end-station MAC Addresses would make the network core much less sensitive to network loops.

3.3.3 Blocked and Unused Links

A disadvantage of a Spanning Tree based network is that traffic in a VLAN can only travel along one path. This path is the Spanning Tree which is forming the loop-free topology. The result of creating the loop-free topology renders several links blocked and therefore unused in normal operations. These blocked links are only used during times of failure in that particular portion of the network. Enterprises are paying a 100% premium for links that may only be used 1% of the time – not a great use of valuable CAPEX.

Spanning Tree is root bridge-based, thus traffic will have to flow potentially from the tree leaf up to the root bridge to reach another leaf in the network. The actual paths through the network can be engineered by setting up multiple Spanning Tree groups, but this once again adds another layer of complexity during design, deployment, and operations.

The optimal model is one that makes efficient use of all network links while still providing resiliency and fast failover. It must also provide optimal (shortest) paths between end points to increase performance and minimize latency.

3.4 Storage and Data Center Bridging

Fibre Channel transport over Ethernet is presently a popular topic in enterprise data center environments. The T11 working group has defined FCoE as a technology to replace native Fibre Channel with Ethernet. The long term vision will have storage traffic as well as user bound LAN traffic converge on one technology. This will reduce both CAPEX and OPEX with a reduction in the number of networks, network adapters, and management.

In order to properly support storage traffic running on Ethernet, new enhancements must be added to address existing deficiencies:

- Today's Ethernet is not lossless

The IEEE is addressing this with the following standards work in the data center bridging working group:

- P802.1Qau: IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks - Amendment 10: Congestion Notification.
- P802.1Qaz: IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks - Amendment: Enhanced Transmission Selection.
- P802.1Qbb: IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks - Amendment: Priority-based Flow Control.
- Today's Ethernet does not provide shortest path switching

As previously discussed, Spanning Tree-based Layer 2 topologies use root tree based forwarding. This results in suboptimal packet forwarding and traffic rarely uses the shortest path to reach its destination. Storage traffic however requires low latency and minimal forwarding hops, thus benefits greatly from a solution that can provide shortest path switching.

4. SPB Solution Details

4.1 Virtualization Standards Evolution

The IEEE has been working on Layer 2 virtualization techniques over the last decade. It had standardized a set of solutions that built on each other and continuously addressed the predecessor's disadvantages.

Standard	Year	Name	Loop free by using:	Service ID's	Provisioning	Virtualization of:
IEEE 802.1Q	1998	Virtual LANs (VLAN Tagging)	Spanning Tree or Switch Clustering	4096	Edge and Core	Layer 2
IEEE 802.1ad	2005	Provider Bridging (Q-in-Q)	Spanning Tree or Switch Clustering	4096x4096	Edge and Core	Layer 2
IEEE 802.1ah	2008	Provider Backbone Bridging MAC-in-MAC	Spanning Tree or Switch Clustering	16 million	Edge and Core	Layer 2

In 1998, IEEE 802.1Q provided a simple way to virtualize Layer 2 broadcast domains by using VLAN tagging to form Virtual LANs. The 12 bits that are available in the 802.1Q defined header provided the ability to separately transport 4096 individual virtual LANs.

The loop free topology had been provided through IEEE 802.1D Spanning Tree and later Rapid Spanning Tree (RSTP) and Multiple Spanning Tree (MSTP) extensions. However, spanning tree is not the technology of choice for large carrier deployments or data center deployments – for details see Section 0.

Avaya introduced Switch Clustering using Split Multi-Link Trunking as a better alternative to spanning tree. Switch Clustering's built-in ability to build large loop free topologies that provided active/active resiliency with sub-second failover proved to be a far superior technology in comparison to all the Spanning Tree options.

Carrier deployments wanted to leverage the cost points of Ethernet and wanted to use the virtual LAN technology. In order to improve scalability, the IEEE introduced the Q-in-Q approach, where the header had been extended to provide a carrier tag attached to a customer tag Q-in-Q. This allowed the carrier to transport customer tagged traffic over its Ethernet based 802.1ad backbone. However in large deployments this technology did not scale well, because the carrier network "saw" the customer MAC addresses. This restricted the carriers in providing a truly robust network solution. For details please refer to Section 0.

In order to overcome this scaling limitation, the IEEE standardized 802.1ah in 2008 which introduced a new header encapsulation to hide the customer MAC Addresses in a backbone MAC Address pair.

In addition to this, the new header also includes a service instance identifier (I-SID) with a length of 24 bits. This I-SID can be used to identify any virtualized traffic across an 802.1ah encapsulated frame. In 802.1ah, these I-SIDs are used to virtualize VLANs across an I-SID-based MAC-in-MAC network. The "hiding/encapsulating" of customer MAC Addresses in backbone MAC Addresses greatly improves network scalability (no end-user MAC

learning required in backbone) and also significantly improves network robustness. This is due to the fact that any customer introduced network loops have no effect on the backbone infrastructure.

4.2 The SPB Model

A recognition of the existing limitations for network virtualization led to the development of a new link-stated based technology known as Provider Link State Bridging (PLSB). Based on IS-IS 2008 and a natural evolution of PBB/PBT, PLSB addressed the growing needs in regard to network virtualization. This technology was introduced into the IEEE standards body and now known as 802.1aq Shortest Path Bridging MAC-in-MAC (SPBM).

Standard	Year	Name	Loop free by using:	Service ID's	Provisioning	Virtualization of:
IEEE 802.1aq	Expected in 2010	Shortest Path Bridging (SPBM)	Link State protocol (IS-IS)	16 million	Only service access points	IEEE: Layer 2 IETF draft: Layer 3 Unicast & Multicast

SPBM is based on the 802.1ah encapsulation schema that does not depend on Spanning Tree to provide a loop free Layer 2 domain, but instead uses the nodal based IS-IS topology protocol. The IEEE is reworking the Spanning Tree specifications 802.1D to include the new SPB solution. The intention is that once the standard is implemented in network products, the network operator will be able to choose a shortest path bridging topology protocol or the legacy root tree-based option.

One of the key advantages of the SPBM protocol is the fact that network virtualization provisioning is achieved by just configuring the edge of the network (service access points), thus the intrusive core provisioning that other Layer 2 virtualization technologies require (including 802.1Q – VLAN tagging) is not needed when new connectivity services are added to an SPBM network. For example, when new virtual server instances are created and need their own VLAN instance, they are provisioned at the network edge only and don't need to be configured throughout the rest of the network infrastructure.

This "edge-only" provisioning model provides a far faster time-to-service on the network side compared to the traditional edge and core provisioning. This is key in order for the network to match the speed improvement of new service instantiations (applications) on virtualized servers.

In addition to the Layer 2 virtualization support that SPB provides, the model is being extended to also support Layer 3 virtualization. A more detailed discussion is provided later in this document in section 0.

The boundary between the MAC-in-MAC SPB domain and 802.1Q domain is handled by the Backbone Edge Bridges (BEBs). At the BEBs, VLANs are mapped into I-SIDs based on the local service provisioning. Redundant connectivity between the VLAN domain and the SPB infrastructure is achieved by operating two SPB switches in Switch Clustering (SMLT).mode. This allows the dual homing of any traditional link aggregation capable device into a SPB network. IEEE also introduced a seamless redundant connection between SPB and Spanning Tree domains as part of the combined Spanning Tree / SPB standard.

4.3 Frame format

SPBM's frame format is based on the packet header that is described in IEEE 802.1ah.

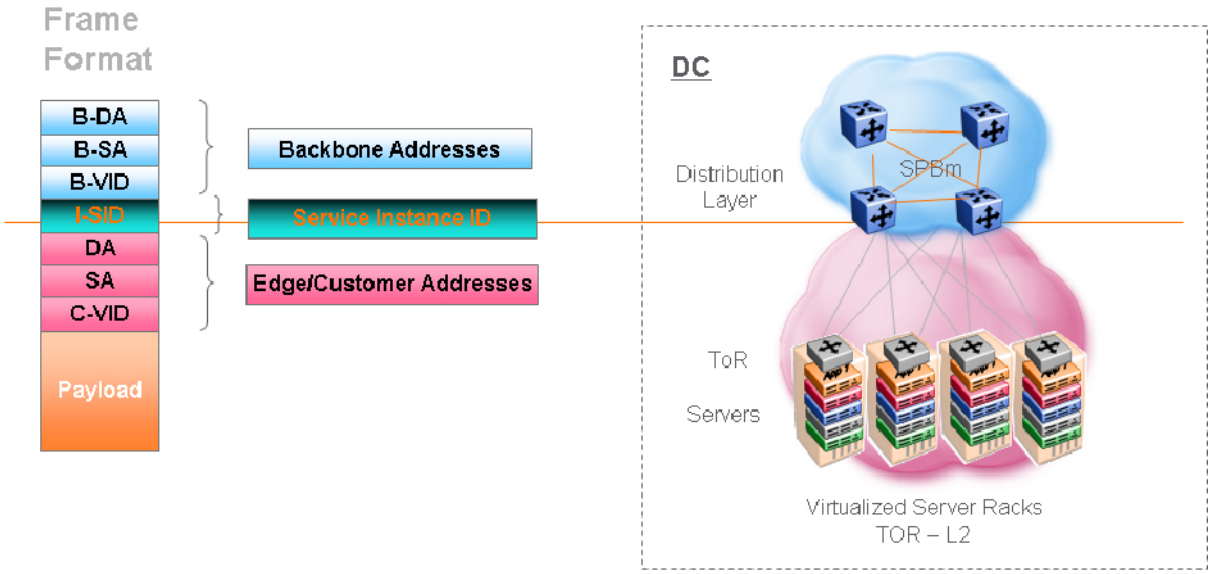


Figure 2 - 802.1ah Frame Format

The blue backbone encapsulation is "hiding" the edge/customer MAC Addresses, thus improving network stability as discussed previously in section 3.3.

4.4 Protocol Infrastructure

Example network topology to run SPB protocol:

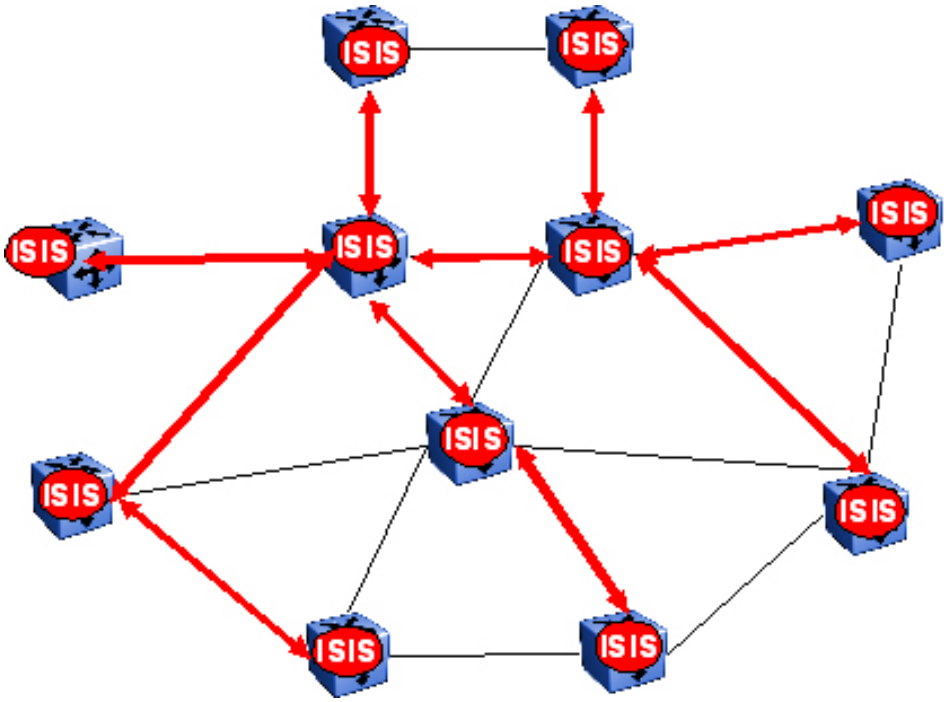


Figure 3 - SPB Topology Discovery

1. Discover network topology

Before calculating the shortest path trees, the network topology needs to be discovered. IS-IS, a natural Layer 2 routing protocol, runs on all nodes of the SPB domain. Through the link-based IS-IS protocol, session topology information is exchanged (similar to OSPF). Each node has a node ID which is used in the topology announcement. Also, each node has one Backbone MAC address (BMAC) which is used as source- respectively destination MAC Address to send traffic to this node across an SPB network.

2. IS-IS nodes automatically build trees from itself to all nodes

As soon as the network topology is discovered and stored in the IS-IS link state database (LSDB), each node calculates shortest path trees based on preconfigured link-metrics for each source node.

Important Properties

- Shortest path tree based on link metrics
- No blocked links
- RPFC to eliminate loops
- Symmetric data path between any two nodes provides closed OAM system
- Unicast path now exists from every node to every other node

3. **Uses IS-IS to advertise new service communities of interest**

When a new service is provisioned, its membership is flooded throughout the topology with an IS-IS advertisement.

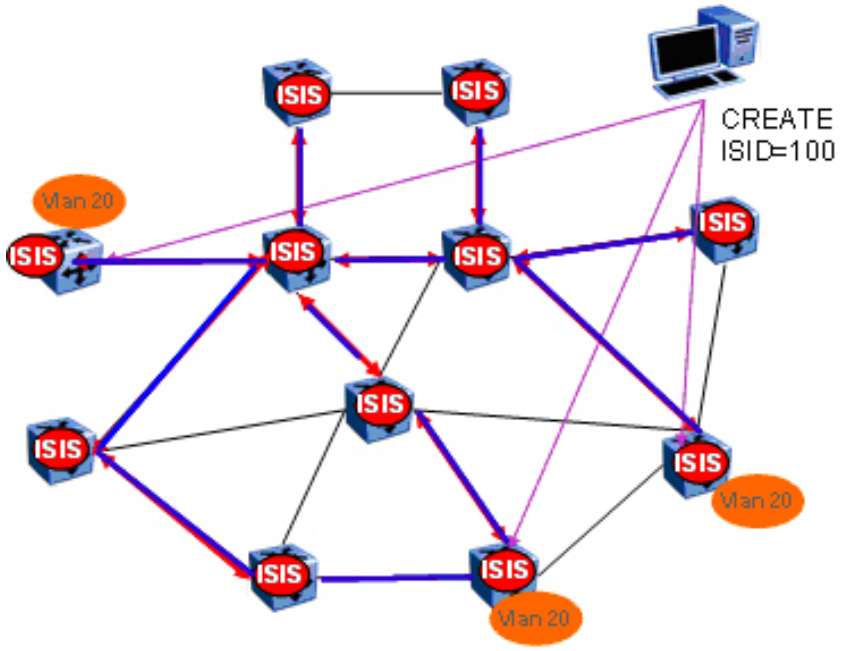


Figure 4 - SPB ISID advertisement

BMAC and ISID information is flooded throughout the network to announce new ISID memberships. In this case VLAN 20 is mapped to ISID 100.

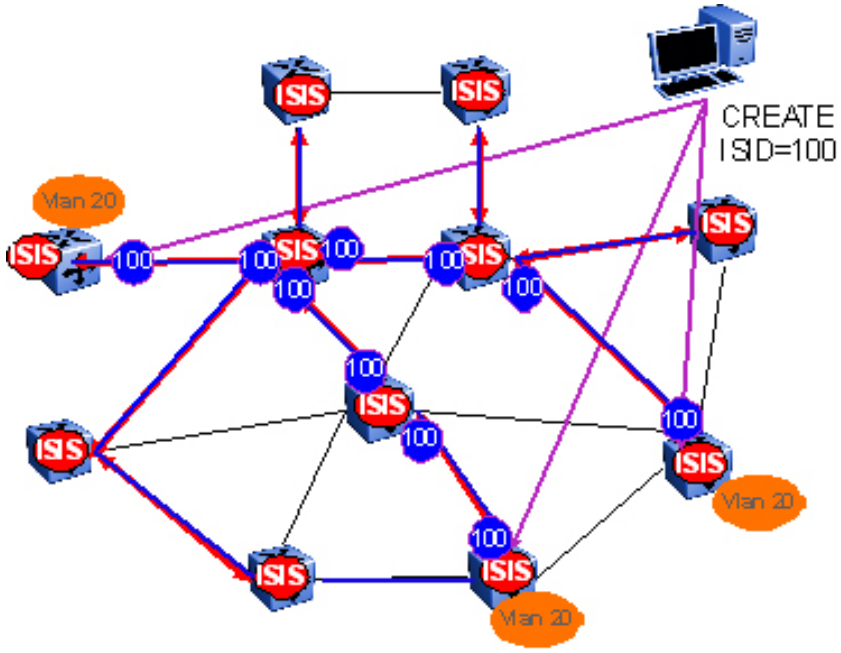


Figure 5 - SPB BMAC/ISID population

Each node populates its FDB with the BMAC information derived from the IS-IS shortest path tree calculations. Thus there is no traditional flooding and learning mechanism in place for the BVLAN, but FDBs are only programmed by the IS-IS protocol.

4. When nodes receive notice of a new service AND they are on the shortest path, update FDB

In this scenario, where there are three source nodes having a membership on ISID 100, there are three shortest path trees calculated (not counting the Equal Cost Trees (ECTs)). The following diagrams depict the traffic flow for this formed ELAN.

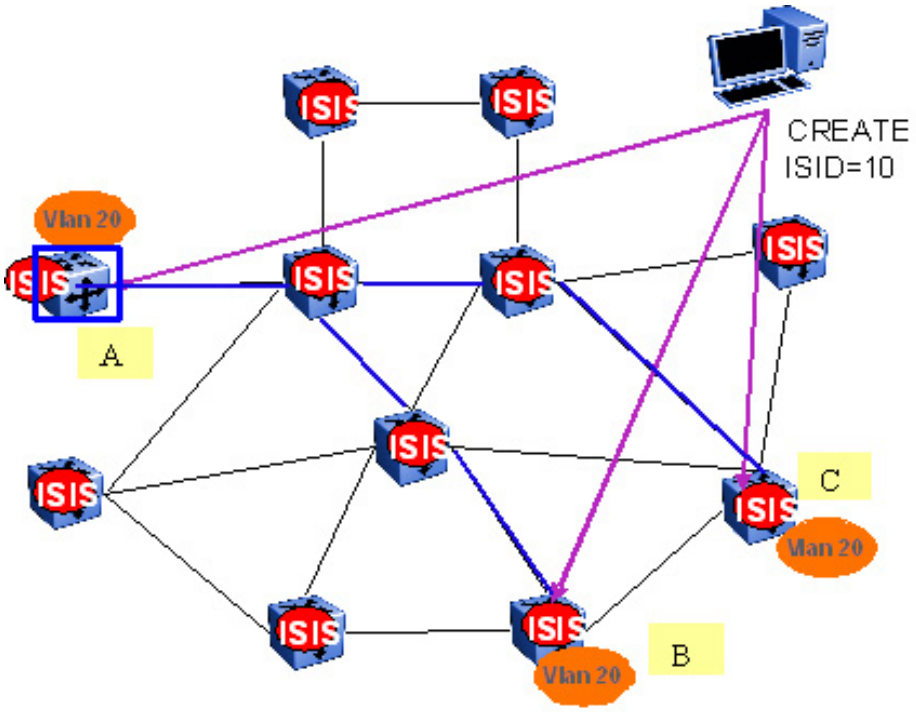


Figure 6 - Shortest Path Tree for Source Node A

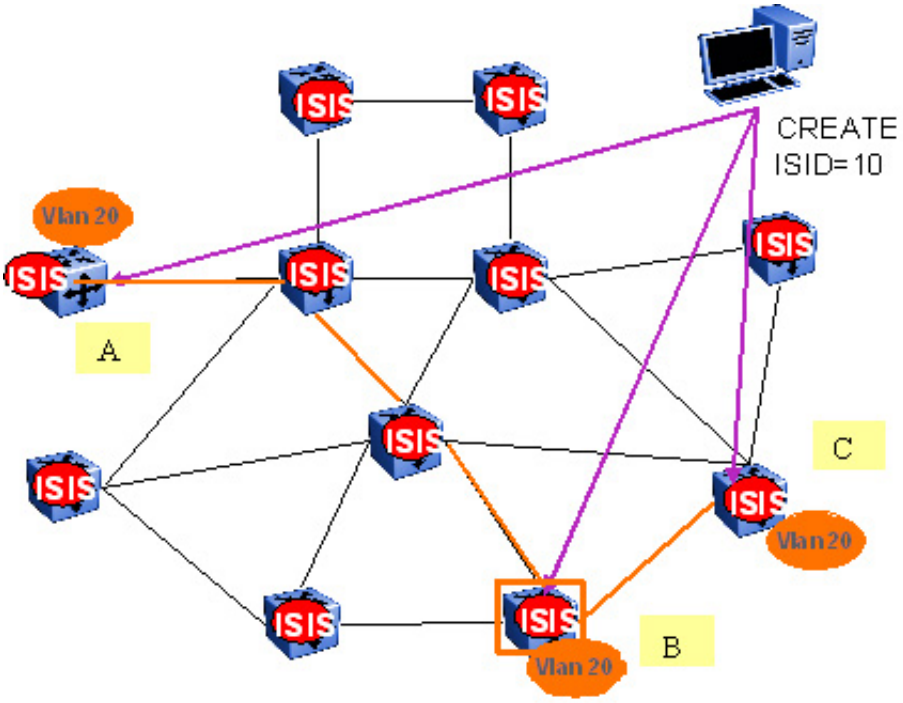


Figure 7 - Shortest Path Tree for Source Node B

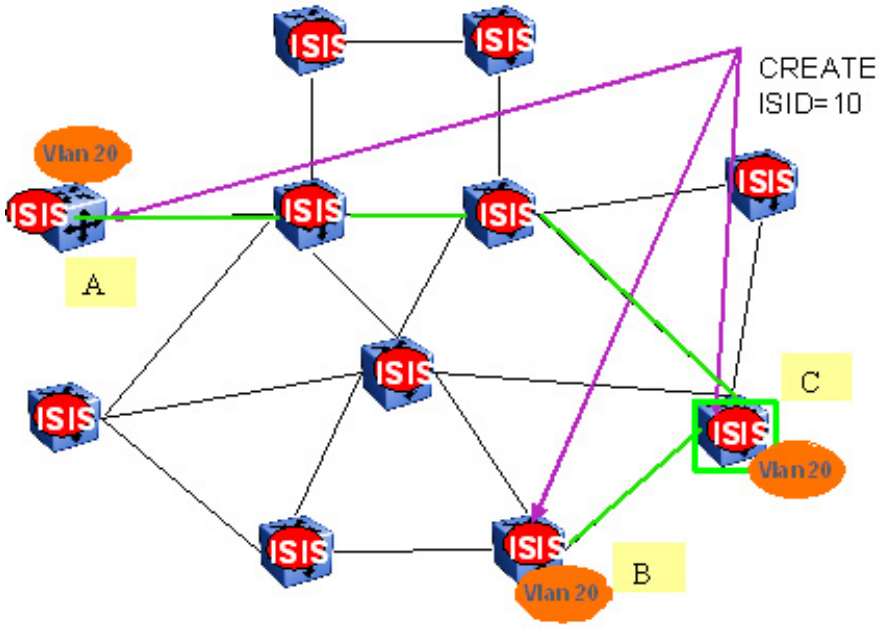


Figure 8 - Shortest Path Tree for Source Node C

The paths between any two nodes are always the shortest paths. Also, the paths in either direction are congruent, thus a bidirectional communication stream can be monitored easily by mirroring ingress and egress on a link to a network analyzer. (As a comparison, the same is true for Spanning Tree based networks, but not true for TRILL-based networks).

VLAN traffic arriving on switch A and VLAN 20 is forwarded following the blue path, traffic arriving on switch B and VLAN 20 the orange path and on switch C VLAN 20 traffic is following the green path. If the destination CMAC is unknown at the SPB ingress node or the traffic is of type broadcast or multicast, then it is flooded to all members of the topology which spans VLAN 20. If the destination CMAC is already known, then the traffic is only forwarded as a unicast to the appropriate destination. In the SPB domain, the traffic is switched on the BMAC header only. The bridge filtering database (FDB) at the VLAN to ISID boundary (backbone edge bridge BEB), maintains a mapping between CMACs and corresponding BMACs. E.g. Switch B learns all CMACs which are on VLAN 20 connected to switch A with the BMAC of A in its FDB and the CMACs which are behind C are learnt with the BMAC of C.

4.5 Service Layer QoS

Quality of Service (QoS) is maintained in a SPB network the same way any IEEE based 802.1Q network is operated. Traffic ingressing a SPB domain which is either already 802.1p bit marked (CMAC), or is being marked by an ingress policy (remarking), is getting its BMAC p-bits marked to the appropriate value. The traffic in a SPB core is scheduled, prioritized and forwarded according to the 802.1p values. In the case where traffic is being routed at any of the SPB nodes, the IP DSCP values are taken into account as well.

Future enhancements to SPB will allow explicit paths through a SPB domain to be predefined. This provides the ability to setup traffic engineered paths through a network, avoiding congested nodes.

4.6 Forwarding Behavior and Security

The following example illustrates the differences between SPB's forwarding behavior and that of MPLS.

Traffic ingressing a SPB domain is forwarded across the backbone by using the Destination BMAC as the tunnel label. The I-SID is then used at the egress node to define which virtualization entity this particular flow belongs to (VLAN, VR, IP MC stream).

In contrary to MPLS, neither BMAC nor I-SIDs are changed throughout the journey of the packet through the backbone. SPB's advantage here is that the intermediate nodes don't need to stitch two identifiers together, which MPLS requires on each node due to its LSP swapping technique. In a potential MPLS Label Distribution Protocol (LDP) error potentially two LSPs are connected together, which don't belong together, as a result this could lead to a security leak in MPLS.

Both technologies employ the use of a tunnel and a service label to separate flows from each other.

4.7 Layer 2 VLAN Extensions

The main application of the standards based SPB solution is to provide robust and scalable VLAN extension across an Ethernet switched network infrastructure. Its shortest path bridging capabilities makes it a powerful replacement for today's spanning tree based solutions. As earlier pointed out, the great improvement in simplicity is achieved by the service access point provisioning model, thus leaving a network core to be a true core without having to worry about new connectivity service being added at the edge.

4.8 Failure and Recovery

Link and nodal failure in a SPB network trigger an IS-IS link state update. Only the nodes that are part of the network affected are recalculating the topology and update their forwarding entries. In typical enterprise networks, a restoration after a failure can be expected within one second to occur. Key to this fast restoration is that a link failure can be detected as quickly as possible. For this either the IEEE based link failure detection 802.1ag (CFM) should be used or similar detection mechanism such as VLACP.

4.9 Scalability

The solution is built for hundreds of SPB nodes running hundreds of services (-ISIDs) throughout the network. While the standard allows addressing 16 million service entries, the nodal limitation is set to 4000 initially.

4.10 Routing Between Extended VLANs

A functionality which is commonly used in traditional 802.1Q environments is the ability to route traffic between VLANs. This capability is also provided in a SPB environment by enabling InterR-ISID routing. This allows the network to use SPB nodes as default gateways/routers for extended VLANs without having to terminate the I-SID at an edge node. This is particularly interesting in a data center deployment where the top of the rack devices are also SPB capable, but are purely Layer 2 devices. In this scenario, the first routing hop is provided at the aggregation layer which exists deeper into the network.

4.11 SPB and Edge Connectivity with Switch Clustering

As earlier described, the boundary between the MAC-in-MAC SPB domain and 802.1Q domain is handled by the Backbone Edge Bridges (BEBs). At the BEBs, VLANs are mapped into I-SIDs based on the local service provisioning. Redundant connectivity between the VLAN domain and the SPB infrastructure is achieved by operating two SPB switches in Switch Clustering (SMLT) mode. This allows dual homing of any traditional link aggregation capable device into a SPB network.

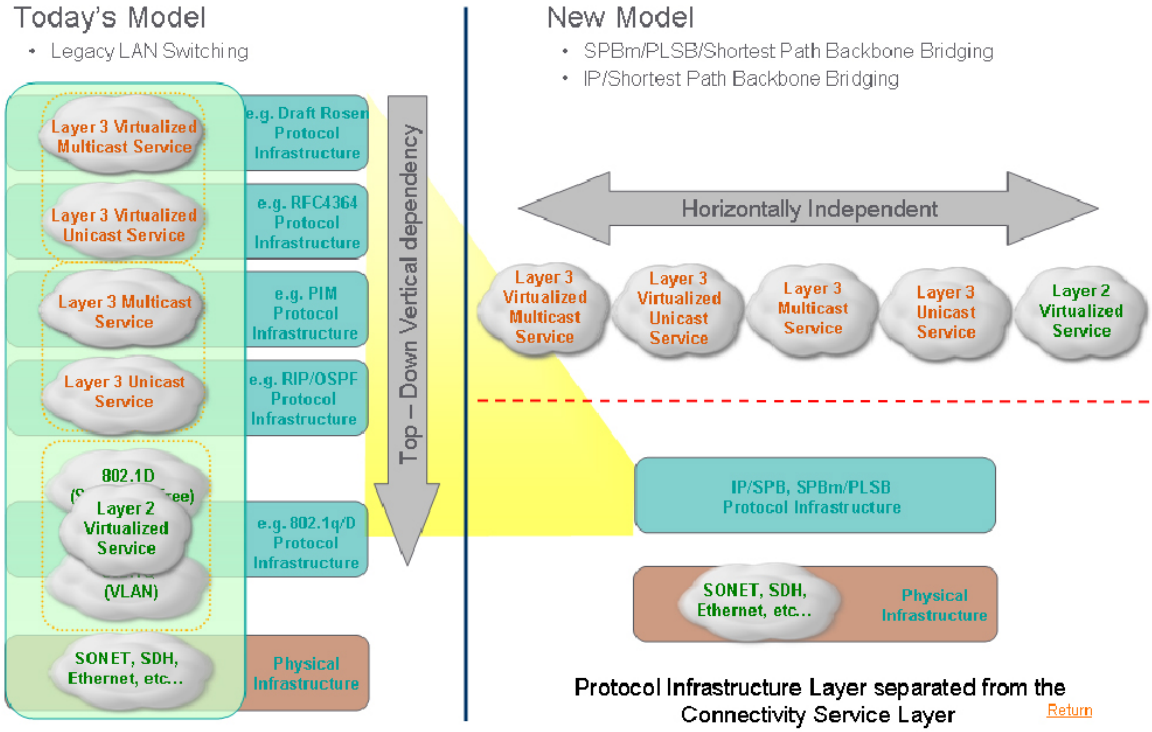
4.12 Layer 3 VRF Extensions

Whether it is an airport authority supporting multiple airlines on its infrastructure or a government IT department in charge of supporting various clients (such as administration, police, education), they have all in common that they want to provide traffic separation on top of one shared network infrastructure. Typically these deployments start with VRF separation, but in most cases those VRFs need to be extended across the network infrastructure.

MPLS-based IP VPNs can be used to provide Layer 3 virtualization support. However the target applications of MPLS based IP VPNs are large scale carrier deployments, for enterprise network operations teams, the layers of complex protocols are a big hurdle for using it.

Draft IP/SPB-Unbehagen describes an extension to SPB that leverages IS-IS to not only build Layer 2 domains, but also provide a very flexible Layer 3 VRF extension capability. This integrated model approach does not require any additional protocol to support Layer 3 virtualization. Typically Layer 3 VRFs can now be provided at any SPB node in the network in parallel to the Layer 2 VLAN extension solution. IS-IS carries the VRF specific route entries in its link state updates. The I-SID is used in this model to provide VRF separation.

4.13 Network Simplification by Protocol Overlay Reduction



SPB simplifies network operation by removing a set of overlay protocols and collapsing them into one link state based protocol: IS-IS. At the same time, it achieves true separation between infrastructure and service layer by leveraging the service ID concept (I-SID). With the inclusion of IP/SPB into the SPB protocol simplified provisioning and operation for Layer 2 and Layer 2 Unicast and Multicast virtualization is achieved. Due to the service separation by I-SID, there is no dependency of SPB virtualized services among themselves occurring. Compared to the traditional model true OPEX savings can be expected due to the protocol simplification.

5. Value Proposition Summary

SBP enables enterprises to improve the delivery of always-on content and simplify the deployment of the private cloud. It is an open, standards-based approach that offers increased reliability, a reduction in time to service from days/months to minutes, better utilization of network resources (no need for blocked ports), and greatly improved manageability and network uptime when compared to alternative models.

With its applicability to both data center and campus applications, SPB delivers a consistent enterprise-wide model for delivering highly resilient access to applications and services with light-touch provisioning. Customers don't have to interconnect differing forms of technology, one for the data center and others for the campus and MAN/WAN - our architecture is truly end-to-end. And crucially SPB can be deployed over any type of network architecture whether it's a ring, full mesh, square etc and can be enabled in parallel with all other protocols presently in use on the network. There is no need to change physical connections or existing configurations; you can migrate to SBP at your pace in the most non-disruptive manner possible.

The benefits of on Shortest Path Bridging solution:

- Avaya's Enhanced SPB not only supports L2 virtualization (E-Line and E-LAN services) but also L3 virtualization.
- End-to-end secure traffic separation: Unlike legacy VPN models, with SPB traffic is not recombined into shared networks within the core.
- Rapid Time to service: a VLAN/VRF extension across the network can be established in no time due to the simple end point provisioning and automatic connectivity establishment.
- Robust Infrastructure and increased network uptime: The link state based infrastructure protocol provides quick failure recovery.
- Dual-homing: Traditional Switches can be attached to the SPB cloud redundantly using Switch Clustering.
- Separation between service and infrastructure: Adding new infrastructure (links, switches), does not require to have knowledge of the services that run on top of the network, as well as any kind of topology is supported, thus network operations and design flexibility is greatly improved compared to traditional networking.
- Smooth migration from traditional networking to SPB based virtualization for both data center and multi-tenant campus requirements.

In summary, with Avaya's Enhanced SPB, enterprises will have the agility and flexibility that they seek of their networking infrastructure. As the pace of change increases and the desire to embrace new enabling application grows, enterprises will have a truly dynamic networking environment, free of the old-world constraints and limitations.

6. Deployment Scenarios

6.1 The Virtualized Data Center

6.1.1 The Requirement

This section depicts a typical larger enterprise network with a backbone (core) and multiple (two) datacenters. The rest of the network (user aggregation) is omitted to simplify the drawings. In this scenario the datacenters are being virtualized with compute virtualization technology such as VMware's ESX infrastructure or a similar technology (e.g. Microsoft Hypervisor). One of the most used functionalities of virtualized server environments is the capability of moving virtual server instances from physical server to another physical server. These servers can reside locally in the same data center, or they can be on a remote location across the backbone infrastructure. In both cases, in order to move and access the server instances the IP subnet, where the server reside on, needs to be extended to the new physical location. This is because the servers retain their IP address even after a move to a new location. Since there is a "one-to-one" mapping between IP subnet and VLAN, the VLAN (broadcast domain) needs to be extended to the new location server location.

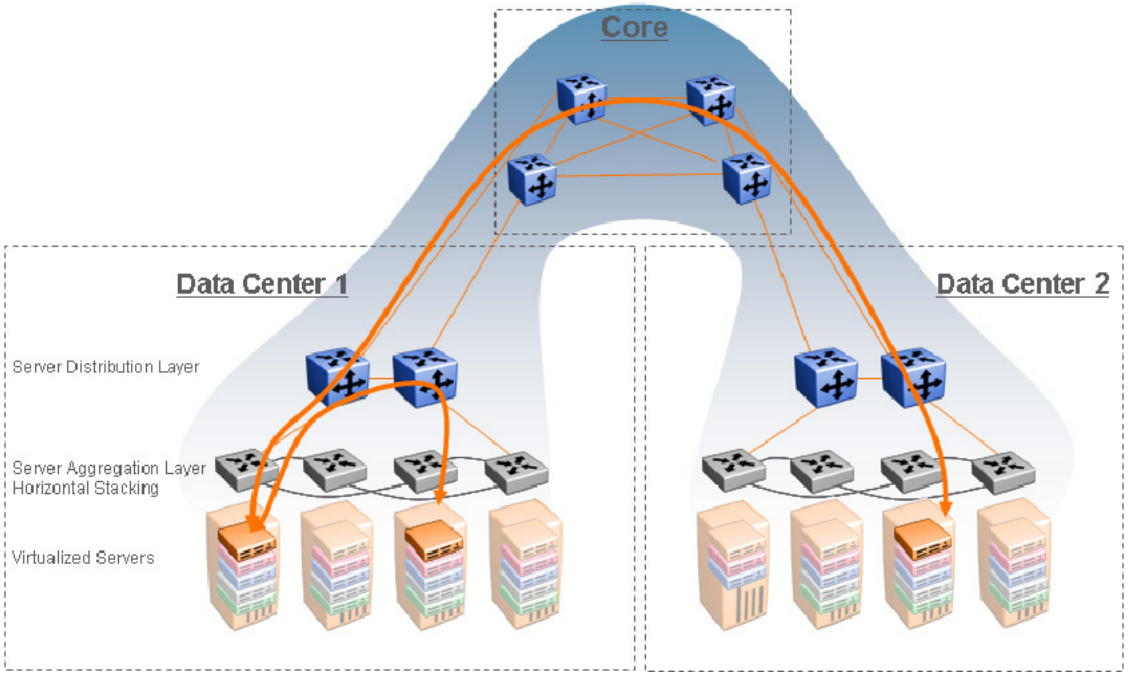


Figure 9 - Data Center Virtualization

This figure depicts the extended VLAN/IP subnet (broadcast domain).

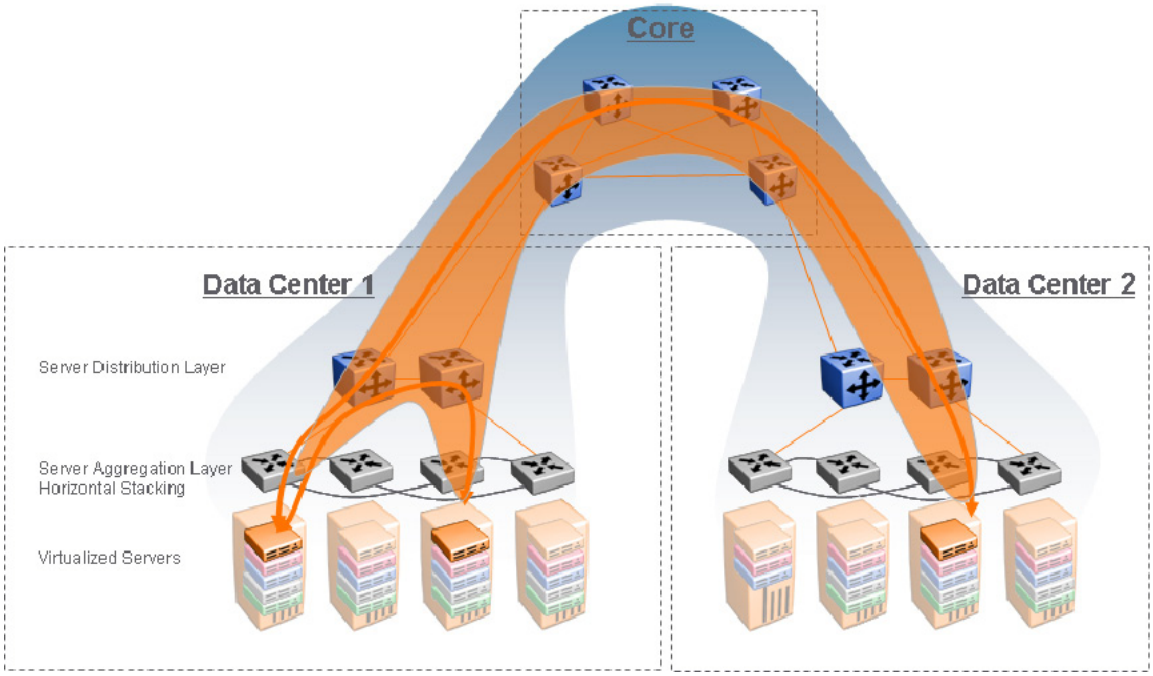


Figure 10 - Extended VLAN

6.1.2 The Solution

The following section explains in a few steps how SPB can be leveraged to enable server move and it also outlines the great benefits SPB provides.

The following picture shows the network infrastructure as well as the virtualized servers, which are connected to ToR switches. The ToR devices can be built as a Horizontal-Stack (HZ), or they can be dual-homed (switch clustering) to the data center server aggregation switches. The SPB infrastructure has been put in place and spans all routing switches. In this example the virtual servers which are to be moved from one physical server to another are on VLAN10 which corresponds to IP subnet 10.10.10.0.

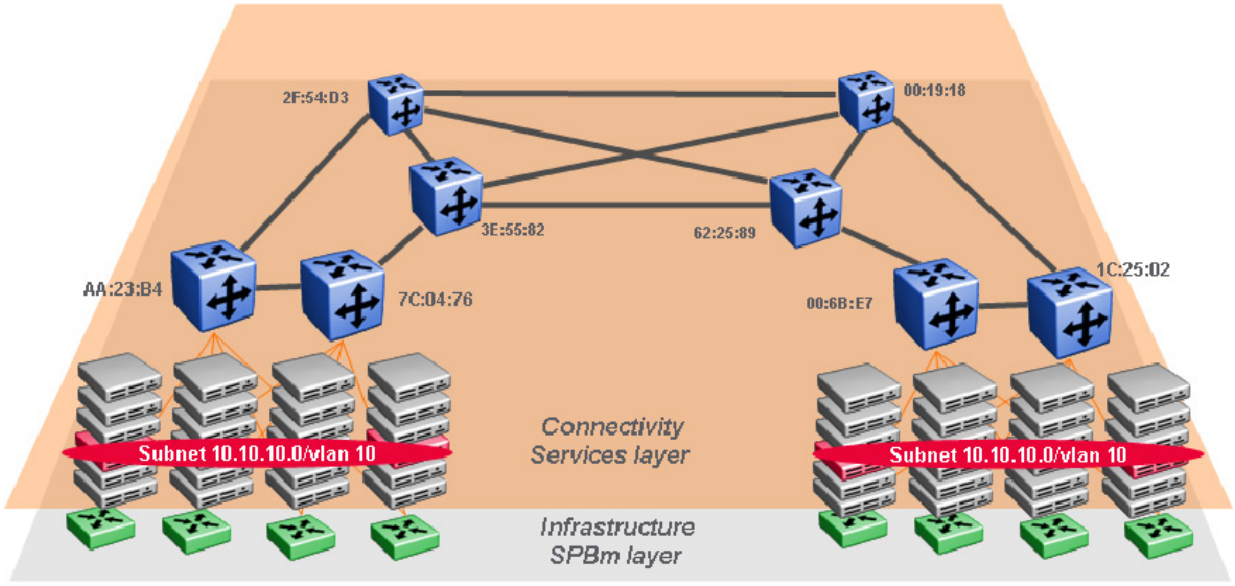


Figure 11 - Data Center Infrastructure and Service Layer

In order to be able to move the virtual server transparently from one physical server to another physical server, the VLAN/IP subnet needs to span across the network.

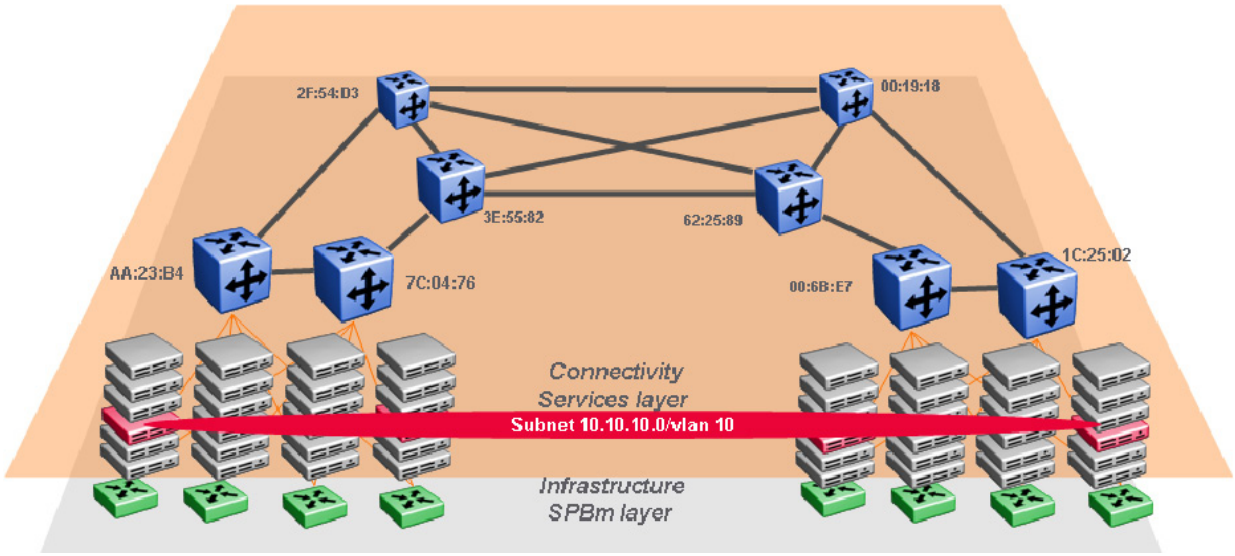


Figure 12 - Extended Subnet

Traditionally this would have been done by configuring a VLAN across the network. Due to the earlier discussed issues, a more robust solution which fulfills the additional requirements should be chosen.

SPB allows spanning VLAN10 from one location to another by binding the VLAN10 at the first SPB-capable node to an I-SID (I-SID 100). The network then automatically connects all the VLAN access points together using ISID 100. The network administrator only configures the service access points.

Dual-homing of user VLANs is also supported using the switch clustering technology.

Integration of SPB into existing networks

It is important to note, that a SPB protocol infrastructure can be put in place without having to change the existing protocol infrastructure. Most Ethernet based networks use 802.1Q tagged interfaces between the routing switches. SPB requires a couple of Backbone VLANs (BVLANs) which are used as the transport instance. A BVLAN is not a tranditional VLAN in the sense that it does not flood Unknown and Broadcast/Multicast traffic, but only forwards based on IS-IS provisioned backbone MAC tables. Once the BVLANs are configured and the IS-IS protocol is operational the services can be mapped to service instances. For migration purposes the user/server VLANs can be moved from the traditional transport to the SPB based transport, one VLAN at the time, enabling a smooth migration.

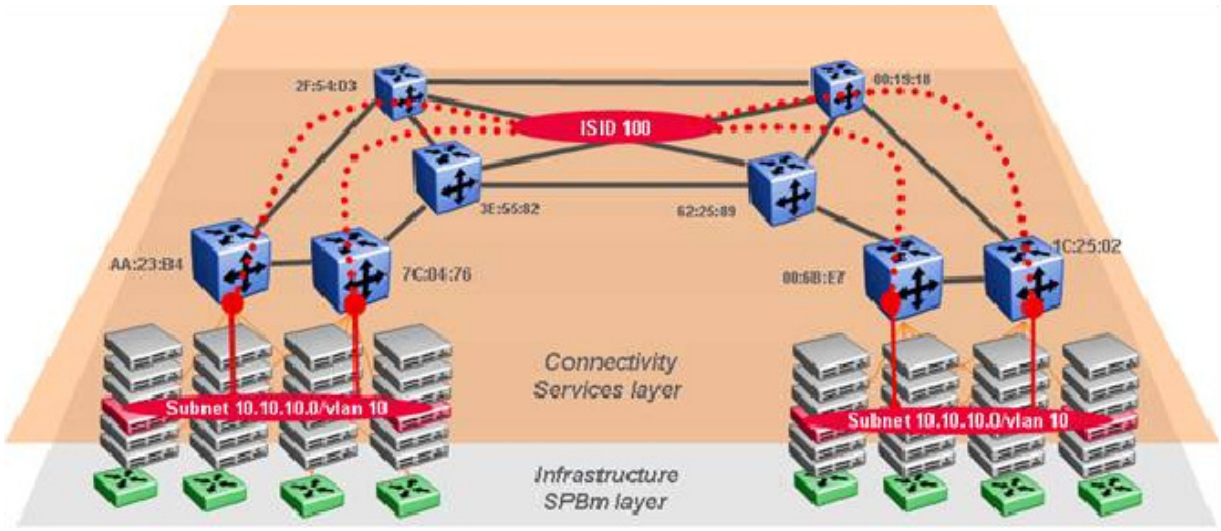


Figure 13 - SBPm VLAN to ISID Mappings

In most scenarios, those IP subnets need to provide an IP routing connection to the rest of the network. The so-called Inter-I-SID routing functionality provides an implicit routing capability on SPB nodes, that allows connecting a routing interface to an I-SID, thus packets are routed directly from I-SID to I-SID without having to be externally looped back to a VLAN-router.

This capability simplifies network topologies and saves cost.

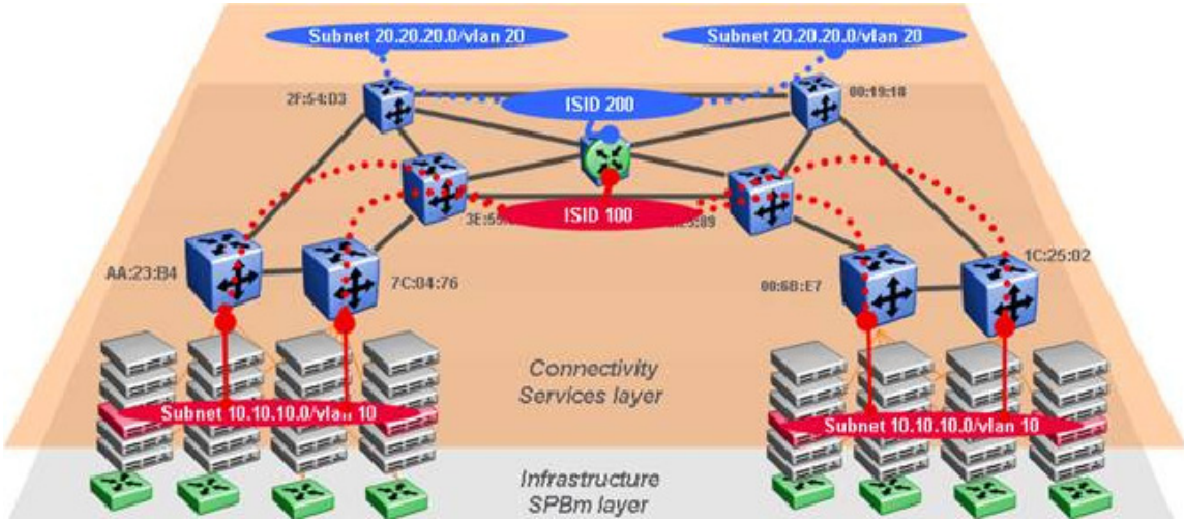


Figure 14 - SPB Inter-I-SID Routing

Summary of benefits:

- Full support for server virtualization technologies: SPB enables easy “moves” of server instances across the network infrastructure.
- Rapid Time to service: a VLAN/IP subnet extension across the network can be established in no time due to the simple end point provisioning and automatic connectivity establishment.
- Robust Infrastructure and increased network uptime: The link state based infrastructure protocol provides quick failure recovery.
- Smooth migration from traditional networking to SPB based virtualization support
- Dual-homing: Traditional Switches can be attached to the SPB cloud redundantly using Switch Clustering.
- Avaya’s SPB not only supports L2 virtualization with SPB but also L3 virtualization.
- Separation between service and infrastructure: Adding new infrastructure (links, switches), does not require to have knowledge of the services that run on top of the network, as well as any kind of topology is supported, thus network operations and design flexibility is greatly improved compared to traditional networking.

6.2 Deployment Scenario: MultiTenant City Network

6.2.1 The Scenario

This section depicts an example of a city network which is providing connectivity services to multiple tenants including: education, administration, government, fire, and police. The tenants are spread across a metro region and are all operated by one network administration. The city operates two data centers, which are using server virtualization techniques that require transparent connectivity between the data centers.

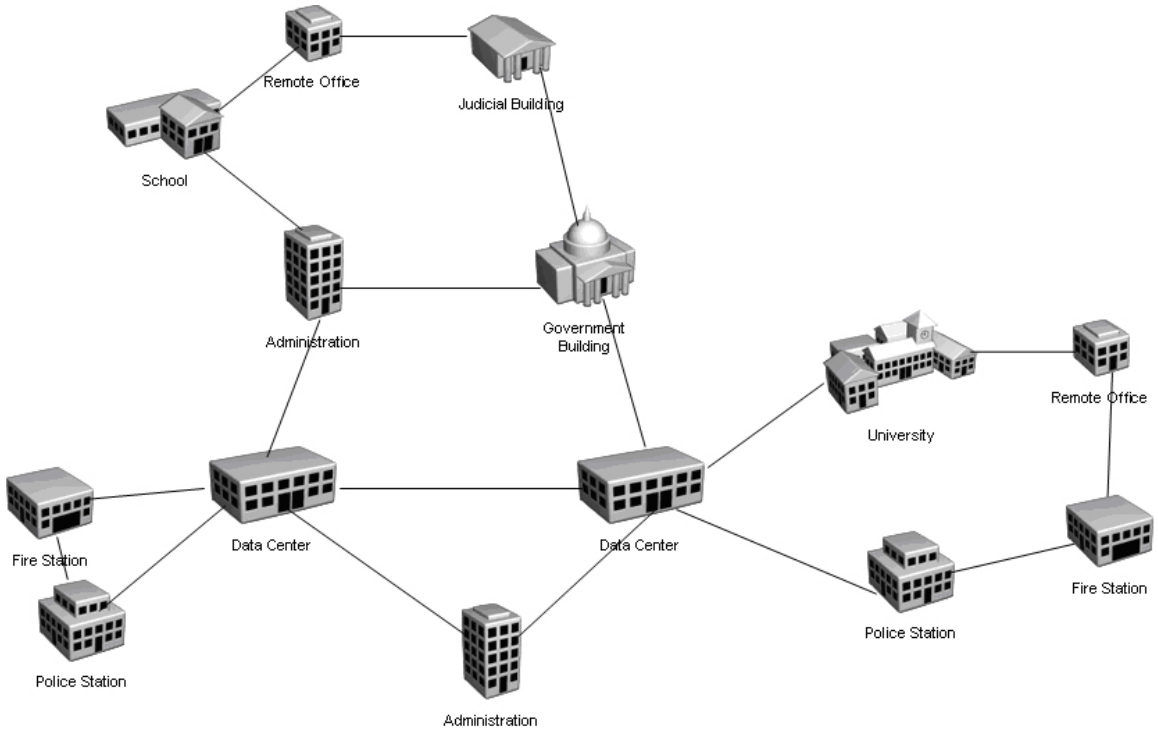


Figure 15 - City Network with Several Departments

6.2.2 Secure network domains

All departments share one backbone infrastructure, but the departments are all operating in their own secure domain. There shall not be any communication between departments without traversing dedicated firewalls which are located in the data centers.

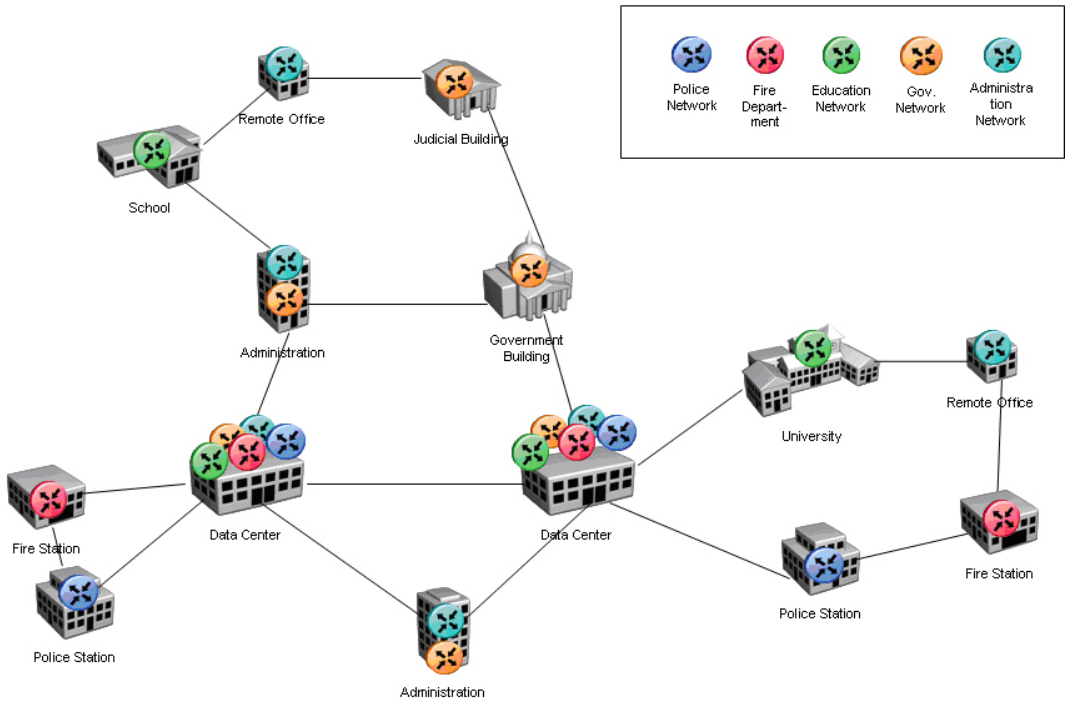


Figure 16 - Secure Domains and their Network Distribution

6.2.3 Core Network – Physical Layout

The blue nodes are SPB capable routing switches which are connected through any type of Ethernet connection.

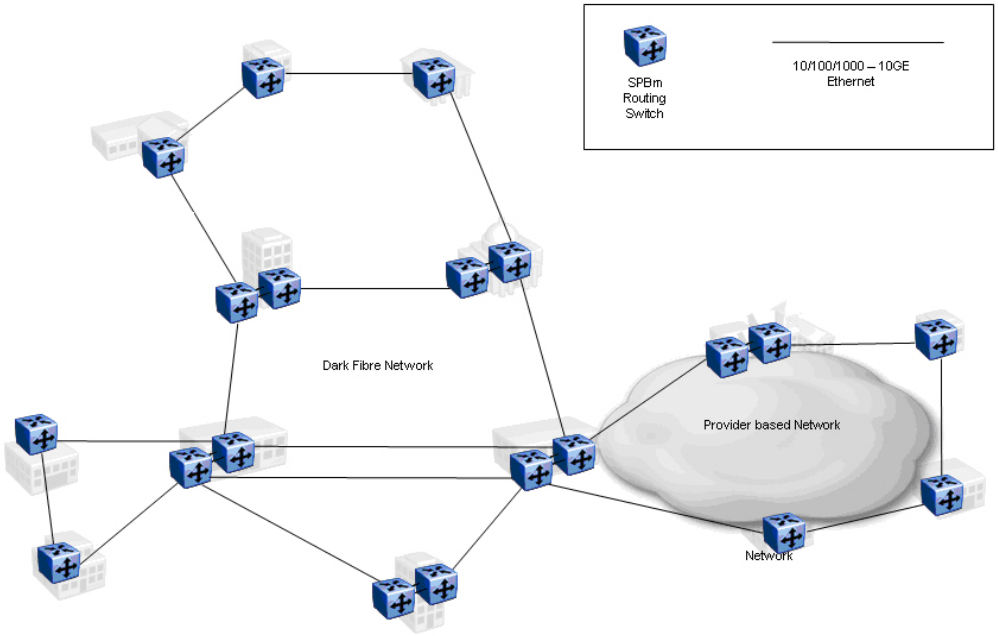


Figure 17 - Physical Layer

The Ethernet can be run over dark fiber, copper, transparent LAN service, carrier E-Line service, CWDM, or DWDM; any type of bridged Ethernet connection as long as the max packet size supports the MAC-in MAC-frames.

The SPB nodes are connected through a set of backbone VLANs (BVLANs), which don't operate in normal bridging mode, but rather have flooding and learning disabled. The SPB-I-SIS protocol is managing the Bridge Filtering Databases (FDBs). One instance of IS-IS is running on all nodes and is responsible for topology discovery as well as virtual network orchestration.

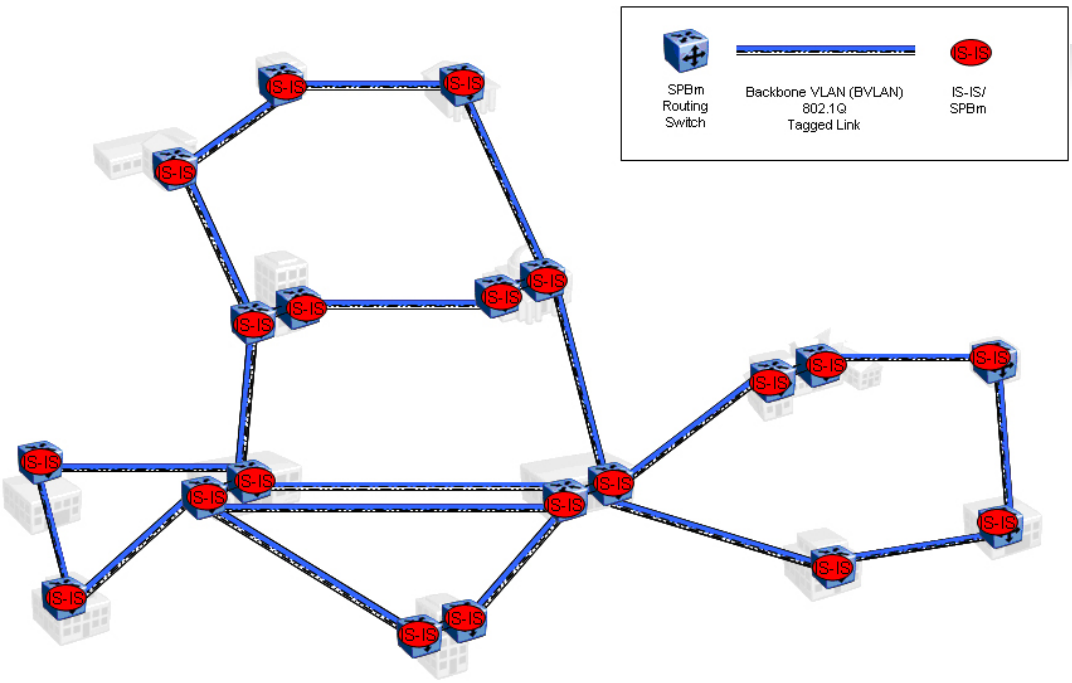


Figure 18 - SPB Protocol Layer

6.2.4 VLAN Extensions

Some of the departments may need to provide Layer 2 extensions across the backbone. SPB provides E-Line point-to-point and E-LAN (any-to-any) connections. Data center bridged connections are frequently required to enable functionalities such as Vmotion from VMware, where server instances can be moved dynamically from one data center to the other. SPB provides a robust and resilient solution for this.

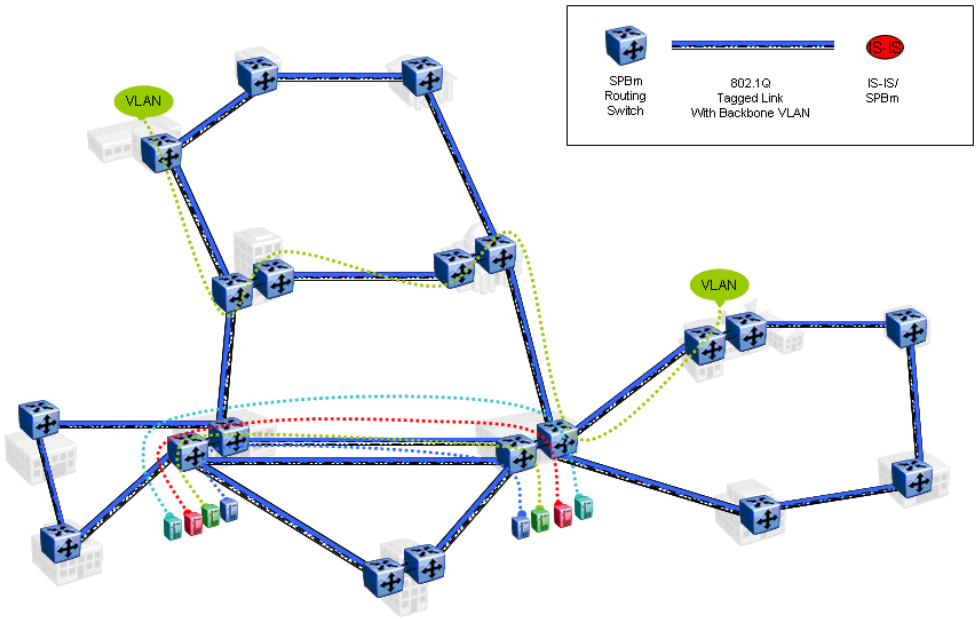


Figure 19 - VLAN Extensions

6.2.5 VRF Extensions

Virtualizing Layer 3 leverages the Virtual Route Forwarder functionality on routing switches. In order to "connect" department specific VRFs across the network, they are assigned to service instances (I-SIDs). IP/SPB then exchanges the VRF specific routing tables automatically and the traffic is forwarded along the VRF-I-SID.

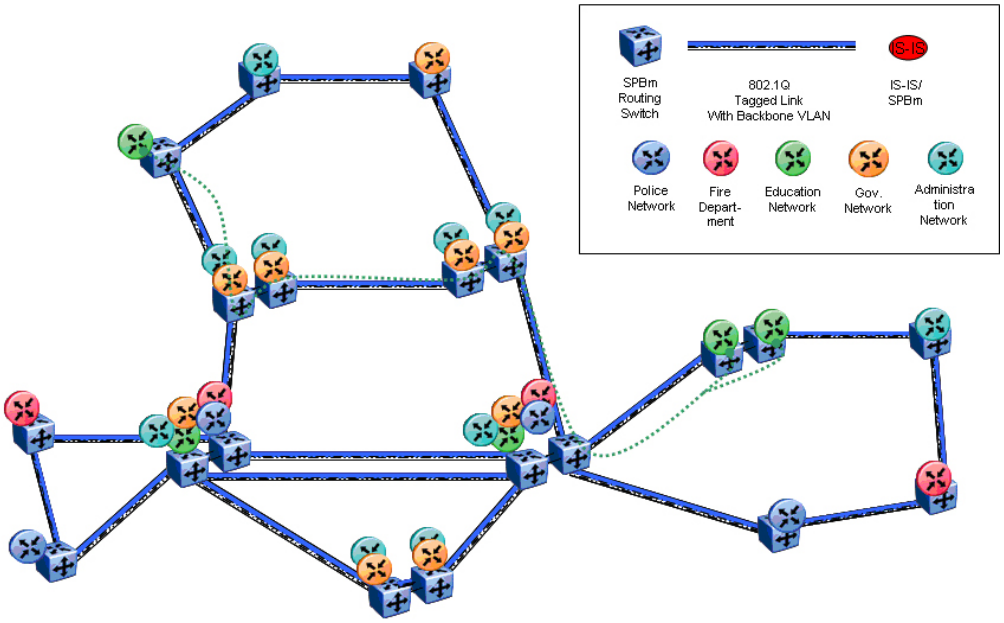


Figure 20 - VRF Extensions

Access Network

The access network is built with traditional network components, the access layer is usually redundantly connect, either to one or two backbone switches.

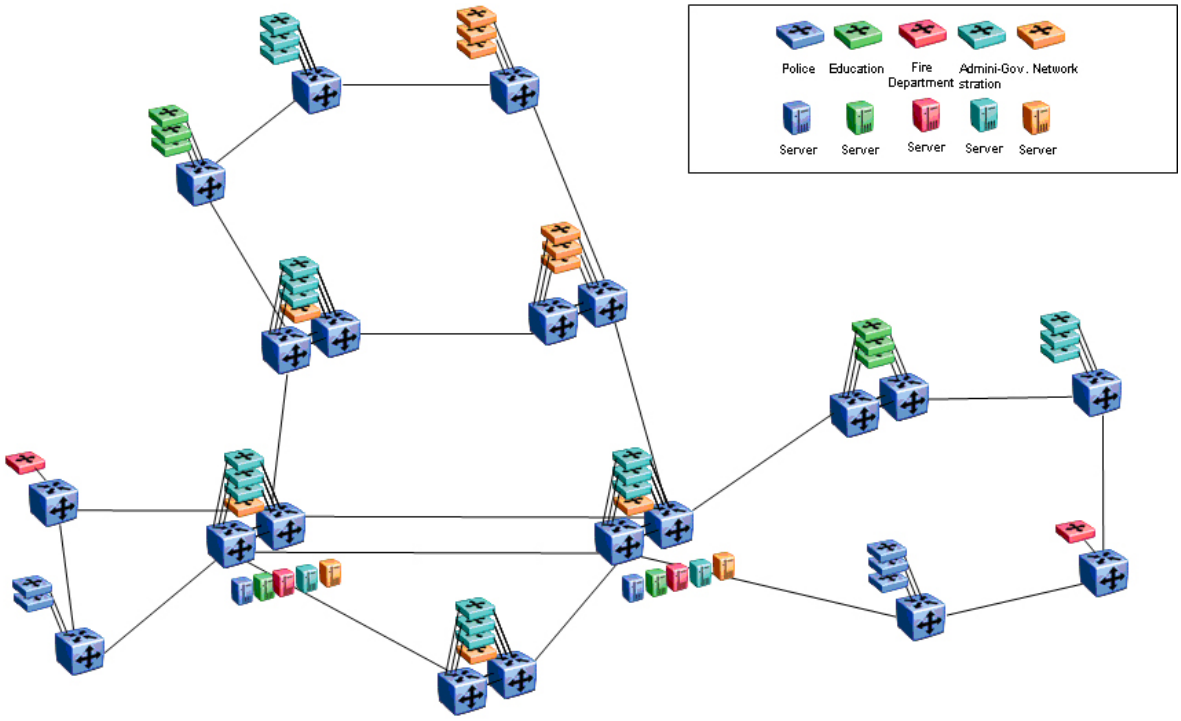


Figure 21 - Access Network

6.2.6 Dual homing access network to SPB backbone

Dual homing of the access network to SPB backbone switches is provided by using the Split Multi-Link Trunking connections directly into the SPB cloud. At the SPB Edge node, VLANs and VRFs are mapped to service instances (I-SIDs). The solution provides a resilient and robust connection into the core.

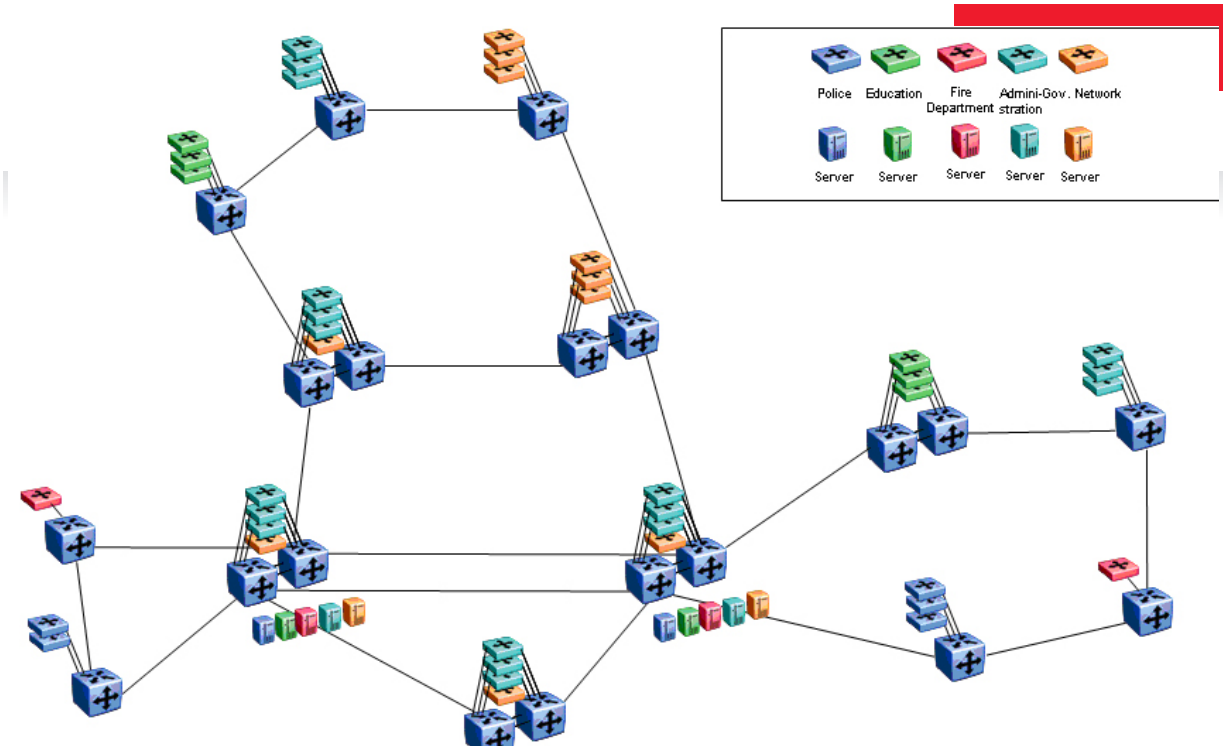


Figure 22 - Dual Homing of Access Switches into SPB Nodes

Traffic ingressing on the grey VLAN is mapped at the SPB node into I-SIDs 100/101 and switched across the SPB core to the destination.

Traffic ingressing on the blue or orange VLAN is routed on the blue or orange VRF, as appropriate, and switched across the SPB network to it's destination on I-SIDs 20x/30x.

VRRP or RSMLT can be used to provide default gateway redundancy.

Traffic on the blue or orange VLAN which is not targeted to be routed by the default gateway can be bridged across the SPB cloud as well (not shown, I-SIDs would have to be assigned to VLAN blue and orange directly in addition to the VRFs.)

7. References

7.1 IEEE

- 802.1D (2004) - MAC Bridges
- 802.1p - Traffic Class Expediting and Dynamic Multicast Filtering (published in 802.1D-1998)
- 802.1Q - Virtual LANs
- 802.1s - Multiple Spanning Trees
- 802.1w - Rapid Reconfiguration of Spanning Tree
- 802.1ag - Connectivity Fault Management
- 802.1ah - Provider Backbone Bridges
- 802.1aq - Shortest Path Bridging

7.2 IETF

- IP/SPB-IP VPNs: http://tools.ietf.org/html/draft-unbehagen-spbb-ip-ipvpn-00

For any comments, edits, corrections, or general feedback, please contact Roger Lapuh (rogerlapuh@avaya.com) or Dan DeBacker (ddebacke@avaya.com).

Acronym Key

Throughout this guide the following acronyms will be used:

BEB: Backbone Edge Bridge (Edge node in a SPB network)
BMAC: Backbone Media Access Control Address
DCB: Data Center Bridging
ECT: Equal Cost Tree
ELAN: Emulated Local Area Network
I-SID: SPB Service ID
IS-IS: Intermediate system to intermediate system Routing Protocol
IP/SPB: IP Shortest Path Bridging
SMLT: Split Multi-Link Trunking
RSMLT: Routed Split Multi-Link Trunking
SPB: Shortest Path Bridging
SPBM: Shortest Path Bridging - MAC-in-MAC
MPLS: Multi Protocol Label Switching
NIC: Network Interface Card
PBB: Provider Backbone Bridging
PBT: Provider Backbone Transport
PLSB: Provider Link State Bridging
VLAN: Virtual Local Area Network
VRF: Virtual Route Forwarder

About Avaya

Avaya is a global leader in enterprise communications systems. The company provides unified communications, contact centers, and related services directly and through its channel partners to leading businesses and organizations around the world. Enterprises of all sizes depend on Avaya for state-of-the-art communications that improve efficiency, collaboration, customer service and competitiveness. For more information please visit www.avaya.com.

AVAYA

INTELLIGENT COMMUNICATIONS

02/11 • DN4469-02

avaya.com