



SPB-based Transportation Network Technical Case Study

Table of Contents

- 1. Introduction3
 - 1.1. Purpose3
 - 1.1. Audience3
 - 1.2. Shortest Path Bridging in Transportation3
- 2. Project Overview3
 - 2.1. System and network overview4
 - 2.2. Operational network overview5
- 3. Operational Network Detailed Design7
 - 3.1. Detailed design topology7
 - 3.2. SPB backbone detailed design7
 - 3.3. SPB service detailed design9
 - 3.4. L3 detailed design10
 - 3.5. L2 multicast detailed design14
 - 3.6. L3 Multicast detailed design15
 - 3.7. QoS detailed design16
 - 3.8. Station attachment detailed design17
- 4. Convergence Time Tests and Results18
- 5. Design Guidelines19
- 6. Conclusion20
- 7. Acronyms20
- 8. Related documents20

1. Introduction

1.1. Purpose

The purpose of this technical case study is to provide a practical example of how ALE's Shortest Path Bridging technology can be applied to a transportation project. This document will present the project requirements along with the Alcatel-Lucent Enterprise Solution Lab validated design and test results.

1.1. Audience

This technical case study is intended for network architects and network engineers involved in the design, implementation and maintenance of networks in the transportation vertical.

1.2. Shortest Path Bridging in Transportation

Rail, highway and airport operators require an IP-based network to support various mission-critical and non-mission critical systems. Each of these systems have unique bandwidth, performance (latency, jitter) and availability requirements. In addition, these systems will communicate various disparate, often proprietary, devices and applications that may be operated and maintained by different groups or vendors and may require communication with third parties.

Shortest Path Bridging (SPB) is an IEEE (802.1aq) standard aimed at addressing various limitations in Spanning Tree Protocol (STP) based Ethernet networks. But SPB is not just the evolution of STP. Like Multiprotocol Label Switching (MPLS), SPB provides virtual private network (VPN) functionality yet is simpler to deploy and maintain, resulting in a lower total cost of ownership (TCO). It is for this reason that SPB is increasingly being considered as an alternative to MPLS across verticals such as transportation.

We can summarize the key reasons why SPB can address the requirements of transportation operators as follows:

Virtualization

SPB VPNs enable secure segregation and bandwidth allocation such that system traffic is isolated and performance requirements are met.

Resiliency

SPB networks can deliver the required level of availability through protected, end-to-end control-plane signaled paths with fast convergence times in any topology.

Operations and Maintenance

SPB networks are simple to operate and maintain because they use a single protocol (IS-IS) at the control plane as opposed to a protocol stack (e.g., BGP/LDP/OSPF etc.) IS-IS builds shortest path trees, distributes service membership information and carries service routes through the backbone.

The Alcatel-Lucent Enterprise Intelligent Fabric technology brings further simplification with plug-n-play and auto-attachment capabilities.

2. Project Overview

This technical case study focuses on a light rail system linking 19 city districts. The light rail system is comprised of:

- Four tramway lines
- 36 stations
- One depot
- Operations Control Center (OCC)
- Backup Operations Control Center (BCC)

2.1. System and network overview

The operation relies on multiple systems which can be classified into three different categories as follows:

Signaling systems

- Automatic Train Control
- Automatic Train Operation
- Automatic Train Protection
- Automatic Train Scheduling
- Wayside Signaling Systems
- Vehicle Location System

Operational systems

- Time Distribution System (TDS)
- Automatic Fare Collection (AFC)
- Facilities and Power Management
- Access Control System (ACS)
- Passenger Information System (PIS)
- Public Address System (PAS)
- Video Surveillance System (VSS) / CCTV
- Telephony (PABX and Help Phone)
- Window Intercom System (WIS)
- Fire Detection and Protection System

Administrative systems

- Advertising Display System
- Email, internet and business processes

Systems in the signaling group are safety-critical and cannot tolerate an outage greater than 50ms. This convergence time is only feasible with SDH or MPLS Fast Re-Route technologies. However, these systems have very low bandwidth requirements.

On the other hand, systems in the operational and administrative groups have higher tolerance to outages (500 ms) but require much more bandwidth.

In addition, signaling and operational systems must be available 24x7 but administrative systems are only used during business hours.

We summarize this in **Table 1** below.

Table 1. System Requirements

	Signaling	Operational	Administrative
Convergence Time	50 ms	500 ms	1 s
Bandwidth	<100 Mbps	<10 Gbps	<10 Gbps
Hours of Operation	24x7	24x7	Business Hours

While a single multi-service MPLS network can meet the requirements of all three groups, the customer opted for having three physically segregated networks instead. This decision is based on the following reasons:

- The single converged network would need to cater to both the high bandwidth requirements of the operational and administrative groups and the low convergence time requirements of the signaling group. The cost of such network is very high.
- Regulations require that signaling systems must use a physically separate network OR, if they use the same network as the other systems, additional security measures are required. This additional security layer adds complexity and cost.
- An administrative network is separate for organizational and operational reasons (e.g., scheduling of maintenance windows).

This document will focus on the Operational network.

2.2. Operational network overview

The Operations Control Centre (OCC) is the primary location where all aspects of the transportation system are supervised and controlled. The OCC hosts applications and databases and interfaces with third parties such as emergency responders. The Backup Control Centre (BCC) hosts redundant infrastructure and resources such that it can replace the OCC in the event of a disaster or during maintenance. OCC and BCC operate in Active/Standby mode.

2.2.1. High level topology

Stations along each of the four tramway lines are linked with the OCC and the BCC in a ring topology. This is the backbone network. **Figure 1** on the following page shows the topology for a single line.

Each station has its own backbone node. The ring topology provides redundant paths in the event of backbone link or backbone node failure. All links use single fiber pairs except for the ones linking OCC and BCC which use double pairs. This is so that traffic between these two sites (such as a database backup or a VM move) does not have to take the longer route over the ring in the event of a link failure.

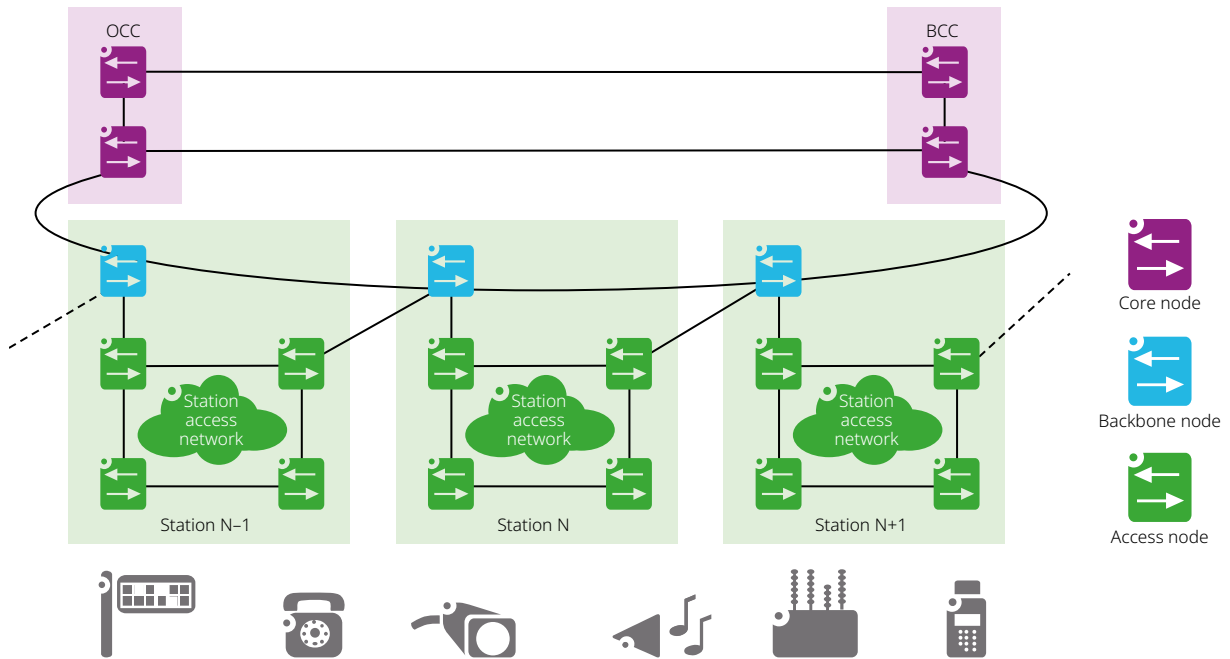
OCC and BCC also have redundant core nodes offering the maximum level of redundancy and no single point of failure.

A station access network connects station devices such as CCTV cameras also in a ring topology. The station access network connects to both the backbone node within the station and also to the next station's backbone node.

As a result, this topology is resilient to the failures below:

- Station node failure
- Station link failure
- Backbone node failure
- Backbone link failure
- Single and double core node failure
- Single and double core link failure

Figure 1. High Level Design Topology



2.2.2. High level design

Backbone network: The backbone network is comprised of backbone nodes and provides L2 SPB services to systems in the operational system category. The backbone nodes are backbone edge nodes acting as point of demarcation between backbone and station networks.

Station access network: The station access network is comprised of hardened switches suitable for track-side deployment (such as Alcatel-Lucent OmniSwitch® 6855/6865) interconnected in a L2 ring topology. The station access nodes connect to redundant backbone nodes. The station access network is outside the scope of this document.

Core network: The core network spans both the OCC and BCC. Dual core nodes are deployed at each of these sites and both sites are linked by redundant fiber. All routing (Unicast and Multicast) is performed at the core nodes. The core nodes are also backbone nodes / backbone edge nodes acting as point of demarcation between the backbone and the data center network.

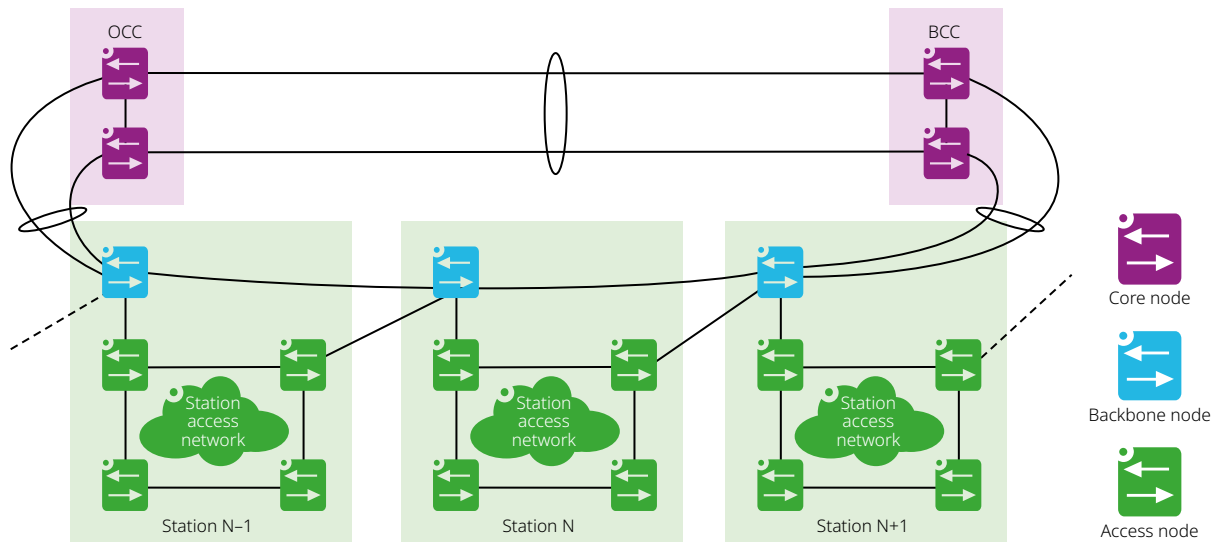
Data center network: The data center network spans both OCC and BCC sites. The data center distribution switches connect to the core nodes at each of these sites. The data center network is a L2 fabric with all routing performed at the core nodes.

3. Operational Network Detailed Design

3.1. Detailed design topology

The detailed design topology is shown in **Figure 2** below.

Figure 2. Detailed Design Topology



A few modifications have been made in this detailed design topology compared to the high level design topology.

- Core nodes are clustered in a virtual chassis configuration. The virtual chassis is a single logical device with a single control and management plane. This ensures that even in the event of a failure of the individual physical nodes, the logical topology and traffic flows will not change. In addition, this means that data center distribution switches can be dual-homed to both core nodes through a LAG, providing redundancy and fast convergence time in the event of link or core node failure.
- Links between OCC and BCC are bundled in an LACP aggregate. Similarly, this means that these two links will behave as a single logical link and topology and traffic flows will not change when one of the physical links fail.
- Links between OCC/BCC and next-hop backbone node are duplicated and also bundled in an LACP aggregate. This ensures that a failure of the individual core nodes will not trigger a topology change or change in traffic flows, resulting in faster convergence time.

3.2. SPB backbone detailed design

The backbone design is straightforward. SPB IS-IS is enabled on all backbone and core node ring interfaces. There are only two possible paths in a ring topology and therefore only two backbone VLANs (BVLANS) are required to load balance service traffic over different equal cost trees (ECTs). These BVLANS are mapped to different equal cost tree IDs (ECT-IDs) to ensure they build different ECTs. The ECT-ID designates the algorithm used by the shortest path tree (SPT) tie-breaking logic. A third BVLAN is used exclusively for SPB (IS-IS) control traffic.

In Section 3.3, we will show how different systems will be mapped to different ISIDs (service instance identifiers) and BVLANS causing them to take different paths. In other words, no link is disabled, all links are used, and the system traffic is balanced over different paths.

The backbone node configuration is shown in the snippet below.

Backbone node

```
! SPB-ISIS:
spb isis bvlan 4000 ect-id 1
spb isis bvlan 4001 ect-id 2
spb isis bvlan 4002 ect-id 3
spb isis control-bvlan 4000
spb isis interface port 1/1/27-28
spb isis admin-state enable
```

The core node configuration is shown in the snippet below.

Core node

```
! SPB-ISIS:
spb isis bvlan 4000 ect-id 1
spb isis bvlan 4001 ect-id 2
spb isis bvlan 4002 ect-id 3
spb isis control-bvlan 4000
spb isis interface linkagg 1
spb isis interface linkagg 3
spb isis admin-state enable
```

In these snippets:

- BVLAN 4000 is used for control
- BVLANS 4001 and 4002 will be used for system traffic
- SPB IS-IS is enabled on backbone ring interfaces (1/1/27 and 1/1/28 on backbone nodes and Linkagg groups 1 and 3 on core nodes)

As seen in the snippets, backbone configuration is independent from service configuration. This means that service moves, adds and changes require no changes in the backbone, simplifying network operations.

It should be noted that LAGs between SPB backbone devices such as those between OCC and BCC and OCC/BCC and the next-hop station are there to provide resiliency but not to provide additional bandwidth. The link metric will not change when one of the physical links fail.

In a LAG, traffic is load balanced by means of a hashing logic that can use either MAC addresses (brief mode) or IP addresses and TCP/UDP port numbers (extended mode). With MAC-in-MAC encapsulation however, MAC addresses are always those of the backbone devices BMACs and IP addresses and port numbers are hidden from the hashing logic. AOS 8.3.1R01 introduced the “tunnel-protocol” option such that traffic can be load balanced according to inner CMACs or IP addresses and TCP/UDP port numbers even when traffic is MAC-in-MAC encapsulated. It is recommended that this option be enabled on all SPB backbone LAGs. The choice of MAC (brief) or IP+TCP/UDP ports (extended) is a global setting which will apply to all LAGs. Please refer to the “AOS Command Line Interface Guide” for further details.

3.3. SPB service detailed design

In order to achieve logical segregation between systems, system VLANs are mapped to SPB services on a 1:1 basis and every system, with all its subnets, is mapped to a virtual routing and forwarding (VRF) instance at the core nodes.

Table 2 below shows mapping of systems to services, VRFs, subnets, VLANs, ISIDs and BVLANS. The use of subnets and VRFs will be discussed in **Section 3.4**.

Configurations in this document are those used during lab testing and therefore the service name does not refer to an actual system or application (such as Automatic Fare Collection, etc.) but rather the service's purpose in the test.

It should be noted that in this design every station has its own set of services and VLANs. In other words, VLANs and services are not shared among stations. Only OCC and BCC core nodes will have all services enabled on them. This is to avoid creating large broadcast domains.

Table 2. Service Mappings

Test Direction	Test Purpose	Nodes	Service #	VRF	Subnet	VLAN	ISID	BVLAN
Station<->OCC/BCC	L2 Unicast	Backbone and core	3001	App1	NA	3001	13001	4001
	L2 Multicast		3101	App1	NA	3101	13101	4001
OCC<->BCC	L3 Unicast A	Core	3201	App1	10.2.1.0/24	3201	13201	4001
	L3 Multicast A		3302	App2	10.3.2.0/24	3302	13302	4002
	L3 Multicast B		3502	App2	10.5.2.0/24	3502	13502	4002
Station<->Station	L3 Unicast B	Backbone and Core	3202	App2	10.2.2.0/24	3202	13202	4002
	L3 Multicast A		3402	App2	10.4.2.0/24	3402	13402	4002
	L3 Multicast B		3602	App2	10.6.2.0/24	3602	13602	4002

Systems are mapped to SPB Services at the backbone node through a service access point (SAP). The SAP is a virtual port that identifies the type of traffic (untagged, single or double VLAN tag, tag range or wildcard) that will be mapped to the SPB service.

The backbone node configuration is shown in the snippet below. In this snippet, port 1/1/13 is the port connecting to local station access node. Configuration of the port connecting to the remote station access node is not shown. As can be seen in the snippet, services are split between two different BVLANS to spread the service load over different shortest path trees, making use of all available links.

Backbone node

```
! SVCMMGR:
service 3001 spb isid 13001 bvlan 4001
service 3101 spb isid 13101 bvlan 4001
service 3202 spb isid 13202 bvlan 4002
service 3402 spb isid 13402 bvlan 4002 multicast-mode tandem
service 3602 spb isid 13602 bvlan 4002 multicast-mode tandem
service 3001 sap port 1/1/13:3001
service 3101 sap port 1/1/13:3101
service 3202 sap port 1/1/13:3202
service 3402 sap port 1/1/13:3402
service 3602 sap port 1/1/13:3602
```

The core node configuration is shown in the snippet below. In this snippet, linkagg 2 is the LAG towards the DC distribution block.

Core node

```

! SVCNMR:
service 3001 spb isid 13001 bvlan 4001
service 3101 spb isid 13101 bvlan 4001 multicast-mode tandem
service 3201 spb isid 13201 bvlan 4001
service 3202 spb isid 13202 bvlan 4002
service 3302 spb isid 13302 bvlan 4002 multicast-mode tandem
service 3402 spb isid 13402 bvlan 4002 multicast-mode tandem
service 3502 spb isid 13502 bvlan 4002 multicast-mode tandem
service 3602 spb isid 13602 bvlan 4002 multicast-mode tandem
service 3001 sap linkagg 2:3001
service 3101 sap linkagg 2:3101
service 3201 sap linkagg 2:3201
service 3302 sap linkagg 2:3302
service 3502 sap linkagg 2:3502

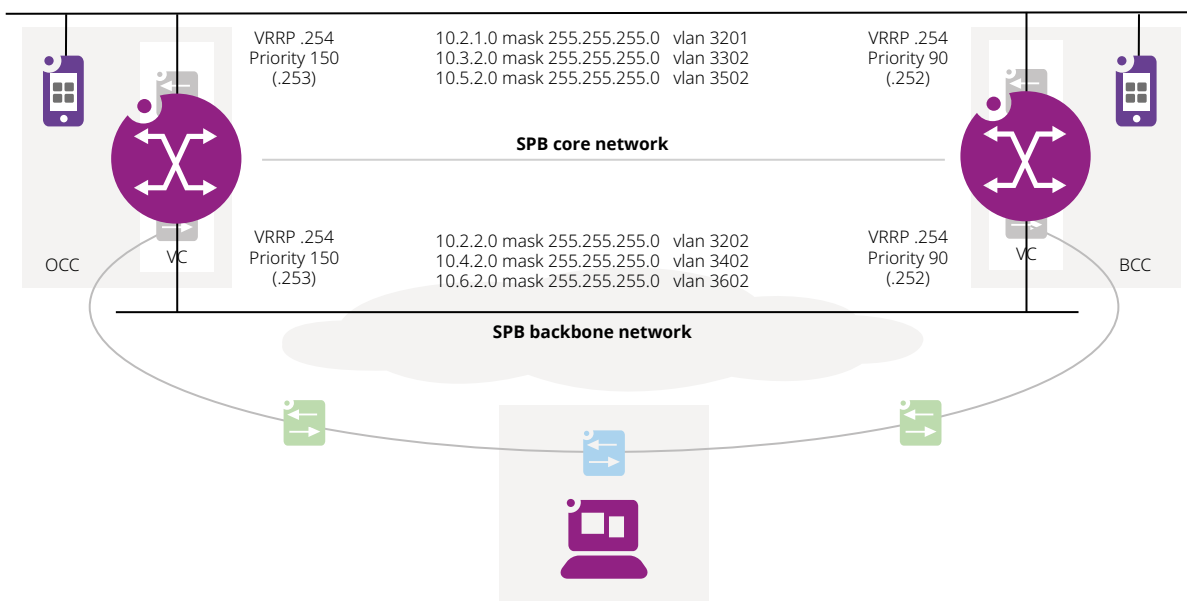
```

Please refer to **Section 3.5** for a discussion of Multicast replication modes.

3.4. L3 detailed design

As explained earlier, this is a L2 SPB VPN deployment and all stations (backbone nodes) as well as OCC and BCC (core nodes) are linked by L2 SPB services. Devices at the stations have their default gateway and all routing is performed at the core nodes. Both core nodes are configured as a VRRP pair to provide default gateway redundancy to station devices in the event of complete core node VC or site failure. Please refer to Figure 3 below.

Figure 3. Default Gateway Redundancy - VRRP



In order to isolate systems at L3, all services belonging to a given system will be grouped in a VRF instance. Only two systems are considered in this document: App1 and App2.

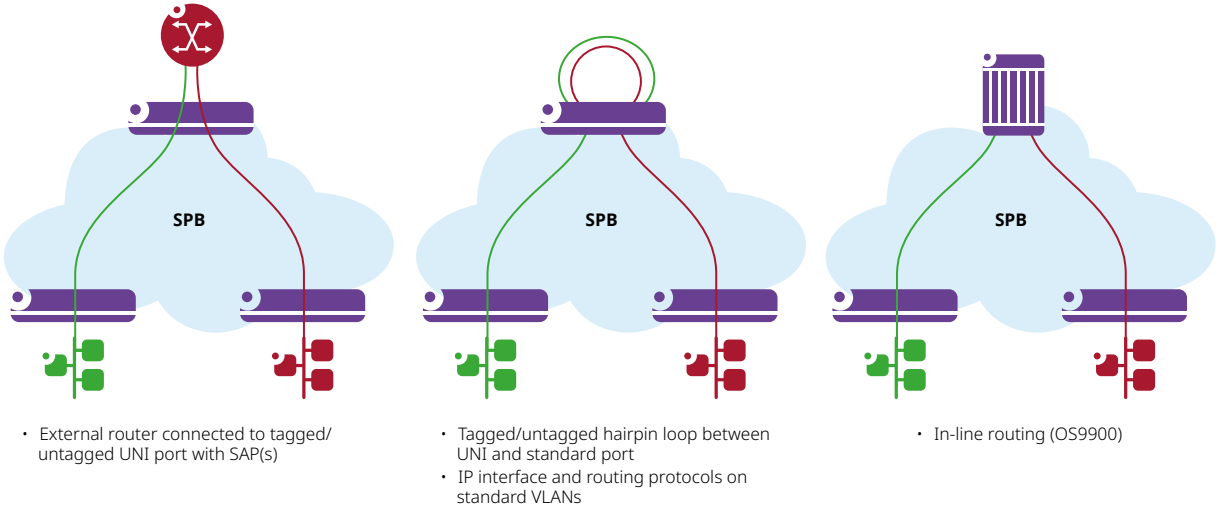
Table 2 on the previous page contains subnet and VRF information for each service.

At this point it is pertinent to explain routing in an SPB environment.

Routing between different SPB ISIDs (services) and between an SPB ISID and a standard VLAN, or routed interface, can be performed in one of three ways as shown in Figure 4. New generation chipsets such as those used on the OmniSwitch 9900 support in-line routing, however, older chipsets can only support external or hairpin routing.

- External router or firewall (router on a stick): The external router or firewall connects to an SPB access port on a BEB through a VLAN-tagged or un-tagged interface. SAPs are created for every SPB service requiring routing.
- Hairpin loop: This is logically equivalent to the external router except the same device is performing both roles. One port is configured as an SPB access port with SAPs and the other port is a standard VLAN-tagged or un-tagged port. IP interfaces are created on the standard VLANs to perform unicast and multicast routing.
- In-line routing: Routing in-line between ISIDs in the SPB domain and between the SPB and VLAN domains in a single operation.

Figure 4. Routing in SPB



In this design the core nodes are OmniSwitch 6900-X72 virtual chassis and use hairpin routing. Each core node VC has a 2-port LAG as a hairpin loop where one side is a standard VLAN-tagged port and the other side is an SPB access port with SAPs for all services requiring routing.

As routing is only performed on core nodes, snippets are shown only for those nodes.

Core node 1

```
! VLAN:
! These standard VLANs are used on the VLAN side of the hairpin loop
vlan 3201 name "VRF.App1_L3.Unicast.Testing"
vlan 3202 name "VRF.App2_L3.Unicast.Testing"
vlan 3302 name "VRF.App2_L3.Multicast.Testing"
vlan 3402 name "VRF.App2_L3.Multicast.Testing"
vlan 3502 name "VRF.App2_L3.Multicast.Testing"
vlan 3602 name "VRF.App2_L3.Multicast.Testing"
! SVCMgr:
! SAPs are created on the SPB access port side of the hairpin loop
service 3201 sap linkagg 102:3201
service 3202 sap linkagg 102:3202
service 3302 sap linkagg 102:3302
service 3402 sap linkagg 102:3402
service 3502 sap linkagg 102:3502
service 3602 sap linkagg 102:3602
! IP:
! VRFs and IP interfaces residing on the standard VLANs are created to route
vrf App1 profile max
vrf App2 profile max
vrf App1 ip interface "vlan3201" address 10.2.1.253 mask 255.255.255.0 vlan
3201
vrf App2 ip interface "vlan3202" address 10.2.2.253 mask 255.255.255.0 vlan
3202
vrf App2 ip interface "vlan3302" address 10.3.2.253 mask 255.255.255.0 vlan
3302
vrf App2 ip interface "vlan3402" address 10.4.2.253 mask 255.255.255.0 vlan
3402
vrf App2 ip interface "vlan3502" address 10.5.2.253 mask 255.255.255.0 vlan
3502
vrf App2 ip interface "vlan3602" address 10.6.2.253 mask 255.255.255.0 vlan
3602
! VRRP:
! Core Node 1 will become the Master because its priority is higher
vrf App2 ip load vrrp
vrf App2 vrrp 4 3202 priority 150 preempt interval 1
vrf App2 vrrp 4 3202 address 10.2.2.254
vrf App2 vrrp 4 3202 admin-state enable
vrf App2 vrrp 5 3302 priority 150 preempt interval 1
vrf App2 vrrp 5 3302 address 10.3.2.254
vrf App2 vrrp 5 3302 admin-state enable
vrf App2 vrrp 6 3402 priority 150 preempt interval 1
vrf App2 vrrp 6 3402 address 10.4.2.254
vrf App2 vrrp 6 3402 admin-state enable
vrf App2 vrrp 7 3502 priority 150 preempt interval 1
vrf App2 vrrp 7 3502 address 10.5.2.254
vrf App2 vrrp 7 3502 admin-state enable
vrf App2 vrrp 8 3602 priority 150 preempt interval 1
vrf App2 vrrp 8 3602 address 10.6.2.254
vrf App2 vrrp 8 3602 admin-state enable
```

Core node 2

```
! VLAN:
! These standard VLANs are used on the VLAN side of the hairpin loop
vlan 3201 name "VRF.App1_L3.Unicast.Testing"
vlan 3202 name "VRF.App2_L3.Unicast.Testing"
vlan 3302 name "VRF.App2_L3.Multicast.Testing"
vlan 3402 name "VRF.App2_L3.Multicast.Testing"
vlan 3502 name "VRF.App2_L3.Multicast.Testing"
vlan 3602 name "VRF.App2_L3.Multicast.Testing"
! SVCMGR:
! SAPs are created on the SPB access port side of the hairpin loop
service 3201 sap linkagg 102:3201
service 3202 sap linkagg 102:3202
service 3302 sap linkagg 102:3302
service 3402 sap linkagg 102:3402
service 3502 sap linkagg 102:3502
service 3602 sap linkagg 102:3602
! IP:
! VRFs and IP interfaces residing on the standard VLANs are created to route
vrf App1 profile max
vrf App2 profile max
vrf App1 ip interface "vlan3201" address 10.2.1.252 mask 255.255.255.0 vlan
3201
vrf App2 ip interface "vlan3202" address 10.2.2.252 mask 255.255.255.0 vlan
3202
vrf App2 ip interface "vlan3302" address 10.3.2.252 mask 255.255.255.0 vlan
3302
vrf App2 ip interface "vlan3402" address 10.4.2.252 mask 255.255.255.0 vlan
3402
vrf App2 ip interface "vlan3502" address 10.5.2.252 mask 255.255.255.0 vlan
3502
vrf App2 ip interface "vlan3602" address 10.6.2.252 mask 255.255.255.0 vlan
3602
! VRRP:
! Core Node 2 will become the Slave because its priority is lower
vrf App2 ip load vrrp
vrf App2 vrrp 4 3202 priority 90 preempt interval 1
vrf App2 vrrp 4 3202 address 10.2.2.254
vrf App2 vrrp 4 3202 admin-state enable
vrf App2 vrrp 5 3302 priority 90 preempt interval 1
vrf App2 vrrp 5 3302 address 10.3.2.254
vrf App2 vrrp 5 3302 admin-state enable
vrf App2 vrrp 6 3402 priority 90 preempt interval 1
vrf App2 vrrp 6 3402 address 10.4.2.254
vrf App2 vrrp 6 3402 admin-state enable
vrf App2 vrrp 7 3502 priority 90 preempt interval 1
vrf App2 vrrp 7 3502 address 10.5.2.254
vrf App2 vrrp 7 3502 admin-state enable
vrf App2 vrrp 8 3602 priority 90 preempt interval 1
vrf App2 vrrp 8 3602 address 10.6.2.254
vrf App2 vrrp 8 3602 admin-state enable
```

While VRFs are used to segregate systems at L3, it is often the case that some communication between different systems is required. This is usually accomplished through an external firewall. In this test bed however, Route Leaking is used to allow communication between specific subnets in a controlled manner.

With Route Leaking, routes can be selectively exported to and imported from one VRF to another by means of route maps. A route map is a special kind of filter that selects routes to be imported into or exported from the routing table through an IP access list.

Core nodes

```
! IP Route Manager:
vrf App1 ip access-list "AL_App1"
vrf App1 ip access-list "AL_App1" address 10.2.2.0/24 action permit redistrib-control no-subnets
vrf App1 ip route-map "RM_App1" sequence-number 10 action permit
vrf App1 ip route-map "RM_App1" sequence-number 10 match ip-address "AL_App1"
vrf App1 ip access-list "AL_App2"
vrf App1 ip access-list "AL_App2" address 10.2.1.0/24 action permit redistrib-control no-subnets
vrf App1 ip route-map "RM_App2" sequence-number 10 action permit
vrf App1 ip route-map "RM_App2" sequence-number 10 match ip-address "AL_App2"
vrf App1 ip export vrf App2 route-map RM_App2
vrf App1 ip import vrf App2 route-map RM_App1
vrf App2 ip access-list "AL_App2"
vrf App2 ip access-list "AL_App2" address 10.2.1.0/24 action permit redistrib-control no-subnets
vrf App2 ip route-map "RM_App2" sequence-number 10 action permit
vrf App2 ip route-map "RM_App2" sequence-number 10 match ip-address "AL_App2"
vrf App2 ip access-list "AL_App1"
vrf App2 ip access-list "AL_App1" address 10.2.2.0/24 action permit redistrib-control no-subnets
vrf App2 ip route-map "RM_App1" sequence-number 10 action permit
vrf App2 ip route-map "RM_App1" sequence-number 10 match ip-address "AL_App1"
vrf App2 ip export vrf App1 route-map RM_App1
vrf App2 ip import vrf App1 route-map RM_App2
```

3.5. L2 multicast detailed design

Various transportation systems rely on Multicast transmission. Video surveillance is usually the system that generates the most multicast traffic both in terms of bandwidth and flows.

In SPB, there are three methods to handle multicast traffic (as well as broadcast and unknown DA traffic): Head-End Replication and Tandem (S,G) or (*,G) Replication.

In the Head-End Replication method, multicast traffic received on a SAP will be replicated at the ingress BEB and a copy will be sent to all BEBs connected to the service (ISID).The Head-End method uses the existing BVLAN's Unicast tree.

In the Tandem (S,G) Replication method, special Multicast trees are created for every BEB and ISID and frames are replicated as they are forwarded down these trees. These Multicast trees are congruent with the BVLAN's Unicast trees, however, are necessary to add ISID and ingress BEB information to the FDB so that BCB nodes know if they are in the shortest path from the ingress BEB for a particular ISID and need to forward (and replicate) the frame.

The Head-End mode of operation is inefficient in terms of bandwidth usage because a copy will be created and sent from the ingress BEB for every other BEB in the same ISID.

The Tandem mode of operation is more efficient in terms of bandwidth consumption. In a ring topology, there are only two possible paths and a maximum of two copies will be created but only a single copy will

be sent over a given link. The downside is that additional Multicast trees are required for every ISID and BEB. These additional trees consume extra table space and CPU cycles. With a relatively small number of ISIDs and BEBs however this is a small price to pay and therefore this is the recommended mode of operation for services carrying high multicast traffic loads such as Video Surveillance.

In order to reduce resource usage, a Tandem (*,G) Mode is also supported. In the (*,G) mode there is not one tree per source BEB and ISID but a single tree per BVLAN. The bridge with the lowest Bridge ID is elected as the root of the tree very much in the same way as Spanning Tree. Traffic from every BEB will use this tree and will be replicated at fork-out points. When using (*,G) mode, the path is not the shortest for every BEB. Also, because Unicast traffic continues to use the SPT, congruency is lost.

Table 3 below compares these three modes of operation and suggested use.

Table 3. Multicast Mode Comparison

	Head-end	Tandem (S,G)	Tandem (*,G)
Bandwidth Efficiency	Low	High	High
Resource Usage	Low	High	Low-Medium
Congruency	Yes	Yes	No
Suggested use	Low multicast bandwidth.	<ul style="list-style-type: none"> • Many receivers, few sources. • High multicast bandwidth. 	<ul style="list-style-type: none"> • When root Bridge is source or receiver of most multicast traffic and congruency is not required • When required to interoperate with third party equipment.

In this design the decision is to use Head-End replication for all services except video surveillance which will use Tandem (S,G) replication. This is a good compromise between bandwidth consumption and switch resource usage.

The choice of multicast replication mode is shown in the snippets presented in **Section 3.3**. Head-End is the default mode of operation. Therefore, only Tandem mode needs to be configured.

3.6. L3 Multicast detailed design

Having discussed Multicast operation at L2, it is necessary to explain Multicast operation at L3. L3 multicast is required because sources and receivers may not be on the same subnet.

L3 Multicast is based on Protocol-Independent Multicast (PIM). PIM uses routing information provided by Unicast routing protocols such as OSPF. In this case, the routing table is populated with local and leaked routes.

PIM is enabled on core nodes where routing takes place. IGMP snooping v3 is enabled on backbone nodes to avoid flooding multicast traffic into SAPs and BEB without active subscribers. Please contact ALE for availability of this feature.

OmniSwitch products support PIM Dense Mode (PIM-DM) and Sparse Mode (PIM-SM) as well as PIM Source Specific Multicast (PIM-SSM). PIM-SSM is chosen for better scalability as required by the Video Surveillance application.

Protocol-Independent Multicast Source-Specific Multicast (PIM-SSM) is a highly-efficient extension of PIM. Using an explicit channel subscription model, PIM-SSM allows receivers to receive multicast traffic directly from the source; a Rendezvous Point (RP) tree model is not used. In other words, a Shortest Path Tree (SPT) between the receiver and the source is created without the use of a RP. Please refer to [1] for an overview of PIM-SSM.

Backbone nodes

```
! IPMS:
ip multicast admin-state enable
ip multicast version 3
```

Core nodes

```
! IPMS:
ip multicast admin-state enable
ip multicast version 3
! IP Multicast:
ip load pim
vrf App2 ip load pim
vrf App2 ip pim interface "vlan3302"
vrf App2 ip pim interface "vlan3402"
vrf App2 ip pim interface "vlan3502"
vrf App2 ip pim interface "vlan3602"
vrf App2 ip pim ssm group 232.0.0.0/8
vrf App2 ip pim sparse admin-state enable
```

3.7. QoS detailed design

In an SPB network, traffic is classified at the point of ingress: The ingress SAP. This classification is used to map traffic to egress queues as it travels across the backbone and exits through another SAP at the destination BEB.

SAPs can be configured as “trusted” or “un-trusted”. When the SAP is trusted, and the incoming traffic is tagged, the CoS markings are copied from the (outer) VLAN tag into the BVLAN tag. When the SAP is trusted but the incoming traffic is not tagged, the CoS is set to the port’s default priority. When the SAP is un-trusted, the CoS marking is configured by the user.

If a SAP is un-trusted and the CoS marking is configured by the user, all traffic received at the SAP is marked with the same CoS. Different CoS markings on the same SAP are only possible if the SAP is configured to trust markings in tags originating from a downstream device. Any

other QoS configuration is disabled on SPB access ports. This includes classification based on L2-L4 policies.

In this design and test, frames will be classified as station access nodes and data center distribution nodes. SAPs on backbone and core nodes are set to the default configuration which is to trust markings on incoming tagged frames. The exception is Service 3202 which will rewrite the CoS to 5 at the SAP for testing purposes. Please refer to configuration snippets below.

Backbone nodes

```
! SVCNMR:
service 3202 sap linkagg 102:3202 un-trusted 5
```

Core nodes

```
! SVCNMR:  
service 3202 sap port 1/1/13:3202 un-trusted 5
```

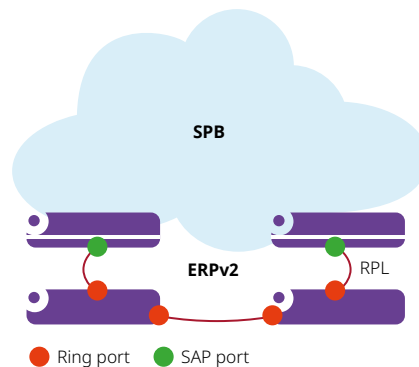
3.8. Station attachment detailed design

The station access network is attached to diverse BEBs for redundancy. The station access network can use Spanning Tree Protocol or Ethernet Ring Protection Version 2 (ERPV2) for loop avoidance. In this case, ERPv2 is chosen.

The station access network constitutes an ERPv2 sub-ring as shown in **Figure 5**. This sub-ring is attached to the two BEBs through SAP ports. The sub-ring is not closed with additional ring or SAP ports and is only closed through the SPB backbone. The sub-ring can use R-APS or non R-APS virtual channel. R-APS PDUs will be tunneled through the SPB backbone to other BEBs connecting the service (ISID). Whether it is tagged or un-tagged, the sub-ring's service VLAN must be matched by SAPs at the BEBs and transported as an SPB service.

For this purpose, VLAN 50 is created as the ERP control VLAN on the station access network and Service 10005 is created to transport it across the SPB backbone. Refer to the snippet below for details. Please note that only details related to the UNI are shown in these snippets. Other configuration details such as RPL are not shown.

Figure 5. ERPv2 Station Access Network Attachment



Backbone node

```
! SVCNMR:  
service 5 spb isid 10005 bvlan 4001  
service 5 sap port 1/1/13:5
```

Access node

```
!ERP  
erp-ring 1 port1 2/49 port2 2/50 service-vlan 5 level 2  
erp-ring 1 rpl-node port 2/49  
erp-ring 1 enable
```

4. Convergence Time Tests and Results

This design was validated at the Alcatel-Lucent Enterprise Solutions Lab in Colombes, France. In particular, re-convergence time was measured in various test cases. Many other tests such as throughput, latency and compliance testing are performed as part of Q&A validation. Please contact ALE if required.

We will not provide details of every test case in this document. We will only focus on two of the most critical test cases as an example: L2 and L3 convergence time in the event of failure of the VC Master unit at the OCC's core node. Please refer to **Figure 6** and **Figure 7** and **Table 4** and **Table 5** below for details. As can be seen in the tables, convergence time results are well below the 500ms mark which was required.

Figure 6. OCC to BCC Primary and Failover Paths

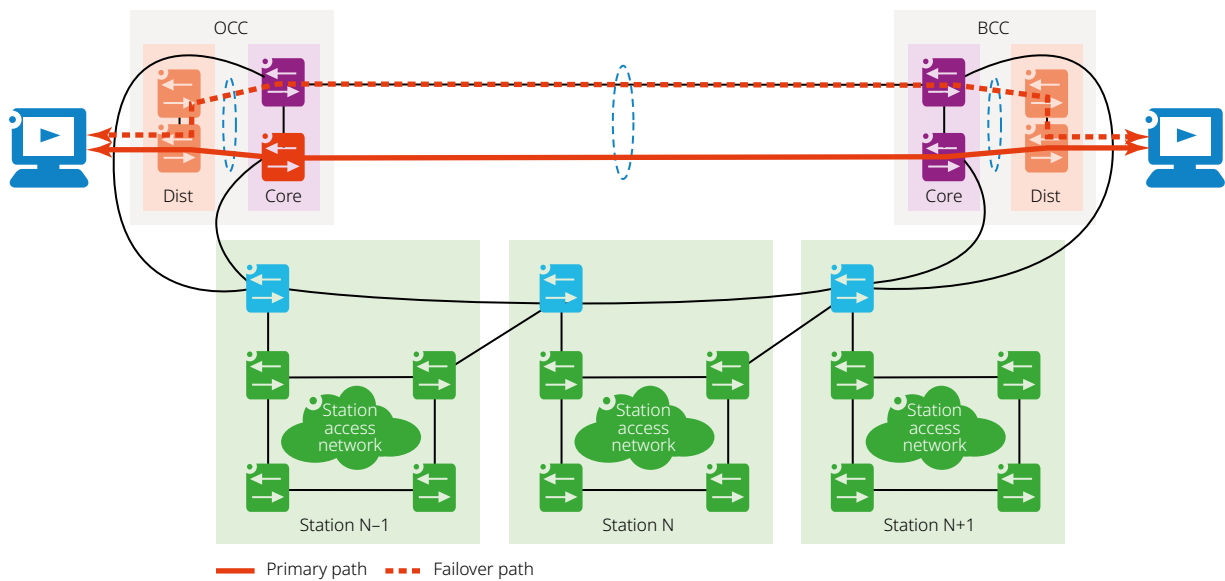


Figure 7. Station to OCC Primary and Failover Paths

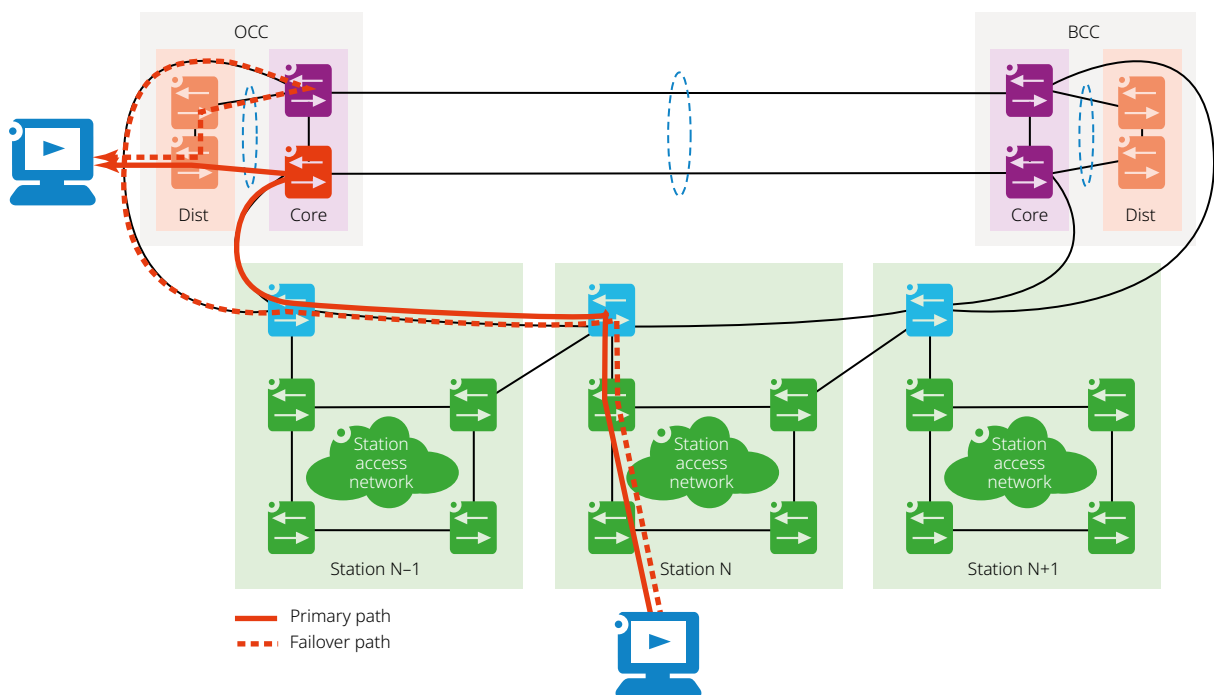


Table 4. L2 Convergence on OCC Core Node Master Down

Test Case: L2 Reconvergence on OCC Core Node Master Down							
	Direction	Service	VRF	VLAN	Transmit	Receive	Convergence Time
L2 Unicast	OCC->BCC	3001	App1	3001	10.0.1.101	10.0.1.102	<300 ms
L2 Unicast	BCC->OCC		App1	3001	10.0.1.102	10.0.1.101	<300 ms
L2 Multicast	OCC->BCC	3101	App1	3101	10.0.1.101/232.1.1.101	10.0.1.102	<300 ms
L2 Multicast	BCC->OCC	3101	App1	3101	10.1.1.122/232.1.1.122	10.1.1.121	<300 ms
L2 Broadcast	Station->	3055	App1	3055	10.0.55.101	Broadcast	<300 ms

Table 5. L3 Convergence on OCC Core Node Master Down

Test Case : L3 Reconvergence on OCC Core Node Master Down										
	Direction	Transmit				Receive				Convergence Time
	Service	VRF	VLAN	IP	Service	Service	VRF	VLAN	IP	
L3 Unicast Route Leaking	OCC->Station	3201	App1	3201	10.2.1.101	3202	App2	3202	10.2.2.101	<300 ms
L3 Unicast Route Leaking	Station->OCC	3202	App2	3202	10.2.2.101	3201	App1	3201	10.2.1.101	
L3 Multicast	OCC->BCC	3302	App2	3302	10.3.2.101/ 232.3.2.101	3402	App2	3402	10.4.2.101	
L3 Multicast	BCC->OCC	3602	App2	3602	10.6.2.101/ 232.6.2.101	3502	App2	3502	10.5.2.101	

5. Design Guidelines

The detailed design was finely tuned based on extensive testing performed at the Alcatel-Lucent Enterprise Solutions Lab. This section will summarize guidelines and recommendations.

- When a BEB or BCB node is a virtual chassis, all units in the VC should have physical links connecting to all SPB neighbors and physical links between the VC and each of the SPB neighbors should be configured as a LACP aggregate. This reduces the need to update tables in the event of VC unit failure and significantly improves the convergence time.
- Even though 16 BVLANS for 16 equal cost trees (ECTs) are supported, there are only 2 different paths in a ring topology and therefore only 2 BVLANS are needed. Every BVLAN builds its own SPT which consumes resources such as table space and CPU cycles. Configuring more than 2 BVLANS would waste those resources.
- An ISID is a broadcast domain. It is recommended that different VLANs be mapped to different ISIDs on a 1:1 basis. This ensures no bridging between VLANs is possible. Mapping multiple VLANs to the same ISID can have undesirable consequences and is only viable in special cases.
- Even though it is possible to group both core nodes (OCC and BCC) in a single VC, this is not recommended. OCC and BCC must have independent Control and Management planes so that control plane failures do not affect both sites simultaneously and service-level redundancy is maintained. This control plane independence facilitates maintenance tasks such as firmware updates which can be performed separately at each site. Use of VC is however recommended for use at each site separately.
- Tandem replication is recommended for services with high multicast traffic load such as video surveillance. Head-end replication is recommended for services with low multicast traffic loads.
- Station access VLANs and services should not be shared among all stations as this would create large broadcast domains.

6. Conclusion

This technical case study has shown with a practical example how ALE's Shortest Path Bridging technology can meet the requirements in the transportation vertical.

SPB natively provides MPLS-like VPN services but is comparatively cheaper and simpler to deploy and maintain. Because of this simplicity, an ALE powered SPB solution offers a lower total cost of ownership.

7. Acronyms

BCB	Backbone Core Bridge
BCC	Backup Control Center
BEB	Backbone Edge Bridge
BVLAN	Backbone Virtual LAN
ECT	Equal-Cost Tree
IGMP	Internet Group Management Protocol
ISID	Service Identifier
IS-IS	Intermediate System to Intermediate System Routing Protocol
LACP	Ling Aggregation Control Protocol (802.3ad)
LAG	Link Aggregation Group
LBD	Loopback detection
MAC-IN-MAC	Ethernet in Ethernet framing as (802.1ah)
MPLS	Multiprotocol Label Switching
OCC	Operational Control Center
Q-in-Q	Double VLAN tagging (802.1ad)
PIM	Protocol Independent Multicast
PIM-DM	PIM Dense Mode
PIM-SM	PIM Sparse Mode
PIM-SSM	PIM Source-Specific Multicast
RP	Rendezvous point
(S,G)	Source and group - identity of a source specific tree
(* ,G)	Any source and group - identity of a shared tree
SP	Shortest Path
SPB	Shortest Path Bridging
SPT	Shortest Path Tree computed by one ECT algorithm
STP	Spanning Tree Protocol
VC	Virtual chassis
VPN	Virtual private network

8. Related documents

[1] RFC 3569, "An Overview of Source-Specific Multicast (SSM)."