INTERNET-DRAFT
Intended Status: Informational
Expires: <September 24, 2012>

Roger Lapuh
Paul Unbehagen
Avaya
Peter Ashwood-Smith
Huawei
Phillip Taylor
Leeds Metropolitan University

March 23, 2012

SPB Deployment Considerations
draft-lapuh-spb-deployment-01

Abstract

Based on live deployments and three interoperability events, this document provides advice to network operators about best practices when implementing IEEE 802.1aq Shortest Path Bridging (SPB) networks. It is principally addressed to system integrators and solution providers, including those that do not yet support SPB. Some advice to implementers is also included. The intention of the advice is to facilitate multi vendor network deployments as well as provide guidance for new installations.

Status of this Memo

Copyright and License Notice

Table of Contents

1 Introduction

This document provides a set of recommendations and reference points for the deployment of IEEE 802.1aq - Shortest Path Bridging (SPB) networks based on MAC in MAC encapsulation. It focuses on the key network design items and does not go into describing the protocol details.

The IEEE 802.1aq standard has been technically frozen since early 2011 before several multi vendor interoperability events had taken place, thus the recommendations described here are valid despite the minor editorial work that has caused non technical change in the IEEE base standard since.

1.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

1.2 MOTIVATION AND BACKGROUND

This document provides a checklist of recommendations which are based on multiple documented multi vendor Interoperability tests [SPBWIKI] and more than 12 months of production deployment experiences. It summarizes the learning's and experience acquired during those activities. New SPB installations can benefit from following the recommendations below.

2. General Deployment Recommendations

All the following described deployments have shown sub second convergence times in case of link or nodal failures within the SPB fabric.

Recommendation 1:

To achieve this, strictly connection oriented point-to-point interfaces are used, and shared segments between SPB fabric nodes have been avoided. Ethernet based mechanisms are used to detect link faults quickly to trigger shortest path calculations in case of a link and nodal failure.

The end-point-only provisioning for network virtualization with SPB has proven very effective in many of the installations described below.

Recommendation 2:

SPB's service ID (I-SID) with its (24 bit) addressing space has helped to keep the VLAN numbering (1 to 4095) local to the respective access network region (e.g. Data Center), avoiding the complexity of managing a global VLAN space out of a range of only 4096. It is recommended to define a global virtualization schema based on I-SIDs, and not tie VLAN ids directly to ISIDs ids in a 1 to 1 relationship throughout the network.

Recommendation 3:

It has been seen that using SPB to keep Spanning Tree regions local to access networks (therefore reducing the impact of network changes) has significantly improved the end user experience.

Besides the need for L2 traffic virtualization for hypervisor migrations, all the deployments were also required to route the virtualized traffic between IP subnets/broadcast domains which are provided with an SPB service.

Recommendation 4:

Routing between services can be done with dedicated routers external to the SPB fabric, but it would be an advantage if SPB nodes could route traffic between services similar to traditional routing switches that are able to perform routing between VLANs/IP subnets without having to leave the Ethernet fabric.

3. INFRASTRUCTURE CONFIGURATION RECOMMENDATIONS

3.1 **IS-IS SYSTEM ID** AND SPB NICKNAME CONFIGURATION RECOMMENDATIONS

As of this writing the IEEE SPB standard defines a single ISIS area for an SPB region, even though large SPB regions can be defined and operated. In the future this will likely be extended to multi-area support.

Recommendation 5:

The interoperability events have shown it is a good practice to manually configure System IDs and SPB Nicknames with a simple

identification scheme, coordinating the system ID numerically with the SPB Nickname for ease of troubleshooting. It is also a good practice to define an area per SPB region. It is recommended to use the upper 24 bits to indicate the area ID. For example, System IDs start with 4900.bb00.1000 for the first node, 4900.bb00.2000 for the second node and so on. In these System IDs, 49 indicates a private address, the "00bb" indicates area "00bb", and 1000, 2000, etc., indicate the node number (1 through n). These System IDs correspond to SPB Nicknames of 1.bb.10, 1.bb,20, 1.bb.30 for nodes 1, 2 and 3 respectively, and so on. Using manually configured BMAC addresses and also coordinating the BMAC with the System ID and SPB Nickname enhances-, ease of identification for management and troubleshooting as well as making possible future PBB-EVPN.

As an additional option, with the goal to reduce configuration tasks, System-IDs could be automatically inherited from the systems chassis MAC addresses. The SPB Nickname could also be derived from the lower bits of the chassis MAC. This approach could be targeted for SPB access switches where a simplified deployment model would be of most interest.

3.2 SPB FABRIC INTERFACE TYPES

Details on Recommendation 1:

SPB Fabric inter-connections in the preceding SPB deployments are all based on point-to-point Ethernet links, optical CWDM/DWDM connections or some sort of transparent E-LINE service. By avoiding connecting SPB over a shared segment (or E-LAN) failure detection and network convergence times have been kept very low. Failure detection and recovery is thus not dependent on IS-IS hello-multiplier intervals but triggered by lower layer protocols.

Such E-LINE services (to interconnect SPB nodes) can be based on any type of transparent Ethernet service (MPLS or PBB based), as long as they are loop free and the service Maximum Transmission Unit (MTU) size allows for a minimum of [MTU] 1544 bytes for non Jumbo Frames.

Tagged IP: = 1522 = 1500(IP MTU)+ 2(Ethertype)+ 12(MAC SA/DA) + 4(TAG) + 4 (CRC) and MacInMac Header = 22 bytes.

On dark-fiber based Ethernet connections, link failures can be detected by the Ethernet remote fault detection mechanisms; however, on service provider based links, there can be multiple active components between two SPB nodes, and thus not all failures can be detected easily. To ensure quick fail-over times across an E-LINE service, an end-to-end connectivity check mechanism such as 802.1ag based Connectivity Check Mechanism (CCM), or similar, is

recommended.

3.3 SPB FABRIC ACCESS

Details on Recommendation 3:

Many networks today still operate with some sort of Spanning Tree (MSTP/RSTP or proprietary versions). SPB can be leveraged to separate Spanning Tree regions into smaller independent domains. Therefore a Spanning Tree root bridge change impacts smaller regions only and is not spread across the whole network. Keeping root bridge elections and the effect of Topology Change Notifications local has proven a significant improvement of network availability in larger Spanning Tree deployments.

3.4 SPB Fabric configuration

Recommendation 6:

In an SPB network the Backbone VLAN IDs (BVIDs) are used to separate and load-spread SPB traffic across multiple paths. The 802.1aq standard defines up to 16 BVIDs. These BVIDs need to be consistently configured across the SPB region. The BVIDs can be selected out of the available VLAN range [1-4095], however, using a pre-defined set of VLANs is recommended.

Usually the lowest 4000 IDs are used by customers for network access VLAN configurations; thus it has been seen as a good practice to use BVLAN numbering that is in the highest upper addressable range, e.g, starting with 4050 for the primary BVLAN and all switches and 4051 to 4065 for the subsequent ones. It is recommended to use at least two BVIDs for load-spreading reasons.

3.5 SPB SERVICES MAPPING

Details on Recommendation 2:

When network virtualization needs to be extended between regions, for example, to support for Virtual Machine movements, it is very important to use a unique virtualization index to achieve this. SPB, with its 802.1ah based Service-ID (I-SID) provides an inherent virtualization technique which allows local VLAN significance and using the I-SID as a global virtualization index. This is especially true in VMWare deployments where it is advantageous for the Portgroup IDs of VCenter instances to correspond with the VM VLAN memberships. With SPB, it is thus easily possible to run a hosted environment with

multiple VCenter instances in parallel on the same infrastructure, without having any VLAN space interference. In the preceding virtual Data Center deployment where multiple domains are interconnected, the VLAN spaces can be kept independent of each other, and the virtualization is achieved by the usage of the I-SIDs.

3.6 SPB AND IP ROUTING

Details on Recommendation 4:

In an SPB network the typical size of the user and server subnets are not being changed from what one is used to with traditional technologies. This means that there is always a need for routing functionality. The best case is if a SPB node can directly route individual IP subnets which consist of I-SIDs, similar to how those nodes can route VLAN based IP subnets. Optimally this routing should be available within the SPB Ethernet fabric between I-SID based services.

4. STANDARD IMPROVEMENT RECOMMENDATIONS

Recommendation 7:

In an SPB network, link-state update propagation is achieved by SPB nodes re-laying topology change notifications through IS-IS on a hop by hop basis in the control plane. With an assumption of 10ms relay latency per node, a ring of 20 SPB nodes could see up to 100ms of propagation latency to reach all nodes in the ring. As an optimization of SPB, the default L2 service instance described in IEEE 802.1aq could be used to flood all propagation changes into this default service, reducing the propagation delay in this example from 100ms to 10ms. The SPB IEEE standard could be enhanced to also include this default-flooding behavior.

5. OA&M

In all deployment experiences, the use of L2 based OAM capabilities have been invaluable in managing the network.

Recommendation 8:

It is recommended that IEEE 802.1ag based connectivity check mechanisms: Layer 2 Ping, Layer 2 Traceroute and Layer 2 Tracetree are being implemented.

6. Tenant Separation Considerations

SPB separates any type of traffic at the edge of the SPB region into its own service instance (I-SID). Classification into the I-SID can be done based on port, vlan or a combination of port/vlan. Once a customer-or application traffic is classified into an I-SID, it is kept separate until it exits the SPB region, very similar to MPLS with its tunnel and service labels. In SPB the "tunnel label" is comprised of the BMAC pair and the "service label", which is the I-SID. Thus SPB is as secure as any other packet switched solution. Today there are many service provider based networks in production using the same 802.1ah (PBB) encapsulation methods as SPB is using.

7 Deployment Experiences

7.1 DEPLOYMENT SCENARIO A

SPB AS INTER-DATACENTER-FABRIC FOR DC REDUNDANCY OR DC MIGRATIONS

Typically in a large enterprise core, it is not viewed as good practice to extend L2 broadcast domains across the backbone network. However, with the advent of server virtualization, it has become a common requirement to extend server VLAN segments between geo-redundant Data Centers to dynamically, efficiently and cost effectively leverage the ability to perform Virtual Machine migrations and run load balancing techniques across multiple Data Centers. The deployment of SPB as a data center connect allows the following challanges to be addressed:

In many cases SPB can be deployed on the existing network architecture with IS-IS running side by side and independently from other routing protocols such as OSPF. OSPF is being used to populate the IP routing table and provide L3 routed connectivity. IS-IS is being used for SPB, bringing the ability to extend server VLANs across the backbone. Typically the server VLANs to be extended across the network are locally configured within the Data Centers, on the server aggregation Top of Rack switch(es) as well as on the distribution layer nodes for the Data Center which aggregate the Top of Rack switches. On the distribution nodes, the server VLANs are assigned to a service ID (I-SID) and thus extended across the SPB network, becoming available in the other data center. Access redundancy is provided with an active-active model which ties the SPB core region to the VLAN based access region. The same distribution nodes can act as a routing gateway for the server VLANs. VRRP is also being used to create a single default gateway IP address for the server VMs. The VRRP instance per Server IP-subnet thus exists on all distribution nodes and provides redundant and distributed default

gateway functionality. Core failure recovery times in the SPB region can be kept well below 1 second and L3 recovery times, depending on the configured VRRP timers.

7.2 DEPLOYMENT SCENARIO B

SPB TO RE-ARCHITECT SECURITY ZONES

There are other valid reasons why it might be necessary to extend L2 segments across the enterprise core. A good example is a major manufacturing plant which has a very rigorous design based on a pure IP routed architecture with a strong focus on firewalling different parts of the network. This is achieved by physically wedging firewalls within the physical topology in such a way as to deny any unwanted interaction between different network zones. The security provided by this model has to be offset by the rigidity it imposes in terms of where devices are allowed to be connected to the network. In this particular example, connecting devices in locations where they were not initially intended to be located was addressed by laying additional cabling, with the costs and delays that this involves. Once deployed, SPB brought to this model the ability to decouple the physical infrastructure from the logical connectivity running above it. This means that it is no longer necessary to wedge firewalls into the physical topology to intercept traffic, but rather let SPB force L2 VLANs to reach the desired firewalls, wherever those firewalls might be located on the network. It is now possible to connect devices anywhere on the physical network infrastructure and simply connect these devices to the VLAN segment to which they need to belong.

7.3 DEPLOYMENT SCENARIO C

SPB FOR CAMPUS VIRTUALIZATION

Another example of where it is useful to extend L2 segments can be found in the health care vertical. An operational challenge, typical of most hospitals, is to be able to support network connectivity for mobile medical equipment which typically needs to connect to a server application hosted in the Data Center. The real challenge with this equipment is often the fact that it is supplied and maintained by separate, often external, technicians with little or no IP skills. As such this equipment is usually not able, or not configured, to use DHCP and instead uses a single flat IP subnet which encompasses the mobile units as well as the server application in the Data Center. The hospital's network team essentially has limited control over the

IP configuration of these devices and hence a desire to segregate such applications within a constrained L2 service. By deploying SPB L2 instances, it is now possible to much more easily manage such applications.

7.4 DEPLOYMENT SCENARIO D

SPB AS MULTI-TENANT FABRIC SOLUTION

In a multi-tenant deployment SPB was leveraged to provide secured and separated services for several tenants. In this implementation SPB leverages 10 Gigabit Ethernet heavily. In the two geo-disbursed data centers LAN and IP connectivity is utilized in a way that makes both appear as one virtual data center. A common 3-tier design is utilized for the entire network. There may be multiple tenants per edge which are then segregated into their own private broadcast domain.

Over 500 L2 services are spread across the network providing IP subnet connectivity to any of the tenants. At the data center those IP subnets are assigned to over a dozen of Virtual Router Forwarding (VRF)instances corresponding to their security requirements. VRRP is used to provide router redundancy.

Layer 3 routing between VRF instances, hence between tenants, to external organizations and to the Internet is performed by stateful firewalls.

This simplified model utilizing Layer 2 I-SIDs, including routing between service instances, across a common SPB backbone allows this solution provider to quickly and effectively extend either Layer 2 services or Layer 3 services to any location in the network for any application.

For Voice over IP a Quality of Service (QoS) framework for traffic prioritization has been employed. IP Differentiated Services (DiffServ) EF DSCP and several specific AF DSCP groups are mapped into the appropriate 802.1p priority classes at the SPB BEB nodes to provide the necessary traffic prioritization within the SPB backbone.

8 Security Considerations

Security implications of SPB deployments are to be discussed in separate documents.

9 IANA Considerations

This document makes no requests to IANA

5 References

5.1 Normative References

5.2 Informative References

[RFC6329] Fedyk, D., "IS-IS Extensions Supporting IEEE 802.1aq Shortest Path Bridging" July 2011

[SPBWIKI] http://en.wikipedia.org/wiki/Shortest_Path_Bridging

Authors' Addresses

Roger Lapuh (editor)
Avaya
Wallisellen, 8304 Switzerland
EMail: rogerlapuh@avaya.com

Paul Unbehagen
Avaya
1300 W. 120th Avenue
Westminster, CO 80234 USA
Email: unbehagen@avaya.com

Peter Ashwood-Smith (editor)
Huawei Technologies Canada Ltd.
303 Terry Fox Drive, Suite 400
Kanata, Ontario, K2K 3J1 CANADA
EMail: Peter.AshwoodSmith@huawei.com

Phillip Taylor
Leeds Metropolitan University
404 Portland Building
Calverley Street
Leeds, LS1 3HE, UK
Email: p.p.taylor@leedsmet.ac.uk

Steven Emert
Avaya
225 South Sixth Street, Suite 4350
Minneapolis, Minnesota 55402-4619 US
Email: semert@avaya.com

Ludovico Stevens (editor)
Avaya
25 Allee Pierre Ziller
06560 Valbonne France
EMail: ludovicostev@avaya.com

Srikanth Keesara
Avaya
600 Technology Park
Billerica MA 01821 US
EMmail: skeesara@avaya.com