

DEMYSTIFYING H-VPLS

Understanding How Hierarchical Virtual Private LAN Services Address Service Provider Needs and Impact Networks

Table of Contents

Introduction	1
VPLS: The Basics	1
What Is H-VPLS?	1
What Is H-VPLS Being Used for Today?	3
“Creative” Applications for H-VPLS	4
A Better Approach to Scale VPLS Services	6
Control Plane Perspective	6
Forwarding Plane Perspective	6
Service Reachability	7
Carrier-Class Ethernet Infrastructure	7
Summary	8
Reference Documents	9
About Juniper Networks	9

Table of Figures

Figure 1: Active and back-up paths	2
Figure 2: Multicast replication at ingress	3
Figure 3: Traffic sent N time over the same link of the ring	4
Figure 4: Dual paths and Spanning Tree	5
Figure 5: Point-to-Multipoint LSP	6
Figure 6: Quadrant comparison	8

Introduction

Virtual private LAN service (VPLS) has become a very attractive technology over the past few years, as it allows service providers to deploy carrier-class services over an Ethernet-based network in a reliable and flexible way. Starting mainly with business services and now with broadband multiplay services, service providers have gotten deployment experience with it, and have also found some of the challenges that this technology brings, especially in terms of scalability. In many cases, the initial answer to those challenges has been hierarchical virtual private LAN service (H-VPLS).

Sometimes technologies tend to take on a life of their own, becoming “the” answer to all problems in the network. This is, to some extent, what is happening with H-VPLS.

There are cases where we could say H-VPLS is thought to be the breakthrough technology that will address all of the network’s problems. But in most cases, the reality is that it will either not address them or not address them in the most effective way, and in the process, will often generate additional problems that should not be minimized.

The purpose of this document is to demystify the use of H-VPLS, explaining its value, identifying the problems it really solves and those it does not, and discussing other approaches or technologies that could address the problems that H-VPLS does not solve. For those not familiar with VPLS, let’s start with some basics on its terminology and the different implementation options.

VPLS: The Basics

VPLS is one of the key MPLS-based services that has developed in the industry over the past few years. As its name implies, the purpose of VPLS is to provide a private multipoint LAN-type Ethernet connectivity service. For those more familiar with technologies like ATM, we could say VPLS is the LAN emulation service for MPLS.

VPLS has special relevance in the service provider space as the way to deliver Layer 2 (L2) multipoint transparent services over an Ethernet infrastructure using MPLS. But what is so special about this? The key point is MPLS. There are different ways or approaches for a service provider to deliver services over an Ethernet infrastructure, but not all of them fit into the requirements that a service provider has in terms of scalability, reliability, service flexibility, and operational complexity. MPLS has become the catalyst that can turn an Ethernet infrastructure into carrier class making it suitable for the service provider, as opposed to a VLAN-based or QinQ operation that has demonstrated through multiple examples that it does not provide what is required in the carrier environment.

VPLS, as the main technology in use in the Metro Ethernet space, has two implementation options that the industry has standardized:

- RFC4761 – BGP-based VPLS
- RFC4762 – LDP-based VPLS

Both of these implementation options are virtually identical from the forwarding plane perspective, but they differ in the control plane, particularly in the protocol they use to signal and establish the pseudowires, BGP or LDP. It is not the intention of this document to discuss the pros and cons of each approach. RFC4762 defines a hierarchical mode of operation for LDP VPLS called H-VPLS.

What Is H-VPLS?

H-VPLS is the hierarchical version of LDP-based VPLS as described in RFC 4762 - Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling. As it is explained in the RFC:

.....

The solution described above requires a full mesh of tunnel LSPs between all the provider edge (PE) routers that participate in the VPLS service. For each VPLS service, $n*(n-1)/2$ pseudowires must be set up between the PE routers. While this creates signaling overhead, the real detriment to large scale deployment is the packet replication requirements for each provisioned pseudowires on a PE router. Hierarchical connectivity, described in this document, reduces signaling and replication overhead to allow large-scale deployment.

.....

.....

In essence you could say that H-VPLS tries to address two different issues:

- The inherent scalability issues of VPLS
- The possibility to extend the VPLS domain to lower end places enabling the use of simpler (cheaper) devices

H-VPLS defines essentially two new roles or functionalities:

- **PE-rs:** This is basically a PE with all its functionality on the VPLS architecture that runs VPLS with other PE-rs (following the same VPLS rules) but which also has pseudowires (it can be based on QinQ access, but this will not be covered) with other devices called MTU-s (the access layer essentially).
- **MTU-s:** This represents the access layer on the H-VPLS architecture and establishes pseudowires to one or two PE-rs through which VPLS traffic is forwarded.

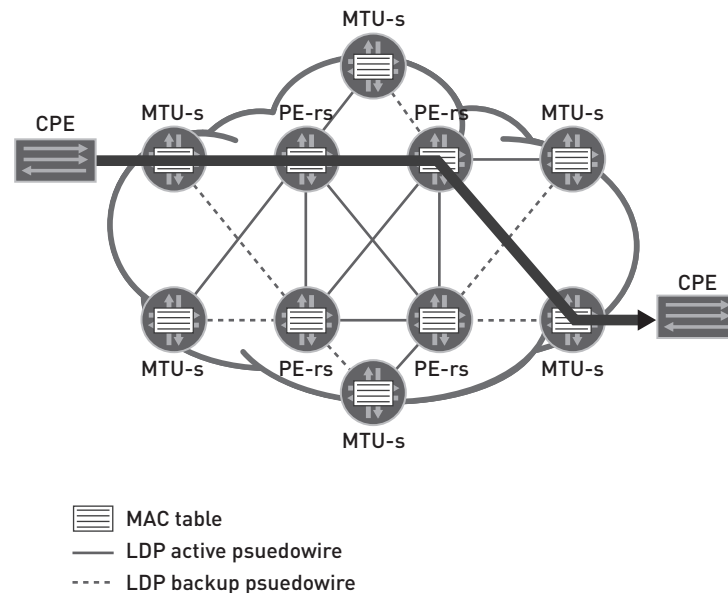


Figure 1: Active and back-up paths

H-VPLS mode of operation: The mode of operation between PE-rs is like normal VPLS. The operation between MTU-s and PE-rs is such that the PE-rs treats the pseudowires like access links, or said differently, the split horizon rule does not apply. If something is received at a PE-rs from an MTU-s, it will be forwarded to the other PE-rs's and to the other MTU-s's connected to the same PE-rs. If something is received at a PE-rs from another PE-rs, it will be forwarded to the MTU-s's connected to it through a pseudowire, but not to the other PE-rs (here the split horizon applies).

By means of this mode of operation, H-VPLS tries to make VPLS more scalable, as you will see later in the document. However, this requires PE-rs's (which would normally be P routers that have no VPLS state) to maintain media access control (MAC) tables and to perform VPLS operations of learning and flooding. Moreover, a PE-rs has to do this for all of the MTU-s's that it has. This could lead to data plane scaling problems, especially in terms of the number and sizes of MAC tables. In summary, H-VPLS creates a control plane hierarchy (in the form of MTU-s and PE-rs), but at the expense of forcing hierarchy in the data plane as well. Therefore, in the process of solving one scalability problem, H-VPLS introduces a new scalability problem, and it does not provide solutions for this new problem.

It is important to highlight, however, that the possibility to extend the VPLS domain to cheaper/simpler devices by establishing pseudowires into a centralized/semi-centralized PE-rs, which is one of the main motivations for using H-VPLS, is not an exclusive capability of LDP-based H-VPLS. Routers may have this capability regardless of the signaling protocol that is being used. This is indeed possible with Juniper Networks® devices, both using LDP or BGP.

What Is H-VPLS Being Used for Today?

Therefore, based on the previous section, H-VPLS is the straight-through mechanism to resolve the N-square scalability problem of VPLS.

This architectural issue of VPLS is reflected in the implementation limits of each vendor, both in the control plane and in the forwarding plane:

- How many targeted LDP sessions does a certain PE support?
- How many hardware ingress replications can be done on a PE?

From the forwarding plane perspective, it is not only the number of hardware replications that the device can perform, but the service implications that it may have:

- If a frame needs to be sent N times over the same link because of the ingress replication, then it will consume N times the expected bandwidth. Therefore, the resource consumption will be much higher.
- If a frame needs to be sent N times over the same link, the Nth frame may have an additional delay of $(N-1) * MTU * 8 * BW$.

Multicast-traffic replicated by Ingress PE

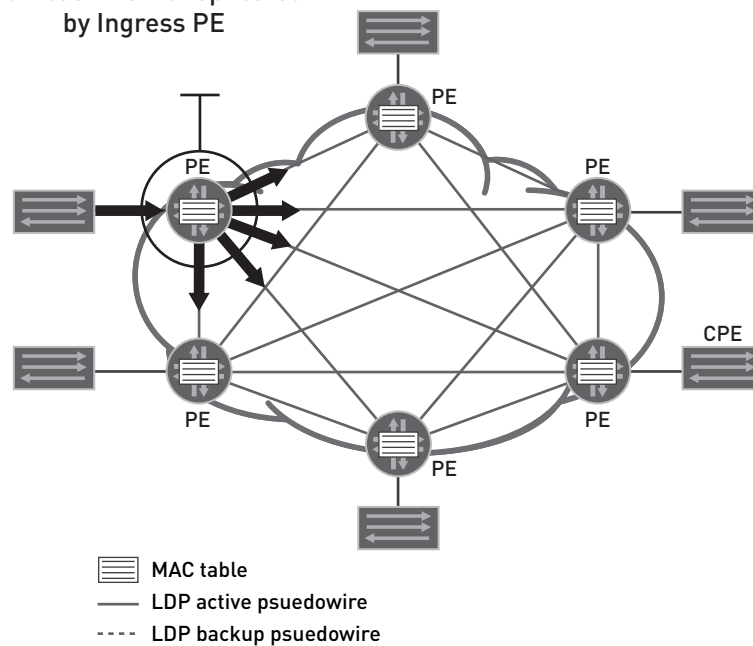


Figure 2: Multicast replication at ingress

- Service provisioning: If you need to add a new site to a VPLS instance, and if that site is on a new PE for that VPLS, you have to update all other PEs. A management system can be developed to do this, which will have its own cost, but changing the configuration on a device can always result in unforeseen consequences.

Therefore, H-VPLS tries to address not only the architectural scaling issues of VPLS, but also the vendor-specific implementation limits of the platforms.

“Creative” Applications for H-VPLS

In addition to the previously described motivations for using H-VPLS, we are seeing how H-VPLS is being positioned and used as the response to LDP VPLS Forwarding Plane “Achilles Heel”: Handling multicast traffic (LDP VPLS Control Plane “Achilles Heel” is the lack of auto-discovery).

It is well known that VPLS relies on the replication of traffic on the ingress PE to deliver the “multipoint” capability of VPLS. This happens for L2 multicast traffic, L2 broadcast traffic, and unknown unicast traffic.

Even though from the service and end customer’s perspective, the service is actually multipoint, the inherent point-to-point nature of this scheme produces serious inefficiencies in the network. The worst case scenarios are the ring topologies, as can be observed in the following diagram, where traffic may need to be replicated over the same link as many times as nodes exist on the ring.

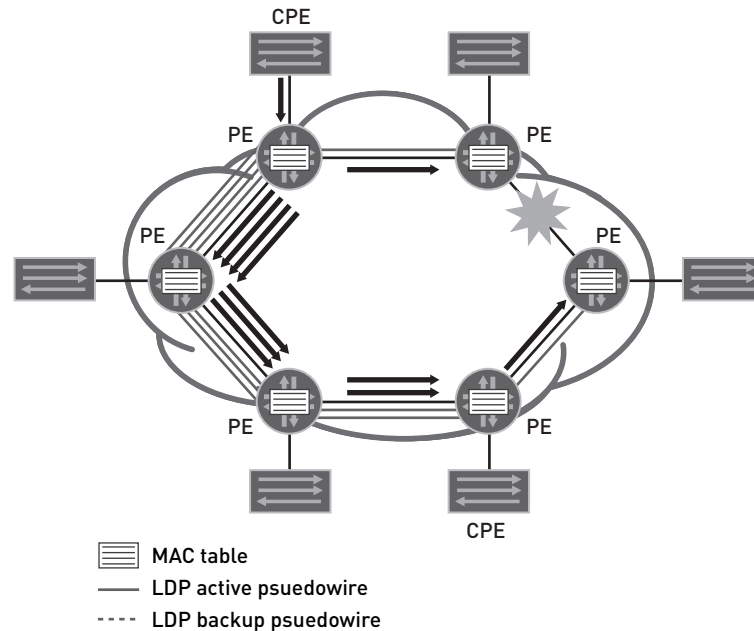


Figure 3: Traffic sent N time over the same link of the ring

So based on these performance attributes, some vendors have found H-VPLS to be “the answer” for this VPLS inefficiency.

Sometimes it is proposed that H-VPLS mechanisms be used to create “artificial” replication points on a VPLS-based network such that it is possible to minimize the amount of ingress replications required to transport multicast traffic over VPLS. The way to create this is by configuring these nodes as H-VPLS hubs, but making them interpret that the device on the other end is an H-VPLS spoke, so effectively the split horizon rule is not applied, and traffic received on a pseudowire is forwarded to the next replication point.

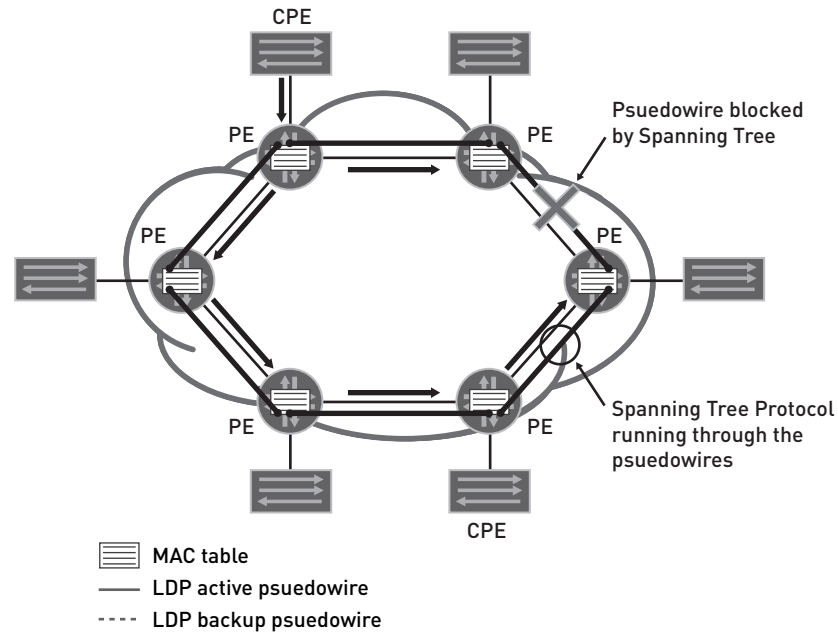


Figure 4: Dual paths and Spanning Tree

This architecture defines single points of failure, in addition to creating unnecessary MAC learning points (which lead to additional scaling issues that require yet another extension to be addressed, as described in draft-cowburnl2vpn-vpls-ldp-mac-hiding-01). The only way to eliminate the single points of failure in these cases is to create dual paths that are somehow connected. Because the service is a multipoint L2 service, this means creating the conditions where L2 broadcast storms may happen. Therefore, for this scenario, some vendors are proposing the use of Spanning Tree Protocol (STP). Interestingly, Ethernet is coming back to its origins.

If service providers discarded STP (and the likes) as a protocol suitable for carrier-class deployments (in many cases after painful experiences), why should they now use Spanning Tree Protocol again for resolving the L2 loops that are created as a result of using H-VPLS (or more particularly the no-split-horizon rule that applies to the communication between H-VPLS spokes and hubs) in a way that was not intended? Needless to say, it creates an additional state on the network due to artificially enabling the bridging function to nodes on the network that were only supposed to do P functionality.

If we look at this whole picture, what is being created is a plain VLAN L2 switching architecture with STP as control protocol where the links between the bridges/switches are not Ethernet interfaces but MPLS pseudowires. So, effectively, this means going back to the Spanning Tree days. Is this the way to go? Should service providers rely on this architecture to transport their IPTV traffic? Each will need to make this judgment based on business requirements.

A Better Approach to Scale VPLS Services

Scalability for any type of service or technology can be viewed from different perspectives, and the more that this is analyzed, the more likely you will achieve the result required. A service provider running VPLS technology or willing to provide VPLS services should care about scaling from the following perspectives:

- Control plane
- Forwarding plane/bandwidth efficiency
- Service reachability (regional, national, international)

Control Plane Perspective

LDP-based VPLS clearly has limitations in its scalability, due to the fact that it is point to point in nature. In LDP-based VPLS, the LDP full mesh is tightly integrated with the pseudowire full mesh required on the forwarding plane. This way, in order to scale the control plane, VPLS requires that the forwarding plane be modified, effectively breaking the forwarding plane into different domains (which is what is done by means of the H-VPLS PE-rs 's). The control plane is also separated into different domains, reducing the N-square stress.

In BGP-based VPLS, the control plane and the forwarding plane are decoupled (see section 3.6 on RFC 4761 Hierarchical BGP VPLS). This means that it is possible to implement mechanisms in the control plane itself to make it scale (use of route reflectors is the most common one) without introducing any change in the forwarding plane. BGP is inherently hierarchical, so the service provider can take advantage of this to scale the control plane. BGP VPLS does not require the route reflectors to maintain MAC tables and do VPLS data plane operations.

Forwarding Plane Perspective

Scaling the forwarding plane will typically be forced by the bandwidth consumption that the ingress replication on VPLS implies. As already described before, we believe the most efficient way to scale the forwarding plane is to add the capability to VPLS to use point-to-multipoint communication label switched paths (LSPs) as opposed to using H-VPLS mechanisms.

If you really think about the cause of the problem itself, you see that it is because of the point-to-point nature of the MPLS LSPs used by VPLS. However, MPLS already resolved this some time ago by the definition and implementation of point-to-multipoint LSPs, either based on LDP or RSVP-TE (each one with its advantages). So why not go to the origin of the problem and solve it? This is what Juniper has been proposing by integrating the use of point-to-multipoint LSPs with VPLS in such a way that broadcast, multicast, and unknown unicast will be forwarded using point-to-multipoint LSP, which will only be replicated on the network where the network really requires it. By adding this capability to VPLS, network and bandwidth efficiency will be achieved, while at the same time preventing the increase of the network complexity with unnecessary MAC/bridging nodes and/or protocols such as Spanning Tree Protocol.

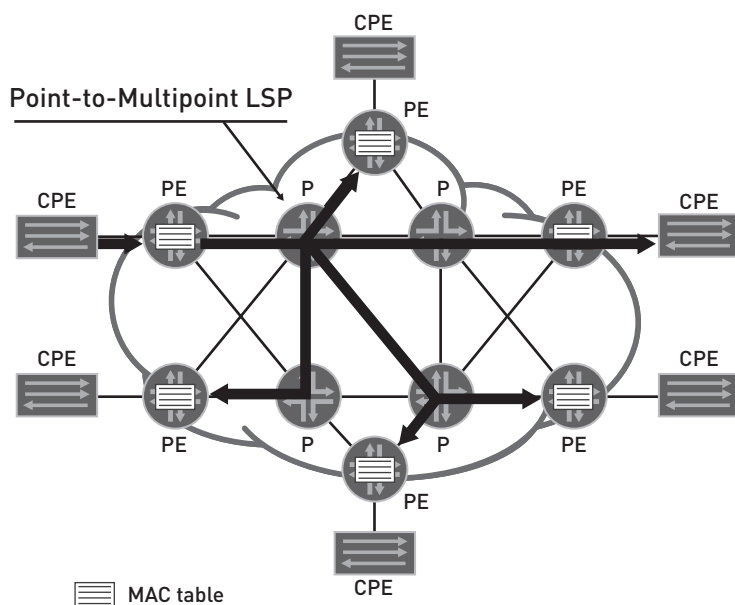


Figure 5: Point-to-Multipoint LSP

Service Reachability

Once you assume that the control plane and the forwarding plane have reasonable scalability properties, the next step is determining how far you can provision your VPLS-based services: Within one region of the country? Nationwide? Internationally? Depending on the answer to this question, a single network frontier may need to be trespassed to provide the service, so Inter-AS VPLS may be required.

As specified in RFC 4762:

10.3. Multi-domain VPLS Service

Hierarchy can also be used to create a large-scale VPLS service within a single domain or a service that spans multiple domains without requiring full mesh connectivity between all VPLS-capable devices. Two fully meshed VPLS networks are connected together using a single LSP tunnel between the VPLS “border” devices. A single spoke pseudowire per VPLS service is set up to connect the two domains together.

When more than two domains need to be connected, a full mesh of inter-domain spokes is created between border PEs. Forwarding rules over this mesh are identical to the rules defined in Section 4.

This creates a three-tier hierarchical model that consists of a hub- and-spoke topology between MTU-s and PE-rs devices, a full-mesh topology between PE-rs, and a full mesh of inter-domain spokes between border PE-rs devices.

BGP-based VPLS, on the other hand, has been designed to support Inter-AS service extension by defining several options, which include redundancy (more details in section 3.5 of RFC 4761), and defining how efficient multicast forwarding can also be extended in an Inter-AS scenario with VPLS (draft-ietf-l2vpn-vpls-mcast-01.txt).

So if you look at the previous dimensions of “scalability,” you can conclude that BGP-based VPLS with point-to-multipoint LSPs can be a better path for scalability compared with LDP-based H-VPLS.

Carrier-Class Ethernet Infrastructure

Ethernet has proven to be a key technology for the service provider’s next-generation services, as it provides the bandwidth and the cost point required for scalability and profitability. There are, however, several ways of using Ethernet to provide services, and some have been described in this document. These different techniques do not have the same properties, and not all of them can be considered carrier class.

Among these properties, we can find: scalability, flexibility, network efficiency, reliability, and operational complexity. We have already analyzed the differences in terms of scalability between these solutions. In terms of flexibility, there is probably not a big difference among all the VPLS-based technologies, so you could probably say that LDP VPLS, H-VPLS, and BGP VPLS have similar properties in terms of flexibility with the exception of the ability to extend the service through other networks, where clearly BGP VPLS has an advantage.

However, let’s analyze the other three carrier-class properties because they have a direct impact on the service provider’s costs:

- **Network efficiency:** A technology that does not use the network resources efficiently will result in additional costs for the service provider, either because extra equipment is needed, or additional links, or it consumes excess bandwidth. So, network efficiency has a clear impact on capital expenditures (CapEx), and probably on operating expenditures (OpEx) as well.
- **Reliability:** The lack of reliability has a direct link with OpEx, so it is clearly a requirement for a carrier-class infrastructure.
- **Operational complexity:** The more difficult a technology is to operate, either because it requires many provisioning points, or many protocols to handle, or particular cases to deal with, the higher the operating expense will be for the service provider.

Therefore, from the perspective of these three attributes, a real carrier-class technology should offer the service provider the following: high network efficiency, high reliability, and low operational complexity. There are always trade-offs and in many cases mutual dependencies exist among them, so increasing one might decrease the other.

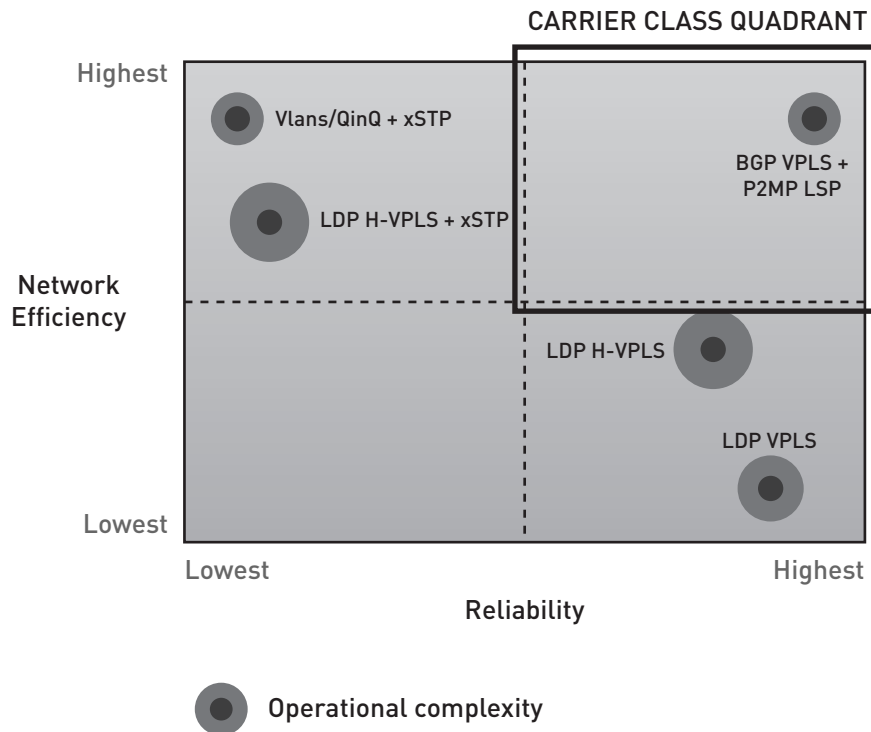


Figure 6: Quadrant comparison

If you analyze, under this framework, the different technologies a service provider can select to operate an Ethernet network, you come to the following conclusions:

1. Although plain VLAN with STP has the least operational complexity among all of the technologies and is very efficient in its use of network resources, the lack of reliability of its control plane clearly makes it non-carrier class, not to mention its scalability limitations, which are not reflected in this diagram.
2. There are a set of trade-offs that play across all of the LDP VPLS/H-VPLS solutions. VPLS provided some additional complexity and much more reliability for a service provider, but at the expense of losing network efficiency. H-VPLS brings some more network efficiency than full mesh VPLS, but because it creates single points of failure, it reduces its reliability while it increases its complexity. H-VPLS, combined with Spanning Tree Protocol, brings a higher level of network efficiency, but at the expense of substantially increased operational complexity and reduced reliability, as it introduces STP again into the equation. Therefore, there is a set of trade-offs that does not allow this technology to enter into the Carrier-Class Quadrant.
3. The only way to break through this barrier is by working with the real driver of network efficiency, which is the point-to-multipoint LSPs. BGP-based VPLS with point-to-multipoint LSPs is the only technology that offers the service provider the highest network efficiency while at the same time maintaining network reliability. This technology also decreases the operational complexity of LDP VPLS by adding autodiscovery on the network, which allows not only the members of the VPLS instances to know each other, but also allows the automatic establishment of the point-to-multipoint LSPs without additional provisioning overhead. You could also add to this analysis the additional operational synergies that BGP-based VPLS offers as most service providers will also be offering BGP-based IP-VPNs.

Summary

H-VPLS tries to address some of the scalability problems that have been described about VPLS. We believe that there are better ways to address them than those defined by LDP-based H-VPLS.

Specifically, if we think about efficiently addressing multicast traffic in a reliable way, we are convinced that H-VPLS does not offer the solution required by service providers.

We believe that H-VPLS is a small step, and not necessarily in the right direction, towards what we would call scalable VPLS service: S-VPLS, which is the real goal for service providers. If we need to find a technology that can really get service providers closer to S-VPLS, then that would be BGP-based VPLS with point-to-multipoint LSPs.

Reference Documents

- M. Lasserre, V. Kompella. RFC4762 *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*
- K. Kompella, Y. Rekhter. RFC4761 *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*
- R. Aggarwal, Y. Kamite, L. Fang. *Multicast in VPLS*, draft-ietf-l2vpn-vpls-mcast-01.txt
- R. Aggarwal, Y. Kamite, L. Fang. *Propagation of VPLS IP Multicast Group Membership Information*, draft-raggarwal-l2vpn-vpls-mcast-ctrl-00.txt
- E. Rosen et. al. *Provisioning, Autodiscovery, and Signaling in L2VPNs*, draft-ietf-l2vpnsignaling-08.txt
- R. Aggarwal, E. Rosen, Y. Rekhter, T. Morin, C. Kodeboniya. *BGP Encodings for Multicast in 2547 VPNs*, draft-ietf-l3vpn-2547bis-mcast-bgp-02.txt
- R. Aggarwal et. al. *Extensions to RSVP-TE for Point to Multipoint TE LSPs*, draft-ietf-mplsrsvp-te-p2mp-07.txt
- Minei et. al. *Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*, draft-ietf-mpls-ldp-p2mp-02.txt
- I. Cowburn. *MAC Hiding in an H-VPLS Environment*, draft-cowburn-l2vpn-vpls-ldp-machiding-01

About Juniper Networks

Juniper Networks, Inc. is the leader in high-performance networking. Juniper offers a high-performance network infrastructure that creates a responsive and trusted environment for accelerating the deployment of services and applications over a single network. This fuels high-performance businesses. Additional information can be found at www.juniper.net.

Corporate and Sales Headquarters

Juniper Networks, Inc.
1194 North Mathilda Avenue
Sunnyvale, CA 94089 USA
Phone: 888.JUNIPER
(888.586.4737)
or 408.745.2000
Fax: 408.745.2100

APAC Headquarters

Juniper Networks (Hong Kong)
26/F, Cityplaza One
1111 King's Road
Taikoo Shing, Hong Kong
Phone: 852.2332.3636
Fax: 852.2574.7803

EMEA Headquarters

Juniper Networks Ireland
Airsides Business Park
Swords, County Dublin,
Ireland
Phone: 35.31.8903.600
Fax: 35.31.8903.601

Copyright 2009 Juniper Networks, Inc. All rights reserved. Juniper Networks, the Juniper Networks logo, JUNOS, NetScreen, and ScreenOS are registered trademarks of Juniper Networks, Inc. in the United States and other countries. JUNOSe is a trademark of Juniper Networks, Inc. All other trademarks, service marks, registered marks, or registered service marks are the property of their respective owners. Juniper Networks assumes no responsibility for any inaccuracies in this document. Juniper Networks reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

To purchase Juniper Networks solutions, please contact your Juniper Networks representative at 1-866-298-6428 or authorized reseller.

