Internet Draft Document   Marc Lasserre
L2VPN Working Group   Xipeng Xiao
                                                        Riverstone Networks

Yetik Serbest   Marc Rapoport
SBC   Completel

Cesar Garrido
Telefonica

draft-ietf-l2vpn-vpls-ldp-applic-00.txt
Expires: October 2004   March 2005

VPLS Applicability

Status of this Memo

Abstract

Virtual Private LAN Service (VPLS) is a layer 2 VPN service that provides multipoint connectivity in the form of an Ethernet emulated

LAN, while usual L2 VPN services are typically point-to-point. Such emulated LANs can span across metropolitan area networks as well as wide area networks.

[VPLS-LDP] defines a method for signaling MPLS connections between member PEs of a VPN and a method for forwarding Ethernet frames over such connections. This document describes the applicability of such procedures to provide VPLS services.

This document also compares the characteristics of this solution against the requirements specified in [L2VPN-REQ]. In summary, there are no architectural limitations to prevent the requirements from being met. But meeting certain requirements (e.g. QoS) is beyond the specification of [VPLS-LDP], and requires careful planning and precise implementation of the Service Provider (SP) networks. This document attempts to capture such issues, presents the potential solutions to these issues, and discusses the pros and cons of each alternative.

This document does not cover the applicability of [VPLS-BGP].

Conventions

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

Placement of this Memo in Sub-IP Area

RELATED DOCUMENTS

www.ietf.org/internet-drafts/draft-ietf-l2vpn-vpls-ldp-04.txt
www.ietf.org/internet-drafts/draft-ietf-l3vpn-applicability-guidelines-00.txt

Table of Contents

VPLS Overview

The primary motivation behind Virtual Private LAN Services (VPLS) is to provide connectivity between geographically dispersed customer sites across MAN/WAN network(s), as if they were connected using a LAN. The intended applications for the end-user can be divided into the following two categories:

- Connectivity between customer routers
- Connectivity between customer Ethernet switches

In addition, VPLS can also be used by the service providers to deliver voice, video , and data services (i.e., triple play, aggregation of IP services over Ethernet access) to connected end-users.

Unlike L3 VPNs such as BGP/MPLS IP VPNs [2547bis] where traffic exchanged between customers and service providers must be IP, VPLS only requires traffic to be Ethernet over which any protocol can be transported, e.g. Netbios or IPX.

The Service Provider Network is a packet switched network (PSN). The PEs are assumed to be fully meshed (note that the mesh can be broken with HVPLS) with transport tunnels over which customer frames that belong to a specific VPLS instance are encapsulated and forwarded. IP-in-IP, L2TPv3, GRE, and MPLS are examples of transport tunnels.

Specific labels used to identify end-to-end paths over such transport tunnels, and these end-to-end paths, which are known as pseudo-wires (PW), are established via targeted LDP [VPLS-LDP].

VPLS defines the bridging rules required for PEs to provide an emulated Ethernet LAN service. In particular, it defines how a loop-free topology must be built, the forwarding rules between PEs, and the signaling method to set up PWs between PEs. The resulting service provides a unique broadcast domain per VPN, with the ability to send unicast, multicast and broadcast traffic (as well as flooding of unknown unicast traffic).

1. Operation of Signaling and Data Planes

1.1. Signaling Plane

As with [PWE3-ETHERNET], [VPLS-LDP] specifies the use of targeted LDP for the signaling of PWs. PWs are established between PEs that are part of the same VPLS instance.

1.2. Data Plane

1.2.1. Ingress Processing

VPLS provides an Ethernet emulated LAN service and hence customer frames are capsulated as Ethernet frames (Ethernet DIX or 802.1). Note that such Ethernet frames can be carried over various access transport technologies (Frame Relay, ATM, etc). Ingress PEs will determine which Forwarding Information Base (FIB) to look up based on the port, VLAN or port/VLAN combination where frames come from. This port to FIB mapping is performed at provisioning time. The destination MAC address is then looked up to determine on which PW this address has been learned from. If the lookup fails, i.e. if this MAC address has not been learned yet, the frame needs to be sent on all the PWs that are part of the corresponding VPLS instance. If the address is known, the frame is sent only over the associated PW. Before actually transmitting the customer frame, it needs to be encapsulated as defined in [PWE3-ETHERNET], and is further encapsulated with the appropriate transport header (e.g. MPLS or GRE).

1.2.2. Egress Processing

Once the tunnel header has been removed, the egress PE determines from the PW label which FIB to look up to determine the egress interface, i.e., VLAN or port/VLAN combination. The original Ethernet frame is then encapsulated with the proper transmission header if necessary (e.g. Frame Relay header) and sent over the corresponding port.

MAC addresses are learned dynamically as traffic is exchanged. New source MAC addresses are learned on a per PW label per VPLS instance basis. An aging timer is used to remove such bindings after a period of time. When user topology changes occur, MAC withdrawal messages in the signaling plane may be used to unlearn MAC addresses to improve convergence time.

Egress PEs might also be configured to perform specific egress encapsulation functions (e.g. VLAN translation).

1.2.3. Intermediate Node Processing

Intermediate nodes (P routers) only act as pure forwarders based on the outer tunnel header. Hence, they do not participate in any VPLS related processing. Only PE routers maintain VPN specific information. This improves the scalability of VPLS service.

2. VPLS vs. Alternative Approaches

2.1. Ethernet Switching

Ethernet can be used to provide multipoint connectivity within small geographical areas such as small metropolitan networks. Pure Ethernet based solutions have scalability issues (e.g. STP limitations, 4095 VLAN limitations). Some enhancements such as QinQ, STP extensions (RSTP, MSTP) provide additional scalability.

VPLS overcomes several limitations of Ethernet based solutions by supporting large numbers of VPNs, better traffic engineering, transport link layer independence and better quality of service.

It is not uncommon for VPLS networks to be complemented with Ethernet switched networks as an aggregation layer.

2.2. BGP/MPLS IP VPN

In metropolitan area networks (MANs), BGP is usually not enabled. MANs provide a transport service to end-users. When multiple sites need to be connected within a metro, VPLS offers the appropriate multipoint transport solution. It is expected that a VPLS instance supports up to O(10^2) site interfaces. When multipoint connectivity is required for a higher number of interfaces sites, with a various range of interface types (e.g. dial-up access, IPSec Tunnels), BGP/MPLS IP VPNs can be more appropriate.

Section 10.1. describes how VPLS and BGP/MPLS IP VPNs can be complementary.

The following sections compare the characteristics of LDP-based VPLS solution against the requirements specified in [L2VPN-REQ]. Key deployment issues that require careful planning and precise implementation of SP networks are highlighted.

3. Provisioning

To provision a VPLS service for a customer, the first step is to create a Virtual Switching Instance (VSI), and assign the customer attachment circuit (AC) (e.g. port, port/VLAN, ATM VC with 1483b encapsulation, etc.) and PWs (including H-VPLS spokes) to it. The PWs interconnect VSIs at different PEs and MTUs together to form an emulated LAN for the customer.

One challenge in doing this is, when a VPLS site needs to be added or removed at a PE, in addition to configuring that particular PE, the network operator needs to find out which other PEs participate in that VPLS instance, and re-configure those PEs. PE auto-discovery can automate this process. The pros and cons of several auto-discovery approaches are discussed in 3.1. .

3.1. PE Auto-Discovery

Currently there are several proposals for PE auto-discovery: the BGP-based approach [VPLS-BGP], the RADIUS-based approach [RADIUS-DIS], and the Provisioning System-based approach.

The BGP and RADIUS-based approaches mandate the use of BGP or RADIUS in every PE, and rely on it to propagate the information of which PEs participate in a VPLS instance (Signaling can automatically happen after the other PEs belonging to the same VPLS instance are discovered). The pros of both approaches are reduced provisioning work and no need for a provisioning system. The con is BGP/RADIUS has to be in every PE, which may not be the case in reality.

With the Provisioning System-based approach, network operators do not configure the PEs. Instead, they specify which PEs participate in which VPLS instances at the Provisioning System. The Provisioning System then translates such service information into PE configuration commands and telnet/ssh to the PEs to execute such commands. Because all information related to every VPLS instance is centralized at the Provisioning System, PE auto-discovery is automatically achieved. To add or remove a PE for a VPLS instance, a network operator simply specifies it at the Provisioning System which will then configure the PEs accordingly.

For VPLS deployments that span across multiple domains, because the

ASBRs (autonomous system border routers) of other domains can be treated as CEs of the current domain, these auto-discovery approaches can all work in the multi-domain case. However, the built-in scalability mechanism in BGP makes the BGP-based auto-discovery more scalable in this scenario [VPLS-BGP].

3.2. Operations and Maintenance

To meet the service level agreement (SLA) with their customers, SPs also need to provision the following:

- Traffic management throughout the network and on customer facing ports in particular
- Traffic Engineering
- Traffic protection (e.g. Fast reroute)
- Service management (e.g. SLA measurement, OAM, accounting, billing, etc)

Manual provisioning for these tasks can be tedious. A provisioning system is highly desirable. If a provisioning system is used, PE auto-discovery may be integrated into it.

4. Migration Impacts

Migration in this document means replacing, or more often, supplementing, an existing metro Ethernet or ATM/Frame Relay network with a VPLS network. There are four likely scenarios:

Interconnecting existing L2 Ethernet islands with a VPLS core
Migrating an existing L2 Ethernet core to a VPLS core;
Interconnecting a new VPLS network with existing ATM/FR networks
Adding VPLS support to an IP routed network

Migration impacts may be mitigated through the use of careful planning when building the network. Also, consideration must be taken when integrating with protocols such as STP/MSTP and how control packets (BPDUs) are handled. In addition, one must also consider ongoing standards efforts within various standards bodies such as the IEEE [802.1ad] and the Metro Ethernet Forum to assess future impact of any changes within the provider network.

4.1. Interconnecting L2 Ethernet Islands with a VPLS Core

Today, many existing metro Ethernet networks are relatively small and cover only specific districts in a metro area. Such networks may simply backhaul traffic to a routing backbone and not interconnected at L2. When metro Ethernet service grows and these networks need to be interconnected at L2, one approach that may be used for a migration strategy is to effectively utilize existing L2 (possibly 802.1Q based or QinQ) networks as "islands" attached to an MPLS based VPLS core network. In this particular case, the L2 network uses predetermined Provider 802.1Q tags (P-tags) to transport a given customers traffic. This P-tag is then utilized as a service delimiter that is then stripped prior to being transported across the MPLS cloud. The service delimiting P-tag is used to identify the VPLS instance to which the traffic should be mapped.

```
                                             ----CE1
                    -------        -------  /      --------
         CE2-      /       \      /       PE1     /        \
             \    /         \    /          \    /          \
              ---|   QinQ    \  /    MPLS/   \  /   QinQ    |
                 |   Domain   PE     VPLS     PE    Domain  |
                  \          /  \   Domain   /  \           /\
                   \        /    \          /    \         /  \
                     -------      ----------      --------     --CE3
```

In this scenario, when different sites of a customer have a mismatch of 802.1Q tags, VLAN translation as defined in [802.1ad] should be applied.

```
                                                   -----
                                                  /  A1 \
     ----                                    ----CE1    |
    /    \          -------        -------  /    |      |
   |  A2 CE2-      /       \      /       PE1     \     /
    \    /   \    /         \    /          \      -----
     ----     ---|   QinQ    \  /   MPLS/   |
                 |   Domain   PE2   VPLS    |
                  \          /   \ Domain   /
           -----   \        /     \        /
           |QinQ|_/  -------        -------
          -|    |
   ----  / ------ ----
  /    \/    \   /    \                 CE = Customer Edge Router
  | A3 CE3    --C4 A4 |                 PE = Provider Edge Router
  \    /         \    /
   ----           ----
```

4.2. Migrating an Existing L2 Ethernet Core to a VPLS Core

```
                                                       CE1
                      -------------------     ------  /
                     /                   \   -|VPLS| /
                    /                     \ / | PE |-
                   /                       \  ------
                  /                         \
                 |        802.1Q/QinQ       |
                  \                         /
           -----   \                       /\  ------
           |VPLS|_/ \                     /  \ |VPLS|
          -| PE |    \                   /    -| PE |-
         / ------     -------------------      ------ \
        /    \                                        \
       CE3    --CE4                                   CE2
```

Providers that have already deployed VLAN based core may choose to build a parallel VPLS core and connect it to the existing 802.1Q/Q-in-Q core. The 802.1Q/Q-in-Q core is effectively treated as a super-island. Then one by one, each individual Ethernet access island is disconnected from the existing core (i.e. super-island) and connected to the VPLS core. The migration issues then become similar to those described in 4.1. and interoperability aspects between .1ad network and MPLS/VPLS network need to be worked out before such migration

A second approach consists in configuring a second VPLS control plane in the existing QinQ PE, hence implementing two virtual networks over a single physical infrastructure. Once the PE/MPLS control plane is running, each customer can be separately migrated through a reconfiguration of its corresponding access ports. For increased stability, the dual control plane approach might require dedicating some links or PEs to the MPLS/VPLS network.

4.3. Interconnecting VPLS networks with ATM/FR networks

If interworking at L2 is needed, the existing ATM/FR networks would need to carry bridge-encapsulated traffic. VPLS can support ATM and Frame Relay (FR) attachment circuits with Ethernet bridge encapsulation. Once the FR/ATM encapsulation has been stripped off, the resulting Ethernet frames can be processed as if they came from an Ethernet link. Therefore, interworking can be naturally achieved.

If the existing ATM/FR networks do not carry bridge-encapsulated traffic, then interworking can only happen at L3. For example, if both VPLS and ATM/FR carry IP traffic, then an IP router can be used to interconnect the two networks.

4.4. Adding VPLS Support to an IP Routed Network

In such a scenario, if existing PEs can support VPLS, then they can continue to serve as PEs. Otherwise, new VPLS PEs need to be added and existing IP routers will serve as Ps or as Layer3-only PEs. Depending on whether the existing IP routers support MPLS or not, MPLS or some other tunneling mechanism such as GRE can be used.

5. Multi-homing

Multi-homing is necessary in order to remove a VPLS PE as a single point of failure for all devices attached to it. There are two instances of multi-homing that apply to VPLS:

- When a CE device is connected to more than one PE,
- In the case of hierarchical VPLS - when an MTU-s device is connected to more than one PE-rs.

In both of these cases, the concern is that a particular MAC address will appear as a source on more than one PE device, causing other PE devices to continuously change their FIBs with regard to the true location of the MAC. This will cause constant table thrashing on the remote PEs, a behavior akin to a Layer 2 switch which participates in a loop.

It is therefore required that any Layer 2 loops, created by multi-homing of a CE or an MTU-s, be resolved within the group of devices participating in that loop. This group includes the multi-homed CE or MTU-s, and all PEs to which it is attached. The PEs involved in such a loop are connected with a full mesh of PWs per VPLS instance.

There are two approaches to resolving the loops created by the multi-homed devices:

- Running an MSTP instance between all devices in the group. In this case, the PEs within the group will need to utilize a P-VLAN for the purposes of running MSTP in the group. This P-VLAN can be re-used on non-overlapping groups of multi-homed CE (or MTU-s) and its PEs. It must be clear that the MSTP process discussed here is a completely different and independent instance of STP than any STP the customer may be running. Such customer STP is always tunneled through the VPLS network, and is never acted upon by the PE or MTU-s devices.

- The MTU-s or the CE can designate its link to one of the PEs it connects to as primary and only send packets for this particular VPLS instance over that link. In this case the MTU-s (CE) is responsible for monitoring the state of that link and for switching to an alternate link if the primary fails. No action is required from the PEs participating in the group, though there should be an indication given from the MTU-s to its connected PEs as to whether the PE is connected to the primary or backup link. This is a very lightweight approach, which is quite useful given the simple and known topology between the CE (MTU-s) and its PEs. With this approach the operator must ensure that PWs in the core remain up, as long as the ingress PE they start from is up. This can typically be ensured with MPLS TE tools, such as fast re-route or back-up LSPs. If there is no available path between the ingress PE and the Egress PEs, a mechanism that monitors the status of the PWs to force the access connection to go down when the PWs are down might be useful. If PWs in the core go down while their ingress PE is up and accepting customer traffic, black-holes can occur.

In each case, the PE nodes are most likely in two different physical locations in the provider network providing network element protection, last mile protection, fiber diversity and provider facility backup. Customer STP traffic is always tunneled through the provider network, and is never acted upon by the PE or MTU-s devices.

Lastly, it should be observed that, since VPLS services provide Ethernet switch-like transport level services, the customer is free to connect any device they desire as a CE. This could be anything from a simple host, hub, L2 switch, or a router. The operator has to be cognizant of the different capabilities of each of those devices to ensure loop-free environment when multi-homed.

6. Loop Prevention

Loops in the core VPLS network are prevented by creating a full mesh of transport circuits between PEs and by applying a split-horizon rule. The split-horizon approach prevents a frame received from the backbone network from being sent out anything other than the customer facing ports belonging to that VPLS instance on the receiving PE. The frame MUST not be forwarded out other PW connecting the receiving PE to other PEs participating in the VPLS instance. This provides the necessary protection, network bandwidth optimization and scalability in the carriers' network as it does not rely on link blocking technologies, like spanning tree type protocols. This forwarding mechanism allows PEs to effectively protect the core network from data loops.

Customer networks need to be able to transparently transport the protocol information that allows their network to properly converge. However, the provider should consider loop protection schemes between the CE and PE that do not affect the customer functions. This would be in addition to spanning tree when the PE connects to a VLAN based L2 metro or when the customer is directly connected to multiple PE nodes.

The provider should look at deploying a loop protection scheme that would intervene automatically when it detects a loop condition on customer access ports. This loop protection scheme serves as an additional line of defense against protocol failures or misconfigurations, which can result in data loops. The concern is that a particular MAC address will appear as a source on more than one PE device, causing other PE devices to continuously update their tables. An external loop protection scheme adds a level of insurance above the customer link protection schemes. Its function is to reduce unnecessary core bandwidth usage when a loop condition occurs in an adjacent network and provide an extra level of protection to multi-homed networks. It is a complement but not a replacement for traditional loop protection mechanisms, like spanning tree. Such a loop protection scheme could be based on the monitoring of the number of Mac addresses moving from one attachment circuit/PW to another circuit/PW.

With directly connected customers, careful consideration needs to be given to backdoor connections. Backdoor connections provide an alternate path around a single provider. If a loop detection scheme is invoked here the customer may be forced to traverse a link that is not desired.

7. Packet Ordering

Normally there is only one transmission path towards a destination with VPLS. So there is no packet re-ordering issue. But if some load balancing mechanism is enabled or if LSPs carrying VPLS traffic are rerouted, packets may be re-ordered inside the PSN. Note that reordering can be avoided when load balancing flows across PWs. Flows can be identified through a number of identifiers in the packet, including MPLS labels, MAC addresses, IP addresses, and UDP/TCP ports.

VPLS data packets use the encapsulation mechanism defined in [PWE3-ETHERNET]. An optional control word which contains a sequence number field can be used to assist in-order delivery. If the user's applications are sensitive to packet re-ordering, this option may be used. However, enabling sequencing usually causes forwarding performance degradation. Another alternative is to avoid load sharing for traffic inside a LSP and pin down LSPs to avoid rerouting.

8. Multi-Domain VPLS Service

As the use of VPLS grows, it is expected that customers will require a single VPLS service delivered by different providers (e.g. either for redundancy or because none of the SPs has the presence to support all the sites of a customer). Different providers would then need to interconnect their VPLS domains for these customers. [VPLS-LDP] has provision for such a requirement, utilizing a full mesh of LSPs among the VPLS gateways of these domains. However, experience of such interconnection is not yet available.

9. Maximum Transmission Unit (MTU) Issues

Because of the encapsulation and transport headers, the MTU for user applications will be smaller than the smallest MTU of all the physical links. In responding to path MTU discovery message, each network device must deduct the total header size from a physical link's MTU. Since path MTU discovery is not always used, SPs must clearly communicate the potential MTU issue to their customers and ask for their cooperation. In reality, most applications will work fine but a small number of them may be affected. This is by no means specific to VPLS. Any networks that put additional header(s) on customer's packets will have the same issue.

10. Interoperability and Interworking

Interoperability should be ensured by proper implementation of the published standards.

10.1. Interworking with BGP/MPLS IP VPN

When interworking VPLS with BGP/MPLS IP VPN, a BGP/MPLS IP VPN (in the backbone) is typically used to interconnect VPLS domains in multiple metros, with such VPLS domains acting as Ethernet aggregation networks for the IP service. In this type of scenario, the BGP/MPLS IP VPN will carry inter-metro traffic whereas VPLS will handle intra-metro traffic.

A useful method for interconnecting a VPLS with a BGP/MPLS IP VPN is to use a "link" to interconnect the VSI and the VRF. Such a "link" can be a physical port, a VLAN spanning across one or multiple physical hops, or 2 LSPs with one in each direction, etc. Analogously, this is like interconnecting a L2 switch with a router, with the VSI as the switch and the VRF as the router.

Access/transport networks such as VPLS can also be interconnected with BGP/MPLS IP VPNs using various mechanisms such as Carrier's Carrier as defined in [RFC-2547].

10.2. Interworking With Frame Relay/ATM Attachment Circuits

Frame Relay (FR) and ATM attachment circuits with Ethernet bridged encapsulation can be terminated within VPLS PEs. The resulting Ethernet frames (i.e. once the FR/ATM encapsulation has been stripped off) are processed as standard Ethernet frames.

In order to support a complete interworking model between FR and Ethernet or between ATM and Ethernet, mapping service profiles and OAM traffic from one to the other are necessary. Additionally, circuit management (e.g. LMI to PW state mapping) between the various technologies are required. Such standards are being defined by other standard organizations such as the MPLS-FR-ATM Alliance.

11. Quality of Service

The provision of appropriate QoS capabilities may require any combination of the following:

- QoS in the access network.
- Admission control by the PE router on the ingress access links.
- Classification by the PE, for traffic arriving from the CE.
  Once the PE classifies a user packet, this classification needs to be preserved in the encapsulation (MPLS EXP or IP DSCP) used to send the packet across the backbone.

- Traffic conditioning (policing or shaping) by the PE router on the ingress access links.
- DSCP/EXP-based queuing and WRED in the VPLS network
- Traffic engineering in the VPLS network.
- Fast reroute in the VPLS network

None of these features are VPLS specific. The ability to support them depends on whether the features are available on the edge and core devices. It is up to the SPs to decide how to use such mechanisms to provide QoS. Such mechanisms can be used to support either the "hose model" or the "pipe model", although the hose model is a more natural fit and is usually the support model by default.

12. Security

12.1. Customer Access Control and Authentication

Control of the customer access can be achieved by controlling physical access to the CEs, the PEs and the links between them. If multiple customers use 802.1Q service delimiting tags in the same trunk link to access VPLS service, and the tags are put on by the customers themselves, ACLs should be used to ensure that each customer only puts on the tag that it is supposed to put on. Packets with other tag(s) must be dropped.

802.1x may be used for CE device authentication.

12.2. Traffic Separation between VPLS Instances

VPLS instances maintain separation of broadcast domains between themselves. Traffic entering a given VPLS instance at a given PE device does not, under any circumstances, cross the boundaries of the VPLS into another instance. VPLS devices (PEs and MTU-s) ensure that by maintaining a FIB table and a full mesh of PWs on a per-VPLS instance basis.

The above statement is correct regardless of the learning mode employed by a particular VPLS instance (qualified or unqualified), or whether or not VLANs are treated as broadcast domain identifiers, or simply as circuit IDs which have no significance in determining the broadcast domain. In either of these cases, the VPLS instance is the outer-most "envelope" which ensures that traffic within it does not "leak" into another VPLS instance.

12.3. Protection of SP Networks

Two types of DoS attacks are of concern with VPLS:

- Attacks against VPLS devices

- Attacks against other devices, for which the VPLS network is a transport.

Attacks of the first type are naturally of greater concern for a VPLS operator, because they can destabilize the VPLS network as a whole, and affect multiple customers. The tunneling nature of VPLS by itself limits the possibilities for attacks via the data plane, simply because such attacks will be tunneled through the VPLS network, and will create the same load on the VPLS equipment as legitimate traffic will.

Operators must watch for exception packet handling in VPLS equipment. In many cases, exception packets are sent to the control plane for handling. If that is the case, the operator must ensure that such exception packets can be rate-limited in a fashion that guarantees that the control plane will not be significantly burdened by them. A SP should limit the amount of traffic that a customer can flood.

The second type of DoS attacks, which use the VPLS network as a transport, are not really a threat to the VPLS devices themselves but are to devices behind them. VPLS PEs may be configured with rate-limiting and rate-shaping capabilities which permit them to limit the amount of traffic allowed into a particular VPLS instance. This prevents a VPLS customer from consuming excessive amount of network resources and from starving other customers. For example, it might be useful to limit the multicast/broadcast/unknown traffic of the customer, considering that the replication of this traffic will create a load in the core proportional to the number of PEs participating to the VPLS instance. Optionally, they can also be tasked with advanced processing of the traffic they tunnel. For example, they may impose access lists which deny traffic from particular sources or protocols.

Such approaches however are highly vendor-specific and outside the scope of [VPLS-LDP]. In addition, they may have significant design and operational repercussions. Alternative approaches which hand-off DoS protection activities to non-VPLS devices (such as customer equipment) are a possibility.

12.4. Protection of User Data

VPLS does not have special provisioning for ensuring user data security. If a customer's traffic is IP traffic, that customer may provide its own user data security by using IPsec. In fact, VPLS is compatible with any use of security by the customer, as long as a clear text Ethernet header is passed from CE to PE.

13. Scalability

As per [L2VPN-REQ], a large SP may eventually require support of up to O(10^4) VPLS instances. In addition, some of these VPLS instances may need to support O(10^2) sites and O(10^3) users/MACs. This section describes the key scalability challenges and how VPLS-LDP addresses them.

13.1. Mesh topology

A full mesh of tunnel LSPs, over which a full mesh of PWs is established, is created between participating PEs. When using hierarchical VPLS constructs, the size of this full mesh can be reduced to hub PEs aggregating point-to-point spokes as described in section 10 of [VPLS-LDP].

This reduces the number of tunnels and PWs from O(N*N) to O(N).

13.2. Signaling

Using HVPLS constructs also allows the total number of targeted LDP sessions to be reduced from O(N*N) to O(N).

13.3. MAC addresses and MAC learning

Depending on the type of CE devices used, i.e. switches or routers, the total number of MAC addresses to be learned by VPLS PEs can vary from one address per site to a large number of MAC addresses.

When Ethernet networks exceed a large number of MAC addresses (e.g. hundreds), routers are introduced to limit the size of such broadcast domains. This reduces the total number of MAC addresses to learn to such routers only.

In the case of large flat Ethernet networks, ingress PEs must be able to limit the number of MAC addresses that can be learned on a per VPLS basis.

13.4. Packet replication

With VPLS, broadcast, multicast and unknown destination frames get replicated by the ingress PEs, i.e. close to the source of the frame. Ideally such frames should be replicated as close to the destination as possible to minimize bandwidth consumption. With hierarchical VPLS, the replication process is distributed between several ingress and egress MTUs and PEs. This helps not only minimizing bandwidth resources but also improving multicast performance and reducing latency.

13.5. Broadcast limiting

Ingress MTUs or PEs may be able to rate limit the amount of broadcast/multicast/unknown traffic generated by end users in order to protect core resources and to prevent a few users from using all the bandwidth available.

13.6. Multicast

In order to optimize the replication of multicast traffic, it is highly desirable for PEs to support multicast snooping techniques in order to only forward traffic where needed. In the case where the CE device is an L2 switch, IGMP snooping would be required, however, if the CE device is a router PIM snooping would be more applicable.

14. Management

The following five major areas in management are discussed bellow:

- Fault
- Configuration
- Accounting
- Provisioning
- Security

VPLS introduces new configurations related to creation and removal of VSIs, etc. VPLS also introduces new provisioning challenges because the service needs to be delivered end-to-end and therefore many things such as access control, QoS, etc need to be provisioned accordingly. Achieving these via manual CLI configuration can be error prone. Therefore, it is advisable to use a provisioning system for configuration and provisioning.

Although VPLS-specific MIBs are still under development, accounting information can usually be achieved via [IF-MIB] and [LSR-MIB]. The important point is that accounting information should be available per service basis. Such information can then be processed by an accounting application to produce the accounting records. Security can be achieved by the measures described in Section 0.

Managing fault with VPLS involves multi-point connectivity verification and locating the fault if there is one. Such mechanism is sometimes referred to as "VPLS OAM" and is discussed below.

Although VPLS OAM is still being defined, one of the approaches has gained momentum. This approach proposes applying Ethernet OAM mechanism that is being standardized by ITU, IEEE and the Metro Ethernet Forum (MEF) to an VPLS environment for L2 connectivity verification and fault locating, and applying MPLS OAM mechanism such as [LSP-PING] or [BFD] or [VCCV] to MPLS connectivity verification and fault locating. Of course, if IP tunnels (e.g. GRE) are used, IP ping and traceroute can be used in the place of MPLS OAM. VPLS OAM is therefore achieved by integrating OAM mechanisms at different layers together.

In summary of this section: management of VPLS services involves many things and can be tedious. A complete suite of management software including EMS, NMS and a provisioning system can therefore be highly desirable.

Acknowledgments

The authors wish to thank the following people for their constructive contributions to the text in this document:

Javier Antich
Ian Cowburn
Richard Foote
Rob Nath
Ali Sajassi
Nick Slabakov

Some text was adapted from the Applicability Statement for BGP/MPLS IP VPNs [AS2547] document.

Copyright Notice

Disclaimer

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

IPR Disclosure Acknowledgement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this

Normative References

[VPLS-LDP] "Virtual Private LAN Services over MPLS", Marc Lasserre, Vach Kompella, et al., draft-ietf-l2vpn-vpls-ldp-05.txt, Work in progress, September 2004

[PWE3-ETHERNET] "Encapsulation Methods for Transport of Ethernet Frames Over IP/MPLS Networks", draft-ietf-pwe3-ethernet-encap-02.txt, Work in progress, February 2003.

[PWE3-CTRL] "Transport of Layer 2 Frames Over MPLS", draft-ietf-pwe3-control-protocol-02.txt, Work in progress, February 2003.
[RFC3036] "LDP Specification", L. Andersson, et al. RFC 3036. January 2001.

[RFC3036] "LDP Specification", L. Andersson, et al. RFC 3036. January 2001.

[802.1D-ORIG] Original 802.1D - ISO/IEC 10038, ANSI/IEEE Std 802.1D-**1993 "MAC Bridges".**

[802.1D-REV] 802.1D - "Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Common specifications - Part 3: Media Access Control (MAC) Bridges: Revision. This is a revision of ISO/IEC 10038: 1993, 802.1j-1992 and 802.6k-1992. It incorporates P802.11c, P802.1p and P802.12e." ISO/IEC 15802-3: 1998.

[802.1Q] 802.1Q - ANSI/IEEE Draft Standard P802.1Q/D11, "IEEE Standards for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks", July 1998.

Informative References

[BGP-VPN] Rosen and Rekhter, "BGP/MPLS VPNs". draft-ietf-ppvpnrfc2547bis-04.txt, Work in Progress, May 2003.

[RADIUS-DISC] " Using Radius for PE-Based VPN Discovery", Juha Heinanen, draft-heinanen-radius-pe-discovery-04.txt, Work in Progress, June 2003.

[BGP-DISC] "Using BGP as an Auto-Discovery Mechanism for Network-based VPNs", Ould-Brahim, et. al., draft-ietf-ppvpn-bgpvpn-auto-05.txt, Work in Progress, May 2003.

[VPLS-BRIDGING] "Bridging and VPLS", draft-finn-ppvpn-bridging-vpls-00.txt, Work in Progress, June 2002.

[L2VPN-SIG] "LDP-based Signaling for L2VPNs", draft-rosen-ppvpn-l2-signaling-03.txt, Work in Progress, May 2003.

[L2FRAME] "L2VPN Framework", draft-ietf-ppvpn-l2-framework-03, Work in Progress, February 2003.

[L2VPN-REQ] "Service Requirements for Layer 2 Provider Provisioned Virtual Private Networks", draft-ietf-ppvpn-l2vpn-requirements-00.txt, Work in Progress, May 2003.

[802.1ad] "IEEE standard for Provider Bridges", Work in Progress, December 2002.

Authors' Addresses

Marc Lasserre
Riverstone Networks
Email: marc@riverstonenet.com

Xipeng Xiao
Riverstone Networks
Email: xxiao@riverstonenet.com

Yetik Serbest
SBC Communications
Yetik_serbest@labs.sbc.com

Cesar Garrido,
Telefonica
cesar.garridosanahuja@telefonica.es

Marc Rapoport
Completel
m.rapoport@completel.fr