Network Working Group
Request for Comments: 5501
Category: Informational

Y. Kamite, Ed.
NTT Communications
Y. Wada
NTT
Y. Serbest
AT&T
T. Morin
France Telecom
L. Fang
Cisco Systems, Inc.

March 2009

Requirements for Multicast Support in Virtual Private LAN Services

Status of This Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This document provides functional requirements for network solutions that support multicast over Virtual Private LAN Service (VPLS). It specifies requirements both from the end user and service provider standpoints. It is intended that potential solutions will use these requirements as guidelines.

Table of Contents

1. Introduction

1.1. Background

VPLS (Virtual Private LAN Service) is a provider service that emulates the full functionality of a traditional Local Area Network (LAN). VPLS interconnects several customer LAN segments over a packet switched network (PSN) backbone, creating a multipoint-to-multipoint Ethernet VPN. For customers, their remote LAN segments behave as one single LAN.

In a VPLS, the provider network emulates a learning bridge, and forwarding takes place based on Ethernet MAC (media access control) learning. Hence, a VPLS requires MAC address learning/aging on a per-PW (pseudowire) basis, where forwarding decisions treat the PW as a "bridge port".

VPLS is a Layer-2 (L2) service. However, it provides two applications from the customer’s point of view:

- LAN Routing application: providing connectivity between customer routers

- LAN Switching application: providing connectivity between customer Ethernet switches

Thus, in some cases, customers across MAN/WAN have transparent Layer-2 connectivity while their main goal is to run Layer-3 applications within their routing domain. As a result, different requirements arise from their variety of applications.

Originally, PEs (Provider Edges) in VPLS transport broadcast/ multicast Ethernet frames by replicating all multicast/broadcast frames received from an Attachment Circuit (AC) to all PW's corresponding to a particular Virtual Switching Instance (VSI). Such a technique has the advantage of keeping the P (Provider Router) and PE devices completely unaware of IP multicast-specific issues. Obviously, however, it has quite a few scalability drawbacks in terms of bandwidth consumption, which will lead to increased cost in large-scale deployment.

Meanwhile, there is a growing need for support of multicast-based services such as IP TV. This commercial trend makes it necessary for most VPLS deployments to support multicast more efficiently than before. It is also necessary as customer routers are now likely to be running IP multicast protocols, and those routers are connected to switches that will be handling large amounts of multicast traffic.

Therefore, it is desirable to have more efficient techniques to support IP multicast over VPLS.

1.2. Scope of This Document

This document provides functional requirements for network solutions that support IP multicast in VPLS [RFC4761] [RFC4762]. It identifies requirements that MAY apply to the existing base VPLS architecture in order to optimize IP multicast. It also complements the generic L2VPN requirements document [RFC4665], by specifying additional requirements specific to the deployment of IP multicast in VPLS.

The technical specifications are outside the scope of this document. In this document, there is no intent to specify either solution-specific details or application-specific requirements. Also, this document does NOT aim to express multicast-inferred requirements that are not specific to VPLS. It does NOT aim to express any requirements for native Ethernet specifications, either.

This document is proposed as a solution guideline and a checklist of requirements for solutions, by which we will evaluate how each solution satisfies the requirements.

This document clarifies the needs from both VPLS customer as well as provider standpoints and formulates the problems that should be addressed by technical solutions while staying solution agnostic.

A technical solution and corresponding service that supports this document's requirements are hereinafter called a "multicast VPLS".

2. Conventions Used in This Document

2.1. Terminology

The reader is assumed to be familiar with the terminology, reference models, and taxonomy defined in [RFC4664] and [RFC4665]. For readability purposes, we repeat some of the terms here.

Moreover, we also propose some other terms needed when IP multicast support in VPLS is discussed.

- ASM: Any Source Multicast. One of the two multicast service models where each corresponding service can have an arbitrary number of senders.

- G: denotes a multicast group.

- MDTunnel: Multicast Distribution Tunnel, the means by which the customer's multicast traffic will be conveyed across the Service Provider (SP) network. This is meant in a generic way: such tunnels can be point-to-point, point-to-multipoint, or multipoint-to-multipoint. Although this definition may seem to assume that distribution tunnels are unidirectional, the wording encompasses bidirectional tunnels as well.

- Multicast Channel: In the multicast SSM (Source Specific Multicast) model [RFC4607], a "multicast channel" designates traffic from a specific source S to a multicast group G. Also denominated as "(S,G)".

- Multicast domain: An area in which multicast data is transmitted. In this document, this term has a generic meaning that can refer to Layer-2 and Layer-3. Generally, the Layer-3 multicast domain is determined by the Layer-3 multicast protocol used to establish reachability between all potential receivers in the corresponding domain. The Layer-2 multicast domain can be the same as the Layer-2 broadcast domain (i.e., VLAN), but it may be restricted to being smaller than the Layer-2 broadcast domain if an additional control protocol is used.

- CE: Customer Edge Device.

- PE: Provider Edge.

- P: Provider Router.

- S: denotes a multicast source.

- SP: Service Provider.

- SSM: Source Specific Multicast. One of the two multicast service models where each corresponding service relies upon the use of a single source.

- U-PE/N-PE: The device closest to the customer/user is called the User-facing PE (U-PE) and the device closest to the core network is called the Network-facing PE (N-PE).

- VPLS instance: A service entity manageable in VPLS architecture. All CE devices participating in a single VPLS instance appear to be on the same LAN, composing a VPN across the SP's network. A VPLS instance corresponds to a group of VSIs that are interconnected using PWs (pseudowires).

- VSI: Virtual Switching Instance. A VSI is a logical entity in a PE that maps multiple ACs (Attachment Circuits) to multiple PWs. The VSI is populated in much the same way as a standard bridge populates its forwarding table. Each PE device may have multiple VSIs, where each VSI belongs to a different VPLS instance.

2.2. 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 [RFC2119] .

3. Problem Statements

3.1. Motivation

Today, many kinds of IP multicast services are becoming available. Over their Layer-2 VPN service, particularly over VPLS, customers would often like to operate their multicast applications to remote sites. Also, VPN service providers using an IP-based network expect that such Layer-2 network infrastructure will efficiently support multicast data traffic.

However, VPLS has a shortcoming as it relates to multicast scalability as mentioned below because of the replication mechanisms intrinsic to the original architecture. Accordingly, the primary goal for technical solutions is to solve this issue partially or completely, and provide efficient ways to support IP multicast services over VPLS.

3.2. Multicast Scalability

In VPLS, replication occurs at an ingress PE (in the hierarchical VPLS (H-VPLS) case, at N-PE) when a CE sends (1) Broadcast, (2) Multicast, or (3) Unknown destination unicast. There are two well-known issues with this approach:

Issue A: Replication to non-member site:

In cases (1) and (3), the upstream PE has to transmit packets to all of the downstream PEs that belong to the common VPLS instance. You cannot decrease the number of members, so this is basically an inevitable situation for most VPLS deployments.

In case (2), however, there is an issue that multicast traffic is sent to sites with no members. Usually, this is caused when the upstream PE does not maintain downstream membership information. The upstream PE simply floods frames to all downstream PEs, and the downstream PEs forward them to directly connected CEs; however, those CEs might not be the members of any multicast group. From the perspective of customers, they might suffer from pressure on their own resources due to unnecessary traffic. From the perspective of SPs, they would not like wasteful over-provisioning to cover such traffic.

Issue B: Replication of PWs on shared physical path:

In VPLS, a VSI associated with each VPLS instance behaves as a logical emulated bridge that can transport Ethernet across the PSN backbone using PWs. In principle, PWs are designed for unicast traffic.

In all cases, (1), (2), and (3), Ethernet frames are replicated on one or more PWs that belong to that VSI. This replication is often inefficient in terms of bandwidth usage if those PWs are traversing shared physical links in the backbone.

For instance, suppose there are 20 remote PEs belonging to a particular VPLS instance, and all PWs happen to be traversing over the same link from one local PE to its next-hop P. In this case, even if a CE sends 50 Mbps to the local PE, the total bandwidth of that link will be to 1000 Mbps.

Note that while traditional 802.1D Ethernet switches replicate broadcast/multicast flows once at most per output interface, VPLS often needs to transmit one or more flows duplicated over the same output interface.

From the perspective of customers, there is no serious issue because they do not know what happens in the core. However, from the perspective of SPs, unnecessary replication brings the risk of resource exhaustion when the number of PWs increases.

In both Issues A and B, these undesirable situations will become obvious with the wide-spread use of IP multicast applications by customers. Naturally, the problem will become more serious as the number of sites grows. In other words, there are concerns over the scalability of multicast in VPLS today.

3.3. Application Considerations

3.3.1. Two Perspectives of the Service

When it comes to IP multicast over VPLS, there are two different aspects in terms of service provisioning. They are closely related to the functional requirements from two technical standpoints:

Layer-2 and Layer-3.

- Native Ethernet service aspect

  This aspect mainly affects Ethernet network service operators. Their main interest is to solve the issue that existing VPLS deployments cannot always handle multicast/broadcast frames efficiently.

  Today, wide-area Ethernet services are becoming popular, and VPLS can be utilized to provide wide-area LAN services. As customers come to use various kinds of content distribution applications that use IP multicast (or other protocols that lead to multicast/broadcast in the Ethernet layer), the total amount of traffic will also grow. In addition, considerations of Operations, Administration, and Management (OAM), security and other related points in multicast in view of Layer-2 are important.

  In such circumstances, the native VPLS specification would not always be satisfactory if multicast traffic is more dominant in total resource utilization than before. The scalability issues mentioned in the previous section are expected to be solved.

- IP multicast service aspect

  This aspect mainly affects both IP service providers and end users. Their main interest is to provide IP multicast services transparently but effectively by means of VPLS as a network infrastructure.

SPs might expect VPLS as an access/metro network to deliver multicast traffic (such as Triple-play (Video, Voice, Data) and Multicast IP VPNs) in an efficient way.

4. General Requirements

We assume the basic requirements for VPLS written in [RFC4665] are fulfilled unless otherwise specified in this document.

4.1. Scope of Transport

4.1.1. Traffic Types

4.1.1.1. Multicast and Broadcast

As described before, any solution is expected to have mechanisms for efficient transport of IP multicast. Multicast is related to both Issues A and B (see Section 3.2); however, broadcast is related to Issue B only because it does not need membership control.

- A multicast VPLS solution SHOULD attempt to solve both Issues A and B, if possible. However, since some applications prioritize solving one issue over the other, the solution MUST identify which Issue (A or B) it is attempting to solve. The solution SHOULD provide a basis for evaluating how well it solves the issue(s) it is targeting, if it is providing an approximate solution.

4.1.1.2. Unknown Destination Unicast

Unknown destination MAC unicast requires flooding, but its characteristics are quite different from multicast/broadcast. When the unicast MAC address is learned, the PE changes its forwarding behavior from flooding over all PWs into sending over one PW. Thereby, it will require different technical studies from multicast/broadcast, which is out of scope of this document.

4.1.2. Multicast Packet Types

Ethernet multicast is used for conveying Layer-3 multicast data. When IP multicast is encapsulated by an Ethernet frame, the IP multicast group address is mapped to the Ethernet destination MAC address. In IPv4, the mapping uses the lower 23 bits of the (32-bit) IPv4 multicast address and places them as the lower 23 bits of a destination MAC address with the fixed header of 01-00-5E in hex. Since this mapping is ambiguous (i.e., there is a multiplicity of 1 Ethernet address to 32 IPv4 addresses), MAC-based forwarding is not ideal for IP multicast because some hosts might possibly receive packets they are not interested in, which is inefficient in traffic

delivery and has an impact on security. On the other hand, if the solution tracks IP addresses rather than MAC addresses, this concern can be prevented. The drawback of this approach is, however, that the network administration becomes slightly more complicated.

Ethernet multicast is also used for Layer-2 control frames. For example, BPDU (Bridge Protocol Data Unit) for IEEE 802.1D Spanning Trees uses a multicast destination MAC address (01-80-C2-00-00-00). Also, some of IEEE 802.1ag [802.1ag] Connectivity Fault Management (CFM) messages use a multicast destination MAC address dependent on their message type and application. From the perspective of IP multicast, however, it is necessary in VPLS to flood such control frames to all participating CEs, without requiring any membership controls.

As for a multicast VPLS solution, it can only use Ethernet-related information, if you stand by the strict application of the basic requirement: "a L2VPN service SHOULD be agnostic to customer's Layer 3 traffic" [RFC4665]. This means no Layer-3 information should be checked for transport. However, it is obvious this is an impediment to solve Issue A.

Consequently, a multicast VPLS can be allowed to make use of some Layer-3-related supplementary information in order to improve transport efficiency. In fact, today's LAN-switch implementations often support such approaches and snoop upper-layer protocols and examine IP multicast memberships (e.g., Protocol Independent Multicast (PIM) snooping and IGMP/MLD (Multicast Listener Discovery) snooping [RFC4541]). This will implicitly suggest that VPLS may adopt similar techniques although this document does NOT state Layer-3 snooping is mandatory. If such an approach is taken, careful consideration of Layer-3 state maintenance is necessary. In addition, note that snooping approaches sometimes have disadvantages in the system's transparency; that is, one particular protocol's snooping solution might hinder other (especially future) protocol's working (e.g., an IGMPv2-snooping switch vs. a new IGMPv3-snooping one). Also, note that there are potential alternatives to snooping:

- Static configuration of multicast Ethernet addresses and ports/ interfaces.

- Multicast control protocol based on Layer-2 technology that signals mappings of multicast addresses to ports/interfaces, such as Generic Attribute Registration Protocol / GARP Multicast Registration Protocol (GARP/GMRP) [802.1D], Cisco Group Management Protocol [CGMP], and Router-port Group Management Protocol (RGMP) [RFC3488].

On the basis described above, general requirements about packet types are given as follows:

- A solution SHOULD support a way to facilitate IP multicast forwarding of the customers. It MAY observe Layer-3 information (i.e., multicast routing protocols and state) to the degree necessary, but any information irrelevant to multicast transport SHOULD NOT be consulted.

- In a solution, Layer-2 control frames (e.g., BPDU, 802.1ag CFM) SHOULD be flooded to all PE/CEs in a common VPLS instance. A solution SHOULD NOT change or limit the flooding scope to remote PE/CEs in terms of end-point reachability.

- In a solution, Layer-2 frames that encapsulate Layer-3 multicast control packets (e.g., PIM, IGMP (for IPv4), and MLD (for IPv6)) MAY be flooded only to relevant members, with the goal of limiting flooding scope. However, Layer-2 frames that encapsulate other Layer-3 control packets (e.g., OSPF, IS-IS) SHOULD be flooded to all PE/CEs in a VPLS instance.

4.1.3. MAC Learning Consideration

In a common VPLS architecture, MAC learning is carried out by PEs based on the incoming frame’s source MAC address, independently of the destination MAC address (i.e., regardless of whether it is unicast, multicast, or broadcast). This is the case with the multicast VPLS solution’s environment too. In this document, the improvement of MAC learning scalability is beyond the scope. It will be covered in future work.

4.2. Static Solutions

A solution SHOULD allow static configuration to account for various operator policies, where the logical multicast topology does not change dynamically in conjunction with a customer’s multicast routing.

4.3. Backward Compatibility

A solution SHOULD be backward compatible with the existing VPLS solution. It SHOULD allow a case where a common VPLS instance is composed of both PEs supporting the solution and PEs not supporting it, and the multicast optimization (both forwarding and receiving) is achieved between the compliant PEs.

Note again that the existing VPLS solutions already have a simple flooding capability. Thus, this backward compatibility will give customers and SPs the improved efficiency of multicast forwarding incrementally as the solution is deployed.

5. Customer Requirements

5.1. CE-PE Protocol

5.1.1. Layer-2 Aspect

A solution SHOULD allow transparent operation of Ethernet control protocols employed by customers (e.g., Spanning Tree Protocol [802.1D]) and their seamless operation with multicast data transport.

Solutions MAY examine Ethernet multicast control frames for the purpose of efficient dynamic transport (e.g., GARP/GMRP [802.1D]). However, solutions MUST NOT assume all CEs are always running such protocols (typically in the case where a CE is a router and is not aware of Layer-2 details).

A whole Layer-2 multicast frame (whether for data or control) SHOULD NOT be altered from a CE to CE(s) EXCEPT for the VLAN ID field, ensuring that it is transparently transported. If VLAN IDs are assigned by the SP, they can be altered. Note, however, when VLAN IDs are changed, Layer-2 protocols may be broken in some cases, such as Multiple Spanning Trees [802.1s]. Also, if the Layer-2 frame is encapsulating a Layer-3 multicast control packet (e.g., PIM/IGMP) and customers allow it to be regenerated at the PE (aka proxy: see Section 5.1.2), then the MAC address for that frame MAY be altered to the minimum necessary (e.g., use PE’s own MAC address as a source).

5.1.2. Layer-3 Aspect

Again, a solution MAY examine the customer’s Layer-3 multicast protocol packets for the purpose of efficient and dynamic transport. If it does, supported protocols SHOULD include:

- PIM-SM (Sparse Mode) [RFC4601], PIM-SSM (Source-Specific Multicast) [RFC4607], bidirectional PIM [RFC5015], and PIM-DM (Dense Mode) [RFC3973].

- IGMP (v1 [RFC1112], v2 [RFC2236], and v3 [RFC3376]) (for IPv4 solutions).

- Multicast Listener Discovery Protocol (MLD) (v1 [RFC2710] and v2 [RFC3810]) (for IPv6 solutions).

A solution MUST NOT require any special Layer-3 multicast protocol packet processing by the end users. However, it MAY require some configuration changes (e.g., turning explicit tracking on/off in the PIM).

A whole Layer-3 multicast packet (whether for data or control), which is encapsulated inside a Layer-2 frame, SHOULD NOT be altered from a CE to CE(s), ensuring that it is transparently transported. However, as for Layer-3 multicast control (like PIM Join/Prune/Hello and IGMP Query/Report packet), it MAY be altered to the minimum necessary if such partial non-transparency is acceptable from point of view of the multicast service. Similarly, a PE MAY consume such Layer-3 multicast control packets and regenerate an entirely new packet if partial non-transparency is acceptable with legitimate reason for customers (aka proxy).

5.2. Multicast Domain

As noted in Section 2.1, the term "multicast domain" is used in a generic context for Layer-2 and Layer-3.

A solution SHOULD NOT alter the boundaries of customer multicast domains. It MUST ensure that the provided Ethernet multicast domain always encompasses the corresponding customer Layer-3 multicast domain.

A solution SHOULD optimize those domains' coverage sizes, i.e., a solution SHOULD ensure that unnecessary traffic is not sent to CEs with no members. Ideally, the provided domain size will be close to that of the customer's Layer-3 multicast membership distribution; however, it is OPTIONAL to achieve such absolute optimality from the perspective of Layer-3.

If a customer uses VLANs and a VLAN ID as a service delimiter (i.e., each VPLS instance is represented by a unique customer VLAN tag carried by a frame through the User Network Interface (UNI) port), a solution MUST provide a separate multicast domain for each VLAN ID. Note that if VLAN ID translation is provided (i.e., if a customer VLAN at one site is mapped into a different customer VLAN at a different site), multicast domains will be created per set of VLAN IDs that are associated with translation.

If a customer uses VLANs but a VLAN ID is not a service delimiter (i.e., the VPN disregards customer VLAN IDs), a solution MAY provide a separate multicast domain for each VLAN ID. An SP is not mandatorily required to provide a separate multicast domain for each VLAN ID, but it may be considered beneficial to do so.

A solution MAY build multicast domains based on Ethernet MAC addresses. It MAY also build multicast domains based on the IP addresses inside Ethernet frames. That is, PEs in each VPLS instance might control forwarding behavior and provide different multicast frame reachability depending on each MAC/IP destination address separately. If IP multicast channels are fully considered in a solution, the provided domain size will be closer to actual channel reachability.

5.3. Quality of Service (QoS)

Customers require that multicast quality of service MUST be at least on par with what exists for unicast traffic. Moreover, as multicast is often used to deliver high-quality services such as TV broadcast, delay-, jitter-, and loss-sensitive traffic MUST be supported over multicast VPLS.

To accomplish this, the solution MAY have additional features to support high QoS such as bandwidth reservation and flow admission control. Also, multicast VPLS deployment SHALL benefit from IEEE 802.1p Class-of-Service (CoS) techniques [802.1D] and Diffserv [RFC2475] mechanisms.

Moreover, multicast traffic SHOULD NOT affect the QoS that unicast traffic receives and vice versa. That is, separation of multicast and unicast traffic in terms of QoS is necessary.

5.4. SLA Parameters Measurement

Since SLA parameters are part of the service sold to customers, they simply want to verify their application performance by measuring the parameters SP(s) provide.

Multicast specific characteristics that may be monitored are, for instance, multicast statistics per stream (e.g., total/incoming/outgoing/dropped traffic by period of time), one-way delay, jitter and group join/leave delay (time to start receiving traffic from a multicast group across the VPN since the join/leave was issued). An operator may also wish to compare the difference in one-way delay for a solitary multicast group/stream from a single, source PE to multiple receiver PEs.

A solution SHOULD provide these parameters with Ethernet multicast group level granularity. (For example, a multicast MAC address will be one of those entries for classifying flows with statistics, delay, and so on.) However, if a solution is aimed at IP multicast transport efficiency, it MAY support IP multicast-level granularity.

(For example, multicast IP address/channel will be entries for latency time.)

In order to monitor them, standard interfaces for statistics gathering SHOULD also be provided (e.g., standard Simple Network Management Protocol (SNMP) MIB Modules).

5.5. Security

A solution MUST provide customers with architectures that give the same level of security both for unicast and multicast.

5.5.1. Isolation from Unicast

Solutions SHOULD NOT affect any forwarding information base, throughput, or resiliency, etc., of unicast frames; that is, they SHOULD provide isolation from unicast.

5.5.2. Access Control

A solution MAY filter multicast traffic inside a VPLS, upon the request of an individual customer, (for example, MAC/VLAN filtering, IP multicast channel filtering, etc.).

5.5.3. Policing and Shaping on Multicast

A solution SHOULD support policing and shaping multicast traffic on a per-customer basis and on a per-AC (Attachment Circuit) basis. This is intended to prevent multicast traffic from exhausting resources for unicast inside a common customer's VPN. This might also be beneficial for QoS separation (see Section 5.3).

5.6. Access Connectivity

First and foremost, various physical connectivity types described in [RFC4665] MUST be supported.

5.7. Multi-Homing

A multicast VPLS MUST allow a situation in which a CE is dual-homed to two different SPs via diverse access networks -- one is supporting multicast VPLS but the other is not supporting it, (because it is an existing VPLS or 802.1Q/QinQ network).

5.8. Protection and Restoration

A multicast VPLS infrastructure SHOULD allow redundant paths to assure high availability.

Multicast forwarding restoration time MUST NOT be greater than the time it takes a customer's Layer-3 multicast protocols to detect a failure in the VPLS infrastructure. For example, if a customer uses PIM with default configuration, the hello hold timer is 105 seconds, and solutions are required to restore a failure no later than this period. To achieve this, a solution might need to support providing alternative multicast paths.

Moreover, if multicast forwarding was not successfully restored (e.g., in case of no redundant paths), a solution MAY raise alarms to provide outage notification to customers before such a hold timer expires.

5.9. Minimum MTU

Multicast applications are often sensitive to packet fragmentation and reassembly, so the requirement to avoid fragmentation might be stronger than the existing VPLS solution.

A solution SHOULD provide customers with enough committed minimum MTU (i.e., service MTU) for multicast Ethernet frames to ensure that IP fragmentation between customer sites never occurs. It MAY give different MTU sizes to multicast and unicast.

5.10. Frame Reordering Prevention

A solution SHOULD attempt to prevent frame reordering when delivering customer multicast traffic. Likewise, for unicast and unknown unicast traffic, it SHOULD attempt not to increase the likelihood of reordering compared with existing VPLS solutions.

It is to be noted that delivery of out-of-order frames is not avoidable in certain cases. Specifically, if a solution adopts some MDTunnels (see Section 6.2) and dynamically selects them for optimized delivery (e.g., switching from one aggregate tree to another), end-to-end data delivery is prone to be out of order. This fact can be considered a trade-off between bandwidth optimization and network stability. Therefore, such a solution is expected to promote awareness about this kind of drawback.

5.11. Fate-Sharing between Unicast and Multicast

In native Ethernet, multicast and unicast connectivity are often managed together. For instance, an 802.1ag CFM Continuity Check message is forwarded by multicast as a periodic heartbeat, but it is supposed to check the "whole" traffic continuity regardless of unicast or multicast, at the same time. Hence, the aliveness of

unicast and multicast is naturally coupled (i.e., fate-shared) in this customer's environment.

A multicast VPLS solution may decouple the path that a customer's unicast and multicast traffic follow through a SP's backbone, in order to provide the most optimal path for multicast data traffic. This may cause concern among some multicast VPLS customers who desire that, during a failure in the SP's network, both unicast and multicast traffic fail concurrently.

Therefore, there will be an additional requirement that makes both unicast and multicast connectivity coupled. This means that if either one of them have a failure, the other is also disabled. If one of the services (either unicast or multicast) becomes operational, the other is also activated simultaneously.

- It SHOULD be identified if the solution can provide customers with fate-sharing between unicast and multicast connectivity for their LAN switching application. It MAY have a configurable mechanism for SPs to provide that on behalf of customers, e.g., aliveness synchronization, but its use is OPTIONAL.

This policy will benefit customers. Some customers would like to detect failure soon at CE side and restore full connectivity by switching over to their backup line, rather than to keep poor half connectivity (i.e., either unicast or multicast being in fail). Even if either unicast or multicast is kept alive, it is just disadvantageous to the customer's application protocols that need both types of traffic. Fate-sharing policy contributes to preventing such a complicated situation.

Note that how serious this issue is depends on each customer's stance in Ethernet operation. If all CEs are IP routers, i.e., if VPLS is provided for a LAN routing application, the customer might not care about it because both unicast and multicast connectivity is assured in the IP layer. If the CE routers are running an IGP (e.g., OSPF/IS-IS) and a multicast routing protocol (e.g., PIM), then aliveness of both the unicast and multicast paths will be detected by the CEs. This does not guarantee that unicast and multicast traffic are to follow the same path in the SP's backbone network, but does mitigate this issue to some degree.

6. Service Provider Network Requirements

6.1. Scalability

The existing VPLS architecture has major advantages in scalability. For example, P-routers are free from maintaining customers' information because customer traffic is encapsulated in PSN tunnels. Also, a PW's split-horizon technique can prevent loops, making PE routers free from maintaining complicated spanning trees.

However, a multicast VPLS needs additional scalability considerations related to its expected enhanced mechanisms. [RFC3809] lists common L2VPN sizing and scalability requirements and metrics, which are applicable in multicast VPLS too. Accordingly, this section deals with specific requirements related to scalability.

6.1.1. Trade-Off of Optimality and State Resource

A solution needs to improve the scalability of multicast as is shown in Section 3:

Issue A: Replication to non-member site.

Issue B: Replication of PWs on shared physical path.

For both issues, the optimization of physical resources (i.e., link bandwidth usage and router duplication performance) will become a major goal. However, there is a trade-off between optimality and state resource consumption.

In order to solve Issue A, a PE might have to maintain multicast group information for CEs that was not kept in the existing VPLS solutions. This will present scalability concerns about state resources (memory, CPU, etc.) and their maintenance complexity.

In order to solve Issue B, PE and P routers might have to have knowledge of additional membership information for remote PEs, and possibly additional tree topology information, when they are using point-to-multipoint (P2MP) techniques (PIM tree, P2MP-LSP (Label Switched Path), etc.).

Consequently, the scalability evaluation of multicast VPLS solutions needs a careful trade-off analysis between bandwidth optimality and state resource consumption.

6.1.2. Key Metrics for Scalability

(Note: This part has a number of similar characteristics to requirements for Layer-3 Multicast VPN [RFC4834].)

A multicast VPLS solution MUST be designed to scale well with an increase in the number of any of the following metrics:

- the number of PEs
- the number of VPLS instances (total and per PE)
- the number of PEs and sites in any VPLS instance
- the number of client VLAN IDs
- the number of client Layer-2 MAC multicast groups
- the number of client Layer-3 multicast channels (groups or source-groups)
- the number of PWs and PSN Tunnels (MDTunnels) (total and per PE)

Each multicast VPLS solution SHALL document its scalability characteristics in quantitative terms. A solution SHOULD quantify the amount of state that a PE and a P device has to support.

The scalability characteristics SHOULD include:

- the processing resources required by the control plane in managing PWs (neighborhood or session maintenance messages, keepalives, timers, etc.)
- the processing resources required by the control plane in managing PSN tunnels
- the memory resources needed for the control plane
- the amount of protocol information transmitted to manage a multicast VPLS (e.g., signaling throughput)
- the amount of Layer-2/Layer-3 multicast information a P/PE router consumes (e.g., traffic rate of join/leave, keepalives, etc.)
- the number of multicast IP addresses used (if IP multicast in ASM mode is proposed as a multicast distribution tunnel)

- other particular elements inherent to each solution that impact scalability

Another metric for scalability is operational complexity. Operations will naturally become more complicated if the number of managed objects (e.g., multicast groups) increases, or the topology changes occur more frequently. A solution SHOULD note the factors that lead to additional operational complexity.

6.2. Tunneling Requirements

6.2.1. Tunneling Technologies

An MDTunnel denotes a multicast distribution tunnel. This is a generic term for tunneling where customer multicast traffic is carried over a provider's network. In the L2VPN service context, it will correspond to a PSN tunnel.

A solution SHOULD be able to use a range of tunneling technologies, including point-to-point (unicast oriented) and point-to-multipoint/ multipoint-to-multipoint (multicast oriented). For example, today there are many kinds of protocols for tunneling such as L2TP, IP, (including multicast IP trees), MPLS (including P2MP-LSP [RFC4875], and P2MP/MP2MP-LSP [LDP-P2MP]), etc.

Note that which variant, point-to-point, point-to-multipoint, or multipoint-to-multipoint, is used depends largely on the trade-offs mentioned above and the targeted network and applications. Therefore, this document does not mandate any specific protocols. A solution, however, SHOULD state reasonable criteria if it adopts a specific kind of tunneling protocol.

6.2.2. MTU of MDTunnel

From the view of an SP, it is not acceptable to have fragmentation/ reassembly so often while packets are traversing a MDTunnel. Therefore, a solution SHOULD support a method that provides the minimum path MTU of the MDTunnel in order to accommodate the service MTU.

6.3. Robustness

Multicast VPLS solutions SHOULD avoid single points of failures or propose technical solutions that make it possible to implement a failover mechanism.

6.4. Discovering Related Information

The operation of a multicast VPLS solution SHALL be as light as possible, and providing automatic configuration and discovery SHOULD be considered a high priority.

Therefore, in addition to the L2VPN discovery requirements in [RFC4665], a multicast VPLS solution SHOULD provide a method that dynamically allows multicast membership information to be discovered by PEs if the solution supports (A), as defined in Section 3.2. This means, a PE needs to discover multicast membership (e.g., join group addresses) that is controlled dynamically from the sites connected to that PE. In addition, a PE needs to discover such information automatically from other remote PEs as well in order to limit flooding scope across the backbone.

6.5. Operation, Administration, and Maintenance

6.5.1. Activation

The activation of multicast enhancement in a solution MUST be possible:

- with a VPLS instance granularity.
- with an Attachment Circuit granularity (i.e., with a PE-CE Ethernet port granularity, or with a VLAN ID granularity when it is a service delimiter).

Also it SHOULD be possible:

- with a CE granularity (when multiple CEs of the same VPN are associated with a common VPLS instance).
- with a distinction between multicast reception and emission.
- with a multicast MAC address granularity.
- with a customer IP multicast group and/or channel granularity (when Layer-3 information is consulted).

Also it MAY be possible:

- with a VLAN ID granularity when it is not a service delimiter.

6.5.2. Testing

A solution MUST provide a mechanism for testing multicast data connectivity and verifying the associated information. Examples that SHOULD be supported that are specific to multicast are:

- Testing connectivity per multicast MAC address
- Testing connectivity per multicast Layer-3 group/channel
- Verifying data plane and control plane integrity (e.g., PW, MDTunnel)
- Verifying multicast membership-relevant information (e.g., multicast MAC-addresses/PW-ports associations, Layer-3 group associations)

Operators usually want to test if an end-to-end multicast user's connectivity is OK before and after activation. Such end-to-end multicast connectivity checking SHOULD enable the end-to-end testing of the data path used by that customer's multicast data packets. Specifically, end-to-end checking will have a CE-to-CE path test and PE-to-PE path test. A solution MUST support the PE-to-PE path test and MAY support the CE-to-CE path test.

Also, operators will want to make use of a testing mechanism for diagnosis and troubleshooting. In particular, a solution SHOULD be able to monitor information describing how client multicast traffic is carried over the SP network. Note that if a solution supports frequent dynamic membership changes with optimized transport, troubleshooting within the SP's network will tend to be difficult.

6.5.3. Performance Management

Mechanisms to monitor multicast-specific parameters and statistics MUST be offered to the SP.

(Note: This part has a number of similar characteristics to requirements for Layer 3 Multicast VPN [RFC4834].)

A solution MUST provide SPs with access to:

- Multicast traffic statistics (total traffic forwarded, incoming, outgoing, dropped, etc., by period of time).

A solution SHOULD provide access to:

- Information about a customer's multicast resource usage (the amount of multicast state and throughput).

- Performance information related to multicast traffic usage, e.g., one-way delay, jitter, loss, delay variations (the difference in one-way delay for a solitary multicast group/stream from a single, source PE to multiple receiver PEs), etc.

- Alarms when limits are reached on such resources.

- Statistics on decisions related to how client traffic is carried on MDTunnels (e.g., "How much traffic was switched onto a multicast tree dedicated to such groups or channels").

- Statistics on parameters that could help the provider to evaluate its optimality/state trade-off.

All or part of this information SHOULD be made available through standardized SNMP MIB Modules (Management Information Base).

6.5.4. Fault Management

A multicast VPLS solution needs to consider those management steps taken by SPs below:

o Fault detection

A solution MUST provide tools that detect group membership/ reachability failure and traffic looping for multicast transport. It is anticipated that such tools are coordinated with the testing mechanisms mentioned in Section 6.5.2.

In particular, such mechanisms SHOULD be able to detect a multicast failure quickly, (on par with unicast cases). It SHOULD also avoid situations where multicast traffic has been in a failure state for a relatively long time while unicast traffic remains operational. If such a situation were to occur, it would end up causing problems with customer applications that depend on a combination of unicast and multicast forwarding.

With multicast, there may be many receivers associated with a particular multicast stream/group. As the number of receivers increases, the number of places (typically nearest the receivers) required to detect a fault will increase proportionately. This raises concerns over the scalability of

fault detection in large multicast deployments. Consequently, a fault detection solution SHOULD scale well; in particular, a solution should consider key metrics for scalability as described in Section 6.1.2.

- o Fault notification

  A solution MUST also provide fault notification and trouble tracking mechanisms (e.g., SNMP-trap and syslog).

  In case of multicast, one point of failure often affects a number of downstream routers/receivers that might be able to raise a notification. Hence, notification messages MAY be summarized or compressed for operators' ease of management.

- o Fault isolation

  A solution MUST provide diagnostic/troubleshooting tools for multicast as well. Also, it is anticipated that such tools are coordinated with the testing mechanisms mentioned in Section 6.5.2.

  In particular, a solution needs to correctly identify the area inside a multicast group impacted by the failure. A solution SHOULD be able to diagnose if an entire multicast group is faulty or if some specific destinations are still alive.

6.6. Security

6.6.1. Security Threat Analysis

In multicast VPLS, there is a concern that one or more customer nodes (presumably untrusted) might cause multicast-related attacks to the SP network. There is a danger that it might compromise some components that belong to the whole system.

This subsection states possible security threats relevant to the system and whether or not they are protected against.

General security consideration about a base VPLS (as part of L2VPNs) is referred to in [RFC4665]. The following is the threat analysis list that is inherent to multicast VPLS.

(a) Attack by a huge amount of multicast control packets.

There is a threat that a CE joins too many multicast groups and causes Denial of Service (DoS). This is caused by sending a large number of join/prune messages in a short time and/or

putting a large variety of group addresses in join/prune messages. This attack will waste PE's control resources (e.g., CPU, memory) that examine customer control messages (for solving Issue A in Section 3.2), and it will not continue expected services for other trusted customers.

(b) Attack by invalid/malformed multicast control packets.

There is a threat that a CE sends invalid or malformed control packets that might corrupt PE, which will cause a DoS attack. In particular, a CE might be spoofing legitimate source/group IP multicast addresses in such control packets (in PIM, IGMP, etc.) and source/destination MAC addresses as Layer-2 frames.

(c) Attack by rapid state change of multicast.

If a malicious CE changes multicast state by sending control packets in an extremely short period, this might affect PE's control resources (e.g., CPU, memory) to follow such state changes. Besides, it might also affect PE/P's control resources if MDTunnel inside the core is dynamically created in conjunction with customer's multicast group.

(d) Attack by high volume of multicast/broadcast data traffic.

A malicious CE might send a very high volume of multicast and/or broadcast data to a PE. If that PE does not provide any safeguards, it will cause excessive replication in the SP network and the bandwidth resources for other trusted customers might be exhausted.

(e) Attack by high volume of unknown destination unicast data traffic.

A malicious CE can send a high volume of unknown unicast to a PE. Generally, according to VPLS architecture, that PE must flood such unknown traffic to all corresponding PEs in the same VPN. A variety of unknown destinations and huge amount of such frames might cause excess traffic in SP network unless there is an appropriate safeguard provided.

6.6.2. Security Requirements

Based on the analysis in the previous subsection, the security requirements from the SP's perspective are shown as follows.

An SP network MUST be invulnerable to malformed or maliciously constructed customer traffic. This applies to both multicast data packets and multicast control packets.

Moreover, because multicast, broadcast, and unknown-unicast need more resources than unicast, an SP network MUST have safeguards against unwanted or malicious multicast traffic. This applies to both multicast data packets and multicast control packets.

Specifically, a multicast VPLS solution SHOULD have mechanisms to protect an SP network from:

(1) invalid multicast MAC addresses

(2) invalid multicast IP addresses

(3) malformed Ethernet multicast control protocol frames

(4) malformed IP multicast control protocol packets

(5) high volumes of

* valid/invalid customer control packets

* valid/invalid customer data packets (broadcast/multicast/ unknown-unicast)

Depending on each solution's actual approach to tackle with Issue A, or B, or both (see Section 3.2.), there are relationships to be highlighted about each item's importance listed above. First off, protection against (3) and (4) becomes significantly important if a solution supports solving Issue A, and PEs are processing customer's Ethernet/IP multicast control messages from CE. Moreover, protection against (2) should also be much focused because PIM/IGMP snooping will usually require that PE's data forwarding be based on IP addresses. By contrast, however, if a solution is solving only Issue B, not A, then PEs might never process the customer's multicast control messages at all; they do not perform IP address-based forwarding, but they do perform native Ethernet forwarding. If so, there is relatively less danger about (2), (3), and (4) compared to the first case.

The following are a few additional guidelines in detail.

For protecting against threat (a), a solution SHOULD support imposing some bounds on the quantity of state used by a VPN to be imposed in order to prevent state resource exhaustion (i.e., lack of memory, CPU etc.). In this case, the bounds MUST be

configurable per VPN basis, not the total of various VPNs so that SP can isolate the resource waste that is caused by any malicious customer.

For protecting against threat (d) and (e), a solution SHOULD support performing traffic policing to limit the unwanted data traffic shown above. In this case, while policing MAY be configurable to the sum of unicast, multicast, broadcast, and unknown unicast traffic, it SHOULD also be configurable to each such type of traffic individually in order to prevent physical resource exhaustion (i.e., lack of bandwidth and degradation of throughput). If the policing limit is configured on total traffic only, there will be a concern that one customer's huge multicast might close other irrelevant unicast traffic. If it can be configured individually, this concern will be avoided. Moreover, such a policing mechanism MUST be configurable per VPN basis, not the total of various VPNs to isolate malicious customer's traffic from others.

For protecting against threat (c), a solution SHOULD be able to limit frequent changes of group membership by customers. For example, PEs might support a dampening mechanism that throttles their multicast state changes if the customers are changing too excessively. Also, if MDTunnel is provided being tightly coupled to dynamic changes of customer's multicast domain, it is also effective to delay building the tunnel when customer's state is changed frequently.

Protecting against threat (b) might not be an easy task. Generally, checking the legitimacy of a customer's IP multicast control packets will eventually require the authentication between PE and CE in Layer-3; however, L2VPN (including VPLS) by its nature does not usually assume Layer-3-based security mechanism supported at PE-CE level.

The ramification of this fact is that there remains a possibility that a PE's control plain might be badly affected by corrupted multicast control packets that the PE is examining. Hence, each PE implementation will need to make an effort to minimize this impact from malicious customers and isolate it from other trusted customers as much as possible.

Nevertheless, it is possible to mitigate this threat to some degree. For example, a PE MAY support a filter mechanism about MAC and IP addresses in a Layer-2/Layer-3 header and a filter mechanism about source/group addresses in the multicast join/prune messages. This will help a PE to validate customers' control messages, to a certain extent.

6.7. Hierarchical VPLS support

A VPLS multicast solution SHOULD allow a hierarchical VPLS (H-VPLS) [RFC4762] service model. In other words, a solution is expected to operate seamlessly with existing hub and spoke PW connectivity.

Note that it is also important to take into account the case of redundant spoke connections between U-PEs and N-PEs.

6.8. L2VPN Wholesale

A solution MUST allow a situation where one SP is offering L2VPN services to another SP. One example here is a wholesale model where one VPLS interconnects other SPs' VPLS or 802.1D network islands. For customer SPs, their multicast forwarding can be optimized by making use of multicast VPLS in the wholesaler SP.

7. Security Considerations

Security concerns and requirements for a base VPLS solution are described in [RFC4665].

In addition, there are security considerations specific to multicast VPLS. Thus, a set of security issues have been identified that MUST be addressed when considering the design and deployment of multicast VPLS. Such issues have been described in Sections 5.5 and 6.6.

In particular, security requirements from the view of customers are shown in Section 5.5. Security requirements from the view of providers are shown in Section 6.6. Section 6.6.1 conducts security threat analysis about the provider's whole system. Section 6.6.2 explains how each threat can be addressed or mitigated.

8. Acknowledgments

The authors thank the contributors of [RFC4834] since the structure and content of this document were, for some sections, largely inspired by [RFC4834].

The authors also thank Yuichi Ikejiri, Jerry Ash, Bill Fenner, Vach Kompella, Shane Amante, Ben Niven-Jenkins, and Venu Hemige for their valuable reviews and feedback.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4665] Augustyn, W. and Y. Serbest, "Service Requirements for Layer 2 Provider-Provisioned Virtual Private Networks", RFC 4665, September 2006.

9.2. Informative References

[802.1D] IEEE Std 802.1D-2004, "IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges", 2004.

[802.1ag] IEEE Std 802.1ag-2007, "Virtual Bridged Local Area Networks - Amendment 5: Connectivity Fault Management", 2007.

[802.1s] IEEE Std 802.1s-2002, "Virtual Bridged Local Area Networks - Amendment 3: Multiple Spanning Trees", 2002.

[CGMP] Farinacci, D., Tweedly, A., and T. Speakman, "Cisco Group Management Protocol (CGMP)", 1996/1997, <ftp://ftpeng.cisco.com/ipmulticast/specs/cgmp.txt>.

[LDP-P2MP] Minei, I., Ed., Kompella, K., Wijnands, I., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", Work in Progress, May 2008.

[RFC1112] Deering, S., "Host extensions for IP multicasting", STD 5, RFC 1112, August 1989.

[RFC2236] Fenner, W., "Internet Group Management Protocol, Version 2", RFC 2236, November 1997.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[RFC2710] Deering, S., Fenner, W., and B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", RFC 2710, October 1999.

[RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.

[RFC3488] Wu, I. and T. Eckert, "Cisco Systems Router-port Group Management Protocol (RGMP)", RFC 3488, February 2003.

[RFC3809] Nagarajan, A., "Generic Requirements for Provider Provisioned Virtual Private Networks (PPVPN)", RFC 3809, June 2004.

[RFC3810] Vida, R. and L. Costa, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, June 2004.

[RFC3973] Adams, A., Nicholas, J., and W. Siadak, "Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)", RFC 3973, January 2005.

[RFC4541] Christensen, M., Kimball, K., and F. Solensky, "Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches", RFC 4541, May 2006.

[RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC 4601, August 2006.

[RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", RFC 4607, August 2006.

[RFC4664] Andersson, L. and E. Rosen, "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, September 2006.

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

[RFC4834] Morin, T., Ed., "Requirements for Multicast in Layer 3 Provider-Provisioned Virtual Private Networks (PPVPNs)", RFC 4834, April 2007.

[RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.

[RFC5015] Handley, M., Kouvelas, I., Speakman, T., and L. Vicisano, "Bidirectional Protocol Independent Multicast (BIDIR-PIM)", RFC 5015, October 2007.

Authors' Addresses

Yuji Kamite (editor)
NTT Communications Corporation
Granpark Tower
3-4-1 Shibaura, Minato-ku
Tokyo 108-8118
Japan
EMail: y.kamite@ntt.com

Yuichiro Wada
NTT
3-9-11 Midori-cho
Musashino-shi
Tokyo 180-8585
Japan
EMail: wada.yuichiro@lab.ntt.co.jp

Yetik Serbest
AT&T Labs
9505 Arboretum Blvd.
Austin, TX 78759
USA
EMail: yetik_serbest@labs.att.com

Thomas Morin
France Telecom R&D
2, avenue Pierre-Marzin
22307 Lannion Cedex
France
EMail: thomas.morin@francetelecom.com

Luyuan Fang
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
USA
EMail: lufang@cisco.com