

Network Working Group
Internet-Draft
Updates: [4761](#) (if approved)
Intended status: Standards Track
Expires: March 16, 2019

B. Kothari
Augtera Networks
K. Kompella
Juniper Networks
W. Henderickx
Nokia
F. Balus
Cisco
J. Uttaro
AT&T
Sep 12, 2018

**BGP based Multi-homing in Virtual Private LAN Service
draft-ietf-bess-vpls-multihoming-02.txt**

Abstract

Virtual Private LAN Service (VPLS) is a Layer 2 Virtual Private Network (VPN) that gives its customers the appearance that their sites are connected via a Local Area Network (LAN). It is often required for the Service Provider (SP) to give the customer redundant connectivity to some sites, often called "multi-homing". This memo shows how BGP-based multi-homing can be offered in the context of LDP and BGP VPLS solutions.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on March 16, 2019.

Copyright Notice

Copyright (c) 2018 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
1.1.	General Terminology	3
1.2.	Conventions	4
2.	Background	4
2.1.	Scenarios	4
2.2.	VPLS Multi-homing Considerations	5
3.	Multi-homing Operation	6
3.1.	Customer Edge (CE) NLRI	6
3.2.	Deployment Considerations	7
3.3.	Designated Forwarder Election	8
3.3.1.	Attributes	8
3.3.2.	Variables Used	9
3.3.3.	Election Procedures	11
3.4.	DF Election on PEs	12
3.5.	Pseudowire and Site-ID Binding Properties	13
4.	Multi-AS VPLS	13
4.1.	Route Origin Extended Community	13
4.2.	VPLS Preference	13
4.3.	Use of BGP attributes in Inter-AS Methods	14
	4.3.1. Inter-AS Method (b): EBGW Redistribution of VPLS Information between ASBRs	15
	4.3.2. Inter-AS Method (c): Multi-Hop EBGW Redistribution of VPLS Information between ASes	16
5.	MAC Flush Operations	16
5.1.	MAC Flush Indicators	16
5.2.	Minimizing the effects of fast link transitions	17
6.	Backwards Compatibility	17
6.1.	BGP based VPLS	18
6.2.	LDP VPLS with BGP Auto-discovery	18
7.	Security Considerations	18
8.	IANA Considerations	18
9.	Contributing Authors	18
10.	Acknowledgments	19
11.	References	19
11.1.	Normative References	19
11.2.	Informative References	19

Authors' Addresses [20](#)

[1.](#) Introduction

Virtual Private LAN Service (VPLS) is a Layer 2 Virtual Private Network (VPN) that gives its customers the appearance that their sites are connected via a Local Area Network (LAN). It is often required for a Service Provider (SP) to give the customer redundant connectivity to one or more sites, often called "multi-homing". [\[RFC4761\]](#) explains how VPLS can be offered using BGP for auto-discovery and signaling; [section 3.5](#) of that document describes how multi-homing can be achieved in this context. [\[RFC6074\]](#) explains how VPLS can be offered using BGP for auto-discovery (BGP-AD) and [\[RFC4762\]](#) explains how VPLS can be offered using LDP for signaling. This document provides a BGP-based multi-homing solution applicable to both BGP and LDP VPLS technologies. Note that BGP MH can be used for LDP VPLS without the use of the BGP-AD solution.

[Section 2](#) lays out some of the scenarios for multi-homing, other ways that this can be achieved, and some of the expectations of BGP-based multi-homing. [Section 3](#) defines the components of BGP-based multi-homing, and the procedures required to achieve this.

[1.1.](#) General Terminology

Some general terminology is defined here; most is from [\[RFC4761\]](#), [\[RFC4762\]](#) or [\[RFC4364\]](#). Terminology specific to this memo is introduced as needed in later sections.

A "Customer Edge" (CE) device, typically located on customer premises, connects to a "Provider Edge" (PE) device, which is owned and operated by the SP. A "Provider" (P) device is also owned and operated by the SP, but has no direct customer connections. A "VPLS Edge" (VE) device is a PE that offers VPLS services.

A VPLS domain represents a bridging domain per customer. A Route Target community as described in [\[RFC4360\]](#) is typically used to identify all the PE routers participating in a particular VPLS domain. A VPLS site is a grouping of ports on a PE that belong to the same VPLS domain. The terms "VPLS instance" and "VPLS domain" are used interchangeably in this document.

A VPLS site is a grouping of ports on a PE that belong to the same VPLS domain. The terms "VPLS instance" and "VPLS domain" are used interchangeably in this document.

If the CE devices that connect to a VPLS site's ports have connectivity to any other PE device then the VPLS site is called a

multi-homed VPLS site. Otherwise, it is called a single-homed VPLS site. The ports are partitioned between VPLS sites such that each port is in no more than one VPLS site. The terms "VPLS site" and "CE site" are used interchangeably in this document.

A BGP VPLS NLRI for the base VPLS instance that has non-zero VE block offset, VE block size and label base is called as VE NLRI in this document. Each VPLS instance is uniquely identified by a VE-ID. VE-ID is carried in the BGP VPLS NLRI as specified in [section 3.2.2 in \[RFC4761\]](#).

A VPLS NLRI with value zero for the VE block offset, VE block size and label base is called as CE NLRI in this document. Section [Section 3.1](#) defines CE NLRI and provides more detail.

A Multi-homed (MH) site is uniquely identified by a CE-ID. Sites are referred to as local or remote depending on whether they are configured on the PE router in context or on one of the remote PE routers (network peers). A single-homed site can also be assigned a CE-ID, but it is not mandatory to configure a CE-ID for single-homed sites. Section [Section 3.1](#) provides detail on CE-ID.

[1.2. 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 [\[RFC2119\]](#).

[2. Background](#)

This section describes various scenarios where multi-homing may be required, and the implications thereof. It also describes some of the singular properties of VPLS multi-homing, and what that means from both an operational point of view and an implementation point of view. There are other approaches for providing multi-homing such as Spanning Tree Protocol, and this document specifies use of BGP for multi-homing. Comprehensive comparison among the approaches is outside the scope of this document.

[2.1. Scenarios](#)

CE1 is a VPLS CE that is dual-homed to both PE1 and PE2 for redundant connectivity.

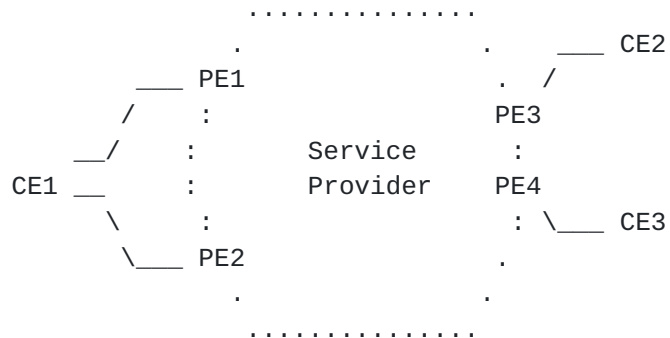


Figure 1: Scenario 1

CE1 is a VPLS CE that is dual-homed to both PE1 and PE2 for redundant connectivity. However, CE4, which is also in the same VPLS domain, is single-homed to just PE1.

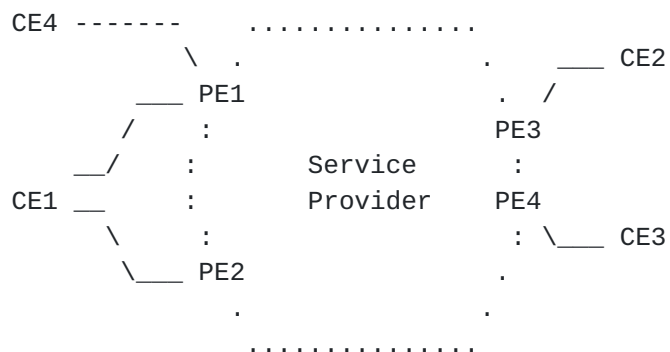


Figure 2: Scenario 2

2.2. VPLS Multi-homing Considerations

The first (perhaps obvious) fact about a multi-homed VPLS CE, such as CE1 in Figure 1 is that if CE1 is an Ethernet switch or bridge, a loop has been created in the customer VPLS. This is a dangerous situation for an Ethernet network, and the loop must be broken. Even if CE1 is a router, it will get duplicates every time a packet is flooded, which is clearly undesirable.

The next is that (unlike the case of IP-based multi-homing) only one of PE1 and PE2 can be actively sending traffic, either towards CE1 or into the SP cloud. That is to say, load balancing techniques will not work. All other PEs MUST choose the same designated forwarder for a multi-homed site. Call the PE that is chosen to send traffic to/from CE1 the "designated forwarder".

In Figure 2, CE1 and CE4 must be dealt with independently, since CE1 is dual-homed, but CE4 is not.

3. Multi-homing Operation

This section describes procedures for electing a designated forwarder among the set of PEs that are multi-homed to a customer site. The procedures described in this section are applicable to BGP based VPLS, LDP based VPLS with BGP-AD or a VPLS that contains a mix of both BGP and LDP signaled PWs.

3.1. Customer Edge (CE) NLRI

[Section 3.2.2 in \[RFC4761\]](#) specifies a NLRI to be used for BGP based VPLS (BGP VPLS NLRI). The format of the BGP VPLS NLRI is shown below.

```

+-----+
| Length (2 octets)                |
+-----+
| Route Distinguisher (8 octets)   |
+-----+
| VE ID (2 octets)                 |
+-----+
| VE Block Offset (2 octets)       |
+-----+
| VE Block Size (2 octets)         |
+-----+
| Label Base (3 octets)            |
+-----+

```

BGP VPLS NLRI

For multi-homing operation, a customer-edge NLRI (CE NLRI) is proposed that uses BGP VPLS NLRI with the following fields set to zero: VE Block Offset, VE Block Size and Label Base. In addition, the VE-ID field of the NLRI is set to CE-ID. Thus, the CE NLRI contains 2 octets indicating the length, 8 octets for Route Distinguisher, 2 octets for CE-ID and 7 octets with value zero.

It is valid to have non-zero VE block offset, VE block size and label base in the VPLS NLRI for a multi-homed site. VPLS operations, including multi-homing, in such a case are outside the scope of this document. However, for interoperability with existing deployments that use non-zero VE block offset, VE block size and label base for multi-homing operation, [Section 6.1](#) provides more detail.

Wherever VPLS NLRI is used in this document, context must be used to infer if it is applicable for CE NLRI, VE NLRI or for both.

3.2. Deployment Considerations

It is mandatory that each instance within a VPLS domain MUST be provisioned with a unique Route Distinguisher value. Unique Route Distinguisher allows VPLS advertisements from different VPLS PEs to be distinct even if the advertisements have the same VE-ID, which can occur in case of multi-homing. This allows standard BGP path selection rules to be applied to VPLS advertisements.

Each VPLS PE must advertise a unique VE-ID with non-zero VE Block Offset, VE Block Size and Label Base values in the BGP NLRI. VE-ID is associated with the base VPLS instance and the NLRI associated with it must be used for creating PWs among VPLS PEs. Any single-homed customer sites connected to the VPLS instance do not require any special addressing. However, an administrator (SP operator) can choose to have a CE-ID for a single-homed site as well. Any multi-homed customer sites connected to the VPLS instance require special addressing, which is achieved by use of CE-ID. A set of customer sites are distinguished as multi-homed if they all have the same CE-ID. The following examples illustrate the use of VE-ID and CE-ID.

Figure 1 shows a customer site, CE1, multi-homed to two VPLS PEs, PE1 and PE2. In order for all VPLS PEs to set up PWs to each other, each VPLS PE must be configured with a unique VE-ID for its base VPLS instance. In addition, in order for all VPLS PEs within the same VPLS domain to elect one of the multi-homed PEs as the designated forwarder, an indicator that the PEs are multi-homed to the same customer site is required. This is achieved by assigning the same VPLS site ID (CE-ID) on PE1 and PE2 for CE1. When remote VPLS PEs receive NLRI advertisement from PE1 and PE2 for CE1, the two NLRI advertisements for CE1 are identified as candidates for designated forwarder selection due to the same CE-ID. Thus, same CE-ID MUST be assigned on all VPLS PEs that are multi-homed to the same customer site.

Figure 2 shows two customer sites, CE1 and CE4, connected to PE1 with CE1 multi-homed to PE1 and PE2. Similar to Figure 1 provisioning model, each VPLS PE must be configured with a unique VE-ID for its base VPLS instance. CE1 which is multi-homed to PE1 and PE2 requires configuration of CE-ID and both PE1 and PE2 MUST be provisioned with the same CE-ID for CE1. CE2 and CE3 are single-homed sites and do not require special addressing. However, an operator must configure a CE-ID for CE4 on PE1. By doing so, remote PEs can determine that PE1 has two VPLS sites, CE1 and CE4. If both CE1 and CE4 connectivity to PE1 is down, remote PEs can choose based on D bit in

VE NLRI not to send multicast traffic to PE1 as there are no VPLS sites reachable via PE1. If CE4 was not assigned a unique CE-ID, remote PEs have no way to know if there are other VPLS sites attached and hence, would always send multicast traffic to PE1. While CE2 and CE3 can also be configured with unique CE-IDs, there is no advantage in doing so as both PE3 and PE4 have exactly one VPLS site.

Note that a CE-ID=0 is invalid and a PE should discard such an advertisement.

Use of multiple VE-IDs per VPLS instance for either multi-homing operation or for any other purpose is outside the scope of this document. However, for interoperability with existing deployments that use multiple VE-IDs, [Section 6.1](#) provides more detail.

3.3. Designated Forwarder Election

BGP-based multi-homing for VPLS relies on standard BGP path selection and VPLS DF election. The net result of doing both BGP path selection and VPLS DF election is that of electing a single designated forwarder (DF) among the set of PEs to which a customer site is multi-homed. All the PEs that are elected as non-designated forwarders MUST keep their attachment circuit to the multi-homed CE in blocked status (no forwarding).

These election algorithms operate on VPLS advertisements, which include both the NLRI and attached BGP attributes. These election algorithms are applicable to all VPLS NLRIs, and not just to CE NLRIs. In order to simplify the explanation of these algorithms, we will use a number of variables derived from fields in the VPLS advertisement. These variables are: RD, SITE-ID, VBO, DOM, ACS, PREF and PE-ID. The notation ADV -> <RD, SITE-ID, VBO, DOM, ACS, PREF, PE-ID> means that from a received VPLS advertisement ADV, the respective variables were derived. The following sections describe two attributes needed for DF election, then describe the variables and how they are derived from fields in VPLS advertisement ADV, and finally describe how DF election is done.

3.3.1. Attributes

The procedures below refer to two attributes: the Route Origin community (see [Section 4.1](#)) and the L2-info community (see [Section 4.2](#)). These attributes are required for inter-AS operation; for generality, the procedures below show how they are to be used. The procedures also outline how to handle the case that either or both are not present.

For BGP-based Multi-homing, ADV MUST contain an L2-info extended community as specified in [[RFC4761](#)]. Within this community are various control flags. Two new control flags are proposed in this document. Figure 3 shows the position of the new 'D' and 'F' flags.

Control Flags Bit Vector

```

0 1 2 3 4 5 6 7
+---+---+---+---+
|D|Z|F|Z|Z|Z|C|S| (Z = MUST Be Zero)
+---+---+---+---+

```

Figure 3

1. 'D' (Down): Indicates connectivity status. In case of CE NLRI, the connectivity status is between a CE site and a VPLS PE. In case of VE NLRI, the connectivity status is for the VPLS instance. In case of CE NLRI, the bit MUST be set to one if all the attachment circuits connecting a CE site to a VPLS PE are down. In case of VE NLRI, the bit must be set to one if the VPLS instance is operationally down. Note that a VPLS instance that has no connectivity to any of its sites must be considered as operationally down.
2. 'F' (Flush): Indicates when to flush MAC state. A designated forwarder must set the F bit and a non-designated forwarder must clear the F bit when sending BGP CE NLRIs for multi-homed sites. A state transition from one to zero for the F bit can be used by a remote PE to flush all the MACs learned from the PE that is transitioning from designated forwarder to non-designated forwarder. Refer to [Section 5](#) for more details on the use case.

[3.3.2.](#) Variables Used

[3.3.2.1.](#) RD

RD is simply set to the Route Distinguisher field in the NLRI part of ADV. Actual process of assigning Route Distinguisher values must guarantee its uniqueness per PE node. Therefore, two multi-homed PEs offering the same VPLS service to a common set of CEs MUST allocate different RD values for this site respectively.

[3.3.2.2.](#) SITE-ID

SITE-ID is simply set to the VE-ID field in the NLRI part of the ADV.

Note that no distinction is made whether VE-ID is for a multi-homed site or not.

3.3.2.3. VBO

VBO is simply set to the VE Block Offset field in the NLRI part of ADV.

3.3.2.4. DOM

This variable, indicating the VPLS domain to which ADV belongs, is derived by applying BGP policy to the Route Target extended communities in ADV. The details of how this is done are outside the scope of this document.

3.3.2.5. ACS

ACS is the status of the attachment circuits for a given site of a VPLS. ACS = 1 if all attachment circuits for the site are down, and 0 otherwise.

ACS is set to the value of the 'D' bit in ADV that belongs to CE NLRI. If ADV belongs to base VPLS instance (VE NLRI) with non-zero label block values, no change must be made to ACS.

3.3.2.6. PREF

PREF is derived from the Local Preference (LP) attribute in ADV as well as the VPLS Preference field (VP) in the L2-info extended community. If the Local Preference attribute is missing, LP is set to 0; if the L2-info community is missing, VP is set to 0. The following table shows how PREF is computed from LP and VP.

VP Value	LP Value	PREF Value	Comment
0	0	0	malformed advertisement, unless ACS=1
0	1 to (2 ¹⁶ -1)	LP	backwards compatibility
0	2 ¹⁶ to (2 ³² -1)	(2 ¹⁶ -1)	backwards compatibility
>0	LP same as VP	VP	Implementation supports VP
>0	LP != VP	0	malformed advertisement

Table 1

3.3.2.7. PE-ID

If ADV contains a Route Origin (RO) community (see [Section 4.1](#)) with type 0x01, then PE-ID is set to the Global Administrator sub-field of the RO. Otherwise, if ADV has an ORIGINATOR_ID attribute, then PE-ID is set to the ORIGINATOR_ID. Otherwise, PE-ID is set to the BGP Identifier.

3.3.3. Election Procedures

The election procedures described in this section apply equally to BGP VPLS and LDP VPLS. A distinction MUST NOT be made on whether the NLRI is a multi-homing NLRI or not. Subset of these procedures documented in standard BGP best path selection deals with general IP Prefix BGP route selection processing as defined in [[RFC4271](#)]. A separate part of the algorithm defined under VPLS DF election is specific to designated forwarded election procedures performed on VPLS advertisements. A concept of bucketization is introduced to define route selection rules for VPLS advertisements. Note that this is a conceptual description of the process; an implementation MAY choose to realize this differently as long as the semantics are preserved.

3.3.3.1. Bucketization for standard BGP path selection

An advertisement

ADV -> <RD, SITE-ID, VBO, ACS, PREF, PE-ID>

is put into the bucket for <RD, SITE-ID, VBO>. In other words, the information in BGP path selection consists of <RD, SITE-ID, VBO> and only advertisements with exact same <RD, SITE-ID, VBO> are candidates for BGP path selection procedure as defined in [[RFC4271](#)].

3.3.3.2. Bucketization for VPLS DF Election

An advertisement

ADV -> <RD, SITE-ID, VBO, DOM, ACS, PREF, PE-ID>

is discarded if DOM is not of interest to the VPLS PE. Otherwise, ADV is put into the bucket for <DOM, SITE-ID>. In other words, all advertisements for a particular VPLS domain that have the same SITE-ID are candidates for VPLS DF election.

3.3.3.3. Tie-breaking Rules

This section describes the tie-breaking rules for VPLS DF election. Tie-breaking rules for VPLS DF election are applied to candidate advertisements by all VPLS PEs and the actions taken by VPLS PEs based on the VPLS DF election result are described in [Section 3.4](#).

Given two advertisements ADV1 and ADV2 from a given bucket, first compute the variables needed for DF election:

```
ADV1 -> <RD1, SITE-ID1, VB01, DOM1, ACS1, PREF1, PE-ID1>
ADV2 -> <RD2, SITE-ID2, VB02, DOM2, ACS2, PREF2, PE-ID2>
```

Note that SITE-ID1 = SITE-ID2 and DOM1 = DOM2, since ADV1 and ADV2 came from the same bucket. Then the following tie-breaking rules MUST be applied in the given order.

1. if (ACS1 != 1) AND (ACS2 == 1) ADV1 wins; stop
 if (ACS1 == 1) AND (ACS2 != 1) ADV2 wins; stop
 else continue
2. if (PREF1 > PREF2) ADV1 wins; stop;
 else if (PREF1 < PREF2) ADV2 wins; stop;
 else continue
3. if (PE-ID1 < PE-ID2) ADV1 wins; stop;
 else if (PE-ID1 > PE-ID2) ADV2 wins; stop;
 else ADV1 and ADV2 are from the same VPLS PE

If there is no winner and ADV1 and ADV2 are from the same PE, a VPLS PE MUST retain both ADV1 and ADV2.

3.4. DF Election on PEs

DF election algorithm MUST be run by all multi-homed VPLS PEs. In addition, all other PEs SHOULD also run the DF election algorithm. As a result of the DF election, multi-homed PEs that lose the DF election for a SITE-ID MUST put the ACs associated with the SITE-ID in non-forwarding state.

DF election result on the egress PEs can be used in traffic forwarding decision. Figure 2 shows two customer sites, CE1 and CE4, connected to PE1 with CE1 multi-homed to PE1 and PE2. If PE1 is the designated forwarder for CE1, based on the DF election result, PE3 can choose to not send unknown unicast and multicast traffic to PE2 as PE2 is not the designated forwarder for any customer site and it has no other single homed sites connected to it.

3.5. Pseudowire and Site-ID Binding Properties

For the use case where a single PE provides connectivity to a set of CEs from which some are multi-homed and others are not, only single pseudowire MAY be established. For example, if PE1 provides VPLS service to CE1 and CE4 which are both part of the same VPLS domain, but different sites, and CE1 is multi-homed, but CE4 is not (as described in figure 2), PE3 would establish only single pseudowire toward PE1. A design needs to ensure that regardless of PE1's forwarding state in respect to DF or non-DF for multi-homed CE1, PE3's access to CE4 is established. Since label allocation and pseudowire established is tied to site-ID, we need to ensure that proper pseudowire bindings are established.

For set of given advertisements with the common DOM but with different Site-ID values, a VPLS PE speaker SHOULD instantiate and bind the pseudowire based on advertisement with the lowest Site-ID value. Otherwise, binding would be completely random and during DF changes for multi-homed site, non-multi-homed CE might suffer traffic loss.

4. Multi-AS VPLS

This section describes multi-homing in an inter-AS context.

4.1. Route Origin Extended Community

Due to lack of information about the PEs that originate the VPLS NLRIs in inter-AS operations, Route Origin Extended Community [[RFC4360](#)] is used to carry the source PE's IP address.

To use Route Origin Extended Community for carrying the originator VPLS PE's loopback address, the type field of the community MUST be set to 0x01 and the Global Administrator sub-field MUST be set to the PE's loopback IP address.

4.2. VPLS Preference

When multiple PEs are assigned the same site ID for multi-homing, it is often desired to be able to control the selection of a particular PE as the designated forwarder. [Section 3.5 in \[RFC4761\]](#) describes the use of BGP Local Preference in path selection to choose a particular NLRI, where Local Preference indicates the degree of preference for a particular VE. The use of Local Preference is inadequate when VPLS PEs are spread across multiple ASes as Local Preference is not carried across AS boundary. A new field, VPLS preference (VP), is introduced in this document that can be used to accomplish this. VPLS preference indicates a degree of preference

for a particular customer site. VPLS preference is not mandatory for intra-AS operation; the algorithm explained in [Section 3.3](#) will work with or without the presence of VPLS preference.

[Section 3.2.4 in \[RFC4761\]](#) describes the Layer2 Info Extended Community that carries control information about the pseudowires. The last two octets that were reserved now carries VPLS preference as shown in Figure 4.

```

+-----+
| Extended community type (2 octets) |
+-----+
| Encaps Type (1 octet)              |
+-----+
| Control Flags (1 octet)            |
+-----+
| Layer-2 MTU (2 octet)              |
+-----+
| VPLS Preference (2 octets)         |
+-----+

```

Figure 4: Layer2 Info Extended Community

A VPLS preference is a 2-octets unsigned integer. A value of zero indicates absence of a VP and is not a valid preference value. This interpretation is required for backwards compatibility. Implementations using Layer2 Info Extended Community as described in ([Section 3.2.4](#)) [[RFC4761](#)] MUST set the last two octets as zero since it was a reserved field.

For backwards compatibility, if VPLS preference is used, then BGP Local Preference MUST be set to the value of VPLS preference. Note that a Local Preference value of zero for a CE-ID is not valid unless 'D' bit in the control flags is set (see [[I-D.kothari-l2vpn-auto-site-id](#)]). In addition, Local Preference value greater than or equal to 2^{16} for VPLS advertisements is not valid.

4.3. Use of BGP attributes in Inter-AS Methods

[Section 3.4 in \[RFC4761\]](#) and [section 4 in \[RFC6074\]](#) describe three methods (a, b and c) to connect sites in a VPLS to PEs that are across multiple AS. Since VPLS advertisements in method (a) do not cross AS boundaries, multi-homing operations for method (a) remain exactly the same as they are within as AS. However, for method (b) and (c), VPLS advertisements do cross AS boundary. This section describes the VPLS operations for method (b) and method (c). Consider Figure 5 for inter-AS VPLS with multi-homed customer sites.

4.3.1. Inter-AS Method (b): EBGP Redistribution of VPLS Information between ASBRs

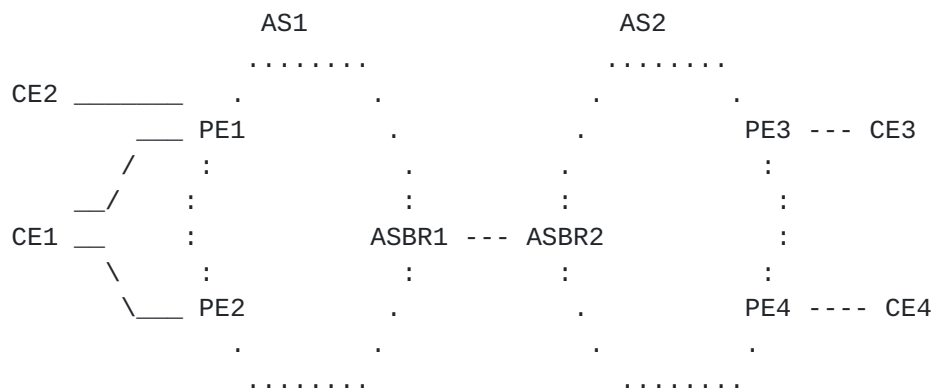


Figure 5: Inter-AS VPLS

A customer has four sites, CE1, CE2, CE3 and CE4. CE1 is multi-homed to PE1 and PE2 in AS1. CE2 is single-homed to PE1. CE3 and CE4 are also single homed to PE3 and PE4 respectively in AS2. Assume that in addition to the base LDP/BGP VPLS addressing (VSI-IDs/VE-IDs), CE-ID 1 is assigned for CE1. After running DF election algorithm, all four VPLS PEs must elect the same designated forwarder for CE1 site. Since BGP Local Preference is not carried across AS boundary, VPLS preference as described in [Section 4.2](#) MUST be used for carrying site preference in inter-AS VPLS operations.

For Inter-AS method (b) ASBR1 will send a VPLS NLRI received from PE1 to ASBR2 with itself as the BGP nexthop. ASBR2 will send the received NLRI from ASBR1 to PE3 and PE4 with itself as the BGP nexthop. Since VPLS PEs use BGP Local Preference in DF election, for backwards compatibility, ASBR2 MUST set the Local Preference value in the VPLS advertisements it sends to PE3 and PE4 to the VPLS preference value contained in the VPLS advertisement it receives from ASBR1. ASBR1 MUST do the same for the NLRIs it sends to PE1 and PE2. If ASBR1 receives a VPLS advertisement without a valid VPLS preference from a PE within its AS, then ASBR1 MUST set the VPLS preference in the advertisements to the Local Preference value before sending it to ASBR2. Similarly, ASBR2 must do the same for advertisements without VPLS Preference it receives from PEs within its AS. Thus, in method (b), ASBRs MUST update the VPLS and Local Preference based on the advertisements they receive either from an ASBR or a PE within their AS.

In Figure 5, PE1 will send the VPLS advertisements with Route Origin Extended Community containing its loopback address. PE2 will do the same. Even though PE3 receives the VPLS advertisements for VE-ID 1 and 2 from the same BGP nexthop, ASBR2, the source PE address contained in the Route Origin Extended Community is different for the CE1 and CE2 advertisements, and thus, PE3 creates two PWs, one for CE1 (for VE-ID 1) and another one for CE2 (for VE-ID 2).

4.3.2. Inter-AS Method (c): Multi-Hop EBGp Redistribution of VPLS Information between ASes

In this method, there is a multi-hop E-BGP peering between the PEs or Route Reflectors in AS1 and the PEs or Route Reflectors in AS2. There is no VPLS state in either control or data plane on the ASBRs. The multi-homing operations on the PEs in this method are exactly the same as they are in intra-AS scenario. However, since Local Preference is not carried across AS boundary, the translation of LP to VP and vice versa MUST be done by RR, if RR is used to reflect VPLS advertisements to other ASes. This is exactly the same as what a ASBR does in case of method (b). A RR must set the VP to the LP value in an advertisement before sending it to other ASes and must set the LP to the VP value in an advertisement that it receives from other ASes before sending to the PEs within the AS.

5. MAC Flush Operations

In a service provider VPLS network, customer MAC learning is confined to PE devices and any intermediate nodes, such as a Route Reflector, do not have any state for MAC addresses.

Topology changes either in the service provider's network or in customer's network can result in the movement of MAC addresses from one PE device to another. Such events can result into traffic being dropped due to stale state of MAC addresses on the PE devices. Age out timers that clear the stale state will resume the traffic forwarding, but age out timers are typically in minutes, and convergence of the order of minutes can severely impact customer's service. To handle such events and expedite convergence of traffic, flushing of affected MAC addresses is highly desirable.

5.1. MAC Flush Indicators

If 'D' bit in the control flags is set in a received VE NLRI, the receiving PE SHOULD flush all the MAC addresses learned from the PE advertising the failure.

Anytime a designated forwarder change occurs, a remote PE SHOULD flush all the MAC addresses it learned from the PE that lost the DF

election (old designated forwarder). If multiple customer sites are connected to the same PE, PE1 as shown in Figure 2, and redundancy per site is desired when multi-homing procedures described in this document are in effect, then it is desirable to flush just the relevant MAC addresses from a particular site when the site connectivity is lost. However, procedures for flushing a limited set of MAC addresses is beyond the scope of this document. Use of either 'D' or 'F' bit in control flags only allows to flush all MAC addresses associated with a PE.

Designated forwarder change can occur in absence of failures, such as when an attachment circuit comes up. Consider the case in Figure 2 where PE1-CE1 link is non-operational and PE2 is the designated forwarder for CE1. Also assume that Local Preference of PE1 is higher than PE2. When PE1-CE1 link becomes operational, PE1 will send a BGP CE advertisement for CE1 to all its peers. If PE3 performs the DF election before PE2, there is a chance that PE3 might learn MAC addresses from PE2 after it was done electing PE1. This can happen since PE2 has not yet processed the BGP CE advertisement from PE1 and as a result continues to send traffic to PE3. This can cause traffic from PE3 to CE1 to black-hole until those MAC addresses are deleted due to age out timers. Therefore, to avoid such race-conditions, a designated forwarder must set the F bit and a non-designated forwarder must clear the F bit when sending BGP CE advertisements. A state transition from one to zero for the 'F' bit can be used by a remote PE to flush all the MACs learned from the PE that is transitioning from designated forwarder to non-designated forwarder.

5.2. Minimizing the effects of fast link transitions

Certain failure scenarios may result in fast transitions of the link towards the multi-homing CE which in turn will generate fast status transitions of one or multiple multi-homed sites reflected through multiple BGP CE advertisements and LDP MAC Flush messages.

It is recommended that a timer to damp the link flaps be used for the port towards the multi-homed CE to minimize the number of MAC Flush events in the remote PEs and the occurrences of BGP state compression for F bit transitions. A timer value more than the time it takes BGP to converge in the network is recommended.

6. Backwards Compatibility

No forwarding loops are formed when PEs or Route Reflectors that do not support procedures defined in this section co exist in the network with PEs or Route Reflectors that do support.

6.1. BGP based VPLS

As explained in this section, multi-homed PEs to the same customer site MUST assign the same CE-ID and related NLRI SHOULD contain the block offset, block size and label base as zero. Remote PEs that lack support of multi-homing operations specified in this document will fail to create any PWs for the multi-homed CE-IDs due to the label value of zero and thus, the multi-homing NLRI should have no impact on the operation of Remote PEs that lack support of multi-homing operations specified in this document.

For compatibility with PEs that use multiple VE-IDs with non-zero label block values for multi-homing operation, it is a requirement that a PE receiving such advertisements must use the labels in the NLRIs associated with lowest VE-ID for PW creation. It is possible that maintaining PW association with lowest VE-ID can result in PW flap, and thus, traffic loss. However, it is necessary to maintain the association of PW with the lowest VE-ID as it provides deterministic DF election among all the VPLS PEs.

6.2. LDP VPLS with BGP Auto-discovery

The BGP-AD NLRI has a prefix length of 12 containing only a 8 bytes RD and a 4 bytes VSI-ID. If a LDP VPLS PEs running BGP AD lacks support of multi-homing operations specified in this document, it SHOULD ignore a CE NLRI with the length field of 17. As a result it will not ask LDP to create any PWs for the multi-homed Site-ID and thus, the multi-homing NLRI should have no impact on LDP VPLS operation. MH PEs may use existing LDP MAC Flush to flush the remote LDP VPLS PEs or may use the MAC Flush procedures as described in [Section 5](#)

7. Security Considerations

No new security issues are introduced beyond those that are described in [\[RFC4761\]](#) and [\[RFC4762\]](#).

8. IANA Considerations

At this time, this memo includes no request to IANA.

9. Contributing Authors

The authos would also like to thank Senad Palislamovic and Wen Lin for their contribution to the development of this document.

Senad Palislamovic
Nokia

Email: senad.palislamovic@nokia.com

Wen Lin

Juniper Networks

Email: wlin@juniper.net

10. Acknowledgments

The authors would like to thank Yakov Rekhter, Nischal Sheth, Mitali Singh, Ian Cowburn and Jonathan Hardwick for their insightful comments and probing questions.

11. References

11.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC4761] Kompella, K., Ed. and Y. Rekhter, Ed., "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", [RFC 4761](#), DOI 10.17487/RFC4761, January 2007, <<https://www.rfc-editor.org/info/rfc4761>>.
- [RFC6074] Rosen, E., Davie, B., Radoaca, V., and W. Luo, "Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)", [RFC 6074](#), DOI 10.17487/RFC6074, January 2011, <<https://www.rfc-editor.org/info/rfc6074>>.

11.2. Informative References

- [I-D.kothari-l2vpn-auto-site-id] Kothari, B., Kompella, K., and T. IV, "Automatic Generation of Site IDs for Virtual Private LAN Service", [draft-kothari-l2vpn-auto-site-id-01](#) (work in progress), October 2008.
- [RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), DOI 10.17487/RFC4271, January 2006, <<https://www.rfc-editor.org/info/rfc4271>>.
- [RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", [RFC 4360](#), DOI 10.17487/RFC4360, February 2006, <<https://www.rfc-editor.org/info/rfc4360>>.

- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", [RFC 4364](#), DOI 10.17487/RFC4364, February 2006, <<https://www.rfc-editor.org/info/rfc4364>>.
- [RFC4762] Lasserre, M., Ed. and V. Kompella, Ed., "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", [RFC 4762](#), DOI 10.17487/RFC4762, January 2007, <<https://www.rfc-editor.org/info/rfc4762>>.

Authors' Addresses

Bhupesh Kothari
Augtera Networks

Email: bhupesh@anvaya.net

Kireeti Kompella
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: kireeti.kompella@gmail.com

Wim Henderickx
Nokia

Email: wim.henderickx@nokia.com

Florin Balus
Cisco

Email: fbalus@gmail.com

James Uttaro
AT&T
200 S. Laurel Avenue
Middletown, NJ 07748
US

Email: uttaro@att.com

