Internet Engineering Task Force (IETF)
Request for Comments: 8614
Updates: 4761
Category: Standards Track
ISSN: 2070-1721

R. Singh
K. Kompella
Juniper Networks
S. Palislamovic
Nokia

June 2019

Updated Processing of Control Flags for BGP Virtual Private LAN Service (VPLS)

Abstract

This document updates the meaning of the Control Flags field in the "Layer2 Info Extended Community" used for BGP Virtual Private LAN Service (VPLS) Network Layer Reachability Information (NLRI) as defined in RFC 4761. This document updates RFC 4761.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8614.

Table of Contents

1. Introduction

"Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling" [RFC4761] describes the concepts and signaling for using the Border Gateway Protocol (BGP) to set up a VPLS. It specifies the BGP VPLS Network Layer Reachability Information (NLRI) by which a Provider Edge (PE) router may require other PEs in the same VPLS to include (or not) the Control Word (CW) and sequencing information in VPLS frames sent to this PE.

The use of the CW helps prevent the misordering of IPv4 or IPv6 Pseudowire (PW) traffic over Equal-Cost Multipath (ECMP) paths or Link Aggregation Group (LAG) bundles. [RFC4385] describes the format for the CW that may be used over point-to-point PWs and over a VPLS. Along with [RFC3985], [RFC4385] also describes sequence number usage for VPLS frames.

However, [RFC4761] does not specify the behavior of PEs in a mixed environment where some PEs support CW/sequencing and others do not.

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Problem Description

[RFC4761] specifies the VPLS BGP NLRI by which a given PE advertises the behavior expected by the multiple PEs participating in the same VPLS. The NLRI indicates the VPLS label that the various PE routers, which are referred to in the NLRI, should use when forwarding VPLS traffic to this PE. Additionally, by using the Control Flags, this PE specifies whether the other PEs (in the same VPLS) should use the CW or sequenced delivery for frames forwarded to this PE. These are indicated by the C-bits and the S-bits, respectively, in the Control Flags, as specified in Section 3.2.4 in [RFC4761].

[RFC4761] requires that if the advertising PE sets the C-bits and S-bits, the receiving PE MUST, respectively, insert a CW and include sequence numbers when forwarding VPLS traffic to the advertising PE.

However, in a BGP VPLS deployment, there would often be cases where a PE receiving the VPLS BGP NLRI may not have the ability to insert a CW or include sequencing information inside PW frames. Thus, the behavior of CW processing and sequencing needs to be further specified.

This document updates the meaning of the Control Flags in the Layer2 Info Extended Community in the BGP VPLS NLRI. It also specifies the forwarding behavior for a mixed-mode environment where not every PE in a VPLS has the ability or the configuration to honor the Control Flags received from the PE advertising the BGP NLRI.

3. Updated Meaning of Control Flags in the Layer2 Info Extended Community

[RFC4761] does not allow for the CW setting to be negotiated. In a typical implementation, if a PE sets the C-bit, it expects to receive VPLS frames with a CW and will send frames the same way. If the PEs at the two ends of a PW do not agree on the setting of the C-bit, the PW does not come up. The behavior is similar for the S-bit.

This memo updates the meaning of the C-bit and the S-bit in the Control Flags.

3.1. Control Word (C-Bit)

If a PE sets the C-bit in its NLRI, it means that the PE has the ability to send and receive frames with a CW.

- If the PEs at both ends of a PW set the C-bit, CWs MUST be used in both directions of the PW.

- If both PEs send a C-bit of 0, CWs MUST NOT be used on the PW.

These two cases behave as before.

However, if the PEs at both ends of the PW do not agree on the setting of the C-bit, CWs MUST NOT be used in either direction on that PW, but the PW MUST NOT be prevented from coming up due to this mismatch. So, the PW will still come up but will not use the CW in either direction. This behavior is changed from the behavior described in [RFC4761] where the PW does not come up.

3.2. Sequence Flag (S-Bit)

If a PE sets the S-bit in its NLRI, it means that the PE has the ability to set sequence numbers as described in Section 4.1 in [RFC4385] and process sequence numbers as described in Section 4.2 in [RFC4385].

- If the PEs at both ends of a PW set the S-bit, non-zero sequence numbers MUST be used in both directions of the PW.

- If both PEs send an S-bit of 0, sequence numbers MUST NOT be used on the PW.

These two cases behave as before.

[RFC4761] does not allow for the S-bit setting to be negotiated either. In a typical implementation, if the PE sets the S-bit in the advertised NLRI, it expects to receive VPLS frames with non-zero sequence numbers and will send outgoing frames over the PW with non-zero sequence numbers.

This memo further specifies the expected behavior when the PEs at the ends of the PW advertise differing S-bit values. If the PEs at both ends of the PW do not agree on the setting of the S-bit, then the PW SHOULD NOT come up. This is to avoid running into out-of-sequence ordering scenarios when the multiple PEs that are enabling multihoming for a site have differing S-bit advertisements as described in Section 4.2 in [RFC4385]. However, if a deployment is known to not utilize multihoming, a user-configurable way to override

this recommendation MAY be provided by an implementation whereby the PW is allowed to come up. In that case, the PE advertising the S-bit as 0 should set sequence numbers in the frames as 0, and the PW receiving the frames should not expect to receive non-zero sequence numbers.

4. Using Point-to-Multipoint (P2MP) LSPs as Transport for BGP VPLS

BGP VPLS can be used over point-to-point Label Switched Paths (LSPs) acting as transport between the VPLS PEs. Alternately, BGP VPLS may also be used over Point-to-Multipoint (P2MP) LSPs with the source of the P2MP LSP rooted at the PE advertising the VPLS BGP NLRI.

In a network that uses P2MP LSPs as transport for a VPLS, there may be some PEs that support the CW while others may not. The behavior is similar for the sequencing of VPLS frames.

In such a setup, a source PE that supports CW should set up two different P2MP LSPs such that:

- One P2MP LSP will transport CW-marked frames to those PEs that advertised the C-bit as 1.

- The other P2MP LSP will transport frames without the CW to those PEs that advertised the C-bit as 0.

Using two different P2MP LSPs to deliver frames with and without the CW to different PEs ensures that a P2MP root PE honors the C-bit advertised by the other P2MP PEs.

However, the set of leaves on the two P2MP LSPs (rooted at the given PE) MUST NOT contain any PEs that advertised a value for the S-bit different from what the root PE itself is advertising. PEs that advertised their S-bit values differently (from what the P2MP root PE advertised) will not be on either of the P2MP LSPs. This ensures that the P2MP root PE is sending VPLS frames only to those PEs that agree on the setting of the S-bit.

The ingress router for the P2MP LSP should send separate NLRIs for the cases of using the CW and for not using the CW.

5. Illustrative Diagram

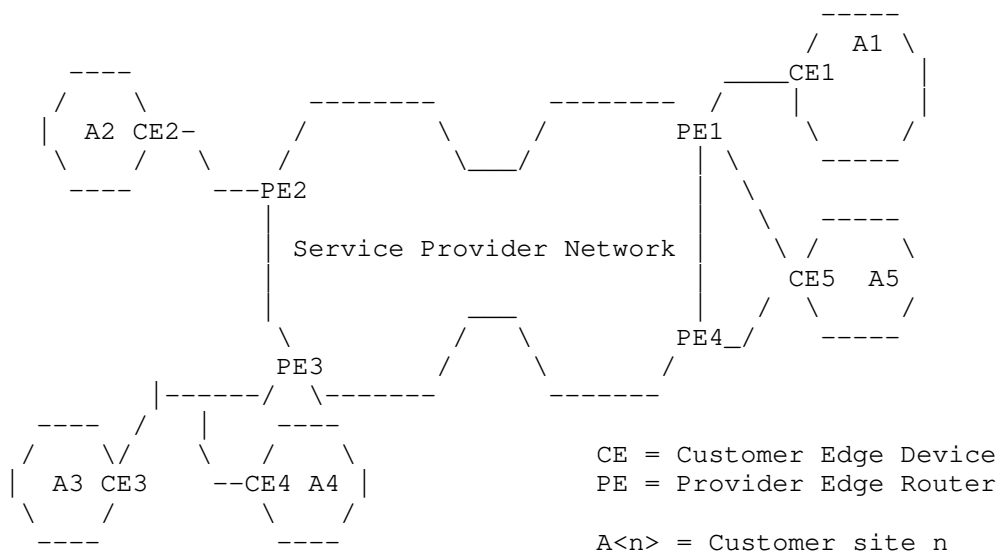

Figure 1: Example of a VPLS

In the above topology, let there be a VPLS configured with the PEs as displayed. Let PE1 be the PE under consideration that is CW enabled and sequencing enabled. Let PE2 and PE3 also be CW enabled and sequencing enabled. Let PE4 not be CW enabled or have the ability to include sequence numbers. PE1 will advertise a VPLS BGP NLRI, containing the C/S-bits marked as 1. PE2 and PE3, on learning of the NLRI from PE1, will include the CW and non-zero sequence numbers in the VPLS frames being forwarded to PE1 as described in Section 4 in [RFC4385]. However, PE4, which does not have the ability to include a CW or include non-zero sequence numbers, will not.

As per [RFC4761], PE1 would expect all other PEs to forward CW-containing frames that have non-zero sequence numbers. That expectation cannot be met by PE4 in this example. Thus, as per [RFC4761], the PW between PE1 and PE4 does not come up.

However, this document addresses how an implementation should support BGP VPLS in a network where a subset of the BGP VPLS PEs support the CW and/or frame sequencing. PE1 will not bring up the PW with PE4 due to the S-bit mismatch, unless overridden by local configuration on PE1 and PE4 as specified in Section 3.2. If PE4 instead was to advertise a C-bit of 0 and an S-bit of 1, then the PW between PE1 and PE4 would come up despite the CW mismatch. Additionally, PE1 would set up its data plane such that it will strip the CW only for those

VPLS frames that are received from PEs that have indicated their desire to receive CW-marked frames. So, PE1 will set up its data plane to strip the CW only for VPLS frames received from PE2 and PE3, and it will expect to process PW frames containing non-zero sequence numbers as described in Section 4.2 in [RFC4385]. PE1 will set up its data plane to not strip the CW from frames received from PE4, and it would expect PE4 to send frames with non-zero sequence numbers. All frames sent by PE4 to PE1 over the PW would have a non-zero sequence number.

6. Treatment of C-Bits and S-Bits in Multihoming Scenarios

6.1. Control Word (C-Bit)

In a multihomed environment, different PEs may effectively represent the same service destination endpoint. It could be assumed that the end-to-end PW establishment process should follow the same rules when it comes to CW requirements, meaning that setting the C-bit would be enforced equally toward both primary and backup designated forwarders.

However, in the multihoming case, each PW SHOULD be evaluated independently. Assuming the network topology specified in Section 5, there could be the case where the PW between PE2 and PE1 could have the CW signaled via the extended community and would be used in the VPLS frame, while the PE2-to-PE4 PW would not insert the CW in the VPLS frame due to a C-bit mismatch. The multihoming behavior of the rest of the PEs should simply follow the rules specified in [VPLS-MULTIHOMING].

6.2. Sequence Flag (S-Bit)

In a multihomed environment, different PEs may effectively represent the same service destination endpoint. In this case, the rules for end-to-end PW establishment SHOULD follow the same behavior as that described in Section 3.2 when it comes to S-bit requirements. Consider the case described in Section 5 with CE5 having a connection to multiple PEs (multihomed) to PE4 and PE1. The PW's behavior is similar to that for the CW scenario such that the S-bit evaluation SHOULD be independent per PW. So, in the case where PE4 does not set the S-bit in its advertised NLRI, there is an S-bit mismatch between PE1 and PE4. This mismatch prevents the PW establishment between PE1 and PE4. So, only one PW -- between PE1 and PE2 -- would be established for the multihomed site shown. Thus, even though CE5 is physically multihomed, due to PE4's lack of support for sending frames with non-zero sequence numbers, there would be no PW between PE2 and PE4. CE5 would effectively not be multihomed.

7. Security Considerations

This document updates the behavior specified in [RFC4761]. The security considerations discussed in [RFC4761] apply. This document essentially addresses BGP VPLS behavior for PEs when the C-bit value, the S-bit value, or both values advertised by a given PE are different from what another PE in the VPLS is advertising. Any bit-flipping media errors leading to causing this mismatch of C/S-bits between PEs do not adversely affect the availability of the PWs. Rather, they cause CWs to not be used or cause the NLRI-advertising PE to not expect non-zero sequenced frames, for the C-bit and the S-bit, respectively, being mismatched across PEs. This is no worse than the previous behavior where any bit-flipping media errors leading to a mismatch of the C/S-bits between PEs would cause the PW to not come up.

8. IANA Considerations

This document has no IANA actions.

9. References

9.1. 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>.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, DOI 10.17487/RFC4385, February 2006, <https://www.rfc-editor.org/info/rfc4385>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

9.2. Informative References

[RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, DOI 10.17487/RFC3985, March 2005, <https://www.rfc-editor.org/info/rfc3985>.

[VPLS-MULTIHOMING] Kothari, B., Kompella, K., Henderickx, W., Balus, F., and J. Uttaro, "BGP based Multi-homing in Virtual Private LAN Service", Work in Progress, draft-ietf-bess-vpls-multihoming-03, March 2019.

Authors' Addresses

Ravi Singh
Juniper Networks
1133 Innovation Way
Sunnyvale, CA 94089
United States of America

Email: ravis@juniper.net

Kireeti Kompella
Juniper Networks
1133 Innovation Way
Sunnyvale, CA 94089
United States of America

Email: kireeti@juniper.net

Senad Palislamovic
Nokia
600 Mountain Avenue
Murray Hill, NJ 07974-0636
United States of America

Email: Senad.palislamovic@nokia.com