Updated processing of Control Flags for BGP VPLS
draft-ietf-bess-bgp-vpls-control-flags-08

Abstract

This document updates the meaning of the Control Flags field in the Layer2 Info Extended Community used for BGP-VPLS NLRI as defined in RFC4761. This document updates RFC4761.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

Copyright and License 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
Table of Contents

1 Introduction ........................................... 3
  1.1 Terminology ........................................ 3
2 Problem ............................................... 3
3 Updated meaning of Control Flags in the Layer2 Info Extended Community ...................................... 4
  3.1 Control word (C-bit) ................................ 4
  3.2 Sequence flag (S-bit) ................................ 4
4 Using Point-to-MultiPoint (P2MP) LSPs as transport for BGP VPLS ...................................... 5
5 Illustrative diagram .................................... 6
6 Treatment of C and S bits in multi-homing scenarios ......................................................... 7
  6.1 Control word (C-bit) ................................ 7
  6.2 Sequence flag (S-bit) ................................. 7
7 Security Considerations ................................ 8
8 IANA Considerations ................................... 8
9 References ............................................. 8
  9.1 Normative References ............................... 8
  9.2 Informative References ............................. 8
Authors’ Addresses ..................................... 9
1 Introduction

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

The use of the Control Word (CW) helps prevent mis-ordering of IPv4 or IPv6 Pseudo-Wire (PW) traffic over Equal Cost Multi-Path (ECMP) paths or Link Aggregation Group (LAG) bundles. [RFC4385] describes the format for CW that may be used over Point-to-Point PWs and over a VPLS. Along with [RFC3985], the document 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 Control Word/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 BCP14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2 Problem

[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 Control Word or sequenced-delivery for frames forwarded to this PE. These are respectively indicated by the C and the S bits in the Control Flags as specified in section 3.2.4 in [RFC4761].

[RFC4761] requires that if the advertising PE sets the C and S bits, the receiving PE MUST, respectively, insert control word (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 processing CW and sequencing needs to be further specified.

This document updates the meaning of the Control Flags in layer2 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

The current specification 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 control word, 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 ability to send and receive frames with a control word. If the PEs at both ends of a PW set the C-bit, control words MUST be used in both directions of the PW. If both PEs send a C-bit of 0, Control Words 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, control words 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 not use control word 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 ability to set sequence numbers as listed in section 4.1 in [RFC4385] and process sequence numbers as listed 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 a S-bit of 0, sequence numbers MUST NOT be used on the PW. These two cases behave as before.
Current BGP VPLS specification does not allow for 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 multi-homing for a site have differing S-bit advertisements as listed in section 4.2 in [RFC4385]. However, if a deployment is known to not utilize multi-homing, 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 S-bit as 0 should set sequence numbers in the frames as zero and the PW receiving the frames should not have an expectation 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-2-point LSPs acting as transport between the VPLS PEs. Alternately, BGP VPLS may also be used over P2MP Label Switched Path (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 CW while others may not. Similarly, for the sequencing of VPLS frames.

In such a setup, a source PE that supports CW should setup 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 CW to those PEs that advertised 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 value 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 S-bit.

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

5 Illustrative diagram

-----
| \ / |
----|
| CE1 |
| CE2 |
| PE1 |
| PE2 |
| PE3 |
| PE4 |
| CE5 |
| A5 |
| A3 |
| CE3 |
| CE4 |
| A4 |

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 NLRI from PE1, will include the CW and non-zero sequence numbers in the VPLS frames being forwarded to PE1 as listed in section 4 in [RFC4385]. However, PE4 which does not have the ability to include CW or include non-zero sequence numbers, will not.

As per [RFC4761], PE1 would have an expectation that all other PEs forward CW-containing frames which 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 to support the mixed-CW and
mixed sequencing-ability of PEs described above. 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 despite the CW mismatch the PW between PE1 and PE4 would come up. Additionally PE1 would setup 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 setup 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 listed in section 4.2 in [RFC4385]. PE1 will setup 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 and S bits in multi-homing scenarios

6.1 Control word (C-bit)

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

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

6.2 Sequence flag (S-bit)

In a multi-homed environment, different PEs may effectively represent the same service destination end-point. In this case, the rules for end-to-end PW establishment SHOULD follow the same behavior as listed in section 3.2 when it comes to sequence bit requirements. Consider the case described in section 5 with CE5 being multi-homed to PE4 and PE1. The PW behavior is similar to the CW scenario so that the insertion of S-bit evaluation SHOULD be independent per PW. However, because S-bit mismatch between two end-point PEs results in no PW establishment, in the case where PE4 doesn’t support S-bit. So, only one PW would be established, between PE1 and PE2. Thus, even though
7 Security Considerations

This document updates the behavior specified in [RFC4761]. The security considerations listed in [RFC4761] apply. This document essentially addresses BGP-VPLS behavior for PEs when the C-bit and/or S-bit value 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 control-words to not be used or cause the NRLI-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 mismatch of C/S bit between PEs would cause the PW to not come up.

8 IANA Considerations

This document does not make any requests from IANA.

9 References

9.1 Normative References


9.2 Informative References

Authors' Addresses

Ravi Singh
Juniper Networks
1133 Innovation Way
Sunnyvale, CA  94089
US
EMail: ravis@juniper.net

Kireeti Kompella
Juniper Networks
1133 Innovation Way
Sunnyvale, CA  94089
US
EMail: kireeti@juniper.net

Senad Palislamovic
Nokia
600 Mountain Avenue
Murray Hill, NJ 07974-0636
US
EMail: Senad.palislamovic@nokia.com