BESS Working Group
INTERNET-DRAFT
Intended Status: Proposed Standard
Updates: 4761 (if approved)

Expires: February 18, 2019

R. Singh
K. Kompella
Juniper Networks
S. Palislamovic
Alcatel-Lucent
August 17, 2018

Updated processing of control flags for BGP VPLS
draft-ietf-bess-bgp-vpls-control-flags-06

Abstract

This document updates the meaning of the "control flags" fields inside the "layer2 info extended community" used for BGP-VPLS NLRI as defined in RFC4761. If approved, this document updates RFC4761.

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 BGP (Border Gateway Protocol) to setup a VPLS (virtual private LAN service). It specifies the BGP VPLS NLRI (network layer reachability information) by which a PE may require other PEs in the same VPLS to include (or not) control-word and sequencing information in VPLS frames sent to this PE.

The use of control word (CW) helps prevent mis-ordering of IPv4 or IPv6 PW traffic over ECMP (equal cost multi-path) paths or LAG (link aggregation group) bundles. [RFC4385] describes the format for control-word that may be used over point-2-point PWs (pseudowires) and over a VPLS. It along with [RFC3985] also describes sequencing of 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", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2 Problem

[RFC4761] specifies the VPLS BGP NLRI by which a given PE advertises the behavior expected from 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, when forwarding VPLS traffic to the PE, the receiving PE MUST insert control word (CW) and by including sequence numbers respectively.

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

Current specification does not allow for the CW setting to be negotiated. Rather, 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 both ends of a pseudowire do not agree on the setting of the C-bit, the PW does not come up. The expected 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 don't agree on the setting of the C-bit, control words MUST NOT be used on that PW but the PW MUST NOT be prevented from coming up due to this mismatch. So, the PW MUST still come up. This behavior is new; the old behavior was that the PW doesn't come up.

3.2 Sequence flag (S-bit)

Current BGP VPLS specification do not allow for S-bit setting to be negotiated either. If the PE sets the S-bit, it expects to receive VPLS frames with sequence numbers, and will send the frames with sequence numbers as well. This memo further specifies the existing behavior. If the PEs on the both ends of the PW set the S-bit, then both PEs MUST include the PW sequence numbers. If the PEs at both ends of the PW do not agree on the setting of the S-bit, the PW SHOULD NOT come up at all.

4 Using p2mp LSP 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 (point to multipoint) LSPs (label switched path) 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 BGP VPLS, in a given VPLS there may be some PEs that support control-word while others do not. Similarly, for sequencing of frames.

In such a setup, a source PE that supports control-word should setup 2 different p2mp LSPs such that:

- one p2mp LSP will carry CW-marked frames to those PEs that advertised C-bit as 1, and
- the other p2mp LSP will carry frames without CW to those PEs that advertised C-bit as 0.

Using 2 different p2mp LSPs to deliver frames with and without CW to different PEs ensures that this PE honors the C-bit advertised by the other PEs.

However, the set of leaves on the 2 p2mp LSPs (rooted at the given PE) MUST NOT contain any PEs that advertised a value for S-bit different from what this PE itself is advertising. PEs that advertised their S-bit value differently (from what this PE advertised) will not be on either of the p2mp LSPs. It is ensured that this PE is sending VPLS frames only to those PEs that agree with this PE on the setting of S-bit.

5 Treatment of C and S bits in multi-homing scenarios

5.1 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 forwarder together.

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

bess-vpls-multihoming-00.

5.2 Sequence flag (S-bit)

In 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 rules when it comes to sequence bit requirements. Consider the case below with CE5 being multi-homed to PE4 and PE1. The PW behavior is similar to the C-word scenario so that the insertion of S-bit evaluation SHOULD be independent per PW. However, because S-bit mismatch between two end-point PEs yields in no PW establishment, in the case where PE4 doesn't support S-bit, only one PW would be established, between PE1 and PE2. Thus, even though CE5 is physically multi-homed, due to PE4's lack of support for S-bit, and no PW between PE1 and PE4, CE5 would not be multi-homed any more.

6 Illustrative diagram

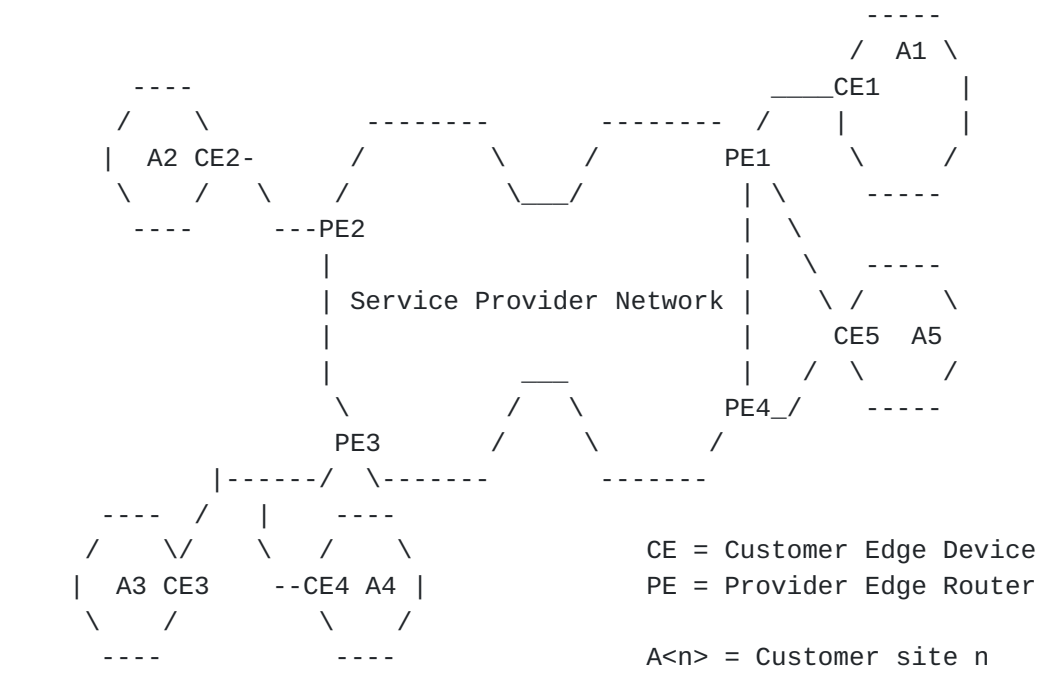

Figure 1: Example of a VPLS

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. Let PE2 and PE3 also be CW enabled. Let PE4 not be CW enabled. PE1 will advertise a VPLS BGP NLRI, containing the C/S bits marked as 1. PE2 and PE3 on learning of NLRI from PE1, shall include the control

word in VPLS frames being forwarded to PE1. However, PE4 which does not have the ability to include control-word.

As per [RFC4761], PE1 would have an expectation that all other PEs forward traffic to it by including CW. 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 environment as above. PE1 will bring up the PW with PE4 despite the CW mismatch. Additionally, it will setup its data-plane such that it will strip the control-word only for those VPLS frames that are received from PEs that are themselves indicating their desire to receive CW marked frames. So, PE1 will setup its data plane to strip-off the CW only for VPLs frames received from PEs PE2 and PE3. PE1 will setup its data plane to not strip CW from frames received from PE4.

7 Security Considerations

This document updates the behavior specified in [RFC4761]. The security considerations listed in [RFC4761] apply. However, there are no new security considerations due to the text of this document.

8 IANA Considerations

This document does not make any requests from IANA.

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.

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

[RFC4385] Bryant, S., Swallow G., Martini L., D. McPherson, Pseudowire Emulation Edge-to-Edge (PWE3) Control Word, RFC 4385, February 2006.

9.2 Informative References

[RFC3985] Bryant, S., P. Pate, Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture, RFC3985, March 2005.

Authors' Addresses

Ravi Singh
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US
EMail: ravis@juniper.net

Kireeti Kompella
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US
EMail: kireeti@juniper.net

Senad Palislamovic
Alcatel-Lucent
EMail: senad.palislamovic@alcatel-lucent.com