Network Working Group
Internet-Draft
Updates: 4761 (if approved)
Intended status: Standards Track
Expires: April 30, 2009

B. Kothari
R. Fernando
Juniper Networks

October 27, 2008

VPLS Flush in BGP-based Virtual Private LAN Service
draft-kothari-l2vpn-vpls-flush-00.txt

Status of this Memo

Abstract

This document defines procedures that allow BGP based Virtual Private LAN Service (VPLS) provider edge (PE) devices to send explicit flush notifications to remote VPLS PEs.

Table of Contents

1. Introduction

[RFC4761] describes mechanisms that allow VPLS PEs to use BGP to automatically discover PE membership in VPLS domains and to signal pseudowires required to carry VPLS traffic. Each VPLS PE maintains state for MAC addresses that it learns from locally attached customer sites. In addition, each VPLS PE also maintains state for MAC addresses that belong to remote customer sites that are attached to remote PEs. MAC addresses of remote customer sites are learned over the pseudowires that are established among all the VPLS PEs. In case of a topology change that teardown pseudowires, VPLS PEs delete MAC addresses that were learned on those pseudowires. However, there are cases when a topology change, such as a failure between a customer site and a PE, does not teardown a pseudowire. In such cases where only local VPLS PE is aware of the topology change, an explicit notification for flushing MAC addresses on remote VPLS PEs is required. In absence of explicit MAC flush notification, stale MAC state might be deleted when MAC age out timer expires, which is typically in the order of minutes. Flushing of MAC addresses increases connectivity restoration time after a failure, and thus, a mechanism to expedite flushing of MAC addresses is highly desirable.

This document describes a new BGP Capability for flush mechanisms in BGP based VPLS. A new BGP message, VPLS-FLUSH, is introduced to carry a list of TLVs that will be used to flush the MAC addresses associated with those TLVs.

BGP is used as the control plane protocol to carry the VPLS-FLUSH message for following reasons:

1. Reuse: Since BGP is already used as the control plane protocol for VPLS service, use of BGP to carry VPLS-FLUSH message eliminates need for service providers to deploy a new protocol for MAC flush notification.

2. Efficient flooding: Since a VPLS PE that triggers the MAC flush operations needs to notify all other VPLS PEs participating in the same VPLS, it needs to efficiently flood the message to only the PEs that are intended recipient of VPLS-FLUSH message. The VPLS-FLUSH message will be propagated to only those routers that would have received the VPLS NLRIs for the same RT that is carried in the VPLS-FLUSH message as well, both in intra-AS and inter-AS deployments. BGP signalled VPLS networks restrict the flow of routing messages to only the interested routers and ASes today by use of Route Target extended communities [RFC4360] and RT constrains [RFC4684].

1.1. General Terminology

VPLS domain: A VPLS domain represents a bridging domain per customer. A Route Target community as described in [RFC4360] is used to identify all the PE routers participating in a particular VPLS domain.

Source PE: A VPLS PE that originates either the VPLS NLRI or VPLS-FLUSH message. The source PE address is carried in Route Origin Extended Community [RFC4360] and use of this community for VPLS advertisements is described in [I-D.kompella-l2vpn-vpls-multihoming].

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. VPLS Flush Capability

To advertise the VPLS Flush Capability to a peer, a BGP speaker uses BGP Capabilities Advertisement [RFC3392]. The Capability code is to be assigned by IANA with a Capability length 0.

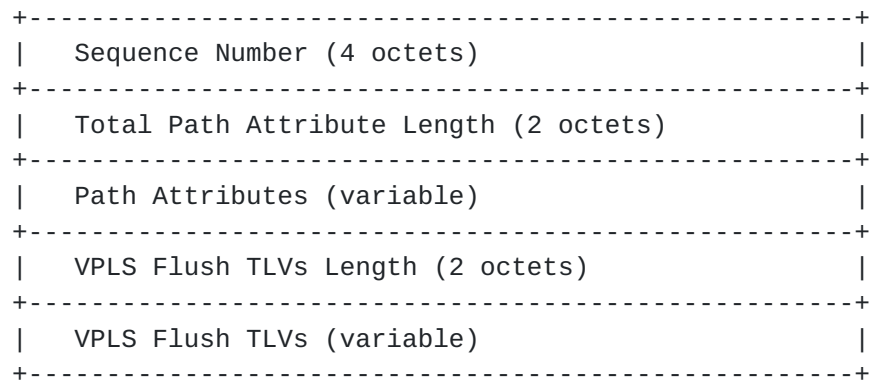
By advertising the VPLS Flush Capability to a peer, a BGP speaker conveys to the peer that the speaker is capable of receiving and properly handling the VPLS-FLUSH message, described in Section 4, from the peer.

A BGP speaker should only send a VPLS Flush Capability to a peer if and only if BGP VPLS address family (Section 3.2.2 [RFC4761]) is also enabled and negotiated with the peer.

3. VPLS-FLUSH Message

The VPLS-FLUSH is a new BGP message that always includes the fixed size BGP header and MUST include the fields shown below. The type is to be assigned by IANA.

Message Format:



Sequence Number:

This 4-octets unsigned integer indicates the current sequence number of the flush message being sent to the remote VPLS PEs.

Total Attribute Length:

This 2-octets unsigned integer indicates the total length of the Path Attributes field in octets. Its value MUST be greater than 0, which implies that at least one attribute must be present.

Attributes:

A variable length sequence of path attributes is present in every VPLS-FLUSH message. The following attributes MUST be present:

- Route Target Community
- AS-PATH Attribute
- ORIGINATOR_ID Attribute

- o CLUSTER_LIST Attribute

- o Route Origin Extended Community

AS-PATH, ORIGINATOR_ID and CLUSTER_LIST attributes are processed and updated exactly like they are in routing messages and are present to make sure VPLS-FLUSH messages never end up in a loop. Note that use of Route Origin Extended Community for VPLS advertisements is described in [I-D.kompella-l2vpn-vpls-multihoming].

VPLS Flush TLVs Length:

This 2-octets unsigned integer indicates the total length of the VPLS Flush TLVs field in octets. Its value MUST be greater than 0.

VPLS Flush TLVs:

This is a variable length field that contains a list of TLVs that indicates what MAC addresses are to be flushed based on the value contained in each TLV. Each TLV is a triple <type, length, value> of variable length. The type is a 2-octet field that identifies one of the possible TLVs defined. Length is a 2-octet field that indicates the TLV value length. Value is of variable length and is encoded according to the TLV type.

If a VPLS PE receives a VPLS-FLUSH message that contains a TLV type that it does not understand, it SHOULD ignore that TLV alone.

The Type is a 2-octet field, with possible values as follows:

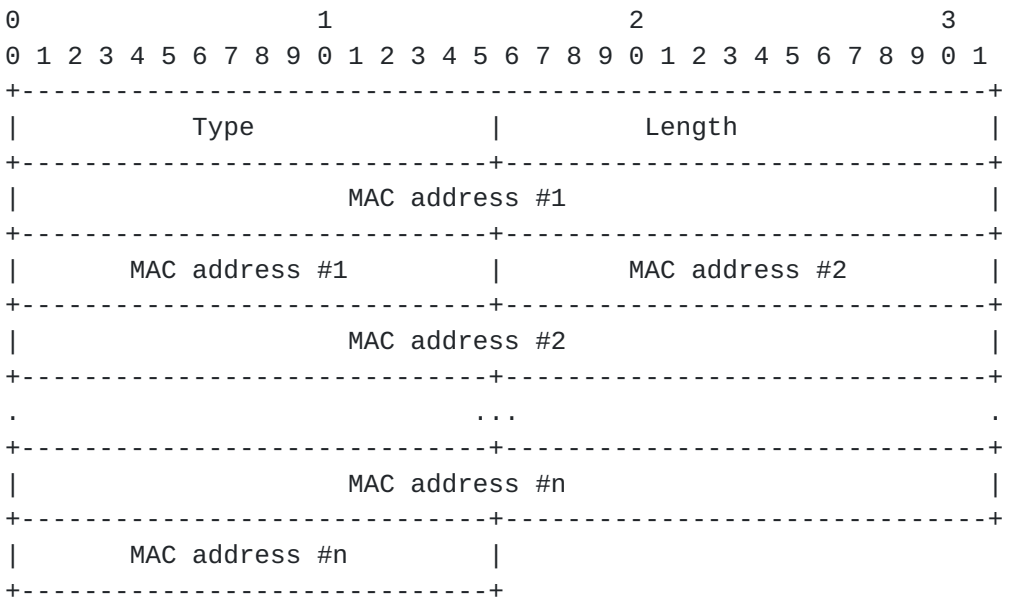
Value	Meaning
0	MAC list TLV

3.1. MAC List TLV

VPLS FLush TLV type 0 indicates that the TLV contains a list of 48 bits MAC addresses that should be flushed by the PE processing the VPLS-FLUSH message.

The length field specifies the total length in octets of the MAC addresses present in the TLV. If the length is 0, which indicates that no MAC addresses are present, then all MAC addresses learned from the source PE (indicated by Route Origin Extended Community) should be flushed. The length MUST be a multiple of 6.

The encoding for MAC list is as follows:



A VPLS PE that receives a VPLS-FLUSH message with a MAC list TLV should delete each MAC address listed in the TLV that it learned from the source VPLS PE for the VPLS domain specified by the Route Target.

4. Operation

A speaker that is willing to receive the VPLS-FLUSH message from its peer should advertise the VPLS-FLUSH capability to its peer.

A speaker may send the VPLS-FLUSH message to its peer only if it has received the VPLS-FLUSH capability from its peer.

A VPLS-FLUSH message originated for a particular VPLS domain should carry the same Route Target (RT) that is used to identify that VPLS domain. The Route Target Extended Communities serve the dual purpose of identifying the member PEs of a VPLS domain as well as limiting the flooding of the VPLS-FLUSH message to be bounded by the member PEs. A router that receives a VPLS-FLUSH message without any RTs MUST neither process it nor propagate it.

A RR or ASBR should not do BGP path selection for VPLS-FLUSH messages. A RR or ASBR MUST process the attributes contained in the VPLS-FLUSH message for loop detection and for RT constrains before propagating the message to other BGP peers, but it should hold no permanent state for a VPLS-FLUSH message.

A PE should not do BGP or VPLS path selection for VPLS-FLUSH messages. A PE should only process VPLS Flush TLVs for the messages that have Route Target that matches one of the VPLS instance configured on the PE router. A PE might receive the same VPLS-FLUSH message from a source PE more than once due to presence of RRs or ASBRs. A PE can use the sequence number field to detect duplicate VPLS-FLUSH messages. It is RECOMMENDED that a PE ignore duplicate VPLS-FLUSH messages. How a PE ignore duplicate VPLS-FLUSH messages is outside the scope of this document. Other than state to detect duplicate flush messages, a PE should hold no other permanent state.

A VPLS-FLUSH message might be lost if there are multiple failures. In such cases, the remote PEs for which the flush message was targeted for will continue to hold stale information unless they age it out or relearn the MAC addresses from a different source PE. If a VPLS-FLUSH message is lost due to a topology change that also teardown the PWs, then the affected PEs SHOULD flush MAC addresses learned over those PWs.

5. Security Considerations

TBD

6. IANA Considerations

TBD

7. Acknowledgments

The authors would like to thank Yakov Rekhter, Nischal Sheth and John Scudder for their comments and suggestions.

8. References

8.1. Normative References

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

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

[I-D.kompella-l2vpn-vpls-multihoming] Kompella, K., Kothari, B., and T. IV, "Multi-homing in BGP-based Virtual Private LAN Service", draft-kompella-l2vpn-vpls-multihoming-01 (work in progress), July 2008.

8.2. Informative References

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[RFC3392] Chandra, R. and J. Scudder, "Capabilities Advertisement with BGP-4", RFC 3392, November 2002.

[RFC5291] Chen, E. and Y. Rekhter, "Outbound Route Filtering Capability for BGP-4", RFC 5291, August 2008.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, November 2006.

[802.1ah] "IEEE Draft P802.1ah/D4.2 Virtual Bridged Local Area Networks, Amendment 6: Provider Backbone Bridges,", March 2008.

Authors' Addresses

Bhupesh Kothari
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: bhupesh@juniper.net

Rex Fernando
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: rex@juniper.net

Full Copyright Statement

Intellectual Property