Abstract

This document specifies protocols and procedures of using BGP as PE-CE control protocol for carrying customer MAC routing information.

Requirements Language

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].

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 http://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 January 5, 2015.

Copyright Notice

Copyright (c) 2014 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 (http://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
1. Introduction

[I-D.ietf-l2vpn-evpn] describes protocols and procedures for BGP MPLS based Ethernet VPNs. BGP is used for MAC learning by exchanging customer MAC routing information between PEs in the control plane instead of MAC learning between PEs in the data plane. It also states that MAC learning between PEs and CEs MAY be done in the control plane, but it does not define the detailed protocols and procedures. This document specifies protocols and procedures of using BGP as PE-CE control protocol for carrying customer MAC routing information. This can provide some benefits such as fast convergence in some situation.

2. Terminology

This document uses terminology described in [I-D.ietf-l2vpn-evpn].

3. Application Scenarios

There are some benefits when control plane is introduced between PE and CE in EVPN network. The following illustrates the benefits with an example of fast convergence in the event of PE to CE network failure.

[I-D.ietf-l2vpn-evpn] defines a mechanism to efficiently and quickly signal, to remote PE nodes, the need to update their forwarding tables upon the occurrence of a failure in connectivity to an Ethernet Segment. This mechanism optimizes the withdrawal of MAC Advertisement routes, and then optimizes the network convergence time.
in the event of PE to CE failures. But it still cannot fully provide convergence time that is independent of the number of MAC addresses learned by the PE. There exist a situation where the network convergence time is dependent on the local MAC learning of PE and the advertisement of them to remote PE.

![Figure 1 Multi-homed EVPN Network](image)

To illustrate this with an example in the Figure 1, consider two PEs (PE1 and PE2) connected to a multi-homed Ethernet Segment ES1. All-Active redundancy mode is assumed. A given MAC address M1 is learned by PE1 but not PE2. On PE3, the following states may arise:

T0- PE3 receives the Ethernet A-D routes per ESI from PE1 and PE2.

T1- When the MAC Advertisement route from PE1 and the Ethernet A-D routes per EVI from PE1 and PE2 are received, PE3 can forward traffic destined to M1 to both PE1 and PE2.

T2- After T1, when the ES1 connected to PE1 fails, PE1 MUST withdraw its Ethernet A-D route per ESI, then PE3 forwards traffic destined to M1 to PE2 only.

T3- After T2, PE1 MUST also withdraw the MAC Advertisement routes (M1) that are impacted by the failure. Before PE2 learns M1 and advertises a MAC Advertisement route for M1, PE3 will treat traffic to M1 as unknown unicast. If the behavior is to drop the unknown unicast based on the administrative policy, the traffic to M1 on PE3 will be interrupted. Until PE2 has also advertised a MAC Advertisement route for M1 before PE1 withdraws its MAC route, then PE3 would have continued forwarding traffic destined to M1.

In the above example, once the local MAC learning of PE was done via control plane, both PE1 and PE2 will advertise a MAC Advertisement route for M1, then PE3 could continue forwarding traffic destined to M1 in the event of ES1 connected to PE1 or PE2 fails. In this case,
the network convergence time is not dependent of the local MAC learning and advertisement of MAC addresses learned by the PE any more.

The benefit can also be achieved in case of single-active redundancy mode.

4. BGP EVPN NLRI Extensions

A new route type is defined for EVPN NLRI to advertise customer MAC route between PE and CE in EVPN:

+ 6 - Customer MAC Advertisement route

A customer MAC Advertisement route type specific EVPN NLRI consists of the following:

+-----------------------------------------+
| Ethernet Segment Identifier (10 octets) |
+-----------------------------------------+
| Ethernet Tag ID (4 octets)              |
+-----------------------------------------+
| MAC Address Length (1 octet)           |
+-----------------------------------------+
| MAC Address (6 octets)                 |
+-----------------------------------------+
| IP Address Length (1 octet)            |
+-----------------------------------------+
| IP Address (4 or 16 octets)            |
+-----------------------------------------+

It should be noted that the Route Distinguisher (RD) is not used since the customer MAC routes are always exchanged in the context of unawareness of Ethernet VPN.

Another solution option is to reuse EVPN MAC Advertisement Route defined in [I-D.ietf-l2vpn-evpn] to exchange MAC route information between CE and PE. In this case RD, MPLS Label1 and MPLS Label2 fields SHOULD be set as 0. In addition, the RT for the route SHOULD also be set as 0.

5. Exchanging C-MAC Routes

This section describes the procedures of exchanging customer MAC routes between PE and CE. This document assumes that a CE and a PE exchange MAC routes over a direct BGP session.
5.1. Originating MAC Route at the CE router

When a CE receives packets in a given VLAN from interfaces, other than interfaces connected to the PE, it learns MAC addresses in the data plane. If the given VLAN is in the setting of VLANs across the Ethernet links attached to a given PE, the CE MAY advertises the MAC addresses it learns in the data plane to the given PE, using MP-BGP and the specific MAC Route, in the control plane. The MAC Route is constructed as follows:

+ The field of the Ethernet Segment Identifier is reserved for future use.

+ The Ethernet Tag ID is set to the VLAN ID from which the MAC addresses are learned.

+ The MAC address length field is in bits and it is typically set to 48. However this specification enables specifying the MAC address as a prefix; in which case, the MAC address length field is set to the length of the prefix. This provides the ability to aggregate MAC addresses if the deployment environment supports that.

+ The MAC address is set to the value of MAC address the CE learned. The encoding of a MAC address MUST be the 6-octet MAC address specified by [802.1D-ORIG] [802.1D-REV]. If the MAC address is advertised as a prefix then the trailing bits of the prefix MUST be set to 0 to ensure that the entire prefix is encoded as 6 octets.

+ The IP Address field is optional. By default, the IP Address Length field is set to 0 and the IP Address field is omitted from the route. When a valid IP address or address prefix needs to be advertised (e.g., for ARP suppression purposes or for inter-subnet switching), it is then encoded in this route. In this case, the IP Address Length field is in bits and it is the length of the IP prefix. This provides the ability to advertise IP address prefixes when the deployment environment supports that.

+ The encoding of an IP Address MUST be either 4 octets for IPv4 or 16 octets for IPv6. When the IP Address is advertised as a prefix, then the trailing bits of the prefix MUST be set to 0 to ensure that the entire prefix is encoded as either 4 or 16 octets. The length field of Ethernet NLRI is sufficient to determine whether an IP address/prefix is encoded in this route and if so, whether the encoded IP address/prefix is IPv4 or IPv6.

+ The Next Hop field of the MP_REACH_NLRI attribute of the route MUST be set to the IPv4 or IPv6 address of the advertising CE.
It should be noted that the BGP advertisement for the MAC route does not need to carry the Route Target (RT) attributes because of its unawareness of Ethernet VPN.

5.2. Receiving a MAC Route by the PE router

When a PE receives a MAC route from a CE, it learns the MAC addresses advertised in the MAC route in the control plane and associates the MAC addresses with the Ethernet Segment from which it can reach to the advertising CE and the VLAN carried in the MAC route.

The PE SHOULD install forwarding state for the associated MAC addresses based on the Ethernet Segment and VLAN inferred from the MAC route.

In addition, the PE SHOULD advertise the MAC addresses it learns from CE in the control plane, to all the other PEs in the associated EVPN instance, using MP-BGP and the MAC Advertisement route defined in [I-D.ietf-l2vpn-evpn]. For example, the PE learns a MAC address M1 on a multi-homed Ethernet Segment (ES1) and on a VLAN 10, and the VLAN 10 is bundled to EVPN A. The PE SHOULD advertise the MAC address M1 to all the other PEs in EVPN A.

The construction of the MAC Advertisement route and procedures of handling the MAC Advertisement route on receiving it are specified in [I-D.ietf-l2vpn-evpn].

6. IANA Considerations

This document requires IANA to assign a new route type value for EVPN NLRI.

7. Security Considerations

There are no additional security aspects beyond those of EVPN ([I-D.ietf-l2vpn-evpn]).

8. Normative References

[I-D.ietf-l2vpn-evpn]

Authors’ Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: lizhenbin@huawei.com

Junlin Zhang
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: jackey.zhang@huawei.com

Shunwan Zhuang
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: zhuangshunwan@huawei.com