Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 30, 2020

G. Liu
H. Wang
Huawei Technologies
October 28, 2019

EVPN ALL-ACTIVE ANYCAST DETECTION
draft-glendon-bess-evpn-all-active-detection-00

Abstract

A principal feature of EVPN is the ability to support universal detection including ping/trace, twamp, 2544, 1564 and so on. This draft specifies a mechanism of valid universal detection in all-active anycast scenes based on connectivity negotiations to avoid detection interruption due to the inconsistency between the request packet path and the response packet path.

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 [RFC2119].

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

A principal feature of EVPN is the ability to support universal detection including ping/trace, twamp, 2544, 1564 and so on. Universal detection may occur in single-active scenes or in all-active scenes. The all-active scene is proposed in EVPN [RFC7432] . The draft draft-eastlake-bess-enhance-evpn-all-active specifies an improvement to load balancing all-active links. But these two documents do not mention the detection method of the all-active anycast scenes. This draft proposes a mechanism of valid universal detection in all-active anycast scenes based on connectivity negotiations to avoid detection interruption due to the inconsistency between the request packet path and the response packet path.

1.1. Situation Anyalisis

Based on [RFC7432] and the draft draft-eastlake-bess-enhance-evpn-all-active, after the request packet is sent from one of the local PEs that are multi-homed by the CE, the remote PE converts the request packet into a response packet. In the case of the anycast route is reachable in anycast VXLAN tunnel or anycast SRV6 tunnel, the response message is sent to any PE that is multi-homed by the CE. If the PE that receives the response packet overlaps with the PE that sends the request packet, the detection is completed normally and the result is valid. If not, the detection process will be interrupted and the result is invalid.

1.2. Alternative Solutions

A possible solution is to use the detection packet as a special data packet to unconditionally copy the packet to all reachable paths. The response packet finally arrives at the initiator of the detection packet, and the detection will still succeed, but the drawbacks of this implementation are also obvious:

1) A large number of redundant protocol packets may occur in a short period of time on the EVPN network. If a packet attack is detected, the effective bandwidth may be heavily occupied.

2) The response packet requires unconditional replication, which may result in a loop between the multi-homed access PE and the remote PE, resulting in continuous degradation of network quality.

1.3. Design Requirement

The connectivity negotiation occurs between the dual-homed PEs. After the negotiation is successful, the detection packet that is extended and carries the multi-homing source address information specifies the endpoint of the detection response packet. The response packet is accurately passed back to the detection initiation PE node.

1.4. 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.

"CE": Customer edge device. It is used to connect a user and a PE device.

"PE": Provider edge device. It is a unique access point for users to access the carrier network.

"the local PE": The detection packet initiator and endpoint. The local PEs may be single-homed or multi-homed by the CE.

"the remote PE": The detection packet reflector used to converting the request packet to the response packet.

2. Solution Overview

The draft includes the following technical points:

1) Detection packets sending and receiving are not the default behavior of the device. It is needed to manually configure the connectivity negotiation enable switch for the route sender and the route receiver. The EVPN neighbours use the inclusive route or the prefix route to implement connectivity negotiation.

2) Detection request packets needs to be extended to include the bypass source ip address based on the bypass channel between one multi-homed PE and the other multi-homed PE, then any multi-homed PE will know the endpoint of the detection response packets.

3) The application scenarios include EVPN over vxlan and EVPN over SRV6. VXLAN includes VXLAN4 and VXLAN6.

3. Conectivity negotiation

Although Connectivity negotiation belongs to the device-level global capability, considering the simplicity of the protocol extension, the connectivity negotiation enable switch may be configured based on EVPN instances. To reduce the number of connectivity negotiation packets, the Layer 3 VXLAN or SRV6 scenario connectivity negotiation function is implemented based on the EVPN prefix route by adding all-active extended community attribute with the IP address equal to 0, the Layer 2 VXLAN or SRV6 scenario connectivity negotiation function is implemented based on the EVPN inclusive route by adding all-active extended community attribute. The C bit of the first "Reserved" field is in correspondence with the negotiation switch. The all-active extended community defined here is defined as follows:

```
+--------------------------------------------+
|  Type (0x06) / Sub-type  (2 octets)        |
+--------------------------------------------+
|  Reserved (2 octets)                       |
+--------------------------------------------+
|  Reserved (2 octets)                       |
+--------------------------------------------+
|  Reserved (2 octets)                       |
+--------------------------------------------+
```

Figure 1: All-Active Extended Community

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|C|             |
+-+-+-+-+-+-+-+-+
```

Figure 2. The C bit of the first "Reserved" field

The first bit in the "Reserved" field is marked as C bit and is used to indicate whether the detection negotiation is enabled on the route sender.

If C bit is set to 1, this flag indicates the connectivity negotiation is enabled by the advertising PE.

If the connectivity negotiation is not enabled, then the C bit must be set to 0.

For the receiving all-active PE, it is necessary to determine whether to allow crossover based on the detection negotiation enable configuration. If only the C bit is set and the detection negotiation is enabled, the received route can finally crossed successfully. From the perspective of route optimization, if the connectivity negotiation is not enabled on the sender, the connectivity negotiation route may not be sent.

4. Detection Protocol Extension

There are various detection protocols. In this chapter, ping packets and 2544 packets are used as examples to describe the extension of the detection packets for EVPN all-active scenes.

4.1. Ping Packets Extension

In order to support ping connectivity detection based on EVPN instances, ping packets need to be extended shown as figure 3.

```
               0             7|8           15|16                        31|
               +--------------------------------------------------------+
               |  Type(0 or 8)|   Code(0)    |       CheckSum           |
               +--------------------------------------------------------+
               |         Identifier          |       Sequence           |
               +--------------------------------------------------------+
               |               bypass ipv4/ipv6 address                 |
               +--------------------------------------------------------+
               |               bypass ipv6 address                      |
               +--------------------------------------------------------+
               |               bypass ipv6 address                      |
               +--------------------------------------------------------+
               |               bypass ipv6 address                      |
               +--------------------------------------------------------+
               |               other option data                        |
               +--------------------------------------------------------+
                 Figure 3. Extended Ping Packets Format
```

The first 4 bytes or 16 bytes of the ICMP header optional data can be used as the source IP address of the multi-homed PE bypass tunnel.

4.2. 2544 packets extension

In order to support 2544 detection based on EVPN instances, 2544 packets need to be extended shown as figure 4.

```
+--------------------------------------------------------------------------------------
+--------------------+
    | ETH(VLAN) |        |        |          | (reserved) | TX_Timestamp |
RX_Timestamp | Bypass  Source    |
    | PPP       |  IPV4  |   UDP  | OPCODE
|-------------------------------------------|   IP Address      |
    | HDLC      |        |        |          |
PAYLOAD                   |                   |

|------------------------------------------------------------------------------------|
4 or 16 bytes    |
    |  14(+4+4) | 20Byte |  8Byte |  1Byte    |   3Byte    |   8Byte      |
8Byte     |                   |
    |  4Byte    |        |        |          |            |
|              |                   |
    |  4Byte    |        |        |          |            |
|              |                   |
```

```
     |------------------------------------------------------ 64Byte
---------------------|                   |
     |---------------------------2544 testing
attributes---------------------------------|-------------------|
```

Figure 4. Extended 2544 Packets Format

The first 4 bytes or 16 bytes immediately following the measuring attributes can be used as the source IP address of the multi-homed PE bypass tunnel.

5. Application Senario

5.1. EVPN over VXLAN

CE is multi-homed to local PE1/PE2 and PE3 is the remote device.

Firstly, On the initial PE1 device, PE1 sends a detection request packet carrying the bypass vxlan source IPV4 address(vxlan4 for IPV4 address, vxlan6 for IPV6 address) to PE3.

Secondly, PE3 terminates the detection request packet, sends a detection response packet which copies the source bypass vxlan IP address in the request packet.

If PE1 receives the response packet, Only when the PE1/PE2 connectivity negotiation succeeds, PE1 compares the source IP address of the bypass vxlan tunnel between PE1 and PE2 with the new IP address carried in the response packet. Because the result is that the two IP addresses are exactly equal, the response packet is sent to the CPU for processing and the final detection is successful.

If PE2 receives the response packet, PE2 also compares the source IP address of the bypass vxlan tunnel between the PE1 and the PE2 with the newly carried IP address in the response packet. Although the result is that the two IP addresses are not equal, fortunately, the bypass VXLAN tunnel between PE1 and PE2 is reachable, PE2 sends the response packet to PE1 since PE1 is the endpoint of the detection response packet indicated by the new IP address carried in the response packet. After the response packet is delivered to PE1, PE1 compares the source IP address of the bypass vxlan tunnel between PE1 and PE2 with the IP address carried in the response packet. What's exciting is that PE1 is a Terminator. The response packet is sent to the CPU for processing and the final detection is successful.

5.2. EVPN over SRV6

The universal detection process of EVPN over SRV6 is very similar to EVPN over vxlan. The difference is that the length of new IP address extended in the response packet must be 16 bytes in the SRV6 scene while the length of new IP address may be 4 bytes or 16 bytes. In addition, the tunnel between the multi-homed PEs is no longer a vxlan tunnel but an SRV6 BE or SRV6 Policy tunnel.

5.3. Limitations

The examples and principles of this draft are focused on the dual-homing scenario. For multi-homing scenarios, vxlan tunnels or srv6 tunnels are established between every two PEs among all-active PEs.

Of course, connectivity negotiation must also occur between every two PEs.

6. Security Considerations

For general EVPN Security Considerations, see [RFC7432].

7. IANA Considerations

This document does not require new codepoints.

8. Contributors

The following individuals gave significant contributions to this document:

Haibo Wang

Huawei Technologies

rainsword.wang@huawei.com

9. References

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

Authors' Addresses

GuoLiang
Huawei Technologies
101 software Avenue, Yuhua District
Nanjing 210012
China

Email: liuguoliang@huawei.com

Haibo
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 10095
China

Email: rainsword.wang@huawei.com