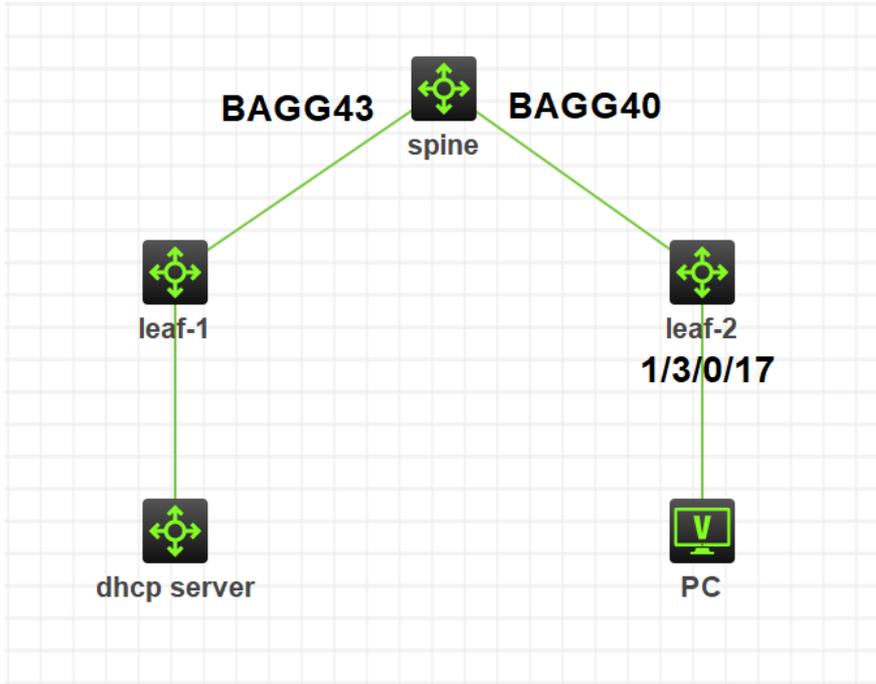


知 某局点S12516F-AF EVPN组网PC无法从远端leaf下的DHCP 服务器获取地址

VxLAN 许家豪 2022-03-17 发表

组网及说明

组网如下图:



问题描述

问题描述: 终端无法从远端dhcp服务器获取IP地址, 但终端静态绑定ip时可ping通dhcp服务器地址

过程分析

过程分析

红框中是远端dhcp server 地址，通过BGP同步过来的路由，下一跳是152.56.249.238

```
Destinations : 23      Routes : 23

Destination/Mask    Proto  Pre  Cost           NextHop             Interface
0.0.0.0/0           BGP    255  0              152.56.59.251      Vsi5
0.0.0.0/32          Direct 0    0              127.0.0.1           InLoop0
127.0.0.0/8         Direct 0    0              127.0.0.1           InLoop0
127.0.0.0/32        Direct 0    0              127.0.0.1           InLoop0
127.0.0.1/32        Direct 0    0              127.0.0.1           InLoop0
127.255.255.255/32 Direct 0    0              127.0.0.1           InLoop0
152.54.17.0/24       Direct 0    0              152.54.17.1         Vsi6
152.54.17.0/32      Direct 0    0              152.54.17.1         Vsi6
152.54.17.1/32      Direct 0    0              127.0.0.1           InLoop0
152.54.17.255/32    Direct 0    0              152.54.17.1         Vsi6
152.54.18.0/24       Direct 0    0              152.54.18.1         Vsi7
152.54.18.0/32      Direct 0    0              152.54.18.1         Vsi7
152.54.18.1/32      Direct 0    0              127.0.0.1           InLoop0
152.54.18.18/32     BGP    255  0              152.56.249.238     Vsi5
152.54.18.19/32     BGP    255  0              152.56.249.238     Vsi5
152.54.18.21/32     BGP    255  0              152.56.249.238     Vsi5
152.54.18.22/32     BGP    255  0              152.56.249.238     Vsi5
152.54.18.23/32     BGP    255  0              152.56.249.238     Vsi5
152.54.18.200/32    BGP    255  0              152.56.249.238     Vsi5
152.54.18.255/32    Direct 0    0              152.54.18.1         Vsi7
224.0.0.0/4         Direct 0    0              0.0.0.0             NULL0
224.0.0.0/24        Direct 0    0              0.0.0.0             NULL0
255.255.255.255/32 Direct 0    0              127.0.0.1           InLoop0
```

报文封装走的是tunnel 10

```
Tunnel10
Current state: UP
Line protocol state: UP
Description: Tunnel10 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 152.56.249.235, destination 152.56.249.238
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 1066 packets, 66204 bytes, 0 drops
Output: 21 packets, 1922 bytes, 0 drops
```

Spine出口到152.56.249.238的下一跳为152.56.249.141，出口为BAGG43

```
<HN2-103-A03B03-Spine-H3C-S12516F-01> dis ip rou
<HN2-103-A03B03-Spine-H3C-S12516F-01>dis ip routing-table 152.56.249.238

Summary count : 2

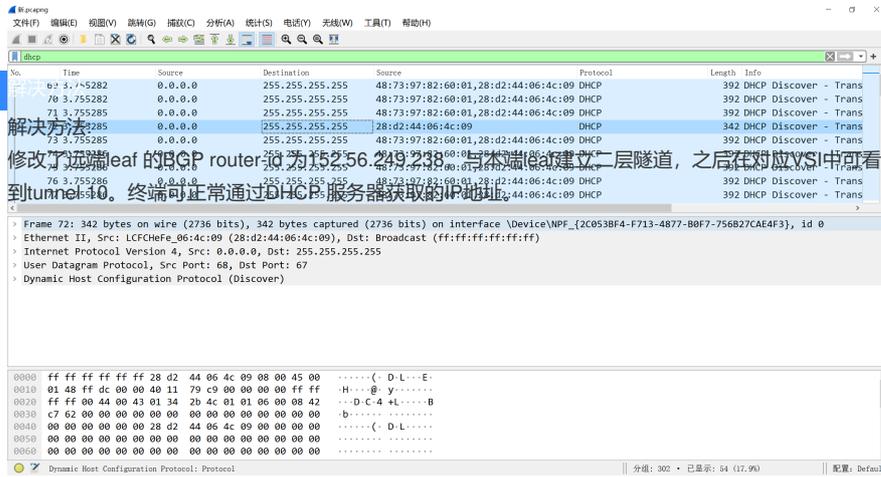
Destination/Mask    Proto  Pre  Cost           NextHop             Interface
0.0.0.0/0           O_ASE1 150 1101          152.56.59.69        Vlan517
152.56.249.238/32   O_INTRA 10   1              152.56.249.141      Vlan583
<HN2-103-A03B03-Spine-H3C-S12516F-01>dis arp 152.56.249.141
Type: S-Static      D-Dynamic          O-OpenFlow         R-Rule             M-Multiport       I-Invalid
IP address          MAC address        VLAN/VSI name      Interface          Aging Type
152.56.249.141      4873-9782-5e01    583                BAGG43
```

ping测试能通

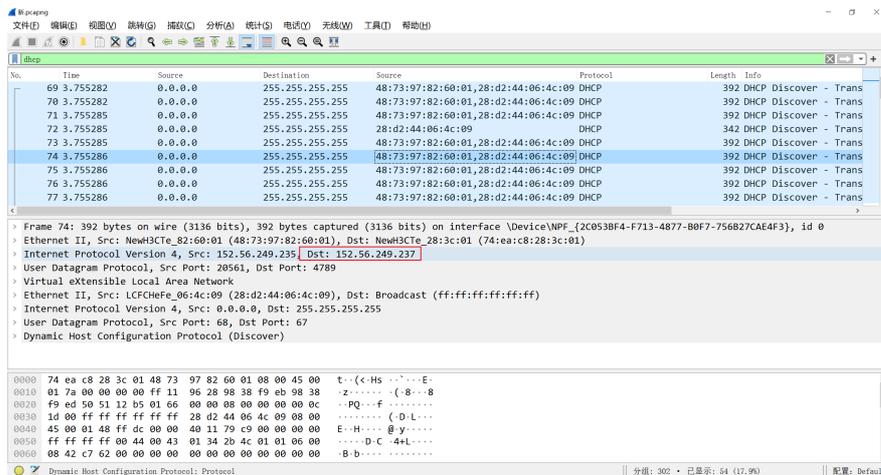
```
No.    Time    Source                Destination           Source                Protocol  Length  Info
--
13    0.700688  152.54.18.220        152.54.18.200        48:73:97:82:60:01,48:73:97:82:60:01  ICMP     124     Echo (ping) request
14    0.701745  152.54.18.200        152.54.18.220        74:ea:c8:28:3c:01,48:73:97:82:5e:01  ICMP     124     Echo (ping) reply
31    1.712659  152.54.18.220        152.54.18.200        48:73:97:82:60:01,48:73:97:82:60:01  ICMP     124     Echo (ping) request
32    1.714100  152.54.18.200        152.54.18.220        74:ea:c8:28:3c:01,48:73:97:82:5e:01  ICMP     124     Echo (ping) reply
52    2.722622  152.54.18.220        152.54.18.200        48:73:97:82:60:01,48:73:97:82:60:01  ICMP     124     Echo (ping) request
53    2.724239  152.54.18.200        152.54.18.220        74:ea:c8:28:3c:01,48:73:97:82:5e:01  ICMP     124     Echo (ping) reply
66    3.734258  152.54.18.220        152.54.18.200        48:73:97:82:60:01,48:73:97:82:60:01  ICMP     124     Echo (ping) request
67    3.735679  152.54.18.200        152.54.18.220        74:ea:c8:28:3c:01,48:73:97:82:5e:01  ICMP     124     Echo (ping) reply
84    4.746449  152.54.18.220        152.54.18.200        48:73:97:82:60:01,48:73:97:82:60:01  ICMP     124     Echo (ping) request
85    4.747549  152.54.18.200        152.54.18.220        74:ea:c8:28:3c:01,48:73:97:82:5e:01  ICMP     124     Echo (ping) reply
104   5.757745  152.54.18.220        152.54.18.200        48:73:97:82:60:01,48:73:97:82:60:01  ICMP     124     Echo (ping) request
105   5.759130  152.54.18.200        152.54.18.220        74:ea:c8:28:3c:01,48:73:97:82:5e:01  ICMP     124     Echo (ping) reply
122   6.770781  152.54.18.220        152.54.18.200        48:73:97:82:60:01,48:73:97:82:60:01  ICMP     124     Echo (ping) request
123   6.772177  152.54.18.200        152.54.18.220        74:ea:c8:28:3c:01,48:73:97:82:5e:01  ICMP     124     Echo (ping) reply

Frame 13: 124 bytes on wire (992 bits), 124 bytes captured (992 bits) on interface \Device\NPF_{2C0538F4-F713-4877-B0F7-75682CAE4F3}, id 0
> Ethernet II, Src: NewH3Cte_82:60:01 (48:73:97:82:60:01), Dst: NewH3Cte_28:3c:01 (74:ea:c8:28:3c:01)
> Internet Protocol Version 4, Src: 152.54.18.220, Dst: 152.54.18.200
> User Datagram Protocol, Src Port: 20640, Dst Port: 4789
> Virtual extensible Local Area Network
> Ethernet II, Src: NewH3Cte_82:60:01 (48:73:97:82:60:01), Dst: NewH3Cte_82:5e:01 (48:73:97:82:5e:01)
> Internet Protocol Version 4, Src: 152.54.18.220, Dst: 152.54.18.200
> Internet Control Message Protocol
```

DHCP discover报文进入到了leaf设备，并在隧道中广播，但未在tunnel 10中广播



没有红框中的地址为152.26.249.238的报文,即报文未进入tunnel 10



查看配置下发dis vxlan tunnel中 tunnel 10隧道未存在于对应vsi中

通过如下命令查看下发,本地leaf与远端238leaf建立隧道时,是通过5类路由建立的,而5类路由建立的隧道是抑制泛洪的,只能转发已知流量,这也解释了为什么终端绑定IP时,Ping是可以通的。

[Leaf-12516f-af]display bgp l2vpn evpn route-type imet

[Leaf-12516f-af]dis bgp l2vpn evpn route-type ip-prefix

查看两端Leaf vxlan的配置都是相同的,因此应该按照3类路由建立二层隧道才对,进一步查看远端leaf 238设备的BGP相关配置发现,BGP router-id配置为152.56.249.237,相当于有两台leaf设备有相同的BGP router-id,因此本端leaf无法与远端leaf以3类路由建立隧道。

在EVPN组网中,不同类路由建立的隧道类型不同,3类路由建立的二层隧道是支持广播的,5类路由建立的隧道是三层隧道无法广播。又因为DHCP discover报文为广播报文,因此未在tunnel 10中向远端leaf广播。

- Ethernet Auto-Discovery Route(RT-1): 在站点多归属组网中通告ES信息,以便实现水平分割、Aliasing和主备备份等特性。
- MAC/IP Advertisement Route(RT-2): 通告MAC/IP地址信息。
- Inclusive Multicast Ethernet Tag Route(RT-3): 通告VTEP及其所属VXLAN,以实现VTEP自动发现、自动建立VXLAN隧道、自动创建VXLAN广播表等。(以router-id建立隧道)
- Ethernet Segment Route(RT-4): 用来通告ES及其连接的VTEP信息,以便发现连接同一ES的VTEP冗余组其他成员,以及在冗余组之间选举指定转发器DF等。
- IP Prefix Advertisement Route(RT-5): IP前缀路由,以IP前缀的形式通告外部路由。

