

知 某局点addc场景业务转发不通问题

ADDc方案 姚智祥 2022-06-21 发表

组网及说明

无

告警信息

无

问题描述

现场leaf交换机下挂的服务器上有2个虚拟机地址分别为10.70.36.71/72，虚拟机的vip地址是10.70.36.41，交换机S6900为网关。客户反馈服务器实地址的业务正常，虚地址.41业务不通，排查发现服务器虚地址到网关都不通，交换机学不到vip的地址10.70.36.41的arp，导致业务转发异常。

过程分析

问题分析:

1. 首选查看arp表项确实没收到41的arp信息:

```
<S>MSR-Z3IDC_3F_F07_40U&F10_40U-SER-ASW-H3C_S6900-364>dis arp vpn-instance 6smnsqfal69rnpf473qq8ou37d
Ip address      Mac address    Vlan/Vsi      Interface      Aging Type
10.70.36.51     fa16-3ee9-031d 0               BAGG6         791 D
10.70.36.52     fa16-3eec-2f29 0               BAGG6         1013 D
10.70.36.55     fa16-3e83-49fe 0               BAGG18        613 D
10.70.36.56     fa16-3ed2-2fd5 0               BAGG18        650 D
10.70.36.57     fa16-3e9d-59eb 0               BAGG4         251 D
10.70.36.58     fa16-3e70-177a 0               BAGG17        1005 D
10.70.36.61     fa16-3e6b-f7f9 0               BAGG6         1194 D
10.70.36.62     fa16-3ebd-5ff6 0               BAGG6         1198 D
10.70.36.63     fa16-3eef-f2e9 0               BAGG6         1196 D
10.70.36.64     fa16-3e98-8df3 0               BAGG6         1191 D
10.70.36.71     fa16-3e55-1100 0               BAGG6         1103 D
10.70.36.72     fa16-3e98-6fcb 0               BAGG6         368 D
```

2. 查看控制器配置, 发现地址均在控制器上上线:

名称	类型	IP地址	MAC地址	虚拟网...	状态	静默时...	主机接入设...	VTEP IP	接入网...	端口编号	操作
---	phy	10.70.36.41	fa:16:3e:54:18:d0	network...	●	30158	172.31.131.13	172.31.255.23	BAGG6	端口编号...	操作

名称	类型	IP地址	MAC地址	虚拟网...	状态	静默时...	主机接入设...	VTEP IP	接入网...	端口编号	操作
---	phy	10.70.36.71	fa:16:3e:54:18:d0	network...	●	19	172.31.131.13	172.31.255.23	BAGG6	端口编号...	操作

名称	类型	IP地址	MAC地址	虚拟网...	状态	静默时...	主机接入设...	VTEP IP	接入网...	端口编号	操作
---	phy	10.70.36.72	fa:16:3e:98:6f:cb	network...	●	2	172.31.131.13	172.31.255.23	BAGG6	端口编号...	操作

3. 查到路由表项:

```
<S>MSR-Z3IDC_3F_F07_40U&F10_40U-SER-ASW-H3C_S6900-364>dis ip routing-table vpn-instance 6smnsqfal69rnpf473qq8ou37d 10.70.36.41
Summary count : 3
Destination/Mask Proto Pre Cost NextHop Interface
0.0.0.0/0 BGP 255 0 172.31.255.1 Vsi30108
10.70.36.0/22 Direct 0 0 10.70.36.1 Vsi736
10.70.36.0/24 BGP 255 0 172.31.255.1 Vsi30108
<S>MSR-Z3IDC_3F_F07_40U&F10_40U-SER-ASW-H3C_S6900-364>dis ip routing-table vpn-instance 6smnsqfal69rnpf473qq8ou37d 10.70.36.71
Summary count : 3
Destination/Mask Proto Pre Cost NextHop Interface
0.0.0.0/0 BGP 255 0 172.31.255.1 Vsi30108
10.70.36.0/22 Direct 0 0 10.70.36.1 Vsi736
10.70.36.0/24 BGP 255 0 172.31.255.1 Vsi30108
```

发现每个地址有3条路由

4. 继续排查fib转发表项:

```
<S>MSR-Z3IDC_3F_F07_40U&F10_40U-SER-ASW-H3C_S6900-364>dis fib vpn-instance 6smnsqfal69rnpf473qq8ou37d 10.70.36.41
Destination count: 1 FIB entry count: 1
Flag:
U:Useable G:Gateway H:Host B:Blackhole D:Dynamic S:Static
R:Relay F:FRR
Destination/Mask NextHop Flag OutInterface/Token Label
10.70.36.0/24 172.31.255.1 UG Vsi30108
<S>MSR-Z3IDC_3F_F07_40U&F10_40U-SER-ASW-H3C_S6900-364>dis fib vpn-instance 6smnsqfal69rnpf473qq8ou37d 10.70.36.71
Destination count: 1 FIB entry count: 1
Flag:
U:Useable G:Gateway H:Host B:Blackhole D:Dynamic S:Static
R:Relay F:FRR
Destination/Mask NextHop Flag OutInterface/Token Label
10.70.36.71/32 10.70.36.71 UH Vsi736
<S>MSR-Z3IDC_3F_F07_40U&F10_40U-SER-ASW-H3C_S6900-364>dis fib vpn-instance 6smnsqfal69rnpf473qq8ou37d 10.70.36.72
Destination count: 1 FIB entry count: 1
Flag:
U:Useable G:Gateway H:Host B:Blackhole D:Dynamic S:Static
R:Relay F:FRR
Destination/Mask NextHop Flag OutInterface/Token Label
10.70.36.72/32 10.70.36.72 UH Vsi736
```

发现vip地址41下一跳是I3的vsi30108, 实地址71/72下一跳是I2的vsi736, 其中leaf网关访问71/72没问题, 访问41不通, 此处开始发现问题所在。通过分析41的下一跳vsi30108发现是到spine S12516-F上。

5. 排查spine 设备:

```
<S>MSR-Z3IDC_3F_G05_4U&G06_4U-SER-CSW-H3C_S12516-162>dis ip routing-table vpn-instance 6smnsqfal69rnpf473qq8ou37d 10.70.36.71
Summary count : 4
Destination/Mask Proto Pre Cost NextHop Interface
0.0.0.0/0 Static 60 0 10.1.1.71 Vlan3034
10.70.36.0/22 BGP 255 0 172.31.255.21 Vsi30108
10.70.36.0/24 Static 60 0 22.22.22.142 Ragg1023.97
10.70.36.71/32 BGP 255 0 172.31.255.23 Vsi30108
```

发现spine上有一条静态的10.70.36.0段的路由

继续查看:

```
<S>MSR-Z3IDC_3F_G05_4U&G06_4U-SER-CSW-H3C_S12516-162>dis cu
<S>MSR-Z3IDC_3F_G05_4U&G06_4U-SER-CSW-H3C_S12516-162>dis current-configuration | in 10.70.36.0
ip route-static vpn-instance 2662b3c8b4848106751mfmfhk 10.70.36.0 22 10.1.1.71 description SDN_ROUTE
ip route-static vpn-instance 410154c193173c0uffe3ff@ 10.70.36.0 22 10.1.1.61 description SDN_ROUTE
ip route-static vpn-instance 6smnsqfal69rnpf473qq8ou37d 10.70.36.0 24 22.22.22.142 description OSCARII_MOSC3-To-OSCARII_MOSC
```

发现spine上被人配置了一条24位的静态路由

综上, 问题原因如下:

在leaf上查看10.70.36.41的地址fib表项, 因为最长掩码匹配优先, 所以交换机并不会匹配直连的22位的路由直接转发到下挂服务器, 而是匹配了evpn从spine传过来24位的10.70.36.0/24的bgp路由, 导致41的流量进行了错误下一跳转发。

解决方法

将配置的24位的静态路由删除后恢复

