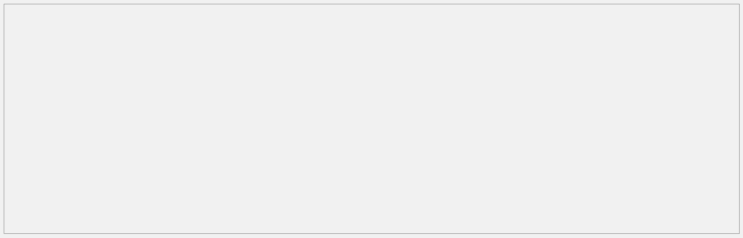
三层转发不通排查分析

彭绍穗 2014-07-27 发表

三层转发不通排查分析

一、组网说明



二、问题描述

查看路由表正常，但是直连ping不通。

三、 三层转发流程

当报文为三层报文时，ip包的目的mac为三层接口mac。设备对于目的mac匹配三层接口mac的报文按照先arp后路由的顺序查找三层硬件表项，找到转发的目的端口和mac，进行目的mac和源mac的替换，ttl减一后向相应端口转发。因此当出现三层转发不通时，可以按照上面的流程逐一排查。

四、 分析过程

1、查看路由表以及fib表，是否有相应的路由表象

三层转发不通自然第一步首先是要检查路由表项以及相应的fib表项学习是否正确。

<H3C>display ip routing-table

Destination/Mask	Proto	Pre	Cost	NextHop	Interface
10.44.11.160/27	Direct	0	0	10.44.11.161	Vlan3130
10.44.11.160/32	Direct	0	0	10.44.11.161	Vlan3130
10.44.11.161/32	Direct	0	0	127.0.0.1	InLoop0
10.44.11.191/32	Direct	0	0	10.44.11.161	Vlan3130
10.44.42.192/26	Direct	0	0	10.44.42.193	Vlan3187
10.44.42.192/32	Direct	0	0	10.44.42.193	Vlan3187
10.44.42.193/32	Direct	0	0	127.0.0.1	InLoop0
10.44.42.255/32	Direct	0	0	10.44.42.193	Vlan3187
10.71.16.152/29	OSPF	150	20	172.16.129.25	Vlan4003
10.71.16.160/29	OSPF	150	20	172.16.129.25	Vlan4003

<H3C>display fib

Destination/Mask	Nexthop	Flag	OutInterface/Token	Label
10.44.11.160/27	10.44.11.161	U	Vlan3130	Null
10.44.11.160/32	10.44.11.161	UBH	Vlan3130	Null
10.44.11.161/32	127.0.0.1	UH	InLoop0	Null
10.44.11.191/32	10.44.11.161	UBH	Vlan3130	Null
10.44.42.192/26	10.44.42.193	U	Vlan3187	Null
10.44.42.192/32	10.44.42.193	UBH	Vlan3187	Null
10.44.42.193/32	127.0.0.1	UH	InLoop0	Null
10.44.42.208/32	10.44.42.208	UH	Vlan3187	Null
10.44.42.255/32	10.44.42.193	UBH	Vlan3187	Null
10.71.16.152/29	172.16.129.25	UDGR	Vlan4003	Null
10.71.16.160/29	172.16.129.25	UDGR	Vlan4003	Null

检查路由表，fib表如果学习不正确，那么需要检查配置是否正确，以便学习到正确的路由。

2、查看arp学习是否正常，检查arp驱动诊断信息

检查完路由表正常后，下一步查看arp表。

<H3C>display arp all

IP address	MAC address	VLAN	Interface	Aging	Type
172.24.172.132	0022-4610-6687	815	GE2/2/0/1	4	D
172.24.172.133	0022-4610-6687	815	GE2/2/0/1	4	D
172.24.172.134	0022-4610-6687	815	GE2/2/0/1	4	D

```
172.24.172.135  0022-4610-6687 815     GE2/2/0/1       4     D
172.24.172.136  0022-4610-6687 815     GE2/2/0/1       4     D
172.24.172.137  0022-4610-6687 815     GE2/2/0/1       4     D
172.24.172.138  0022-4610-6687 815     GE2/2/0/1       4     D
172.24.172.139  0022-4610-6687 815     GE2/2/0/1       4     D
172.24.172.140  0022-4610-6687 815     GE2/2/0/1       4     D
172.24.172.141  0022-4610-6687 815     GE2/2/0/1       4     D
172.24.172.142  0022-4610-6687 815     GE2/2/0/1       4     D
10.44.42.208    de06-8c00-3558 3187    GE2/2/0/24      20    D
```

然后查看相对应的arp是否下发到硬件中去了。

```
[H3C-diagnose]debug ipv4-drv show arp 0 10.44.42.208 slot 2
--- UNIT: 1 ---
- VRF:           0
- IP ADDR:       10.44.42.208
- LOCATION:      defip
- EGRESS ID:     100002
- IPMC PTR:      0
- FLAGS:         0x10000
- EGRESS FLAGS:  0x0
- INTF NUM:      0
- MAC ADDR:      de06-8c00-3558
- VLAN:          0
- DMOD:          20
- DPORT:         1
- TRUNK:         0
- FRR LABEL:     0
```

如若没有下发到硬件中去检查arp表项是否已经超出规格

```
[H3C-diagnose]debug l3intf-drv show statistics slot 1
- ARP
      SPECIFICATION:       131072                  //单板的arp规格
      COUNT:               0                       //下发到该单板的arp
      NHCOUNT:             58
- IPV4 ROUTE
      SPECIFICATION:       16384
      COUNT:               16384
- ND
      SPECIFICATION:       65536
      COUNT:               0
- IPV6 ROUTE
      SPECIFICATION:       8192
      ROUTE COUNT:         0
- ARP LOCATION:            ARP&DEFIP
- ND LOCATION:             ND
- IPV4 PROXY MODE:         NO PROXY
- IPV6 PROXY MODE:         NO PROXY
Notes: One IPv6 record equals two IPv4 records.
```

如果出现arp学习异常，请尝试静态绑定arp进行测试，如果arp超出规格，请减少arp的条目。

3、三层转发底层表项

如果arp检查也无异常。则要检查三层转发表项。三层转发底层表项主要包括host表，defip表、下一跳egress表以及三层接口intf表。三层转发流程如图所示，分别检查转发流程是否有相对应的表项。

对于目的ip为直连路由的报文先检查host表项，得到下一跳表的索引值

[H3C-diagnose]bcm 1 0 l3/l3table/show，

Entry VRF IP address Mac Address INTF MOD PORT CLASS HIT

187 0 172.24.173.200 00:00:00:00:00:00 105848 0 0 32 n

188 0 10.44.39.109 00:00:00:00:00:00 105693 0 0 32 n

189 0 10.44.39.104 00:00:00:00:00:00 100466 0 0 32 n

190 0 10.44.42.208 00:00:00:00:00:00 100836 0 0 32 n //100836

191 0 172.24.172.130 00:00:00:00:00:00 100654 0 0 32 n

192 0 10.44.5.247 00:00:00:00:00:00 100002 0 0 32 n

对于目的ip为非直连路由的报文检查host没有命中，紧接着便检查defip表项。

[H3C-diagnose]bcm 1 0 l3/defip/show

Entry VRF Net addr Next Hop Mac INTF MODID PORT CLASS HIT vl
an

1451 6 10.19.141.184/29 00:00:00:00:00:00 102664 0 0 0 32 n

1452 6 10.19.140.48/29 00:00:00:00:00:00 102664 0 0 0 32 n

1452 0 8.71.16.40/29 00:00:00:00:00:00 102463 0 0 0 32 n

1453 0 10.71.16.152/29 00:00:00:00:00:00 102463 0 0 0 32 n

1453 0 10.71.16.160/29 00:00:00:00:00:00 102463 0 0 0 32 n

1454 0 172.16.255.144/29 00:00:00:00:00:00 104737 0 0 0 32 n

1454 0 172.16.88.24/29 00:00:00:00:00:00 100242 0 0 0 32 n

1455 0 172.16.146.104/29 00:00:00:00:00:00 101459 0 0 0 32 n

1455 0 172.16.240.128/29 00:00:00:00:00:00 101444 0 0 0 32 n

1456 0 172.16.240.144/29 00:00:00:00:00:00 101444 0 0 0 32 n

1456 6 172.24.224.0/29 00:00:00:00:00:00 104580 0 0 0 32 n

根据下一跳索引号查找下一跳表，可以得到路由表下一跳的MAC、出VLAN、三层接口索引号、mod和port 。

====bcm chassis 2 slot 2 chip 0 l3/egress/show====

Entry Mac Vlan INTF PORT MOD MPLS_LABEL

100829 00:0f:e2:22:ef:40 4003 143 3 72 1146

100830 00:10:f3:32:87:50 2003 38 23 65 -1

100831 00:10:5c:c6:f7:c8 801 42 4 68 -1

100832 00:0f:e2:22:ef:40 4003 148 3 72 11631

100833 00:0f:e2:22:ef:40 4003 149 3 72 19038

100836 de:06:8c:00:35:58 3187 235 23 72 -1 //235

100834 00:0f:e2:22:ef:40 4003 148 3 72 11634

100835 00:25:ab:2d:27:14 3166 4 20 68 -1

根据下一跳表查到三层出接口表，得到的出接口mac即为二层封装时候的源mac。

[H3C-diagnose]bcm 1 0 l3/intf/show

Unit Intf VRF Group VLAN Source Mac MTU TTL Tunnel InnerVlan

0 0 0 0 4095 0c:da:41:b5:d0:6f 16383 0 0 0

0 1 0 0 3 0c:da:41:b5:d0:6f 16383 0 0 0

0 2 6 0 3164 0c:da:41:b5:d0:6f 16383 0 0 0

0 3 6 0 3165 0c:da:41:b5:d0:6f 16383 0 0 0

0 4 6 0 3166 0c:da:41:b5:d0:6f 16383 0 0 0

0 5 6 0 3167 0c:da:41:b5:d0:6f 16383 0 0 0

0 6 6 0 3168 0c:da:41:b5:d0:6f 16383 0 0 0

0 235 0 0 3187 0c:da:41:b5:d0:6f 16383 0 0 0

如若底层表项中没有找到相关的路由信息查看是否路由超规格或者资源不足

debug l3intf-drv show statistics slot 2

- ARP

SPECIFICATION: 131072

COUNT: 0
NHCOUNT: 58
- IPV4 ROUTE
SPECIFICATION: 16384 //单板的ipv4路由规格
COUNT: 16383 //实际下发到单板的ipv4路由
- ND
SPECIFICATION: 65536
COUNT: 0
- IPV6 ROUTE
SPECIFICATION: 8192
ROUTE COUNT: 0
- ARP LOCATION: ARP&DEFIP
- ND LOCATION: ND
- IPV4 PROXY MODE: NO PROXY
- IPV6 PROXY MODE: NO PROXY
Notes: One IPv6 record equals two IPv4 records.

如若出现路由超规格，检查local log会有相关的路由下发失败的记录。
local logbuffer slot 0 display
Feb 20 2014 11:13:41:353178:
LINE:2933-TASK:kfib/1-FUNC:drv_l3uc_sdk_add_ipv4_defip:
Fail to add defip!Unit=0,iRv=-6,ip=0xa23b300,mask=0xfffffe0,vrf=6,intf=105997.
Feb 20 2014 11:13:41:353208:
LINE:1847-TASK:kfib/1-FUNC:drv_ipv4_uc_hard_addroute:
drv_l3uc_sdk_add_ipv4_defip return 0x4001000b
Feb 20 2014 11:14:17:324977:
LINE:14629-TASK:mIPC-FUNC:drv_mac_set_statinfo:
DRV_DEVM_GetUnitID ERR, modid(126), unit 0, vid 1, mac 00e0-fc0f-8c25, group 0.
Feb 20 2014 11:15:53:442708:
LINE:2116-TASK:kfib/1-FUNC:drv_ipv4_uc_shim_addroute:
Call drv_ipv4_uc_hard_addroute return 0x4001000b, vrf=6, ip=ac11fb00, mask=fffffffe0
至此问题分析完毕。