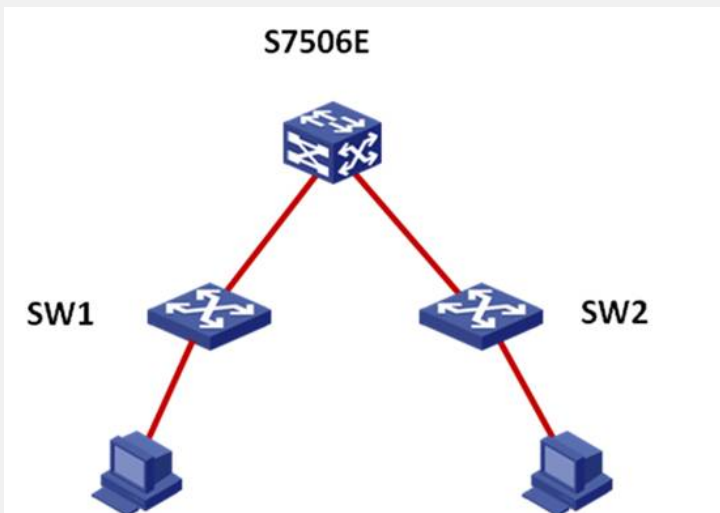


某局点S7506E 迭代路由问题导致转发不通 案例分析

一、 组网：



网络中使用S7506E作为三层网关设备，下挂PC通过S7506E实现三层转发。

S7506E版本是R6703。

二、 问题描述：

客户反馈最近S7506E设备上增加了部分业务，其中部分业务出现不通的问题。现场通过两台PC 10.129.244.244和PC 10.129.10.1直连S7506E跨三层转发测试不通，从S7506E ping两个PC地址都可以通，而从PC 10.129.244.244 ping 59.205.255.191地址发现也无法ping通。

三、 过程分析：

从现场采集的信息分析来看：

1、PC 10.129.244.244和PC 10.129.10.1 ping不通

在S7506E上通过display arp可以看到两个PC的ARP地址，软件学习ARP正常。

10.129.244.244 0023-8b4a-eacd 244 GE2/0/14 N/A D

10.129.10.1 00e0-4c87-7d8d 10 GE2/0/3 N/A D

检查交换机三层转发硬件表项，发现：

- IPv4 ARP Information Slot 2

--- UNIT: 0 ---

- VRF: 0

- IP ADDR: 10.129.244.244

- LOCATION: host

- EGRESS ID: 100209

- IPMC PTR: 0

- FLAGS: 0xc

- HITBIT: Source Dest

- EGRESS FLAGS: 0x0

- INTF NUM: 4095

```
- MAC ADDR:      0000-0000-0000      //此表项不正常
- VLAN:          4095
- DMOD:          8
- DPORT:         31
- TRUNK:         0
- FRR LABEL:     0
```

```
- IPv4 ARP Information Slot 2
```

```
--- UNIT: 0 ---
```

```
- VRF:           0
- IP ADDR:       10.129.10.1
- LOCATION:      host
- EGRESS ID:     100204
- IPMC PTR:      0
- FLAGS:         0xc
- HITBIT:        Source Dest
- EGRESS FLAGS:  0x0
- INTF NUM:      8
```

```
- MAC ADDR:      00e0-4c87-7d8d      //这个表项是正常的
- VLAN:          10
- DMOD:          8
- DPORT:         2
- TRUNK:         0
- FRR LABEL:     0
```

display arp看到的是软件表项，是正确的，而硬件中10.129.244.244的表项是有问题的，没有指定目的MAC 地址和端口，这样流量是无法转发出去的。

从S7506E上ping两台PC查找的是软件表项，所以可以ping通，而两个PC之间三层转发互通是查找硬件表项，因为硬件表项学的有问题所以会不通。

检查设备上记录的异常信息，发现所有槽位都频繁的报出下发硬件表项失败的记录信息，并有提示资源不足：

```
Jan 29 2013 17:38:39:0740:
```

```
LINE:2632, File platform_bcm/drv/l3/ipv4/drv_ipv4_uc.c-TASK:FIB-FUNC::
```

```
ArpCount=251, ArpNHCount=251, RouteCount=127749, EcmpRouteCount=127009, ulDrvflag=0.
```

```
Jan 29 2013 17:38:39:0743:
```

```
LINE:2400, File platform_bcm/drv/l3/ipv4/drv_ipv4_shim.c-TASK:FIB-FUNC::
```

```
No resource to add ecmp route entry!ulRouteTotalCount=127749, ulUnitID=1Jan 29 2013 17:38:39:0743:
```

```
LINE:921, File platform_bcm/drv/l3/ipv4/drv_ipv4_shim.c-TASK:FIB-FUNC:
```

```
Call drv_ipv4_shim_uc_ipv4_add_ecmp_route return 11,ip = 0x3bcdfe45, mask = 0xffffffff
```

```
Jan 29 2013 17:38:39:0743:
```

```
LINE:1305, File platform_bcm/drv/l3/ipv4/drv_ipv4_shim.c-TASK:FIB-FUNC::
```

```
No resource to add route entry!ulRouteTotalCount=127749, ulUnitID=1
```

```
Jan 29 2013 17:38:39:0743:
```

```
LINE:968, File platform_bcm/drv/l3/utils/drv_ipv4_shim.c-TASK:FIB-FUNC:
:
Call drv_ipv4_shim_uc_ipv4_add_forward_route return 11,ip = 0xac0000
0,mask = 0xffff0000
Jan 29 2013 17:38:39:0743:
LINE:417, File platform_bcm/drv/l3/utils/drv_ipv4_shim.c-TASK:FIB-FUNC:
:
Call drv_ipv4_shim_ipv4_addroute_hard error,Ret=11,ip = 0xac00000,m
ask = 0xffff0000
Jan 29 2013 17:38:39:0743:
LINE:3462, File platform_bcm/drv/l3/ipv4/drv_ipv4_uc_intf.c-TASK:FIB-FU
NC::
ulRet=11
Jan 29 2013 17:38:39:0743:
LINE:2617, File platform_bcm/drv/l3/ipv4/drv_ipv4_uc.c-TASK:FIB-FUNC::
V5 param:ulVrfIndex=0, ulDstMask=fff00000, ulIfIndex=0x93a0fa0, uIA
TIndex=0
继续检查设备的表项的资源使用情况，发现等价和普通路由的规格都已经满了：
Multipath Egress Object 200000
Ecmp Cnt: 1
Interfaces: 100006
Multipath Egress Object 200001
Ecmp Cnt: 1
Interfaces: 100007
Multipath Egress Object 200002
Ecmp Cnt: 1
Interfaces: 100005
Multipath Egress Object 200003
. . . . .
*****
- IPv4 Config Slot 0
*****
- ARP SIZE: 16384
- ArpCanNotSetToHW: NO
- IPV4 ROUTE SIZE: 128000
- ECMP SIZE: 16
- SDK ECMP SIZE: 16
- ECMP GROUP SIZE: 127
普通路由也已经达到规格了：
*****
- IPv4 Statistics Slot 0
*****
- ROUTE TOTAL COUNT: 127865
- ROUTE DMAFAIL COUNT: 0
- ECMP ROUTE COUNT: 127127
- ECMP DMAFAIL COUNT: 0
*****
- IPv4 Config Slot 0
*****
```

```

- ARP SIZE:      16384
- ArpCanNotSetToHW: NO
- IPV4 ROUTE SIZE: 128000
- ECMP SIZE:      16
- SDK ECMP SIZE: 16
- ECMP GROUP SIZE: 127
- ND SIZE:        8192
再去检查设备上学习到的路由情况，的确有不少的等价路由和迭代路由
10.58.0.0/16      O_ASE 150 1800      59.205.128.5  Vlan4002
                  O_ASE 150 1800      59.205.128.1  Vlan4001
10.59.0.0/16      O_ASE 150 1800      59.205.128.5  Vlan4002
                  O_ASE 150 1800      59.205.128.1  Vlan4001
. . . . .
. . . . .

```

```

192.168.237.96/28 59.205.128.5  UDGR  Vlan4001 Null  Invalid
10.112.0.0/16     59.205.128.1  UDGR  Vlan4001 Null  Invalid
10.3.0.0/16       59.205.128.1  UDGR  Vlan4001 Null  Invalid
59.205.255.86/32  59.205.128.1  UDGR  Vlan4001 Null  Invalid
10.221.0.0/16     59.205.128.1  UDGR  Vlan4001 Null  Invalid
192.168.212.0/24  59.205.128.1  UDGR  Vlan4001 Null  Invalid
59.205.255.100/32 59.205.128.1  UDGR  Vlan4001 Null  Invalid
10.214.0.0/16     59.205.128.1  UDGR  Vlan4001 Null  Invalid
59.205.255.237/32 59.205.128.1  UDGR  Vlan4001 Null  Invalid
10.21.0.0/16      59.205.128.1  UDGR  Vlan4001 Null  Invalid
192.168.22.0/24   59.205.128.1  UDGR  Vlan4001 Null  Invalid

```

到此已经找到了问题的原因，在R6703版本上，S7506E新增了路由迭代的新功能，在有等价路由和迭代路由的情况下，会下发大量的硬件表项，从而易出现路由表项的超规格情况，而大量路由表项也会占用非常很多的出接口Egress表项，从而导致ARP表项也无法再下发下去。

在使用等价路由进行负载分担的场合，建议不要使用迭代模式；在需要路由表项数目较多的场合，也可以配置为非迭代模式来保证路由规格。这时可以通过下面配置，配置为非迭代模式。

```

switch-mode { route-iterative | route-normal }    缺省情况下，为迭代模式route-iterative

```

2、从PC 10.129.244.244 ping 59.205.255.191地址无法ping通也是这样的原因，59.205.255.191/32地址的路由由于硬件规格满而无法下发到硬件表项中，导致路由无法生效。

```

Routing Table : Public
Summary Count : 4
Destination/Mask  Proto Pre Cost    NextHop      Interface
0.0.0.0/0         Static 60   0    10.129.7.253 Vlan7
59.192.0.0/10     O_ASE 150 402    59.205.128.5 Vlan4002
                  O_ASE 150 402    59.205.128.1 Vlan4001
59.205.255.191/32 OSPF  10  100    59.205.128.1 Vlan4001

```

四、 解决方法：

现场由于路由表项较多并有迭代、等价路由，需要将S7506E设备配置为非迭代模式：

```

switch-mode route-normal

```

