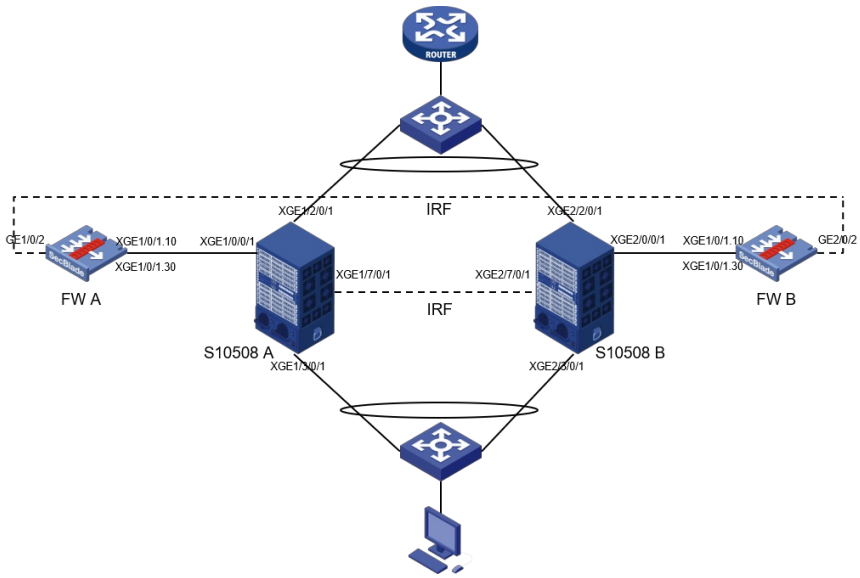
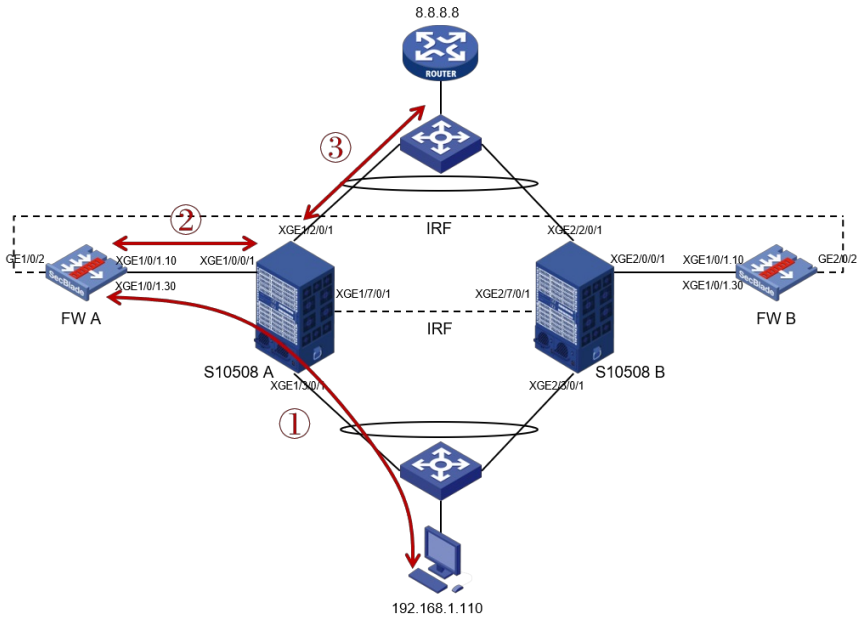
H3C S10508（V7）交换机配合SecBlade FW III单板三层转发部分流量不通问题

单板 流量统计 转发不通 丁犁 2018-03-22 发表

现场采用两台S10508交换机IRF2 + 两块SecBlade FW 插卡组网部署，如下图所示：



FW插卡IRF2后，部署冗余接口Reth3（XGE1/0/1.30为主，XGE2/0/1.30为备），作为内网用户的网关；部署冗余接口Reth2（XGE1/0/1.10为主，XGE2/0/1.10为备），作为与S10508三层互联接口。
S10508 IRF2后，部署二层聚合口（XGE1/3/0/1和XGE2/3/0/1为聚合成员接口），保证内网终端流量进入S10508设备后二层转发给FW插卡Reth3接口；部署三层聚合接口（XGE1/2/0/1和XGE2/2/0/1为聚合成员接口），保证FW将流量从Reth2接口发给S10508互联三层SVI接口后，三层转发给上游路由器。
当管理员基本转发部署操作完成后，按照部署方式，内网终端访问外网流量走向为下图所示：

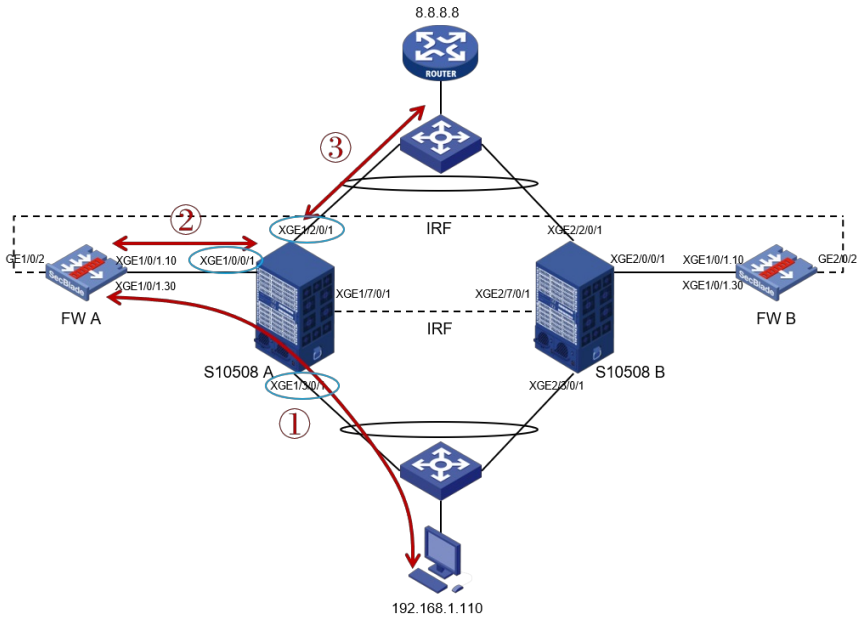


PC访问外网的流量：PC--①-->②-->③
其中：
流量① 经过S10508 A交换机时，S10508为二层转发；
流量在FWA上经过三层转发；
流量② ③ 经过S10508 A交换机时，S10508为三层转发。
出现的故障现象：
内网多台PC访问外网资源时，出现访问部分外网资源正常，部分外网资源不成功的故障。（PC上能ping通正常访问的外网资源地址，不能ping通无法访问的外网资源地址）

出现问题后，因为涉及转发流量的设备类型、数量较多，因此我们第一时间在S10508交换机上采用流

量统计的方法，缩小定位的范围（对于交换机部署流量统计的方法可参考http://kms.h3c.com/View.aspx?id=41506案例，本案例中不再重复介绍）。

对于从内到外的流量，分别在下图S10508交换机蓝圈接口上部署流量统计：



当在蓝圈接口部署流量统计Qos策略后，根据统计结果，PC无法访问某外网资源时可得出如下结论：

1、PC发出N个报文，S10508 XGE1/3/0/1收到了N-M个报文，说明有M个报文被S10508下游内网丢弃；

2、PC发出N个报文，S10508 XGE1/3/0/1收到N个报文，XGE1/0/0/1仅发出N-M个报文，说明有M个报文被S10508丢弃（S10508存在故障）；

3、PC发出N个报文，S10508 XGE1/3/0/1收到N个报文，XGE1/0/0/1发出N个报文，但XGE1/0/0/1收到N-M个报文，说明有M个报文被FW A丢弃（FW A存在故障）；

4、PC发出N个报文，S10508 XGE1/3/0/1收到N个报文，XGE1/0/0/1发出N个报文，XGE1/0/0/1收到N个报文，XGE1/2/0/1发出N-M个报文，说明有M个报文被S10508丢弃（S10508存在故障）；

5、PC发出N个报文，S10508 XGE1/3/0/1收到N个报文，XGE1/0/0/1发出N个报文，XGE1/0/0/1收到N个报文，XGE1/2/0/1发出N个报文，说明丢包位于S10508上游的网络中。

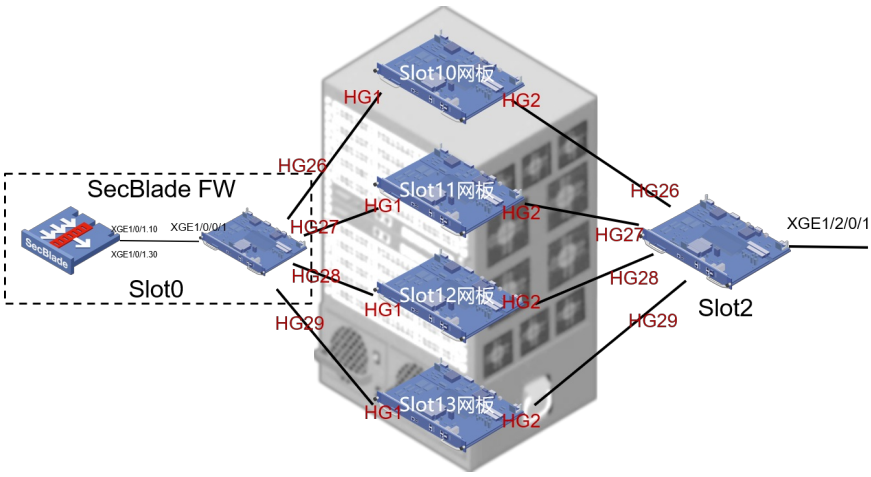
管理员实际测试发现，内网PC（192.168.1.110）无法ping通外网资源8.8.8.8后，查看S10508交换机流量统计结果为：

PC发送5个ICMP请求后，XGE1/3/0/1 inbound方向统计到5个报文，XGE1/0/0/1 inbound/outbound方向分别统计到5个报文，XGE1/2/0/1 outbound方向统计到0个报文。由此统计结果，可确认造成内网PC（192.168.1.110）无法ping通外网资源8.8.8.8的原因为，S10508交换机将报文丢弃。

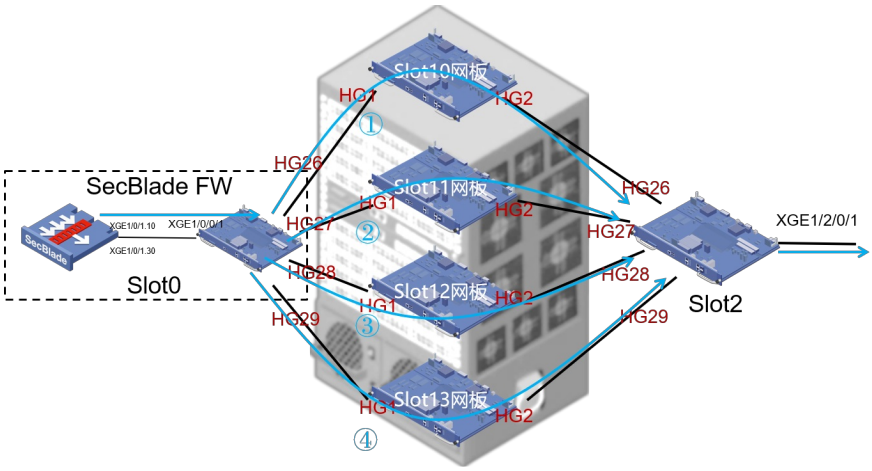
```
<S10508>display qos policy interface
Interface: Ten-GigabitEthernet1/0/0/1
  Direction: Inbound
  Policy: 1
   Classifier: 1
     Operator: AND
     Rule(s) :
      If-match acl 3500
     Behavior: 1
      Accounting enable:
        5 (Packets)
......
Interface: Ten-GigabitEthernet1/2/0/1
  Direction: Outbound
  Policy: 1
   Classifier: 1
     Operator: AND
     Rule(s) :
      If-match acl 3500
     Behavior: 1
      Accounting enable:
        0 (Packets)
```

对于分布式交换机存在跨板流量转发丢包时（流量从Slot0 XGE1/0/0/1到Slot1 XGE1/2/0/1），进一步需要明确报文具体丢弃在哪个业务板上（是Slot 0还是SLot 1）。因此需要管理人员明确不同单板在交

换机内部的互联方式，同时在S10508部署交换机内部流量统计策略（查看单板内部连接，及内部流量统计方法，可参考http://kms.h3c.com/View.aspx?id=52324案例，本案例中不再重复介绍）。
通过查看S10508 Slot0与Slot1之间的内部结构，可得到下面的连接示意图：



其中：
1、Slot 0 SecBlade FW III分为前插卡（FW左）和后插卡（交换板卡右），前后插卡通过内部连接线（单板外观上无法看到）互联。在S10508设备上查看到的XGE1/0/0/1接口为后插卡上的接口，其与前FW插卡XGE1/0/1接口内部互联。
2、Slot 0 SecBlade FW III后插卡，通过内部HG接口（HG26、HG27、HG28、HG29）分别与交换机Slot10、11、12、13交换网板的HG1接口互联，实现流量跨单板转发。
3、Slot 2交换业务单板，通过内部HG接口（HG26、HG27、HG28、HG29）分别与交换机Slot10、11、12、13交换网板的HG2接口互联，实现流量跨单板转发。
对于Slot 0 FW前插卡转发出来的流量，需要从Slot 2 XGE1/2/0/1接口转发出去，其内部流量走向应为下图蓝色线路所示：（①②③④路径随机选择一条路径转发）



通过之前在S10508交换机物理接口上部署流量统计发现XGE1/0/0/1接口收到了5个报文，但是从XGE1/2/0/1发出了0个报文的情况，再结合单板硬件内部互联HG接口互联的拓扑，分别在如下HG接口上部署统计策略，进一步明确报文丢弃在哪个硬件模块上（具体部署内部流量统计方法，可参考http://kms.h3c.com/View.aspx?id=52324案例，本案例中不再重复介绍）：
Slot 0后插卡HG26、HG27、HG28、HG29的outbound方向；
Slot10、Slot11、Slot12、Slot13 HG1的inbound方向；
Slot10、Slot11、Slot12、Slot13 HG2的outbound方向；
Slot 2插卡HG26、HG27、HG28、HG29的inbound方向。
通过部署内部流量统计，并查看相关流量转发不通后的，内部统计结果，发现Slot 0 HG26、HG27、HG28、HG29 outbound方向统计值为0

[S10508-probe]debug qacl show packet pattern chassis 1 slot 0 chip 0 28 out
========
Acl-Type Statistics based PktPattern, Stage EFP, SinglePort, Installed, Active
Prio Mjr/Sub 256/3, Group 9 [9], Slice/Idx 1/1, Entry 1348, Single: 257
Rule Match --------
Out Port: 28 //HG28接口
Source IP: 192.168.1.110, 255.255.255.255
Dest IP: 8.8.8.8, 255.255.255.255

Actions --------
Account mode packets, green and non-green
Accounting: Hi 0, **LO 0 //统计为0**

[S10508-probe]debug qacl show packet pattern chassis 1 slot 0 chip 0 26 out
========
Acl-Type Statistics based PktPattern, Stage EFP, SinglePort, Installed, Active
Prio Mjr/Sub 256/3, Group 9 [9], Slice/Idx 1/0, Entry 1347, Single: 256
Rule Match --------
Out Port: 26 //HG26接口
Source IP: 192.168.1.110, 255.255.255.255
Dest IP: 8.8.8.8, 255.255.255.255
Actions --------
Account mode packets, green and non-green
Accounting: Hi 0, **LO 0 //统计为0**

[S10508-probe]debug qacl show packet pattern chassis 1 slot 0 chip 0 27 out
========
Acl-Type Statistics based PktPattern, Stage EFP, SinglePort, Installed, Active
Prio Mjr/Sub 256/3, Group 9 [9], Slice/Idx 1/2, Entry 1349, Single: 258
Rule Match --------
Out Port: 27 //HG27接口
Source IP: 192.168.1.110, 255.255.255.255
Dest IP: 8.8.8.8, 255.255.255.255
Actions --------
Account mode packets, green and non-green
Accounting: Hi 0, **LO 0 //统计为0**

[S10508-probe]debug qacl show packet pattern chassis 1 slot 0 chip 0 29 out
========
Acl-Type Statistics based PktPattern, Stage EFP, SinglePort, Installed, Active
Prio Mjr/Sub 256/3, Group 9 [9], Slice/Idx 1/3, Entry 1350, Single: 259
Rule Match --------
Out Port: 29 //HG29接口
Source IP: 192.168.1.110, 255.255.255.255
Dest IP: 8.8.8.8, 255.255.255.255
Actions --------
Account mode packets, green and non-green
Accounting: Hi 0, LO **0 //统计为0**
由上述统计结果，便可明确相关报文被Slot 0 FW后插卡所丢弃。
由于故障流量（内网访问外网不通的流量）在S10508交换机上为三层转发，因此当锁定了报文丢弃的相关硬件后，管理员需要查看相关硬件（插卡）的三层转发资源是否足够。
在S10508上通过debug l3intf-drv show statistics chassis 1 slot 0 命令进行查看：

[S10508-probe]probe //进入诊断视图
[S10508-probe]debug l3intf-drv show statistics chassis 1 slot 0 //查看1框slot0（FW后插卡）三层资源使用情况
****
- L3INTF Statistics Chassis 1 Slot 0

PInfo: LM=0 U=0 CNT=0 C=0(1207011)
Dpc: L3=0 VX=0 EVX=0
Dbg: L3=0 VX=0 EVX=0
****
- ARP
SPECIFICATION: 16384
COUNT: 0
NHCOUNT: 61
- IPV4 ROUTE
SPECIFICATION: 65536 //该单板具有65535条IPv4路由资源
6To4 RELAY COUNT: 0
COUNT: 65536 //当前单板已使用65535条IPv4路由资源
- ND
SPECIFICATION: 8192

COUNT: 0
- IPV6 ROUTE
SPECIFICATION: 8192
ROUTE COUNT: 56
- 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.
.....

通过上述命令查看，发现Slot0 FW（后插卡）的IPv4路由资源已经被用完，因此就会造成部分三层转发流量不通。

观察S10508交换机的运行模式（dis playswitch-mode status ），发现chassis 1 slot 0单板采用了缺省的混杂（MIX-Bridging-Routing）运行模式：

<S10508>dis switch-mode status chassis 1
LPU switch mode:
Slot Current Config
0 MIX NONE
.....

在MIX混杂模式下单板仅能提供64K的FIB IPv4转发资源，因此需要管理员通过命令将slot 0单板模式修改为Routing模式，增大IPv4转发资源数量到128K，来满足实际网络的需要。

在S10508交换机上，调整单板转发模式，重启相关单板，使模式修改生效后，问题得以解决。
[S10508]switch-mode routing chassis 1 slot 0
[S10508]switch-mode routing chassis 2 slot 0 //FW插卡在两个S10508机框上均部署，且进行了IRF2，转发资源表项会同步，因此2框对应的单板也需要调整

对于交换机网络转发丢包问题，可采用如下排查部署，先流量统计缩小故障范围，再检查转发资源数量的方式进行排查：
1、Qos流量统计http://kms.h3c.com/View.aspx?id=41506
2、内部流量统计http://kms.h3c.com/View.aspx?id=52324
3、检查单板硬件转发资源数量