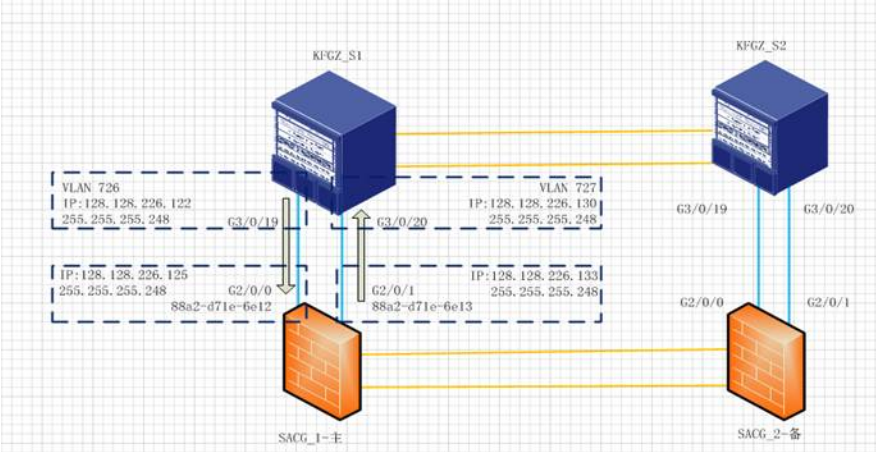
# S12508由于PBR引流配置不当导致网关不可达案例

俞彦飞 2016-01-16 发表

**组网说明：**




组网说明：两台12508通过聚合口1互联，开启VRRP协议，分别做内网网关（KFGZ_S1为主）和防火墙网关（KFGZ_S2为主），上行连接两台华为防火墙，构成口字型接口，华为防火墙配置主备模式，正常情况下仅主设备工作；

**问题描述：**

现网S12508通过配置PBR，匹配内网/24位网段地址，下一跳为SACG_1防火墙的G2/0/0口IP，以实现引流；作为PBR策略后，客户发现现网业务均正常，但内网终端无法和网关互ping，内网网段：128.128.176.0/24，网关IP：128.128.176.254；

```
acl number 3000
 rule 1 permit ip source 128.128.164.0 0.0.1.255
 rule 2 permit ip source 128.128.166.0 0.0.1.255
 rule 3 permit ip source 128.128.177.0 0.0.0.255
 rule 4 permit ip source 128.128.176.0 0.0.0.255
 rule 5 permit ip source 128.128.178.0 0.0.0.255
 rule 6 permit ip source 128.128.186.0 0.0.1.255
 rule 7 permit ip source 128.128.188.0 0.0.1.255

policy-based-route sacg permit node 10
  if-match acl 3000
  apply ip-address next-hop 128.128.226.124 track 1
```

1、测试PC的IP地址是128.128.176.45，VLAN260是网关，VRRP主是在KFGZ_S1上，从KFGZ_S1上的ARP学习情况来看是学习在两台125的互联口上，所以终端PC去ping网关，流量过KFGZ_S2，然后送到KFGZ_S1上去做三层处理：

```
  <KFGZ_S1>dis arp 128.128.176.45
    Type: S-Static    D-Dynamic     A-Authorized    M-Multiport
   IP Address       MAC Address     VLAN ID   Interface          Aging Type
   128.128.176.45   2c41-389e-3377   260       BAGG1              16    D

  Interface Vlan-interface260
   VRID          : 33             Adver Timer : 3
   Admin Status  : Up             State       : Master
   Config Pri    : 20             Running Pri : 20
   Preempt Mode  : Yes            Delay Time  : 0
   Auth Type     : None
   Virtual IP    : 128.128.176.254
   Virtual MAC   : 0000-5e00-0121
   Master IP     : 128.128.176.252
```

2、因为PC的源地址是128.128.176.45，会匹配ACL3000的rule4做PBR动作，而PBR的下一跳是128.128.226.124，即报文会从GE3/0/19送给FW：

```
      128.128.226.124  0000-5e00-0102  726      GE3/0/19            N/A   D
```

interface Vlan-interface726
ip address 128.128.226.122 255.255.255.248
vrrp vrid 6 virtual-ip 128.128.226.121
vrrp vrid 6 priority 10
vrrp vrid 6 timer advertise 5

acl number 3000
rule 4 permit ip source 128.128.176.0 0.0.0.255

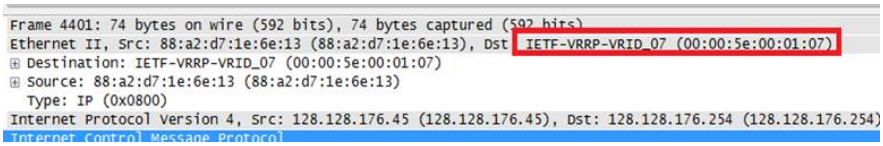
policy-based-route sacg permit node 10
if-match acl 3000
apply ip-address next-hop 128.128.226.124 track 1

3、按照125上的抓包，以及FW的抓包分析，ICMP的报文会从3/0/20送回给KFGZ_S1这台125，VLAN是727：

#
interface GigabitEthernet3/0/20
port link-mode bridge
description connect_SACG1_G2/0/1
port access vlan 727

4、但是VLAN726和727的VRRP主是在KFGZ_S2上，而不是在KFGZ_S1上，FW回应过来的报文在KFGZ_S1只是做二层转发，匹配目的MAC 0000-5e00-0107送给KFGZ_S2做三层转发：

0000-5e00-0107 727 Learned Bridge-Aggregation1 AGING
128.128.226.129 0000-5e00-0107 727 BAGG1 15 D



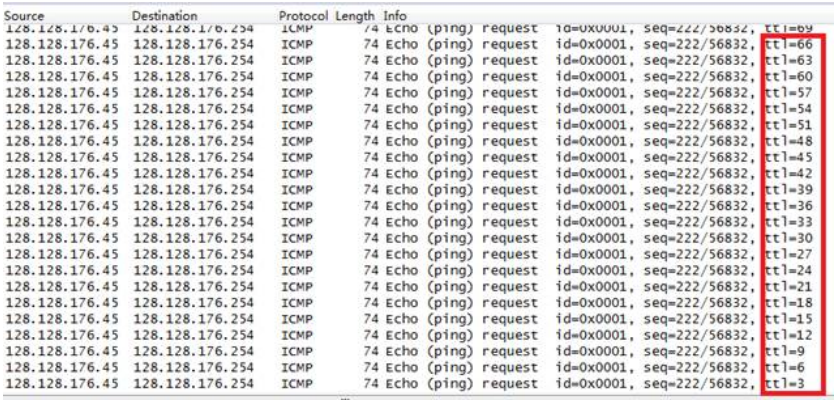

Interface Vlan-interface726
VRID : 6 Adver Timer : 5
Admin Status : Up State : Backup
Config Pri : 10 Running Pri : 10
Preempt Mode : Yes Delay Time : 0
Become Master : 13400ms left
Auth Type : None
Virtual IP : 128.128.226.121
Master IP : 128.128.226.123

Interface Vlan-interface727
VRID : 7 Adver Timer : 5
Admin Status : Up State : Backup
Config Pri : 10 Running Pri : 10
Preempt Mode : Yes Delay Time : 0
Become Master : 11500ms left
Auth Type : None
Virtual IP : 128.128.226.129
Master IP : 128.128.226.131

5、PC回应的报文的目的地址是KFGZ_S1上VLAN260的是地址 128.128.176.252；在KFGZ_S2上到128.128.176.252直连，即会从VLAN727三层转发到VLAN260又送回给KFGZ_S1；但是KFGZ_S1上的VLAN260是做了PBR，又回到了PC回应报文的处理流程：

#
interface Vlan-interface260
ip address 128.128.176.252 255.255.255.0
vrrp vrid 33 virtual-ip 128.128.176.254
……
ip policy-based-route sacg

这样报文TTL在KFGZ_S1、KFGZ_S2、和FW各减1，抓取到的报文的TTL差值是3；
直至减完丢弃：



| Source | Destination | Protocol | Length | Info |
|---|---|---|---|---|
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=69 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=66 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=63 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=60 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=57 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=54 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=51 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=48 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=45 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=42 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=39 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=36 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=33 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=30 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=27 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=24 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=21 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=18 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=15 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=12 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=9 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=6 |
| 128.128.176.45 | 128.128.176.254 | ICMP | 74 | Echo (ping) request id=0x0001, seq=222/56832, ttl=3 |

针对部署PBR没法PING通情况，可以在PBR中识别出这类要送给125自己的报文，PERMIT通过，避免PBR做动作；