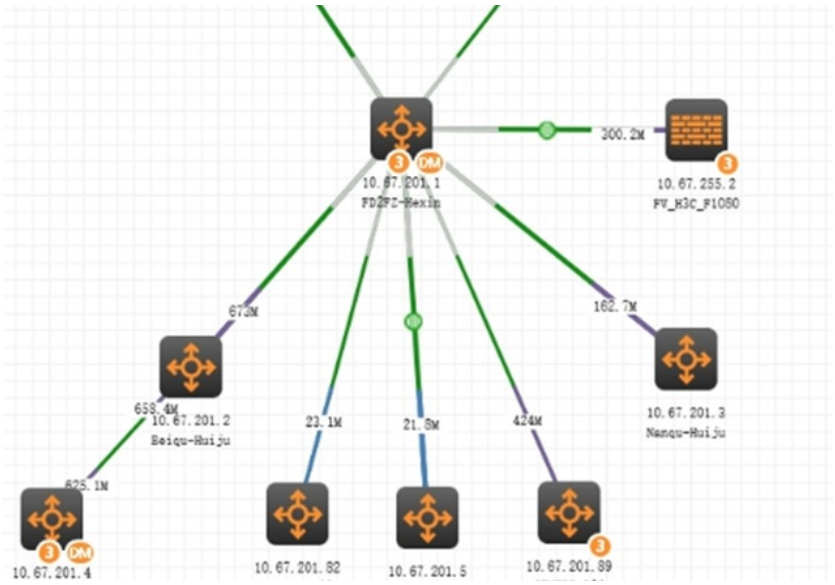
某局点S10504 ICMP报文不匹配策略路由的经验案例

ACL 叶靖 2020-10-28 发表

组网及说明



现场拓扑如上， 现场防火墙旁挂在核心交换机10504上，现场通过配置策略路由，想让内网终端访问交换机本身的流量也到防火墙上过一遍。

问题描述

现场流量路径大概如下，

1、终端流量上来到达interface Vlan-interface12接口，匹配ip policy-based-route to_fw的node20下一跳转到防火墙

```
interface Vlan-interface12
 description TO-DHCP-client
 ip address 10.67.96.1 255.255.240.0
 ip policy-based-route to_fw

policy-based-route to_fw permit node 10
 if-match acl 3900
#
policy-based-route to_fw permit node 20
 if-match acl 3901
 apply next-hop 10.67.255.2
 apply default-next-hop 10.67.255.2
```

2、防火墙配置默认路由，指到交换机，流量回到交换机，

```
#
interface Route-Aggregation99//交换机与防火墙互联接口
 ip address 10.67.255.1 255.255.255.252
 ip policy-based-route fw_in
#
```

3、交换机回包时，匹配ip local policy-based-route local_to_fw，再指到防火墙上

```
policy-based-route local_to_fw permit node 20
 if-match acl 3801
 apply next-hop 10.67.255.2
 apply default-next-hop 10.67.255.2
```

4、防火墙转回交换机上之后，交换机查路由转发给终端。

但是现场测试，终端ping不通交换机，但是SSH交换机是能通的。

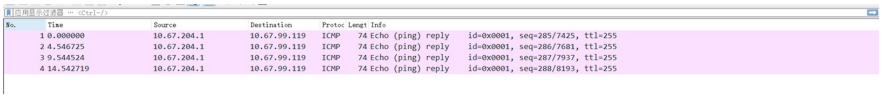
```
   连接特定的 DNS 后缀 . . . . . . . :
   本地链接 IPv6 地址. . . . . . . . : fe80::2cab:8da6:4ceb:55c9%14
   IPv4 地址 . . . . . . . . . . . . : 10.67.99.119
   子网掩码 . . . . . . . . . . . . : 255.255.240.0
   默认网关. . . . . . . . . . . . . : 10.67.96.1

C:\Users\43684>ping 10.67.204.1

正在 Ping 10.67.204.1 具有 32 字节的数据:
请求超时。
请求超时。
请求超时。
请求超时。

10.67.204.1 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 0, 丢失 = 4 (100% 丢失),

C:\Users\43684>ping 10.67.96.1

正在 Ping 10.67.96.1 具有 32 字节的数据:
请求超时。
请求超时。
请求超时。
请求超时。

10.67.96.1 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 0, 丢失 = 4 (100% 丢失),

C:\Users\43684>_
```

且在核心上面取消本地策略路由之后就可以正常ping通。

```
C:\Users\43684>ping 10.67.96.1

正在 Ping 10.67.96.1 具有 32 字节的数据:
来自 10.67.96.1 的回复: 字节=32 时间=8ms TTL=255
来自 10.67.96.1 的回复: 字节=32 时间=3ms TTL=255
来自 10.67.96.1 的回复: 字节=32 时间=5ms TTL=255
来自 10.67.96.1 的回复: 字节=32 时间=3ms TTL=255

10.67.96.1 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
    最短 = 3ms, 最长 = 8ms, 平均 = 4ms

C:\Users\43684>_
 rule 35 permit ip source 10.67.212.0 0.0.1.255
 rule 40 permit ip source 10.67.217.0 0.0.0.255
<FD2FZ-Hexin>sys
System View: return to User View with Ctrl+Z.
[FD2FZ-Hexin]undo ip loc
[FD2FZ-Hexin]undo ip local p
[FD2FZ-Hexin]undo ip local policy-based-route
[FD2FZ-Hexin]
```

过程分析

在终端ping不通交换机时，在防火墙上抓包，发现只有ICMP的回包



No.	Time	Source	Destination	Protoc	Lengt	Info
1	0.000000	10.67.204.1	10.67.99.119	ICMP	74	Echo (ping) reply id=0x0001, seq=285/7425, ttl=255
2	4.546725	10.67.204.1	10.67.99.119	ICMP	74	Echo (ping) reply id=0x0001, seq=286/7681, ttl=255
3	9.544524	10.67.204.1	10.67.99.119	ICMP	74	Echo (ping) reply id=0x0001, seq=287/7937, ttl=255
4	14.542719	10.67.204.1	10.67.99.119	ICMP	74	Echo (ping) reply id=0x0001, seq=288/8193, ttl=255

SSH交换机的时候，在防火墙上能看到完整的会话信息。即SSH报文正确匹配网关接口PBR转到了防火墙，且交换机回包也正常转给了防火墙，路径正确。

现在怀疑ICMP报文没有匹配网关接口上的PBR，直接处理，但是回包匹配Local PBR，发给了防火墙，防火墙直接丢弃，导致不通。

后续经研发确认：**到达本地的ICMP报文确实不会匹配接口的PBR，会直接上送交换机CPU处理。**

1． SSH和ICMP都是通过acl上送cpu的。

2．是否被pbr匹配上，需要确认协议对应的底层acl是否有取消三层转发字段L3Switch Cancel L3Switch NextHopIndex 0x4001，icmp有该字段，不会被pbr抓走，而ssh无该字段，会被PBR抓走。

Ssh和Telnet规则

[ZLYY_WK_2F_IDC_Core_SW-probe]debug qacl show chassis 1 slot 0 c 0 verbose 0 sysidx 65

========
Acl-Type RX IPv4 Middle, Stage IFP, Pipe 0, Global, Installed, Active
Prio Mjr/Sub 524/18, Group 1 [1], Slice/Idx 8/42, Entry 46, Double: 6186/6698
Rule Match --------
    Ports: 0x000000000000001fffe; 0x60000000000007fffff
    Lookup: VLAN ID valid[y], STP forwarding, 0x1c, 0x1c
    IP protocol: tcp
    IP Type: Any IPv4 packet
    L4 Dst Port: 23, 0xffff
    Dest Port: CPU
    DropBit: 0x0, Mask : 0x1
    L3 Dest Class id: 0x20
Actions --------
    CAR cir 0x200, cbs 0x800, pir 0x200, pbs 0x800, mode srTCM color blind,Bytes
    Account mode packets, green and non-green
    Change CPU pkt COS 27
    Red Deny
    Red_Copy_to_cpu : No
    Yel Deny
    Yel_Copy_to_cpu : No
MatchedName:65, TELNET/SSH
Accounting: Hi 0, LO 0

Imcp规则

[ZLYY_WK_2F_IDC_Core_SW-probe]debug qacl show chassis 1 slot 0 c 0 verbose 0 sysidx 44

========
Acl-Type RX IPv4 Middle, Stage IFP, Pipe 0, Global, Installed, Active
Prio Mjr/Sub 524/18, Group 1 [1], Slice/Idx 8/39, Entry 38, Double: 6183/6695
Rule Match --------
    Ports: 0x000000000000001fffe; 0x60000000000007fffff
    Lookup: VLAN ID valid[y], STP forwarding, 0x1c, 0x1c
    IP protocol: icmp
    IP Type: Any IPv4 packet
    Dest Port: CPU
    DropBit: 0x0, Mask : 0x1
    SysmRule Index : 44
    L3 Dest Class id: 0x20
    My Station Hit
Actions --------
    CAR cir 0x200, cbs 0x800, pir 0x200, pbs 0x800, mode srTCM color blind,Bytes
    Account mode packets, green and non-green
    L3Switch Cancel L3Switch NextHopIndex 0x4001
    Change CPU pkt COS 22
    Red Deny
    Red_Copy_to_cpu : No
    Yel Deny
    Yel_Copy_to_cpu : No

解决方法

现场想要终端能够正常ping通交换机的话，可以让访问本地的ICMP流量不经过防火墙，直接通过查找路由进行转发即可。可以修改 ip local policy-based-route local_to_fw，添加如下配置：
policy-based-route local_to_fw permit node 10
 if-match acl 3800

acl advanced 3800
rule 0 permit icmp
添加如上配置之后，终端访问交换机的ICMP流量到达交换机后，直接上送CPU进行处理，之后回包时，直接匹配本地PBR（ip local policy-based-route local_to_fw ）的node10节点，通过查找路由转发，直接回复给终端，从而内网终端可以ping通交换机。