S7500E/S10500产品PBR不生效问题分析案例

策略路由 王晨 2014-07-28 发表

S7500E/S10500产品PBR不生效问题分析案例

一、 组网：

策略路由是一种依据用户制定的策略进行路由转发的机制。策略路由可以基于报文的源地址等信息灵活地控制报文的发送。报文到达设备后，系统首先根据策略路由转发，若没有配置策略路由或配置了策略路由但找不到匹配的表项时，再根据路由表来转发报文。

（1） 策略路由与路由的区别

我们知道，路由是用来指导报文转发时的选路，它是目的地址和下一跳的映射关系：想去什么地方，请走什么路；或者说要去哪里，请先去哪里。路由关心的是报文的目的地址。如果说路由像一个指路牌，那么策略路由更像一座立交桥：火车请走火车道、汽车请走汽车道、行人请走人行道，等等。策略路由更多的是关心报文的特征，它可以灵活的为不同特征的报文选择不同的线路来发送。这些特征可以是报文的源地址，目的地址，四层端口号，报文协议类型等等，为满足一定条件的报文制定一定的转发策略，这就是策略路由的功能。

（2） 策略路由与路由策略区别

路由策略（Routing Policy）是为了改变网络流量所经过的途径而修改路由信息的技术，主要通过改变路由属性（包括可达性）来实现。路由策略的主要功能为控制路由的发布与接收，管理引入的路由并设置路由的属性。如果说策略路由像一座立交桥一样用以指导报文的转发，路由策略就像是交通局——用以对路由进行管理。两者完全是不同层面的概念，不可混淆，更不可相互比较。

S7500E/S10500系列交换机支持两种策略路由配置方式：PBR和QOS。本文只讲述PBR的使用。

PBR方式（policy-based-route）通过ACL制定匹配规则，支持对报文的下一跳，优先级及缺省下一跳进行设置，目前只支持IPv4单播策略路由。

二、 问题描述：

用户配置PBR的方式下发策略路由后，发现配置没有生效，数据流依然按照路由进行转发，配置的PBR没有生效。

三、 过程分析：

交换机的PBR均是通过软件层配置后下发在硬件底层来实现，因此PBR出现问题，可能和配置、硬件资源、报文类型、软件版本等有关，下面我们来从几个方面分3.1 硬件资源不足导致策略路由无法生效

在部署PBR之前，应对整个设备的ACL资源进行评估，PBR主要占用ACL IFP资源，如果PBR中总的匹配ACL的RULE条数是N，要在M个VLAN上下发，则需要ACL的资源为N*M，由于各种单板的匹配能力不同，有的情况下需要ACL资源数为N*M*2。如下举例，当IFP的资源要大于N*M*2才能保证PBR下发时不出现硬件资源不足的告警，出现资源不足就会导致PBR的部分策略无法生效。

```
dis acl resource slot 6
Interface:
  GE6/0/1 to GE6/0/24
  ------------------------------------------------------------------------
  Type        Total     Reserved   Configured Remaining  Usage
  ------------------------------------------------------------------------
  VFP ACL     1024      0          0          1024       0%
  IFP ACL     4096      1024       0          3072       25%
    IFP Meter   2048      512        0          1536       25%
  IFP Counter 2048      512        0          1536       25%
  EFP ACL     512       0          0          512        0%
  EFP Meter   256       0          0          256        0%
    EFP Counter 512       0          0          512        0%
```

1）PBR是基于VLAN三层接口配置，通过全局下发到各个单板中，所以即使有的单板端口上没有允许下发PBR的VLAN通过，也会在这块单板上下发PBR策略，因为单板与内部互联的端口上默认包含设备已经创建的VLAN。

2）S75E的1CB主控板匹配能力较弱，在下发PBR时易出现资源不足的告警。由于1CB是S75E的主控板，没有出接口，虽然出现告警但可以不必管它。

3）S75E的EA单板匹配能力较弱，不建议使用PBR，如果做PBR，不建议在EA单板上做PBR的VLAN接入。

4）S75E的SA单板ACL资源过少，不建议做PBR，在操作手册中已经明确SA不支持策略路由。

5）PBR是基于VLAN下发，具有全局的概念，因此整个设备上ACL资源规格最小的单板在下发PBR时，最容易出现硬件资源不足。

3．2 在一个VLAN三层接口上只能下发一个PBR，若重复下发，只有最后一个PBR生效。

3．3 一个PBR可以包含多个Node的，Node数值越小优先级越高，优先匹配。即使两个Node匹配相同的ACL内容，由于Node不同，仍然会下发两份。

3．4 避免相同的ACL在PBR不同的NODE中下发，这样会多占资源，下发在后面NODE中的ACL实际没有意义。

```
acl number 3000
rule 1 deny ip source 10.0.0.0 0.255.255.255 destination 10.0.0.0 0.255.255.255 counting
rule 2 deny ip source 10.0.0.0 0.255.255.255 destination 192.168.0.0 0.0.255.255
rule 3 deny ip source 10.0.0.0 0.255.255.255 destination 172.16.0.0 0.0.255.255
rule 4 deny icmp source 10.0.0.0 0.255.255.255 destination 10.0.0.0 0.255.255.255
rule 5 deny icmp source 10.0.0.0 0.255.255.255 destination 192.168.0.0 0.0.255.255
rule 6 deny icmp source 10.0.0.0 0.255.255.255 destination 172.16.0.0 0.0.255.255
rule 7 deny ip source 10.0.0.0 0.255.255.255 destination 172.27.0.0 0.0.255.255
rule 8 deny icmp source 10.0.0.0 0.255.255.255 destination 172.27.0.0 0.0.255.255
rule 10 permit ip source 10.0.0.0 0.255.255.255
acl number 3002
rule 1 permit ip source 10.0.0.0 0.255.255.255 destination 10.0.0.0 0.255.255.255
rule 2 permit ip source 10.0.0.0 0.255.255.255 destination 192.168.0.0 0.0.255.255
rule 3 permit ip source 10.0.0.0 0.255.255.255 destination 172.16.0.0 0.0.255.255
rule 4 permit icmp source 10.0.0.0 0.255.255.255 destination 10.0.0.0 0.255.255.255
```

ACL 3000和ACL 3002分别与同一PBR的不同NODE关联，由于RULE 1和RULE 2完全一致，可以只留一个，简化配置。

3.5 能否合并的IP地址，通过掩码表示，通过合并来节省资源，例如：

```
rule 0 permit ip source 106.120.176.36 0 logging
rule 1 permit ip source 106.120.176.37 0 logging
rule 2 permit ip source 106.120.176.38 0 logging
rule 3 permit ip source 106.120.176.39 0 logging
```

如上的规格可以通过rule 0 permit ip source 106.120.176.36 0.0.0.3 logging一条替代，简化配置。

3.6 PBR基于VLAN三层接口下发，只对三层转发流量生效，这里说的三层转发流量并不仅仅是三层单播转发报文会生效。对于组播报文也会生效。因为组播报文本身没有二层还是三层转发的概念，在S7500E/S10500产品的路由策略中，只要组播报文可以匹配上策略路由规则，就会按照规则执行相应的动作。并不是所有的组播都可以配置，下面几种情况的组播报文就不会匹配上PBR规格

```
MULTICAST:
IPV4:
1. good checksum
2. no conditions below:
SIP or DIP is zero
SIP or DIP is of the form 127.xx.xx.xx (loop back address) SIP is class D (src multicast) DIP is NOT a MC address Lower 23(32) bits of mac DA and ipv4(ipv6) dip does not  match
IPv4 packets with options
3. DIP is non broadcast IP(0xffffffff)
```

3.7 PBR基于绑定VPN的端口下发，可以生效。但如果VLAN三层接口是连接的PE或P设备，报文已经是标签报文，PBR功能无法生效。

四、解决办法：

策略路由问题，大多可以通过配置解决，如果配置无法规避的，建议咨询中低端交换产品线。

策略路由的维护，有以下注意事项：

PBR功能一旦在有业务的设备上已经部署完毕，就尽量不要随意改变，否则操作可能会引起业务中断，必须要进行PBR调整时，同样要注意ACL资源的使用情况，并应掌握以下的使用技巧。

4.1 避免在PBR功能已经生效的情况下，取消if-match acl的命令和apply的重定向动作。比如：

```
policy-based-route SNAT permit node 8
  if-match acl 3700
  apply ip-address next-hop 10.21.255.20
```

如果去掉if-match acl 3700，NODE 8的规则就变成了所有流量都重定向到10.21.255.20，如果再去掉apply ip-address next-hop 10.21.255.20，NODE 8的规则就变成了所有流量都按照正常路由转发。

4．2 避免在PBR功能已经生效的情况下，任意增减ACL RULE，除非你对新增的ACL规则非常有把握。必须要增加或减少，应对原有的流量匹配顺序进行评估，尽量在业务低峰期进行操作。

4．3 避免在PBR功能已经生效的情况下，增加或删除策略路由的NODE。必须要增加或删除的，操作要尽量要快，否则在下发NODE的过程中流量转发路径就会变化。比如在原有PBR策略中增加NODE 8：

```
    policy-based-route SNAT permit node 8
      if-match acl 3700
      apply ip-address next-hop 10.21.255.20
```

就算是通过拷贝转贴的方式，上面三条命令执行的瞬间，流量也是变化的，会引起短时流量路径变化。