

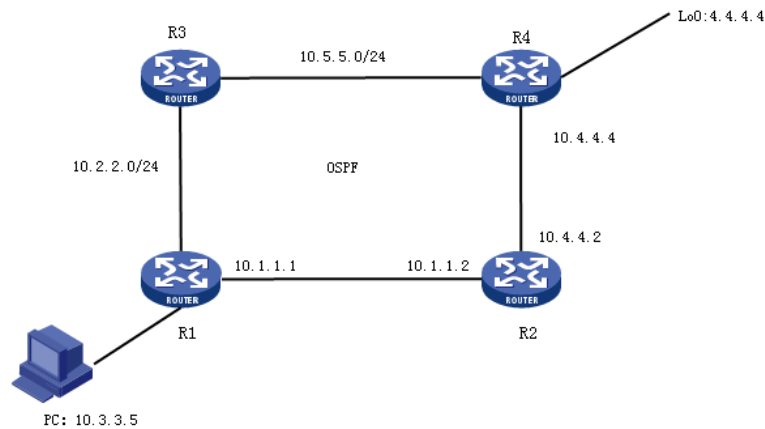
知 OSPF、PBR组网中路由环路问题分析

策略路由

朱尘场

2015-12-12 发表

故障组网拓扑如下：



组网情况

R1、R2、R3、R4所有接口加入OSPF域,包括R4上联出口4.4.4.4和R1下联PC终端所在接口
所有链路开销一致

组网需求

针对R1上来的部分流量做PBR策略, 匹配策略的流量走R2 (PC-R1-R2-R4)

R2上联R4的链路断了之后, 将流量再扔回给R1 (PC-R1-R2-R1-R3-R4), 即流量走另外一条备份链
路

故障时发现产生路由环路 (即R2上联R4的链路断了)

```
traceroute to 4.4.4.4 (4.4.4.4), 30 hops at most, 40 bytes each packet
 1  10.3.3.1 (10.3.3.1)  2.000 ms  1.000 ms  1.000 ms
 2  10.1.1.2 (10.1.1.2)  3.000 ms  2.000 ms  1.000 ms
 3  10.1.1.1 (10.1.1.1)  2.000 ms  2.000 ms  4.000 ms
 4  10.1.1.2 (10.1.1.2)  3.000 ms  2.000 ms  2.000 ms
 5  10.1.1.1 (10.1.1.1)  2.000 ms  2.000 ms  2.000 ms
 6  10.1.1.2 (10.1.1.2)  3.000 ms  2.000 ms  5.000 ms
 7  10.1.1.1 (10.1.1.1)  6.000 ms  3.000 ms  *
 8  10.1.1.2 (10.1.1.2)  5.000 ms  6.000 ms  3.000 ms
```

查看R1上的快速转发表, 发现随着PC不停地ping测试, R1的快速转发表越来越多

```
[R1]display ip fast-forwarding cache
Total number of fast-forwarding entries: 42
SIP      I      SPort DIP      DPort Pro Input_If Output_If Flg
4.4.4.4  1      33440 10.3.3.5  32963 17  GE0/0    GE0/2    1
4.4.4.4  1      33445 10.3.3.5  32963 17  GE0/0    GE0/2    1
10.3.3.5  1      32963 4.4.4.4   33447 17  GE0/2    GE0/0    1
4.4.4.4  1      33454 10.3.3.5  32963 17  GE0/0    GE0/2    1
10.3.3.5  1      32963 4.4.4.4   33445 17  GE0/2    GE0/0    1
10.3.3.5  1      32963 4.4.4.4   33446 17  GE0/2    GE0/0    1
4.4.4.4  1      33441 10.3.3.5  32963 17  GE0/0    GE0/2    1
4.4.4.4  1      33450 10.3.3.5  32963 17  GE0/0    GE0/2    1
4.4.4.4  1      33439 10.3.3.5  32963 17  GE0/0    GE0/2    1
10.1.1.2  1      0      10.3.3.5  2816  1  GE0/0    GE0/2    1
10.1.1.1  1      0      10.3.3.5  2816  1  InLoop0  GE0/2    1
10.3.3.5  1      32963 4.4.4.4   33449 17  GE0/2    GE0/0    1
4.4.4.4  1      33448 10.3.3.5  32963 17  GE0/0    GE0/2    1
10.3.3.5  1      32963 4.4.4.4   33450 17  GE0/2    GE0/0    1
4.4.4.4  1      33453 10.3.3.5  32963 17  GE0/0    GE0/2    1
10.3.3.5  1      32963 4.4.4.4   33439 17  GE0/2    GE0/0    1
10.3.3.5  1      32963 4.4.4.4   33448 17  GE0/2    GE0/0    1
10.3.3.5  1      32963 4.4.4.4   33437 17  GE0/2    GE0/0    1
```

1、查看R1和R2的路由表均正常, 按照正常需求转发路径: PC-R1-R2-R1-R3-R4

```
[R1]display ospf routing
```

OSPF Process 1 with Router ID 10.3.3.1 Routing Table					
Routing for network					
Destination	Cost	Type	NextHop	AdvRouter	Area
10.2.2.0/24	1	Transit	0.0.0.0	10.5.5.3	0.0.0.0
100.100.100.100/32	2	Stub	10.2.2.3	100.100.100.100	0.0.0.0
10.3.3.0/24	1	Stub	0.0.0.0	10.3.3.1	0.0.0.0
4.4.4.4/32	2	Stub	10.2.2.3	100.100.100.100	0.0.0.0
10.5.5.0/24	2	Transit	10.2.2.3	100.100.100.100	0.0.0.0
10.1.1.0/24	1	Transit	0.0.0.0	10.4.4.2	0.0.0.0

```
[R2]display ospf routing
```

OSPF Process 1 with Router ID 10.4.4.2 Routing Table					
Routing for network					
Destination	Cost	Type	NextHop	AdvRouter	Area
10.2.2.0/24	2	Transit	10.1.1.1	10.5.5.3	0.0.0.0
100.100.100.100/32	3	Stub	10.1.1.1	100.100.100.100	0.0.0.0
10.3.3.0/24	2	Stub	10.1.1.1	10.3.3.1	0.0.0.0
4.4.4.4/32	3	Stub	10.1.1.1	100.100.100.100	0.0.0.0
10.5.5.0/24	3	Transit	10.1.1.1	100.100.100.100	0.0.0.0
10.1.1.0/24	1	Transit	0.0.0.0	10.4.4.2	0.0.0.0

2、路由转发优先级：快速转发表>策略路由>路由协议

3、快速转发负载均衡 关闭快速转发负载均衡功能后，将会根据入接口的不同对五元组标识的数据流再次做出区分，即将入接口作为区分数据流的另一特征标识；开启快速转发负载均衡功能后，当一条数据流从不同入接口上来进行转发时，不再根据入接口不同区分数据流，根据五元组标识一条数据流。

缺省情况下，快速转发负载均衡功能处于开启状态

4、根据如上信息，很明显路由环路是快速转发负载均衡造成的

1、解决方案一

V7：关闭快速转发负载均衡功能

V5：关闭快速转发

效果截图：

```
<H3C>tracert 4.4.4.4
traceroute to 4.4.4.4 (4.4.4.4), 30 hops at most, 40 bytes
 1 10.3.3.1 (10.3.3.1) 1.087 ms 0.489 ms 0.660 ms
 2 10.1.1.2 (10.1.1.2) 1.986 ms 1.133 ms 0.929 ms
 3 10.1.1.1 (10.1.1.1) 2.375 ms 1.286 ms 1.697 ms
 4 10.2.2.3 (10.2.2.3) 3.708 ms 1.610 ms 1.782 ms
 5 4.4.4.4 (4.4.4.4) 2.143 ms 4.201 ms 5.466 ms
```

方案一关闭快速转发负载均衡或者快速转发后，影响路由器的转发效率，同时路由转发路径也不是最优路径，最优路径应该是PC-R1-R3-R4

2、优化解决方案二

PBR+路由迭代（即R1配置PBR的时候，应用的next-hop地址配置为下一跳的下一跳地址，即本案例中的10.4.4.4地址）

3、优化解决方案三

PBR+TRACK+NQA（通过NQA的方式，去探测下一跳的下一跳是否存活，探测路径R1-R2-R4；如果R2和R4之间的链路断了，使PBR的下一跳失效从来去查找路由表转发）

```
nqa entry admin h3c
type icmp-echo
destination ip 10.4.4.4
frequency 100
next-hop 10.1.1.2
reaction 1 checked-element probe-fail threshold-type consecutive 5 action-type trigger-only
#
nqa schedule admin h3c start-time now lifetime forever
```

```
track 1 nqa entry admin h3c reaction 1
```

```
[H3C]display ip policy-based-route
Policy name: h3c
node 1 permit:
if-match acl 3001
apply next-hop 10.1.1.2 track 1
```

1、V7设备上关闭快速转发负载均衡即可，V5设备需要关闭快速转发

2、尽量使用优化方法，实现最佳路由转发路径，同时又不影响路由器转发效率