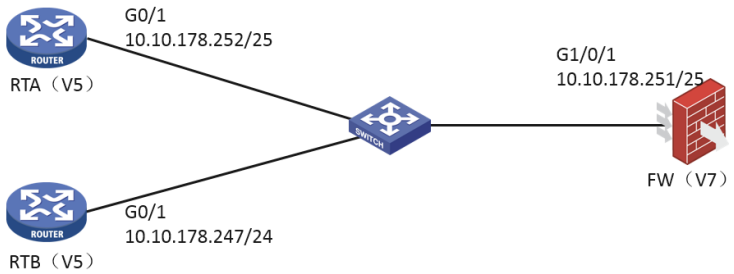
某局点BGD+BFD联动，BFD会话DOWN经验案例

BFD 郑标 2019-03-16 发表

组网及说明



如上图所示，两台V5的路由器和一台V7的防火墙通过一台二层交换机互联，彼此建立IBGP邻居；通过BGP和BFD的联动，以实现链路的快速收敛。

问题描述

相关配置完成后，发现RTA和防火墙FW之间的BFD session状态一直是down的状态，其它各IBGP邻居间的BFD session均是正常的up状态，如下显示：

```
<DG-769DCC-F1005>dis bfd se verbose
 Total Session Num: 2     Up Session Num: 1     Init Mode: Active

 IPv4 Session Working Under Ctrl Mode:
       Local Discr: 129                 Remote Discr: 12
        Source IP: 10.10.178.251      Destination IP: 10.10.178.247
     Session State: Up                      Interface: N/A
      Min Tx Inter: 400ms                 Act Tx Inter: 400ms
      Min Rx Inter: 400ms                 Detect Inter: 2000ms
          Rx Count: 17242                    Tx Count: 376063
      Connect Type: Indirect            Running Up for: 01:54:56
         Hold Time: 1664ms                   Auth mode: None
       Detect Mode: Async                        Slot: 1
          Protocol: BGP
           Version: 1
         Diag Info: No Diagnostic

       Local Discr: 130                 Remote Discr: 0
        Source IP: 10.10.178.251      Destination IP: 10.10.178.252
     Session State: Down                    Interface: N/A
      Min Tx Inter: 400ms                 Act Tx Inter: 1000ms
      Min Rx Inter: 400ms                 Detect Inter: 5000ms
          Rx Count: 0                        Tx Count: 363224
      Connect Type: Indirect            Running Up for: 00:00:00
         Hold Time: 0ms                      Auth mode: None
       Detect Mode: Async                        Slot: 1
          Protocol: BGP
           Version: 1
         Diag Info: No Diagnostic
```

过程分析

首先，RTA ping防火墙互联地址可以通，路由可达，彼此IBGP邻居也都是established的；

然后，仔细整理了设备侧相关配置，如下：

FW：

```
bgp 64573
 router-id 10.10.178.251
 peer 10.10.178.247 as-number 64573
 peer 10.10.178.247 connect-interface GigabitEthernet1/0/1
 peer 10.10.178.247 bfd
 peer 10.10.178.247 password cipher $c$3$LleCbqG4kg1dvrJYOddkW2yuJElqCtgW6WMw
 peer 10.10.178.252 as-number 64573
 peer 10.10.178.252 connect-interface GigabitEthernet1/0/1
 peer 10.10.178.252 bfd
 peer 10.10.178.252 password cipher $c$3$9hcbCdgmwEBDNFa+mLgcGH/D2xraCmfzMX3O
```

RTA：

```
bgp 64573
 router-id 10.10.178.252
 undo synchronization
 peer 10.10.178.247 as-number 64573
 peer 10.10.178.251 as-number 64573
 peer 10.10.178.247 password cipher $c$3$E0Djv5GohGqts7xEspxmHOki0Oa7xo6TGxJK
 peer 10.10.178.247 next-hop-local
 peer 10.10.178.247 connect-interface GigabitEthernet0/1
 peer 10.10.178.247 bfd
 peer 10.10.178.251 password cipher $c$3$OemVzu7SaRyRyIHyuCWoMSzBQIyODlE60twG
```

peer 10.10.178.251 next-hop-local
peer 10.10.178.251 connect-interface GigabitEthernet0/1
peer 10.10.178.251 bfd

RTB:

bgp 64573
router-id 10.10.178.247
undo synchronization
peer 10.10.178.251 as-number 64573
peer 10.10.178.252 as-number 64573
peer 10.10.178.251 password cipher c3$24rwMOOagmkI5pmdwWjLQCazofrvFxNQ5Ukk
peer 10.10.178.251 next-hop-local
peer 10.10.178.251 connect-interface GigabitEthernet0/1
peer 10.10.178.251 bfd
peer 10.10.178.252 password cipher c3$EPGG2WlJyMfwyfi6mJWw8FDLsyd+XLem9MDW
peer 10.10.178.252 next-hop-local
peer 10.10.178.252 connect-interface GigabitEthernet0/1
peer 10.10.178.252 bfd

通过核对设备侧配置并查阅相关命令手册发现，V7设备在使用peer bfd命令时，后面是建议跟上**multi-hop** 或者**single-hop**参数，若在未指定该参数的情况下，是采用BFD**多跳方式**检测本地路由器和指定IBGP对等体之间的链路，而本地路由器和BGP对等体采用的BFD检测方式（单跳或多跳）必须相同，否则无法建立BFD会话。对于现场组网来讲，属于单跳的情况。

到此时，故障原因有了眉目，建议现场在防火墙侧peer bfd后面加上**single-hop**参数后，RTA和防火墙间的BFD session恢复up状态。

但是，疑问又来了，同样的配置和组网情况下，为什么防火墙和RTB间BGP BFD联动没有指定**single-hop**参数，之间的BFD session也是正常up的状态，于是建议现场在防火墙侧抓包反馈如下：

防火墙收到的报文：

```
   11 1.599848   10.10.178.247      10.10.178.251      BFD Control       66 Diag: No Diagnostic, State: Up, Flags: 0x08
Frame 11: 66 bytes on wire (528 bits), 66 bytes captured (528 bits)
Ethernet II, Src: 9c:06:1b:99:18:a8 (9c:06:1b:99:18:a8), Dst: 88:df:9e:0a:d9:d1 (88:df:9e:0a:d9:d1)
Internet Protocol Version 4, Src: 10.10.178.247 (10.10.178.247), Dst: 10.10.178.251 (10.10.178.251)
   Version: 4
   Header length: 20 bytes
 ⊞ Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00: Not-ECT (Not ECN-Capable Transport))
   Total Length: 52
   Identification: 0x913a (37178)
 ⊞ Flags: 0x00
   Fragment offset: 0
   Time to live: 254
   Protocol: UDP (17)
 ⊞ Header checksum: 0xb0b7 [correct]
   Source: 10.10.178.247 (10.10.178.247)
   Destination: 10.10.178.251 (10.10.178.251)
   [Source GeoIP: Unknown]
   [Destination GeoIP: Unknown]
User Datagram Protocol, Src Port: 49168 (49168), Dst Port: bfd-control (3784)
   Source port: 49168 (49168)
   Destination port: bfd-control (3784)
```

防火墙发出的报文：

```
   12 1.692694   10.10.178.251      10.10.178.247      BFD Control       66 Diag: No Diagnostic, State: Up, Flags: 0x08
Frame 12: 66 bytes on wire (528 bits), 66 bytes captured (528 bits)
Ethernet II, Src: 88:df:9e:0a:d9:d1 (88:df:9e:0a:d9:d1), Dst: 9c:06:1b:99:15:68 (9c:06:1b:99:15:68)
Internet Protocol Version 4, Src: 10.10.178.251 (10.10.178.251), Dst: 10.10.178.247 (10.10.178.247)
   Version: 4
   Header length: 20 bytes
 ⊞ Differentiated Services Field: 0xe0 (DSCP 0x38: Class Selector 7; ECN: 0x00: Not-ECT (Not ECN-Capable Transport))
   Total Length: 52
   Identification: 0xa867 (43111)
 ⊞ Flags: 0x00
   Fragment offset: 0
   Time to live: 64
   Protocol: UDP (17)
 ⊞ Header checksum: 0x576b [correct]
   Source: 10.10.178.251 (10.10.178.251)
   Destination: 10.10.178.247 (10.10.178.247)
   [Source GeoIP: Unknown]
   [Destination GeoIP: Unknown]
User Datagram Protocol, Src Port: 49285 (49285), Dst Port: bfd-multi-ctl (4784)
   Source port: 49285 (49285)
   Destination port: bfd-multi-ctl (4784)
```

V7设备判断是否是直连的条件是端口号为3784并且ttl为255，现场抓包显示3784端口号的报文ttl为254，因此V7设备当做非直连处理，所以可以V7侧可以正常建立。如果V5侧收到端口号为4784、ttl为64的报文当作直连处理，就可以在V5侧也建立。

对于防火墙侧为什么收到报文的ttl值是254而不是255，怀疑是报文路径有问题，于是进一步查看设备侧的路由表信息确认。

RTA接口地址　10.10.178.252/25

RTB接口地址　10.10.178.247/**24**

防火墙接口地址　10.10.178.251/25

RTB路由表:

10.10.176.128/25　BGP　255 0　　10.10.178.251　GE0/1

10.10.177.0/24　　Static 60　0　　10.10.178.253　GE0/1

10.10.178.0/24　　Direct 0　0　　10.10.178.247　GE0/1

10.10.178.128/25　OSPF　10　400　　10.10.178.252　GE0/1（优先）

防火墙路由表:

10.10.178.0/24　BGP　255 0　　10.10.178.247　GE1/0/1

10.10.178.128/25　Direct　0　0　　10.10.178.251　GE1/0/1

RTA路由表:

10.10.178.128/25 Direct 0 0 10.10.178.252 GE0/1

由路由表可以看出，由于接口地址掩码长度的不同，RTB去往防火墙的路由会根据最长匹配原则，优选ospf学来的下一跳为RTA接口地址的路由，然后从RTA再匹配直连路由到达防火墙，导致报文根据路由转发经过了两跳（多跳），也就解释了上述现象。

至此，故障问题水落石出。建议现场修改了RTB的接口地址掩码为25后，复现了前面RTA和防火墙BFD 会话down的故障现象，然后在V7防火墙侧peer bfd后加上**single-hop**参数，问题解决。

解决方法

1、V7设备侧peer bfd后加上**single-hop**参数，缺省是**multi-hop**方式；

2、规范互联接口地址掩码，保证路由选路得当。