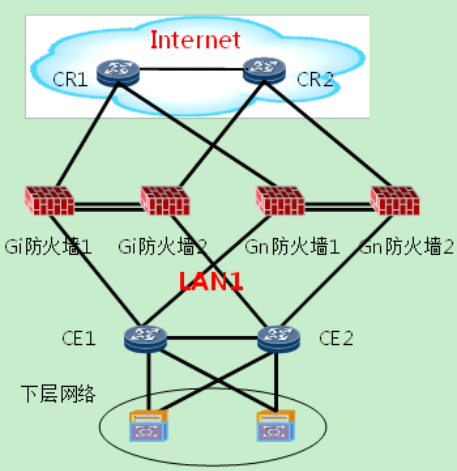
CR16K设备多出口负载分担不均衡问题

陈丹威 2015-11-03 发表

两台CR16018设备CE1和CE2作为下联网络的网关，通过两条链路上行到防火墙设备。由于客户的业务流量不断增加，两条链路的上行带宽无法满足客户的业务需求，需要新增三条链路来做负载分担，达到五条并行链路上行。

具体组网如图所示：



新增三条上行的等价路由后，业务流量应该能够基本平均的分配到五条上行的链路上。但是在实际运行中，流量只能通过其中的两条链路进行转发，其余的三条链路上并没有任何的业务流量，导致扩容的链路浪费，并未解决现有的两条链路上行带宽不够的问题。

通过配置不同的VPN，将下层网络的流量分别送到不同的防火墙设备上去。

```
#
ip vpn-instance Gn
 route-distinguisher 100:1
 vpn-target 100:1 export-extcommunity
 vpn-target 100:1 import-extcommunity
#
ip vpn-instance Gi
 route-distinguisher 200:1
 vpn-target 200:1 export-extcommunity
 vpn-target 200:1 import-extcommunity
#
interface Ten-GigabitEthernet7/0/5
 port link-mode route
ip binding vpn-instance Gi
 ipv6 address 2409:8099:8420::AF9:65C5/126
 ip address 10.249.101.197 255.255.255.252
 ospf cost 1000
 ospf timer hello 1
 ospf timer dead 4
 ospf network-type p2p
#
interface Ten-GigabitEthernet8/0/5
 port link-mode route
ip binding vpn-instance Gi
 ipv6 address 2409:8099:8420::AF9:65B9/126
 ip address 10.249.101.185 255.255.255.252
ospf cost 1000
 ospf timer hello 1
 ospf timer dead 4
 ospf network-type p2p
#
ospf 2 router-id 10.249.101.129 vpn-instance Gi
 import-route direct type 1
 maximum load-balancing 6
 vpn-instance-capability simple
```

```
 area 0.0.0.0
  network 10.249.101.176 0.0.0.3
  network 10.249.101.128 0.0.0.15
  network 10.249.101.184 0.0.0.3
  network 10.249.101.196 0.0.0.3
#
```

查看设备的OSPF路由进程2，可以看到设备的OSPF路由表如下：

```
[CR16018]display  ospf  2  peer
OSPF Process 2 with Router ID 10.249.101.129
                    Neighbor Brief Information
 Area: 0.0.0.0
Router ID       Address          Pri  Dead-Time  Interface       State
10.0.0.3        10.249.101.186   1      4          XGE8/0/5        Full/ -
10.0.0.3        10.249.101.198   1      4          XGE7/0/5        Full/ -
10.0.0.3        10.249.104.38    1      4          XGE5/0/8        Full/ -
10.0.0.3        10.249.104.46    1      4          XGE6/0/8        Full/ -
10.0.0.3        10.249.104.54    1      4          XGE8/0/8        Full/ -
[CR16018]display ip routing-table vpn-instance Gi
Routing Tables: Gi
        Destinations : 34755    Routes : 34759
 Destination/Mask    Proto  Pre  Cost         NextHop         Interface
 0.0.0.0/0           O_ASE  150  1001         10.249.101.186  XGE8/0/5
                     O_ASE  150  1001         10.249.101.198  XGE7/0/5
                     O_ASE  150  1001         10.249.104.38   XGE5/0/8
                     O_ASE  150  1001         10.249.104.46   XGE6/0/8
                     O_ASE  150  1001         10.249.104.54   XGE8/0/8
```

通过查看路由表确认，形成了五条等价的OSPF路由条目，并且在VPN路由表中都已经生效，应该都可以作为正常路径参与转发。

但是通过查看接口流量发现，接口5/0/8、6/0/8、8/0/8只有入方向的流量，却没有出方向的。实际结果却是，路由表中显示的五条等价路由，只有其中的两条参与了转发，而其中的三条只是可以看到路由条目，并未参与流量转发。

```
[CR16018-ospf-2]display interface Ten-GigabitEthernet6/0/8
Ten-GigabitEthernet6/0/8 current state: UP
Line protocol current state: UP
Description: TO:[SZHFW204BEr] - GE-3/0/0 for Gi traffic to GiFW-link 3
The Maximum Transmit Unit is 1500
Link delay is 1(sec)
Internet Address is 10.249.104.45/30 Primary
IP Packet Frame Type: PKTFMT_ETHNT_2,  Hardware Address: 0cda-4140-cf44
IPv6 Packet Frame Type: PKTFMT_ETHNT_2,  Hardware Address: 0cda-4140-cf44
Media type is optical fiber, Port hardware type is 10G_BASE_LR_XFP
Ethernet port mode: LAN
Loopback is not set
10Gbps-speed mode, full-duplex mode
Last 300 seconds input:  23850 packets/sec 22209010 bytes/sec  2%
Last 300 seconds output: 0 packets/sec 1080 bytes/sec  0%
Input  (total): 130300267 packets, 120313392363 bytes
        2 broadcasts, 17272 multicasts, - pauses
Input  (normal): 130300267 packets, 120313392363 bytes
        2 broadcasts, 17272 multicasts, 0 pauses
Input: 0 input errors, 0 runts, 0 giants, 0 throttles
        0 CRC, 0 frame, 0 overruns, - aborts
        - ignored, - parity errors
Output  (total): 22709 packets, 5627666 bytes
        3 broadcasts, 15434 multicasts, - pauses
Output  (normal): 22709 packets, 5627666 bytes
        3 broadcasts, 15434 multicasts, 0 pauses
Output: 0 output errors, - underruns, - buffer failures
        0 aborts, 0 deferred, 0 collisions, 0 late collisions
        - lost carrier, - no carrier
Peak value of input: 37836170 bytes/sec, at 2015-05-27 00:53:04
Peak value of output: 3580 bytes/sec, at 2015-05-27 00:52:03
```

经确认设备的款型发现，现场使用的是SPC类单板。按照规格，SPC类单板支持0.0.0.0/0的缺省路由只能hash到2个接口上。

解决方案：可以使用明细路由，或是手工拆分缺省路由分布到各个接口上。
参考如下：
ip route-static vpn-instance Gi 0.0.0.0 128.0.0.0 10.249.101.186
ip route-static vpn-instance Gi 0.0.0.0 128.0.0.0 10.249.101.198
ip route-static vpn-instance Gi 0.0.0.0 128.0.0.0 10.249.104.38
ip route-static vpn-instance Gi 0.0.0.0 128.0.0.0 10.249.104.46
ip route-static vpn-instance Gi 0.0.0.0 128.0.0.0 10.249.104.54
将上述缺省路由改为以下明细路由：
ip route-static vpn-instance Gi 128.0.0.0 128.0.0.0 10.249.101.186
ip route-static vpn-instance Gi 128.0.0.0 128.0.0.0 10.249.101.198
ip route-static vpn-instance Gi 128.0.0.0 128.0.0.0 10.249.104.38
ip route-static vpn-instance Gi 128.0.0.0 128.0.0.0 10.249.104.46
ip route-static vpn-instance Gi 128.0.0.0 128.0.0.0 10.249.104.54
这样流量就可以通过五条链路做负载分担进行转发。