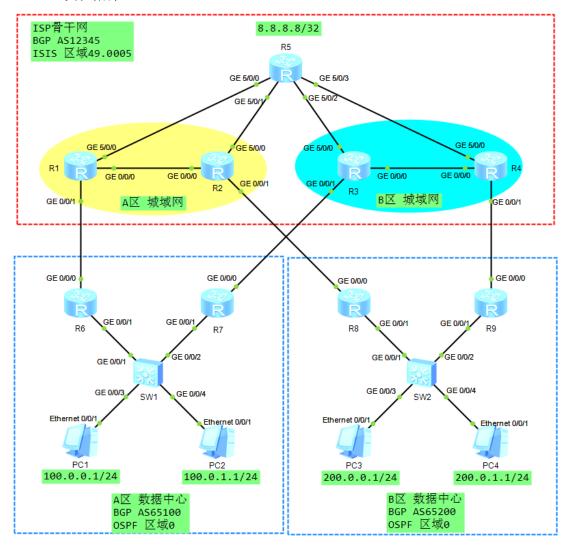
【HCIP 实验 14】BGP 综合选路

一、 实验拓扑



二、实验需求及解法

本实验模拟某市ISP骨干网与两个数据中心互联的网络,完成以下配置需求:

- 1.预配置包括:
- 1.1 所有设备互联IP已配置,且所有设备都有Loopback0地址。

```
1.2 SW1与SW2已创建vlan、划分vlan、并创建vlanif。
1.3 PC已配置IP和网关。
2.在所有设备的系统视图下配置全局router id为Loopback0地址。
使得OSPF和BGP协议自动选择该RID,而无需另行配置。
R1-R9
Router id 5.x.x.x
SW1
Router id 5.11.11.11
SW2
Router id 5.22.22.22
3.内部网关协议IGP
3.1 ISP骨干网中运行ISIS进程1,满足以下需求:
3.1.1 ISIS区域49.0005
3.1.2 系统ID规划如下:
R1:0000.0000.0001
R2:0000.0000.0002
R3:0000.0000.0003
R4:0000.0000.0004
R5:0000.0000.0005
3.1.3 所有路由器均为Level-2
3.1.4 与其他AS互联的接口不激活ISIS。
3.1.5 确保ISP骨干网内部互通。
R1
isis 1
is-level level-2
network-entity 49.0005.0000.0000.0001.00
#
R2:
isis 1
is-level level-2
network-entity 49.0005.0000.0000.0002.00
R3:
isis 1
is-level level-2
network-entity 49.0005.0000.0000.0003.00
R4:
isis 1
```

```
is-level level-2
 network-entity 49.0005.0000.0000.0004.00
#
R1/2/3/4:
interface LoopBack0
 isis enable 1
interface GigabitEthernet0/0/0
 isis enable 1
interface GigabitEthernet5/0/0
 isis enable 1
#
R5:
isis 1
 is-level level-2
 network-entity 49.0005.0000.0000.0005.00
interface LoopBack0
 isis enable 1
interface GigabitEthernet5/0/0
 isis enable 1
interface GigabitEthernet5/0/1
 isis enable 1
interface GigabitEthernet5/0/2
 isis enable 1
interface GigabitEthernet5/0/3
 isis enable 1
[R5]dis ip routing-table protocol isis
```

Destination/Mask	Proto	Pre	Cost	Flags	NextHop	Interface
5.0.12.0/24	ISIS-L2		20	D	5.0.15.1	GigabitEthernet5/0/0
	ISIS-L2	15	20	D	5.0.25.2	GigabitEthernet5/0/1
5.0.34.0/24	ISIS-L2	15	20	D	5.0.35.3	GigabitEthernet5/0/2
	ISIS-L2	15	20	D	5.0.45.4	GigabitEthernet5/0/3
5.1.1.1/32	ISIS-L2	15	10	D	5.0.15.1	GigabitEthernet5/0/0
5.2.2.2/32	ISIS-L2	15	10	D	5.0.25.2	GigabitEthernet5/0/1
5.3.3.3/32	ISIS-L2	15	10	D	5.0.35.3	GigabitEthernet5/0/2
5.4.4.4/32	ISIS-L2	15	10	D	5.0.45.4	GigabitEthernet5/0/3

R1/2/3/4自行查路由表并测试。

```
3.2 两个数据中心分别运行OSPF进程1,满足以下需求:
3.2.1 router-id自动选择,不要手动配置。
3.2.2 全部属于区域0. (两个数据中心OSPF并不互通。)
3.2.3 使用network命令宣告, 通配符使用0.0.0.0
3.2.4 不允许宣告与其他AS互联的接口。
3.2.5 确保两个数据中心内部各自互通。
R6:
ospf 1
area 0.0.0.0
 network 5.0.116.6 0.0.0.0
 network 5.6.6.6 0.0.0.0
#
R7:
ospf 1
area 0.0.0.0
 network 5.0.117.7 0.0.0.0
 network 5.7.7.7 0.0.0.0
#
SW1:
ospf 1
area 0.0.0.0
 network 5.11.11.11 0.0.0.0
 network 5.0.116.11 0.0.0.0
 network 5.0.117.11 0.0.0.0
 network 100.0.0.254 0.0.0.0
 network 100.0.1.254 0.0.0.0
#
R8:
ospf 1
area 0.0.0.0
 network 5.0.228.8 0.0.0.0
 network 5.8.8.8 0.0.0.0
#
R9:
ospf 1
area 0.0.0.0
 network 5.0.229.9 0.0.0.0
 network 5.9.9.9 0.0.0.0
#
SW2:
```

```
ospf 1
 area 0.0.0.0
  network 5.22.22.22 0.0.0.0
  network 5.0.228.22 0.0.0.0
  network 5.0.229.22 0.0.0.0
  network 200.0.0.254 0.0.0.0
  network 200.0.1.254 0.0.0.0
```

在PC上测试如下:

```
FPC1
                                                                                                                                                                  PC3
                                                                         seq=1 ttl=254 time=47 ms
seq=2 ttl=254 time=31 ms
seq=3 ttl=254 time=47 ms
seq=4 ttl=254 time=31 ms
                 5.0.116.6: bytes=32 seq=4 ttl=254 time=31 ms
5.0.116.6: bytes=32 seq=5 ttl=254 time=31 ms
                                                                                                                                                                           - 5.0.228.8 ping statistics ---
5 packet(s) transmitted
5 packet(s) received
0.00% packet loss
round-trip min/avg/max = 31/37/47 ms
       -- 5.0.116.6 ping statistics ---
5 packet(s) transmitted
5 packet(s) received
0.00% packet loss
round-trip min/avg/max = 31/37/47 ms
                                                                                                                                                                         C>ping 5.0.229.9
      C>ping 5.0.117.7
                                                                                                                                                                      Ping 5.0.229.9: 32 data bytes, Press Ctrl_C to brown 5.0.229.9: bytes=32 seq=1 ttl=254 time=31 ms From 5.0.229.9: bytes=32 seq=2 ttl=254 time=31 ms From 5.0.229.9: bytes=32 seq=3 ttl=254 time=46 ms From 5.0.229.9: bytes=32 seq=4 ttl=254 time=31 ms
   Ping 5.0.117.7: 32 data bytes,
```

4.建立BGP邻居

- 4.1 ISP骨干网属于AS12345, R1/2/3/4/5均运行BGP, 满足以下需求:
- 4.1.1 R5为路由反射器 (RR), R1/2/3/4为客户端 (RR-client)。
- 4.1.2 R1/2/3/4分别使用Loopback0与R5建立ibgp邻居关系。
- 4.1.3 R1/2/3/4之间不建立bgp邻居关系。
- 4.1.4 配置必要的next-hop-local命令。

R1/2/3/4:

```
bgp 12345
```

peer 5.5.5.5 as-number 12345

peer 5.5.5.5 connect-interface LoopBack0

peer 5.5.5.5 next-hop-local

#

R5:

bgp 12345

peer 5.1.1.1 as-number 12345

peer 5.1.1.1 connect-interface LoopBack0

peer 5.1.1.1 reflect-client

peer 5.2.2.2 as-number 12345

```
peer 5.2.2.2 connect-interface LoopBack0
 peer 5.2.2.2 reflect-client
 peer 5.3.3.3 as-number 12345
peer 5.3.3.3 connect-interface LoopBack0
 peer 5.3.3.3 reflect-client
peer 5.4.4.4 as-number 12345
 peer 5.4.4.4 connect-interface LoopBack0
peer 5.4.4.4 reflect-client
4.2 A区数据中心属于AS65100, B区数据中心属于AS65200。
两个数据中心都采用了同城异地灾备方案,即双出口连接到不同城域网,可以避免本地城
域网故障导致的网络瘫痪。
数据中心配置BGP,满足以下需求:
4.2.1 使用LoopbackO建立ibgp邻居:
A区数据中心:SW1不运行BGP, R6/7不建立ibgp邻居。
B区数据中心: SW2 - R8 - R9 全互联
4.2.2 配置必要的next-hop-local命令
SW2:
bgp 65200
peer 5.8.8.8 as-number 65200
peer 5.8.8.8 connect-interface LoopBack0
peer 5.9.9.9 as-number 65200
peer 5.9.9.9 connect-interface LoopBack0
R8:
bgp 65200
peer 5.9.9.9 as-number 65200
peer 5.9.9.9 connect-interface LoopBack0
peer 5.9.9.9 next-hop-local
peer 5.22.22.22 as-number 65200
peer 5.22.22.22 connect-interface LoopBack0
peer 5.22.22.22 next-hop-local
#
R9:
bgp 65200
peer 5.8.8.8 as-number 65200
peer 5.8.8.8 connect-interface LoopBack0
peer 5.8.8.8 next-hop-local
peer 5.22.22.22 as-number 65200
 peer 5.22.22.22 connect-interface LoopBack0
```

```
peer 5.22.22.22 next-hop-local
4.2.3 使用物理接口建立以下ebgp邻居:
R6 - R1
R7 - R3
R8 - R2
R9 - R4
R6:
bgp 65100
peer 5.0.16.1 as-number 12345
R1:
bgp 12345
 peer 5.0.16.6 as-number 65100
#
R7:
bgp 65100
peer 5.0.37.3 as-number 12345
R3:
bap 12345
peer 5.0.37.7 as-number 65100
#
R8:
bgp 65200
peer 5.0.28.2 as-number 12345
R2:
bgp 12345
peer 5.0.28.8 as-number 65200
R9:
bgp 65200
peer 5.0.49.4 as-number 12345
R4:
bgp 12345
 peer 5.0.49.9 as-number 65200
5.发布BGP路由
R5使用network命令发布8.8.8.8/32
R5:
bgp 12345
  network 8.8.8.8 32
数据中心按照以下要求发布路由:
```

```
5.1 SW2使用network命令发布200.0.0/24和200.0.1.0/24两条路由
SW2:
bgp 65200
 network 200.0.0.0
 network 200.0.1.0
5.2 R7使用network命令发布100.0.0.0/24和100.0.1.0/24两条路由
R7:
bgp 65100
 network 100.0.0.0 255.255.255.0
 network 100.0.1.0 255.255.255.0
5.3 R6使用import-route命令发布100.0.0/24和100.0.1.0/24两条路由,使用以下策略:
5.3.1 ip-prefix, 名称为100
index 10 匹配 100.0.0.0/24
index 20 匹配 100.0.1.0/24
5.3.2 route-policy, 名称为OSPFtoBGP (注意大小写), node 10, 调用prefix。
5.3.3 将OSPF引入BGP时调用策略。
R6:
ip ip-prefix 100 index 10 permit 100.0.0.0 24
ip ip-prefix 100 index 20 permit 100.0.1.0 24
route-policy OSPFtoBGP permit node 10
if-match ip-prefix 100
#
bgp 65100
 import-route ospf 1 route-policy OSPFtoBGP
5.4 在R6和R7上将BGP引入OSPF, 确认SW1学习到路由。
R6/7:
ospf 1
import-route bgp
[SW1]dis ospf routing
```

Routing for ASES	5				
Destination	Cost	Type	Tag	NextHop	AdvRouter
8.8.8.8/32	1	Type2	1	5.0.116.6	5.6.6.6
200.0.0.0/24	1	Type2	1	5.0.116.6	5.6.6.6
200.0.1.0/24	1	Type2	1	5.0.116.6	5.6.6.6

6.协议优先级选路

查看SW1上收到的其他AS路由如8.8.8/32,发现下一跳只有一个。

尝试分析原因.

此时R6/R7形成了BGP和OSPF的双点双向引入路由环境。

当R6将BGP引入到OSPF时,除SW1外,R7也会收到OSPF外部路由。

R7 从R3收到EBGP路由 协议优先级为255;

SPOTO 全球 培训 ● 项目 ● 人才

SPOTO IT 人才培训机构 ● IT 人才解决方案 ● CCIE 实验室 ● BOOTCAMP 全真 IT 项目实战 WEB: http://www.SPOTO.net BBS: (网络)http://bbs.SPOTO.net ● 以伙伴关系帮助客户成功,帮助员工成功,帮助合作伙伴成功。 —— 我们共创未来!



从R6收到OSPF外部路由 协议优先级为150。

由于150<255, 所以R7会使用OSPF外部路由。

R7上把BGP引入OSPF时, 8.8.8.8/32不属于BGP路由, 所以R7的引入无效。

只有当R7优选R3的EBGP路由时,才会把8.8.8.8/32引入到OSPF。

所以需要修改R7从R3收到路由的协议优先级,小于150。

R7正常后, R6会有同样的问题, 所以也需要修改。

并在R6/7部署以下策略:

- 6.1 进入BGP进程,使用preference命令修改BGP协议优先级。
- 6.2 EBGP路由协议优先级为50、IBGP和本地路由使用默认值255。

R6/7:

bgp 65100

preference 50 255 255

6.3 确认SW1收到的AS外部路由可以看到R6/7两个下一跳。

[SW1]dis ospf routing

Routing for ASES	5				
Destination	Cost	Type	Tag	NextHop	AdvRouter
8.8.8.8/32	1	Type2	1	5.0.117.7	5.7.7.7
8.8.8.8/32	1	Type2	1	5.0.116.6	5.6.6.6
200.0.0.0/24	1	Type2	1	5.0.117.7	5.7.7.7
200.0.0.0/24	1	Type2	1	5.0.116.6	5.6.6.6
200.0.1.0/24	1	Type2	1	5.0.117.7	5.7.7.7
200.0.1.0/24	1	Type2	1	5.0.116.6	5.6.6.6

7. 权重 (PrefVal) 选路

在R5上查看去往AS65100的路由,发现只有R3一个边界传递了路由,而R1没有发送路由。 尝试分析原因,

<R5>dis bgp routing-table

	5.15 15 9 p 1 5 5 5 1 1 1 3 5 5 5 1 5	. •				
	Network	NextHop	MED	LocPrf	PrefVa	l Path/Ogn
*>	8.8.8.8/32	0.0.0.0	0		0	i
*>i	100.0.0.0/24	5.3.3.3	2	100	0	65100i
*>i	100.0.1.0/24	5.3.3.3	2	100	0	65100i

尝试分析原因:

R1上同时收到R3和R6发来的路由,R3路由起源属性为"i",R6起源属性为"?",故而选择了R3作为最佳下一跳。由于ibgp邻居发来的路由不能再传递给其他ibgp邻居,所以R1没有将路由发送给R5。

为预防此类现象发生, 在所有AS边界路由器上部署以下策略进行优化:

- 7.1 不允许使用route-policy
- 7.2 修改ebgp邻居发来的路由权重(PrefVal)为1.

R1:

bgp 12345

peer 5.0.16.6 preferred-value 1

R2:

bgp 12345

peer 5.0.28.8 preferred-value 1

R3:

bgp 12345

peer 5.0.37.7 preferred-value 1

R4:

bgp 12345

peer 5.0.49.9 preferred-value 1

R6:

bgp 65100

peer 5.0.16.1 preferred-value 1

R7:

bgp 65100

peer 5.0.37.3 preferred-value 1

R8:

bgp 65200

peer 5.0.28.2 preferred-value 1

R9:

bgp 65200

peer 5.0.49.4 preferred-value 1

7.3 确认R5可以同时收到R1和R3发来AS65100的路由。

[R5]dis bgp routing-table

	Network	NextHop	MED	LocPrf	PrefVa	l Path/Ogn
*>	8 8 8 8/32	0 0 0 0	0		0	i
*>i	100.0.0.0/24	5.3.3.3	2	100	0	65100i
* i		5.1.1.1	2	100	0	65100?
*>i	100.0.1.0/24	5.3.3.3	2	100	0	65100i
* i		5.1.1.1	2	100	0	65100?

- 8.本地优先(LocPrf)选路
- 8.1 ISP访问A区数据中心优先走R1。

修改R1默认本地优先(LocPrf)为150。R5上查看路由验证。

R1:

bgp 12345

default local-preference 150

<R5>dis bgp routing-table

	Network	NextHop	MED	LocPrf	PrefVa	l Path/Ogn
*>	8.8.8.8/32	0.0.0.0	0		0	i
*>i	100.0.0/24	5.1.1.1	2	150	Ö	65100?
* i		5.3.3.3	2	100	0	65100i
*>i	100.0.1.0/24	5.1.1.1	2	150	0	65100?
* 1		5.3.3.3	2	100	0	65100i

8.2 B区数据中心访问ISP优先走R9

修改R9默认本地优先(LocPrf)为150。SW2上查看路由验证。

R9:

bgp 65200

default local-preference 150

[SW2]dis bgp routing-table

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i * i	8.8.8.8/32	5.9.9.9		150	0	12345i
*>i	100.0.0.0/24	5.8.8.8 5.9.9.9		100 150	0	12345i 12345 65100?
* i *>i	100.0.1.0/24	5.8.8.8 5.9.9.9		100 150	0	12345 65100? 12345 65100?
* 1		5.8.8.8		100	Ŏ	12345 65100?

8.3 B区数据中心访问A区数据中心优先走R8, R8上部署以下策略:

8.3.1 ip-prefix 名称100

index 10 匹配 100.0.0.0/24

index 20 匹配 100.0.1.0/24

8.3.2 route-policy 名称toSW2

node 10, 引用prefix, 修改本地优先 (LocPrf)为180。

node 100, 允许其他路由。

8.3.3 R8发送路由给SW2时调用策略。

8.3.4 在SW2查看bgp路由验证。

R8:

ip ip-prefix 100 index 10 permit 100.0.0.0 24

ip ip-prefix 100 index 20 permit 100.0.1.0 24

#

route-policy toSW2 permit node 10

if-match ip-prefix 100

apply local-preference 180

#

route-policy toSW2 permit node 100

#

bgp 65200

peer 5.22.22.22 route-policy toSW2 export

#

[SW2]dis bgp routing-table

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i	8.8.8.8/32	5.9.9.9		150	0	12345i
* i		5.8.8.8		100	0	12345i
*>i	100.0.0.0/24	5.8.8.8		180	0	12345 65100?
* i		5.9.9.9		150	0	12345 65100?
*>i	100.0.1.0/24	5.8.8.8		180	0	12345 65100?
* i		5.9.9.9		150	0	12345 65100?

9.MED选路

B区数据中心管理员希望外部AS数据流量从R9进入本AS。

9.1 在R5上查看AS65200的路由,发现选择R2作为最佳路径,流量会从R8进入AS65200。 尝试分析原因。

<R5>dis bgp routing-table

	Network	NextHop	MED	LocPrf	PrefVa	Path/Ogn
*>	8.8.8.8/32	0.0.0.0	0		0	i
*>i	100.0.0.0/24	5.1.1.1	2	150	0	65100?
* i		5.3.3.3	2	100	0	65100i
*>i	100.0.1.0/24	5.1.1.1	2	150	0	65100?
* i		5.3.3.3	2	100	0	65100i
*>i	200.0.0.0	5.2.2.2		100	0	65200i
* i		5.4.4.4		100	0	65200i
*>i	200.0.1.0	5.2.2.2		100	0	65200i
* i		5.4.4.4		100	0	65200i

按照选路13条原则,前11条都完全相同,第12条比较RID小的作为最佳路径。所以R5选择5.2.2.2

9.2 B区数据中心原管理员计划使用MED值影响AS12345优选R4作为最佳路径,使得流量从R9进入本AS。

方案如下:

R8部署策略,增加发送给R2路由的MED值

route-policy toR2 permit node 10

apply cost 200

#

bgp 65200

peer 5.0.28.2 route-policy toR2 export

原管理员还没来得及部署该策略,就被ISP骨干网领导挖走,也没来得及完成工作交接。

B区数据中心新管理员不会使用route-policy, 担心配置错误没有按照原管理员方案执行。

但是当他仔细查看R8的OSPF路由表和BGP路由表发现:

R8使用ospf学习到200.0.0.0/24和200.0.1.0/24, cost为2.

R8使用bgp学习到200.0.0/24和200.0.1.0/24, MED为0.

于是他灵机一动,R8上只新增了两条命令,在不影响其他路由的情况下完成了这个需求。 R8:

bgp 65200

network 200.0.0.0 24

network 200.0.1.0 24

#R8直接宣告OSPF路由,OSPF的cost值会写入MED,发送给R2的路由MED自动变成2。 [R2]dis bgp routing-table

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i	8.8.8.8/32	5.5.5.5	0	100	0	i
*>i	100.0.0.0/24	5.1.1.1	2	150	0	65100?
*>i	100.0.1.0/24	5.1.1.1	2	150	0	65100?
*>	200.0.0.0	5.0.28.8	2		1	65200i
* j		5.4.4.4		100	0	65200i
*>	200.0.1.0	5.0.28.8	2		1	65200i
* i		5.4.4.4		100	0	65200i
[R2]						

R2上看,R8的MED为2,R4MED默认为0,但是由于需求7中修改了权重(PrefVal),所以依旧选择R8。

<R5>dis bgp routing-table

	Network	NextHop	MED	LocPrf	PrefVa	1 Path/Ogn
* >	8.8.8.8/32	0.0.0.0	0		0	i
*>i	100.0.0.0/24	5.1.1.1	2	150	0	65100?
* i		5.3.3.3	2	100	0	65100i
*>i	100.0.1.0/24	5.1.1.1	2	150	0	65100?
* j	<u> </u>	5.3.3.3	2	100	0	65100i
*>i	200.0.0.0	5.4.4.4		100	0	65200i
* j		5.2.2.2	2	100	0	65200i
*>i	200.0.1.0	5.4.4.4		100	0	65200i
* j		5.2.2.2	2	100	0	65200i

R5上看, R2的MED为2, R4的MED为默认0, 所以选择R4。

9.3 原管理员得知新管理员的办法后,表示这个办法看似简单,实则将来更麻烦,尝试分析原因。

新管理的方法必须逐条使用network命令通告路由,将来B区数据中心扩容,IP地址增多会导致更多的network宣告,如果忘记宣告还可能导致流量从R8进入AS65200。

原管理员的方法直接匹配所有路由,新增路由自动修改MED,具备很好的扩展性。