

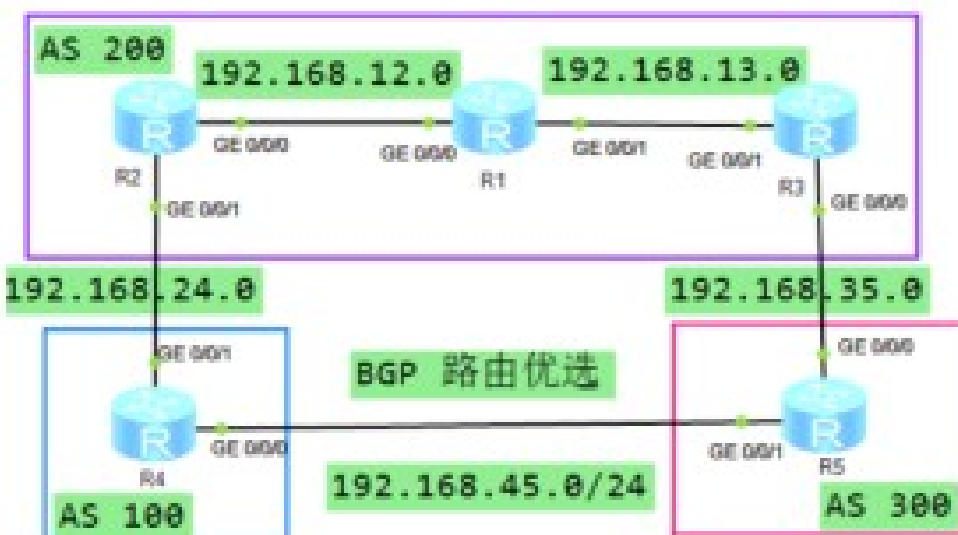
## BGP 路由优选实验

实验拓扑如图所示，所有设备均创建 Loopback0 接口，IP 地址为 192.168.x.x ( x 为设备编号 )，所有设备使用环回口地址作为 Router ID。

AS200 内运行 OSPF，在内部互联接口 ( 不包含连接外部 AS 的接口 )、Loopback 接口上激活 OSPF。

AS 200 内部基于 Loopback0 接口建立 IBGP 对等体关系，AS 之间基于直连接口建立 EBGP 对等体关系。

R4、R5 上存在相同的网段：192.168.45.0/24，通过 import-route 命令将该网段的直连路由注入到 BGP，用于验证 BGP 路由优选规则。



### 基本配置

R1:

```
undo ter mo
```

```
sys
sysname R1
user-interface console 0
idle-timeout 0 0
int loop 0
ip add 1.1.1.1 24
int g0/0/0
ip add 192.168.12.1 24
int g0/0/1
ip add 192.168.13.1 24
q
ospf router-id 1.1.1.1
area 0
net 1.1.1.1 0.0.0.0
net 192.168.12.1 0.0.0.0
net 192.168.13.1 0.0.0.0
q
bgp 200
router-id 1.1.1.1
peer 2.2.2.2 as-n 200
peer 2.2.2.2 con loo0
peer 3.3.3.3 as-n 200
peer 3.3.3.3 con loo0
q
```

```
R2:
undo ter mo
sys
sysname R2
user-interface console 0
```

```
idle-timeout 0 0
int loop 0
ip add 2.2.2.2 24
int g0/0/0
ip add 192.168.12.2 24
int g0/0/1
ip add 192.168.24.2 24
q
ospf router-id 2.2.2.2
area 0
net 2.2.2.2 0.0.0.0
net 192.168.12.2 0.0.0.0
q
bgp 200
router-id 2.2.2.2
peer 1.1.1.1 as-n 200
peer 1.1.1.1 con loo0
peer 3.3.3.3 as-n 200
peer 3.3.3.3 con loo0
peer 192.168.24.4 as-n 100
q
```

R3:

```
undo ter mo
sys
sysname R3
user-interface console 0
idle-timeout 0 0
int loop 0
ip add 3.3.3.3 24
```

```
int g0/0/0
ip add 192.168.35.3 24
int g0/0/1
ip add 192.168.13.3 24
q
ospf router-id 3.3.3.3
area 0
net 3.3.3.3 0.0.0.0
net 192.168.13.3 0.0.0.0
q
bgp 200
router-id 3.3.3.3
peer 1.1.1.1 as-n 200
peer 1.1.1.1 con loo0
peer 2.2.2.2 as-n 200
peer 2.2.2.2 con loo0
peer 192.168.35.5 as-n 300
q
```

```
R4:
undo ter mo
sys
sysname R4
user-interface console 0
idle-timeout 0 0
int loop 0
ip add 4.4.4.4 24
int g0/0/0
ip add 192.168.45.4 24
int g0/0/1
```

```
ip add 192.168.24.4 24
q
bgp 100
router-id 4.4.4.4
peer 192.168.24.2 as-n 200
peer 192.168.45.5 as-n 300
q
acl 2000
rule permit source 192.168.45.0 0.0.0.255
route-policy 10 permit node 10
if-match acl 2000
bgp 100
import-route direct route-policy 10
q
```

R5:

```
undo terminal
sys
sysname R5
user-interface console 0
idle-timeout 0 0
int loop 0
ip add 5.5.5.5 24
int g0/0/0
ip add 192.168.35.5 24
int g0/0/1
ip add 192.168.45.5 24
q
bgp 300
router-id 5.5.5.5
```

```

peer 192.168.35.3 as-n 200
peer 192.168.45.4 as-n 100
q
acl 2000
rule permit source 192.168.45.0 0.0.0.255
route-policy 10 permit node 10
if-match acl 2000
bgp 300
import-route direct route-policy 10
q

```

=====

### 验证 BGP 选路规则

丢弃下一跳不可达的路由

```
[R1]dis bgp routing-table
```

Network	NextHop	MED	LocPrf
PrefVal Path/Ogn			
i 192.168.45.0	192.168.24.4	0	100
0 100?			
i	192.168.35.5	0	100
0 300?			

配置完成后，路由是不可用的，需要改变下一跳

R2 :

```
bgp 200
```

```

peer 1.1.1.1 next-hop-local
peer 3.3.3.3 next-hop-local
q
R3 :
bgp 200
peer 1.1.1.1 next-hop-local
peer 2.2.2.2 next-hop-local

```

```
[R1]dis bgp routing-table
```

```

Total Number of Routes: 2
Network          NextHop    MED      LocPrf
PrefVal Path/Ogn

  * > i  192.168.45.0          2.2.2.2          0
100      0      100?
  * i          3.3.3.3          0
100      0      300?

```

选择 R2 的原因是 router-id 更小

```
[R1]display bgp routing-table 192.168.45.0
```

```

BGP local router ID : 1.1.1.1
Local AS number : 200
Paths: 2 available, 1 best, 1 select
BGP routing table entry information of
192.168.45.0/24:
From: 2.2.2.2 (2.2.2.2)
Route Duration: 00h54m20s

```

Relay IP Nexthop: 192.168.12.2  
Relay IP Out-Interface: GigabitEthernet0/0/0  
Original nexthop: 2.2.2.2  
Qos information : 0x0  
AS-path 100, origin incomplete, MED 0,  
localpref 100, pref-val 0, valid, intern  
al, best, select, active, pre 255, IGP cost 1  
Not advertised to any peer yet

BGP routing table entry information of  
192.168.45.0/24:  
From: 3.3.3.3 (3.3.3.3)  
Route Duration: 00h19m51s  
Relay IP Nexthop: 192.168.13.3  
Relay IP Out-Interface: GigabitEthernet0/0/1  
Original nexthop: 3.3.3.3  
Qos information : 0x0  
AS-path 300, origin incomplete, MED 0,  
localpref 100, pref-val 0, valid, internal, pre  
255, IGP cost 1, not preferred for router ID  
Not advertised to any peer yet

=====

### 修改 Preferred-Value

使用 preferred-value 命令修改 R3 通告的 BGP 路由其 Preferred-V  
alue 为 100 ，优于 R2 通告 BGP 路由的默认 Preferred-Value ，R1  
将会优选 R3 通告的 BGP 路由

R1 :

bgp 200



```
peer 3.3.3.3 preferred-value 100
q
```

```
[R1]dis bgp routing-table
```

```
*>i 192.168.45.0      3.3.3.3      0
100      100      300?
* i      2.2.2.2      0
100      0      100?
```

R3 通告的 BGP 路由拥有更高的 Preferred-Value ( 100 ) ,  
因此 R1 将会优选 R3 通告的 BGP 路由

=====

### 修改 Local\_Preference

R1 : 删除掉首选值的修改

```
bgp 200
und peer 3.3.3.3 preferred-value
q
```

R3:

```
acl 2000
rule permit source 192.168.45.0 0.0.0.255
route-policy 10 permit node 10
if-match acl 2000
apply local-preference 200
route-policy 10 permit node 20
bgp 200
peer 1.1.1.1 route-policy 10 export
q
```

下一跳可达、相同 Preferred-Value 的情况下将会比较 Local\_Preference，R3 通告的 BGP 路由 Local\_Preference 值为 200，高于 R2 通告的 BGP 路由，R1 将会优选 R3 通告的 BGP 路由。

```
*>i 192.168.45.0 3.3.3.3 0
200 0 300?
* i 2.2.2.2 0
100 0 100?
```

=====

优选本地生成的路由

### 手动聚合

R3: 删除掉 ACL 2000，route-policy 10

R3:

```
und acl 2000
```

```
und route-policy 10
```

```
bgp 200
```

```
und peer 1.1.1.1 route-policy 10 export
```

q

为了在 R3 上进行手动聚合，在 R3 上配置两条指向 null0 的静态路由，用于注入到 BGP。

R3 上配置两条静态路由，将静态路由通过 import-route 注入到 BGP，并通过 aggregate 命令进行手动聚合，同时增加关键字 detail-suppressed 抑制明细路由的对外通告。

R3:

```
ip route-static 192.168.45.0 25 NULL 0
```

```
ip route-static 192.168.45.128 25 NULL 0
```

```
bgp 200
```

```
aggregate 192.168.45.0 24 detail-suppressed
```

```
import-route static
q
```

R3 上查看 BGP 路由表存在两条 BGP 路由 10.0.45.0/24 :

本地产生的：静态路由注入到 BGP 中，由手动聚合产生

对等体通告：由对等体 R5 ( 10.0.35.5 ) 通告

在 R3 上这两条路由都不存在 local\_preference、Preferred-Value 值，此时比较路由来源：手动聚合最优，R3 将会优选本地手动聚合产生的 BGP 路由。

```
*> 192.168.45.0      127.0.0.1
0      ?
*
      192.168.35.5      0
0      300?
s> 192.168.45.0/25      0.0.0.0      0
0      ?
s> 192.168.45.128/25  0.0.0.0      0
0      ?
```

```
[R3]dis bgp routing-table 192.168.45.0 24
```

BGP local router ID : 3.3.3.3

Local AS number : 200

Paths: 2 available, 1 best, 1 select

BGP routing table entry information of  
192.168.45.0/24:

Aggregated route.

Route Duration: 00h05m57s

Direct Out-interface: NULL0

Original nexthop: 127.0.0.1

Qos information : 0x0  
AS-path Nil, origin incomplete, pref-val 0,  
valid, local, best, select, active,  
pre 255  
Aggregator: AS 200, Aggregator ID 3.3.3.3,  
**Atomic-aggregate**

### 自动聚合

R3 :  
ip route-static 10.0.45.0 255.255.255.128 null0  
ip route-static 10.0.45.128 255.255.255.128 null0  
bgp 200  
summary automatic  
q

R5:  
int loo1  
ip add 10.0.0.1 8  
bgp 300  
net 10.0.0.0 8  
q

R3 将会优选本地自动聚合产生的 BGP 路由。

```
*> 10.0.0.0 127.0.0.1
0 ?
* 192.168.35.5 0
0 300i
```

R3 : 再做一个手工聚合

bgp 200  
aggregate 10.0.0.0 255.0.0.0 detail-suppressed

q

优选的依旧是本地产生的 BGP 路由,但是可以看到本地产生的 BGP 路由有两条,从该表项无法判断出优选的为手动聚合还是自动聚合产生的 BGP 路由。

```
*> 10.0.0.0 127.0.0.1
0 ?
* 127.0.0.1
0 ?
* 192.168.35.5 0
0 300i
```

```
[R3]dis bgp routing-table 10.0.0.0
```

```
BGP local router ID : 3.3.3.3
Local AS number : 200
Paths: 3 available, 1 best, 1 select
BGP routing table entry information of
10.0.0.0/8:
Aggregated route.
Route Duration: 00h02m34s
Direct Out-interface: NULL0
Original nexthop: 127.0.0.1
Qos information : 0x0
AS-path Nil, origin incomplete, pref-val 0,
valid, local, best, select, active,
pre 255
Aggregator: AS 200, Aggregator ID 3.3.3.3,
Atomic-aggregate
```

R3 上通过 display bgp routing-table 10.0.0.0 查看 BGP 路由 10.0.0.0/8 的详细信息，存在三条有效路由，其中最优的条目由聚合产生，并且存在 Atomic-aggregate 属性，由此可以看出该聚合条目为手动聚合产生的条目。

自动聚合产生的聚合路由并不会携带 Atomic-aggregate 属性。

=====

### 优选 AS\_Path 最短

R3:

```
bgp 200
undo summary automatic
undo aggregate 192.168.45.0 24 detail-suppressed
q
```

R2:

```
ip ip-prefix aspath permit 192.168.45.0 24
route-policy 10 permit node 10
if-match ip-prefix aspath
apply as-path 400 additive
route-policy 10 permit node 20
bgp 200
peer 1.1.1.1 route-policy 10 export
```

R3 通告的 BGP 路由拥有更短的 AS\_Path，在前几条优选规则一致的情况下，R1 优选 R3 通告的 BGP 路由。

*>i	192.168.45.0	3.3.3.3	0	100
0	300?			
* i		2.2.2.2	0	100
0	400	100?		

=====

### 修改 Origin 属性

R2:

```
und route-policy 10
bgp 200
und peer 1.1.1.1 route-policy 10 export
q
```

在 R5 上修改注入路由的方式为 network ,之后在 R1 上再次查看 BGP 路由表。

R5:

```
bgp 300
und import-route direct
net 192.168.45.0 24
q
```

此时 R5 注入的 BGP 路由 192.168.45.0/24 其 Origin 属性为“i”，在前几条优选规则相同情况下，起源类型为“i”的 BGP 路由成为优选路由。

```
*>i  192.168.45.0      3.3.3.3      0
100      0      300i
```

=====

### 优选 MED 最小

R5:

```
bgp 300
und net 192.168.45.0
import-route direct route-policy 10
q
```

R2:

```
ip ip-prefix med permit 192.168.45.0 24
route-policy 10 permit node 10
if-match ip-prefix med
apply cost 20
route-policy 10 permit node 20
bgp 200
peer 1.1.1.1 route-policy 10 export
q
```

R1:

```
bgp 200
compare-different-as-med
q
```

R2 上通过路由策略修改通告给 R1 的 BGP 路由其 MED 属性值。  
默认情况下 BGP 只会对来自同一个 AS 的相同路由比较 MED 值，  
可以通过命令开启来自不同 AS 的相同路由也比较 MED 值。

R4 发布的 BGP 路由 MED 值为 20，R5 发布的 BGP 路由不携带 MED 值（不携带默认为 0），R5 发布的 BGP 路由拥有更小的 MED 值，R1 优选 R5 发布的 BGP 路由。

*>i	192.168.45.0	3.3.3.3	0
100	0	300?	
* i		2.2.2.2	20
100	0	100?	

=====

优选从 EBGp 对等体学来的路由

R2 :

```
undo ip ip-prefix med
```



```
undo route-policy 10
bgp 200
undo peer 1.1.1.1 route-policy 10 export
```

在 R1 上创建一条 192.168.45.0/24 的静态路由（指向 null0），将该条路由发布到 BGP，同时为了保证 R1、R5 通告给 R3 的 BGP 路由 AS\_Path 长度相同，使用路由策略为 R1 通告给 R3 的路由加上 AS\_Path 属性，其值为：500。

R1:

```
ip route-static 192.168.45.0 24 NULL 0
ip ip-prefix e permit 192.168.45.0 24
route-policy 10 permit node 10
if-match ip-prefix e
apply as-path 500 additive
route-policy 10 permit node 20
bgp 200
import-route static
peer 3.3.3.3 route-policy 10 export
```

q

此时比较通告路由的对等体类型，R5 为 EBGP 对等体，R1 为 IBGP 对等体，EBGP 对等体通告的 BGP 路由优于 IBGP 对等体通告的 BGP 路由，R3 优选 R5 通告的 BGP 路由。

```
*> 192.168.45.0 192.168.35.5 0
0 300?
* i 1.1.1.1 0
100 0 500?
```

AS-path 500, origin incomplete, MED 0,  
localpref 100, pref-val 0, valid, internal,  
pre 255, IGP cost 1, not preferred for peer

type

=====

优选 IGP Cost 值最小

R1:

```
und ip route-static 192.168.45.0 24 NULL 0
```

```
undo ip ip-prefix e
```

```
und route-policy 10
```

```
bgp 200
```

```
und peer 3.3.3.3 route-policy 10 export
```

q

当前 7 条优选规则无法比较出优选 BGP 路由时将会比较前往下一跳地址的 IGP cost 值。

R1:

```
int g0/0/0
```

```
ospf cost 9
```

q

```
AS-path 100, origin incomplete, MED 0,  
localpref 100, pref-val 0, valid, intern  
al, pre 255, IGP cost 9, not preferred for IGP  
cost
```

=====

负载分担

通过配置 BGP 负载分担，可以使得设备同时将多条等代价的 BGP 路由加载到路由表，实现流量负载均衡，减少网络拥塞。

R1:

```
int g0/0/0
```

```
undo ospf cost
q
```

AS\_Path 属性完全相同，需要把 R4 改变一下，放在 AS 300 内  
R4：

```
und bgp 100
bgp 300
router-id 4.4.4.4
peer 192.168.45.5 as-n 300
peer 192.168.24.2 as-n 200
acl 2000
rule permit sourc 192.168.45.0 0.0.0.255
route-policy 10 permit node 10
if-match acl 2000
bgp 300
import-route direct route-policy 10
q
```

R2:

```
bgp 200
und peer 192.168.24.4
peer 192.168.24.4 as-n 300
q
```

R5:

```
bgp 300
und peer 192.168.45.4
peer 192.168.45.4 as-n 300
```

R1:

```
bgp 200
```

maximum load-balancing ibgp 2

q

IP 路由表中出现了到达 192.168.45.0/24 的等价路由

BGP 路由表中依旧只有一条最优的路由

dis ip rout

192.168.45.0/24 IBGP 255 0 RD 3.3.3.3

GigabitEthernet0/0/1

IBGP 255 0 RD 2.2.2.2

GigabitEthernet0/0/0

\*>i 192.168.45.0 2.2.2.2 0 100

0 300?

\* i 3.3.3.3 0 100

0 300?

=====

优选 Cluster\_List 最短

只在 R5 上将 192.168.45.0/24 发布到 BGP

配置 R1 为 RR，R3 为 R1 的客户端。

R2、R3 之间基于环回口建立 IBGP 对等体关系

R2 上将收到 R3 通告的 BGP 路由 192.168.45.0/24、R1 反射的 BGP 路由 192.168.45.0/24。

默认配置下，前面介绍的规则无法比较出优选路由，此时将根据 Cluster\_List 进行优选。

R4:

bgp 300

und import-route direct

q

```
R1:
bgp 200
peer 3.3.3.3 reflect-client
reflector cluster-id 1
q
```

```
R2:dis bgp routing-table 192.168.45.0 24
```

```
Original nexthop: 3.3.3.3
AS-path 300, origin incomplete, MED 0,
localpref 100, pref-val 0, valid, internal, pre 255, IGP cost 2, not preferred for
Cluster List
Originator: 3.3.3.3
Cluster list: 0.0.0.1
Not advertised to any peer yet
```

经由 R1 反射的路由不是最优路由，原因也被标出：

not preferred for Cluster List

R3 直接通告给 R2 的 BGP 路由因为没有经过路由反射器，不存在 Cluster\_List 属性，即被认为 Cluster\_List 长度为 0，小于由 R1 反射的 BGP 路由其 Cluster\_List 长度（1），所以 R3 通告的 BGP 路由为优选路由。