BGP路由优选规则详解

朱仕耿

www.huawei.com

Author / Email : Zhushigeng 261992 / zhushigeng@huawei.com

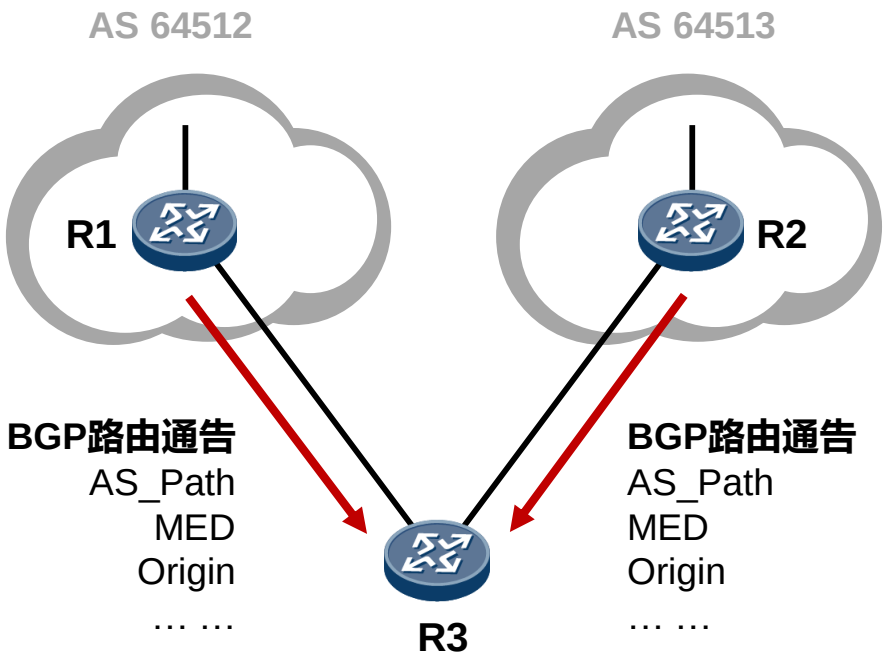
Version 1.2 (2016-04-26)

HUAWEI

课程目标

- 深入理解BGP路由优选规则；
- 掌握利用选路规则进行路由策略部署的方法。

技术背景



- 当BGP设备学习到去往同一个目的网络的多条BGP路由（路径）时，设备将这些路由都装载到BGP路由表，并在这些条目中进行路由优选，最终决策出最优（Best）的路由，将该BGP路由加载到全局路由表中，作为数据转发的依据。
- 当存在多路径时，BGP只会将其选择出来的最优路由通告给其他对等体。
- BGP定义了一系列路由优选规则，从而使得设备能够在多条路由中选择出最优的路由。BGP在选择路由时严格按照先后顺序比较路由的属性，如果通过当前的属性就可以选出最优路由，BGP将不再进行后面的比较。
- BGP的选路规则与BGP路径属性及路由策略息息相关，它们使得BGP拥有了强大的路由操控能力。

BGP路由表示例

[R4] display bgp routing-table

BGP Local router ID is 4.4.4.4

Status codes: * - valid, > - best, d - damped,

h - history, i - internal, s - suppressed, S - Stale

Origin : i - IGP, e - EGP, ? - incomplete

Total Number of Routes: 4

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
***>i**	**100.0.1.0/24**	**3.3.3.3**	**0**	**100**	**10**	**100i**
* i		5.5.5.5	0	100	0	200i

关于100.0.1.0/24这个目的网络，存在两条BGP路由（路径），设备在这两条路由中进行决策，选择出最优的路由，最优的路由将出现“>”符号，它将被加载到设备的全局路由表中。

在BGP路由表中的路由必须首先是可用的（Valid），可用的路由在表项行首存在“*”号，可用意味着该BGP路由的Next_hop是路由可达的，设备在其全局路由表中查询到了去往该Hext_hop地址的路由，即认为该BGP路由可用。如果BGP路由不可用，则不会被优选。

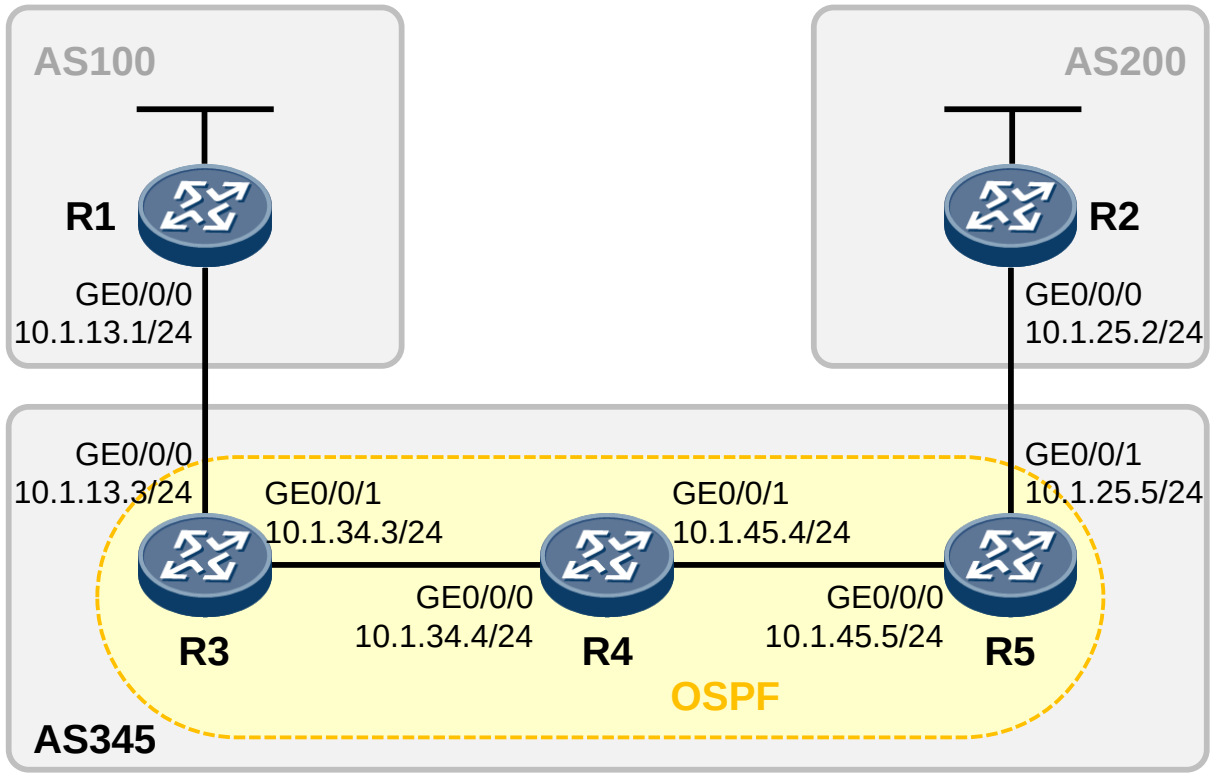
BGP路由优选规则概览

1. 优选具有最大Preferred-Value的路由
2. 优选具有最大Local_Preference的路由
3. 优选起源于本地的路由
4. 优选AS_Path最短的路由
5. Origin（IGP > EGP > Incomplete）
6. 优选MED最小的路由
7. 优选EBGP对等体所通告的路由
8. 优选到Next_Hop的IGP度量值最小的路由
9. BGP路由负载分担
10. 优选Cluster_List 最短的路由
11. 优选Router-ID最小的BGP对等体发来的路由
12. 优选Peer-IP地址最小的对等体发来的路由

BGP路由优选规则也被称为BGP选路规则，不同厂商的设备在BGP选路上存在细微差异，本文档以华为VRP V8版本中实现的选路规则（常用规则）进行讲解。

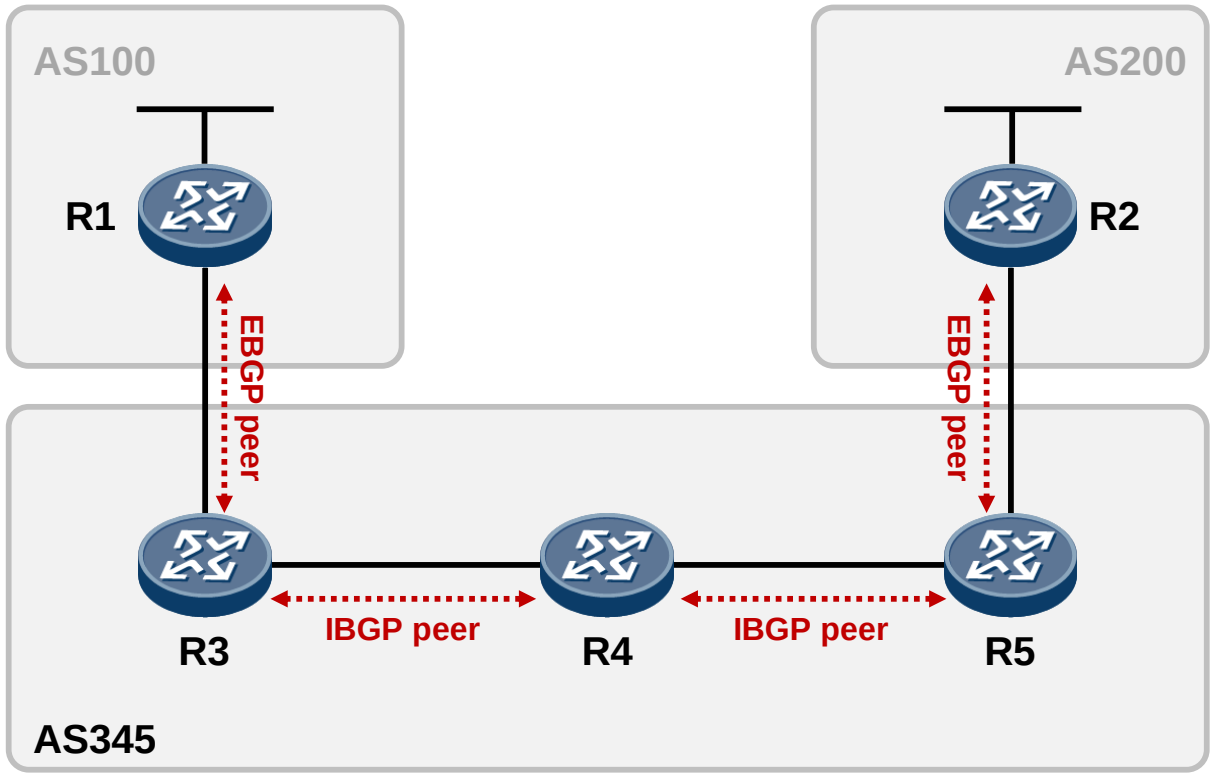
HUAWEI

用于验证选路规则的拓扑



- AS规划、设备互联IP规划如上图所示；所有路由器均创建Loopback0接口，IP为x.x.x.x/32，其中x为设备编号；
- AS345内，R3、R4及R5运行OSPF，在相关接口上激活OSPF（包括Loopback0接口）；
- EBGP对等体关系基于直连接口建立；IBGP对等体关系基于Loopback0接口建立。

用于验证选路规则的拓扑（续）



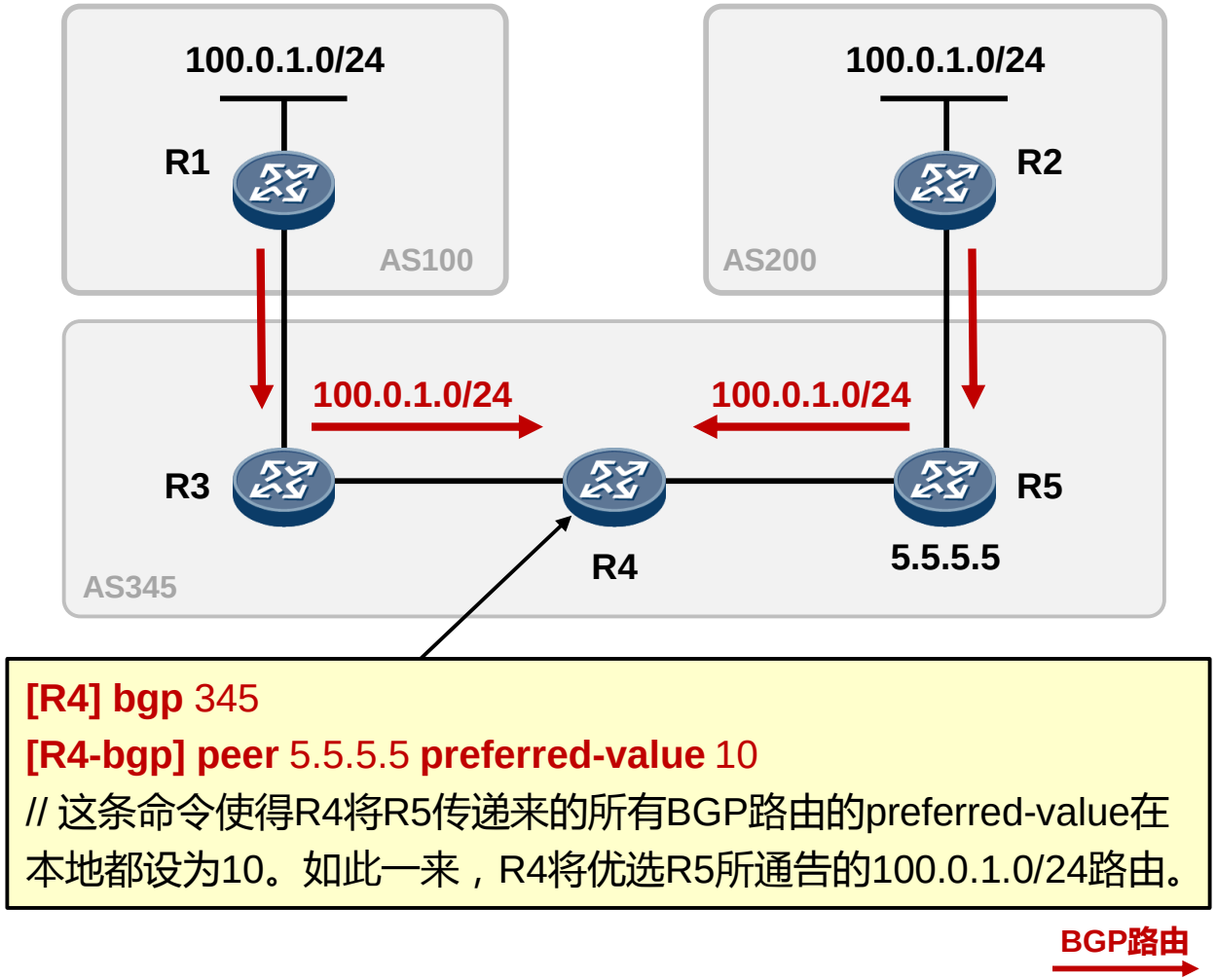
目录

HUAWEI

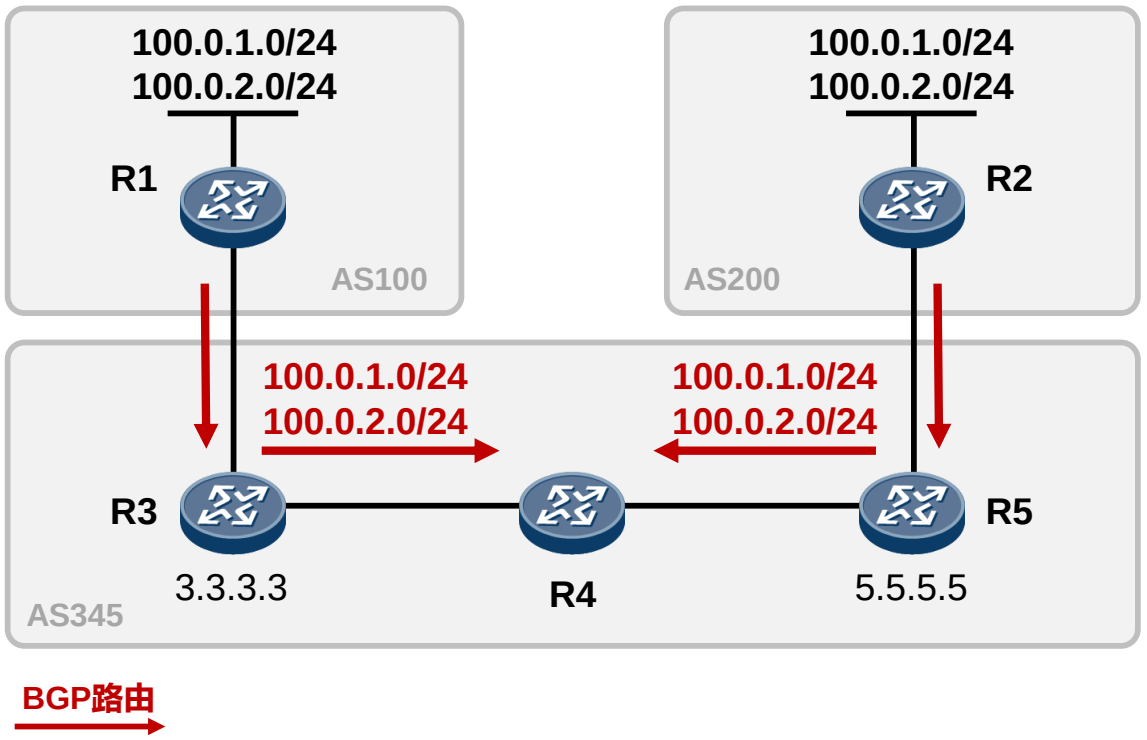
关于Preferred-Value

- 华为私有的路径属性，相当于路由的权重值，取值范围：0~65535；该值越大，则路由越优先。
- Preferred-Value只能在路由器本地配置，而且只影响本设备的路由优选。该属性不会传播给任何BGP对等体。
- 路由器本地始发的BGP路由默认的Preferred-Value为0，从其他BGP对等体学习到的路由默认Preferred-Value也为0。

修改从特定对等体收到的“所有路由”的Preferred-Value



使用route-policy修改Preferred-Value



R4的关键配置如下：

```
ip ip-prefix 1 permit 100.0.1.0 24
ip ip-prefix 2 permit 100.0.2.0 24

route-policy RP1 permit node 10
  if-match ip-prefix 1
  apply preferred-value 10
route-policy RP1 permit node 20

route-policy RP2 permit node 10
  if-match ip-prefix 2
  apply preferred-value 10
route-policy RP2 permit node 20

bgp 345
  peer 3.3.3.3 route-policy RP1 import
  peer 5.5.5.5 route-policy RP2 import
```

在R4上部署路由策略，通过Preferred-Value值的调控，使得它优选R3所通告的100.0.1.0/24路由，而优选R5所通告的100.0.2.0/24路由。

使用route-policy修改Preferred-Value（续）

[R4]display bgp routing-table

BGP Local router ID is 4.4.4.4

Status codes: * - valid, > - best, d - damped,

h - history, i - internal, s - suppressed, S - Stale

Origin : i - IGP, e - EGP, ? - incomplete

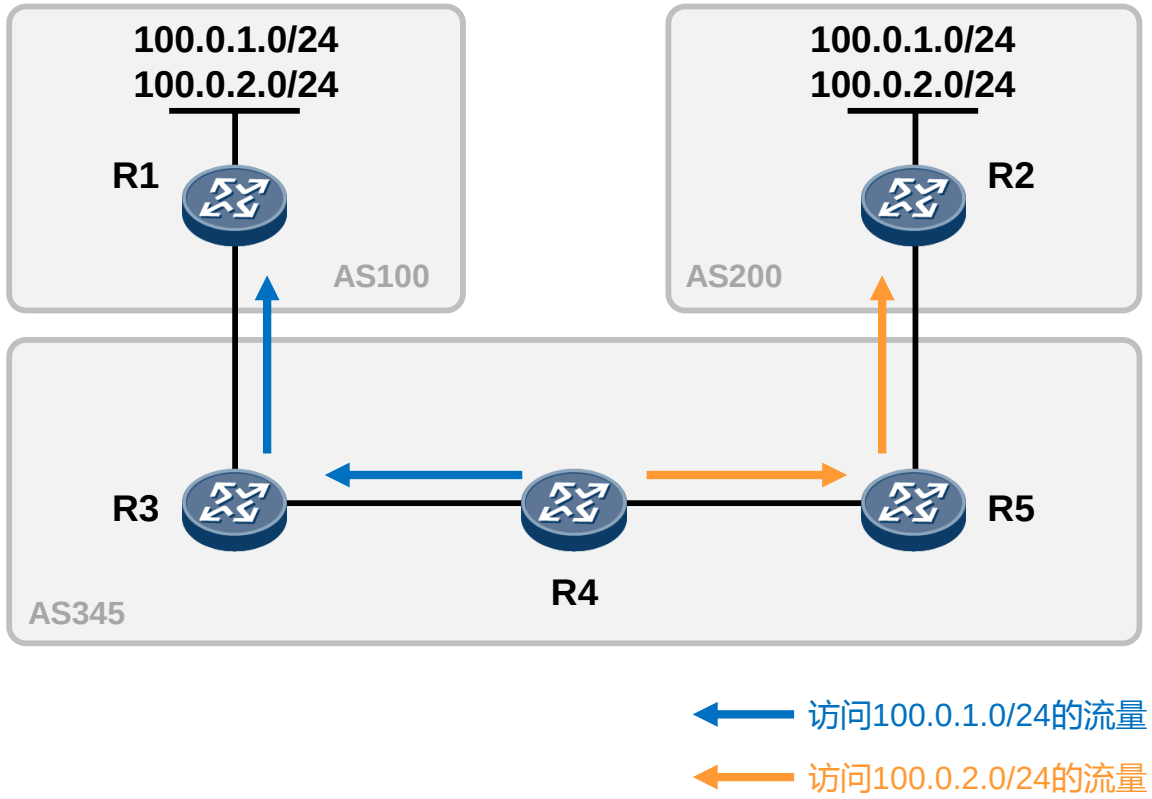
关于100.0.1.0/24路由，优选了R3所通告的路径

关于100.0.2.0/24路由，优选了R5所通告的路径

Total Number of Routes: 4

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
***>i**	**100.0.1.0/24**	**3.3.3.3**	**0**	**100**	**10**	**100i**
* i		5.5.5.5	0	100	0	200i
***>i**	**100.0.2.0/24**	**5.5.5.5**	**0**	**100**	**10**	**200i**
* i		3.3.3.3	0	100	0	100i

使用route-policy修改Preferred-Value（续）

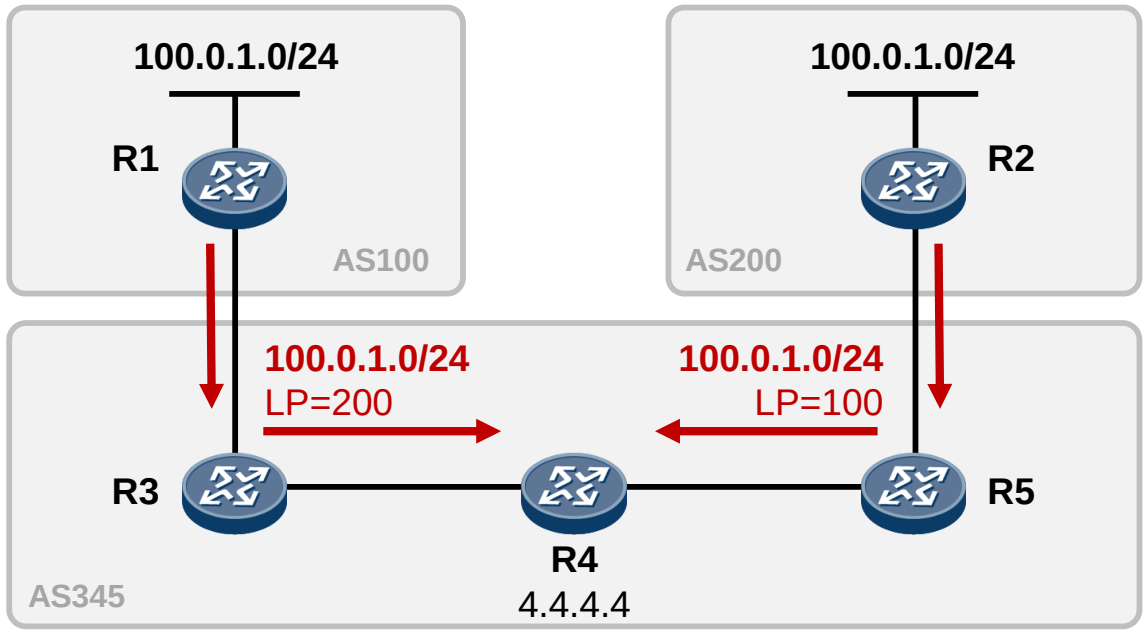


通过操控BGP路由优选，实现了数据的分流。

目录

HUAWEI

通过route-policy修改Local_Preference



R3的关键配置如下：

```
ip ip-prefix 1 permit 100.0.1.0 24

route-policy RP permit node 10
  if-match ip-prefix 1
  apply local-preference 200
route-policy RP permit node 20

bgp 345
  peer 4.4.4.4 route-policy RP export
```

在R3上执行路由策略，当其向R4通告100.0.1.0/24路由时，将该路由的LP属性值设置为200，使得R4优选R3所通告的100.0.1.0/24路由。

通过route-policy修改Local_Preference（续）

```
[R4]display bgp routing-table
BGP Local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete

 Total Number of Routes: 4
      Network          NextHop        MED        LocPrf    PrefVal    Path/Ogn

*>i  100.0.1.0/24     3.3.3.3        0          200       0          100i
* i                   5.5.5.5        0          100       0          200i
```

在其他条件相同的情况下，由于R3所通告的100.0.1.0/24路由的Local_Preference属性值大于R5所通告的100.0.1.0/24路由的属性值，因此R4将优选R3所通告的该条路由。

目录

HUAWEI

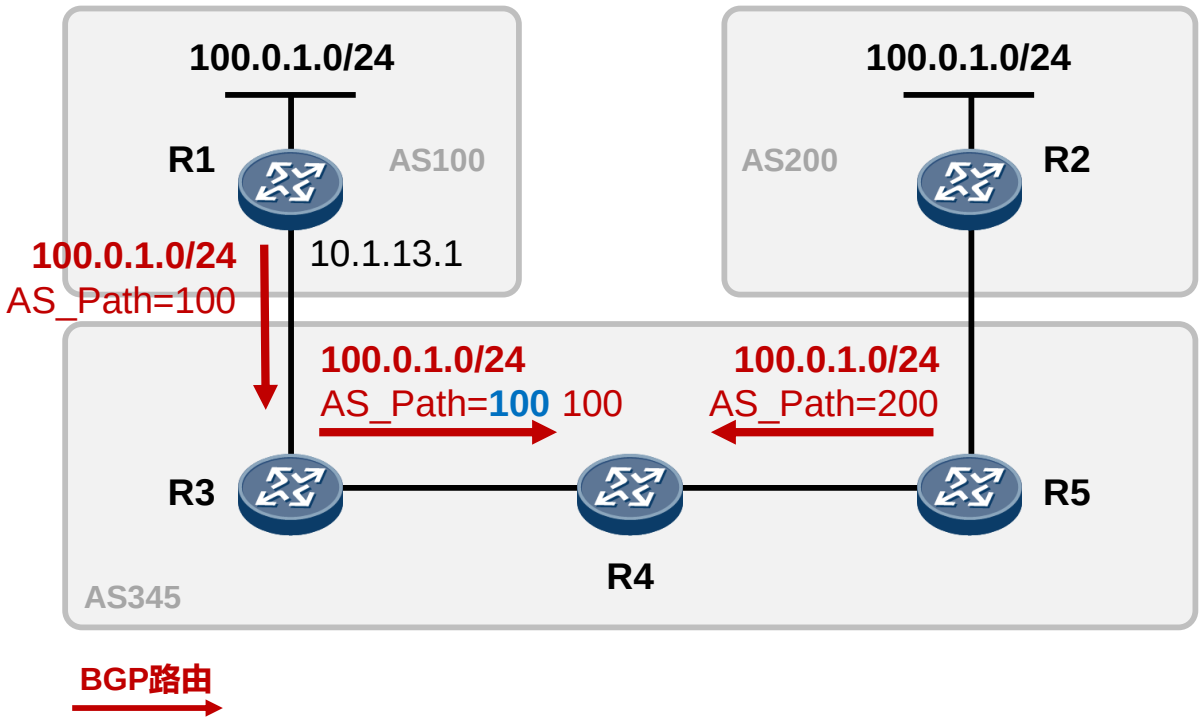
规则描述

- 在其他条件相同的情况下，优选本地生成的路由（本地生成的路由优先级高于从邻居学来的路由）。
- 本地生成的路由包括通过network或import-route命令引入的路由、手工汇总路由和自动汇总路由。这些本地生成的路由之间的优选如下：
  1. 优选汇总路由（汇总路由优先级高于非汇总路由）。
  2. 通过aggregate命令生成的手动汇总路由的优先级高于通过summary automatic命令生成的自动汇总路由。
  3. 通过network命令引入的路由的优先级高于import-route命令引入的路由。

目录

HUAWEI

通过route-policy修改AS_Path属性



R3的关键配置如下：

```
ip ip-prefix 1 permit 100.0.1.0 24

route-policy RP permit node 10
  if-match ip-prefix 1
  apply as-path 100 additive
route-policy RP permit node 20

bgp 345
  peer 10.1.13.1 route-policy RP import
```

在R3上对R1执行入站（Import）方向的路由策略，使得其在收到对方通告的100.0.1.0/24路由后，在该路由的AS_Path属性值前面插入一个AS号（100），从而将AS_Path属性值的长度增加1，使得R4优选R5通告的路由。

通过route-policy修改AS_Path属性（续）

[R4]display bgp routing-table

BGP Local router ID is 4.4.4.4

Status codes: * - valid, > - best, d - damped,

h - history, i - internal, s - suppressed, S - Stale

Origin : i - IGP, e - EGP, ? - incomplete

Total Number of Routes: 2

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
***>i**	**100.0.1.0/24**	**5.5.5.5**	**0**	**100**	**0**	**200i**
* i		3.3.3.3	0	100	0	100 100i

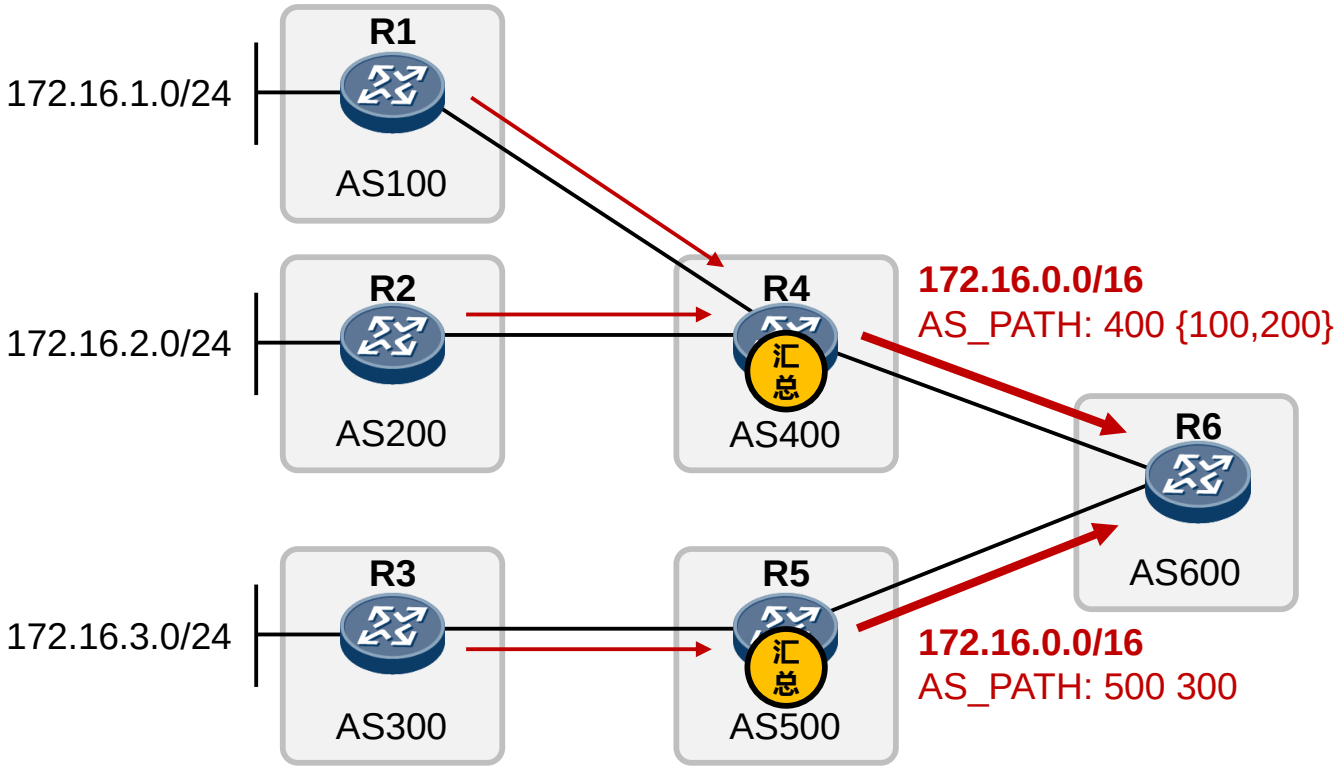
在其他条件相同的情况下，由于R3所通告的100.0.1.0/24路由的AS_Path属性值的长度比R5所通告的100.0.1.0/24路由的AS_Path长度更长，因此R4将优选R5所通告的路由。

通过route-policy修改AS_Path属性（续）

- **使用route-policy修改BGP路由的AS_Path：**
  - **apply as-path** xxx **additive**　　在已有AS_Path基础上追加xxx
  - **apply as-path** xxx **overwrite**　　将已有AS_Path值替换（覆盖）成xxx
  - **apply as-path none overwrite**　　清空路由的AS_Path属性
- 使用route-policy修改BGP路由的AS_Path时，可以在EBGP对等体之间改变EBGP路由的AS_Path属性，从而影响BGP路由的优选。在华为路由器上，在IBGP对等体之间，也可以使用route-policy修改BGP路由的AS_Path。无论何种场景，改变BGP路由的AS_Path都必须十分谨慎。
- Bestroute as-path-ignore命令用来配置BGP在选择最优路由时忽略AS路径属性。配置该命令后，BGP将不比较AS路径的长度。缺省情况下，长度更小者优。

AS_Path选路规则扩展1

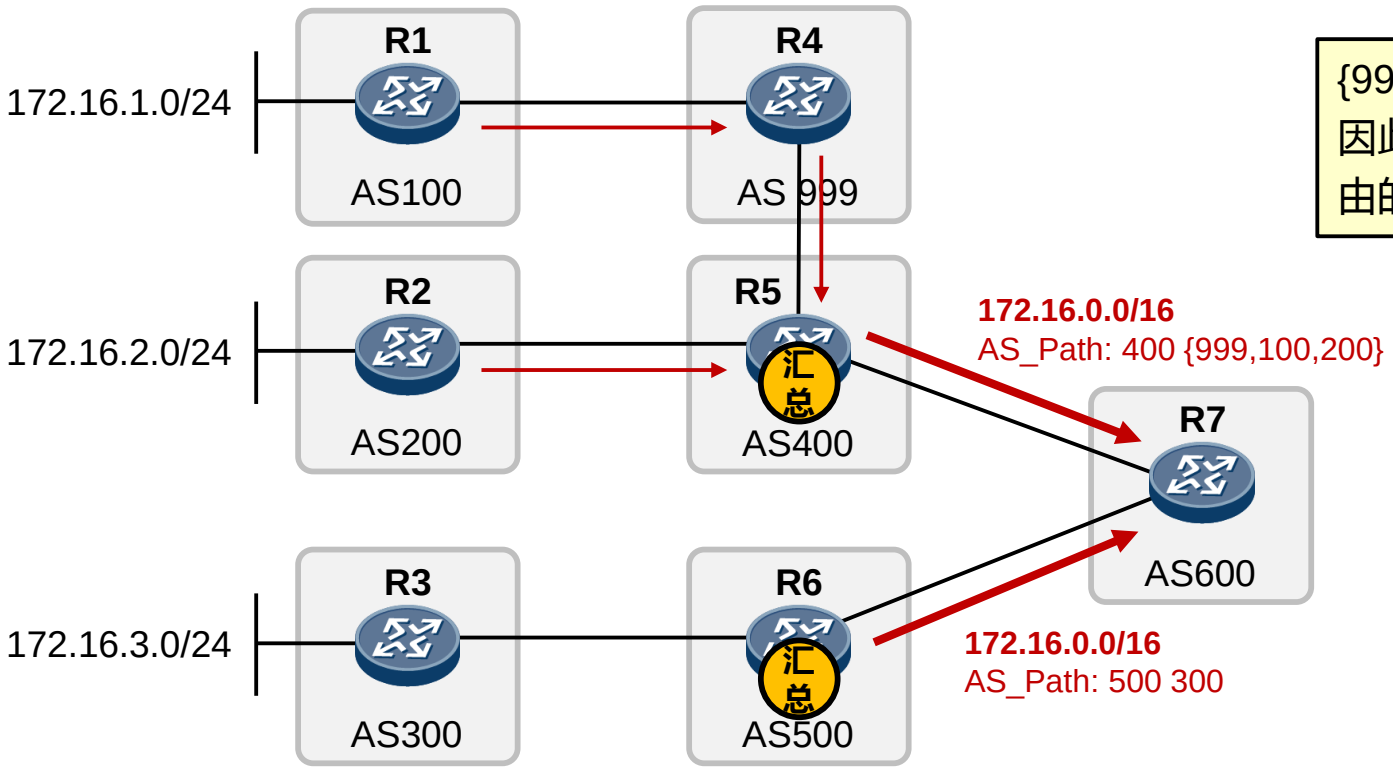
- 规则补充：AS_Set的长度为1，无论AS_Set中包括多少AS号，长度仅当做1来计算。



AS_Path属性中，as-set类型的AS号，也就是图中{100,200}部分在AS_Path长度计算时，只作为1跳AS。因此R6上所收到的两条172.16.0.0/16路由AS_Path长度相等，R6需通过其他属性进行路由优选决策。

AS_Path选路规则扩展1（续）

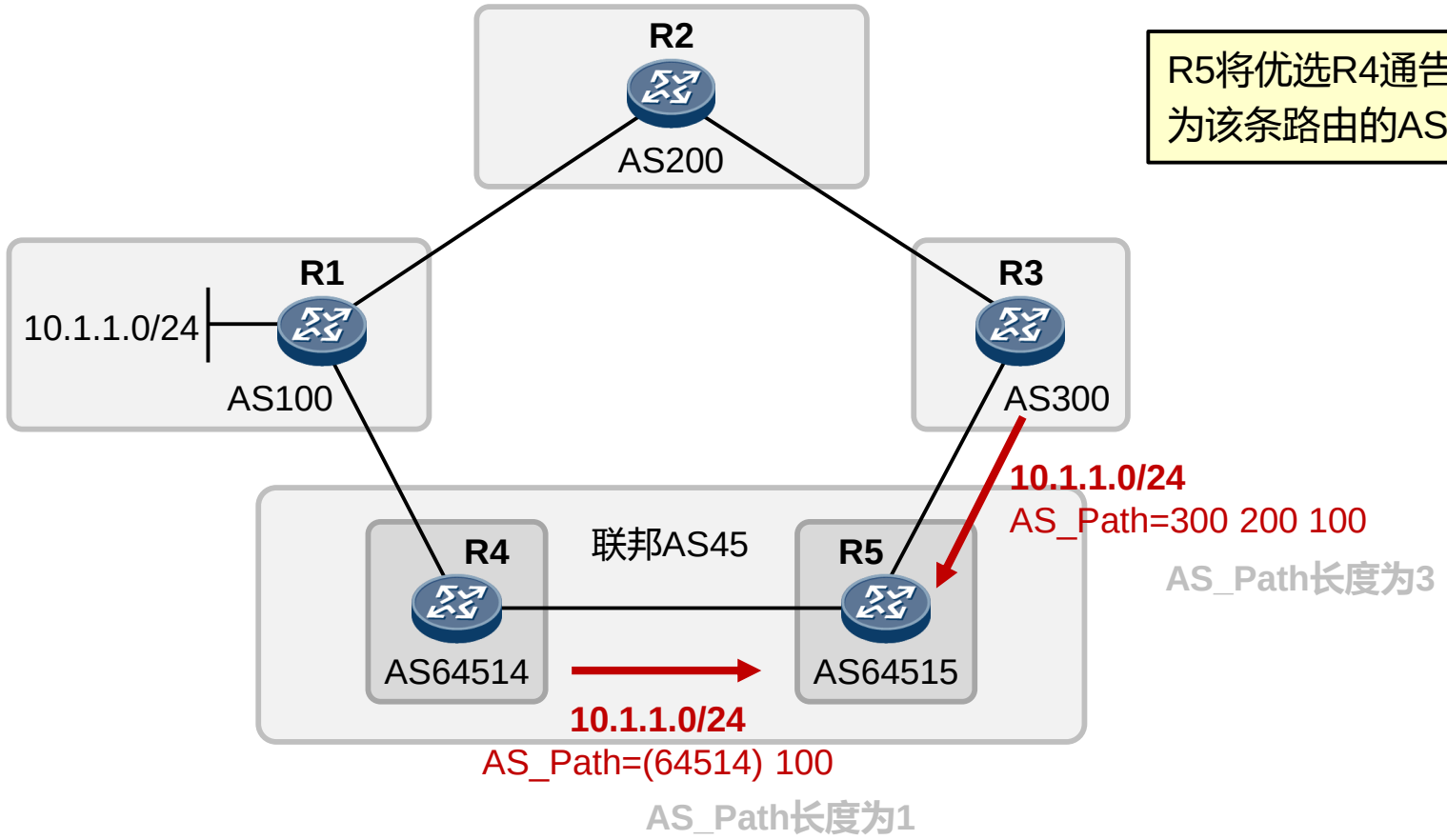
- 规则补充：AS_Set的长度为1，无论AS_Set中包括多少AS号，长度仅当做1来计算。



{999,100,200}是AS_Set类型，长度为1，因此R6上所收到的两条172.16.0.0/0/16路由的AS_Path长度相等。

AS_Path选路规则扩展2

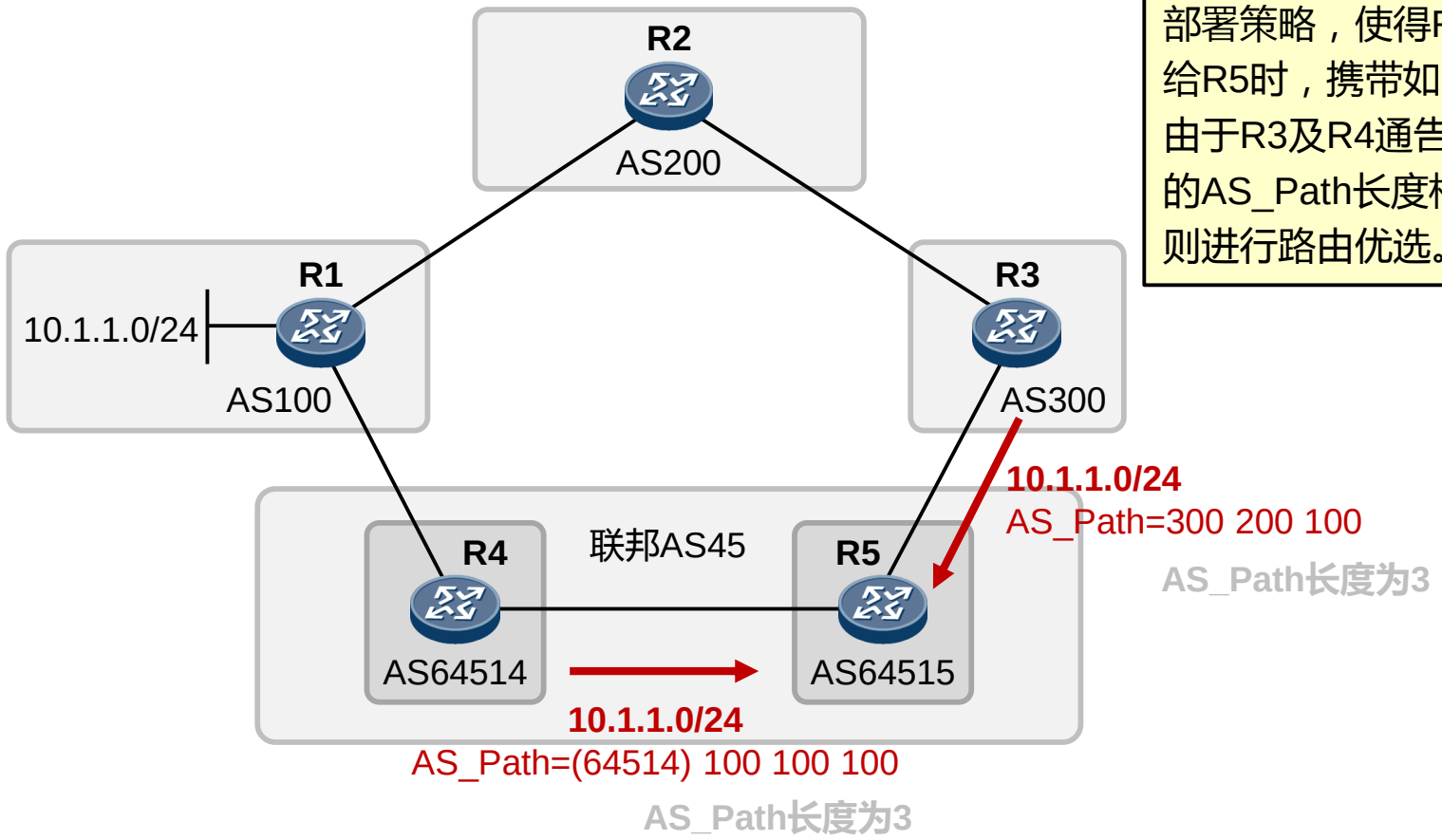
- 规则补充：AS_confed_seq和AS_confed_set类型不参与AS_Path长度计算。



R5将优选R4通告的10.1.1.0/24路由，因为该条路由的AS_Path长度更短。

AS_Path选路规则扩展2（续）

- 规则补充：AS_confed_seq和AS_confed_set类型不参与AS_Path长度计算。



部署策略，使得R4将10.1.1.0/24路由通告给R5时，携带如图所示的AS_Path，此时由于R3及R4通告给R5的10.1.1.0/24路由的AS_Path长度相同，因此需通过其他规则进行路由优选。

目录

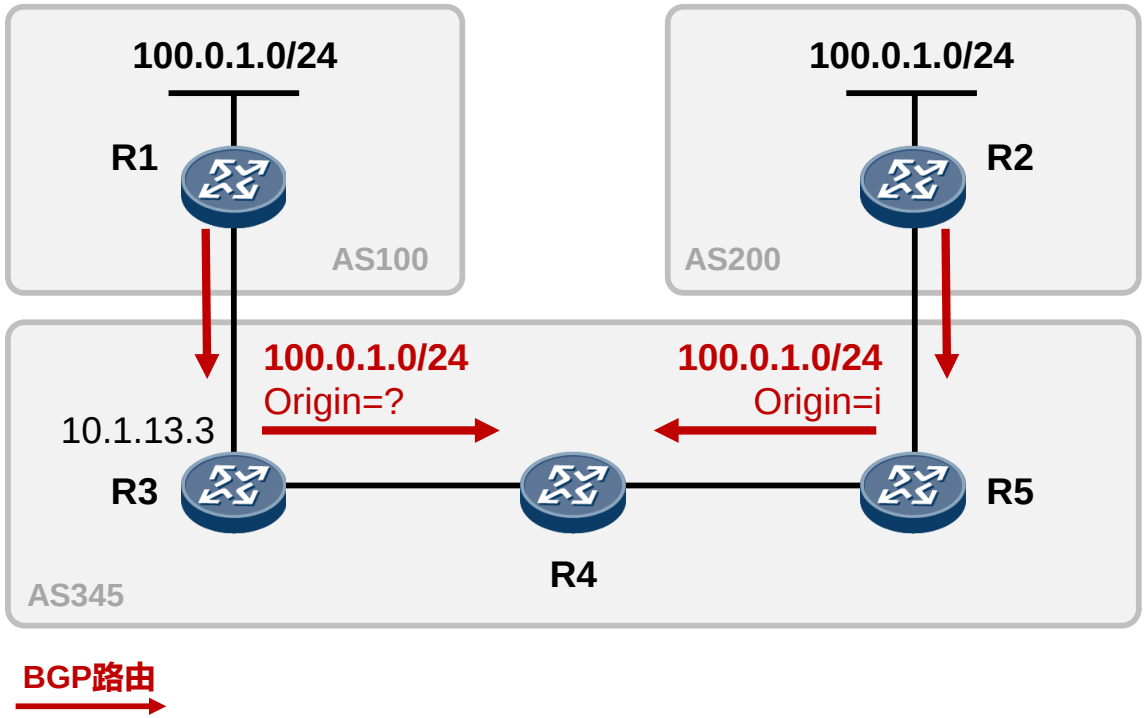
HUAWEI

Origin属性

- 该属性为公认必遵属性，它标识了BGP路由的起源。如下表所示，根据路由被引入BGP的方式不同，存在三种类型的Origin。
- 当去往同一个目的地存在多条不同Origin属性的路由时，在其他条件都相同的情况下，BGP将按如Origin的下顺序优选路由：IGP > EGP > Incomplete。

名称	标记	描述
IGP	i	通过BGP network的路由，也就是起源于IGP的路由，Origin为igp。因为BGP network必须保证该网络在路由表中。
EGP	e	如果BGP路由是由EGP 这种早期的协议重发布而来，那么其Origin为egp。
Incomplete	?	通过Import命令，从其他协议引入到BGP的路由，其Origin为Incomplete（确认该路由来源的信息不完全）。

使用route-policy修改路由Origin属性值



R1的关键配置如下：

```
ip ip-prefix 1 permit 100.0.1.0 24

route-policy RP permit node 10
  if-match ip-prefix 1
  apply origin incomplete
route-policy RP permit node 20

bgp 100
  network 100.0.1.0 24
  peer 10.1.13.3 as-number 345
  peer 10.1.13.3 route-policy RP export
```

R1及R2通过network的方式将100.0.1.0/24路由通告到BGP。在R1上执行路由策略，使得其通告给R3的该条路由的Origin属性值被修改为incomplete，如此一来，R4将优选R5所传递的100.0.1.0/24路由。

使用route-policy修改路由Origin属性值（续）

[R4]display bgp routing-table

BGP Local router ID is 4.4.4.4

Status codes: * - valid, > - best, d - damped,

h - history, i - internal, s - suppressed, S - Stale

Origin : i - IGP, e - EGP, ? - incomplete

Total Number of Routes: 2

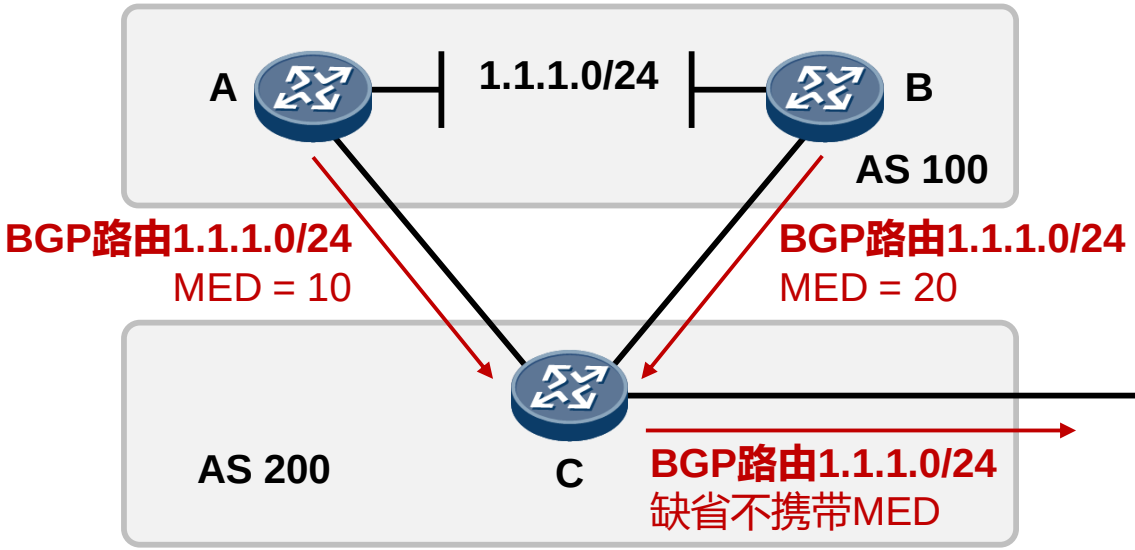
	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i	100.0.1.0/24	5.5.5.5	0	100	0	200 i
* i		3.3.3.3	0	100	0	100 **?**

目录

HUAWEI

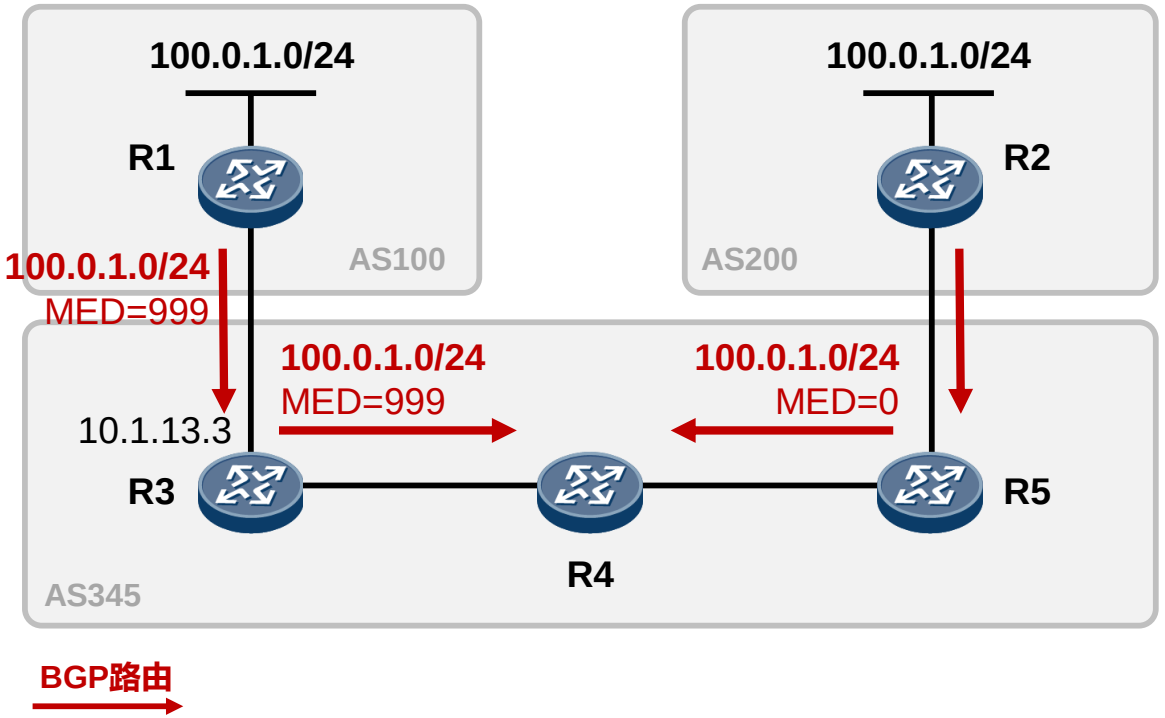
优选MED最小的路由

- MED（Multi Exit Discriminator）是可选非传递属性，是一种度量值，用于向外部对等体指出进入本AS的首选路径，即当进入本AS的入口有多个时，AS可以使用MED动态地影响其他AS选择进入的路径。
- MED属性值越小则BGP路由越优。
- MED主要用于在AS之间影响BGP的选路。MED被传递给EBGP对等体后，对等体在其AS内传递路由时，携带该MED值，但将路由传递给其EBGP对等体时，缺省不会携带MED属性。



在A、B上部署策略，使得A通告给C的BGP路由1.1.1.0/24的MED值为10，而B通告的1.1.1.0/24路由的MED值为20。当其他条件相同时，C将优选A传递过来的BGP路由。

使用route-policy修改路由的MED属性值



R1的关键配置如下：

```
ip ip-prefix 1 permit 100.0.1.0 24

route-policy RP permit node 10
  if-match ip-prefix 1
  apply cost 999
route-policy RP permit node 20

bgp 100
  network 100.0.1.0 24
  peer 10.1.13.3 as-number 345
  peer 10.1.13.3 route-policy RP export
```

R1及R2通过network的方式将100.0.1.0/24路由通告到BGP。在R1上执行路由策略，使得其通告给R3的该条路由的MED属性值被设置为999，如此一来，R4将优选R5所传递的100.0.1.0/24路由，因为该条路由的MED属性值更小。

使用route-policy修改路由的MED属性值（续）

[R4]display bgp routing-table

BGP Local router ID is 4.4.4.4

Status codes: * - valid, > - best, d - damped,

h - history, i - internal, s - suppressed, S - Stale

Origin : i - IGP, e - EGP, ? - incomplete

Total Number of Routes: 2

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i	100.0.1.0/24	5.5.5.5	0	100	0	100i
* i		3.3.3.3	**999**	100	0	100i

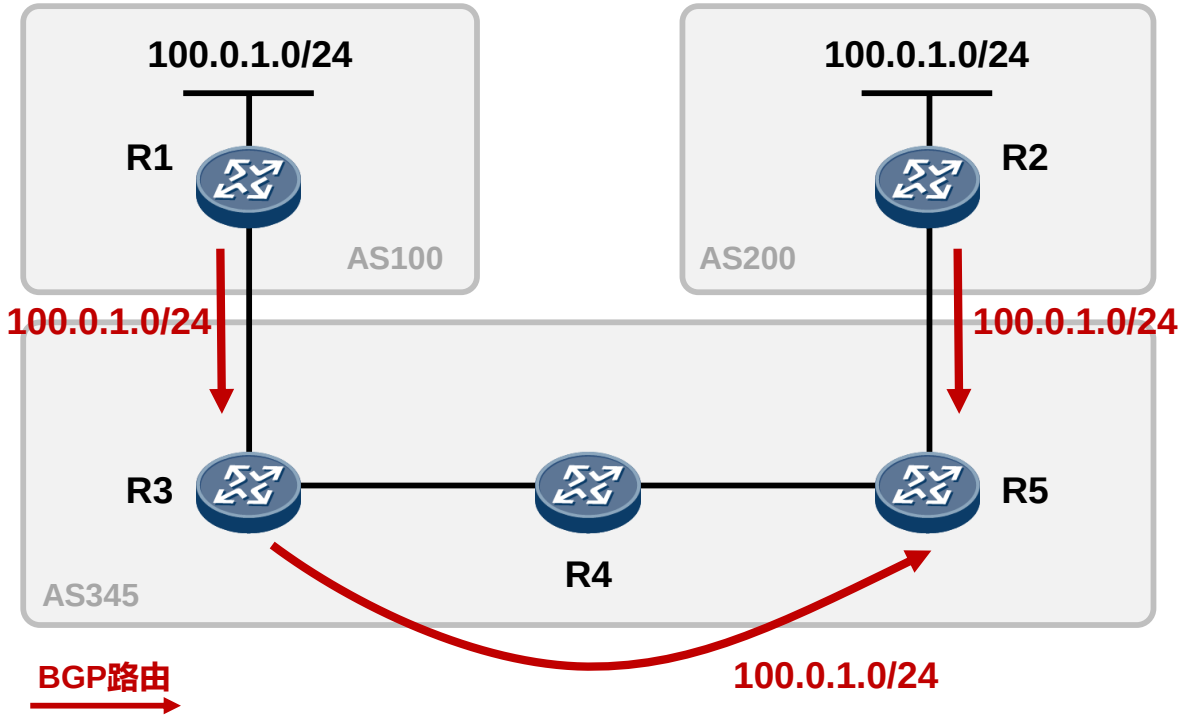
关于MED对BGP路由优选的影响

- 一般情况下，BGP设备只比较来自同一AS（不同对等体）的路由的MED属性值。可以通过配置命令来允许BGP比较来自不同AS的路由的MED属性值。执行compare-different-as-med命令后，系统将比较来自不同AS中的对等体的路由的MED值。
- 如果路由没有MED属性，BGP选路时将该路由的MED值按缺省值0来处理；执行bestroute med-none-as-maximum命令后，BGP选路时将该路由的MED值按最大值4294967295来处理。
- BGP路由属性AS_Path按一定次序记录了某条路由从本地到目的地址所要经过的所有自治系统号。配置bestroute med-confederation命令后，只有当AS_Path中不包含外部自治系统（不在联盟范围内的自治系统）号时才比较MED值的大小。如果AS_Path中包含外部自治系统号，则不进行比较。
- 使能deterministic-med功能，在对从多个不同AS收到的相同前缀的路由进行选路时，首先会按路由的AS_Path最左边的AS号进行分组。在组内进行比较后，再用组中的优选路由和其他组中的优选路由进行比较，消除了选路的结果和路由接收顺序的相关性。

目录

1. 优选具有最大Preferred-Value的路由
2. 优选具有最大Local_Preference的路由
3. 优选起源于本地的路由
4. 优选AS_Path最短的路由
5. Origin（IGP > EGP > Incomplete）
6. 优选MED最小的路由
7. 优选EBGP对等体所通告的路由
8. 优选到Next_Hop的IGP度量值最小的路由
9. BGP路由负载分担
10. 优选Cluster_List 最短的路由
11. 优选Router-ID最小的BGP对等体发来的路由
12. 优选Peer-IP地址最小的对等体发来的路由

优选EBGP对等体所通告的路由（相对于IBGP对等体）



R5从IBGP对等体R3及EBGP对等体R2都学习到了BGP路由100.0.1.0/24，在其他条件相同的情况下，R5优选EBGP对等体R2传递过来的BGP路由。

注意：在本规则的验证中，需在R3及R5之间增加IBGP对等体关系，该对等体关系基于双方的Loopback0接口建立，而且需在R3上对R5配置next-hop-local。

优选EBGP对等体所通告的路由（续）

```
[R5-bgp]display bgp routing-table 100.0.1.0
 BGP routing table entry information of 100.0.1.0/24:
 From: 10.1.25.2 (2.2.2.2)
 Original nexthop: 10.1.25.2
 AS-path 200, origin igp, MED 0, pref-val 0, valid, external, best, select, active, pre 255
……
BGP routing table entry information of 100.0.1.0/24:
 From: 3.3.3.3 (10.1.13.3)
 Route Duration: 00h01m22s
 Original nexthop: 3.3.3.3
AS-path 100, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, pre255, IGP
cost 2, not preferred for peer type
```

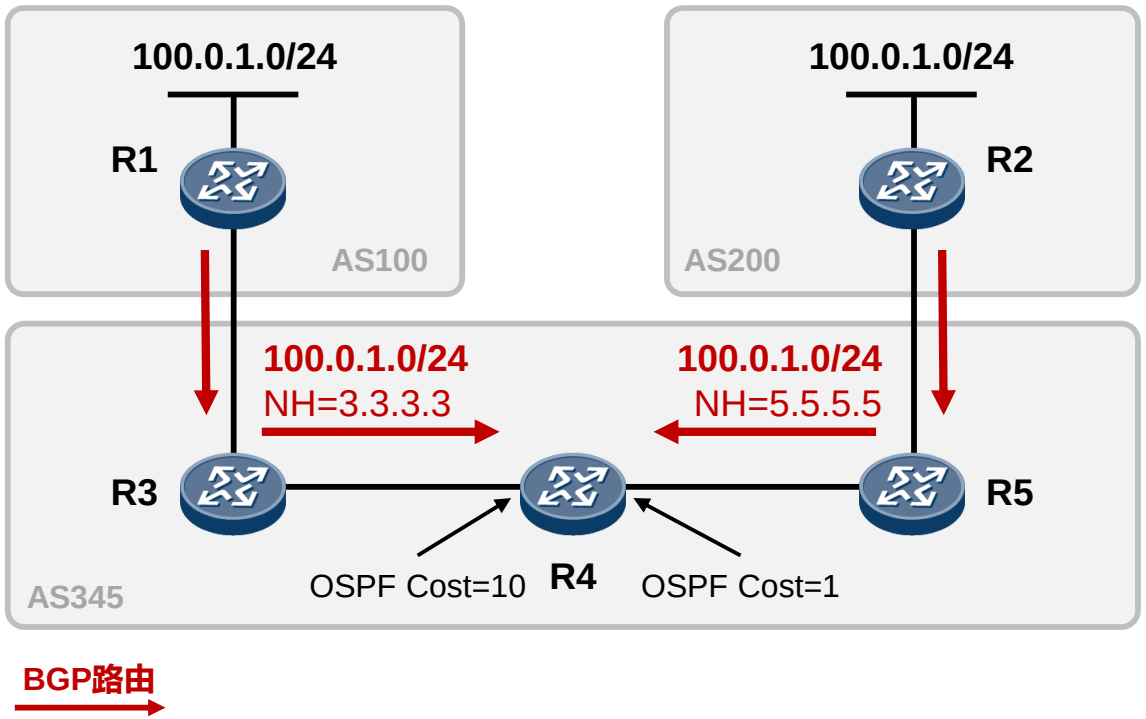
描述出了该条路由没有被优选的原因。

目录

HUAWEI

优选到Next_Hop的IGP度量值最小的路由

- 场景示例1



- R3与R4，R4与R5均维护IBGP对等体关系并且都使用各自的Loopback接口作为更新源。AS345内运行OSPF，使得所有路由器都能学习到其他路由器的Loopbback接口路由。
- R4同时从R3及R5学习到100.0.1.0/24的BGP路由，Next_Hop分别为3.3.3.3及5.5.5.5，在其他条件相同的情况下，R4将比较其到达这两个Next_Hop的Cost，由于到5.5.5.5的Cost更小，因此R4优选R5通告的100.0.1.0/24路由。

优选到Next_Hop的IGP度量值最小的路由

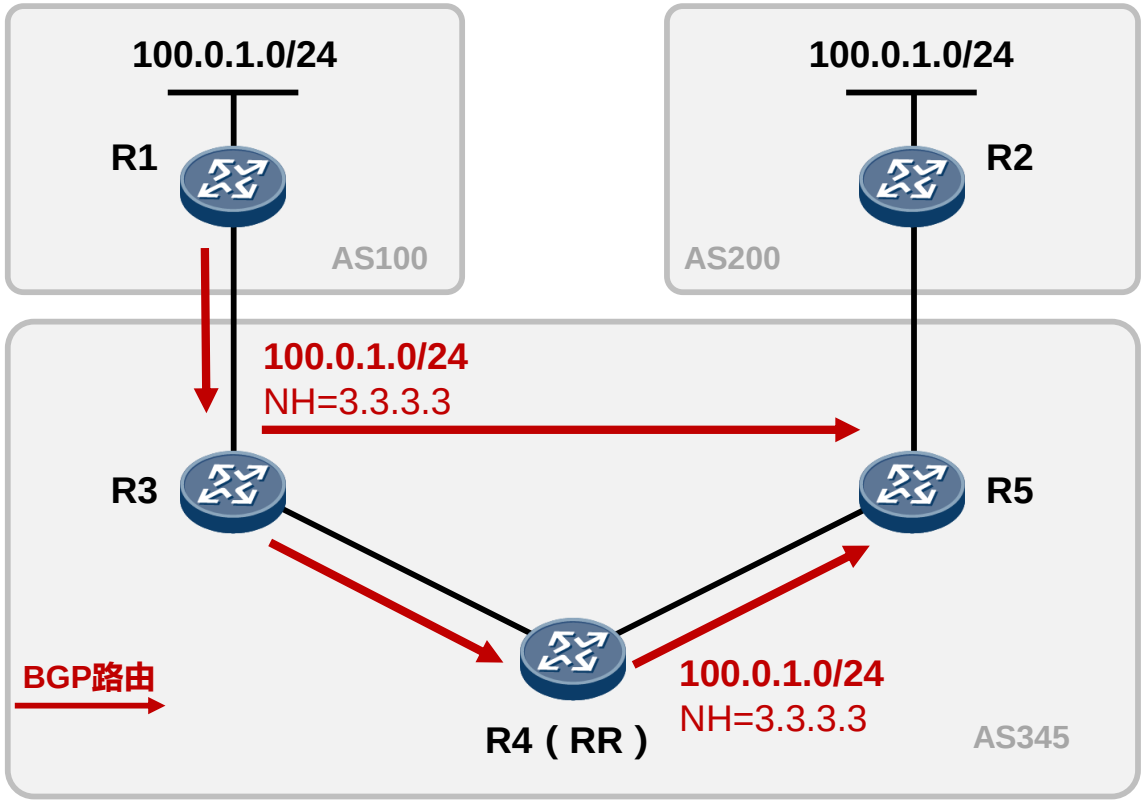
- 场景示例1（续）

[R4]display bgp routing-table 100.0.1.0
BGP local router ID : 4.4.4.4
Local AS number : 345
Paths: 2 available, 1 best, 1 select
BGP routing table entry information of 100.0.1.0/24:
From: 5.5.5.5 (5.5.5.5) #路径1
Route Duration: 00h00m07s
Relay IP Nexthop: 10.1.45.5
Relay IP Out-Interface: GigabitEthernet0/0/1
Original nexthop: 5.5.5.5
Qos information : 0x0
AS-path 200, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, best
, select, active, pre 255, **IGP cost 1**
Not advertised to any peer yet
… …接右

… …
BGP routing table entry information of 100.0.1.0/24:
From: 3.3.3.3 (3.3.3.3) #路径2
Route Duration: 00h00m23s
Relay IP Nexthop: 10.1.34.3
Relay IP Out-Interface: GigabitEthernet0/0/0
Original nexthop: 3.3.3.3
Qos information : 0x0
AS-path 100, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, pre
255, **IGP cost 10, not preferred for IGP cost**
Not advertised to any peer yet

优选到Next_Hop的IGP度量值最小的路由

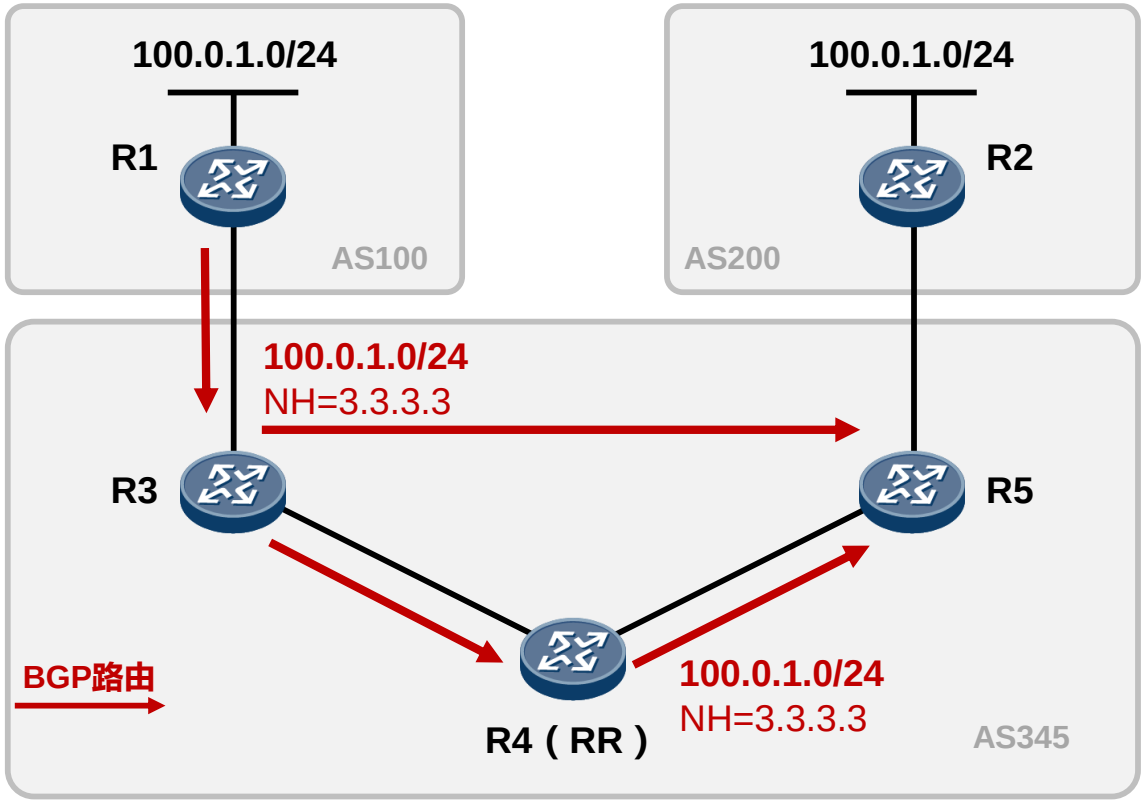
- 场景示例2（思考）



- R3与R4，R4与R5，R3与R5均维护IBGP对等体关系并且都使用各自的Loopback接口作为更新源。AS345内运行OSPF，使得所有路由器都能学习到其他路由器的Loopbback接口路由。
- 配置R4为RR，R3为该RR的Client 。
- R5会同时从R3及R4学习到100.0.1.0/24的BGP路由，它将如何选路？哪一条选路规则生效？

优选到Next_Hop的IGP度量值最小的路由

- 场景示例2（解答）

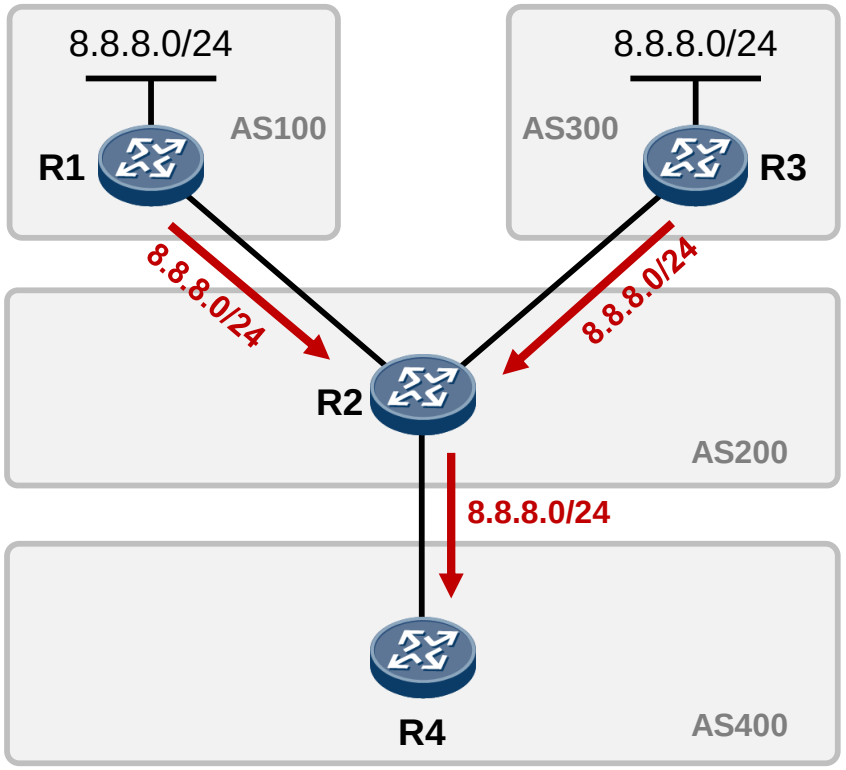


- 首先“优选到Next_Hop的IGP度量值最小的路由”规则并不适用，因为两条BGP路由的Next_Hop属性值相等，都是3.3.3.3，不具有可比性。
- 最后通过“选路规则10”——优选Cluster_List长度最短的路由，作出决策，优选R3通告的100.0.1.0/24路由。

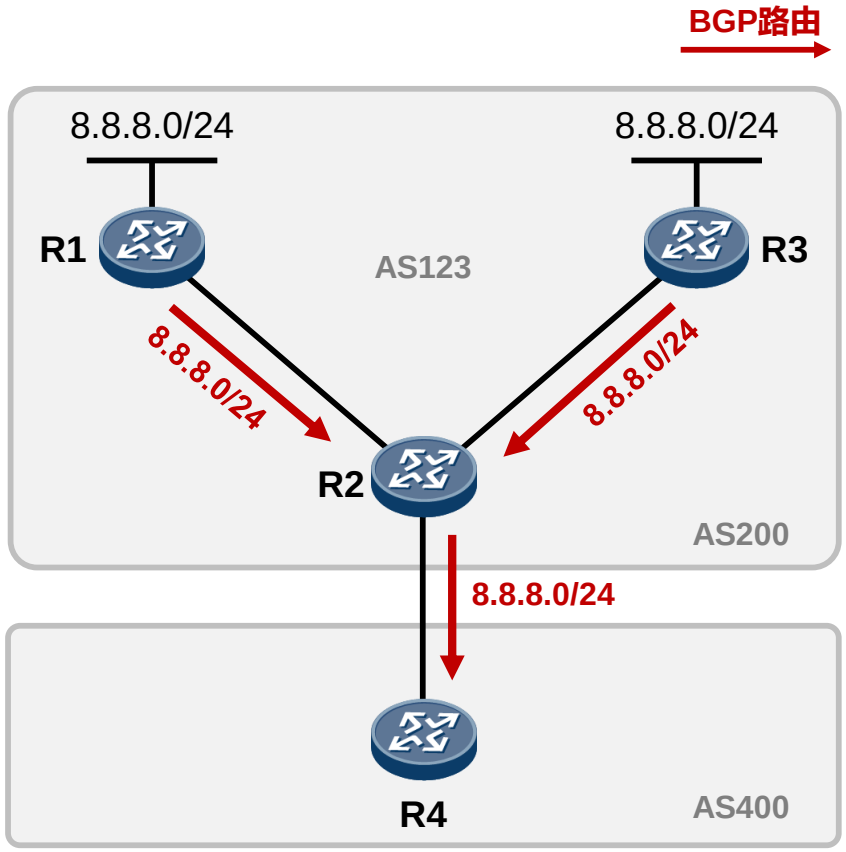
目录

HUAWEI

没有配置BGP路由负载分担时



R2学习到两条去往8.8.8.0/24网段的EBGP路由，它只会优选1条最优的路由，将该路由加载到路由表使用。而且只将最优路由传递给R4。



R2学习到两条去往8.8.8.0/24网段的IBGP路由，它只会优选1条最优的路由，将该路由加载到路由表使用。而且只将最优路由传递给R4。

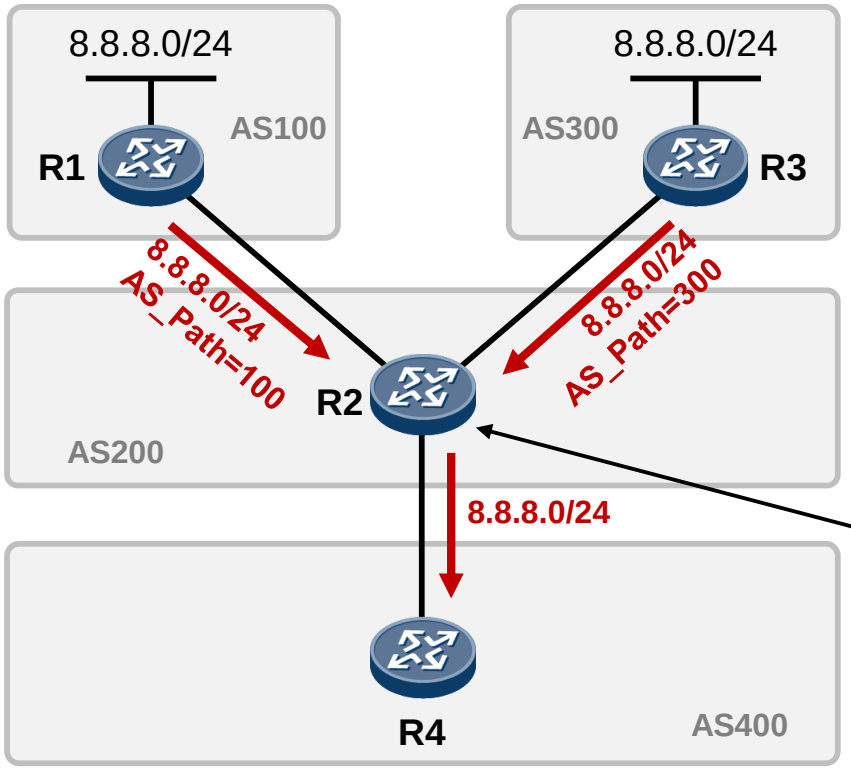
BGP路由负载分担

- 在大型网路中，到达同一目的地通常会存在多条有效BGP路由，设备只会优选一条最优的BGP路由，将该路由加载到路由表中使用，并且只将最优路由发布给对等体，这一特点往往会造成很多流量负载不均衡的情况。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。
- 值得注意的是，尽管配置了BGP负载分担，设备依然只会在多条到达同一目的地的BGP路由中优选一条路由，并只将这条路由通告给其他对等体。
- 形成BGP等价负载分担的条件是“BGP路由优选规则”的1至8条规则中需要比较的属性完全相同，例如相同的Preferred_Value、Local_Preference、AS_Path（包括长度及值）、MED、Origin、到达Next_Hop的IGP度量值、路由类型（IBGP或EBGP）等。

BGP路由负载分担（续）

- 如果实现了BGP负载分担，则不论是否配置了peer next-hop-local命令，本地设备向IBGP对等体组发布路由时都先将下一跳地址改变为自身地址。
- 在公网中到达同一目的地的路由形成负载分担时，系统会首先判断最优路由的类型。若最优路由为IBGP路由则只是IBGP路由参与负载分担，若最优路由为EBGP路由则只是EBGP路由参与负载分担，即公网中到达同一目的地的IBGP和EBGP路由不能形成负载分担。
- 如果到达目的地址存在多条路由，但是这些路由分别经过了不同的AS，缺省情况下，这些路由不能形成负载分担。如果用户需要这些路由参与负载分担，就可以执行load-balancing as-path-ignore命令。配置load-balancing as-path-ignore命令后会改变路由参与负载分担的条件，路由形成负载分担时不再比较AS-Path属性，配置时需要慎重考虑。
- load-balancing as-path-ignore命令和bestroute as-path-ignore命令互斥，不能同时使能。

配置了BGP路由负载分担之后1



[R2-bgp] **maximum load-balancing ebgp** 2
//修改EBGP路由负载分担的最大等价路由条数为2。缺省时该值为1，也就是不执行负载分担。
[R2-bgp] **load-balancing as-path-ignore**
//由于R1及R3所通告的两条路由的AS_Path属性值不同，因此需配置该条命令，使得路由在形成负载分担时不比较路由的AS-Path属性，该命令需谨慎配置，否则可能会引起路由环路。

R2学习到两条去往8.8.8.0/24网段的EBGP路由（除AS_Path属性外，其他路径属性相同），它只会优选1条最优的路由。而且只将最优路由传递给R4。但是由于R2配置了maximum load-balancing ebgp 2，因此它会将两条等代价的EBGP路径都加载到路由表中，执行路由负载分担。

配置了BGP路由负载分担之后1（续）

```
<R2>display ip routing-table protocol bgp
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        8.8.8.0/24  EBGP    255  0         D     10.1.12.1       GigabitEthernet0/0/0
                    EBGP    255  0         D     10.1.23.3       GigabitEthernet0/0/1
```

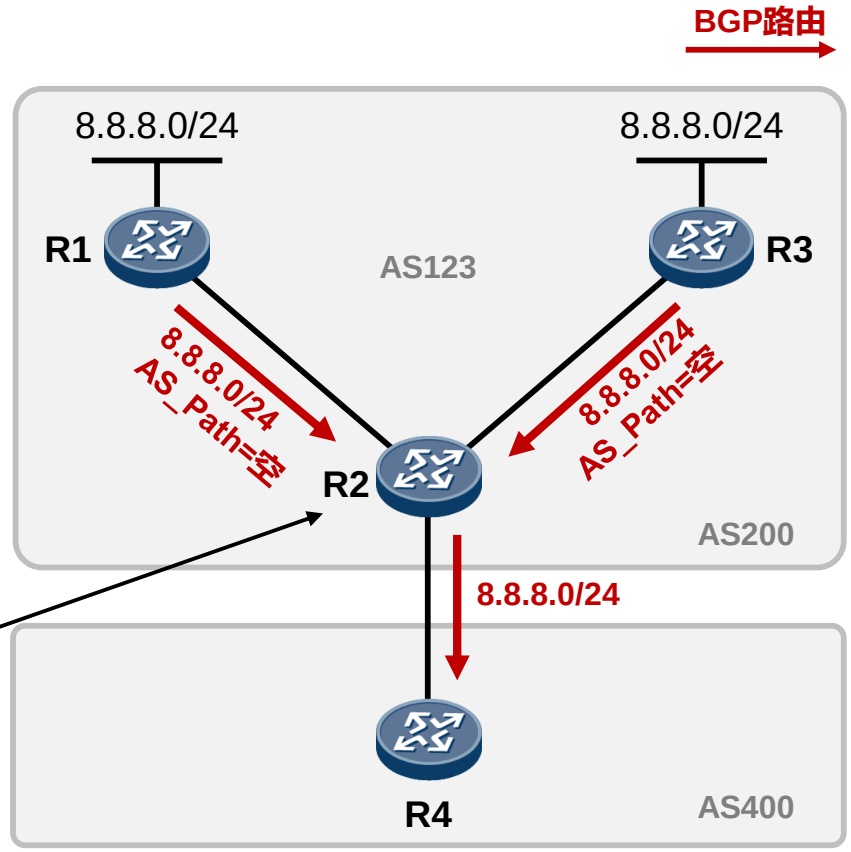
R2的路由表中出现到达8.8.8.0/24网段路由的等价负载分担。

```
<R2>display bgp routing-table
 BGP Local router ID is 10.1.12.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
 Total Number of Routes: 2
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.8.8.0/24         10.1.12.1      0                    0       100i
 *                       10.1.23.3      0                    0       300i
```

R2的BGP表中，两条BGP路由依然仅有一条被优选。该条路由被传递给R4。

配置了BGP路由负载分担之后2



[R2-bgp] **maximum load-balancing ibgp** 2
//修改IBGP路由负载分担的最大等价路由条数为2。缺省时该值为1，也就是不执行负载分担。

R2学习到两条去往8.8.8.0/24网段的IBGP路由（路径属性相同），它只会优选1条最优的路由。而且只将最优路由传递给R4。但是由于R2配置了maximum load-balancing ibgp，因此它会将两条等代价的IBGP路径都加载到路由表中，执行路由负载分担。

配置了BGP路由负载分担之后2（续）

<R2>display ip routing-table protocol bgp

Route Flags: R - relay, D - download to fib

Destination/Mask	Proto	Pre	Cost	Flags	NextHop	Interface
8.8.8.0/24	IBGP	255	0	RD	1.1.1.1	GigabitEthernet0/0/0
	IBGP	255	0	RD	3.3.3.3	GigabitEthernet0/0/1

R2的路由表中出现到达8.8.8.0/24网段路由的等价负载分担。

<R2>display bgp routing-table

BGP Local router ID is 10.1.12.2

Status codes: * - valid, > - best, d - damped,

h - history, i - internal, s - suppressed, S - Stale

Origin : i - IGP, e - EGP, ? - incomplete

Total Number of Routes: 2

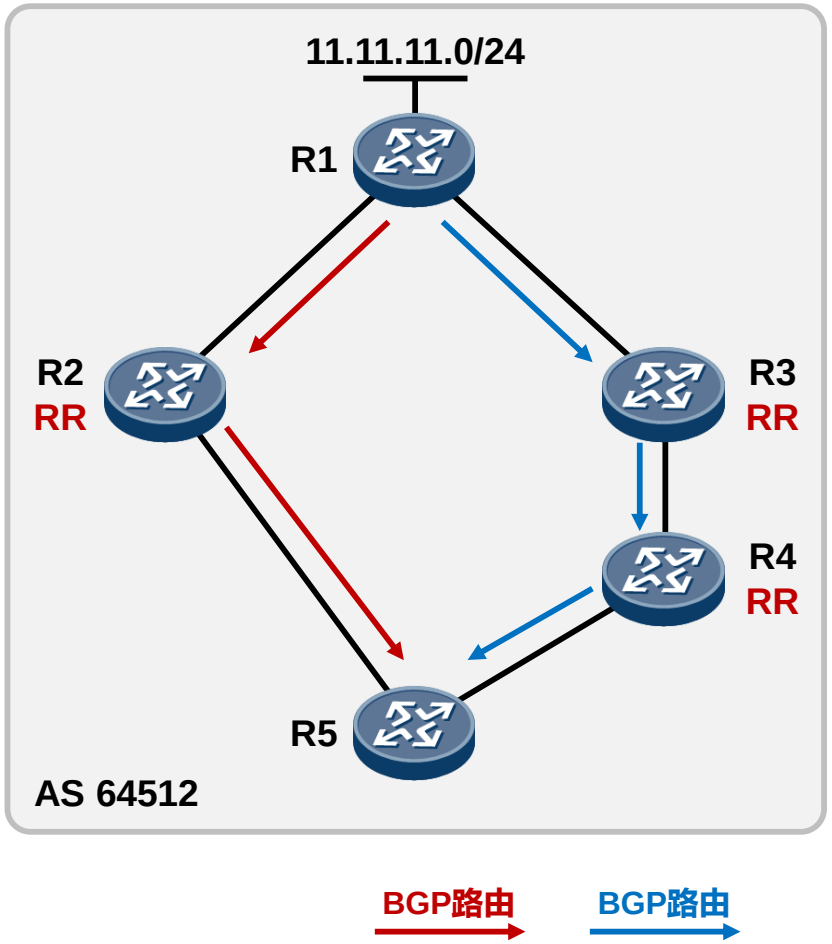
Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i 8.8.8.0/24	1.1.1.1	0	100	0	i
* i	3.3.3.3	0	100	0	i

R2的BGP表中，两条BGP路由依然仅有一条被优选。该条路由被传递给R4。

目录

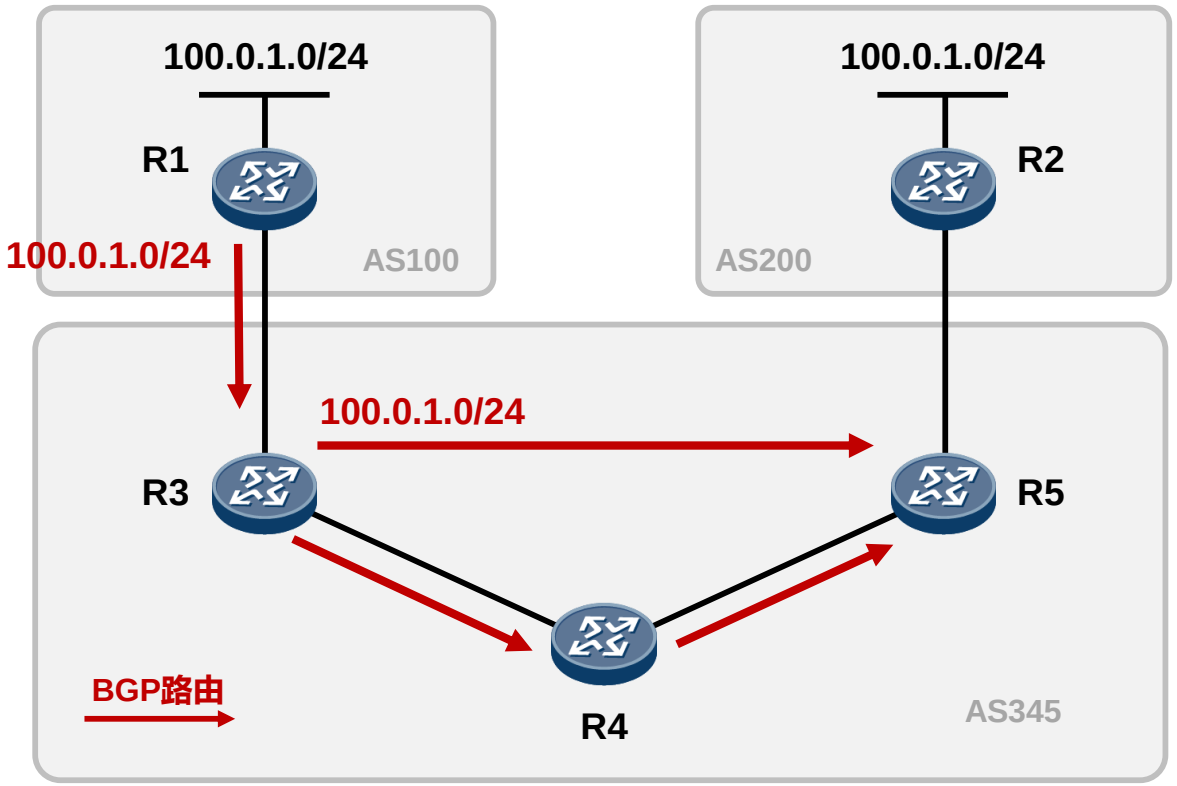
HUAWEI

优选Cluster_List最短的路由 案例1



- R1-R2；R1-R3；R3-R4；R2-R5；R4-R5基于Loopback接口建立IBGP对等体关系。
- R2为RR，R1为它的Client。
- R3为RR，R1为它的Client。
- R4为RR，R3为它的Client。
- R1将路由11.11.11.0/24发布到BGP。
- R5将分别从R2及R4学习到去往11.11.11.0/24的BGP路由，根据前面几条“路由优选规则”，R5无法做出优选，最终R5将根据本条规则，优选Cluster_List最短的路径（R2所通告的路由）。

优选Cluster_List最短的路由 案例2



- R3-R4 ； R4-R5 ； R3-R5 都基于各自的Loopback接口建立IBGP对等体关系。
- 配置R4为RR，R3为其lient 。
- R1在BGP通告100.0.1.0/24路由，R5将分别从R3及R4学习到去往100.0.1.0/24的BGP路由，它将作何优选？

优选Cluster_List最短的路由 案例2（续）

[R5-bgp]display bgp routing-table 100.0.1.0

BGP routing table entry information of 100.0.1.0/24:
From: 3.3.3.3 (3.3.3.3)
……
AS-path 100, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, best
, select, active, pre 255, IGP cost 2

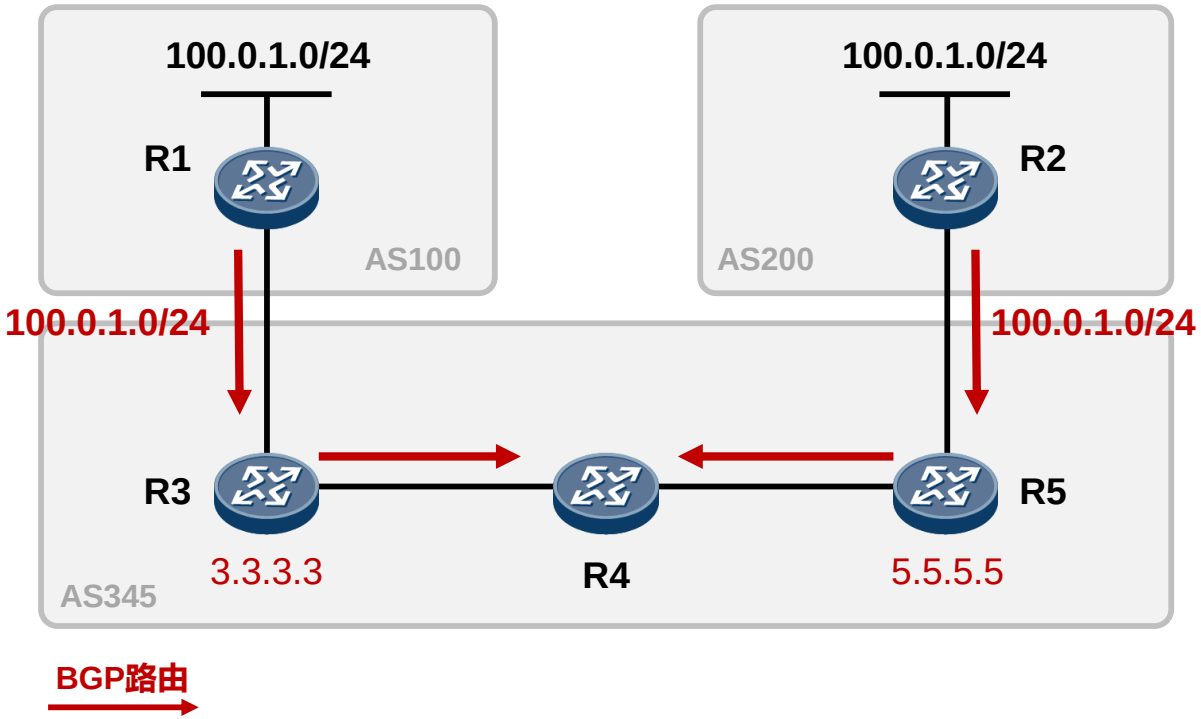
BGP routing table entry information of 100.0.1.0/24:
From: 4.4.4.4 (4.4.4.4)
……
AS-path 100, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, pre
255, IGP cost 2, **not preferred for Cluster List**
Originator: 3.3.3.3
Cluster list: 4.4.4.4

R4所通告的路径没有被优选的原因

目录

HUAWEI

优选Router-ID最小的BGP对等体所通告的路由



在该环境中，如果不部署任何BGP路由策略，R4将从R3及R5都学习到100.0.1.0/24的BGP路由，而且根据前面10条优选规则无法作出决策。最终R4将根据本条规则，优选Router-ID最小的对等体（也就是R3）所通告的路由。

优选Router-ID最小的BGP对等体所通告的路由

[R4-bgp]display bgp routing-table 100.0.1.0

... ...

Paths: 2 available, 1 best, 1 select

BGP routing table entry information of 100.0.1.0/24:

From: 3.3.3.3 (**3.3.3.3**)

... ...

AS-path 12, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, **best**, select, active, pre 255, IGP cost 1

Router-ID PK

BGP routing table entry information of 100.0.1.0/24:

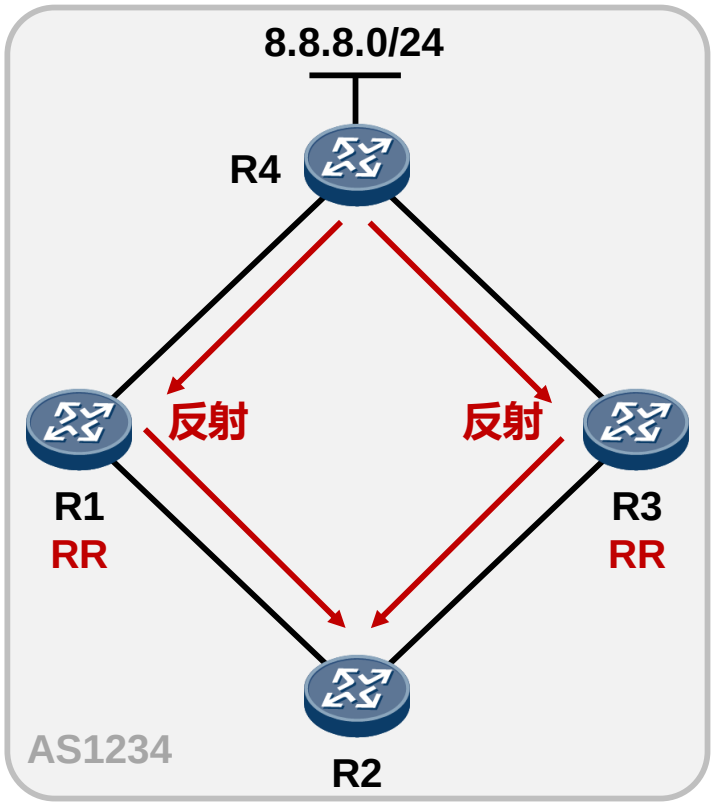
From: 5.5.5.5 (**5.5.5.5**)

... ...

AS-path 12, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, pre 255, IGP cost 1, **not preferred for router ID**

规则补充

- 规则补充：如果路由携带Originator_ID属性，则在本条规则的选路过程中，将比较Originator_ID的大小（不再比较Router-ID），并优选Originator_ID最小的路由。

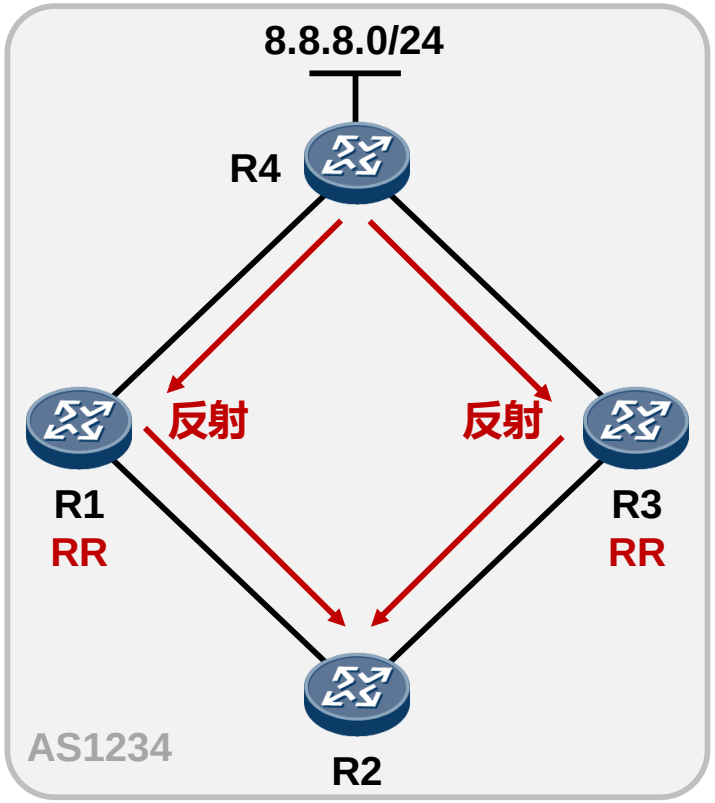


1. Prefered_Value属性值相等；Local_Preference属性值相等；
2. 都不是起源于本地；
3. AS_Path长度一致；
4. Origin属性相同；
5. MED属性值相等；
6. 都是IBGP路由；
7. 两条路由的Next_Hop属性值相同；
8. Cluster_List的长度相等；
9. **由于路由携带Originator_ID，因此本条规则不比较对等体的Router-ID，而是比较路由的Originator_ID属性值。然而两条路由该属性相同。**
10. **继续下一跳规则的比较。**

目录

HUAWEI

优选Peer-IP地址最小的BGP对等体发送过来的路由



1. Prefered_Value属性值相等；Local_Preference属性值相等；
2. 都不是起源于本地；
3. AS_Path长度一致；
4. Origin属性相同；
5. MED属性值相等；
6. 都是IBGP路由；
7. 两条路由的Next_Hop属性值相同；
8. Cluster_List的长度相等；
9. 由于路由携带Originator_ID，因此本条规则不比较对等体的Router-ID，而是比较路由的Originator_ID属性值。然而两条路由该属性相同。
10. **优选Peer-IP地址最小的BGP对等体发送过来的路由，由于R1的地址更小，因此R1通告的8.8.8.0/24路由被优选。**

注意：此处的Peer-IP指的是R2的BGP配置视图中，通过peer命令配置对等体R1及R3时所指定的IP地址。

优选Peer-IP地址最小的BGP对等体发送过来的路由（续）

```
[R2]dis bgp ro 44.44.44.0

BGP routing table entry information of 44.44.44.0/24:
 From: 1.1.1.1 (1.1.1.1)
.....
 AS-path Nil, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, best
, select, active, pre 255, IGP cost 2
 Originator:  4.4.4.4
 Cluster list: 1.1.1.1

 BGP routing table entry information of 44.44.44.0/24:
 From: 3.3.3.3 (3.3.3.3)
.....
 AS-path Nil, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, pre
255, IGP cost 2, not preferred for peer address
 Originator:  4.4.4.4
 Cluster list: 3.3.3.3
```

R3所通告的路径没有被优选的原因

Thank you

www.huawei.com