边界网关协议(BGP)最佳路由反射

目录

简介
背景信息
网络图
理论
IOS-XR实施
配置
配置示例
根路由器上的MPLS流量工程
故障排除
结论

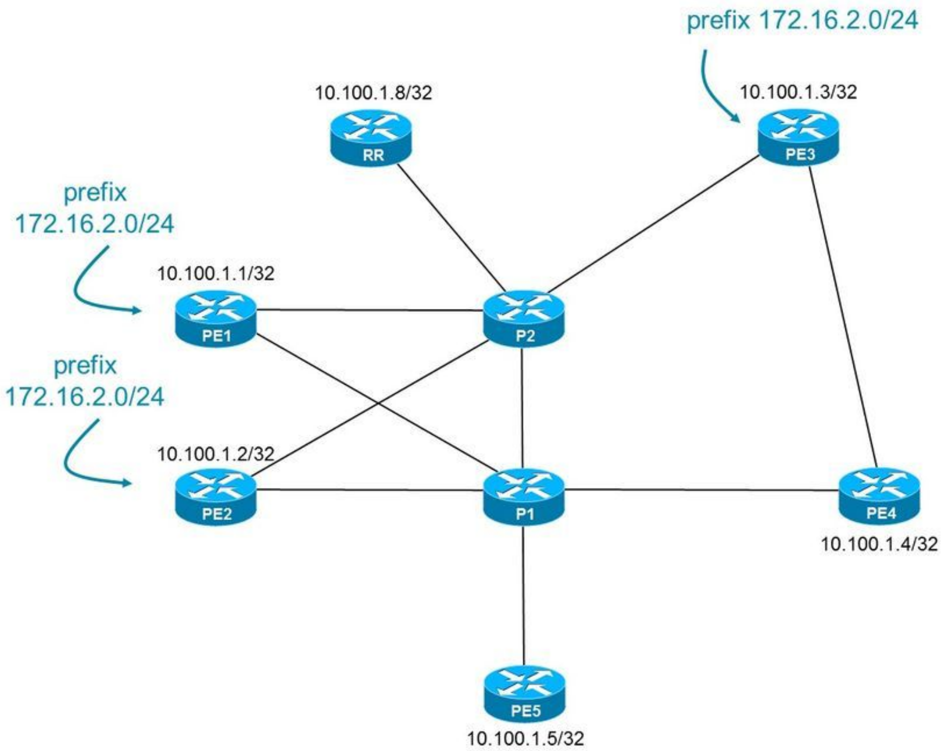
简介

本文档介绍当网络中存在一个或多个路由反射器(RR)以避免iBGP路由器之间出现全网状时，如何影响路由。

背景信息

BGP最佳路径选择算法中的第8步是首选具有最低IGP度量的路径而不是BGP下一跳。因此，如果步骤8之前的所有步骤都相等，则步骤8可以成为决定RR上最佳路径的决定因素。然后，RR的位置决定从RR到通告iBGP路由器的IGP开销。默认情况下，RR仅向其客户端通告最佳路径。根据RR的位置，通告路由器的IGP成本可能更小或更大。如果路径的所有IGP开销相同，则可能最终导致通告路由器具有最低BGP路由器ID的绑定。

网络图



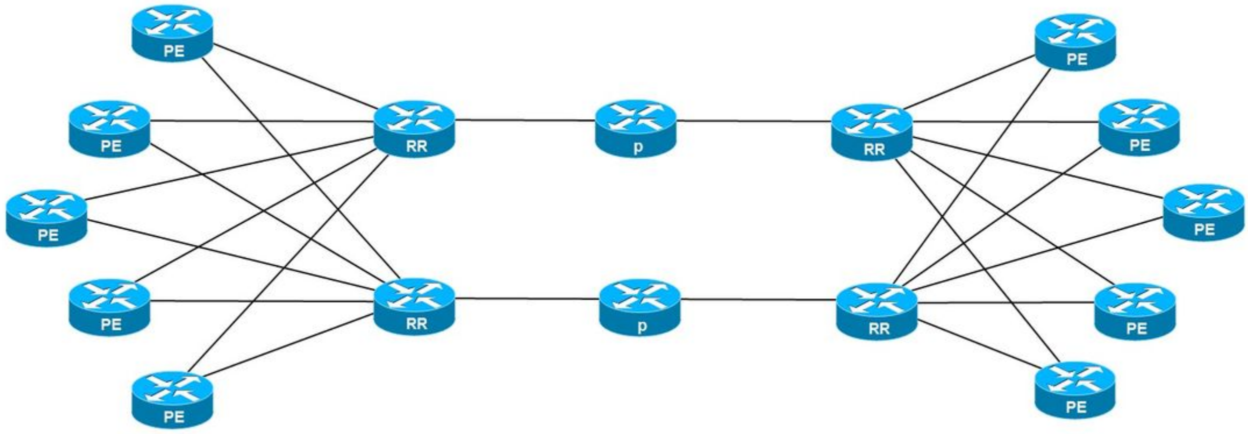
路由器PE1、PE2和PE3通告前缀172.16.2.0/24。如果链路的所有IGP开销相同，则RR将看到来自PE1、PE2和PE3的路径，其IGP开销为2。最后，RR选择来自PE1的路径，因为它的BGP路由器ID更低。这是BGP最佳路径选择算法中的步骤11。结果是，所有PE路由器（包括PE4）都会选择PE1作为前缀172.16.2.0/24的出口PE路由器。从PE4的角度来看，到任何出口PE路由器的较短的IGP路径是到PE3的路径，IGP开销为1。IGP开销到任何其他PE路由器2.对于许多网络，以尽可能短的方式通过中转网络传输流量这一事实非常重要。这称为火薯路由。

RR从PE1选择最佳路径的另一个可能原因是。如果在图像中，链路P2-PE3的内部网关协议(IGP)开销为10，而所有其它链路的IGP开销仍为1，则即使PE3有，RR也不会从PE3选择最佳路径最低的BGP路由器ID。

如果此网络的管理员想要进行热门路由，则必须建立一种机制，以便当网络中有RR时，入口路由器仍可以获知到iBGP网络中最近的出口路由器的路径。BGP功能Add Path可以实现此目的。但是，使用该功能，路由反射器和边界路由器必须有更新的代码来理解该功能。由于BGP最佳路由反射的功能，这不是要求。此功能将允许RR根据RR认为从入口BGP路由器的角度是最佳路径的内容，将最佳路径发送到入口BGP边界路由器。

部署RR时允许热门土豆路由的另一种解决方案是RR的串接放置。这些RR不是专用RR，只运行BGP和IGP。这些内联RR位于转发路径中，并放置在网络中，以便它们拥有自己的一组RR客户端，以便它们能够反映到每个RR客户端的最佳路径，这也是该RR客户端的最佳路径。

如此图所示，RR放在网络中，以便它们可以服务的一小组附近RR客户端。由于网络设计，RR客户端从RR处接收最佳路径，即从其角度来看的最佳路径，以便在网络中进行火薯路由。



理论

BGP最佳路由反射在IETF草案draft-ietf-idr-bgp-optimal-route-reflection中描述。

BGP最佳路由反射解决方案允许路由反射器向特定BGP边界路由器发送特定最佳路径。路由反射器可以选择向不同的BGP边界路由器或一组边界路由器发送不同的最佳路径。边界路由器必须是路由反射器的路由反射器客户端。目标是每个入口BGP边界路由器可以具有不同的出口BGP路由器，用于相同的前缀。如果入口边界路由器始终可以将流量转发到关闭的AS出口路由器，则这允许进行热播路由。

问题是RR通常只向每个BGP边界路由器发送相同的最佳路径，这会阻止热播路由。为了解决此问题，您需要RR能够根据入口BGP边界路由器计算相同前缀的不同最佳路径。RR上的最佳路径计算基于入口BGP边界路由器的位置，因此，RR将从入口边界路由器的角度执行BGP最佳路径计算。只能执行此操作的RR是从IGP的角度，从RR和入口边界路由器所在的位置，掌握网络拓扑完整图的RR。要满足此要求，IGP必须是链路状态路由协议。

在这种情况下，路由反射器可以以入口边界路由器作为树的根来运行最短路径优先(SPF)计算，并计算到其他每台路由器的开销。这样，就知道从入口边界路由器到所有其他出口边界路由器的开销。此特殊SPF计算与另一台路由器作为根桥，称为反向SPF(rSPF)。 只有RR从所有BGP边界路由器获取所有BGP路径时，才能执行此操作。运行rSPF的数量可能与运行RR客户端的数量一样多。这会在一定程度上增加RR上的CPU负载。

该解决方案允许基于BGP最佳路径选择算法进行最佳路径计算，这将导致RR从RR发送路径到的入口边界路由器的角度选择最佳路径。这意味着将根据到BGP下一跳的最短IGP开销选择最佳路径。该解决方案还允许根据某些已配置的策略选择最佳路径。入口边界路由器可以根据某些已配置的策略选择其最佳路径，而不是以最低的IGP开销。该解决方案允许路由反射器在IGP开销（网络上的位置）或某些已配置策略上实施最佳路由反射，或同时实施这两种策略。如果同时部署了这两种策略，则首先应用策略，然后在其余路径上发生基于IGP的最佳路由反射。

IOS-XR实施

IOS-XR实施最多允许三个根节点进行rSPF计算。如果一个更新组中有多个RR客户端，则如果这些RR客户端对不同出口BGP边界路由器具有相同的策略和/或相同的IGP开销，则无需对每个RR客户端计算一个rSPF。后者通常意味着RR客户端共置（可能位于同一POP中）。 如果是这样，则无需将每个RR客户端配置为根。IOS-XR实施允许为冗余目的配置每组RR客户端的三个主根、辅助根和第三根。要使BGP ORR功能应用于任何RR客户端，必须将该RR客户端配置为ORR策略组的一部分。

BGP ORR功能按地址系列启用。

需要链路状态协议。可以是OSPF或IS-IS。

IOS XR仅根据BGP下一跳的IGP开销实施BGP ORR功能，而不是根据某些已配置的策略。

具有相同出站策略的BGP对等体将放置在同一更新组中。RR上的iBGP通常是这种情况。启用功能BGP ORR后，来自不同ORR组的对等体将位于不同的更新组中。这是合乎逻辑的，因为从RR发送到不同BGP ORR组中的RR客户端的更新将不同，因为BGP最佳路径不同。

rSPF计算的结果存储在数据库中。

ORSPF是IOS-XR中BGP ORR功能所需的新组件。ORSPF会注意：

1. 收集链路状态信息并维护链路状态数据库
2. 按策略组运行rSPF并维护SPT
3. 将前缀从SPT下载到带度量的RIB

数据库直接从链路状态IGP或从BGP-LS获取其链路状态信息。

rSPF计算会生成拓扑，显示从RR客户端到区域/级别中任何其他路由器的最短路径。

在拓扑中，挂断每台路由器的路由存储在ORR组策略和AFI/SAFI的特殊RIB表中。此表由RSI创建。该表由rSPF计算的路由填充，其中主根作为根。如果主根变得不可用，则辅助根是根并填充ORR RIB表中的路由。这同样适用于第三根。

配置

所需的最低配置：

1. 需要为BGP的地址系列、特定BGP邻居组启用ORR
2. 对于每组BGP邻居，至少需要配置一个根。或者，可以配置辅助根和第三根。
3. 需要启用ORR路由从IGP重分发到BGP。

配置示例

如第一个映像所示，RR是具有BGP ORR功能的IOS-XR路由器。

所有其他路由器都运行IOS。这些路由器没有BGP ORR功能。

PE1、PE2和PE3在AFI/SAFI 1/1（IPv4单播）中通告前缀172.16.2.0/24。 RR与PE1和PE2的距离相比与PE3的距离相同。所有链路的IGP开销为1。此前缀的最佳路径是R1作为下一跳的路径，因为R1的BGP路由器ID最低。

```
RP/0/0/CPU0:RR#show bgp ipv4 unicast 172.16.2.0/24 bestpath-compare
BGP routing table entry for 172.16.2.0/24
Versions:
 Process           bRIB/RIB  SendTblVer
 Speaker                 34          34
Last Modified: Mar  7 20:29:48.156 for 11:36:44
Paths: (3 available, best #1)
 Advertised to update-groups (with more than one peer):
   0.3
 Path #1: Received by speaker 0
 Advertised to update-groups (with more than one peer):
   0.3
 Local, (Received from a RR-client)
   10.100.1.1 (metric 3) from 10.100.1.1 (10.100.1.1)
     Origin IGP, metric 0, localpref 100, valid, internal, best, group-best
     Received Path ID 0, Local Path ID 1, version 34
     best of local AS, Overall best
  Path #2: Received by speaker 0
 Not advertised to any peer
 Local, (Received from a RR-client)
   10.100.1.2 (metric 3) from 10.100.1.2 (10.100.1.2)
     Origin IGP, metric 0, localpref 100, valid, internal, add-path
     Received Path ID 0, Local Path ID 6, version 33
     Higher router ID than best path (path #1)
  Path #3: Received by speaker 0
 ORR bestpath for update-groups (with more than one peer):
   0.1
 Local, (Received from a RR-client)
   10.100.1.3 (metric 5) from 10.100.1.3 (10.100.1.3)
     Origin IGP, metric 0, localpref 100, valid, internal, add-path
     Received Path ID 0, Local Path ID 7, version 34
     Higher IGP metric than best path (path #1)
```

PE4将收到PE1作为下一跳的路径。因此，PE4没有热土豆路由。

如果要在PE4上进行热播路由，然后对于PE1、PE2和PE3通告的前缀(例如前缀172.16.2.0/24)，则PE1应将PE3作为出口点。这意味着PE4上的路径应是PE3作为下一跳的路径。我们可以使RR使用此ORR配置将具有下一跳PE3的路由发送到PE4。

```
router ospf 1
distribute bgp-ls
 area 0
 interface Loopback0
 !
 interface GigabitEthernet0/0/0/0
  network point-to-point
 !
 !
!

router bgp 1
 address-family ipv4 unicast
 optimal-route-reflection ipv4-orr-group 10.100.1.4
 !
 address-family vpnv4 unicast
 !
 neighbor 10.100.1.1
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  route-reflector-client
 !
```

```
 !
 neighbor 10.100.1.2
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  route-reflector-client
 !
 !
 neighbor 10.100.1.3
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  route-reflector-client
 !
 !
 neighbor 10.100.1.4
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group
   route-reflector-client
 !
 !
 neighbor 10.100.1.5
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  route-reflector-client
 !
 !
!
```

如果IGP为IS-IS:

```
router isis 1
net 49.0001.0000.0000.0008.00
 distribute bgp-ls
address-family ipv4 unicast
metric-style wide
!
interface Loopback0
address-family ipv4 unicast
!
!
interface GigabitEthernet0/0/0/0
address-family ipv4 unicast
!
!
!
```

注意：地址系列链路状态无需全局配置或在BGP邻居下配置。

根路由器上的MPLS流量工程

RR需要在IGP数据库中找到已配置的根地址，以运行rSPF。在ISIS中，路由器ID存在于ISIS数据库中。对于OSPF，OSPF LSA中不存在路由器ID。解决方案是让根路由器通告与路由反射器上已配置的根地址匹配的多协议标签交换(MPLS)TE路由器ID。

对于OSPF，根路由器需要额外配置才能使BGP ORR正常工作。要通告此MPLS TE路由器ID，任何根路由器上都需要最低的MPLS TE配置。确切的最小命令集取决于根路由器的操作系统。根路由

器上的MPLS TE配置需要启用MPLS TE的最小配置，以便OSPF在不透明区域LSA（类型10）中通告MPLS TE路由器ID。

当RR具有MPLS TE路由器ID与已配置的根路由器地址匹配的不透明区域LSA时，rSPF可以运行，并且RR上的BGP可以通告最佳路由。

如果根路由器是IOS路由器，则根路由器上OSPF所需的最低配置为：

```
!
interface GigabitEthernet0/2
 ip address 10.1.34.4 255.255.255.0
 ip ospf network point-to-point
mpls traffic-eng tunnels
!

router ospf 1
mpls traffic-eng router-id Loopback0
 mpls traffic-eng area 0
 router-id 10.200.1.155
 network 10.0.0.0 0.255.255.255 area 0
!
```

请注意：

- MPLS TE在特定OSPF区域中启用
- MPLS TE路由器ID的配置与RR上配置的根地址匹配
- MPLS TE在至少一个接口上配置
- 无需配置RSVP-TE
- 无需在区域内的任何其他路由器上配置MPLS TE

如果根路由器是IOS-XR路由器，则根路由器上OSPF所需的最低配置为：

```
!
router ospf 1
 router-id 5.6.7.8
 area 0
 mpls traffic-eng
  interface Loopback0
 !
 interface GigabitEthernet0/0/0/0
  network point-to-point
 !
 !
mpls traffic-eng router-id 10.100.1.11
!
mpls traffic-eng
!
```

如果根路由器上已配置上述配置，则RR应在OSPF数据库中具有MPLS TE路由器ID。

```
RP/0/0/CPU0:RR#show ospf 1 database


           OSPF Router with ID (10.100.1.99) (Process ID 1)

               Router Link States (Area 0)

Link ID         ADV Router      Age         Seq#       Checksum Link count
```

```
10.1.12.1       10.1.12.1       1297        0x8000002b 0x006145 3
10.100.1.2      10.100.1.2      646         0x80000025 0x00fb6f 7
10.100.1.3      10.100.1.3      1693        0x80000031 0x003ba9 5
10.100.1.99     10.100.1.99     623         0x8000001e 0x00ade1 3
10.200.1.155    10.200.1.155    28          0x80000002 0x009b2e 5

               Type-10 Opaque Link Area Link States (Area 0)

Link ID         ADV Router      Age         Seq#       Checksum Opaque ID
1.0.0.0         10.200.1.155    34          0x80000001 0x00a1ad        0

1.0.0.3         10.200.1.155    34          0x80000001 0x0057ff        3

RP/0/0/CPU0:RR#show ospf 1 database opaque-area adv-router 10.200.1.155


           OSPF Router with ID (10.100.1.99) (Process ID 1)

               Type-10 Opaque Link Area Link States (Area 0)

  LS age: 184
 Options: (No TOS-capability, DC)
 LS Type: Opaque Area Link
 Link State ID: 1.0.0.0
 Opaque Type: 1
 Opaque ID: 0
 Advertising Router: 10.200.1.155
 LS Seq Number: 80000001
 Checksum: 0xa1ad
 Length: 28

   MPLS TE router ID : 10.100.1.4

   Number of Links : 0

 LS age: 184
 Options: (No TOS-capability, DC)
 LS Type: Opaque Area Link
 Link State ID: 1.0.0.3
 Opaque Type: 1
 Opaque ID: 3
 Advertising Router: 10.200.1.155
 LS Seq Number: 80000001
 Checksum: 0x57ff
 Length: 132

   Link connected to Point-to-Point network
     Link ID : 10.100.1.3      (all bandwidths in bytes/sec)
     Interface Address : 10.1.34.4
     Neighbor Address : 10.1.34.3
     Admin Metric : 1
     Maximum bandwidth : 125000000
     Maximum reservable bandwidth global: 0
     Number of Priority : 8
     Priority 0 :                  0  Priority 1 :                  0
     Priority 2 :                  0  Priority 3 :                  0
     Priority 4 :                  0  Priority 5 :                  0
     Priority 6 :                  0  Priority 7 :                  0
     Affinity Bit : 0
     IGP Metric : 1

   Number of Links : 1
```
请注意，MPLS TE路由器ID(10.100.1.4)和OSPF路由器ID不同。

PE4将PE3作为前缀的下一跳（使用正确的IGP度量到下一跳）：

```
PE4#show bgp ipv4 unicast 172.16.2.0
BGP routing table entry for 172.16.2.0/24, version 37
Paths: (1 available, best #1, table default)
 Not advertised to any peer
 Refresh Epoch 1
 Local
   10.100.1.3 (metric 2) from 10.100.1.8 (10.100.1.8)
     Origin IGP, metric 0, localpref 100, valid, internal, best
     Originator: 10.100.1.3, Cluster list: 10.100.1.8
     rx pathid: 0, tx pathid: 0x0
```

PE5仍将PE1作为前缀的下一跳（使用正确的IGP度量到下一跳）：

```
PE5#show bgp ipv4 unicast 172.16.2.0/24
BGP routing table entry for 172.16.2.0/24, version 13
Paths: (1 available, best #1, table default)
 Not advertised to any peer
 Refresh Epoch 1
 Local
   10.100.1.1 (metric 3) from 10.100.1.8 (10.100.1.8)
     Origin IGP, metric 0, localpref 100, valid, internal, best
     Originator: 10.100.1.1, Cluster list: 10.100.1.8
     rx pathid: 0, tx pathid: 0x0
```

故障排除

检验RR上的前缀：

```
RP/0/0/CPU0:RR#show bgp ipv4 unicast 172.16.2.0
BGP routing table entry for 172.16.2.0/24
Versions:
 Process           bRIB/RIB  SendTblVer
 Speaker                 19          19
Last Modified: Mar  7 16:41:20.156 for 03:07:40
Paths: (3 available, best #1)
 Advertised to update-groups (with more than one peer):
   0.3
 Path #1: Received by speaker 0
 Advertised to update-groups (with more than one peer):
   0.3
 Local, (Received from a RR-client)
   10.100.1.1 (metric 3) from 10.100.1.1 (10.100.1.1)
     Origin IGP, metric 0, localpref 100, valid, internal, best, group-best
     Received Path ID 0, Local Path ID 1, version 14
 Path #2: Received by speaker 0
 Not advertised to any peer
 Local, (Received from a RR-client)
   10.100.1.2 (metric 3) from 10.100.1.2 (10.100.1.2)
     Origin IGP, metric 0, localpref 100, valid, internal, add-path
     Received Path ID 0, Local Path ID 4, version 14
 Path #3: Received by speaker 0
 ORR bestpath for update-groups (with more than one peer):
   0.1
 Local, (Received from a RR-client)
   10.100.1.3 (metric 5) from 10.100.1.3 (10.100.1.3)
     Origin IGP, metric 0, localpref 100, valid, internal, add-path
     Received Path ID 0, Local Path ID 5, version 19
```

请注意，add-path已添加到其他非最佳路径，因此除了最佳路径外，还可以通告它们。RR及其客户端之间不使用添加路径功能：不使用路径标识符通告路径。

检验路由是否（仍）通告给特定BGP邻居。

对于邻居PE4，下一跳是前缀172.16.2.0/24的PE3:

```
RP/0/0/CPU0:RR#show bgp ipv4 unicast neighbors 10.100.1.4 advertised-routes
Network            Next Hop        From            AS Path
172.16.1.0/24      10.100.1.5      10.100.1.5      i
172.16.2.0/24      10.100.1.3      10.100.1.3      i

Processed 2 prefixes, 2 paths
```

对于邻居PE5，下一跳是前缀172.16.2.0/24的PE1:

```
RP/0/0/CPU0:RR#show bgp ipv4 unicast neighbors 10.100.1.5 advertised-routes
Network            Next Hop        From            AS Path
172.16.1.0/24      10.100.1.8      10.100.1.5      i
172.16.2.0/24      10.100.1.1      10.100.1.1      i
```

邻居10.100.1.4位于其自己的更新组中，因为ORR策略已实施：

```
RP/0/0/CPU0:RR#show bgp ipv4 unicast update-group
Update group for IPv4 Unicast, index 0.1:
 Attributes:
   Neighbor sessions are IPv4
   Internal
   Common admin
   First neighbor AS: 1
   Send communities
   Send GSHUT community if originated
   Send extended communities
   Route Reflector Client
   ORR root (configured): ipv4-orr-group; Index: 0
   4-byte AS capable
   Non-labeled address-family capable
   Send AIGP
   Send multicast attributes
   Minimum advertisement interval: 0 secs
 Update group desynchronized: 0
 Sub-groups merged: 0
 Number of refresh subgroups: 0
 Messages formatted: 8, replicated: 8
 All neighbors are assigned to sub-group(s)
   Neighbors in sub-group: 0.1, Filter-Groups num:1
    Neighbors in filter-group: 0.3(RT num: 0)
     10.100.1.4

Update group for IPv4 Unicast, index 0.3:
 Attributes:
   Neighbor sessions are IPv4
   Internal
   Common admin
   First neighbor AS: 1
   Send communities
   Send GSHUT community if originated
```

```
   Send extended communities
   Route Reflector Client
   4-byte AS capable
   Non-labeled address-family capable
   Send AIGP
   Send multicast attributes
   Minimum advertisement interval: 0 secs
 Update group desynchronized: 0
 Sub-groups merged: 1
 Number of refresh subgroups: 0
 Messages formatted: 12, replicated: 42
 All neighbors are assigned to sub-group(s)
   Neighbors in sub-group: 0.3, Filter-Groups num:1
    Neighbors in filter-group: 0.1(RT num: 0)
     10.100.1.1                10.100.1.2                10.100.1.3
10.100.1.5
```

show orspf database命令显示ORR组及其根，

```
RP/0/0/CPU0:RR#show orrspf database

ORR policy: ipv4-orr-group, IPv4, RIB tableid: 0xe0000012
Configured root: primary: 10.100.1.4, secondary: NULL, tertiary: NULL
Actual Root: 10.100.1.4


Number of mapping entries: 1
```

带有detail关键字的同一命令为同一OSPF区域中的路由器/前缀提供rSPF根的开销：

```
RP/0/0/CPU0:RR#show orrspf database detail

ORR policy: ipv4-orr-group, IPv4, RIB tableid: 0xe0000012
Configured root: primary: 10.100.1.4, secondary: NULL, tertiary: NULL
Actual Root: 10.100.1.4

Prefix                                      Cost
10.100.1.6                                  2
10.100.1.1                                  3
10.100.1.2                                  3
10.100.1.3                                  2
10.100.1.4                                  0
10.100.1.5                                  3
10.100.1.7                                  3
10.100.1.8                                  4

Number of mapping entries: 9
```

表ID由RSI为ORR组和AFI/SAFI分配：

```
RP/0/0/CPU0:RR#show rsi table-id 0xe0000012

TBL_NAME=ipv4-orr-group, AFI=IPv4, SAFI=Ucast TBL_ID=0xe0000012 in VRF=default/0x60000000 in
VR=default/0x20000000
Refcnt=1
VRF Index=4 TCM Index=1
Flags=0x0 LST Flags=(0x0) NULL

RP/0/0/CPU0:RR#show rib tables

Codes: N - Prefix Limit Notified, F - Forward Referenced
```

```
     D - Table Deleted, C - Table Reached Convergence

VRF/Table              SAFI  Table ID     PrfxLmt   PrfxCnt TblVersion  N F D C
default/default        uni   0xe0000000   5000000        22        128  N N N Y
**nVSatellite/default  uni   0xe0000010   5000000         2          4  N N N Y
default/ipv4-orr-grou  uni   0xe0000012   5000000         9         27  N N N Y
default/default        multi 0xe0100000   5000000         0          0  N N N Y
```

rSPF的根(R4/10.100.1.4)到另一台路由器的开销与PE4上的**show ip route ospf**的开销相同：

```
PE4#show ip route ospf
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
      D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
      N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
      E1 - OSPF external type 1, E2 - OSPF external type 2
      i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
      ia - IS-IS inter area, * - candidate default, U - per-user static route
      o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
      a - application route
      + - replicated route, % - next hop override, p - overrides from PfR

Gateway of last resort is not set

     10.0.0.0/8 is variably subnetted, 20 subnets, 2 masks
O        10.100.1.1/32 [110/3] via 10.1.7.6, 4d05h, GigabitEthernet0/1
O        10.100.1.2/32 [110/3] via 10.1.7.6, 4d05h, GigabitEthernet0/1
O        10.100.1.3/32 [110/2] via 10.1.8.3, 4d06h, GigabitEthernet0/2
O        10.100.1.5/32 [110/3] via 10.1.7.6, 4d05h, GigabitEthernet0/1
O        10.100.1.6/32 [110/2] via 10.1.7.6, 4d05h, GigabitEthernet0/1
O        10.100.1.7/32 [110/3] via 10.1.8.3, 4d06h, GigabitEthernet0/2
                      [110/3] via 10.1.7.6, 4d05h, GigabitEthernet0/1
O        10.100.1.8/32 [110/4] via 10.1.8.3, 4d05h, GigabitEthernet0/2
                      [110/4] via 10.1.7.6, 4d05h, GigabitEthernet0/1
```

BGP ORR组的RIB:

```
RP/0/0/CPU0:RR#show route afi-all safi-all topology ipv4-orr-group

IPv4 Unicast Topology ipv4-orr-group:
-------------------------------------

Codes: C - connected, S - static, R - RIP, B - BGP, (>) - Diversion path
      D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
      N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
      E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
      i - ISIS, L1 - IS-IS level-1, L2 - IS-IS level-2
      ia - IS-IS inter area, su - IS-IS summary null, * - candidate default
      U - per-user static route, o - ODR, L - local, G  - DAGR, l - LISP
      A - access/subscriber, a - Application route
      M - mobile route, r - RPL, (!) - FRR Backup path

Gateway of last resort is not set

o    10.100.1.1/32 [255/3] via 0.0.0.0, 14:14:52, Unknown
o    10.100.1.2/32 [255/3] via 0.0.0.0, 14:14:52, Unknown
o    10.100.1.3/32 [255/2] via 0.0.0.0, 00:04:53, Unknown
o    10.100.1.4/32 [255/0] via 0.0.0.0, 14:14:52, Unknown
o    10.100.1.5/32 [255/3] via 0.0.0.0, 14:14:52, Unknown
o    10.100.1.6/32 [255/2] via 0.0.0.0, 14:14:52, Unknown
o    10.100.1.7/32 [255/3] via 0.0.0.0, 14:14:52, Unknown
o    10.100.1.8/32 [255/4] via 0.0.0.0, 14:14:52, Unknown
```

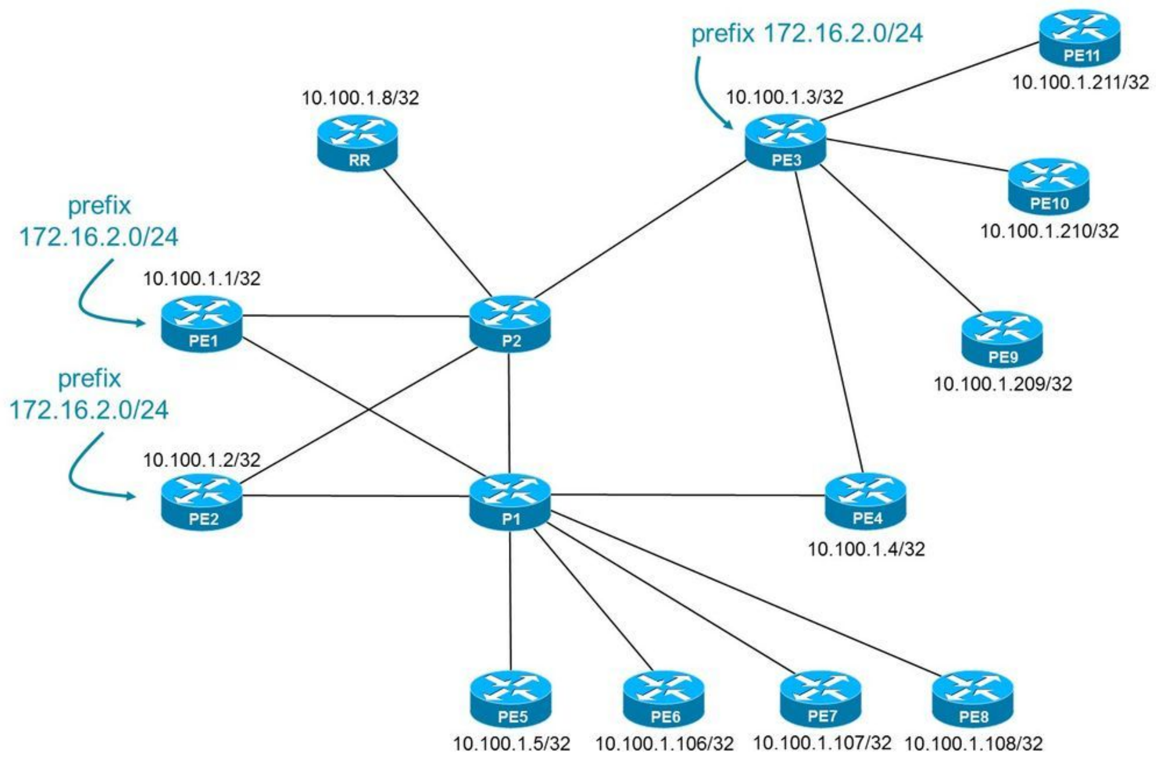
```
RP/0/0/CPU0:RR#show rsi table name ipv4-orr-group

VR=default:

TBL_NAME=ipv4-orr-group, AFI=IPv4, SAFI=Ucast TBL_ID=0xe0000012 in VRF=default/0x60000000 in
VR=default/0x20000000
Refcnt=1
VRF Index=4 TCM Index=1
Flags=0x0 LST Flags=(0x0) NULL
```

show bgp neighbor命令显示对等体是否为ORR根：

```
RP/0/0/CPU0:RR#show bgp neighbor 10.100.1.4

BGP neighbor is 10.100.1.4
 Remote AS 1, local AS 1, internal link
 Remote router ID 10.100.1.4
 Cluster ID 10.100.1.8
 BGP state = Established, up for 01:17:41
 NSR State: None
 Last read 00:00:52, Last read before reset 01:18:30
 Hold time is 180, keepalive interval is 60 seconds
 Configured hold time: 180, keepalive: 60, min acceptable hold time: 3
 Last write 00:00:34, attempted 19, written 19
 Second last write 00:01:34, attempted 19, written 19
 Last write before reset 01:17:43, attempted 19, written 19
 Second last write before reset 01:18:43, attempted 19, written 19
 Last write pulse rcvd  Mar  8 10:20:13.779 last full not set pulse count 12091
 Last write pulse rcvd before reset 01:17:42
 Socket not armed for io, armed for read, armed for write
 Last write thread event before reset 01:17:42, second last 01:17:42
 Last KA expiry before reset 01:17:43, second last 01:18:43
 Last KA error before reset 00:00:00, KA not sent 00:00:00
 Last KA start before reset 01:17:43, second last 01:18:43
 Precedence: internet
 Non-stop routing is enabled
 Multi-protocol capability received
 Neighbor capabilities:
   Route refresh: advertised (old + new) and received (old + new)
   4-byte AS: advertised and received
   Address family IPv4 Unicast: advertised and received
 Received 6322 messages, 0 notifications, 0 in queue
 Sent 5782 messages, 4 notifications, 0 in queue
 Minimum time between advertisement runs is 0 secs
 Inbound message logging enabled, 3 messages buffered
 Outbound message logging enabled, 3 messages buffered

 For Address Family: IPv4 Unicast
 BGP neighbor version 41
 Update group: 0.1 Filter-group: 0.1  No Refresh request being processed
 Route-Reflector Client
 ORR root (configured): ipv4-orr-group; Index: 0
  Route refresh request: received 0, sent 0
 0 accepted prefixes, 0 are bestpaths
 Cumulative no. of prefixes denied: 0.
 Prefix advertised 2, suppressed 0, withdrawn 0
 Maximum prefixes allowed 1048576
 Threshold for warning message 75%, restart interval 0 min
 AIGP is enabled
 An EoR was received during read-only mode
 Last ack version 41, Last synced ack version 0
 Outstanding version objects: current 0, max 2
 Additional-paths operation: None
```

```
 Send Multicast Attributes
 Advertise VPNv4 routes enabled with  option
 Advertise VPNv6 routes is enabled with Local with stitching-RT option

 Connections established 6; dropped 5
 Local host: 10.100.1.8, Local port: 25176, IF Handle: 0x00000000
 Foreign host: 10.100.1.4, Foreign port: 179
 Last reset 01:17:42, due to User clear requested (CEASE notification sent - administrative
reset)
 Time since last notification sent to neighbor: 01:17:42
 Error Code: administrative reset
 Notification data sent:
   None
```

如本图所示，配置了多组RR客户端



有一组RR客户端连接到PE3，另一组连接到P1。每组RR客户端与任何出口BGP边界路由器的距离相等。

```
router bgp 1
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group-1 10.100.1.4 10.100.1.209 10.100.1.210
  optimal-route-reflection ipv4-orr-group-2 10.100.1.5 10.100.1.106 10.100.1.107
 !
…
neighbor 10.100.1.4
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group-1
   route-reflector-client
 !
 !
 neighbor 10.100.1.5
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
```

```
  optimal-route-reflection ipv4-orr-group-2
   route-reflector-client
 !
 !
neighbor 10.100.1.106
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group-2
   route-reflector-client
 !
 !
 neighbor 10.100.1.107
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group-2
   route-reflector-client
 !
 !
 neighbor 10.100.1.108
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group-2
   route-reflector-client
 !
 !
 neighbor 10.100.1.209
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group-1
   route-reflector-client
 !
 !
 neighbor 10.100.1.210
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group-1   route-reflector-client
 !
 !
 neighbor 10.100.1.211
 remote-as 1
 update-source Loopback0
 address-family ipv4 unicast
  optimal-route-reflection ipv4-orr-group-1
   route-reflector-client
 !
 !
!
```

两个组的orspf数据库：

```
RP/0/0/CPU0:RR#show orrspf database detail

ORR policy: ipv4-orr-group-1, IPv4, RIB tableid: 0xe0000012
Configured root: primary: 10.100.1.4, secondary: 10.100.1.209, tertiary: 10.100.1.210
Actual Root: 10.100.1.4

Prefix                                        Cost
10.100.1.1                                    3
```

```
10.100.1.2                                  3
10.100.1.3                                  2
10.100.1.4                                  0
10.100.1.5                                  3
10.100.1.6                                  2
10.100.1.7                                  3
10.100.1.8                                  4
10.100.1.106                                3
10.100.1.107                                3
10.100.1.108                                3
10.100.1.209                                3
10.100.1.210                                3
10.100.1.211                                3
ORR policy: ipv4-orr-group-2, IPv4, RIB tableid: 0xe0000013
Configured root: primary: 10.100.1.5, secondary: 10.100.1.106, tertiary: 10.100.1.107
Actual Root: 10.100.1.5

Prefix                                      Cost
10.100.1.1                                  3
10.100.1.2                                  3
10.100.1.3                                  4
10.100.1.4                                  3
10.100.1.5                                  0
10.100.1.6                                  2
10.100.1.7                                  3
10.100.1.8                                  4
10.100.1.106                                3
10.100.1.107                                3
10.100.1.108                                3
10.100.1.209                                5
10.100.1.210                                5
10.100.1.211                                5

Number of mapping entries: 30
```

如果对于组，主根关闭或无法访问，则辅助根将是实际使用的根。在本例中，组ipv4-orr-group-1的主根不可达。次根成为实际根：

```
RP/0/0/CPU0:RR#show orrspf database ipv4-orr-group-1

ORR policy: ipv4-orr-group-1, IPv4, RIB tableid: 0xe0000012
Configured root: primary: 10.100.1.4, secondary: 10.100.1.209, tertiary: 10.100.1.210
Actual Root: 10.100.1.209

Prefix                                      Cost
10.100.1.1                                  4
10.100.1.2                                  5
10.100.1.3                                  2
10.100.1.5                                  5
10.100.1.6                                  4
10.100.1.7                                  3
10.100.1.8                                  4
10.100.1.106                                5
10.100.1.107                                5
10.100.1.108                                5
10.100.1.209                                0
10.100.1.210                                3
10.100.1.211                                3

Number of mapping entries: 14
```

结论

BGP最佳路由反射(ORR)功能允许在存在路由反射器时在iBGP网络中进行热播路由，而无需边界路由器上使用更新的操作系统软件。前提条件是IGP是链路状态路由协议。