

排除BGP路由抖动故障（递归路由故障）

目录

[简介](#)

[先决条件](#)

[要求](#)

[使用的组件](#)

[背景信息](#)

[规则](#)

[问题](#)

[症状](#)

[递归路由失败](#)

[递归路由失败的原因](#)

[解决方案](#)

[路由惩罚](#)

[相关信息](#)

简介

本文档介绍如何排除因递归路由故障而造成的抖动边界网关协议 (BGP) 路由问题。

先决条件

要求

本文档没有任何特定的要求。

使用的组件

本文档不限于特定的软件和硬件版本。

本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您的网络处于活动状态，请确保您了解所有命令的潜在影响。

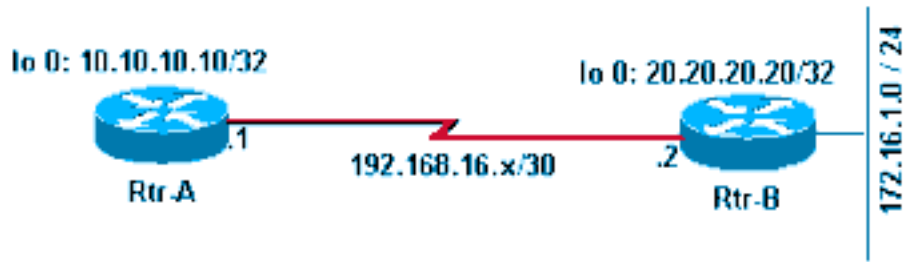
背景信息

本文档介绍如何排除因递归路由故障而造成的抖动边界网关协议 (BGP) 路由问题。

BGP 中递归路由故障的常见症状有：

- 不断在路由表中删除和重新插入 BGP 路由。
- 失去到通过 BGP 获知的目的地的连接。

使用本文档时请参阅此网络图：



网络图

在使用此文档时，请参考以下配置：

Rtr-A

```
hostname RTR-A
!
interface Loopback0
 ip address 10.10.10.10 255.255.255.255
!
interface Serial8/0
 ip address 192.168.16.1 255.255.255.252
!
router bgp 1
 bgp log-neighbor-changes
 neighbor 10.20.20.20 remote-as 2
 neighbor 10.20.20.20 ebgp-multihop 2
 neighbor 10.20.20.20 update-source Loopback0
!
ip route 10.20.20.0 255.255.255.0 192.168.16.2
```

Rtr-B

```
hostname RTR-B
!
interface Loopback0
 ip address 10.20.20.20 255.255.255.255
!
interface Ethernet0/0
 ip address 172.16.1.1 255.255.255.0
!
!
interface Serial8/0
 ip address 192.168.16.2 255.255.255.252
!
router bgp 2
 no synchronization
 bgp log-neighbor-changes
 network 10.20.20.20 mask 255.255.255.255
 network 172.16.1.0 mask 255.255.255.0
 neighbor 10.10.10.10 remote-as 1
 neighbor 10.10.10.10 ebgp-multihop 2
 neighbor 10.10.10.10 update-source Loopback0
 no auto-summary
!
ip route 10.10.10.0 255.255.255.0 192.168.16.1
!
```

规则

有关文档规则的详细信息，请参阅 Cisco 技术提示规则。

问题

症状

发生递归路由故障时可观察到两种症状：

- IP 路由表中通过 BGP 获知的路由持续抖动。为查看抖动，请持续观察路由表几分钟。

```
RTR-A#show ip route
```

```
Codes: C - connected, S - static, I - IGRP, 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, E - EGP
       i - IS-IS, L1 - ISIS level-1, L2 - ISIS level-2, ia - ISIS inter are
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route
```

```
Gateway of last resort is not set
```

```
10.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
B    10.20.20.20/32 [20/0] via 10.20.20.20, 00:00:35
S    10.20.20.0/24 [1/0] via 192.168.16.2
172.16.0.0/24 is subnetted, 1 subnets
B    172.16.1.0 [20/0] via 10.20.20.20, 00:00:35
10.0.0.0/32 is subnetted, 1 subnets
C    10.10.10.10 is directly connected, Loopback0
192.168.16.0/30 is subnetted, 1 subnets
C    192.168.16.0 is directly connected, Serial8/0
```

注：使用 `show ip route | include , 00:00` 命令，以便在处理大型路由表时观察摆动路由。

在等待大约一分钟后，`show ip route` 命令的结果更改为：

```
RTR-A#show ip route
```

```
[...]
```

```
Gateway of last resort is not set
```

```
10.0.0.0/24 is subnetted, 1 subnets
S    10.20.20.0 [1/0] via 192.168.16.2
10.0.0.0/32 is subnetted, 1 subnets
C    10.10.10.10 is directly connected, Loopback0
192.168.16.0/30 is subnetted, 1 subnets
C    192.168.16.0 is directly connected, Serial8/0
```

注：之前的路由表中缺少 BGP 路由。

- 当路由表中存在 BGP 路由时，到这些网络的连接将失败。为了观察此现象，我们在 Rtr-A 的路由表中包含通过 BGP 获知的路由 172.16.1.0/24 时，对有效主机 172.16.1.1 执行 ping 操作，可以发现操作失败。

```
RTR-A#show ip route 172.16.1.0
```

```
Routing entry for 172.16.1.0/24
```

```
Known via "bgp 1", distance 20, metric 0
```

```
Tag 2, type external
Last update from 10.20.20.20 00:00:16 ago
Routing Descriptor Blocks:
* 10.20.20.20, from 10.20.20.20, 00:00:16 ago
  Route metric is 0, traffic share count is 1
  AS Hops 1
```

```
RTR-A#ping 172.16.1.1
```

```
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 172.16.1.1, timeout is 2 seconds:
.....
Success rate is 0 percent (0/5)
RTR-A#
```

递归路由失败

在 Rtr-A 中，观察到 BGP 对等体 10.20.20.20 的路由。两个下一跳之间的路由大约每分钟必然抖动一次。

```
RTR-A#show ip route 10.20.20.20
Routing entry for 10.20.20.20/32
  Known via "bgp 1", distance 20, metric 0
  Tag 2, type external
  Last update from 10.20.20.20 00:00:35 ago
  Routing Descriptor Blocks:
  * 10.20.20.20, from 10.20.20.20, 00:00:35 ago
    Route metric is 0, traffic share count is 1
    AS Hops 1
```

通向BGP对等IP地址的路由是通过BGP本身获取的；因此，它会造成递归路由故障。

在大约一分钟之后，路由更改为：

```
RTR-A#show ip route 10.20.20.20
Routing entry for 10.20.20.0/24
  Known via "static", distance 1, metric 0
  Routing Descriptor Blocks:
  * 192.168.16.2
    Route metric is 0, traffic share count is 1
```

递归路由失败的原因

以下步骤介绍递归路由故障的原因：

1. 请参阅[Rtr-A](#)的配置。在此配置中，静态路由10.20.20.0/24配置为指向直连的下一跳 192.168.16.2。使用此静态路由，将建立与对等体 Rtr-B 10.20.20.20 之间的 BGP 会话。
2. Rtr-B 声明到 Rtr-A 的 BGP 路由 172.16.1.0/24 和 10.20.20.20/32，以其环回 IP 地址 10.20.20.20 用作下一跳。
3. Rtr-A 收到 Rtr-B 声明的 BGP 路由并尝试安装 10.20.20.20/32。这比已在 Rtr-A 中配置为静态路由的 10.20.20.0/24 更为具体。由于首选最长的匹配路由，因此 10.20.20.20/32 优先于 10.20.20.0/24。有关详细信息，请参阅 [Cisco 路由器的路由选择](#)。已安装的路由 10.20.20.20/32的路由表中有下一跳10.20.20.20 (Rtr-B对等IP地址)。由于到 10.20.20.20/32 的路由以自身作为下一跳，因此这会导致递归路由故障。为了了解在此特殊情况下发生递归路由故障的原因，需要了解路由算法的工作原理。对于路由表中下一跳 IP 地址不是路由器直接

连接接口的任何非直接连接路由器，算法都会递归查找路由表，直到找到可以将数据包转发到的直接连接接口。在此特定情况下，Rtr-A 获知到非直接连接网络 10.20.20.20/32 的路由，其非直接连接的下一跳为 10.20.20.20（其本身）。路由算法由于找不到可将发往 10.20.20.20/32 的数据包发送到的任何直接连接接口，因此遇到递归路由环路故障。

4. 路由器检测到该非直连路由 10.20.20.20/32 存在递归路由故障，并从路由表中撤消 10.20.20.20/32。相应地，也会从路由表中撤消通过 BGP 获知的下一跳 IP 地址为 10.20.20.20 的所有路由。
5. 整个过程从 [步骤 1 重复执行](#)。如果发出 `debug ip routing` 命令，则可确认这一点。注：在运行任何 debug 命令之前，请针对特定网络的访问控制列表 (ACL) 运行 debug 命令，以限制 debug 的输出。在本例中，我们配置 ACL 以限制调试输出。

```
RTR-A(config)#access-list 1 permit 10.20.20.20
RTR-A(config)#access-list 1 permit 172.16.1.0
RTR-A(config)#end
```

```
RTR-A#debug ip routing 1
```

```
IP routing debugging is on for access list 1
```

```
00:29:50: RT: add 10.20.20.20/32 via 10.20.20.20, bgp metric [20/0]
00:29:50: RT: add 172.16.1.0/24 via 10.20.20.20, bgp metric [20/0]
00:30:45: RT: recursion error routing 10.20.20.20 - probable routing loop
00:30:45: RT: recursion error routing 10.20.20.20 - probable routing loop
00:30:45: RT: recursion error routing 10.20.20.20 - probable routing loop
00:30:46: RT: recursion error routing 10.20.20.20 - probable routing loop
00:30:46: RT: recursion error routing 10.20.20.20 - probable routing loop
00:30:48: RT: recursion error routing 10.20.20.20 - probable routing loop
00:30:48: RT: recursion error routing 10.20.20.20 - probable routing loop
00:30:50: RT: del 10.20.20.20/32 via 10.20.20.20, bgp metric [20/0]
00:30:50: RT: delete subnet route to 10.20.20.20/32
00:30:50: RT: del 172.16.1.0/24 via 10.20.20.20, bgp metric [20/0]
00:30:50: RT: delete subnet route to 172.16.1.0/24
```

6. 如果路由递归不断失败，则会出现以下错误消息：

```
%COMMON_FIB-SP-6-FIB_RECURSION: 10.71.124.25/32 has too many (8) levels of
recursion during setting up switching info
```

```
%COMMON_FIB-SP-STDBY-6-FIB_RECURSION: 10.71.124.25/32 has too many (8)
levels of recursion during setting up switching info
```

这是因为在启用 MPLS 的网络上发生了 TCP 重新传输。如果 BGP keepalive 消息由于传输链路关闭而无法发送至 BGP 对等体，那么即使 TCP 通过备用路径重新传送失败的消息，邻居 BGP 对等体也不再接受任何 keepalive 数据包，并且最终会导致 BGP 对等体因保持时间到期而关闭。只有在 Catalyst6500 或 Cisco7600 中配置 MPLS 时，才会出现此问题。此信息包含在 Cisco Bug ID [CSCsj89544](#) 中。注意：只有注册的思科用户才能访问内部 Bug 信息和其他工具。

解决方案

下面详细介绍此问题的解决方案。

在 Rtr-A 中为 BGP 对等 IP 地址（在此例中为 10.20.20.20）添加一个具体的静态路由。

```
RTR-A#configure terminal
```

```
Enter configuration commands, one per line. End with CNTL/Z.
```

```
RTR-A(config)#ip route 10.20.20.20 255.255.255.255 192.168.16.2
```

为前缀 10.20.20.20/32 配置静态路由可确保不会在路由表中安装动态获知的 BGP 路由

10.20.20.20/32，从而避免递归路由环路情况。有关详细信息，请参阅 Cisco 路由器的路由选择。

注意：当EBGP对等体配置为使用默认路由相互访问时，不会显示BGP邻居关系。这是为了避免路由抖动和路由环路。

对 172.16.1.1 的 ping 操作可确认该解决方案的有效性。

```
RTR-A#ping 172.16.1.1
```

```
Type escape sequence to abort.
```

```
Sending 5, 100-byte ICMP Echos to 172.16.1.1, timeout is 2 seconds:
```

```
!!!!
```

```
Success rate is 100 percent (5/5), round-trip min/avg/max = 20/24/40 ms
```

路由惩罚

路由惩罚一项 BGP 功能，旨在最大程度地减少抖动路由在互联网上的传播。ISP 建议的值是 Cisco IOS® 的默认值，您只需配置此命令即可启用它。

```
router bgp <AS number>  
  bgp dampening
```

The `bgp dampening` command sets default values for the dampening parameters 诸如半小时= 15分钟、重复使用= 750、抑制时间= 2000和最大抑制时间= 60。这些值可由用户配置，但 Cisco 建议保持这些默认值。

相关信息

- [“#%BGP-3-INSUFCHUNKS: Insufficient chunk pools for aspath”错误消息意味着什么？](#)
- [BGP 邻居为什么在空闲、连接及活动状态之间切换？](#)
- [IP路由/BGP支持](#)
- [思科技术支持和下载](#)

关于此翻译

思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。