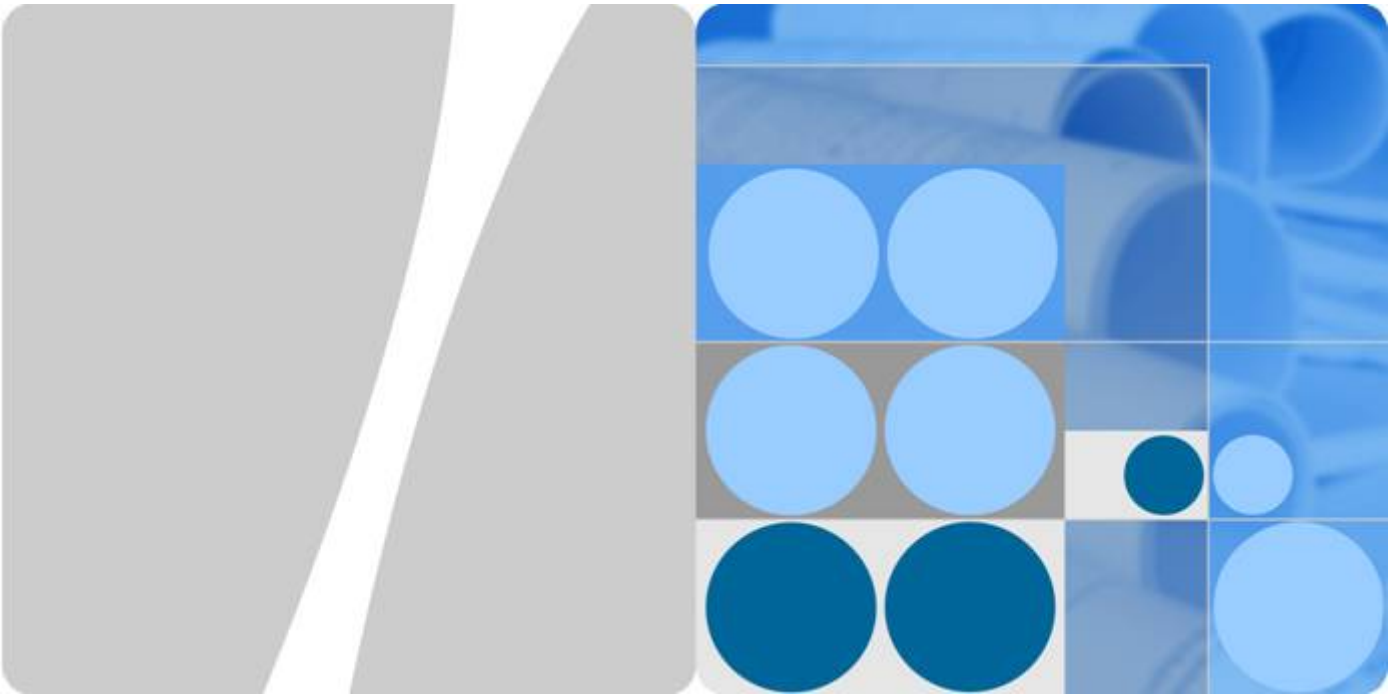


HUAWEI NetEngine80E/40E 路由器

V600R003C00

故障处理-IP 转发和路由

文档版本　02

发布日期　2011-09-10

华为技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： http://www.huawei.com

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

前 言

概述

说明

- 手册中所使用的链路接口编号和链路类型以 NE40E-X8 为例，实际使用时以现网设备为准。
- NE80E/40E 系列中的非 X1/X2 设备的线路处理板称为 LPU，交换网板称为 SFU；X1/X2 设备没有 LPU 和 SFU，由 NPU 集中实现报文交换和转发功能。

本文档针对的 HUAWEI NetEngine80E/40E 各类业务，从常见故障及其处理方法、故障处理案例、FAQ 等方面分析介绍了故障的处理过程。

本文档提供了 HUAWEI NetEngine80E/40E 故障的处理流程和方法。

产品版本

与本文档相对应的产品版本如下所示。

产品名称	产品版本
HUAWEI NetEngine80E/40E 路由器	V600R003C00

读者对象

本文档主要适用于以下工程师：

- 系统维护工程师
- 调测工程师
- 网络监控工程师

符号约定

在本文中可能出现下列标志，它们所代表的含义如下。

符号	说明
 危险	以本标志开始的文本表示有高度潜在危险，如果不能避免，会导致人员死亡或严重伤害。
 警告	以本标志开始的文本表示有中度或低度潜在危险，如果不能避免，可能导致人员轻微或中等伤害。
 注意	以本标志开始的文本表示有潜在风险，如果忽视这些文本，可能导致设备损坏、数据丢失、设备性能降低或不可预知的结果。
窍门	以本标志开始的文本能帮助您解决某个问题或节省您的时间。
说明	以本标志开始的文本是正文的附加信息，是对正文的强调和补充。

命令行格式约定

格式	意义
粗体	命令行关键字（命令中保持不变、必须照输的部分）采用**加粗**字体表示。
斜体	命令行参数（命令中必须由实际值进行替代的部分）采用*斜体*表示。
[ ]	表示用“[ ]”括起来的部分在命令配置时是可选的。
{ x | y | ... }	表示从两个或多个选项中选取一个。
[x | y | ...]	表示从两个或多个选项中选取一个或者不选。
{ x | y | ... } *	表示从两个或多个选项中选取多个，最少选取一个，最多选取所有选项。
[x | y | ...] *	表示从两个或多个选项中选取多个或者不选。
&<1-n>	表示符号&的参数可以重复 1 ～ n 次。
#	由“#”开始的行表示为注释行。

修订记录

修改记录累积了每次文档更新的说明。最新版本的文档包含以前所有文档版本的更新内容。

文档版本 02 (2011-09-10)

第二次正式发布，相对于上一版本无变更。

文档版本 01 (2011-05-30)

第一次正式发布

目 录

1 ARP 问题

关于本章

1.1 以太链路 ARP 问题的定位思路

1.1.1 常见原因

本类故障的常见原因主要包括：

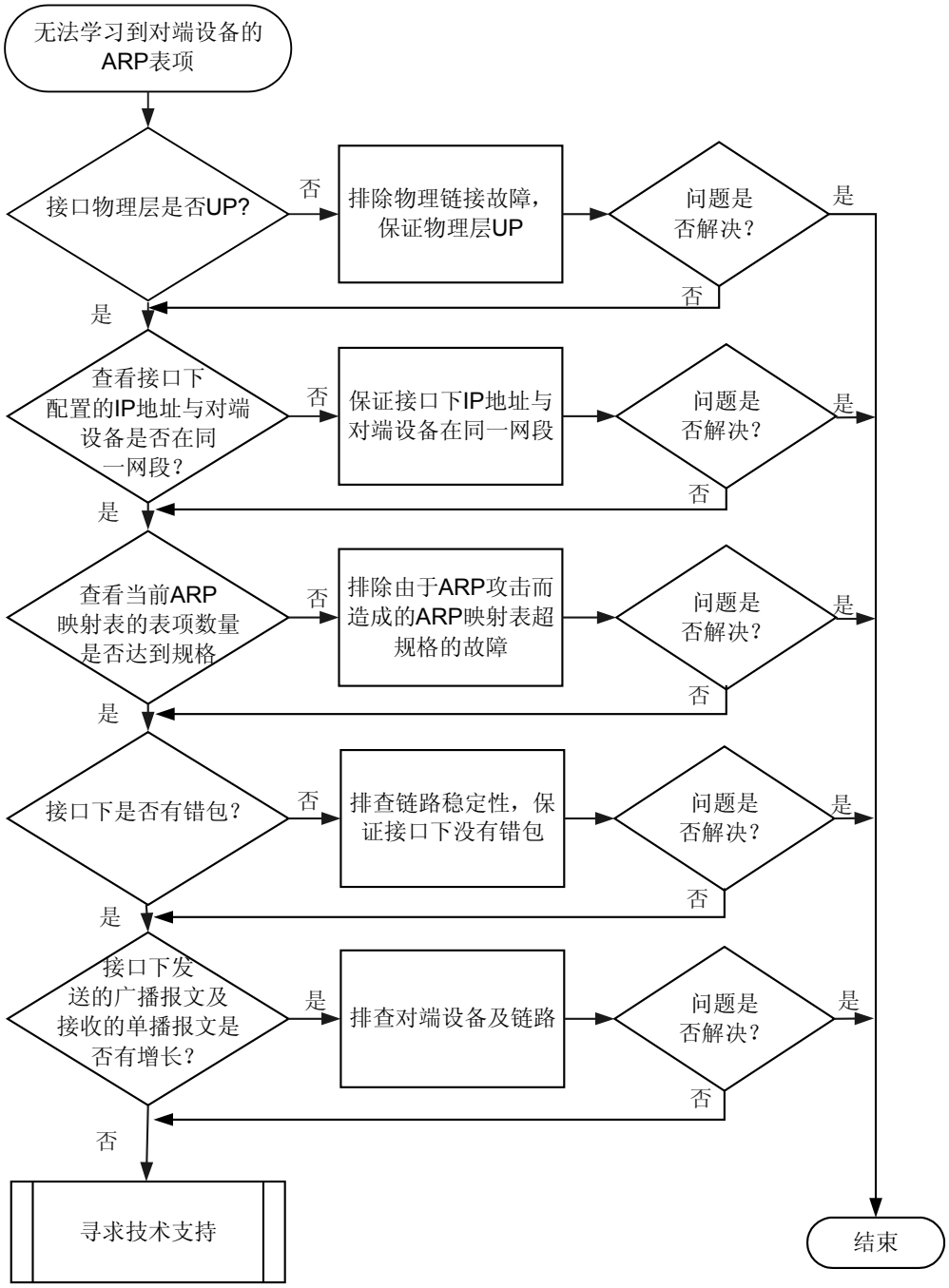
- 接口物理层未正常 Up。
- 接口下配置的 IP 地址与对端设备不在同一网段。
- 设备受到 ARP 攻击。
- 链路传输不稳定或光功率不足。
- 设备软件故障。

1.1.2 故障诊断流程

NE80E/40E 与对端设备相连，发现无法学习到对端设备的 ARP 表项。

详细处理流程如图 1-1 所示。

图 1-1 无法学习到对端设备的 ARP 表项的故障诊断流程图



1.1.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 在用户视图下执行命令 **display interface** *interface-type interface-number* 查看接口物理层状态：

说明

执行命令 **display interface** *interface-type interface-number* 可完成以下步骤中接口信息的查看。

- 如果显示为 Down，请参考相关故障处理案例进行排查。
- 如果显示为 Up，请执行步骤 2。

步骤 2 查看本端和对端接口信息中的接口 IP 地址字段“Internet Address”：

- 如果两端接口的地址不在同一网段，请重新配置接口地址，使其在同一网段。
- 如果两端接口的地址在同一网段，请执行步骤 3。

步骤 3 在用户视图下配置命令 **display arp slot** *slot-id* ，查看指定接口板的 ARP 映射表，看当前表项数量是否达到规格。如果出现达到规格的情况，请参考相关故障处理案例排查设备是否受到 ARP 攻击。

如果故障仍不能排除，请执行步骤 4。

步骤 4 查看接口下的状态信息中的“CRC”字段计数是否有增长，如果有增长则表明接口下存在大量的错包。请对链路质量进行排查，如是否存在传输不稳定或发送光功率不足等情况。

如果故障仍不能排除，请执行步骤 5。

步骤 5 查看接口信息中“output”发送的广播报文字段“Broadcast”是否有增长。

在查看以下信息之前，需要首先在本端设备上 ping 对端设备的 IP 地址，用来触发本端设备发送 ARP 请求报文。

- 如果广播报文没有增长，则说明 ARP 请求报文没有发出，即本端设备存在故障，请执行步骤 7。
- 如果广播报文有增长，请排查对端设备及链路：
  - 如果对端设备及链路故障，请具体定位并排除故障。
  - 如果对端设备及链路均正常，请执行步骤 6。

步骤 6 查看接口信息中接收的单播报文字段“Unicast”是否有增长。

- 如果单播报文没有增长，则说明设备没有收到 ARP 响应报文，可能是由于本端设备发出的 ARP 单播报文不符合规范而被对端丢弃，请执行步骤 7。
- 如果单播报文有增长，则说明设备接收到了 ARP 响应报文，则可判断为本端设备软件内部处理模块的故障，请执行步骤 7。

步骤 7 请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

1.1.4 相关告警与日志

相关告警

无

相关日志

无

1.2 相关案例

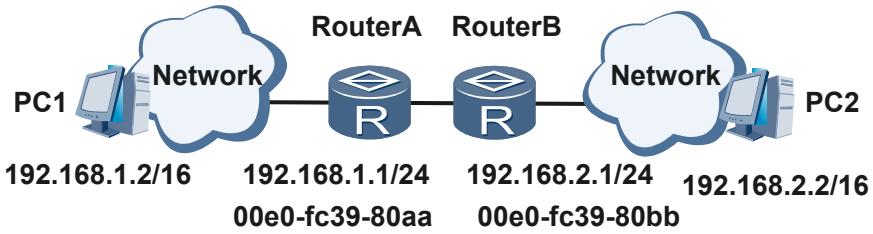
1.2.1 未使能 ARP proxy 导致同一网段中的 PC 不能互访

网络环境

在图 1-2 的网络中，PC1 和 PC2 同属于 192.168.0.0/16 网段，两台路由器上均配置了到对方网段的静态路由。

配置完成后，发现 PC1 ping 不通 PC2。

图 1-2 同网段 PC 不能互访组网图



故障分析

1. 在 PC1 上执行命令 **arp -a** 查看所有 ARP 表项，发现没有 PC2 的 IP 地址与 MAC 地址的映射。说明执行 ping 命令时，ARP 表项并没有自动学习到。这是因为 RouterA 接收到 PC1 的 ARP 请求报文后，发现请求报文的目标 IP 地址不是本地接口的 IP 地址，将 ARP 请求报文丢弃。

   说明

   通常情况下，当路由器收到 ARP 请求报文时，将进行检查，看该 ARP 请求的目的地址是否是自己。如果是，发出 ARP 应答报文；如果不是，丢弃该报文。

   若使能路由式 Proxy ARP 功能后，路由器接收到目的地址不是自己的 ARP 请求报文，并不立即丢弃该报文，而是查找路由表。如果有到达该目的地址的路由，在满足代理条件时，将自己的 MAC 地址发送给 ARP 请求方。ARP 请求方就将到该目的地址的报文发送给路由器，路由器再将其转发出去。

操作步骤

步骤 1 在 RouterA 和 RouterB 上执行命令 **system-view**，进入系统视图。

步骤 2 在 RouterA 和 RouterB 上执行命令 **interface** *interface-type interface-number*，进入路由器与 PC 相连的接口的接口视图。

步骤 3 在 RouterA 和 RouterB 上执行命令 **arp-proxy enable**，使能接口的路由式 Proxy ARP 功能。

步骤 4 在 PC1 上执行命令 **ping 192.168.2.2**，ping 对端 PC2 的 IP 地址。然后在 PC1 上执行命令 **arp －a**，发现 PC2 的 IP 地址对应的 MAC 地址为 RouterA 与 PC1 相连的接口的 MAC 地址。

```
C:\Documents and Settings\Administrator>arp -a
Interface: 192.168.1.2 --- 0x2
Internet Address      Physical Address      Type
192.168.2.2            00e0-fc39-80aa       dynamic
```

完成上述操作后，在 PC1 上 ping 对端 PC2，可以 ping 通，故障排除。

----结束

案例总结

Proxy ARP 主要是通过代理的方式来解决网络互通问题，分为如下三种 Proxy ARP 方式，不同的方式可以解决不同问题。

- 路由式 Proxy ARP：解决同一网段不同物理网络上计算机的互通问题。
- VLAN 内 Proxy ARP：解决相同 VLAN 内，且 VLAN 配置用户隔离后的网络上计算机互通问题。
- VLAN 间 Proxy ARP：解决不同 VLAN 之间对应计算机的二层互通问题。

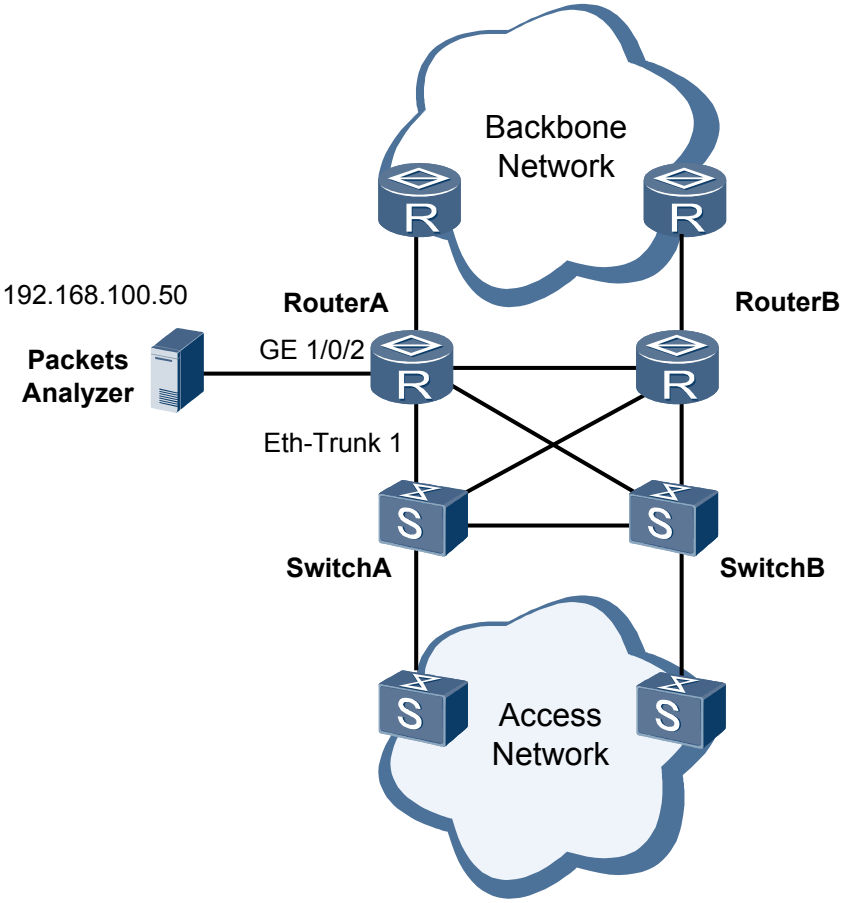
1.2.2 静态 ARP 配置不合理，导致割接后网页无法打开

网络环境

用户通过 RouterA 和 RouterB 接入骨干网，访问网络。接入侧流量通过 Eth-Trunk 1 上送到 RouterA。

在 RouterA 上通过 GE1/0/2 连接一台报文分析设备，用以分析接入侧与骨干网的报文情况。该报文分析设备完成分析后，会将接收到的报文从入接口原样返回。

图 1-3 静态 ARP 配置不合理导致割接后网页无法打开组网图



当 RouterA 割接完成后，出现网页无法打开的现象。

故障分析

1. 在 SwitchA 上，查看路由信息。

   发现 SwitchA 的路由表正常。出现目标网络 IP 地址的报文通过负载分担至 RouterA。

2. 在 RouterA 上，执行命令 **display current-configuration** 检查配置。发现相应接口下配置了流策略：

```
[RouterA] display current-configuration
#
interface Eth-Trunk1
 description HUAWEI-RouterA
 ip address 10.1.1.1 255.255.255.252
 traffic-policy Fivedai inbound
 pim sm
 ospf cost 1000
 mpls
 mpls ldp
#
traffic policy Fivedai
classifier Fivedai behavior Fivedai
traffic classifier Fivedai operator or
  if-match acl 3100
traffic behavior Fivedai
  redirect ip-nexthop 192.168.100.50
```

```
acl number 3100
  rule 5 permit tcp destination-port eq www
```

所有从 SwitchA 上送的访问网页报文都会被重定向到报文分析设备进行分析。

3. 执行命令 **display traffic policy statistics**，查看该流策略命中情况。发现所有访问网页的报文，都匹配了该流策略：

```
[RouterA] display traffic policy statistics interface GigabitEthernet 1/0/0
inbound verbose rule-based
    Info: The statistics is shared because the policy is shared.
    Interface: GigabitEthernet1/0/0
    Traffic policy inbound: Fivedai
    Traffic policy applied at 2009-07-02 21:04:40
    Statistics enabled at 2009-07-10 02:39:28
    Statistics last cleared: Never
    Rule number: 5 IPv4, 0 IPv6
    Current status: OK!

Classifier: Fivedai operator or
    if-match ACL 3100
      rule 5 permit tcp destination-port eq www
       584,574,637 bytes, 4,505,760 packets
Last 30 seconds rate 1,858 pps, 2,270,232 bps
```

4. 在 RouterA 上，执行命令 **display current-configuration**，检查 GE1/0/2 和报文分析设备的相关配置。

```
[RouterA] display current-configuration
#
arp static 192.168.100.50 00e0-1111-1111
```

发现 RouterA 上配置了报文分析设备的 MAC 地址和 IP 地址的静态 ARP 表项。

当报文从 GE1/0/2 发往报文分析设备时，报文的目的 MAC 为 00e0-1111-1111。报文分析设备将报文原样返回后，GE1/0/2 收到的报文目的 MAC 地址仍为 00e0-1111-1111，不等于 GE1/0/2 的接口 MAC 地址（2222-2222-2222），因此会将报文丢弃。

至此，问题定位。

操作步骤

步骤 1 在 RouterA 上，执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **undo arp static 192.168.100.50 00e0-1111-1111**，删除原静态 ARP 配置。

步骤 3 执行命令 **arp static 192.168.100.50 2222-2222-2222**，重新配置静态 ARP，将 MAC 地址修改为 GE1/0/2 的 MAC 地址。

配置完成后，GE1/0/2 发送给报文分析设备的报文目的 MAC 地址为 GE1/0/2 的 MAC 地址。因此，报文分析设备原样返回的报文 MAC 地址等于 GE1/0/2 的 MAC 地址，报文可以被接收并继续转发。用户可以正常打开网页，故障排除。

----结束

案例总结

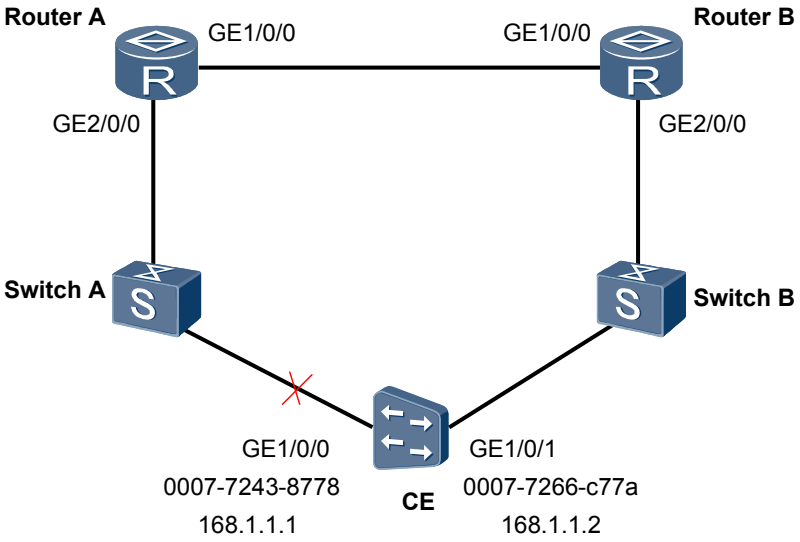
在配置与特殊设备相关的静态 ARP 时，要考虑到特殊设备对报文的处理造成的影响。

1.2.3 由于接收到不符合 RFC 规范的免费 ARP 报文导致 NE80E/40E 的 ARP 表项无法及时更新

网络环境

其他厂商的 CE 设备通过两台交换机以 VRRP 的方式连接到两台 NE80E/40E 上，RouterA 和 RouterB 分别作为主设备和备设备；CE 设备上配置了主备接口备份，GE1/0/0 和 GE1/0/1 分别为主、备接口。组网图如下：

图 1-4 VRRP 典型组网图



当 SwitchA 和 CE 之间的链路断开后，业务出现了中断，20 分钟后才恢复，且在业务中断期间 CE 设备 ping 外网服务器或 VRRP 主设备时有丢包现象。

故障分析

1. 为检查 RouterA 与 CE 设备之间的链路连通性，首先修改 RouterA 的接口 GE1/0/0、RouterB 的接口 GE1/0/0 和 GE2/0/0 的 MTU 值为 2000，然后在 RouterA 上以 100 个 Jumbo 帧（长度大于 1518 的帧）ping CE 设备的接口 GE1/0/1 的 IP 地址，发现丢弃了一个。

   说明

   由于现网 Jumbo 帧数量较少，用它进行 ping 包检查设备的连通性能够更清晰定位；增大设备的 MTU 值是为了能够使 Jumbo 帧被顺利转发，但在修改 MTU 值之前需要首先评估对现网业务是否有影响，如果有影响，请不要采用此方法。

```
<RouterA>ping -c 100 -s 2000 -m 10 -vpn cnc_signal 168.1.1.2
  PING 168.1.1.2: 2000  data bytes, press CTRL_C to break
    Request time out
......
  --- 168.1.1.2 ping statistics ---
    100 packet(s) transmitted
    99 packet(s) received
    1.00% packet loss
    round-trip min/avg/max = 4/15/928 ms
```

2. 在 RouterB 上查看接口 GE1/0/0 和 GE2/0/0 的状态信息。从计数情况来看，RouterB 已经正确接收到了报文，并全部发送给 CE 设备，但回程方向只有 99 个报文进入

RouterB，并且 RouterB 全部将其转发给了 RouterA，至此可判断两台设备转发正常，CE 设备有丢包。

```
<RouterB>display interface GigabitEthernet 1/0/0
GigabitEthernet1/0/0 current state : UP
Description : Cc-Cc#GE#2585.2/0/8-HX.1/0/4-B, Switch Port
The Maximum Transmit Unit is 2000 bytes
Statistics last cleared: 2009-08-04 04:18:10
    Last 30 seconds input rate: 12240 bits/sec, 18 packets/sec
    Last 30 seconds output rate: 830976 bits/sec, 476 packets/sec
    Input: 257912 bytes, 910 packets
    Output: 4222228 bytes, 18610 packets
    Input:
      Unicast: 354, Multicast: 471
      Broadcast: 85, Jumbo: 100
      CRC: 0, Symbol: 0
      Overrun: 0 , InRangeLength: 0
      LongPacket: 0 , Jabber: 0,  Alignment: 0
      Fragment: 0, Undersized Frame: 0
      RxPause: 0
    Output:
      Unicast: 18220, Multicast: 105
      Broadcast: 285, Jumbo: 99
      Lost: 0, Overflow: 0, Underrun: 0
      TxPause: 0
<RouterB>display interface GigabitEthernet 2/0/0
GigabitEthernet2/0/0 current state : UP
Description : CC-CC#GE#2585.H40.1/0/4-2511.H7810.0/0/11-B, Switch Port
The Maximum Transmit Unit is 2000 bytes
Statistics last cleared: 2009-08-04 04:18:10
    Last 30 seconds input rate: 307248 bits/sec, 156 packets/sec
    Last 30 seconds output rate: 304576 bits/sec, 156 packets/sec
    Input: 1458695 bytes, 5993 packets
    Output: 1448674 bytes, 5992 packets
    Input:
      Unicast: 5783, Multicast: 1
      Broadcast: 209, Jumbo: 99
      CRC: 0, Symbol: 0
      Overrun: 0 , InRangeLength: 0
      LongPacket: 0 , Jabber: 0,  Alignment: 0
      Fragment: 0, Undersized Frame: 0
      RxPause: 0
    Output:
      Unicast: 5853, Multicast: 111
      Broadcast: 28, Jumbo: 100
      Lost: 0, Overflow: 0, Underrun: 0
      TxPause: 0
```

3. 在 RouterB 上利用命令 **display arp interface gigabitethernet 2/0/0** 查看接口的动态 ARP 映射表，显示未将 CE 设备接口 GE1/0/1 的 MAC 地址刷新至映射表。
4. 在 RouterB 上抓包，发现 CE 设备在不间断的发送免费 ARP 报文，但报文的格式不符合 RFC 协议规定，我司设备收到这种报文后会直接丢弃，而无法更新 ARP 表项，致使设备到 20 分钟的老化时间后才会自动更新 ARP 表项，业务得以恢复。

操作步骤

步骤 1 在用户视图下执行命令 **reset arp all**，直接复位 ARP 表项，使之强制性的重新学习，由于 CE 设备响应 ARP 请求报文没有问题，所以业务能够恢复正常。

请在 RouterA 和 RouterB 上进行以上操作。

步骤 2 执行命令 **system-view**，进入系统视图。

步骤 3 执行命令 **interface gigabitethernet 1/0/0**，进入接口视图。

步骤 4 执行命令 **mtu**，修改接口的 MTU 值为初始配置。

----结束

案例总结

免费 ARP 报文的不规范会造成其他互连设备 ARP 表项不能及时更新。

2 IP 转发故障处理

关于本章

2.1 Ping 不通问题的定位思路

2.1.1 常见原因

Ping 不通指的是在源端发送请求报文后，在一定的时间范围内没有收到目的端对该请求的回应。

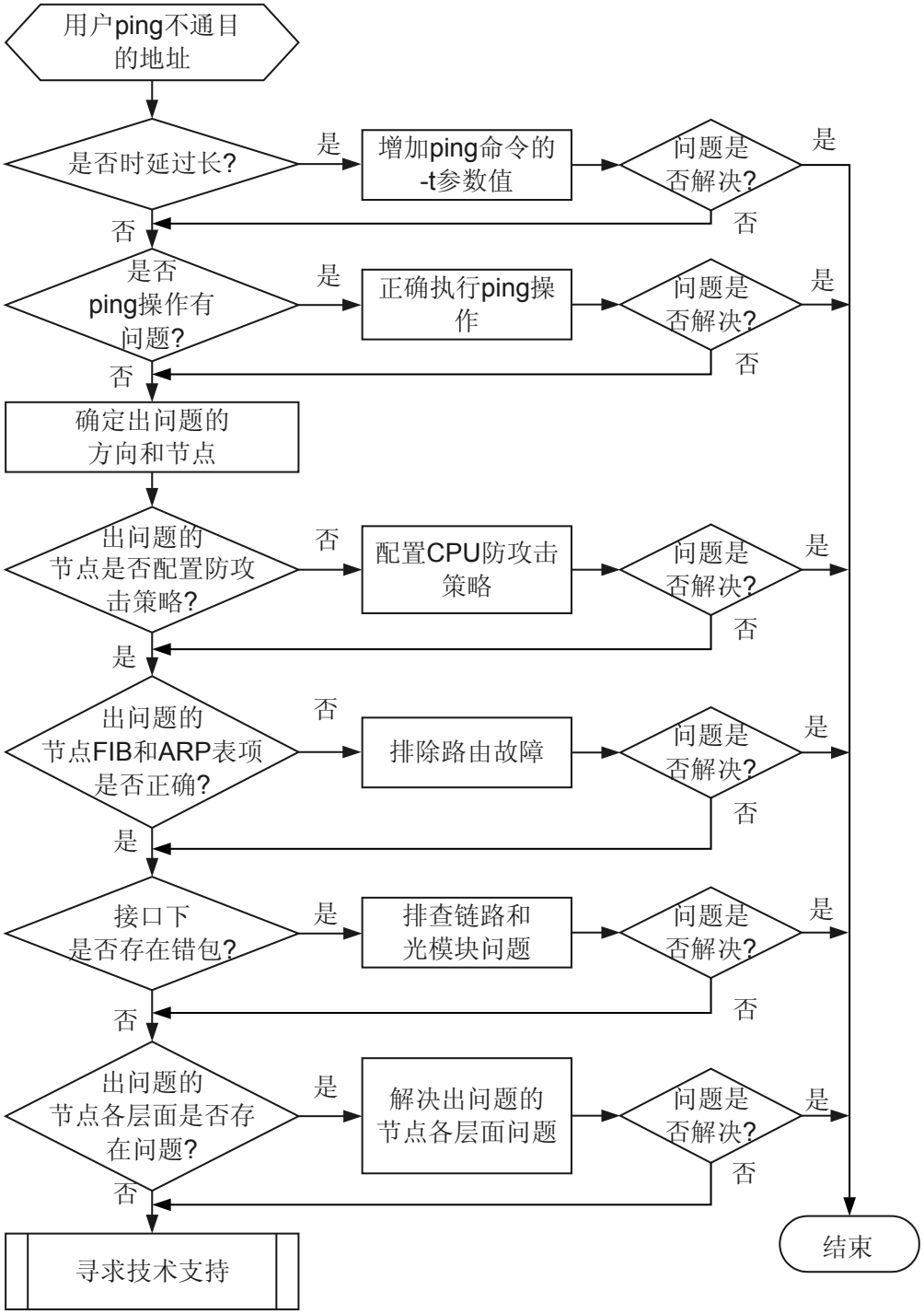
本类故障的常见原因主要包括：

- 链路传输时延较长。由于传输时延长，虽然源端接收到了目的端的回应报文，但已经超过等待时限而造成的 PING 不通的现象。
- 操作不当。例如当 Ping 报文过大时，报文的出接口 MTU 值较小，但是又设置了不可分片的功能等。
- 路由表项或 ARP 表项（ARP 表项只针对以太链路）有问题。
- 硬件故障。

2.1.2 故障诊断流程

可按照故障诊断流程图 2-1 排除此类故障。

图 2-1 IP 转发不通故障诊断流程图



2.1.3 故障处理步骤

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查是否链路传输时延较长导致 Ping 不通

执行 **ping -t** *time-value* **-v** *destination-address* 命令确认是否链路传输时延较长导致 Ping 不通。

说明

-t 参数用来设置等待目的端响应报文的超时时间，默认为 2000ms；**-v** 参数用来显示接收到的非期望回应报文，缺省是不显示。

Ping 的原理是在特定时间内收到回应报文就表示能 Ping 通，否则就表示 Ping 不通。因此首先通过设置 Ping 的**-t** 和**-v** 参数排除由于传输时延较长造成的 Ping 不通。如果是传输时延较长导致的丢包会打印如下信息：

```
<HUAWEI> ping -v -t 1 10.1.1.1
  PING 10.1.1.1: 56  data bytes, press CTRL_C to break
    Request time out
Error:  Sequence number = 1 is less than the correct = 2!
```

出现如上提示信息则说明是链路传输时延较长造成的 Ping 不通，请增大**-t** 参数的值。

如果**-t** 值较大时才能 Ping 通，请检查设备的状态和链路情况，排除网络和设备异常导致的 Ping 不通情况。

如果增大**-t** 参数的值，仍 Ping 不通，请执行步骤 2。

说明

如果在 PE 端 Ping 私网地址，需使用命令 **ping -vpn-instance** *vpn-name destination-address*，其中的**-vpn-instance** *vpn-name* 指 Ping 的目的地址所属的 VPN 实例。

步骤 2 检查是否操作错误

1. 检查是否执行了 **ping -f**，如果执行此操作，则该 Ping 报文不支持分片，此时需要检查路径上出接口的 MTU 值是否小于 Ping 的报文大小，如果 MTU 小于 Ping 报文大小，则丢失为正常现象，请更改 Ping 报文大小小于 MTU 值，否则请执行子步骤 b。查看接口的 MTU 值可执行如下命令：

   ```
   <HUAWEI> display interface gigabitethernet 1/0/0
   GigabitEthernet1/0/0 current state : UP
   Line protocol current state : UP
   Last line protocol up time: 2008-08-30 10:56:22
   Description:HUAWEI, GigabitEthernet6/2/0 Interface
   Route Port,The Maximum Transmit Unit is 1500, Hold timer is 10(sec)
   ```

2. 请检查是否是执行了 **ping -i**，即指定出接口。如果指定的出接口是以太链路等广播类型接口，只支持 Ping 的目的地址是直连接口地址的情况。如果不满足此条件，请更改 Ping 的操作。如果操作无误后故障仍然存在，请执行步骤 3。

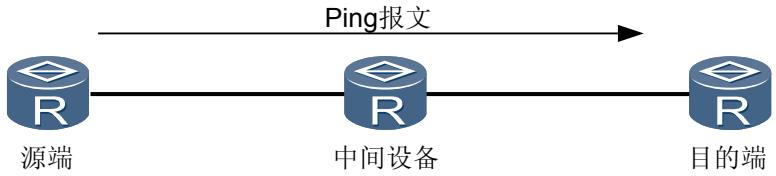
说明

f 参数用来设置该 Ping 报文不支持分片。**-i** *interface-name* 参数用来指定 Ping 报文的出接口，此时会把目的 IP 地址作为下一条地址进行处理。

步骤 3 确定出问题的方向

Ping 应用场景包含三个角色：Ping 报文发起端（源端），中间设备和 Ping 报文接收端（目的端）。故障可能发生在其中任何一个设备的发送或接收方向，因此要确定出故障的方向和节点，缩小定位范围。

图 2-2 Ping 应用场景



确定报文是在从源端至目的端的路径出现问题，还是在反方向出现问题。在源端和目的端停止 Ping 操作，通过 **display icmp statistics** [**slot** *slot-id*]查看 ICMP 报文收发情况，如下：

```
<HUAWEI> display icmp statistics
  Input: bad formats              0     bad checksum                    0
         echo                    36     destination unreachable         9
         source quench            0     redirects                      43
         echo reply              18     parameter problem               0
         timestamp                0     information request             0
         mask requests            0     mask replies                    0
         time exceeded            6
         Mping request            0     Mping reply                     0
  Output:echo                    20     destination unreachable     71438
         source quench            0     redirects                       0
         echo reply              36     parameter problem               0
         timestamp                0     information reply               0
         mask requests            0     mask replies                    0
         time exceeded            0
         Mping request            0     Mping reply                     0
```

说明

在发送端，使用命令 **display icmp statistics** 命令查看主控板的报文统计信息。

在接收端，使用命令 **display icmp statistics** **slot** *slot-id* 命令查看对应接口板的报文统计信息。

- 如果 ICMP 报文计数没有增长，则单板或设备上没有其他上送的 ICMP 报文（比如网管的报文）。请执行如下步骤：

  执行 **ping** 操作，再次使用 **display icmp statistics** 命令查看 ICMP 报文收发情况。

  根据统计信息中 Input/Output 包的数量可以确定 Ping 出现问题的方向，如下：

  - 源端 Output:echo 值正常增加，Input:echo 没有增加；目的端 Input/Output 都没有变化。说明源端发出了请求但是没有收到回应，而在目的端没有收到请求，因此可以确定 Ping 在源端->目的端方向出现问题。
  - 源端 Output:echo 值正常增加，Input:echo 没有增加；目的端 Input/Output:echo 都正常增加。说明源端发出了请求但是没有收到回应，而在目的端收到请求，同时发出了回应，因此可以确定 Ping 包在目的端->源端方向出现问题。

  确定了出问题的方向后，请执行步骤 4。

- 如果 ICMP 报文计数仍在增长，则单板或是设备上有其他上送的 ICMP 报文。请执行如下步骤：

  说明

  此方法需要在以下前提下进行:

  - 确保不影响现网业务。
  - 相应接口下相应方向没有应用流量策略。

  1. 依次在每台设备上配置 ACL，通过源和目的 IP 地址匹配 Ping 报文。

     配置文件如下：

```
statistics enable
#
acl number 3000
rule 5 permit ip source 1.1.1.1 0 destination 1.1.1.2 0
#
traffic classifier 3000 operator or
if-match acl 3000
#
traffic behavior 3000
#
traffic policy 3000
statistics enable
classifier 3000 behavior 3000
```

2. 在接口视图下使用命令 **traffic-policy**，依次在接口上应用 ACL。
   - 对于 Ping 的发起端和接收端：在接口的 inbound 方向应用该流量策略。
   - 对于中间设备：在接口的 inbound 和 outbound 方向都应用该流量策略。

   配置文件举例如下：

```
#
interface GigabitEthernet2/0/0
 ip address 1.1.1.2 255.255.255.252
 traffic-policy 3000 inbound
#
interface GigabitEthernet3/0/0
 traffic-policy 3001 outbound
#
display traffic policy statistics interface
```

   说明

   当应用流量策略的接口是 Trunk 或 VLANIF 时，流量策略需要配置在成员物理接口下。

3. 使用命令 **display traffic policy statistics interface**，依次在每台设备的接口上查看 ACL 的命中情况。

```
<HUAWEI> display traffic policy statistics interface gigabitethernet 1/0/0 inbound
Interface: GigabitEthernet1/0/0
inbound: test
Traffic policy applied at 2007-08-30 18:30:20
Traffic policy Statistics enabled at 2007-08-30 18:30:20
Statistics last cleared: Never
Rule number: 7 IPv4, 1 IPv6
Current status: OK!
Item                              Packets                    Bytes
-------------------------------------------------------------------
Matched                             1,000                  100,000
  +--Passed                           500                   50,000
  +--Dropped                          500                   50,000
    +--Filter                         100                   10,000
    +--URPF                           100                   10,000
    +--CAR                            300                   30,000
Missed                              500                   50,000
Last 30 seconds rate
```

   - 如果 ACL 完全命中，则说明 Ping 报文发送或接收正常。如果仍无法 Ping 通，请保留上述信息，联系华为技术工程师。
   - 如果中间设备的 inbound 和 outbound 方向的 ACL 完全命中，则说明中间设备正常。需要排查发起端或目的端问题。
   - 如果某设备的 inbound 方向没有命中，则为 Ping 报文相应方向的上游设备故障。请在故障设备上执行步骤 5。

步骤 4 确定出问题的节点

从出问题方向顺序定位。

- 源端->目的端方向出现了问题，可以按照下面的方法确定出问题的节点，先从源端检查。
- 目的端->源端方向出现问题时方法一样，从目的端检查。

执行 **tracert** *dest-ip-address* 命令确定报文丢失的位置。

```
<HUAWEI> tracert 1.1.1.1
  traceroute to 1.1.1.1 (1.1.1.1), max hops: 30, packet length: 40, press CTRL_C to break
    1 30.1.1.1 5  ms  4 ms  3 ms
    2 89.0.0.2 10 ms  11 ms  8
    3   *  *  *
    .....
```

上面所示在 89.0.0.2 10 的下一跳设备（即显示为“3 * * *”的节点）出了问题。确定了出问题的设备后请执行步骤 5。

说明

Tracert VPN 时，请使用 **tracert -vpn-instance** *vpn-name destination-address* 来检测。其中的**-vpn-instance** *vpn-name* 是指 Tracert 目的地址所属的 VPN 实例。

步骤 5 检查出问题的节点上是否配置了本机防攻击策略

因有的设备有受到过 ICMP 报文的攻击，为了防止攻击，将 ICMP 报文上送 CPU 的速率改小或将 ICMP 报文直接丢弃（Drop），从而导致了 Ping 不通的情况。

使用命令 **display current-configuration | include cpu-defend**，检查设备配置文件中是否存在 **cpu-defend policy** 配置。

- 如果存在 CPU 防攻击策略，使用命令行 **display cpu-defend policy** *policy-number* 和 **display cpu-defend car** 检查：
  - 是否配置了 Ping 相关 IP 地址的黑名单。
  - 是否配置了 CAR。如果配置了 CAR，请确认 CAR 的带宽参数是否过小，导致 Ping 报文无法处理。

  如果上述两种情况中的任何一种符合，都将导致 Ping 不通或丢包。请根据业务情况分析，如需继续执行 Ping 操作，请执行 undo 命令删除相应配置后再次执行 Ping 命令。如仍不能 Ping 通，请执行步骤 6。
- 如果没有配置 CPU 防攻击策略，请执行步骤 6。

步骤 6 在出问题的节点检查 FIB 和 ARP 表项是否正确

在出问题的节点执行 **display fib** *slot-number destination-address*，检查是否存在到目的地址的路由，如果路由不存在请参见或进行处理。

如果路由存在并且报文所经链路是以太链路，请执行 **display arp slot** *slot-number*，查看所需的 ARP 表项是否存在，如果不存在请执行步骤 9，否则请执行步骤 7。

说明

对于 Ping VPN 的情况，请使用 **display fib** *slot-number* **vpn-instance** *vpn-name destination-address* 命令查看 FIB 表项。其中的 **vpn-instance** *vpn-name* 是指 Ping 目的地址所属的 VPN 实例。

步骤 7 在出问题的节点检查接口下是否存在错包

执行命令 **display interface** *interface-type interface-number*，查看接口的报文计数信息。检查如下信息：

- 以太网接口的显示信息中 CRC 计数在两次执行该命令之间是否有增长。

- POS 接口的显示信息中 SDH alarm 和 SDH error 字段的错误或告警在执行两次该命令之间是否有增长。
- 如果接口下错包或告警计数有增长，请排查链路和光模块问题。
- 如果接口下错包或告警计数没有增长，请继续执行步骤 8。

步骤 8 确定出问题的层面

请通过下面的方法和步骤在出问题的设备上继续定位出问题的层面：

1. 查看 ICMP 报文是否正常接收。

```
<HUAWEI> display icmp statistics
Input: bad formats       0     bad checksum                0
        echo             0       destination unreachable     0
        source quench    0       redirects                   0
        echo reply       0       parameter problem           0
        timestamp        0       information request         0
        mask requests    0       mask replies                0
        time exceeded    0
        Mping request    0       Mping reply                 0
Output:echo              0      destination unreachable 476236
        source quench    0       redirects                   0
        echo reply       0       parameter problem           0
        timestamp        0       information reply           0
        mask requests    0       mask replies                0
        time exceeded    0
        Mping request    0       Mping reply                 0
```

   若没有收到 ICMP 报文，或是收到有错包，请执行步骤 9。

   若 ICMP 报文接收正常，请执行子步骤 b。

2. 检查 IP 层面是否正常

   通过 **display ip statistics** [**slot** *slot-num*]命令查看 IP 层面的统计信息以确认是否是 IP 层面出了问题，如下所示：

```
<HUAWEI> display ip statistics slot 2
   Input:     sum                123174     local                   0
              bad protocol           0      bad format              0
              bad checksum           0      bad options             0
              discard srr            0      TTL exceeded            0
   Output:    forwarding             0      local              268816
              dropped                0      no route                0
   Fragment: input                   0      output                  0
              dropped                0
              fragmented             0      couldn't fragment       0
  Reassembling:sum                   0      timeouts                0
```

   如果上面的统计信息显示的错误统计计数（如 bad protocol、 bad format、bad checksum、 bad options、discard srr、TTL exceeded、dropped、no route、couldn't fragment）有增加，那么就有表明有错误报文到达了 IP 层面，IP 经合法性判断后将其丢弃。

   - 如果发生这种情况，说明本机的单板可能有故障，请执行步骤 9。
   - 如果统计计数正常，请执行子步骤 c。

3. 查看 ICMP 报文是否从 IP 层正常下发

   通过配置 ACL 的方式，确认报文是否下发到接口板。

   ACL 配置文件举例如下：

```
acl number 3000
rule 5 permit icmp source 1.1.1.1 0 destination 1.1.1.2 0
```

   打开 IP 报文的 debug 开关：

注意

打开 debug 开关会对系统性能造成一定影响，请确认后再操作。

```
<HUAWEI> debugging ip packet acl 3000
<HUAWEI> terminal monitor
<HUAWEI> terminal debugging
```

执行 Ping 操作，如 Ping 5 个报文。在终端上查看：是否显示发送了 5 个报文。如果没有看到发送 5 个报文，则说明 ICMP 报文没有下发到接口板。请执行步骤 9。

步骤 9 请收集如下信息，并联系华为技术工程师

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

2.1.4 相关告警与日志

相关告警

无

相关日志

无

2.2 Tracert 不通问题的定位思路

2.2.1 常见原因

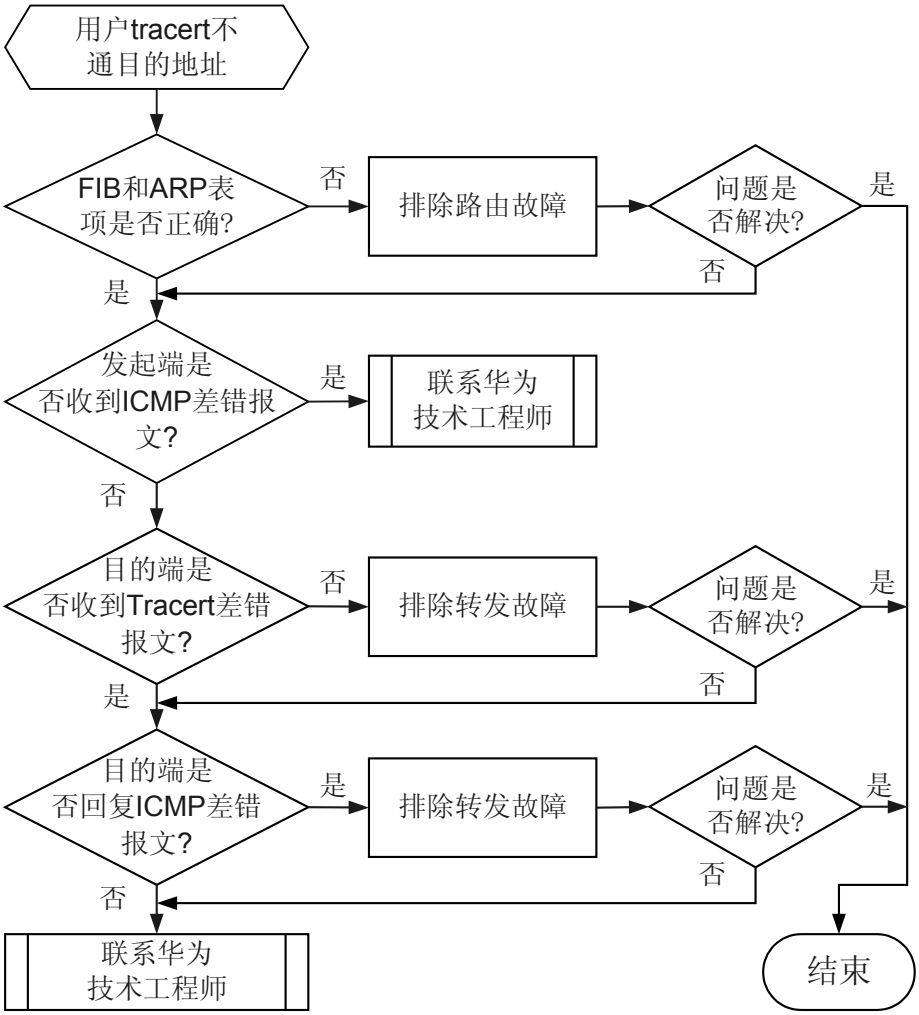
本类故障的常见原因主要包括：

- 路由或者 ARP 表项有问题。
- Tracert 报文被改写导致 IP 层面进行合法性检查失败，丢弃报文。

2.2.2 故障诊断流程

可按照故障诊断流程图 2-3 排除此类故障。

图 2-3 tracert 不通故障诊断流程图



2.2.3 故障处理步骤

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查 FIB 表项和 ARP 表项是否正确

在不能回应 ICMP 差错报文的设备上面执行 **display fib** *slot-number dest-ip-address*，检查是否存在到目的地址的路由。

- 如果路由不存在请参考 **3 OSPF 故障处理**或者 **4 IS-IS 故障处理**，排除路由问题。
- 如果路由存在并且报文所经链路是以太链路，请执行 **display arp slot** *slot-number*，查看 Tracert 的下一跳地址对应的 ARP 表项是否存在，如果不存在请执行步骤 3，否则请执行步骤 2。

步骤 2 检查 Tracert 发起端是否收到 ICMP 差错报文

在 Tracert 发起端执行 **display icmp statistics** 命令查看发起端是否收到 ICMP 差错报文，如下显示：

```
<HUAWEI> display icmp statistics
  Input: bad formats            0     bad checksum                  0
         echo                  13     destination unreachable      18
         source quench          0     redirects                    43
         echo reply           697      parameter problem            0
         timestamp              0     information request           0
         mask requests          0     mask replies                 0
         time exceeded         12
         Mping request          0     Mping reply                  0
  Output:echo                 704      destination unreachable   93326
         source quench          0     redirects                    0
         echo reply            13     parameter problem            0
         timestamp              0     information reply            0
         mask requests          0     mask replies                 0
         time exceeded          0
         Mping request          0     Mping reply                  0
```

在 Tracert 过程中多次执行该命令并查看结果，如果 Input 项目里面的 destination unreachable 和 time exceeded 两项的计数增加的个数和发起的 Tracert 报文的个数相等则表明 Tracert 发起端收到了 ICMP 差错报文，该回复报文在本机转发时被丢弃，请联系华为技术支持工程师排除转发故障。否则，请执行步骤 3。

步骤 3 请收集如下信息，并联系华为技术支持工程师

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

2.2.4 相关告警与日志

相关告警

无

相关日志

无

3 OSPF 故障处理

关于本章

3.1 OSPF 邻居 Down 的定位思路

3.1.1 常见原因

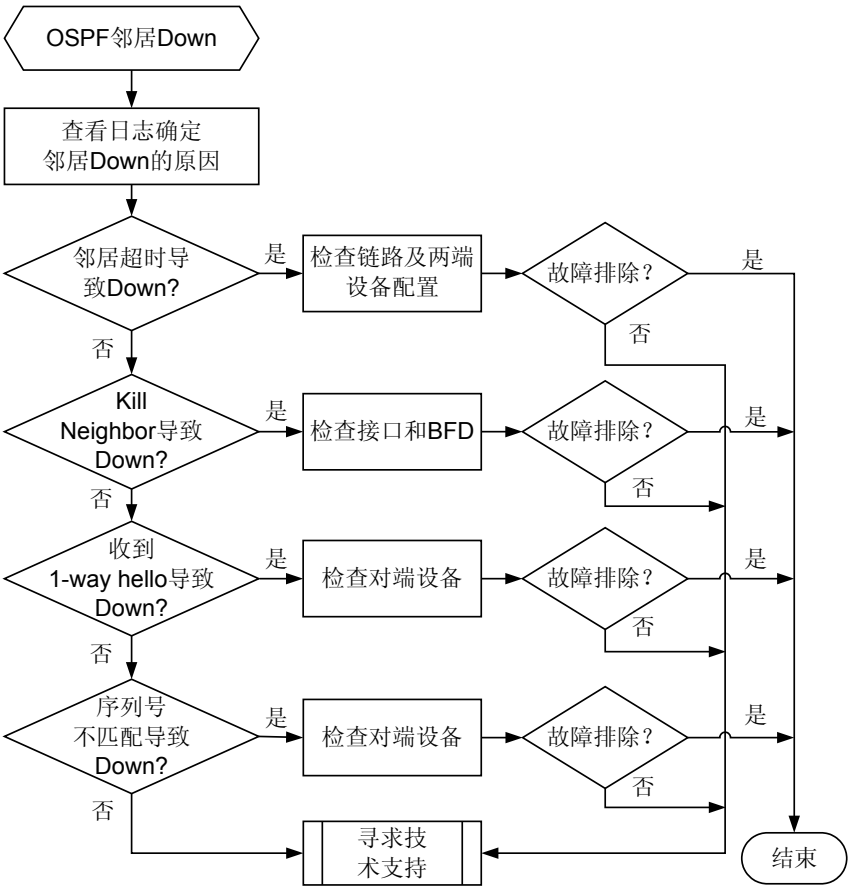
本类故障的常见原因主要包括：

- BFD 故障。
- 对端设备故障。
- CPU 利用率过高。
- 链路故障。
- 接口没有 Up。
- 两端 IP 地址不在同一网段。
- RouterID 配置冲突。
- 两端区域类型配置不一致。
- 两端 OSPF 参数配置不一致。

3.1.2 故障诊断流程

在配置 OSPF 后发现 OSPF 邻居 Down，可按照故障诊断流程图 3-1 排除故障。

图 3-1 OSPF 邻居 Down 故障诊断流程图



3.1.3 故障处理步骤

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 通过日志查看 OSPF 邻居 Down 的原因

执行 **display logbuffer** 命令，查看如下日志信息。

```
NBR_DOWN_REASON(l): Neighbor state leaves full or changed to Down. (ProcessId=[USHORT],
NeighborRouterId=[IPADDR], NeighborAreaId=[ULONG], NeighborInterface=
[STRING],NeighborDownImmediate reason=[STRING], NeighborDownPrimeReason=[STRING],
NeighborChangeTime=[STRING])
```

重点关注关键字 NeighborDownImmediate reason，此关键字记录的是 OSPF 邻居 Down 的原因。OSPF 邻居 Down 的原因一般会有以下几种：

- Neighbor Down Due to Inactivity

  表示在 deadtime 时间内没有收到 Hello 报文导致 OSPF 邻居 Down，出现这种情况请执行**步骤 2**。

- Neighbor Down Due to Kill Neighbor

  表示因为接口 Down、BFD Down 或执行了 **reset ospf process** 操作。此时，可以通过查看 NeighborDownPrimeReason 字段判断具体原因：

  - 如果是 Physical Interface State Change 则表示接口状态发生了改变，请执行 **display interface** [*interface-type* [*interface-number*]]命令查看接口状态，排查接口故障。详细的故障处理方法请参见《故障处理-二层网络》的定位。
  - 如果是 BFD Session Down，则表示 BFD 会话状态变成 Down，请排查 BFD 故障，详细步骤请参见 **BFD 会话无法 Up 的定位思路**。
  - 如果是 OSPF Process Reset，则表示执行了 **reset ospf process** 的操作，OSPF 进程正在重启，请等待 OSPF 重新建立邻居关系。

- Neighbor Down Due to 1-Wayhello Received 或 Neighbor Down Due to SequenceNum Mismatch

  表示因为对端 OSPF 状态首先变成 Down，从而向本端发送 1-Wayhello，导致本端 OSPF 状态也变成 Down。这种情况请先排查对端设备的原因。

- 其他情况请执行**步骤 9**。

步骤 2 检查链路是否故障

请检查设备链路是否故障（包括传输设备故障）。详细的故障处理方法请参见物理对接类的定位。如果链路正常，请执行**步骤 3**。

步骤 3 检查 CPU 利用率是否过高

请执行 **display cpu-usage** 命令检查故障设备的主控板和接口板的 CPU 利用率 ROUT 字段值是否超过 60%。如果 CPU 利用率过高会导致 OSPF 无法正常收发协议报文从而导致邻居振荡。如果 CPU 利用率超过 60%则执行**步骤 9**，否则执行**步骤 4**。

步骤 4 检查接口状态是否为 Up

请执行 **display interface** [*interface-type* [*interface-number*]]命令查看接口物理层状态，如果接口物理层状态为 Down 请先处理接口故障问题。详细的故障处理方法请参见《HUAWEI NetEngine80E/40E 路由器 - 二层网络》的定位。

如果接口物理层状态是 Up，请执行 **display ospf interface** 查看接口在 OSPF 协议下状态是否为 Down。接口在 OSPF 协议下正常状态可能为 DR、BDR、DROther 或 P2P 等。

```
<HUAWEI> display  ospf interface
            OSPF Process 1 with Router ID 1.1.1.1
                    Interfaces
 Area: 0.0.0.0           (MPLS TE not enabled)
 IP Address      Type          State     Cost    Pri    DR             BDR
 192.1.1.1       Broadcast     DR        1       1      192.1.1.1      0.0.0.0
```

- 如果接口在 OSPF 协议下状态为 Down，请执行命令 **display ospf cumulative** 检查 OSPF 进程下使能的接口数是否超出了规格，如果超出规格则减少 OSPF 使能的接口数。详细的规格请参见产品的 PAF/License 文件。

```
<HUAWEI> display ospf cumulative
          OSPF Process 1 with Router ID 1.1.1.1
                    Cumulations
  IO Statistics
             Type        Input     Output
            Hello            0         86
   DB Description            0          0
   Link-State Req            0          0
Link-State Update            0          0
   Link-State Ack            0          0
  SendPacket Peak-Control: (Disabled)
  ASE: (Disabled)
  LSAs originated by this router
  Router: 1
  Network: 0
  Sum-Net: 0
  Sum-Asbr: 0
  External: 0
  NSSA: 0
  Opq-Link: 0
  Opq-Area: 0
  Opq-As: 0
  LSAs Originated: 1  LSAs Received: 0
  Routing Table:
    Intra Area: 1  Inter Area: 0  ASE: 0
  Up Interface Cumulate: 1
```

- 如果接口在 OSPF 协议下状态不是 Down，请执行**步骤 5**。

步骤 5 如果接口连接的是广播网络或 NBMA 网络，检查两端 IP 地址是否在同一网段。

- 如果 IP 地址不在同一网段，请修改两端的 IP 地址，使其在同一网段。
- 如果 IP 地址处于同一网段，请执行**步骤 6**。

步骤 6 检查各接口的 MTU 是否一致

如果在接口上使能了 **ospf mtu-enable**，则要求接口的 MTU 一致，否则 OSPF 邻居无法协商成功。

- 如果接口的 MTU 值配置不一致，请在接口视图下执行 **mtu** *mtu* 命令，修改链路两端的 MTU 值为一致。
- 如果接口的 MTU 值配置一致，请执行**步骤 7**。

步骤 7 检查各接口的优先级是否非零

对于 Broadcast 和 NBMA 类型的网段，各接口的优先级至少有一个是非零的，以确保能够正确的选举出 DR，否则两边的邻居状态只能达到 2-Way。

执行命令 **display ospf interface**，查看接口的优先级。

```
<HUAWEI> display ospf interface
        OSPF Process 100 with Router ID 1.1.1.41
                Interfaces
 Area: 0.0.0.0          (MPLS TE not enabled)
 IP Address      Type         State     Cost  Pri    DR             BDR
 1.1.1.41        Broadcast    DR        1     1      1.1.1.41       0.0.0.0
```

步骤 8 检查两端 OSPF 的配置是否有错误

1. 检查两端 OSPF RouterID 配置是否相同

```
<HUAWEI> display ospf brief
         OSPF Process 1 with Router ID 1.1.1.1
                 OSPF Protocol Information
```

如果相同则执行 **ospf router-id*router-id*** 命令修改配置使 Router ID 在 AS 域内唯一，否则继续执行以下检查。

2. 检查两端 OSPF Area 配置是否一致

```
<HUAWEI> display ospf  interface
         OSPF Process 1 with Router ID 111.1.1.1
                 Interfaces
 Area: 0.0.0.0          (MPLS TE not enabled)
 IP Address      Type         State     Cost  Pri    DR             BDR
 111.1.1.1       Broadcast    BDR       1     1      111.1.1.2      111.1.1.1
```

如果不一致则修改配置使两端 OSPF Area 一致，否则继续执行以下检查。

3. 检查两端 OSPF 的其他配置是否一致

每 10 秒钟执行一次命令 **display ospf error**，持续 5 分钟。

```
<HUAWEI> display ospf error
         OSPF Process 1 with Router ID 1.1.1.1
                 OSPF error statistics
General packet errors:
 0      : IP: received my own packet     0      : Bad packet
 0      : Bad version                  0      : Bad checksum
 0      : Bad area id                   0      : Drop on unnumbered interface
 0      : Bad virtual link              0      : Bad authentication type
 0      : Bad authentication key        0      : Packet too small
 0      : Packet size > ip length       0      : Transmit error
 0      : Interface down                0      : Unknown neighbor
HELLO packet errors:
 0      : Netmask mismatch              0      : Hello timer mismatch
 0      : Dead timer mismatch           0      : Extern option mismatch
 0      : Router id confusion           0      : Virtual neighbor unknown
 0      : NBMA neighbor unknown         0      : Invalid Source Address
```

- 查看 **Bad authentication type** 字段，如果这个字段对应的计数值一直增长，表示建立邻居的两台设备配置的 OSPF 认证类型不一致，需要在两端设备上执行 **area-authentication-mode** 命令配置相同认证的类型。
- 查看 **Hello timer mismatch** 字段，如果这个字段对应的计数值一直在增长，表示接口上 hello timer 配置不一致，需要通过检查两端设备接口配置，执行 **ospf timer hello** 命令将 hello timer 间隔配置一致。
- 查看 **Dead timer mismatch** 字段，如果这个字段对应的计数值一直在增长，表示接口的 dead timer 配置不一致，需要通过检查两端设备接口配置，执行 **ospf timer dead** 命令将 dead timer 间隔配置一致。
- 查看 **Extern option mismatch** 字段，如果这个字段对应的计数值一直在增长，表示区域类型配置不一致（一端配置为普通区域，另一端配置为 stub 或 nssa 区域），需要将两端区域类型配置一致（在 OSPF 区域视图下，如果有 **stub** 命令，表示区域类型为 stub；如果有 **nssa** 命令，表示区域类型为 nssa）。

如果故障仍然存在，请执行**步骤 9**。

步骤 9 请收集如下信息，并联系华为技术支持工程师

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

3.1.4 相关告警与日志

相关告警

OSPF_1.3.6.1.2.1.14.16.2.2 ospfNbrStateChange

相关日志

OSPF/4/NBR_DOWN_REASON

3.2 OSPF 邻居无法达到 FULL 状态的定位思路

3.2.1 常见原因

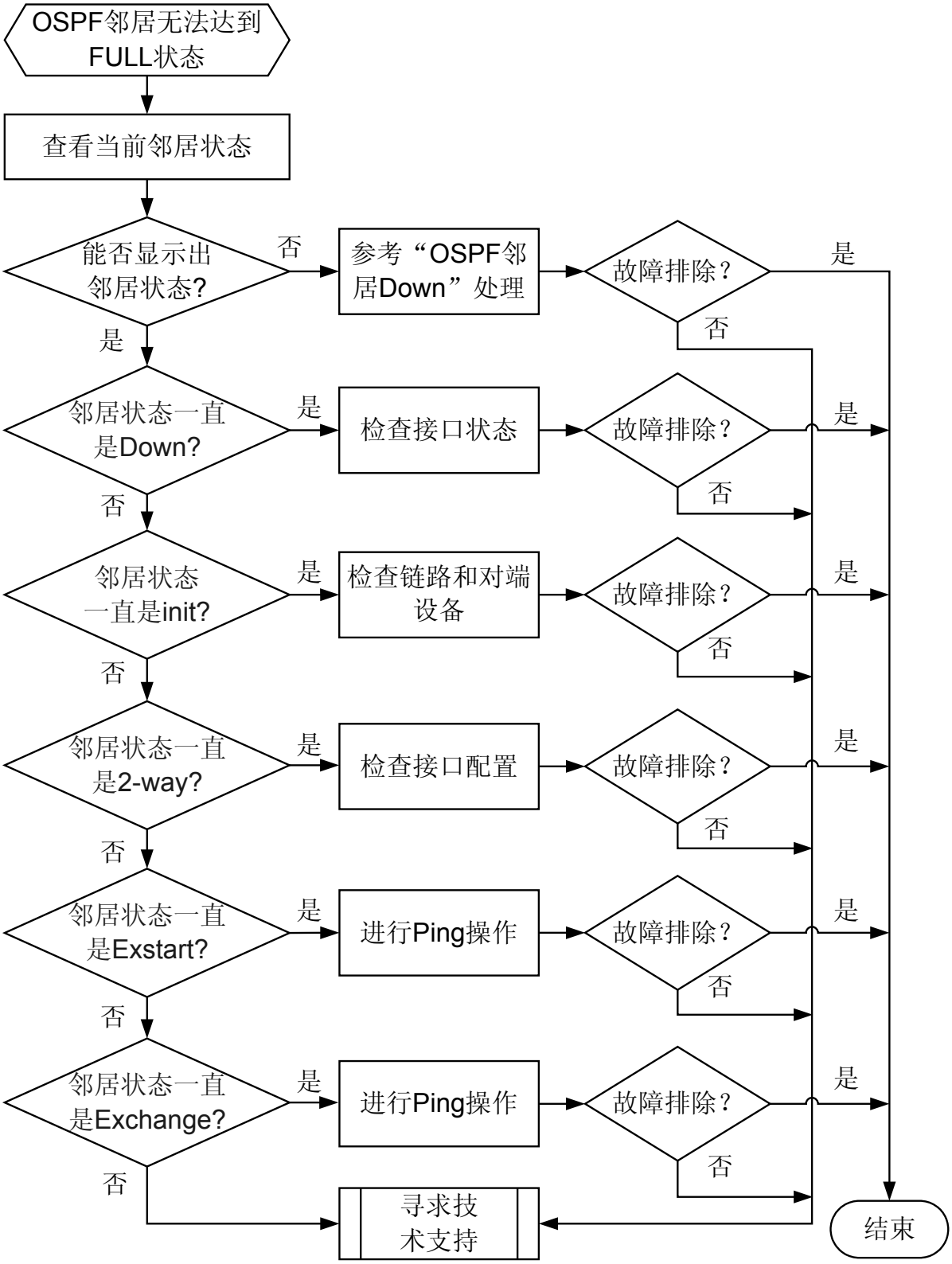
本类故障的常见原因主要包括：

- 链路故障，OSPF 报文被丢弃。
- 接口的 dr-priority 配置不合理。
- 两端配置的 OSPF MTU 值不相等。

3.2.2 故障诊断流程

可按照故障诊断流程图 3-2 排除故障。

图 3-2 OSPF 邻居无法达到 FULL 状态故障诊断流程图



3.2.3 故障处理步骤

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 根据不同的邻居状态进行相应的处理

- 无法显示 OSPF 邻居

  如果查看邻居状态时显示不出 OSPF 邻居，请参见 **OSPF 邻居 Down 故障处理**。

- 邻居状态一直是 Down

  请执行 **display interface** [*interface-type* [*interface-number*]]命令查看接口物理层状态，如果接口物理层状态为 Down 请先处理接口故障问题。

  如果接口物理层状态是 Up，请执行 **display ospf interface** 查看接口在 OSPF 协议下状态是否为 Up（接口 Up 状态为 DR、BDR、DROther 或 P2P）。

```
<HUAWEI> display  ospf interface
            OSPF Process 1 with Router ID 1.1.1.1
                    Interfaces
Area: 0.0.0.0           (MPLS TE not enabled)
IP Address      Type          State     Cost    Pri   DR              BDR
192.1.1.1       Broadcast     DR        1       1     192.1.1.1       0.0.0.0
```

  - 如果 OSPF 下的接口为 Up，请执行**步骤 2**
  - 如果 OSPF 下的接口为 Down，请执行命令 **display ospf cumulative** 检查 OSPF 进程下使能的接口数是否超出了规格，如果超出规格则减少 OSPF 使能的接口数。

```
<HUAWEI> display ospf cumulative
          OSPF Process 1 with Router ID 1.1.1.1
                  Cumulations
  IO Statistics
             Type        Input     Output
            Hello            0         86
    DB Description           0          0
    Link-State Req           0          0
Link-State Update            0          0
    Link-State Ack           0          0
  SendPacket Peak-Control: (Disabled)
  ASE: (Disabled)
  LSAs originated by this router
  Router: 1
  Network: 0
  Sum-Net: 0
  Sum-Asbr: 0
  External: 0
  NSSA: 0
  Opq-Link: 0
  Opq-Area: 0
  Opq-As: 0
  LSAs Originated: 1  LSAs Received: 0
  Routing Table:
    Intra Area: 1  Inter Area: 0  ASE: 0
  Up Interface Cumulate: 1
```

- 邻居状态一直是 init

  如果查看邻居状态时显示一直是 init，表示对端设备收不到本端发送的 hello 报文，此时请排查链路和对端设备是否故障。

- 邻居状态一直是 2-way

  如果查看邻居状态一直是 2-way，则执行命令 **display ospf interface** 查看设备在 OSPF 下面使能的接口配置的 dr-priority 是否为 0。

```
<HUAWEI> display ospf interface
         OSPF Process 1 with Router ID 111.1.1.1
                 Interfaces

 Area: 0.0.0.0          (MPLS TE not enabled)
 IP Address      Type          State     Cost  Pri   DR              BDR
 111.1.1.1       Broadcast     DROther   1     0     111.1.1.2       0.0.0.0
```

- 如果 OSPF 下使能的接口配置的 dr-priority 是 0 且 State 为 DROther，则说明他们都不是 DR 或 BDR，两者之间不需要交换 LSA，2-way 为正常状态，无需处理；
- 如果不是 0，请执行**步骤 2**

- 邻居状态一直是 Exstart

  如果查看邻居状态一直是 Exstart，表示设备一直在进行 DD 协商，但无法进行 DD 同步，出现该情况有两种可能性：

  - 超大报文包无法正常收发

    可以通过执行命令 **ping -s** *1500 neighbor-address* 查看超大报文收发情况。如果无法 Ping 通，请先解决链路问题。

  - OSPF MTU 值配置不同

    如果 OSPF 接口下配置了 **ospf mtu-enable**，请检查两端的 OSPF MTU 值是否相等，如果不相等则修改接口下的 MTU 值。

  如果故障没有解决，请执行**步骤 2**。

- 邻居状态一直是 Exchange

  如果查看邻居状态一直是 Exchange，表示设备在进行 DD 交换，请参见邻居状态一直是 init 状态处理。如果问题没有解决请执行**步骤 2**。

- 邻居状态一直是 Loading

注意

重启 OSPF 会导致该 OSPF 进程下所有邻居重新建立，并会导致业务暂时中断。

如果查看邻居状态一直是 Loading，可以尝试执行命令 **reset ospf** *process-id* **process** 重启 OSPF 进程。

如果问题没有解决请执行**步骤 2**。

步骤 2 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

3.2.4 相关告警与日志

相关告警

OSPF_1.3.6.1.2.1.14.16.2.2 ospfNbrStateChange

OSPF_1.3.6.1.2.1.14.16.2.8 ospfIfRxBadPacket

OSPF_1.3.6.1.2.1.14.16.2.16 ospfIfStateChange

相关日志

无

3.3 相关案例

3.3.1 由于接口配置重定向导致 OSPF 邻居状态无法建立

网络环境

如图 3-3 所示，Router A 双归属至 Router C 和 Router D。因网络调整，将 Router A 的上行至 Router C 的链路调整为接口 GE1/0/2 所在链路，如图 3-4 所示。

图 3-3 OSPF 邻居无法 Full 案例组网图（链路调整前）

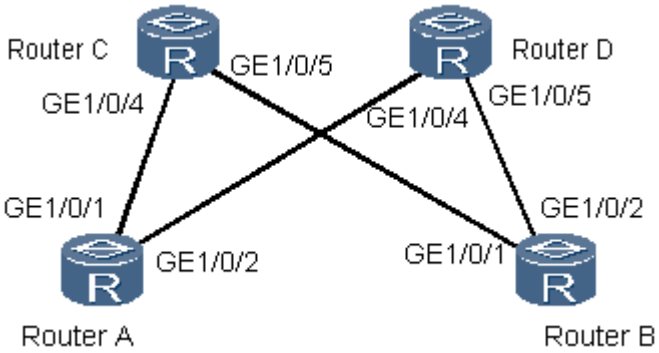
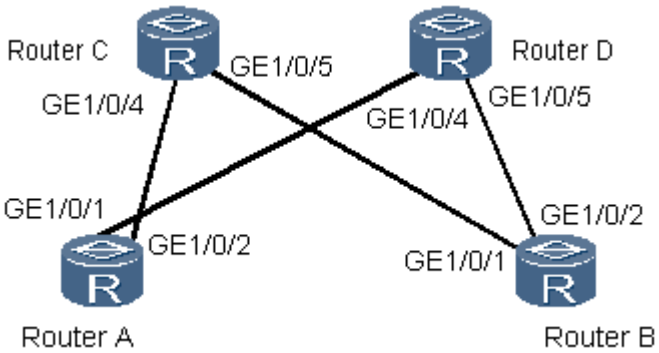


图 3-4 OSPF 邻居无法 Full 案例组网图（链路调整后）



网络调整后，Router C 与 Router A 的 OSPF 邻居均无法建立，Router C 的邻居状态一直处于 ExStart 状态。

故障分析

1. 执行命令 **display ospf peer**，查看显示信息中的 State 字段，发现当前的 OSPF 邻居状态为 ExStart。

   因为 Router C 的 OSPF 邻居状态一直 ExStart，说明已经收到对端的 Hello 报文，并发送了 DD 报文，但没有收到对端的 DD 报文。

2. 因为已经收到了对端的 Hello 报文，说明链路无故障。因此怀疑 Router C 接收到对端的 DD 报文后，将 DD 报文丢弃，没有上送 CPU。

   执行命令 **interface** *interface-type interface-number*，进入接口视图。执行命令 **display this**，发现接口下配置了流策略。

```
<HUAWEI> system-view
[HUAWEI] interface gigabitethernet 1/0/4
[HUAWEI-GigabitEthernet1/0/4] display this
#
interface gigabitethernet 1/0/4
 undo shutdown
 ip address 222.61.1.185 255.255.255.252
 traffic-policy Redirect inbound
#
```

   查看流策略以及对应的流分类和流行为，发现流分类中的 ACL 3000 能够匹配 Router A 的接口地址，报文被重定向至接口 GE1/0/3，而没有上送 CPU。

```
[HUAWEI] display traffic policy user-defined Redirect
  User Defined Traffic Policy Information:
  Policy: Redirect
   Share-mode
   Classifier: default-class
     Behavior: be
      -none-
   Classifier: user
     Behavior: toCR
      Redirecting:
        Redirect Ip-NextHop 222.61.1.241 Interface GigabitEthernet1/0/3
[HUAWEI] display traffic classifier user-defined user
  User Defined Classifier Information:
   Classifier: user
    Operator: OR
    Rule(s) : if-match acl 3000
[HUAWEI] display acl 3000
Advanced ACL 3000, 1 rule
Acl's step is 5
 rule 15 permit ip source 222.61.0.0 0.0.63.255
```

   由于 Hello 报文的目的地址为保留组播地址，此类报文无法进入重定向流程，能够正常上送 CPU 处理。而 DD 报文是单播报文，携带的是接口地址，因此被重定向至其它接口，导致无法上送 CPU 处理。

操作步骤

步骤 1 执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **interface** *interface-type interface-number*，进入接口视图。

步骤 3 执行命令 **undo traffic-policy inbound**，删除流策略。

删除流策略后，可以执行 **display ospf peer**，查看到 OSPF 邻居状态为 Full，故障恢复。

说明

待 OSPF 邻居状态为 Full 后，可执行 **traffic-policy** 命令将流策略重新应用在接口上，不会对 OSPF 状态造成影响。但是此时需注意，若因链路变化、插拔光纤等操作导致 OSPF 邻居状态为 DOWN 后，将无法再次达到 Full 状态。

----结束

案例总结

由于 Hello 报文为保留组播报文，而 DD 报文为单播报文。保留组播报文，不会被重定向，能够正常上送 CPU，而 DD 报文却被重定向至其它接口，因为导致设备的 OSPF 邻居状态一直处于 ExStart 状态。

3.3.2 OSPF 5 类 LSA FA 问题导致下挂设备路由不正常

网络环境

在图 3-5 的网络中，RouterC 是其他厂商设备，RouterA 和 RouterB 两台路由器上各有两个上行的 GE 接口，并分别配置两条静态路由，如下：

- RouterA

```
[RouterA] ip route-static 0.0.0.0 0.0.0.0 192.168.0.69
[RouterA] ip route-static 0.0.0.0 0.0.0.0 192.168.0.65
```

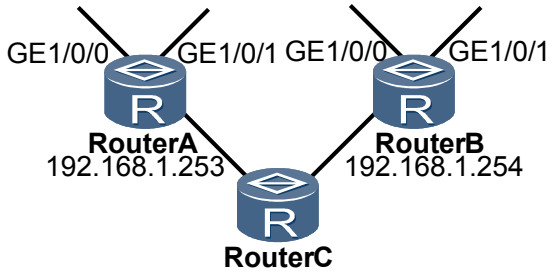
- RouterB

```
[RouterB] ip route-static 0.0.0.0 0.0.0.0 192.168.0.5
[RouterB] ip route-static 0.0.0.0 0.0.0.0 192.168.0.1
```

两台路由器都在 OSPF 进程中非强制发布默认路由给 RouterC，测试中发现 RouterC 上故障现象如下：正常时 RouterC 有两条 OSPF 默认外部路由指向两台路由器，但是如下两种情况时，RouterC 上只有一条 OSPF 默认路由指向两台路由器中的一台。

- 在 RouterA 上删除 192.168.0.65 的静态路由，其他保持不变。此时，在 RouterC 上只有一条 OSPF 默认路由指向 RouterB；
- 在 RouterB 上删除 192.168.0.1 的静态路由，其他保持不变。此时，RouterC 上只有一条 OSPF 默认路由指向 RouterA。

图 3-5 OSPF 5 类 LSA FA 问题导致下挂设备路由不正常组网图



故障分析

1. 在 RouterA 上执行 **undo ip route-static** *0.0.0.0 0.0.0.0 192.168.0.65*，然后在 RouterC 上查看对应 LSA 详细信息时，发现 FA 地址被 RouterA 置错，此时 RouterC 上只有

一条 OSPF 默认路由指向 RouterB，因为 RouterC 上 OSPF 的 SPF 计算时发现 192.168.0.69 地址不可达。

2. 在 RouterB 上执行 **undo ip route-static** *0.0.0.0 0.0.0.0 192.168.0.1*，然后在 RouterC 上查看对应 LSA 详细信息时，发现 FA 地址被 RouterB 置错，此时 RouterC 上只有一条 OSPF 默认路由指向 RouterA，因为 RouterC 上 OSPF 的 SPF 计算时发现 192.168.0.5 地址不可达。
3. 从如上故障现象中，发现 RouterC 上出现 OSPF 路由学习不是预期的结果，根本的原因是上面 RouterA 和 RouterB 将 Forwarding Address（FA）设置错误。

   路由器填写 5 类 LSA 的 FA 地址及其路由计算的规则如下：

   - FA 填写为 0.0.0.0 时：

     当一个 5 类 LSA 中的 FA 为 0.0.0.0 时，接收该 LSA 的路由器按照 Adv Rtr（也就是 ASBR）来计算下一跳。

   - FA 填写为非 0.0.0.0 时：

     同时满足如下条件时，ASBR 会在 5 类 LSA 的 FA 域内填写非 0.0.0.0 的转发地址，接收 LSA 的路由器按照该非 0.0.0.0 地址计算下一跳。

     (1) OSPF 在 ASBR 与外部网络连接的下一跳接口启动；

     (2) ASBR 与外部网络连接的下一跳接口没有被设置为被动接口；

     (3) ASBR 与外部网络连接的下一跳接口不是 OSPF P2P 或 P2MP 类型的；

     (4) ASBR 与外部网络连接的下一跳接口地址是落在 OSPF 协议中发布的网络范围之内。

     不满足如上四点条件的，FA 都填写为 0.0.0.0。

操作步骤

步骤 1 如下几种方式可以解决此问题：

- 检查 RouterA 和 RouterB 的数据配置发现：
  - RouterA 上 OSPF 进程中配置了 **network** *192.168.0.68 0.0.0.3*，而没有配置 **network** *192.168.0.64 0.0.0.3*；
  - RouterB 上 OSPF 进程中配置了 **network** *192.168.0.4 0.0.0.3*，而没有配置 **network** *192.168.0.0 0.0.0.3*。

  分别在 RouterA 和 RouterB 上 OSPF 进程内，将对应静态路由下一跳网段的 **network** 配置删除，问题解决。
- 不影响正常业务的情况下，在 RouterA 和 RouterB 上 **network** 命令指定的接口下，分别执行 **ospf network-type p2p**，对端接口也如此修改，问题解决。
- 在 RouterA 和 RouterB 上将对应接口设置为 slience 接口，或者让 RouterA 和 RouterB 的所有静态路由的下一跳 IP 地址在 RouterC 上都是路由可达，都可以解决此问题。

----结束

案例总结

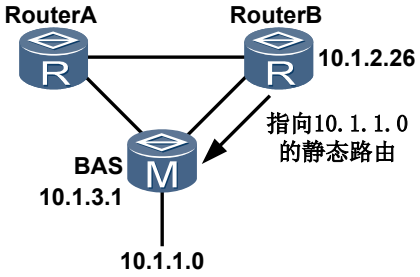
通过正确指定运行 OSPF 协议的接口的 IP 地址位于的网段和配置接口类型，使路由器必须按照规则填写 5 类 LSA 的 FA 地址及其路由计算。

3.3.3 路由器收到两条相同 LSID 的 LSA 但其中一条不能计算出路由

网络环境

在图 3-6 的网络中，由于到 BAS 下的流量不均匀，需要让 RouterA 到 BAS 下目的网段的路由通过“RouterA--BAS--目的”和“RouterA--RouterB--BAS--目的”来形成负载分担均衡流量。

图 3-6 收到两条相同 LSID 的 LSA 但其中一条不能计算出路由组网图



下面以到目的网段为 10.1.1.0 为例。

用户在 RouterB 上配置了一条到 10.1.1.0 的静态路由，并且配置 OSPF 引入静态路由，RouterA 上收到 RouterB 发来的 LS ID 为 10.1.1.0 的 ASE LSA，同时 RouterA 上也收到从 BAS 发来的 LS ID 为 10.1.1.0 的 ASE LSA。结果，BAS 发来 LSA 生效计算出路由，RouterB 发来 LSA 并没有计算出路由。

故障分析

出现上述故障，可能有如下原因：

1. 配置问题。
2. RouterB 发来 LSA 中的 Forwarding Address: 10.1.2.26 置位，怀疑为 FA 问题导致 LSA 没被计算。
3. 生成负载分担路由条件不具备。

对上述原因进行一一排查和确认，结果如下：

1. 通过检查配置未发现问题。
2. 检查 FA 置位的 LSA，发现 LSA 符合计算路由条件。如下：

```
<RouterA> ping 10.1.3.1
  PING 10.1.3.1: 56  data bytes, press CTRL_C to break
    Reply from 10.1.3.1: bytes=56 Sequence=1 ttl=255 time=1 ms
    Reply from 10.1.3.1: bytes=56 Sequence=2 ttl=255 time=1 ms
    Reply from 10.1.3.1: bytes=56 Sequence=3 ttl=255 time=1 ms
    Reply from 10.1.3.1: bytes=56 Sequence=4 ttl=255 time=1 ms
    Reply from 10.1.3.1: bytes=56 Sequence=5 ttl=255 time=1 ms

  --- 10.1.3.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/1/1 ms
<RouterA> display ip routing-table 10.1.3.1
Route Flags: R - relay, D - download to fib
```

```
------------------------------------------------------------------------------
Routing Table : Public
Summary Count : 2

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

     10.1.3.1/32  O_ASE   150  1           D  10.1.2.45     GigabitEthernet1/0/5
                  O_ASE   150  1           D  10.1.2.49     GigabitEthernet1/0/6
<RouterA> ping 10.1.2.26

    Reply from 10.1.2.26: bytes=56 Sequence=1 ttl=254 time=1 ms
    Reply from 10.1.2.26: bytes=56 Sequence=2 ttl=254 time=1 ms

    0.00% packet loss
    round-trip min/avg/max = 1/1/1 ms
<RouterA> display ip routing-table 10.1.2.26

    10.1.2.24/30  OSPF    10   101         D  10.1.2.45     GigabitEthernet1/0/5

                  OSPF    10   101         D  10.1.2.49     GigabitEthernet1/0/6
```

3. 在该网络中，LSA 的 cost 都是 1，则需要比较到 ASBR 的 cost 以及 FA 的 cost。

   对于 Type2 的 ASE LSA，OSPF 形成等价路由的比较方式如下：

   (1) 比较 LSA 的 cost，如果相等，进行下一步比较；

   (2) 比较到 ASBR/FA 的 cost，如果相等，形成等价路由。

   发现到 FA 转发地址的 cost 值为 101。

   - 对于 FA 为 0 的 LSA，其到 ASBR 10.1.3.1 的 cost 为 1；
   - 对于 FA 不为 0 的 LSA，其到 FA 10.1.2.26 的 cost 为 101；

   FA 置位的 LSA 由于优先级较低，所以没有被计算，因此无法形成等价路由。

操作步骤

步骤 1

此组网形成等价路由的办法为：

在 BAS 上，执行 **network** 命令使能 10.1.1.0 对应路由的下一跳。并执行 **ospf cost** 命令将该接口 cost 配置为 100，使其发布带 FA 的 LSA，FA 地址为接口地址。

这样在 RouterA 上，看到的两个 LSA 都有 FA，且到两个 FA 的 cost 都为 101，形成等价路由。

----结束

案例总结

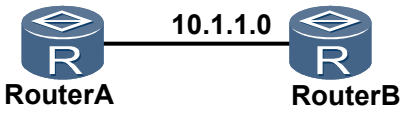
配置 OSPF 形成负载分担，需要正确配置带相同 FA 的 LSA，且配置 LSA 的相同 cost 值。

3.3.4 OSPF 邻居因链路问题无法建立

网络环境

在图 3-7 的组网中，RouterA 上 OSPF 邻居无法建立，状态为 State:Exchange。

图 3-7 OSPF 邻居因链路问题无法建立组网图



故障分析

出现上述故障，可能有如下原因：

- OSPF 配置问题。
- 两端设备的 OSPF 接口的相关参数不匹配。
- OSPF 协议报文被丢弃。

检查 RouterA 的 OSPF 配置，确认 RouterA 的 OSPF 配置没有问题。

检查两端设备的接口的 OSPF 相关参数，都匹配，也没有问题。

在 RouterB 上执行 **debugging ospf packet dd** 发现是 MTU 值协商不成功造成的。在两端设备上检查的 MTU 值都为 4470，但是 debug 信息发现 RouterB 收到的 MTU 值为“0”，即没有收到 RouterA 的 MTU 值。说明链路方面存在不畅通的情况。

在 RouterA 上 PING 对端设备直连接口地址，发现有丢包：

```
<RouterA> ping 10.1.1.0
  PING 10.1.1.0: 56  data bytes, press CTRL_C to break
    Request time out
    Reply from 10.1.1.0: bytes=56 Sequence=2 ttl=255 time=5 ms
    Reply from 10.1.1.0: bytes=56 Sequence=3 ttl=255 time=5 ms
    Reply from 10.1.1.0: bytes=56 Sequence=4 ttl=255 time=5 ms
    Request time out

    40.00% packet loss
```

首先经过传输侧确认中间的链路没有问题。然后在 RouterA 上做流量统计，发现数据包是在 RouterA 接口之外丢掉的，也就是说数据包有可能是在对端设备单板上或者链路上丢掉的。

经过在对端设备上做流量统计，确认为 RouterB 单板问题。

操作步骤

步骤 1 更换 RouterB 的故障单板。

----结束

案例总结

有时 OSPF 的报文无法正确接收，原因有很多，首先要检查链路层是否畅通。可以打开 OSPF 的 debug 开关来查。Debug 命令有 **debugging ospf packet**、**debugging ospf event** 等，还可以通过 **display ospf error** 来看各种 OSPF 的错误统计信息。如果 OSPF 的信息正确，可以通过打开 **debugging ip packet** 来检查 IP 层是否转发成功。

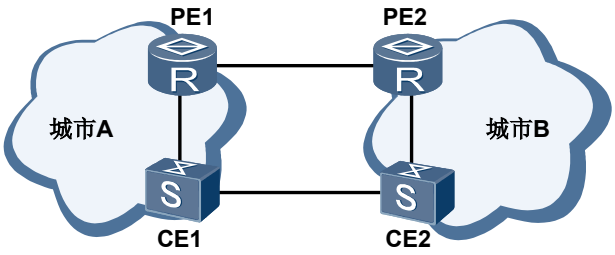
3.3.5 Router-ID 冲突导致 OSPF 路由环路

网络环境

在图 3-8 的组网中，PE 和 CE 之间运行 OSPF 多实例，CE 为其他厂商的三层交换机，PE 下发 OSPF 缺省路由引导两地业务的互通。PE1 和 PE2 都有另外一个接口连接到同一台 UMG 设备，并且两台 PE 上连接 UMG 的两个接口 IP 都配置为 10.1.1.33，并且绑定到了上面的 VPN 实例中。正常情况下由于 UMG 到备用 PE2 的接口不发光，所以两台 PE 上关于 10.1.1.33 的接口不会同时 UP。

现象：本地 CE1 和 CE2 设备 PING 本地直连 PE 都正常，但是 PING 远端的 CE 和业务 IP 会出现偶尔不规则的丢包。

图 3-8 Router-ID 冲突导致 OSPF 路由环路组网图



故障分析

1. 由于在两边 PE 绑定的 VPN 实例中，10.1.1.33 为最大的一个 IP 地址。并且 OSPF 多实例的配置为：

   〈PE1〉**ospf 4 vpn-instance www**

   所以导致 PE1、PE2 的 OSPF 进程 4 都选择 10.1.1.33 做为 Router-ID。
2. 在 CE1、CE2 上查看两边 PE 的 Router-ID 都为 10.1.1.33。
3. 在 CE 上查看 debug 相关信息后发现，Router-ID 为 10.1.1.33 的设备不断发送 LSA，频率为 5 秒一次，而且 seq 值递增，不稳定。
4. CE 交换机均收到相同 Router-ID 的两台设备发送的 LSA，所以查看路由表看到的 OSPF 缺省路由信息就会不断变动。而当 CE1 的缺省路由从 CE2 中学到，CE2 的缺省路由又从 CE1 中学到时，就形成了路由环路，因此出现路由不可达，造成丢包。

操作步骤

步骤 1 在两台 PE 上分别执行命令，强制指定该 OSPF 多实例的 Router-ID 为 PE 本机上唯一的地址。

```
[PE1] ospf 4 router-id 10.2.2.9 vpn-instance www
[PE2] ospf 4 router-id 10.2.2.10 vpn-instance www
```

步骤 2 重启两台 PE 上设备该 VPN 实例的 OSPF 进程，业务恢复。

----结束

案例总结

建议在 PE 上强制指定 OSPF 多实例的 Router-ID 为 PE 本机上唯一的地址。

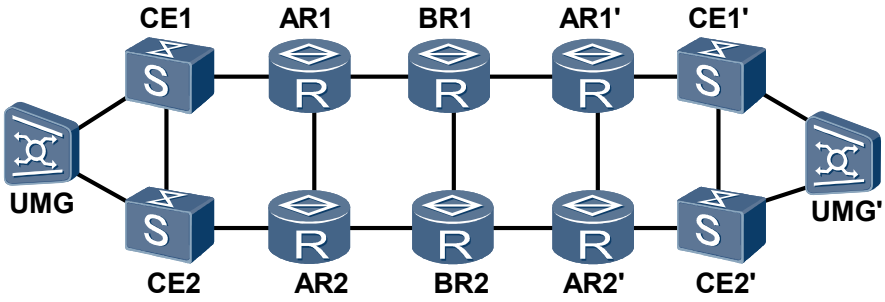
3.3.6 承载网备用平面设备主备倒换导致主用平面业务中断

网络环境

图 3-9 所示的承载网有主备两个平面，主用平面流量模型为：UMG→CE1→AR1→BR1→AR1'→CE1'，返回路径相同。

故障现象：在对备用平面的 AR2 设备进行主备倒换时，发现主用平面流量有少量丢包现象（1 秒左右）。同样，对主用平面 AR1 做主备倒换时，发现备用平面也有丢包现象。

图 3-9 承载网备用平面设备主备倒换导致主用平面业务中断组网图



故障分析

从现网拓扑和路由情况来分析，主备平面是相互独立的，不可能出现一个平面主备倒换影响到另一个平面的情况。经过分析配置发现，在现网所有相关 AR 上，OSPF 多实例下都配置了路由聚合命令：

```
ospf 1 vpn-instance 123
 asbr-summary 10.0.0.0 255.0.0.0
```

在 BGP 上通过 network 方式把路由发布出去。这样，到远端路由会被聚合成一条 10.0.0.0/8 的路由。根据 OSPF ABR 的聚合原则，聚合后的路由的 cost 值为所有被聚合具体路由中 cost 值最大者（ASBR 和 ABR 聚合后一样选 cost 值最大的下发）。举个例子，假设有如下三条路由：

- 10.1.1.1/24 cost 10；
- 10.2.1.1/24 cost 100；
- 10.3.1.1/24 cost 1000；

那么聚合后路由为：10.0.0.0/8 cost 1000。

在主用平面主备倒换后，AR2 上的私网路由将重新收敛；假设 AR2 先收到一条 cost 值小于 200 的 10.x.x.x 的路由，此时 AR2 向 CE2 发布的 10.0.0.0/8 的聚合路由的 cost 值比原来的小，通过 OSPF 协议扩散到 CE1。主用平面的流量模型变为：UMG→CE1→CE2→AR2→BR2→AR2'→CE2'。

由于网络规模较大，AR2 还未完全收敛，即 AR2 还没有目的地对应的明细路由，因此发生如上现象。

操作步骤

步骤 1 在 OSPF 视图下执行 **asbr-summary** 命令配置路由聚合时，指定聚合路由的 cost 值，这样可以避免先收到 cost 值较小的路由时发生选路问题。

```
ospf 1 vpn-instance 123
 asbr-summary 10.0.0.0 255.0.0.0 cost 300
```

配置如上命令后，经过多次验证，CE1 与 CE2 之间的链路上不再有流量经过，主用平面的流量能够一直保持在主用平面转发。

----结束

案例总结

在网络部署时，要注意避免双平面相互影响，如 IS-IS 协议的 checksum-error 问题、OSPF 协议的 cost 值问题、IS-IS 协议的 cost 值问题。

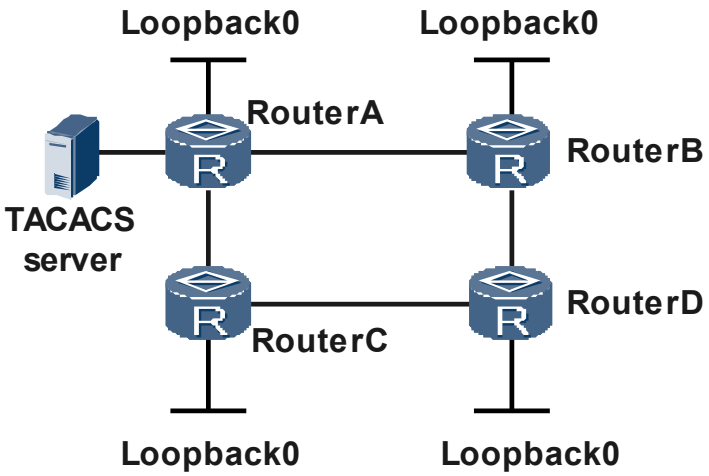
3.3.7 正确的用户名和密码不能通过 HWTACACS 认证

进行网络调整时，更换的路由协议中没有发布原 Loopback 接口的 IP 地址，导致使用该地址作为源 IP 地址的 HWTACACS 客户端和服务器之间无法互通。

网络环境

在图 3-10 所示的网络，在网络的核心节点部署了路由协议、AAA、QoS、SNMP 等业务，其中四台设备属于同一个 AS 域，路由协议采用 IBGP、ISIS。现按照客户规划新的私有 AS 号重新配置路由器，将 IBGP 改为 EBGP，将 IGP 的 ISIS 改为 OSPF 协议。其中 ISIS 协议中只包括互连接口和 Loopback 接口的 IP 地址。

图 3-10 核心网 TACACS 认证组网图



配置完成后，原来正确的 HWTACACS 用户名和密码不能通过 HWTACACS 认证。

故障分析

1. 检查 TACACS Server 记录的用户名和密码与用户使用的是否一致，发现用户名和密码正确。
2. 在 RouterA 上执行 **ping** 命令，检查路由器和 TACACS Server 是否互通，发现能够 ping 通。
3. 在 RouterA 上执行 **display current-configuration** 命令，检查 HWTACACS 的配置是否正确。发现在 HWTACACS 服务器模板中配置了如下命令：

   ```
   hwtacacs-server source-ip 192.168.1.227
   ```

   其中，192.168.1.227 是 RouterA 的 Loopback 接口地址。

   由于删除的 ISIS 协议中包括 Loopback 接口的 IP 地址，并且 HWTACACS 使用 RouterA 的 Loopback 接口地址作为源 IP，因此考虑可能是路由器无法收到 TACACS SERVER 返回的以 192.168.1.227 为目的地址的认证响应报文，导致 HWTACACS 认证失败。
4. 在 RouterA 上执行 **ping -a 192.168.1.227 100.1.1.245** 命令（100.1.1.245 是 TACACS Server 的 IP 地址），检查该 Loopback 地址和 TACACS Server 是否互通，发现不能 ping 通。
5. 在 RouterA 上执行 **display ip routing-table** 命令，检查路由协议是否发布了该 Loopback 接口的 IP 地址，发现 Loopback0 接口的 IP 地址没有发布。

   因此，确认是网络调整中删除 ISIS 协议，发布 Loopback 接口的配置也被删除，且 OSPF 协议中没有发布该 Loopback 接口的地址，路由器无法接收 TACACS Server 返回的认证响应报文，导致认证失败。

操作步骤

步骤 1 执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **ospf** *process-id*，进入 OSPF 视图。

步骤 3 执行命令 **area** *area-id*，进入 OSPF 区域视图。

步骤 4 执行命令 **network** *address wildcard-mask*，发布该 Loopback 接口的 IP 地址。

完成上述操作后，使用该用户名和密码，可以正常登录，故障排除。

----结束

案例总结

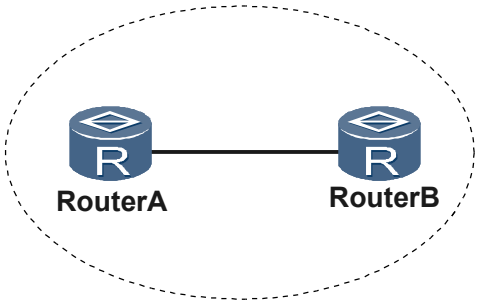
网络设备协议数据调整前，请记录之前的相关配置。协议数据调整后，检查调整后数据是否满足协议调整前的需求，并且检查是否对其他配置产生影响。

3.3.8 链路两端 MTU 不一致导致 OSPF 邻居状态不能达到 Full 状态

网络环境

如图 3-11 所示，RouterA 和 RouterB 是直连设备，其中 RouterB 是其他厂商设备，二者之间建立 OSPF 邻居。在 RouterA 上配置 MTU 为 1520 字节后，发现 Ping 不通 RouterB，OSPF 邻居状态一直为 Exchange，不能达到 Full 状态。

图 3-11 链路两端 MTU 不一致导致 OSPF 邻居状态不能达到 Full 状态组网图



故障分析

1. 在 RouterA 上，执行命令 **ping -s** *1500 host*，可以 Ping 通 RouterB。
2. 在 RouterA 上执行命令 **display ospf peer** 查看 OSPF 邻居状态及 OSPF MTU 值，发现 OSPF 邻居为 Exchange 状态（“State”字段），MTU 值为 1506 字节（MTU 字段）。
3. 查看 RouterB 的 OSPF 邻居状态及 OSPF MTU 值，发现 MTU 值为 1520 字节。

因为 RouterB 是其他厂商设备，其 MTU 实现机制与 RouterA 不一致，RouterB 接口的 MTU 包括 14 字节的二层报文头，而 RouterA 的接口 MTU 不包含二层报文头。在 RouterA 上，配置与 RouterB 接口相同数值的 MTU（1520 字节）时，RouterA 接口的实际 MTU 值（1520 字节）大于 RouterB 接口的实际 MTU 值（1506 字节），RouterA 发送报文时不分片，RouterB 不能识别收到的报文，因此导致 OSPF 邻居协商不通过而无法达到 Full 状态。

操作步骤

步骤 1 在 RouterA 上执行命令 **system-view**，进入系统视图。

步骤 2 在 RouterA 上执行命令 **interface** *interface-type interface-number*，进入指定的接口视图。

步骤 3 在 RouterA 的指定的接口视图下执行命令 **mtu** *1506*，将该接口的 MTU 值调节为 1506 字节。

步骤 4 在 RouterA 上执行命令 **display ospf peer**，发现 OSPF 邻居 Full 状态，故障排除。

----结束

案例总结

OSPF 不能进入 Full 状态时，主要有以下常见原因：

- 链路故障。
- OSPF 邻居的配置问题。

另外，不同厂商设备对 MTU 的计算存在差异，因此不同厂商设备相互对接时，需要确认对端设备的 MTU 计算方式，以保证两端设备 MTU 的一致。

如果 OSPF 接口下配置了命令 **ospf mtu-enable**，请检查两端的 OSPF MTU 值是否相等，如果不相等则修改接口下的 MTU 值。

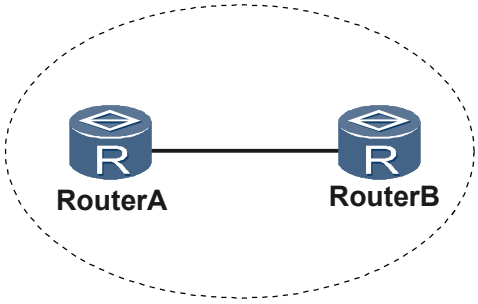
- 如果接收到的报文中携带的 MTU 值大于本端接口配置的 MTU 值，丢弃收到的 DD 报文，邻居状态就一直处于 ExStart 状态。
- 如果接收到的报文中携带的 MTU 值小于本端接口配置的 MTU 值，可以接受收到的 DD 报文，邻居状态就达到 Exchange 状态。

3.3.9 使能 Opaque LSA 能力后，OSPF 邻居无法建立

网络环境

如图 3-12 所示，RouterA 与 RouterB 之间建立 OSPF 邻居，并运行 MPLS TE 业务。其中 RouterA 为其他厂商设备，缺省使能 Opaque LSA 能力；RouterB 为华为设备，且在 RouterB 上的 OSPF 进程中配置了命令 **opaque-capability enable**。现发现 OSPF 邻居无法建立，RouterA 端到达 Full 状态，而 RouterB 则一直处于 Loading 状态。

图 3-12 使能 Opaque LSA 能力的 OSPF 邻居无法建立组网图



故障分析

1. 在 RouterB 上执行 **display ospf peer** 命令，发现 OSPF 邻居处于 Loading 状态（“State”字段为“Loading”），表明请求列表中存在需要向邻居请求的 LSA。
2. 在 RouterB 上执行 **display ospf request-queue** 命令查看是否在请求 Opaque LSA，发现 RouterB 一直在请求 Opaque-Area LSA。
3. 在 RouterB 上执行 **debugging ospf packet update filter src** *acl-number* 命令，发现 RouterA 一直在发送仅带头部信息的 LSA。

由于 RouterB 是华为设备，OSPF 不兼容仅带头部信息的 LSA。在与其他厂商设备建立 OSPF 邻居时，当其他厂商设备在 DD 报文中发送仅带头部信息的 Opaque LSA 时，华为设备会向其他厂商设备请求这些 LSA；当其他厂商设备在 LSU 报文中发送这些 LSA 时，华为设备会丢弃这些 LSA，并继续向其他厂商设备请求，导致邻居长期处于 Loading 状态。

操作步骤

步骤 1 在 RouterB 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **ospf** *process-id*，进入指定的 OSPF 进程。

步骤 3 执行命令 **undo opaque-capability enable**，禁止 Opaque LSA 能力，故障排除。

----结束

案例总结

使能 Opaque LSA 能力能力后与其他厂商设备互通时，需确认其他厂商设备的是否缺省使能 Opaque LSA 能力。

4 IS-IS 故障处理

关于本章

4.1 IS-IS 邻居无法建立的定位思路

4.1.1 常见原因

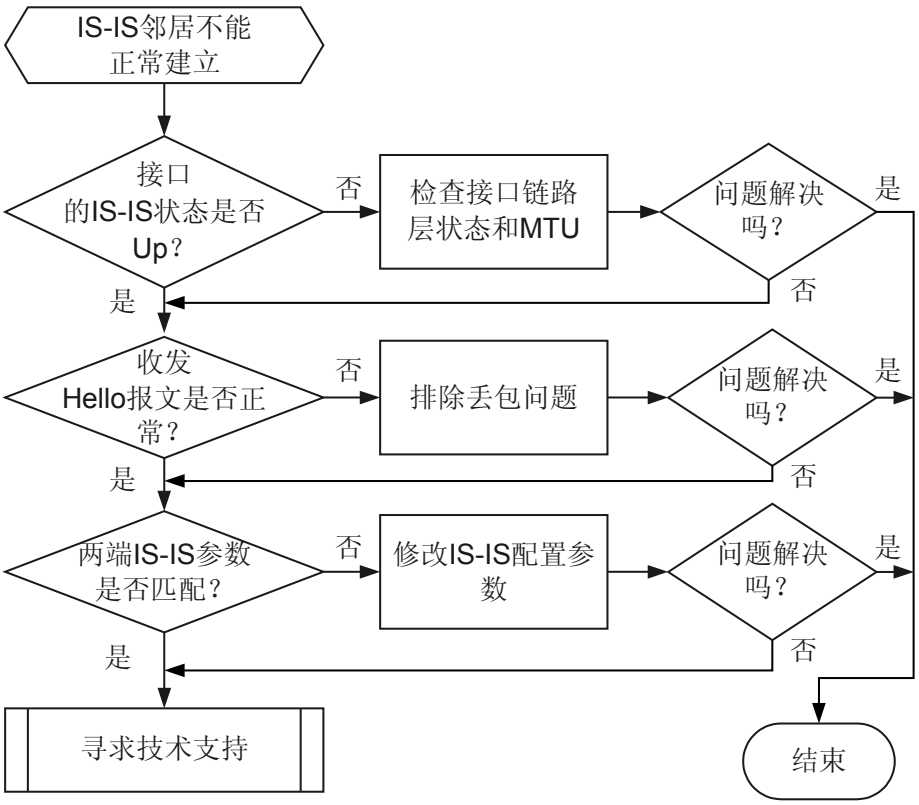
本类故障的常见原因主要包括：

- 设备底层故障或者链路故障导致 IS-IS 无法正常的收发 Hello 报文；
- 链路两端的设备配置的 System ID 相同；
- 链路两端的接口的 MTU 设置不一致或者接口的 MTU 小于发送的 Hello 报文的长度；
- 链路两端的接口的 IP 地址不在同一网段；
- 链路两端的 IS-IS 接口认证方式不匹配；
- 链路两端的 IS-IS Level 不匹配；
- 建立 IS-IS Level-1 邻居时，链路两端设备的区域地址不匹配；

4.1.2 故障诊断流程

可按照故障诊断流程图 4-1 排除故障。

图 4-1 IS-IS 邻居无法建立故障诊断流程图



4.1.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查 IS-IS 接口的状态

执行 **display isis interface** 命令，检查使能了 IS-IS 的接口的状态（“IPv4.State”或“IPv6.State”字段）。

- 如果状态为 **Mtu:Up/Lnk:Dn/IP:Dn**，请执行**步骤 2**。
- 如果状态为 **Mtu:Dn/Lnk:Up/IP:Up**，执行 **display current-configuration interface** *interface-type* [*interface-number*]，检查两端接口的 MTU 的设置。执行 **display current-configuration configuration isis** 命令，检查 IS-IS 进程的 LSP 的长度设置。

  说明

  如果执行 **display current-configuration configuration isis** 命令无法查询到 LSP 的长度的设置，则表示 LSP 的长度设置为缺省值。该缺省值可以通过命令 **display default-parameter isis** 查看，显示字段 **LSP-Originate-Length** 和 **LSP-Receive-Length** 分别表示生成 LSP 的最大长度和接收 LSP 的最大长度。

  如果执行 **display current-configuration interface** *interface-type* [*interface-number*] 命令无法查询到 MTU 的设置，则表示 MTU 为缺省值。POS 接口的缺省 MTU 值为 4470 字节，其余接口缺省 MTU 值为 1500 字节。

  对于 P2P 接口，需要保证 LSP 的长度不大于接口的 MTU 值；对于广播网接口，需要保证 MTU 值减 LSP 的长度大于等于 3。如果不满足该条件，请在 IS-IS 视图下执行 **lsp-length** 命令修改 LSP 的长度，或者修改 MTU 值。如果两端接口 MTU 值不同，请在对应接口视图下将其修改为相同值。

  如果故障仍未排除，请执行**步骤 4**。
- 如果状态为 **Down**，执行 **display current-configuration configuration isis** 检查是否配置了 NET，如果没有配置，请执行 **network-entity** 命令配置 NET。

  如果故障仍未排除，请执行**步骤 2**。
- 如果状态是 **Up**，请执行**步骤 4**。

步骤 2 检查接口是否 Up

执行 **display ip interface** [*interface-type* [*interface-number*]]命令，查看指定接口的状态。

- 如果接口链路层协议状态（**Line protocol current state** 字段）不是 Up，请处理接口故障，使接口链路层协议状态为 Up。详细的故障处理方法请参见物理对接类和接口协议层问题的定位。

  如果故障仍未排除，请执行**步骤 3**。
- 如果接口状态是 Up，请执行**步骤 3**。

步骤 3 检查链路两端接口的 IP 地址是否在同一网段

- 如果 IP 地址不在同一网段，请修改两端的 IP 地址，保证两端的 IP 地址在同一网段。如果故障仍未排除，请执行**步骤 4**。
- 如果 IP 地址在同一网段，请执行**步骤 4**。

步骤 4 检查 IS-IS 收发 Hello 报文是否正常

执行 **display isis statistics packet** [**interface** *interface-type interface-number*]命令，检查 IS-IS 收发报文是否正常。

说明

IS-IS 发送 Hello 报文的缺省间隔是 10 秒，每隔 10 秒执行一次上述命令，查看对应的报文计数是否增长。

对于广播网接口，IS-IS 的 Hello 报文区分 Level，可以根据建立邻居的 Level 查看对应的 Hello 报文计数（**L1 IIH** 或者 **L2 IIH**）；对于 P2P 类型的接口，IS-IS 的 Hello 报文不区分 Level，都记录在 **L2 IIH** 中。

- 如果接收 Hello 报文的计数一直没有增长，请检查 IS-IS 报文是否被丢弃。
  - 对于广播网类型接口，执行 **debugging ethernet packet isis** *interface-type interface-number* 命令。如果有类似如下的信息，表示接口能正常收发 IS-IS 报文。

```
*0.75124950 HUAWEI ETH/7/eth_rcv:Slot=3;Receive an Eth Packet, interface : Ethernet1/0/0,
eth format: 3, length: 60, protoctype: 8000 isis, src_eth_addr: 00e0-fc37-08c1,
dst_eth_addr: 0180-c200-0015
*0.75124950 HUAWEI ETH/7/eth_send:Slot=3;Send an Eth Packet, interface : Ethernet1/0/0, eth
format: 3, length: 112, protoctype: 8000 isis, src_eth_addr: 00e0-fc26-f9d9, dst_eth_addr :
0180-c200-0015
```

  - 对于 P2P 类型接口，执行 **debugging ppp osi-npdu packet** *interface-type interface-number* 命令。如果有类似如下的信息，表示接口能正常收发 IS-IS 报文。

```
*0.85102199 HUAWEI PPP7/debug2:Slot=2;
  PPP Packet:
  Pos2/0/0 Output OSI-NPDU(0023) Pkt, Len 1004
*0.85102199 HUAWEI PPP7/debug2:Slot=2;
  PPP Packet:
  Pos2/0/0 Input OSI-NPDU(0023) Pkt, Len 1501
```

  说明

  display isis interface *interface-type interface-number* 命令的 DIS 字段如果显示“--”表示 P2P 类型的接口；如果显示的不是“--”，表示广播网类型接口。

  如果接口无法正常收发 IS-IS 报文，请执行**步骤 9**。
- 如果设备能够正常接收 Hello 报文，则执行**步骤 5**。

步骤 5 检查链路两端的设备配置的 System ID 是否相同

执行 **display current-configuration configuration isis** 查看链路两端设备的 IS-IS 配置的 System ID 是否相同。

- 如果两端 System ID 相同，请修改配置，使两端的 System ID 不同。
- 如果两端 System ID 不相同，请执行**步骤 6**。

步骤 6 检查链路两端的设备的 IS-IS Level 是否匹配

执行 **display current-configuration configuration isis | include is-level** 命令查看两端 IS-IS 进程的 Level，执行 **display current-configuration interface** *interface-type interface-number* **| include isis circuit-level** 命令，查看接口的 IS-IS Level 的配置，需要保证链路两端的 Level 匹配才能建立起 IS-IS 邻居。

说明

如果使用命令 **display current-configuration interface** *interface-type interface-number* | **include isis circuit-level** 无法查询到接口的 Level 级别，则表示该接口级别为缺省值。该缺省值可以使用命令 **display default-parameter isis** 查看字段 **Circuit-Level**。

接口 Level 匹配的原则如下：

- 如果本端接口 Level 级别为 **Level-1**，则对端接口 Level 级别必须为 **Level-1** 或 **Level-1-2**。
- 如果本端接口 Level 级别为 **Level-2**，则对端接口 Level 级别必须为 **Level-2** 或 **Level-1-2**。
- 如果本端接口 Level 级别为 **Level-1-2**，则对端接口 Level 级别可以为 **Level-1**、**Level-2** 或 **Level-1-2**。

- 如果链路两端 Level 不匹配，请在 IS-IS 视图下使用命令 **is-level** 修改设备的 IS-IS 级别，或者在接口视图下使用命令 **isis circuit-level** 修改接口的 Level 级别。
- 如果链路两端 Level 匹配，请执行**步骤 7**。

步骤 7 检查链路两端设备的区域地址是否匹配

区域地址不匹配时，会出现 IS-IS 区域地址不匹配的告警 **ISIS_1.3.6.1.3.37.2.0.12 isisAreaMismatch**。

说明

如果链路两端建立 Level-1 邻居，需要保证链路两端设备在同一个区域内。

一个 IS-IS 进程最多可以配置 3 个区域地址，两端只要有一个区域地址相同，即可建立 Level-1 邻居。

建立 IS-IS Level-2 邻居时，不需要判断区域地址是否匹配。

- 如果链路两端无相同区域地址，请在 IS-IS 视图下使用命令 **network-entity** 修改设备的区域地址。
- 如果链路两端区域地址匹配，请执行**步骤 8**。

步骤 8 检查链路两端设备的认证方式是否匹配

认证方式不匹配时，会出现 ISIS 认证类型不匹配的告警 **ISIS_1.3.6.1.3.37.2.0.9 isisAuthenticationTypeFailure** 或者认证失败的告警 **ISIS_1.3.6.1.3.37.2.0.10 isisAuthenticationFailure**。

执行 **display current-configuration interface** *interface-type interface-number* | **include isis authentication-mode** 命令查看两端接口的 IS-IS 认证配置。

- 如果两端认证类型不匹配，请在链路两端的 IS-IS 接口视图下执行命令 **isis authentication-mode**，将链路两端设置为相同的认证类型。
- 如果两端认证密码不匹配，请在链路两端的 IS-IS 接口视图下执行命令 **isis authentication-mode**，将链路两端设置为相同的认证密码。
- 如果两端认证匹配，请执行**步骤 9**。

步骤 9 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

4.1.4 相关告警与日志

相关告警

ISIS_1.3.6.1.3.37.2.0.12 isisAreaMismatch

ISIS_1.3.6.1.3.37.2.0.9 isisAuthenticationTypeFailure

ISIS_1.3.6.1.3.37.2.0.10 isisAuthenticationFailure

相关日志

无

4.2 设备学习不到 IS-IS 路由的定位思路

4.2.1 常见原因

本类故障的常见原因主要包括：

- 其它路由协议也发布了相同的路由，并且路由协议优先级比 IS-IS 协议高；
- 引入的外部路由优先级低，没有被优选；
- IS-IS 开销值类型不匹配；
- IS-IS 邻居没有正常建立；
- 两台设备的 System ID 配置相同；
- LSP 报文认证不匹配；
- 设备底层故障或者链路故障，造成 LSP 报文丢失。

4.2.2 故障诊断流程

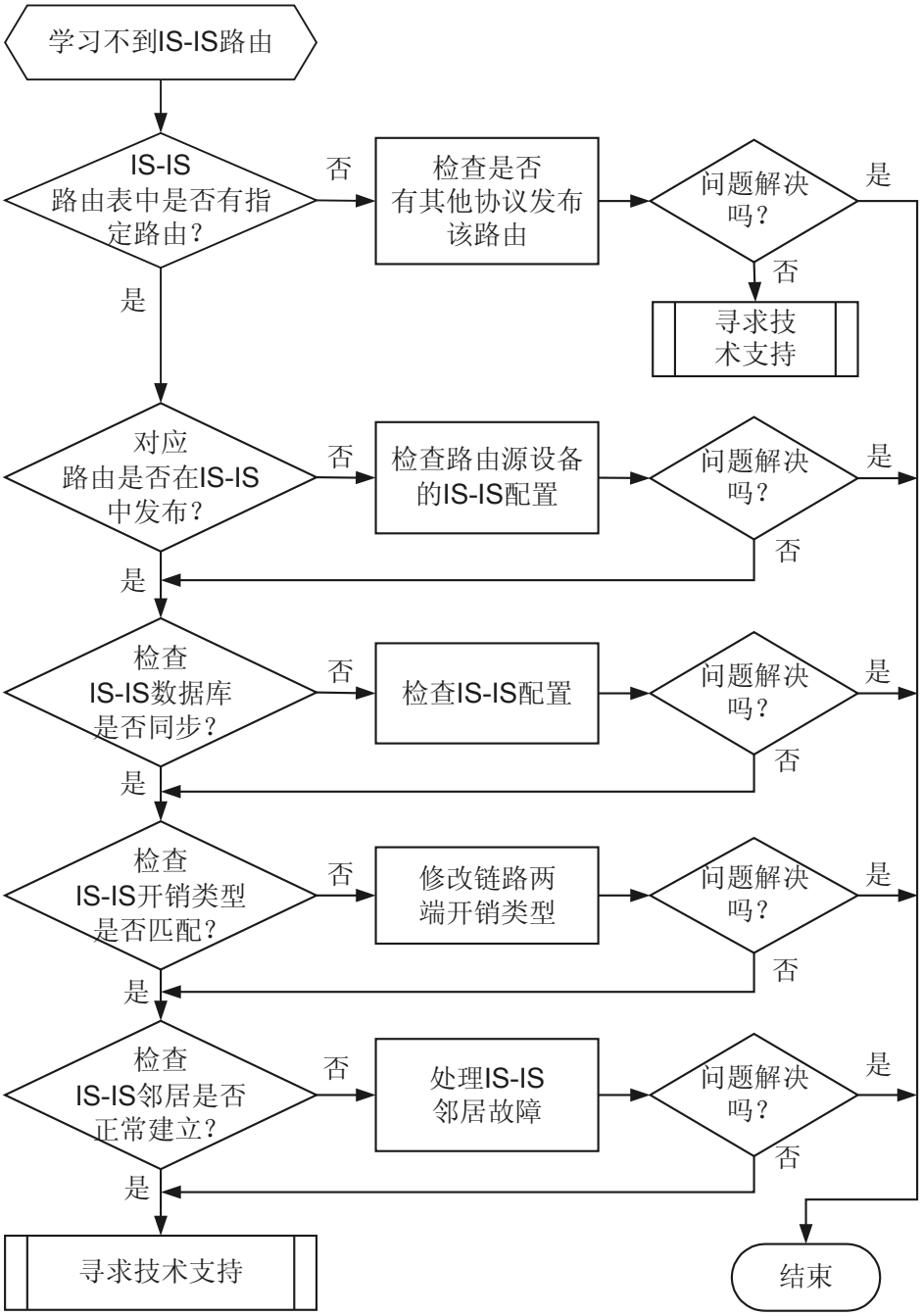
在配置 IS-IS 后发现设备学习不到 IS-IS 路由。

故障诊断思路：

- 检查是否有其它协议也学到了指定路由。
- 检查 IS-IS 是否计算出路由。
- 检查 IS-IS 的 LSDB 数据库是否同步。
- 检查 IS-IS 的配置是否正确。

可按照故障诊断流程图 4-2 排除故障。

图 4-2 设备学习不到 IS-IS 路由的故障诊断流程图



4.2.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查 IS-IS 路由表是否正确

执行 **display isis route** 命令，查看 IS-IS 路由表。

- 如果 IS-IS 路由表中存在指定的路由，执行 **display ip routing-table** *ip-address* [*mask* | *mask-length*] **verbose** 命令查看 IP 路由表中是否存在协议优先级比 IS-IS 高的路由。

  说明

  State 字段为 **Active Adv** 表示该路由为活跃的路由，如果存在相同前缀的多个协议的路由，协议优先级高的路由优选为活跃的路由。

  - 如果存在，请根据网络规划调整配置。
  - 如果不存在，请执行**步骤 6**。
- 如果 IS-IS 路由表中不存在指定的路由，请执行**步骤 2**。

步骤 2 检查指定的 IS-IS 路由是否发布

在发布指定路由的设备上，执行 **display isis lsdb local verbose**，查看本地产生的 LSP 报文中是否携带了指定路由。

- 如果 LSP 报文中没有携带指定的路由，请检查配置是否正确，例如接口是否使能 IS-IS。

  说明

  如果是引入的外部路由，执行 **display ip routing-table protocol** *protocol* **verbose** 命令查看外部路由是否是活跃的。
- 如果 LSP 报文中携带了指定的路由，请执行**步骤 3**。

步骤 3 检查 IS-IS 的数据库是否同步

在学习不到 IS-IS 路由的设备上，执行 **display isis lsdb**，查看是否收到发布指定路由的设备的 LSP 报文。

说明

其中，**LSPID** 是一条 LSP 的标识，**Seq Num** 是报文的序列号，序列号越大表示报文越新。

- 如果 LSDB 数据库中不存在指定的 LSP 报文。
  - 如果产生告警信息 **ISIS_1.3.6.1.3.37.2.0.9 isisAuthenticationTypeFailure** 或 **ISIS_1.3.6.1.3.37.2.0.10 isisAuthenticationFailure**，则表示配置的 LSP 报文认证的认证类型或认证密码不匹配，请修改配置。
  - 如果未产生以上告警信息，请排查设备底层和中间链路是否存在故障。
- 如果 LSDB 数据库中存在指定的 LSP 报文，但 **Seq Num** 与 **display isis lsdb local verbose** 命令显示的不一致，并且 **Seq Num** 不停的增长，则网络中存在其他设备与发布指定路由的设备的 System ID 配置相同，这种情况下会产生告警信息 **ISIS_1.3.6.1.3.37.2.0.8 isisSequenceNumberSkip**，请排查并网络中设备的 IS-IS 配置。
- 如果 LSDB 数据库中存在指定的 LSP 报文，但 **Seq Num** 不一致，并且一直保持不变，可能是 LSP 报文在传输过程中被丢弃，请排查设备底层和中间链路是否存在故障。
- 如果 LSDB 数据库中存在指定的 LSP 报文，并且 **Seq Num** 一致，请执行**步骤 4**。

步骤 4 检查 IS-IS 开销类型是否匹配

分别在发布路由的设备和学习不到路由的设备上，执行 **display current-configuration configuration isis** 命令，查看 IS-IS 的开销类型配置（**cost-style** 命令）。

说明

只有开销类型相同，才能学到路由。

IS-IS 的开销类型可以配置为以下 5 种模式:

- narrow：接收和发送开销值类型为 narrow 的报文。
- narrow-compatible：可以接收开销值类型为 narrow 和 wide 的报文，但却只发送 narrow 的报文。
- compatible：可以接收或发送开销值类型为 narrow 和 wide 的报文。
- wide-compatible：可以接收开销值类型为 narrow 和 wide 的报文，但却只发送 wide 的报文。
- wide：接收或发送开销值类型为 wide 的报文。

如果一端配置是 narrow，另一端配置为 wide 或者 wide-compatible，则两端不能互通。

如果一端配置是 narrow-compatible，另一端配置为 wide，则两端不能互通。

- 如果链路两端的设备的 IS-IS 开销类型不匹配，请执行 **cost-style** 命令修改配置。
- 如果两端的设备的 IS-IS 开销类型匹配，请执行**步骤 5**。

步骤 5 检查 IS-IS 邻居是否正常建立

在路径上的每一台设备上执行 **display isis peer**，查看 IS-IS 邻居是否都正常建立。

- 如果 State 字段不是 Up，请参见 **IS-IS 邻居无法建立的定位思路**。
- 如果 State 字段是 Up，请执行**步骤 6**。

步骤 6 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

4.2.4 相关告警与日志

相关告警

ISIS_1.3.6.1.3.37.2.0.8 isisSequenceNumberSkip

ISIS_1.3.6.1.3.37.2.0.9 isisAuthenticationTypeFailure

ISIS_1.3.6.1.3.37.2.0.10 isisAuthenticationFailure

相关日志

无

4.3 IS-IS 邻居震荡的定位思路

4.3.1 常见原因

本类故障的常见原因主要包括：

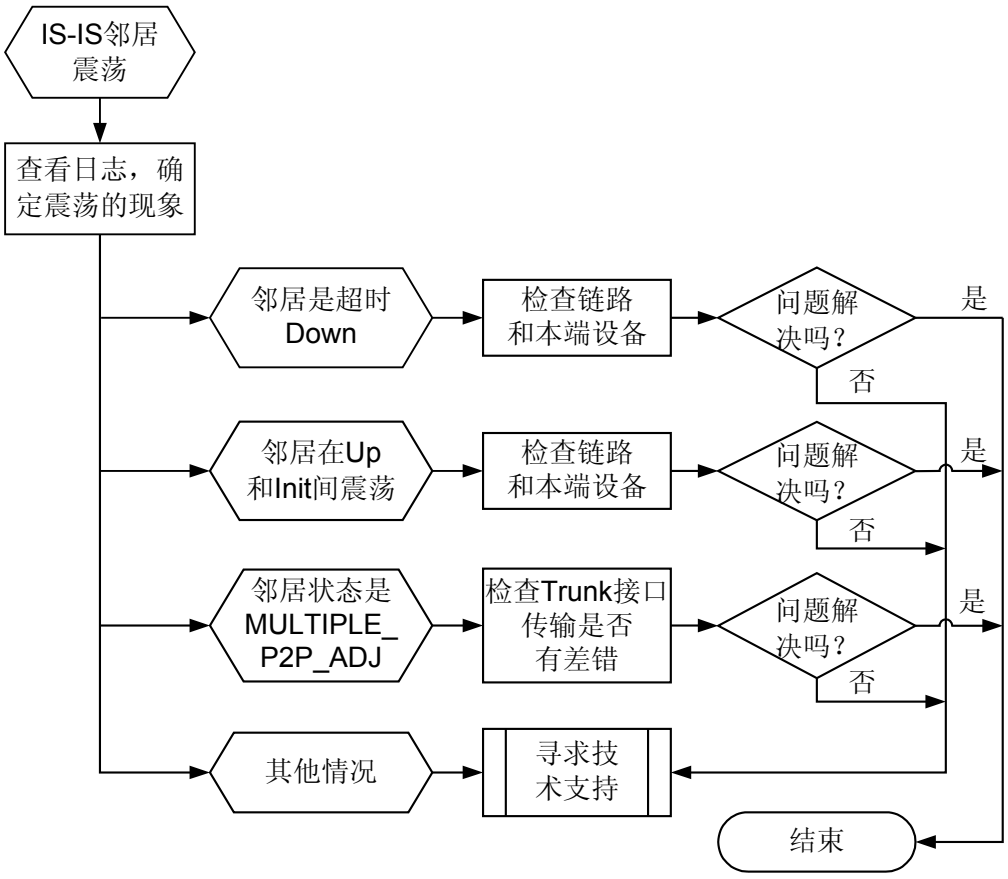
- 链路不稳定或者设备异常，造成丢失报文；
- Trunk 接口的成员口的传输线插错位置。

4.3.2 故障诊断流程

在配置 IS-IS 后发现 IS-IS 邻居震荡。

可按照故障诊断流程图 4-3 排除故障。

图 4-3 IS-IS 邻居震荡的故障诊断流程图



4.3.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查 IS-IS 邻居变化的情况

如果 IS-IS 邻居状态发生变化，就会出现 IS-IS 邻居变化的告警 **ISIS_1.3.6.1.3.37.2.0.17 isisAdjacencyChange** 和日志 **ISIS/4/ADJ_CHANGE_LEVEL**。

说明

只有在 ISIS 进程下配置了 **log-peer-change** 命令后，才会记录 **ISIS/4/ADJ_CHANGE_LEVEL** 的日志信息。

- 如果在 IS-IS 进程下配置了 **log-peer-change** 命令，可以查看日志信息中 **ChangeType** 字段的值。
  - 如果 **ChangeType** 是 HOLDTIMER_EXPIRED，说明是本端设备不能稳定的收到对端设备的 Hello 报文，请排查中间链路和本端设备底层是否存在丢包问题。
  - 如果 **ChangeType** 在 3_WAY_INIT 和 3_WAY_UP 之间震荡（针对 P2P 类型接口），或者 **ChangeType** 都是 NEW_L1_ADJ 或者 NEW_L2_ADJ（针对广播网类型接口），表明邻居状态是在 Up 和 Init 之间震荡，这是对端设备不能稳定的收到本端设备的 Hello 报文导致的，请排查中间链路和对端设备底层是否存在丢包问题。
  - 如果 **ChangeType** 是 MULTIPLE_P2P_ADJ，并且接口是 IP-Trunk 接口，请检查 Trunk 接口下绑定的接口是否插错线。
  - 其他情况，请执行**步骤 2**。
- 如果没有配置 **log-peer-change**，可以连续执行 **display isis peer** 命令，观察 **State** 和 **HoldTime** 字段的值，确定 IS-IS 邻居变化的情况。
  - 如果邻居震荡时，**State** 字段值不变，**HoldTime** 字段值一直减小，直到减小到 0 后邻居关系被删除，说明是本端设备不能稳定的收到对端设备的 Hello 报文，请排查中间链路和本端设备底层是否存在丢包问题。
  - 如果邻居震荡时，**State** 在 Up 和 Init 之间变化，这是对端设备不能稳定的收到本端设备的 Hello 报文导致的，请排查中间链路和对端设备底层是否存在丢包问题。
  - 其他情况，请执行**步骤 2**。

步骤 2 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

4.3.4 相关告警与日志

相关告警

ISIS_1.3.6.1.3.37.2.0.17 isisAdjacencyChange

相关日志

ISIS/4/ADJ_CHANGE_LEVEL

4.4 IS-IS 路由震荡的定位思路

4.4.1 常见原因

本类故障的常见原因主要包括：

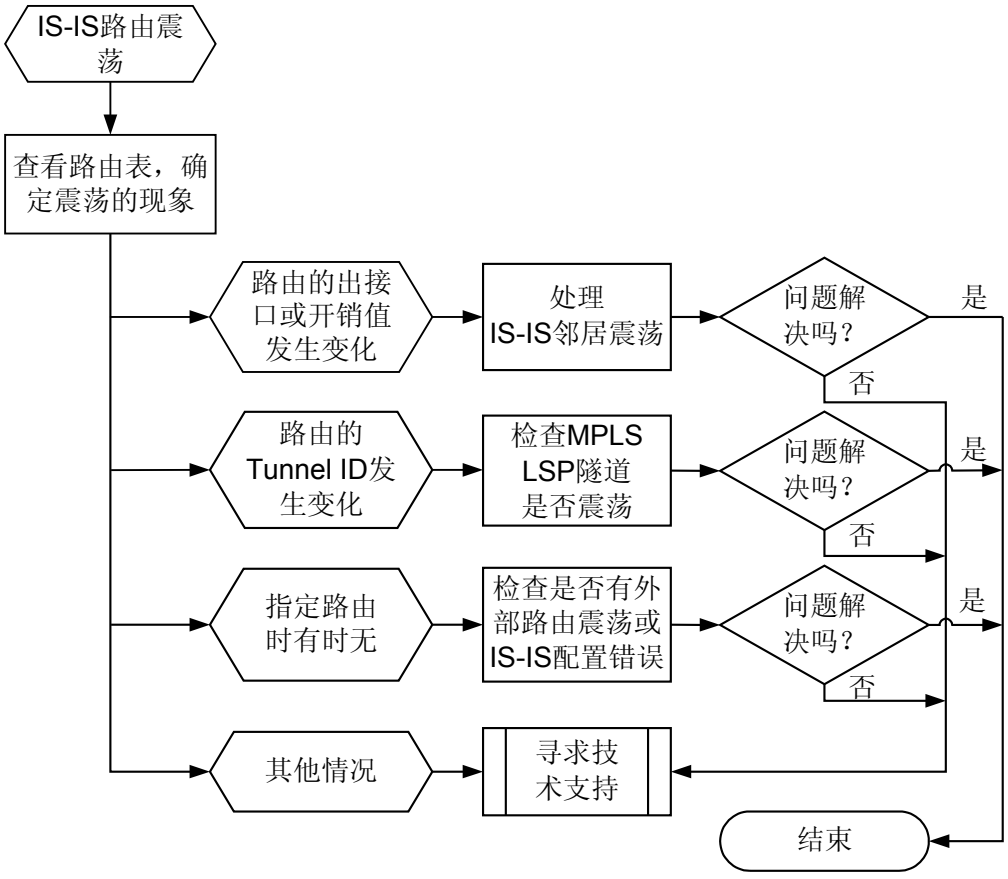
- IS-IS 邻居震荡；
- MPLS LSP 隧道震荡；
- 两台设备的 IS-IS 引入了相同的外部路由，并且外部路由的优先级比 IS-IS 协议的优先级低；
- 网络上两台设备的 System ID 配置相同。

4.4.2 故障诊断流程

在配置 IS-IS 后发现 IS-IS 路由震荡。

可按照故障诊断流程图 4-4 排除故障。

图 4-4 IS-IS 路由震荡的故障诊断流程图



4.4.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查路由震荡的情况

执行 **display ip routing-table** *ip-address* **verbose** 命令，查看路由震荡的具体情况，例如路由表中 **Active** 的路由是从哪个协议学到的，路由震荡的时候，哪些属性发生了变化。

- 如果路由震荡的前后，**TunnelID** 字段发生了变化，请检查 MPLS LSP 隧道是否存在震荡（执行 **display trapbuffer** 命令查看告警信息中是否存在多条 **LSPM_1.3.6.1.2.1.10.166.2.0.2 mplsXCDown** 和 **LSPM_1.3.6.1.2.1.10.166.2.0.1 mplsXCUp**）。如果 MPLS LSP 隧道振荡，请参考 **LDP LSP 振荡的定位思路**或 **TE Tunnel 由 Up 突然变 Down 的定位思路**排查 LSP 振荡问题。
- 如果路由的 **Cost** 或者 **Interface** 字段发生变化，请检查该路由路径上的 IS-IS 邻居是否在震荡，详细的故障处理方法请参考 **IS-IS 邻居震荡故障处理**。
- 如果路由在路由表中时有时无（**Age** 字段在震荡），执行 **display isis lsdb verbose** 命令，确定路由是由哪条 LSP 报文携带的，并且根据查看到的 LSP 报文，执行 **display isis lsdb** *lsp-id* **verbose** 查看这条 LSP 的更新情况。
  - 如果 LSP 中一直携带指定的路由，请检查该路由路径上是否存在 IS-IS 邻居震荡，详细的故障处理方法请参考 **IS-IS 邻居震荡故障处理**。
  - 如果 LSP 的 **Seq Num** 字段值在不停的增加，请检查网络中是否有两台设备配置了相同的 System ID。
  - 如果 LSP 的 **Seq Num** 字段值在不停的增加，并且 LSP 更新前后，指定的路由时有时无，请在产生该 LSP 的设备上执行**步骤 2**。

    说明

    display isis lsdb*lsp-id* **verbose** 命令的显示信息中，IP-Internal 字段或+IP-Internal 字段对应的 IP 地址所在的设备就是产生该 LSP 的设备。
- 如果路由的 **Protocol** 字段发生变化，请执行**步骤 2**。

步骤 2 检查 IS-IS 引入外部路由的配置

如果指定的路由是作为外部路由引入到 IS-IS 的，在引入到 IS-IS 的设备上，执行 **display ip routing-table** *ip-address* **verbose** 命令，查看路由震荡的具体情况。

- 如果路由表中 **Active** 的路由是 IS-IS 路由，而不是 IS-IS 要引入的外部路由，说明网络中还存在其它设备的 IS-IS 也发布了相同的路由，请根据网络规划修改路由协议的优先级，或者在 IS-IS 下配置路由过滤策略，控制下发到 IP 路由表的路由。
- 其它情况，请执行**步骤 3**。

步骤 3 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

4.4.4 相关告警与日志

相关告警

无

相关日志

无

4.5 IS-IS 组播多拓扑中路由信息不正确的定位思路

介绍配置 IS-IS 组播多拓扑时出现 IS-IS 组播拓扑中路由信息不正确的故障原因、处理流程和详细的故障处理步骤。

4.5.1 常见原因

本类故障的常见原因主要包括：

- IS-IS 接口上组播配置错误，导致 IS-IS 邻居组播拓扑状态不正确；
- 接口没有使能 PIM-SM，导致没有在 IS-IS 组播拓扑中发布相关路由；
- 组播没有使用 IS-IS 的 *multicast* 拓扑。

4.5.2 故障诊断流程

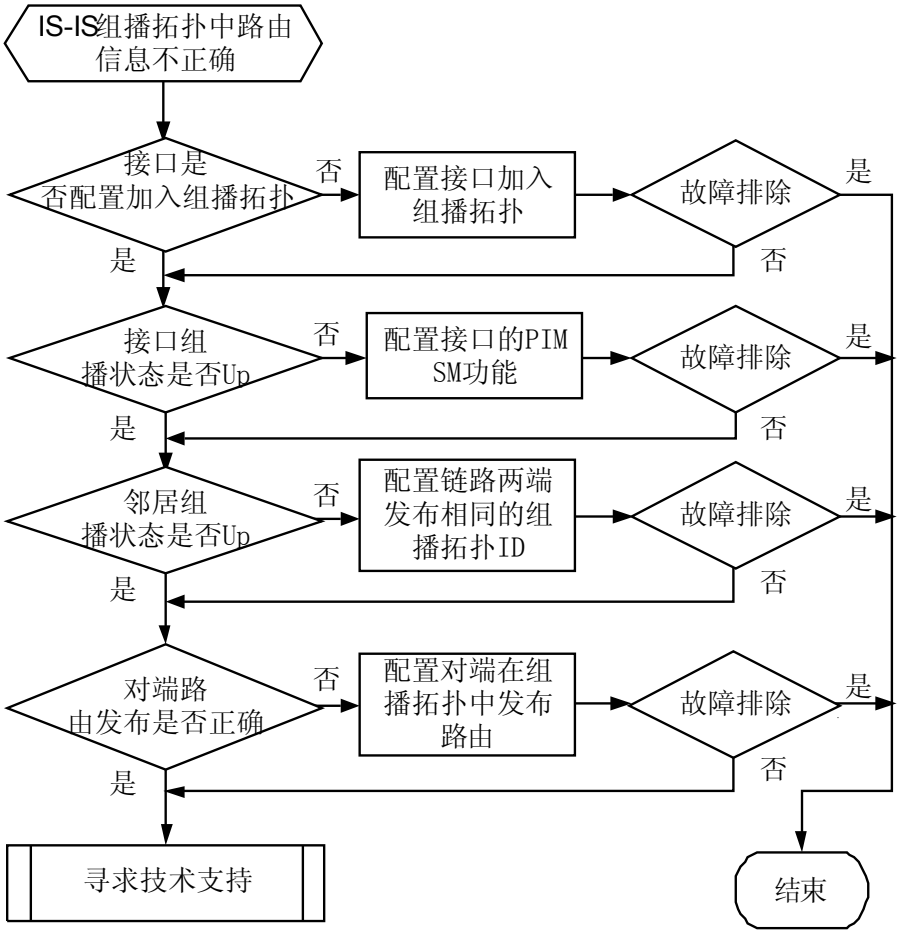
在配置 IS-IS 后发现 IS-IS 组播拓扑中路由信息不正确。

故障定位思路：

- 检查 IS-IS 接口是否配置加入组播拓扑。
- 检查 IS-IS 接口组播状态是否 Up。
- 检查 IS-IS 邻居的组播状态是否为 Up。
- 检查 IS-IS 的组播拓扑路由是否正确。

可按照故障诊断流程图 4-5 排除故障。

图 4-5 IS-IS 组播多拓扑中路由信息不正确的诊断流程图



4.5.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查 IS-IS 接口是否加入组播拓扑

执行 **display isis interface verbose** 命令，查看 IS-IS 接口是否已经加入组播拓扑。

- 如果显示信息中未包含字段 **IPv4 MT 3**，则该接口未加入组播拓扑。请在指定接口视图下执行命令 **isis topology** *multicast*，配置接口加入组播拓扑。

说明

在配置将接口加入组播拓扑之前，需要首先执行：

- 在系统视图下执行命令 **ip topology** *multicast* 创建组播拓扑。
- 在 IS-IS 视图下执行命令 **cost-style wide** 或 **cost-style wide-compatible** 修改 IS-IS 的开销类型。
- 在 IS-IS 视图下执行命令 **topology** *multicast* **topology-id** *3* 使能 IPv4 IS-IS 组播拓扑。
- 在接口视图下执行命令 **ip topology** *multicast* **enable** 使接口与组播拓扑实例绑定。
- 在接口视图下执行命令 **isis topology** *multicast* 使能 IS-IS 接口与 IS-IS 组播拓扑实例绑定。

- 如果显示信息中包含字段 **IPv4 MT 3**，则该接口已经加入组播拓扑。请执行步骤 2。

步骤 2 检查 IS-IS 接口的组播拓扑状态是否 Up

执行 **display isis interface verbose** 命令，查看 IS-IS 接口的组播拓扑状态是否 Up。

说明

I 如果显示信息字段 **IPv4 MT 3** 值为 **multicast**，则 IS-IS 接口的组播拓扑状态为 **Up**；如果显示信息字段 **IPv4 MT 3** 值为 **multicast (NO PIM)**，则 IS-IS 接口的组播拓扑状态没有 **Up**。

- 如果 IS-IS 接口的组播拓扑状态没有 Up，请在指定接口视图下执行命令 **pim sm** 使能 PIM-SM。
- 如果 IS-IS 接口的组播拓扑状态已经 Up，请执行步骤 3。

步骤 3 检查 IS-IS 的邻居组播拓扑状态是否 Up

执行 **display isis peer verbose** 命令，检查 IS-IS 的邻居组播拓扑状态是否 Up。

说明

如果显示信息字段 **MT IDs supported** 中包含 **3(UP)**，则 IS-IS 的邻居组播拓扑状态为 **Up**；否则 IS-IS 的邻居组播拓扑状态没有 **Up**。其中数字 **3** 为 IPv4 IS-IS 组播拓扑 ID。

- 如果 IS-IS 的邻居组播拓扑状态没有 **Up**，请在对端设备上执行步骤 1 和步骤 2 进行检查，保证组播拓扑中的所有设备及接口的配置正确，包括产生源路由的接口也同样需要加入 IS-IS 组播拓扑，并且执行命令 **pim sm** 使能 PIM-SM。
- 如果 IS-IS 邻居的组播拓扑状态已经 **Up**，请执行步骤 4。

步骤 4 检查 IS-IS 路由表中的组播拓扑路由是否正确

执行 **display isis route topology** *multicast* 命令，检查 IS-IS 组播拓扑中的路由信息是否正确。

- 如果路由信息正确，则故障已经排除。
- 如果路由信息不正确，请执行步骤 5。

步骤 5 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

4.5.4 故障处理步骤补充（华为工程师专用）

本手册只供华为工程师使用，严禁外传。

说明

本节是对“故障处理步骤”的补充。如果华为工程师在执行“故障处理步骤”的步骤后仍不能解决故障，可以根据本节提供的内部定位方法继续定位。

执行 **display current-configuration** 命令，检查组播拓扑中 RPF 检查的应用方案，并确认该方案是否正确。

执行 **display isis spf-tree** 命令，检查组播拓扑中 IS-IS 的 SPF 计算出的 SPT（Shortest Path Tree，最短路径树）是否正确。

4.5.5 相关告警与日志

相关告警

无

相关日志

无

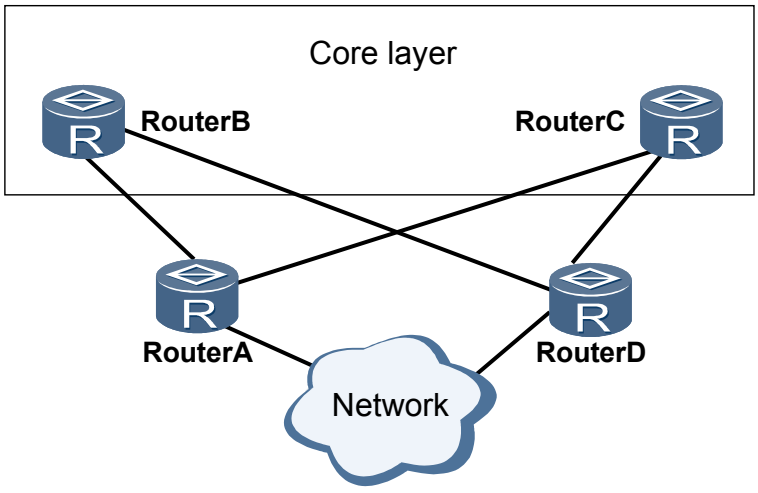
4.6 相关案例

4.6.1 由于 IS-IS 路由引入类型与其他厂商设备不一致导致上层设备无法学习 IS-IS 路由

网络环境

在图 4-6 的网络中，RouterB 和 RouterC 两台设备位于核心层，下挂两台 SR 设备 RouterA 和 RouterD，其中，RouterD 为其他厂商设备。所有设备部署 IS-IS，都为 Level-2 设备。为实现负载分担，在 RouterA 和 RouterD 连接了相同的网络，并在 IS-IS 进程下引入直连路由和静态路由到 IS-IS。配置后发现核心层两台设备 RouterB 和 RouterC 上只能从 RouterD 学习到路由。

图 4-6 设备无法学习 IS-IS 路由



故障分析

由于 RouterD 引入静态路由到 IS-IS 时，缺省引入类型为 **internal**，**cost** 为路由的原有开销值，而 RouterA 引入静态路由到 IS-IS 时缺省为 **external**，**cost** 为路由的原有开销值+64。由于开销值不一样，RouterB 和 RouterC 两台设备优选从 RouterD 学到路由。

说明

此问题只有 cost-style 为 narrow 才可能出现。

操作步骤

步骤 1 在 RouterA 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **isis** *process-id*，进入 IS-IS 视图。

步骤 3 执行命令 **import-route direct cost-type internal**, 引入直连路由，设定 **cost-type** 为 **internal**。

步骤 4 执行命令 **import-route static cost-type internal**，引入静态路由，设定 **cost-type** 为 **internal**。

说明

将 **cost-type** 由 **external** 改为 **internal** 后，路由的 **cost** 为原有开销值而不是原有开销值+64。

完成上述操作后，在 RouterB 和 RouterC 设备上使用命令 **display isis route** 查看路由信息，可以看到有两条到相同 IP 网段的 IS-IS 路由，RouterA 和 RouterD 进行负载分担。

----结束

案例总结

在多厂商设备的组网环境中，要注意各厂商设备之间的实现差异。

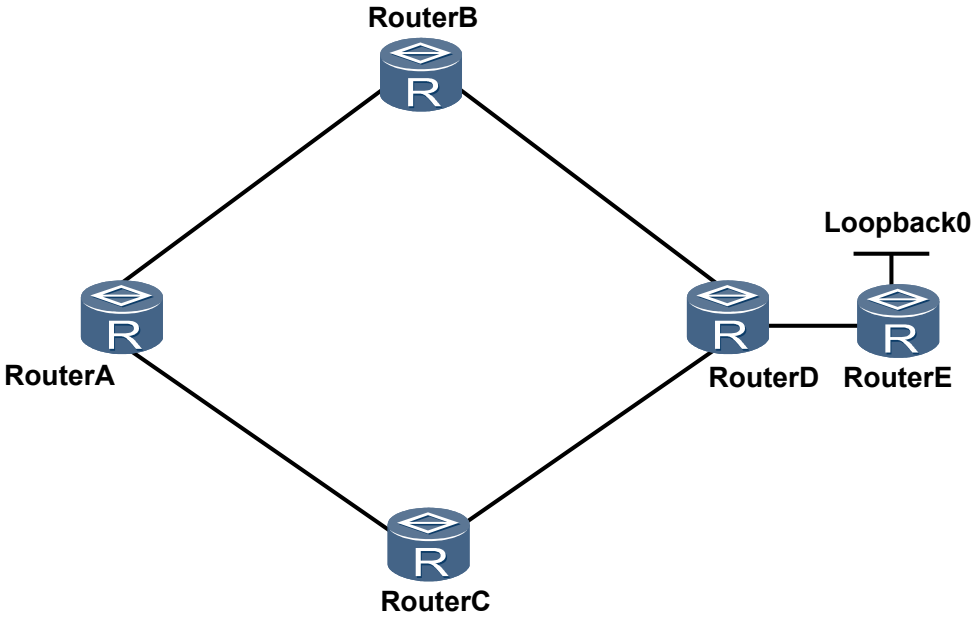
4.6.2 IS-IS 组播拓扑中邻居状态不能 Up 导致路由不正确

网络环境

在图 4-7 的网络中，所有设备运行 IS-IS 协议，其中 RouterA、RouterB、RouterD 和 RouterE 运行组播拓扑业务，其中 RouterE 的 Loopback0 接口也运行组播拓扑业务。

配置完成后，发现在 RouterA 上无法看到 RouterE 的 Loopback0 接口的路由。

图 4-7 IS-IS 组播拓扑中邻居状态不能 Up 导致路由不正确组网图



故障分析

1. 在 RouterA 上执行 **display isis interface verbose** 命令，可以看到显示信息中包含字段 **IPv4 MT 3**，即 RouterA 的接口已经加入了组播拓扑。
2. 在 RouterA 上执行 **display isis interface verbose** 命令，发现 **IPv4 MT 3** 字段值为 **NO PIM SM**，即该接口未使能 PIM SM。因为 IS-IS 接口上组播拓扑未激活，导致 IS-IS 邻居组播拓扑状态不能 Up，无法建立组播拓扑最短路径树，最终无法计算组播拓扑路由。

操作步骤

步骤 1 在 RouterA 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **interface interface-type** *interface-number*，进入接口视图。

步骤 3 执行命令 **pim sm**，使能接口的 PIM SM 功能。

完成上述操作后，在 RouterA 上执行命令 **display isis route** 查看路由信息，可以看到 RouterE 的 Loopback0 接口的路由。

----结束

案例总结

由于 IS-IS 组播多拓扑需要与 PIM SM 联动，所以配置组播业务时，需保证接口上已经配置了 PIM SM，IS-IS 组播多拓扑才能正确运行。

5 BGP 故障处理

关于本章

5.1 BGP 邻居无法建立的定位思路

5.1.1 常见原因

BGP 邻居无法建立是指 BGP 邻居状态无法到达 Established 状态。

本类故障的常见原因主要包括：

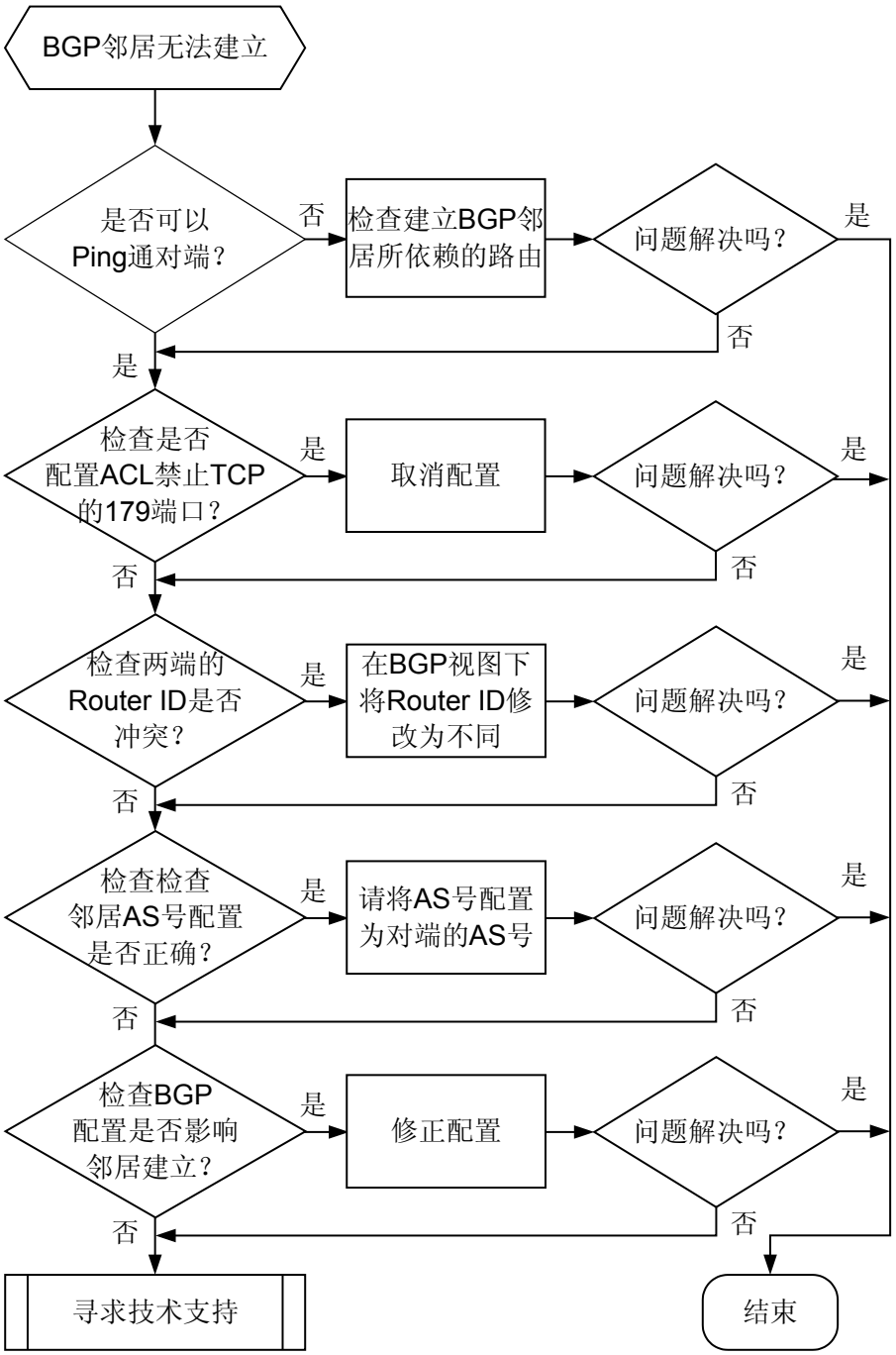
- BGP 报文转发不通
- ACL 过滤了 TCP 的 179 端口
- 邻居的 Router ID 冲突
- 配置的邻居的 AS 号错误
- 用 Loopback 口建立邻居时没有配置 **peer connect-interface**
- 用 Loopback 口建立 EBGP 邻居未配置 **peer ebgp-max-hop**
- **peer valid-ttl-hops** 配置错误。
- 对端发送的路由数量是否超过 **peer route-limit** 命令设定的值。
- 对端配置了 **peer ignore**
- 两端的地址族不匹配

5.1.2 故障诊断流程

在配置 BGP 协议后发现 BGP 邻居无法建立。

可按照故障诊断流程**图 5-1** 排除故障。

图 5-1 BGP 邻居无法建立故障诊断流程图



5.1.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 使用 **ping** 命令检测 BGP 邻居之间是否可以 Ping 通

- 如果可以 Ping 通，则说明 BGP 邻居之间有可达的路由并且链路传输也没有问题，请执行**步骤 2**。

  说明

  请使用命令 **ping** **a** *source-ip-address* **s** *packetsize* *host* 来检测两端的互通性，因为带源地址可以同时检测两端路由是否正常，指定 ping 的字节可以检查大包在链路上传输是否正常。

- 如果不能 Ping 通，请参见 **Ping 不通问题**检查两端的路由表中是否存在对端路由。

步骤 2 检查是否配置 ACL 禁止 TCP 的 179 端口

在两端执行 **display acl all** 命令查看是否禁止 TCP 的 179 端口。

```
<HUAWEI> display acl all
 Total nonempty ACL number is 1

Advanced ACL 3001, 2 rules
Acl's step is 5
 rule 5 deny tcp source-port eq bgp
 rule 10 deny tcp destination-port eq bgp
```

- 如果有禁止 TCP 的 179 端口的 ACL，请执行 **undo rule** *rule-id* **destination-port** 和 **undo rule** *rule-id* **source-port** 命令取消配置。
- 如果没有禁止 TCP 的 179 端口的 ACL，请执行**步骤 3**。

步骤 3 检查邻居的 Router ID 是否冲突

在两端分别查看无法建立的 BGP 邻居的情况，例如 ipv4 单播邻居无法建立可以执行 **display bgp peer** 命令，查看 Router ID 是否冲突。显示 Router ID 信息的命令行示例如下，该例中本端的 Router ID 是 **223.5.0.109**。

```
<HUAWEI> display bgp peer
 BGP local router ID : 223.5.0.109
 Local AS number : 41976
 Total number of peers : 12                 Peers in established state : 4

  Peer            V          AS  MsgRcvd  MsgSent  OutQ  Up/Down       State PrefRcv
  8.9.0.8         4         100     1601     1443     0 23:21:56 Established   10000
  9.10.0.10       4         200     1565     1799     0 23:15:30 Established    9999
```

说明

查看 BGP-VPNv4 地址族或 BGP-VPN 实例地址族的邻居可以使用命令 **display bgp vpnv4 all peer**。

- 如果 Router ID 冲突，请在 BGP 视图下运行命令 **router id** 将 Router ID 修改为不同（一般会用 Loopback 口的地址作为本端的 Router ID）。
- 如果 Router ID 没有冲突，请执行**步骤 4**。

步骤 4 检查邻居 AS 号配置是否正确

在两端分别执行 **display bgp peer**，检查邻居的 AS 号是否是对端的 AS 号。

```
<HUAWEI> display bgp peer
 BGP local router ID : 223.5.0.109
 Local AS number : 41976
 Total number of peers : 12                 Peers in established state : 4

  Peer            V          AS  MsgRcvd  MsgSent  OutQ  Up/Down       State PrefRcv
  8.9.0.8         4         100     1601     1443     0 23:21:56 Established   10000
  9.10.0.10       4         200     1565     1799     0 23:15:30 Established    9999
```

说明

查看 BGP-VPNv4 地址族或 BGP-VPN 实例地址族的邻居可以使用命令 **display bgp vpnv4 all peer**。

- 如果 AS 号配置错误，请将 AS 号配置为对端的 AS。
- 如果 AS 号配置没有错误，请执行**步骤 5**。

步骤 5 检查 BGP 配置是否影响邻居建立

通过 **display current-configuration configuration bgp** 查看 BGP 的配置，进行如下检查。

检查项	说明
peer connect-interface *interface-type interface-number*	如果邻居两端使用 Loopback 口建立邻居，则需要使用命令 **peer connect-interface** 指定相应的 Loopback 口为发送 BGP 报文的源接口。
peer ebgp-max-hop *hop-count*	如果直连设备用 Loopback 口建立 EBGP 邻居，或者非直连多跳设备建立 EBGP 邻居，则需要配置命令 **peer ebgp-max-hop** 指定允许的最大跳数 *hop-count*。 ● 直连设备使用 Loopback 口建立连接时，*hop-count* 只要大于 1 即可。 ● 非直连设备建立连接时需要指定 *hop-count* 为相应的跳数。
peer valid-ttl-hops *hops*	如果有该配置，请确认 **peer valid-ttl-hops** *hops* 是否正确：如果配置为 *hops*，则被检测的报文的 TTL 值有效范围为[255 - *hops*+1, 255]。其中 *hops* 是 BGP 会话两端之间的跳数值，直连设备之间的 *hops* 为 1。 **说明** 命令 **peer valid-ttl-hops** 的配置是对称的，即需要在 BGP 会话两端同时使能该命令。
peer route-limit *limit*	如果有该配置时，请确认对端发送的路由数量是否超过 **peer route-limit** *limit*，其中 *limit* 表示限制的路由数量。如果是，则需要降低对端发送过来的路由数量，并在本端使用 **reset bgp** *ip-address* 命令复位相应的 BGP 连接来触发 BGP 重新建立连接。
peer ignore	如果对端配置了 **peer ignore**，说明由于某种原因对端暂时不想和本端建立邻居。如果想建立邻居时，执行 **undo peer ignore** 命令去使能对端的配置即可。
地址族能力	请检查 BGP 会话两端的地址族能力是否匹配。例如，建立 BGP VPNv4 邻居时，需要两端都要在 BGP-VPNv4 地址族下配置命令 **peer enable**。如果一端已配置而另一端没有配置时，没有配置的一端 BGP 邻居状态为“No neg”。

步骤 6 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

5.1.4 相关告警与日志

相关告警

BGP_1.3.6.1.2.1.15.7.2 bgpBackwardTransition

相关日志

BGP/3/STATE_CHG_UPDOWN

BGP/3/WRONG_ROUTERID

BGP/3/WRONG_AS

5.2 BGP 公网流量中断的定位思路

5.2.1 常见原因

BGP 公网流量中断是指在 BGP 邻居关系正常的情况下，依赖 BGP 公网路由建立起来的流量的中断。

本类故障的常见原因主要包括：

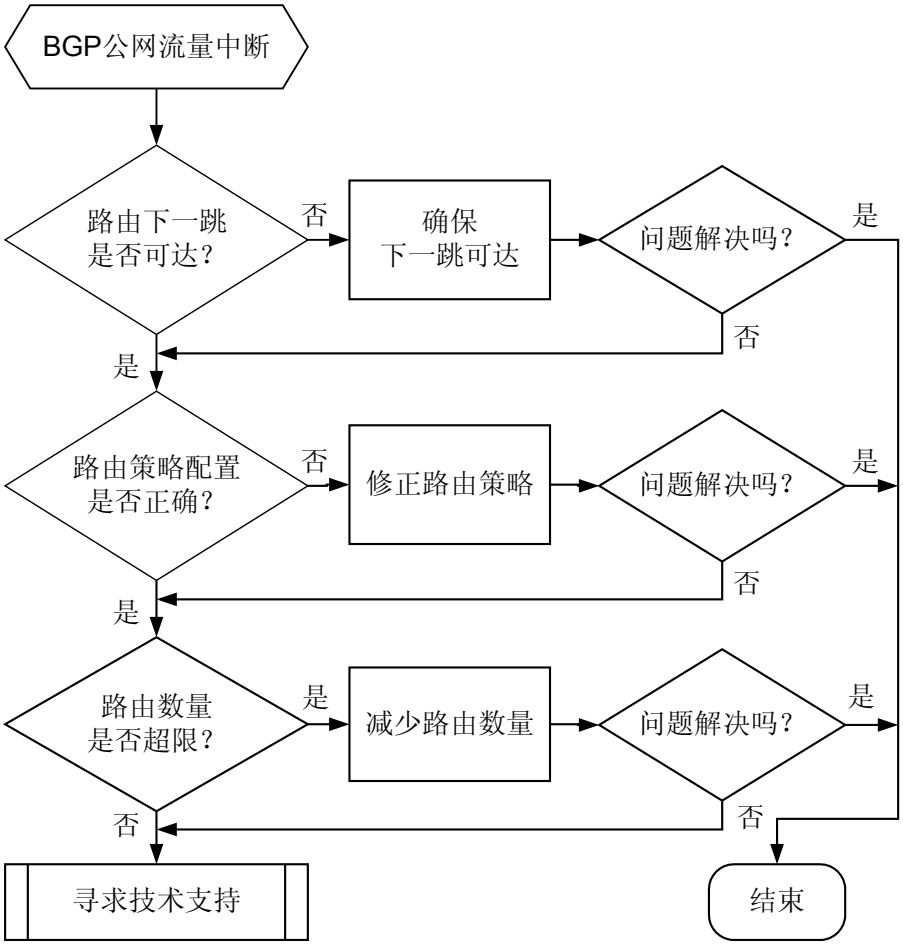
- 路由下一跳不可达导致路由不活跃。
- 路由策略配置不当导致路由无法发布/接收。
- 路由数量超限导致收到的路由被丢弃。

5.2.2 故障诊断流程

在配置 BGP 协议后发现 BGP 公网流量中断。

可按照故障诊断流程图 5-2 排除故障。

图 5-2 BGP 公网流量中断故障诊断流程图



5.2.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查路由下一跳是否可达

在路由的发送端执行 **display bgp routing-table** *network* { *mask* | *mask-length* }命令查看目标路由（*network* 表示目标路由前缀），确认路由是否活跃，并且查看此路由是否已经被发送给路由接收端。命令示例如下：

以 13.0.0.0/8 这条路由举例，显示此路由是活跃的（valid）和优选的（best），并且发送给了邻居 3.3.3.3，此路由的 BGP 下一跳为 1.1.1.1（Original nexthop），经过迭代后的下一跳为 172.1.1.1（Relay IP Nexthop）。

```
<HUAWEI> display bgp routing-table 13.0.0.0 8
```

```
 BGP local router ID : 23.1.1.2
 Local AS number : 100
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 13.0.0.0/8:
 From: 1.1.1.1 (121.1.1.1)
 Route Duration: 4d21h29m39s
 Relay IP Nexthop: 172.1.1.1
 Relay IP Out-Interface: GigabitEthernet1/0/2
 Original nexthop: 1.1.1.1
 Qos information : 0x0
 AS-path Nil, origin incomplete, localpref 100, pref-val 0, valid, internal, best, select, active,
pre 255
 Aggregator: AS 100, Aggregator ID 121.1.1.1
 Advertised to such 1 peers:
    3.3.3.3
```

- 如果目标路由不活跃，请确认 IP 路由表中是否存在到 BGP 下一跳（Original nexthop）的路由，如果不存在说明 BGP 路由不发布是由于路由下一跳不可达导致，请确认为何没有到 BGP 下一跳（Original nexthop）的路由（一般属于 IGP 或静态路由问题）。
- 如果目标路由活跃，却没有被优选（没有 best），请确认 IP 路由表中是否有其他协议优先级（preference）更高的路由存在。如果有请确认是否需要将此路由引入到 BGP 中或调整其协议优先级。 如果没有请联系华为技术工程师。

  说明

  在 BGP 路由表中同一前缀可能有多条路由，其中最多只有 1 条路由会被优选（best），并且只有被优选的路由才会被添加到 IP 路由表中并发送给其他邻居。BGP 路由与其他协议路由进行比较时是靠协议优先级（preference）来决定哪个更优的。

- 如果目标路由活跃且被优选，但没有显示发送给路由接收端，请执行**步骤 2**（重点检查路由发送端的出口策略）。

在路由接收端执行 **display bgp routing-table** *network* { *mask* | *mask-length* }查看是否收到目标路由。

- 如果收到目标路由，请重复执行**步骤 1** 判断路由下一跳是否可达并且是否被优选。
- 如果没有收到目标路由，请执行**步骤 2**（重点检查路由接收端的入口策略）。

步骤 2 检查路由策略是否正确

在路由的发送端/接收端执行 **display current-configuration configuration bgp** 命令查看 BGP 配置，确认是否配置邻居的出口/入口策略。

```
<HUAWEI> display current-configuration configuration bgp
#
bgp 100
 peer 1.1.1.1 as-number 100
 #
 ipv4-family unicast
  undo synchronization
  filter-policy ip-prefix aaa import
  filter-policy ip-prefix aaa export
  peer 1.1.1.1 enable
  peer 1.1.1.1 filter-policy acl-name acl-name import
  peer 1.1.1.1 filter-policy acl-name acl-name export
  peer 1.1.1.1 as-path-filter 1 import
  peer 1.1.1.1 as-path-filter 1 export
  peer 1.1.1.1 ip-prefix prefix-name import
  peer 1.1.1.1 ip-prefix prefix-name export
  peer 1.1.1.1 route-policy policy-name import
  peer 1.1.1.1 route-policy policy-name export
 #
 ipv4-family vpnv4
  policy vpn-target
  peer 1.1.1.1 enable
```

```
#
return
```

- 如果两端配置了出口/入口策略，则需要确认这些策略是否会把目标路由过滤掉，导致该路由无法正常收发。路由策略的具体配置请参见《HUAWEI NetEngine80E/40E 路由器 配置指南-IP 路由》。
- 如果两端没有配置相应的出口/入口策略，请直接执行**步骤 3**。

步骤 3 检查路由是否超限

在路由接收端执行 **display current-configuration configuration bgp | include peer** *destination-address* 和 **display current-configuration configuration bgp | include peer** *group-name*（如果 Peer 被加入到对等体组中）命令查看 BGP 配置，确认是否配置邻居路由限制。

例如，限制只能从邻居 1.1.1.1 收 5 条路由，超限之后将丢弃路由并记录日志。

```
<HUAWEI> display current-configuration configuration bgp | include peer 1.1.1.1
 peer 1.1.1.1 as-number 100
 peer 1.1.1.1 route-limit 5 alert-only
  peer 1.1.1.1 enable
```

如果 BGP 邻居被加入到组中，显示信息中有可能没有 route-limit 的配置。

```
<HUAWEI> display current-configuration configuration bgp | include peer 1.1.1.1
 peer 1.1.1.1 as-number 100
 peer 1.1.1.1 group IBGP
  peer 1.1.1.1 enable
  peer 1.1.1.1 group IBGP
```

这种情况下，需要使用 **display current-configuration configuration bgp | include peer** *group-name* 来查看该对等体组的配置。

```
<HUAWEI> display current-configuration configuration bgp | include peer IBGP
 peer IBGP route-limit 5 alert-only
  peer IBGP enable
```

如果流量中断时，产生了路由超限日志 **BGP/3/ROUTPRIX_EXCEED**，表示路由超限导致目标路由被丢弃，则需要扩大本端的路由限制数值。

说明

修改 BGP 邻居限制的最大路由数量时会中断邻居，建议在路由发送端通过路由聚合以减少路由数量来解决。

步骤 4 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

5.2.4 相关告警与日志

相关告警

BGP_1.3.6.1.4.1.2011.5.25.177.1.3.1 hwBgpPeerRouteNumThresholdExceed

相关日志

BGP/3/ROUTPRIX_EXCEED

5.3 私网流量中断的定位思路

5.3.1 常见原因

BGP 私网流量中断是指在 BGP 邻居正常的情况下依赖 BGP 私网路由的流量的中断。

本类故障的常见原因主要包括：

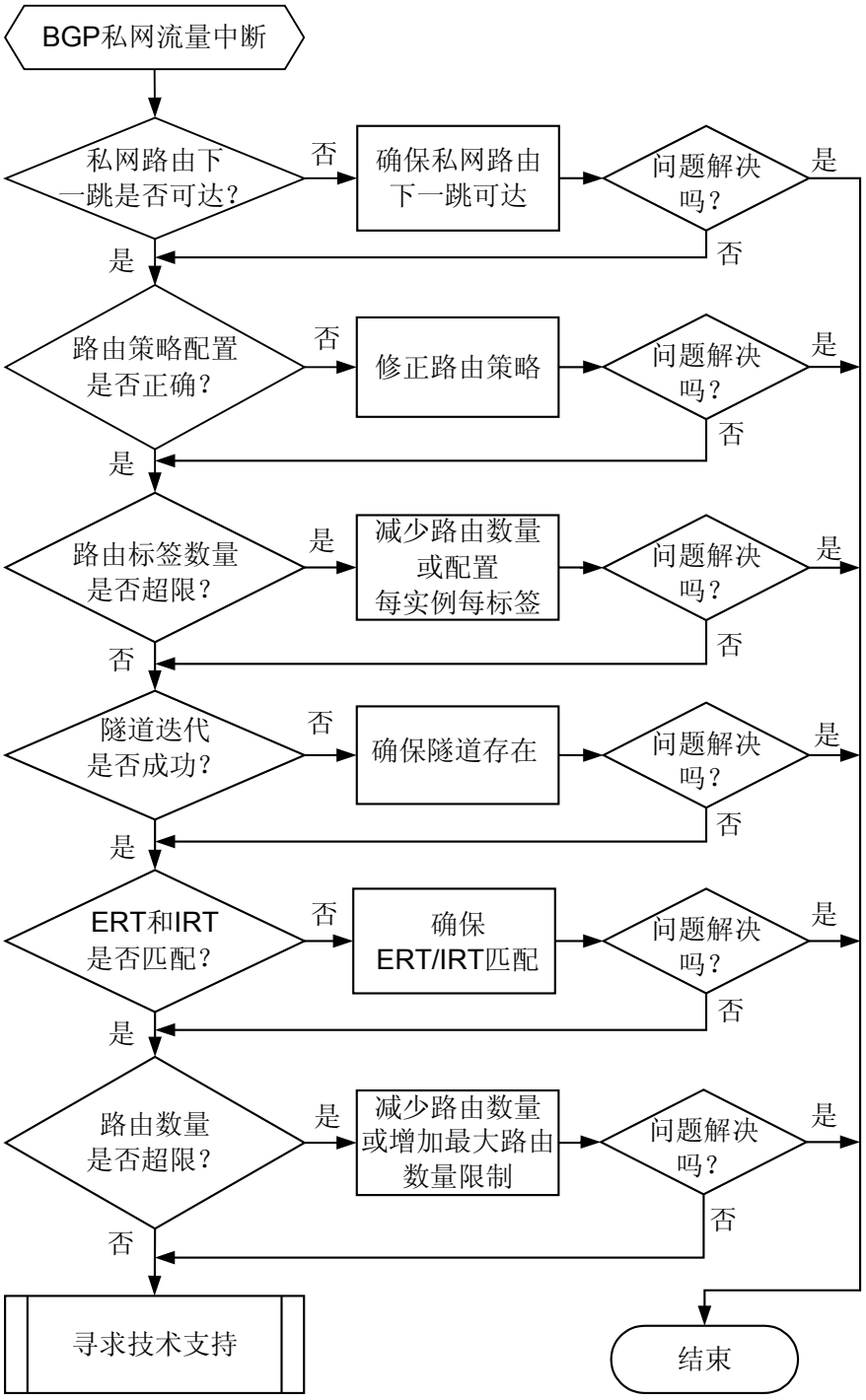
- 路由下一跳不可达导致路由不活跃。
- 路由策略配置不当导致路由无法发布/接收。
- 标签超限导致私网路由无法发布。
- 私网路由迭代不到隧道导致路由不活跃。
- ERT/IRT 不匹配导致路由无法交叉到私网路由表中。
- 路由超限导致收到的路由被丢弃。

5.3.2 故障诊断流程

在配置 BGP 协议后发现 BGP 私网流量中断。

可按照故障诊断流程图 5-3 排除故障。

图 5-3 BGP 私网流量中断故障诊断流程图



5.3.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查路由下一跳是否可达

在路由的发送端(本端 PE)执行 **display bgp vpnv4 vpn-instance** *vpn-instance-name* **routing-table** *ipv4-address* [*mask* | *mask-length*]命令查看目标路由（*ipv4-address* 表示目标路由前缀），确认路由是否存在。

- 如果路由不存在，请确认 CE 路由是否发布到 PE。
- 如果路由存在，请按照下面示例确认路由是否活跃。

以 1.1.1.1/32 这条路由举例，下面命令显示此路由是活跃的（valid）优选的（best），此路由的 BGP 下一跳为 3.3.3.3（Original nexthop），经过迭代后的下一跳为 20.1.1.2（Relay IP Nexthop）。

```
<HUAWEI> display bgp vpnv4 vpn-instance vpna routing-table 1.1.1.1

 BGP local router ID : 20.1.1.2
 Local AS number : 100
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 1.1.1.1/32:
 From: 20.1.1.1 (1.1.1.1)
 Route Duration: 00h00m03s
 Relay IP Nexthop: 20.1.1.2
 Relay IP Out-Interface: Pos1/0/0
 Original nexthop: 3.3.3.3
 Qos information : 0x0
 AS-path Nil, origin incomplete, MED 0, localpref 100, pref-val 0, valid, internal, best, select,
active, pre 255
 Not advertised to any peer yet
```

- 如果目标路由不活跃，请确认 IP 路由表中是否存在到 BGP 下一跳（Original nexthop）的路由，如果不存在说明 BGP 路由不发布是由于路由下一跳不可达导致，请确认为何没有到 BGP 下一跳（Original nexthop）的路由（一般属于 IGP 或静态路由问题）。
- 如果目标路由活跃，却没有被优选（没有 best），请确认 IP 路由表中是否有其他协议优先级（preference）更高的路由存在。如果有请确认是否需要将此路由引入到 BGP 中或调整提协议优先级。 如果没有请联系华为技术工程师。

  说明

  在 BGP 路由表中同一前缀可能有多跳路由，其中最多只有 1 条路由会被优选（best），并且只有被优选的路由才会被添加到 IP 路由表中并发送给其他邻居。BGP 路由与其他协议路由进行比较时是靠协议优先级（preference）来决定哪个更优的。
- 如果目标路由活跃且被优选，但没有显示发送给路由接收端，请执行**步骤 2**（重点检查路由发送端的出口策略）。

在路由接收端执行 **display bgp vpnv4 all routing-table** *network* { *mask* | *mask-length* }查看是否收到目标路由。

- 如果收到目标路由，请重复执行**步骤 1** 判断路由下一跳是否可达并且是否被优选。
- 如果没有收到目标路由，请执行**步骤 2**（重点检查路由接收端的入口策略）。

步骤 2 检查路由策略是否正确

在路由的发送端/接收端执行 **display current-configuration configuration bgp** 命令查看 BGP 配置，确认是否配置邻居的出口/入口策略。

说明

由于是私网流量中断，只需要关注 BGP-VPNv4 地址族或 BGP-VPN 实例地址族下的邻居。

```
<HUAWEI> display current-configuration configuration bgp
#
bgp 100
 peer 1.1.1.1 as-number 200
 #
 ipv4-family unicast
  undo synchronization
  peer 1.1.1.1 enable
 #
 ipv4-family vpnv4
  policy vpn-target
  peer 1.1.1.1 enable
  peer 1.1.1.1 filter-policy acl-name acl-name import
  peer 1.1.1.1 filter-policy acl-name acl-name export
  peer 1.1.1.1 as-path-filter 1 import
  peer 1.1.1.1 as-path-filter 1 export
  peer 1.1.1.1 ip-prefix prefix-name import
  peer 1.1.1.1 ip-prefix prefix-name export
  peer 1.1.1.1 route-policy policy-name import
  peer 1.1.1.1 route-policy policy-name export
 #
 ipv4-family vpn-instance vpna
  peer 10.1.1.1 as-number 300
  peer 10.1.1.1 filter-policy acl-name acl-name import
  peer 10.1.1.1 filter-policy acl-name acl-name export
  peer 10.1.1.1 as-path-filter 1 import
  peer 10.1.1.1 as-path-filter 1 export
  peer 10.1.1.1 ip-prefix prefix-name import
  peer 10.1.1.1 ip-prefix prefix-name export
  peer 10.1.1.1 route-policy policy-name import
  peer 10.1.1.1 route-policy policy-name export
#
return
```

- 如果两端配置了出口/入口策略，则需要确认这些策略是否会把目标路由过滤掉，导致该路由无法正常收发。路由策略的具体配置请参见《HUAWEI NetEngine80E/40E 配置指南-IP 路由》。
- 如果两端没有配置相应的出口/入口策略，请直接执行**步骤 3**。

步骤 3 检查是否迭代不到隧道导致路由不活跃

在路由的接收端（远端 PE）执行 **display bgp vpnv4 all routing-table** *ipv4-address* [*mask* | *mask-length*]命令查看目标路由，确认 VPNv4 路由是否可以迭代到隧道。

以路由 50.1.1.2/32 为例，显示信息中 Relay Tunnel Out-Interface 和 Relay token 字段不为空表示该路由可以迭代到隧道。

```
<HUAWEI> dis bgp vpnv4 all routing-table 50.1.1.2
BGP local router ID : 2.2.2.2
 Local AS number : 100

 Total routes of Route Distinguisher(1:2): 1
 BGP routing table entry information of 50.1.1.2/32:
 Label information (Received/Applied): 13316/NULL
 From: 1.1.1.1 (1.1.1.1)
 Route Duration: 00h00m08s
 Relay IP Nexthop: 20.1.1.1
 Relay IP Out-Interface: Pos1/0/0

 Relay Tunnel Out-Interface: Pos1/0/0

 Relay token: 0x1002
 Original nexthop: 1.1.1.1
 Qos information : 0x0
 Ext-Community:RT <1 : 1>
 AS-path Nil, origin incomplete, MED 0, localpref 100, pref-val 0, valid, internal, best, select,
pre 255
 Not advertised to any peer yet
```

```
Total routes of vpn-instance vpna: 1
BGP routing table entry information of 50.1.1.2/32:
Label information (Received/Applied): 13316/NULL
From: 1.1.1.1 (1.1.1.1)
Route Duration: 00h00m07s
Relay Tunnel Out-Interface: Pos1/0/0

Relay token: 0x1002
Original nexthop: 1.1.1.1
Qos information : 0x0
Ext-Community:RT <1 : 1>
AS-path Nil, origin incomplete, MED 0, localpref 100, pref-val 0, valid, internal, best, select,
active, pre 255
Not advertised to any peer yet
```

- 如果迭代不到隧道，请执行 **display ip vpn-instance verbose [*vpn-instance-name*]**命令检查 Tunnel Policy 字段。如果没有显示该字段，表示没有为 VPN 实例配置隧道策略，VPN 实例使用的隧道为 LDP LSP。如果 VPN 实例使用 MPLS-TE 隧道需要配置隧道策略。Tunnel Policy 字段值表示 VPN 实例使用隧道策略，可以在隧道策略视图下执行 **display this** 检查隧道策略的配置。

  ```
  [HUAWEI-tunnel-policy-p1] display this
  #
  tunnel-policy p1
   tunnel select-seq cr-lsp load-balance-number 1
  #
  ```

  说明

  如果隧道策略下配置了 **tunnel binding destination** *dest-ip-address* **te** { **tunnel** *interface-number* }，还需要在 Tunnel 接口下使能 **mpls te reserved-for-binding** 命令。

  如果隧道没有 Up，请参考 **LDP LSP Down 的定位思路**或者 **TE Tunnel 状态为 Down 的定位思路**继续定位，使隧道状态 Up。
- 如果迭代到隧道，请直接执行**步骤 4**。

步骤 4 检查是否 ERT/IRT 不匹配导致路由无法交叉到私网路由表中

在路由的发送端（本端 PE）/接收端（远端 PE）执行 **display current-configuration configuration vpn-instance** 命令查看是否本端 VPN 实例的 ERT 与远端 VPN 实例的 IRT 不匹配，导致路由发送到远端 PE 后无法交叉到远端 VPN 实例中。

export-extcommunity 表示 ERT， import-extcommunity 表示 IRT。

```
<HUAWEI> display current-configuration configuration vpn-instance
#
ip vpn-instance vpna
 route-distinguisher 1:1
 apply-label per-instance
 vpn-target 1:1 export-extcommunity
 vpn-target 1:1 import-extcommunity
ip vpn-instance vpnb
 route-distinguisher 1:2
 vpn-target 1:1 export-extcommunity
 vpn-target 1:1 import-extcommunity
#
return
```

- 如果 ERT 和 IRT 不匹配，请在 VPN 实例下配置匹配的 vpn-target。
- 如果 ERT 和 IRT 匹配，请执行**步骤 5**。

步骤 5 检查是否标签超限

首先在路由发送端（本端 PE）确认是否使能了 mpls。然后，使用 **display bgp vpnv4 all routing-table** *ipv4-address* [*mask* | *mask-length*]查看目标路由，确定该目标路由是否分到私网标签。

如果显示信息中没有 Label information 字段，则可能是标签资源不足，导致无法为该路由申请到标签而不会给其它对等体。

```
<HUAWEI> display bgp vpnv4 all routing-table 100.1.1.1

 BGP local router ID : 10.1.1.2
 Local AS number : 100

 Total routes of Route Distinguisher(1:1): 1
 BGP routing table entry information of 100.1.1.0/24:
 Imported route.
 Label information (Received/Applied): NULL/13312


 From: 0.0.0.0 (0.0.0.0)
 Route Duration: 00h21m24s
 Direct Out-interface: NULL0
 Original nexthop: 0.0.0.0
 Qos information : 0x0
 Ext-Community:RT <1 : 1>
 AS-path Nil, origin incomplete, MED 0, pref-val 0, valid, local, best, select, pre 255
 Advertised to such 1 peers:
    1.1.1.1

 Total routes of vpn-instance vpna: 1
 BGP routing table entry information of 100.1.1.0/24:
 Imported route.
 From: 0.0.0.0 (0.0.0.0)
 Route Duration: 00h21m24s
 Direct Out-interface: NULL0
 Original nexthop: 0.0.0.0
 Qos information : 0x0
 AS-path Nil, origin incomplete, MED 0, pref-val 0, valid, local, best, select, pre 60
 Not advertised to any peer yet
```

- 如果是标签不足，可在 VPN 实例视图下通过命令 **apply-label per-instance** 配置每实例每标签，来减少标签的使用量。也可以通过路由聚合来减少路由数量。
- 如果标签没有超限，请执行**步骤 6**。

步骤 6 检查路由是否超限

在路由接收端执行 **display current-configuration configuration bgp | include peer** *destination-address* 和 **display current-configuration configuration bgp | include peer** *group-name*（如果 Peer 被加入到对等体组中）命令查看 BGP 配置，确认是否配置邻居路由限制。

例如，限制只能从邻居 1.1.1.1 收 5 条路由，超限之后将丢弃路由并记录日志。

```
<HUAWEI> display current-configuration configuration bgp | include peer 1.1.1.1
 peer 1.1.1.1 as-number 100
 peer 1.1.1.1 route-limit 5 alert-only
  peer 1.1.1.1 enable
```

如果 BGP 邻居被加入到组中，显示信息中有可能没有 route-limit 的配置。

```
<HUAWEI> display current-configuration configuration bgp | include peer 1.1.1.1
 peer 1.1.1.1 as-number 100
 peer 1.1.1.1 group IBGP
  peer 1.1.1.1 enable
  peer 1.1.1.1 group IBGP
```

这种情况下，需要使用 **display current-configuration configuration bgp | include peer** *group-name* 来查看该对等体组的配置。

```
<HUAWEI> display current-configuration configuration bgp | include peer IBGP
 peer IBGP route-limit 5 alert-only
  peer IBGP enable
```

如果流量中断时，产生了路由超限日志 **BGP/3/ROUTPRIX_EXCEED**，表示路由超限导致目标路由被丢弃，则需要扩大本端的路由限制数值。

说明

修改 BGP 邻居限制的最大路由数量时会中断邻居，建议在路由发送端通过路由聚合以减少路由数量来解决。

步骤 7 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

5.3.4 相关告警与日志

相关告警

BGP_1.3.6.1.4.1.2011.5.25.177.1.3.1 hwBgpPeerRouteNumThresholdExceed

相关日志

BGP/3/ROUTPRIX_EXCEED

5.4 BGP ORF 本端（路由发送者）无法收到对端（路由接收者）的 ORF 信息故障的定位思路

介绍了 BGP ORF 本端无法收到对端的 ORF 信息故障的定位思路和案例。

5.4.1 常见原因

本类故障的常见原因主要包括：

- BGP IPv4 单播邻居建立不成功；
- BGP ORF 能力没有协商成功；
- 对端（路由接收者）没有配置 peer 上的 ip-prefix 入口策略；
- 对端（路由接收者）没有配置 peer 上的 ip-prefix 入口策略对应的前缀列表。

5.4.2 故障诊断流程

在配置 BGP ORF 特性后发现本端无法收到对端的 ORF 信息，通过命令 **display bgp peer** *ipv4-address* **orf ip-prefix** 查看，没有显示任何 ip-prefix 信息。

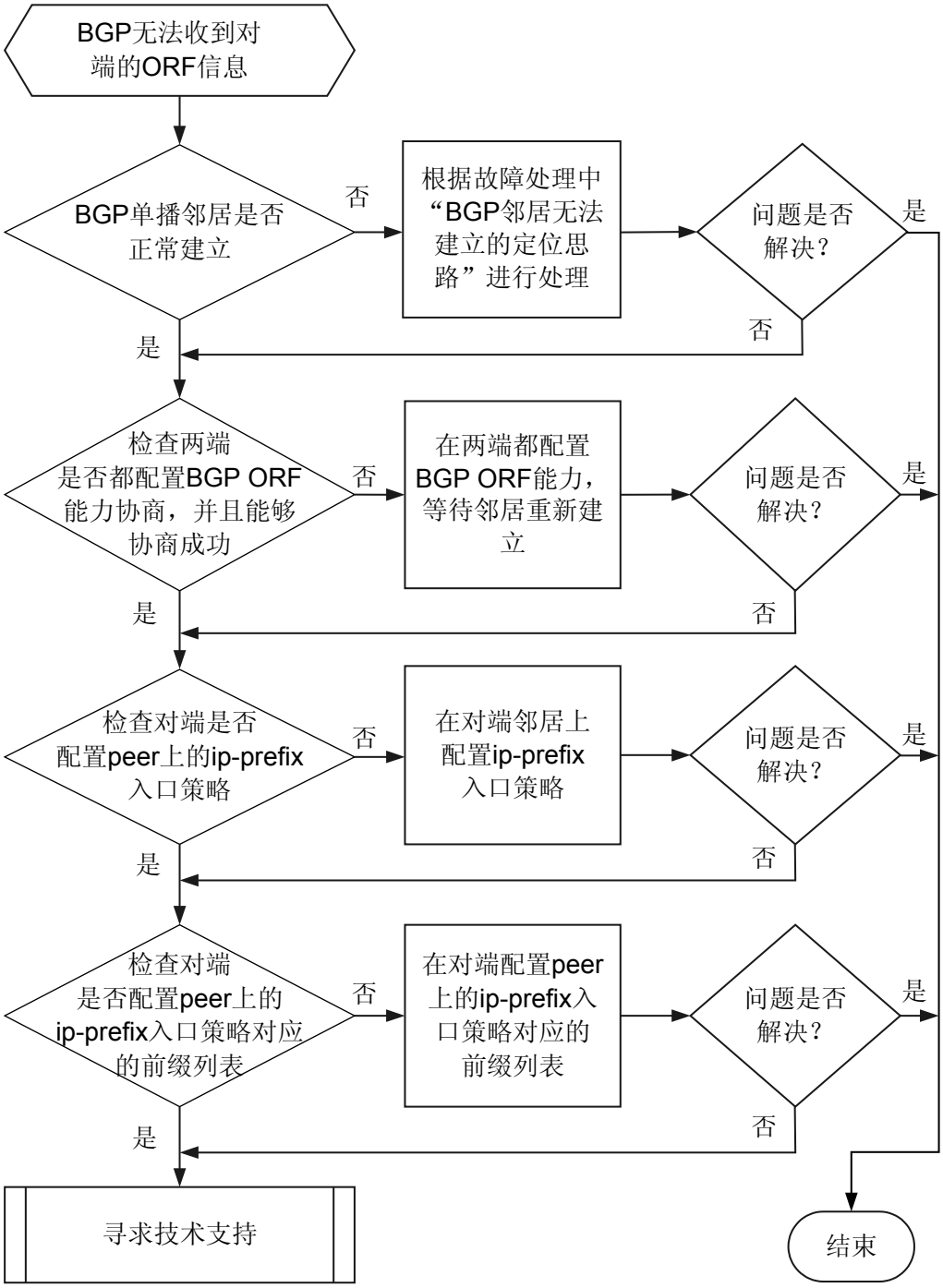
故障定位思路：

- 检查 BGP 单播邻居是否正常建立。
- 检查 BGP ORF 能力是否协商成功。
- 检查对端（路由接收者）是否配置 peer 上的 ip-prefix 入口策略。

- 检查对端（路由接收者）是否配置 peer 上的 ip-prefix 入口策略对应的前缀列表。

可按照故障诊断流程图 5-4 排除故障。

图 5-4 BGP ORF 无法收到对端 ORF 信息故障诊断流程图



5.4.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查 BGP 单播邻居是否正常建立

执行 **display bgp peer** 命令，查看 BGP 对等体状态是否是 Established 状态。

- 如果不是 Established 状态，请参考故障处理中“**BGP 邻居无法建立的定位思路**”进行处理。
- 如果 BGP 对等体状态已经是 Established 状态，请执行**步骤 2**。

步骤 2 检查 BGP 对等体两端是否都配置了 BGP ORF 能力，并且能够协商成功

在两端执行 **display current-configuration configuration** *bgp* 命令，检查显示信息中 IPv4 单播地址族中是否都包含 peer *ipv4-address* capability-advertise orf ip-prefix 配置。

```
<HUAWEI> display current-configuration configuration bgp
#
bgp 100
peer 7.1.1.1 as-number 100
#
 ipv4-family unicast
  undo synchronization
  peer 7.1.1.1 ip-prefix in import
  peer 7.1.1.1 capability-advertise orf ip-prefix both
#
```

说明

BGP ORF 有三种模式：send，receive，both。send 模式说明本端可以发送 ORF 信息；receive 模式说明本端可以接收 ORF 信息；both 模式说明本端既可以发送也可以接收 ORF 信息。如果要让本地能够接收 ORF ip-prefix 前缀信息，则本地必须配置 both 或者 receive，对端必须配置 both 或者 send。

- 如果任何一端没有配置 BGP ORF 能力，则进入 BGP IPv4 单播地址族视图，执行 **peer** *ipv4-address* **capability-advertise orf ip-prefix** 命令配置 BGP ORF 能力。在本端配置时需要在该配置命令中增加命令关键字 **both** 或者 **receive**，在对端配置时需要增加命令关键字 **both** 或者 **send**。

  ```
  <HUAWEI> system-view
  [HUAWEI] bgp 100
  [HUAWEI-bgp] ipv4-family unicast
  [HUAWEI-bgp-af-ipv4] peer 7.1.1.1 capability-advertise orf ip-prefix both
  ```

  确认两端都配置了 BGP ORF 能力，等待对等体关系重新建立之后，通过命令 **display bgp peer** *ipv4-address* **verbose** 查看 BGP ORF 能力是否协商成功，从该命令中能看到接收到的对端的 ORF 能力和本地配置的 ORF 能力。

  ```
  <HUAWEI> display bgp peer 7.1.1.1 verbose | include Address-Prefix
           Support Address-Prefix: IPv4-UNC address-family, rfc-compatible, both
  Enable Address-Prefix: IPv4-UNC address-family, rfc-compatible, both
  ```

  说明

  上面的显示信息中前面部分是接收到的对端的 ORF 能力，后面部分是本地配置的 ORF 能力。由于其他厂商设备的 ORF 能力码与 RFC 规定的能力码可能不同，因此为了与其他厂商设备互通，增加了新的兼容命令，需要确认两端使能了相同的模式（都是 cisco-compatible 模式或者都是 rfc-compatible 模式）。

- 如果两端都配置了 BGP ORF 能力，并且 BGP ORF 能力协商成功，请执行**步骤 3**。

步骤 3 检查对端（路由接收者）是否配置 ip-prefix 入口策略。

在对端执行 **display current-configuration configuration** *bgp* 命令，检查显示信息中 IPv4 单播地址族中是否有 peer *ipv4-address* ip-prefix *ip-prefix-name* import 配置。

```
<HUAWEI> display current-configuration configuration bgp
#
bgp 100
peer 7.1.1.1 as-number 100
#
 ipv4-family unicast
  undo synchronization
  peer 7.1.1.1 ip-prefix in import
  peer 7.1.1.1 capability-advertise orf ip-prefix both
#
```

- 如果对端没有配置 peer 上的 ip-prefix 入口策略，则进入 BGP IPv4 单播地址族视图，执行 **peer** *ipv4-address* **ip-prefix** *ip-prefix-name* **import** 命令，在对端配置 peer 上的 ip-prefix 入口策略，下面是配置前缀列表 **in** 的示例。

  ```
  <HUAWEI> system-view
  [HUAWEI] bgp 100
  [HUAWEI-bgp] ipv4-family unicast
  [HUAWEI-bgp-af-ipv4] peer 7.1.1.1 ip-prefix in import
  ```

- 如果对端已经配置 peer 上的 ip-prefix 入口策略，但是本端仍然不能收到对端的 ORF 前缀信息，请执行步骤 4。

步骤 4 检查对端（路由接收者）是否配置 peer 上的 ip-prefix 入口策略对应的前缀列表

在对端执行 **display ip ip-prefix** *ip-prefix-name* 命令，检查对应 BGP 邻居 ip-prefix 入口策略的前缀列表是否配置。

```
<HUAWEI> display ip ip-prefix in
Info: The specified filter list does not exist.
```

出现上述提示信息，表示前缀列表 in 没有配置成功。

请进入系统视图，执行 **ip ip-prefix** *ip-prefix-name* **index** *index-number* **permit** *ipv4-address mask-length* 命令配置前缀列表。

```
<HUAWEI> system-view
[HUAWEI] ip ip-prefix in index 10 permit 10.1.1.0 24
```

配置完成后，在对端执行 **display ip ip-prefix** *ip-prefix-name* 命令，检查对应 BGP 邻居 ip-prefix 入口策略的前缀列表是否配置成功。

```
<HUAWEI> display ip ip-prefix in
Prefix-list in
Permitted 0
Denied 0
        index: 10                 permit  10.1.1.0/24
```

出现上述信息，表示前缀列表 in 已经配置成功。上述操作之后，如果本端仍无法收到对端的 ORF 信息，请执行步骤 5。

步骤 5 请收集如下信息，并联系华为技术支持工程师

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

5.4.4 相关告警与日志

相关告警

BGP_1.3.6.1.2.1.15.7.2 bgpBackwardTransition

BGP_1.3.6.1.2.1.15.7.1 bgpEstablished

相关日志

BGP/3/STATE_CHG_UPDOWN

BGP/3/WRONG_ROUTERID

BGP/3/WRONG_AS

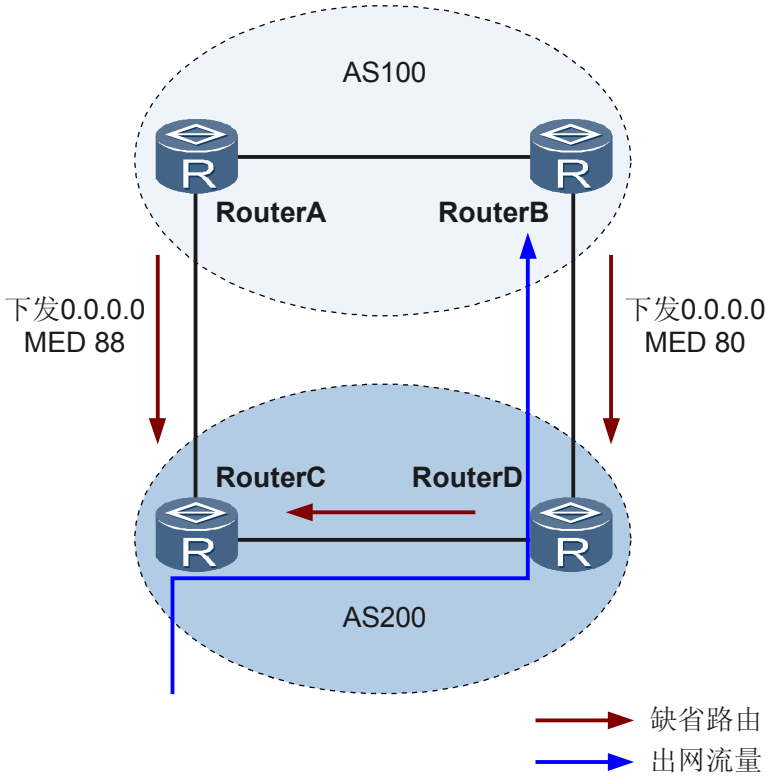
5.5 相关案例

5.5.1 BGP 下发缺省路由的 MED 值不同，导致对端 AS 出口设备间流量穿越

网络环境

在图 5-5 的网络中，RouterA 和 RouterB 为骨干网设备。AS100 和 AS200 间配置了 EBGP 对等体。AS 内的设备间配置了 IBGP 对等体。RouterA 和 RouterB 下发缺省路由后，在 RouterC 上查看 BGP 缺省路由的详细信息，发现 AS200 的出网流量全部指向了 RouterD，即 BGP 缺省路由的下一跳是 RouterD。流量穿越了 RouterC。

图 5-5 AS 出口设备间流量穿越组网图



故障分析

在 RouterC 上执行 **display bgp routing-table 0.0.0.0** 命令查看 BGP 缺省路由的详细信息，发现 RouterA 和 RouterB 设置的 MED 值不同，导致 AS200 的出网流量穿越了 RouterC。

操作步骤

步骤 1 在 RouterA 或 RouterB 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **bgp** *as-number*，进入 BGP 视图。

步骤 3 执行命令 **ipv4-family unicast**，进入 BGP-IPv4 单播地址族视图。

步骤 4 执行命令 **default med** *med*，修改 BGP 路由的缺省 MED 值，使 RouterA 和 RouterB 一致。

完成上述操作后，在 RouterC 上执行 **display bgp routing-table 0.0.0.0** 命令查看 BGP 缺省路由的详细信息，AS200 的出网流量通过 RouterC，故障排除。

----结束

案例总结

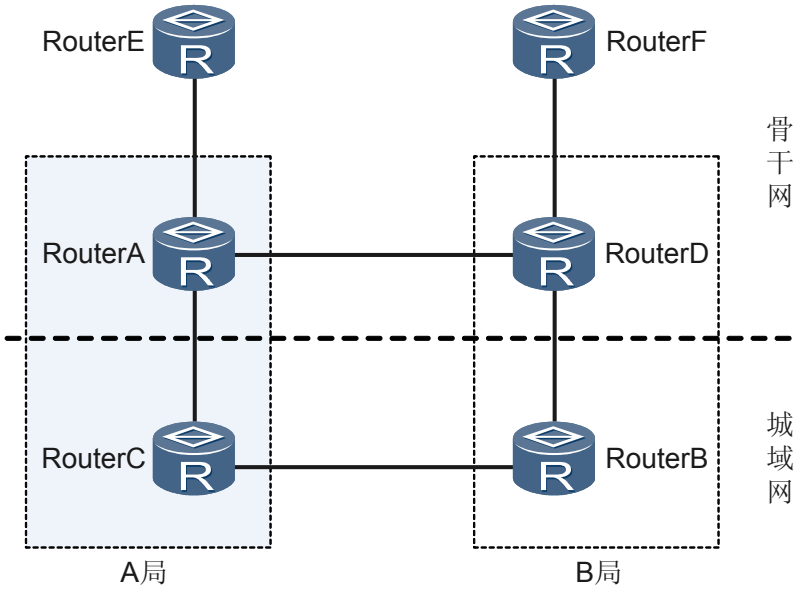
两个 AS 间存在多个出口设备时，需要将其下发缺省路由的 MED 值配置一致。由于 local-preference、MED 等值都一致，BGP 对等体会优选从 EBGP 学来的路由，避免流量穿越。

5.5.2 BGP 路由振荡导致城域网用户无法上网

网络环境

某运营商的骨干网和城域网之间的连接如图 5-6 所示。

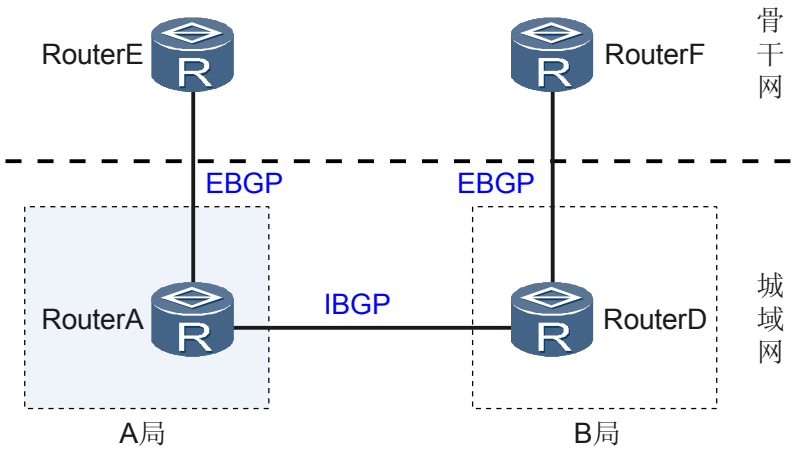
图 5-6 运营商城域网割接前组网图



为适应业务的发展，需对网络进行如下调整（调整后的组网如图 5-7 所示）：

- A 局的 RouterA 从骨干网出口路由器变更为城域网出口路由器，且与 RouterE 相连并建立 EBGP 邻居关系。
- B 局的 RouterD 直接与 F 连接并建立 EBGP 邻居关系。
- RouterA 与 RouterD 之间建立 IBGP 邻居关系。
- 原 RouterC 和 RouterB 下线，不再在现网中使用。

图 5-7 运营商城域网割接后目标组网图



A 局割接成功后准备割接 B 局。此时先升级 RouterD 的版本，在 RouterD 重启后发现大量网吧用户和 PPPoE 用户不能访问 Internet。

故障分析

1. 在城域网内无法 Ping 通该运营商 DNS，通过使用 **tracert** 命令 Tracert 该 DNS，发现流量到达 RouterA 后便中断。
2. 在 RouterA 上使用命令 **display bgp peer** 查看 BGP 对等体状态，状态正常。在 RouterA 上可以 Ping 通该运营商 DNS。
3. 在 RouterA 上重复使用命令 **display ip routing-table statistics** 查看 BGP 路由更新条数来判断是否存在路由环路，发现不存在环路。
4. 在骨干网的 RouterE 上使用命令 **display ip routing-table** 查看路由，发现没有到该城域网的 BGP 路由。
5. 在 RouterE 上使用命令 **display bgp peer** 查看 BGP 对等体状态（“State”字段），发现和 RouterA 的 EBGP 邻居状态不是“Established”。
6. 在 RouterE 上查看日志，发现由于 BGP 路由频繁更新，造成去往 RouterA 的 EBGP 路由被抑制，导致城域网流量无法出网。

   由于 BGP 路由频繁更新造成 BGP 路由被抑制，其中路由频繁更新的原因是：

   (1) RouterA 通过 **network** 方式向 RouterE 发布 BGP 路由，并且是逐条进行更新，而这些更新都被 RouterE 记录。

(2) RouterD 在版本升级过程中进行重启，造成 BGP 路由又一次更新，而此时的路由更新次数恰好达到了 RouterE 设置的 BGP 路由抑制阈值。

操作步骤

步骤 1 在 RouterE 上执行命令 **system-view**，进入系统视图。

步骤 2 在 RouterE 上执行命令 **bgp** *as-number*，进入 BGP 视图。

步骤 3 在 RouterE 上执行命令 **undo dampening**，取消 BGP 路由抑制。

步骤 4 在 RouterE 上执行命令 **display bgp peer**，发现和 RouterA 的 EBGP 邻居关系正常（“State”字段为“Established”）。此时，BGP 路由重新收敛，故障排除。

步骤 5 在 RouterE 上执行命令 **dampening** [*half-life-reach reuse suppress ceiling* | **route-policy** *route-policy-name*] *，恢复原来的 BGP 路由抑制配置。

----结束

案例总结

路由不稳定的主要表现形式是路由振荡，即路由表中的某条路由反复消失和重现。一般情况下，BGP 都应用于复杂的网络环境中，路由变化比较频繁，而频繁的路由振荡会消耗大量的带宽资源和 CPU 资源，严重时会影响到网络的正常工作。通过 BGP 路由振荡抑制，可防止持续路由振荡带来的不利影响。

在网络调整过程中，如果涉及有大量的 BGP 路由更新时，建议先取消原来的 BGP 路由抑制，以免由于路由的频繁振荡导致路由抑制而影响业务运行，待网络调整结束后再恢复原来的路由抑制。

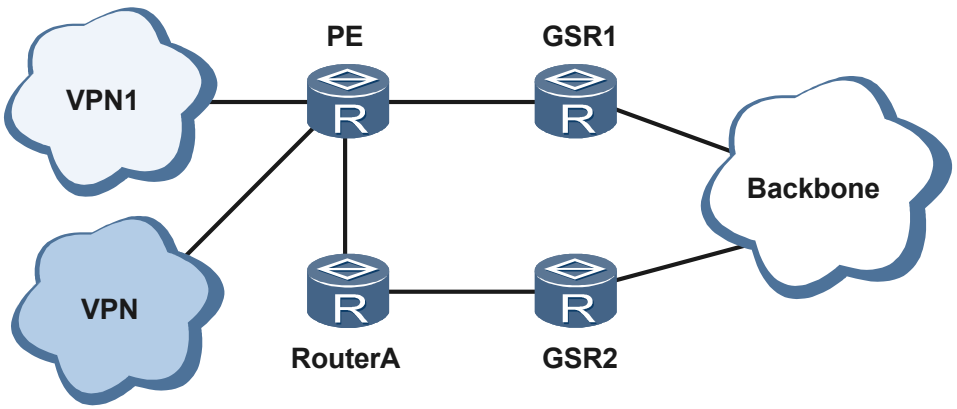
抑制 BGP 路由振荡有三种方法：路由聚合、路由衰减、设置最小路由更新时间间隔。

5.5.3 存在多条同名无效路由策略导致 PE 下发路由策略失效

网络环境

在图 5-8 所示的网络中，PE 接入 VPN1，上行到 GSR1，需要将 PE 的 VPN1 路由发布到骨干网。PE 通过路由策略控制发布到 GSR1 的路由，即在 GSR1 上只需要学习到 PE 发布的 VPN1 的汇总路由，无需学习到明细路由。配置完成后发现 GSR1 不仅学习到了 VPN1 的汇总路由，还学习到了明细路由。

图 5-8 PE 下发路由策略组网图



故障分析

1. 在 PE 上执行命令 **display current-configuration** 检查路由策略相关配置，没有发现异常。
2. 根据故障现象初步判断，可能是 VPN1 下发的路由策略没有生效，造成 GSR1 学习到 PE 发布的 VPN1 的明细路由。
3. 在 PE 上执行命令 **display bgp vpnv4 vpn-instance** *vpn-instance-name* **routing-table peer** *peer-address* { **advertised-routes** | **received-routes** [**active**] }，查看在 PE 上接入的其它 VPN 路由。在 GSR1 上学到的路由均为这些 VPN 的汇总路由，由此可以确定是 VPN1 的路由策略发布出了问题。
4. 经过进一步检查 PE 的配置文件，发现 PE 在 VPN1 实例下发布路由时引用的路由策略有冗余，即下发了同名的三条路由策略，其中第一条路由策略引用的 ip-prefix NGN-A 被定义了，引用有效，其它两条路由策略分别引用的 ip-prefix NGN-A1 和 ip-prefix NGN-A2 未被定义，引用无效。即：

   ```
   ipv4-family vpn-instance CDMA-NGN
   peer 10.247.0.1 route-policy PE_NGN_OUT_MASTER export
   route-policy PE_NGN_OUT_MASTER permit node 10
    if-match ip-prefix NGN-A
   route-policy PE_NGN_OUT_MASTER permit node 20
    if-match ip-prefix NGN-A1
   route-policy PE_NGN_OUT_MASTER permit node 30
    if-match ip-prefix NGN-A2
   ip ip-prefix NGN-A index 10 permit 10.247.0.0 21
   ```

   根据引用路由策略原则，这三条同名的路由策略之间互为或的关系，即只要有一条路由策略引用有效即可，但判断 VPN1 下发路由策略失效可能是多余的无效路由策略导致的。
5. 在 PE 上删除无用的后两条路由策略后，发现 GSR1 学习到的路由只有一条汇总路由，即 ip-prefix NGN-A 中的路由。问题解决。

操作步骤

步骤 1 在 PE 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **undo route-policy** *route-policy-name* [**node** *node*]，删除多余的两条路由策略。

步骤 3 执行命令 **display bgp vpnv4 vpn-instance** *vpn-instance-name* **routing-table peer** *peer-address* **advertised-routes**，查看 PE 接入的 VPN1 路由，发现只有一条汇总路由。故障排除。

----结束

案例总结

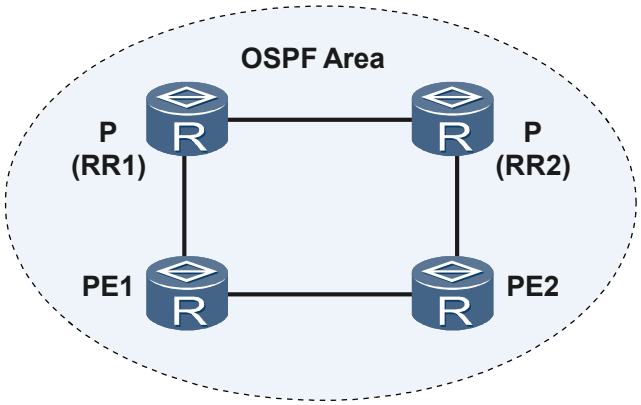
根据引用路由策略的规则，几条同名的路由策略之间互为或的关系，需要将多余的无效路由策略删除，并尽可能简化路由策略，以避免产生问题。

5.5.4 IGP 路由路径错误导致 PE 无法生成公网 LSP

网络环境

在图 5-9 所示的网络中，PE1 可以收到 RR1 反射的 VPNv4 路由，但是路由无法写入 VPN 实例路由表。即在 PE1 的 BGP VPNv4 路由表中可以查看到相关路由信息，但不能在 IPv4 VPN 实例路由表中查看到。

图 5-9 IGP 路由路径错误导致 PE 无法生成公网 LSP 组网图



故障分析

1. 在 PE1 上执行命令 **display bgp vpnv4 all routing-table** 查看 BGP 路由表，看到 BGP 路由表可以学到对端 PE 的路由，说明 BGP 邻居关系正常，通过路由表进一步确认私网标签分发正常。
2. 在 PE1 上执行命令 **display ip routing-table vpn-instance** *vpn-instance-name ip-address* **verbose** 查看私网路由的详细信息，发现 Interface 字段为 NULL0，说明私网路由没有正确的公网迭代出口，即私网路由无效，所以不会被写入 VPN 实例的路由表。
3. 在 PE1 上执行命令 **display mpls ldp session** 查看公网 LDP 会话，状态正常，说明两台 P 设备之间的 LDP 会话可以建立。
4. 在 PE1 上执行命令 **display mpls ldp lsp** *destination-address mask-length* 查看公网 LSP 的标签分发情况，发现 In/OutLabel 显示为 Null，Next-Hop 显示为空，说明 PE 和 P 设备之间虽然可以建立 LDP 会话，但是不能分配标签。
5. 在 PE1 上执行命令 **display ip routing-table** *ip-address* 查看 P 设备 Loopback 地址的 IGP 路由信息，可以看到 P 设备（RR1）的 32 位 Loopback 路由信息是从 PE2 学到的，而不是从 P 设备本身学到的，所以虽然可以建立 LDP LSP 会话，但无法触发公网标签分配。而且两台 PE 间没有配置 LDP，如果配置了 LDP，也可以完成公网标签分配以及生成 VPN 实例路由。
6. 执行命令 **display current configuration** 和 **display ip routing-table** *ip-address* 检查组网中设备的 IGP 相关配置以及路由发布信息，发现 RR1 没有在与 PE1 互连的接口上正确启用 OSPF。

操作步骤

步骤 1 在 RR1 与 PE1 互连的接口上正确配置 OSPF，达到更正 IGP 学习路径的目的。更正后路由信息从 PE 与 P 设备互连接口学习到。

步骤 2 在 PE1 上执行命令 **display ip routing-table vpn-instance** *vpn-instance-name ip-address* **verbose** 查看 MPLS 标签分发情况以及私网路由的出口迭代情况，已恢复正常。Interface 字段为正常的正确的公网迭代出口。

步骤 3 在 PE1 上执行命令 **display mpls ldp lsp** *destination-address mask-length* 查看公网 LSP 的标签分发情况，公网标签分配正常。

步骤 4 在 PE1 上执行命令 **display ip routing-table vpn-instance** *vpn-instance-name* 查看 VPN 实例的路由表，可以看到相关联的私网路由。故障排除。

----结束

案例总结

私网路由的写入依赖于公网的 LSP 是否正常。若公网标签的分配以及 LSP 的生成有问题，则需要注意 IGP 路由的学习路径是否可以触发标签分配，检查 IGP 路由是否正常。

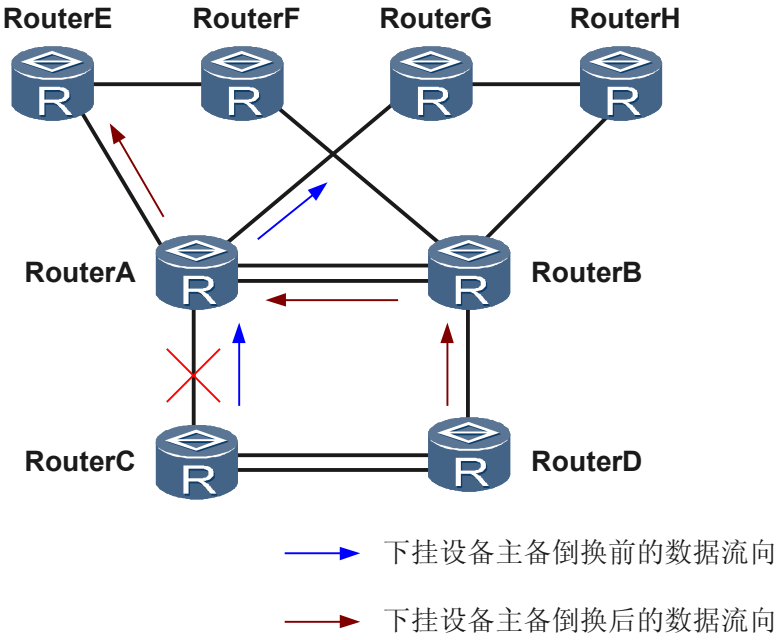
5.5.5 下挂的其他厂商设备主备切换后导致上行流量的链路下一跳改变

网络环境

在图 5-10 所示的网络中，RouterA 和 RouterB 配置了 VRRP，两者之间用心跳线连接，其余均为其他厂商设备。下挂的流量均衡设备 RouterC 和 RouterD 互为冗余备份。RouterA 与 RouterC 相连的接口上配置了路由策略，使上行流量从指定接口转发，下一跳为 RouterG。

下挂的主设备 RouterC 发生故障后 CPU 达到 100%。由于配置了冗余备份，自动主备倒换，流量切换到 RouterD 上，发现 RouterA 上行流量的下一跳为 RouterE，没有按已配置的路由策略指定的下一跳转发。

图 5-10 上行流量的链路下一跳改变组网图



故障分析

其他厂商设备主备倒换之后，上行流量路径为：RouterD->RouterB（Backup）->心跳线->RouterA（Master），因为 RouterA 和 RouterB 之间的心跳线接口上没有配置路由重定向，所以上行流量的链路下一跳就不是路由策略指定的 RouterG，而是 RouterE 了，之前配置的路由策略失效。

故障排除思路：在 RouterA 与 RouterB 的心跳线接口上配置路由策略，使报文沿指定路径转发，路径与主备倒换前相同。即在心跳线接口上配置路由重定向，强制更改下一跳为 RouterG。

操作步骤

步骤 1 在 RouterA 上执行命令 **system-view**，进入系统视图。

步骤 2 定义 ACL 规则。

1. 执行命令 **acl number** *acl-number* 创建一个 ACL 并进入 ACL 视图。
2. 执行命令 **rule** 配置 ACL 规则。
3. 执行命令 **quit** 退出 ACL 视图。

步骤 3 匹配 ACL 规则进行路由重定向。

1. 执行命令 **traffic classifier** *classifier-name* **operator or** 定义一个类并进入流分类视图。
2. 执行命令 **if-match acl** *acl-number* 匹配 ACL 规则。
3. 执行命令 **quit** 退出流分类视图。
4. 执行命令 **traffic behavior** *behavior -name* 定义一个流行为并进入流行为视图。

5. 执行命令 **redirect ip-nexthop** *ip-address* 进行路由重定向，指定下一跳为 RouterG 上与 RouterA 相连的接口地址。
6. 执行命令 **quit** 退出流行为视图。
7. 执行命令 **traffic policy** *policy-name* 定义一个流策略并进入流策略视图。
8. 执行命令 **classifier** *classifier-name* **behavior** *behavior-name* 在流策略中为类制定采用的动作。

步骤 4 在 RouterA 的心跳线接口下应用 traffic-policy。

1. 执行命令 **interface eth-trunk** *trunk-id* 命令进入心跳线接口视图。
2. 执行命令 **traffic-policy** *policy-name* **inbound** 应用 traffic-policy。

步骤 5 执行命令 **display ip routing-table** 查看 RouterA 的上行流量，下一跳为 RouterG。故障排除。

----结束

案例总结

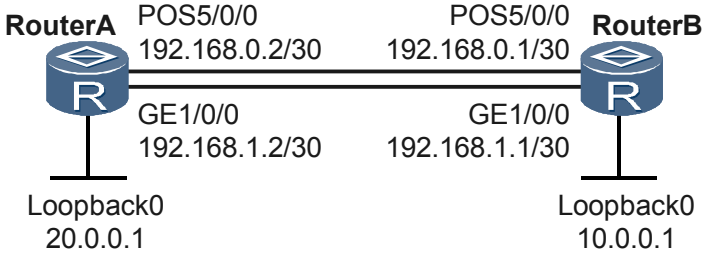
在网络中的设备出现主备倒换之后，需要注意对现网流量的影响。如果上行流量未按照已配置的路由策略沿指定路径转发，可配置路由重定向来指定链路下一跳。

5.5.6 路由迭代导致 BGP 邻居 Down

网络环境

RouterA 通过 POS 接口和 GE 接口双上行到其他厂商设备 RouterB。两端设备通过 Loopback 接口建立 BGP 邻居关系。RouterA 的 POS 接口 Down 掉后，RouterA 和 RouterB 的 BGP 邻居 Down，一直处于 OpenSent 状态。但是从 RouterA 上可以 ping 通对端 RouterB 的 Loopback 地址。

图 5-11 路由迭代导致 BGP 邻居 Down 组网图



故障分析

1. 发现 RouterA 的 POS5/0/0 接口 Down 掉后，在 RouterA 上执行命令 **display ip routing-table** *ip-address* 查看到公网的等价路由，NextHop 为 10.0.0.1 的路由有两条，出接口分别为 GE1/0/0 和 NULL0。而 POS5/0/0 原来没有 Down 时，可以查看到公网的等价路由，NextHop 为 10.0.0.1 的路由出接口分别为 GE1/0/0 和 POS5/0/0。

   在 RouterA 上执行命令 **display bgp peer**，地址为 10.0.0.1 的 BGP 邻居状态为 OpenSent。

2. 等价路由的出接口发生改变，应该是因为发生了路由迭代。如果没有发生路由迭代，POS5/0/0 接口 Down 掉后，原来的两条等价路由上行，应该只有一条出接口为 GE1/0/0 的路由。

3. 检查 RouterA 的配置，分析出接口迭代到 NULL0 的原因。RouterA 上配置了指向 RouterB 的 Loopback 接口地址 10.0.0.1 的 32 位掩码的静态路由。

```
ip route-static 10.0.0.1 255.255.255.255 192.168.1.1
ip route-static 10.0.0.1 255.255.255.255 192.168.0.1
```

RouterA 的 POS5/0/0 接口 Down 掉后，如上的静态路由配置导致 RouterA 进行路由迭代，查找路由表中是否存在到达 192.168.0.1 的路由。通过查看配置文件，发现有如下的静态路由配置：

```
ip route-static 192.168.0.0 255.255.255.0 NULL0 preference 255
```

因此双上行的两条等价路由其中一条下一跳变为 NULL 口。

4. 再分析出接口为 NULL0 和 BGP 邻居 Down 的关系。POS5/0/0 接口 Down 后，RouterA 的双上行路由变为：

```
Destination/Mask    Proto   Pre  Cost      NextHop         Interface
      10.0.0.1/32   BGP     100  0         10.0.0.1        GigabitEthernet1/0/0
                    BGP     100  0         10.0.0.1        NULL0
```

此时从 RouterA 上可以 ping 通 RouterB 的 Loopback 接口地址 10.0.0.1。一般情况下 BGP 邻居不应该 Down。但由于 RouterA 是双路由上行，发包存在 Hash 问题。执行不带源地址的 **ping** 命令，Hash 的结果是出接口为 GE1/0/0，因此可以 ping 通。如果在 RouterA 上执行以 Loopback 地址 20.0.0.1 作为源地址的 **ping** 命令，Hash 结果就是出接口为 POS5/0/0，导致 ping 不通。而 Loopback 地址正是 RouterA 和 RouterB 建立 BGP 邻居的源地址和目的地址，POS5/0/0 现在迭代到的路由下一跳是 NULL0，因此 RouterA 上的 BGP 邻居 Down。

故障排除思路：更改静态路由配置，防止 RouterA 上错误的路由迭代。

操作步骤

步骤 1 在 RouterA 上执行命令 **system-view**，进入系统视图。

在 RouterA 的 POS5/0/0 接口 Down 掉后，迭代到如上静态路由就迭代中止了，宣布出接口为 POS5/0/0 的静态路由不可达，从路由表中删除该路由，到达 RouterB 的所有报文只能有唯一的出口 GE1/0/0。

步骤 2 执行命令 **undo ip route-static 10.0.0.1 255.255.255.255 192.168.1.1** 和 **undo ip route-static 10.0.0.1 255.255.255.255 192.168.0.1**，删除原有的静态路由配置。

步骤 3 执行命令 **ip route-static 10.0.0.1 255.255.255.255 gigabitethernet 1/0/0 192.168.1.1** 和 **ip route-static 10.0.0.1 255.255.255.255 pos 5/0/0 192.168.0.1**，配置静态路由并指定下一跳和对应的出接口。

步骤 4 执行命令 **display bgp peer**，查看到地址为 10.0.0.1 的 BGP 邻居状态为 Established。BGP 邻居正常，故障排除。

----结束

案例总结

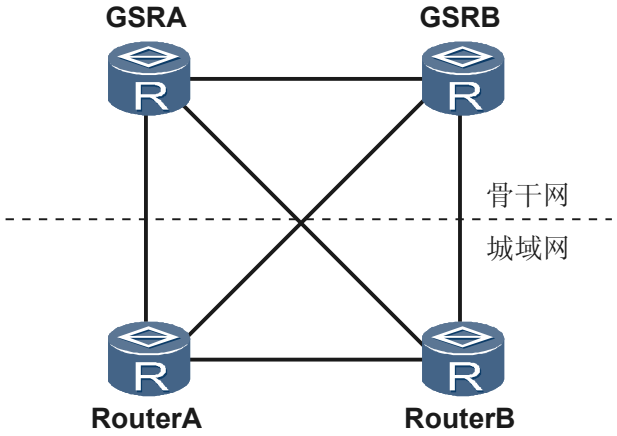
缺省情况下，路由迭代是使能的。在实际网络中，需要分析路由迭代是否会引起不期望的结果。

5.5.7 接收含超长 As_Path 属性的路由导致 BGP 邻居频繁闪断

网络环境

城域网两台出口设备与骨干网两台 GSR 成口字型组网，每台城域网出口设备与连接的骨干设备之间通过 Loopback 地址建立 EBGP 邻居关系。GSR 为其他厂商设备。

图 5-12 BGP 邻居频繁闪断



城域网与骨干设备之间的 EBGP 邻居同时中断后，且频繁 Up/Down，之后故障自动恢复。

故障分析

可能产生故障的原因有：

- 攻击导致 CPU 使用率高，引发 BGP 协议中断。
- 链路问题。
- 城域网出口设备软件问题。
- GSR 设备软件问题。

操作步骤

步骤 1 在城域网出口设备上使用隐藏命令**_display bgp discard peer** *peer-address* 查看 BGP 错误日志，发现大量 BGP 丢包告警。显示如下：

```
2009-Feb-17    00:23:37    Neighbor: 219.158.2.152
ErrInfo: Aspath attribute type error.
Errdata: 50 02 02 04 02 01 12 E5 12 E5 12 E5 0C B9 71 B9 BA FC BA FC BA FC BA FC BA FC BA FC
BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA
FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC
BA FC BA FC
```

步骤 2 分析告警信息发现，城域网出口设备收到了携带有长度超过 255 的 As_Path 属性的路由更新报文。具体分析如下：

```
50 02  //As_Path属性，长度为2个字节
02 04  //As_Path属性长度，因为1个AS号是2个字节，所以总长度是514字节
02 01  //AS-Path，类型为Sequence，个数为1
```

```
12 E5 12 E5 12 E5 0C B9 71 B9 BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA
FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC
BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC BA FC  //全部是AS号
```

因为 BGP 的错误报文只记录 200 字节，所以超出部分没有记录。从这里的 AS 号可以看到，前面 Sequence 类型的 AS 号应该是 1 个，但是后面有很多明显不止一个的 AS 号，该报文显然是一个错误报文。而且从该报文来看，AS 号总长度加上类别是 514 字节，去掉 2 字节，剩下的纯 AS 号的长度应该是 512 字节，也就是 256 个 AS 号，而 BGP 一个 AS 序列最多只能打包 255 个，超过 255 就需要另起一个 AS 序列，所以 256 个 AS 应该打包成 2 个 AS 序列，也就是说应该打包成如下格式：

```
02 01 12 E5  //报文一
02 FF 12 E5 12 E5……  //报文二
```

步骤 3 依据 RFC 定义，对于超过 255 个 As_Path 属性的路由更新报文，需要对 As_Path 属性分不同的 AS 序列来通告。收到不符合定义的更新报文，需要发送 Notification 消息给对端，提示删除 BGP 邻居。经确认本次故障是海外某运营商填充 252 个重复 AS 号导致。

----结束

案例总结

BGP 协议中断后，可以通过_**display bgp discard peer** *peer-address* 命令来查看 BGP 错误日志记录，分析引起中断的原因。

BGP 接收报文时，首先对报文进行解析，只有保证报文的正确性，才会根据路由策略、前缀列表等手段进行路由过滤、属性修改等动作。BGP 的特点是变化才更新。如果某邻居发布路由时打包出现了错误，BGP 只是简单的丢弃报文而不断开连接，之后该邻居将不再把路由重新发布过来。这样就可能导致业务中断，而 BGP 重新建立连接后，对等体重新打包过来的报文就可能正确。

5.5.8 由于迭代深度问题导致静态路由不生效

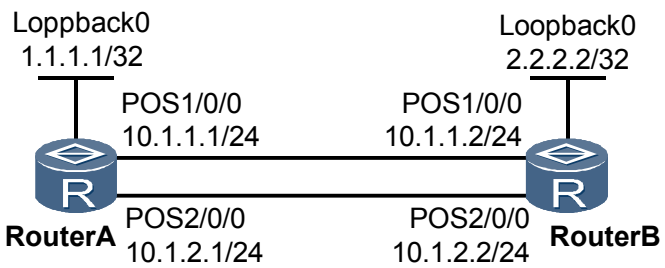
网络环境

在图 5-13 所示的网络中，RouterA 与 RouterB 采用两条 POS 链路相连，并且建立 EBGP 邻居。RouterA 上配置两条静态路由：

```
ip route-static 2.2.2.2 255.255.255.255 pos1/0/0 10.1.1.2
ip route-static 2.2.2.2 255.255.255.255 10.1.2.2
```

查看路由表，去往 RouterB 的路由只有一个下一跳出接口 POS1/0/0。

图 5-13 由于迭代深度问题导致静态路由不生效组网图



故障分析

由于 RouterA 上配置的路由 **ip route-static** “2.2.2.2 255.255.255.255” **pos1/0/0 10.1.1.2** 指定了出接口，不需要迭代，迭代深度为 0；而另一条路由 **ip route-static 2.2.2.2 255.255.255.255 10.1.2.2** 没有指定出接口，需要进行 1 次迭代，迭代深度为 1。

BGP 选择迭代深度最小的静态路由，因此，选中上述第一条迭代深度为 0 的，所以 BGP 路由出接口都为 POS1/0/0。

操作步骤

步骤 1 在 RouterA 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **undo ip route-static 2.2.2.2 255.255.255.255 10.1.2.2**，删除静态路由。

步骤 3 执行命令 **ip route-static 2.2.2.2 255.255.255.255 pos2/0/0 10.1.2.2**，配置静态路由，并指定出接口。

完成上述操作后，BGP 选择迭代深度最小的静态路由，两条静态路由同时命中，所以在 RouterA 上查看路由表，可以看到两个出接口 POS1/0/0 和 POS2/0/0。

----结束

案例总结

配置静态路由时指定出接口，可以避免由于迭代深度不同造成某些静态路由不生效。

5.5.9 未使能 BGP 负载分担导致出流量不均衡

网络环境

在图 5-14 所示的网络中，RouterA、RouterB、RouterC 和 RouterD 为 4 台城域网出口设备，RouterE、RouterF、RouterG 和 RouterH 为 4 台省骨干网设备。4 台城域网出口设备与 4 台省骨干网设备按照图 5-14 所示通过运行 EBGP 协议互连。

因业务需要，对网络进行调整。如图 5-14 所示，4 台城域网出口设备与 4 台省骨干网设备通过运行 EBGP 协议交叉互连。此时，发现流量不均衡，每台城域网出口设备的双出口流量不均衡，大部分流量只通过其中一条链路传输。

图 5-14 割接前组网图

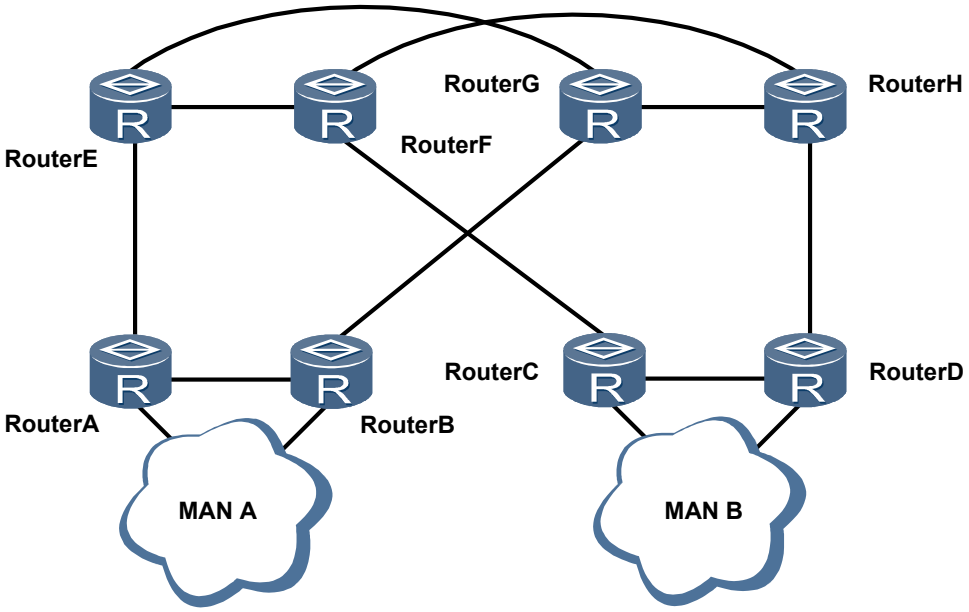
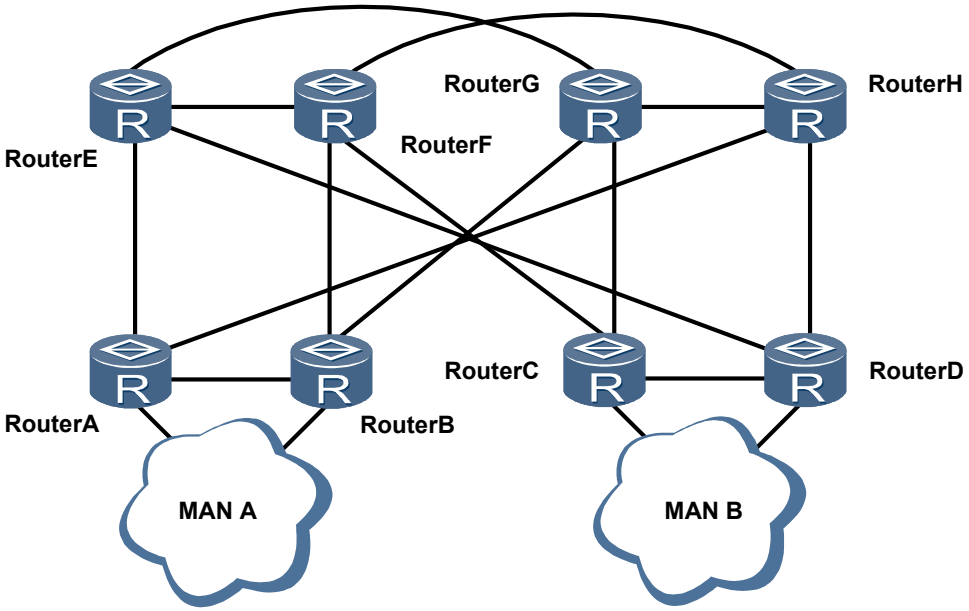


图 5-15 割接后组网图



故障分析

此时，每台城域网出口设备上产生两条等值路由，但是由于没有打开 BGP 负载分担功能，所以大部分流量会通过其中一条链路传输，造成流量不平衡。通过使能 4 台城域出口设备的负载分担功能，可以解决该问题。

操作步骤

步骤 1 在 RouterA 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **bgp** *as-number*，进入 BGP 视图。

步骤 3 执行命令 **maximum load-balancing** *number*，配置最大等价路由条数大于等于 2。

完成上述操作后，在 RouterB、RouterC 和 RouterD 上进行同样的操作。

----结束

案例总结

由于网络的快速发展，各个城域网和骨干网都存在着类似的组网。当属于两个城域网的设备进行互访时，就将产生等价路由。如果没有在城域网设备上使能 BGP 负载分担，根据 BGP 路由优选原则，将会选择 **router id** 最小的一条路由，便会产生本案例中描述的问题。

建议使能城域网出口设备的负载分担功能，配置最大等价路由条数为最大值，以便于后期扩容。当组网中存在其他厂商设备时，需要考虑其他厂商设备支持配置的最大等价路由条数。

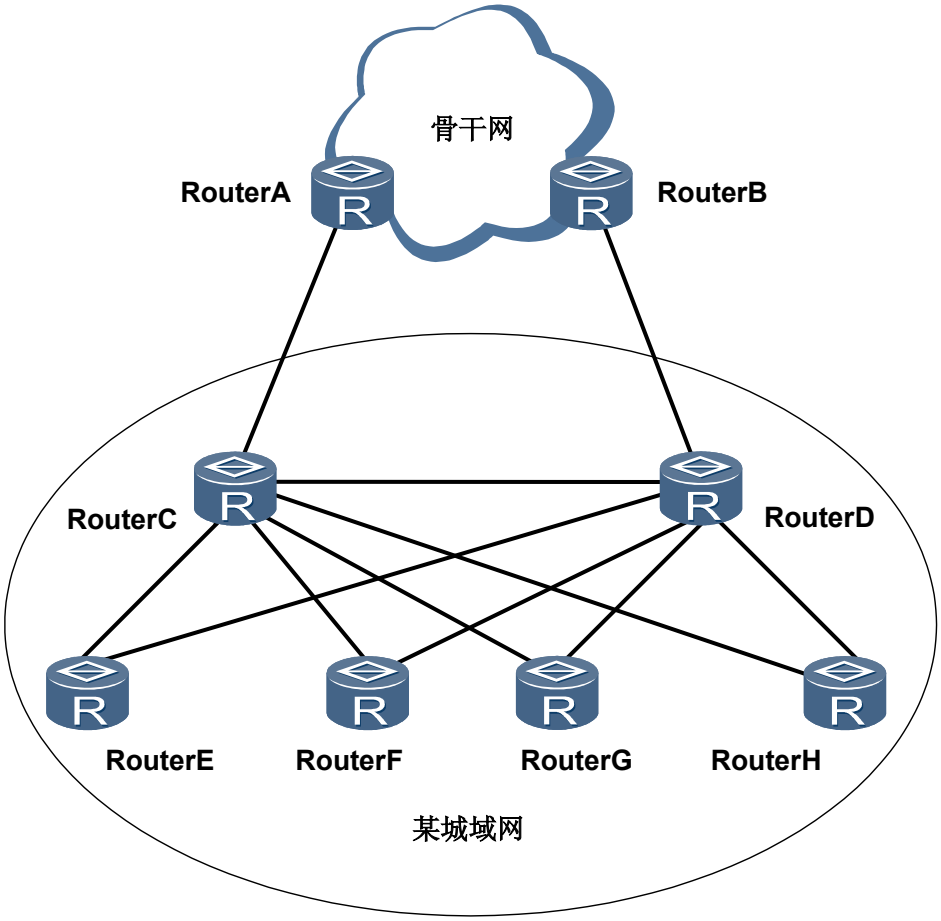
5.5.10 由于不同路由协议优先级规划不合理导致 EBGP 发布的聚合路由频繁震荡

网络环境

在图 5-16 所示的网络中，某城域网 RouterC、RouterD 为某城域网出口设备，通过运行 EBGP 协议与骨干网设备 RouterA、RouterB 互连。骨干网设备对城域网出口设备配置了 EBGP 路由抑制。在城域网中，RouterC、RouterD 与下挂的设备运行 IS-IS 协议互连，并且建立 IBGP 邻居。由于城域网出口设备的互连链路上存在部分穿越流量，为避免链路故障导致穿越流量环路，RouterC 与 RouterD 通过接口地址建立 IBGP 邻居关系，并且使用 **network** 方式和静态路由黑洞向骨干网设备发布城域网内部路由。

此时，由于单板或链路故障，RouterC 与 RouterD 的 IBGP 邻居关系频繁 Up 或 Down，整个城域网业务中断。

图 5-16 EBGP 邻居发布的聚合路由频繁震荡组网图



故障分析

导致路由振荡的条件有：

- 修改了相关的策略，包括本端和对端的策略，主要是人为操作导致。
- 连续两次添加和删除路由（主要是发布的汇总路由）。
- 静态路由和动态协议优先级规划不合理，导致 BGP 发布汇总路由并非完全采用 **network** 和黑洞路由方式。

通过查看设备日志，没有人为操作修改相关策略及删除或添加路由。

城域网出口设备采用 **network** 和黑洞路由方式发布路由，不可能存在路由的添加或删除。通过查看汇总路由发现其生存时间很长，不可能发生中断。

由于在 RouterC 和 RouterD，BGP 协议优先级配置为 20，黑洞静态路由缺省为 60。因此，在静态路由和 IBGP 路由都存在的情况下，发布汇总路由会优先选择 IBGP 路由，这样当 RouterC 与 RouterD 之间的 IBGP 邻居关系振荡时，会导致设备发布的汇总路由振荡，从而造成城域网业务中断。

通过调整 BGP 和静态路由的优先级，使 IBGP 的优先级低于静态路由，可解决该问题。

操作步骤

步骤 1 在 RouterC 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **bgp** *as-number*，进入 BGP 视图。

步骤 3 执行命令 **preference** *preference*，配置 BGP 协议的优先级值大于 60。

完成上述操作后，在 RouterD 上进行同样的操作，修改 BGP 协议的优先级，城域网业务恢复正常。

----结束

案例总结

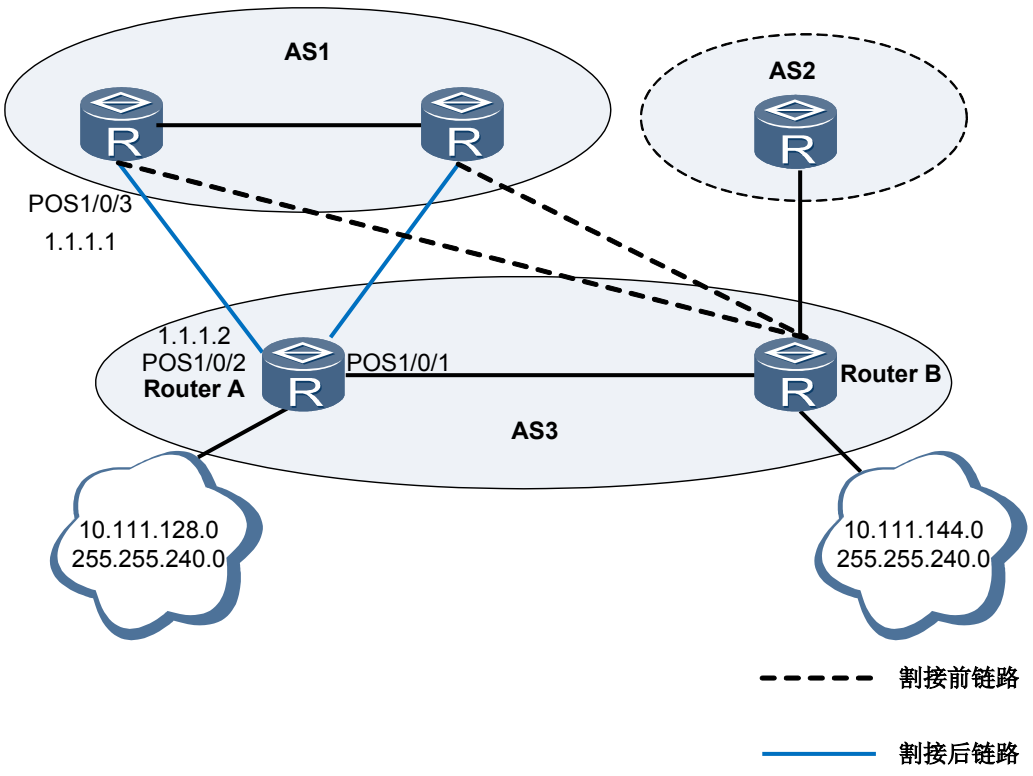
如果采用 **network** 方式和黑洞路由方式发布城域网路由，应该不存在路由振荡的问题，而此案例主要是由于不同路由协议的优先级配置不当，导致城域网设备路由未按计划发布。

5.5.11 对端未配置负载分担导致两条链路的流量不均衡

网络环境

AS1、AS2 和 AS3 分属三个不同的运营商。AS1 通过两条链路与 RouterB 相连,两条链路上的流量比较均衡，大约为 2：1。将两条链路割接至 RouterA 后，从 AS1 流向 AS3 的流量中，接口 POS1/0/2 的流量达到 120M，而 POS1/0/1 上的流量却只有 1 ～ 3M，严重不均衡。

图 5-17 负载分担典型组网图



故障分析

由于 AS1 为其他运营商的设备，判断其设备判断未对通过 AS3 学到的来自 AS2 的路由进行负载分担。从而 AS1 中的路由器只根据最优路径进行转发，造成此故障现象。此时可以通过修改 RouterA 的 MED 值从而影响对端的选路。

操作步骤

步骤 1 在 RouterA 上执行命令 **system-view**，进入系统视图。

步骤 2 执行命令 **ip ip-prefix med-prefix index 10 permit 10.111.128.0 20**，配置名为 med-prefix 的地址前缀列表，只允许 10.111.128.0/20 网段内的路由通过。

步骤 3 执行命令 **route-policy med-500 permit node 10**，创建名为 med-500、节点号为 10 的 Route-Policy，匹配模式为允许。

步骤 4 执行命令 **if-match ip-prefix med-prefix**，设置基于 IP 地址前缀列表 med-prefix 的匹配规则。

步骤 5 执行命令 **apply cost 500**，设置 10.111.128.0/20 网段的路由开销为 500。

步骤 6 执行命令 **quit**，退出 Route-Policy 视图。

步骤 7 执行命令 **route-policy med-500 permit node 15**，创建名为 med-500、节点号为 15 的 Route-Policy，匹配模式为允许。

步骤 8 执行命令 **apply cost 0**，设置其他网段的路由开销为 0。

步骤 9 执行命令 **quit**，退出 Route-Policy 视图。

步骤 10 执行命令 **bgp 3**，进入 BGP 视图。

步骤 11 执行命令 **ipv4-family unicast**，进入 IPv4 单播地址族视图。

步骤 12 执行命令 **peer 1.1.1.1 route-policy med-500 export**，对向对等体发布的路由应用名为 med-500 的 Route-Policy。

步骤 13 执行命令 **display interface pos 1/0/1**，发现 POS1/0/1 的入流量较之前已经大幅增加，说明策略已经生效。如需继续将流量分担至 POS1/0/1，请在已经配置的地址前缀列表里直接增加网段进行调整。

----结束

案例总结

MED（Multi-Exit-Discriminator）属性仅在相邻两个 AS 之间传递，收到此属性的 AS 一方不会再将其通告给任何其他第三方 AS。 MED 属性相当于 IGP 使用的度量值（Metrics），它用于判断流量进入 AS 时的最佳路由。当一个运行 BGP 的 Router 通过不同的 EBGP 对等体得到目的地址相同但下一跳不同的多条路由时，在其它条件相同的情况下，将优先选择 MED 值较小者作为最佳路由。

6 RIP 故障处理

关于本章

6.1 RIP 没有学到部分或全部路由的定位思路

6.1.1 常见原因

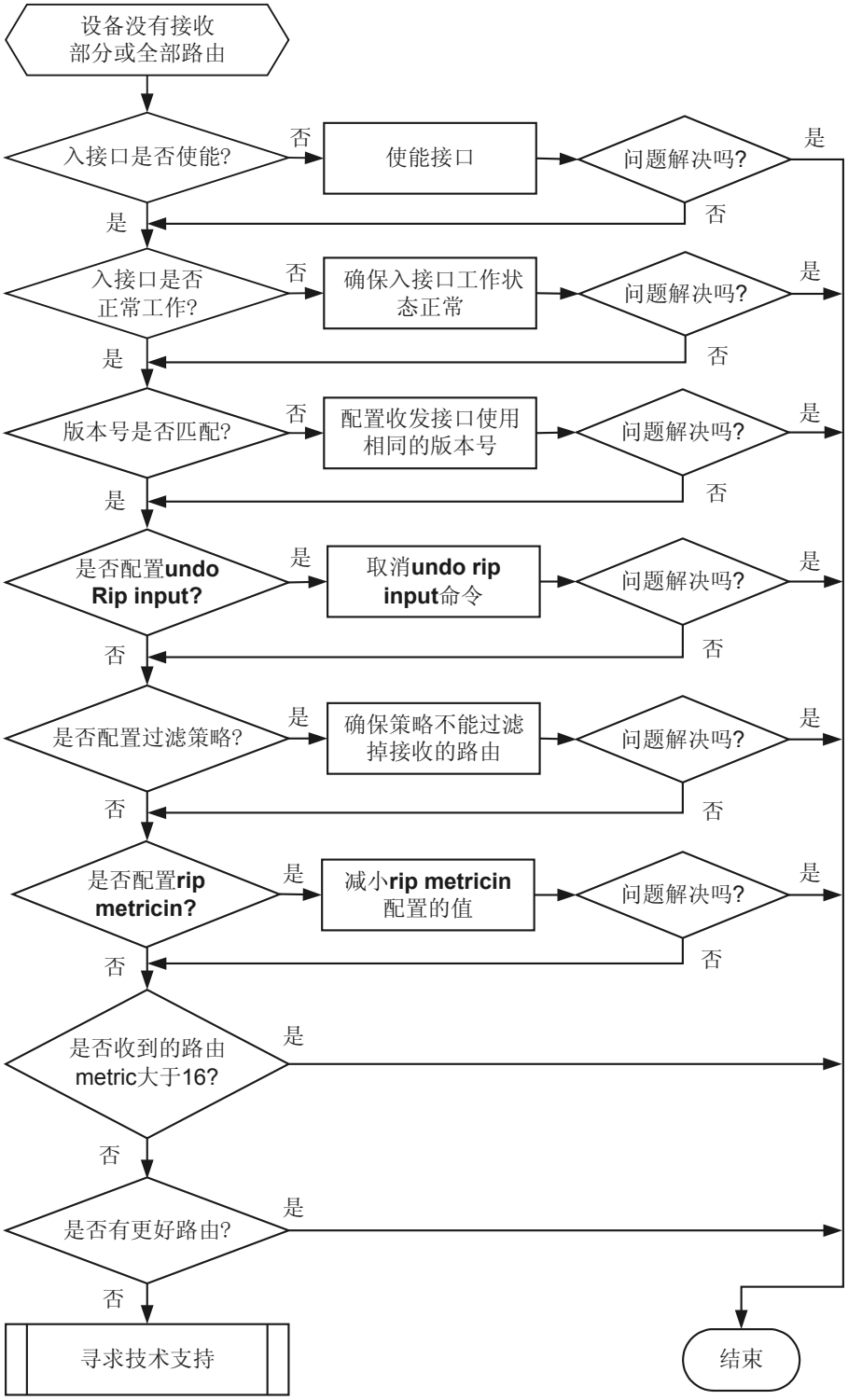
本类故障的常见原因主要包括：

- 接口未使能 RIP
- 接口状态不是 Up
- 对端发送 RIP 协议报文的版本号和本地接口接收的 RIP 协议报文版本号不一致
- 接口上配置了禁止接收 RIP 报文
- 在 RIP 中配置了策略，过滤掉收到的 RIP 路由
- 收到的路由度量值大于 16
- 路由表中存在其它协议学到的相同路由
- 路由超限

6.1.2 故障诊断流程

在配置各路由器后，发现部分或全部路由没有接收，或 **display ip routing-table** 显示信息中没有 RIP 学到的路由。请使用下面的故障诊断流程，如**图 6-1** 所示。

图 6-1 RIP 路由接收故障诊断流程图



6.1.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查入接口是否在 RIP 中使能

network 命令用来使能指定接口网段，只有使能了 RIP 协议的接口才会进行 RIP 路由的接收、发送。使用命令 **display current-configuration configuration rip** 可以看到当前使能 RIP 的网段信息，检查入接口是否在其中。

network 命令使能的网络地址，必须是自然网段的地址。

步骤 2 检查入接口工作是否正常

使用 **display interface** 命令，查看入接口的工作状态：

- 如果接口当前物理状态为 Down 或 Administratively Down，那么 RIP 将无法从这个接口接收到路由。
- 如果接口当前协议状态为 Down，那么 RIP 已经从该接口学到的路由的 cost 值先变为 16，再被清除。

因此，必须确保接口的工作状态正常。

步骤 3 检查对方发送版本号和本地接口接收的版本号是否匹配

缺省情况下，接口只发送 RIP-1 报文，但可以接收 RIP-1 和 RIP-2 报文。当入接口与收到的 RIP 报文使用不同的版本号时，有可能造成 RIP 路由不能被正确的接收。

步骤 4 检查入接口是否配置了 **undo rip input** 命令

rip input 命令用来控制允许指定接口接收 RIP 报文。**undo rip input** 命令用来禁止指定接口接收 RIP 报文。如果在入接口配置了 **undo rip input**，则从这接口上来的 RIP 报文都得不到处理，导致收不到路由。

步骤 5 检查在 RIP 中是否配置了策略，过滤掉收到的 RIP 路由

filter-policy import 命令用来过滤接收的 RIP 路由信息。如果使用 ACL 过滤路由，通过命令 **display current-configuration configuration acl-basic** 可以查看从邻居来的 RIP 路由是否被过滤掉；如果使用 IP 地址前缀列表过滤路由，使用 **display ip ip-prefix** 查看配置策略。

如果被路由策略过滤掉，请正确地配置路由策略。

步骤 6 检查入接口是否配置了 **rip metricin** 命令，使得接收到得路由的度量值大于 16

rip metricin 命令用来设置接口接收 RIP 报文时给路由增加的度量值。如果最终的度量值超过了 16，则认为该路由不可达，从而不会将该路由加到路由表。

步骤 7 检查收到的路由度量值是否大于 16

同上，如果接收到的 RIP 路由的度量值超过 16，则认为该路由不可达，从而不会将该路由加到路由表。

步骤 8 检查在路由表中是否有其它协议学到的相同路由

通过 **display rip** *process-id* **route** 查看是否从邻居接收到了路由。可能的情况是：RIP 路由已经正确的接收了，同时本地还从其它的协议学到了相同的路由，比如 OSPF 或者 IS-IS。这时，OSPF 或 IS-IS 的协议权重一般大于 RIP，路由管理将优先选择通过 OSPF 或 IS-IS 学到的路由。通过命令 **display ip routing-table protocol rip verbose** 应该可以看到该路由，状态应该是非激活的。

步骤 9 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

6.1.4 相关告警与日志

相关告警

无

相关日志

无

6.2 设备没有发送部分或全部 RIP 路由的定位思路

6.2.1 常见原因

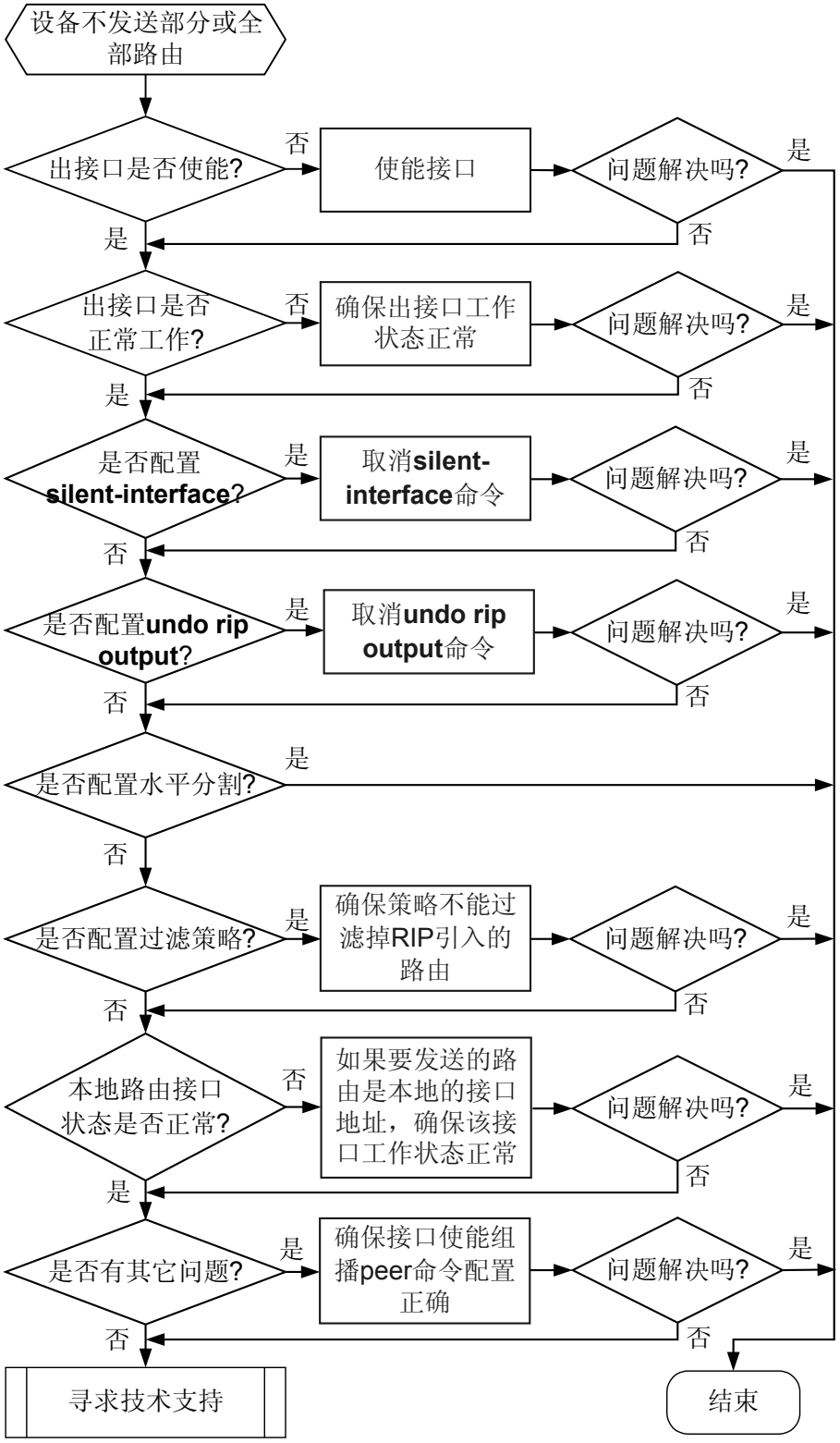
本类故障的常见原因主要包括：

- 接口未使能 RIP
- 接口状态不是 Up
- 接口下配置了 **silent-interface** 命令，被抑制发送 RIP 报文
- 接口下配置了 **undo rip output** 命令，被禁止发送 RIP 报文
- 接口上没有使能水平分割
- RIP 中是否配置了策略，过滤掉引入到 RIP 的路由
- 端口的物理状态是“Down”或“Administratively Down”，或者接口出方向协议的当前状态是“Down”。因此，接口的 IP 地址不能够加到 RIP 的发布路由表中。
- 出接口不支持组播，而要发送的报文是发送到组播地址；或者如果出接口不支持广播，而要发送的报文是发送到广播地址

6.2.2 故障诊断流程

在配置各路由器后发现路由器不发送部分或全部路由。请使用下面的故障诊断流程，如图 6-2 所示。

图 6-2 RIP 路由发送故障诊断流程图



6.2.3 故障处理步骤

背景信息

说明

请保存以下步骤的执行结果，以便在故障无法解决时快速收集和反馈信息。

操作步骤

步骤 1 检查出接口是否在 RIP 中使能

network 命令用来使能指定接口网段，只有使能了 RIP 协议的接口才会进行 RIP 路由的接收、发送。使用命令 **display current-configuration configuration rip** 可以查看当前使能 RIP 的网段信息，检查入接口是否在其中。

network 命令使能的网络地址，必须是自然网段的地址。

步骤 2 检查出接口工作是否正常

使用 **display interface** 命令，查看出接口的工作状态。如果接口当前物理状态为 Down 或 Administratively Down，或者当前协议状态为 Down，那么 RIP 将不能在该接口上正常工作。因此，必须确保接口的工作状态正常。

步骤 3 检查出接口是否配置了 **silent-interface** 命令

silent-interface 命令用来抑制接口使其不发送 RIP 报文。使用命令 **display current-configuration configuration rip** 查看出接口是否被抑制。如果是，则取消对该接口的抑制。

步骤 4 检查出接口是否配置了 **undo rip output** 命令

在出接口上使用命令 **display current-configuration** 查看是否配置了 **rip output**。**rip output** 命令用来允许接口发送 RIP 报文。**undo rip output** 命令用来禁止接口发送 RIP 报文。如果显示出接口配置了 **undo rip output**，则将不能从该接口发送 RIP 报文。

步骤 5 检查出接口是否配置了水平分割命令

在出接口上使用命令 **display current-configuration** 查看是否配置了 **rip split-horizon**。缺省情况下，出接口都使能了水平分割，该命令的显示信息中没有关于水平分割的配置项；但对于 NBMA（NonBroadcast Multiple Access）网络连接的出接口（如 X.25、FR），如果没有显示关于水平分割的配置项，则表明在该接口上没有使能水平分割。

水平分割是指：从一个接口学到的路由，将不能再从该接口对外发布。水平分割机制是用于避免相临邻居间的路由循环。所以不要轻易取消接口的水平分割。

步骤 6 检查在 RIP 中是否配置了策略，过滤掉引入到 RIP 的路由

filter-policy export 命令用来配置全局出口过滤策略，只有通过过滤策略的路由才能被加入 RIP 的通告路由表中，并通过更新报文发布出去。

步骤 7 如果要发送的路由是本地的接口地址，检查该接口的状态

使用 **display interface** 命令，查看接口的工作状态。如果显示接口当前物理状态为 Down 或 Administratively Down，或者出接口的当前协议状态为 Down，则该接口的 IP 地址将不会被加入 RIP 的通告路由表。从而不会发给邻居。

步骤 8 检查是否有其它特殊问题

如果出接口不支持组播，而要发送的报文是发送到组播地址；或者如果出接口不支持广播，而要发送的报文是发送到广播地址，将会出现故障。这时候可以先排除接口的问题，然后在 RIP 模式下配置 **peer** 命令，使用单播地址进行发送，可以避免此故障发生。

步骤 9 如果故障仍未排除，请收集如下信息，并联系华为技术支持工程师。

- 上述步骤的执行结果。
- 设备的配置文件、日志信息、告警信息。

----结束

6.2.4 相关告警与日志

相关告警

无

相关日志

无