

# 边界网关协议基本问题故障排除

## 目录

[简介](#)

[先决条件](#)

[要求](#)

[使用的组件](#)

[背景信息](#)

[拓扑](#)

[场景和问题](#)

[邻接关系断开](#)

[无连接](#)

[配置问题](#)

[TCPS会话问题](#)

[邻接退回](#)

[接口摆动](#)

[保持计时器已过期](#)

[AFI/SAFI suses](#)

[路径安装和选择](#)

[下一跳](#)

[RIB故障](#)

[竞争条件](#)

[其他问题](#)

[BGP慢速对等体](#)

[内存问题](#)

[高 CPU 利用率](#)

[相关信息](#)

## 简介

本文档介绍如何解决边界网关协议(BGP)最常见的问题，并提供基本解决方案和指南。

## 先决条件

### 要求

本文档没有任何特定的前提条件。基本BGP协议知识非常有用，有关详细信息，请参阅[BGP配置指南](#)。

### 使用的组件

本文档不限于特定的软件和硬件版本，但命令适用于Cisco IOS®和Cisco IOS® XE。

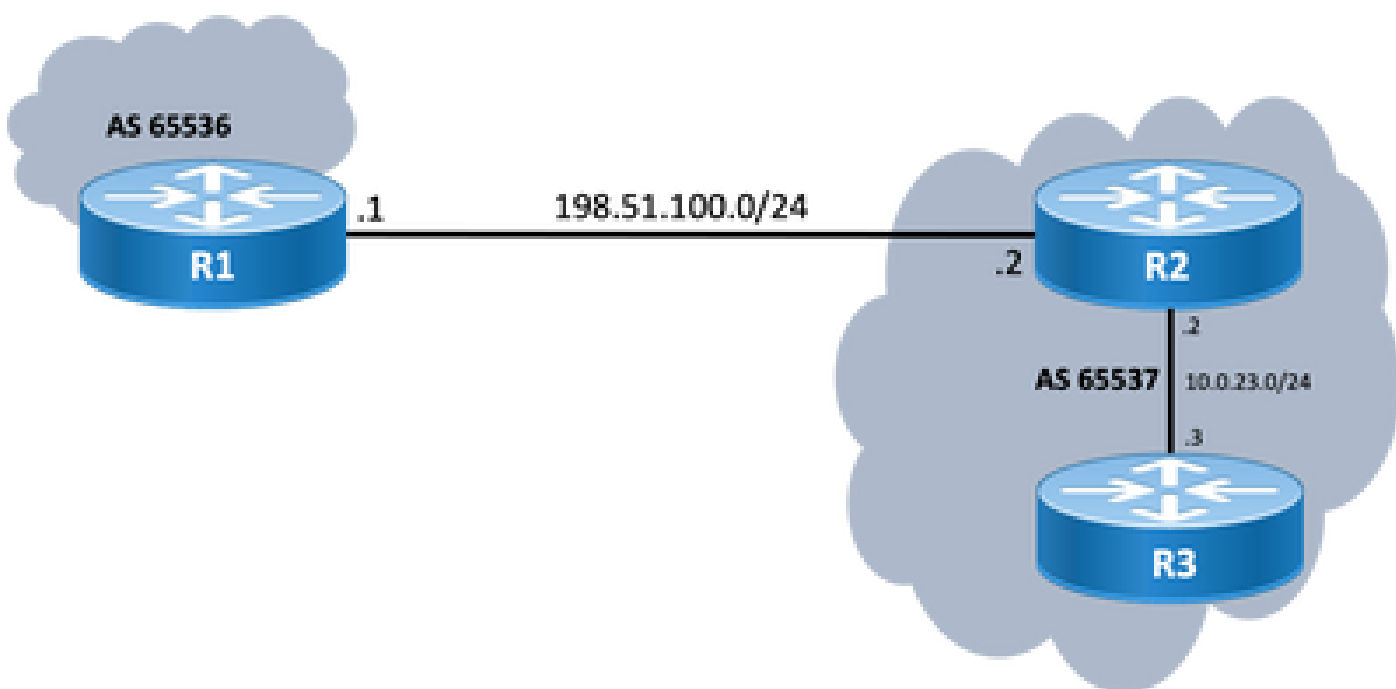
本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您的网络处于活动状态，请确保您了解所有命令的潜在影响。

## 背景信息

本文档介绍用于解决边界网关协议(BGP)中最常见问题的基本指南，提供纠正措施、用于检测问题根本原因的有用命令/调试以及避免潜在问题的最佳实践。请记住，思科TAC不能考虑所有可能的变量和场景，可能需要进行更深入的分析。

## 拓扑

使用此拓扑图作为本文档中提供的输出的参考。



## 场景和问题

### 邻接关系断开

如果BGP会话关闭且未启动，请发出 `show ip bgp all summary` command. 您可以在此处找到会话的当前状态：

- 如果会话处于not up状态，它可能在IDLE和ACTIVE之间变化（取决于有限状态机进程）。
- 如果会话处于启用状态，您将看到收到的前缀数量。

```
<#root>
```

```
R2#
```

```
show ip bgp all summary
```

```
For address family: IPv4 Unicast
```

```

BGP router identifier 198.51.100.2, local AS number 65537
BGP table version is 19, main routing table version 19
18 network entries using 4464 bytes of memory
18 path entries using 2448 bytes of memory
1/1 BGP path/bestpath attribute entries using 296 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 7208 total bytes of memory
BGP activity 18/0 prefixes, 18/0 paths, scan interval 60 secs
18 networks peaked at 11:21:00 Jun 30 2022 CST (00:01:35.450 ago)

```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
10.0.23.3	4	65537	6	5	19	0	0	00:01:34	18
198.51.100.1	4	65536	0	0	1	0	0	never	Idle

## 无连接

必须确保两个对等体之间的连接，以便能够建立端口179上的TCP会话。它们是否直接相连。简单的ping操作对于此问题很有用。如果在环回接口之间建立对等，则必须执行环回至环回ping。如果在没有特定环回作为源接口的情况下执行ping测试，则将传出物理接口IP地址用作数据包的源IP地址，而不是路由器的环回IP地址。

如果ping不成功，请考虑以下原因：

- 没有连接的路由对等体或根本没有路由： `show ip route peer_IP_address` 可用。
- 第1层问题：需要考虑物理接口、SFP（连接器）、电缆或外部问题（传输和提供商，如果适用）。
- 检查可以阻止连接的任何防火墙或访问列表。

如果ping成功，请考虑以下因素：

## 配置问题

- 错误的IP地址或AS配置：错误的IP地址，不显示此类消息，但请确保正确配置完成。对于错误的AS，您必须看到一条消息，例如 `show logging` 命令。

```
%BGP-3-NOTIFICATION: sent to neighbor 198.51.100.1 passive 2/2 (peer in wrong AS) 2 bytes 1B39
```

检查两端的BGP配置以更正AS编号或对等IP地址。

- 路由器 ID 重复:

```
%BGP-3-NOTIFICATION: sent to neighbor 198.51.100.1 passive 2/3 (BGP identifier wrong) 4 bytes 0A0A0A0A
```

通过检查两端的BGP标识符 `show ip bgp all summary` 并更正重复问题。这可以通过全局命令手动实现 `bgp router-id X.X.X.X` 在bgp路由器配置下。作为一种最佳实践，请确保将路由器ID手动设置为唯一编号。

- BGP源和TTL:

大多数iBGP会话是通过通过IGP可访问的环回接口配置的。此环回接口必须显式定义为源，请使用命令执行此操作 `neighbor ip-address update-source interface-id` .

对于eBGP对等体，直接连接的接口通常用于对等连接，并且会检查是否有Cisco IOS/Cisco IOS XE来实现此目的，或者检查是否有 甚至不尝试建立会话。如果在直连路由器上尝试将eBGP从环回转换为环回，则可以通过禁用对两端的特定邻居的此检查 `neighbor ip-address disable-connected-check` .

但是，如果eBGP对等体之间存在多个跃点，则需要适当的跃点计数，请确保 `neighbor ip-address ebgp-multihop [hop-count]`配置了正确的跳数，因此可以建立会话。

如果未指定跳数，则iBGP会话的默认TTL值为255，而eBGP会话的默认TTL值为1。

## TCP会话问题

测试端口179的一个有用操作是从一个对等设备到另一个对等设备的手动telnet:

```
<#root>
R1#
telnet 198.51.100.2 179
Trying 198.51.100.2, 179 ... Open
[Connection to 198.51.100.2 closed by foreign host]
```

打开/连接关闭或远程主机拒绝连接表示数据包到达远程端，然后，确保远端的控制平面没有问题。否则，如果存在无法到达的目标，请检查可以阻止TCP端口179、BGP数据包或路径上任何数据包丢失的任何防火墙或访问列表。

如果出现身份验证问题，您会看到以下消息：

```
%TCP-6-BADAUTH: Invalid MD5 digest from 198.51.100.1(179) to 198.51.100.2(20062) tableid - 0
%TCP-6-BADAUTH: No MD5 digest from 198.51.100.1(179) to 198.51.100.2(20062) tableid - 0
```

检查身份验证方法、密码和相关配置，并进一步排除故障，请参阅[BGP对等体之间的MD5身份验证配置示例](#)。

如果TCP会话未启动，您可以使用以下命令进行隔离：

```
show tcp brief all
show control-plane host open-ports
debug ip tcp transactions
```

## 邻接退回

如果会话处于打开和关闭状态，请查找 `show log` 您可以看到一些场景。

## 接口摆动

```
%BGP-5-ADJCHANGE: neighbor 198.51.100.2 Down Interface flap
```

如消息所示，此故障的原因是接口关闭情况，请查找端口/SFP、电缆或断开连接上的任何物理问题。

## 保持计时器已过期

```
%BGP-3-NOTIFICATION: sent to neighbor 198.51.100.2 4/0 (hold time expired) 0 bytes
```

这是一种非常常见的情况；这意味着路由器在保持计时器过期之前未接收或处理keepalive消息或任何更新消息。设备发送通知消息并关闭会话。此问题最常见的原因如下：

- 接口问题：查找两个对等体的连接接口上的任何输入错误、输入队列丢弃或物理问题；`show interface` 可用于此目的。
- 传输中的数据包丢失：有时，Hello数据包可能在传输过程中被丢弃，这是确保在接口级别捕获数据包的最佳方法。
  - 您可以在Cisco IOS [和](#) Cisco IOS [XE](#)设备上使用嵌入式数据包捕获。
  - 如果在接口级别看到数据包，您需要确保它们到达控制平面(EPC) 在控制平面上，或 `debug bgp [vrf name] ipv4 unicast keepalives` 非常有用。
- 高CPU：高CPU状况可能导致控制平面发生丢弃，`show processes cpu [sorted|history]` 对于确定问题非常有用。基于该平台，您可以找到使用CPU参考文档进行故障排除的[下一步骤](#)
- CoPP策略问题：故障排除方法因平台而异，超出本文档的范围。
- MTU不匹配：如果路径中存在MTU差异，并且如果在从源到目标的路径中阻止ICMP消息，则

PMTUD不起作用且可能导致会话抖动。使用协商的MSS值和DF位集发送更新。如果路径中的设备或甚至目标设备无法接受具有较高MTU的数据包，它会向BGP扬声器发送一条ICMP错误消息。目标路由器会等待BGP keepalive或BGP更新数据包更新其抑制计时器。

- 。您可以检查与协商的MSS `show ip bgp neighbors ip_address`.

使用df集对特定邻居执行ping测试可以显示这样的MTU是否沿路径有效：

```
<#root>
```

```
ping 198.51.100.2 size
```

```
max_seg_size
```

```
df
```

如果发现MTU问题，则必须对配置进行准确检查，以确保MTU值在整个网络中保持一致。

---

注：有关MTU的详细信息，请参阅使用MTU[排除BGP邻居抖动故障](#)。

---

## AFI/SAFI问题

```
%BGP-5-ADJCHANGE: neighbor 198.51.100.2 passive Down AFI/SAFI not supported
```

```
%BGP-3-NOTIFICATION: received from neighbor 198.51.100.2 active 2/8 (no supported AFI/SAFI) 3 bytes 000
```

地址系列标识符(AFI)是由多协议BGP(MP-BGP)添加的功能扩展。它涉及特定网络协议（例如IPv4、IPv6等），并通过后续地址系列标识符(SAFI)（例如单播和组播）与额外粒度相关。MBGP通过BGP路径属性(PA)MP\_REACH\_NLRI和MP\_UNREACH\_NLRI实现此分离。这些属性在BGP更新消息内传输，用于传输不同地址系列的网络可达性信息。

该消息将为您提供IANA注册的这些AFI/SAFI的编号：

- [IANA地址系列号](#)
- [后续地址系列标识符\(SAFI\)参数](#)
- 检查两端的地址系列的BGP配置，以更正任何不必要的地址系列。
- 使用 `neighbor ip-address dont-capability-negotiate` 两端。有关详细信息，请参阅[不支持的功能导致BGP对等体故障](#)。

## 路径安装和选择

有关BGP如何工作以及选择最佳路径的详细信息，请参阅[BGP最佳路径选择算法](#)。

## 下一跳

对于要安装到路由表中的路由，需要到达下一跳，否则，即使前缀位于我们的Loc-RIB BGP表中，它也不会进入RIB。作为环路避免规则，在Cisco IOS/Cisco IOS XE上，iBGP不会更改下一跳属性，并且在eBGP重写下一跳并预置其AS\_PATH时，仅保留AS\_PATH。

您可以使用以下命令检查下一跳 `show ip bgp [prefix]`。它为您提供下一跳和无法访问的词。在本示例中，这是R1通过eBGP向R2通告的前缀，由R3通过R2的iBGP连接获取该前缀。

```
<#root>
```

```
R3#
```

```
show ip bgp 192.0.2.1
```

```
BGP routing table entry for 192.0.2.1/32, version 0
```

```
Paths: (1 available, no best path)
```

```
Not advertised to any peer  
Refresh Epoch 1  
65536
```

```
198.51.100.1 (inaccessible)
```

```
from 10.0.23.2 (10.2.2.2)  
Origin incomplete, metric 0, localpref 100, valid, internal  
rx pathid: 0, tx pathid: 0  
Updated on Jul 1 2022 13:44:19 CST
```

在输出中，下一跳是R1的传出接口，而R3不知道该接口。要解决此问题，您可以通过IGP通告下一跳、静态路由或使用 `neighbor ip-address next-hop-self` 命令，修改下一跳IP（直接连接）。在图示示例中，此配置需要在R2上；指向R3的邻居(`neighbor 10.0.23.3 next-hop-self`)。

因此，下一跳会更改(在 `clear ip bgp 10.0.23.2 soft`)连接到直接连接的接口（可访问），且已安装前缀。

```
<#root>
```

```
R3#
```

```
show ip bgp 192.0.2.1
```

```
BGP routing table entry for 192.0.2.1/32, version 24
```

```
Paths: (1 available, best #1, table default)
```

```
Not advertised to any peer  
Refresh Epoch 1  
65536
```

```
10.0.23.2
```

```
from 10.0.23.2 (10.2.2.2)
```

```
Origin incomplete, metric 0, localpref 100, valid, internal, best
rx pathid: 0, tx pathid: 0x0
Updated on Jul 1 2022 13:46:53 CST
```

## RIB故障

当路由无法安装到全局RIB中时，就会发生这种情况，这会导致RIB故障。常见原因是，对于管理距离较短的另一个路由协议，RIB上已经存在相同的前缀，但是使用show ip bgp rib-failure命令可以看到RIB故障的确切原因。有关更详细的说明，请参阅以下链接：

---

注：您可以按照[了解BGP RIB-failure和bgp suppress-inactive命令中的说明识别和更正此类问题。](#)

---

## 竞争条件

最常见的问题是，在相互重分发场景中，IGP优先于eBGP。当IGP路由重分发到BGP时，它被视为由BGP在本地生成，默认情况下权重为32768。默认情况下，将从BGP对等体接收的所有前缀分配的本地权重为0。因此，如果必须比较相同的前缀，根据BGP最佳路径选择过程将权重较高的前缀添加到路由表中，这就是将IGP路由安装到RIB上的原因。

此问题的解决方案是，在路由器bgp配置下，为从BGP对等体接收的所有路由设置更高的权重：

```
<#root>
neighbor
ip-address
  weight 40000
```

---

注意：有关详细说明，请参阅[了解网络故障切换场景中BGP权重路径属性的重要性。](#)

---

## 其他问题

### BGP慢速对等体

它是一种对等体，无法跟上发送方生成更新消息的速率。对等体出现此问题的原因有很多：其中一个对等体的CPU使用率较高、链路上的流量过大或流量丢失、带宽资源等。

---

注意：要帮助识别和纠正慢速对等体问题，请参阅[使用BGP“慢速对等体”功能解决慢速对等体问题。](#)

---



## 内存问题

BGP使用分配给Cisco IOS进程的内存，以维护网络前缀、最佳路径、策略和所有相关的配置，使其正常运行。通过命令查看整体进程 `show processes memory sorted`：

<#root>

R1#

`show processes memory sorted`

Processor Pool Total: 2121414332 Used: 255911152 Free: 1865503180

reserve P Pool Total: 102404 Used: 88 Free: 102316

lsmapi\_io Pool Total: 3149400 Used: 3148568 Free: 832

PID TTY Allocated Freed

Holding

	Getbufs	Retbufs	Process			
0	0	266231616	81418808	160053760	0	0 *Init*
662	0	34427640	51720	34751920	0	0 SBC main process
85	0	9463568	0	8982224	0	0 IOSD ipc task
0	0	34864888	25213216	8513400	8616279	0 *Dead*
504	0	696632	0	738576	0	0 QOS_MODULE_MAIN
518	0	940000	8616			

613760

0 0

BGP Router

228	0	856064	345488	510080	0	0 mDNS
82	0	547096	118360	417520	0	0 SAMsgThread
0	0	0	0	395408	0	0 *MallocLite*

处理器池是使用的内存；在本例中大约为2.1 GB。接下来，您必须查看“暂挂”列以标识暂挂其大部分的子进程。然后，您需要检查您拥有的BGP会话、接收的路由数以及使用配置。

减少BGP内存占用的常见步骤：

- BGP过滤：如果不需要接收完整的BGP表，请使用策略过滤路由并仅安装所需的前缀。
- 软重新配置：在BGP配置下查找`neighbor ip_address soft-reconfiguration inbound`；此命令允许您查看在任何入站策略(Adj-RIB-in)之前收到的所有前缀。但是，此表需要当前BGP本地RIB表的大约一半来存储此信息，因此您可以避免此配置，除非是强制要求的，或者当前前缀很少。

---

注意：有关如何优化BGP的详细信息，请参阅[配置BGP路由器以实现最佳性能和减少内存消耗](#)。

---

## 高 CPU 利用率

路由器使用不同的进程来运行BGP。要验证BGP进程是CPU使用率较高的原因，请使用 `show process cpu sorted` 命令。

```
<#root>
```

```
R3#
```

```
show processes cpu sorted
```

```
CPU utilization for five seconds: 0%/0%; one minute: 0%; five minutes: 0%
```

PID	Runtime(ms)	Invoked	uSecs	5Sec	1Min	5Min	TTY	Process
163	36	1463	24	0.07%	0.00%	0.00%	0	ADJ background
62	28	132	212	0.07%	0.00%	0.00%	0	Exec
2	39	294	132	0.00%	0.00%	0.00%	0	Load Meter
1	0	4	0	0.00%	0.00%	0.00%	0	Chunk Manager
3	27	1429	18	0.00%	0.00%	0.00%	0	

```
BGP Scheduler
```

4	0	1	0	0.00%	0.00%	0.00%	0	R0 Notify Timers
63	4	61	65	0.00%	0.00%	0.00%	0	

```
BGP I/O
```

83	924	26	35538	0.00%	0.03%	0.04%	0	
----	-----	----	-------	-------	-------	-------	---	--

```
BGP Scanner
```

96	142	11651	12	0.00%	0.00%	0.00%	0	Tunnel BGP
7	0	1	0	0.00%	0.00%	0.00%	0	DiscardQ Backgro

以下是克服由于BGP而导致的CPU使用率较高的常见进程、原因和常规步骤：

- BGP路由器：每秒运行一次以保护更快的融合。它是最重要的线索之一。它读取bgp更新消息，验证前缀/网络和属性，更新每个AFI/SAFI网络/前缀表和属性表，执行最佳路径计算以及其他许多任务。

大规模路由变动是导致这种情况的一个非常常见的场景。

- BGP扫描程序：默认情况下每60秒运行一次的低优先级进程。此过程检查整个BGP表以验证下一跳可达性，并相应地更新BGP表，以防路径发生任何更改。它通过路由信息库(RIB)进行重分发。

检查平台规模，随着安装更多前缀和路由以及TCAM的使用、需要更多资源，通常会出现设备过载的情况。

---

注：有关如何对这两个进程进行故障排除的详细信息，请参阅[对由BGP扫描程序或路由器进程导致的CPU使用率过高进行故障排除。](#)

---

- BGP I/O：在收到BGP控制数据包时运行，并管理BGP数据包的排队和处理。如果BGP队列中长期接收的数据包过多，或者TCP存在问题，则路由器会显示BGP I/O进程导致的CPU使用率较高的症状。(通常，在这种情况下，BGP路由器也很高。查看消息计数以标识对等设备，并捕获数据包以标识这些消息的来源。)
- BGP打开：会话建立过程中使用的进程。除非会话停滞在“打开”状态，否则不会出现常见的CPU使用率较高的问题。
- BGP事件：负责下一跳处理。查找接收的前缀上的下一跳摆动。

## 相关信息

- [技术支持和文档 - Cisco Systems](#)
- [BGP配置指南](#)
- [在 BGP 对等体之间进行 MD5 认证的配置示例](#)
- [嵌入式数据包捕获](#)
- [使用MTU的BGP邻居抖动故障排除](#)
  
- [IANA地址系列号](#)
  
- [后续地址系列标识符\(SAFI\)参数](#)
  
- [不支持的功能导致BGP对等体故障](#)
- [BGP 最佳路径选择算法](#)
- [了解BGP RIB-failure和bgp suppress-inactive命令](#)
- [了解 BGP Weight 路径属性在故障切换场景中的重要性](#)
- [使用BGP“慢速对等体”功能解决慢速对等体问题](#)
- [配置BGP路由器以实现最佳性能并减少内存消耗](#)
- [对BGP扫描程序或路由器进程导致的CPU使用率过高进行故障排除](#)

## 关于此翻译

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

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

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