目录

简介

如输出show process cpu命令所显示，本文描述Cisco IOS路由器也许体验高CPU利用率由于边界网关协议(BGP)路由器进程或BGP扫描程序进程的情况。CPU 使用率过高这一情况的持续时间取决于多个条件，尤其是 Internet 路由表的大小和特定路由器在其路由和 BGP 表中保留的路由数。

show process cpu 命令可显示过去五秒钟、一分钟和五分钟的 CPU 平均使用率。CPU 使用率数值显示出使用率与流入负载并不具有真实的线性关系。以下是一些主要原因：

- 在实际的全球网络中，CPU 必须处理网络管理等多种系统维护功能。
- CPU 必须处理定期的和事件触发的路由更新。
- 还存在与流量负载不成比例的其他内部系统开销操作，例如对资源可用性的轮询。

您能也使用show processes cpu命令为了得到CPU活动的某个征兆。

一旦读本文，您应该了解每BGP进程角色，并且，当每进程运行。另外，您应该了解BGP收敛和技术优化收敛时间。

开始使用前

规则

有关文档规则的详细信息，请参阅 Cisco 技术提示规则。

先决条件

本文档要求您了解如何解读 **show process cpu** 命令。请参阅参考资料对 Cisco 路由器上的 CPU 使用率过高进行故障排除。

使用的组件

本文档中的信息根据Cisco IOS软件版本12.0。

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。如果您使用的是真实网络，请确保您已经了解所有命令的潜在影响。

了解 BGP 进程

Cisco IOS进程，一般来说，包括执行任务，例如系统维护，交换信息包的各自的线索和相关的数据和实现路由协议。在路由器以启用BGP执行的几Cisco IOS进程运行。请使用 **show process cpu|** 命令查看由 BGP 进程导致的 CPU 使用率的数字。

下表列出了 BGP 进程的功能，并显示出每个进程根据其处理的任务在不同的时间运行。由于 BGP 扫描程序和 BGP 路由器进程负责处理大量计算，因此，您可能会发现由于其中某一进程导致 CPU 使用率过高。以下部分将更详细地讨论这些进程。

进程名	说明	间隔
BGP Open	执行 BGP 对等体的建立。	初始化时，与 BGP 对等体建立 TCP 连接时。
BGP I/O	处理 BGP 数据包（例如更新和 KEEPALIVE）的排队和处理。	收到 BGP 控制数据包时。
BGP Scanner	扫描 BGP 表，并确认下一跳的可达性。BGP 扫描程序还会检查条件通告，以确定 BGP 是否应该通告条件前缀并/或执行路由衰减。在 MPLS VPN 环境中，BGP 扫描程序将路由导入和导出到特定的 VPN 路由和转发实例 (VRF) 中。	每分钟执行一次。
BGP Rout	计算最佳的 BGP 路径，并处理所有路由“波动”。它发送并且接收路由，也设立对等体，并且与路由信息库(RIB)呼应。	每秒钟执行一次，在添加、删除 BGP 对等体或通过软件对其进行重新配置时也会执行。

er		

BGP 扫描程序引起的高 CPU 使用率

当路由器上的 Internet 路由表较大时，BGP 扫描程序进程将导致在较短持续时间内 CPU 使用率过高。BGP 扫描程序每分钟扫描一次 BGP RIB 表并执行重要的维护任务。这些任务包括检查路由器 BGP 表中引用的下一跳，以及验证是否可到达下一跳设备。因此，扫描和验证大型 BGP 表需要花费相当长的时间。

由于 BGP 扫描程序进程将扫描整个 BGP 表，因此，CPU 使用率过高这一情况的持续时间根据邻居数量和每个邻居获知的路由数而有所不同。请使用 show ip bgp summary 和 show ip route summary 命令获取此信息。

BGP 扫描程序进程扫描 BGP 表以更新所有数据结构，并扫描路由表以进行路由重分配。(亦称在此上下文，路由表是路由信息库(RIB)，路由器输出，当您执行show ip route命令)。这两个表分别存储在路由器内存中，可能会非常大，因此很占用 CPU 周期。

以下示例是 debug ip bgp updates 命令的输出，捕获了 BGP 扫描程序的执行结果，此扫描程序从最小的前缀编号或者 0.0.0.0 开始扫描。

当 BGP 扫描程序运行时，优先级低的进程需要等待更长的时间才可访问 CPU。一个低优先级进程流程控制互联网控制消息协议(ICMP)数据包例如ping。由于 ICMP 进程必须等在 BGP 扫描程序之后，因此发往或源自路由器的数据包可能会遇到比预期时间更长的延时。周期是 BGP 扫描程序运行一段时间并自行暂停，然后 ICMP 运行。相反，应该通过思科快速转发(CEF)交换通过路由器被发送的ping，并且不应该忍受任何另外的延迟。对定期出现的延时峰值进行故障排除时，请将通过路由器转发的数据包转发时间与由路由器 CPU 直接处理的数据包转发时间进行比较。

注意： 指定 record route 等 IP 选项的 ping 命令也需要由 CPU 直接进行处理，可能会遇到更长的转发延时。

请使用 **show process|** 命令查看 CPU 优先级。以下示例输出中的“Lsi”值使用“L”表示优先级较低的进程。

```
6513# show processes | include BGP Scanner 172 Lsi 407A1BFC       29144     29130     1000
8384/9000   0 BGP Scanner
```

BGP 路由器进程引起的高 CPU 使用率

BGP 路由器进程约每秒钟运行一次，以检查工作。BGP 收敛定义了首个 BGP 对等体的建立时间和 BGP 收敛时间之间的持续时间。为确保收敛时间尽可能最短，BGP 路由器将会占用所有空闲的 CPU 周期。但是在开始之后，它将会间歇性地释放（或暂停）CPU。

收敛时间是对 BGP 路由器在 CPU 上所花费时间（而非总时间）的直接测量。此步骤显示高效CPU使用状况在BGP收敛期间并且交换与两外部BGP (EBGP)对等体的BGP前缀。

1. 在开始测试之前首先获取 CPU 正常使用率的基准。`router# show process cpu  CPU utilization for five seconds: 0%/0%; one minute: 4%; five minutes: 5%`
2. 测试开始之后，CPU 使用率达到 100%。**show process cpu** 命令显示 CPU 使用率过高这一情况是由 BGP 路由器导致的，在以下输出中表示为 139（BGP 路由器的 IOS 进程 ID）。
   `router# show process cpu  CPU utilization for five seconds: 100%/0%; one minute: 99%; five minutes: 81%!--- Output omitted.139 6795740 1020252 6660 88.34% 91.63% 74.01% 0 BGP Router`
3. 通过在事件期间捕获 **show ip bgp summary** 和 show process cpu 命令的多个输出监控路由器

。**show ip bgp summary** 命令可捕获 BGP 邻居的状态。`router# **show ip bgp summary**`
`Neighbor V AS MsgRcvd MsgSent TblVer InQ OutQ Up/Down State/PfxRcd 10.1.1.1 4 64512 309453 157389 19981 0 253 22:06:44 111633 172.16.1.1 4 65101 188934 1047 40081 41 0 00:07:51 58430`

4. 当路由器与其 BGP 对等体完成前缀交换时，CPU 使用率应返回到正常水平。计算出的一分钟和五分钟平均数也将回落，并且可能会显示该数字在五秒以上的时间内高于正常水平。`router# **show process cpu** CPU utilization for five seconds: 3%/0%; one minute: 82%; five minutes: 91%`
5. 请使用所捕获的以上 show 命令输出计算 BGP 收敛时间。**具体而言，请使用 show ip bgp summary 命令的“Up/Down”列，并比较 CPU 使用率过高这一情况的开始和停止时间。**通常情况下，当交换较大的 Internet 路由表时，BGP 收敛可能会花费数分钟。

注意： 在设备的高CPU能也归结于BGP表的不稳定性。如果路由器接收路由表的两复制，一个从有ISP的EBGP对等体和其他从在网络的IBGP同位体。此的根本原因是在设备的内存数量。思科推荐RAM至少1 Gig互联网路由表的单一副本的。要避免此不稳定性，请增加在设备的RAM或过滤前缀，以便它和内存占用的BGP表被解除。

性能改进

随着 Internet 路由表中路由数的增加，BGP 收敛所花费的时间也相应增加。通常，收敛可定义为使所有路由表达成一致状态的进程。如果满足以下条件，则认为 BGP 已收敛：

- 已接受所有路由。
- 所有路由都已安装在路由表中。
- 所有对等体的表版本与 BGP 表的表版本相同。
- 所有对等体的 InQ 和 OutQ 都为零。

本部分介绍了为缩短 BGP 收敛时间而进行的某些 IOS 性能改进，这些性能改进可减少由 BGP 进程导致的 CPU 使用率较高的情况。

TCP 对等连接排队

BGP 现在不再每秒钟对数据进行一次排队，而是主动将数据从 BGP OutQ 排队到每个对等体的 TCP 套接字中，直至完全清空 OutQ。由于 BGP 现在的发送速率更快，因此 BGP 的收敛速度也更快。

BGP 对等体组

BGP 对等体组不但有助于简化 BGP 配置，同时还可以提高可扩展性。所有对等体组成员必须共享一个公用出站策略。因此，可向每个组成员发送同一更新数据包，以减少 BGP 向对等体通告路由所需的 CPU 周期数。换言之，使用对等体组时，BGP 仅扫描对等体组引导路由器上的 BGP 表，通过出站策略过滤前缀，并生成更新，然后向对等体组引导路由器发送更新。接着，引导路由器将更新复制到与其同步的组成员中。不使用对等体组时，BGP 必须扫描每个对等体的表，通过出站策略过滤前缀，并生成更新，然后仅向一个对等体发送更新。

路径 MTU 和 ip tcp path-mtu-discovery 命令

单个数据包中可传输的字节数存在一个限值，所有 TCP 会话都受此限值的限定。默认情况下此限制，叫作最大分段尺寸(MSS)，是536个字节。换言之，将数据包向下传递到 IP 层之前，TCP 首先会将传输队列中的数据包划分为大小为 536 个字节的块。**请使用 show ip bgp neighbors|** 显示 BGP 对等体的 MSS：

```
Router# show ip bgp neighbors | include max data Datagrams (max data segment is 536 bytes):
Datagrams (max data segment is 536 bytes): Datagrams (max data segment is 536 bytes): Datagrams
(max data segment is 536 bytes):
```

536 个字节的 MSS 所具有的优点是：由于大多数链路使用的 MTU 至少为 1500 个字节，因此，在通往目的地的路径上 IP 设备不可能将数据包分段。缺点是较小的数据包将会增加用于传输的带宽开销。由于 BGP 建立了到所有对等体的 TCP 连接，因此，536 个字节的 MSS 将会影响 BGP 收敛时间。

解决方案是使用 ip tcp path-mtu-discovery 命令启用路径 MTU (PMTU) 功能。可以使用此功能动态确定 MSS 值的大小，而不创建需要进行分段的数据包。通过 PMTU，TCP 可以确定 TCP 会话所有链路中的最小 MTU 大小。然后，TCP 将使用此 MTU 值（减去 IP 报头和 TCP 报头占用的空间）作为会话的 MSS。如果 TCP 会话仅通过以太网网段传输，则 MSS 将为 1460 个字节。如果它只横断SONET上的分组(POS)分段，则MSS将是4430个字节。MSS 从 536 个字节增加到 1460 或 4430 个字节可减少 TCP/IP 开销，这有助于 BGP 更快收敛。

启用 PMTU 之后，请再次使用 **show ip bgp neighbors|** 查看每个对等体的 MSS 值：

```
Router# show ip bgp neighbors | include max data Datagrams (max data segment is 1460 bytes):
Datagrams (max data segment is 1460 bytes): Datagrams (max data segment is 1460 bytes):
Datagrams (max data segment is 1460 bytes):
```

增加接口输入队列

如果 BGP 向多个对等体通告数千个路由，则 TCP 必须在短时间内传输数千个数据包。BGP 对等体将接收这些数据包，并向发通告的 BGP 扬声器发送 TCP 确认，这将导致 BGP 扬声器在短时间内收到大量 TCP ACK。如果 ACK 到达路由处理器的速率过高，则数据包会在入站接口队列中备份。默认情况下，路由器接口使用的输入队列大小为 75 个数据包。另外，特殊控制数据包例如BGP更新以选择性数据包丢弃(SPD)使用一个特殊队列。此特殊队列可存放 100 个数据包。在 BGP 收敛期间，TCP ACK 可能会迅速填充输入缓冲区的 175 个位置，新到达的数据包则必须丢弃。在包含 15 个或更多 BGP 对等体并交换完整 Internet 路由表的路由器上，每分钟每个接口上可能发生 10,000 次以上丢包。以下为重新启动 15 分钟之后一个路由器的示例输出：

```
Router# show interface pos 8/0 | include input queue Output queue 0/40, 0 drops; input queue
0/75, 278637 drops Router#
```

增加接口输入队列深度（使用 hold-queue <1-4096> in 命令）有助于减少丢弃的 TCP ACK 数量，从而减少 BGP 为执行收敛而必须承担的工作量。通常情况下，如果值为 1000，则可以解决由输入队列丢包导致的问题。

注意： Cisco 12000 系列当前使用的 SPD Headroom 默认值为 1000。它保留了输入队列大小的默认值 75。请使用 **show spd** 命令查看这些特殊输入队列。

IOS 12.0(19)S 的其他改进

IOS 12.0(19)S 包括对 BGP 对等体组代码的多项优化，以改进更新打包和复制。在讨论这些改进之前，我们将更仔细地了解一下更新打包和复制。

共享属性的该组合的BGP更新包括属性的组合(例如MED = 50和LOCAL_PREF = 120)和网络层列表可达性信息(NLRI)前缀。BGP 可在单个更新中列出的 NLRI 前缀越多，由于减少了开销（如 IP、TCP 和 BGP 报头），BGP 收敛的速度就越快。“更新打包”是指将 NLRI 打包到 BGP 更新中。例如，一个 BGP 表存放了包含 15,000 个唯一属性组合的 100,000 个路由，则如果以 100% 的效率打包 NLRI，BGP 将仅需要发送 15,000 个更新。

注意： 打包效率为 0% 意味着 BGP 需要在此环境中发送 100,000 个更新。

请使用 show ip bgp peer-group 命令查看 BGP 更新的效率。

如果对等体组成员是“同步的”，则 BGP 路由器将选择已为对等体组引导路由器格式化的更新消息，并为该成员复制该消息。复制对等体组成员的更新比重新格式化更新的效率更高。例如，假设对等体组包含 20 个成员，并且所有成员都需要接收 100 个 BGP 消息。百分之百复制意味着 BGP 路由器将为对等体组引导路由器格式化 100 个消息，然后将这些消息复制到其他 19 个对等体组成员中。要确认复制改进，请将复制的消息数与格式化的消息数（如 **show ip bgp peer-group** 命令所示）进行比较。改进可使收敛时间产生显著变化，并允许 BGP 支持更多的对等体。请看以下示例。

请使用 **show ip bgp peer-group** 命令检查更新打包和更新复制的效率。以下输出是对 6 个对等体组进行收敛测试的结果，其中前 5 个对等体组（eBGP 对等体）每组包含 20 个对等体，第 6 个对等体组（内部 BGP (iBGP) 对等体）包含 100 个对等体。此外，所使用的 BGP 表包含 36,250 个属性组合。

以下示例是在运行 IOS 12.0(18)S 的路由器上执行 **show ip bgp peer-group|** 命令的输出，显示信息如下：

```
Router# show interface pos 8/0 | include input queue Output queue 0/40, 0 drops; input queue
0/75, 278637 drops Router#
```

要计算每个对等体组的复制率，请将复制的更新数除以格式化的更新数：

1668500/836500 = 1.99 1455000/1050000 = 1.38 1844500/660500 = 2.79 1849000/656000 = 2.81 2003750/501250 = 3.99 12114785/2476715 = 4.89

如果 BGP 完全复制更新，则由于对等体组包含 20 个对等体，每个 eBGP 对等体组的复制率将为 19。应该为对等体组引导路由器格式化更新，然后将其复制到其他 19 个对等体中，这样可以实现最佳复制率 19。由于存在 100 个对等体，因此，iBGP 对等体组的理想复制率为 99。

如果 BGP 将更新完全打包，则只需格式化 36,250 个更新。由于 BGP 表中的属性组合数为 36,250，因此只需为每个对等体组生成 36,250 个更新。单个 iBGP 对等体组格式化约 2,500,000 个更新，而每个 eBGP 对等体组生成的更新数则在 500,000 至 1,000,000 范围内不等。

在运行 IOS 12.0(19)S 的路由器上，**show ip bgp peer-group|** 命令提供以下信息：

```
Router# show interface pos 8/0 | include input queue Output queue 0/40, 0 drops; input queue
0/75, 278637 drops Router#
```

注意： 更新打包最佳。为每个对等体组格式化的更新数正好是 36,250 个。

688750/36250 = 19 688750/36250 = 19 688750/36250 = 19 688750/36250 = 19 688750/36250 = 19 3588750/36250 = 99

注意： 更新复制同样具有很好的效果。

排除故障

使用这些步骤为了排除故障高CPU由于BGP扫描程序或BGP路由器：

- 收集 BGP 拓扑相关信息。确定 BGP 对等体数以及每个对等体通告的路由数。根据您的环境，CPU 使用率过高这一情况的持续时间是否合理？
- 确定 CPU 使用率过高所发生的时间。它是否与 BGP 表的定期扫描时间一致？
- 执行 show ip bgp flap-statistics 命令。CPU 使用率过高是否发生在接口抖动之后？
- 通过路由器执行 Ping 操作，然后从路由器发出 ping。将 ICMP Echo 作为低优先级进程进行处

理。文档了解 Ping 和 Traceroute 命令详细介绍了此内容。确保常规转发不受影响。
- 检查是否已在入站和出站接口上启用了快速交换和/或 CEF，以确保数据包可沿着快速转发路径进行转发。保证您看不到no ip route-cache cef命令在接口或**no ip cef命令**在全局配置。为了启用在全局配置模式的CEF，请使用ip cef命令。
- 从这些命令得到输出：
- 验证在路由器的DRAM。根据建议，应该有DRAM空间至少512 MB每个发送完全互联网路由表的BGP对等体。如果路由器有运行完全互联网路由表的两EBGP对等体，则1 GB DRAM空间推荐最低。被提及的此处DRAM空间是为BGP要求的内存。在路由器运行的其它特性将要求另外的DRAM空间。

相关信息

- BGP 支持页
- IP 路由支持页
- 技术支持 - Cisco Systems