對由BGP掃描程式或路由器進程引起的高CPU問題進行故障排除

目錄

簡介

本文說明如何使用BGP掃描器或路由器時，對CPU讀數偏高的原因進行疑難排解。

必要條件

需求

本文檔需要瞭解如何解釋**show process cpu**命令。

採用元件

本檔案中的資訊是根據Cisco IOS®軟體版本12.0。

本文中的資訊是根據特定實驗室環境內的裝置所建立。文中使用到的所有裝置皆從已清除（預設）的組態來啟動。如果您的網路運作中，請確保您瞭解任何指令可能造成的影響。

慣例

如需文件慣例的詳細資訊，請參閱思科技術提示慣例。

背景資訊

本檔案將說明**show process cpu** 命令的輸出所示，由於邊界閘道通訊協定(BGP)路由器進程或BGP掃描程式進程，Cisco IOS路由器可能會遇到高CPU使用率的情況。高CPU條件的持續時間取決於多個條件，特別是Internet路由表的大小以及特定路由器在其路由和BGP表中擁有的路由數。**show process cpu** 命令顯示過去五秒、一分鐘和五分鐘內的平均CPU利用率。CPU使用率數字不能提供有關所提供負載的真實線性使用率指示。

以下是一些主要原因：

- 在現實的網路中，CPU必須處理各種系統維護功能，例如網路管理。
- CPU必須處理定期和事件觸發的路由更新。
- 還有其它內部系統開銷操作，如資源可用性的輪詢，這些操作與流量負載不成比例。

您還可以使用**show processes cpu** 命令來獲取CPU活動的某些指示。

> **附註**：有關show命令的詳細資訊，請參閱Cisco IOS配置基礎命令參考

瞭解BGP進程

Cisco IOS進程通常包含執行任務的單個執行緒和相關資料，例如系統維護、交換資料包和實施路由協定。路由器上執行的幾個Cisco IOS進程使BGP能夠運行。使用**show process cpu |包括BGP命**令以檢視由於BGP進程而產生的CPU使用量。

下表列出了BGP進程的功能，並顯示了每個進程在不同時間運行，這些時間取決於所處理的任務。由於BGP掃描程式和BGP路由器進程負責大量計算，因此您可以看到由於其中任何一個進程導致的CPU使用率高。接下來的部分將更詳細地討論這些過程。

進程名稱	**說明**	**間隔**
BGP開啟	執行BGP對等體建立。	初始化時，建立與BGP對等體的TCP連線時。
BGP I/O	處理隊列中且已處理的BGP資料包，例如UPDATES和KEEPALIVE。	接收BGP控制封包時。
BGP掃描器	遍歷BGP表並確認下一跳的可達性。BGP掃描程式還會檢查條件通告，以確定BGP是否通告條件字首並執行路由減少。在MPLS VPN環境中，BGP掃描器將路由匯入和匯出到特定VPN路由和轉發例項(VRF)。	一分鐘一次
BGP路由器	計算最佳BGP路徑並處理任何路由流**動。它還會傳送和接收路**由，建立對等體，並與路由資訊庫(RIB)互動。	每秒一次，以及在新增、刪除或置一個BGP對等體時。

BGP掃描器導致的CPU使用率高

在承載大型Internet路由表的路由器上，由於BGP掃描程式進程導致的CPU使用時間較短。BGP掃描程式每分鐘掃描BGP RIB表並執行一次重要的維護任務。這些任務包括檢查路由器BGP表中引用的下一跳，並檢驗是否可以到達下一跳裝置。因此，大型BGP表需要花費相當多的時間進行遍歷和驗證。

因為BGP掃描程式進程運行在整個BGP表中，所以高CPU條件的持續時間會因鄰居數量和每個鄰居獲知的路由數量而異。使用**show ip bgp summary**和**show ip route summary**命令捕獲此資訊。BGP掃描程式進程會遍歷BGP表以更新任何資料結構，並遍歷路由表以重分佈路由。(在此上下文中，路由表也稱為路由資訊庫(RIB)，當您執行；show ip route命令時，路由器輸出該資訊)。 這兩個表分別儲存在路由器的記憶體中，可能很大，並且會消耗CPU週期。

debug ip bgp updates 命令的下一個示例輸出捕獲了BGP掃描程式的執行：

```
router#
2d17h: BGP: scanning routing tables
2d17h: BGP: 10.0.0.0 computing updates, neighbor version 8,
     table version 9, starting at 0.0.0.0
2d17h: BGP: 10.0.0.0 update run completed, ran for 0ms, neighbor
     version 8, start version 9, throttled to 9, check point net 0.0.0.0
2d17h: BGP: 10.1.0.0 computing updates, neighbor version 8,
     table version 9, starting at 0.0.0.0
2d17h: BGP: 10.1.0.0 update run completed, ran for 4ms, neighbor
     version 8, start version 9, throttled to 9, check point net 0.0.0.0
router#
```

當BGP掃描程式運行時，低優先順序進程需要等待更長時間才能訪問CPU。一個低優先順序的程式會控制網際網路控制訊息通訊協定(ICMP)封包，例如ping。目的地為路由器或源自路由器的封包可能會遇到比預期更高的延遲，因為ICMP程式必須晚於BGP掃描程式。週期是BGP掃描程式運行一段時間並暫停其本身，然後運行ICMP。相反，透過路由器傳送的ping必須透過思科快速轉送(CEF)進行交換，且不會遇到任何額外延遲。當您對延遲的週期性高峰進行故障排除時，請將通過路由器轉發的資料包的轉發時間與路由器上CPU直接處理的資料包進行對比。

> **附註**：指定IP選項（如記錄路由）的ping命令也要求CPU直接處理這些選項，這會導致更長的轉發延遲。

使用show**過程 | include bgp scanner命令**以檢視CPU優先順序。下一個示例輸出中的**Lsi**值使用**L**來表示低優先順序進程。

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

BGP路由器進程導致的CPU使用率高

BGP路由器進程大約每秒運行一次以檢查工作。BGP收斂定義建立第一個BGP對等體時的時間和BGP收斂的時間之間的持續時間。為了確保儘可能短的收斂時間，BGP路由器會消耗所有空閒CPU週期。但是啟動後，它會間歇性地放棄（或暫停）CPU。

收斂時間是BGP路由器在CPU上花費時間的直接度量，而不是總時間。此程式在BGP聚合期間顯示高CPU條件，並與兩個外部BGP(eBGP)對等點交換BGP字首。

1. 開始測試之前，請捕獲正常CPU使用率基線。
   ```
   router#show process cpu
     CPU utilization for five seconds: 0%/0%; one minute: 4%; five minutes: 5%
   ```
2. 測試開始後，CPU利用率達到100%。**show process** cpu命令顯示高CPU狀態是由BGP路由器引起的，在下一個輸出中以139（BGP路由器的Cisco IOS進程ID）表示。
   ```
   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. 此時，您可以監視和捕獲**show ip bgp summary**和**show process cpu**命令的多個輸出。**show ip bgp summary**命令會擷取BGP鄰居的狀態。

```
router#show ip bgp summary
  Neighbor    V    AS  MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down State/PfxRcd
  10.0.0.0    4  64512  309453  157389    19981    0  253 22:06:44 111633
  10.1.0.0  4  65101  188934    1047    40081   41    0 00:07:51 58430
```

4. 當路由器完成與其BGP對等體的字首交換時，CPU利用率將返回到正常水準。計算得出的1分鐘和5分鐘平均數同樣可以穩定下來，並會在比5秒速率更長的時段內顯示出高於正常水準的消息。

```
router#show process cpu
  CPU utilization for five seconds: 3%/0%; one minute: 82%; five minutes: 91%
```

5. 使用之前的show命令捕獲的輸出計算BGP收斂時間。特別是，使用**show ip bgp summary**命令的**Up/Down列，比較高CPU條件的開始和停止時間。通常，當一個大型網際網路路由表時，BGP收斂可能需要幾分鐘時間。已交換**

   附註：裝置上的CPU使用率較高也可能是由於BGP表的不穩定性引起的。如果路由器收到兩個路由表副本（一個來自與ISP對等的EBGP，另一個來自網路中的IBGP對等的），就會發生這種情況。根本原因是裝置上的記憶體量。Cisco建議對網際網路路由表的單個副本使用至少1 Gig的RAM。要避免這種不穩定性，請增加裝置上的RAM或過濾字首，以緩解BGP表及其佔用的記憶體。

效能改進

隨著Internet路由表中路由數量的增加，BGP收斂所需的時間也會增加。一般情況下，收斂定義為使所有路由表達到一致狀態的過程。當以下條件為真時，BGP會視為收斂：

- 已接受所有路由。
- 所有路由都已安裝到路由表中。
- 所有對等體的表版本等於BGP表的表版本。
- 所有對等體的InQ和OutQ均為零。

本節介紹為縮短BGP收斂時間而進行的某些Cisco IOS效能改進，這些效能改進可減少由BGP進程導致的CPU使用率較高的情況。

TCP對等連線的隊列

BGP現在會主動將資料從BGP OutQ排入每個對等體的TCP套接字中，直到OutQs完全耗盡。由於BGP現在以較快的速率傳送，因此BGP收斂得更快。

BGP對等體組

雖然它們有助於簡化BGP配置，但BGP對等體組還可以增強可擴充性。所有對等組成員必須共用一個公共出站策略。因此，可以將相同的更新資料包傳送到每個組成員，從而減少BGP向對等體通告路由所需的CPU週期數。換句話說，對於對等組，BGP只在對等組領導上遍歷BGP表，通過出站策略過濾字首，並生成更新，並將其傳送到對等組領導。反過來，引線會將更新複製到與其同步的組成員。如果沒有對等體組，BGP必須遍歷每個對等體的表，通過出站策略過濾字首，並生成僅傳送到一個對等體的更新。

路徑MTU和ip tcp path-mtu-discovery指令

所有TCP會話都受單個資料包中可傳輸位元組數的限制。此限制稱為最大片段大小(MSS)，預設為

536位元組。換句話說，TCP會在將封包傳遞到IP層之前，將傳輸佇列中的封包分解為536位元組的區塊。使用**show ip bgp neighbors | include max data命令**以顯示BGP對等體的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位元組MSS的優勢在於，由於大多數連結使用的MTU至少為1500位元組，因此封包不大可能會在前往目的地的路徑上的IP裝置上進行分段。缺點是較小的資料包會增加用於傳輸開銷的頻寬量。由於BGP建立到所有對等體的TCP連線，因此536位元組的MSS會影響BGP收斂時間。

解決方式為使用**ip tcp path-mtu-discovery** 指令啟用路徑MTU(PMTU)功能。您可以使用此功能以動態方式判斷MSS值可以有多大，同時，請不要建立需要分段的資料包。PMTU允許TCP確定TCP作業階段中所有連結中的最小MTU大小。接著，TCP會使用此MTU值（減去IP和TCP標頭的空間）作為作業階段的MSS。如果TCP作業階段僅遍歷乙太網路區段，則MSS為1460位元組。如果它僅遍歷SONET(POS)封包，則MSS為4430位元組。MSS從536增加到1460或4430位元組，降低了TCP/IP額外負荷，從而有助於BGP更快收斂。

啟用PMTU後，再次使用**show ip bgp neighbors | include max data命令**以檢視每個對等體的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):
```

增加介面輸入隊列

如果BGP向許多對等點通告數千個路由，則TCP必須在短時間內傳輸數千個資料包。BGP對等體會收到這些封包並向播發的BGP發言者傳送TCP確認，這會導致BGP發言者在較短時間內收到大量的TCP ACK。如果ACK到達的速率對於路由處理器來說太高，則資料包將在入站介面隊列中備份。預設情況下，路由器介面使用的輸入隊列大小為75個資料包。此外，特殊控制封包（例如BGP UPDATES）使用具有選擇性封包捨棄(SPD)的特殊佇列。 此特殊隊列可容納100個資料包。當BGP收斂時，TCP ACK可以快速填滿175個輸入緩衝點，到達的新資料包必須被丟棄。在具有15個或更多的BGP對等體並交換完整Internet路由表的路由器上，每分鐘可以看到超過10,000個丟包。以下是重新引導15分鐘後路由器輸出的範例：

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

如果您增加介面輸入佇列深度(使用**hold-queue in** 指令)，將有助於減少捨棄的TCP ACK數量，這減少了BGP進行收斂必須執行的工作量。通常，值1000可解決輸入佇列捨棄所導致的問題。

> **附註**：請注意這一點，因為輸入隊列增量可能會增加一些延遲。

Cisco IOS中的其他改進

Cisco IOS包括對BGP對等組代碼的多個最佳化，以改進更新打包和複製。在檢查這些改進之前，請更詳細地檢查更新打包和複製。

BGP更新包含一系列屬性（例如MED = 50和LOCAL_PREF = 120）和一系列網路層可達性資訊(NLRI)字首，它們共用這些屬性組合。BGP在單一更新中列出的NLRI字首越多，BGP收斂速度就越快，因為開銷（例如IP、TCP和BGP標頭）會減少。 **更新包**裝是指將NLRI打包到BGP更新中。例如，如果BGP表包含100,000個具有15,000個唯一屬性組合的路由，那麼如果NLRI以100%的效率打包，BGP只需要傳送15,000個更新。

附註：零的打包效率意味著BGP需要在此環境中傳送100,000個更新。

使用**show ip bgp peer-group** 命令檢視BGP更新的效率。

如果對等組成員處於同步狀態，BGP路由器會接收針對對等組領導進行格式化的更新消息，並為該成員複製該消息。複製對等組成員的更新比重新格式化更新要有效得多。例如，假設對等組有20個成員，並且所有成員都需要接收100條BGP消息。100%的複製意味著BGP路由器為對等組領導格式化了100條消息，並將這些消息複製到其他19個對等組成員。要確認複製改進，請將複製的消息數與格式化的消息數進行比較，如**show ip bgp peer-group** 命令所示。這些改進顯著縮短了收斂時間，允許BGP支援更多對等體。

例如，使用**show ip bgp peer-group** 命令檢查更新打包和更新複製的效率。下一個輸出來自具有6個對等體組的收斂測試，前5個對等體組（eBGP對等體）中的每個對等體組中有20個對等體，第6個對等體組(內部BGP(iBGP)對等體組中有100個對等體。 此外，使用的BGP表有36,250個屬性組合。

show ip bgp peer-group的下一個輸出示例 | include replicated命令在運行Cisco IOS 12.0(18)S的路由器上顯示以下資訊：

```
Update messages formatted 836500, replicated 1668500
Update messages formatted 1050000, replicated 1455000
Update messages formatted 660500, replicated 1844500
Update messages formatted 656000, replicated 1849000
Update messages formatted 501250, replicated 2003750

!-- The first five lines are for eBGP peer groups. Update messages formatted 2476715, replicated
12114785 !-- The last line is for an iBGP peer group.
```

為了計算每個對等組的複製速率，請將複製的更新數除以格式化的更新數：

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完全複製，則eBGP對等體組的複製速率均為19，因為對等體組中有20個對等體。更新是為對等組領導設定格式，然後複製到其他19個對等體。這提供了最佳複製速率19。iBGP對等體組的理想複製速率是99，因為有100個對等體。
- 如果BGP打包的更新非常完美，則只有36,250個格式化更新。您只需要為每個對等組生成36,250個更新，因為這是BGP表中的屬性組合數。僅iBGP對等體組就格式化了近2,500,000個更新，而eBGP對等體組每個都生成從500,000到1,000,000個更新的任意位置。

在執行Cisco IOS 12.0(19)S的路由器上，**show ip bgp peer-group | include** replicatedcommand提供以下資訊：

```
Update messages formatted 36250, replicated 688750
Update messages formatted 36250, replicated 688750
Update messages formatted 36250, replicated 688750
Update messages formatted 36250, replicated 688750
Update messages formatted 36250, replicated 688750
```

```
Update messages formatted 36250, replicated 3588750
```

附註：更新裝箱為最佳。每個對等組有恰好36,250個更新的格式。688750/36250 = 19 688750/36250 = 19 688750/36250 = 19 688750/36250 = 19 688750/36250 = 19 3588750/36250 = 99

附註：更新複製也非常完美。

疑難排解程式

使用以下步驟對因BGP掃描器或BGP路由器而導致的高CPU問題進行故障排除：

- 收集有關BGP拓撲的資訊。確定BGP對等體的數量和每個對等體通告的路由數。基於您的環境，高CPU條件的持續時間是否合理？
- 確定高CPU發生的時間。它是否與BGP表的定期步行一致？
- 高CPU是否遵循介面翻動？如果啟用了阻尼功能，則可以使用命令**show ip bgp dampening flap-statistics** 命令。
- 通過路由器ping，然後從路由器ping。ICMP回應作為低優先順序進程處理。瞭解Ping和Traceroute指令檔案對此有更詳細的說明。確保常規轉發不受影響。
- 檢查入站和出站介面上是否啟用了快速交換和/或CEF時，需要確保資料包可以遵循快速轉發路徑。請確保在介面上未看到 **no ip route-cache cef** 命令，在全域組態上未看到**no ip cef**命令。要在全域性配置模式下啟用CEF，請使用**ip cef** 命令。
- 檢查平台擴展，因為在大多數情況下，它是由於裝置過載導致出現此類情況所致。此外，還確保路由器上有適當的三重內容可定址儲存器(TCAM)空間。

- 驗證路由器上是否有足夠的記憶體。根據建議，每個BGP對等體至少要有1 GB的DRAM分配給傳送完整網際網路路由表的Cisco IOS空間。此處提到的DRAM空間只是BGP所需的記憶體。路由器上運行的其他功能可能需要額外的空間。

相關資訊

- **IP 路由支援頁面**
- **技術支援 - Cisco Systems**