邊界網關協定基本問題故障排除

目錄

簡介

本文說明如何對邊界閘道通訊協定(BGP)的最常見問題進行疑難排解，並提供基本解決方案和准則。

必要條件

需求

本文件沒有特定先決條件。基本BGP通訊協定知識很有用，您可以參閱BGP組態設定指南以瞭解詳細資訊。

採用元件

本檔案所述內容不限於特定軟體和硬體版本，但命令適用於Cisco IOS®和Cisco IOS-XE®。

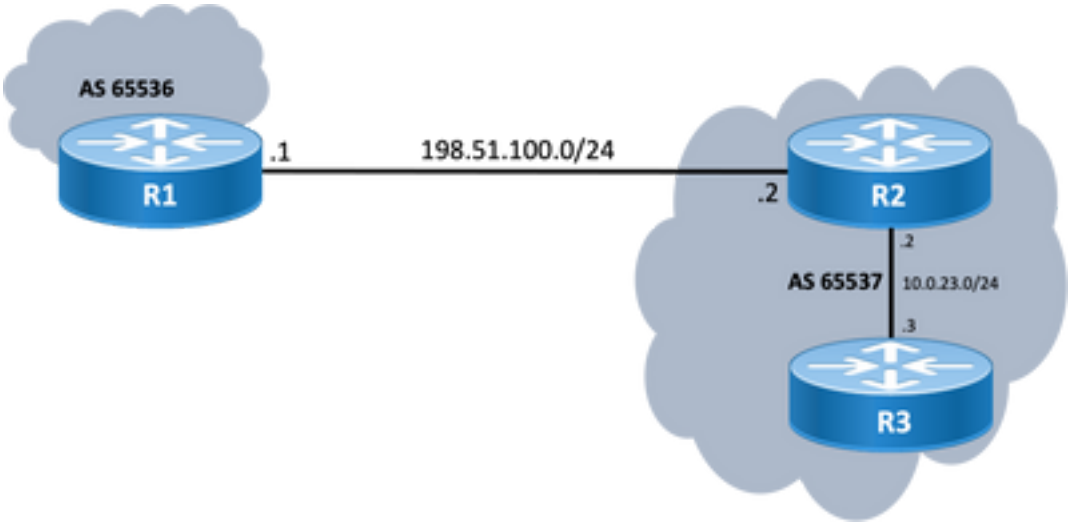
本文中的資訊是根據特定實驗室環境內的裝置所建立。文中使用到的所有裝置皆從已清除（預設）的組態來啟動。如果您的網路運作中，請確保您瞭解任何指令可能造成的影響。

背景資訊

本文描述對邊界網關協定(BGP)中最常見問題進行故障排除的基本指南，提供了糾正措施、用於檢測問題的根本原因的有用命令/調試以及避免潛在問題的最佳做法。請記住，思科TAC不能考慮所有可能的變數和情景，因此需要更深入的分析。

拓撲

請使用此拓撲圖作為本文檔中提供的輸出的參考。



情景和問題

鄰接關係關閉

如果BGP會話已關閉但未啟動，請發出 show ip bgp all summary command. 您可以在此處找到作業階段的目前狀態：

- 如果會話處於NOT up狀態，則狀態可能在IDLE和ACTIVE之間變化（取決於有限狀態機進程）。
- 如果會話處於開啟狀態，您將看到收到的字首數。

```
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
```

無連線

必須保證的第一個要求是兩個對等體之間的連線，這樣可以在埠179上建立TCP會話，無論它們是否直接連線。簡單ping命令對此事很有用。如果在環回介面之間建立了對等，則必須執行環回至環回ping。如果在未將特定環回作為源介面的情況下執行ping測試，則將傳出物理介面IP地址用作資料包的源IP地址，而不是路由器的環回IP地址。

如果ping不成功，請考慮以下原因：

- 沒有連線的路由對等體或根本沒有路由： show ip route *peer_IP_address* 可以使用。
- 第1層問題：需要考慮物理介面、SFP（聯結器）、電纜或外部問題（傳輸和提供商，如果適用）。
- 檢查可以阻止連線的任何防火牆或訪問清單。

如果ping成功，請考慮以下情況：

配置問題

- IP地址或AS配置錯誤：IP錯誤 地址，則不顯示此類消息，但請確保完成正確的配置。對於錯誤的AS，您必須看到這樣的消息： show logging 指令。

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

檢查兩端的BGP配置以更正AS編號或對等IP地址。

- 重複的路由器ID:

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

檢查兩端的BGP識別符號，通過 show ip bgp all summary 並更正重複問題，這可以通過使用全域性命令手動實現 bgp router-id X.X.X.X 在bgp路由器配置下。最佳作法是手動將路由器ID設定為唯一編號。

- BGP來源和TTL:

大多數iBGP會話是通過通過IGP可訪問的環回介面配置的。此環回介面必須顯式定義為源，請使用命令執行此操作 neighbor *ip-address* update-source *interface-id*．

對於eBGP對等體，直接連線的介面通常用於對等操作，並且會檢查Cisco IOS/Cisco IOS-XE是否達到此目的或確實達到此目的 甚至不嘗試建立會話。如果在直接連線的路由器上嘗試將eBGP從環回改為環回，則可以通過禁用兩端的特定鄰居的此檢查 neighbor *ip-address* disable-connected-check．

但是，如果eBGP對等體之間存在多個躍點，則需要正確的躍點計數，請確保 **neighbor *ip-address* ebgp-multihop [*hop-count*]** 配置了正確的跳數，因此可以建立會話。

如果未指定跳數，則iBGP會話的預設TTL值為255，而eBGP會話的預設TTL值為1。

TCP作業階段問題

測試連線埠179的有用操作是手動從另一個對等點進行telnet:

```
R1#telnet 198.51.100.2 179
Trying 198.51.100.2, 179 ... Open

[Connection to 198.51.100.2 closed by foreign host]
```

開啟/連線關閉或遠端主機拒絕的連線表示資料包到達遠端端，然後，確保遠端的控制平面沒有問題。否則，如果存在無法到達目的地址，請檢查可以阻止TCP埠179或BGP資料包或路徑上任何資料包丟失的任何防火牆或訪問清單。

如果驗證出現問題，您可以看到以下訊息：

```
%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
```

檢查身份驗證方法、密碼和相關配置，並進一步排除故障，請參閱BGP對等體之間的MD5身份驗證配置示例。

如果TCP會話未啟動，可以使用以下命令進行隔離：

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

鄰接回退

如果會話處於開啟和關閉狀態，請查詢 show log 我們可以看到一些場景。

介面翻動

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

如消息所示，此故障的原因是介面關閉情況，請查詢埠/SFP、電纜或斷開連線上的任何物理問題。

保留計時器已過期

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

這種情況非常常見；它表示路由器在保持計時器過期前未收到或處理keepalive消息或任何更新消息。裝置傳送通知消息並關閉會話。此問題最常見的原因如下：

- **介面問題：在**兩個對等點的連線介面上尋找任何輸入錯誤、輸入佇列捨棄或實體問題；show interface 可用於此目的。
- **傳輸中的資料包丟失**：有時，Hello資料包可能會在傳輸過程中被丟棄，這是確保在介面級別捕獲資料包的最佳方法。 您可以在Cisco IOS和Cisco IOS-XE裝置上使用內嵌式封包擷取。 如果在介面級別看到資料包，我們需要確保它們到達控制平面(EPC) 在控制平面或 debug bgp [vrf name] ipv4 unicast keepalives 非常有用。
- **高CPU：高**CPU狀況可能導致控制平面上的丟棄， show processes cpu [sorted|history] 對於發現問題非常有用。基於該平台，您可以使用CPU Reference文檔找到進行故障排除的下一步驟
- **CoPP策略問題：**故障排除方法因平台而異，超出本文檔的範圍。

- **MTU不相符**：如果路徑中存在MTU差異，且如果在從來源到目的地的路徑中封鎖了ICMP訊息，PMTUD無法運作且可能會導致作業階段翻動。傳送更新時，會使用交涉的MSS值和DF位元設定。如果路徑中的裝置或甚至目的地無法接受具有更高MTU的封包，便會將ICMP錯誤訊息傳送回BGP發言者。目的地路由器會等待BGP keepalive或BGP更新封包更新其抑制計時器。 您可以檢查與交涉的MSS show ip bgp neighbors *ip_address*.

對已設定df的特定鄰居執行Ping測試可以顯示這類MTU是否沿路徑有效：

```
ping 198.51.100.2 size max_seg_size df
```

如果發現MTU問題，則必須準確檢視配置，以確保整個網路中的MTU值一致。

注意：有關MTU的詳細資訊，請參閱使用MTU的BGP鄰居翻板故障排除。

AFI/SAFI問題

```
%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 000000
```

地址系列識別符號(AFI)是由多協定BGP(MP-BGP)新增的功能擴展，它涉及特定網路協定，例如IPv4、IPv6等，並通過後續地址系列識別符號(SAFI)，例如單播和組播，來增加粒度。MBGP透過BGP路徑屬性(PAs)MP_REACH_NLRI和MP_UNREACH_NLRI實現此分離。這些屬性攜帶在BGP更新消息中，用於攜帶不同地址系列的網路可達性資訊。

該消息為您提供了IANA註冊的這些AFI/SAFI的編號：

- IANA地址系列號
- 後續地址系列識別符號(SAFI)引數

- 檢查兩端的地址系列的BGP配置，以更正任何不想要的地址系列。
- 使用 neighbor *ip-address* dont-capability-negotiate 兩端。有關詳細資訊，請參閱不支援的功能導致BGP對等體故障。

路徑安裝和選擇

如需有關BGP如何運作並選擇最佳路徑的更好說明，請參閱BGP最佳路徑選取演演算法。

下一跳

對於要安裝到路由表中的路由，需要到達下一跳，否則，即使字首位於我們的Loc-RIB BGP表中，它也不會進入RIB。 作為回圈避免規則，在Cisco IOS/Cisco IOS-XE上，iBGP不會變更下一個躍點屬性，且在eBGP重寫下一個躍點並預置其AS_PATH時只保留AS_PATH。

您可以通過以下方式檢查下一跳： show ip bgp [prefix]，則會為您提供下一跳和無法訪問的單詞。在示例中，這是R1通過eBGP通告到R2的字首，由R3通過R2的iBGP連線獲取該字首。

```
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
```

在輸出中，下一跳是R3不知道的R1的傳出介面。為了解決此問題，您可通過IGP、靜態路由或使用neighbor *ip-address next-hop-self* 命令修改iBGP對等路由器上的下一跳IP（直接連線）。在圖示示例中，此配置需要在R2上；指向R3的鄰居(neighbor 10.0.23.3 next-hop-self)。

因此，下一跳會改變(在 clear ip bgp 10.0.23.2 soft)連線到直連介面（可訪問），且已安裝字首。

```
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

當無法將路由安裝到全域性RIB中時會發生這種情況，這會導致RIB故障。常見的原因是，對於管理距離較小的另一個路由協定，RIB上已經存在相同的字首，但使用show ip bgp rib-failure命令可以看到RIB故障的確切原因。有關更詳細的說明，請參閱以下連結：

> **注意**：您可以按照瞭解BGP RIB-failure和bgp suppress-inactive命令中的說明來識別和糾正此類問題。

競爭條件

最常見的問題是，在相互重分發的情況下，IGP優先於eBGP。將IGP路由重新分發到BGP中時，BGP會將其視為本地生成，預設情況下會獲得32768的權重。預設情況下，從BGP對等體接收的所有字首都分配了本地權重0。因此，如果必須比較相同的字首，則根據BGP最佳路徑選擇過程將在路由表中安裝權重較高的字首，這就是在RIB上安裝IGP路由的原因。

此問題的解決方法是將路由器bgp配置下從BGP對等點收到的所有路由的權重設定為更高：

```
neighbor ip-address weight 40000
```

> **注意**：有關詳細說明，請參閱瞭解網路故障轉移場景中BGP權重路徑屬性的重要性。

其他問題

BGP慢速對等點

此對等體無法跟上傳送方生成更新消息的速率。對等體出現此問題的原因有多種：其中一個對等體的CPU高、鏈路上的流量過剩或流量丟失、頻寬資源等。

注意：要幫助識別和糾正慢速對等體問題，請參閱使用BGP「慢速對等體」功能解決慢速對等體問題。

記憶體問題

BGP使用分配給Cisco IOS進程的記憶體來維護網路字首、最佳路徑、策略和所有相關的配置，以正常運行。通過命令檢視整體流程 show processes memory sorted:

```
R1#show processes memory sorted
Processor Pool Total: 2121414332 Used: 255911152 Free: 1865503180
reserve P Pool Total:     102404 Used:         88 Free:     102316
 lsmpi_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*
```

處理器池是使用的記憶體；在本例中大約為2.1 GB。接下來，我們必須檢視「暫掛」列，以確定暫掛大部分子流程的子流程。 然後，我們需要檢查我們擁有的BGP會話、已接收的路由數以及使用的配置。

減少BGP記憶體佔用的常見步驟：

- **BGP過濾：如**果不需要接收完整的BGP表，請使用策略過濾路由並僅安裝所需的字首。
- **軟重新配置**：在BGP配置下查詢**neighbor *ip_address* soft-reconfiguration inbound**；此命令允許您檢視在任何入站策略(Adj-RIB-in)之前收到的所有字首。但是，此表需要大約一半的當前BGP本地RIB表來儲存此資訊，因此您可以避免此配置，除非是強制要求的，或者您當前的字首很少。

  註：有關如何最佳化BGP的更多資訊，請參閱配置BGP路由器以獲得最佳效能和減少的記憶體消耗。

高CPU

路由器為BGP運行使用不同的進程。要驗證BGP進程是CPU使用率較高的原因，請使用 show process cpu sorted 指令。

```
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
 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 RO 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
```

以下是克服由於BGP而導致的CPU使用率較高的常見進程、原因和一般步驟：

- **BGP路由器：每**秒運行一次以保護更快的收斂。它是最重要的執行緒之一，它讀取bgp更新消息，驗證字首/網路和屬性，更新每個AFI/SAFI網路/字首表和屬性表，執行最佳路徑計算以及其他許多工。
  大規模路由變動是導致這種情況的一種非常普遍的情況。
- **BGP掃描程式**：預設情況下每60秒運行一次的低優先順序進程。此程式會檢查整個BGP表以驗證下一個躍點的可達性，並在路徑發生任何變更時相應地更新BGP表。它通過路由資訊庫(RIB)進行重分發。
  檢查平台規模，隨著安裝更多字首和路由以及TCAM的使用，需要更多資源，並且通常會出現裝置超載的情況。

  註：有關如何對這兩個進程進行故障排除的詳細資訊，請參閱對由BGP掃描程式或路由器進程導致的CPU使用率過高進行故障排除。

- **BGP I/O：在接**收BGP控制資料包時運行，並管理BGP資料包的排隊和處理。如果BGP隊列中長期接收的資料包過多，或者如果TCP出現問題，則由於BGP I/O進程，路由器會顯示高CPU的症狀。（通常，在這種情況下，BGP路由器也很高。）檢視消息計數以標識對等裝置，並捕獲資料包以標識這些消息的來源。）
- **BGP開啟：建立**會話時使用的進程。除非會話停滯在「開啟」狀態，否則不會出現常見的高CPU問題。
- **BGP事件**：負責下一跳處理。在接收的字首上查詢下一跳擺動。

相關資訊

- 技術支援與文件 - Cisco Systems
- BGP配置指南
- BGP對等體之間的MD5身份驗證配置示例
- 嵌入式封包擷取
- 使用MTU的BGP鄰居翻動疑難排解
- IANA地址系列號
- 後續地址系列識別符號(SAFI)引數
- 不支援的功能導致BGP對等體故障
- BGP 最佳路徑選取演算法
- 瞭解BGP RIB-failure和bgp suppress-inactive指令
- 瞭解在網路容錯移轉案例中， BGP 加權路徑屬性的重要性
- 使用BGP「慢速對等體」功能解決慢速對等體問題
- 配置BGP路由器以獲得最佳效能並減少記憶體消耗
- 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

關於此翻譯

思科已使用電腦和人工技術翻譯本文件，讓全世界的使用者能夠以自己的語言理解支援內容。請注意，即使是最佳機器翻譯，也不如專業譯者翻譯的內容準確。Cisco Systems, Inc. 對這些翻譯的準確度概不負責，並建議一律查看原始英文文件（提供連結）。