排除C8000v路由器的效能問題

目錄

簡介

本文檔介紹如何解決C8000v企業路由器在公共雲和ESXi方案中的效能問題。

採用元件

本檔案中的資訊是根據以下硬體和軟體元件：

- 執行17.12版的C8000v
- ESXI版本7.0 U3

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。

一般疑難排解

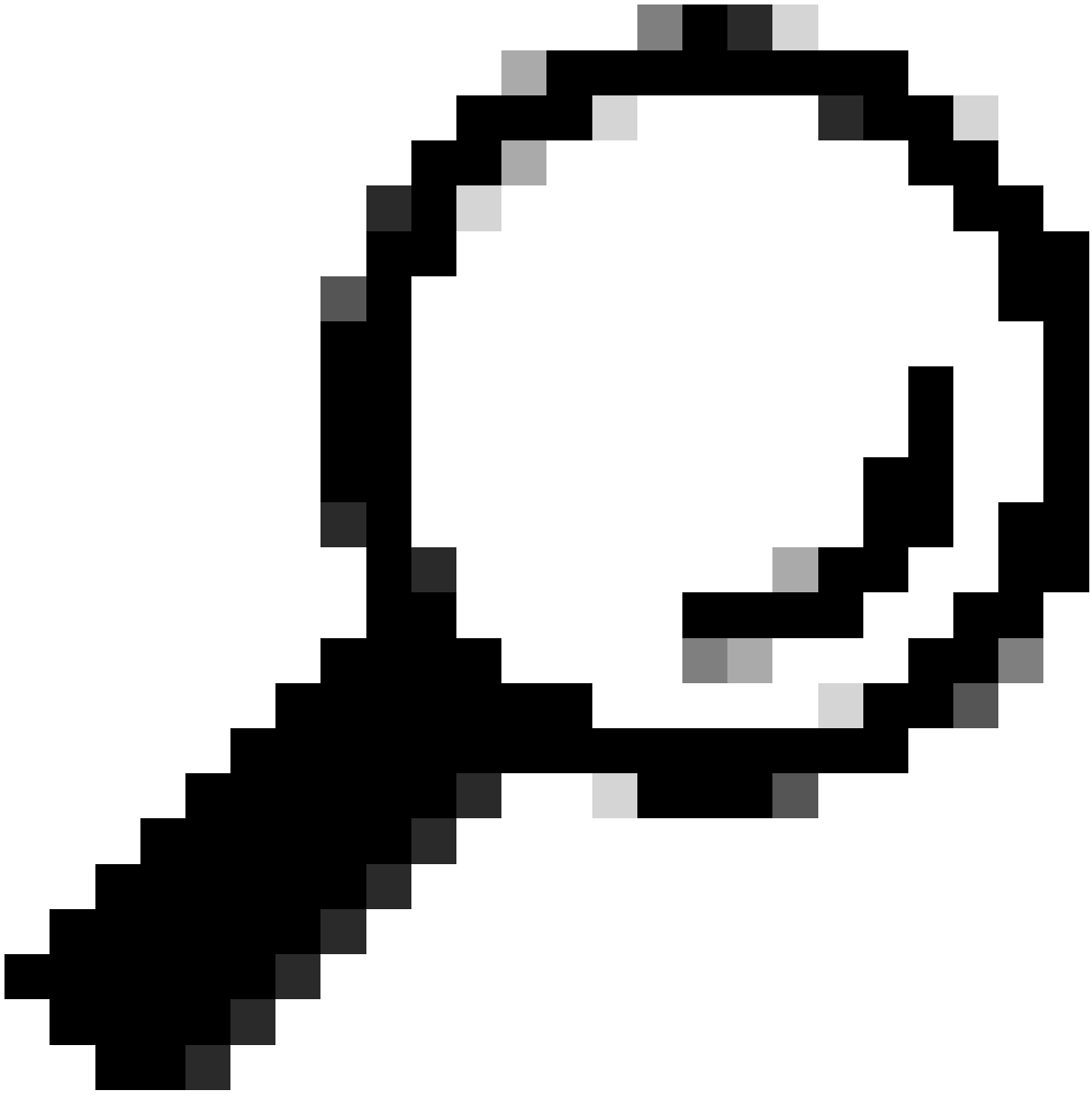
雖然您可以在不同的環境中託管C8000v，但無論託管C8000v的位置，您仍然可以完成一些相同的故障排除步驟。讓我們從基本知識開始。首先需要確認的是裝置是否達到其容量限制。為此，您可以先檢查以下兩個輸出：

1.show platform hardware qfp active datapath util summary — 此命令為您提供C8000v從每個埠接收和傳輸的輸入/輸出的完整資訊。您必須將注意力集中在處理負荷百分比上。如果您處於達到100%的情況，則意味著您將達到容量限制

```
------------------ show platform hardware qfp active datapath utilization summary  ------------------

  CPP 0:                     5 secs        1 min        5 min       60 min
Input:     Total (pps)        93119        92938        65941        65131
                 (bps)    997875976   1000204000    708234904    699462016
Output:    Total (pps)        93119        92949        65944        65131
                 (bps)   1052264704   1054733128    746744264    737395744
Processing: Load (pct)           14           14           10           10
```

2.show platform hardware qfp active datapath infrastructure sw-cio — 將此命令視為上面命令的一個更深入的版本。它提供了有關各個核心（包括IO和加密核心，它們不是QFP利用率的一部分）的更詳細的資訊。在您希望檢視特定資料平面核心是否導致瓶頸的場景中，它非常有用。



提示：使用此命令時的一個非常重要的詳細資訊，請始終運行兩次。計算執行命令之間的

核心資源利用率百分比時。

```
------------------ show platform hardware qfp active datapath infrastructure sw-cio  ------------------

Credits Usage:

  ID      Port  Wght  Global WRKR0  WRKR1  WRKR2  WRKR3  WRKR4  WRKR5  WRKR6  WRKR7  WRKR8  WRKR9  WRKR
   1      rcl0    16:    492      0      0      0      0      0      0      0      0      0
   1      rcl0    32:    496      0      0      0      0      0      0      0      0      0
   2       ipc     1:    489      0      0      0      0      0      0      0      0      0
   3 vxe_punti     4:    490      0      0      0      0      0      0      0      0      0
   4       Gi1     4:   1999      0      0      0      0      0      0      0      0      0
   5       Gi2     4:   1991      0      0      0      0      0      0      0      0      0
   6       Gi3     4:   1991      0      0      0      0      0      0      0      0      0
   7       Gi4     4:   1993      0      0      0      0      0      0      0      0      0
   8       Gi5     4:   2009      0      0      0      0      0      0      0      0      0
   9       Gi6     4:   2015      0      0      0      0      0      0      0      0      0
  10       Gi7     4:   2002      0      0      0      0      0      0      0      0      0
  11      vpg0   400:    490      0      0      0      0      0      0      0      0      0

Core Utilization over preceding 107352.2729 seconds
-------------------------------------------------
      ID:       0       1       2       3       4       5       6       7       8       9      10
    % PP:    2.98    2.01    1.81    1.67    1.60    1.53    1.35    1.30    1.25    1.19    2.19    1.
    % RX:    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.
    % TM:    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.00    0.
  % IDLE:   97.02   97.99   98.19   98.33   98.40   98.47   98.65   98.70   98.75   98.81   97.81   98.
```

現在，您已確定您是否達到平台限制。下一步是檢查丟包。這些與效能問題有內在聯絡。根據發生位置的不同，可以考慮三種型別的丟包。

- 超支：這種型別的丟包發生在Rx端。之所以發生這種情況，是因為已超出一個或多個核心的處理能力。
- 功能丟棄：這種型別的丟包發生在PPE中。它們與路由器中的功能（例如ACL或QoS）相關。
- Taildrops:這種型別的丟包發生在Tx端。它們是由於Tx緩衝區的擁塞引起的。

要確定您正在經歷的丟棄，可以使用以下輸出：

- show platform hardware qfp active drop state clear
- 顯示介面
- show policy map interface

您將驗證如何確定您面臨哪些丟包以及如何緩解這些丟包。不過，本文的最大關注點將是被稱為Taildrops的丟包，因為它們對於在虛擬路由器中進行故障排除尤其棘手。

超支

當網路介面接收封包的速度超過其處理或儲存封包的緩衝區時，Cisco IOS XE會發生溢位捨棄。具體而言，介面（FIFO隊列）的內部緩衝區會變滿，因為傳入的資料速率超過了硬體處理它的能力。結果，新的傳入資料包無法儲存並被丟棄，從而增加了超限計數器。這基本上是介面暫時不堪重負而導致的資料包丟失。

這種型別的丟包發生在Rx端。發生這種情況是因為已超出一個或多個核心的處理能力，並且Rx執行緒無法將傳入資料包分發到相關PP執行緒，並且入口緩衝區已滿。打個簡單的比方，你可以把它想象成在收銀檯排隊的隊，由於封包到達的速度比收銀員（介面硬體）所能提供的速度要快，因此收銀檯排的隊太滿了。當隊伍排滿時，新顧客必須離開而得不到服務 — 這些是超支的投放。

儘管本節中提到了硬體，但C8000v是基於軟體的路由器。在這種情況下，超支可能是由以下原因造成的：

- 資料平面利用率高：如果資料平面利用率高，則無法足夠快地輪詢資料包，從而導致超載。例如，「大象流」（大型、連續的資料流）的存在可能會使處理資源飽和，並導致介面上發生溢位。
- 裝置模板不正確：使用不適當的裝置模板會導致低效的緩衝區管理和溢位。可以使用show platform software cpu alloc命令檢查此情況，也可以使用platform resource <template>命令進行更改。

每個介面都分配了一個有限的信用池，此機制可防止介面繁忙以及系統資源過載。每次新資料包到達資料平面時，都需要信用額度。資料包處理完成後，將返回積分，以便Rx執行緒可以再次使用它。如果沒有可用於介面Rx環中資料包需要等待的介面的信用。通常，由於已超過一個或多個核心的處理能力，您預計與丟棄相關的效能限制將為Rx溢位。

要識別溢位，通常需要檢查介面統計資訊中的輸入錯誤，特別是溢位計數器：

- 使用命令show platform hardware qfp active datapath infrastructure sw-cio確定核心利用率，以及特定介面的信用數量是否已超過。
- 使用命令show interface <interface-name>並在輸出中查詢溢位計數。

超限顯示為輸入錯誤的一部分，例如：

```
Gig2 is up, line protocol is up
241302424557 packets input, 168997587698686 bytes, 0 no buffer
20150 input errors, 0 CRC, 0 frame, 20150 overrun, 0 ignored <<<<<<<<<<<
```

讓我們假設Gig2觀察由超限引起的效能問題的情況。要確定與此介面關聯的工作執行緒，您可以使用以下命令：

```
#show platform hardware qfp active datapath infra binding
Port Instance Bindings:

ID Port IOS Port WRKR 2
1 rcl0 rcl0 Rx+Tx
2 ipc ipc Rx+Tx
3 vxe_punti vxe_puntif Rx+Tx
4 Gi1 GigabitEthernet1 Rx+Tx
5 Gi2 GigabitEthernet2 Rx+Tx <<< in this case, WRKR2 is the thread responsible for both Rx and Tx
```

然後，您可以分析負責此介面的Rx流量的特定執行緒的利用率及其信用數。

在Gig2觀察到由於超限導致的效能問題的情況下，可以預期PP#2持續地充分利用（空閒= 0%），並且介面Gig2的信用值為低/零：

```
#show platform hardware qfp active datapath infrastructure sw-cio
Credits Usage:

ID Port Wght Global WRKR0 WRKR1 WRKR2 Total
1 rcl0 16: 487 0 0 25 512
1 rcl0 32: 496 0 0 16 512
2 ipc 1: 490 0 0 21 511
3 vxe_punti 4: 459 0 0 53 512
4 Gi1 4: 477 0 0 35 512
5 Gi2 4: 474 0 0 38 512 <<< low/zero credits for interface Gig2:

Core Utilization over preceding 1.0047 seconds
-----------------------------------------------
ID: 0 1 2
% PP: 0.77 0.00 0.00
% RX: 0.00 0.00 0.44
% TM: 0.00 0.00 5.63
% IDLE: 99.23 99.72 93.93 <<< the core ID relevant in this case would be PP#2
```

功能丟棄

資料包由任何可用資料平面執行緒處理，並且嚴格基於QFP核心的可用性通過軟體Rx函式(x86) — 基於負載的分配(LBD)進行分配。到達PPE的資料包可以丟棄具有特定QFP丟棄原因的資料包，可以使用以下輸出進行檢查：

```
#show drops
------------------ show platform hardware qfp active statistics drop detail ------------------

Last clearing of QFP drops statistics : never

-------------------------------------------------------------------------------
   ID  Global Drop Stats                          Packets                Octets
-------------------------------------------------------------------------------
  319  BFDoffload                                     403                 31434
  139  Disabled                                       105                  7487
   61  Icmp                                           135                  5994
   94  Ipv4NoAdj                                        1                   193
   33  Ipv6NoRoute                                   2426                135856
  215  UnconfiguredIpv4Fia                        1937573             353562196
  216  UnconfiguredIpv6Fia                        8046173            1057866418


------------------ show platform hardware qfp active interface all statistics drop_summary ------------

-----------------------------------------------------------------
Drop Stats Summary:
note: 1) these drop stats are only updated when PAL
```

```
         reads the interface stats.
      2) the interface stats include the subinterface

Interface                                          Rx Pkts            Tx Pkts
--------------------------------------------------------------------------------
GigabitEthernet1                                   9980371                  0
GigabitEthernet2                                      4012                  0
```

下降的原因多種多樣，通常不言自明。為了進一步研究，可以使用封包追蹤軌跡。

Taildrops

如前所述，當裝置嘗試傳輸資料包但傳輸緩衝區已滿時，就會發生尾部丟棄。

在此小節中，您將看到在遇到此類情況時可以檢查的輸出。您能從中看到的價值意味著什麼，以及您可以做些什麼來緩解這個問題。

首先，你需要知道如何識別它們。其中一種方式是隻檢視show interface。留意任何輸出滴數的增加：

```
GigabitEthernet2 is up, line protocol is up
Hardware is vNIC, address is 0050.56ad.c777 (bia 0050.56ad.c777)
Description: Connected-To-ASR Cloud Gateway
Internet address is 10.6.255.81/29
MTU 1500 bytes, BW 10000000 Kbit/sec, DLY 10 usec,
reliability 255/255, txload 2/255, rxload 3/255
Encapsulation ARPA, loopback not set
Keepalive set (10 sec)
Full Duplex, 10000Mbps, link type is force-up, media type is Virtual
output flow-control is unsupported, input flow-control is unsupported
ARP type: ARPA, ARP Timeout 04:00:00
Last input 00:00:00, output 00:00:00, output hang never
Last clearing of "show interface" counters 03:16:21
Input queue: 0/375/0/0 (size/max/drops/flushes); Total output drops: 7982350 <<<<<<<<
Queueing strategy: fifo
Output queue: 0/40 (size/max)
5 minute input rate 150449000 bits/sec, 20461 packets/sec
5 minute output rate 89116000 bits/sec, 18976 packets/sec
```

此命令對於瞭解您是否遇到擁塞尤其有用：

- show platform hardware qfp active datapath infrastructure - HQF代表「Hierarchical Queueing Framework」。此功能可使用模組化QoS命令列介面(MQC)實現不同級別（物理、邏輯和類）的服務品質(QoS)管理。 它顯示當前的RX和TX成本。 TX隊列已滿時，輸出顯示（1959年已滿）

```
pmd b1689fc0 device Gi1
RX: pkts 5663120 bytes 1621226335 return 0 badlen 0
Out-of-credits: Hi 0 Lo 0
```

```
pkts/burst 1 cycl/pkt 1565 ext_cycl/pkt 1173
Total ring read 12112962299, empty 12107695202
TX: pkts 8047873582 bytes 11241140363740
pri-0: pkts 8047873582 bytes 11241140363740
pkts/send 3
Total: pkts/send 3 cycl/pkt 452
send 2013612969 sendnow 1810842
forced 2013274797 poll 724781 thd_poll 0
blocked 2197451 retries 7401 mbuf alloc err 0
TX Queue 0: full 1959 current index 0 hiwater 224
```

輸出表明底層硬體跟不上資料包的傳送速度。要調試基礎介面，可能需要檢視C8000v外部和運行C8000v的基礎環境，以檢視基礎物理介面上是否報告任何其他錯誤。

為了檢查環境，在檢查C8000v路由器運行哪個虛擬機器監控程式之前，您可以執行一個步驟。這是為了檢查show controller指令輸出。然而，你可能會迷失在每一個櫃檯的意義或去哪裡看東西上。

首先，檢視此輸出時需要牢記的一個重要細節是，資訊主要源自vNIC本身。每個NIC驅動程式都有其使用的特定計數器集，這些計數器集因驅動程式而異。不同的虛擬機器管理程式對顯示的內容也有一定的影響。某些計數器（如mbuf計數器）是DPDK驅動程式的統計資訊。不同的DPDK驅動程式可能有所不同。實際計數通常由虛擬化層的虛擬機器監控程式完成。

```
GigabitEthernet2 - Gi2 is mapped to UIO on VXE
rx_good_packets 1590
tx_good_packets 1402515
rx_good_bytes 202860
tx_good_bytes 1857203911
rx_missed_errors 0
rx_errors 0
tx_errors 0
rx_mbuf_allocation_errors 0
rx_q0_packets 1590
rx_q0_bytes 202860
rx_q0_errors 0
tx_q0_packets 1402515
tx_q0_bytes 1857203911
rx_q0_drop_total 0
rx_q0_drop_err 0
rx_q0_drop_fcs 0
rx_q0_rx_buf_alloc_failure 0
tx_q0_drop_total 976999540797
tx_q0_drop_too_many_segs 0
tx_q0_drop_tso 0
tx_q0_tx_ring_full 30901211518
```

請花幾分鐘時間瞭解如何解釋和閱讀這些計數器：

1. 如果看到subX，則表示它是子介面 — 主介面的邏輯劃分。sub0通常是主/預設值。當涉及多個VLAN時，通常使用這些功能。
2. 然後是「rx = receiving」和「tx = transmitting」。
3. 最後， q0是指該介面使用的第一個/預設隊列

雖然沒有對每個計數器的描述，但文章介紹了其中一些可能與您的故障排除有關的計數器：

- 當NIC緩衝區(Rx FIFO)過滿時，出現「RX_MISSED_ERRORS：」。這種情況會導致丟包和延遲增加。一個可能的解決方案是增加NIC緩衝區（在我們的情況下是不可能的）或更改NIC驅動程式。
- "tx_q0_drop_total"和"tx_q0_tx_ring_full":這可以告訴您主機正在丟棄資料包，並且C8000v在C8000v中遇到尾部丟棄，因為主機正在對C8000v進行反壓

在上方輸出中，我們沒有看到任何「rx_missed_errors」。但是，由於我們關注的是尾滴，因此我們確實看到「tx_q0_drop_total」和「tx_q0_tx_ring_full」。 這樣，我們可以斷定，確實存在由主機底層硬體導致的擁塞。

如前所述，每個虛擬機器監控程式都會對顯示的內容產生某種影響。在下一部分中，我們將重點介紹在C8000v可以託管到的不同虛擬機器監控程式之間的差異。您也可以找到不同的建議，嘗試緩解每一次此類問題。

虛擬機器監控程式

hypervisor是一種軟體層，它通過管理和分配硬體資源（如CPU、記憶體和儲存）到每個VM，允許多個作業系統（稱為虛擬機器或虛擬機器）在單個物理硬體主機上運行。它確保這些虛擬機器彼此獨立運行而不會相互干擾。

在Cisco Catalyst 8000V(C8000v)環境中，虛擬機器監控程式是託管C8000v虛擬機器的平台。如何確定託管C8000v的虛擬機器監控程式？有一個相當有用的輸出為我們提供了這些資訊。此外，您還可以檢查我們的虛擬路由器可以訪問的資源型別：

```
C8000v#show platform software system all
Processor Details
=================
Number of Processors : 8
Processor : 1 - 8
vendor_id : GenuineIntel
cpu MHz : 2593.906
cache size : 36608 KB
Crypto Supported : Yes
model name : Intel(R) Xeon(R) Platinum 8272CL CPU @ 2.60GHz

Memory Details
==============
Physical Memory : 32817356KB

VNIC Details
============
Name Mac Address Driver Name Status Platform MTU
GigabitEthernet1 0022.480d.7a05 net_netvsc UP 1500
GigabitEthernet2 6045.bd69.83a0 net_netvsc UP 1500
GigabitEthernet3 6045.bd69.8042 net_netvsc UP 1500

Hypervisor Details
==================
Hypervisor: AZURE
Manufacturer: Microsoft Corporation
```

```
Product Name: Virtual Machine
Serial Number: 0000-0002-0201-5310-5478-4052-71
UUID: 8b06091c-f1d3-974c-85a5-a78dfb551bf2
Image Variant: None
```

VMware ESXi

ESXi是由VMware開發的一種第1類虛擬機器監控程式，它直接安裝在物理伺服器上以實現虛擬化。它通過提取硬體資源並將其分配給每個虛擬機器，使多個虛擬機器(VM)可以在單個物理伺服器上運行。C8000v路由器是其中一個VM。

您可以從出現擁塞的常見場景開始。這一點可以通過檢查tx_q0_tx_ring_full計數器確認：

範例：

```
------------------ show platform software vnic-if interface-mapping ------------------

---------------------------------------------------------------
Interface Name Driver Name Mac Addr
---------------------------------------------------------------
GigabitEthernet3 net_vmxnet3 <-- 0050.5606.2239
GigabitEthernet2 net_vmxnet3 0050.5606.2238
GigabitEthernet1 net_vmxnet3 0050.5606.2237
---------------------------------------------------------------

GigabitEthernet3 - Gi3 is mapped to UIO on VXE
rx_good_packets 99850846
tx_good_packets 24276286
rx_good_bytes 78571263015
tx_good_bytes 14353154897
rx_missed_errors 0
rx_errors 0
tx_errors 0
rx_mbuf_allocation_errors 0
rx_q0packets 99850846
rx_q0bytes 78571263015
rx_q0errors 0
tx_q0packets 24276286
tx_q0bytes 14353154897
rx_q0_drop_total 0
rx_q0_drop_err 0
rx_q0_drop_fcs 0
rx_q0_rx_buf_alloc_failure 0
tx_q0_drop_total 160945155
tx_q0_drop_too_many_segs 0
tx_q0_drop_tso 0
tx_q0_tx_ring_full 5283588 <------
```

當C8000V嘗試通過VMXNET3介面傳送資料包時，會發生這種擁塞。但是，緩衝區環中已經滿是資料包，這些資料包最終會延遲或丟棄。

在這些情況下，這些丟棄發生在我們前面提到的虛擬機器監控程式端。如果符合所有建議，建議與VMware支援人員聯絡，以瞭解網絡卡上發生的情況。

下面是有關如何提高效能的一些建議：

- 使用專用vSwitch和上行鏈路以獲得最佳效能
- 通過將C8000V分配給由自己的物理上行鏈路支援的專用vSwitch，我們可以將其流量與雜訊鄰居隔離，並避免共用資源瓶頸。

有些命令值得從ESXi端檢視。例如，要檢查來自ESXi介面的資料包丟失，可以執行以下操作：

1. 啟用SSH。
2. 使用SSH連線到ESXi。
3. 運行esxtop。
4. 輸入n。

如果虛擬機器的網路驅動程式耗盡Rx緩衝區記憶體，esxtop命令可以顯示虛擬交換機上丟棄的資料包。即使esxtop顯示封包在虛擬交換器上遭捨棄，但實際上這些封包在虛擬交換器和訪客作業系統驅動程式之間遭捨棄。

搜尋%DRPTX和%DRPRX下丟棄的任何資料包：

```
12:34:43pm up 73 days 16:05, 907 worlds, 9 VMs, 53 vCPUs; CPU load average: 0.42, 0.42, 0.42

PORT-ID USED-BY TEAM-PNIC DNAME PKTTX/s MbTX/s PSZTX PKTRX/s MbRX/s PSZRX %DRPTX %DRPRX
67108870 Management n/a vSwitch-to-9200 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
67108872 Shadow of vmnic1 n/a vSwitch-to-9200 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
67108876 vmk1 vmnic1 vSwitch-to-9200 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
67108890 2101719:c8kv-gw-mgmt vmnic1 vSwitch-to-9200 76724.83 792.35 1353.00 16180.39 9.30 75.00 0.00 0
100663305 Management n/a vSwitch-to-Cisc 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
100663307 Shadow of vmnic0 n/a vSwitch-to-Cisc 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
100663309 vmk0 vmnic0 vSwitch-to-Cisc 3.64 0.01 280.00 3.29 0.00 80.00 0.00 0.00
100663310 2100707:gsoaresc-On_Prem vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 2.43 0.00 60.00 0.00 0.00
100663311 2100993:cats-vmanage void vSwitch-to-Cisc 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
100663312 2100993:cats-vmanage void vSwitch-to-Cisc 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
100663313 2100993:cats-vmanage vmnic0 vSwitch-to-Cisc 5.38 0.01 212.00 9.71 0.01 141.00 0.00 0.00
100663314 2101341:cats-vsmart void vSwitch-to-Cisc 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
100663315 2101341:cats-vsmart vmnic0 vSwitch-to-Cisc 2.60 0.00 164.00 6.94 0.01 124.00 0.00 0.00
100663316 2101522:cats-vbond vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 0.00 0.00 0.00 0.00 100.00
100663317 2101522:cats-vbond vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 0.00 0.00 0.00 0.00 100.00
100663318 2101522:cats-vbond vmnic0 vSwitch-to-Cisc 4.33 0.01 174.00 7.80 0.01 162.00 0.00 0.00
100663319 2101522:cats-vbond vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 4.16 0.00 90.00 0.00 0.00
100663320 2101547:gdk-backup vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 3.12 0.00 77.00 0.00 0.00
100663321 2101703:sevvy vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 3.12 0.00 77.00 0.00 0.00
100663323 2101719:c8kv-gw-mgmt vmnic0 vSwitch-to-Cisc 16180.91 9.09 73.00 76755.87 792.44 1353.00 0.00 0
100663324 2137274:telemetry-server vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 3.12 0.00 77.00 0.00 0.00
100663335 2396721:netlab vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 3.12 0.00 77.00 0.00 0.00
2214592519 vmnic1 - vSwitch-to-9200 76727.26 792.38 1353.00 16182.64 9.30 75.00 0.00 0.00
2248146954 vmnic0 - vSwitch-to-Cisc 16189.05 9.32 75.00 76736.97 792.38 1353.00 0.00 0.00
```

此命令列出主機上配置的所有NIC:

```
esxcli network nic list
Name PCI Device Driver Admin Status Link Status Speed Duplex MAC Address MTU Description
------ ------------ ------ ------------ ----------- ----- ------ ----------------- ---- -----------
vmnic0 0000:01:00.0 igbn Up Up 1000 Full fc:99:47:49:c5:0a 1500 Intel(R) I350 Gigabit Network Connectio
vmnic1 0000:01:00.1 igbn Up Up 1000 Full fc:99:47:49:c5:0b 1500 Intel(R) I350 Gigabit Network Connectio
vmnic2 0000:03:00.0 ixgben Up Up 1000 Full a0:36:9f:1c:1f:cc 1500 Intel(R) Ethernet Controller 10 Gigab
vmnic3 0000:03:00.1 ixgben Up Up 1000 Full a0:36:9f:1c:1f:ce 1500 Intel(R) Ethernet Controller 10 Gigab
```

還有一個有用的命令可用於檢查分配給特定VM的vNIC的狀態。

```
esxcli network vm list
World ID Name Num Ports Networks
-------- ---------------- --------- --------
2137274 telemetry-server 1 Cisco Backbone 10.50.25.0/24
2101703 sevvy 1 Cisco Backbone 10.50.25.0/24
2396721 netlab 1 Cisco Backbone 10.50.25.0/24
2101547 gdk-backup 1 Cisco Backbone 10.50.25.0/24
2101522 cats-vbond 4 VPN0, VPN0, VPN0, VPN0
2101719 c8kv-gw-mgmt 2 c8kv-to-9200l, c8kv-to-cisco
2100707 gsoaresc-On_Prem 1 Cisco Backbone 10.50.25.0/24
2100993 cats-vmanage 3 VPN0, VPN0, VPN0
2101341 cats-vsmart 2 VPN0, VPN0
[root@localhost:~]
```

檢視c8kv-gw-mgmt（C8000v虛擬機器），分配了2個網路：

- c8kv-to-9200l
- c8kv到cisco

您可以使用全球ID查詢有關此VM的詳細資訊：

```
[root@localhost:~] esxcli network vm port list -w 2101719
Port ID: 67108890
vSwitch: vSwitch-to-9200L
Portgroup: c8kv-to-9200l
DVPort ID:
MAC Address: 00:0c:29:31:a6:b6
IP Address: 0.0.0.0
Team Uplink: vmnic1
Uplink Port ID: 2214592519
Active Filters:

Port ID: 100663323
vSwitch: vSwitch-to-Cisco
Portgroup: c8kv-to-cisco
DVPort ID:
```

```
MAC Address: 00:0c:29:31:a6:ac
IP Address: 0.0.0.0
Team Uplink: vmnic0 <----
Uplink Port ID: 2248146954
Active Filters:
[root@localhost:~]
```

一旦您獲得了此資訊，您就可以確定vSwitch分配到的網路。

要檢查分配給vSwitch的物理NIC的一些流量統計資訊，可使用以下命令：

```
# esxcli network nic stats get -n <vmnic>
```

此命令顯示接收的資料包、接收的位元組數、丟棄的資料包和接收的錯誤等資訊。這有助於確定網絡卡上是否發生任何丟包。

```
[root@localhost:~] esxcli network nic stats get -n vmnic0
NIC statistics for vmnic0
Packets received: 266984237
Packets sent: 123640666
Bytes received: 166544114308
Bytes sent: 30940114661
Receive packets dropped: 0
Transmit packets dropped: 0
Multicast packets received: 16773454
Broadcast packets received: 36251726
Multicast packets sent: 221108
Broadcast packets sent: 1947649
Total receive errors: 0
Receive length errors: 0
Receive over errors: 0
Receive CRC errors: 0
Receive frame errors: 0
Receive FIFO errors: 0
Receive missed errors: 0
Total transmit errors: 0
Transmit aborted errors: 0
Transmit carrier errors: 0
Transmit FIFO errors: 0
Transmit heartbeat errors: 0
Transmit window errors: 0
```

通過修改主機和虛擬機器上的設定，可以驗證一些可以提高Cisco Catalyst 8000V在ESXi環境中運行的效能的配置：

- 設定虛擬硬體：CPU保留設定設定為最大值。
- 在虛擬硬體中保留所有訪客記憶體：記憶體。

- 從虛擬硬體中選擇VMware Paravirtual:SCSI控制器。
- 從虛擬硬體：網路介面卡：介面卡型別選項，為支援的NIC選擇SR-IOV
- 將General Guest OS Version > VM Options選項設定為Other 3.x或更高版本Linux（64位）。
- 將Advanced Latency Sensitivity下的VM Options選項設定為High。
- 在VM Options > Advanced Edit Configuration下，將「numa.nodeAffinity」新增到與SRIOV NIC相同的NUMA節點
- 啟用虛擬機器監控程式效能設定。
- 通過在支援的物理NIC上啟用SR-IOV來限制vSwitch的開銷。
- 將VM的vCPU配置為與物理NIC運行在同一NUMA節點上。
- 將VM Latency Sensitivity（VM延遲靈敏度）設定為High（高）。

AWS

C8000v通過在Amazon Virtual Private Cloud(VPC)中作為Amazon Machine Image(AMI)啟動，支援在AWS上部署，允許使用者為其網路資源預配AWS雲中邏輯隔離的部分。

多TX佇列

在AWS上運行的C8000v中，一項關鍵功能是使用多TX隊列（多TXQ）。 這些隊列有助於減少內部處理開銷並提高可擴充性。使用多個隊列可以將傳入和傳出資料包分配到正確的虛擬CPU(vCPU)更快速更簡單。

與某些每個vCPU分配RX/TX隊列的系統不同，在C8000v中，這些隊列按介面分配。RX（接收）和TX（傳輸）隊列充當Catalyst 8000V應用程式與AWS基礎設施或硬體之間的連線點，管理網路流量的傳送和接收方式。AWS根據例項型別控制每個介面可用的RX/TX隊列的數量和速度。

要建立多個TX隊列，Catalyst 8000V需要多個介面。當啟用多個TX隊列時，裝置使用基於流5元組（源IP、目標IP、源埠、目標埠和協定）的雜湊方法來保持資料包流的順序。 此雜湊決定每個流使用哪個TX隊列。

使用者可以使用連線到AWS例項的同一物理網路介面卡(NIC)在Catalyst 8000V上建立多個介面。這是通過配置環回介面或新增輔助IP地址完成的。

使用多TXQ時，有多個傳輸隊列來處理傳出流量。在示例中，有12個TX隊列（編號為0到11）。 此設定允許您單獨監視每個隊列，以檢視是否有隊列變滿。

通過檢視輸出，您可以看到TX Queue 8具有很高的「full」計數器(56,406,998)，這意味著其緩衝區經常滿載。其他TX佇列的「full」計數器顯示零，表示它們並未擁塞。

```
Router#show platform hardware qfp active datapath infrastructure sw-cio
pmd b17a2f00 device Gi2
RX: pkts 9525 bytes 1229599 return 0 badlen 0
Out-of-credits: Hi 0 Lo 0
pkts/burst 1 cycl/pkt 560 ext_cycl/pkt 360
Total ring read 117322273, empty 117312792
```

```
TX: pkts 175116324 bytes 246208197526
pri-0: pkts 157 bytes 10238
pkts/send 1
pri-1: pkts 75 bytes 4117
pkts/send 1
pri-2: pkts 91 bytes 6955
pkts/send 1
pri-3: pkts 95 bytes 8021
pkts/send 1
pri-4: pkts 54 bytes 2902
pkts/send 1
pri-5: pkts 75 bytes 4082
pkts/send 1
pri-6: pkts 104 bytes 8571
pkts/send 1
pri-7: pkts 74 bytes 4341
pkts/send 1
pri-8: pkts 175115328 bytes 246208130411
pkts/send 2
pri-9: pkts 85 bytes 7649
pkts/send 1
pri-10: pkts 106 bytes 5784
pkts/send 1
pri-11: pkts 82 bytes 7267
pkts/send 1
Total: pkts/send 2 cycl/pkt 203
send 68548581 sendnow 175024880
forced 1039215617 poll 1155226129 thd_poll 0
blocked 2300918060 retries 68534370 mbuf alloc err 0
TX Queue 0: full 0 current index 0 hiwater 0
TX Queue 1: full 0 current index 0 hiwater 0
TX Queue 2: full 0 current index 0 hiwater 0
TX Queue 3: full 0 current index 0 hiwater 0
TX Queue 4: full 0 current index 0 hiwater 0
TX Queue 5: full 0 current index 0 hiwater 0
TX Queue 6: full 0 current index 0 hiwater 0
TX Queue 7: full 0 current index 0 hiwater 0
TX Queue 8: full 56406998 current index 224 hiwater 224 <<<<<<<<<<
TX Queue 9: full 0 current index 0 hiwater 0
TX Queue 10: full 0 current index 0 hiwater 0
TX Queue 11: full 0 current index 0 hiwater 0
```

監控TX隊列的「滿」計數器有助於識別是否有任何傳輸隊列超載。對特定TX隊列持續增加的「full」計數指向給裝置帶來壓力的流量。解決此問題可能需要流量平衡、調整配置或擴展資源以提高效能。

超出指標

AWS在例項級別設定了某些網路限制，以確保不同例項大小之間一致的高品質網路效能。這些限制有助於為所有使用者保持穩定的網路。

您可以在裝置上使用show controllers命令檢查這些限制和相關統計資訊。輸出包含許多計數器，但此處我們只關注用於監控網路效能的最重要計數器：

```
c8kv-2#sh control | inc exceed
<snipped>
bw_in_allowance_exceeded 0
bw_out_allowance_exceeded 0
pps_allowance_exceeded 0
conntrack_allowance_exceeded 0
linklocal_allowance_exceeded 0
<snipped>
```

現在，您可以深入檢視這些計數器具體指什麼：

- bw_in_allowance_exceeded:由於傳入頻寬超出例項限制而排隊或丟棄的資料包數。
- bw_out_allowance_exceeded:由於傳出頻寬超出例項限制而排隊或丟棄的資料包數。
- pps_allowance_exceeded:由於每秒資料包總數(PPS)超過例項限制而排隊或丟棄的資料包數。
- contrack_allowance_exceeded:跟蹤的連線數達到例項型別允許的最大值。
- linklocal_allowance_exceeded:由於本地代理服務（如Amazon DNS、例項後設資料服務和時間同步服務）的流量超過網路介面的PPS限制而丟棄的資料包數。這不會影響自定義DNS解析程式。

這對您的C8000v效能意味著什麼：

- 如果您注意到這些計數器不斷增加並且遇到效能問題，並不一定意味著C8000v路由器是問題所在。相反，它通常表示您使用的AWS例項已達到其容量限制。您可以檢查AWS例項的規格，以確保其能夠處理您的流量需求。

Microsoft Azure

在本節中，瞭解Microsoft Azure和思科C8000v虛擬路由器如何相結合，在雲中提供可擴展、安全且高效能的虛擬網路解決方案。

瞭解加速網路(AN)和資料包分段如何影響效能。以及檢查對Microsoft Azure使用受支援的例項型別的重要性。

加速網路

在C8000v託管在Microsoft Azure雲中的效能問題中。不能忽略的一個方面是加速網路是否已啟用。因為它極大地提高了路由器的效能。簡而言之，加速網路可在虛擬機器（例如Cisco Catalyst 8000V VM）上實現單一根I/O虛擬化(SR-IOV)。加速的網路路徑會繞過虛擬交換機，提高網路流量的速度，改善網路效能，並降低網路延遲和抖動。

檢查是否已啟用「加速網路」的方法非常簡單。也就是檢查show controllers輸出，並驗證是否有特定計數器存在：

```
------------------ show controllers ------------------

GigabitEthernet1 - Gi1 is mapped to UIO on VXE
  rx_good_packets 6497723453
```

```
tx_good_packets 14690462024
rx_good_bytes 2271904425498
tx_good_bytes 6276731371987
```

```
rx_q0_good_packets 58576251
rx_q0_good_bytes 44254667162
```

```
vf_rx_good_packets 6439147188
vf_tx_good_packets 14690462024
vf_rx_good_bytes 2227649747816
vf_tx_good_bytes 6276731371987
```

您正在尋找的計數器是以vf開頭的計數器，例如vf_rx_good_packets。如果您驗證這些計數器存在，可以絕對確保已啟用加速網路。

Azure和分段

碎片可能會對效能造成負面影響。影響效能的主要原因之一是分段和重組資料包的CPU/記憶體影響。當網路裝置需要將資料包分段時，它必須分配CPU/記憶體資源來執行分段。

重組資料包時也會發生同樣的情況。網路裝置必須儲存所有片段，直到收到這些片段為止，這樣它才能將它們重組為原始資料包。

Azure不會使用加速網路處理分段的資料包。當VM收到分段的資料包時，非加速路徑會對其進行處理。因此，分段的資料包會錯失加速網路帶來的好處，例如更低的延遲、更低的抖動和更高的每秒資料包數。因此，建議儘可能避免分段。

預設情況下，Azure會丟棄以無序方式到達VM的分段資料包，這意味著這些資料包與來自源端點的傳輸序列不匹配。當封包通過Internet或其他大型WAN傳輸時，就可能會發生此問題。

Microsoft Azure支援的例項型別

根據思科標準，C8000v使用支援的例項型別非常重要。請參閱Cisco Catalyst 8000V Edge軟體安裝和設定指南。

這是因為該清單中的例項型別是正確測試C8KV的例項型別。現在，如果C8000v在未列出的例項型別上運行，會出現一個有效的問題。答案很可能是肯定的。但是，當您對像效能問題這樣的複雜問題進行故障排除時，不希望將其他未知因素新增到問題中。僅出於此原因，Cisco TAC始終建議您保留在受支援的例項型別中。

其他資源

只有此時發生了效能問題，才能真正排除故障。然而，這很難理解，因為這可能發生在任何特定時刻。因此，我們提供此EEM指令碼。它有助於在資料包開始遭到丟棄和效能問題出現時捕獲重要輸出：

```
ip access-list extended TAC
permit ip host  host



permit ip host


          host
```

```
conf t
event manager applet CONNECTIONLOST1 authorization bypass
event track 100 state down maxrun 500
action 0010 syslog msg "Logging information to file bootflash:SLA-DROPS.txt and bootflash:FIASLA_Decode
action 0020 cli command "enable"
action 0021 cli command "term length 0"
action 0022 cli command "term exec prompt timestamp"
action 0023 cli command "term exec prompt expand"
action 0095 cli command "show clock | append bootflash:SLA-DROPS.txt"
action 0096 cli command "show platform hardware qfp active statistics drop detail | append bootflash:SL
action 0097 cli command "show logging | append bootflash:SLA-DROPS.txt"
action 0099 cli command "show interfaces summary | append bootflash:SLA-DROPS.txt"
action 0100 cli command "show interfaces | append bootflash:SLA-DROPS.txt"
action 0101 cli command "show platform hardware qfp active statistics drop clear"
action 0102 cli command "debug platform packet-trace packet 2048 fia-trace"
action 0103 cli command "debug platform packet-trace copy packet both"
action 0104 cli command "debug platform condition ipv4 access-list TAC both"
action 0105 cli command "debug platform condition start"
action 0106 cli command "show platform hardware qfp active data infrastructure sw-cio | append bootflas
action 0110 wait 60
action 0111 cli command "debug platform condition stop"
action 0112 cli command "show platform packet-trace packet all decode | append bootflash:FIASLA_Decode.
action 0120 cli command "show platform packet-trace statistics | append bootflash:FIASLA_Decode.txt"
action 0121 cli command "show platform packet-trace summary | append bootflash:FIASLA_Decode.txt"
action 0122 cli command "show platform hardware qfp active datapath utilization summary | append bootfl
action 0123 cli command "show platform hardware qfp active statistics drop detail | append bootflash:SL
action 0124 cli command "show platform hardware qfp active infrastructure bqs queue output default all
action 0125 cli command "show platform software status control-processor brief | append bootflash:SLA-D
action 0126 cli command "show platform hardware qfp active datapath infrastructure sw-pktmem | append b
action 0127 cli command "show platform hardware qfp active infrastructure punt statistics type per-caus
action 0128 cli command "show platform hardware qfp active statistics drop | append bootflash:SLA-DROPS
action 0129 cli command "show platform hardware qfp active infrastructure bqs queue output default all
action 0130 cli command "show platform hardware qfp active data infrastructure sw-hqf config 0 0 | appe
action 0131 cli command "show platform hardware qfp active feature lic-bw oversubscription | append boo
action 0132 cli command "show platform hardware qfp active data infrastructure sw-hqf config 0 0 | appe
action 0133 cli command "show platform hardware qfp active data infrastructure sw-cio | append bootflas
action 0134 cli command "show platform hardware qfp active data infrastructure sw-hqf sched | append bo
```

```
action 0135 cli command "show platform hardware qfp active data infrastructure sw-dist | append bootfla
action 0136 cli command "show platform hardware qfp active data infrastructure sw-nic | append bootflas
action 0137 cli command "show platform hardware qfp active data infrastructure sw-pktmem | append bootf
action 0138 cli command "show controllers | append bootflash:SLA-DROPS.txt"
action 0139 cli command "show platform hardware qfp active datapath pmd controllers | append bootflash:
action 0140 cli command "show platform hardware qfp active datapath pmd system | append bootflash:SLA-D
action 0141 cli command "show platform hardware qfp active datapath pmd static-if-config | append bootf
action 0150 cli command "clear platform condition all"
action 0151 cli command "clear platform packet-trace statistics"
action 0152 cli command "clear platform packet-trace configuration"
action 0153 cli command "show log | append bootflash:througput_levelinfoSLA.txt"
action 0154 cli command "show version | append bootflash:througput_levelinfoSLA.txt"
action 0155 cli command "show platform software system all | append bootflash:througput_levelinfoSLA.tx
action 0156 syslog msg "EEM script and FIA trace completed."
action 0180 cli command "conf t"
action 0181 cli command "no event manager applet CONNECTIONLOST1"
end
```

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）。