Dépannage des problèmes de performances des routeurs C8000v

Table des matières

Introduction

Composants utilisés

Dépannage général

- Dépassements
- Abandons de fonctionnalités
- Gouttes De Queue

Hyperviseurs

- VMware ESXi
- AWS
  - Files d'attente multiémission
  - Mesures dépassées
- Microsoft Azure
  - Mise en réseau accélérée
  - Azure et fragmentation
  - Types d'instance pris en charge pour Microsoft Azure

Ressources supplémentaires

Introduction

Ce document décrit comment dépanner les problèmes de performances dans les routeurs d'entreprise C8000v à travers les clouds publics et les scénarios ESXi.

Composants utilisés

Les informations contenues dans ce document sont basées sur les composants matériels et logiciels suivants :

- C8000v exécutant la version 17.12
- ESXI version 7.0 U3

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. Si votre réseau

est en ligne, assurez-vous de bien comprendre l’incidence possible des commandes.

Dépannage général

Bien que votre C8000v puisse être hébergé dans différents environnements, il existe encore quelques étapes de dépannage que vous pouvez suivre et qui sont identiques, quel que soit l'emplacement où le C8000v est hébergé. Commençons par les bases. La première chose que vous devez vérifier est si le périphérique atteint ses limites de capacité ou non. Pour cela, vous pouvez commencer par vérifier les deux sorties suivantes :

1. show platform hardware qfp active datapath util summary - Cette commande vous donne les informations complètes des entrées/sorties que le C8000v reçoit et transmet depuis chaque port. Vous devez vous concentrer sur le pourcentage de charge de traitement. Si vous êtes dans un scénario où vous atteignez 100 %, cela signifie que vous atteignez la limite de capacité

```
------------------ 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 - Considérez cette commande comme une version plus approfondie de celle ci-dessus. Il fournit de meilleurs détails sur les coeurs individuels, y compris les coeurs d'E/S et de chiffrement qui ne font pas partie du numéro d'utilisation du QFP. Il est très utile dans un scénario où vous voulez voir si un coeur de plan de données spécifique provoque un goulot d'étranglement.



Conseil : Si vous utilisez cette commande, exécutez-la toujours deux fois. Lors du calcul du pourcentage d'utilisation du coeur entre l'exécution de la commande.

```
------------------ 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      0
   5       Gi2      4:    1991     0      0      0      0      0      0      0      0      0      0
   6       Gi3      4:    1991     0      0      0      0      0      0      0      0      0      0
   7       Gi4      4:    1993     0      0      0      0      0      0      0      0      0      0
   8       Gi5      4:    2009     0      0      0      0      0      0      0      0      0      0
   9       Gi6      4:    2015     0      0      0      0      0      0      0      0      0      0
  10       Gi7      4:    2002     0      0      0      0      0      0      0      0      0      0
  11      vpg0    400:     490     0      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.
```

Maintenant, vous avez déterminé si vous atteignez ou non la limite de la plate-forme. L'étape suivante consiste à rechercher les gouttes. Ils sont intrinsèquement liés à des problèmes de performances. Il existe trois types de chutes que vous pouvez envisager en fonction de l'endroit où elles se produisent.

- Dépassements : Ce type de perte de paquets se produit à l'extrémité Rx. Elles se produisent parce que la capacité de traitement d'un ou de plusieurs coeurs a été dépassée.
- Abandons de fonctionnalités : Ce type d'abandon de paquet se produit dans l'EPI. Elles sont liées aux fonctions du routeur, telles qu'une liste de contrôle d'accès ou une qualité de service.
- Taildrops : Ce type d'abandon de paquet se produit à l'extrémité Tx. Ils se produisent en raison de la congestion dans les tampons Tx.

Pour identifier les abandons que vous rencontrez, vous pouvez utiliser les sorties suivantes :

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

Vous vérifiez comment identifier les chutes que vous rencontrez et comment les atténuer. Néanmoins, le point le plus important de cet article sera les chutes connues sous le nom de chutes de queue, car elles sont particulièrement difficiles à dépanner dans les routeurs virtuels.

Dépassements

Une perte de dépassement dans Cisco IOS XE se produit lorsque l'interface réseau reçoit des paquets plus rapidement qu'elle ne peut les traiter ou les stocker dans sa mémoire tampon. Plus précisément, les tampons internes des interfaces (file d'attente FIFO) sont saturés car le débit des

données entrantes dépasse la capacité du matériel à les gérer. Par conséquent, les nouveaux paquets entrants ne peuvent pas être stockés et sont abandonnés, ce qui incrémente le compteur de dépassement. Il s'agit essentiellement d'une perte de paquets provoquée par la saturation temporaire de l'interface.

Ce type de perte de paquets se produit à l'extrémité Rx. Elles se produisent parce que la capacité de traitement d'un ou plusieurs coeurs a été dépassée et que le thread Rx ne peut pas distribuer les paquets entrants au thread PP approprié et que les tampons d'entrée sont déjà pleins. Pour faire une simple analogie, vous pouvez le voir comme une file d'attente à un comptoir de caisse qui devient trop plein parce que les paquets arrivent plus vite que la caisse (matériel d'interface) peut les servir. Lorsque la file d'attente est pleine, les nouveaux clients doivent partir sans être servis. Il s'agit des chutes de débordement.

Bien que le matériel soit mentionné dans cette section, le C8000v est un routeur logiciel. Dans ce cas, les dépassements peuvent être causés par :

- Utilisation élevée du plan de données : Si l'utilisation du plan de données est élevée, les paquets ne peuvent pas être interrogés assez rapidement, ce qui entraîne des dépassements. Par exemple, la présence de « flux éléphants » (flux de données volumineux et continus) peut saturer les ressources de traitement et provoquer des dépassements sur les interfaces.
- Modèle de périphérique incorrect : L'utilisation d'un modèle de périphérique inapproprié peut entraîner une gestion inefficace de la mémoire tampon et des dépassements. Ceci peut être vérifié avec la commande show platform software cpu alloc et il peut être modifié avec la commande platform resource <template>.

Chaque interface reçoit un pool limité de crédits, ce mécanisme empêche une interface occupée et surcharge les ressources système. Chaque fois qu'un nouveau paquet arrive dans le plan de données, un crédit est requis. Lorsque le traitement des paquets est terminé, le crédit est retourné afin que le thread Rx puisse l'utiliser à nouveau. S'il n'y a pas de crédit disponible pour l'interface, le paquet doit attendre dans l'anneau Rx de l'interface. Généralement, vous vous attendez à ce que les pertes liées à la limite de performances soient des dépassements de Rx parce que la capacité de traitement d'un ou plusieurs coeurs a été dépassée.

Pour identifier les dépassements, vous vérifiez généralement les statistiques de l'interface pour les erreurs d'entrée, en particulier le compteur de dépassements :

- Utilisez la commande show platform hardware qfp active datapath infrastructure sw-cio pour identifier l'utilisation principale et si le nombre de crédits pour une interface spécifique a été dépassé.
- Utilisez la commande show interface <interface-name> et recherchez le nombre de dépassements dans le résultat.

Les dépassements sont affichés dans le cadre des erreurs d'entrée, par exemple :

```
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 <<<<<<<<<<<
```

Supposons un cas où Gig2 observe des problèmes de performances causés par des dépassements. Pour déterminer le thread de travail associé à cette interface, vous pouvez utiliser la commande suivante :

```
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 bo
```

Ensuite, vous pouvez analyser l'utilisation du thread spécifique responsable du trafic Rx de cette interface et son nombre de crédits.

Dans un scénario où Gig2 observe des problèmes de performances dus à des dépassements de capacité, vous pouvez vous attendre à ce que le PP#2 soit constamment entièrement utilisé (inactif = 0 %) et à des crédits faibles/nuls pour l'interface Gig2 :

```
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
```

Abandons de fonctionnalités

Les paquets sont gérés par n'importe quel thread de plan de données disponible et sont distribués strictement en fonction de la disponibilité des coeurs QFP via la fonction logicielle Rx (x86) - Load Based Distribution (LBD). Les paquets qui arrivent dans l'PPE peuvent être abandonnés avec une raison d'abandon QFP spécifique, qui peut être vérifiée à l'aide de la sortie suivante :

```
------------------ 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
```

Les raisons de ces chutes sont diverses et sont généralement explicites. Pour approfondir l'étude, une trace de paquet peut être utilisée.

Gouttes De Queue

Comme nous l’avons mentionné précédemment, les abandons arrière se produisent lorsque le périphérique tente de transmettre des paquets, mais que les tampons de transmission sont saturés.

Dans cette sous-section, vous allez examiner les résultats que vous pouvez examiner lorsque vous êtes confronté à ce type de situation. Quelles valeurs vous pouvez voir dans les significations et ce que vous pouvez faire pour atténuer le problème.

Tout d'abord, vous devez savoir comment les identifier. L'une d'elles consiste à simplement regarder l'interface show. Surveillez l'augmentation des gouttes de sortie :

```
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 >>> Output drops.
  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
```

Cette commande est particulièrement utile pour déterminer si vous rencontrez ou non des problèmes d'encombrement :

- show platform hardware qfp active datapath infrastructure - HQF signifie « Hierarchical Queueing Framework ». Il s'agit d'une fonctionnalité qui permet la gestion de la qualité de service (QoS) à différents niveaux (physique, logique et classe) à l'aide de l'interface de ligne de commande (MQC) QoS modulaire. Il indique les coûts RX et TX actuels. Lorsque la file d'attente TX est pleine, comme l'indique le résultat (full 1959)

```
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 >>>
```

Le résultat suggère que le matériel sous-jacent ne suit pas le rythme de l'envoi de paquets. Pour déboguer l'interface sous-jacente, vous devez potentiellement rechercher en dehors de C8000v et dans l'environnement sous-jacent sur lequel le C8000v est en cours d'exécution pour voir si des erreurs supplémentaires sont signalées sur les interfaces physiques sous-jacentes.

Afin de vérifier l'environnement, vous pouvez effectuer une étape avant de vérifier dans quel hyperviseur le routeur C8000v est en cours d'exécution. Ceci est pour vérifier la sortie de la commande show controller. Néanmoins, vous pouvez vous perdre sur ce que chaque comptoir signifie ou où regarder.

Tout d'abord, un détail important que vous devez garder à l'esprit lorsque vous examinez ce résultat est que les informations proviennent principalement des vNIC elles-mêmes. Chaque pilote de carte réseau dispose d'un ensemble spécifique de compteurs qu'il utilise, qui peut varier naturellement en fonction du pilote. Différents hyperviseurs ont également un certain effet sur ce qui est présenté. Certains compteurs, tels que les compteurs mbuf, sont des statistiques du pilote DPDK. Ceux-ci peuvent varier pour différents pilotes DPDK. Le comptage réel est généralement effectué par l'hyperviseur au niveau de la couche de virtualisation.

```
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
```

Prenez une minute ici pour apprendre à interpréter et lire ces compteurs :

1. Si vous voyez subX, cela signifie qu'il s'agit d'une sous-interface - une division logique de l'interface principale. Le sub0 est généralement le port principal/par défaut. Ils sont souvent

utilisés lorsque plusieurs VLAN sont impliqués.
2. Ensuite, vous avez "rx = réception" et "tx = transmission".
3. Enfin, q0 fait référence à la première file d'attente/file d'attente par défaut utilisée par cette interface

Bien qu'il n'y ait pas de description pour chaque compteur, l'article en décrit quelques-uns, qui peuvent être pertinents pour votre dépannage :

- La mention « RX_MISSED_ERRORS: » s'affiche lorsque la mémoire tampon de la carte réseau (Rx FIFO) est saturée. Cette condition entraîne des pertes et une augmentation de la latence. Une solution possible consiste à augmenter la mémoire tampon de la carte réseau (ce qui est impossible dans notre cas) ou à modifier le pilote de la carte réseau.
- "tx_q0_drop_total" et "tx_q0_tx_ring_full" : Ceux-ci peuvent vous indiquer que l'hôte abandonne des paquets, et que le C8000v connaît des pertes de queue dans le C8000v parce que l'hôte fait une contre-pression sur le C8000v

Dans le résultat ci-dessus, nous ne voyons pas de «rx_miss_errors». Cependant, comme nous nous concentrons sur les points de terminaison, nous voyons à la fois "tx_q0_drop_total" et "tx_q0_tx_ring_full". Avec cela, nous pouvons conclure qu'il y a en effet un encombrement causé par le matériel sous-jacent de l'hôte.

Comme indiqué précédemment, chaque hyperviseur a un effet sur ce qui est présenté. L'article se concentre sur ce point dans la section suivante, car il passe en revue les différences entre les différents hyperviseurs où le C8000v peut être hébergé. Vous pouvez également trouver les différentes recommandations pour tenter d'atténuer ce type de problème dans chacune d'entre elles.

Hyperviseurs

Un hyperviseur est une couche logicielle qui permet à plusieurs systèmes d'exploitation (appelés machines virtuelles ou VM) de s'exécuter sur un hôte matériel physique unique en gérant et en allouant les ressources matérielles telles que le processeur, la mémoire et le stockage à chaque VM. Il garantit que ces machines virtuelles fonctionnent indépendamment sans interférer les unes avec les autres.

Dans le contexte de Cisco Catalyst 8000V (C8000v), l'hyperviseur est la plate-forme qui héberge la machine virtuelle C8000v. Comment savoir quel hyperviseur héberge votre C8000v ? Il y a une sortie plutôt utile qui nous donne cette information. En outre, vous pouvez également vérifier à quel type de ressources notre routeur virtuel a accès :

```
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 est un hyperviseur de type 1 développé par VMware et installé directement sur les serveurs physiques pour permettre la virtualisation. Elle permet à plusieurs machines virtuelles (VM) de s'exécuter sur un seul serveur physique en faisant abstraction des ressources matérielles et en les allouant à chaque VM. Le routeur C8000v est l'une de ces machines virtuelles.

Vous pouvez commencer par passer en revue un scénario courant où la congestion se produit. Pour le confirmer, vérifiez le compteur tx_q0_tx_ring_full :

Exemple :

```
------------------ 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 <------
```

Cet encombrement se produit lorsque le C8000V tente d'envoyer des paquets via l'interface VMXNET3. Cependant, l'anneau tampon est déjà plein de paquets, qui se terminent soit par des retards, soit par des pertes.

Dans ces conditions, ces chutes se produisent du côté de l'hyperviseur comme nous l'avons mentionné précédemment. Si toutes les recommandations sont respectées, il est recommandé de vérifier auprès du support VMware pour comprendre ce qui se passe sur la carte réseau.

Voici quelques suggestions sur la façon d'améliorer les performances :

- Utiliser un commutateur virtuel dédié et une liaison ascendante pour des performances optimales
- En attribuant le C8000V à un commutateur virtuel dédié soutenu par sa propre liaison ascendante physique, nous pouvons isoler son trafic des voisins bruyants et éviter les goulots d'étranglement des ressources partagées.

Quelques commandes valent la peine d'être consultées du côté d'ESXi. Par exemple, pour vérifier la perte de paquets à partir de l'interface ESXi, nous pouvons procéder comme suit :

1. Activer le protocole SSH.
2. Connectez-vous à ESXi à l'aide de SSH.
3. Exécutez esxtop.
4. Tapez n.

La commande esxtop peut afficher les paquets abandonnés au niveau du commutateur virtuel si le pilote réseau de la machine virtuelle est à court de mémoire tampon Rx. Même si esxtop affiche les paquets comme abandonnés au niveau du commutateur virtuel, ils sont en fait abandonnés entre le commutateur virtuel et le pilote du système d'exploitation invité.

Recherchez les paquets abandonnés sous %DRPTX et %DRPRX :

```
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
67108870    Management                  n/a         vSwitch-to-9200      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
67108876    vmk1                       vmnic1      vSwitch-to-9200      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
100663305   Management                  n/a         vSwitch-to-Cisc      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
100663309   vmk0                       vmnic0      vSwitch-to-Cisc      3.64     0.01   280.00      3.29
100663310   2100707:gsoaresc-On_Prem   vmnic0      vSwitch-to-Cisc      0.00     0.00     0.00      2.43
100663311   2100993:cats-vmanage       void        vSwitch-to-Cisc      0.00     0.00     0.00      0.00
100663312   2100993:cats-vmanage       void        vSwitch-to-Cisc      0.00     0.00     0.00      0.00
100663313   2100993:cats-vmanage       vmnic0      vSwitch-to-Cisc      5.38     0.01   212.00      9.71
100663314   2101341:cats-vsmart        void        vSwitch-to-Cisc      0.00     0.00     0.00      0.00
100663315   2101341:cats-vsmart        vmnic0      vSwitch-to-Cisc      2.60     0.00   164.00      6.94
100663316   2101522:cats-vbond         vmnic0      vSwitch-to-Cisc      0.00     0.00     0.00      0.00
100663317   2101522:cats-vbond         vmnic0      vSwitch-to-Cisc      0.00     0.00     0.00      0.00
100663318   2101522:cats-vbond         vmnic0      vSwitch-to-Cisc      4.33     0.01   174.00      7.80
100663319   2101522:cats-vbond         vmnic0      vSwitch-to-Cisc      0.00     0.00     0.00      4.16
100663320   2101547:gdk-backup         vmnic0      vSwitch-to-Cisc      0.00     0.00     0.00      3.12
100663321   2101703:sevvy              vmnic0      vSwitch-to-Cisc      0.00     0.00     0.00      3.12
100663323   2101719:c8kv-gw-mgmt       vmnic0      vSwitch-to-Cisc  16180.91     9.09    73.00  76755.87
100663324   2137274:telemetry-server   vmnic0      vSwitch-to-Cisc      0.00     0.00     0.00      3.12
100663335   2396721:netlab             vmnic0      vSwitch-to-Cisc      0.00     0.00     0.00      3.12
2214592519  vmnic1                     -           vSwitch-to-9200  76727.26   792.38  1353.00  16182.64
2248146954  vmnic0                     -           vSwitch-to-Cisc  16189.05     9.32    75.00  76736.97
```

Cette commande répertorie toutes les cartes réseau configurées sur un hôte :

```
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
```

Il existe également une commande utile pour vérifier l'état de la vNIC attribuée à une VM particulière.

```
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
```

Si l'on examine la machine virtuelle c8kv-gw-mgmt , qui est une machine virtuelle C8000v, 2 réseaux sont affectés :

- c8kv-à-9200l
- c8kv-to-cisco

Vous pouvez utiliser l'ID mondial pour rechercher des informations supplémentaires sur cette VM :

```
[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:~]
```

Une fois que vous disposez de ces informations, vous pouvez identifier le réseau auquel le commutateur virtuel est affecté.

Pour vérifier certaines statistiques de trafic de la carte réseau physique attribuée au commutateur virtuel, nous utilisons la commande suivante :

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

Cette commande affiche des informations telles que les paquets reçus, les octets reçus, les paquets abandonnés et les erreurs reçues. Cela peut aider à identifier si des pertes se produisent sur la carte réseau.

```
[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
```

Il y a quelques configurations à vérifier qui peuvent améliorer les performances du Cisco Catalyst 8000V exécuté sur un environnement ESXi en modifiant les paramètres sur l'hôte et la machine virtuelle :

- Définition du matériel virtuel : Le paramètre de réservation du processeur est défini sur Maximum.
- Réservez toute la mémoire des invités dans le matériel virtuel : Mémoire.

- Sélectionnez VMware Paravirtual dans Virtual Hardware : Contrôleur SCSI.
- À partir du matériel virtuel : Carte réseau : Adapter Type, sélectionnez SR-IOV pour les cartes réseau prises en charge
- Définissez l'option General Guest OS Version > VM Options sur Other 3.x ou ultérieur Linux (64 bits).
- Définissez l'option VM Options sous Advanced Latency Sensitivity sur High.
- Sous VM Options > Advanced Edit Configuration, ajoutez « numa.nodeAffinity » au même noeud NUMA que la carte réseau SRIOV
- Activez les paramètres de performance de l'hyperviseur.
- Limitez la surcharge de vSwitch en activant SR-IOV sur les cartes réseau physiques prises en charge.
- Configurez les vCPU de la VM pour qu'ils s'exécutent sur le même noeud NUMA que les cartes réseau physiques.
- Définissez la sensibilité à la latence des VM sur Élevée.

AWS

Le C8000v prend en charge le déploiement sur AWS en se lançant en tant qu'image de machine Amazon (AMI) dans un cloud privé virtuel (VPC) Amazon, ce qui permet aux utilisateurs de provisionner une section logiquement isolée du cloud AWS pour leurs ressources réseau.

Files d'attente multiémission

Dans un C8000v exécuté sur AWS, l'utilisation de files d'attente multi-TX (Multi-TXQ) est une fonctionnalité clé. Ces files d'attente permettent de réduire la charge de traitement interne et améliorent l'évolutivité. Le fait d'avoir plusieurs files d'attente rend plus rapide et plus simple l'affectation des paquets entrants et sortants au processeur virtuel (vCPU) approprié.

Contrairement à certains systèmes où les files d'attente RX/TX sont attribuées par vCPU, dans le C8000v, ces files d'attente sont attribuées par interface. Les files d’attente RX (réception) et TX (transmission) servent de points de connexion entre l’application Catalyst 8000V et l’infrastructure ou le matériel AWS, gérant ainsi l’envoi et la réception du trafic réseau. AWS contrôle le nombre et la vitesse des files d'attente RX/TX disponibles pour chaque interface, en fonction du type d'instance.

Pour créer plusieurs files d'attente TX, le Catalyst 8000V doit disposer de plusieurs interfaces. Lorsque plusieurs files d'attente TX sont activées, le périphérique conserve l'ordre des flux de paquets en utilisant une méthode de hachage basée sur le 5-uplet du flux (IP source, IP de destination, port source, port de destination et protocole). Ce hachage détermine la file d'attente TX à utiliser pour chaque flux.

Les utilisateurs peuvent créer plusieurs interfaces sur le Catalyst 8000V en utilisant la même carte réseau physique connectée à l'instance AWS. Pour ce faire, vous devez configurer des interfaces de bouclage ou ajouter des adresses IP secondaires.

Avec les files d'attente multitâche, il existe plusieurs files d'attente de transmission pour gérer le trafic sortant. Dans l'exemple, il y a douze files d'attente TX (numérotées de 0 à 11). Cette configuration vous permet de surveiller chaque file d'attente individuellement pour voir si certaines sont saturées.

En regardant la sortie, vous pouvez voir que la file d'attente TX 8 a un compteur « full » très élevé (56 406 998), ce qui signifie que sa mémoire tampon se remplit fréquemment. Les autres files d'attente TX affichent zéro pour le compteur « complet », ce qui indique qu'elles ne sont pas encombrées.

```
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
```

La surveillance des compteurs « complets » des files d'attente TX permet d'identifier si une file d'attente de transmission est surchargée. Un nombre « complet » croissant de manière constante sur une file d'attente TX particulière indique un flux de trafic qui stresse le périphérique. Pour y remédier, il peut s'agir d'équilibrer le trafic, de régler les configurations ou d'adapter les ressources pour améliorer les performances.

Mesures dépassées

AWS définit certaines limites réseau au niveau de l'instance afin de garantir des performances réseau homogènes et de haute qualité sur différentes tailles d'instance. Ces limites permettent de maintenir la stabilité du réseau pour tous les utilisateurs.

Vous pouvez vérifier ces limites et les statistiques associées en utilisant la commande show controllers sur votre périphérique. Le résultat inclut de nombreux compteurs, mais nous nous concentrons ici uniquement sur les plus importants pour surveiller les performances du réseau :

```
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>
```

Vous pouvez maintenant plonger dans et voir à quoi ces compteurs font référence exactement :

- bw_in_allow_beyond : Nombre de paquets mis en file d'attente ou abandonnés parce que la bande passante entrante a dépassé la limite de l'instance.
- bw_out_allow_beyond : Nombre de paquets mis en file d'attente ou abandonnés parce que la bande passante sortante a dépassé la limite de l'instance.
- pps_allow_beyond : Nombre de paquets mis en file d'attente ou abandonnés car le nombre total de paquets par seconde (PPS) a dépassé la limite de l'instance.
- conntrack_allow_beyond : Nombre de connexions suivies ayant atteint le maximum autorisé

pour le type d'instance.
- dépassement de la limite linklocal_allow : Nombre de paquets abandonnés parce que le trafic vers les services proxy locaux (comme Amazon DNS, Instance Metadata Service et Time Sync Service) a dépassé la limite PPS pour l'interface réseau. Cela n'affecte pas les résolveurs DNS personnalisés.

Ce que cela signifie pour vos performances C8000v :

- Si vous remarquez que ces compteurs augmentent et que vous rencontrez des problèmes de performances, cela ne signifie pas toujours que le problème vient du routeur C8000v. Au lieu de cela, il indique souvent que l'instance AWS que vous utilisez a atteint ses limites de capacité. Vous pouvez vérifier les spécifications de votre instance AWS pour vous assurer qu'elle peut gérer vos besoins en matière de trafic.

Microsoft Azure

Dans cette section, découvrez comment Microsoft Azure et le routeur virtuel Cisco C8000v s'associent pour fournir des solutions de mise en réseau virtuelles évolutives, sécurisées et hautes performances dans le cloud.

Découvrez comment la mise en réseau accélérée (AN) et la fragmentation des paquets peuvent affecter les performances. En plus de revoir l'importance d'utiliser un type d'instance pris en charge pour Microsoft Azure.

Mise en réseau accélérée

Dans les cas de problèmes de performances où le C8000v est hébergé dans le cloud Microsoft Azure. Un aspect que vous ne pouvez pas négliger est de savoir si le réseau accéléré est activé ou non. Comme il augmente considérablement les performances du routeur. En bref, la mise en réseau accélérée permet la virtualisation des E/S à racine unique (SR-IOV) sur des machines virtuelles telles qu'une machine virtuelle Cisco Catalyst 8000V. Le chemin réseau accéléré contourne le commutateur virtuel, augmente la vitesse du trafic réseau, améliore les performances réseau et réduit la latence et la gigue du réseau.

Il existe un moyen très simple de vérifier si le réseau accéléré est activé. C'est-à-dire vérifier la sortie de show controllers et vérifier si un certain compteur est présent ou non :

```
------------------ 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
```

Les compteurs que vous recherchez sont ceux qui commencent par vf tels que vf_rx_good_packets. Si vous vérifiez que ces compteurs sont présents, vous pouvez être absolument sûr que la mise en réseau accélérée est activée.

Azure et fragmentation

La fragmentation peut avoir des conséquences négatives sur les performances. L'une des principales raisons de l'effet sur les performances est l'effet CPU/mémoire de la fragmentation et du réassemblage des paquets. Lorsqu'un périphérique réseau doit fragmenter un paquet, il doit allouer des ressources CPU/mémoire pour effectuer la fragmentation.

La même chose se produit lorsque le paquet est réassemblé. Le périphérique réseau doit stocker tous les fragments jusqu'à leur réception afin de pouvoir les réassembler dans le paquet d'origine.

Azure ne traite pas les paquets fragmentés avec la mise en réseau accélérée. Lorsqu'une machine virtuelle reçoit un paquet fragmenté, le chemin non accéléré le traite. Par conséquent, les paquets fragmentés ne bénéficient pas des avantages de la mise en réseau accélérée, tels qu'une latence plus faible, une gigue réduite et des paquets plus élevés par seconde. C'est pourquoi il est recommandé d'éviter autant que possible la fragmentation.

Azure, par défaut, abandonne les paquets fragmentés qui arrivent à la VM dans le désordre, ce qui signifie que les paquets ne correspondent pas à la séquence de transmission du point de terminaison source. Ce problème peut se produire lorsque des paquets circulent sur Internet ou sur d'autres grands réseaux étendus.

Types d'instance pris en charge pour Microsoft Azure

Il est important que le C8000v utilise un type d'instance pris en charge conformément aux normes Cisco. Ils sont disponibles dans le Guide d'installation et de configuration du logiciel Cisco Catalyst 8000V Edge.

La raison en est que les types d'instance dans cette liste sont ceux où le C8KV a été correctement testé. Maintenant, il y a la question valide si le C8000v fonctionne sur un type d'instance qui n'est pas répertorié ? La réponse est très probablement oui. Cependant, lorsque vous dépannez un problème aussi complexe que les problèmes de performances, vous ne voulez pas ajouter un autre facteur inconnu au problème. Pour cette seule raison, le TAC Cisco vous recommande toujours de rester dans un type d'instance pris en charge.

Ressources supplémentaires

Un problème de performances ne peut être réellement résolu que lorsqu'il se produit sur le moment. Cependant, cela peut être difficile à attraper car cela peut arriver à tout moment. Pour cette raison, nous fournissons ce script EEM. Il permet de capturer des sorties importantes au moment où les paquets commencent à être abandonnés et où des problèmes de performances surviennent :

```
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
```

À propos de cette traduction
Cisco a traduit ce document en traduction automatisée vérifiée par une personne dans le cadre d'un service mondial permettant à nos utilisateurs d'obtenir le contenu d'assistance dans leur propre langue.

Il convient cependant de noter que même la meilleure traduction automatisée ne sera pas aussi précise que celle fournie par un traducteur professionnel.