Solucionar problemas de desempenho dos roteadores C8000v

Contents

Introdução

Este documento descreve como solucionar problemas de desempenho em roteadores corporativos C8000v em nuvens públicas e cenários ESXi.

Componentes Utilizados

As informações neste documento são baseadas nestes componentes de hardware e software:

- C8000v executando a versão 17.12
- ESXI Versão 7.0 U3

As informações neste documento foram criadas a partir de dispositivos em um ambiente de laboratório específico. Todos os dispositivos utilizados neste documento foram iniciados com uma configuração (padrão) inicial. Se a rede estiver ativa, certifique-se de que você entenda o impacto potencial de qualquer comando.

Troubleshooting Geral

Embora o C8000v possa ser hospedado em ambientes diferentes, ainda existem algumas etapas de solução de problemas que podem ser realizadas e que são idênticas, independentemente de

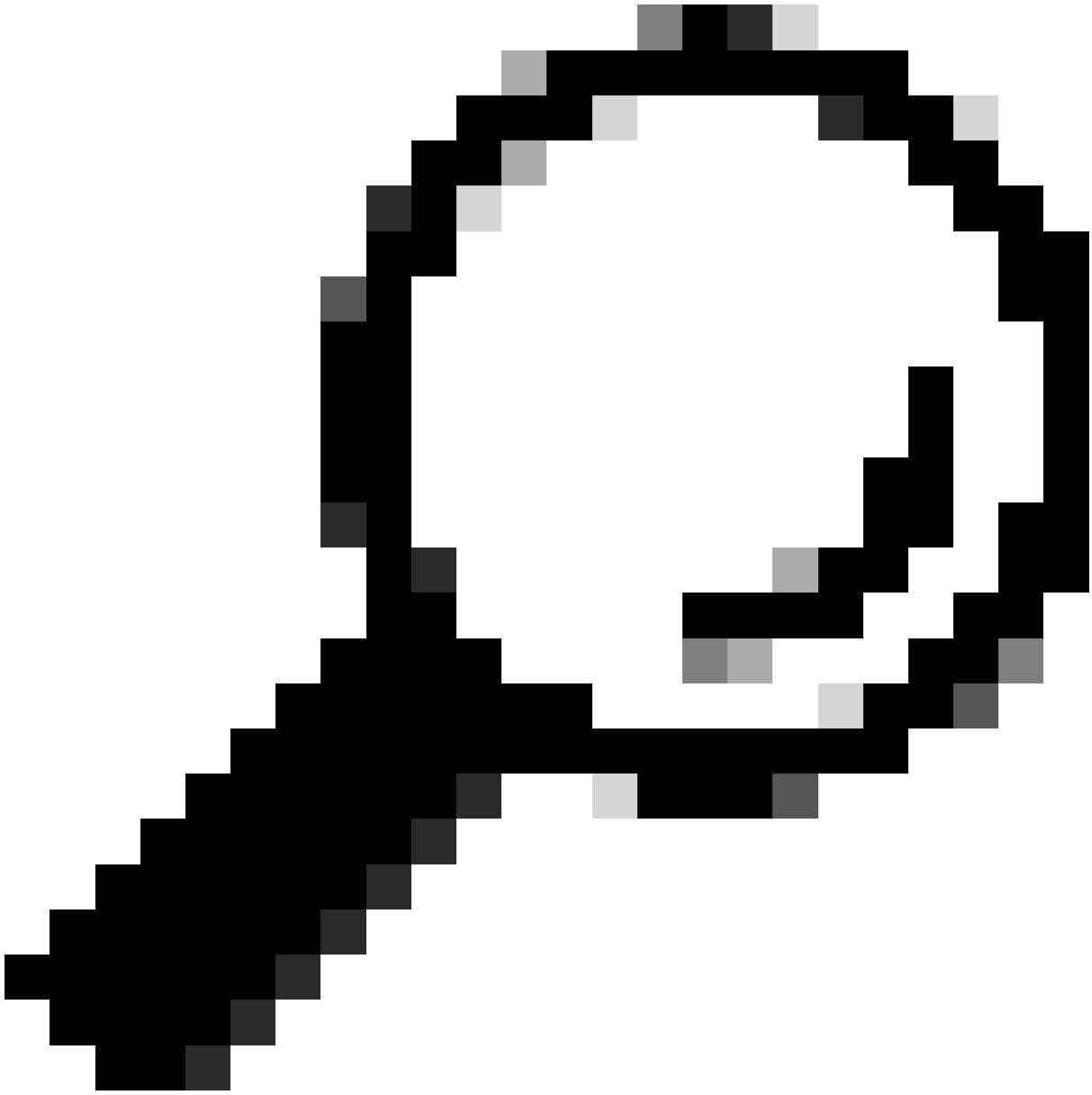
onde o C8000v esteja hospedado. Vamos começar com os fundamentos. A primeira coisa que você precisa verificar é se o dispositivo está atingindo seus limites de capacidade ou não. Para isso, você pode começar verificando estas duas saídas:

1. show platform hardware qfp ative datapath util summary - esse comando fornece informações completas sobre a entrada/saída que o C8000v está recebendo e transmitindo de cada porta. Você deve concentrar sua atenção na porcentagem de carga de processamento. Se você estiver em um cenário em que está atingindo 100%, isso significa que você está atingindo o limite de capacidade

```
------------------ 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 ative datapath infrastructure sw-cio - Pense neste comando como uma versão mais detalhada da versão acima. Ele fornece mais detalhes sobre os núcleos individuais, incluindo os núcleos de E/S e de criptografia que não fazem parte do número de utilização do QFP. É muito útil em um cenário em que você deseja ver se um núcleo de plano de dados específico está causando um gargalo.



Tip: Um detalhe muito importante ao usar este comando, sempre execute-o duas vezes. À medida que calcula a porcentagem de utilização do núcleo entre o momento em que o comando foi executado.

```
------------------ 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      0
   1      rcl0    32:    496      0      0      0      0      0      0      0      0      0      0
   2       ipc     1:    489      0      0      0      0      0      0      0      0      0      0
   3 vxe_punti     4:    490      0      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      1
    % 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.
```

Agora, você determinou se está atingindo o limite da plataforma ou não. A próxima etapa seria verificar se há quedas. Eles estão inerentemente conectados a problemas de desempenho. Há três tipos de quedas que podem ser considerados, dependendo de onde estejam ocorrendo.

- Superações: Esse tipo de queda de pacote ocorre na extremidade Rx. Eles ocorrem porque a capacidade de processamento de um ou mais núcleos foi excedida.
- Quedas de recursos: Esse tipo de queda de pacote ocorre no PPE. Eles estão relacionados aos recursos do roteador, como uma ACL ou QoS.
- Quedas de energia: Esse tipo de queda de pacote ocorre na extremidade Tx. Elas acontecem devido ao congestionamento nos buffers de Tx.

Para identificar quais descartes você está experimentando, você pode usar estas saídas:

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

Você verifica como identificar quais descartes você está enfrentando e como reduzi-los. No entanto, o foco maior neste artigo será nos descartes conhecidos como Taildrops, pois eles são particularmente difíceis de solucionar em roteadores virtuais.

Overruns

Uma queda de saturação no Cisco IOS XE ocorre quando a interface de rede recebe pacotes mais rápido do que pode processá-los ou armazená-los em seu buffer. Especificamente, os buffers internos das interfaces (fila FIFO) ficam cheios porque a taxa de dados de entrada excede a capacidade do hardware de manipulá-los. Como resultado, novos pacotes de entrada não podem ser armazenados e são descartados, o que incrementa o contador de saturação. Isso é essencialmente uma perda de pacotes causada pela sobrecarga temporária da interface.

Esse tipo de descarte de pacote ocorre na extremidade Rx. Eles acontecem porque a capacidade de processamento de um ou mais núcleos foi excedida e o thread Rx é incapaz de distribuir pacotes de entrada para o thread PP relevante e os buffers de entrada já estão cheios. Para fazer uma simples analogia, você pode pensar nela como uma fila em um caixa que fica muito cheia porque os pacotes estão chegando mais rápido do que o caixa (hardware de interface) pode

atendê-los. Quando a fila está cheia, os novos clientes precisam sair sem serem atendidos - esses são os pontos de sobrecarga.

Embora o hardware seja mencionado nesta seção, o C8000v é um roteador baseado em software. Neste caso, as derrapagens podem ser causadas por:

- Alta utilização do plano de dados: Se a utilização do plano de dados for alta, os pacotes não poderão ser interrogados com a rapidez necessária, levando a sobrecargas. Por exemplo, a presença de "fluxos de elefantes" (fluxos de dados grandes e contínuos) pode saturar os recursos de processamento e causar saturações nas interfaces.
- Modelo de dispositivo incorreto: O uso de um modelo de dispositivo inadequado pode resultar em gerenciamento ineficiente de buffer e saturações. Isso pode ser verificado com o comando show platform software cpu alloc e pode ser alterado com o comando platform resource <template>.

Cada interface recebe um conjunto limitado de créditos, esse mecanismo impede uma interface ocupada e sobrecarrega os recursos do sistema. Cada vez que um novo pacote chega ao dataplane, um crédito é necessário. Quando o processamento do pacote estiver concluído, o crédito será retornado para que o thread Rx possa usá-lo novamente. Se não houver crédito disponível para a interface, o pacote precisará aguardar no anel Rx da interface. Em geral, você espera que as quedas relacionadas ao limite de desempenho sejam saturações de Rx porque a capacidade de processamento de um ou mais núcleos foi excedida.

Para identificar saturações, você normalmente verifica as estatísticas da interface em busca de erros de entrada, especificamente o contador de saturação:

- Use o comando show platform hardware qfp ative datapath infrastructure sw-cio para identificar a utilização principal e se o número de créditos para uma interface específica tiver sido excedido.
- Use o comando show interface <interface-name> e procure a contagem de saturação na saída.

As saturações são mostradas como parte dos erros de entrada, por exemplo:

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

Vamos supor um caso em que Gig2 está observando problemas de desempenho causados por saturações. Para determinar o thread de trabalho associado a esta interface, você pode usar este comando:

```
#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
```

Em seguida, você pode analisar a utilização do segmento específico responsável pelo tráfego Rx dessa interface e seu número de créditos.

Em um cenário em que o Gig2 está observando problemas de desempenho devido a saturações, você pode esperar que o PP#2 seja constantemente totalmente utilizado (Ocioso = 0%) e créditos baixos/zero para a 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
```

Quedas de recursos

Os pacotes são tratados por qualquer thread de plano de dados disponível e são distribuídos estritamente com base na disponibilidade de núcleos QFP através da função Rx do software (x86) - Distribuição Baseada em Carga (LBD). Os pacotes que chegam no PPE podem ser descartados com uma razão de descarte QFP específica, que pode ser verificada usando esta saída:

```
#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
```

As razões para as gotas são diversas e geralmente são autoexplicativas. Para investigar mais, um rastreamento de pacote pode ser usado.

Taildrops

Como foi mencionado antes, as quedas de cauda ocorrem onde o dispositivo está tentando transmitir pacotes, mas os buffers de transmissão estão cheios.

Nesta subseção, você examinará quais saídas podem ser examinadas diante desse tipo de situação. Quais valores você pode ver neles significam e o que você pode fazer para mitigar o problema.

Primeiro, você precisa saber como identificá-los. Uma dessas maneiras é simplesmente observar o show interface. Fique atento a qualquer queda de saída que aumente:

```
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
```

Esse comando é particularmente útil para entender se você está passando por congestionamento ou não:

- show platform hardware qfp ative datapath infrastructure - HQF significa ‘Hierarchical Queueing Framework`. Este é um recurso que permite o gerenciamento da Qualidade de Serviço (QoS - Quality of Service) em diferentes níveis (físico, lógico e de classe) usando a interface de linha de comando QoS modular (MQC - QoS Command-Line Interface). Ele mostra os custos atuais de RX e TX. Quando a fila TX está cheia, como mostra a saída (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
```

A saída sugere que o hardware subjacente não está acompanhando o envio de pacotes. Para depurar a interface subjacente, você precisa potencialmente olhar para fora do C8000v e para o ambiente subjacente em que o C8000v está sendo executado para ver se há erros adicionais relatados nas interfaces físicas subjacentes.

Para verificar o ambiente, há uma etapa que você pode executar antes de verificar em qual hipervisor o roteador C8000v está sendo executado. Isso serve para verificar a saída do comando show controller. No entanto, você pode se ver perdido no que cada contador significa ou onde olhar.

Primeiro, um detalhe importante que você precisa ter em mente ao observar essa saída é que as informações são originadas principalmente dos próprios vNICs. Cada driver NIC tem um conjunto específico de contadores que eles usam, que podem variar naturalmente de acordo com o driver. Hipervisores diferentes têm algum tipo de efeito sobre o que é apresentado também. Alguns contadores, como os contadores mbuf, são estatísticas do driver DPDK. Eles podem variar de acordo com os diferentes drivers DPDK. A contagem real geralmente é feita pelo hipervisor na camada de virtualização.

```
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
```

Dedique um minuto aqui para saber como interpretar e ler estes contadores:

1. Se você vir subX, significa que é uma sub-interface - uma divisão lógica da interface principal. O sub0 é geralmente o principal/padrão. Eles são frequentemente usados quando várias VLANs estão envolvidas.
2. Em seguida, você terá "rx = recebendo" e "tx = transmitindo".
3. Finalmente, q0 refere-se à primeira fila/padrão usada por essa interface

Embora não haja uma descrição para cada contador, o artigo descreve alguns deles, que podem ser relevantes para a solução de problemas:

- "RX_MISSED_ERRORS:" é visto quando o buffer da NIC (Rx FIFO) fica sobrecarregado. Essa condição leva a quedas e um aumento na latência. Uma possível solução para isso é aumentar o buffer da placa de rede (o que é impossível no nosso caso) ou alterar o driver da placa de rede.
- "tx_q0_drop_total" e "tx_q0_tx_ring_full": Eles podem dizer que o host está descartando pacotes, e o C8000v está experimentando quedas de cauda no C8000v porque o host está pressionando novamente o C8000v

Na saída acima, não vemos nenhum "rx_missing_errors". No entanto, como estamos nos concentrando em taildrops, vemos "tx_q0_drop_total" e "tx_q0_tx_ring_full". Com isso, podemos concluir que há realmente um congestionamento causado pelo hardware subjacente do host.

Como mencionado anteriormente, cada hipervisor tem algum tipo de efeito sobre o que é apresentado. O artigo aborda isso na próxima seção, à medida que aborda as diferenças entre os diferentes hipervisores onde o C8000v pode ser hospedado. Você também pode encontrar as diferentes recomendações para tentar atenuar esse tipo de problema em cada uma delas.

Hipervisores

Um hipervisor é uma camada de software que permite que vários sistemas operacionais (chamados de máquinas virtuais ou VMs) sejam executados em um único host de hardware físico, gerenciando e alocando os recursos de hardware, como CPU, memória e armazenamento, para cada VM. Ele garante que essas máquinas virtuais operem de forma independente, sem interferir umas nas outras.

No contexto do Cisco Catalyst 8000V (C8000v), o hipervisor é a plataforma que hospeda a máquina virtual do C8000v. Como descobrir qual hipervisor está hospedando seu C8000v? Há uma saída bastante útil que nos dá essas informações. Além disso, você também pode verificar a que tipo de recursos nosso roteador virtual tem acesso:

```
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

O ESXi é um hipervisor tipo 1 desenvolvido pela VMware que é instalado diretamente em servidores físicos para permitir a virtualização. Ele permite que várias máquinas virtuais (VMs) sejam executadas em um único servidor físico abstraindo os recursos de hardware e alocando-os a cada VM. O roteador C8000v é uma dessas VMs.

Você pode começar analisando um cenário comum em que o congestionamento está ocorrendo.

Isso pode ser confirmado verificando-se o contador tx_q0_tx_ring_full:

Exemplo:

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

Esse congestionamento ocorre quando o C8000V tenta enviar pacotes através da interface VMXNET3. No entanto, o anel de buffer já está cheio de pacotes, que acabam em atrasos ou quedas.

Nessas condições, essas quedas estão acontecendo no lado do hipervisor, como mencionamos antes. Se todas as recomendações forem atendidas, é recomendável consultar o suporte da VMware para entender o que está acontecendo na placa de rede.

Aqui estão algumas sugestões sobre como melhorar o desempenho:

- Use um vSwitch e um uplink dedicados para obter o desempenho ideal
- Ao atribuir o C8000V a um vSwitch dedicado apoiado por seu próprio uplink físico, podemos isolar seu tráfego de vizinhos ruidosos e evitar gargalos de recursos compartilhados.

Há alguns comandos que valem a pena observar no lado do ESXi. Por exemplo, para verificar a

perda de pacotes da interface ESXi, podemos fazer o seguinte:

1. Ative o SSH.
2. Conecte-se ao ESXi usando SSH.
3. Execute esxtop.
4. Digite n.

O comando esxtop pode mostrar pacotes descartados no switch virtual se o driver de rede da máquina virtual ficar sem memória de buffer Rx. Mesmo que esxtop mostre os pacotes como descartados no switch virtual, eles são realmente descartados entre o switch virtual e o driver do sistema operacional convidado.

Procure qualquer pacote descartado em %DRPTX e %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 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 (
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
```

Este comando lista todas as NICs configuradas em um host:

```
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
```

Há também um comando útil para verificar o status do vNIC atribuído a uma VM específica.

```
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:~]
```

Observando o c8kv-gw-mgmt , que é uma VM do C8000v, há duas redes atribuídas:

- c8kv-to-9200l
- c8kv-para-cisco

Você pode usar a ID mundial para procurar mais informações sobre esta 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:~]
```

Depois de obter essas informações, você pode identificar a qual rede o vSwitch está atribuído.

Para verificar algumas estatísticas de tráfego da NIC física atribuída ao vSwitch, temos este comando:

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

Esse comando exibe informações como pacotes recebidos, bytes recebidos, pacotes descartados e erros recebidos. Isso pode ajudar a identificar se há quedas acontecendo na placa de rede.

```
[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
```

Há algumas configurações a serem verificadas que podem melhorar o desempenho do Cisco Catalyst 8000V executado em um ambiente ESXi modificando as configurações no host e na máquina virtual:

- Defina o hardware virtual: Configuração de reserva de CPU para Máximo.
- Reserve toda a memória do convidado no Hardware virtual: Memória.
- Selecione VMware Paravirtual no hardware virtual: Controladora SCSI.
- Na página Hardware virtual: Adaptador de rede: Tipo de adaptador, selecione SR-IOV para as placas de rede suportadas
- Defina a opção General Guest OS Version > VM Options como Other 3.x ou posterior Linux (64 bits).

- Defina a opção Opções de VM em Sensibilidade de latência avançada como Alta.
- Em Opções da VM > Configuração de Edição Avançada, adicione "numa.nodeAffinity" ao mesmo nó NUMA da NIC SRIOV
- Habilite as configurações de desempenho do hipervisor.
- Limite a sobrecarga do vSwitch habilitando o SR-IOV nas placas de rede físicas suportadas.
- Configure os vCPUs da VM para execução no mesmo nó NUMA que as NICs físicas.
- Defina a sensibilidade de latência da VM como Alta.

AWS

O C8000v suporta a implantação no AWS, iniciando como uma Amazon Machine Image (AMI) dentro de uma Amazon Virtual Private Cloud (VPC), permitindo que os usuários provisionem uma seção logicamente isolada da nuvem do AWS para seus recursos de rede.

Filas Multi-TX

Em um C8000v executado em AWS, um recurso importante é o uso de Filas Multi-TX (Multi-TXQs). Essas filas ajudam a reduzir a sobrecarga de processamento interno e melhoram a escalabilidade. Ter várias filas torna mais rápido e simples atribuir pacotes de entrada e saída à CPU virtual (vCPU) correta.

Diferentemente de alguns sistemas em que as filas RX/TX são atribuídas por vCPU, no C8000v, essas filas são atribuídas por interface. As filas RX (recepção) e TX (transmissão) servem como pontos de conexão entre o aplicativo Catalyst 8000V e a infraestrutura ou hardware AWS, gerenciando como o tráfego de rede é enviado e recebido. O AWS controla o número e a velocidade das filas RX/TX disponíveis para cada interface, dependendo do tipo de instância.

Para criar várias filas TX, o Catalyst 8000V precisa ter várias interfaces. Quando várias filas TX são ativadas, o dispositivo mantém a ordem dos fluxos de pacotes usando um método de hash baseado na tupla 5 do fluxo (IP origem, IP destino, porta origem, porta destino e protocolo). Esse hash decide qual fila de TX usar para cada fluxo.

Os usuários podem criar várias interfaces no Catalyst 8000V usando a mesma placa de interface de rede (NIC) física conectada à instância do AWS. Isso é feito por meio da configuração de interfaces de loopback ou da adição de endereços IP secundários.

Com Multi-TXQs, há várias filas de transmissão para lidar com o tráfego de saída. No exemplo, há doze filas TX (numeradas de 0 a 11). Essa configuração permite monitorar cada fila individualmente para ver se alguma está ficando cheia.

Observando a saída, você pode ver que a Fila TX 8 tem um contador "completo" muito alto (56.406.998), o que significa que seu buffer está sendo preenchido com frequência. As outras filas TX mostram zero para o contador "cheio", indicando que não estão congestionadas.

```
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
```

A monitoração dos contadores "completos" das filas TX ajuda a identificar se alguma fila de transmissão está sobrecarregada. Uma contagem "completa" em aumento constante em uma fila TX específica aponta para um fluxo de tráfego que está sobrecarregando o dispositivo. Abordar isso pode envolver o balanceamento de tráfego, o ajuste de configurações ou o dimensionamento de recursos para melhorar o desempenho.

Métricas Excedidas

O AWS define certos limites de rede no nível da instância para garantir desempenho de rede consistente e de alta qualidade em diferentes tamanhos de instância. Esses limites ajudam a manter uma rede estável para todos os usuários.

Você pode verificar esses limites e as estatísticas relacionadas usando o comando show controllers em seu dispositivo. A saída inclui muitos contadores, mas aqui nos concentramos apenas nos mais importantes para monitorar o desempenho da rede:

```
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>
```

Agora você pode se aprofundar e ver exatamente a que esses contadores se referem:

- bw_in_permit_beyond: Número de pacotes na fila ou descartados porque a largura de banda de entrada ultrapassou o limite da instância.
- bw_out_permit_beyond: Número de pacotes enfileirados ou descartados porque a largura de banda de saída ultrapassou o limite da instância.
- pps_permit_beyond: Número de pacotes enfileirados ou descartados porque o total de pacotes por segundo (PPS) excedeu o limite da instância.
- conntrack_permit_beyond: Número de conexões rastreadas que atingiram o máximo permitido para o tipo de instância.
- linklocal_permit_beyond: Número de pacotes descartados porque o tráfego para serviços proxy locais (como Amazon DNS, Instance Metadata Service e Time Sync Service) excedeu o limite de PPS para a interface de rede. Isso não afeta os resolvedores DNS personalizados.

O que isso significa para o desempenho do seu C8000v:

- Se você observar que esses contadores aumentam e apresentam problemas de desempenho, isso nem sempre significa que o roteador C8000v é o problema. Em vez disso, geralmente indica que a instância do AWS que você está usando atingiu seus limites de capacidade. Você pode verificar as especificações de sua instância do AWS para garantir que ela possa lidar com suas necessidades de tráfego.

Microsoft Azure

Nesta seção, explore como o Microsoft Azure e o roteador virtual Cisco C8000v se combinam para fornecer soluções de rede virtual escaláveis, seguras e de alto desempenho na nuvem.

Veja como a Rede Acelerada (AN) e a fragmentação de pacotes podem afetar o desempenho. Além de revisar a importância de usar um tipo de instância com suporte para o Microsoft Azure.

Rede acelerada

Em casos de problemas de desempenho em que o C8000v está hospedado na Nuvem do Microsoft Azure. Um aspecto que não pode ser ignorado é se a Accelerated Network está ativada ou não. À medida que aumenta muito o desempenho do roteador. Resumindo, a rede acelerada permite a virtualização de E/S de raiz única (SR-IOV) em VMs como uma VM do Cisco Catalyst 8000V. O caminho de rede acelerado ignora o switch virtual, aumenta a velocidade do tráfego de rede, melhora o desempenho da rede e reduz a latência e o jitter da rede.

Há uma maneira muito simples de verificar se a rede acelerada está ativada. Isso serve para verificar a saída do comando show controllers e se um determinado contador está presente ou não:

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

Os contadores que você está procurando são aqueles que começam com vf como vf_rx_good_packets. Se você verificar que esses contadores estão presentes, poderá ter certeza absoluta de que a rede acelerada está habilitada.

Azure e Fragmentação

A fragmentação pode ter implicações de desempenho negativas. Uma das principais razões para o efeito no desempenho é o efeito CPU/memória da fragmentação e remontagem de pacotes. Quando um dispositivo de rede precisa fragmentar um pacote, ele precisa alocar recursos de

CPU/memória para executar a fragmentação.

O mesmo acontece quando o pacote é remontado. O dispositivo de rede deve armazenar todos os fragmentos até que sejam recebidos para que possa reagrupá-los no pacote original.

O Azure não processa pacotes fragmentados com a Rede Acelerada. Quando uma VM recebe um pacote fragmentado, o caminho não acelerado o processa. Como resultado, os pacotes fragmentados perdem os benefícios da rede acelerada, como menor latência, atraso de sincronismo reduzido e pacotes mais altos por segundo. Por esse motivo, a recomendação é evitar a fragmentação, se possível.

O Azure, por padrão, descarta pacotes fragmentados que chegam à VM fora de ordem, o que significa que os pacotes não correspondem à sequência de transmissão do ponto de extremidade de origem. Esse problema pode ocorrer quando os pacotes trafegam pela Internet ou por outras WANs grandes.

Tipos de Instância com Suporte para o Microsoft Azure

É importante que o C8000v esteja usando um tipo de instância compatível de acordo com os padrões da Cisco. Eles podem ser encontrados no Guia de instalação e configuração do software Cisco Catalyst 8000V Edge.

A razão para isso é porque os tipos de instância nessa lista são aqueles em que o C8KV foi devidamente testado. Agora, há uma pergunta válida se o C8000v funciona em um tipo de instância que não está listado? A resposta é provavelmente sim. No entanto, quando estiver solucionando problemas tão complexos quanto problemas de desempenho, você não desejará adicionar outro fator desconhecido ao problema. Só por isso, o Cisco TAC sempre recomenda que você permaneça em um tipo de instância compatível.

Outros recursos

Um problema de desempenho só pode ser realmente solucionado quando estiver acontecendo no momento. No entanto, isso pode ser difícil de capturar, pois pode acontecer a qualquer momento. Por esse motivo, fornecemos este script EEM. Ele ajuda a capturar saídas importantes no momento em que os pacotes começam a ser descartados e problemas de desempenho ocorrem:

```
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
```

Sobre esta tradução
A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês (link fornecido) seja sempre consultado.