Risoluzione dei problemi di prestazioni dei router C8000v

Sommario

Introduzione

In questo documento viene descritto come risolvere i problemi di prestazioni dei router aziendali C8000v in cloud pubblici e scenari ESXi.

Componenti usati

Le informazioni di questo documento si basano sui seguenti componenti hardware e software:

- C800v con versione 17.12
- ESXI versione 7.0 U3

Le informazioni discusse in questo documento fanno riferimento a dispositivi usati in uno specifico ambiente di emulazione. Su tutti i dispositivi menzionati nel documento la configurazione è stata ripristinata ai valori predefiniti. Se la rete è operativa, valutare attentamente eventuali conseguenze derivanti dall'uso dei comandi.

Risoluzione dei problemi generali

Sebbene sia possibile ospitare il server C8000v in ambienti diversi, è comunque possibile eseguire alcune operazioni di risoluzione dei problemi identiche indipendentemente dalla

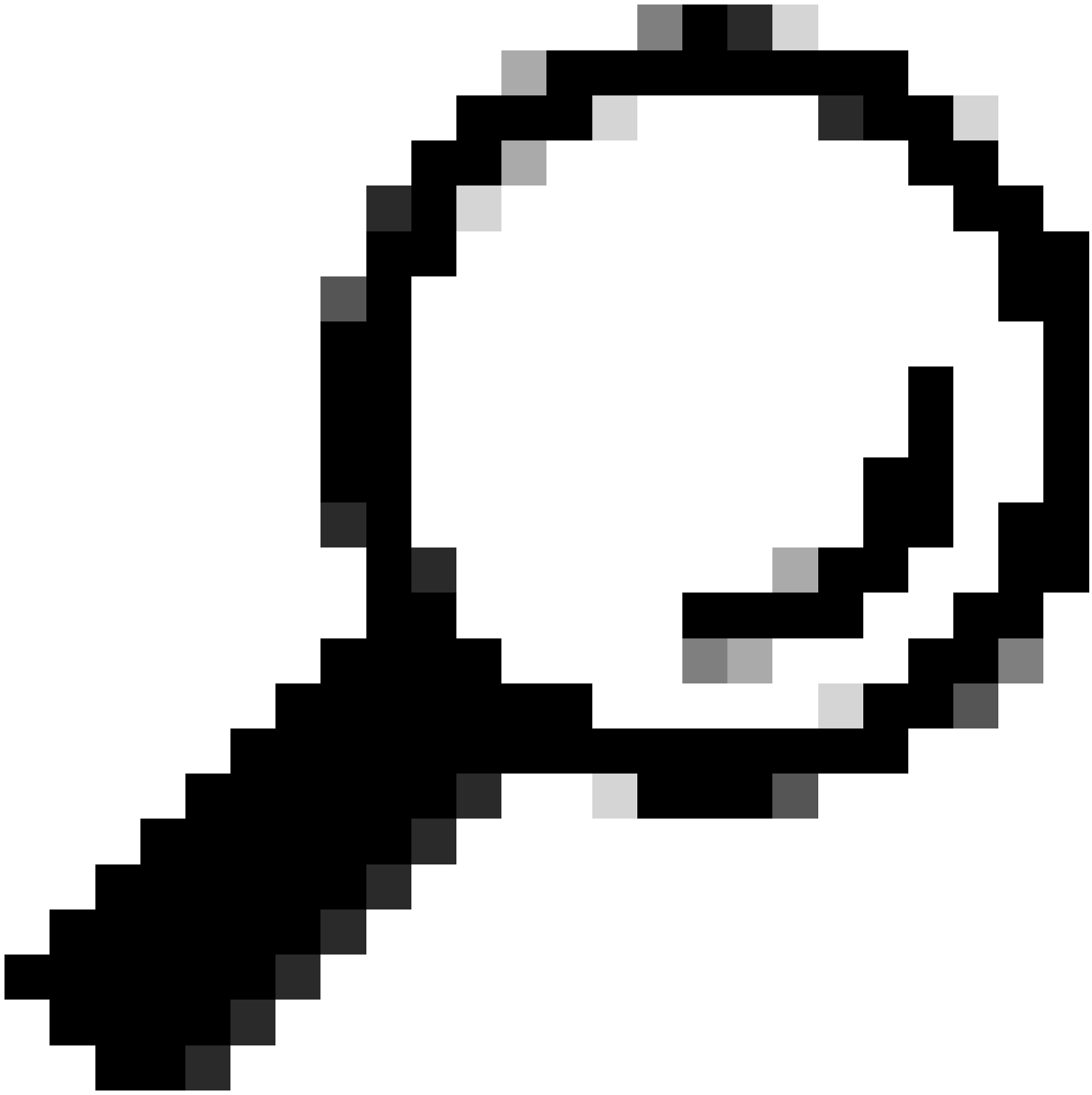
posizione in cui è ospitato il server C8000v. Iniziamo con le nozioni di base. La prima cosa da verificare è se il dispositivo sta raggiungendo o meno i limiti di capacità. A tale scopo, è possibile iniziare verificando i seguenti due output:

1. show platform hardware qfp active datapath summary - Questo comando fornisce le informazioni complete sugli input/output che il C8000v sta ricevendo e trasmettendo da ogni porta. È necessario focalizzare l'attenzione sulla percentuale di carico di elaborazione. In uno scenario in cui si raggiunge il 100%, significa che si sta raggiungendo il limite di capacità

```
------------------ 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 - Considerare questo comando come una versione più dettagliata di quella precedente. Fornisce maggiori dettagli sui singoli core, compresi i core I/O e i core di crittografia che non fanno parte del numero di utilizzo QFP. È molto utile in uno scenario in cui si desidera verificare se un core del piano dati specifico causa un collo di bottiglia.



Suggerimento: Quando si utilizza questo comando, è consigliabile eseguirlo sempre due volte. Calcola la percentuale di utilizzo dei core tra l'ora di esecuzione del comando.

```
------------------ 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
    % 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.
```

Ora, avete determinato se state raggiungendo o meno il limite della piattaforma. Il passo successivo sarebbe controllare se ci sono cadute. Questi sono intrinsecamente connessi a problemi di prestazioni. Esistono tre tipi di gocce che è possibile prendere in considerazione a seconda della posizione in cui si verificano.

- Sovraccarichi: Questo tipo di perdita di pacchetto si verifica sull'estremità Rx. Si verificano perché è stata superata la capacità di elaborazione di uno o più core.
- Perdite di funzionalità: Questo tipo di perdita di pacchetti si verifica nel DPI. Sono correlate a funzionalità del router, ad esempio un ACL o una QoS.
- Condutture: Questo tipo di perdita di pacchetti si verifica nell'estremità Tx. Sono causate dalla congestione dei buffer Tx.

Per identificare le cadute che si verificano, è possibile utilizzare i seguenti output:

- mostra stato di rilascio attivo qfp hardware della piattaforma deselezionato
- show interface
- mostra interfaccia mappa criteri

Verificate come identificare le gocce che state affrontando e come mitigarle. Tuttavia, l'obiettivo principale di questo articolo saranno le gocce di carico note come "Taildrops", in quanto sono particolarmente difficili da risolvere nei router virtuali.

Sovraccarichi

Una perdita di sovraccarico in Cisco IOS XE si verifica quando l'interfaccia di rete riceve i pacchetti più rapidamente di quanto possa elaborarli o memorizzarli nel proprio buffer. In particolare, i buffer interni delle interfacce (coda FIFO) diventano pieni perché la velocità dei dati in ingresso supera la capacità dell'hardware di gestirli. Di conseguenza, i nuovi pacchetti in arrivo non possono essere archiviati e vengono scartati, incrementando così il contatore di sovraccarico. Si tratta essenzialmente di una perdita di pacchetti causata dal sovraccarico temporaneo dell'interfaccia.

Questo tipo di pacchetti vengono scartati sull'estremità Rx. Questo problema si verifica perché la capacità di elaborazione di uno o più core è stata superata e il thread Rx non è in grado di distribuire i pacchetti in arrivo al thread PP pertinente e i buffer in entrata sono già pieni. Per fare una semplice analogia, si può pensare a una coda a un banco di check-out che diventa troppo piena perché i pacchetti stanno arrivando più velocemente di quanto il cassiere (hardware di

interfaccia) può servire loro. Quando la coda è piena, i nuovi clienti devono partire senza ricevere assistenza - queste sono le cadute di sovraccarico.

Anche se l'hardware è menzionato in questa sezione, il C8000v è un router basato su software. In questo caso, il superamento può essere causato da:

- Elevato utilizzo del piano dati: Se l'utilizzo del piano dati è elevato, non è possibile eseguire il polling dei pacchetti abbastanza velocemente, con conseguenti sovraccarichi. Ad esempio, la presenza di "flussi di elefanti" (grandi flussi di dati continui) può saturare le risorse di elaborazione e causare sovraccarichi sulle interfacce.
- Modello di dispositivo non corretto: L'utilizzo di un modello di dispositivo non appropriato può causare una gestione inefficiente del buffer e sovraccarichi. Per controllare questa condizione, usare il comando show platform software cpu alloc e modificarla con il comando platform resource<template>.

A ogni interfaccia viene assegnato un pool limitato di crediti. Questo meccanismo impedisce che un'interfaccia occupata sovraccarichi le risorse di sistema. Ogni volta che un nuovo pacchetto arriva nel piano dati, è necessario un credito. Al termine dell'elaborazione del pacchetto, il credito viene restituito in modo che il thread Rx possa utilizzarlo di nuovo. Se non è disponibile alcun credito per l'interfaccia che il pacchetto deve attendere nell'anello Rx dell'interfaccia. In genere, ci si aspetta che le perdite relative al limite delle prestazioni siano sovraccaricate da Rx perché è stata superata la capacità di elaborazione di uno o più core.

Per identificare i sovraccarichi, è in genere necessario controllare le statistiche dell'interfaccia per individuare eventuali errori di input, in particolare il contatore di sovraccarico:

- Utilizzare il comando show platform hardware qfp active datapath infrastructure sw-cio per identificare l'utilizzo principale e se è stato superato il numero di crediti per un'interfaccia specifica.
- Usare il comando show interface <nome-interfaccia>e cercare il conteggio dei sovraccarichi nell'output.

I sovraccarichi vengono visualizzati come parte degli errori di input, ad esempio:

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

Supponiamo che Gig2 stia osservando problemi di prestazioni causati da sovraccarichi. Per determinare il thread di lavoro associato all'interfaccia, è possibile utilizzare il comando seguente:

```
#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
```

Quindi, è possibile analizzare l'utilizzo del thread specifico responsabile per il traffico Rx di questa interfaccia e il relativo numero di crediti.

In uno scenario in cui Gig2 osserva problemi di prestazioni dovuti a sovraccarichi, è possibile prevedere che PP#2 sia costantemente utilizzato (inattivo = 0%) e crediti minimi/nulli per l'interfaccia 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
```

Rilasci funzionalità

I pacchetti vengono gestiti da qualsiasi thread del piano dati disponibile e vengono distribuiti esclusivamente in base alla disponibilità dei core QFP tramite la funzione Rx del software (x86) - Load Based Distribution (LBD). I pacchetti che arrivano nel DPI possono essere scartati con uno specifico motivo di rilascio QFP, che può essere controllato utilizzando questo output:

```
#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
```

Le ragioni per le gocce sono diverse e di solito sono auto-esplicative. Per ulteriori informazioni, è possibile utilizzare una traccia del pacchetto.

Gocce

Come accennato in precedenza, quando il dispositivo tenta di trasmettere pacchetti si verificano perdite di dati, ma i buffer di trasmissione sono pieni.

In questa sezione vi presenteremo i risultati che potrete ottenere quando affronterete questo tipo di situazione. Quali valori si possono vedere nel loro significato e cosa si può fare per mitigare il problema.

Innanzitutto, devi sapere come identificarli. Uno di questi metodi è semplicemente guardare l'interfaccia dello show. Tieni d'occhio le cadute in aumento:

```
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
```

Questo comando è particolarmente utile per capire se si verifica o meno una congestione:

- show platform hardware qfp active datapath infrastructure - HQF sta per 'Hierarchical Queueing Framework'. Questa funzionalità consente la gestione QoS (Quality of Service) a diversi livelli (fisico, logico e di classe) tramite l'interfaccia della riga di comando (MQC) QoS modulare. Mostra i costi correnti di RX e TX. Quando la coda TX è piena, come mostra l'output (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
```

L'output suggerisce che l'hardware sottostante non sta tenendo il passo con l'invio dei pacchetti. Per eseguire il debug dell'interfaccia sottostante, è necessario esaminare l'ambiente esterno a C8000v e l'ambiente sottostante su cui è in esecuzione C8000v per verificare se sono stati segnalati errori aggiuntivi sulle interfacce fisiche sottostanti.

Per controllare l'ambiente, è possibile eseguire una sola operazione prima di controllare in quale hypervisor è in esecuzione il router C800v. In questo modo, è possibile controllare l'output del comando show controller. Tuttavia, ci si può trovare persi su cosa ogni contatore significa o dove guardare.

In primo luogo, un dettaglio importante da tenere presente quando si osserva questo output è che le informazioni provengono principalmente dalle vNIC stesse. Ogni driver di scheda NIC dispone di un set specifico di contatori che utilizzano, che possono variare in base al driver. Hypervisor diversi hanno una sorta di effetto anche su ciò che viene presentato. Alcuni contatori, ad esempio i contatori mbuf, sono statistiche del driver DPDK. Queste impostazioni possono variare a seconda dei driver DPDK. Il conteggio effettivo viene in genere eseguito dall'hypervisor a livello di virtualizzazione.

```
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
```

Per informazioni su come interpretare e leggere questi contatori, utilizzare la pagina seguente:

1. Se viene visualizzato subX, significa che si tratta di una sottointerfaccia, ovvero una divisione logica dell'interfaccia principale. Il valore sub0 è in genere quello principale/predefinito. Queste VLAN vengono spesso usate quando sono coinvolte più VLAN.
2. Quindi, avete "rx = ricevente" e "tx = trasmittente".
3. q0 fa infine riferimento alla prima coda predefinita utilizzata dall'interfaccia

Sebbene non esista una descrizione per ogni contatore, nell'articolo ne vengono descritti alcuni, che possono essere rilevanti per la risoluzione dei problemi:

- "RX_MISSED_ERRORS:" viene visualizzato quando il buffer della scheda NIC (Rx FIFO) è pieno. Questa condizione determina una riduzione e un aumento della latenza. Una possibile soluzione a questo problema consiste nell'aumentare il buffer della scheda NIC (che nel nostro caso è impossibile) o nel modificare il driver della scheda NIC.
- "tx_q0_drop_total" e "tx_q0_tx_ring_full": Questi elementi possono indicare che l'host sta scartando i pacchetti e che il C8000v sta subendo un calo della coda nel C8000v perché l'host sta eseguendo una contropressione nel C8000v

Nell'output precedente, non viene visualizzato alcun "rx_missing_errors". Tuttavia, mentre ci concentriamo sulle code, vediamo sia "tx_q0_drop_total" che "tx_q0_tx_ring_full". In questo modo, si può concludere che esiste effettivamente una congestione causata dall'hardware sottostante dell'host.

Come accennato in precedenza, ogni hypervisor ha una sorta di effetto su ciò che viene presentato. L'articolo affronta questo argomento nella sezione successiva, analizzando le differenze tra i diversi hypervisor in cui può essere ospitato il C8000v. È inoltre possibile trovare i diversi consigli per cercare di mitigare questo tipo di problema in ciascuno di essi.

Hypervisor

Un hypervisor è un livello software che consente l'esecuzione di più sistemi operativi (denominati macchine virtuali o VM) su un singolo host hardware fisico tramite la gestione e l'allocazione di risorse hardware quali CPU, memoria e storage a ciascuna VM. Garantisce che queste macchine

virtuali funzionino in modo indipendente senza interferire tra loro.

Nel contesto di Cisco Catalyst 8000V (C8000v), l'hypervisor è la piattaforma che ospita la macchina virtuale C8000v. Come è possibile individuare l'hypervisor che ospita il server C800v? C'è un output piuttosto utile che ci dà quell'informazione. È inoltre possibile verificare il tipo di risorse a cui il router virtuale ha accesso:

```
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 è un hypervisor di tipo 1 sviluppato da VMware e installato direttamente sui server fisici per consentire la virtualizzazione. Consente l'esecuzione di più macchine virtuali (VM) su un singolo server fisico, astragendo le risorse hardware e allocandole a ciascuna VM. Il router C8000v è una di queste VM.

È possibile iniziare passando in rassegna uno scenario comune in cui si verifica una congestione. Ciò può essere confermato controllando il contatore tx_q0_tx_ring_full:

Esempio:

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

Questa congestione si verifica quando il C8000V tenta di inviare pacchetti tramite l'interfaccia VMXNET3. Tuttavia, l'anello del buffer è già pieno di pacchetti, che finiscono per essere ritardati o scartati.

In queste condizioni, queste gocce si stanno verificando sul lato dell'hypervisor, come accennato in precedenza. Se tutte le raccomandazioni sono state soddisfatte, si consiglia di consultare il supporto VMware per comprendere cosa sta succedendo sulla scheda NIC.

Ecco alcuni suggerimenti su come migliorare le prestazioni:

- Utilizzo di uno switch vSwitch e di un uplink dedicati per prestazioni ottimali
- Assegnando il C8000V a uno switch vSwitch dedicato supportato da un uplink fisico, è possibile isolare il traffico proveniente dai vicini più rumorosi ed evitare i colli di bottiglia delle risorse condivise.

Vi sono alcuni comandi che meritano di essere esaminati dal lato ESXi. Ad esempio, per verificare la perdita di pacchetti dall'interfaccia ESXi, è possibile effettuare le seguenti operazioni:

1. Abilitare SSH.
2. Collegarsi a ESXi utilizzando SSH.
3. Eseguire esxtop.

4. Digitare n.

Il comando esxtop può visualizzare i pacchetti ignorati nel commutatore virtuale se il driver di rete del computer virtuale ha esaurito la memoria buffer Rx. Anche se exxtop mostra i pacchetti come scartati allo switch virtuale, in realtà vengono scartati tra lo switch virtuale e il driver del sistema operativo guest.

Cercare i pacchetti ignorati in %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 0
100663324 2137274:telemetry-server vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 3.12 0.00 77.00 0.00 0.00
100663335 2396721:netlab vmnic0 vSwitch-to-Cisc 0.00 0.00 0.00 3.12 0.00 77.00 0.00 0.00
2214592519 vmnic1 - vSwitch-to-9200 76727.26 792.38 1353.00 16182.64 9.30 75.00 0.00 0.00
2248146954 vmnic0 - vSwitch-to-Cisc 16189.05 9.32 75.00 76736.97 792.38 1353.00 0.00 0.00
```

Con questo comando vengono elencate tutte le schede NIC configurate su un 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
```

È inoltre disponibile un comando utile per controllare lo stato della vNIC assegnata a una VM

specifica.

```
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:~]
```

Se si esamina la VM c8kv-gw-mgmt, che è una VM C8000v, vengono assegnate 2 reti:

- da c8kv a 9200l
- da c8kv a cisco

È possibile utilizzare l'ID mondiale per cercare ulteriori informazioni su questa 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:~]
```

Dopo aver ottenuto queste informazioni, è possibile identificare la rete a cui è assegnato lo switch vSwitch.

Per controllare alcune statistiche del traffico della scheda NIC fisica assegnata allo switch vSwitch,

si dispone di questo comando:

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

Con questo comando vengono visualizzate informazioni quali pacchetti ricevuti, byte ricevuti, pacchetti ignorati e errori ricevuti. In questo modo è possibile identificare eventuali cadute nella scheda NIC.

```
[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
```

Esistono alcune configurazioni da verificare che possono migliorare le prestazioni di Cisco Catalyst 8000V in esecuzione su un ambiente ESXi modificando le impostazioni sull'host e sul computer virtuale:

- Impostare l'hardware virtuale: Impostazione di prenotazione CPU su Massimo.
- Riservare tutta la memoria guest nell'hardware virtuale: Memoria.
- Selezionare VMware Paravirtual da Hardware virtuale: Controller SCSI.
- Dall'hardware virtuale: Scheda di rete: Tipo scheda, selezionare SR-IOV per le schede NIC supportate
- Impostare l'opzione General Guest OS Version > VM Options (Versione generale del sistema operativo guest > Opzioni VM) su Other 3.x or later Linux (64 bit) (Altro Linux 3.x o versioni successive).
- Impostare l'opzione Opzioni VM in Sensibilità latenza avanzata su Alta.

- In Opzioni VM > Advanced Edit Configuration, aggiungere "numa.nodeAffinity" allo stesso nodo NUMA della scheda NIC SRIOV
- Abilitare le impostazioni relative alle prestazioni dell'hypervisor.
- Limitare il sovraccarico di vSwitch abilitando SR-IOV sulle schede di interfaccia di rete fisiche supportate.
- Configurare le vCPU della VM in modo che vengano eseguite sullo stesso nodo NUMA delle schede di interfaccia di rete fisiche.
- Impostare la sensibilità della latenza della VM su Alta.

AWS

Il C8000v supporta l'installazione su AWS lanciando come Amazon Machine Image (AMI) all'interno di Amazon Virtual Private Cloud (VPC), consentendo agli utenti di effettuare il provisioning di una sezione del cloud AWS logicamente isolata per le proprie risorse di rete.

Code Multi-TX

In un C800v in esecuzione su AWS, una caratteristica chiave è l'uso di code Multi-TX (Multi-TXQ). Queste code consentono di ridurre il sovraccarico di elaborazione interno e di migliorare la scalabilità. La presenza di più code rende più rapida e semplice l'assegnazione dei pacchetti in entrata e in uscita alla CPU virtuale (vCPU) corretta.

A differenza di alcuni sistemi in cui le code RX/TX sono assegnate per vCPU, nel C8000v queste code sono assegnate per interfaccia. Le code RX (ricezione) e TX (trasmissione) fungono da punti di connessione tra l'applicazione Catalyst 8000V e l'infrastruttura o l'hardware AWS, gestendo le modalità di invio e ricezione del traffico di rete. AWS controlla il numero e la velocità delle code RX/TX disponibili per ciascuna interfaccia, a seconda del tipo di istanza.

Per creare più code TX, Catalyst 8000V deve avere più interfacce. Quando sono abilitate più code TX, il dispositivo mantiene l'ordine dei flussi di pacchetti utilizzando un metodo di hashing basato sulla 5-tupla del flusso (IP di origine, IP di destinazione, porta di origine, porta di destinazione e protocollo). Questo hashing decide quale coda TX utilizzare per ogni flusso.

Gli utenti possono creare più interfacce su Catalyst 8000V utilizzando la stessa scheda di interfaccia di rete fisica (NIC) collegata all'istanza AWS. A tale scopo, è possibile configurare le interfacce di loopback o aggiungere indirizzi IP secondari.

Con i Multi-TXQ, esistono più code di trasmissione per gestire il traffico in uscita. Nell'esempio sono presenti dodici code TX (numerate da 0 a 11). Questa impostazione consente di monitorare singolarmente ogni coda per verificare se si sta esaurendo.

Guardando l'output, si può vedere che TX Queue 8 ha un contatore "full" molto alto (56.406.998), il che significa che il suo buffer si sta riempiendo frequentemente. Le altre code TX mostrano zero per il contatore "full", a indicare che non sono congestionate.

```
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
```

Il monitoraggio dei contatori "completi" delle code TX consente di identificare se una coda di trasmissione è sovraccarica. Un conteggio "completo" in costante aumento su una particolare coda TX indica un flusso di traffico che sta causando lo stress del dispositivo. Per risolvere questo problema, è possibile eseguire il bilanciamento del traffico, l'adeguamento delle configurazioni o la scalabilità delle risorse per migliorare le prestazioni.

Metriche superate

AWS imposta alcuni limiti di rete a livello di istanza per garantire prestazioni di rete coerenti e di alta qualità su diverse dimensioni di istanza. Questi limiti consentono di mantenere una rete stabile per tutti gli utenti.

È possibile controllare questi limiti e le statistiche correlate utilizzando il comando show controller sul dispositivo. L'output include molti contatori, ma qui ci concentriamo solo su quelli più importanti per il monitoraggio delle prestazioni della rete:

```
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>
```

A questo punto è possibile eseguire un'immersione e vedere a cosa si riferiscono esattamente questi contatori:

- bw_in_ALLOW_exceeded: Numero di pacchetti in coda o scartati perché la larghezza di banda in ingresso ha superato il limite dell'istanza.
- bw_out_ALLOW_exceeded: Numero di pacchetti in coda o scartati perché la larghezza di banda in uscita ha superato il limite dell'istanza.
- pps_ALLOW_exceeded: Numero di pacchetti in coda o scartati perché il numero totale di pacchetti al secondo (PPS) ha superato il limite dell'istanza.
- conntrack_alloweded_exceeded: Numero di connessioni rilevate che hanno raggiunto il numero massimo consentito per il tipo di istanza.
- linklocal_ALLOW_exceeded: Numero di pacchetti ignorati perché il traffico verso i servizi proxy locali (come Amazon DNS, Instance Metadata Service e Time Sync Service) ha superato il limite PPS per l'interfaccia di rete. Ciò non influisce sui resolver DNS personalizzati.

Conseguenze per le prestazioni del server C8000v:

- Se questi contatori aumentano e si verificano problemi di prestazioni, ciò non sempre indica che il router C8000v è il problema. Al contrario, indica spesso che l'istanza AWS in uso ha raggiunto i limiti di capacità. È possibile controllare le specifiche dell'istanza AWS per verificare che possa gestire le esigenze di traffico.

Microsoft Azure

In questa sezione viene illustrato come Microsoft Azure e il router virtuale Cisco C800v si combinano per offrire soluzioni di rete virtuale scalabili, sicure e ad alte prestazioni nel cloud.

Ecco come la funzionalità di rete accelerata (LAN, Accelerated Networking) e la frammentazione dei pacchetti possono influire sulle prestazioni. Oltre a esaminare l'importanza dell'utilizzo di un

tipo di istanza supportato per Microsoft Azure.

Reti accelerate

Nei casi di problemi di prestazioni in cui C8000v è ospitato nel cloud di Microsoft Azure. Un aspetto che non è possibile ignorare è se la rete accelerata è abilitata o meno. Aumenta notevolmente le prestazioni del router. In poche parole, una rete accelerata consente la virtualizzazione dell'I/O con un'unica radice (SR-IOV) su VM come Cisco Catalyst 8000V. Il percorso di rete accelerato ignora lo switch virtuale, aumenta la velocità del traffico di rete, migliora le prestazioni di rete e riduce la latenza e l'instabilità della rete.

Esiste un modo molto semplice per verificare se la funzionalità Rete accelerata è abilitata. Ciò significa controllare l'output show controllers e verificare se un certo contatore è presente o meno:

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

I contatori che si stanno cercando sono quelli che iniziano con vf, ad esempio vf_rx_good_packets. Se si verifica la presenza di questi contatori, si è certi che la funzionalità di rete accelerata sia abilitata.

Azure e frammentazione

La frammentazione può avere implicazioni negative sulle prestazioni. Uno dei motivi principali che hanno influito sulle prestazioni è l'effetto della frammentazione e del riassemblaggio dei pacchetti

sulla CPU o sulla memoria. Quando un dispositivo di rete deve frammentare un pacchetto, deve allocare le risorse CPU/memoria per eseguire la frammentazione.

La stessa cosa accade quando il pacchetto viene ricomposto. Il dispositivo di rete deve memorizzare tutti i frammenti finché non li riceve per poterli ricomporre nel pacchetto originale.

Azure non elabora i pacchetti frammentati con le reti accelerate. Quando una VM riceve un pacchetto frammentato, il percorso non accelerato lo elabora. Di conseguenza, i pacchetti frammentati non traggono vantaggio dalle funzionalità di rete accelerata, quali una latenza inferiore, un jitter ridotto e pacchetti più alti al secondo. Per questo motivo si raccomanda di evitare, se possibile, la frammentazione.

Per impostazione predefinita, Azure elimina i pacchetti frammentati che arrivano alla VM in modo non corretto, ovvero i pacchetti non corrispondono alla sequenza di trasmissione dall'endpoint di origine. Questo problema può verificarsi quando i pacchetti viaggiano su Internet o su altre WAN di grandi dimensioni.

Tipi di istanza supportati per Microsoft Azure

È importante che il C8000v utilizzi un tipo di istanza supportato in base agli standard Cisco. È possibile trovarli nella Guida di installazione e configurazione del software Cisco Catalyst 8000V Edge.

Ciò è dovuto al fatto che i tipi di istanza presenti in tale elenco sono quelli in cui il C8KV è stato correttamente testato. A questo punto, è possibile stabilire se C8000v funziona su un tipo di istanza non elencato. La risposta è molto probabilmente sì. Tuttavia, quando si risolvono problemi complessi come quelli relativi alle prestazioni, non si desidera aggiungere un altro fattore sconosciuto al problema. Solo per questo motivo, Cisco TAC consiglia sempre di utilizzare un tipo di istanza supportato.

Ulteriori risorse

Un problema relativo alle prestazioni può essere effettivamente risolto solo quando si verifica in quel momento. Tuttavia, questo può essere difficile da scoprire come può accadere in ogni dato momento. Per questo motivo, forniamo questo script EEM. Aiuta a catturare output importanti nel momento in cui i pacchetti vengono scartati e i problemi di prestazioni ne derivano:

```
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
```

Informazioni su questa traduzione

Cisco ha tradotto questo documento utilizzando una combinazione di tecnologie automatiche e umane per offrire ai nostri utenti in tutto il mondo contenuti di supporto nella propria lingua. Si noti che anche la migliore traduzione automatica non sarà mai accurata come quella fornita da un traduttore professionista. Cisco Systems, Inc. non si assume alcuna responsabilità per l’accuratezza di queste traduzioni e consiglia di consultare sempre il documento originale in inglese (disponibile al link fornito).