



PROBLEMA BANDA ASR1004/ASR1006 ESP40

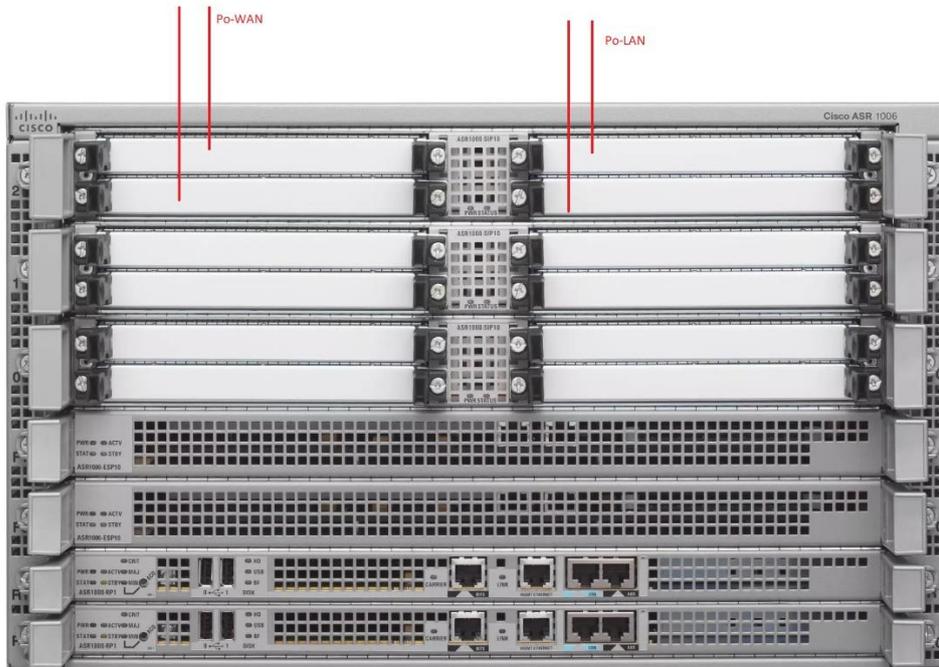
Responsável: Daniel Vieceli

Data: 02/06/2020

Cenário de análise e testes:

Um LACP com duas interfaces 10G para WAN

Um LACP com duas interfaces 10G para LAN



Software:

Cisco IOS XE Software, Version 03.16.10.S - Extended Support Release

Cisco IOS Software, ASR1000 Software (X86_64_LINUX_IOSD-ADVIPSERVICESK9-M), Version 15.5(3)S10, RELEASE SOFTWARE (fc3)

Chassis type: ASR1004

Slot	Type	State	Insert time (ago)
------	------	-------	-------------------

0	ASR1000-SIP40	ok	3w3d
0/0	SPA-1X10GE-L-V2	ok	3w3d
0/1	SPA-1X10GE-L-V2	ok	3w3d
0/2	SPA-1X10GE-L-V2	ok	3w3d
0/3	SPA-1X10GE-L-V2	ok	3w3d
R0	ASR1000-RP2	ok, active	3w3d
F0	ASR1000-ESP40	ok, active	3w3d
P0	ASR1004-PWR-AC	ps, fail	3w3d
P1	ASR1004-PWR-AC	ok	3w3d

Slot	CPLD Version	Firmware Version
------	--------------	------------------

0	00200800	15.4(2r)S
R0	08103002	15.4(2r)S
F0	1003190E	15.4(2r)S

Após uma análise de vários sistemas ASR1004/1006 especificamente com ESP40 foi identificado uma situação onde não era alcançado mais de 16Gbps de tráfego IN+OUT sendo encaminhado pela placa ESP40.

Observamos nessa situação um incremento de overrun na interface quando atingida essa banda e nesse momento é quando se inicia o descarte de frames, apresentando o mesmo valor de input erros e overrun

```

TenGigabitEthernet0/1/0 is up, line protocol is up
Hardware is SPA-1X10GE-L-V2, address is 5c71.0d3f.ebc0 (bia 5c71.0d3f.eb10)
Description: Pol-WAN
MTU 9216 bytes, BW 10000000 Kbit/sec, DLY 10 usec,
    reliability 255/255, txload 1/255, rxload 48/255
Encapsulation ARPA, loopback not set
Keepalive not supported
Full Duplex, 10000Mbps, link type is force-up, media type is 10GBase-SR/SW
output flow-control is on, input flow-control is on
ARP type: ARPA, ARP Timeout 04:00:00
Last input 00:00:05, output 00:00:09, output hang never
Last clearing of "show interface" counters 3w3d
Input queue: 0/32384/0/0 (size/max/drops/flushes); Total output drops: 0
Queueing strategy: fifo
Output queue: 0/40 (size/max)
30 second input rate 1903201000 bits/sec, 202959 packets/sec
30 second output rate 0 bits/sec, 0 packets/sec
483152298731 packets input, 606081836370061 bytes, 0 no buffer
Received 1 broadcasts (0 IP multicasts)
 0 runs, 0 giants, 0 throttles
26050 input errors, 0 CRC, 0 frame, 26050 overrun, 0 ignored
 0 watchdog, 141793 multicast, 0 pause input
76449 packets output, 9479676 bytes, 0 underruns
 0 output errors, 0 collisions, 2 interface resets
70637 unknown protocol drops
 0 babbles, 0 late collision, 0 deferred
 0 lost carrier, 0 no carrier, 0 pause output
 0 output buffer failures, 0 output buffers swapped out
    
```

Diagnostico/Testes

Arquitetura SIP40 e ESP40

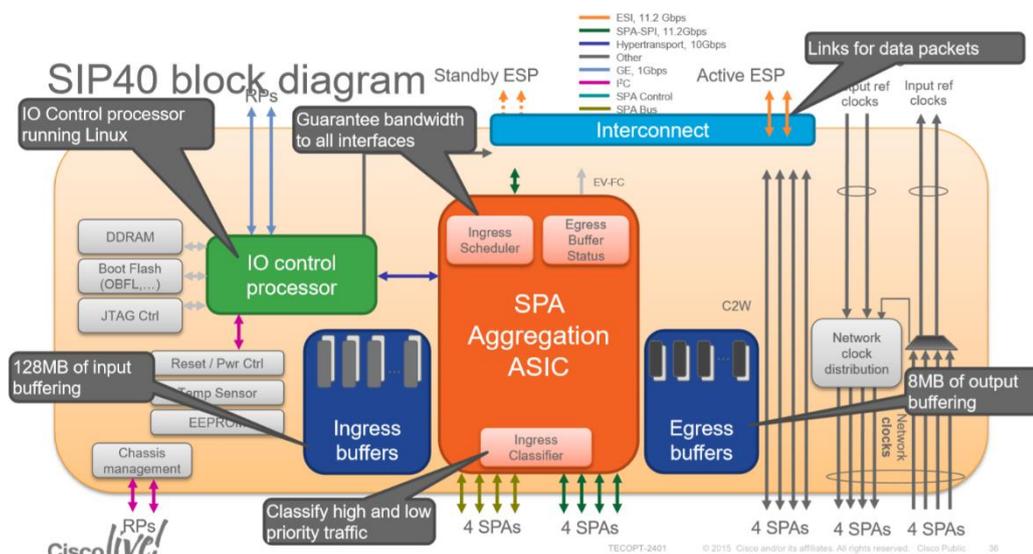


Diagrama de Bloco SIP40

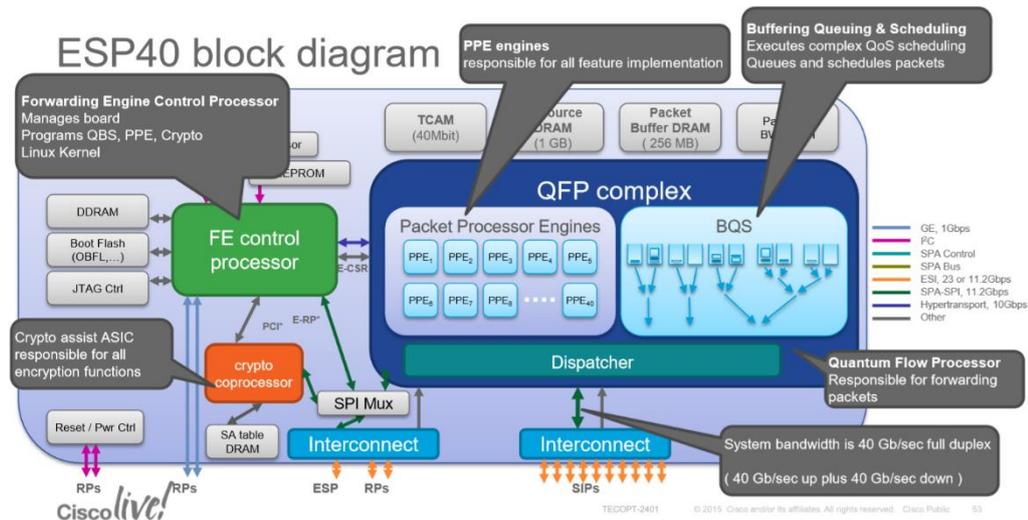


Diagrama de Bloco ESP40

Como observado na arquitetura do sistema o pacote entre a comunicação da SIP com a ESP frame é encaminhado para a GPM (General Packet Memory).

Essa memória é dividida em 16 partes sendo assim temos 16 filas de FIFO.

Para as SPA existem dois BUS de comunicação um de baixa prioridade (Low) e um de Alta Prioridade (High)

O Canal de Low aloca 18.27% da GPM, e o canal de High aloca 19.24% da GPM caso a GPM estiver cheia o modulo SPA irá deixar esse pacote buffer, (Hold-Queue) quando esse buffer acaba os pacotes serão descartados gerando o incremento do erro.

Porém observamos que isso ocorre longe da capacidade da ESP.

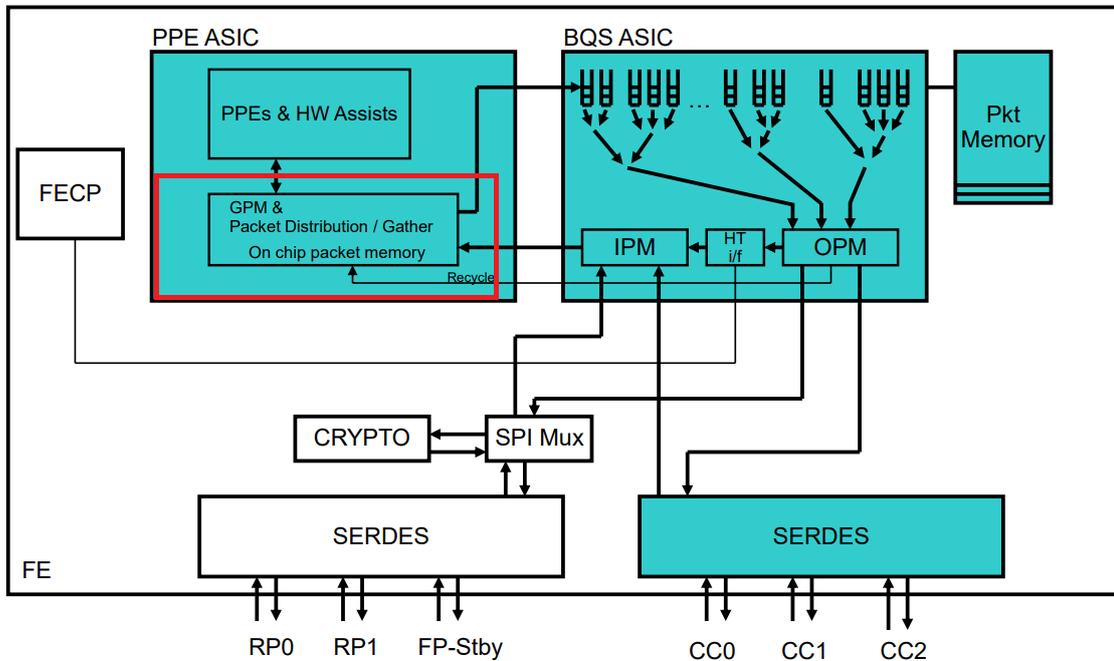
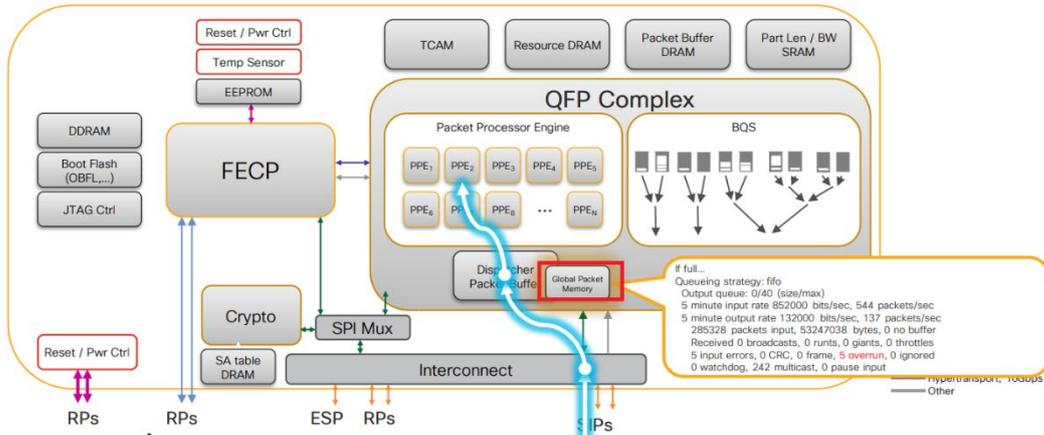
Fazendo uma análise nesses buffers da GPM, foi observado que somente o buffer de Low é utilizado por default em praticamente todos os frames de Input.

- **Global packet memory (GPM):** This is a small size on-chip SRAM, where the packet will remain stored during the entire processing by the packet-processing element (PPE) and will only be transferred to the traffic manager when the processing is complete. GPM can receive packets from all the SIPs and RPs in the system.

Descrição da GPM

Detalhamento da memória GPM (Global Packet Memory)

Ingress Packet Through ESP



```

show platform hardware qfp active infrastructure exmem qfp 0 resource
show platform hardware qfp active infrastructure bqs sorter memory available
show platform hardware qfp active infrastructure bqs queue output default interface
tenGigabitEthernet 0/0/0
show platform hardware slot 0 plim buffer settings detail
  
```

Podemos observar por default um baixo valor de pacotes na fila High.

show platform hardware qfp active datapath utilization

```
bras-cda-c#show platform hardware port 0/1/0 plim qos input map counters
Interface 0/1/0
RX High Pkts : 709241865      Bytes : 499253080648
RX Low Pkts  : 483379638082   Bytes : 606673138450735

bras-cda-c#show platform hardware qfp active datapath utilization
CPP 0: Subdev 0      5 secs      1 min      5 min      60 min
Input: Priority (pps)  1032        939        987        677
      (bps)  4410808    3608344    4130208    2466792
      Non-Priority (pps) 764293      775801     786547     704099
      (bps)  5535767216 5629718616 5714164224 5101130128
      Total (pps)  765325      776740     787534     704776
      (bps)  5540178024 5633326960 5718294432 5103596920
Output: Priority (pps)  213         210        210        210
      (bps)  113920    112264     112368     112328
      Non-Priority (pps) 755935      768110     778218     696490
      (bps)  5467557160 5569616872 5651535480 5042116152
      Total (pps)  756148      768320     778428     696700
      (bps)  5467671080 5569729136 5651647848 5042228480
Processing: Load (pct)  21          21         21         19
```

Packet Flow da arquitetura ASR1000:

ASR1k High-level Packet Flow

- Frame is received and **classified** ('hi' / 'lo') by either SPA or SIP
- Frames are scheduled based on priority and sent to QFP over ESI 'hi' or 'lo' priority channel
- Entire L2 frame is received by QFP **Input Packet Module** (IPM) and stored in **Global Packet Memory** (GPM)
- A free **PPE thread** is assigned to process the packet
- Packet remains in on chip memory (GPM) while it is processed by one of the PPEs
- The PPE thread runs through a **Feature Chain** in software. It can access resources like the HW-assists and TCAM and perform deep packet inspection, e.g. NBAR

Workaround:

Após essa análise entendemos que existe uma limitação no GPM de Low que não permite que todo o tráfego seja encaminhado em intervenção.

A solução encontrada foi marcar idealmente 50% do tráfego para fila de High, e aumentar o buffer do modulo SIP.

Aumentar buffer da SPA:

```
hold-queue 32384 in  
hold-queue 32384 out
```

Encaminhar pacotes de uma subinterface para a fila High:

```
plim qos input map cos enable  
plim qos input map cos 0 1 2 3 4 5 queue strict-priority
```

***Necessário ser em subinterface**

Encaminhar pacote para uma interface física para a Fila High:

```
plim qos input weight 10000  
plim qos input policer bandwidth 10000000 strict-priority  
plim qos input map ip all queue strict-priority
```

Existem outras formas de marcar esse tráfego na PLIM, por exemplo marcando o DSCP na entrada e usando esse valor na PLIM.

Esse tipo de artifício só foi testado em laboratório, por ser inviável balancear de forma efetiva os buffers.

```
ingress-class-map 1  
map ip dscp-based  
map ip dscp 43 queue strict-priority
```

```
class-map match-any QOS_MARK  
match access-group name ACL_QOS
```

```
policy-map QOS_IN  
class QOS_MARK  
set dscp af43  
set discard-class 0
```

```
ip access-list extended ACL_QOS  
permit icmp any any  
permit icmp any 192.168.254.0 0.0.0.3  
permit ip any 192.168.254.0 0.0.0.3  
permit ip any any
```

```

interface tenGigabitEthernet 0/1/0
  plim qos input map ip dscp-based
  plim qos input map ip dscp 38 queue strict-priority
  service-policy input QOS_IN

```

Após isso notamos um grande incremento em pacotes na fila de High:

```

router-asr-ppp#show platform hardware qfp active datapath utilization
  CPP 0: Subdev 0          5 secs          1 min          5 min          60 min
Input: Priority (pps)      270             262            262            263
      (bps)              224248         217936         218192         218416
  Non-Priority (pps)      888631         836161         806916         762179
      (bps)              6705405608     6286998736    6067961216    5708699728
  Total (pps)             888901         836423         807178         762442
      (bps)              6705629856     6287216672    6068179408    5708918144
Output: Priority (pps)    294             284            284            282
      (bps)              153480         147104         147632         145256
  Non-Priority (pps)     886667         833806         804460         759790
      (bps)              6706344728     6286278192    6066164296    5704061688
  Total (pps)            886961         834090         804744         760072
      (bps)              6706498208     6286425296    6066311928    5704206944
Processing: Load (pct)   22              21              20              19

```

```

router-asr-ppp#show platform hardware qfp active datapath utilization
  CPP 0: Subdev 0          5 secs          1 min          5 min          60 min
Input: Priority (pps)     431005         349630         70136          6085
      (bps)              4028761600     3278043200    655783200     54848760
  Non-Priority (pps)     371068         454951         749123         762801
      (bps)              2006553272     2791403896    5508539208    5703846504
  Total (pps)            802073         804581         819259         768886
      (bps)              6035314872     6069447096    6164322408    5758695264
Output: Priority (pps)    318             311            290            283
      (bps)              176392         170736         152848         145864
  Non-Priority (pps)     796446         800299         816407         766183
      (bps)              5996099264     6052088960    6161481808    5753595000
  Total (pps)            796764         800610         816697         766466
      (bps)              5996275656     6052259696    6161634656    5753740864
Processing: Load (pct)   20              20              21              19

```

Solução aplicada após a análise:

Remoção do LACP
 Criação de 2 Subinterfaces
 Uma interface encaminhando na fila Low
 Outra interface encaminhando na fila High



Foi observado que o tráfego ficou bem distribuído entre os dois buffers

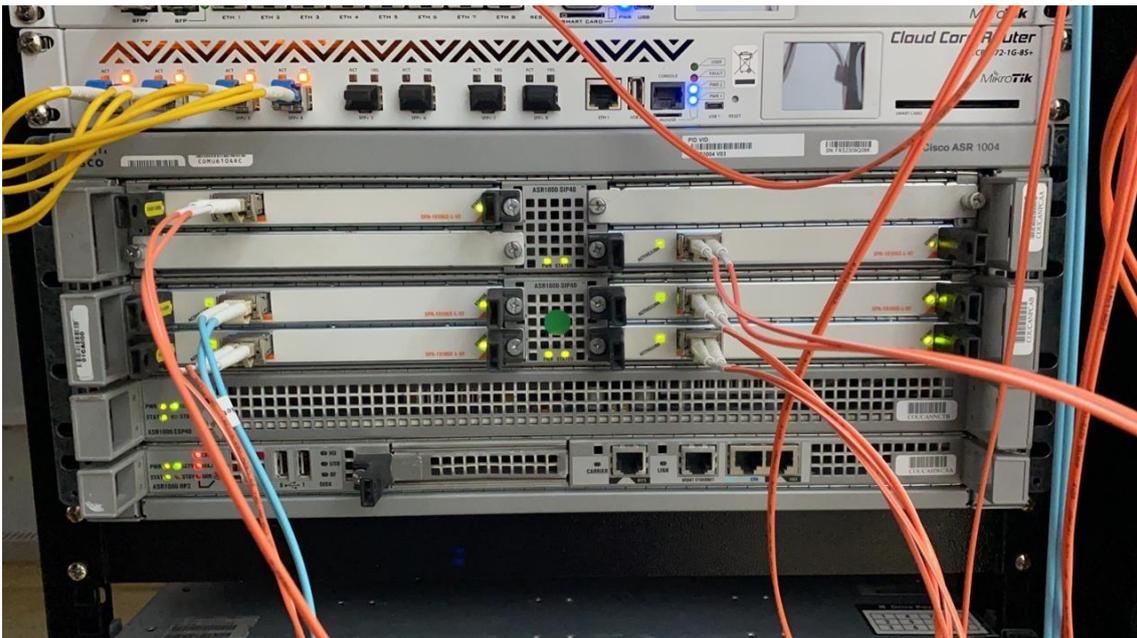
```

CPP 0: Subdev 0      5 secs      1 min      5 min      60 min
Input: Priority (pps) 547418      543705      546279      629104
      (bps) 6037014832 6002256128 6025692280 6990049632
      Non-Priority (pps) 817240      830831      851007      941724
      (bps) 4830878792 4966380880 5117376472 5584521288
      Total (pps) 1364658      1374536      1397286      1570828
      (bps) 10867893624 10968637008 11143068752 12574570920
Output: Priority (pps) 270      266      266      270
      (bps) 150112      148320      149680      152752
      Non-Priority (pps) 1348100      1357164      1379188      1551384
      (bps) 10725963888 10819194912 10990417936 12408546864
      Total (pps) 1348370      1357430      1379454      1551654
      (bps) 10726114000 10819343232 10990567616 12408699616
Processing: Load (pct) 38      38      38      43
  
```

Também não houve mais incremento de overruns na interface, e com isso não ocorreram descartes.

Teste de tráfego realidade após a alteração dos buffers:

Foi adicionado mais um modulo SIP40 e duas SPA10 no sistema assim ficando com 3x10G de IN e 3x10G de OUT



```

bras-cda-c#show platform hardware qfp active datapath utilization
CPP 0: Subdev 0      5 secs      1 min      5 min      60 min
Input: Priority (pps) 1058697      1053592      1015193      672022
      (bps) 11722044624 11649368344 11184137824 6852364560
      Non-Priority (pps) 1510011      1507385      1365893      904195
      (bps) 13631995328 13602260880 12041891632 6644139368
      Total (pps) 2568708      2560977      2381086      1576217
      (bps) 25354039952 25251629224 23226029456 13496503928
Output: Priority (pps) 331      346      342      344
      (bps) 172064      179936      177392      178936
      Non-Priority (pps) 2397272      2394227      2256034      1543272
      (bps) 23493001024 23445935992 21899584712 13199774888
      Total (pps) 2397603      2394573      2256376      1543616
      (bps) 23493173088 23446115928 21899762104 13199953824
Processing: Load (pct) 64      63      58      40
  
```

```

ras-cda-c#show interfaces port-channel 1 human-readable
Port-channel1 is up, line protocol is up
Hardware is 10GChannel, address is 5c71.0d3f.ebc0 (bia 5c71.0d3f.ebc0)
Description: Po1-UPLINK-HUAWEI
MTU 9216 bytes, BW 30000000 Kbit/sec, DLY 10 usec,
    reliability 255/255, txload 6/255, rxload 205/255
Encapsulation 802.1Q Virtual LAN, Vlan ID 1., loopback not set
Keepalive set (10 sec)
ARP type: ARPA, ARP Timeout 04:00:00
  No. of active members in this channel: 3
    Member 0 : TenGigabitEthernet0/0/0 , Full-duplex, 10000Mb/s
    Member 1 : TenGigabitEthernet0/1/0 , Full-duplex, 10000Mb/s
    Member 2 : TenGigabitEthernet1/0/0 , Full-duplex, 10000Mb/s
  No. of PF_JUMBO supported members in this channel : 3
Last input 00:00:43, output never, output hang never
Last clearing of "show interface" counters 00:01:21
Input queue: 0/97152/0/0 (size/max/drops/flushes); Total output drops: 0
Queueing strategy: fifo
Output queue: 0/120 (size/max)
30 second input rate 24.96 giga-bits/sec , 2.13 Mpps
30 second output rate 751.52 mega-bits/sec , 418.57 Kpps
168,782,051 packets input, 238,104,916,747 bytes, 0 no buffer
Received 0 broadcasts (0 IP multicasts)
0 runs, 0 giants, 0 throttles
0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored
0 watchdog, 19 multicast, 0 pause input
33,477,621 packets output, 7,376,534,165 bytes, 0 underruns
0 output errors, 0 collisions, 0 interface resets
1 unknown protocol drops
0 babbles, 0 late collision, 0 deferred
0 lost carrier, 0 no carrier, 0 pause output
0 output buffer failures, 0 output buffers swapped out

```

Teste de tráfego sintético:

```

Port-channel1 is up, line protocol is up
Hardware is 10GChannel, address is 5c71.0d3f.ebc0 (bia 5c71.0d3f.ebc0)
Description: Po1-UPLINK-HUAWEI
MTU 9216 bytes, BW 30000000 Kbit/sec, DLY 10 usec,
    reliability 247/255, txload 10/255, rxload 239/255
Encapsulation 802.1Q Virtual LAN, Vlan ID 1., loopback not set
Keepalive set (10 sec)
ARP type: ARPA, ARP Timeout 04:00:00
  No. of active members in this channel: 3
    Member 0 : TenGigabitEthernet0/0/0 , Full-duplex, 10000Mb/s
    Member 1 : TenGigabitEthernet0/1/0 , Full-duplex, 10000Mb/s
    Member 2 : TenGigabitEthernet1/0/0 , Full-duplex, 10000Mb/s
  No. of PF_JUMBO supported members in this channel : 3
Last input 00:00:37, output never, output hang never
Last clearing of "show interface" counters 00:01:56
Input queue: 0/97152/0/0 (size/max/drops/flushes); Total output drops: 0
Queueing strategy: fifo
Output queue: 0/120 (size/max)
30 second input rate 28.58 giga-bits/sec , 2.58 Mpps
30 second output rate 1.24 giga-bits/sec , 785.40 Kpps
293,41,248 packets input, 402,841,889,021 bytes, 0 no buffer
Received 0 broadcasts (0 IP multicasts)
0 runs, 0 giants, 0 throttles
3,805,808 input errors, 0 CRC, 0 frame, 8,805,808 overrun, 0 ignored
0 watchdog, 26 multicast, 0 pause input
93,829,353 packets output, 19,114,107,392 bytes, 0 underruns
0 output errors, 0 collisions, 0 interface resets
2 unknown protocol drops
0 babbles, 0 late collision, 0 deferred
0 lost carrier, 0 no carrier, 0 pause output
0 output buffer failures, 0 output buffers swapped out
bras-cda-c#show platform hardware qfp active datapath utilization
CPP 0: Subdev 0      5 secs      1 min      5 min      60 min
Input: Priority (pps) 1302337      1295972    1248434    1007307
      (bps) 13719521104 13678883680 13072851544 10092692464
  Non-Priority (pps) 1980141      1989382    1750117    1387473
      (bps) 15293903800 15257278408 12816381832 9780487672
      Total (pps) 3282478      3285354    2998551    2394780
      (bps) 29013424904 28936162088 25889233376 19873180136
Output: Priority (pps) 370      350      348      351
      (bps) 192552      182720    181072    183456
  Non-Priority (pps) 2954227      2991910    2858530    2360868
      (bps) 25425862408 25735247040 24432805464 19602875032
      Total (pps) 2954597      2992260    2858878    2361219
      (bps) 25426054960 25735429760 24432986536 19603058488
Processing: Load (pct) 87      87      79      62

```