

Comportamento do FlexPod Nexus 5k em vPC durante a interrupção

Contents

[Introduction](#)

[Topologia](#)

[Testes](#)

[Perda de enlace de dados](#)

[Atualização ou recarregamento com interrupções](#)

[Link peer do vPC desativado](#)

[Atualização de software em serviço \(ISSU\)](#)

[Problemas conhecidos com ISSU](#)

[Perda do link de keepalive de peer](#)

[Desativação do recurso vPC](#)

[Conclusão](#)

[Discussões relacionadas da comunidade de suporte da Cisco](#)

Introduction

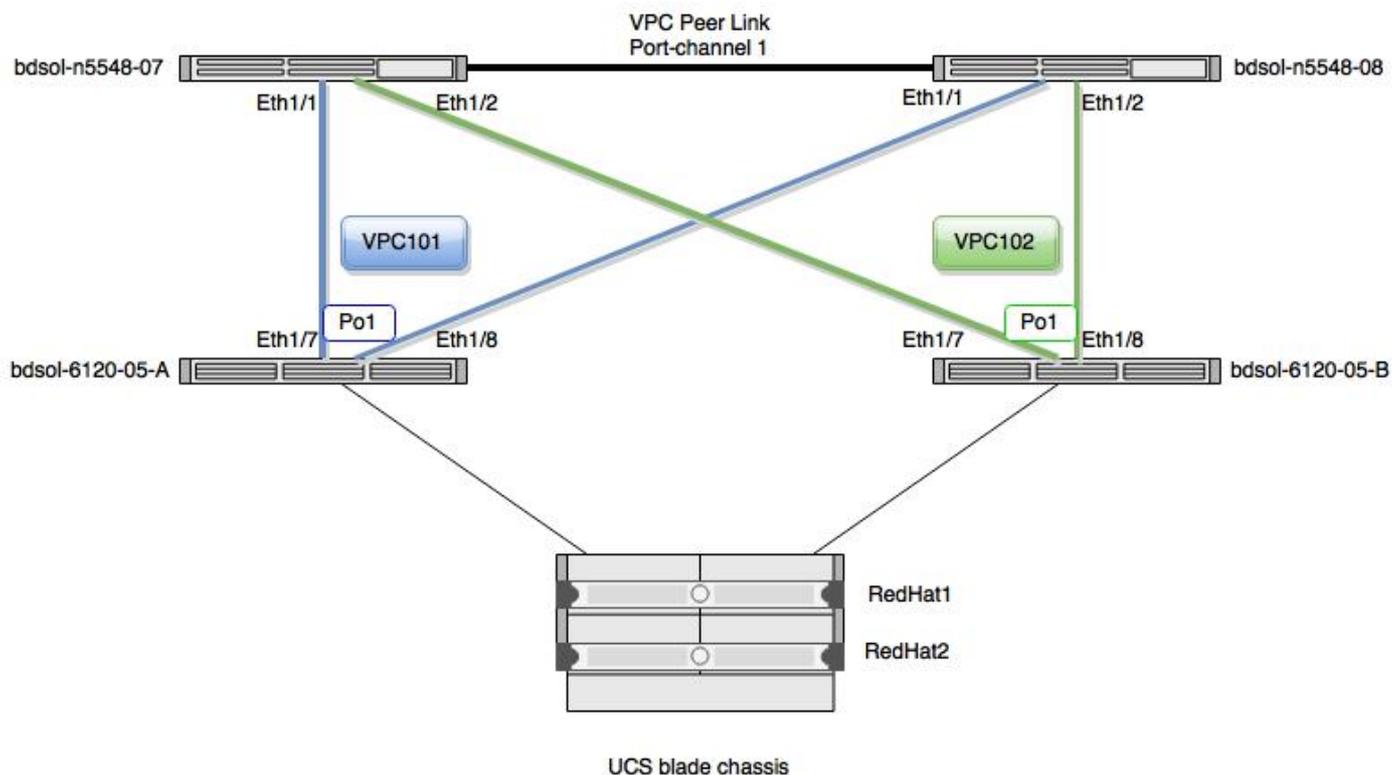
Este documento descreverá o comportamento da rede em reação a diferentes interrupções, concentrando-se no Virtual Port-Channel (vPC).

Uma interrupção típica seria: recarga, perda de link ou perda de conectividade.

O objetivo deste documento é demonstrar a perda de pacotes durante cenários comuns.

Topologia

Durante o teste, a menos que haja outra indicação de topologia a seguir.



As linhas verde e azul indicam um canal de porta vPC de cada uma das interconexões de estrutura para ambos os switches Nexus.

A rede de gerenciamento fora de banda não está delineada.

É uma topologia simplificada comumente recomendada em implantações FlexPod, como visto, por exemplo, em:

http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/flexpod_esxi51_ucsm2.html

Componentes Utilizados

Dois switches Nexus 5548P.

Duas interconexões em malha do Unified Computing System (UCS) 6120 executando software 2.2(4b).

Um chassi 5108 UCS.

Dois blades B200M3 com adaptador VIC 1240 executando software 2.2(4).

Para executar e verificar os testes de conectividade, dois blades foram instalados e o sistema operacional RedHat Enterprise Linux 7.1 está instalado.

Configuração.

A configuração do vPC e do portchannel está usando o padrão.

```
vpc domain 75
role priority 3000
peer-keepalive destination 10.48.43.79 source 10.48.43.78
delay restore 150
peer-gateway
```

```
interface port-channel1
description vPC Peer-Link
switchport mode trunk
spanning-tree port type network
vpc peer-link
```

Exemplo de vPC levando ao UCS Fabric Interconnect (FI) neste caso bdsol-6120-05—A

```
interface port-channel101
description bdsol-6120-05-A
switchport mode trunk
spanning-tree port type edge trunk
vpc 101
```

Testes

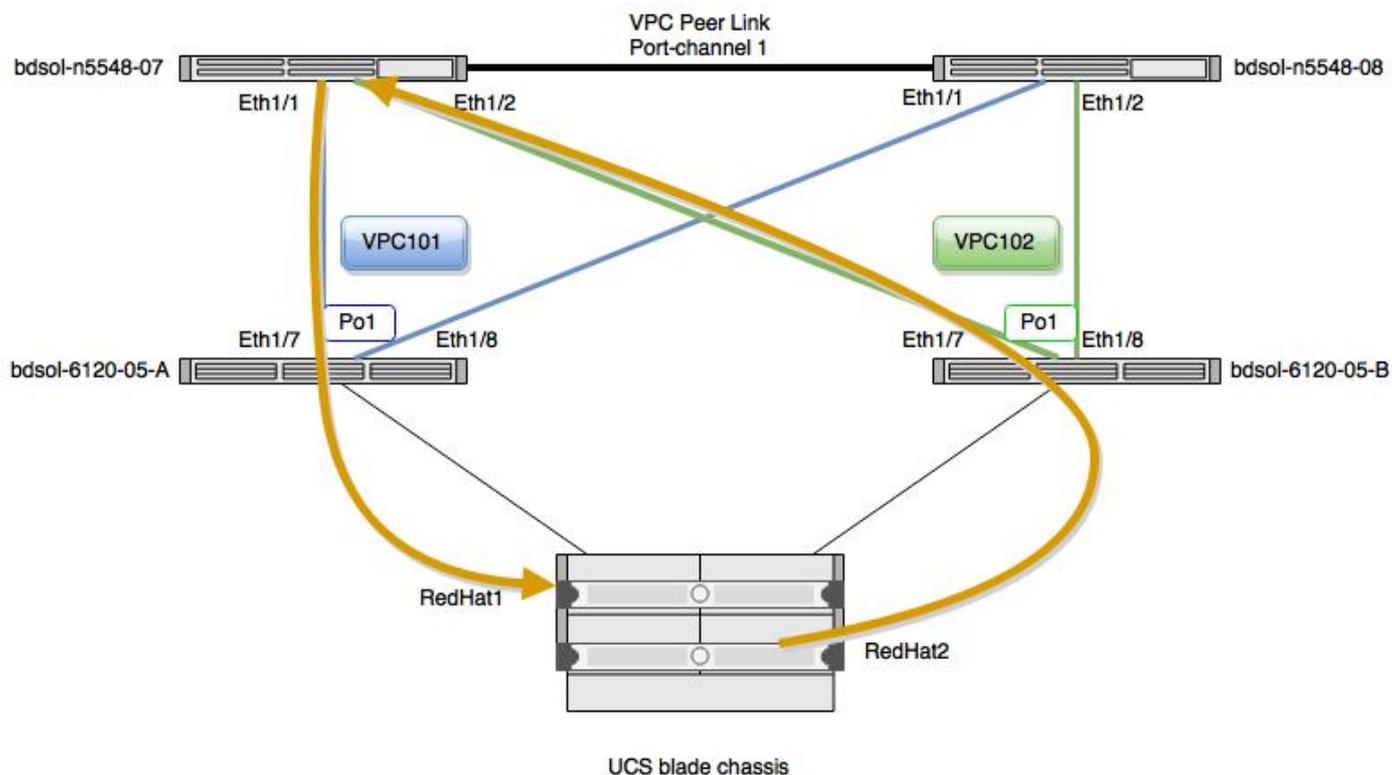
O seguinte teste será realizado.

- Perda de enlace de dados.
- Atualização com interrupções
- Atualização de software em serviço (ISSU)
- Perda de peer keepalive link - interface mgmt0 no caso dessa topologia/configuração.
- Perda de peer portchannel - Port-channel 1 nesta configuração.
- Desabilitando o recurso vPC

Fluxo de tráfego básico.

Uma única sessão iperf3 é usada para gerar 6,5 gigabits por segundo de tráfego TCP de teste para verificar a perda de quadros durante as transições.

O RedHat2 é fixado ao Fabric Interconnect B, enquanto o RedHat1 é fixado ao fabric interconnect A - isso resulta em tráfego que precisa cruzar a parte de switching.



Parâmetros Iperf3:

- Servidor: `iperf3 -s -i 1`
- Cliente `iperf3 -c 10.37.9.131 -t 0 -i 1 -w 1M -V`

Os parâmetros acima foram escolhidos para permitir alta taxa de tráfego e fácil detectar a perda de pacotes.

A janela TCP é pressionada para evitar rajadas de dados pelas quais o iperf é conhecido. Permitir que o iperf execute o modo não apertado pode resultar em quedas ocasionais em buffers de entrada ao longo do caminho - dependendo da configuração de QoS. Os parâmetros acima permitem uma taxa sustentada de 6 a 7 Gbps sem perda de quadros.

Para verificar, podemos verificar a taxa cumulativa de tráfego nas interfaces.

```
bdsol-n5548-07# show interface ethernet 1/1-2 | i rate
30 seconds input rate 5612504 bits/sec, 9473 packets/sec
30 seconds output rate 7037817832 bits/sec, 578016 packets/sec
input rate 5.60 Mbps, 9.38 Kpps; output rate 7.01 Gbps, 576.10 Kpps
30 seconds input rate 7037805336 bits/sec, 578001 packets/sec
30 seconds output rate 5626064 bits/sec, 9489 packets/sec
input rate 7.01 Gbps, 575.71 Kpps; output rate 6.56 Mbps, 9.79 Kpps
```

A saída acima mostra 7 Gbps de tráfego entrando na interface Ethernet 1/2 e deixando na interface Ethernet 1/1.

Perda de enlace de dados

Este teste é designado para testar como os dados se comportarão se um link que faz parte do vPC for desligado.

Este exemplo usará a Ethernet 1/1, a interface de saída para o tráfego de dados, e será desligado usando a linha de comando.

```
bdsol-n5548-07# conf t
Enter configuration commands, one per line. End with CNTL/Z.
bdsol-n5548-07(config)# int et1/1
bdsol-n5548-07(config-if)# shut
```

Nesse caso, apenas um único pacote foi perdido, fora de um fluxo de 6,5 Gbps.

[4]	2025.01-2026.01	sec	800	MBytes	6.71	Gbits/sec	0	272	KBytes
[4]	2026.01-2027.01	sec	800	MBytes	6.71	Gbits/sec	0	272	KBytes
[4]	2027.01-2028.01	sec	801	MBytes	6.72	Gbits/sec	0	272	KBytes
[4]	2028.01-2029.01	sec	798	MBytes	6.69	Gbits/sec	0	272	KBytes
[4]	2029.01-2030.01	sec	799	MBytes	6.70	Gbits/sec	0	272	KBytes
[4]	2030.01-2031.01	sec	799	MBytes	6.71	Gbits/sec	0	272	KBytes
[4]	2031.01-2032.01	sec	799	MBytes	6.70	Gbits/sec	0	272	KBytes
[4]	2032.01-2033.01	sec	799	MBytes	6.70	Gbits/sec	0	272	KBytes
[4]	2033.01-2034.01	sec	800	MBytes	6.71	Gbits/sec	0	272	KBytes
[4]	2034.01-2035.01	sec	799	MBytes	6.70	Gbits/sec	0	272	KBytes
[4]	2035.01-2036.01	sec	799	MBytes	6.70	Gbits/sec	0	272	KBytes
[4]	2036.01-2037.01	sec	634	MBytes	5.32	Gbits/sec	1	272	KBytes
[4]	2037.01-2038.01	sec	792	MBytes	6.65	Gbits/sec	0	272	KBytes
[4]	2038.01-2039.01	sec	795	MBytes	6.67	Gbits/sec	0	272	KBytes

O tráfego é quase imediatamente balanceado entre os links restantes no portchannel no UCS, nesse caso usando a porta Ethernet 1/8 (a única restante) do UCS FI B até o Nexus 5548 B, daí ele será transportado para o UCS FI A usando Ethernet 1/1.

```
bdsol-n5548-08# show interface ethernet 1/1-2 | i rate
30 seconds input rate 5575896 bits/sec, 9413 packets/sec
30 seconds output rate 6995947064 bits/sec, 574567 packets/sec
input rate 2.21 Mbps, 3.70 Kpps; output rate 2.78 Gbps, 227.99 Kpps
30 seconds input rate 6995940736 bits/sec, 574562 packets/sec
30 seconds output rate 5581920 bits/sec, 9418 packets/sec
input rate 2.78 Gbps, 227.99 Kpps; output rate 2.22 Mbps, 3.71 Kpps
```

Atualização ou recarregamento com interrupções

Uma interrupção combinada de dados e plano de controle pode ser emulada executando uma atualização revolucionária do bdsol-n5548-07 (vPC principal).

Espera-se perda de tráfego.

Funcionalmente, esse teste é o mesmo que recarregar um peer vPC.

```
bdsol-n5548-07# install all kickstart bootflash:n5000-uk9-kickstart.7.1.0.N1.1a.bin system
bootflash:n5000-uk9.7.1.0.N1.1a.bin
(...)
```

```
Compatibility check is done:
Module bootable Impact Install-type Reason
-----
1 yes disruptive reset Incompatible image
(...)
```

Switch will be reloaded for disruptive upgrade.

Do you want to continue with the installation (y/n)? [n] y

Install is in progress, please wait.

Performing runtime checks.

[#####] 100% -- SUCCESS

Setting boot variables.

[#####] 100% -- SUCCESS

Performing configuration copy.

[#####] 100% -- SUCCESS

Finishing the upgrade, switch will reboot in 10 seconds.

Após os 10 segundos mencionados, ocorre a perda de pacotes.

Durante esse período, somente 55 pacotes são perdidos (fora do fluxo de 6,6 Gbps).

```
[ 4] 3571.01-3572.01 sec 800 MBytes 6.71 Gbits/sec 0 272 KBytes
[ 4] 3572.01-3573.01 sec 801 MBytes 6.72 Gbits/sec 0 272 KBytes
[ 4] 3573.01-3574.01 sec 800 MBytes 6.71 Gbits/sec 0 272 KBytes
[ 4] 3574.01-3575.01 sec 638 MBytes 5.35 Gbits/sec 55 211 KBytes
[ 4] 3575.01-3576.01 sec 805 MBytes 6.75 Gbits/sec 0 242 KBytes
[ 4] 3576.01-3577.01 sec 801 MBytes 6.72 Gbits/sec 0 260 KBytes
[ 4] 3577.01-3578.01 sec 801 MBytes 6.72 Gbits/sec 0 269 KBytes
[ 4] 3578.01-3579.01 sec 799 MBytes 6.70 Gbits/sec 0 272 KBytes
[ 4] 3579.01-3580.01 sec 797 MBytes 6.68 Gbits/sec 0 272 KBytes
[ 4] 3580.01-3581.01 sec 800 MBytes 6.71 Gbits/sec 0 272 KBytes
^C[ 4] 3581.01-3581.83 sec 656 MBytes 6.71 Gbits/sec 0 272 KBytes
-----
Test Complete. Summary Results:
[ ID] Interval          Transfer          Bandwidth          Retr
[ 4]  0.00-3581.83 sec 3002125194048 bits 6.71 Gbits/sec    55
[ 4]  0.00-3581.83 sec  0.00 Bytes 0.00 bits/sec
CPU Utilization: local/sender 34.6% (0.5%u/34.1%s), remote/receiver 0.0% (0.0%u/0.0%s)
iperf3: interrupt - the client has terminated
```

Se o iperf3 foi reiniciado imediatamente, o operador pode verificar se o tráfego realmente mudou para bdsol-n5548-08.

```
bdsol-n5548-08# show interface ethernet 1/1-2 | i rate
30 seconds input rate 5601392 bits/sec, 9455 packets/sec
30 seconds output rate 7015307760 bits/sec, 576159 packets/sec
input rate 2.25 Mbps, 3.77 Kpps; output rate 2.81 Gbps, 231.14 Kpps
30 seconds input rate 7015303696 bits/sec, 576152 packets/sec
30 seconds output rate 5605280 bits/sec, 9462 packets/sec
input rate 2.81 Gbps, 231.14 Kpps; output rate 2.25 Mbps, 3.77 Kpps
```

A taxa de tráfego está mostrando abaixo de 6 Gbps, já que o contador de taxa tem uma média de 30 segundos.

Link peer do vPC desativado

Neste exemplo, o link do peer do vPC é desativado, disparado por uma alteração de configuração.

Nesse momento, o tráfego é tratado pelo bdsol-n5548-07, atuando como secundário do vPC.

A sequência de eventos.

O canal de porta 1 fica inativo.

2015 Jul 10 15:00:25 bdsol-n5548-07 %ETHPORT-5-IF_DOWN_CFG_CHANGE: A interface port-channel1 está inoperante (alteração de configuração)

Como o bdsol-n5548-07 está agindo como secundário, ele suspenderá seus vPCs, pois não pode garantir a topologia sem problemas:

```
2015 Jul 10 15:00:28 bdsol-n5548-07 %VPC-2-VPC_SUSP_ALL_VPC: Peer-link going down, suspending all vPCs on secondary
2015 Jul 10 15:00:28 bdsol-n5548-07 %ETHPORT-5-IF_DOWN_INITIALIZING: Interface port-channel928 is down (Initializing)
2015 Jul 10 15:00:28 bdsol-n5548-07 %ETHPORT-5-IF_DOWN_INITIALIZING: Interface port-channel102 is down (Initializing)
2015 Jul 10 15:00:28 bdsol-n5548-07 %ETHPORT-5-IF_DOWN_INITIALIZING: Interface port-channel101 is down (Initializing)
```

Durante esse período, o iperf3 perdeu uma parte do tráfego - 90 pacotes.

```
[ 41] 5871.01-5872.01 sec 798 MBytes 6.70 Gbits/sec 0 272 KBytes
[ 41] 5872.01-5873.01 sec 798 MBytes 6.70 Gbits/sec 0 272 KBytes
[ 41] 5873.01-5874.01 sec 801 MBytes 6.72 Gbits/sec 0 272 KBytes
[ 41] 5874.01-5875.01 sec 801 MBytes 6.72 Gbits/sec 0 272 KBytes
[ 41] 5875.01-5876.01 sec 800 MBytes 6.71 Gbits/sec 0 272 KBytes
[ 41] 5876.01-5877.01 sec 796 MBytes 6.68 Gbits/sec 0 272 KBytes
[ 41] 5877.01-5878.01 sec 796 MBytes 6.68 Gbits/sec 0 272 KBytes
[ 41] 5878.01-5879.01 sec 599 MBytes 5.03 Gbits/sec 90 272 KBytes
[ 41] 5879.01-5880.01 sec 800 MBytes 6.71 Gbits/sec 0 272 KBytes
[ 41] 5880.01-5881.01 sec 799 MBytes 6.70 Gbits/sec 0 272 KBytes
[ 41] 5881.01-5882.01 sec 798 MBytes 6.70 Gbits/sec 0 272 KBytes
[ 41] 5882.01-5883.01 sec 798 MBytes 6.69 Gbits/sec 0 272 KBytes
[ 41] 5883.01-5884.01 sec 801 MBytes 6.72 Gbits/sec 0 272 KBytes
[ 41] 5884.01-5885.01 sec 800 MBytes 6.71 Gbits/sec 0 272 KBytes
```

Mas foi capaz de se recuperar bem rápido.

Como os vPCs estão suspensos em bdsol-n5548-07, todo o tráfego é tratado por bdsol-n5548-08

```
bdsol-n5548-08# show int ethernet 1/1-2 | i rate
30 seconds input rate 5623248 bits/sec, 9489 packets/sec
30 seconds output rate 7036030160 bits/sec, 577861 packets/sec
input rate 2.83 Mbps, 4.74 Kpps; output rate 3.54 Gbps, 290.64 Kpps
30 seconds input rate 7036025712 bits/sec, 577854 packets/sec
30 seconds output rate 5627216 bits/sec, 9498 packets/sec
input rate 3.54 Gbps, 290.64 Kpps; output rate 2.83 Mbps, 4.75 Kpps
```

Novamente, a taxa não mostra 6,5 gigabits por segundo imediatamente devido ao cálculo da média de carga.

Recuperação do link vPC inativo.

Quando o link par do vPC volta ativo, o tráfego pode ser rebalanceado entre os enlaces e pode ser esperada uma perda de pacote de vida curta devido à alteração na topologia.

No caso deste teste de laboratório, 1 pacote foi perdido.

Atualização de software em serviço (ISSU)

Neste teste, uma atualização de ISSU foi realizada para verificar a interrupção do tráfego.

As funções do vPC durante este teste são as seguintes:

bdsol-n5548-07 - principal

bdsol-n5548-08 - secundário.

Para executar um critério definido por ISSU, é necessário que seja atendido.

Para encontrar informações sobre os comandos usados para verificar esses critérios e executar um ISSU, foi usado o seguinte guia:

http://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus5500/sw/upgrade/705_N1_1/n5500_upgrade_downgrade_700.html#pgfid-727913.

Depois de executar um ISSU primeiro no primário e depois no peer vPC secundário, nenhum pacote foi perdido.

Isso se deve ao fato de que ISSU todas as funcionalidades do plano de dados permanecem ininterruptas e somente o tráfego do plano de controle será afetado.

```
[ 4] 1096.01-1097.01 sec 798 MBytes 6.69 Gbits/sec 0 272 KBytes
[ 4] 1097.01-1098.01 sec 800 MBytes 6.71 Gbits/sec 0 272 KBytes
[ 4] 1098.01-1099.01 sec 798 MBytes 6.69 Gbits/sec 0 272 KBytes
[ 4] 1099.01-1100.01 sec 799 MBytes 6.70 Gbits/sec 0 272 KBytes
^C[ 4] 1100.01-1100.71 sec 563 MBytes 6.72 Gbits/sec 0 272 KBytes
-----
Test Complete. Summary Results:
[ ID] Interval Transfer Bandwidth Retr
[ 4] 0.00-1100.71 sec 856 GBytes 6.68 Gbits/sec 0 sender
[ 4] 0.00-1100.71 sec 0.00 Bytes 0.00 bits/sec receiver
CPU Utilization: local/sender 35.0% (0.5%u/34.5% s), remote/receiver 0.0% (0.0%u/0.0% s)
iperf3: interrupt - the client has terminated
```

Problemas conhecidos com ISSU

Recursos e licenças da camada 3.

Durante o teste do ISSU, vários problemas precisaram ser resolvidos. O "show install all impact..." pode fornecer saída de que ISSU não pode ser executado com a seguinte explicação: "Não há suporte para instalação sem interrupções se L3 estiver habilitado." No ambiente de teste, isso se deve ao LAN_BASE_SERVICES_PKG sendo usado no arquivo de licença instalado.

LAN_BASE_SERVICES_PKG inclui a funcionalidade L3 e para executar o ISSU este pacote deve ser não utilizado e o arquivo de licença deve ser limpo do dispositivo usando o comando "clear license LICENSEFILE". É possível que o arquivo de licença esteja atualmente em uso pelo dispositivo. Para limpar esse arquivo de licença, é importante verificar quais pacotes estão em uso usando o "show license usage" e desabilitando os recursos desses pacotes.

Portas STP fora da borda

Durante o teste, também foi necessário desligar o canal de porta ascendente, já que ele não passou na verificação de "show spanning-tree issue-impact" non-edge, Critério 3, e isso levaria a uma atualização sem interrupções. Esse canal de porta ascendente não estava listado como um vPC Edge no comando "show spanning-tree vlan 1".

Perda do link de keepalive de peer

Após a perda do link peer keepalive mgmt0, nenhuma interrupção no tráfego foi registrada. Nesta topologia, a interface de gerenciamento (mgmt0) é usada como link keepalive, portanto, não afeta o tráfego de dados gerado durante o teste.

```
[ 41] 71.01-72.01 sec 793 MBytes 6.65 Gbits/sec 0 272 KBytes
[ 41] 72.01-73.01 sec 794 MBytes 6.66 Gbits/sec 0 272 KBytes
[ 41] 73.01-74.01 sec 791 MBytes 6.63 Gbits/sec 0 272 KBytes
[ 41] 74.01-75.01 sec 793 MBytes 6.65 Gbits/sec 0 272 KBytes
[ 41] 75.01-76.01 sec 793 MBytes 6.65 Gbits/sec 0 272 KBytes
^C[ 41] 76.01-76.02 sec 8.13 MBytes 5.83 Gbits/sec 0 272 KBytes
-----
Test Complete. Summary Results:
[ ID] Interval      Transfer      Bandwidth      Retr
[ 41] 0.00-76.02 sec 58.6 GBytes 6.62 Gbits/sec 0          sender
[ 41] 0.00-76.02 sec 0.00 Bytes 0.00 bits/sec          receiver
CPU Utilization: local/sender 35.2% (0.5%u/34.7%u), remote/receiver 0.0% (0.0%u/0.0%u)
iperf3: interrupt - the client has terminated
```

Os dispositivos percebem que a interface mgmt0 está ficando inativa e que as keepalives de peer estão falhando, mas como o link de peer está ativo, a comunicação de local de dados pode continuar.

```
2015 Jul 14 12:11:28 bdsol-n5548-07 %IM-5-IM_INTF_STATE: mgmt0 is DOWN in vdc 1
2015 Jul 14 12:11:32 bdsol-n5548-07 %VPC-2-PEER_KEEP_ALIVE_RECV_FAIL: In domain 75, VPC peer
keep-alive receive has failed
2015 Jul 14 12:12:07 bdsol-n5548-07 %IM-5-IM_INTF_STATE: mgmt0 is UP in vdc 1
```

Desativação do recurso vPC

Este teste descreverá o que acontece quando o vPC é desabilitado em um dos switches durante a transferência de dados ao vivo.

O recurso VPC pode ser desabilitado usando o seguinte comando no modo de configuração global:

```
bdsol-n5548-07(config)# no feature vpc
```

A desativação do recurso vPC no par vPC principal ou secundário leva à perda instantânea de conectividade de dados. Isso se deve à natureza baseada em peer do vPC. Assim que o recurso é desabilitado, toda a funcionalidade do vPC no switch deixa de funcionar, o link par fica inativo, o status de keepalive do vPC é Suspenso e o canal de porta 101 do ambiente de teste é desativado. Isso é evidente na saída do comando show vPC do peer switch, que ainda tem o recurso vPC ativado.

```
bdsol-n5548-07# show vpc
```

Legend:

(*) - local vPC is down, forwarding via vPC peer-link

```
vPC domain id          : 75
Peer status            : peer link is down
vPC keep-alive status  : Suspended (Destination IP not reachable)
...
```

vPC status

```
-----
id Port Status Consistency Reason Active vlans
-----
```

```
101 Po101 down success success -
```

A interrupção do tráfego, como antes, tem duração curta.

Nas condições de teste mencionadas acima, 50 a 80 pacotes foram perdidos de uma única sessão.

Remover o comando "feature vpc" também fez com que a configuração do vPC fosse removida dos canais de porta.

Essa configuração precisa ser lida.

Conclusão

O recurso vPC tem o objetivo de oferecer desempenho de resiliência dividindo o tráfego de dados em um canal de porta entre vários dispositivos.

Essa ideia simples requer implementações complicadas de plano de controle.

Os testes acima foram feitos para mostrar interrupções no plano de controle e de dados que podem ocorrer durante o ciclo de vida do recurso.

Como esperado, as interrupções do plano de dados foram detectadas e corrigidas quase imediatamente - com pacotes únicos perdidos em testes.

As interrupções do plano de controle testadas mostram que o vPC ainda mantém o tempo de convergência de menos de um segundo, mesmo quando o plano de controle é afetado.

O teste mais impactante realizado - link de peer do vPC sendo desligado - potencialmente combina dados e falha do plano de controle. Ainda foi demonstrado um tempo de convergência rápido.