Identificar e solucionar problemas de sincronia do BGP entre o núcleo de pacotes ultra e o switch Nexus devido à configuração incorreta

Contents

Introduction

Este documento descreve a solução para os flaps de Border Gateway Protocol (BGP) entre o Cisco Ultra Packet Core (UPC) e o switch Nexus 9000 configurado com a conexão BGP redundante.

Problema

Flaps de BGP são acionados quando uma das interfaces redundantes entre o Cisco Ultra Packet Core e o switch Nexus oscila.

Condições

O nó UPC (Ultra Packet Core) está conectado à folha A e à folha B do Nexus em portas separadas. Os peers BGP IPv6 são estabelecidos e as rotas padrão são instaladas no nó UPC. A Figura 1 mostra o diagrama de rede de alto nível com caminho redundante para switches Leaf.

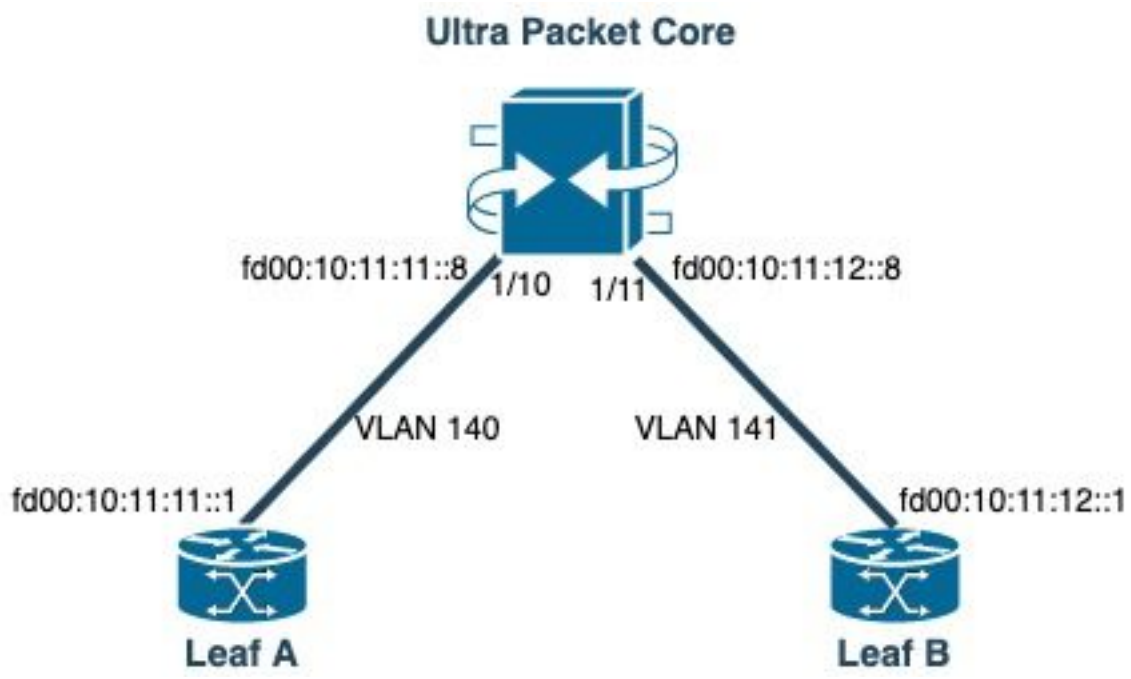

Figura 1: Diagrama de rede

Configuração

Configuração de porta UPC com VLAN e ligação de interface:

```
port ethernet 1/10
    no shutdown
    vlan 140
        no shutdown
        bind interface saegw_vlan140_1/10 saegw
#exit

#exit
port ethernet 1/11
    no shutdown
    vlan 141
        no shutdown
        bind interface saegw_vlan141_1/11 saegw
#exit
#exit
end
```

Configuração da interface UPC com endereços IP:

```
interface saegw_vlan140_1/10
  ip address 10.11.11..8 255.255.255.0
  ipv6 address fd00:10:11:11::8/64 secondary
  bfd interval 300 min_rx 300 multiplier 3
#exit
interface saegw_vlan141_1/11
  ip address 10.11.12.8 255.255.255.0
  ipv6 address fd00:10:11:12::8/64 secondary
  bfd interval 300 min_rx 300 multiplier 3
#exit
```

Configuração de BGP UPC:

```
router bgp 25949
  router-id 172.19.20.30
  maximum-paths ebgp 4
  neighbor 10.11.11..1 remote-as 25949
  neighbor 10.11.11..1 fall-over bfd
  neighbor 10.11.12.1 remote-as 25949
  neighbor 10.11.12.1 fall-over bfd
  neighbor fd00:10:11:11::1 remote-as 25949
  neighbor fd00:10:11:12::1 remote-as 25949
  address-family ipv4
    neighbor 10.11.11..1 route-map accept_default in
    neighbor 10.11.11..1 route-map gw-1-OUT out
    neighbor 10.11.12.1 route-map accept_default in
    neighbor 10.11.12.1 route-map gw-1-OUT out
    redistribute connected
#exit
address-family ipv6
  neighbor fd00:10:11:11::1 activate
  neighbor fd00:10:11:11::1 route-map accept_v6_default in
  neighbor fd00:10:11:11::1 route-map allow_service_ips_v6 out
  neighbor fd00:10:11:12::1 activate
  neighbor fd00:10:11:12::1 route-map accept_v6_default in
  neighbor fd00:10:11:12::1 route-map allow_service_ips_v6 out
  redistribute connected
#exit

ipv6 prefix-list name accept_v6_default_routes seq 10 permit ::/0
route-map accept_v6_default permit 10
  match ipv6 address prefix-list accept_v6_default_routes
#exit
```

Configuração do switch Nexus 9000:

```
Interface vlan140
ipv6 address fd00:10:11:11::1/64
no ipv6 redirects

interface vlan141
ipv6 address fd00:10:11:12::1/64
no ipv6 redirects

vrf upc
address-family ipv4 unicast
advertise l2vpn evpn
maximum-paths ibgp 2
address-family ipv6 unicast
advertise l2vpn evpn
maximum-paths ibgp 2
neighbor fd00:10:11:12::5
remote-as 25949
address-family ipv6 unicast
neighbor fd00:10:11:12::6
remote-as 25949
address-family ipv6 unicast
neighbor fd00:10:11:12::8
remote-as 25949
address-family ipv6 unicast
```

Análise

Inicialmente, uma comunicação BGP normal entre uma das interfaces UPC (fd00:10:11:12::8) e o switch Nexus (fd00:10:11:12::1 pertence à vlan141) é observada e inclui mensagens TCP ACK:

```
2023-01-01 01:01:59.000000 fd00:10:11:12::8 -> fd00:10:11:12::1 TCP 35813 > bgp [ACK] Seq=250
Ack=8664 Win=31744 Len=0 TSV=2412344062 TSER=531234647
2023-01-01 01:01:59.000087 fd00:10:11:12::8 -> fd00:10:11:12::1 TCP 35813 > bgp [ACK] Seq=250
Ack=11520 Win=37376 Len=0 TSV=2412344062 TSER=531234647
2023-01-01 01:01:59.000162 fd00:10:11:12::8 -> fd00:10:11:12::1 TCP 35813 > bgp [ACK] Seq=250
Ack=14376 Win=43008 Len=0 TSV=241234062 TSER=531234647
2023-01-01 01:01:59.000281 fd00:10:11:12::8 -> fd00:10:11:12::1 TCP 35813 > bgp [ACK] Seq=250
Ack=17232 Win=49152 Len=0 TSV=2412344062 TSER=531234647
2023-01-01 01:01:59.000936 fd00:10:11:12::8 -> fd00:10:11:12::1 TCP 35813 > bgp [ACK] Seq=250
Ack=20663 Win=48640 Len=0 TSV=2412344063 TSER=531234647
```

Em caso de falha da interface Leaf-B em direção ao UPC, um comportamento incorreto é visto nos registros onde uma nova tentativa de conexão BGP é iniciada pelo UPC (source: fd00:10:11:12::8) em direção ao Leaf-A na interface fd00:10:11:11::1, que pertence a uma VLAN diferente, vlan140.

```
2023-01-01 22:36:12.370117 fd00:10:11:12::8 -> fd00:10:11:11::1 TCP 41987 > bgp [SYN] Seq=0
Win=14400 Len=0 MSS=1440 TSV=2412347369 TSER=0 WS=9
```

Essa mensagem BGP SYN inválida enviada na interface incorreta resulta na inatividade do BGP. Quando o Nexus anuncia sua própria rota conectada e o UPC recebe uma rota para a interface que estava inoperante sobre o BGP, o UPC tenta a conexão através de outra interface com um IP de saída diferente/errado.

Solução

Devido à configuração referida na seção Condição deste artigo, como o UPC recebe as informações de rota conectada de ambos os Leafs de ambas as interfaces, quando uma das interfaces está inativa, o UPC tenta se comunicar com essa Leaf através da outra interface.

Para evitar que o UPC envie mensagens de estabelecimento de conexão BGP da interface errada, aqui estão as alterações de configuração a serem consideradas:

1. Na configuração UPC, adicione update-source para o vizinho. Essa configuração impede a conexão BGP de uma interface diferente, se a interface principal estiver inoperante. Por exemplo, quando saegw_vlan140_1/10 (fd00:10:11:11::1/64) está desativado, o nó não pode usar a interface de saída saegw_vlan141_1/11 para o par BGP fd00:10:11:11::8.
   Esta é uma configuração de exemplo:

   ```
   neighbor fd00:10:11:11::1 update-source fd00:10:11:11::8
   neighbor fd00:10:11:12::1 update-source fd00:10:11:12::8
   ```
2. Na configuração do Nexus, bloqueie os prefixos das interfaces erradas.
   Por exemplo, negamos rotas para a folha redundante sobre o vizinho fd00:10:11:11::1

   ```
   neighbor fd00:10:11:11::1
   update prefix list to deny fd00:10:11:12::8/64
   ```
3. No switch Nexus, o peering EBGP do VTEP para um nó externo sobre VXLAN deve estar em um VRF de locatário e deve usar o update-source de um loopback interface (peering sobre VXLAN), conforme recomendado no Guia de configuração do Cisco Nexus 9000

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.