RUB

RUHR-UNIVERSITÄT BOCHUM

CONFIGURING VPC FABRIC PEERING

in a VXLAN-EVPN fabric with Cisco Nexus 9k

Introduction

- Robin Därmann
- Technically responsible for the data network of the Ruhr-University Bochum
- NOC Team with seven members
- Campus network with ~3.000 devices, ~140.000 Ethernet access ports, ~2.300 wireless APs (>1.000 more to come)
- A lot of self-made automation (perl mostly)
- Datacenter built in 2021/2022 with 210 Nexus 9300/9500 switches
  - (will get DIN EN 50600 certified)

About vPC Fabric Peering

- vPC = „virtual Port-Channel“
- Multichassis Link-Aggregation feature of Cisco Nexus series / NX-OS
- Traditionally, connection of two switches / management planes with direct attached links

- vPC Fabric Peering eliminates the need for direct links between the switches

- First shipped 2019 in NX-OS 9.2(3)
  - We are running 9.3(10)

Advantages

- No direct links between vPC member switches, traffic goes through the underlay network instead
- Additional ports available on every ToR switch
- Cost savings, as you don‘t need the cables and/or transceivers for direct connection
  - Saved us about 50.000 EUR (~200 cables * 250 EUR)
- Simple configuration
- vPC Keep-Alive link can also run over the fabric (simply use loopback addresses)
  - ...but we use direct links between the mgmt interfaces instead

Disadvantages

- vPC Port-channel can‘t be used as a backup uplink anymore
- Requires QoS on the spine layer
  - vPC traffic is important, vPC domains will break in situations of packet loss
- Only few platforms are supported (only Nexus 9300-FX/FXP/FX2/FX3)
- Not supported on virtual Nexus 9000v, testing requires real hardware

vPC behaviour changes for VXLAN-EVPN

- Anycast VTEP
  - For Dual-homed hosts, an anycast IP address is used in EVPN
  - Shared (secondary) loopback IP address on both vPC members

- advertise-pip
  - Command in the BGP l2vpn evpn address-family
  - Use the primary VTEP IP address for single-homed hosts
  - Traffic gets directed to the correct vPC member switch and does not need to traverse the (virtual) peer-link

Configuration – QoS on spines (N9K-9508)

```
class-map type qos match-all CFS-class
  match dscp 56

policy-map type qos CFS-pmap
  class CFS-class
    set qos-group 7

interface Ethernet x/y
  service-policy type qos input CFS-pmap
```

Configuration – leaf preparation

- TCAM carving is necessary:

```
hardware access-list tcam region ing-racl 0
hardware access-list tcam region ing-flow-redirect 512
```

→ **Reload the leafs afterwards!**

Configuration – vPC Domain on leafs

- Peer-Link goes between underlay loopback addresses of the leafs

Leaf 1:

```
vpc domain 3

  virtual peer-link destination 10.96.0.4 source 10.96.0.3 dscp 56
```

Leaf 2:

```
vpc domain 3

  virtual peer-link destination 10.96.0.3 source 10.96.0.4 dscp 56
```

vPC Configuration with differences on both member switches

```
feature vpc

vpc domain 3
  virtual peer-link destination 10.96.0.4 source 10.96.0.3 dscp 56
                                                  virtual peer-link destination 10.96.0.3 source 10.96.0.4 dscp 56
interface port-channel500
  description vPC Peer Link
  switchport
  switchport mode trunk
  spanning-tree port type network
  vpc peer-link

interface nve1
  advertise virtual-rmac
  source-interface loopback1

interface loopback0
  description Routing loopback interface
  ip address 10.96.0.3/32                         ip address 10.96.0.4/32

interface loopback1
  description VTEP loopback interface
  ip address 10.96.2.3/32                         ip address 10.96.2.4/32
  ip address 10.96.4.3/32 secondary

router bgp 65000
  address-family l2vpn evpn
    advertise-pip
```

Alternatives

- EVPN-Multihoming (EVPN-MH)
  - Not at well/at all supported by Cisco/on our gear
  - Only older devices with broadcom chipsets seem to be supported
- Arista seems to support EVPN-MH
- Juniper too
- Dell? Does anyone use Dell switches in a datacenter and can tell?

That‘s it

- Questions?
- Contact: robin.daermann@rub.de
- Cisco Documentation: https://l.rub.de/22a904ab