Seamless Data Center Interconnect options using EVPN & various tunnel stitching ID 4469

Michal Styszynski

October 18th, 2022 – NANOG 86, Hollywood, CA

Agenda

- Why ‘Seamless EVPN-VxLAN stitching’ ?
- L2 and L3 DCI Overlay design options
- Seamless EVPN-VxLAN stitching – packet walk
  - EVPN-VXLAN to EVPN-VXLAN EVPN Type-2 & Type-5
  - EVPN-VXLAN to EVPN-MPLS

Why seamless tunnel stitching in DC and DCI use-case ?

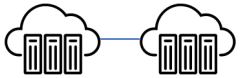
Multi-pod or DCI evpn-vxlan tunnel scale control	The vxlan-to-vxlan stitching allows to reduce and control the number of vxlan tunnels between DC rooms and sites, when running EVPN-VXLAN fabric
Easier architecture extensions when workloads are growing	Adding another pod into the existing DC architecture is simpler with no impact on the existing pod leaf operation
More efficient Ethernet flooding	Better flooding between DC rooms or sites when just selected L2& L3 domain are extended
Additional virtualization & security option	Each DC site can have different set of VNIs to secure tenant while still deliver same broadcast domain
Improved operations	The operator/admin can decide which workloads get extended

Main use-case for tunnel stitching

Seamless tunnel stitching

1 multi-pod L2 / L3 DC

- larger dc fabrics
with selective L2 VNI and RVNI stretch

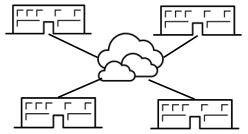


Multi-Pod L2/L3
Data Center

- EVPN-VXLAN to EVPN-VXLAN
- Server overlays gateway to the fabric

2 L2 / L3 DCI

- multiple DC sites (2 or more)
- medium/large DC



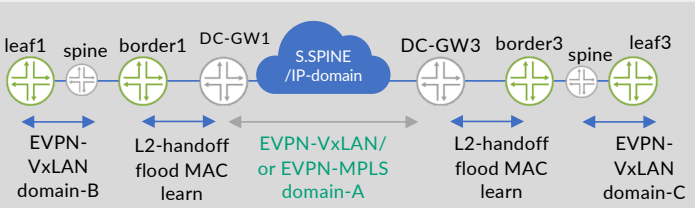
Multi-site L2/L3 DCI

- EVPN-VXLAN to EVPN-VXLAN
- EVPN-VXLAN to EVPN-MPLS
- EVPN-VXLAN to EVPN-SRv6

Simplified DCI design options

Option 1

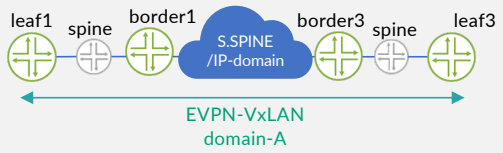
Decoupled Interconnect model using L2 VLAN handoff – L2 DCI



- Harder to manage - more devices/equipment
- Traditional demarcation points
- Flood based MAC learning

Option 2

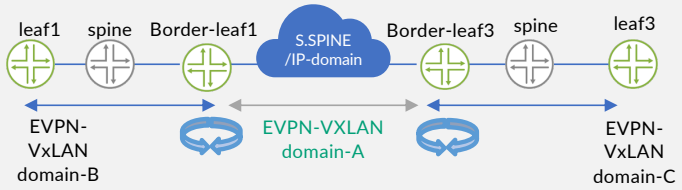
OTT – over the top full-mesh interconnect – L2 DCI



- Intermediate scale and control
- Lack of clear demarcation points
- One EVPN-VXLAN domain between DC sites

Option 3

Seamless EVPN-VxLAN to EVPN-VXLAN seamless stitching - L2 DCI



- **Unified EVPN layer 2 solution**
- **Pure overlay with controlled scaling**
- **Higher scale—more DC sites and pods**
- **rfc9014**

NANOG™

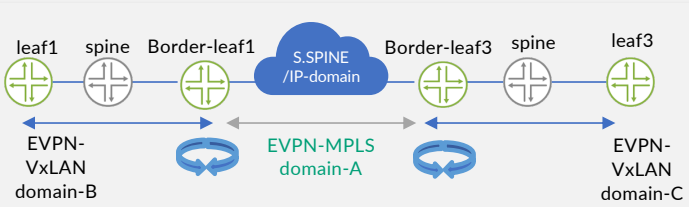
© 2022 Juniper Networks

Juniper Networks

Simplified DCI design options

Option 4

Seamless EVPN-VxLAN to EVPN-MPLS seamless stitching – L2 DCI

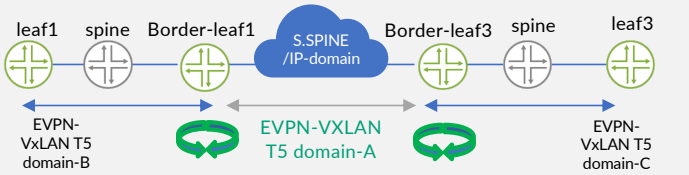


- Unified EVPN layer 2 solution
- Pure overlay with controlled scaling
- Higher scale—more DC sites and pods
- MPLS direct connect from border-leaf

Option 5

EVPN Type-5 to EVPN Type-5 VxLAN to VxLAN stitching

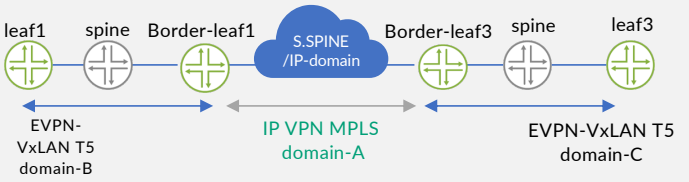
L3 DCI



- Unified EVPN layer 3 solution
- Pure overlay with controlled scaling
- Higher scale—more DC sites and pods
- VXLAN for IP only DCI

Option 6

EVPN-VXLAN Type-5 to IPVPN MPLS – L3 DCI



- Unified EVPN layer 3 solution
- Pure overlay with controlled scaling
- Higher scale—more DC sites and pods
- MPLS IPVPN using IP only DCI

© 2022 Juniper Networks

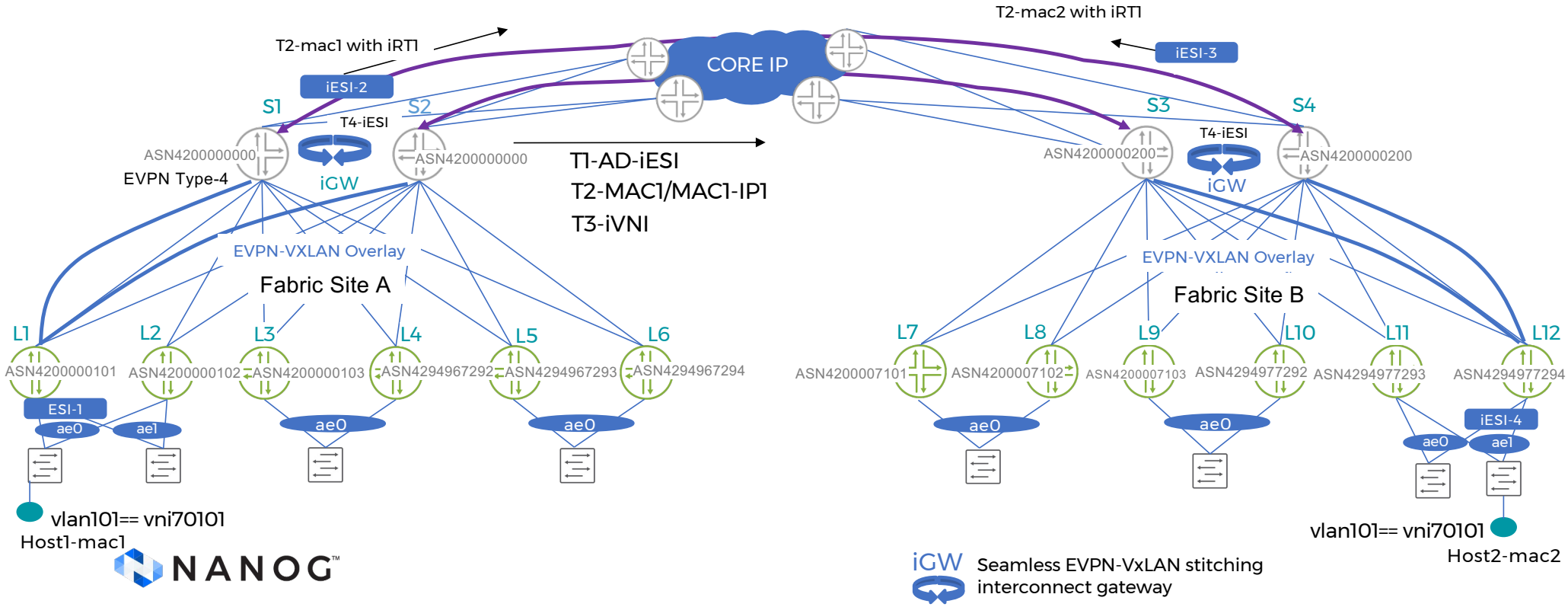
EVPN-VXLAN to EVPN-VXLAN tunnel stitching – EVPN T2/T2 & T5/T5

NANOG™

Juniper Public

SEAMLESS EVPN-VXLAN STITCHING - USING BORDER SPINE

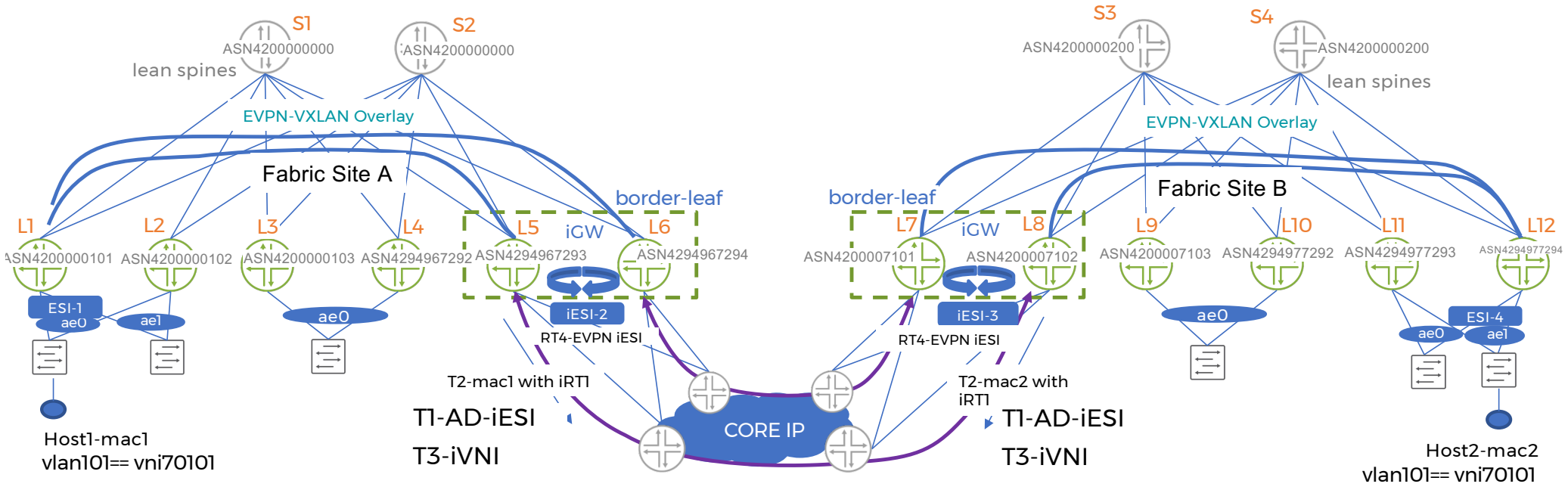
- By default, the NH of the MAC@ is changed in case of seamless stitching overlay DCI
- RT/RD rewrite operation for RT1-->RT3



NANOG™

SEAMLESS EVPN-VXLAN STITCHING - USING BORDER-LEAF

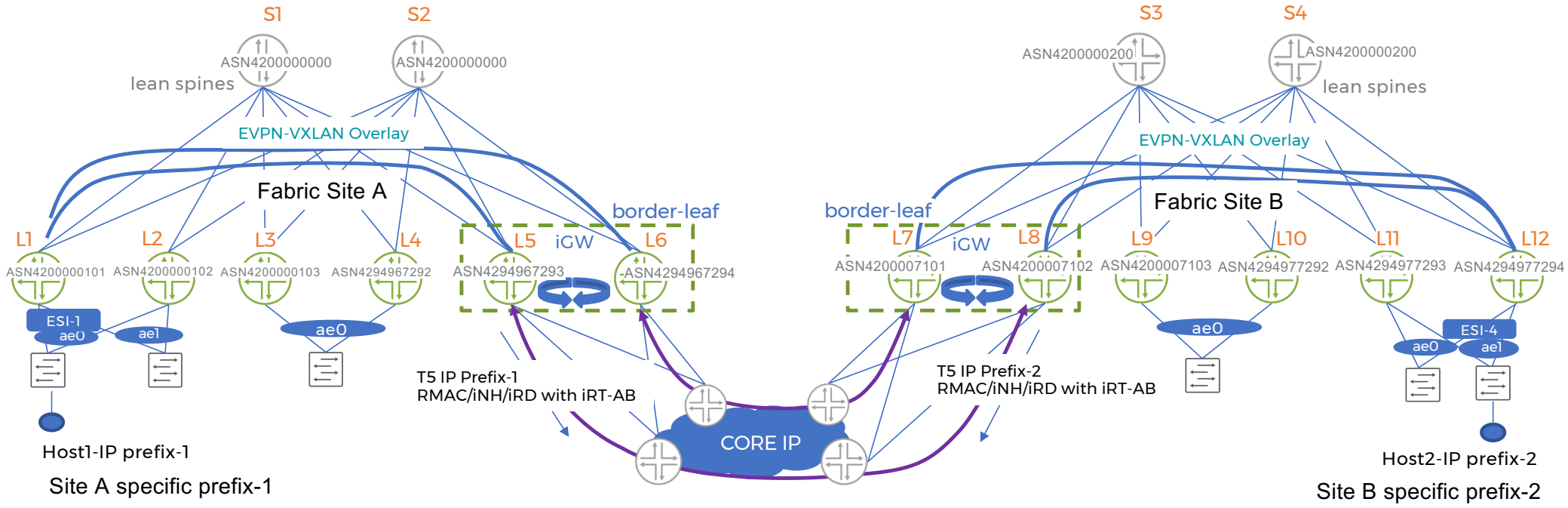
- By default, the NH of the MAC@ is changed in case of seamless stitching overlay DCI
- RT/RD rewrite operation for RT1-->RT3
- RT4 for DF/nDF election per iESI



iGW Seamless EVPN-VxLAN stitching interconnect gateway

SEAMLESS EVPN-VXLAN STITCHING - PURE TYPE-5 EVPN

- By default, the NH of the IP prefix is changed in case of seamless stitching overlay DCI
- RT/RD rewrite operation for EVPN RT5
- No local IFL interfaces creations



NANOG™

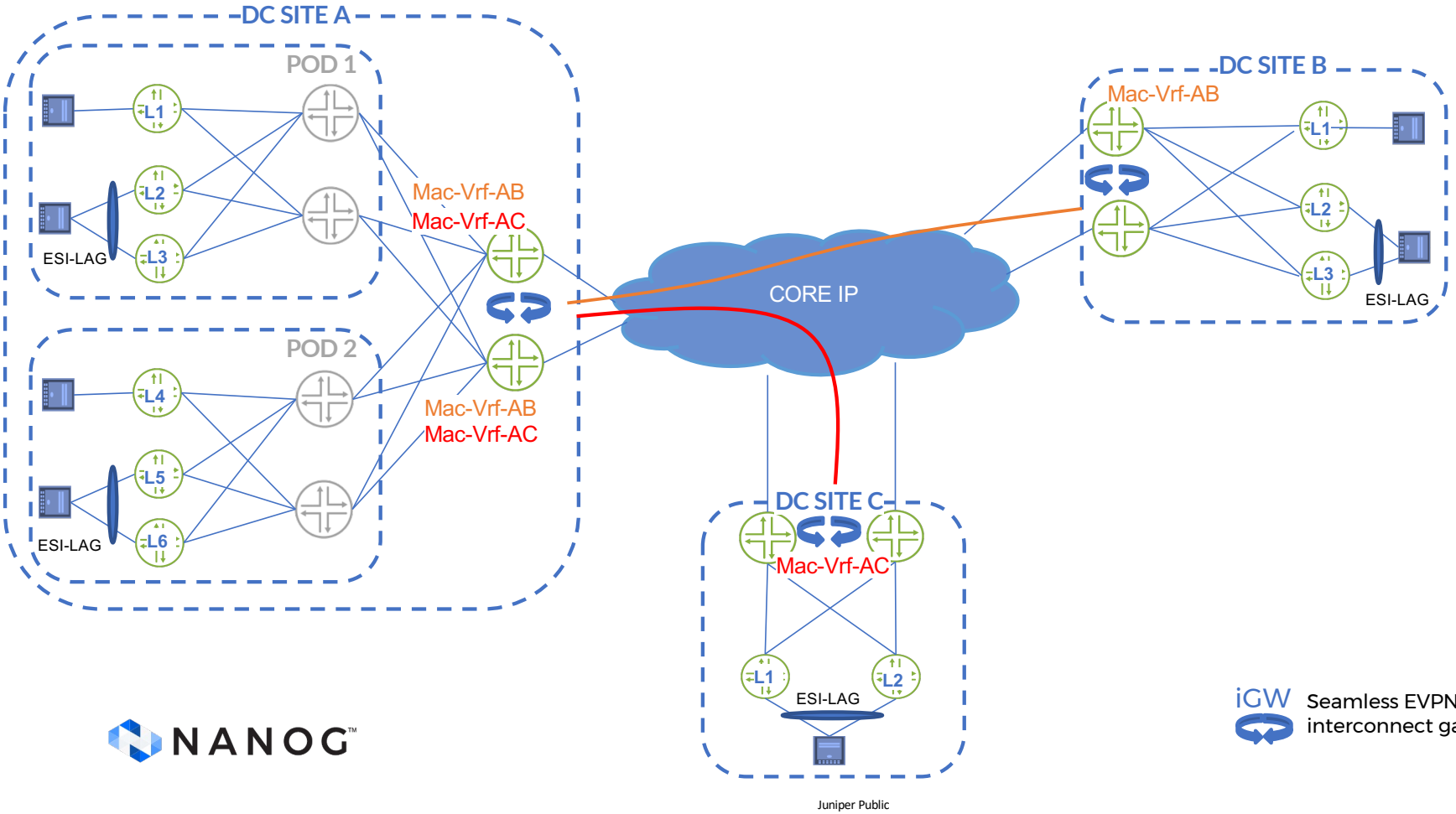
iGW Seamless EVPN-VxLAN stitching interconnect gateway

Juniper Public

MAC-VRF
and Seamless tunnel stitching

Juniper Public

MAC-VRF ISOLATION and SEAMLESS TUNNEL STITCHING



iGW Seamless EVPN-VxLAN stitching interconnect gateway

NANOG™

Type-2 to Type-2 EVPN-VXLAN stitching implementation example inside MAC-VRF (EVI)

```
macvrfAB-vlan-aware {
    instance-type mac-vrf;
    protocols {
        evpn {
            encapsulation vxlan;
            default-gateway no-gateway-community;
            extended-vni-list [ 5055 5056 ];
            interconnect {
                vrf-target target:1:55;
                route-distinguisher 172.16.7.11:55;
                esi {
                    00:00:99:99:99:99:99:99:55:55;
                    all-active;
                }
                interconnected-vni-list [ 6055 ];
            }
        }
    }
    vtep-source-interface lo0.0;
    service-type vlan-aware;
    route-distinguisher 172.16.7.11:155;
    vrf-target target:1:155;
    vlans {
        vlan55 {
            vlan-id 55;
            vxlan {
                vni 5055;
                translation-vni 6055;
            }
        }
        vlan56 {
            vlan-id 56;
            vxlan {
                vni 5056;
            }
        }
    }
}
```

New tunnel stitching configuration block

VNI translation specific to interconnect function

Type-5 to Type-5 EVPN-VXLAN stitching implementation example inside Type-5 IP VRF

```
root@border-leaf1> show configuration routing-instances T5-VRF1
instance-type vrf;
routing-options {
    static {
        route 10.10.100.113/32 discard;
    }
    multipath;
}
protocols {
    evpn {
        interconnect {
            vrf-target target:102:102;
            route-distinguisher 172.16.7.113:102;
        }
        ip-prefix-routes {
            advertise direct-nexthop;
            encapsulation vxlan;
            vni 1100;
            export my-t5-export-vrf1;
        }
    }
}
interface lo0.1;
route-distinguisher 172.16.7.113:100;
vrf-target target:1100:1100;
vrf-table-label;

root@border-leaf1>
```

DCI new dedicated block in Type-5 instance

```
root@border-leaf2> show configuration routing-instances T5-VRF1
instance-type vrf;
routing-options {
    static {
        route 10.10.100.114/32 discard;
    }
    multipath;
}
protocols {
    evpn {
        interconnect {
            vrf-target target:102:102;
            route-distinguisher 172.16.7.114:102;
        }
        ip-prefix-routes {
            advertise direct-nexthop;
            encapsulation vxlan;
            vni 1100;
            export my-t5-export-vrf1;
        }
    }
}
interface lo0.1;
route-distinguisher 172.16.7.114:100;
vrf-target target:1100:1100;
vrf-table-label;

root@border-leaf2>
```

EVPN-VXLAN stitching – MAC origin and Endpoint type

```
root@border-leaf4> show evpn database instance macvrf101 origin dci-remote
Instance: macvrf101
VLAN  DomainId  MAC address        Active source                  Timestamp        IP address
     61001      00:00:01:01:01:01  00:00:11:11:11:11:11:11:11:11  Sep 09 11:59:42  10.10.0.1
     61001      00:50:56:ab:01:01  00:00:11:11:11:11:11:11:11:11  Sep 09 11:59:42  10.10.0.101
     61001      00:50:56:ab:01:02  00:00:11:11:11:11:11:11:11:11
     61001      00:50:56:ab:01:03  00:00:11:11:11:11:11:11:11:11
     61002      00:00:01:01:01:02  00:00:11:11:11:11:11:11:11:11
     61002      c0:03:80:1c:7e:e0  00:00:11:11:11:11:11:11:11:11
     61002      c0:03:80:1c:b5:e0  00:00:11:11:11:11:11:11:11:11  Sep 09 11:59:42  10.10.1.1

root@border-leaf4>
```

All remote-DC MAC@ behind same remote iESI

Interconnect ESI enabled on remote-site

```
root@border-leaf4# run show mac-vrf forwarding vxlan-tunnel-end-point remote summary
Logical System Name       Id  SVTEP-IP         IFL   L3-Idx    SVTEP-Mode    ELP-SVTEP-IP
<default>                 0   172.16.7.216     lo0.0    0
 RVTEP-IP      IFL-Idx     Interface    NH-Id   RVTEP-Mode  ELP-IP        Flags
 172.16.7.113    19309     vtep.32773   55058   Wan-VTEP
 172.16.7.114    19308     vtep.32772   55043   Wan-VTEP
 172.16.7.206    19307     vtep.32771   55022   RNVE
 172.16.7.215    19312     vtep.32774   55063   I-ESI-Peer
 RVTEP-IP         L2-RTT                 IFL-Idx   Interface    NH-Id   RVTEP-Mode  ELP-IP        Flags
 172.16.7.114     macvrf-techfest22      671547396 vtep-56.32772 55043  RNVE
 172.16.7.215     macvrf-techfest22      671547398 vtep-56.32774 55063  I-ESI-Peer
 172.16.7.113     macvrf101              671555589 vtep-57.32773 55058  Wan-VTEP
 172.16.7.114     macvrf101              671555588 vtep-57.32772 55043  Wan-VTEP
 172.16.7.206     macvrf101              671555587 vtep-57.32771 55022  RNVE
 172.16.7.215     macvrf101              671555590 vtep-57.32774 55063  I-ESI-Peer

[edit]
root@border-leaf4#
```

shows which VTEP is for DCI

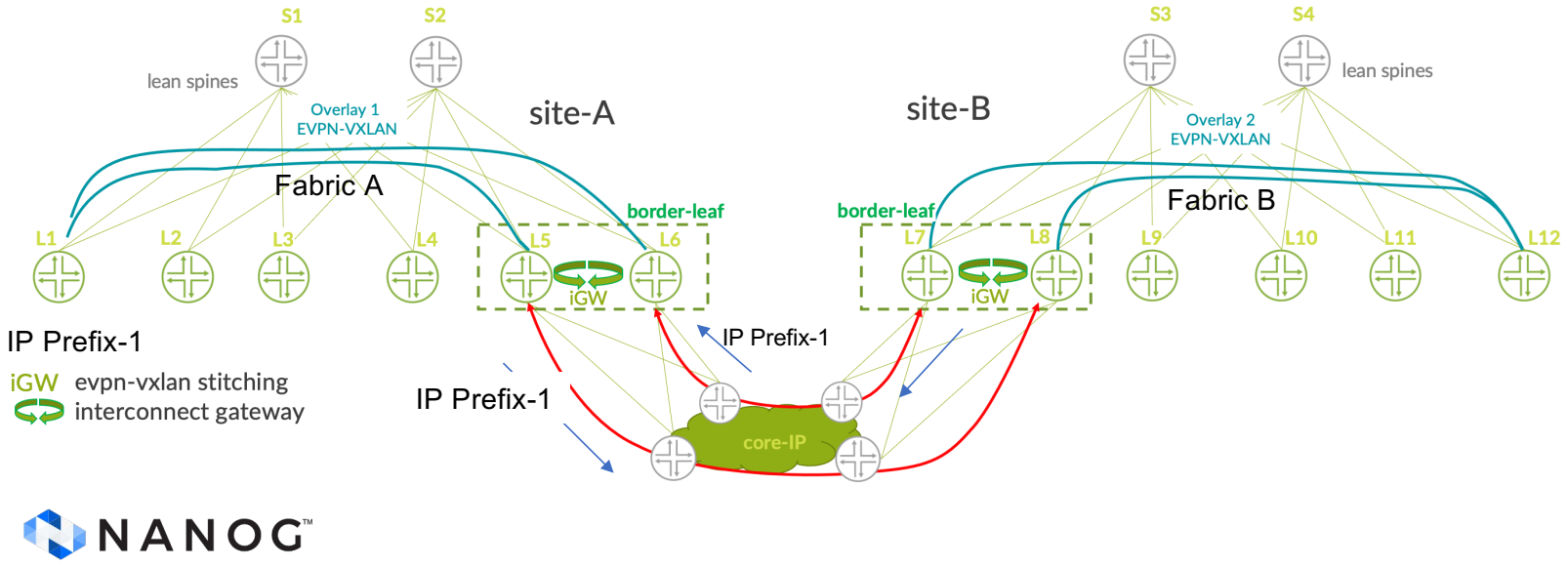
What are the main differences between the T5 and T2 tunnel stitching ?

- Type-5 stitching is not creating local IFL interfaces
- There's no interconnect ESI concept in pure Type-5
- Configuration is done inside the T5 IPVPN instance
- Load balancing is done using IP ECMP overlay in pure Type-5 (aka ipvrf-to-ipvrf / interface-less model – rfc9136)

Seamless tunnel stitching - D-PATH new BGP attribute

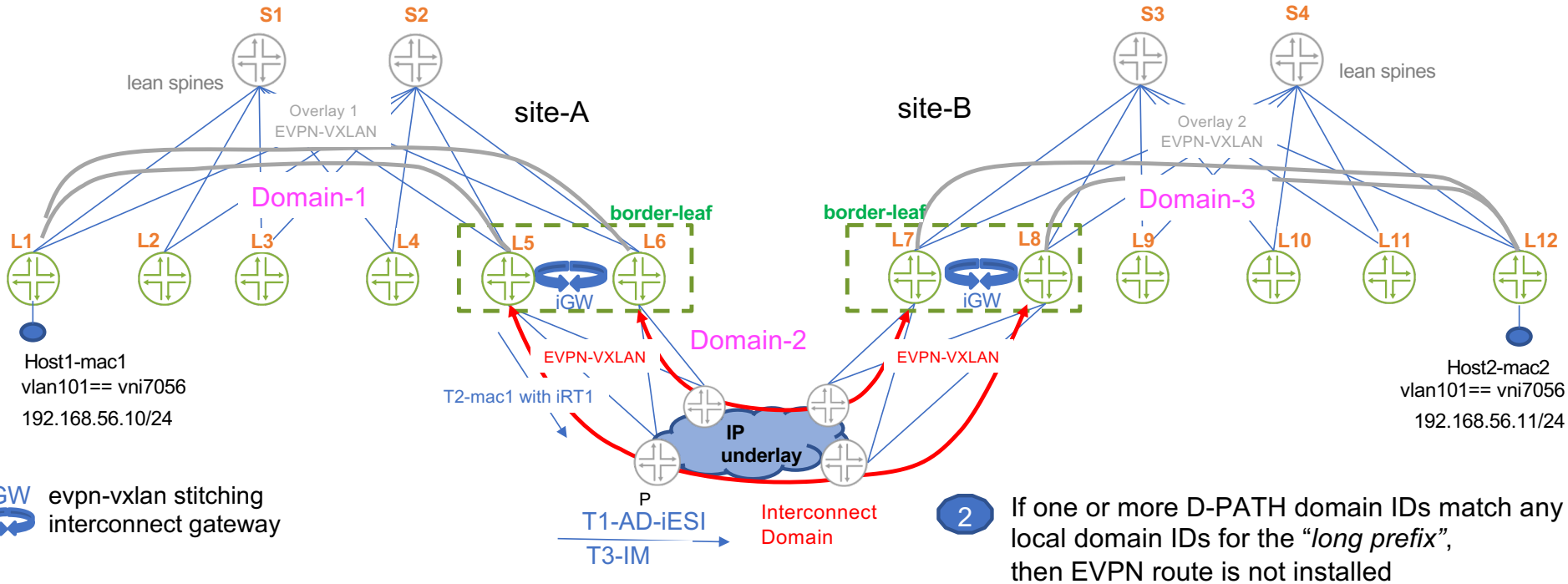
DCI EVPN & routing loops – problem statement

- EVPN routing loop may occur when routes originated in the interconnect domain are learned back from LAN side and advertised back to the interconnect (WAN) side
- Sub-optional path forwarding may be created in consequence
- Managing the problem at the community level and policy is too complex to manage from implementation point of view



D-PATH AS A NEW BGP ATTRIBUTE FOR DCI

1 All domain IDs included in the D-PATH BGP attribute are compared with the local domain IDs after the route-target verification passed



iGW evpn-vxlan stitching interconnect gateway

2 If one or more D-PATH domain IDs match any local domain IDs for the "*long prefix*", then EVPN route is not installed

NANOG™

Juniper Public

Seamless tunnel stitching summary notes

NANOG™

Juniper Public

Seamless tunnel stitching – summary notes

- Suitable for L2 and/or L3 DCI and multi-pod designs
  when LAN Fabric is running EVPN-VXLAN
- Reduces overlay network load – remote site learns DCI routes only
- Improves DC and DCI control – overlay domain creations
- Simplifies day to day operations – pre-established DCI tunnels
- Normalized scale of tunneling – local site gateways next-hops
- Integrated DCI security at the edge of the fabric
  – GBP and flood rate checks at the border-leaf level

Thank you

NANOG™ OCT 18th 2022, Hollywood, CA

Juniper Public