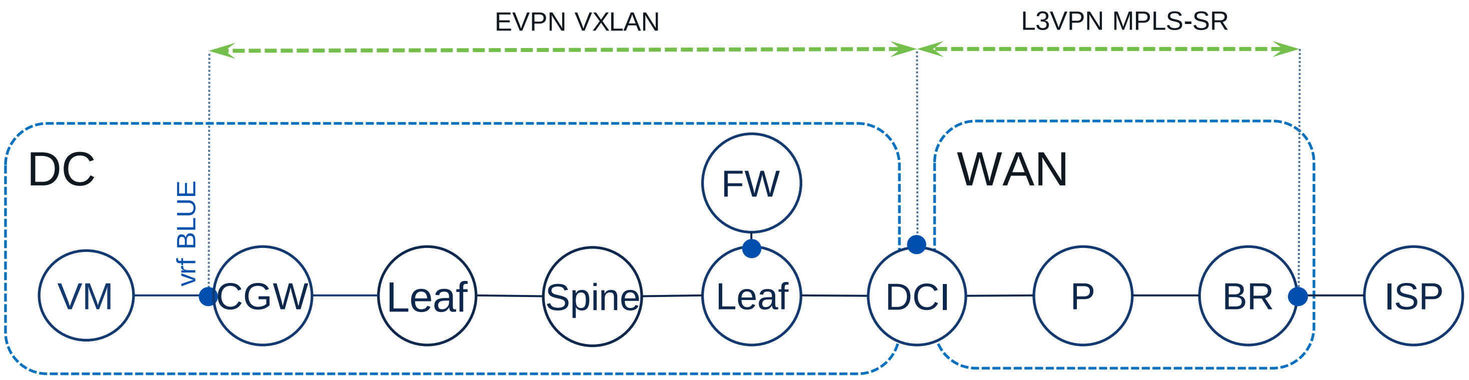
NEBIUS

Case Study: SRv6 uSID DC Frontend to Peering

Alexey Gorovoy
Network Engineer @ NEBIUS

cisco Live!

Current architecture of the Frontend network



- IPv6 only infrastructure in the Data Center
- Multivendor DC and WAN networks approach
- CGW (Cloud Gateway) and FW are NFV’s running on hosts. Nebius develops them
- VXLAN based overlay between CGW and DCI
- DCI does “stitching” between EVPN VXLAN and L3VPN MPLS-SR

Current architecture - evaluation

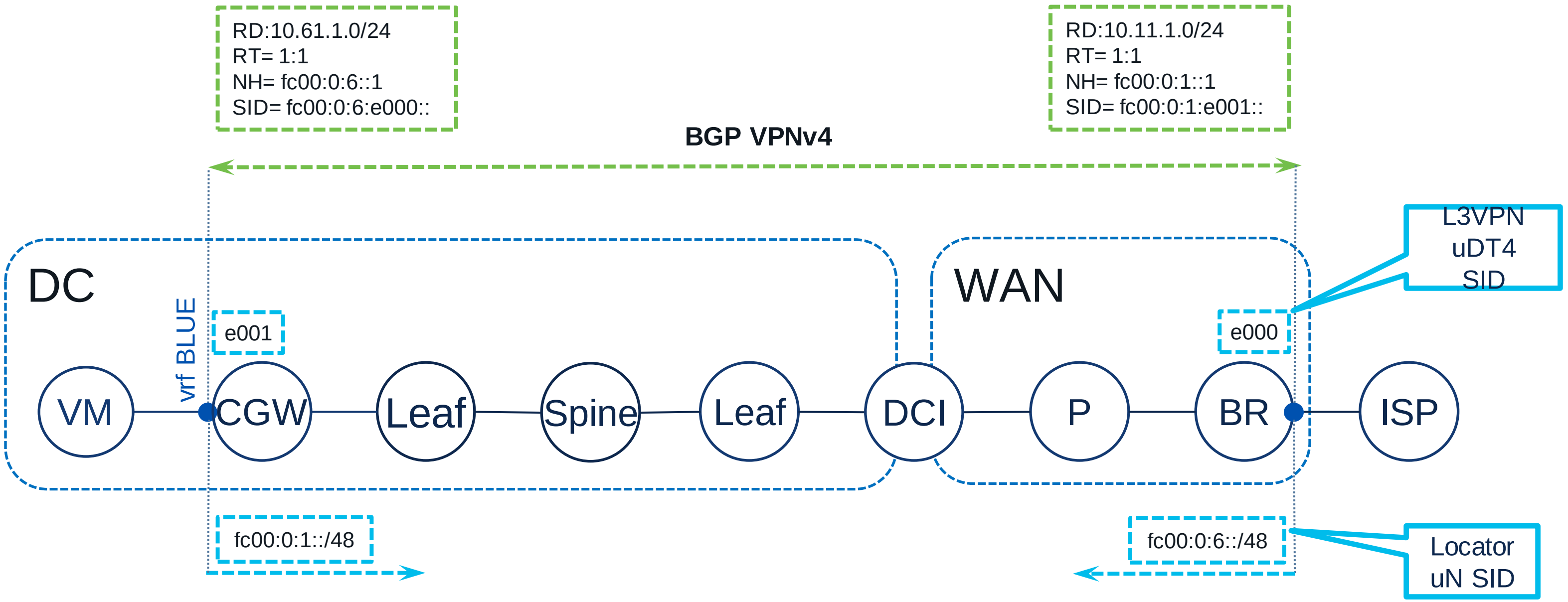
- Pros:
  - VXLAN EVPN has wide industry support and adoption
  - DC fabric is overlay agnostic thus scalable, simple and efficient
  - EVPN provides rich variety of network services
  - MPLS-SR is a mature technology with good multivendor interoperability for VPN and TE applications
- Cons:
  - No traffic engineering capabilities inside the Data Center
  - Service chaining with VXLAN requires specific routing design (PBR, Default GW, VRF/VLAN hand-off, etc.)
  - Majority vendor implementations of VXLAN still require IPv4 loopbacks in the Underlay
  - MPLS-SR lacks native Data Center optimisations and not applicable in the DC domain
  - Requires ”stitching” gateway functionality at the DCI routers to interconnect WAN and DC domains

SRv6 addresses all of them!

Transition to SRv6

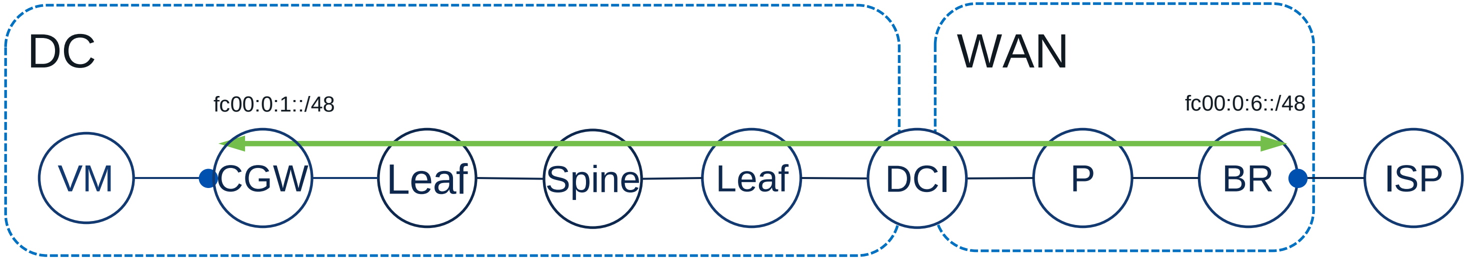
- Bridges both DC and WAN domains together in efficient and simple way
- Creates unified data plane based on IPv6 protocol only
- Allows to build end-to-end overlay service across DC and WAN without stitching fuctionality on the intermidiate devices
- Offers true traffic engineering capabilities initiated from the source of an application allowing efficient service chainings creation

Overlay with SRv6 uSID



- IPv6 in DC and WAN
- SRv6 only required on CGW and BR
- CGW and BR act as SRv6 L3VPN PEs

Overlay with SRv6 uSID – Packet walk



IPv4 | SA = 10.11.1.11
DA = 10.61.1.61
Payload

→

IPv6 | SA = fc00:0:1::1
DA = fc00:0:6:e000::
IPv4 | SA = 10.11.1.11
DA = 10.61.1.61
Payload

→

IPv4 | SA = 10.11.1.11
DA = 10.61.1.61
Payload

IPv4 | SA = 10.61.1.61
DA = 10.11.1.11
Payload

←

IPv6 | SA = fc00:0:6::1
DA = fc00:0:1:e001::
IPv4 | SA = 10.61.1.61
DA = 10.11.1.11
Payload

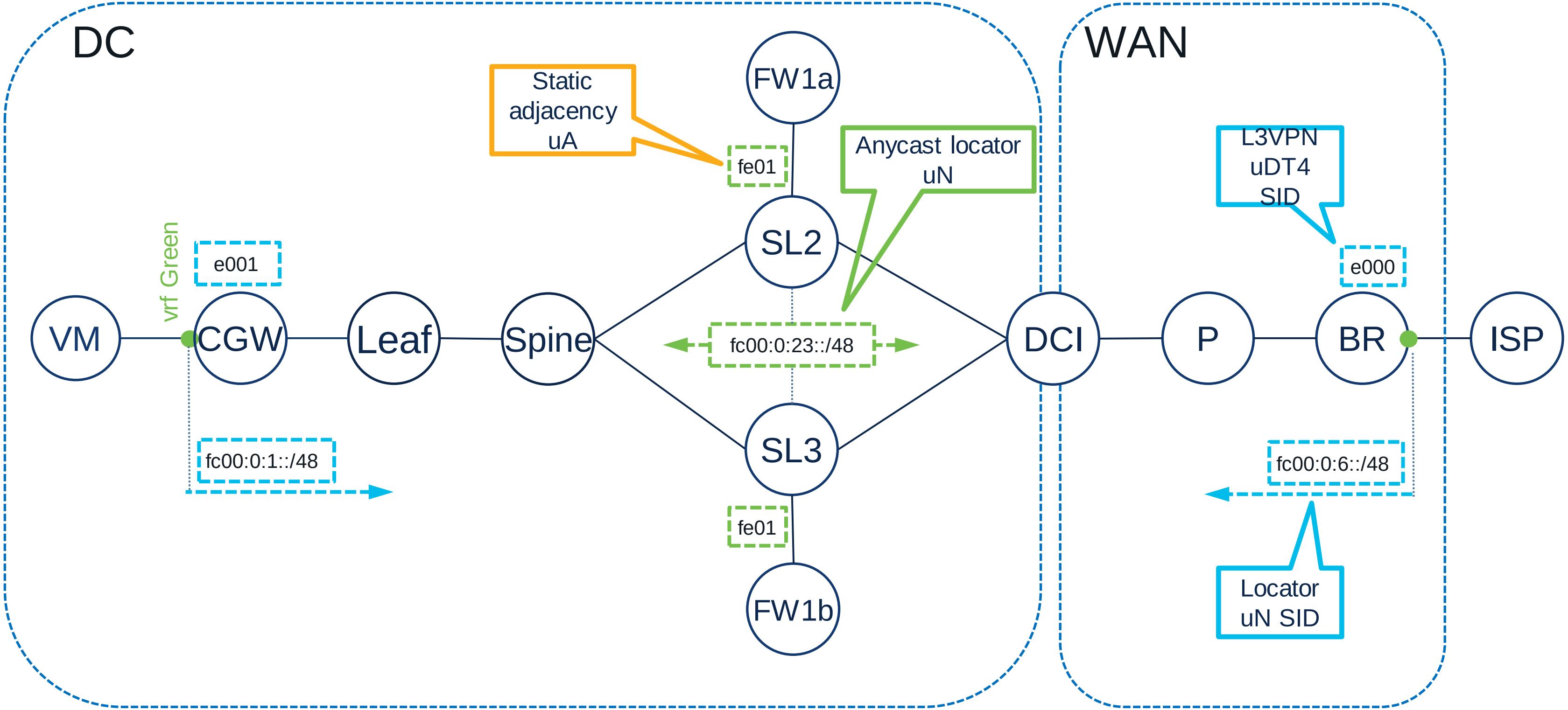
←

IPv4 | SA = 10.61.1.61
DA = 10.11.1.11
Payload

Services Chaining with SRv6 uSID (e.g., Firewall)

- Current design proposed solution:
    - FW is a cluster of sync'd nodes
    - Deployed behind dedicated physical nodes - Service Leaves
    - FW service inspects the inner packet, does not change the outer IP header
        - No encap/decap at SL’s
        - SL’s are SRv6 enabled routers
        - FW is a plain IPv6 forwarder
- Future goal:
    - FW is SRv6 enabled VNF, attached anywhere in the plain IPv6 forwarding network
        - Scaling FW service per any network segment, customer or application

Firewall Insertion

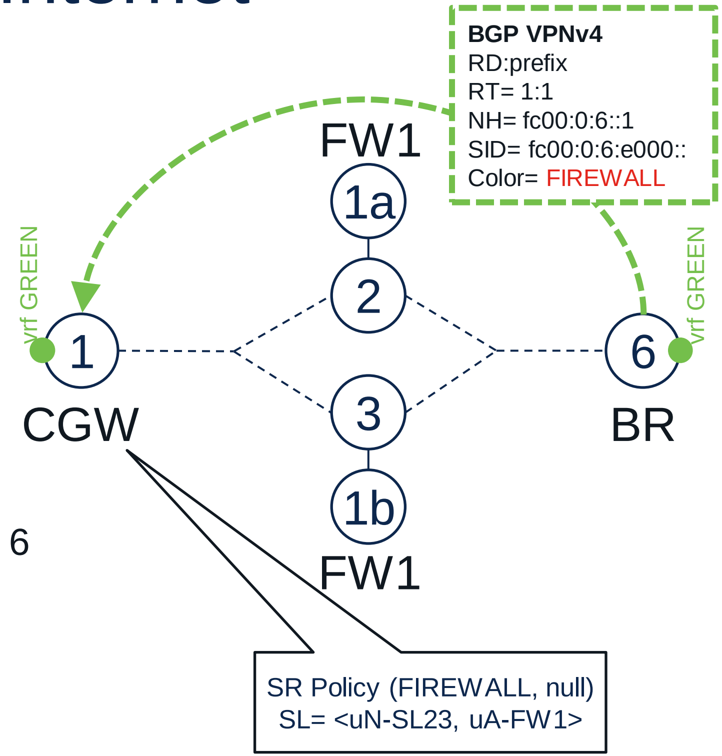


Firewall insertion – From VM to Internet

- BR advertises Internet routes in VRF GREEN with a color “FIREWALL”
  - Individual prefixes, aggregates, default route
- CGW uses BGP AS into a color-only SR Policy
- CGW steers into SR Policy (FIREWALL, null) with SID list <fc00:0:<uN-SL23>:<uA-FW>::>
  - E.g., CGW 1 steers to FW1a/b with SID list <fc00:0:23:fe01::>
  - E.g., CGW 33 may steer it to FW33a/b with SID list <fc00:0:ab:fe33::>
- CGW 1 sends the FIREWALL service packets destined for BR 6 with DA= fc00:0:23:fe01:6:e000::

SL uN + FW uA

uN+uDT4 BR 6



BGP VPNv4
RD:prefix
RT= 1:1
NH= fc00:0:6::1
SID= fc00:0:6:e000::
Color= FIREWALL

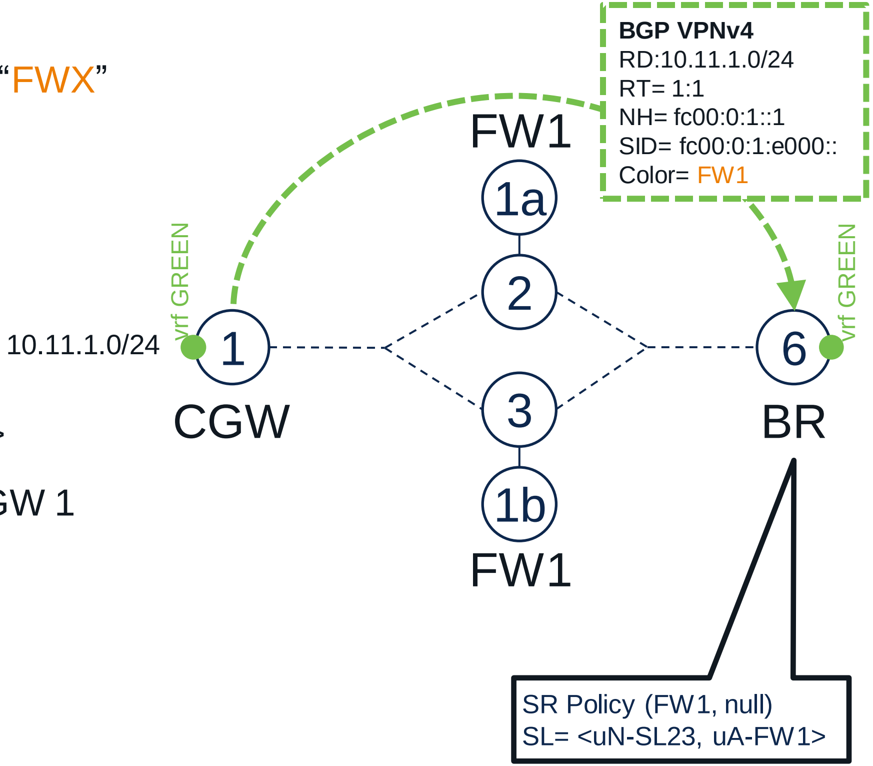
SR Policy (FIREWALL, null)
SL= <uN-SL23, uA-FW1>

Firewall insertion – From Internet to VM

- CGW advertises its VRF GREEN routes with a color “FWX”
  - E.g., CGW 1 advertises 10.11.1.0/24 with color “FW1”
  - E.g., CGW 33 may advertise its prefixes with color “FW33”
- BR steers the service routes into the matching SR Policy (FWX, 0.0.0.0) with SID list <fc00:0:<uN-FWX>:<uA-FWX>::>
  - E.g., BR 6 steers to FW1a/b with SID list <fc00:0:23:fe01::>
- BR 6 sends the FW1 service packets destined for CGW 1 with DA= fc00:0:23:fe01:1:e001::

uN+uA SL23/FW1

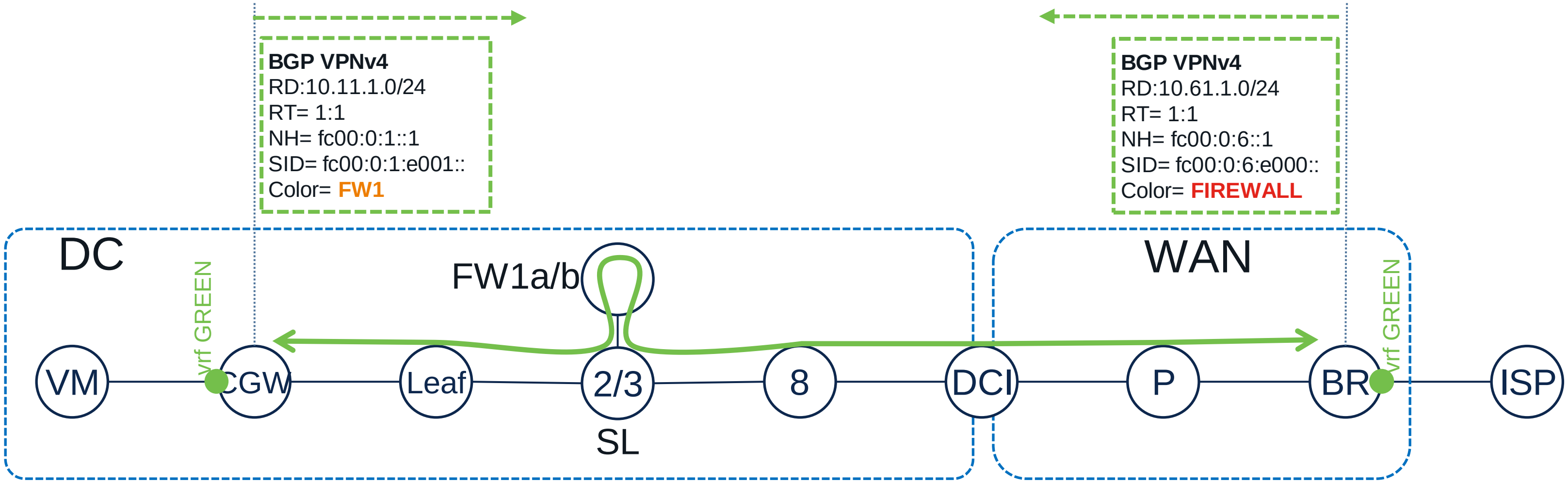
uN+uDT* CGW 1



BGP VPNv4
RD:10.11.1.0/24
RT= 1:1
NH= fc00:0:1::1
SID= fc00:0:1:e000::
Color= FW1

SR Policy (FW1, null)
SL= <uN-SL23, uA-FW1>

Firewall insertion – Packet walk



BGP VPNv4
RD:10.11.1.0/24
RT= 1:1
NH= fc00:0:1::1
SID= fc00:0:1:e001::
Color= **FW1**

BGP VPNv4
RD:10.61.1.0/24
RT= 1:1
NH= fc00:0:6::1
SID= fc00:0:6:e000::
Color= **FIREWALL**

IPv4 | SA = 10.11.1.11 | DA = 10.61.1.61 | Payload

IPv6 | SA = fc00:0:1::1 | DA = fc00:0:23:fe01:6:e000:: | IPv4 | SA = 10.11.1.11 | DA = 10.61.1.61 | Payload

IPv6 | SA = fc00:0:1::1 | DA = fc00:0:6:e000:: | IPv4 | SA = 10.11.1.11 | DA = 10.61.1.61 | Payload

IPv4 | SA = 10.11.1.11 | DA = 10.61.1.61 | Payload

IPv4 | SA = 10.61.1.61 | DA = 10.11.1.11 | Payload

IPv6 | SA = fc00:0:6::1 | DA = fc00:0:1:e001:: | IPv4 | SA = 10.61.1.61 | DA = 10.11.1.11 | Payload

IPv6 | SA = fc00:0:6::1 | DA = fc00:0:23:fe01:1:e001:: | IPv4 | SA = 10.61.1.61 | DA = 10.11.1.11 | Payload

IPv4 | SA = 10.61.1.61 | DA = 10.11.1.11 | Payload

SRv6 Benefits: simplicity and unification

- Unified solution across all domains
- Operational and configuration simplicity
- Gaining scalability

Acknowledgements

- Team Nebius
  - Andrew Tikhonov, Senior Network Engineer, Nebius
  - Samvel Vartapetov, Senior Software Developer, Nebius
- Team Cisco
  - Clarence Filsfils, Fellow, Cisco
  - Kris Michielsen, Technical Leader Engineering, Cisco
  - Pablo Camarillo, Technical Leader Engineering, Cisco