draft-sharma-bess-multi-site-evpn-03

L. Krattiger (Cisco), A. Banerjee(Cisco),), A. Sajassi (Cisco), R, Sharma, R. Shivaramu, K. Ananthamurthy (Cisco

IETF 118, Nov 2023

Prague

Agenda

- Why Multisite
- Multisite Overview
- Comparison with RFC9014
- Border Gateway provisioning
- Multisite Procedures
- Host Mobility
- MVPN with Multi-site EVPN
- Feedback

Why Multisite?

- Sperate EVPN domains to address scaling of number of routes, TEPs by breaking into smaller administrative domains
- Interconnect these EVPN domains with IP backbone without using traditional DCI technologies
- Simple gateway redundancy model to interconnect EVPN domain with IP based ECMP without adding any new requirements to any of the surrounding TEPs
- Flexible VNI allocation across the sites and stitching
- BUM Optimization

Multisite Overview

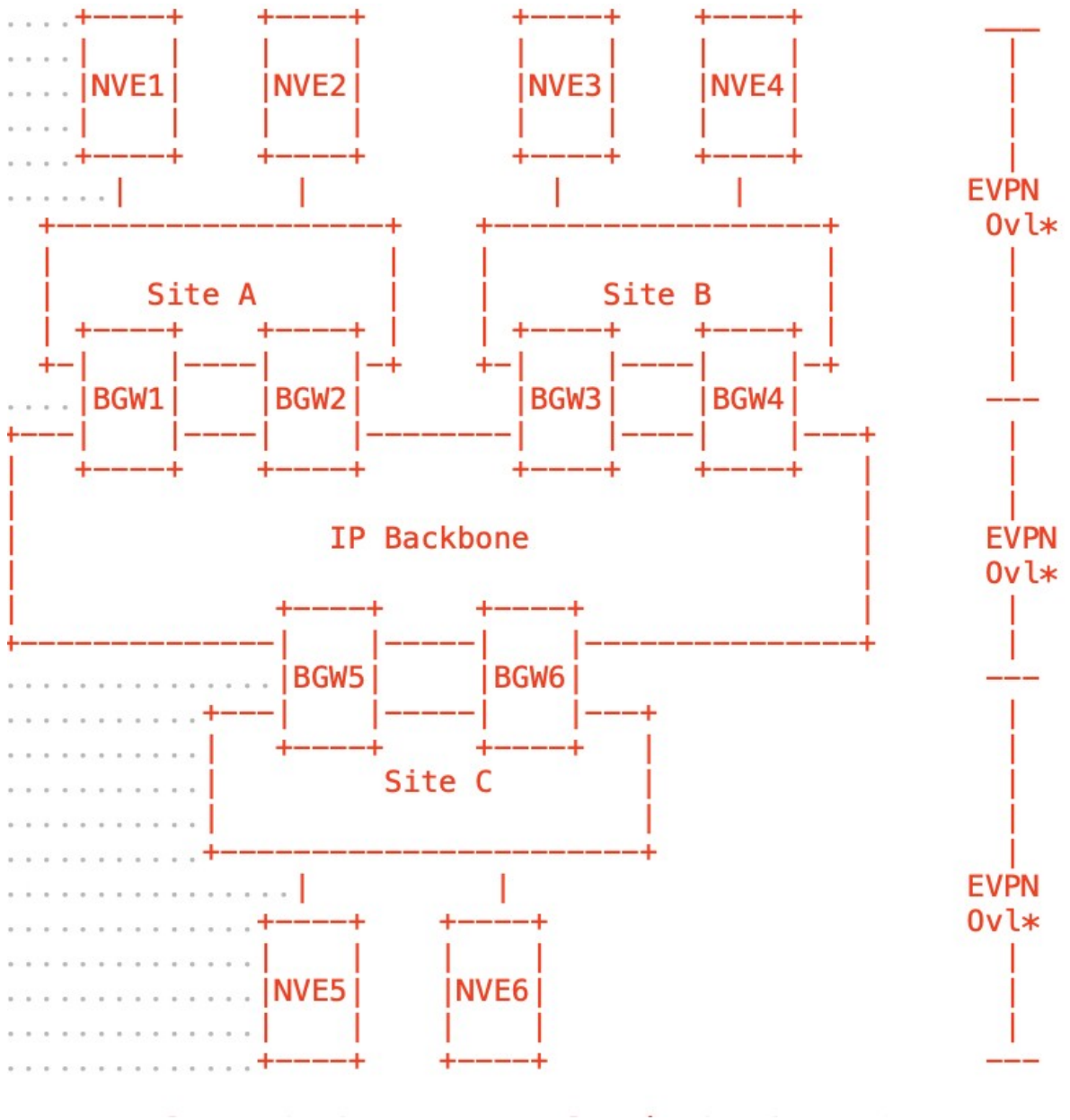
- Site is an EVPN domain consisting of set of EVPN nodes front ended by gateways.
- Border Gateway (BGW) – Gateway Node that separates sites/fabrics/PODs/DCs
  - Acts as entry/exit to a site
  - Anycast BGW provides minimal state sharing, and it can scale out widely by sharing a same anycast/virtual IP address
  - Many TEPs within a DC site/domain are masqueraded behind a single Anycast TEP which represents the gateway between the DC internal and DC external domain
  - The Anycast gateway approach alleviates the hardware of performing multi-path for overlay reachability and respectively reduces control plane paths.
- Site-ID represents a site which is known only to BGWs
- Interconnects two or more EVPN sites/fabrics/DCs in a scalable fashion over an IP-only
- Multisite to interop with a DC domain that interconnects with a RFC9014 - multi-path gateway

Comparison with RFC 9014

	DCI-EVPN-Overlay (RFC 9014)	Multi-Site EVPN (draft-sharma-bess-multi-site-evpn)	
Interconnect	Integrated (1-Box), Decoupled (2-Box)	Integrated (1-Box)	
DCI Encap	VPLS, PBB-VPLS, EVPN-MPLS, PBB-EVPN, VXLAN	VXLAN	
Gateway Mode	Multipath PIP	Anycast VIP	Multipath PIP
ECMP	Underlay and Overlay	Underlay	Underlay and Overlay
EVPN RT-1	Consumed and Generated	None	Consumed and Generated
EVPN RT-2	Re-Originated with I-ESI	Re-Originated with ESI 0	Re-Originated with I-ESI
EVPN RT-3	Consumed and Generated	Consumed and Generated	Consumed and Generated
EVPN RT-4	Consumed and Generated	Consumed and Generated	Consumed and Generated
EVPN RT-5	- (not part of RFC)	Re-Originated	Re-Originated
Route Distinguisher (RD)	Separate RD for Intra and Inter DC	Separate RD for VIP and PIP	
Route-Target (RT)	Separate RT for Intra and Inter DC	Same RT for Intra and Inter DC	
VNI Allocation	Global and Downstream	Global and Downstream	
DF Election	Based on EVPN RT-4	Based on EVPN RT-4	
Identifier	I-ESI	I-ESI (= Site-ID)	
Split Horizon	Local Bias	Local Bias	
ESI-Type	Type 0 (Operator Managed)	Type 3 (MAC Based) or Type 5 (AS based)	
BUM Tree #	2, GW stitched (Intra and Inter DC)	2, GW stitched (Intra and Inter DC)	

Border Gateway Provisioning

- Anycast BGW
  - No RT-1
  - No ESI in RT-2
  - No overlay Multipath
  - Nexthop will be rewritten to Anycast IP
- Multi-path Border Gateway
  - Nexthop will be rewritten to Unique IP (PIP)
  - The Multi-path Border Gateway follows the model of the interconnect ESI (I-ESI) as described in [RFC9014].
  - RT-2 with I-ESI
  - RT-1 used for route resolution
  - RT-5 requires no additional resolution as per [EVPN- IPVPN].



Multisite Procedures

- MP-iBGP EVPN for internal site peers
- MP-eBGP EVPN for between Multiple Sites
- BGW is aware of site internal and external EVPN peers
- DF Election
  - DF election based on Site ESI label (I-ESI) using site identifier
- Remote BGW discovery
  - RT2-/RT-5
- RT-2/RT-5 route will be terminated and re-originated
  - Nexthop will be rewritten
  - Carry downstream assigned VNI labels
  - Flexibility to provision different MAC-VRF or IP-VRF to VNI mapping in different sites
- RT-1, RT-3 and RT-4 from other sites will be terminated at BGWs

Host Mobility

- No new procedures and follows RFC7432 host mobility procedures

Convergence

- Fabric to Border Gateway Failure
  - If BGW is lost, NH will be withdrawn for RT-2/RT-5 routes
  - Per-VNI DF election will be triggered to chose new DF.
  - New DF winner will become forwarder of Multi-destination inter-site traffic.
- Border Gateway to Border Gateway Failures
  - NH reachability to BGW will lost and handled as regular underlay failure

MVPN with Multi-site EVPN

- Inter-Site MI-PMSI
- Stitching of customer multicast trees across sites
- RP placement across sites
- Inter-Site S-PMSI

Next Steps

- Request feedback/comments from WG members