Build Hierarchical Fabrics with VXLAN EVPN Multi-Site

Multi-Site as the evolution of the overlay fabric control plane

Virtual Extensible LAN (VXLAN) Ethernet VPN (EVPN) Multi-Site marks an important milestone in the evolution of fabric overlays. The introduction of BGP EVPN Control Plane for VXLAN enabled a more scalable overlay solution than traditional flood-and-learn VXLAN. With integrated layer-2 and layer-3 forwarding, multi-tenancy and workload mobility, BGP EVPN Control Plane pairs more efficiency with greater scalability. VXLAN EVPN Multi-Site continues that journey of efficient overlays. EVPN Multi-Site incorporates proven networking design principles of hierarchical networks, fault containment, as well as network control boundaries for building scalable overlays fabrics.

Customers interconnect their VXLAN overlays to make optimal use of available capacity or for disaster avoidance and recovery. They need layer-2 extensions to support workload mobility across multiple data center fabrics. Earlier they had two options to do this. They either built a **multipod** solution with a single flat, end-to-end overlay or used **multifabric** with handoff from a VXLAN fabric to a traditional data center interconnect technology. Now, with VXLAN EVPN Multi-Site, they can interconnect these fabrics with multiple, separated control and data plane domains using an integrated handoff approach.

Solution description

The VXLAN EVPN Multi-Site solution uses border gateways in either anycast or virtual port-channel configuration in the data plane to terminate and interconnect overlay domains.

The border gateways provide the network control boundary that is necessary for traffic enforcement and failure containment functionality.

In the control plane, BGP sessions between the border gateways rewrite the next-hop information of EVPN routes and reoriginate them. VXLAN Tunnel Endpoints (VTEPs) see only their overlay domain internal neighbors including the border gateways. All routes external to the fabric will have a next hop on the border gateways for Layer 2 and Layer 3 traffic.

This is an open solution that extends BGP-EVPN control plane to provide this hierarchical multi-site connectivity.

Challenges

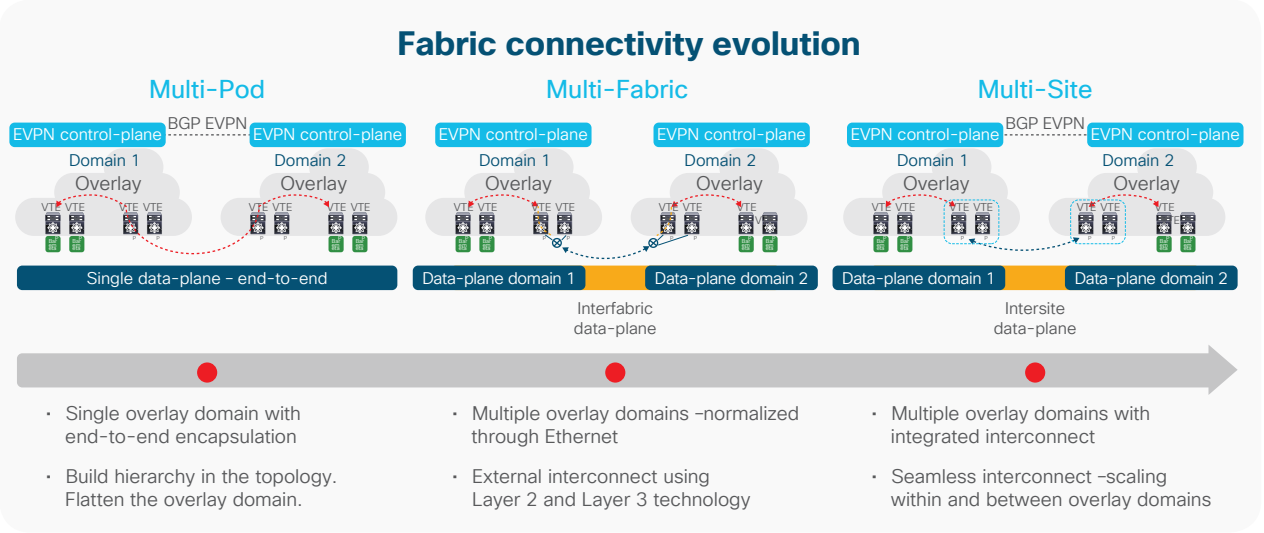
Multi-Pod uses a **single** overlay domain (end-to-end encapsulation). This approach results in challenges with **scale**, **fate sharing** and **operational restrictions**.

Since this is a **single** overlay domain spread across multiple pods, the total **scale** supported in terms of tunnel endpoints (VTEP) is limited to the **scale** of a single overlay domain.

With a single overlay there is a single control-plane domain, single broadcast domain and hence there is fate-sharing with a single failure domain. A failure introduced in one pod percolates across the **multipod** topology.

Operational aspects have to be considered as with the need of end-to-end encapsulation, end-to-end reachability becomes necessary. Network operators do not have the freedom to address the pods independently. They are also restricted to selecting single replication mode per pod - either ingress replication or multicast replication has to be used end-to-end.

Any benefit gained from the hierarchical topology is lost with a **single** flat overlay domain.



Benefits of EVPN Multi-Site

VXLAN EVPN Multi-Site continues the journey of efficient overlay networking. Multi-Site incorporates proven networking design principles of hierarchical networks, fault containment, as well as network control boundaries for building scalable overlays fabrics. EVPN Multi-Site facilitates both Layer 2 and Layer 3 extension beyond a single overlay domain and extends the build of hierarchical topologies into the overlay domains. With this approach, multi-stage fat-tree topologies become feasible for overlay designs without trade-offs. While the benefits apply to the build out of scalable data center fabrics, they equally apply to the interconnect of geographically dispersed data centers. The increase in scalability, the reduction of fate sharing with simplified operation and transport independence are key to modern data center designs.

Multi-Fabric brings back not only the hierarchical topology but also the concept of multiple overlay domains. Even with this advantage, challenges exist with **operational complexity**, **transport dependency** and the need for **additional equipment** and **technology**. In **multifabric**, different technologies are involved to extend the individual overlay domains and interconnect them. This results in greater operational complexity through the introduction of additional state and technology disaggregation. **Additional equipment** is required to achieve the interconnect for the targeted domain separation.

Last but not least there is more architecture rigidity with **transport dependency** on the hand-off requirements between the fabric and the interconnecting devices.

These are the solution's benefits:

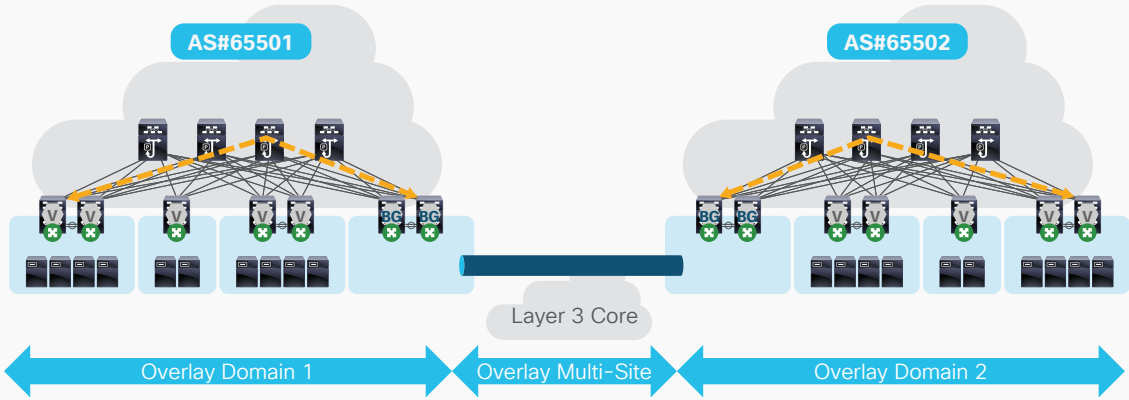
- **Scalable:** Multisite allows building hierarchical overlay domains with seamless interconnection.
- **Failure containment:** Multisite allows granular network control boundaries for Layer 2 Broadcast, Unknown Unicast and Multicast (BUM) traffic.
  - Aggregated Storm control rate limits can be applied from overlay domain to overlay domain
  - Unknown unicast flooding can be controlled towards the overlay domain interconnection
  - Broadcast storms can be limited or mitigated by enabling or rate limiting this type of traffic
  - The network control boundary can choose what type of BUM traffic to forward
- **Flexible and independent** topology configuration. The multiple topologies are isolated and translated on the border gateway (BGW).
- **Resiliency:** Automated multihoming with up to 4-way border gateways.
- **No additional boxes:** Seamless integrated interconnection at the border gateways stitching individual overlay domains by preserving multi tenancy and state aggregation.
- **Loop prevention**
- **Transport independence:** The transport network between overlay domains can use any IP based transport; no requirements to multi-destination traffic replication.
- **No fate sharing:** Flexibility in configuring per-overlay domain multi-destination replication. Ingress replication or multicast can be chosen per overlay domain.
- **Just-in-time scale as you grow:** Build the right-size of overlay domains and extend as demand grows. Add additional capacity by building new overlay domains and interconnecting them seamlessly using EVPN Multisite. This incremental approach allows simplified capacity planning over time and avoids the need for single large overlay domains.
- **Network visibility** across multiple sites, from one host in one site to another host in another site leveraging VXLAN operations, administration, and management (OAM).
- **Brownfield integration:** By inserting Border Gateways in front of brownfield networks, it is possible to integrate them as part of a Multi-Site solution with other vxlan greenfield deployments.
- **Open:** This BGP EVPN based solution is open and has been submitted to IETF for standardization.

Platforms supported

The border gateway support required for the Multi-Site solution is available on Cisco Nexus® 9300-EX and 9300-FX switches.

For more information

Cisco Nexus 9000 Series Switches (Data Sheets and Literature)



C45-739422-00 09/17