
# Link State Vector Routing

ENOG-16 2019

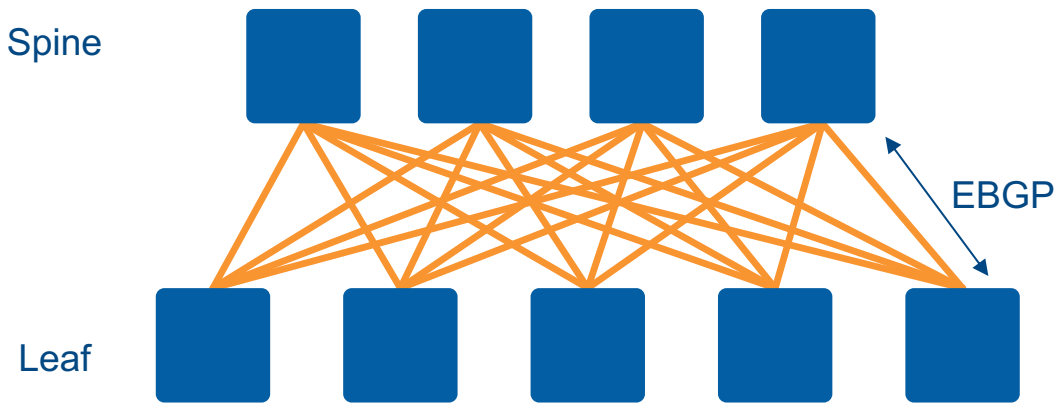
June 3rd 2019

Keyur Patel, CTO & Founder, Arrcus Inc.

# Agenda

- Background and Motivation
- LSVR Solution
- LSVR Benefits & Takeaways

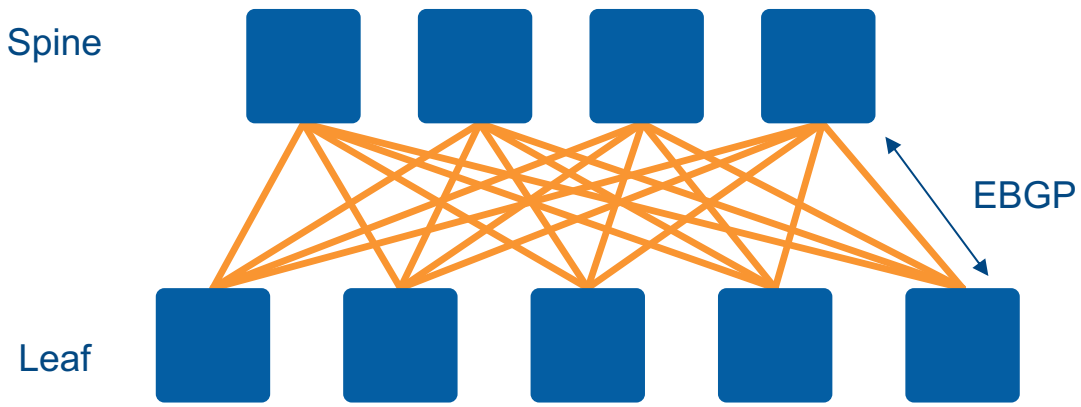
# Modern Data Centers




**IP CLOS Fabric**

- IP CLOS - Leaf-Spine fabric
- Layer3 - BGP Routing Protocol
  ("Use BGP for Routing in Large-Scale Data Centers" in RTGWG)
- Goals - Scale, Simplicity, Resiliency
- What's next
  - Growing ECMP scale
  - Increasing CLOS tiers
  - New applications

# BGP CLOS Architecture - Attributes




**IP CLOS Fabric**

- Multiple Tiers
- BGP Hop by Hop Peering
- BGP configuration has direct co-relation with ECMPs – configuration explosion
- Propagation of routing table to each hop
- BFD for link level liveness

# CLOS IP fabric - Routing protocol comparison

## IGP

+ Complete fabric topology at each node => Path computations – SPF, TE, CSPF, LFA;  etc
+ Faster convergence

- Flooding of link-state
- Complexity - State machine, LSDB

## BGP

+ Simplicity – operational, troubleshooting (BRIB vs LSDB)
+ Incremental updates, no flooding and selective filtering
+ Reliable transport
+ Per-hop TE (UCMP)

- Convergence (per prefix, best path computation at every hop prior to advertisement)
- Configuration

# LSVR – combining the best attributes of BGP and IGP

+ Complete fabric topology at each node => Path computations – SPF, TE, CSPF, LFA;  etc
+ Faster convergence

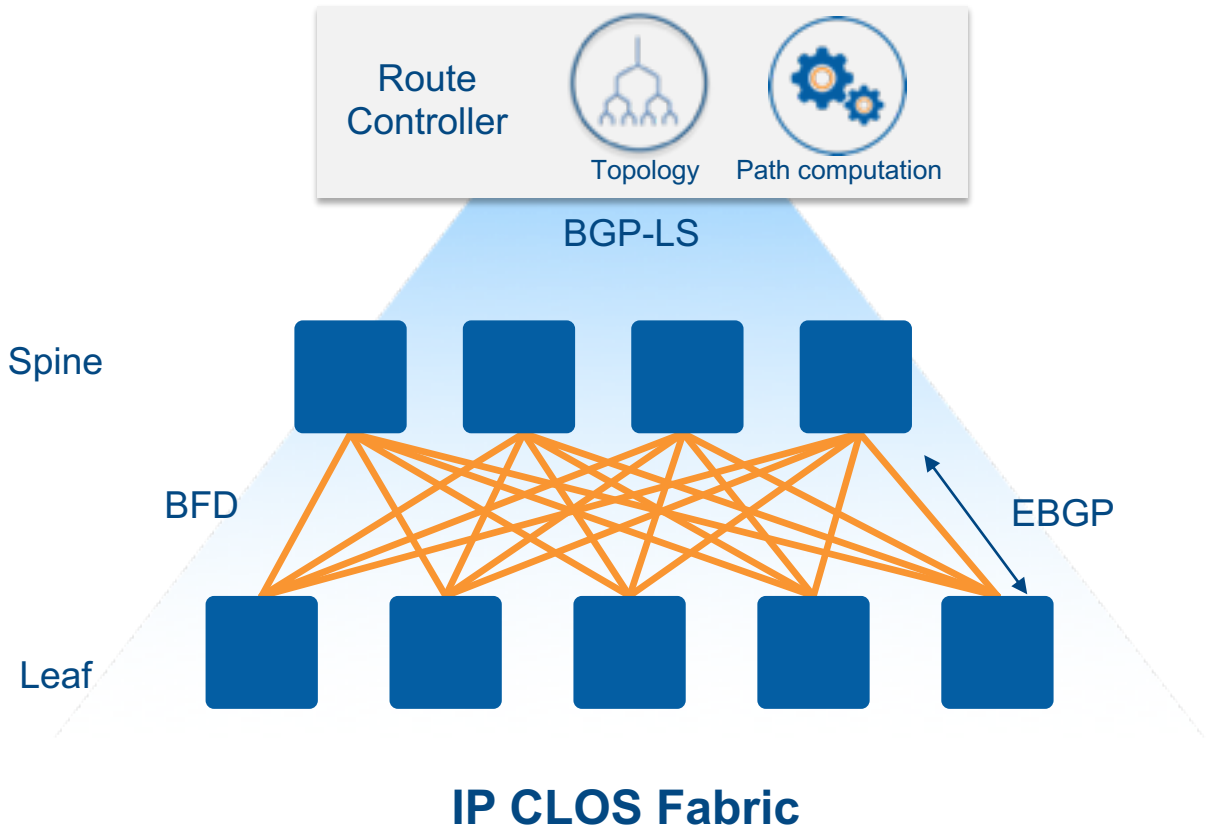
+

+ Simplicity – operational, troubleshooting (BRIB vs LSDB)
+ Incremental updates, no flooding and selective filtering
+ Reliable transport

Use BGP as base protocol
Add the best of IGP characteristics

***Link State Vector Routing (LSVR)***

# LSVR Solution - Highlights

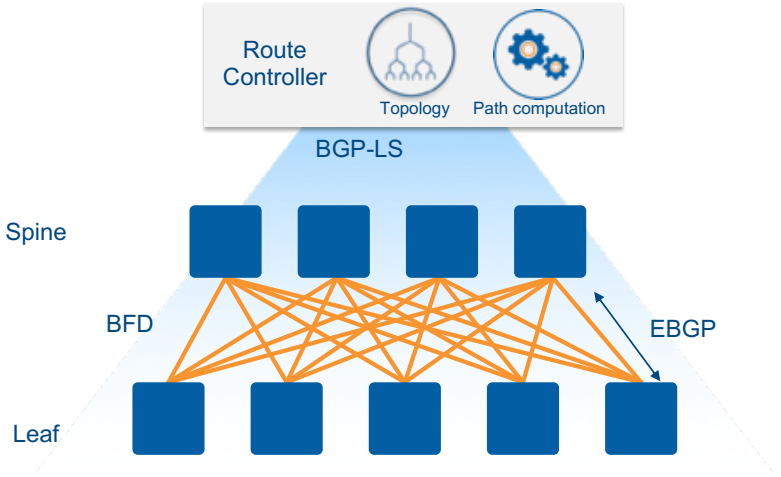



**IP CLOS Fabric**

- **IP CLOS Architecture**
  - Multiple Tiers
  - BGP peering with Route Controllers (/RRs)
    - Simplified protocol configuration
  - BGP configuration has direct co-relation with number of controllers *not* ECMP
    - Control Plane Flooding optimized
  - Route Controllers (/RRs) merely reflecting route updates
    - No head of line blocking for update announcements
  - Switches performing SPF algorithm to create graphs
    - Distributed computing
  - Overlay AFI/SAFIs (EVPN) can follow the same model with a traditional BGP
  - BFD for Link layer liveness

Enabling new DC applications – flexibility, control, scale

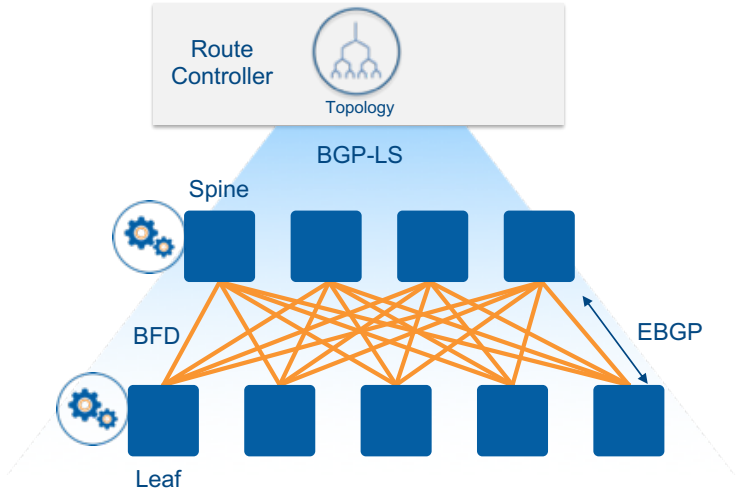
# LSVR Solution – Flexible operational models




**IP CLOS Fabric**

- Centralized out-of-path Route Controller (/RR) – topology and path computation
- Simplified management, single point for policy enforcement, enhanced path computation (CSPF) for traffic engineering, ORR
- Can be coupled with BGP-SR in underlay for SR-TE

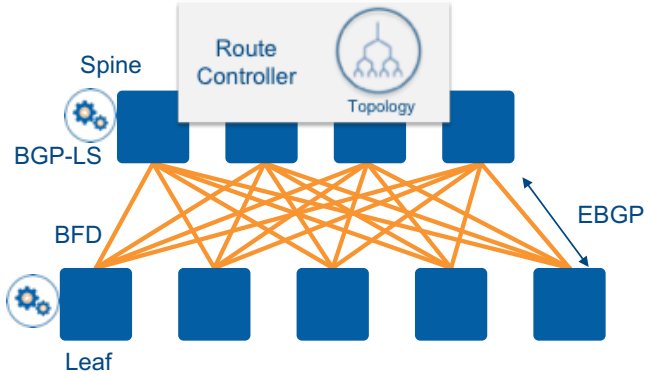
**Centralized path computation**




**IP CLOS Fabric**

- Centralized out-of-path Route Controller (/RR) – topology
- Simplified management, single point for policy enforcement
- Network nodes (Leaf-Spine) – distributed path computation
- All nodes have full topology
- Loop-free Alternates (LFA) computation, Fast Convergence

**Distributed path computation**




**IP CLOS Fabric**

- Inline Route Controller (/RR) on Spine – topology
- Network nodes (Leaf-Spine) – distributed path computation
- Leverage compute cores on Spine

**Inline Route controller**

# LSVR Solution – Protocol Enhancements

- Define a new SAFI
  - NLRI format is exactly same as BGP LS Address Family to carry link state information
- BGP MP Capability and BGP LS Node attribute to assure compatibility
- Multiple peering models supported
- BGP runs Dijkstra instead of Bestpath decision process

# LSVR Solution - Best-Path Changes

- Next-Hop and Path attributes announced as part of RFC4271
- BGP Decision Process Phases 1 and 2 replaced by SPF algorithm
- Decision Process Phase 3 may be short-circuited since NLRI is unique per BGP speaker
- Need to assure the most recent version of NLRI is always used and re-advertised
  - Augmented with support of sequence numbers

# LSVR Solution - SPF

- Starting with greatly simplified SPF with P2P only links in single area (i.e., SPT)
- Will scale very well to many use cases
- Could support computation of LFAs, Segment Routing SIDs, and other IGP features
  - BGP-LS format includes necessary Link-State
- Link-State AF is dual stack AF since both IPv4 and IPv6 addresses/prefixes advertised
  - BGP-LS format also supports VPNs but SPF behavior not defined
  - Work needed to define interaction with existing unicast Afs
    - Matter of local implementation policy

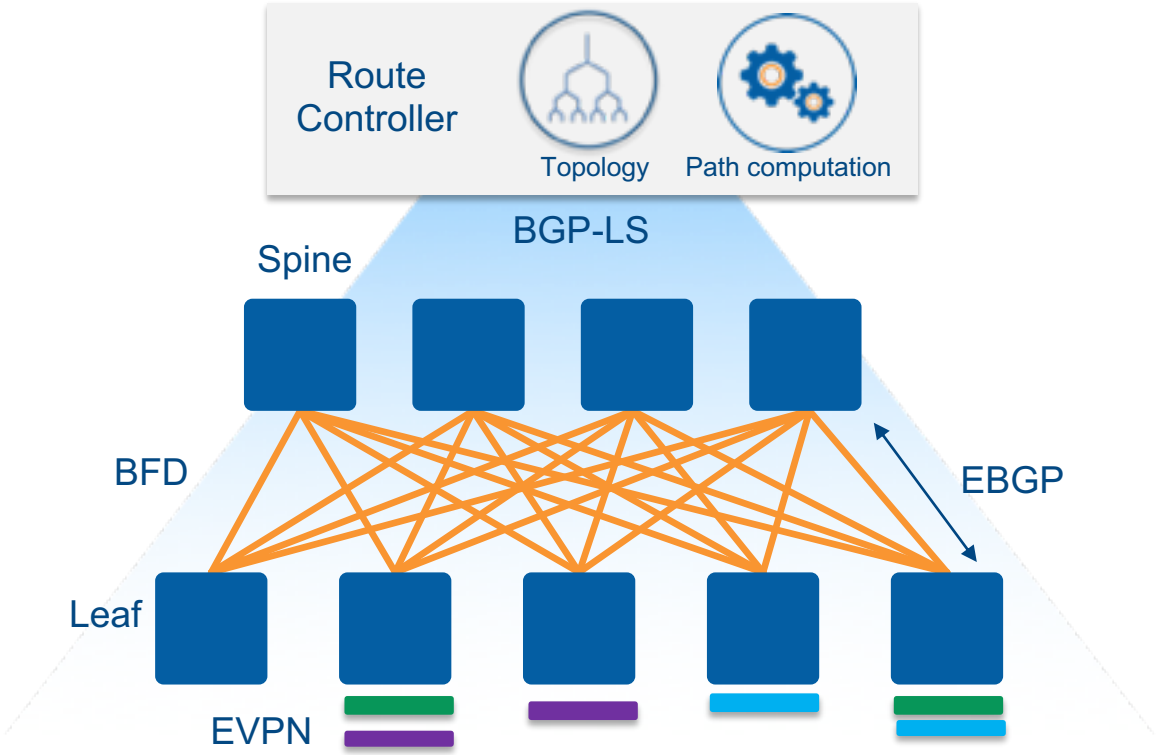
# LSVR - Route Origination

- Routes originated by redistribution
- Discovered using a new Discovery protocol - Layer 3 Liveness and Discovery protocol (L3DL)
  - Used to discover unique identities of port/link including IP, MAC, Label binding
  - Discover each other's unique endpoint identification
  - Discover configuration information
  - Has layer2 keepalive for session continuity
  - Standards being developed in LSVR WG at IETF
  - Open source software efforts going on

# LSVR Solution - Convergence Improvements

- BGP Link NLRI attribute BGP SPF Status TLV added
  - Signals link down event without immediately withdrawing Link NLRI
- BGP Prefix NLRI attribute BGP SPF Status TLV added
  - Signals Prefix down event
- NLRI Implicit withdrawal delay
  - Avoids problem of flooding path discrepancies causing unnecessary route flaps
  - Specifically, avoids flap when on fastest flooding path fails and NLRI is implicitly withdrawn before being received on others
- Link Liveness is done using BFD

# LSVR Benefits & Takeaways




**IP CLOS and Multi-tenant BGP EVPN overlay**

- Builds on top of existing BGP protocol
- Enables faster scale and convergence in growing ECMP IP CLOS fabrics
- Enables centralized Route Controller architectures
- Enables Traffic Engineering of underlay paths
- When combined with overlay multi-tenant fabric solution like EVPN, LSVR enables Automation, Programmability and Resiliency
- Ability to discover enhance link capabilities when combined with L3DL – facilitates route origination as well

ARRCUS
NETWORK DIFFERENT™

# Thank You!