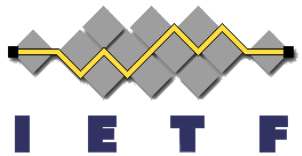
LSRV BGP SPF Applicability IETF 101, London

Keyur Patel, Arrcus
Acee Lindem, Cisco
Shawn Zandi, LinkedIn
Gaurav Dawra, LinkedIn



Agenda

- Data Center Applicability
- Full Peering Model
- Sparse Peering Model
- Security
- Operational Simplicity

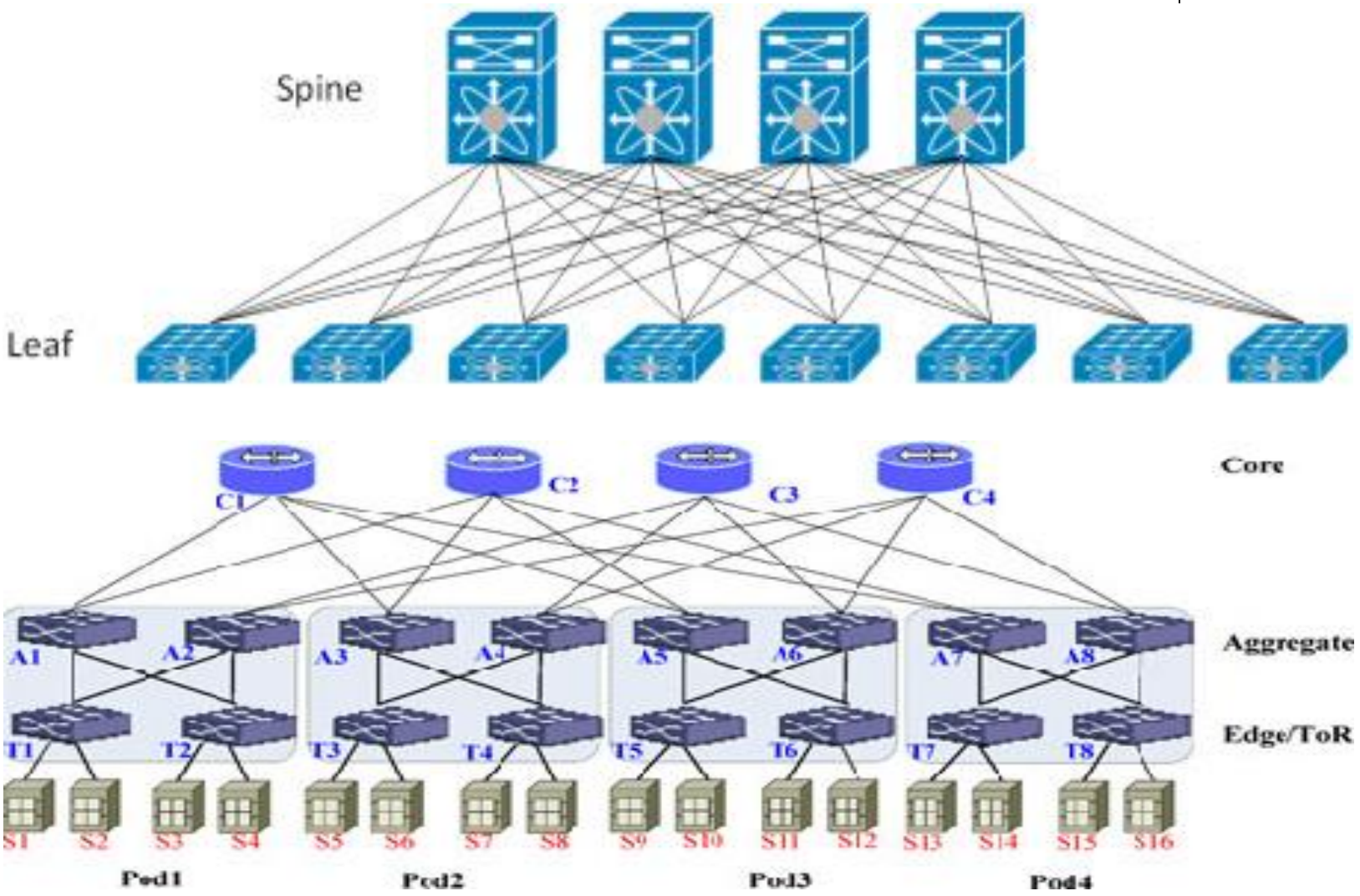
Data Center Applicability 1/2

- Massively Scalable Data Centers (MSDCs) have implemented simplified layer 3 routing
- Centralized route control using some controller-based solution for simplified management
- Operational simplicity has lead MSDCs to converge on BGP as their routing protocol
  - RFC 7938 - Use of BGP for Routing in Large-Scale Data Centers

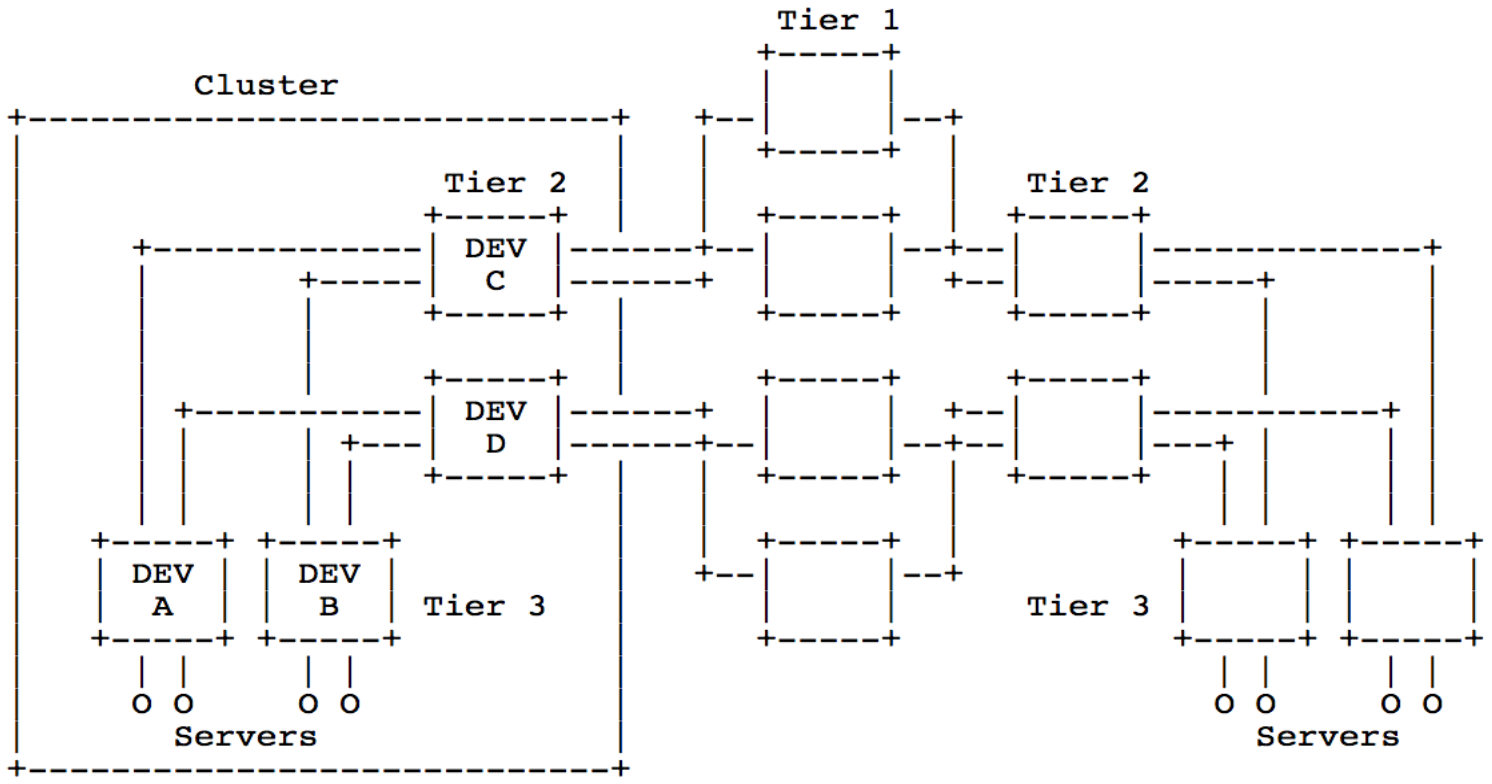
Data Center Applicability 2/2

- Route Controller has a similar functionality as a Route Reflector
  - May Reflect Routes
  - Central Database for policy enforcements, management, etc.
- However Route Reflector assumes a presence of IGP that help resolve next hop sand its adjacencies for its clients
- BGP based MSDCs solve this problem by establishing hop-by-hop peering sessions
- Proposed solution helps towards deployment of Route Controllers and yet preserve operational simplicity by using BGP

Data Center CLOS and FAT Tree Applicability



Full BGP Peering, ala RFC 7938



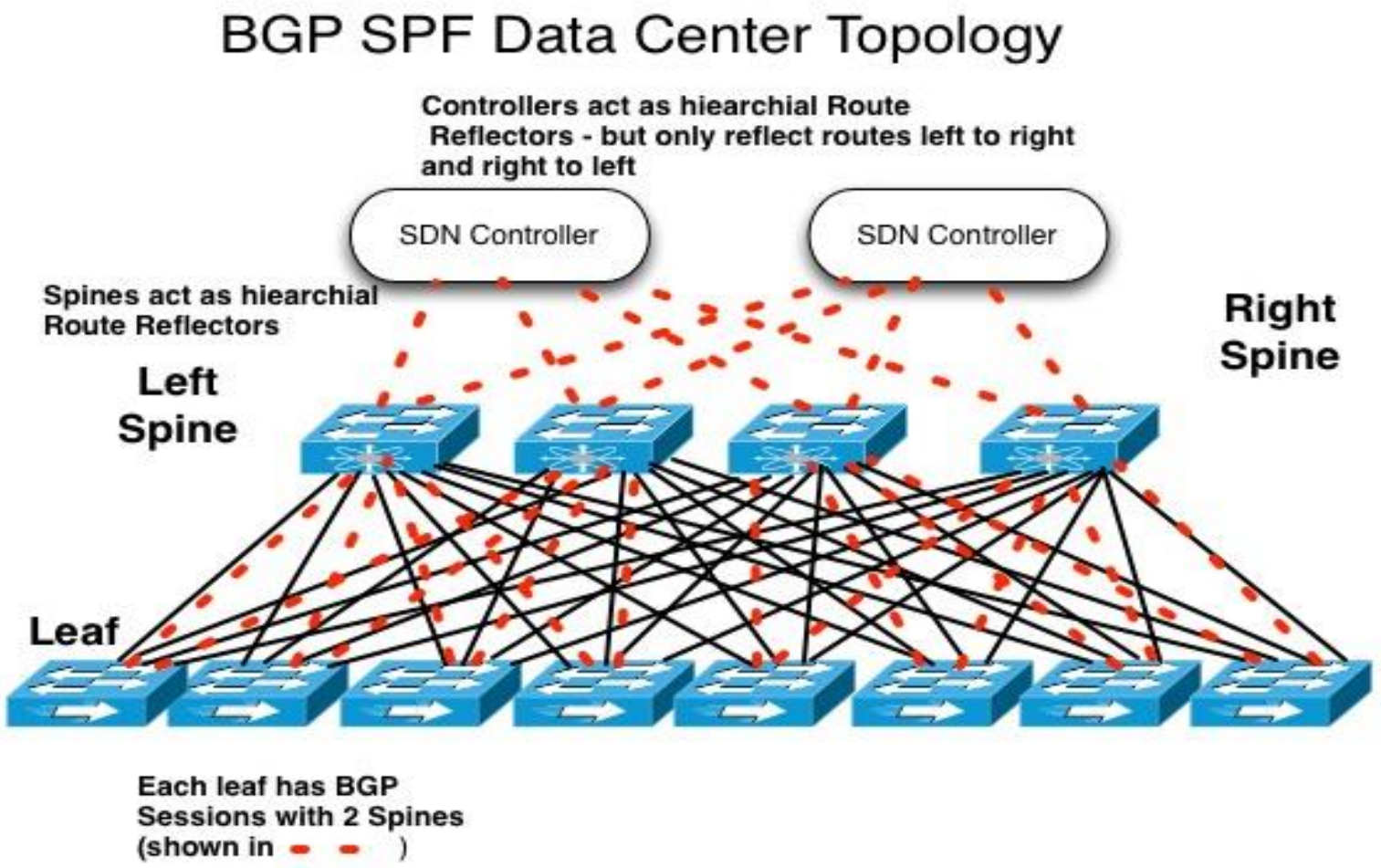
Full BGP Peering

- Peering on every link in fabric – Same topology as RFC 7938
- BGP SPF used for IPv4 and IPv6 Unicast AFs in underlay – some MSDCs are flat with no overlay.
- BFD recommended for faster link/node down detection rather than aggressive BGP keep-alive and hold timers.
- Drawback one will have a separate BGP RIB (BRIB) copy of the complete topology for every northbound peer.
  - BGP Session results in discard of NLRI from peer.
  - Especially CLOS Topologies can support extremely dense ECMP
  - Not really all the bad, since only best-path is propagated.
- Do not use BGP Add-Path (RFC 7911) since BGP speakers have the full topology!

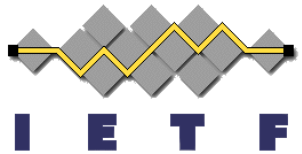
Sparse BGP Peering

- Liveness detection for links done outside of BGP (i.e., based solely on link status or using BFD)
- Leaves peer with subset of spines (e.g., only 2 to offer redundancy)
  - Spines act as Route Reflector
  - Savings in sessions depends on the number of spines to which leaves are connected
  - Redundancy trade-off versus copies of advertisements
- Spines peer with controllers
  - Controllers reflect between spines that peer with a unique set of leaves

BGP SPF Data Center Sparse Peering Example

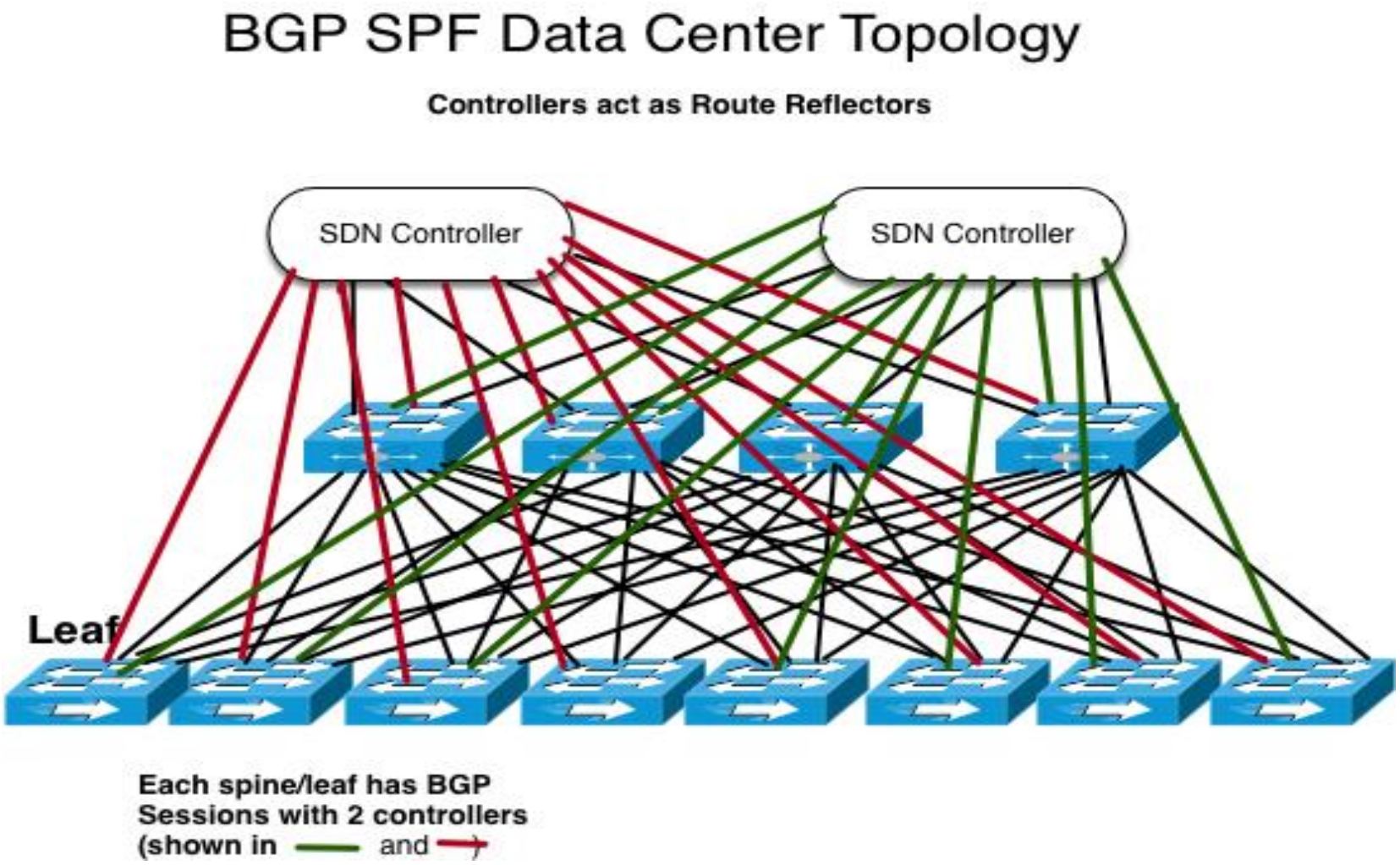


Sparse BGP Alternate Peering Option



- Use local discovery mechanism to install northbound routes to controller.
  - BFD used to determine status of northbound routes.
- Removes hierarchal route reflection with all BGP speakers in the fabric peering with controllers.
  - Only Best-Path NLRI need be advertised
  - Trade-off with convergence and NLRI updates
- Not fully baked yet

BGP SPF Data Center Alternate Sparse Peering



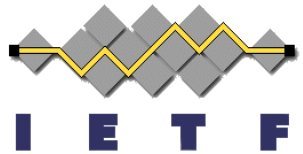
SDN Controller Role

- Selective hierarchal route reflection between groups of spine nodes
- Provision Overlay Services
  - EVPN for L2 and L3 VPNs
- Use BGP-LS Based topology to provision traffic engineered routes
  - BGP SR-TE could be used for this provisioning

BGP SPF Security

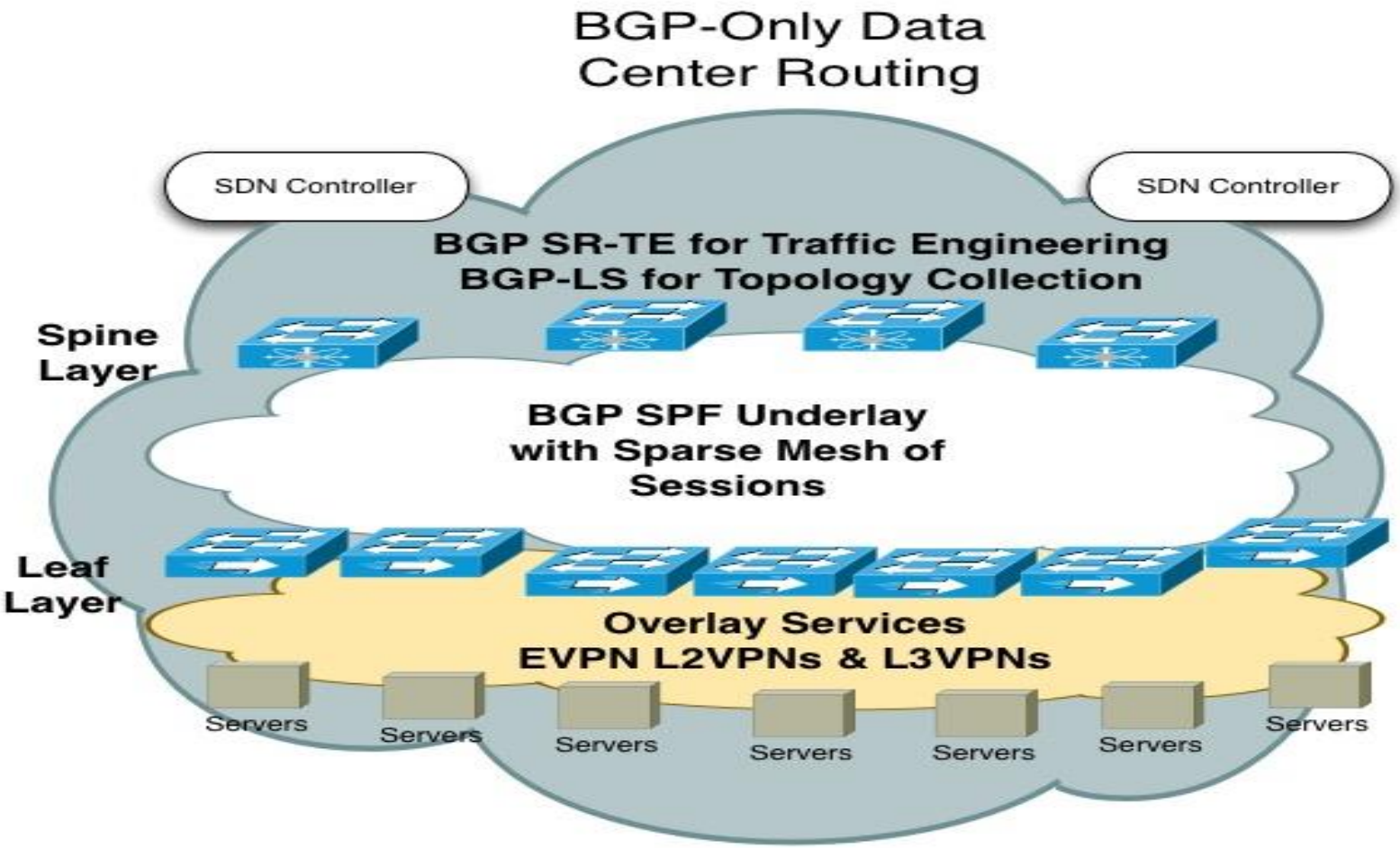
- Really no different than classic BGP underlay security
  - Simple for both full and sparse peering
  - Tolerance required for alternate sparse peering model
- Use of TTL security on intra-fabric BGP sessions (RFC 5082)
- If BGP fabric is not isolated, recommend control plane protection as well (RFC 6192)
- If BGP fabric may be subverted, TCP-AO (RFC 5925) is recommended (MD5 - RFC 2385 if unavailable)
  - Keys should support key-chain rollover via the YANG model as described in RFC 8177 and be changed periodic or when there is potential for a breach.

Operational Simplicity with Single DC Protocol



- BGP SPF for underlay in data center fabric
  - BGP-LS encodings used for link-state advertisement
  - Segment Routing SIDs can be advertised using existing SR encodings
- BGP EVPN for L2VPN and L3VPN Services
  - EVPN for Virtual VLANs (classic RFC 7432)
  - EVPN Type 5 Route for L3VPNs (draft)
  - EVPN Extended Community for VPWS (draft)
- BGP SR-TE for Traffic Engineering
  - BGP-LS NLRI can be leveraged for traffic engineering as well

BGP-Only Data Center Routing



Next Steps

- New revision of the document with more meat (or more tofu for you vegetarians)
- WG Review and Discussion