VPLS P2MP and NGEN MVPN in JunOS

Rahul Aggarwal
Distinguished Engineer
rahul@juniper.net

Juniper™ NETWORKS

Evolution of MPLS Multicast in JunOS

- The beginning with P2MP MPLS TE
- Enhancements to P2MP MPLS TE
- VPLS P2MP
- NGEN MVPN

The beginning – JunOS 6.4

- P2MP MPLS TE
- Driven by video over MPLS
- Deployed in multiple networks

P2MP MPLS TE after JunOS 6.4

- Several enhancements already made
  - P2MP CSPF
  - P2MP Link Protection
  - P2MP Transit Graceful Restart
  - P2MP Integration with IP multicast on egress LSRs
- Other enhancements on JunOS roadmap
  - P2MP TE ingress PE redundancy
  - P2MP TE node protection

What about other applications ?

- VPLS multicast
- NGEN MVPNs
- Where is JunOS ?
  - This presentation focuses on this

VPLS Multicast and P2MP LSPs

- Intra-AS P2MP LSPs for VPLS multicast
  - JunOS 8.3

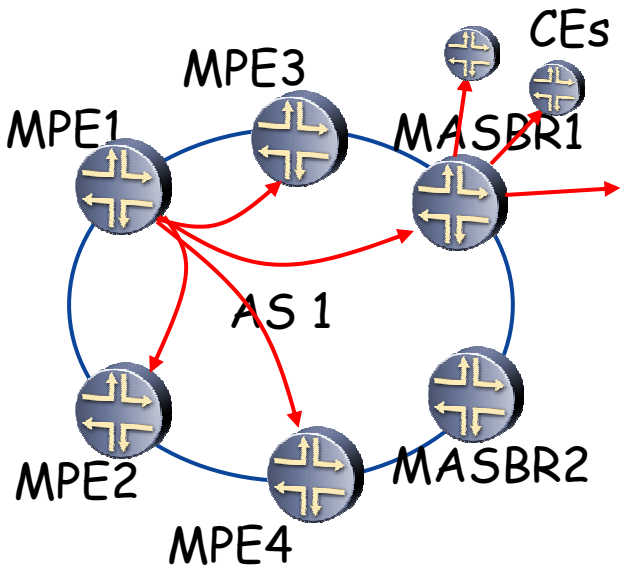
P2MP LSPs for Intra-AS VPLS

- Consider a VPLS with all the sites in a single AS
- BGP-VPLS in the AS
  - With VPLS support for P2MP LSPs as described in draft-ietf-l2vpn-vpls-mcast-01.txt
- Each PE with a site in a VPLS is the root of a P2MP LSP other PEs in the AS that have sites in that VPLS
  - One P2MP LSP per VPLS
  - We will discuss "aggregation" later

P2MP LSPs for Intra-AS VPLS…

- All the multicast/broadcast/unknown unicast traffic received from the VPLS site is sent over the P2MP LSP
- BGP P-Tunnel Attribute, carried in the BGP-VPLS updates, signals the VPLS-P2MP LSP binding
  - From the root of the P2MP LSP to the leaves of the P2MP LSP

P2MP RSVP-TE LSPs in AS1 Control Plane



- Let us look at the control plane mechanisms that allow MPE1 to setup a RSVP-TE P2MP LSP for VPLS multicast/broadcast/unknown unicast
  - The red arrows show the logical flow of BGP and hop-by-hop RSVP-TE control information from MPE1 to MPE2, 3, 4 and MASBR1
- Consider MPE1, 2, 3, 4, and MASBR1 with sites in VPLS1

P2MP RSVP-TE LSPs in AS1 Control Plane

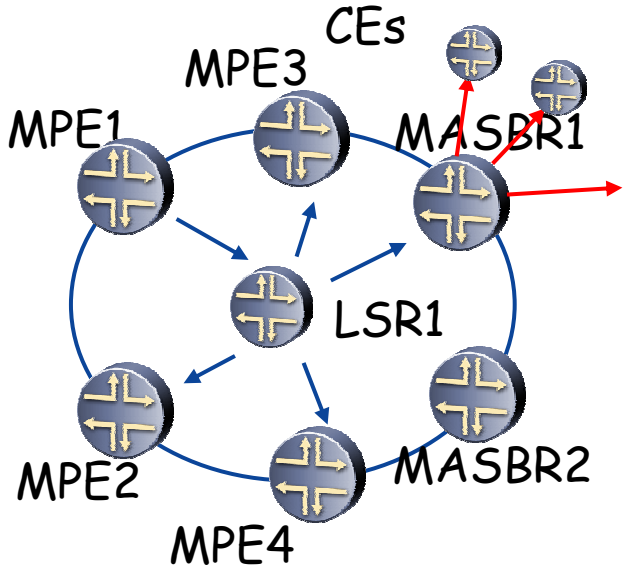
- MPE1, 2, 3,4 and MASBR1, MASBR2 advertise the BGP-VPLS or BGP Auto-Discovery Route for VPLS1
  - MPE1 learns that MPE2, 3, 4 and MASBR1 have sites in VPLS1
- MPE1 signals a RSVP-TE P2MP LSP (P2MP LSP1) with MPE2, 3, 4 and MASBR1 as leaves
  - Draft-ietf-mpls-rsvp-te-p2mp-07.txt
- MPE1 advertises the RSVP-TE P2MP LSP identifier in the updated BGP-VPLS or BGP Auto-Discovery route for VPLS1
- MPE1 programs its VPLS MAC destination table to forward VPLS multicast/broadcast/unknown unicast traffic on P2MP LSP1

P2MP RSVP-TE LSPs in AS1 Control Plane…

- MPE2, 3, 4 and MASBR1 receive VPLS1 to P2MP LSP1 binding from MPE1
- MPE2, 3, 4 and MASBR1 program their forwarding path to forward traffic received over P2MP LSP1 using the VSI corresponding to VPLS1
- Further MPE2, 3, 4 learn the source MAC addresses for packets received over P2MP LSP1
  - The PW to MPE1 is used as the next-hop to forward traffic for the learned source MAC address

P2MP RSVP-TE LSPs in AS1 Data Plane

- P2MP LSP1 outgoing label at MPE1 is say 4556 (received from LSR1)
  - Black arrows show RSVP-TE P2MP physical topology
- P2MP LSP1 label advertised by MASBR1 is say 6778
  - No penultimate hop popping



P2MP RSVP-TE LSPs in AS1
Data Plane

- MPE1 receives a multicast packet from a site in VPLS1
- MPE1 forwards this packet over P2MP LSP1 by encapsulating it in label 4556
- LSR1 replicates the incoming packet (multicast replication) with label 4556 and sends it to MPE2, 3, 4, and MASBR1.
- MASBR1 receives the packet with label 6778
- MASBR1 forwards the packet using VPLS1 MAC table

P2MP LSPs in AS1 Scalability

- Consider **N** MPEs in AS1 with **M** VPLSs on each on average
- So far we have discussed a model where each MPE in a VPLS is the root of a P2MP LSP
  - Let us call this intra-AS VPLS P2MP Phase 1 (JunOS 8.3)
- The number of P2MP LSPs signaled by a MPE equals **M**
- The number of total P2MP LSPs in AS1 equals **M * N**
  - Note that all these P2MP LSPs may not transit a single LSR

P2MP LSPs in AS1 Scalability…

- Aggregation is the ability to use a single P2MP LSP for **multiple** VPLSs
  - Let us call this intra-AS P2MP VPLS phase 2
  - On JunOS roadmap
- Aggregation uses upstream MPLS label assignment
  - Draft-ietf-mpls-upstream-label
- Aggregation reduces the number of P2MP LSPs signaled by a MPE for the VPLSs connected to the MPE
  - If a MPE can use a single P2MP LSP for all VPLSs connected to it, aggregation reduces the number of P2MP LSPs rooted at the MPE from **M to 1**
- This slide set is not currently describing all the control plane and data plane details of aggregation

P2MP LSPs in AS1 for intra-AS VPLS Scalability (Example)

- Consider 100 VPLSs in the AS with 10 sites i.e. 10 MPEs, in the AS, on an average in each VPLS and the same set of MPEs with all 100 VPLSs
- VPLS P2MP phase 1, number of P2MP LSPs rooted at each MPE = 100 (one per VPLS)
- VPLS P2MP phase 1, number of P2MP LSPs in the AS = 100 (per MPE) * 10 (number of MPEs) = 1000
  - All 1000 P2MP LSPs may not transit the same LSR in the AS
- VPLS P2MP phase 2 i.e. aggregation, number of P2MP LSPs rooted at each MPE = 1 (one P2MP LSP for all VPLSs)
  - Note that this number may be greater than 1 depending on the number of VPLSs aggregated on one P2MP LSP
- VPLS P2MP phase 2 i.e. aggregation, number of P2MP LSPs in the AS = 1 (per MPE) * 10 (number of MPEs) = 10
- Bottomline – The potential to aggregate leaves sufficient head room for growth

P2MP RSVP-TE LSPs in AS1 JunOS Implementation Notes

- The leaves of the P2MP RSVP-TE LSP are auto-discovered using VPLS auto-discovery
  - Do not have to be configured
- The P2MP RSVP-TE LSPs are by default setup using no bandwidth
  - One configuration knob per VPLS per PE
  - Use dynamic path computation for P2MP LSPs

P2MP RSVP-TE LSPs in AS1 JunOS Implementation Notes…

- The P2MP RSVP-TE LSPs may be configured with TE parameters
  - Using a LSP template introduced in JunOS 8.3
- The P2MP RSVP-TE LSP may also be configured statically

Migration from Ingress Replication to P2MP LSPs

- Consider migrating a given AS to use P2MP LSPs from ingress replication on each MPE in a VPLS
- The first step is to upgrade the software on each MPE to the new version supporting VPLS P2MP
  - This upgrade can occur one each MPE at a time, impacting only the CEs on the MPE during the upgrade
  - The assumption is that the AS core supports P2MP RSVP-TE day one. If this is not the case, the core needs to be upgraded as well

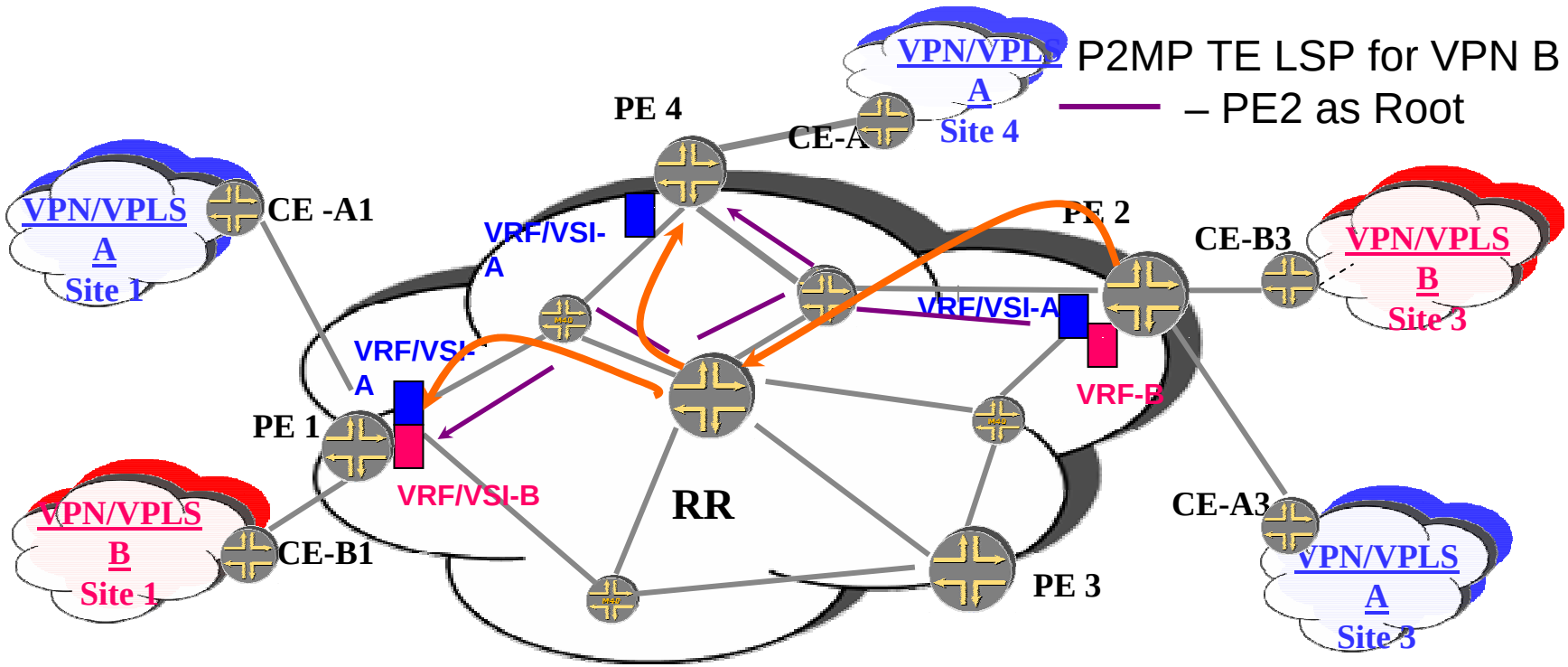
Migration from Ingress Replication to P2MP LSPs…

- The next step is to enable P2MP LSPs **one VPLS at a time**
- Enabling the P2MP VPLS configuration per MPE per VPLS
  - This enables the ingress MPE to perform the control and data plane functionality for VPLS-P2MP
  - The egress MPEs can receive data on both P2MP LSPs and PWs
  - The multicast/broadcast/unknown unicast traffic is switched over by the ingress MPE to the P2MP LSP when the P2MP LSP to all leaves is up

NGEN MVPNs

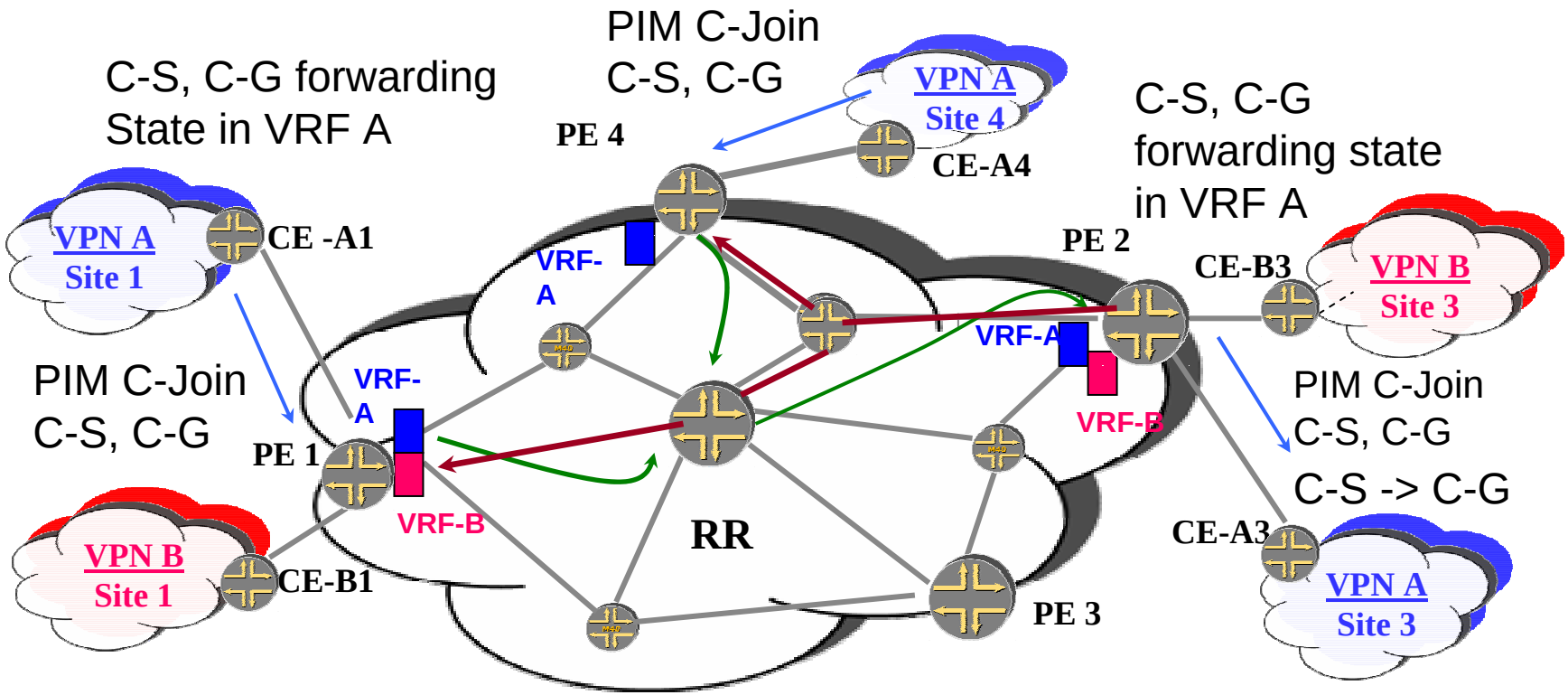
- This presentation will focus on JunOS implementation
  - For technology details see:
  - "Advanced Topics in NGEN MVPNs", Rahul Aggarwal, MPLS WC 2007
  - MVPN Tutorial by Yakov Rekhter, MPLS WC 2007
  - "Next-Generation Solution for Multicast in 2547 VPNs and VPLS", Rahul Aggarwal, MPLS WC 2006

NGEN MVPN Intra-AS Overview (Auto-Discovery, Inclusive Mapping)



BGP Signaled MVPN AD and Tree Binding by PE2:
<RD, PE2, VPN A RT, P2MP LSP Identifier>

NGEN MVPN Intra-AS Procedure Overview (PIM-SM in SSM Mode)



BGP MVPN C-Mcast Routing Information:
<RD, C-S, C-G, PE2's VRF A Route Import RT>

P2MP TE Inclusive P-Tnl: PE2 as Root

RR Aggregates C-Multicast Routes
- no explicit tracking by default

NGEN MVPNs JunOS Phase 1

- NGEN MVPN Intra-AS BGP PE-PE Control plane
- PIM-SM based Intra-AS Inclusive P-Tunnels
- IETF drafts supported
  - Draft-ietf-l3vpn-2547bis-mcast-bgp-03.txt
  - Draft-ietf-l3vpn-2547bis-mcast-02.txt

NGEN MVPNs JunOS Phase 1…

- Allows a SP to use the same PE-PE control plane protocol and infrastructure for BGP-MPLS VPN unicast and multicast
  - No need to run the draft-rosen PIM based virtual router control plane for multicast
- However the P-tunnel data plane is based on PIM-SM GRE tunnels
  - This limitation will be lifted in phase 2 with the support of P2MP MPLS TE LSPs

NGEN MVPNs JunOS Phase 1
BGP PE-PE Control plane

- MVPN Auto-Discovery/Inclusive Binding
  - Granularity of <PE, MVPN>
  - Binding one MVPN to a P-tunnel
- C-multicast routing information exchange among PEs
  - Granularity of <PE, C-S, C-G>
  - Mechanisms to exchange C-Control traffic through the SP network is independent from the mechanisms to exchange C-Data traffic

NGEN MVPNs JunOS Phase 1 BGP PE-PE Control plane…

- Introduces MCAST-VPN NLRI in JunOS BGP
  - Intra-AS Auto-Discovery (AD) routes
  - C-multicast Source Join route
    - Sufficient for C-PIM-SM in both SSM and ASM mode
  - Source Active (SA) routes
- Leverages JunOS VPN BGP features
  - E.g. Route Target Constrain is supported for MCAST-VPN NLRI
- Leverages JunOS VPN policy features
  - Auto-export is supported to enable MVPN extranet

BGP PE-PE Control Plane - Configuration

- Minimal requirements in addition to configuring 2547 unicast are
  - Enabling protocol mvpn in the routing instance
  - Enabling BGP to exchange family inet-mvpn routes
- By default the same RTs are used for intra-AS AD and SA routes as those configured for unicast
  - For certain topologies different RTs may be configured – more on this later
- Route Import Extended Communities are auto-generated

P-Tunnel - Configuration

- Required in a mvpn routing instance only if PE is connected to one or more sites in the MVPN that have sources
- JunOS Phase 1 supports only PIM-SM P-Tunnels
- P2MP TE LSP P-Tunnels planned for phase 2 in a following release

JunOS NGEN MVPN Phase 1 C-PIM Observations

- No explicit tracking by default
  - Only RR does "explicit tracking" for the RR cluster
- Route Import RTs are advertised in RT-constrain
  - Ensures that C-multicast route distribution is constrained

JunOS NGEN MVPN Phase 1 Extranet

- The ability for a CE that is in a particular VRF to receive multicast traffic from sources in another VRF.
  - This is also termed as overlapping MVPNs at times.
- BGP MVPN control plane lends itself naturally to supporting extranets
  - No extra configuration beyond that required for BGP-MPLS VPN unicast extranet is required
- Due to lack of time this presentation will not be able to describe the mechanism details

JunOS NGEN MVPN Phase 1
Hub and spoke

- Important functionality in BGP-MPLS VPN unicast
- An important requirement for MVPNs as well
  - Almost the same configuration as unicast
- Spokes send multicast traffic to the hub PE using a P-tunnel that has only the hub PE as the leaf
- Hub PE has two VRFs for the MVPN
  - Spoke to hub VRF receives traffic from the spokes
  - Hub to spoke VRF sends traffic to the spokes using a PIM-SM P-tunnel
- Hub PE advertises SA routes
  - C-RP is on the hub site
- Due to lack of time this presentation will not be able to describe all the details

NGEN MVPN JunOS Phase 1 Migration from Draft-rosen

- JunOS phase 1 has a goal of supporting migration to BGP-MVPN from an existing draft-rosen network
  - Details are being hashed out
- A given MVPN on a PE could run both draft-rosen [with some PEs] and BGP-MVPN [with other PEs]
  - Ships in the night model in the network per MVPN
- Thus a SP could migrate to BGP-MVPN one MVPN one PE at a time seamlessly
  - Data plane continues to be PIM-SM P-Tunnels

NGEN MVPN JunOS Phase 1 Migration from Draft-rosen…

- Upgrade the software to the new JunOS version supporting BGP-MVPN one PE at a time
  - Impacts only the customers on that PE
  - No need to upgrade the P-routers
- Configure BGP-MVPN one MVPN one PE at a time
  - Leave the draft-rosen configuration in place
- Once all the PEs are configured with BGP-MVPN, for a MVPN deconfigure draft-rosen on each PE for that MVPN
- Minimal disruption in service

NGEN MVPN JunOS Phase 1 Scalability

- Let us look at some of the scalability issues
  - Not exhaustive by any means
- PE control plane
- Route Reflector Scaling
- P-data plane

NGEN MVPN JunOS Phase 1 PE Control Plane Scalability

- PE – PE MVPN control plane scales similar to the PE-PE VPN unicast control plane
  - Number of MVPNs on a PE
  - Number of other PEs in a MVPN
- PE-CE control plane
  - Depends on PE-CE PIM control plane
    - Number of MVPNs on a PE are limited by the number of PIM sessions
    - Number of PIM C-multicast routes received from local CEs is limited by periodic PIM refresh of C-Joins
- Number of BGP C-multicast routes
  - The real limit is the total number of routes on a PE

NGEN MVPN JunOS Phase 1 Route Reflector Scaling

- A handful of BGP sessions for MCAST-VPN NLRI to a handful for Route Reflectors
  - RR scaling is a function of the memory consumption for C-multicast routes and the processing required to maintain these routes
  - This load can be spread across RRs
  - RRs can be dedicated to only MVPNs

NGEN MVPN JunOS Phase 1
P-Data plane Scaling

- Similar to VPLS P2MP scaling considerations
- Potential to aggregate leaves headroom for growth

Conclusion

- JunOS is continuing to build on video over P2MP MPLS TE solution
- JunOS is taking the lead in multicast optimization in VPLS
- JunOS is answering the need for a a BGP PE-PE control plane for NGEN MVPNs and a MPLS forwarding plane

Thank you!