Deploying Next-Generation Multicast VPN

Emil Gągała
PLNOG, Warsaw, 5.03.2010

Agenda

Introduction to Next-Generation Multicast VPN (NG-MVPN)

How to migrate smoothly from draft-Rosen to NG-MVPN

IPTV NG-MVPN case study

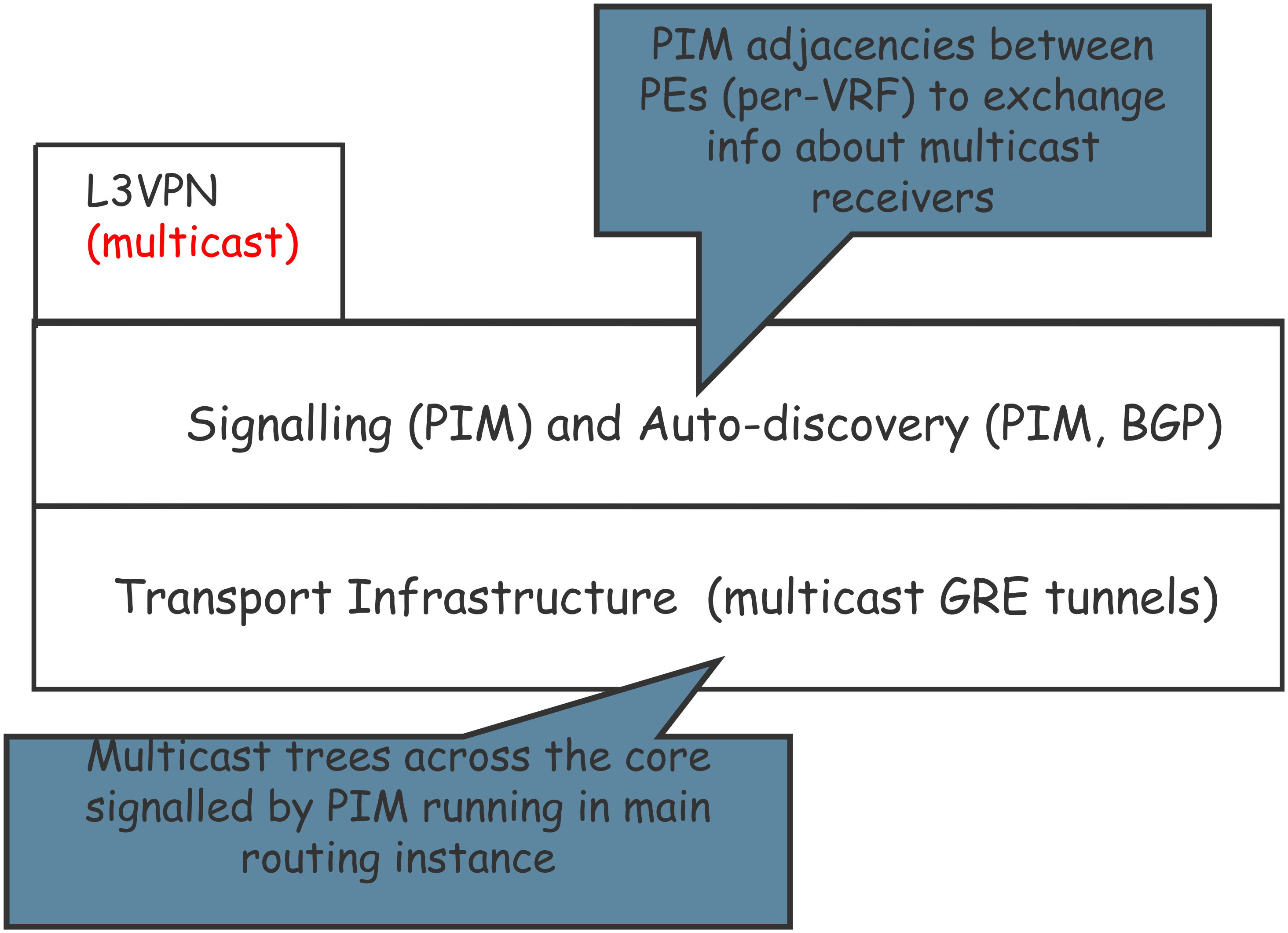
Multicast VPN

- Layer 3 BGP-MPLS VPNs are widely deployed in today's networks for forwarding VPN unicast traffic only.
- An "incremental" approach for deploying Multicast services can use the same technology as used for deploying Layer 3 VPN for unicast services.
- This approach can reduce the operational and deployment effort.
- Multicast applications, such as IPTV and multimedia collaboration, gain popularity
- There is demand for a scalable, reliable MVPN service over a shared MPLS infrastructure merging different service needs

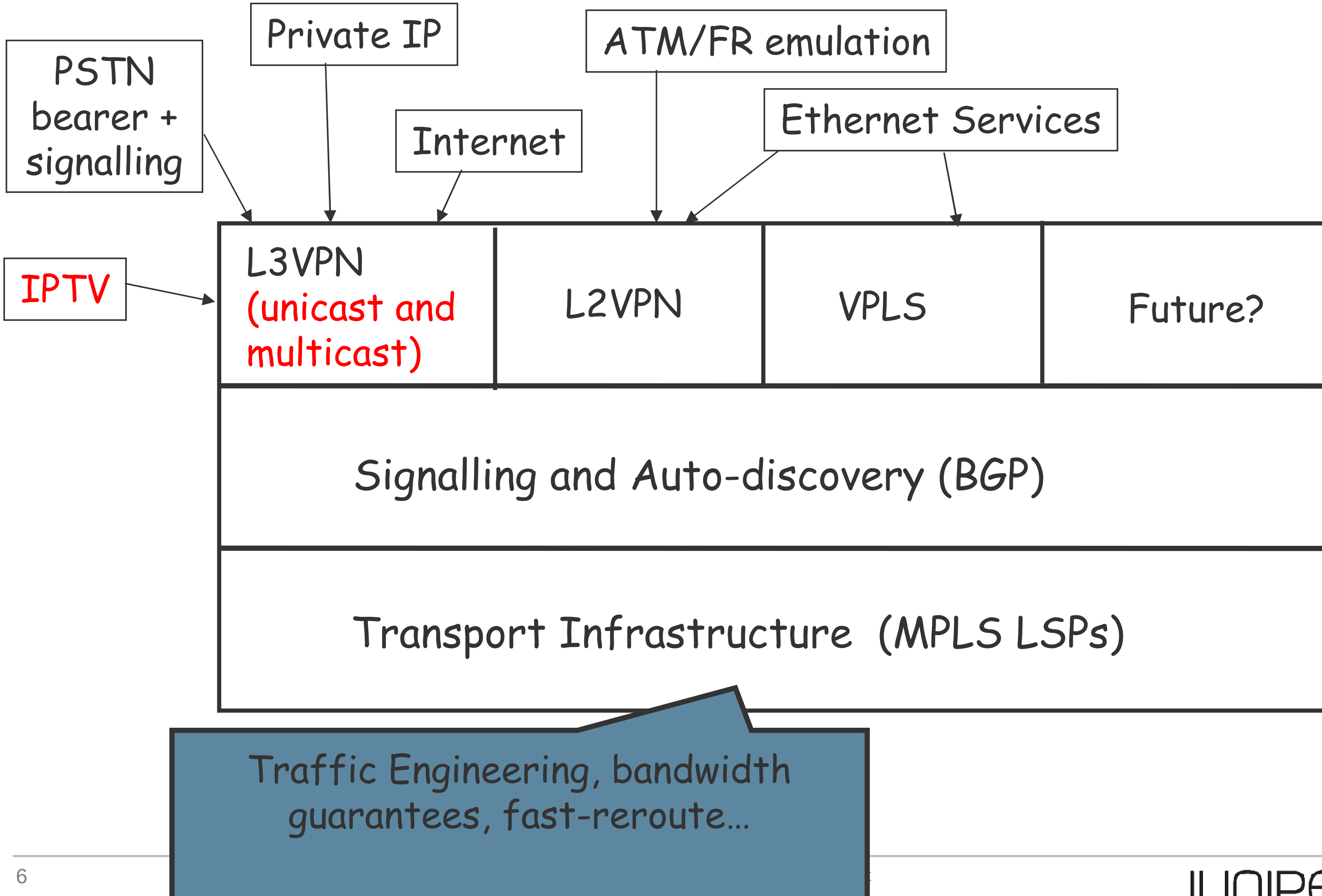
MCAST VPN Alternatives

- CE-CE GRE Overlay Tunnels
  - No multicast routing in the ISP’s core
  - Customer’s groups can overlap
  - Not scalable design – full mesh tunnels between CEs for each customer
  - Optimal multicast routing not achieved
- Draft Rosen Multicast VPN
  - Introducing Multicast VRF type
  - Based on native IP multicast (PIM SM/SSM mode) in the ISP’s core network – customer’s multicast is tunneled within ISP’s core native IP multicast using multicast GRE tunnels
  - Customer’s PIM adjancency with PE routers
  - Based on draft-rosen-vpn-mcast-[xy].txt (the latest is draft-rosen-vpn-mcast-12.txt)
- Next Generation Multicast VPN
  - In the past there was no way of carrying multicast traffic over MPLS but this all changed with the invention of “Point-to-Multipoint (P2MP) LSPs”
  - NG MVPN main architecture draft defined by draft-ietf-l3vpn-2547bis-mcast-08.txt

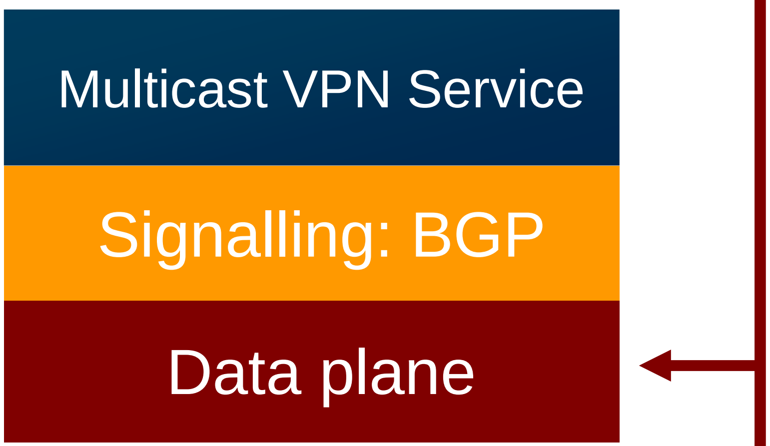
Draft Rosen MVPN Scheme



NextGen MVPN Scheme



Next Generation Multicast VPN



Wide choice of data-planes:

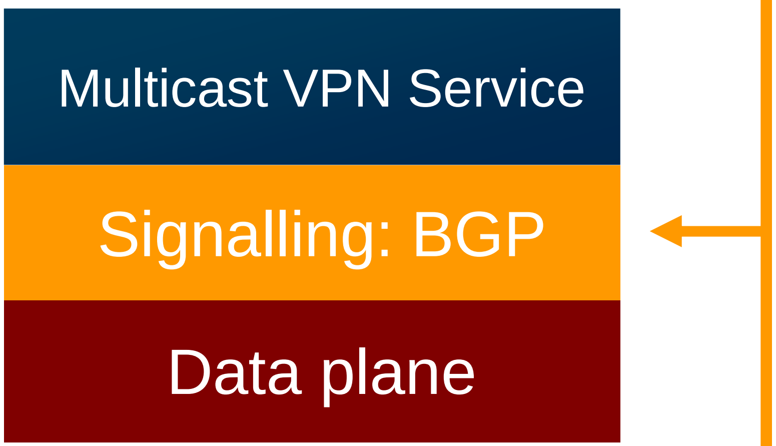
- PIM-GRE tunnels
- RSVP-P2MP LSP
- LDP-P2MP LSP
- Ingress Replication

} PIM-free core (RSVP-P2MP LSP, LDP-P2MP LSP, Ingress Replication)

P2MP LSP data-plane:

- MPLS encapsulation just like for unicast
- RSVP-P2MP gives Traffic Engineering, MPLS Fast Reroute, Path Diversity, Admission Control

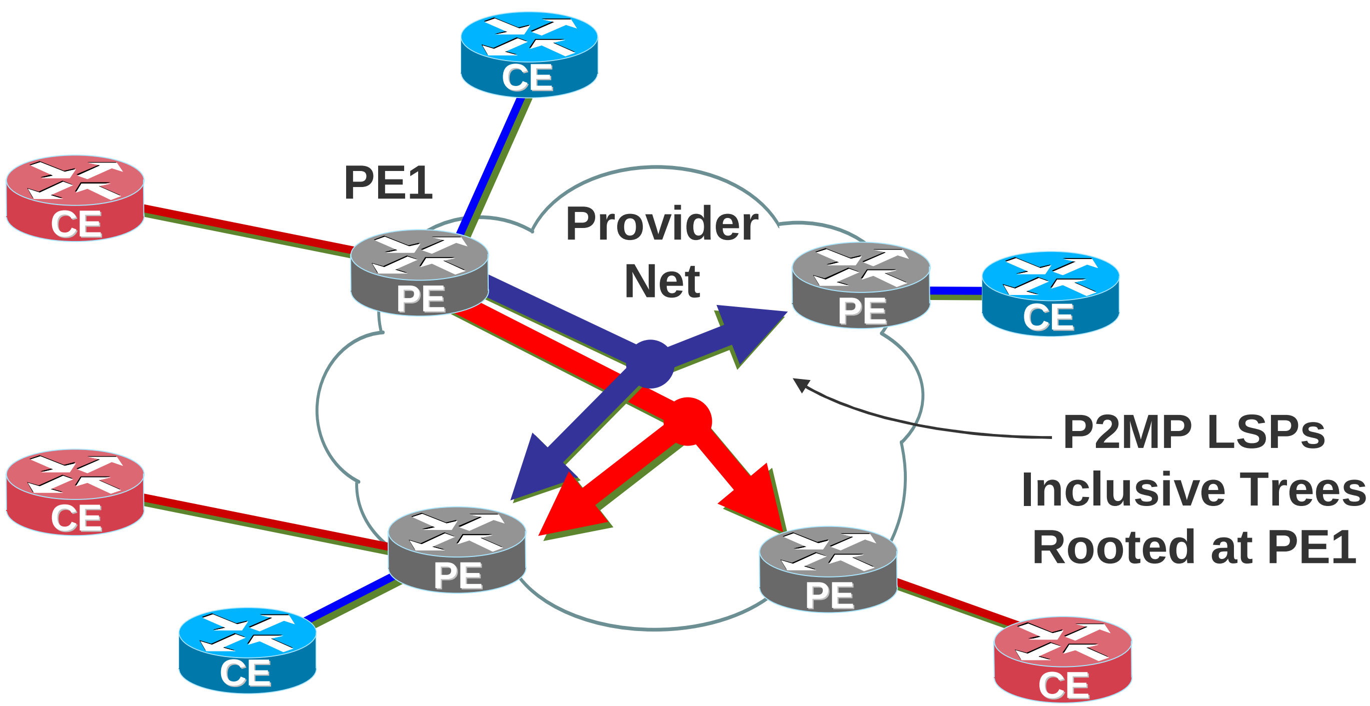
Next Generation Multicast VPN



- Same BGP control plane as used for L3VPN unicast, BGP-L2VPN, BGP-VPLS…
- Can use same RRs and BGP sessions if desired
- More scalable than draft-Rosen VR model
- Cleaner Inter-provider schemes
- Easy to build Extranets, using same technique as unicast L3VPN Extranets
- Fine-grain single forwarder election

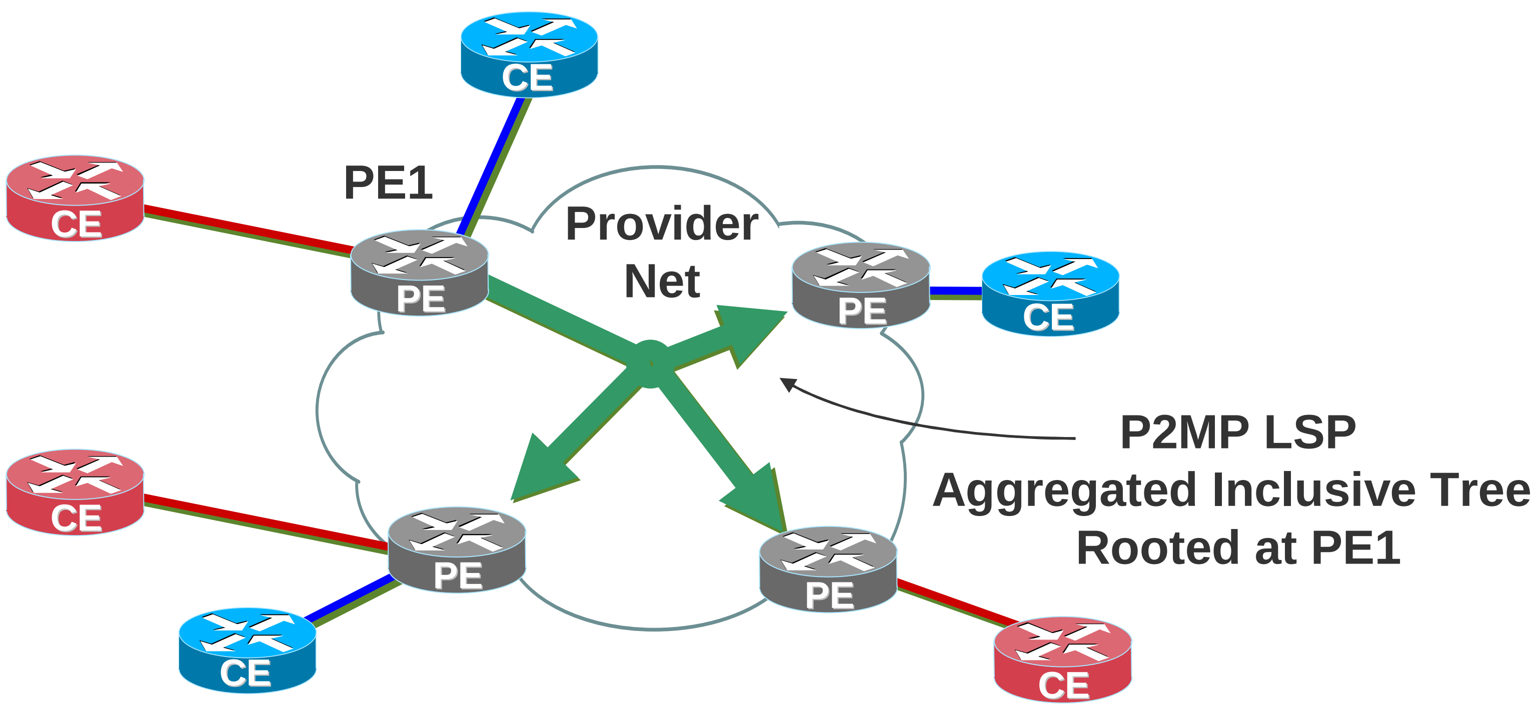
Inclusive Tree

So called **Inclusive Trees** - analogous to Default-MDT in draft-Rosen



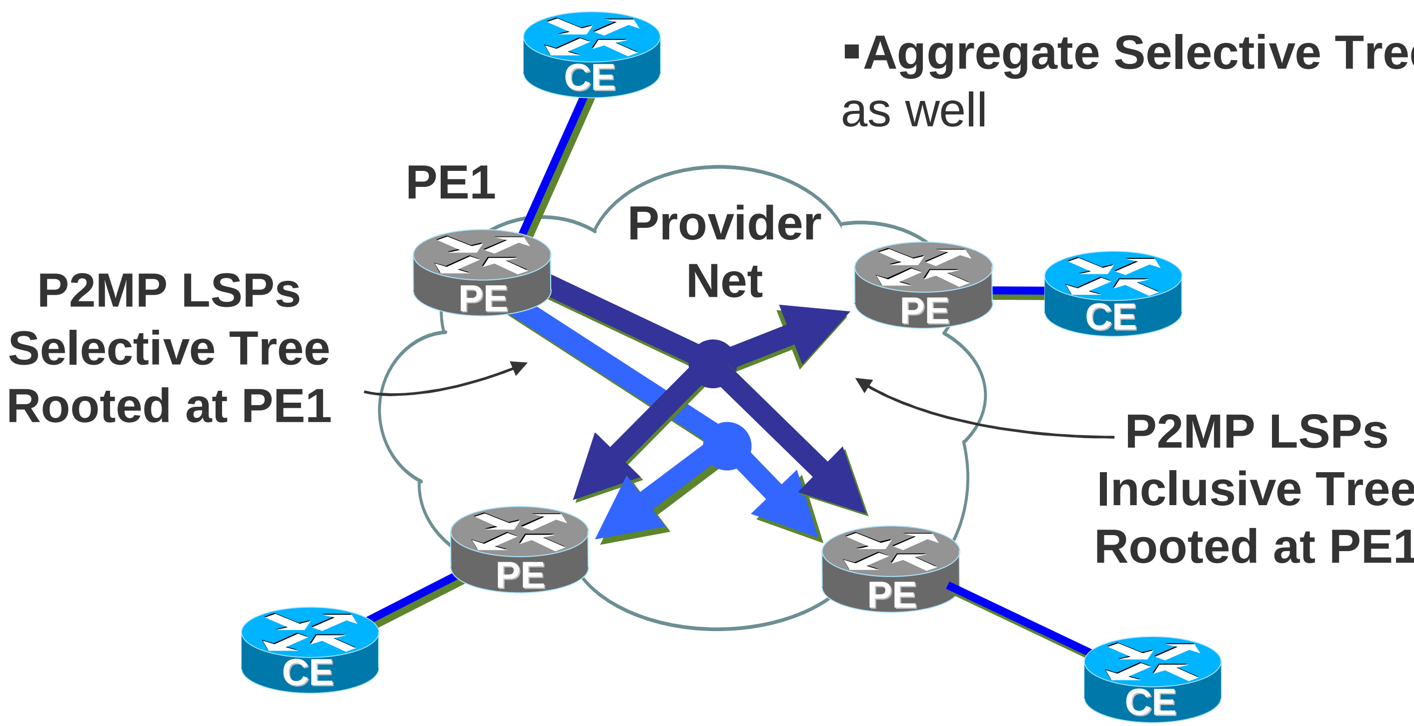
Aggregate Inclusive Tree

All the multicast groups in more than one MVPN use the same shared tree!



Selective Tree

- **Selective Tree** - analogous to Data-MDT in draft-Rosen
- Serves particular selected multicast group(s) from a given MVPN with Active Receivers
- **Aggregate Selective Tree** is possible as well



BGP Control Plane Functions

- **MVPN Membership Autodiscovery** - Discovery of which PEs are members of each MVPN and communication between PEs (NextGen VPN Alternatives are PIM based or BGP based – preferred one)

- **MVPN to Tunnel Mapping** - A PE router needs to know what type of tunnel and identifier to use for sending (and receiving) multicast data for a particular MVPN.

- **PE-PE C-multicast Route Exchange** - A PE router participates in the customer multicast (C-multicast) routing protocol by forming multicast routing adjacencies over its VPN interface.

BGP MCAST-VPN Address Family

- The new BGP address family (SAFI 5) is called MCAST-VPN and used for distributing MVPN control information between PE routers – so called “**mvpn routes**”
- There are seven types of mvpn routes:
  - Type 1 - Intra-AS auto-discovery route (A-D route)
  - Type 2 - Inter-AS auto-discovery route (inter-AS A-D route)
  - Type 3 - S-PMSI (Selective P-Multicast Service Interface) A-D route
  - Type 4 - Intra-as leaf A-D route
  - Type 5 - Source Active A-D route (or SA route)
  - Type 6 – Shared Tree Join Route (C-multicast route)
  - Type 7 – Source Tree Join Route (C-multicast route)
- The first 5 mvpn routes can be considered as the auto-discovery routes while last two are used for C-multicast routing exchange between PE routers of an MVPN.

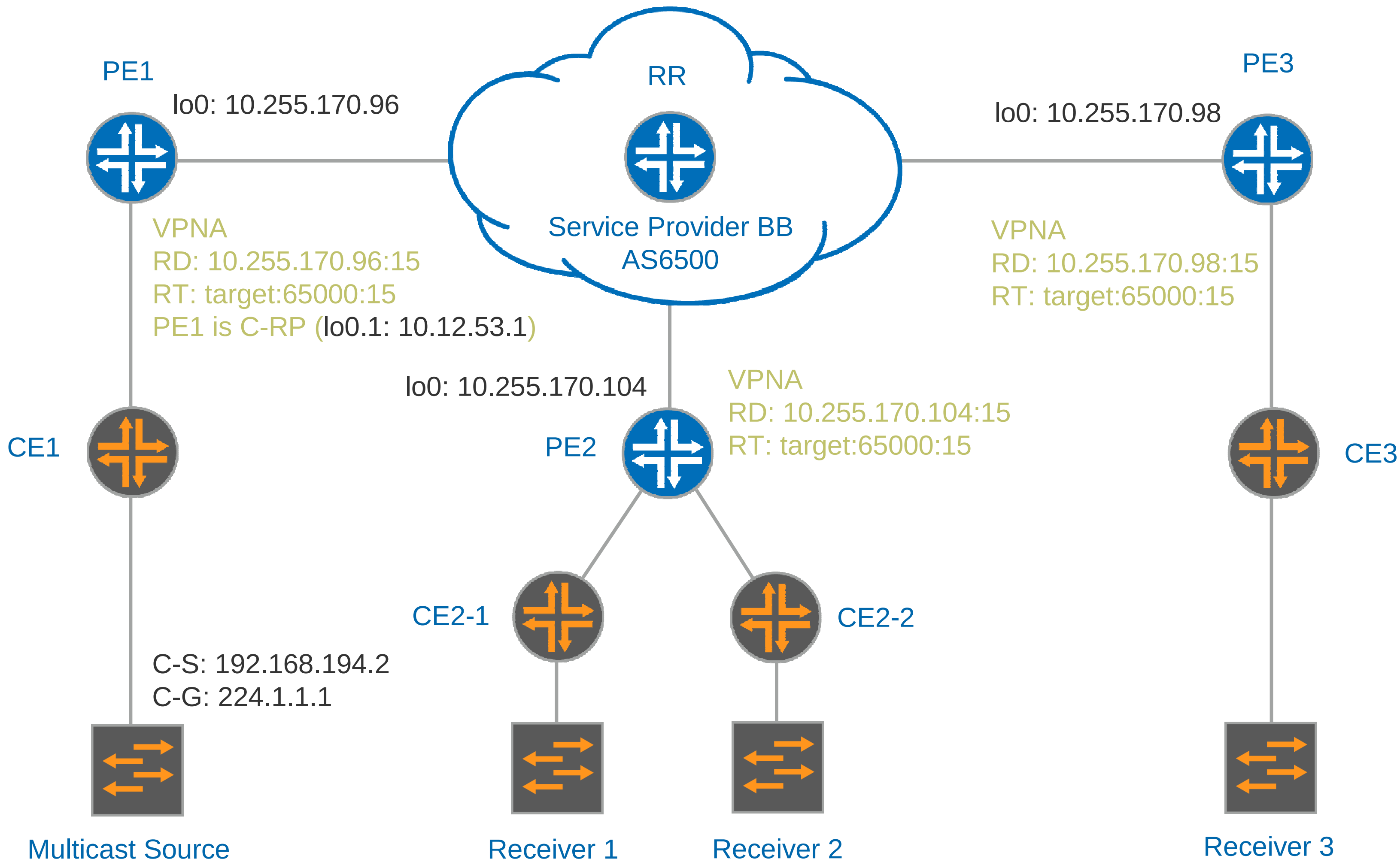
Acronyms and analogies

Conceptual Route Distribution

▪ PIM Hello	→	BGP AD Route
▪ PIM Join	→	BGP C-Multicast route
▪ PIM Register	→	BGP AD-SA route
▪ MSDP SA	→	BGP AD-SA route

NG-MVPN Control Plane - Continued

Reference Network



JUNOS NG-MVPN Implementation

PE1

NG-MVPN Routing Information Flow Summary

PE2/3

iBGP

Step 1: PE1/PE2/PE3 IBGP sessions are established with INET-VPN and MCAST-VPN NLRIs.

PE1

PE2/3

INET-VPN NLRI

Step 2: PE1 advertises VPN-IP unicast routes (including route to C-S) to PE2/PE3 via INET-VPN NLRI: 10.255.170.96:15:192.168.194.0/24 It attaches three communities to this route: target:65000:15 src-as:65000 *rt-import:10.255.170.96:3*

PE1

PE2/3

INET-VPN NLRI

Step 3: All PEs originate and advertise an A-D route: PE1 A-D Route: 1:10.255.170.96:15:10.255.170.96 PE2 A-D Route: 1:10.255.170.104:15:10.255.170.104 PE3 A-D Route: 1:10.255.170.98:15:10.255.170.98 PEs attach RT to A-D routes: target:65000:15 PE1 also attaches a PMSI attribute to the A-D route based on P-tunnel configuration.

JUNIPER NETWORKS

JUNOS NG-MVPN Implementation

PE1

NG-MVPN Routing Information Flow Summary

PE2/3

Step 4: PE1 binds MVPN to the tunnel advertised via PMSI
PE2 and PE3 join (or create necessary state to receive traffic) through the tunnel identified in the PMSI attribute.

PE2/3

CE2-1/2-2
CE3

(C-*, C-G) Join

Step 5: Receivers come online
PE2/PE3 receive (C-*,C-G) from CEs: (*, 224.1.1.1)

PE2/3

Step 6: PE2/PE3 does a route lookup in the VRF unicast table for C-RP, C-RP RD, Source AS and VRF Route Import communities: C-RP: 10.12.53.1 RD:10.255.170.96:15 src-as:65000 *rt-import:10.255.170.96:3*

PE2/3

Step 7: PE2/PE3 constructs Type 6 Shared Tree C-multicast route: 6:10.255.170.96:15:65000:32:10.12.53.1:32:224.1.1.1

JUNOS NG-MVPN Implementation

NG-MVPN Routing Information Flow Summary

PE1 ← **C-Multicast Data** ← CE1

Step 8: Source becomes active PE1 receives data for (C-*,C-G) from CE1: (*, 224.1.1.1)

PE1 → **MCAST-VPN NLRI** → PE2/3

Step 9: PE1 (C-RP) originates an SA A-D route and advertises it to PE2/PE3: 5:10.255.170.96:15:32:192.168.194.2:32:224.1.1.1

PE1 ← **MCAST-VPN NLRI** ← PE2/3

Step 10: PE2 and PE3 both originate and advertise a Type7 C-multicast route to PE1 and each other:
7:10.255.170.96:15:65000:32:192.168.194.2:32:224.1.1
Only PE1 accepts the route because of the unique RT the route carries: *rt-import:10.255.170.96:3*
PE2/PE3 discard the route they received from each other due to non matching RT values

JUNOS NG-MVPN Implementation

PE1

NG-MVPN Routing Information Flow Summary

Step 11: PE1 compares RT of C-multicast mvpn route to the special mvpn RT community (whose value is set to VRF Route Import community). If there is a match, the C-multicast route is accepted and (C-S,C-G) is passed to C-multicast protocol on PE1/VPNA to be processed.

PE1

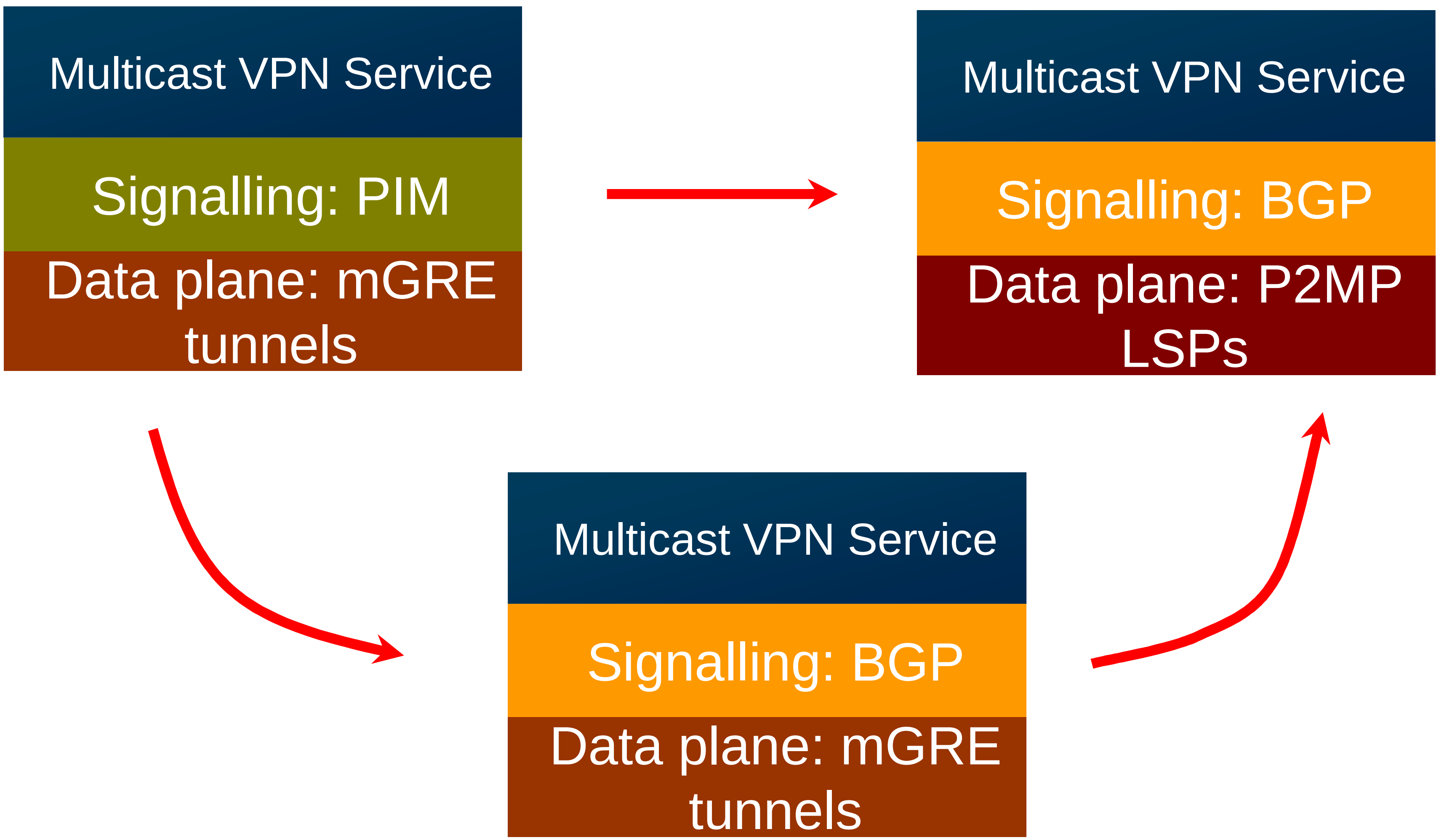
(C-S, C-G) Join

CE1

Step 12: PE1 creates state in C-PIM database and propagates (C-S, C-G) to CE1 towards the source.

Migration to NG-MVPN

Draft-Rosen to NG-MVPN migration options

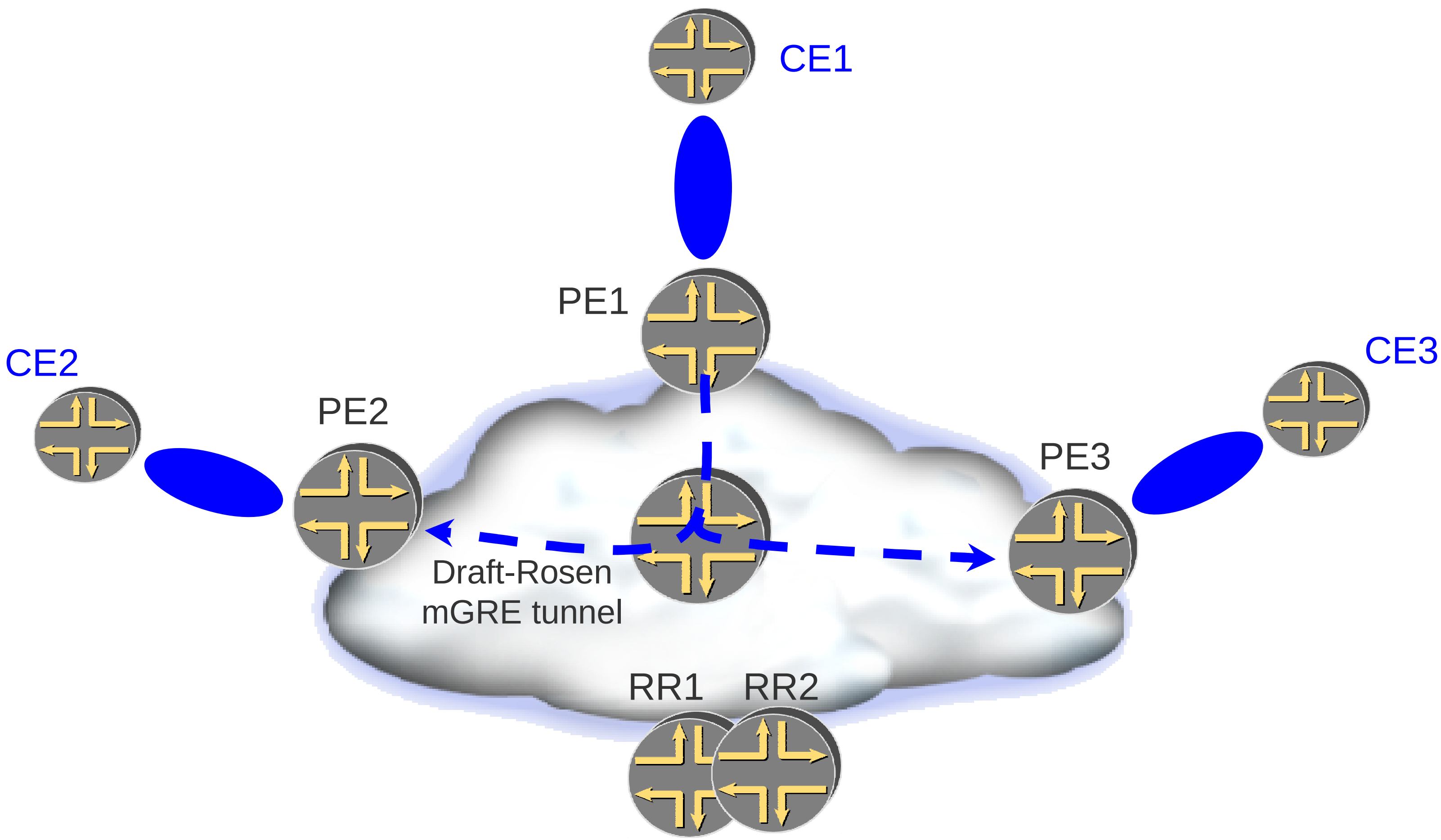


Migrating from draft-Rosen to NG-MVPN

How can we smoothly migrate a given mVPN from draft-Rosen to NG-MVPN

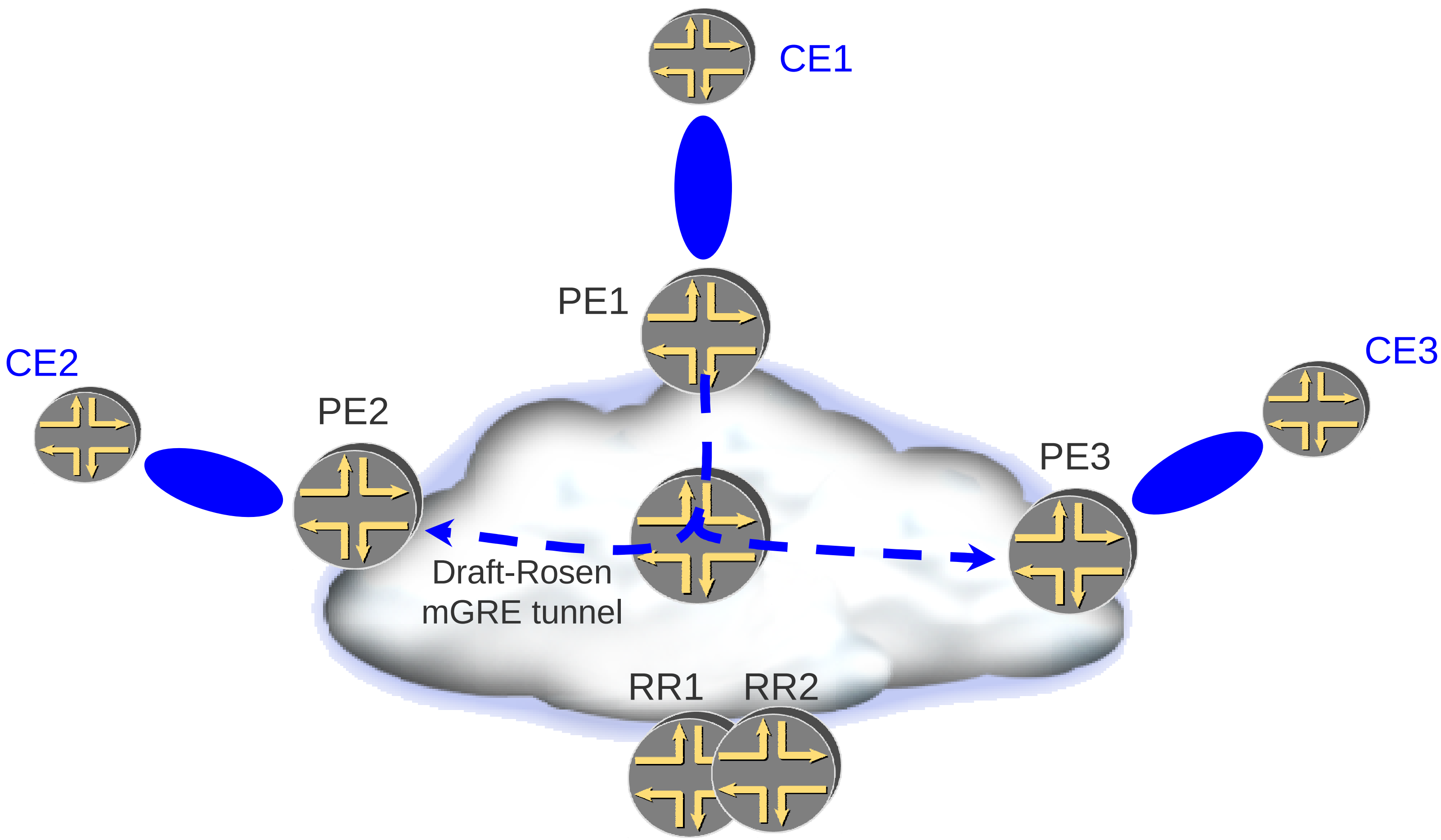
- With minimal traffic disruption?
- Avoiding having to change the configuration of all the PEs involved in the VPN simultaneously?

Migration to NG-MVPN: initial state



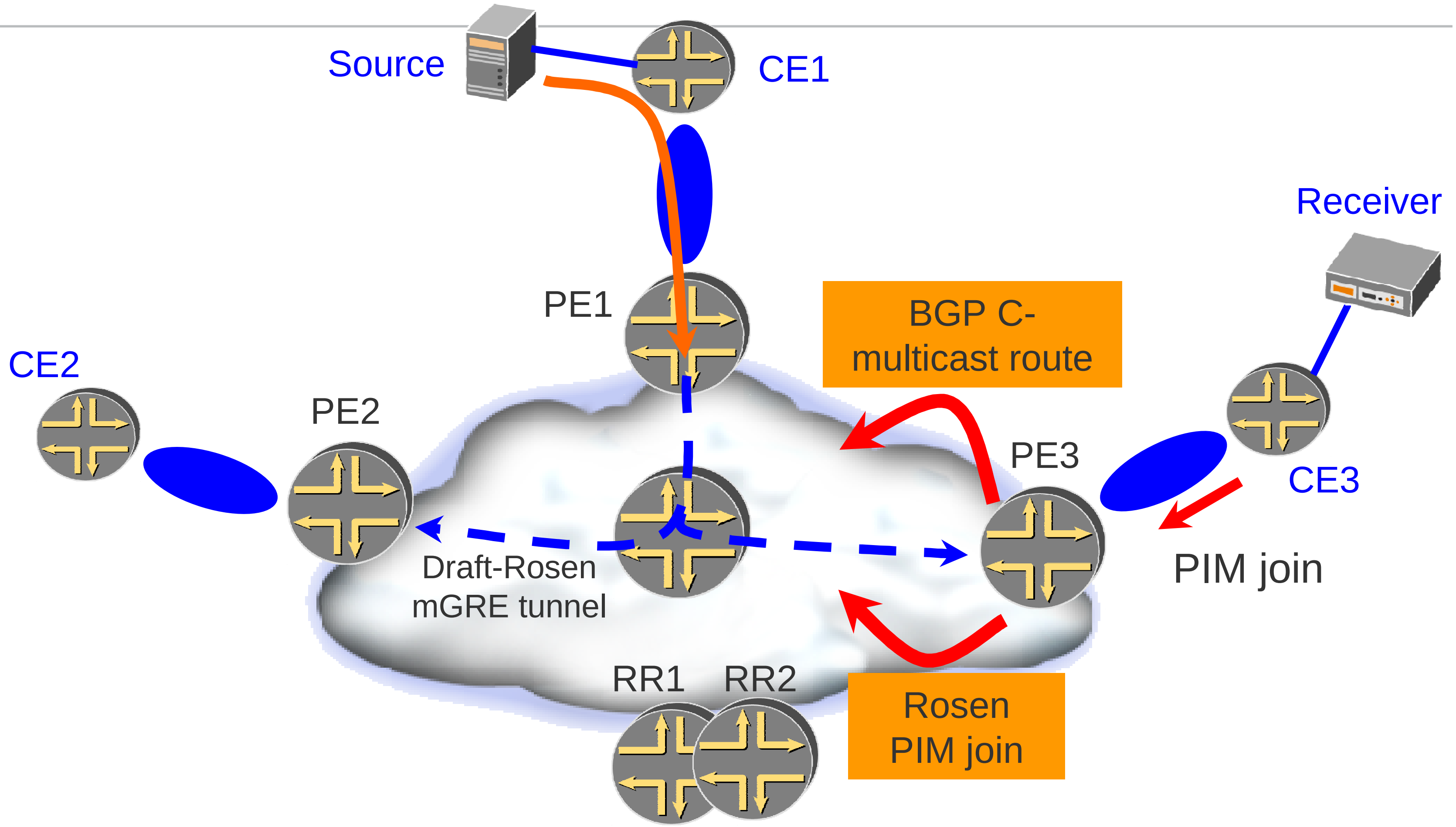
Blue VRFs all have draft-Rosen configuration (PIM control plane and mGRE data plane)

Migration step 1: NG-MVPN control plane



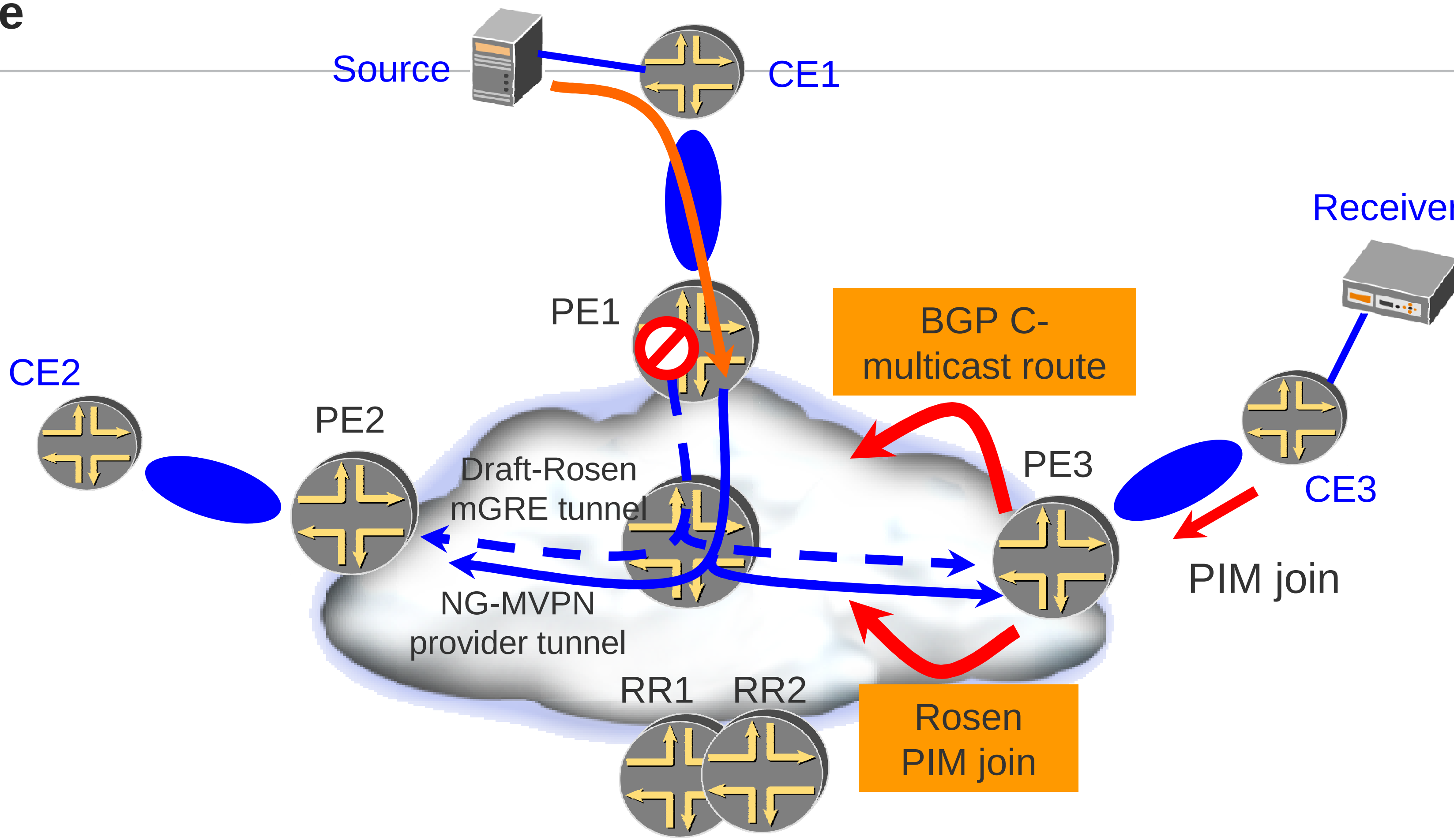
- Turn on NG-MVPN BGP address family on the BGP sessions.
- Turn on NG-MVPN control plane in the blue VRFs.
- (Note: All blue VRFs still have configuration for draft-Rosen after these steps)

After activating NG-MVPN BGP control plane



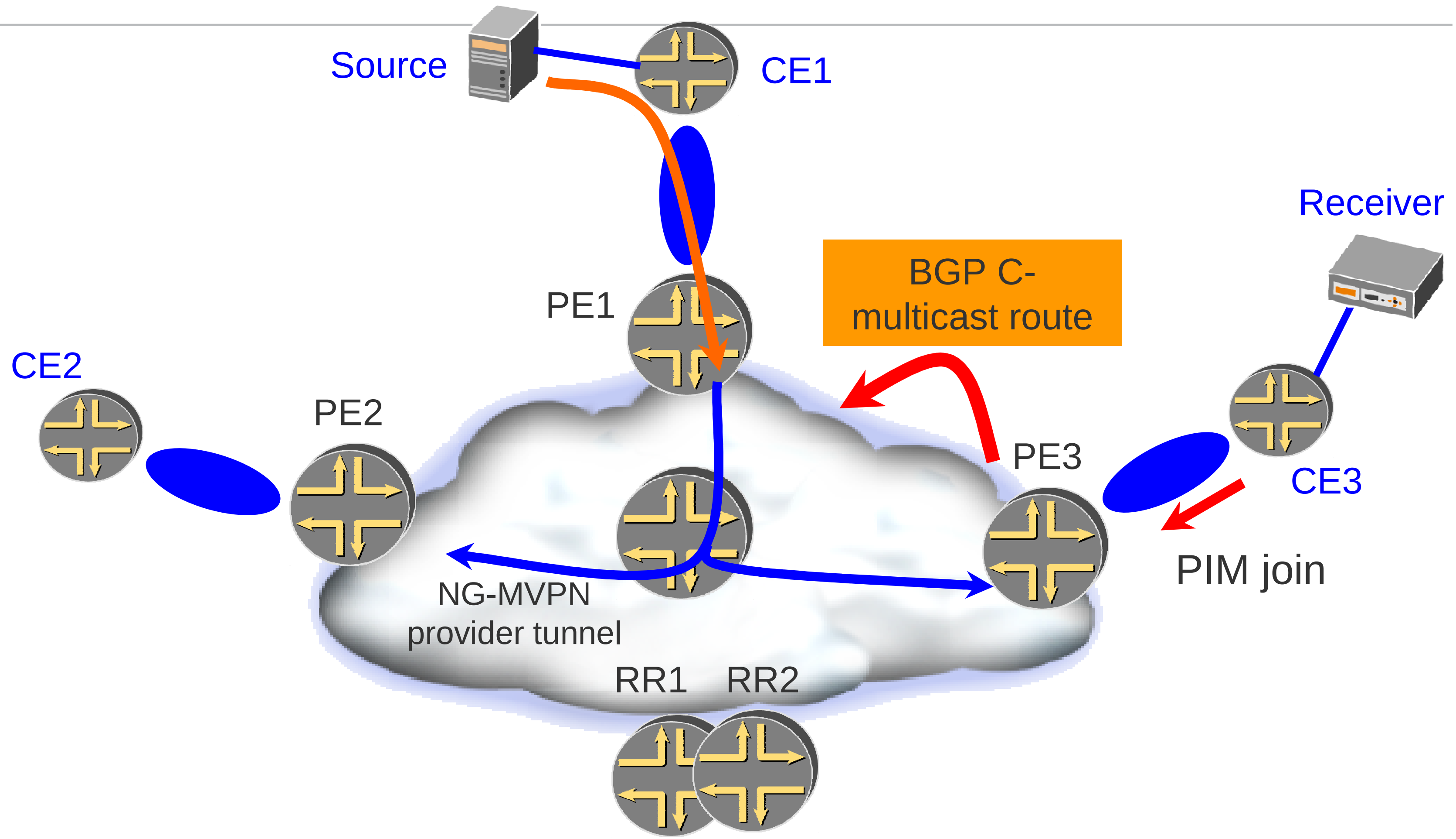
- At this stage, the blue VPN is running *both* Rosen and NG-MVPN *control plane* simultaneously.
- Rosen mGRE tunnel *data plane* is still used at this stage.

Migration step 2: activate NG-MVPN provider tunnel in each sender site



- Activate NG-MVPN provider tunnel in each blue VRF. This is preferred by an ingress PE (e.g. PE1) over the Rosen tunnel, so it moves its traffic onto it.
- During this migration step, a PE might be receiving traffic on NG-MVPN provider tunnels from some PEs and on Rosen tunnels from other PEs.

Migration step 3: remove Rosen config from blue VRFs



- Finally, Rosen configuration is removed from each of the blue VRFs.
- If mGRE provider tunnels are being used for the NG-MVPN data plane, they can be replaced by P2MP LSP provider tunnels once all the P-routers support P2MP LSPs

Next-Generation MVPN deployments

NG-MVPN deployments are exercising a wide spectrum of the tool-kit

- BGP Control Plane
- P2MP RSVP Data Plane
  - Shortest Path Trees/Minimum cost trees
  - Online path computation using CSPF/Offline path computation
- Inclusive trees/Selective trees
- PIM-SSM in C-domain/PIM-ASM in C-domain
- MPLS FRR link protection
- Single Forwarder Election
  - Default (PE with highest IP address) versus fine-grain (BGP selection rules)
- Live-Live/Live-Standby
  - P2MP-RSVP data plane makes it easy to ensure path diversity for live-live traffic

Next-Generation MVPN deployments at MPLS World Congress 2010

NG-MVPN live deployment examples

- BT - distribute all of the UK Digital Terrestrial TV channels to the transmitter sites
- BGC Partners - use of NG-MVPN for real-time financial transactions
- Cox Communications - use of NG-MVPN for Cable TV and Video-on-Demand distribution infrastructure
- FT Orange Business Services - discuss the use of NG-MVPN as a service offering to MNC and Enterprise customers

NG-MVPN IPTV distribution Case Study

Thanks to Rafał Szarecki!

Customer

One of top 3 in a country

Owned by other Telco form neighbor country

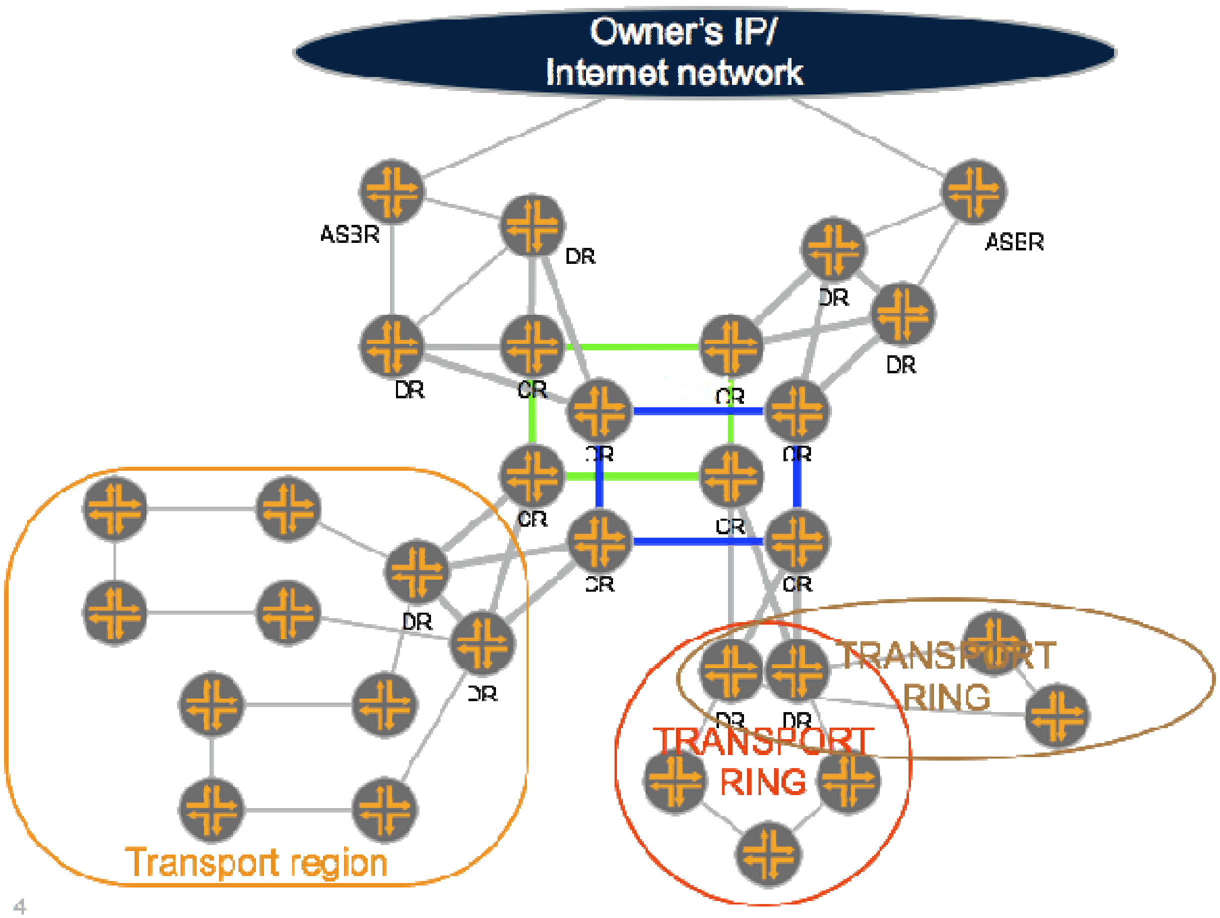
Offers Data, Telephony and Mobile service for residential and businesses. Address IPTV too.

Huge project to unify and modernize IP/MPLS network in last year.

- (PLNOG3: 200 PE network presentation talk obout same case)

It runs IPTV testing on old network base on PIM and IP multicast.

NETWORK



L2 (Eth) DSLAMs are connected to routers on transport rings.

IGMPv2 only from DSLAM/CPE.

DSLAM supports IGMP proxy feature.

Internet is running in other VPN

IPTV – what is given

IPTV head end is out of network – in neighboring country, owned and operated by owner TELCO

- IPTV streams delivered over Internet interconnect in the same context. (same VLAN)
- PIM-SM required.

However MiddleWare local

- Private addresses
- Separation from internet is most welcome.

Separate VLAN/VC for Internet and for IPTV

- Single VLAN (50) per DSLAM for multicast
- DSLAM do IGMP proxy to send requested group to interested subscriber only.

VoD planned. Same context as MW.

~100 channels.

Requirements

Avoid PIM and IP multicast – it is hard to manage.

Avoid extensive per-router configurations

- Static (*,G)->(S,G) mapping

Failures are destructive – couple hundreds of miliseconds (couple of I-frame lost). Avoid/limit impact if possible.

IPTV observations

DSLAM has > 1000 subscribers, and few DSLAM is connected to single router. High probability that at least one watcher per channel will be active on local DSLAMs.

Not all (200) routers handle DSLAMs, rather ~40.(Co-location with PTT)

Due to topology – transport rings – router may need to handle IPTV traffic even there is no watcher on its DSLAMs.

With 99% prob. all channels need to be delivered to every router.

Solution

MPLS P2MP address sub second traffic restoration (~50ms, observed even faster)

NG-MVPN configured on this routers where DSLAM are connected (once) – it allow for auto-discovery.

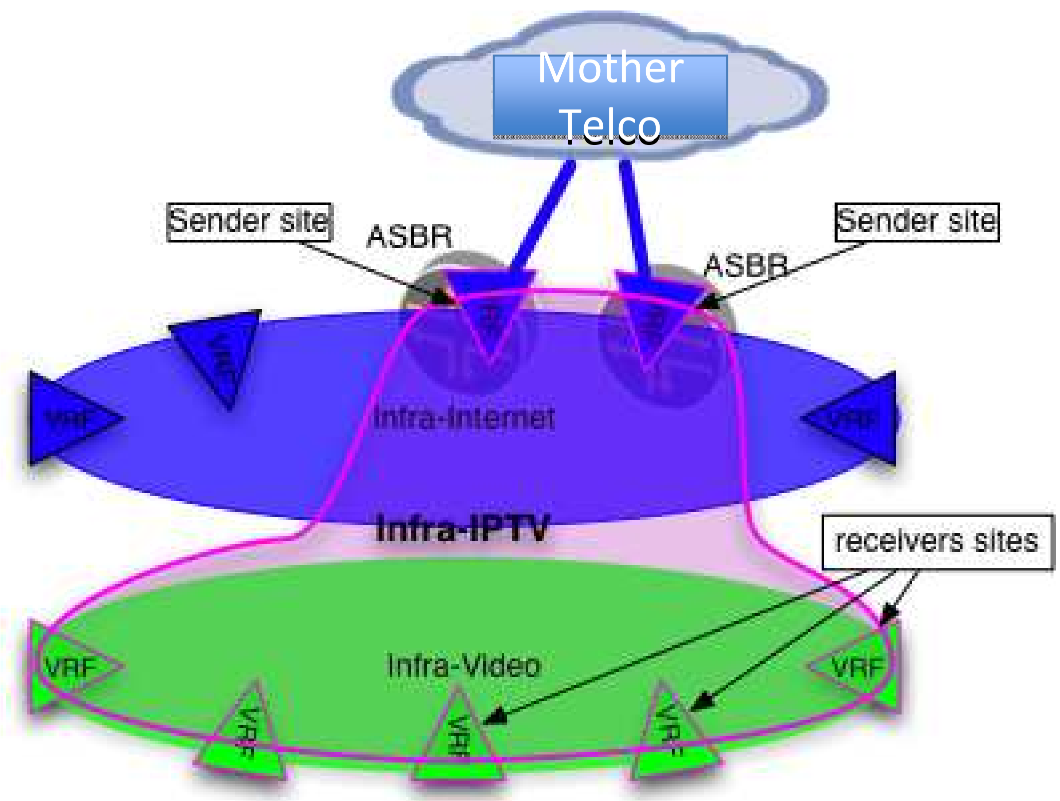
I-PMSI is used

- Limit number of states – One P2MP LSP for all channels (x 2)
- Minimal BW waste due to high congruency (see above)
- Use of templates to build branches.

NG-MVPN

- allows for separation of IPTV/VoD related unicast traffic form Internet traffic
- Multicast is separated by definition – separate LSPs
- However IPTV and Internet shares VRF on ASBRs.

Design



Overlapping VPN – same as RFC 2547bis (RT driven)

- Infra-IPTV is NG-MVPN for carry multicast traffic.
- Infra-Video is unicast VPN, and may have sites not belonging to infra-IPTV (e.g. VoD)

Note that blue and yellow/purple VRF are on same router (except ASBR)

How it works:
After provisioning

All receiver (and two source) sites are discovered by iBGP

P2MP LSP is signalled by each ASBR (source site). Designation IP of each branch is extracted form iBGP discovery phase.

No M-cast traffic so fare. Not even on NNI

How it works:
First watcher connects to channel (G)

TR (receiver site) receives IGMPv2 (*,G) report in VRF context

- Receiver site converse it to (S,G)
- Single Designated Forwarder for Source address is selected.
- Receiver site TR sends MVPN (type 7) C-(S,G) update in iBGP.

Both ASBR receives update. One of them is Designated Forwarder.

- Designated Forwarder creates PIM Join for (S,G) and sends over NNI
- Designated Forwarder receives M-cast traffic on NNI (S,G)
- Designated Forwarder forward M-cast to I-PMSI (P2MP LSP)

<u>All</u> receiver sites (all TR configured for Infra-Video VPN) receives M-cast traffic

- high congruency expected !
- Reduces zapping time

Receiver site where watcher's DSLAM is connected forward M-cast traffic.

- Only to watcher's DSLAM interface/vlan. (DSLAM is proxing traffic to watcher only)
- No forwarding to DSLAMs connected to other interfaces

Other receiver sites (other TR) drops (S,G) – no forwarding state.

How it works:
Next watchers connect to channel (G)

On same DSLAM as previous

- No action on TR and MVPN

On same TR (receiver site) but other DSLAM

- IGMPv2 is received on receiver site VRF, and forwarding entry is created.
- No MVPN signaling
- Receiver site VRF forward traffic on additional interface.

On other TR.

- IGMPv2 is received on receiver site VRF, and forwarding entry is created.
- Receiver site VRF forward traffic on interface (zapping time !)
- Single Designated Forwarder for Source address is selected.
- Receiver site TR sends MVPN (type 7) C-(S,G) update in iBGP.
  - Suppressed by RR (in case RR are used)
  - Used in case other TR () withdraw his C-(S,G) path

SUMMARY

Summary

Low number of states – only one P2MP LSP

<50ms restoration – MPLS FRR in case of link failure. Lossless recovery after repair of link.

No per group / per (S,G) provisioning

Automatic P2MP LSP endpoint discovery

Service separation

No PIM in network (except NNI)

References

- *"Multicast in MPLS/BGP IP VPNs",* draft-ietf-l3vpn-2547bis-mcast-10.txt
- *"BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs",* draft-ietf-l3vpn-2547bis-mcast-bgp-08.txt
- *"Mandatory Features in a Layer 3 Multicast BGP/MPLS VPN Solution",* draft-ietf-l3vpn-mvpn-considerations-05.txt
- *"Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC4875*
- *"Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*", draft-ietf-mpls-ldp-p2mp-08

everywhere