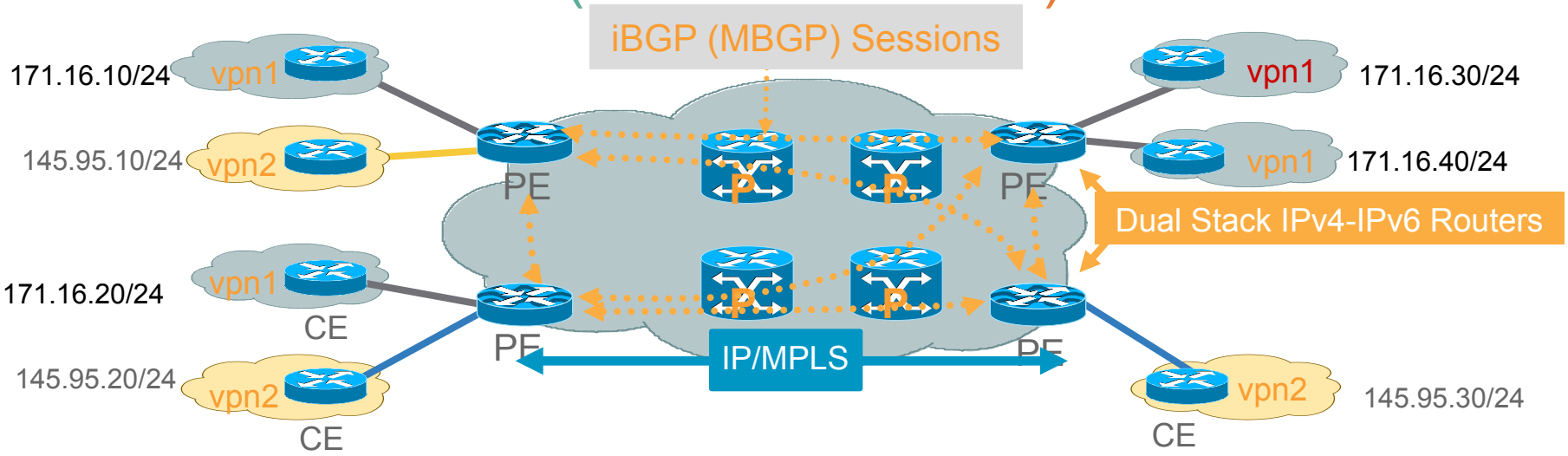
Next Generation Multicast VPN

Hernán Contreras G.
Consulting Systems Engineer
Cisco Systems

Next Gen MVPN Introduction

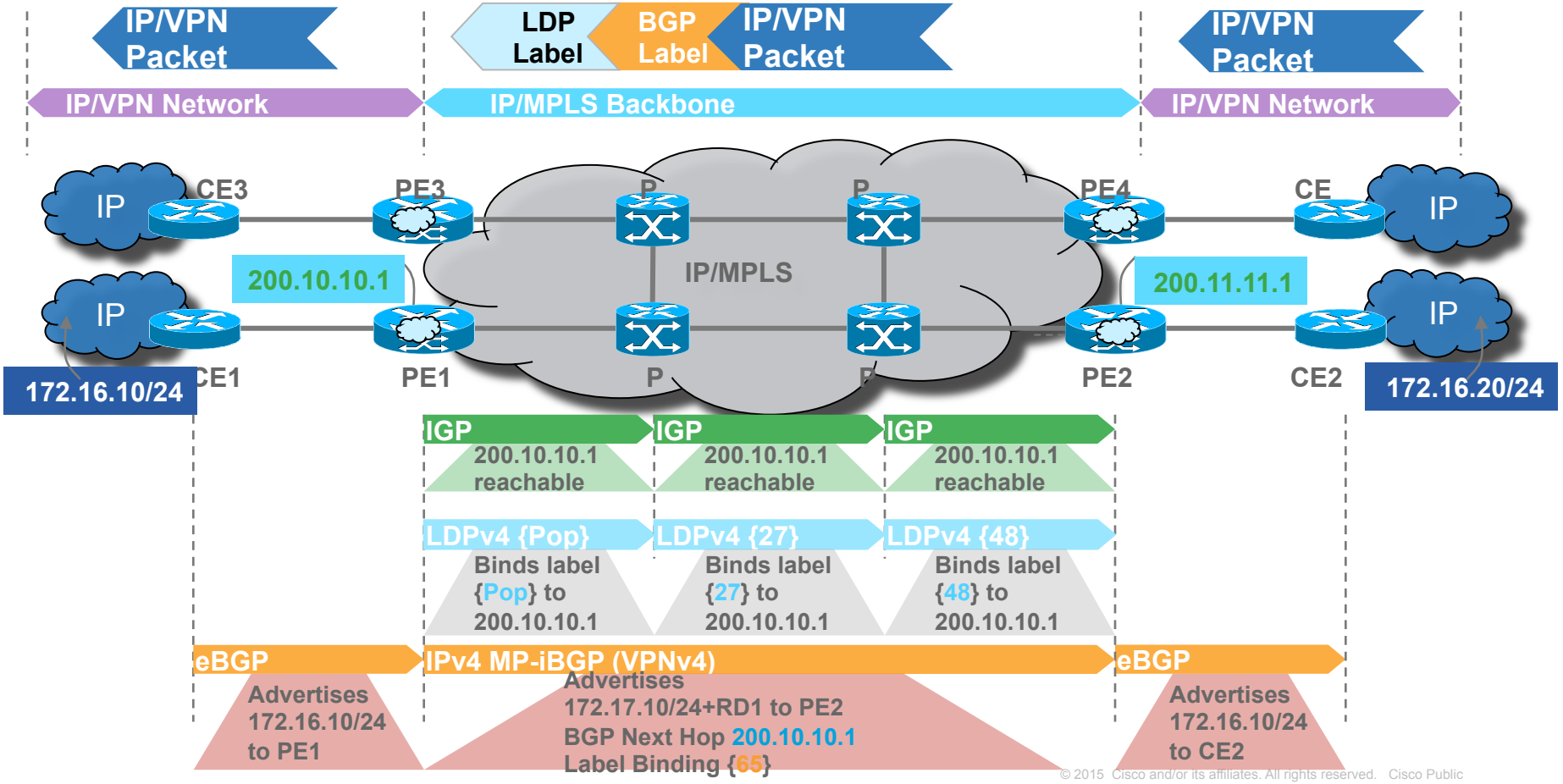
- Solutions to support forwarding multicast traffic in the context of VRFs
- Current MVPN Deployments are based on RFC 6037 (“Rosen Model”).
- NextGen (NG) MVPN refers to Support for RFC 6513/6514 based MVPNs
  - Superset of Rosen MVPN
- The new MVPN Architecture define MVPNs in terms of the following components
  - Core-Tree Protocol
  - Auto-Discovery
  - C-multicast Routing Protocol
  - MDT Model

BGP/MPLS IP Virtual Private Networks
Connection Model (RFC 4364/2547)



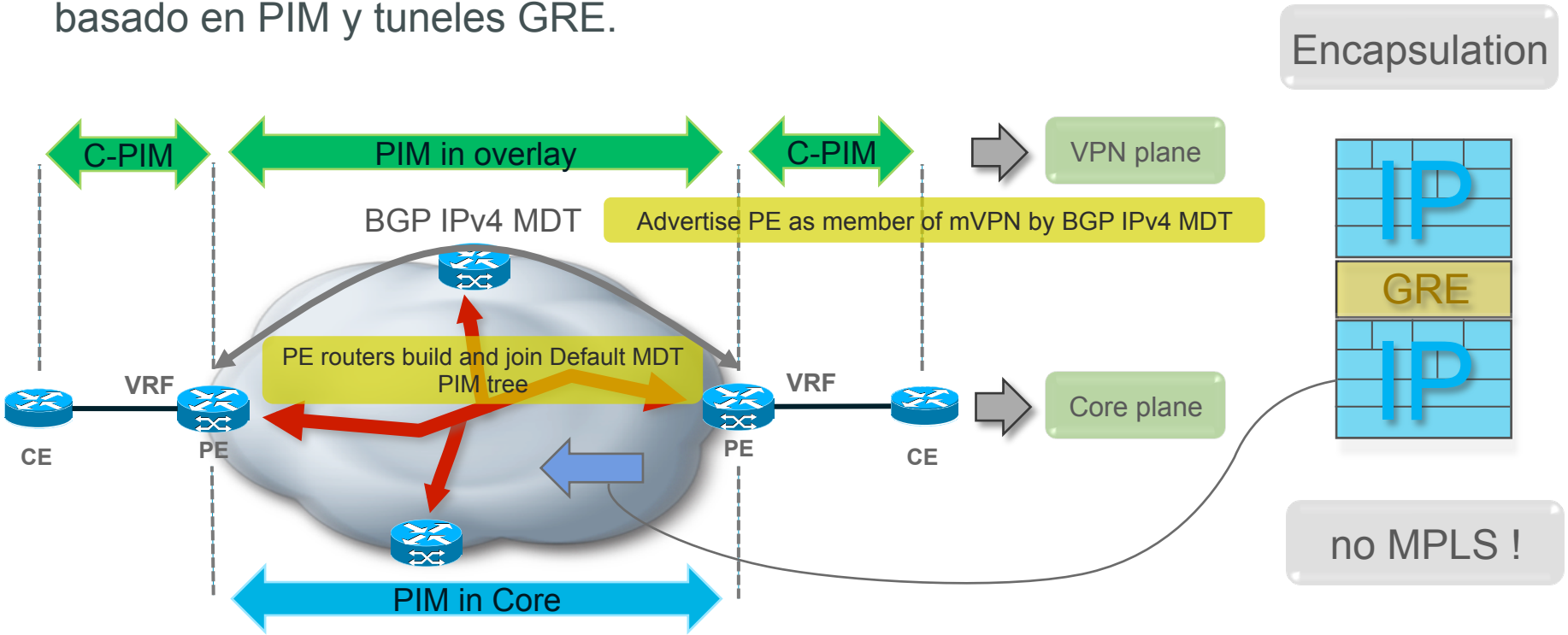
- P and PEs must support IP/MPLS (IGP+LDP)
- PEs support virtualization of IP/VPN domain into VRFs (Virtual Routing Forwarding instances)
- IP/VPN reachability exchanged among PEs via iBGP (MP-BGP)
  - VPNv4 (IPv4+RD) + Label used to exchange prefixes between PEs
- Import and Export information for IP prefixes at VRF level controlled by Route Target communities
- Extended to support IPv6 via VPNv6 address-family (RFC 4659)

MPLS VPN Forwarding Model (RFC 4354)

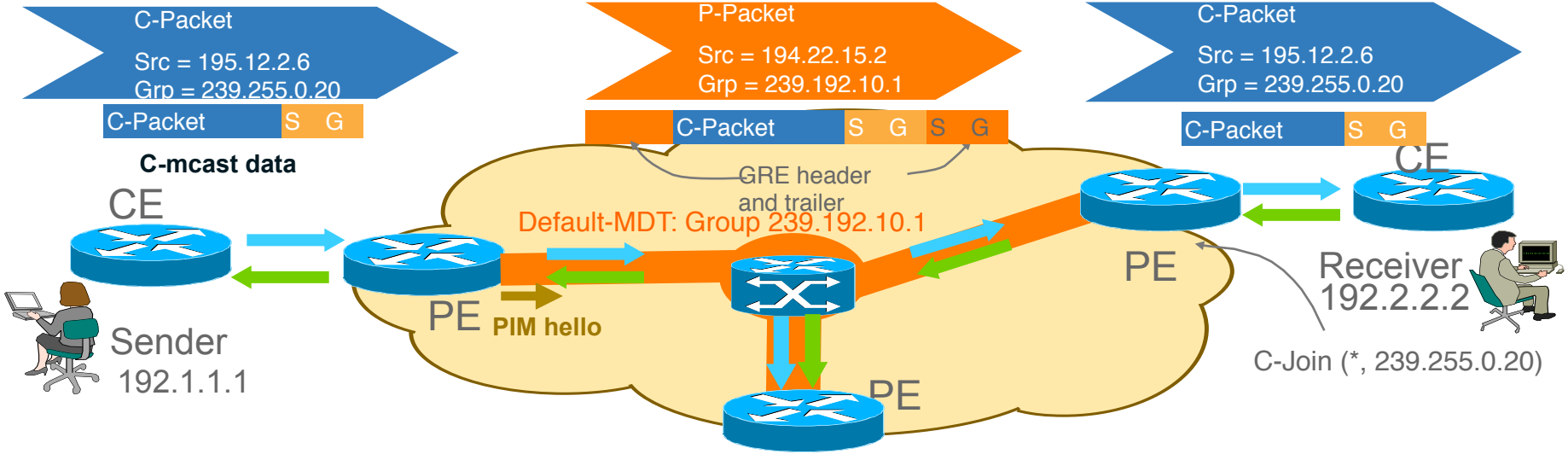


Multicast in BGP/MPLS IP VPNs
Connection Model (RFC 6037) aka draft-Rosen

- Implementando un plano de control y forwarding restringido solo para Multicast basado en PIM y tuneles GRE.

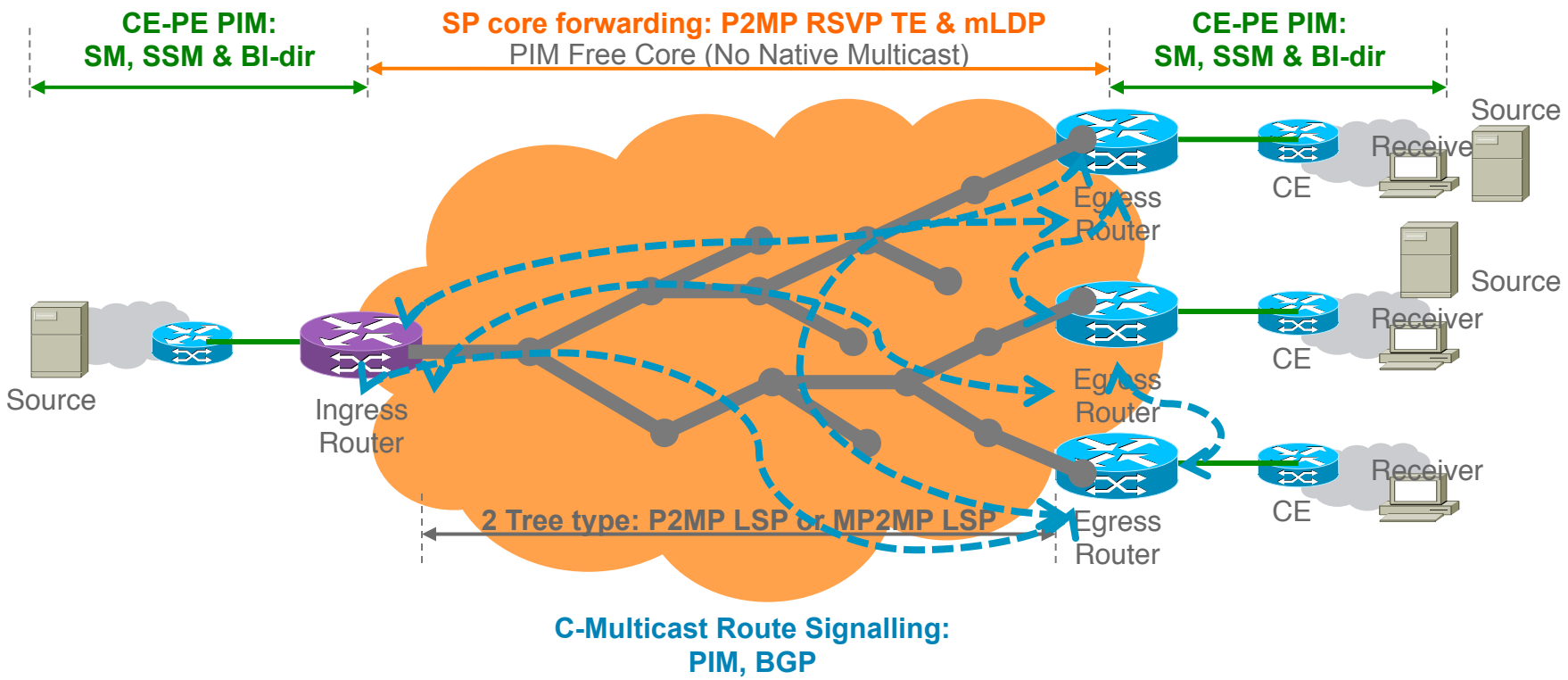


Multicast in BGP/MPLS IP VPNs
Connection Model (RFC 6037) aka draft-Rosen



- Default MDT connects all mVPN PE routers and carries all PIM signaling and all mcast traffic by default
  - Optimized Tree (MDT-Data) only for PE with receivers is possible for high data rates Sources
- Customer´s multicast packets will get encapsulated into GRE using P-multicast groups (Default or Data MDT)
  - MPLS labels are NOT used in core

LSM: Label Switch Multicast
The building block for NG-MVPN



Why Labeled Switch Multicast (LSM)?

Classic MVPN

- Only GRE encapsulation
- Only PIM in core
- Default MDT / Data MDT

LSM / NG MVPN

- Leverage MPLS encapsulation
  - Share data plane with unicast
- Leverage new core tree protocols
  - mLDP, P2MP TE, BIER
- Fast RestoRation (FRR)
- More flexible designs per VPN
- Manageability: no need to track Multicast Groups per VPN/Default/Data MDT

- Single Forwarding Plane (MPLS Labels) for both Unicast and Multicast at IP-VPN

LSM Protocols-at-a-Glance

P2MP RSVP-TE

- Deterministic static trees
  - Desired latency & bw
- Suitable for subset of multicast applications:
  - Video Distribution
- FRR inherent capability
- Head end static setup
- Hop by hop protocol

mLDP

- Dynamic tree building
  - Based on IGP info
- Suitable for broad set of multicast applications:
  - mVPN, Native Multicast
- LFA and FRR optional capability
- Receiver driven tree setup
- Hop by hop protocol

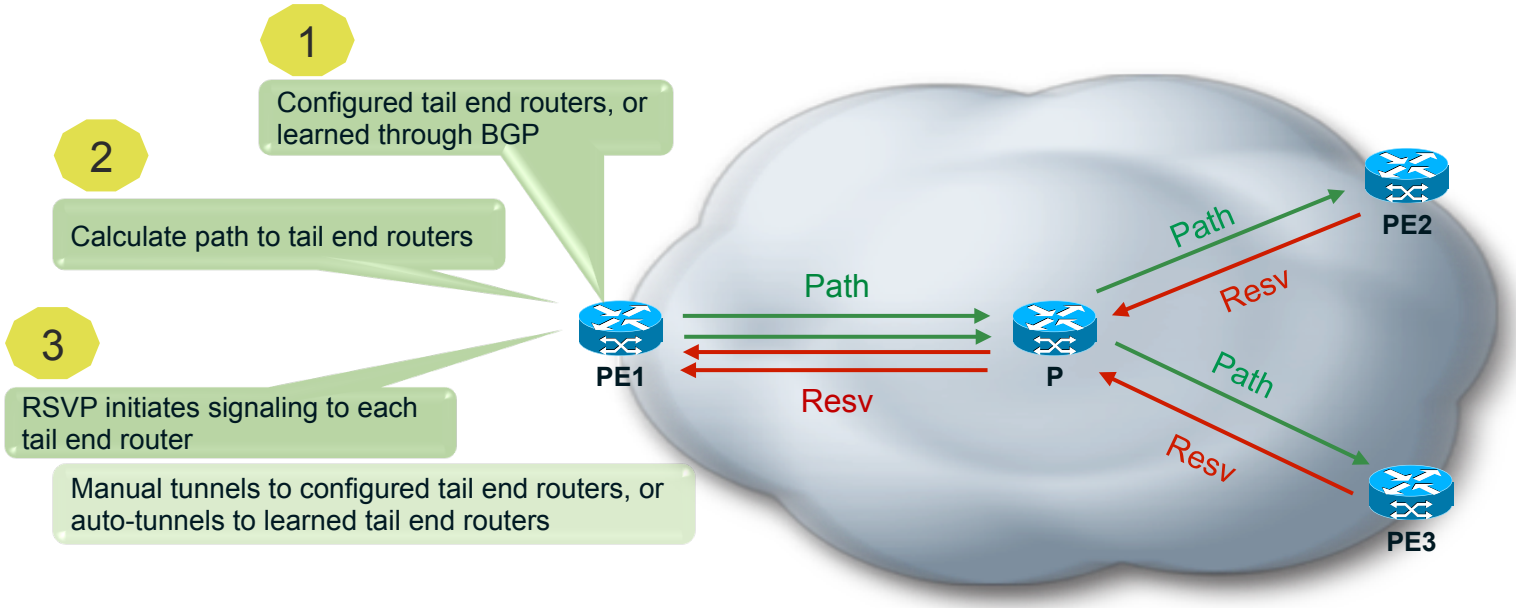
Both completely eliminate PIM from the core

Same forwarding plane for both unicast and multicast traffic

Multicast LSPs can be Fast ReRoute protected

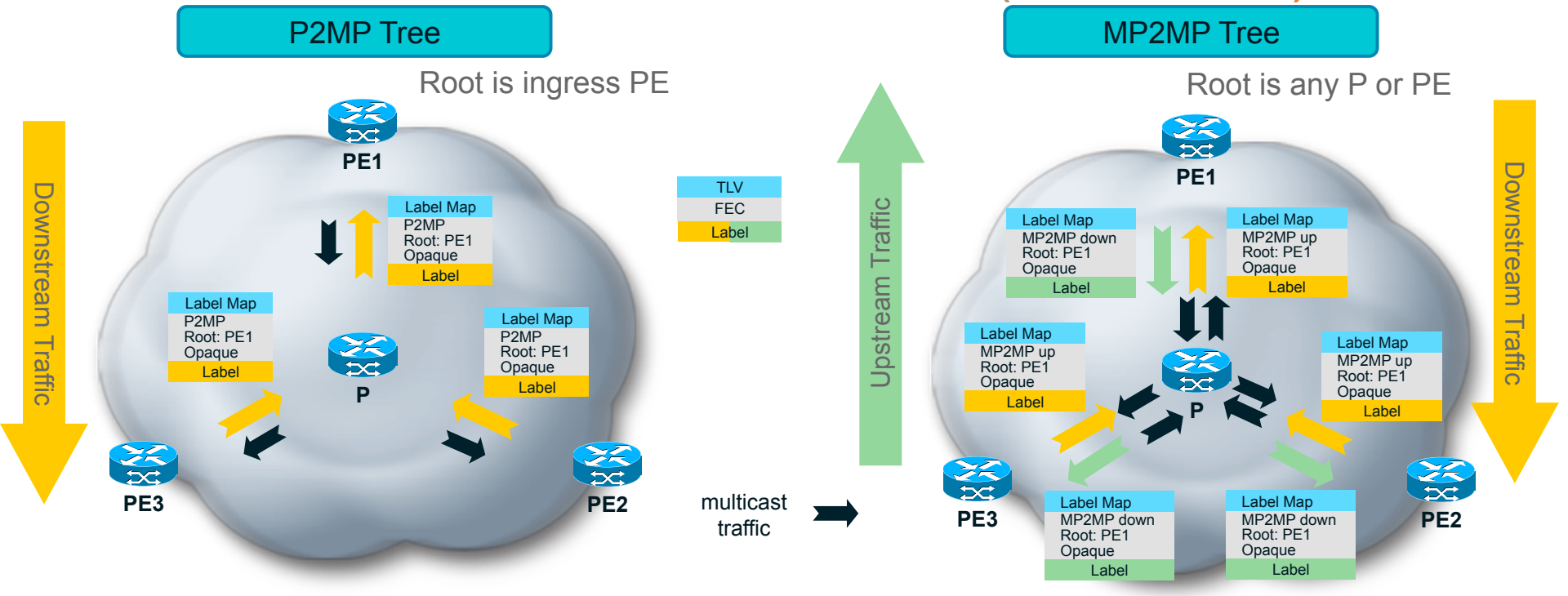
P2MP Traffic Engineering (TE) Extensions for RSVP-TE and IGP (RFC 4875)

- P2MP tunnel signaled by RSVP, to multiple tail end routers



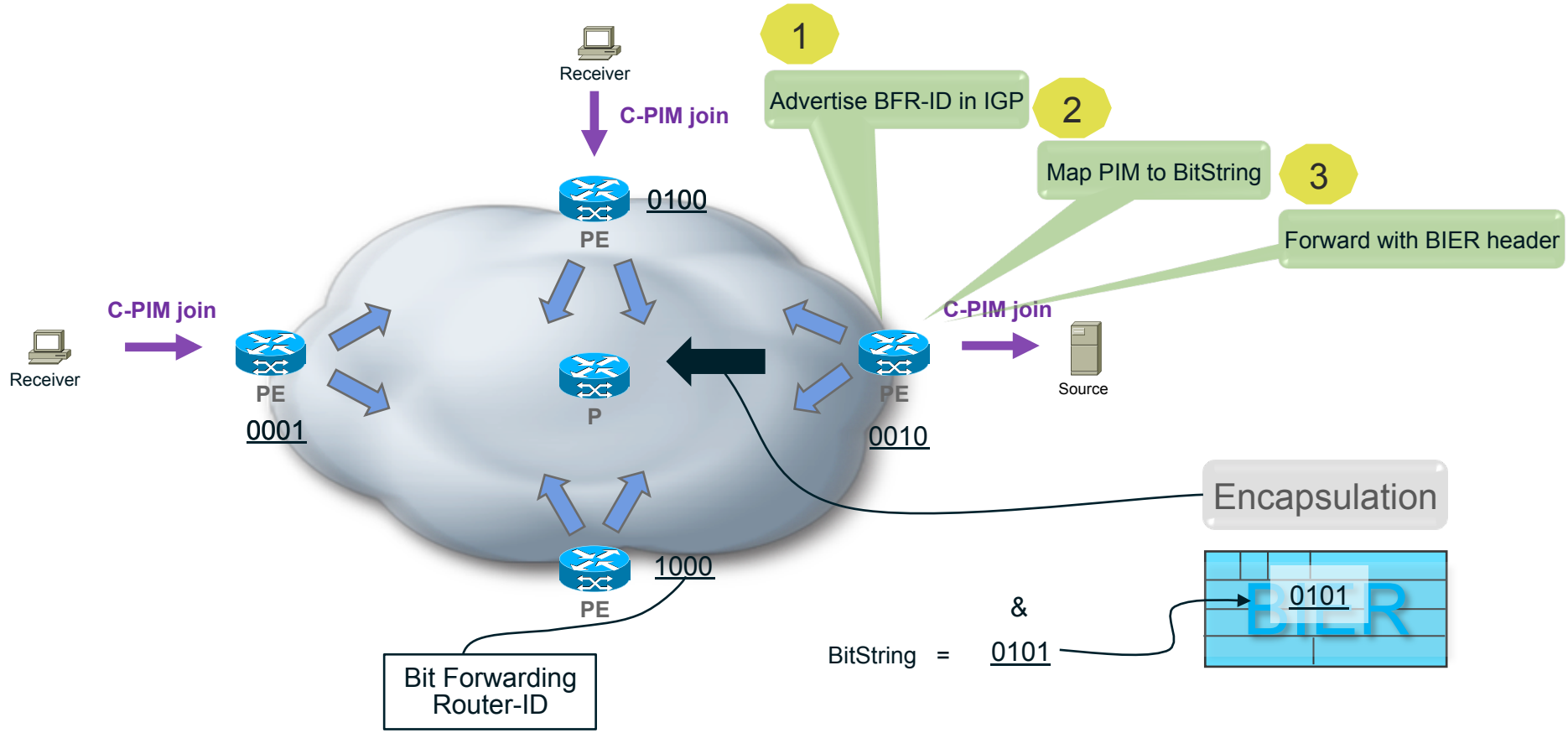
Explicit (source) routing, Bandwidth reservation, Fast ReRoute (FRR) protection

mLDP, Multicast LDP Extensions for P2MP and MP2MP (RFC 6388)



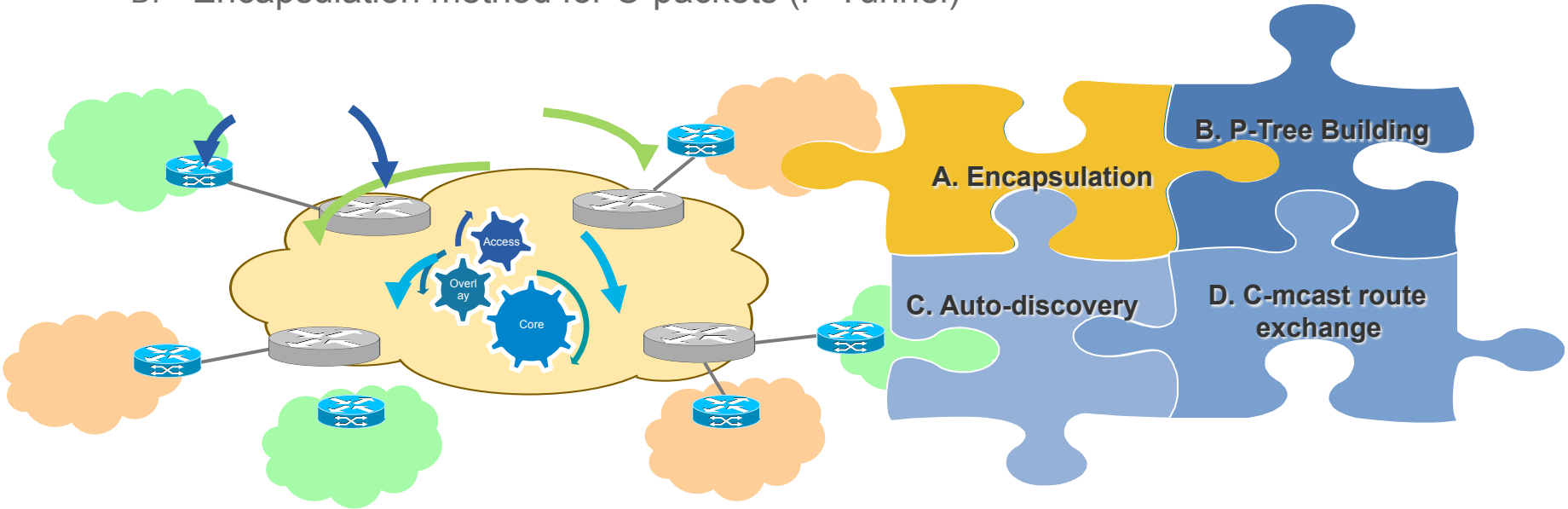
Receiver driven (signaled toward Root), holds FEC elements (Type of tree, Root, Opaque value likes (S,G), MDT number, LSP ID)

Bit Indexed Explicit Replication (BIER)



MVPN Components

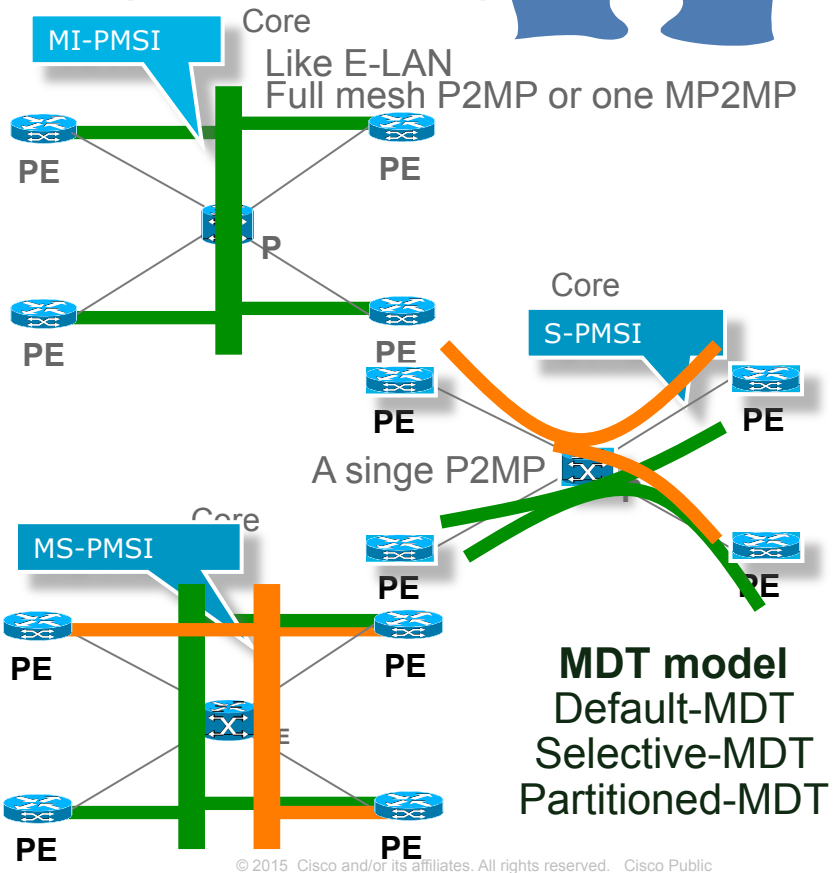
- Decompose into major components
  - A. Tree-building method in the P-core
  - B. Autodiscovery of the PE routers involved in a given MVPN
  - C. Distribution of C-mroutes among the PEs
  - D. Encapsulation method for C-packets (P-Tunnel)



Demystifying MVPNs - the PMSI (RFC 6513)

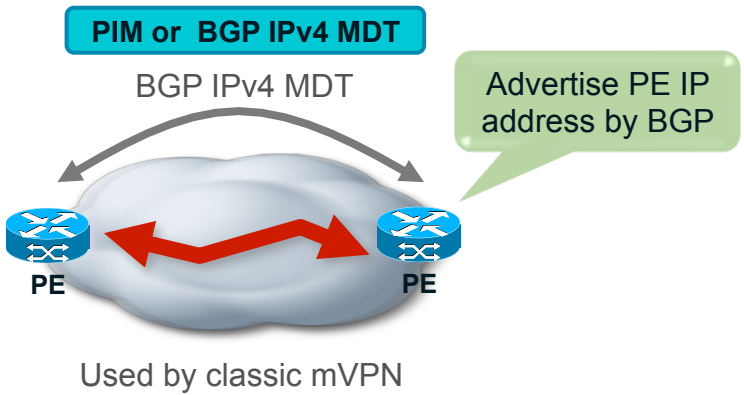
- PMSI (Provider Multicast Service Interface) is an abstract multicast service, which serves a single MVPN to carry C-multicast data traffic through P-tunnels
- Flavors of PMSI:
  - Multidirectional Inclusive (**MI-PMSI**): from all PEs to all PEs of MVPN (default-MDT)
  - Unidirectional Inclusive (**UI-PMSI**): from one PE to all PEs of MVPN
  - Selective (**S-PMSI**): from one PE to select subset of PEs in MVPN (Data-MDT)
  - MS-PMSI, aka Partitioned-MDT, a single MP2MP
- PMSIs from several MVPNs may share a single tunnel

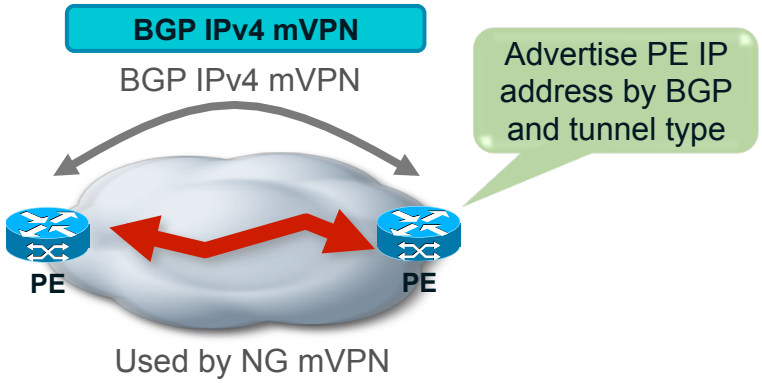
A PMSI is a conceptual "overlay" on P-network.

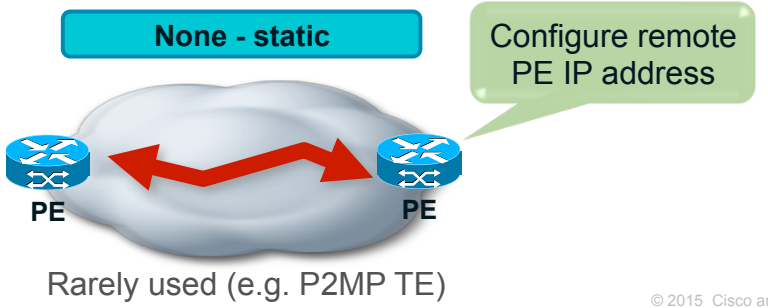


MVPN Components: Auto-Discovery

- Auto Discovery (AD)
  - The process of discovering all the PEs with members in a given mVPN
  - In order to build the MDT

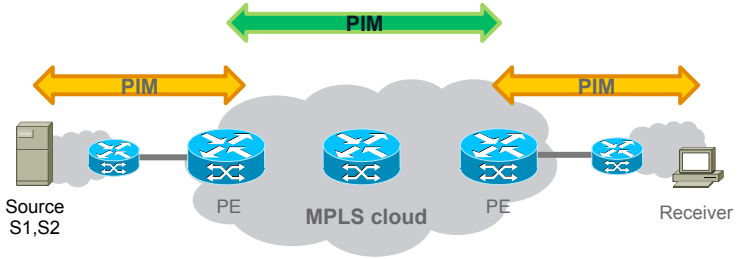




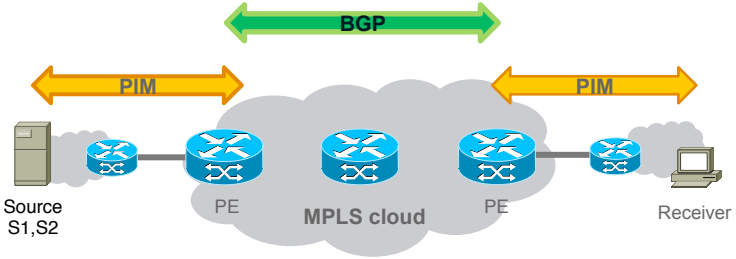


MVPN Componets: Overlay Signaling Possibilities

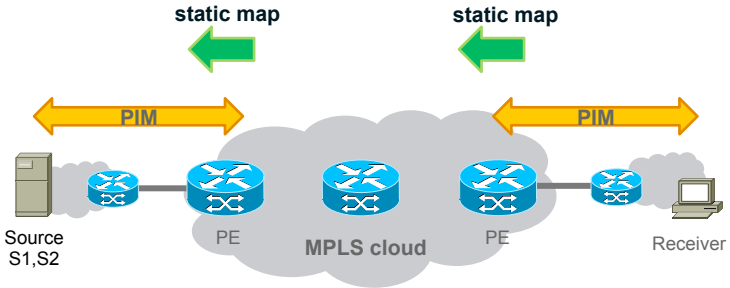
D. C-mcast route exchange



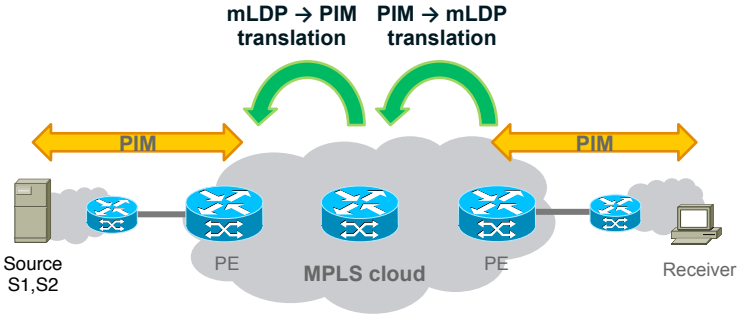
PIM in Overlay



BGP in Overlay



Static

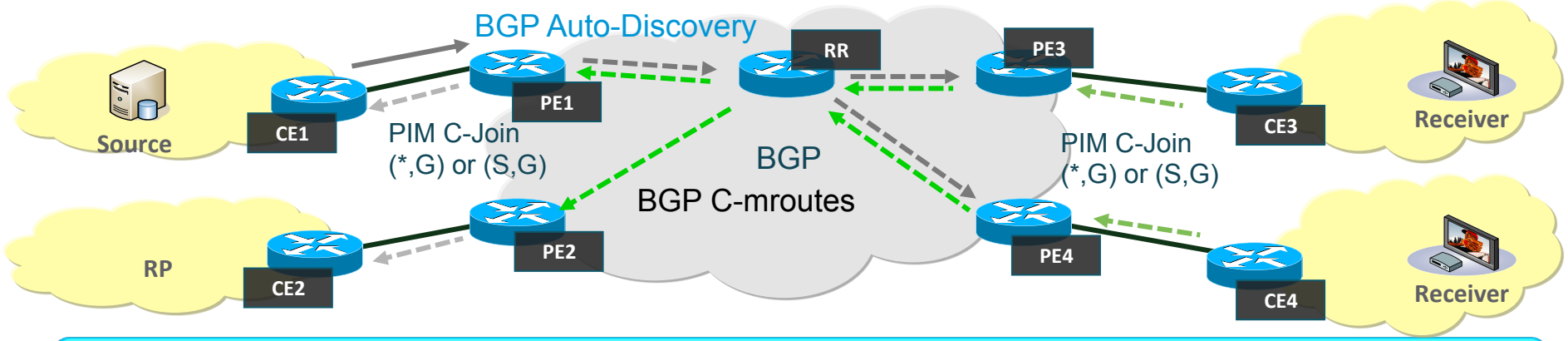


In-band

Multicast VPN – BGP Signaling

Route Type (1 octet)
Length (1 octet)
Route Type specific (variable)

New address family MCAST-VPN NLRI used for
- MVPN auto-discovery
- Advertising Tunnel Binding (MVPN to I-PMSI or (C-S,C-G) to S-PMSI)
- Signal multicast information:
  - (*,G) or (S,G)
  - Which tunnel to use (core tree protocol and tunnel type)



BGP as overlay allows Service Providers to capitalize on a single protocol and common control plane for both Ucast and Mcast at MPLS IP-VPN

MCAST VPN BGP UPDATE message (RFC 6514)

Withdrawn routes length	Withdrawn routes length
Withdrawn routes	Length in octets of the Route Type specific field of MCAST-VPN NLRI
Total path attribute length	Total path attribute length
Path attributes	Attribute x
	PMSI Tunnel Attribute (RFC6514 section 5)
	MP-REACH/UNREACH-NLRI attribute
	PPMP MPLS label attribute (3 octets)
NLRI	Network Layer Reachability Information (variable length)

Flags (1 octet)	0-6 – Reserved 7 – L – Leaf Information required
Tunnel type (1 octet)	0 - No tunnel information present (used for explicit tracking purpose with L bit set) 1 - RSVP-TE P2MP LSP (defined as LSP SESSION Object in RFC-4875) 2 - mLDP P2MP LSP (defined in RFC-6388) 3 - PIM-SSM Tree 4 - PIM-SM Tree 5 - PIM-Bidir Tree 6 - Ingress Replication 7 - mLDP MP2MP LSP (defined in RFC-6388)
MPLS label (3 octets)	MPLS label (label 20 bits, exp 3 bits, BSB 1 bit)
Tunnel Identifier (variable)	1. RSVP-TE P2MP LSP - <P2MP ID, Reserved Tunnel ID, Ext Tunnel ID> (RFC4675 SESSION object) 2 - mLDP P2MP LSP - P2MP FEC Element [mLDP] 3- PIM-SSM Tree - <P- Root Node Address, P-Multicast Group> 4 – PIM-SM Tree - <Sender Address, P-Multicast Group> 5 - PIM-Bidir Tree - <Sender Address, P-Multicast Group> 6 – Ingress replication - unicast tunnel endpoint IP address of the local PE that is to be this PE's receiving endpoint address for the tunnel 7 - mLDP MP2MP LSP - MP2MP FEC Element [mLDP]

AFI: 1 for v4, 2 for ipv6
SAFI: MVPN
Length of network address of next hop (1 octet)
Address of next hop (variable length)
MCAST VPN NLRIs (RFC6514 section 3)

Route type (1 octet)	1 - Intra-AS I-PMSI A-D route; - 4.1.1 support 2 - Inter-AS I-PMSI A-D route; 3 - S-PMSI A-D route; - 4.1.1 support 4 - Leaf A-D route; 5 - Source Active A-D
Length (1 octet)	Length in octets of the Route Type specific field of MCAST-VPN NLRI
Route type specific (variable length)	**1 - Intra-AS I-PMSI A-D :** RD (8 octets) + originating router's IP address **2 – Inter-AS I-PMSI :** RD (8 octets) + source AS (4 octets) **3 - S-PMSI A-D :** RD (8 octets) + MCAST source length (1 octet) + MCAST source (variable) + MCAST group length (1 octet) + MCAST group (variable) + originating router's IP address **4 – Leaf AD :** (Route Key (variable) + originating router's IP address **5 – Source active :** RD (8 octets) + MCAST source length (1 octet) + MCAST source (variable) + MCAST group length (1 octet) + MCAST group (variable) **6/7 – Shared/Source tree join :** RD (8 octets) + MCAST source length (1 octet) + MCAST source (variable) + MCAST group length (1 octet) + MCAST group (variable)

BGP Address Family MCAST-VPN (RFC 6514)

Signal multicast information

(*,G) (S,G)

mcast vpn NLRI

Tunnel information

PMSI Tunnel Attribute (PTA)

Route Type	Meaning	Usage
1	Intra-AS I-PMSI A-D route	AD Signaling
2	Inter-AS I-PMSI A-D route	AD Signaling
3	S-PMSI A-D route	AD Signaling
4	Leaf A-D route	AD Signaling
5	Source Active A-D route	AD Signaling
6	Shared Tree Join route	C-signaling
7	Source Tree Join route	C-signaling

Encoding can be : RD (8 octets) , MCAST source length (1 octet), MCAST source (variable) , MCAST group length (1 octet), MCAST group (variable), Originating router's IP address

Tunnel Type	Meaning	Info encoded
0	No tunnel info present	-
1	P2MP TE tunnel	Ext tunnel ID / Tunnel ID / P2MP ID
2	mLDP P2MP	P2MP FEC Element
3	PIM SSM	Root address / P-Group
4	PIM Sparse Mode	Sender Address / P-Group
5	PIM BiDirectional	Sender Address / P-Group
6	Ingress Replication	Unicast tunnel endpoint address
7	mLDP MP2MP	MP2MP FEC Element
8	Transport Tunnel	Source PE address / local number

BGP MCAST-VPN address family

- Specified in RFC-6514/RFC-6513 using BGP Multiprotocol Extensions [RFC-4760] with an AFI of 1 or 2 and an SAFI of MCAST-VPN.
- Used for advertisement of the following AD routes:

Type	Name	Use
1	Intra I-PMSI	Advertise default lmdt from encap PE
2	Inter-AS I-PMSI A-D route	Segmented Inter-AS
3	S-PMSI	Advertise data mdt from encap PE
4	Leaf A-D route	Generated by receivers PE as a response to type 3 route with leaf info required flag set
5	Source-Active	Originated on a encap PE with active source
6	Shared-tree join route	To advertise *,G join from decap PE
7	Source-tree join route	To advertise S,G join from decap PE

MCAST VPN BGP message

NLRI

1) **MCAST-VPN NLRI**

ROUTE TYPE [1 BYTE]
LENGTH [1 BYTE]
ROUTE TYPE SPECIFIC [VAR LENGTH]

1) INTRA-AS I-PMSI A-D ROUTE
2) INTER-AS I-PMSI A-D ROUTE
3) S-PMSI A-D ROUTE
4) LEAF A-D ROUTE
5) SOURCE ACTIVE A-D ROUTE
6) SHARED TREE JOIN ROUTE
7) SOURCE TREE JOIN ROUTE

ORIGINATED BY ALL MVPN PES

RD [8 BYTES]
ORIGINATING PE's IP ADD [4 BYTES]

ORIGINATED BY MVPN ASBRs

RD [8 BYTES]
SOURCE AS [4 BYTES]

ORIGINATED BY ROOT PEs

RD [8 BYTES]
MCAST C-SRC LEN [1 BYTE]
MCAST C-SRC ADD [VAR]
MCAST C-GRP LEN [1 BYTE]
MCAST C-GRP ADD [VAR]
ORIGINATING PE's IP ADD [4 BYTES]

ORIGINATED BY TAIL PEs

ROUTE KEY[VAR]
ORIGINATING PE's IP ADD [4 BYTES]

ORIGINATED BY ACTIVE SOURCE PEs

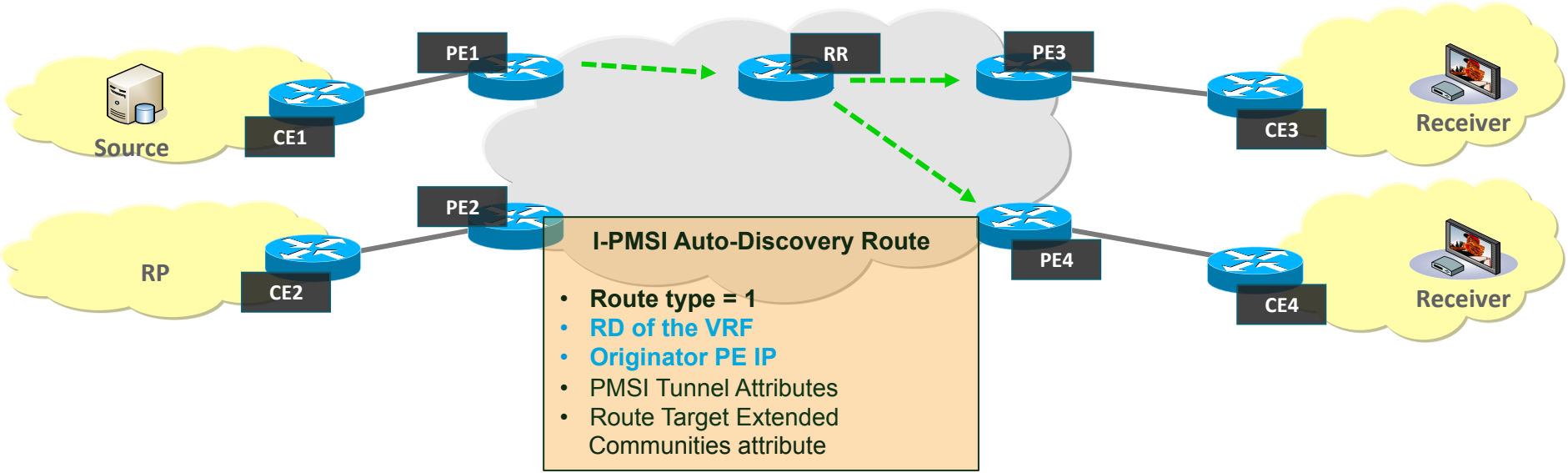
RD [8 BYTES]
MCAST C-SRC LEN [1 BYTE]
MCAST C-SRC ADD [VAR]
MCAST C-GRP LEN [1 BYTE]
MCAST C-GRP ADD [VAR]

ORIGINATED BY RECEIVER PEs

ORIGINATED BY RECEIVER PEs

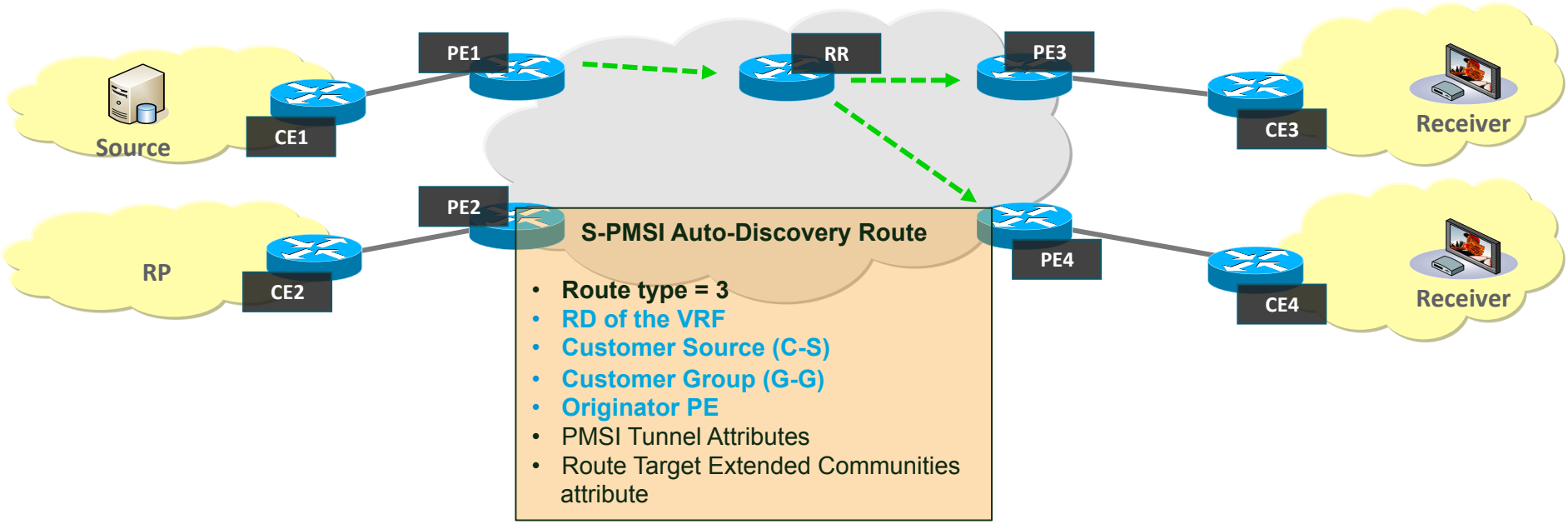
RD [8 BYTES]
SOURCE AS [4 BYTES]
MCAST C-SRC LEN [1 BYTE]
MCAST C-SRC ADD [VAR]
MCAST C-GRP LEN [1 BYTE]
MCAST C-GRP ADD [VAR]

BGP Address Family MCAST-VPN (RFC 6514) Intra-AS I-PMSI Auto-Discovery



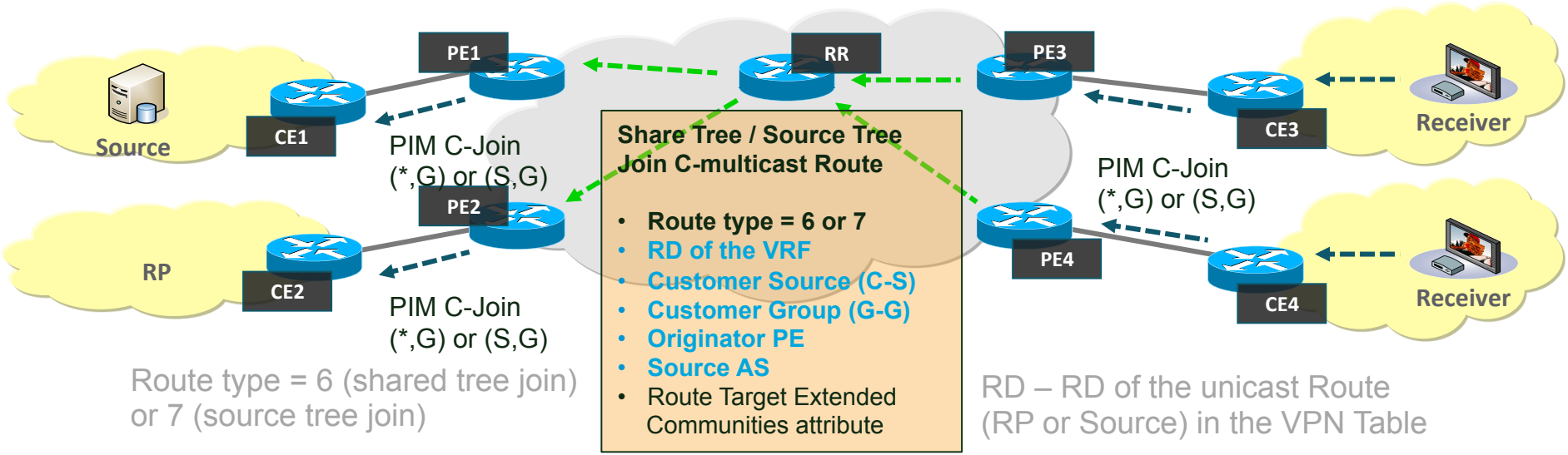
- This route is advertised by a PE router to indicate its membership in a given MVPN.
- This route is used for auto-discovery of all PEs within an AS, and for auto-discovery of all PEs across all ASes when using non-segmented P-Tunnels.
- Replacement for Draft Rosen MDT-SAFI (SAFI 66)

BGP Address Family MCAST-VPN (RFC 6514) Intra-AS S-PMSI Auto-Discovery



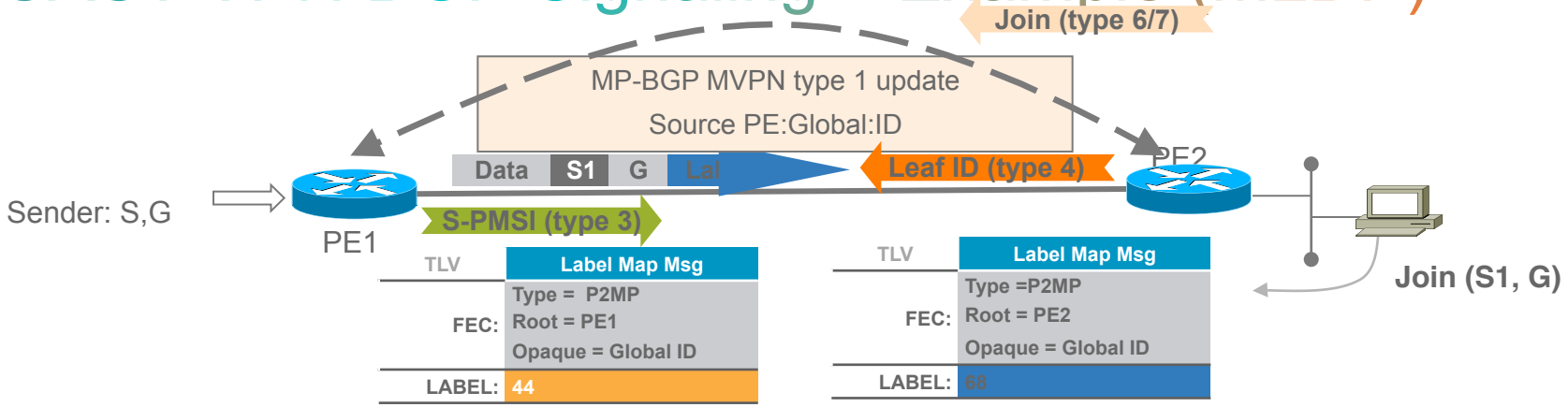
- This route is advertised by a PE router when it wants to bind a (C-S,C-G)-flow to a P-Tunnel.
- Used for moving C-flows from I-PMSIs to S-PMSIs

BGP Address Family MCAST-VPN (RFC 6514) Customer Multicast Routes



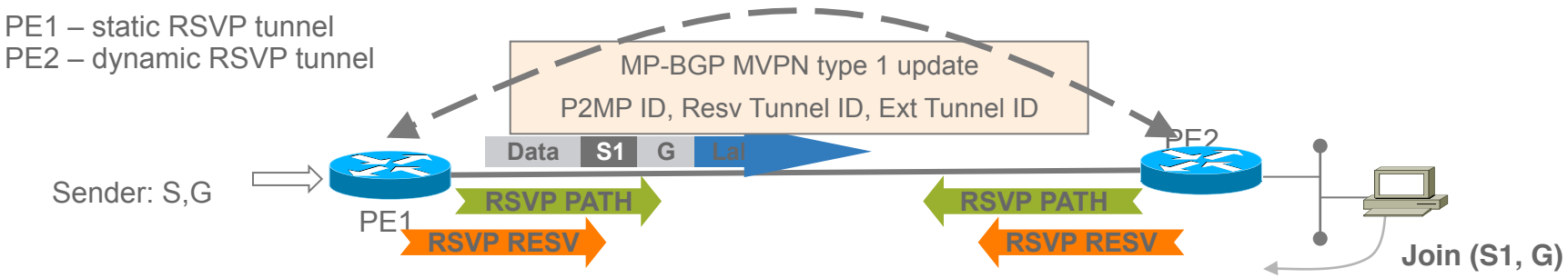
- C-multicast routes are originated as a result of updates in (C-S, C-G), or (C-*, C-G) state learnt by a PE via the C-multicast protocol.
- A C-multicast shared tree join route is advertised by a PE router when it wants to propagate a C-Join for a (C-*,C-G) flow upstream.
- Similarly, a C-multicast source tree join route is advertised when a C-Join for a (C-S, C-G) flow needs to be propagated
- On receiving the C-multicast join route, the receiving PE adds S-PMSI to oif-list if one exists for (C-S, C-G) and not already added, else adds I-PMSI to oif-list if not already added.

MCAST VPN BGP Signaling – Example (mLDP)



1. PE1 and PE2 advertises BGP mvpn I-PMSI with P2MP PMSI Type 1 – Source-PE:Global-ID
2. All routers join advertised trees using MLDP opaque type advertised in I-PMSI routes. After process is complete we have full mesh of P2MP tunnels acting internally as LAN.
3. After customer join is received type6/7 route is generated by the egress PE2.
4. If the first join is received by the PE1 it starts sending traffic over default-mdt
5. If S-PMSI is configured PE1 advertises type 3 routes
6. PE2 joins the S-PMSI tree based on the information from the BGP route (if leaf info required bit is set in the type 3 route PE2 sends type 4 route towards PE1 as well)
7. After receiving MLDP FEC PE1 starts forwarding traffic using S-PMSI tree (no type 4 route is required to start traffic forwarding)

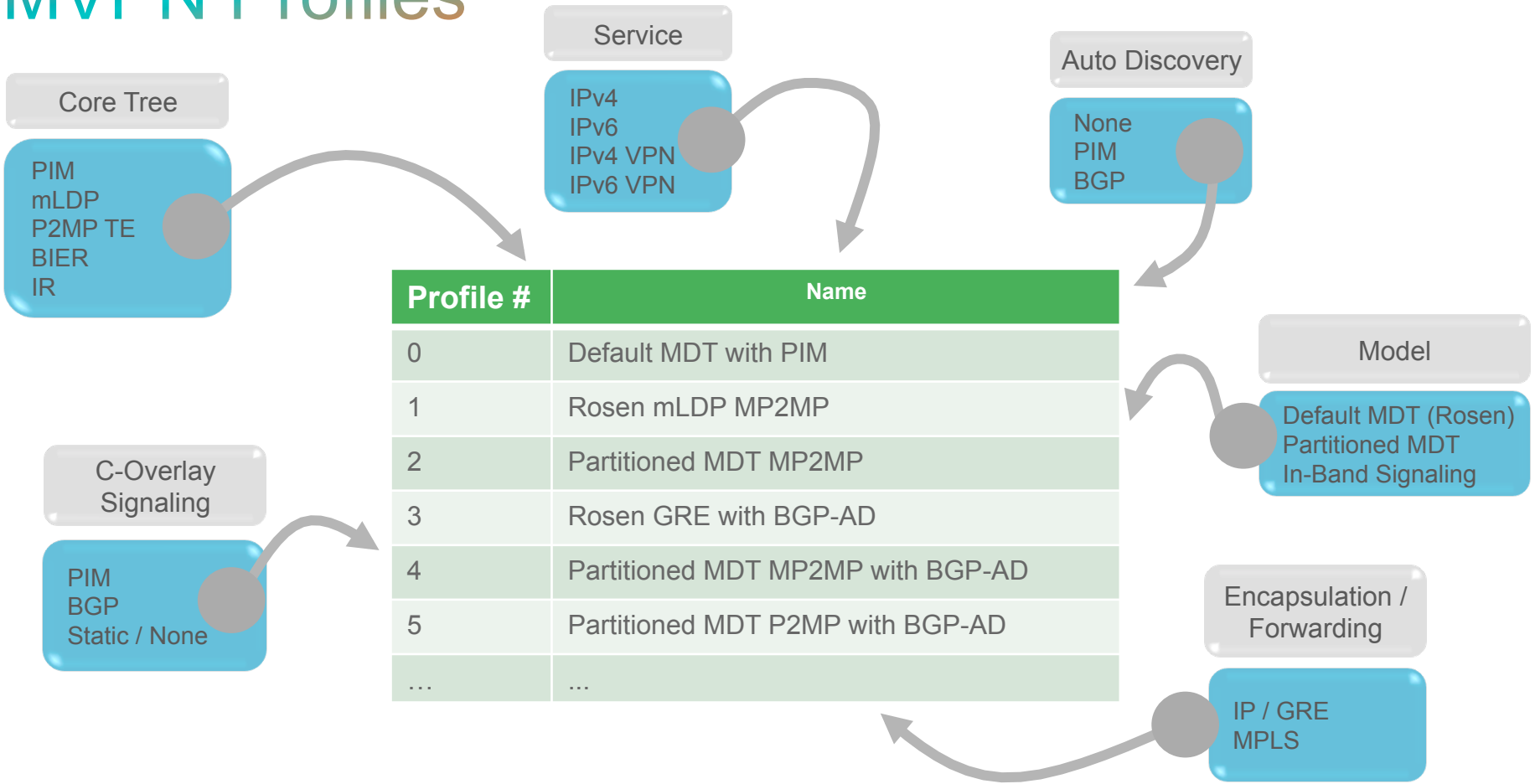
MCAST VPN BGP Signaling – Example (RSVP-TE)



1. PE1 and PE2 advertises BGP mvpn I-PMSI with RSVP-TE PMSI tunnel attribute (contains PE1/PE2 address coded as Ext Tunnel ID).
2. Static tunnel – PE1 sets up RSVP-TE tunnel sending RSVP path option message to PE2 based on statically configured tunnel-p2mp interface

   Dynamic tunnel – after receiving I-PMSI PE2 sends RSVP path option towards PE1 based on Ext Tunnel ID info (PE1 address)
3. PE1/PE2 after receiving RSVP PATH messages responds with the corresponding RESV message (upstream label allocation). Unidirectional RSVP-TE tunnels are created between PE1 and PE2.
4. PE2 receives join and forwards it towards PE1 via PIM using default tree (P18) or BGP mvpn type 6/7 route (P16).
5. After receiving join PE1 starts forwarding traffic over default tree (RSVP-TE tunnel)

MVPN Profiles



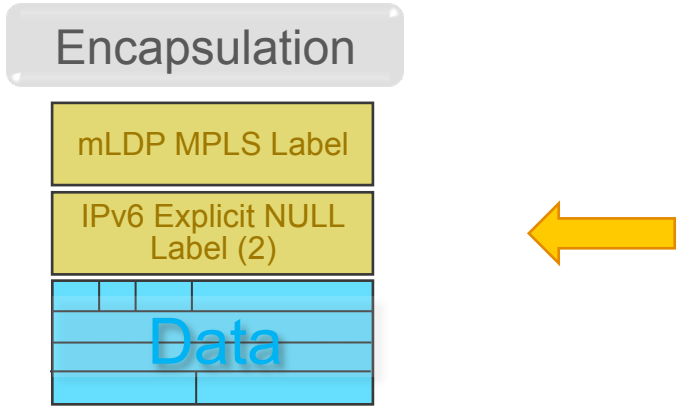
Profile #	Name
0	Default MDT with PIM
1	Rosen mLDP MP2MP
2	Partitioned MDT MP2MP
3	Rosen GRE with BGP-AD
4	Partitioned MDT MP2MP with BGP-AD
5	Partitioned MDT P2MP with BGP-AD
…	...

IPv6 MCAST VPN (RFC 6516)

- No core tree support for IPv6 (PIM, mLDP, P2MP TE, BIER)
- IPv4 core tree re-used for IPv6
  - PIM, mLDP, P2MP-TE, BIER
- Overlay signaling supports IPv6
  - PIMv6
  - BGP (IPv6 mvpn)

```
router bgp 1
…
 address-family ipv4 mvpn
  neighbor 10.100.1.4 activate
  neighbor 10.100.1.4 send-community both
!
 address-family ipv6 mvpn
  neighbor 10.100.1.4 activate
  neighbor 10.100.1.4 send-community both
```

- Note: Encapsulation of IPv6 over mLDP : explicit null label at the bottom to differentiate between IPv4 & IPv6 mcast on the same MDT



Next-Gen MVPN Summary

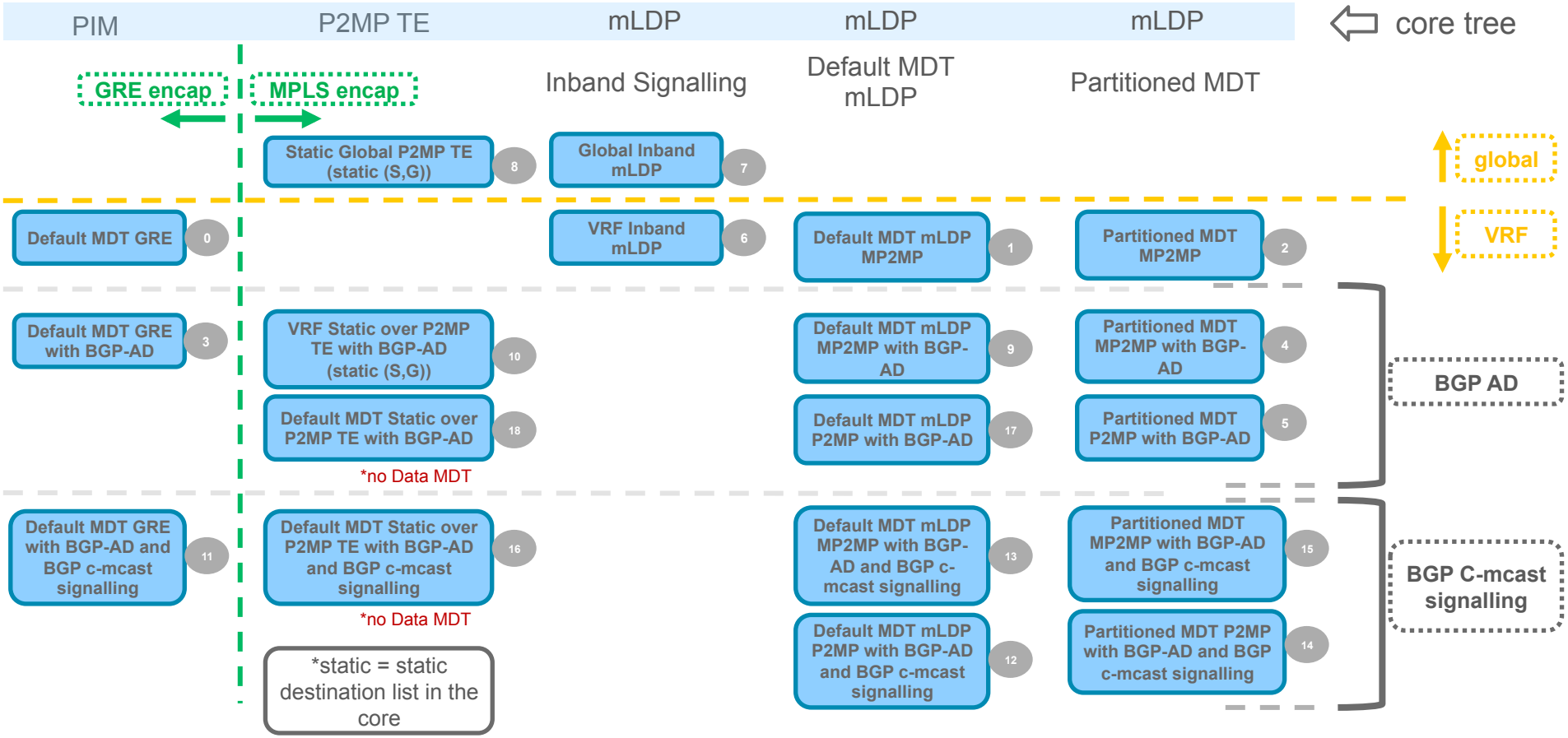
- Next Gen MVPN brings BGP and LSM together
  - Common forwading paradigm for both ucast and mcast L3-VPN (MPLS Label)
  - Common BGP-based control plane for both ucast and mcast L3-VPN
- Multiple solutions supported based on the core protocols:
  - MPLS forwarding in the core based on MLDP protocol (LDP extensions)
  - MPLS forwarding in the core based on the RSVP-TE tunnels
  - Ingress replication using unicast MPLS forwarding
- Multicast is label switched edge to edge
  - PIM on edge of core network
- Customer signalling solutions:
  - PIM based – customer joins are forwarded over the pre-established tunnels in the core using PIM protocol
  - BGP based – customer joins send using BGP mvpn address family
- Multiple profiles as combination of Core protocol, MDT Tree, Signaling, Discovery
- Compatible with previous draft-rosen PIM-GRE Model (MDT-Default with PIM Profile)

References for Next Gen MVPN

- RFC 4364, "BGP/MPLS IP Virtual Private Networks (VPNs)", Rosen, E. and Y. Rekhter, February 2006.
- RFC 4659, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", De Clercq, J., Ooms, D., Carugi, M., and F. Le Faucheur, September 2006.
- RFC 4760, "Multiprotocol Extensions for BGP-4", Bates, T., Chandra, R., Katz, D., and Y. Rekhter, January 2007.
- RFC 6037, "Cisco Systems' Solution for Multicast in BGP/MPLS IP VPNs", E. Rosen, Y. Cai, I. Wijnands, October 2010.
- RFC 4875, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., May 2007.
- RFC 6388, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B. Thomas, November 2011.
- RFC 6513, "Multicast in MPLS/BGP IP VPNs", Rosen, E., Ed., and R. Aggarwal, Ed., February 2012.
- RFC 6514, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, February 2012.
- RFC 6515, "IPv4 and IPv6 Infrastructure Addresses in BGP Updates for Multicast VPN", R. Aggarwal and E. Rosen, February 2012.
- RFC 6516, "IPv6 Multicast VPN (MVPN) Support Using PIM Control Plane and Selective Provider Multicast Service Interface (S-PMSI) Join Messages", Y. Cai, E. Rosen, IJ. Wijnands, February 2012.

Thank you

MVPN Profiles



MVPN Profiles (Cont)