RIPE65 – Amsterdam, NL
September 24, 2012

JUNIPER NETWORKS

SCALING MPLS – SEAMLESSLY

RESILIENT SERVICE ENABLEMENT AT MASSIVE SCALE USING STANDARD PROTOCOLS

Christian Martin
Sr. Director, Network Architecture
Office of the CTO – Platform Systems Division, Juniper Networks

ACKNOWLEDGEMENTS

Many thanks to **Maciek Konstantynowicz**, **Kireeti Kompella**, **Yakov Rekhter**, **Nitin Bahadur** and many others from Juniper for their contribution to the developments of technologies described in this presentation.

AGENDA

Network design evolution

“Seamless” MPLS

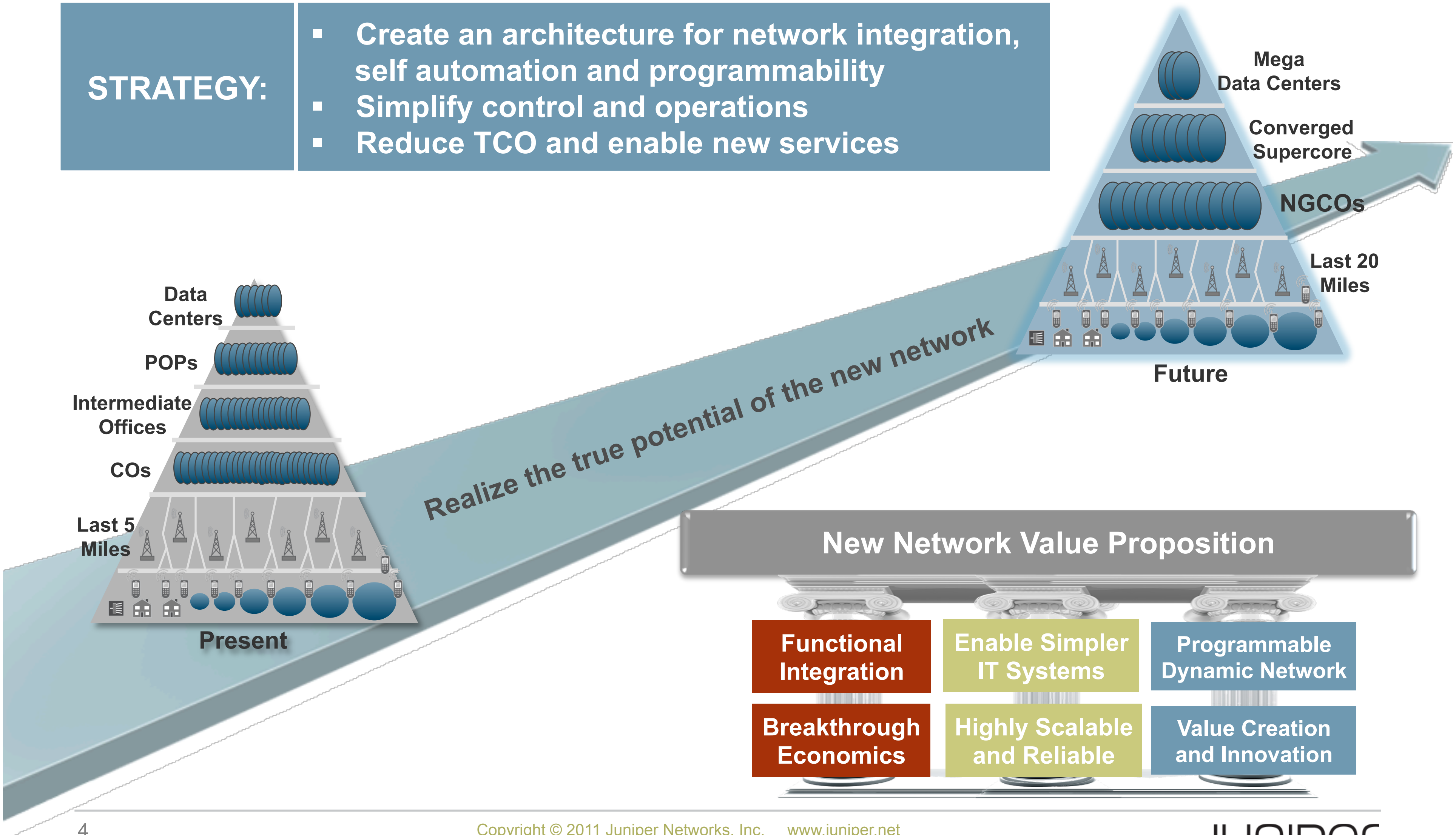
- Architecture
- Design use cases
- MPLS in the access

Universal Edge with MPLS access

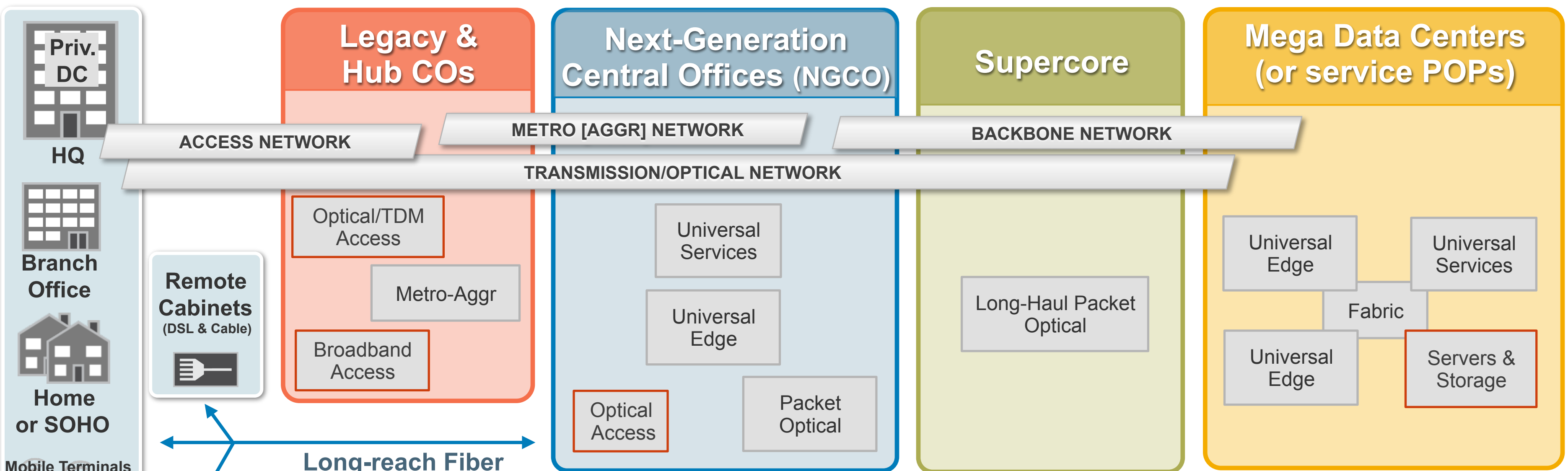
NEW NETWORK GOALS

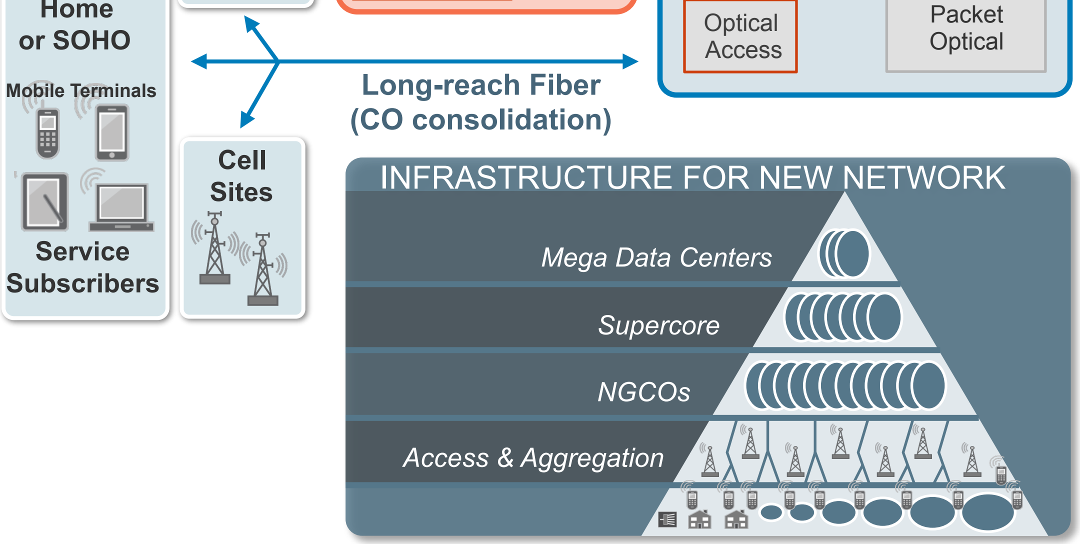
STRATEGY:

- Create an architecture for network integration, self automation and programmability
- Simplify control and operations
- Reduce TCO and enable new services



NEW NETWORK TOPOLOGY



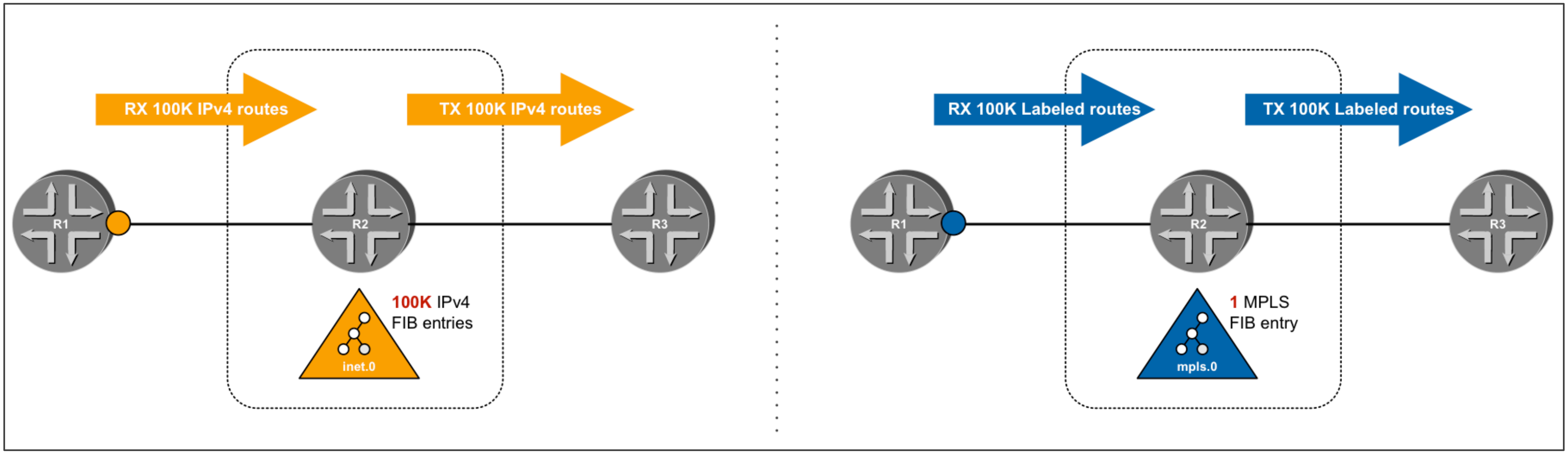


The All-IP NGN new network vision:

- Eliminate silos, consolidate and streamline the access & metropolitan part of the SP networks
- Optimize service delivery (network, content, applications)
- Simplify network and service control and operation, enable streamlined IT Systems
- Service innovation with software programmable network, leverage self-organizing network
- Further integrate packet and optical network layers

SEAMLESS MPLS - ARCHITECTURE

FIRSTLY - WHY IS MPLS USEFUL ?



Control plane and data plane separation

Unified data plane

- ***Universal platform*** for Services

Support for arbitrary hierarchy

- Stack of MPLS labels
- Used for ***Services***, ***Scaling*** and fast service ***Restoration***

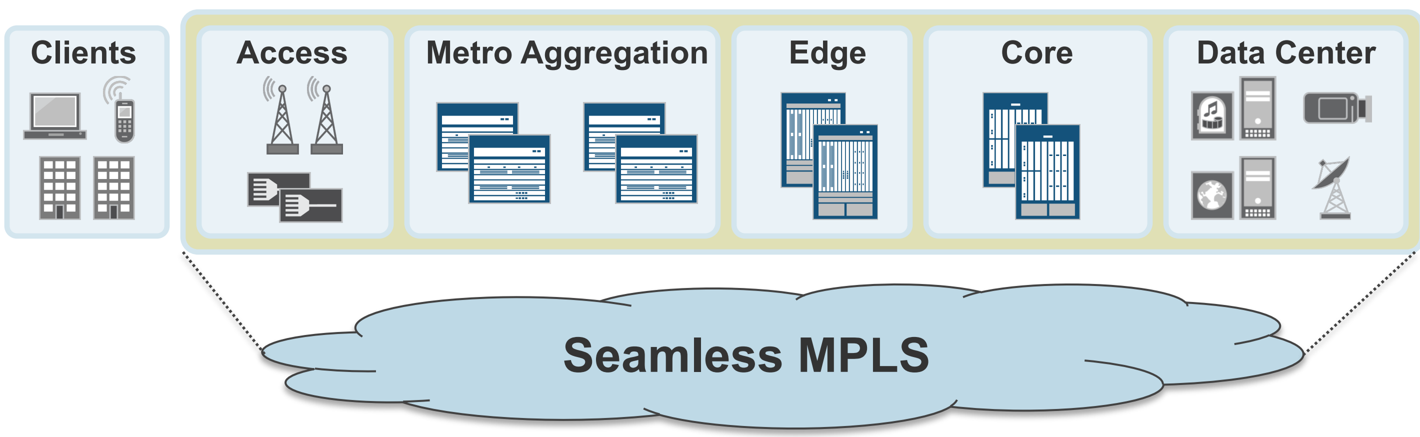
IMPLEMENTATION: SEAMLESS MPLS FOUNDATION FOR THE CONVERGED NETWORK

Network Scale and End-to-End service restoration

- MPLS in the access, 100,000s of devices in ONE packet network
- Seamless service recovery from any failure event (Sub-50ms)

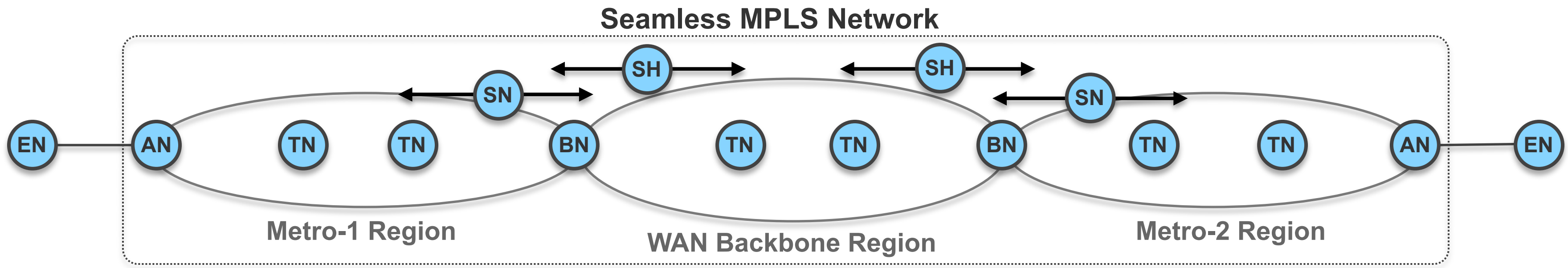
Decoupled network and service architectures

- Complete virtualization of network services
- Flexible topological placement of services – enabler for per service de-centralization
- Minimized number of provisioning points, simplified end-to-end operation



Networking at scale without boundaries

SEAMLESS MPLS FUNCTIONAL BLUEPRINT



Devices and their roles

- Access Nodes – terminate local loop from subscribers (e.g. DSLAM, MSAN)
- Transport Nodes – packet transport within the region (e.g. Metro LSR, Core LSR)
- Border Nodes – enable inter-region packet transport (e.g. ABR, ASBR)
- Service Nodes – service delivery points, with flexible topological placement (e.g.BNG, IPVPN PE)
- Service Helpers – service enablement or control plane scale points (e.g. Radius, BGP RR)
- End Nodes – represent customer network, located outside of service provider network

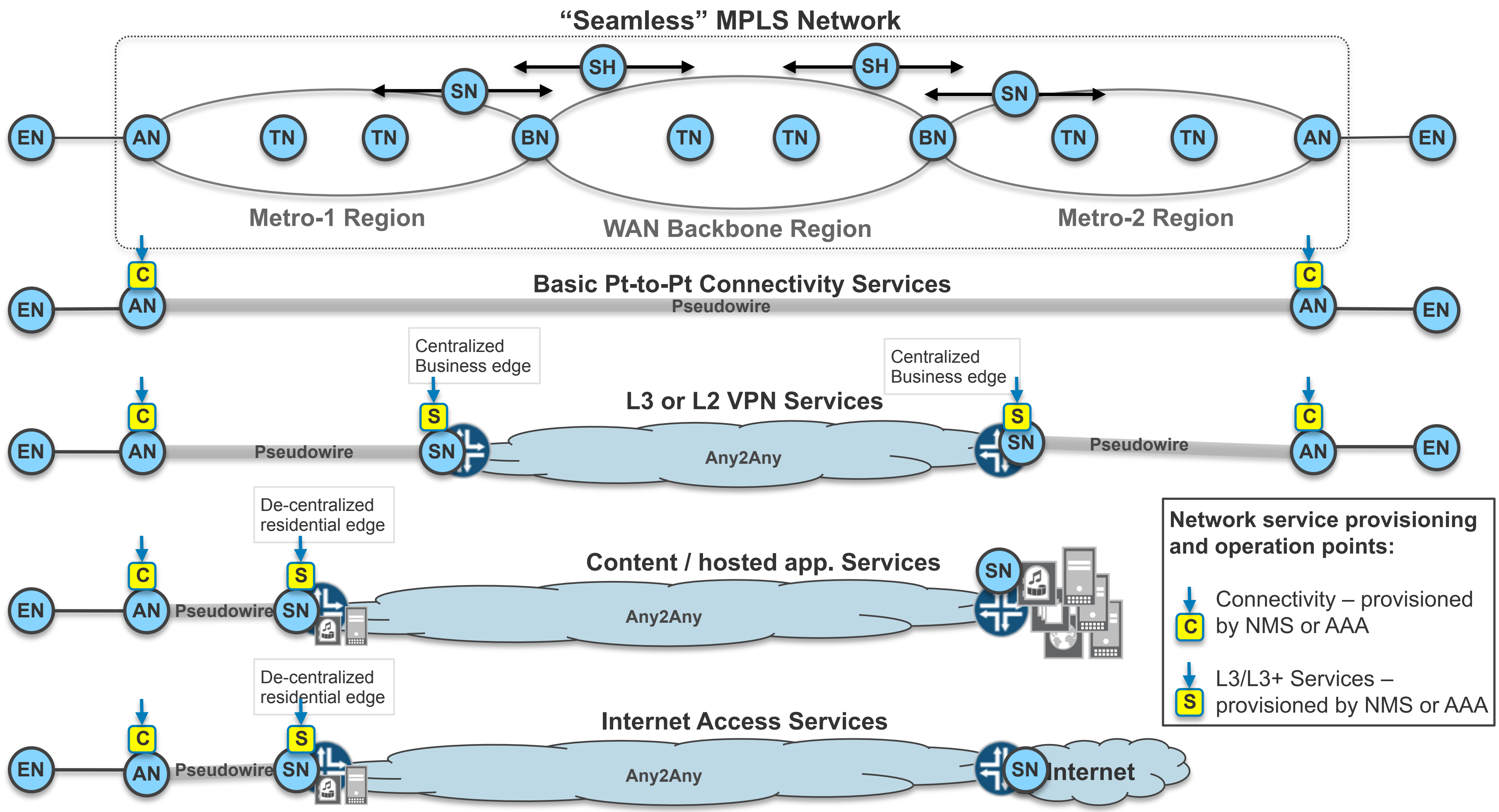
Regions

- A single network divided into regions: multiple Metro regions (leafs) interconnected by WAN backbone (core)
- Regions can be of different types: (i) IGP area, (ii) IGP instance, (iii) BGP AS
- All spanned by a single MPLS network, with any to any MPLS connectivity blueprints (AN to SN, SN to SN, AN to AN, etc)

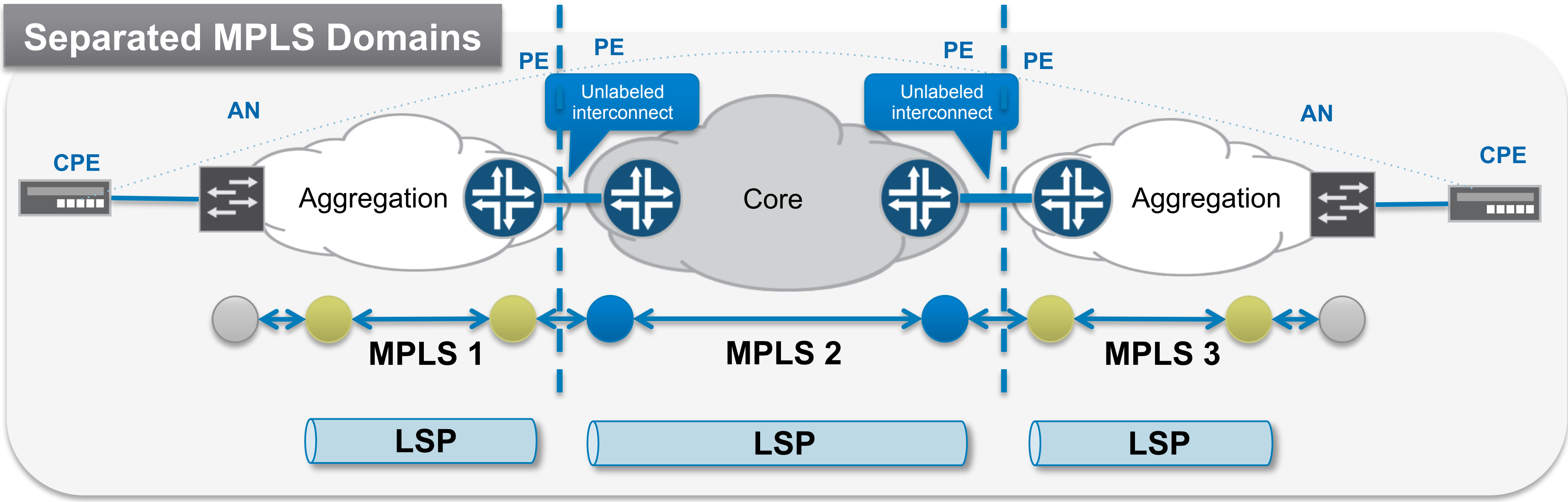
Decoupled architectures

- Services architecture – defines where & how the services are delivered, incl. interaction between SNs and SHs
- Network architecture – provides underlying connectivity for services

SEAMLESS MPLS ARCHITECTURE CONNECTIVITY AND SERVICES BLUEPRINT



CURRENT NETWORK ENVIRONMENT



Segmented inter-domain LSP signaling

- Intra-domain LSP signaling only

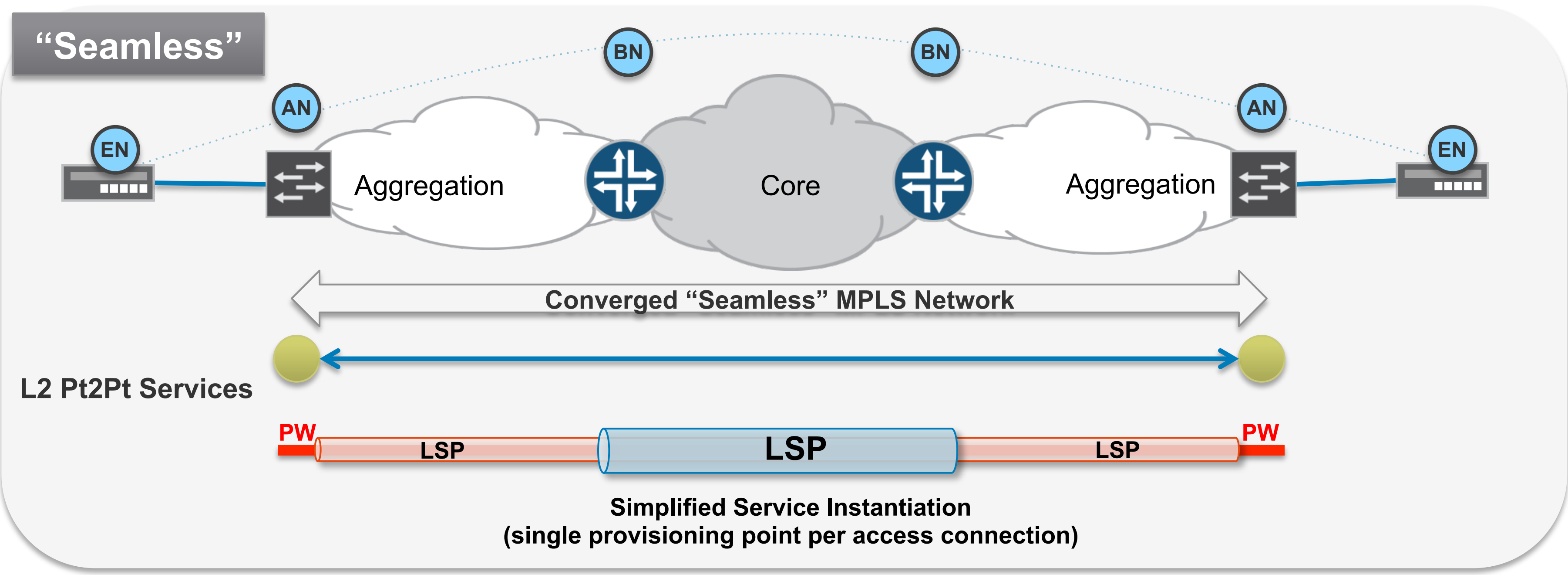
Inflexible end-to-end service stitching points

No end-to-end service protection/restoration

- Or difficult and expensive..

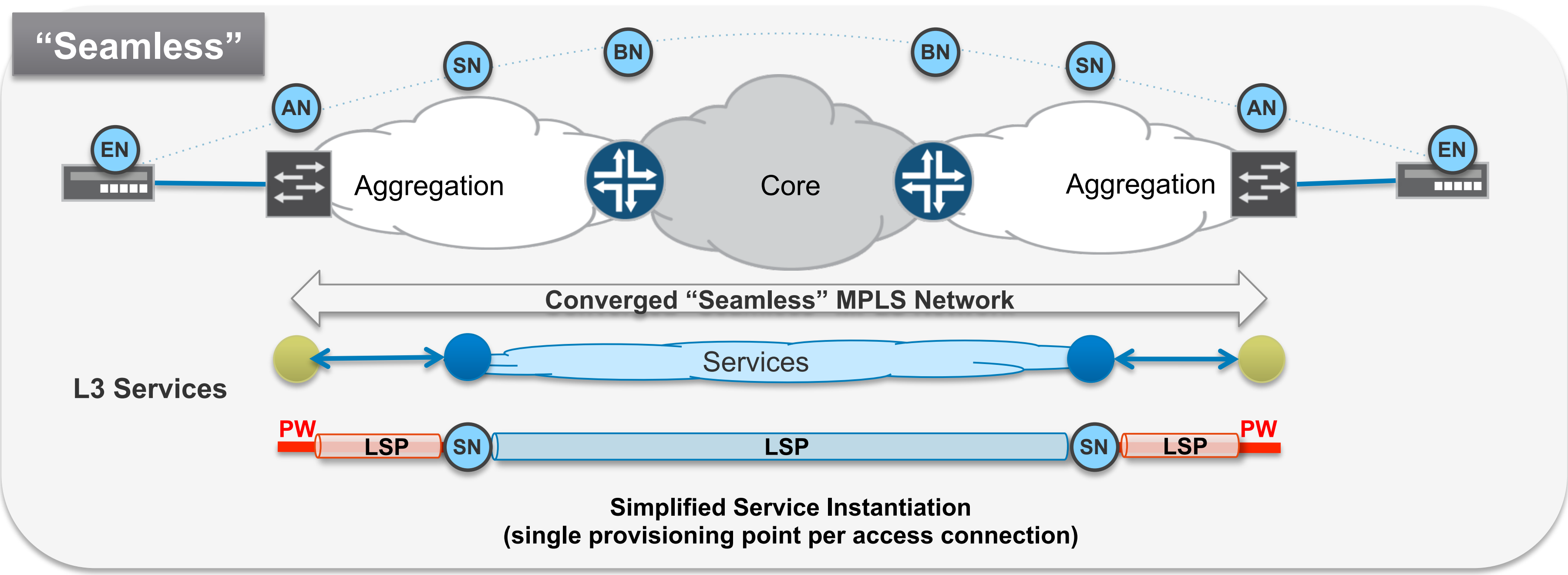
SEAMLESS MPLS – END-TO-END CONTINUITY

- End-to-end single MPLS domain, inter-area LSP signaling
- Inter-area independence through LSP hierarchy
- End-to-end service continuity (service agnostic)

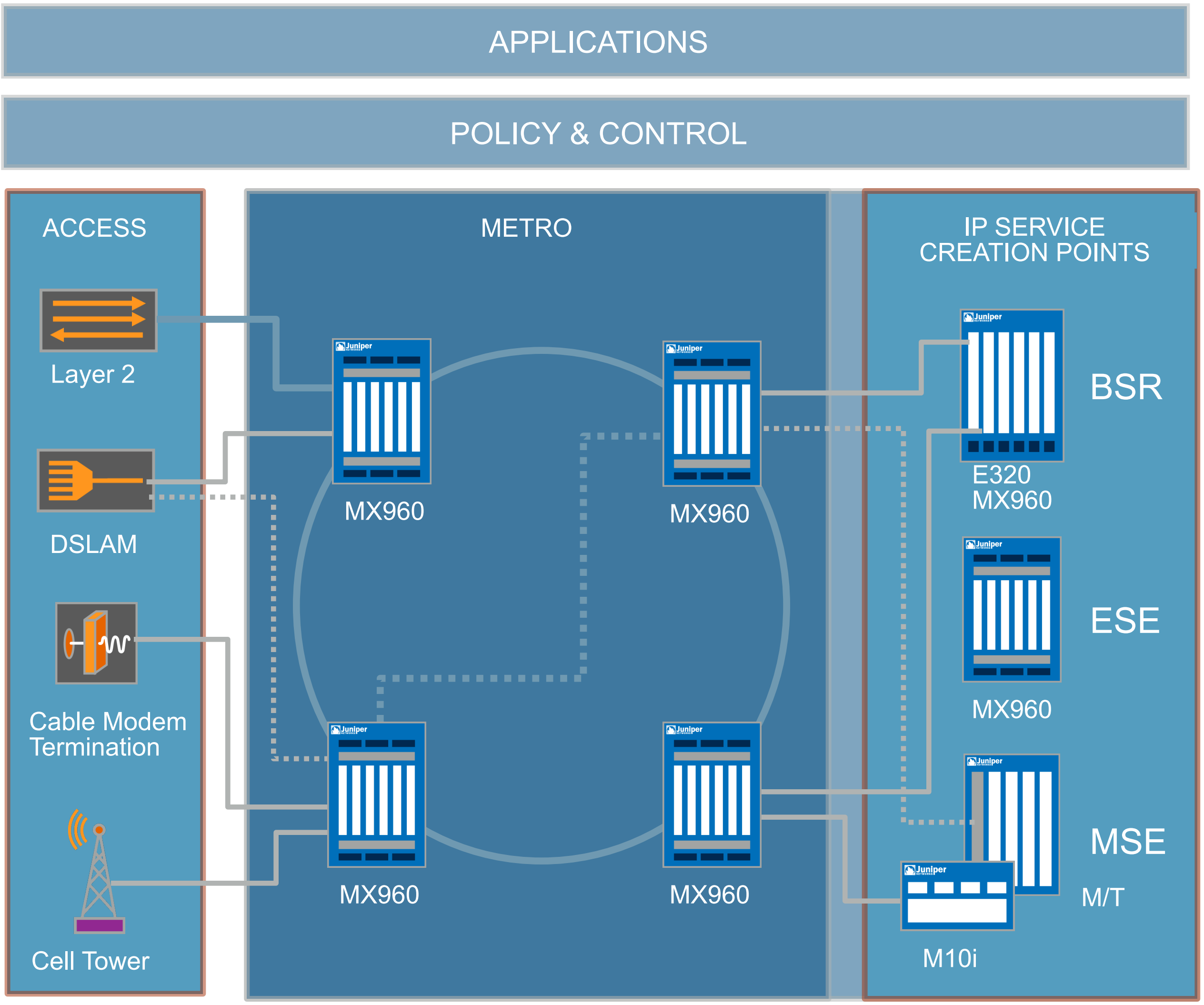


SEAMLESS MPLS – SERVICE FLEXIBILITY

- End-to-end single MPLS domain, inter-area LSP signaling
- Pseudowire access to L2/L3 network services
- Flexible topological service placement



FLEXIBILITY TO CHOOSE LOCATION OF SERVICE EDGE



- Customize location of service edge based on:
  - Scalability requirements
  - Network topology
  - Maturity of service
  - Success of service
  - Degree of location customization

SEAMLESS MPLS – DESIGN USE CASES

SEAMLESS MPLS – DESIGN USE CASE NETWORK SCALE

Design

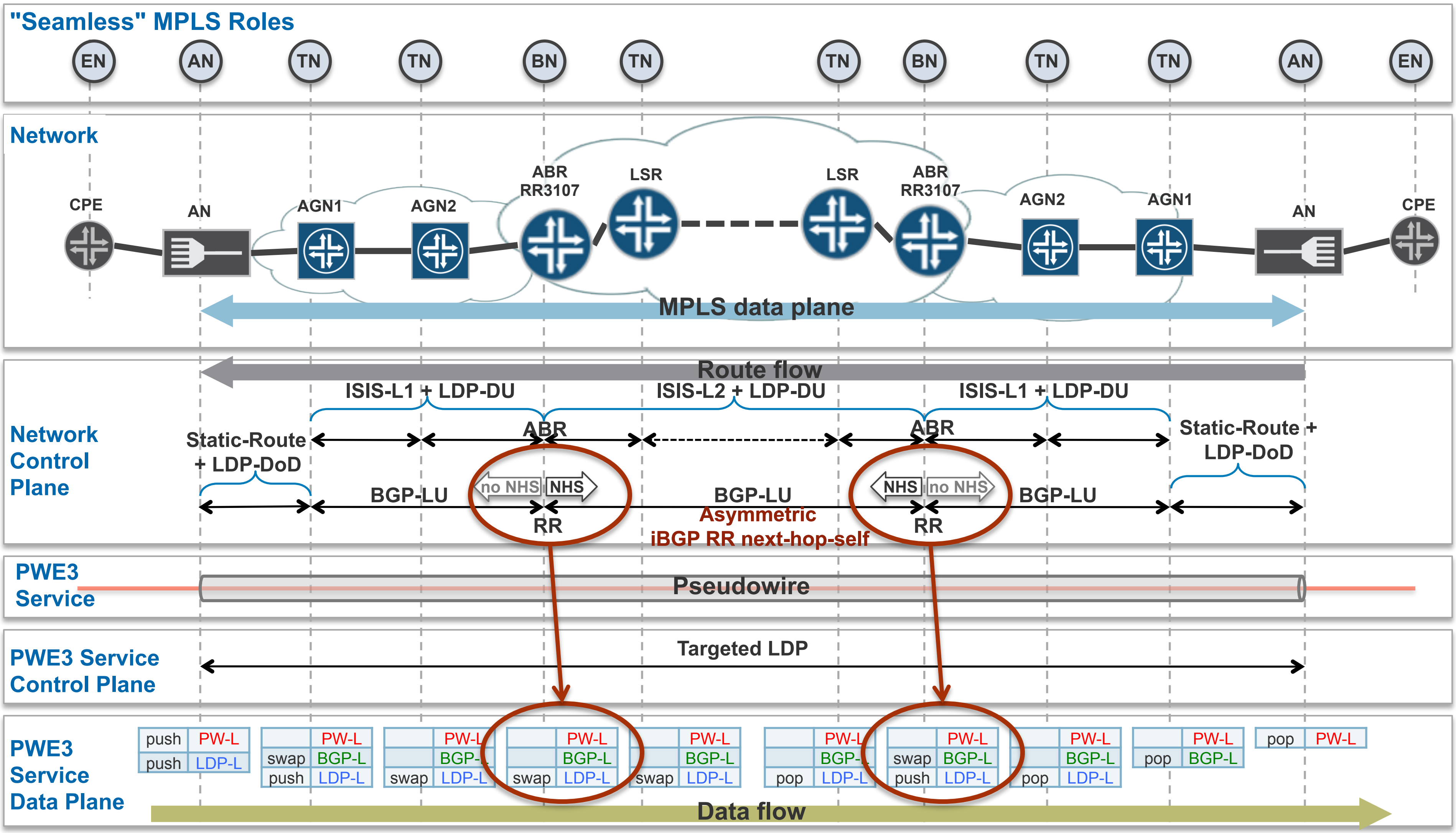
- Split the network into regions: access, metro/aggregation, edge, core
- Single IGP with areas per metro/edge and core regions
- Hierarchical LSPs to enable e2e LSP signaling across all regions
- IGP + LDP for intra-domain transport LSP signaling
  - RSVP-TE alternative to LDP
- BGP labeled unicast for cross-domain hierarchical LSP signaling
- LDP Downstream-on-Demand for LSP signaling to/from access devices
- Static routing on access devices

Properties

- Large scale achieved with hierarchical design
- BGP labeled unicast enables any-to-any connectivity between >100k devices – no service dependencies (e.g. no need for PW stitching for VPWS service)
- A simple MPLS stack on access devices (static routes, LDP DoD)

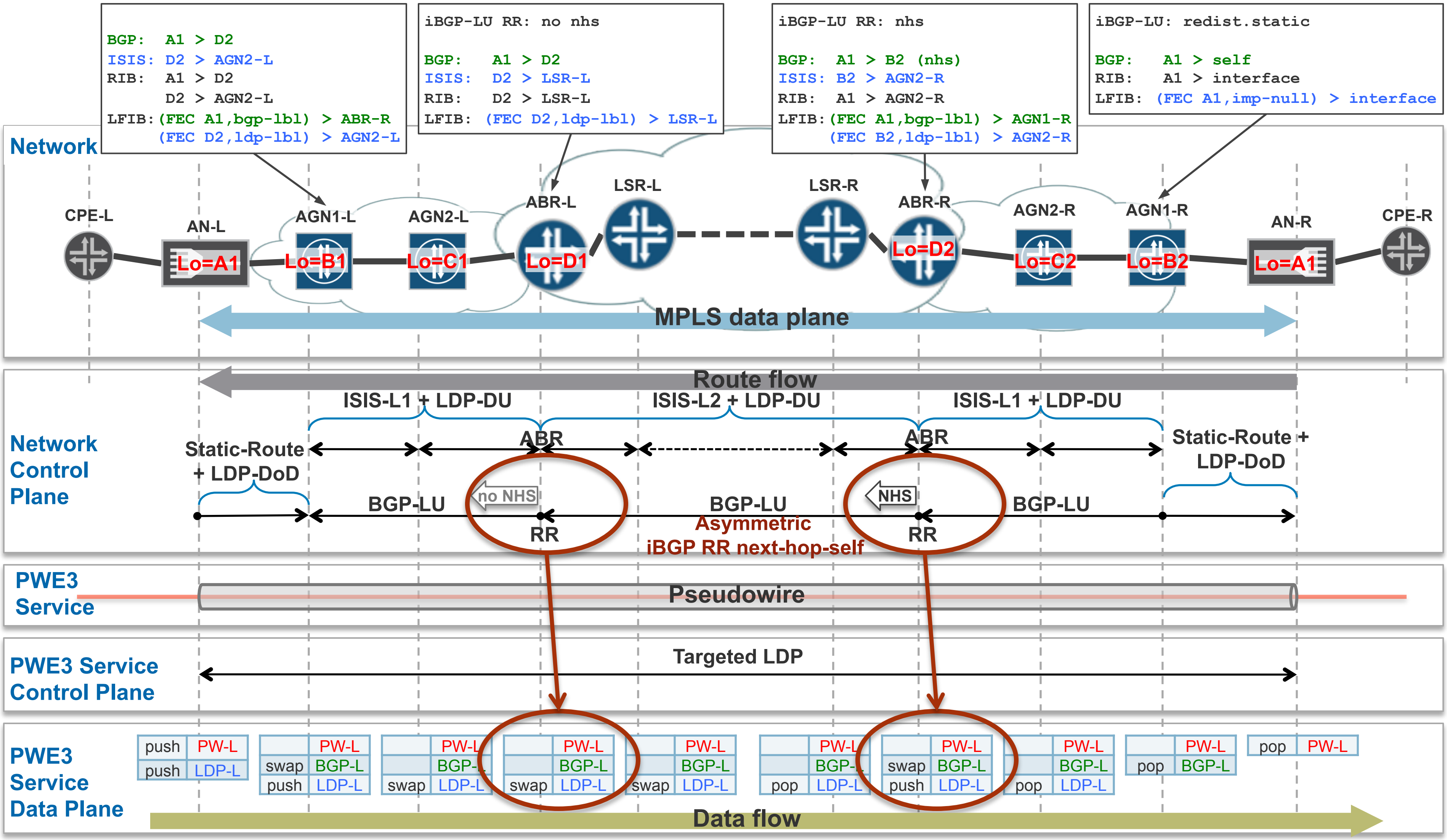
SEAMLESS MPLS – USE CASE 1*
CONTROL AND DATA PLANE LAYOUT

LDP DoD – LDP Downstream on Demand, RFC5036
LDP DU – LDP Downstream Unsolicited, RFC5036
BGP LU – BGP Label Unicast, RFC3107
NHS – BGP next-hop-self



* IP/MPLS control plane protocol stack and MPLS dataplane per "Deployment Scenario #1" in draft-mpls-seamless-mpls-00.

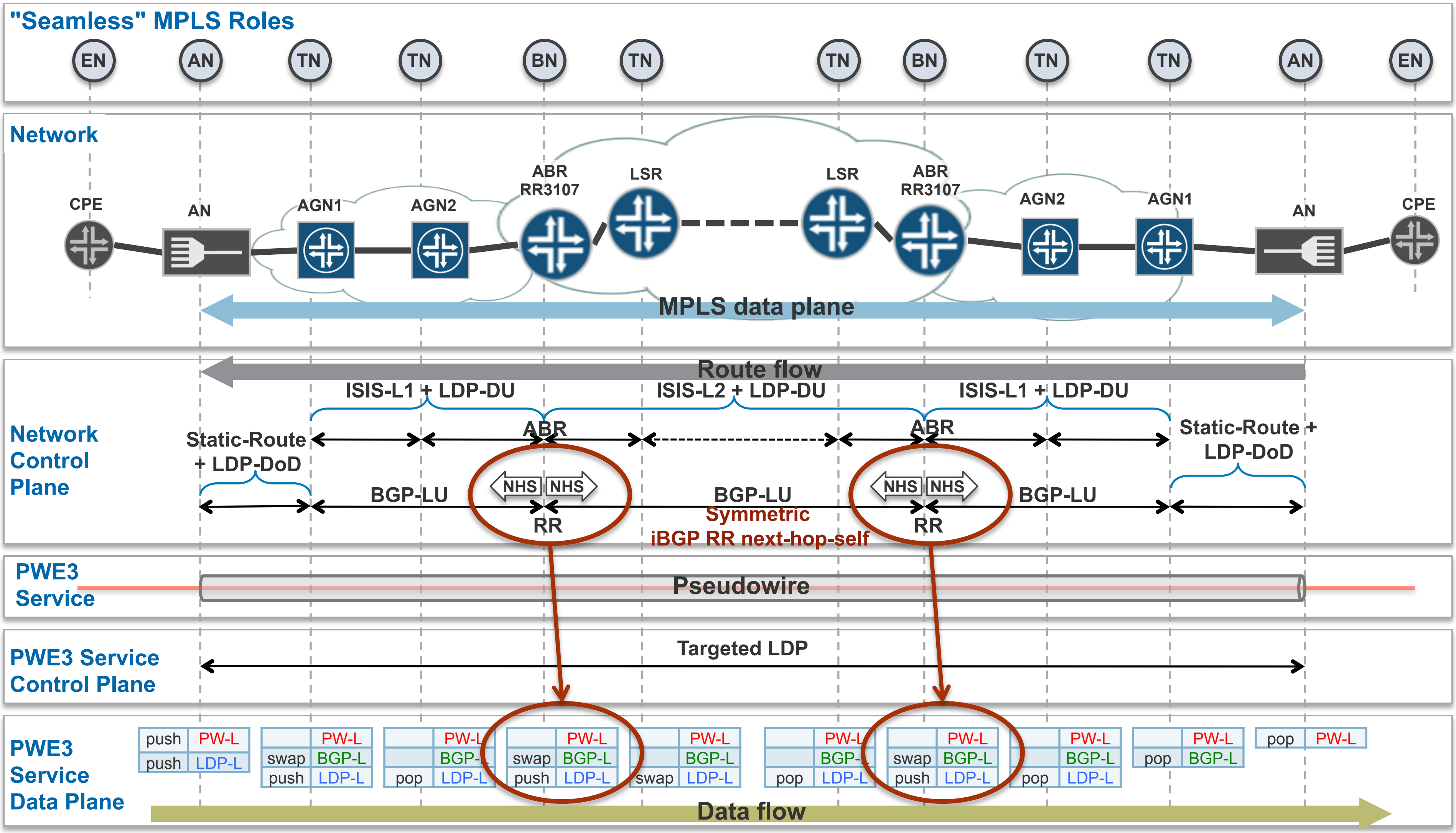
SEAMLESS MPLS – USE CASE 1*
ROUTE DISTRIBUTION EXAMPLE



* IP/MPLS control plane protocol stack and MPLS dataplane per "Deployment Scenario #1" in draft-mpls-seamless-mpls-00.

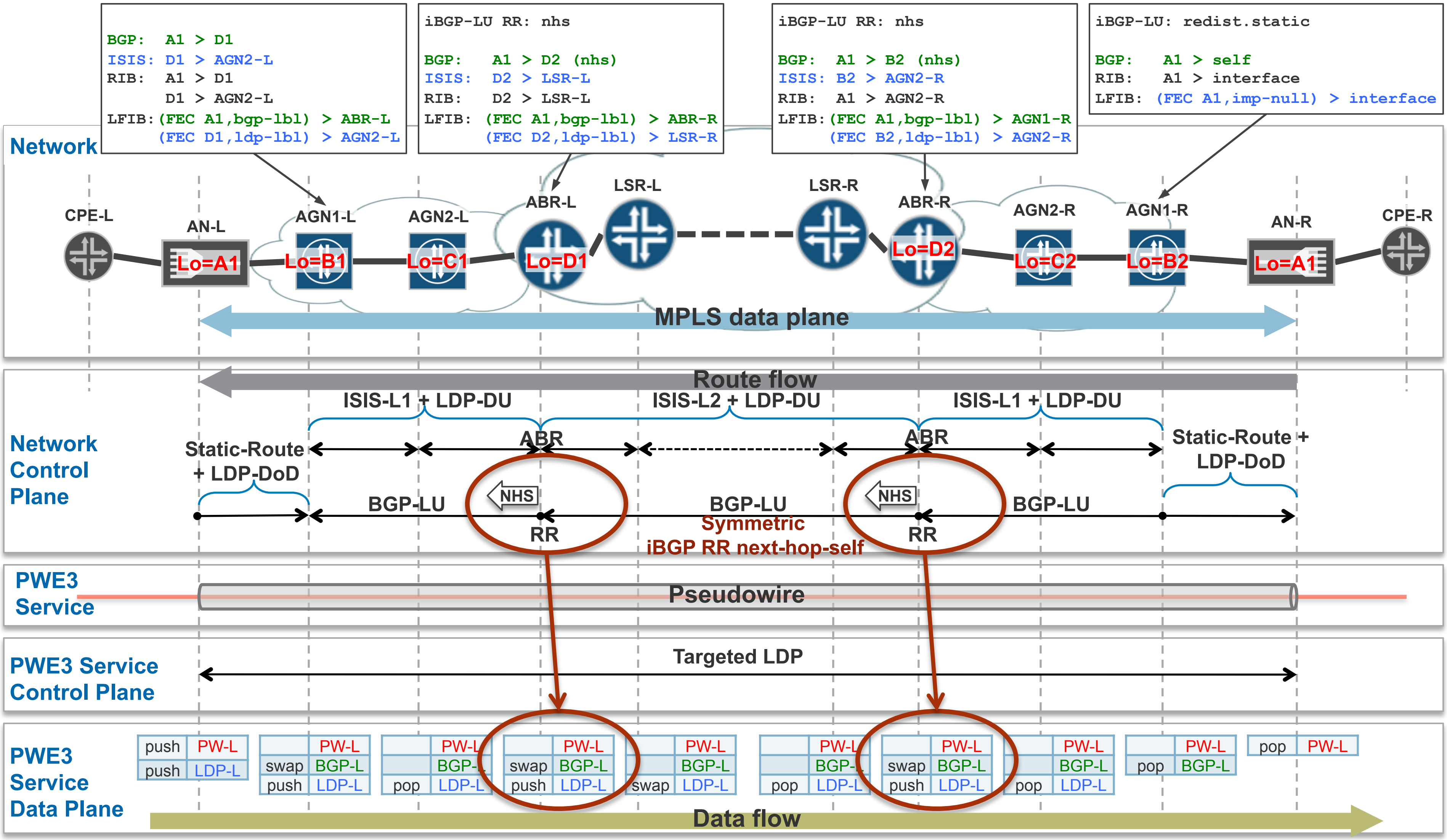
SEAMLESS MPLS – USE CASE 2*
CONTROL AND DATA PLANE LAYOUT

LDP DoD – LDP Downstream on Demand, RFC5036
LDP DU – LDP Downstream Unsolicited, RFC5036
BGP LU – BGP Label Unicast, RFC3107
NHS – BGP next-hop-self



* IP/MPLS control plane protocol stack and MPLS dataplane per "Deployment Scenario #1" in draft-mpls-seamless-mpls-00.

SEAMLESS MPLS – USE CASE 2*
ROUTE DISTRIBUTION EXAMPLE



* IP/MPLS control plane protocol stack and MPLS dataplane per “Deployment Scenario #1” in draft-mpls-seamless-mpls-00.

ENABLING IP/MPLS SCALE WITH BGP LABELED UNICAST (RFC3107)

BGP-LU enables distribution of /32 router loopback MPLS FECs

- Used between Seamless MPLS regions for any2any MPLS reachability
- Enables large scale MPLS network with hierarchical LSPs

Not all MPLS FECs have to be installed in the data plane

- Separation of BGP-LU control plane and LFIB data plane
- ***Only required MPLS FECs are placed in LFIB***
  - E.g. on RR BGP-LU FECs with next-hop-self
  - E.g. FECs requested by LDP-DoD by upstream
- Enables scalability with minimum impact on data plane resources
  - ***use what you need !***

ENABLING IP/MPLS SCALE
LDP DOWNSTREAM-ON-DEMAND (LDP DOD)

IP/MPLS routers implement LDP Downstream Unsolicited (LDP DU) label distribution

- Advertising MPLS labels for all routes in their RIB
- This is very insufficient for Access Nodes
  - Mostly stub nodes, can rely on static routing and need reachability to a small subset of total routes (labels)

AN requirement addressed with LDP DoD

- LDP DoD enables on-request label distribution ensuring that only required labels are requested, provided and installed

LDP DoD is described in RFC5036

- Seamless MPLS use cases for LDP DoD in a new IETF draft
  - draft-beckhaus-ldp-dod-01

SEAMLESS MPLS - MPLS IN THE ACCESS

GENERAL REQUIREMENTS OF ACCESS NODES SUMMARY

- **Challenge**
  - Need to enable Access Nodes integration into the MPLS network but without the need to implement the full MPLS edge node capability set
- **Requirements**
  - Access Nodes should only use the required labels
  - The solution has to support general routing capability between access and aggregation
  - The solution has to support all the required access topologies
  - The solution must not change the MPLS deployment within the rest of the network behind the border aggregation nodes
- **Use defined standard MPLS protocols**
  - No or minimal changes to standard protocols and network operation

ADDRESSING THE REQUIREMENTS OF ACCESS

- **Approach**
  - Apply an access *“subscription model”* to marry a high number of access MPLS devices with a large-scale any-to-any MPLS network
  - Employ a common MPLS label distribution protocol in a *“request mode”*
- **Solution**
  - Use *LDP Downstream-on-Demand* (DoD) MPLS label advertisement for providing only the requested labels to Access Nodes (RFC 5036)
  - Integrate LDP DoD with routing using *ordered label distribution control* (RFC 5036)
  - Enable simple access configuration and operation with default routes and *inter-area LDP* (RFC 5283)

MPLS LDP DOD IN ACCESS AND AGGREGATION USE CASES AND LDP DOD PROCEDURES

Seamless MPLS access use cases drive the required LSR LDP DoD procedures for Access Nodes and border Aggregation Nodes

I-D.draft-ietf-ldp-dod lists the access use cases and maps LDP DoD procedures against them

LDP DoD use cases (AN, AGN)

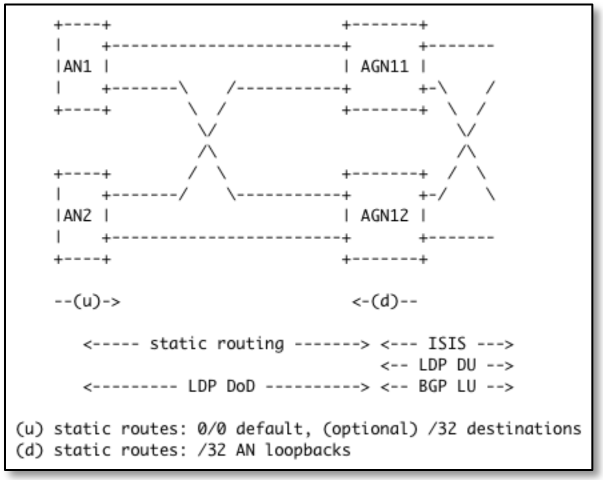
1) (AN, AGN) Initial network setup
2) (AN) Service provisioning, activation
3) (AN) Service changes, decommissioning
4) (AN) Service failure
5) (AN, AGN) Network transport failures

LDP DoD procedures (Access LSR)

a) LDP DoD session negotiation
b) Label request, mapping
c) Label withdraw
d) Label release
e) Local repair

REFERENCE ACCESS TOPOLOGIES
WITH ACCESS STATIC ROUTES AND ACCESS IGP

V



Topologies with access static routes*

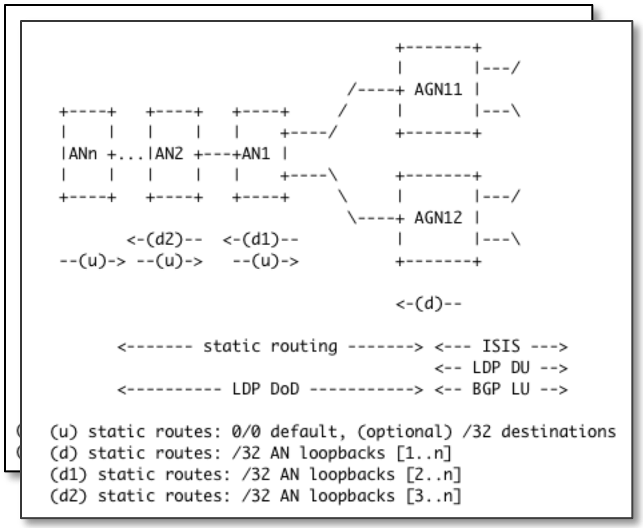
- **I1** - a single AN homed to a single AGN
- **I** - multiple ANs daisy-chained to a single AGN
- **V** - a single AN dual-homed to two AGNs
- **Y** - multiple ANs daisy-chained to two AGNs
- **U2** - two ANs dual-homed to two AGNs

*Access topology references from draft-beckhaus-ldp-dod-01.

REFERENCE ACCESS TOPOLOGIES
WITH ACCESS STATIC ROUTES AND ACCESS IGP

V

Y



```
                                    +-------+
                                    |       |---/
                               /----+ AGN11 |
 +----+   +----+   +----+     /     |       |---\
 |    |   |    |   |    +----/      +-------+
 |ANn +...|AN2 +---+AN1 |
 |    |   |    |   |    +----\      +-------+
 +----+   +----+   +----+     \     |       |---/
                               \----+ AGN12 |
        <-(d2)--  <-(d1)--          |       |---\
 --(u)-> --(u)->   --(u)->          +-------+

                                    <-(d)--

       <------- static routing -------> <--- ISIS --->
                                        <-- LDP DU -->
       <---------- LDP DoD -----------> <-- BGP LU -->

 (u) static routes: 0/0 default, (optional) /32 destinations
 (d) static routes: /32 AN loopbacks [1..n]
 (d1) static routes: /32 AN loopbacks [2..n]
 (d2) static routes: /32 AN loopbacks [3..n]
```

Topologies with access static routes*

- **I1** - a single AN homed to a single AGN
- **I** - multiple ANs daisy-chained to a single AGN
- **V** - a single AN dual-homed to two AGNs
- **Y** - multiple ANs daisy-chained to two AGNs
- **U2** - two ANs dual-homed to two AGNs

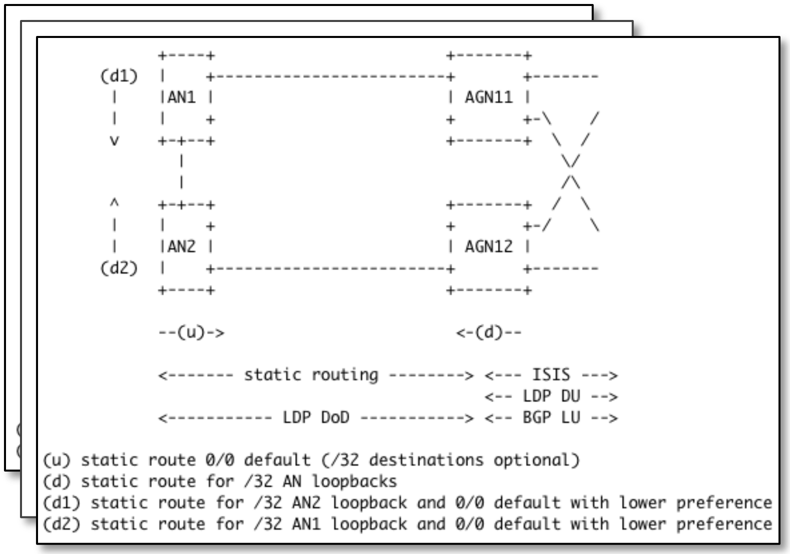
*Access topology references from draft-beckhaus-ldp-dod-01.

REFERENCE ACCESS TOPOLOGIES
WITH ACCESS STATIC ROUTES AND ACCESS IGP

V

Y

U2



Topologies with access static routes*

- **I1** - a single AN homed to a single AGN
- **I** - multiple ANs daisy-chained to a single AGN
- **V** - a single AN dual-homed to two AGNs
- **Y** - multiple ANs daisy-chained to two AGNs
- **U2** - two ANs dual-homed to two AGNs

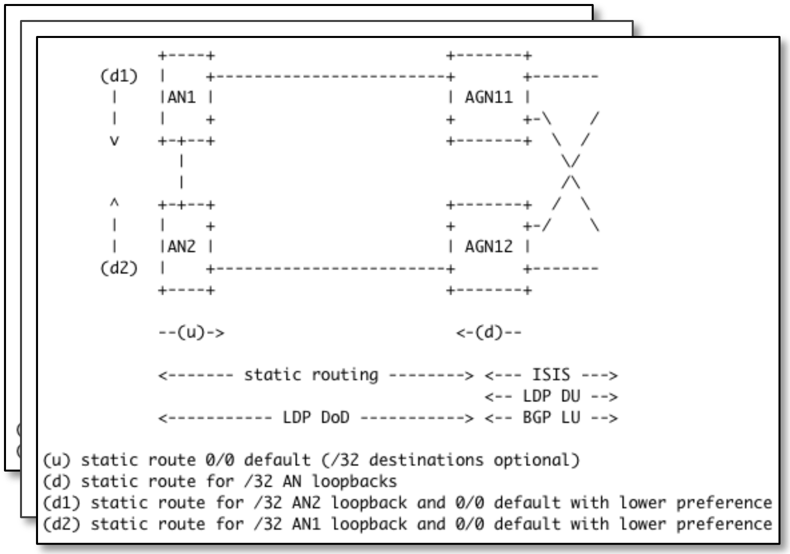
*Access topology references from draft-beckhaus-ldp-dod-01.

REFERENCE ACCESS TOPOLOGIES WITH ACCESS STATIC ROUTES AND ACCESS IGP

V

Y

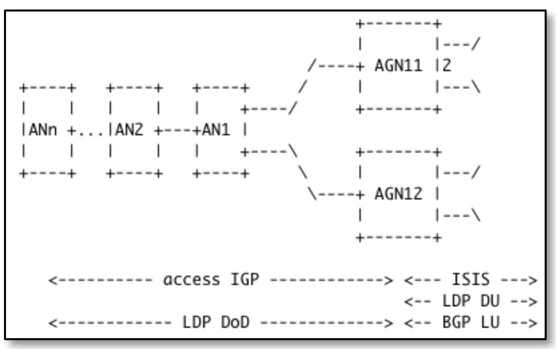
U2



Topologies with access static routes*

- **I1** - a single AN homed to a single AGN
- **I** - multiple ANs daisy-chained to a single AGN
- **V** - a single AN dual-homed to two AGNs
- **Y** - multiple ANs daisy-chained to two AGNs
- **U2** - two ANs dual-homed to two AGNs

Y



- **Topologies with access IGP***
  - **Y** - multiple ANs daisy-chained to two AGNs
  - **U** - multiple ANs in a horseshoe, dual-homed to two AGNs

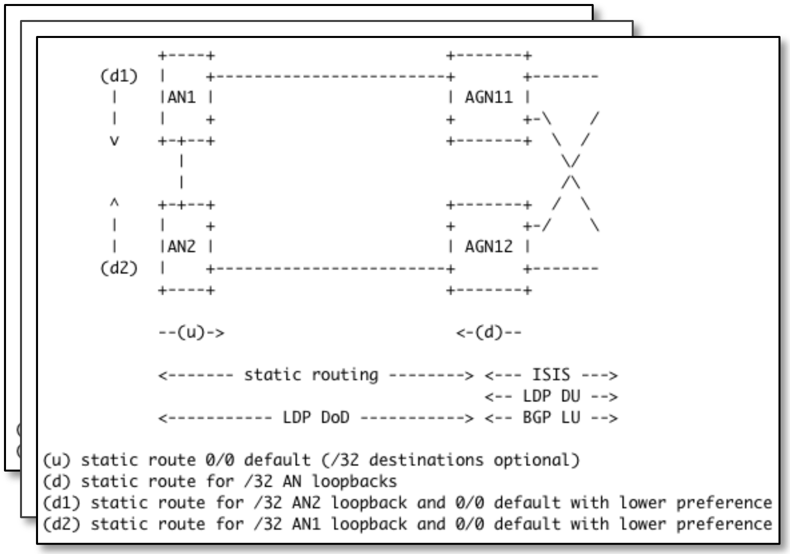
*Access topology references from draft-beckhaus-ldp-dod-01.

REFERENCE ACCESS TOPOLOGIES WITH ACCESS STATIC ROUTES AND ACCESS IGP

V

Y

U2



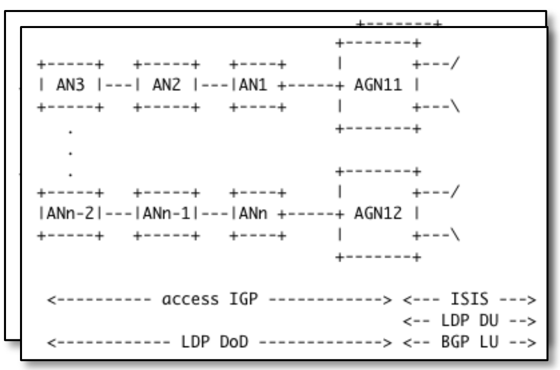
```
          +----+                        +-------+
   (d1)   |    +------------------------+       +-------
     |    |AN1 |                        | AGN11 |
     |    |    +                        +       +-\    /
     v    +-+--+                        +-------+  \  /
            |                                       \/
            |                                       /\
     ^    +-+--+                        +-------+  /  \
     |    |    +                        +       +-/    \
     |    |AN2 |                        | AGN12 |
   (d2)   |    +------------------------+       +-------
          +----+                        +-------+

          --(u)->                        <-(d)--

          <------- static routing --------> <--- ISIS --->
                                            <-- LDP DU -->
          <----------- LDP DoD -----------> <-- BGP LU -->

(u) static route 0/0 default (/32 destinations optional)
(d) static route for /32 AN loopbacks
(d1) static route for /32 AN2 loopback and 0/0 default with lower preference
(d2) static route for /32 AN1 loopback and 0/0 default with lower preference
```

Topologies with access static routes*

- **I1** - a single AN homed to a single AGN
- **I** - multiple ANs daisy-chained to a single AGN
- **V** - a single AN dual-homed to two AGNs
- **Y** - multiple ANs daisy-chained to two AGNs
- **U2** - two ANs dual-homed to two AGNs

Y

U



```
                               +-------+
+-----+   +-----+   +----+     |       +---/
| AN3 |---| AN2 |---|AN1 +-----+ AGN11 |
+-----+   +-----+   +----+     |       +---\
                               +-------+
   .
   .
   .                           +-------+
+-----+   +-----+   +----+     |       +---/
|ANn-2|---|ANn-1|---|ANn +-----+ AGN12 |
+-----+   +-----+   +----+     |       +---\
                               +-------+

 <---------- access IGP ------------> <--- ISIS --->
                                      <-- LDP DU -->
 <------------ LDP DoD -------------> <-- BGP LU -->
```

- **Topologies with access IGP***
  - **Y** - multiple ANs daisy-chained to two AGNs
  - **U** - multiple ANs in a horseshoe, dual-homed to two AGNs

*Access topology references from draft-beckhaus-ldp-dod-01.

SEAMLESS MPLS USE CASE WITH LDP DOD AND ACCESS STATIC ROUTES

① **AN** – provisioned network static routes, default* or /32 destination

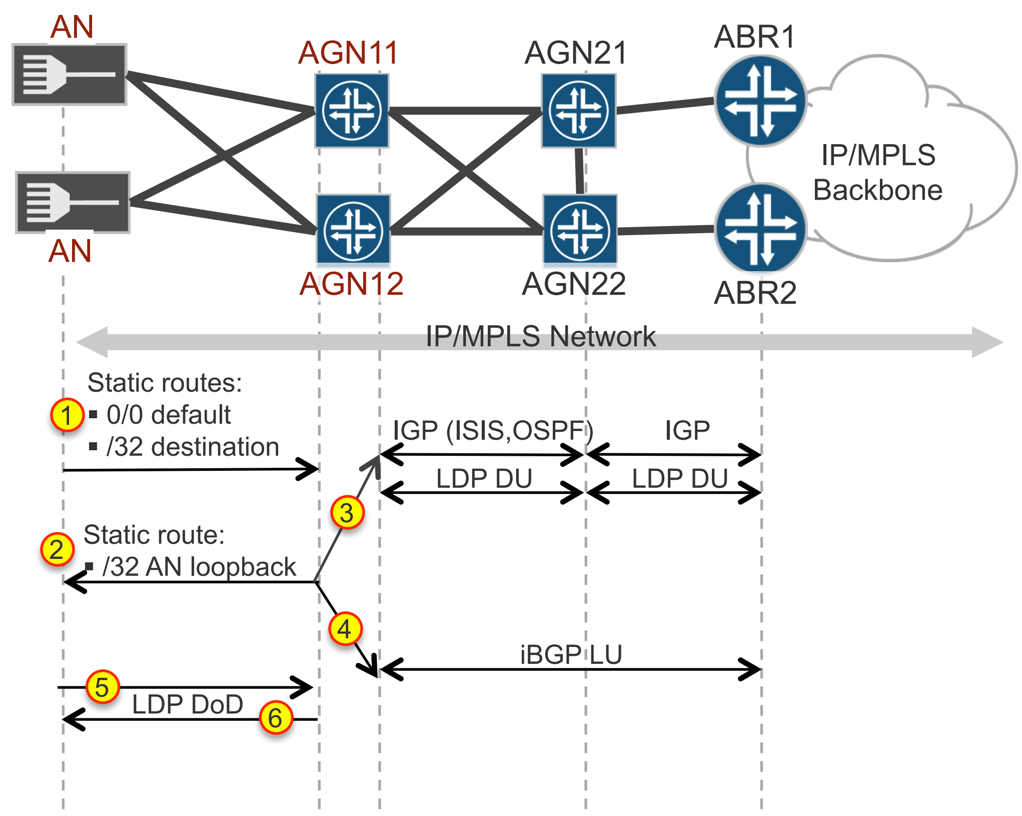
② **AGN1x** – provisioned access /32 static routes

③ **AGN1x** – (option1) access /32 statics redistributed into IGP, LDP-DU

④ **AGN1x** – (option2) access /32 statics redistributed into BGP-LU

⑤ **AN** – LDP DoD lbl requests for FECs associated with svc destinations* or configured /32 static routes

⑥ **AGN1x** – LDP DoD lbl requests for FECs associated with access /32 static routes



LDP DoD – Label Distribution Protocol, Downstream on Demand distribution, RFC 5036
LDP DU – Label Distribution Protocol, Downstream Unsolicited distribution, RFC 5036
BGP LU – Border Gateway Protocol, Label Unicast extensions, RFC 3107

(*) Requires inter-area LDP (RFC 5283), match on longest prefix in RIB.

SEAMLESS MPLS USE CASE WITH LDP DOD AND ACCESS IGP

① **AN** – provisioned access IGP instance

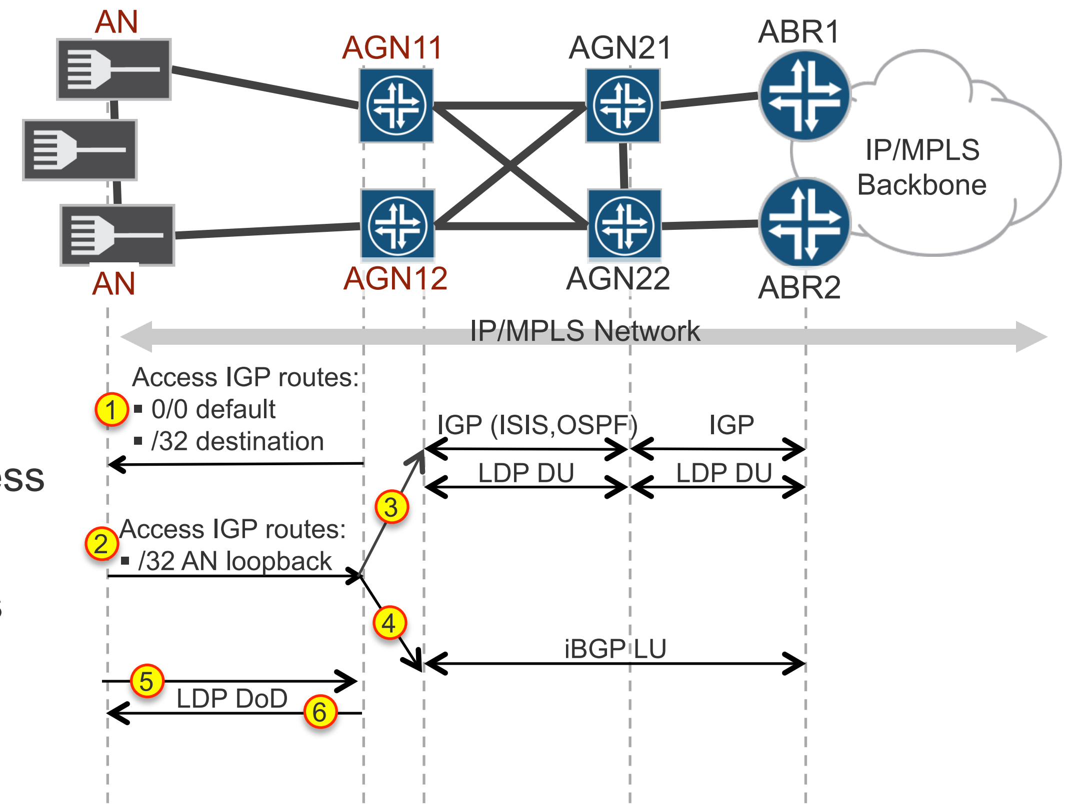
② **AGN1x** – provisioned access IGP

③ **AGN1x** – (option1) access IGP routes redistributed into IGP, LDP-DU

④ **AGN1x** – (option2) access IGP routes redistributed into BGP-LU

⑤ **AN** – LDP DoD lbl requests for FECs associated with svc destinations* or access IGP /32 routes

⑥ **AGN1x** – LDP DoD lbl requests for FECs associated with access IGP /32 routes



LDP DoD – Label Distribution Protocol, Downstream on Demand distribution, RFC 5036
LDP DU – Label Distribution Protocol, Downstream Unsolicited distribution, RFC 5036
BGP LU – Border Gateway Protocol, Label Unicast extensions, RFC 3107

(*) Requires inter-area LDP (RFC 5283), match on longest prefix in RIB.

ENABLING IP/MPLS SCALE WITH LDP LDP DOD – SUMMARY

In the Seamless MPLS design, scaling into the access does introduce *new functional and operational requirements*

- *LDP DoD* approach provides *a simple yet very effective solution* for access network in a large scale MPLS design
- The solution meets all of the requirements and relies on *defined standard IP/MPLS* protocols

LDP DoD design can be adopted to other large scale IP/MPLS deployments e.g. MPLS to cell site gateways

UNIVERSAL EDGE WITH MPLS ACCESS

THE BASIC IDEA IS TO USE MPLS IN METRO AND ACCESS ...

MPLS is already in the core and in most metros

- Now MPLS can be scaled up into the access too

Next step is to use it for services ☺

- Use MPLS pseudowires between access and edge nodes
- Enable service edge to natively terminate MPLS on the access side...

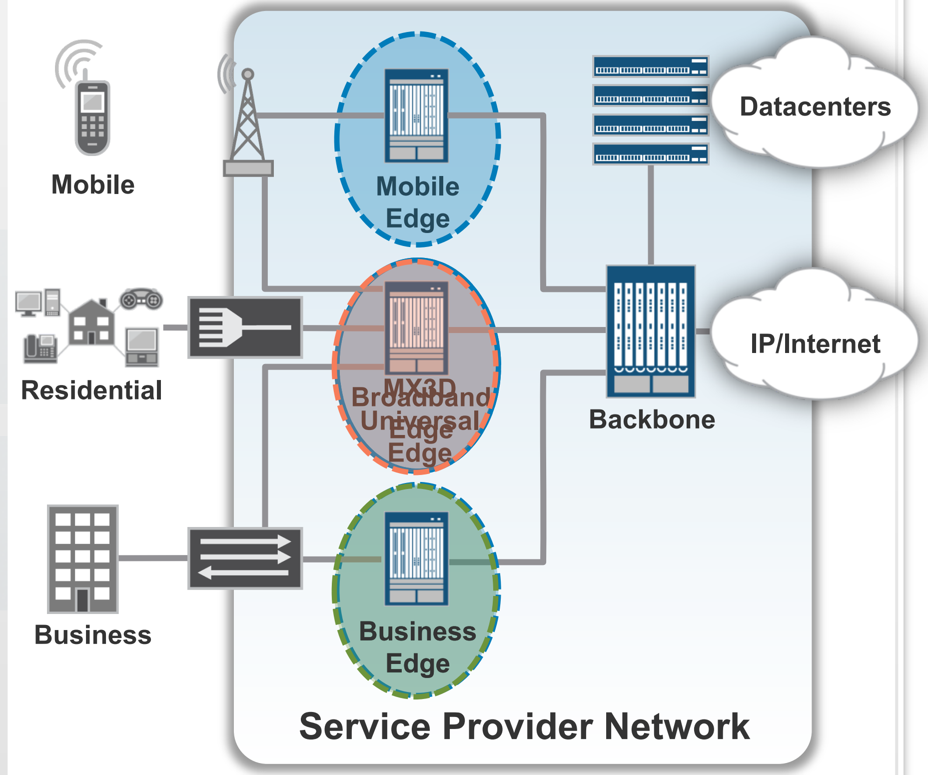
Immediate benefits

- No multiple breakouts in/from Ethernet VLAN trunks
- Greater flexibility of service edge placement
- Simpler e2e design

... AND INTEGRATE WITH A UNIVERSAL EDGE

Universal Edge Example Benefits

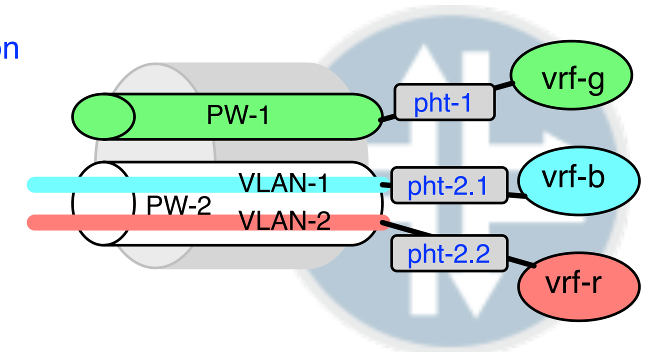
- Convergence of wireline residential and business edge
- Convergence of wireline and wireless edge services
- Convergence of policy services
- TCO reductions via single Universal Edge



CONNECTIVITY AND SERVICE INTERFACES FOR BUSINESS AND BROADBAND EDGE

Service Edge with Pseudowire Headend Termination (PHT)

PW Type	Encapsulation and PHT classification
4 or 5	PW-lbl | MAC | IPH | IP-payload
4 or 5	PW-lbl | MAC | VLAN | IPH | IP-payload



PW-lbl – Pseudowire label
MAC – MAC header
VLAN – 802.1Q or QinQ/.3AD tag
IPH – IP header

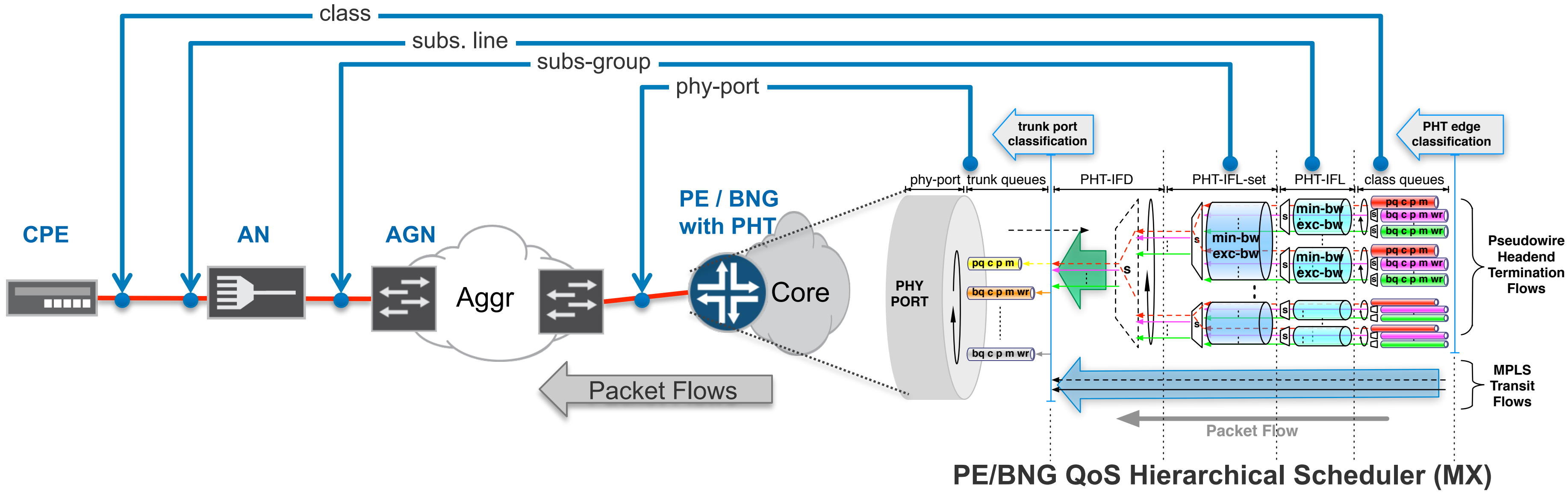
PHT for business services

- Pseudowire per subscriber (customer) line, carries a single service or bundle of services (service per VLAN, multiple VLANs)

PHT for residential broadband services

- Pseudowire per access node (DSLAM, OLT), carries multiple subscriber lines and sessions

ADDING HIERARCHICAL QOS – A SAMPLE DEPLOYMENT MAPPING



phy-port	local port of the PE / BNG; may or may not be oversubscribed.
access node or subscriber group	subscribers served by a single Access Node (AN e.g. DSLAM, OLT), multiple subscriber groups may be present on a single AN, associated shape rate reflects either the BW of AGN-to-AN link or part thereof that is “carved out” for specific subscriber group.
access port	subscriber line (copper or fibre) terminated on AN, associated shape-rate reflects the BW of this line or sub-rate thereof based on specific subscriber SLA.
QoS class	QoS forwarding class, associated with service and/or application, multiple classes per session/line (4 to 8).

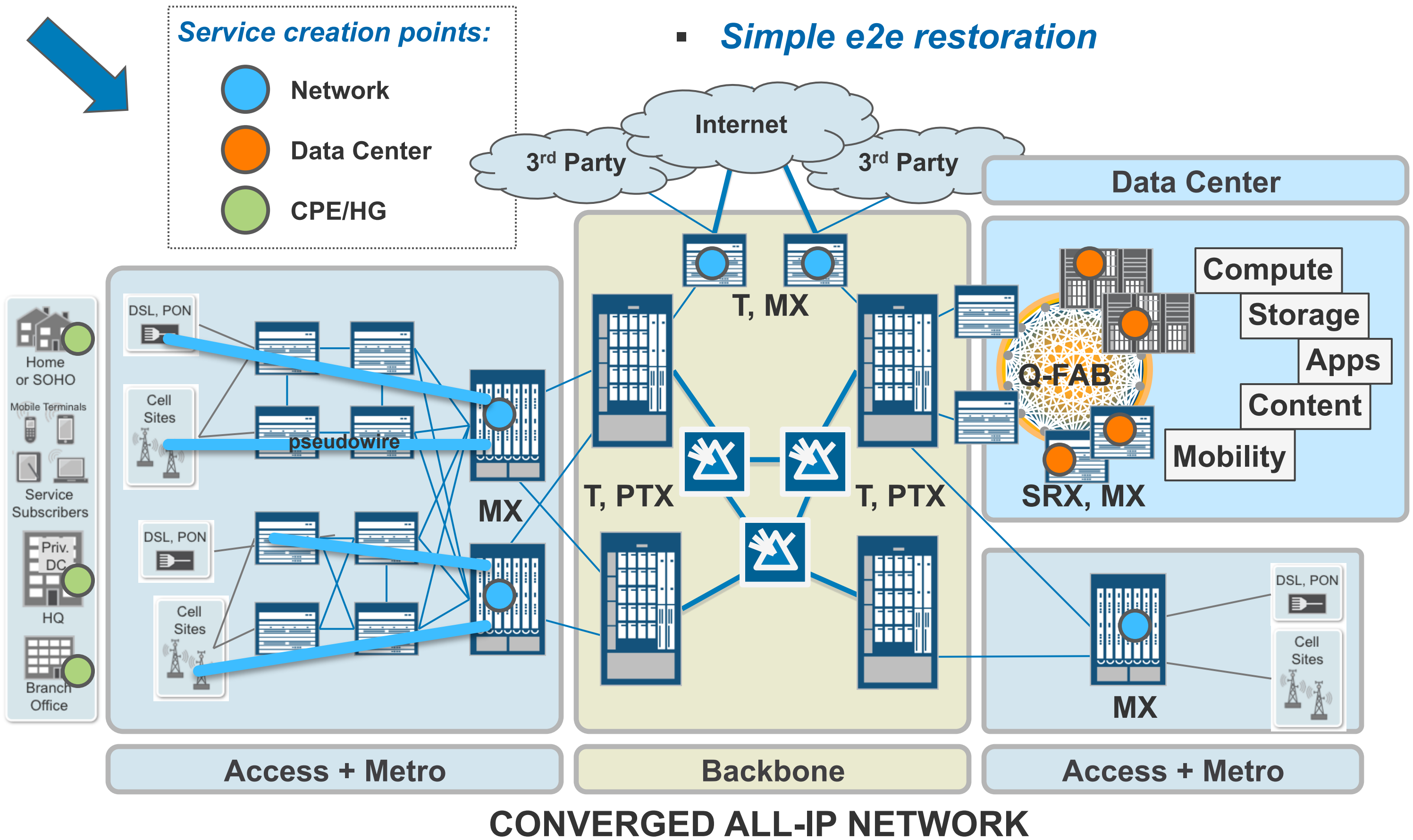
SIMPLER SERVICE DELIVERY WITH CENTRALIZED UNIVERSAL EDGE - MX

Simplicity principle driving the design

- ***Simple protocol stack***
- ***Simple service creation***
- ***Simple e2e restoration***

***Residential** - automate service provisioning and operation*:
- Network port
- Socket in the wall
- User
- Service

***Business** - uniform service provisioning and operation:*
- L3 and L2 services



JUNIPER NETWORKS

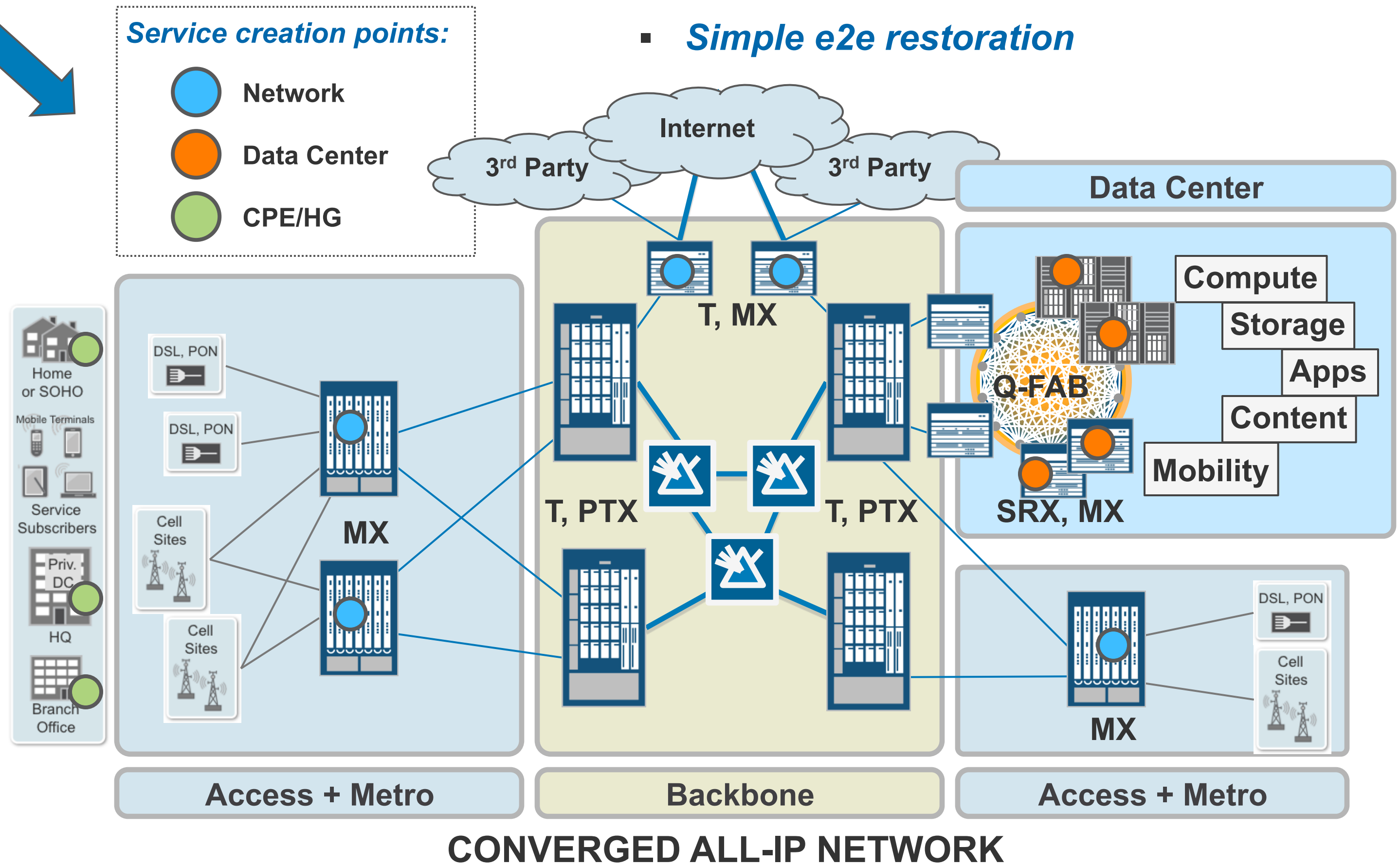
SIMPLER SERVICE DELIVERY WITH DE-CENTRALIZED UNIVERSAL EDGE - MX

Simplicity principle driving the design

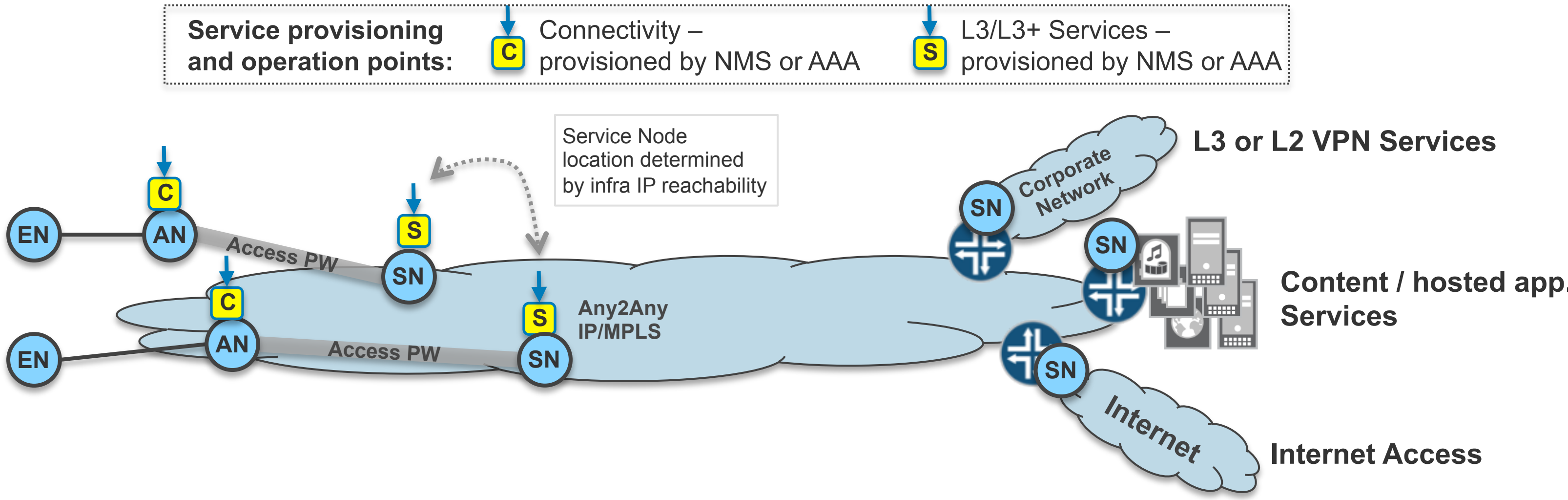
- ***Reduced number of network elements***
- ***Simple protocol stack***
- ***Simple service creation***
- ***Simple e2e restoration***

***Residential** - automate service provisioning and operation*:
- Network port
- Socket in the wall
- User
- Service

***Business** - uniform service provisioning and operation:*
- L3 and L2 services



POTENTIAL FUTURE DEVELOPMENTS:
SERVICE RELOCATION AND AUTO-PROVISIONING



Service touchpoint locations placed flexibly within the IP/MPLS network

- Service access ports bound to Service Nodes, that are programmed within the network
- Service access flows switched to Service Nodes based on self automated network reachability

Reduction of provisioning points with SW programmable self- automated network

- Access node programmed with virtual access ports bound to service access ports
- Network routes virtual access ports and flows to statically or dynamically allocated Service Node location(s)
- Applicable to all service types with dynamic (or static) optimization of service and network capacity

SUMMARY

SUMMARY

Seamless MPLS architecture meets converged network goals

- support for all packet services
- support for a large scale network
  - ***MPLS LSP hierarchy*** with BGP-LU
  - ***MPLS in access*** with LDP DoD

Seamless MPLS can be combined with ***Juniper Universal Edge***

- a single platform (MX) for business and residential services over PHT
- enables ***flexible topological placement*** of Universal Edge

REFERENCES

- draft-mpls-seamless-mpls, N.Leymann et al, May 2011.
- draft-beckhaus-ldp-dod, T.Beckhaus, M.Konstantynowicz et al, October 2011.
- “Enabling Seamless MPLS using LDP DoD”, T.Beckhaus, M.Konstantynowicz, MPLS2011 conference.
- “"Seamless" MPLS”, K.Kompella, MPLS WC 2009.

BACKUP: END-TO-END SERVICE PROTECTION & RESTORATION

SCALE ENABLERS
BGP LABELED UNICAST (RFC3107)

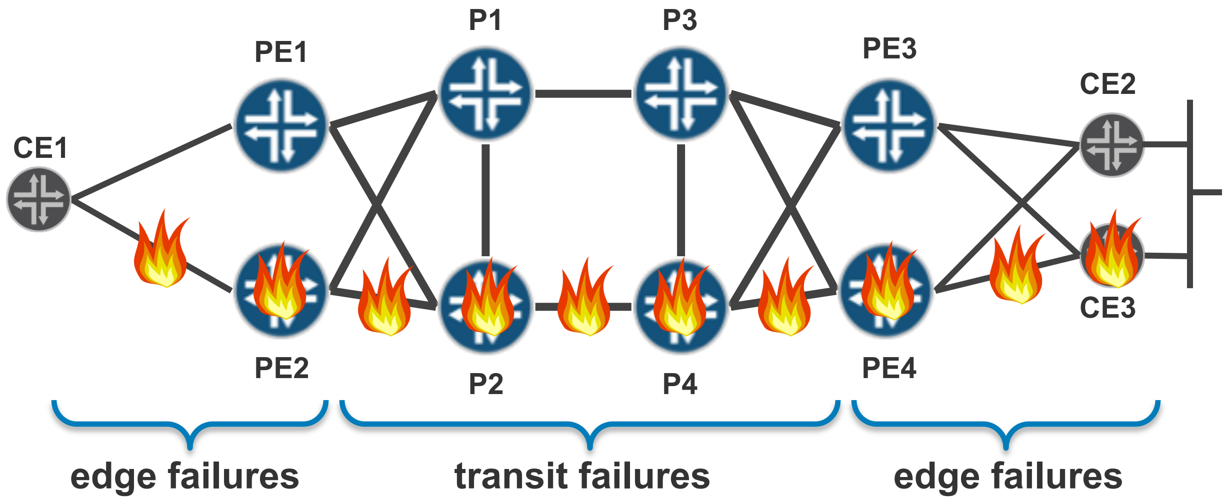
BGP-LU enables distribution of /32 router loopback MPLS FECs

- Used between Seamless MPLS regions for any2any MPLS reachability
- Enables large scale MPLS network with hierarchical LSPs

Not all MPLS FECs have to be installed in the data plane

- Separation of BGP-LU control plane and LFIB
- Only required MPLS FECs are placed in LFIB
  - E.g. on RR BGP-LU FECs with next-hop-self
  - E.g. FECs requested by LDP-DoD by upstream
- Enables scalability with minimum impact on data plane resources – use what you need approach

NETWORK FAILURE EVENT TYPES



General categories of network failures:

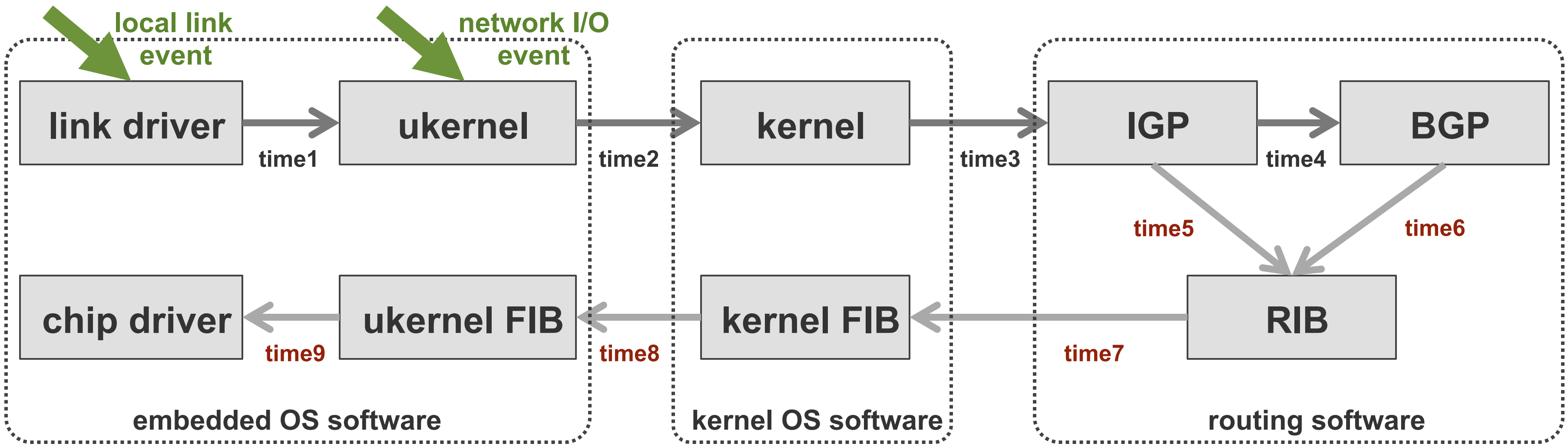
- ***Transit failures*** – link / node
  - Requires alternate network paths to be propagated or pre-programmed
  - Fairly easy to protect against, subject to topology
- ***Edge failures*** – ingress to / egress from the network
  - Requires edge state to be propagated or pre-programmed
  - Harder to protect against

ANATOMY OF NETWORK CONVERGENCE

Four Basic stages in the following order

1. Detection of Failure
   - Link flap
   - Linecard failure
   - Router crash
   - Metric Change (Administrative or Learnt from peer)
2. Flooding of event
   - Link Status
   - Routing Updates
3. Computation of Alternate Path
   - Short Path Calculations
   - RSVP FRR
   - Loop Free Alternates
4. Forwarding Plane Update

FAILURE EVENT PROPAGATION INSIDE THE ROUTER OBSERVATIONS



Service restoration is ***"Event"*** driven

An ***"Event"*** may be any local or remote change in the ***Network***

Arrival of an ***"Event"*** at service router is ***non-deterministic***

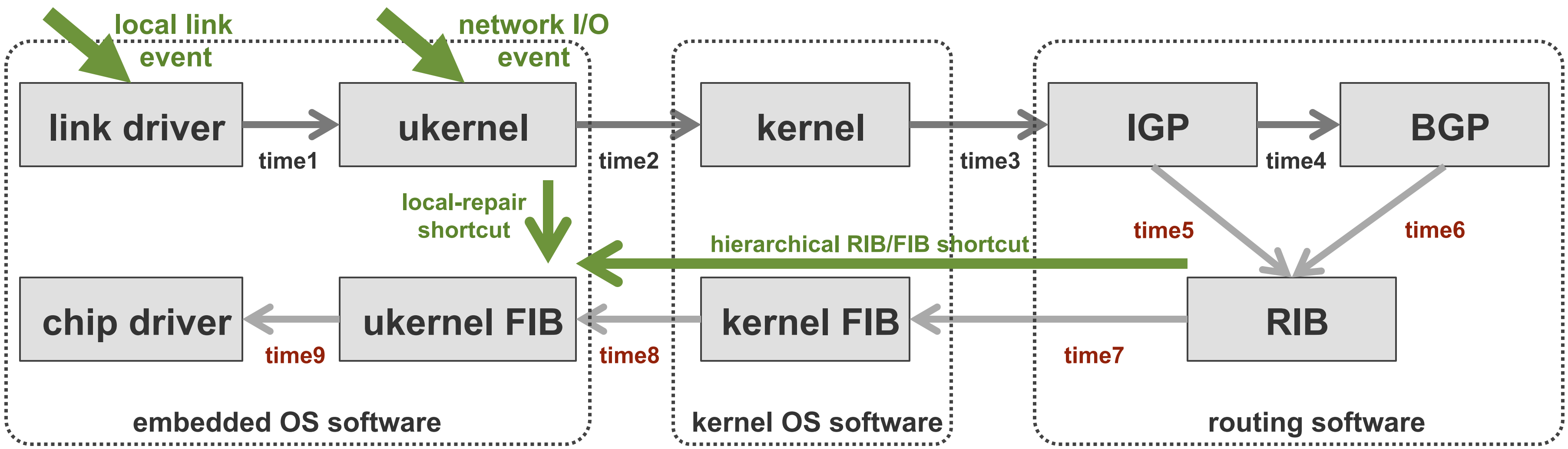
Impact of an ***"Event"*** through the router system is ***non-deterministic***

FAILURE EVENT PROPAGATION INSIDE THE ROUTER
Shortcuts are key

Many vendors have implemented ***"Shortcuts"***

- Fast path invalidation (Data-plane or Control-plane driven)
- Local repair (Data-plane driven)

"Level" of shortcut (the lower the better) is key for ***Deterministic FIB updates***



-> Less is more !

END-TO-END RESTORATION
Local vs. Global Repair

Local-repair

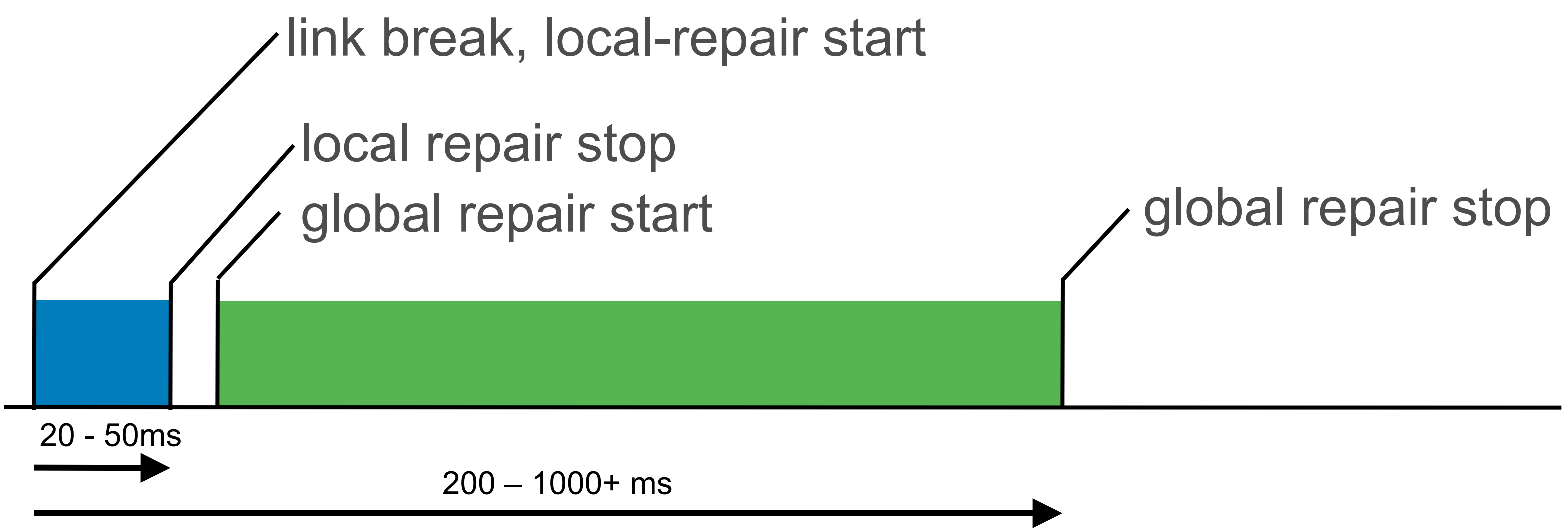
- Based on the pre-computed local backup forwarding state - provides sub-50msec restoration

Global-repair

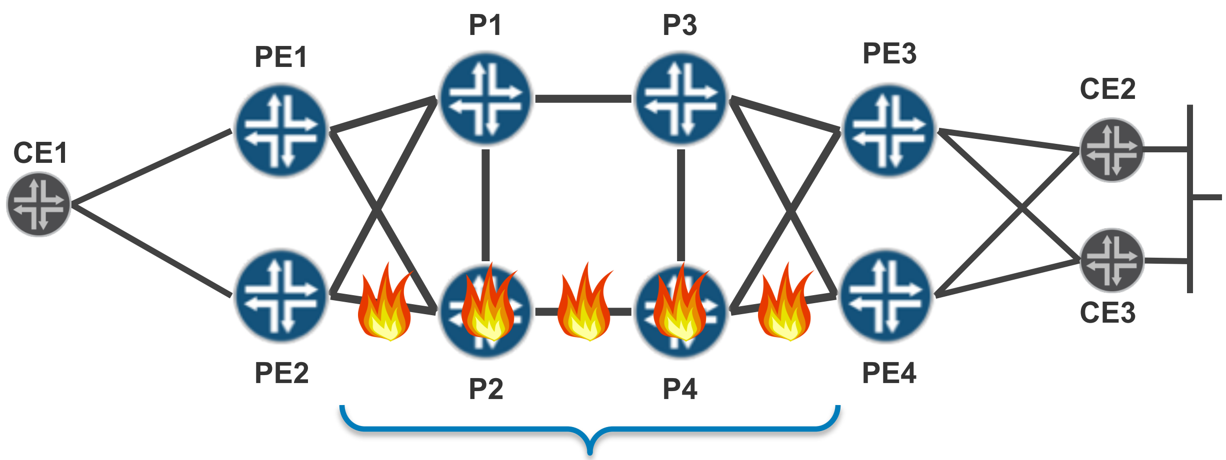
- Requires signaling to take place after failure detection - can provide sub-1sec or longer restoration times

Local-repair *complements* Global-repair

- Local-repair keeps traffic flowing while
- Global-repair gets things right
- Variation of "Make before break"



RESTORATION FROM TRANSIT FAILURES



Global-repair

Fast IGP - IS-IS, OSPF

- Fast failure detection
  - PHY/LOS, OAM/BFD
- Failure and topology change flooding
- Local SPF calcs, local RIB/FIB updates

Local-repair

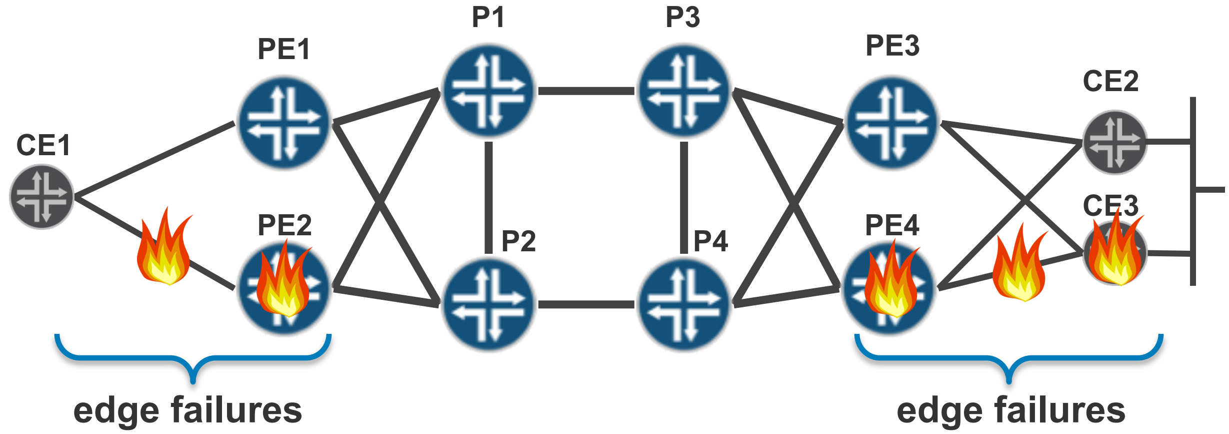
RSVP-TE FRR

- Link and node protection with bypass LSPs
- 100% topology coverage

IP FRR

- Loop Free Alternates with pre-programmed alternate next-hops
- Good but not complete native topology coverage
  - RSVP bypass tunnels for complete coverage

RESTORATION FROM EDGE FAILURES



Global-repair

Fast IGP - IS-IS, OSPF

- As for transit failures
- Used as a trigger for BGP next-hop change

Hierarchical FIB

- Hierarchical FIB with pre-programmed alternate BGP next-hops
- Based on the Junos indirect- and composite-next-hop technologies

Local-repair

PE-CE link failure

- (1) Vrf-table-label with IP lookup
- (2) PE-CE link protection

Egress PE node protection !

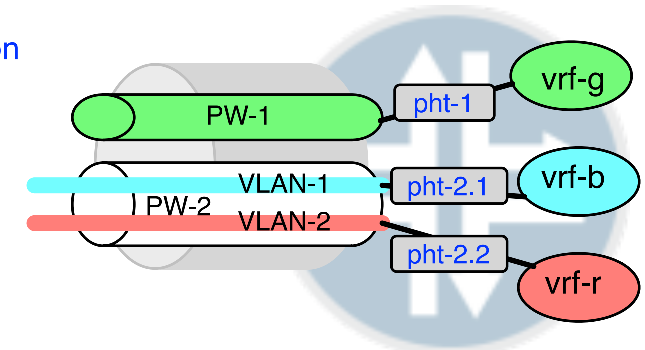
- LSP tailend protection with context label lookup
  - Local-repair by PLR* transit router

*PLR – point of local repair

PSEUDOWIRE HEADEND TERMINATION (PHT) FOR BUSINESS AND BROADBAND SERVICES

PW Type	Encapsulation and PHT classification
4 or 5	PW-lbl | MAC | IPH | IP-payload
4 or 5	PW-lbl | MAC | VLAN | IPH | IP-payload

Service Node



PW-lbl – Pseudowire label
MAC – MAC header
VLAN – 802.1Q or QinQ/.3AD tag
IPH – IP header

Business Edge

- Pseudowire per subscriber (customer) line, carries a single service or bundle of services (service per VLAN, multiple VLANs)
- Implementation based on JUNOS LT, later on Pseudowire Services IFD

Broadband Edge

- Pseudowire per access node (DSLAM), carries multiple subscriber lines and sessions
- Implementation based on JUNOS Pseudowire Services IFD

PHT FAILURE HANDLING
LOCAL LINK FAILURE

Local link failure is handled by native local-repair

- IS-IS LFA with MPLS LDP
- RSVP TE-FRR
- L2 LAG

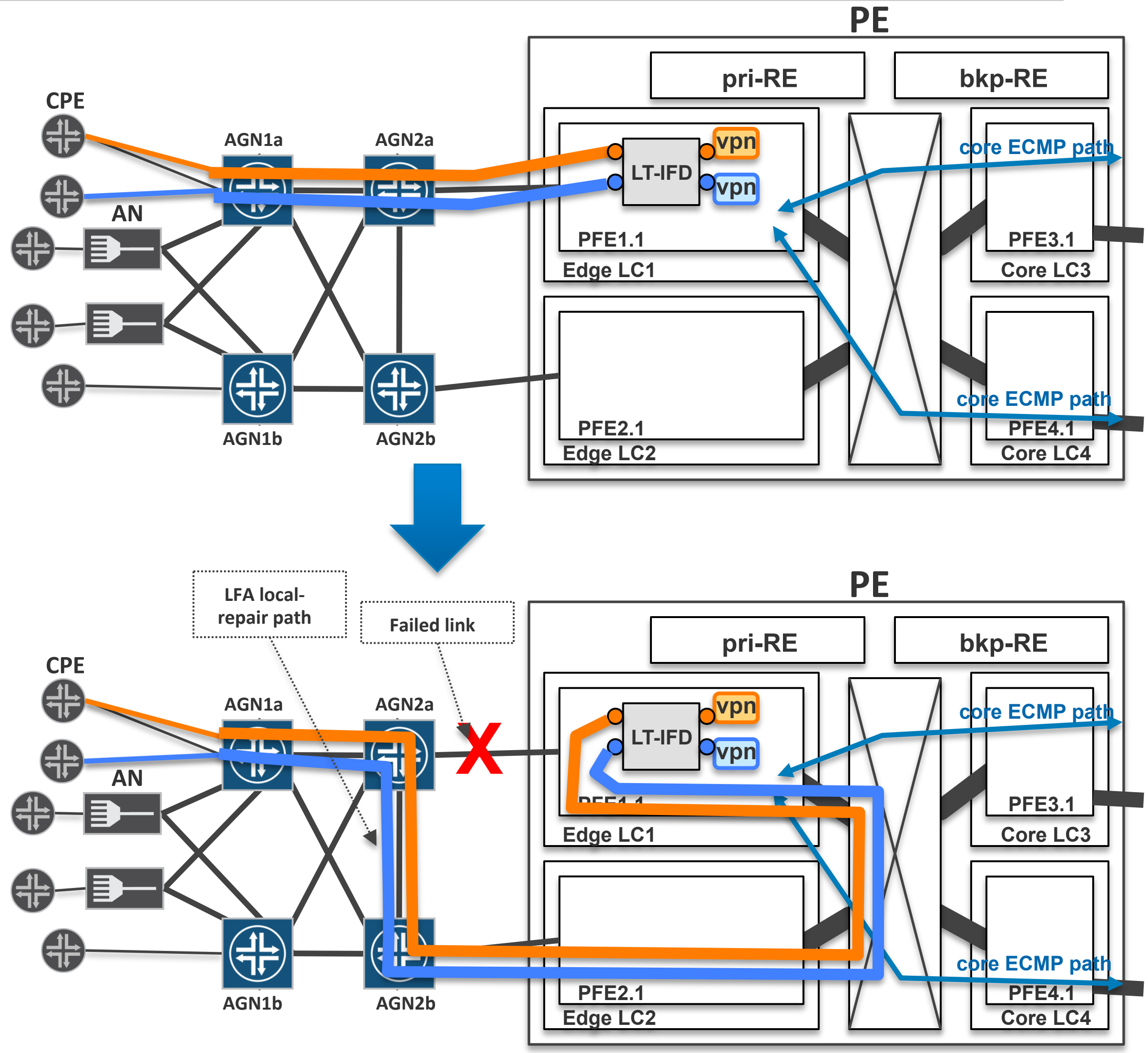
PHT traffic forwarding in case of link failure

- via backup link to PE
- then internally via fabric to PFE hosting associated LT
- apart from local-repair no other impact on service traffic

IP redundancy - once IS-IS converges traffic directed by global-repair

- no impact on access PW traffic

Same scheme applies to the adjacent AGN2 node failure

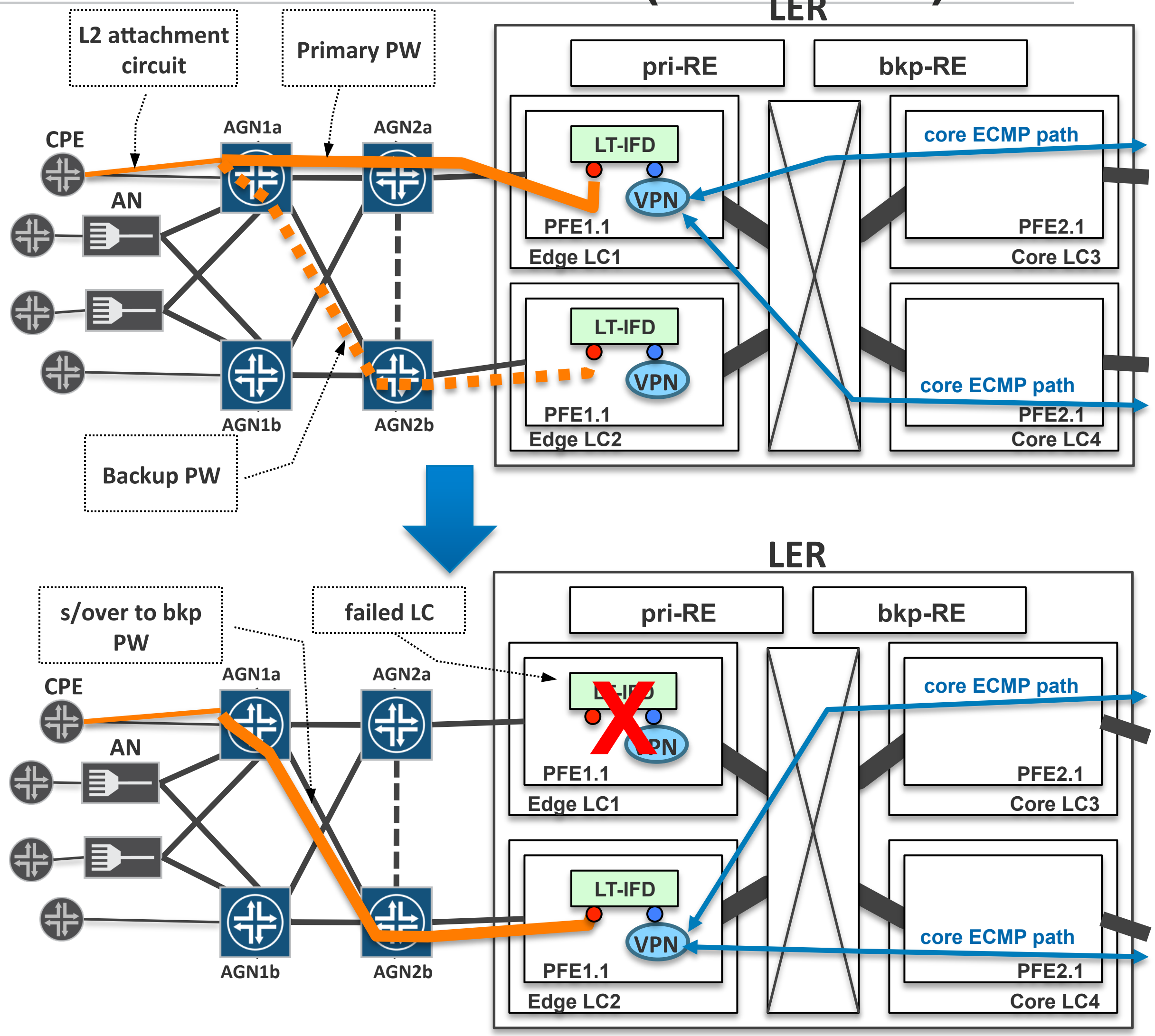


PHT FAILURE HANDLING
LER LINECARD WITH PW REDUNDANCY (ACT/BKP)

Linecard with LT failure handled by pre-provisioned backup LT and backup PW

PHT traffic forwarding in case of LT linecard failure

- via backup PW to backup LT
- restoration time dependent on PW down detection time, activation of backup PW and routing convergence to backup LT-IFD



PHT FAILURE HANDLING
PE EDGE LINECARD FAILURE – PFE REDUNDANCY

Linecard with LT failure handled by pre-programmed redundant LT (rLT) and native local-repair

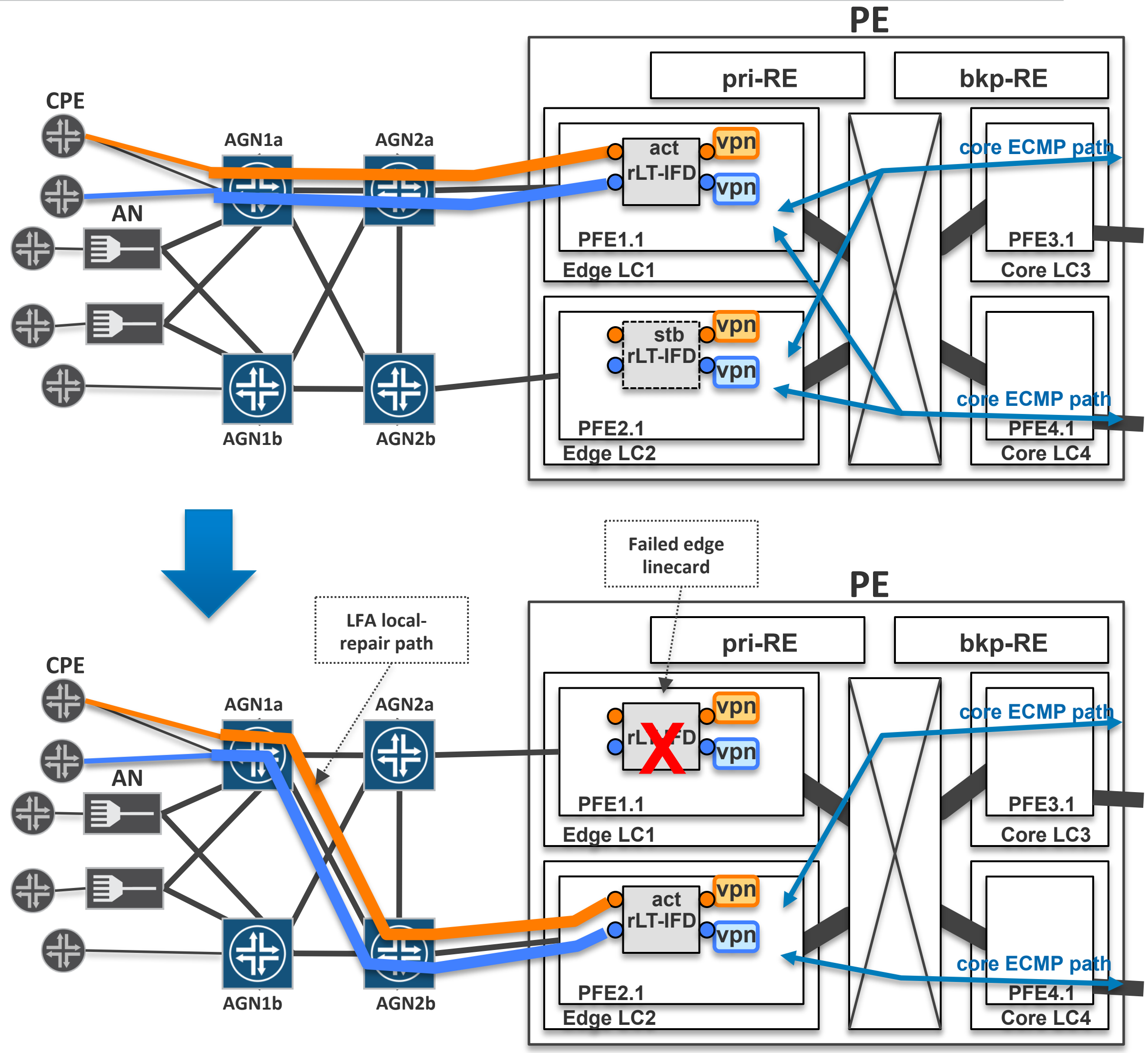
- IS-IS LFA with MPLS LDP
- RSVP TE-FRR
- L2 LAG

PHT traffic forwarding in case of LT linecard failure

- via backup link to PE
- then to rLT-IFD

IP redundancy - once IS-IS converges traffic directed by global-repair

- no impact on service traffic



JUNIPER NETWORKS

E2E RESTORATION
IP/MPLS LOCAL-REPAIR COVERAGE – 100%!

Ingress: CE-PE link, PE node failure

- ECMP, LFA

Transit: PE-P, P-P link, P node failure

- LFA based on IGP/LDP; if no 100% LFA coverage, delta with RSVP-TE
- RSVP-TE FRR

Egress: PE-CE link failure

- BGP PE-CE link local protection

Egress: PE node failure (new)(*)

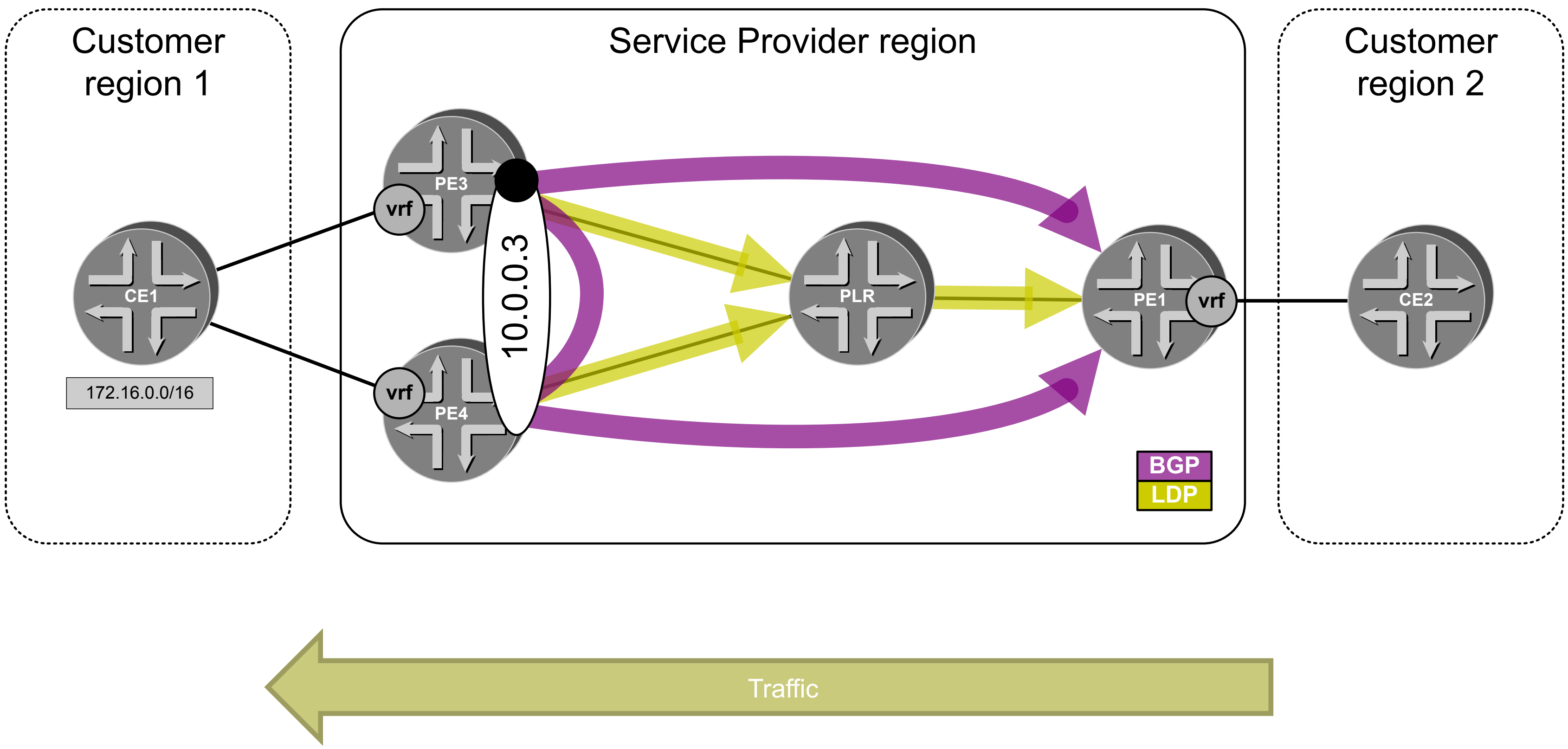
- LSP tailend protection with context label lookup on the backup PE
- Failure repaired locally by adjacent P router using LFA (or TE-FRR)

Packet based networks finally can provide E2E service protection similar to SDH 1:1 protection, regardless of network size and service scale

This provides **network layer failure transparency to service layers**, becoming a major enabler for network consolidation

(*) "High Availability for 2547 VPN Service", Y.Rekhter, MPLS&Ethernet World Congress, Paris 2011.

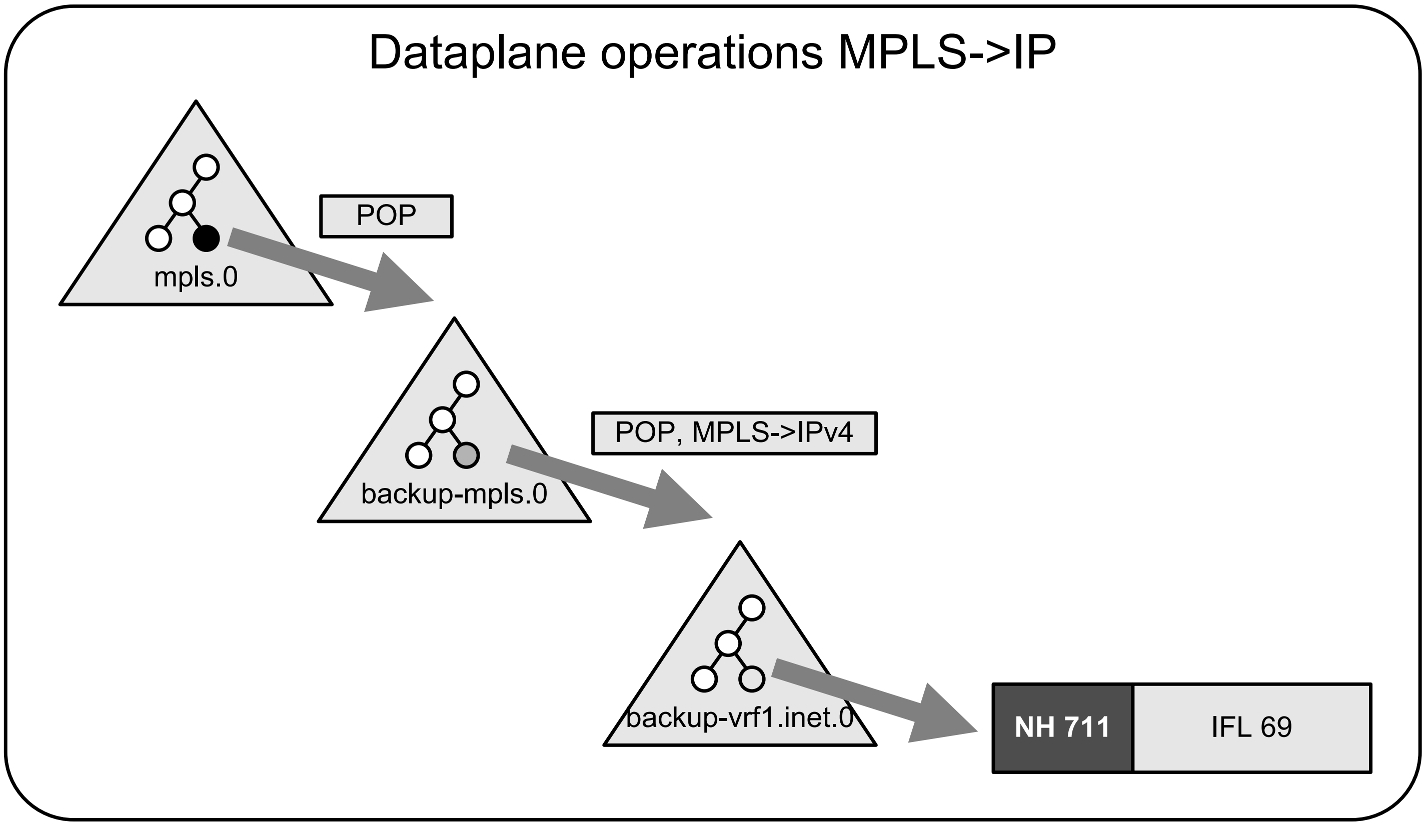
PROTECTING A (SERVICE) TUNNEL ENDPOINT



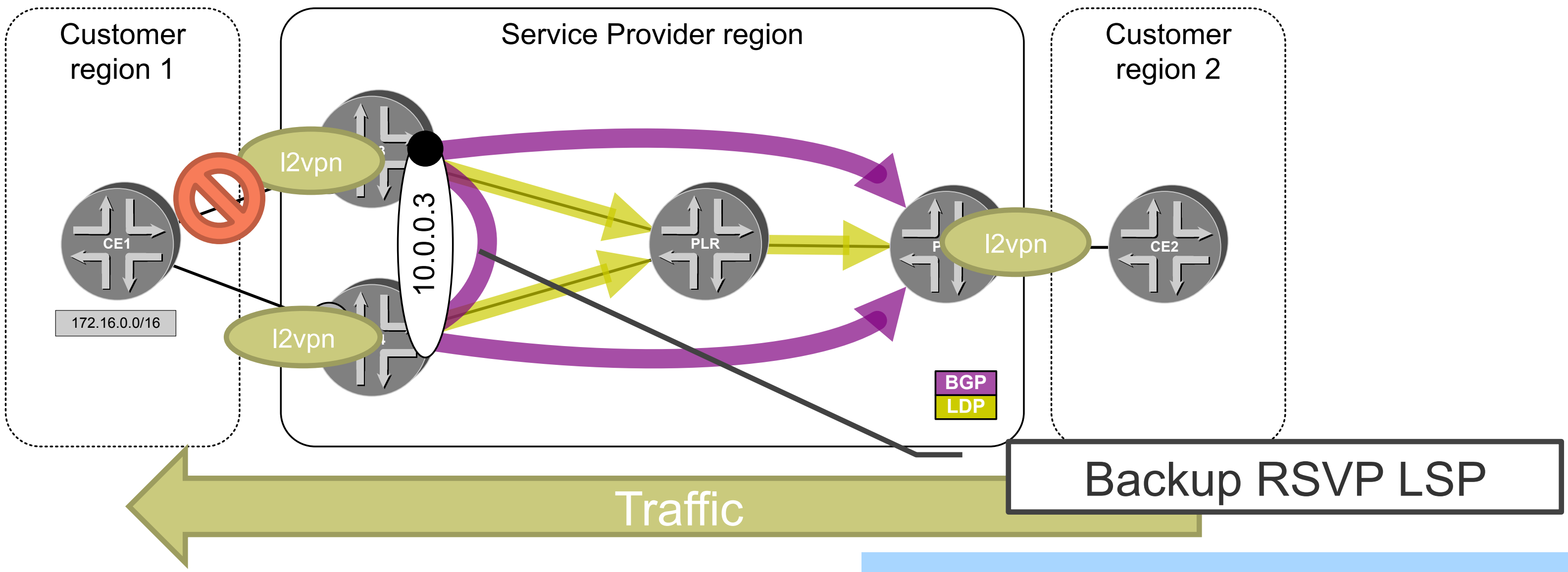
PLR: Point of Local Repair – this is one hop from the point of failure

LSP TAIL END PROTECTION – BACKUP PE LOOKUP

Backup label	Service LSP	IP Payload



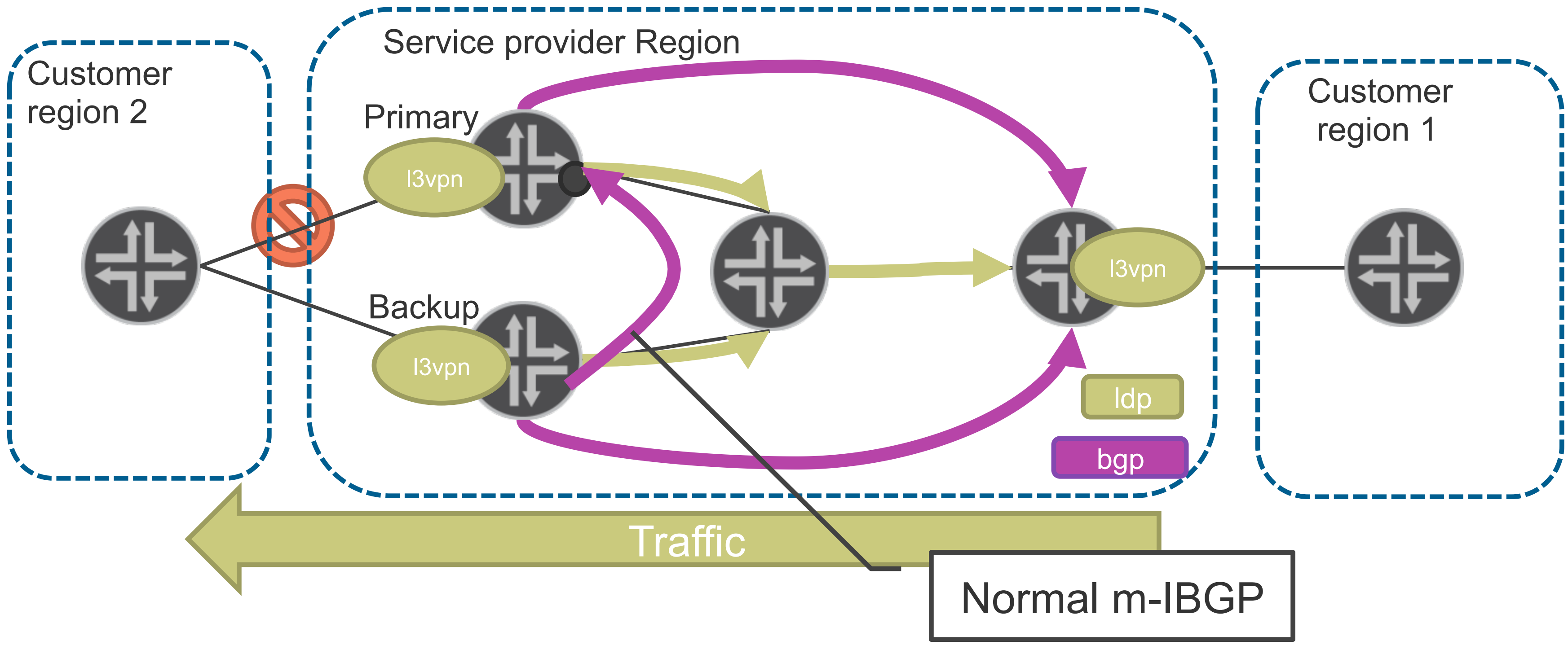
STEP #1 – L2CIRCUIT LINK PROTECTION (AVAILABLE IN 10.4)




```
Configuration on PE1:

protocols {
    l2circuit {
        neighbor 1.1.1.3 {
            interface fe-1/0/1.1001 {
                egress-protection {
                    protector-pe 1.1.1.2
           context-identifier 10.0.0.3;
                }
            }
        }
    }
}
```

```
Configuration on PE2:

protocols {
    l2circuit {
        neighbor 1.1.1.4 {
            interface fe-1/0/2.1001 {
                egress-protection {
                    protected-l2circuit PW31 {
                        ingress-pe 1.1.1.3;
                        egress-pe 1.1.1.1;
                        virtual-circuit-id 13;
                    }
                }
            }
        }
    }
}
```

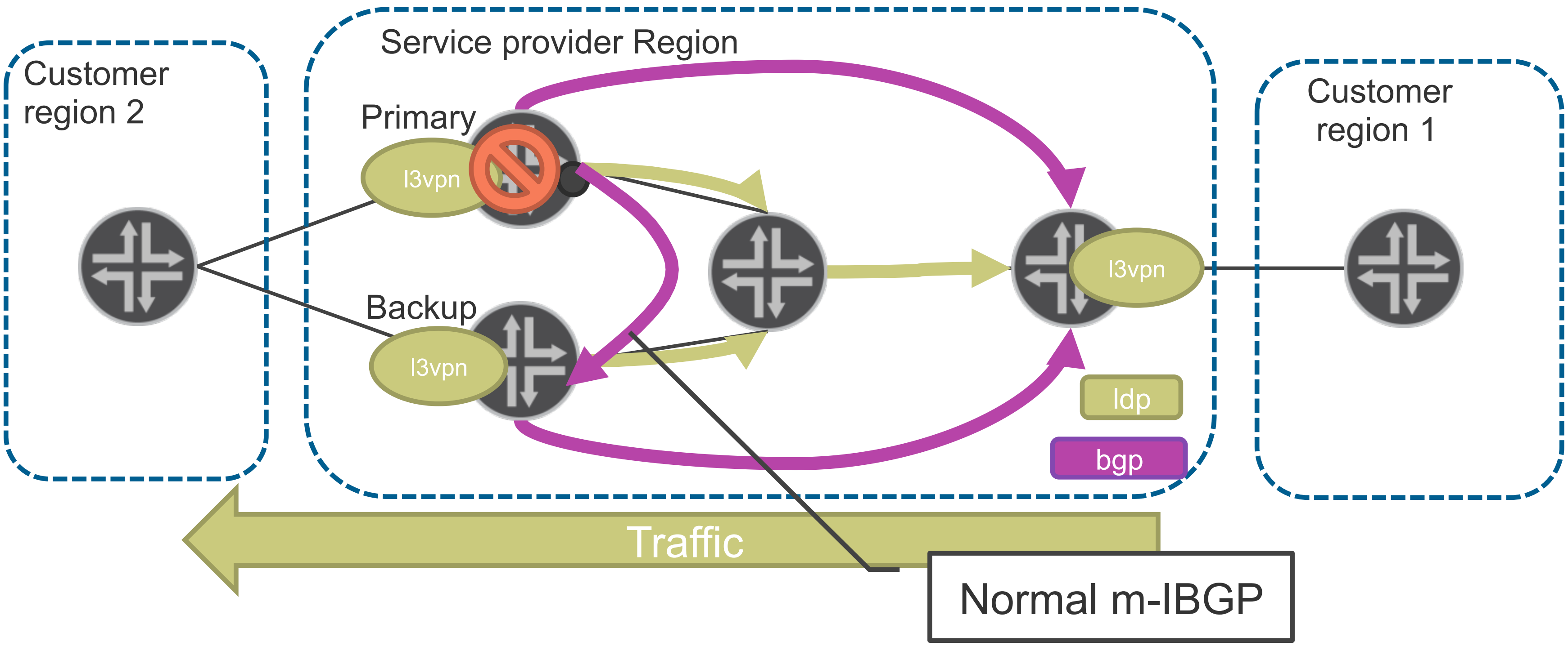
L3VPN PE-CE LINK PROTECTION (11.4)




```
Configuration on PE1:

[edit routing-instances vpn-xy]
routing-options {
    forwarding-table {
        link-protection;
    }
}
```

```
[edit routing-instances vpn-xy]
routing-options {
               l3vpn-composite-nexthop;
    multipath vpn-unequal-cost equal-external-
internal;
}
```

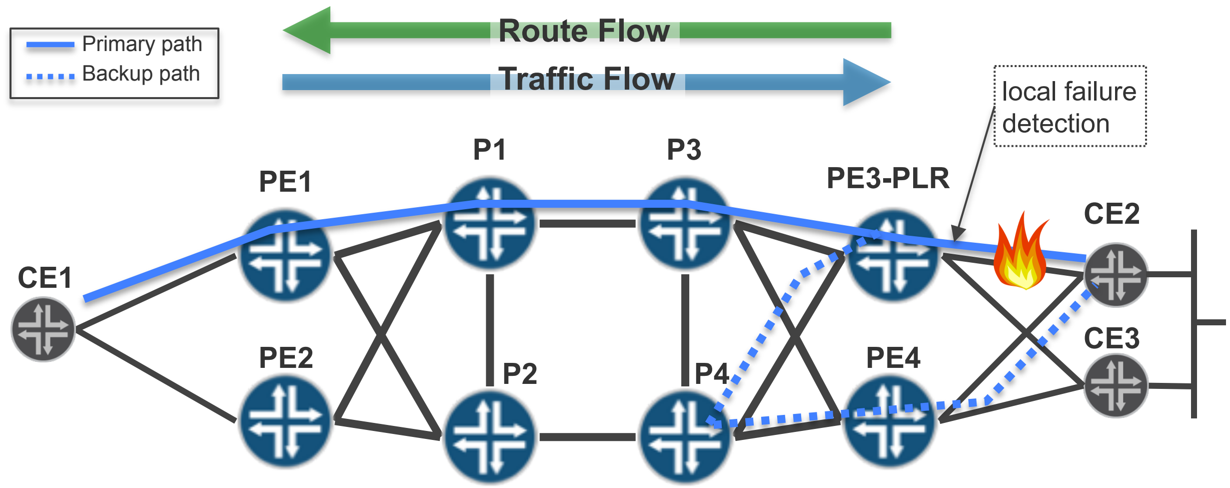
L3VPN PE NODE PROTECTION (TARGET 1H-2012)




```
Configuration on Backup PE:

[edit routing-instances vpn-xy]
interface ge-0/1/0.200 {
    egress-protection {
        context-identifier 10.0.0.3;
    }
}
```

```
protocols {
    bgp {
        group internal
            family [inet-vpn|inet6-vpn|iso-vpn]
            unicast {
                egress-protection {
                    context-identifier 10.0.0.3;
                }
            }
        }
    }
}
```

PE-CE LINK FAILURE LOCAL-REPAIR - SOLUTION



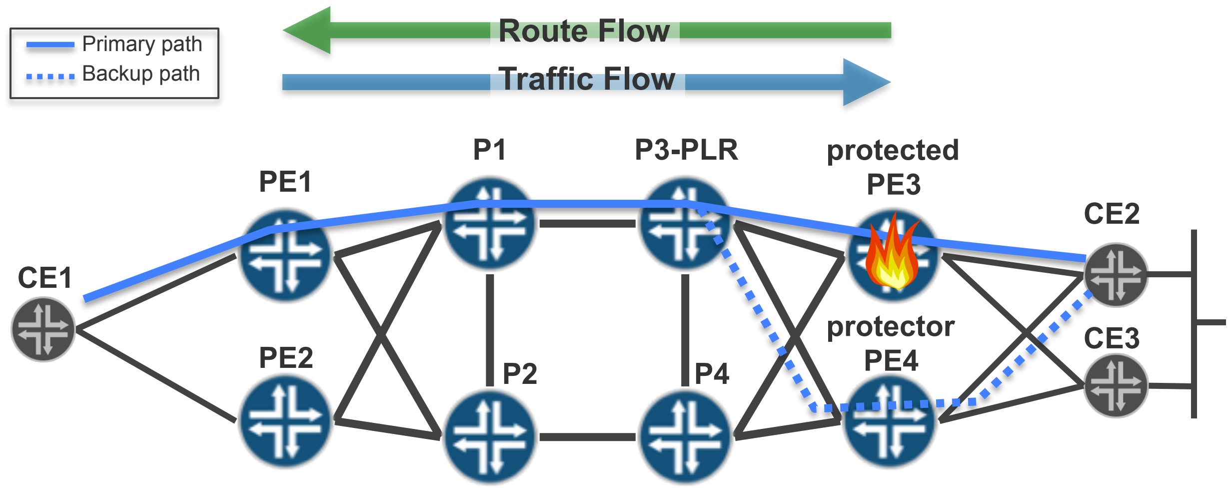
Choices for handling egress PE-CE link failure

- Use PE-CE link protection for any label allocation mode

PE-CE link protection (local-repair)

- Core facing nexthop(s) installed in FIB as alternate (backup) for CE facing routes
- Upon local PE-CE failure FIB in-place modification of CE routes to use alternate nexthop(s), using JUNOS indirect-next-hop
- Support for both BGP uni-path and multi-path

EGRESS PE NODE FAILURE
LOCAL-REPAIR – LSP TAILEND PROTECTION*



- Protector PE4 maintains a “mirror image” of the protected PE3 service label table – a context specific label space identified by a context-id (an IP address) present on both protected and protector PEs
- Protected PE3 “owns” the context-id address, advertising it in the BGP Next_Hop attribute (context-id is never used for control plane peerings)
- In case of protected PE3 failure, P3-PLR diverts the traffic destined to the context-id address to the protector PE4 using TE FRR or IP FRR procedures
- Protector PE4 looks up received packets in the context-specific label table for PE3 (identified by the label associated with PE3 context-id), and forwards packets to the right destination

* draft-minto-2547-egress-node-fast-protection

IPFRR – LFA VS. NOTVIA VS. PQ

LFA is useful and networks are being designed to improve coverage (draft-ietf-rtgwg-lfa-applicability)

- but ***LFA doesn’t guarantee*** 100% coverage.
- Increasing Demand for IP/LDP Fast-Reroute with 100% Coverage

NotVia can guarantee coverage but requires ***significant*** network state

- Research done to reduce it, but nothing sufficiently practical & it’s been years

PQ tunnels (aka remote LFA) cannot guarantee 100% coverage.

Requires explicit tunnels

requires targeted LDP sessions for FEC label bindings.

Topologies change due to failures and growth.

- 100% Coverage gives ***protection always*** –
- not just until the first maintenance event.

=> Increasing Requirement and Demand for IP/LDP Fast-Reroute with 100% Coverage everywhere