Implementing Seamless MPLS

Maciek Konstantynowicz, maciek@juniper.net

Internet Engineering, Juniper Networks

March 2011

SP Priorities

Minimize recurring operational costs

- Simplify provisioning, management and troubleshooting
- In many cases IT Systems struggling with keeping up with next-gen network evolution

Improve service velocity

- Reduce time to service, lower new service intro cost
- Improve effectiveness of service delivery
- Differentiate service offering in the market place

No compromise on service quality

- E2E restoration with transparency of network failures to service layer
- Appropriate QoS SLAs and quality of experience per service type

… Must factor in bandwidth and service mix uncertainties, not an easy task

More about service trends..

Most edge services today are delivered at the boundary of metro and backbone – in the IP POP

- **Some services are even more centralized incl. mobile and Data Center**

However service dynamics are changing rapidly due to following factors

- Compute power of end devices, incl. increased number of mobile devices/tablets
- Increasing network access speeds with fiber access
- High volume video content, both walled garden and over-the-top

De-centralization of selected services is becoming reality

- Residential services, location-based services, video caching
- But some services likely to stay more centralized (e.g. business L3VPN)

Target packet network architecture must enable this seamlessly

Translating into Requirements for Converged Packet Network

A single converged packet network for delivery of all services

- Wireline, mobile, residential, business, wholesale

Support for large network and services scale

- A single packet network spanning core, edge, aggregation and access

Deterministic availability SLAs regardless of scale

- Fast service restoration and transparency of network failures to service layer are key drivers for network consolidation
- Enables migration away from SONET/SDH

Greatly simplified service introduction, delivery and operations

- Flexible topological placement of service delivery points
- Minimized number of service provisioning points

A Word on packets vs. circuits

Common perception

- MPLS ports expensive, OTN ports cheap

Reality check

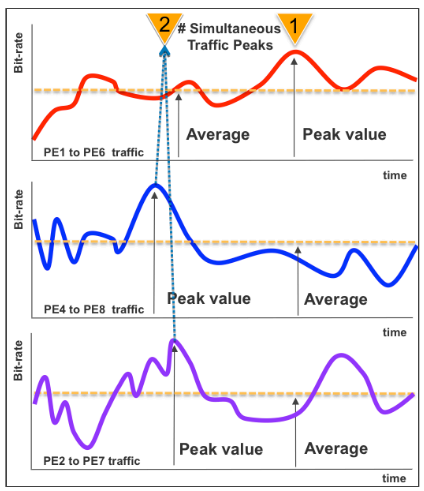
- Depends on service demands and topology, quantifying statistical multiplexing gain is key

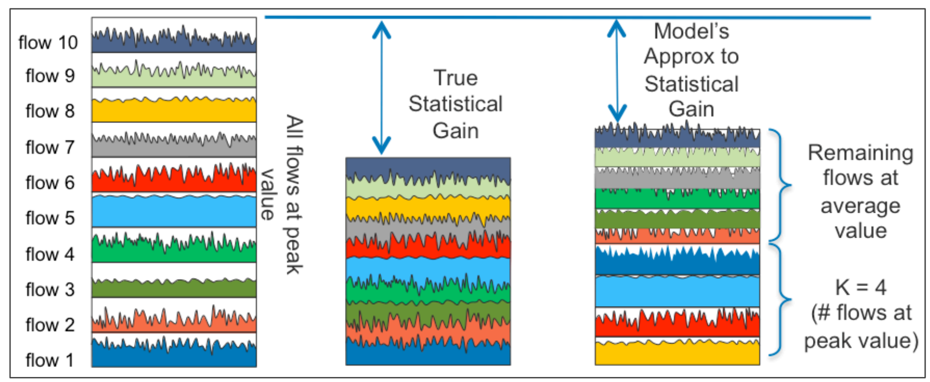
Need network cost modeling for multi-layer optimization

- New modeling approach enables cost optimization based on realistic packet flow approximations[(1)(2)]
  - Packet demands modeled more accurately taking into account avg BW, peak BW and concurrent peaks
  - Cost optimization problem formulated as a linear optimization using MPLS and OTN ports and devices
- Result: a provable optimal solution, i.e., there is a guarantee that no better solution exists

A new model for packet traffic:

- Add averages, but consider some number of simultaneous peaks
- Describe packet demands as function of three parameters:
  - **B** = average bandwidth
  - **α** = peak:average ratio
  - **K** = # of simultaneous peaks



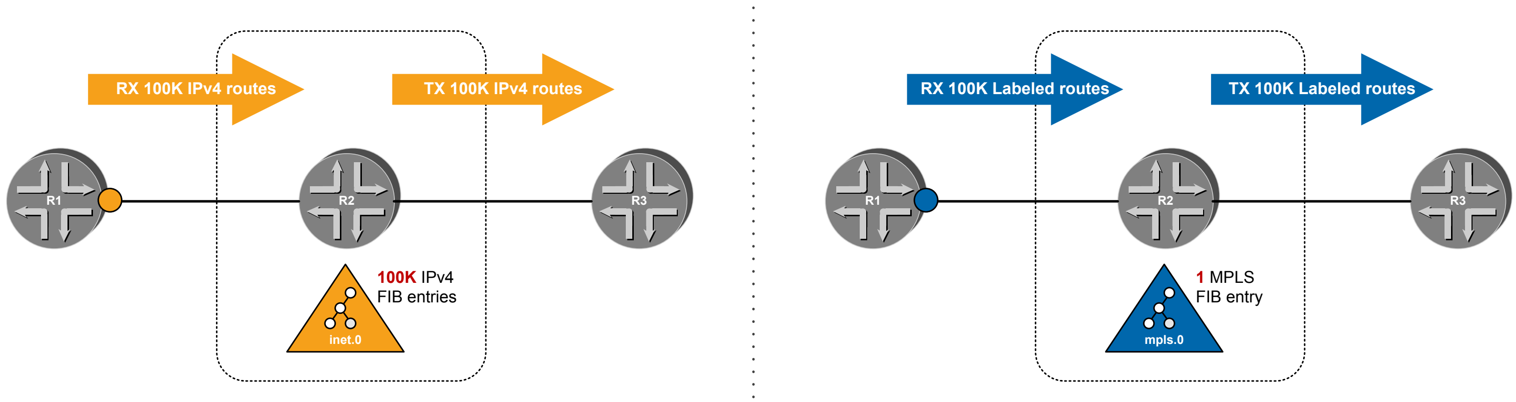


References:
(1) "MPLS vs. OTN", K.Kompella, Juniper, MPLS & Ethernet WC Paris, February 2011.
(2) "Robust optimization models for networks with statistical multiplexing", P.Belotti et al, Clemson University, September 2010.

JUNIPER NETWORKS

MPLS Overview: Why is MPLS useful at all ?

Separation between Control Plane and Data plane



Unified Data plane

- ***"Magic Carpet"*** for Services

Support for arbitrary Hierarchy

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

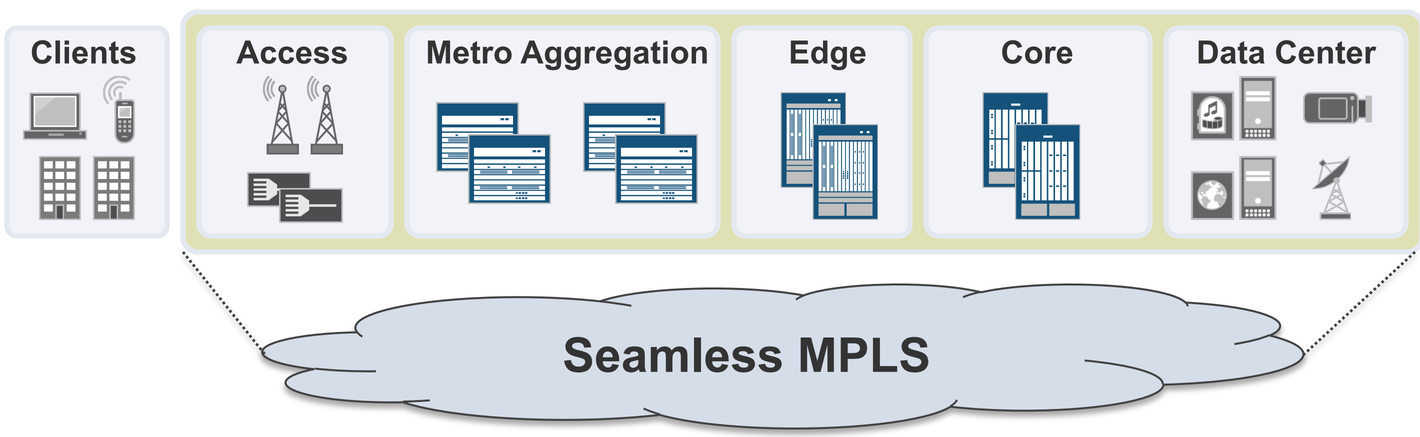
Implementation: Seamless MPLS Foundation for One Converged network

Network Scale and End-to-End service restoration

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

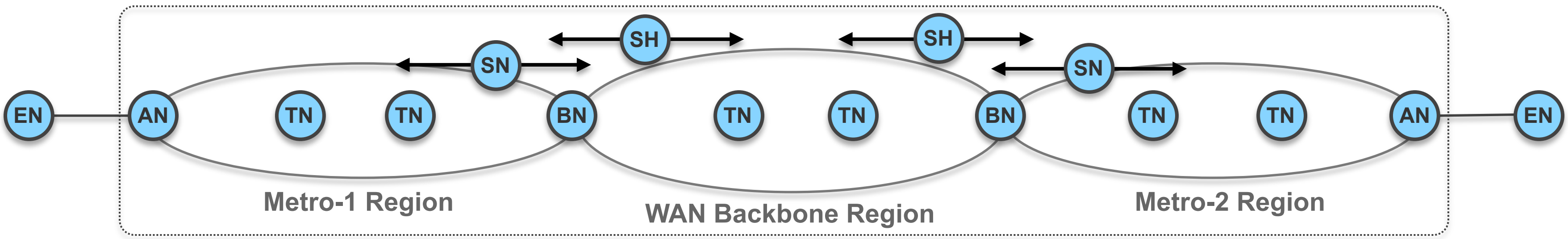
Decoupled network and service architectures

- 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

Juniper's 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

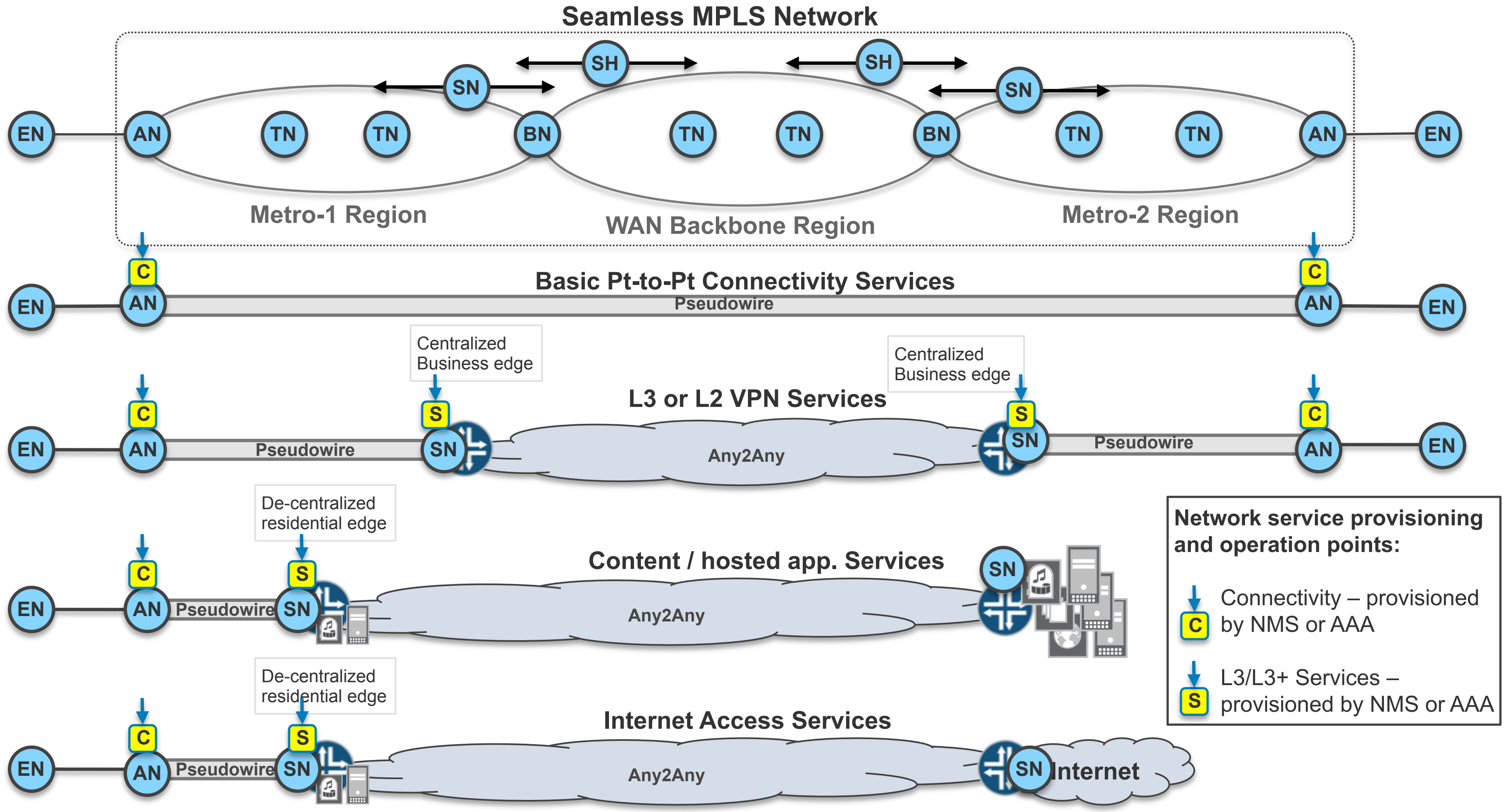
Regions

- 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)

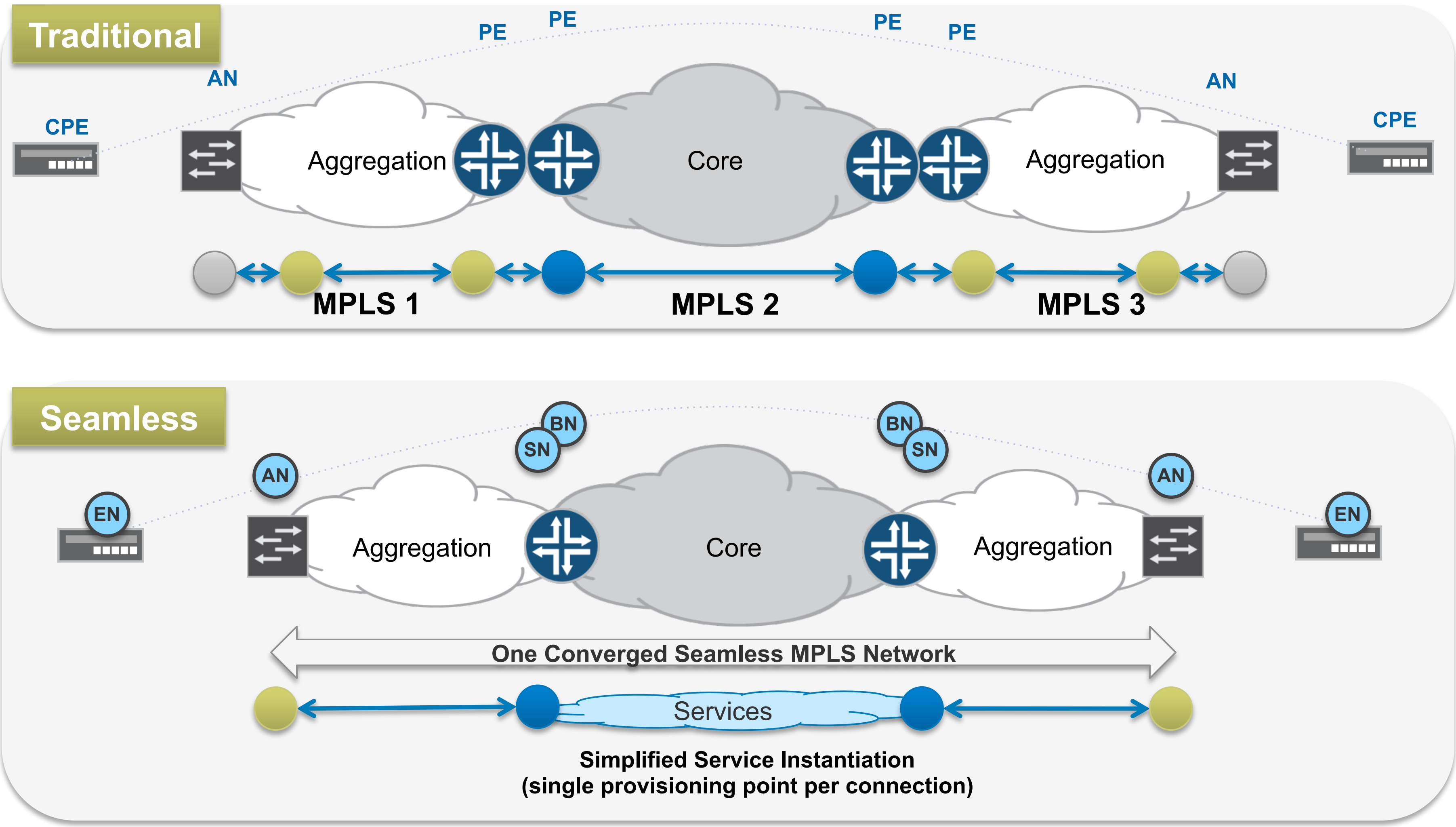
Decoupled architectures

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

Juniper's Seamless MPLS Architecture Connectivity and Services Blueprint



Seamless MPLS
Simplified Service Delivery



Juniper Seamless MPLS
Service and Network Architecture

Requirements addressed across the three main architectural dimensions

(1) Scale – enables 100,000s of devices in ONE PSN network

- Large network scale via MPLS LSP hierarchy and robust network protocol stack (IGP, BGP)
- No service dependency whatsoever – all packet services supported
- Low-cost/low-end access devices accommodated natively without adding complexity (MPLS labels on demand)

(2) E2E service restoration – enables sub-50ms recovery from any event

- Service restoration made independent of scale, services and failure types
- Achieved with full coverage of local-repair mechanisms for sub-50ms restoration
- Deterministic for any failure domain size / radius

(3) Decoupled network and service architectures

- Flexible topological placement of services enabled via MPLS Pseudowire Termination into Services
- E2E virtualization of network service delivery with tight integration of Ethernet, IP and MPLS
- Minimized number of provisioning points, simplifying service delivery and IT systems(!)

Seamless MPLS – EU Use Case
Network Scale

Design

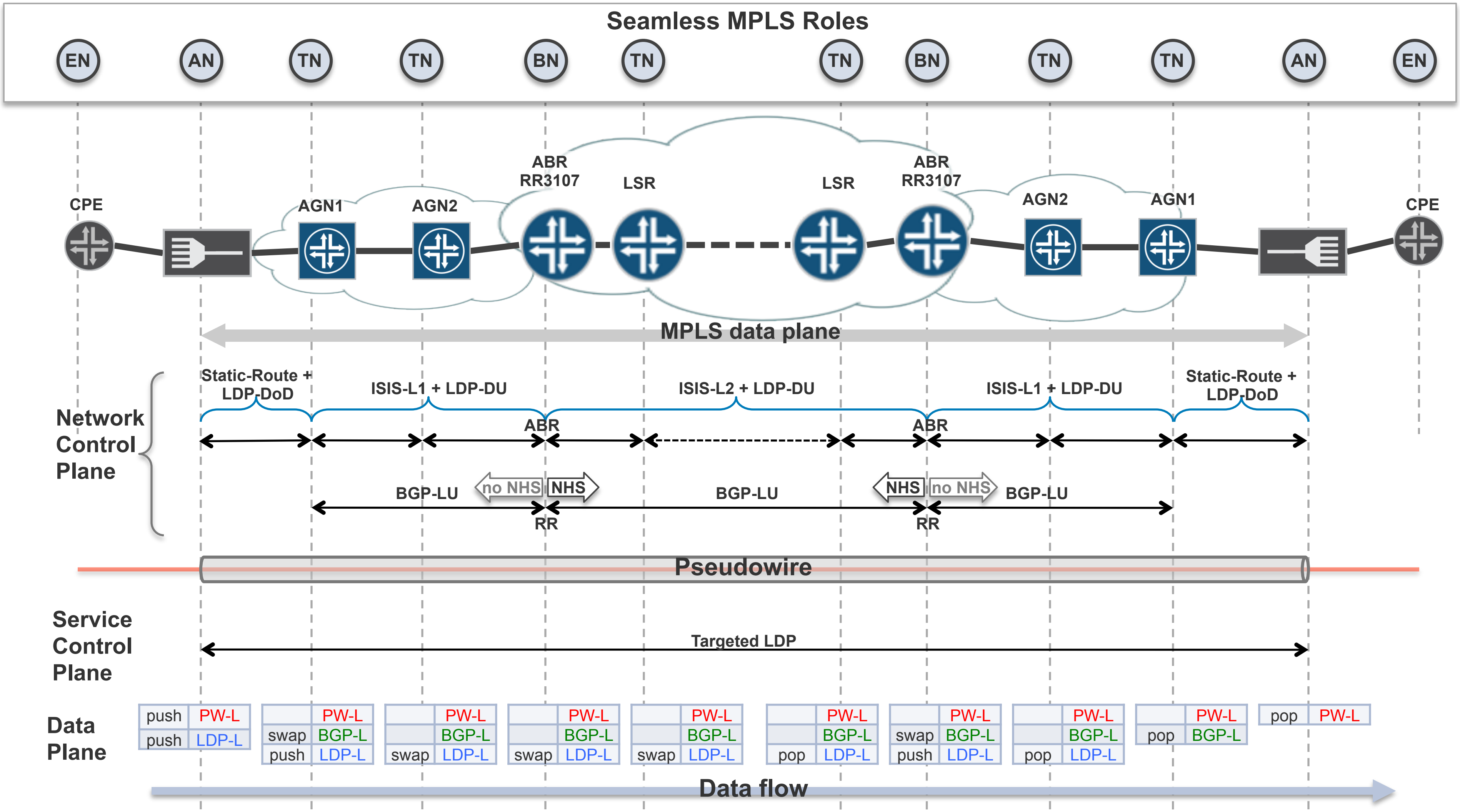
- Split the network into regions: access, metro/aggregation, edge, core
- Single IGP with areas per metro/edge and core regions
- Hierarchical LSPs to enable e2e LSP signaling across all regions
- IGP + LDP for intra-domain transport LSP signaling
  - RSVP-TE as alternative
- BGP labeled unicast for 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 base VPWS service)
- A simple MPLS stack on access devices (static routes, LDP DoD)

Seamless MPLS – EU Use Case
IP/MPLS Network Infrastructure

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



Juniper NETWORKS

Scale Enablers
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

- But not widely available in IP/MPLS routers apart from MPLS over ATM/FR
- This is being fixed now ☺

LDP DoD – Seamless MPLS Use Case Configuration and Operation

① **AN:** provisioned static routes

② **AGN1:** provisioned static routes

③ **AGN1:** statics redistributed into IGP (optional)

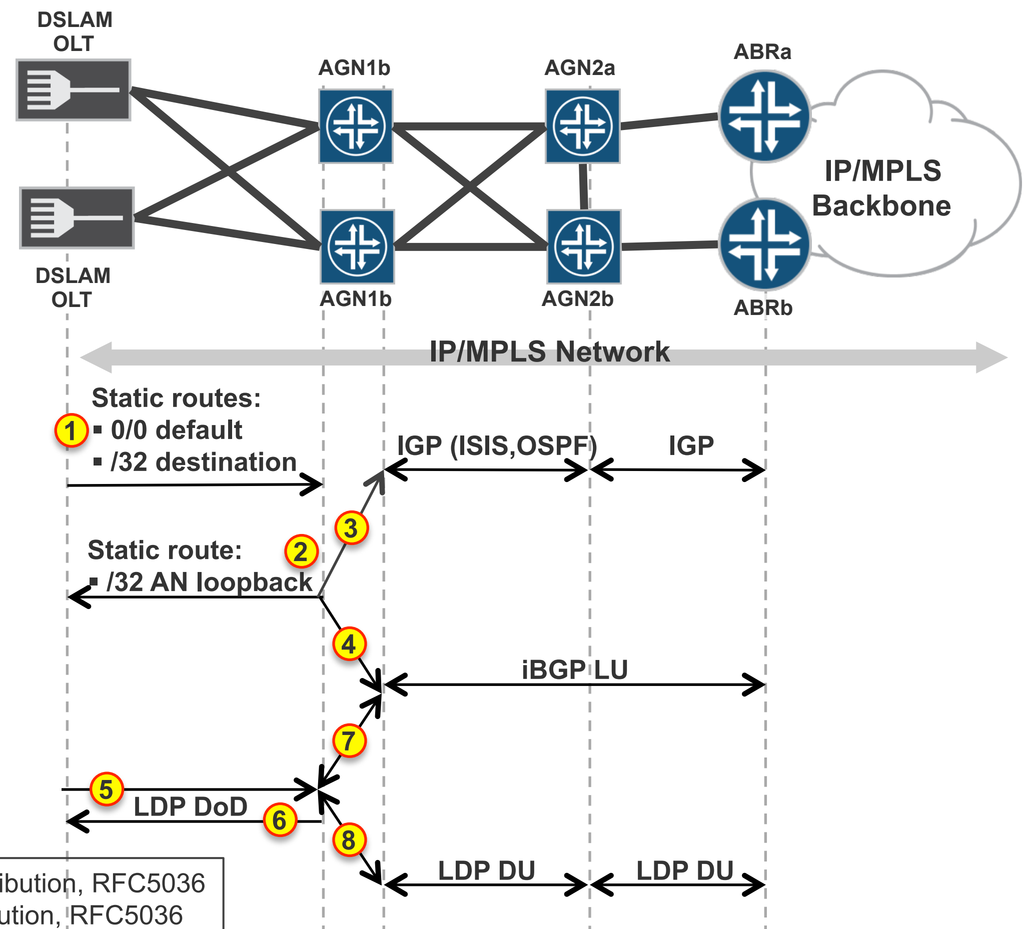
④ **AGN1:** statics redistributed into BGP-LU

⑤ **AN:** LDP DoD lbl mapping requests for FECs associated with /32 static routes and configured services using /32 routes matching default route(*)

⑥ **AGN1:** LDP DoD lbl mapping requests for static route /32 FECs

⑦ **AGN1:** AN loopbacks advertised in iBGP LU

⑧ **AGN1:** if (3) AN loopbacks advertised in LDP DU



LDP DoD – Label Distribution Protocol, Downstream on Demand distribution, RFC5036
LDP DU – Label Distribution Protocol, Downstream Unsolicited distribution, RFC5036
BGP LU – Border Gateway Protocol, Label Unicast extensions, RFC3107

(*) Requires LDP support for longest match prefix in RIB (in addition to the exact match) as per RFC5283.

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

Seamless MPLS – EU Use Case
E2E Service Restoration

Design

- IPFRR/LFA for local-repair of transit MPLS link and node failures
  - TE FRR as alternative to LFA
- LSP tail-end protection for egress PE node failures (IP, L3VPN, L2VPN, BGP-LU, RR-NHS)
- Optimized global-repair as fall-back if local-repair not feasible (e.g. no LFA cover)
  - Note: LFA cover can be extended with RSVP-TE
- BGP PE-CE link local-repair protection for BGP edge link failures (IP, L3VPN, L2VPN, BGP3107)

Properties

- Local-repair for all PE access links, PE and P nodes
- Local-repair for all PE/P transit links, topology independent (albeit certain topologies may introduce increased complexity e.g. RSVP-TE if no LFA coverage)
- E2E restoration in O(50ms) achievable, regardless of network and service scale

End-to-End Restoration
Local vs. Global Repair

Local-repair

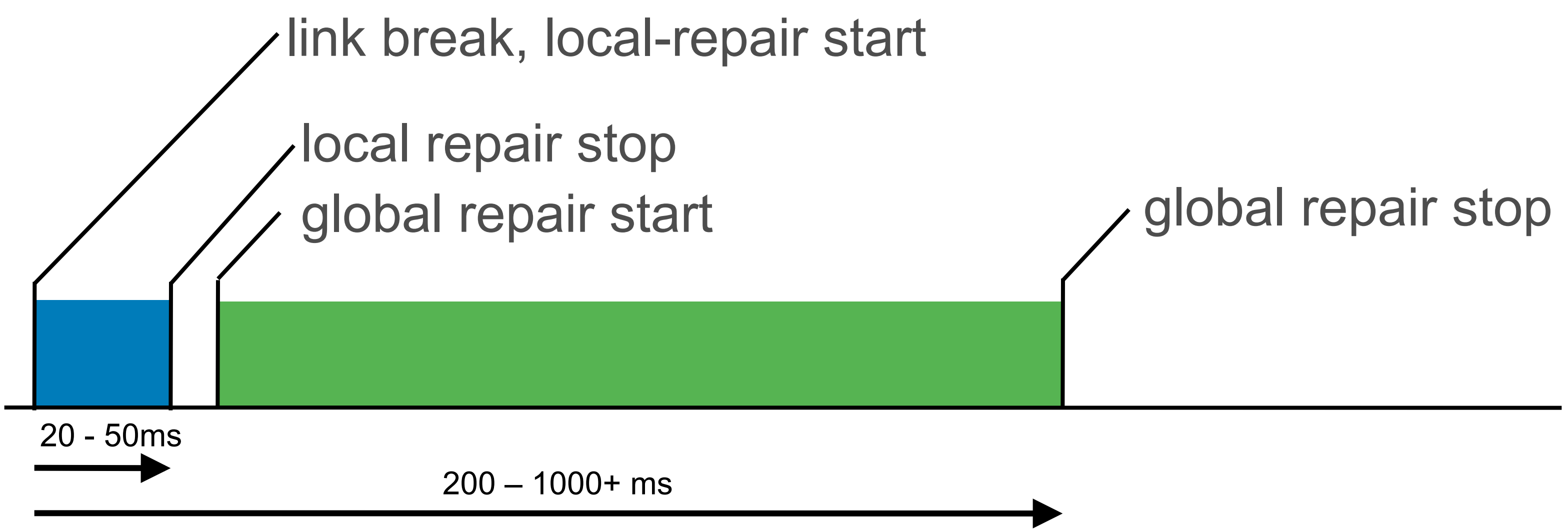
- Based on the pre-computed local backup forwarding state - provides sub-50msec restoration

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"



End-to-End Restoration
IP/MPLS Local-Repair Coverage – 100% Achieved!

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.

Seamless MPLS – EU Use Case
Decoupled Network and Service Architectures

Design

- Use MPLS transport pseudowires (PW) to virtualize access for L2 and L3 services
- Service Node (SN e.g. PE, BNG) to support a PW Headend access interface with all required data plane and control plane functions (HQoS, security, OAM, PE-CE routing)
- Combined SN and TN - enable co-existence of IP/MPLS Service and Transport functions on the same physical node and the same physical links

Properties

- L1/L2 access interface on SN replaced with PW Headend virtual interface
- Access side SN reachability govern by IP/MPLS – decoupled from L1/L2 interfaces
- Reduced number of access provisioning points vs. present mode of operation
- SN support for all L2 and L3 services with appropriate scale
- E2E service restoration with local-repair for SN failures incl. access PW, node, transit links

Pseudowire Termination into L3 services

Flexible topological L3 edge placement

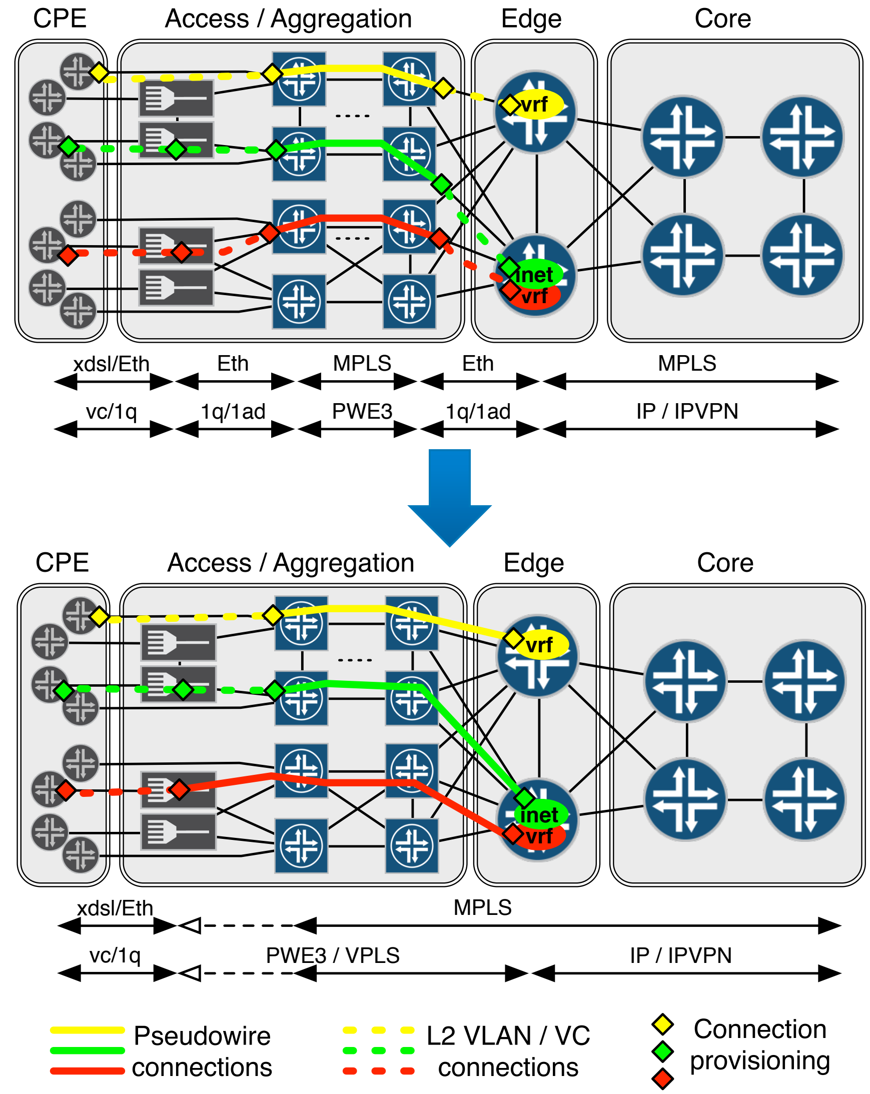
- Virtualized pseudowire access interface enables L3 edge insertion anywhere within the MPLS cloud
- Services can be placed optimally based on network, services and operation economics – no impact on the network architecture

Simplify network operations

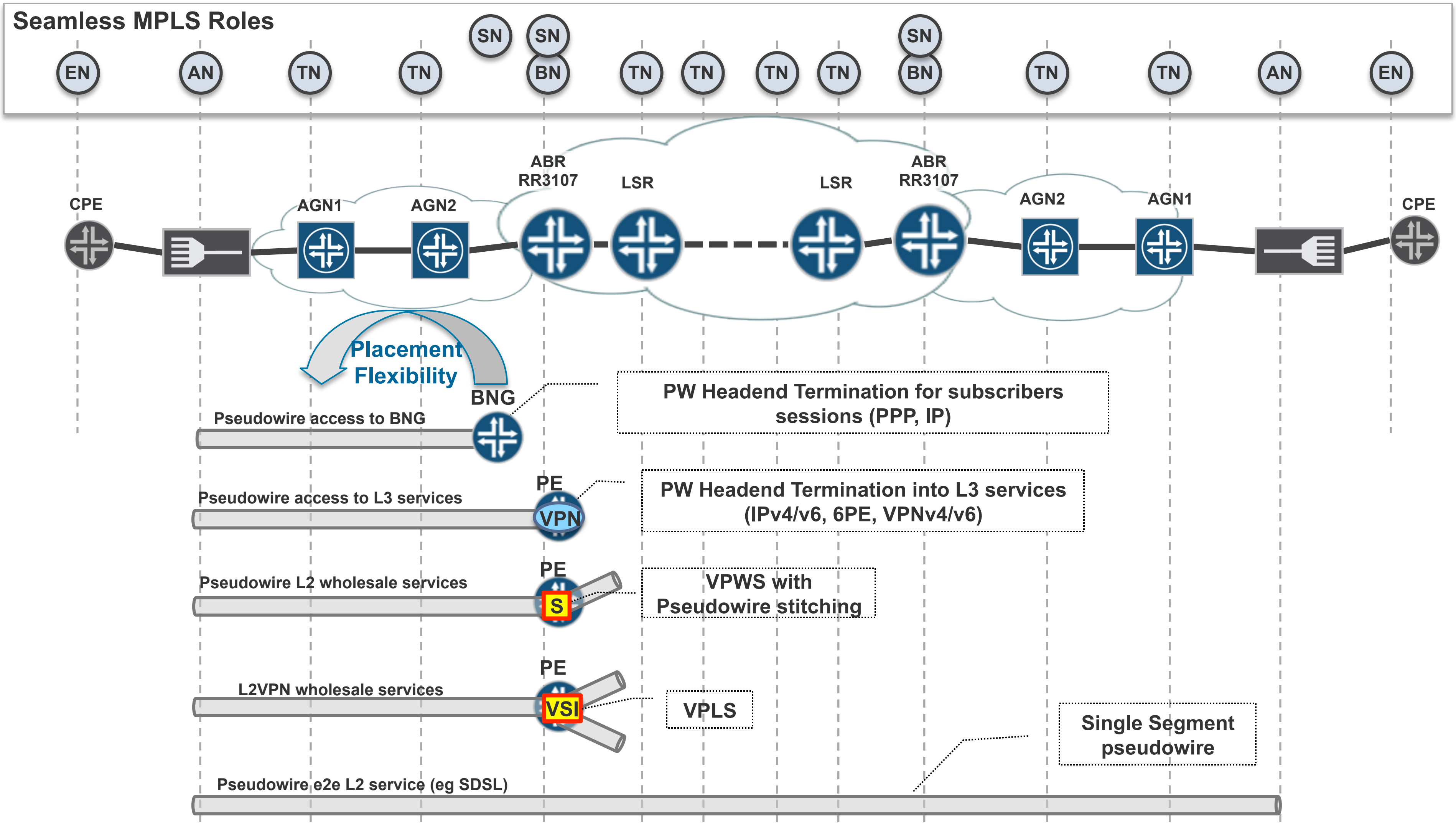
- End-to-end MPLS across Aggregation and Core domains
- No Ethernet L2 interconnect with associated complexity (provision, assure, protect)
- No complex L2/L3 redundancy schemes (ie no MC-LAG)
- Simplified provisioning

Unify L3 edge

- Standardize on pseudowire access with Ethernet and IP encapsulations only
- Support legacy access (ATM, FR, SONET/SDH) thru IP i/working function distributed into access/aggr
- Results in a uniform L3 Edge for all access



Seamless MPLS – EU Use Case
Service Architecture and Connectivity



In Conclusion…

Seamless MPLS approach addresses all key requirements for converged packet network design

- Support for all packet services across fixed, mobile, business, residential, wholesale
- Support for large scale incl. high number of low end access devices
- E2E fast restoration sub-50msec for all network failures
- Simplified service delivery with flexible topological placement

References

- draft-leymann-mpls-seamless-mpls-03, N.Leymann et al, March 2011.
- “High Availability for 2547 VPN Service”, Y.Rekhter, MPLS&Ethernet WC, Paris 2011.
- “LSP Tailend Proteciton”, H.Gredler,
- “Seamless MPLS”, K.Kompella, MPLS WC 2009.
- “MPLS vs. OTN”, K.Kompella, MPLS & Ethernet WC Paris, February 2011.
- “Robust optimization models for networks with statistical multiplexing”, P.Belotti et al, Clemson University, September 2010.

everywhere