Building a Fixed-Line Broadband Network in 2019

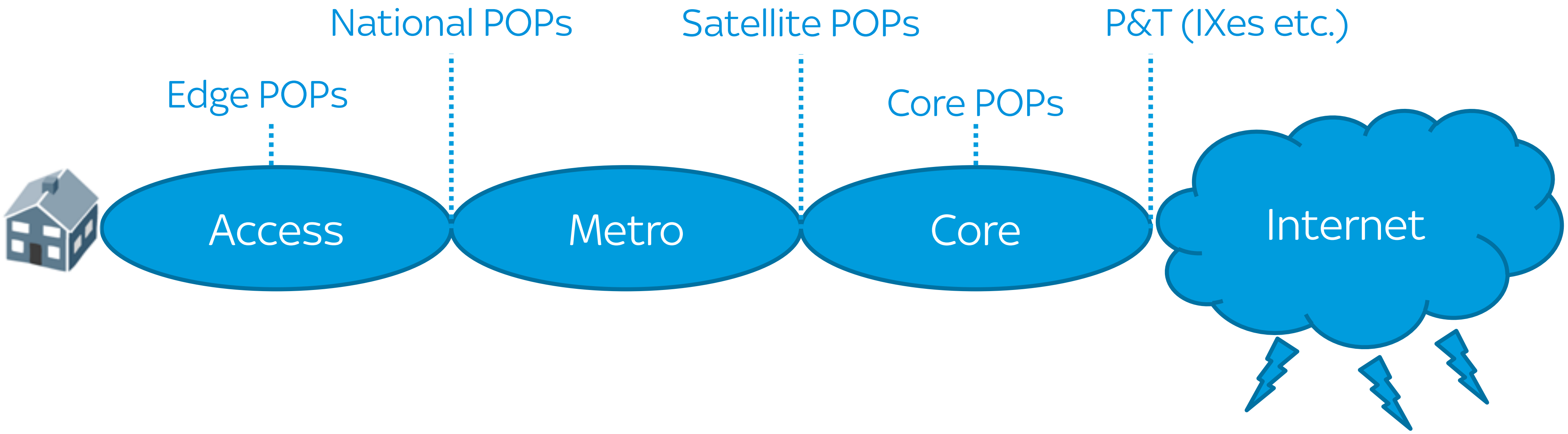
Westworld 2

sky atlantic

NetLdn – 11th July 2019

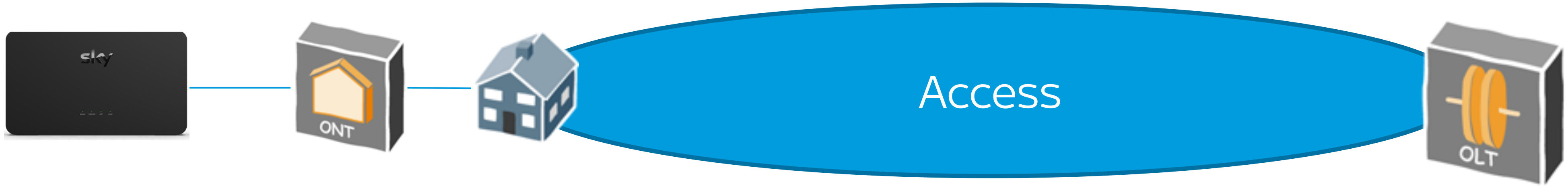
High-Level

End-to-End Network



Access

Fibre to the Home with
Gigabit Passive Optical Network (GPON)

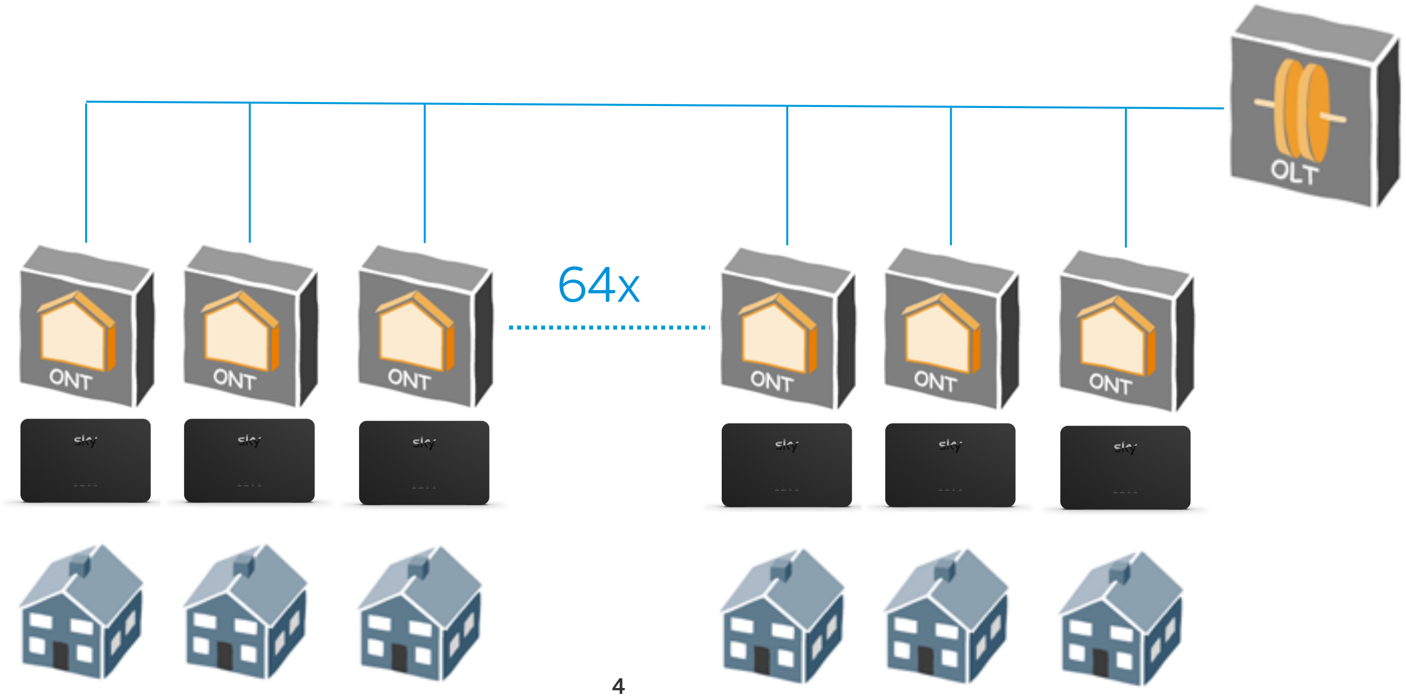


- Optical Line Terminal (OLT) @ Edge POP
- Optical Network Terminal (ONT) @ Customer Premise
    - 2 box solution with Sky Hub 4 CPE router

Access

GPON

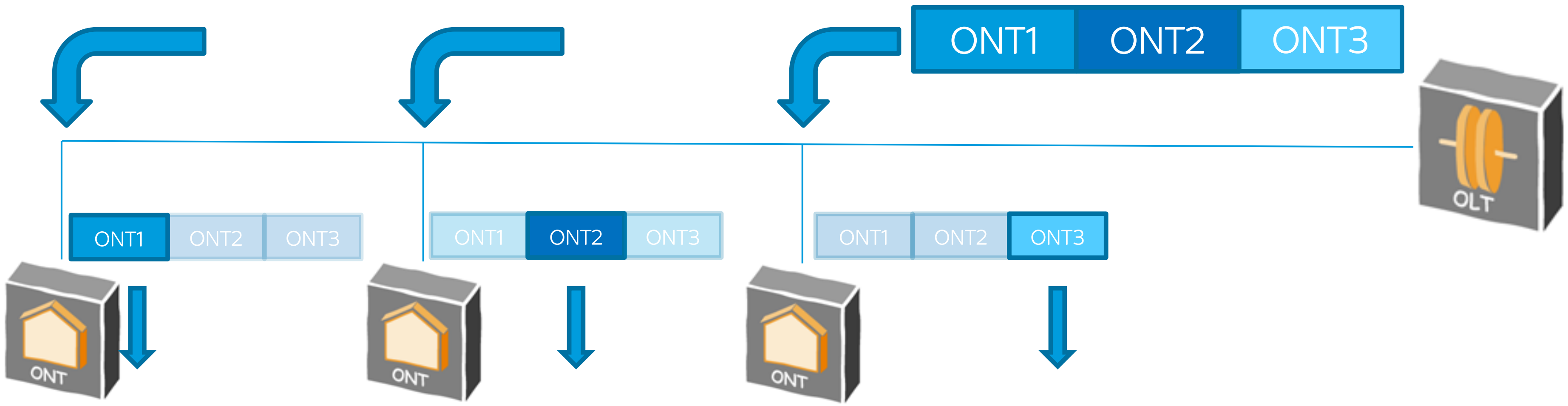
- Single-strand single-mode fibre
    - BiDi Optics
- Point-to-Multipoint using passive splitters
    - Commonly @ 32:1 or 64:1 but up to 128:1
- 2.5Gbps down
- 1.25Gbps up



Access

GPON - Downstream

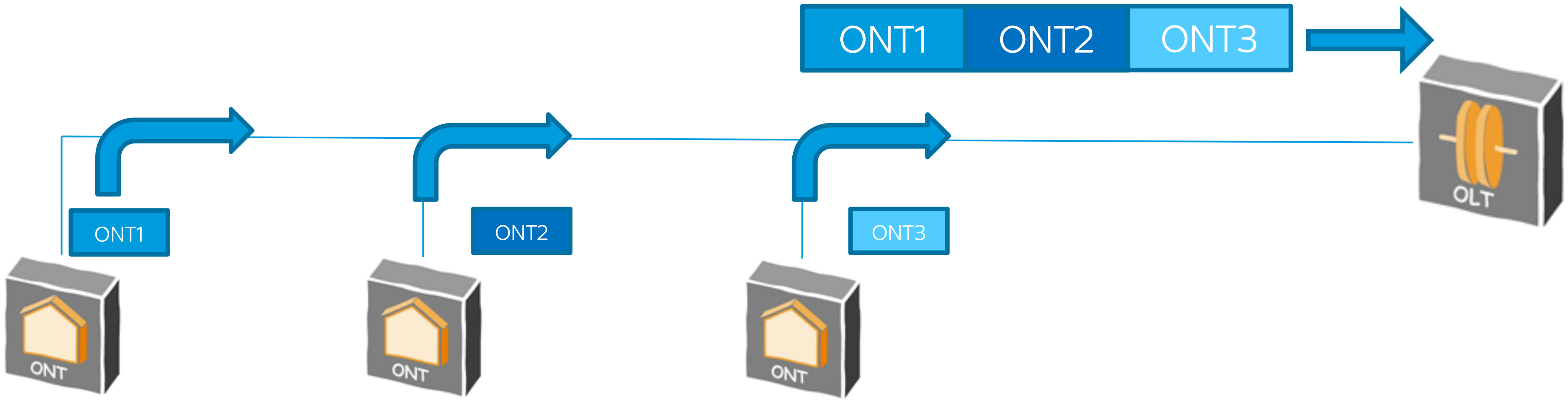
- ~1490nm Downstream
- Broadcast to all ONTs
- AES Encrypted



Access

GPON - Upstream

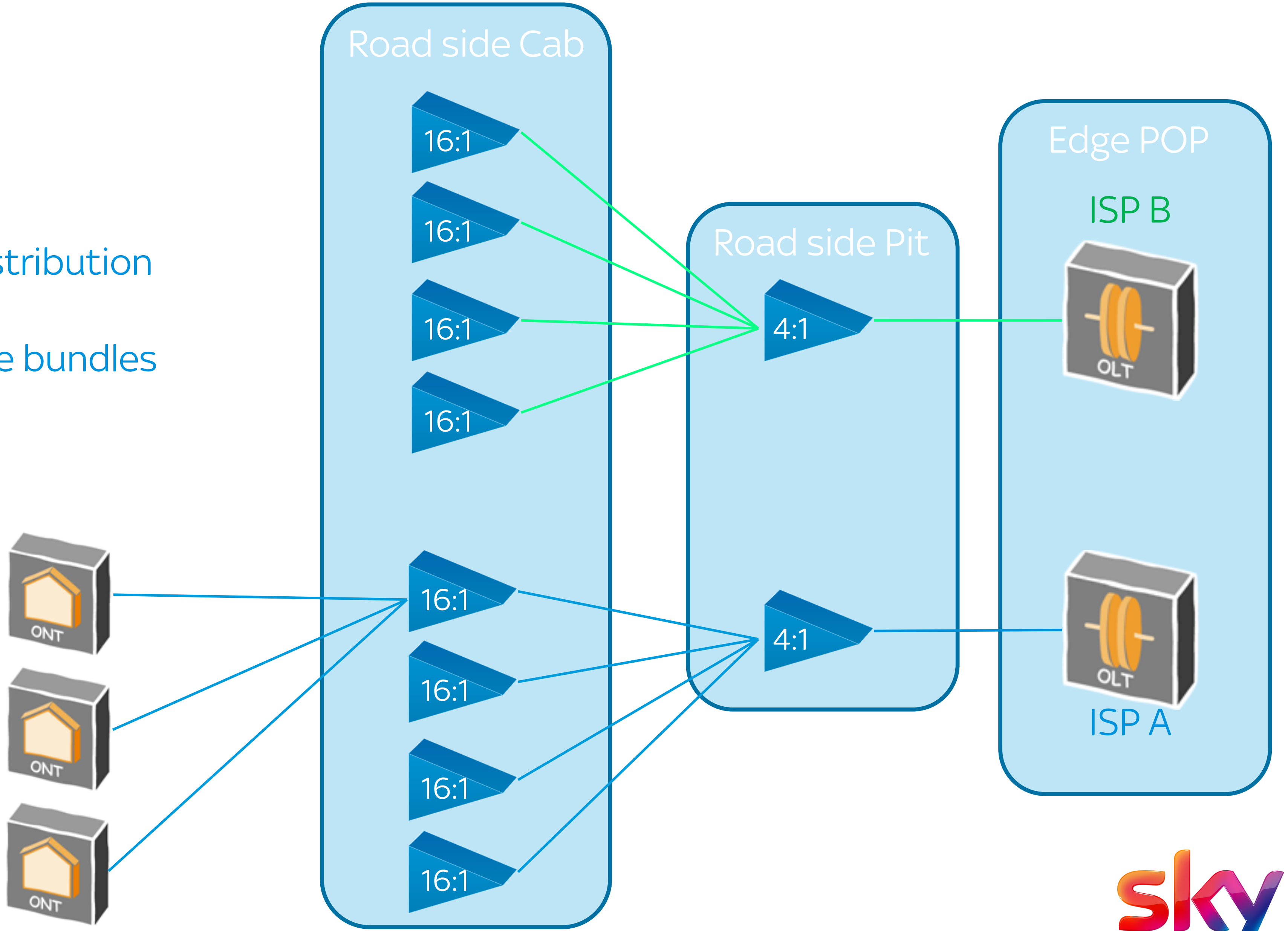
- ~1310nm Upstream
- OLTs “range” to find each ONT’s distance from OLT
- Time Division Multiplexed (TDM)
- Cells then converge on the wire without collision



Access

Unbundled GPON

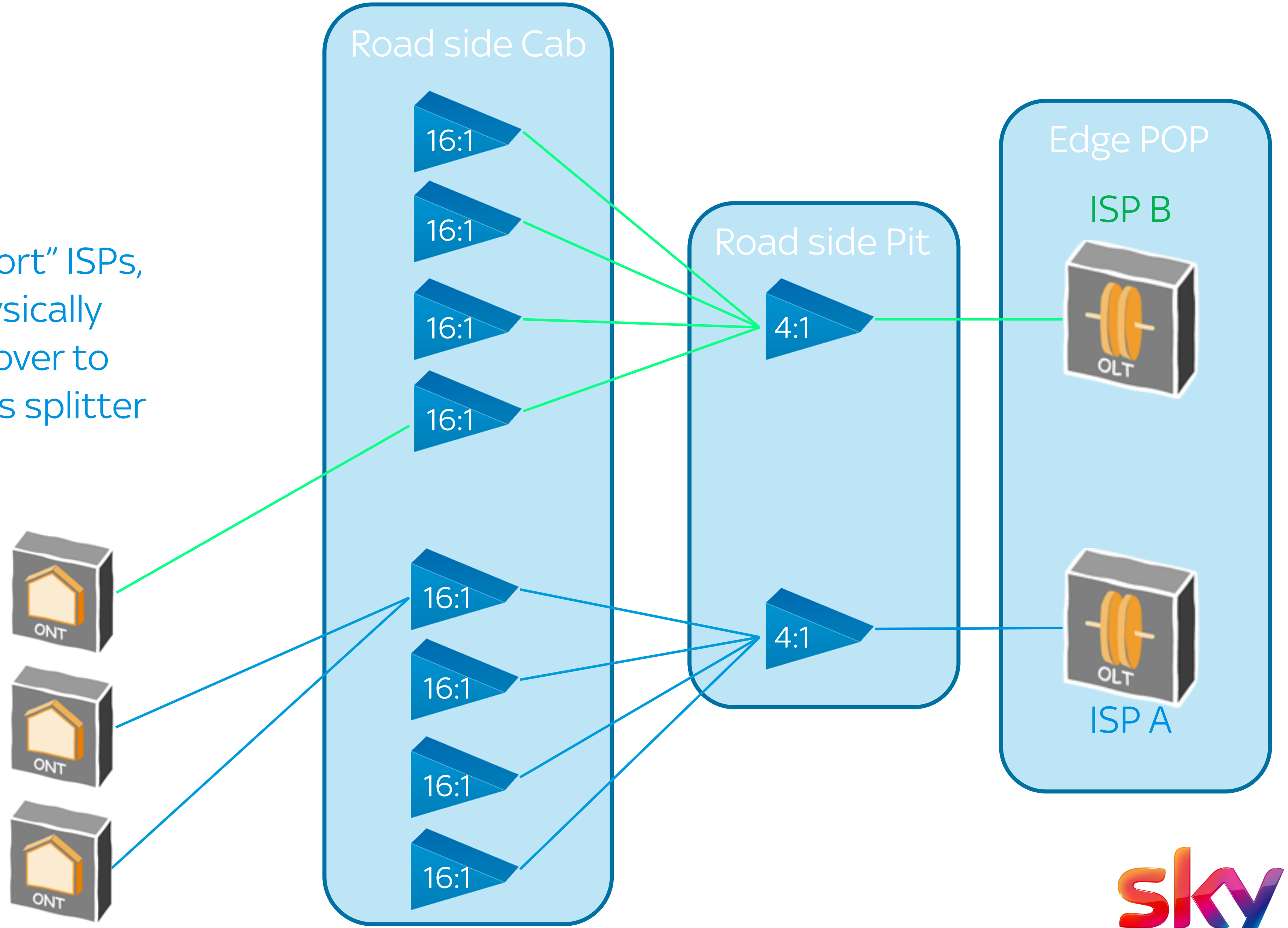
- Distinct Optical Distribution Network (ODN)
- Shared ducts / fibre bundles



Access

Unbundled GPON

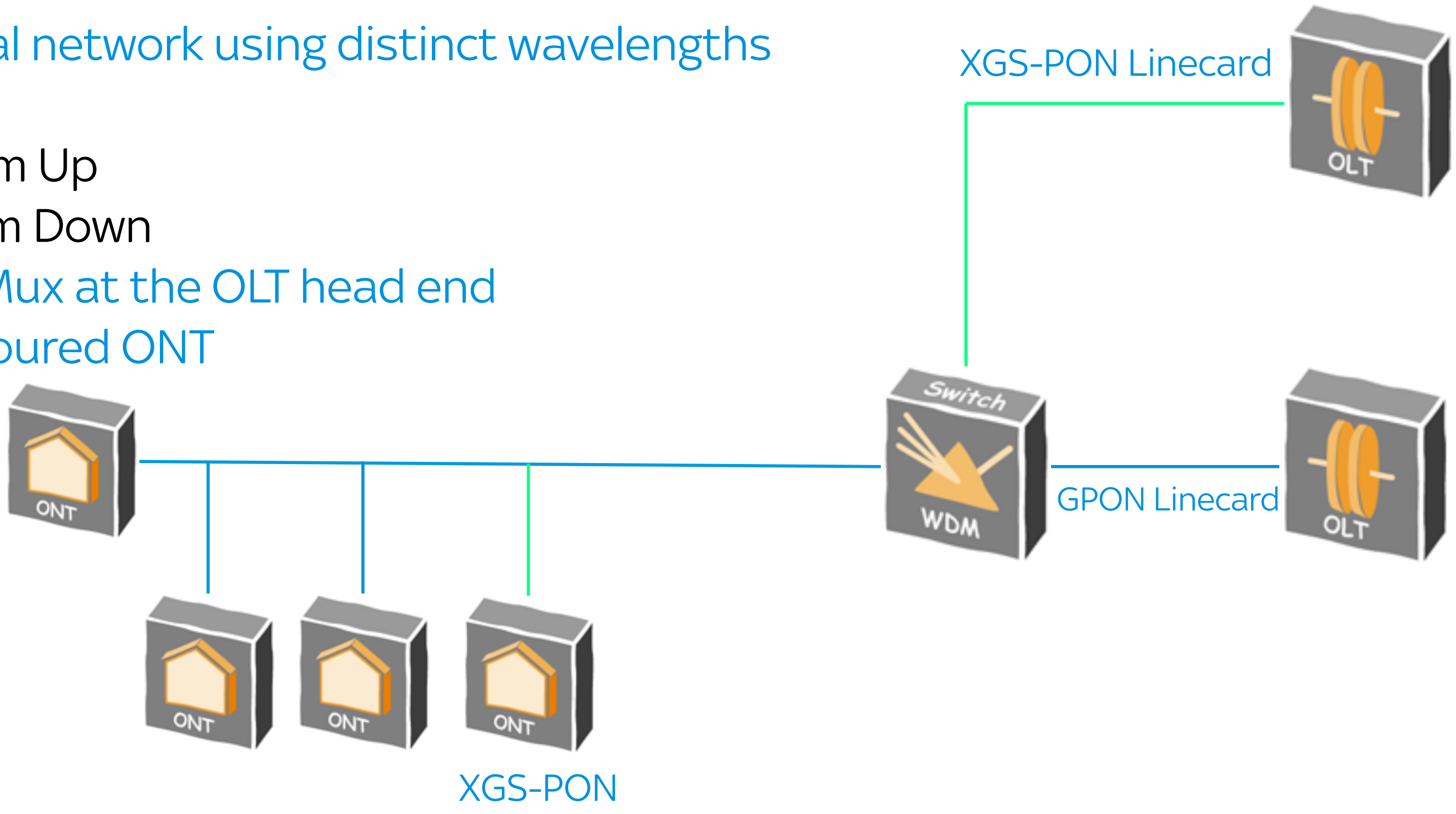
- A Customer can "port" ISPs, by an engineer physically patching the ONT over to the other provider's splitter



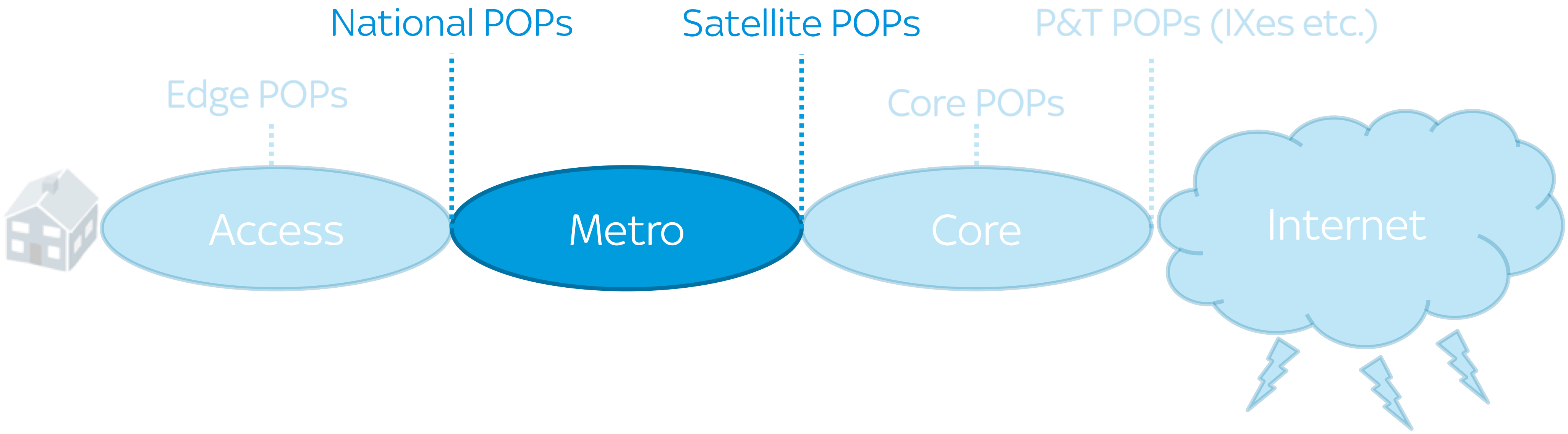
Access

Future XGS-PON

- Shared optical network using distinct wavelengths
- XGS-PON
    - ~1270nm Up
    - ~1577nm Down
- Wavelength Mux at the OLT head end
- XGS-PON coloured ONT



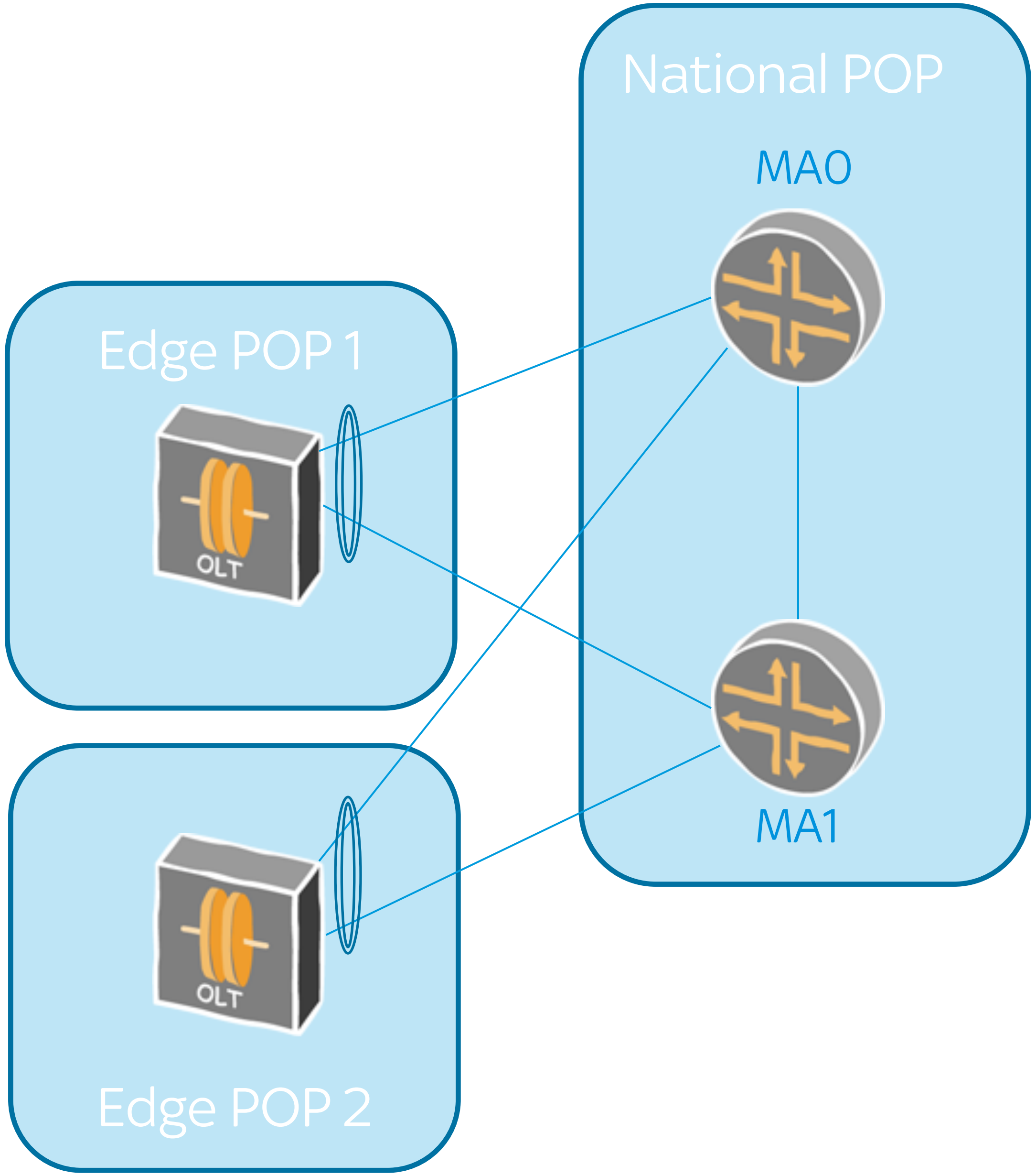
Metro Network



Metro Network

Metro Aggregation

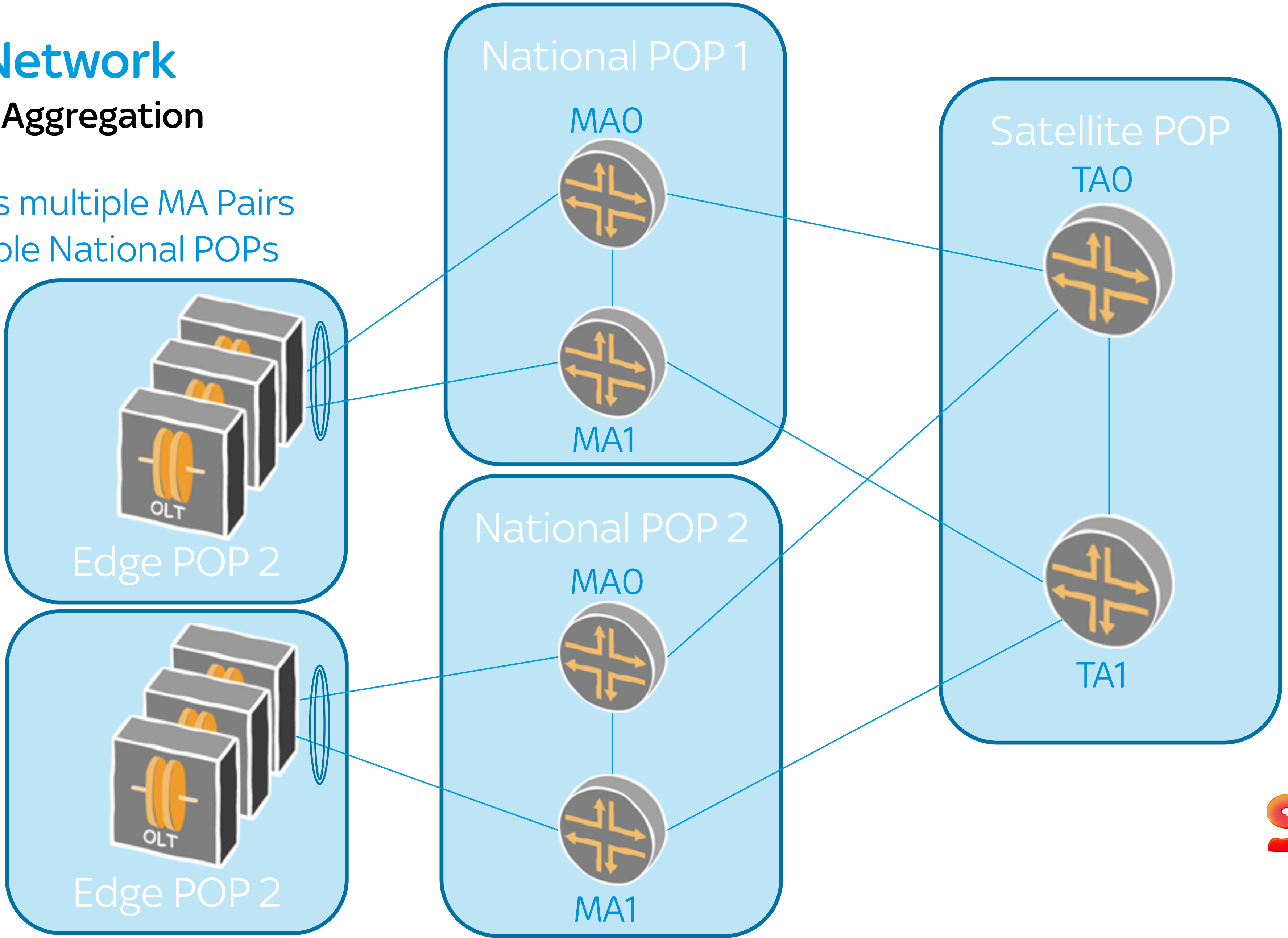
Aggregates multiple OLTs from multiple Edge POPs



Metro Network

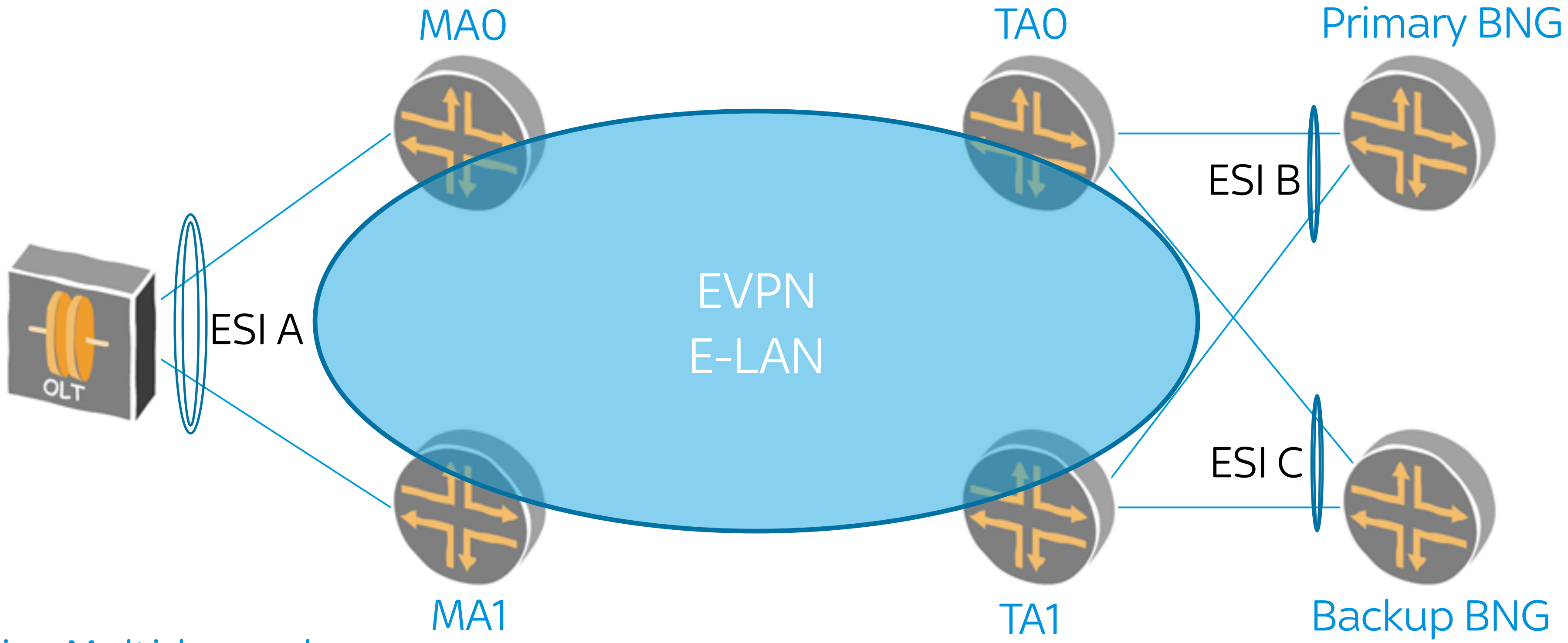
Transport Aggregation

Aggregates multiple MA Pairs
from multiple National POPs



Metro Network

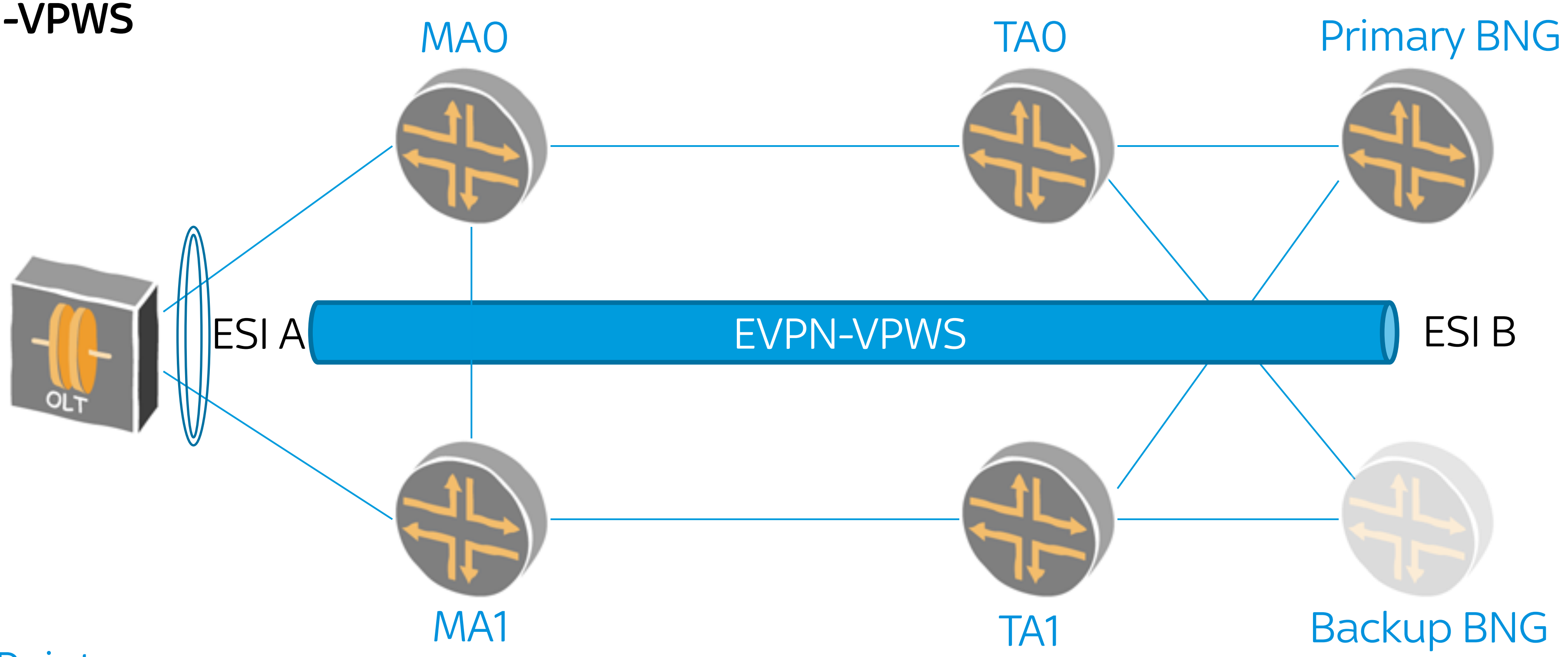
EVPN E-LAN



- Active/Active Multi-homed
- No need for MC-LAG or STP
- BGP-Signaled
- Efficient MAC-learning
- Efficient handling of BUM packets

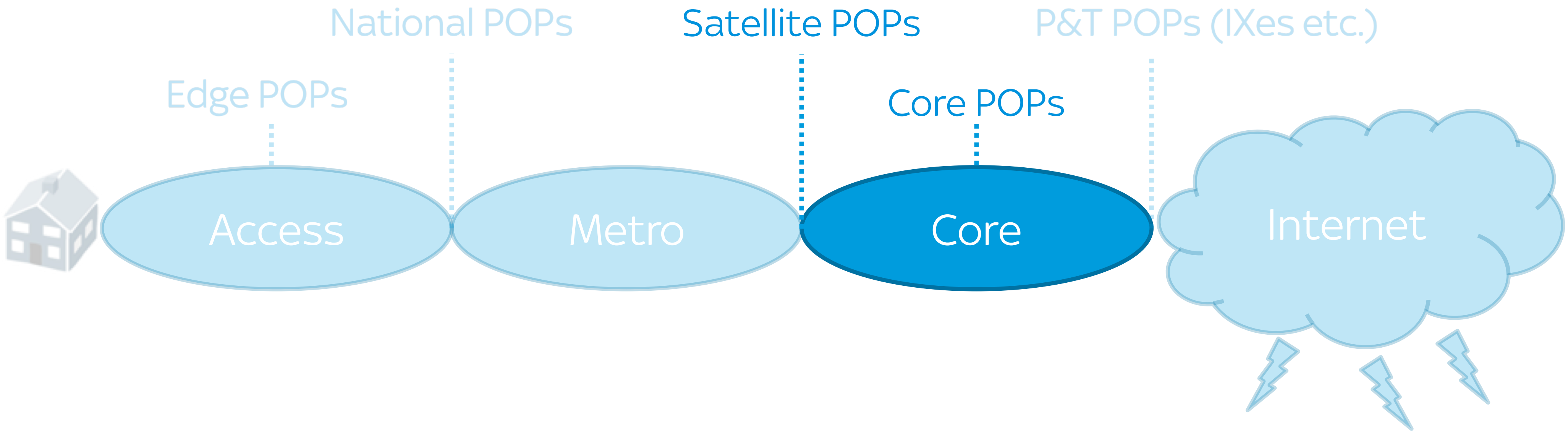
Metro Network

Future EVPN-VPWS



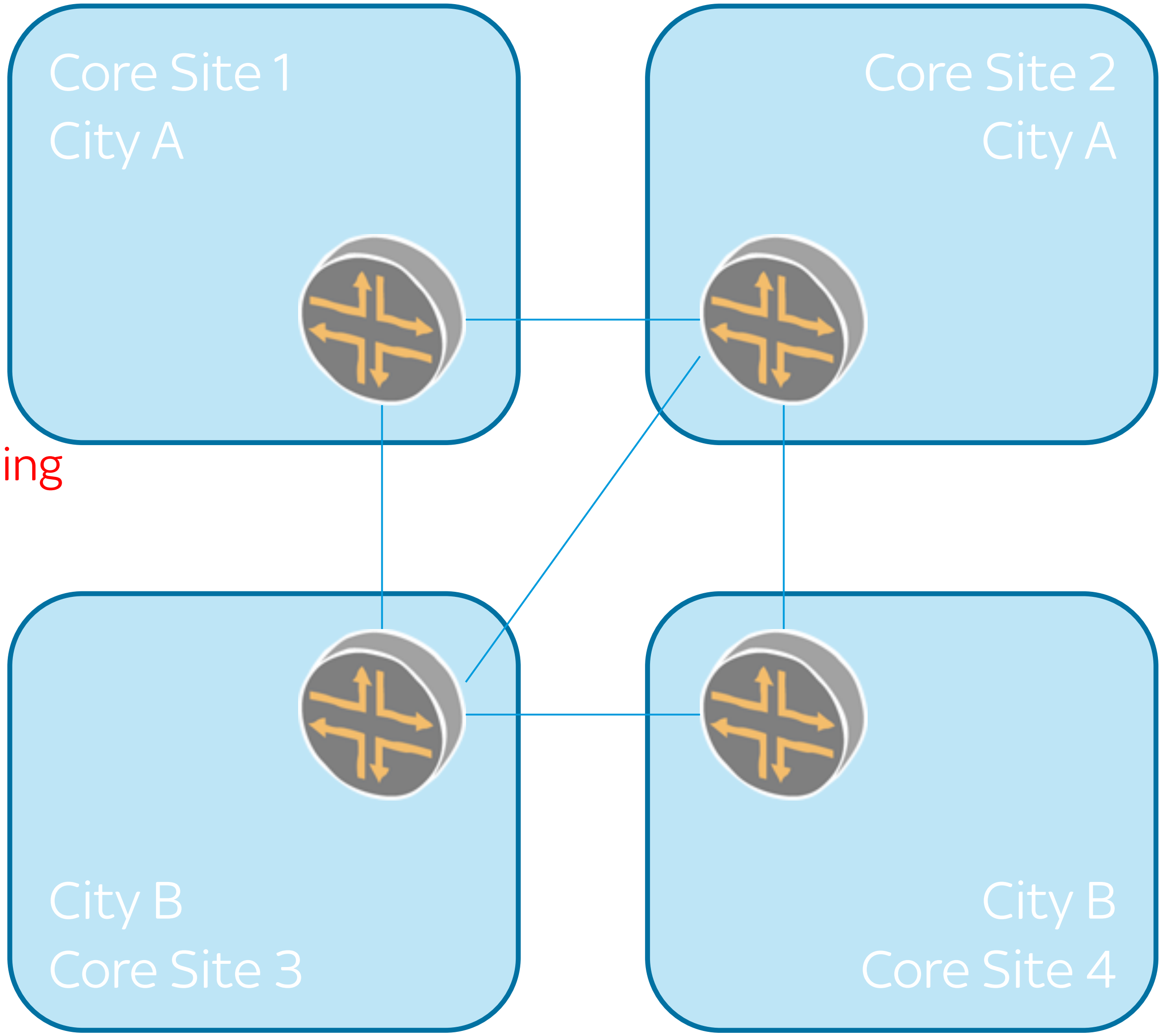
- Point-to-Point
  - No MAC-learning
- Single Active on BNG side
  - Using “Backup” control flag (RFC8214)

Core Network



Core Network

- Merchant Silicon
    - Cheaper per Gbps cost
    - High Port Density
    - Rigid Feature Set / Packet Processing
    - Bugs in Vendor Implementations
- Segment Routing (MPLS)
    - No LDP or RSVP-TE
- Virtualised Route Reflectors
- ECMP Everywhere

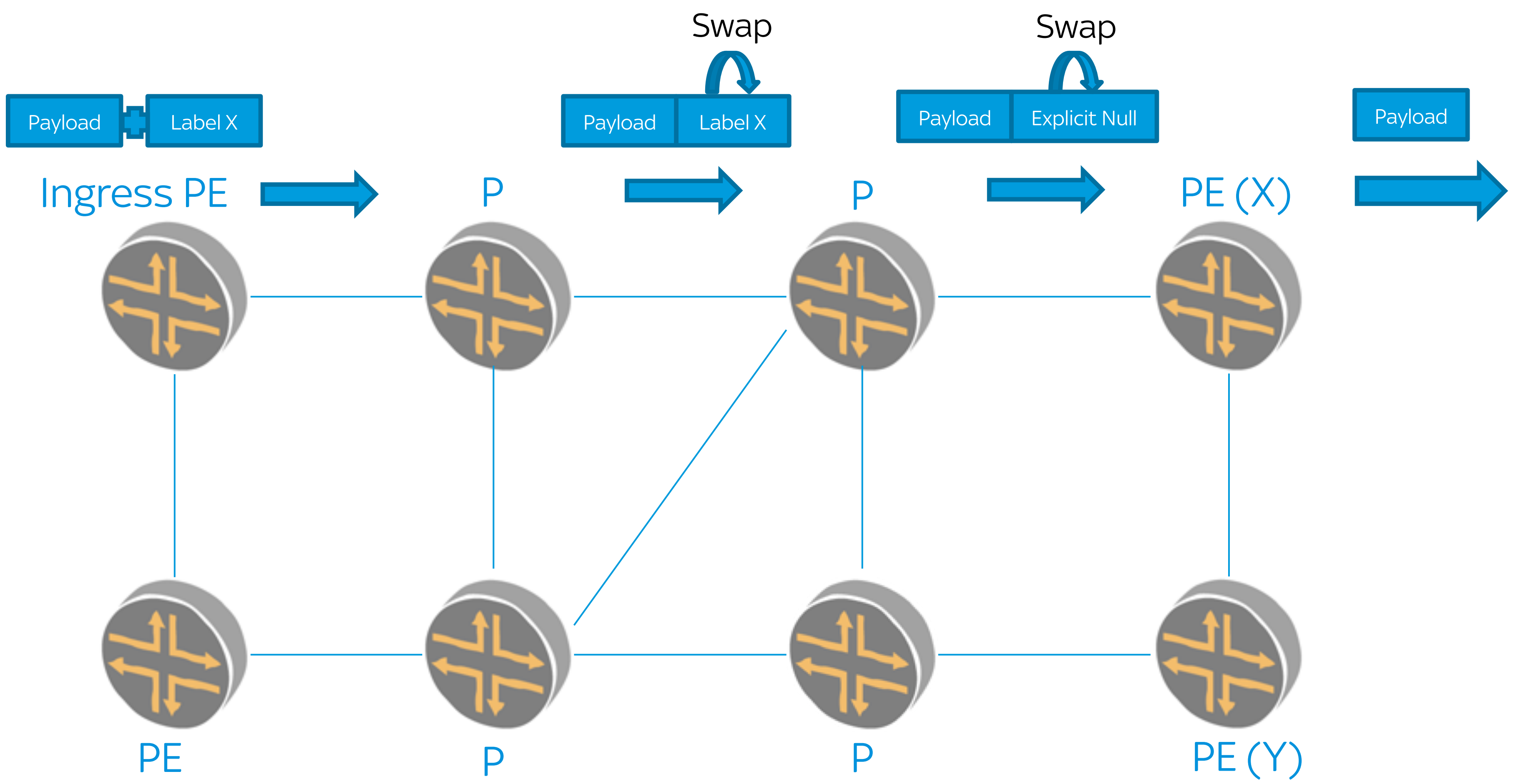


Segment Routing

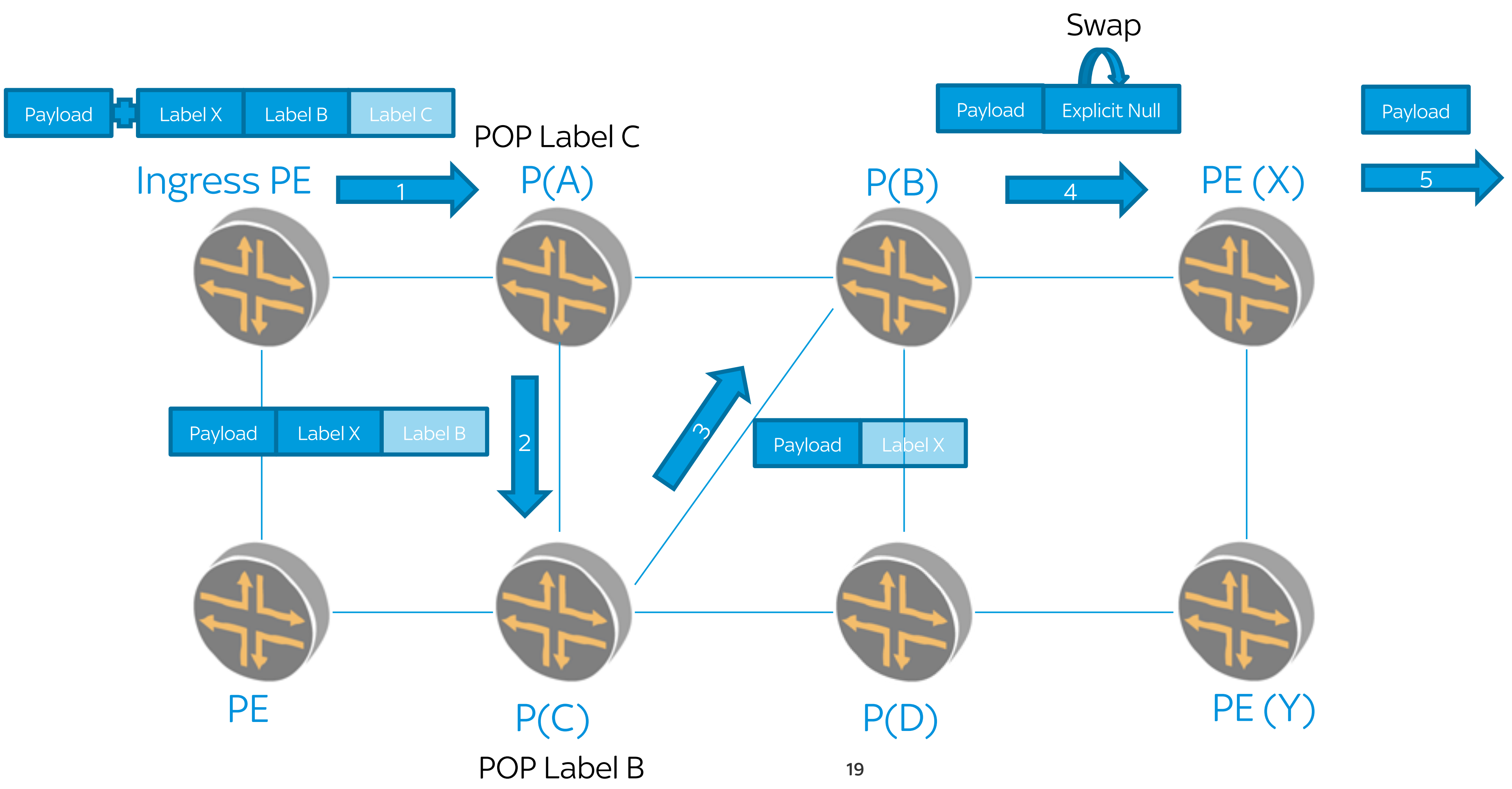
- Extensions to IGP for label distribution
  - Not a distinct new “protocol”
- Uses existing MPLS label switching
- No Stateful Label Switch Paths (LSPs)
  - State is kept in the packet header
  - Scalable
- Supports Traffic Engineering
  - Label Stack or “Segment List”
  - Offline Path Computation
  - Dynamic TE using Link Delay metrics signaled in ISIS (RFC8570)

Segment Routing

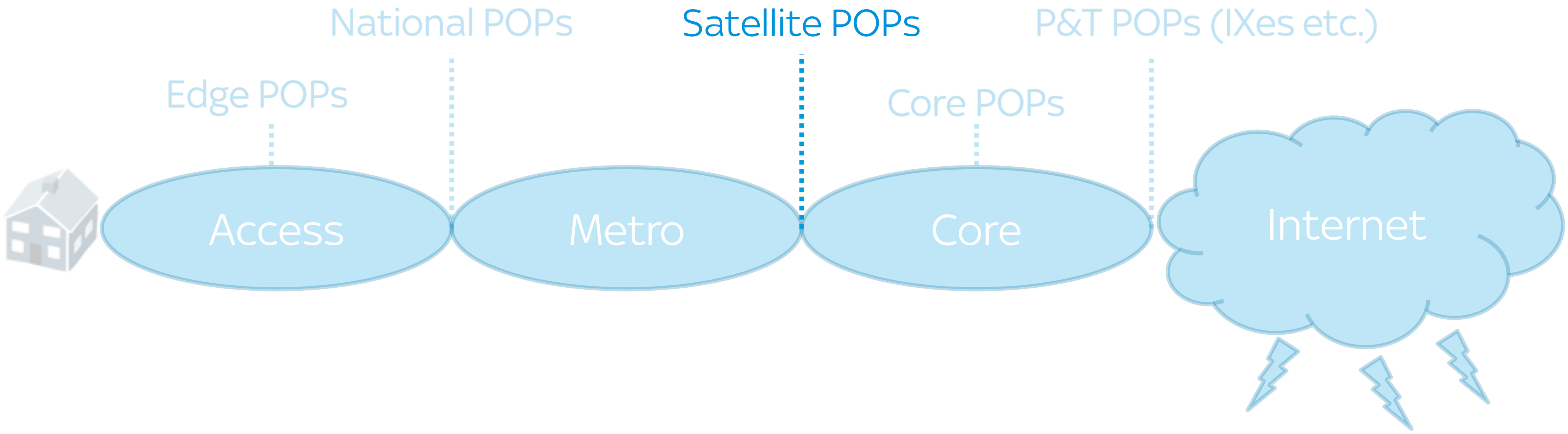
Shortest Path



Segment Routing
Traffic Engineering

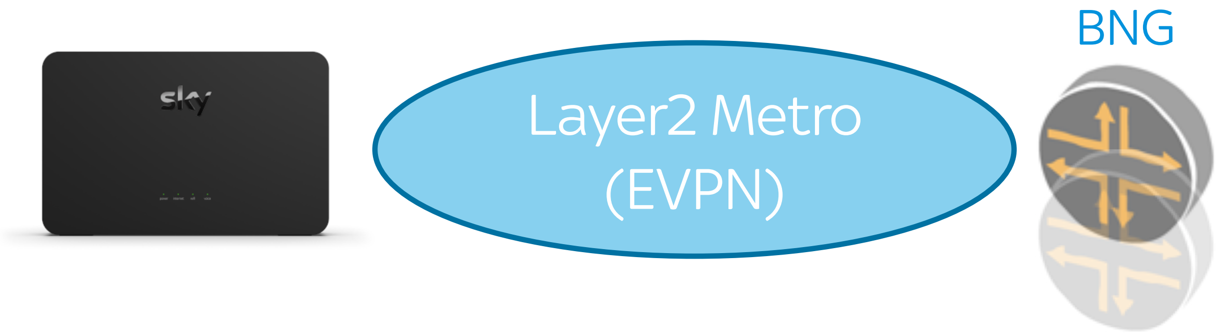


Subscriber Termination



Subscriber Termination

Broadband Network Gateways (BNG)



- IPoE – Less encapsulation overhead compared to PPPoE
- Native IPv6 - /48 PDs
- Port-based Authentication – DHCPv6 Option 37 Remote-ID
  - Inserted by the OLT's Lightweight DHCPv6 Relay Agent (LDRA)
- Redundant BNG – Proprietary vendor magic session state syncing, plus VRRP
- IPv4aaS – Mapping of Address and Port (MAP)

Future:

- Subscriber Termination directly on EVPN-VPWS
  - Vendor magic ties BNG backup state with EVPN state

IPv4 Addressing

- Starting with zero
- Join RIPE as a Local Internet Registry (LIR)
  – €2000 joining fee + €1400pa
  – Gets you a /22 or 1024 IPv4 addresses
  – Maybe this will be enough for infrastructure?
- Buy off the open market @ >$15USD / IP
- So you will probably also want to do some form of IPv4 address sharing
  – We've chosen to use MAP-T

MOST RECENT TRANSFERS COMPLETED

Individual buyers and sellers will agree to a specific nominated currency such as USD, EUR, GBP, etc. in the following table, prices per IP are illustrated in USD for comparative regional purposes.

Block Size*	/24	/23	/22	/21	/20	/19	/18	/17	/16
Price/IP (USD)	26.00	23.00	20.00	20.00	19.50	19.50	19.00	19.50	19.00+ depending on quality

Source: http://ipv4marketgroup.com/broker-services/buy/
11th July 2019

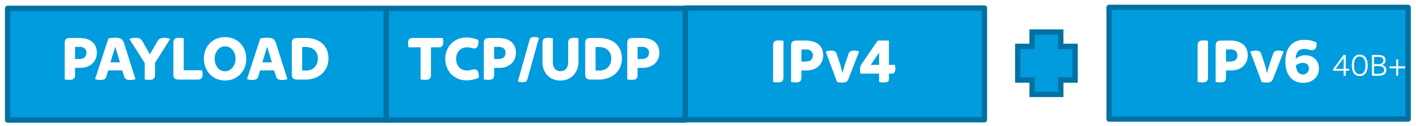
Mapping of Address and Port (MAP)

- IPv4aaS.
  - IPv4 over the top of IPv6 transport.
  - IPv6-only Access Layer.
  - Reduces operational overhead.
- Allows IPv4 address sharing, or 1:1.
- No DNS synthesizing required.
- Doesn’t require an agent on end-hosts.
- Can operate in either encapsulation or translation modes.
- **Stateless.**

Mapping of Address and Port (MAP)

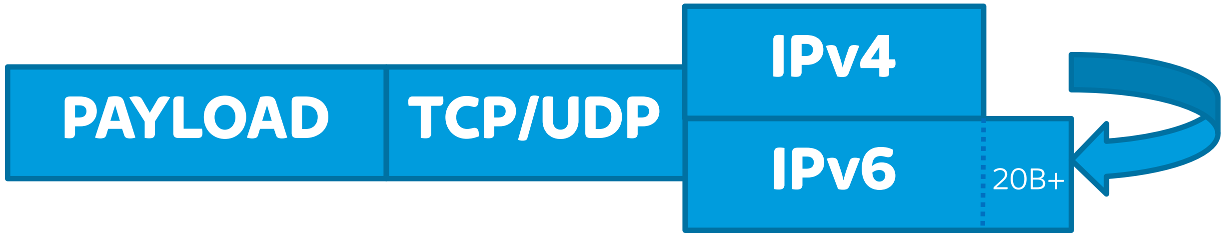
RFC7597: MAP-E

- Encapsulation
  - Larger per-packet overhead.
  - IPv4 header remains intact.



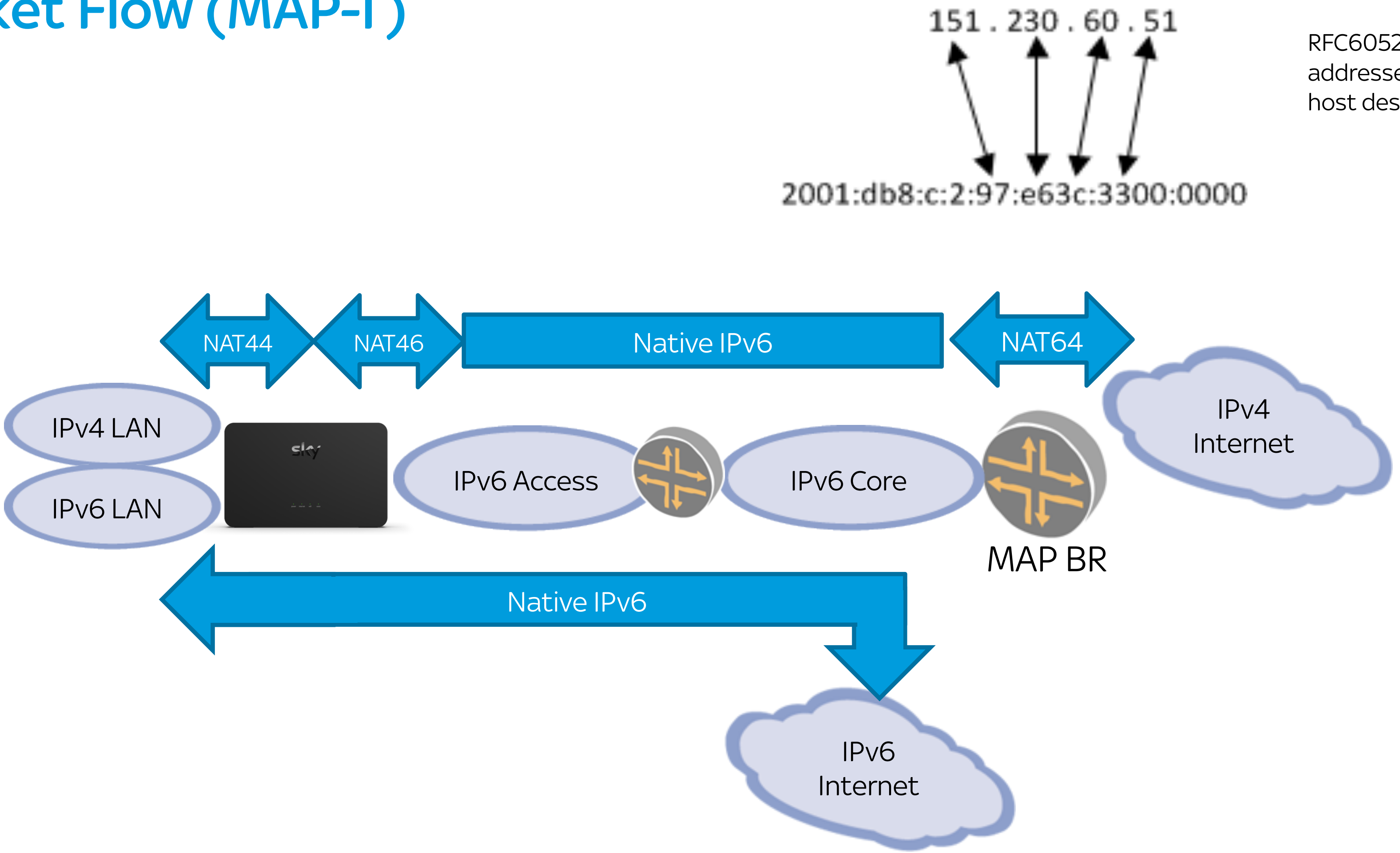
RFC7599: MAP-T

- Translation
  - Less per-packet overhead (not zero!)
  - Loses IPv4-only header attributes.
  - 5-tuple hashing. (E.g., with ECMP or over LAGs)
  - Border relay-bypass.

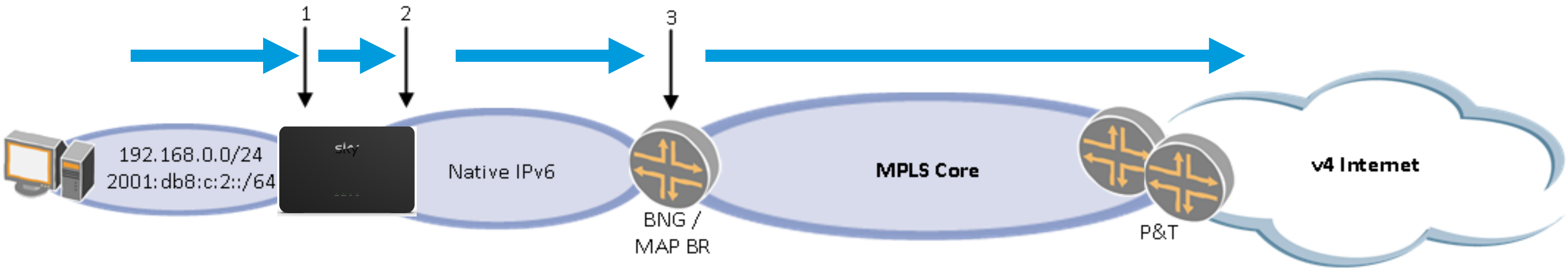


Packet Flow (MAP-T)

RFC6052 IPv4-embedded IPv6 addresses used for external host destination address.



Packet Flow Example (MAP-T)



TCP/UDP Header		IP Header	
SPORT: 6783	DPORT: 80	SOURCE: 192.168.0.123	DEST: 8.8.8.8

1. Stateful NAPT on CPE translates both source IPv4 address and source port (when oversubscribing) based on the BMR

TCP/UDP Header		IP Header	
SPORT: 1200	DPORT: 80	SOURCE: 151.230.60.51	DEST: 8.8.8.8

2. MAP-T agent on CPE translates source IP address to v6 based on the BMR, and destination IP address to v6 based on the DMR (and RFC6052)

TCP/UDP Header		IP Header	
SPORT: 1200	DPORT: 80	SOURCE: 2001:db8:c:2:97:e63c:3300:0000	DEST: 2001:db8:ffff:0:8:808:800:0000

3. MAP-T Border Router translates source IP address back to IPv4 based on the BMR, and destination IP based on DMR (and RFC6052)

TCP/UDP Header		IP Header	
SPORT: 1200	DPORT: 80	SOURCE: 151.230.60.51	DEST: 8.8.8.8

Note: Source address translation is BMR-dependent, it would not use RFC6052 as shown in this example.

Mapping of Address and Port (MAP)

Stateless

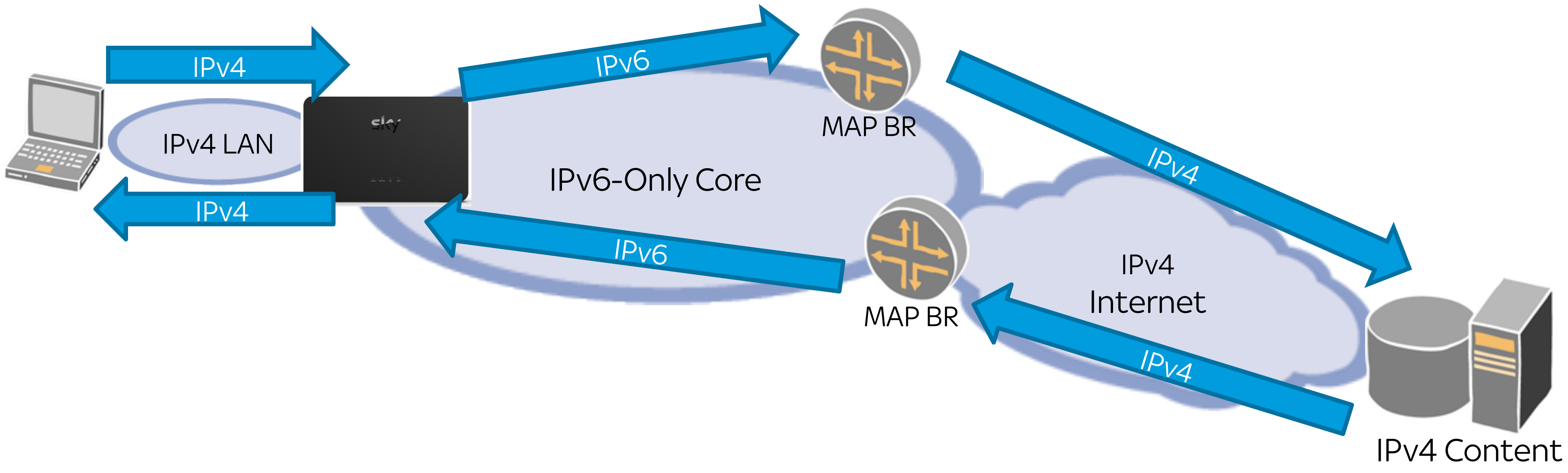
- No need to keep track of every flow
- Efficient packet processing
- Cheaper, more scalable hardware
- Already supported on existing linecards from some vendors.

- Some jurisdictions may require 5-tuple logging for compliance reasons, which stateless IPv4aaS methods don't provide.

Appendix A – Additional MAP-T Awesomeness

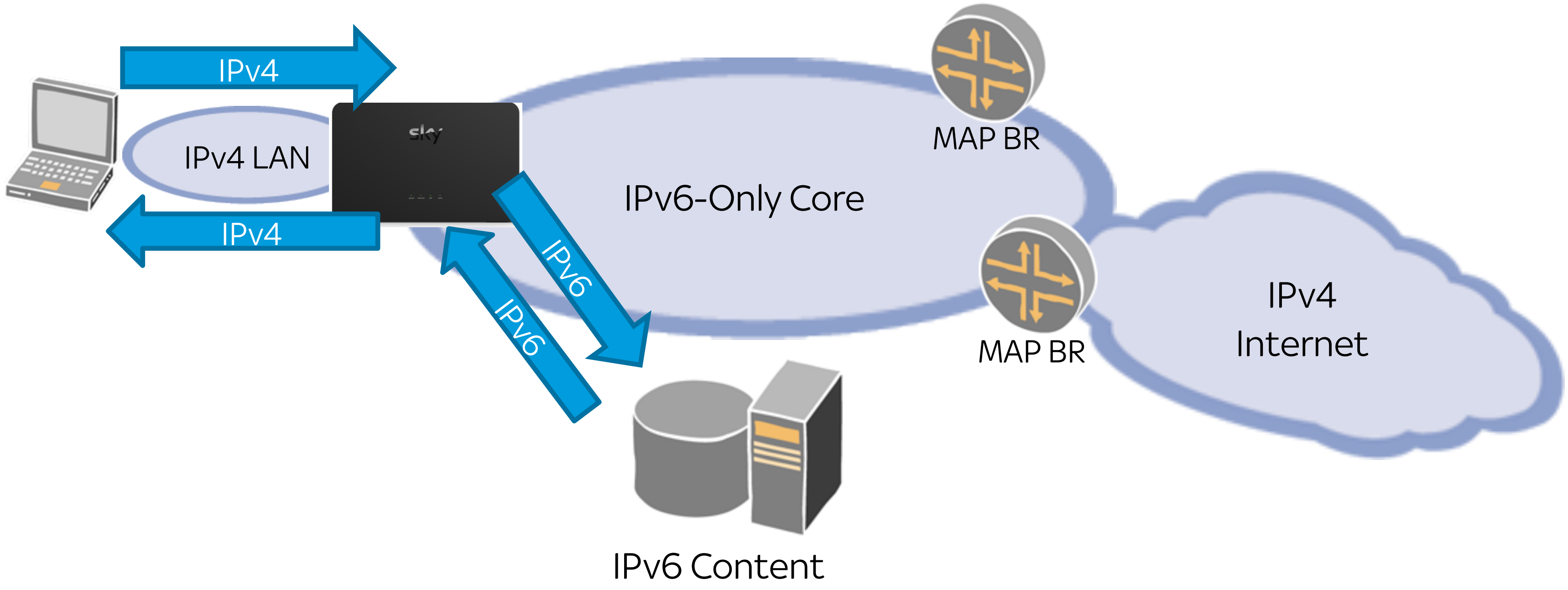
MAP Border Relay Anycasting

- DMR IPv6 prefixes can be anycasted internally
- Public IPv4 prefixes can be anycasted externally.
- Stateless translation/encapsulation allows for asymmetric packet flows.



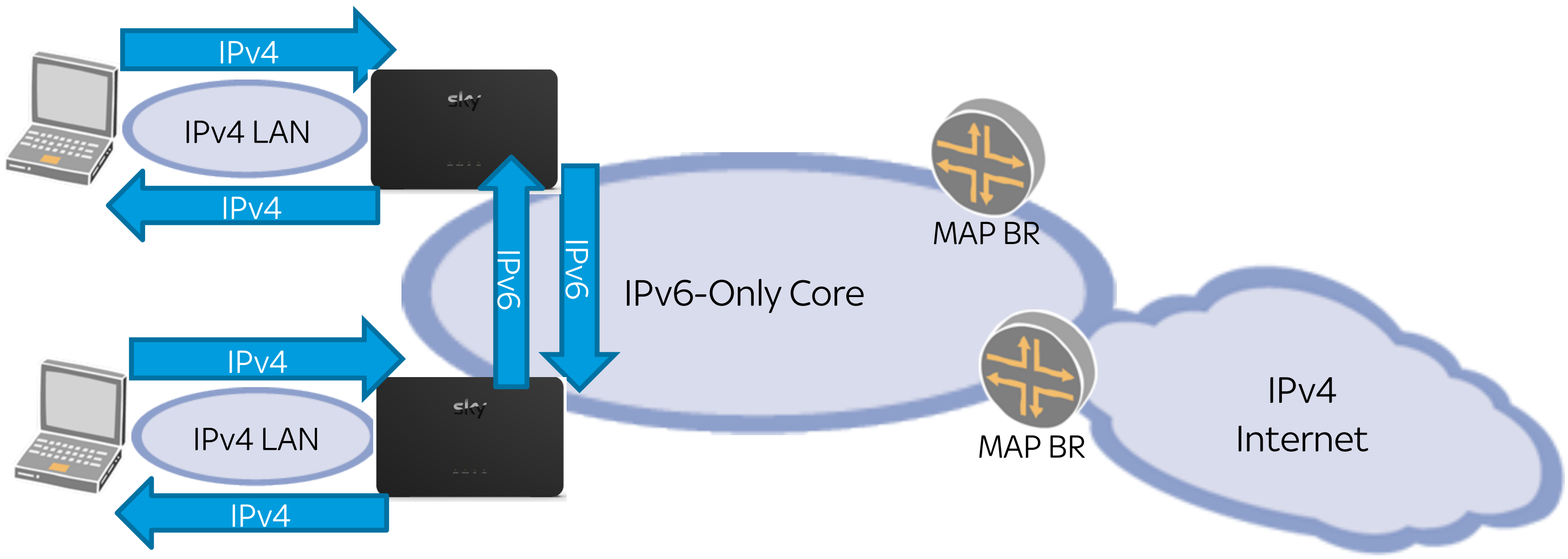
MAP Border Relay Bypass for CDN

- On-net content servers can be numbered from within the IPv6 DMR prefix, allowing for Border Relay-bypass, using more specific destination-based routing.
- Allows for serving of IPv4-only clients from IPv6-capable CDNs.



MAP Forward Mapping Rules

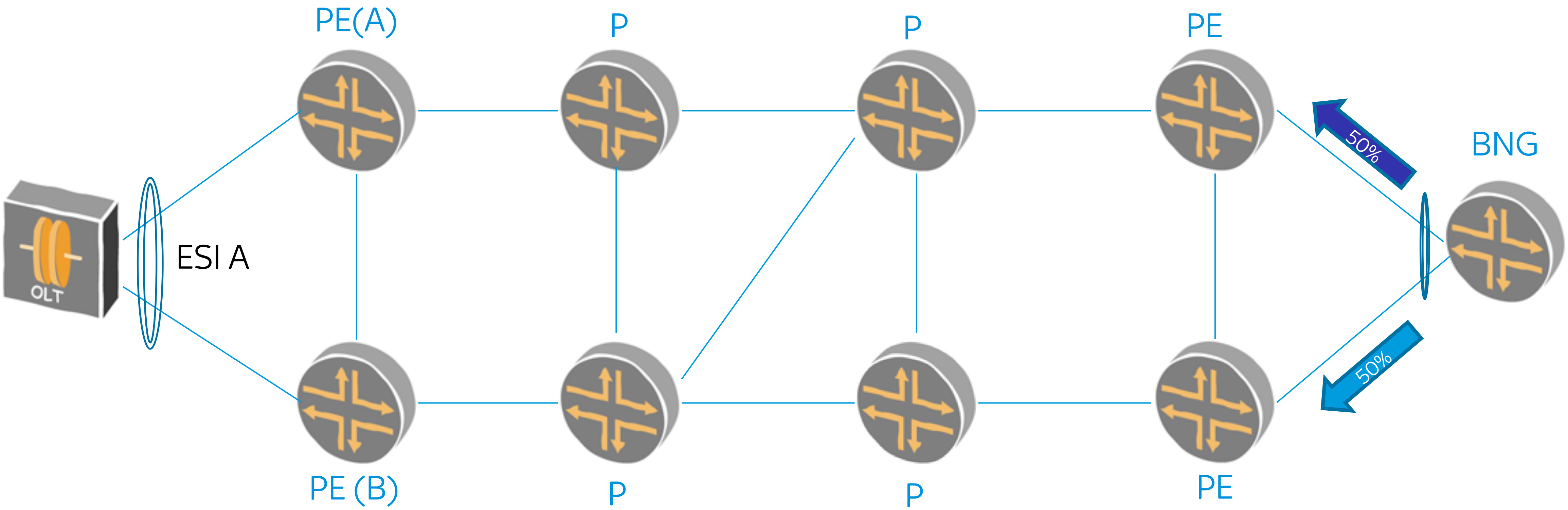
- Allows direct CPE <-> CPE communication, bypassing Border Relays.



Appendix B – All the Load Balancing

Load-balancing with EVPN + ECMP

Ingress LAG Hashing

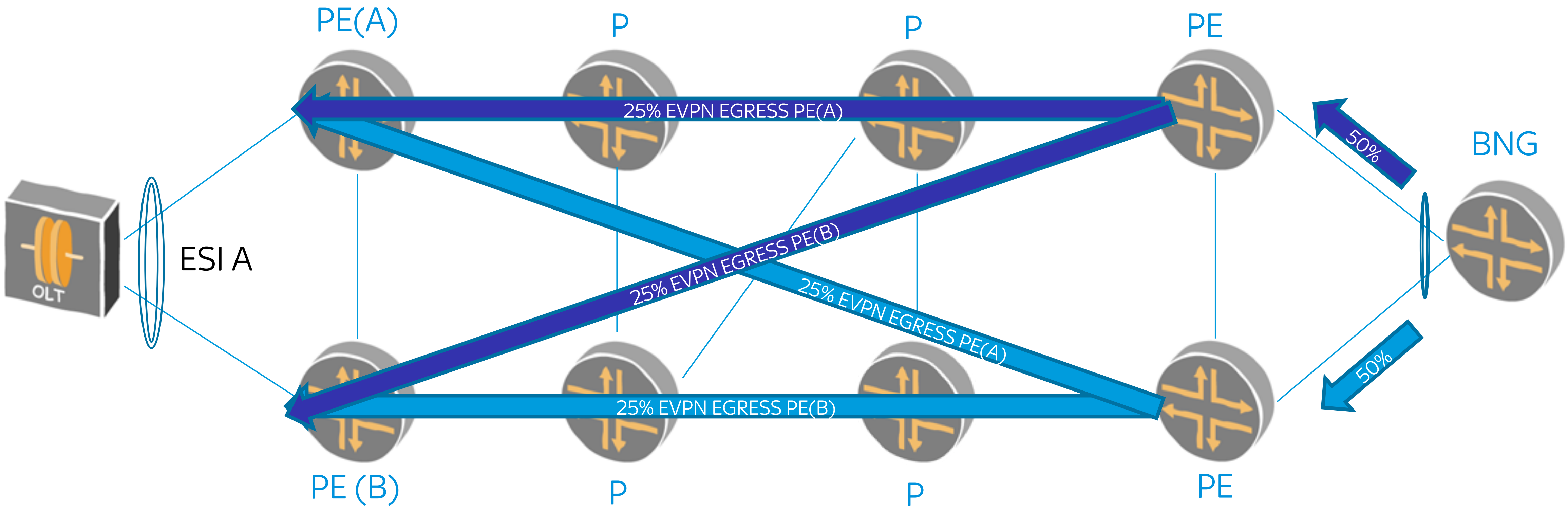


Step 1: Ingress LAG Hashing

Hashed across all LAG members

Load-balancing with EVPN + ECMP

EVPN “Aliasing”

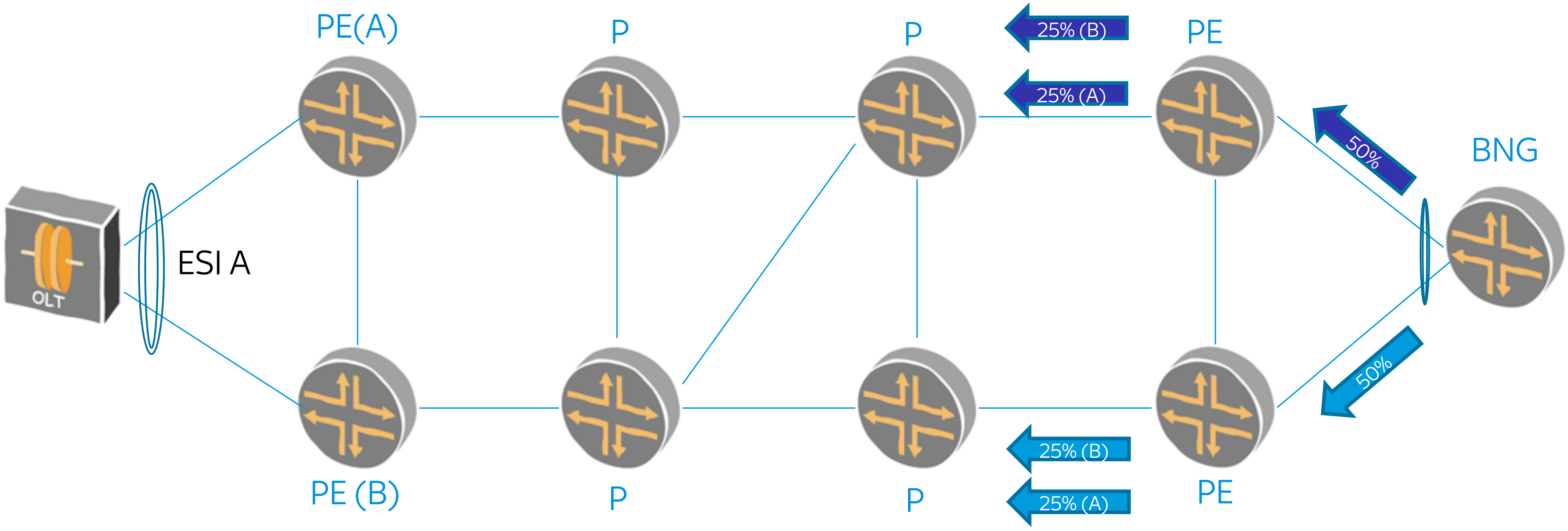


Step 2: EVPN Active/Active Load Balancing

Both ingress PEs see 2 egress PEs advertising ESI A

Load-balancing with EVPN + ECMP

EVPN "Aliasing"

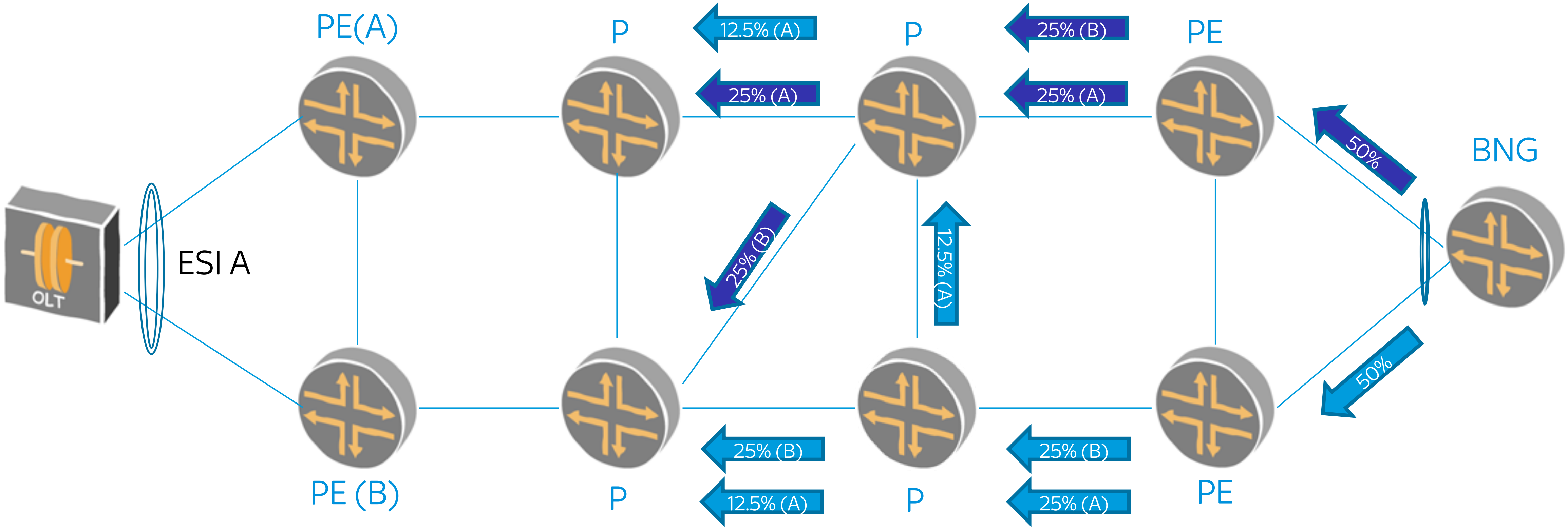


Step 2: EVPN Active/Active Load Balancing

Both ingress PEs see 2 egress PEs advertising ESI A

Load-balancing with EVPN + ECMP

ECMP of Egress PE Loopbacks in ISIS

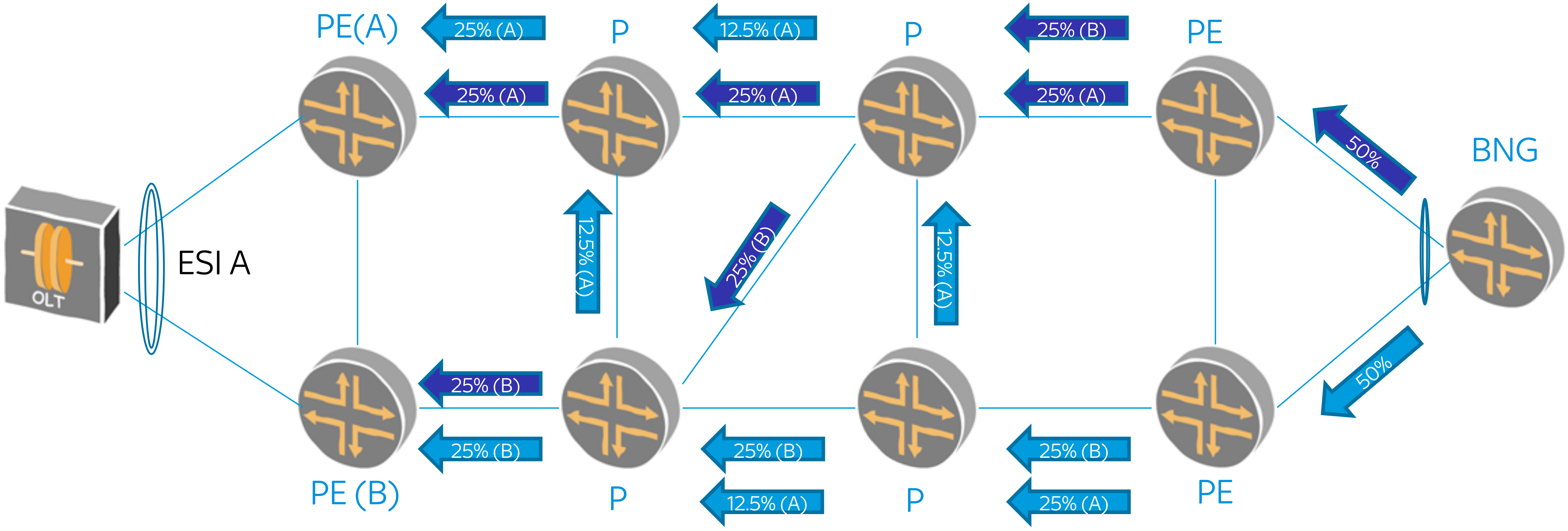


Step 3: ECMP Next Hops

LSRs see multiple paths for both egress PE next hops.

Load-balancing with EVPN + ECMP

ECMP of Egress PE Loopbacks in ISIS

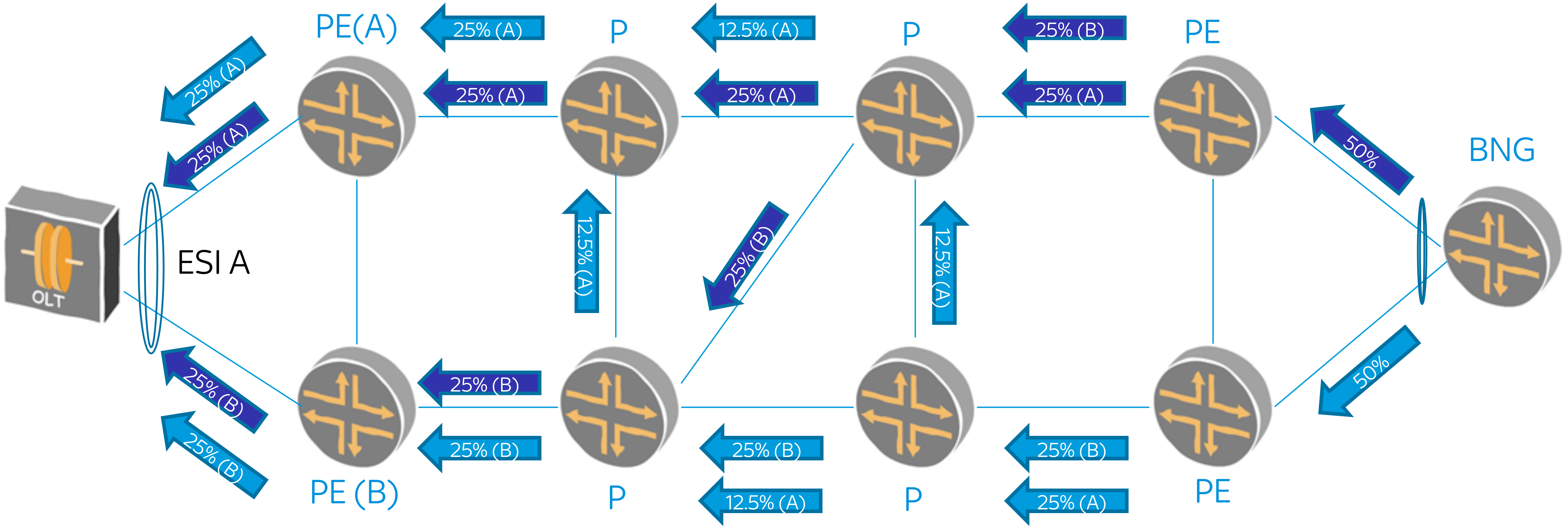


Step 3: ECMP Next Hops

LSRs see multiple paths for both egress PE next hops.

Load-balancing with EVPN + ECMP

Egress LAG Hashing



Step 4: Egress LAG Hashing

End result is the same in this instance. But interesting paths through the core