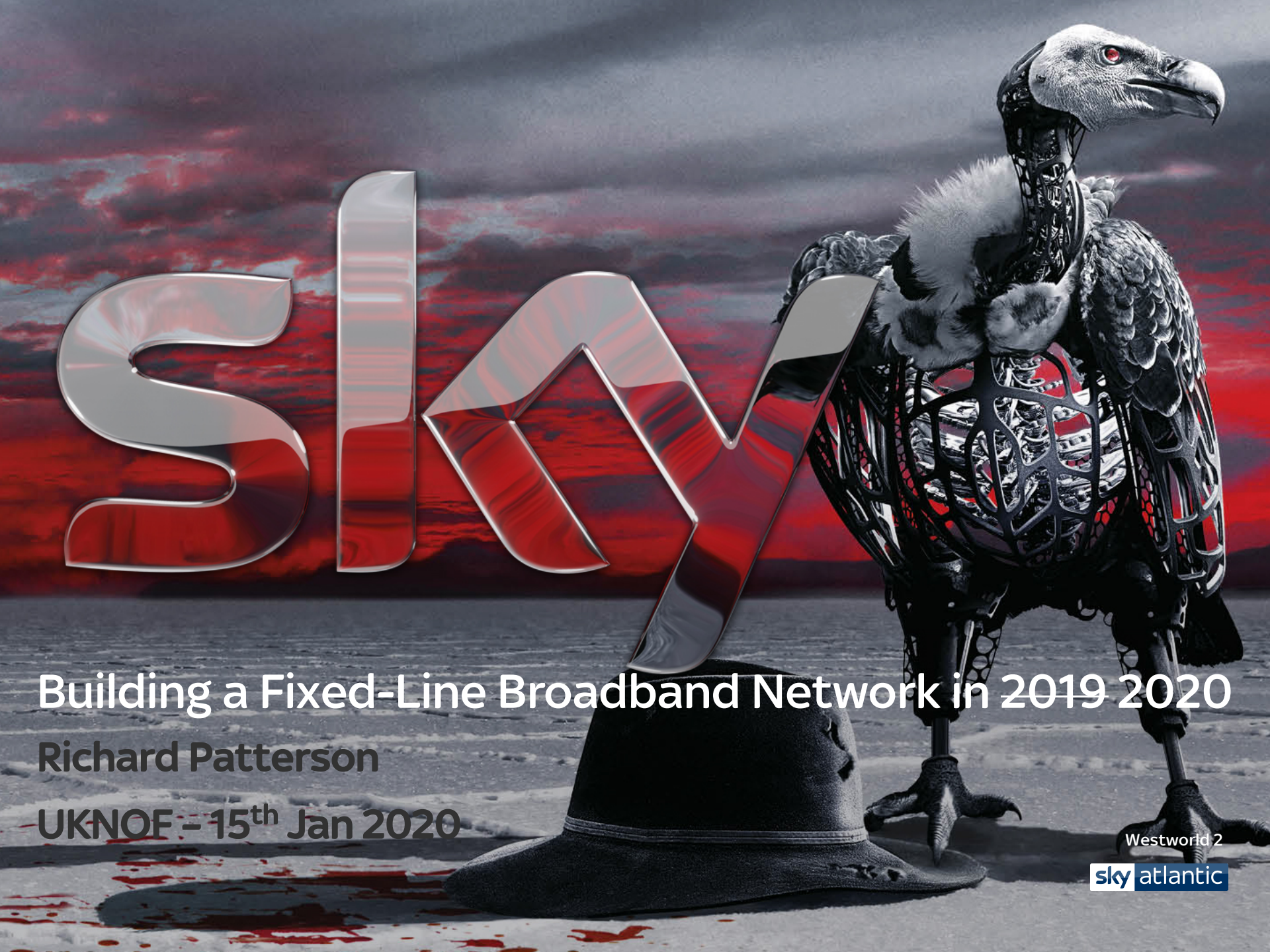


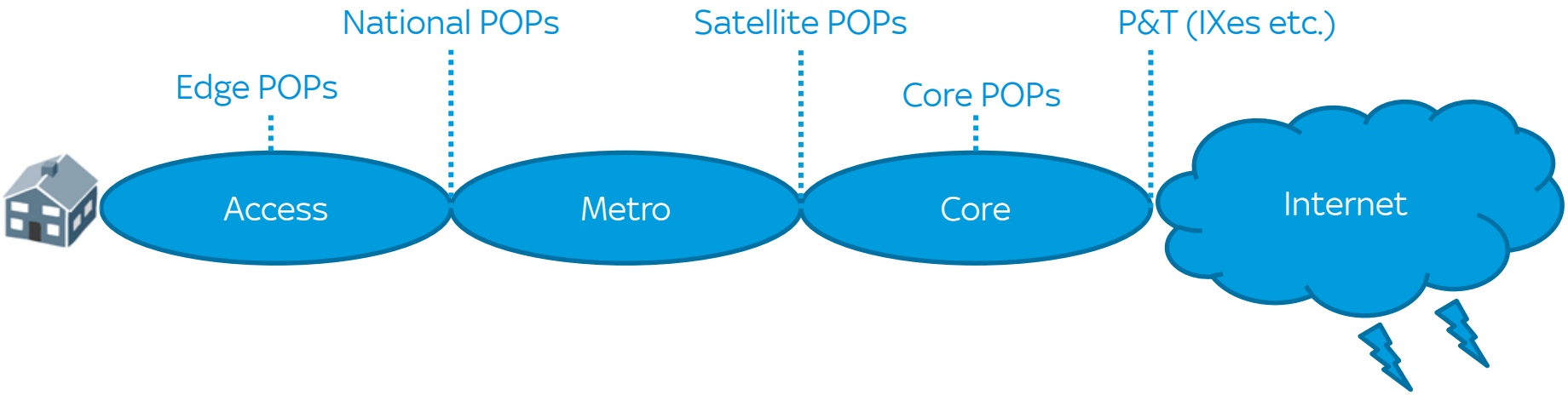
Building a Fixed-Line Broadband Network in ~~2019~~ 2020

Richard Patterson

UKNOF - 15th Jan 2020

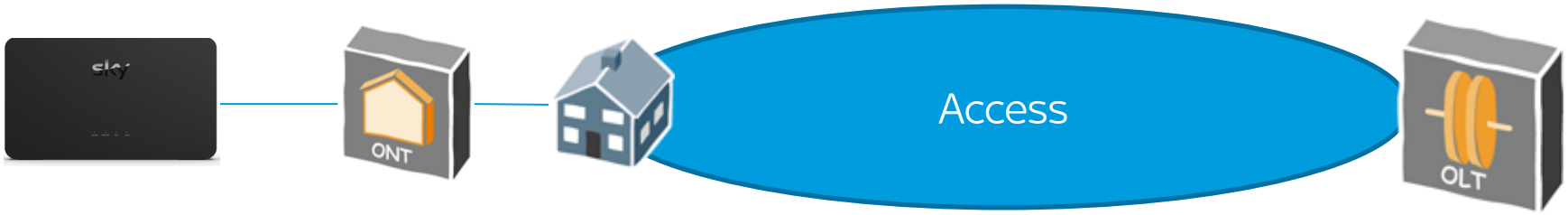
High-Level

End-to-End Network



Access

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

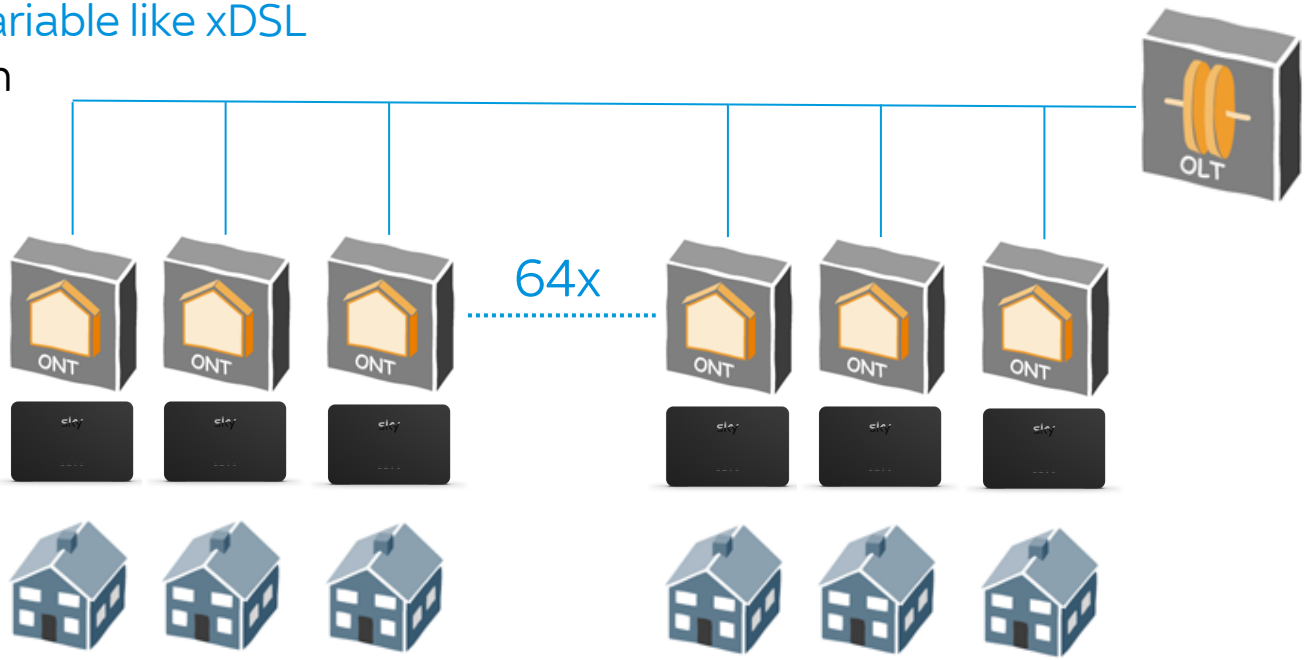


- Optical Line Terminal (OLT) @ Edge POP
- Optical Network Terminal (ONT) @ Customer Premise
  - 2 box solution with Sky Hub 4 CPE router

Access

GPON

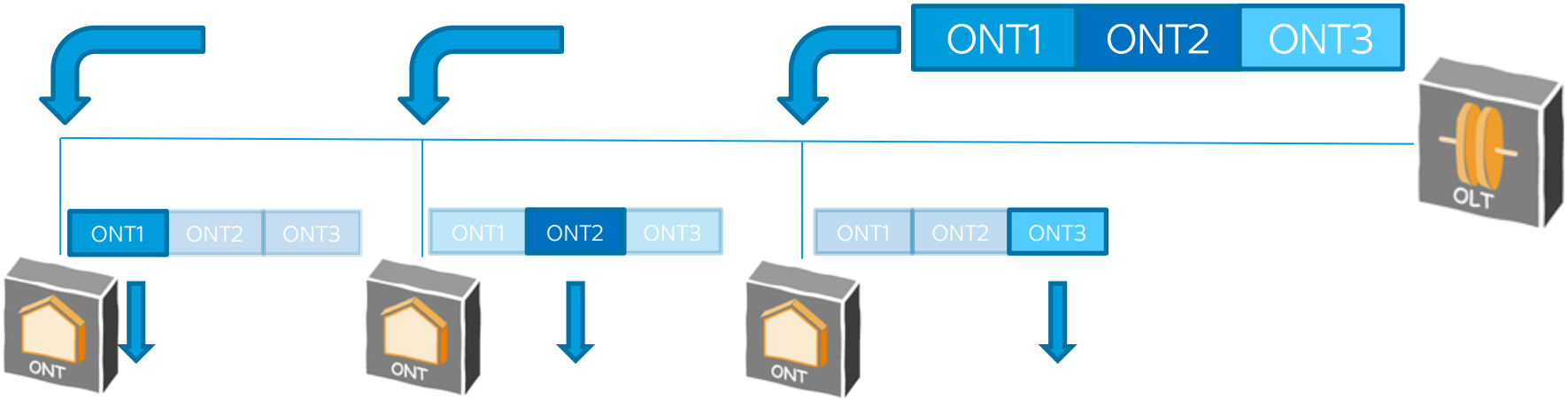
- Single-strand single-mode fibre
  - BiDi Optics
- Point-to-Multipoint using passive splitters
  - Commonly @ 32:1 or 64:1 but up to 128:1
- Performance isn't variable like xDSL
  - 2.5Gbps down
  - 1.25Gbps up
- Optical Budget
  - Splitters
  - Splicing
  - Connectors
  - Distance
  - Etc.



Access

GPON - Downstream

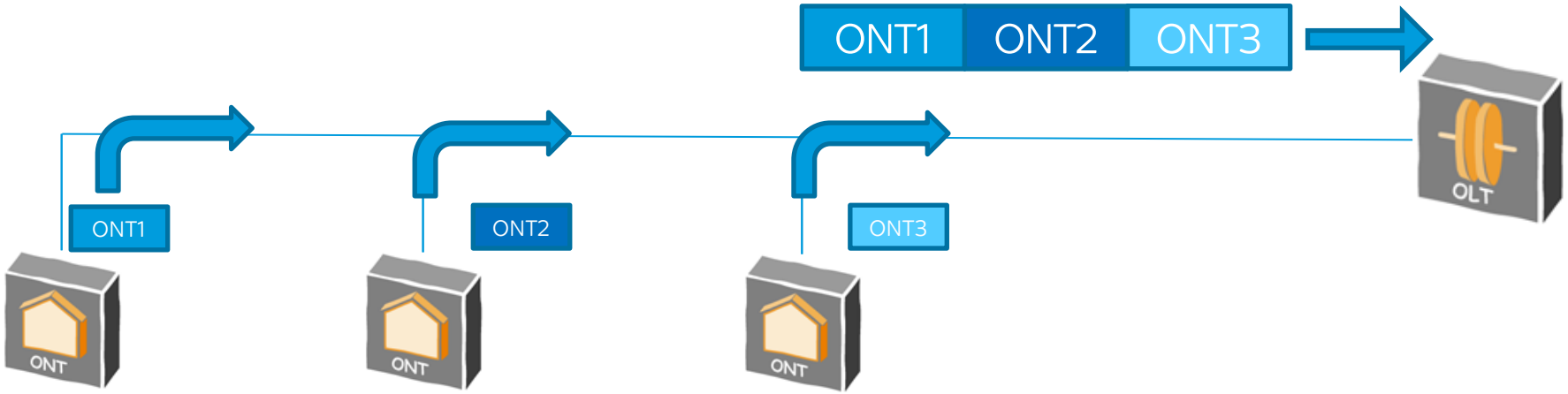
- ~1490nm Downstream
- Broadcast to all ONTs
- Distinguished by Port-ID in the GEM header
- AES128 Encrypted



Access

GPON - Upstream

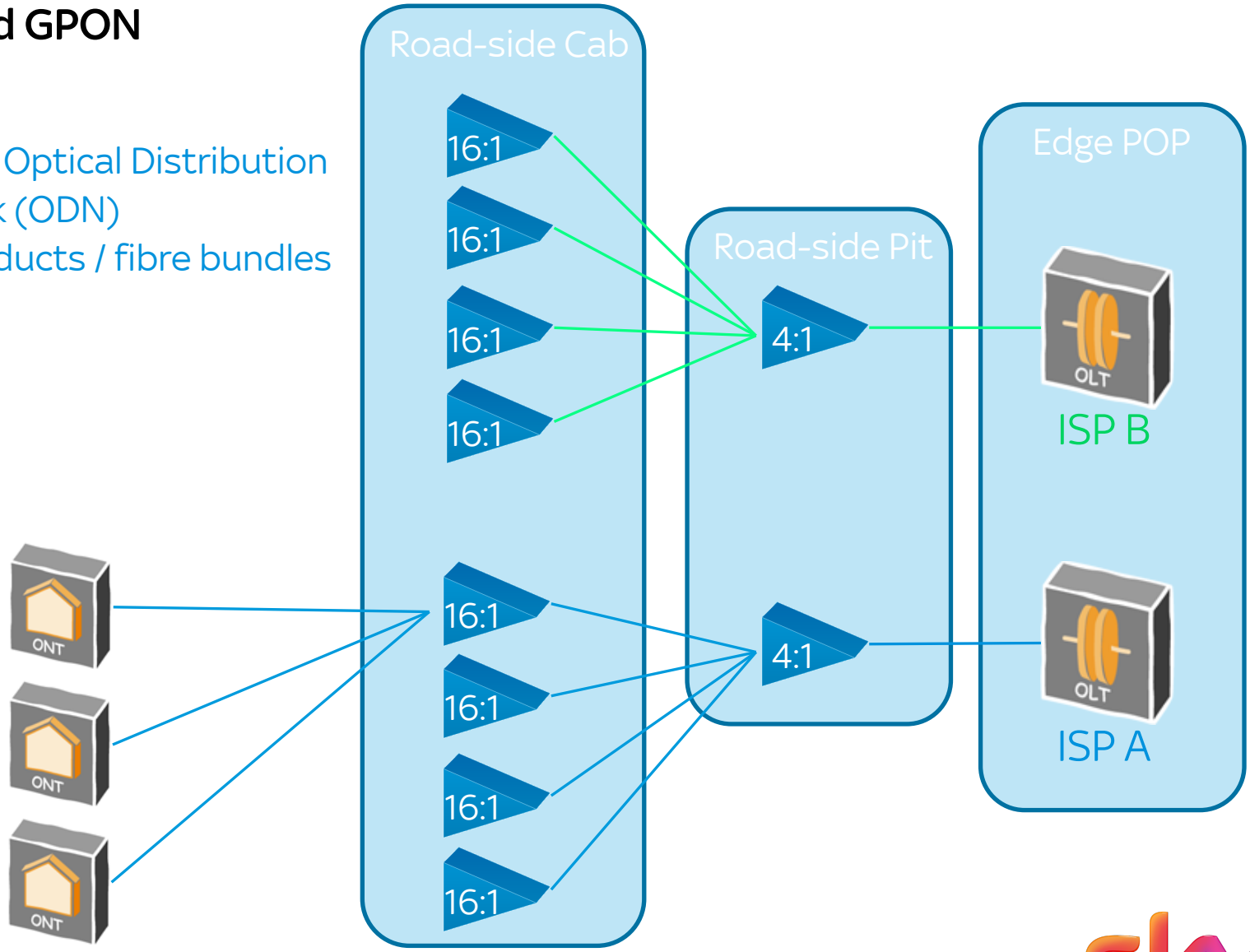
- ~1310nm Upstream
- Time-Division Multiple Access
- OLTs “range” to find each ONT’s distance from OLT
- Frames converge on the fibre without collision



Access

Unbundled GPON

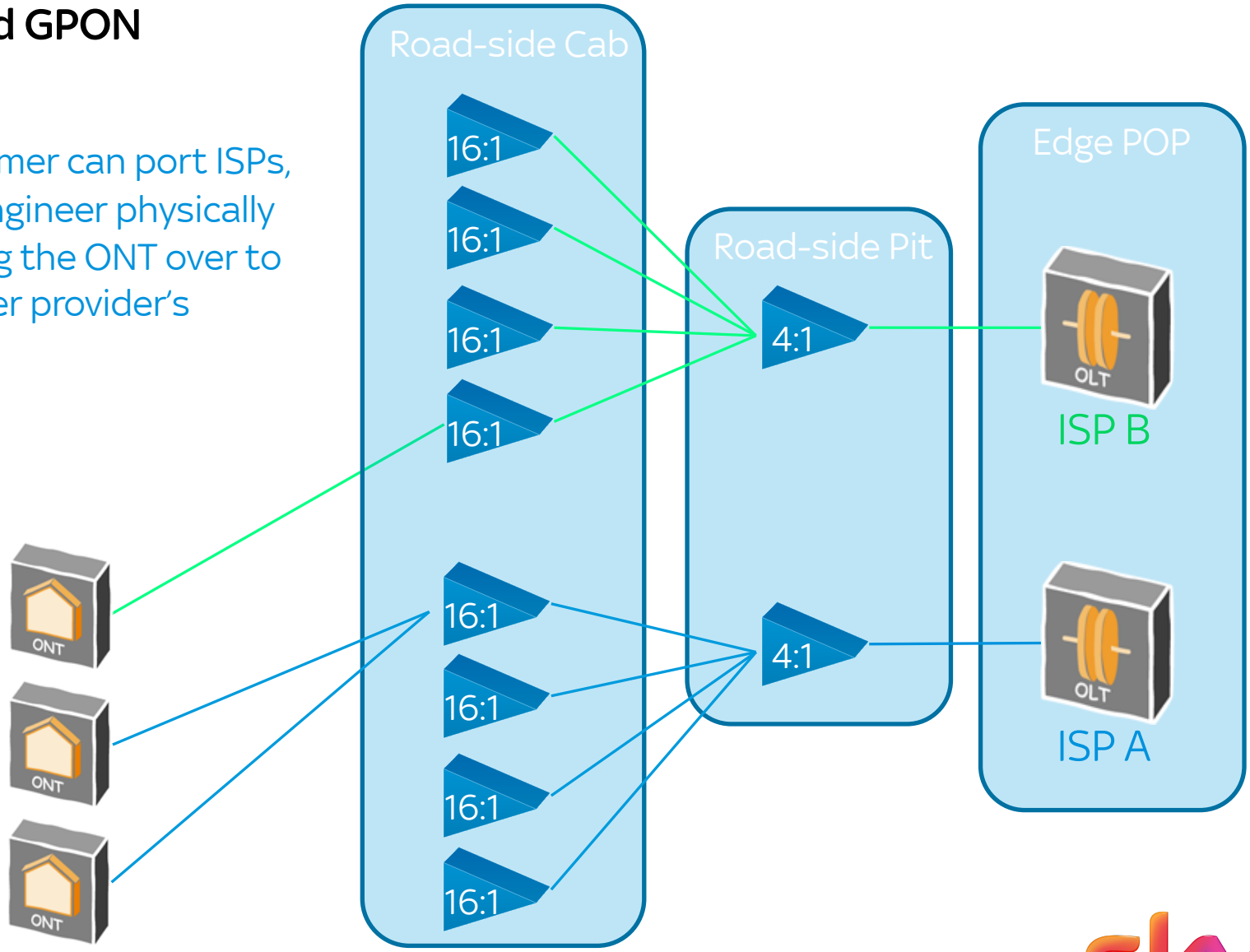
- Distinct Optical Distribution Network (ODN)
- Shared ducts / fibre bundles



Access

Unbundled GPON

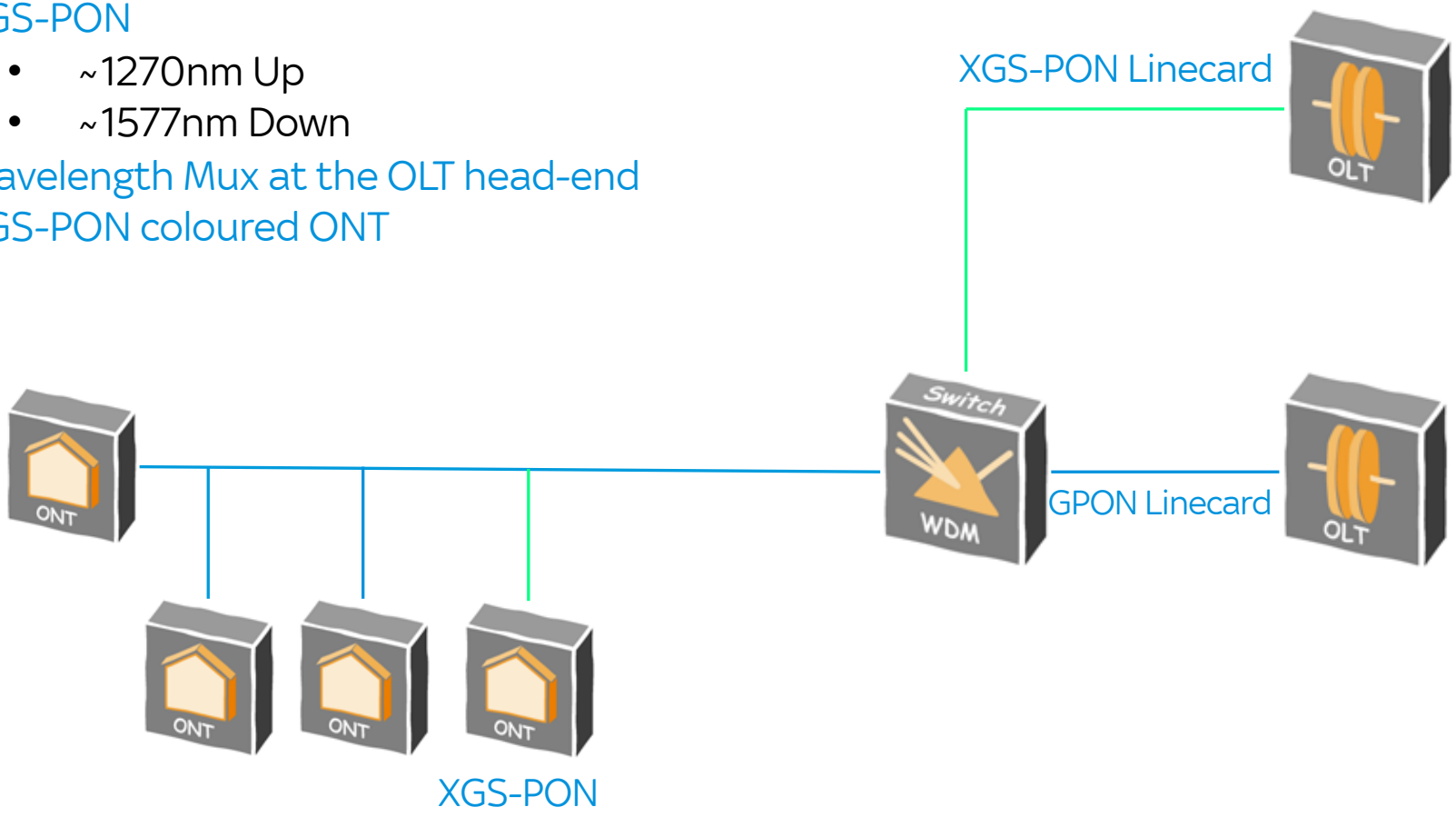
- A Customer can port ISPs, by an engineer physically patching the ONT over to the other provider's splitter



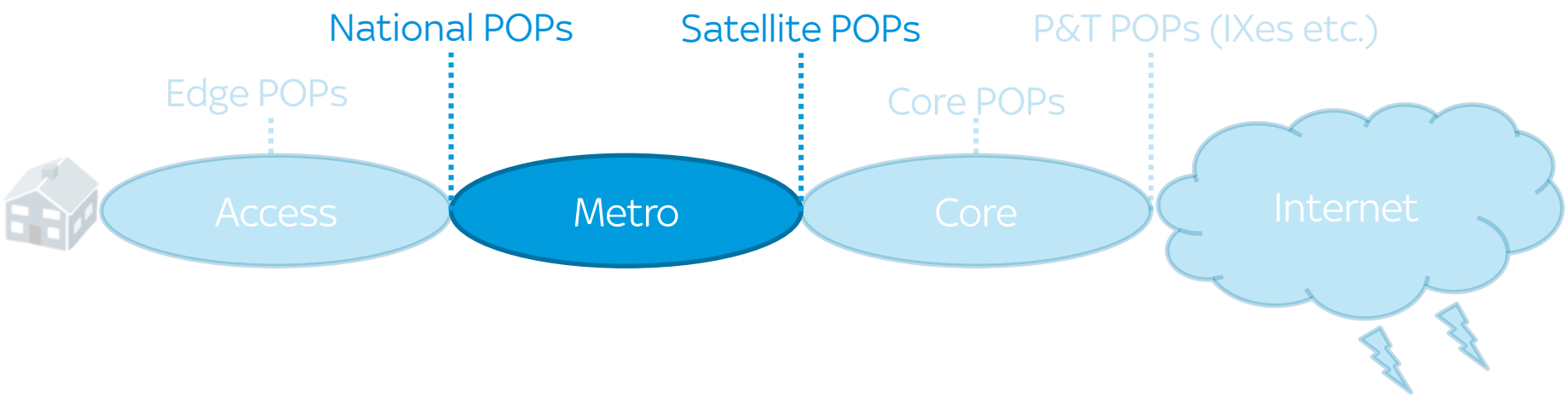
Access

Future-proofed: 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



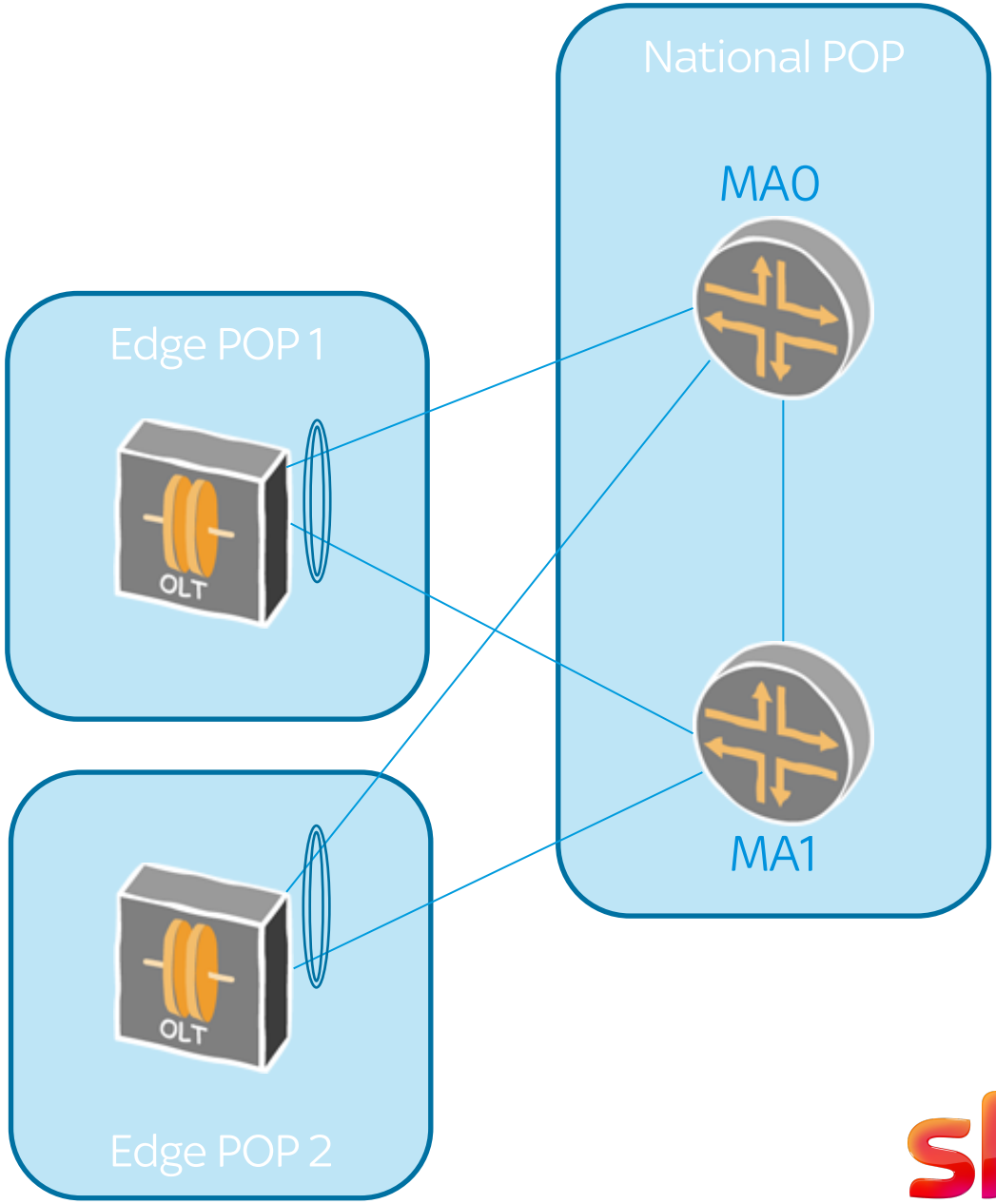
Metro Network



Metro Network

Metro Aggregation

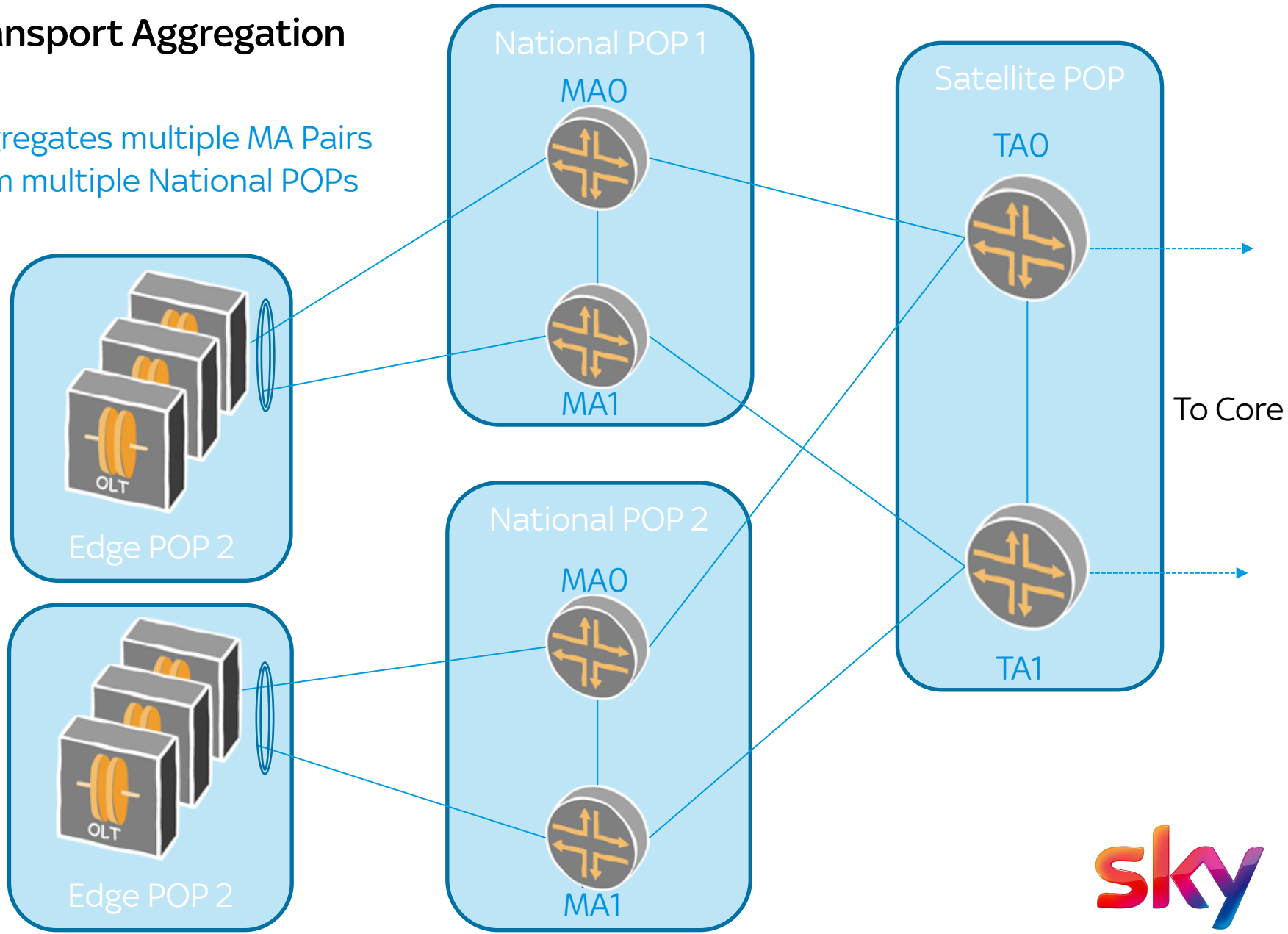
Aggregates multiple OLTs
from multiple Edge POPs



Metro Network

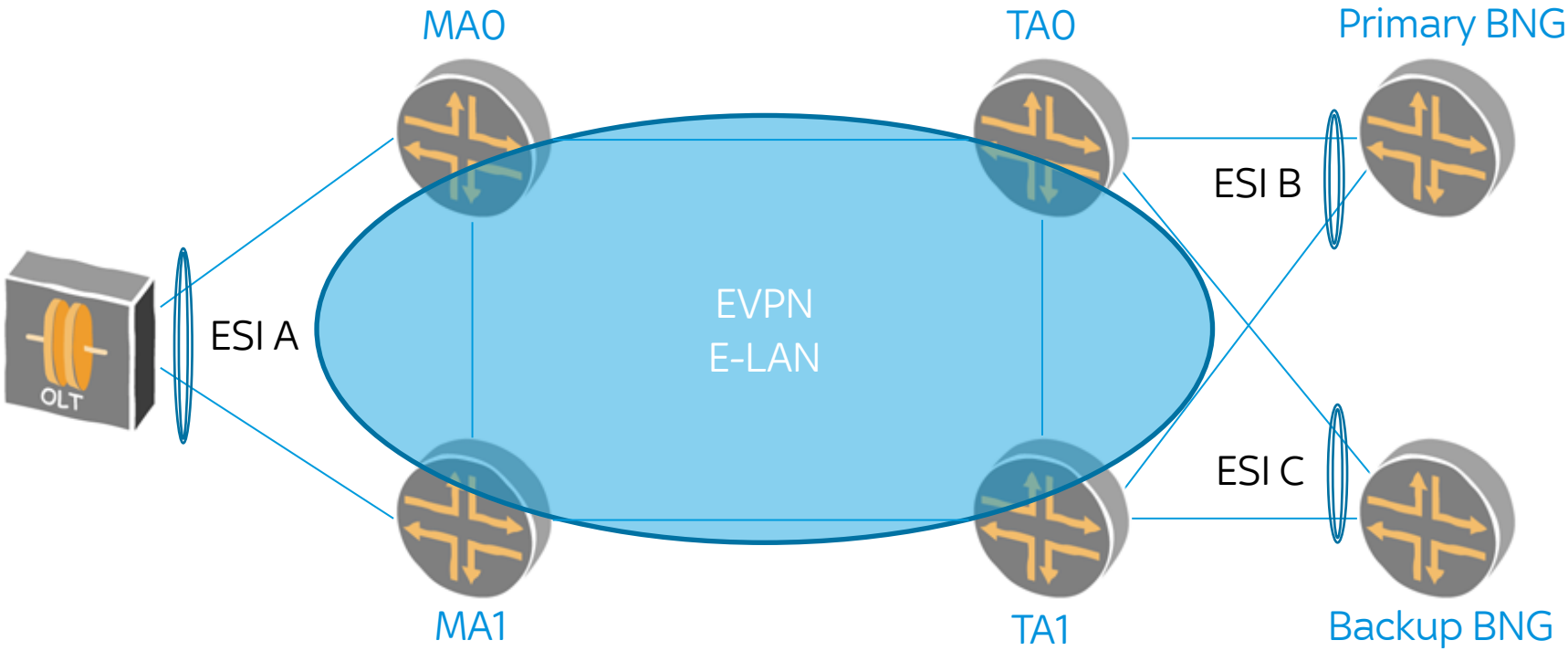
Transport Aggregation

Aggregates multiple MA Pairs
from multiple National POPs



Metro Network

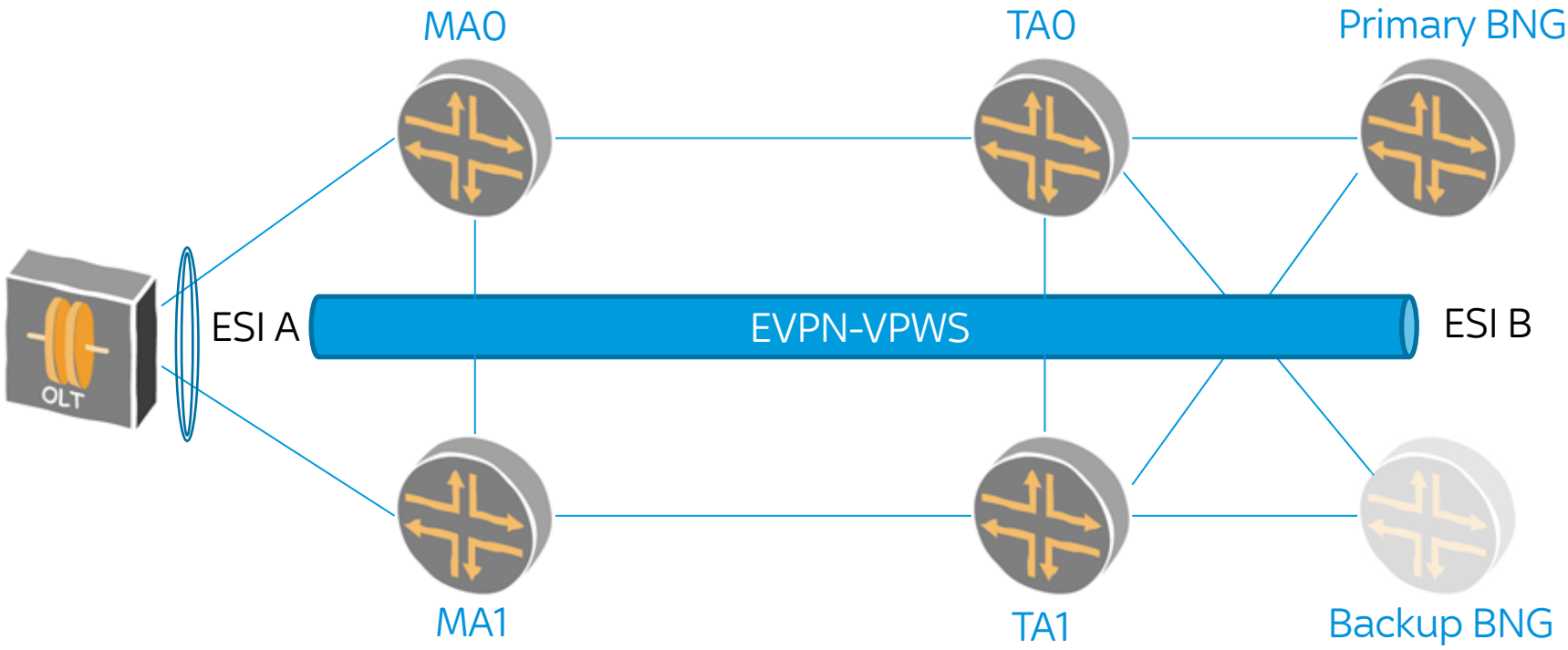
EVPN E-LAN



- BGP-Signaled L2VPN
- Efficient MAC-learning
- Active/Active Multi-homed
- No need for MC-LAG or STP
- Efficient handling of BUM packets

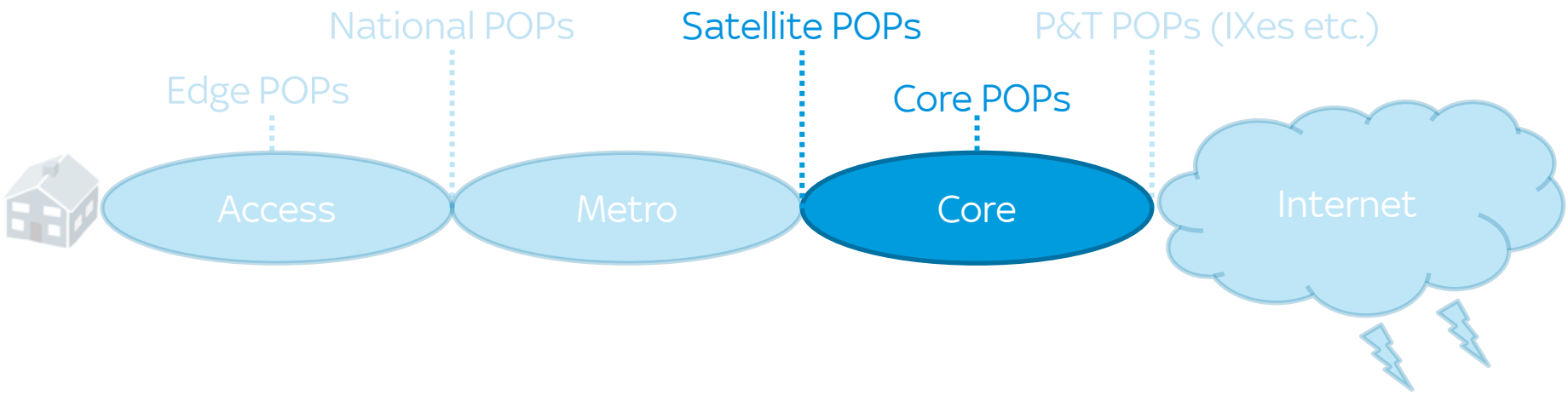
Metro Network

Future: EVPN-VPWS



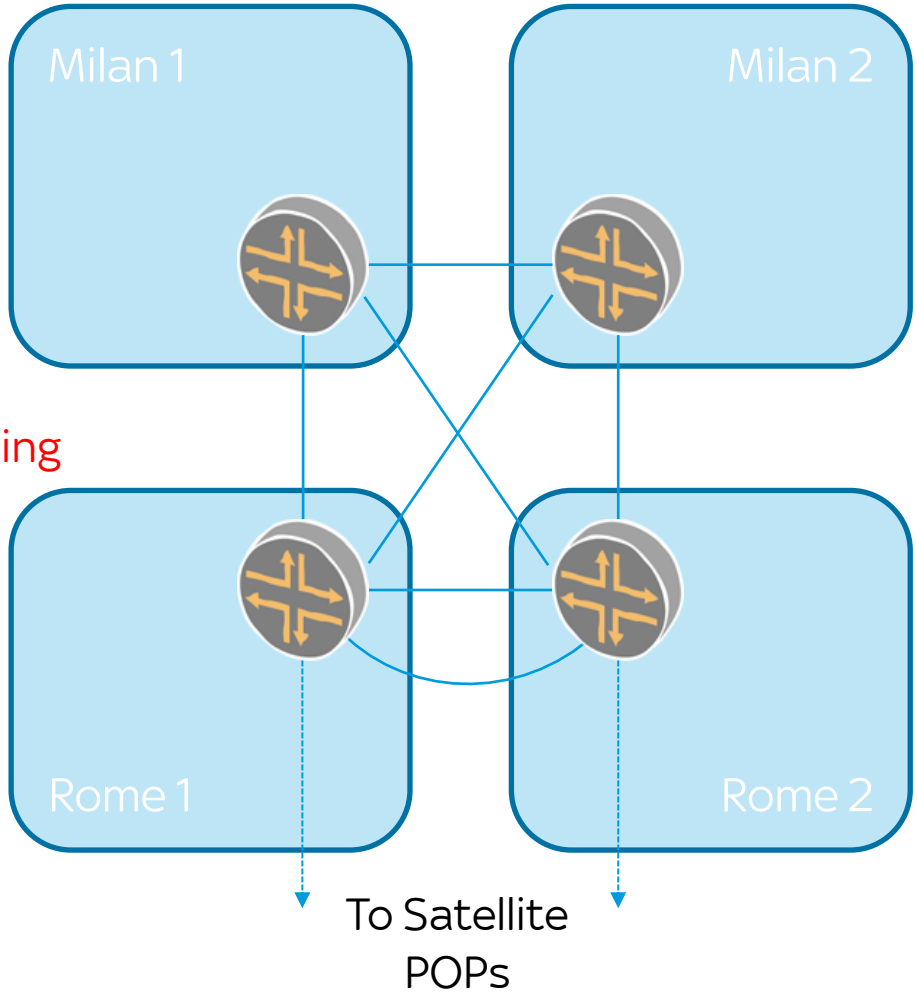
- Point-to-Point
  - No MAC-learning
- Single-Active multi-homing on BNG side
  - Uses the "Backup" control flag (RFC8214 §3.1)

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

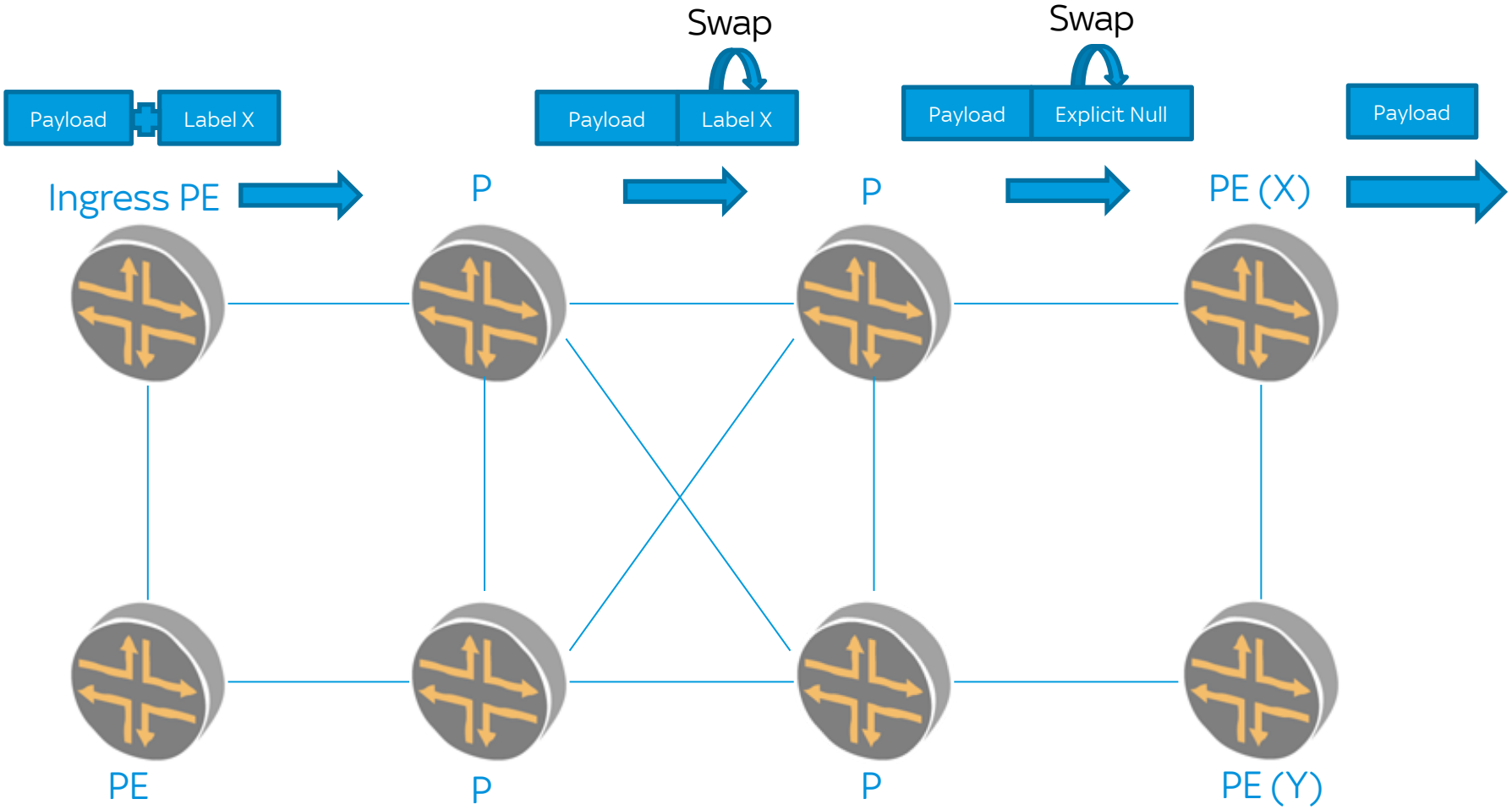


Segment Routing (MPLS)

- Uses existing MPLS packet forwarding
- Extensions to IGP for label distribution
  - Not a distinct new “protocol”
  - No need for LDP or RSVP-TE
- 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)

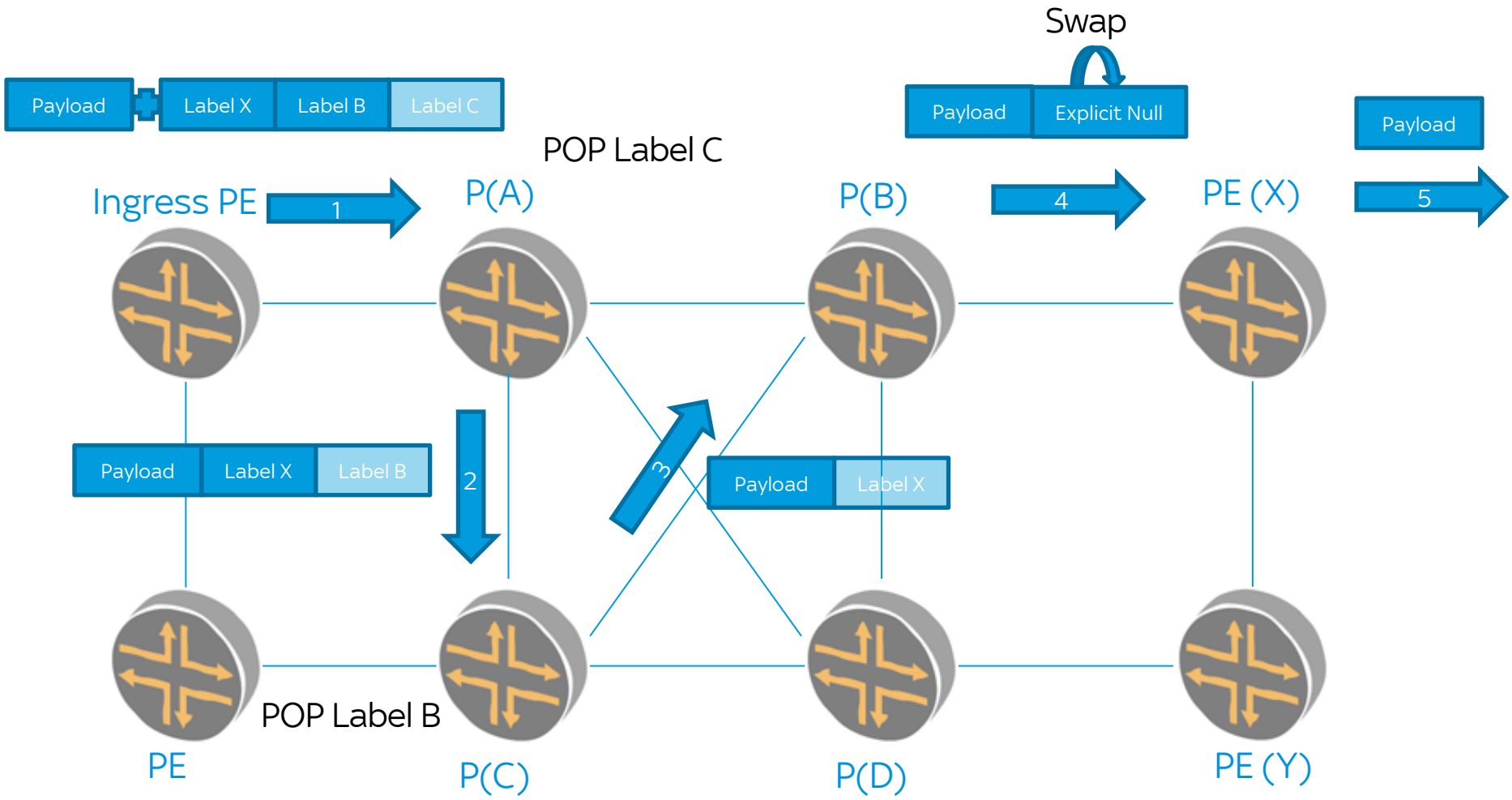
Segment Routing

Shortest Path

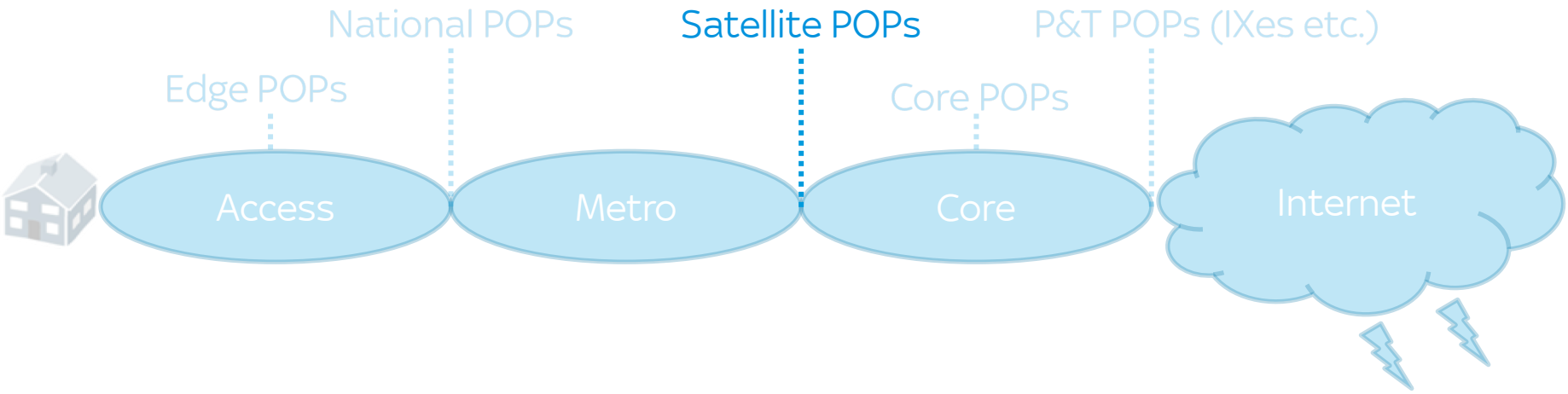


Segment Routing

Traffic Engineering

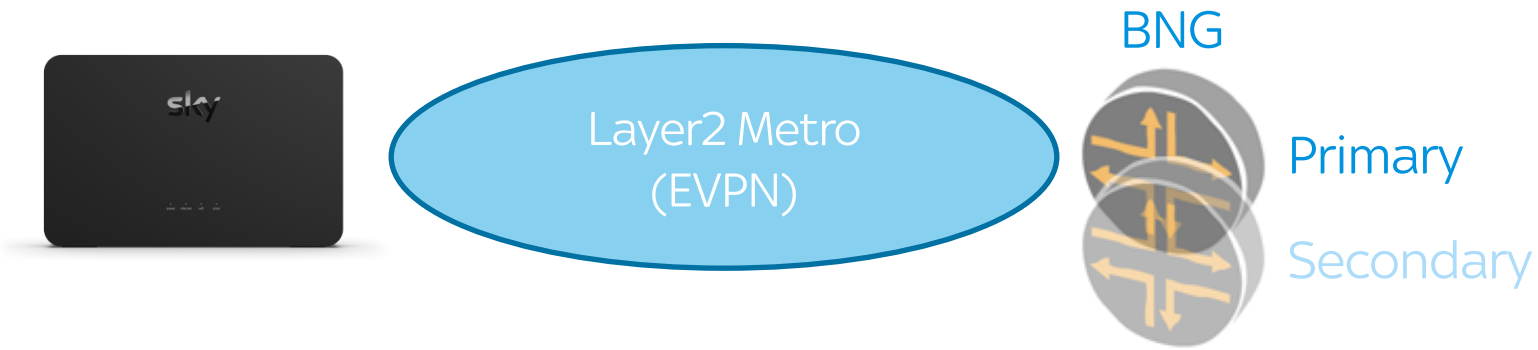


Subscriber Termination



Subscriber Termination

Broadband Network Gateways (BNG)



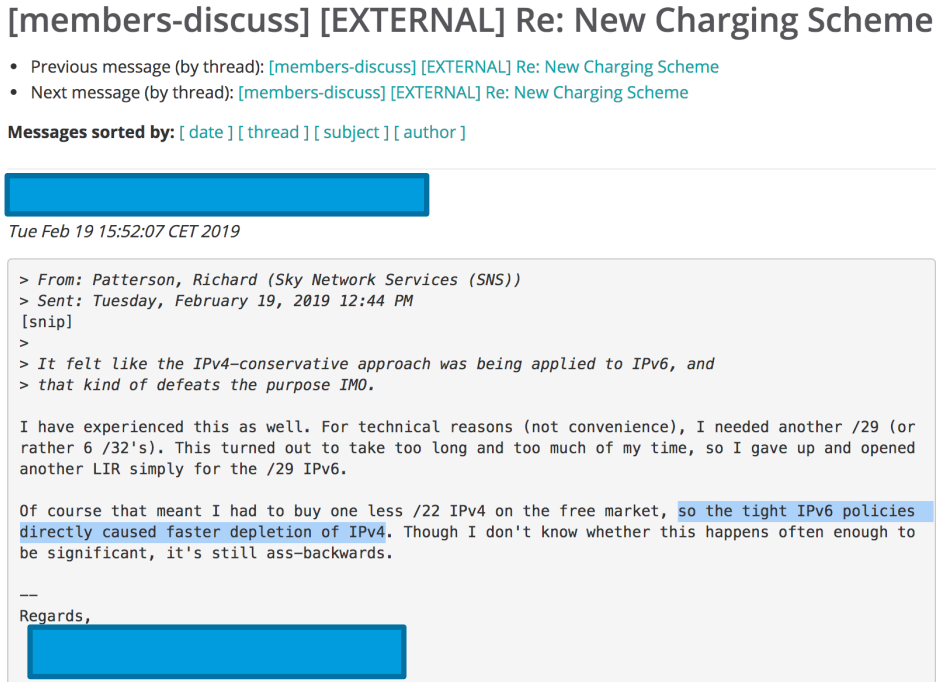
- IPoE – Less encapsulation overhead compared to PPPoE
- Port-based Authentication – DHCPv6 Option 37 Remote-ID
    - Inserted by the OLT's Lightweight DHCPv6 Relay Agent (LDRA)
- Redundant BNG – Proprietary vendor-magic session-state syncing, plus VRRP
- Native IPv6 - /48 Prefix Delegations

Future:

- Subscriber Termination directly on EVPN-VPWS
    - Vendor-magic glues BNG backup state with EVPN state
    - "Single-Active" bit set in the ESI Label Extended Community

IPv6 Addressing

- RIPE NCC allocates an LIR up to a /29 without question
- Enough for ~500K subscribers @ /48 PD
  - As recommended in RIPE-690 BCOP
  - Not enough for our projected growth
- >/29 available with questions
  - Lots of questions
  - So many questions
  - Some RIPE NCC members decide that spinning up new LIRs is more efficient



[members-discuss] [EXTERNAL] Re: New Charging Scheme

- Previous message (by thread): [members-discuss] [EXTERNAL] Re: New Charging Scheme
- Next message (by thread): [members-discuss] [EXTERNAL] Re: New Charging Scheme

Messages sorted by: [date] [thread] [subject] [author]

[redacted]

Tue Feb 19 15:52:07 CET 2019

> *From: Patterson, Richard (Sky Network Services (SNS))*
> *Sent: Tuesday, February 19, 2019 12:44 PM*
[snip]
>
> *It felt like the IPv4-conservative approach was being applied to IPv6, and*
> *that kind of defeats the purpose IMO.*

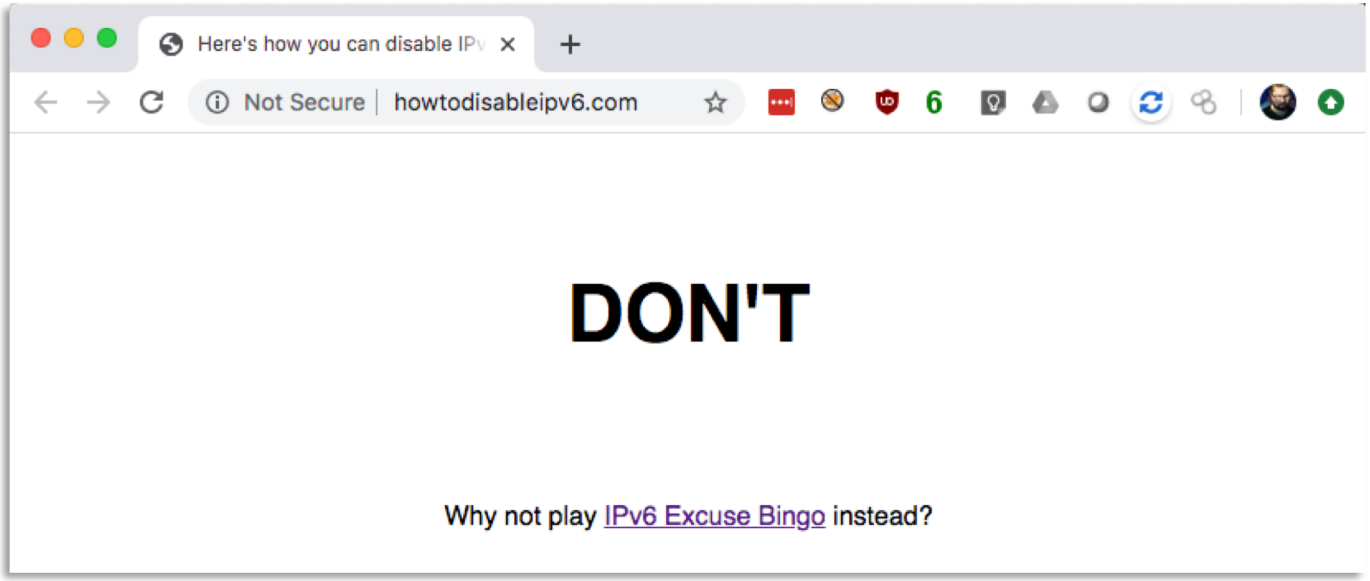
I have experienced this as well. For technical reasons (not convenience), I needed another /29 (or rather 6 /32's). This turned out to take too long and too much of my time, so I gave up and opened another LIR simply for the /29 IPv6.

Of course that meant I had to buy one less /22 IPv4 on the free market, so the tight IPv6 policies directly caused faster depletion of IPv4. Though I don't know whether this happens often enough to be significant, it's still ass-backwards.

--
Regards,
[redacted]

IPv6 Addressing

- IPv6-only Recursive DNS
- OTT Voice delivered over IPv6



IPv6 is no longer an optional feature

IPv4 Addressing

- Started with zero IPv4 addresses
- Join RIPE as a Local Internet Registry (LIR)
  - €2000 joining fee + €1400pa
  - ~~Gets you a /22 or 1024 IPv4 addresses~~
    - Last of RIPE's IPv4 allocated on 25th November 2019
    - Maybe this will be enough for infrastructure?
- Buy off the open market @ >$20USD / IP
- Still need some form of IPv4 address sharing
  - We've chosen MAP-T (RFC7599)

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

Block Size*	/24	/23	/22	/21	/20	/19	/18	/17	/16
Price/IP (USD)	30.00	26.00	22.00	20.00	20.00	20.00	20.00	20.00	21.00

**Prices in this table are based on the most recent transfer agreed upon by a previous seller and buyer. Each price by block size is indicative of current seller expectation, and buyers can reference these prices to support internal planning and budgeting.*

Source: https://ipv4marketgroup.com/broker-services/buy/
23rd Dec 2019

Mapping of Address and Port (MAP)

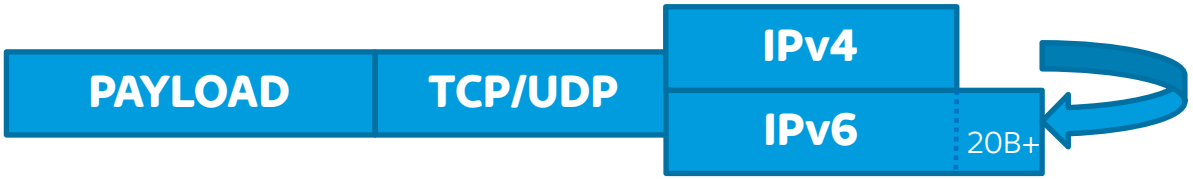
- IPv4aaS
  - IPv4 delivered via IPv6 transport
  - IPv6-only Access Layer
  - Reduced operational overhead
- Allows IPv4 address sharing, or 1:1
- No DNS synthesizing required
- Doesn't require an agent on end-hosts.
- Can operate in either encapsulation or translation modes
- **Stateless**

Mapping of Address and Port (MAP)

RFC7597: MAP-E

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

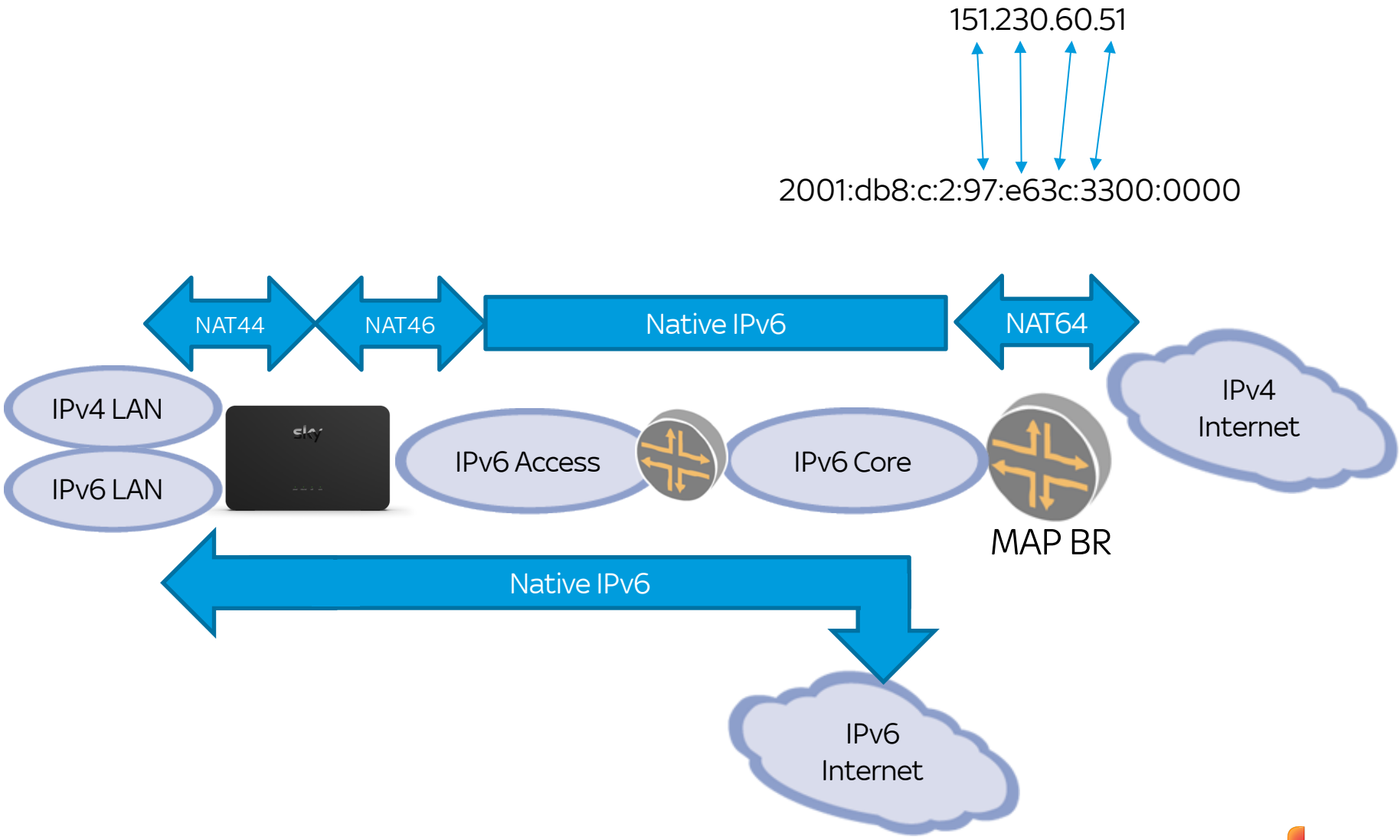
RFC7599: MAP-T



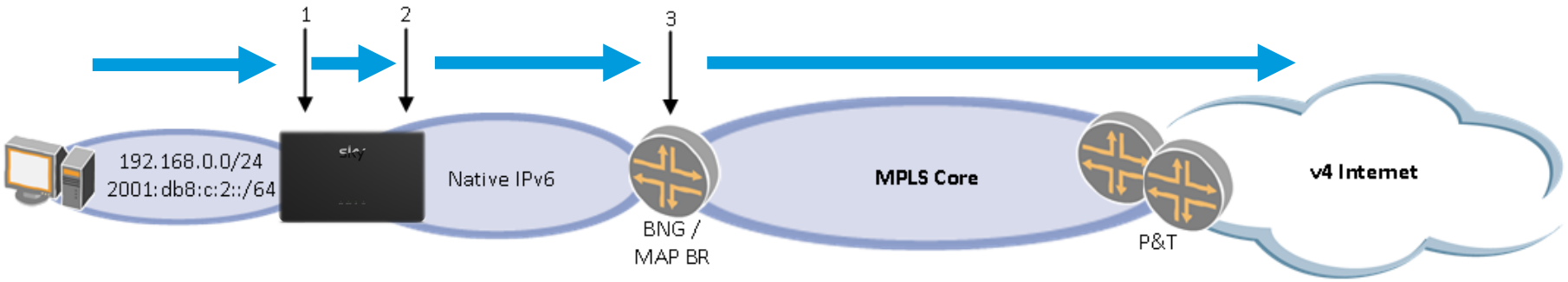
- Translation
  - Less per-packet overhead (not zero!)
  - 5-tuple hashing. (E.g., with ECMP or over LAGs)
  - Border relay-bypass
  - Loses IPv4-only header attributes

Packet Flow (MAP-T)

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



Life of a Packet (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

Algorithmically translated/encapsulated based on predefined rules

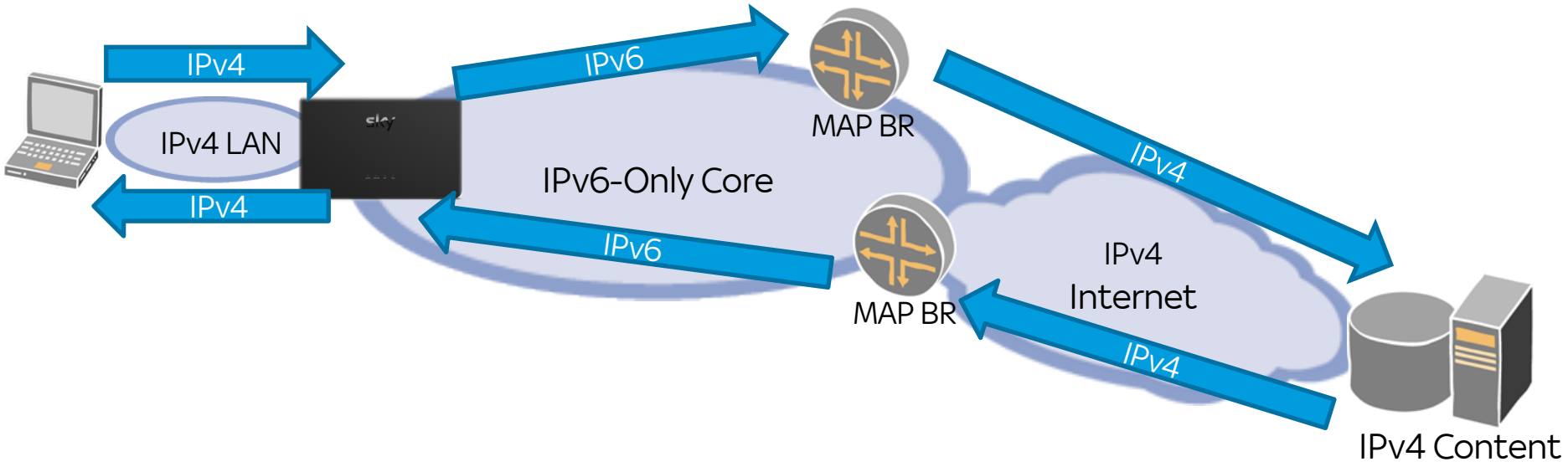
- No need to keep track of every flow
- Efficient packet processing
- Cheaper, more scalable hardware
- Already supported on existing linecards by some vendors

- IPv4 and DHCPv6 pool management becomes complicated
- Some jurisdictions regulate 5-tuple logging for compliance reasons, which stateless IPv4aaS methods don't provide

sky

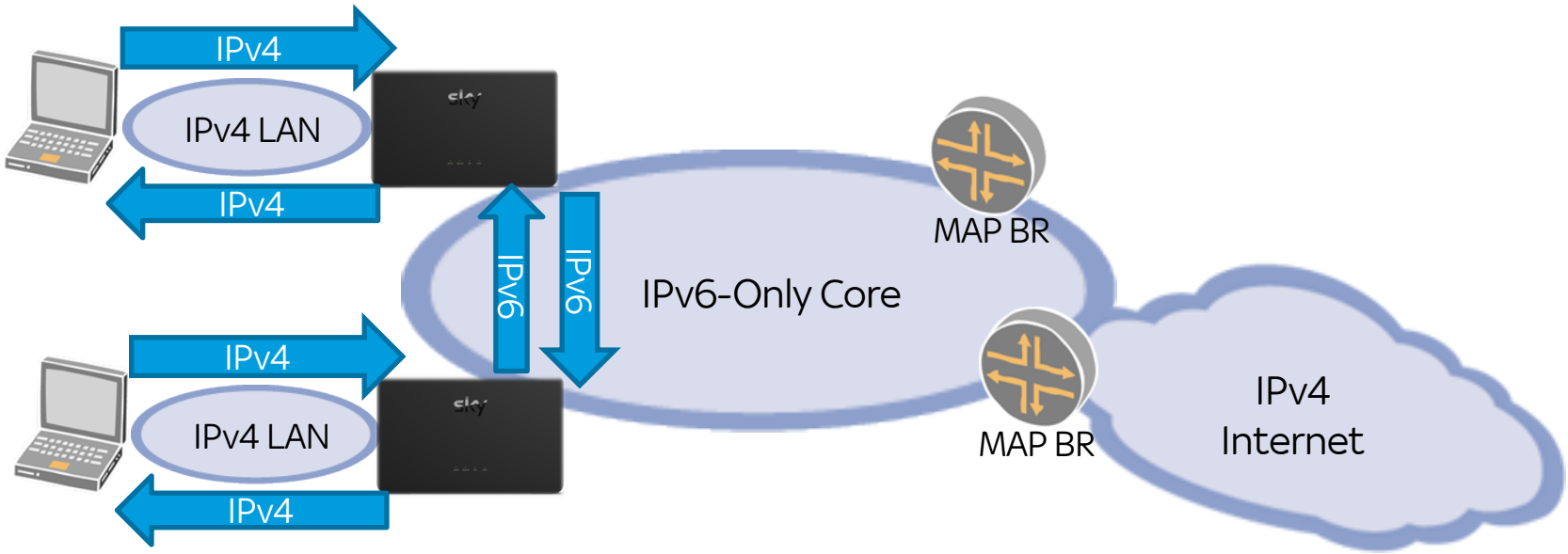
MAP Border Relay Anycasting

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



MAP Forward Mapping Rules

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



MAP Border Relay Bypass for CDN

- On-net content servers could 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.