



# IPv6-only with IPv4aaS (MAP-T)

UK IPv6 Council Day – 19<sup>th</sup> Nov 2024

~5 Years Ago.....



#### UKNOF45 - Building a Greenfield Fixed-line ISP in 2019



UKNOFconf  
2.11K subscribers

Subscribed

32



Share

Download

Clip

Save



2,145 views 19 Jan 2020 [ETC.VENUES 155 BISHOPSGATE](#)

Speaker: Richard Patterson (Sky)

<http://uknof.uk/45/>

An architectural overview of how Sky Italia's broadband network was built, the technologies used, and the decisions made based on previous learnings.

# What is IPv4aaS and MAP-T?

## Mapping of Address and Port using Translation

“IPv4-as-a-Service” (IPv4aaS)

- Provides IPv4 connectivity using IPv6 transport
- Allows IPv4 address sharing

## Comparison with other IPv4aaS

|                             | 464XLAT                         | MAP-T                | MAP-E                | lw4o6                | DS-Lite              |
|-----------------------------|---------------------------------|----------------------|----------------------|----------------------|----------------------|
| Data Plane (NAT464 or 6in4) | <b>Translation</b>              | <b>Translation</b>   | <b>Encapsulation</b> | <b>Encapsulation</b> | <b>Encapsulation</b> |
| NAT64                       | <b>Stateful</b>                 | <b>Stateless</b>     | <b>N/A</b>           | <b>N/A</b>           | <b>N/A</b>           |
| NAT46 (CPE)                 | <b>Stateless<br/>(optional)</b> | <b>Stateless</b>     | <b>N/A</b>           | <b>N/A</b>           | <b>N/A</b>           |
| IPv4 Address Sharing        | <b>Stateful CGN</b>             | <b>Stateless A+P</b> | <b>Stateless A+P</b> | <b>Stateless A+P</b> | <b>Stateful CGN</b>  |

# Encapsulation vs Translation

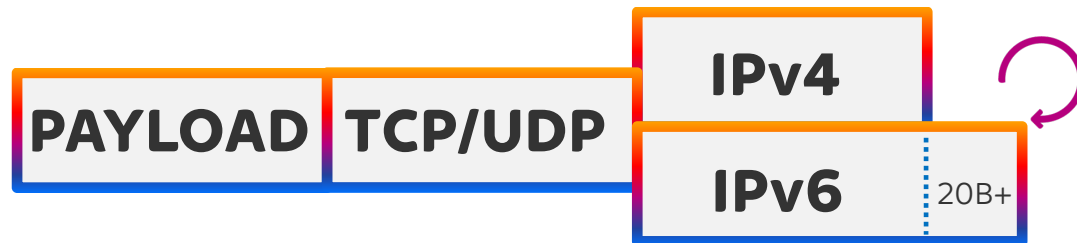
## RFC7597: MAP-E

- Encapsulation
  - Larger per-packet overhead. (40 bytes)
  - IPv4 header remains intact.

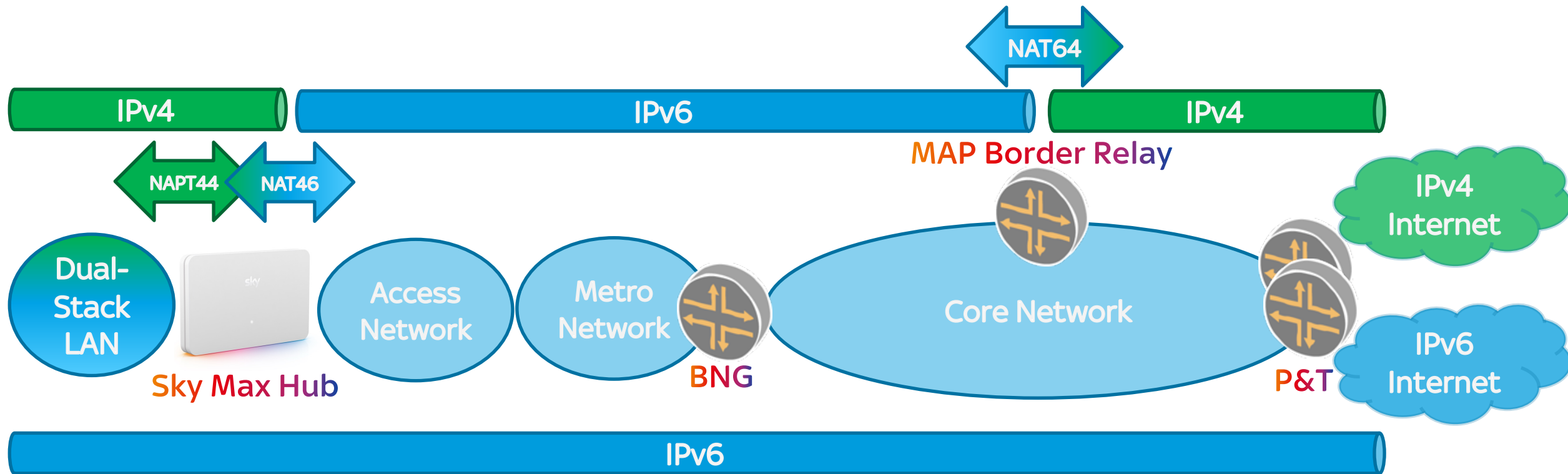


## RFC7599: MAP-T

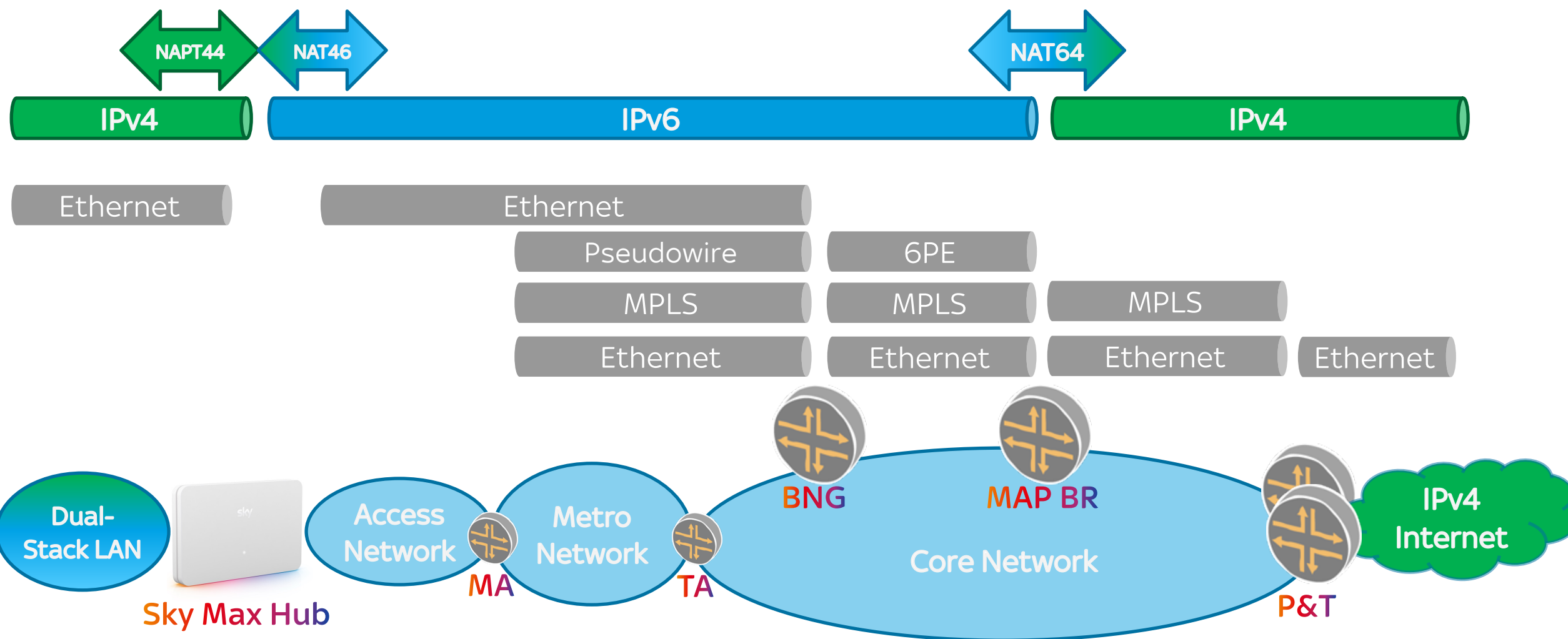
- Translation
  - Fewer bytes of overhead. (20 bytes)
  - Loses IPv4-only header attributes.
  - 5-tuple hashing.
  - Border relay-bypass.



# Sky UK – MAP-T Topology



# Sky UK – MAP-T Transport

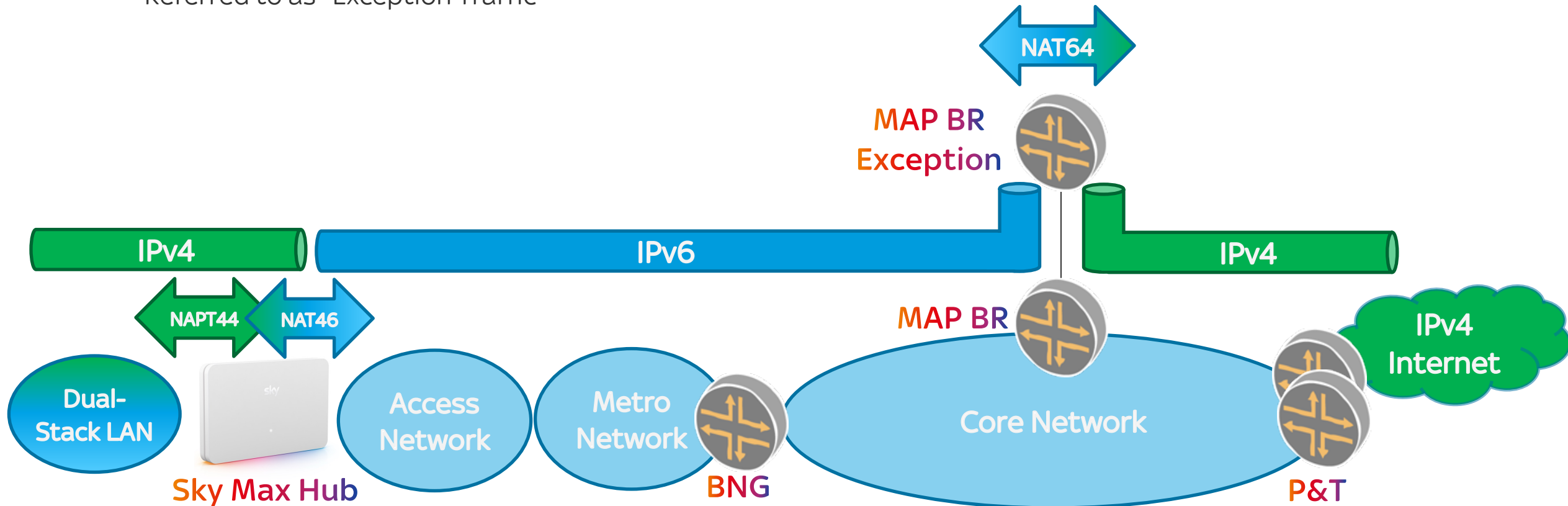




# MAP-T Exception Border Relays

## 2x Box Border Relay Solution

- >99% of traffic translated by a Cisco ASR9903 based on Lightspeed+ ASIC
- <1% of traffic translated by a Cisco Catalyst 8500 based on 3<sup>rd</sup> gen Quantum Flow Processor (QFP)
  - Referred to as “Exception Traffic”



# Possible Exception Traffic

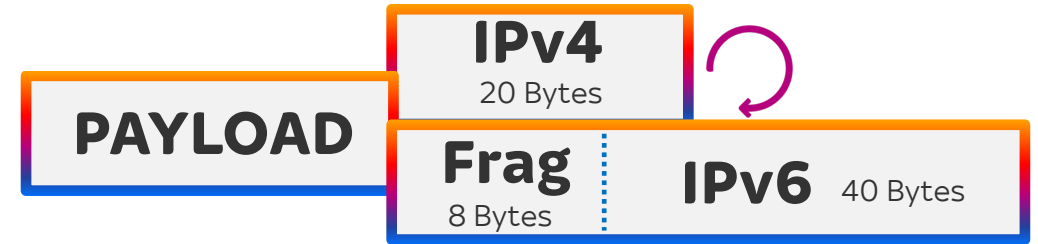
|   | ASR9K (LSP+ ) | Cat8500   | Notes:   |
|---|---------------|-----------|--|
| Translation of Non-TCP/UDP/ICMP layer 4 protocols: <ul style="list-style-type: none"> <li>GRE (IP proto 47)</li> <li>ESP (IP proto 50)</li> <li>AH (IP proto 51)</li> <li>IP-in-IP (IP proto 4)</li> <li>L2TP (IP proto 115)</li> </ul> | ✓             | ✓         | When configured without IPv4 address sharing       |
| Translation of ICMPv4: <ul style="list-style-type: none"> <li>Type 0 &amp; 8: Echo &amp; Replies</li> <li>Type 3: Destination Unreachable (Code 0 - 4; incl. Fragmentation Needed)</li> <li>Type 11: Time Exceeded</li> </ul>           | ✓             | ✓         |  |
| Translation of ICMPv6: <ul style="list-style-type: none"> <li>Type 1: Destination Unreachable (Code 0 - 4)</li> <li>Type 2: Packet Too Big</li> <li>Type 3: Time Exceeded</li> <li>Type 128 &amp; 129: Echo &amp; Replies</li> </ul>    | ✓             | ✓         | Fragmented echo request & replies on C8500 roadmap |
| Generation of ICMPv4 Type 3 Code 4 Fragmentation Needed when resulting IPv6 packet is larger than IPv6 MTU & IPv4 DF=1  | ✓             | Not Req'd |  |
| Fragmentation of packets when resulting IPv6 packet is larger than IPv6 MTU & IPv4 DF=0   | ⤵ Redirect    | ✓         |  |
| Translation of IPv4 fragments   | ⤵ Redirect    | ✓         |  |
| Translation of IPv6 fragments   | ⤵ Redirect    | ✓         |  |
| Translation of UDP4 packets with zero checksum.   | ✓             | ✓         | Both calculate new csum                            |
| Adjust TCP MSS value  | Road-mapped   | Not Req'd |  |
| Translation of IPv4 Packets with IP Options   | ✗             | ✗         |  |



# MTU & Fragmentation

## IPv6 Fragment Header (8 bytes)

- Added to fragmented packets
  - 1528 byte IPv6 MTU required to avoid fragmentation



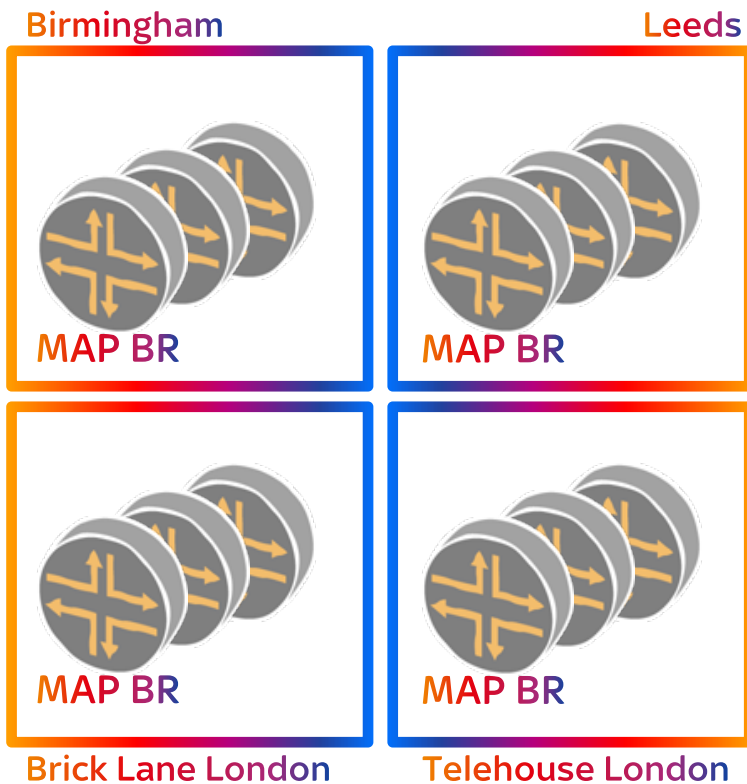
## Openreach GEA

- Officially<sup>[1]</sup> only supports 1500 byte IP payloads.
  - Unofficially 1900 byte packets are forwarded through most DSLAM/OLTs excluding ECI.
- Fragmented packets within the MAP domain
  - Additional processing on CPE (upstream) & Border Relays (downstream)
- Can be avoided with:
  - TCP Max Segment Size (MSS) adjusting on CPE/BR; &
  - Path MTU Discovery (PMTUD)

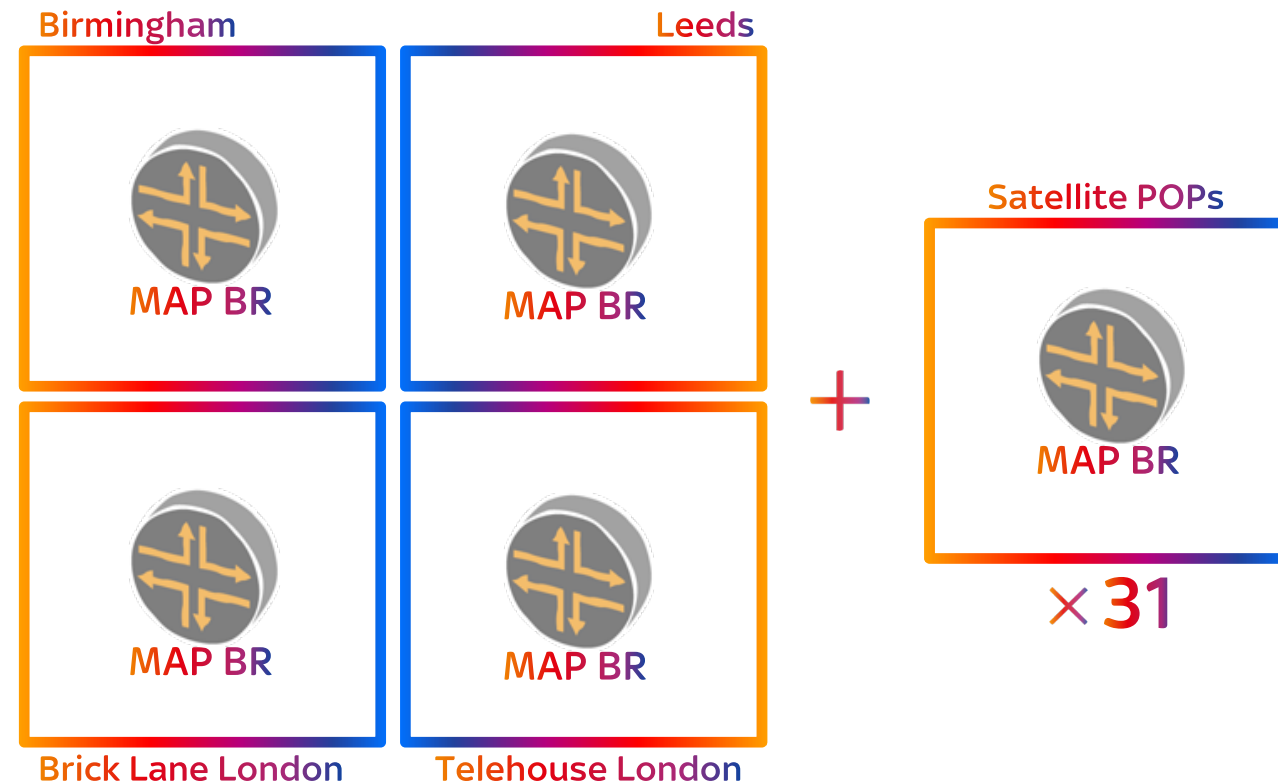
# Centralised or Distributed

MAP Border Relays have initially been deployed at 4x centralised “Supercore” POPs

- They can be scaled horizontally at the 4 central POPs; or
- Additional BRs can be distributed to 31x “Satellite” POPs
  - Where BNGs also exist



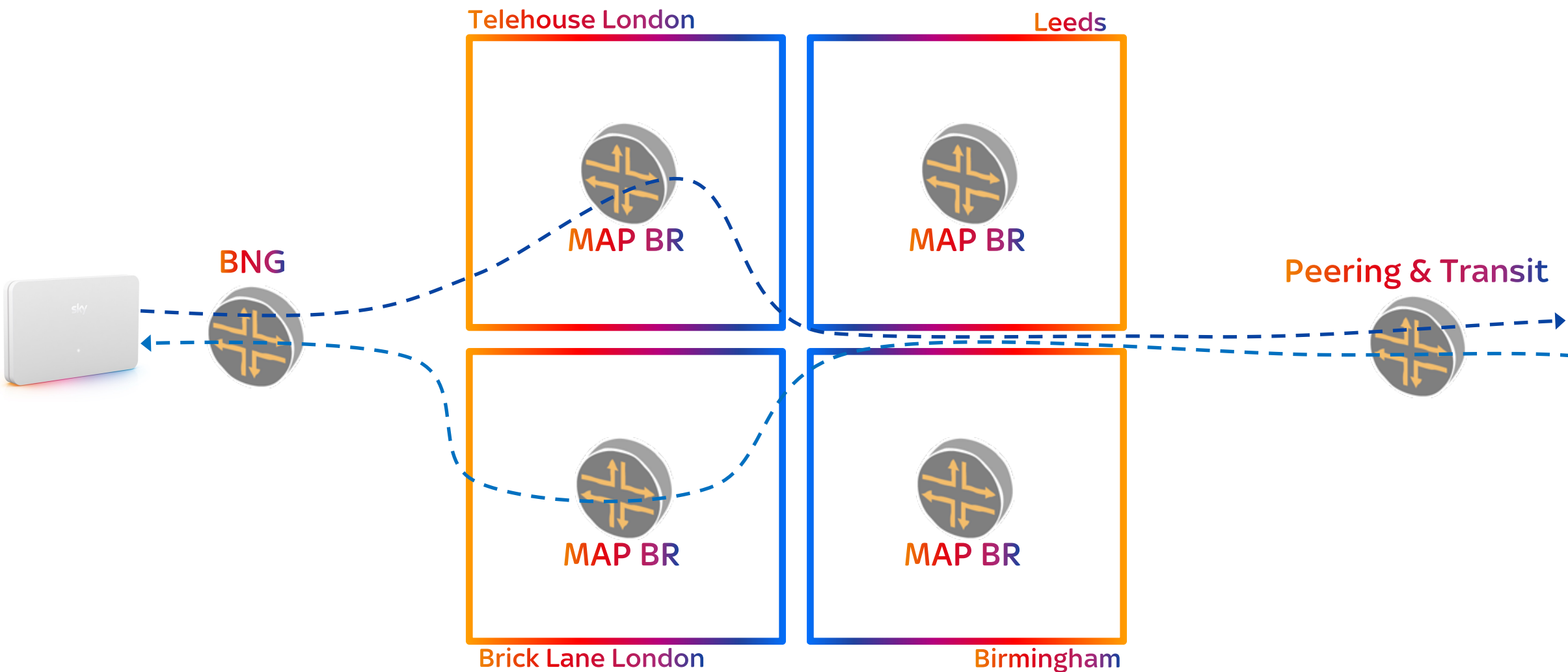
Or



# Anycast

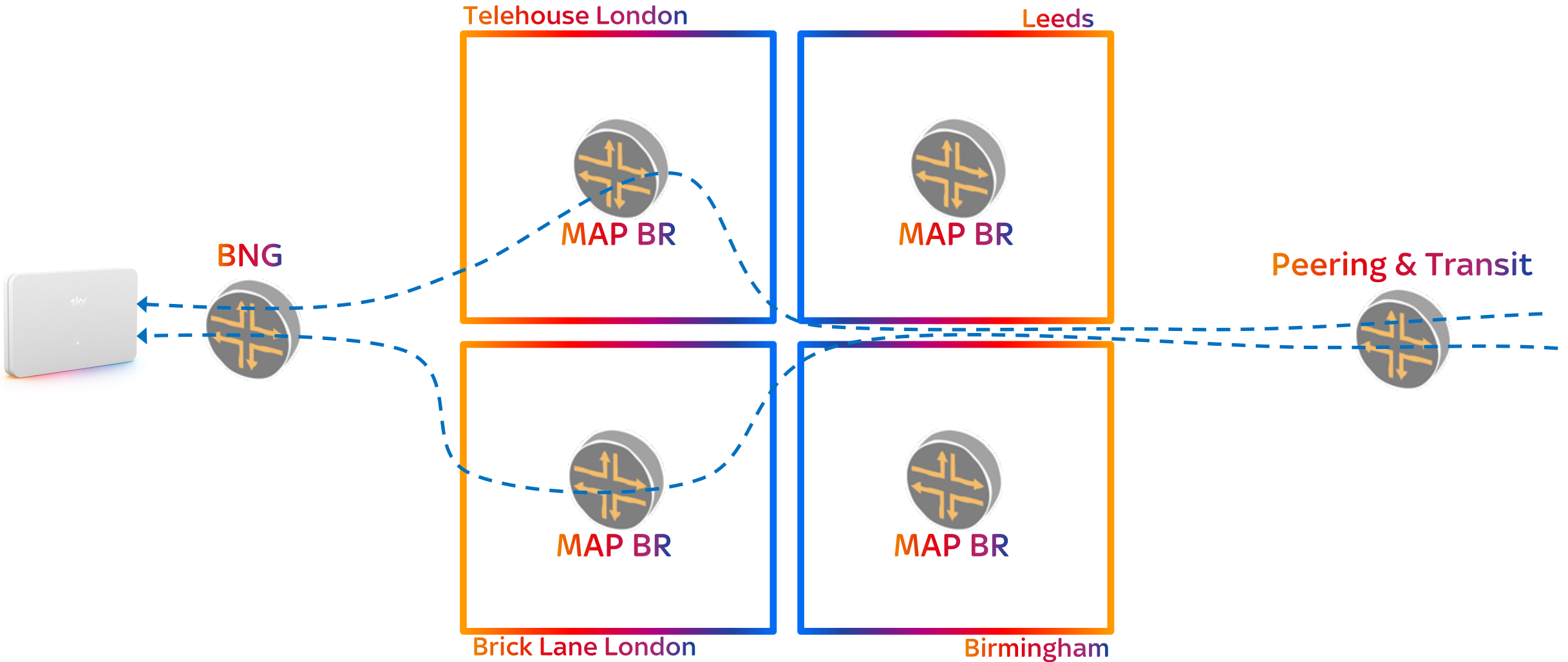
(Resilience & Asymmetry)

- All IPv4 & IPv6 prefixes are anycast from all border relays
- Any border relay can translate any packet
- Ingress & Egress flows may go via different border relays



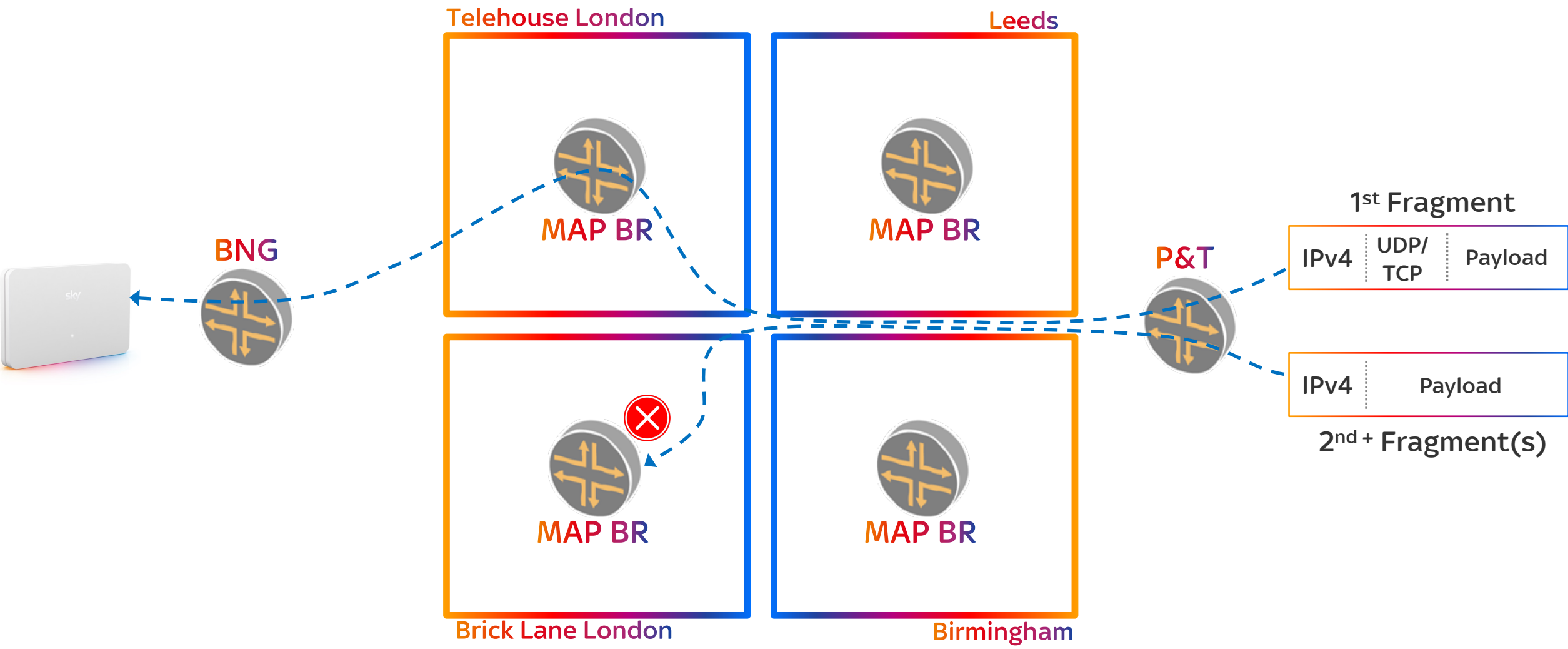
# Load-balancing (Equal Cost Multi-Path)

Leveraging anycast and ECMP, a subscriber may have different flows hashed via multiple border relays



# Hashing of Fragments

L3-only hashing to avoid fragments arriving out-of-order, or on entirely different BRs, preventing translation.



# BNG <-> MAP Border Relay Mapping Examples

Whilst any centralized MAP BR can translate packets for a subscriber on any BNG, routing policies are applied to prefer certain MAP BRs based on BNG location and POP parenting

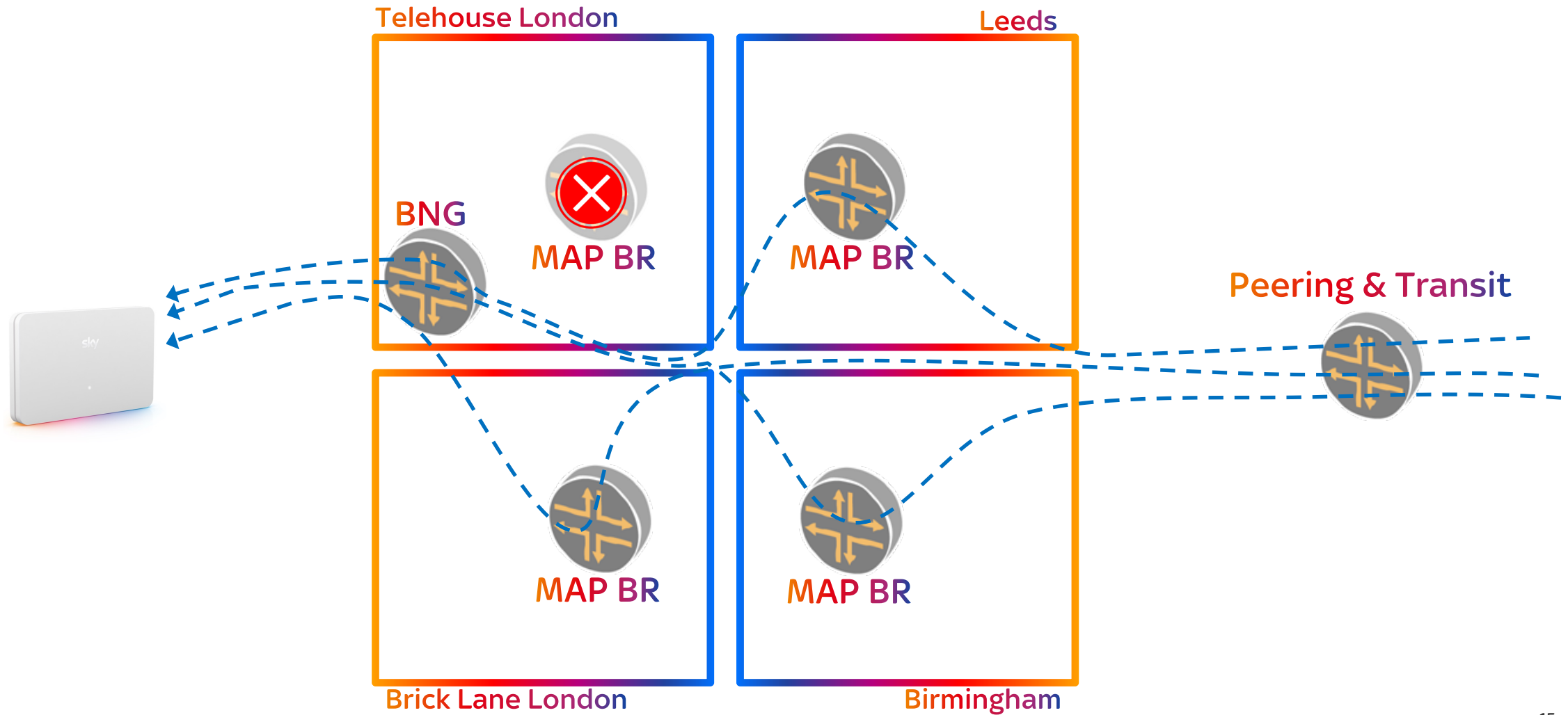
|                | BNG Location | Primary BR(s)        | Secondary BR(s) | Tertiary BR(s)      |
|----------------|--------------|----------------------|-----------------|---------------------|
| Supercore POPs | BLLON        | BLLON                |                 | THLON, ENLBA, HOBIR |
|                | THLON        | THLON                |                 | BLLON, ENLBA, HOBIR |
|                | ENLBA        | ENLBA                |                 | THLON, ENLBA, HOBIR |
|                | HOBIR        | HOBIR                |                 | BLLON, THLON, HOBIR |
| Satellite POPs | ENHMI        | ENHMI <sup>[1]</sup> | BLLON, THLON    | ENLBA, HOBIR        |
|                | ENBEL        | ENBEL <sup>[1]</sup> | BLLON, ENLBA    | THLON, HOBIR        |
|                | ENBAS        | ENBAS <sup>[1]</sup> | THLON, ENLBA    | BLLON, HOBIR        |
|                | ENEDI        | ENEDI <sup>[1]</sup> | ENLBA, HOBIR    | BLLON, THLON        |
|                | Etc.         |                      |                 |                     |

<sup>[1]</sup> Planned future expansion, distributed model



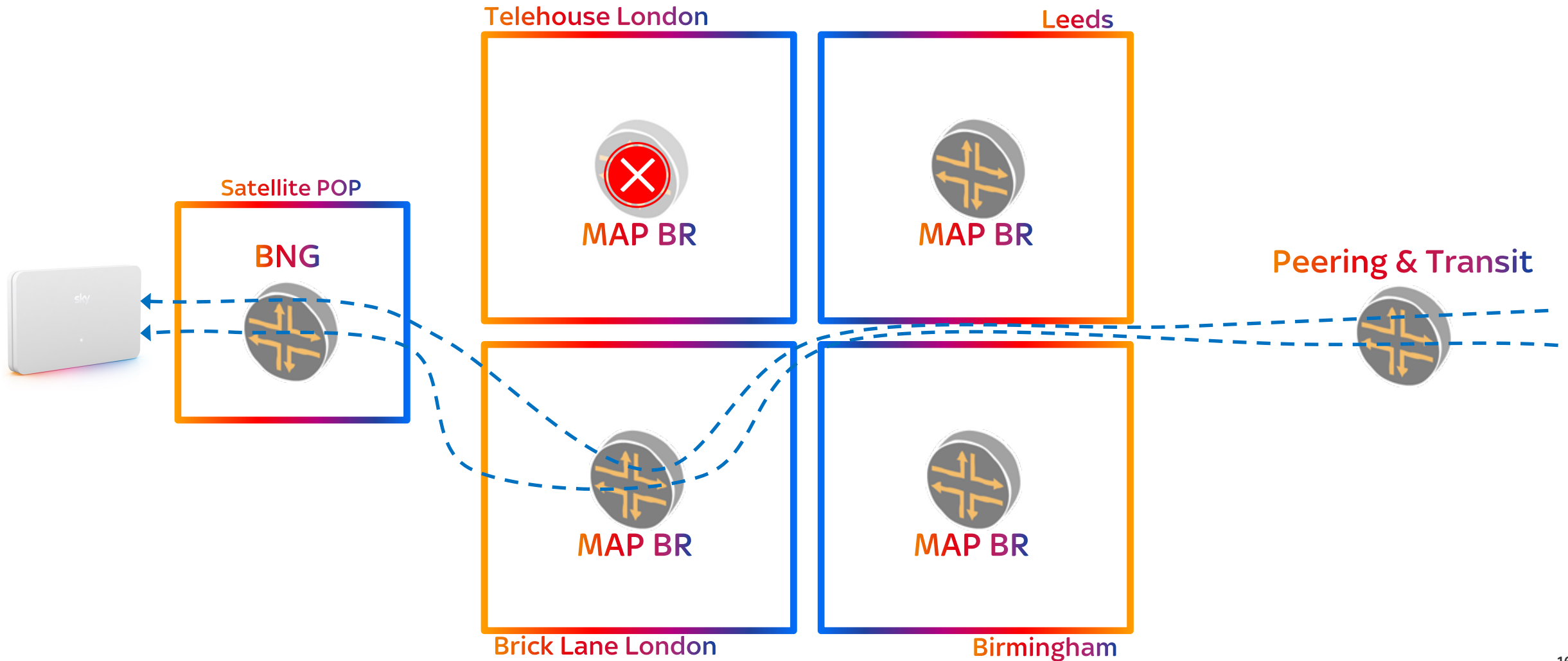
# Failure Modes

## Supercore POP BNG w/ Centralised Model



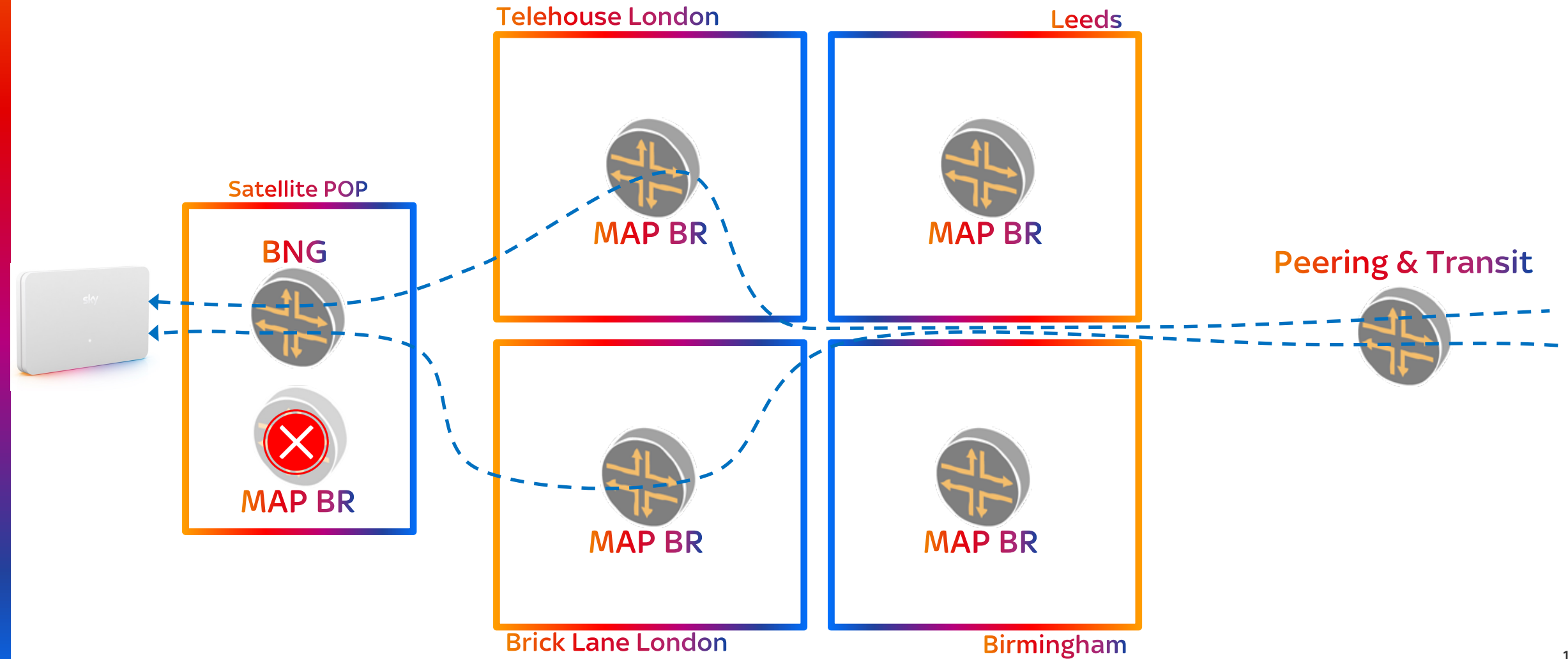
# Failure Modes

## Satellite POP BNG w/ Centralised Model

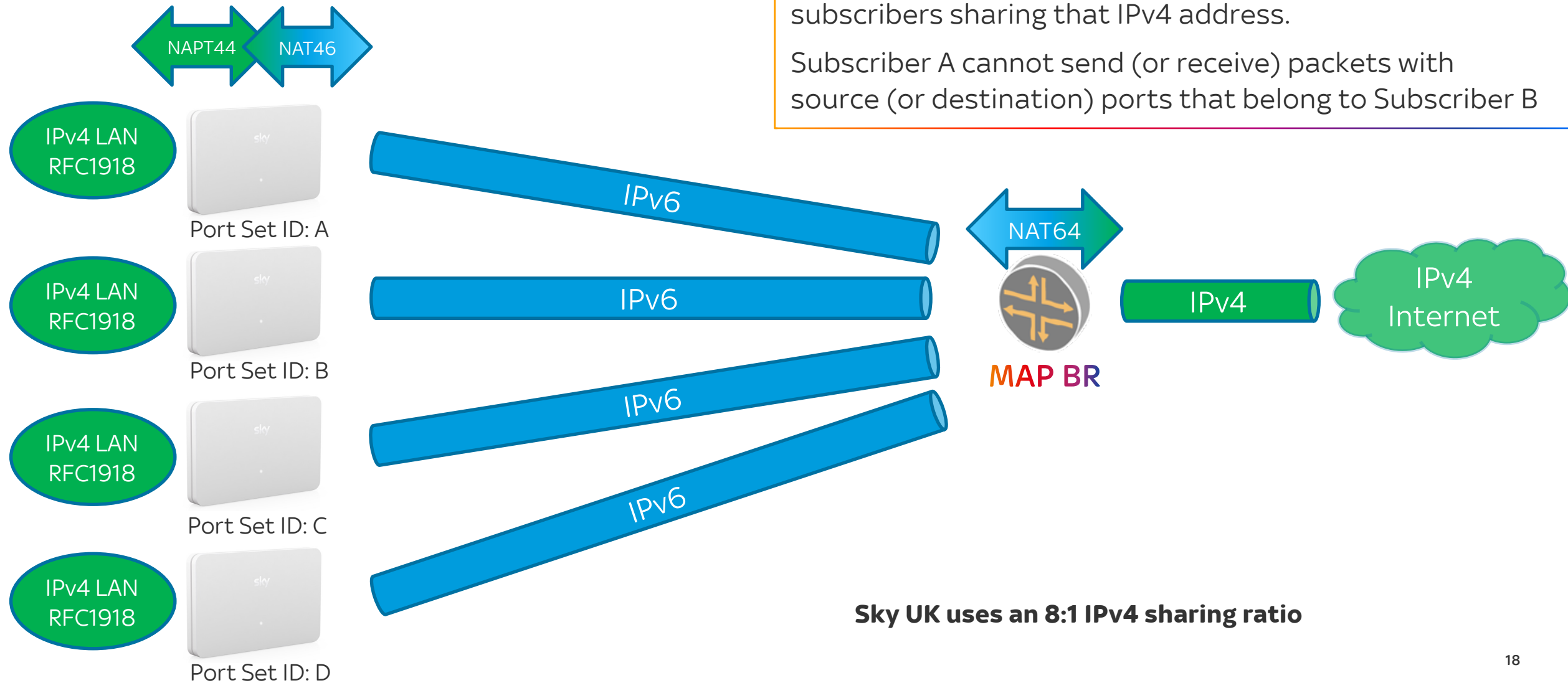


# Failure Modes

## Satellite POP BNG w/ Distributed Model



# IPv4 Address Sharing



# IPv4 Address Sharing Opt-out

IPv4 address sharing is enabled by default for all [MAP-T] subscribers.

Subscribers may be opted-out of IPv4 address sharing by two methods, proactively, or reactively.


## Proactive Opt-out

To minimise impact to customer experience, we listen for notifications sent by the Sky Hub or the My Sky app, when a user enables a known-incompatible feature.

- Universal Plug-n-Play (UPnP)
  - Game consoles, Peer2Peer apps, etc.
- IPv4 DMZ & Port Forwarding
  - Server hosting
- Port Triggering
  - Obscure feature. Dynamic port forwarding.

## Reactive Opt-out

Following an inbound call or digital journey, cases may be escalated to an agent who can manually opt a subscriber out of IPv4 address sharing.

 Most common user experience issues related to IPv4 address sharing require UPnP or port forwarding anyway.

# Automation of IP Pools

MAP-T replaces forwarding plane complexity, with administration complexity and overhead.

Basic Mapping Rules need to be carefully planned, dimensions and ideally automated to maximise efficiency.

## Previously:

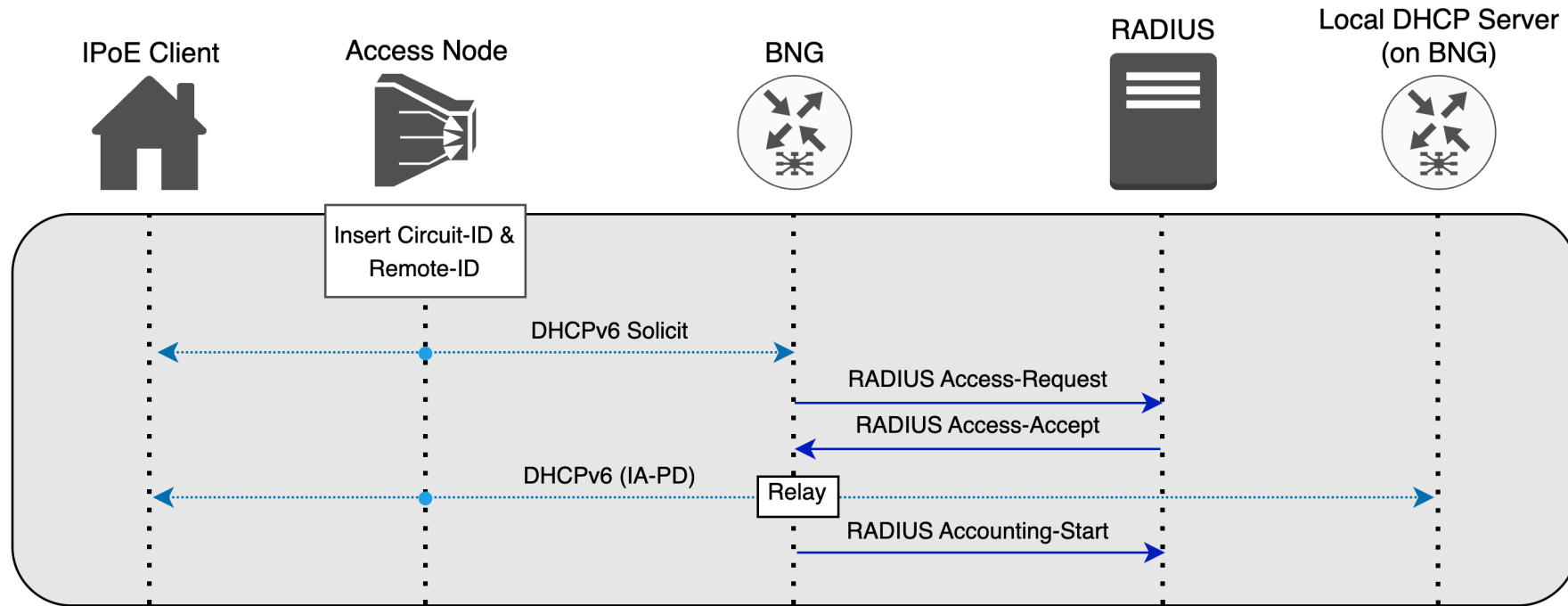
- DHCPv4 pool management was automated with granular prefixes and regular adds & removes to optimise efficiency
- DHCPv6 pool management was **not automated**; large prefixes were overprovisioned manually once upon initial BNG installation.

## With MAP-T:

- IPv4 consumption is directly tied to DHCPv6 pool.
  - Overprovisioning DHCPv6 means overprovisioning IPv4
- Automate DHCPv6 pool managed on BNG
- Automate MAP BMR management on Border Relays



# Authentication – DHCPv6



Sky uses port-based authentication for subscribers, using a string inserted into DHCP messages by Openreach's access nodes. Specifically, into the Remote-ID options, DHCPv6 Option 37 and DHCPv4 Option 82 sub-option 2.

With dual-stack, either DHCPv4 or DHCPv6 can be used for authentication; but MAP-T subscribers [should] only use DHCPv6.

# DHCPv6 Option 95

The Sky Hub 6 is currently Sky UK's only CPE that supports MAP-T; it requests MAP-T by including option code 95, within the Option Request Option (ORO)<sup>[1]</sup> of the DHCPv6 Solicit.

This is how the BNG (instructed by RADIUS) knows that it should give the Sky Hub a DHCPv6 lease with MAP-T rules.

```
> Frame 10: 208 bytes on wire (1664 bits), 208 bytes captured (1664 bits) on interface en12, id 0
> Ethernet II, Src: SkyUk_fe:95:e3 (00:a3:88:fe:95:e3), Dst: IPv6mcast_01:00:02 (33:33:00:01:00:02)
> Internet Protocol Version 6, Src: fe80::2a3:88ff:fefe:95e3, Dst: ff02::1:2
> User Datagram Protocol, Src Port: 546, Dst Port: 547
< DHCPv6
  Message type: Solicit (1)
  Transaction ID: 0xceea84
  > Client Identifier
  > Identity Association for Prefix Delegation
  > Reconfigure Accept
  > Elapsed time
  > User Class
  > Vendor Class
  < Option Request
    Option: Option Request (6)
    Length: 18
    Requested Option code: DNS recursive name server (23)
    Requested Option code: Domain Search List (24)
    Requested Option code: INF_MAX_RT (83)
    Requested Option code: Simple Network Time Protocol Server (31)
    Requested Option code: Vendor-specific Information (17)
    Requested Option code: S46 MAP-T Container (95)
    Requested Option code: SOL_MAX_RT (82)
    Requested Option code: User Class (15)
    Requested Option code: Vendor Class (16)
```

<sup>[1]</sup> [RFC 7598](#) – DHCPv6 Options for Configuration of Softwire and Port-Mapped Clients

# Rogue DHCPv6 Option 95

CPEs that do not support MAP-T should not include DHCPv6 Option 95, however some 3<sup>rd</sup> party CPEs mistakenly do.

CPEs with misbehaving DHCPv6 clients will obtain a DHCPv6 lease with MAP-T, and because we expect them to use MAP-T, a DHCPv4 lease with a non-working IPv4 address, resulting in broken IPv4 connectivity.

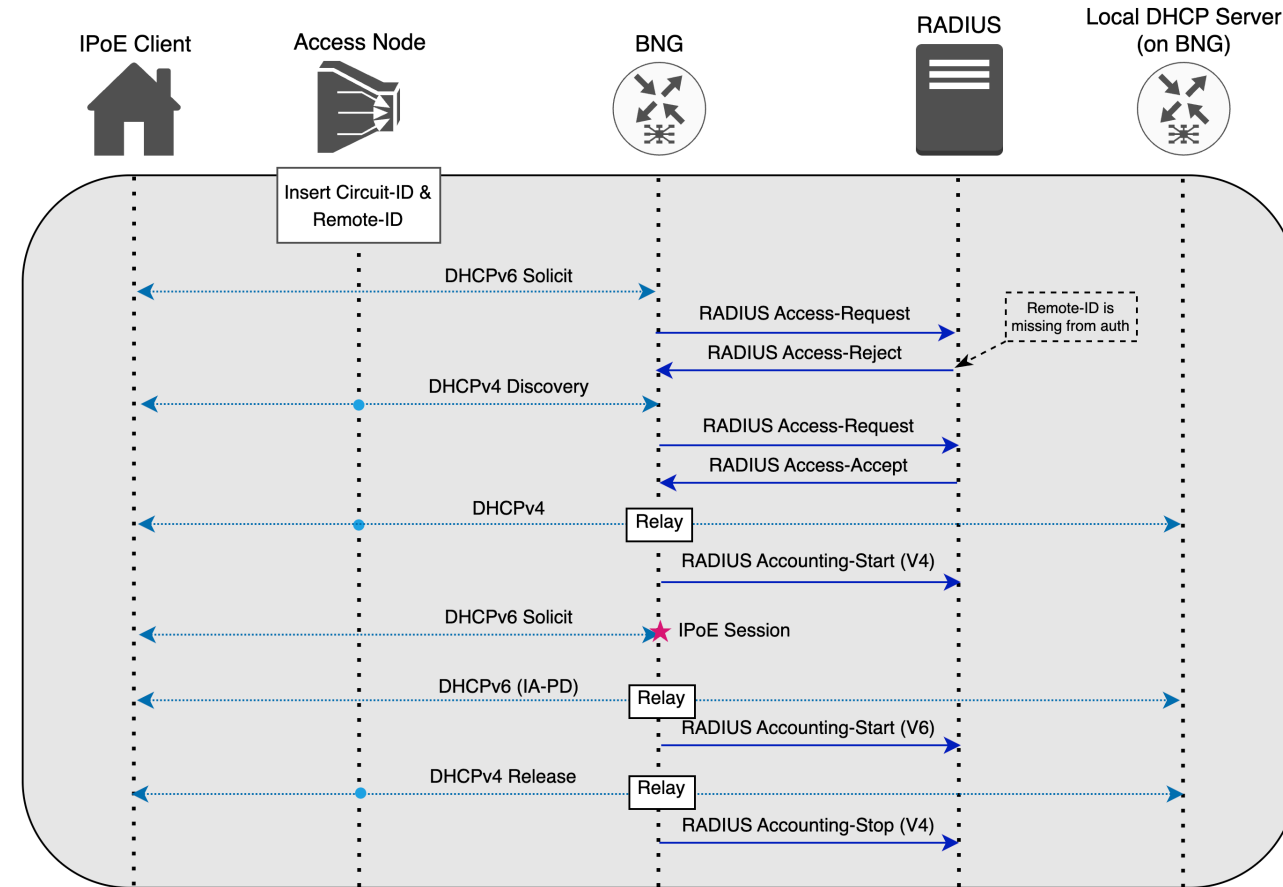
## **Notable examples:**

- OpenWRT's odhcp6c
  - Greedy with options by default
  - User fix: "opkg install map"
  - Long-term fix: Update default behaviour
- Ubiquiti's Unifi routers also use odhcp6c
  - No proper fix. Requires a call center escalation to disable MAP-T; or
  - Disable IPv6 to force DHCPv4-based authentication ☹

# Authentication – DHCPv4

Only ~70% of Openreach's footprint<sup>[1]</sup> supports the LDRA<sup>[2]</sup> feature required to insert Remote-ID into DHCPv6 messages, so the remaining ~30% rely on DHCPv4 to trigger authentication, and to bootstrap a subscriber session on the BNG.

These temporary DHCPv4 leases will be given a private<sup>[3]</sup> **non-usable** IPv4 address from 100.64.0.0/10.



<sup>[1]</sup> Everything but the ECI access nodes

<sup>[2]</sup> Lightweight DHCPv6 Relay Agent

<sup>[3]</sup> [RFC 6598](#) - IANA-Reserved IPv4 Prefix for Shared Address Space

# DHCPv4 Option 60

RADIUS identifies clients with MAP-T support to determine if it should return the MAP-T DHCPv6 pool during authentication, but.....

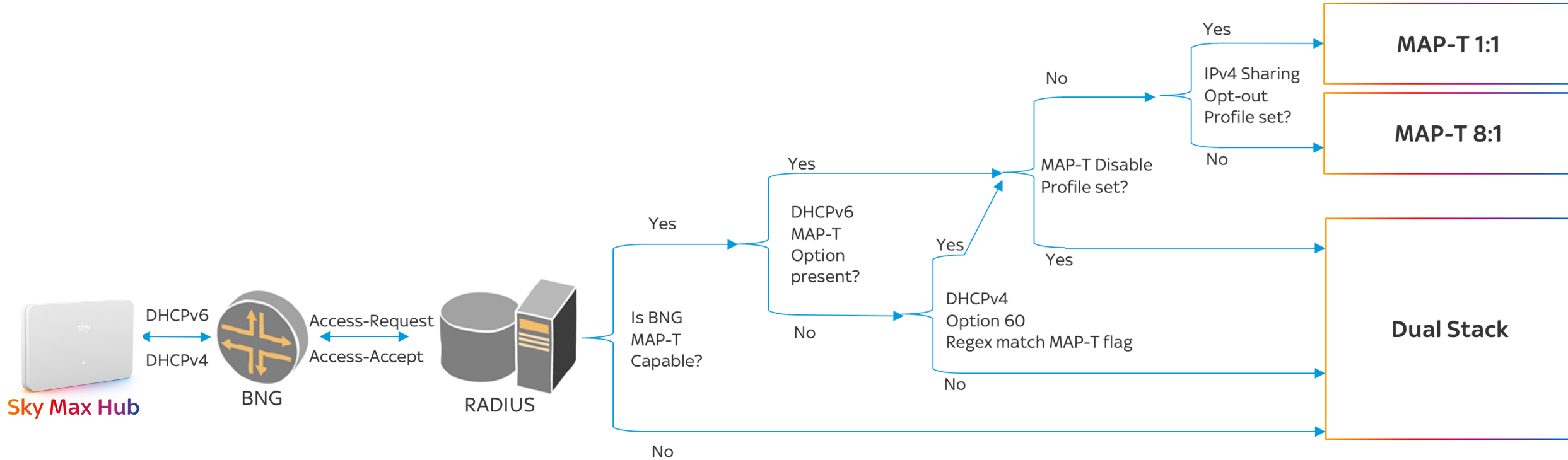
MAP-T option code 95 does not exist in DHCPv4.

To allow DHCPv4 to bootstrap a MAP-T session, the Sky Hub signals to RADIUS (via the BNG), that it is MAP-T capable by including the string "MAPT1" within DHCPv4 Option 60 - Vendor Class ID

```

Dynamic Host Configuration Protocol (Discover)
  Message type: Boot Request (1)
  Hardware type: Ethernet (0x01)
  Hardware address length: 6
  Hops: 0
  Transaction ID: 0x525564b9
  Seconds elapsed: 0
  > Bootp flags: 0x0000 (Unicast)
  Client IP address: 0.0.0.0
  Your (client) IP address: 0.0.0.0
  Next server IP address: 0.0.0.0
  Relay agent IP address: 0.0.0.0
  Client MAC address: SkyUk_fe:8c:c3 (00:a3:88:fe:8c:c3)
  Client hardware address padding: 00000000000000000000
  Server host name not given
  Boot file name not given
  Magic cookie: DHCP
  > Option: (53) DHCP Message Type (Discover)
  > Option: (57) Maximum DHCP Message Size
  > Option: (55) Parameter Request List
  > Option: (60) Vendor class identifier
    Length: 42
    Vendor class identifier: "6.10.0.12|001|SR213|310622CA001048|MAPT1"
  > Option: (61) Client identifier
  > Option: (82) Agent Information Option
  > Option: (255) End
```

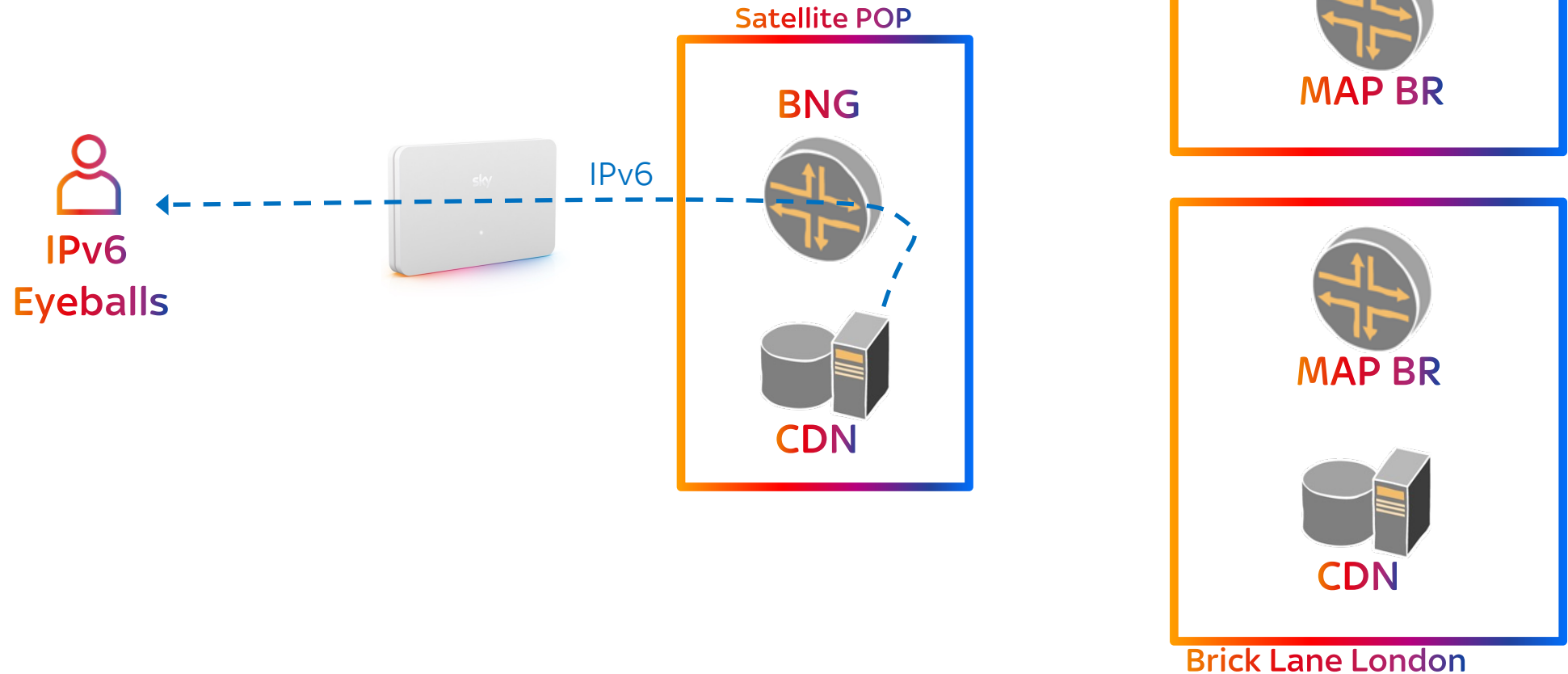
# Authentication Logic





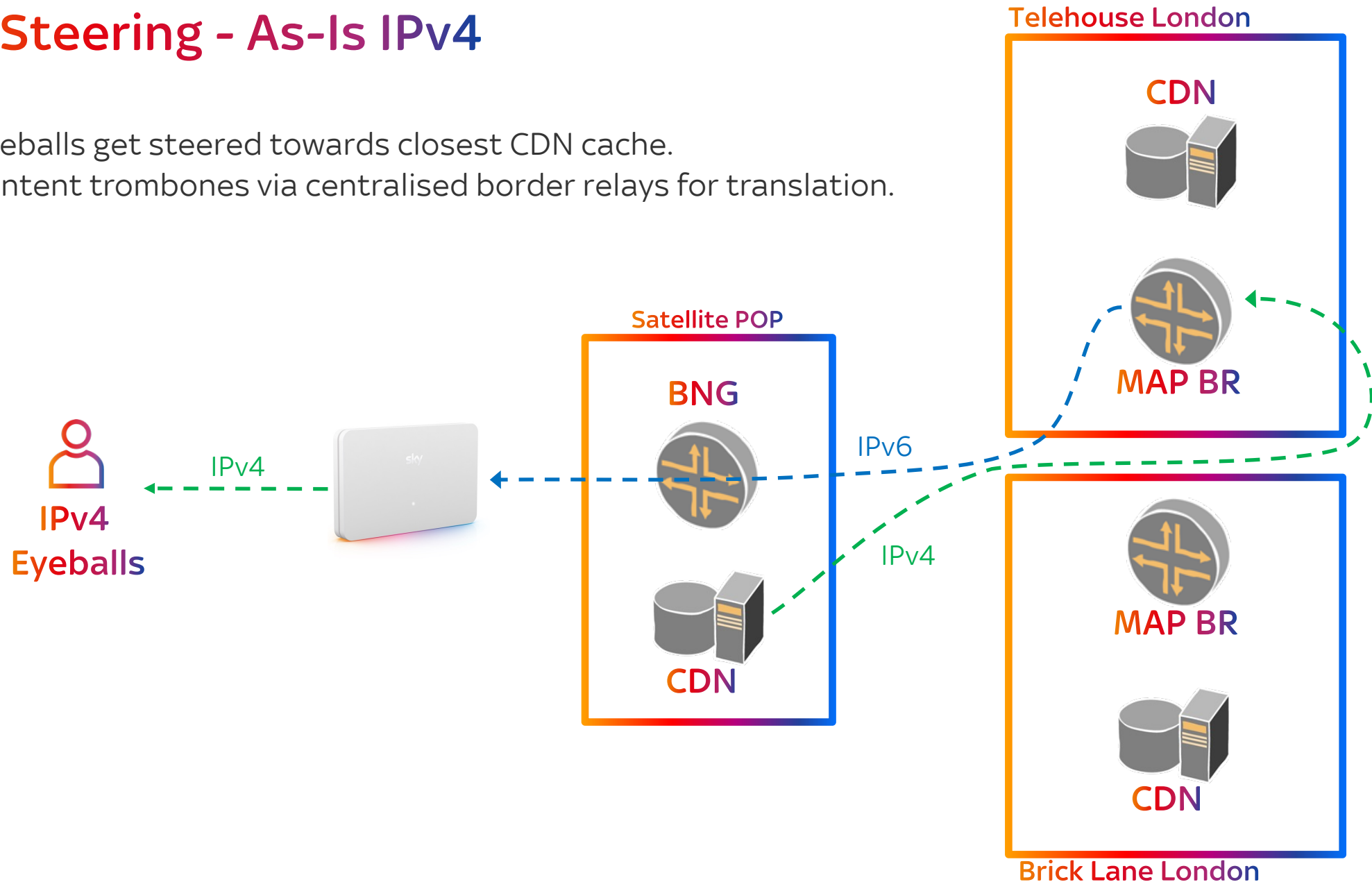
# CDN Steering - As-Is IPv6

IPv6 content routed directly from local CDN



# CDN Steering - As-Is IPv4

IPv4 eyeballs get steered towards closest CDN cache.  
IPv4 content trombones via centralised border relays for translation.



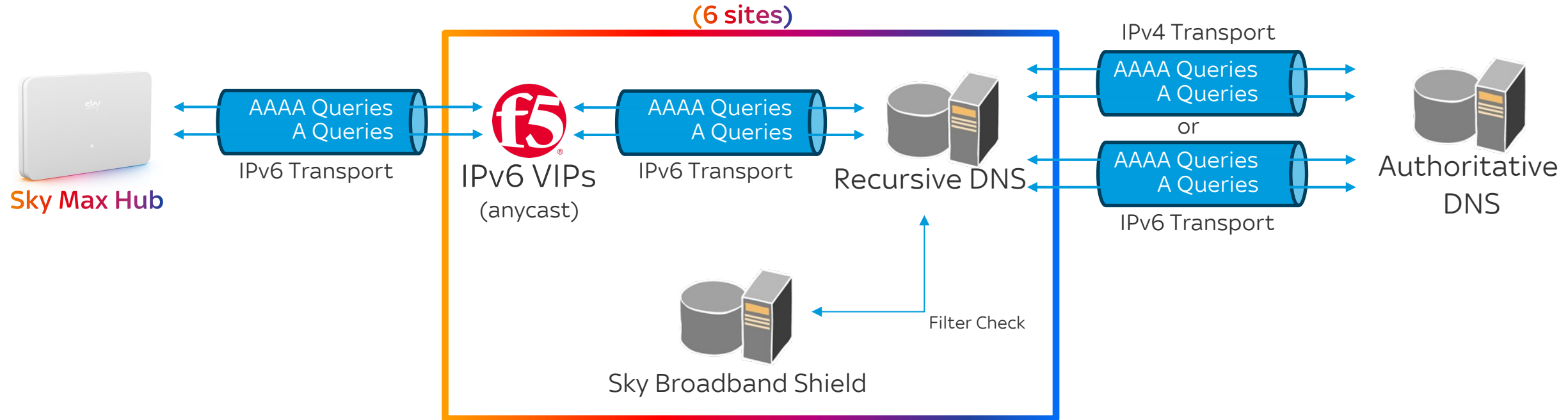
## CDN Steering – DNS-based

The source address of a DNS query is no longer appropriate for making steering decisions. (was it ever?)  
EDNS0 Client Subnet (ECS) by itself doesn't help.

CDNs need to be aware that IPv4 and IPv6 topologies are no longer the same and make different mapping & steering decisions accordingly.

**Authoritative NS needs to make decisions based on ECS value + QueryType (A vs AAAA),** but until such time.....

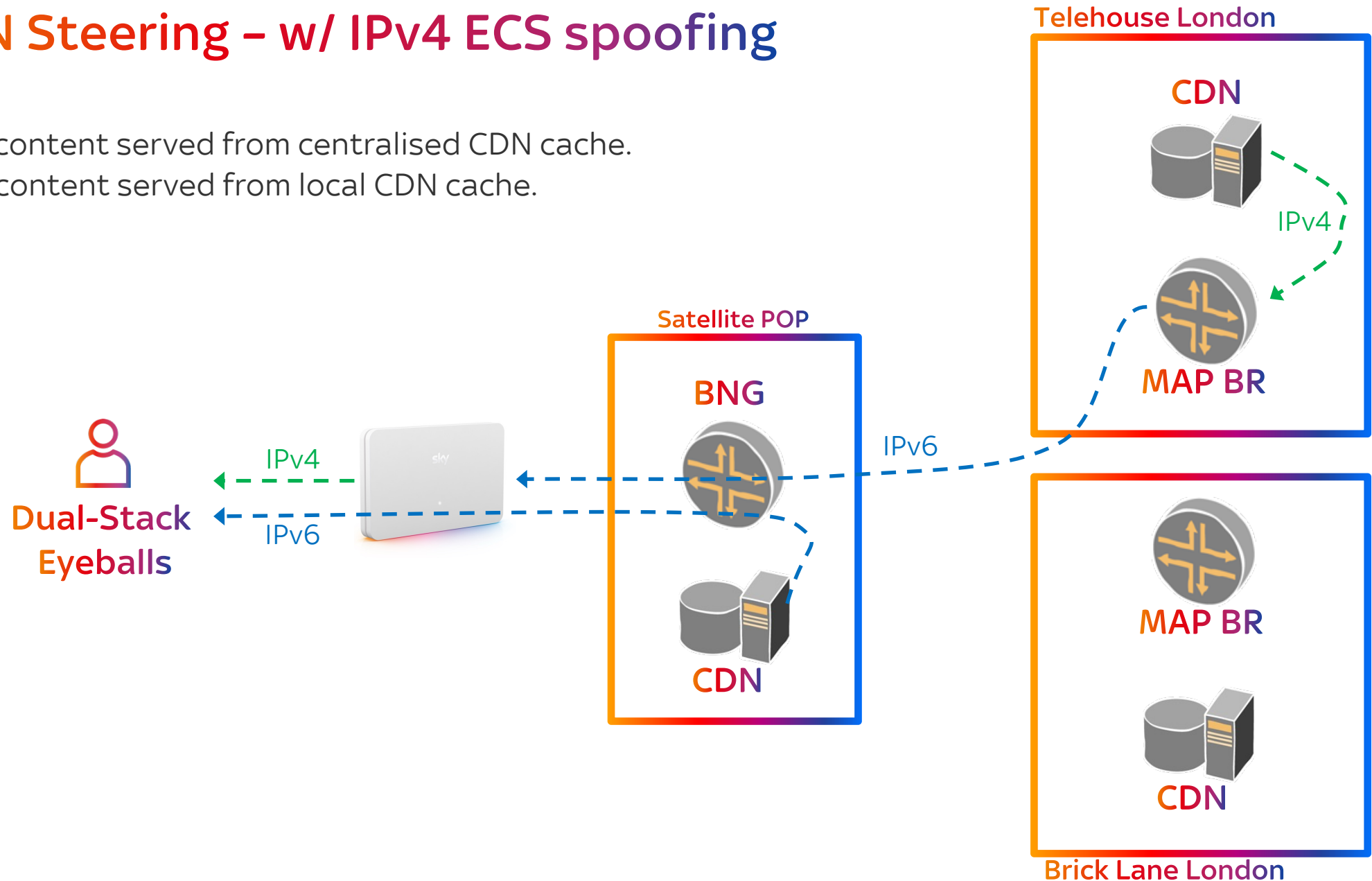
# Recursive DNS, CDN Steering & Sky Broadband Shield



- EDNS0 Client Subnet (ECS) inserted by Recursive DNS to aid CDN steering.
  - AAAA Queries for CDN domains, have IPv6 ECS inserted natively.
  - A Queries for CDN domains, have IPv4 ECS values synthesized and inserted.
- IPv6 source address of DNS query is used by Recursive DNS for filtering purposes.

# CDN Steering - w/ IPv4 ECS spoofing

IPv4 content served from centralised CDN cache.  
IPv6 content served from local CDN cache.



## Sky Glass & Stream – IPv6 Support



IPv6 support from software version QS031+



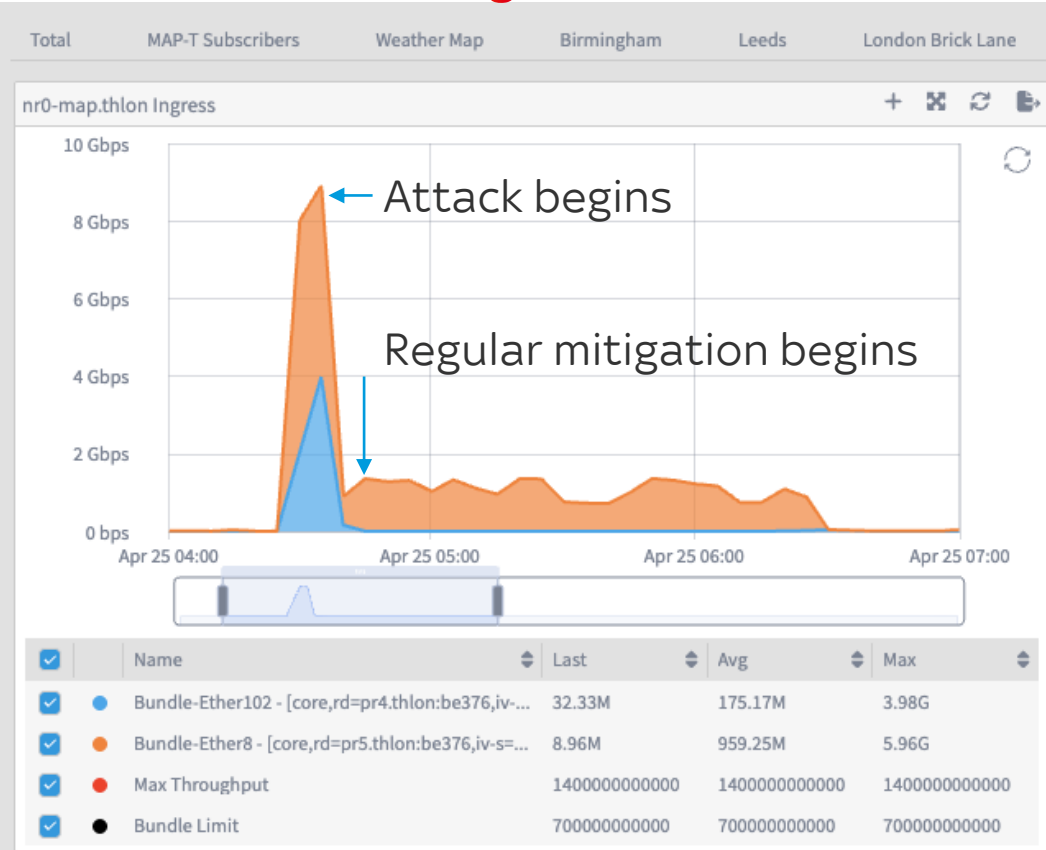
# Alternative Off-the-Shelf MAP-T CPEs

- RDK-B
- OpenWRT
  - Although “map” package not installed by default
  - Migration to nftables introduced a bug (#14449) that limits SNAT flows to one port range
- Keenetic
  - Keenetic OS 3.8+
- FRITZ!Box
  - FRITZ!Box 5590, 5530 Fiber
  - FRITZ!Box 7590 AX, 7530 AX, 7530, 7520, 7510
  - FRITZ!Box 6660, 6591 Cable
  - FRITZ!Box 4040
- TP-Link
  - Aginet Wi-Fi 6 & 7 CPEs from ISP Aginet portfolio.
  - EX820v & HX710 Pro
- ZyXel
  - (allegedly)

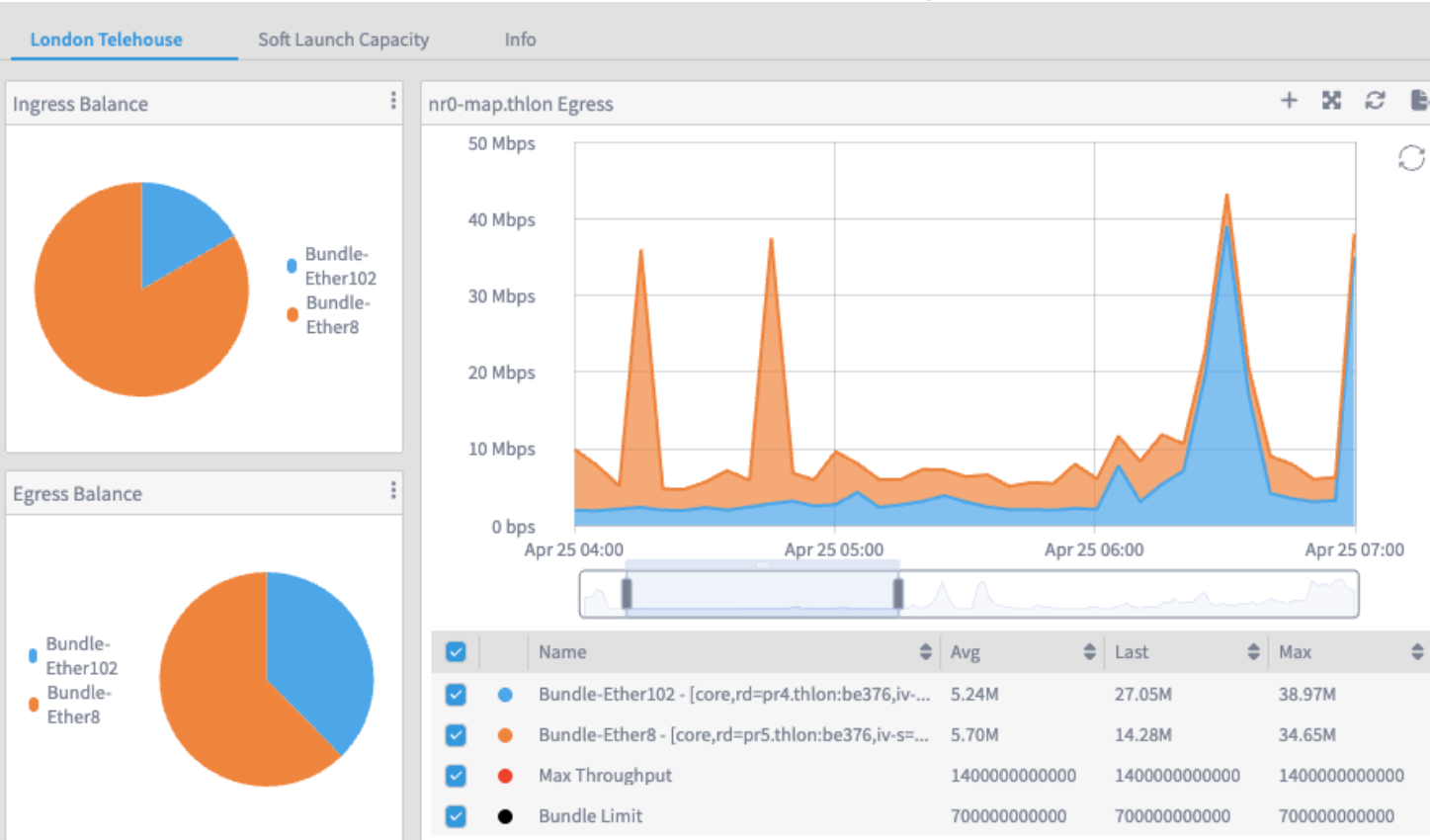
# Unexpected DDoS Protection

Malformed packets are not translated by the Border Relays

## Ingress



## Egress



Q&A

# DHCPv6 Handshake

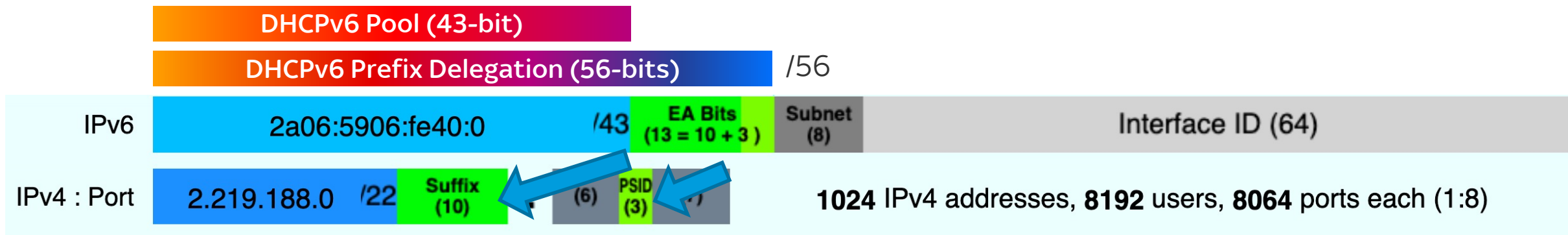


```
√ DHCPv6
  Message type: Solicit (1)
  Transaction ID: 0xceea84
  > Client Identifier
  > Identity Association for Prefix Delegation
  > Reconfigure Accept
  > Elapsed time
  > User Class
  > Vendor Class
  √ Option Request
    Option: Option Request (6)
    Length: 18
    Requested Option code: DNS recursive name server (23)
    Requested Option code: Domain Search List (24)
    Requested Option code: INF_MAX_RT (83)
    Requested Option code: Simple Network Time Protocol Server (31)
    Requested Option code: Vendor-specific Information (17)
    Requested Option code: S46 MAP-T Container (95)
    Requested Option code: SOL_MAX_RT (82)
    Requested Option code: User Class (15)
    Requested Option code: Vendor Class (16)
```

```
√ DHCPv6
  Message type: Advertise (2)
  Transaction ID: 0xceea84
  > Server Identifier
  > Client Identifier
  √ Identity Association for Prefix Delegation
    Option: Identity Association for Prefix Delegation (25)
    Length: 41
    IAID: 00000000
    T1: 1800
    T2: 2880
  √ IA Prefix
    Option: IA Prefix (26)
    Length: 25
    Preferred lifetime: 3600
    Valid lifetime: 3600
    Prefix length: 56
    Prefix address: 2a06:5904:fc40:200::
  √ S46 MAP-T Container
    Option: S46 MAP-T Container (95)
    Length: 39
  √ S46 Rule
    Option: S46 Rule (89)
    Length: 22
    > Flags: 0x00
    EA-bit length: 13
    IPv4 prefix length: 22
    IPv4 prefix: 2.123.240.0
    IPv6 prefix length: 43
    IPv6 prefix: 2a06:5904:fc40::
  √ S46 Port Parameters
    Option: S46 Port Parameters (93)
    Length: 4
    Offset: 6
    PSID length: 0
    PSID: 0
  √ S46 DMR
    Option: S46 DMR (91)
    Length: 9
    IPv6 prefix length: 64
    IPv6 prefix: 2a02:c79:701:ffe2::
  √ DNS recursive name server
    Option: DNS recursive name server (23)
    Length: 32
    1 DNS server address: 2001:4860:4860::8888
    2 DNS server address: 2001:4860:4860::8844
```

# MAP-T Basic Mapping Rule (BMR)

- The same BMRs are applied to both CPEs and Border Relays alike; no custom per-CPE configuration is required.
- BMRs are communicated to the CPE via DHCPv6 options within the lease (RFC 7598)
- The bits between the DHCPv6 pool supernet, and the Prefix Delegation leased to the CPE, informs the CPE which IPv4 address and layer 4 ports are available.



# MAP-T Default Mapping Rule (DMR)

## IPv4-Embedded IPv6 Address Format [RFC 6052]

8.8.8.8  
↑ ↑ ↑ ↑  
2001:db8:ffff:0:8:808:800:0000

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| +--+ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|