IPv6-Only with MAP-T

Sky Italia – AS210278

Semplice. Potente. Spettacolare.

Richard Patterson
RIPE NCC Open House - May 2021

Why MAP-T?

Pros:

- IPv4aaS.
  - IPv6-only access layer.
  - Reduce operational overhead.
- Allows IPv4 address sharing, or 1:1.
- Fewer bytes of overhead compared to encapsulation.
- Layer 4 header exposed for 5-tuple hashing.
- No DNS synthesizing required.
- Stateless.

Cons:

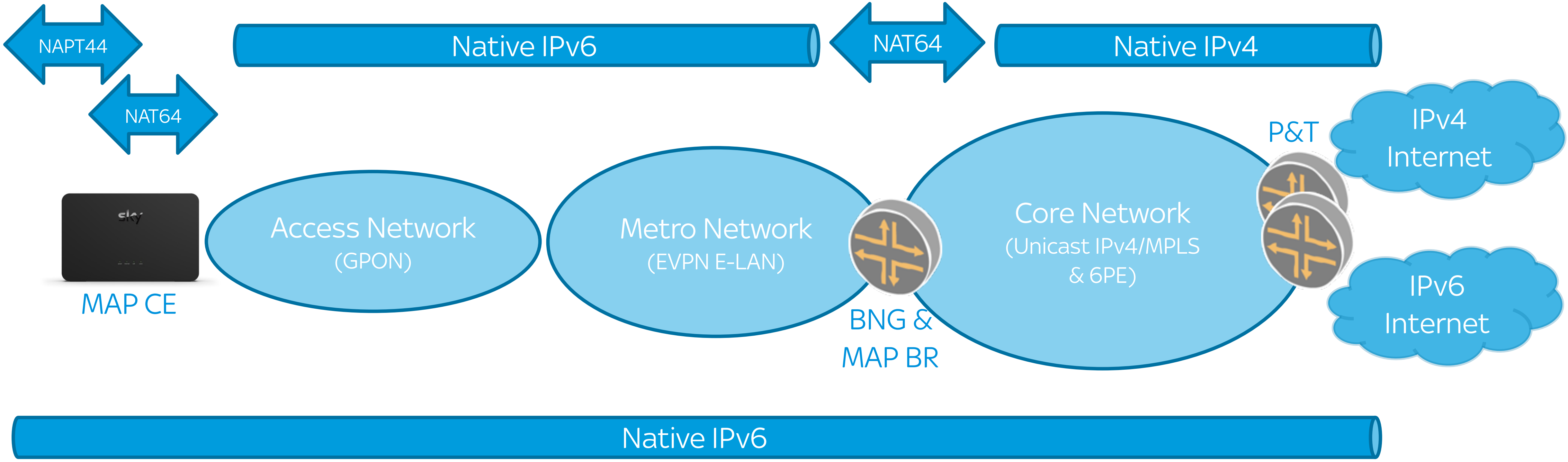
- No vendor could provide a real-world reference customer with a large deployment.
- Lack of CPE Support.
- IP pool management becomes more complicated.

MAP Border Relay

Vendor Selection

- Cisco ASR9K w/ Tomahawk
  - Near line-rate performance.
  - Probably want the Virtualized Services Module (VSM). [ICMP PTB, fragment handling, etc.]
- A10 Thunder
  - Good implementation & complete feature set.
  - High bps cost.
- Nokia vSR / ESA
  - Good implementation & complete feature set.
  - x86-only. No FP-based implementation.
- **Huawei CX600-X8 w/ LPUI-480**
  - Near line-rate performance and complete feature set.
  - No additional hardware required for ICMP generation, fragment handling.
  - Selected for BNG function, MAP BR feature came for the cost of a licence

MAP-T Network Topology



IPv4 Allocations

“But I thought this was an IPv6-only talk?”

- Greenfield network starting with 0 IPv4 addresses.
- Registered a new LIR, got a /22.
- We also bought a /13 from the open market.
  - Still not enough for our subscriber forecast, let alone addressing infrastructure as well.
- Dual Stack subscribers initially to de-risk the product launch.
- Borrowed additional IPv4 from Sky UK.
  - Borrowed IP space for use with Dual Stack only.
  - The new /13 remained untouched to simplify IP planning for the MAP domains.

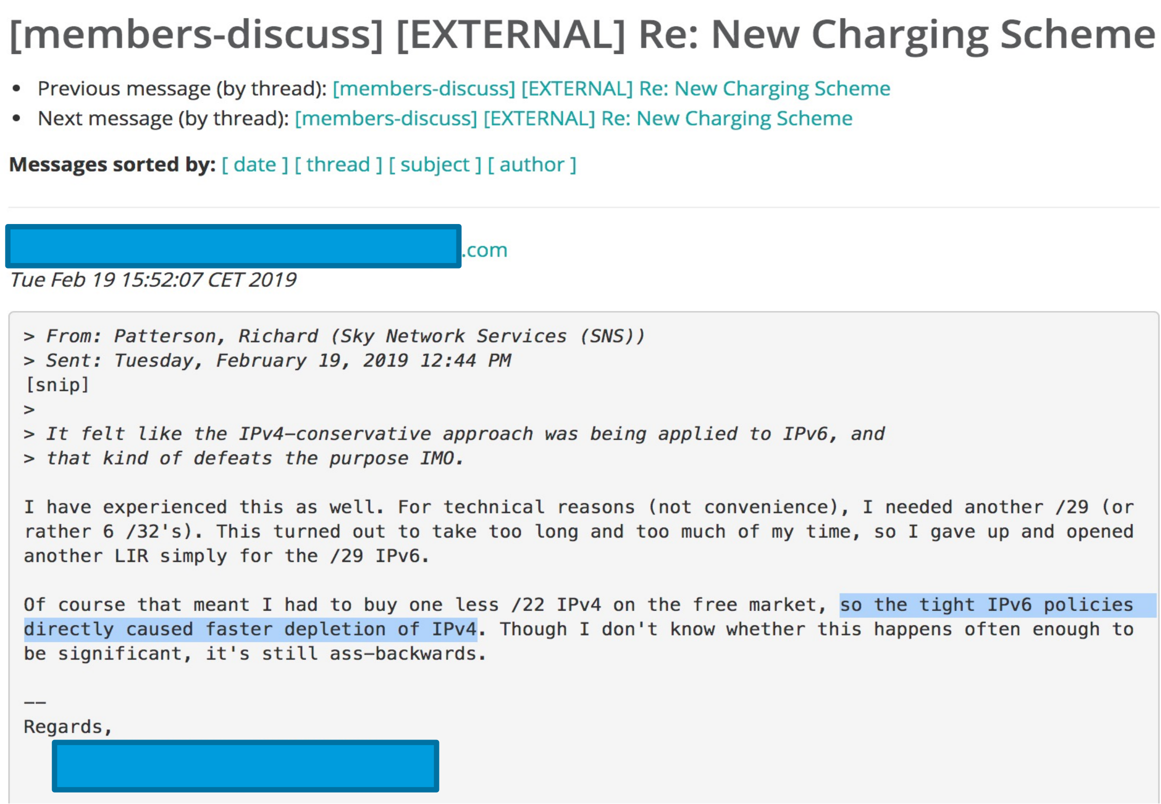
Dimensioning

IPv4 Usage / MAP-T Rules

101.56.0.0/13				
/14		/14		
/15	/15	/15	/15	
			/16	/16
Subscribers	**Subscribers**	**Subscribers**	**Infrastructure**	
Fixed ratio 16:1	Fixed ratio 1:1	Reserved	CDN	SPARE
32x blocks of ~65K Subs (/20)	32x blocks of 4K Subs (/20)		CSP	
~2.1M Subscribers Total	130K Subs Total		Loopbacks, etc.	

IPv6 Allocations

- RIPE NCC allocates an LIR up to a /29 without question.
- Enough for ~500K subscribers with /48-sized PDs.
  - As recommended in RIPE-690 BCOP
  - Not enough for our projected growth
- >/29 available with justification
  - Lots of back-n-forth emails.
  - IPv6 Transition technology constraints are excluded as justification.
  - We almost went with /56-sized PDs.
  - Some RIPE NCC members decide that spinning up a new LIR is the path of least resistance.



[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].com
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]
```

Dimensioning

IPv6 Usage / DHCPv6 Pools

2a0e:400::/25					
15 x /29		/29			
13x /29	2x /29	/31	/31	/31	/31
Subscribers		**Infrastructure**			
External		**Private**	**Internal**	**External**	**Spare**
/48 per subscriber = ~8M		Loopbacks	Point-to-Points	Public Servers	
~ 6.8 Million	~1M	Management	Intranet	CDNs	
104 blocks of /32	256 blocks of /36	Backend Servers	Middleware Servers	Enterprise / Corporate	Future Use
IPv4 Sharing Ratio 16:1	1:1				

IP Pool Management

Previously (in the UK):

- We over-provisioned DHCPv6 pools without fear of running out.
- DHCPv4 pools were tightly managed by automation to allow for efficient usage.

With MAP:

- IPv4-usage is now directly tied to DHCPv6 pools.
  - DHCPv6 Prefix Delegation + MAP Basic Mapping Rule = IPv4 Address + Layer 4 Ports.
- Over-provisioning DHCPv6 means wasting, or at least inefficient IPv4-usage.
- We still haven't automated it like we have automated our UK DHCPv4 pools.

IPv4 Address Sharing

- ~95% of subscribers on a MAP profile with a 16:1 sharing ratio.
- ~5% of subscribers on an “Opt-Out” MAP profile with sharing ratio 1:1 to allow:
  - Port forwarding
  - DMZ
  - Non-port-based layer 4 protocols
    - GRE
    - ESP
- Proactively detect opt-out triggers using WebPA.
  - DMZ enablement.
  - Port forwarding / firewall rules.
  - ~~UPnP AddPortMapping requests for ports used by known-problematic applications.~~
  - Direct cost impact. Could be abused, needs to be monitored.
  - Proactively opt-back-in when no longer required.

Regulatory Compliance

- Stateless translation = No per-flow logging.
  - Some jurisdictions expect 5-tuple logging when sharing IPv4 addresses.
  - Some Border Relays can still support per-flow logging. (A10)
- AGCOM, the local regulator, specifies a maximum IPv4 address sharing ratio.
  - 16:1 for Fixed-line.
  - 32:1 for Mobile.
- Lawful Intercept & Additional Mandatory Obligations
  - Location of LI & AMO functions in relation to the the MAP Border Relay function.
- Our Broadband Network Gateway (BNG) is also our MAP Border Relay
  - Custom solution to enrich RADIUS Accounting session data with MAP rules.
  - RFC8658: RADIUS Attributes for Softwires, support to come.

Customer Premise Equipment

MAP CE

- In-house developed Sky Hub 4.
  - Based on RDK-B with a Broadcom SoC.
- Initial trials run using CERNET's ivi implementation. (Incl. in Broadcom's SDK)
  - Integrated stateful NAPT44
  - Couldn't use existing iptables rules for NAT or IPv4 firewalling.
    - Hooks in to Netfilter on PREROUTING before conntrack/mangle/nat.
  - Port forwarding broken. (Broadcom patched ivictl with rudimentary support)
  - Non-port-based layer 4 protocols broken.
- Migrated to Andrew Yourtchenko's NAT46 kernel module. [1]
  - Used by OpenWRT.
  - Broadcom patched with support for their hardware acceleration.

[1] https://github.com/ayourtch/nat46

Customer Premise Equipment

Cont’d.

Netfilter w/ port-restricted SNAT

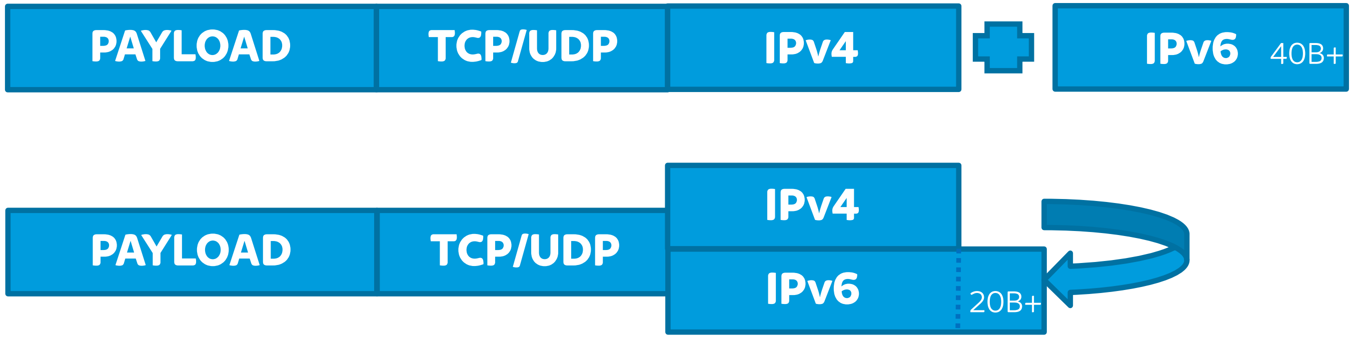
- Source ports can be re-used when the destination IP and port are different.
  - However Netfilter’s SNAT target isn’t built with multiple non-contiguous sport ranges in mind.
    - Support removed in 2.6.11-rc1
- Netfilter’s Connlimit match used to fall-through multiple SNAT rules with different port ranges.
  - Broadcom patched with daddr & dport matching for more efficient sport usage.
- Even high-speed fixed-line broadband usage can make do with very few external source ports.
  - Regulations mean we didn’t push this beyond (65536-1024)/16 = 4,032 ports per subscriber.
  - Left as an exercise for the reader to quantify. Sorry. ☺

MTU & Fragmentation

Encapsulation:

Vs

Translation:



- Build your access & transport networks to handle the additional overhead to avoid unnecessary fragmentation.
- Varying frame-size support from different wholesale access providers.
  – Who may in turn aggregate multiple access-types from 3rd parties, also with varying frame-size support.
- Don't forget the IPv6 Fragmentation Header (+8 bytes)
  – Used to signal if IPv4 DF=0, even when there aren't IPv6 fragments.

IPv6 First

- Recursive DNS over IPv6-only.
- Voice over IPv6-only.
- CPE management must be IPv6 capable.
  - WebPA *(akin to TR.069 / ACS)*
  - NTP
  - Firmware Upgrade Server
- Plume Pod WiFi extenders updated to support IPv6.
  - Internal GRE tunnel over IPv6.
    - Using ULA endpoints for stability even when WAN is down.
  - Cloud management over IPv6 using Opensync 2.0 [1]
    - Our first use-case for a 2nd /64 on the LAN.

[1] https://www.opensync.io/about

CDN, Steering & Analytics

- IPv6 where possible to avoid translation.
  - Border Relay being co-located with BNG makes this somewhat moot for us.
- EDNS0 Client Subnet
  - IPv6-only recursive DNS + DNS proxy on CPE = ECS all IPv6.
  - Simplifies ECS summarisation and topology mapping.
- IPv4 topology may be different to IPv6 topology
  - Location of Border Relays.
  - Anycasting Border Relay prefixes.
  - MAP domain design decisions. Single large domain or many smaller ones.
- Application owners & 3rd parties may want a feed of MAP rules to understand the IPv4 address sharing behaviour.

Dual Stack

- Wasted effort resolving dual stack-related bugs and complexities.
  - RADIUS Accounting & dealing with multiple independent sessions.
  - Wholesale access provider hit a vendor bug with DHCPv4.[1]
    - DHCPv6 was unaffected.
- Consumes IPv4 space which you will need for planning MAP domains.
- Customers used to dual stack may get a surprise when forced to use MAP-T.
  - Majority won’t notice as they use Sky-provided CPE.
  - Small number with 3rd party CPEs that don’t support MAP-T.
  - An even smaller number (**0.085%**) are 3rd party CPEs connected with single stack IPv4-only.

[1] (CSCvt83520)

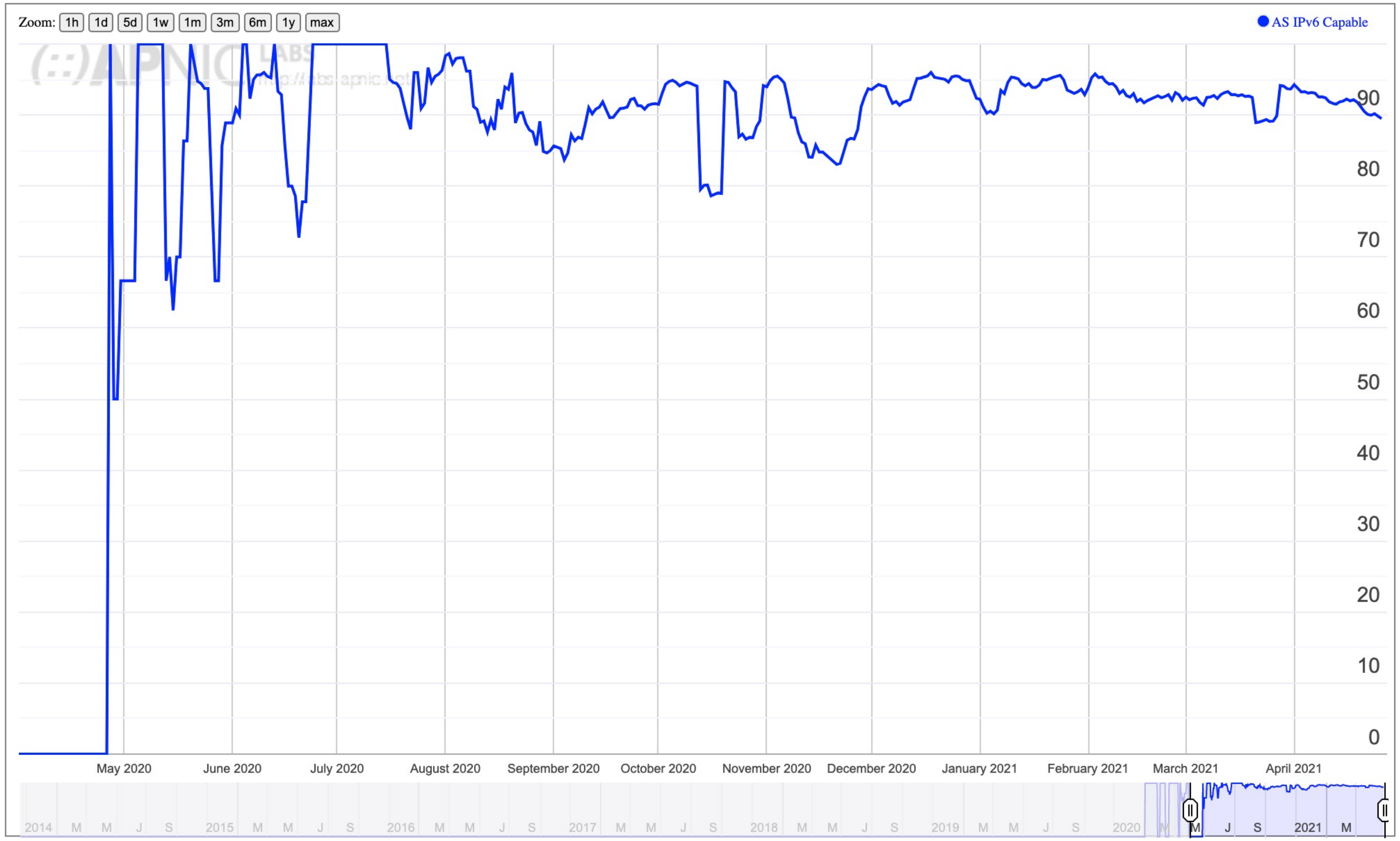
Where Are We At?

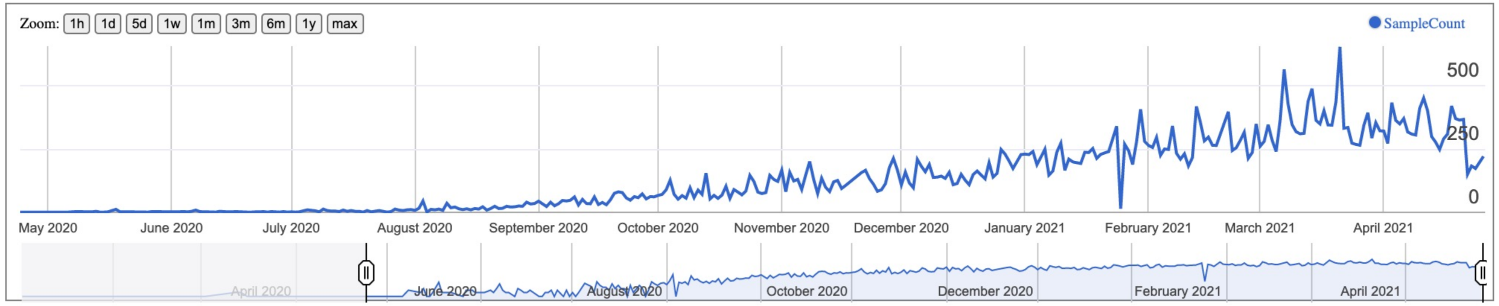
- Currently still in a staff trial phase, with >500 subscribers.
- Testing new Sky Hub firmware with nat46 integration.
- DHCPv6 Server S46 PortParams Option bug.
  - Sky Hubs unaffected, but OpenWRT is.
- MTU Problems
  - Unexpected IPv6 Fragmentation Header being added when IPv4 DF=0.
- Rollout targeted for July.
- MAP-T default on for all new subscribers by August.

AS201278

APNIC Labs Stats

IPv6 Per-Country Deployment for AS210278: SKYIT-BB, Italy (IT)


Zoom: 1h 1d 5d 1w 1m 3m 6m 1y max
AS IPv6 Capable
May 2020
June 2020
July 2020
August 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021
March 2021
April 2021
2014 M M J S 2015 M M J S 2016 M M J S 2017 M M J S 2018 M M J S 2019 M M J S 2020 M M J S 2021 M


Zoom: 1h 1d 5d 1w 1m 3m 6m 1y max
SampleCount
May 2020
June 2020
July 2020
August 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021
March 2021
April 2021
April 2020
June 2020
August 2020
October 2020
December 2020
February 2021
April 2021