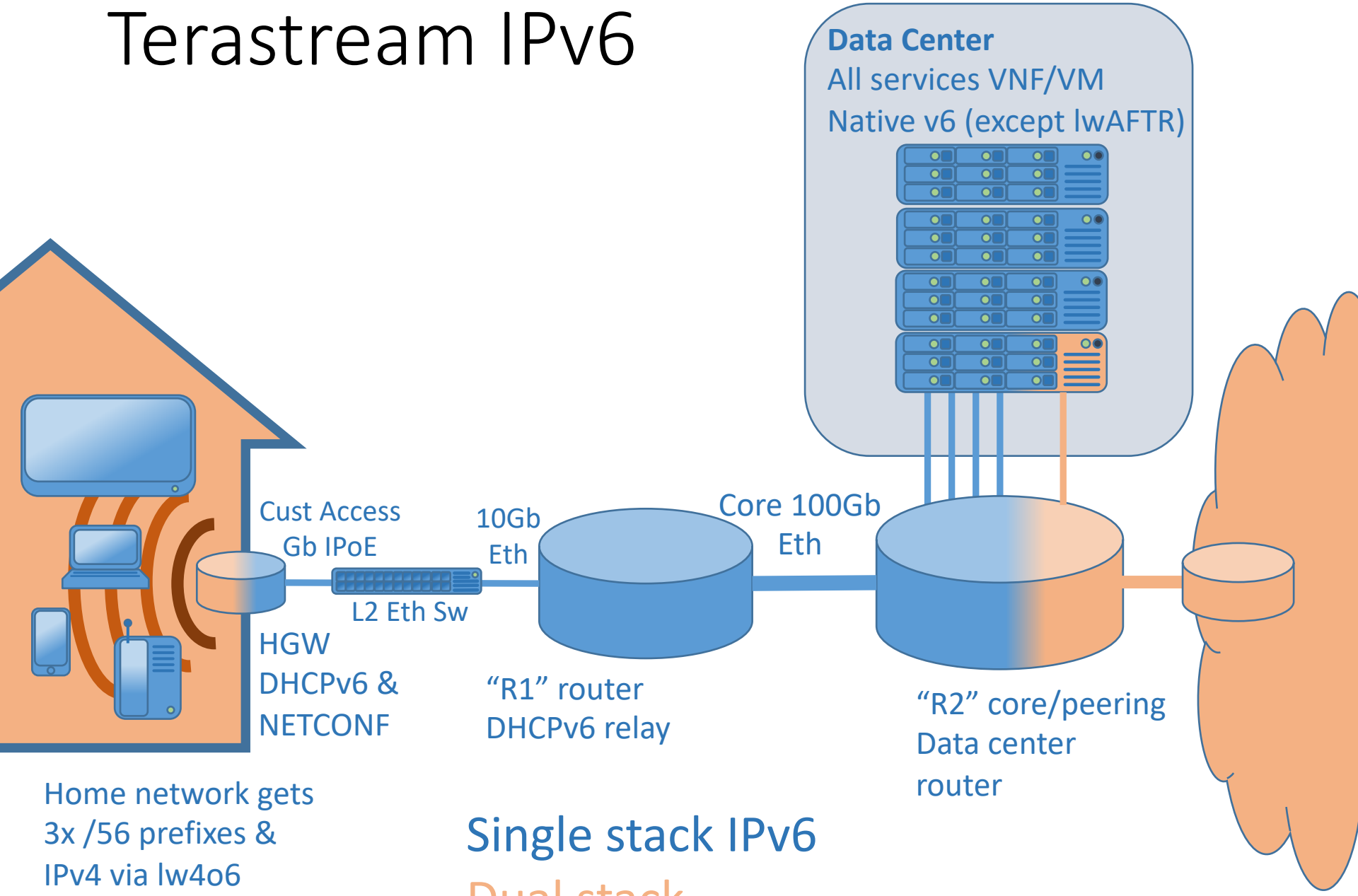
Deutsche Telekom "Terastream"

M. Abrahamsson, I. Farrer

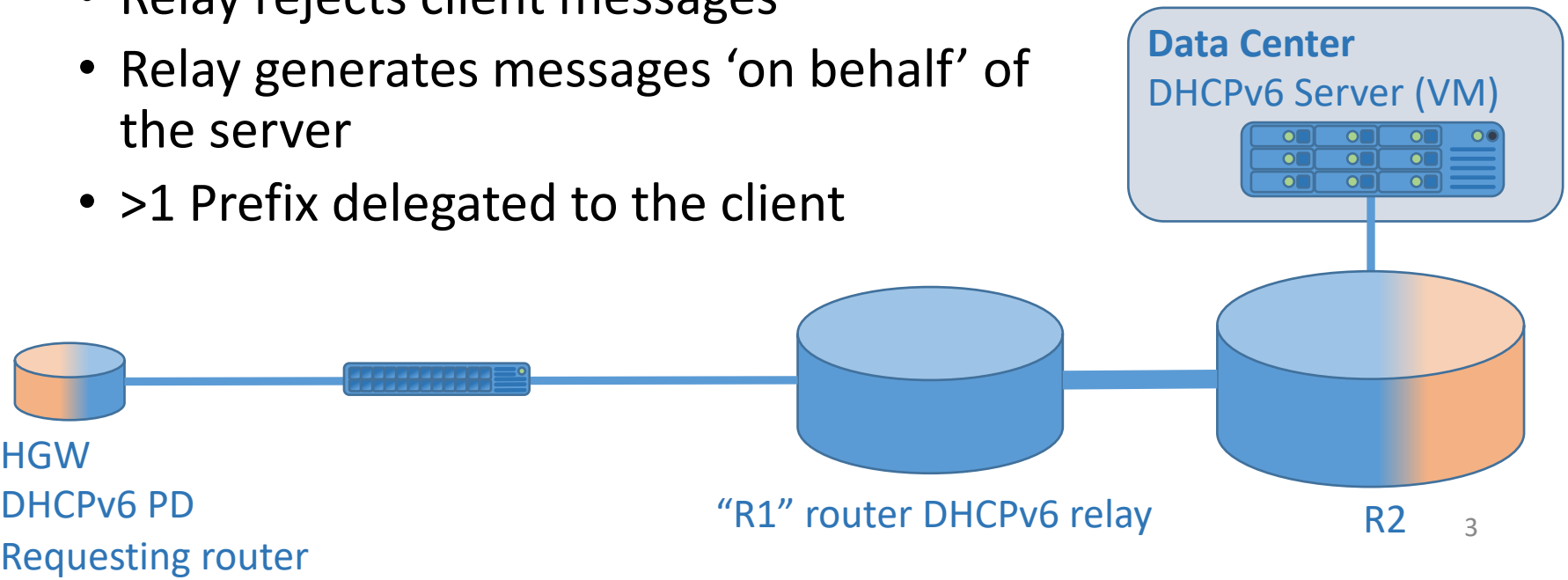
IETF104 Prague

Terastream IPv6



DHCPv6 Relay with PD Implementations 1

- At one time or another we've had problems with every major network vendor's DHCPv6 relay implementation when used with PD
- General problems –
  - Client/relay/server out of sync
  - Relay rejects client messages
  - Relay generates messages 'on behalf' of the server
  - >1 Prefix delegated to the client



DHCPv6 Relay with PD Implementations 2

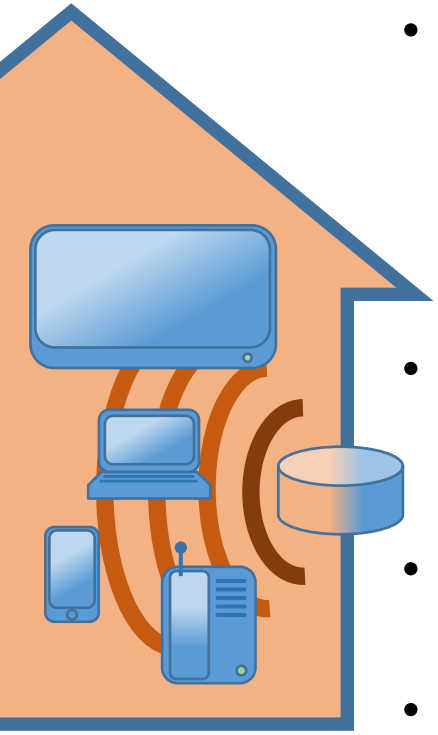
- RFC8415 is sketchy on how this is meant to work (section 19.1.3):

A relay agent forwards messages containing prefix delegation options in the same way as it would relay addresses (i.e., per Sections 19.1.1 and 19.1.2).

If a server communicates with a client through a relay agent about delegated prefixes, the server may need a protocol or other out-of-band communication to configure routing information for delegated prefixes on any router through which the client may forward traffic.

- This is true, but incomplete – the relay needs to implement a state machine synchronized with the server and client
- This undefined behavior has resulted in vendor implementation problems

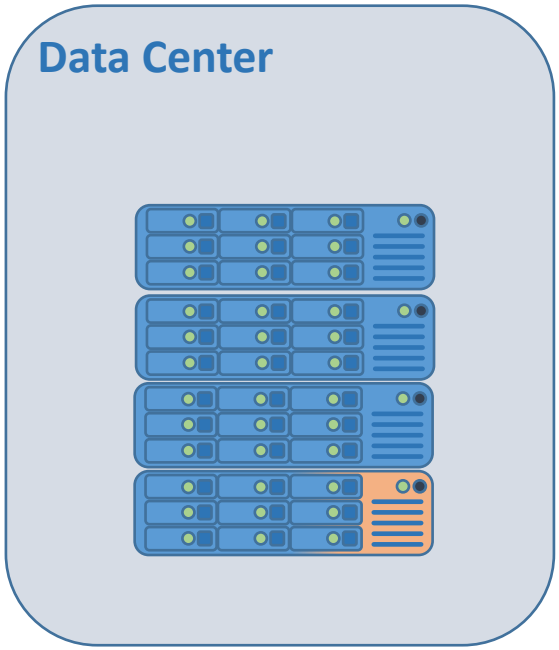
Multiple, service specific IPv6 Prefixes to the Host



Home network gets 3x /56 prefixes

- The Terastream architecture provides multiple prefixes to the client, currently:
  - Video
  - Voice
  - Best Efforts (BE)
- These are used to identify traffic throughout the network to identify traffic (for QoS, ACL etc.) without needing DPI for setting TOS etc.
- This requires user hosts to select correct source address for the traffic type. 1 device may use >1 prefix
- Provisioning Domain (MIF) IETF WG was chartered to solve this but was not successful
- draft-ietf-intarea-provisioning-domains looks like it will be very helpful here (INTAREA)

Multi Attached Data Center Host



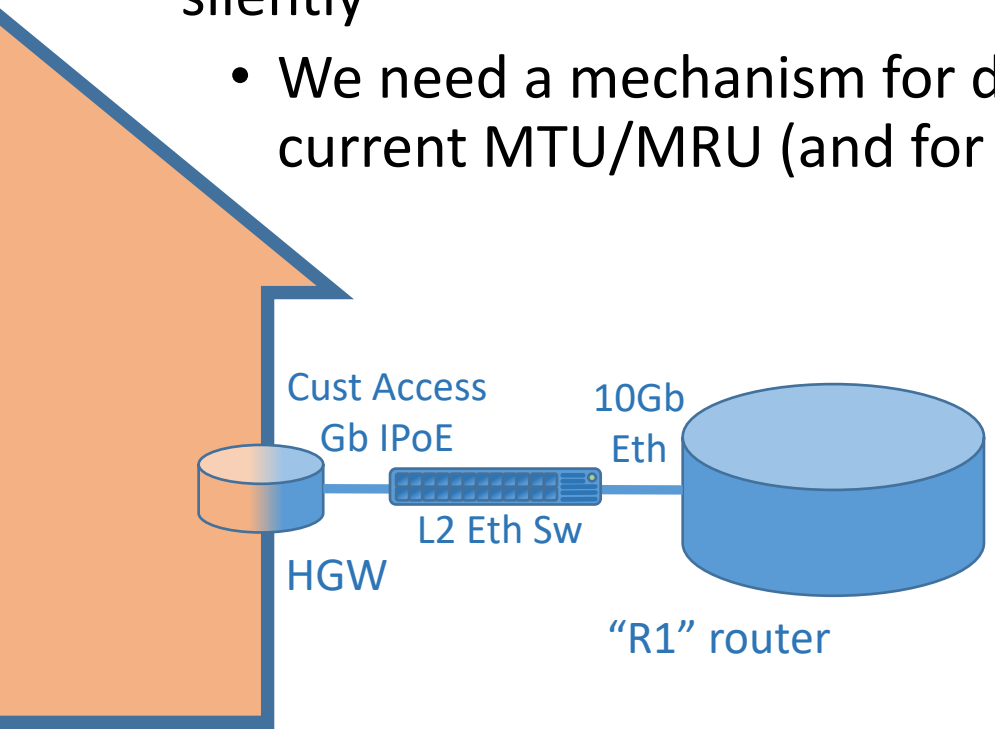
- As part of the multi-prefix addressing model, data center VMs/VNFs have multiple interfaces connecting to different domains (data plane, signaling/m2m, management etc.)
- The interaction between:
  - Destination address selection
  - Source address selection
  - Route selection
  - Source based routing
  - Strong/weak host model

Are unpredictable, vary from OS to OS and version to version

- To solve this, we have needed to supply hosts with a lot of fragile, static configuration

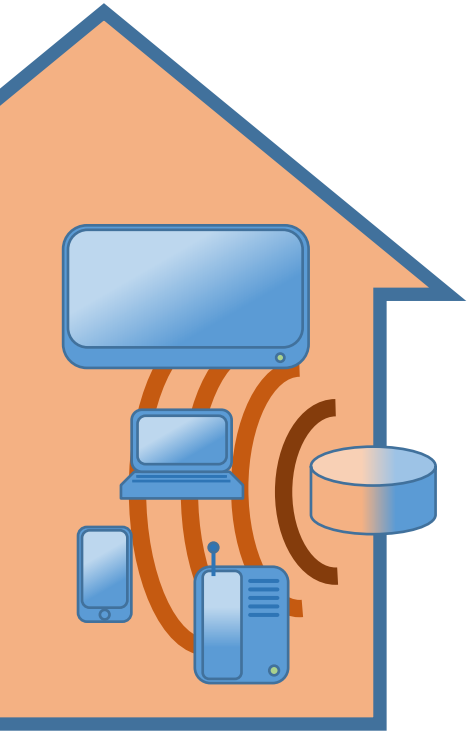
MTU problems HGW WAN

- We configure 9000 on R1, and it sends RA MTU=9000
- Common HGW WAN interface MTUs: 1500, ~2300, 9000 (HW limit)
- Some devices will use 1500, some will configure 9000 but not have MRU of more than 2300 meaning > 2300 will be dropped, silently
  - We need a mechanism for devices to announce their current MTU/MRU (and for their claims to be verified)



draft-van-beijnum-multi-mtu could be applicable. This problem space is shared by Internet exchanges.

MTU problems HGW LAN



- For the LAN we currently use only 1500 bytes (IPv4 and IPv6)
- We'd like to support MTU 9000
- Wi-Fi chips commonly only support ~2300
- Most operating systems come with a default MTU 1500 (in some cases this is the largest supported)
- We need to support legacy and enhanced hosts on same LAN (mixed MTU).

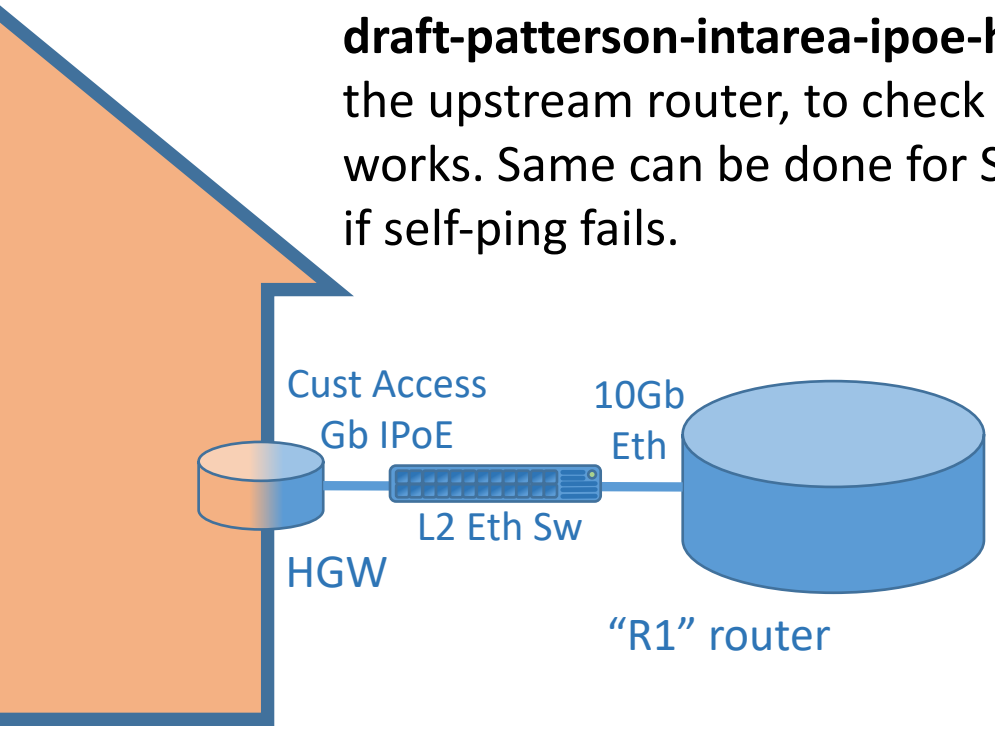
Again, **draft-van-beijnum-multi-mtu** would work.

This needs to be incrementally deployable.

WAN uplink working?

- With IPoE there is no built-in mechanism to check if the L3 connectivity is working (problem for both IPv4 and IPv6). If the L2 switch-R1 link goes down then the HGW cannot detect it.
- If it has a secondary uplink, it can't figure out that it needs to use it.
- With ND/RA and DHCPv4/DHCPv6 as ships in the night, there is no standardized way to handle certain events.

draft-patterson-intarea-ipoe-health suggests pinging yourself via the upstream router, to check that DHCPv6-PD forwarding plane works. Same can be done for SLAAC based addresses. Perform action if self-ping fails.



Another way could be to use ND/NUD and trigger some action if the upstream router becomes unreachable (ND fails).

Summary…

- There's been quite a lot of much bigger problems with IPv6 (and its implementations) that we've found and resolved in the last 6 years
- The issues described in this presentation are still outstanding points
- BUT – they are mostly relatively minor gremlins rather than barriers to deployment