# DHCPv4 Option for IPv4 routes with IPv6 nexthops

`draft-equinox-intarea-dhcpv4-route4via6-02`

Tobias Fiebig (Max-Planck Institut für Informatik),
David ‘equinox’ Lamparter (NetDEF / FRRouting)

IETF 124 · Montreal, CA · 2025-11-04

# required context

- `RFC8950`
  Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop (this is for BGP specifically)
- `draft-ietf-intarea-v4-via-v6-03`
  IPv4 routes with an IPv6 next hop
- TL;DR: “just use gw IPv6 addr + ND instead of gw IPv4 addr + ARP to find the gateway’s MAC address”

# works great on/between routers

control plane support (BGP):

- JunOS
- Arista
- Cisco
- ExaBGP (no FIB)
- FRR
- Bird
- probably more by now

actually sending packets:

- JunOS
- Arista
- Cisco
- Linux (netlink)
- FreeBSD

Euro-IX WG is looking at using this at internet exchange points, cf.
https://github.com/euro-ix/rfc8950-ixp

# what about the end hosts?

current options:

- full-on subnet (/30 or larger) – wasteful & tricky isolation
  (yes, subnets with multiple hosts are less wasteful depending on utilization)
- per-host /31 ⇒ 50% “guaranteed wastage”
- host /32 + shared /32 on router
  special case: https://datatracker.ietf.org/doc/draft-ipv6-resolved-gateway/

# what about the end hosts?

current options:

- full-on subnet (/30 or larger) – wasteful & tricky isolation
  (yes, subnets with multiple hosts are less wasteful depending on utilization)
- per-host /31 ⇒ 50% “guaranteed wastage”
- host /32 + shared /32 on router
  special case: https://datatracker.ietf.org/doc/draft-ipv6-resolved-gateway/

⇒ host /32 + IPv6 gateway?

# IPv6 gw is a config/deployment problem

- Linux & FreeBSD kernels happily accept the route
  - covers a lot of VMs & containers in clouds
  - if you have the access, you(r orchestration) can “just set it”
  - if you don’t: you’re probably using DHCP and have no way to signal this
- also need to make the router→host direction work

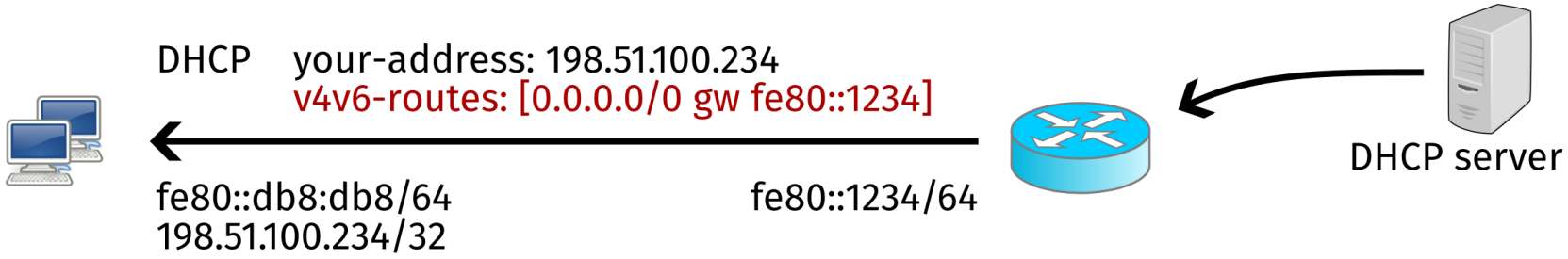
# ⇒ goals

- use DHCP to assign individual IPv4 /32 addrs as before
- get IPv4 routes (or at least default gateway) with IPv6 nexthops into end hosts using DHCP ⇒ no IPv4 addr on router
- minimize ARP usage

- DHCP client considered alterable for now

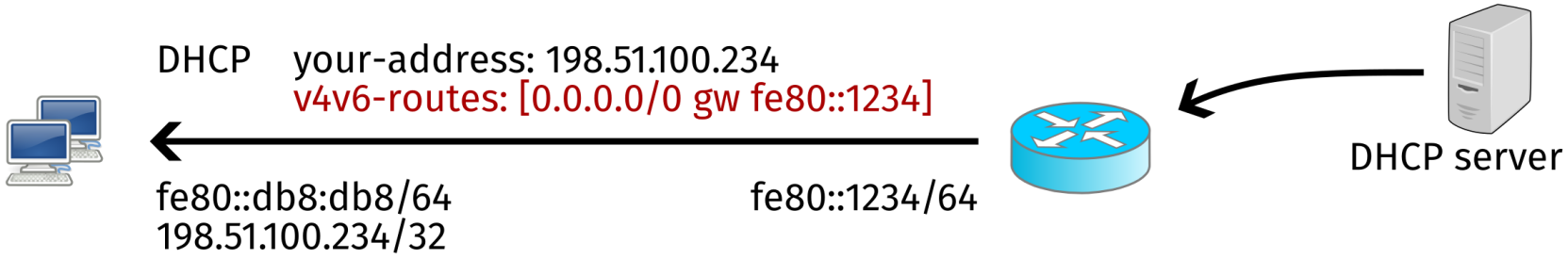
# 1st half: new DHCP(v4) option for v6 routes

- plain new DHCP (**IPv4**) option to carry IPv4 routes with IPv6 nexthops
- similar to option 121 (CIDR static routes)
  - list of (v4 prefix + v6 nexthop)



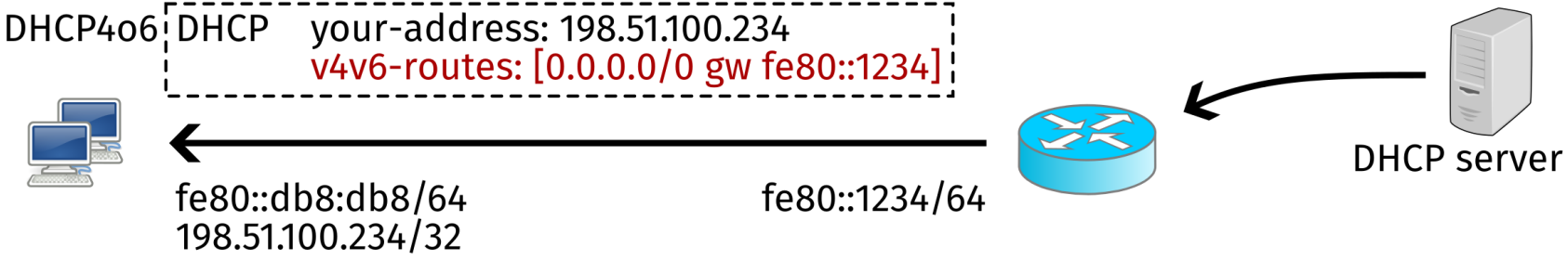
# 1st half: — not quite sufficient — fragile

- what's the DHCP (IPv4) relay address?
- router doesn't know host's IPv6 address
  - can't use v4-over-v6 route in the R→H direction
  - grab MAC from DHCP? use ARP? with what source IPv4?



# 2nd half: use DHCP 4o6

- RFC7341: “put DHCP(v4) in DHCP(v6) option”
- router now has an IPv6 address for the host!
- DHCPv6 relaying is much cleaner



# caveats

- router must DHCP-snoop assigned IPv4 & inject routes
- EVPN ARP-snooping won't work if there's no ARP
- fill DHCP IPv4 values for backwards compatibility where needed (e.g. use dummy address from Remco's draft)
  - TL;DR: router ARP-responds for 192.0.0.x address
    ⇒ use that as default gateway
  - works with unmodified DHCP clients, but less flexible & clean

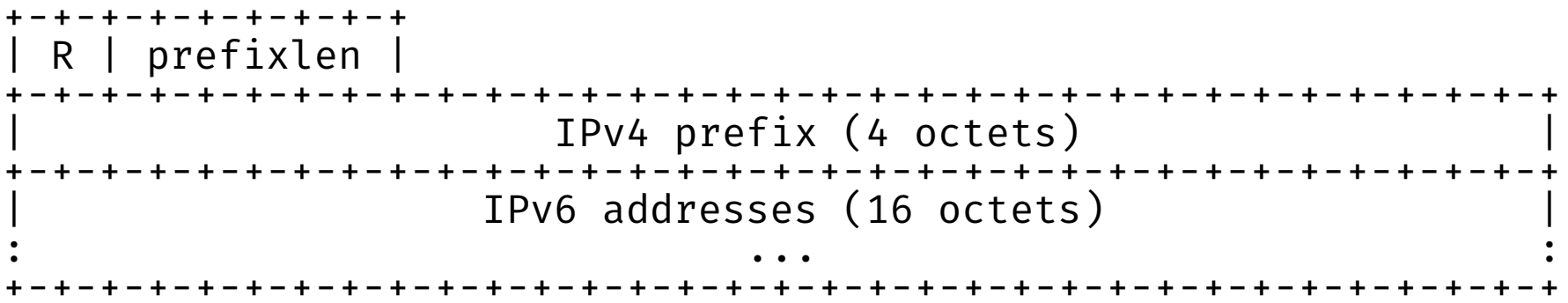
# questions?

just to repeat: DHCPv4 option for IPv4 route with IPv6 nexthop

our question: is this worth doing?

- especially if we can do Remco's reserved 192.0.0.x address instead?
- actually had working code for an earlier revision
- we don't really care about encoding details for now

# current option encoding (backup slide)




- default route is 0.0.0.0/0
- some IPv6 addresses have special meanings
  - ::0/128 ⇒ use DHCP 4o6 source IPv6 address
  - 100::/64 ⇒ unreachable network (RFC6666 DISCARD IPv6 address block)

# preexisting shared-IP approach (backup slide)

- DHCP offer contains:

```
your-IP-address:    198.51.100.123
subnet-mask:        255.255.255.255
router:             203.0.113.234
cidr-route:    /32 203.0.113.234 0.0.0.0
```

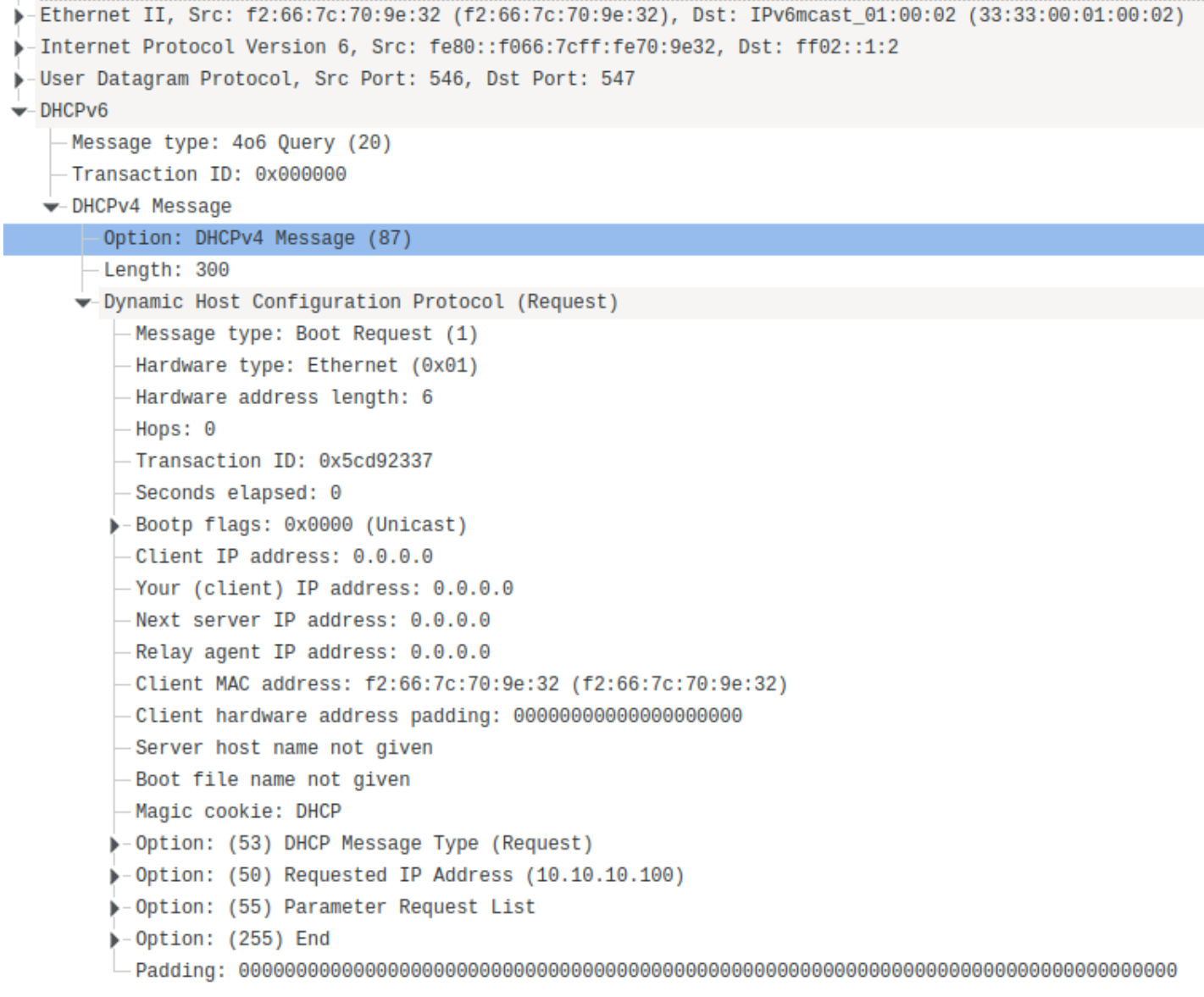
- client assumes 203.0.113.234 is on-link
  - some clients need the cidr-route for this, some don’t
- this was (to our knowledge) never specified anywhere
  - but works quite widely

# DHCP4o6

(backup slide)

originated in DS-Lite provisioning to CPE devices, but agnostic of actual use case

could be terminated at router (wouldn't recommend that) or passed all the way to DHCP server



# ICMP / error reporting (backup slide)

- routers theoretically don't need any IPv4 addresses
- this breaks PMTUD & traceroute

  ⇒ use 192.0.0.8 dummy address
  + draft-intarea-extended-icmp-nodeid

- looks a little bit like NAT64 scenarios if you tilt your head sideways, but it's ICMPv4 here, not ICMPv6
  - the PMTUD/traceroute triggering packet is pure IPv4

# IPv4 route w/ IPv6 nexthop on Linux (backup slide)

```
# ip -4 route add 198.51.100.0/24 dev veth0 via inet6 fe80::1234
# ip -4 route list
198.51.100.0/24 via inet6 fe80::1234 dev veth0
```

If your Linux kernel is newer than ca. 2020-ish, it supports this.