Routujeme IPv4 provoz prostřednictvím IPv6 sousedů

Radek Zajíc, radek@zajic.v.pytli.cz • Den IPv6, 6. 6. 2025


IPV6

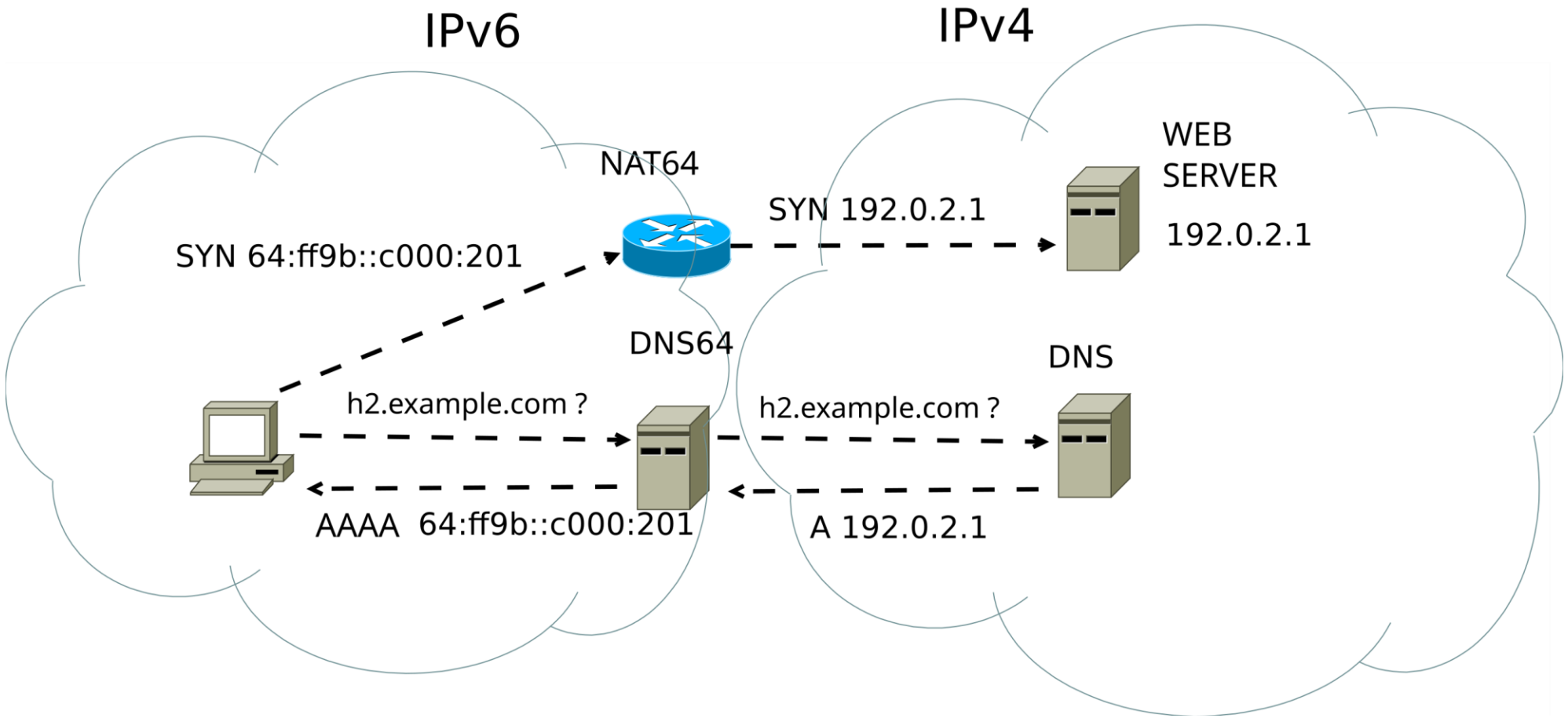

Máte
IPv6?

Dual-Stack je jen dočasným řešením

Dual IP Layer Operation

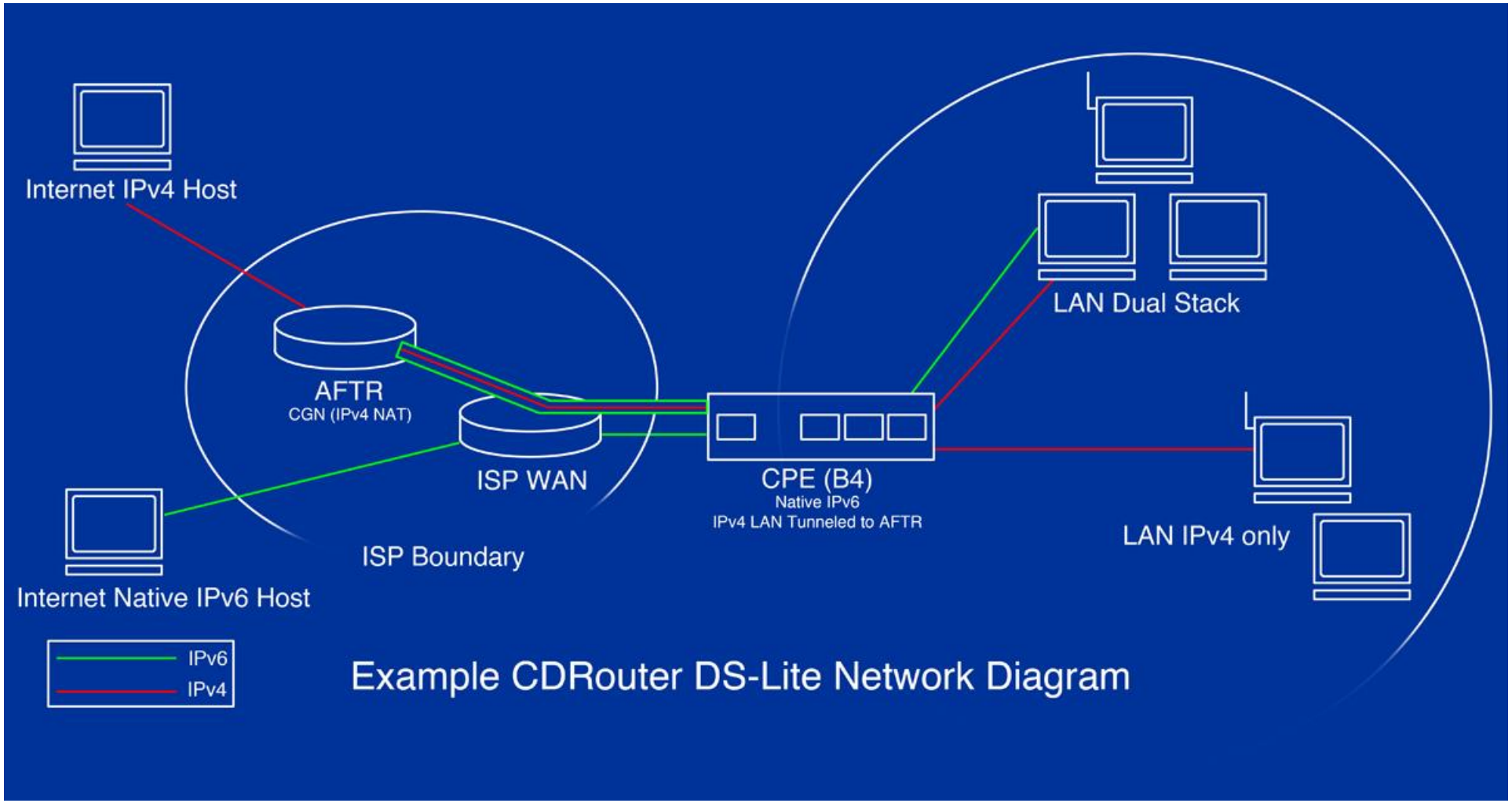
The most straightforward way for IPv6 nodes to remain compatible with IPv4-only nodes is by providing a complete IPv4 implementation. IPv6 nodes that provide a complete IPv4 and IPv6 implementations are called "IPv6/IPv4 nodes." IPv6/IPv4 nodes have the ability to send and receive both IPv4 and IPv6 packets. They can directly interoperate with IPv4 nodes using IPv4 packets, and also directly interoperate with IPv6 nodes using IPv6 packets. (RFC 2893, **Transition Mechanisms for IPv6 Hosts and Routers**, 08/2000)

Přechodové mechanismy – NAT64, DNS64 (2011)



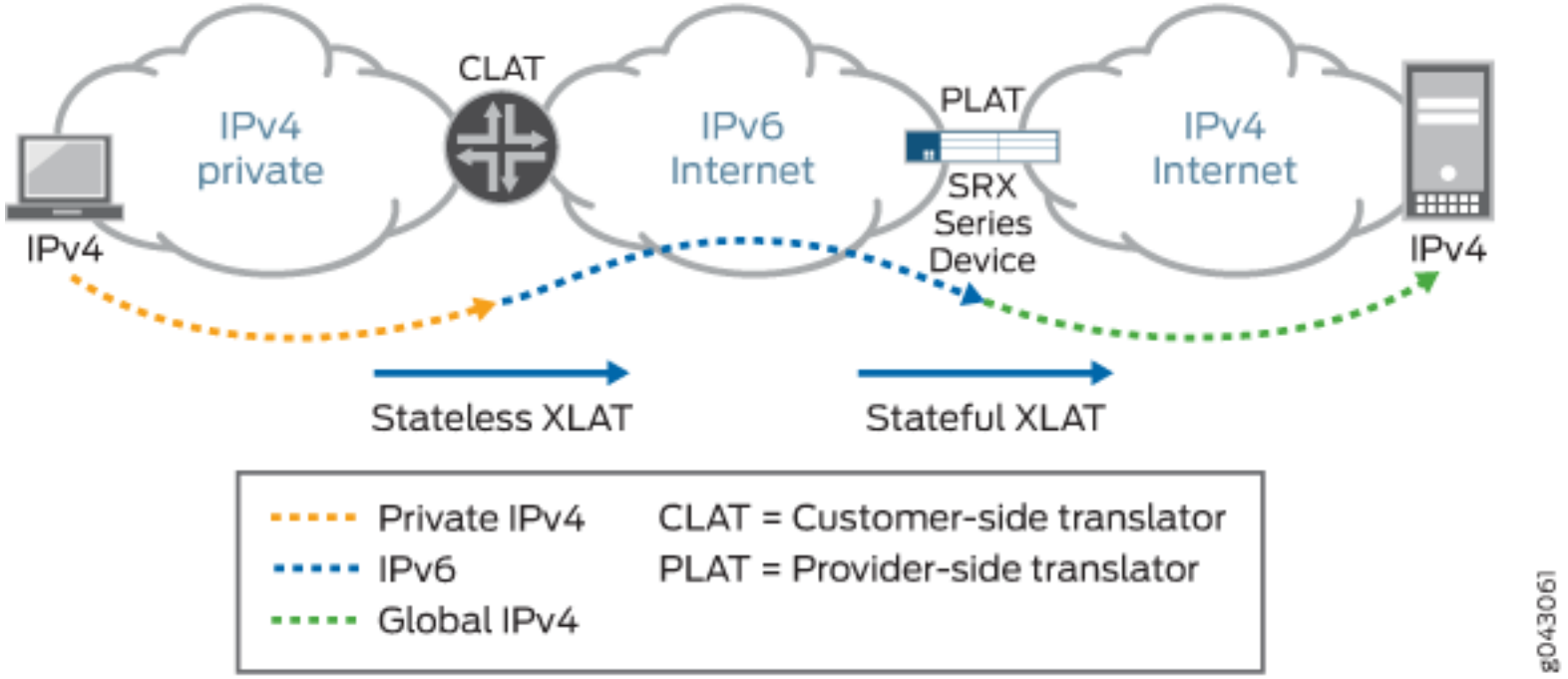
Wikipedia: Mro, CC BY-SA 3.0; https://en.wikipedia.org/wiki/NAT64

Přechodové mechanismy – DS-Lite (2011)



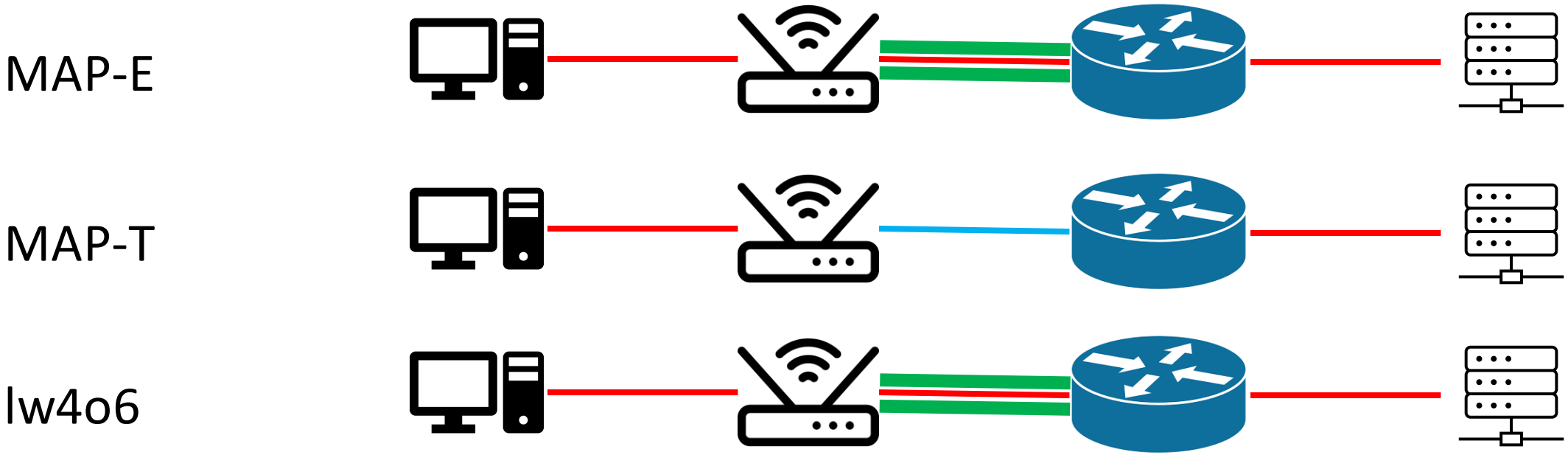
https://support.qacafe.com/cdrouter/knowledge-base/ds-lite-testing-with-cdrouter/

Přechodové mechanismy – 464XLAT (2013)



https://www.juniper.net/documentation/us/en/software/junos/alg/topics/concept/security-xlat-alg-traffic-support-understanding.html

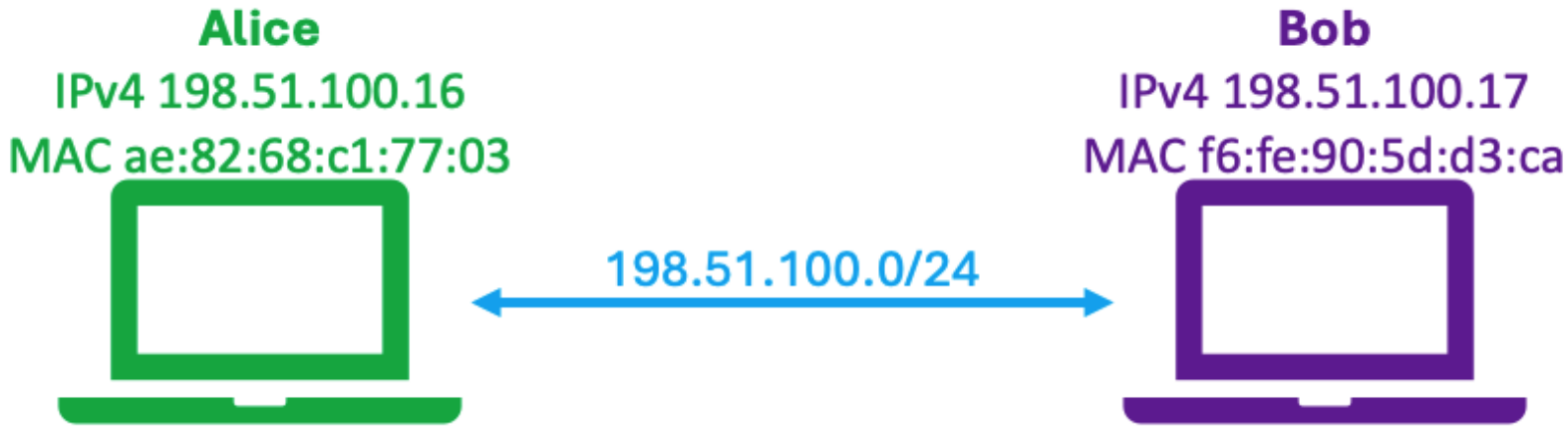
Přechodové mechanismy – MAP-T, MAP-E, lw4o6 (2015)



Přechodové mechanismy – IPv6-mostly

Přechodové mechanismy – IPv6 next hops

IPv4 Routing 101

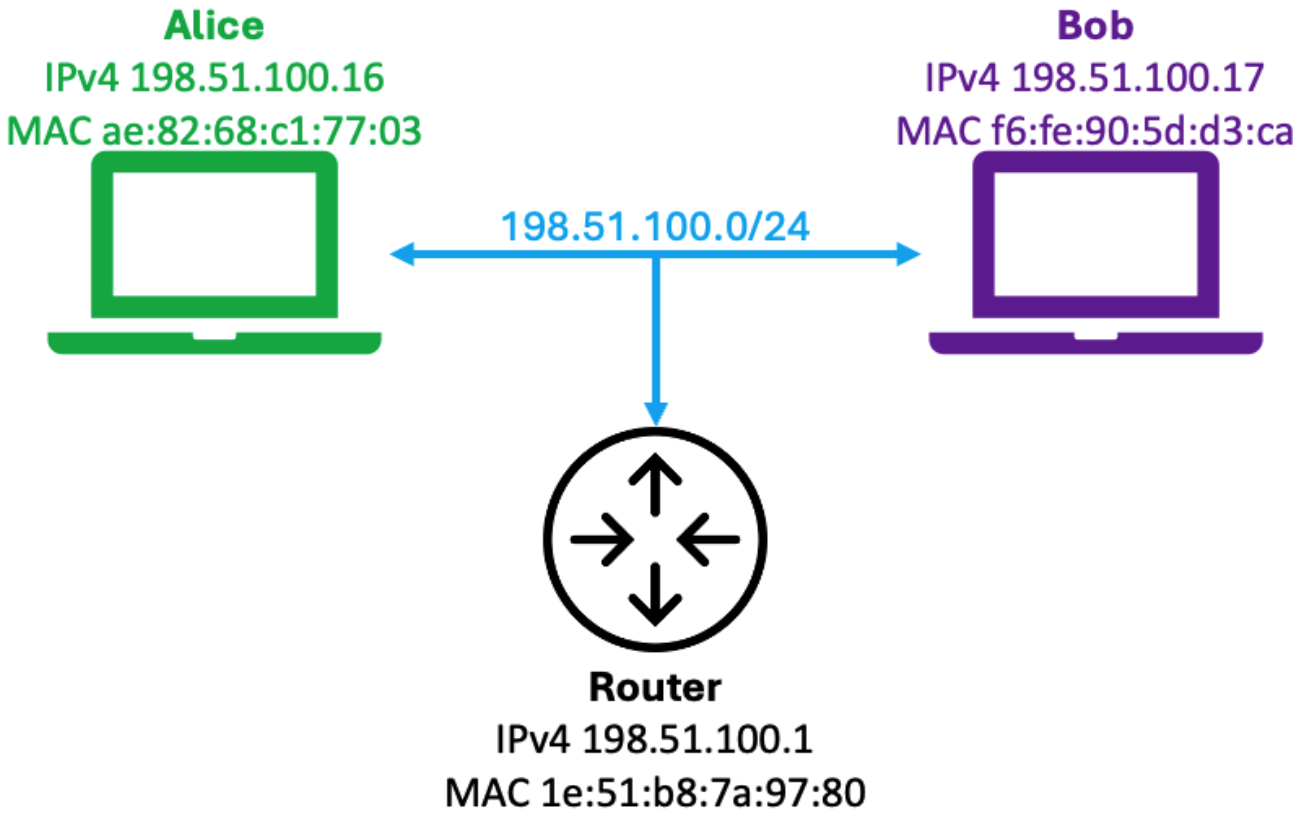


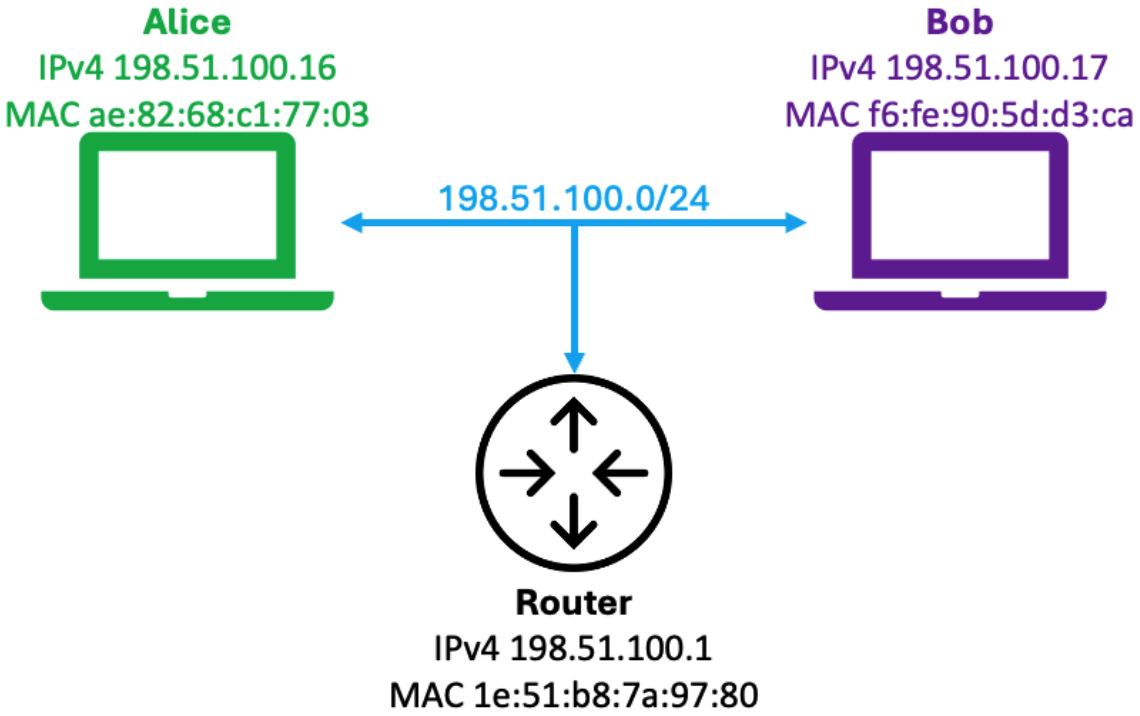

```
Alice: IP 198.51.100.16, MAC ae:82:68:c1:77:03
Bob:   IP 198.51.100.17, MAC f6:fe:90:5d:d3:ca

bob@bob:~# ip route
198.51.100.0/24 dev from_bob proto kernel scope link src 198.51.100.17
```

Kde je sakra Alice?

```
f6:fe:90:5d:d3:ca > ff:ff:ff:ff:ff:ff, ARP: Request who-has 198.51.100.16
tell 198.51.100.17
ae:82:68:c1:77:03 > f6:fe:90:5d:d3:ca, ARP: Reply 198.51.100.16 is-at
ae:82:68:c1:77:03
```

```
bob@bob:~# ip route
198.51.100.0/24 dev from_bob proto kernel scope link src 198.51.100.17

bob@bob:~# ping 8.8.8.8
ping: connect: Network is unreachable
```



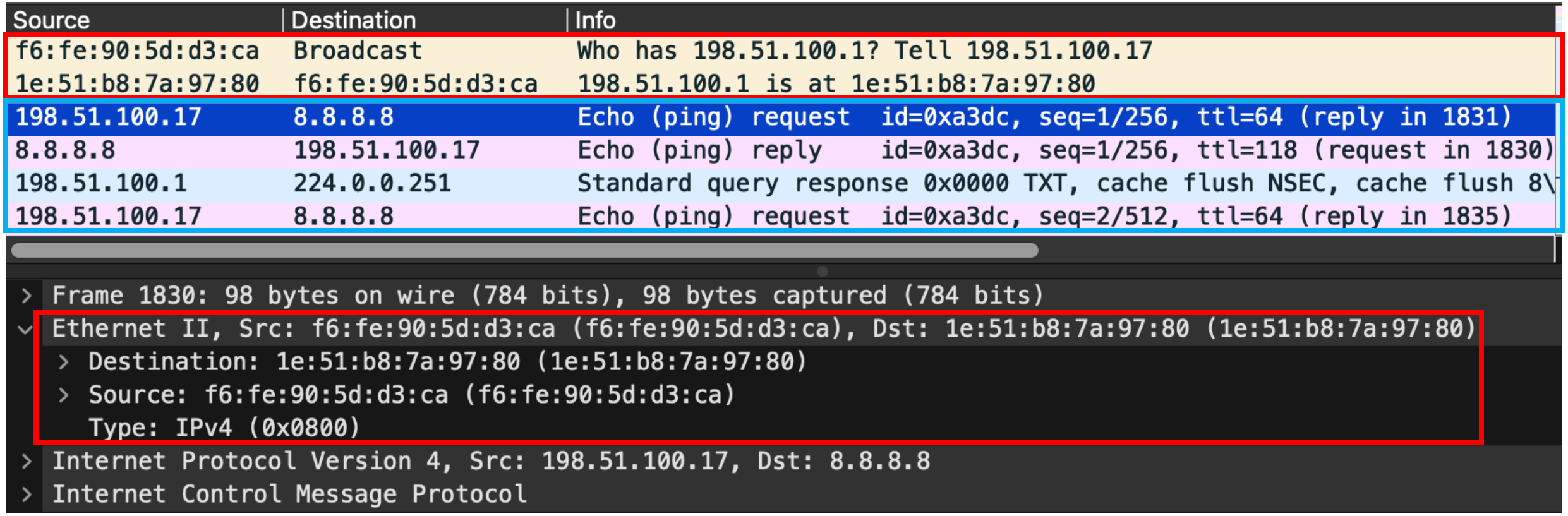



```
Router IP 198.51.100.1,  MAC 1e:51:b8:7a:97:80
Alice  IP 198.51.100.16, MAC ae:82:68:c1:77:03
Bob    IP 198.51.100.17, MAC f6:fe:90:5d:d3:ca
```

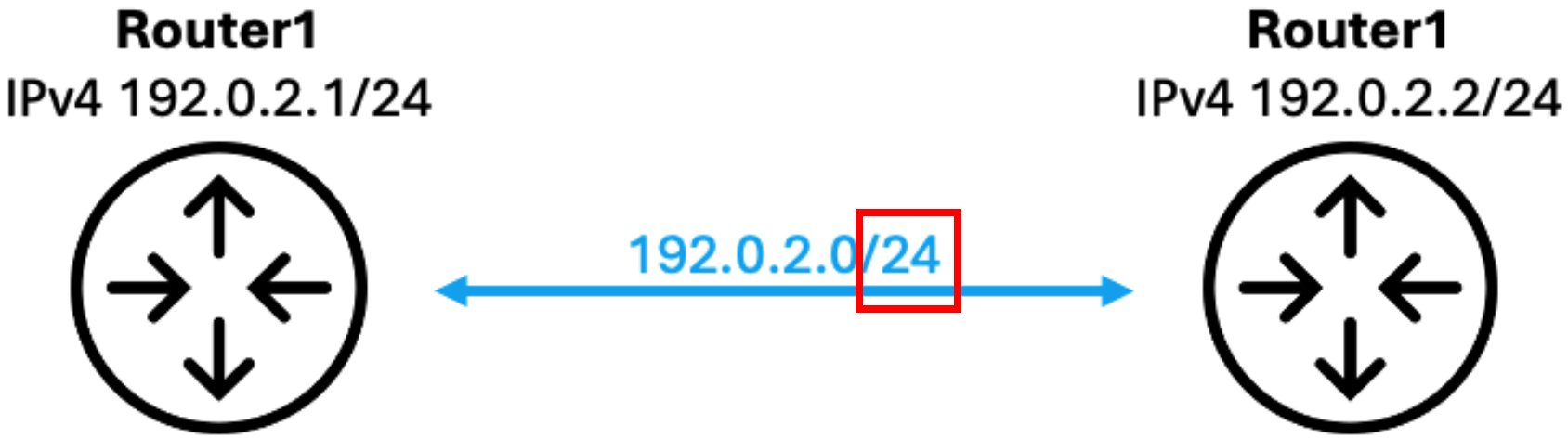
```
bob@bob:~# ip route add 0.0.0.0/0 via 198.51.100.1
bob@bob:~# ip route
default via 198.51.100.1 dev from_bob
198.51.100.0/24 dev from_bob proto kernel scope link src 198.51.100.17
```

IPv4 Neighbor Discovery, IPv4 provoz

```
Router IP 198.51.100.1,  MAC 1e:51:b8:7a:97:80
Bob    IP 198.51.100.17, MAC f6:fe:90:5d:d3:ca
```

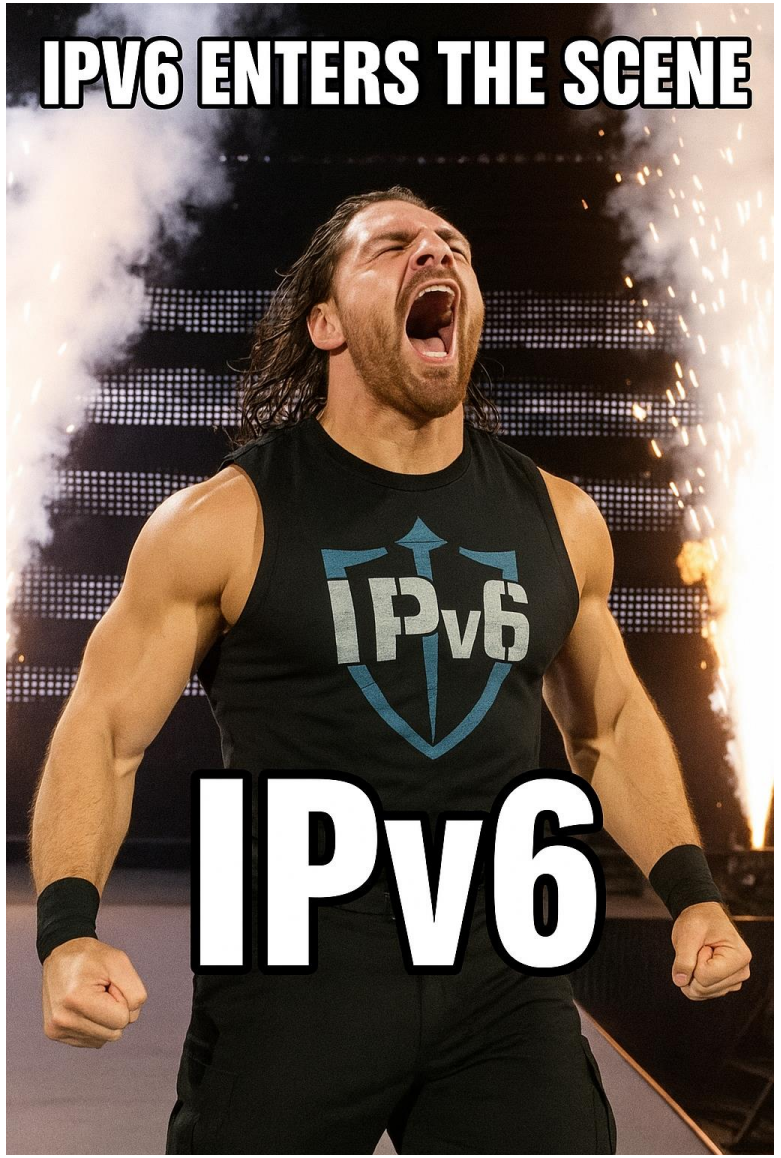


Sítě mezi routery



- /32 Ne ☺
- /31 RFC 3021 (2000)
- /30
- /29
- atd.

Využití masky /32 v IPv4 – Loopback/dummy



Obrázek vytvořen pomocí promptu "Vyrob mi meme, jak IPv6 přichází na scénu!" s využitím ChatGPT-4o (OpenAI, 2025).

Network Working Group
Request for Comments: 5549
Category: Standards Track

F. Le Faucheur
E. Rosen
Cisco Systems
May 2009

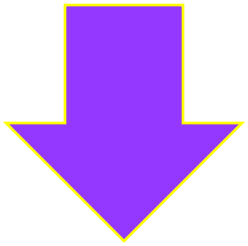
Advertising IPv4 Network Layer Reachability Information with an IPv6 Next Hop

Internet Engineering Task Force (IETF)
Request for Comments: 8950
Obsoletes: 5549
Category: Standards Track
Published: November 2020
ISSN: 2070-1721

S. Litkowski
Cisco
S. Agrawal
Cisco
K. Ananthamurthy
Cisco
K. Patel
Arrcus

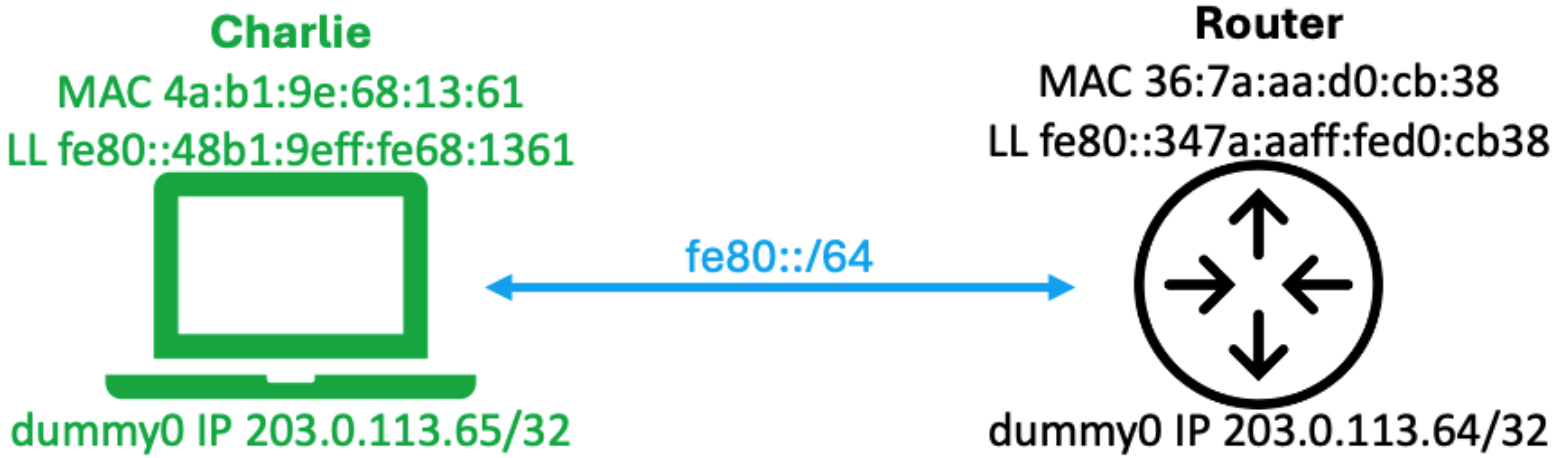
Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop

```
bob@bob:~# ip route add 0.0.0.0/0 via 198.51.100.1
bob@bob:~# ip route
default via 198.51.100.1 dev from_bob
198.51.100.0/24 dev from_bob proto kernel scope link src 198.51.100.17
```



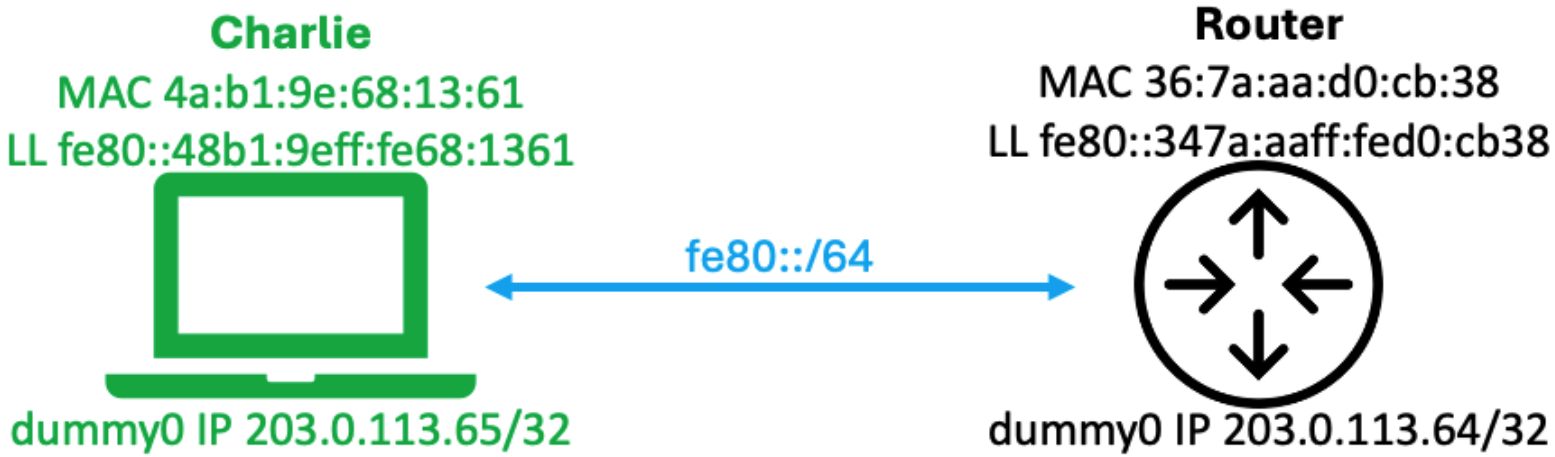
```
bob@bob:~# ip route add 0.0.0.0/0 via inet6 fe80::1 dev from_bob
bob@bob:~# ip route
default via inet6 fe80::1 dev from_bob
198.51.100.0/24 dev from_bob proto kernel scope link src 198.51.100.17
```

Konfigurace rozhraní typu Loopback/dummy



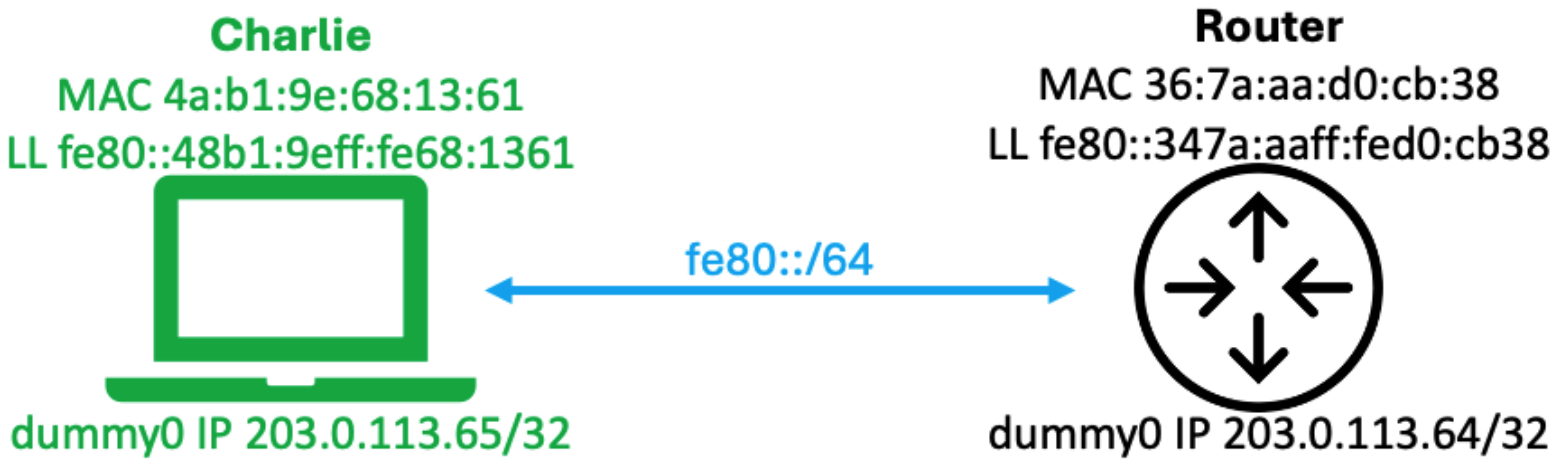

```
charlie@charlie:~# ip a s dev dummy0
3: dummy0: <BROADCAST,NOARP,UP,LOWER_UP> mtu 1500
    link/ether 22:89:1c:bb:97:fe brd ff:ff:ff:ff:ff:ff
    inet 203.0.113.65/32 scope global dummy0
       valid_lft forever preferred_lft forever
operator@router:~# ip a s dev dummy0
76: dummy0: <BROADCAST,NOARP,UP,LOWER_UP> mtu 1500
    link/ether 06:7e:d5:e2:ea:47 brd ff:ff:ff:ff:ff:ff
    inet 203.0.113.64/32 scope global dummy0
       valid_lft forever preferred_lft forever
```

Výchozí brána na zařízení „Charlie“




```
charlie@charlie:~# ip route

charlie@charlie:~# ping 8.8.8.8
ping: connect: Network is unreachable

charlie@charlie:~# ip -4 route add 0.0.0.0/0 via inet6 \
    fe80::347a:aaff:fed0:cb38 dev from_charlie src 203.0.113.65

charlie@charlie:~# ip -4 route
default via inet6 fe80::347a:aaff:fed0:cb38 dev from_charlie src 203.0.113.65
```

Host route na zařízení „Router“

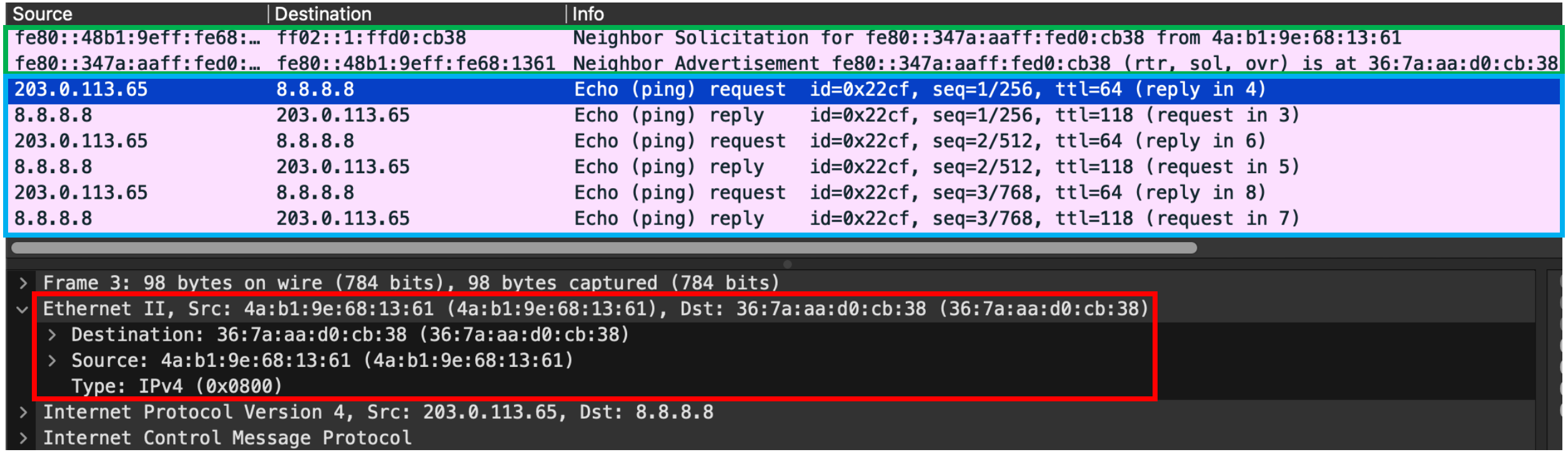



```
operator@router:~# ip route | grep 203.0.113.65
203.0.113.65 dev to_charlie scope link

operator@router:~# arp -n | grep 203.0.113.65  # (arp_ignore == 0)
203.0.113.65  ether  4a:b1:9e:68:13:61  C  to_charlie


10:19:01.537002 ARP, Request who-has 203.0.113.65 tell 203.0.113.64, length 28
10:19:01.537006 ARP, Reply 203.0.113.65 is-at 4a:b1:9e:68:13:61, length 28
```

IPv6 next-hop na zařízení „Router“

```
operator@router:~# ip route add 203.0.113.65 via inet6 \
                   fe80::48b1:9eff:fe68:1361 dev to_charlie

operator@router:~# ip route | grep 203.0.113.65
203.0.113.65 via inet6 fe80::48b1:9eff:fe68:1361 dev to_charlie

10:21:29.516907 IP6 fe80::347a:aaff:fed0:cb38 > ff02::1:ff68:1361: ICMP6,
neighbor solicitation, who has fe80::48b1:9eff:fe68:1361
10:21:29.516914 IP6 fe80::48b1:9eff:fe68:1361 > fe80::347a:aaff:fed0:cb38:
ICMP6, neighbor advertisement, tgt is fe80::48b1:9eff:fe68:1361

charlie@charlie:~# ip -6 neigh
fe80::347a:aaff:fed0:cb38 dev from_charlie lladdr 36:7a:aa:d0:cb:38 router
REACHABLE
```

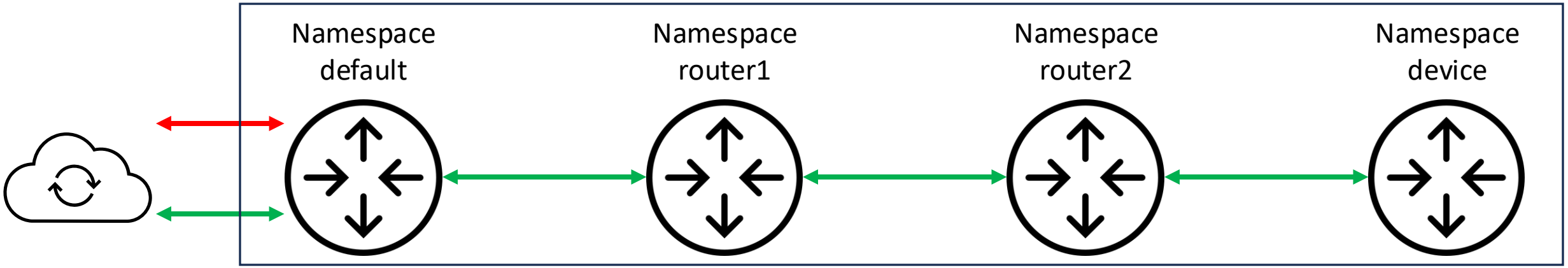
IPv6 Neighbor Discovery, IPv4 provoz



IPv6 a IPv4 next-hopy lze kombinovat

```
root@router:~# ip route | egrep '(default|203.0.113.65)'
default dev ppp0 scope link metric 9
default via 192.0.2.1 dev wan1 metric 10
203.0.113.65 via inet6 fe80::48b1:9eff:fe68:1361 dev to_charlie
```

Chained routers using IPv6 next hops: Hetzner VM



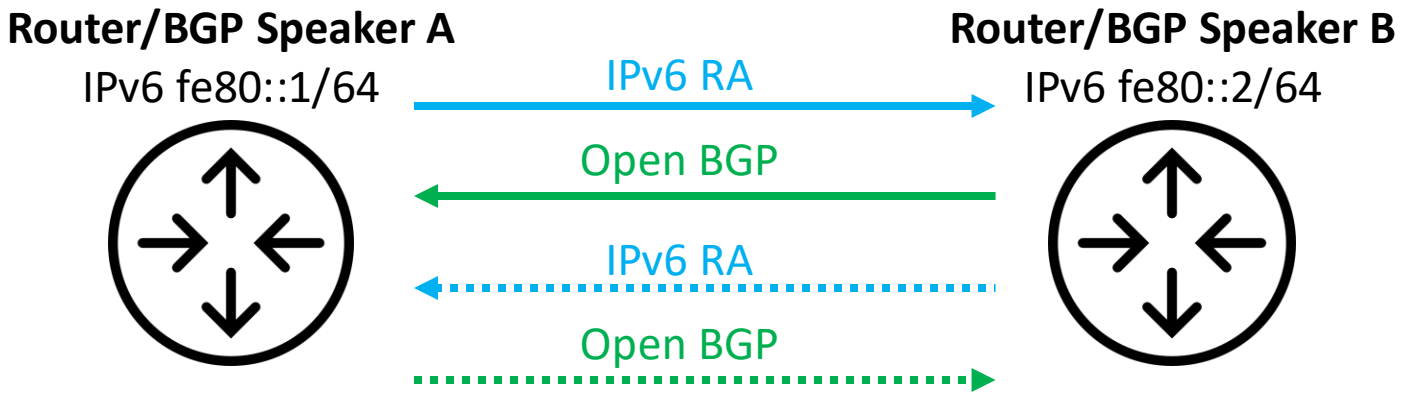

```
 7.|-- spine15.cloud2.fsn1.hetzner.com (2a01:4f8:0:3::5bd)
 8.|-- spine3.cloud2.fsn1.hetzner.com (2a01:4f8:0:e172::a14e)
 9.|-- ???
10.|-- 20371.your-cloud.host (2a01:4f8:0:e172::2635)
11.|-- 2a01:4f8:1c17:7039::1    # default
12.|-- 2a01:4f8:1c17:7039::1:2  # router1
13.|-- 2a01:4f8:1c17:7039::2:2  # router2
14.|-- 2a01:4f8:1c17:7039::4:1  # device
```

```
 7.|-- spine16.cloud2.fsn1.hetzner.com (213.239.227.221)
 8.|-- spine3.cloud2.fsn1.hetzner.com (213.239.239.154)
 9.|-- ???
10.|-- 20371.your-cloud.host (49.12.26.24)
11.|-- ???                 # default
12.|-- ???                 # router1
13.|-- ???                 # router2
14.|-- 188.245.88.112      # device
```

Demo time?

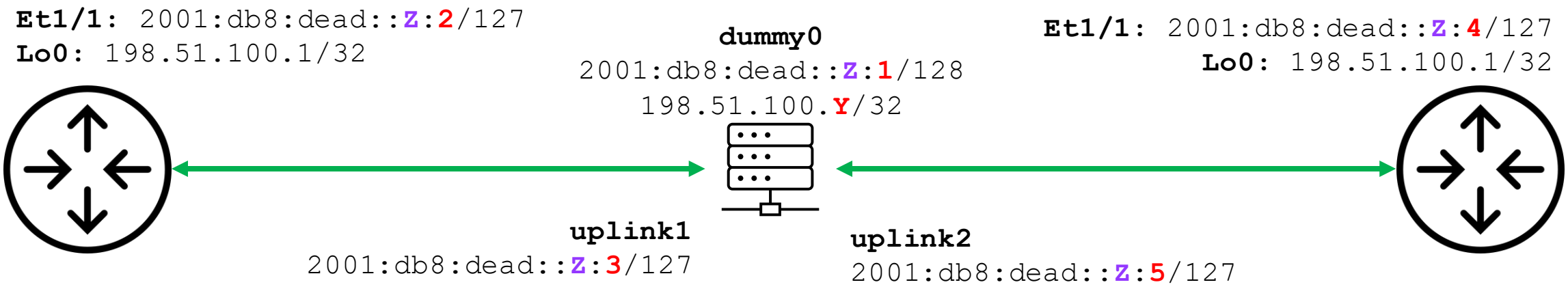
BGP

BGP Unnumbered



- Pro komunikaci protokolu BGP se používají IPv6 link-local adresy místo globálních adres
- Router A vyšle IPv6 router advertisement
- Router B zaslechne IPv6 router advertisement a nakonfiguruje s adresou, ze které přišel RA, BGP session
- Bird umí fungovat pouze v roli Router A (vysílá RA), ale už ne v roli Router B (naslouchá/čeká na RA a sestaví BGP)
- Arista umí fungovat v roli Router B i Router A

Adresní plán



Adresa	Použití
2001:db8:dead::Z:1/128	IPv6 Loopback na serveru
2001:db8:dead::Z:2/127	L3 interface na první Aristě
2001:db8:dead::Z:3/127	Uplink1 interface na serveru
2001:db8:dead::Z:4/127	L3 interface na druhé Aristě
2001:db8:dead::Z:5/127	Uplink2 interface na serveru
198.51.100.Y/32	IPv4 Loopback na serveru
198.51.100.1/32	IPv4 Loopback0 na první Aristě
198.51.100.2/32	IPv4 Loopback0 na druhé Aristě **(ip address unnumbered Loopback0)**

Konfigurace sítě na serveru

```
uplink1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500
    link/ether 1c:34:da:56:3a:d0 brd ff:ff:ff:ff:ff:ff
    inet6 2001:db8:dead::Z:3/127 scope global
    inet6 fe80::f8a8:96ff:febe:58b0/64 scope link

uplink2: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500
    link/ether 1c:34:da:56:3a:d1 brd ff:ff:ff:ff:ff:ff
    inet6 2001:db8:dead::Z:5/127 scope global
    inet6 fe80::1e34:daff:fe56:3ad1/64 scope link

dummy0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500
    inet 198.51.100.Y/32 scope global dummy0
    inet6 2001:db8:dead::Z:1/128 scope global
```

Konfigurace sítě na serveru

```
$ ip route
default proto bird src 198.51.100.Y metric 32
        nexthop via inet6 fe80::d089:20ff:fe64:f39a dev uplink1 weight 1
        nexthop via inet6 fe80::5001:68ff:fef4:170e dev uplink2 weight 1

$ ip -6 route
::1 dev lo proto kernel metric 256 pref medium
2001:db8:dead::Z:1 dev dummy0 proto kernel metric 256 pref medium
2001:db8:dead::Z:2/127 dev uplink1 proto kernel metric 256 pref medium
2001:db8:dead::Z:4/127 dev uplink2 proto kernel metric 256 pref medium
fe80::/64 dev dummy0 proto kernel metric 256 pref medium
fe80::/64 dev uplink2 proto kernel metric 256 pref medium
fe80::/64 dev uplink1 proto kernel metric 256 pref medium
default proto bird src 2001:db8:dead::Z:1 metric 32 pref medium
        nexthop via fe80::d089:20ff:fe64:f39a dev uplink1 weight 1
        nexthop via fe80::5001:68ff:fef4:170e dev uplink2 weight 1
default via 2001:db8:dead::Z:2 dev uplink1 proto static metric 100
default via 2001:db8:dead::Z:4 dev uplink2 proto static metric 200
```

Bird

```
router id 198.51.100.Y;

protocol device {
    scan time 10;
}

# Direct, IPv4
protocol direct {
    ipv4;                           # Connect to default IPv4 table
    interface "dummy*";
}

# Direct, IPv4
protocol direct {
    ipv6;                           # ... and to default IPv6 table
    interface "dummy*";
}
```

Bird

```
router id 198.51.100.Y;

# Kernel, IPv4
protocol kernel {
    ipv4 {
        export filter {
            krt_prefsrc = 198.51.100.Y;
            if (source = RTS_BGP) then accept;
            reject;
        };
        import none;
    };
    persist;
    scan time 20;
    merge paths on;    # Add multipath to route to kernel e.g. for ECMP egress
}
```

Bird

```
router id 198.51.100.Z;

# Kernel, IPv6
protocol kernel {
    ipv6 {
        export filter {
            krt_prefsrc = 2001:db8:dead::Z:1;
            if (source = RTS_BGP) then accept;
            reject;
        };
        import none;
    };
    persist;
    scan time 20;
    merge paths on;    # Add multipath to route to kernel e.g. for ECMP egress
}
```

Bird

```
# BGP to advertise the loopback (dummy) addresses
template bgp spine {
  direct;
  # ...
  dynamic name "uplink";
  local as 4200000000;
  # IP unnumbered, using router advertisements to tell Aristas we're a router
  neighbor range fe80::/64 as 64511;
  # (...)
  ipv4 {
    extended next hop;
    import all;
    export filter {
      if (net.len = 32 && source = RTS_DEVICE) then accept;
      reject;
    };
  };
# (...)
```

Bird

```
# BGP to advertise the loopback (dummy) addresses
# (... continuing)
  ipv6 {
    import all;
    export filter {
      if (net.len = 128 && source = RTS_DEVICE) then accept;
      reject;
    };
  };
}

protocol bgp spine1 from spine {
    interface "uplink1";
}

protocol bgp spine2 from spine {
    interface "uplink2";
}
```

Bird

```
# This is necessary for the unnumbered BGP sessions to work
# Aristas wait for the RA and then initiate connection to us
protocol radv {
    propagate routes no;
    ipv6 {
        import none;
        export none;
    };
    interface "uplink*" {
        max ra interval 15;
    };
}
```

Arista – L3 port

```
ipv6 dhcp relay option link-layer address

interface Ethernet1/1
   description aristaX-server1
   no switchport
   ip proxy-arp
   ip address unnumbered Loopback0
   ipv6 dhcp relay destination 2001:db8:babe::2
   ipv6 address 2001:db8:dead::Z:X/127
   ipv6 nd ra mtu suppress
   ipv6 nd managed-config-flag
   ipv6 nd reachable-time 900000
   ipv6 nd prefix 2001:db8:dead::Z:X/127 no-onlink no-autoconfig

interface Loopback0
   ip address 198.51.100.1/32
   ipv6 address 2001:db8:dead::1/128
```

Arista

```
router bgp 64511
   neighbor SERVERS peer group
   neighbor SERVERS remove-private-as
   neighbor SERVERS graceful-restart
   neighbor interface Et1/1 peer-group SERVERS remote-as 4200000000
   !
   address-family ipv4
      neighbor SERVERS activate
      neighbor SERVERS default-originate always
      neighbor SERVERS next-hop address-family ipv6 originate
   !
   address-family ipv6
      neighbor SERVERS activate
      neighbor SERVERS default-originate always
!
```

Arista

```
# sh ip bgp summary | include fe80::f8a8:96ff:febe:58b0
  fe80::f8a8:96ff:febe:58b0%Et1/1 4 4200000000  0    0  42d17h Estab   1       1

# show ipv6 bgp peers fe80::f8a8:96ff:febe:58b0%Et1/1 received-routes
BGP routing table information for VRF default
Router identifier 198.51.100.1, local AS number 64511

    Network                Next Hop                            Path
* > 198.51.100.Z/32       fe80::f8a8:96ff:febe:58b0%Et1/1  4200000000 i
* > 2001:db8:dead::Z:1/128 fe80::f8a8:96ff:febe:58b0%Et1/1  4200000000 i
```

Q & A

linkedin.com/in/radek-zajic/

@zajDee