amsix
amsterdam internet exchange

# RFC 5549

## BGP IPv4 NLRIs with an IPv6 next hop

RIPE-65 Amsterdam 25-09-2012

arien.vijn@ams-ix.net

stefan.plug@ams-ix.net

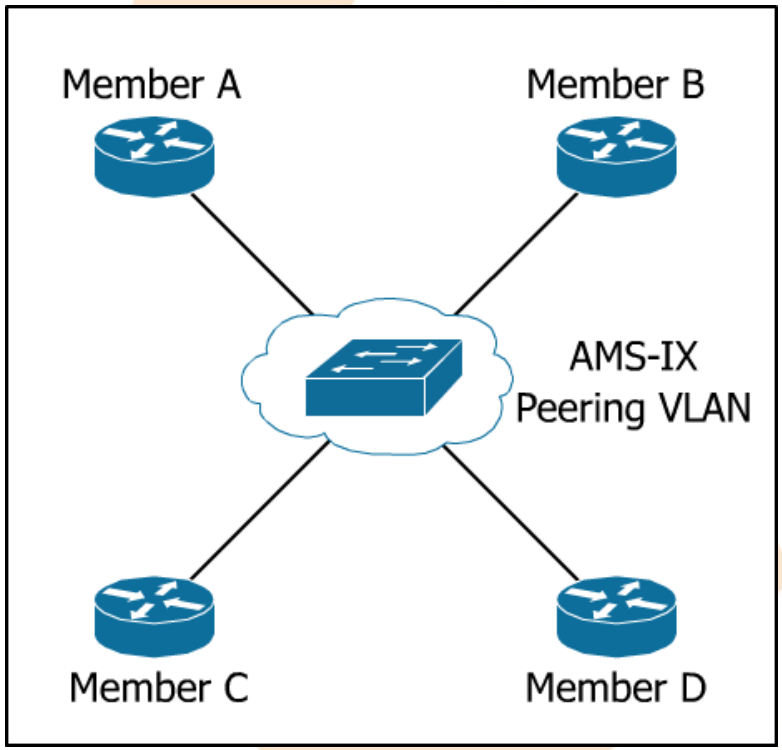
# Overview



- No more IPv4!
- Possible solutions
- RFC 5549

# No more IPv4!

- AMS-IX Internet peering VLAN:
- 1024 IPv4 (/22)
- 18,446,744,073,709,551,616 IPv6 (/64)

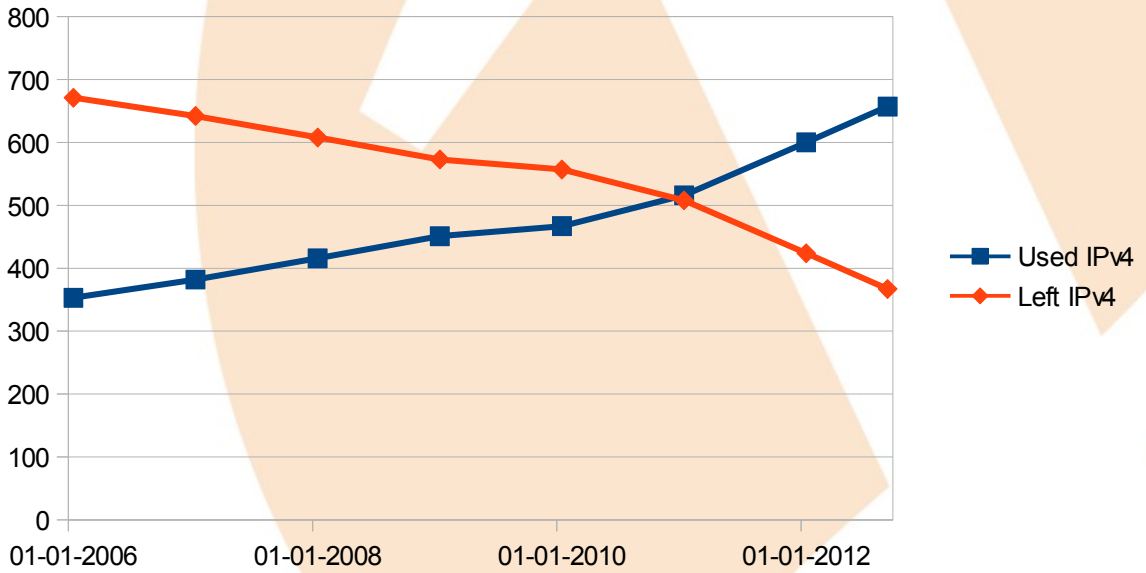


# No more IPv4!

## IPv4 usage 2006-2012

Latest data: 11-sept-2012

- Used: 657
- Left: 367



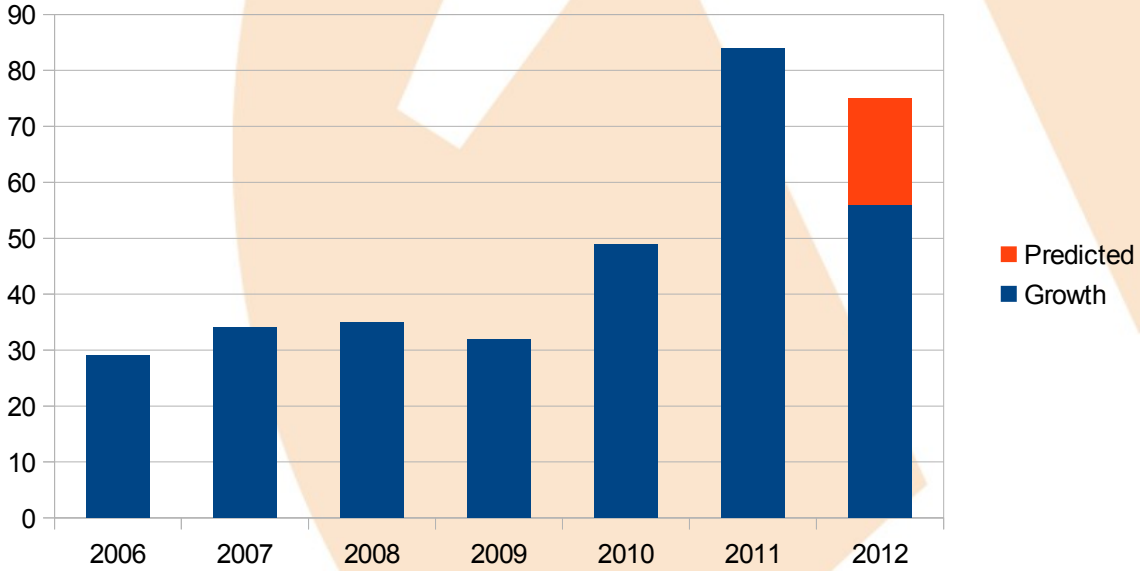
# No more IPv4!

## Growth of IPv4 usage

Latest data: 11-sept-2012

- 56, but we still have 4 months to go..

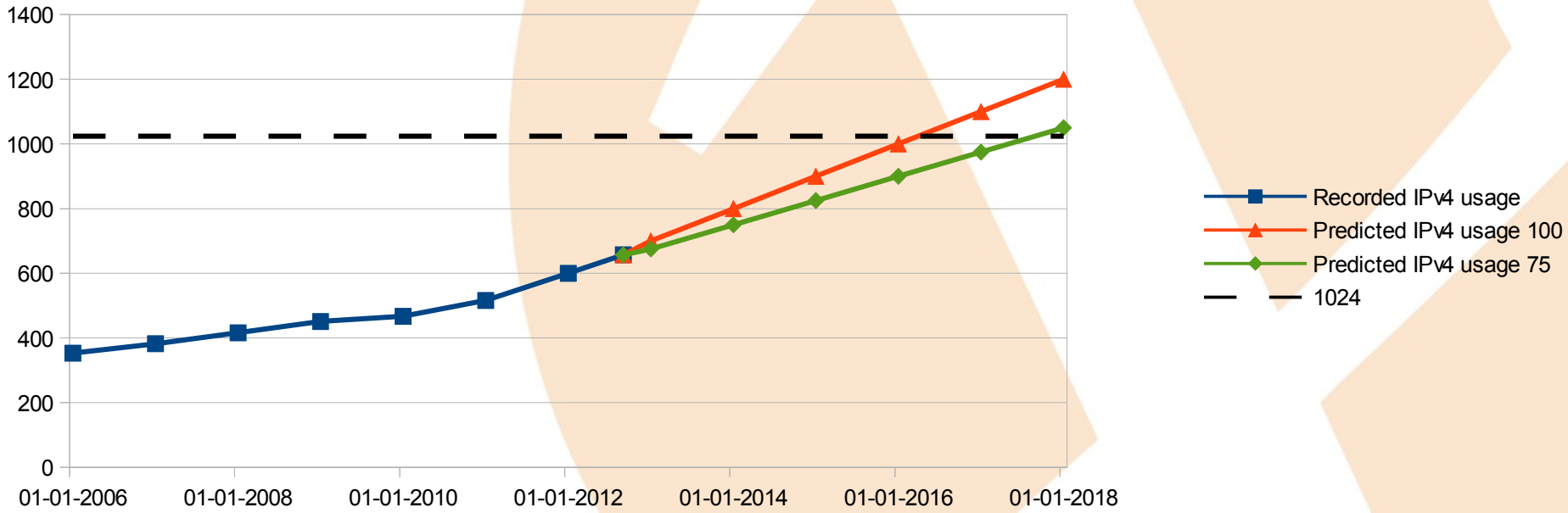
Probable reason: AMS-IX reseller program



# No more IPv4!

## Predicted growth of IPv4 usage

- Worst case: 100 each year, depletion **2016**
- Best case: 75 each year, depletion in **2017**



# Overview

- No more IPv4!
- Possible solutions
- RFC 5549

# Possible solutions

## #1: Bigger subnet (/21)

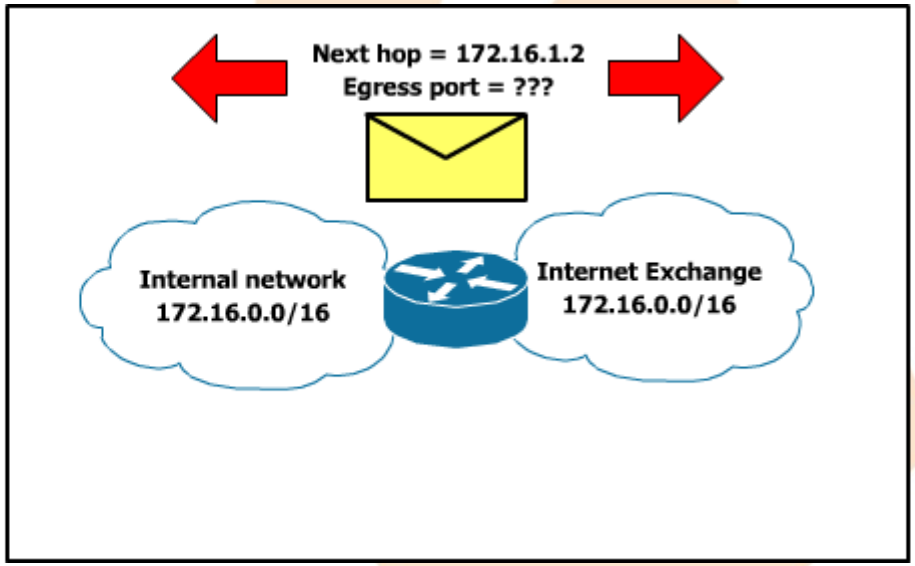
- Because IPv4 address space is unlimited!... Oh wait...
- Could give us +-10 more years
- Bigger IPv4 broadcast domain == more ARP problems
- **Not a solution, just a workaround**

# Possible solutions



## #2: Private address space

- No member (500+) could use that private range anymore
- Including management interfaces
- ACLs deny private range traffic

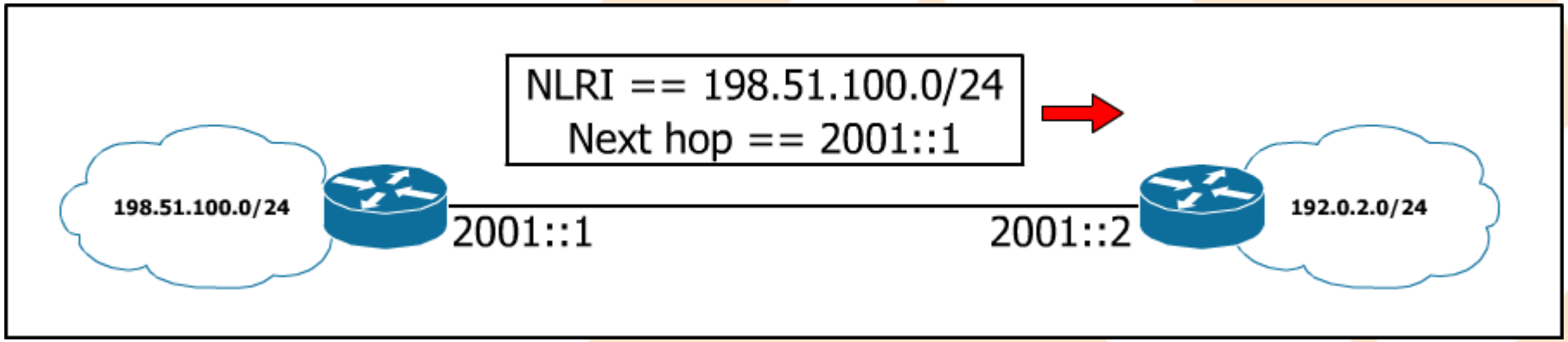


# Possible solutions



## #3: Advertising IPv4 next-hops

Could we use an IPv6 address as a next hop for its IPv4 routes?

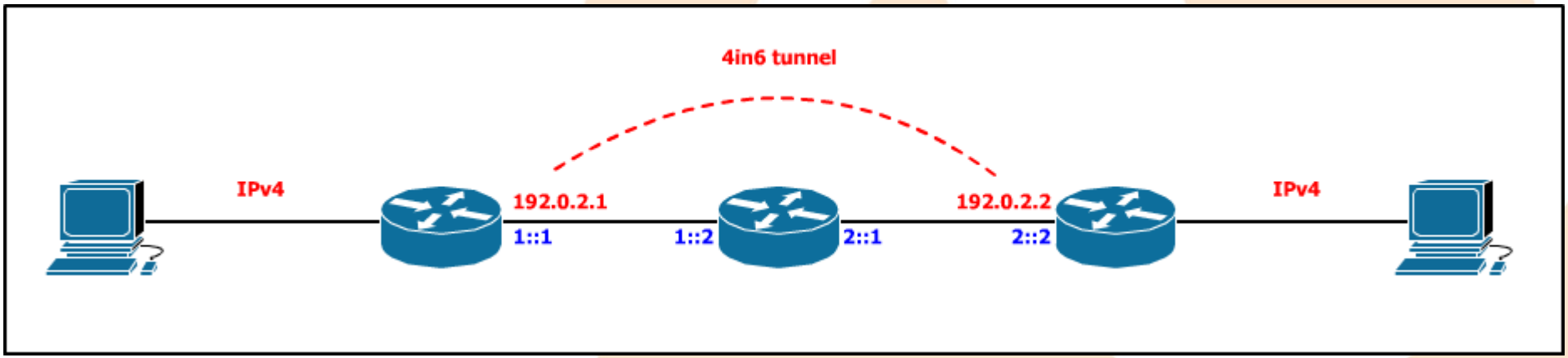


# Possible solutions



## #3.1: 4in6 tunnel

- With what IPv4 addresses as termination points?
- Wouldn't solve our problem

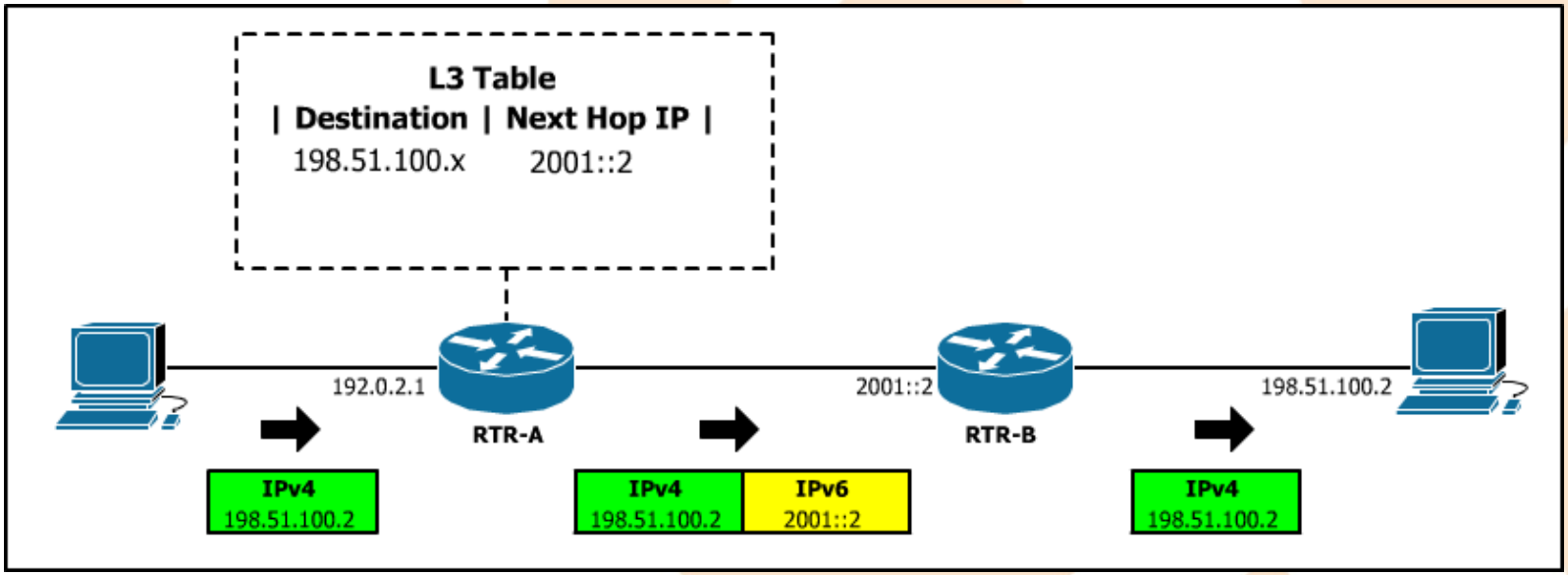


# Possible solutions



## #3.2: '4in6' Softwire mesh tunnel [RFC 5565]

- Tunnel terminates in the router itself
- Encapsulate every IPv4 packet (+40B per packet)
- Non-IPv4 members can participate in IPv4 routing

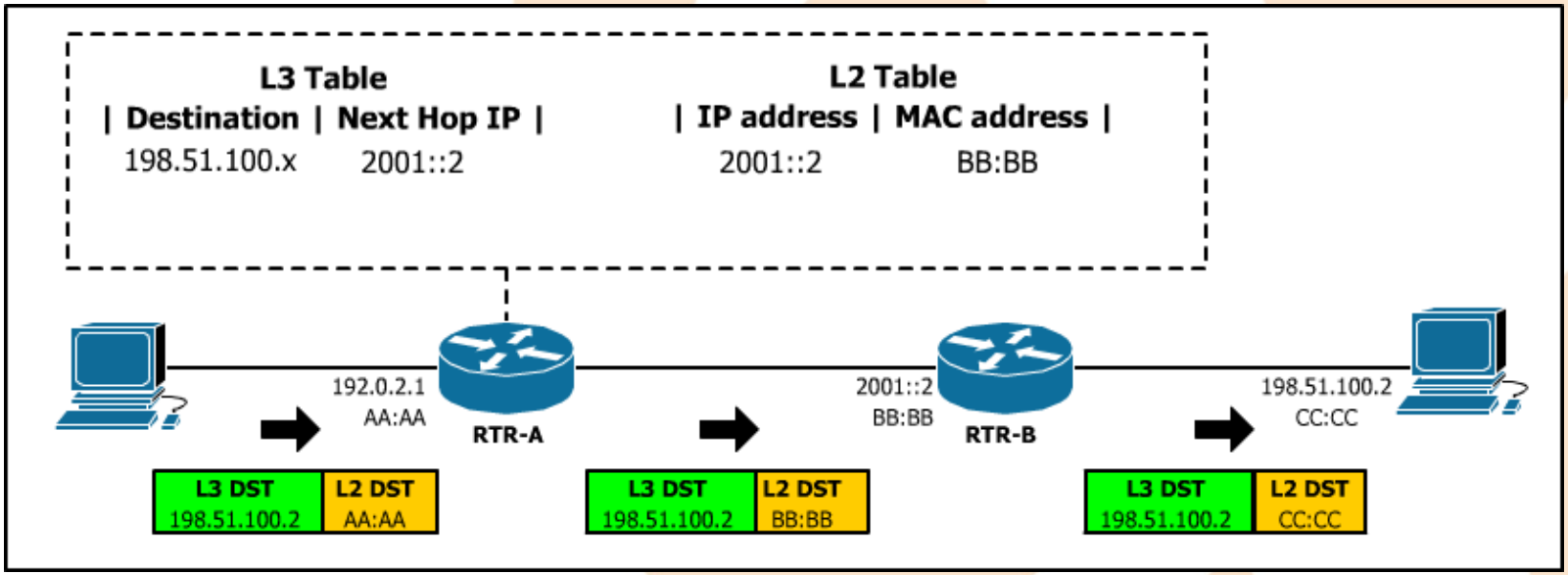


# Possible solutions



## #3.3: Just send out the IPv4 packet

- We actually only need a L2 address for the next hop
- No need for encapsulation == no overhead
- But what if the next hop is not dual stack?



# RFC 5549

- No more IPv4!
- Possible solutions
- RFC 5549

# RFC 5549

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

# RFC 5549



***Receive an MP_BGP update message***

***Look at the MP_REACH_NLRI attribute***

```
IF ((Update AFI == IPv4)
   &&
   (Length of next hop == 16 Bytes || 32 Bytes))
{
   This is an IPv4 route with an IPv6 next hop;
}
```

# RFC 5549



## Example MP_REACH_NLRI attribute:

- AFI: = 1 (IPv4)
- SAFI: = 1 (unicast)
- Next hop length = 32
- Next hop address = 2001::2 & FE80::1
- NLRI = 192.0.2.0/24

# RFC 5549

```
▼ MP_REACH_NLRI (44 bytes)
  ▶ Flags: 0x80 (Optional, Non-transitive, Complete)
    Type code: MP_REACH_NLRI (14)
    Length: 41 bytes
    Address family: IPv4 (1)
    Subsequent address family identifier: Unicast (1)
  ▼ Next hop network address (32 bytes)
      Next hop: Optional, Non-transitive, Complete32.1.0.0 (4)
      Next hop: Optional, Non-transitive, Complete32.1.0.00.0.0.0 (4)
      Next hop: Optional, Non-transitive, Complete32.1.0.00.0.0.00.0.0.0 (4)
      Next hop: Optional, Non-transitive, Complete32.1.0.00.0.0.00.0.0.00.0.0.2 (4)
      Next hop: Optional, Non-transitive, Complete32.1.0.00.0.0.00.0.0.00.0.0.2254.128.0.0 (4)
      Next hop: Optional, Non-transitive, Complete32.1.0.00.0.0.00.0.0.00.0.0.2254.128.0.00.0.0.0 (4)
      Next hop: Optional, Non-transitive, Complete32.1.0.00.0.0.00.0.0.00.0.0.2254.128.0.00.0.0.010.0.39.255 (4)
      Next hop: Optional, Non-transitive, Complete32.1.0.00.0.0.00.0.0.00.0.0.2254.128.0.00.0.0.010.0.39.255254.206.11.23 (4)
    Subnetwork points of attachment: 0
  ▼ Network layer reachability information (4 bytes)
    ▼ 192.168.2.0/24
        MP Reach NLRI prefix length: 24
        MP Reach NLRI IPv4 prefix: 192.168.2.0 (192.168.2.0)

0  88 d0 80 18 06 f9 40 83  00 00 01 01 08 0a 00 ee   ......@. ........
0  40 da 00 3a 12 50 ff ff  ff ff ff ff ff ff ff ff   @..:.P.. ........
0  ff ff ff ff ff ff 00 58  02 00 00 00 41 40 01 01   .......X ....A@..
0  00 50 02 00 06 02 01 00  00 00 02 80 04 04 00 00   .P...... ........
0  00 00 80 0e 29 00 01 01  20 20 01 00 00 00 00 00   ....)...  ......
0  00 00 00 00 00 00 00 00  02 fe 80 00 00 00 00 00   ........ ........
0  00 0a 00 27 ff fe ce 0b  17 00 18 c0 a8 02         ...'.... ......
```

# RFC 5549



## BGP capability advertisement

Used to signal BGP capabilities between peers in OPEN message

| | | |
|---|---|---|
| Capability code: | 5 | Extended Next Hop Encoding |
| NLRI AFI: | 1 | IPv4 |
| NLRI SAFI: | 1 | Unicast |
| Nexthop AFI: | 2 | IPv6 |

# RFC 5549



```
▼ Border Gateway Protocol
  ▼ OPEN Message
      Marker: 16 bytes
      Length: 63 bytes
      Type: OPEN Message (1)
      Version: 4
      My AS: 2
      Hold time: 180
      BGP identifier: 2.2.2.2
      Optional parameters length: 34 bytes
    ▼ Optional parameters
      ▶ Capabilities Advertisement (8 bytes)
      ▼ Capabilities Advertisement (10 bytes)
          Parameter type: Capabilities (2)
          Parameter length: 8 bytes
        ▼ Unknown capability (8 bytes)
            Capability code: Unknown capability (5)
            Capability code: Unknown (5)
            Capability length: 6 bytes
            Capability value: Unknown
      ▶ Capabilities Advertisement (4 bytes)
      ▶ Capabilities Advertisement (4 bytes)
      ▶ Capabilities Advertisement (8 bytes)

0020  a5 00 67 77 00 04 20 01  07 f8 00 01 00 00 00 00   ..gw.. . ........
0030  a5 04 55 40 00 01 8e f4  00 b3 f4 bd 06 fb 21 7e   ..U@.... ......!~
0040  96 04 50 18 38 40 57 0f  00 00 ff ff ff ff ff ff   ..P.8@W. ........
0050  ff ff ff ff ff ff ff ff  ff ff 00 3f 01 04 00 02   ........ ...?....
0060  00 b4 02 02 02 02 22 02  06 01 04 00 02 00 01 02   ......". ........
0070  08 05 06 00 01 00 01 00  02 02 02 80 00 02 02 02   ........ ........
0080  00 02 06 41 04 00 00 00  02                       ...A.... .
```

# RFC 5549



## Possible NLRI AFI/SAFI combinations

AFI= IPv4 (1)

SAFI

- Unicast (1)
- Multicast (2)
- Include an MPLS Label (4)
- MPLS-labeled VPN address (128)

Next hop = IPv6

# RFC 5549

## Pros:

- IXPs wouldn't need IPv4 anymore
- No more ARP flooding

## Cons:

- Forwarding might be complicated
- No real implementations yet (as fas as we know)
- Members need to implement this, not AMS-IX


Questions/comments/brilliant solutions:

arien.vijn@ams-ix.net
stefan.plug@ams-ix.net