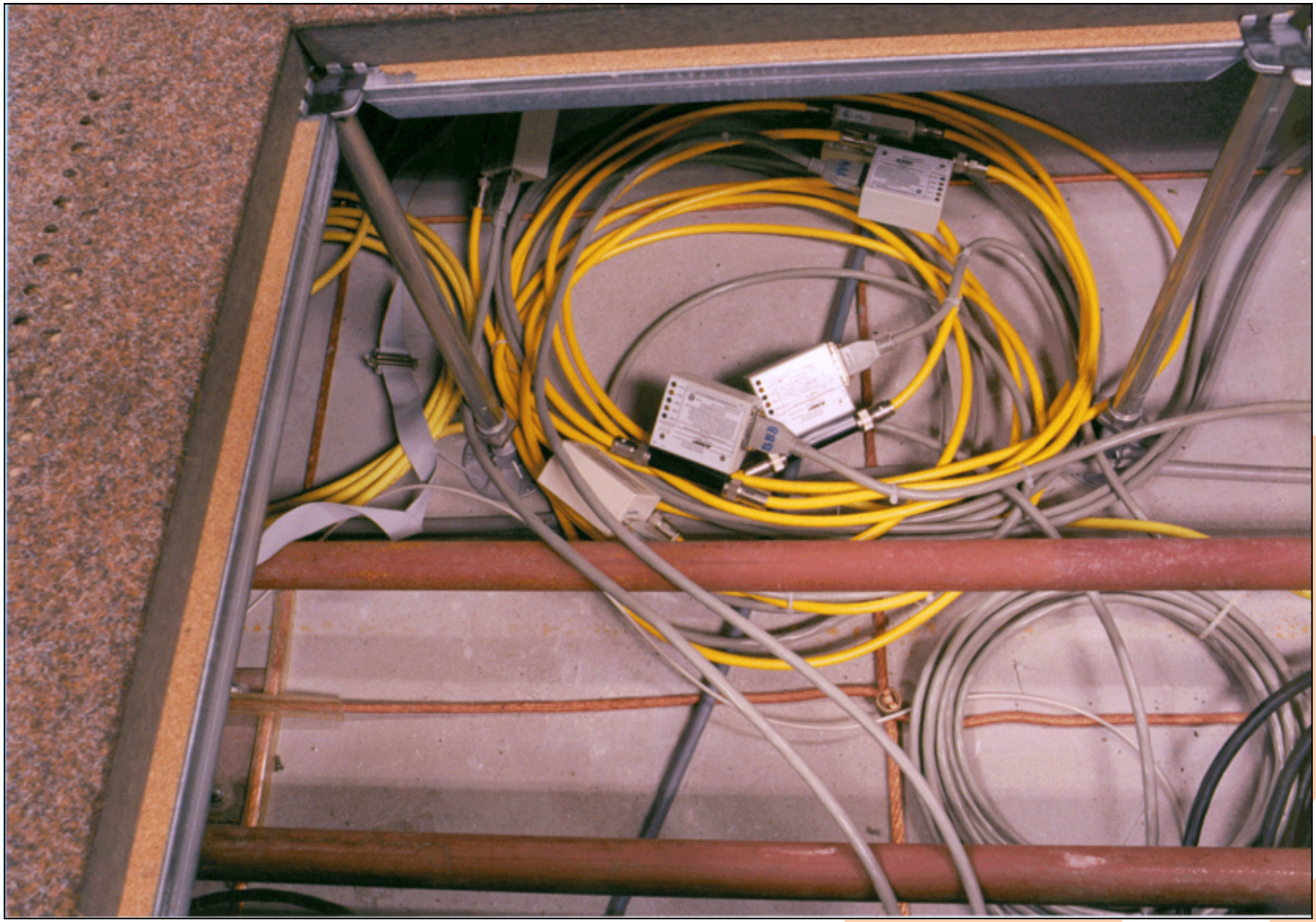
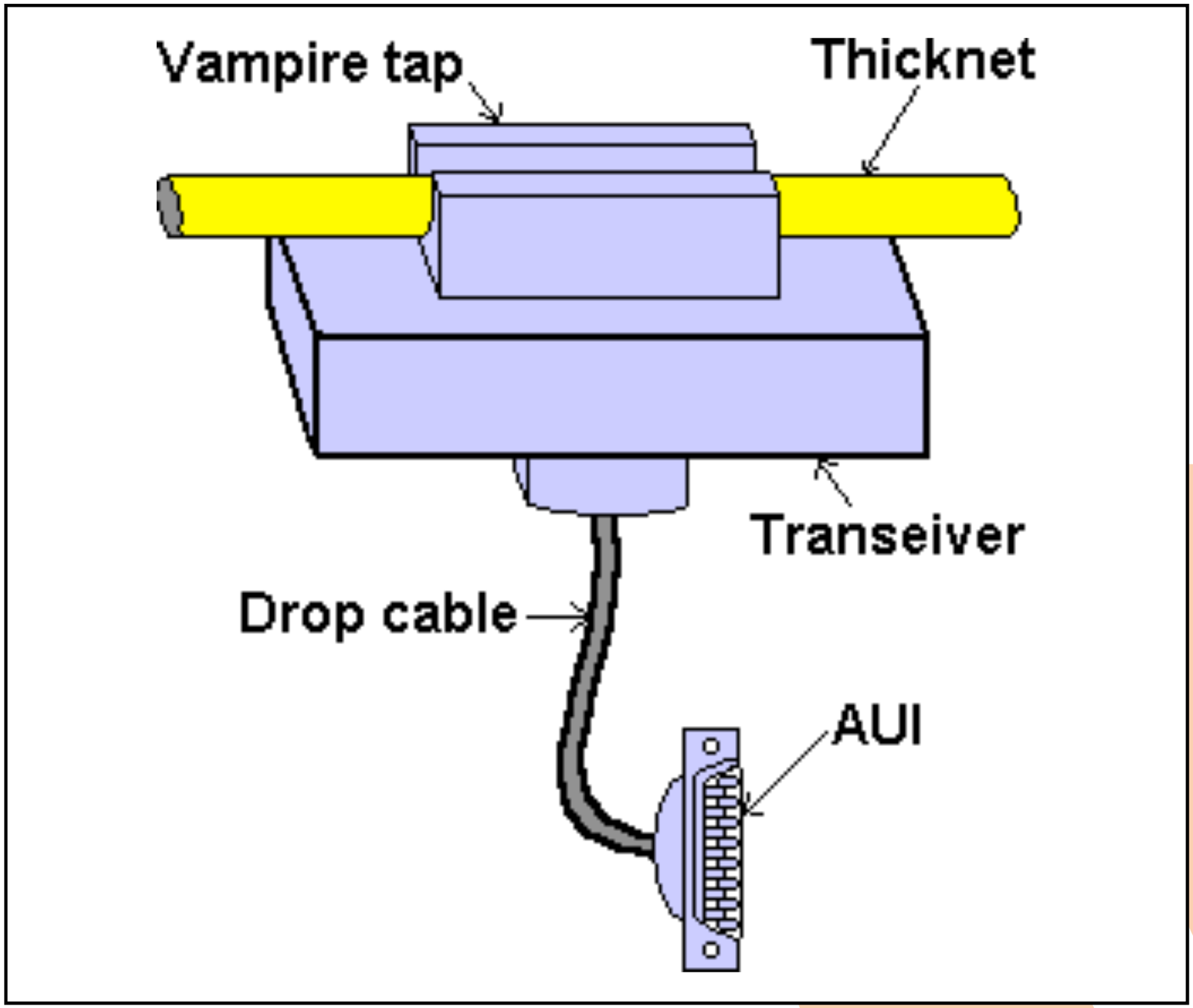
RFC5549

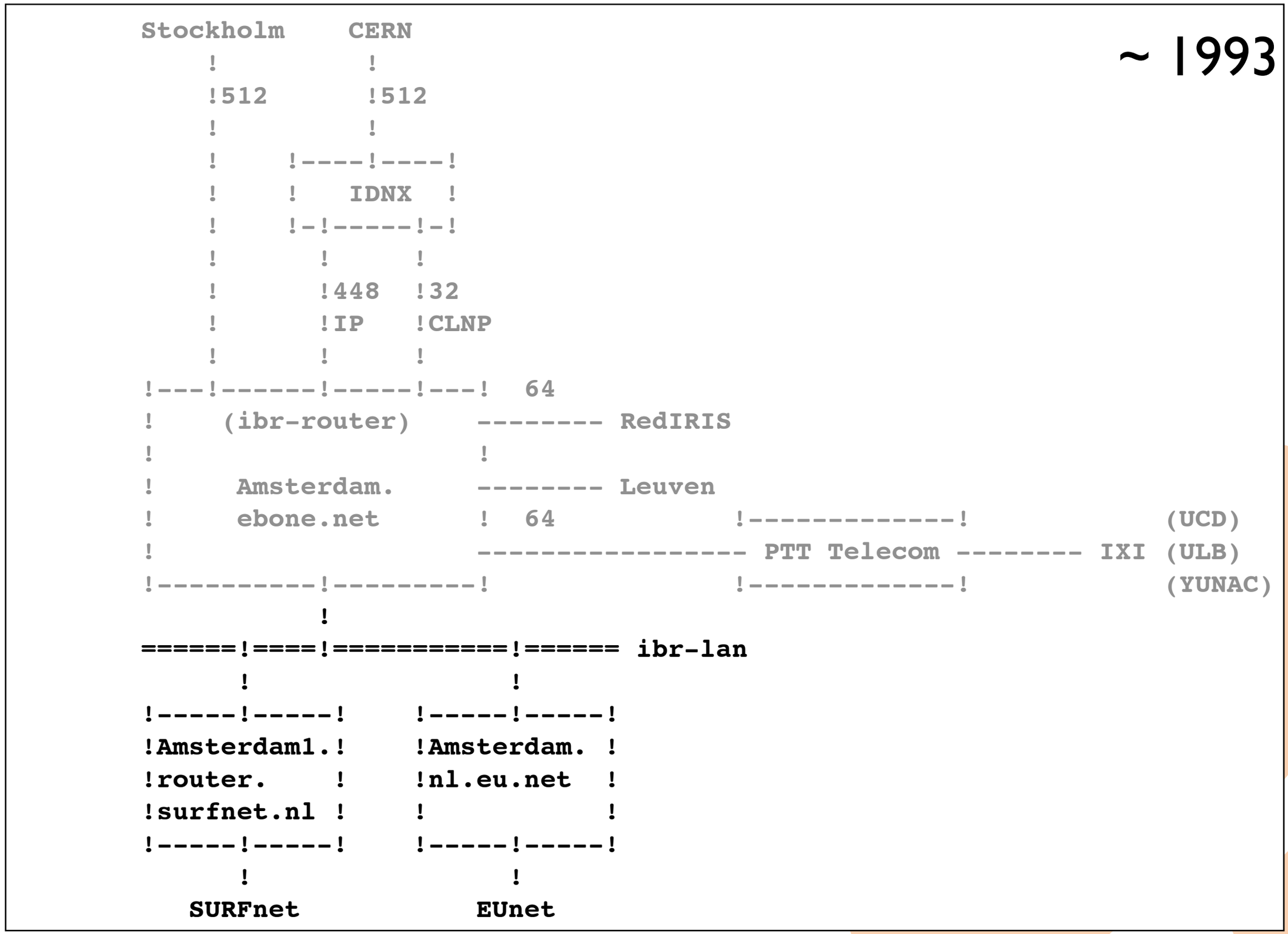
Advertising IPv4 NLRI with an IPv6 Next Hop

Ariën Vijn
arien.vijn@ams-ix.net




Vampire tap
Thicknet
Transeiver
Drop cable
AUI

~ 1993


Stockholm
CERN
512
512
IDNX
448
IP
32
CLNP
64
(ibr-router)
RedIRIS
Amsterdam.
ebone.net
Leuven
64
PTT Telecom
IXI
(UCD)
(ULB)
(YUNAC)
ibr-lan
Amsterdam1.
router.
surfnet.nl
Amsterdam.
nl.eu.net
SURFnet
EUnet

Agenda

- AMS-IX
- Problem statement
- Possible solutions
- RFC5549
- Proof of concept implementation

Agenda

- AMS-IX
- Problem statement
- Possible solutions
- RFC5549
- Proof of concept implementation

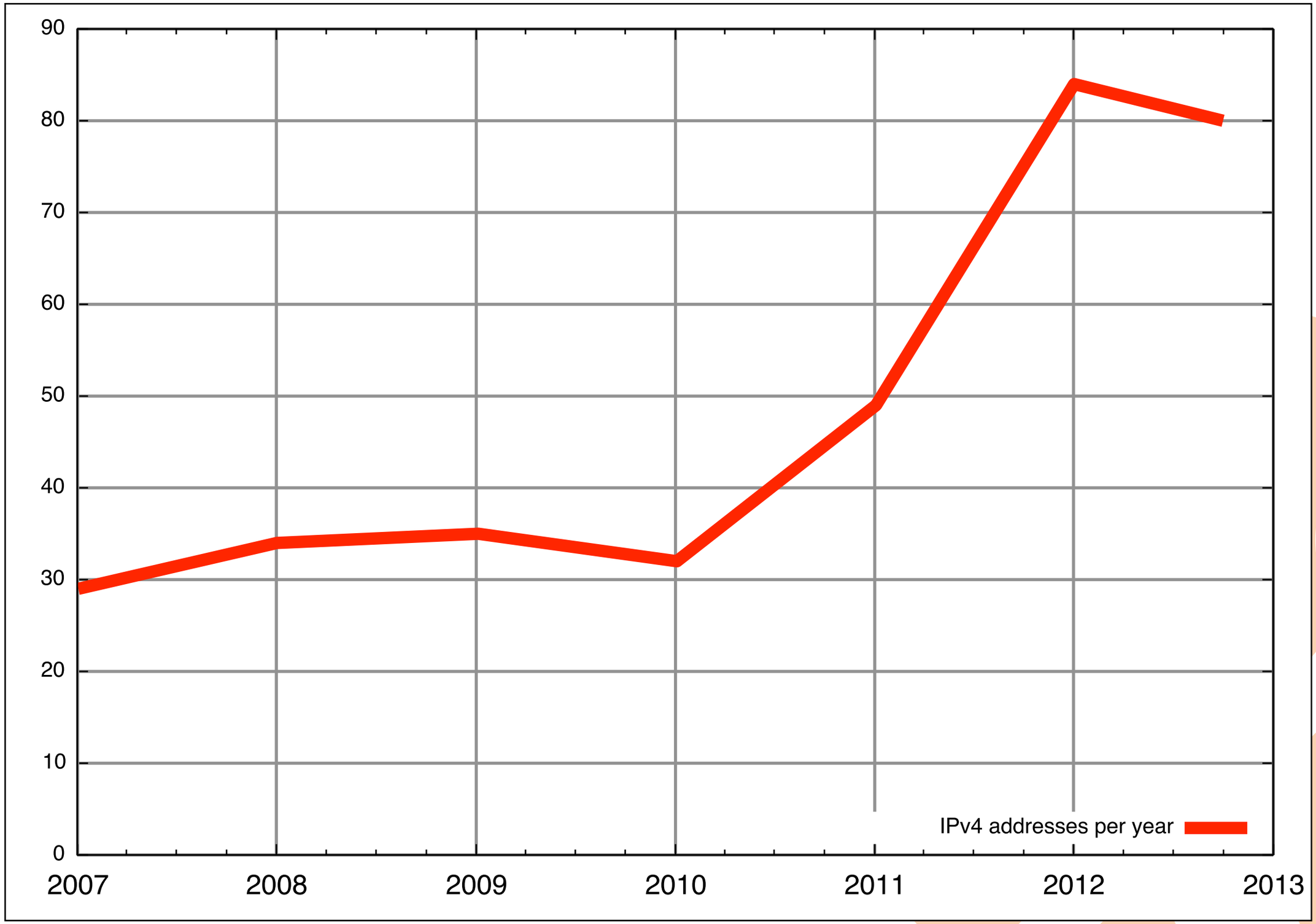
AMS-IX

- Amsterdam Internet Exchange
  - Not for profit IXP.
  - 521 Networks (ASes) connected.
  - 1903Gbit/s peak
    - 6Gbit/s native IPv6
  - 1003 customer ports.
  - 11 Operational sites.
    - Equinix AM3 almost ready.

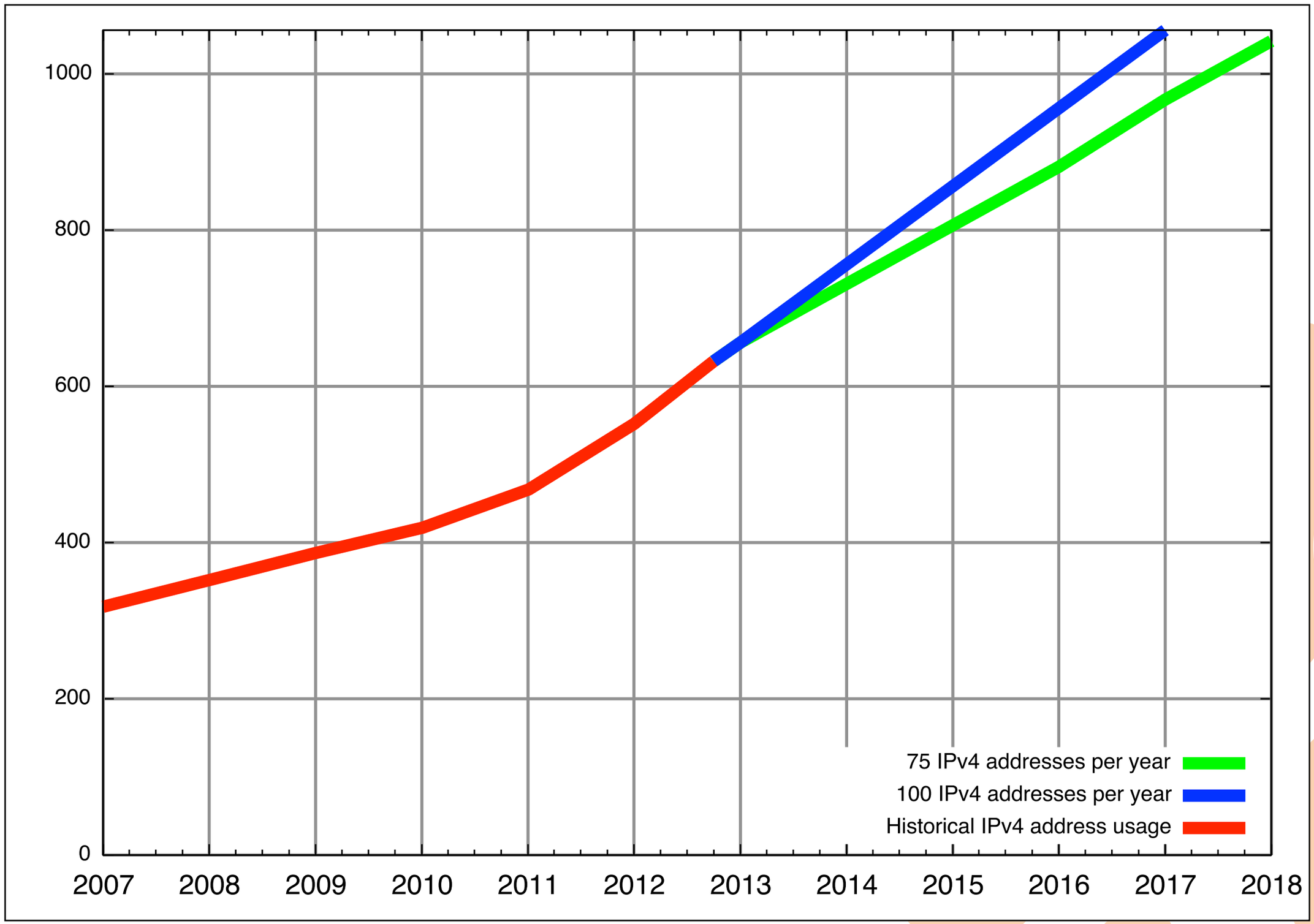
Agenda

- AMS-IX
- Problem statement
- Possible solutions
- RFC5549
- Proof of concept implementation

IPv4 addresses per year



Predictions



Problem statement

- We are facing IPv4 addresses depletion.
  - 100 per year, depletion 2016
  - 75 per year, depletion in 2017

Agenda

- AMS-IX
- Problem statement
- Possible solutions
- RFC5549
- Proof of concept implementation

Solution #1

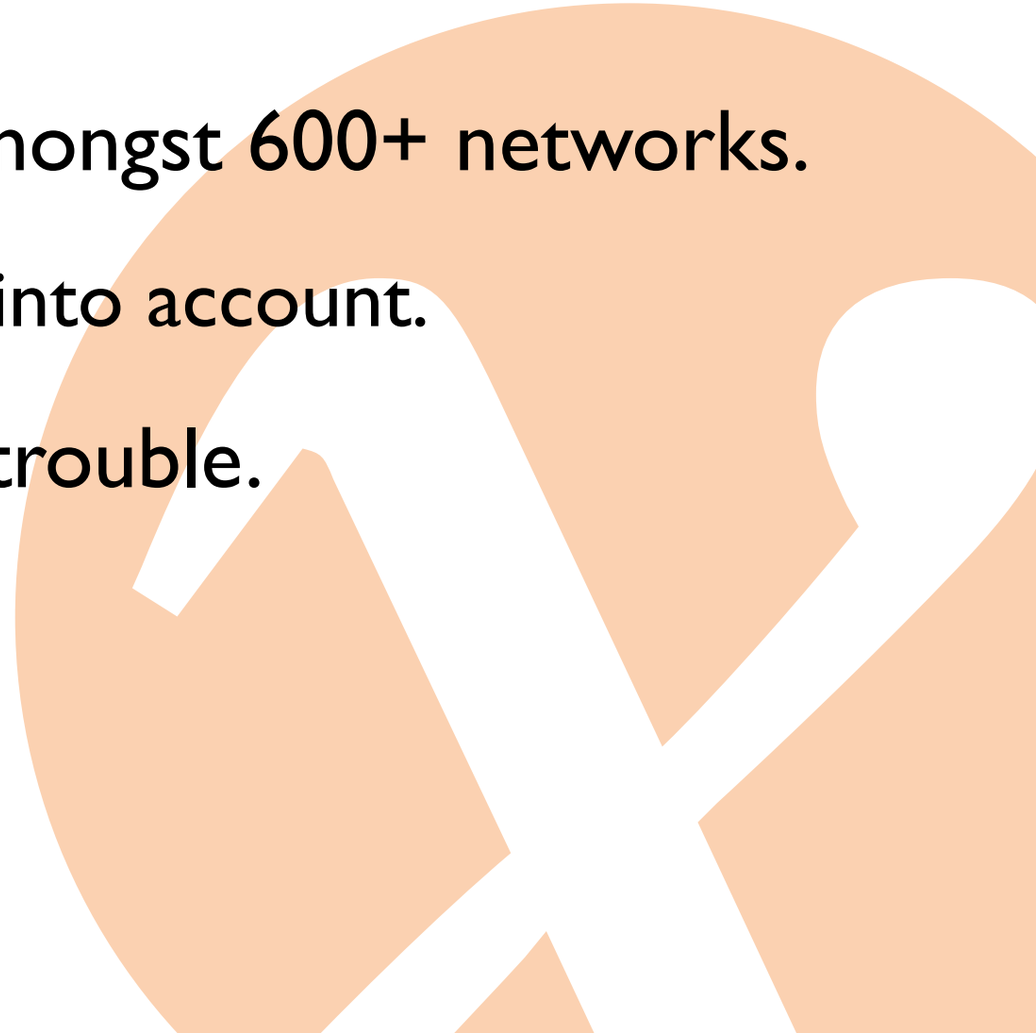
- Secure more unique addresses when it is still possible.
  - Initial allocation for a new LIR was a /21.
  - Works... but no fun.
  - ARP broadcast rate might be a problem.

Solution #2

- Abandon the one shared medium.
  - Non-contiguous address blocks.
    - Allows for small blocks.
    - Migration to 802.1q tagged interfaces.
    - Route server usage is growing.
      - Currently route servers 537 IPv4 sessions.
      - Second route servers LAN.
        - Large and small route servers.

Solution #3

- RFC1918 (10/8, 172.16/12, 192.168/16)
  - These address blocks are in use within the connected networks.
  - Find a “unique” or “usable” /21 amongst 600+ networks.
    - New networks have to take this into account.
  - More specifics in IGPs will cause trouble.
  - Breaks traceroute.



Solution #4

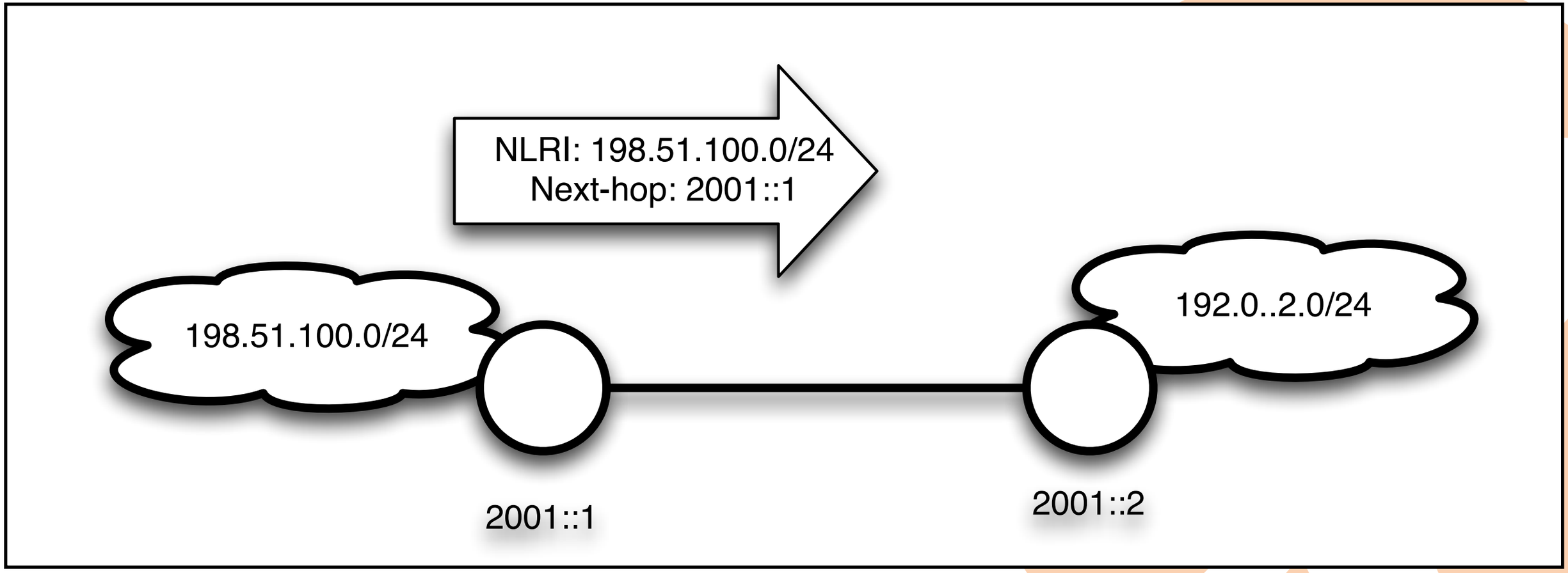
- RFC3927 (169.254/16)
  - "This specification is intended for use with small ad hoc networks on a single link containing only a few hosts."
- Might be in use by connected networks.
  - Unlike link-local IPv6 (fe80::/64) there is no interface scope.
  - draft-kato-bgp-ipv6-link-local-00.txt only applies to IPv6.

Solution #4

- RFC3927 (169.254/16)
  - Many routers are connected to IXP peering LANs.
    - MAC addresses on AMS-IX, DECIX, LINX en NL-IX:
      - 95 addresses on 2 of 4 IXPs
      - 14 addresses on 3 of 4 IXPs
      - 2 addresses on all 4 IXps.
- Breaks traceroute.

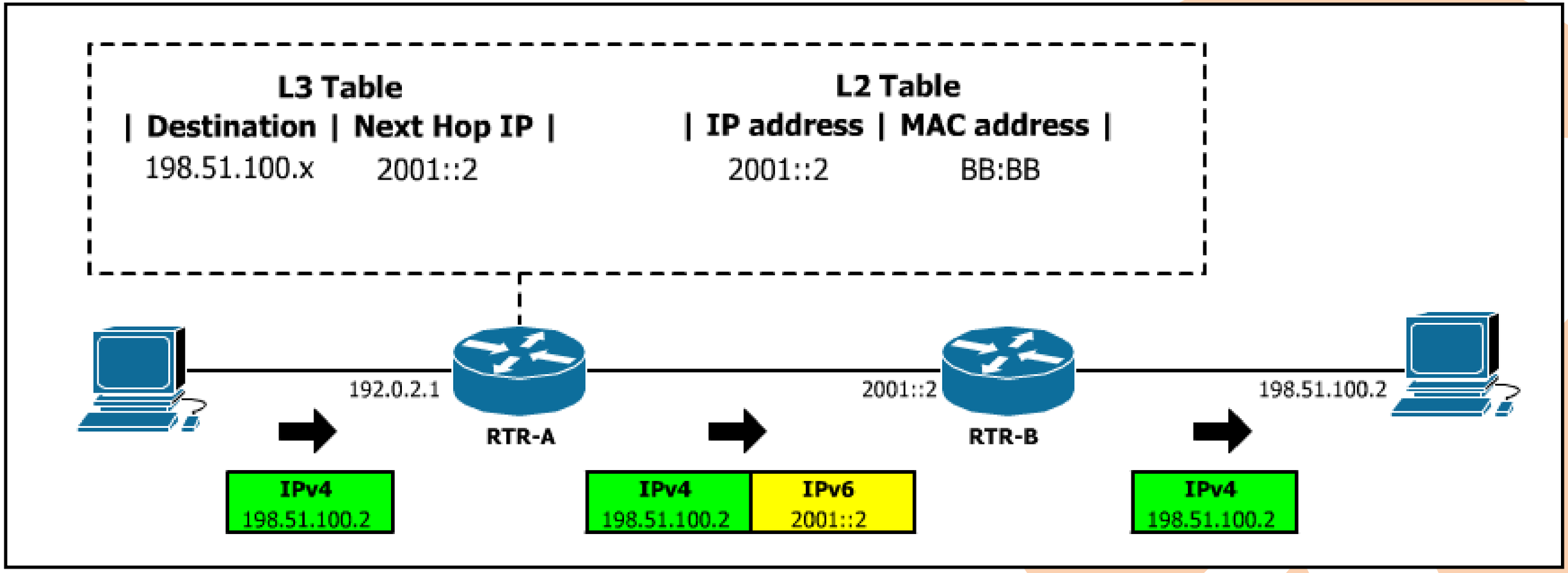
Solution #5 (routing)

- Use an IPv6 address as next hop for an IPv4 route.
  - RFC5549



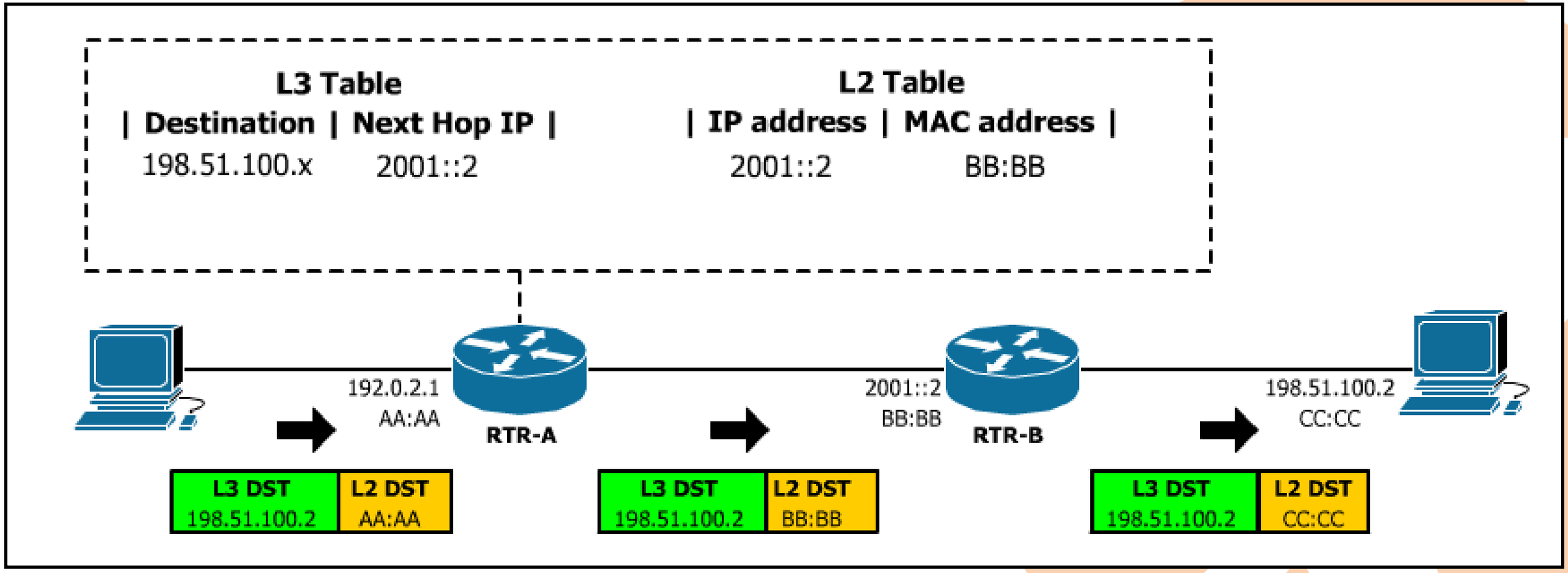
Solution #5 (forwarding)

- '4in6' Softwire mesh tunnel [RFC 5565]
  - Tunnel terminates in the router itself
  - Encapsulate every IPv4 packet (+40B per packet)



Solution #5 (forwarding)

- Direct forwarding
  - Using the L2 (destination MAC address).
  - No overhead.



Solution #5

- Breaks traceroute.

Agenda

- AMS-IX
- Problem statement
- Possible solutions
- RFC5549
- Proof of concept implementation

RFC5549

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

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

RFC5549

- MP_REACH_NLRI attribute:
  - AFI: Address Family ID
    - 1=IPv4, 2=IPv6
  - SAFI: Subsequent AFI
    - 1=Unicast, 2=Multicast
  - LNHA: Length Next Hop Address
  - NANH: Network Address Next Hop
  - NLRI: Network Layer Reachability Information

Pseudo code

```
*Receive an MP_BGP UPDATE message*
IF ((Update AFI == IPv4)
    && (Length of next hop == 16 Bytes
    || 32 Bytes))
    {
      This is an IPv4 route, but
      with an IPv6 next hop;
    }
```

Agenda

- AMS-IX
- Problem statement
- Possible solutions
- RFC5549
- Proof of concept implementation

POC implementation

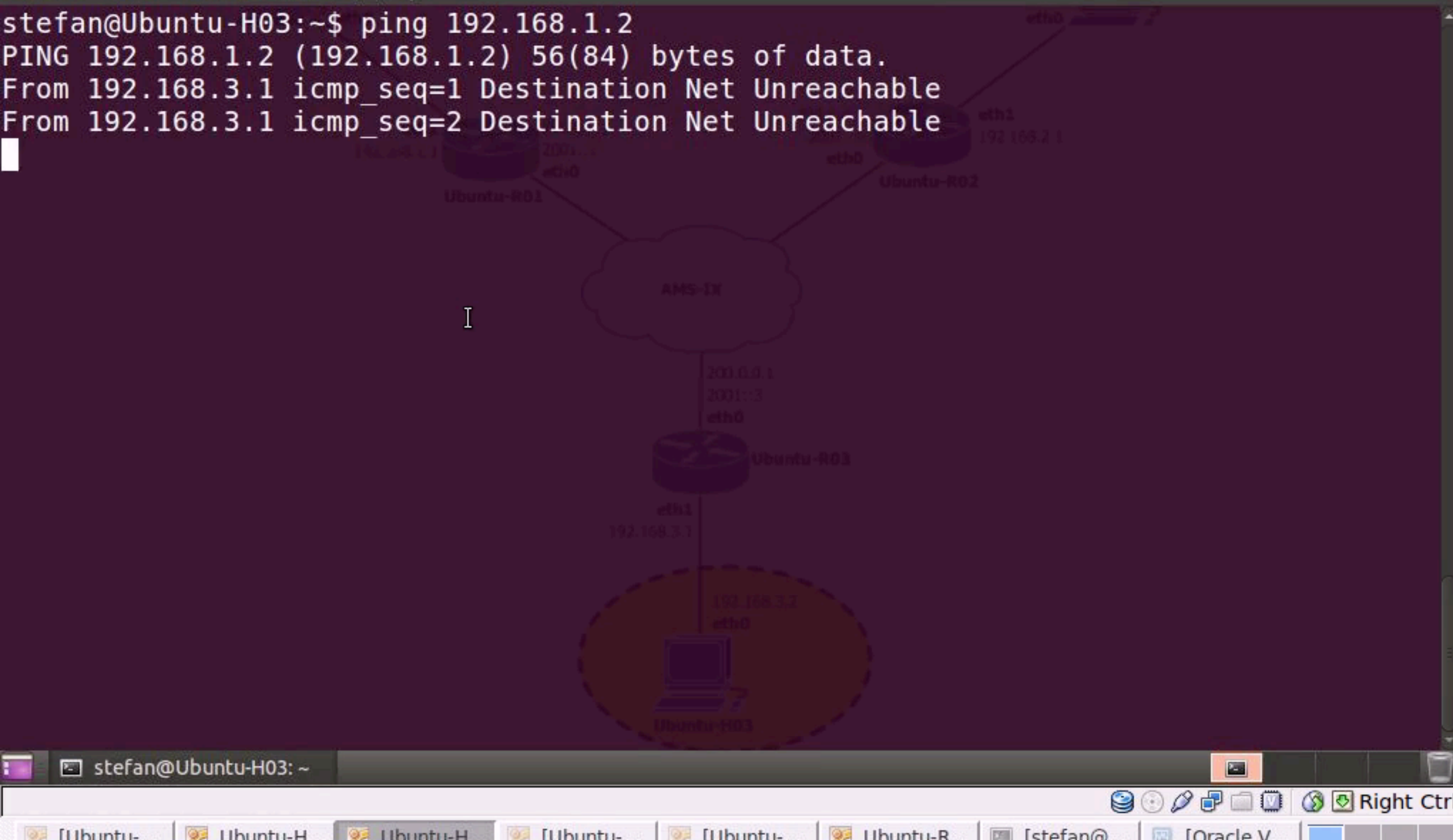
- Based on “direct forwarding” or “native interprotocol forwarding”.
- Based on Quagga v0.99.20
  - Not near production grade code.
  - https://www.ams-ix.net/downloads/RFC5549/

Ubuntu-H03 [Running] - Oracle VM VirtualBox

stefan@Ubuntu-H03: ~

File Edit View Search Terminal Help

```
stefan@Ubuntu-H03:~$ ping 192.168.1.2
PING 192.168.1.2 (192.168.1.2) 56(84) bytes of data.
From 192.168.3.1 icmp_seq=1 Destination Net Unreachable
From 192.168.3.1 icmp_seq=2 Destination Net Unreachable
```



Applications Places Wed May 23, 12:40 AM stefan

Ubuntu-H03 [Running] - Oracle VM VirtualBox

Machine View Devices Help

Applications Places System 6:40 PM stefan

stefan@Ubuntu-H03: ~

File Edit View Search Terminal Help

```
stefan@Ubuntu-H03:~$ ping 192.168.1.2
PING 192.168.1.2 (192.168.1.2) 56(84) bytes of data.
From 192.168.3.1 icmp_seq=1 Destination Net Unreachable
From 192.168.3.1 icmp_seq=2 Destination Net Unreachable
From 192.168.3.1 icmp_seq=3 Destination Net Unreachable
From 192.168.3.1 icmp_seq=4 Destination Net Unreachable
^C
--- 192.168.1.2 ping statistics ---
4 packets transmitted, 0 received, +4 errors, 100% packet loss, time 3000ms
```

```
stefan@Ubuntu-H03:~$ ping 192.168.1.2
PING 192.168.1.2 (192.168.1.2) 56(84) bytes of data.
64 bytes from 192.168.1.2: icmp_req=1 ttl=62 time=5.12 ms
64 bytes from 192.168.1.2: icmp_req=2 ttl=62 time=1.36 ms
64 bytes from 192.168.1.2: icmp_req=3 ttl=62 time=1.34 ms
```

stefan@Ubuntu-H03: ~

Right Ctrl

[Ubuntu-... Ubuntu-H... Ubuntu-H... [Ubuntu-... [Ubuntu-... Ubuntu-R... [stefan@... [Oracle V...

End

Comments & Questions