IPv6: Protocol, Transition Tools and Deployment

Prof. Wing C. Lau

wclau@ie.cuhk.edu.hk

IERG5090

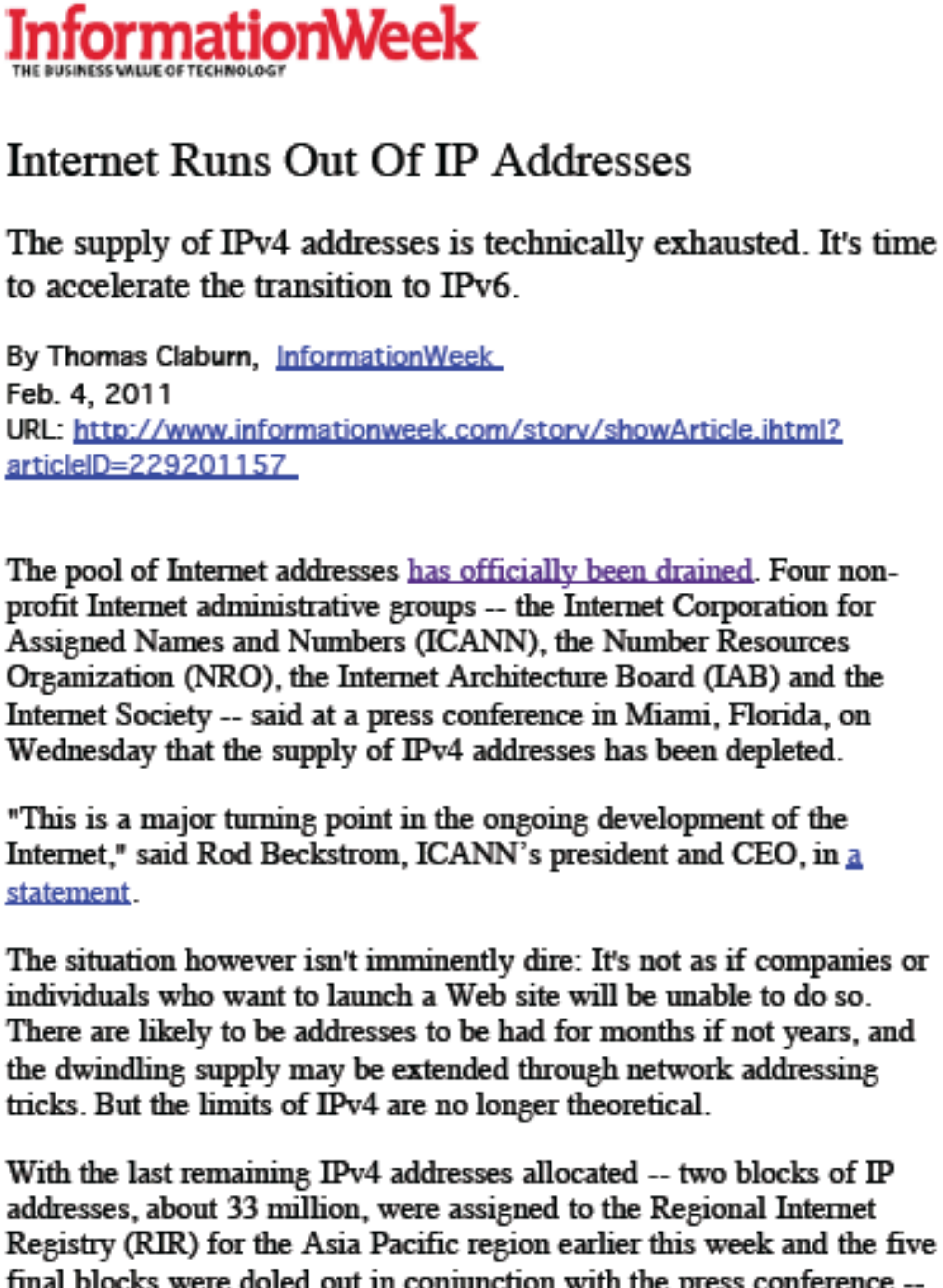
Spring, 2017

Acknowledgements

- Many of the slides used in this lecture are adapted from the following sources:
  - "Computer Networking: A Top-down approach featuring the Internet" by James F. Kurose and Keith W. Ross 4th Edition, all material copyright 1996-2007, J.F Kurose and K.W. Ross, All Rights Reserved.
  - Jeff Doyle, "Issues in IPv6 Deployment", Juniper Networks.
  - Henning Schulzrinne, lecture notes, Columbia University.
  - Paul Francis, "NAT and IPv6: We meet at last", Cornell University
  - Quincy Wu, "Teredo - Tunneling IPv6 through NATs", National Chiao Tung University
  - Geoff Huston, "IPv4 Unallocated Address Space Exhaustion," APNIC
  - Cecil Goldstein, "Depletion of IPv4 free address pool - IPv6 deployment, The Day After", APNIC
  - Japan IPv6 Promotional Council, Jordi Palet Martinez of Consulintel, Merike Kaeo of Double Shot Security, Philip Smith of Cisco, Randy Bush (IIJ), Paul Wilson (APNIC), "Internet in Crisis", APNIC Seminar
- The instructor acknowledges with thanks and appreciation the contribution of the original authors. All copyrights belong to the authors of the original material.

Outline

- Motivation
- Understand IPv6 basics
  - Header format
  - Addressing
  - Autoconfiguration
- IPv6 Transitioning Strategies and Tools
- IPv6 deployment status



InformationWeek
THE BUSINESS VALUE OF TECHNOLOGY

Internet Runs Out Of IP Addresses

The supply of IPv4 addresses is technically exhausted. It's time to accelerate the transition to IPv6.

By Thomas Claburn, InformationWeek
Feb. 4, 2011
URL: http://www.informationweek.com/story/showArticle.jhtml?articleID=229201157

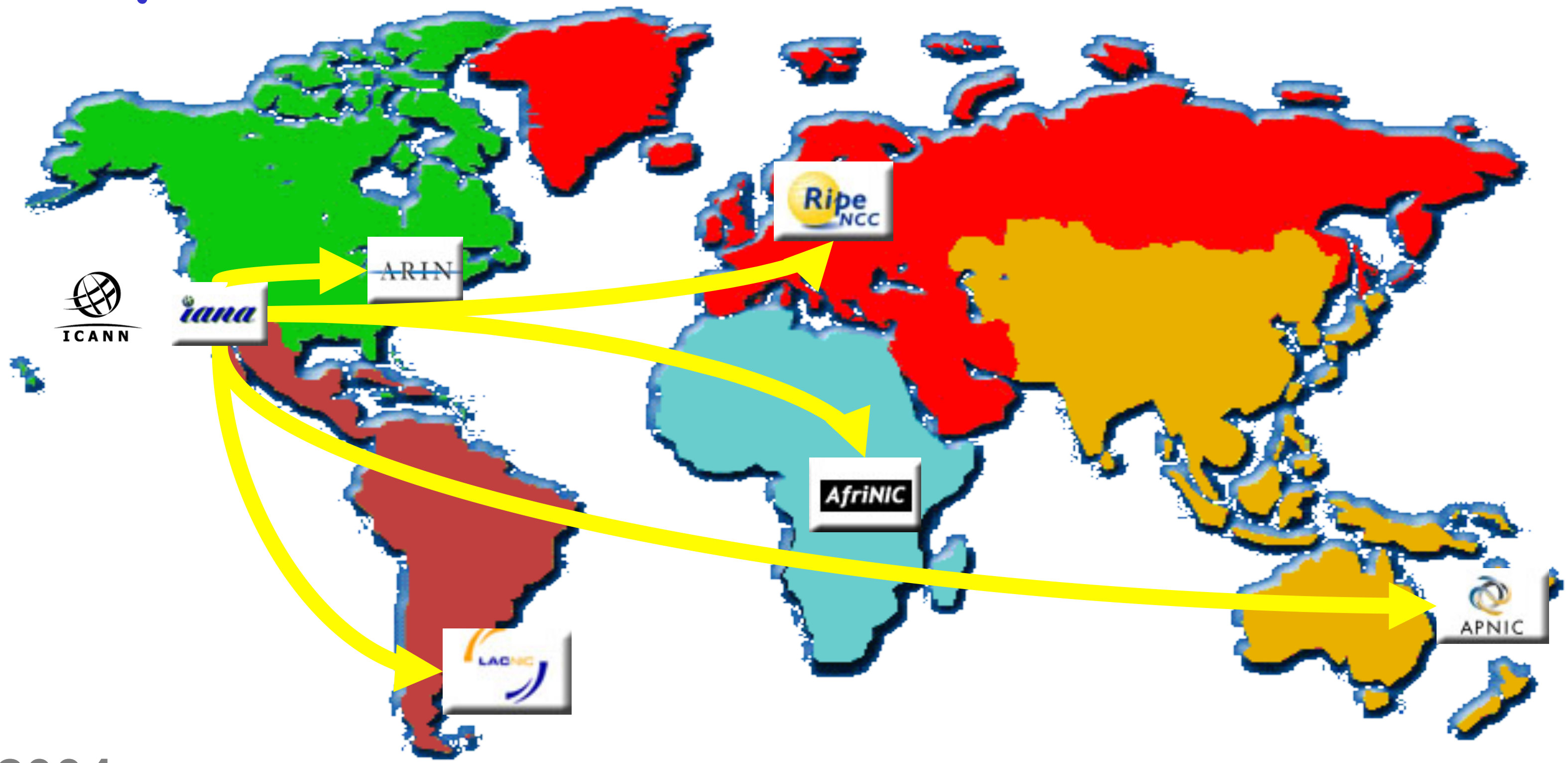
The pool of Internet addresses has officially been drained. Four non-profit Internet administrative groups -- the Internet Corporation for Assigned Names and Numbers (ICANN), the Number Resources Organization (NRO), the Internet Architecture Board (IAB) and the Internet Society -- said at a press conference in Miami, Florida, on Wednesday that the supply of IPv4 addresses has been depleted.

"This is a major turning point in the ongoing development of the Internet," said Rod Beckstrom, ICANN's president and CEO, in a statement.

The situation however isn't imminently dire: It's not as if companies or individuals who want to launch a Web site will be unable to do so. There are likely to be addresses to be had for months if not years, and the dwindling supply may be extended through network addressing tricks. But the limits of IPv4 are no longer theoretical.

With the last remaining IPv4 addresses allocated -- two blocks of IP addresses, about 33 million, were assigned to the Regional Internet Registry (RIR) for the Asia Pacific region earlier this week and the five final blocks were doled out in conjunction with the press conference --

The address infrastructure today



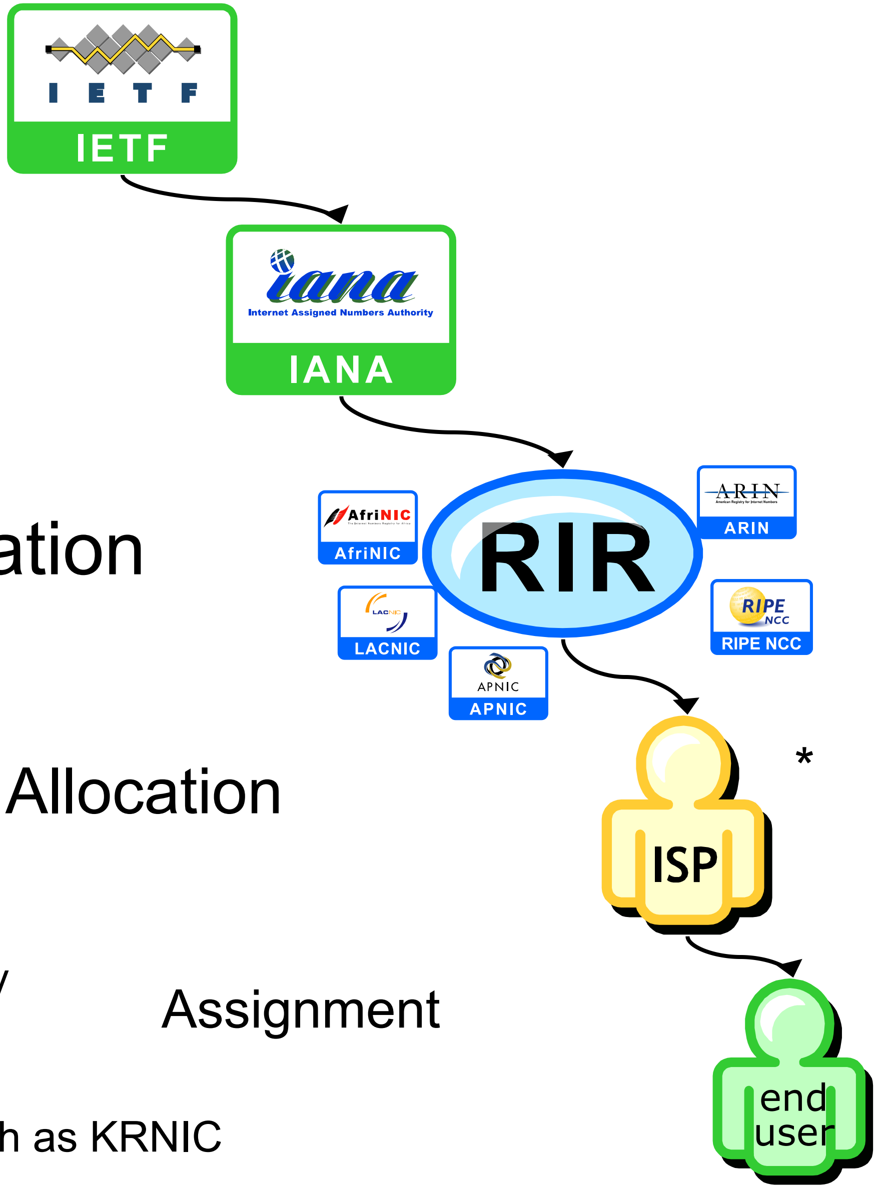
2004:

Number Resource Organization

NRO

IP Addresses - what makes the internet The Internet

IPv4 \IPv6



IETF

IANA

Internet Assigned Numbers Authority

RIR

AfriNIC

ARIN

LACNIC

APNIC

RIPE NCC

ISP

*

end user

Allocation

Allocation

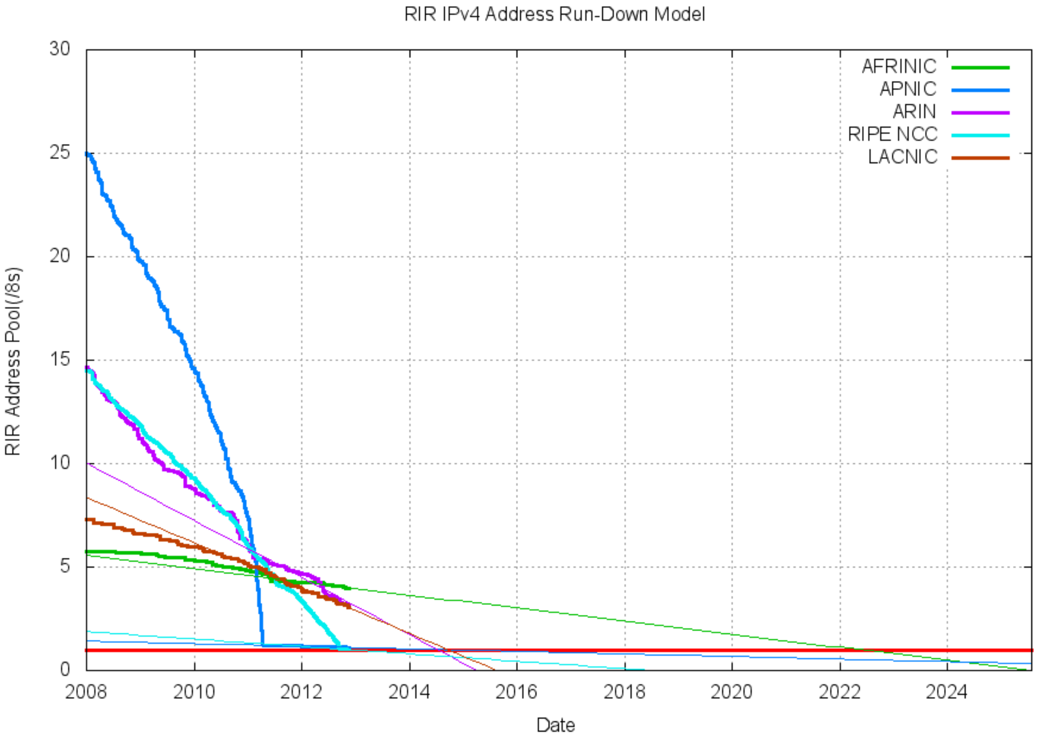
Assignment

RIR = Regional Internet Registry

NIC = Network Info. Center

* In some cases via an NIR such as KRNIC

IPv4 Address Exhaustion Date



IPv4 Address Report

Projected RIR Address Pool Exhaustion Dates:

RIR	Projected Exhaustion Date	Remaining Addresses in RIR Pool (/8s)
APNIC:	**19-Apr-2011** (actual)	0.9023
RIPE NCC:	**14-Sep-2012** (actual)	0.9632
ARIN:	**16-Sep-2013**	3.1179
LACNIC:	**15-Jun-2015**	3.0039
AFRINIC:	**02-Sep-2019**	3.9563

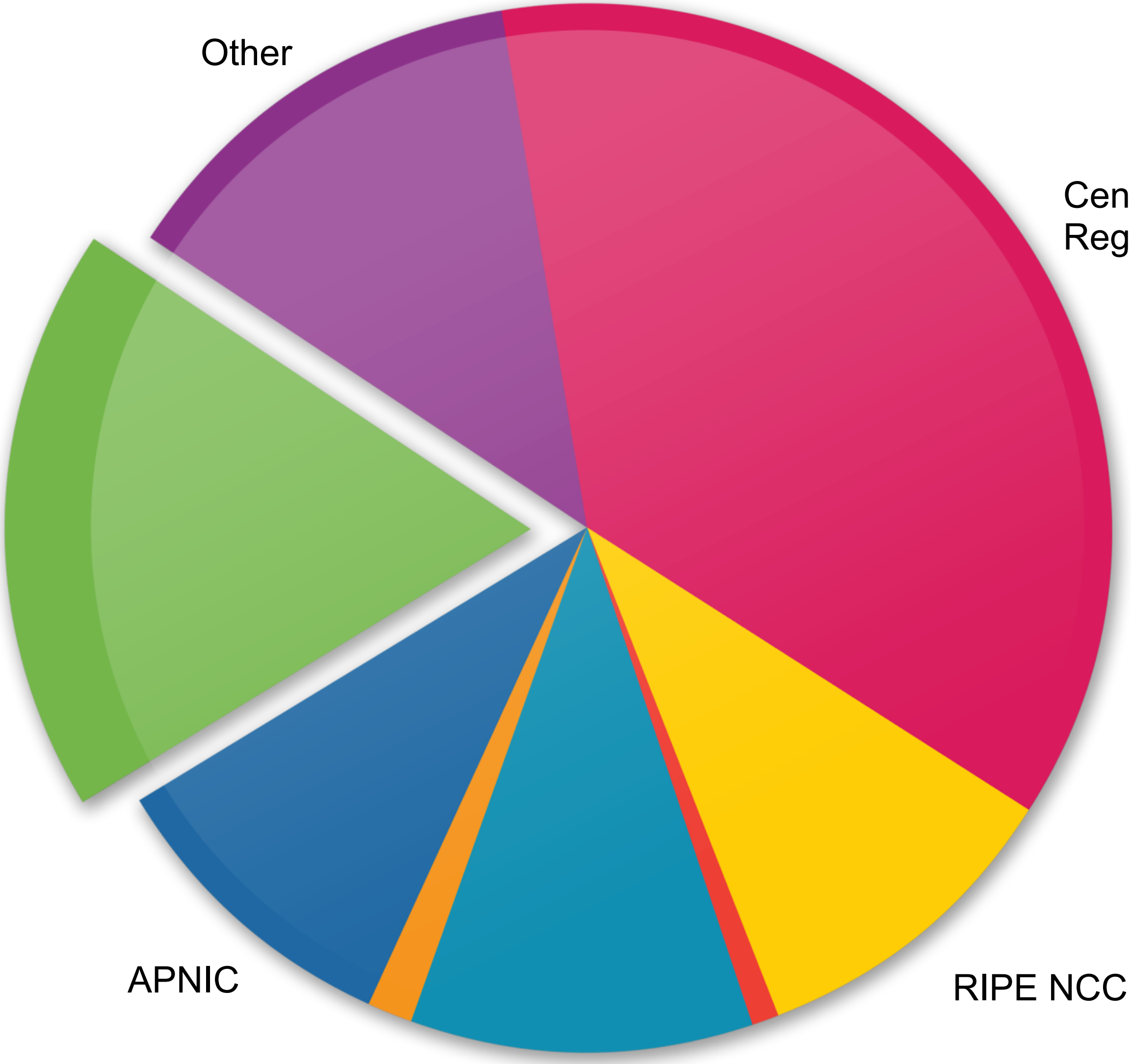
IANA Unallocated Address Pool Exhaustion: 3-Feb-2011

http://www.potaroo.net/tools/ipv4

IPv4 IANA address pool - Mar, 2008

16%

available



source: http://potaroo.net

So what will happen after the exhaustion?

- The Internet will not stop but its growth will be impacted
- Who will be impacted?
  - ISPs
    - Sustaining their business models will become more difficult unless you have huge IPv4 address blocks
  - End users
    - Cost of access to the Internet will increase

Some possible scenarios

- What will happen after the IPv4 unallocated address space exhaustion?
  - Persist in IPv4 networks using more NATs
  - Address markets emerging for IPv4
    - In March 2011, Microsoft paid US$7.5M for 666K IPv4 addresses from the bankrupted Nortel
  - Routing fragmentation
  - IPv6 deployment/ transition

Ref: IPv4 unallocated address space exhaustion by Geoff Huston, Sept 2007

How can we cope with it?

CIDR

- Classless Inter Domain Routing
  - Developed to cope with a rapid IPv4 address consumption (around 1994 - 1995 period)
    - Before CIDR, people used Classfull address architecture
      - Class A, B and C
      - A very inflexible architecture
  - CIDR allows to assign IP addresses in a much more flexible manner
    - Classless address architecture
    - CIDR allows us to extend the IPv4 address space more than we expected

IPv4 Network Address Translators (NATs) today

- Today NATs are largely externalised costs for ISPs
  - Customers buy and operate NATs
  - Applications are tuned to single-level-NAT traversal
  - Static public addresses typically attract a traffic premium in the real market
    - For retail customers, IP addresses already have a market price!

Ref: IPv4 unallocated address space exhaustion by Geoff Huston, Sept 2007

The “Just” add more NATs option

- Demand for increasing NAT “intensity”
  - Shift ISP infrastructure to private address realms
  - Multi-level NAT deployment both at the customer edge and within the ISP network
    - This poses issues in terms of application discovery and adaptation to NAT behaviours
- Some NAT Traversal solutions already exist:
  - STUN:(Simple Traversal of UDP through NAT) RFC3489
    - Aka the “Classic STUN”
  - STUN+TURN+ICE
    - Session Traversal Utilities for NAT (STUN) RFCs 5389,
    - Traversal Using Relay around NATs (TURN)
    - Interactive Connectivity Establishment (ICE)
  - STUNT: STUN to support TCP traversal via NATs

    => ICE-TCP
- Defining new requirement for new NAT boxes
  - IETF BEHAVE Working Group

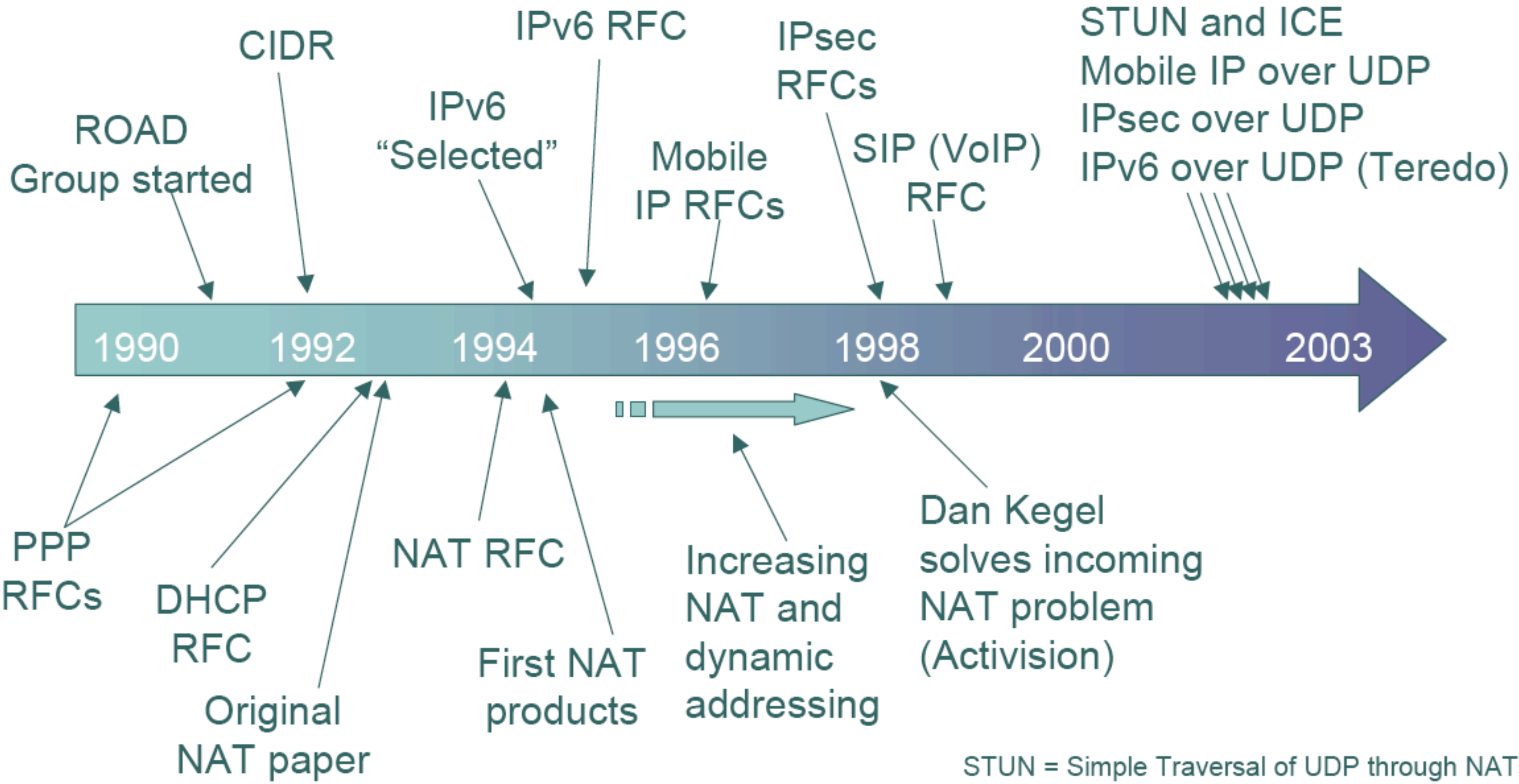
Ref: IPv4 unallocated address space exhaustion by Geoff Huston, Sept 2007

The "Just" add more NATs option (cont'd)

- How far can NATs scale?
  - Not well known
  - What are the critical resources here
    - NAT biding capability and state maintenance, NAT packet throughput, private address pool sizes and application complexity
- End cost for static public addresses may increase

Ref: IPv4 unallocated address space exhaustion by Geoff Huston, Sept 2007

Timeline for IPv6 vs. Private IPv4 +NAT traversal



ROAD Group started

CIDR

IPv6 “Selected”

IPv6 RFC

Mobile IP RFCs

IPsec RFCs

SIP (VoIP) RFC

STUN and ICE
Mobile IP over UDP
IPsec over UDP
IPv6 over UDP (Teredo)

1990 1992 1994 1996 1998 2000 2003

PPP RFCs

DHCP RFC

Original NAT paper

NAT RFC

First NAT products

Increasing NAT and dynamic addressing

Dan Kegel solves incoming NAT problem (Activision)

STUN = Simple Traversal of UDP through NATs
ICE = Interactive Connectivity Establishment

Recovering unused IPv4 address space

- 46 x /8 (in various prefixes) un-routed address spaces existing
  - APNIC and LACNIC have active reclamation processes
  - However, recovery of such address space is not easy
    - Most of historical address space exist in USA
    - Historical address space: address distributed before the RIR mechanism kicked into the system
    - Reclamation processes are not only likely to be lengthy and difficult, but also expensive
    - Most likely "address market" will emerge
  - Amount of recoverable address space is relatively insignificant
  - Fragmented address blocks
    - Increase injection to the global routing table
- Only provides limited solutions

Ref: APster Issues 23 – September 2007, "Responses to IPv4 address space consumption" By Paul Wilson

Transition to IPv6

- But IPv6 is not backward compatible with IPv4 on the wire

Ref: IPv4 unallocated address space exhaustion by Geoff Huston, Sept 2007

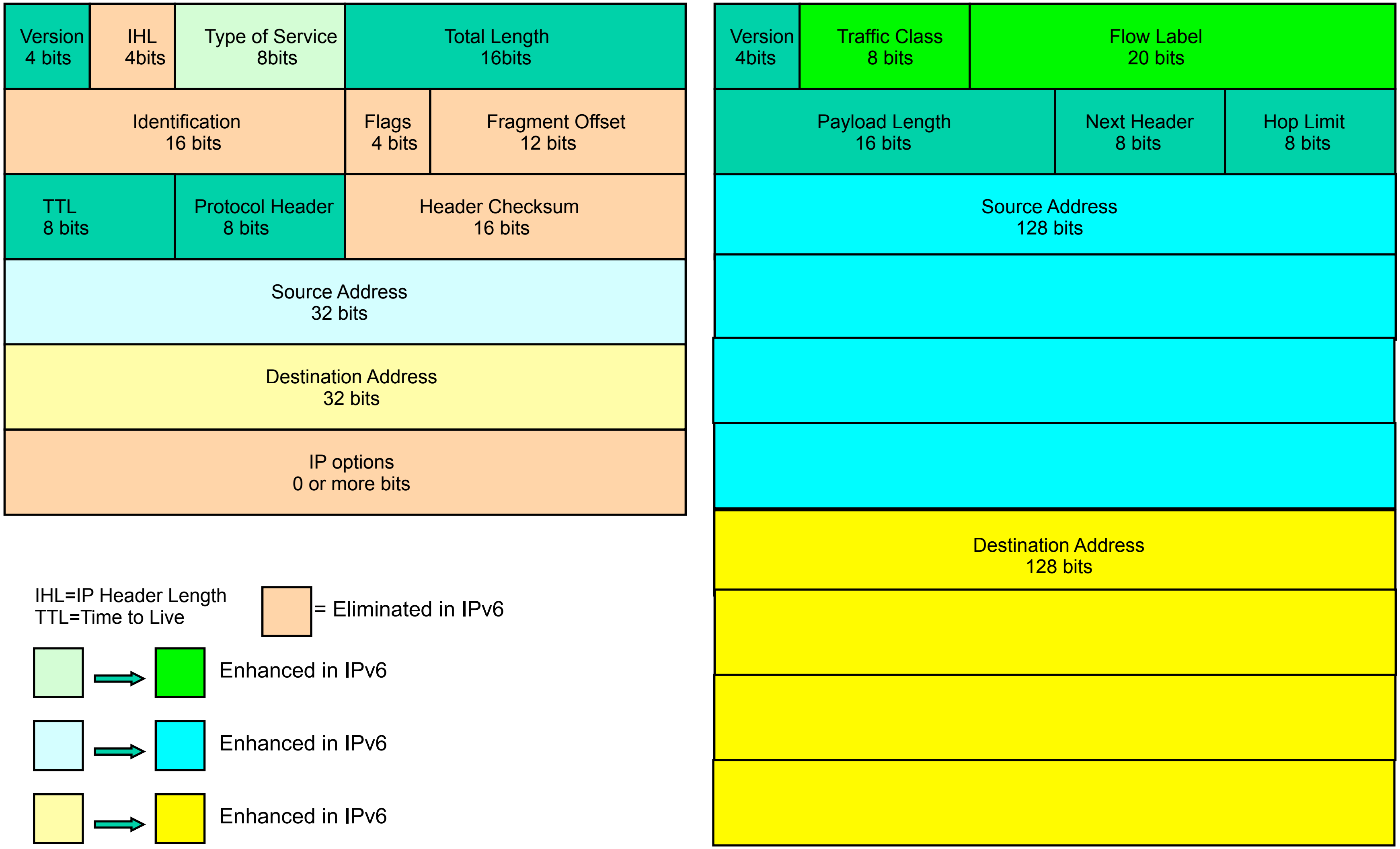
So, what is IPv6?

IPv6 header

- Comparison between IPv4 header and IPv6 header

IPv4 Header

IPv6 Header



IPv4 Header

Version 4 bits

IHL 4bits

Type of Service 8bits

Total Length 16bits

Identification 16 bits

Flags 4 bits

Fragment Offset 12 bits

TTL 8 bits

Protocol Header 8 bits

Header Checksum 16 bits

Source Address 32 bits

Destination Address 32 bits

IP options 0 or more bits

IPv6 Header

Version 4bits

Traffic Class 8 bits

Flow Label 20 bits

Payload Length 16 bits

Next Header 8 bits

Hop Limit 8 bits

Source Address 128 bits

Destination Address 128 bits

IHL=IP Header Length
TTL=Time to Live

= Eliminated in IPv6

Enhanced in IPv6

Enhanced in IPv6

Enhanced in IPv6

IPv6 Key Features

- Increased address space
  - 128 bits = 340 trillion trillion trillion addresses
  - (2^{128} = 340,282,366,920,938,463,463,374,607,431,768,211,456)
  - = 67 billion addresses per cm^2 of the planet surface
- More efficient header architecture
  - Eliminate IP Header checksum
  - No packet fragmentation allowed => Required e2e MTU discovery instead

=> More efficient per-packet processing/forwarding at router

- Provision for Protocol Extension
  - fixed-length 40 byte header for common case ;
  - Replace rigid IP-option fields with “Next Header” extensions

  => makes the protocol more adaptable
- Neighbor Discovery and Autoconfiguration
  - No more ARP for IPv6
  - Improved operational efficiency
  - Easier network changes and renumbering
  - Simpler network applications (Mobile IPv6)

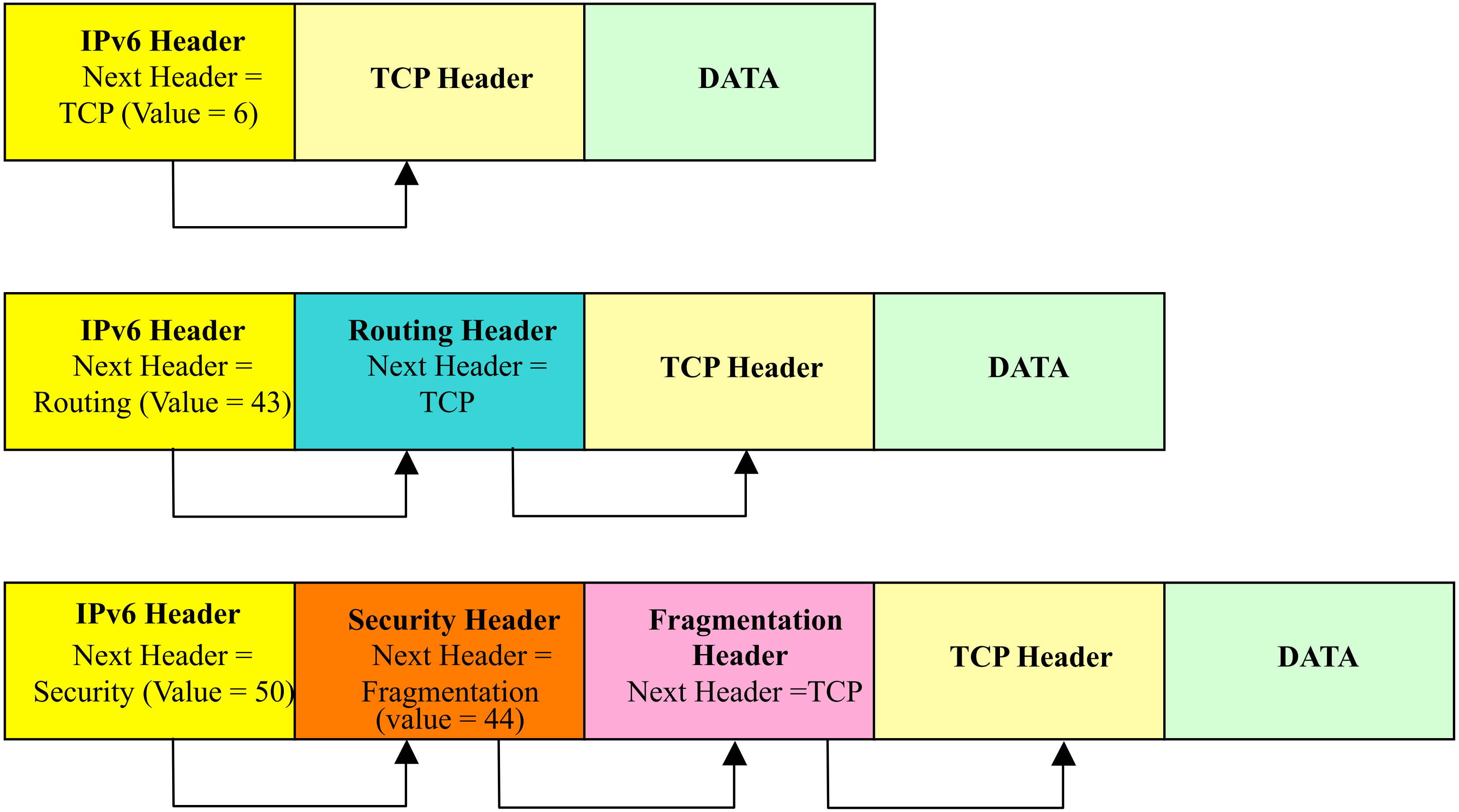
  Some of these features are supported via a new version ICMP aka ICMPv6
- Header changes to facilitate QoS (via Flow Labels)
- Mandatory and Integrated security features (via IPsec support)

IPv6 header

- IPv6 header is considerably simpler than IPv4
  - IPv4: 12 fields + options , IPv6: 8 fields + options
- IPv4 header less flexible – cannot exceed 60 bytes
- Eliminated fields in IPv6
    - Header Length
    - Identification
    - Flag
    - Fragmentation Offset
    - Checksum
- Enhanced fields in IPv6
    - TOS =>Traffic Class
    - Time to Live => Hop Limit
    - Protocol => Next header (extension headers)
    - New Flow Label
- Authentication and privacy capabilities

Extension headers

- Next header field



IPv6 Header
Next Header =
TCP (Value = 6)

TCP Header

DATA

IPv6 Header
Next Header =
Routing (Value = 43)

Routing Header
Next Header =
TCP

TCP Header

DATA

IPv6 Header
Next Header =
Security (Value = 50)

Security Header
Next Header =
Fragmentation
(value = 44)

Fragmentation Header
Next Header =TCP

TCP Header

DATA

Flow Support in IPv6

- **A flow** is a sequence of related packets sent from a source to a unicast, anycast, or multicast destination
  - E.g. transport connections, media streams, but not necessarily 1:1
- **Flow labeling** with the **Flow Label field** enables classification of packets belonging to a specific flow
  - Without the flow label the classifier must use transport next header value and port numbers
    - Less efficient (need to parse the option headers)
    - May be impossible (due to fragmentation or IPsec ESP)
    - Layer violation may hinder introduction of new transport protocols
- **Flow state** is established in a subset or all of the IP nodes on the path
  - Includes the **flow classifier**
  - Defines the **flow-specific treatment** the packets should receive
  - Can be signaled, or configured (administratively or manually)
  - Can also be determined algorithmically in some cases (e.g. load spreading)

Jarno Rajahalme et al, IETF 53

IPv6 addressing

IPv6 addressing

- 128 bits of address space
- Hexadecimal values of eight 16 bit fields
    - X:X:X:X:X:X:X:X (X=16 bit number, ex: A2FE)
    - 16 bit number is converted to a 4 digit hexadecimal number
- Example:
    - 2001:DB8:124C:C1A2:BA03:6735:EF1C:683D
  - Abbreviated form of address
    - 2001:DB8:0023:0000:0000:036E:1250:2B00

    →2001:DB8:23:0:0:36E:1250:2B00

    →2001:DB8:23::36E:1250:2B00

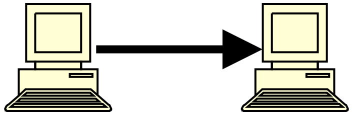
    (Null value can be used only once)

IPv6 addressing model

- **IPv6 Address type**

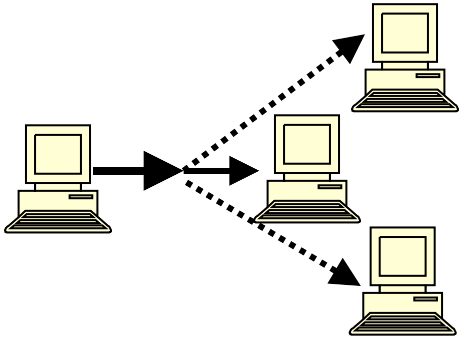
RFC 4291

- Unicast
  - An identifier for a single interface

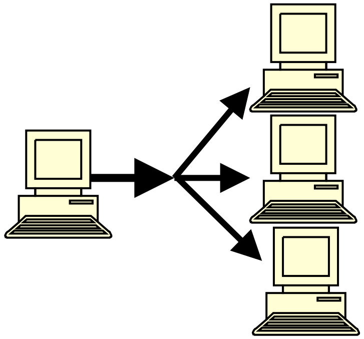


- Anycast
  - An identifier for a set of interfaces

  e.g. the nearest video server



- Multicast
  - An identifier for a group of nodes



Unicast address

- Address given to interface for communication between host and router
  - Global unicast address currently delegated by IANA

001 FP 3bits	Global routing prefix 45 bits	Subnet ID 16 bits	Interface ID 64 bits

FP = Format Prefix

  - Local use unicast address
    - Link-local address (starting with FE80::)

1111111010 10 bits	000…….0000 54 bits	Interface ID 64 bits

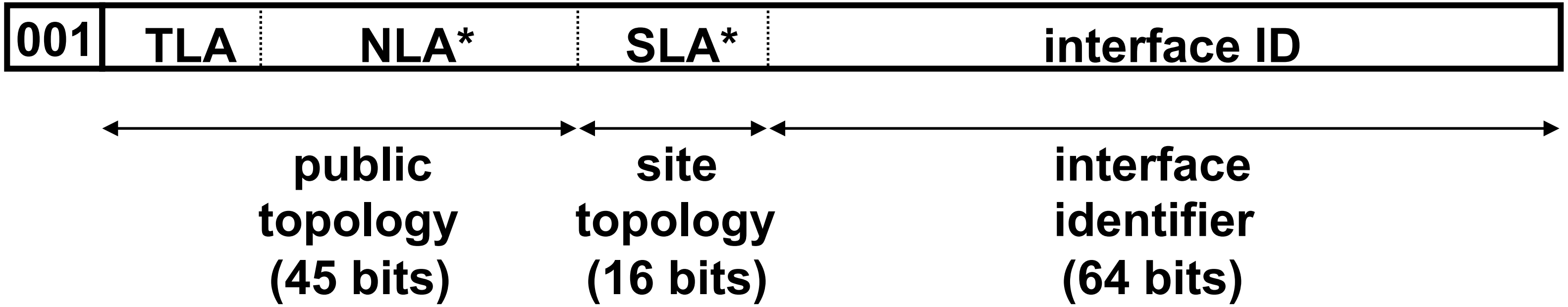
    - Site-local address (starting with FEC0::)

~~1111111011~~ ~~10 bits~~	~~Subnet-ID~~ ~~54 bits~~	~~Interface ID~~ ~~64 bits~~

Aggregatable global unicast address - deprecated

~~RFC 2374~~

- RFC 2374 – deprecated



001	TLA	NLA*	SLA*	interface ID

public topology (45 bits)

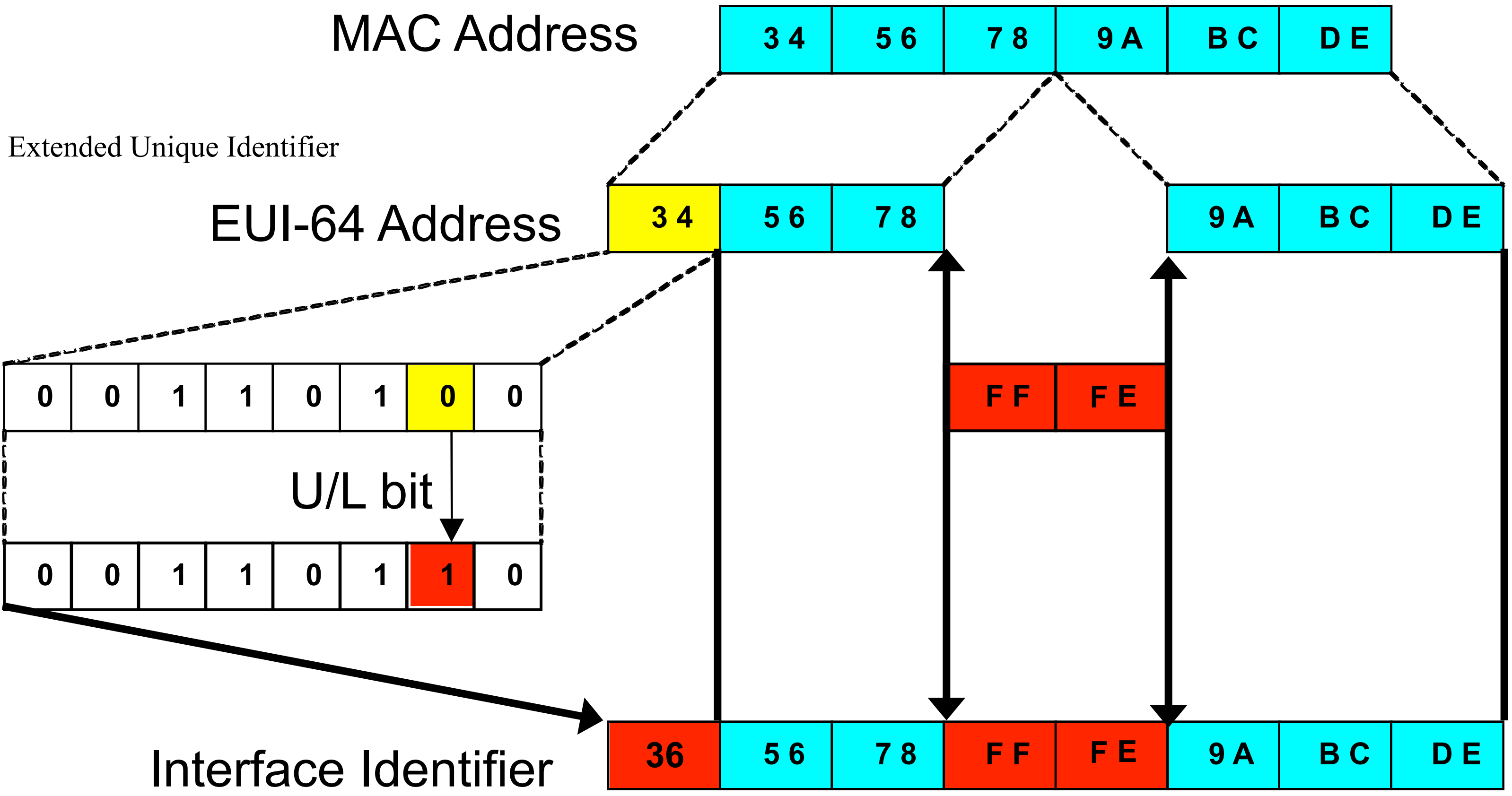
site topology (16 bits)

interface identifier (64 bits)

- TLA = Top-Level Aggregator
  NLA = Next-Level Aggregator(s)
  SLA = Site-Level Aggregator(s)
- This scheme has been replaced by a coordinated allocation policy defined by RIR.
- You may see them in text books, but remember they are deprecated!

Interface ID: EUI-64

EUI = Extended Unique Identifier

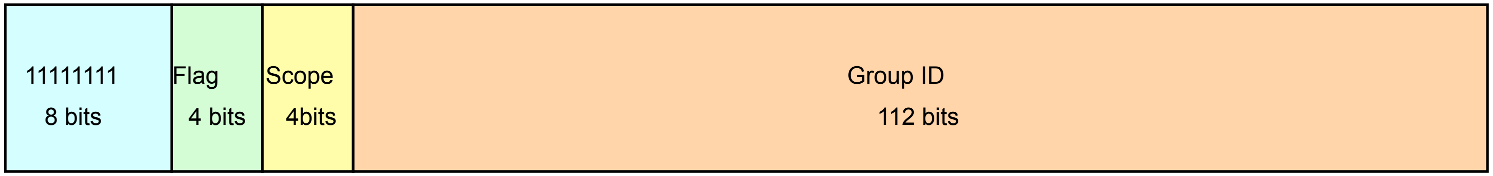


U/L bit = 0 if non-unique MAC address (A MAC address may be not unique if the administrator changes the MAC address of the Interface.)
U/L bit = 1 if unique MAC address

Anycast

- IPv4 has unicast, broadcast, multicast
- IPv6 has no broadcast (uses multicast)
- But there's the new ***anycast***
- Sends a packet to the "best" destination amongst a set of hosts
    - Delivery to a single interface
    - e.g. "send this packet to the nearest router that has a connection to the Internet"
- Cannot be used as the source address
- Need more widespread experience in the future

Multicast address



11111111 8 bits	Flag 4 bits	Scope 4bits	Group ID 112 bits

- ❑ First 8 bits identifies multicast address
  - ❖ 11111111 (FF)
- ❑ Flags
  - ❖ 0000 = a permanently-assigned (well-known) multicast address
  - ❖ 0001 = a non-permanently-assigned (transient) multicast address
- ❑ Scope (indicates the scope of the multicast group)
  - ❖ 1= node local
  - ❖ 2= link local
  - ❖ 3= site local
  - ❖ 8= organization local
  - ❖ E= global
- ❑ Group ID
  - ❖ Identifies the multicast group within the specified scope
- ❑ Well-known multicast addresses
  - ❖ FF02:0:0:0:0:0:0:1 All-nodes address with Link-local scope
  - ❖ FF02:0:0:0:0:0:0:2 All-routers address with Link-local scope

Autoconfiguration (of IPv6 addr)

IPv6 autoconfiguration

RFC 2462

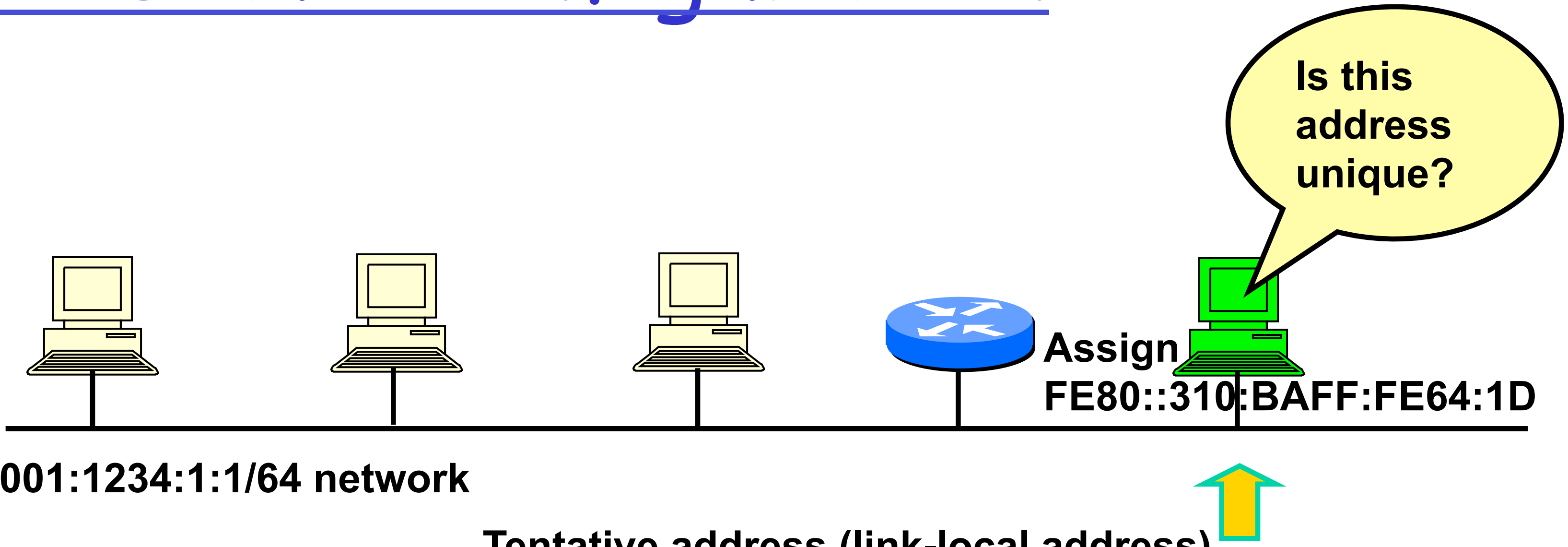
How an IPv6 host get its address(es) ?

- Stateless mechanism
  - For a site not concerned with the exact addresses
  - No manual configuration required
  - Minimal configuration of routers
  - No additional servers
- Stateful mechanism
  - For a site requires tighter control over exact address assignments
  - Need DHCP server
  - DHCPv6
- Enable "Plug and play"

Plug and Play

- IPv6 link local address
  - Even if no servers/routers exists to assign an IP address to a device, the device can still auto-generate an IP address
    - Allow interfaces on the same link to communicate with other devices
- Stateless
  - No control/record of which interface associated with an assigned IP address
    - Possible security issues
- Stateful
  - Remember information about interfaces that are assigned IP addresses

IPv6 autoconfiguration



Is this address unique?

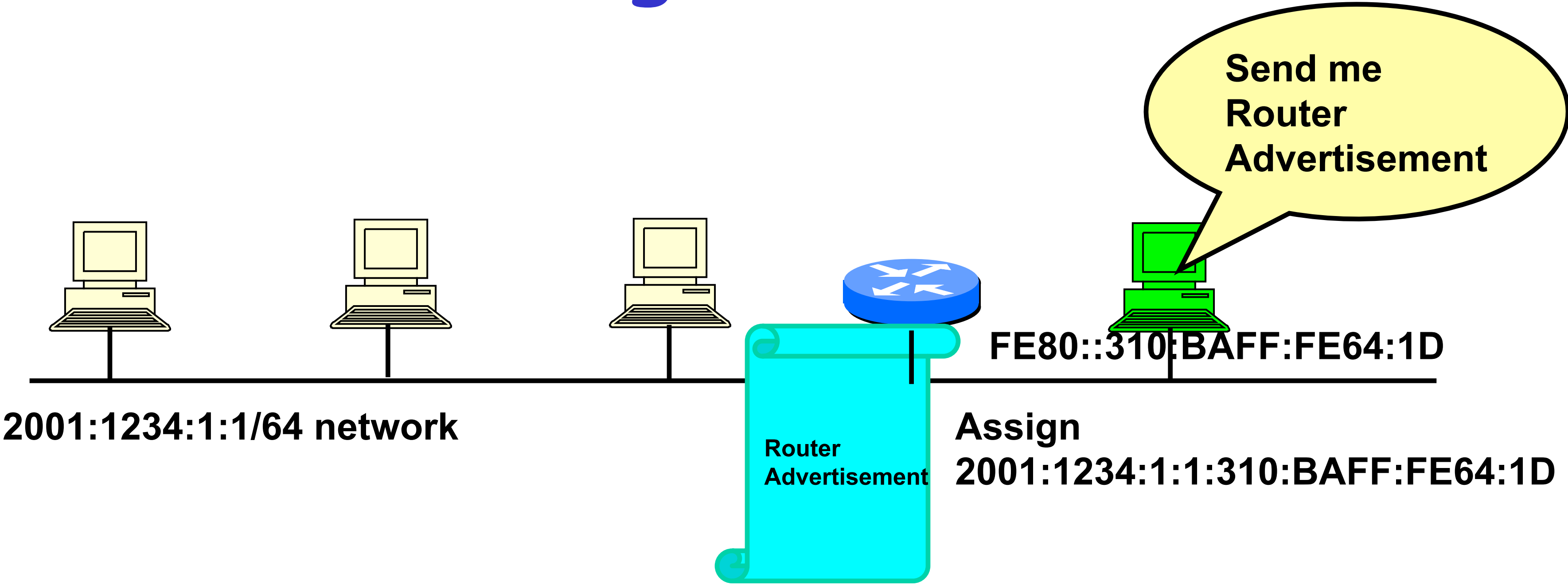
Assign FE80::310:BAFF:FE64:1D

2001:1234:1:1/64 network

Tentative address (link-local address)
Well-known link local prefix +Interface ID (EUI-64)
Ex: FE80::310:BAFF:FE64:1D

1. A new host is turned on.
2. Tentative address will be assigned to the new host.
3. Duplicate Address Detection (DAD) is performed. First the host transmit a Neighbor Solicitation (NS) message to all-nodes multicast address (FF02::1)
5. If no Neighbor Advertisement (NA) message comes back then the address is unique.
6. FE80::310:BAFF:FE64:1D will be assigned to the new host.

IPv6 autoconfiguration



1. The new host will send Router Solicitation (RS) request to the all-routers multicast group (FE02::2).
2. The router will reply Routing Advertisement (RA).
3. The new host will learn the network prefix. E.g, 2001:1234:1:1/64
4. The new host will assigned a new address Network prefix+Interface ID E.g, 2001:1234:1:1:310:BAFF:FE64:1D

IPv6 autoconfiguration (cont'd)

- Keeps end user costs down
  - No need for manual configuration
  - In conjunction with the possibility of low cost network interface
- Helpful when residential networks emerge as an important market
- But the address not automatically registered into the DNS
- Security issues need to be considered, e.g.
  - How to prevent Rogue Router sending faked RAs ?
    - IETF SEcure Neighbor Discovery (SEND) protocol (RFC3971)
  - How to enforce address assignment policy to hosts ?

IPv4 to IPv6 transition

Key Reference:

http://www.6net.org/book/deployment-guide.pdf

Transition overview

- Not all routers can be upgraded simultaneous
  - no "flag days"
  - How will the network operate with mixed IPv4 and IPv6 routers?
- How to get connectivity from an IPv6 host to the global IPv6 Internet?
  - Via an native connectivity
  - Via an IPv6-in-IPv4 tunnelling techniques
- IPv6-only deployment are rare
- Practical reality
  - Sites deploying IPv6 will not transit to IPv6-only, but transit to a state where they support both IPv4 and Ipv6 (dual-stack)

http://www.6net.org/book/deployment-

Types of Transition Mechanisms

- Dual Stacks
  - IPv4/IPv6 coexistence in a device (host/router) & network
- Tunnels
  - For tunneling IPv6 across IPv4 clouds
  - Later, for tunneling IPv4 across IPv6 clouds
- Build a new IPv6 network and use v6<->v4 Protocol Translators
  - Translators enable IPv6 only nodes to communicate with IPv4 only nodes

Different demands of hosts and networks to be connected to IPv6 networks will determine the best way of transition

Transition Strategies

- Edge-to-core
  - When services are important
  - When addresses are scarce
  - User (customer) driven
- Core-to-edge
  - Good ISP strategy
- By routing protocol area
  - When areas are small enough
- By subnet
  - Probably too incremental

Transition overview

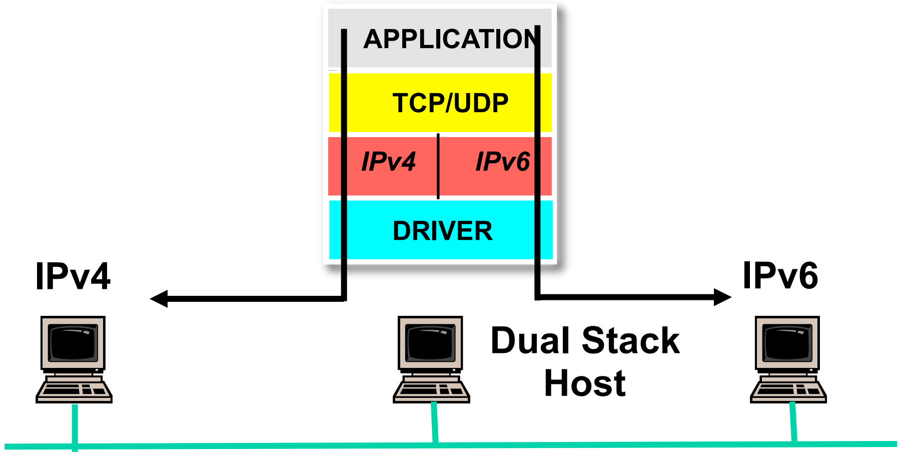
- Once the internal networking is determined,
- The next step is to arrange external connectivity for the whole site
  - Involves external routing issues
  - Either natively or via some tunnelling mechanism

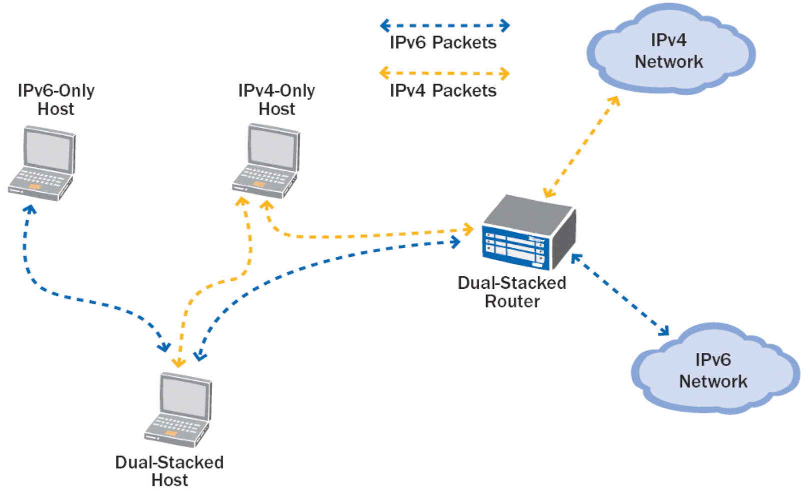
http://www.6net.org/book/deployment-

Dual stack transition

RFC 4213

- Dual stack = TCP/IP protocol stack running both IPv4 and IPv6 protocol stacks simultaneously
  - Application can talk to both
  - Network, Transport, and Application layers do not necessarily interact without further modification or translation
- Useful at the early phase of transition





Dual stack

- A host or a router is equipped with both IPv4 and IPv6 protocol stacks in the OS
- Each node (an IPv4/IPv6 node) is configured with both IPv4 and IPv6 addresses
- Therefore it can both send and receive datagrams belonging to both protocols
- The simplest and the most desirable way for IPv4 and IPv6 to coexist
- BUT Network, Transport, and Application layers do not necessarily interact without further modification or translation

http://www.6net.org/book/deployment-

Dual stack

- Some challenges, e.g.
  - If you use OSPFv2 for your IPv4 network you need to run OSPFv3 in addition to OSPFv2

  => or use IS-IS instead (which runs over the link-layer directly and can support both IPv4/v6 addresses)
  - How to manage the interaction of the two protocols ? e.g.
    - How failover is handled between IPv4 vs. IPv6 ?

http://www.6net.org/book/deployment-

Dual stack

- DNS is used with both protocol versions to resolve names and IP addresses
  - An dual stack node needs a DNS resolver that is capable of resolving both types of DNS address records
    - DNS A record to resolve IPv4 addresses
    - DNS AAAA record to resolve IPv6 addresses
      - DNS A6 record as an alternative, but Experimental only (RFC3363)
      - AAAA record is preferred for production environment.
      - AAAA vs A6 records comparisons in RFC3364
- Dual stack network
  - Is an infrastructure in which both IPv4 and Ipv6 forwarding is enabled on routers ?

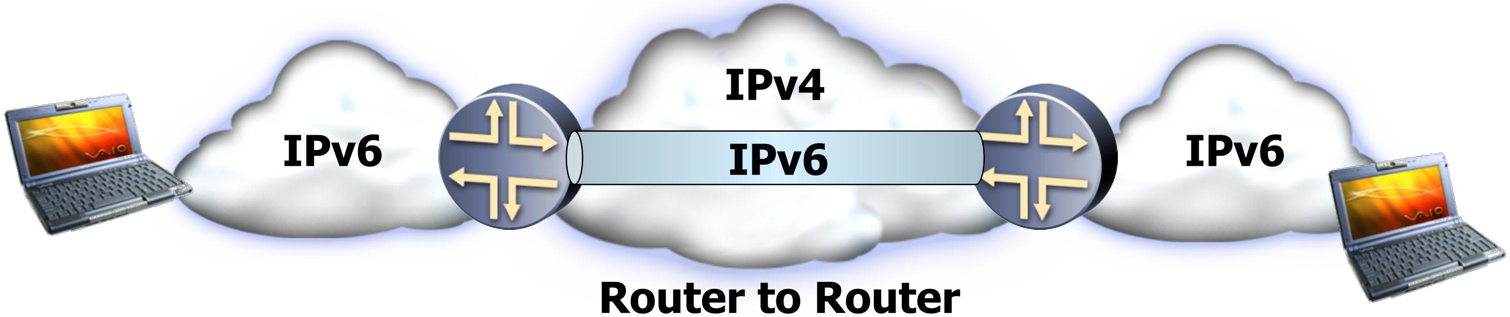
IPv6 essentials by Silvia Hagen, p256

Tunnels

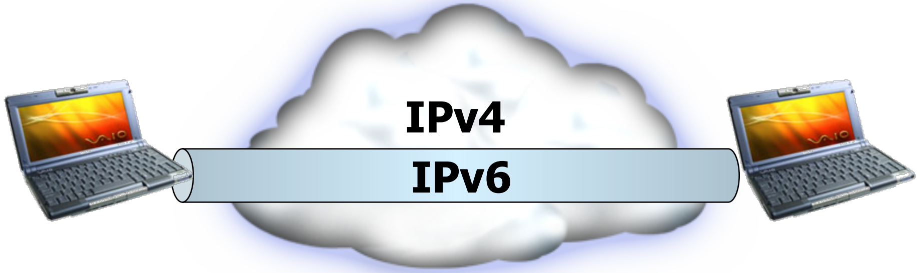
- ❑ Additional IPv6 infrastructure
  - Tunneling techniques used on top of the present IPv4 infrastructure without having to make any changes to the IPv4 routing or the routers
  - Tunneling is often used by networks not yet capable of offering native IPv6 functionality
  - Often used as a first step to test the new protocol and to start integration of IPv6
- ❑ Manual, automatic, semi-automatic configured tunnels are available

http://www.6net.org/book/deployment-

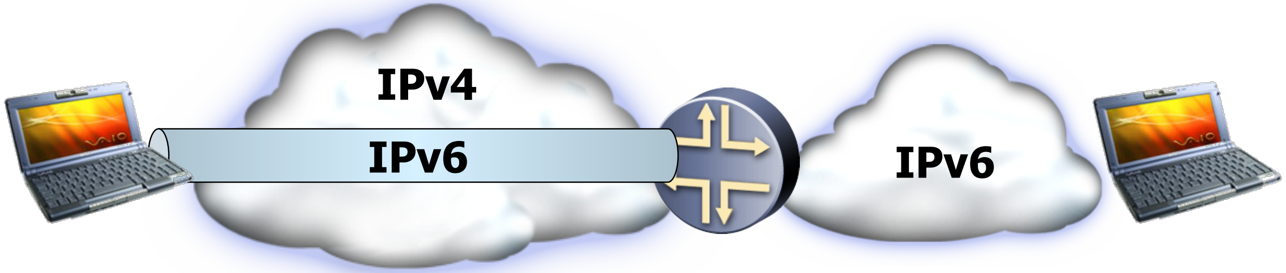
Tunneling Scenarios



Router to Router



Host to Host



Host to Router / Router to Host

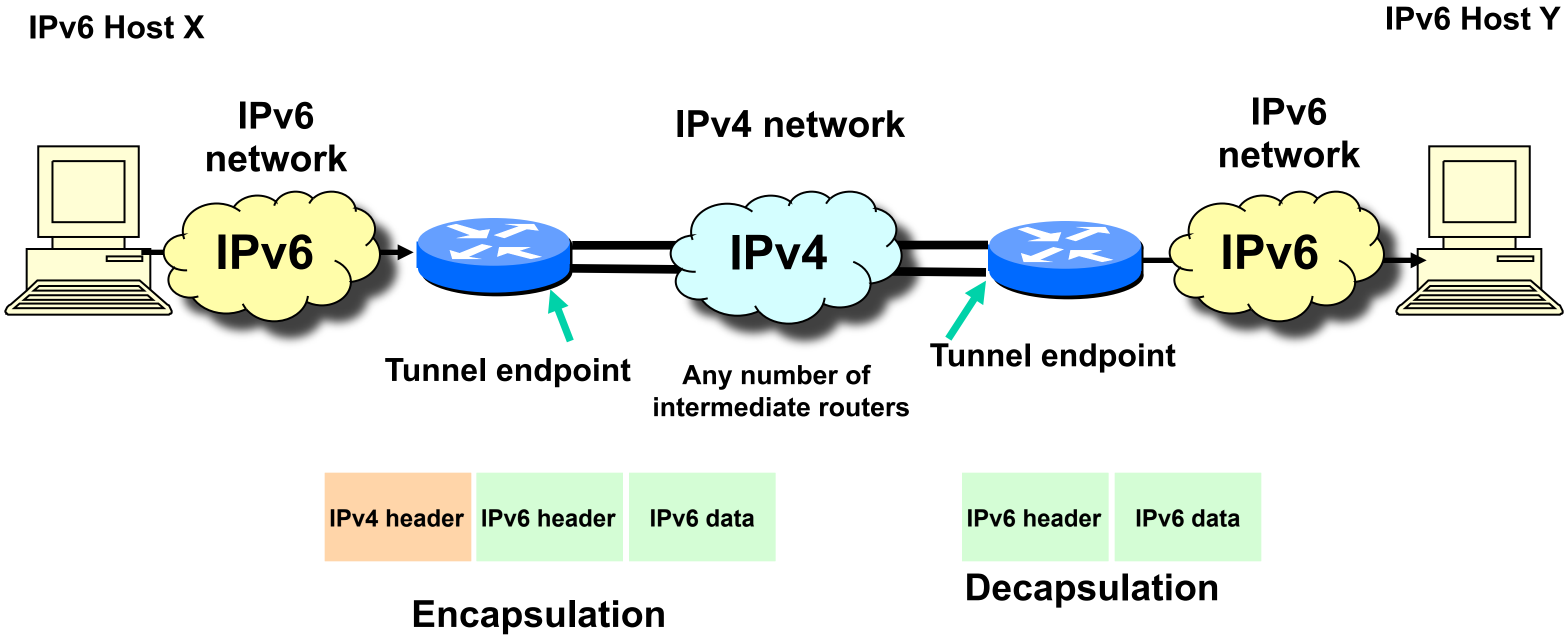
Tunneling - general concept

- A tunnel can be configured in four different ways:
  - Router to router
    - Spans one segment of the end-to-end path between two hosts. Probably the most common method
  - Host to router
    - Spans the first segment of the end-to-end path between two hosts. Many be found in the tunnel broker model
  - Host to host
    - Spans the entire end-to-end path between two hosts
  - Router to host
    - Spans the last segment of the end-to-end path between two hosts
- Other Dimensions:
  - Whether private IPv4 addr./ NAT is involved in one or both ends of the Tunnel ?
  - What’s the trust assumption amongst different devices ?

http://www.6net.org/book/deployment-

Tunneling - general concept

- Tunneling can be used by routers and hosts
  - IPv6-over-IPv4 tunneling
  - Involves three steps
    - Encapsulation, decapsulation, and tunnel management



From Cisco training material "IPv6 Seminar"

IPv6 essentials by Silvia Hagen, p256

Tunnel encapsulation

- The steps for the encapsulation of the IPv6 packet
  - The entry point of the tunnel decrements the IPv6 hop limit by one
  - Encapsulates the packet in an IPv4 header
  - Transmits the encapsulated packet through the tunnel
  - The exit point of tunnel receives the encapsulated packet
    - In some cases, the IPv4 packet may be fragmented
  - The exit point checks whether the source of the packet (tunnel entry point) is an acceptable source (according to its configuration)
    - If the packet is fragmented, the exit point reassembles it
  - The exist point removes the IPv4 header
  - Then it process IPv6 packet to its original destination

IPv6 essentials by Silvia Hagen, p258

Some Tunnel Types

- Manually Configured tunnels
  - Router to router
- Automatic/semi-automatic tunnels
  - Tunnel Brokers (RFC 3053) + Tunnel Setup Protocol (TSP)
    - Server-based automatic tunneling
  - 6to4 (RFC 3056)
    - Router to router
  - ISATAP (Intra-Site Automatic Tunnel Addressing Protocol)
    - Host to router, router to host
    - Maybe host to host
    - Can' t work with private IPv4 addr and NATs
  - 6over4 (RFC 2529) (little adoption)
    - Host to router, router to host
  - Teredo (basis of Microsoft P2P Toolkit)
    - For tunneling IPv6 packets through IPv4 networks with NAT
  - IPv64 (few adoption)
    - For mixed IPv4/IPv6 environments
  - DSTM (Dual Stack Transition Mechanism) (little adoption)
    - IPv4 in IPv6 tunnels

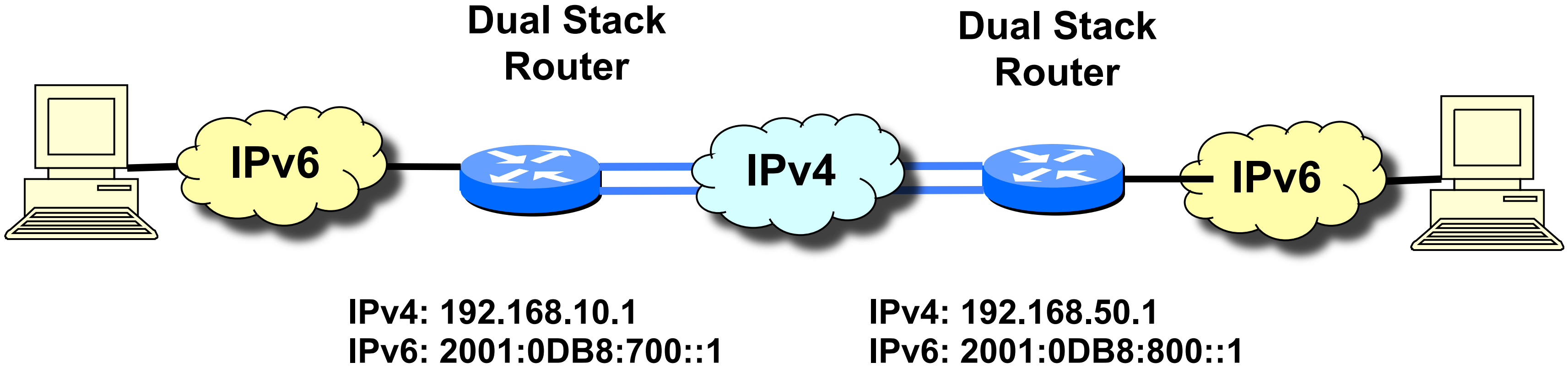
Configuring tunnels

- The IPv4 tunnel's endpoint address is determined by configuring information on the encapsulating node
  - Therefore the encapsulating node must keep information about all the tunnel endpoints addresses
  - Manual configuration
    - The administrative work is higher than with automatic tunnels
- For control of the tunnel paths, and to reduce the potential for tunnel relay DoS attacks
  - Manually configured tunnels can be advantageous over automatically configured tunnels
    - More secure

http://www.6net.org/book/deployment-

Manual configuration

RFC 4213



Manually configured tunnels require:

- Dual stack end points
- Explicit configuration with both IPv4 and IPv6 addresses at each end

from Cisco, Training material “Ipv6 Seminar” delivered at South Asian IPv6 Summit, Jan 2004

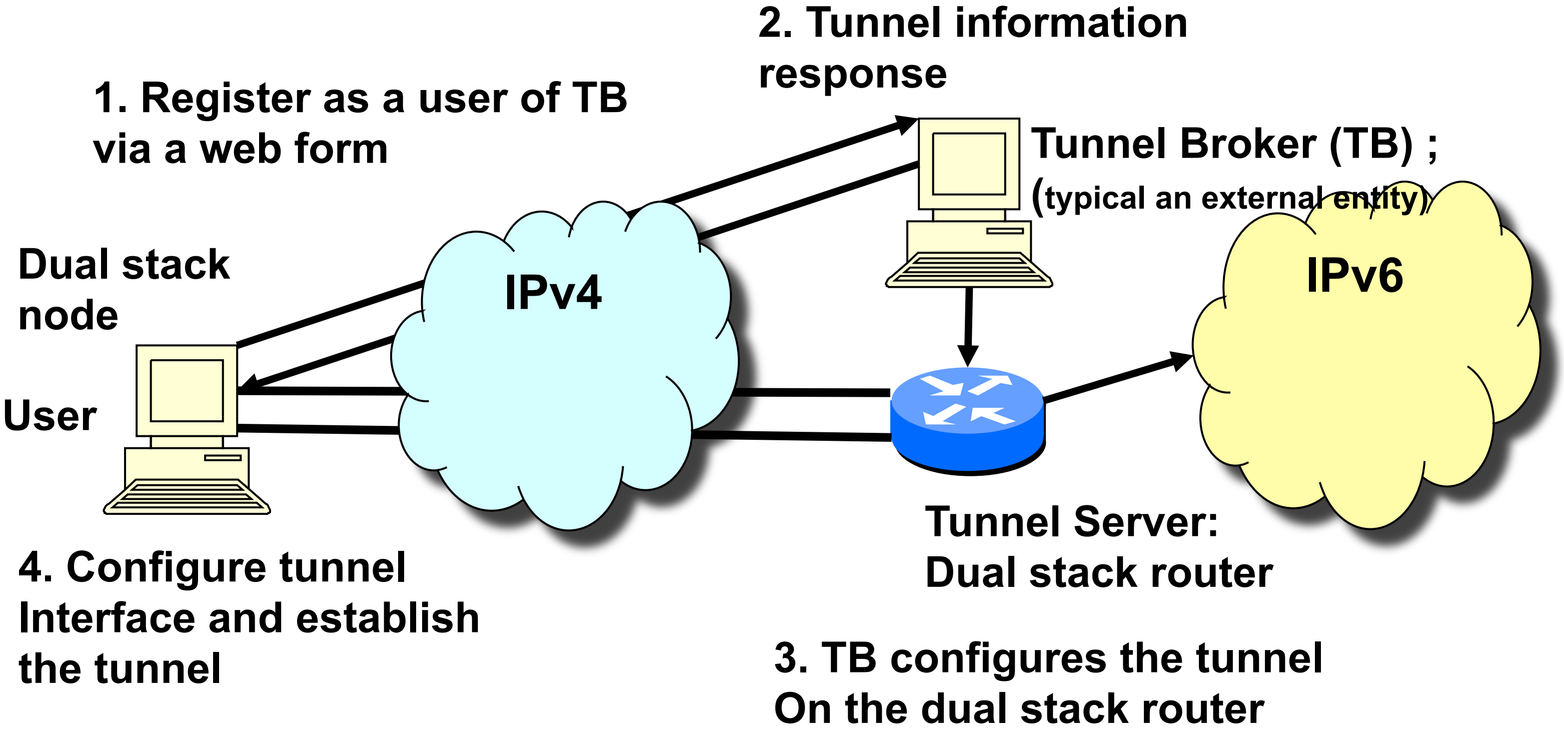
Tunnel broker

- Semi-automatic alternative to manual configuration
- Useful when:
  - A dual stack host in an IPv4-only network wishing to gain IPv6 connectivity
- The basic idea of a tunnel broker
  - It allows a user to connect to a web server
  - Enter some authentication details
  - Receive back a short script to run
  - Establish an IPv6-in-IPv4 tunnel to the tunnel broker server

http://www.6net.org/book/deployment-

Tunnel Broker

RFC 3053


2. Tunnel information response
1. Register as a user of TB via a web form
Tunnel Broker (TB) ; (typical an external entity)
Dual stack node
IPv4
IPv6
User
4. Configure tunnel Interface and establish the tunnel
Tunnel Server: Dual stack router
3. TB configures the tunnel On the dual stack router

- Three basic components:
  - Client: Dual-stacked host or router, tunnel end-point
  - Tunnel Broker: Dedicated server for automatically managing tunnel requests from users, sends requests to Tunnel Server
  - Tunnel Server: Dual-stacked Internet-connected router, other tunnel end point

Tunnel Broker

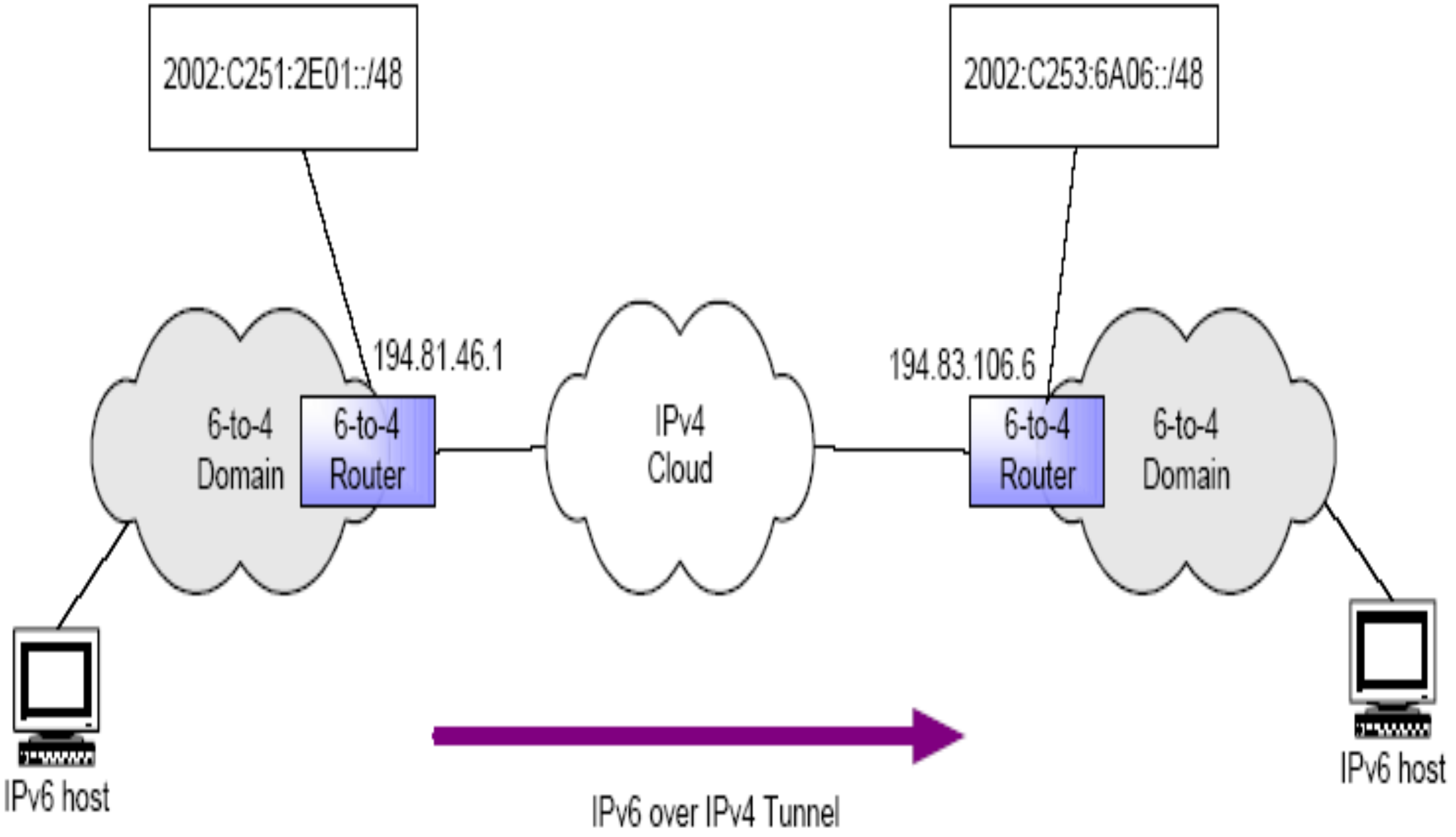
- RFC 3053 describes general architecture, not a specific protocol
- Designed for small sites and isolated IPv6 hosts to connect to an existing IPv6 network

- Free TB services are available:
  - CERNET/Nokia [China] (www.tb.6test.edu.cn)
  - Internet Initiative Japan (www.iij.ad.jp)
  - Hurricane Electric [USA] (www.tunnelbroker.com)
  - Many others…
  - See https://en.wikipedia.org/wiki/List_of_IPv6_tunnel_brokers for details

Tunnel Setup Protocol (TSP)

- Proposed control protocol for negotiating tunnel parameters (IETF RFC 5572 (Experimental), Feb 2010)
    - Applicable to several IPv6 tunneling schemes
    - Can negotiate either IPv6 or IPv4 tunnels
    - Uses XML messages over TCP session
    - Adopted by quite a few IPv6 Tunnel Broker
- Example tunnel parameters:
    - IP addresses
    - Prefix information
    - Tunnel endpoints
    - DNS delegation
    - Routing information
- Three TSP phases:
    1. Authentication Phase
    2. Command Phase (client to server)
    3. Response Phase (server to client)

6to4



http://www.6net.org/book/deployment-

6to4

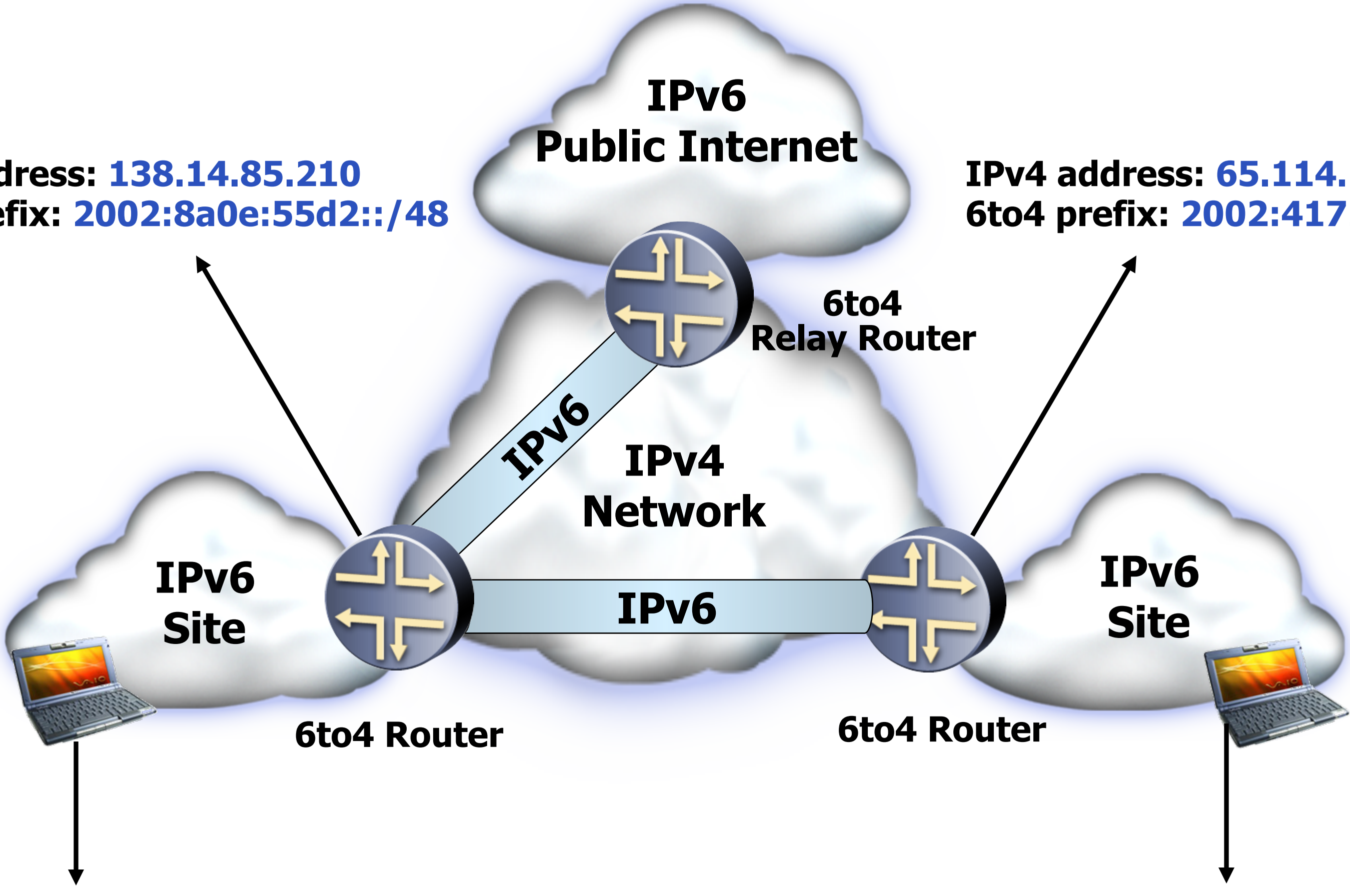
- A form of automatic router-to-router tunneling
  - Tunnel-endpoint (IPv4 address) is already embedded in the IPv6 address
  - Uses the IANA-assigned IPv6 prefix 2002::/16
- Designed for site-to-site and site-to-existing IPv6 network connectivity
- Site border router must have at least one globally-unique IPv4 address
- Uses IPv4 embedded address

Example:

Reserved 6to4 TLA-ID:	**2002::/16**
IPv4 address:	**138.14.85.210 = 8a0e:55d2**
Resulting 6to4 prefix:	**2002:8a0e:55d2::/48**

- Router advertises 6to4 prefix to hosts via RAs
- Embedded IPv4 address allows discovery of tunnel endpoints

6to4



IPv4 address: 138.14.85.210
6to4 prefix: 2002:8a0e:55d2::/48

IPv6
Public Internet

IPv4 address: 65.114.168.91
6to4 prefix: 2002:4172:a85b::/48

6to4
Relay Router

IPv6

IPv4
Network

IPv6
Site

IPv6

IPv6
Site

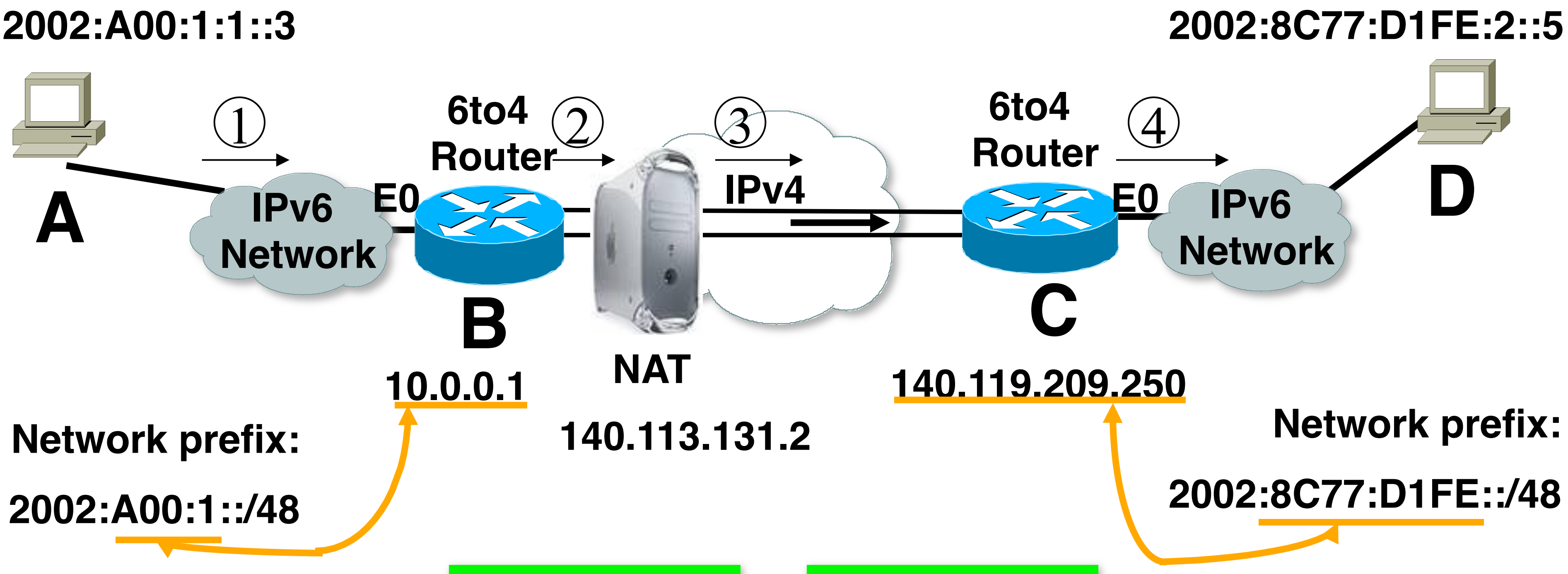
6to4 Router

6to4 Router

6to4 address:
2002:8a0e:55d2::a4ff:fea0:bc97

6to4 address:
2002:4172:a85b::4172:a85b

IPv6 Tunneling Problem (1/2)



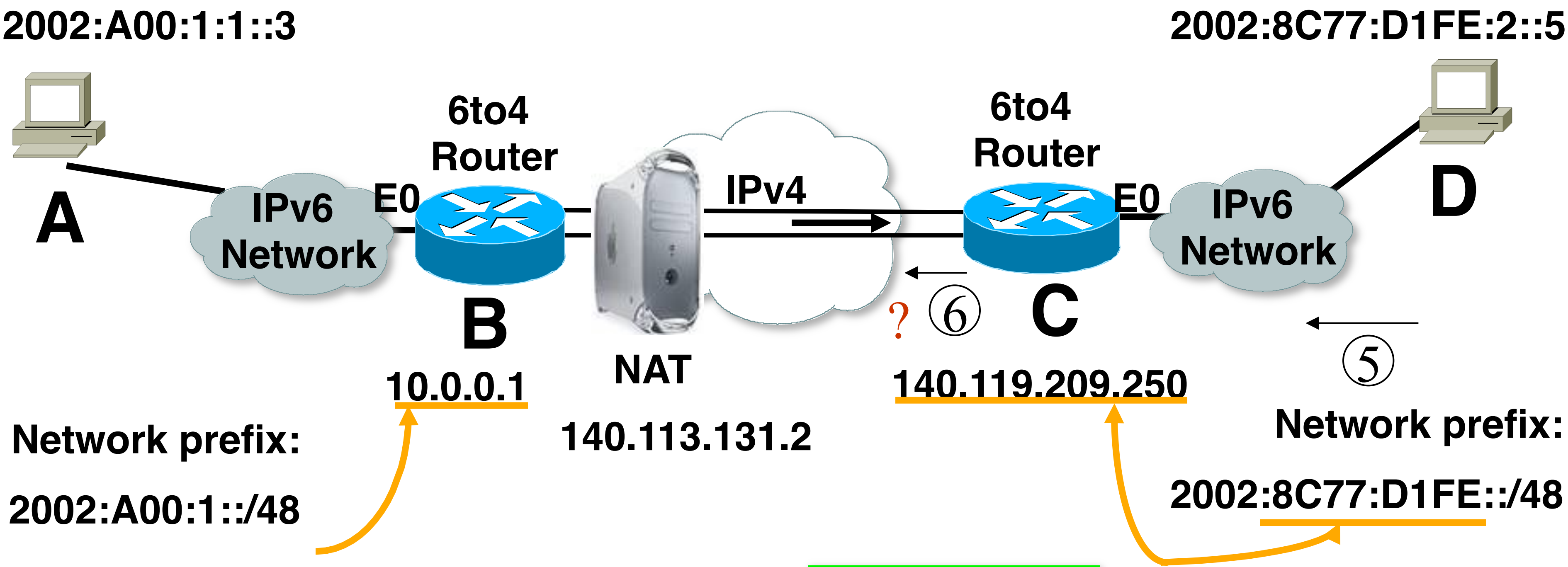
IPv6 SRC 2002:A00:1:1::3
IPv6 DEST 2002:8C77:D1FE:2::5
Data

IPv4 SRC 10.0.0.1
IPv4 DEST 140.119.209.250
IPv6 SRC 2002:A00:1:1::3
IPv6 DEST 2002:8C77:D1FE:2::5
Data

IPv4 SRC 140.113.131.2
IPv4 DEST 140.119.209.250
IPv6 SRC 2002:A00:1:1::3
IPv6 DEST 2002:8C77:D1FE:2::5
Data

IPv6 SRC 2002:A00:1:1::3
IPv6 DEST 2002:8C77:D1FE:2::5
Data

IPv6 Tunneling Problem (2/2)



2002:A00:1:1::3

2002:8C77:D1FE:2::5

A

IPv6 Network

E0

6to4 Router

B

10.0.0.1

NAT

140.113.131.2

IPv4

6to4 Router

E0

C

140.119.209.250

IPv6 Network

D

? ⑥

⑤

Network prefix:

2002:A00:1::/48

Network prefix:

2002:8C77:D1FE::/48

Destination is
Private Address!

IPv4 SRC
140.119.209.250

IPv4 DEST
10.0.0.1

IPv6 SRC
2002:8C77:D1Fe:2::5

IPv6 DEST
2002:A00:1:1::3

Data

IPv6 SRC
2002:8C77:D1Fe:2::5

IPv6 DEST
2002:A00:1:1::3

Data

Teredo (aka "shipworm")

Teredo navalis

- A form of automatic tunnelling intended to provide IPv6 connectivity to IPv4 hosts located behind a NAT
  - Considered as the "last resort" for an isolated Dual-stack host to get onto the IPv6 Internet
  - The host does not possess permanent, global-scope IPv4 addresses
    - Other tunneling solutions require global IPv4 address, and so do not work from behind NAT
  - Host to host automatic tunneling mechanism
    - With NAT hole-punching built-in
  - Provide IPv6 connectivity by encapsulating IPv6 packets in IPv4-based UDP messages
    - Allows pass through most NAT devices
- Require a certain amount of infrastructure
  - Teredo server and Teredo relay
  - UDP port 3544 is used by the Teredo server to listen for requests from the Teredo clients

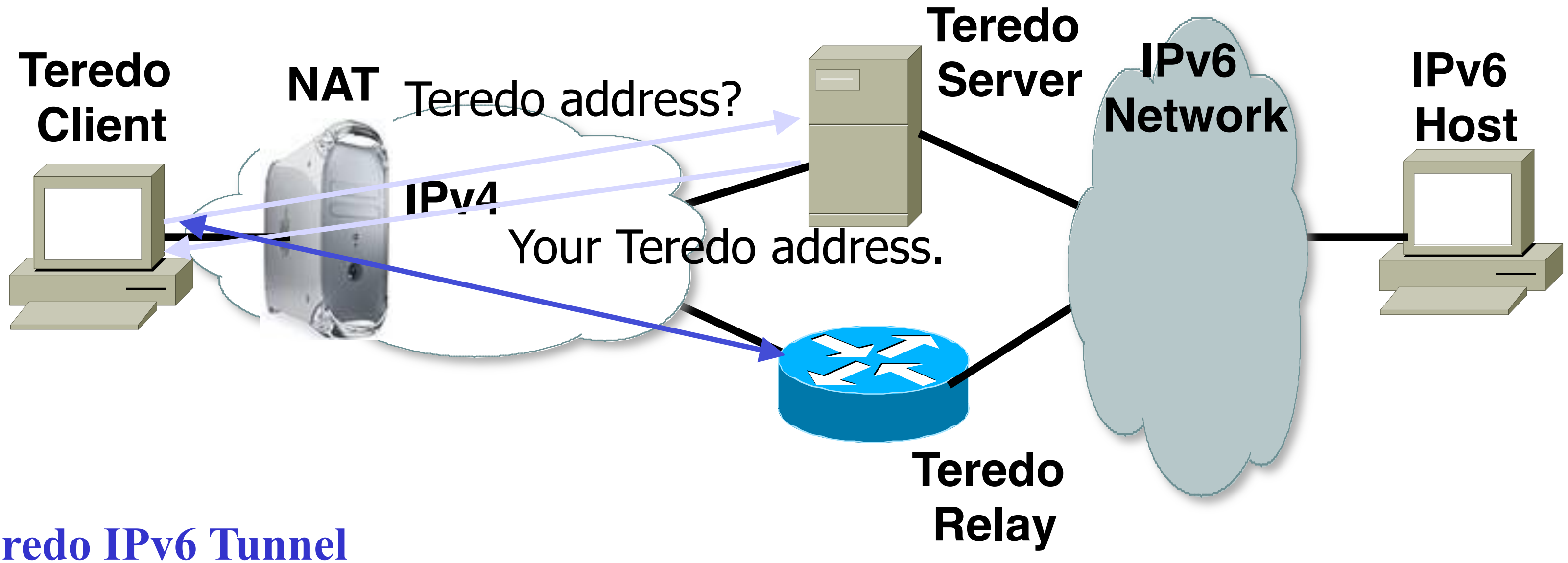
Teredo

- Teredo servers
  - To facilitate the addressing of and communication between Teredo clients and Teredo relays
  - They must be on the public IPv4 Internet
- Teredo relays
  - Gateways between the IPv6 Internet and the Teredo clients
  - To forward the data packets
  - They must be on the IPv4 and IPv6 Internet

IPv6 Network administration, pp 70 - 71

Teredo Operation Model

- Teredo Client gets its Teredo IPv6 address from Teredo Server.
- Use Teredo Relay as Relay router.



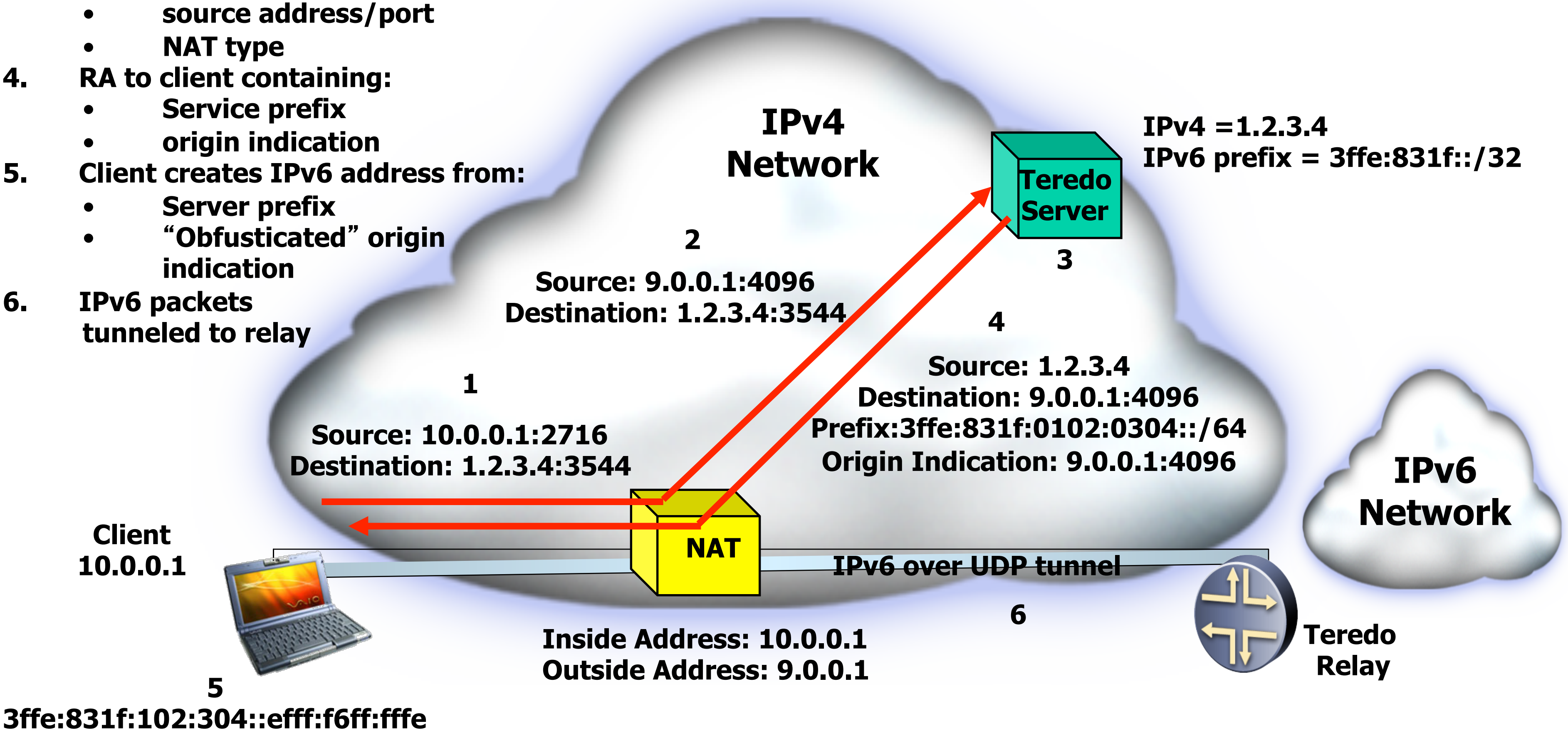
Teredo IPv6 Tunnel

IPv4 Header	UDP Header	Teredo Header	IPv6 packet

Teredo Operation Model (Detail)

1. RS to server
2. NAT maps inside address/port to outside address/port
3. TS notes:
   - source address/port
   - NAT type
4. RA to client containing:
   - Service prefix
   - origin indication
5. Client creates IPv6 address from:
   - Server prefix
   - "Obfusticated" origin indication
6. IPv6 packets tunneled to relay

- ❑ TSP can be used in place of RS/RA for:
  - ❖ Stateful tunnel
  - ❖ Authentication



IPv4 Network

Teredo Server

IPv4 =1.2.3.4
IPv6 prefix = 3ffe:831f::/32

3

2

Source: 9.0.0.1:4096
Destination: 1.2.3.4:3544

4

Source: 1.2.3.4
Destination: 9.0.0.1:4096
Prefix:3ffe:831f:0102:0304::/64
Origin Indication: 9.0.0.1:4096

1

Source: 10.0.0.1:2716
Destination: 1.2.3.4:3544

IPv6 Network

Client
10.0.0.1

NAT

IPv6 over UDP tunnel

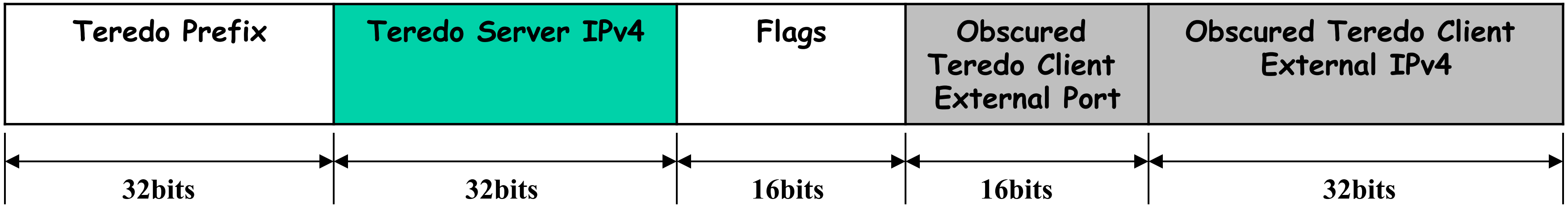
6

Teredo Relay

Inside Address: 10.0.0.1
Outside Address: 9.0.0.1

5
3ffe:831f:102:304::efff:f6ff:fffe

Teredo Address Encoding

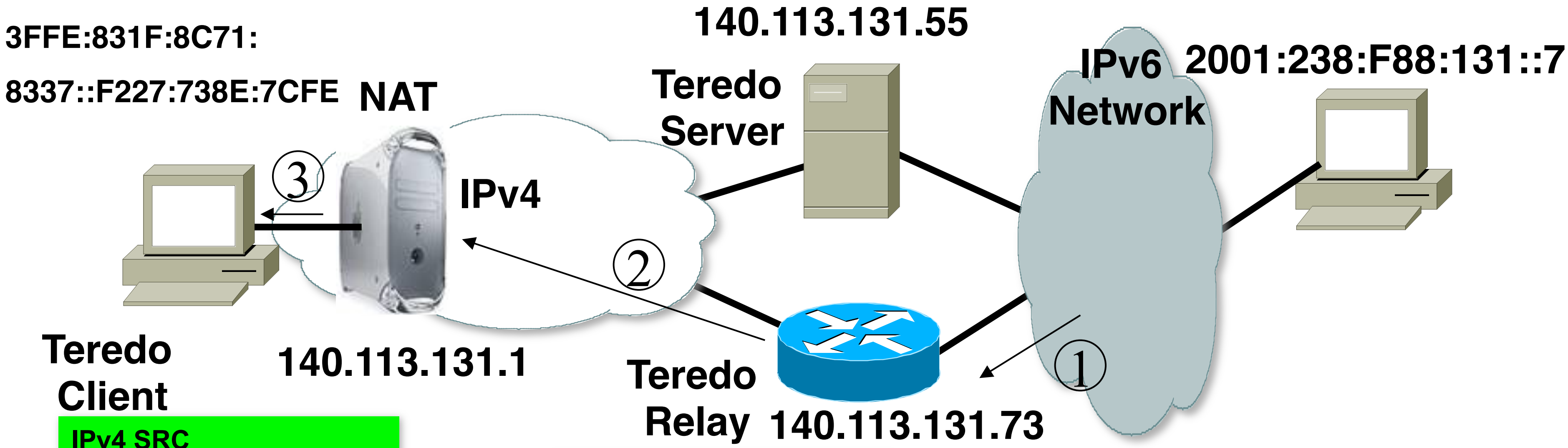


Teredo Prefix	Teredo Server IPv4	Flags	Obscured Teredo Client External Port	Obscured Teredo Client External IPv4
32bits	32bits	16bits	16bits	32bits

- ❑ Teredo Prefix: 32 bit Teredo service prefix.
  - ❖ 3FFE:831F::/32
- ❑ Teredo Server IPv4: IPv4 address of the Teredo server.
- ❑ Flags: 16 bits that document type of address and NAT.
  - ❖ Bit pattern: “C00000UG00000000”
  - ❖ C=1 if NAT is cone.
  - ❖ UG should set to “00”.
- ❑ Obscured Teredo Client External Port: mapped UDP port of the client
- ❑ Obscured Teredo Client External IPv4: mapped IPv4 address of the client

Obfuscated: XOR every bits in the field with 1, prevent “clever” NAT’s translation.

Teredo Tunnel: To host behind NAT



IPv4 SRC
140.113.131.73

IPv4 DEST
10.0.0.1

UDP SRC
3544

UDP DEST
3544

IPv6 SRC
2001:238:F88:131::7

IPv6 DEST 3FFE:831F:
8C71:8337::F227:738E:
7CFE

Data

IPv4 SRC
140.113.131.73

IPv4 DEST
140.113.131.1

UDP SRC
3544

UDP DEST
54392

IPv6 SRC
2001:238:F88:131::7

IPv6 DEST 3FFE:831F:
8C71:8337::F227:738E:
7CFE

Data

IPv6 SRC
2001:238:F88:131::7

IPv6 DEST 3FFE:831F:
8C71:8337::F227:738E:
7CFE

Data

Many other Protocol Translators proposed (but few actually deployed)

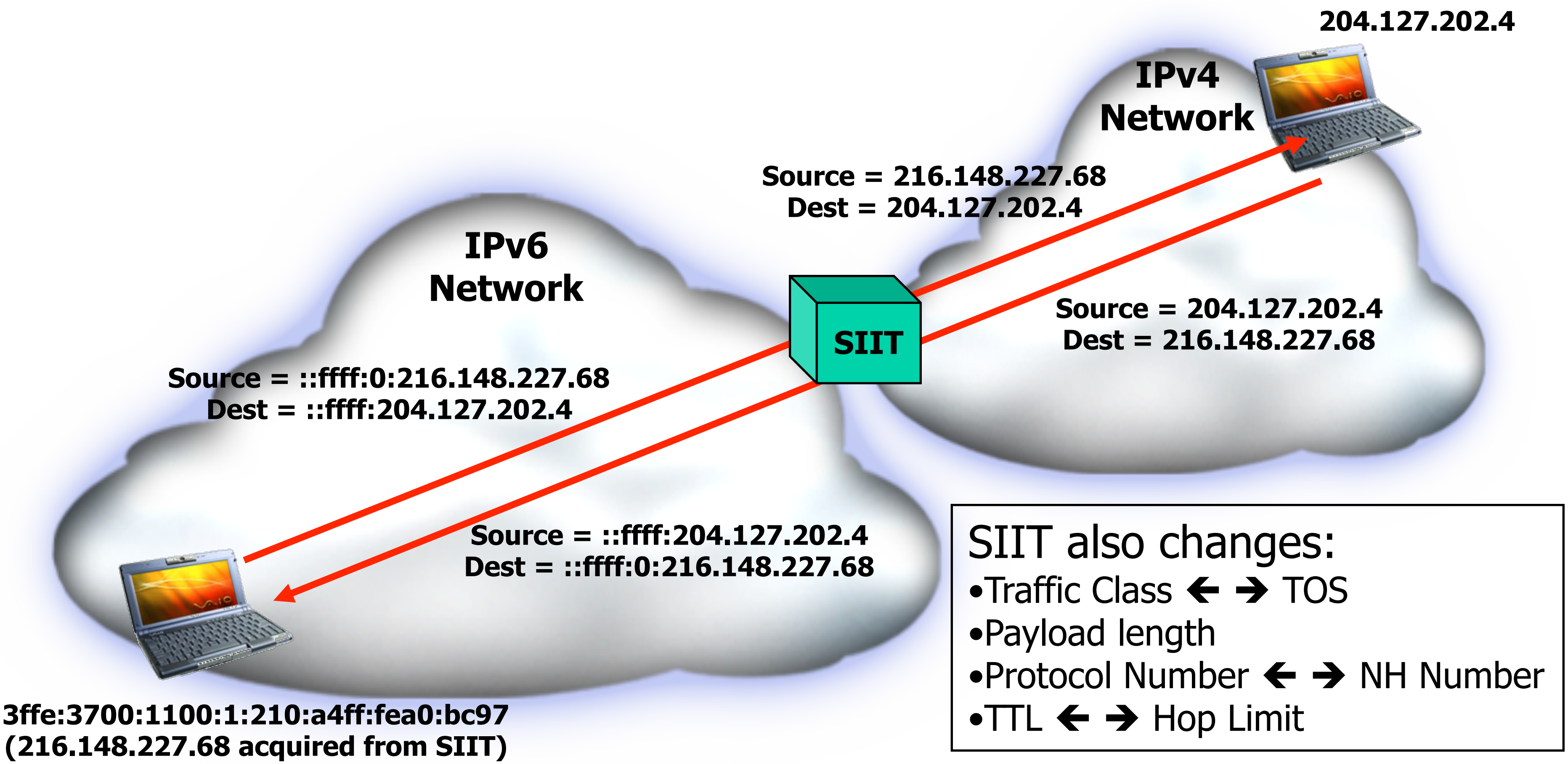
- Network level translators
  - Stateless IP/ICMP Translation Algorithm (SIIT)(RFC 2765)
  - NAT-PT (RFC 2766,4966)
  - Bump in the Stack (BIS) (RFC 2767)
- Transport level translators
  - Transport Relay Translator (TRT) (RFC 3142)
- Application level translators
  - Bump in the API (BIA)(RFC 3338)
  - SOCKS64 (RFC 3089)
  - Application Level Gateways (ALG)

Translation Types

- Network level translators
  - Stateless IP/ICMP Translation Algorithm (SIIT) (RFC 2765)
  - Network Address Translation - Protocol Translation (NAT-PT) (RFC 2766)
  - ~~Bump in the Stack (BIS) (RFC 2767)~~
- ~~Transport level translators~~
  - ~~Transport Relay Translator (TRT) (RFC 3142)~~
- ~~Application level translators~~
  - ~~Bump in the API (BIA)(RFC 3338)~~
  - ~~SOCKS64 (RFC 3089)~~
  - ~~Application Level Gateways (ALG)~~

IETF: NAT-PT should be considered a specialized solution to be applied only when there are no other practical alternatives.

Stateless IP/ICMP Translation (SIIT)



204.127.202.4

IPv4 Network

Source = 216.148.227.68
Dest = 204.127.202.4

IPv6 Network

SIIT

Source = 204.127.202.4
Dest = 216.148.227.68

Source = ::ffff:0:216.148.227.68
Dest = ::ffff:204.127.202.4

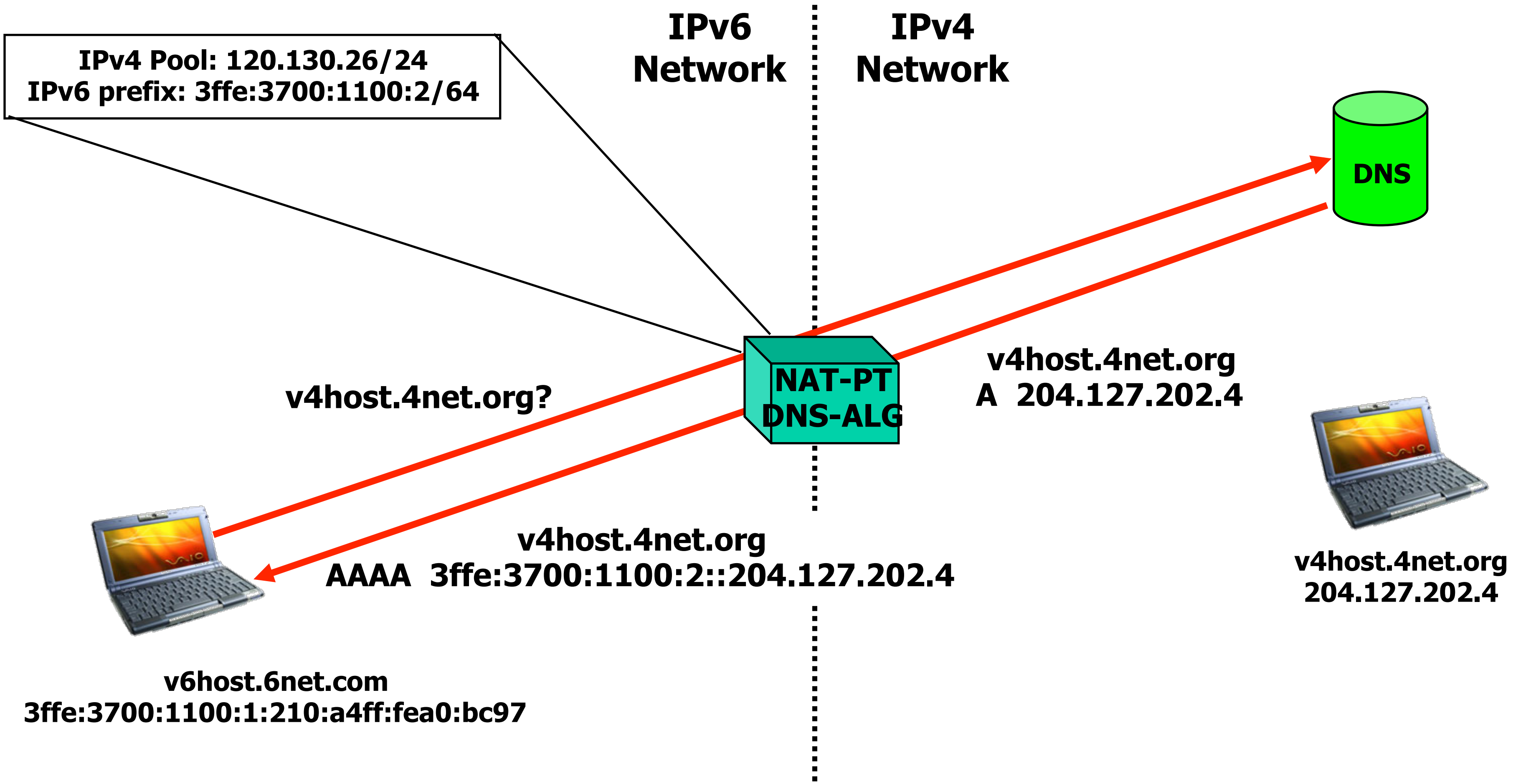
Source = ::ffff:204.127.202.4
Dest = ::ffff:0:216.148.227.68

3ffe:3700:1100:1:210:a4ff:fea0:bc97
(216.148.227.68 acquired from SIIT)

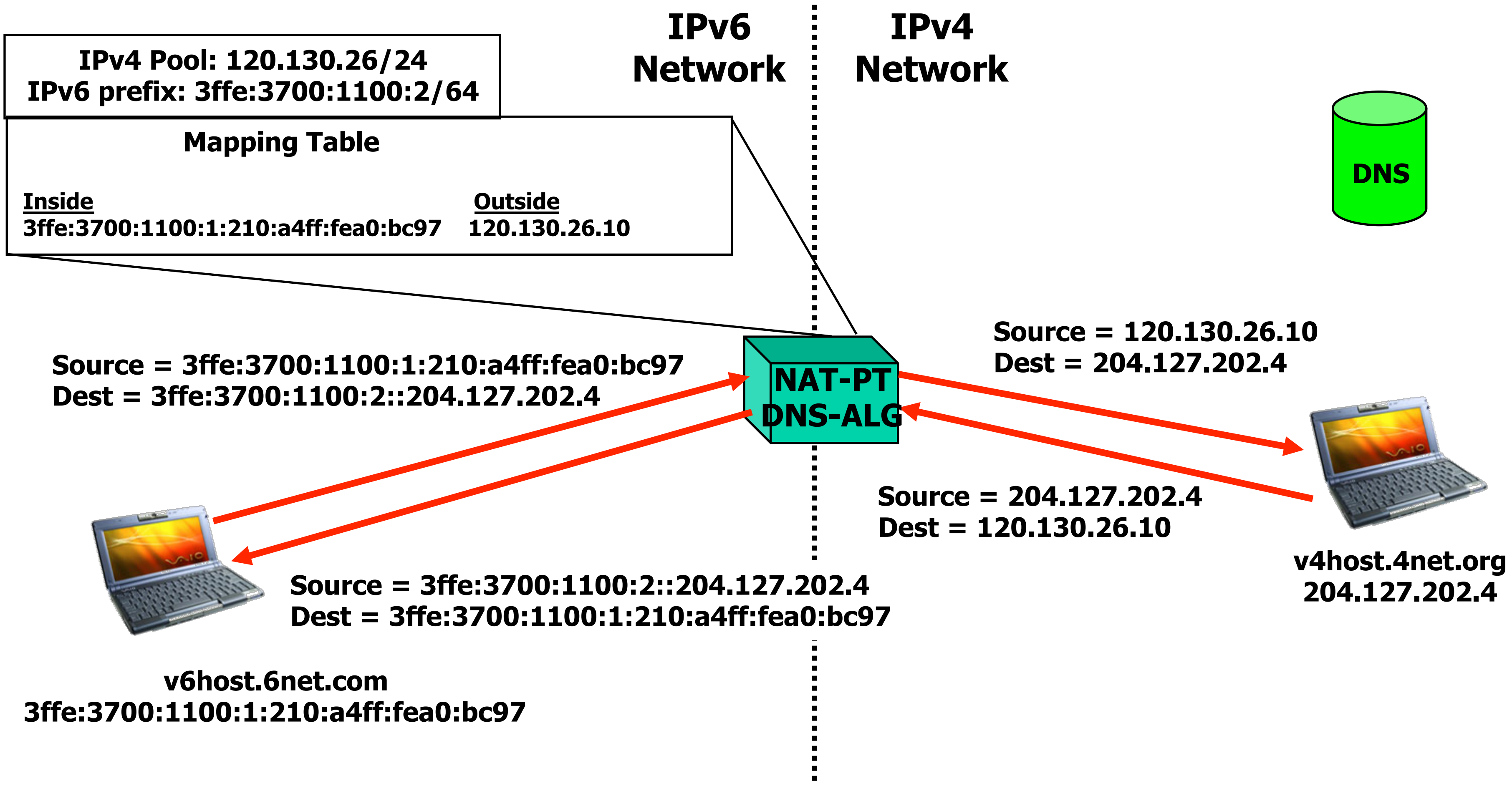
SIIT also changes:

- Traffic Class ← → TOS
- Payload length
- Protocol Number ← → NH Number
- TTL ← → Hop Limit

Network Address Translation - Protocol Translation (NAT-PT)



Network Address Translation - Protocol Translation (NAT-PT)



IPv4 Pool: 120.130.26/24
IPv6 prefix: 3ffe:3700:1100:2/64

Mapping Table

Inside	Outside
3ffe:3700:1100:1:210:a4ff:fea0:bc97	120.130.26.10

Source = 3ffe:3700:1100:1:210:a4ff:fea0:bc97
Dest = 3ffe:3700:1100:2::204.127.202.4

Source = 120.130.26.10
Dest = 204.127.202.4

Source = 204.127.202.4
Dest = 120.130.26.10

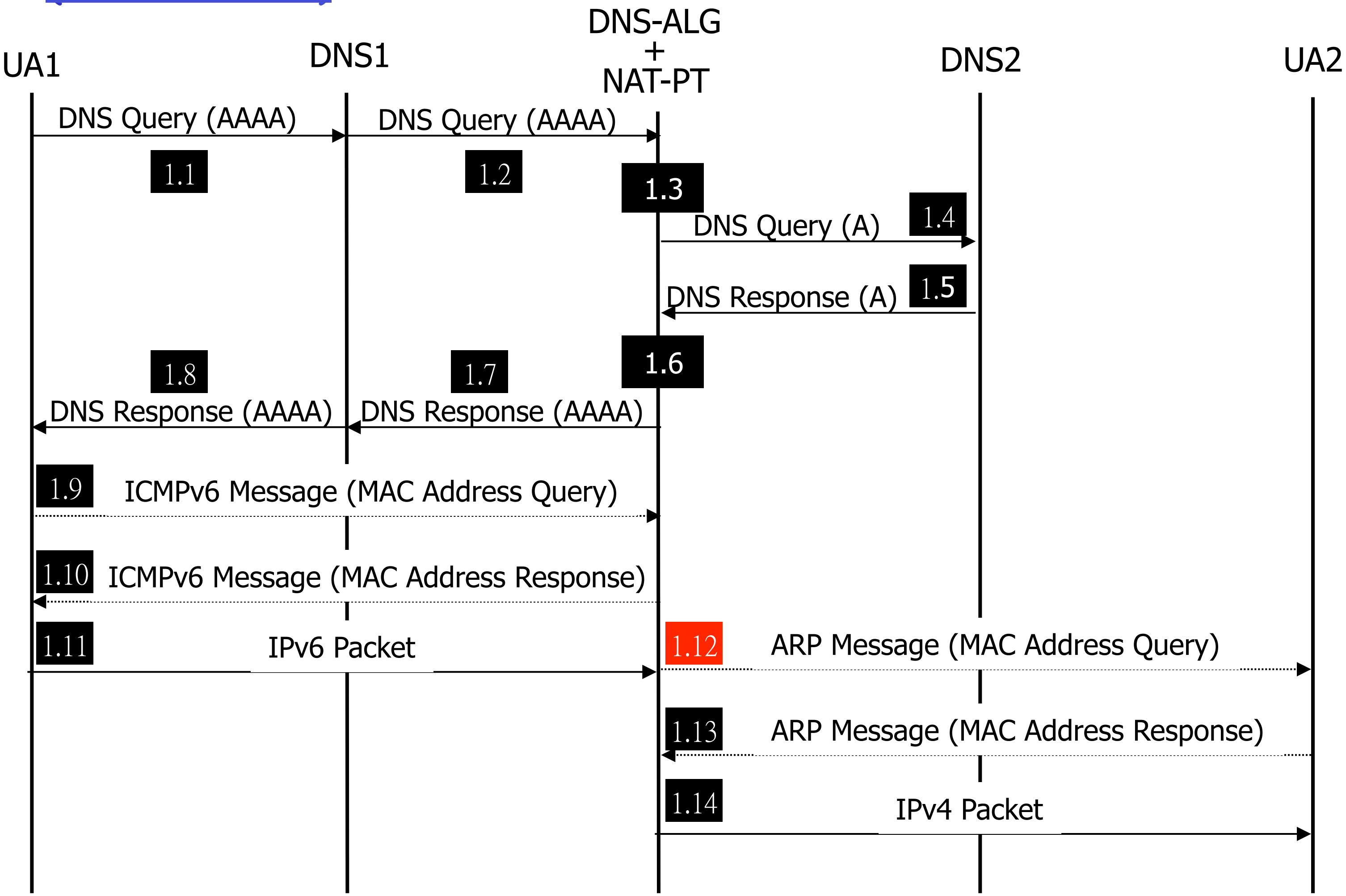
Source = 3ffe:3700:1100:2::204.127.202.4
Dest = 3ffe:3700:1100:1:210:a4ff:fea0:bc97

v6host.6net.com
3ffe:3700:1100:1:210:a4ff:fea0:bc97

v4host.4net.org
204.127.202.4

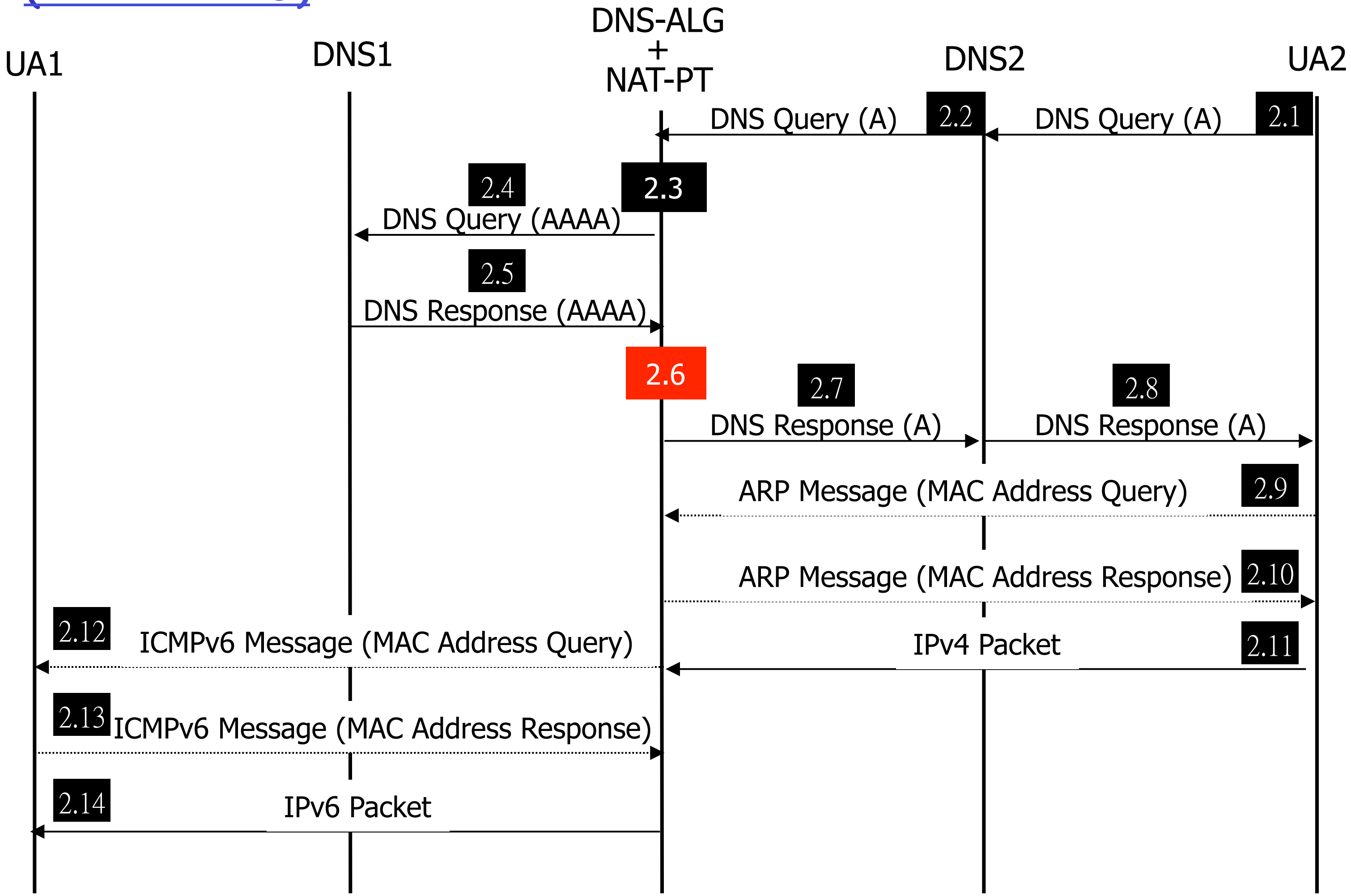
NAT-PT Operations with DNS-ALG

(IPv6→IPv4)



NAT-PT Operations with DNS-ALG

(IPv4→IPv6)



NAT-PT Issues

- NAT-PT falls into sharp critics. **The essential drawback of the legacy translation mechanisms lies in adding extra states for recording address correspondence in the translators.**
- Network Topology Constraints Implied by NAT-PT
  - NAT-PT box also has to be the default IPv6 router for the site so that the DNS-ALG is able to examine all DNS requests made over IPv6. On sites with both IPv6 and dual-stack nodes, this will result in all traffic flowing through the NAT-PT with consequent scalability concerns.
- Issues with Lack of Address Persistence
  - Using the DNS-ALG to create address bindings requires that the translated address returned by the DNS query is used for communications before the NAT-PT binding state is timed out. Applications will not normally be aware of this constraint, which may be different from the existing lifetime of DNS query responses. This could lead to "difficult to diagnose" problems with applications.
  - For example, an application that might have long "silent" periods under normal operation might be required to implement some sort of keepalive mechanisms to insure that NAT-PT does not time out its address binding.
- DoS Attacks on Memory and Address/Port Pools
  - NAT-PT may create dynamic NAT bindings, each of which consumes memory resources as well as an address (or port if NAPT-PT is used) from an address (or port) pool. A number of documents, including [RFC2766] and [NATPT-SEC] discuss the possible denial of service (DoS) attacks on basic NAT-PT and NAPT-PT that would result in a resource depletion associated with address and port pools. NAT-PT does not specify any authentication mechanisms; thus, an attacker may be able to create spurious bindings by spoofing addresses in packets sent through NAT-PT.
- NAT-PT does not discuss a way to map multicast addresses between IPv4 and IPv6.

IETF has now deprecated NAT-PT (RFC2766) to historical status and recommends finding alternate solutions. (RFC4966)

Other Technical Issues for IPv4-to-IPv6 Transition

- Security
  - Firewalls must become smarter
  - Transition vulnerabilities need to be better understood
- Management
  - IPv6 must be managed in conjunction with IPv4
  - Long-term, IPv6 networks should be cheaper to manage than IPv4 networks
- IPv6 Multihoming
  - Myriad proposed solutions, e.g. shim6, see multi6 for details

IPv6 deployment

Issues and concerns

IPv6 current deployment status

- Only used by small fraction of production networks
  See http://www.google.com/ipv6/statistics.html
  - Not many business cases
  - Some symbolic gesture/ recent progress:
    - World IPv6 Launch Day on June 6, 2012 http://www.worldipv6launch.org/
    - "Virtually all" Google Services available on IPv6
      http://googleblog.blogspot.jp/2012/01/ipv6-countdown-to-launch.html
- Quite a few research and experimental networks, e.g.
  - Internet2 IPv6
    - http://ipv6.internet2.edu/
    - /32 address block (2001:468::/32). "Abilene allocates /40 address blocks to Connectors and /48 address blocks to directly connected Participants."
  - The EU 6net
    - http://www.6net.org
- Some statistics to review
  - http://mnlab-ipv6.seas.upenn.edu/
  - http://www.ipv6actnow.org/info/statistics/

What will happen ?

No more room in v4

Quite empty v6

- The real driver for IPv6 in this first decade of the 21st century:
  - **Countries MUST GET enough addresses to support the continued growth of the Internet and IP services into the foreseeable future**
  - **Great needs and incentive for the “late-comers” to the IP game who have huge, growing demand: China, Japan, especially their Wireless Service Providers**

NATs remain to be a Necessary Evil

- Russ Housley, **IETF Chair**, says
  - "NATs are necessary for a smooth transition from IPv4 to IPv6."
- Fred Baker, a Cisco Fellow who was **Chair of the IETF when IPv6 was designed**, says
  - "When the chips are down, NATs may be the only way we are going to get IPv6 added to the Internet. If we have IPv4-only and IPv6-only networks, both of which we have now, NATs are the only way they will connect."
- In particular, the implications of deploying multiple layers of IPv4 address translation need to be considered, as well as those associated with translation between IPv4 and IPv6 which led to the deprecation of [RFC2766] as detailed in [RFC4966].
  - The BEHAVE working group started out with a charter to provide some standard specifications for the behaviour of IPv4 to IPv4 NAT units, but in recent times this has been expanding to encompass the examination of the role of NATs in various IPv6 transition scenarios.
- Dual Stack as a transition mechanism has been effectively abandoned
  - draft-arkko-townsley-coexistence-00