Introduction to IPv6

Yang Hsiung (熊暘)
ZyXEL Communications, Inc

yhsiung@zyxel.com
(408) 735-9736 Ext. 107
VOIP: *51103

ZyXEL

Objectives

- **Length: 2 hours**
- **Objectives:**
  - ✓ **Understand why and what is IPv6**
  - ✓ **Understand IPv6 Addressing**
  - ✓ **Understand the IPv6 basic protocols**
  - ✓ **Understand IPv6 migration mechanisms (depends our time)**

IPv4 Issues

- Severe shortage of IP addresses on Internet
  - With strong growth of 3G wireless devices (mobile phones, PDAs, cars, home appliances) which require global IP addresses to connect Internet—IPv4 is not an option (public shortage & private via NAT).
    - Mobile phone usage: 405M in 2000, 1B+ by 2005
  - The number of Internet users dramatically increasing
    - 320M in 2000, 550M by 2005
  - Growth of "always-on" access technologies (xDSL, cable)
  - Very serious in Asia
    - Stanford University has more IPv4 addresses than China.
    - America owns 78% of the IPv4 addresses
    - Japan is the fastest IPv6 promotion country

IPv4 Issues (2)

- Rapid increase of routing information
  - Subnetting leads to more routing entries
  - Slow down the routing table lookup & memory usage
  - Network routing today is inefficient
- System management manually is costly
  - Versus IPv6 plug-n-play
- Inefficient IP Header Construction
  - Variable length IP header options
  - IP header checksum calculation
  - Packet fragmentation (at routers)
- No support for new applications
  - Mobility
  - QoS (packet classification)
  - Security

Temporary Solutions

- PPP
  - address sharing (via dial-up)
- DHCP
  - address sharing
- NAT
  - private addressing
- CIDR (Classless Inter-Domain Routing)
  - Supernetting (reduce routing table entries)

NAT Issues

- Breaks End-to-End (or P2P) communication
  - The end-to-end (peer-to-peer) communication is destroyed by NAT which was introduced to ease the shortage of addresses.
  - One-way access – outgoing direction based on destination address
- Breaks globally unique address model
- Breaks always-on model
- Single point of failure (NAT device)
- Performance bottleneck
  - Address translation on packet basis
- Breaks the golden rule - It alters the data
  - Changes the IP header and TCP/UDP header, which Conflicts with IPSec.
  - Resolved by "NAT-Traversal" (NAT device in-between)

What Is IPv6?

- Why not IPv5?
  - Used by Internet Stream Protocol (ST2) – experimental protocol
  - RFC 1819
- Internet Protocol Version 6
  - Formerly “IPng” – IP Next Generation
- IPv6 is a set of specifications from the Internet Engineering Task Force (IETF)
- IPv6 was designed as an evolutionary set of improvements to the current IP Version 4
  - Scalability (128-bit addressing)
  - Efficiency
  - Extensibility (more features via Extension Header)

IPv6 Features

- Huge IP Address Space
  - 128 bits IPv6 address vs. 32 bits IPv4 address
    - 2^{128} = 3.4 x 10^{38} addresses vs. 4B
    - 6.65 x 10^{23} addresses per square meter of the Earth's surface
  - Multiple levels of subnetting (network prefix)
  - No longer needs NAT
- "Simple" architecture for high-speed networks
  - Aggregatable Address (route aggregation: prefix::/n)
  - Extension Header
- Plug-n-Play
  - Stateless (Serverless) Address Auto-configuartion
- Native Authentication and Security
  - Built-in IPSec

IPv6 Features (2)

- **Seamless (True) Mobility Support**
  - **Built-in Mobile IPv6**
- **Maintain End-to-End (E2E) Internet Nature**
  - **Chat/VIOP/Video Conference**
  - **Network Games**
  - **Other Peer-to-Peer (P2P) Applications**
  - **Two-way Access based on global unique destination address**
- **Better support for Quality-of-Service (QoS)**
  - **Improve IP telephony, videoconferencing, and multicasting, which is sending data, voice, and video from a central place to many different sites.**
  - **Use flow-label management to provide differentiated services**

IPv6 Features (3)

- **Routing Improvement**
  - Reduction in routing load based on the use of a hierarchical address
  - Core routers have much smaller routing table, less memory usage
  - Speed-up routing table look-up
- **New Header Format**
  - Keep header overhead to minimum by moving both non-essential fields and options to extension header
  - No limitation (< MTU packet size) on options vs. 40-byte of options in IPV4
  - More efficient on header processing
    - Fixed size IPv6 header, less fields in the main header
    - Processing most of the options only at final destination
    - Simplify header processing at intermediate routers

IPv6 Features (4)

- New Neighbor Node Interaction Protocol
  - Neighbor Discovery Protocol (RFC 2461)
- Extensibility
  - Easily be extended for new features by adding extension headers
- Multicast Support
  - Built-in implementation as native communication mode
  - IPv4 multicast is optional
- Address Auto-configuration (RFC 2462)
  - Auto-configuration of IP addresses for a "plug-and-play" environment
  - Stateless and stateful address configuration

IPv6 – what's changed

- Expanded Address Space
  - 32 bits => 128 bits
- Header Format Simplification
  - IPv4 length is 20 bytes + various options
  - IPv6 length is 40 bytes, optional headers are daisy-chained
- No checksum at IPv6 header
  - Will it be an issue?
  - Left to transport and data link layers
  - Data links are more reliable these days
  - Upper layer checksums are mandatory (TCP, UDP, ICMPv6)
  - No need to check/recalculate each hop (performance)

IPv6 – what's changed (2)

- No hop-by-hop fragmentation (except at source)
  - Path MTU discovery on the sending node
  - Reduce loads on routers
  - Easier to implement in hardware
  - Easier for L3 switching
- Addressing Auto-Configuration
  - Stateless
  - Stateful (DHCPv6)
- 64 bits aligned Header/Options
- IP Mobility
  - MIP6 is mandated
- Authentication and Privacy Capabilities
  - IPSec is mandated
- No more IP broadcast
  - Use multicast

IPv6 Addressing

IPv6 Addressing

- The size of an IPv6 address is 128 bits (16 bytes).
- Consists of a 64-bit (Global Routing Prefix + Subnet ID) and a 64-bit interface ID.
- A subnet mask is not used (always 16 bits)
- IPv6 addresses are assigned to interfaces, not nodes.
- A single interface may be assigned multiple IPv6 addresses of any type.

IPv6 Address Notations

- The 128-bit address is divided into 8 16-bit blocks
- Each 16-bit block is converted to hex and delimited with colons :
  - 21DA:00D3:0000:0000:02AA:00FF:FF28:9C5A
- Suppress the leading zeros within each 16-bit block
  - 21DA:D3:0:0:2AA:FF:FF28:9C5A
- A single contiguous sequence of 16-bit blocks set to 0 can be compressed to double colon (::)
  - 21DA:D3::2AA:FF:FF28:9C5A

Zero Compression

- To determine how many 0 bits are represented by the "::"
  - Subtract the number of blocks in the compressed address from 8 (blocks) and them multiply by 16 (bits)

- Example:
  - FF02::2 – the number of bits represented by "::" is 96
  - (8 blocks – 2 blocks) x 16 bits = 96 bits

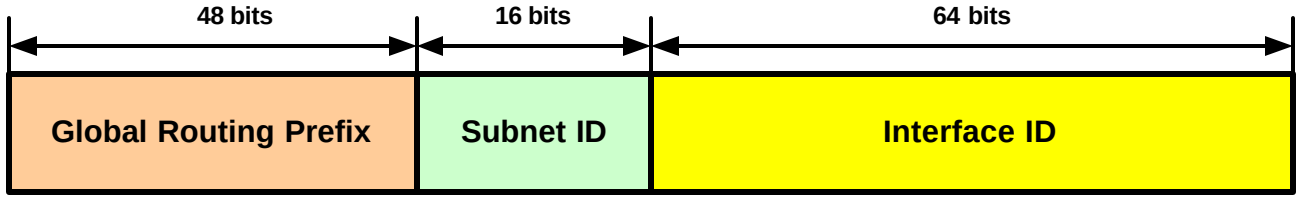
- Zero compression can only be used once in a given address
  - 3ffe:0000:0000:cd30:0000:0000:0000:1234
  - 3ffe:0:0:cd30:0:0:0:1234
  - 3ffe::cd30:0:0:0:1234
  - 3ffe:0:0:cd30::1234

More IPv6 Address Notations

- **Special addresses**
  - 0:0:0:0:0:0:0:1 (::1) – Loopback (IPv4 – 127.0.0.1)
  - 0:0:0:0:0:0:0:0 (::) – Unspecified Address (IPv4 – 0.0.0.0)
- **Prefix representation (CIDR notation – prefix/length)**
  - 3ffe:0000:0000:cd30:0000:0000:0000:0000/60
  - 3ffe::cd30:0:0:0:0/60
  - 3ffe:0:0:cd30::/60
- **Literal Address in URL (RFC 2732)**
  - http://[2001::4:fedc:ba98:7654:210]:80/index.html
  - Use [ ] to delimit the address and port

IPv6 Address Format

- **Global Routing Prefix (old name: Format Prefix)**
  - Identify special addresses (unicast, multicast).
- **Subnet ID**
  - 16-bit (65536 subnets), identify a link within a site.
  - A subnet ID is associated with one link.
  - No subnet mask (always 16 bits)
- **Interface ID**
  - Identify an unique interface on a link.
  - 64-bit ID that follows the IEEE EUI-64 (Extended Unique Identifier) format.



IPv6 General Address Format

Types of IPv6 Addresses

- **Unicast (1-to-1)**
  - Identify a single interface within the scope of the address type.
- **Multicast (1-to-M)**
  - Identify a group of interfaces, such that a packet sent to a multicast address is delivered to all of the interfaces in the group.
  - No broadcast addresses in IPv6 (superseded by multicast).
  - Optional in IPv4.
- **Anycast (1-to-nearest)**
  - RFC 1546
  - Identify a set of interfaces, such that a packet sent to an anycast address will be delivered to one member (nearest) of the set.
  - Optional in IPv4 and only used by IPv4 routers.
  - Allocated out of the unicast address space.
  - No way to distinguish unicast and anycast addresses.

IPv6 Unicast Addresses (Scope)

- **Aggregatable global unicast address (routable)**
  - Equivalent to IPv4 public address
  - Scope is entire IPv6 Internet
- **Link-local address (not routable)**
  - Equivalent to Automatic Private IP Addressing (APIPA) addresses autoconfigured on Windows (169.254.0.0)
  - Scope is the local link (e.g., LAN)
- **Site-local address (intranet routable)**
  - Equivalent to IPv4 private address (10.0.0.0, 172.16.0.0, 192.168.0.0)
  - Scope is the site

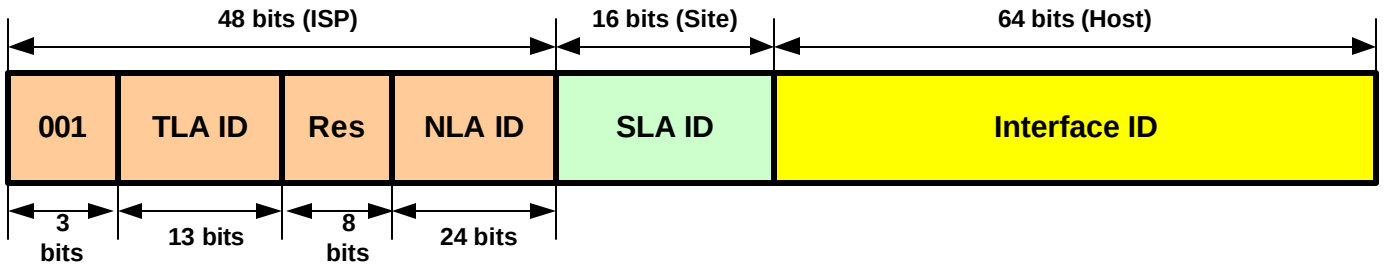
IPv6 Unicast Addresses (2)

- Special address
  - Unspecified address (::)
    - Equivalent to IPv4 0.0.0.0
    - Must never be assigned to any node
    - Represents absence of an address
    - Must never be used as destination address in IPv6 packets
    - Used for autoconfiguartion DAD
  - Loopback address (::1)
    - Equivalent to IPv4 127.0.0.1
    - Can never be assigned to any physical interface
    - Used by nodes to send packets to themselves
    - Traffic destined to loopback address must never leave the sending node

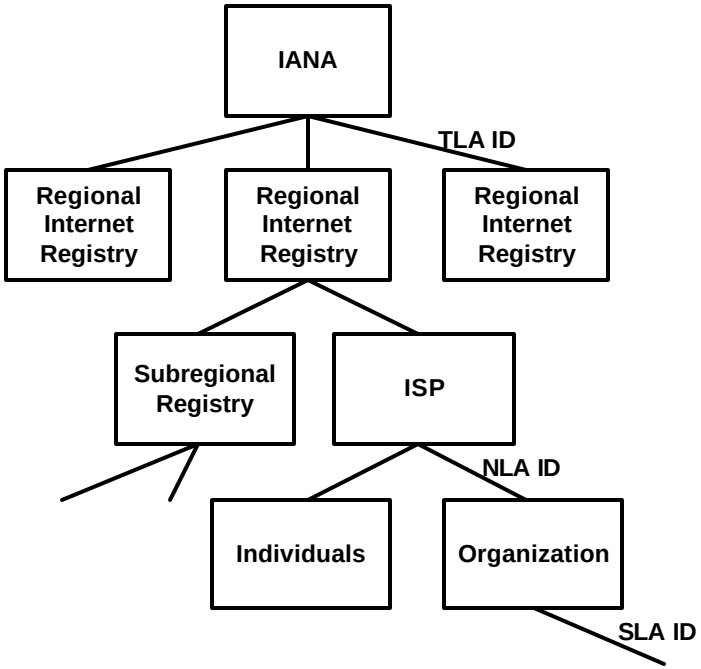
IPv6 Unicast Addresses (3)

- Compatibility address
  - Used for dual-stack nodes with v4 and v6 migration
  - IPv6 address assignment is based on v4 address
  - Used for automatic tunnels
  - IPv4-compatible address
    - 0:0:0:0:0:0:w.x.y.z
    - Intend to tunnel IPv6 packets through IPv4 routers
    - Where w.x.y.z is the IPv4 address
  - IPv4-mapped address
    - 0:0:0:0:0:FFFF:w.x.y.z
    - Used by IPv6 devices sending to nodes only supporting IPv4
  - 6over4 address
    - [64-bit prefix]:0:0:wwxx:yyzz
  - 6to4 address
    - 2002:wwxx:yyzz:[SLA ID]:[Interface ID]
  - ISATAP address
    - [64-bit prefix]:0:5EFE:w.x.y.z
    - w.x.y.z can be either public or private IPv4 address

Aggregatable Global Unitcast

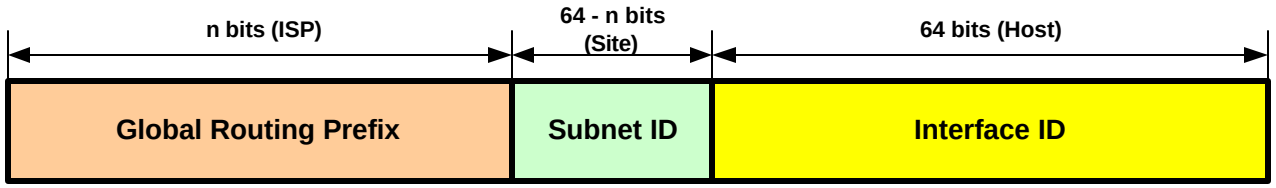


- RFC 2374 – obsolete
- TLA: Top-Level Aggregation
- NLA: Next-Level Aggregation
- Global Routing Prefix = TLA + NLA
- SLA: Site-Level Aggregation
- IANA: Internet Assigned Numbers Authority
- 48-bit public topology (external routing prefix)
- 16-bit site topology



New Global Unitcast

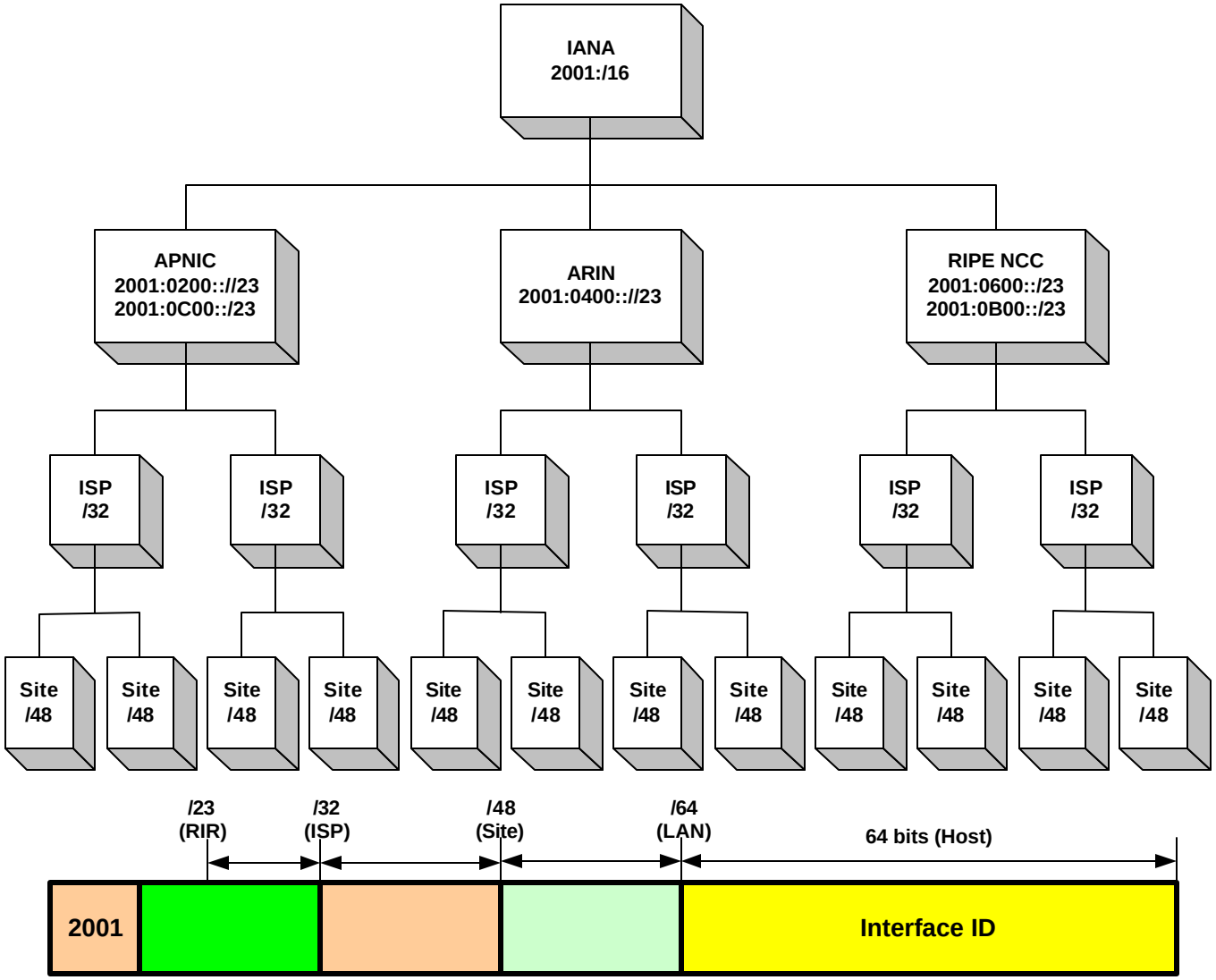
- Basic unicast address format defined in IPv6 Global Unicast Address RFC 3513, RFC 3587 (replace RFC 2374).
- TLA/NLA: disappears
- Global Routing Prefix = TLA + NLA
- SLA maintained under subnet ID



Aggregatbale Global Unitcast

- They begin with (where *"x"* is any hex character)
  - 2*xxx*: 0010
  - 3xxx: 0011
- There are some further subtypes defined, see below:
  - 6bone test addresses (pseudo address)
    - 3ffe
    - IPv6 backbone is a test-bed network for validating IPv6 standards and implementations.
  - 6to4 addresses (6to4 tunnel address)
    - 2002
  - Assigned by provider for hierarchical routing (production address)
    - 2001

Another View

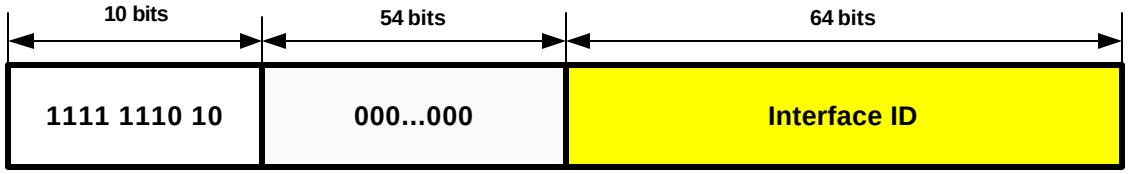


Interface ID

- Lowest-order 64-bit field of unicast address may be assigned in several different ways:
  - Auto-configured from a 64-bit EUI-64 MAC address
    - Most common approach
  - Auto-generated pseudo-random number (to address privacy concerns)
    - Temporary address for IPv6 host client application
    - Run DAD (Duplicate Address Detection) before using it
    - IPv6 Address Privacy (RFC 3041)
    - Rate of change based on local policy
  - Manually configured

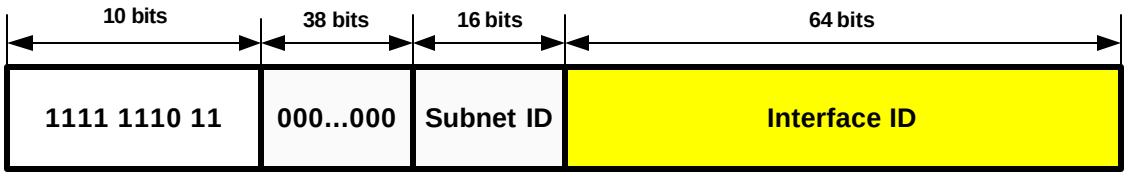
Link-local Unitcast

- These are special addresses which will only be valid on a single link.
- Link-local addresses for use during auto-configuration and when no routers are present (no site-local and global addresses yet).
- An address with this prefix is found on each IPv6-enabled interface after stateless auto-configuration.
- Using this address as destination the packet would never pass through a router.
- It's used for link communications via NDP such as:
  - anyone else here on this link? (neighbor discovery)
  - anyone here with a special address (e.g. router discovery, DAD, auto-config)?
- They begin with (where *"x"* is any hex character, normally *"0"*)
  - fe8*x:* <- *currently the only one in use.*
  - fe9*x:*
  - fea*x:*
  - feb*x:*

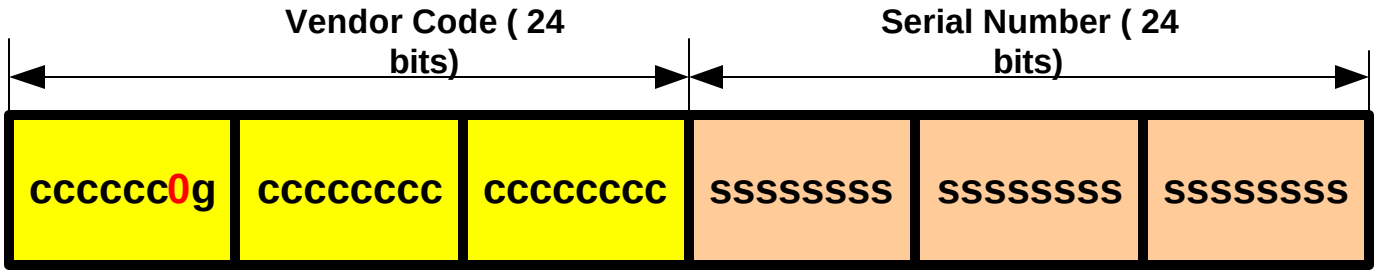


Site-local Unitcast

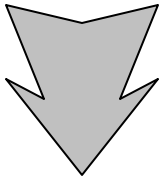
- These are addresses similar to the IPv4 private address today.
- They are not globally reachable.
- To be used within a site only.
- Edge routers must keep site-local traffic within site.
- They begin with (where *"x"* is any hex character, normally *"0"*)
  - fec*x:* <- *most commonly used.*
  - fed*x:*
  - fee*x:*
  - fef*x:*

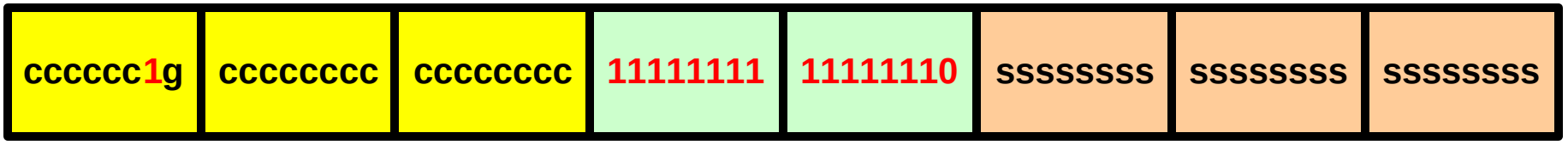


EUI-64 Interface ID (RFC 2373)



- Global/Local -> complement
- Insert FF, FE
- E.g. 00:40:05:34:D9:7F
- => 0240:05ff:fe34:d97f
- => fe80::0240:05ff:fe34:d97f



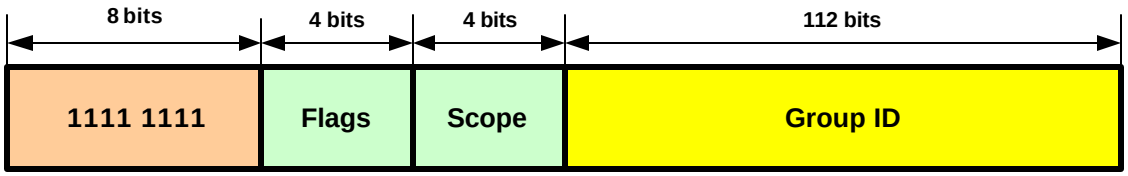


IPv6 Multicast Addresses

- IPv6 nodes can listen to multiple multicast addresses at the same time.
- IPv6 nodes can join and leave a multicast group at any time.
- IPv6 multicast addresses have the Format Prefix of 1111 1111 (0xFF).
- Multicast addresses cannot be used as source addresses or as intermediate destinations in a Routing header.
- IPv6 does not have IP broadcast address.
- Multicast addresses are scoped (see next slide)

Multicast Address Format (RFC 3513)

- Flags
  - Transient (T) flag – well-known permanent address (0), temporary address or Group ID (1).
  - Prefix (P) flag – multicast address assigned based on the network prefix (1), not based on prefix (0).
- Scope
  - The scope of this multicast address covers
  - Node-local (1), Link-local (2), Site-local (5), Organization-local (8), Global (E), reserved (others).
- Group ID
  - Identify the multicast group and is unique within the scope.
  - All nodes (1), all routers (2), all DHCP server (0x1000)



Fixed Multicast Addresses

- http://www.iana.org/assignments/ipv6-multicast-addresses
  - FF01::1 (node-local scope all-nodes)
    - Node local scope is applicable within a node. Such scope is defined because IPv6 nodes can have multiple addresses.
  - FF02::1 (link-local scope all-nodes)
  - FF01::2 (node-local scope all-routers)
  - FF02::2 (link-local scope all-routers)
  - FF05::2 (site-local scope all-routers)
- Solicited-Node Multicast Address
  - Efficient querying for link-layer address resolution
  - Replace multicast of link-local scope
  - FF02::1:FF00:0/104 + 24-bit of unicast IPv6 address
  - Node A is assigned the link-local address of FE80::2AA:FF:FE28:9C5A, is also listening FF02::1:FF28:9C5A

Fixed Multicast Addresses (2)

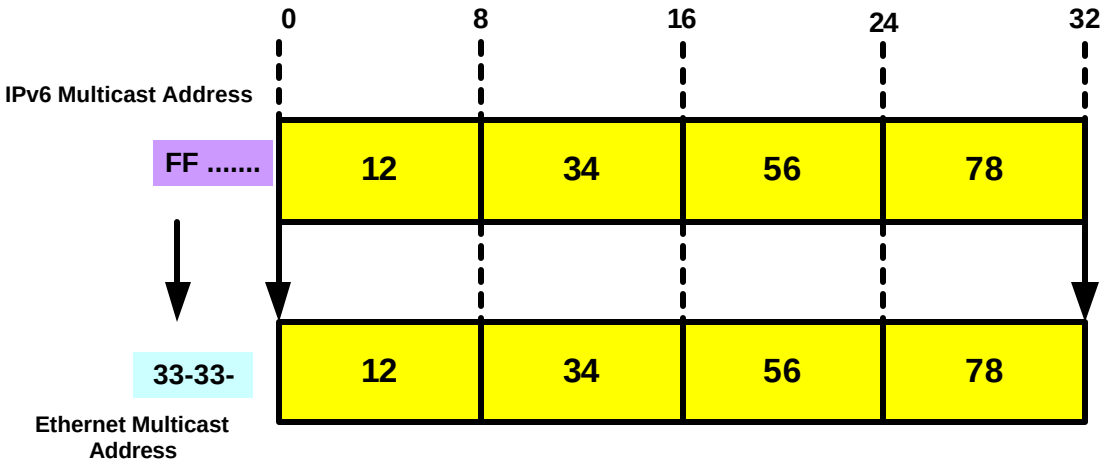
- At boot time, every IPv6 node must join 2 special multicast groups for each network interface:
  - All-nodes multicast group - ff02::1
  - Solicited-node multicast group - ff02::1ffxx:xxxx (derived from the lower 24-bit of the node's address)
  - During ND processing, these two multicast addresses will be used quite often (will be covered in ND)
- How about IPv6 routers?
  - All-routers multicast group – ff02::2

Multicast Address Mapping
(multicast IP => multicast MAC)

FF02::1:FF02:6EA5 => 33-33-FF-02-6E-A5

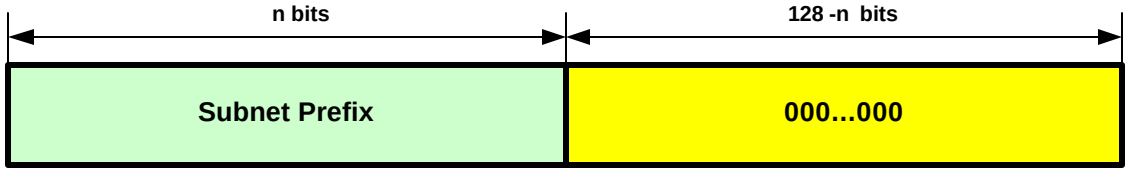
FF02::1=>33-33-00-00-00-01 (link-local scope all-nodes)

What NIC driver should do?



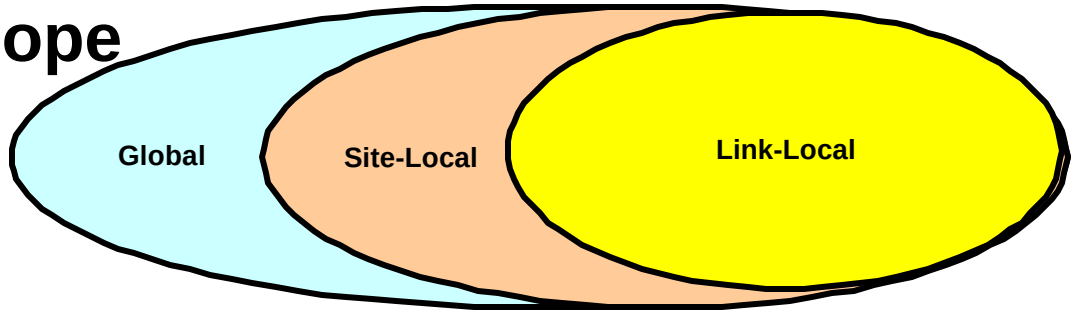
IPv6 Anycast Addresses

- Anycast addresses are used only as destination addresses
- Anycast addresses are assigned only to routers
- Allocated out of the unicast address space.
- It is not possible to determine if a given destination unicast address is also an anycast address.
- Used to address multiple interfaces on different nodes with the SAME IPv6 address.
- Example, Subnet-Router Anycast Address
  - All router interfaces attached to a subnet are assigned the Subnet-Router anycast address for that subnet.
  - It is used to communicate with the nearest router connected to a specified subnet.



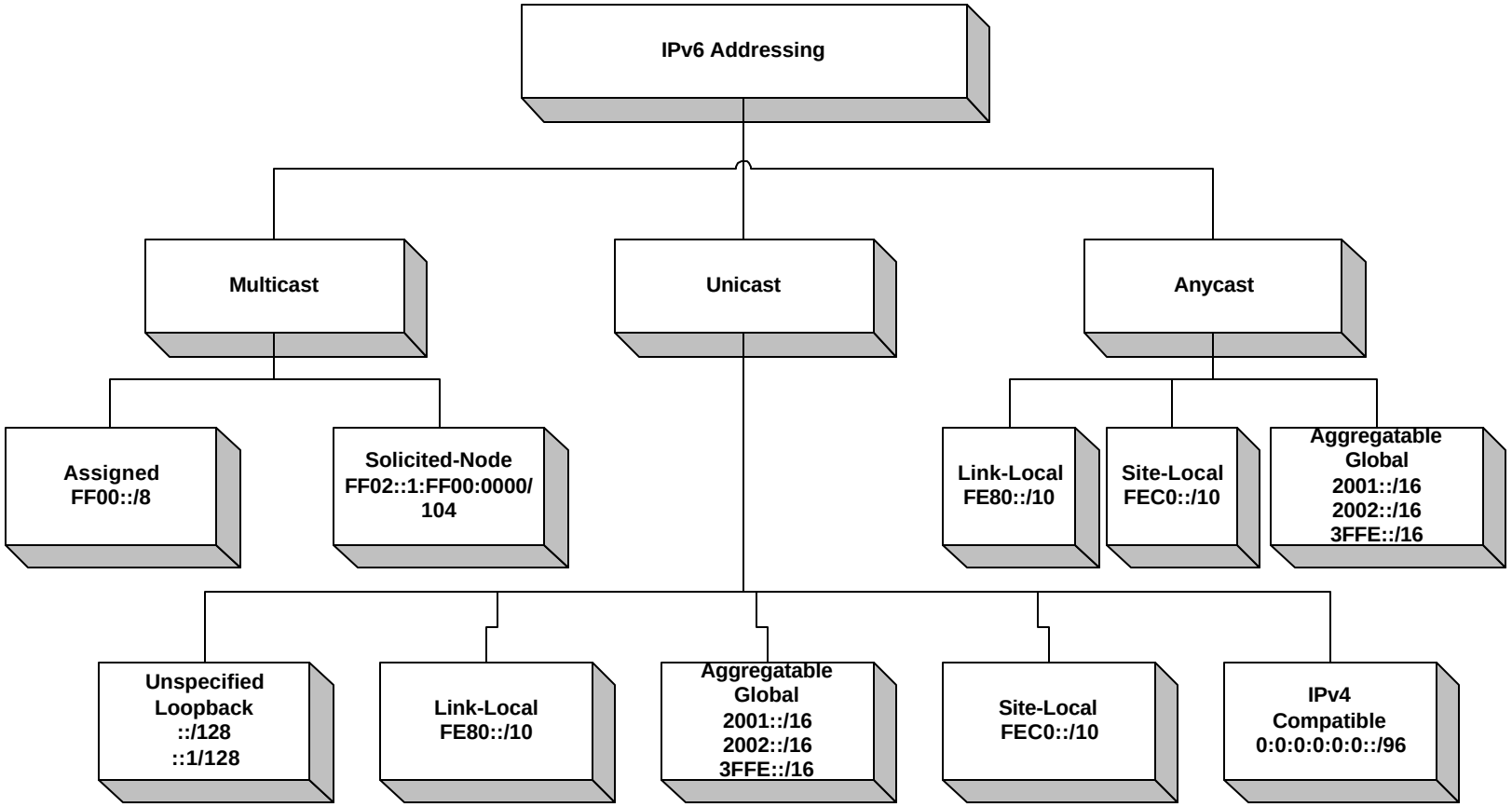
IPv6 Addressing Model (Summary)

- Addresses are assigned to interfaces
  - Interface have multiple addresses (change from IPv4)
- Addresses have scope
  - Link Local
  - Site Local
  - Global



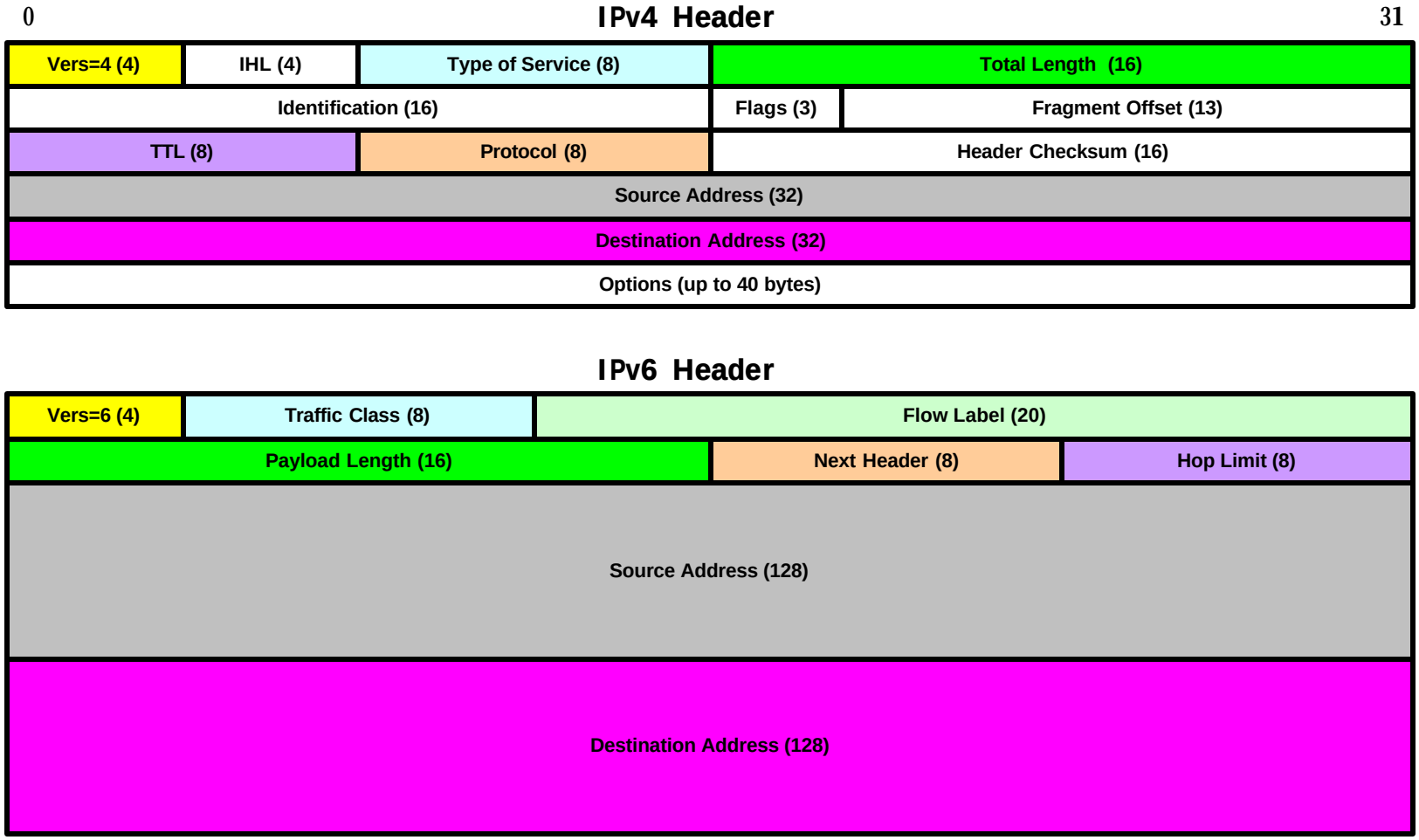
- Addresses are formed through the combination of:
  - Routing Prefix: where are you connected to?
  - Interface ID: where you are?
- Addresses have lifetime
  - Valid and Preferred lifetime

IPv6 Addressing Architecture



IPv6 Header
&
Extensions

IPv4 vs. IPv6 Headers



What's New

- Revised fields
  - Payload length (extension + data) vs. Total length (header + data)
    - 16-bit (64K bytes), use jumbogram option if > 64K
  - Next headers (Extension Header or Transport) vs. Protocol type
  - Hop Limit (no time concept) vs. TTL (Time To Live)
    - Maximum number of links over which the IPv6 packet can travel before being discarded.
- New fields
  - Traffic Class (vs. TOS)
    - To support differentiated services (e.g., prioritized best effort queuing)
    - Values are not defined by RFC. It is up to applications.
  - Flow Label
    - A flow is a sequence of packets sent from a particular source to a particular (unicast or multicast) destination for which the source desires special handling by the intervening routers

IPv6 Support for Int-Serv

- 20-bit “Flow Label” field to identify specific flows needing special QoS
  - Each source chooses its own Flow Label values
  - A flow label is assigned to a flow by the flow's source node. New flow labels must be chosen randomly and uniformly from the range 1 to FFFFF hex.
  - Flow label value of 0 used when no special QoS requested (the common case today)
  - Mutable (changeable) field for IPSec
  - Routers use “source address” + “destination address” + non-zero “flow label” to identify distinct flows
  - The Flow Label field suitable for use as a hash key by routers, for looking up the state associated with the flow.
  - Stateful architecture
  - Use RSVP signaling
- RFC 3697 (March 2004) – IPv6 Flow Label Spec.

IPv6 Support for Diff-Serv

- 8-bit “Traffic Class” field to identify specific classes of packets needing special QoS
  - Same as new definition of IPv4 “Type of Service” byte
  - May be initialized by source or by router enroute
  - May be rewritten by routers enroute
  - Stateless architecture
  - Mutable field for IPSec (ICV calculation)
  - Traffic Class value of 0 used when no special QoS requested (the common case today)
- RFC 2474: Definition of the Differentiated Services Field
- RFC 2475: An Architecture for Differentiated Service

What's Gone?

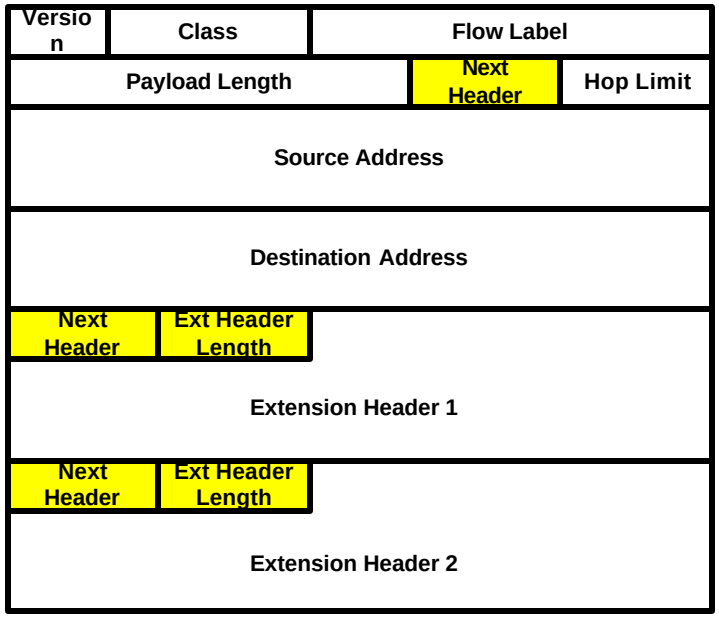
- Why "IHL" is gone?
  - In IPv4, the "Options" length is various (up to 40 bytes)
  - In IPv6, the header is fixed (40 bytes), the length of the option header(s) is included in payload length.
- Why "Header Checksum" is gone?
  - In IPv6, the link layer performs bit-level error detection for the entire IPv6 packet.
  - TCP, UDP, ICMPv6 checksum calculation becomes mandatory.
- Why "Fragment Offset", "ID" and "Flags" are gone?
  - In IPv4, all of these fields are used for fragmentation.
  - In IPv6, the fragmentation is contained in Fragment extension header.
  - In IPv6, routers no longer perform the fragmentation
  - Fragmentation was considered CPU intensive processing

IPv6 Extension Headers

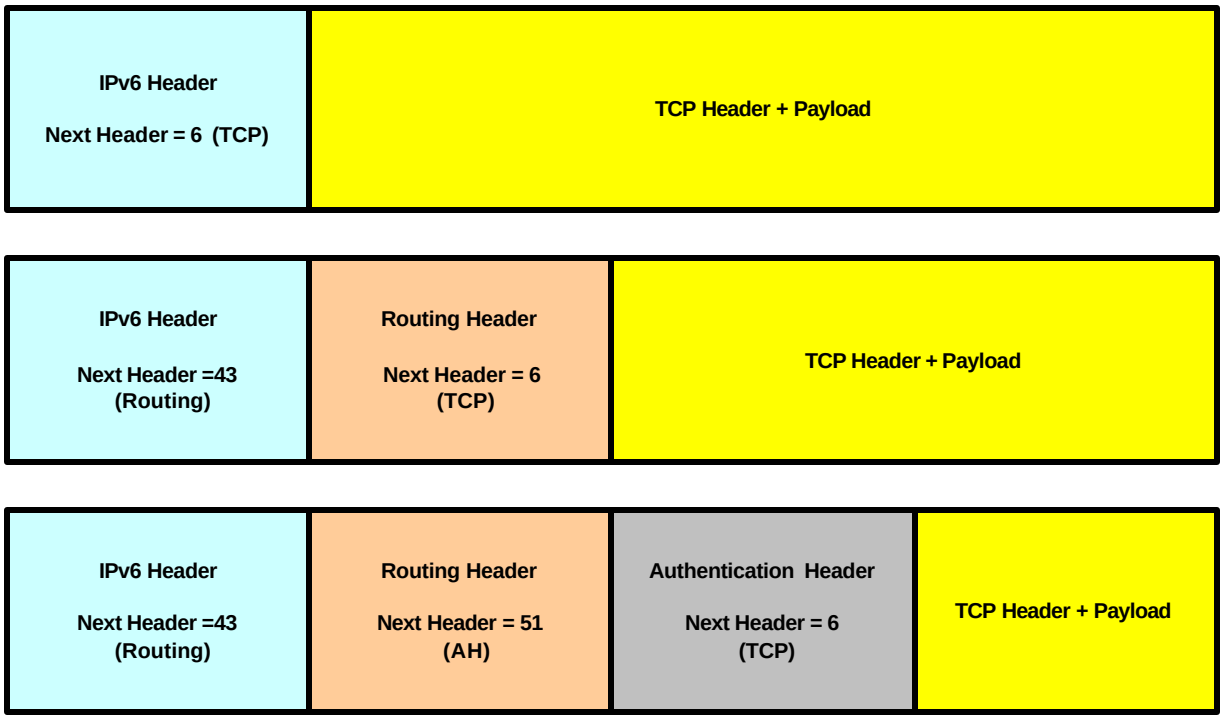
- IPv6 options are placed in separate extension headers (daisy-chained)
- The last header of a chain is the upper-layer protocol (e.g., TCP, UDP, ICMPv6) carrying the payload
- Will not limited by option length (IPv4), extensibility.
- Most IPv6 extension headers are not examined or processed by any router along a packet's delivery path until it arrives at its final destination. (except Hop-by-Hop, Destination (intermediate), Routing options).
- IPv6 extension headers can be of arbitrary length.
- IPv6 options are always an integer multiple of 8 bytes long to retain this alignment for subsequent headers. (to improve performance).
- The order of the IPv6 extension headers are important.
- RFC 2460 describes the IPv6 Extension Header formats
- RFC 3452 describes how to program extension headers via socket API.

Next Header Field

Value	Header
0	Hop-by-Hop Options Header
6	TCP
17	UDP
41	Encapsulated IPv6 Header
43	Routing Header
44	Fragment Header
50	Encapsulating Security Payload Header
51	Authentication Header
58	ICMPv6
59	No Next Header
60	Destination Options Header



IPv6 Extensions Example



Benefits of IPv6 Extensions

- IPv4 options drawbacks
  - IPv4 options required special treatment in routers
  - Options had negative impact on router's forwarding performance, therefore rarely used
  - No way to extend (limited size of option field)

- Benefits of IPv6 extension headers
  - Extension headers are external to IPv6 header
  - Routers do not look at these options except for Hop-by-Hop, Routing options etc.
  - No negative impact on router's forwarding performance
  - Easy to extend with new headers and options

IPv6 Extension Order

1. **Hop-by-Hop Options header (0)**
2. **Destination (intermediate destination) Options header (60)**
3. **Routing header (43)**
4. **Fragmentation header (44)**
5. **Authentication header (AH) (51)**
6. **Encapsulating Security Payload header (ESP) (50)**
7. **Destination (for the final destination) Options header (60)**

- **The Hop-by-Hop must be first because it is processed by every node on the path. The value is set to 0 because it is easier to test by router hardware. This header is used for Jumbo-gram packets and the Router Alert (e.g., RSVP)**
- **Destination can appear twice, the first occurs before Routing header, the second occurs before upper-layer header. This header can be use to exchange registration messages between MN and HA for mobile IP.**
- **No next header (59)**

Hop-by-Hop Options Header

- **Jumbo Payload Option (option type 194)**
  - Is used to indicate a payload size > 65535 bytes
  - Up to 4G bytes (32-bit) payload size
  - More efficient transfers with fewer interrupts
  - Only good at link MTU > 65535 + 40 bytes (IPv6 header)
  - RFC 2675
- **Router Alert Option (option type 5)**
  - Is used to indicate to a router that the contents of the packet require additional processing
  - E.g., RSVP processing etc.

Destination Options Header

- If the Destination Option Header is present right before the Routing header, it should be processed by the intermediate destination nodes (routers). Otherwise, it should by processed by the final destination node.
- Binding Update Option (option type = 198)
  - Is used by a MN to update another node with its new CoA
- Binding Acknowledgement Option (option type = 7)
  - Is used to acknowledge the receipt of a binding update
- Binding Request Option (option type = 8)
  - Is used to request the binding from a MN
- Home Address Option (option type = 201)
  - Is used to indicate to the home address of the MN.

- The Mobile IPv6 uses this option to exchange registration messages between MN and HA
- We will cover this header in the IPv6 Mobility

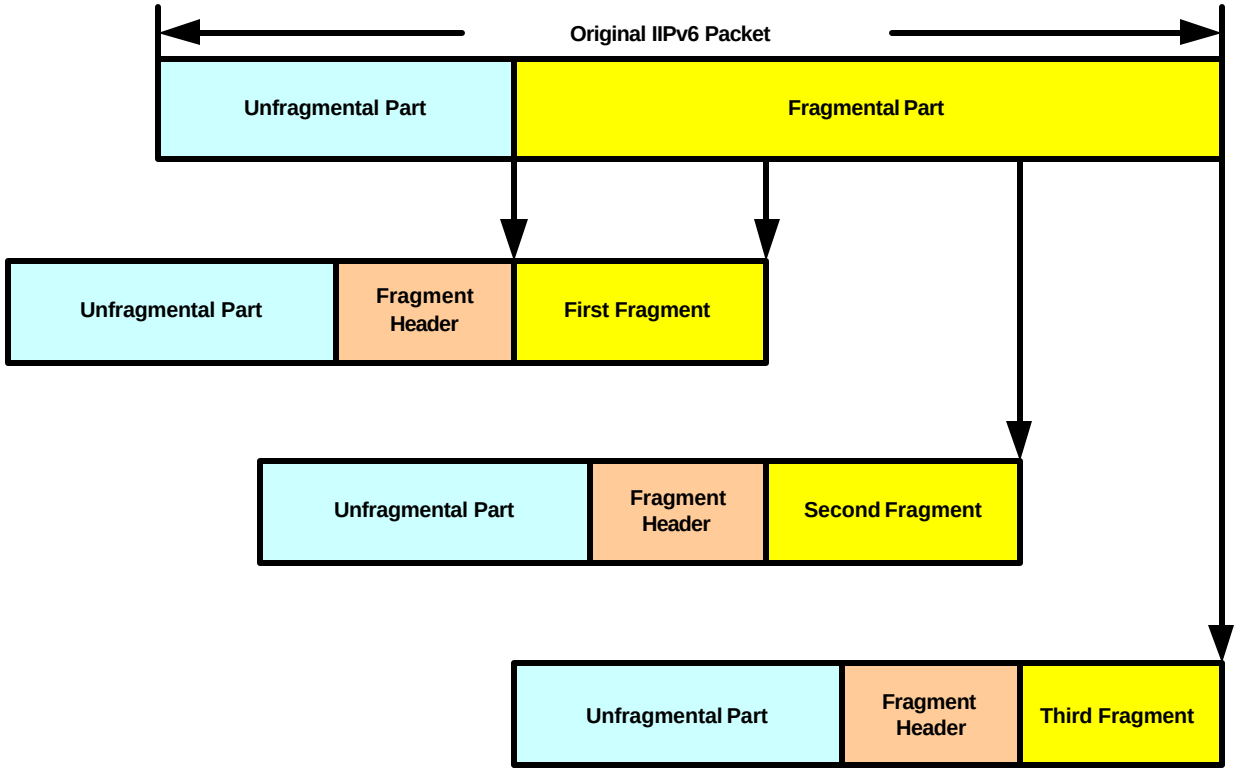
Routing Header

- **Routing header can be used by source node to force a packet to pass through specific routers (i.e., source-routing) on the way to its destination. This is similar to the "Loose Source Route" option in IPv4.**
- **Mobile IPv6 is an example that uses the Routing header (for efficiency) when a node is away from its home network.**
- **Use "Segment Left" field and "Routing Type Specific Data" to control the path.**
- **When Routing Header is processed by an intermediate destination as:**
  - **Routers that are not mentioned in the source route simply forward the packet without processing the Routing Header**
  - **The current destination address and the address in the (N – Segment Left + 1) position in the address list are swapped, where N is total number of addresses in the list.**
  - **The "segment left" is decremented**
  - **The packet is forwarded**
  - **When the packet arrives at final destination, the "segment left" is 0.**

Fragment Header

- In IPv6, only the source nodes can fragment payload. Therefore, this header is only used by source when packet is fragmented.
- Each fragmented packet has a Fragment Header.
- “Fragment Offset” and “Identification” fields in the Fragment Header have the same meaning as IPv4.
- If the payload from the upper-layer protocol is > Path MTU, then IPv6 fragments the payload and uses Fragment header to provide assembly information.
- An IPv6 router will never fragment an IPv6 packet being forwarded.
- The unfragmentable part of the original IPv6 packet must be processed by the intermediate nodes.
- This part consists of IPv6 header, Hop-by-Hop Options header, Destination header (intermediate), Routing header.

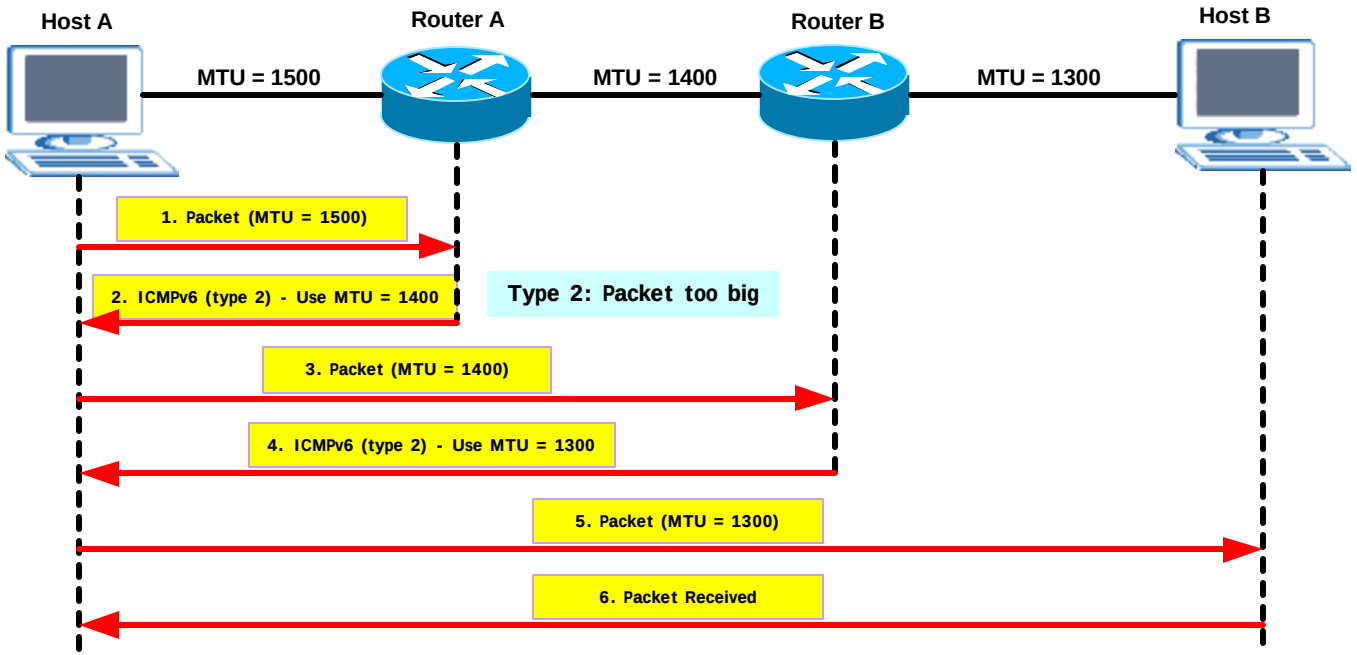
IPv6 Fragmentation

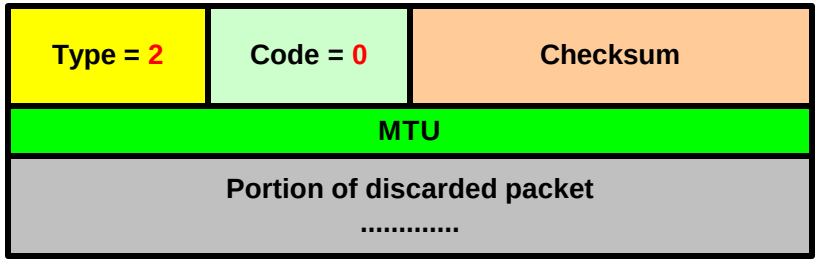


IPv6 Path MTU Discovery

- IPv6 routers do not fragment packets.
- Link MTU: max packet length can be transmitted on a given link without fragmentation
- PMTU: The smallest link MTU between two nodes across multiple networks. PMTU = min(LMTU) of a given path.
- If the upper-layer sends packets > PMTU, IPv6 sender needs to fragment them to PMTU size (not recommended by spec.)
- In order to avoid the processing of fragmentation and reassembly, the sending nodes should perform the Path MTU Discovery (more efficient).
- The Path MTU Discovery is achieved via the ICMPv6 Packet Too Big (type 2) message (contains link MTU of the interface) LMTU < PMTU
- IPv4 uses ICMP Destination Unreachable Error (Type 3) with Fragmentation needed but DF bit set (Code 4) approach.
- In IPv4, the minimum link MTU size is 68 bytes. IPv6 requires the link layer support a minimum size of 1280 bytes (excluding link layer header and trailer). Hosts do not support PMTU discovery option should send 1280B.

IPv6 Path MTU Discovery (2)

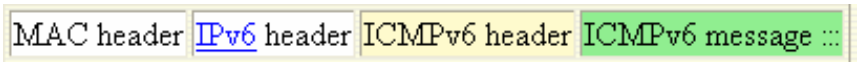


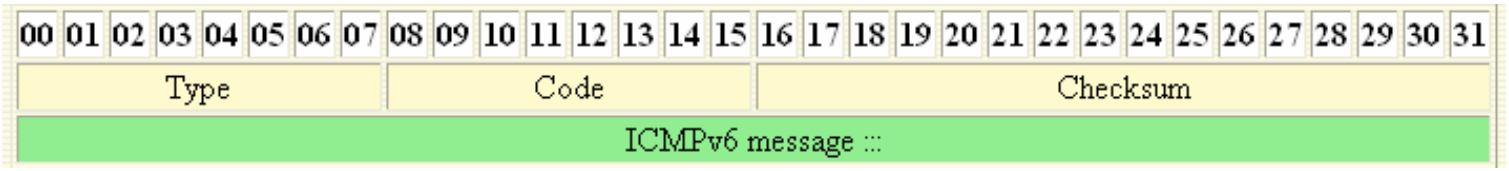


ICMPv6

ICMPv6 (RFC2463)

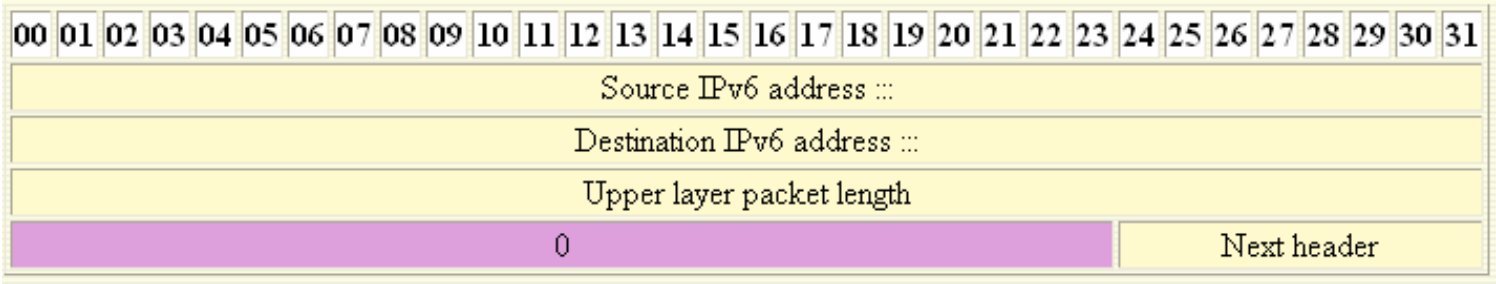
- IPv6 Header Next Header: 58
- Based on IPv4 ICMP (same header format) with a number of changes (message types)
- Type (8-bit): specify the ICMP message type
- Code (8-bit): further qualify ICMP message





ICMPv6 (2)

- **Checksum (16-bit): This field contains the 16-bit one's complement of the one's complement sum of the entire ICMPv6 message starting with the ICMPv6 message type field, prepended with a pseudo-header of IPv6 header fields (double-check that data has arrived at the correct destination).**



ICMPv6 Messages

- **MLD (RFC 2710)**
  - **Multicast Listener Query – ICMP type 130**
  - **Multicast Listener Report (join) – ICMPv6 type 131**
  - **Multicast Listener Done (leave) – ICMPv6 type 132**
- **Mobile IPv6 (RFC 3775)**

Type	Description	References
0		
1	Destination unreachable.	RFC 2463
2	Packet too big.	RFC 2463
3	Time exceeded.	RFC 2463
4	Parameter problem.	RFC 2463
5 - 127		
128	Echo request.	RFC 2463
129	Echo reply.	RFC 2463
130	Group Membership Query.	
131	Group Membership Report.	
132	Group Membership Reduction.	

Type	Description	References
133	Router Solicitation.	RFC 2461
134	Router Advertisement.	RFC 2461
135	Neighbor Solicitation.	RFC 2461
136	Neighbor Advertisement.	RFC 2461
137	Redirect.	RFC 2461
138	Router Renumbering.	RFC 2894
139	ICMP Node Information Query.	
140	ICMP Node Information Response.	
141	Inverse Neighbor Discovery Solicitation Message.	RFC 3122
142	Inverse Neighbor Discovery Advertisement Message.	RFC 3122
143	Home Agent Address Discovery Request Message.	
144	Home Agent Address Discovery Reply Message.	
145	Mobile Prefix Solicitation.	
146	Mobile Prefix Advertisement.	

Neighbor Discovery Protocol (RFC 2461)

Neighbor Discovery Goals

- IPv6 nodes which share the same physical medium (link) use Neighbor Discovery Protocol (NDP) to:
  - Address Resolution
    - Resolve a neighbor's IPv6 address to its link-layer (MAC) address.
    - It is equivalent to ARP in IPv4.
  - Duplicate Address Detection (DAD)
    - Determines that an address for use is not already in use by a neighbor node.
    - It is equivalent to Gratuitous ARP frames in IPv4.
  - Neighbor Unreachable Detection (NUD)
    - Determines that the IPv6 layer of a neighbor is no longer receiving packets.
    - Might not be the final destination but the reachability of the first hop of the destination.

Neighbor Discovery Goals (2)

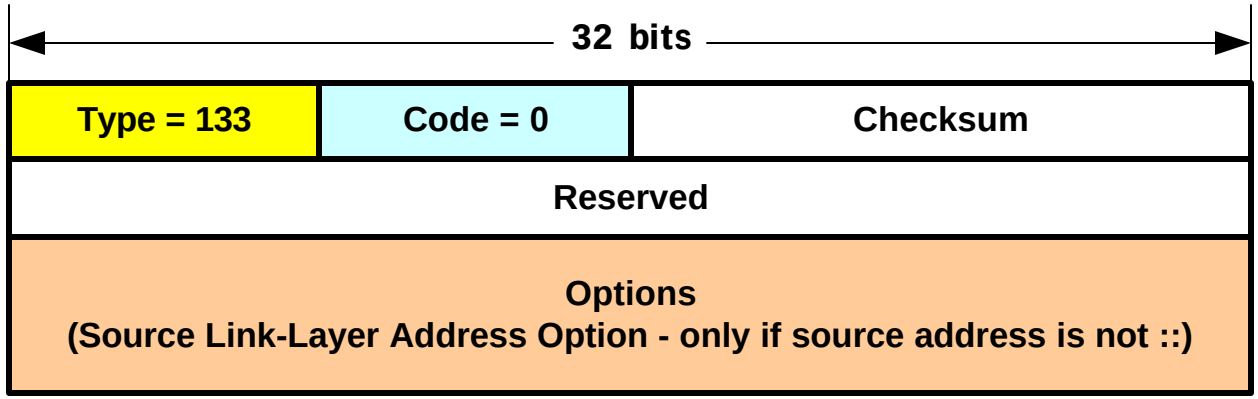
- Router Discovery
  - A host discovers the local router(s) on an attached link.
  - Determine which local router is a default gateway
  - Switch to backup default router if the primary one is unavailable
    - Router Lifetime expiration
    - Neighbor Unreachability Detection (NUD)
  - Network Prefix(es) discovery
  - Parameters discovery (link MTU, Max Hop Limit, auto-config …)
  - It is equivalent to ICMPv4 Router Discovery.
- Next-hop Determination (based on destination address in the packet)
  - From destination cache (contains Next-hop address)
  - From prefix list (contains a range of IP addresses for destinations)
  - From default router list (contains default router IP addresses)
- Redirect Function
  - Process of a router inform a host of a better first-hop IPv6 address to reach a destination
  - It is equivalent to ICMPv4 Redirect Message

Neighbor Discovery Messages

- To achieve the goals, the following ICMPv6 messages are defined:
  - Router Solicitation – RS (ICMPv6 Type 133)
    - Sent by host who needs RA immediately (at boot time)
  - Router Advertisement – RA (ICMPv6 Type 134)
    - Sent by router periodically or in response to RS
  - Neighbor Solicitation – NS (ICMPv6 Type 135)
    - Sent by host & router to determine link layer address, DAD, NUD etc.
  - Neighbor Advertisement – NA (ICMPv6 Type 136)
    - Sent by host & router in response to NS or advertise the change of link address
  - Redirect (ICMPv6 Type 137)
    - Sent by router for a better route

Router Solicitation

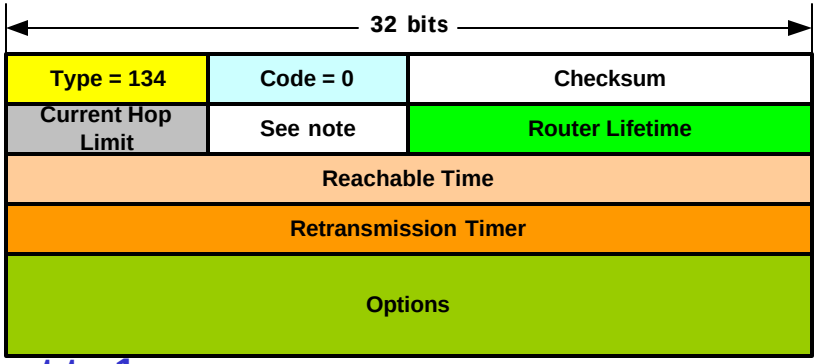
- Sent by host to speed up learning of link-local routers rather than waiting for unsolicited RA periodically.
- Destination address is all-routers multicast address (FF02::2)



Router Advertisement

- Sent by routers periodically or in response to a RS to provide information necessary for a node to configure itself.
- Destination address
  - Unicast address of a node that sent an RS, or
  - Link-scope all-nodes multicast address (FF02::1)
- Possible options:
  - Source link-layer address (of this router)
  - MTU (of this link)
  - Prefix information (for auto-configuration)
  - Advertisement Interval (the interval of unsolicited RA sent by this router)
  - Home Agent Information (preference and lifetime of home agent)
  - Route Information (routes to be added to local routing table for efficiency)

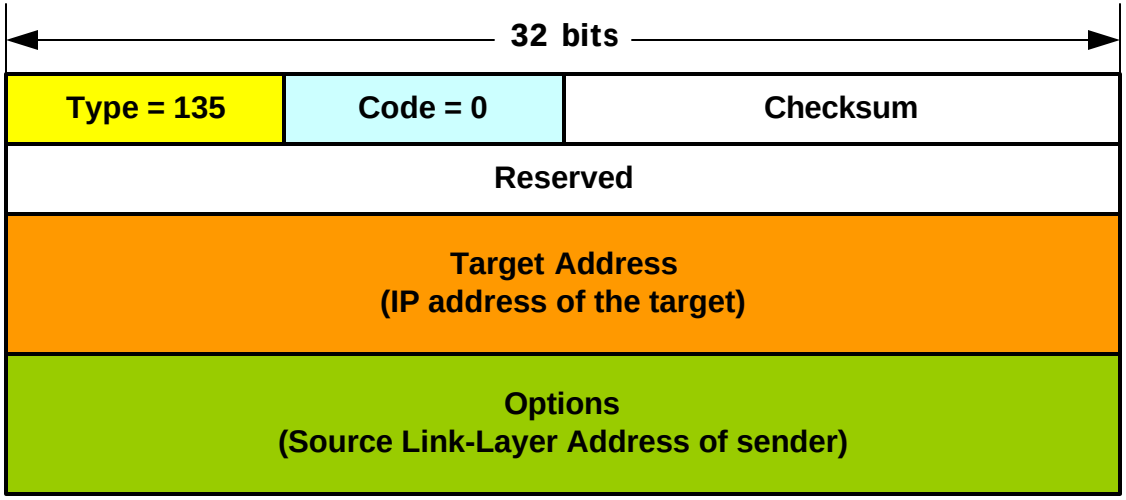
Router Advertisement Format



- Current Hop Limit
  Default hop limit for packets sent by host
- Managed Address Configuration (M) Flag (1-bit)
  Use DHCPv6 to configure address if set to 1
- Other Stateful Configuration (O) Flag (1-bit)
  Use DHCPv6 to obtain non-address information if set to 1
- Home Agent (H) Flag (1-bit)
  The router is also functioning as a home agent for IPv6 Mobility if set to 1
- Default Router Preference (2-bit)
  The level of preference for this router as the default router in a multiple routers network
- Router Lifetime
  The lifetime (in seconds) of the router as the default router
- Reachable Time
  The amount of time (in ms) that a node can consider reachable after receiving confirmation
- Retransmission Timer
  The amount of time (in ms) between retransmission of NS for neighbor unreachability detection

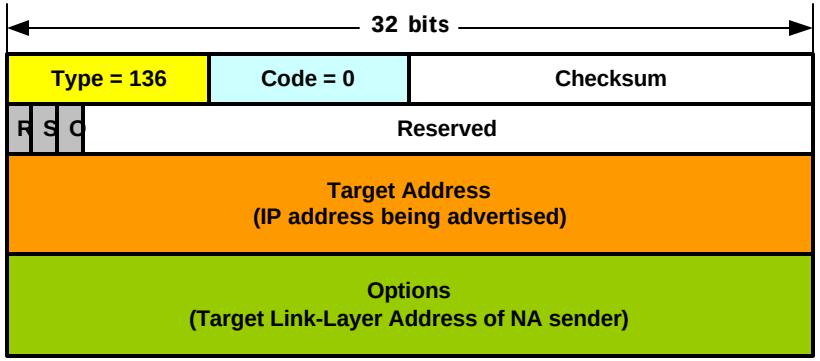
Neighbor Solicitation

- Used to provide/obtain link-layer address to/of a neighbor (DAD/address resolution)
- Used to verify neighbor reachability (NUD)
- Destination address
  - Solicited-node multicast address of itself or target (DAD/address resolution)
  - Unicast address of the target (reachability verification)



Neighbor Advertisement

- Sent in response to NS or unsolicited to immediately propagate new information
- Destination address
  - For solicited advertisements
    - Source address of the solicitation
    - If solicitation's address is unspecified:
      - all-nodes multicast address (ff02::1)
  - For unsolicited advertisements
    - All-nodes multicast (ff02::1)



- Router flag – sender is a router if set to 1
- Solicited flag – in response to NS if set to 1
- Override flag – override the link-layer address in neighbor cache (~ARP cache) with the link-layer address in Target link-layer address option

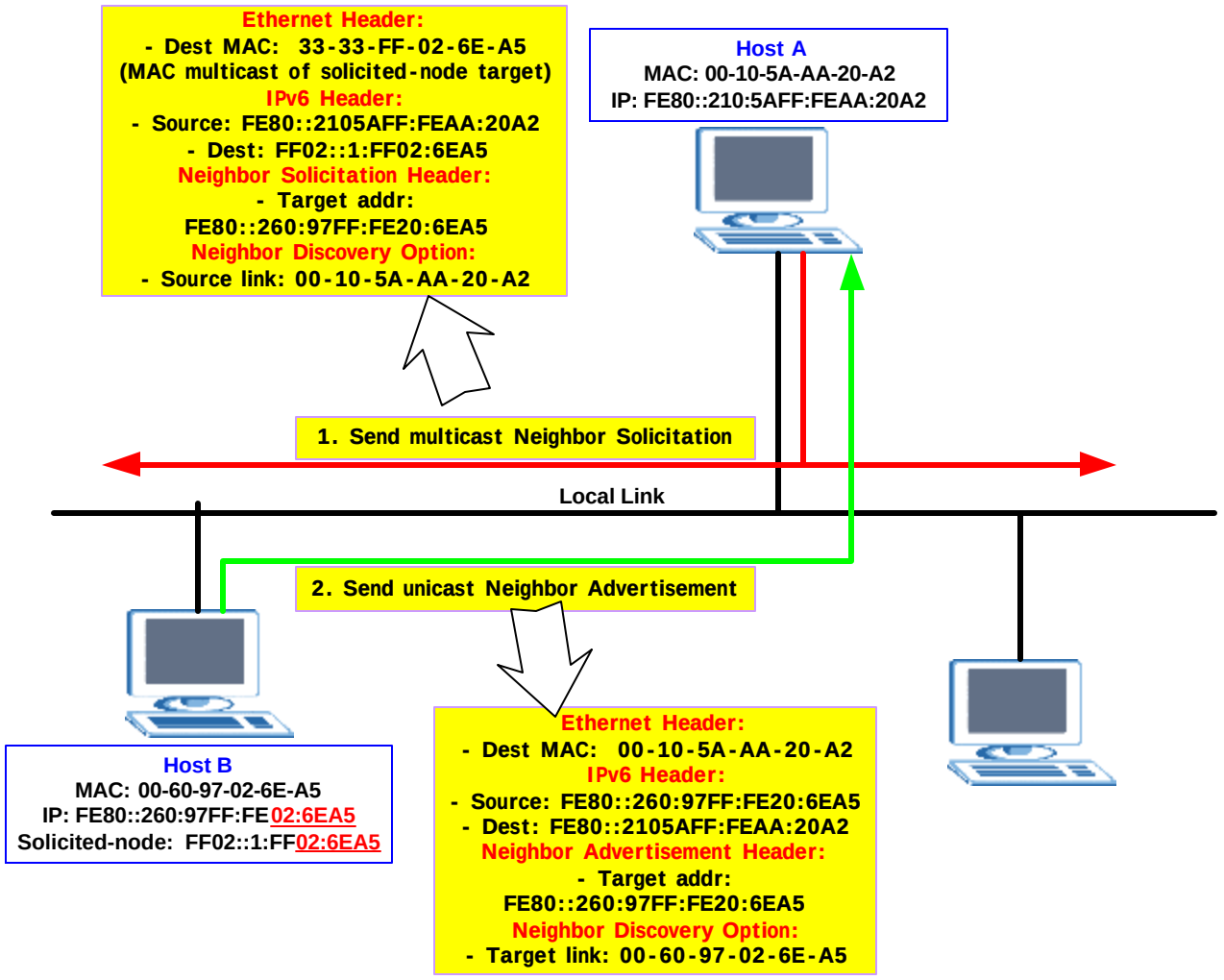
Address Resolution

- Replace IPv4 ARP
- To resolve the link-layer address of the on-link next-hop address for a given destination.
- Accomplished by using Neighbor Solicitation (multicast) and Neighbor Advertisement messages.
- Node checks Neighbor Cache (contains IP address, MAC address, reachable state) first.
- If no entry exists, node creates IP entry with state INCOMPLETE.
- Node then sends NS to solicited-node multicast address
  - Instead of using local-link scope all-nodes multicast (FF02::1), every node listens on solicited-node multicast address (FF02::1:FF00:0/104 + last 24-bit of unicast IPv6 address).
  - The destination address is the solicited-node multicast address derived from the target IP address.
  - Source address of NS is the sender's unicast address.

Address Resolution (2)

- Receiving node responds with NA indicating it's own link-layer address
- After receiving the NA from the target, the sending host updates its neighbor cache for the link-layer address from the Target Link-Layer Address option.
- Sending node updates Neighbor Cache entry from INCOMPLETE to REACHABLE state upon reception of NA.
- RFC 2461 – Neighbor Discovery for IPv6 (describes the conceptual data structures).
  - Destination Cache
  - Neighbor Cache
  - Prefix List
  - Default Router List

Address Resolution Scenario



Duplicate Address Detection

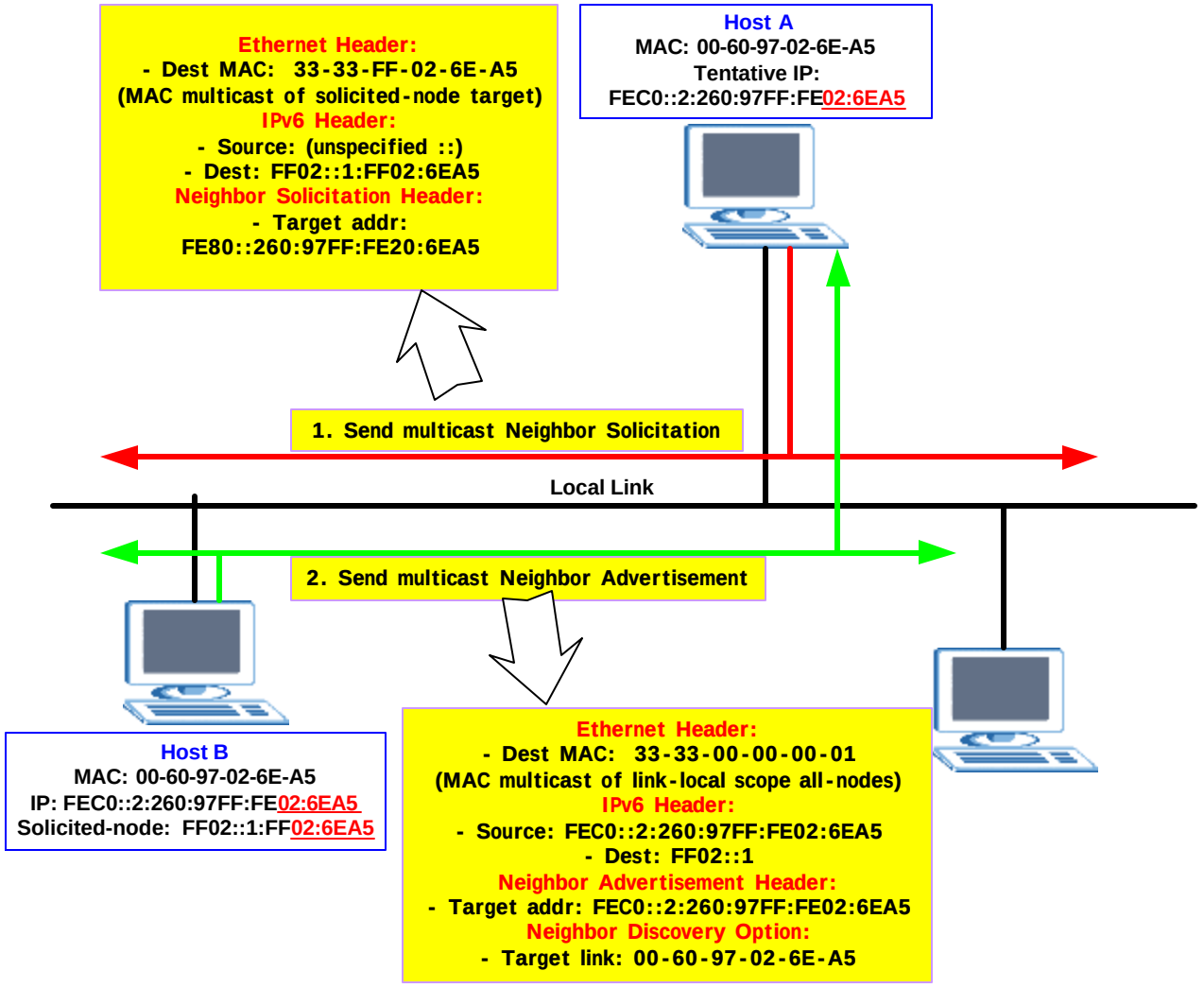
- Replace IPv4 ARP request and Gratuitous ARP
- What is Gratuitous ARP?
  - IPv4 sets both source and target with the same IP address of the sender in ARP request.
  - If it receives ARP reply, then the IP address is duplicate.

- Must be performed by all nodes (hosts & routers).
- Performed before assigning a unicast address to an Interface.
- Performed on interface initialization.
- Not performed for anycast address.
- Link must be multicast capable.

Duplicate Address Detection (2)

- Accomplished by using Neighbor Solicitation (multicast) and Neighbor Advertisement messages.
- Node sends NS with:
  - Source address is unspecific address (::)
  - Destination address is tentative solicited-node address
  - Target address field is set to tentative IP address
  - The Source Link-layer Address option is not used.
- If address already exist, the particular node sends a NA reply with:
  - The destination address of NA is set to FF02::1.
  - The solicited flag is 0 because NS is not using the desired IP address, it cannot receive unicast NA.

DAD Scenario

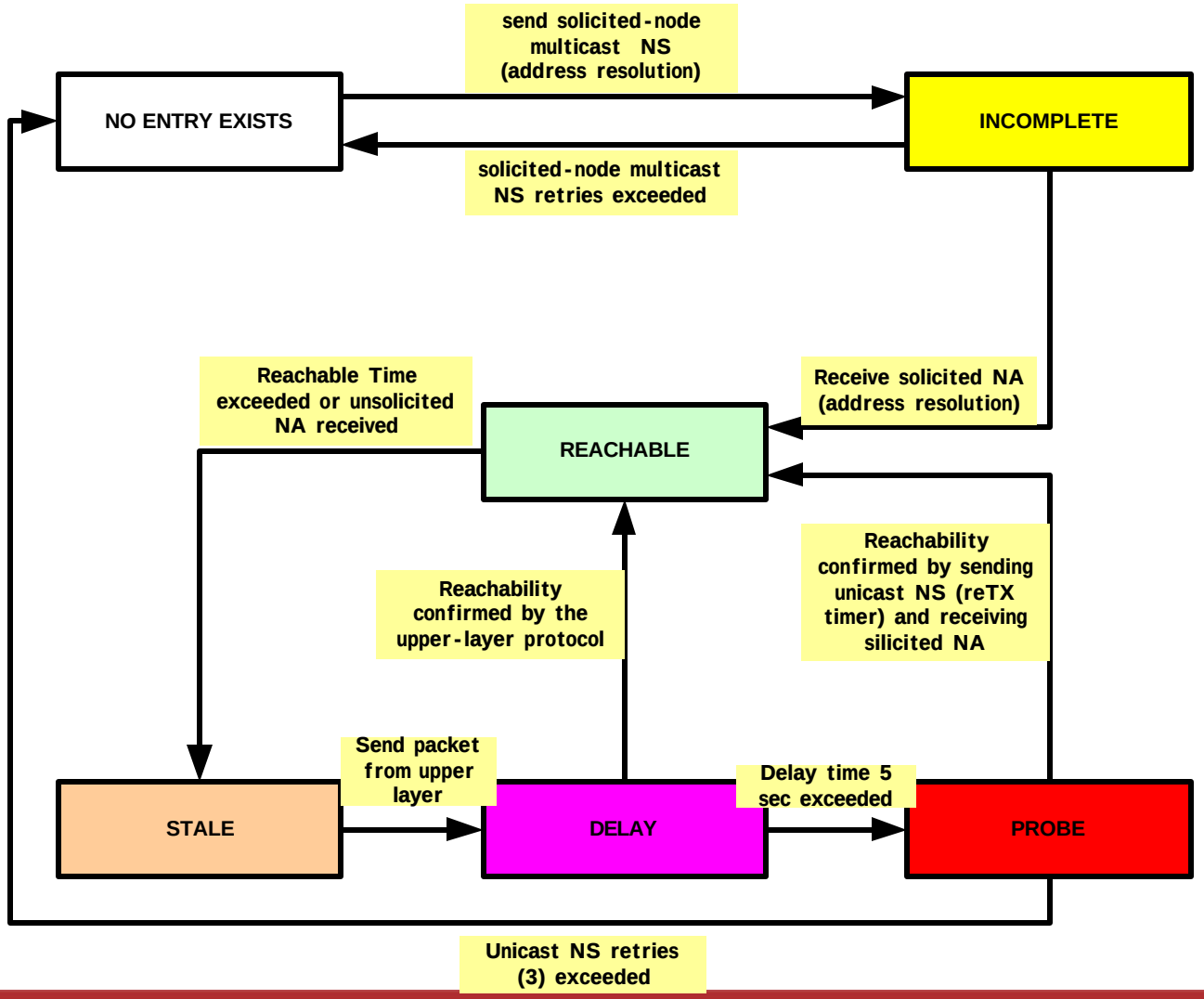


Neighbor Unreachability Detection

- Two ways to verify neighbor reachability:
  - Using hints from upper-layer protocols
  - From responses to Neighbor Solicitations

- Neighbor cache (~ ARP cache) stores information about neighbors:
  - IP address
  - Link-layer address
  - Reachability states
    - INCOMPLETE
    - REACHABLE
    - STALE
    - DELAY
    - PROBE

Neighbor Cache States



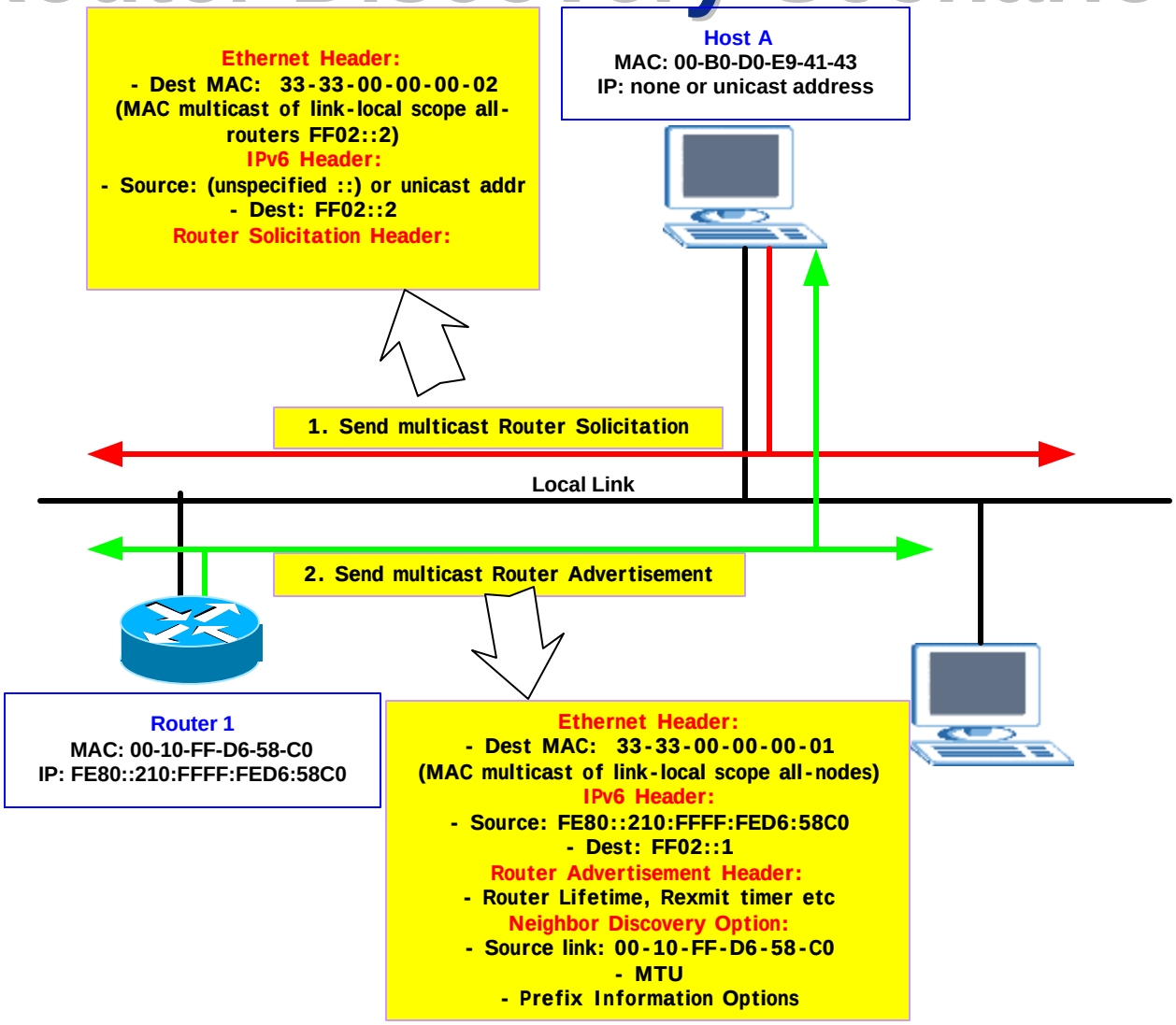
Neighbor Unreachability Detection

- Does not necessarily verify the end-to-end reachability of the destination.
- Verifies only the reachability of the first hop to the destination.
- Accomplished by sending a unicast Neighbor Solicitation and the receipt of a solicited (solicited flag = 1) Neighbor Advertisement message.

Router Discovery

- Attempts to discover the set of routers on the local links.
- Similar to IPv4 ICMP router discovery (RFC 1256).
- In IPv6 RA messages, the Router Lifetime field indicates the time that router can be considered a default router.
- Accomplished by sending a multicast Router Solicitation (FF02::2) and the receipt of a multicast Router Advertisement (FF02::1) message.
- If the router becomes unavailable, the condition is detected via neighbor unreachability detection instead of the Router Lifetime in the RA messages.
- A new default router is chosen from default router list or the host sends a RS message to determine a new default router.

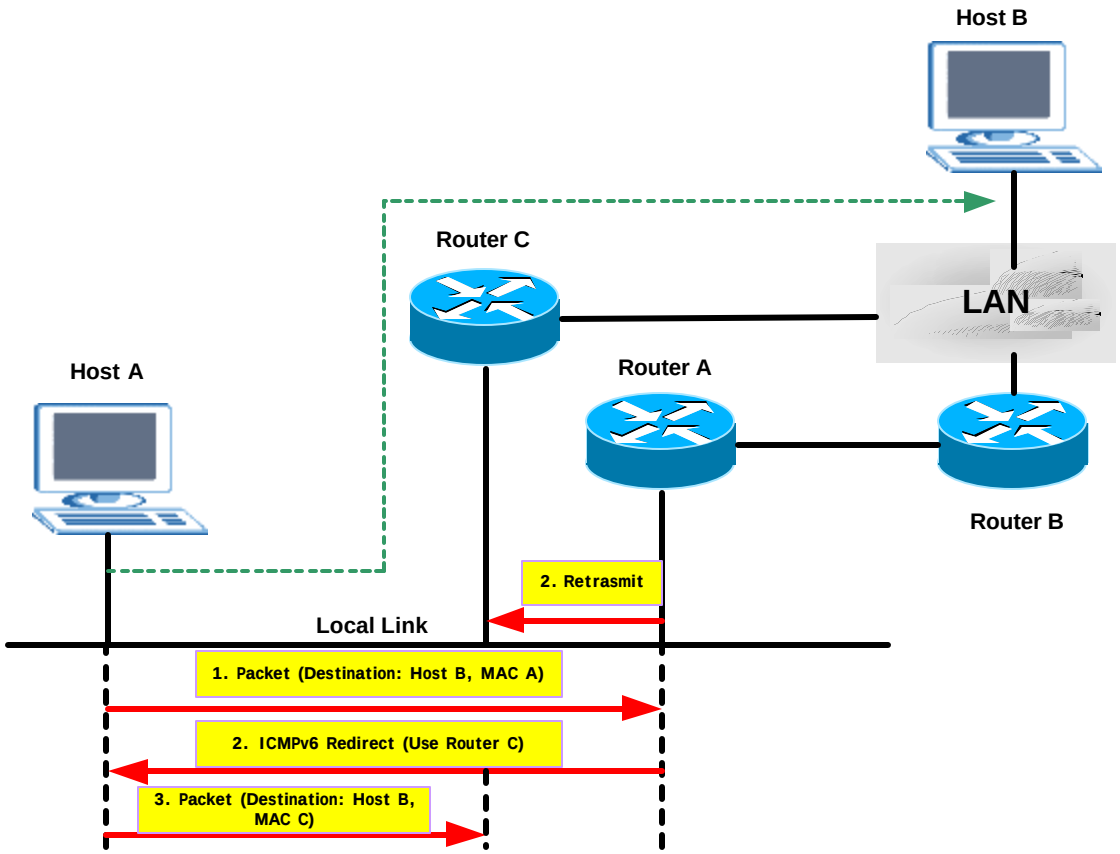
Router Discovery Scenario



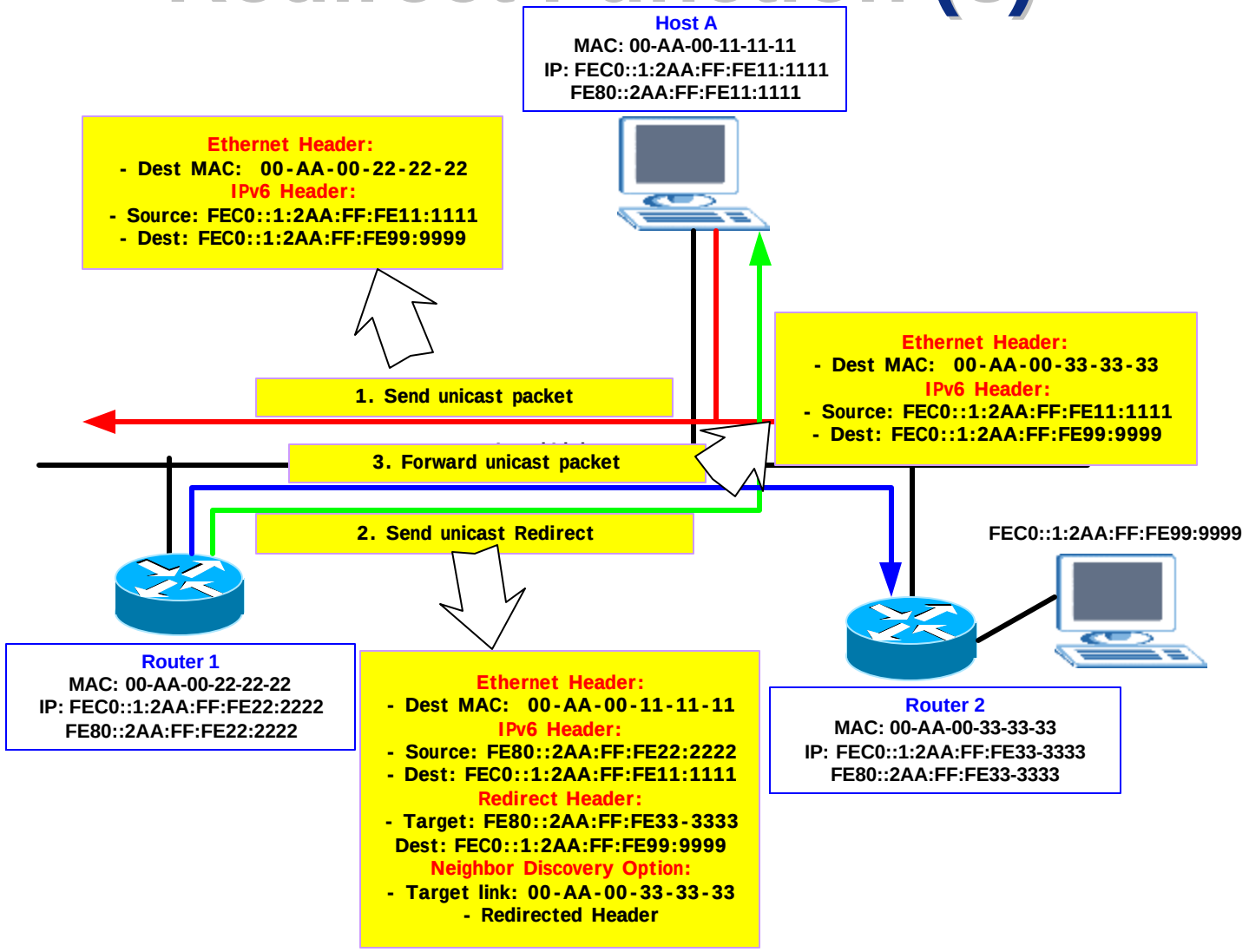
Redirect Function

- Sometimes hosts will pick the wrong next-hop
  - There are several routers
  - Send to a router although destination is connected to the same link
- The router that receives the packet
  - Will retransmit to the correct hop
  - Send a Redirect message to the sender
- Next message send to that destination travels only once to the correct router

Redirect Function (2)



Redirect Function (3)



Address Auto-configuration

Auto-configuration Goals

- Designed for hosts
  - Stateless auto-configuration does not apply to routers
  - It is assumed that routers are configured by some other mean: e.g. router renumbering
- Plug-and-Play Capability
- No manually addressing of hosts
- Generates routable addresses
- Automate network address host renumbering

Auto-configuration Mechanism

- Requires each host learn its address
- Methods for obtaining addresses:
  - Link-local address
    - No router or server required
    - Assigned by node itself followed by DAD processing
  - Stateless mechanism
    - Router Advertisements provide prefix
    - Prefix + EUI64(MAC @)
  - Stateful mechanism
    - DHCPv6 Server provides address
  - Updates to Naming Service
    - Not automatic
    - Dynamic updated to DNS via DDNS mechanism

Address Auto-configuration

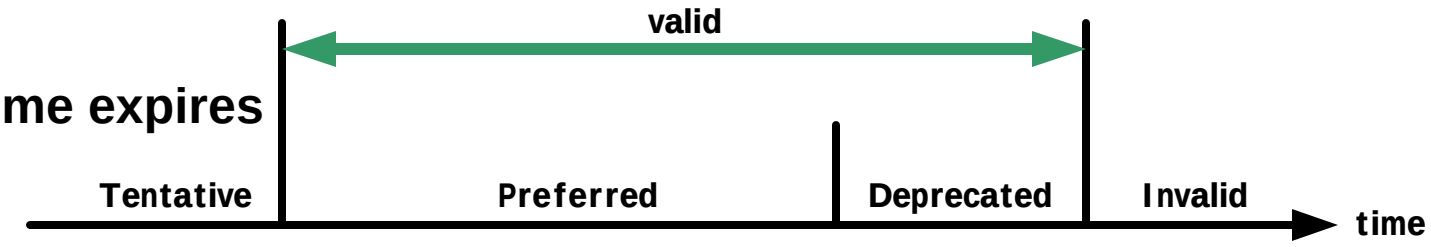
- Stateless Address Auto-configuration (RFC 2462)
  - Based on ICMPv6
    - Router Solicitation
    - Router Advertisement
  - Creation of global and site-local addresses
    - Based on the receipt of Router Advertisement messages
  - Creation of link-local addresses
    - Determine by host With Duplicate Address Detection (DAD)
    - Assumes that each interface can provide a unique ID
- Stateful Address Auto-configuration (RFC 3315)
  - Based on DHCPv6
  - Obtain network information from DHCP server
  - Servers maintain a database with
    - Client addresses pool
    - Client's state
    - Other configuration information (Prefix Delegation, DNS Update)

Stateless Auto-config Process

- Host assign link-local address
  - Interface Initialization
  - Perform DAD process, if fails then auto-configuration stops. Manually configuration required.
- Find routers
  - Wait periodic or on-demand “Router Advertisement” message
  - Host sends all-routers multicast “Router Solicitation” message
  - Router responds “Router Advertisement” message with “Network Prefix”.
  - M-bit set to 0 tells host to use stateless address auto-configure
  - O-bit set to 0 tells host to use stateless auto-configure for other parameters
- Host builds global or site-local IPv6 address from prefix
  - Using EUI-64 interface identifier (assume Ethernet)
  - Build an on-link prefix-list
  - Know the link MTU
- Finish auto-configuration

Auto-configured Address States

- Tentative
  - In the process of being verified as unique through DAD.
  - Cannot receive unicast traffic
  - Can receive multicast Neighbor Advertisement in DAD process
- Valid
  - Preferred
    - No limitation for traffic
  - Deprecated
    - Still can send and receive unicast traffic on the existing sessions but its use is discourage for new session.
- Invalid
  - Valid lifetime expires



IPv6 Address Lifetime

- Address must have a preferred and valid lifetime
  - Each RA received, restarts timer
  - Link-Local address has a infinite preferred and valid lifetime
- When the preferred lifetime expires
  - The address must be set to the Deprecated state on its interface.
- When the address enters the Deprecated state
  - Discourage to use it for new sessions
- When the valid lifetime expires
  - The address is deleted from the interface

Stateless Auto-configuration

- Creation of Link-Local Address
  - A tentative address (before DAD)
  - FE80::/64 + EUI64 Interface ID
- Duplicate Address Detection
  - For stateless, DAD is only performed on link-local address
  - Joins the all-nodes multicast
  - Joins solicited-node multicast address of the tentative address
  - Sends Neighbor Solicitation message with tentative address as the target address
  - Receive Neighbor Advertisement message
    - Too bad, address is in use cannot be assigned to interface
    - Otherwise, no response means the address is unique (how long?)

Stateless Auto-configuration (2)

- Creation of Global and Site-Local Address
  - Host sends Router Solicitation message
    - Destination: All-routers multicast address
  - Router sends Router Advertisement message
    - Destination: All-nodes multicast address
    - Prefix information
    - Associated lifetime
      - Preferred lifetime
      - Valid lifetime
  - Absence of Router Advertisement message
    - Stateful Auto-configuration (DHCPv6)

Stateful Auto-configuration

- Clients obtain IPv6 address (if M-bit = 1) and other configuration parameters (if O-bit = 1) from DHCP server (M/O bits in RA)
- DHCP server maintains the database and has a tight control over address assignments
- Client sends DHCP Solicitation to all-DHCPv6-agents (FF02::1:2)
  - M-bit or O-bit set to 1
  - No IPv6 router is found
- DHCP server replies with DHCP Advertisement
- http://sourceforge.net/projects/dhcpv6 (DHCPv6 project for Linux)

Policy of Stateless or Stateful?

- Site administrators specify which type of auto-configuration to use
  - Setting or clearing flags (M-bit and O-bit) in the ICMPv6 Router Advertisement message
  - If the IPv6 router is absent, DHCP stateful configuration is used
- Stateless and stateful can co-exist
  - Address could come from stateless (M-bit)
  - Additional configuration information could be provided by DHCPv6 (O-bit)

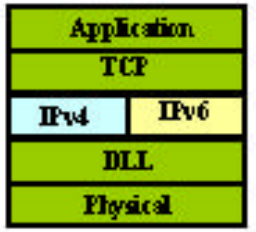
IPv6 Migration

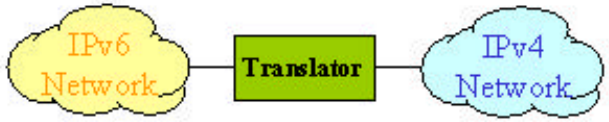
IPv6 Transition Assumptions

- No “Flag Day”
  - Last Internet transition was 1983 (NCP -> TCP, see RFC 801)
- Transition will be incremental
  - Possible over several years
- No IPv4/IPv6 barriers at any time
- No transition dependencies
  - No requirement of node X before node Y
- Must be easy for end users
  - Transition from IPv4 to dual stack must not break anything
- IPv6 is designed with transition in mind
  - Assumption of IPv4/IPv6 coexistence
- Many different transition technologies

IPv6 Migration Mechanisms

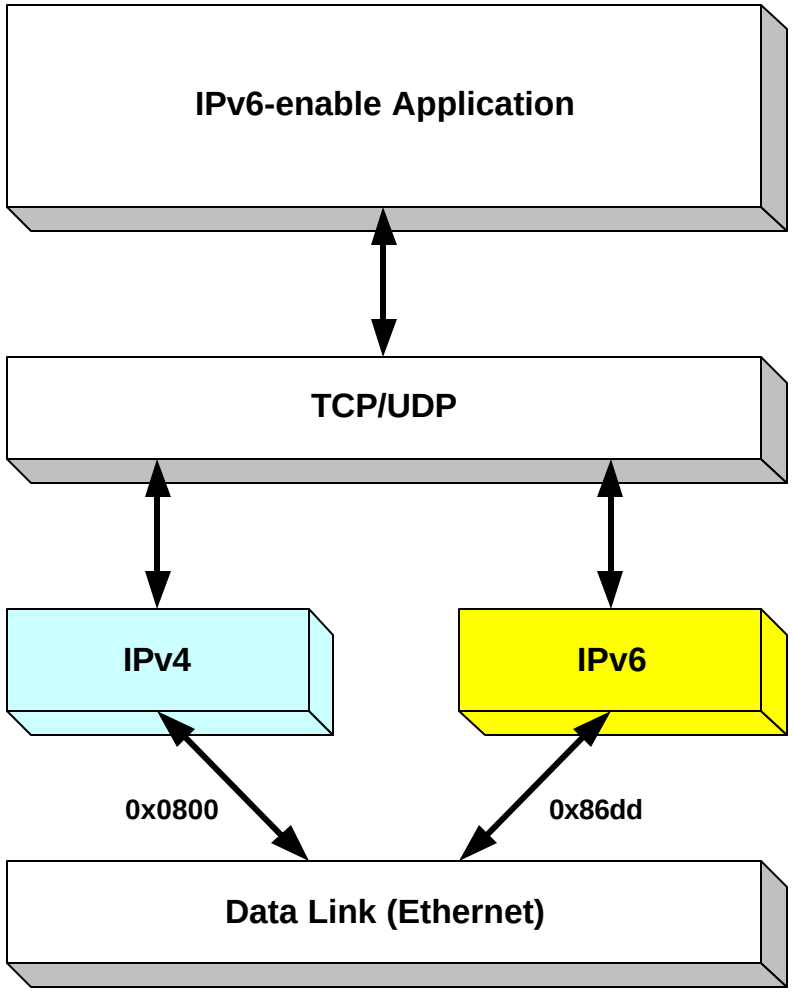
- **Dual-Stack**
  - A node will be able to understand both IPv4 and IPv6 protocols.
  - IPv4 stack and IPv6 stack coexist on one device.
- **Tunneling**
  - Allows making a link between two IPv6 networks via IPv4 network and vice-versa.
  - IPv6 <-> IPv4 <-> IPv6
  - IPv4 <-> IPv6 <-> IPv4
- **Translation**
  - Allows IPv6-only devices to communicate with IPv4-only devices and vice-versa.
  - IPv6 <-> IPv4





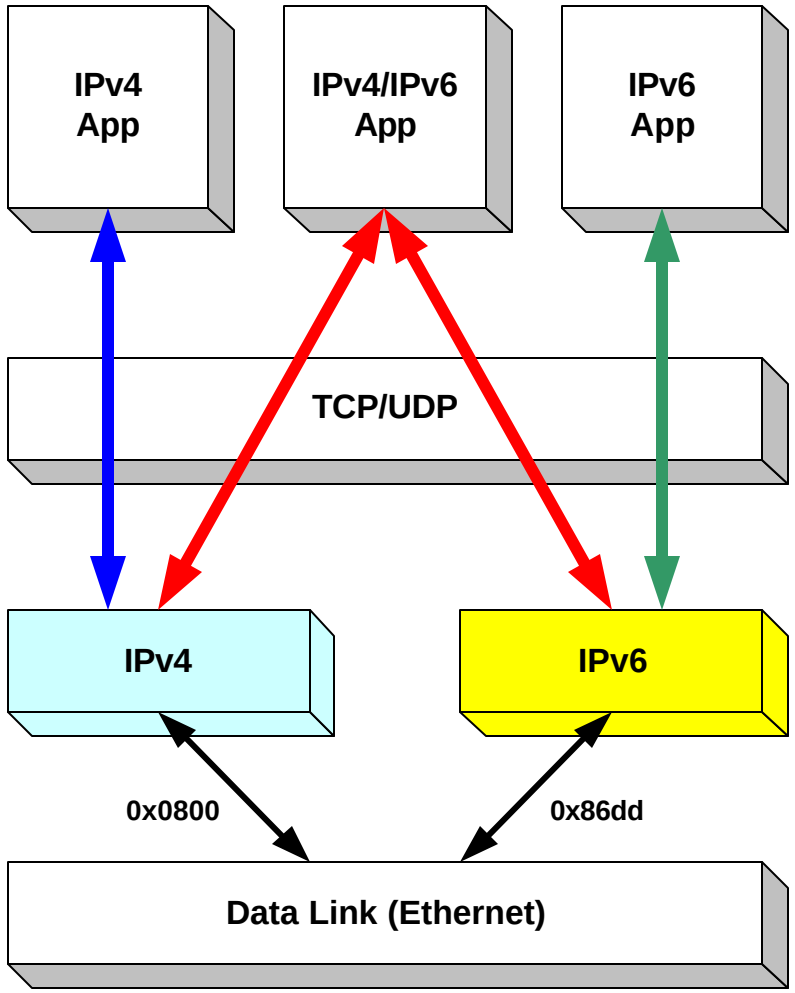
Dual-IP Layer Technique

- Only IP layer is dual, not the whole stack
- Both IPv4 and IPv6 stacks enabled
- Applications can talk to both
- Choice of the IP version is based on name lookup and application preference
- Dual-stacked routers and hosts if both IPv4 and IPv6 addresses configured at the interface(s)



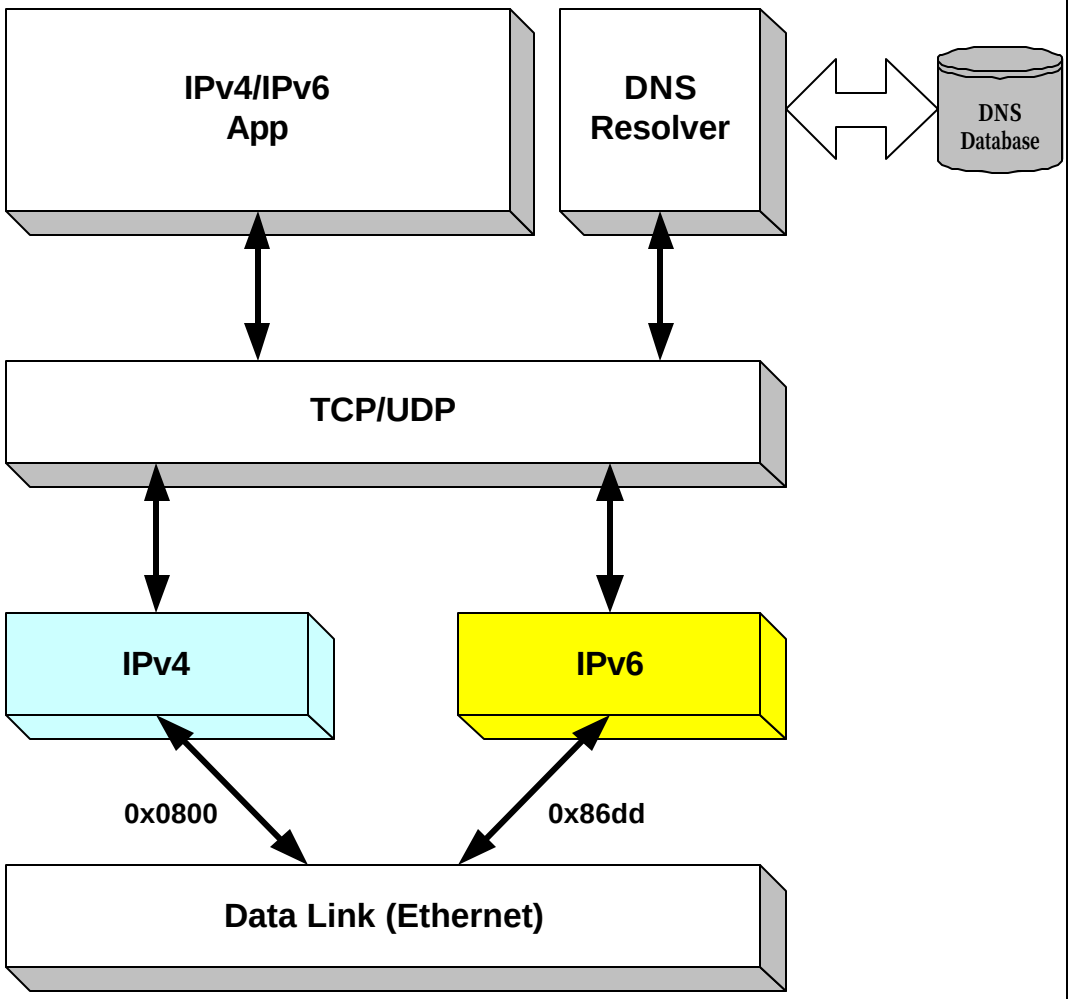
Dual-Stack Applications

- Existing IPv4 applications continue to run
- IPv6 applications can be introduced
- Interoperation of IPv4 and IPv6 is another issue
  - Applications to be modified to handle both



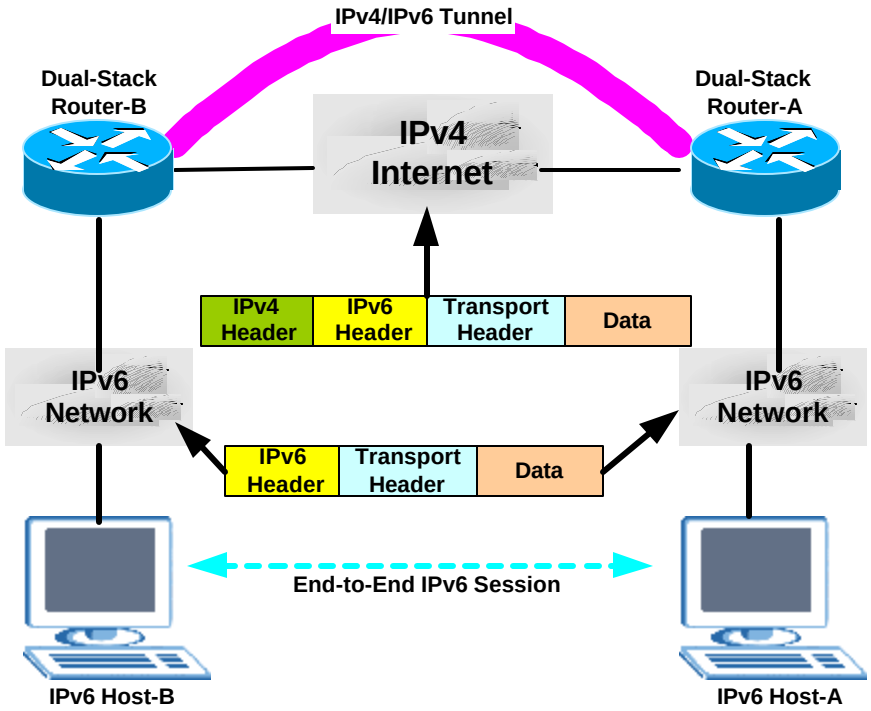
IPv6/IPv4 Dual-Stack Operation

- Have both IPv6 and IPv4 addresses
- Include resolver libraries capable of dealing with A, AAAA, A6 records
- Query to DNS for a name
- DNS could return
  - A record
  - AAAA/A6 record
  - both
- Resolver gives answers to application
- Application uses IPv6 or IPv4 depending on the answers received and their order

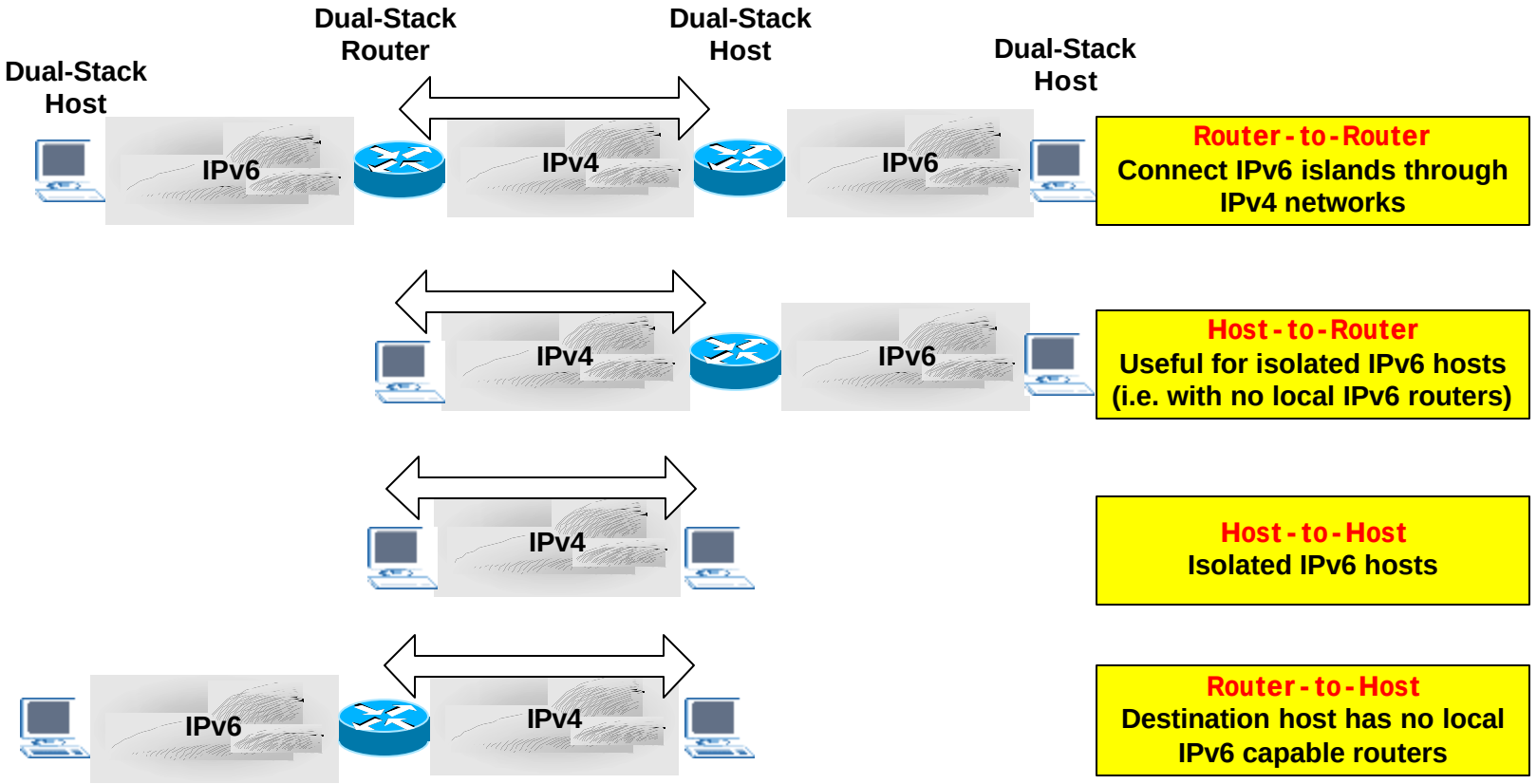


Tunneling Technique

- Carries incompatible protocols or specific data over an existing network
- Encapsulating the IPv6 packet in the IPv4 packet
- Tunneling can be used by routers and hosts
- This picture shows router-to-router tunneling



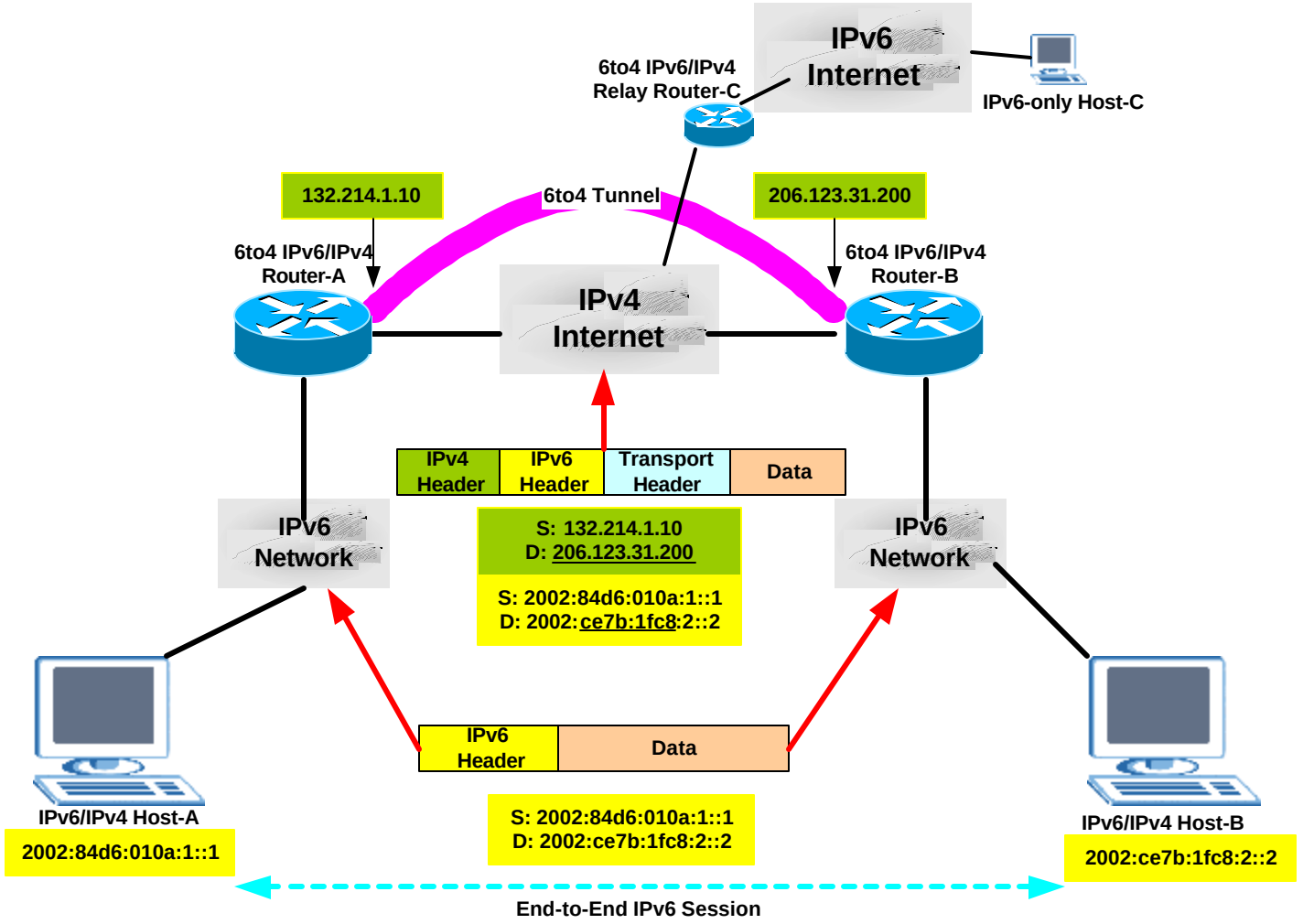
Tunnel Scenarios



6to4 Tunnel

- 2002::/16 is assigned to 6to4
- Requires one global IPv4 address on each end
- Site prefix is derived from the IPv4 address of the edge router
- Eg., 1.2.3.4 => 2002:0102:0304::/48
- Specify the SLA ID xxxx in radvd.conf => 2002:0102:0304:xxxx::/64
- Each host in IPv6 site auto-configure using that prefix
- You don't need to manually configure the other end of the IPv4 address because 6to4 router can always derive this address from the destination 2002::/48 IPv6 address
- Deployment Consideration:
  - The 6to4 edge router's IPv6 address might change over time, which will force the renumbering of the whole 6to4 site
  - Blocking or filtering inbound 6to4 traffic based on source IPv4 addresses breaks the 6to4 model

6to4 Tunnel (2)



6to4 Relay Router

- The 6to4 routers can only forward the IPv6 packets to any destination with 2002::/16 prefix over IPv4 network.
- Other aggregatable global addresses (e.g., 2001::/16, 3ffe::/16) are unreachable unless the 6to4 router is also a 6to4 relay router.
- A 6to4 router providing traffic forwarding to the native IPv6 Internet is called a 6to4 relay router.
- A 6to4 relay router is generally at the border of the IPv4 Internet and IPv6 Internet.
- To use a 6to4 relay router, a 6to4 router must add a default route in its routing table pointing to the 6to4 relay router.
- In Linux,
  - ip –6 route add 2000::/3 via ::192.88.99.1 dev tun6to4
  - IANA assigned the 6to4 relay anycast prefix 192.88.99.0/24 for routing 6to4 packets to the nearest 6to4 relay
  - You can configure a specific 6to4 relay IP address if you know it

Translators

- IPv6-only hosts to IPv4-only hosts communication
- IP layer
  - NAT-PT (RFC 2766)
  - SIIT (Stateless IP/ICMP Translation) (RFC 2765)
  - BIS (Bump-In-the-Stack) (RFC 2767)
    - Takes NAT-PT and SIIT functionality and moves it to the OS protocol stack within each host
- Transport layer
  - TCP-UDP Relay (RFC 3142)
    - (a.k.a Transport Relay Translator – TRT)
- Application layer
  - BIA (Bump-In-the-API) (RFC 3338)
  - SOCKS-based Gateway (RFC 3089)
  - Application Level Gateways (ALG)

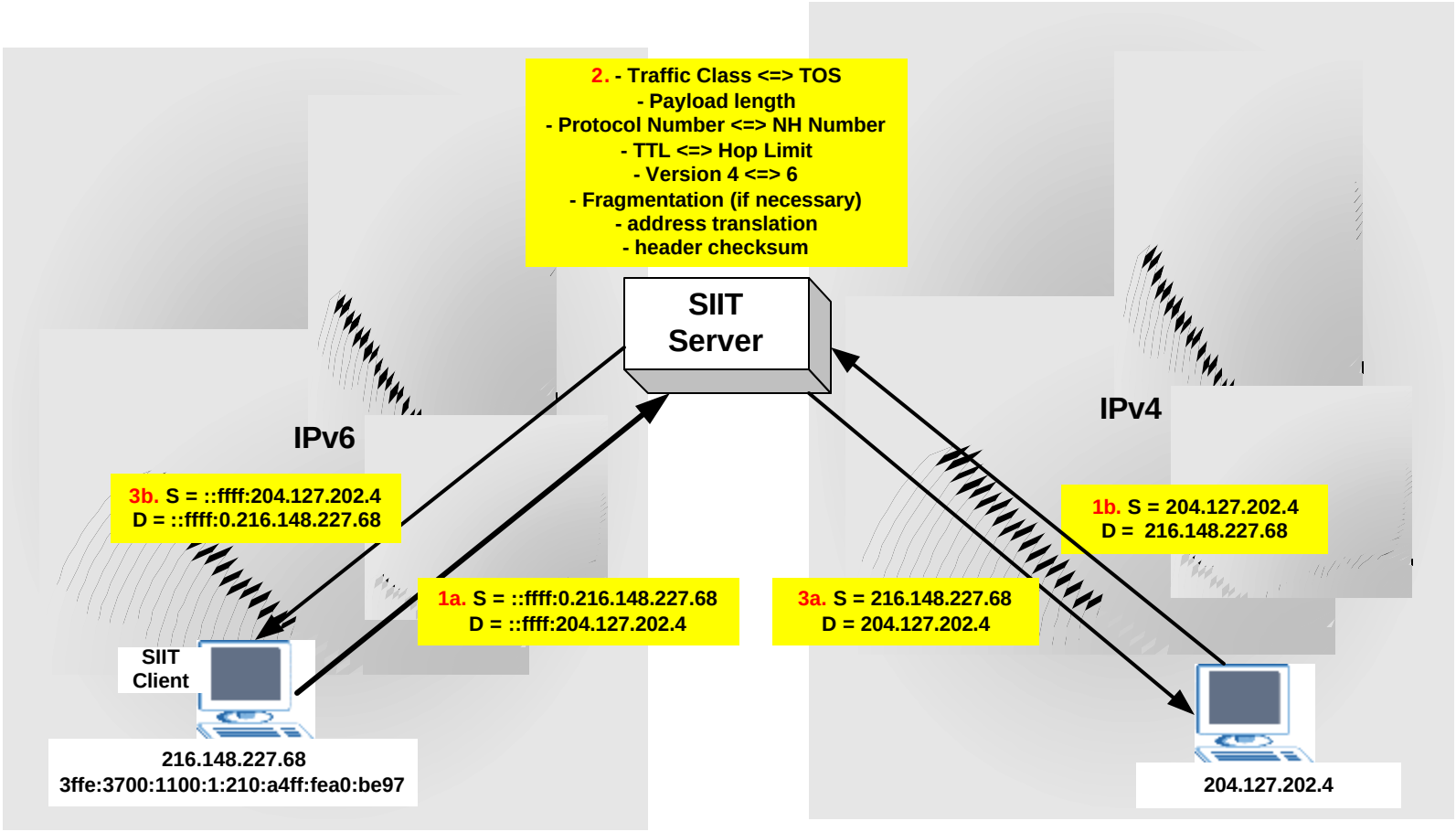
SIIT Server

- Stateless IP/ICMP Translation (RFC 2765)
- It is stateless operation
  - Session traffic is not necessary traverse the same NAT-PT translator
- Translator replaces headers IPv4 ⇔ IPv6
- Translate ICMP messages
  - Contents of message translated
  - ICMP pseudo-header checksum added
  - Headers and type values are translated
  - Some discarded, if no equivalent counterpart is defined
- Fragments IPv4 messages to fit IPv6 MTU when necessary
- SIIT does not translate
  - Routing headers
  - Hop-by-hop options
  - Destination options
  - End-to-end AH header
  - Multicast

SIIT Client

- Required IPv6-enabled hosts to acquire a translatable source IPv4 address from a shared pool; the translatable destination address is made from IPv4 DNS entry.
- Uses IPv4-mapped addresses to refer to IPv4-only nodes (destination)
  - 0:0:0:0:0:ffff/96 + 32-bit IPv4 address
- Uses IPv4-translated addresses to refer to IPv6-enabled nodes (source)
  - 0:0:ffff:0:0:0/96 + 32-bit IPv4 address
- Creates IPv6 packets and sends them to SIIT Sever

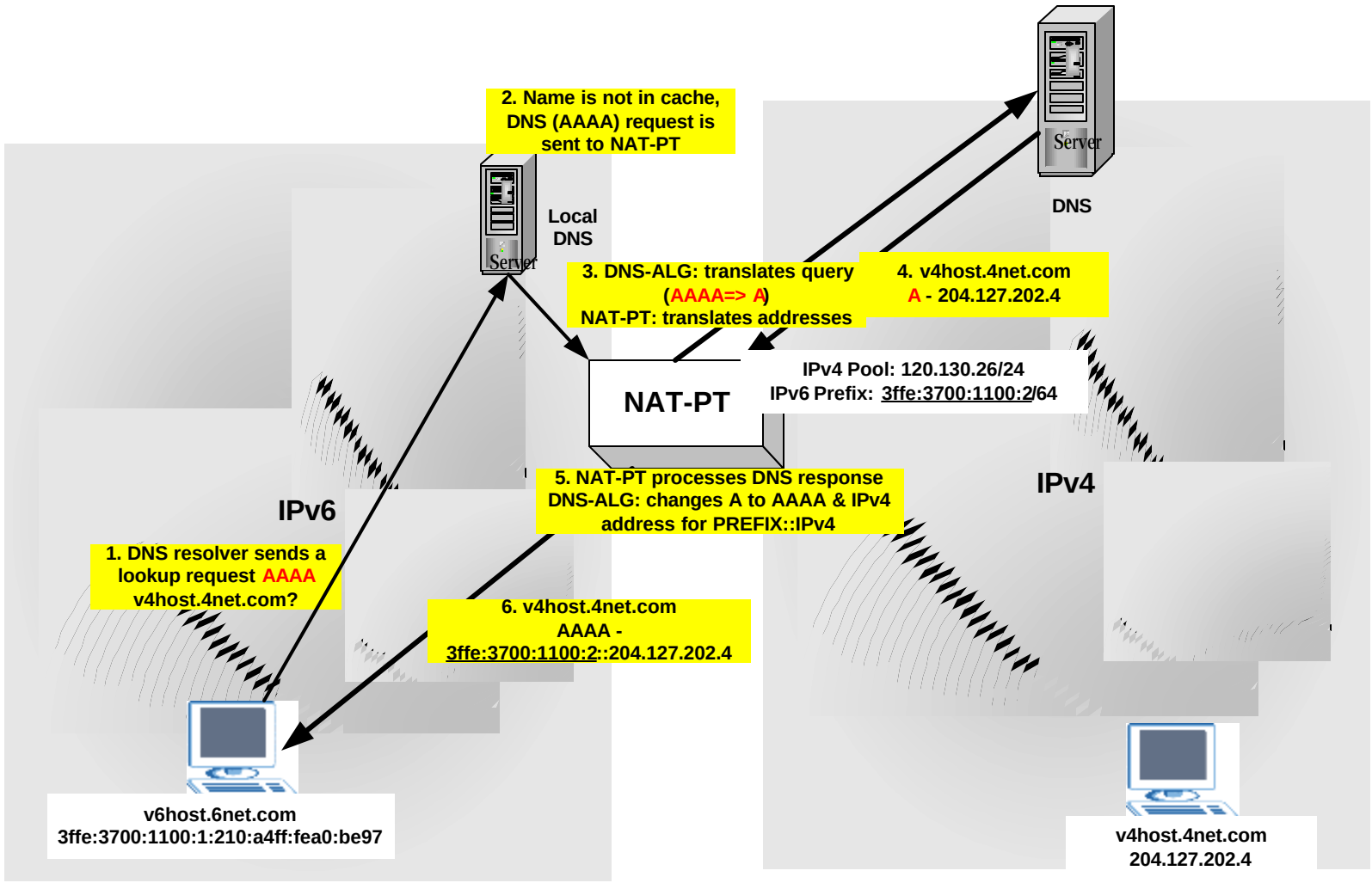
SIIT Scenario



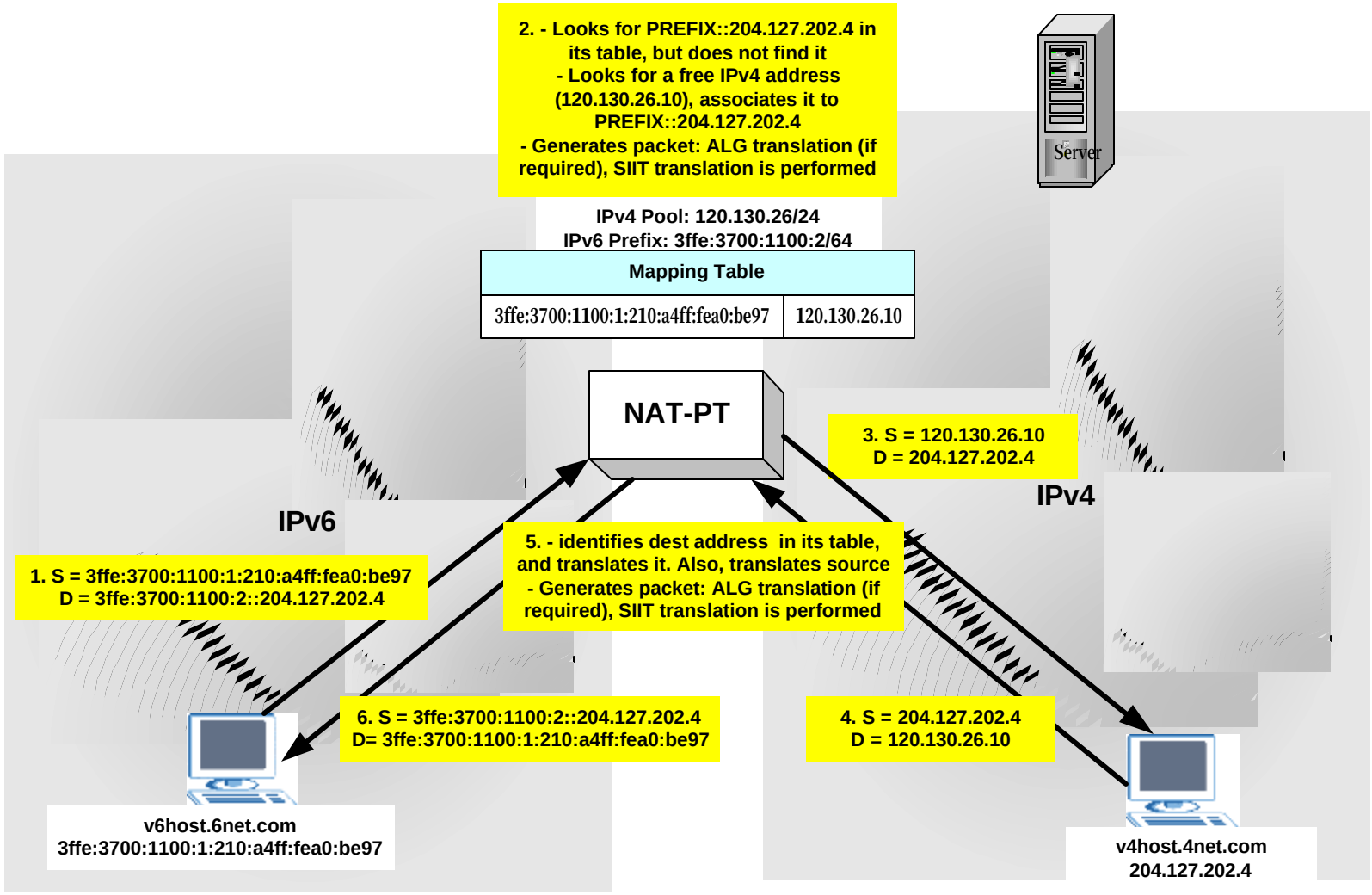
NAT-PT

- Network Address Translation – Protocol Translation (RFC 2766)
- Stateful operation
  - Session traffic must traverse the same NAT-PT translator
  - Once an association between addresses is established, it is kept for a given time
- Allows IPv6 hosts to communicate with IPv4 hosts
- NAT refers to translation of an IPv4 address into an IPv6 address and vice-versa.
- PT means the translation of the IPv4 packet into an equivalent IPv6 packet and vice-versa. (It uses SIIT for protocol translation)
- A NAT-PT device resides at the boundary between an IPv6 and IPv4 network. It is also a dual stack border router node with two Ethernet interfaces.
- It uses a pool of IPv4 addresses for assigning to IPv6 nodes dynamically.
- Address translation is sometimes required at application level
  - Application Layer Gateway (payload has embedded IPv4/v6 address)
  - For applications that transport addresses: DNS-ALG, FTP-ALG, SIP-ALG

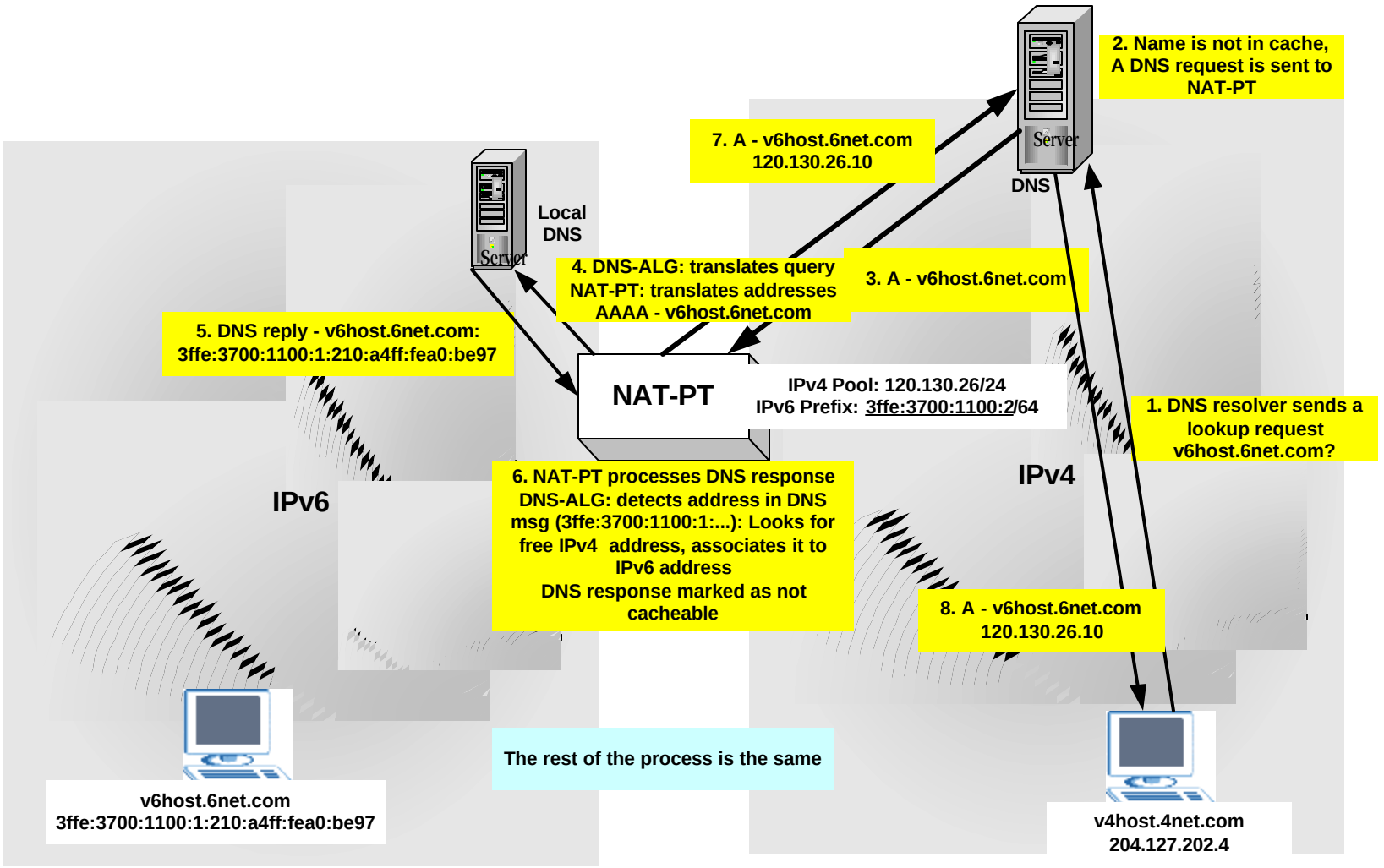
NAT-PT Scenario (1) – IPv6 Initiated



NAT-PT Scenario (2) - Data



NAT-PT Scenario (3) – IPv4 initiated



Q & A