

# IPv6 Neighbor Discovery DeepDive

Jen Linkova aka Furry  
furry13@gmail.com  
Nov 2015

# Most Important Slide of This Talk

IPv6 is

**NOT**

“like IPv4 but longer addresses”

# What Can Be Improved in IPv6?

- Just one control protocol?
  - IPv4: ARP for L2, ICMP for Internet
- Less chatty protocol?
  - ARP using broadcast
- do more than just map IP <> L2 addresses?
  - ARP does not confirm reachability

# Protocol Choice

- ICMPv6 is already here as a control protocol
- No reason to use non-IP protocol
- Flexible
  - new message types can be created
- Like in ARP, 2 messages are needed:
  - **Neighbor Solicitation**, "Who has this IPv6 address"?
  - **Neighbor Advertisement**, "I have this IPv6 address"

# ARP Is Too Chatty

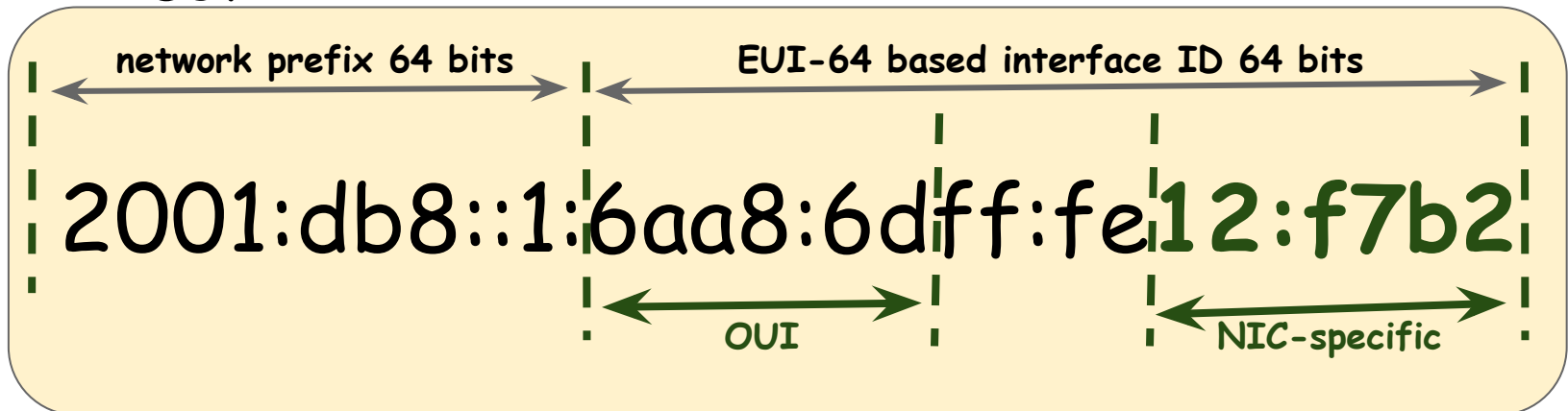
- Broadcast
  - creates noise
  - consumes network resources
  - Kills the battery on mobile devices
- NS message should be received by?
  - all hosts? => broadcast ❌
  - or hosts which might have that address? 😊



**MULTICAST**

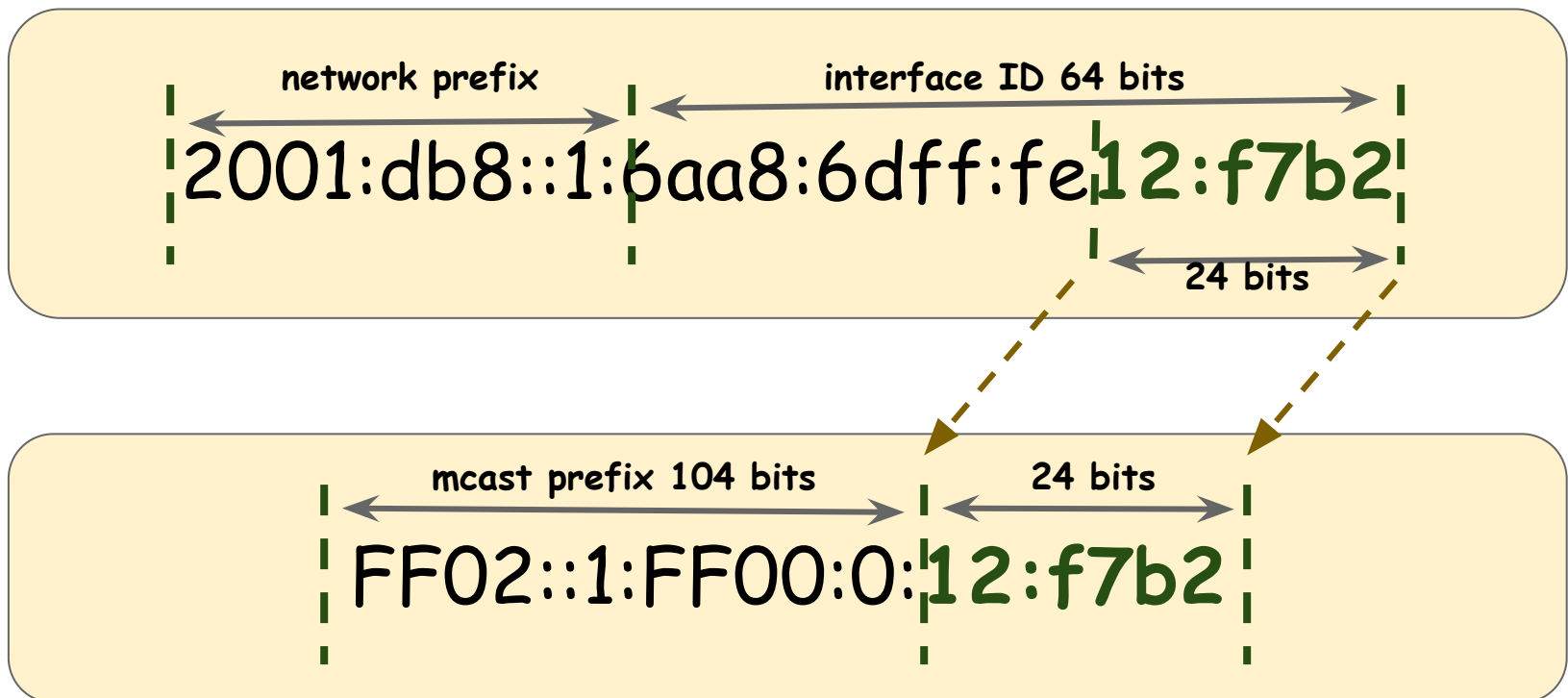
# Multicast Group Address

- What about: one IPv6 address - one mcast group (ff02::- Too many mcast groups
- Local resolution to L2 address - only 32 bits of L3 addresses going to L2 mcast address
- EUI-64: highest 24 bits are OUI, then fffe
- Manually assigned interface\_id: no highest bits set



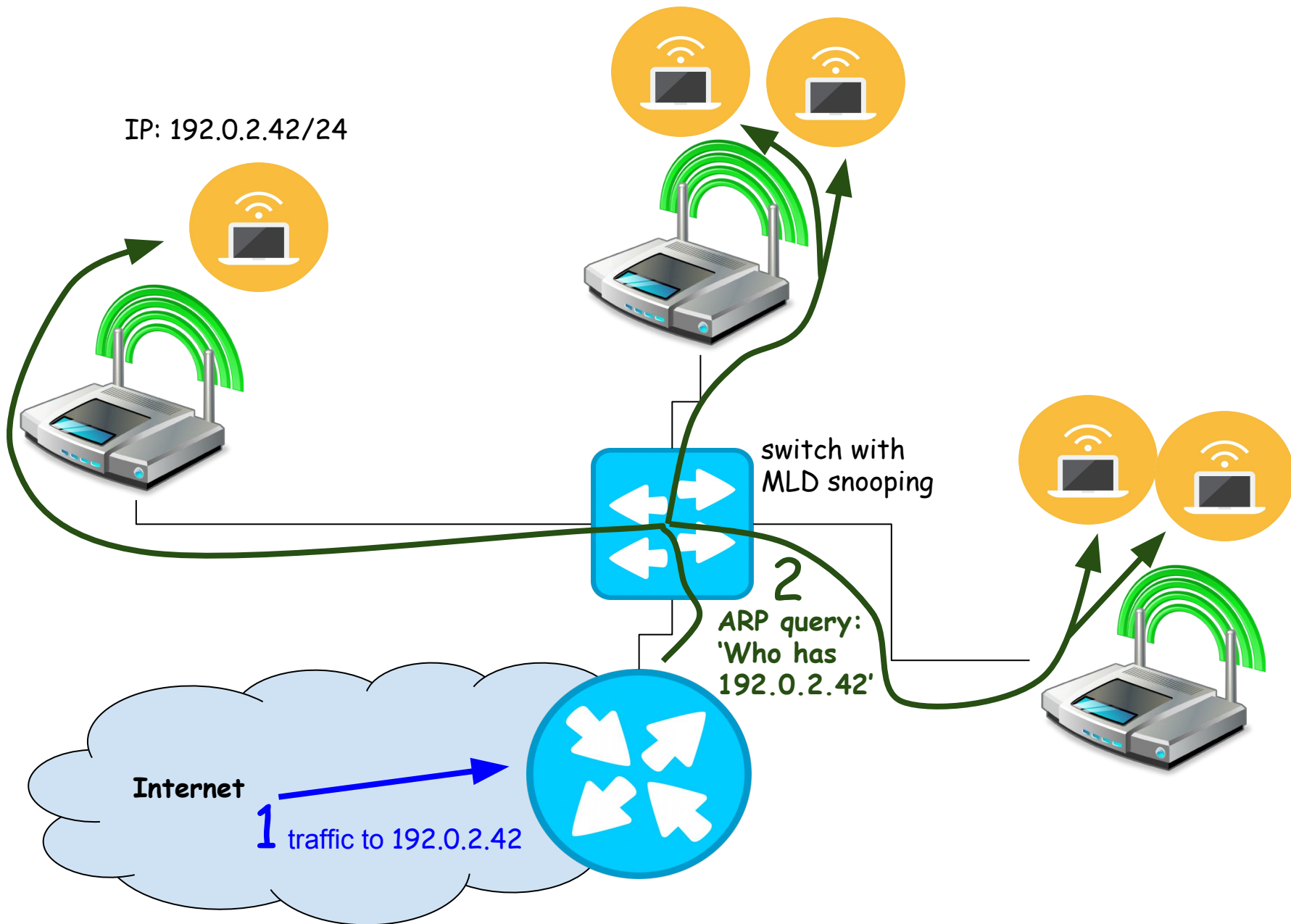
# Solicited Node MCast Address

- Lower 24 bits of IPv6 address
- Solicited-node multicast address format:
  - Globally-assigned prefix **FF02::1:FF00:0:/104**
  - low-order 24 bits of a node address



solicited node multicast address

# IPv4/ARP: WiFi





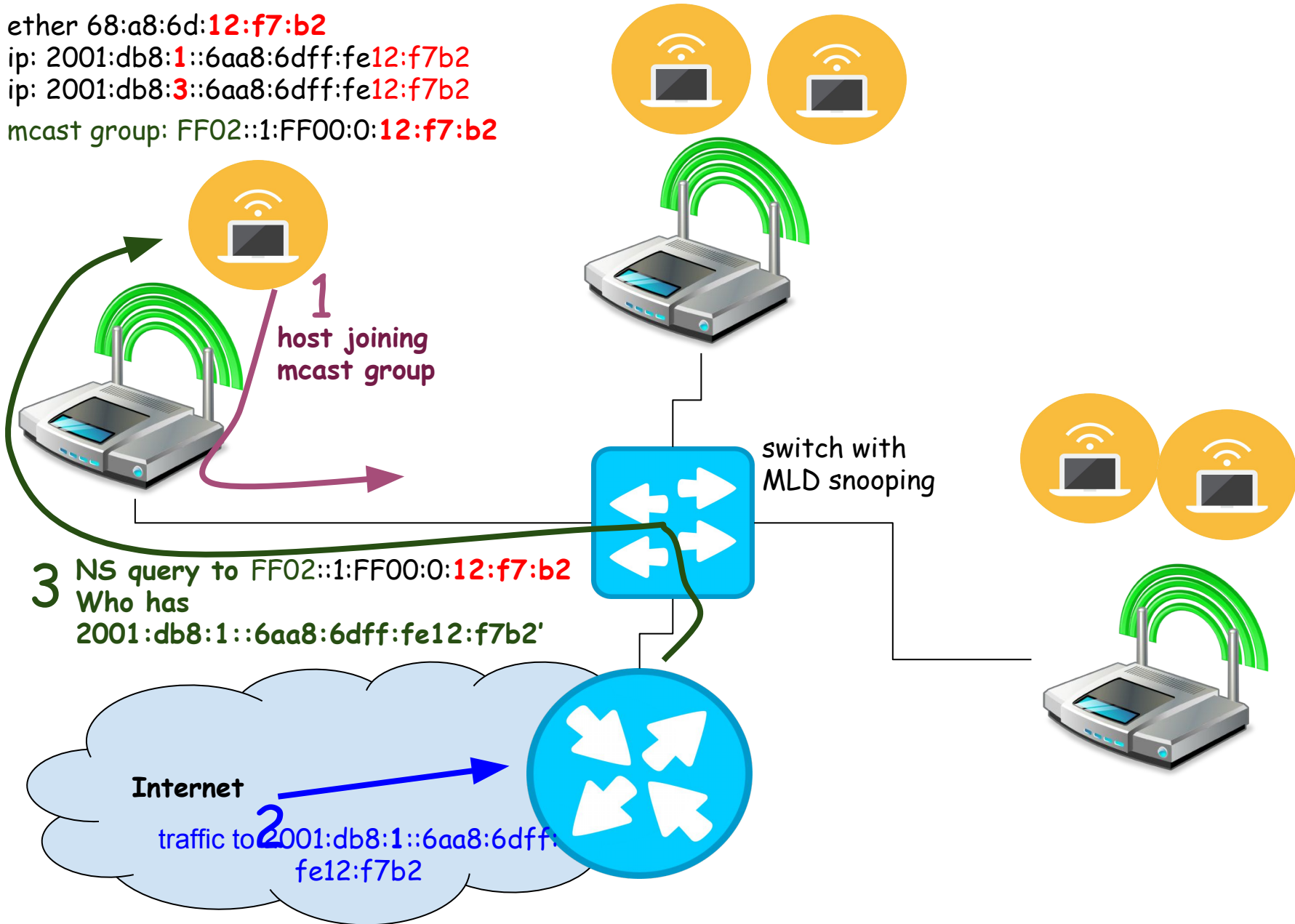
# Multicast, ND and MLD: WiFi

ether 68:a8:6d:12:f7:b2

ip: 2001:db8:1::6aa8:6dff:fe12:f7b2

ip: 2001:db8:3::6aa8:6dff:fe12:f7b2

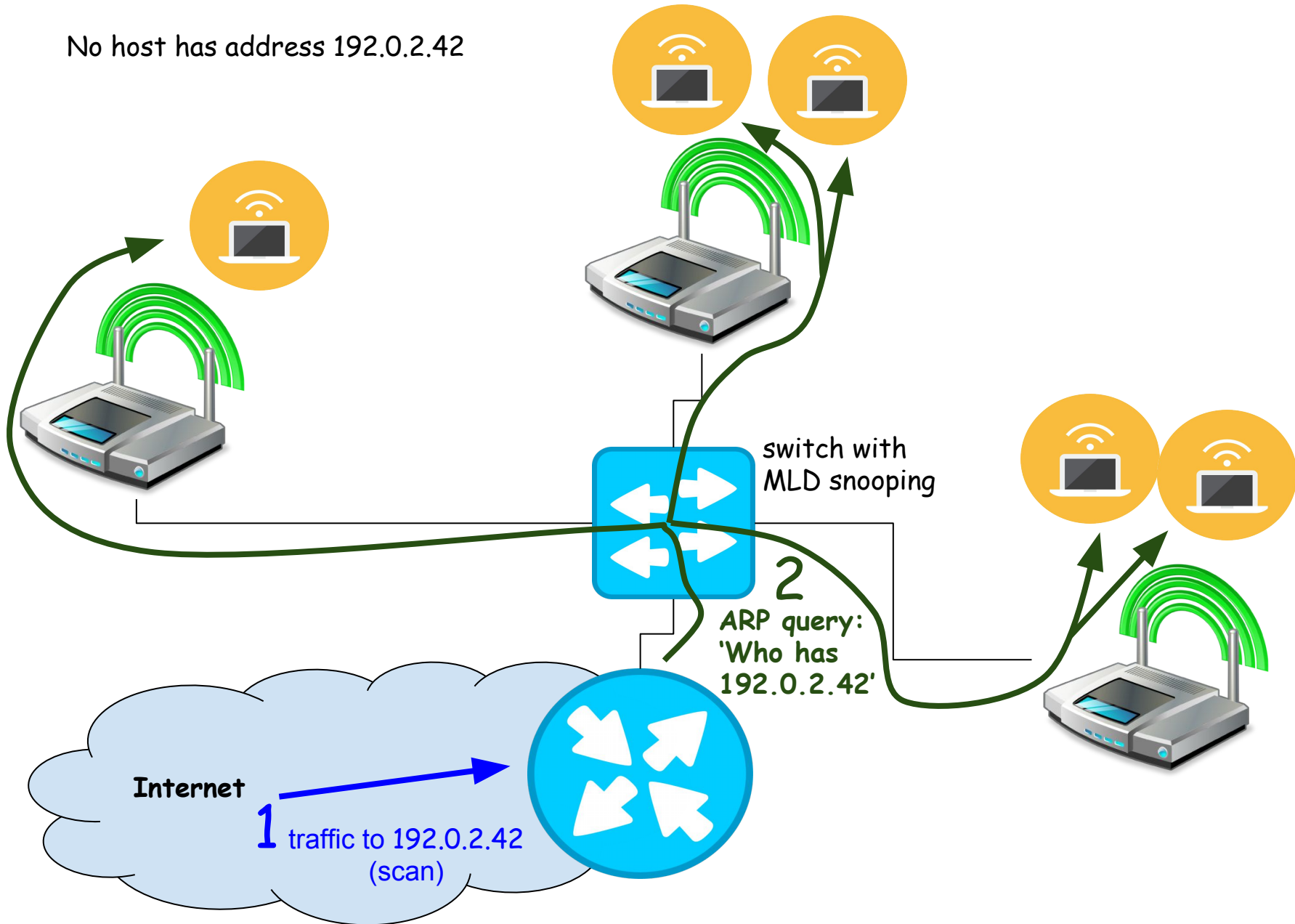
mcast group: FF02::1:FF00:0:12:f7:b2



3 NS query to FF02::1:FF00:0:12:f7:b2  
Who has  
2001:db8:1::6aa8:6dff:fe12:f7b2'

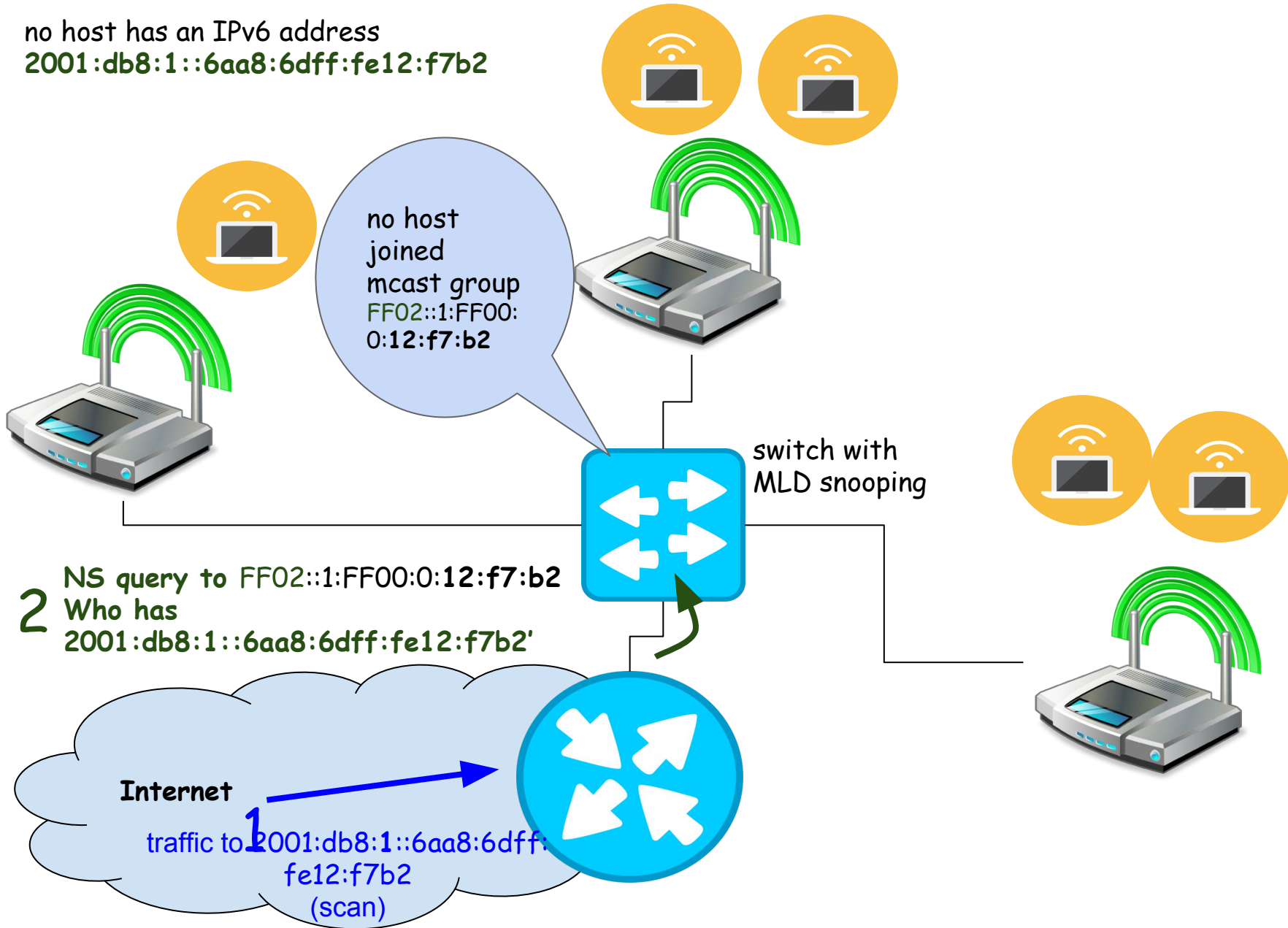
# Non-Existent Address & WiFi: V4

No host has address 192.0.2.42

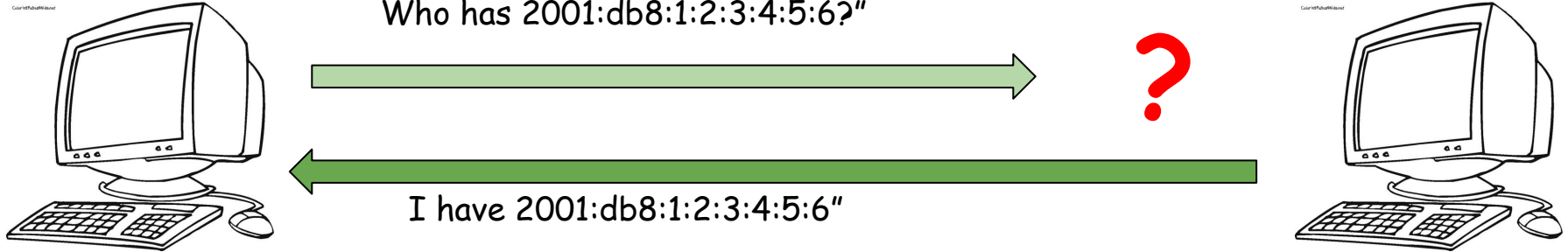


# Non-existent Address and WiFi: V6

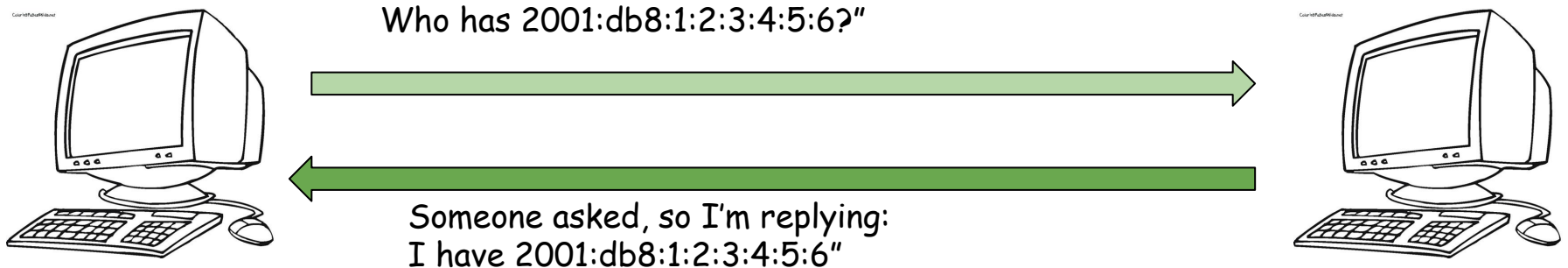
no host has an IPv6 address  
2001:db8:1::6aa8:6dff:fe12:f7b2



# Improve ND Robustness (1)



**one-way communication possible**

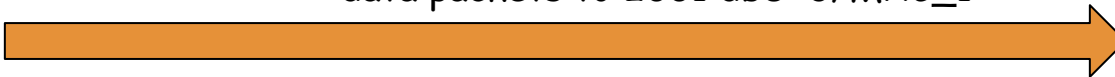


**most likely communication is bi-directional**

# Improve ND Robustness (2)

I have traffic for 2001:db8::3. The cache entry for 2001:db8::3 is very old and needs updating. But maybe L2 address has not changed and it is still MAC\_1?

data packets to 2001:db8::3/MAC\_1



from: fe80::3 to 2001:db8::3 Who has 2001:db8::3?"



I have 2001:db8::3, my L2 address is MAC\_1"

2001:db8::3

cool, 2001:db8::3 still has MAC\_1, refreshing my cache entry for 2001:db8::3



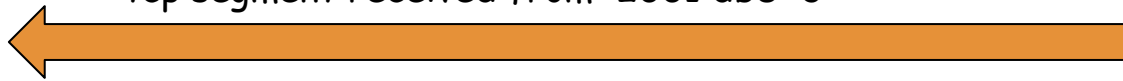
# Upper-Layer Integration

The cache entry for 2001:db8::3 is about to expire and needs updating. Oh, but TCP layer has just confirmed bi-directional communication and it is using the existing `MAC_1` - so it seems to be correct one!



2001:db8::2

tcp segment received from 2001:db8::3



2001:db8::3

# Neighbor Caches Entry Lifecycle

REACHABLE

“good to use” IPv6 <> MAC entry



ReachableTime, random value normally 15-45 secs



STALE

last known to be “good to use” entry



Used to be: unlimited time until new packet needs to be sent;  
Now: just long timer



DELAY

“let's try to use it...it might work..”



DELAY\_FIRST\_PROBE timer, default = 5 sec



PROBE

let's send unicast NS to that address



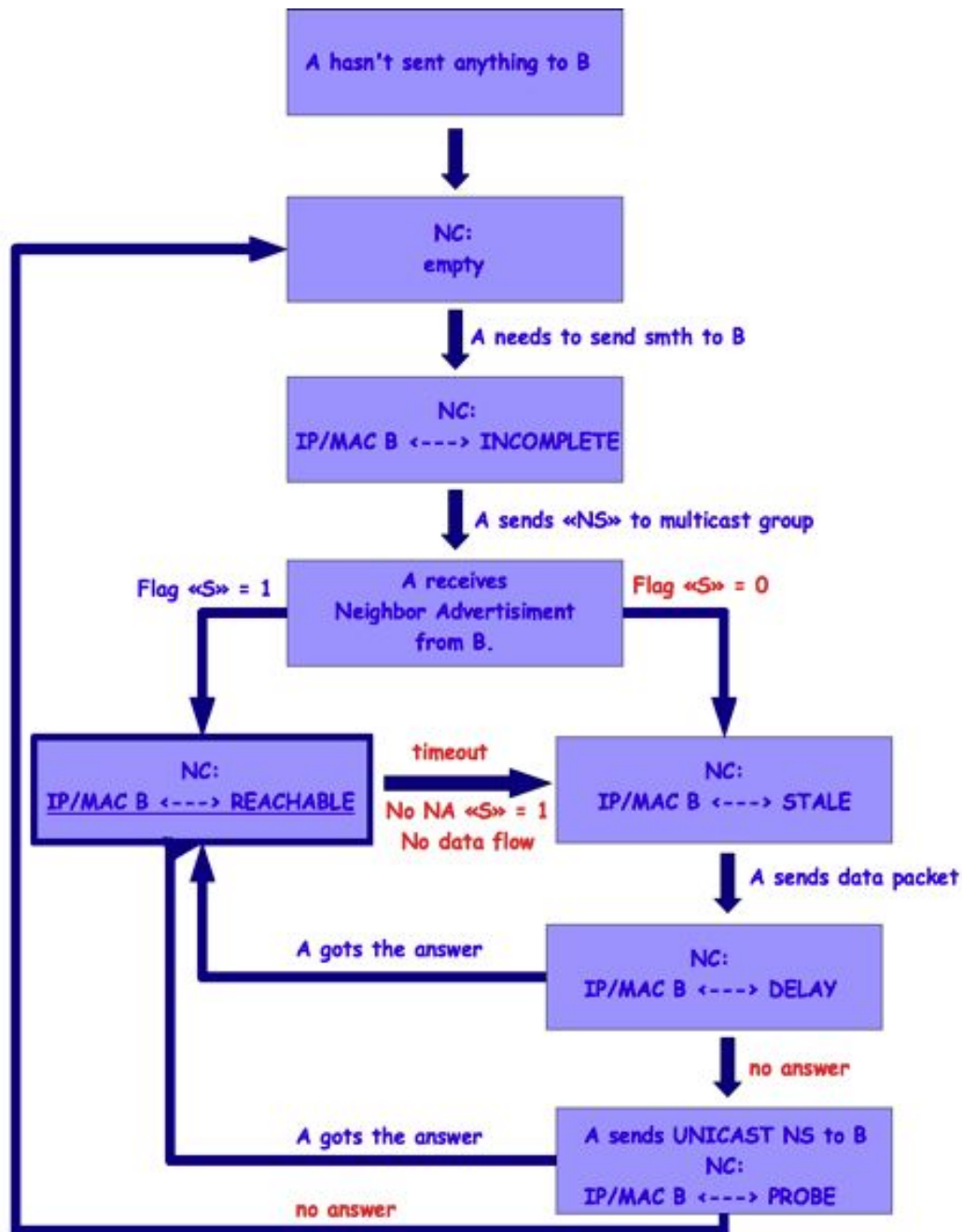
MAX\_UNICAST\_SOLICIT times  
RETRANS\_TIMES, default = 3x1 sec



NON-EXISTENT

old entry does not work, delete it!







# IPv6 Autoconfiguration

# How to Configure a Host

- Manual configuration
- Controlled by external system (DHCP)
  - stateful or stateless
- **StateLess Address Auto-Configuration (SLAAC)**
  - network prefix
    - routers know prefixes configured
    - well-known link-local fe80::/10
  - host can generate interface id:
    - EUI-64
    - random
    - other options

# Interface ID: Avoiding Duplicates

interface id generated by a host might not be unique



before using an address - check for duplicates

???HOW???

host needs to know if there is another owner for that IP



Neighbor Discovery locates owners of IP addresses

# Duplicate Address Detection

# Duplicate Address Detection (DAD)

Host A generates an IP 'A' for interface 'I'

Host A joins mcast groups: 'all nodes on the link' (ff02::1)  
solicited-node multicast group for address A (FF02::1:FF00:0:A)

NS queries for A received?  
dst FF02::1:FF00:0:A', src ::

YES

NO

Host A sends multicast NS for address A

NA received for A?

YES

NO

Address A is NOT unique

Address A is unique

# Duplicate Address Detection (contd)

- DAD **MUST** be performed for all addresses (ex. anycast)
- Heavily relies on multicast so false negatives might happen:
  - on busy wifi networks
  - if multicast is broken

After DAD failure manual actions are required.

# Prefix and Router Discovery

## Critical Info to Discover

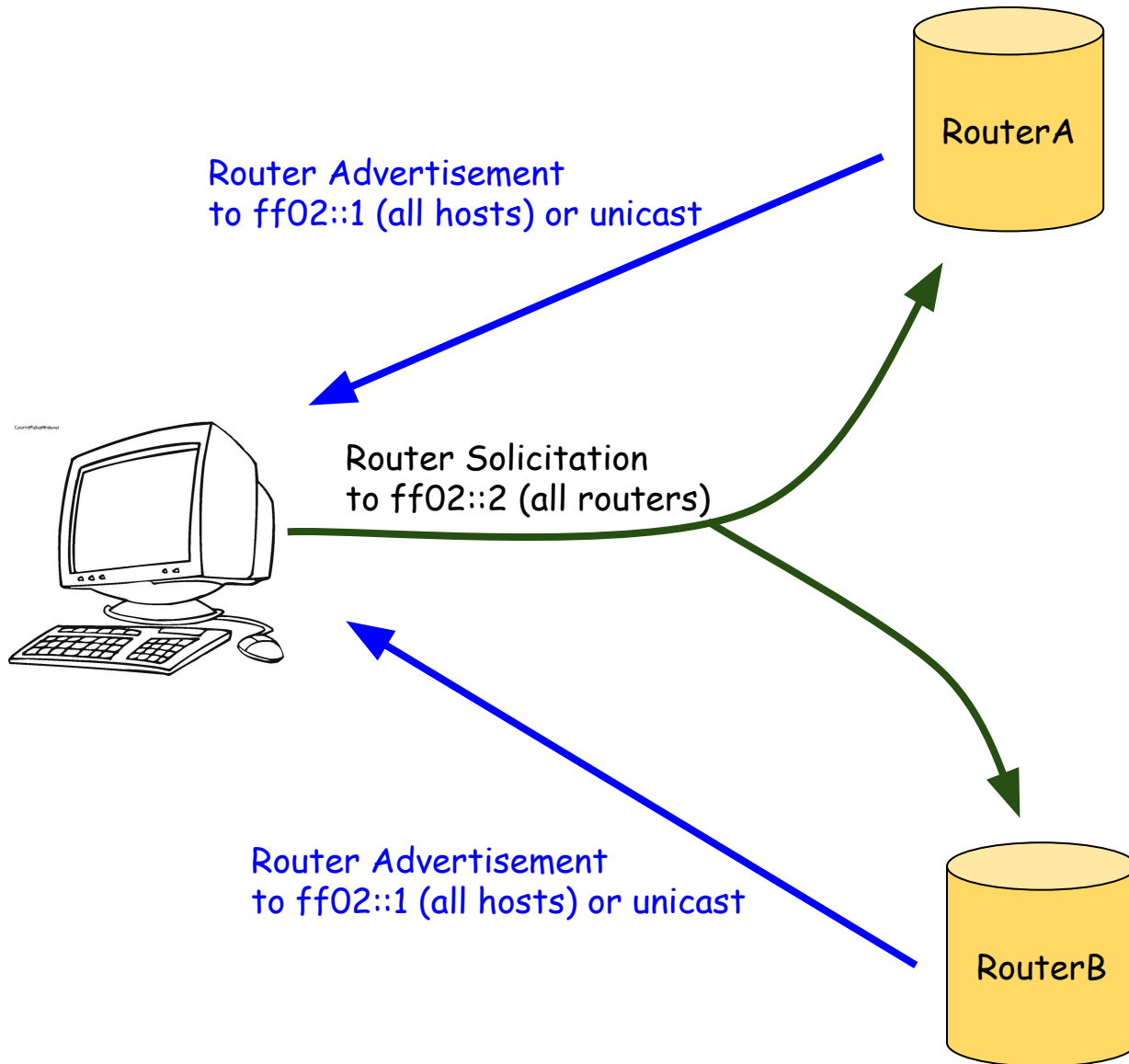
- Network prefix (to complete the address configuration)
- Default router(s)
- Routes to some destinations
- DNS servers and search list

Routers have most of this info

Routers are neighbors => ND can be used!



# Router Solicitation and Advertisement



# Router Advertisement Message

		8	10	16	32
type = 134		code = 0		checksum	
hop limit	M	O	reserved	router lifetime	
reachable time					
retransmit timer					
options (variable length)					

Src IP = link-local, Dst IP = the source IP of the RS query or FF02::1

- **M, O flags:** indicate that addresses (M) or other configuration info (O) is available via DHCPv6
- **Router lifetime** (in seconds) - the lifetime associated with the default router (0 - the router isn't default router, shouldn't appear on the default router list)
- **Reachable time** (millisecs) - how long the neighbor is reachable after receiving a reachability confirmation (NC record goes from Reachable -> Stale then)
- **Retransmit timer** (millisecs) - the interval between retransmitted NS messages

# RAs: What Can Possibly Go Wrong #1

- Solicited RAs: in response to RS but rate-limited to  $1 / \text{MIN\_DELAY\_BETWEEN\_RAS}$  sec (default: 3 secs)
- Large-scale wifi network:
  - every 3 secs
  - kills the battery on mobile devices
- Solution: to send solicited RAs unicast

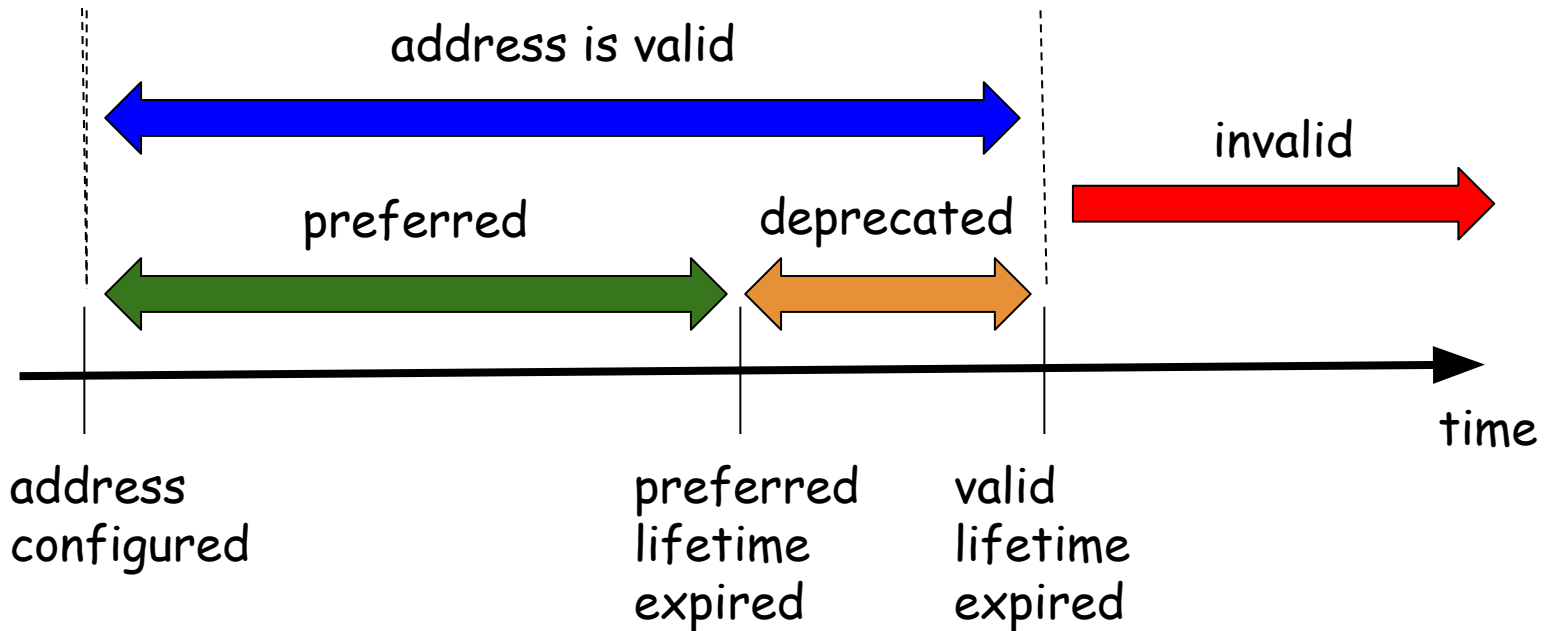
see [draft-ietf-v6ops-reducing-ra-energy-consumption](#)

# Prefix Information Option (PIO)

0	7	15	23	26	31			
TYPE = 3		LENGTH = 4		PREFIX LENGTH		L	A	RESERVED
VALID LIFETIME								
PREFERRED LIFETIME								
RESERVED								
PREFIX (128 bit)								

- L - prefix is on-link
- A - prefix can be used for SLAAC

# It's All About Timers

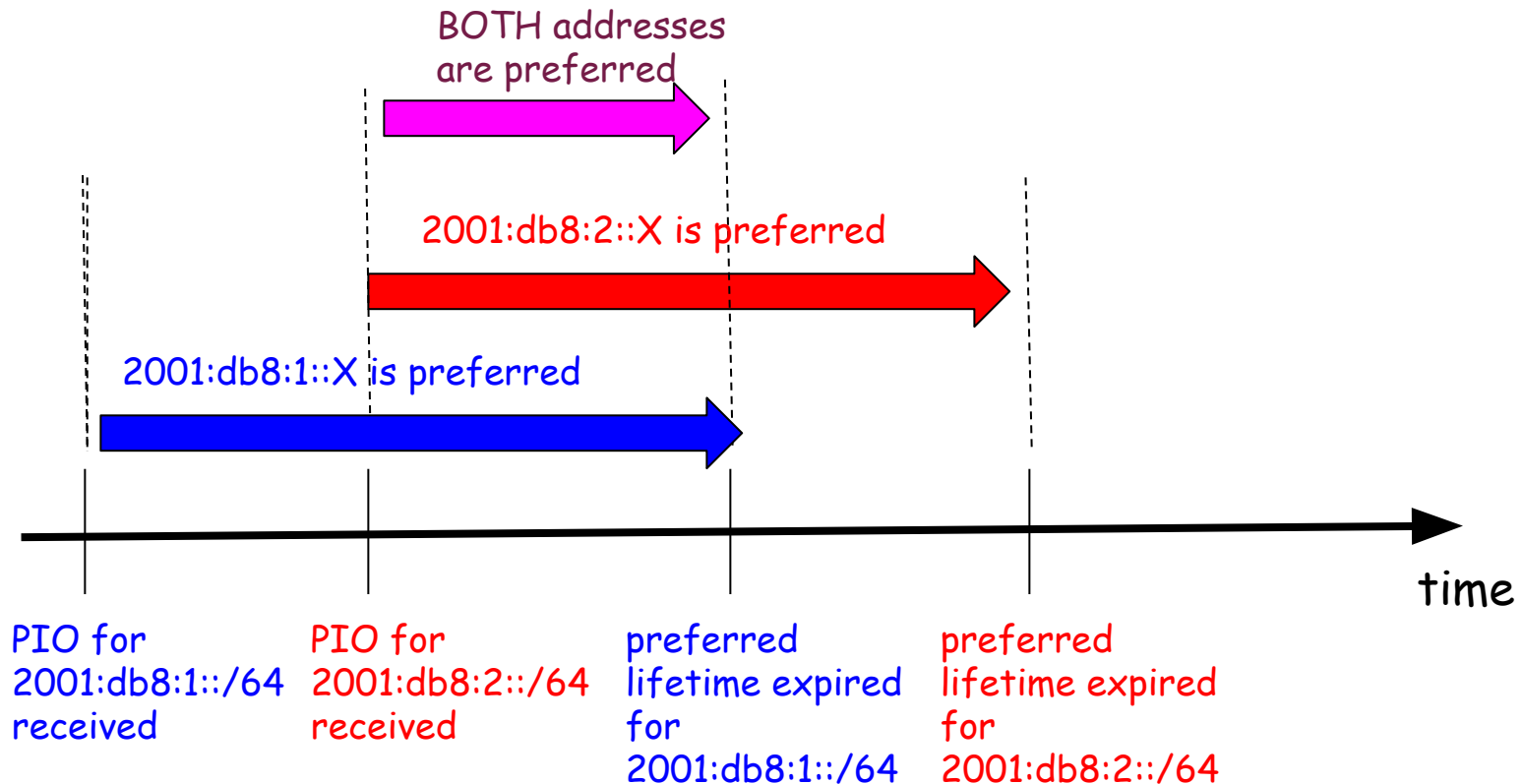


- Deprecated address
  - SHOULD be used for existing communications
  - SHOULD NOT be used for new ones
- Preferred lifetime  $\leq$  Valid lifetime
- Can not set valid lifetime  $<$  2hrs

Default values:

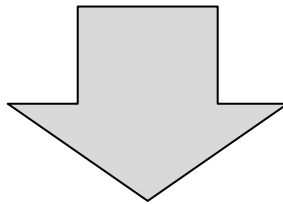
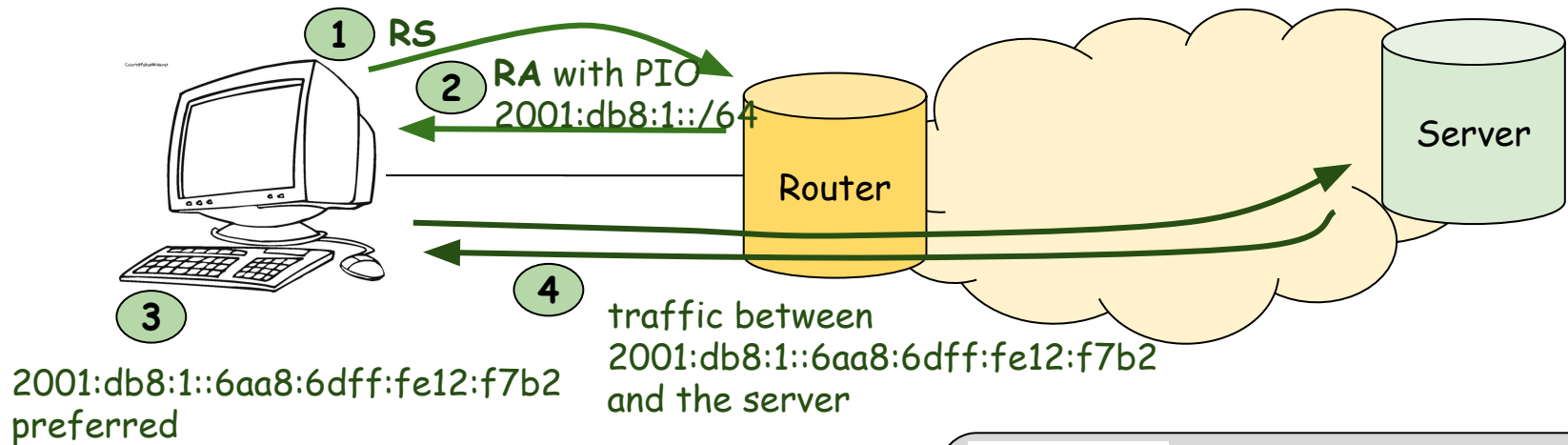
- preferred lifetime - 7d
- valid lifetime - 30d

# Multiple PIOs

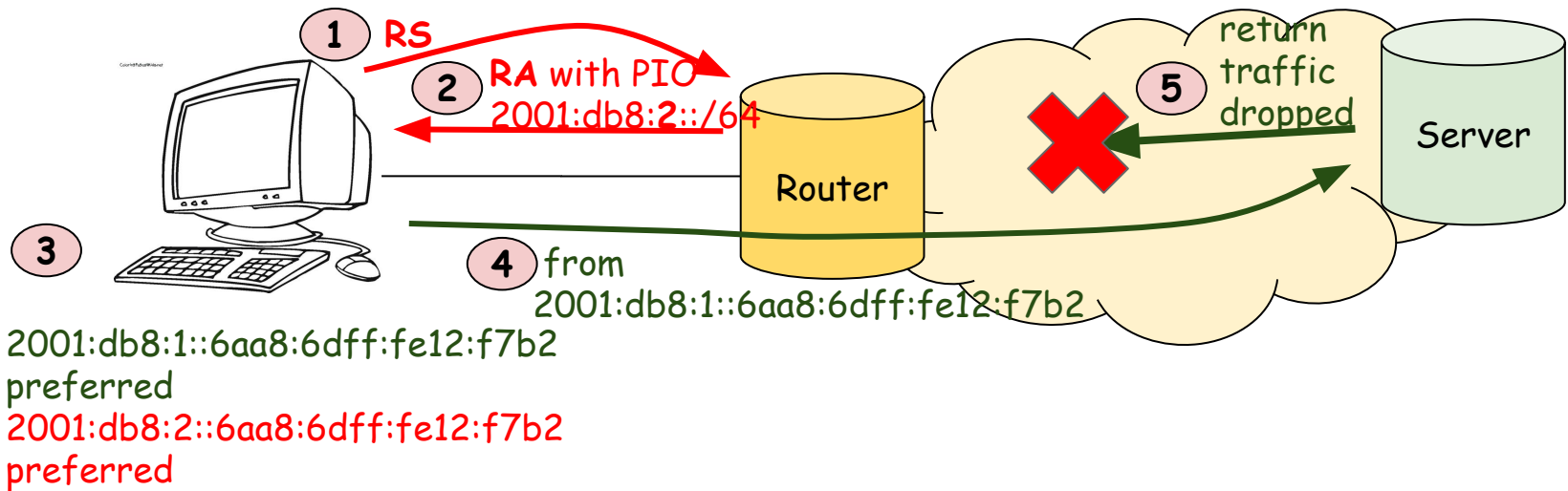


- **Removing a prefix from router configuration DOES NOT mean hosts stop using it**
- To renumber a new RA with PIO with preferred lifetime = 0 needs to be sent

# How NOT TO Renumber



```
# show | compare  
[edit protocols router-advertisement interface ae1]  
  
+ prefix 2001:db8:2::/64;  
- prefix 2001:db8:1::/64;
```



# Second Most Important Slide

Never

Ever

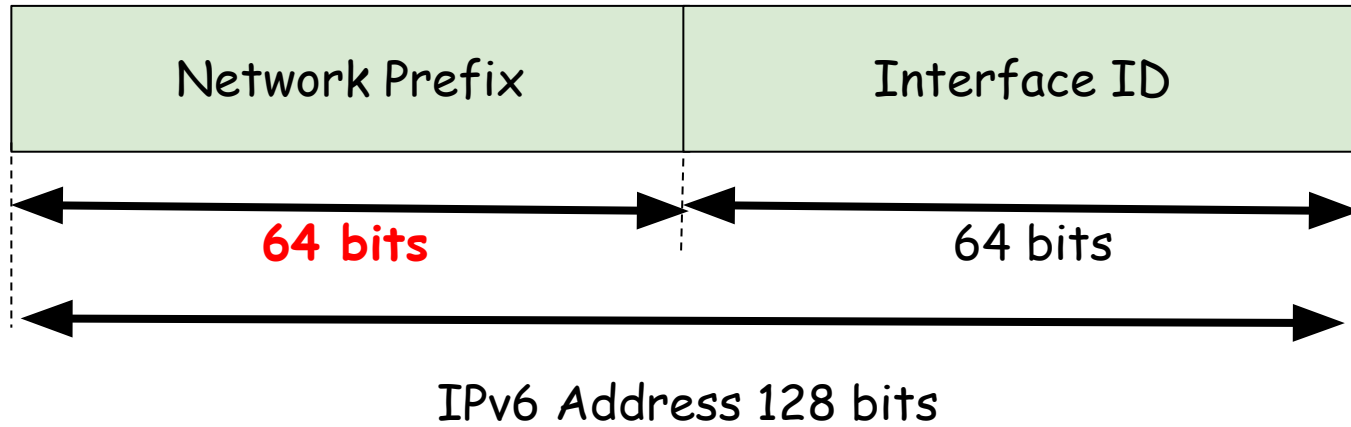
Use

Prefixlength

between 64 and 127 bytes



# RAs, PIOs and Prefix Length



**SLAAC DOES NOT** work with PIOs if prefix length  $\neq 64$

What the prefix length field is for then?

# The Host, the Link, and the Subnet

# What Does "Subnet Mask" Mean

- IPv4: IP address + netmask => on-link prefix
- IPv6: On-link prefix & address assignment are separated!
  - Link-local - always on-link
  - Any other addresses - if explicitly told so

*See "IPv6 Subnet Model: The Relationship between Links and Subnet Prefixes", RFC5942*

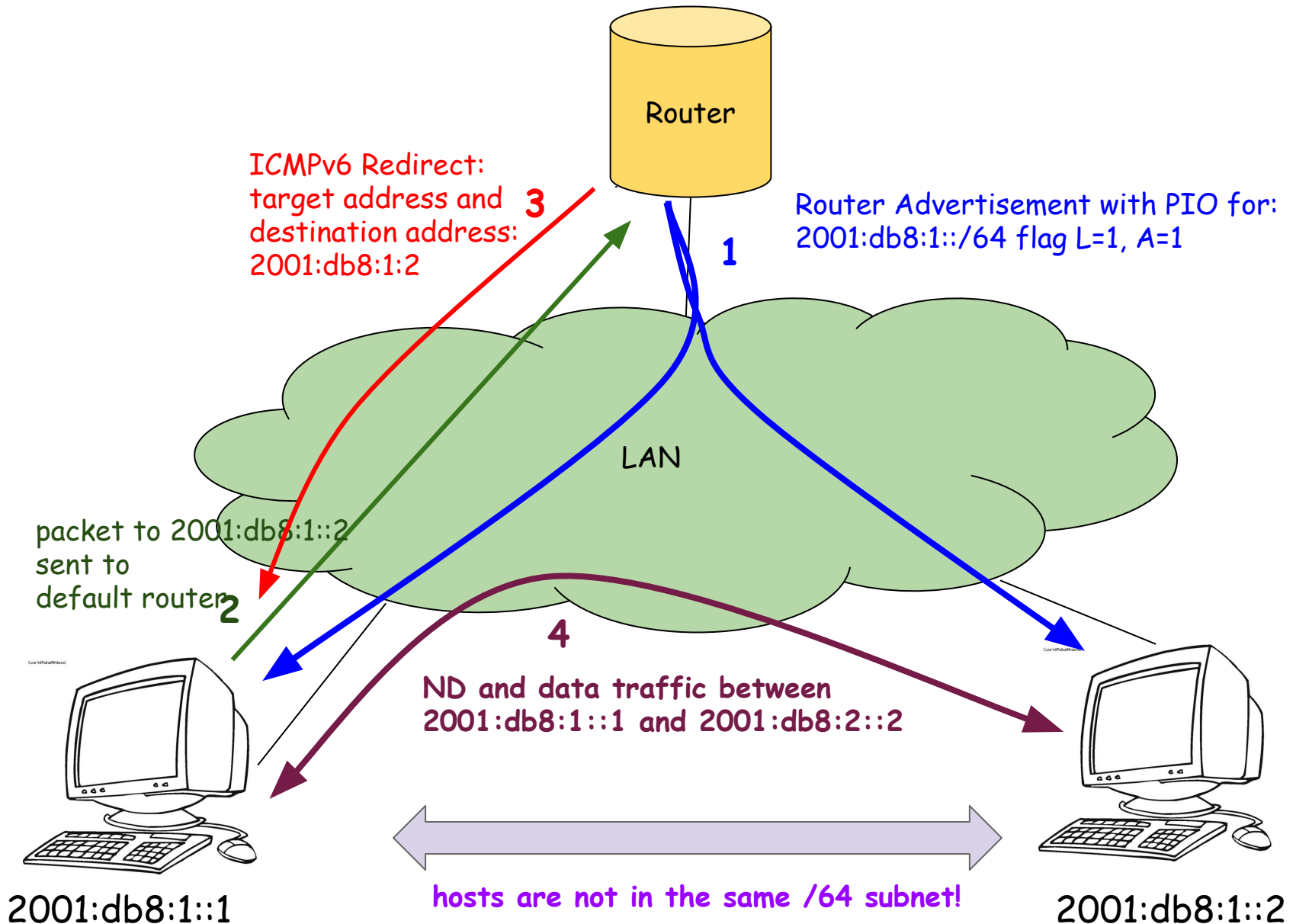
# On-link Prefix List

Prefix is on-link if

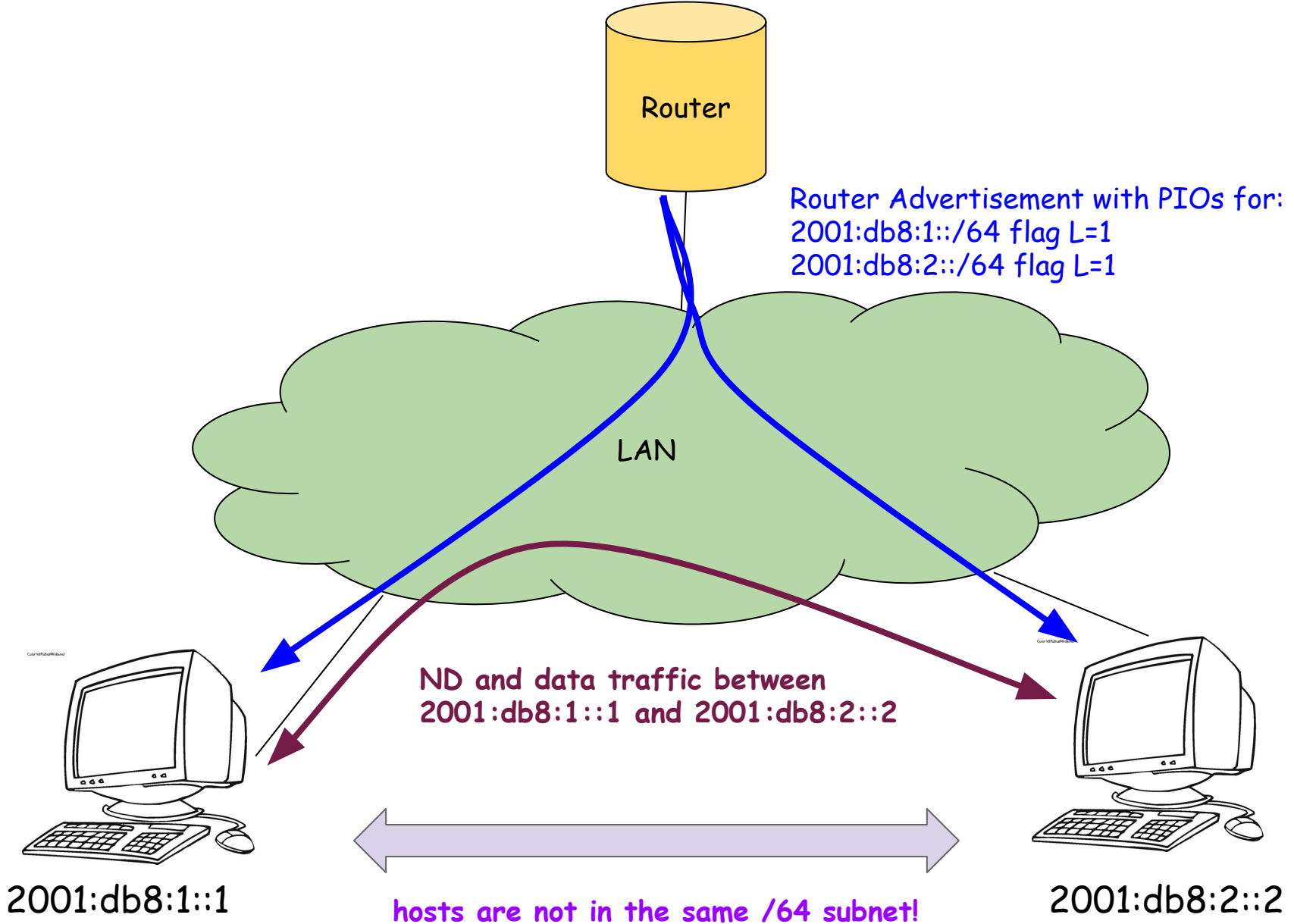
- PIO with L flag received
- ICMPv6 Redirect received from a router
- a Neighbor Advertisement message is received for the (target) address
- any Neighbor Discovery message is received from the address
- it is link-local fe80::/10

Everything else is OFF-LINK

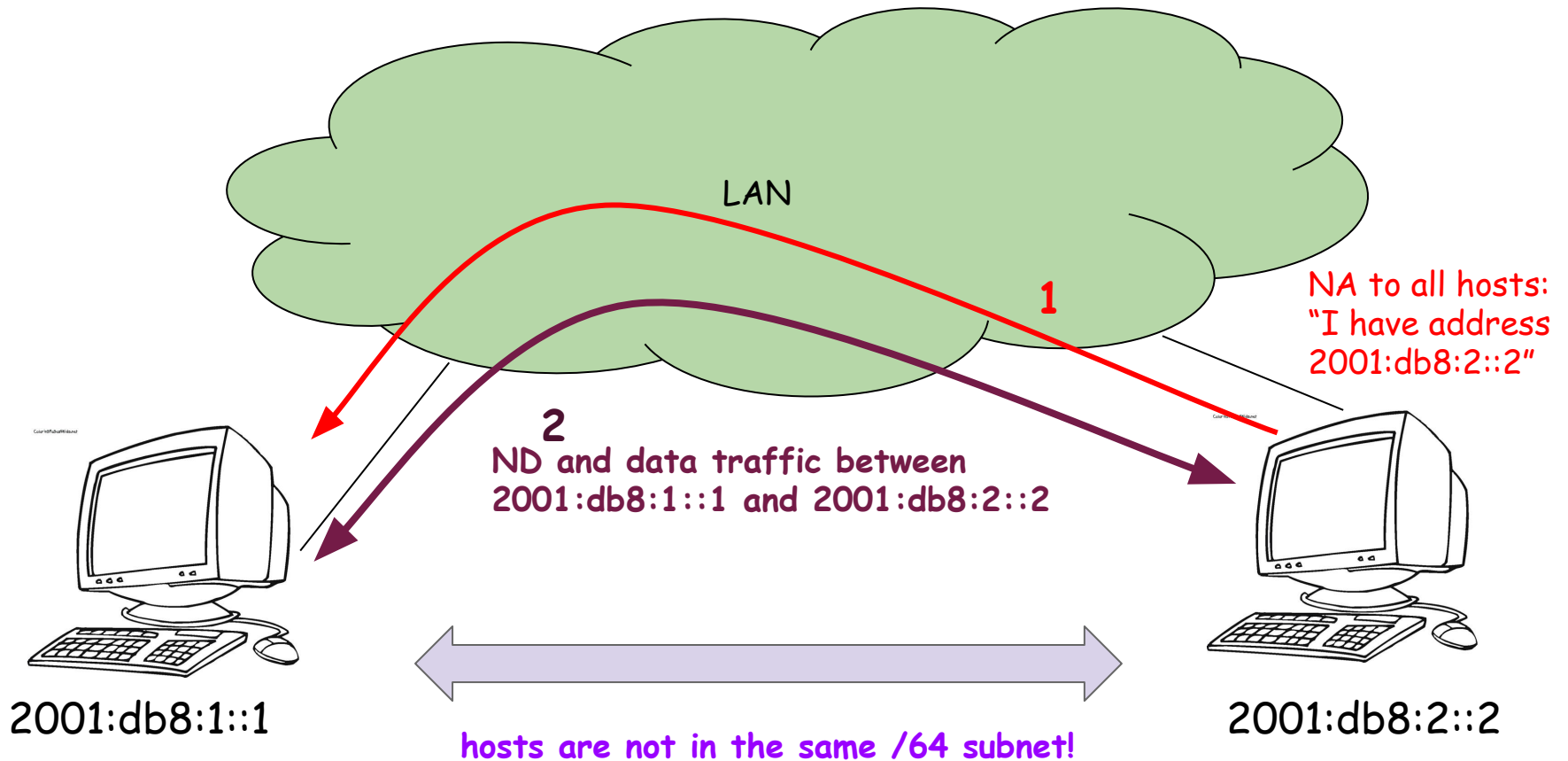
# On-link Prefixes: Examples



# On-link Prefix: Examples



# On-link Prefix: Examples



# Other RA Options

- Route Information Option
  - more specific routes
- MTU
- DNS
  - Recursive DNS Server
  - Search List



**Questions?**