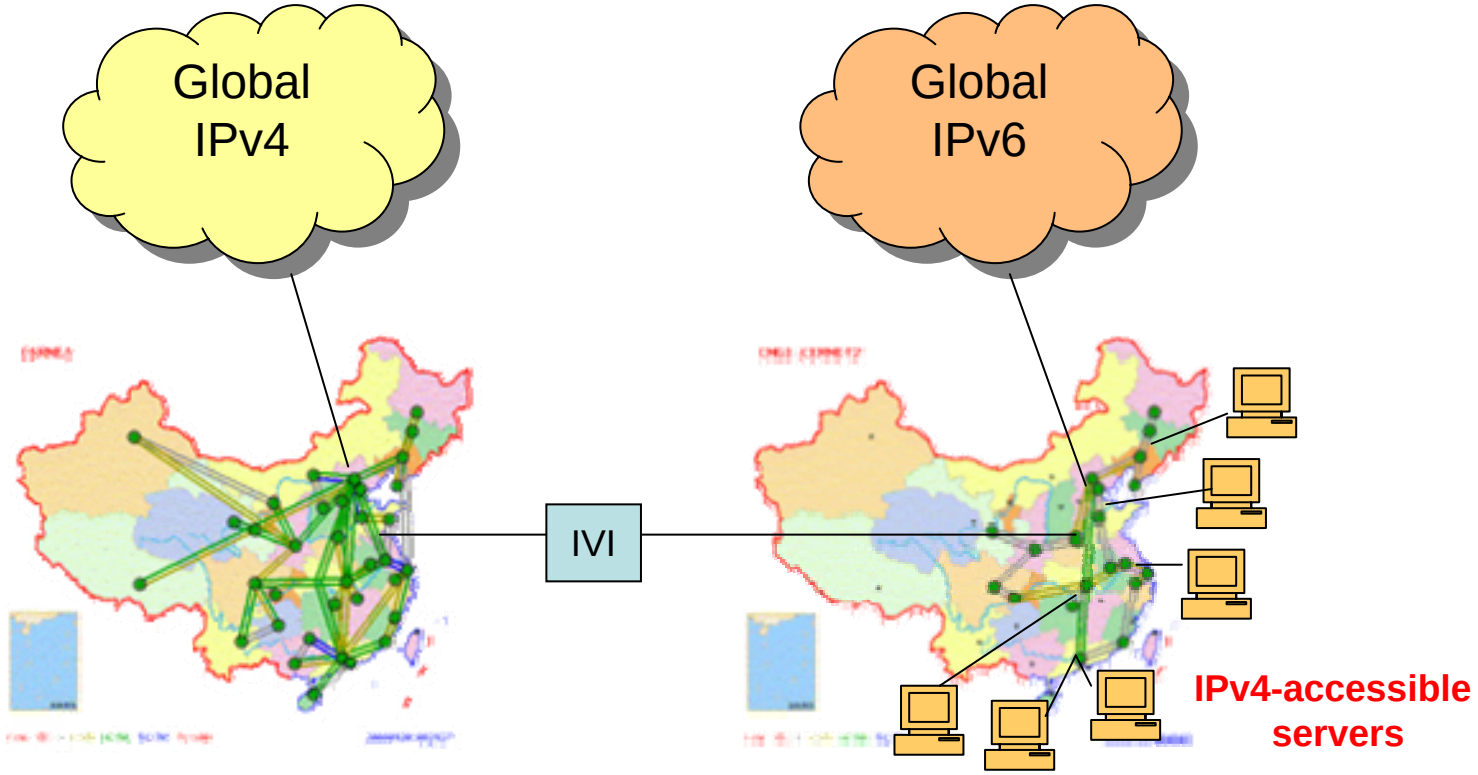
# IVI Update to SIIT and NAT-PT

Xing Li, Congxiao Bao,
Fred Baker, Kevin Yin
2008-10-02

# Outline

- Background
- Objectives
- The IVI model
  - IVI address format
  - DNS service in IVI networks
  - Stateless (1:1) Operation
  - Stateful (1:n) Operation
- Reflections on RFC 4966
- Transition plan

# Background




CERNET (IPv4)

2,000 universities
connected
20M users

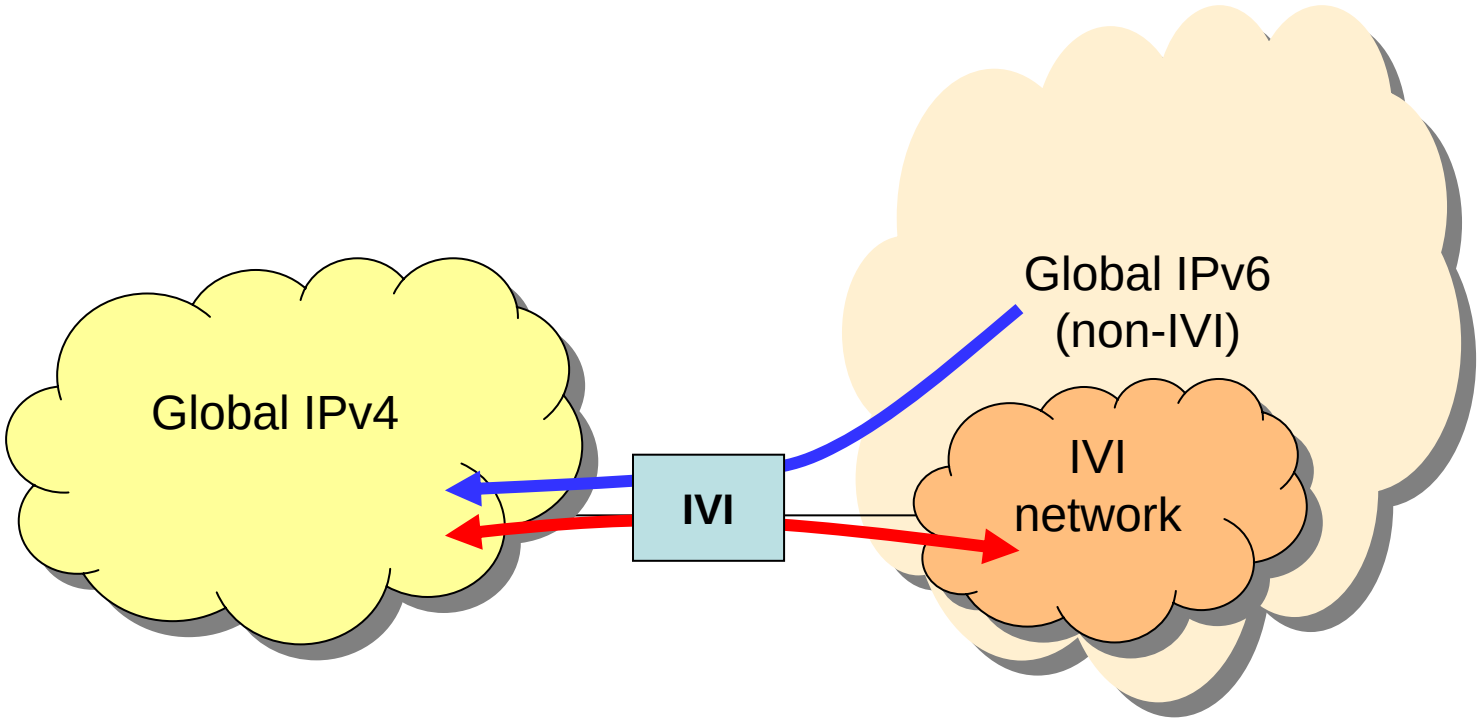
CNGI-CERNET2 (IPv6)

100 universities
connected
400K users

# The lessons learned

- The only viable option for future Internet is IPv6
  - The transitions can only starts when the part of it is pure IPv6
- The scenarios of building new IPv6 network for the unwired population
  - The cost-effective way for building a new infrastructure
- The natural transition
  - Construction and operation single stack costs less than dual-stack
  - Construction and operation simple (stateless) network costs less than complex (stateful) network
- The resources should be shared via inter-communication
  - The IPv6 servers should be IPv4 accessible
  - The IPv4 servers should be IPv6 accessible

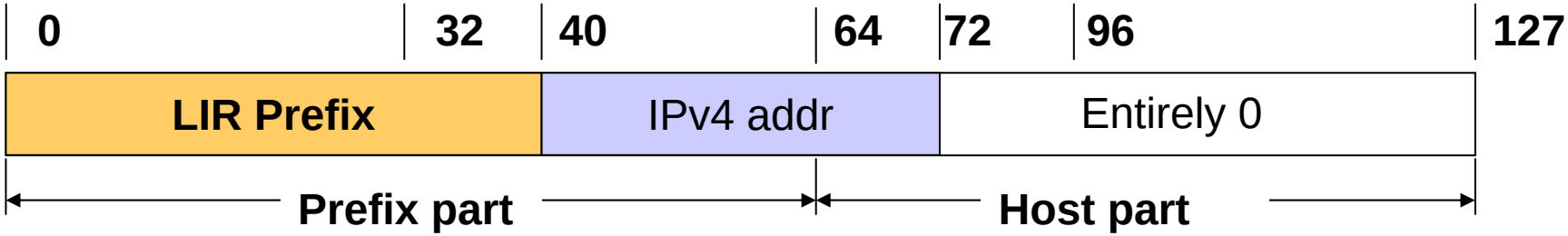
# The IVI model



# Objectives

- Native connectivity (1)
  - IPv4 ←→ IPv4
  - IPv6 ←→ IPv6
- Native connectivity (2)
  - Dual-stack ←→ single-stack (use same address family)
- Translation (1)
  - IPv4 → IPv6 IVI (SIIT extension, stateless)
- Translation (2)
  - IPv6 IVI → IPv4 (SIIT extension, stateless)
  - IPv6 → IPv4 (update to NAT-PT, stateful)

# IVI address format



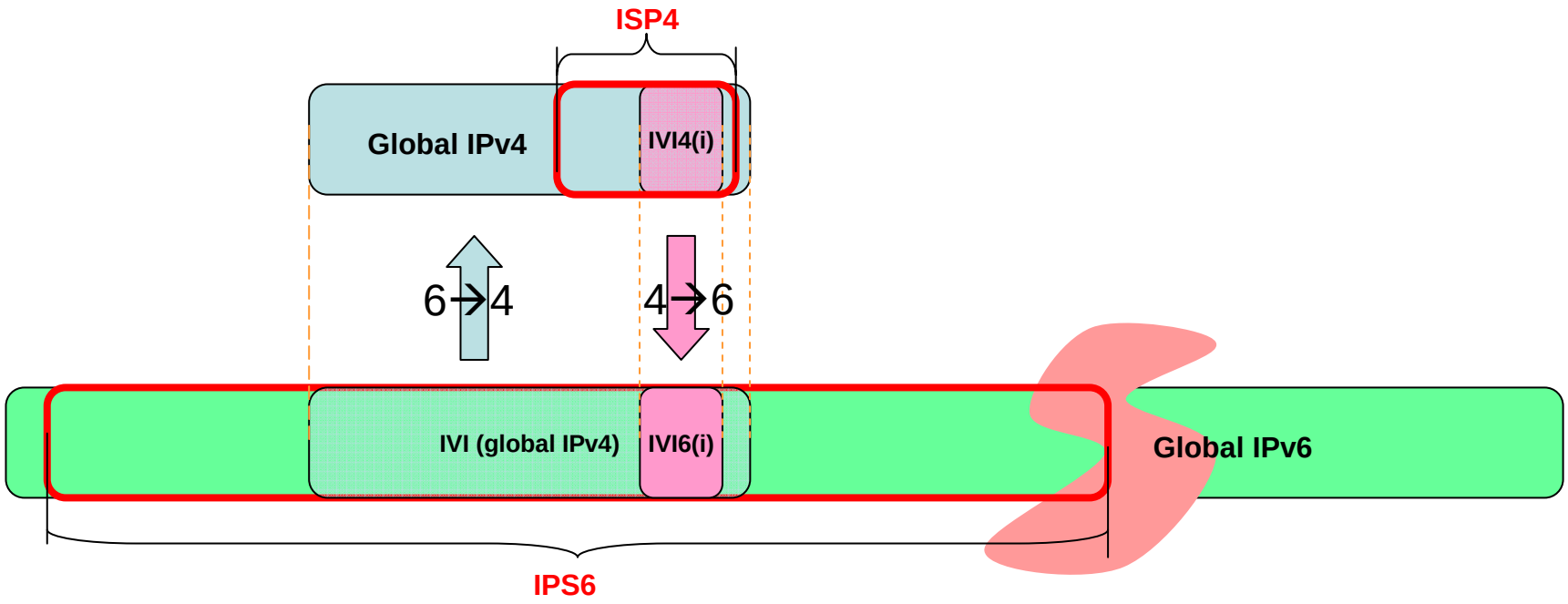

**For example**

**LIR consists of ISP prefix (usually /32) and IVI flag**

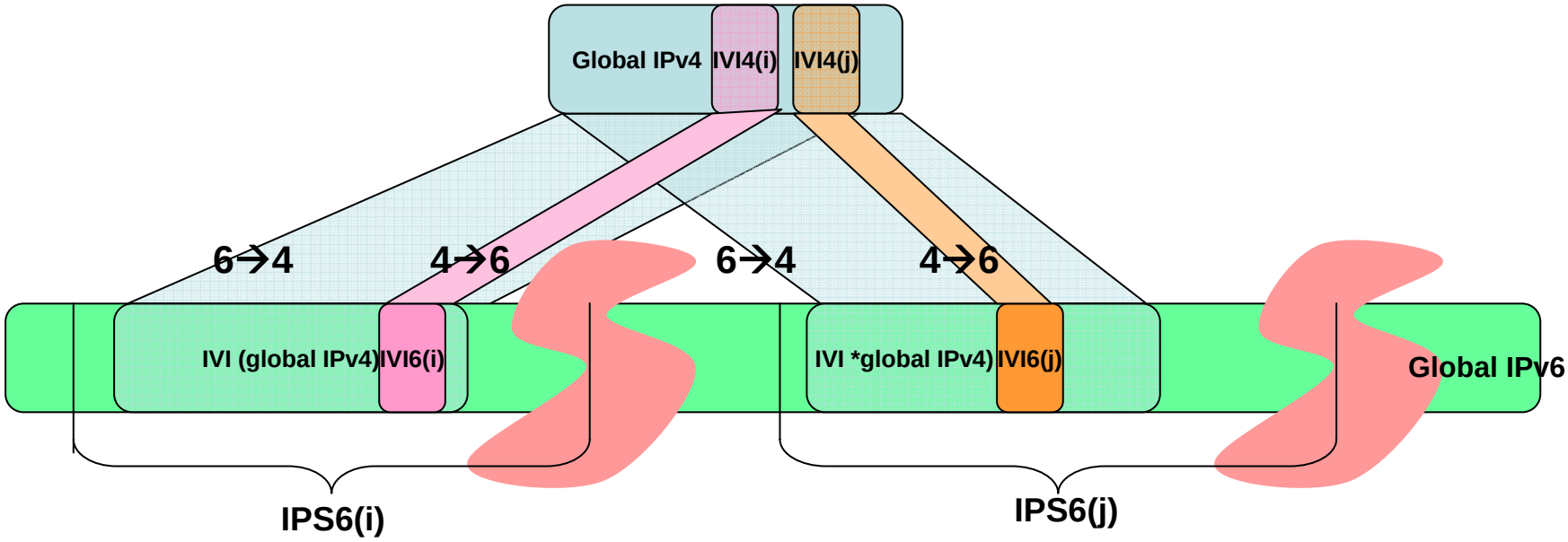
**CERNET/CNGI-CERNET2's selection**

- LIR = 2001:da8:ff00::/40
- **ISP's IVI service IPv4 address mapping**
  - **202.38.108.0/24 → 2001:250:ffca:266c:0000::/64**
- ISP's non-IVI service IPv4 address mapping
  - 202.38.96.0/20 → 2001:250:ffca:2660:0000::/60
- Other ISP's IPv4 address mapping
  - **0.0.0.0 → 2001:da8:ff00::/40**
  - 18.181.0.31/32 → 2001:250:ff12:b500:1f00::/72

# Address space overlay (1)



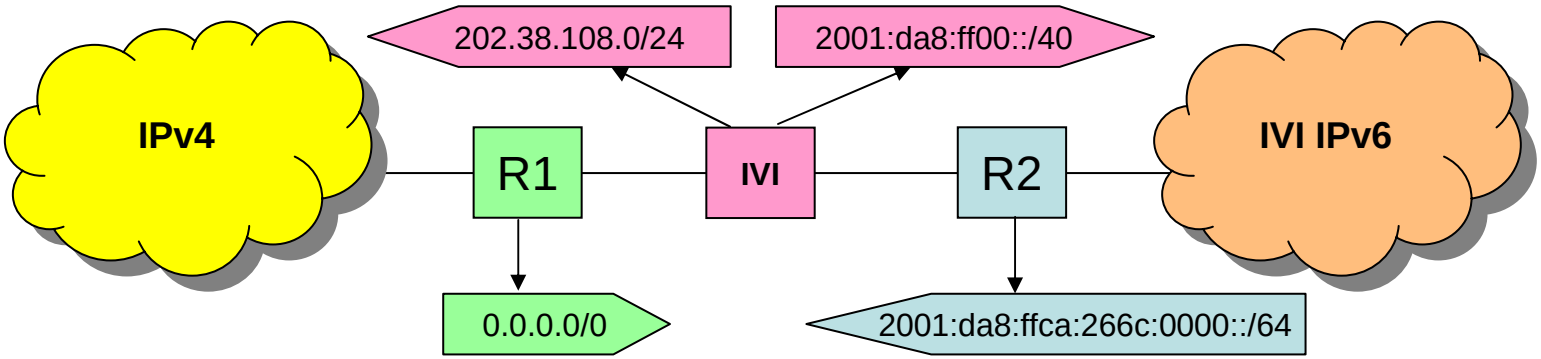
# Address space overlay (2)



# Routing in IVI networks

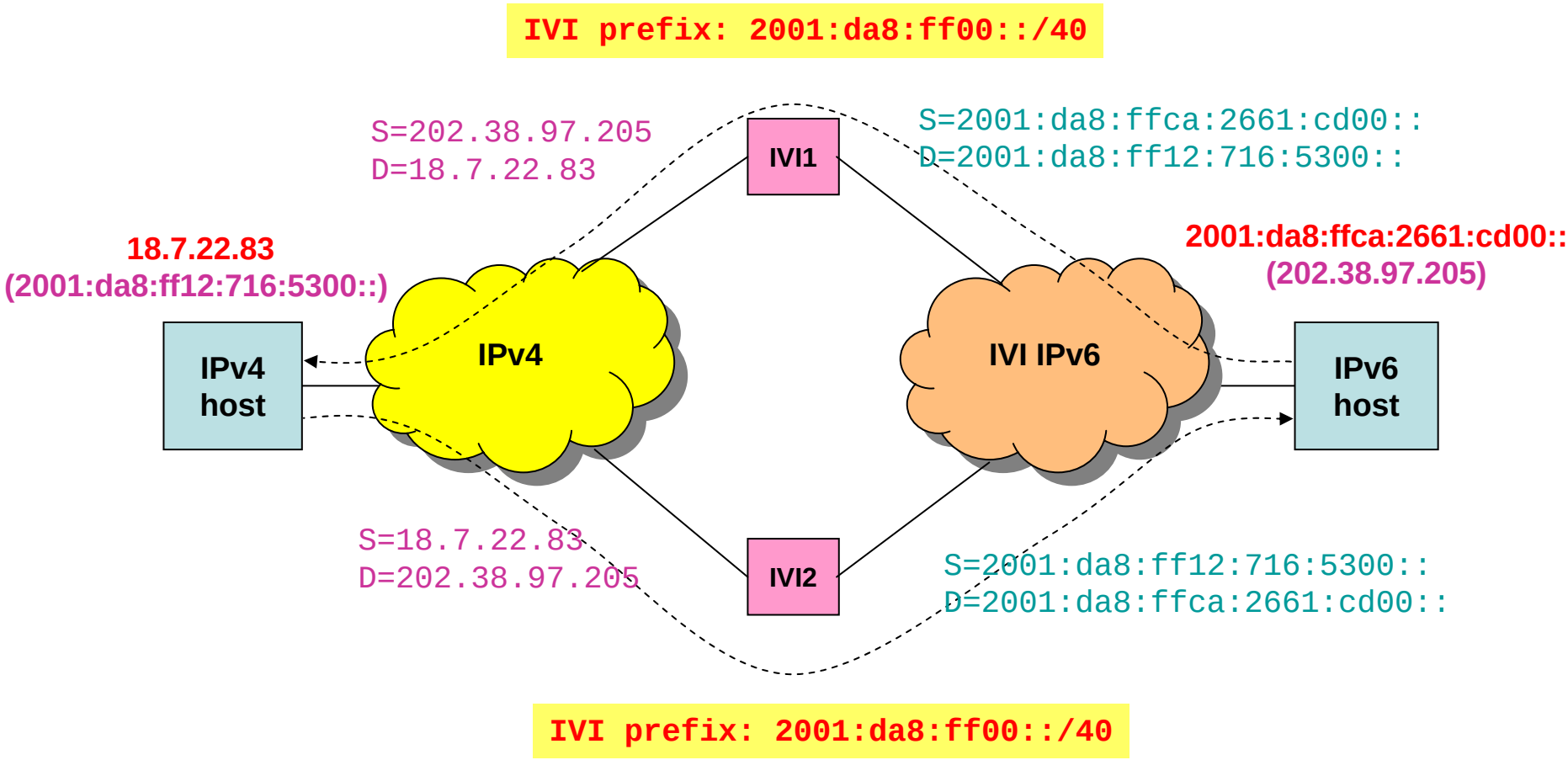
- One IVI gateway
  - Consistent with the Internet routing practice
    - In IPv4, the IVI gateway advertises
      - `202.38.108.0/24 (2001:250:ffca:266c:0000::/64)`
    - In IPv6, the IVI gateway advertises
      - `2001:da8:ff00::/40 (0.0.0.0)`
    - In the IPv6 domain, the routers or hosts advertise
      - `2001:250:ffca:266c:0000::/64 (202.38.108.0/24)`
    - In the IPv4 domain, the router advertises
      - `0.0.0.0`
- Multiple IVI gateways
  - Supporting multihoming
- Multiple IVI domains
  - Supporting incremental deployment

# Prefix announcement of IVI




**Longest prefix match**

# Multiple IVI gateways



# DNS service in IVI networks

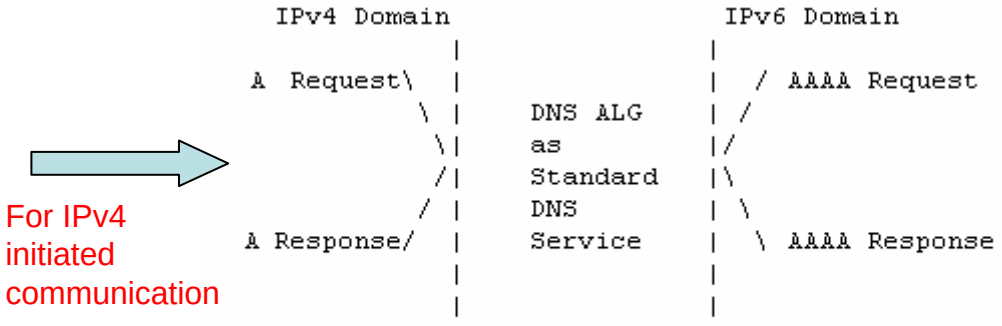



Figure 4: Normal DNS Service

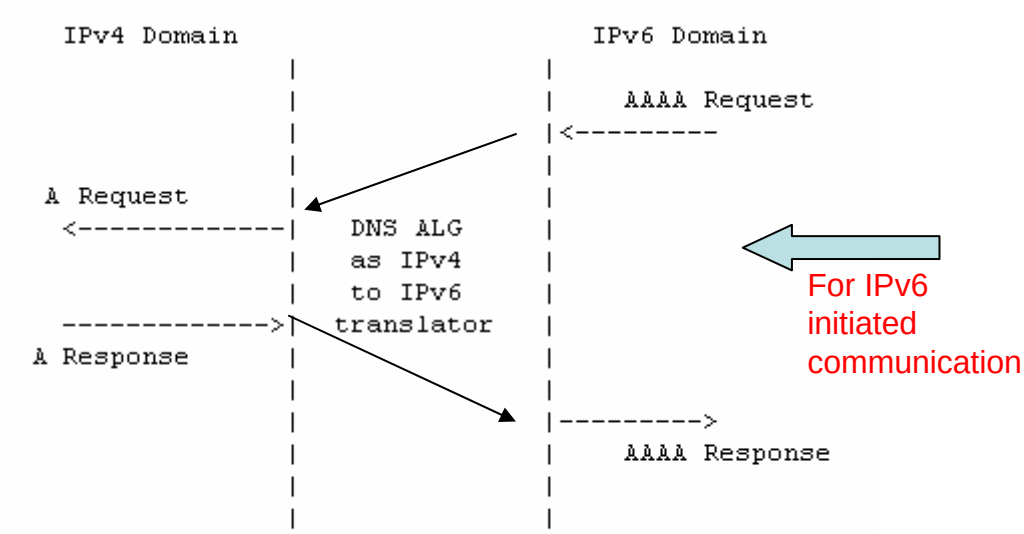



Figure 5: DNS Record Translation Service

- Normal DNS service
  - Based on algorithm, not state
- DNS ALG as IPv4 to IPv6 translator
  - Based on algorithm, not state

# Stateless (1:1) Operation

- In the stateless mode, the IVI gateway translates datagram exchanged between IPv4 systems and IPv6 systems that have an IVI address.
  - SIIT extension
    - The address format is prefix specific address (e.g. not IPv4 Compatible address)
    - The transformation between IPv4 and IPv6 communication is entirely algorithmic and requires no long-term state in either the hosts or the gateway.

# Stateful (1:n) Operation

- In the stateful mode, the IVI gateway operates as a standard Network Address Translator, but between IPv4 and IPv6 domains.
  - Replacement of NAT-PT (NAPT-PT)
    - The address format is prefix specific address (e.g. not IPv4 Compatible address)
    - IPv4 addresses and port numbers are mapped to IPv6 addresses in a stateful manner
    - It is unidirectional (IPv6 initiated communication)
      - The source port in an IPv6 → IPv4 translation may have to be changed to provide adequate flow identification,
      - The source port in the IPv4 → IPv6 direction does not need to be changed

# Operation of the IVI Gateway

- Native (IPv6 ←→ IPv6 or IPv4 ←→ IPv4) communications are preferable to any form of translation
  - This derives from the End-to-End principle discussed in [Saltzer] - the utility of the network to the applications that use it is generally maximized by staying out of their way.
- Stateless translation is preferable to stateful translation
  - The Simplicity Principle discussed in [RFC3439]; given an easy and a hard way to do something, and given equivalence of outcome, the easy way is generally better for all concerned
  - Stateful operation requires supporting dynamically-created per-flow tables in the gateway while stateless operation transforms datagram algorithmically without per-flow state.

# IVI Reachability

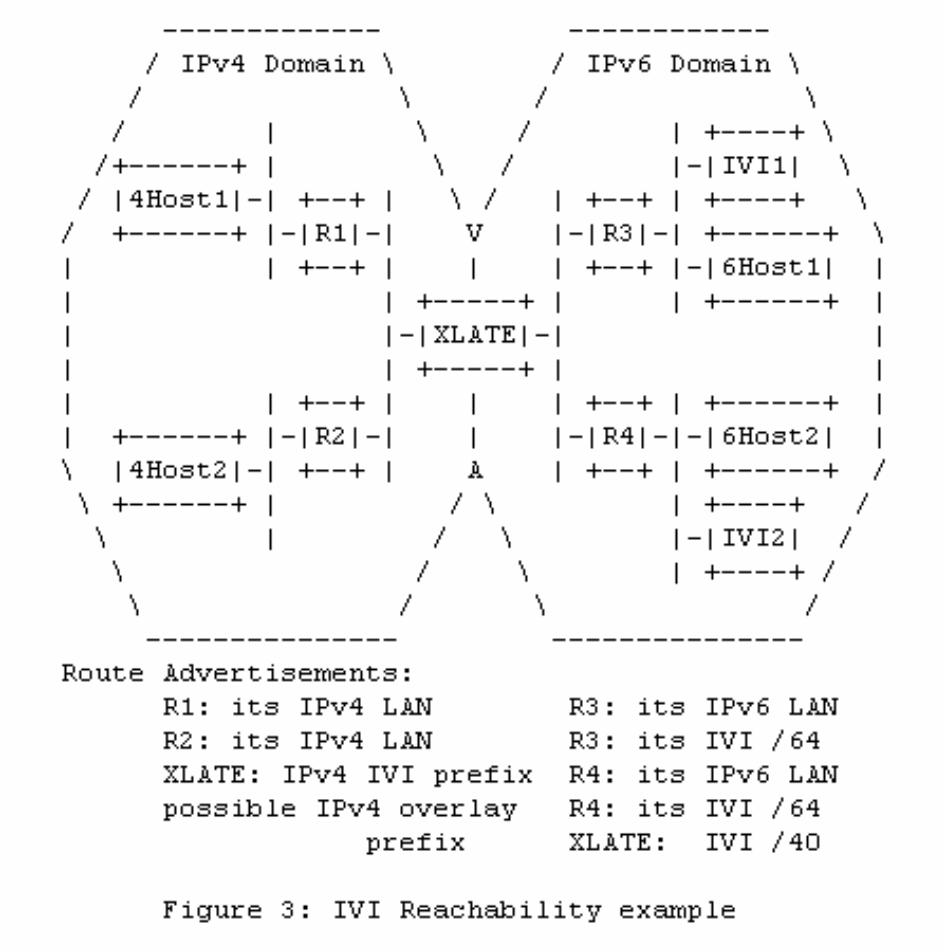



Figure 3: IVI Reachability example

- 4Host1 → IVI1 (stateless)
- IVI1 → 4Host1 (stateless)

- 4Host1 —X 6Host1
- 6Host1 → 4Host1 (stateful)

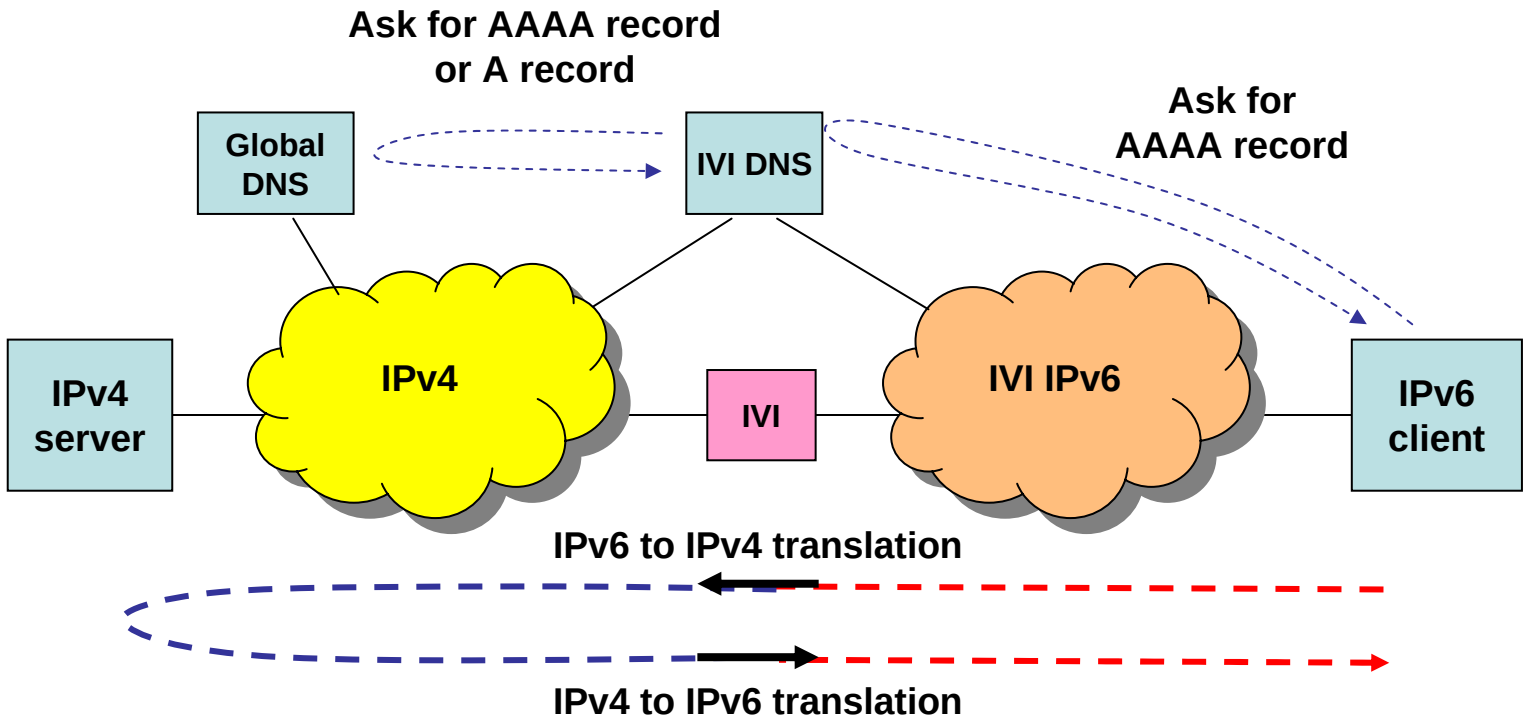
- 4Host1 → 4Host2
- 4Host2 → 4Host1

- 6Host1 → IVI1
- IVI1 → 6Host1

- 6Host1 → 6Host2
- 6Host2 → 6Host1

- IVI1 → IVI2
- IVI2 → IVI1

# IPv6 initiated communication (1:1)



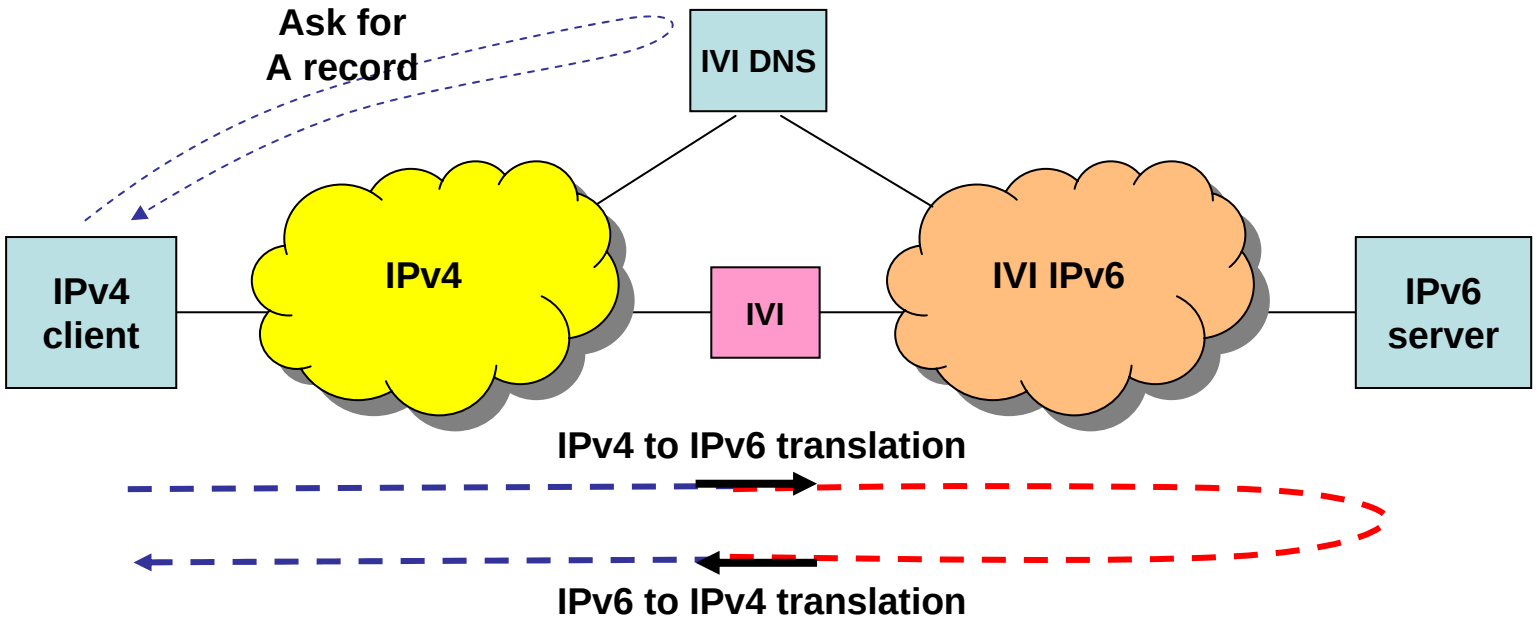
# IPv6 initiated communication (1:1)

- Goal
  - IPv6 client (2001:250:ffca:266c:0500::) → IPv4 server (www.mit.edu)

- DNS query
  - www.mit.edu → A record (18.7.22.83)
  - 18.7.22.83 → AAAA record (2001:250:ff12:0716:5300::)
- In IPv6 network
  - src=2001:250:ffca:266c:0500:: dst=2001:250:ff12:0716:5300::
- In IVI gateway
  - src=2001:250:ffca:266c:0500:: dst=2001:250:ff12:0716:5300::
  - src=202.38.108.5 dst=18.7.22.83
- In IPv4 network
  - src=202.38.108.5 dst=18.7.22.83
  - src=18.7.22.83 dst=202.38.108.5
- In IVI gateway
  - src=18.7.22.83 dst=202.38.108.5
  - src=2001:250:ff12:0716:5300:: dst=2001:250:ffca:266c:0500::

- Repeats, until the session terminates.

# IPv4 initiated communication (1:1)



# IPv4 initiated communication (1:1)

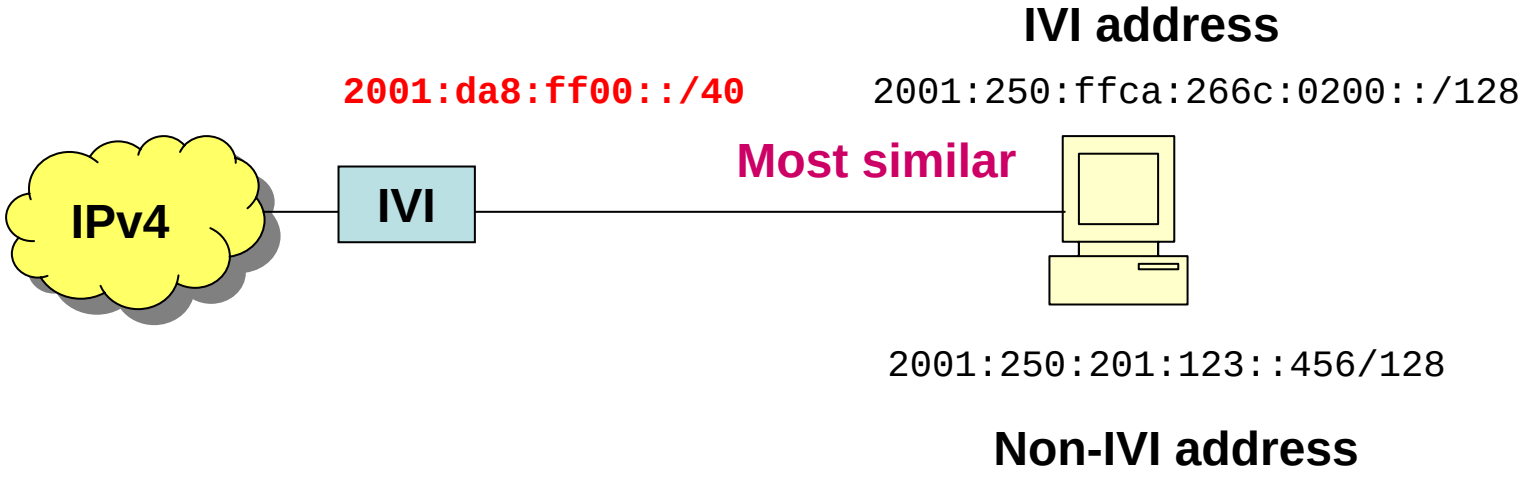
- Goal
  - IPv4 client (59.66.24.42) → ivi.sasm3.net

- DNS query
  - ivi.sasm3.net → A record (202.38.114.1)
  - ivi.sasm3.net → AAAA record (2001:250:ffca:2672:0100::0)
- In IPv4 network
  - src=59.66.24.42 dst=202.38.114.1
- In IVI gateway
  - src=59.66.24.42 dst=202.38.114.1
  - src=2001:250:ff3b:4218:2a00:: dst=2001:250:ffca:2672:0100::0
- In IPv6 network
  - src=2001:250:ff3b:4218:2a00:: dst=2001:250:ffca:2672:0100::0
  - src=2001:250:ffca:2672:0100::0 dst=2001:250:ff3b:4218:2a00::
- In IVI gateway
  - src=2001:250:ffca:2672:0100::0 dst=2001:250:ff3b:4218:2a00::
  - src=202.38.114.1 dst=59.66.24.42

- Repeats, until the session terminates.

# Host operation in IVI networks

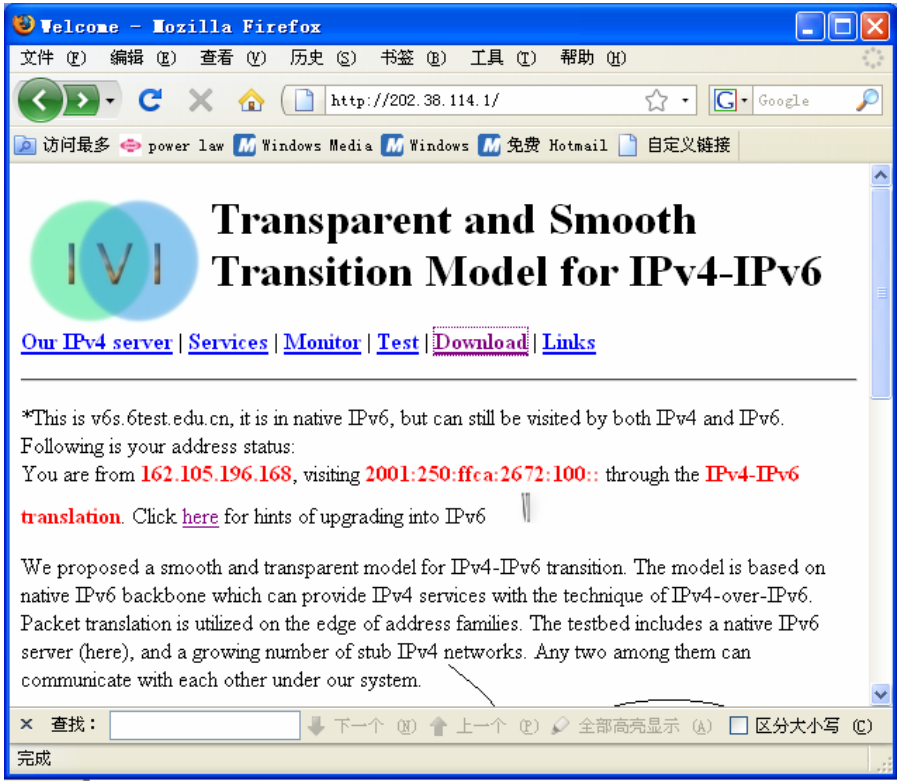
- Interaction of IVI Addresses with RFC3484 Address Selection
  - IPv6 systems should select source and destination addresses that are as similar as possible.
    - IPv6-only systems with IVI addresses to connect from their IVI address when communicating with IPv4-only systems. This is important, because it promotes stateless translation operation.
    - IVI systems may also find the IVI address pair "most similar" when communicating with other systems with IVI addresses.

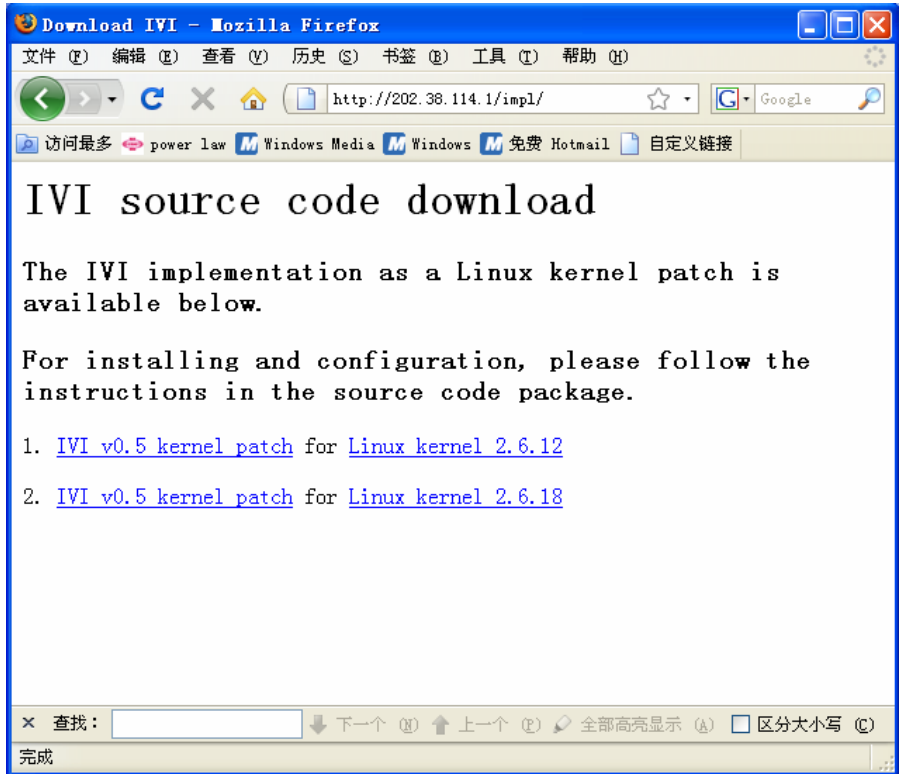


# Reflections on RFC 4966

- The DNS Application layer Gateway
  - The form of the relationship between A/MX and AAAA records is algorithmic and fixed, and the translation is done on the fly without saved state.
- The stateless data plane translation algorithm is essentially that of SIIT apart from the address format.
  - A prefix that lets the LIR specify the upper bits gives the operator the flexibility.
  - Moving the IPv4-mapped portion of the IVI address into the upper 64 bits of the address retains an address format familiar from other IPv6 addresses.
  - The stateless mode applies to every session initiated from the IPv4 side of the gateway.
- The stateful data plane algorithm is similar to that performed in normal IPv4/IPv4 NATs.

# Running code



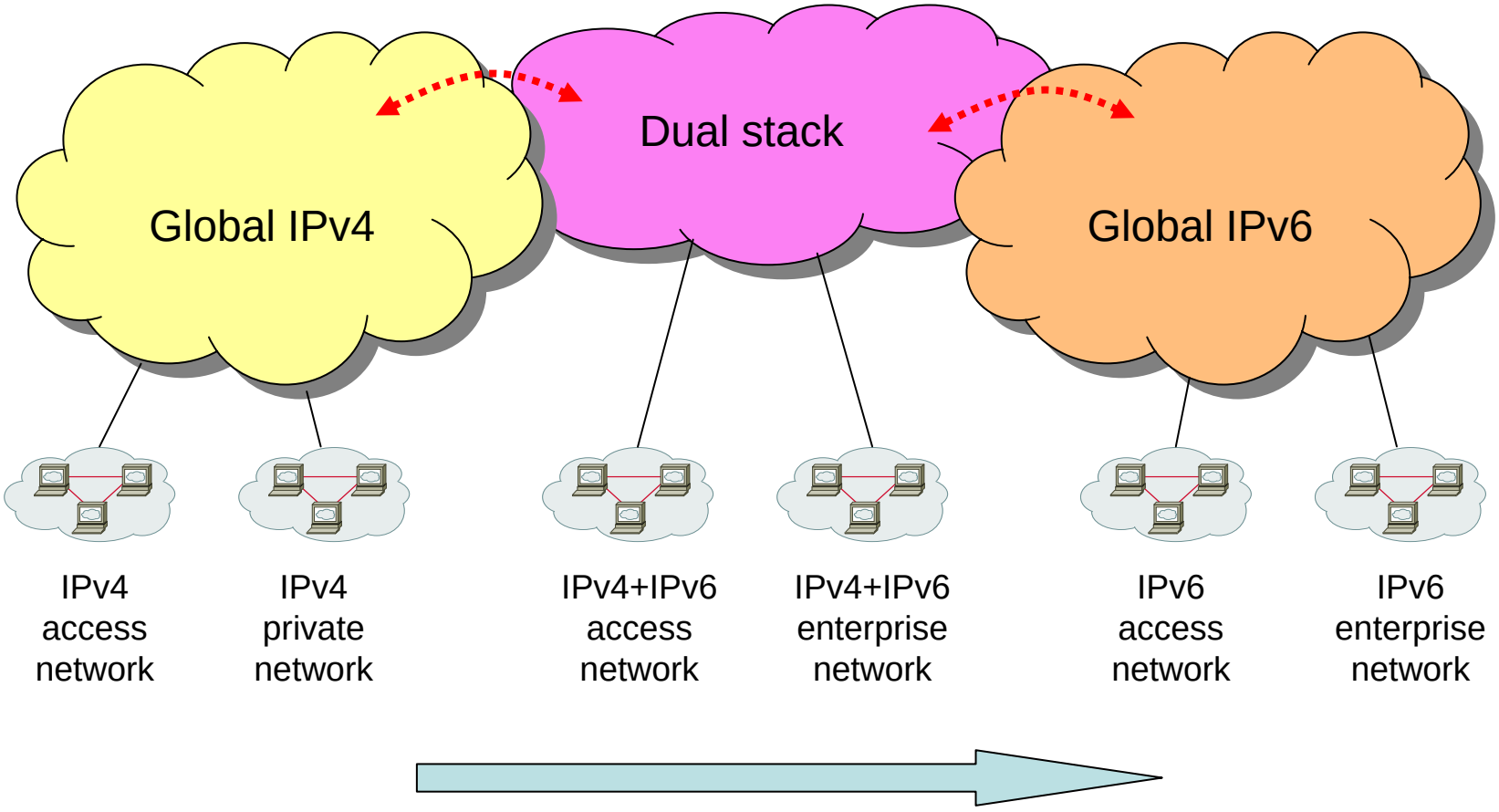




- These three algorithms have two years of operational experience behind them in the CERNET/CNGI-CERNET2 network. In a world of "rough consensus and running code", these are running code.

# When IPv4 addresses are becoming expensive

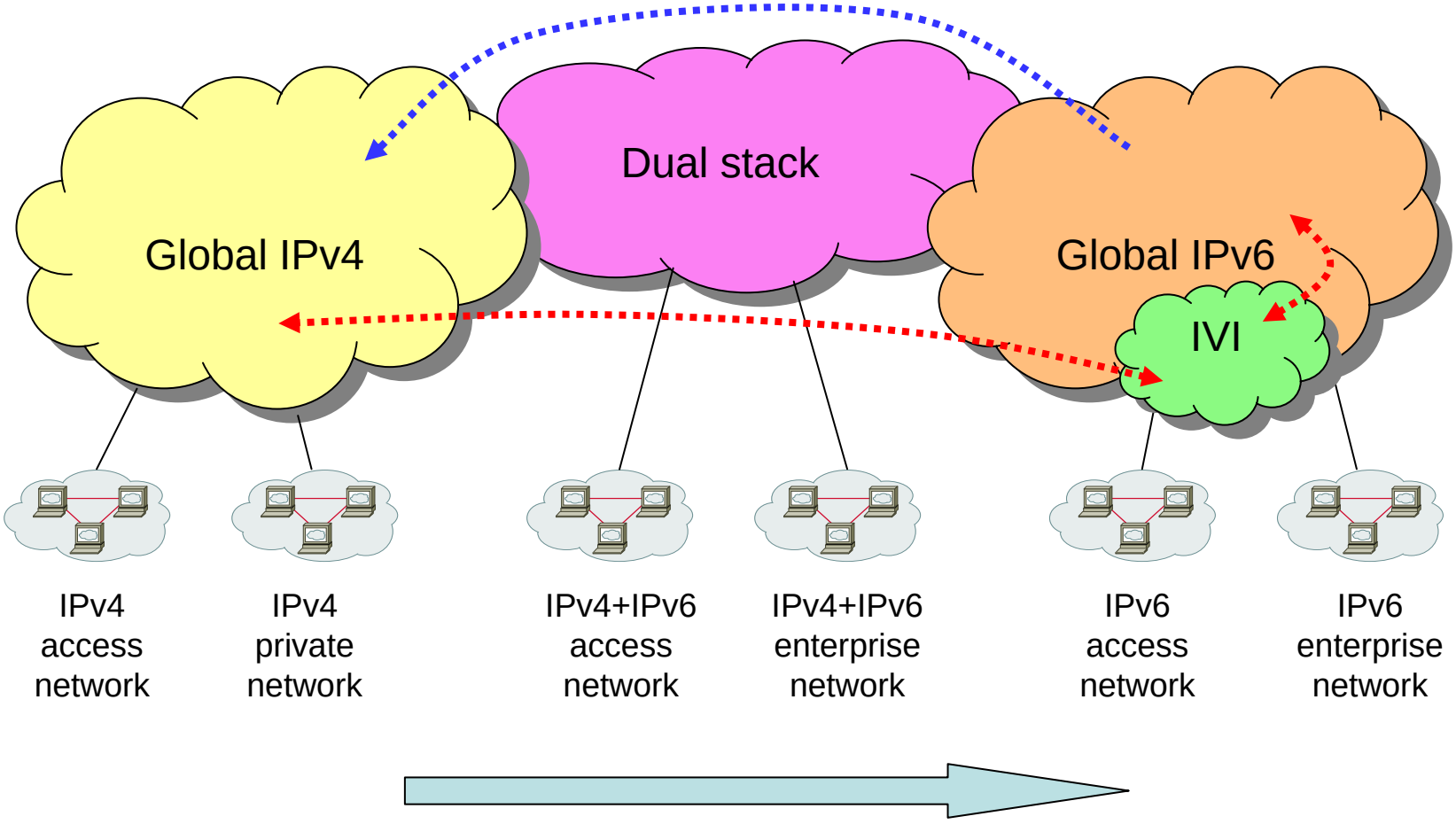
- Four choices:
  - Get no IPv4 addresses, just deploy IPv6
    - CNGI-CERNET2 developed IVI because this proved unacceptable.
  - Get IPv4 addresses and use them to deploy NAT
    - This introduces the state and complexity and is  a short term solution.
  - Get IPv4 addresses and use them to deploy a dual stack network
    - The ISPs and edge networks, each presuming the other will deploy the dual stack and bear the cost.
  - Get  IPv4 addresses and use them to sell an IVI + general IPv6 service
    - Our proposal: IPv6+IPv4-accessible Network. This allows one to deploy IPv6 to everyone and IVI addresses to those relatively few systems that require accessibility from the IPv4 network. It provides a middle ground that is economically more feasible.
    - Servers need 1:1 stateless

# The ideal next step



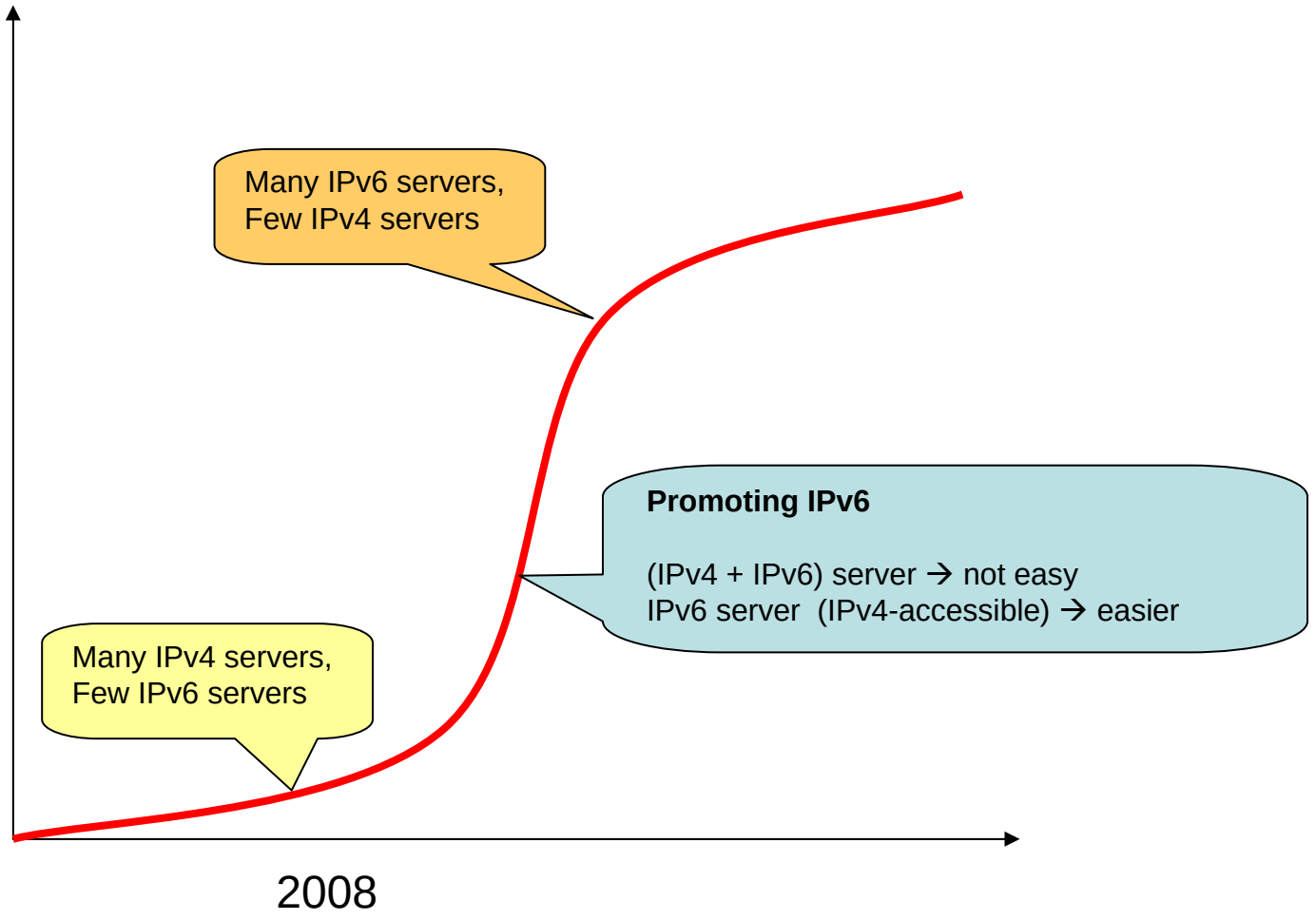

- The ISPs and edge networks, each presuming the other will deploy the dual stack and bear the cost.
- IPv4 servers inaccessible by IPv6
- IPv6 servers inaccessible from IPv4

# The IVI next step




•Deploy an IPv6-only network in which some systems have IPv4-accessible addresses (IVI).

# IPv6 adoption curve



# Summary

- The IVI is needed to update NAT-PT entirely and enable the start of IPv6 transition.
- The IVI bi-directional communication is the key point to encourage migration to IPv6.
- The IVI is used for accessing legacy IPv4 service during entire transition.
- The IVI fits in scenario 3, 4, 5.