IVI - The Transition to IPv6

Xing Li,

2009-04-25

大纲

- 背景
- IVI模型
- 过渡方案
- 结论

CERNET IPv6 历史

- May 3, 1998, IPv6 on FreeBSD at Tsinghua University
- Sept. 1998 first BGP4+ peer to Sprint (MRTd)
- Nov. 1998, First 6Bone backbone node in China
- Mar., 2000 IPv6 BBS--6th Galaxy opened
- April 26, 2000 obtained sTLA (2001:250::/35) from APNIC
- Aug., 2000《IPv6原理与实践》published
- Sep. 25, 2000 CERNET joined IPv6 Forum
- Jan. 2002 DRAGONTAP IPv6 IX is running
- Jun. 2002 NSFCNET double stack is running
- Sept. 2003 CERNET2 test nodes are running
- ……

中国网络IP地址告急

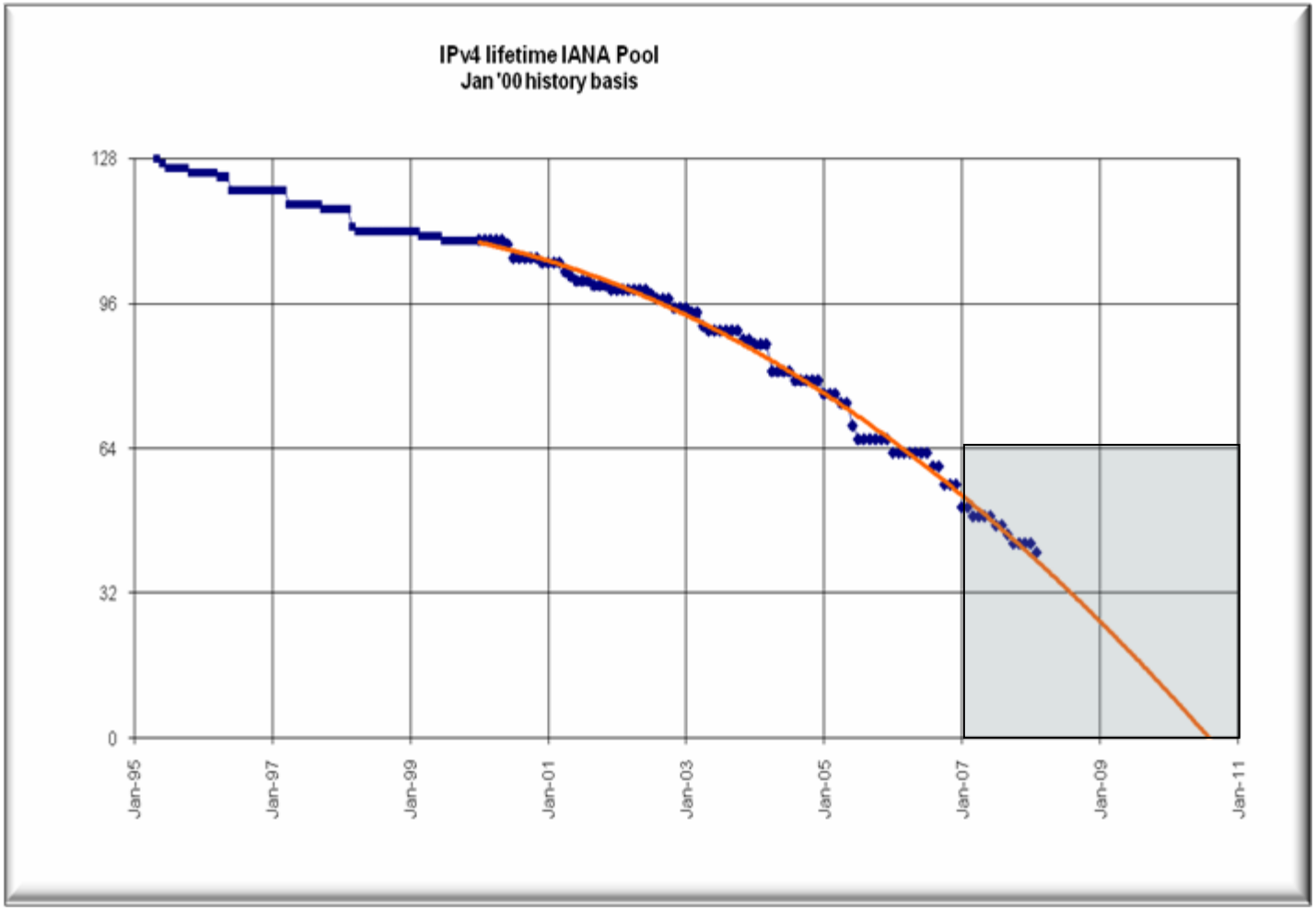
2008年09月23日09:57 来源：《重庆晚报》

近日，由中国互联网络信息中心(CNNIC)等主办的2008IP地址资源研讨会重庆召开。会议透露，IPv4地址资源按照目前的分派速度只剩下830多天。届时，如不采取措施，新网民将无法正常上网。

CNNIC国际业务部IP组业务主管李凯解释，网民要正常上网就必须要有一个IP地址，通过IP地址才能解析域名浏览网页。目前中国绝大部分网络都是使用IPv4的网络地址，作为互联网的基础资源，IPv4的资源是有限的，目前已经用掉了80%。其中中国的使用量近来超过了日本，仅次于美国居世界第二位。按照目前情况，IPv4的网络地址资源只剩下了830多天，这意味着大约到2010年时，如果不使用新的地址资源，新网民将无法正常上网，网络运营商的业务也无法拓展。

李凯介绍，目前美国已经开发了IPv6网络地址，这是一种没有上限的网络基础资源。但是目前中国使用这个地址资源的只有教育网。如果要使用IPv6的网络地址，意味着运营商要使用新的设备，而旧的设备都要被淘汰掉，这需要一笔很大的资金，“我们现在到处都开研讨会，就是要告诉网络运营商尽快申请剩下的IP地址，储备起来，另外要提前为网民准备提供IPv6的IP地址”。

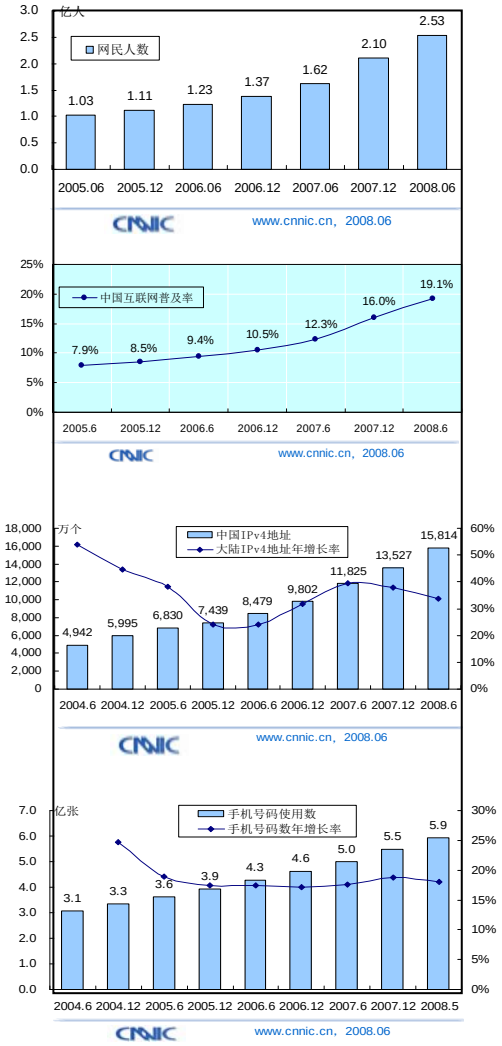
剩余IPv4地址池

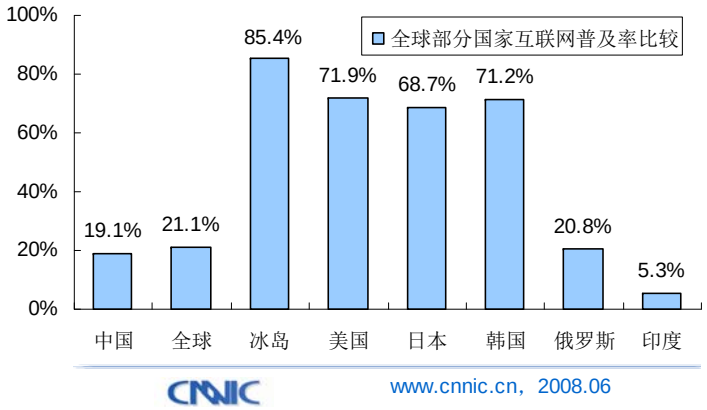


http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_8-3/ipj_8-3.pdf

Tony Hain

需求分析





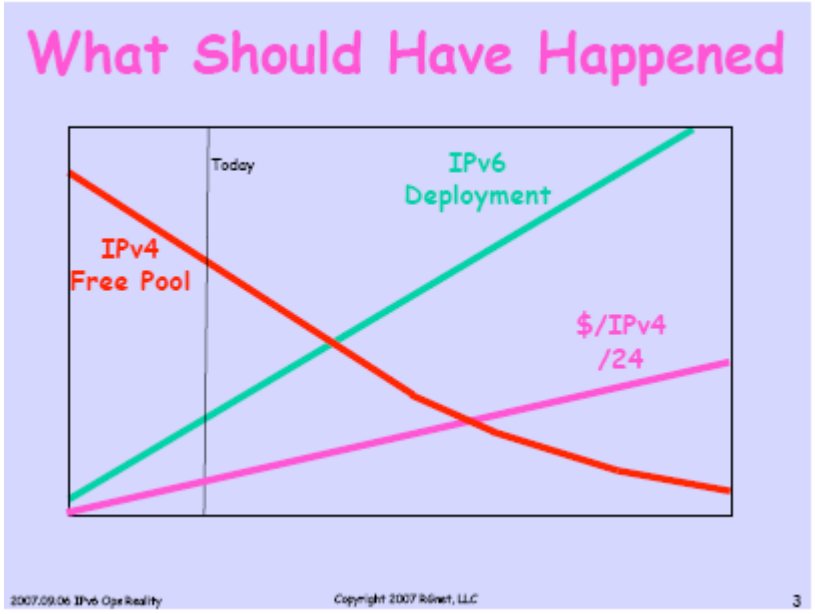
	2008年6月总量（个）	年增长率	每万人拥有量（个/万居民）	每万网民拥有量（个/万网民）
IPv4地址（个）	158,141,184	33.7%	1,197	6,251

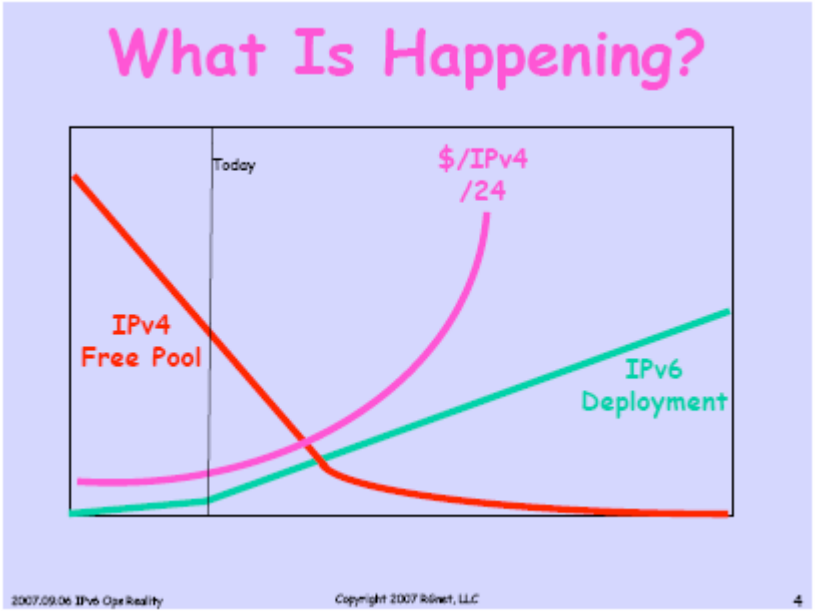
中国网民数：	2.53亿
中国网络普及率：	19.1%
中国IPv4地址数：	1.581亿
中国人均IPv4地址数：	0.6251个
中国手机号码数：	5.9亿
中国人口：	13.2亿

Transition

- IPv4 and IPv6 is NOT compatible
- There is NO flag day
- Transition is REALLY difficult

Unexpected situation





Why Is This Happening?

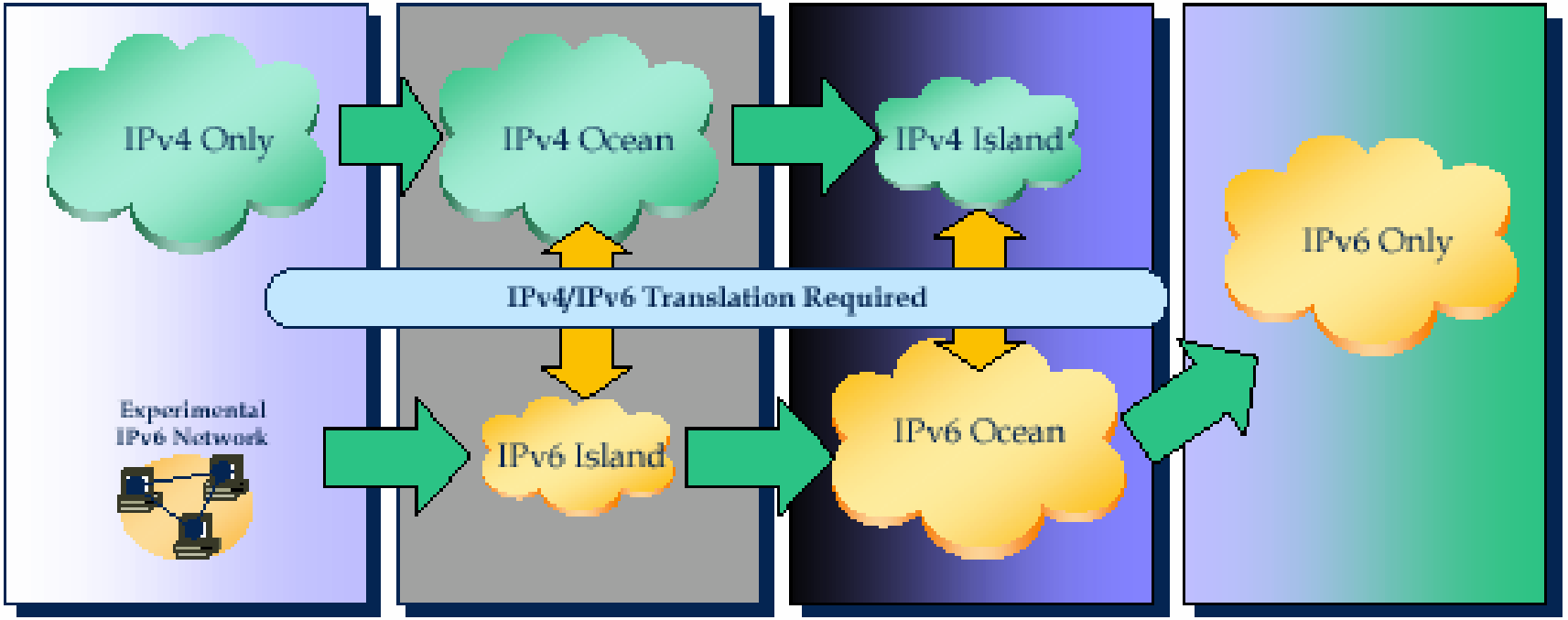
No transition plan
Declared victory before the hard part started
No real long term plan
No realistic estimation of costs
No support for the folk on the front lines
Victory will be next month

Why Is This Happening?

No transition plan
Declared victory before the hard part started
No real long term plan
No realistic estimation of costs
No support for the folk on the front lines
Victory will be next month
This Describes:

a - The invasion of Iraq
b - IPv6
c - DNSSec
d - All of the above

IPv6 过渡


IPv4 Only
IPv4 Ocean
IPv4 Island
IPv6 Only
IPv4/IPv6 Translation Required
Experimental
IPv6 Network
IPv6 Island
IPv6 Ocean

Current transition methods

- Dual stack approach
  - IPv4 address depletion problem
- Tunneled architectures
  - No communication between two address families
- Translation architectures
  - Not scalable, lost end-to-end

The IP infrastructure at crossroad

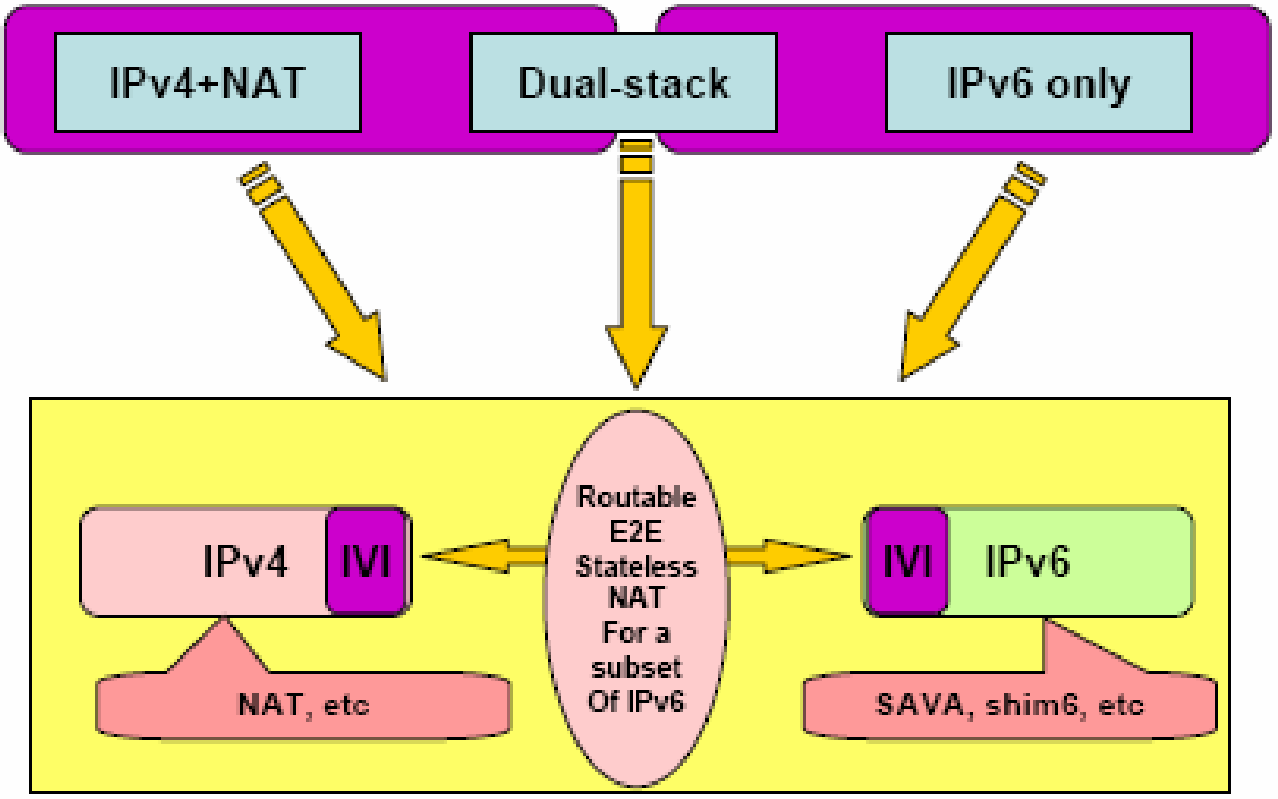
IPv6 only

IPv4/NAT

Something new coming along?

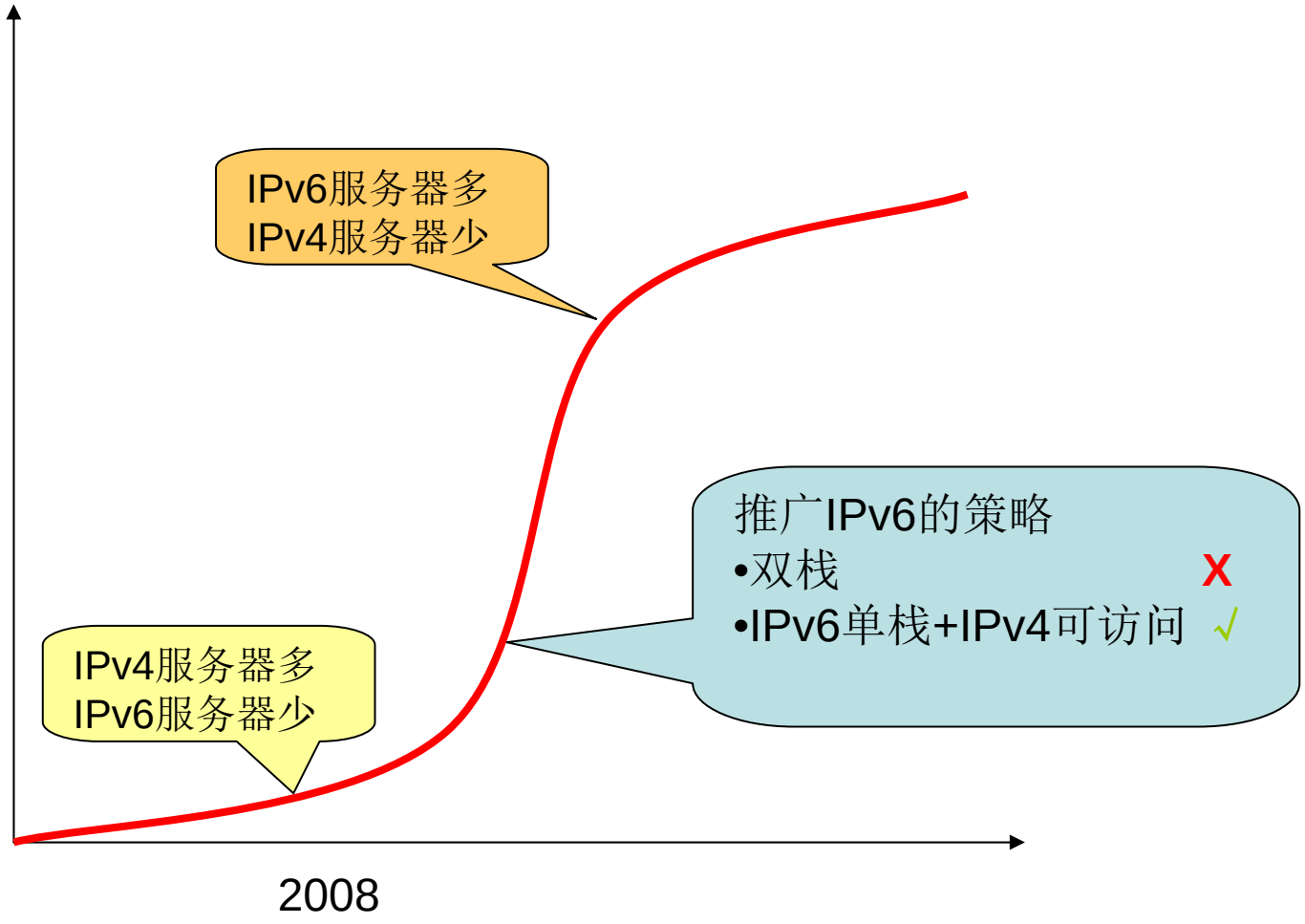
IP4-6 coexistence

Crossroad

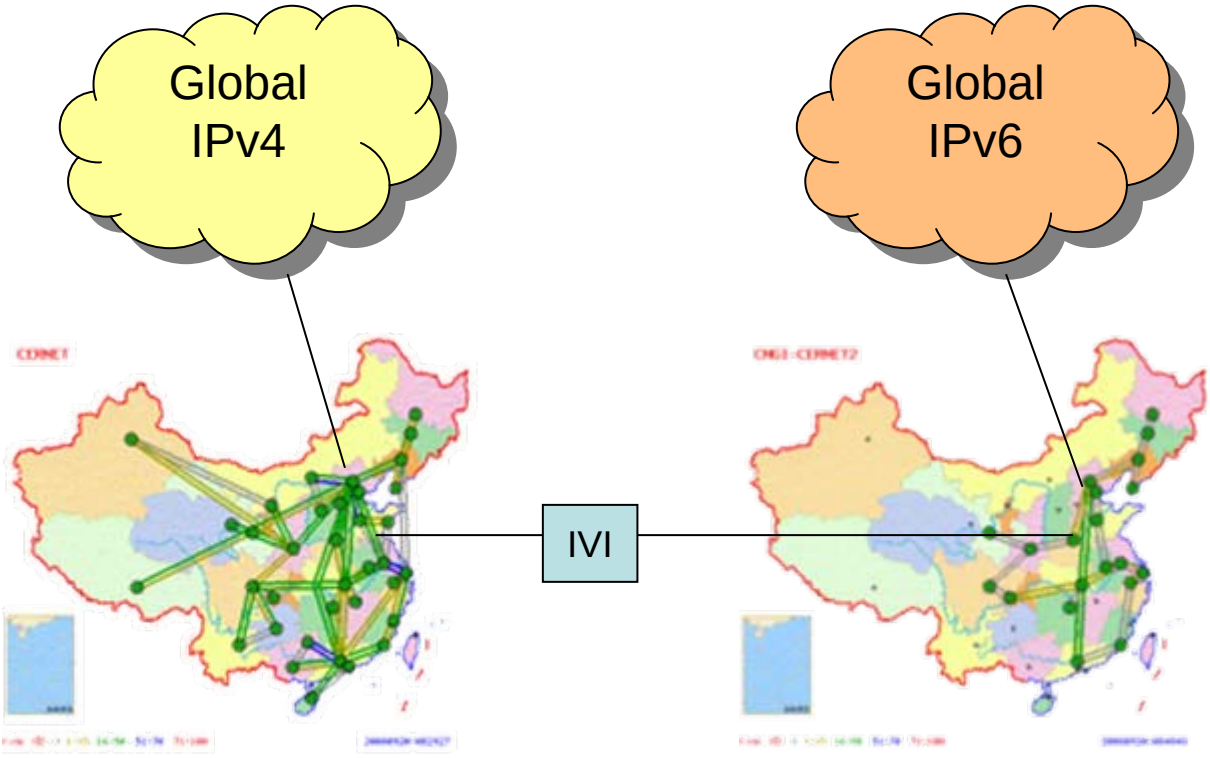


IPv6 Workshop, Sigcomm 2007

IPv6发展曲线


IPv6服务器多
IPv4服务器少
推广IPv6的策略
•双栈 X
•IPv6单栈+IPv4可访问 √
IPv4服务器多
IPv6服务器少
2008

CNGI-CERNET2的实践



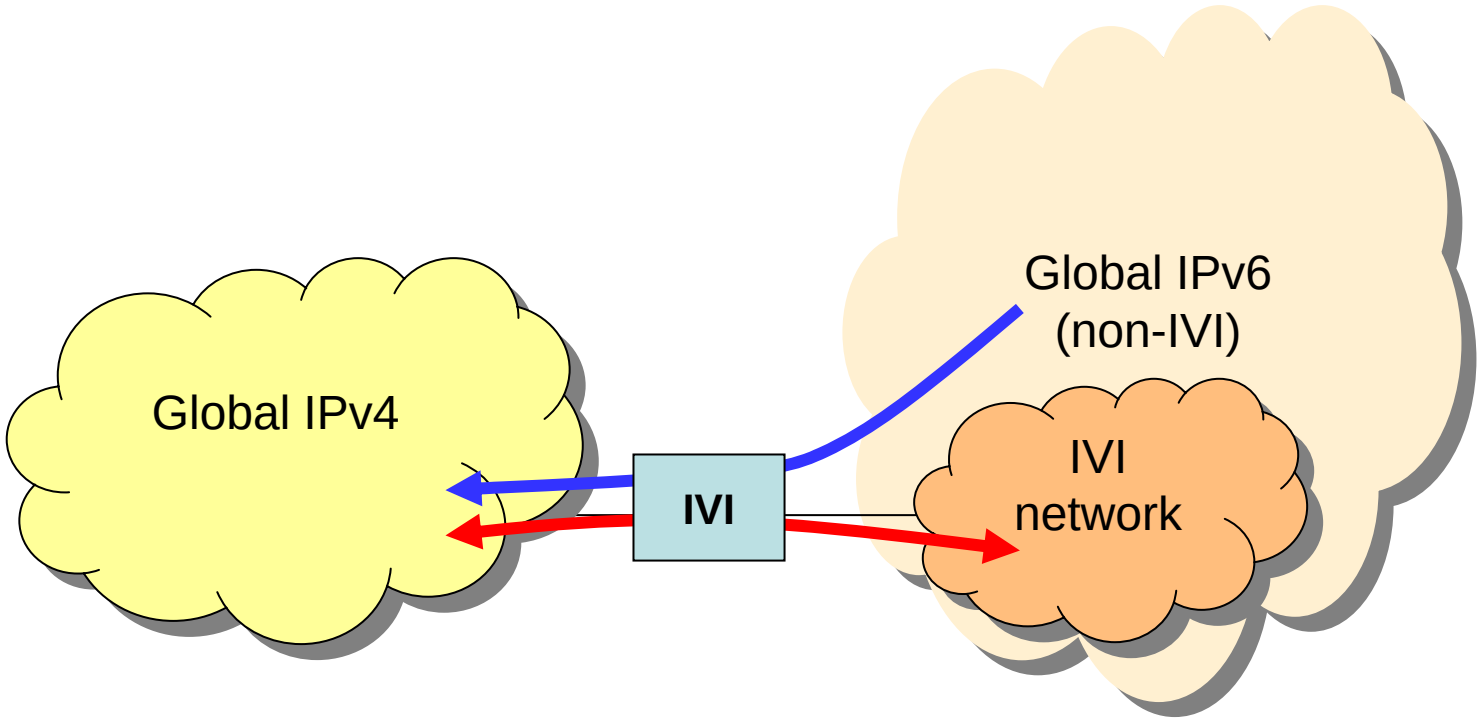
CERNET (IPv4)

CNGI-CERNET2 (IPv6)

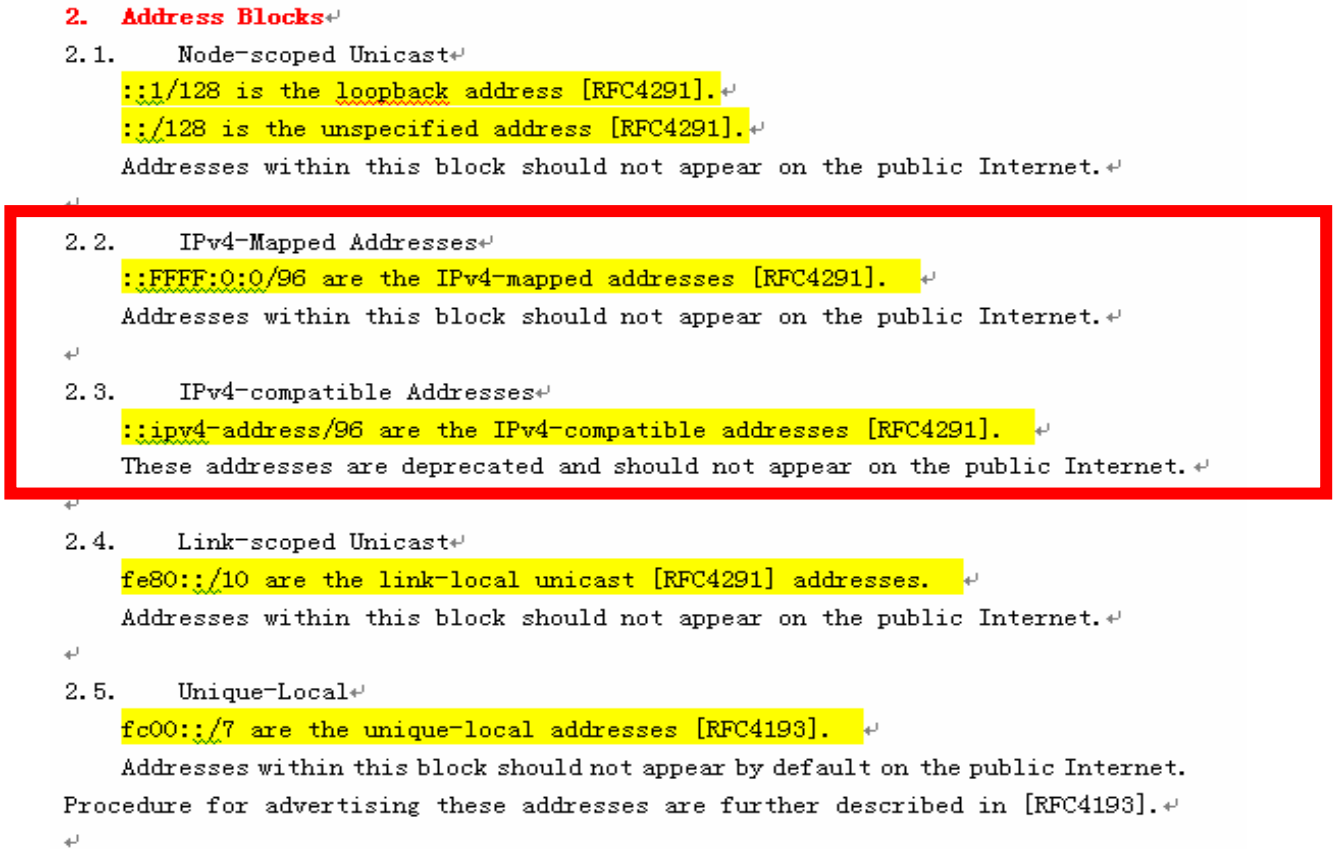
The lessons learned

- 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

IVI模型



The existing mapping methods



2. Address Blocks

2.1. Node-scoped Unicast

::1/128 is the loopback address [RFC4291].

::/128 is the unspecified address [RFC4291].

Addresses within this block should not appear on the public Internet.

2.2. IPv4-Mapped Addresses

::FFFF:0:0/96 are the IPv4-mapped addresses [RFC4291].

Addresses within this block should not appear on the public Internet.

2.3. IPv4-compatible Addresses

::ipv4-address/96 are the IPv4-compatible addresses [RFC4291].

These addresses are deprecated and should not appear on the public Internet.

2.4. Link-scoped Unicast

fe80::/10 are the link-local unicast [RFC4291] addresses.

Addresses within this block should not appear on the public Internet.

2.5. Unique-Local

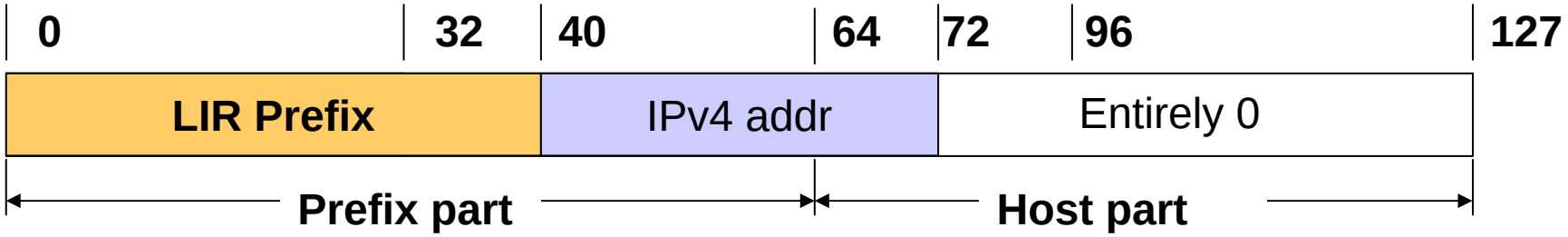
fc00::/7 are the unique-local addresses [RFC4193].

Addresses within this block should not appear by default on the public Internet.

Procedure for advertising these addresses are further described in [RFC4193].

- IPv4-mapped and IPv4-compatible addresses are not routable and cannot achieve the new transition goal

IVI 地址格式

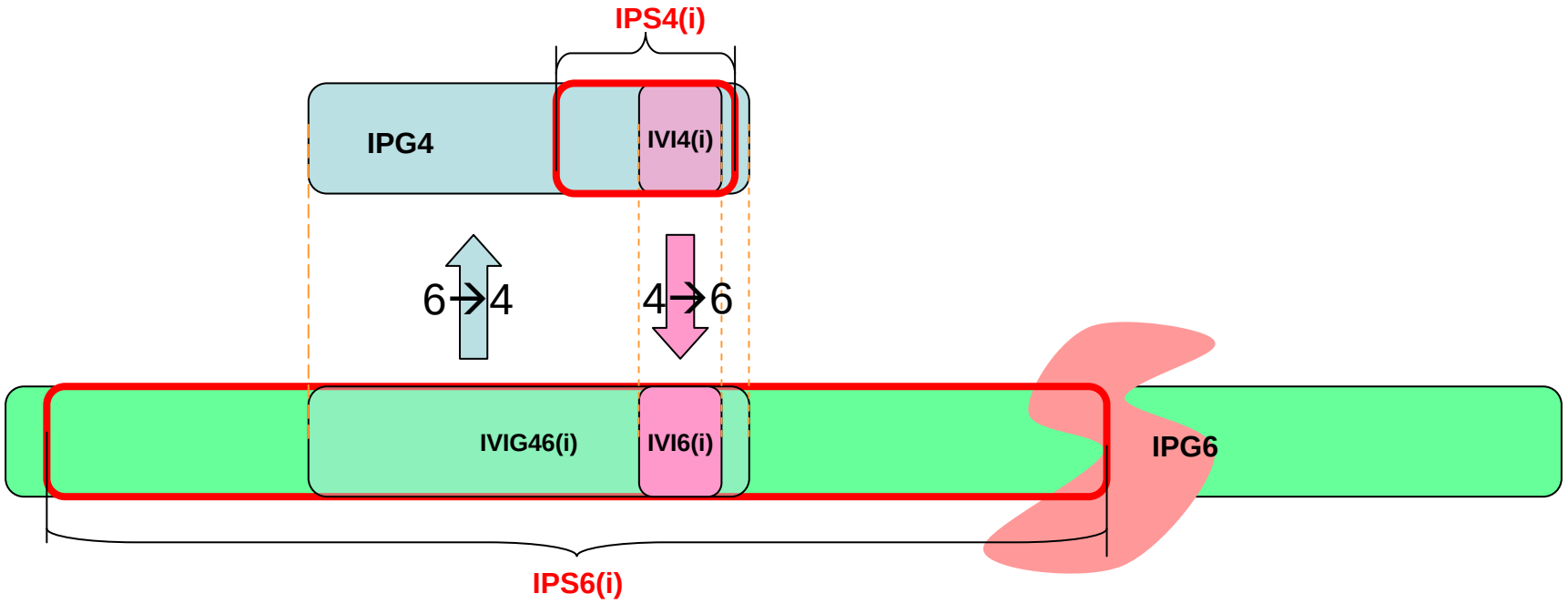


For example

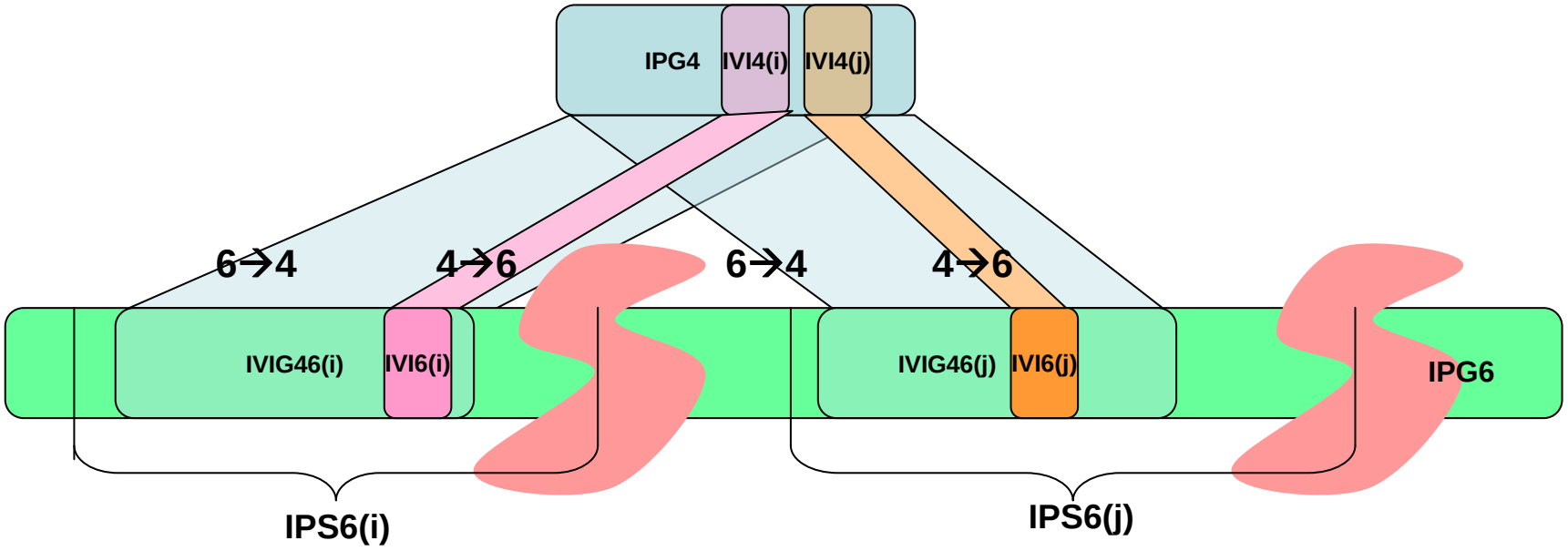
LIR consists of ISP prefix (usually /32) and IVI flag
CERNET/CNGI-CERNET2's selection

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

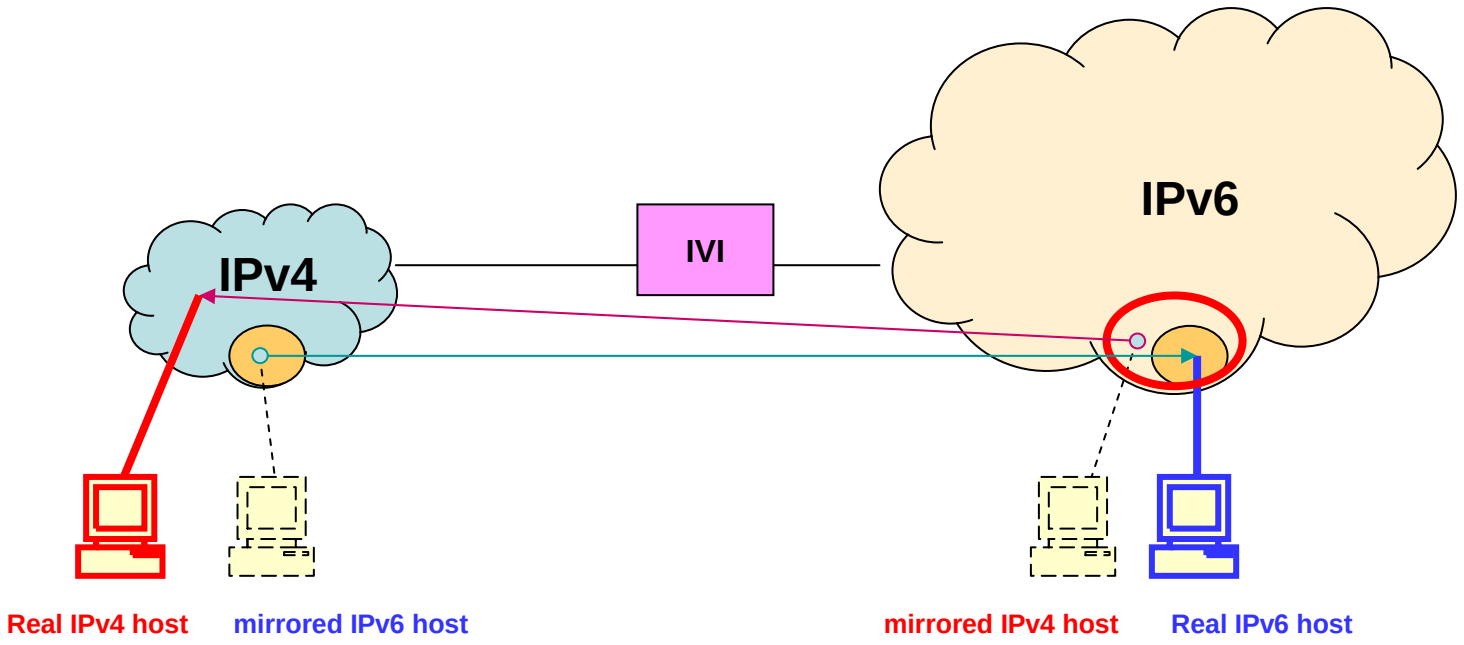
Address Mapping (1)



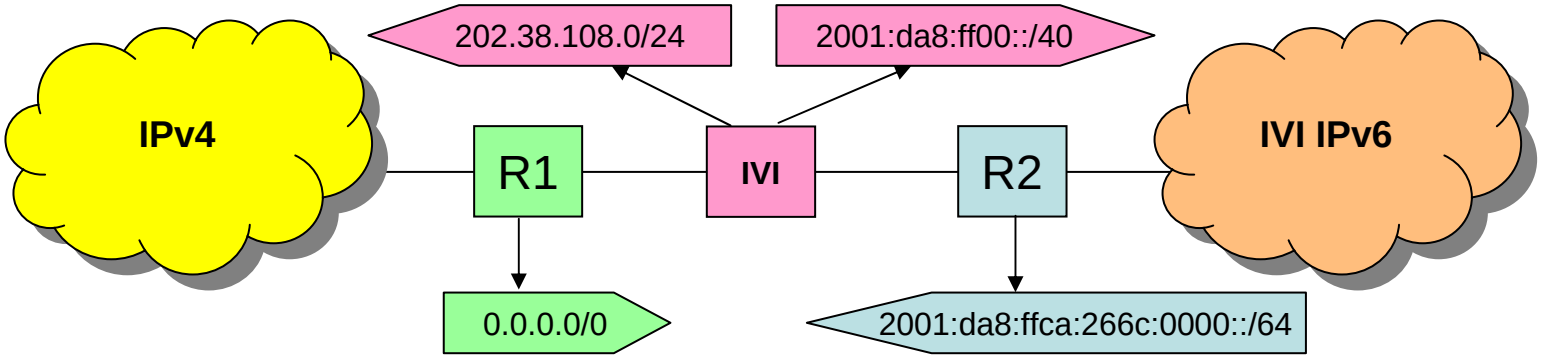
Address Mapping (2)



Conceptual example

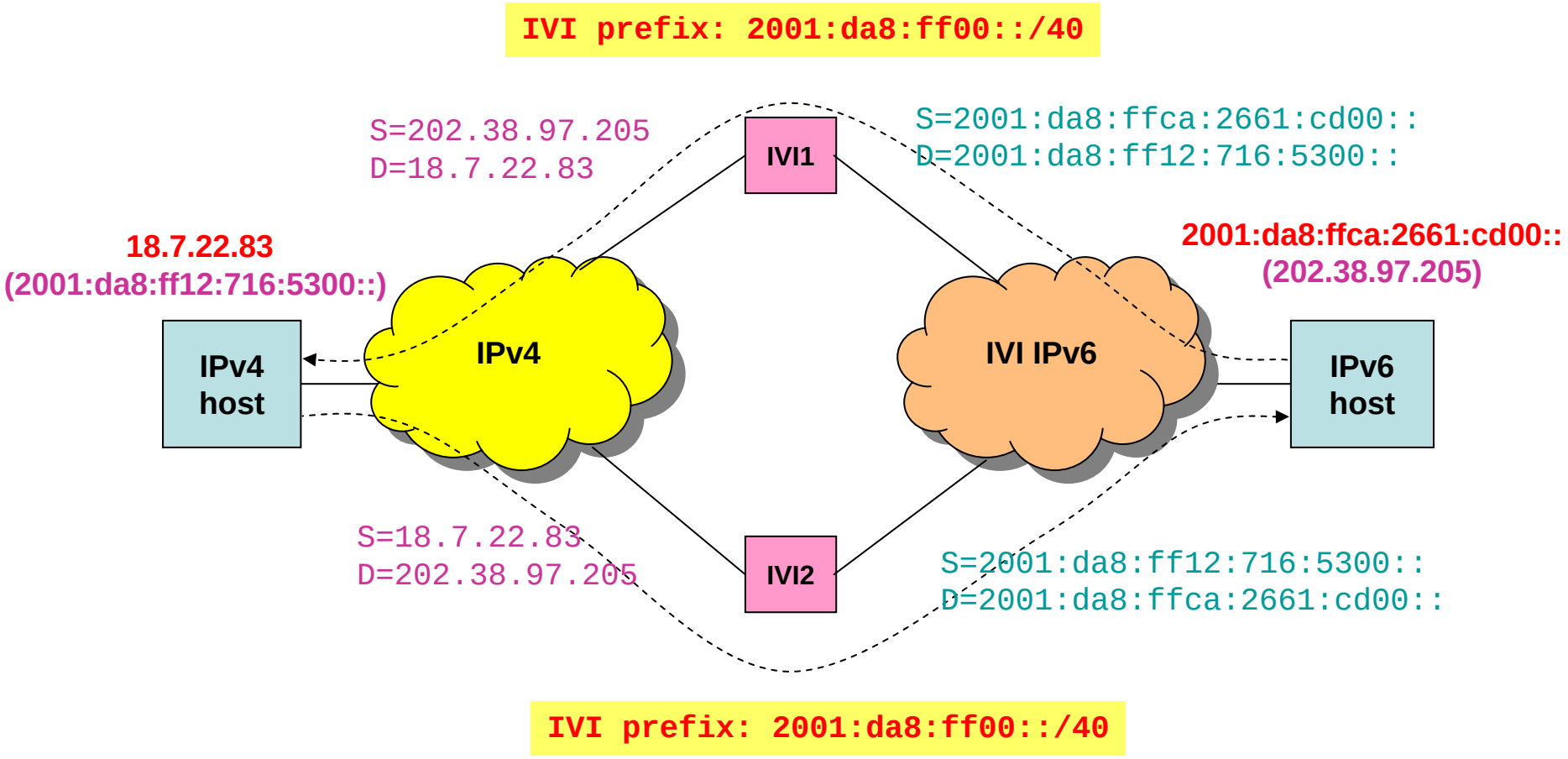


单个IVI网关

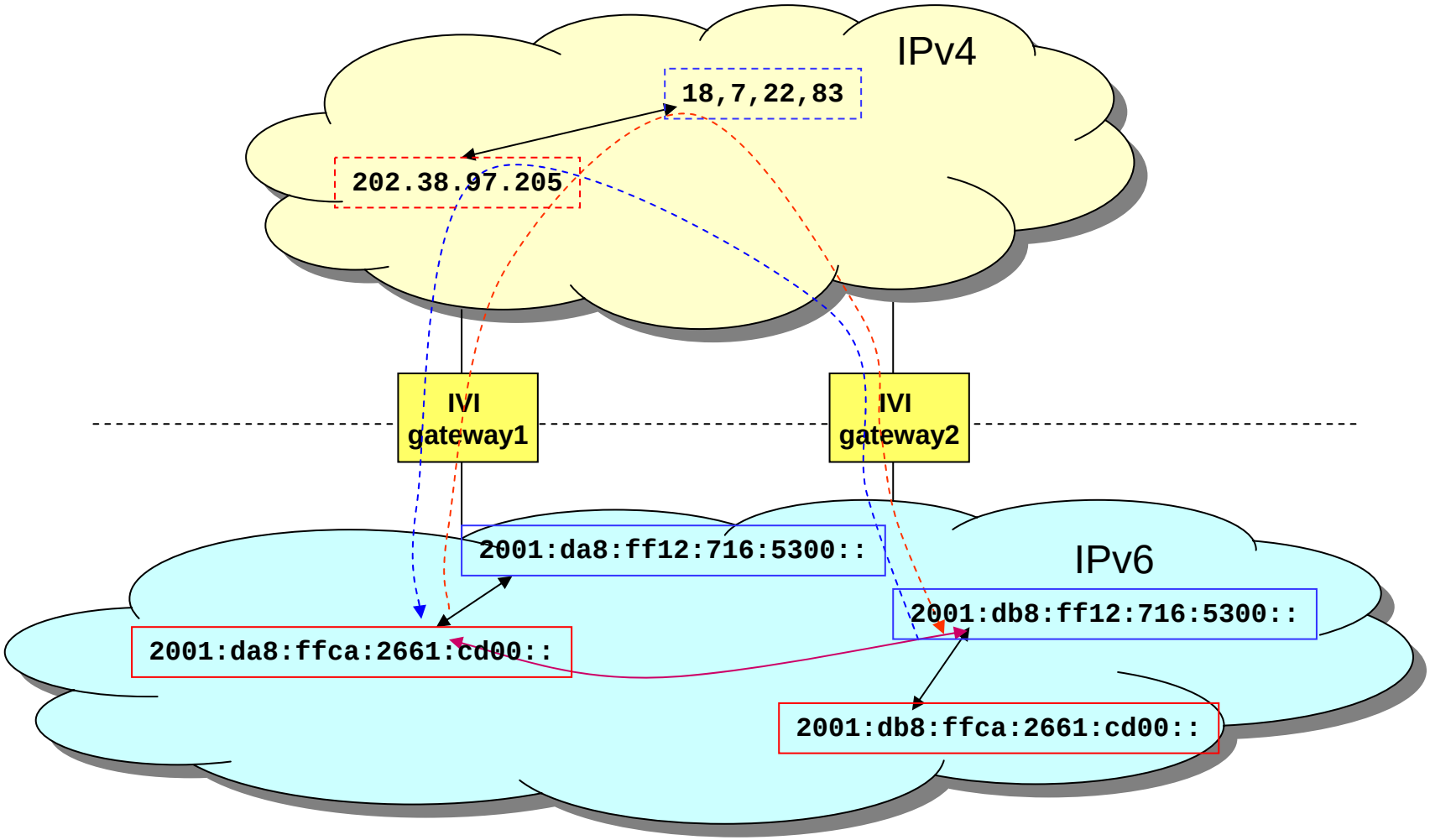


Longest prefix match

多个IVI网关


IVI prefix: 2001:da8:ff00::/40
S=202.38.97.205
D=18.7.22.83
IVI1
S=2001:da8:ffca:2661:cd00::
D=2001:da8:ff12:716:5300::
18.7.22.83
(2001:da8:ff12:716:5300::)
2001:da8:ffca:2661:cd00::
(202.38.97.205)
IPv4 host
IPv4
IVI IPv6
IPv6 host
S=18.7.22.83
D=202.38.97.205
IVI2
S=2001:da8:ff12:716:5300::
D=2001:da8:ffca:2661:cd00::
IVI prefix: 2001:da8:ff00::/40

多个IVI域


IPv4
18,7,22,83
202.38.97.205
IVI
gateway1
IVI
gateway2
2001:da8:ff12:716:5300::
IPv6
2001:db8:ff12:716:5300::
2001:da8:ffca:2661:cd00::
2001:db8:ffca:2661:cd00::

Reachability

	IPG4	IVI	IPG6
IPG4	OK	OK	NO
IVI	OK	OK	OK
IPG6	NO	OK	OK

IVI DNS service

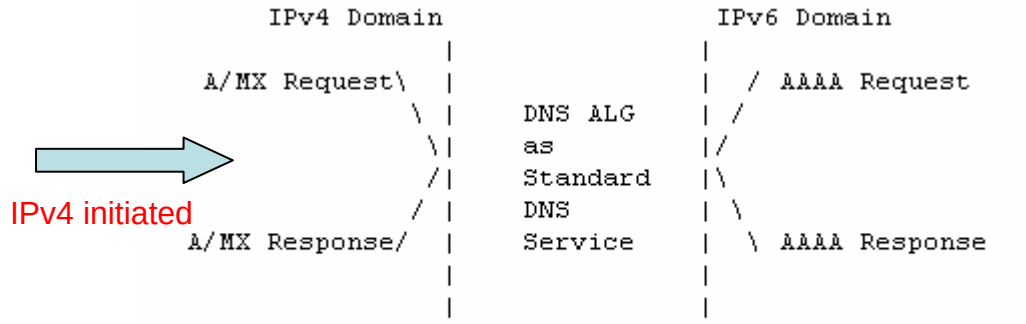


Figure 4: Normal DNS Service

- Normal DNS
  - Algorithm based

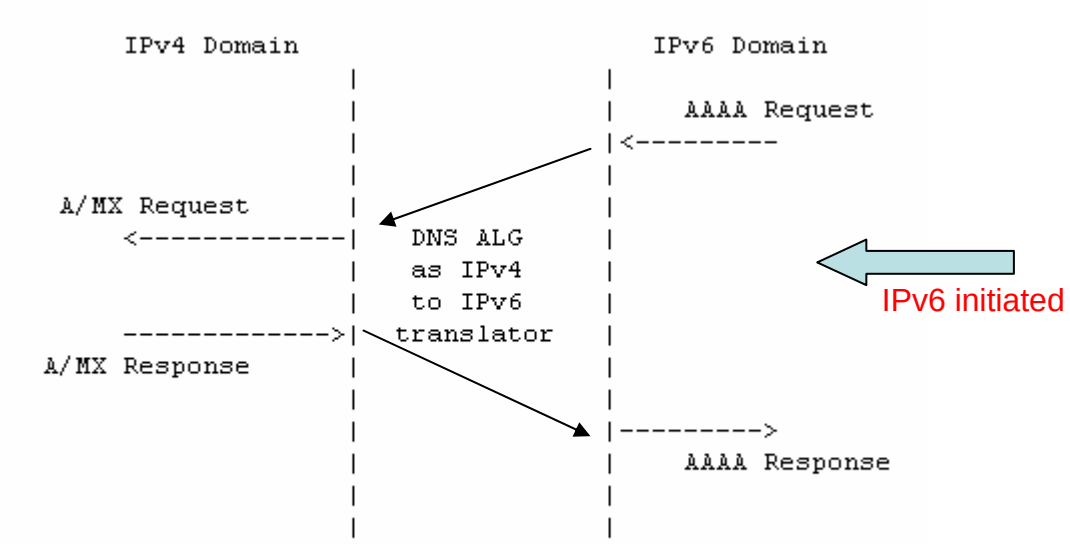


Figure 5: DNS Record Translation Service

- DNS translation
  - Algorithm based

IVI 运行模式

- 优选同一地址类的通信
  - 基于端对端的原则
    - IPv6 ←→ IPv6
    - IPv4 ←→ IPv4
- 优选无状态
  - 基于简单性原理
    - 由算法确定转换，不是由状态决定

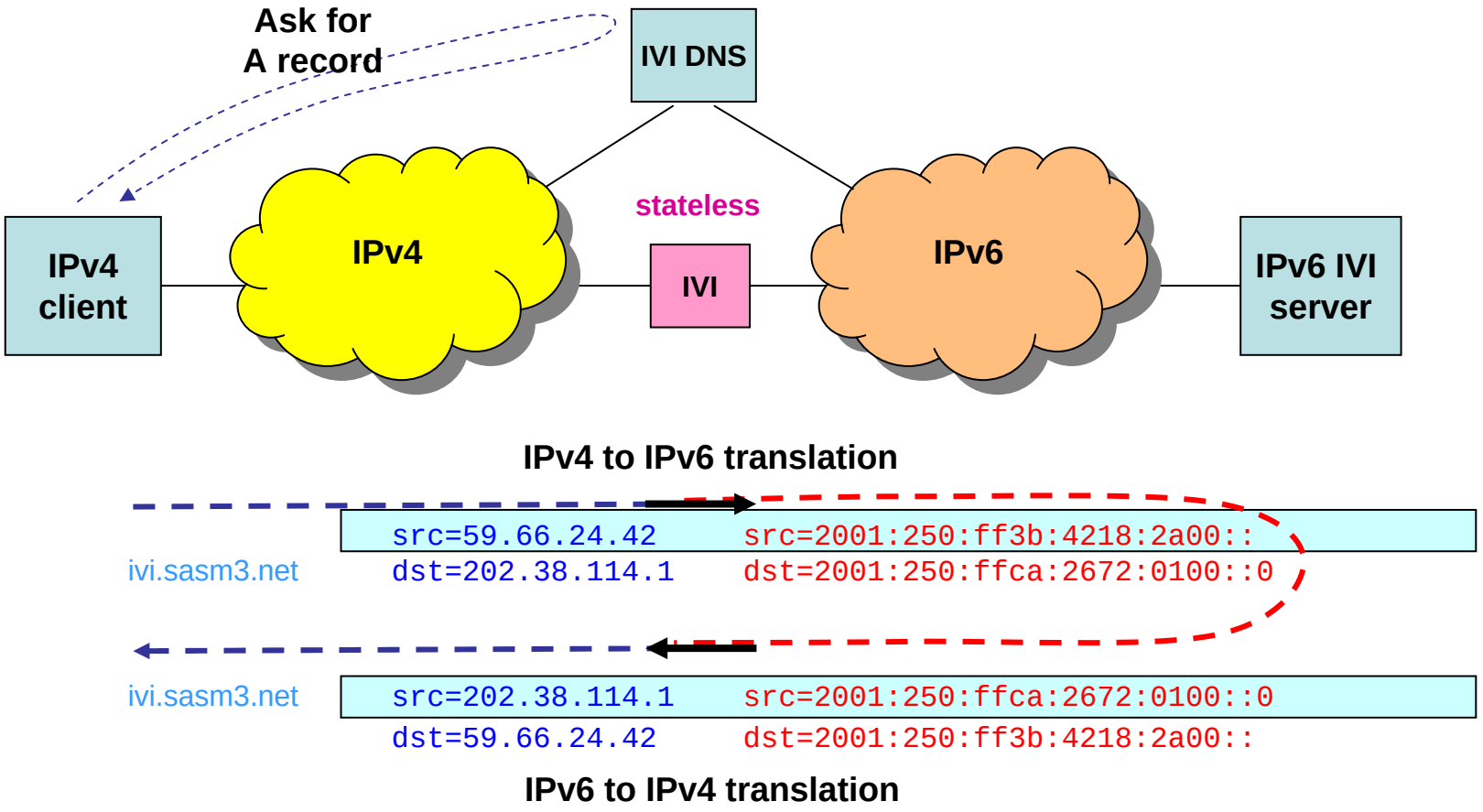
无状态 (1:1)运行模式

- 在无状态情况下，对IPv4和IVI IPv6 分组进行转换
  - SIIT扩展
    - 基于运营商的前缀
    - IPv4和 IVIIPv6地址之间的映射完全基于算法
    - 支持IPv6发起的通信，也支持IPv4发起的通信

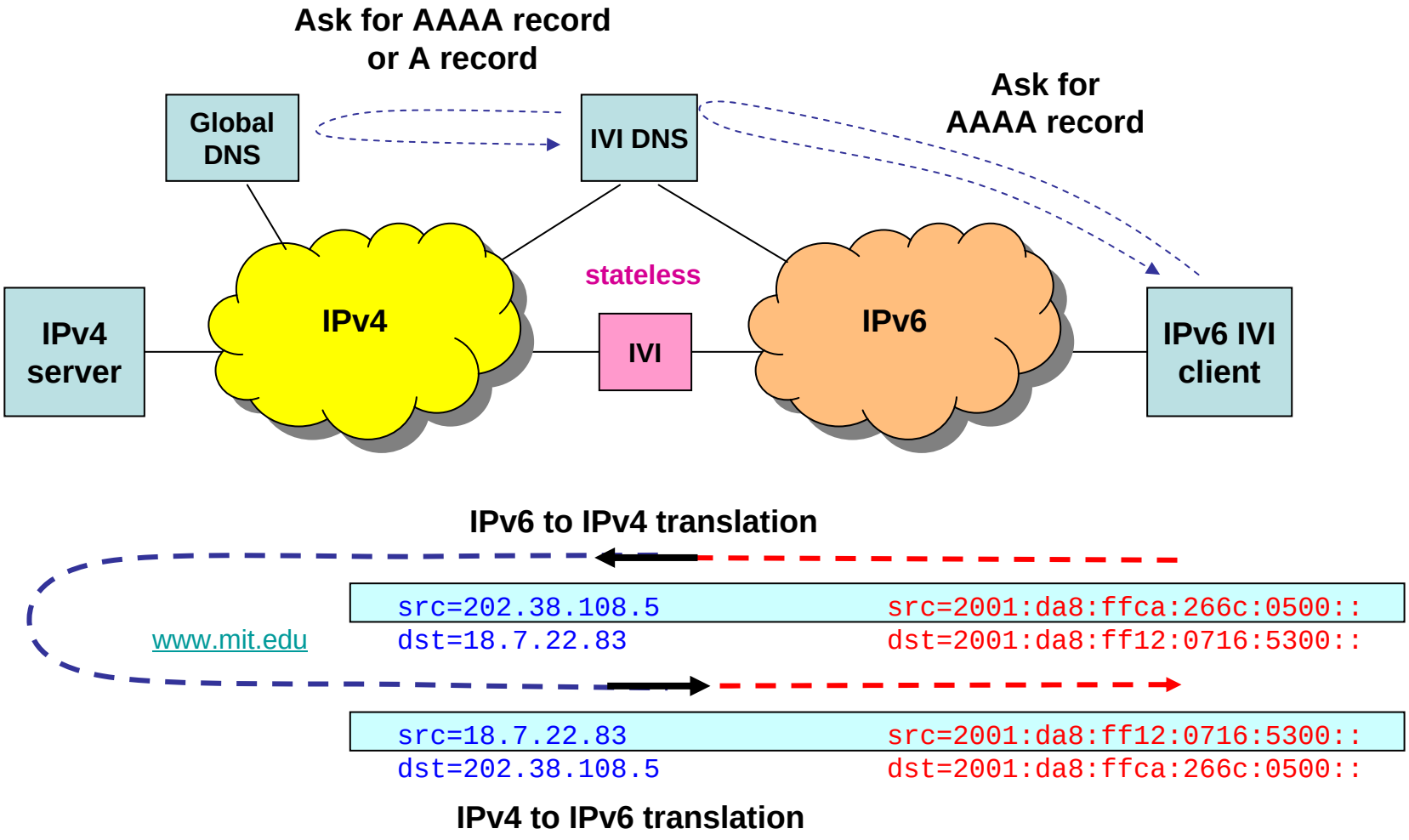
有状态 (1:N)运行模式

- 在有状态的情况下，对IPv4和通用IPv6分组进行转换
  - NAT-PT (NAPT-PT)的改进
    - IPv4地址和端口进行基于状态的复用
    - 基于运营商的前缀
    - 仅支持IPv6发起的通信

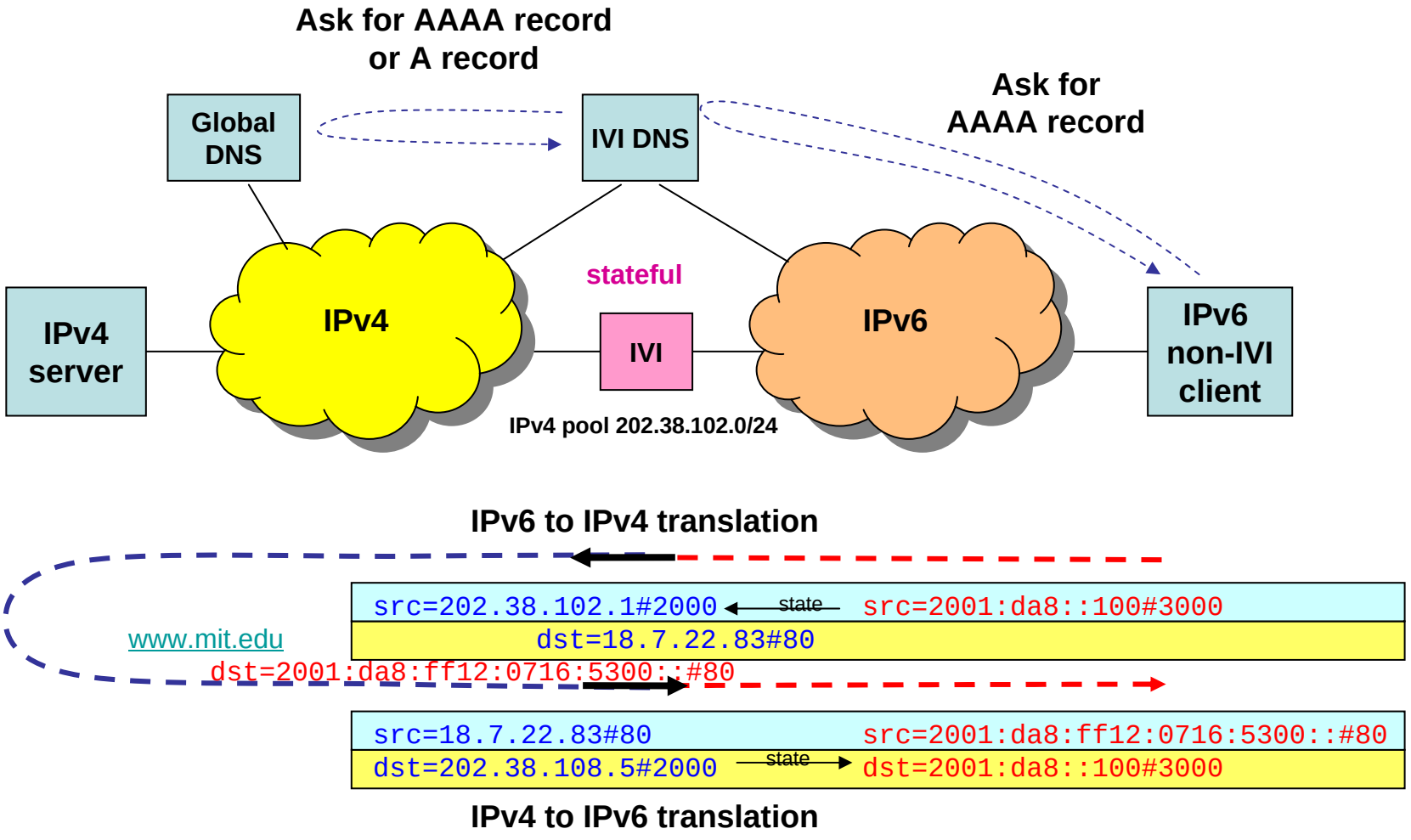
IPv4 initiated communication (1:1)



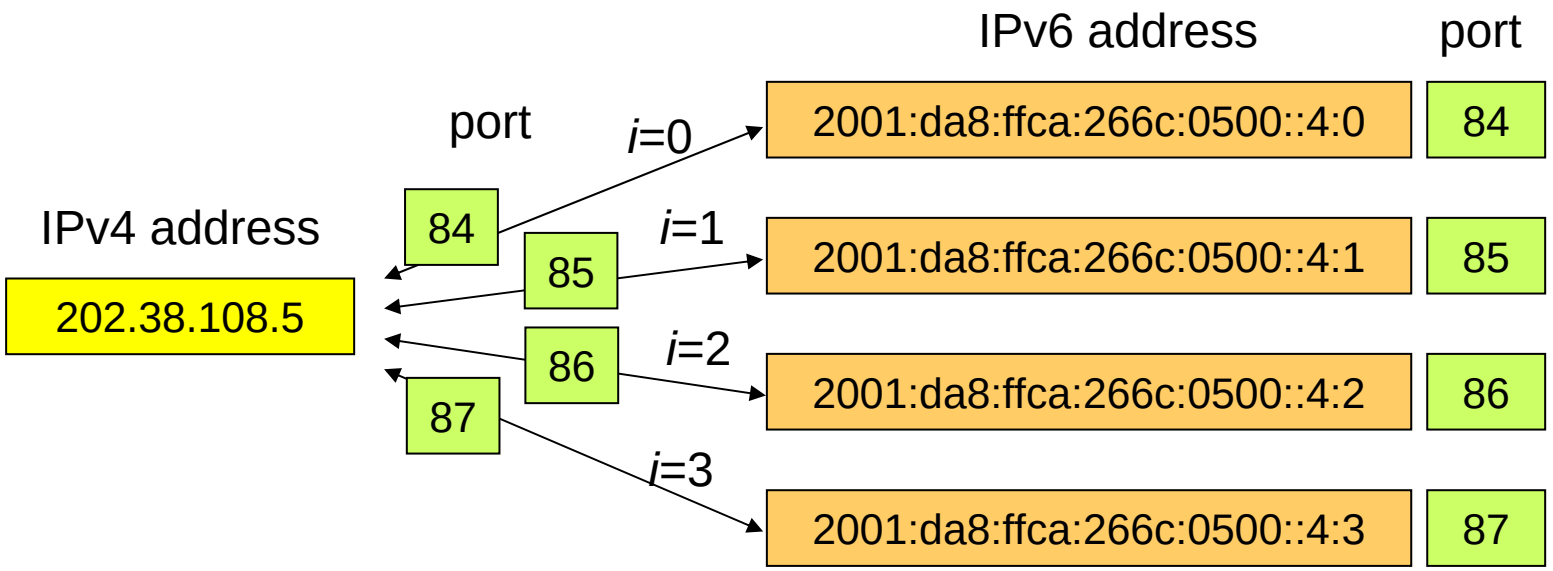
IPv6 initiated communication (1:1)



IPv6 initiated communication (1:N)



无状态1:N运行模式



Port number keep the same

在IVI中IPv4地址需求计算

- 根据测量一个客户计算机的并发通信进程的上限小于512，则复用率可选为128。
- 一个IPv4 /8（A类地址）的地址数为16,777,216，选用复用率128，则2个IPv4 /8可以支持43亿IPv6主机，数量与目前整个IPv4空间一样大。这些IPv6主机可以与所有的IPv4地址通信，也可以与所有的IPv6地址通信。

IVI multicast

- IVI supports PIM SSM
  - Group address mapping
  - RPF → mapped IPv6 address
  - PIM Spare-mode ALG

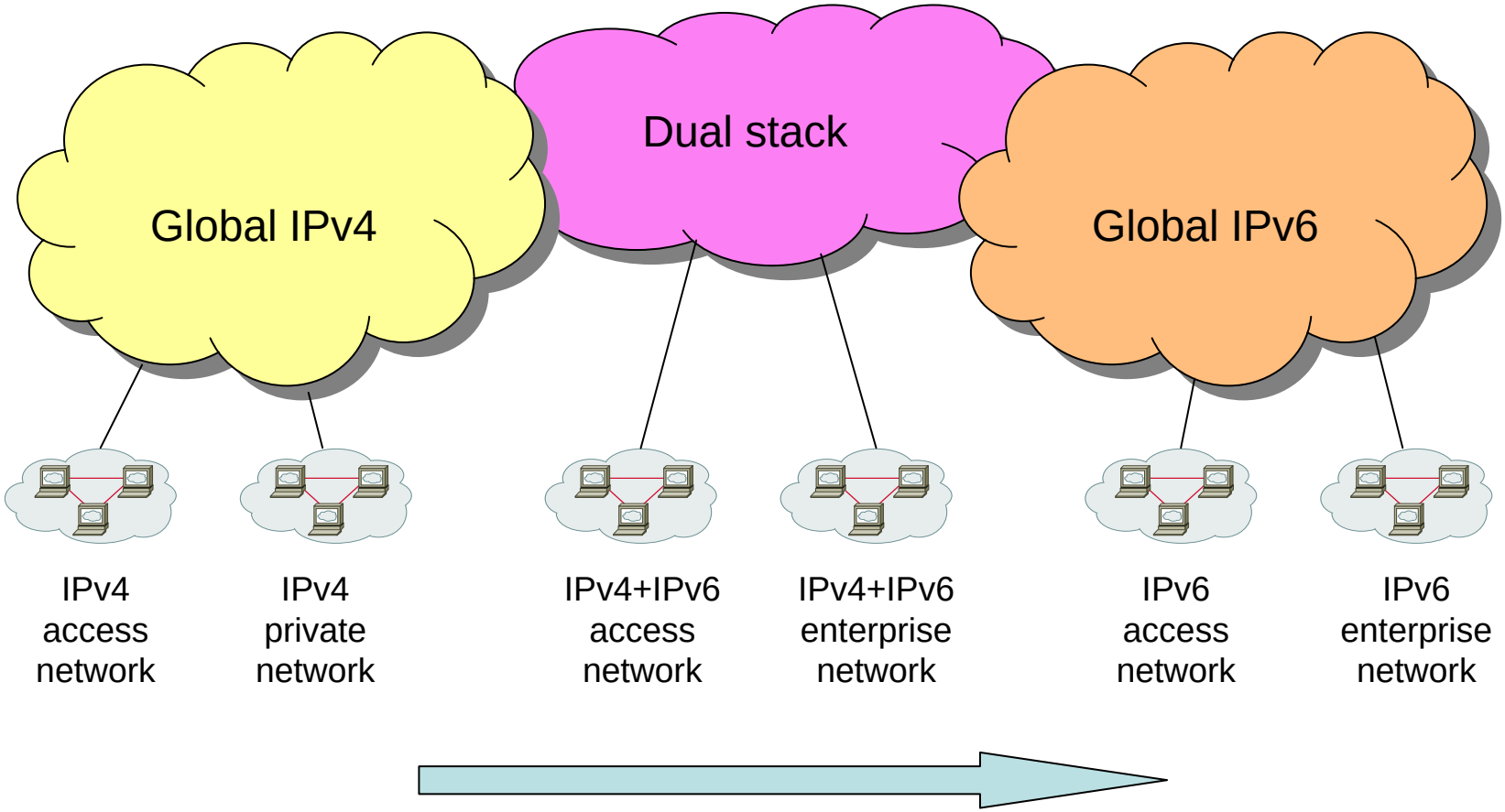
IVI Multicast Group Address Mapping

IPv4 Group Address	IPv6 Group Address
232.0.0.0/8	ff3e:0:0:0:0:0:f000:0000/96
232.255.255.255/8	ff3e:0:0:0:0:0:f0ff:ffff/96

过渡机制

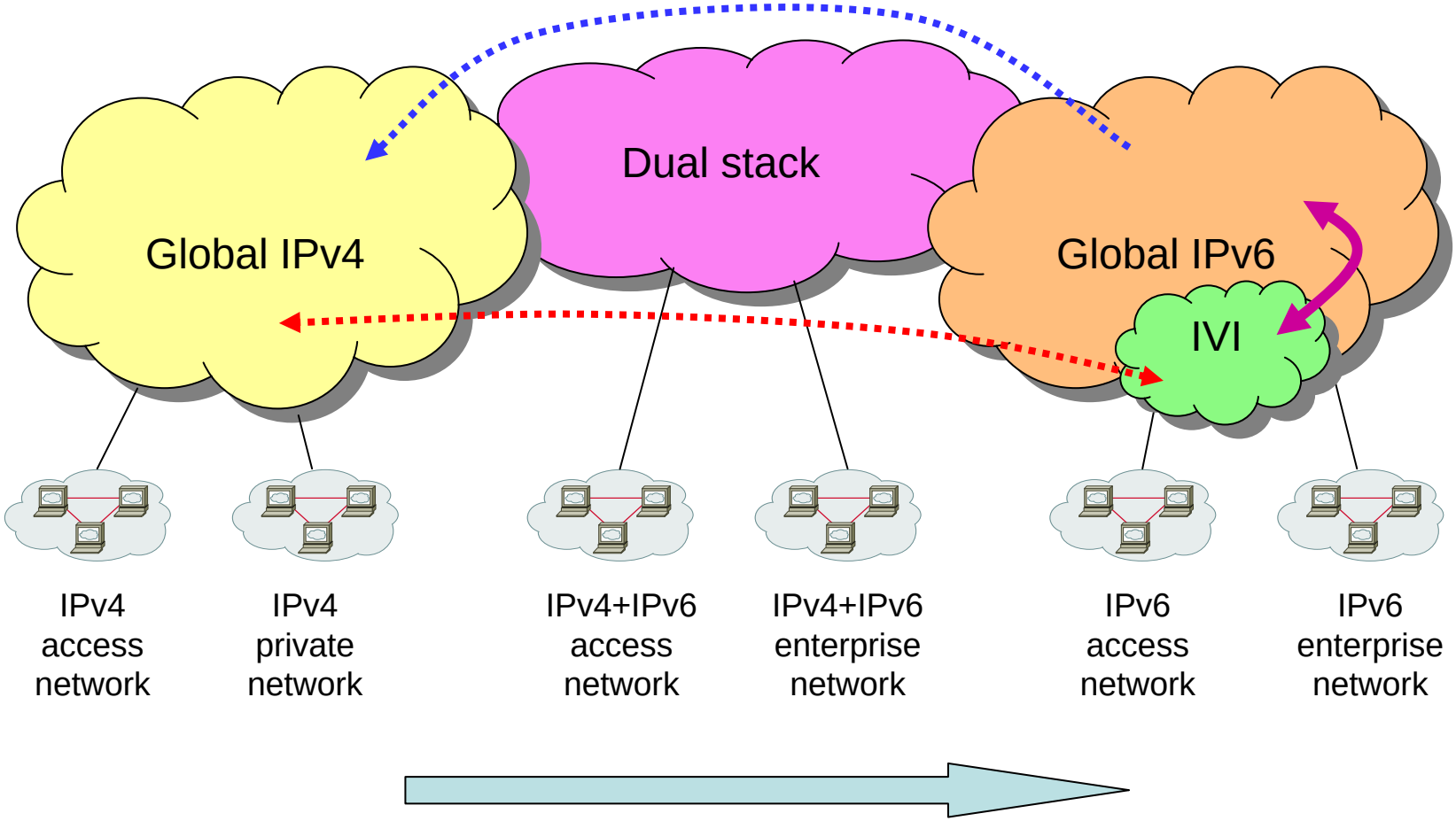
- 当IPv4地址变成稀缺资源时，有四种选择
  - 使用IPv4地址建NAT网络
    - 短期解决方案
  - 不用IPv4地址，建纯IPv6网络
    - 对于用户不可接受，因为互联网的全球可达性是基本需求
  - 使用IPv4地址建设双栈网络
    - 大多数运营商希望他人建双栈，自己搭便车
  - 使用IPv4地址建IVI IPv6单栈网络
    - 使用IVI IPv6的服务器可以被IPv6访问，也可以被IPv4访问（1:1）。
    - 使用IPv6的主机可以访问IPv4（1:n）。

原来的设计



- 大多数运营商希望他人建双栈，自己搭便车

IVI 模型

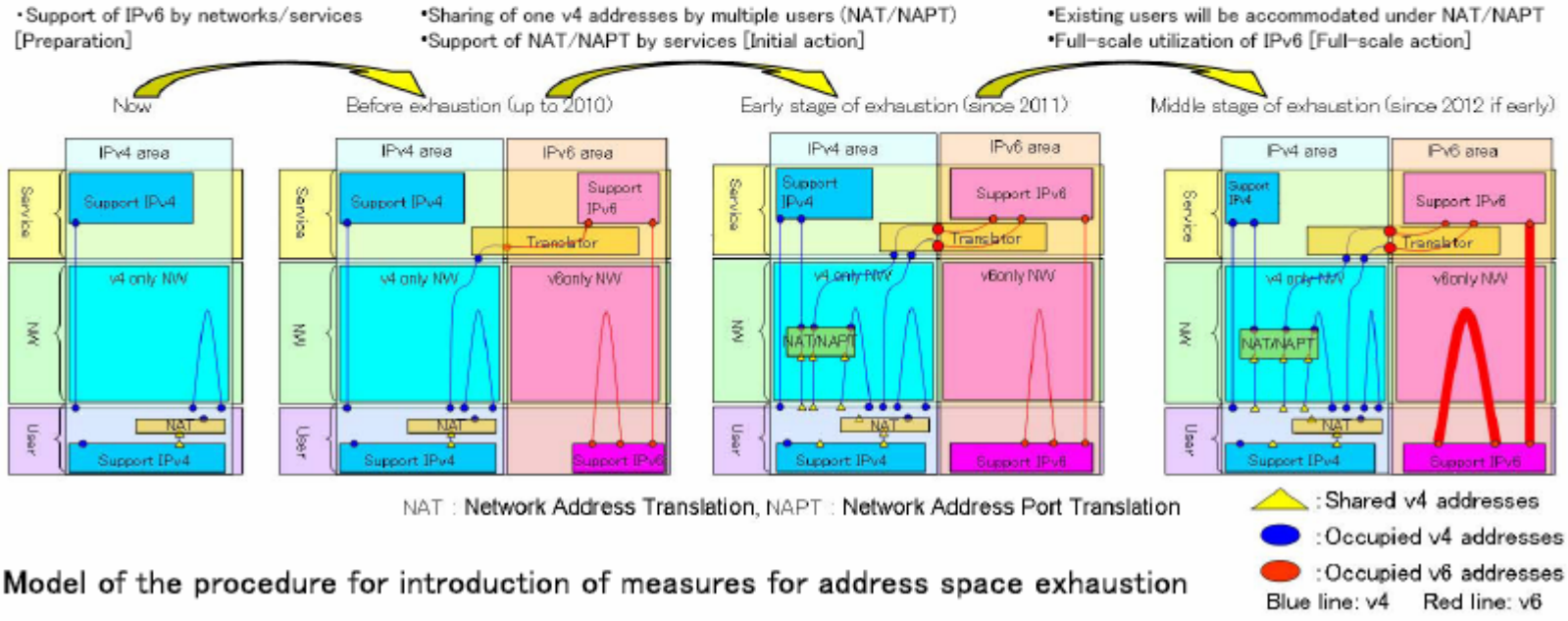


- 使用IVI IPv6的服务器可以被IPv6访问，也可以被IPv4访问（1:1），

使用IPv6的主机可以访问IPv4（1:n）

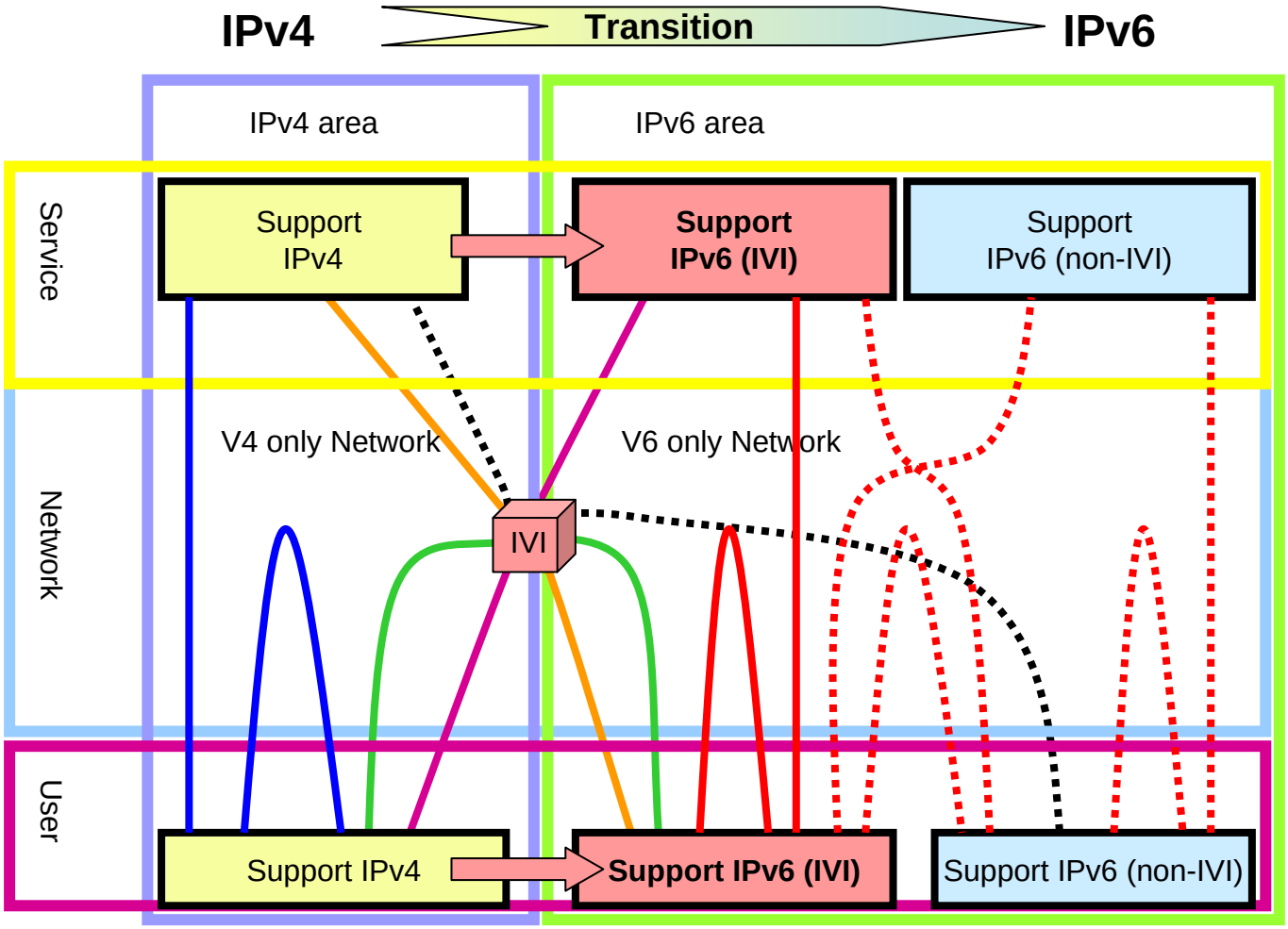
Measures for address space exhaustion

1. For continuous development of the Internet since 2011, the combination of **the transition to a new address system (IPv6)** and **sharing of one address by multiple users (using NAT/NAPT)** must be performed from three viewpoints of *feasibility within a time limit, continuity of service on the Internet, and continuance of effect*,
2. It is appropriate to **introduce the measures in three stages**: before exhaustion, early, and middle stages of exhaustion.



Model of the procedure for introduction of measures for address space exhaustion

服务器和用户的过渡


IPv4
Transition
IPv6
IPv4 area
IPv6 area
Service
Support IPv4
Support IPv6 (IVI)
Support IPv6 (non-IVI)
Network
V4 only Network
V6 only Network
IVI
User
Support IPv4
Support IPv6 (IVI)
Support IPv6 (non-IVI)

Actions for address space exhaustion

	Utilization of NAT/NAPT (Sharing of IPv4 addresses)	Reallocation of the assigned IPv4 addresses	Transition to IPv6	**IVI**
Feasibility within a time limit	✓	Doubtful	Extremely difficult	√
Service continuity	Limited	✓	✓	√
Permanent effect	Doubtful	NG	✓	√

Actually, IVI combines the methods of NAT/NAPT, reallocation of the assigned IPv4 addresses and transition to IPv6 in a natural and harmonious way.

http://www.ivi2.org/IVI/

IVI source code download

The IVI IPv4/IPv6 packet translation implementation as a Linux kernel patch is available below.

- IVI v0.5 kernel patch for Linux kernel 2.6.12
- IVI v0.5 kernel patch for Linux kernel 2.6.18

The IVI A/AAAA DNS proxy implementation is available below.

- IVIDNS v0.1 C code

For installing and configuration, please follow the instructions in the source code packages.

IVI test servers

- Access single-stack IPv6 server [2001:250:ffca:2672:100::] via IPv4
- Access single-stack IPv6 sever [2001:250:ffca:2672:100::] via IPv6
- Access IPv4 server (202.38.114.129) cross single-stack IPv6 network

---

IVI references

- IVI Update to SIIT and NAT-PT
- Prefix-specific and Stateless Address Mapping (IVI) for IPv4/IPv6 Coexistence and Transition

---

Additional notes

IVI ICMP扩展

```
ivitraceroute6 www.mit.edu

src_ivi4=202.38.97.205 src_ivi6=2001:da8:ffca:2661:cd00::
dst_host=www.mit.edu
dst_ip4=18.7.22.83 dst_ivig=2001:da8:ff12:716:5300::

traceroute to 2001:da8:ff12:716:5300:: (2001:da8:ff12:716:5300::)
30 hops max, 40 byte packets to not_ivi

1  2001:da8:ff0a:0:100::      0.304 ms 0.262 ms 0.190 ms
   10.0.0.1
2  2001:da8:ffca:7023:fe00::  0.589 ms * *
   202.112.35.254
3  2001:da8:ffca:7035:4900::  1.660 ms 1.538 ms 1.905 ms
   202.112.53.73
4  2001:da8:ffca:703d:9e00::  0.371 ms 0.530 ms 0.459 ms
   202.112.61.158
5  2001:da8:ffca:7035:1200::  0.776 ms 0.704 ms 0.690 ms
   202.112.53.18
6  2001:da8:ffcb:b5c2:7d00::  89.382 ms 89.076 ms 89.240 ms
   203.181.194.125
7  2001:da8:ffc0:cb74:9100::  204.623 ms 204.685 ms 204.494 ms
   192.203.116.145
8  2001:da8:ffcf:e7f0:8300::  249.842 ms 249.945 ms 250.329 ms
   207.231.240.131
9  2001:da8:ff40:391c:2d00::  249.891 ms 249.936 ms 250.090 ms
   64.57.28.45
10 2001:da8:ff40:391c:2a00:: 259.030 ms 259.110 ms 259.086 ms
   64.57.28.42
11 2001:da8:ff40:391c:700::  264.247 ms 264.399 ms 264.364 ms
   64.57.28.7
12 2001:da8:ff40:391c:a00::  271.014 ms 269.572 ms 269.692 ms
   64.57.28.10
13 2001:da8:ffc0:559:dd00::  274.300 ms 274.483 ms 274.316 ms
   192.5.89.221
14 2001:da8:ffc0:559:ed00::  274.534 ms 274.367 ms 274.517 ms
   192.5.89.237
15 * * *
16 2001:da8:ff12:a800:1900:: 276.032 ms 275.876 ms 276.090 ms
   18.168.0.25
17 2001:da8:ff12:716:5300::  276.285 ms 276.370 ms 276.214 ms
   18.7.22.83
```

- 运行和管理
  - IPv4 → IPv6
  - IPv6 → IPv4

```
ivitraceroute 202.38.108.2

1  202.112.0.65 6 ms 2 ms 1 ms
2  202.112.53.73 4 ms 6 ms 12 ms
3  202.112.53.178 1 ms 1 ms 1 ms
4  202.112.61.242 1 ms 1 ms 1 ms
5  202.38.17.186 1 ms 1 ms 1 ms
   202.38   AS4538
6  202.38.17.186 1 ms 1 ms 1 ms
   202.38   AS4538
7  202.38.17.186 2 ms 2 ms 2 ms
   202.38   AS4538
8  202.38.17.186 2 ms 2 ms 2 ms
   202.38   AS4538
9  202.38.17.186 4 ms 4 ms 3 ms
   202.38   AS4538
10 202.38.108.2 2 ms 3 ms 3 ms
```

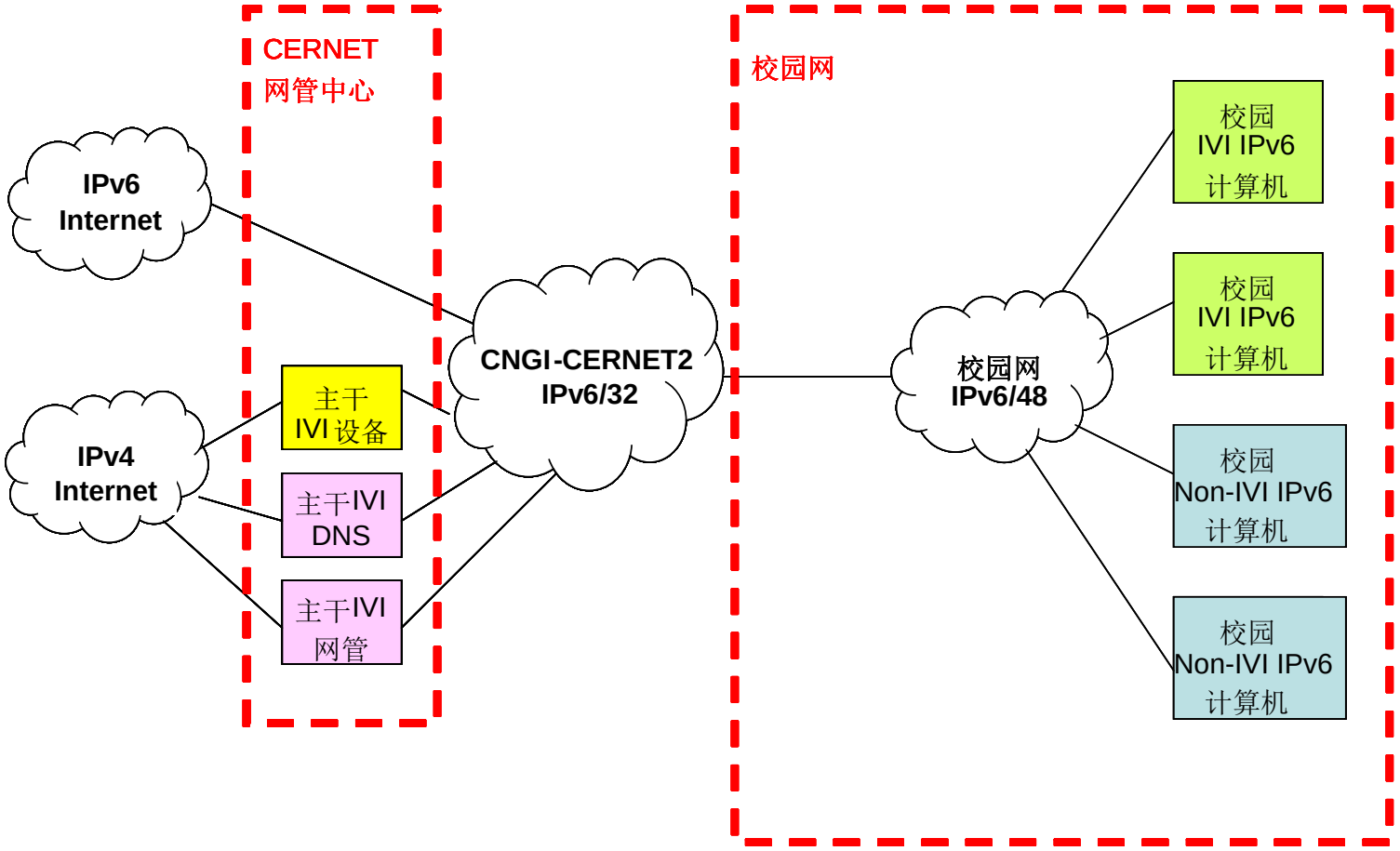
Dual-stack

- bc35207:xing % traceroute6 www.apnic.net
- traceroute to www.apnic.net (2001:dc0:2001:0:4608:20::), 30 hops max, 40 byte packets
- 1 cernet2.net (2001:da8:ffca:2661:cf00::) 0.515 ms * *
- 2 2001:250:1::1 (2001:250:1::1) 0.446 ms 0.506 ms 0.576 ms
- 3 2001:252:0:1::101 (2001:252:0:1::101) 8.314 ms 12.232 ms 16.767 ms
- 4 2001:252:0:101::2 (2001:252:0:101::2) 38.405 ms * *
- 5 hurricaneelectric-RGE.hkix.net (2001:7fa:0:1::ca28:a19e) 41.639 ms 41.647 ms 41.661 ms
- 6 v1026.core1.sjc1.he.net (2001:470:0:c3::1) 217.248 ms 212.821 ms 212.892 ms
- 7 10gigabitethernet2-3.core1.pao1.he.net (2001:470:0:54::2) 206.595 ms 206.598 ms 206.873 ms
- 8 2001:504:d::d (2001:504:d::d) 209.230 ms 208.085 ms 208.341 ms
- 9 2001:268:ff00::1 (2001:268:ff00::1) 216.445 ms 214.376 ms v6-oteCORE02.kddnet.ad.jp (2001:268:ff00::2) 314.355 ms
- 10 2001:200:0:fe00::12a9:0 (2001:200:0:fe00::12a9:0) 214.018 ms v6-oteCORE01.kddnet.ad.jp (2001:268:ff00:5::1) 206.059 ms 2001:200:0:fe00::12a9:0 (2001:200:0:fe00::12a9:0) 214.121 ms
- 11 2001:dc0:2001:255::1 (2001:dc0:2001:255::1) 345.619 ms 2001:200:0:fe00::12a9:0 (2001:200:0:fe00::12a9:0) 206.494 ms 2001:dc0:2001:255::1 (2001:dc0:2001:255::1) 346.695 ms
- 12 2001:dc0:2001:255::1 (2001:dc0:2001:255::1) 337.573 ms 336.493 ms www.apnic.net (2001:dc0:2001:0:4608:20::) 347.640 ms

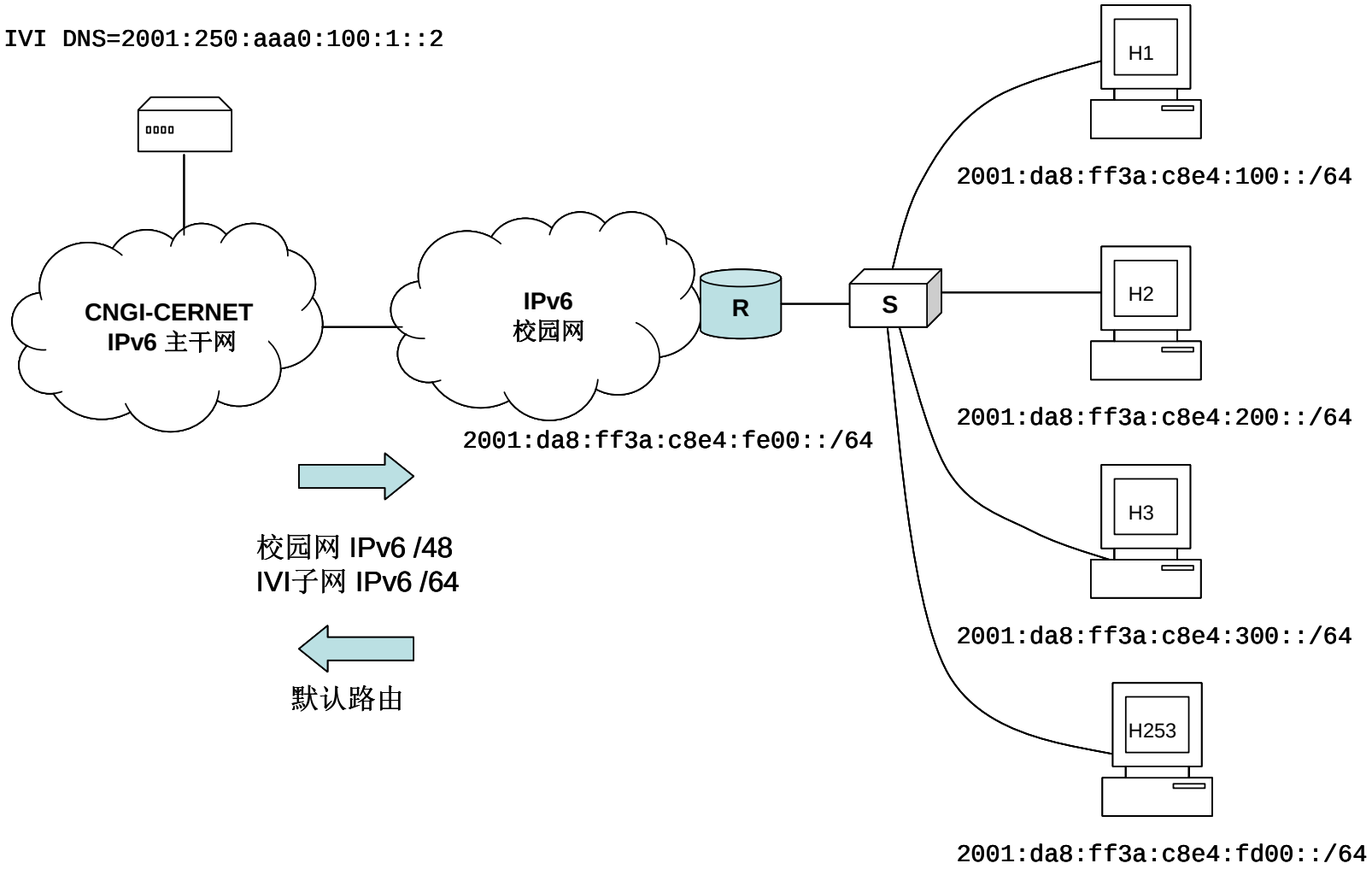
IVI

```
bc35207:xing % traceroute6 www.mit.edu
traceroute to www.mit.edu (2001:da8:ff12:716:5300::), 30 hops max, 40 byte packets
 1  cernet2.net (2001:da8:ff0a:0:100::)  0.510 ms  0.502 ms  0.495 ms
 2  * * *
 3  cernet2.net (2001:da8:ffca:7035:4900::)  7.664 ms  9.145 ms  10.590 ms
 4  cernet2.net (2001:da8:ffca:703d:9e00::)  1.166 ms  1.153 ms  1.171 ms
 5  cernet2.net (2001:da8:ffca:7035:1200::)  1.708 ms  1.691 ms  1.697 ms
 6  cernet2.net (2001:da8:ffcb:b5c2:7d00::)  90.171 ms  89.684 ms  89.669 ms
 7  cernet2.net (2001:da8:ffc0:cb74:9100::)  210.906 ms  210.985 ms  211.030 ms
 8  cernet2.net (2001:da8:ffcf:e7f0:8300::)  205.689 ms  205.765 ms  205.809 ms
 9  cernet2.net (2001:da8:ff40:391c:2d00::)  240.182 ms  237.486 ms  236.847 ms
10  cernet2.net (2001:da8:ff40:391c:2a00::)  261.744 ms  260.396 ms  260.171 ms
11  cernet2.net (2001:da8:ff40:391c:700::)  276.300 ms  273.889 ms  273.810 ms
12  cernet2.net (2001:da8:ff40:391c:1200::)  278.728 ms  278.903 ms  279.235 ms
13  cernet2.net (2001:da8:ffc0:559:dd00::)  294.181 ms  284.200 ms  284.170 ms
14  cernet2.net (2001:da8:ffc0:559:ed00::)  284.001 ms  283.921 ms  283.922 ms
15  cernet2.net (2001:da8:ffcf:d28f:6e00::)  283.961 ms  283.855 ms  283.869 ms
16  cernet2.net (2001:da8:ff12:a800:1900::)  284.264 ms  284.257 ms  284.247 ms
17  cernet2.net (2001:da8:ff12:716:5300::)  284.420 ms  284.254 ms  284.550 ms
```

CNGI-2期


CERNET
网管中心
IPv6
Internet
IPv4
Internet
主干
IVI 设备
主干IVI
DNS
主干IVI
网管
CNGI-CERNET2
IPv6/32
校园网
校园网
IPv6/48
校园
IVI IPv6
计算机
校园
IVI IPv6
计算机
校园
Non-IVI IPv6
计算机
校园
Non-IVI IPv6
计算机

Testbed


IVI DNS=2001:250:aaa0:100:1::2
CNGI-CERNET
IPv6 主干网
IPv6
校园网
R
S
H1
H2
H3
H253
2001:da8:ff3a:c8e4:100::/64
2001:da8:ff3a:c8e4:200::/64
2001:da8:ff3a:c8e4:300::/64
2001:da8:ff3a:c8e4:fd00::/64
2001:da8:ff3a:c8e4:fe00::/64
校园网 IPv6 /48
IVI子网 IPv6 /64
默认路由

End system configuration

如 H1 为 Linux 计算机，其配置举例为：

```
% cat /etc/rc.local
#!/bin/sh -e
/sbin/ip -6 addr  add 2001:da8:ff3a:c8e4:100::/64 dev eth0
/sbin/ip -6 route add 2000::/3 via add 2001:da8:ff3a:c8e4:fe00::
```

如 H2 为 Windows XP 计算机，其配置举例为：

1. Click **Start**, point to **Programs**, point **Accessories**, and click **Command Prompt**.
2. Type **ipv6 install**, and then type ENTER. After a few minutes you should get a `success' message.
3. Type **netsh**, and then press ENTER.
4. Type **interface ipv6**, and then press ENTER.
5. Type **add address "Wireless Network Connection" 2001:da8:ff3a:c8e4:200::**, and then press ENTER.
6. Type **add route ::/0 "Wireless Network Connection" 2001:da8:ff3a:c8e4:fe00:: persistant**, and then press ENTER.

注：在中文系统中，请键入对应的命令和相应的网络接口名称。

```
% cat /etc/resolv.conf
nameserver 2001:250:aaa0:100:1::2
```

IVI地址分配方案

IVI 地址分配方案

	单位名称	城市	IPv4 地址	IPv6 地址
1	北京大学	北京	58.200.128.0/24	2001:da8:ff3a:c880::/64
2	清华大学	北京	58.200.129.0/24	2001:da8:ff3a:c881::/64
3	北京航空航天大学	北京	58.200.130.0/24	2001:da8:ff3a:c882::/64
4	北京邮电大学	北京	58.200.131.0/24	2001:da8:ff3a:c883::/64
5	中国人民大学	北京	58.200.132.0/24	2001:da8:ff3a:c884::/64
6	北京理工大学	北京	58.200.133.0/24	2001:da8:ff3a:c885::/64
7	北京师范大学	北京	58.200.134.0/24	2001:da8:ff3a:c886::/64
8	中央民族大学	北京	58.200.135.0/24	2001:da8:ff3a:c887::/64
9	中国地质大学	北京	58.200.136.0/24	2001:da8:ff3a:c888::/64
10	北京交通大学	北京	58.200.137.0/24	2001:da8:ff3a:c889::/64
11	对外经济贸易大学	北京	58.200.138.0/24	2001:da8:ff3a:c88a::/64
12	北京科技大学	北京	58.200.139.0/24	2001:da8:ff3a:c88b::/64
13	北京工业大学	北京	58.200.140.0/24	2001:da8:ff3a:c88c::/64
14	中国农业大学	北京	58.200.141.0/24	2001:da8:ff3a:c88d::/64
15	北京化工大学	北京	58.200.142.0/24	2001:da8:ff3a:c88e::/64