

IPv6 Part 1: protocol and implementation status

Jun-ichiro Itojun Hagino, Ph.D
KAME project/ijlab
itojun@ijlab.net

Tutorial overview

- Part 1
 - IPv6 protocol itself (packet format, ...)
 - Impact/relationship to other internet protocols (such as TCP/HTTP)
 - Implementation status

- Part 2
 - Deployment status
 - Operational issues, clues, tips

Prerequisite/references

- Knowledge of
 - IPv4, TCP, UDP, routing in the internet
- SOI courses (<http://www.soi.wide.ad.jp/>)
 - IPv6 tutorial by kazu/itojun
- Books (not mandatory to read, but are informative)
 - Huitema, "IPv6: the new internet protocol"
 - Stevens, "TCP/IP illustrated vol.1"
 - Stevens, "UNIX network programming"
 - Hagino, "IPv6 network programming" (in Japanese)

Internet protocol version 6 (IPv6)

What is IPv6? (1)

- The answer to IPv4 address space problem
 - IP address changed from 32bit to 128bit
 - 32bit is smaller than worldwide human population
 - IPv4 address shortage is a bondage to new application areas
 - IPv4 addresses will run out around 2010?
 - 4×10^9 -> 3.4×10^{38}
 - 4 billion -> 340 undecillion
- A new starting point
 - IPv4: used for 20 years, minimal implementation lacks almost everything
 - IPv6: new standard features like multicast, PMTUD, IPsec, autoconfig
 - Leverage the availability of advanced technologies
- An answer to routing table size issue
 - IPv4: portable address (->CIDR), result in 50K entries
 - IPv6: full CIDR + well-structured address assignment
 - restricts DFZ routes to 8K

What is IPv6? (2)

- Say a long good-bye to NAT, say hi again to end-to-end model
 - Has been used as temporary cure for address space issue
 - NAT breaks bidirectional communication
 - NAT is a bondage to application protocol designers
 - NAT does not let multicast/IPsec through
 - NAT does not work with proprietary protocols
 - NAT does not improve security
 - NAT box is a single point of failure
 - If you reboot NAT box, all connection will be gone
 - NAT is a barrier to scalability of the Internet as a whole
 - prevents deployment of peer-to-peer apps

IPv6 features

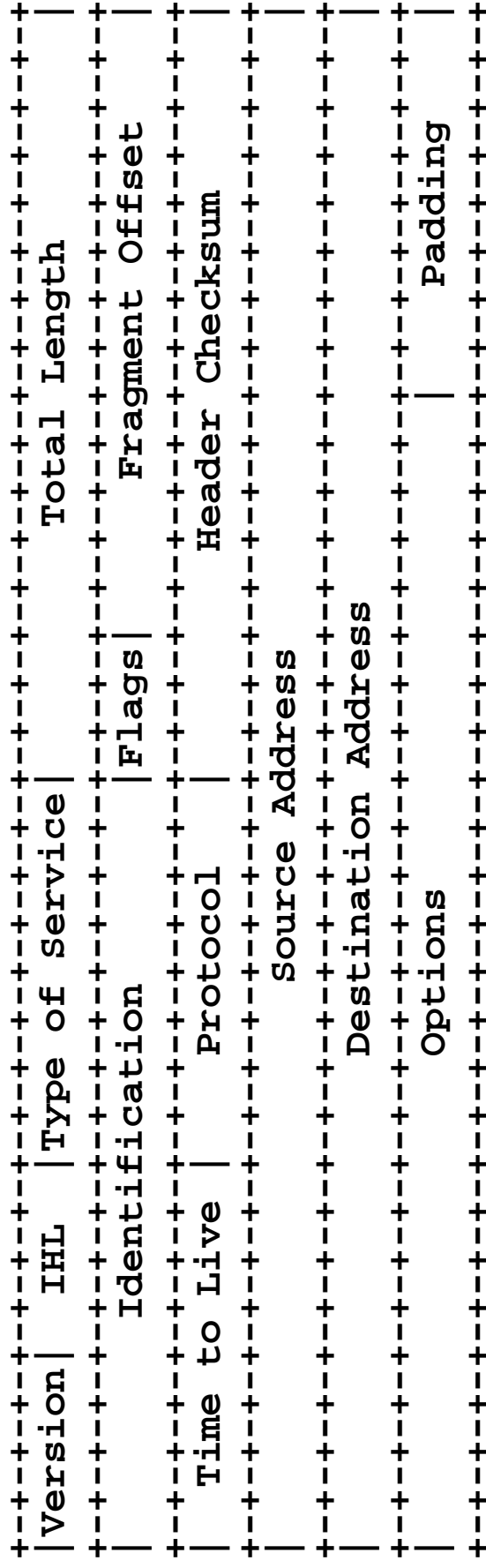
- Can coexist with IPv4 network
 - There's no flag day, you can use existing devices
 - "Dual stack" - devices that can use IPv4/IPv6 at the same time
- Only needs software update in most cases
 - You do not need to buy new routers
- Vast amount of address space
 - 2^{32} or 2^{13} route entries in ISP
 - 2^{64} nodes on a subnet
 - 2^{16} subnets to a site
- Aggregation friendly
 - Core routes are limited to 8192 (2^{13}) by addressing architecture
 - No portable address available
- Designed for less operation cost in routers
 - Mandates PMTUD, no fragmentation in intermediate routers
- Autoconfiguration works beautiful
 - No server necessary, no wacky state management
 - Autoconfig available everywhere: IPv4 autoconfig required DHCP server

IPv6 INCORRECT rumors

- Fixes all Internet problems
 - How? :-P
- Noone uses it
 - There are so many users
 - (in Japan commercial IPv6 ISP services are available too)
 - Asia and Europe are more active than the US
 - US DoD/DoC will require IPv6 on their procurement
- QoS problem is solved
 - Fields are reserved for QoS
 - QoS technology is still in its infancy, and needs more time to be stabilized
- Security problem is solved
 - IPsec is mandatory
 - NAT will be gone, so IPsec can be used everywhere
 - However, IPsec itself has very difficult issue
- IPv6 is not necessary since we have NAT
 - BIG NO!
- Need hardware upgrade
 - In most cases, no. Software upgrade is enough

IPv4 header

- Many fields
 - TOS: left unused in most cases
 - IP header checksum: high cost to maintain, cost/benefit?
 - IPv4 options: not really used (most of routers do not permit options)
- Usually 20 bytes (8 bytes are for src/dst)



How to write IPv6 addresses

- 16 groups of 4-digit hexadecimal
 - 3ffe:0507:0000:0001:0200:86ff:fe05:80fa
 - 0000:0000:0000:0000:0000:0000:0000:0001
- You can omit starting "0" in each of the groups
 - 3ffe:507:0:1:200:86ff:fe05:80fa
 - 0:0:0:0:0:0:1
- In only one place, you can use "::" to denote continuous "0"
 - 3ffe:507::1:200:86ff:fe05:80fa
 - ::1
- Usually you do not write numeric IPv6 addresses, use DNS names

Upper layer protocols

- Upper layer
 - ICMPv6: ICMP for IPv6
 - Very similar
 - TCP, UDP: same as IPv4
 - Only pseudo-header checksum is different
 - Protocols on TCP/UDP: same
 - HTTP, FTP will work just like it was on IPv4
 - If we embed IP address into upper-layer protocol header, we need to modify it (FTP)

Modifications to upper layer protocols

- FTP
 - PORT, PASV passes IPv4 address, and does not support other types
 - EPSV, EPRT command - RFC2428
- HTTP
 - host:port form is ambiguous if host is an IPv6 address
 - [host]:port notation - RFC2732

IPv4 packet example

□ TCP packet toward port 22, in IPv4

```
16:09:25.113204 202.232.15.102.63472 > 202.232.15.98.22: P 20:40(20) ack 21
win 17520 <nop,nop,timestamp 245605 38398> (ttl 64, id 13795)
    4500 0048 35e3 0000 4006 9034 cae8 0f66
    cae8 0f62 f7f0 0016 8520 7612 1672 89e4
    8018 4470 4060 0000 0101 080a 0003 bf65
    0000 95fe 0000 000a 0eef be30 a72a 41f1
    7be1 2f9b 3ccd b5b0
16:09:25.113364 202.232.15.102.63472 > 202.232.15.98.22: P 20:40(20) ack 21
win 17520 <nop,nop,timestamp 245605 38398> (ttl 64, id 13795)
    4500 0048 35e3 0000 4006 9034 cae8 0f66
    cae8 0f62 f7f0 0016 8520 7612 1672 89e4
    8018 4470 4060 0000 0101 080a 0003 bf65
    0000 95fe 0000 000a 0eef be30 a72a 41f1
    7be1 2f9b 3ccd b5b0
```

IPv6 packet example

□ TCP packet toward port 22, in IPv6

```
15:43:17.093542 3ffe:507:0:1:200:86ff:fe05:80fa.49226 >  
3ffe:507:0:1:260:97ff:fe07:69ea.22: . ack 1093 win 17519 <nop,nop,timestamp  
242469 35276> [flowlabel 0x62073] (len 32, hlim 64)
```

```
6006 2073 0020 0640 3ffe 0507 0000 0001  
0200 86ff fe05 80fa 3ffe 0507 0000 0001  
0260 97ff fe07 69ea c04a 0016 3ee4 92b2  
29d4 2455 8010 446f 80da 0000 0101 080a  
0003 b325 0000 89cc
```

```
15:43:17.093648 3ffe:507:0:1:200:86ff:fe05:80fa.49226 >  
3ffe:507:0:1:260:97ff:fe07:69ea.22: . ack 1093 win 17519 <nop,nop,timestamp  
242469 35276> [flowlabel 0x62073] (len 32, hlim 64)
```

```
6006 2073 0020 0640 3ffe 0507 0000 0001  
0200 86ff fe05 80fa 3ffe 0507 0000 0001  
0260 97ff fe07 69ea c04a 0016 3ee4 92b2  
29d4 2455 8010 446f 80da 0000 0101 080a  
0003 b325 0000 89cc
```

IPv6 extension headers

- Hop-by-Hop Options
 - Routers will look at it
- Routing
 - Source routes
- Fragment
 - Fragmentation, ONLY at originating node
- Destination Options
 - Pass info to final destination
- IPsec (AH, ESP)
- Header chain
 - Header parsing code becomes somewhat different from IPv4

IPv6(nxt=TCP) TCP payload

IPv6(nxt=routing) Routing(nxt=TCP) TCP payload

IPv6(nxt=fragment) Fragment(nxt=TCP) TCP payload

IPv6 extension headers example (1)

- IPv6 echo request, without AH and with AH

```
16:14:31.750102 ::1 > ::1: icmp6: echo request (len 16, hlim 64)
6000 0000 0010 3a40 0000 0000 0000 0000
0000 0000 0000 0001 0000 0000 0000 0000
0000 0000 0000 0001 8000 7f1c d802 1000
57ea e137 d471 0b00
```

```
16:14:57.060181 ::1 > ::1: AH(spi=9999, sumlen=16, seq=0x7): icmp6: echo
request (len 40, hlim 64)
```

```
6000 0000 0028 3340 0000 0000 0000 0000
0000 0000 0000 0001 0000 0000 0000 0000
0000 0000 0000 0001 3a04 0000 0000 270f
0000 0007 5a3c 2d0d efbf 3370 d029 1ed5
8000 c3a3 dd02 0300 71ea e137 88ea 0000
```

Lower layer

- Ethernet protocol type: different from IPv4
 - IPv4: 0x0800
 - IPv6: 0x86dd
- Hardware address resolution
 - IPv4: medium dependent, ARP for ethernet
 - Simple protocol
 - IPv6: medium independent, NDP on ICMPv6
 - Manages more state than ARP, more complex (but efficient)
 - Implements DAD, Duplicated Address Detection
- Uses link-layer multicast heavily
 - No use of broadcast
 - Link-layer broadcast will wake everyone up on the medium

Ethernet encapsulation example (1)

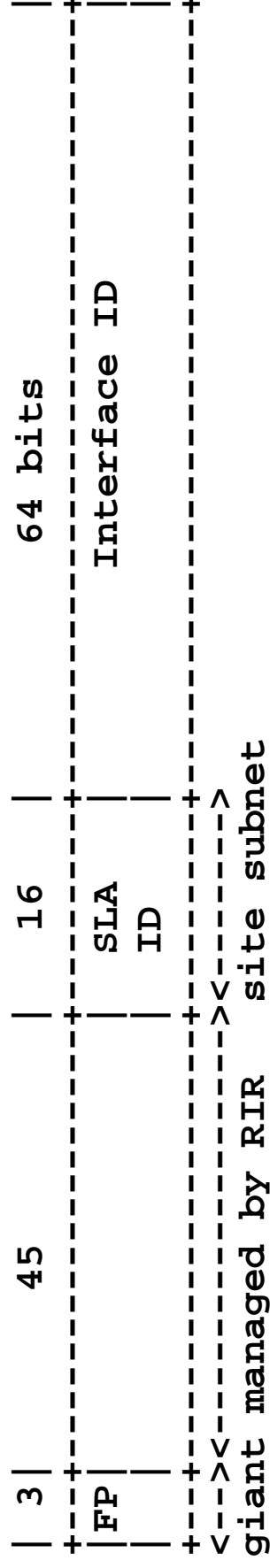
□ all-nodes ICMPv4/v6 echo request (IPv4: broadcast, IPv6: multicast)

```
15:46:30.651062 0:0:86:5:80:fa ff:ff:ff:ff:ff:ff 0800 98: 202.232.15.102 >
202.232.15.103: icmp: echo request (ttl 255, id 9961)
    4500 0054 26e9 0000 ff01 e021 cae8 0f66
    cae8 0f67 0800 b14d 02a5 0000 c6e3 e137
    a7ee 0900 0809 0a0b 0c0d 0e0f 1011 1213
    1415 1617 1819 1a1b 1c1d 1e1f 2021 2223
    2425 2627 2829 2a2b 2c2d 2e2f 3031 3233
    3435 3637
```

```
15:44:52.080190 0:0:86:5:80:fa 33:33:0:0:0:1 86dd 70:
fe80::200:86ff:fe05:80fa > ff02::1: icmp6: echo request (len 16, hlim 64)
    6000 0000 0010 3a40 fe80 0000 0000 0000
    0200 86ff fe05 80fa ff02 0000 0000 0000
    0000 0000 0000 0001 8000 ebd9 9c02 0100
    64e3 e137 aa38 0100
```

IPv6 addressing architecture (1)

- Global address: 2000:: - Hierarchy design to encourage route aggregation
 - Interface ID to help autoconfiguration (later)
 - Site boundary is fixed to /48 to encourage ISP change and site-local routing
 - A site can always have 2¹⁶ subnets
- Current RIR/ISP practice
 - Large ISP gets /32 from RIR (like 2001:200:: - Smaller ISP gets /40 suballocation from upstream ISP
 - End site (universities, companies) gets /48



IPv6 addressing architecture (2)

- **Link-local: fe80::/10**
 - Addresses used for address resolution and bootstrap
 - Unique only on single link

- **Multicast: ff00::/8**
 - Multicast addresses are scope as well
 - IPv4 - controlled by TTL, a whole mess
 - ff02::/16 - link, ff05:: - site

- **Global: the rest of the address**
 - Similar to IPv4 global address, worldwide uniqueness

IPv6 autoconfiguration

- Stateless address autoconfiguration
 - No resource management thanks to address architecture
 - Routers advertise info about the subnet
 - Hosts receive the information and configures itself
- DHCPv6
 - Stateful, needs server
- Stateless address autoconfiguration is much easier, and will be available everywhere

IPv6 autoconfig details

- End host transmits router solicitation
 - "I want to know the configuration for the subnet"
- Router transmits router advertisement
 - "Okay, here's information on subnet"
- End host configures itself

```
15:40:16.590444 fe80::200:86ff:fe05:80fa > ff02::2: icmp6: router
solicitation (src lladdr: 0:0:86:5:80:fa) (len 16, hlim 255)
    6000 0000 0010 3aff fe80 0000 0000 0000
    0200 86ff fe05 80fa ff02 0000 0000 0000
    0000 0000 0000 0002 8500 6d2e 0000 0000
    0101 0000 8605 80fa

15:40:16.819647 fe80::260:97ff:fe07:69ea > ff02::1: icmp6: router
advertisement (chlim=64, router_ltime=1800, reachable_time=30000,
retrans_time=1000) (src lladdr: 0:60:97:7:69:ea) (mtu: mtu=1500) (prefix
info: LA valid_ltime=3600000, preferred_ltime=3600000,
prefix=3ffe:507:0:1::/64) (len 64, hlim 255)
    6000 0000 0040 3aff fe80 0000 0000 0000
    0260 97ff fe07 69ea ff02 0000 0000 0000
    0000 0000 0000 0001 8600 4625 4000 0708
    0000 7530 0000 03e8 0101 0060 9707 69ea
    0501 0000 0000 05dc 0304 40c0 0036 ee80
    0036 ee80 0000 0000 3ffe 0507 0000 0001
    0000 0000 0000 0000
```

IPv6 and DNS

□ DNS IPv6 support: payload issue and transport issue

□ IPv6 in DNS payload

- AAAA record: similar to A record
- PTR record: use "ip6.arpa." tree

- "ip6.int" tree has been used, transition ongoing

`turneric.itojun.org. IN AAAA 3ffe:507:0:1:200:86ff:fe05:80fa`

`a.f.0.8.5.0.e.f.f.6.8.0.0.2.0.1.0.0.0.0.0.0.0.7.0.5.0.e.f.f.3.ip6.arpa.`

`IN PTR turneric.itojun.org.`

□ IPv6 in DNS transport

- BIND8 supports queries via IPv4 TCP/UDP only
- BIND9 supports queries over IPv6 TCP/UDP as well

IPv6 transition strategy

- There's no flag day, we will transition gradually
- IPv4/v6 dual stack
 - Ethernet protocol type is different - they can coexist
 - IPv4/v6 dual stack node can speak with IPv4-only node
 - You can even use NAT for IPv4 communication
 - NAT network for IPv4, global network access for IPv6
 - Gradually IPv6 will become dominating
- Tunnelling
 - At this moment, internet infrastructure uses IPv4
 - Tunnel IPv6 packet into IPv4
 - RFC2893 defines IPv6-over-IPv4 tunnelling
 - Can be used just like IPv6 p2p link
 - We run IPv6 routing protocol on top of it

Deploying IPv6-only network

- Happens only in far future, or when admins got tired of maintaining IPv4
- IPv6-only node needs to be able to talk with IPv4-only node
 - Need to click www.yahoo.com...
- Translator
 - Application layer: web proxy, fwtk, sendmail
 - TCP layer: socks, fwtk
 - IP layer: header translation
- Has the same problems with NAT (not end-to-end, single point of failure)

Tunnelling example

□ RFC2893 tunnelling

```
16:20:56.718074 210.163.17.131 > 202.232.15.98:
3ffe:507:102:0:260:8cff:fec8:b43b.22 >
3ffe:50a:ffff:100:210:4bff:fe0a:b89.1021: P 249:357(108) ack 380 win 17080
[flowlabel 0x2c8df] (len 128, hlim 64) (ttl 20, id 8872)
4500 00bc 22a8 0000 1429 c500 d2a3 1183
cae8 0f62 6002 c8df 0080 0640 3ffe 0507
0102 0000 0260 8cff fec8 b43b 3ffe 050a
ffff 0100 0210 4bff fe0a 0b89 0016 03fd
e7a9 375e 1fee a61c 5018 42b8 6270 0000
0000 0062 8d0a 2d88 6d8b 9ee9 4fdd a123
36a2 5b2d fef2 a3ec 882e 543b 4cbb 5fe8
8d31 8257 a02d 345b cbf6 491f 476e df02
1526 91fa 5e59 52fb f970 209a 24cc 4a45
f4d6 5e9e 81df 0c38 9514 27a5 2de5 b55c
2d41 c98f 06d2 2e06 4830 6990 dff2 d99b
3630 9a69 e0c7 476a 67be 0a9e
```

Tiny demo

```
$ ifconfig sm1
sm1: flags=8863<UP,BROADCAST,NOTRAILLERS,RUNNING,SIMPLEX,MULTICAST> mtu 1500
    media: Ethernet 10baseT
    inet 0.0.0.0 netmask 0xff000000 broadcast 255.255.255.255
    inet6 fe80::200:86ff:fe05:80fa prefixlen 64 scopeid 0x01
    inet6 3ffe:501:4819:2000:200:86ff:fe05:80fa prefixlen 64

$ ifconfig lo0
lo0: flags=8009<UP,LOOPBACK,MULTICAST> mtu 32976
    inet 127.0.0.1 netmask 0xff000000
    inet6 fe80::1 prefixlen 64 scopeid 0x02
    inet6 ::1 prefixlen 128

$ ping6 ::1
PING6(56=40+8+8 bytes) ::1 --> ::1
16 bytes from ::1, icmp_seq=0 hlim=64 time=0.137 ms
16 bytes from ::1, icmp_seq=1 hlim=64 time=0.112 ms
$ telnet localhost
Trying ::1...
Connected to localhost.
Escape character is '^]'.

FreeBSD (banana.kame.net) (ttyp8)

login:
```

IPv6 Implementations

IPv6 implementations (1)

- Layer 2
 - Nothing tricky, you just need normal switch/hub
 - Some of ethernet card/driver have problem with multicast
 - Multicast MUST work properly
- Routers (like Cisco)
 - Many vendors are shipping IPv6 in their firmware, just a matter of config
 - shipping: Cisco, Juniper, Hitachi, Fujitsu, NEC, Yamaha, IJ, Allied-Telesys
 - testing: Extreme, Foundary, ...
 - Dialup servers/DSL devices are lagging behind

IPv6 implementations (2)

- End node
 - Linux: kernel is IPv6 ready (2.1 or later)
 - Need to reinstall most of the userlands
 - Status depends on the distro you are using
 - BSD: all *BSDs are ready
 - Windows NT: MS Research release
 - Windows 95/98: Hitachi Toolnet6, Trumpet Fanfare
 - Windows XP: disabled by default (for developers), near-future service pack will enable it
 - Solaris: ready since Solaris 7
 - MacOS X: 10.2 is IPv6 ready (for developers), 10.3 is fully IPv6-ready
 - DEC/Compaq, IBM AIX
- Many of the implementation can become routers
- IPsec on IPv6
 - KAME, NRL, MSR IPv6, IBM AIX

IPv6 implementations (3)

- Applications: anything you want!
 - Either from master distribution, or patch
 - apache2 bind9 squid mozilla lynx grail w3m
 - kaffe(java interpreter) perl ruby python
 - icecast ...
- Routing daemons
 - Similar to IPv4, as we have limited set of algorithms
 - RIPng: similar to RIP
 - BGP4+: multiprotocol version of BGP
 - OSPFv6 (OSPFv3): similar to OSPF
 - zebra/quagga
- Advanced features
 - There are inherently difficult problem remains
 - Experimental code is available
 - diffserv, IPsec policy control, mobile-ipv6

Home appliances and other devices

- New applications will boost IPv6 deployment
- Home appliances
 - Lots of interests, research ongoing, some real products
- Building controllers
 - Lots of devices, autoconfig, need to be future-proven, long lifetime
- Cellphones
 - Nokia/Ericsson are pushing backbone using IPv6, handset using IPv6/VoIP
- Other applications
 - "IP in every taxicab" experiment (Nagoya Japan)
 - IPv6 multicast video streaming events

BSD implementation - KAME project

- IPv4: BSD UNIX kernel was the reference
- IPv6: documents comes first, no de-facto reference code
 - Interop, availability, education...
- Goals: provide IPv6/IPsec reference code to the world
 - BSD license: can be incorporated into product
 - Incorporate experimental protocols and recent technologies
 - mobile-ip4/6, IP over satellite, traffic shaping/diffserv
- 10 core members from 8 companies, including IJ
- Suborganization of WIDE consortium

KAME status

- IPv6
 - Latest spec + corrections to spec
- IPsec + IKE, all made in Japan
- Good note PC support, flow control, ATM PVC
- Routing daemons
- IPv6 multicast
- Bunch of IPv6-ready applications (ports/pkgsrc)
- Various physical medium
 - Ethernet, ATM PVC, sync ppp, async ppp
- SCTP, DCCP, mobile-ip6, queueing policy (diffserv), and other interesting things
- Integrated into NetBSD, OpenBSD, FreeBSD, BSD/OS, MacOS X
 - For normal use, plain NetBSD/OpenBSD/FreeBSD/MacOS X installation is okay
- Used as a base code for Juniper JunOS, Windriver VxWorks
- Available from <ftp.kame.net>
 - Weekly SNAP kit: for those who want more hacking than stability

Using KAME code

- Most applications are seamless
 - ftp, telnet, rsh, ssh, ...
- IPv6 becomes visible only when you use numeric addresses
 - ping6 ::1
 - telnet ::1
- Socket APIs are updated with **AF_INET6** and **sockaddr_in6**
 - Basic applications are trivial
 - Routing socket, **setsockopt**, and other tricky parts need update

URLs of interest

<http://www.wide.ad.jp/>

<http://www.v6.wide.ad.jp/>

<http://www.kame.net/>

<http://www.6bone.net/>

<http://www.ipv6.org/>

<http://playground.ijlab.net/> (the slides will be available here)

Homework

- Configure an IPv6 node and IPv6 network
- what you have done?
- what was it like compared to IPv4 operations
 - even if you feel no difference, it's okay
- For testing: <http://www.kame.net/> is available over IPv4 and IPv6
 - If you click the page over IPv6, the turtle will dance
- Depends on the implementation you are using
 - If you don't have any of these, install NetBSD 1.6 onto your machine

Homework - Windows

- Make sure to use Windows XP (95, 98, NT are difficult)
- invoke "ipv6 install" from the command line
- various command line tools should become IPv6 capable
- Try to click <http://www.kame.net/>

Homework - Macintosh

- Make sure to use MacOS 10.3
- Enable IPv6 via System Preference - Network
- try using ssh/telnet/ftp
- see IPv6 service is enabled with "netstat -an"

Homework - Linux

<http://www.bieringer.de/linux/IPv6/status/IPv6+Linux-status-distributions.html>

- Enable IPv6 services by `/etc/inetd.conf`
- `"telnet ::1"` or `"ssh ::1"` should be possible even without IPv6 external connectivity
- Configure 6to4 network
- Install Mozilla and try <http://www.kame.net/>

Homework - BSD UNIX

- Use the latest version possible (FreeBSD 4.9, NetBSD 1.6.2, OpenBSD 3.4)
- Enable IPv6 services by `/etc/inetd.conf`
- "`telnet ::1`" or "`ssh ::1`" should be possible even without IPv6 external connectivity
- Configure 6to4 by looking at <http://www.netbsd.org/Documentation/network/ipv6/>
- Install Mozilla and try <http://www.kame.net/>