ELEC / COMP 177 – Fall 2011

Computer Networking
→ Future of the Internet

Some slides from Kurose and Ross, *Computer Networking*, 5th Edition

Upcoming Schedule

- Today
  - *Ethernet: 10G and Beyond*
  - *IPv6*
  - Project 3 – **Due today at 11:55pm**
    - Questions?
    - More time?
- Thursday
  - Exam review

Final Exam

- **Tuesday, December 13th – 8am-11am**
- Short answer format
- Comprehensive – covers entire semester
  - No paper resources (books, notes, …)
  - No electronic resources (computer, phone, …)
  - No human resources (except for you!)
- Time limited – 3 hours max
- Just you, your pencil, and paper
  - *You can bring a calculator if you want to convert from binary<->decimal*

HTTP Proxy Discussion

- Handling incomplete sends()
  - See Beej's Guide section 7.3
- How long does the socket between the web browser and proxy stay open?
- How long does the socket between the proxy and web server stay open?
- What does my proxy do with all the headers sent to it by the web server?
- What should my proxy do if the web server sends a 301 redirect? A 404 error?

Homework #4 and #5 Review

- *Solutions posted in Sakai resources folder*
- Discuss selected problems now…

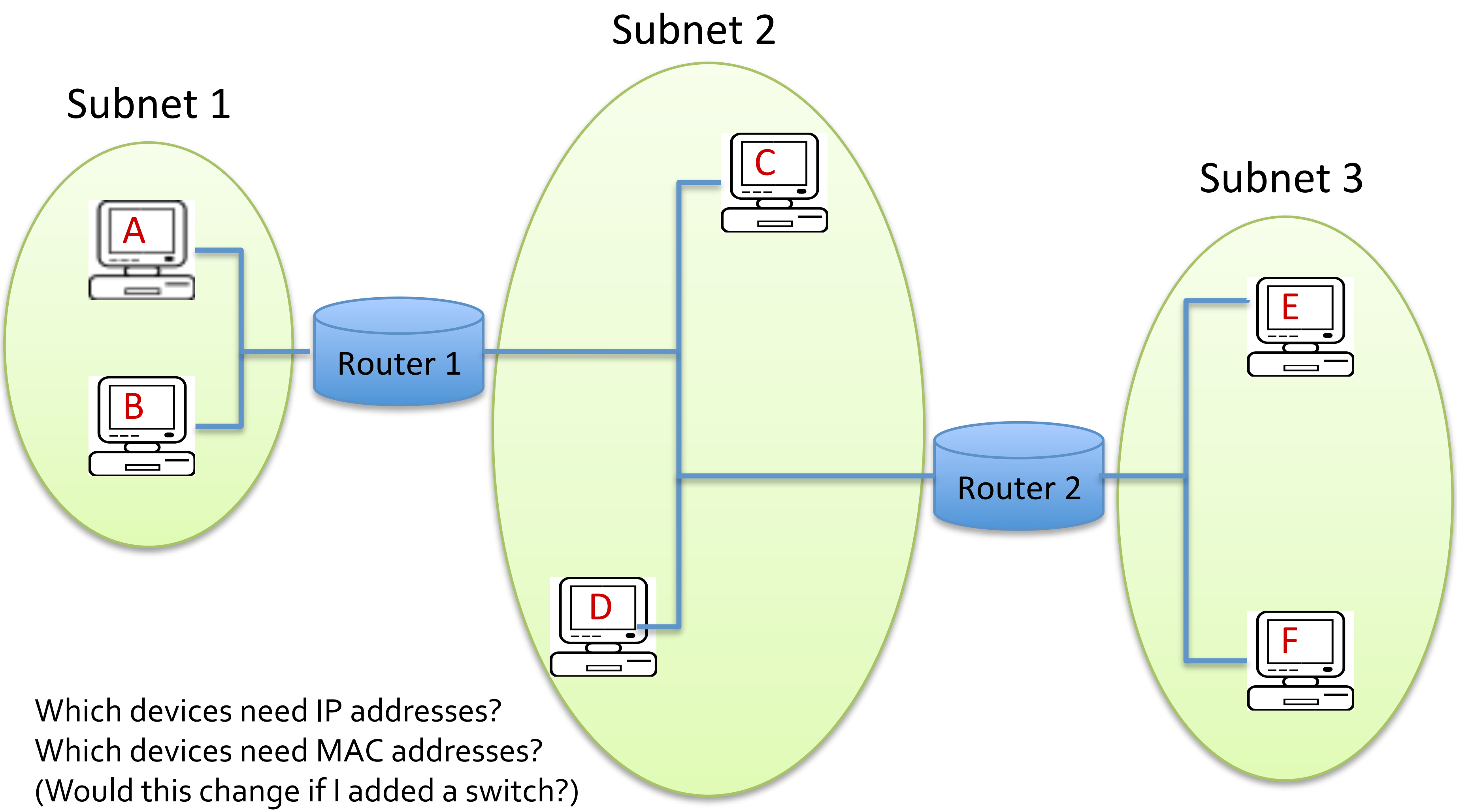
Homework #4 – Problem 2

- IP Address: 87.15.213.53
  - In binary: `01010111.00001111.11010101.00110101`
- Netmask: 255.255.255.128
  - In binary: `11111111.11111111.11111111.10000000` *(25 1's!)*
- **Length of subnet?**
  - **25 bits**
- **Subnet address?**
  - AND IP address with netmask: **87.15.213.0** (or **87.15.213.0/25**)
- **Broadcast address?**
  - All 1's in the host field (i.e. after subnet address): `01010111.00001111.11010101.01111111` or **87.15.213.127**
- **Lowest and highest valid host IP?**
  - Lowest: 1 above subnet address: **87.15.213.1**
  - Highest: 1 below broadcast address: **87.15.213.126**

Homework #4 – Problem 3

- **How does the host distinguish between hosts on its local subnet, and hosts <u>not</u> on its local subnet?**
  - Bitwise AND host IP with netmask
  - Bitwise AND destination IP with netmask
  - Compare – If equal, destination is on same subnet
    - *Essentially, we're comparing subnet addresses*
- Your computer does this for every outgoing IP packet
  - **Who do I send the packet to if it is on the local subnet?**
    - **Directly to the destination**
  - **Who do I send the packet to if it is <u>not</u> on the local subnet?**
    - To your **default gateway** (or another router, if configured)

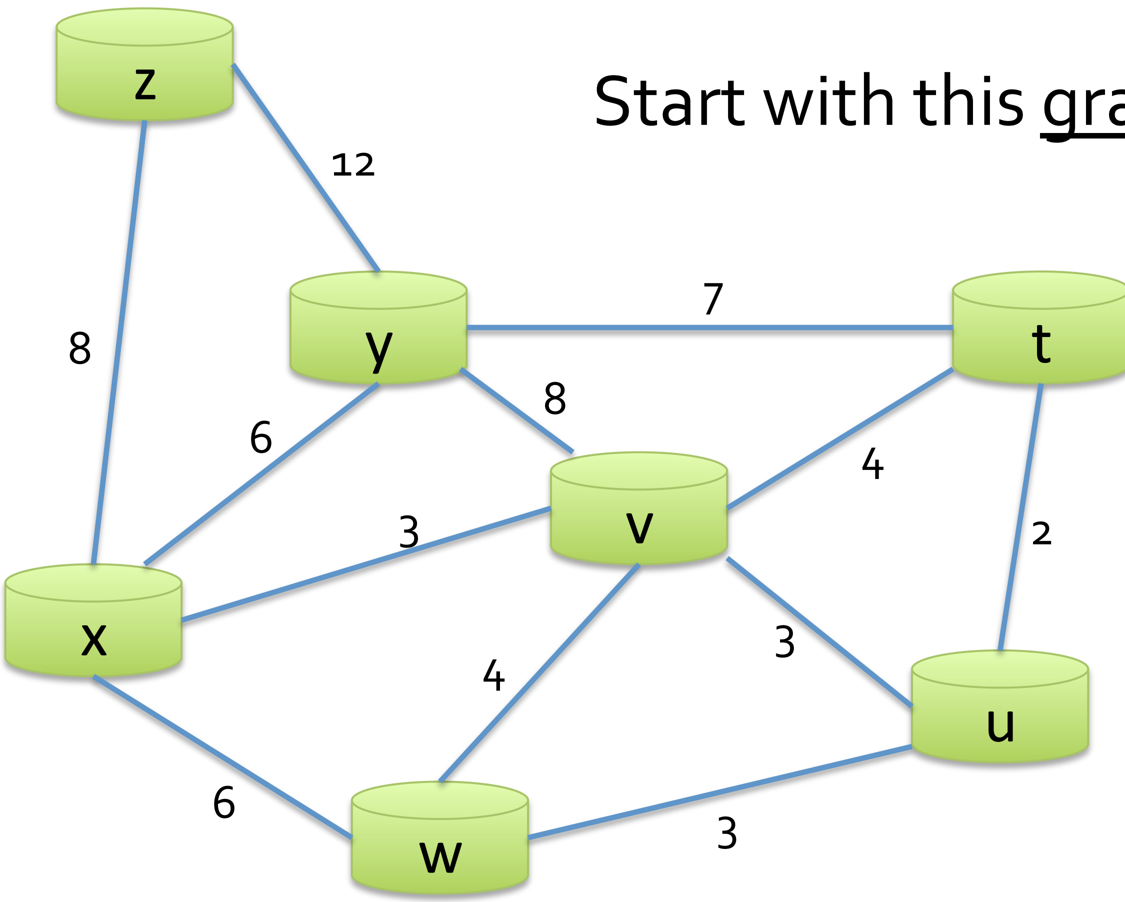
Homework #4 – Problem 5



- Which devices need IP addresses?
- Which devices need MAC addresses?
- (Would this change if I added a switch?)
- Packet from E to B with ARP cache in E empty…

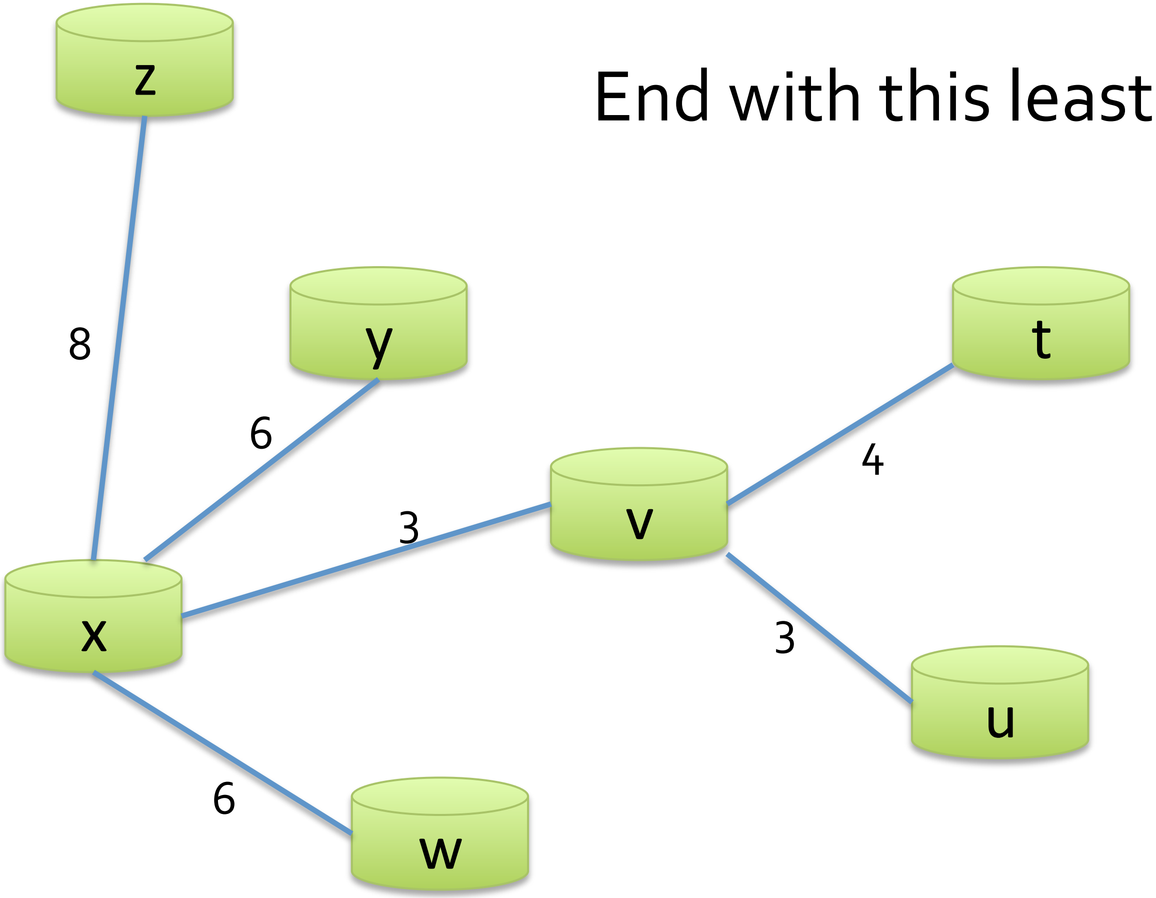
Homework #5 – Problem 1

Start with this graph...



Homework #5 – Problem 1

End with this least-cost tree



Homework #5

Link state / Dijkstra's Algorithm
D(t) = Least cost to t
p(t) = Predecessor node on path to t

Step	N'	D(t), p(t)	D(u), p(u)	D(v), p(v)	D(w), p(w)	D(y), p(y)	D(z), p(z)
0	x	Inf	Inf	3,x	6,x	6,x	8,x
1	x,v						
2							
3							
4							
5							
6							

Future of the Internet

Ethernet

Future of Ethernet

- Ethernet standard (802.3) first published in 1983
- Much of the original standard has been discarded at 1Gbps and above:
  - No more shared bus or thick coax cable, only point-to-point links
  - No more Carrier Sense Multiple Access or collisions
  - No more Manchester encoding

"Today's Ethernet technology is extremely diverse and has very little in common with what appeared in '74. The good news is that they still call it Ethernet, and that's my word."

Bob Metcalfe, 2003

Future of Ethernet

- Some parts remain
  - Ethernet frame format
  - Business model
    - Companies compete with proprietary designs
    - IEEE standards ensure interoperability
    - Standards evolve rapidly (but with backwards compatibility)

"What Ethernet is today is more than a packet format or media access algorithm--it is a business model"

"If they want to call 802.11 wireless Ethernet, I'm all for it, especially because it's reminiscent of the Aloha network from which 802.11 is derived"

Bob Metcalfe, 2003

10 Gbps Ethernet

- Diversity of physical layer options
  (just like previous Ethernet!)
  - 6 fiber optic standards + 3 copper standards
- Marketplace will determine which survive
  - Possibly 10GBASE-T, which can use normal twisted-pair Ethernet cables
  - Currently expensive! (NIC $500+, small switch $10k+)

Image from
http://www.myricom.com/

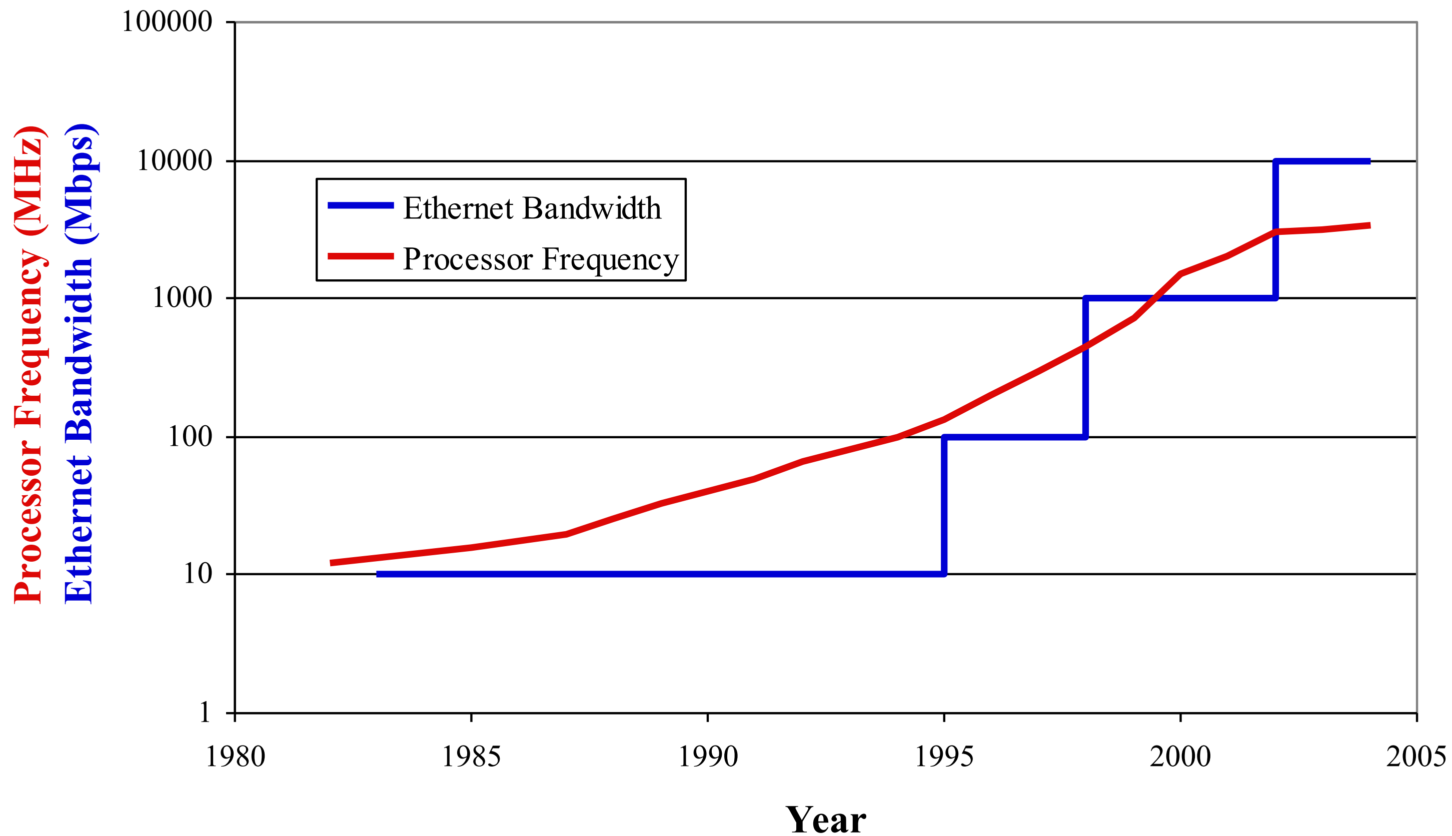
40/100 Gbps Ethernet

- Data centers are already aggregating multiple 10Gbps lines together – Always demand for more bandwidth!
- Design objectives set in 2006 – research ongoing
  - 100 Gbit/s (at the client interface)
  - Full-duplex operation only
  - Preserve 802.3 / Ethernet frame format at the MAC level
  - Preserve current minimum and maximum frame size
- Current status
  - Standard ratified in 2010
  - Commercial routers shipping ($$$$)
    - No NICs yet…

Network Performance – Achievable?

- In 2012, will I be able to buy a 100Gbps NIC, plug it into my computer, and expect to get 100Gbps of throughput?
  - Not even close!
  - Challenging to produce/consume that much data
  - Challenging to produce/consume headers for that many frames
    - 81,300 frames/s at 1Gbps
    - 813,000 frames/s at 10Gbps
    - 8,130,000 frames/s at 100Gbps

Network vs Processor Performance



Network performance is outpacing processor (single core) performance

Network Performance – Achievable?

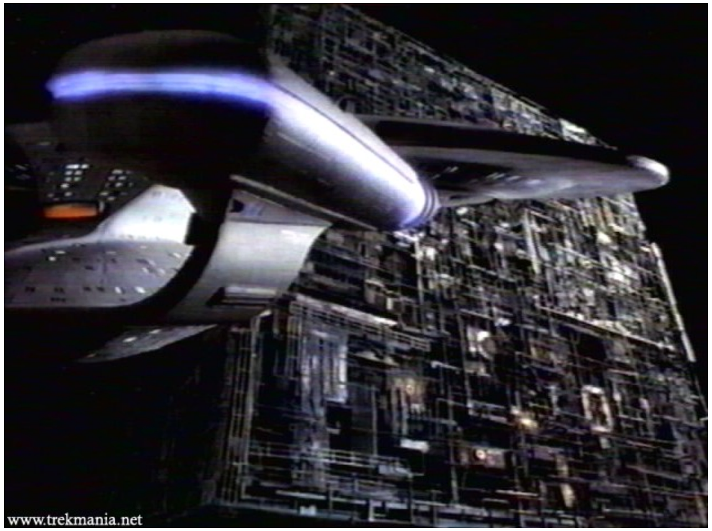
- Better NIC designs are needed
  - Transmit data path:
    TCP Segmentation Offload (TSO)
    - Send the NIC a large buffer (64kB)
    - Have NIC segment data into multiple packets
  - Receive data path:
    Large Receive Offload (LRO)
    - More efficient for network stack to process a large buffer of data (from a single stream) than many small buffers
    - Data must be aggregated either on the NIC or in software
      - Surprisingly, even software method can improve performance by reducing overhead of higher layers of the network stack

Network Performance – Achievable?

- Better OS architectures are needed
  - Multicore is a given
  - Efficiently parallelized network stacks are required
  - How many cores should the network subsystem scale to?
  - How do we divide the work?
    - Per connection?
    - Per message?

Future of Ethernet

- Battle plan
  - Attack on all fronts!
- Targets
  - Backplane Technology – Blade servers
    - Short distance, < 1 meter
  - SAN - Storage Area Networks
    - Short distance, inside datacenter
  - MAN - Metropolitan Area Network
    - Long distance, tens of km
- Marketplace will determine if these new products succeed



Scaling Ethernet

- Can I have a single switched Ethernet network spanning the entire world?
  - Commercial switches only have ~16,000 entry forwarding table
  - How do the switches find the destination computer?
    - Broadcast to every computer in the world?
  - Ethernet scalability has limits
- Routing / Higher-Layer Protocols Needed
  - Partition network into discrete LANs
    - Link to other LANs may also be Ethernet, but link is not accessible via a switch, but instead a router

Future of the Internet

IPv6

IP Versions

ID	Description
0-3	Unused: Development versions of IP
4	Current network-layer protocol
5	Unused: Experimental stream protocol – ST
6	New network-layer protocol (1996)
7-9	Unused: Experimental protocols – TP/IX, PIP, TUBA
10-15	Not allocated

Why Replace IPv4?

- Performance?
  - Quality of Service?
- Security?
  - Authenticating users?
  - Tracking criminals / spammers?
  - Denial of service?
- New applications?
  - Mobile devices?
- Routing is scaling out of control?

- IPv6 does not attempt to solve all problems with computer networks
  - Actually, it only really solves one of them!

Why Replace IPv4?

- The problem
  - IPv4 has ~4.3 billion addresses
  - World has ~6.6 billion people!
    - How many internet-capable devices per person?
- IP address exhaustion
  - Internet will not "collapse", but new devices / networks will not be able to join
- When? Now – 2011!
  - Consuming about one /8 block (16.78 million addresses) per month

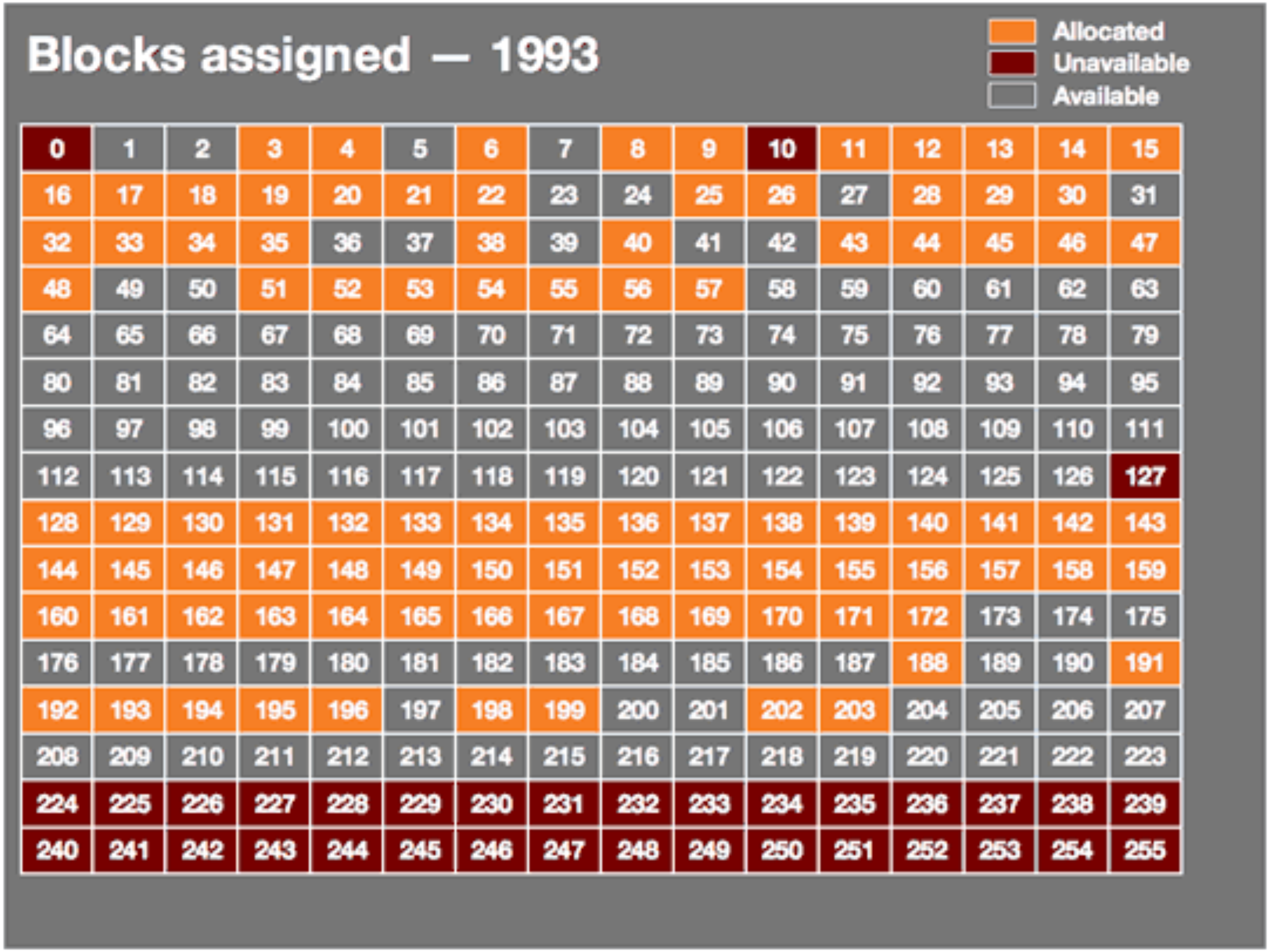


Figure from http://arstechnica.com/articles/paedia/IPv6.ars

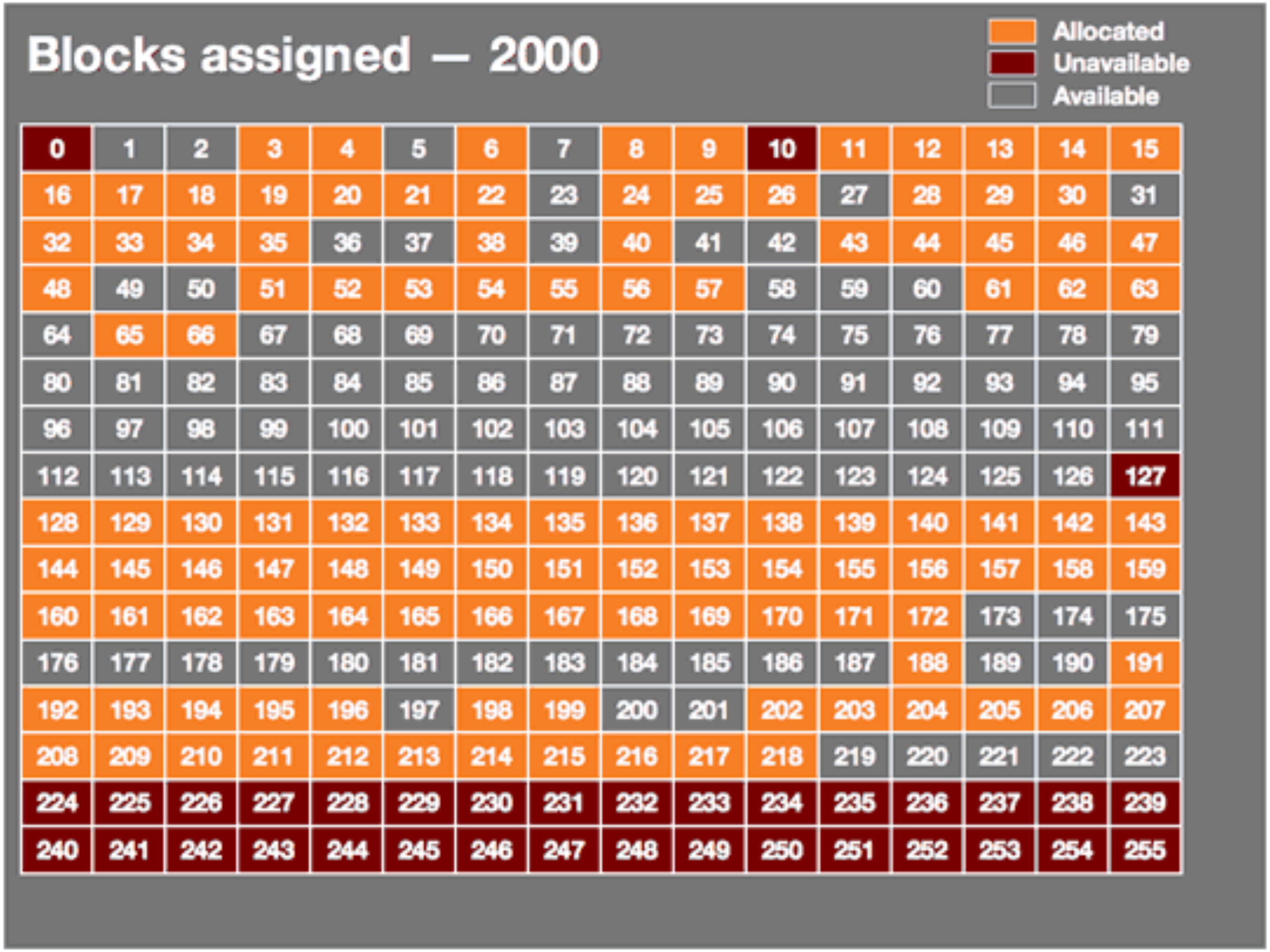


Figure from http://arstechnica.com/articles/paedia/IPv6.ars

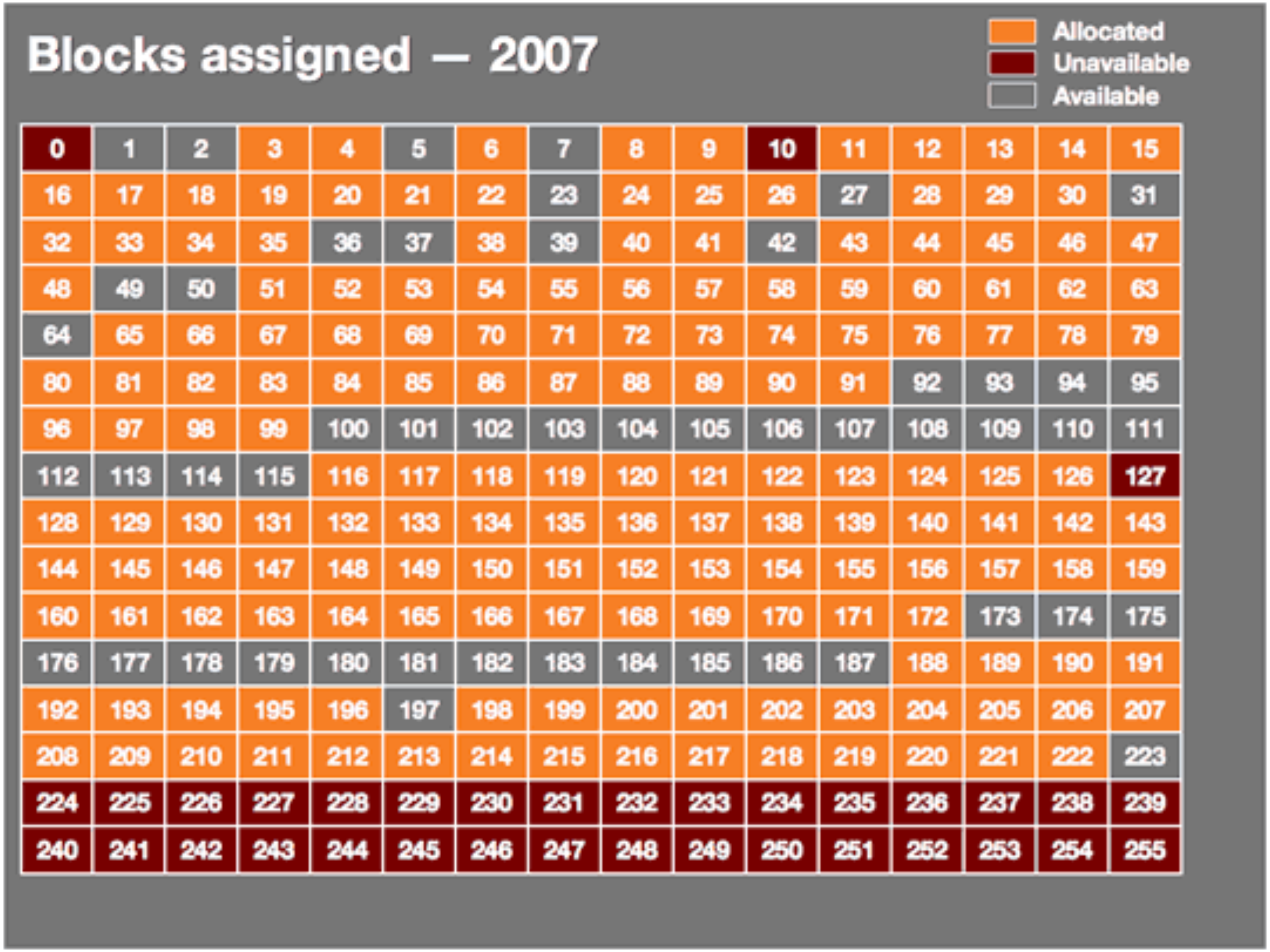
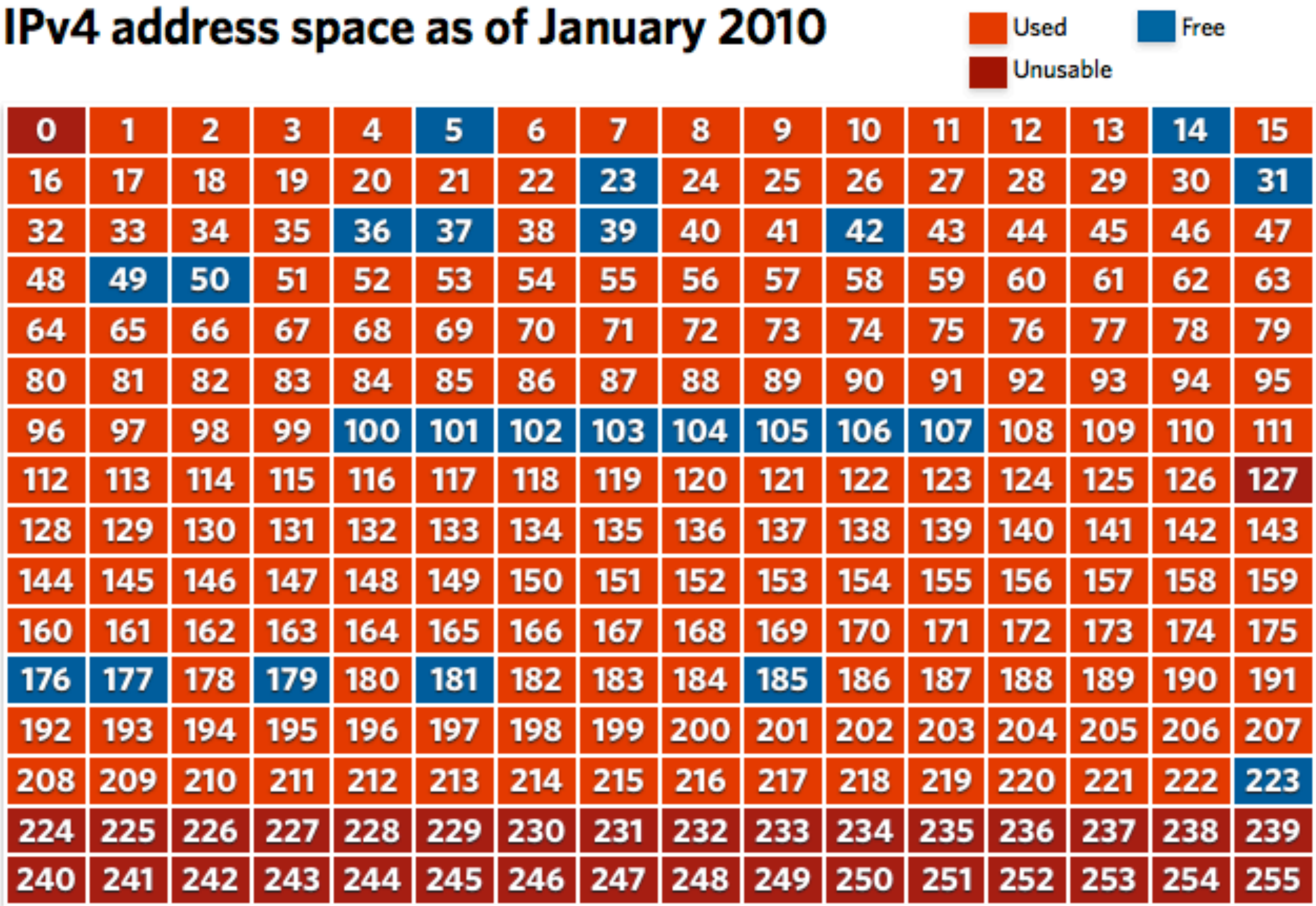
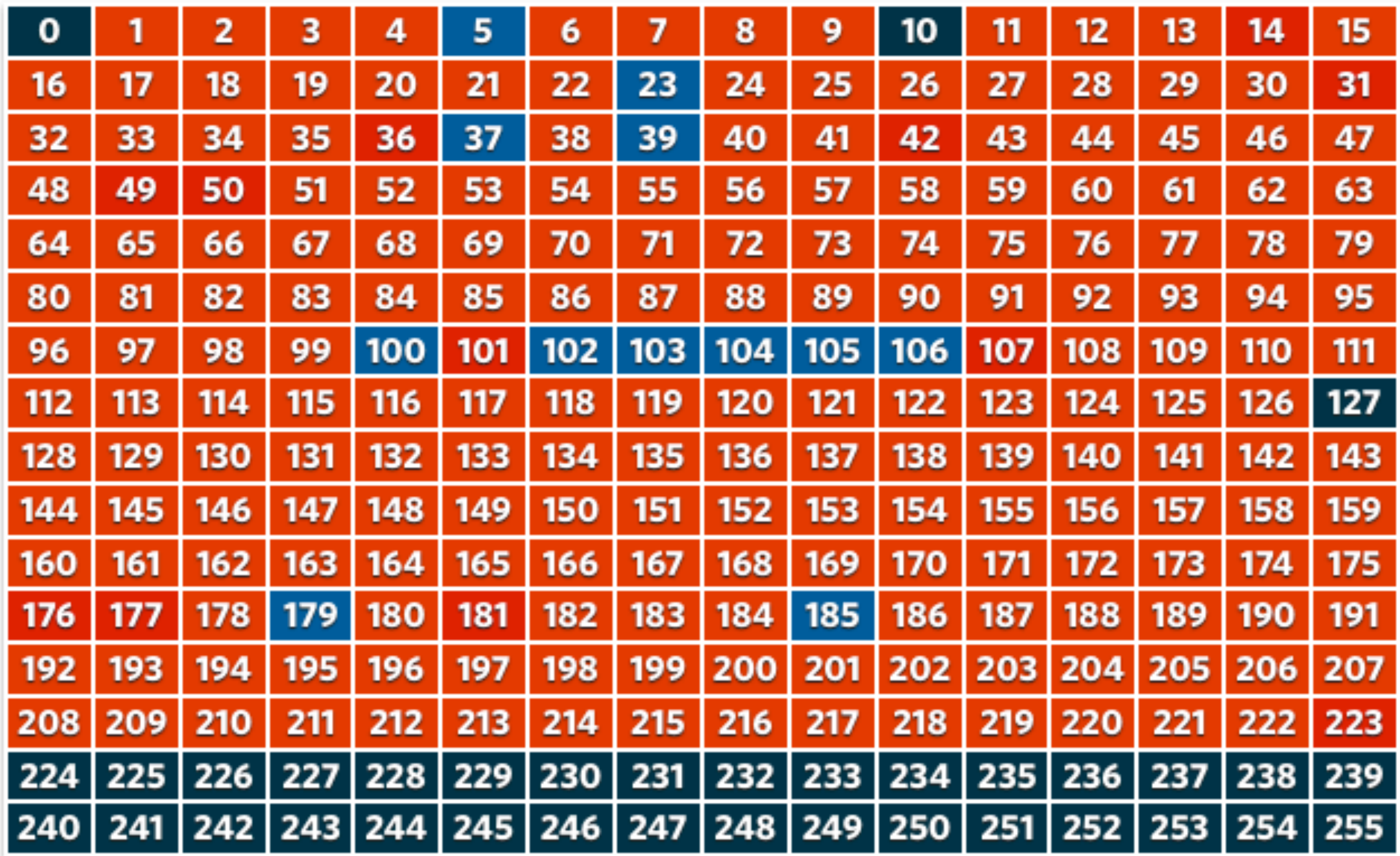


Figure from http://arstechnica.com/articles/paedia/IPv6.ars



http://arstechnica.com/tech-policy/news/2010/01/90-of-ipv4-address-space-used-ipv6-move-looking-messy.ars

IPv4 address space as of October 18, 2010



IPv4 Address Space

- Unavailable Addresses
  - 10.x – Private Addresses
    - Along with 192.168.x and 172.16.x to 172.31.x
  - 127.x – Local Loopback Addresses
    - Why an entire /8?
  - 224.x to 239.x — Multicast groups
  - 240.x to 254.x — Reserved for "future use"
    - Waste of address space
    - Impossible to re-use today because most IP software flags these addresses as invalid
  - 91 entities with entire class A's (Govt, IBM, GE, HP, MIT, …)
- Current Allocation
  - http://www.iana.org/assignments/ipv4-address-space

IPv4 vs IPv6 - Similarities

- Datagram
  - Each packet is individually routed
  - Packets may be fragmented or duplicated
- Connectionless
  - No guarantee of delivery in sequence
- Unreliable
  - No guarantee of delivery
  - No guarantee of integrity of data
- Best effort
  - Only drop packets when necessary
  - No time guarantee for delivery

IPv4 vs IPv6 - Differences

- Address Length
  - IPv4 – 32 bits (2^{32} = ~4 billion)
  - IPv6 – 128 bits (2^{128}= ~340 trillion, trillion, trillion)
- Security – IPSec support required in IPv6
  - IPSec encrypts each IP packet independently
- Reliability – No Header Checksum in IPv6
  - Easier for routers – No need to update checksum after decrementing TTL
  - Relies on link-level error checking
- Quality of Service
  - Label data flows for special priority levels at routers
- Simplified Header Format
  - Infrequently used fields are optional

IPv6 Address Notation

- 128 bits – 8 groups of 4 hex digits
  - `2001:0db8:85a3:08d3:1319:8a2e:0370:7334`
- User friendly! Easy to remember!
- “Helpful” Shortcuts:
  - Omit leading zeros in a group (`0005:0db8:`… is equivalent to `5:db8:`…)
  - Collapse groups of all-zeros with :: (`2001:0000:0000:0000:0000:8a2e:0370:7334` is equivalent to `2001::8a2e:0370:7334`)

IPv6 – Routing

- **How can having bigger IP addresses (128 bits) make routing easier?**
  - Larger address space allows more intelligent network organization
    - Addresses match physical network organization
    - Collapse routing table entries
- Several competing proposals on the details…
- Basic constraint on all designs
  - Use upper 64 bits for routing
  - Use lower 64 bits for interface ID
    (clients pick this randomly or based on MAC address)

Routing

- **Besides the address layout, how does IPv6 make routing easier?**
  - No checksum calculation
  - No fragmentation
  - Infrequently used headers are optional
- **How does IPv6 make routing harder?**
  - Forwarding table entries 2x-4x larger
  - Need to route both IPv4 and IPv6 for the foreseeable future

Deployment

- Why should I deploy IPv6 today?
  - My customers can reach anywhere on the Internet today
  - Google, CNN, etc... will always be reachable
  - Only new applications / users will suffer
- How do I deploy IPv6?
  - Flip a switch across the internet?
- Legacy routers may not be upgradeable
  - Hardware implementations are fixed
  - Software implementations may be insufficiently capable (either incapable or only at low performance)
- Islands of IPv6 in the sea of IPv4
  - Dual network stacks support both IPv4 and IPv6
  - Tunnel IPv6 across IPv4 networks
- Need to upgrade other systems
  - DHCP
  - DNS (9 of 13 root nameservers as of Dec 2011)
  - Firewalls, traffic shapers, etc.

IPv6 – Original Plan

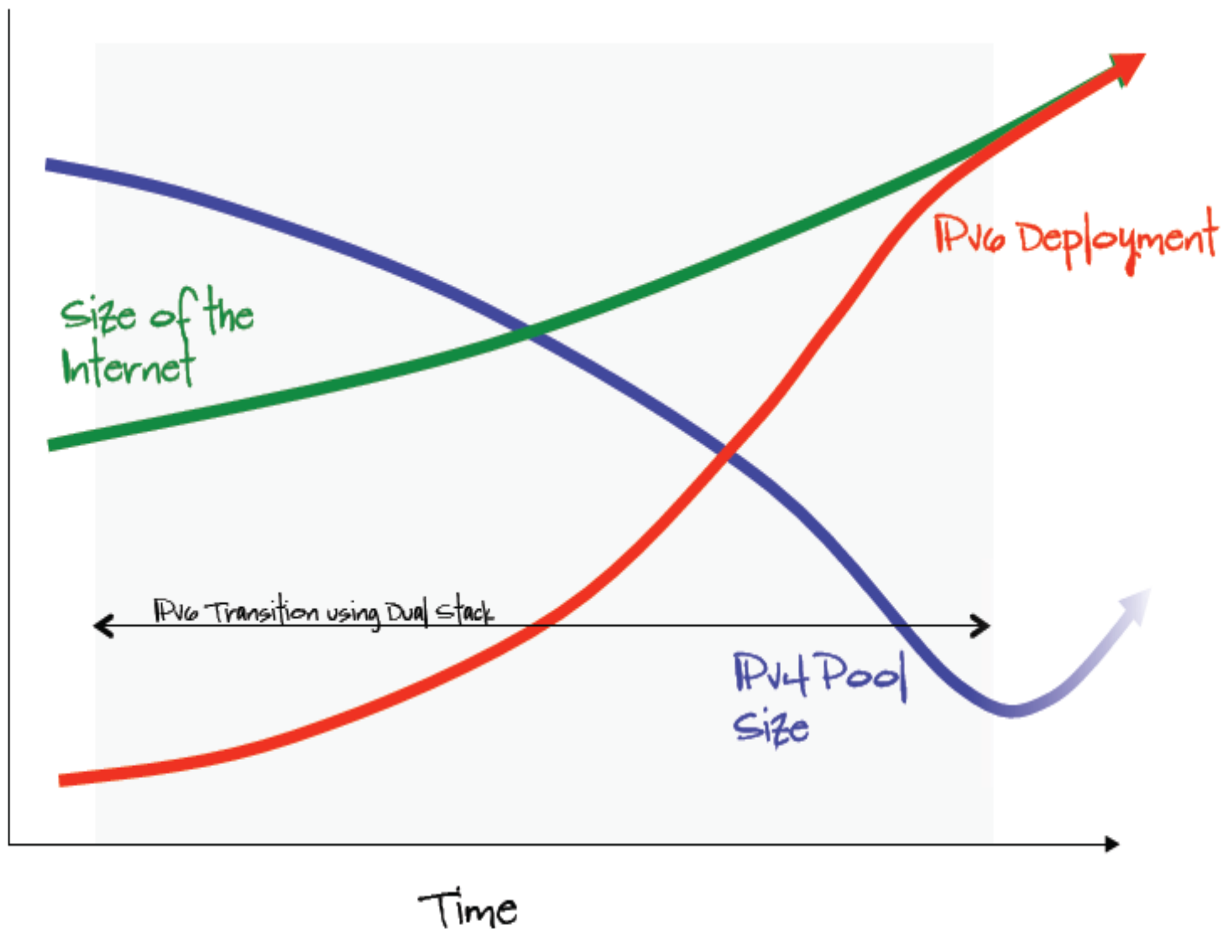
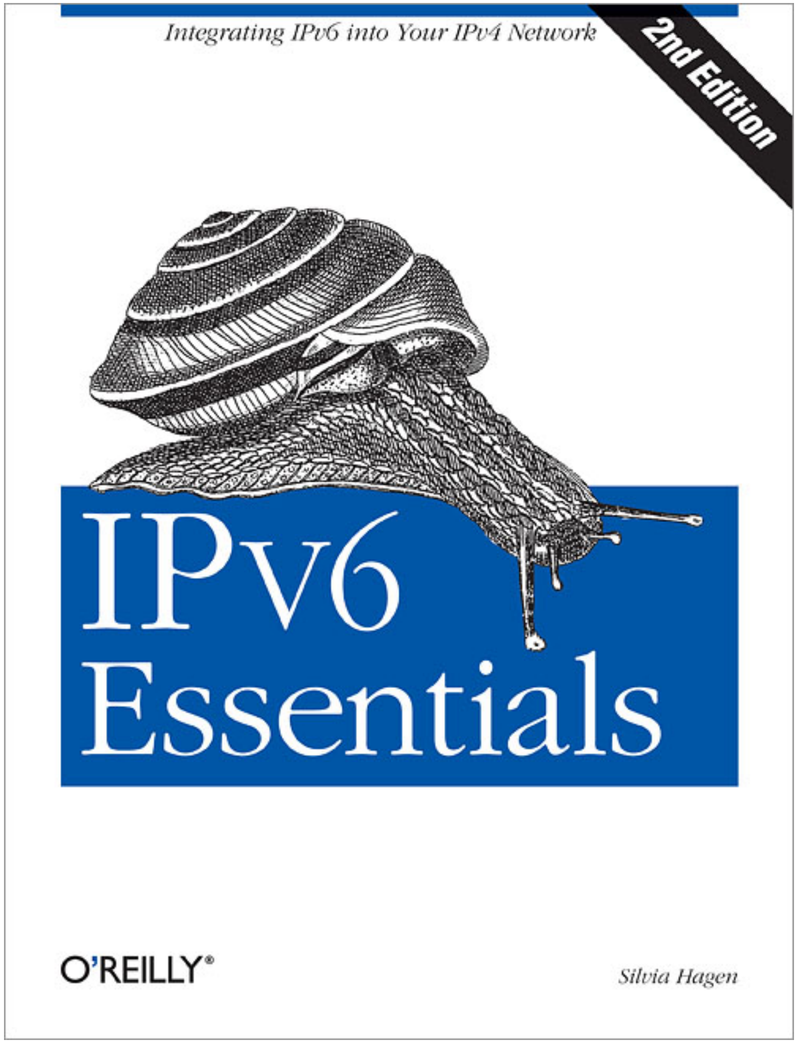


Figure from http://www.potaroo.net/presentations/2008-11-17-ipv6-failure.pdf

IPv6 – Current Status



IPv6 – The New "Plan" (?)

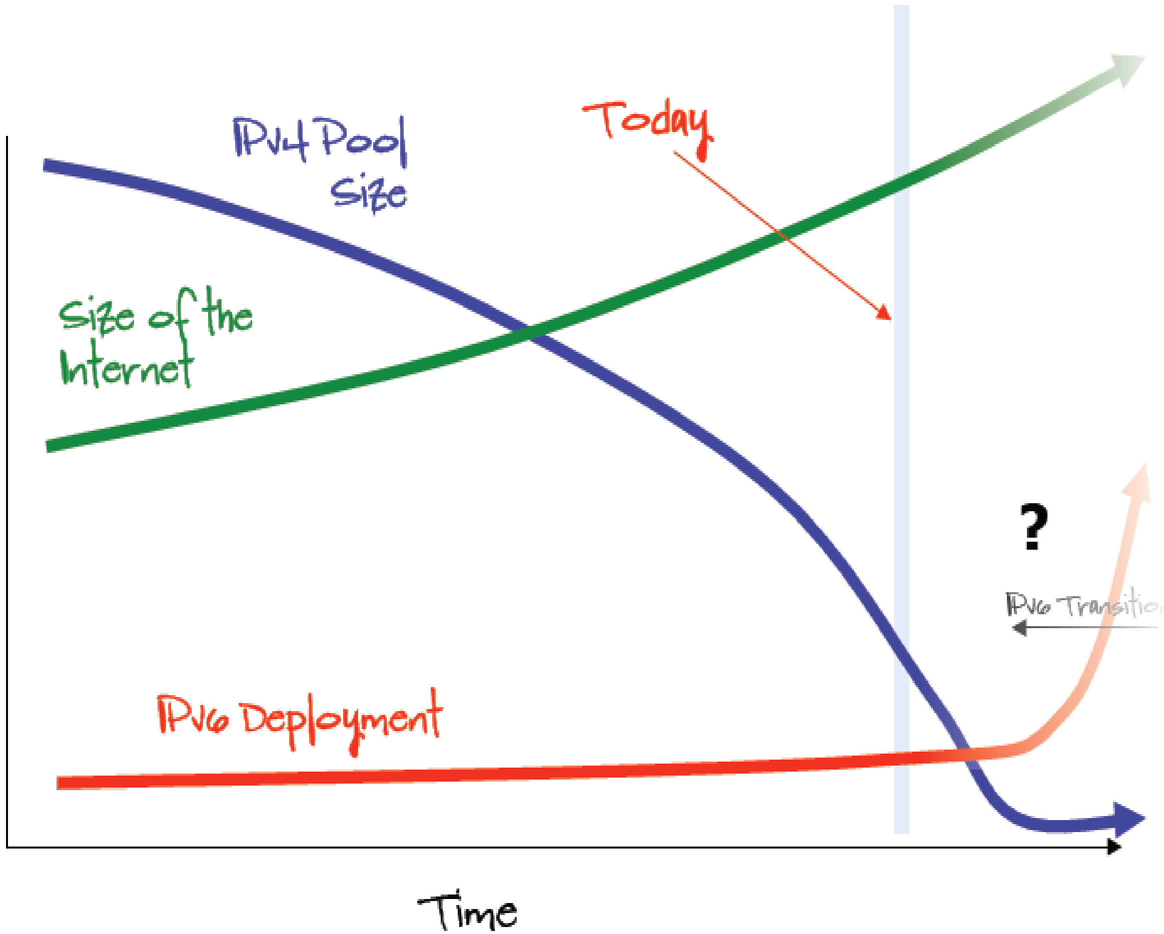
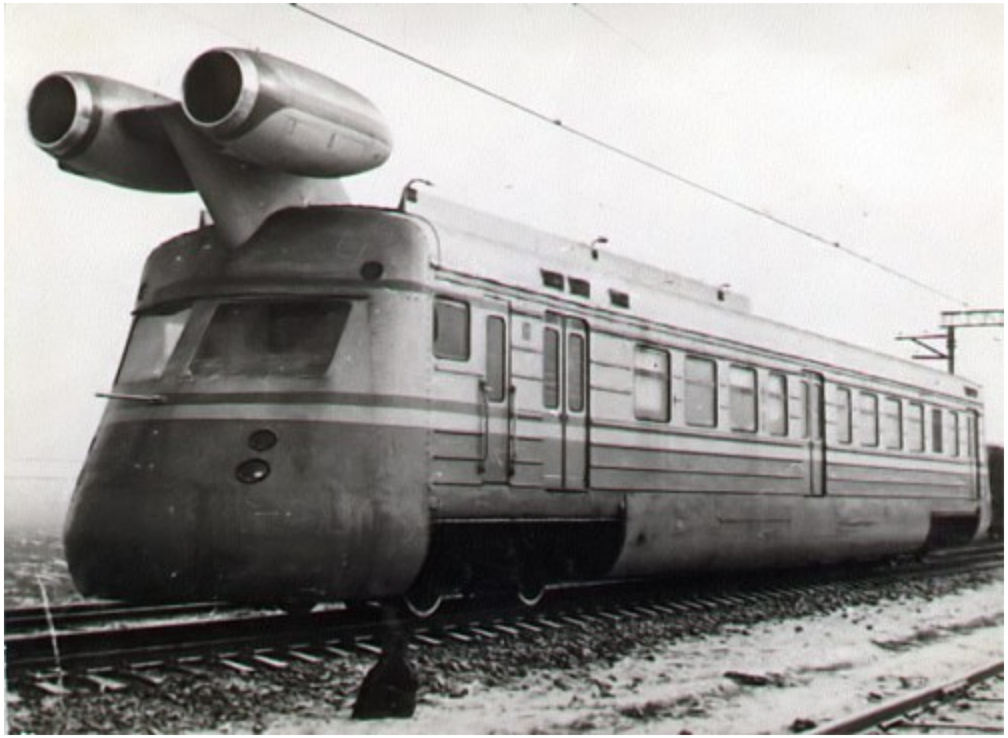


Figure from http://www.potaroo.net/presentations/2008-11-17-ipv6-failure.pdf

IPv6 – Failure is an Option



Is this IPv6?

IPv6 – Failure is an Option

- What happens if IPv6 "fails"?
  - Failure is defined as anything less than a complete migration from IPv4 to IPv6
  - Do we stop allowing new hosts to connect to the internet?
- What about using NAT? (address translation)
  - Observation: Only 5-20% of assigned IPs are actually used by hosts.
  - Solution: Use lots of NAT to reclaim unused addressed
- What happens if this works, and we build "carrier-grade" NAT everywhere?
  - No more end-to-end connectivity?
  - Need coordination with ISP to deploy new services?
  - New opportunities for ISPs to filter traffic and charge for services?

http://www.potaroo.net/presentations/2008-11-17-ipv6-failure.pdf

IPv6 – Failure is an Option

- If an organization deploys NAT extensively, how can you get them to give up the reclaimed addresses?
- IP Address Marketplace
  - Can we create a marketplace? (Currently "forbidden" to sell IP blocks)
  - Imagine: *"For Sale: One Lightly-Used IP Block (only used by grandma to check email on Sunday)"*
  - Same problems as buying a used car:
    - Does the person selling the IP block actually "own" it?
    - What is the condition of the IP block? (If used for spam or illicit activity, IP block may be in blacklists worldwide)

http://www.potaroo.net/presentations/2008-11-17-ipv6-failure.pdf