15-441/641: Computer Networks
Intradomain Routing

15-441 Spring 2019
Profs **Peter Steenkiste** & Justine Sherry

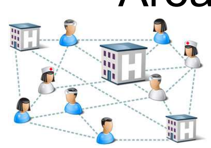
Fall 2019
https://computer-networks.github.io/sp19/

Carnegie Mellon University

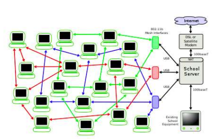
Outline

- IP design goals
- Traditional IP addressing
  - Addressing approaches
  - Class-based addressing
  - Subnetting
  - CIDR
- Packet forwarding

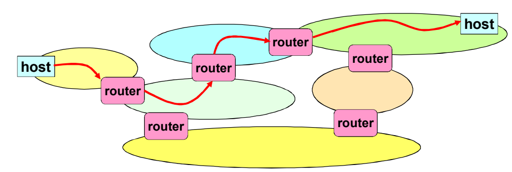
So far you know how to build a Local Area Network



How do we get them to talk to each other?



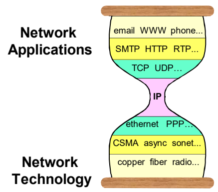
Logical Structure of an Internet



- Interconnection of separately managed networks using routers
  - Individual networks can use different (layer 1-2) technologies
- Packet travels from source to destination by hopping through networks
  - “Network” layer responsibility
- How do routers connect heterogeneous network technologies?

Solution: Internet Protocol (IP)

- Inter-network connectivity provided by the Internet protocol
- Hosts use Internet Protocol to send packets destined across networks.
- IP creates abstraction layer that hides underlying technology from network application software
  - Allows range of current & future technologies
  - WiFi, traditional and switched Ethernet, personal area networks, …



The Packet as an Envelope

But need a local addressing header to travel between routers

IP address identifies final destination

Host wants to send…

Local Address Header (Ethernet)
To: Destination Host
From: Sender Host

IP Header
To: 123.45.67.89 (Destination Host)
From: 169.229.49.157 (Sender Host)

HTTP Packet Payload:
GET nyan.cat….

The Packet as an ~~Envelope~~ Set of envelopes

But need a local addressing header to travel between routers

IP address identifies final destination

Host wants to send…

Local Address Header (Token Ring)
To: Destination Host
From: Local Router

IP Header
To: 123.45.67.89 (Destination Host)
From: 169.229.49.157 (Sender Host)

HTTP Packet Payload:
GET nyan.cat….

Datalink headers may differ across networks

But IP header remains unchanged!

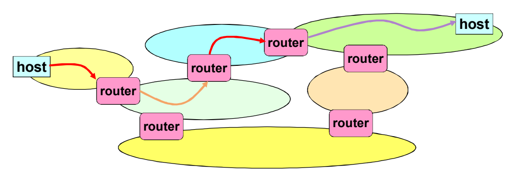
Traveling through the Internet

- Source adds all headers (HTTP, transport -> PHY)
- Each router:
  - Removes datalink layer
  - Uses IP header to make forwarding decision
  - Adds data link layer header for next network
- Destination removes all headers (PHY -> HTTP)

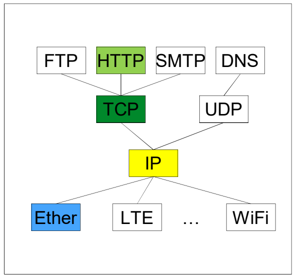
Datalink Header
To: Next hop router
From: This router

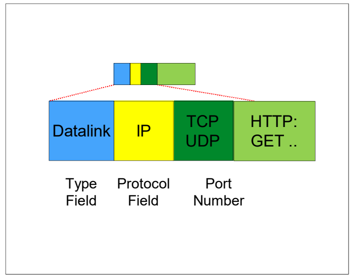
IP Header
To: 123.45.67.89
From: 169.229.49.157

HTTP Packet Payload:
GET nyan.cat….



Protocol Demultiplexing





- What layers do not need a protocol field?

What are the Goals?

- LANs: "Connect hosts" → switching:
  - "Wire" abstraction: behaves like Ethernet – helps manageability
  - Only has to scale up a "LAN size"
  - Availability
- Internet: "Connect networks" → routing:
  - Scalability
  - Manageability of individual networks – contributes to scalability
  - Availability
- Affects addressing, protocols, routing

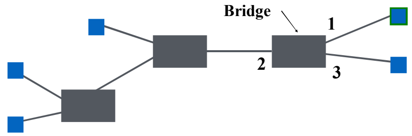
Outline

- IP design goals
- Traditional IP addressing
  - Addressing approaches
  - Class-based addressing
  - Subnetting
  - CIDR
- Packet forwarding

Addressing and Forwarding

- Flat address space with smart routers
  - Packets carry destination
  - Routers know location of every host
- Flat address space with dumb routers
  - Packet carries a path
- Heirarchical Routing Space
  - What we actually do in IP
- (Table of virtual circuits ids)
  - More on this later, but not today

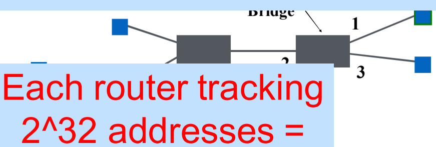
Flat Address Forwarding



MAC Address	Port	Age
A21032C9A591	1	36
99A323C90842	2	01
8711C98900AA	2	15
301B2369011C	2	16
695519001190	3	11

- Bridge/switch has a table that shows for each MAC Address which port to use for forwarding
- For every packet, the bridge "looks up" the entry for the packets destination MAC address and forwards the packet on that port.
  - Other packets are broadcast – why?
- Timer is used to flush old entries

Why is this not a good solution for the Internet? arding



MAC Address	Port	Age
A21032C9A591	1	36
99A323C90842	2	01
8711C98900AA	2	15
301B2369011C	2	16

Each router tracking 2^32 addresses = scalability nightmare

- B for eac MAC Address which port to use for forwardi
- For every packet, the bridge "looks up" the e destination MAC address and forwards the p
  - Other packets are broadcast – why?
- Timer is used to flush old entries

Source Routing

- List entire path in packet
  - Driving directions (north 3 hops, east, etc..)
- Router processing
  - Strip first step from packet
  - Examine next step in directions and forward
- Defined for IPv4 but rarely used
  - End points need to know a lot about network
  - Economic and security concerns
  - Variable header size

Hierarchical Addressing

- Flat addresses – one address for every host
  - Peter Steenkiste: 123-45-6789
  - Does not scale – router table size explodes
  - 630M (1/09) entries, doubling every 2.5 years
  - Why does it work for Ethernet?
- Hierarchical – add structure
  - Pennsylvania / Pittsburgh / Oakland / CMU / Gates / 9th fl / Steenkiste
  - Common "trick" to simplify forwarding, reduce forwarding table
- What type of Hierarchy do we need for the Internet?
  - How many levels?
  - Same hierarchy depth for everyone?
  - Who controls the hierarchy?

IP Addresses (IPv4)

- Unique 32-bit number associated with a host

00001100 00100010 10011110 00000101

- Represented with the "dotted quad" notation
  - e.g., 12.34.158.5

12	34	158	5
00001100	00100010	10011110	00000101

Hierarchy in IP Addressing

- 32 bits are partitioned into a prefix and suffix components
- Prefix is the network component: CMU
- Suffix is host component: Prof. Sherry's laptop at CMU

12	34	158	5
00001100	00100010	10011110	00000101

Network (23 bits) | Host (9 bits)

- Interdomain routing operates on the network prefix
- Destination network operates on the host component

History of Internet Addressing

- Always dotted-quad notation
- Always network/host address split
- But nature of that split has changed over time

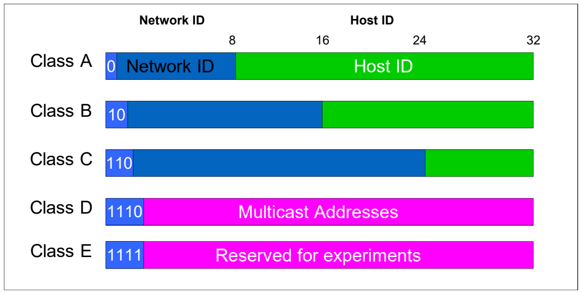
Original Internet Addresses

- First eight bits: network component
- Last 24 bits: host component

Assumed 256 networks were more than enough!

IP Address Structure, ca 1981

Routers know how to get to network ID, but not individual hosts.



IP Route Lookup, ca 1981

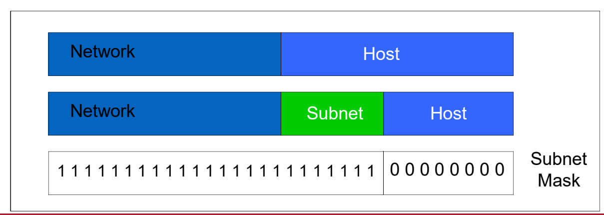
- Address specifies prefix for forwarding table
  - Extract address type and network ID
- Forwarding table structure reflects address structure
  - Logically, a separate forwarding table for each address class
  - For unicast address (classes A-C) entries contain
    - The prefix for a destination network (length 8/16/24)
    - Information on how to forward the packet, e.g., exit port, ..
- www.cmu.edu address 128.2.11.43
  - Class B address – class + network is 128.2
  - Lookup 128.2 in forwarding table for class B
- Tables are still large!
  - 2 Million class C networks

Outline

- IP design goals
- Traditional IP addressing
  - Addressing approaches
  - Class-based addressing
  - Subnetting
  - CIDR
- Packet forwarding

Subnetting

- Add another layer to hierarchy
- Variable length subnet masks
  - Could subnet a network internally into several chunks
- Subnetting is done internally in the organization
  - It is not visible outside – important for management



Subnet Addressing RFC917 (1984)

- Some "LANs" are very big
  - Large companies, universities, …
  - Internet became popular quickly
- Cannot manage this as a single LAN
  - Hard to manage, becomes inefficient
- Need simple way to partition large networks
  - Partition into multiple IP networks that share the same prefix – called a "subnet", part of a network
- CMU case study in RFC
  - Chose not to adopt – concern that it would not be widely supported ☺

IP Address Problem (1991)

- Address space depletion
  - Suppose you need 2^{16} + 1 addresses?
  - Class A too big for all but a few domains
  - Class C too small for many domains but they don't need a class B address
  - Class B address pool allocated at high rate
  - Many allocated address block are sparsely used
- Developed a strategy based on a three solutions
  - Switch to a "classless" addressing model – this lecture
  - Network address translation (NAT) – later in the course
  - Definition of IPv6 with larger IP addresses – next lecture

Today's Addressing: CIDR

- CIDR = Classless Interdomain Routing
- Idea: Flexible division between network and host addresses
  - Not limited to three sizes 8/16/24
  - Prefix can be any size
- Motivation: offer a better tradeoff between size of the forwarding table and efficient use of the IP address space

CIDR (example)

- Suppose a network has fifty computers
  - allocate 6 bits for host addresses (since $2^5 < 50 < 2^6$)
  - remaining 32 - 6 = 26 bits as network prefix
- Flexible boundary means the boundary must be explicitly specified with the network address!
  - Informally, "slash 26" → 128.23.9/26
  - Formally, prefix represented with a 32-bit mask: 255.255.255.192 where all network prefix bits set to "1" and host suffix bits to "0"

Classful vs. Classless addresses

- Example: an organization needs 500 addresses.
  - A single class C address not enough (254 hosts).
  - Instead a class B address is allocated (~65K hosts)
  - That's overkill, a huge waste!
- CIDR allows an arbitrary prefix-suffix boundary
  - Hence, organization allocated a single /23 address (equivalent of 2 class C's)
- Maximum waste: 50%

Hence, IP Addressing: Hierarchical

- CIDR allows more efficient use of the IP address space
  - Helps (at least for a while) with the high demand for IP addresses
- But how does this help with the growth of forwarding tables?
  - Number of destination networks is growing as well!
- Solution has two complementary parts:
  - Allocation of IP addresses is done hierarchically
  - Routers will combine forwarding entries for destinations "in the same general direction"

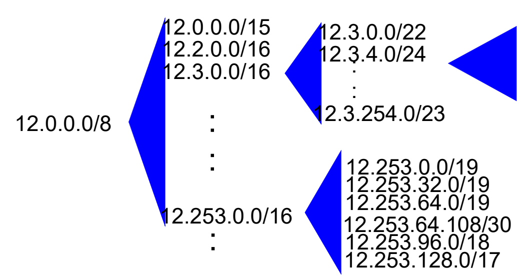
Allocation Done Hierarchically

- Historically assignment of prefixes was "first come first serve"
- With CIDR: Internet Corporation for Assigned Names and Numbers (ICANN) gives large blocks to…
- Regional Internet Registries, such as the American Registry for Internet Names (ARIN), which give blocks to…
- Large institutions (ISPs), which give addresses to …
- Individuals and smaller institutions
- FAKE Example:

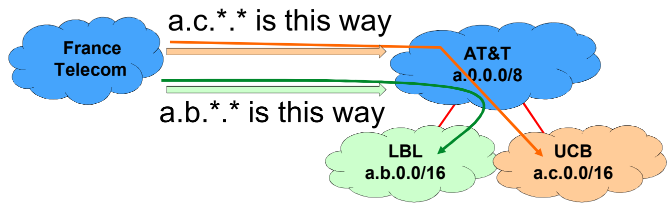
  ICANN ➔ ARIN ➔ AT&T ➔ UCB ➔ EECS

CIDR: Addresses allocated in contiguous prefix chunks

Recursively break down chunks as get closer to host



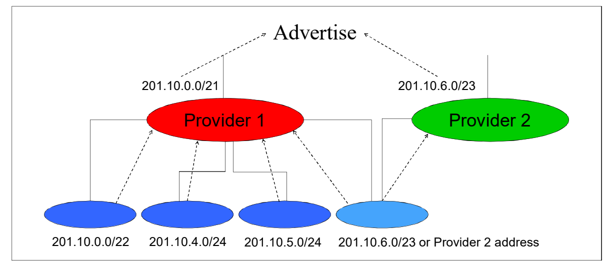
IP Addressing → Scalable Forwarding?



- How many forwarding entries does France Telecom need for LBL/UCB destinations? Two
- How about if all a.0.0.0/8 addresses are served by AT&T - it "own" a.0.0.0/8, right? One
- In practice, it is complicated …

CIDR Implication: Longest Prefix Match

- How to deal with multi-homing, legacy addresses, …



How LPM Works

- Routing protocols aggregate forwarding entries to reduce table size
  - E.g., 3 forwarding entries A/B/C 01010011.xy/10 can be combined into 01010011/8 if they forward through the same port
  - A fourth entry D that uses a different egress port has its own entry
- Works correctly because of longest prefix match (LPM)
  - Packets to A/B/C will match only the 01010011/8
  - Packets to D will match entries but will prefer the short "/10" entry
- Legacy prefixes (e.g., 128.2) also often have their own entry

Filling in Some Router Details

- How do routing protocols learn the prefix size?
  - Routing advertisements include the prefix size; for destination addresses in packets, the prefix size is not relevant
  - For stub networks (subnetting): routers are configured by admin
- But a router now needs ~30 forwarding tables?
  - No – forwarding uses a single tree data structure (called a trie)
  - Very efficient algorithms exist for look up both in HW and SW
- How do routers know the prefix size for destination addresses?
  - They do not need them because of how LPM look up works

Outline

- IP design goals
- Traditional IP addressing
  - Addressing approaches
  - Class-based addressing
  - Subnetting
  - CIDR
- Packet forwarding

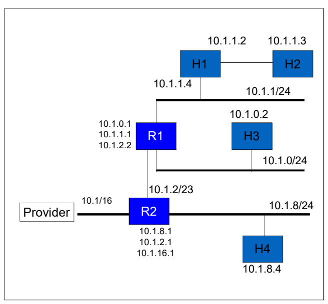
Host Routing Table Example

Destination	Gateway	Genmask	Iface
128.2.209.100	0.0.0.0	255.255.255.255	eth0
128.2.0.0	0.0.0.0	255.255.0.0	eth0
127.0.0.0	0.0.0.0	255.0.0.0	lo
0.0.0.0	128.2.254.36	0.0.0.0	eth0

- From "netstat –rn"
- Host 128.2.209.100 when plugged into CS ethernet
- Dest 128.2.209.100 → routing to same machine
- Dest 128.2.0.0 → other hosts on same ethernet
- Dest 127.0.0.0 → special loopback address
- Dest 0.0.0.0 → default route to rest of Internet
  - Main CS router: gigrouter.net.cs.cmu.edu (128.2.254.36)

Routing to the Network

- Packet to 10.1.1.2 arrives
- Path is R2 – R1 – H1 – H2
- H1 serves as a router for the 10.1.1.2/31 network

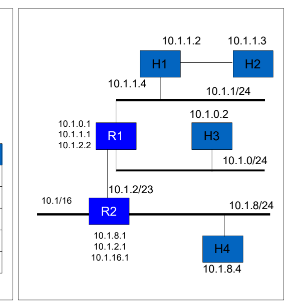


Routing Within the Subnet

- Packet to 10.1.1.3
- Matches 10.1.0.0/23

Routing table at R2

Destination	Next Hop	Interface
127.0.0.1	-	lo0
Default or 0/0	provider	10.1.16.1
10.1.8.0/24	-	10.1.8.1
10.1.2.0/23	-	10.1.2.1
10.1.0.0/23	10.1.2.2	10.1.2.1

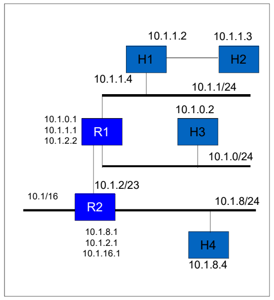


Routing Within the Subnet

- Packet to 10.1.1.3
- Matches 10.1.1.2/31
  - Longest prefix match

Routing table at R1

Destination	Next Hop	Interface
127.0.0.1	-	lo0
Default or 0/0	10.1.2.1	10.1.2.2
10.1.2.0/23	10.1.2.1	10.1.2.2
10.1.0.0/24	-	10.1.0.1
10.1.1.0/24	-	10.1.1.1
10.1.1.2/31	10.1.1.4	10.1.1.1

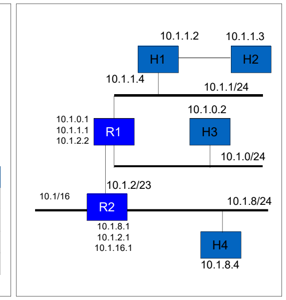


Routing Within the Subnet

- Packet to 10.1.1.3
- Direct route
  - Longest prefix match

Routing table at H1

Destination	Next Hop	Interface
127.0.0.1	-	lo0
Default or 0/0	10.1.1.1	10.1.1.4
10.1.1.0/24	-	10.1.1.1
10.1.1.2/31	-	10.1.1.2



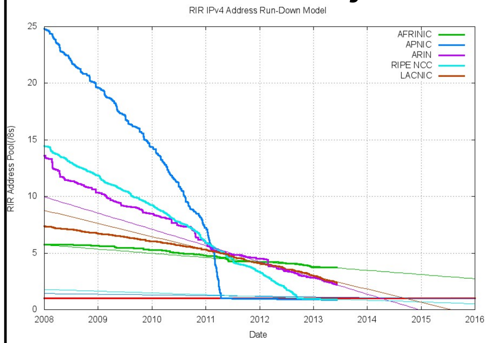
Important Concepts

- Hierarchical addressing critical for scalable system
  - Don’t require everyone to know everyone else
  - Reduces number of updates when something changes
- Classless inter-domain routing supports more efficient use of address space
  - Adds complexity to routing, forwarding, …
    - But it is Scalable!

IP Addresses: How to Get One?

- How does an ISP get block of addresses?
  - From **Regional Internet Registries** (RIRs)
    - ARIN (North America, Southern Africa), APNIC (Asia-Pacific), RIPE (Europe, Northern Africa), LACNIC (South America)
- How about a single host?
  - Assigned by sys admin (static or dynamic)
  - DHCP: Dynamic Host Configuration Protocol: dynamically get address: “plug-and-play”
    - Host broadcasts “DHCP discover” msg
    - DHCP server responds with “DHCP offer” msg
    - Host requests IP address: “DHCP request” msg
    - DHCP server sends address: “DHCP ack” msg

IP Address Availability Remains a Major Challenge



- Some are in big trouble!
- APNIC: Asia
- AFRINIC: Africa
- ARIN: North America
- LACNIC: Latin America
- RIPE NCC: Europe, Middle East, parts of central Asia