Module M2

CPSC 317

September 16, 2022

Learning Goals

A. IPv4

- Know the format of an IPv4 address and what it names.
- Know the difference between classful and classless addressing (CIDR addressing).
- Given a range of addresses determine an appropriate CIDR representation for a network.
- For a given network, assign a host name using slash notation and a mask for the network/host parts.
- Explain how organizations gets IP addresses for its use.
- Define what is meant by a non-routable IP address and give an example.
- Understand how and why the IPv6 header is different from IPv4
- Describe the general information contained in packet headers and the role of this information
- Describe IP protocol header and the purpose of the fields.
- Describe the relationship between IP addresses and routing in the Internet
- Perform longest prefix matching
- Given a router, its routing tables and an incoming packet determine the link the packet will go out on
- Given a collection of routers, their routing tables, and a packet trace the packet through the network
- Understand the concepts of subnetting and super-netting (or aggregation).
- Be able to decompose a network into subnetworks of a specific size.
- Know how to use aggregation to reduce the number of advertised networks.

B. BGP

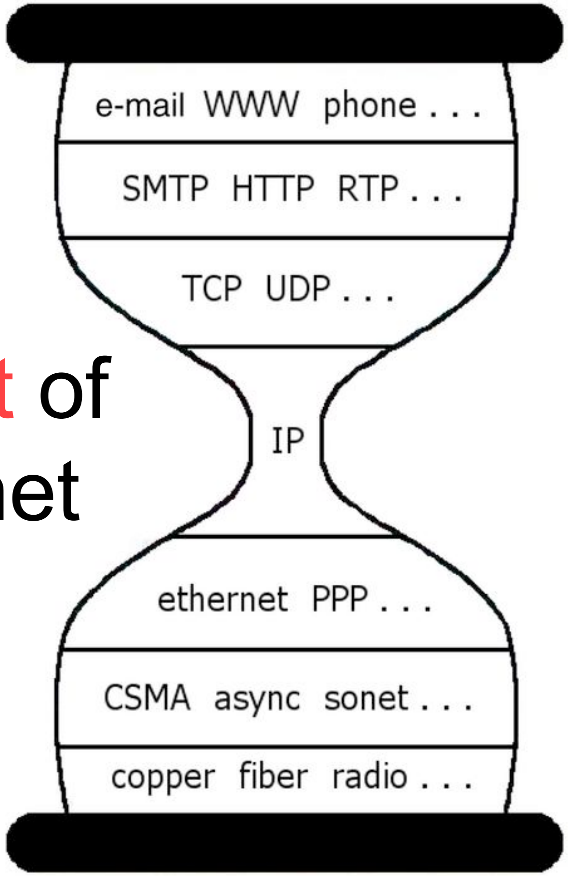
- Define the purpose of an AS
- Explain how routing decisions are made from the perspective of the AS
- List the types of information exchanged by eBGP
- Given multiple routes to a destination enumerate the factors that go into the router's decision to route a particular way.
- Understand the terminology related to IGP, EGP, iBGP, eBGP, BGP connection, peering, transit, border, exchange point. OSPF (just as an example of an IGP).
- Explain, using hot potato routing, how a packet is forwarded from a router in one AS to its destination in another AS.

C. ICMP tools traceroute and ping

- What fields appear in a ICMP header and what is the purpose of the protocol.
- How are ICMP messages sent in the Internet.
- Use traceroute to map-out parts of the Internet
- Define the term congestion and explain its implication with respect to latency and packet loss
- Given a traceroute identify where there might be congestion
- Given a traceroute identify links with large latency
- Given a traceroute describe the journey a packet takes to get from the source to the destination.

IP

TCP/IP Hourglass



The waist of the Internet

IP sends packets (Datagrams) from source to destination…

NAMING

- Purpose is to route messages from source to destination.
- What's in a Name?
  - Format
  - Semantics
  - Properties
- What is an "address"?

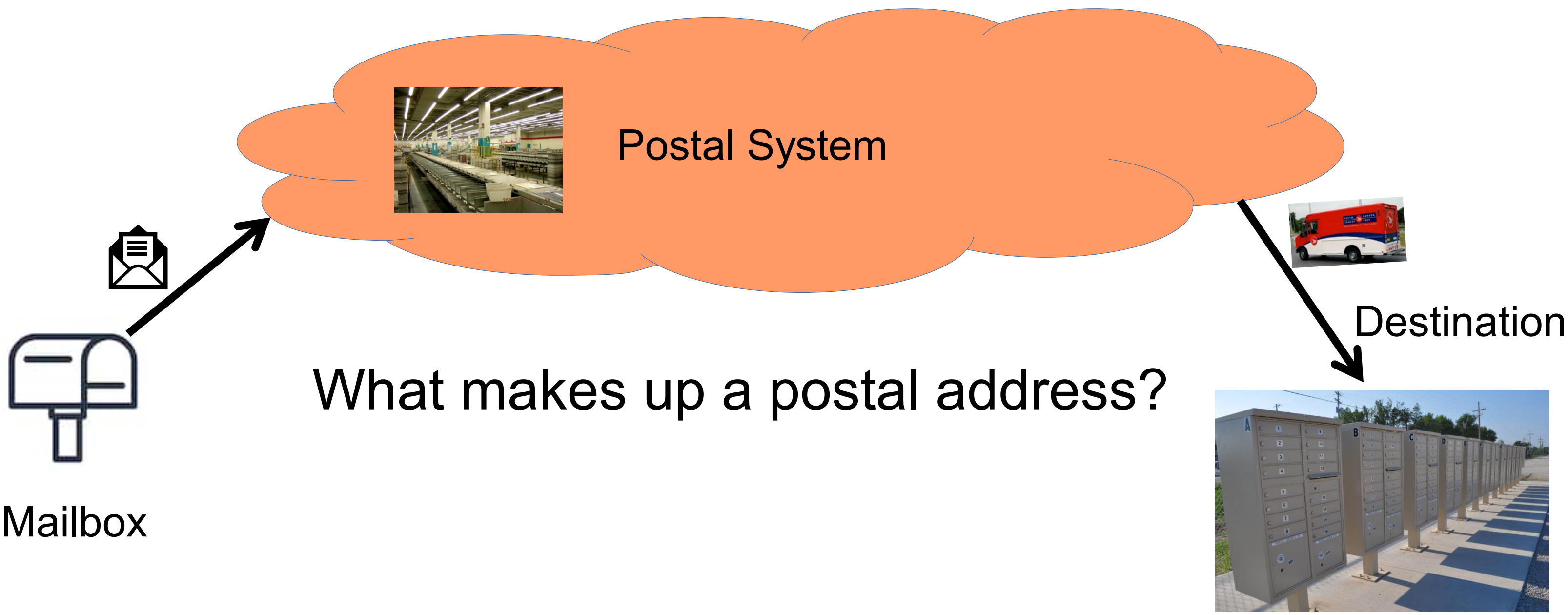
Addresses and Names?

- My Name
  - Telephone number (office, home, cell)
  - Mail address (office, home)
  - Email
  - Piazza

What does it name?

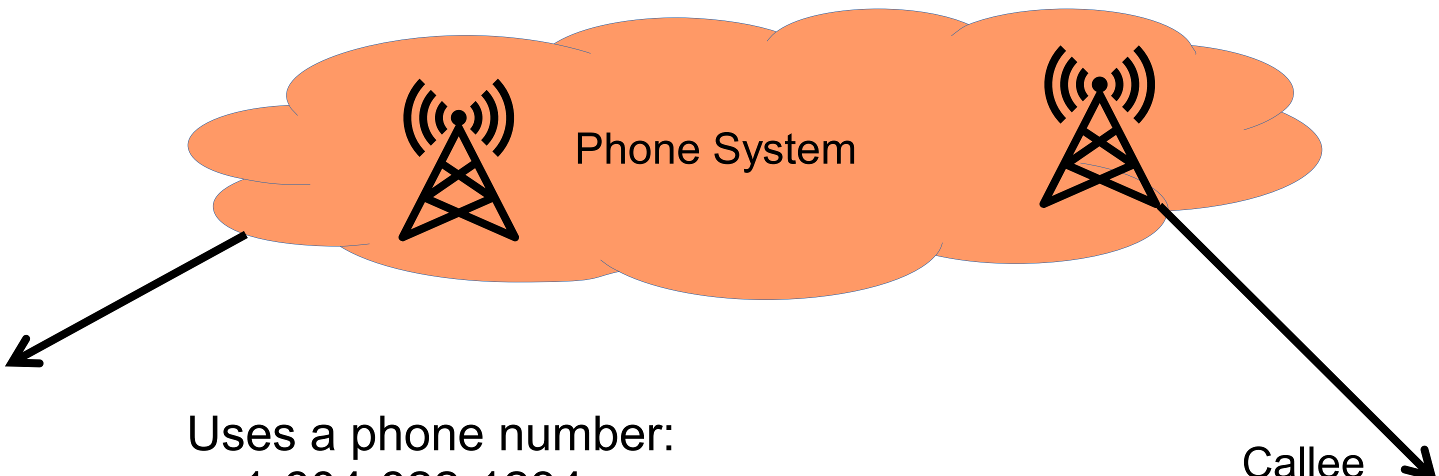
person, thing, location (address)

Mail is delivered from one location to another



What makes up a postal address?

A call is delivered to a phone



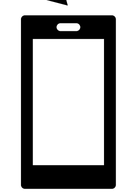
Caller

Uses a phone number:

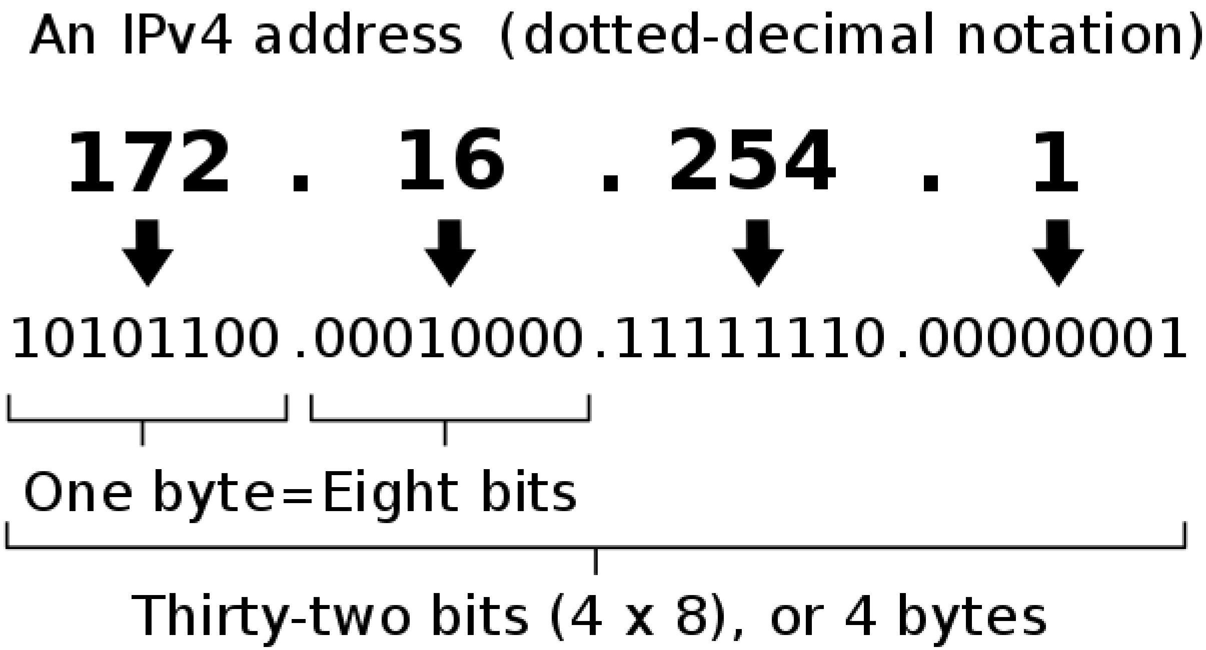
1-604-822-1234

- Region code – 1
- Area code 604 – broad area within the region
- Telephone exchange/switch 822
- The phone line on that switch 1234

Callee

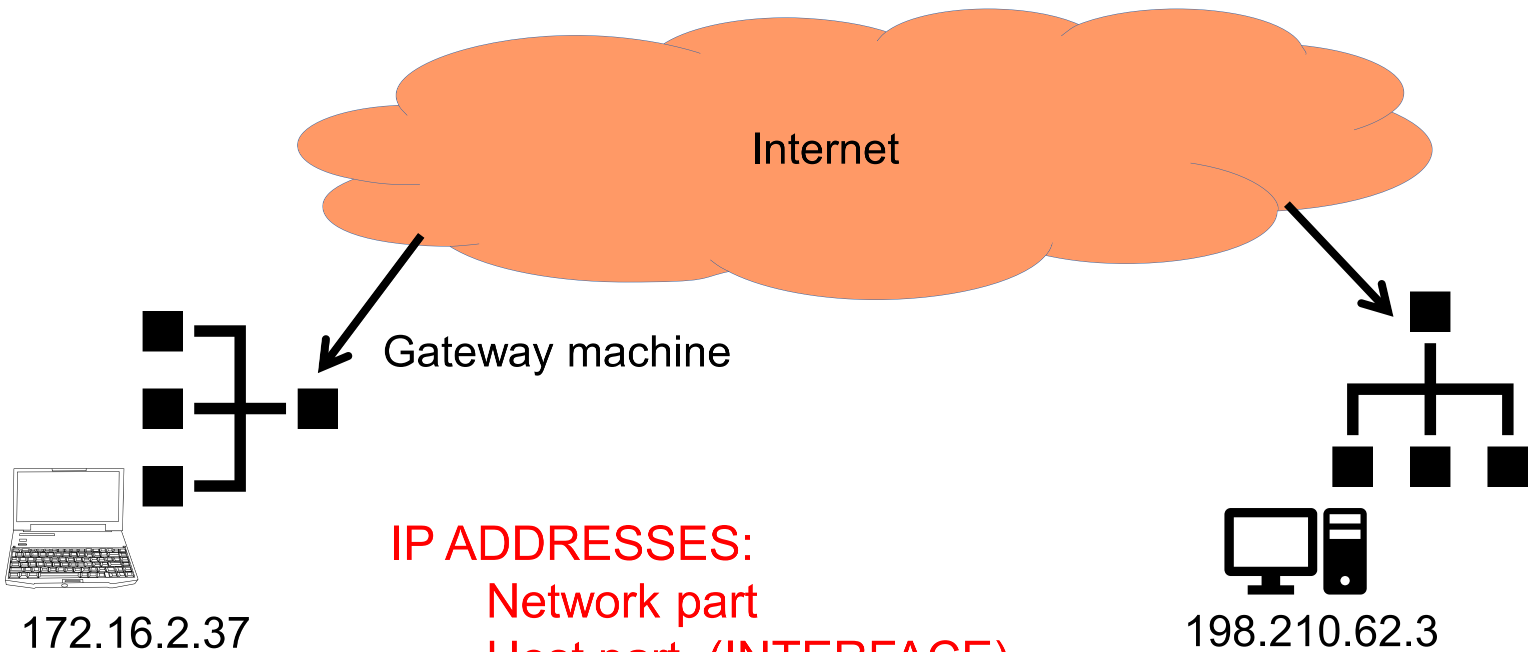


IPv4 Address



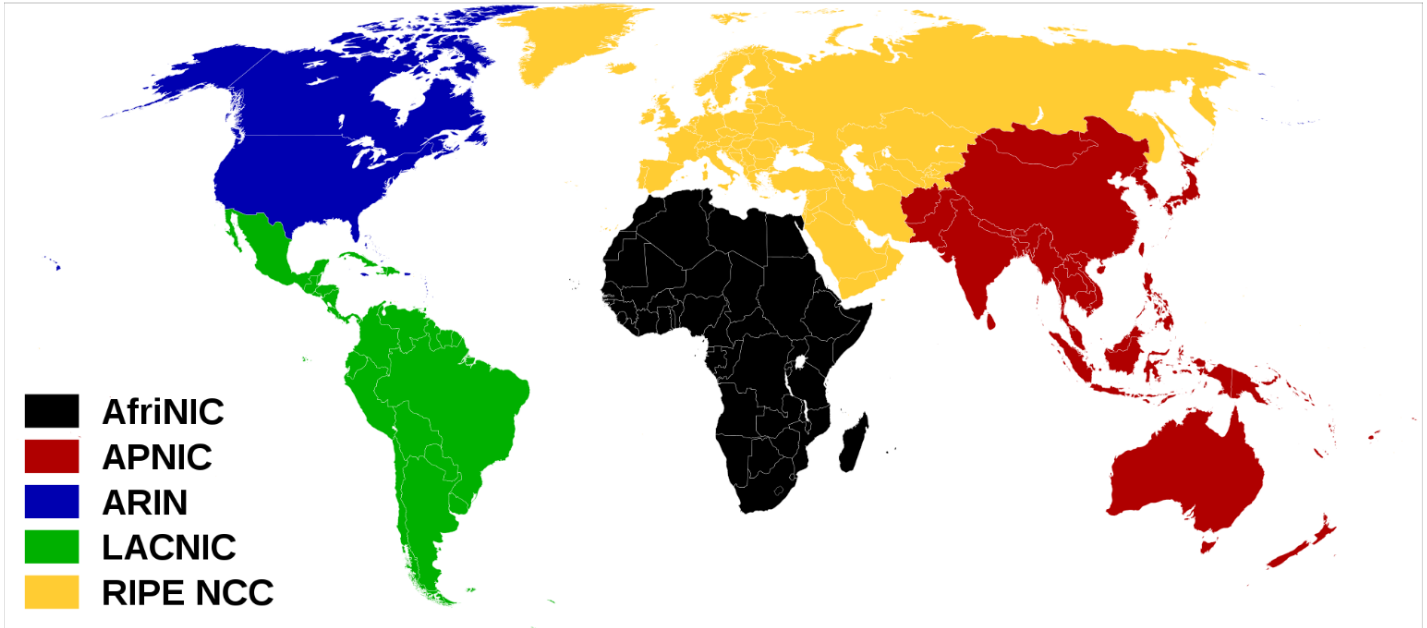
What is an OCTET?

Data is sent from one computer to another



IP ADDRESSES:
- Network part
- Host part (INTERFACE)

Regional Internet Registries (RIR)



From http://commons.wikimedia.org/wiki/File:Regional_Internet_Registries_world_map.svg

IPv4 address exhaustion

- First top-level exhaustion happened on Jan 31, 2011. North American exhaustion happened on Sept. 24th, 2015.
- CIDR addressing (classless)
- IPv6 with 128 bit addresses

https://en.wikipedia.org/wiki/IPv4_address_exhaustion

NAMING NETWORKS

IP Addressing: CIDR

CIDR: Classless InterDomain Routing

- network portion of address of arbitrary length
- address format: a.b.c.d/x, where x is # bits in network portion of address

Example: 200.23.18.4/22

network part — host part

11001000 00010111 00010010 00000100

IP Network Address Prefix

Replace the host bits with zero (don't cares).
Routers only care about the network part.

Example: 200.23.18.4/22

network part | host part

11001000 00010111 00010000 00000000

200.23.16.0/22

or 200.23.16/22

IP Network Addressing

Example: 200.23.18.4/22

network part ← → host part

11001000 00010111 00010010 00000100

- Given the CIDR address of the host. Determine its network address (i.e. prefix)

- Apply mask to address (bitwise AND operation)

Configuring an interface

Need to specify both the network part and the host part:

host IP: 200.23.17.129

network mask: 255.255.254.0 – slash part

network part   host part

11001000 00010111 00010000 00000000

200.23.16.0/23

Network Size

network part | host part

11001000 00010111 00010000 00000000

200.23.16.0/23

How many hosts can exist on this network?:

host number (smallest) 200.23.16.0

host number (largest) 200.23.17.255

But avoid all zeros, all ones: 200.23.16.1 to 200.23.17.254 or $2^9 = 512-2 = 510$

Note: All ones and all zeros avoided (all zeros is the IP machines use when they haven't been assigned an address, and multicast uses all ones, confusing to not follow convention). Routers usually assigned as 1 or max-1 in the domain.

ACTIVITY 1

Try the following network address

172.16.129.72

Original classfull addressing

Prefix	Network Size	Name
0	8	Class A
10	16	Class B
110	24	Class C
1110	32	Class D
11110	32	reserved

network	host
192.168.10	**52**

IPv4 Reserved Addresses

List of Reserved IPv4 Address ranges

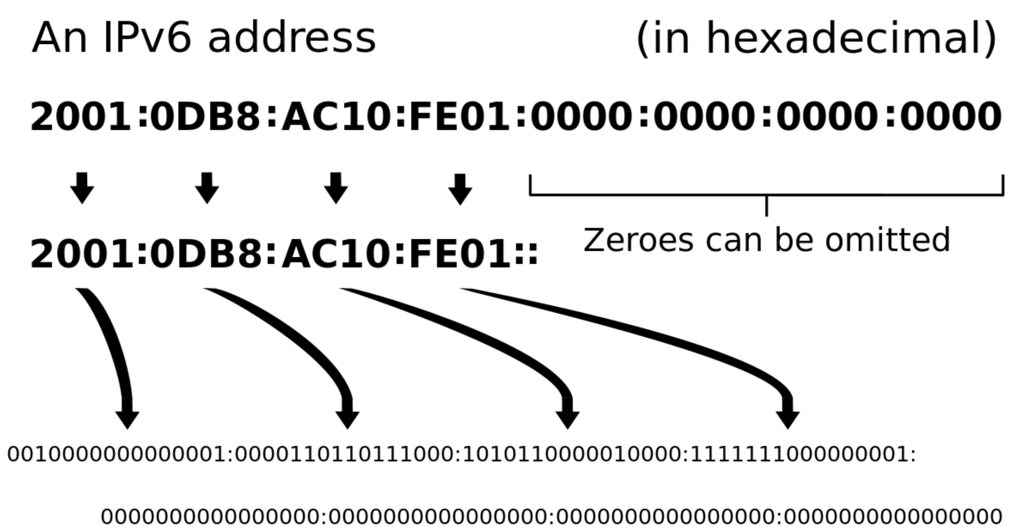
Address Range	RFC	Suitable for Internal Network
0.0.0.0/8	RFC1122	no ("any" address)
10.0.0.0/8	RFC1918	yes
100.64.0.0/10	RFC6598	yes (with caution: If you are a "carrier")
127.0.0.0/8	RFC1122	no (localhost)
169.254.0.0/16	RFC3927	yes (with caution: zero configuration)
172.16.0.0/12	RFC1918	yes
192.0.0.0/24	RFC5736	no (not used now, may be used later)
192.0.2.0/24	RFC5737	yes (with caution: for use in examples)
192.88.99.0/24	RFC3068	no (6-to-4 anycast)
192.168.0.0/16	RFC1918	yes
198.18.0.0/15	RFC2544	yes (with caution: for use in benchmark tests)
198.51.100.0/24	RFC5737	yes (with caution: test-net used in examples)
203.0.113.0/24	RFC5737	yes (with caution: test-net used in examples)
224.0.0.0/4	RFC3171	no (Multicast)

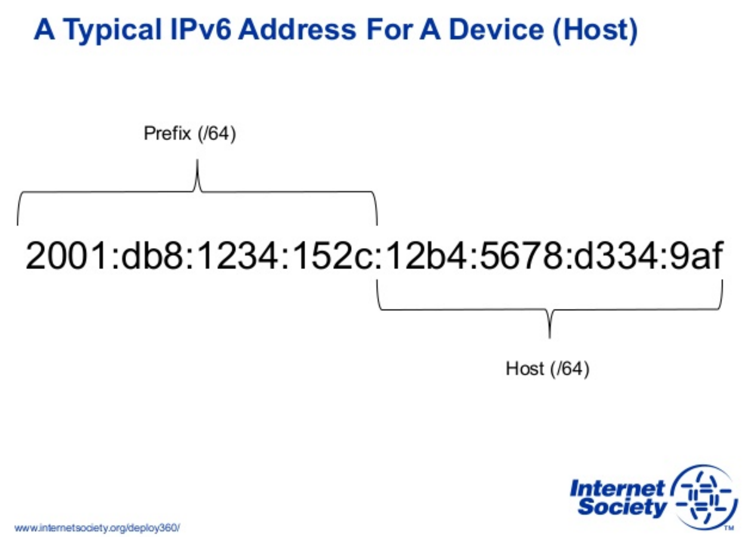
Reserved Addresses

Address Block	What it represents	Where is it used
127.0.0.1/8	Loopback address (the host)	Same as localhost for sending messages to self
192.168.0.0/16 10.0.0.0/8 172.16.0.0/12	Private networks, non-global	Used to implement your own private network
All ones	Broadcast address on local network	Used on a local area network
All zeros	Represents “this network”	Can only be a source address
244.0.0.0/4	IP multicast	Used for multicast (not widely used)

IPv6 format

- 128 bits, 32 Hex-digits (8 hextets) separated by colons
- Zero suppression
- Separated into network+subnet (48+16), host (64)



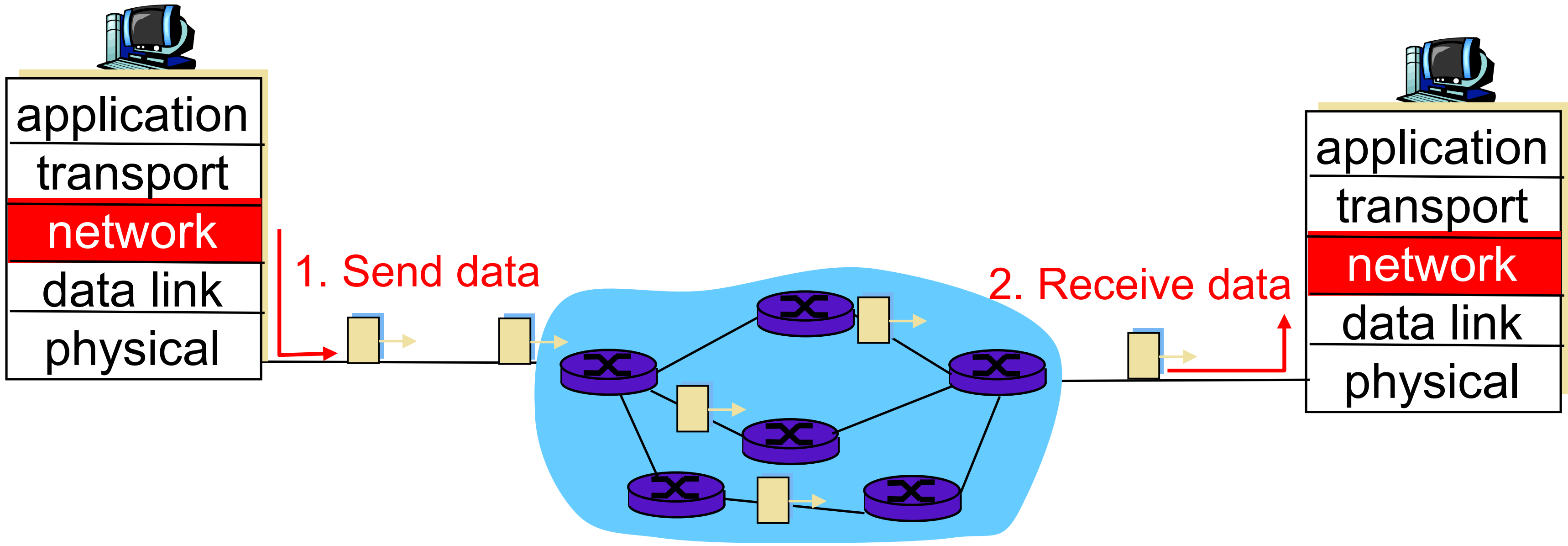


Zero compression and suppressing leading zeros

IP HEADER

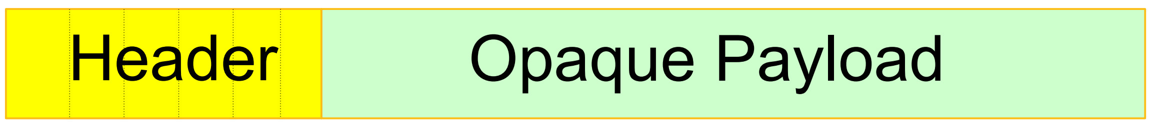
Packet (or Datagram) Network

- no call setup at network layer
- routers: no state about end-to-end connections
  - no network-level concept of "connection"
- packets forwarded using destination host address
  - packets between same source-destination pair may take different paths



Protocol Design

- Syntax: format of packet
  - Nontrivial part: packet “header”
  - Rest is opaque payload *(why opaque?)*



- Semantics: meaning of header fields
  - Required processing
  - Like an interface, what function are you trying to perform

What do we have to do?

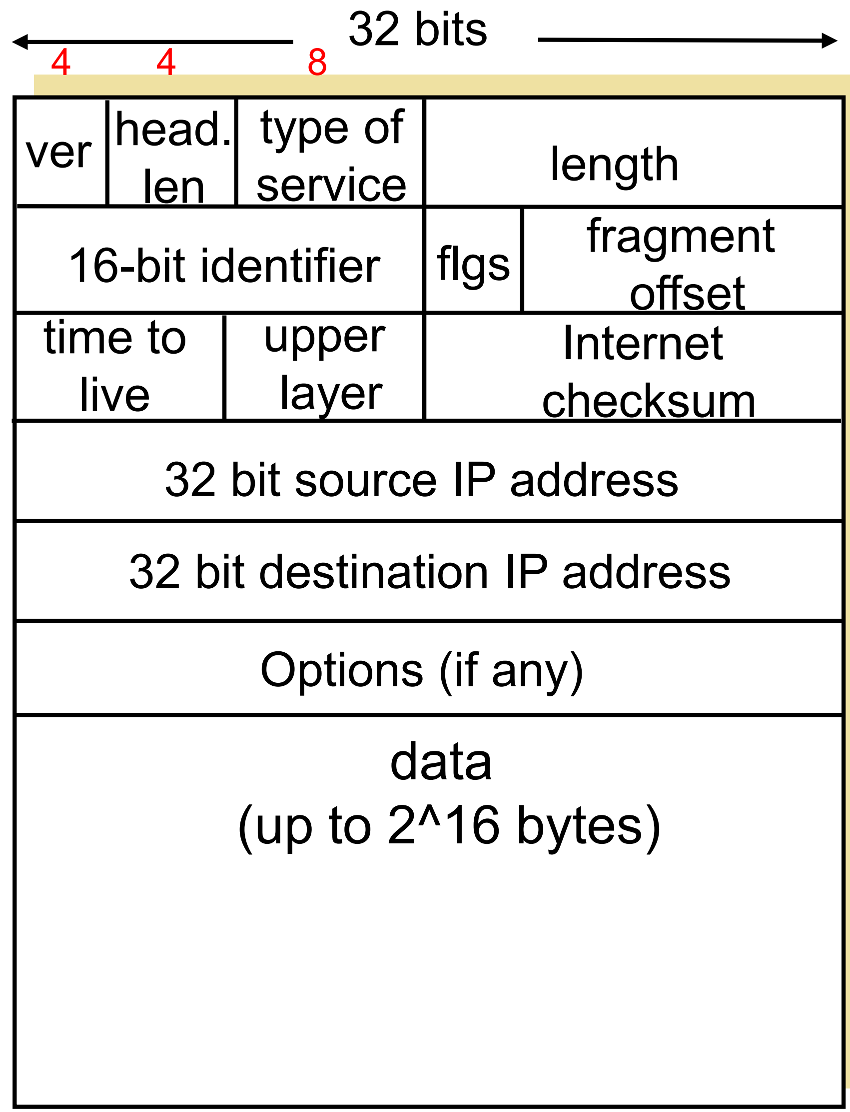
- Read packet correctly
- Get packet to the destination
- Get responses to the packet back to source
- Carry data
- Tell host what to do with packet once arrived
- Specify any special network handling of the packet
- Deal with problems that arise along the path

IPv4 Datagram Header

Header Length Field: number of 32-bit **words**.

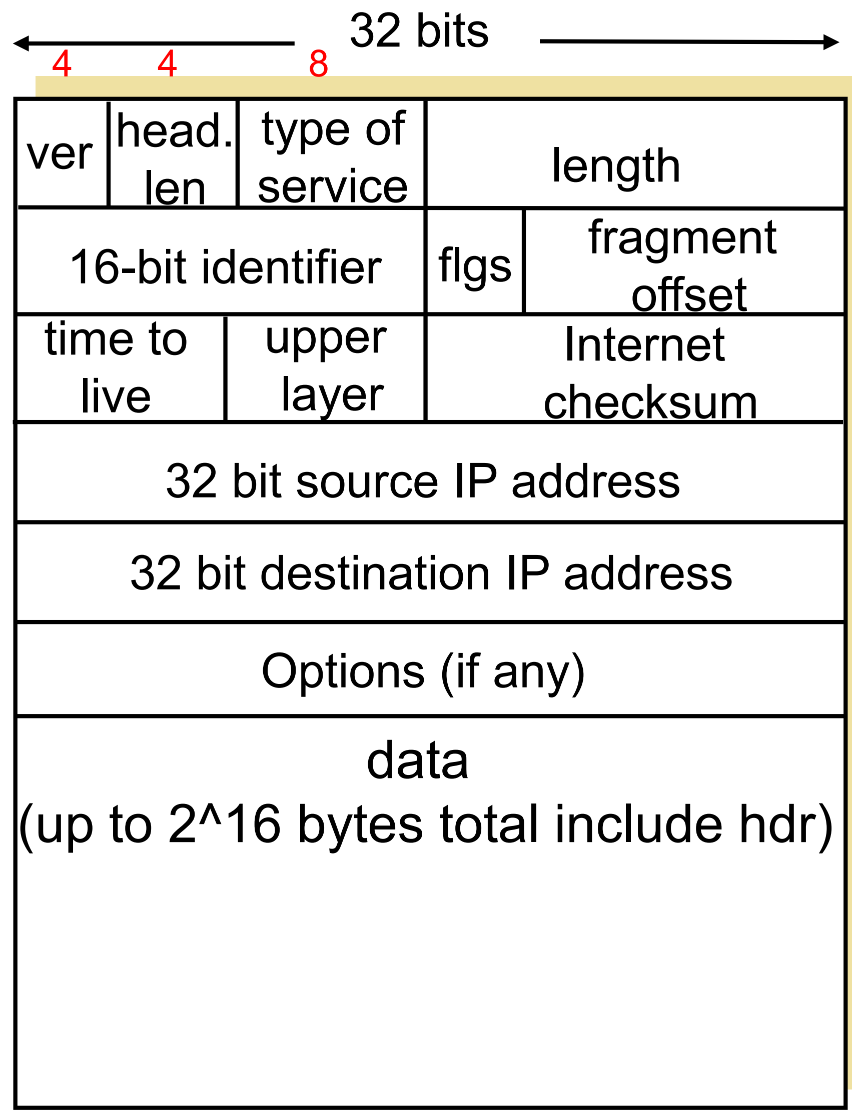
In computing, a **word** is the natural unit of data used by a particular processor design.

Words: 32 bits
Byte: 8 bits (or octet)
Nibble: 4 bits



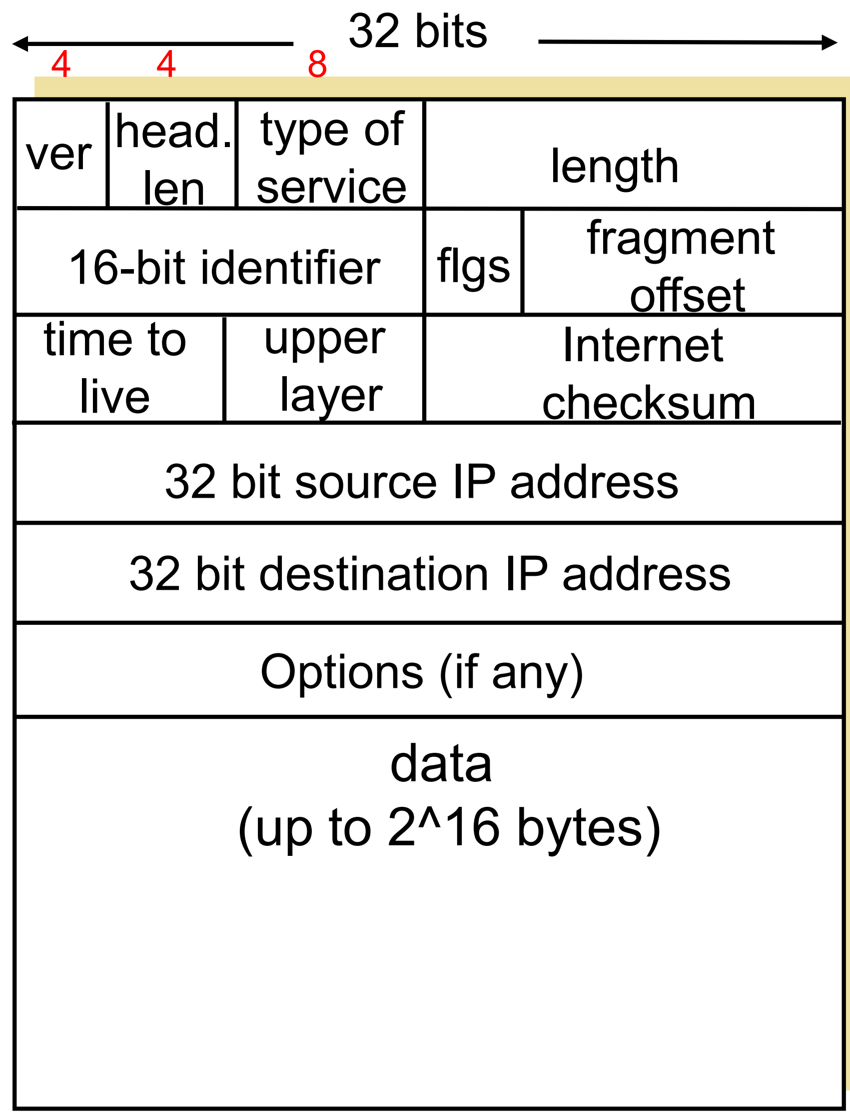
Reading datagram correctly?

- Where does header end?
- Where does packet end?
- What version of IP?
  - *Why is this so important?*



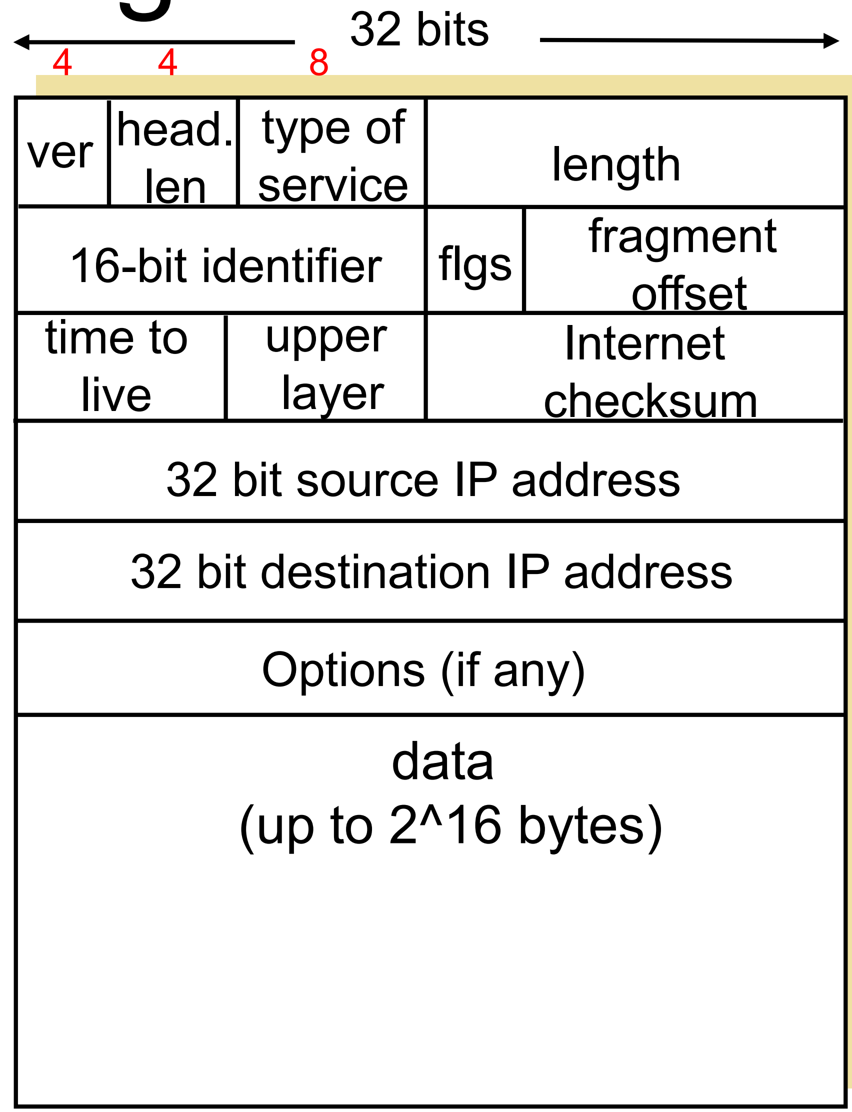
Getting to the destination?

- Provide destination address (duh!)
- Should this be location (addr) or identifier (name)?
  - And what's the difference?
- If a host moves, should its address change?
  - If not, how can you build scalable Internet?



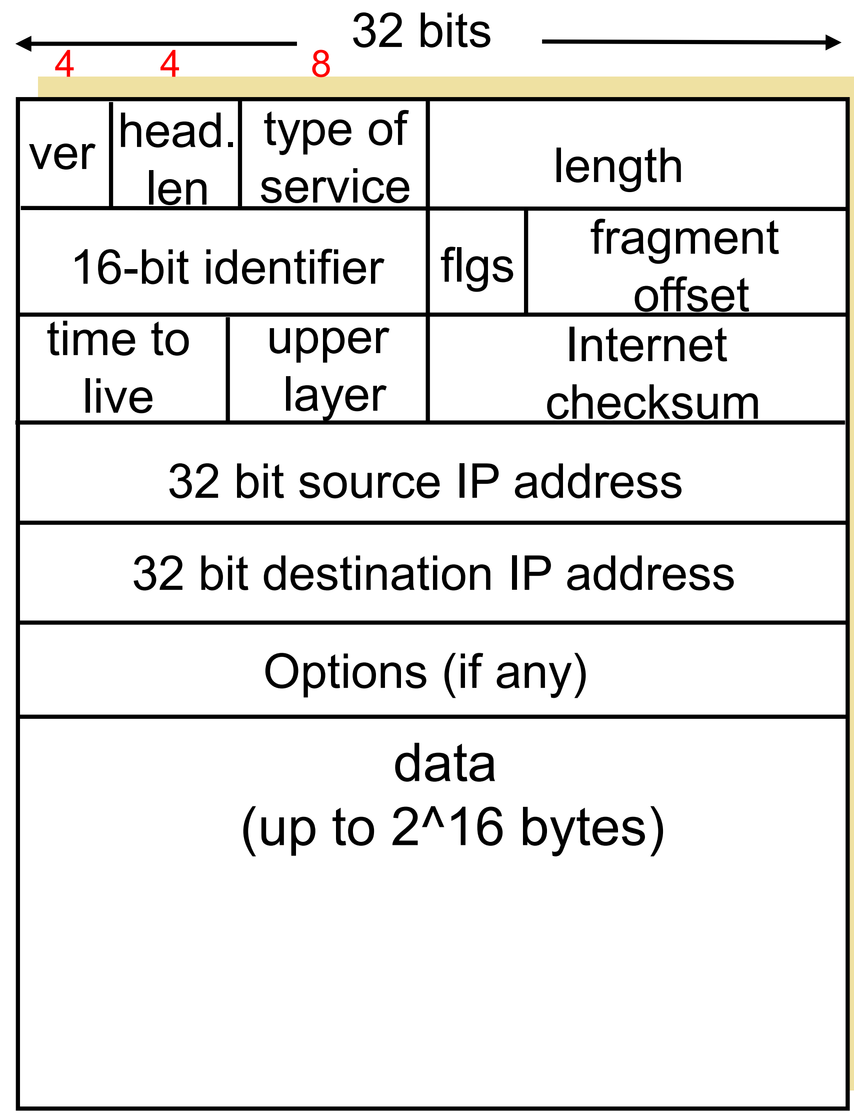
Do we have to get back?

- Provide Source address (duh!)
- Other ways for destination to get back to source
  - Source address can be in packet payload
  - So you could eliminate source address for this reason
- But source address is necessary for routers to respond to source with errors



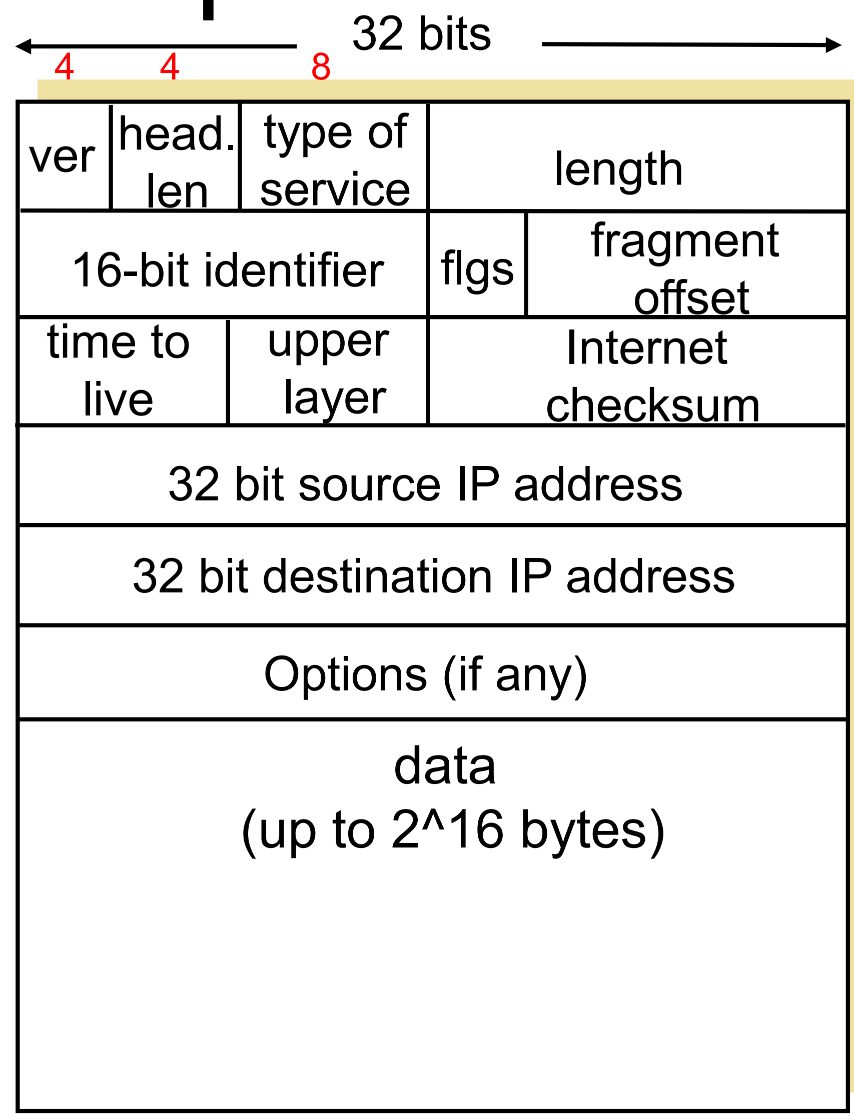
What to do with data at DST

- Indicate which protocols should handle packet
- What layer should this protocol be in?
- What are some options for this today?
- How does the source know what to enter here?



How to deal with problems

- Is packet caught in loop?
  - TTL
- Header Corrupted:
  - Detect with Checksum
  - What about payload checksum?
- Packet too large?
  - Deal with fragmentation
  - Split packet apart
  - Keep track of how to put together



IPv4 Datagram Header

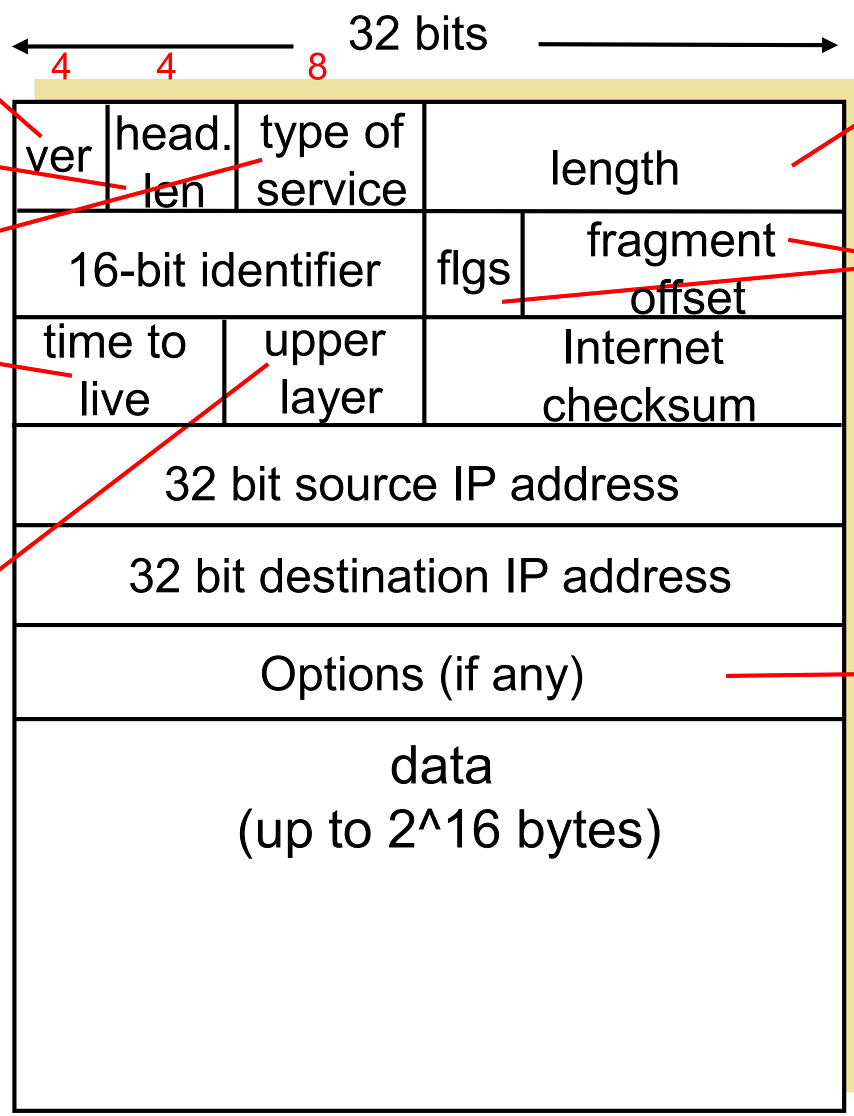
IP protocol version number

header length (words)

"type" of data

max number remaining hops (decremented at each router)

upper layer protocol to deliver payload to



total datagram length (bytes)

for fragmentation/ reassembly

E.g. timestamp, record route taken, specify list of routers to visit.

- 20 bytes of IP header (5 words)
- At most 15 words

IP Fragmentation

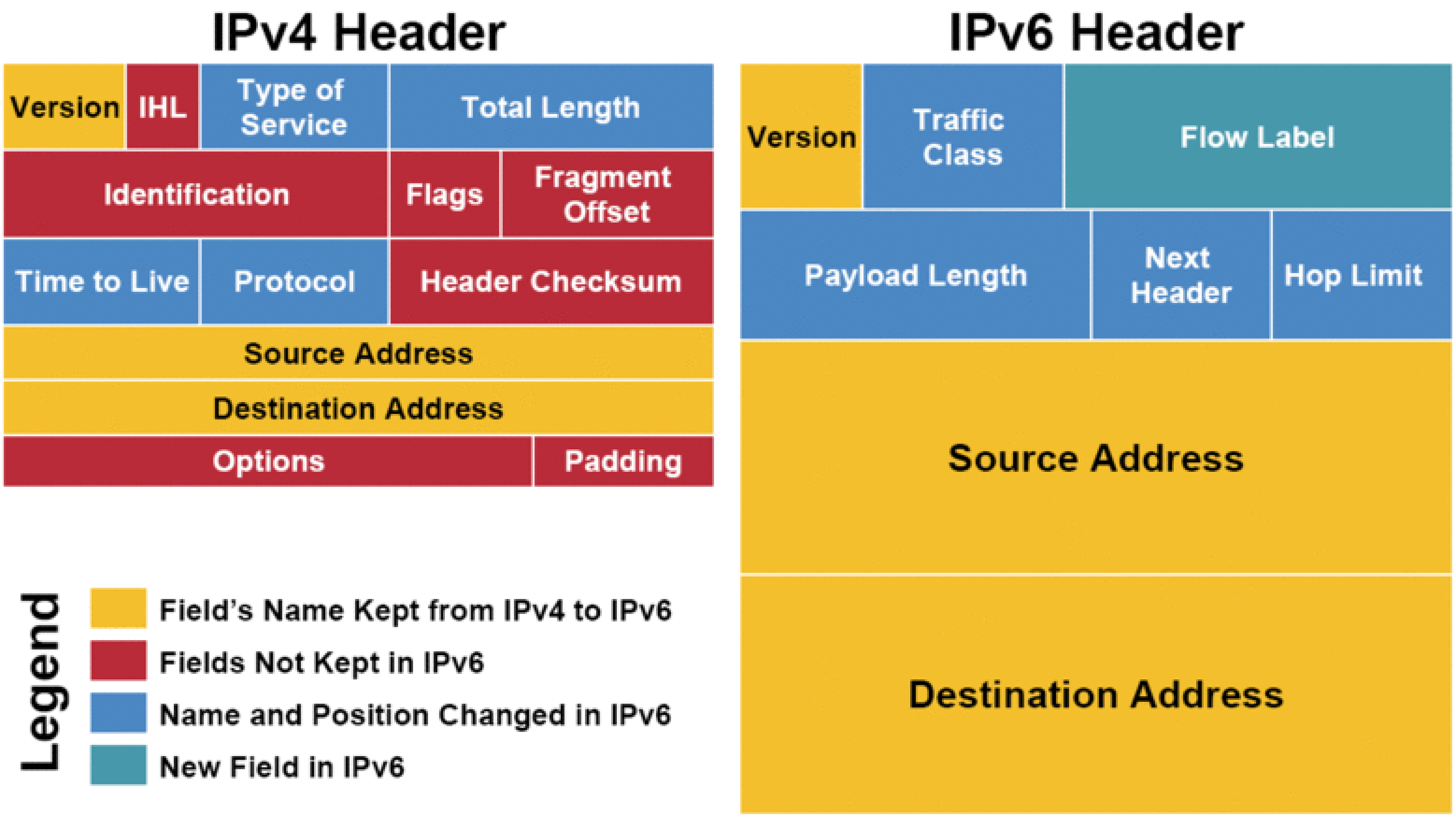
- MTU (Maximum Transmission Unit)
  - Size of a packet that can be transmitted
  - Example, Ethernet frame is 1500 bytes.
- Every packet leaving the host is given an ID, usually just incremental (all fragmented packets keep the same ID).
- For re-assembly give the offset (8 byte blocks)
- Reassembled at the host

IP Header Checksum

- Set the checkpoint field to all zeros.
- Divide the header into 16 bit chunks.
- Add up the chunks using 1's complement arithmetic (end-around carry)
- Complement (1->0,0->1), put in checkpoint field.

RFC 1071

IPv4 Datagram Header



Words, Bytes and Bits

IPv4 Header

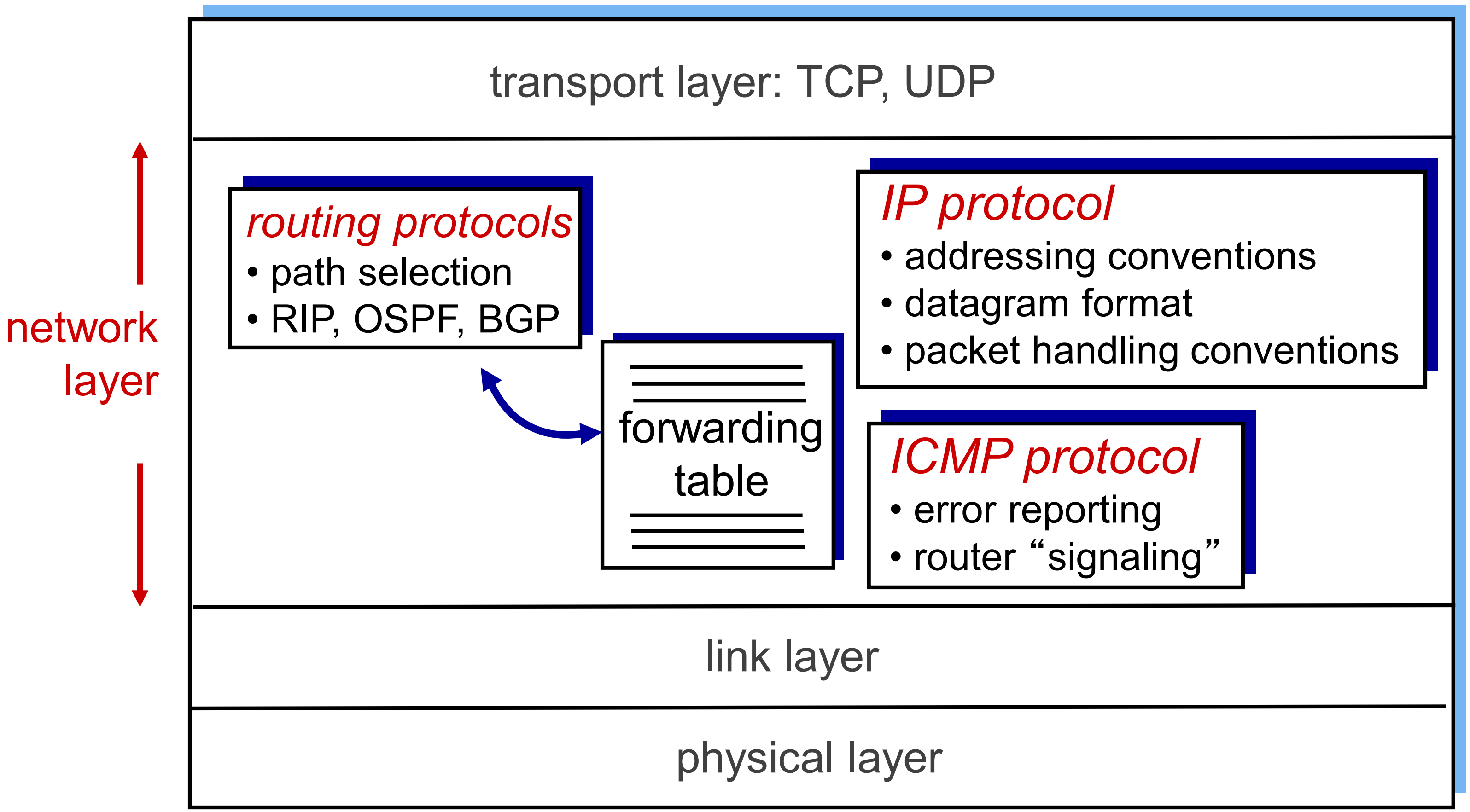
<table>
<tr><th>Byte
Row</th><th colspan="2">0</th><th>1</th><th colspan="2">2</th><th>3</th></tr>
<tr><td>0</td><td>Version</td><td>IHL</td><td>Type of Service</td><td colspan="3">Total Length</td></tr>
<tr><td>1</td><td colspan="3">Identification</td><td>Flags</td><td colspan="2">Fragment Offset</td></tr>
<tr><td>2</td><td colspan="2">Time to Live</td><td>Protocol</td><td colspan="3">Header Checksum</td></tr>
<tr><td>3</td><td colspan="6">Source Address</td></tr>
<tr><td>4</td><td colspan="6">Destination Address</td></tr>
<tr><td>5</td><td colspan="5">Options</td><td>Padding</td></tr>
</table>

In computing, a **word** is the natural unit of data used by a particular processor design.

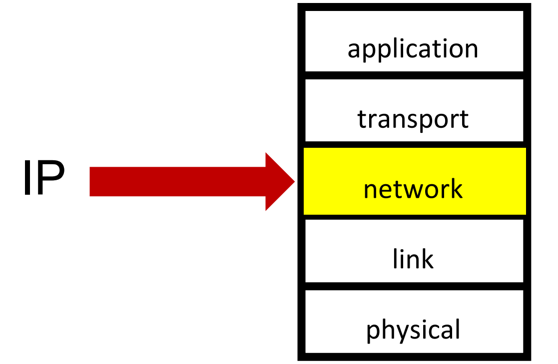
Header Length Field: number of 32-bit **words**.

Words: 32 bits
Byte: 8 bits
Nibble: 4 bits

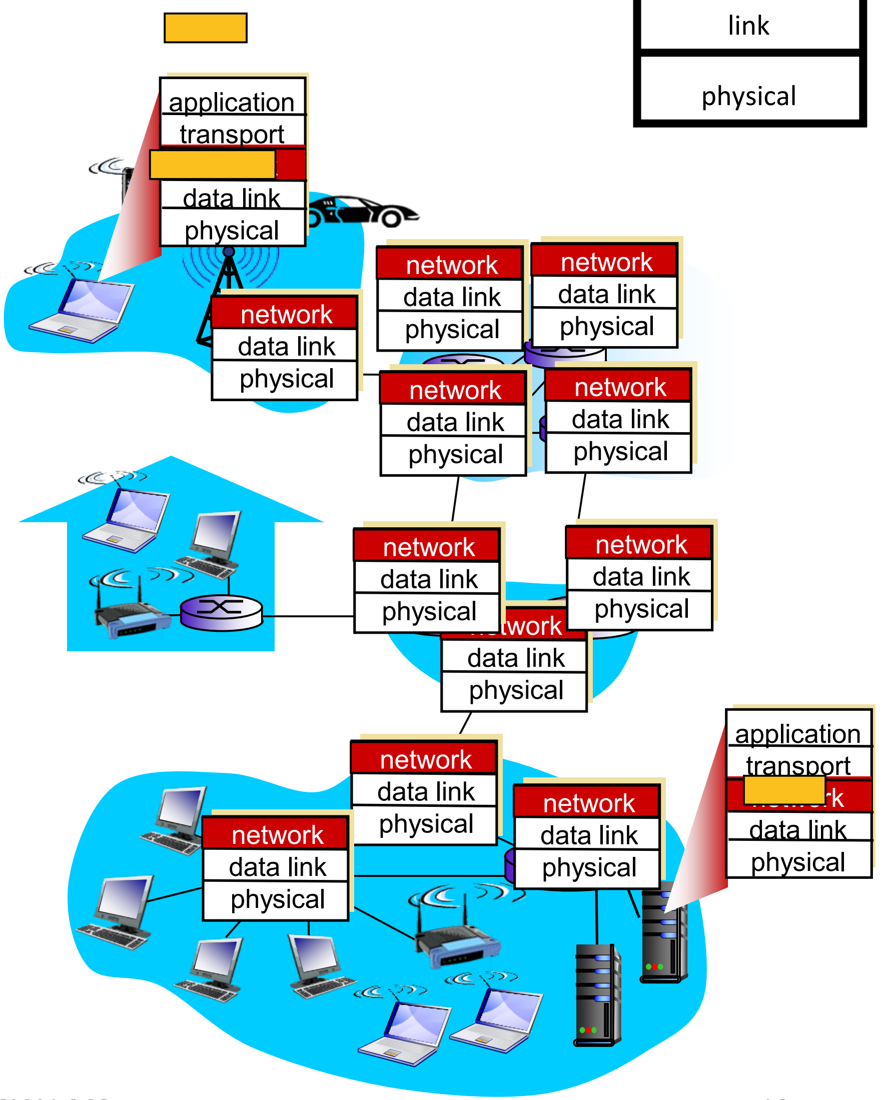
Network Layer



Network layer



- Transport segment from sending to receiving host
- On sending side encapsulates segments into datagrams
- On receiving side, delivers segments to transport layer
- Network layer protocols in *every* host, router
- Router examines header fields in all IP datagrams passing through it



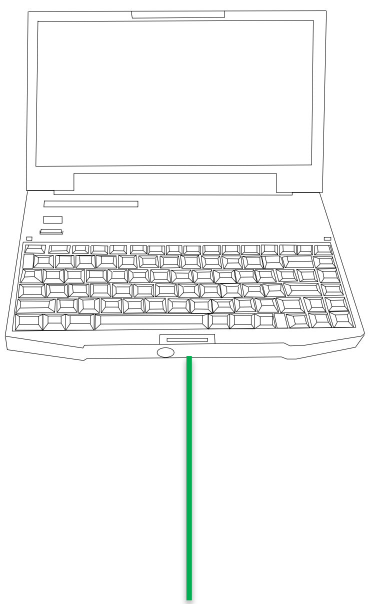
ROUTING

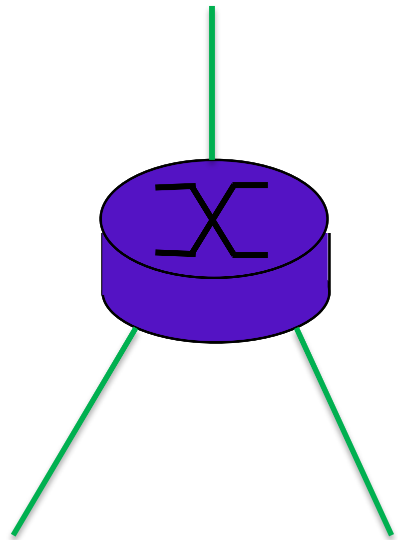
What does a router do?

- Needs a “routing table” (Routing Information Base – RIB).
- Based on the routing table, it needs to look-up and forward packets.
- A routing protocol to maintain the routing table; creating and updating the table based on network conditions.

IDENTIFYING NETWORKS

Adapters NOT Hosts

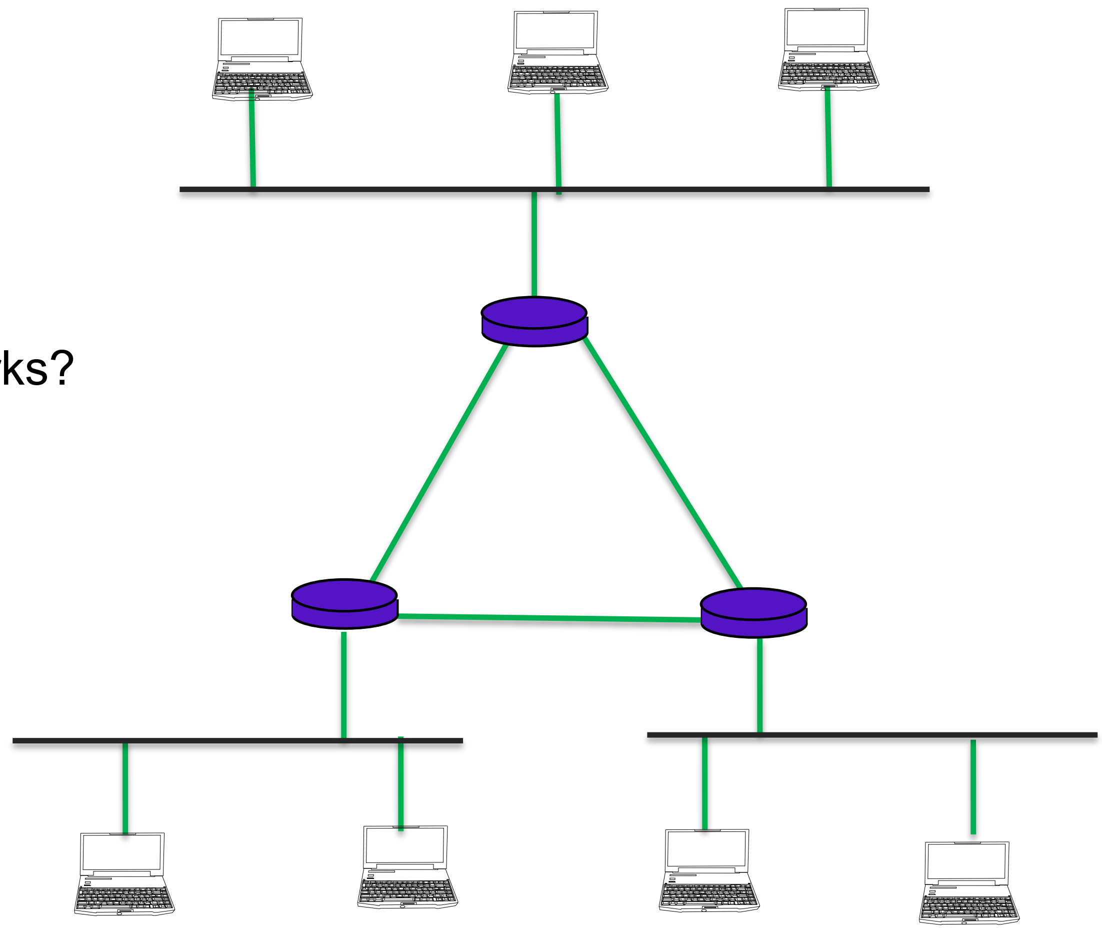




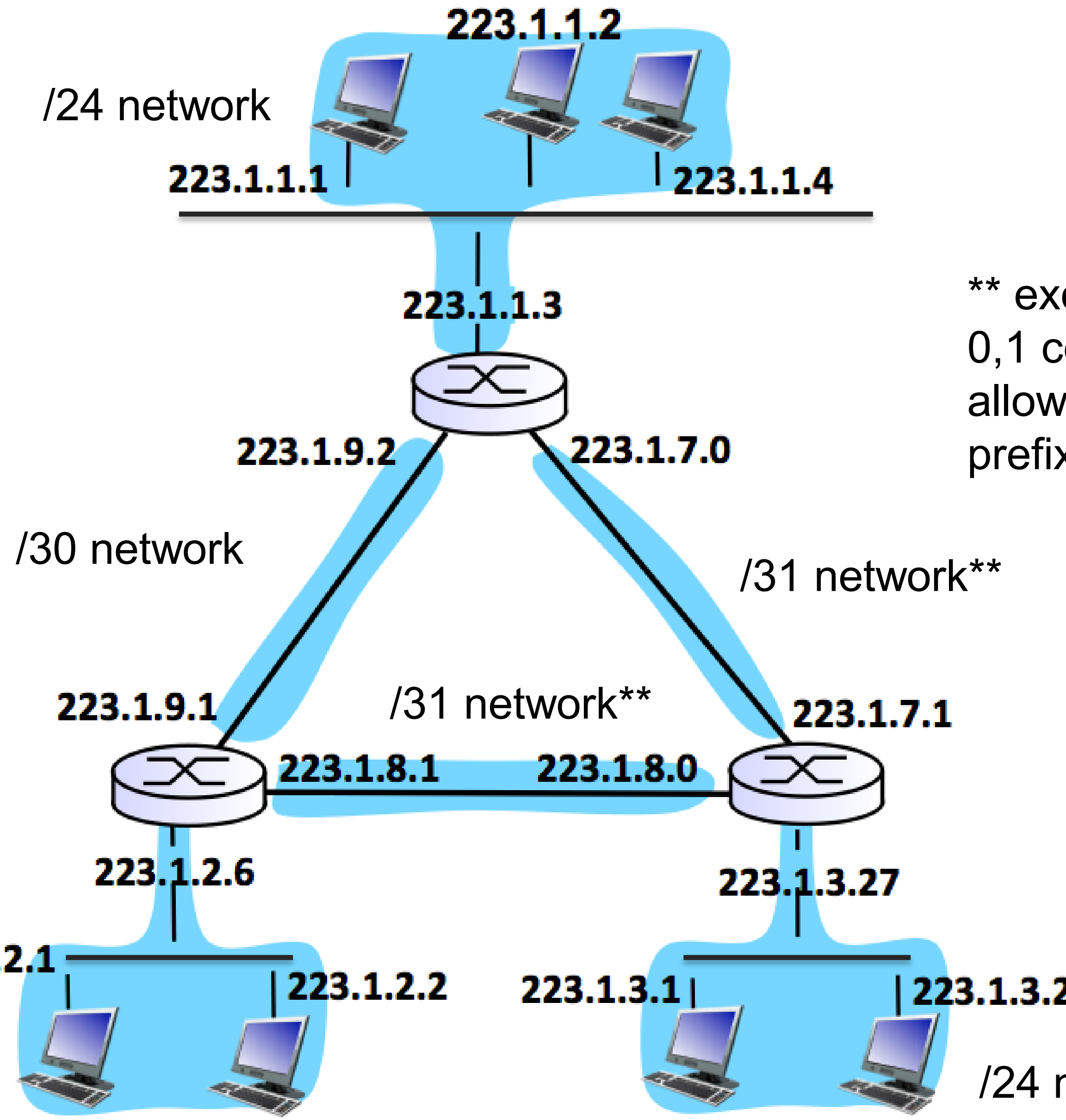
Router ports

Switch ports – same network

Example



How many networks?

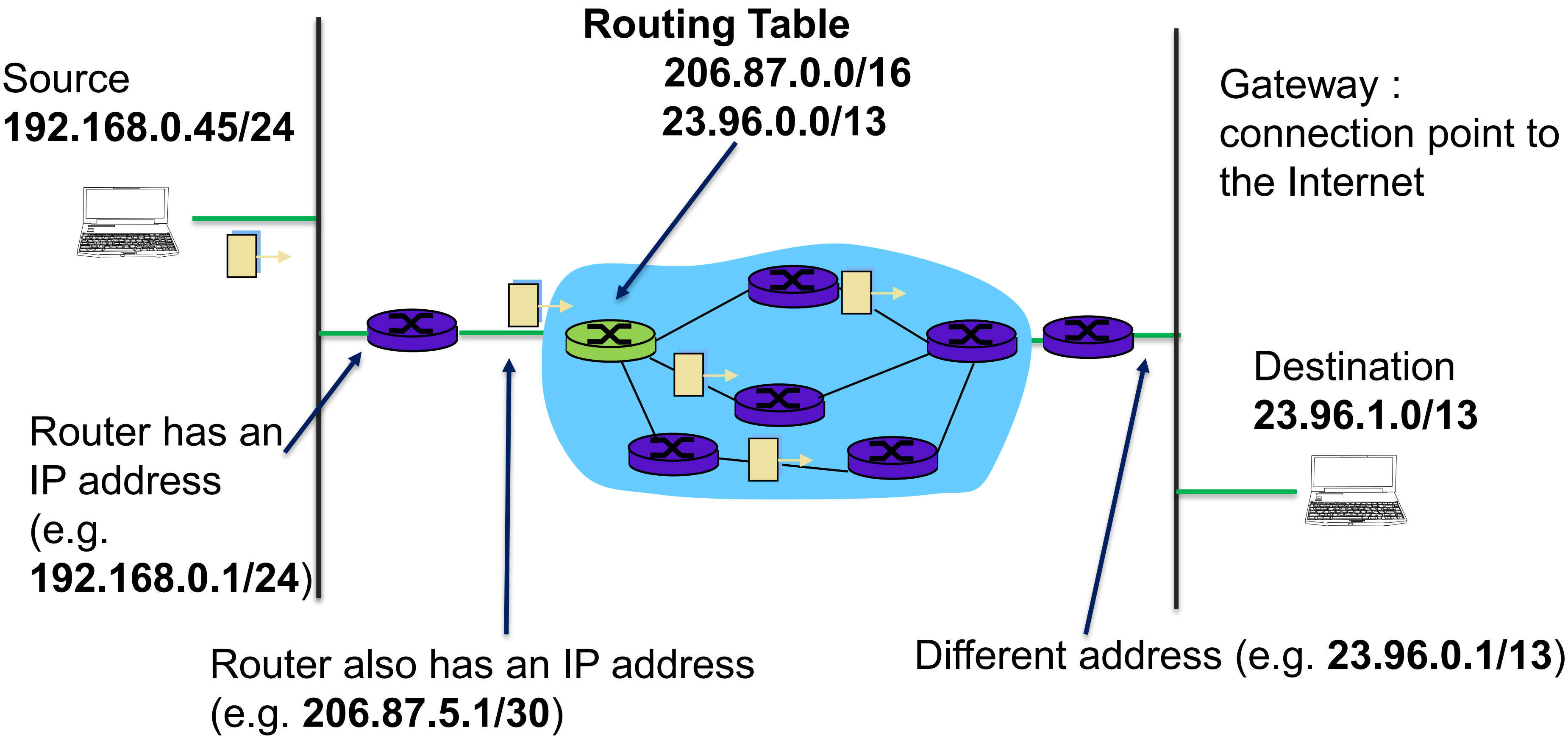


** exception to 0,1 convention allowing /31 prefix RFC 3032

Network names?

LOOK-UP and FORWARDING

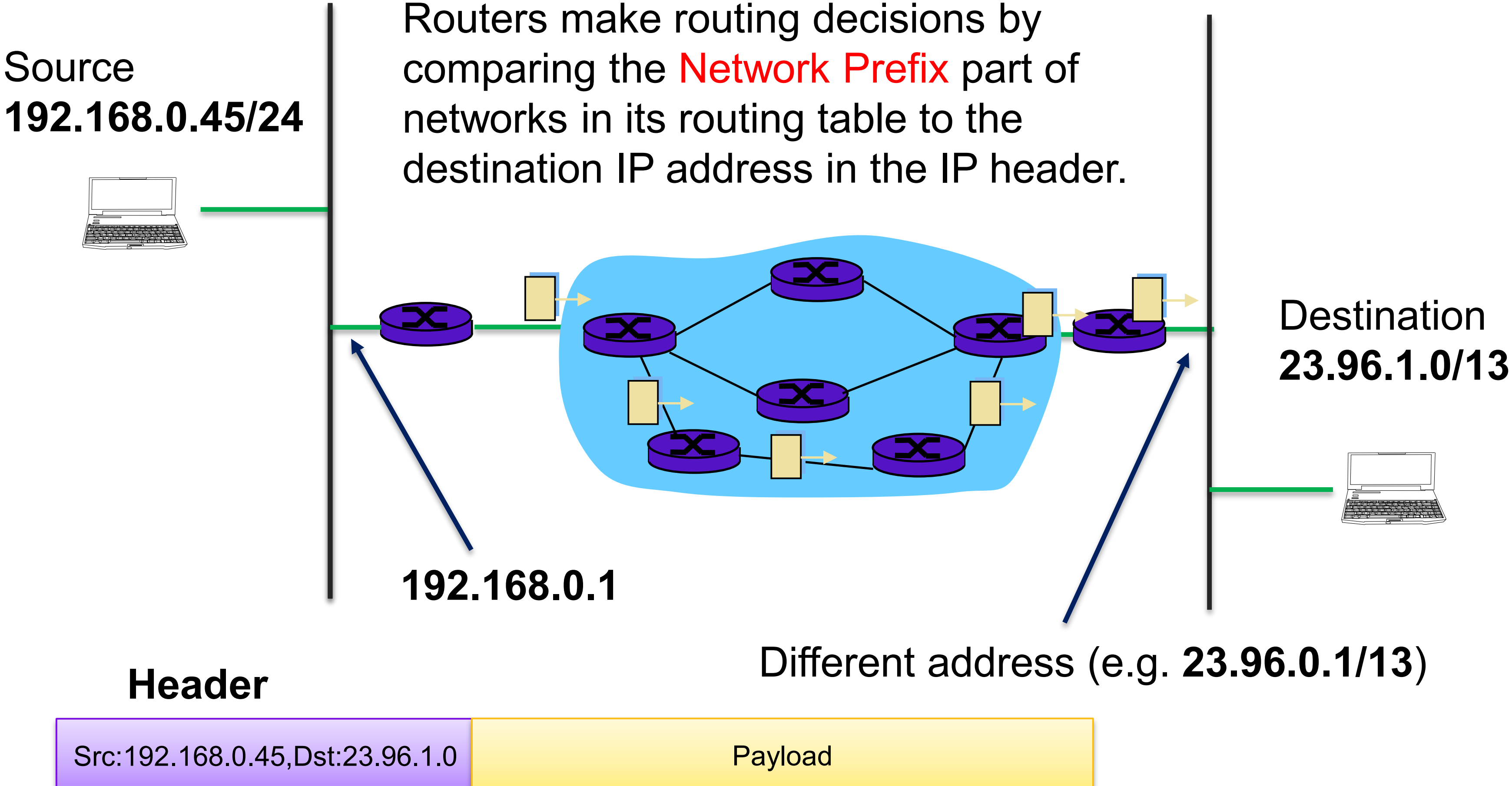
Network Routing – One hop



Routers make routing decisions by comparing the network part of each entry in its routing table with the destination IP address in the datagram. (e.g. at the destination network the first 13 bits of 23.96.0.0/13 match the first 13 bits of 23.96.1.0.

Hopping across the Network

Routers make routing decisions by comparing the Network Prefix part of networks in its routing table to the destination IP address in the IP header.



ROUTING

Two key network-layer functions

network-layer functions:

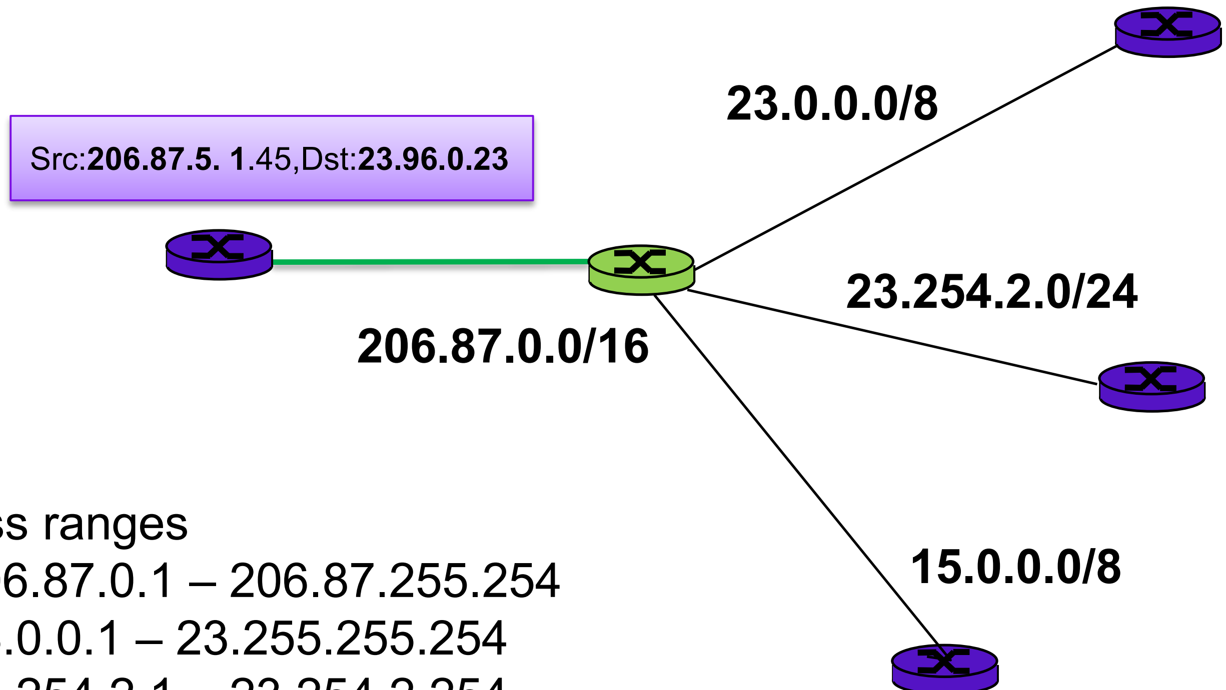
- *forwarding:* move packets from router's input to appropriate router output
- *routing:* determine route taken by packets from source to destination
  - *routing algorithms*

analogy: taking a trip

- *forwarding:* process of getting through single interchange
- *routing:* process of planning trip from source to destination

Each router forwards the datagram to the next hop

DST IP address -- 23.96.0.23

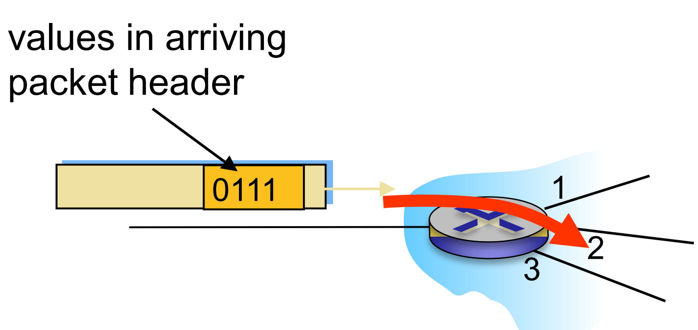


Address ranges
- 206.87.0.1 – 206.87.255.254
- 23.0.0.1 – 23.255.255.254
- 23.254.2.1 – 23.254.2.254
- 15.0.0.1 – 15.255.255.254

Network layer: data plane, control plane

Data plane

- local, per-router function
- determines how datagram arriving on router input port is forwarded to router output port
- forwarding function

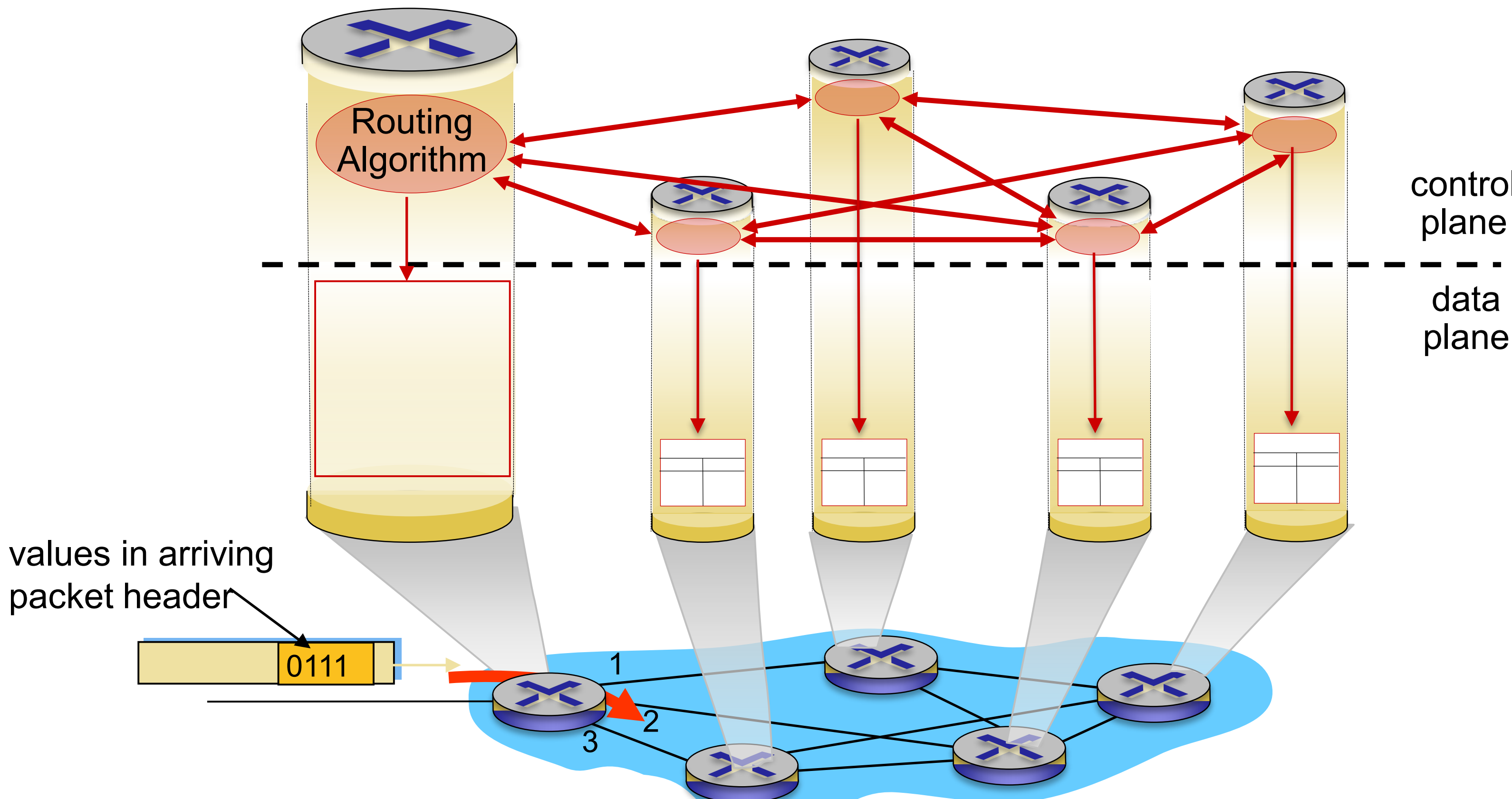


Control plane

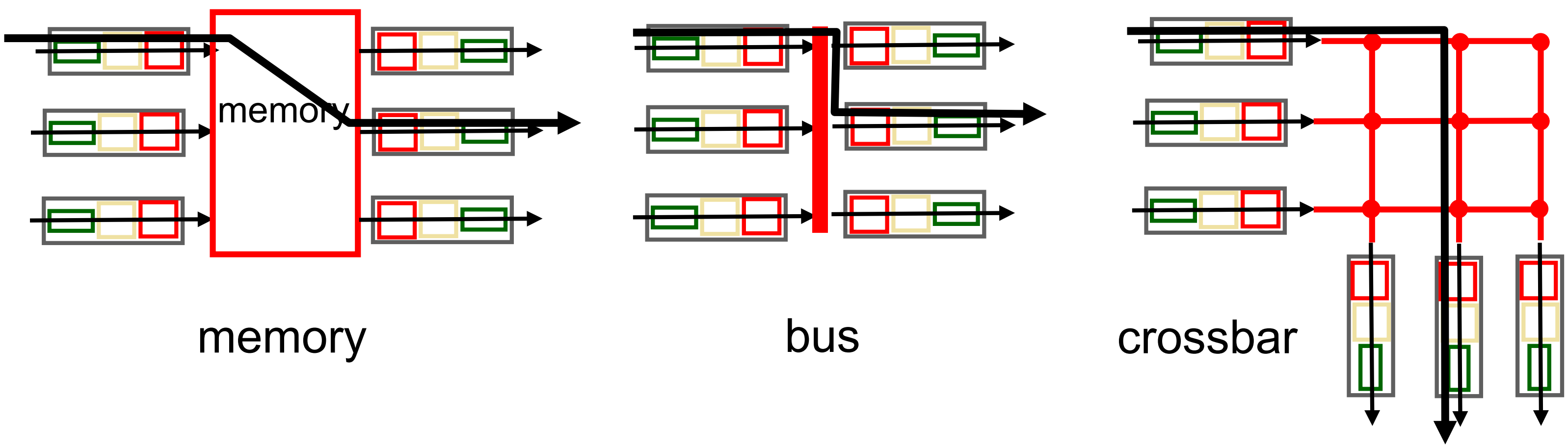
- network-wide logic
- determines how datagram is routed among routers along end-end path from source host to destination host
- two control-plane approaches:
  - *traditional routing algorithms:* implemented in routers
  - *software-defined networking (SDN)*: implemented in (remote) servers

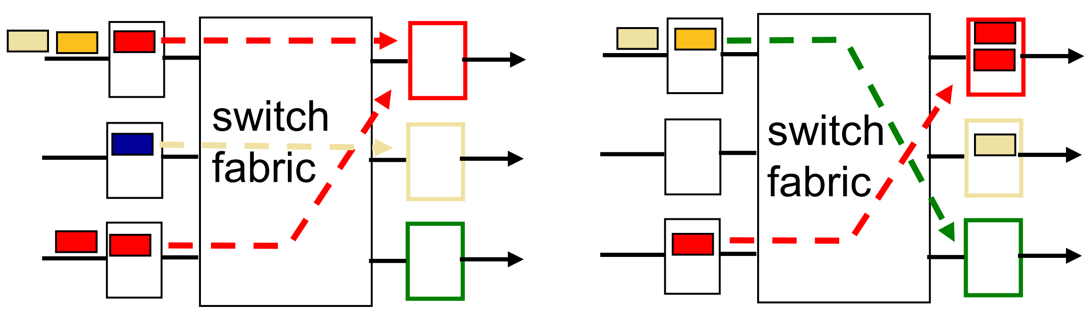
Per-router control plane

Individual routing algorithm components *in each and every router* interact in the control plane



Switching Hardware (extra) (virtual circuits)





Routing/Forwarding

For the routing table we consider:

- path: sequence of routers packets will traverse in going from given initial source host to given final destination host
- “good”: least “cost”, “fastest”, “least congested”
- Prefix, port, metric, … misc other info.

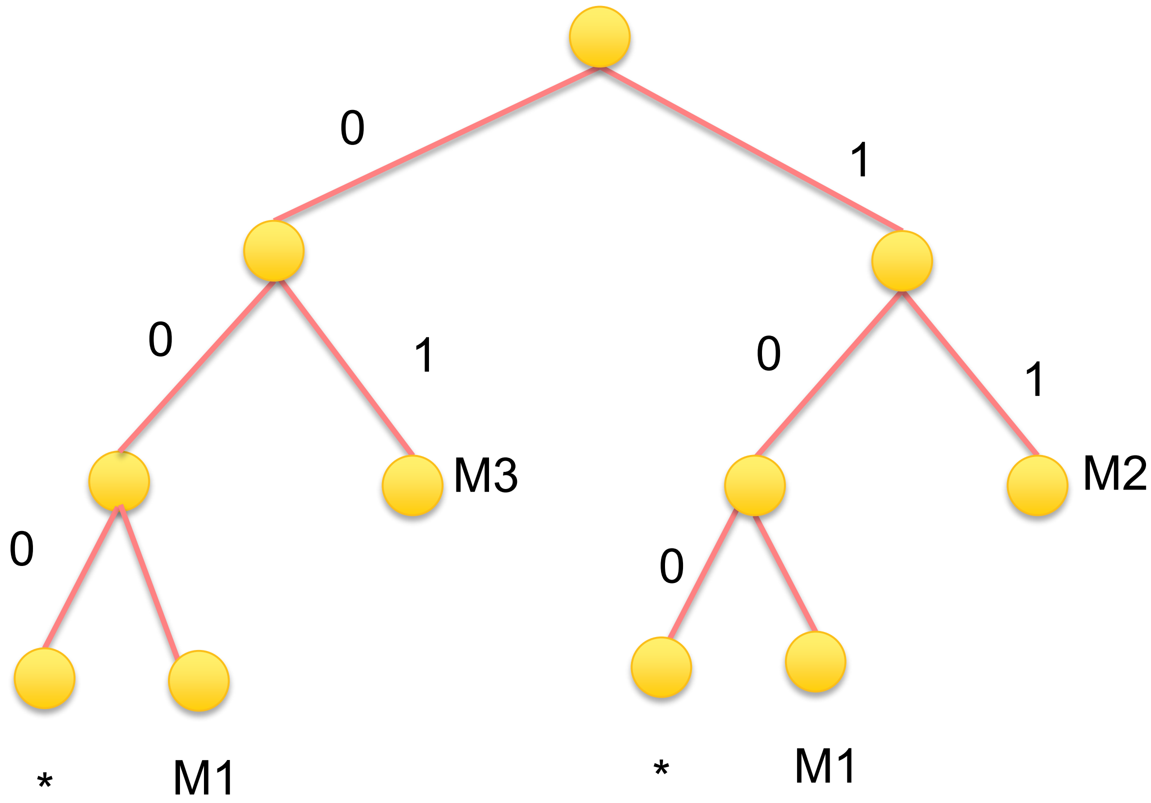
Forwarding Table:

- Simple prefix:port pairs..

Longest Prefix Match

- Table of CIDR network addresses of different length? Which do you match?
  - 23/8
  - 23.15/16
- Table matching
  - Convert the advertised address to binary.
  - Strip off any bits past the X bits (A.B.C.D/x)
  - Always a default that matches everything (default value 0.0.0.0 (or /0), see next bullet)
  - See how many bits of the prefix match the destination IP. (use a MASK with prefix 1's and AND)

Longest Prefix Match (extra)



ROUTING PROTOCOL

Routing Protocol

GOAL: determine “good” paths (equivalently, routes), from sending hosts to receiving host, through network of routers

- path: sequence of routers packets will traverse in going from given initial source host to given final destination host
- “good”: least “cost”, “fastest”, “least congested”

Routing Protocol

- Configured statically by network operators.
- Configured centrally by software (Software Defined Network).
- Dynamically by having the routers exchange messages (i.e. routing protocol)
  - Construct an entire network map
  - Sign-post, only know about the next hop
  - Failure, convergence to the same view of the network.
  - Distributed and decentralized

Interior Gateway Routing Protocols (IGP – OSPF, IGRP, RIP)

- Link State
  - Broadcast link information to all routers.
  - Create a “map” of the network.

- Distance Vector
  - Send local information to neighbouring routers.
  - Create “signposts” for routing.

Obtain state information by sending route advertisements.

Interior Gateway Protocol (IGP)

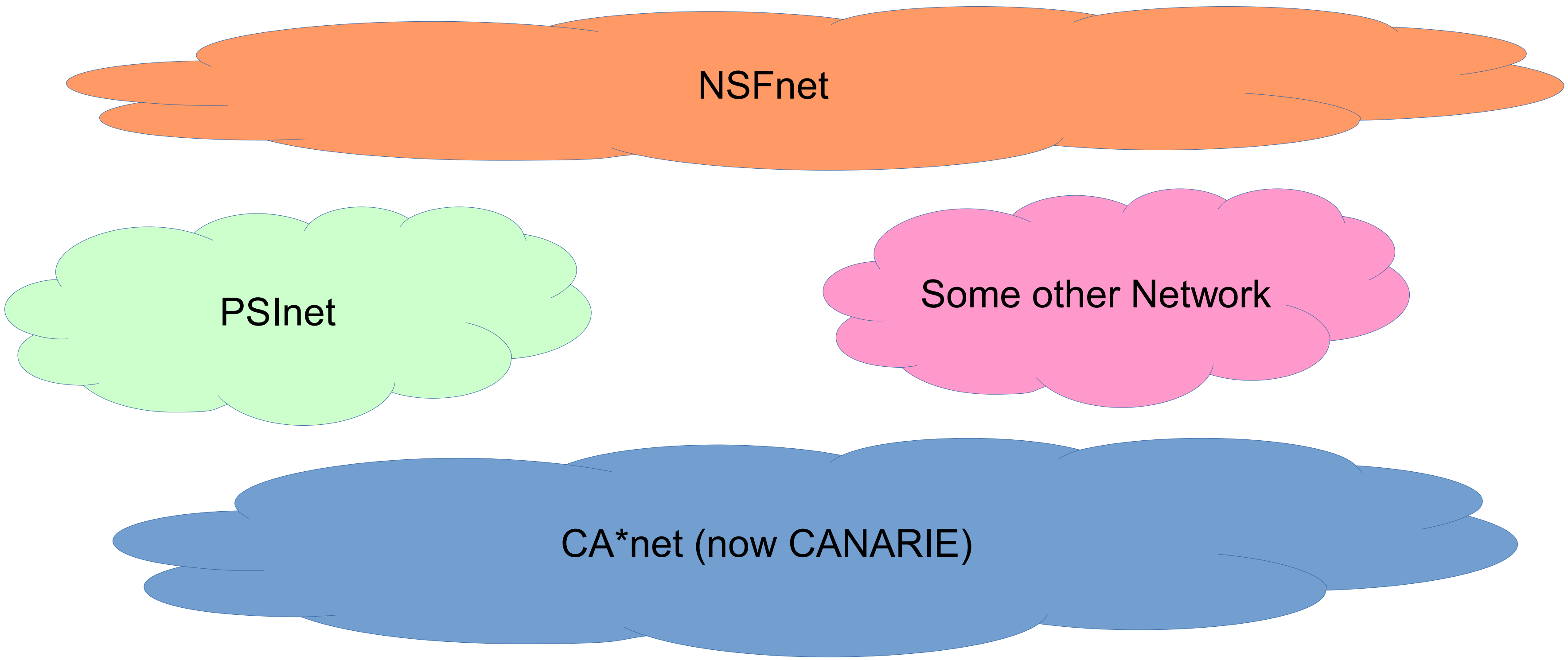
- OSPF, open-shortest-path-first.
- OSPF supports areas as well and an area can support hundreds of routers (depends on hardware)
- Within an area it uses a distributed shortest path algorithm to discover routes, floods information. (Link State)
- Between areas it uses summarization to reduce table sizes. (covered in later slides)

NETWORK of NETWORKS

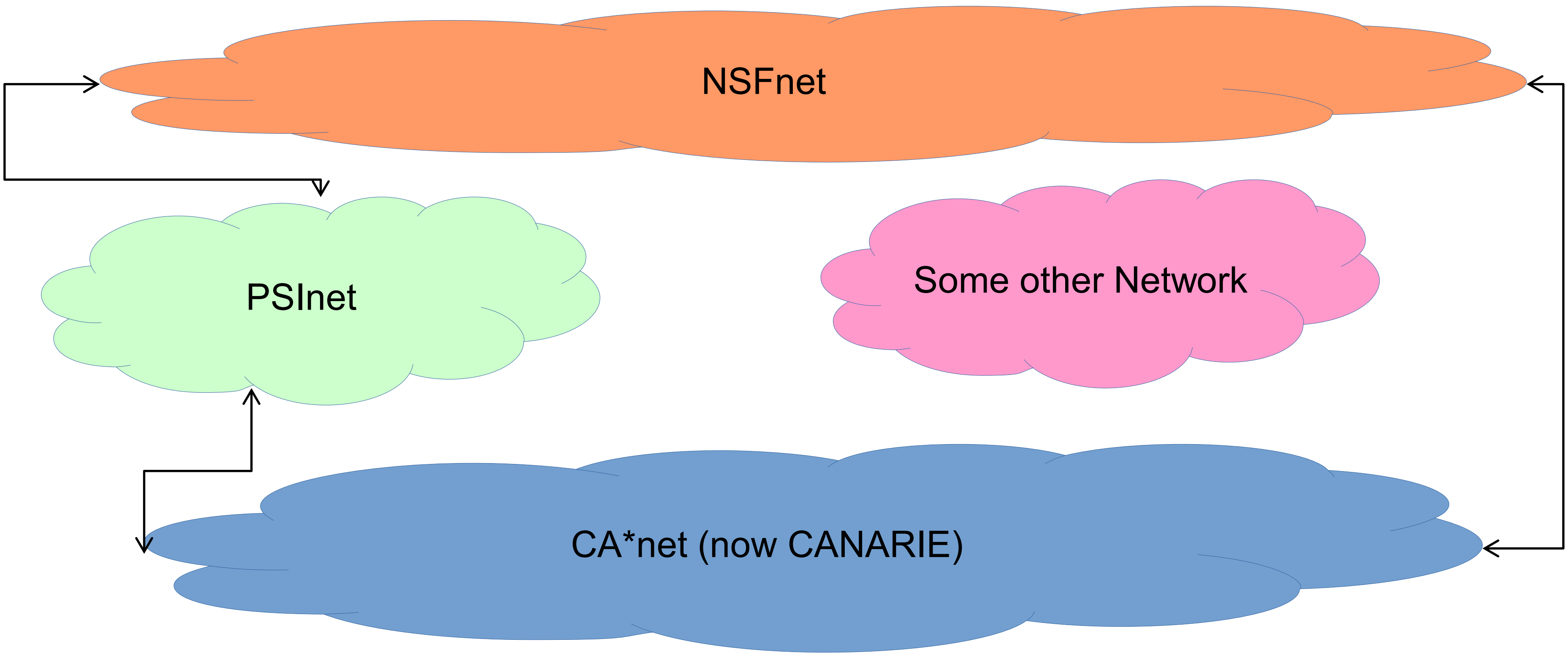
Interior Routing Protocols (next hop)

- Do not scale
- Do not account for administrative differences (administrative autonomy)
  - Political
  - Company
  - International boundaries
- Did not differentiate (stub, transit)
  - i.e. who will pay for transit across oceans?

In the beginning, imagine a bunch of private networks



Start joining them



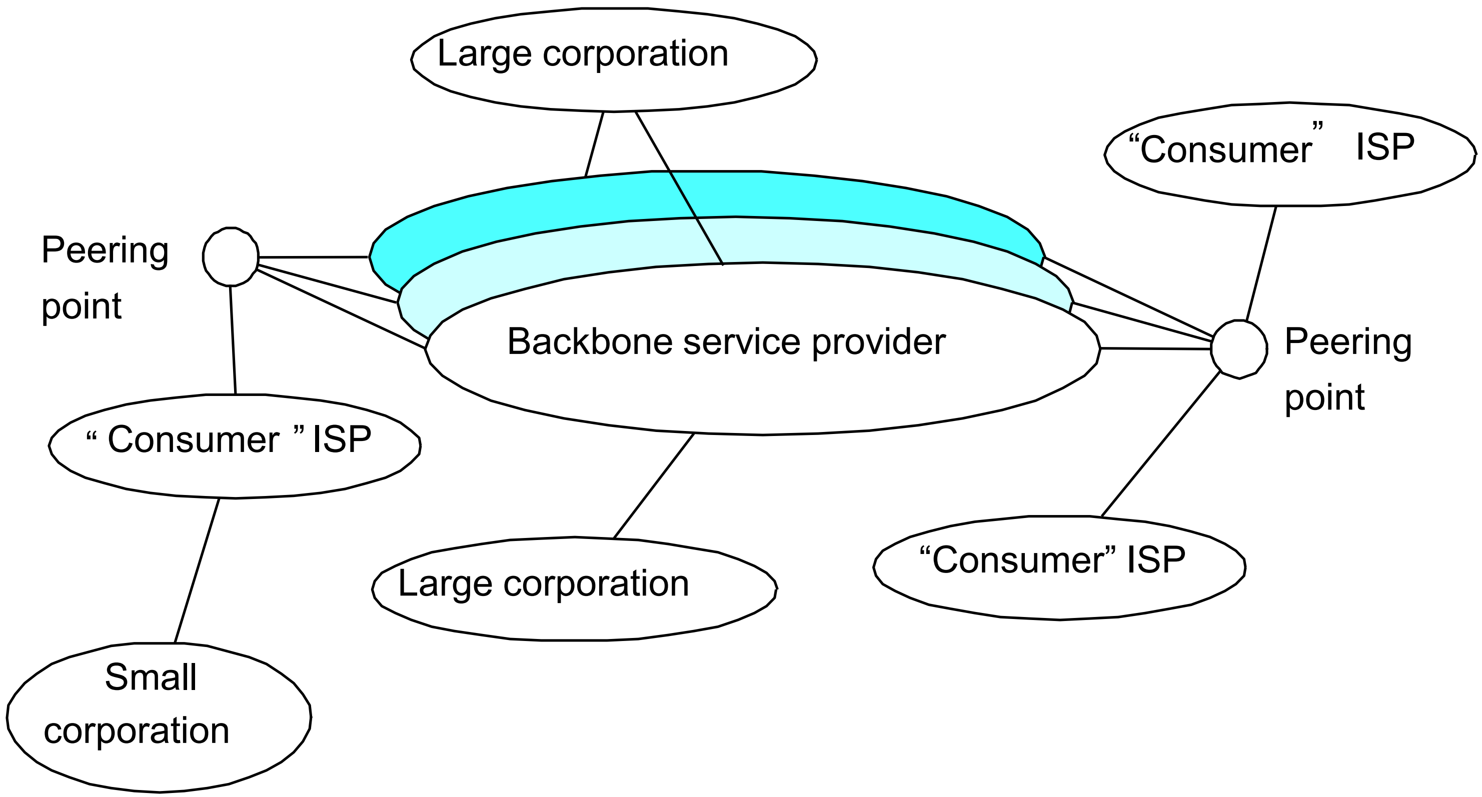
AS ORGANIZATION

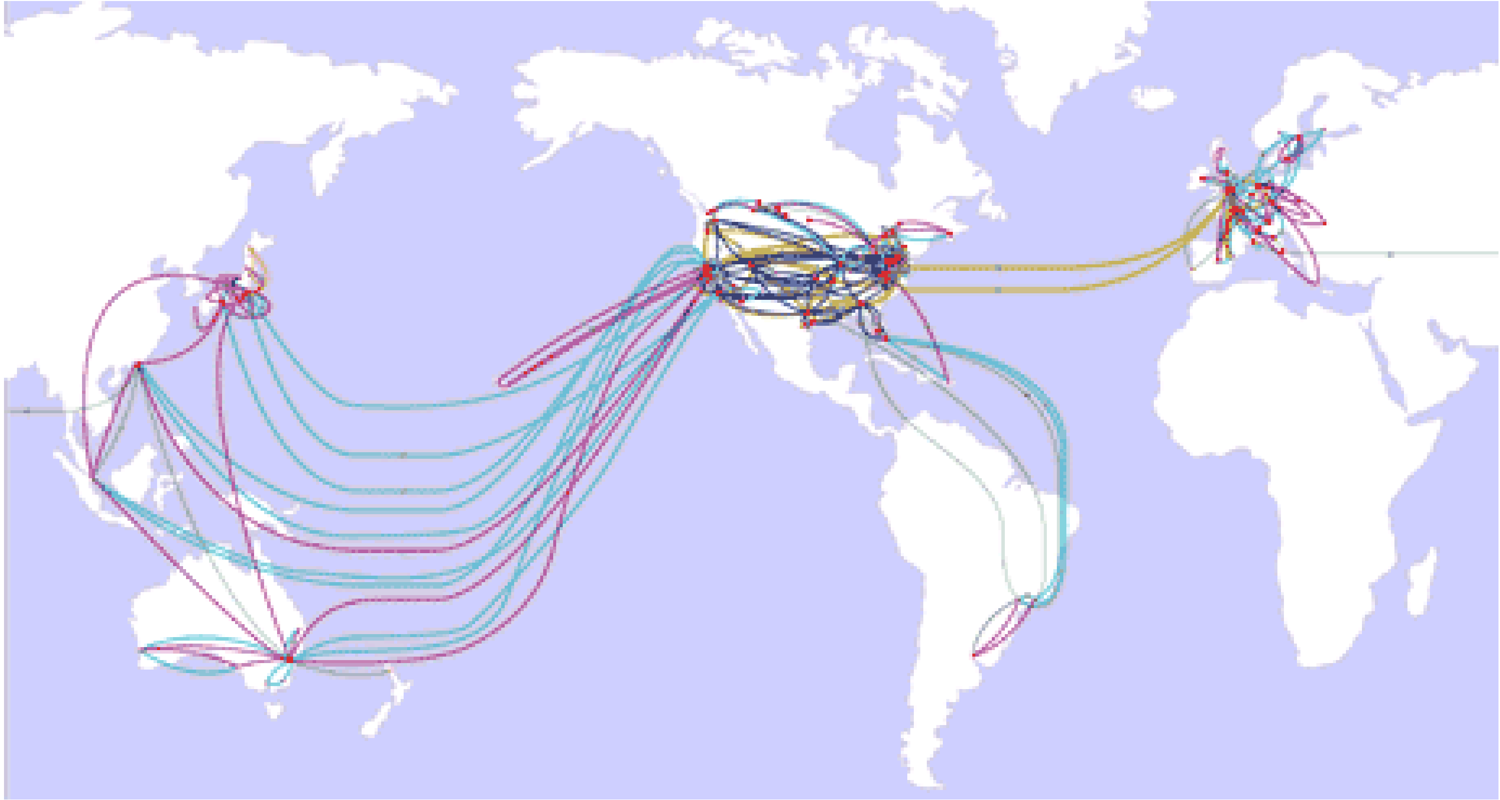
Autonomous System (AS)

- An **autonomous system** (**AS**) is a collection of connected Internet Protocol (IP) routing prefixes under the control of one or more network operators on behalf of a single administrative entity or domain that presents a common, clearly defined routing policy to the Internet.[1]
- A unique ASN is allocated to each AS for use routing. The ASN uniquely identifies each network on the Internet.
- Until 2007, AS numbers were defined as 16-bit integers, but are now 32-bit numbers (still supporting the old style).

wikipedia

Today's Internet

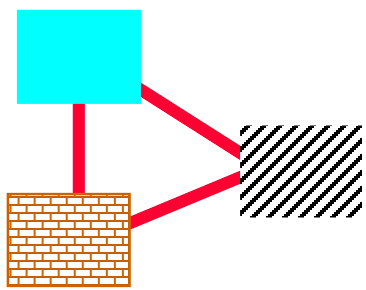




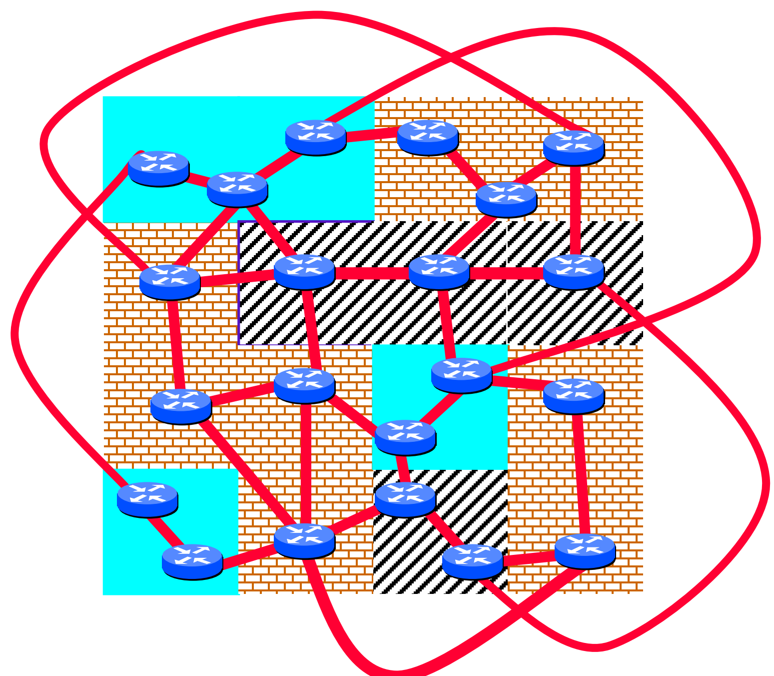
- 64 Kbps
- T1/E1 (1.5 Mbps/2 Mbps)
- E3/T3/DS3 (35 Mbps/45 Mbps)
- T2 (6 Mbps)
- OC3c/STM1 (155 Mbps)
- OC12c/STM4 (622 Mbps)
- OC48c/STM16 (2.5 Gbps)
- OC192c/STM64 (10 Gbps)
- Single Hub City
- Multiple Hubs City
- Data Center Hub

… the administration of an AS appears to other ASes to have a single coherent interior routing plan and presents a consistent picture of what networks are reachable through it.

RFC 1930: Guidelines for creation, selection, and registration of an Autonomous System



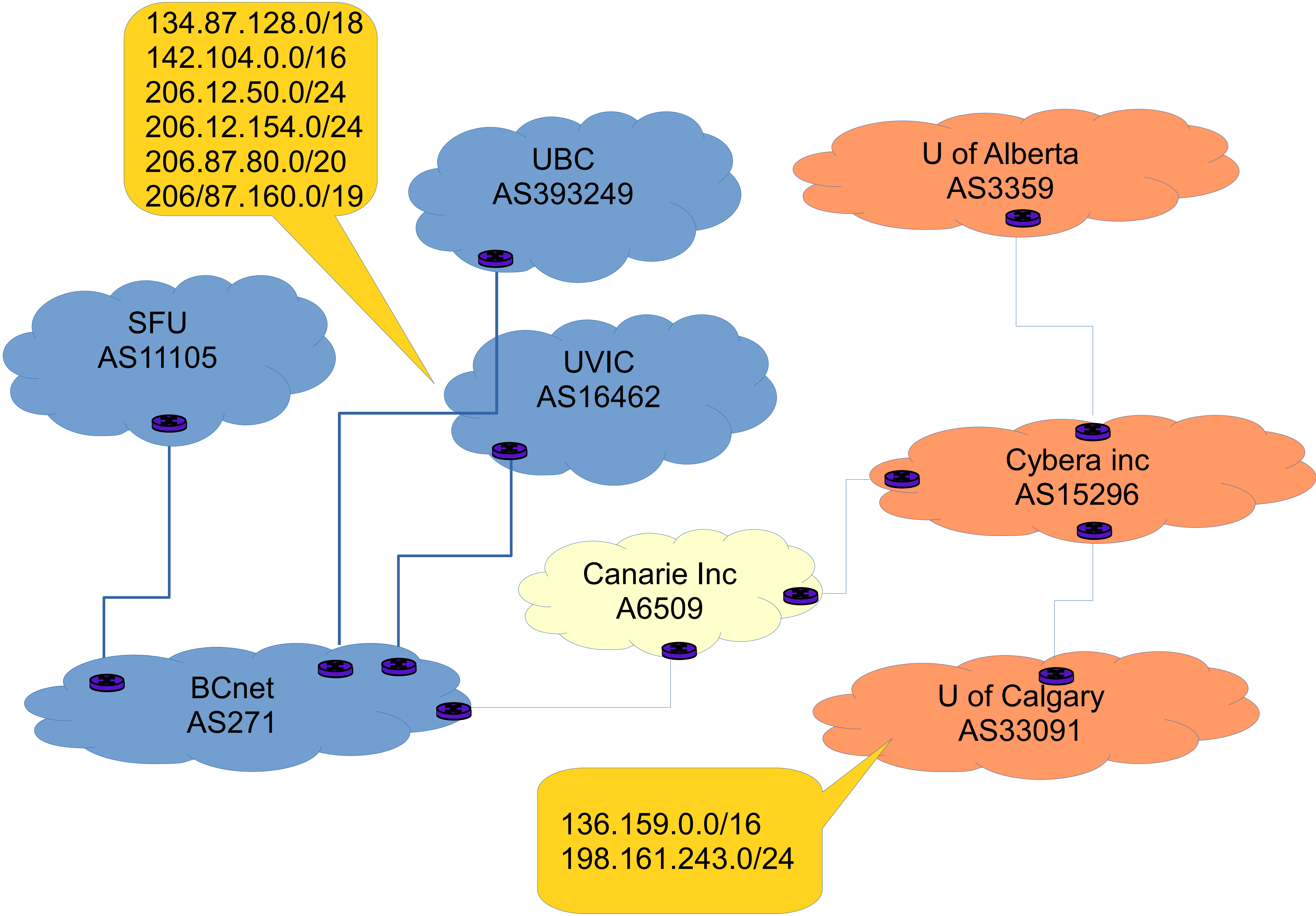
The AS graph may look like this.



Reality may be closer to this…

Peering and Transit

- Peering
  - Two ISPs pass traffic between each other for their “customers”
- Transit
  - Passing traffic across an AS to get to a different AS
- Stub network, single provider as Internet Gateway


134.87.128.0/18
142.104.0.0/16
206.12.50.0/24
206.12.154.0/24
206.87.80.0/20
206/87.160.0/19
UBC
AS393249
U of Alberta
AS3359
SFU
AS11105
UVIC
AS16462
Cybera inc
AS15296
Canarie Inc
A6509
BCnet
AS271
U of Calgary
AS33091
136.159.0.0/16
198.161.243.0/24

Routing

- Two levels of routing
  - Routing within a single AS, under the control of a single administrative entity
  - Routing between different AS, where we have no control over the routing policies of another AS.
- Internal Gateway Protocols (IGP)
  - LS and DV (OSPF, IS-IS, RIP …), do NOT scale.
- Another protocol for ASes (External Gateway Protocol) – called inter-domain rather than intra-domain routing protocol.

Border Gateway Protocol

What problem is BGP solving?

How do we route among organizations with different policies, business models and trust?

- Border Gateway Protocol (BGP-4)
- De-facto standard inter-domain routing protocol (inter AS routing)
- Enables policies in routing decisions

InterAS Routing protocol?

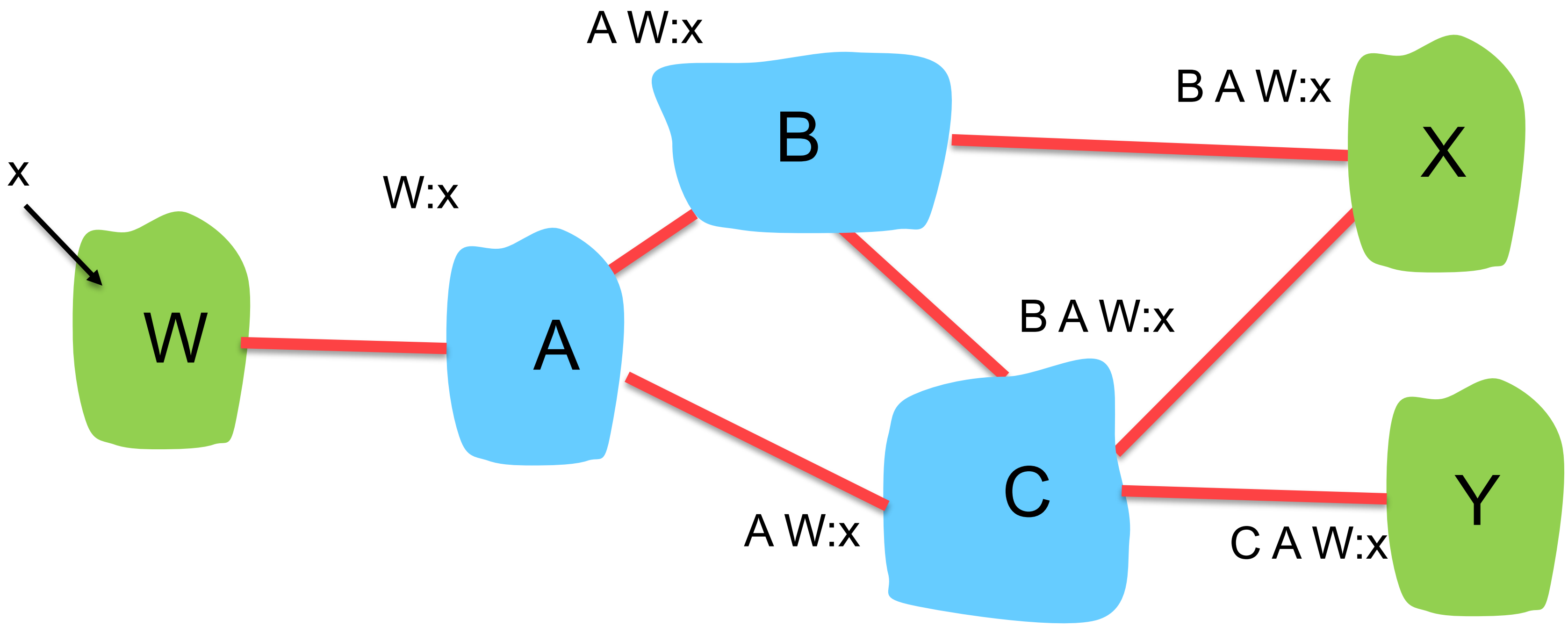
- Obtain prefix reachability information from neighbouring ASes (advertising

- Determine the “best” routes to the prefixes.

BGP uses a *path vector* routing protocol (see RFC 1322)

BGP

BGP – Path Vector Routing

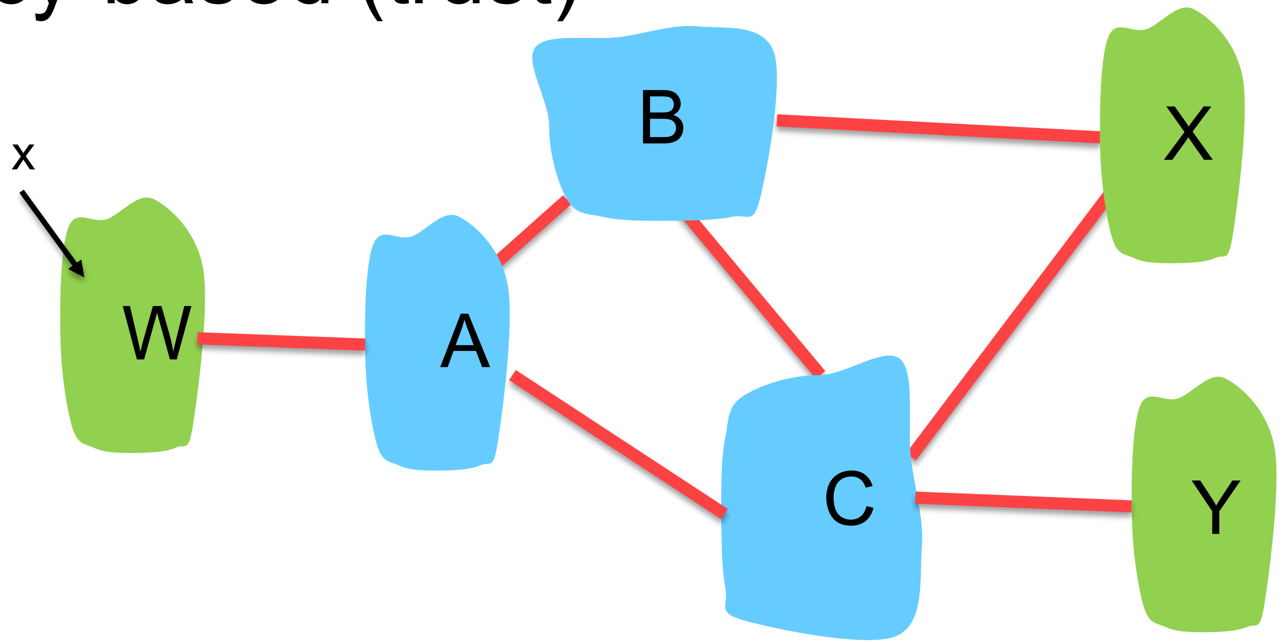


How to find out about x?

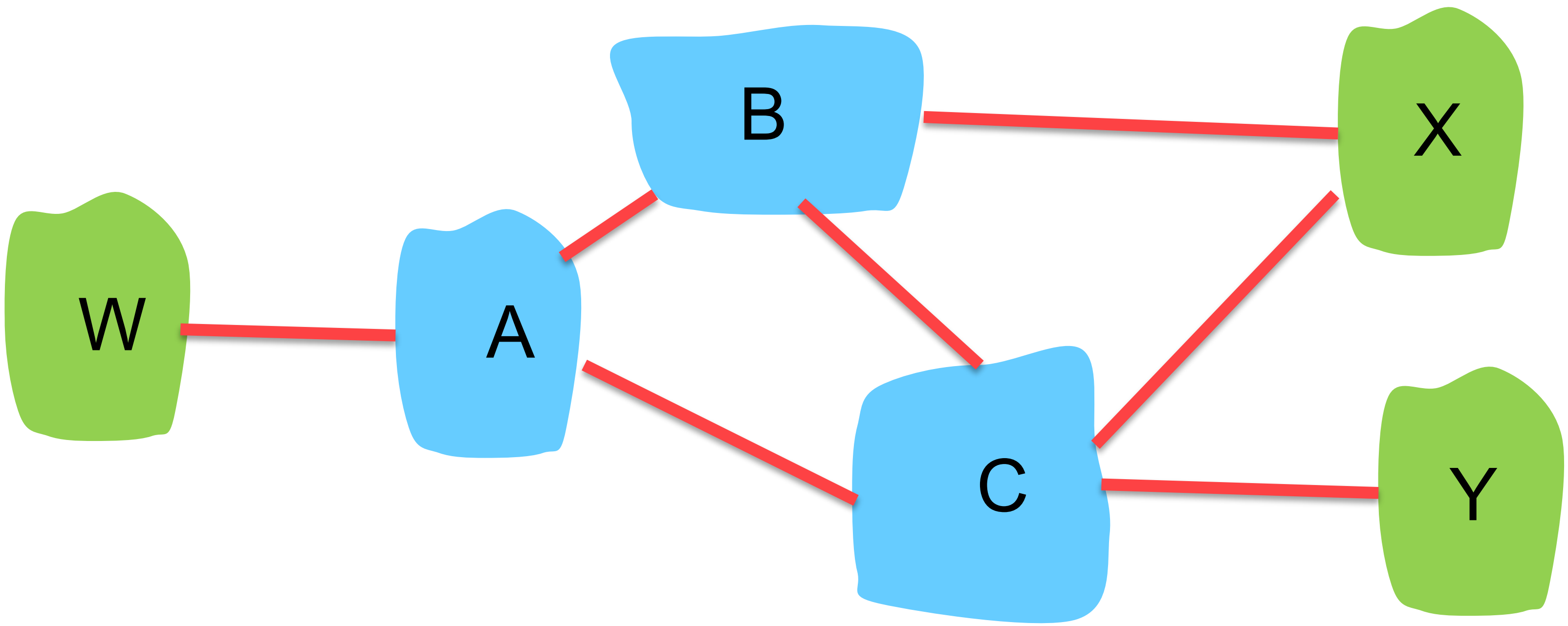
Advertise reachability of x to other ASes.

Path Vector

- Reachability information to x
- Loop Detection
- Policy-based (trust)



Selective Advertising



Advertisement to Cybera?

Advertises:
AS271:AS16462
AS6509: AS271 AS16462 x

Canarie Inc
AS6509

Cybera Inc
AS15296

BGP Route Advertisements

Scalable Design

- Interior gateway protocols cannot scale to the Internet
- Routing Table sizes is a problem in the core.
- BGP routes between prefixes not networks.

Route Aggregation

Route Aggregation

- Route summarization
- Reduces the size of the routing table
- Reduces the number of advertisements

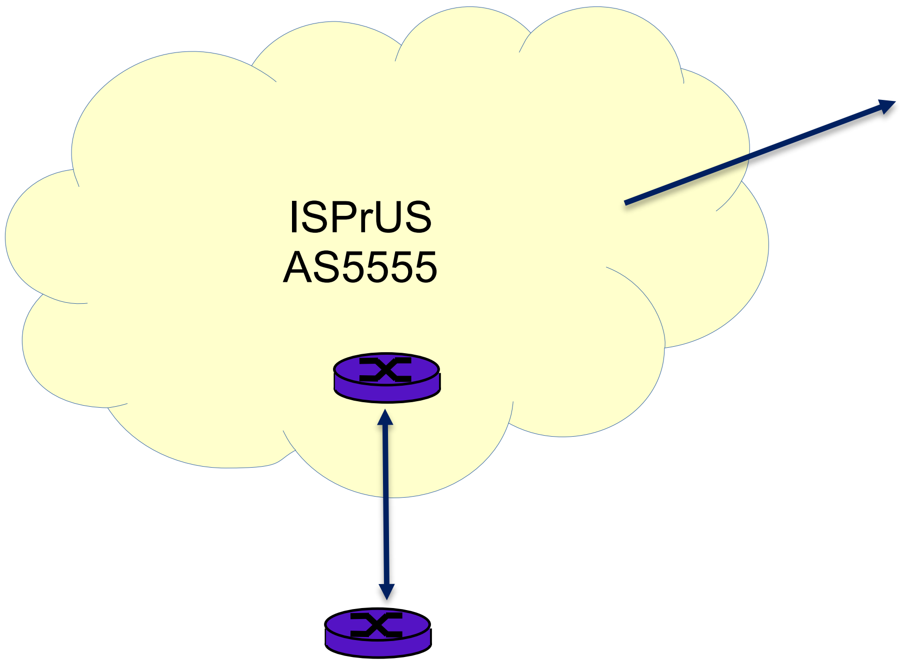
- Inside an AS it is called supernetting (the 200.23.26.0/21 is a supernet)

Exterior Gateway Protocol

- BGP (Border Gateway Protocol).
- Border Routers
- Between ASes it uses summarization to reduce table sizes.
- iBGP and eBGP (uses TCP)

Advertises:
Send to AS271: msg AS16462 x   -- network x exists and is here
Send to AS6509: msg AS271 AS16462 x

Example



200.23.64.0/20
200.23.20.0/22
200.23.16.0/22

ISPrUS

- ISPrUS agrees to host (advertise) our IP range (ISP and AS – autonomous system)

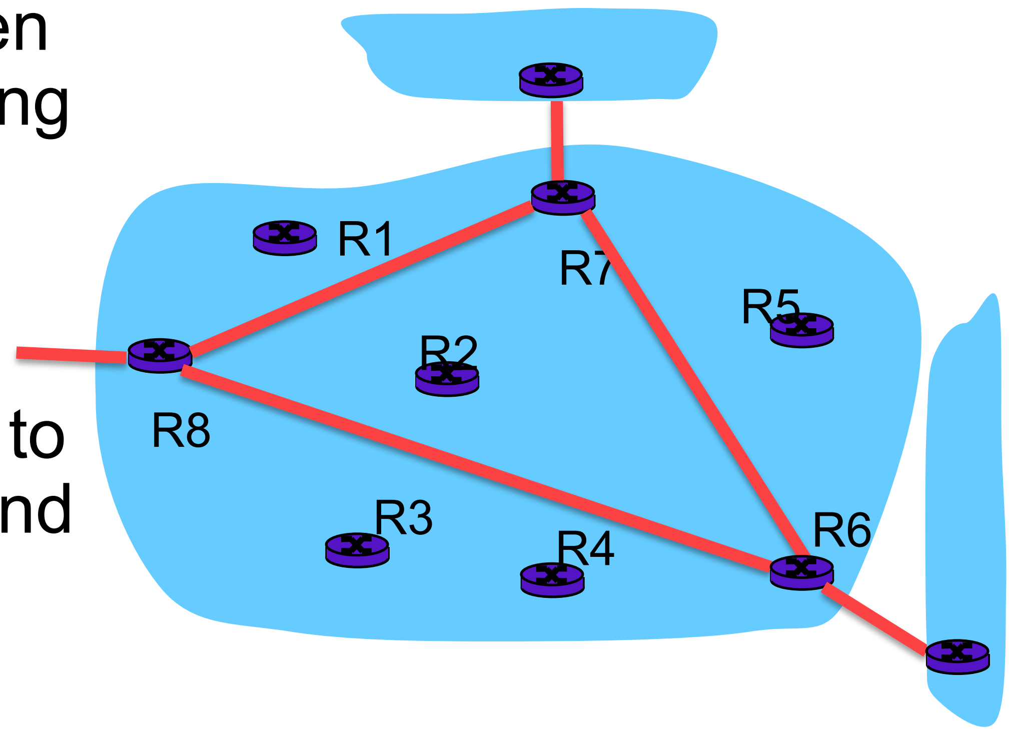
Advertise the following IP ranges:

200.23.64.0/20
200.23.20.0/22

BGP DETAILS
(connecting inside to outside)

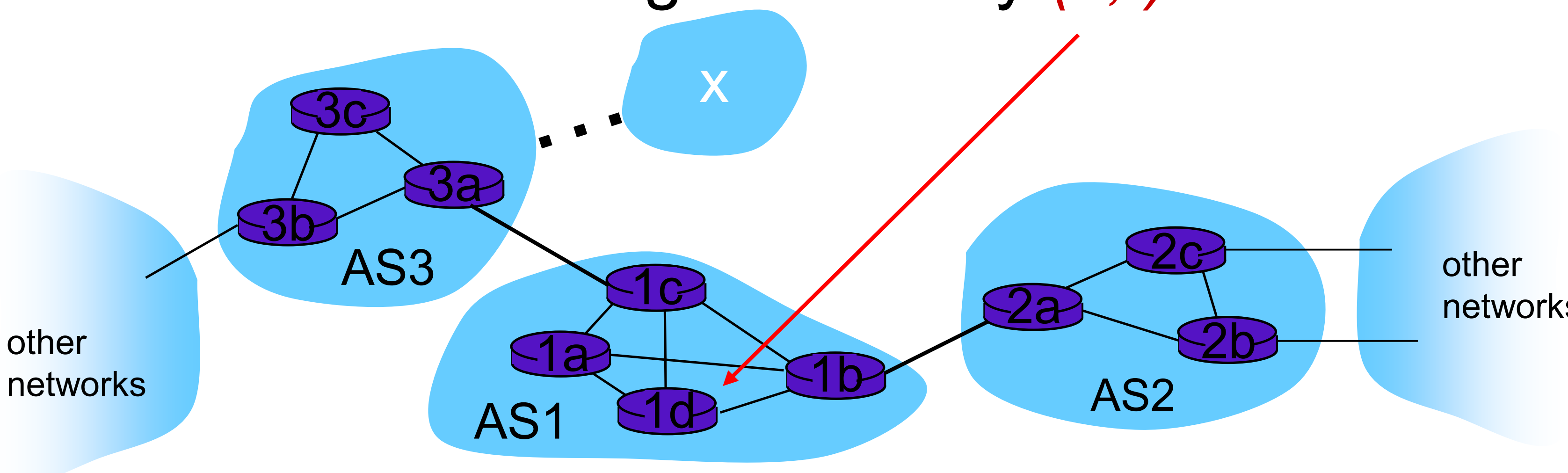
BGP Terminology

- BGP speaker – creates a TCP connection between two routers both speaking BGP
- Border router
- iBGP: propagate reachability information to all AS-internal routers and advertise prefix
- eBGP: obtain subnet reachability information from neighboring ASs.



What about internal routers?

- suppose AS1 learns (via 1c, eBGP) that subnet *x* is reachable via AS3 (gateway 1c), but not via AS2
  - Routers use iBGP to distribute info to all routers
- router 1d determines from iBGP that its interface *l* is on the least cost path to 1c (and not 1b)
  - installs forwarding table entry *(x,l)*



What information is passed?

- BGP speakers, eBGP, iBGP
- BGP advertises routes and uses path-vector routing (advertising the entire path)

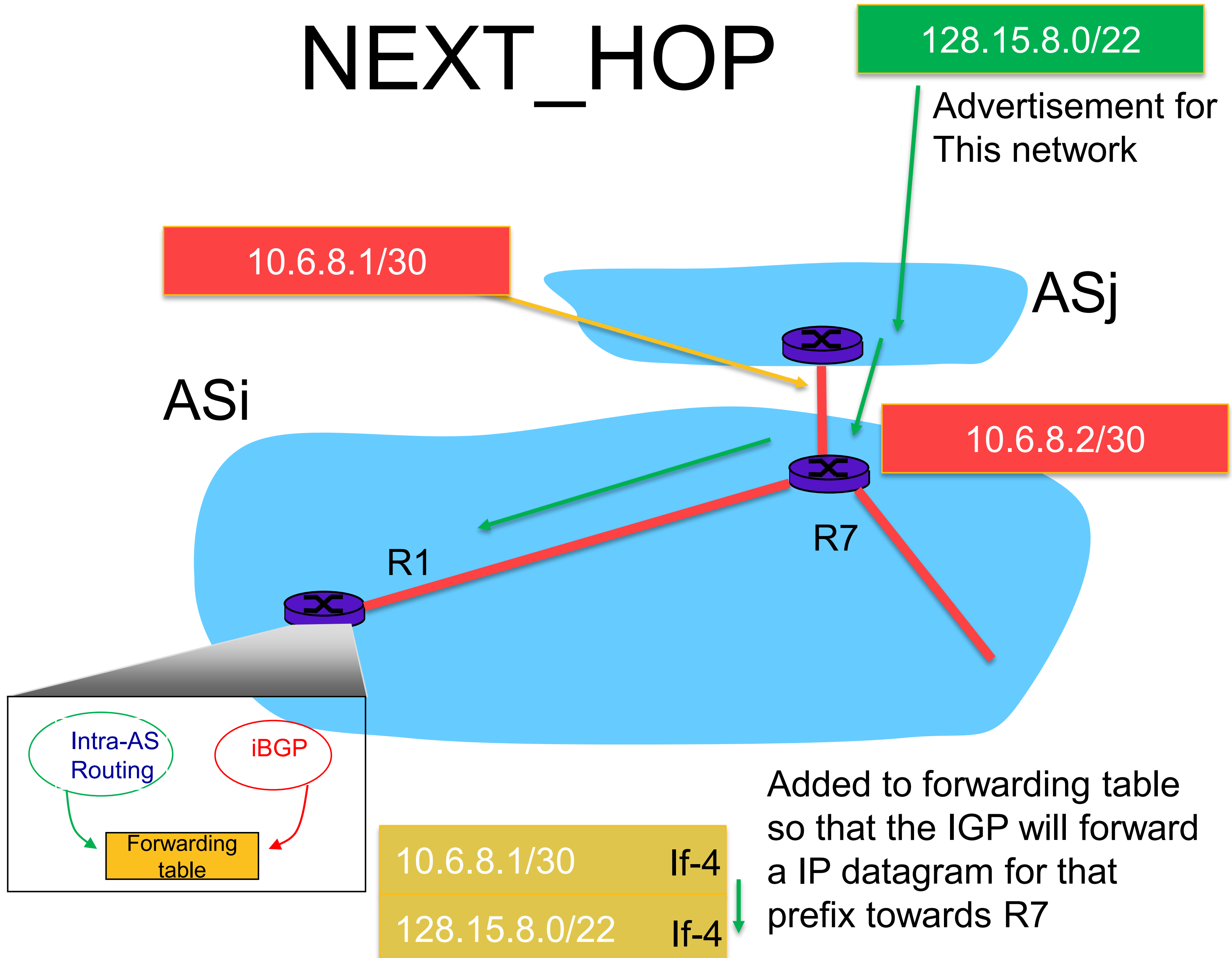
Advertisement consists of:

a. Network prefix of a destination network

b. Route   x: A -> B -> C … (a path vector)

c. Next Hop (IP address of interface that begins the AS path)

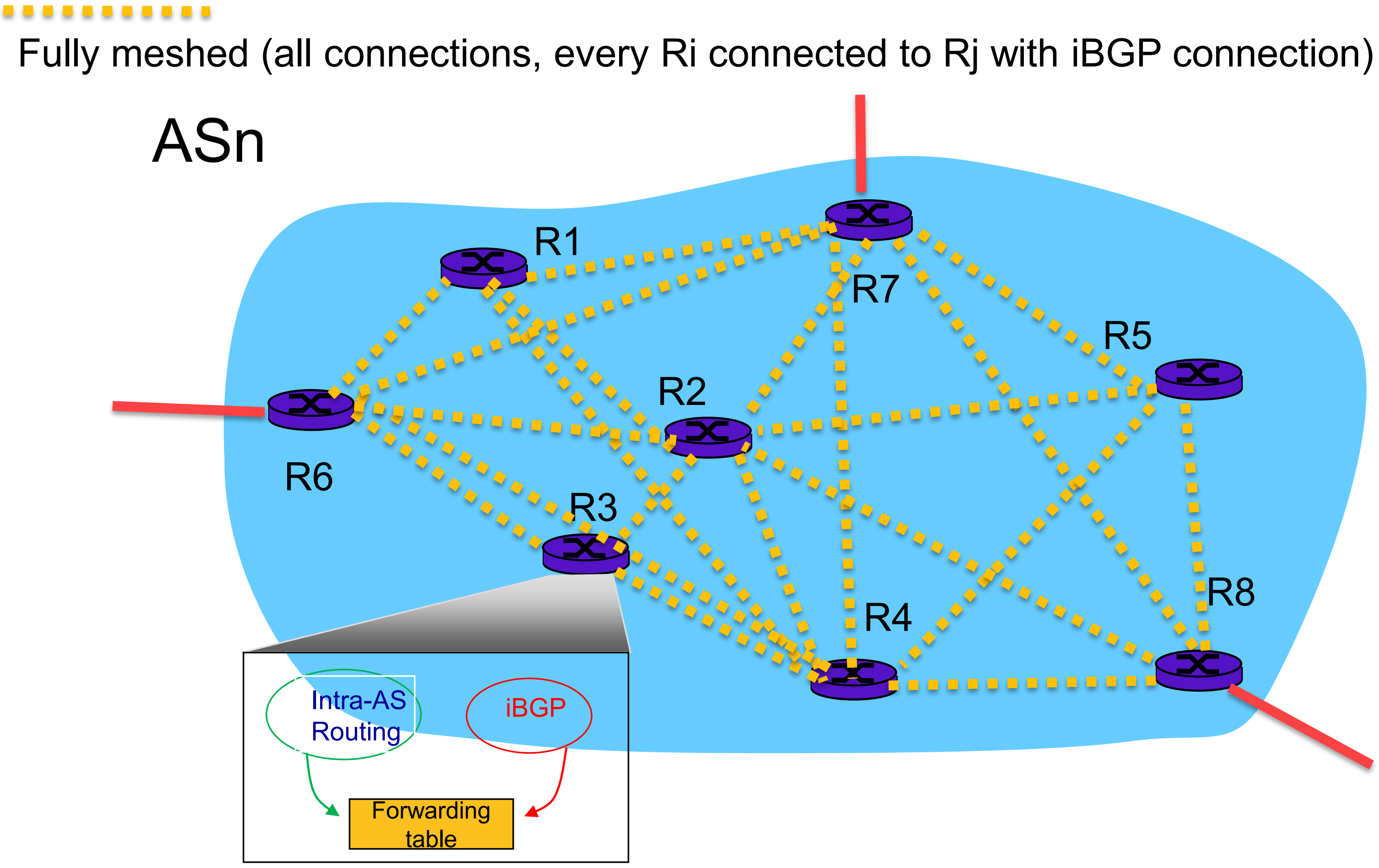
NEXT_HOP



Added to forwarding table so that the IGP will forward a IP datagram for that prefix towards R7

Hot Potato Routing

Fully meshed (all connections, every Ri connected to Rj with iBGP connection)



BGP Routing

- Policy
- Number of AS hops
- Shortest IGP route
- Some other identifier

BGP Route Advertisements (100,000 or even 500,000)

Remote Sites

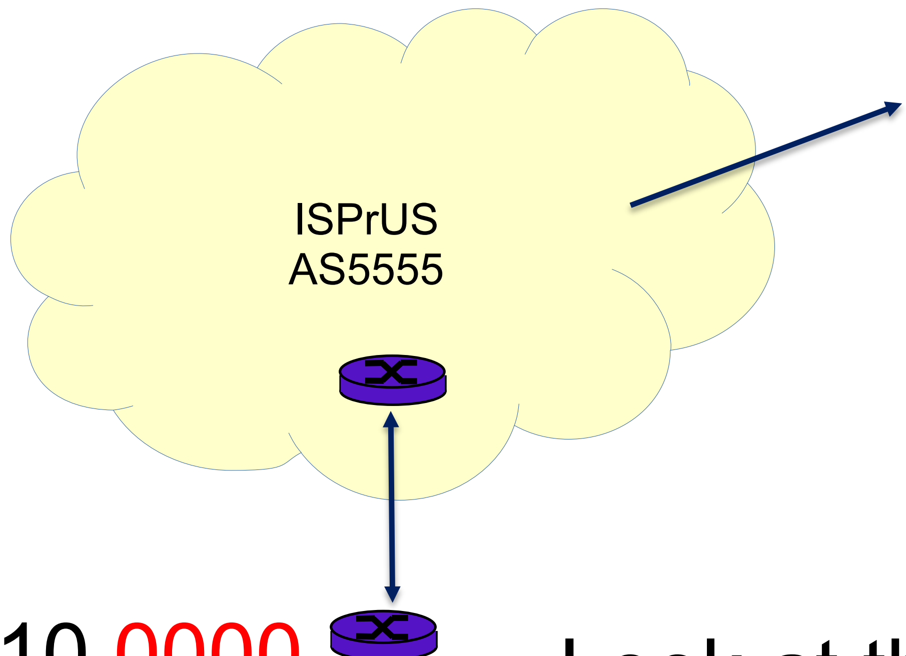
- https://tools.keycdn.com/bgp-looking-glass
  Show ip bgp summary
    - Whose BGP feeds do the router take?
  - Show ip bgp
    - Prefix
    - Origin AS
    - AS Path
- Collected at http://archive.routeviews.org/
- Other BGP table collections are:
  - http://www.ripe.net/projects/ris/rawdata.html

Scalable Design

- Interior gateway protocols cannot scale to the Internet
- Routing Table sizes is a problem in the core.
- Lets' allow BGP route between prefixes not just networks. (what does this mean?)

Route Aggregation (summarization, supernetting)

Example



200.23.32.0/20
200.23.20.0/22
200.23.16.0/22

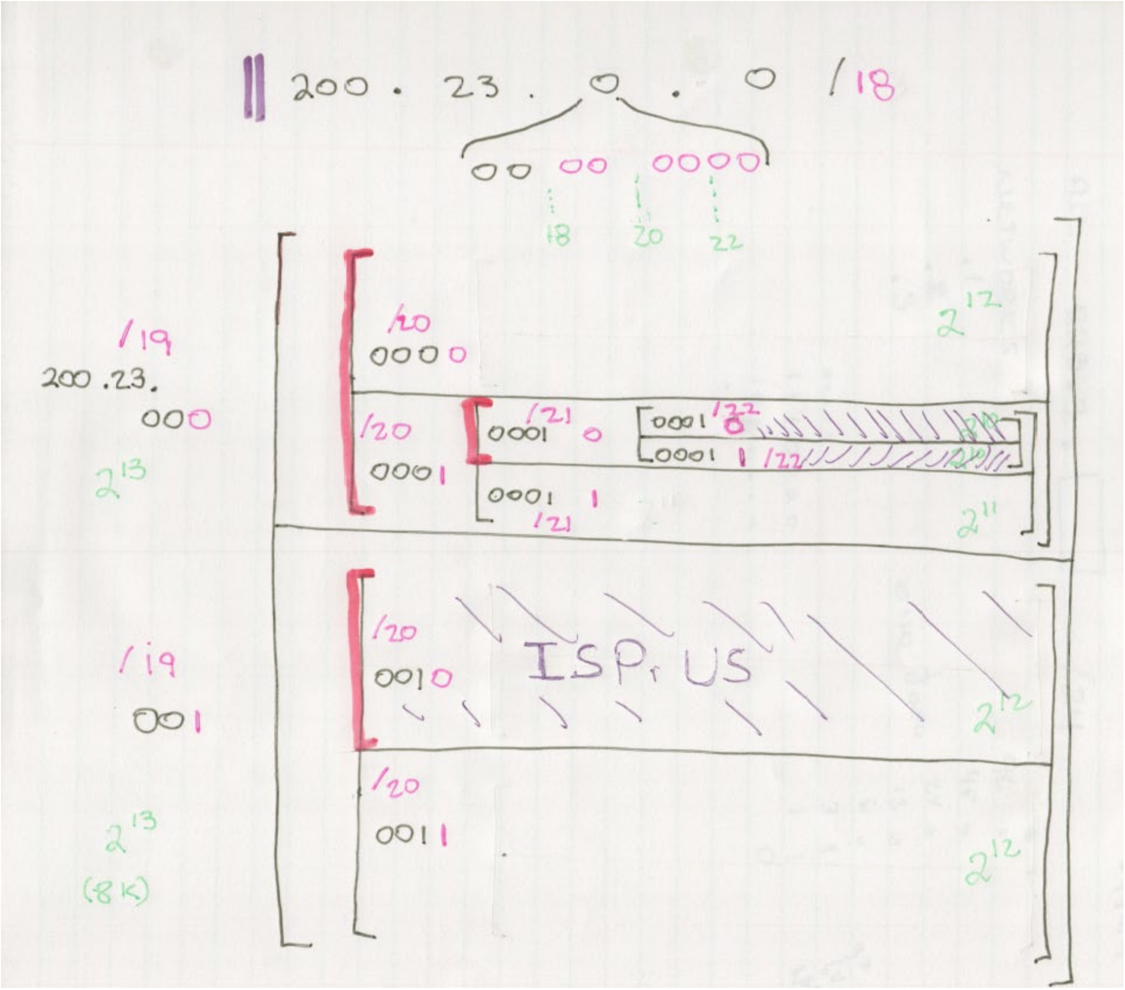
0010 0000
0001 0100
0001 0000

Look at these two blocks:
200.23.20.0/22
200.23.16.0/22

Blocks of IP addresses

Blocks of IP addresses advertised by the AS:
200.23.32.0/20
200.23.20.0/22
200.23.16.0/22

The picture shows each bit and how it splits the block into two parts. We are showing the third octet



ISPrUS

- ISPrUS agrees to host (advertise) our IP range (ISP and AS – autonomous system)

Advertise the following IP ranges:

200.23.32.0/20
200.23.20.0/22
200.23.16.0/22

Does it have to advertise both blocks? NO

200.23.16.0/21

Route Aggregation

- Route summarization
- Reduces the size of the routing table
- Reduces the number of advertisements

- Inside an AS it is called supernetting (the 200.23.16.0/21 is a supernet)

Network Discovery Tools

Network Discovery

- ICMP (Internet Control Management Protocol) RFC 792
- Ping
- Traceroute (tracert on windows)

Network Discovery

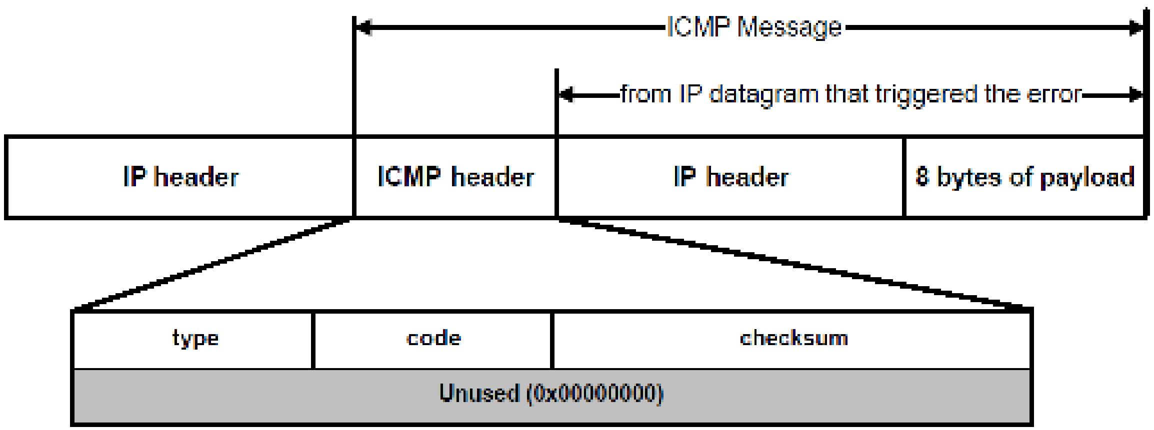
- Ping
  - measure the time for a packet to travel to a remote host and back
  - The server sends back an acknowledgment when the packet arrives
- Traceroute
  - list the router hops between the client host and a remote host.
  - The IP address and domain name (if there is one) of each router is returned to the client

ICMP message types

- used by hosts and routers to communicate network-level information
  - error reporting: unreachable host, network, port, protocol
  - echo request/reply (used by ping)
- network-layer "above" IP:
  - ICMP messages carried in IP datagrams
- ICMP message: type, code plus first 8 bytes of IP datagram causing error

Type	Code	description
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

ICMP



ICMP: Message Types

Type	Message
0	Echo reply
3	Destination unreachable
4	Source quench
5	Redirect
8	Echo request
11	Time exceeded
12	Parameter unintelligible
13	Time-stamp request
14	Time-stamp reply
15	Information request
16	Information reply
17	Address mask request
18	Address mask reply

Ping in Wireshark

```
→ 380 23.943634   198.162.52.230   142.103.6.5      ICMP   98 Echo (ping) request  id=0x1aee, seq=1/256, ttl=128 (reply in 381)
← 381 23.944666   142.103.6.5      198.162.52.230   ICMP   98 Echo (ping) reply    id=0x1aee, seq=1/256, ttl=62 (request in 380)
  382 23.945796   198.162.52.230   142.103.6.6      DNS    84 Standard query 0x5c7f PTR 5.6.103.142.in-addr.arpa

> Frame 380: 98 bytes on wire (784 bits), 98 bytes captured (784 bits) on interface 0
> Ethernet II, Src: Microsof_9a:64:79 (58:82:a8:9a:64:79), Dst: All-HSRP-routers_00 (00:00:0c:07:ac:00)
v Internet Protocol Version 4, Src: 198.162.52.230, Dst: 142.103.6.5
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
  > Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 84
    Identification: 0x7d92 (32146)
  > Flags: 0x00
    Fragment offset: 0
    Time to live: 128
    Protocol: ICMP (1)
    Header checksum: 0x0000 [validation disabled]
    [Header checksum status: Unverified]
    Source: 198.162.52.230
    Destination: 142.103.6.5
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
v Internet Control Message Protocol
    Type: 8 (Echo (ping) request)
    Code: 0
    Checksum: 0xcc29 [correct]
    [Checksum Status: Good]
    Identifier (BE): 6894 (0x1aee)
    Identifier (LE): 60954 (0xee1a)
    Sequence number (BE): 1 (0x0001)
    Sequence number (LE): 256 (0x0100)
    [Response frame: 381]
    Timestamp from icmp data: Jan 14, 2018 09:40:50.000000000 Pacific Standard Time
    [Timestamp from icmp data (relative): 0.074431000 seconds]
  v Data (48 bytes)
      Data: d323010000000000101112131415161718191a1b1c1d1e1f...
      [Length: 48]
```

Ping reply

380	23.943634	198.162.52.230	142.103.6.5	ICMP	98 Echo (ping) request id=0x1aee, seq=1/256, ttl=128 (reply in 381)
381	23.944666	142.103.6.5	198.162.52.230	ICMP	98 Echo (ping) reply id=0x1aee, seq=1/256, ttl=62 (request in 380)
382	23.945796	198.162.52.230	142.103.6.6	DNS	84 Standard query 0x5c7f PTR 5.6.103.142.in-addr.arpa

```
> Frame 381: 98 bytes on wire (784 bits), 98 bytes captured (784 bits) on interface 0
> Ethernet II, Src: Cisco_46:2c:00 (00:1e:f6:46:2c:00), Dst: Microsof_9a:64:79 (58:82:a8:9a:64:79)
v Internet Protocol Version 4, Src: 142.103.6.5, Dst: 198.162.52.230
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
  > Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 84
    Identification: 0x6bbc (27580)
  > Flags: 0x00
    Fragment offset: 0
    Time to live: 62
    Protocol: ICMP (1)
    Header checksum: 0x80f8 [validation disabled]
    [Header checksum status: Unverified]
    Source: 142.103.6.5
    Destination: 198.162.52.230
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
v Internet Control Message Protocol
    Type: 0 (Echo (ping) reply)
    Code: 0
    Checksum: 0xd429 [correct]
    [Checksum Status: Good]
    Identifier (BE): 6894 (0x1aee)
    Identifier (LE): 60954 (0xee1a)
    Sequence number (BE): 1 (0x0001)
    Sequence number (LE): 256 (0x0100)
    [Request frame: 380]
    [Response time: 1.032 ms]
    Timestamp from icmp data: Jan 14, 2018 09:40:50.000000000 Pacific Standard Time
    [Timestamp from icmp data (relative): 0.075463000 seconds]
  v Data (48 bytes)
      Data: d32301000000000010111213141516171819 1a1b1c1d1e1f...
      [Length: 48]
```

How traceroute works

- Uses TTL of IP (both ICMP for IPv4 and IPv6)
- UDP version: sends a packet to a port it hopes isn't in use,
  - sets TTL to 1, sends UDP datagram
  - Sets TTL to 2, sends UDP datagram
  - … continues until TTL is large enough to reach host.
- Depends on fact that when a router decrements the TTL to 0 the router sends an ICMP “time exceeded” packet back to the source of the IP datagram.
- Sender times how long it takes to get the ICMP “time exceeded” packet for each router for each TTL value.
- Some routers don't send an ICMP response when TTL gets to 0 (security reasons), sometimes you see stars.

Traceroute to China

```
[alan]: traceroute gov.ch
traceroute to gov.ch (80.74.156.100), 64 hops max, 52 byte packets
 1  198.162.52.253 (198.162.52.253)  3.674 ms  0.469 ms  0.377 ms
 2  137.82.73.5 (137.82.73.5)  0.701 ms  1.355 ms  0.695 ms
 3  a0-a1.net.ubc.ca (142.103.78.250)  7.829 ms  5.647 ms  6.430 ms
 4  anguborder-a0.net.ubc.ca (137.82.123.137)  0.547 ms  0.536 ms  0.624 ms
 5  347-tx-cr1-ubcab.vncv1.bc.net (134.87.30.158)  1.986 ms  2.033 ms  1.895 ms
 6  v559.core1.yvr1.he.net (184.105.148.149)  18.094 ms  1.606 ms  1.643 ms
 7  100ge10-2.core1.sea1.he.net (184.105.64.109)  5.122 ms  5.017 ms  4.684 ms
 8  * * *
 9  ae-1-16.bar1.zurich1.level3.net (4.69.142.129)  171.800 ms  171.824 ms  171.945 ms
10  ae-1-16.bar1.zurich1.level3.net (4.69.142.129)  171.667 ms  171.747 ms  171.860 ms
11  l3-tengig03-cr2.ch-meta.net (213.242.82.90)  171.986 ms  171.958 ms  172.160 ms
12  nova.metanet.ch (80.74.156.100)  158.376 ms  158.227 ms  158.279 ms
```

Reading Traceroute results

- Typical reply … 3 replies about equal

- Replies vary a lot

- No reply

```
*        *        *        Request timed out.
```

Traceroute ru.ac.za

```
[alan]: traceroute ru.ac.za
traceroute to ru.ac.za (146.231.128.43), 64 hops max, 52 byte packets
 1  198.162.52.253 (198.162.52.253)  15.577 ms  24.589 ms  0.449 ms
 2  137.82.73.5 (137.82.73.5)  2.592 ms  0.789 ms  0.661 ms
 3  a0-a1.net.ubc.ca (142.103.78.250)  4.623 ms  5.465 ms  6.586 ms
 4  anguborder-a0.net.ubc.ca (137.82.123.137)  0.542 ms  0.706 ms  0.664 ms
 5  343-oran-cr2-ubcab.vncv1.bc.net (134.87.2.54)  3.543 ms  1.087 ms  1.242 ms
 6  cr1-bb3900.vantx1.bc.net (206.12.0.33)  3.926 ms  4.992 ms  1.262 ms
 7  vncv1rtr1.canarie.ca (205.189.32.172)  1.678 ms  1.497 ms  1.538 ms
 8  clgr2rtr1.canarie.ca (205.189.32.175)  12.978 ms  12.454 ms  12.388 ms
 9  wnpg1rtr1.canarie.ca (205.189.32.177)  26.933 ms  26.442 ms  26.552 ms
10  toro1rtr1.canarie.ca (205.189.32.181)  47.670 ms  47.765 ms  49.683 ms
11  mtrl2rtr1.canarie.ca (205.189.32.193)  54.181 ms  53.931 ms  54.064 ms
12  unknown.uni.net.za (196.32.209.225)  135.183 ms  140.673 ms  134.996 ms
13  196.32.209.77 (196.32.209.77)  134.934 ms  135.061 ms  140.751 ms
14  196.32.209.174 (196.32.209.174)  322.110 ms  308.501 ms  308.579 ms
15  be1-cpt1-pe1.net.tenet.ac.za (155.232.64.69)  321.174 ms  321.045 ms  320.946 ms
16  te0-0-0-0-plz1-pe1.net.tenet.ac.za (155.232.6.42)  321.162 ms  321.064 ms  321.056 ms
17  te0-0-0-0-grh1-pe1.net.tenet.ac.za (155.232.5.5)  321.413 ms  321.532 ms  321.282 ms
18  strubenedge-tenet.net.ru.ac.za (192.42.99.252)  320.713 ms  320.705 ms  320.650 ms
19  strubencore-strubenedge.net.ru.ac.za (192.42.99.247)  321.897 ms  321.816 ms  321.681 ms
20  datacentres-0-strubencore.net.ru.ac.za (146.231.0.37)  323.434 ms  323.254 ms  322.917 ms
21  vhost.ru.ac.za (146.231.128.43)  321.603 ms  321.530 ms  321.317 ms
[alan]: 
```

Traceroute from other places

- http://www.traceroute.org
  - Remote traceroute servers
  - Hundreds of them
  - Limited probe rate
  - Not always available
- http://www.caida.org/tools/measurement/skitter/
  - Dedicated remote traceroute monitors
  - Almost unlimited probe rate
  - Only a couple of dozens of them

Tips

- Sometimes ICMP packets get through when UDP packets do not and vice versa, so it may be worth trying more than one version of traceroute
- Location:
  - If there is no hostname or the hostname does not indicate a location try looking up the IP address or hostname or parts of the hostname in Google
  - Try using IP address location tools, but beware these are not always accurate
  - Use a whois server (E.g. the one on www.DNSstuff.com) to look up the organisation which owns the IP address. This will sometimes indicate the country in which the router is located
- Time:
  - If the RTT makes a big jump (50 - 150 milliseconds (ms)) the route is probably going over a long fibre cable (possibly submarine)
  - If the RTT jumps by more than 230 ms, the route may be going over a satellite link
  - Under 50ms, likely within 500 kms, a few ms likely same location