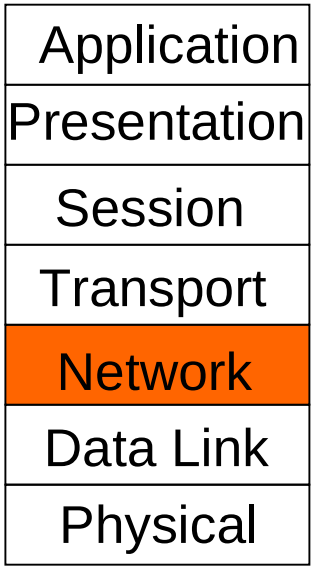
CSE/EE 461 – Lecture 8

David Wetherall

djw@cs.washington.edu

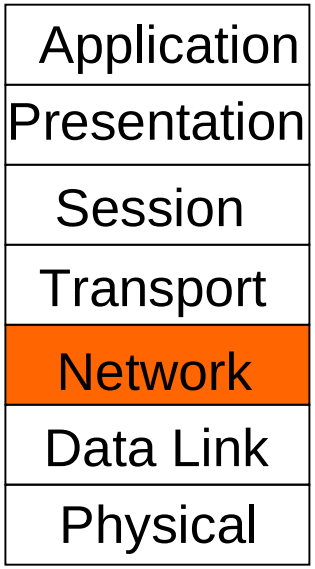
Last Time

- Focus
  - How do we calculate routes for packets?
- Routing Algorithms
  - Distance Vector routing (RIP)
  - Link State routing (OSPF)
  - Cost Metrics



This Time

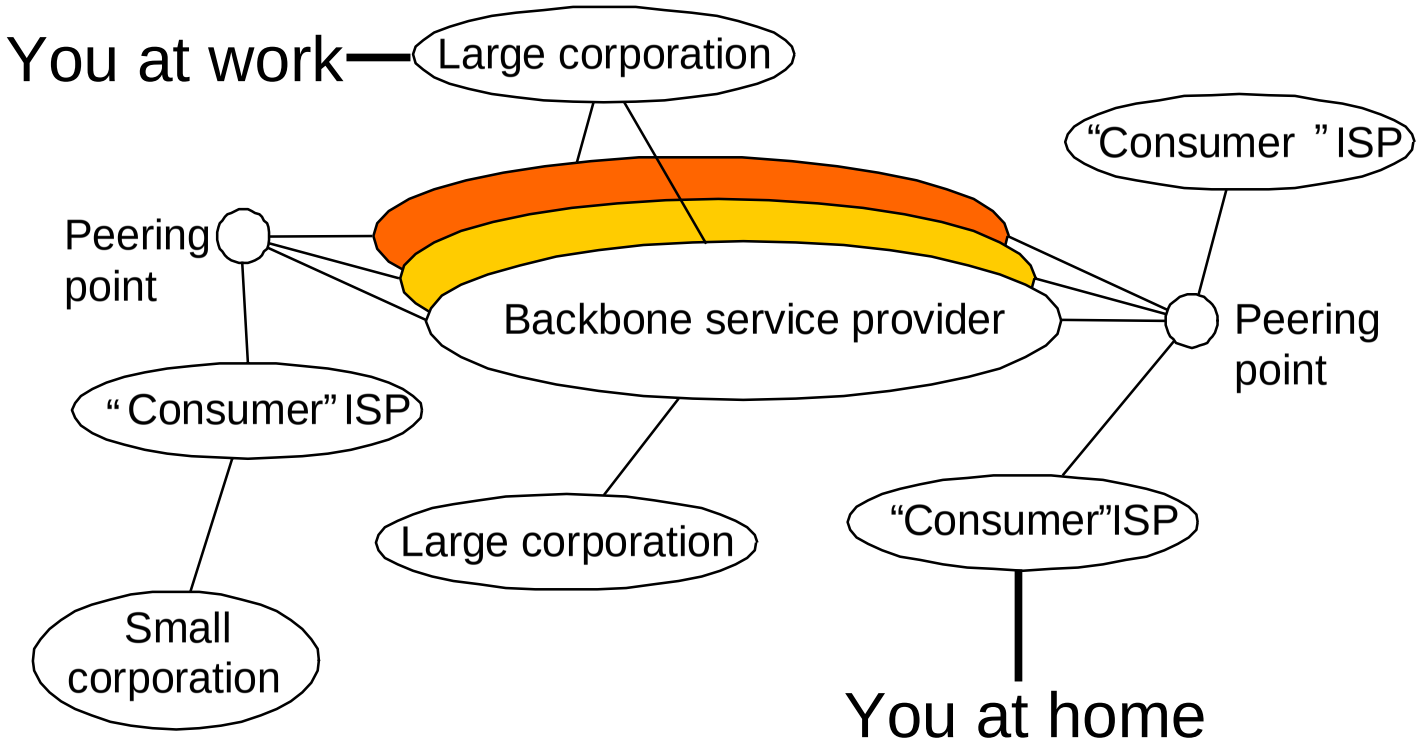
- Focus
  - How do we make routing scale?
- Topics
  - IP Addressing (ARP, DHCP, CIDR, subnets)
  - Inter-domain routing (EGP, BGP)



Scalability Concerns

- Routing burden grows with size of an internetwork
  - Size of routing tables
  - Volume or routing messages
  - Amount of routing computation
- RIP/OSPF do not scale to the size of the Internet
- We must apply further techniques:
  - Careful address allocation
  - Use of hierarchy
  - Route aggregation

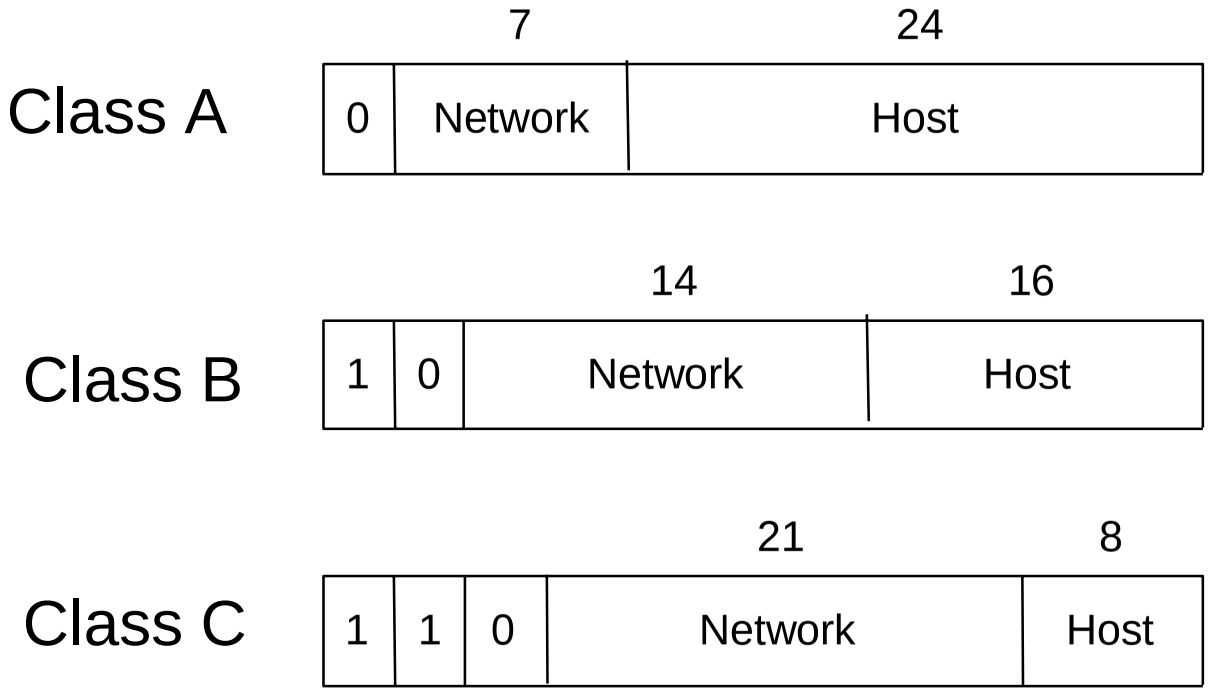
Structure of the Internet



IP Addresses

- Reflect location in topology; used for scalable routing
  - Unlike “flat” Ethernet addresses
- Interfaces on same network share prefix
  - Prefix administratively assigned (IANA or ISP)
  - Addresses globally unique
- Routing only advertises entire networks
  - Local delivery within a single “network” doesn’t involve router
  - (will make “network” precise later on)

IPv4 Address Formats



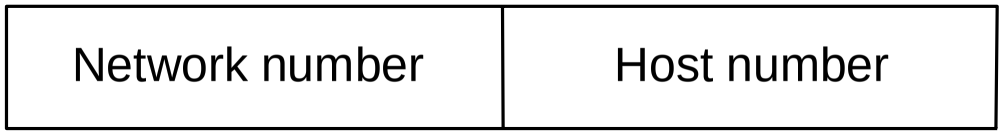
- 32 bits written in “dotted quad” notation, e.g., 18.31.0.135

IPv6 Address Format

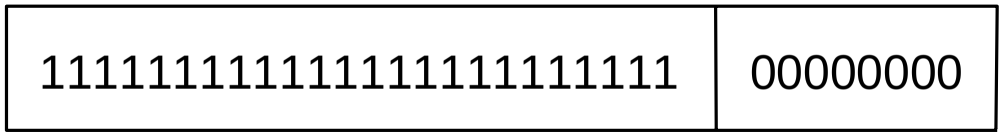
- 128 bits written in 16 bit hexadecimal chunks
- Still hierarchical, just more levels

Subnetting

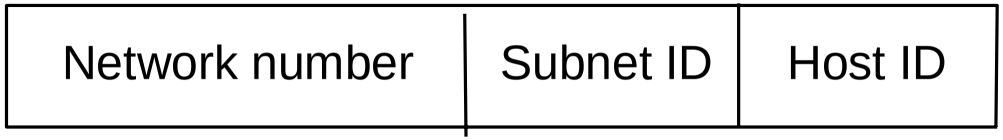
- Split up one network number into multiple physical networks
- Internal structure isn't propagated
- Helps allocation efficiency



Class B address

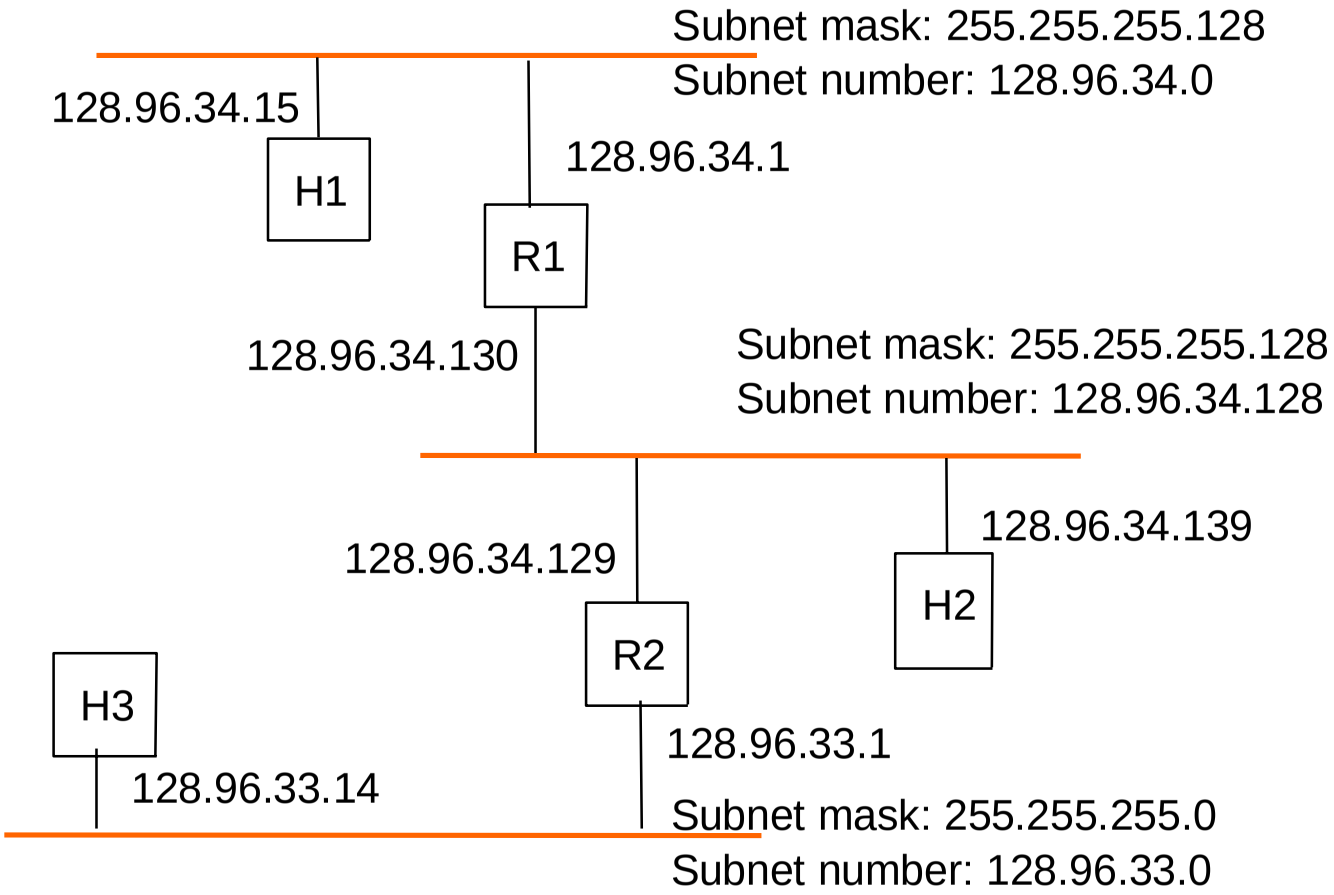


Subnet mask (255.255.255.0)



Subnetted address

Subnet Example



Getting an IP address

- Old fashioned way: sysadmin configured each machine
- Dynamic Host Configuration Protocol (DHCP)
  - One DHCP server with the bootstrap info
    - Host address, gateway address, subnet mask, ...
    - Find it using broadcast
  - Addresses may be leased; renew periodically
- "Stateless" Autoconfiguration (in IPv6)
  - Get rid of server – reuse Ethernet addresses for lower portion of address (uniqueness) and learn higher portion from routers

Address Resolution Protocol (ARP)

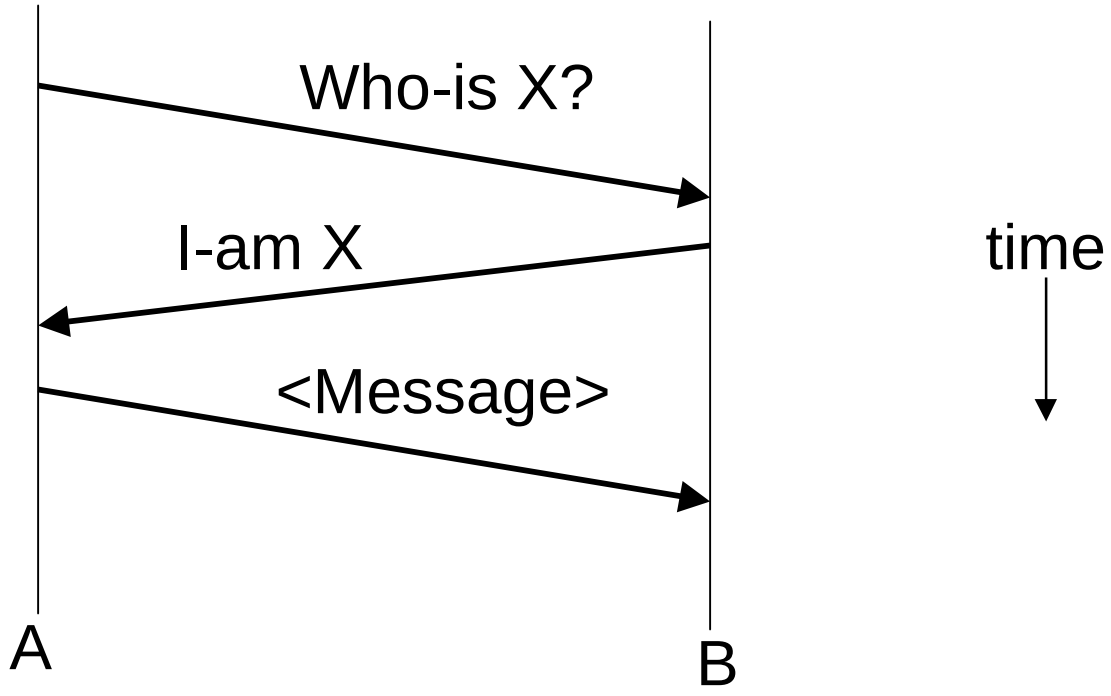
- On a single link, need Ethernet addresses to send a frame … source is a given, but what about destination?
  - Requires mapping from IP to MAC addresses

- ARP is a dynamic approach to learn mapping
  - Node A sends broadcast query for IP address X
  - Node B with IP address X replies with its MAC address M
  - A caches (X, M); old information is timed out (~15 mins)

  - Also: B caches A's MAC and IP addresses, other nodes refresh

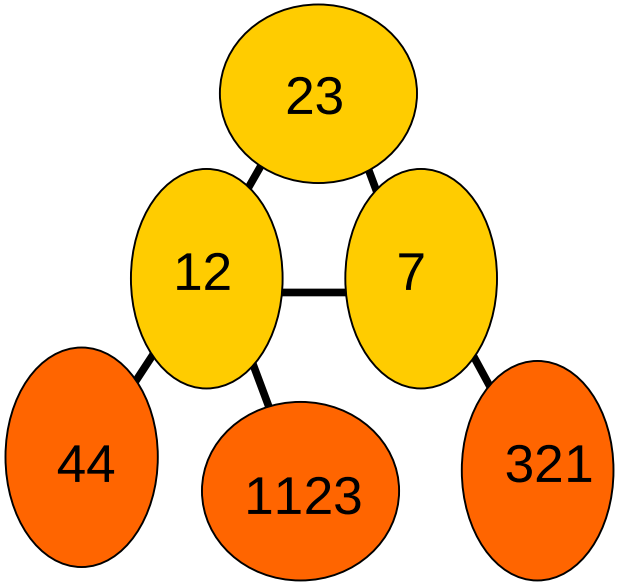
ARP Example

- To send a message … common case doesn't require ARP



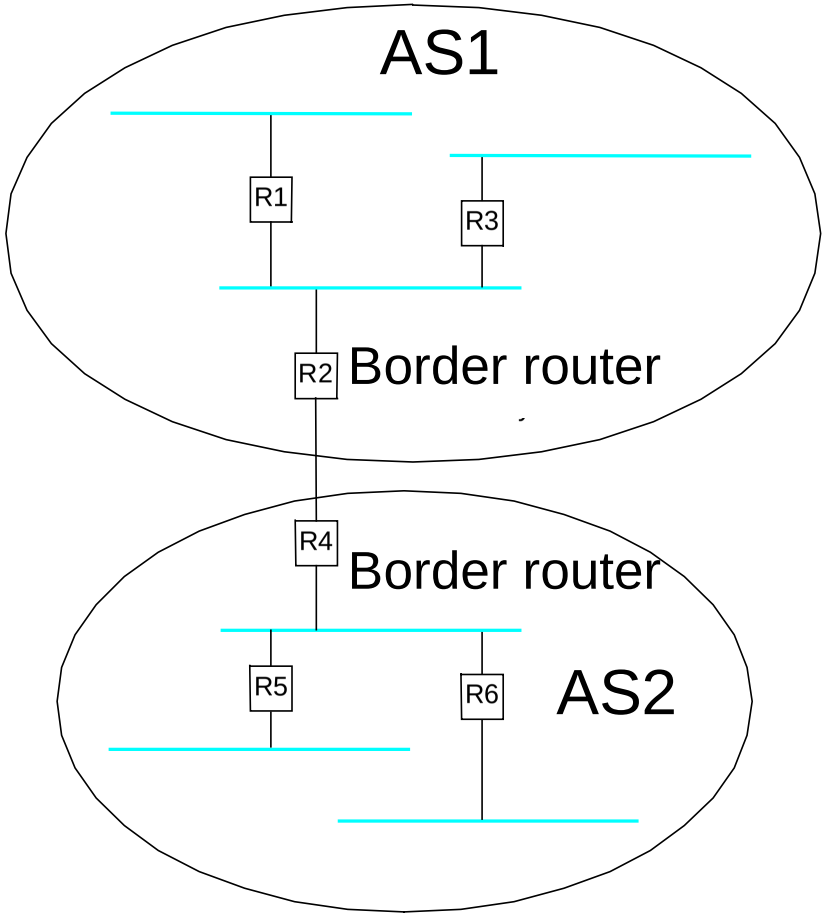
Inter-Domain Routing

- Network comprised of many Autonomous Systems (ASes) or domains
- To scale, use hierarchy: separate inter-domain and intra-domain routing
- Also called interior vs exterior gateway protocols (IGP/EGP)
  - IGP = RIP, OSPF
  - EGP = EGP, BGP



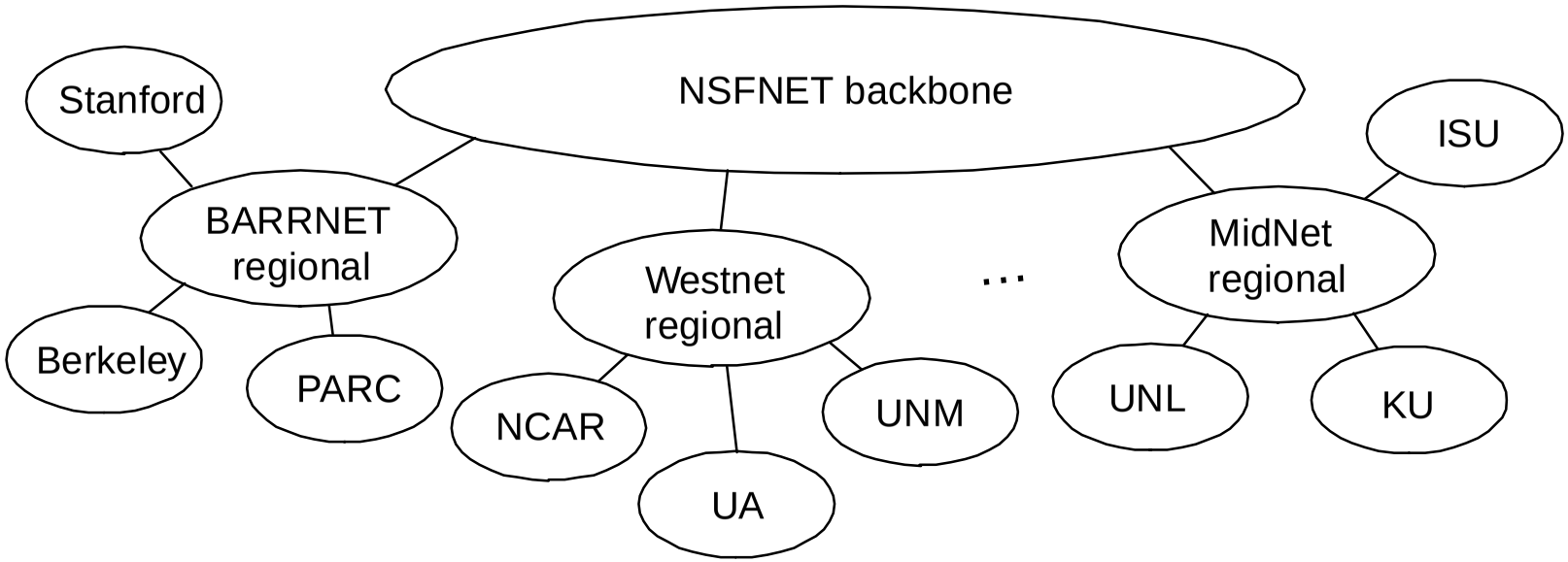
Inter-Domain Routing

- Border routers summarize and advertise internal routes to external neighbors and vice-versa
- Border routers apply policy
- Internal routers can use notion of default routes
- Core is “default-free”; routers must have a route to all networks in the world



Exterior Gateway Protocol (EGP)

- First major inter-domain routing protocol
- Constrained Internet to tree structure; no longer in use



Border Gateway Protocol (BGP-4)

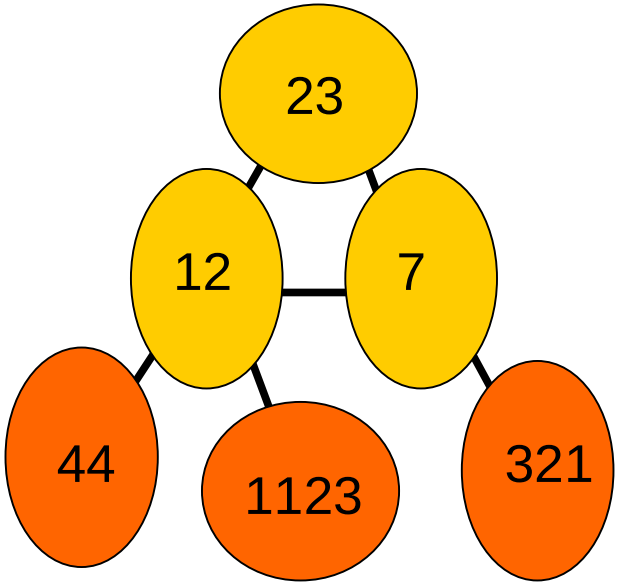
- EGP used in the Internet backbone today

- Features:
  - Path vector routing
  - Application of policy
  - Operates over reliable transport (TCP)
  - Works with CIDR

Path Vectors

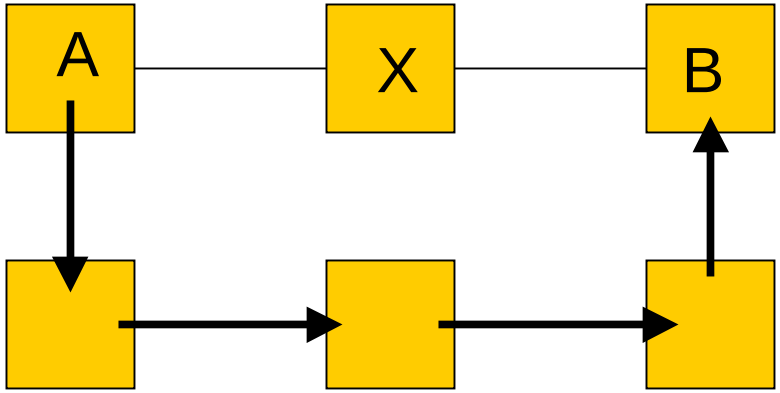
- Similar to distance vector, except send entire paths
  - e.g. 321 hears [7,12,44]
  - stronger avoidance of loops
  - supports policies (later)

- Modulo policy, shorter paths are chosen in preference to longer ones
- Reachability only – no metrics



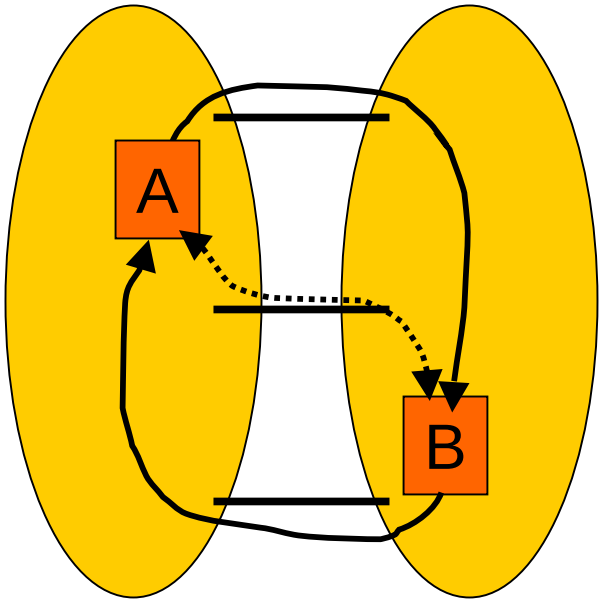
Policies

- Choice of routes may depend on owner, cost, AUP, ...
  - Business considerations
- Local policy dictates what route will be chosen and what routes will be advertised!
  - e.g., X doesn't provide transit for B, or A prefers not to use X



Impact of Policies – Example

- Early Exit / Hot Potato
  - "if it's not for you, bail"
- Combination of best local policies not globally best
- Side-effect: asymmetry



Operation over TCP

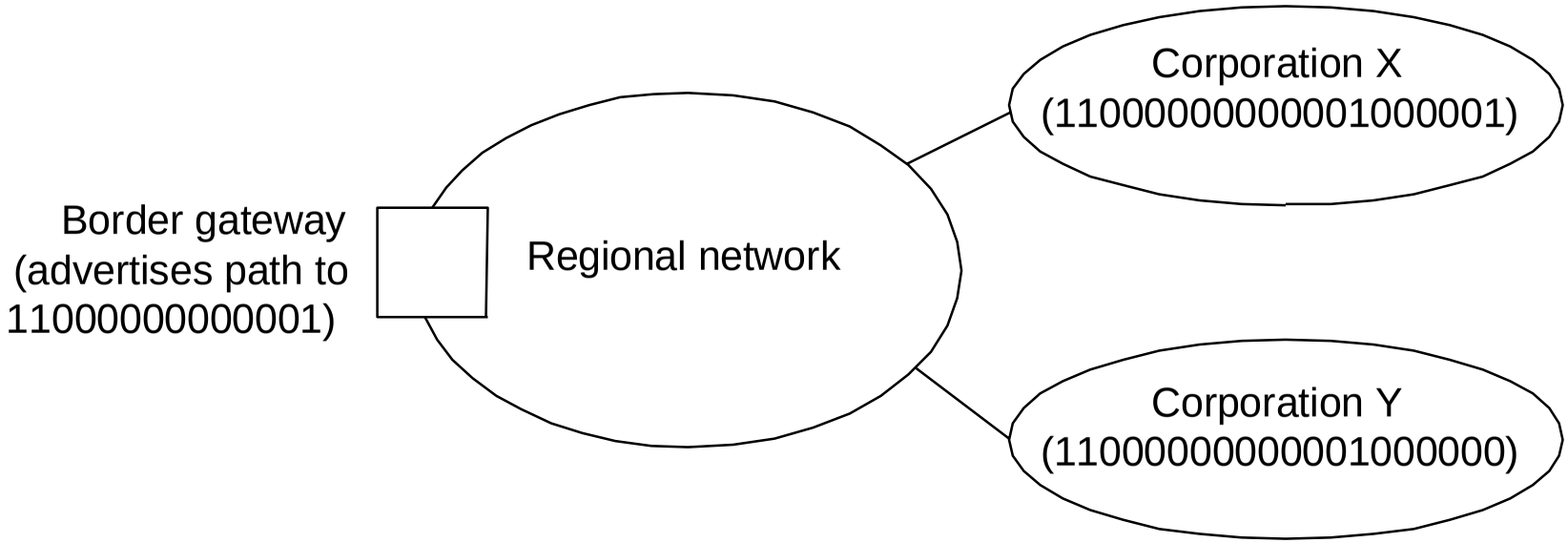
- Most routing protocols operate over UDP/IP
- BGP uses TCP
  - TCP handles error control; reacts to congestion
  - Allows for incremental updates
- Issue: Data vs. Control plane
  - Shouldn't routing messages be higher priority than data?

CIDR (Supernetting)

- CIDR = Classless Inter-Domain Routing
- Aggregate advertised network routes
  - e.g., ISP has class C addresses 192.4.16 through 192.4.31
  - Really like one larger 20 bit address class ...
  - Advertise as such (network number, prefix length)
  - Reduces size of routing tables
- But IP forwarding is more involved
  - Based on Longest Matching Prefix operation

CIDR Example

- X and Y routes can be aggregated



Key Concepts

- Scalable routing is requires that we minimize:
  - Size of routing tables
  - Exchange of routing messages
- Scalable routing is based on:
  - Careful address allocation
  - Use of hierarchy and route aggregation
- Internet is a collection of Autonomous Systems (ASes)
  - Policy dominates routing at the AS level