

**CSE 461  
Module 9**

**Aggregation & Hierarchy  
(& Inter-domain Routing)**

---

John Zahorjan  
zahorjan@cs.washington.edu

**This Lecture**

---

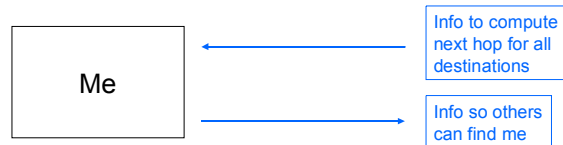
- Focus
  - How do we make routing scale?
- Approaches
  - Aggregating
    - Reduce the amount others need to know
  - Hierarchy
    - Reduce the amount I need to know
- Inter-domain routing
  - ASes and BGP

Application
Presentation
Session
Transport
<b>Network</b>
Data Link
Physical

## Preliminaries

---

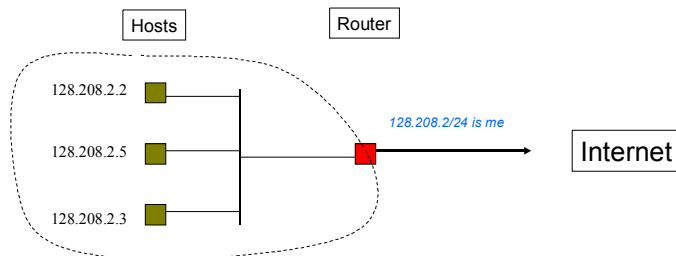
- Basic issue is how much information is required to effect routing
  - To scale, we want to be able to control it, at the least



## Aggregation

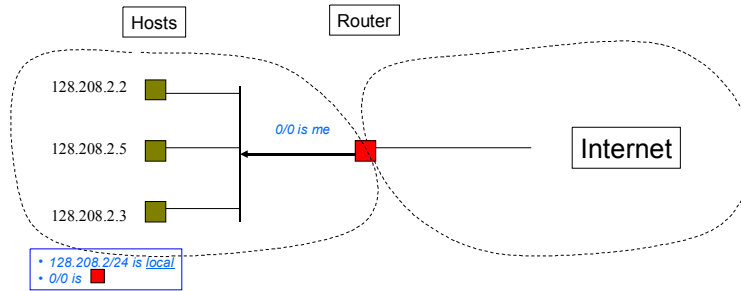
---

- We've already seen an example: forwarding tables index networks, not individual hosts



## Hierarchy

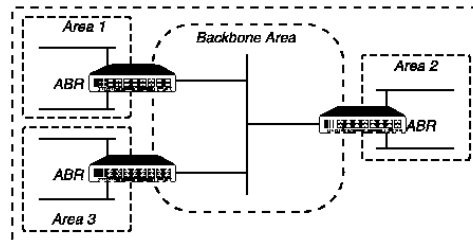
- We've already seen an example: host gateways



CSE 461 09wi

m9.5

## Generalizing: Routing Areas



- Routers within an area (only) exchange full link state information
  - Limit cost of link state traffic / computation
  - (Different areas could have different cost metrics)
- Area border routers (ABRs) summarize area to other ABRs
- ABRs summarize rest of world to an area
- (Areas can have more than one ABR.)

CSE 461 09wi

m9.6

## Inter-domain routing

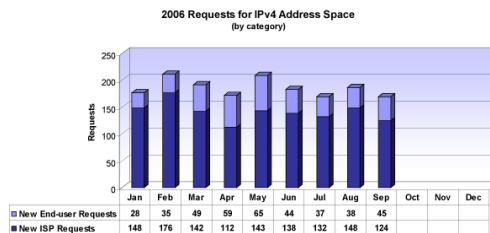
---

- A *domain* is an administrative entity
  - A corporation, a university, ...
- Synonym: *autonomous system* (AS)
- AS's are the basic building block of the Internet
  - AS's have id's (because we need to be able to name them, as we'll see)
- IP address space assignment is largely hierarchical
  - The Internet Assigned Numbers Authority owns everything
  - It assigns blocks of addresses to Regional Internet Registries (RIRs)
  - They assign to ISPs (reallocators) and end-users (non-reallocators)

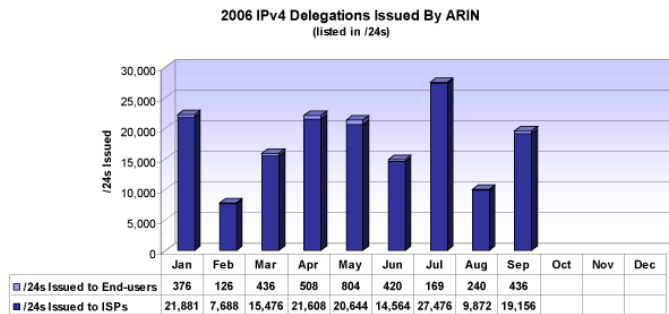
## Example: IANA ⇒ ARIN ⇒ ...

---

(ARIN = American Registry for Internet Numbers)



## Example (cont.)

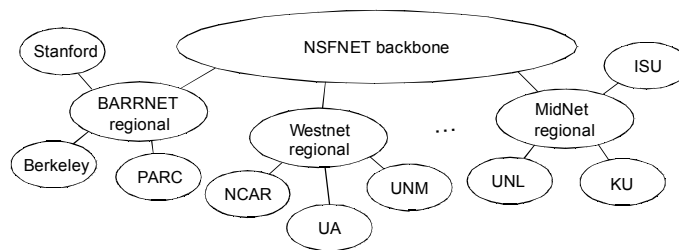


CSE 461 09wi

m9.9

## Original Structure of the Internet

- Like address assignment: hierarchical



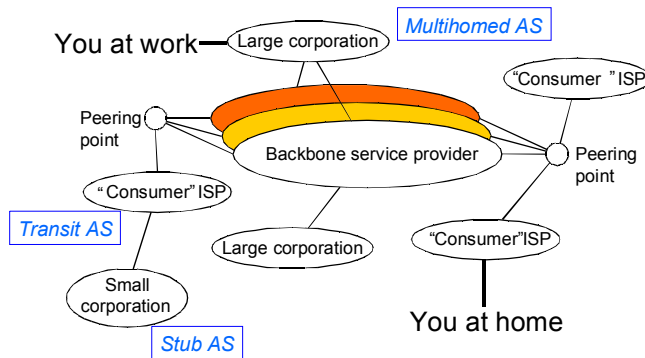
- What's "wrong" with this?

CSE 461 09wi

m9.10

## Current Structure

- Inter-domain versus intra-domain routing

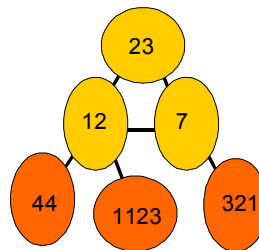


CSE 461 09wi

m9.11

## 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



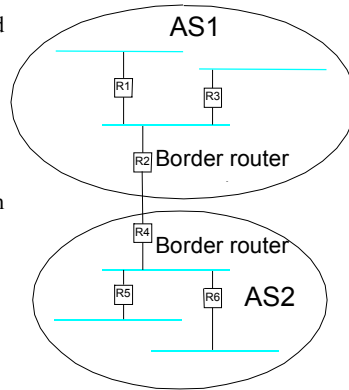
CSE 461 09wi

m9.12

## 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



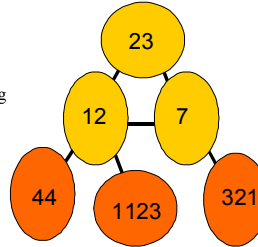
## Border Gateway Protocol (BGP-4)

---

- BGP used in the Internet backbone today
- Features:
  - Path vector routing
  - Application of policy
  - Operates over reliable transport (TCP)
  - Uses route aggregation (CIDR)

## Path Vectors

- Similar to distance vector, except send entire paths
  - reachability only; no metrics (but AS hop count)
  - e.g., 7 hears [12,44], advertises [7,12,44] to 321
    - No requirement to advertise to everyone
  - strong avoidance of loops
- AS can choose whatever path it wants for forwarding
- No information about internal networks exchanged
- Goal: support (business) policies
- Modulo policy, shorter paths are chosen in preference to longer ones

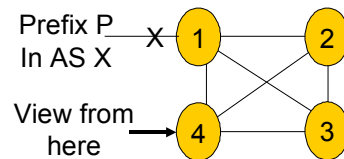


CSE 461 09wi

m9.15

## An Ironic Twist on Convergence

- Recently, it was realized that BGP convergence can undergo a process analogous to count-to-infinity!



- AS 4 uses path 4 1 X. A link fails and 1 withdraws 4 1 X.
- So 4 uses 4 2 1 X, which is soon withdrawn, then 4 3 2 1 X, ...
- Result is many invalid paths can be explored before convergence

CSE 461 09wi

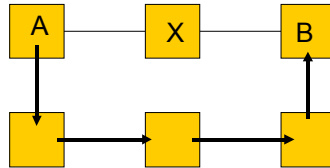
m9.16



## 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



CSE 461 09wi

m9.17

## Simplified Policy Roles

---

- Providers sell Transit to their customers
  - Customer announces path to their prefixes to providers in order for the rest of the Internet to reach their prefixes
  - Providers announces path to all other Internet prefixes to customer C in order for C to reach the rest of the Internet
- Additionally, parties Peer for mutual benefit
  - Peers A and B announce path to their customer's prefixes to each other but do not propagate announcements further
  - Peering relationships aren't transitive
  - Tier 1s peer to provide global reachability

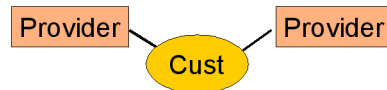
CSE 461 09wi

m9.18

## Multi-Homing

---

- Connect to multiple providers for reliability, load sharing

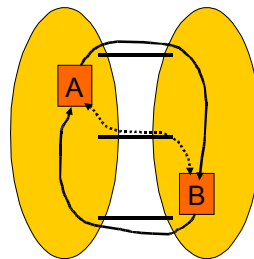


- Choose the best outgoing path to P out of any of the announcements to P that we hear from our providers
  - Easy to control outgoing traffic, e.g, for load balancing
- Advertise the possible routes to P to our providers
  - Less control over what paths other parties will use to reach us

## 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?

## Key Concepts

---

- Internet is a collection of Autonomous Systems (ASes)
  - Policy dominates routing at the AS level
- Structural hierarchy helps make routing scalable
  - BGP routes between autonomous systems (ASes)