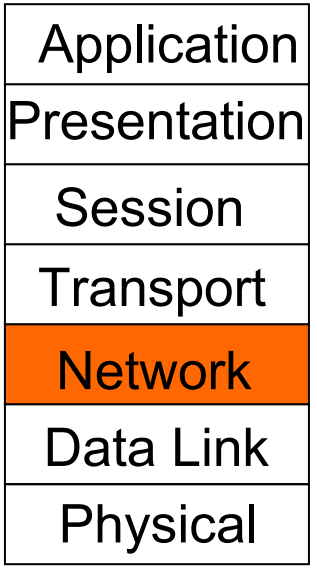
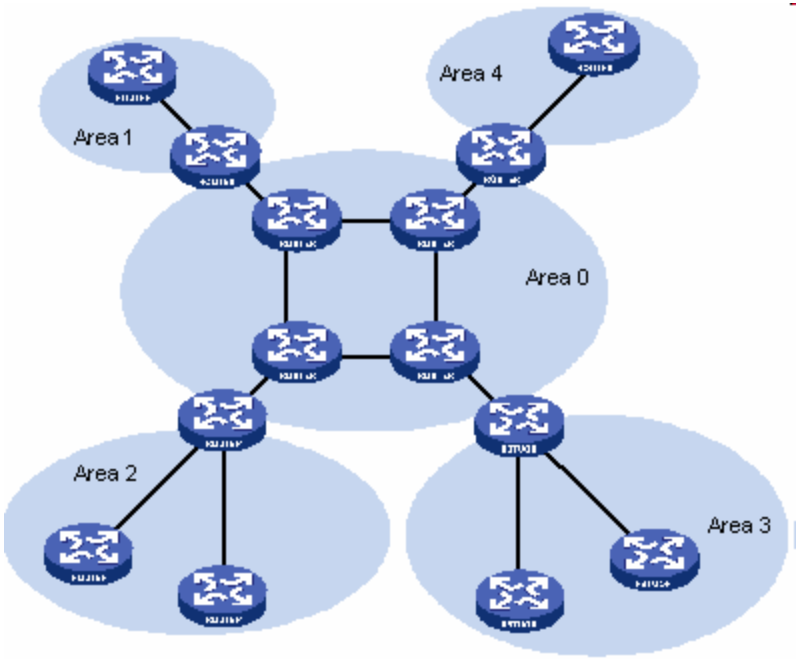
Routing Across the Interet

- Focus
  - How do we make routing scale?
  - We already saw IP address prefixes
- Inter-domain routing
  - Uses hierarchy and aggregation
  - ASes and BGP



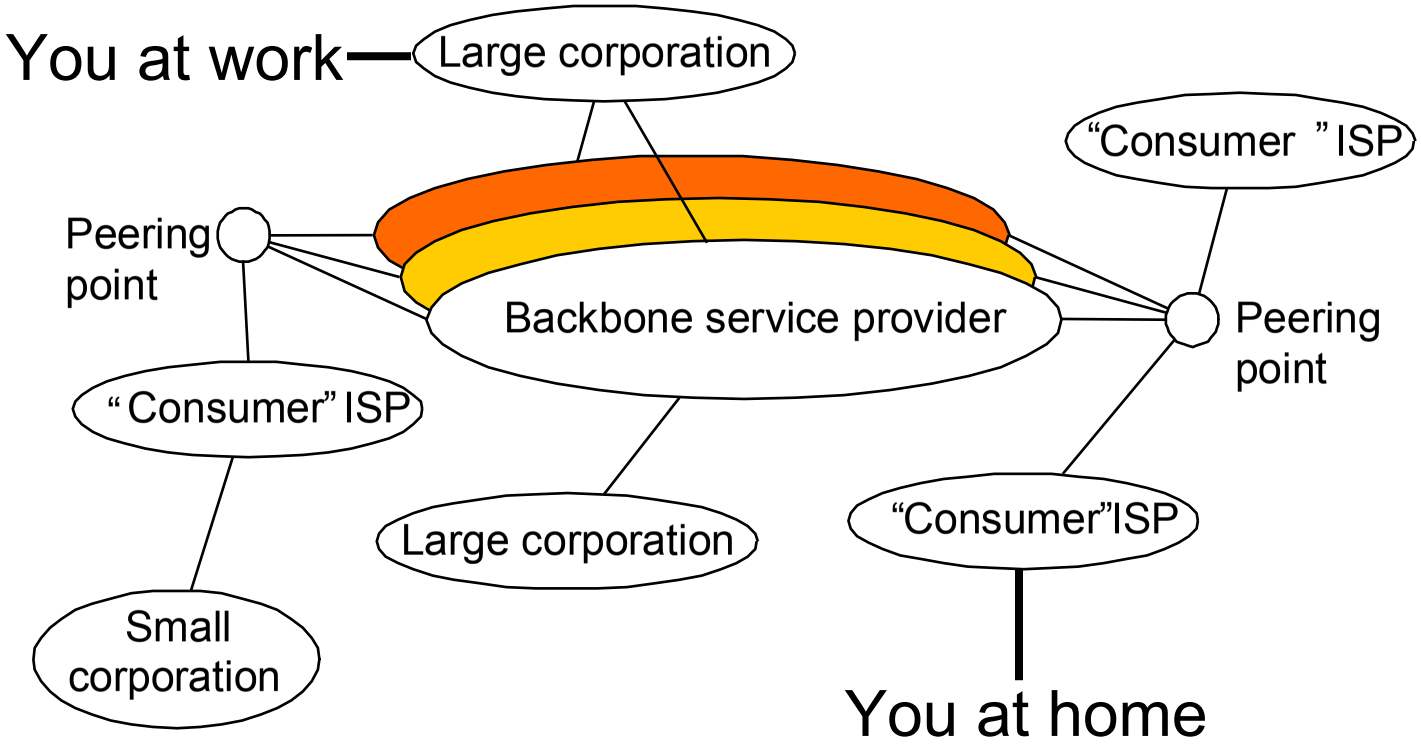
Scaling OSPF with Areas

- Split a large network into "areas"
  - Areas connect via border routers
  - Backbone area connects to all
- Border routers send a summary of the area routes to other areas
  - Hides internal area detail
- Example of applying hierarchy



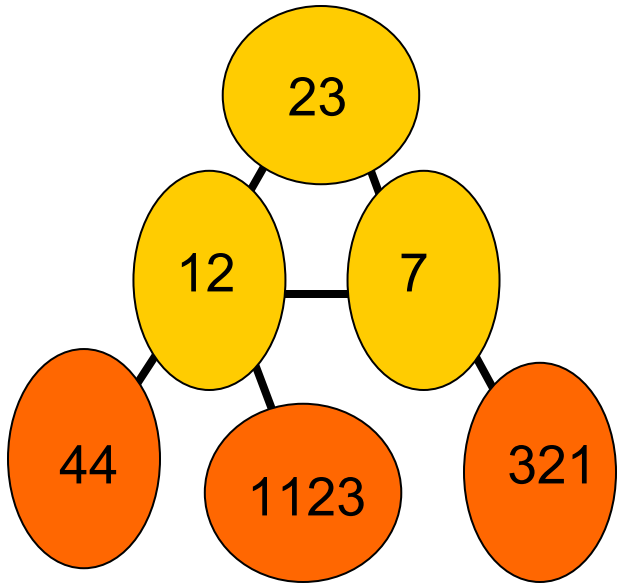
Structure of the Internet

- Inter-domain versus intra-domain routing



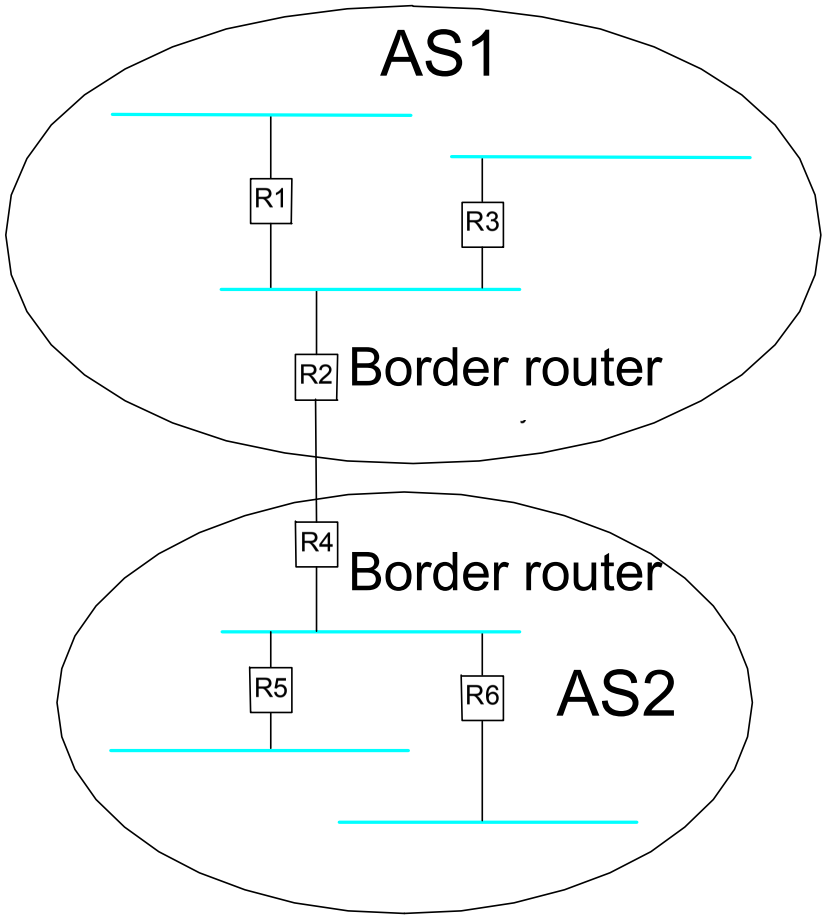
Inter-Domain Routing

- 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



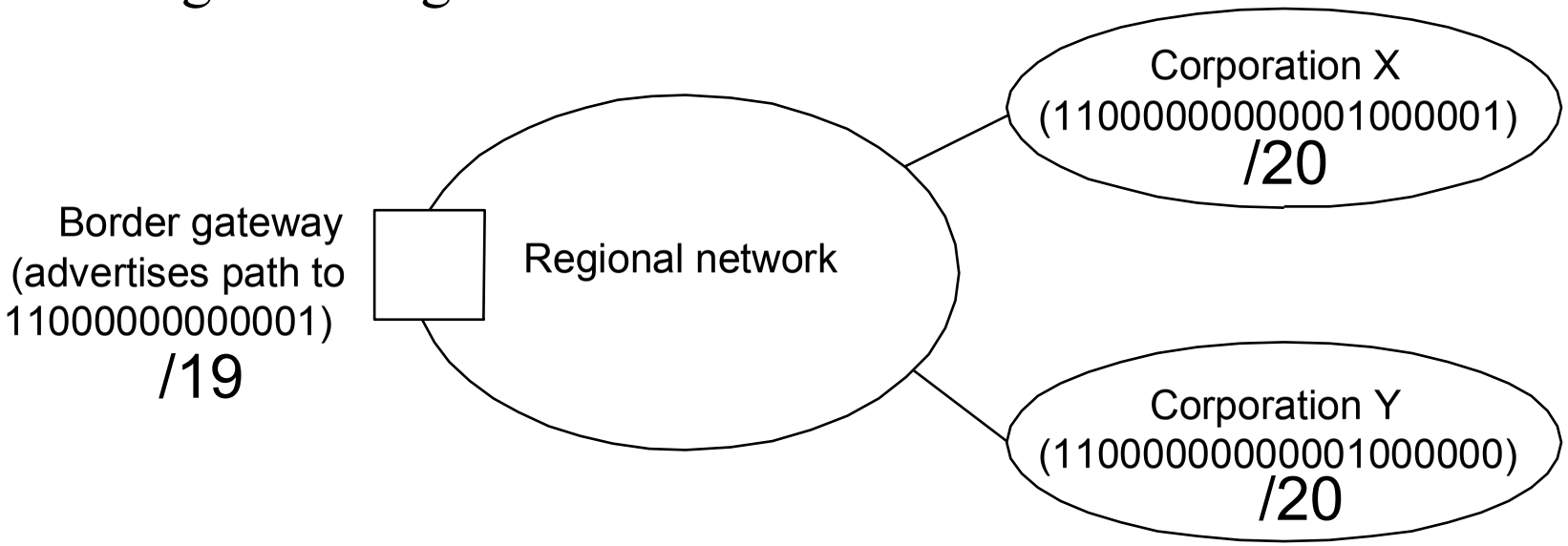
Inter-Domain Routing

- 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



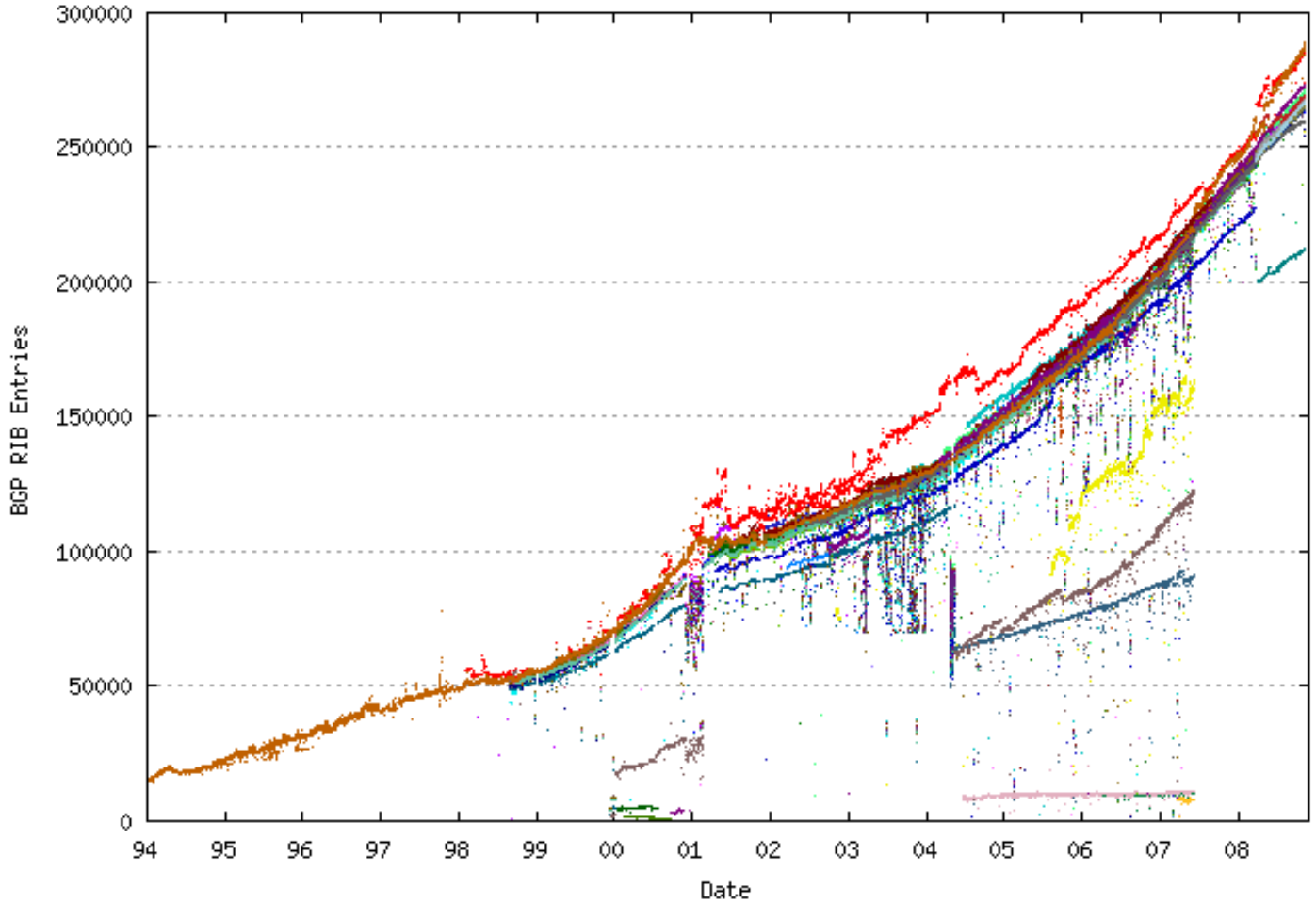
CIDR Example

- X and Y routes can be aggregated because they form a bigger contiguous range.



- But aggregation isn't always possible. Why?

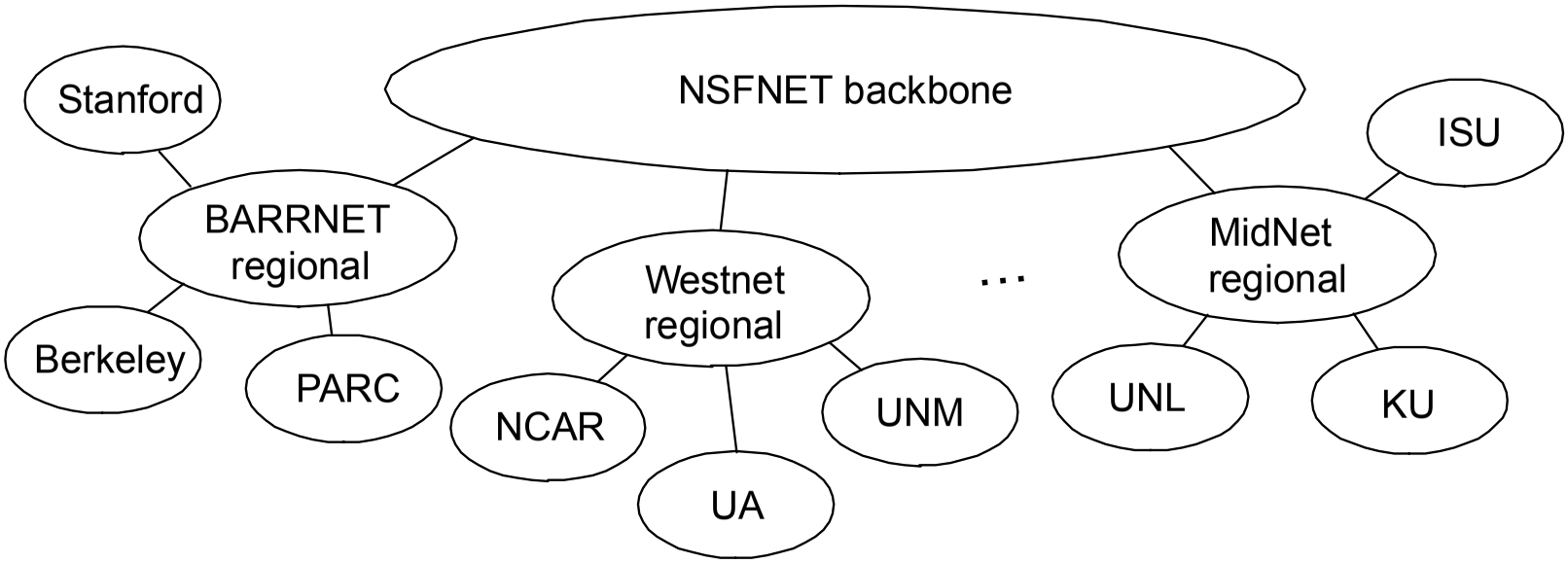
Core BGP Table Growth 1994 - 2008



www.cidr-report.org November 2008

Exterior Gateway Protocol (EGP)

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



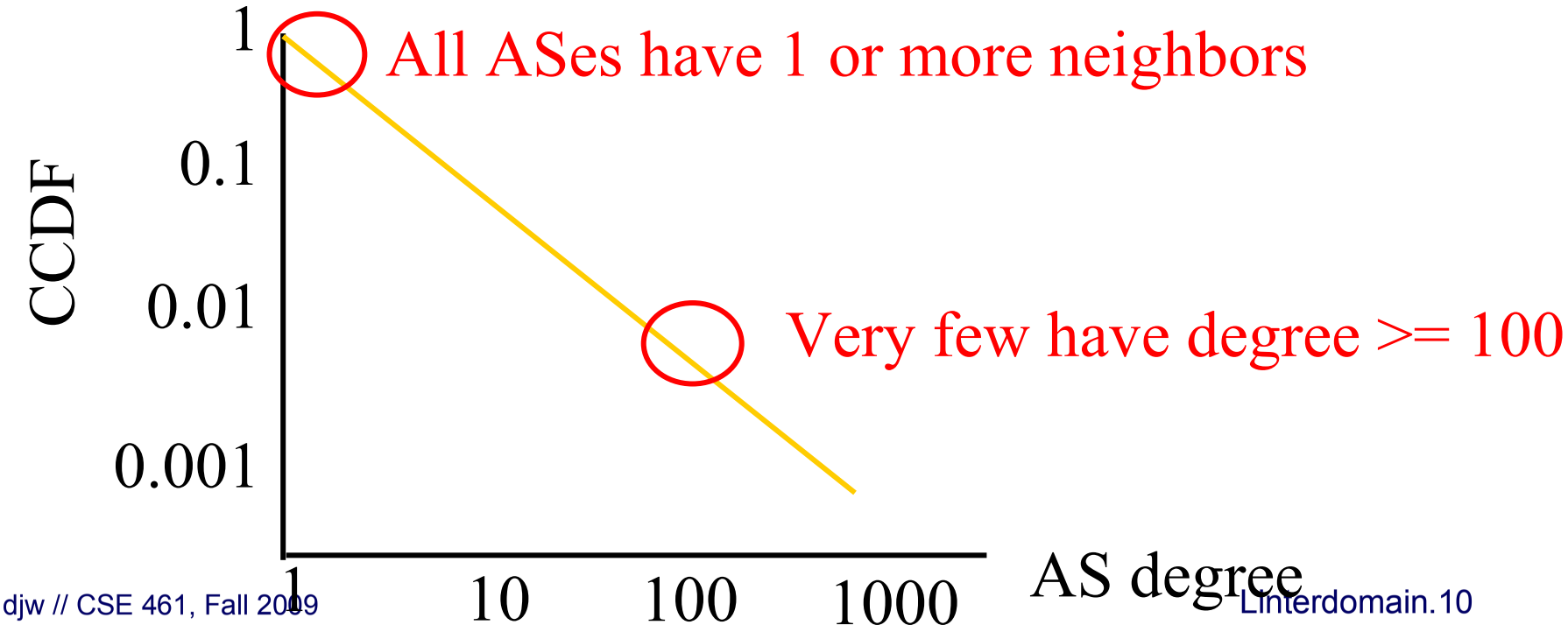
Border Gateway Protocol (BGP-4)

- EGP used in the Internet backbone today

- Features:
  - Path vector routing
  - Application of policy
  - Operates over reliable transport (TCP)
  - Uses route aggregation (CIDR)

Measurements of the AS Graph

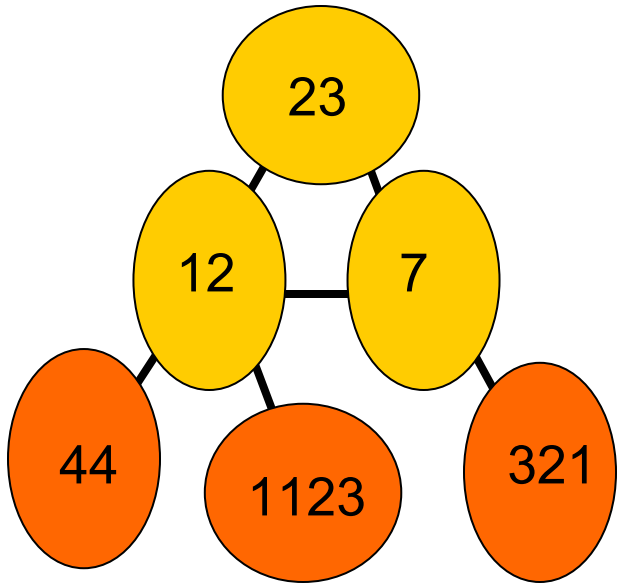
- AS graph structure
  - High variability in node degree ("power law")
  - A few very highly-connected ASes
  - Many ASes have only a few connections



Path Vectors

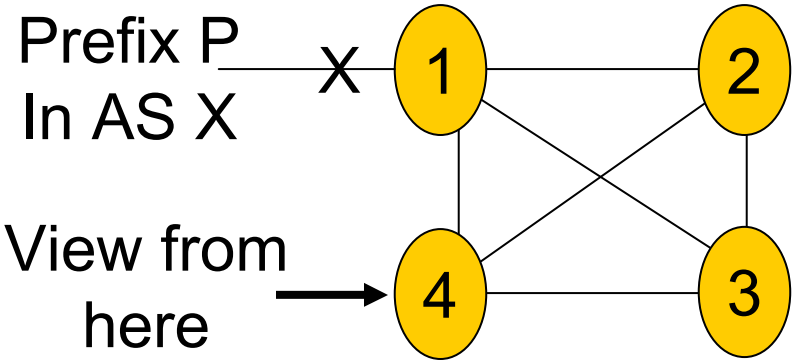
- Similar to distance vector, except send entire paths
  - e.g. 321 hears [7,12,44]
  - stronger avoidance of loops
  - supports policies (later)

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



An Ironic Twist on Convergence

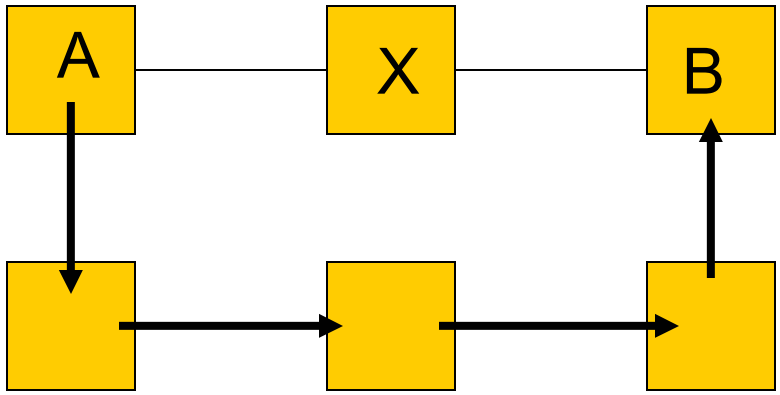
- Recently, it was realized that BGP convergence can undergo a process analogous to count-to-infinity!



- 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

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

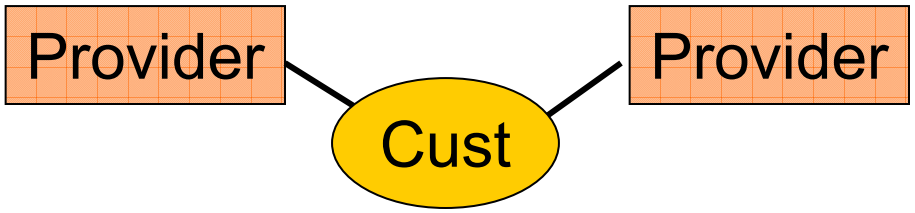


Simplified Policy Roles

- 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

Multi-Homing

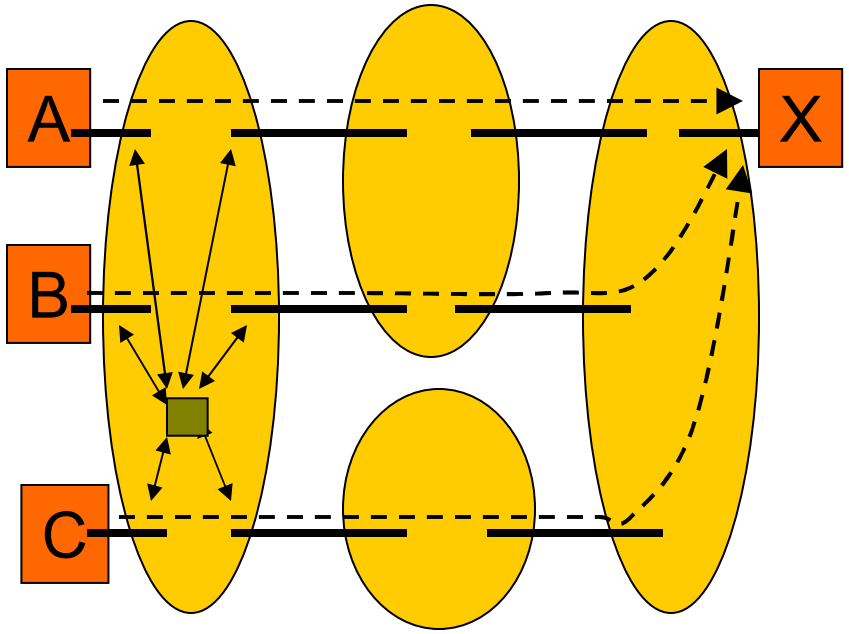
- Connect to multiple providers for reliability, load sharing



- 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

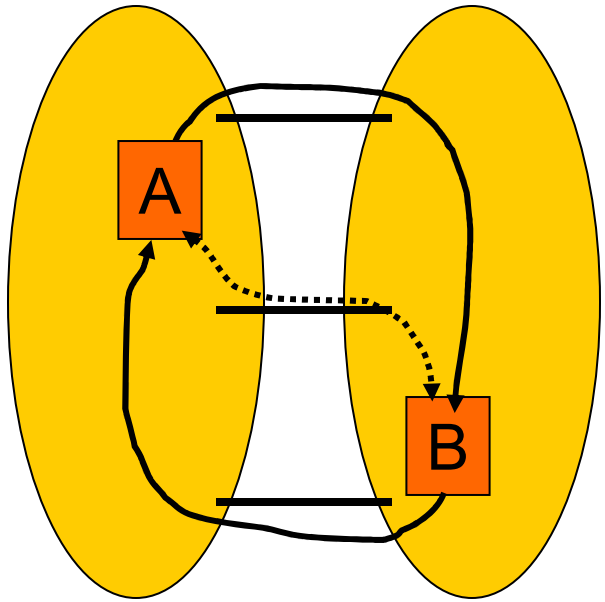
Integration of Intra- and Inter-domain

- Each location makes its own decision based on the routes it sees; not one decision per ISP.
  - e.g, paths A→X, B→X, C→X
- Externally facing routers at exchanges hear routes from other ISPs
- Internally, routers share the routes they know about (green box).



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?