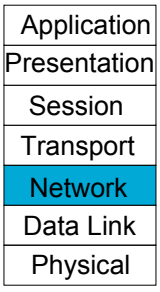
IP Addressing & Interdomain Routing

Next Topic

- IP Addressing
  - Hierarchy (prefixes, class A, B, C, subnets)
- Interdomain routing



Scalability Concerns

- Routing burden grows with size of an inter-network
  - Size of routing tables
  - Volume of routing messages
  - Amount of routing computation
- To scale to the size of the Internet, apply:
  - Hierarchical addressing
  - Route aggregation

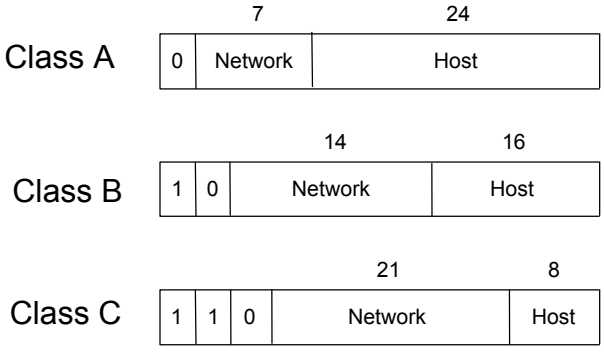
IP Addresses

- 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 by prefix
  - Local delivery in a single "network" doesn't involve router

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

IPv4 Address Formats



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

IPv6 Address Format

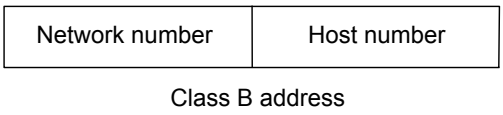
- 128 bits written in 16 bit hexadecimal chunks
- Still hierarchical, just more levels

Updated Forwarding Routine

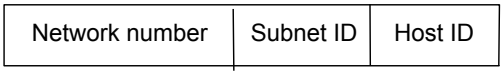
- Used to be "look up destination address for next hop"
- Now addresses have network and host portions:
  - Source host:
    - if destination network is the same as the host network, then deliver locally (without router)
    - Otherwise send to the router
  - Intermediate router:
    - look up destination network in routing table to find next hop and send to next router.
    - If destination network is directly attached then deliver locally.
- (Note that it will get a little more complicated later)

Subnetting – More Hierarchy

- Split up one network number into multiple physical networks
- Helps allocation efficiency -- can hand out subnets
- Rest of internet does not see subnet structure
  - subnet is purely internal to network
  - aggregates routing info

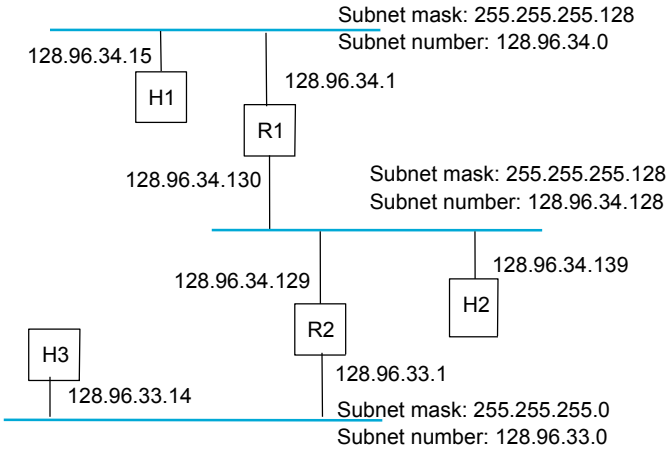


Class B address



Subnetted address

Subnet Example

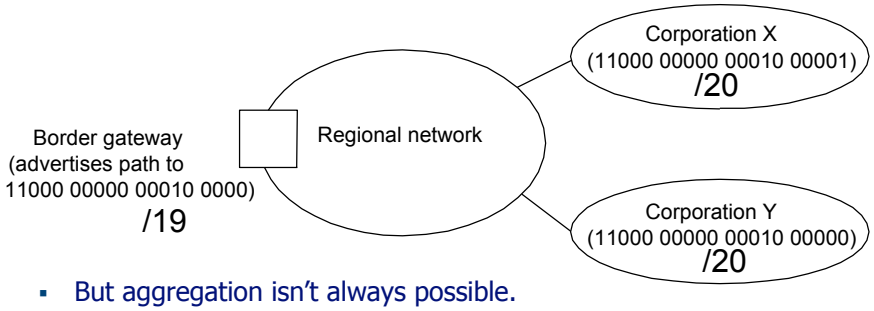


CIDR (Supernetting)

- CIDR = Classless Inter-Domain Routing
- Generalize class A, B, C into prefixes of arbitrary length; now must carry prefix length with address
- Aggregate adjacent 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 because they form a bigger contiguous range.



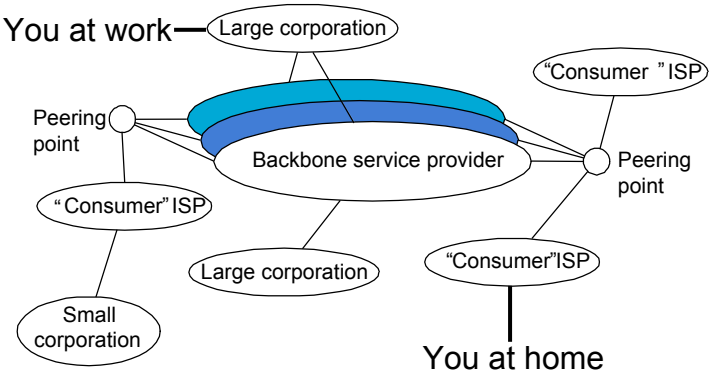
- But aggregation isn't always possible.
  - can only aggregate power of 2

IP Forwarding Revisited

- Routing table now contains routes to "prefixes"
  - IP address and length indicating what bits are fixed
- Now need to "search" routing table for longest matching prefix, only at routers
  - Search routing table for the prefix that the destination belongs to, and use that to forward as before
  - There can be multiple matches; take the longest prefix
- This is the IP forwarding routine used at routers.

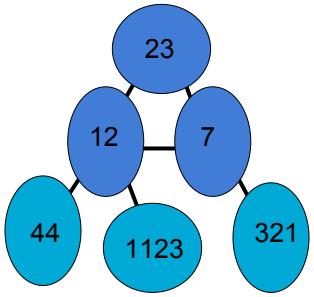
Structure of the Internet

- Inter-domain versus intra-domain routing



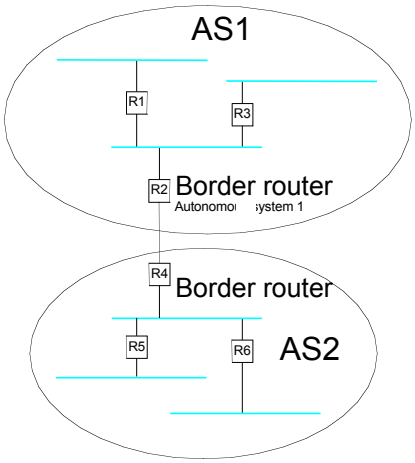
Inter-Domain Routing

- 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

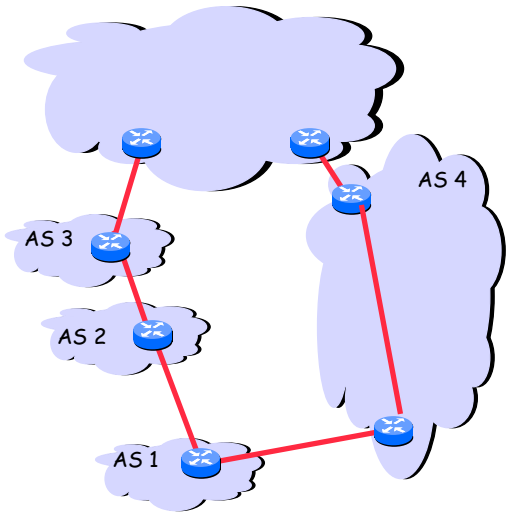


Inter-Domain Routing

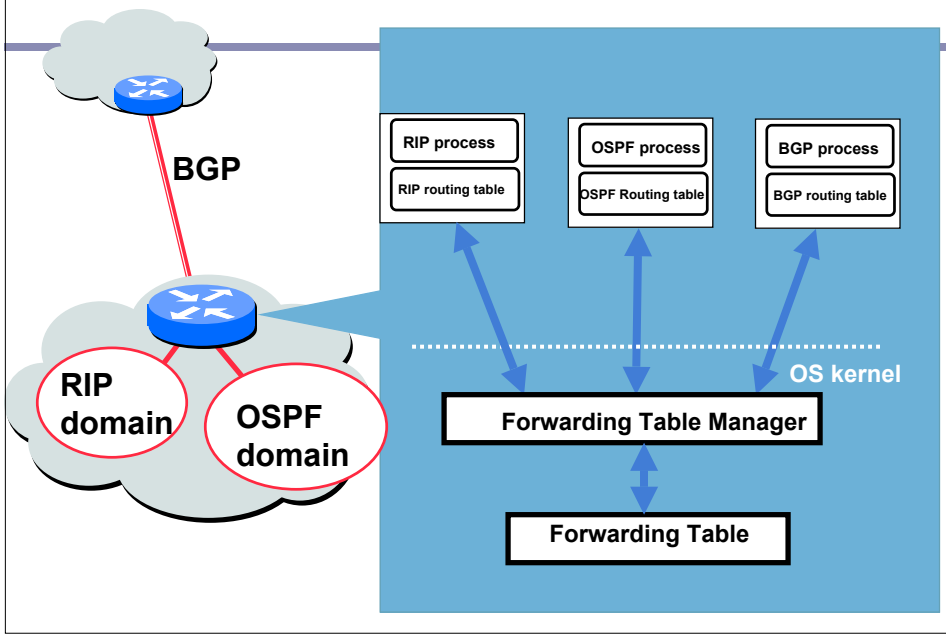
- Border routers summarize and advertise internal routes to external neighbors and vice-versa
- Border routers apply policy
- Internal routers can use notion of default routes
- Core is "default-free"; routers must have a route to all networks in the world



Hierarchical Routing May Pay a Price for Path Quality



Many Routing Processes Can Run on a Single Router



Border Gateway Protocol (BGP-4)

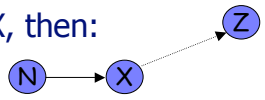
- Features:
  - Path vector routing
  - Application of policy
  - Operates over reliable transport (TCP)
  - Uses route aggregation (CIDR)

Internet Interdomain Routing: BGP

- BGP (Border Gateway Protocol): *the* de facto standard
- **Path Vector** protocol:
  - similar to Distance Vector protocol
  - a border gateway sends to a neighbor *entire path* (i.e., a sequence of ASes) to a destination, e.g.,
    - gateway X sends to neighbor N its path to dest. Z:

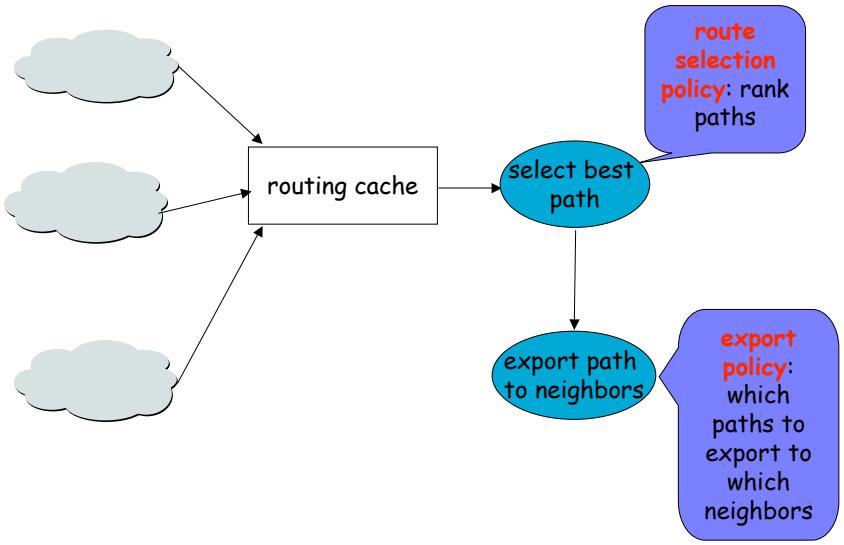
      path (X,Z) = X,Y1,Y2,Y3,…,Z
  - if N selects path(X, Z) advertised by X, then:

    path (N,Z) = N, path (X,Z)

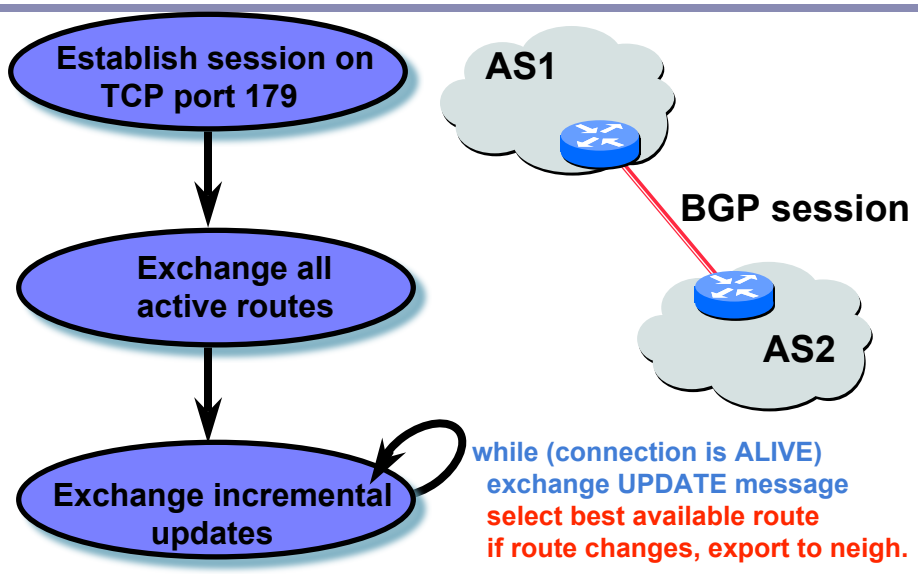


Question: what are the implications of path vector?

BGP Routing Decision Process



BGP Operations (Simplified)



BGP Messages

- Four types of messages
  - OPEN: opens TCP connection to peer and authenticates sender
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE keeps connection alive in absence of UPDATES; also ACKs OPEN request
  - NOTIFICATION: used to close connection

Internet Routing Architecture

- Divided into Autonomous Systems
  - Distinct regions of administrative control
  - Routers/links managed by a single "institution"
  - Service provider, company, university, …
- Hierarchy of Autonomous Systems
  - Large, tier-1 provider with a nationwide backbone
  - Medium-sized regional provider with smaller backbone
  - Small network run by a single company or university
- Interaction between Autonomous Systems
  - Internal topology is not shared between ASes
  - … but, neighboring ASes interact to coordinate routing

Autonomous System Numbers

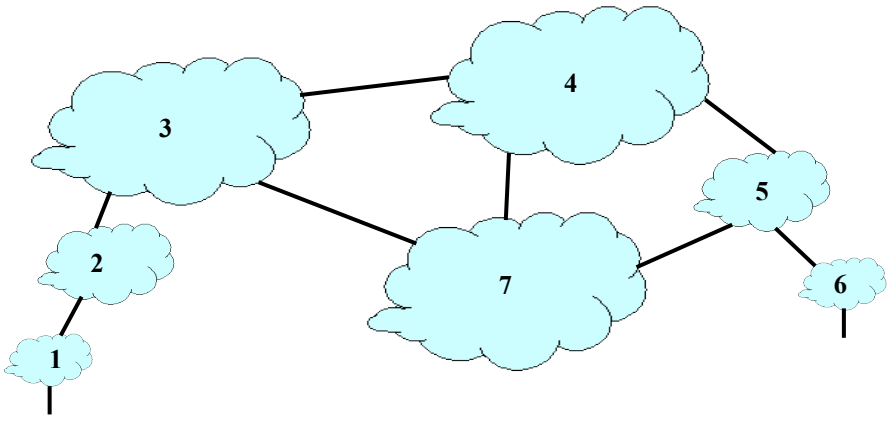
AS Numbers are 16 bit values.

Currently just over 20,000 in use.

- **Level 3: 1**
- **MIT: 3**
- **Washington AS: 73**
- **Princeton: 88**
- **AT&T: 7018, 6341, 5074, …**
- **UUNET: 701, 702, 284, 12199, …**
- **Sprint: 1239, 1240, 6211, 6242, …**
- **…**

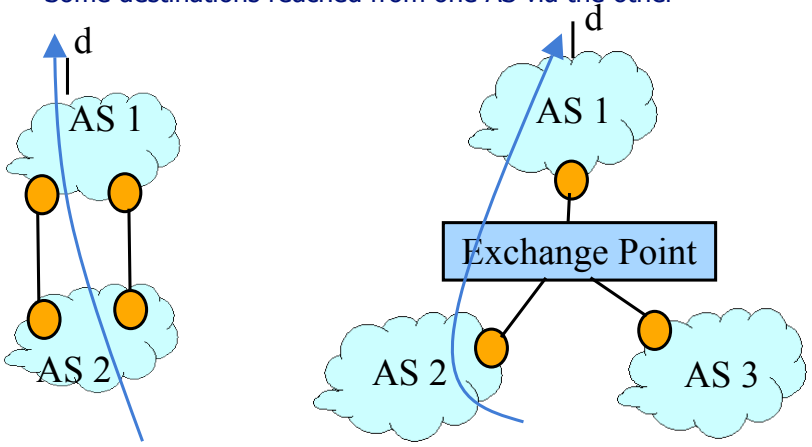
AS Topology

- Node: Autonomous System
- Edge: Two ASes that connect to each other



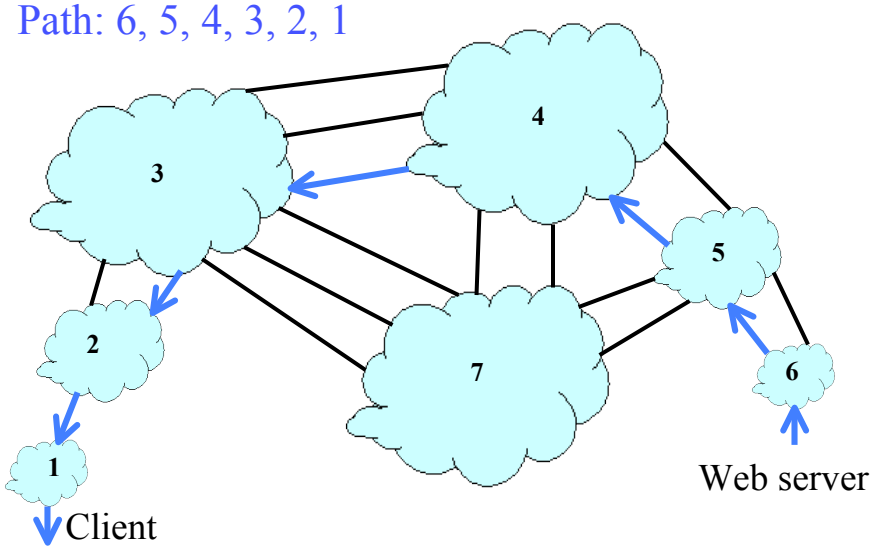
What is an Edge, Really?

- Edge in the AS graph
  - At least one connection between two ASes
  - Some destinations reached from one AS via the other



Interdomain Paths

Path: 6, 5, 4, 3, 2, 1



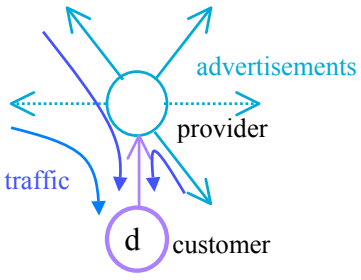
Business Relationships

- Neighboring ASes have business contracts
  - How much traffic to carry
  - Which destinations to reach
  - How much money to pay
- Common business relationships
  - Customer-provider
    - E.g., UW is a customer of NTT
    - E.g., MIT is a customer of Level 3
  - Peer-peer
    - E.g., AT&T is a peer of Sprint

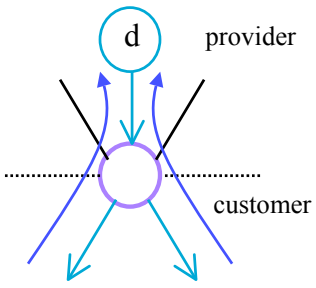
Customer-Provider Relationship

- Customer needs to be reachable from everyone
  - Provider tells all neighbors how to reach the customer
- Customer does not want to provide transit service
  - Customer does not let its providers route through it

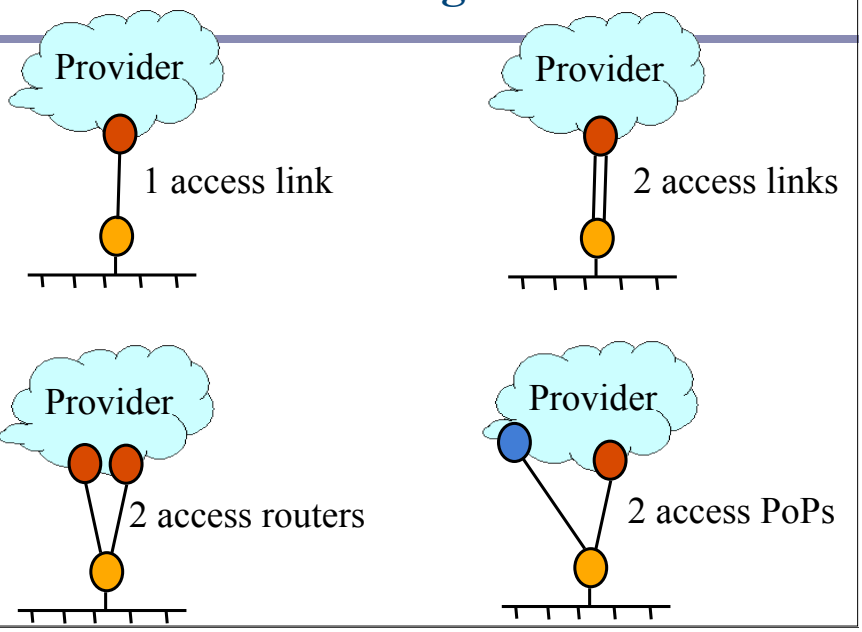
Traffic **to** the customer



Traffic **from** the customer

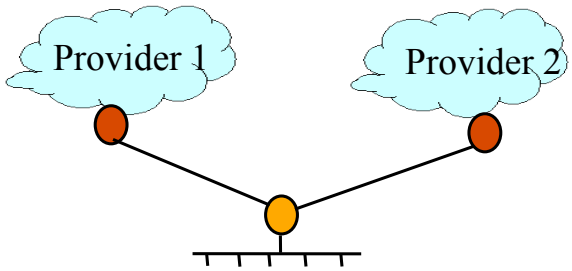


Customer Connecting to a Provider



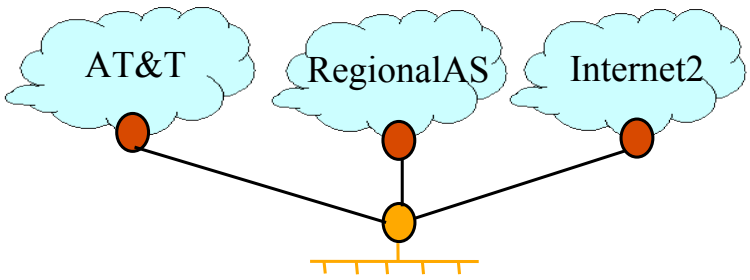
Multi-Homing: Two or More Providers

- Motivations for multi-homing
  - Extra reliability, survive single ISP failure
  - Financial leverage through competition
  - Better performance by selecting better path
  - Gaming the 95th-percentile billing model



Example

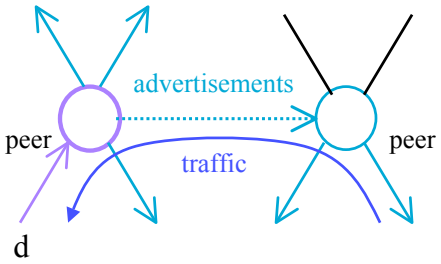
- Internet: customer of AT&T and RegionalAS
- Research universities/labs: customer of Internet2



Peer-Peer Relationship

- Peers exchange traffic between customers
  - AS exports *only* customer routes to a peer
  - AS exports a peer's routes *only* to its customers
  - Often the relationship is settlement-free (i.e., no $$$)

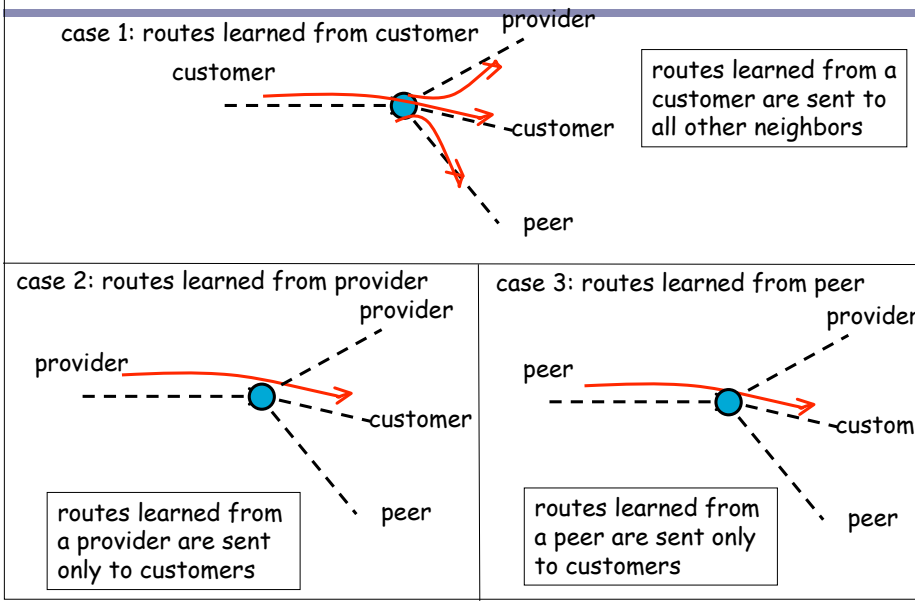
Traffic to/from the peer and its customers



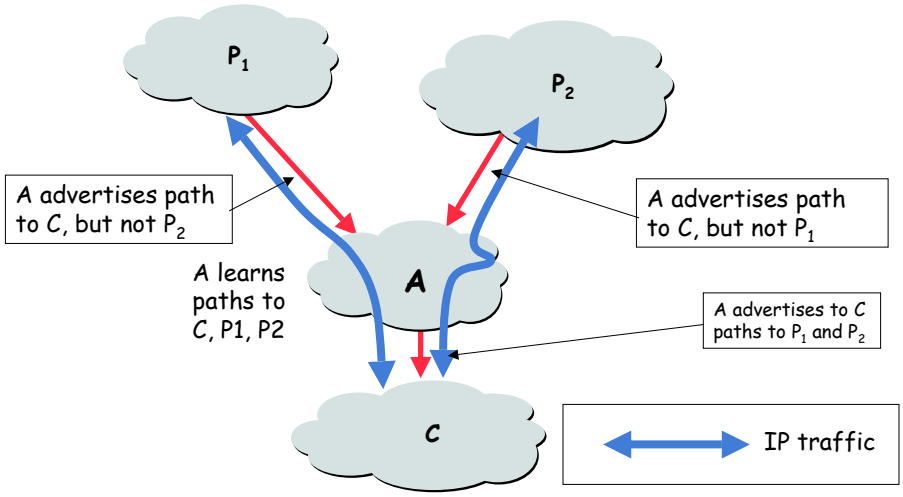
Implication of Business Relationship on Policies

- Route selection (ranking) policy:
  - the typical route selection policy is to prefer customers over peers/providers to reach a destination, i.e., Customer > Peer > Provider
- Route export policy:
  - since the export of a path to a neighbor is an indication that the AS is willing to transport traffic for the neighbor, an AS may not export routes to all neighbors

Typical Export Policies



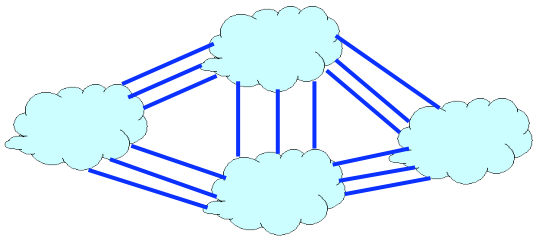
Example Export Policy: No-Valley Routing



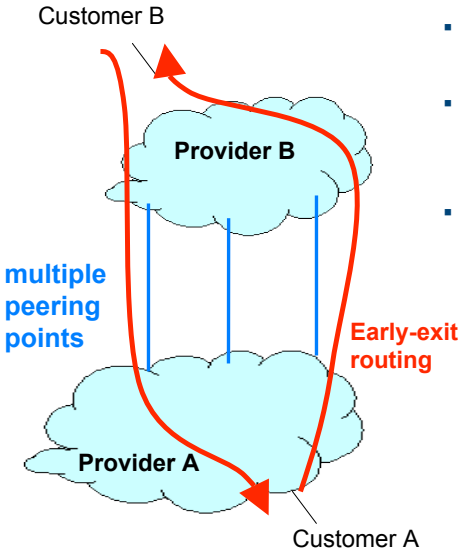
Suppose P_1 and P_2 are providers of *A*; *A* is a provider of *C*

AS Structure: Tier-1 Providers

- Tier-1 provider
  - Has no upstream provider of its own
  - Typically has a national or international backbone
  - UUNET, Sprint, AT&T, Level 3, …
- Top of the Internet hierarchy of 9-15 ASes
  - Full peer-peer connections between tier-1 providers



Efficient Early-Exit Routing



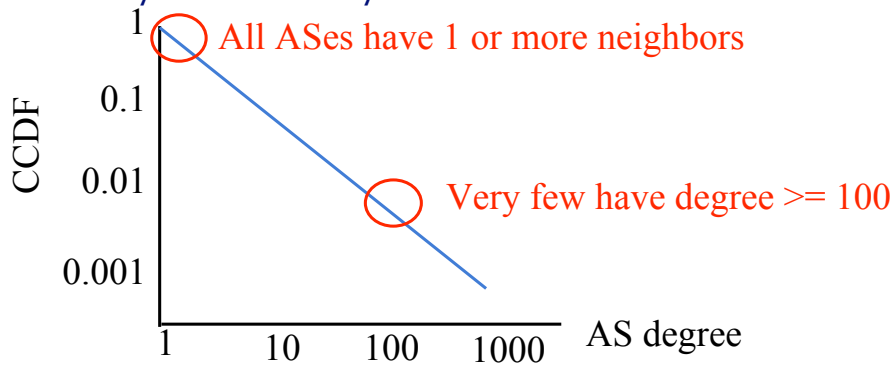
- Diverse peering locations
- Comparable capacity at all peering points
  - Can handle even load
- Consistent routes
  - Same destinations advertised at all points
  - Same AS path length for a destination at all points

AS Structure: Other ASes

- Tier-2 and Tier-3 providers
  - Provide transit service to downstream customers
  - … but, need at least one provider of their own
  - Typically have national or regional scope
  - E.g., Minnesota Regional Network
  - Includes a few thousand of the Ases
- Stub ASes
  - Do not provide transit service to others
  - Connect to one or more upstream providers
  - Includes vast majority (e.g., 85-90%) of the ASes

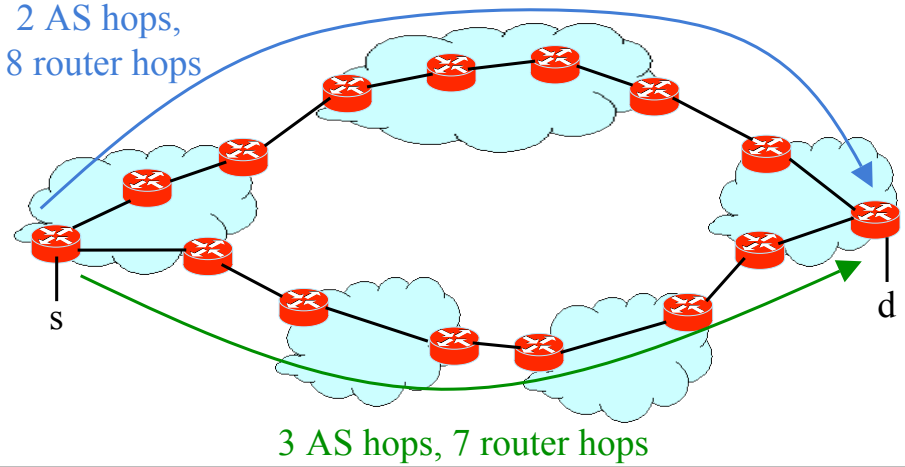
Characteristics of the AS Graph

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



Characteristics of AS Paths

- AS path may be longer than shortest AS path
- Router path may be longer than shortest path

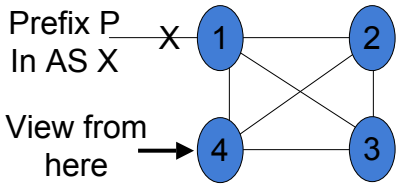


Shared Risks

- Co-location facilities ("co-lo hotels")
  - Places ISPs meet to connect to each other
  - … and co-locate their routers, and share space & power
  - E.g., 32 Avenue of the Americas in NYC
- Shared links
  - Fiber is sometimes leased by one institution to another
  - Multiple fibers run through the same conduits
  - … and run through the same tunnels, bridges, etc.
- Difficult to identify and accounts for these risks
  - Not visible in network-layer measurements
  - E.g., traceroute does not tell you links in the same ditch

Convergence

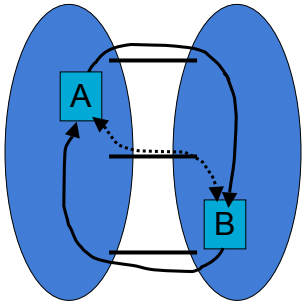
- Recently, it was realized that BGP convergence can undergo a process analogous to count-to-infinity!



- 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

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



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)