Interdomain Routing Policy

COS 461: Computer Networks
Spring 2011

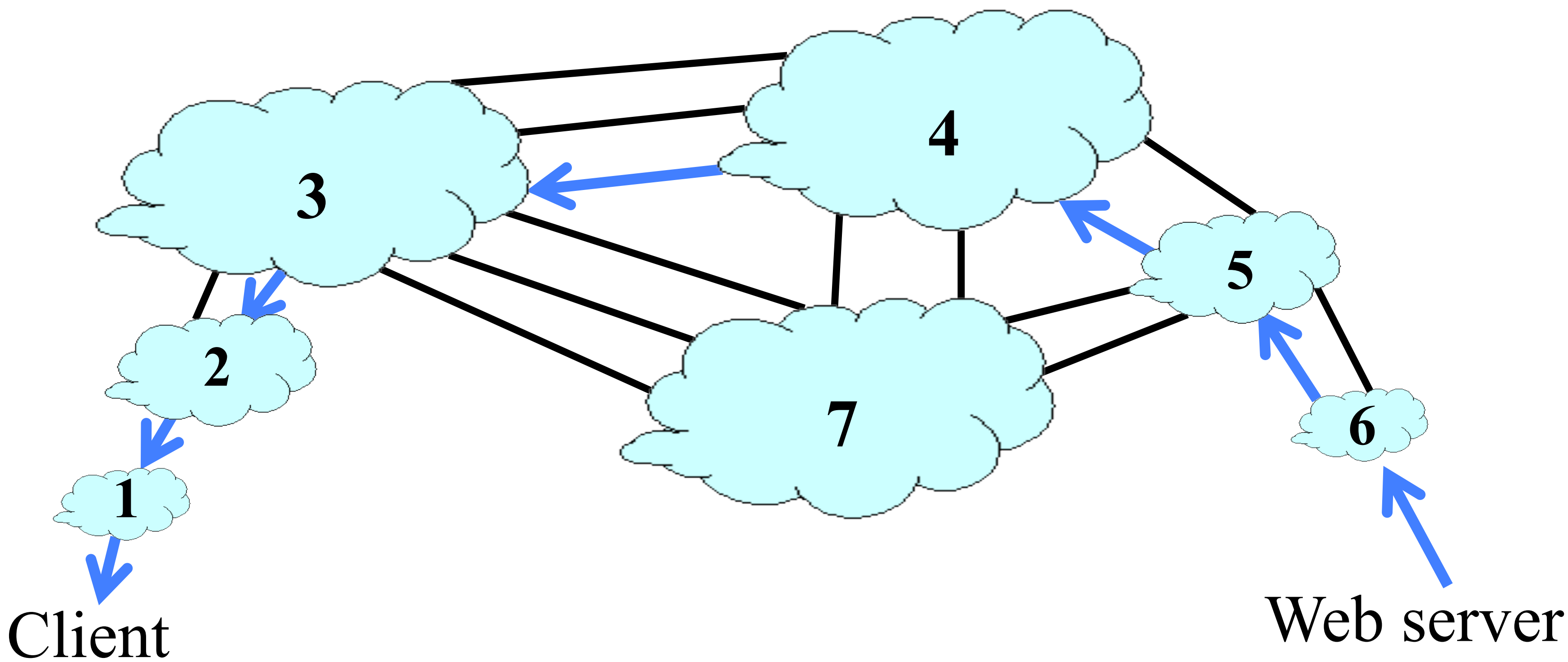
Mike Freedman
http://www.cs.princeton.edu/courses/archive/spring11/cos461/

Goals of Today's Lecture

- Business relationships between ASes
  - Customer-provider: customer pays provider
  - Peer-peer: typically settlement-free
- Realizing routing policies
  - Import and export filtering
  - Assigning preferences to routes
- Multiple routers within an AS
  - Disseminated BGP information within the AS
  - Combining with intradomain routing information

Interdomain Routing

- AS-level topology
  - Destinations are IP prefixes (e.g., 12.0.0.0/8)
  - Nodes are Autonomous Systems (ASes)
  - Edges are links and business relationships



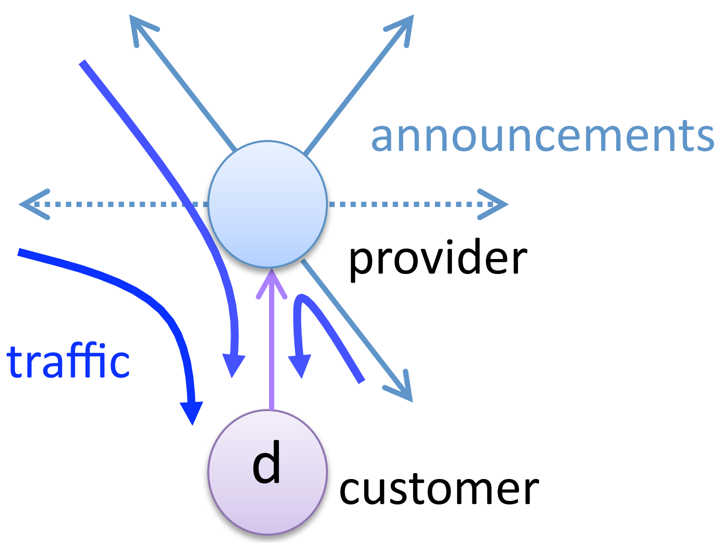
Business Relationships

- Neighboring ASes have business contracts
  - How much traffic to carry
  - Which destinations to reach
  - How much money to pay
- Common business relationships
  - Customer-provider: Customer pays provider for transit
    - E.g., Princeton is a customer of USLEC
    - E.g., MIT is a customer of Level3
  - Peer-peer: No money changes hands
    - E.g., UUNET is a peer of Sprint
    - E.g., Harvard is a peer of Harvard Business School

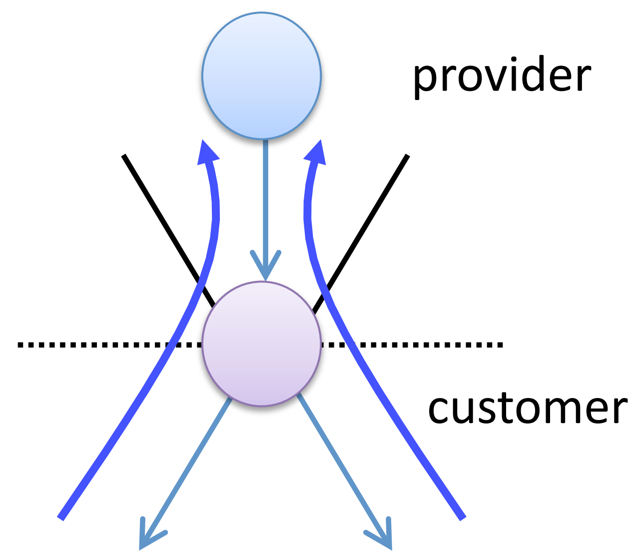
Customer-Provider Relationship

- 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

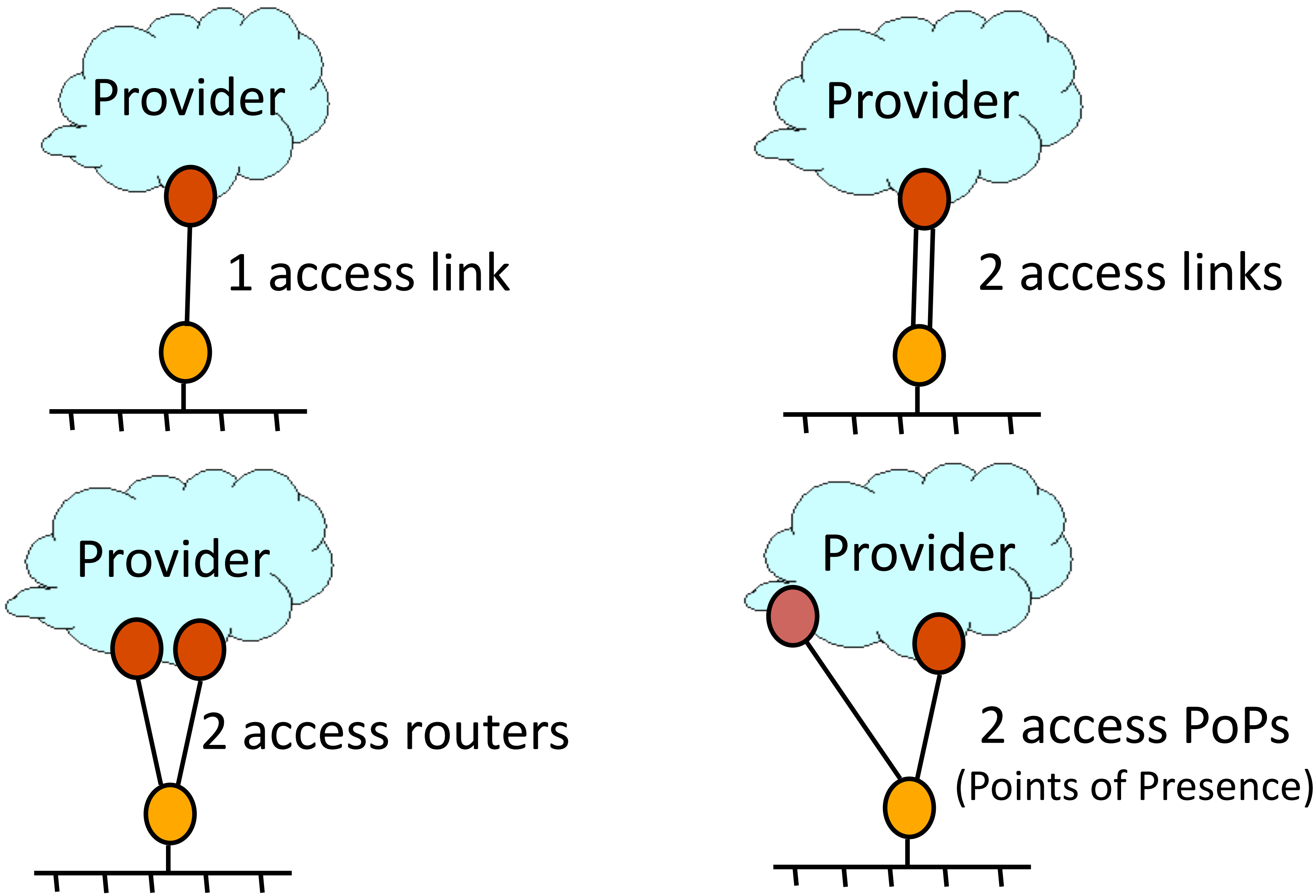
Traffic **to** the customer



Traffic **from** the customer

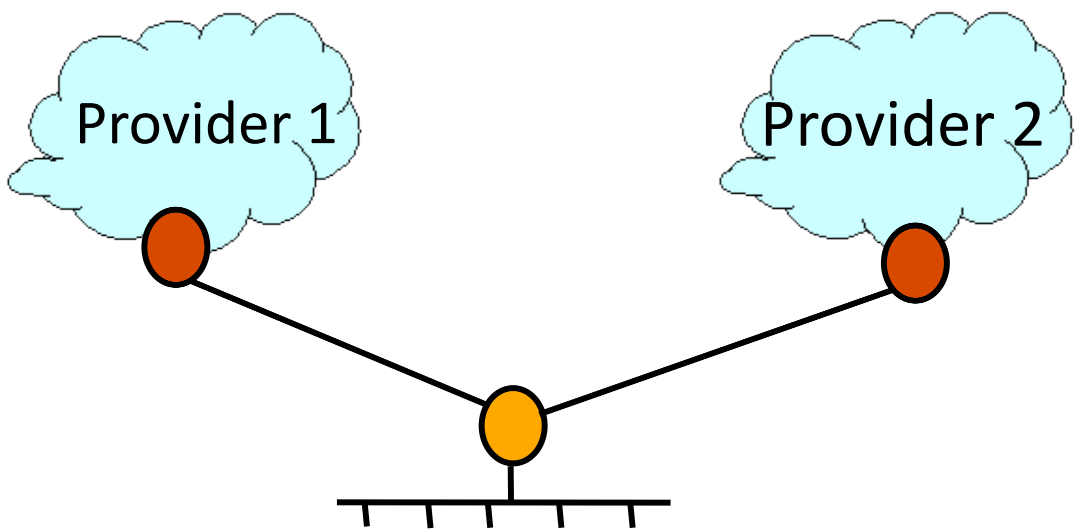


Customer Connecting to a Provider



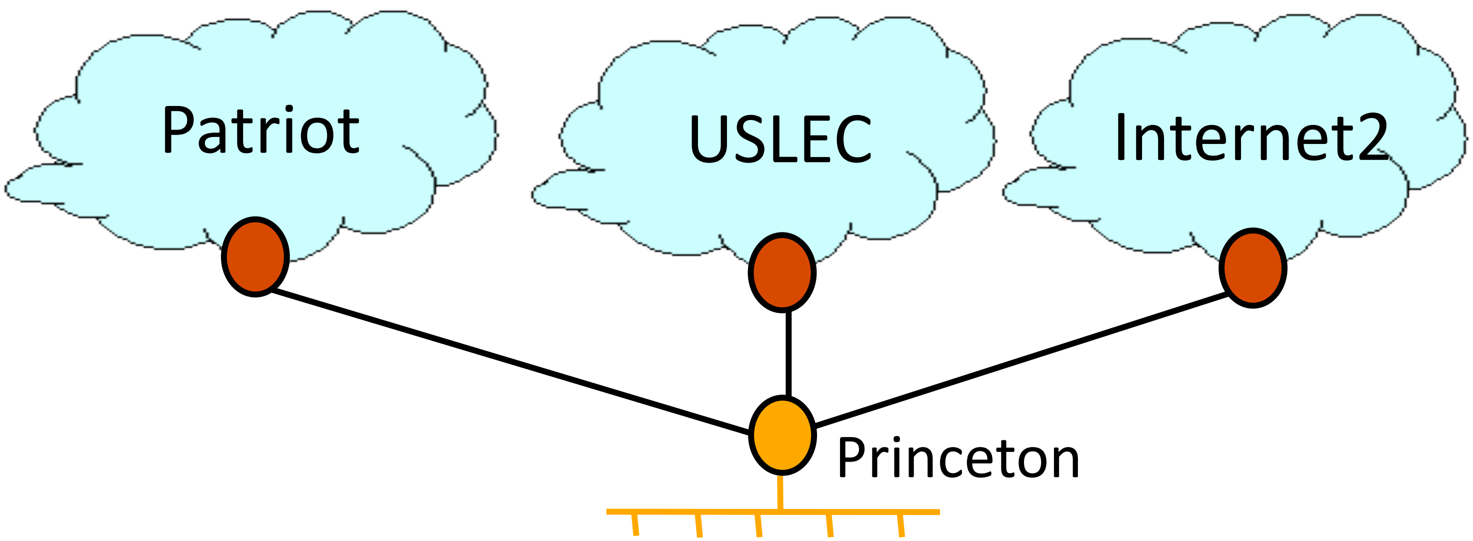
Multi-Homing: Two or More Providers

- 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

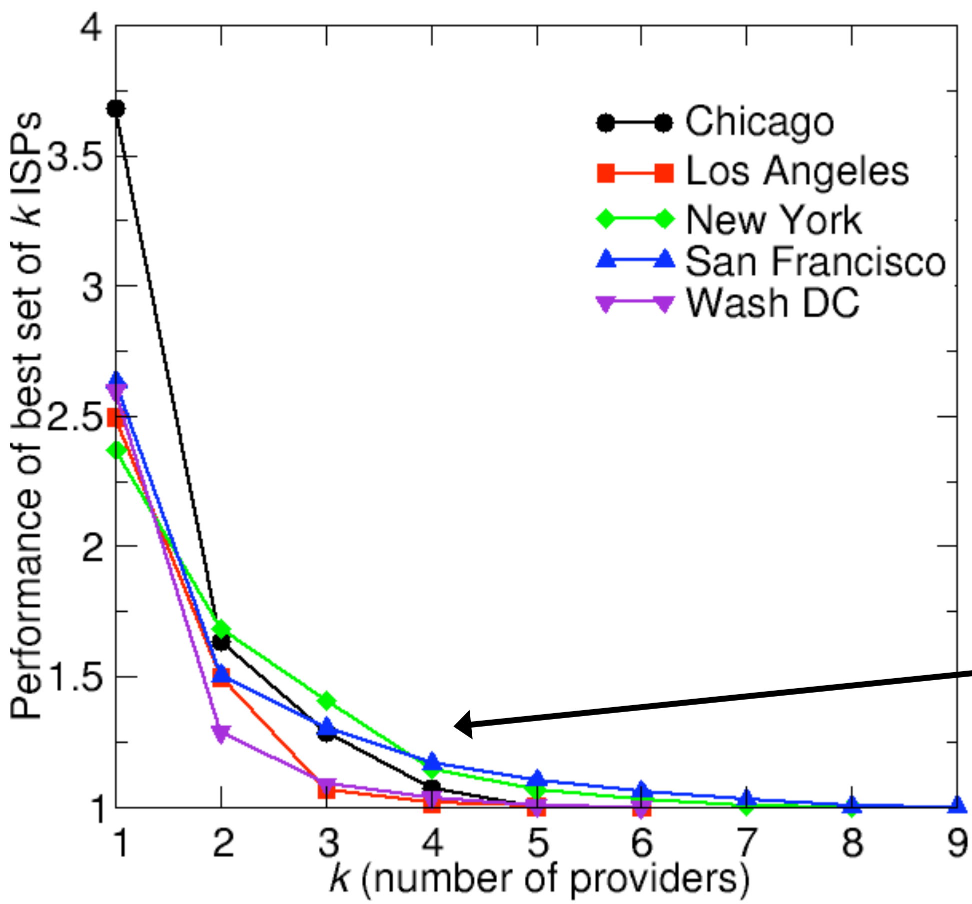


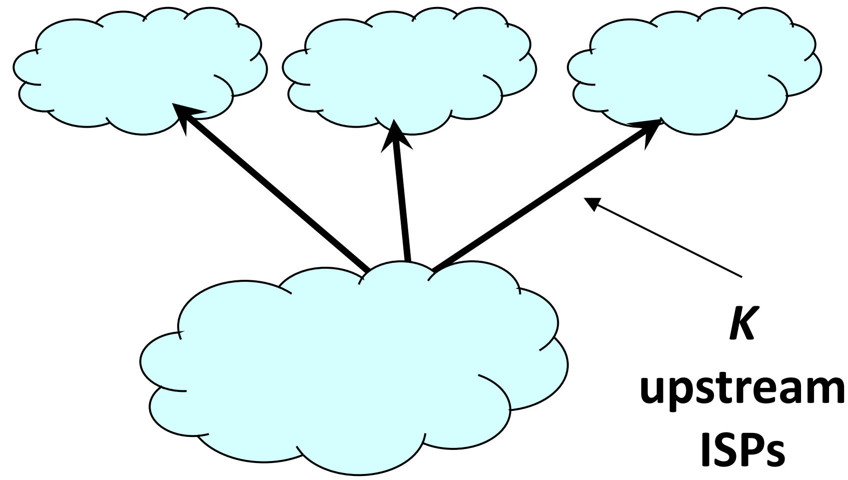
Princeton Example

- Internet: customer of USLEC and Patriot
- Research universities/labs: customer of Internet2
- Local non-profits: provider for several non-profits



How many links are enough?





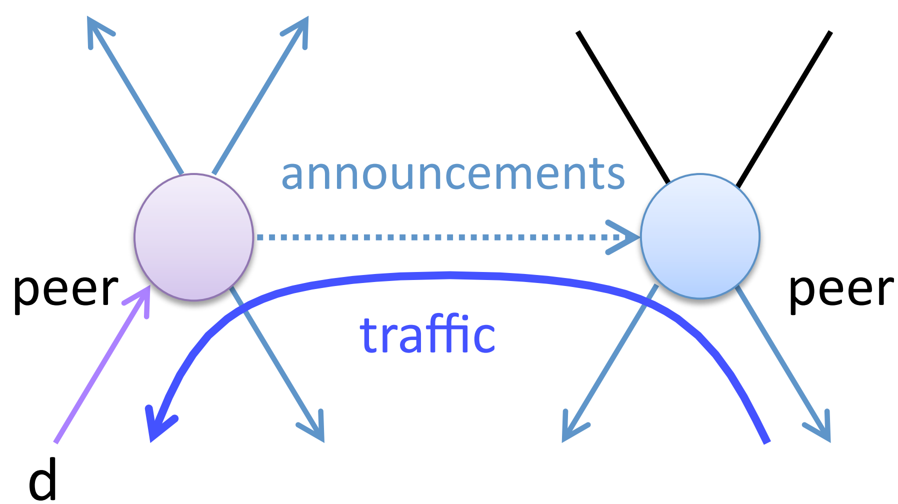
Not much benefit beyond 4 ISPs

Akella *et al.*, "Performance Benefits of Multihoming", *SIGCOMM 2003*

Peer-Peer Relationship

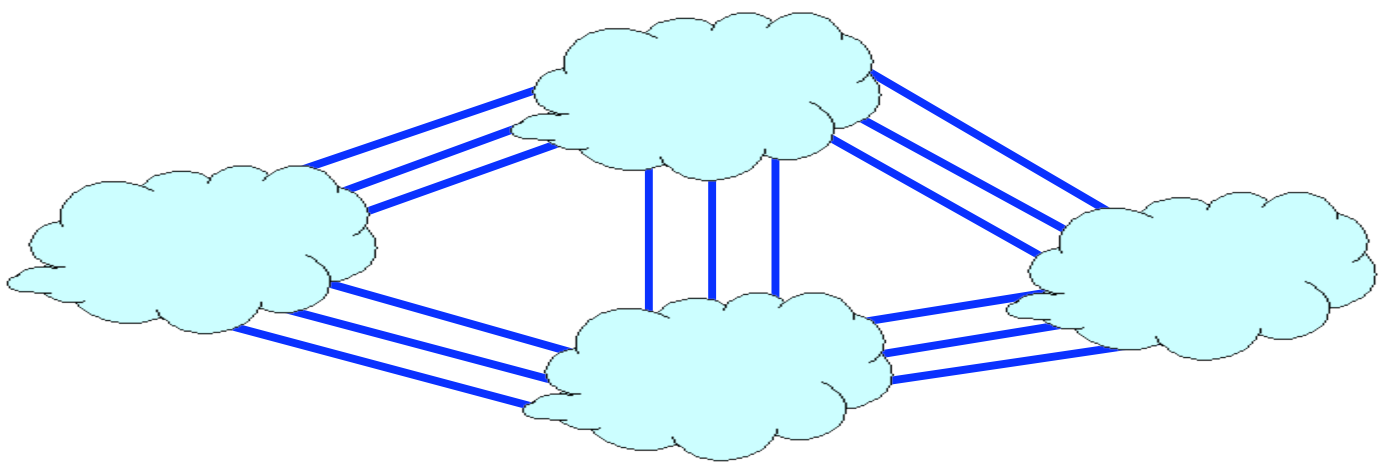
- Peers exchange traffic between customers
  - AS exports *only* customer routes to a peer
  - AS exports a peer's routes *only* to its customers
  - Often the relationship is settlement-free (i.e., no $$$)

Traffic to/from the peer and its customers



AS Structure: Tier-1 Providers

- Tier-1 provider
  - Has no upstream provider of its own
  - Typically has a national or international backbone
- Top of the Internet hierarchy of ~10 ASes
  - AOL, AT&T, Global Crossing, Level3, UUNET, NTT, Qwest, SAVVIS (formerly Cable & Wireless), and Sprint
  - Full peer-peer connections between tier-1 providers



AS Structure: Other ASes

- Other providers
  - Provide transit service to downstream customers
  - … but, need at least one provider of their own
  - Typically have national or regional scope
  - Includes several thousand ASes
- Stub ASes
  - Do not provide transit service to others
  - Connect to one or more upstream providers
  - Includes the vast majority (e.g., 85-90%) of the ASes

The Business Game and Depeering

- Cooperative competition (brinksmanship)
- Much more desirable to have your peer's customers
  – Much nicer to get paid for transit
- Peering "tiffs" are relatively common

31 Jul 2005: Level 3 Notifies Cogent of intent to disconnect.

16 Aug 2005: Cogent begins massive sales effort and mentions a 15 Sept. expected depeering date.

31 Aug 2005: Level 3 Notifies Cogent again of intent to disconnect (according to Level 3)

5 Oct 2005 9:50 UTC: Level 3 disconnects Cogent. Mass hysteria ensues up to, and including policymakers in Washington, D.C.

7 Oct 2005: Level 3 reconnects Cogent

During the "outage", Level 3 and Cogent's singly homed customers could not reach each other. (~ 4% of the Internet's prefixes were isolated from each other)

Depeering Continued

Resolution…

Level 3 and Cogent Reach Agreement on Equitable Peering Terms

Friday October 28, 7:00 am ET

BROOMFIELD, Colo. and WASHINGTON, Oct. 28 /PRNewswire-FirstCall/ -- Level 3 Communications (Nasdaq: LVLT - News) and Cogent Communications (Amex: COI - News) today announced that the companies have agreed on terms to continue to exchange Internet traffic under a modified version of their original peering agreement. The modified peering arrangement allows for the continued exchange of traffic between the two companies' networks, and includes commitments from each party with respect to the characteristics volume of traffic to be exchanged. Under the terms of the agreement, the companies have agreed to the settlement-free exchange of traffic subject to specific payments if certain obligations are not met.

…but not before attempt to steal customers!

As of 5:30 am EDT, October 5th, Level(3) terminated peering with Cogent without causes …. Cogent has left the peering circuits open in the hope that Level(3) will change its mind and allow traffic to be exchanged between our networks. We are extending a special offering to single homed Level 3 customers.

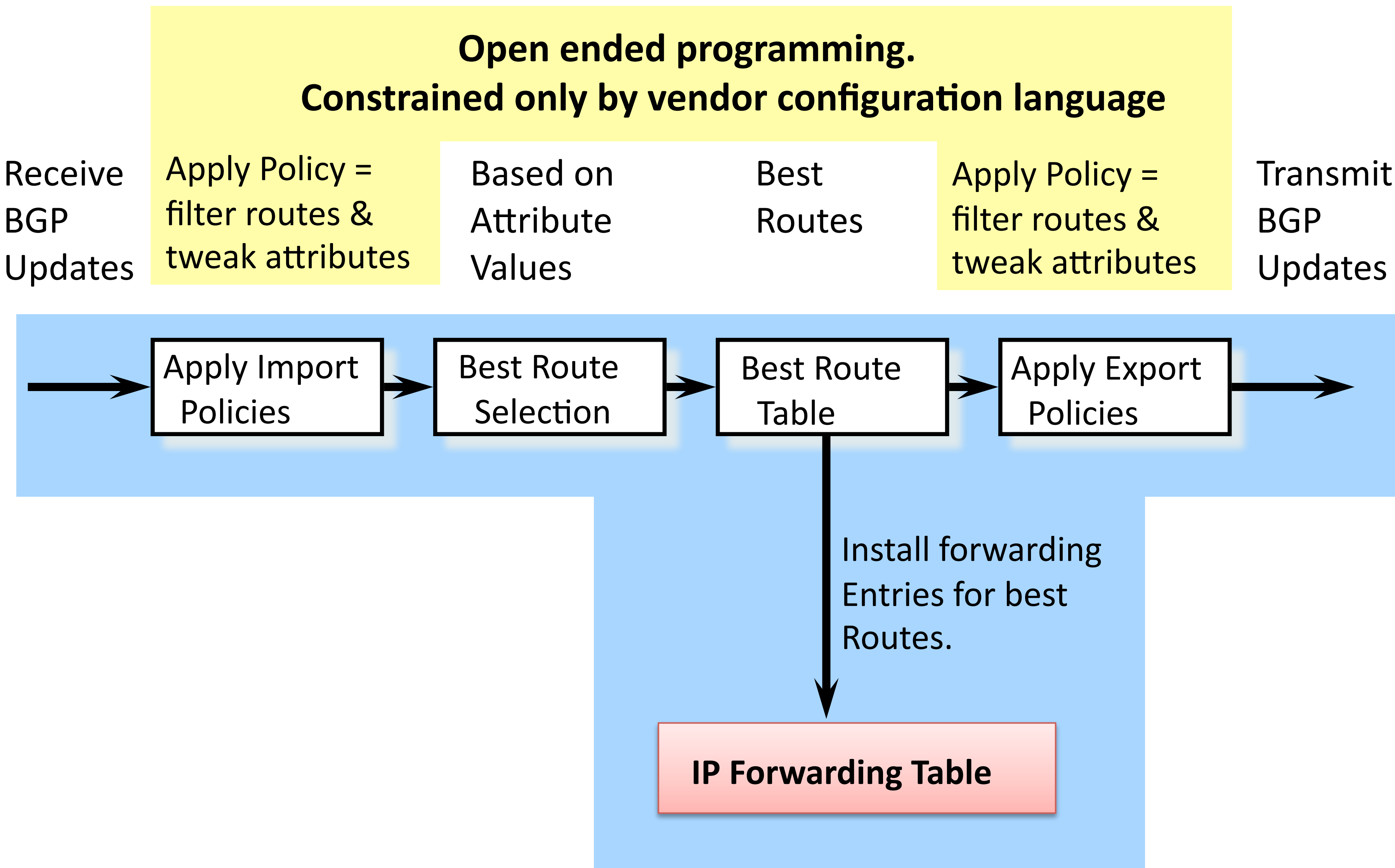
"Cogent will offer any Level 3 customer, who is single homed to the Level 3 …, one year of full Internet transit free of charge at the same bandwidth…. Cogent will provide this connectivity in over 1,000 locations."

Realizing BGP Routing Policy

BGP Policy: Applying Policy to Routes

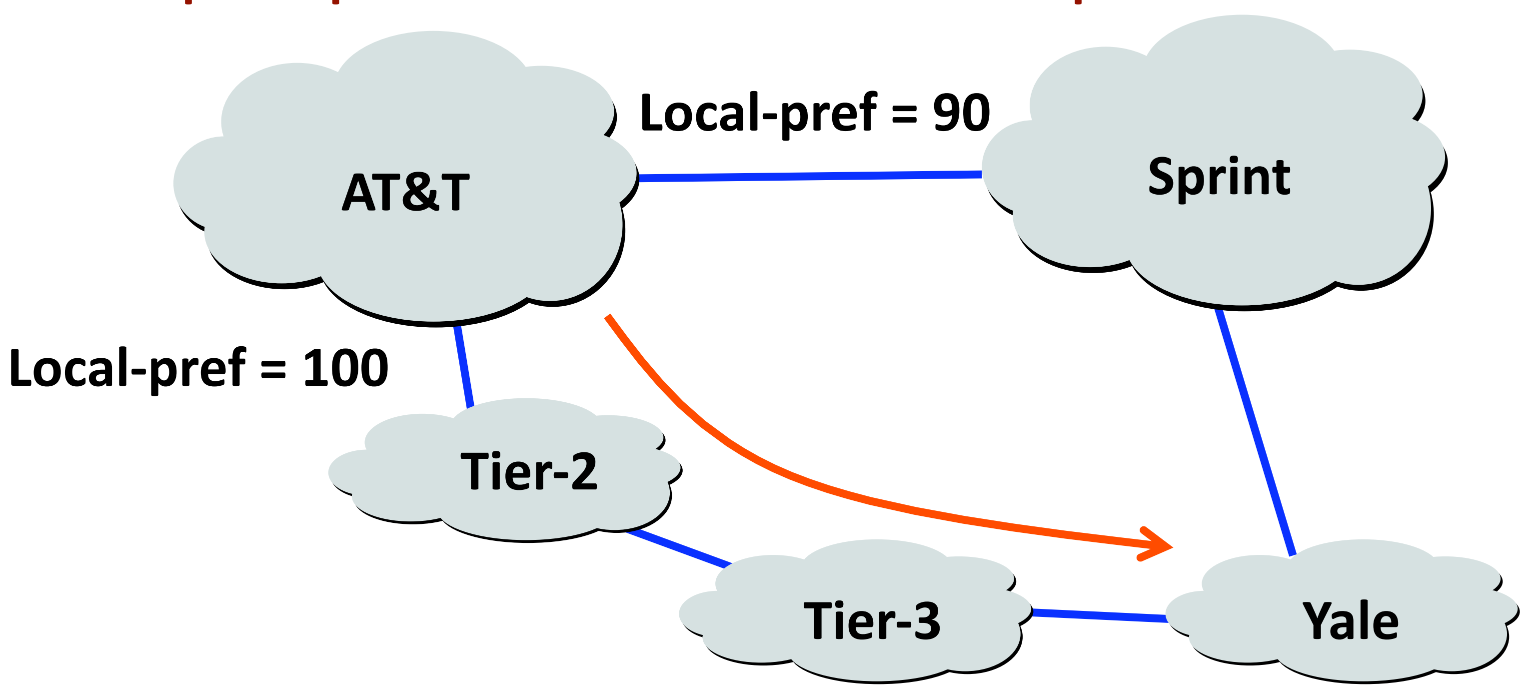
- Import policy
  - Filter unwanted routes from neighbor
    - E.g. prefix that your customer doesn't own
  - Manipulate attributes to influence path selection
    - E.g., assign local preference to favored routes
- Export policy
  - Filter routes you don't want to tell your neighbor
    - E.g., don't tell a peer a route learned from other peer
  - Manipulate attributes to control what they see
    - E.g., make a path look artificially longer than it is

BGP Policy: Influencing Decisions



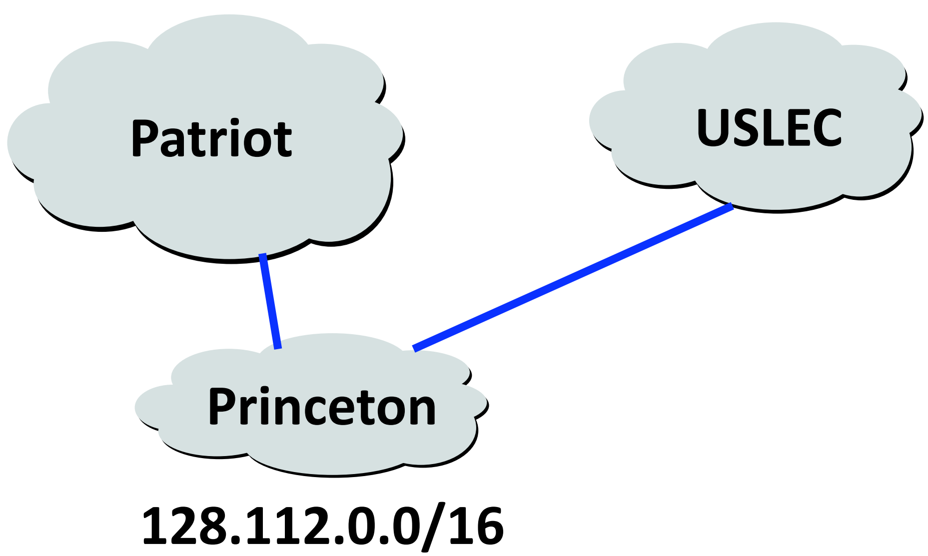
Import Policy: Local Preference

- Favor one path over another
  - Override the influence of AS path length
  - Apply local policies to prefer a path
- Example: prefer customer over peer



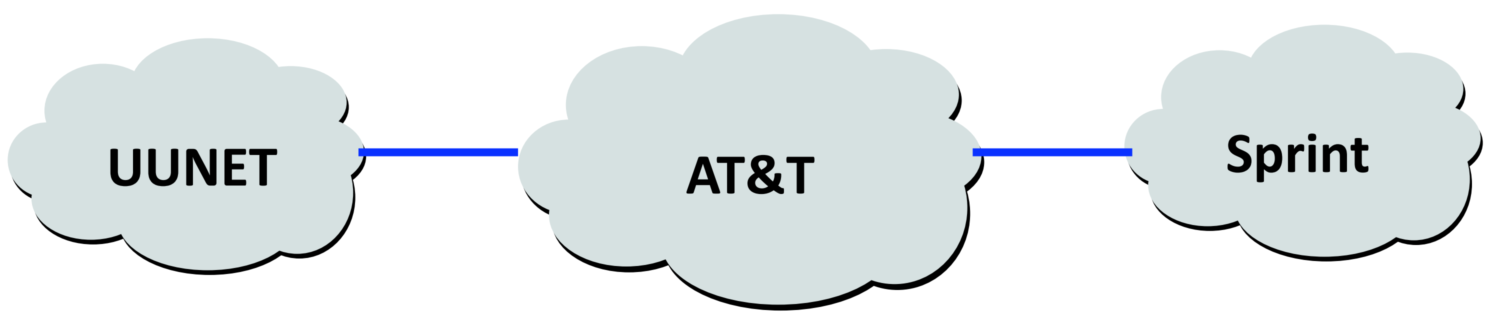
Import Policy: Filtering

- Discard some route announcements
  - Detect configuration mistakes and attacks
- Examples on session to a customer
  - Discard route if prefix not owned by the customer
  - Discard route that contains other large ISP in AS path



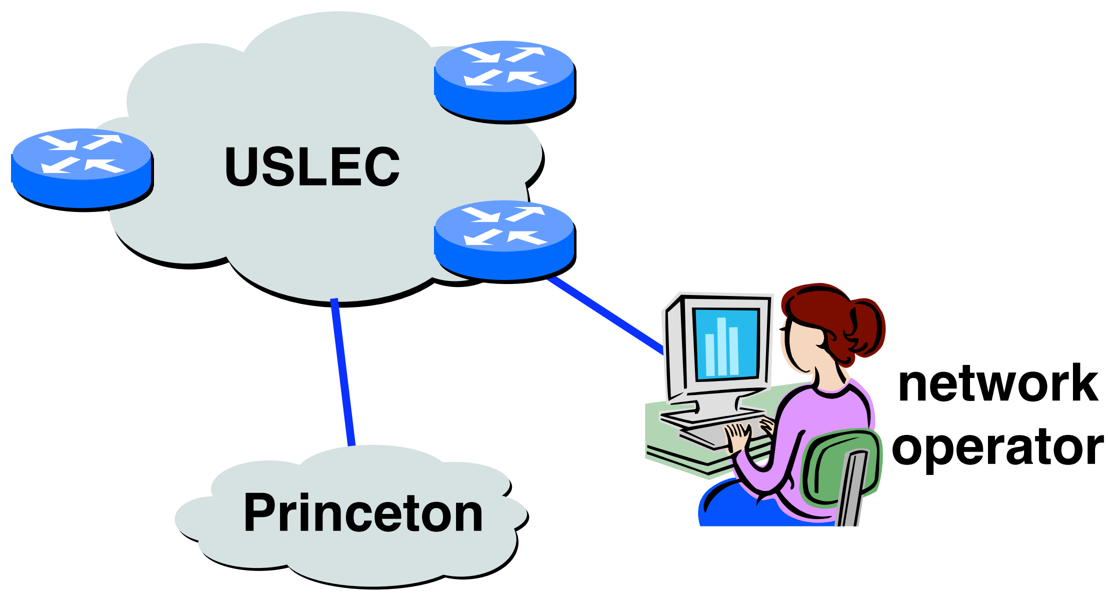
Export Policy: Filtering

- Discard some route announcements
  - Limit propagation of routing information
- Examples
  - Don't announce routes from one peer to another



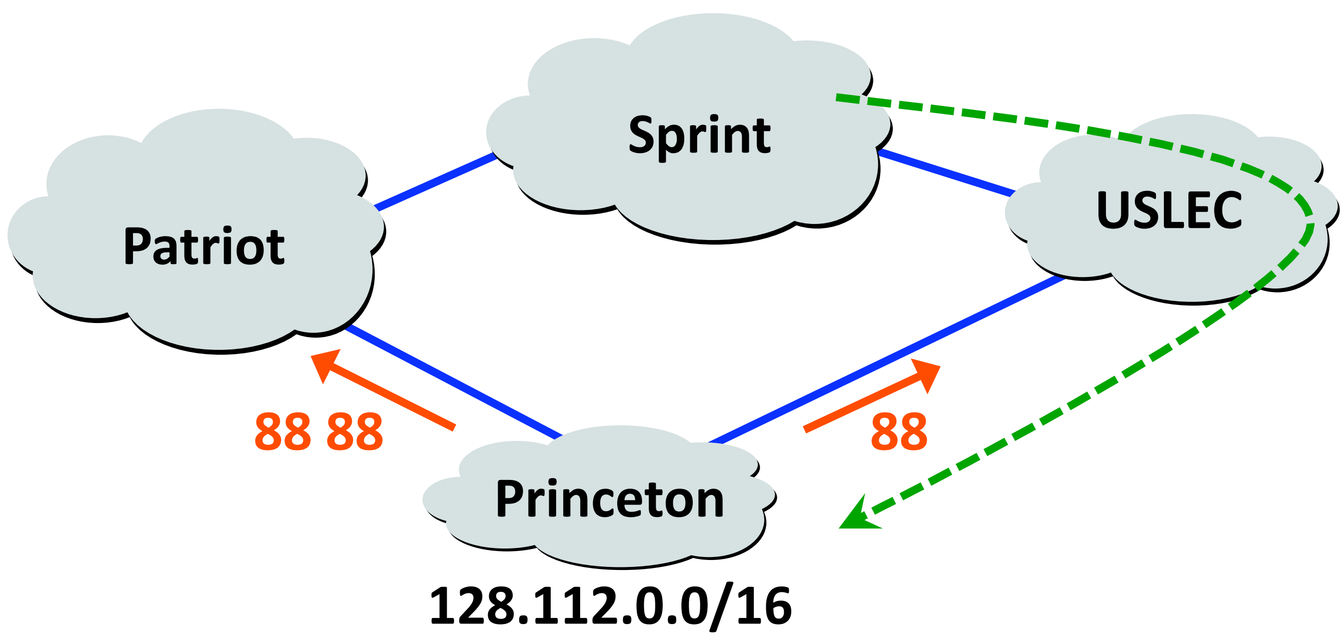
Export Policy: Filtering

- Discard some route announcements
  - Limit propagation of routing information
- Examples
  - Don't announce routes for network-management hosts or the underlying routers themselves



Export Policy: Attribute Manipulation

- Modify attributes of the active route
  - To influence the way other ASes behave
- Example: AS prepending
  - Artificially inflate the AS path length seen by others
  - To convince some ASes to send traffic another way



BGP Policy Configuration

- Routing policy languages are vendor-specific
  - Not part of the BGP protocol specification
  - Different languages for Cisco, Juniper, etc.
- Still, all languages have some key features
  - Policy as a list of clauses
  - Each clause matches on route attributes
  - … and either discards or modifies the matching routes
- Configuration done by human operators
  - Implementing the policies of their AS
  - Business relationships, traffic engineering, security, …

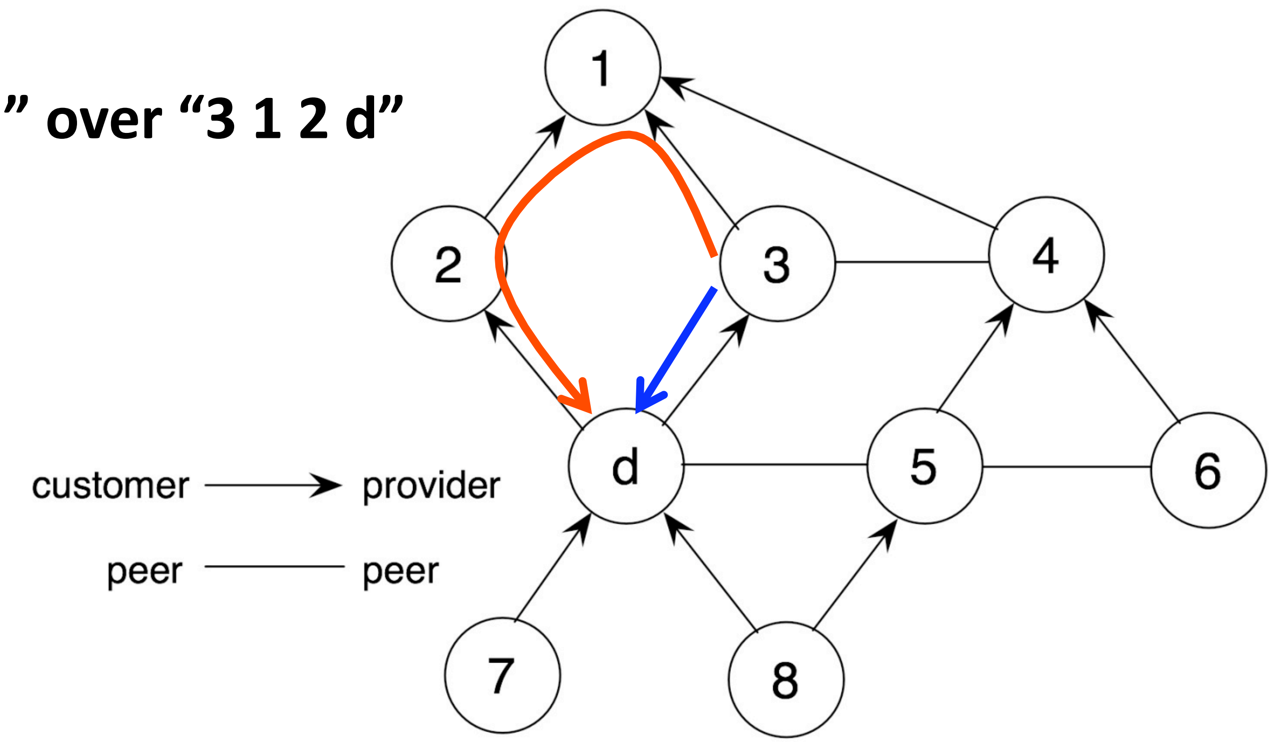
Why Is The Internet Generally Stable?

- Mostly because of $$ ☺
- Policy configurations based on ISPs' bilateral business relationships
  - Customer-Provider
    - Customers pay provider for access to the Internet
  - Peer-Peer
    - Peers exchange traffic free of charge
- Most well-known result reflecting this practice: "Gao-Rexford" stability conditions

The "Gao-Rexford" Stability Conditions

- Preference condition
  - Prefer customer routes over peer or provider routes

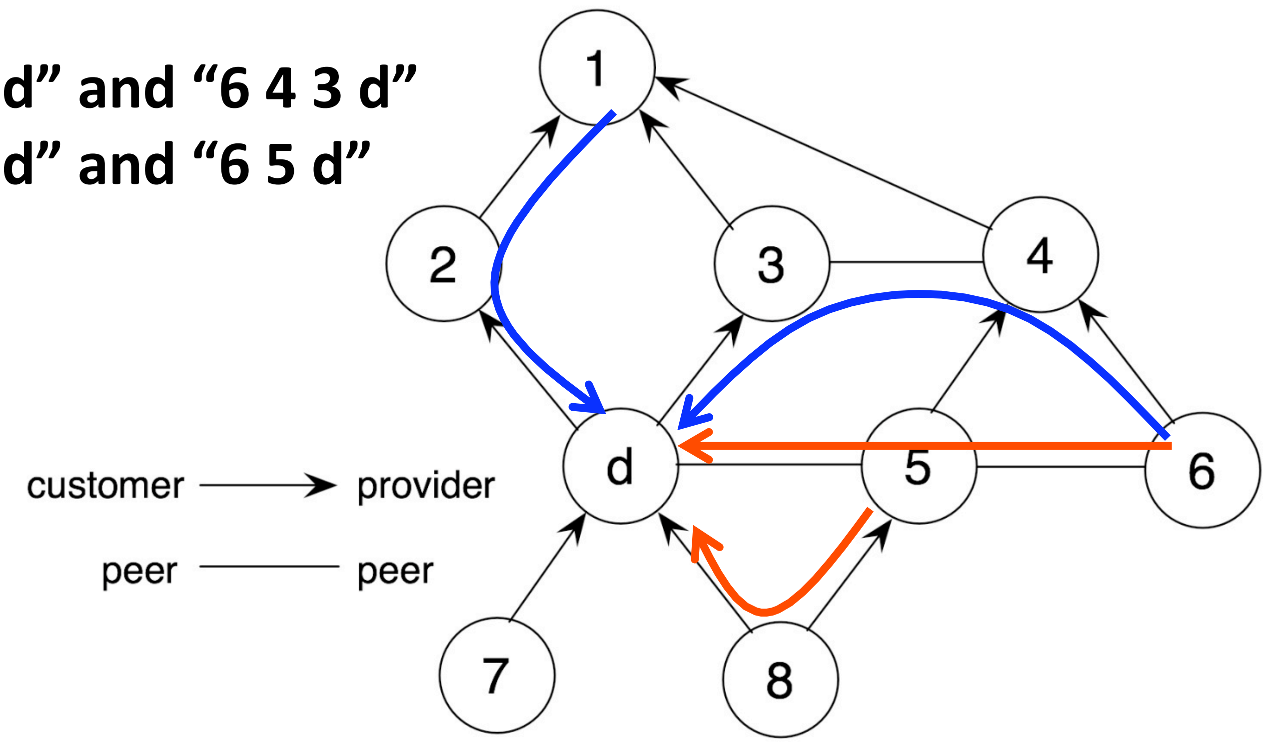
Node 3 prefers "3 d" over "3 1 2 d"



The "Gao-Rexford" Stability Conditions

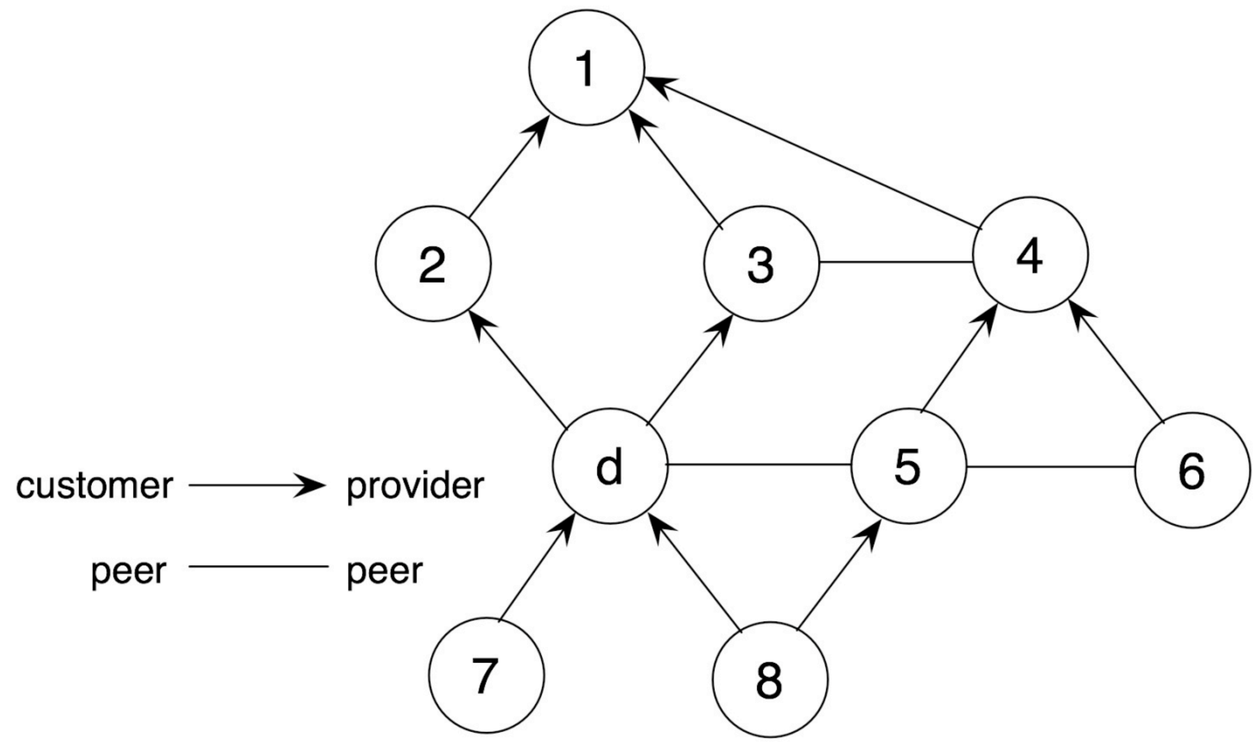
- Export condition
  - Export only customer routes to peers or providers

Valid paths: "1 2 d" and "6 4 3 d"
Invalid path: "5 8 d" and "6 5 d"



The "Gao-Rexford" Stability Conditions

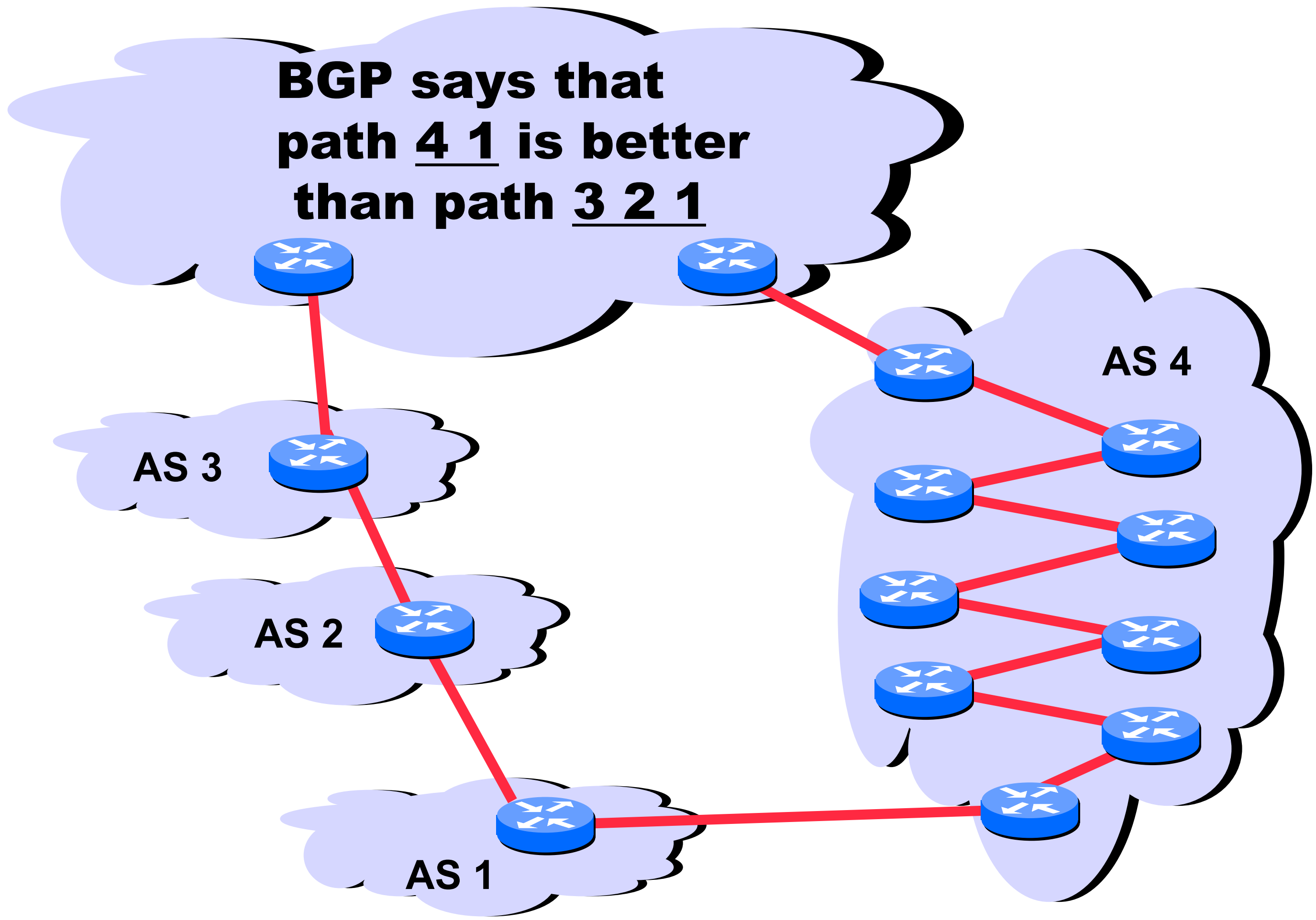
- Topology condition (acyclic)
  - No cycle of customer-provider relationships



BGP and
Multiple Routers in an AS

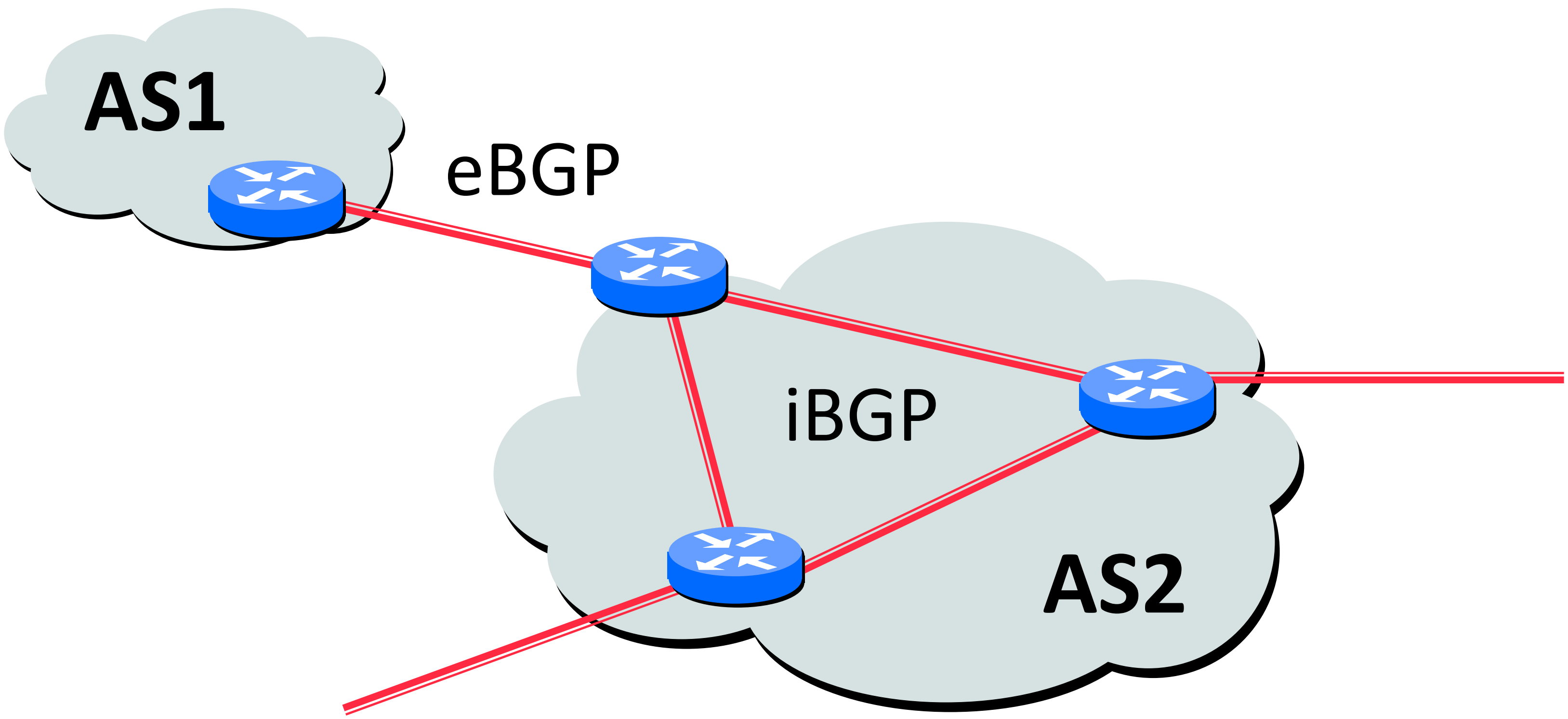
An AS is Not a Single Node

- AS path length can be misleading
  - An AS may have many router-level hops



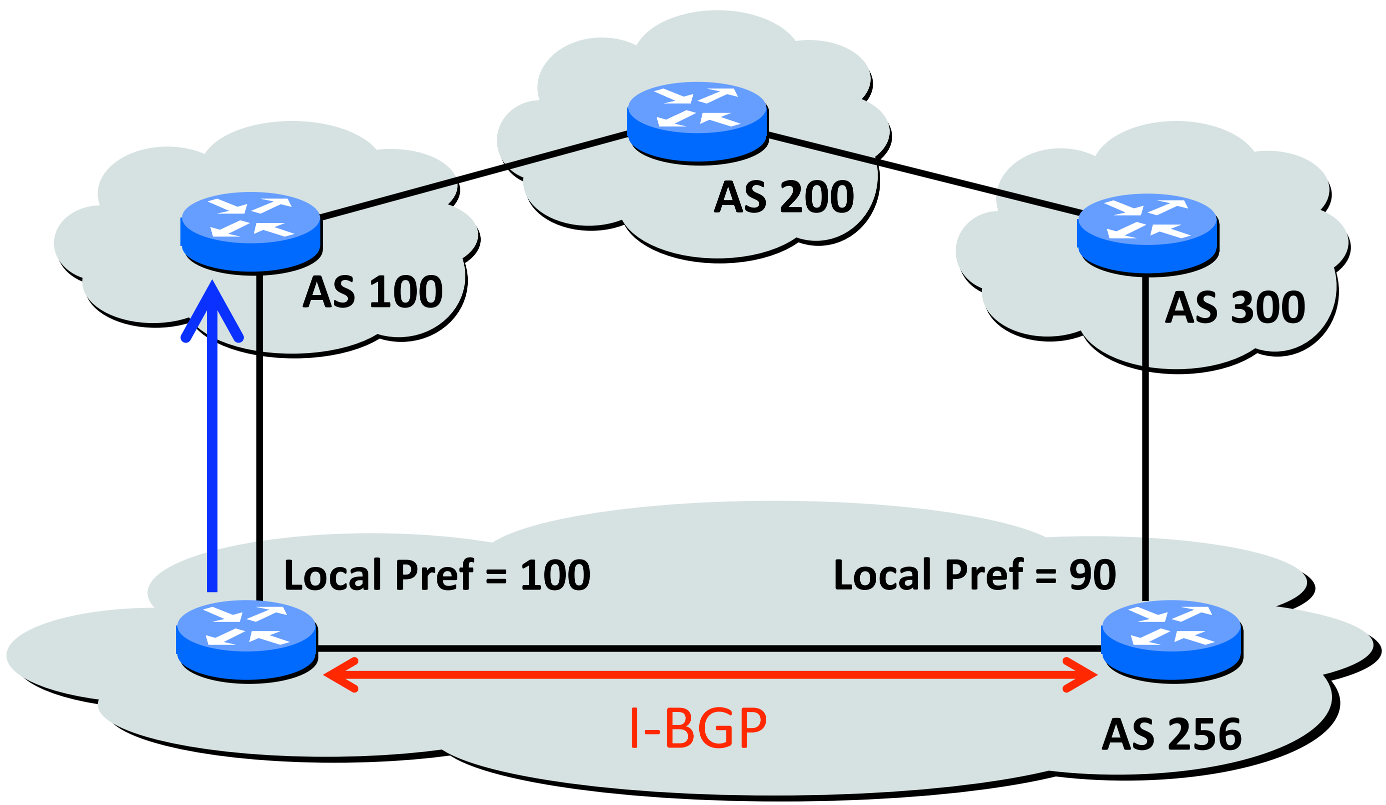
An AS is Not a Single Node

- Multiple routers in an AS
  - Need to distribute BGP information within the AS
  - Internal BGP (iBGP) sessions between routers



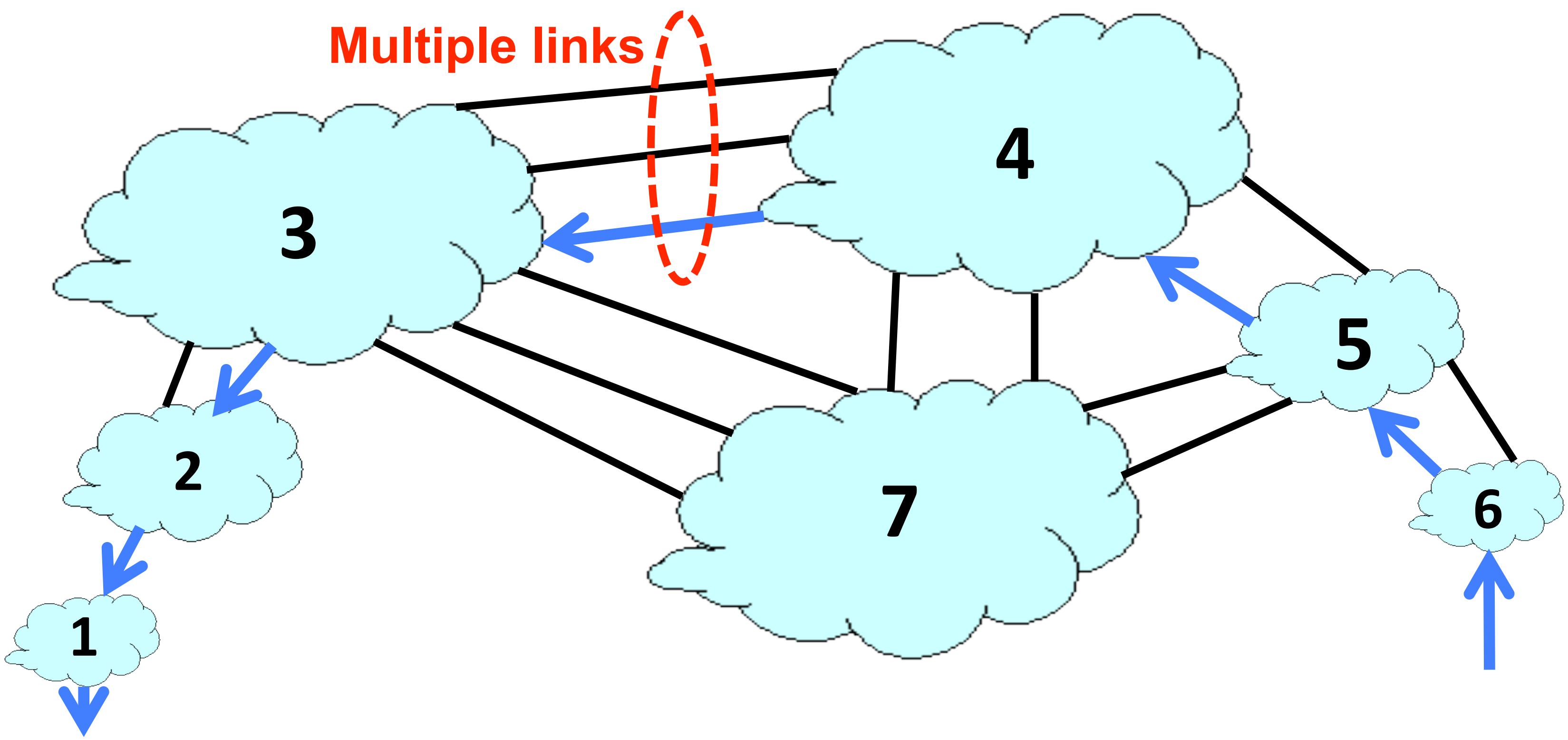
Internal BGP and Local Preference

- Example
  - Both routers prefer path through AS 100 on the left
  - … even though right router learns an external path

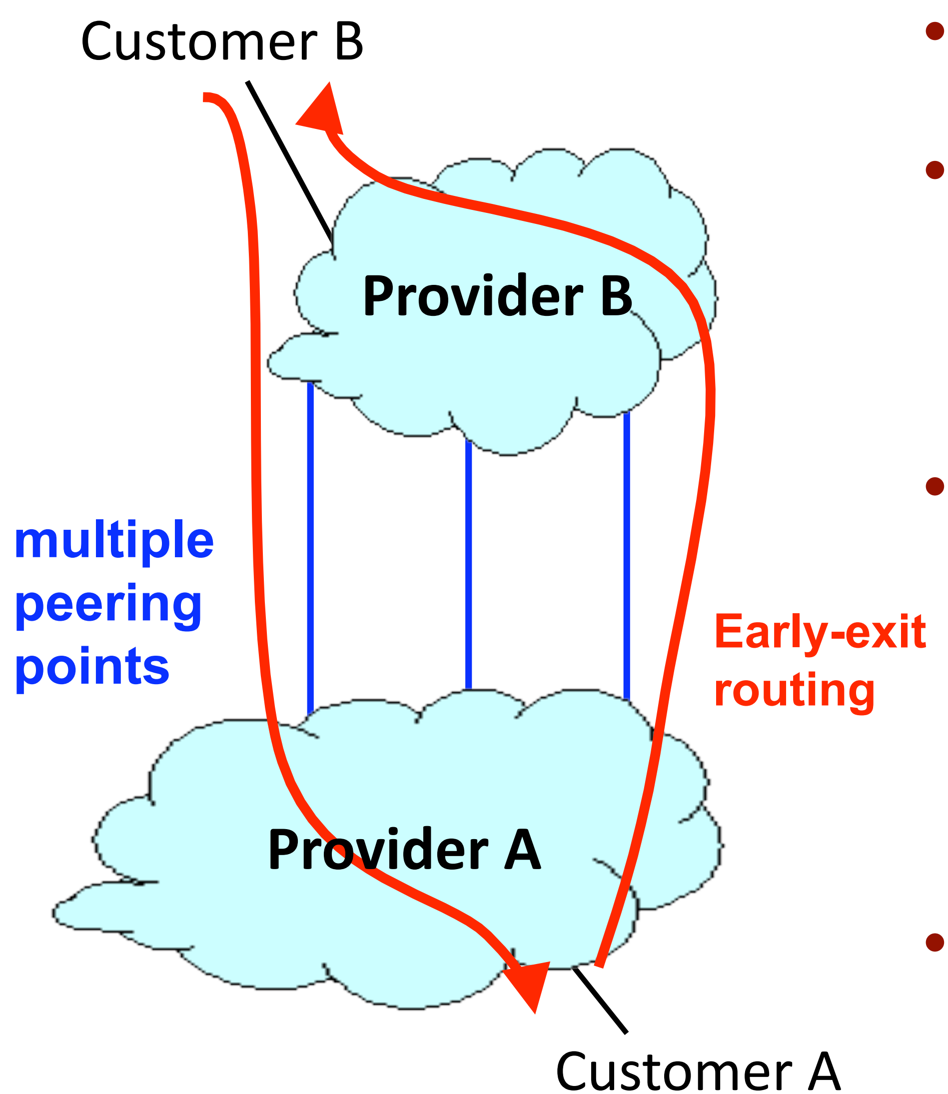


An AS is Not a Single Node

- Multiple connections to neighboring ASes
  - Multiple border routers may learn good routes
  - … with the same local-pref and AS path length



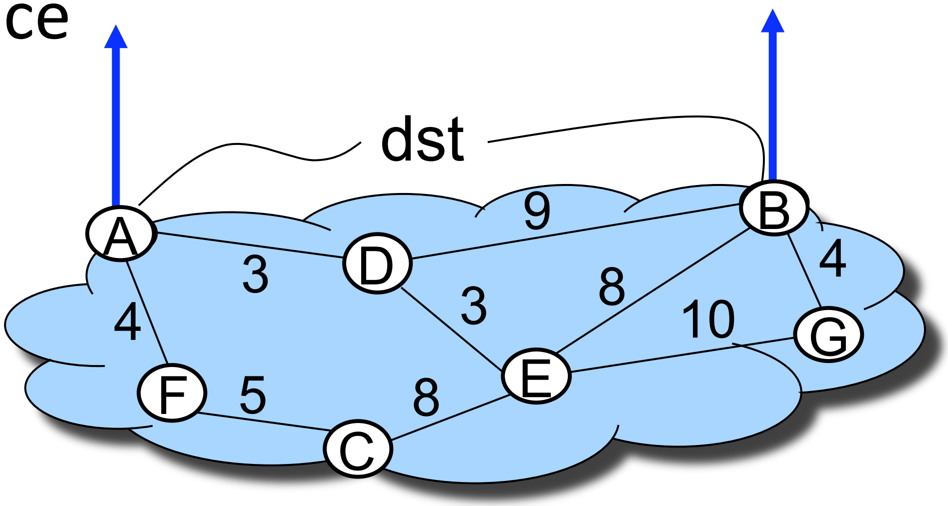
Early-Exit or Hot-Potato Routing



- 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
- Why not push wide-area routing to peer?

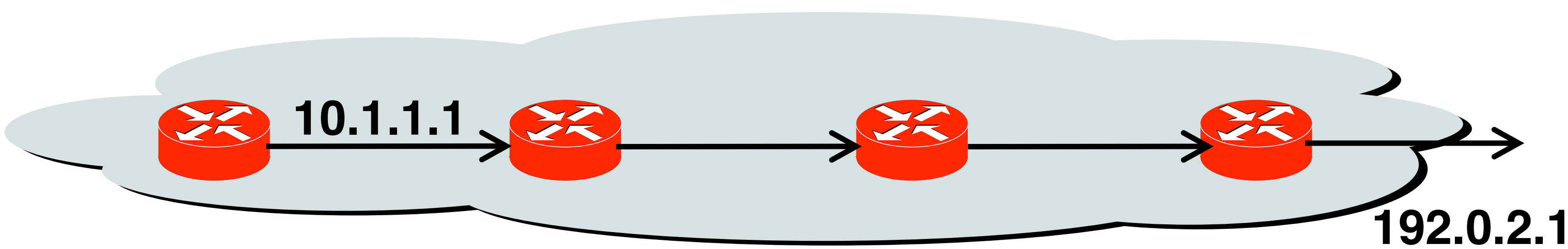
Realizing Hot-Potato Routing

- Hot-potato routing
  - Each router selects the closest egress point
  - … based on the path cost in intra-domain protocol
- BGP decision process
  - Highest local preference
  - Shortest AS path
  - Closest egress point
  - Arbitrary tie break

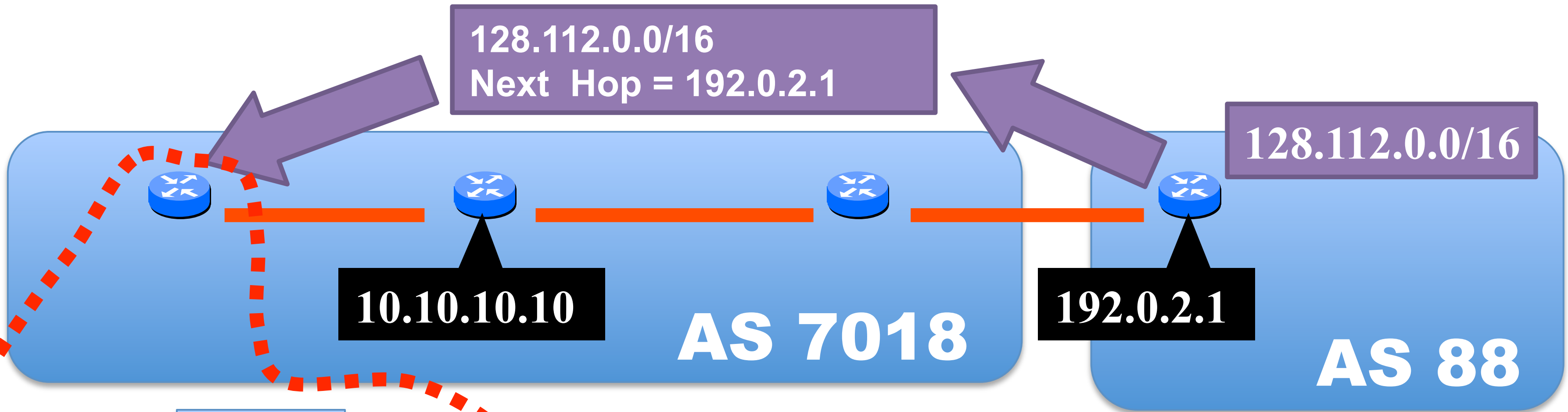


Joining BGP and IGP Information

- Border Gateway Protocol (BGP)
  - Announces reachability to external destinations
  - Maps a destination prefix to an egress point
    - 128.112.0.0/16 reached via 192.0.2.1
- Interior Gateway Protocol (IGP)
  - Used to compute paths within the AS
  - Maps an egress point to an outgoing link
    - 192.0.2.1 reached via 10.1.1.1



Joining BGP with IGP Information



IGP

destination	next hop
192.0.2.0/30	10.10.10.10

+

BGP

destination	next hop
128.112.0.0/16	192.0.2.1

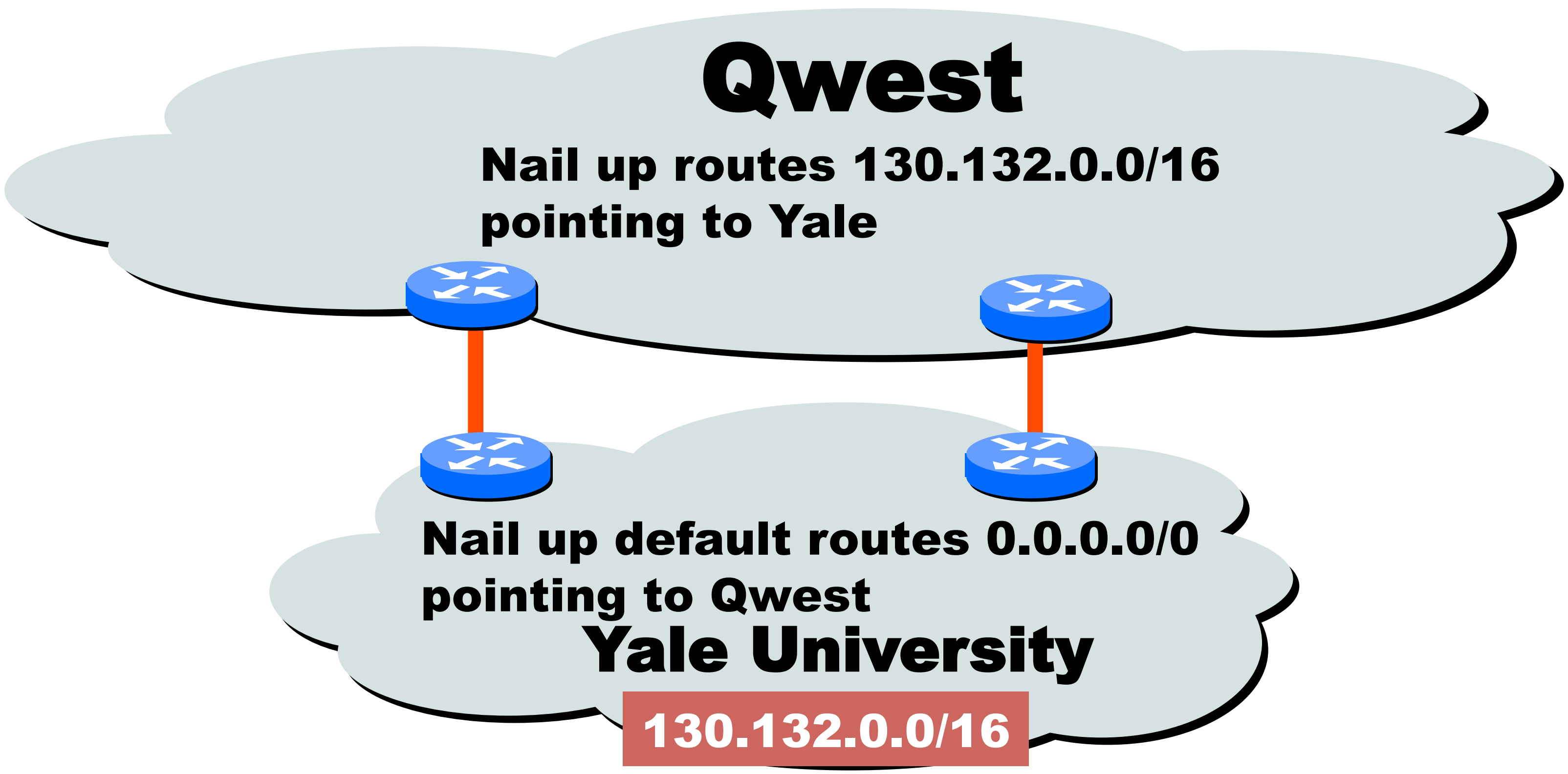
→

Forwarding Table

destination	next hop
128.112.0.0/16	10.10.10.10
192.0.2.0/30	10.10.10.10

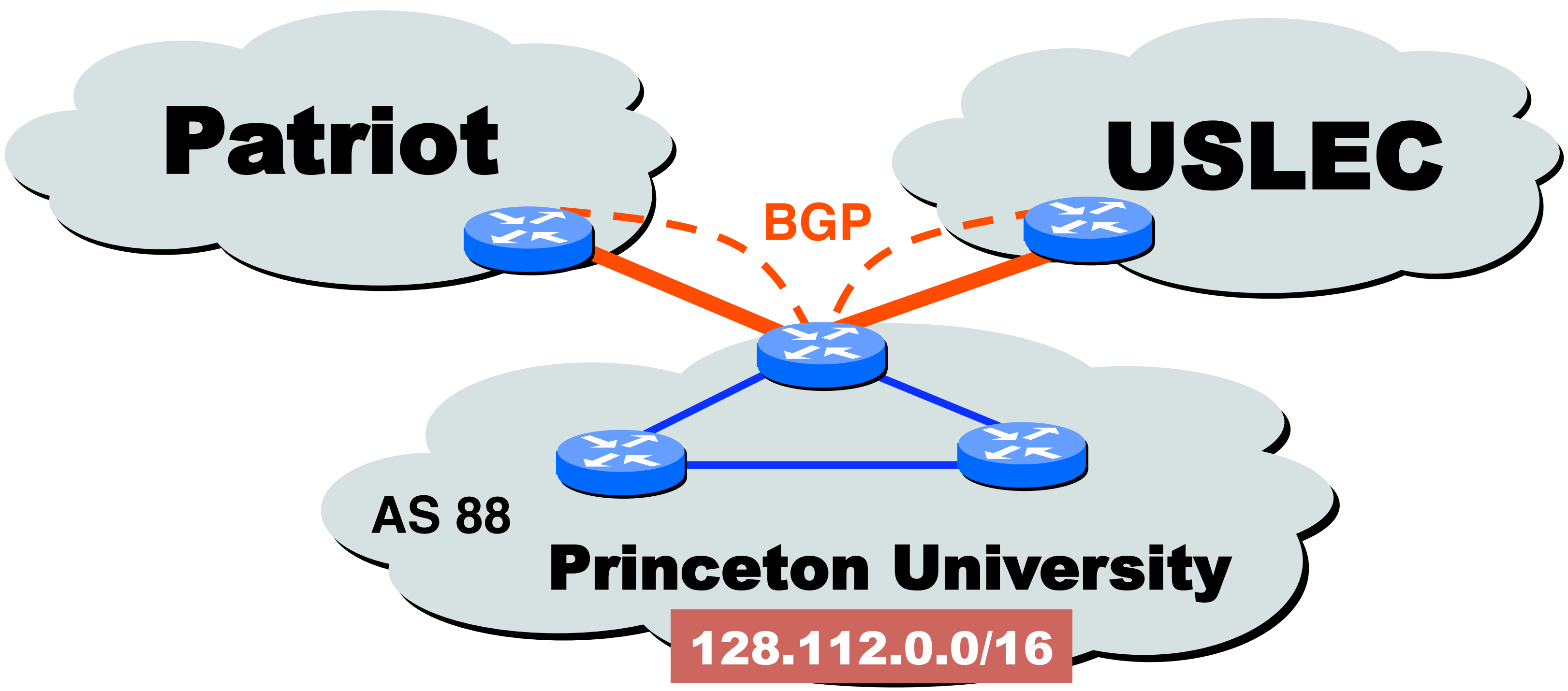
Some Routers Don't Need BGP

- Customer that connects to a single upstream ISP
  - The ISP can introduce the prefixes into BGP
  - … and customer can simply default-route to the ISP



Some Routers Don't Need BGP

- Routers inside a "stub" network
  - Border router may speak BGP to upstream ISPs
  - But, internal routers can simply "default route"



Conclusions

- BGP is solving a hard problem
  - Routing protocol operating at a global scale
  - Tens of thousands of independent networks
  - Each have own policy goals; all want convergence

- Key features of BGP
  - Prefix-based path-vector protocol
  - Incremental updates (announcements and withdrawals)
  - Policies applied at import and export of routes
  - Internal BGP to distribute information within an AS
  - Interaction with the IGP to compute forwarding tables