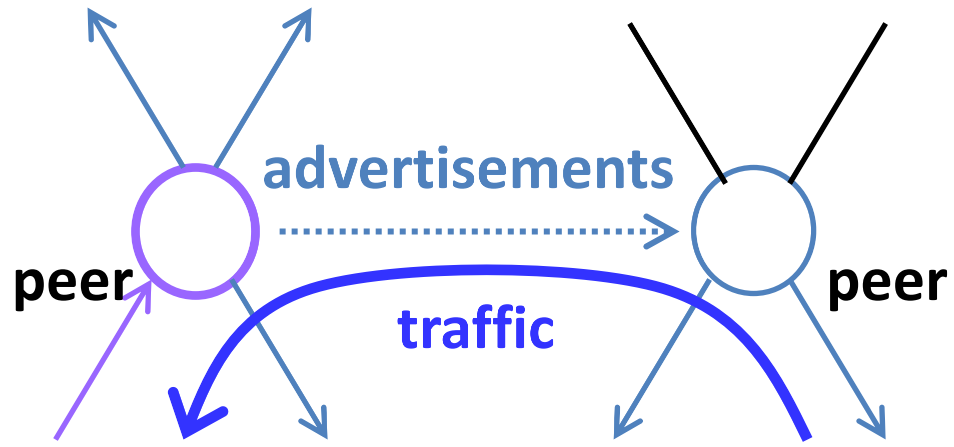


Interdomain Routing

Kyle Jamieson

Lecture 11

COS 461: Computer Networks

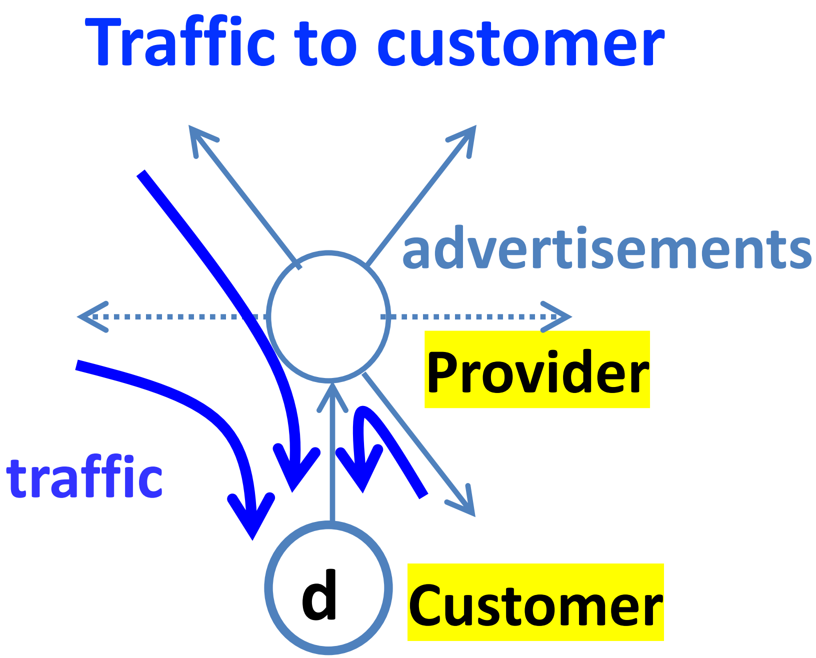
How to avoid BGP Instability

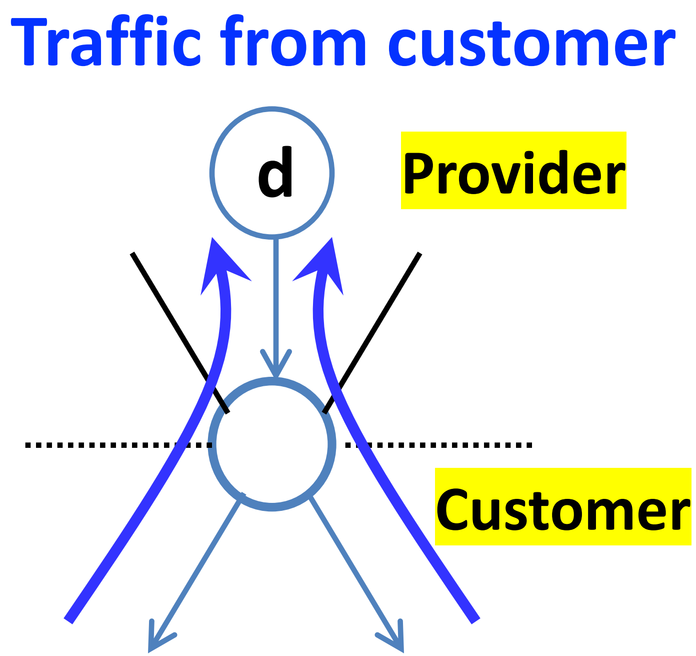
- Detecting conflicting policies
  - Con: Computationally expensive
  - Con: Requires too much cooperation
- Detecting oscillations
  - Observing the repetitive BGP routing messages
  - Con: Requires dynamic, stateful analysis
- **Restricted routing policies and topologies**
  - **Policies based on business relationships**

AS (Autonomous System) Business Relationships

Customer-Provider Relationship

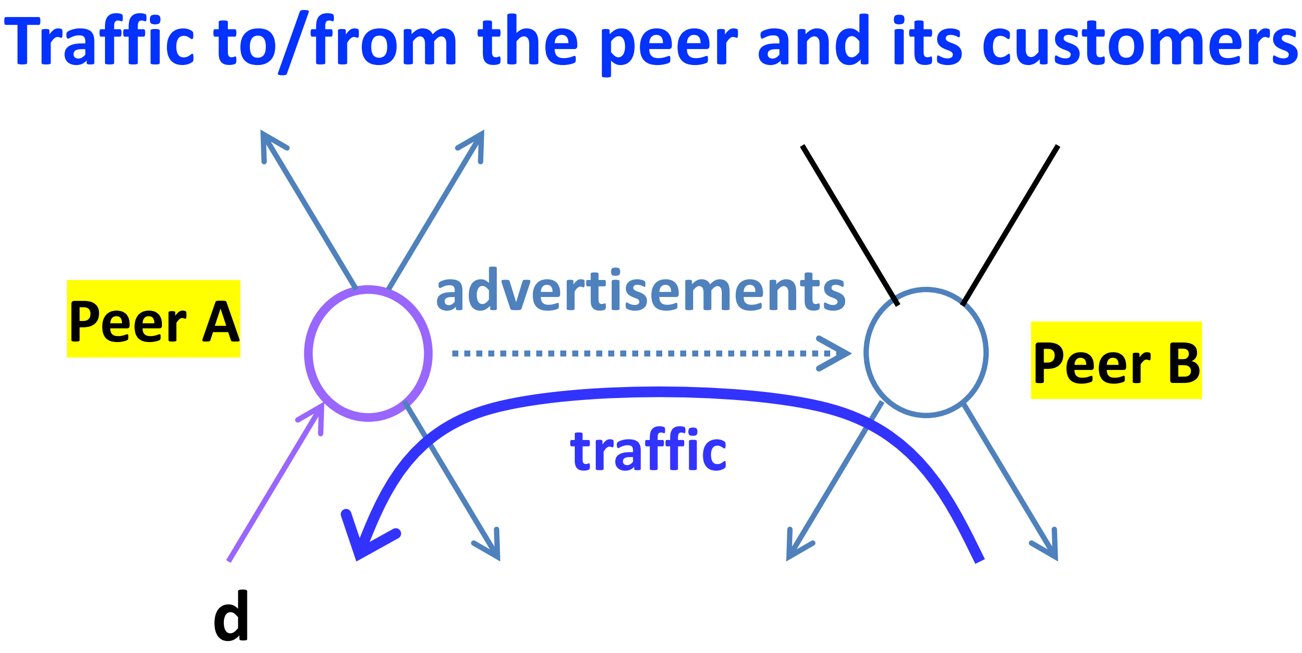
- Customer pays provider for access to Internet
  - Provider exports its customer routes to everybody
  - Customer exports provider routes only to its customers





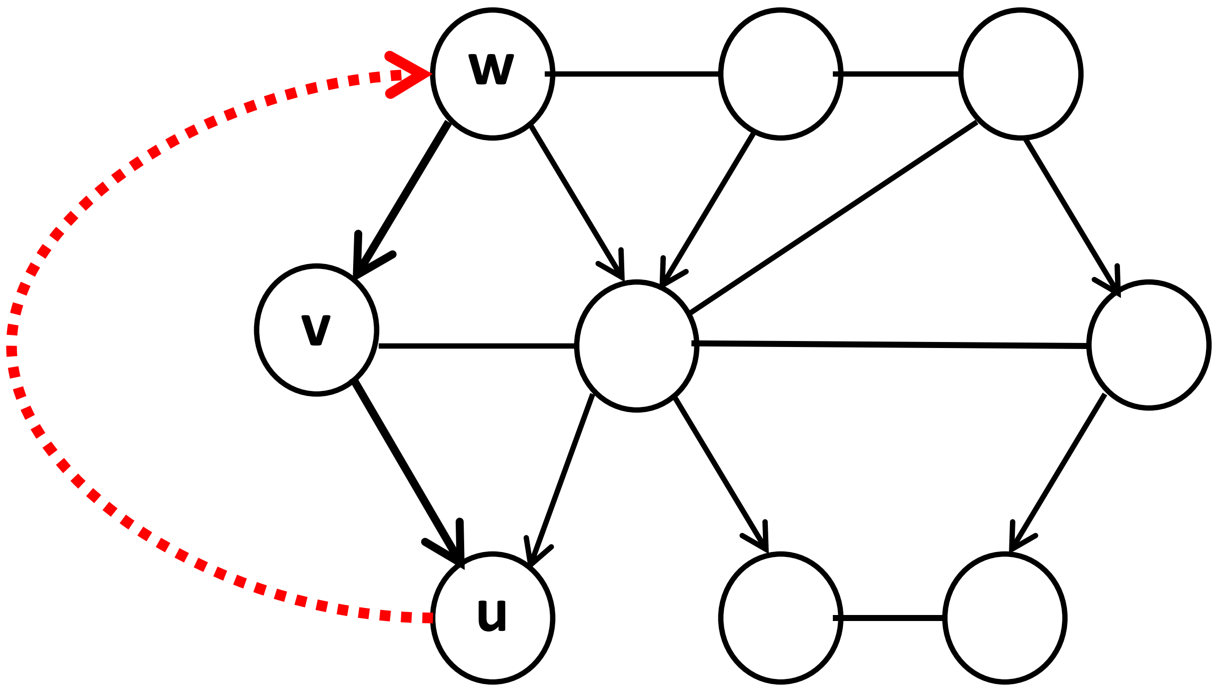
Peer-Peer Relationship

- Peers exchange traffic between their customers
  - AS exports **only customer** routes to a peer AS
  - AS exports a peer AS's routes **only** to **its customers**



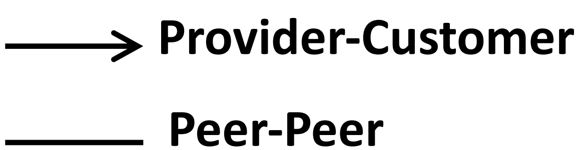
Hierarchical AS Relationships

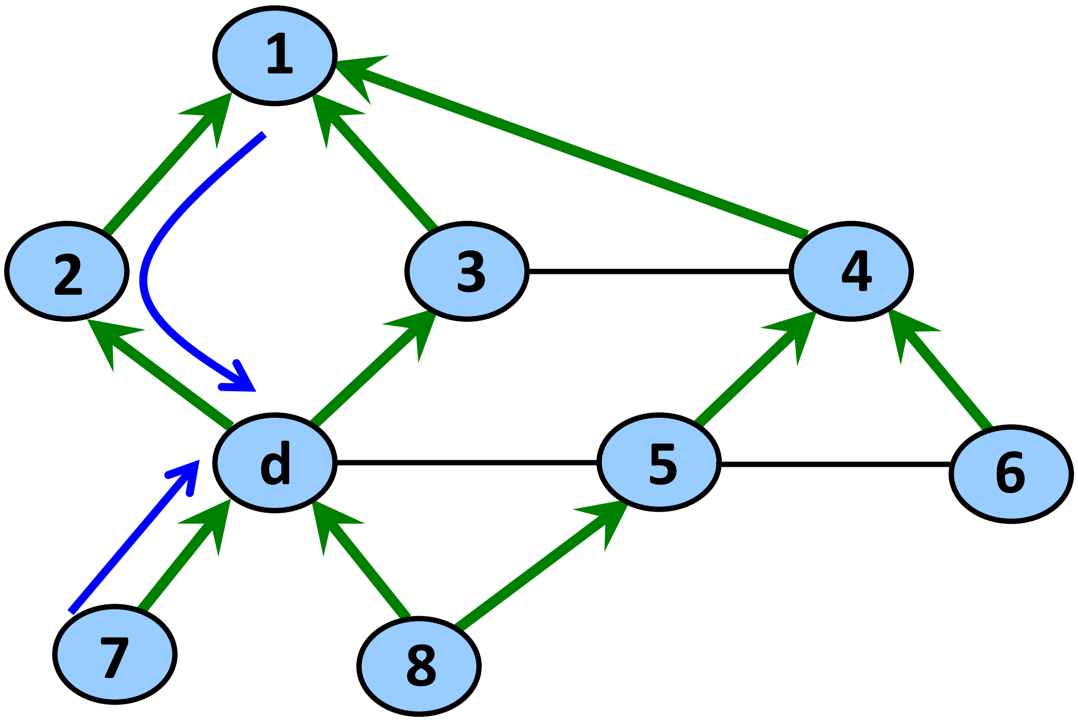
- Provider-customer graph is directed and acyclic
  - If u is a customer of v and v is a customer of w
  - … then w is not a customer of u



Valid and Invalid Paths

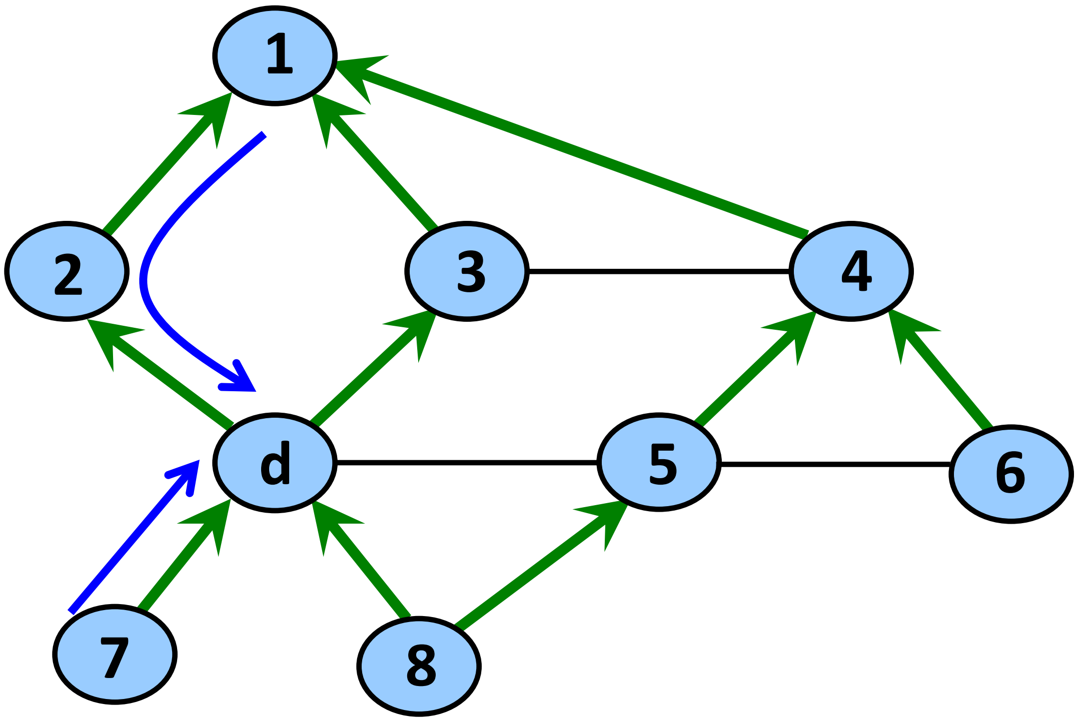
Path 1 2 d
Path 7 d
Path 5 8 d
Path 6 4 3 d
Path 8 5 d
Path 6 5 d
Path 1 4 3 d





Valid and Invalid Paths

Path 1 2 d	Valid
Path 7 d	Valid
Path 5 8 d	Invalid
Path 6 4 3 d	Valid
Path 8 5 d	Valid
Path 6 5 d	Invalid
Path 1 4 3 d	Invalid



→ Provider-Customer

— Peer-Peer

Local Control, Global Stability: "Gao-Rexford Conditions"

1. Route export
   – Don't export routes learned from a peer or provider to another peer or provider
2. Global topology
   – Provider-customer relationship graph is acyclic
   – E.g., my customer's customer is not my provider
3. Route selection
   – Prefer routes through customers over routes through peers and providers

Guaranteed to converge to unique, stable solution

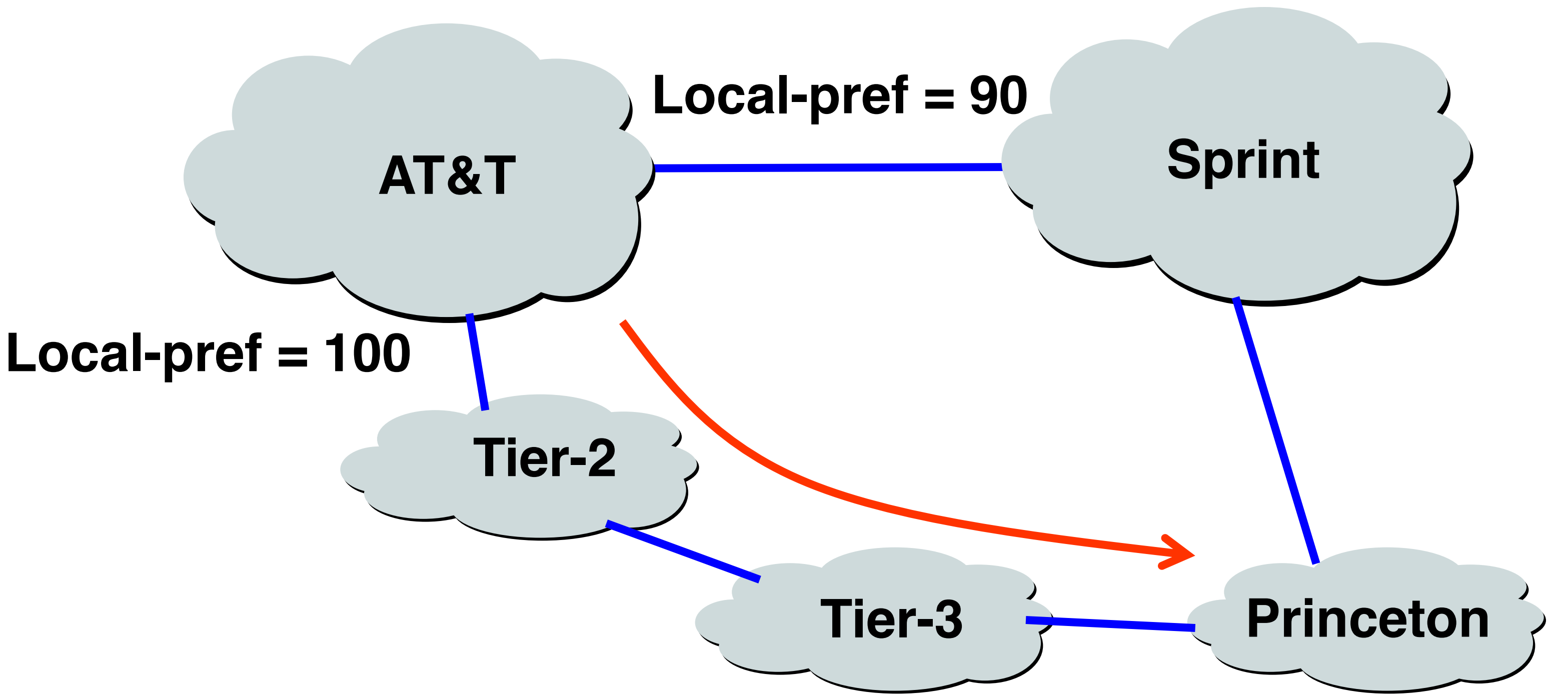
How do we implement Interdomain Routing Policy?

Selecting a Best Path

- Routing Information Base
  - Store all BGP routes for each destination prefix
  - Withdrawal: remove the route entry
  - Announcement: update the route entry
- BGP decision process
  - Highest local preference
  - Shortest AS path
  - Closest egress point
  - Arbitrary tie break

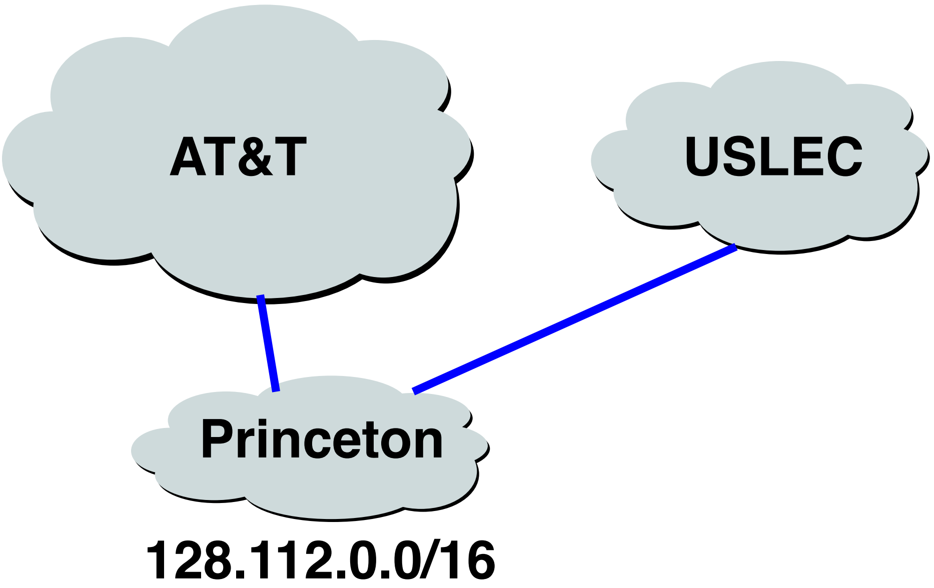
Import Policy: Local Preference

- Favor one path over another
  - Override the influence of AS path length
- Example: prefer customer over peer



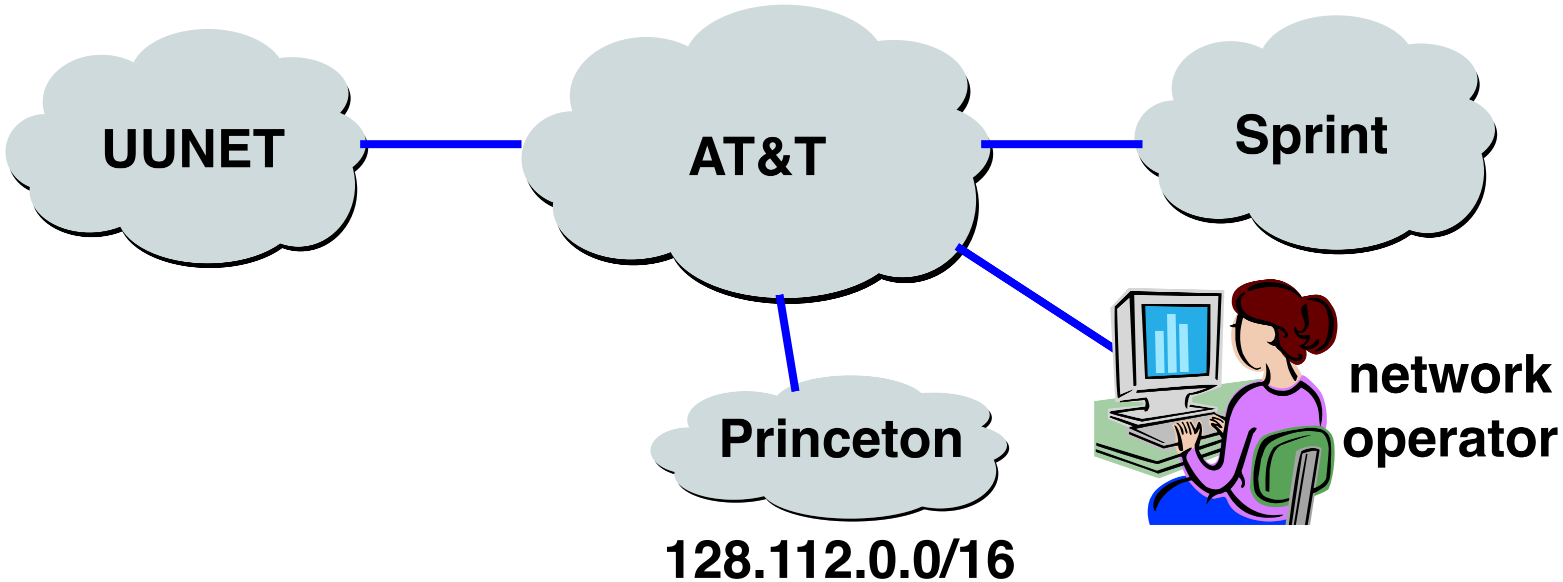
Import Policy: Filtering

- 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 with other large ISP in the AS path



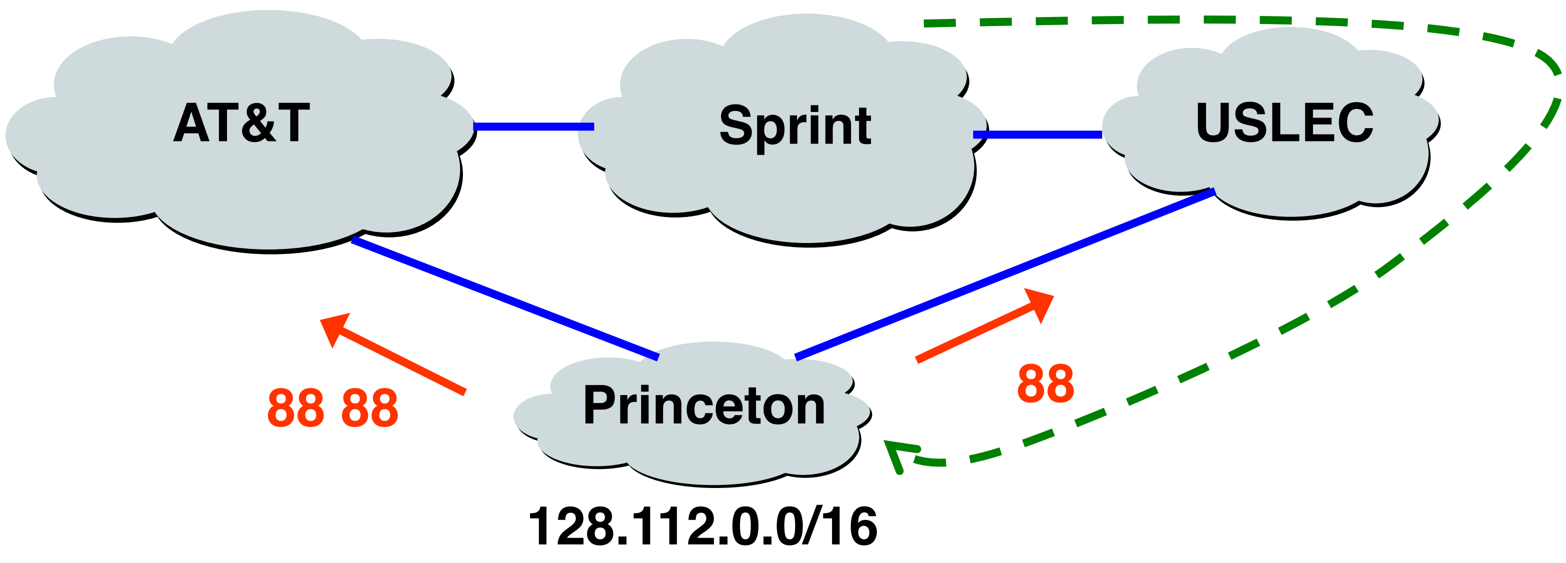
Export Policy: Filtering

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



Export Policy: Attribute Manipulation

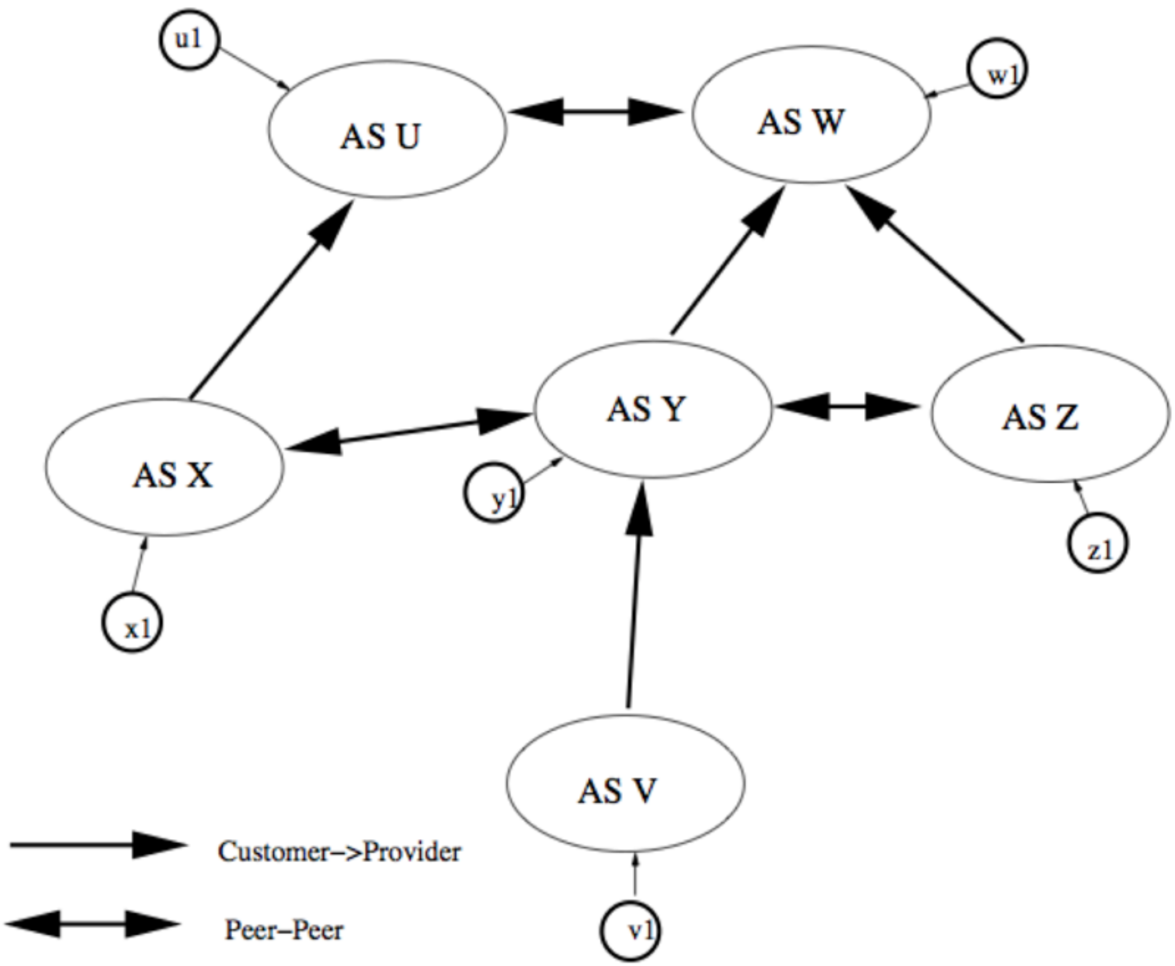
- Modify attributes of the active route
  - To influence the way other ASes behave
- Example: AS prepending
  - Artificially inflate AS path length seen by others
  - Convince some ASes to send traffic another way



Reflect Business Relationships

- Common relationships
  - Customer-provider
  - Peer-peer
  - Backup, sibling, …
- ISP terminology:
  - Tier-1 (~15 worldwide): No settlement or transit
  - Tier-2 ISPs: Widespread peering, still buy transit
- Policies implementing in BGP, e.g.,
  - Import: Ranking customer routes over peer routes
  - Export: Export only customer routes to peers and providers

BGP Policy



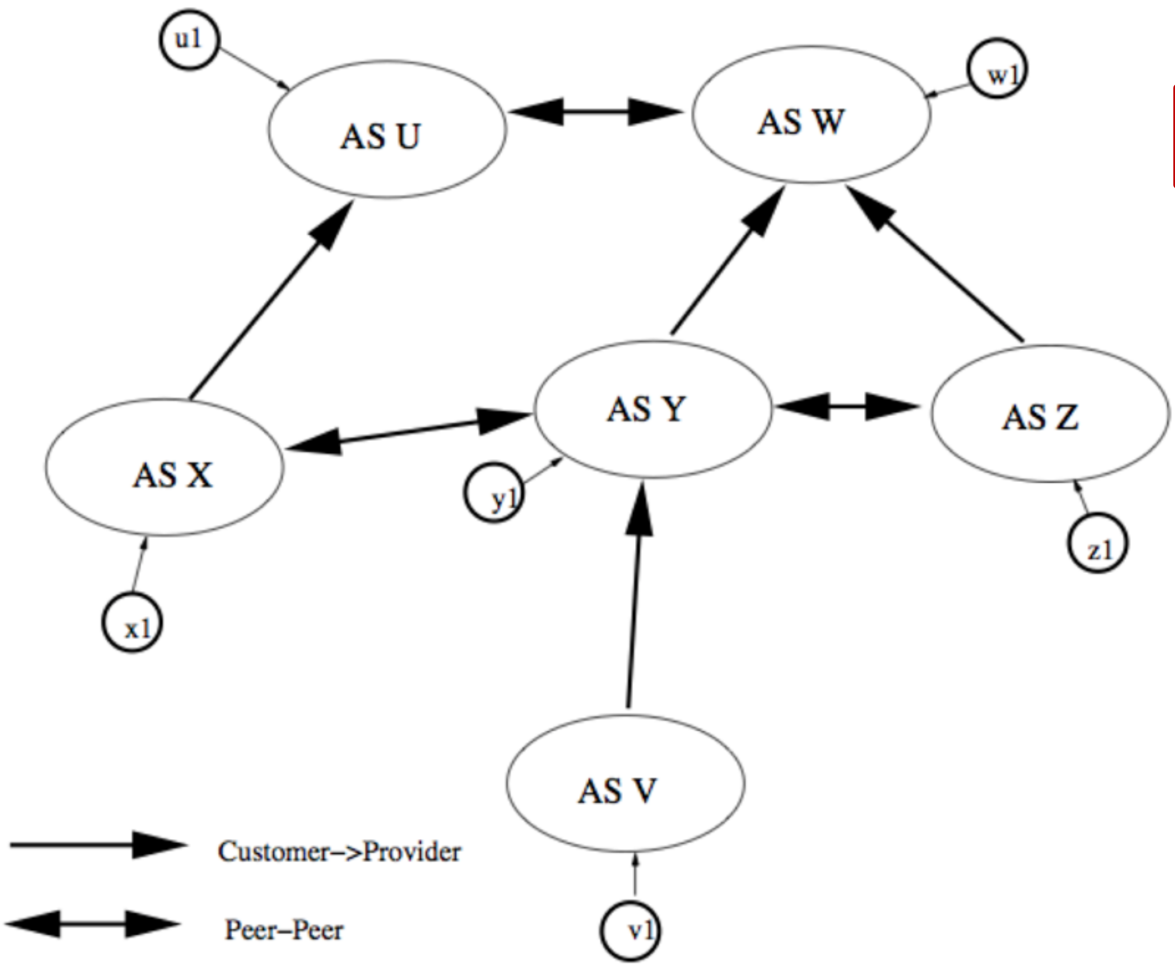
Tier 1 ISPs?

Y. U, W
M. U, X
C. X, Y, Z

Which path may packets take (given commercial policies)?

Y. Red
M. Blue
C. Green
A. Orange

BGP Policy



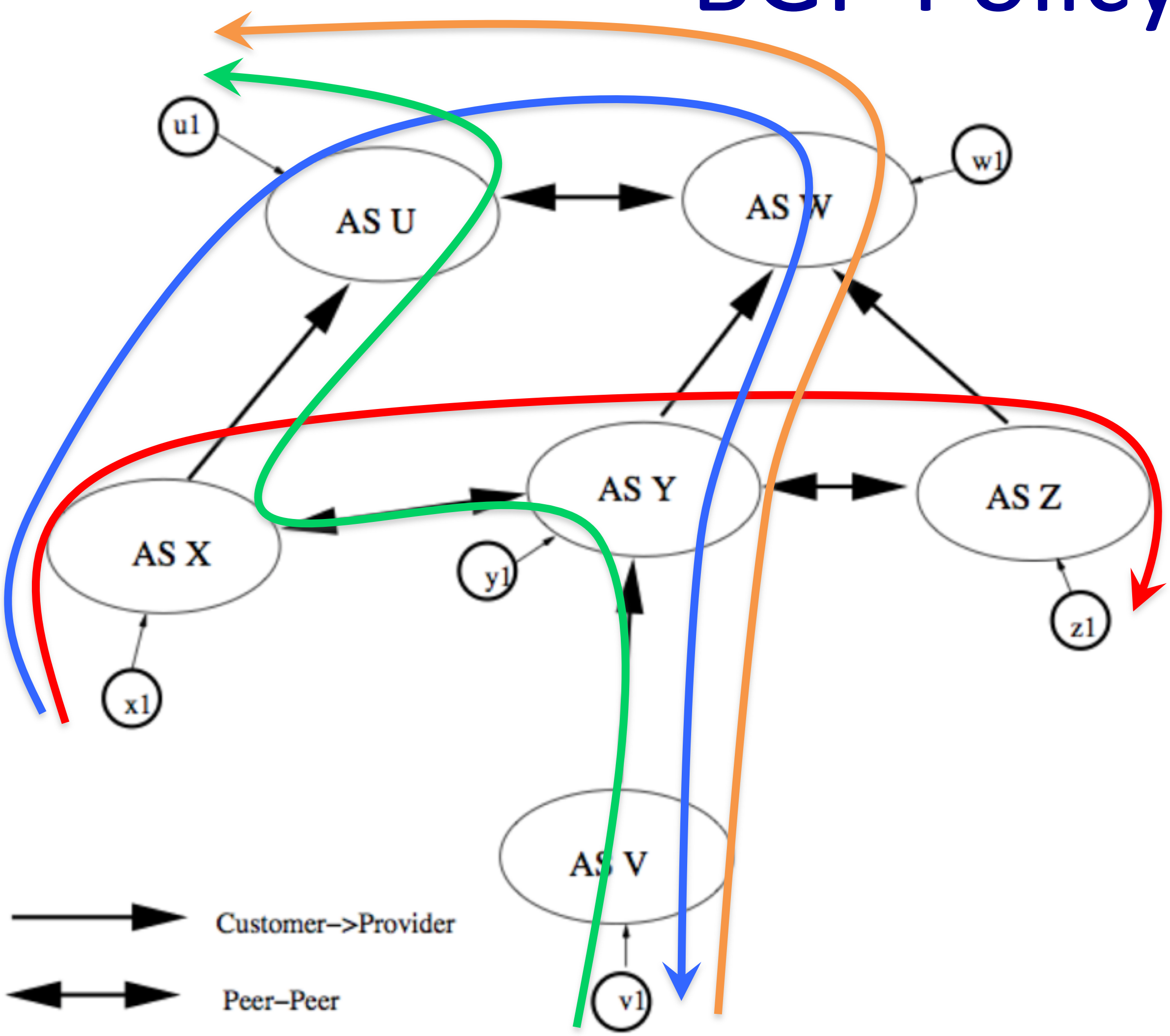
Tier 1 ISPs?

Y. U, W
M. U, X
C. X, Y, Z

Which path may packets take (given commercial policies)?

Y. Red
M. Blue
C. Green
A. Orange

BGP Policy



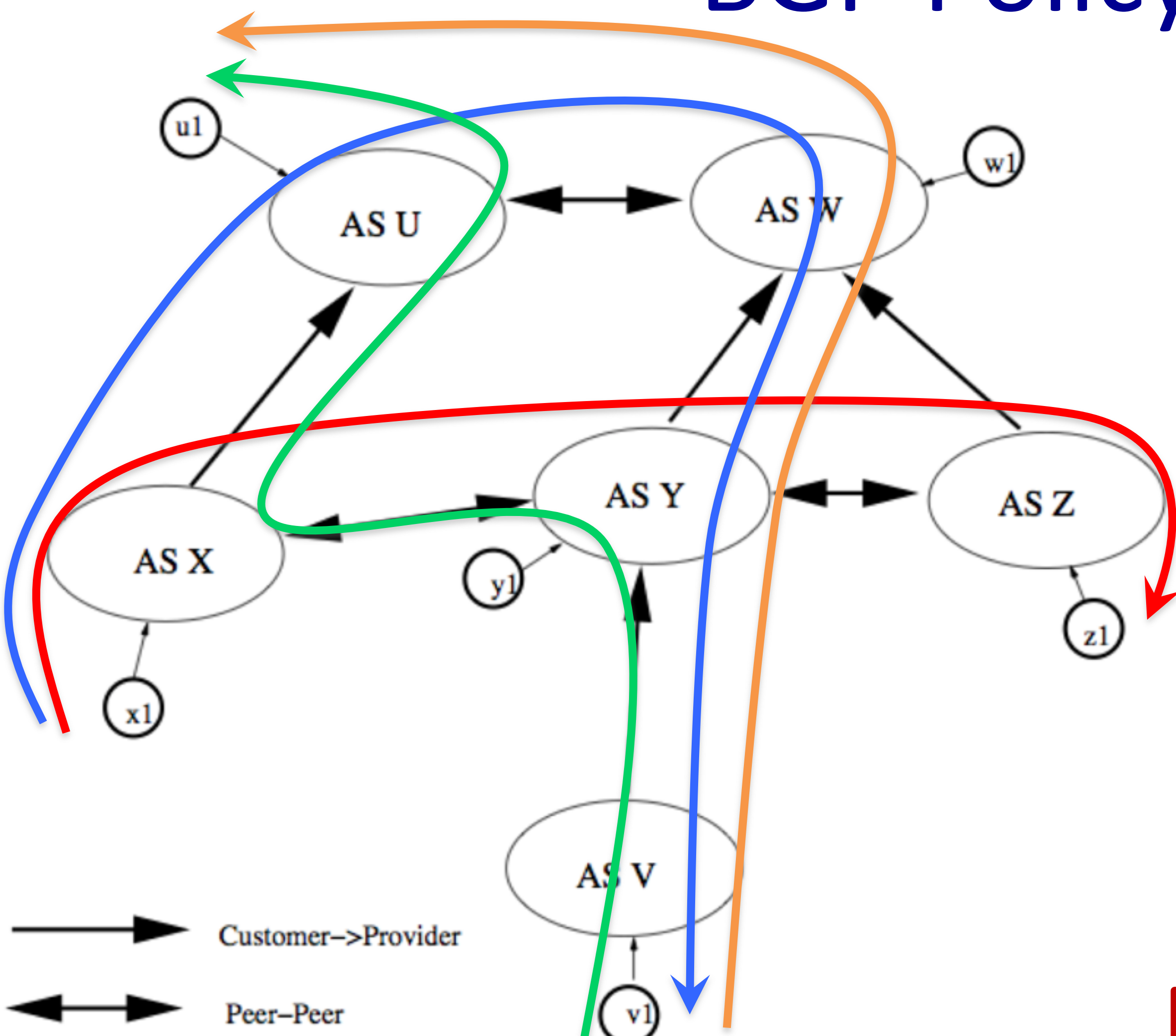
Tier 1 ISPs?

Y. U, W
M. U, X
C. X, Y, Z

Which path may packets take (given commercial policies)?

Y. Red
M. Blue
C. Green
A. Orange

BGP Policy



Tier 1 ISPs?

Y. U, W
M. U, X
C. X, Y, Z

Which path may packets take (given commercial policies)?

Y. Red
M. Blue
C. Green
A. Orange

BGP Policy Configuration

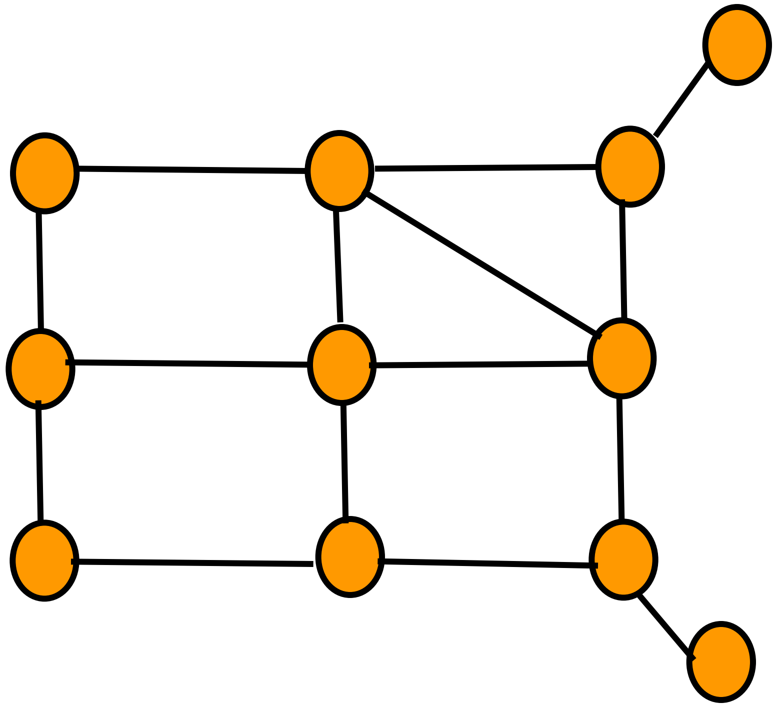
- Routing policy languages are vendor-specific
  – Not part of the BGP protocol specification
- Still, all languages have some key features
  – List of clauses matching on route attributes
  – … and discarding or modifying the matching routes
- Configuration done by human operators
  – Implementing the policies of their AS
  – Business relationships, traffic engineering, security

How do backbone ASs operate?

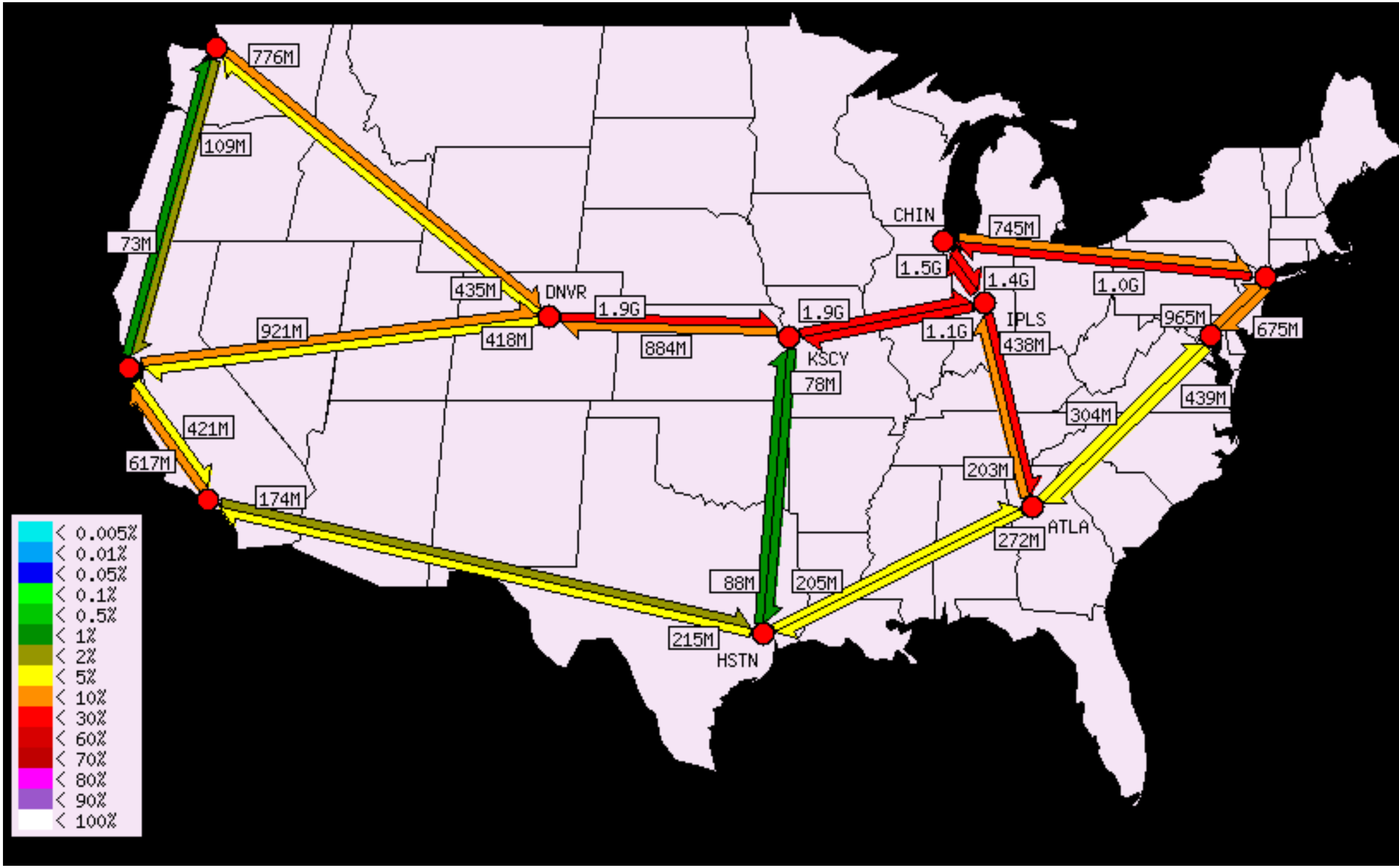
Backbone Topology

Backbone Networks

- Backbone networks
  - Multiple Points-of-Presence (PoPs)
  - Lots of communication between PoPs
  - Accommodate traffic demands and limit delay

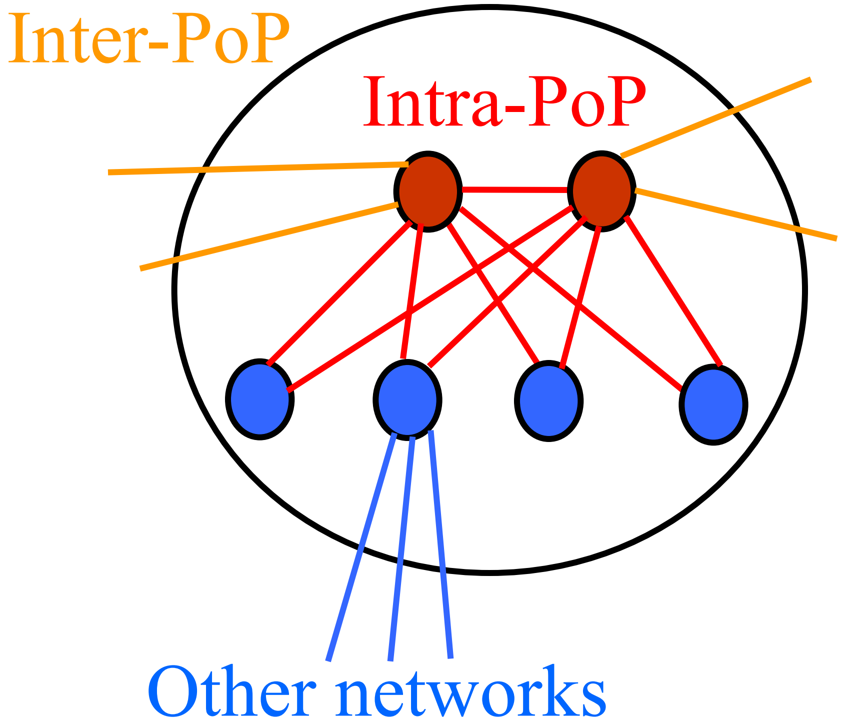


Abilene Internet2 Backbone



Points-of-Presence (PoPs)

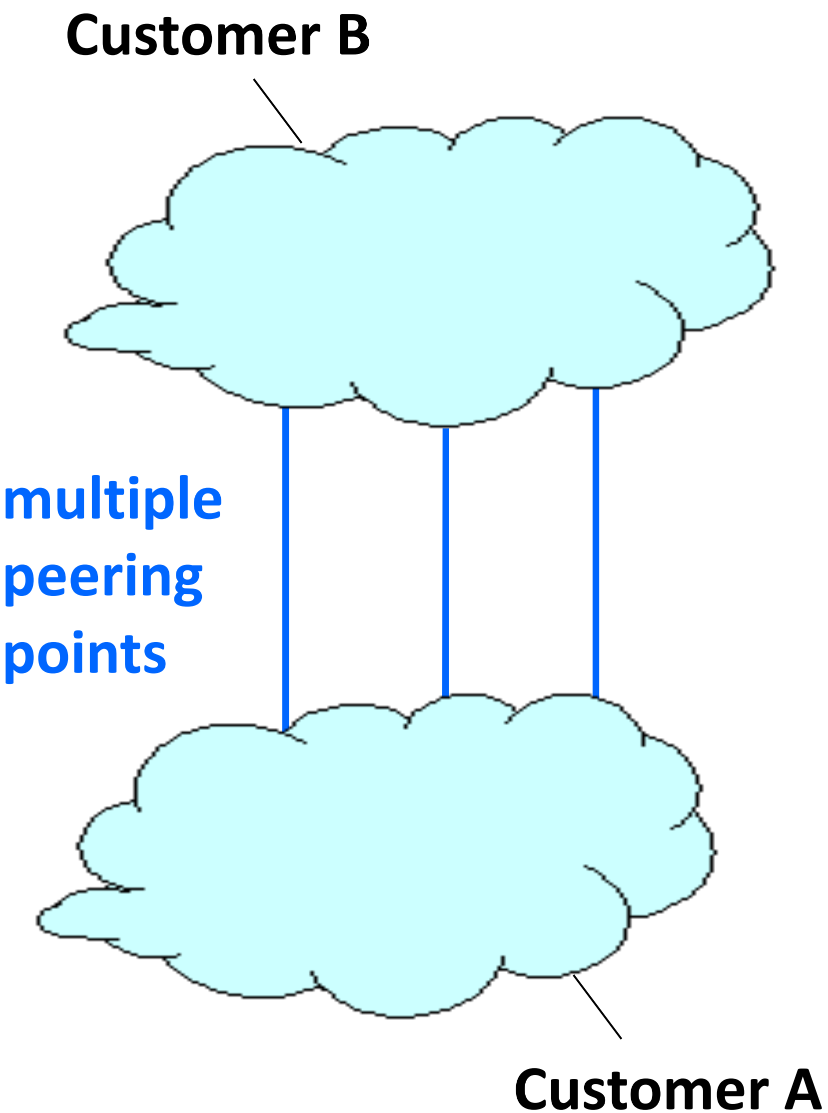
- Inter-PoP links
  - Long distances
  - High bandwidth
- Intra-PoP links
  - Short cables between racks or floors
  - Aggregated bandwidth
- Links to other networks
  - Wide range of media and bandwidth



Where to Locate Nodes and Links

- Placing Points-of-Presence (PoPs)
  – Large population of potential customers
  – Other providers or exchange points
  – Cost and availability of real-estate
  – Mostly in major metropolitan areas
- Placing links between PoPs
  – Already fiber in the ground
  – Needed to limit propagation delay
  – Needed to handle the traffic load

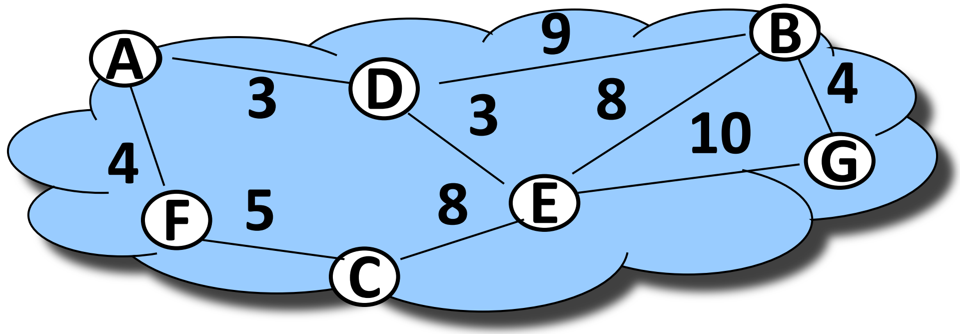
Peering



- Exchange traffic between customers
  - Settlement-free
- Diverse peering locations
  - Both coasts, and middle
- Comparable capacity at all peering points
  - Can handle even load

Combining Intradomain and Interdomain Routing

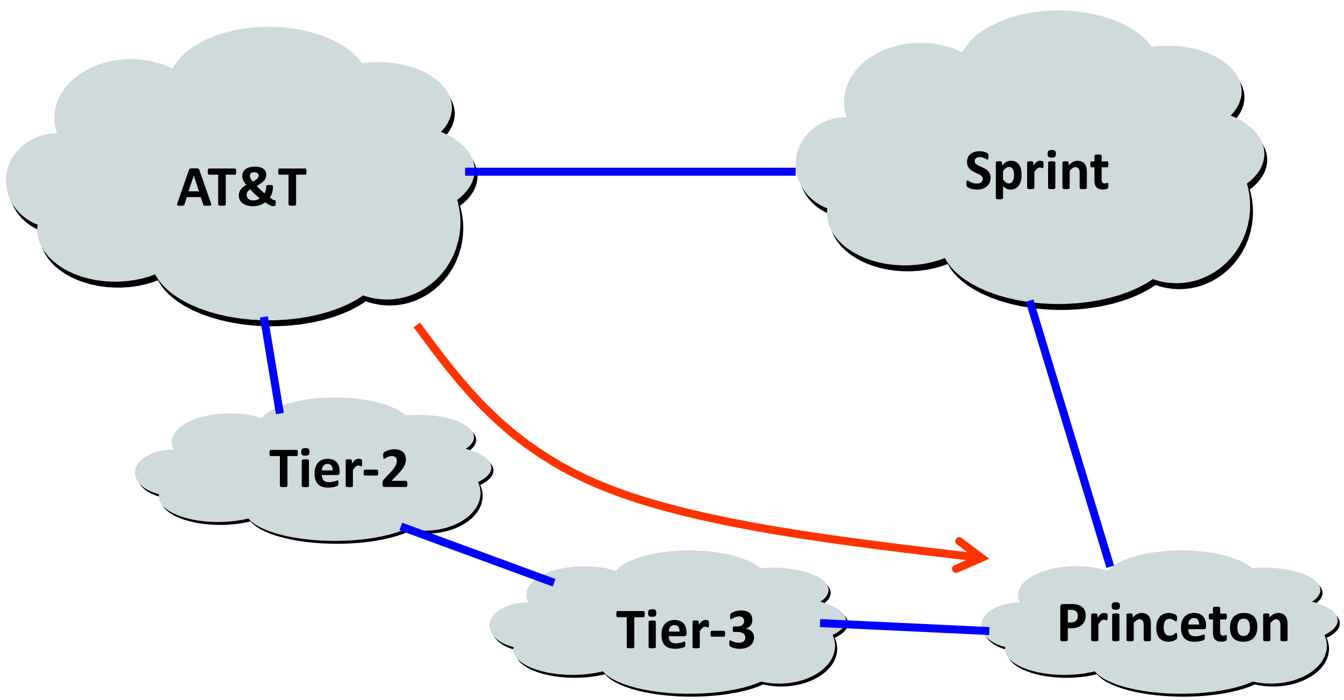
Intradomain Routing



- ***Interior Gateway Protocol (IGP)*** computes shortest paths between routers in same AS
  - Router C takes path C-F-A to router A
- Using link-state routing protocols
  - E.g., OSPF, IS-IS

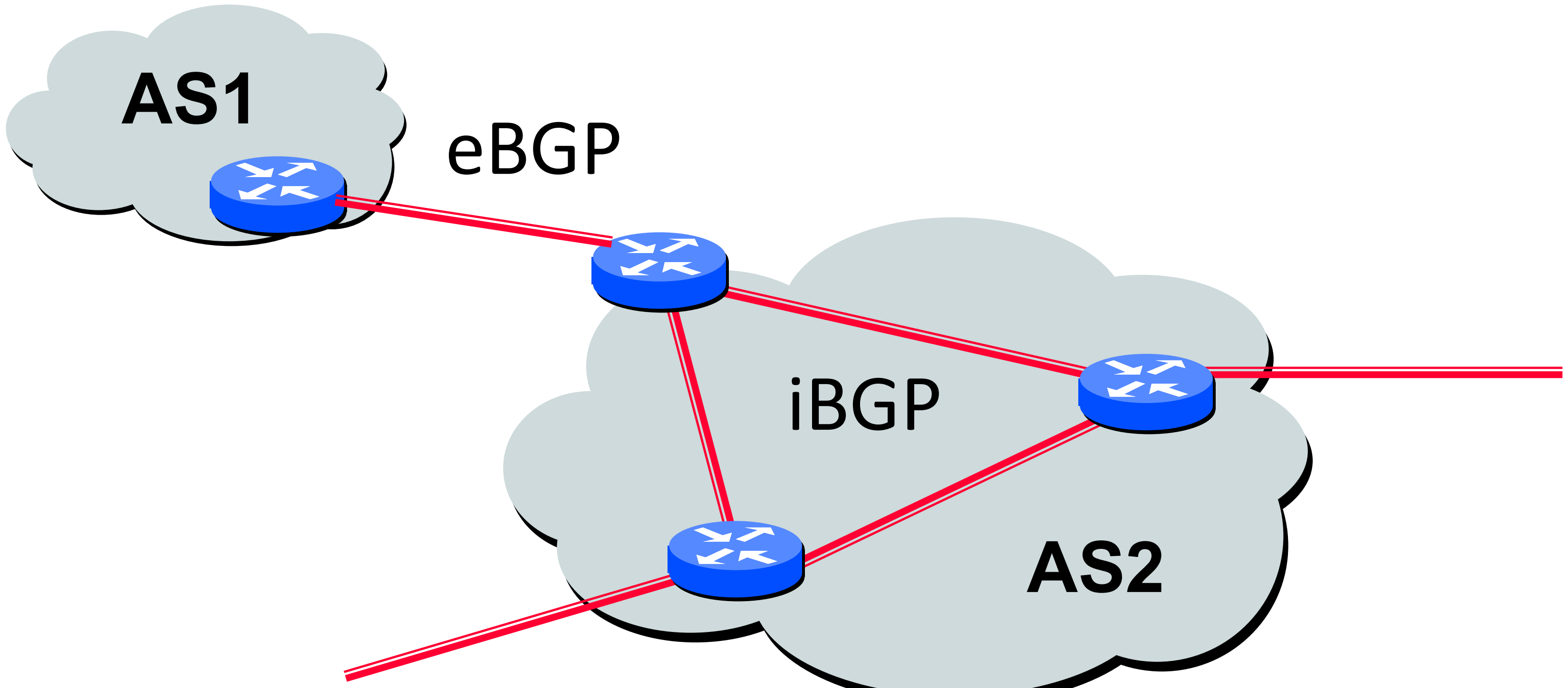
Interdomain Routing

- Learn paths to remote destinations
  - AT&T learns two paths to Princeton
- Applies local policies to select a best route



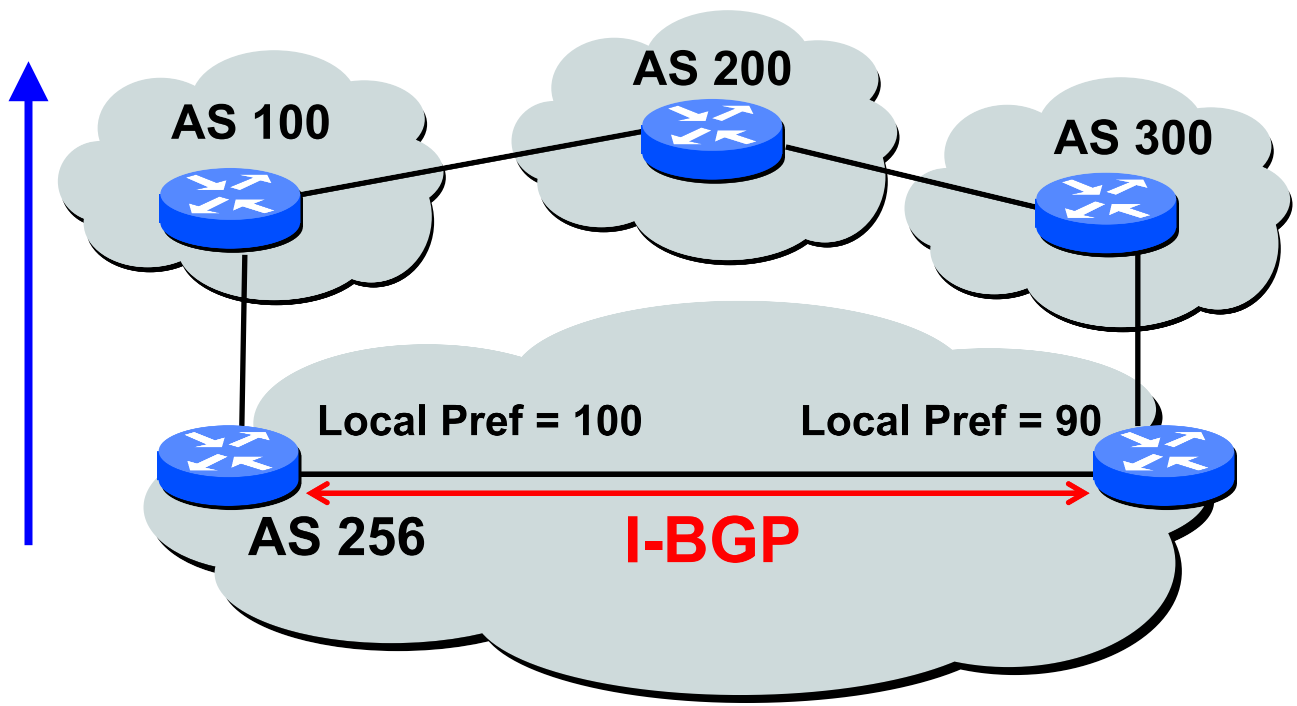
An AS is Not a Single Node

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



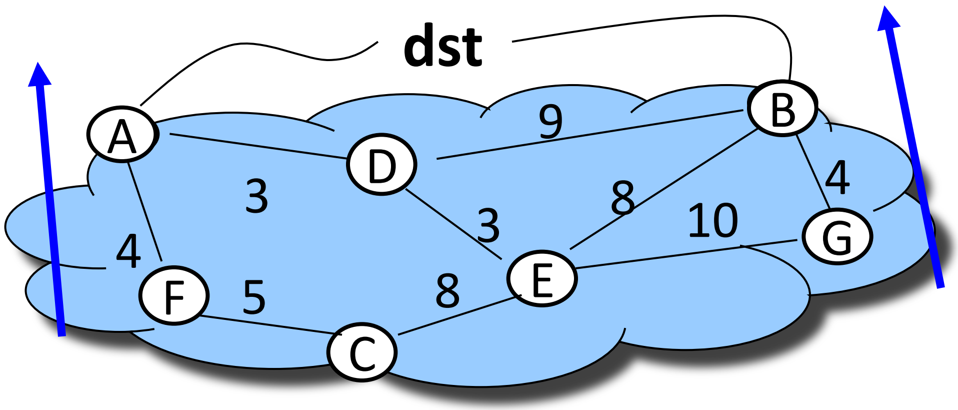
Internal BGP and Local Preference

- Both routers prefer path through AS 100
- … even though router on right-hand-side learns external path

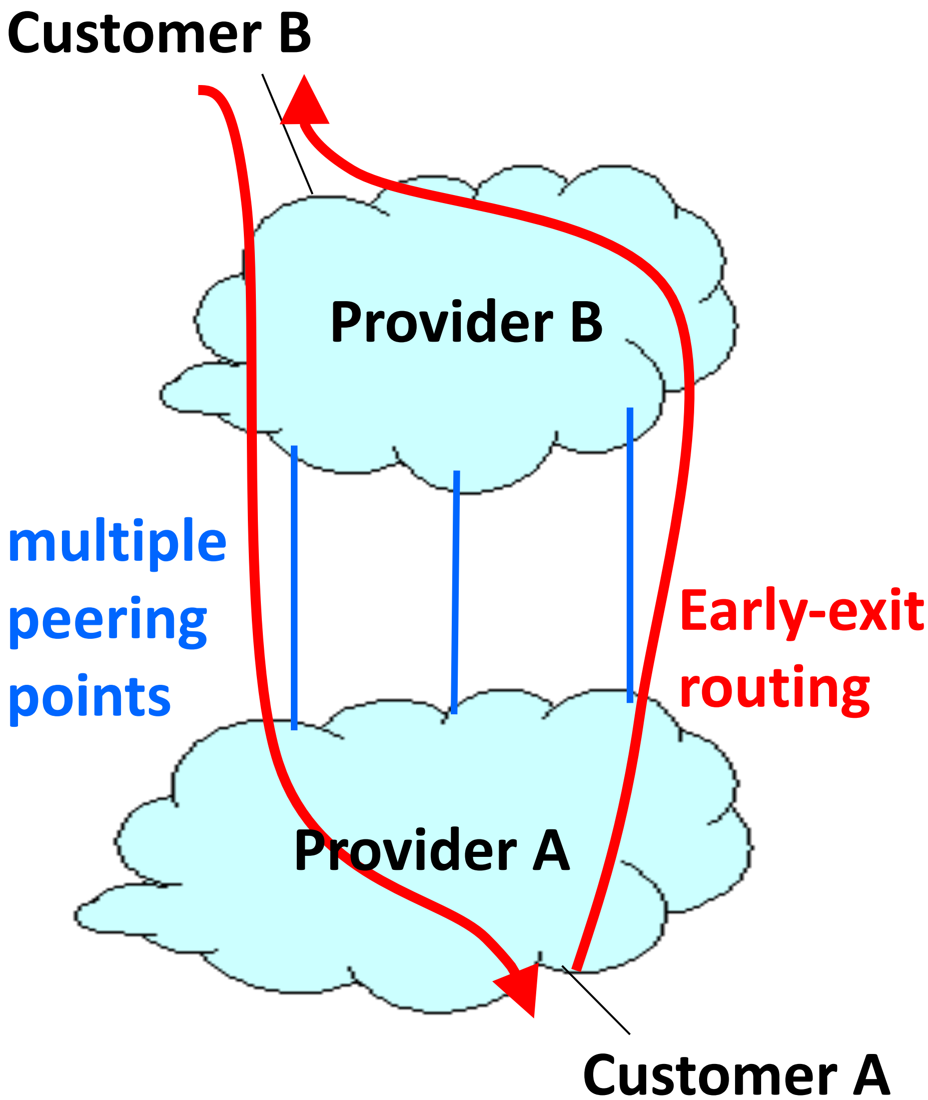


Hot-Potato (Early-Exit) Routing

- Hot-potato routing
  - Each router selects closest egress point based on IGP path cost
- BGP decision process
  - Highest local preference
  - Shortest AS path
  - Closest egress point
  - Arbitrary tie break

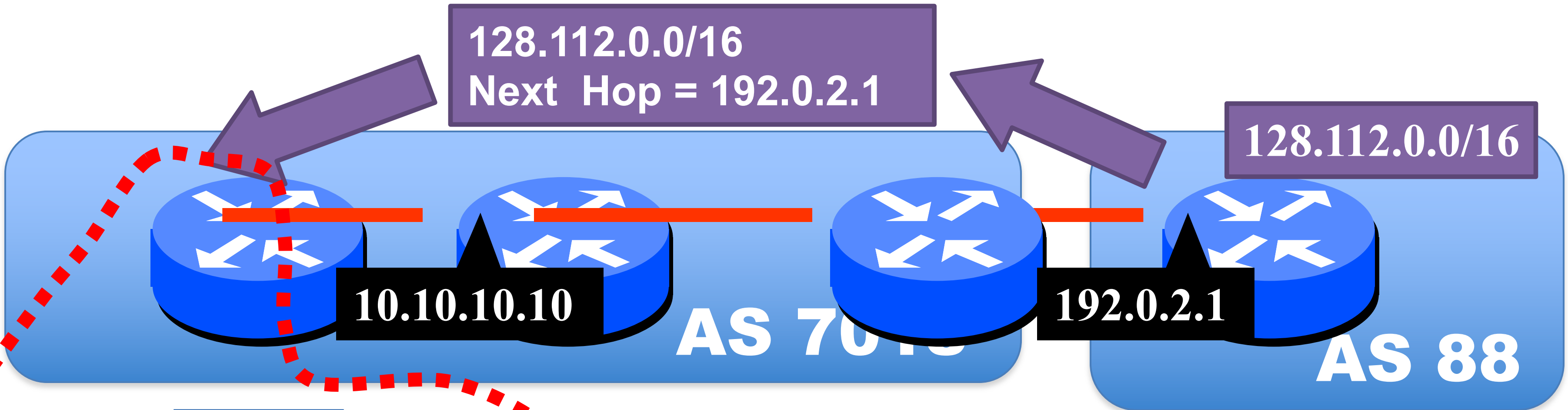


Hot-Potato Routing



- Selfish routing
  - Each provider dumps traffic on the other
  - As early as possible
- Asymmetric routing
  - Traffic does not flow on same path in both directions

Joining BGP with IGP Information



IGP

destination	next hop
192.0.2.0/30	10.10.10.10

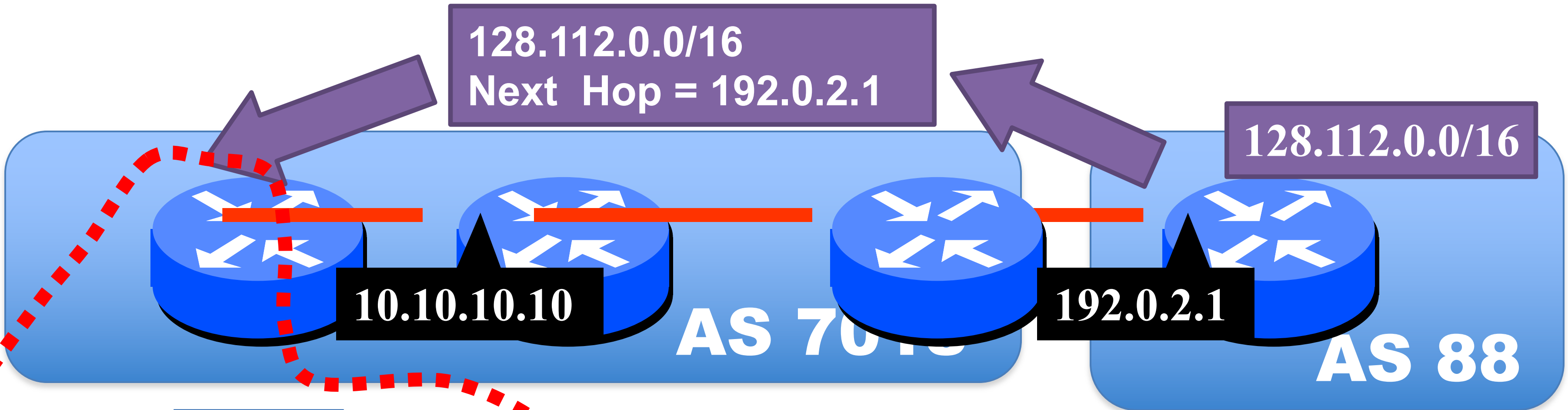
+

BGP

destination	next hop
128.112.0.0/16	192.0.2.1

- The FIB of internal routers are of size O(all dest prefixes known to ISP)
- The FIB of internal routers point to border router to neighbor ISP

Joining BGP with IGP Information



IGP

destination	next hop
192.0.2.0/30	10.10.10.10

+

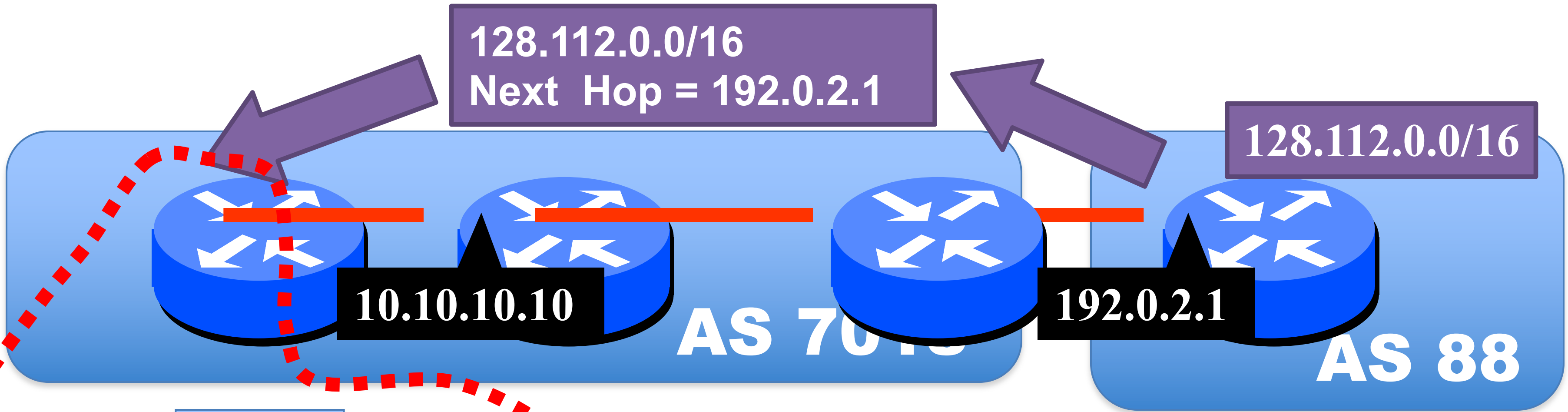
BGP

destination	next hop
128.112.0.0/16	192.0.2.1

The FIB of internal routers are of size O(all dest prefixes known to ISP) **TRUE**

- The FIB of internal routers point to border router to neighbor ISP **FALSE**

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

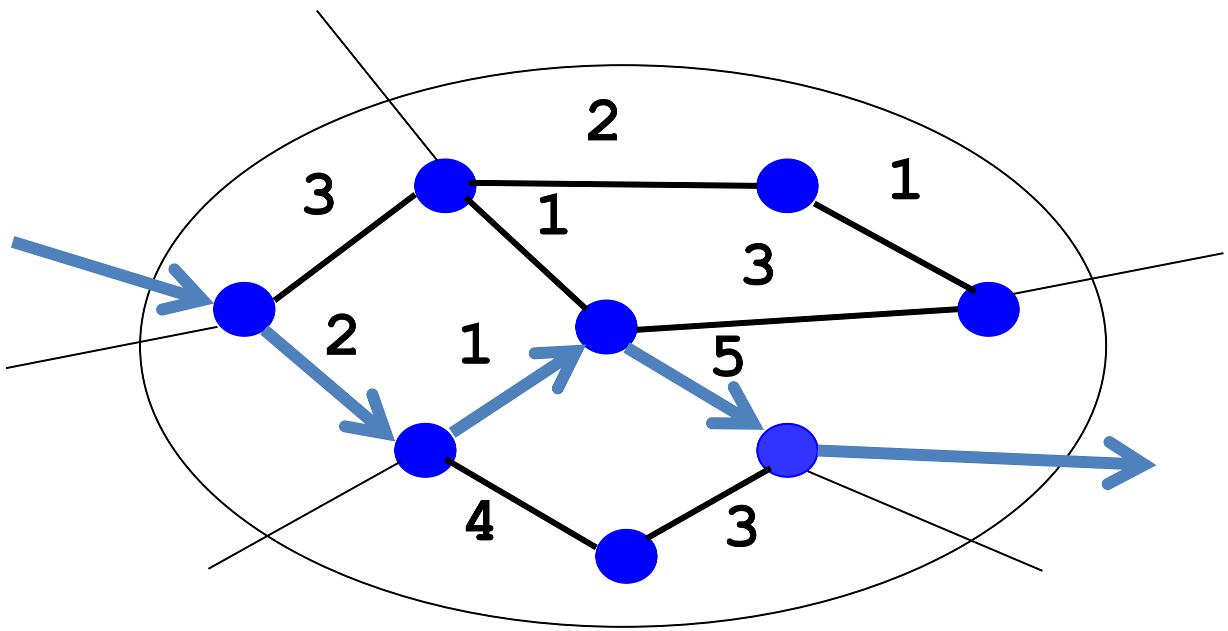
Forwarding Table

destination	next hop
128.112.0.0/16	10.10.10.10
192.0.2.0/30	10.10.10.10

Backbone Traffic Engineering

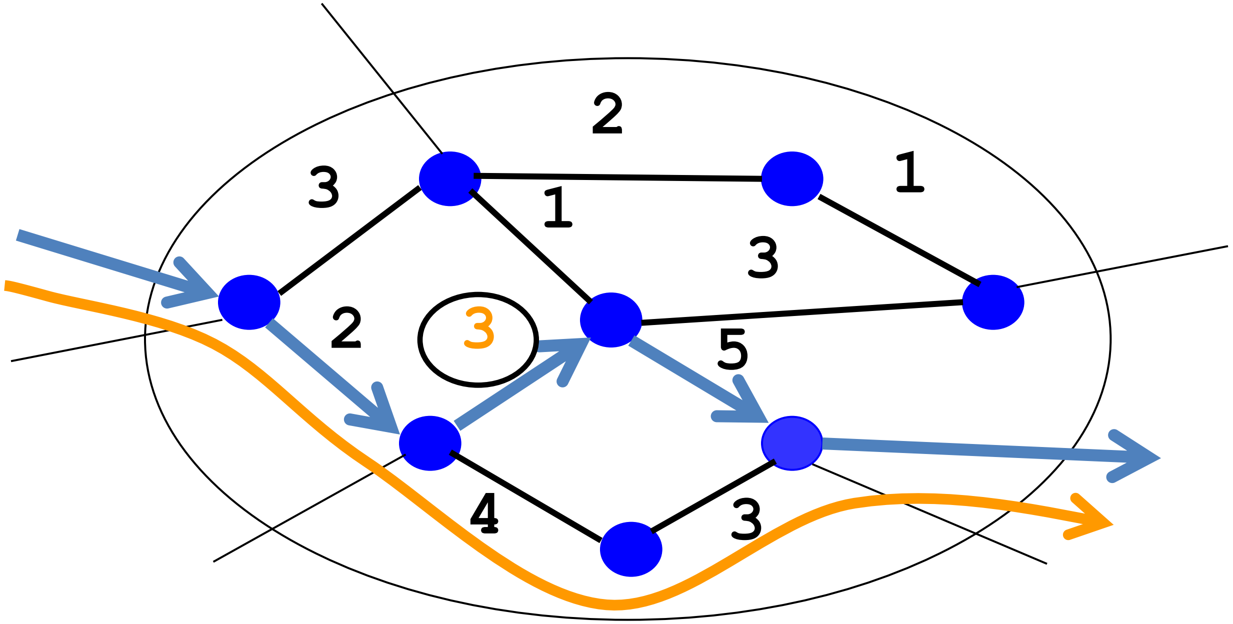
Routing With "Static" Link Weights

- Routers flood information to learn topology
  - Determine "next hop" to reach other routers…
  - Compute shortest paths based on link weights
- Link weights configured by network operator

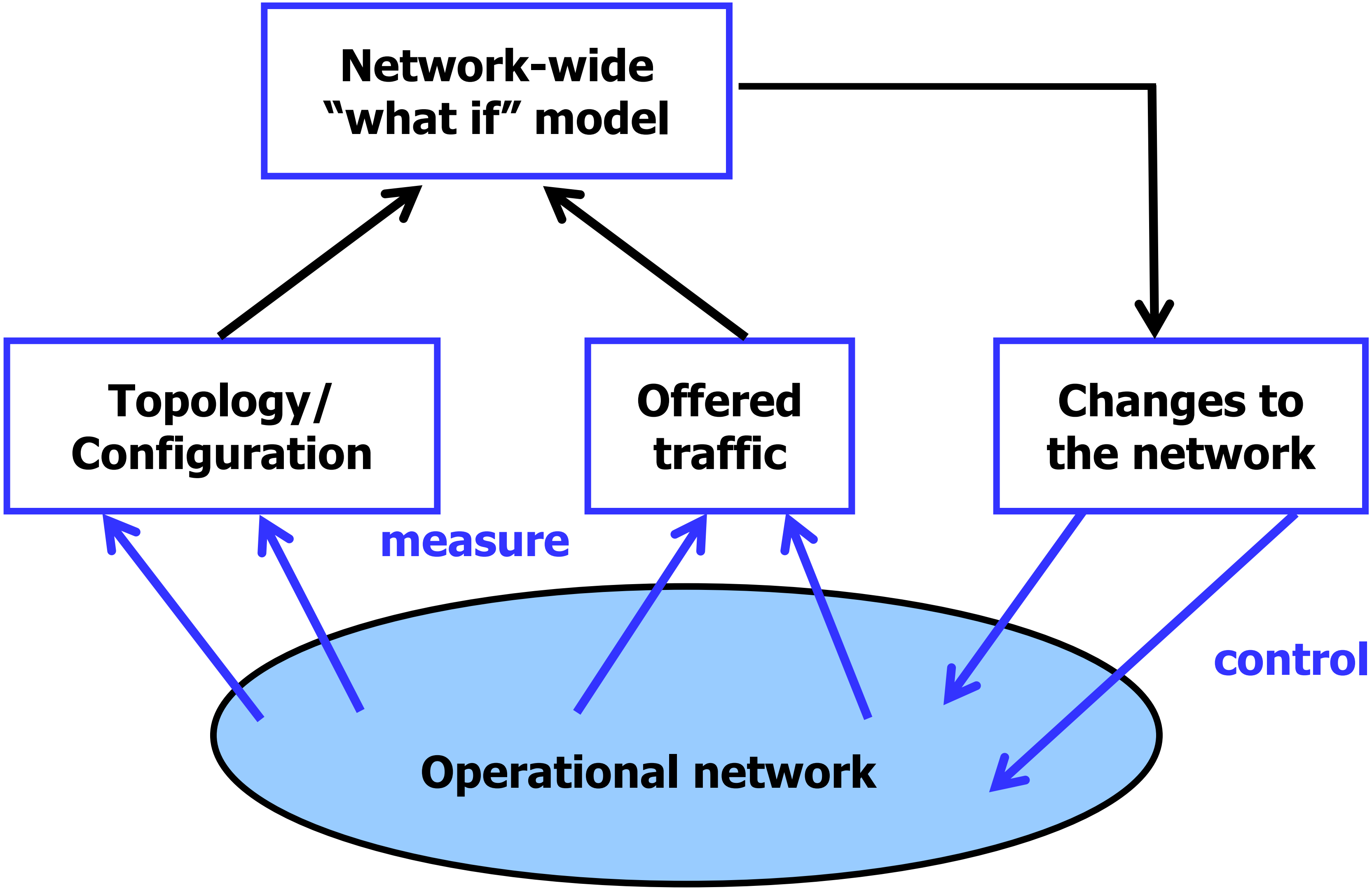


Setting the Link Weights

- How to set the weights
  - Inversely proportional to link capacity?
  - Proportional to propagation delay?
  - Network-wide optimization based on traffic?



Measure, Model, and Control



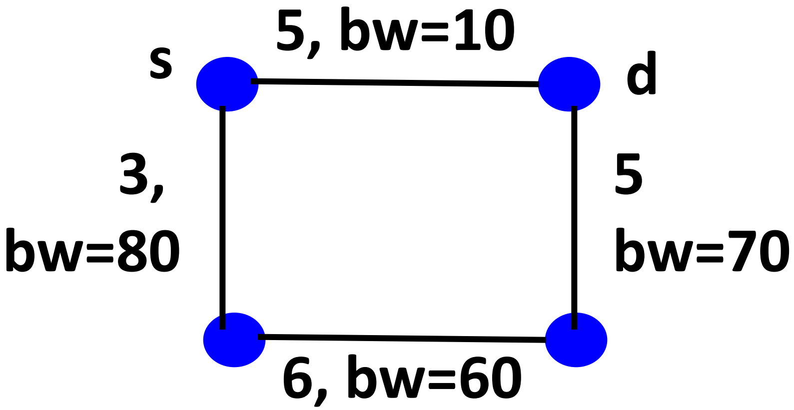
Limitations of Shortest-Path Routing

- Sub-optimal traffic engineering
  - Restricted to paths expressible as link weights
- Limited use of multiple paths
  - Only equal-cost multi-path, with even splitting
- Disruptions when changing the link weights
  - Transient packet loss and delay, and out-of-order
- Slow adaptation to congestion
  - Network-wide re-optimization and configuration
- Overhead of the management system

Constrained Shortest Path First

- Run a link-state routing protocol
  - Configurable link weights
  - Plus other metrics like available bandwidth

- Constrained shortest-path computation
  - Prune unwanted links (e.g., not enough bw)
  - Compute shortest path on the remaining graph



Conclusions

- Interdomain routing
  - Business relationships reflected in interdomain routing, leads to more stable paths
  - Peering and transit key ideas between providers, peers, and customer AS
- Backbone networks
  - Transit service for customers
  - Combine inter and intradomain routing
  - Glue that holds the Internet together