Lecture 14: Interdomain Routing

CSE 123: Computer Networks
Alex C. Snoeren

HW 4 out next time

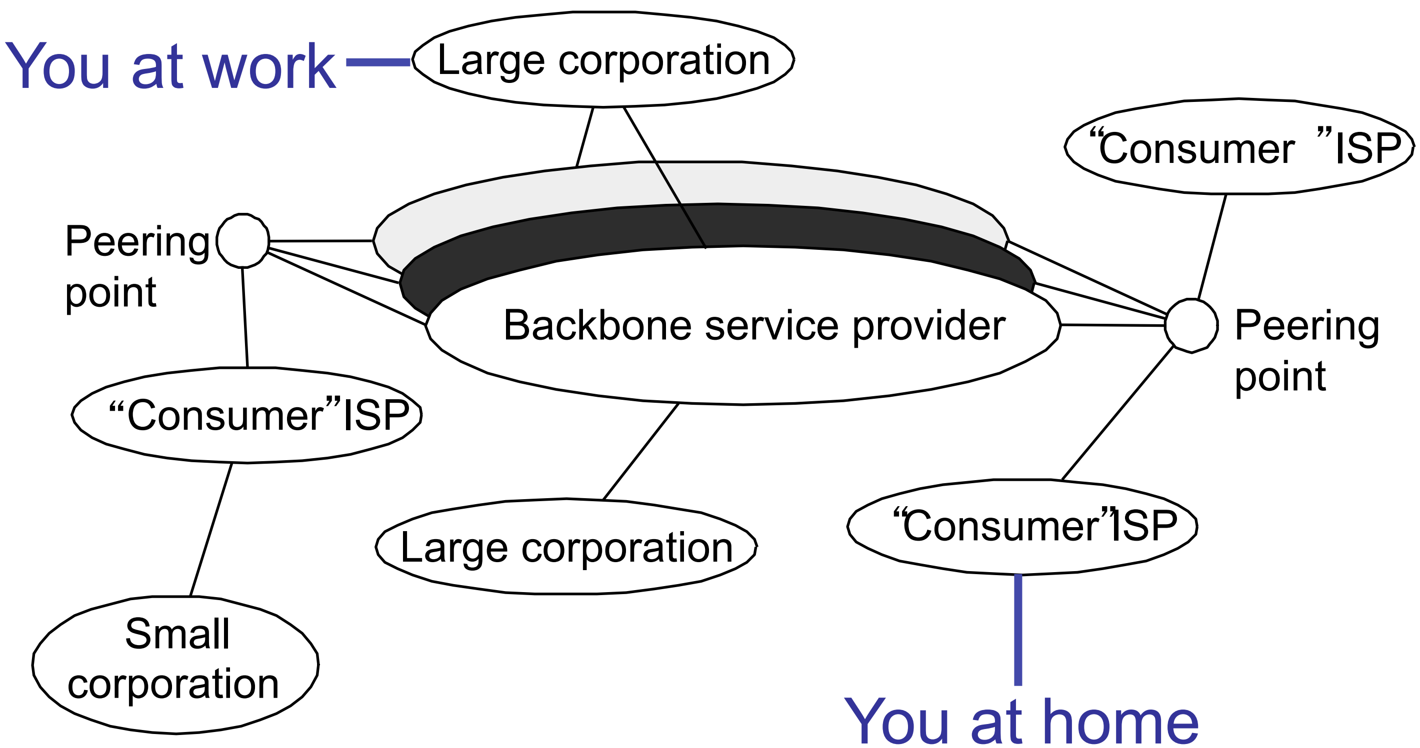
UCSDCSE

Lecture 14 Overview

- Autonomous Systems
  - Each network on the Internet has its own goals
- Path-vector Routing
  - Allows scalable, informed route selection
- Border Gateway Protocol
  - How routing gets done on the Internet today

The Internet is Complicated

- Inter-domain versus intra-domain routing

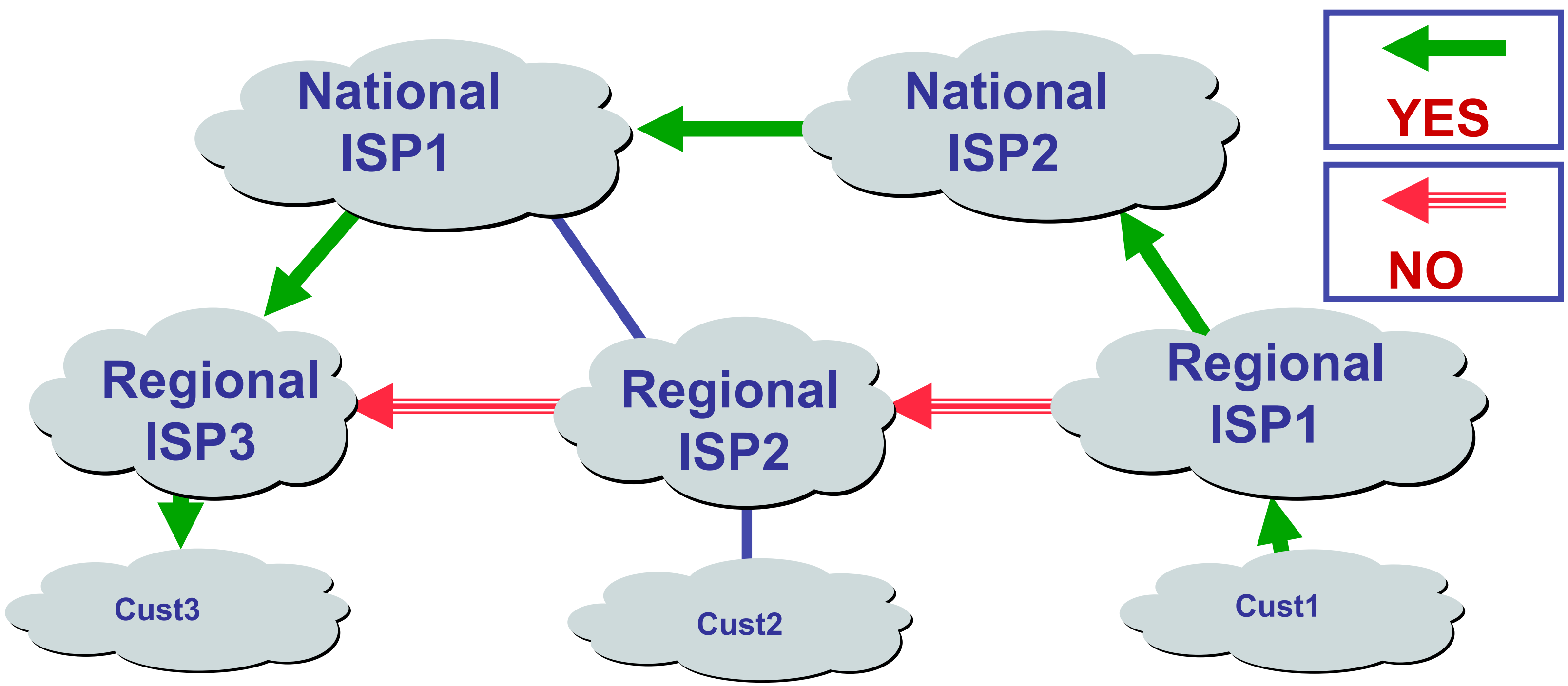


A Brief History

- Original ARPAnet had single routing protocol
  - Dynamic DV scheme, replaced with static metric LS algorithm
- New networks came on the scene
  - NSFnet, CSnet, DDN, etc…
  - The total number of nodes was growing exponentially
  - With their own routing protocols (RIP, Hello, ISIS)
  - And their own rules (e.g. NSF AUP)
- New requirements
  - Huge scale: millions of routers
  - Varying routing metrics
  - Need to express business realities (policies)

Shortest Path Doesn't Work

- All nodes need common notion of link costs
- Incompatible with commercial relationships



A Technical Solution

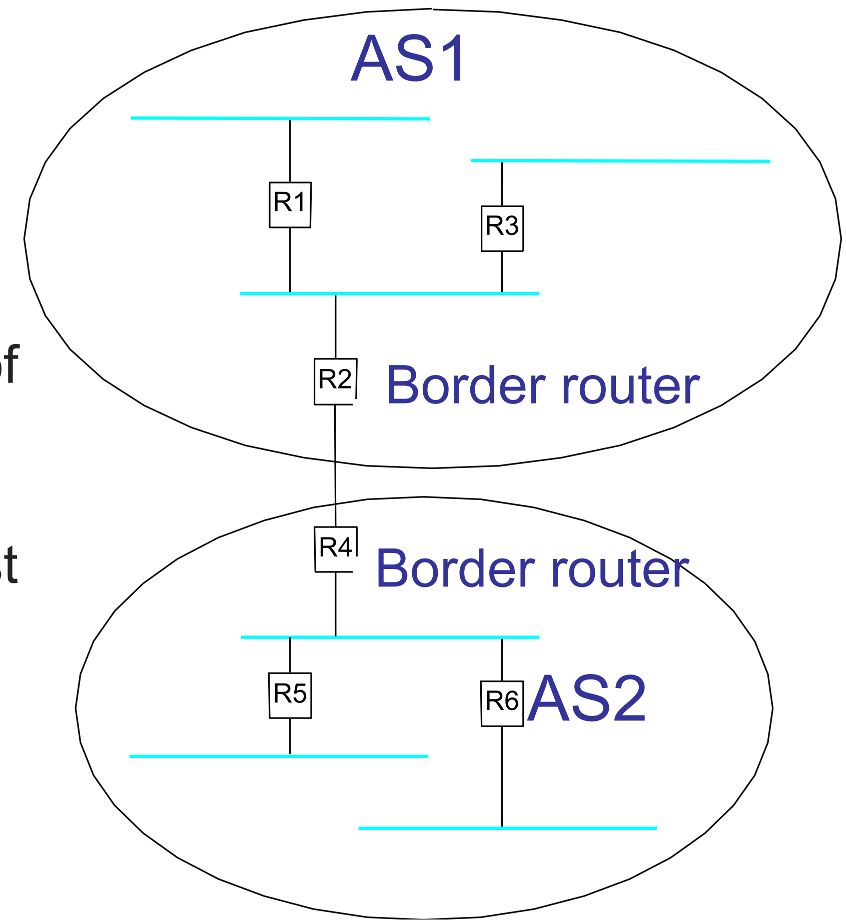
- Separate routing inside a domain from routing between domains
  - Inside a domain use traditional interior gateway protocols (RIP, OSPF, etc)
    - » You've seen these before
  - Between domains use Exterior Gateway Protocols (EGPs)
    - » Only exchange reachability information (not specific metrics)
    - » Decide what to do based on local policy
- What is a domain?

Autonomous Systems

- Internet is 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

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
- But what routing protocol?



Issues with Link-state

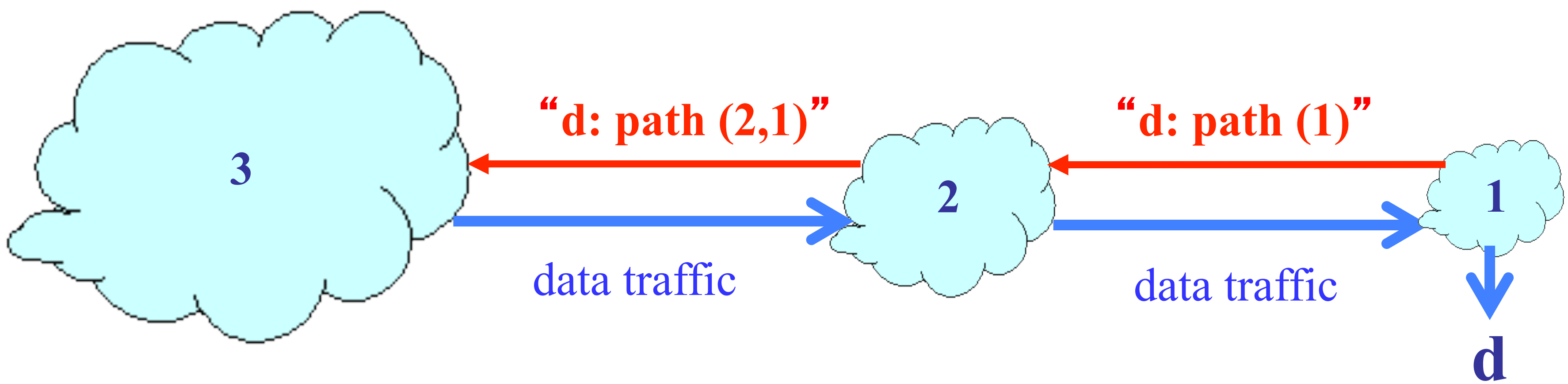
- Topology information is flooded
  - High bandwidth and storage overhead
  - Forces nodes to divulge sensitive information
- Entire path computed locally per node
  - High processing overhead in a large network
- Minimizes some notion of total distance
  - Works only if policy is shared and uniform
- Typically used only inside an AS
  - E.g., OSPF and IS-IS

Distance Vector *almost there*

- Advantages
  - Hides details of the network topology
  - Nodes determine only "next hop" toward the destination
- Disadvantages
  - Minimizes some notion of total distance, which is difficult in an interdomain setting
  - Slow convergence due to the counting-to-infinity problem ("bad news travels slowly")
- Idea: extend the notion of a distance vector
  - To make it easier to detect loops

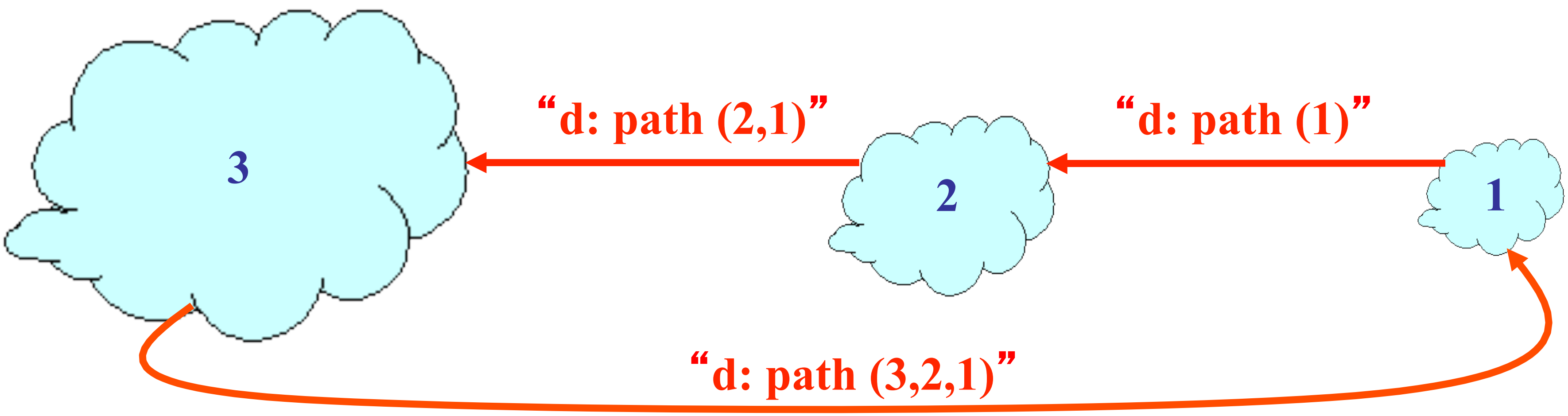
Path-vector Routing

- Extension of distance-vector routing
  - Support flexible routing policies
  - Avoid count-to-infinity problem
- Key idea: advertise the entire path
  - Distance vector: send *distance metric* per destination
  - Path vector: send the *entire path* for each destination



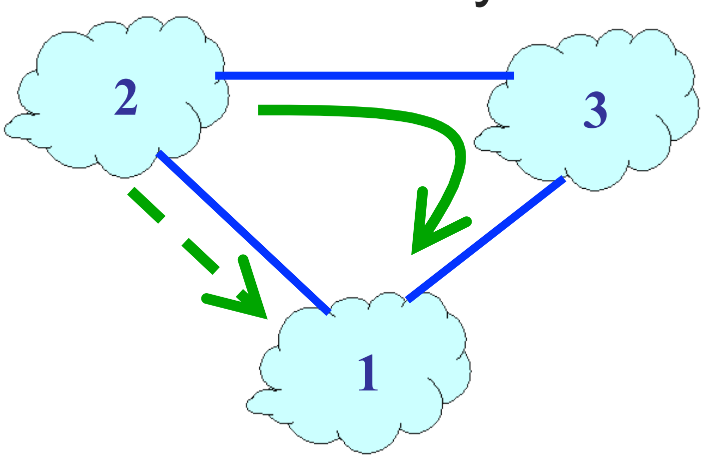
Loop Detection

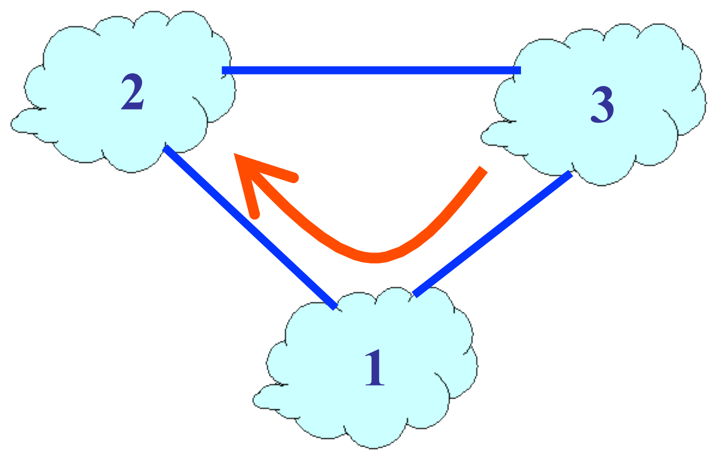
- Node can easily detect a loop
  - Look for its own node identifier in the path
  - E.g., node 1 sees itself in the path "3, 2, 1"
- Node can simply discard paths with loops
  - E.g., node 1 simply discards the advertisement



Policy Support

- Each node can apply local policies
  - Path selection: Which path to use?
  - Path export: Which paths to advertise?
- Examples
  - Node 2 may prefer the path "2, 3, 1" over "2, 1"
  - Node 1 may not let node 3 hear the path "1, 2"



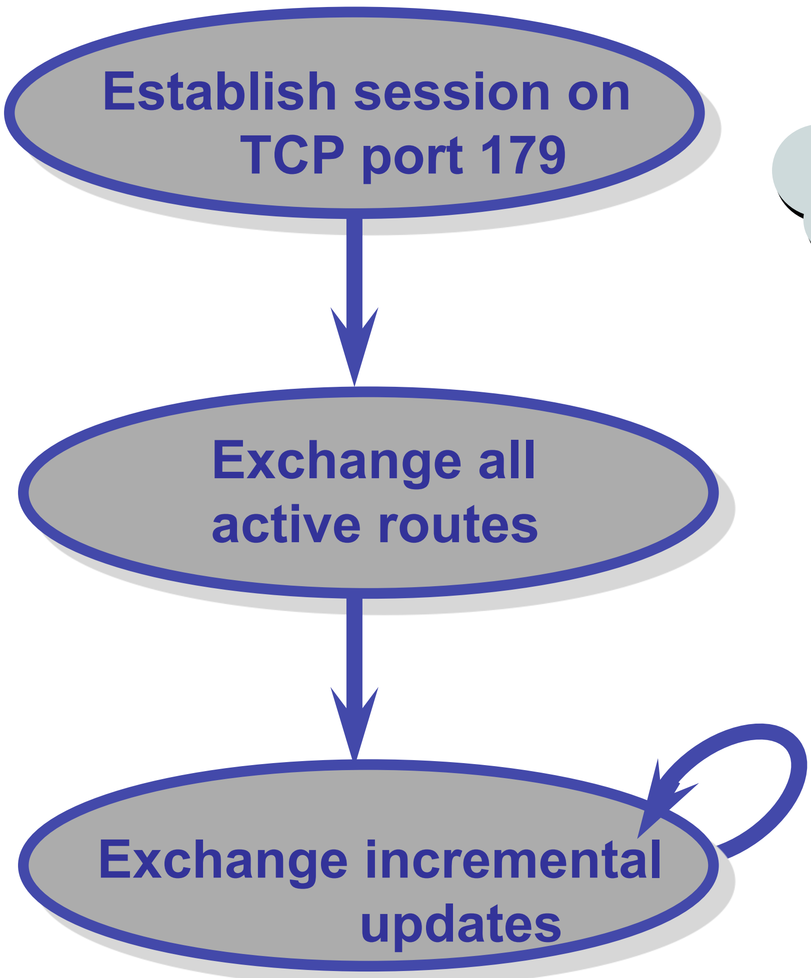


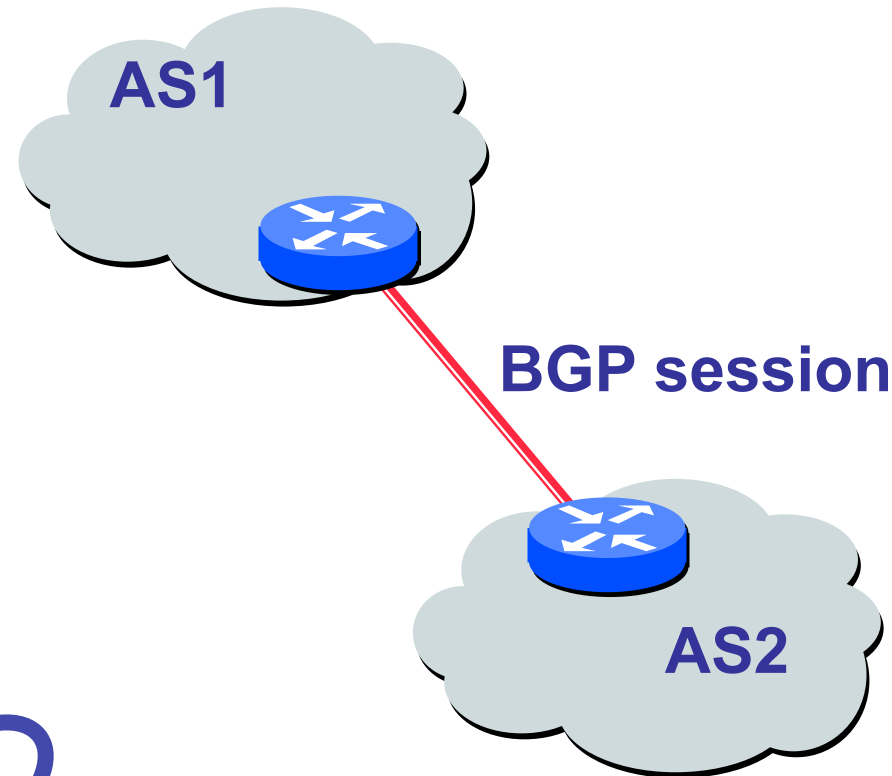
Border Gateway Protocol

- Interdomain routing protocol for the Internet
  - Prefix-based path-vector protocol
  - Policy-based routing based on AS Paths
  - Evolved during the past 18 years

- **1989 : BGP-1 [RFC 1105], replacement for EGP**
- **1990 : BGP-2 [RFC 1163]**
- **1991 : BGP-3 [RFC 1267]**
- **1995 : BGP-4 [RFC 1771], support for CIDR**
- **2006 : BGP-4 [RFC 4271], update**

Basic BGP Operation





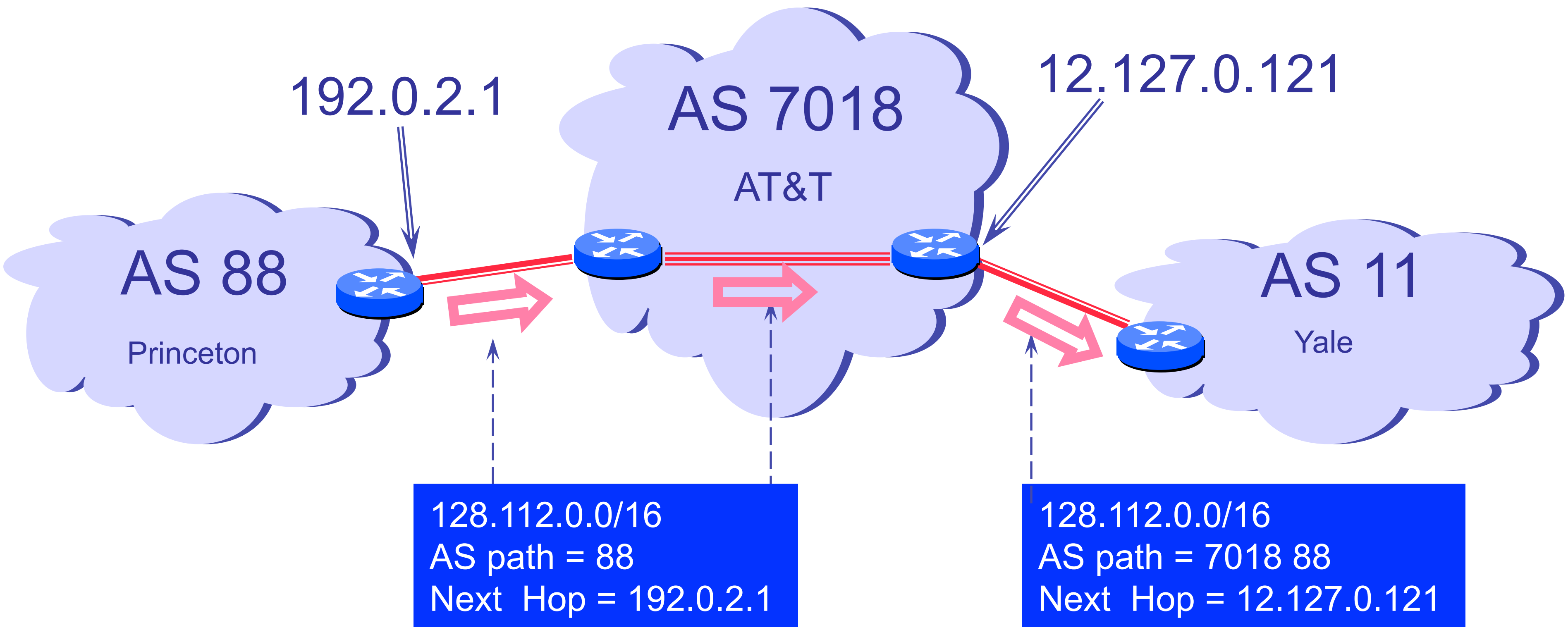
While connection
is ALIVE exchange
route UPDATE messages

Step-by-Step

- A node learns multiple paths to destination
  - Stores all of the routes in a routing table
  - Applies policy to select a single active route
  - … and may advertise the route to its neighbors
- Incremental updates
  - Announcement
    - » Upon selecting a new active route, add node id to path
    - » … and (optionally) advertise to each neighbor
  - Withdrawal
    - » If the active route is no longer available
    - » … send a withdrawal message to the neighbors

A Simple BGP Route

- Destination prefix (e.g., 128.112.0.0/16)
- Route attributes, including
  - AS path (e.g., “7018 88”)
  - Next-hop IP address (e.g., 12.127.0.121)

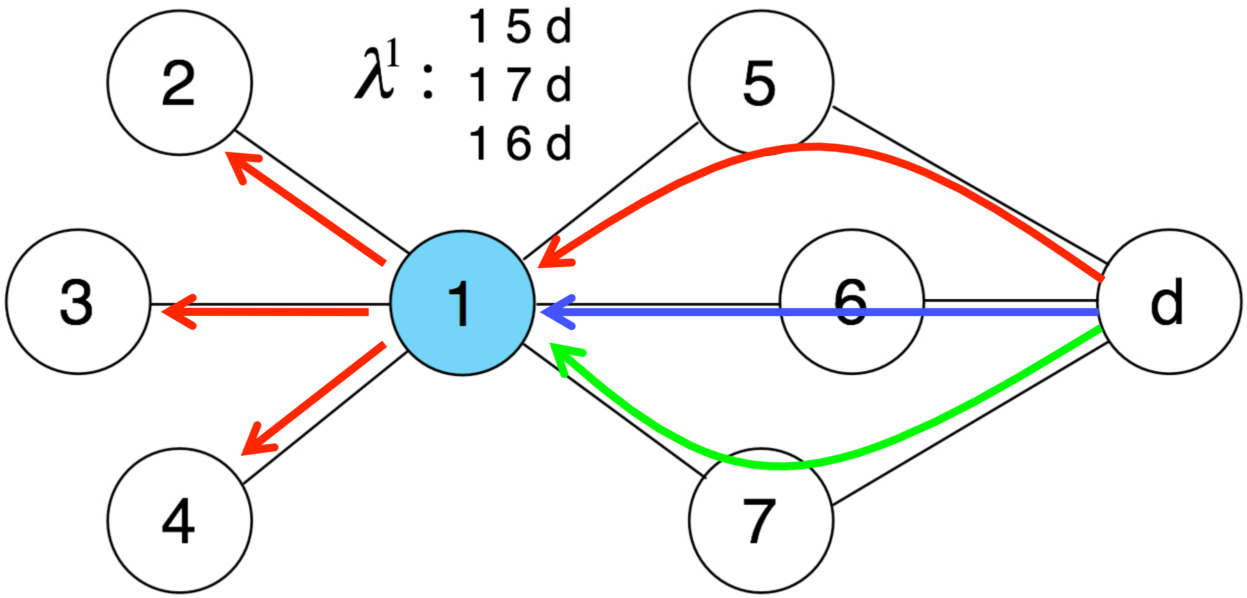


BGP Attributes

- **Local pref:** Statically configured ranking of routes within AS
- **AS path:** ASs the announcement traversed
- **Origin:** Route came from IGP or EGP
- **Multi Exit Discriminator:** preference for where to *exit* network
- **Community:** opaque data used for inter-ISP policy
- **Next-hop**: where the route was heard from

Export Active Routes

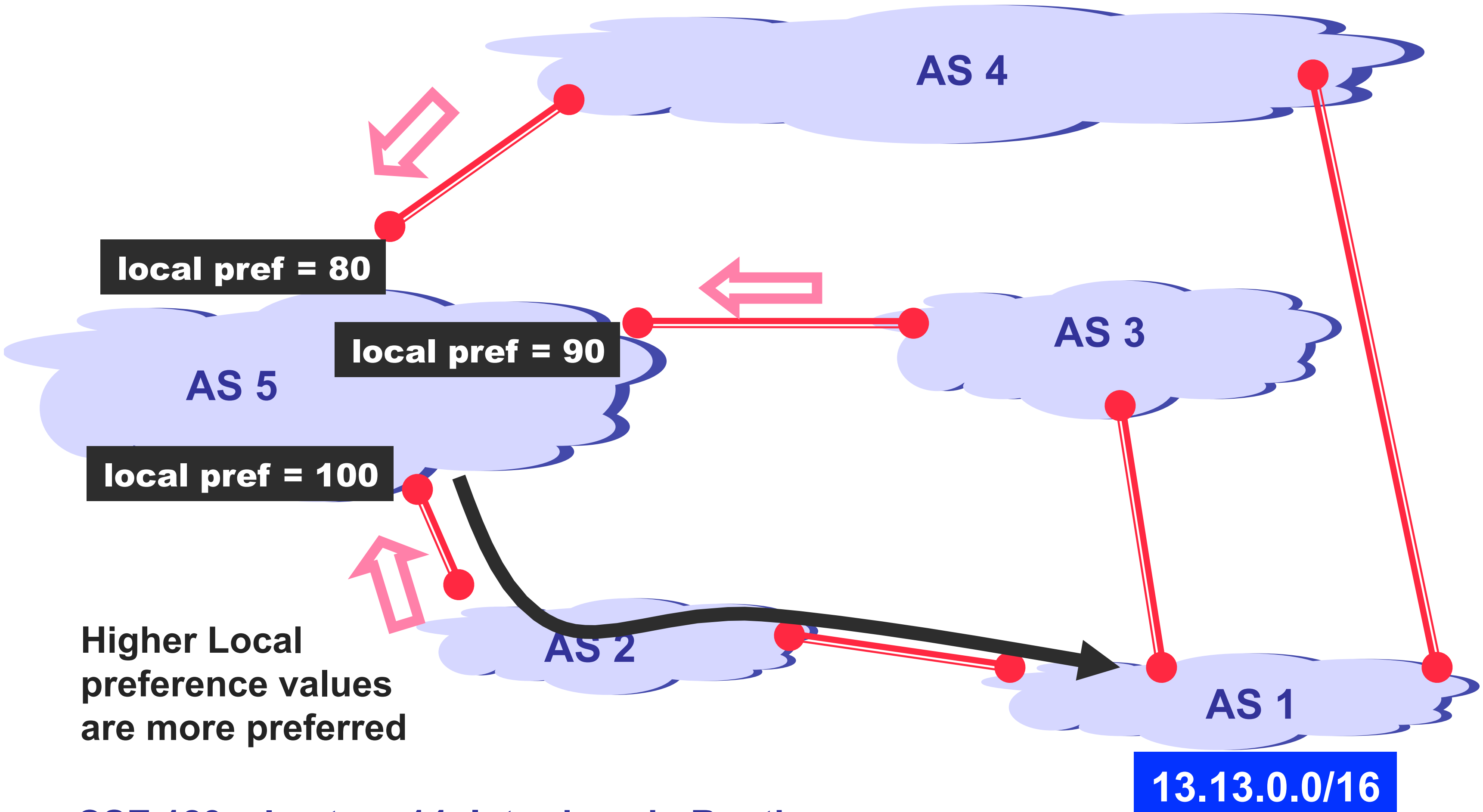
- In conventional path vector routing, a node has one ranking function, which reflects its routing policy



BGP Decision Process

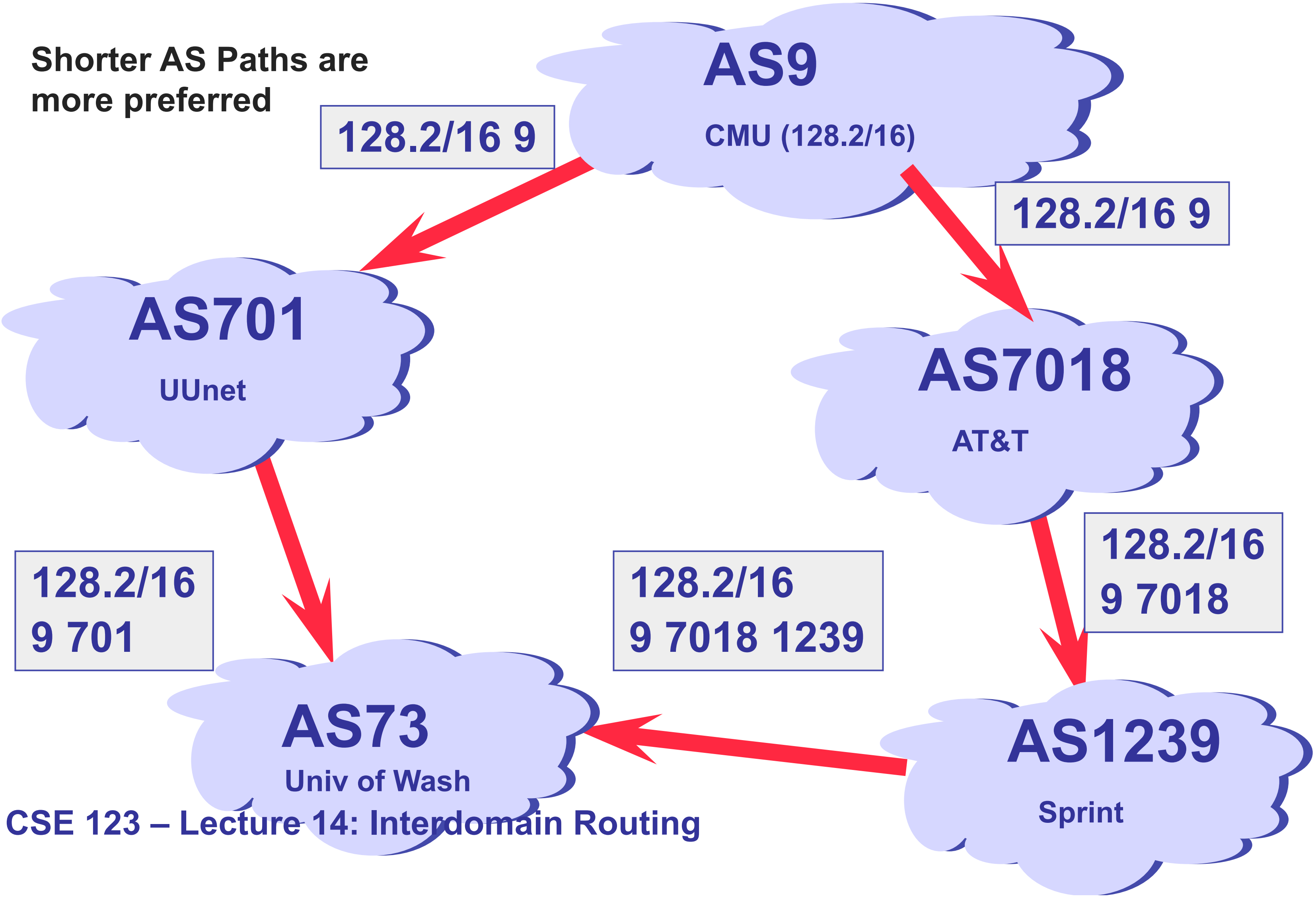
- Default decision for route selection
  - Highest local pref, shortest AS path, lowest MED, prefer eBGP over iBGP, lowest IGP cost, router id
- Many policies built on default decision process, but…
  - Possible to create arbitrary policies in principal
    - » Any criteria: BGP attributes, source address, prime number of bytes in message, …
    - » Can have separate policy for inbound routes, installed routes and outbound routes
  - Limited only by power of vendor-specific routing language

Example: Local Pref

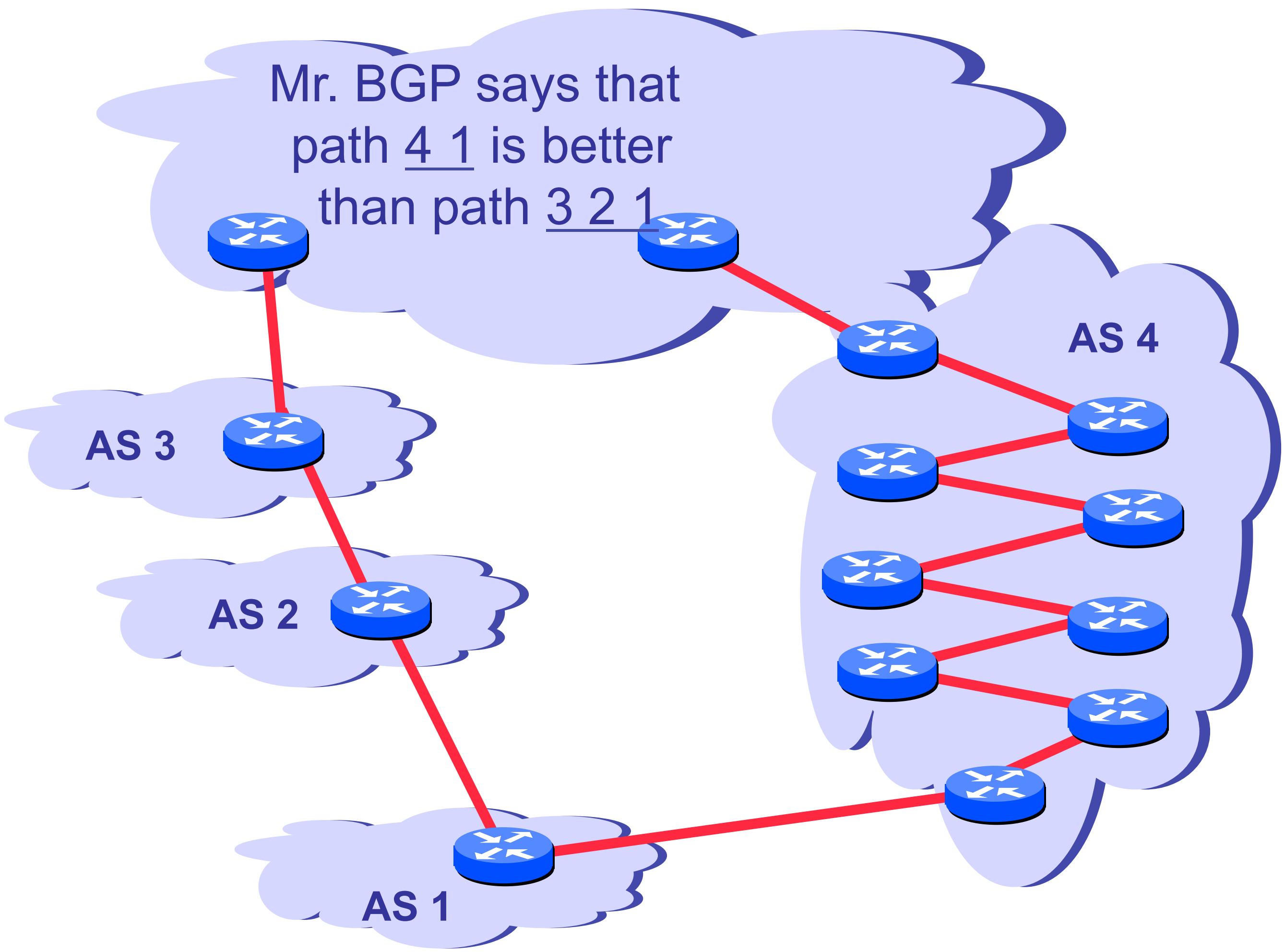


Higher Local preference values are more preferred

Example: Short AS Path



AS Paths *vs.* Router Paths



BGP Has Lots of Problems

- Instability
  - Route flapping (network x.y/z goes down… tell everyone)
  - Long AS-path decision criteria defaults to DV-like behavior (bouncing)
  - Not guaranteed to converge, NP-hard to tell if it does
- Scalability still a problem
  - ~300,000 network prefixes in default-free table today
  - Tension: Want to manage traffic to very specific networks (eg. multihomed content providers) but also want to aggregate information.
- Performance
  - Non-optimal, doesn't balance load across paths

A History of Settlement

- The telephone world
  - LECs (local exchange carriers)
  - IXCs (inter-exchange carriers)
- LECs MUST provide IXCs access to customers
  - This is enforced by laws and regulation
- When a call goes from one phone company to another:
  - Call billed to the caller
  - The money is split up among the phone systems – this is called "settlement"

Business Relationships

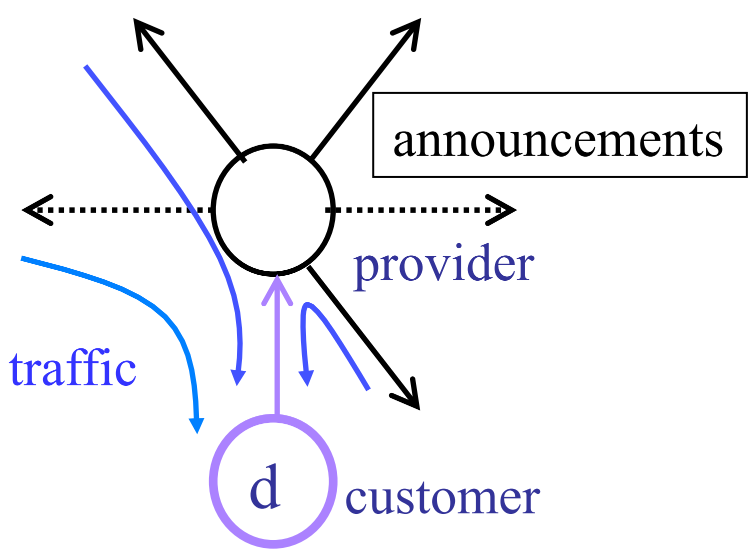
- Neighboring ASes have business contracts
  - How much traffic to carry
  - Which destinations to reach
  - How much money to pay

- Common business relationships
  - Customer-provider
    - E.g., Princeton is a customer of USLEC
    - E.g., MIT is a customer of Level3
  - Peer-peer
    - E.g., UUNET is a peer of Sprint
    - E.g., Harvard is a peer of Harvard Business School

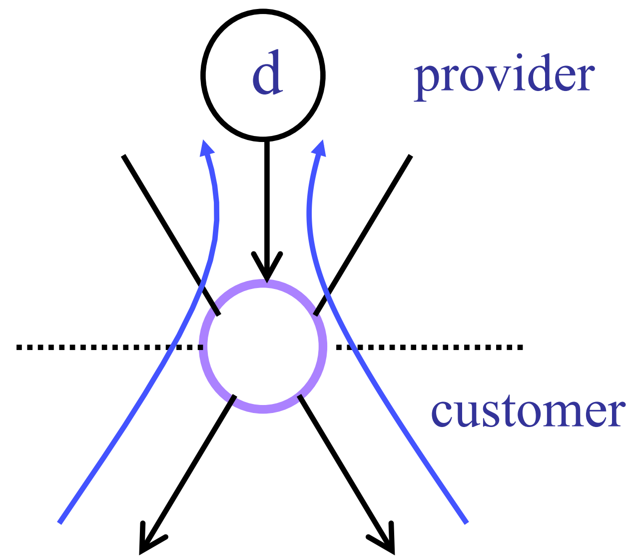
Customer/Provider

- 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

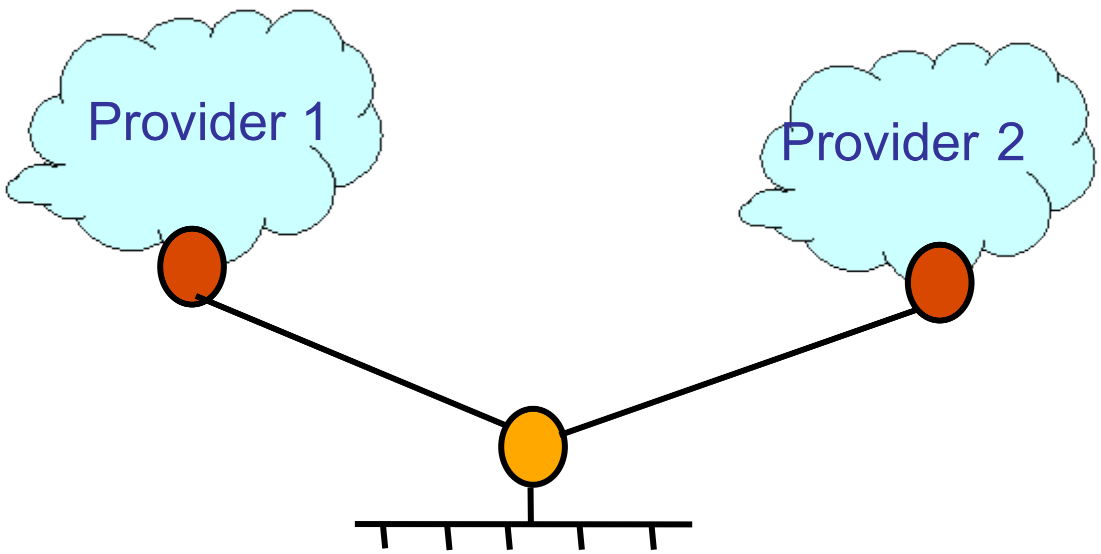


Traffic **from** the customer



Multi-Homing

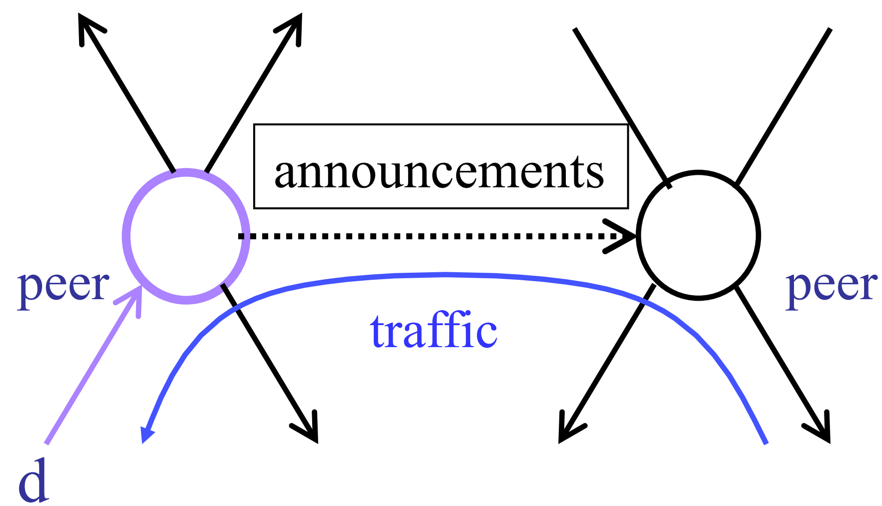
- Customers may have more than one provider
  - Extra reliability, survive single ISP failure
  - Financial leverage through competition
  - Better performance by selecting better path
  - Gaming the 95th-percentile billing model



Peer-to-Peer Relationship

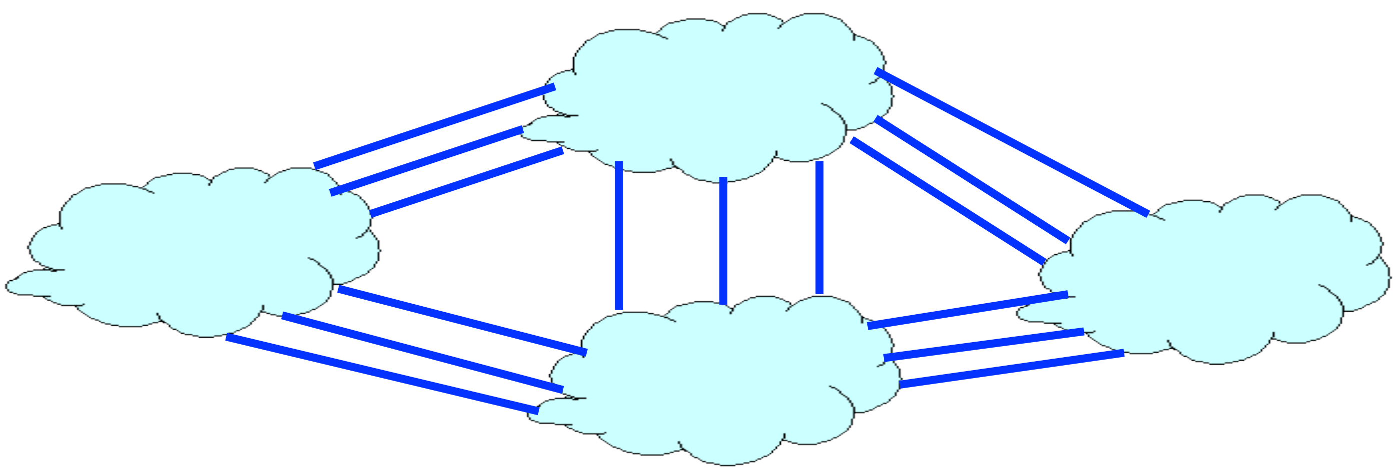
- Peers exchange traffic between customers
  - AS exports *only* customer routes to a peer
  - AS exports a peer's routes *only* to its customers
  - Often the relationship is settlement-free (i.e., no $$$)

Traffic to/from the peer and its customers



Tier-1 Providers

- Make up the "core" of the Internet
  - 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



Summary

- Interdomain-routing
  - Exchange reachability information (plus hints)
  - BGP is based on path vector routing
  - Local policy to decide which path to follow
- Traffic exchange policies are a big issue $$$
  - Complicated by lack of compelling economic model (who creates value?)
  - Can have significant impact on performance

For next time…

- Read Ch. 6.2,.6.5 in P&D
- Keep moving on Project 2