

Lecture 17:

Border Gateway Protocol

CSE 123: Computer Networks

Aaron Schulman



Some figures courtesy Mike Freedman



Lecture 17 Overview

- Path vector routing
- Border Gateway Protocol (BGP)
 - The canonical path vector protocol
 - How routing gets done on the Internet today
- AS relationships
 - Customer/Provider
 - Multihoming
 - Peer-to-peer

Distance Vector *almost there*

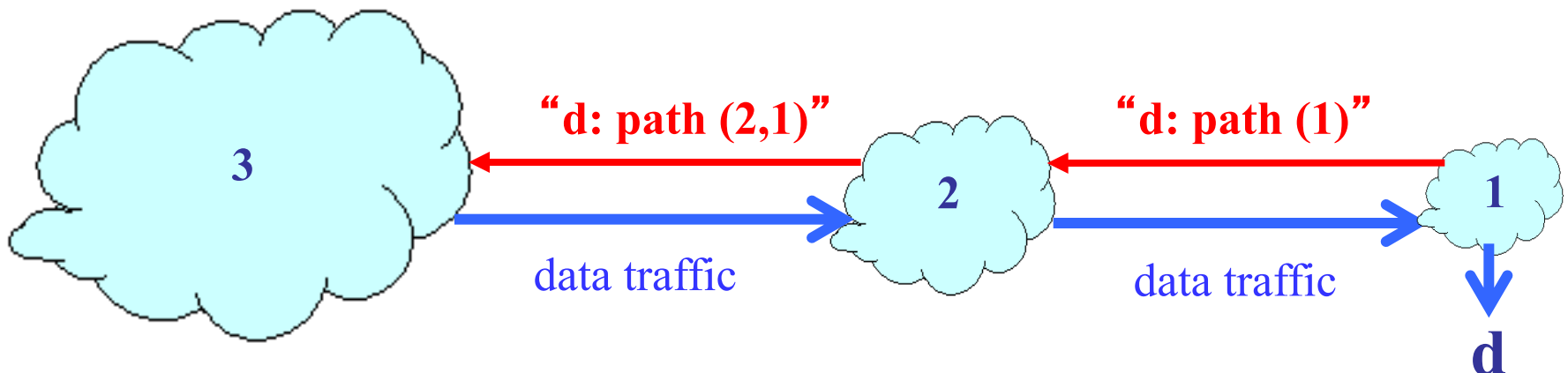


- Advantages
 - ◆ Hides details of the network topology
 - ◆ Nodes determine only “next hop” toward the destination
- Disadvantages
 - ◆ Minimizes some notion of total distance, which is difficult in an interdomain setting
 - ◆ Slow convergence due to the counting-to-infinity problem (“bad news travels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



Path-vector Routing

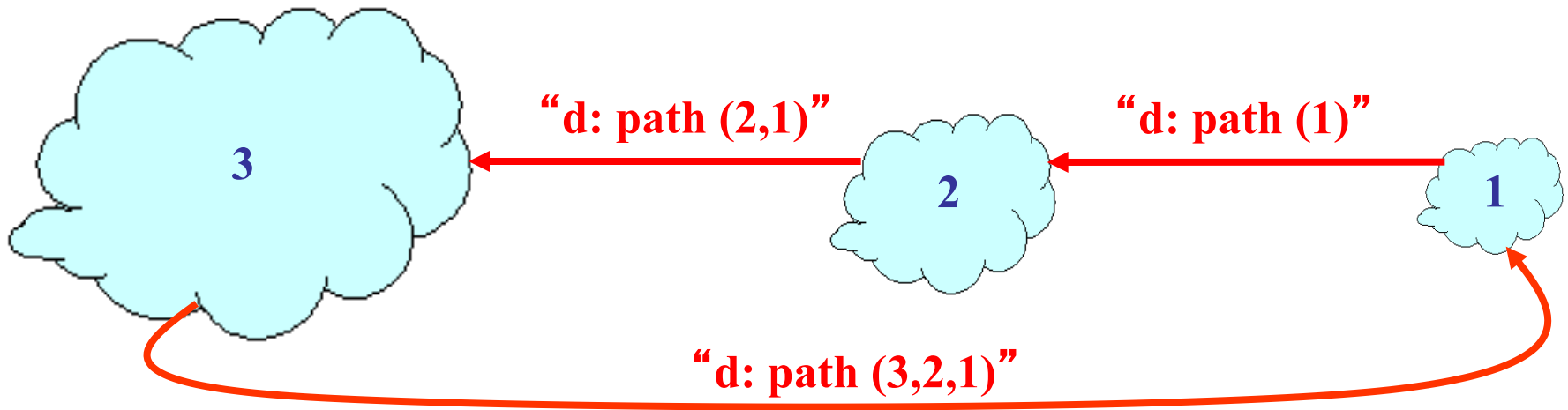
- Extension of distance-vector routing
 - ◆ Support flexible routing policies
 - ◆ Avoid count-to-infinity problem
- Key idea: advertise the entire path
 - ◆ Distance vector: send *distance metric* per destination
 - ◆ Path vector: send the *entire path* for each destination





Loop Detection

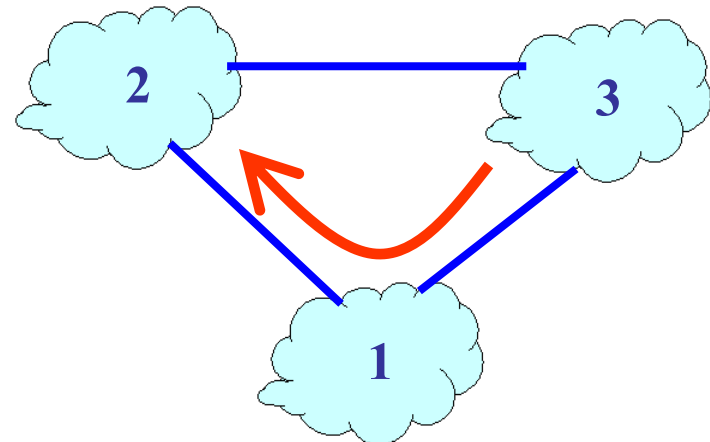
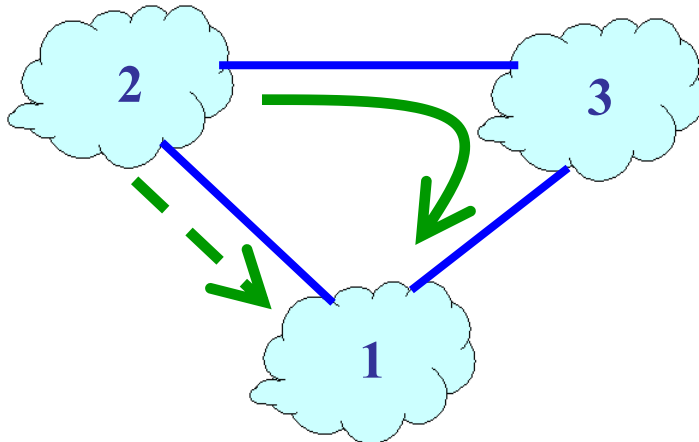
- Node can easily detect a loop
 - ◆ Look for its own node identifier in the path
 - ◆ E.g., node 1 sees itself in the path “3, 2, 1”
- Node can simply discard paths with loops
 - ◆ E.g., node 1 simply discards the advertisement





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”





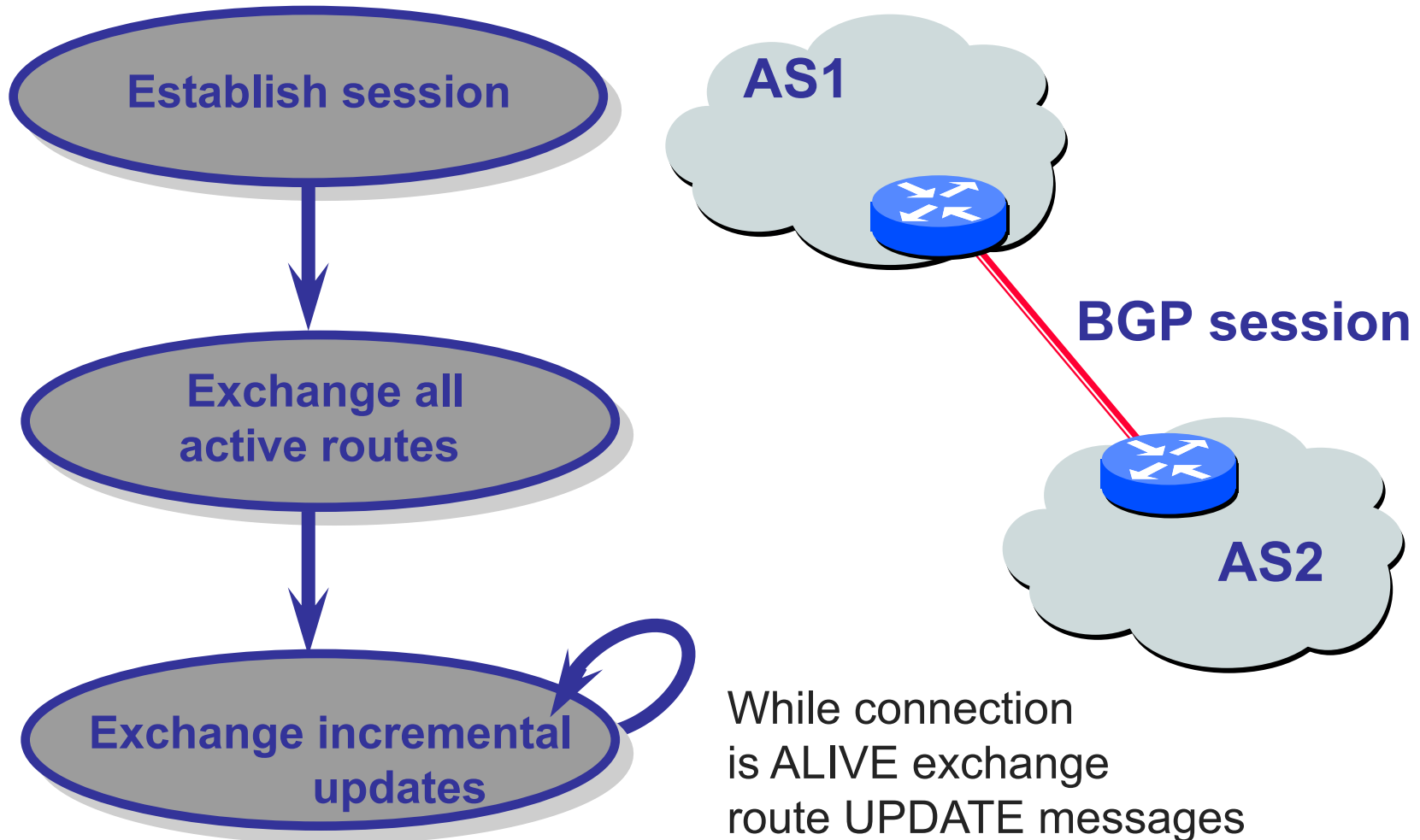
Border Gateway Protocol

- Interdomain routing protocol for the Internet
 - ◆ Prefix-based path-vector protocol
 - ◆ Policy-based routing based on AS Paths
 - ◆ Evolved during the past 28 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



Step-by-Step

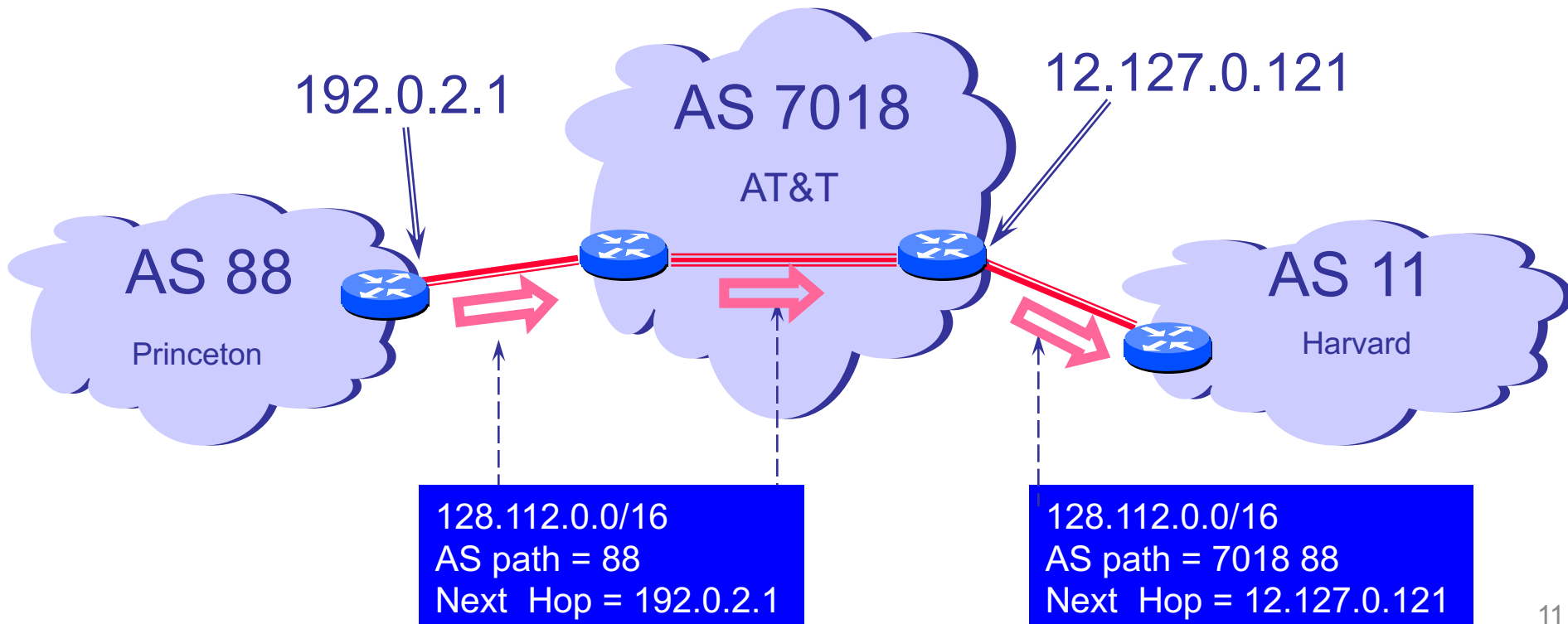


- A node learns multiple paths to destination
 - ◆ Stores all of the routes in a routing table
 - ◆ Applies policy to select a single active route
 - ◆ ... and may advertise the route to its neighbors
- Incremental updates
 - ◆ Announcement
 - » Upon selecting a new active route, add own node id to path
 - » ... 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)





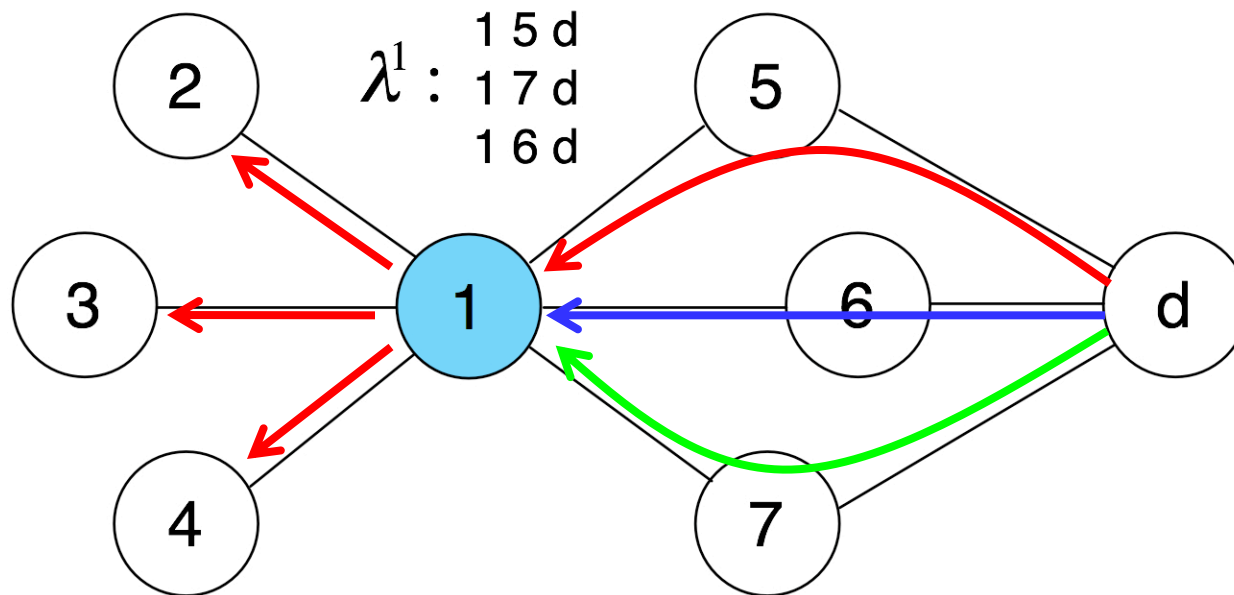
(some) BGP Attributes

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



Export Active Routes

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



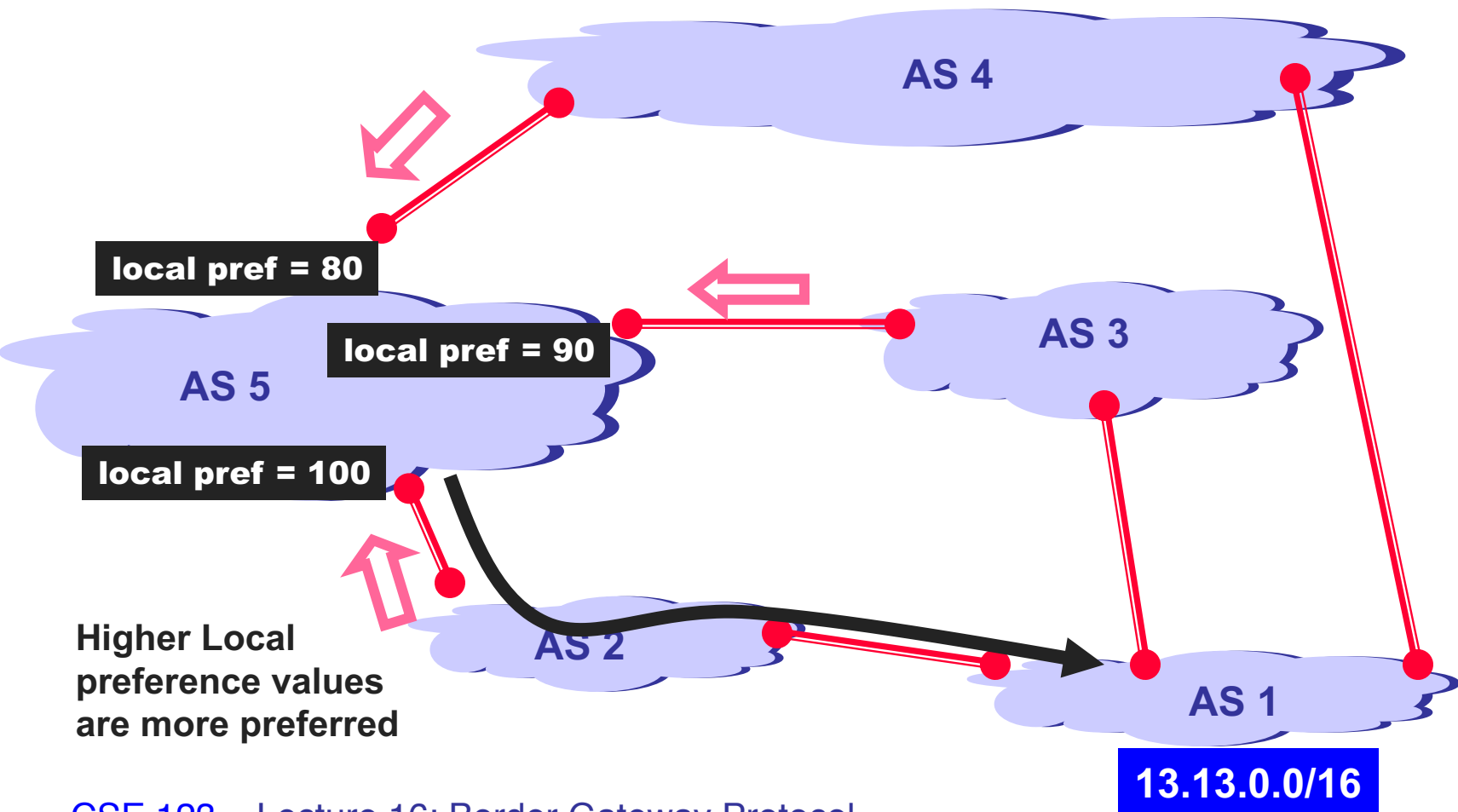


BGP Decision Process

- 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



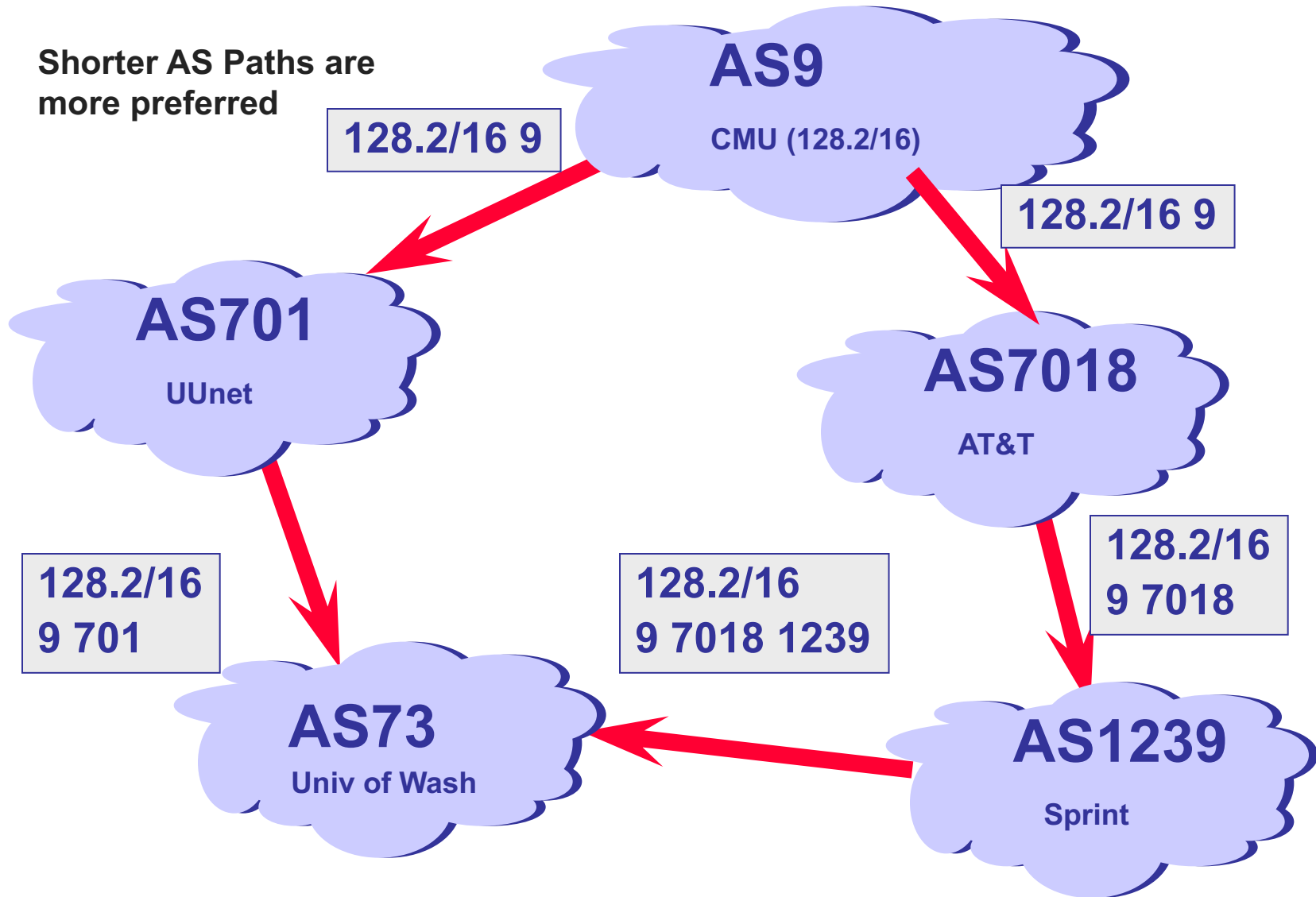
Example: Local Pref



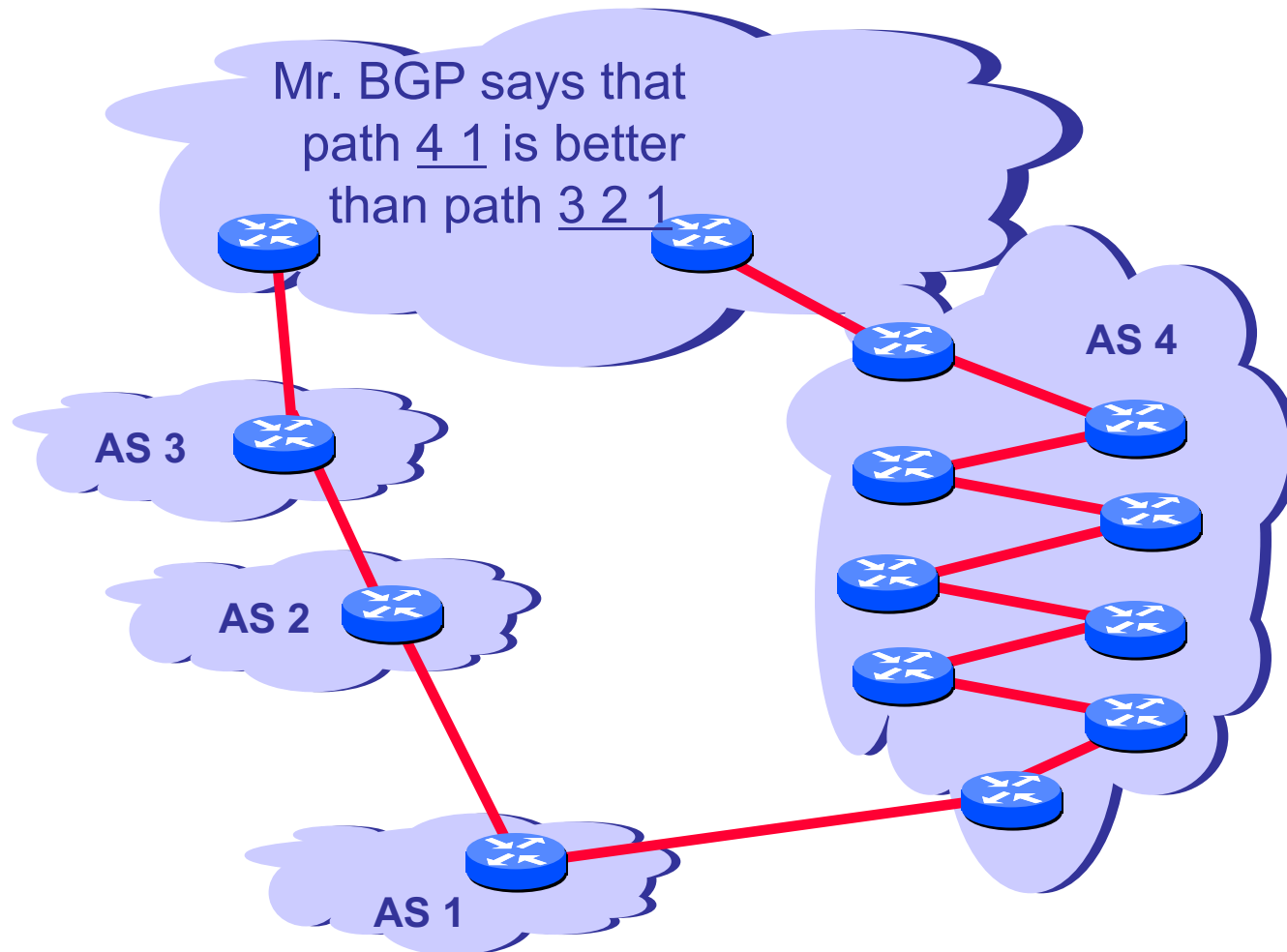


Example: Short AS Path

Shorter AS Paths are more preferred



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
 - ◆ >500,000 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) (e.g., PacBell, NYNEX)
 - ◆ IXC (inter-exchange carriers) (e.g., Sprint, AT&T)
- 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

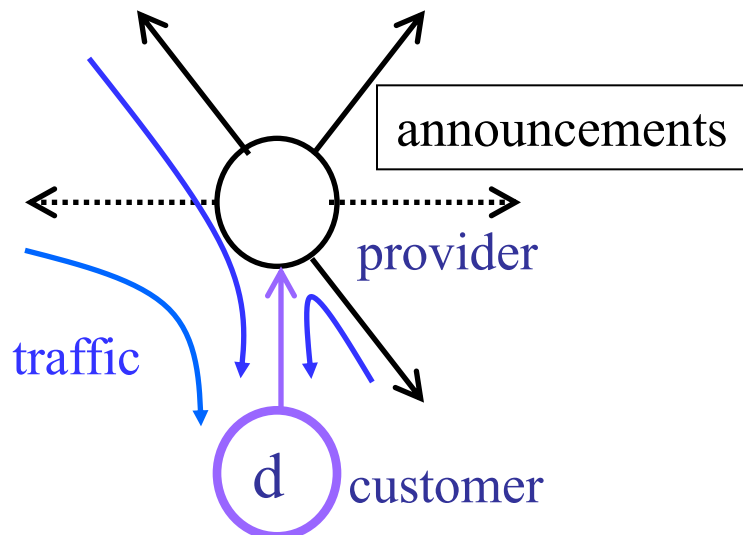
- Neighboring ASes have business contracts
 - ◆ How much traffic to carry
 - ◆ Which destinations to reach
 - ◆ How much money to pay
- Common business relationships
 - ◆ Customer-provider
 - » E.g., Princeton is a customer of USLEC
 - » E.g., MIT is a customer of Level3
 - ◆ Peer-peer
 - » E.g., UUNET is a peer of Sprint
 - » E.g., Harvard is a peer of Harvard Business School



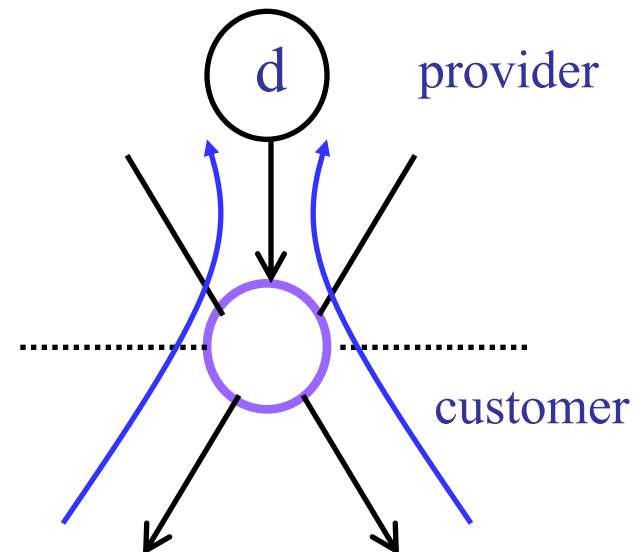
Customer/Provider

- 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



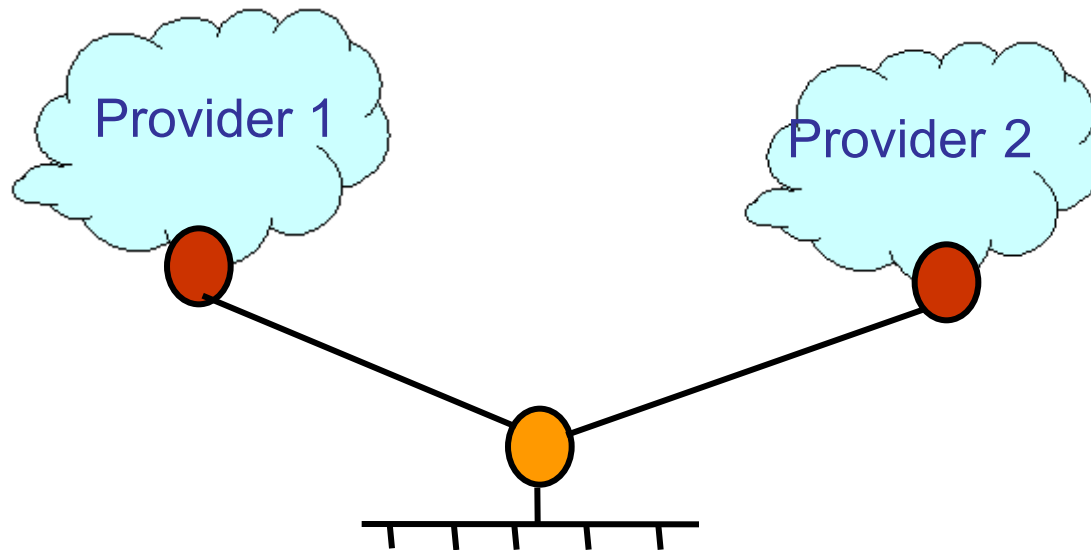
Traffic **from** the customer





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





For Next Time

- Read P&D 3.4
- Homework 3 and Project 2 out Monday