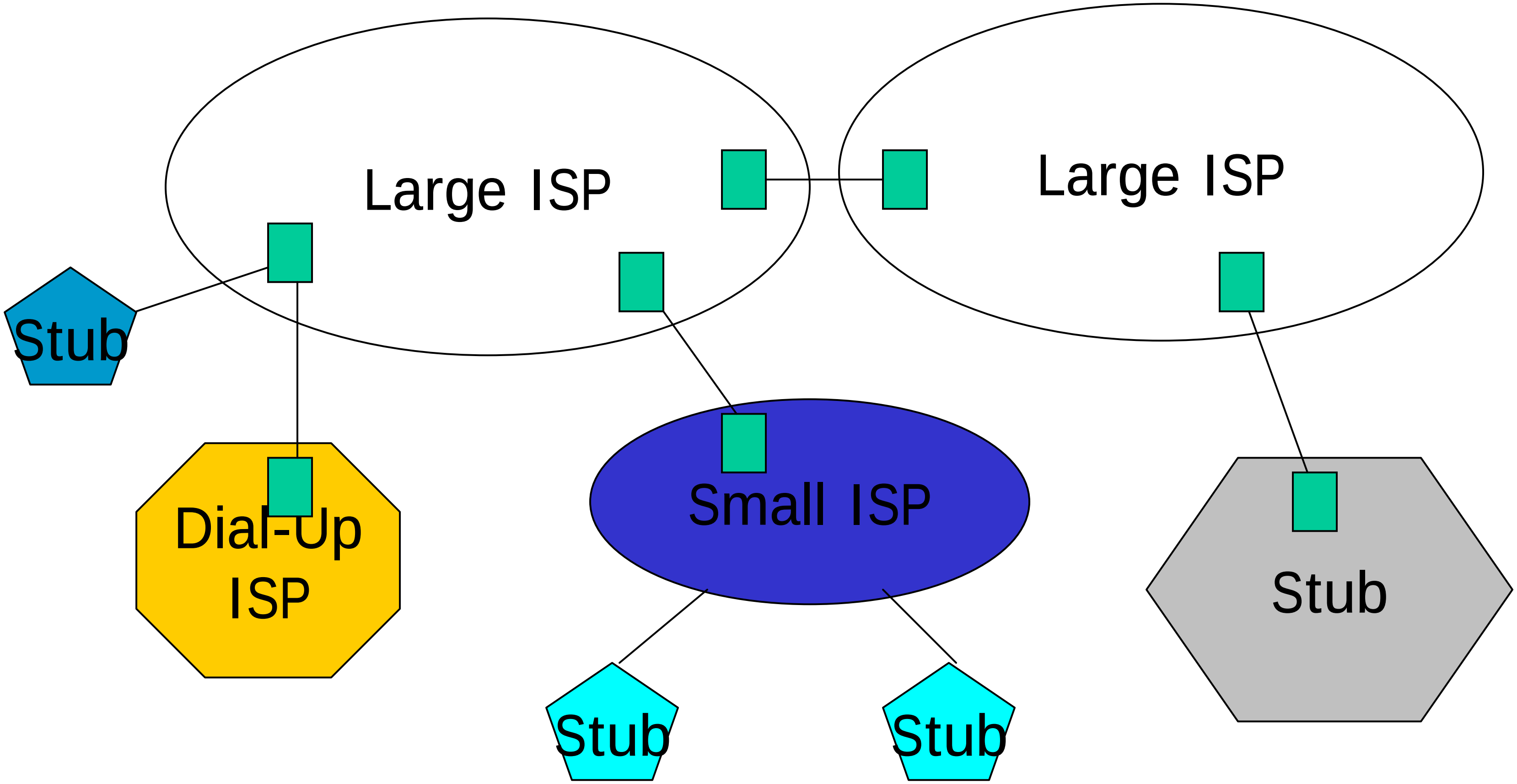
COMP/ELEC 429
Introduction to Computer Networks

Lecture 11: Inter-domain routing

Slides used with permissions from Edward W. Knightly, T. S. Eugene Ng, Ion Stoica, Hui Zhang

Internet Structure



Autonomous Systems (AS)

- Internet is not a single network!
- The Internet is a collection of networks, each controlled by different administrations
- An autonomous system (AS) is a network under a single administrative control

AS Numbers (ASNs)

ASNs are 16 bit values.
64512 through 65535 are “private”

Currently over 11,000 in use.

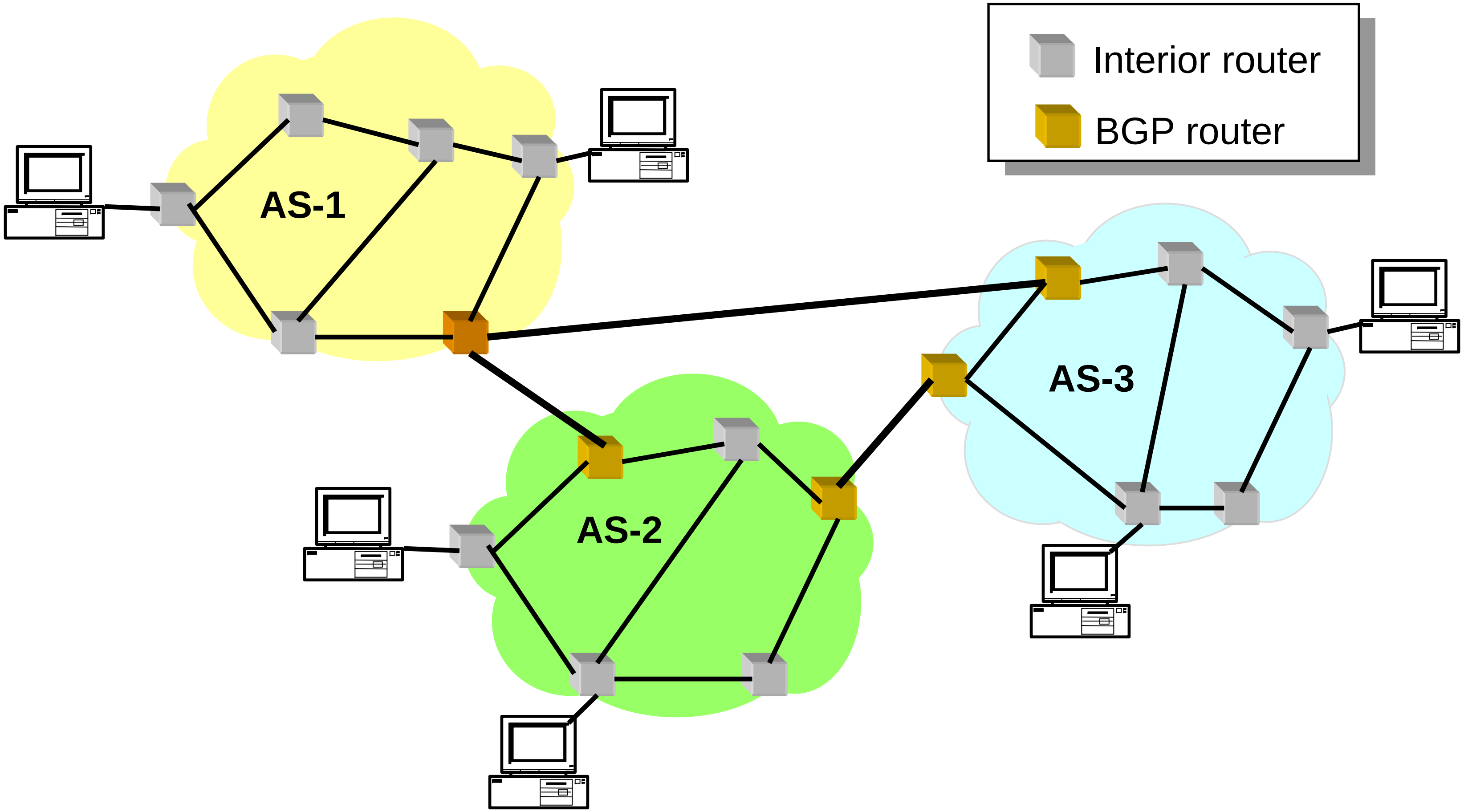
- Genuity: 1
- AT&T: 7018, 6341, 5074, …
- UUNET: 701, 702, 284, 12199, …
- Sprint: 1239, 1240, 6211, 6242, …
- …

Implications

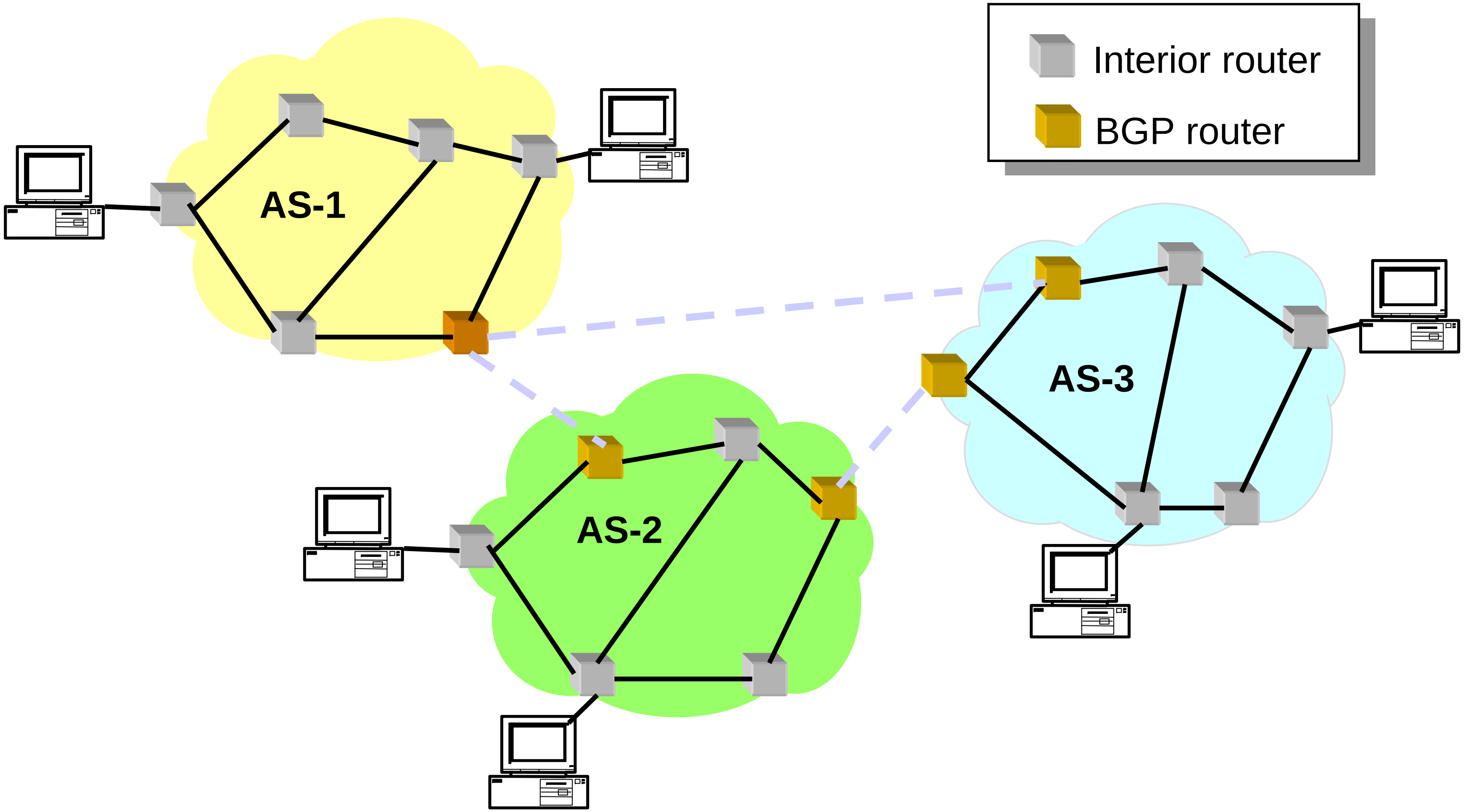
- ASs want to choose own local routing algorithm
  - AS takes care of getting packets to/from their own hosts
  - Intradomain routing: RIP, OSPF, etc

- ASs want to choose own non-local routing policy
  - Interdomain routing must accommodate this
  - BGP is the current interdomain routing protocol
  - BGP: Border Gateway Protocol

Example

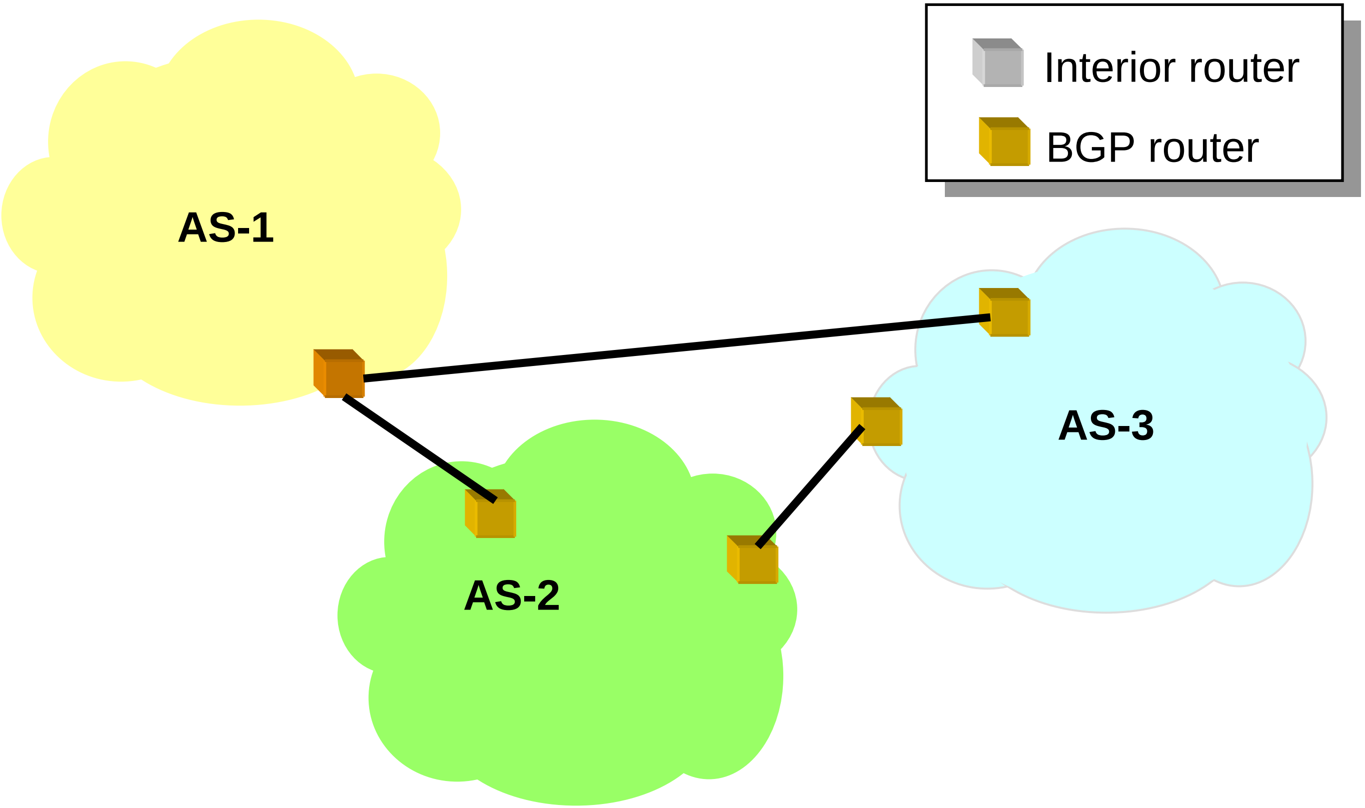


Intra-Domain



Intra-domain routing protocol aka Interior Gateway Protocol (IGP), e.g. OSPF, RIP

Inter-Domain



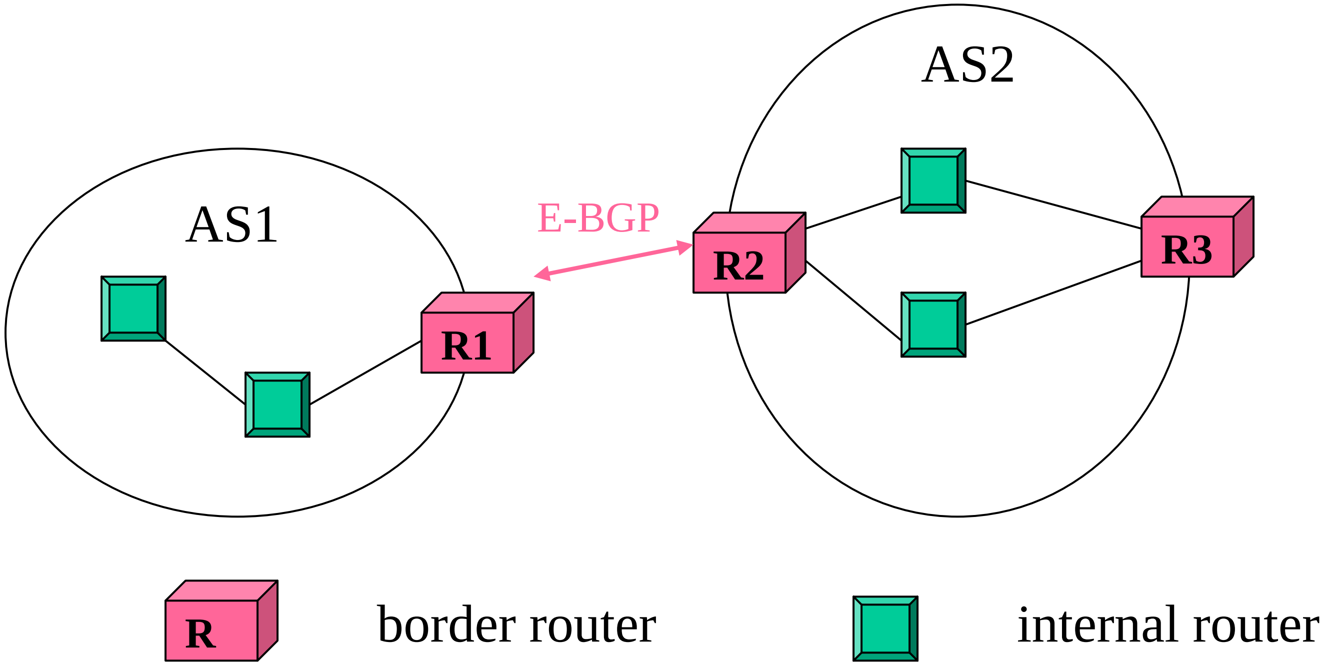
Inter-domain routing protocol aka Exterior Gateway Protocol (EGP), e.g. BGP

Inter-Domain Routing

- Global connectivity is at stake
- Inevitably leads to one single protocol that everyone must speak
  - Unlike many choices in intra-domain routing
- What are the requirements?
- Scalability
- Flexibility in choosing routes
- If you were to choose, link state based or distance vector based?

- BGP is sort of a hybrid: Path vector protocol

Border Gateway Protocol Part I: E-BGP



- Two types of routers
  - Border router (Edge), Internal router (Core)

Purpose of E-BGP

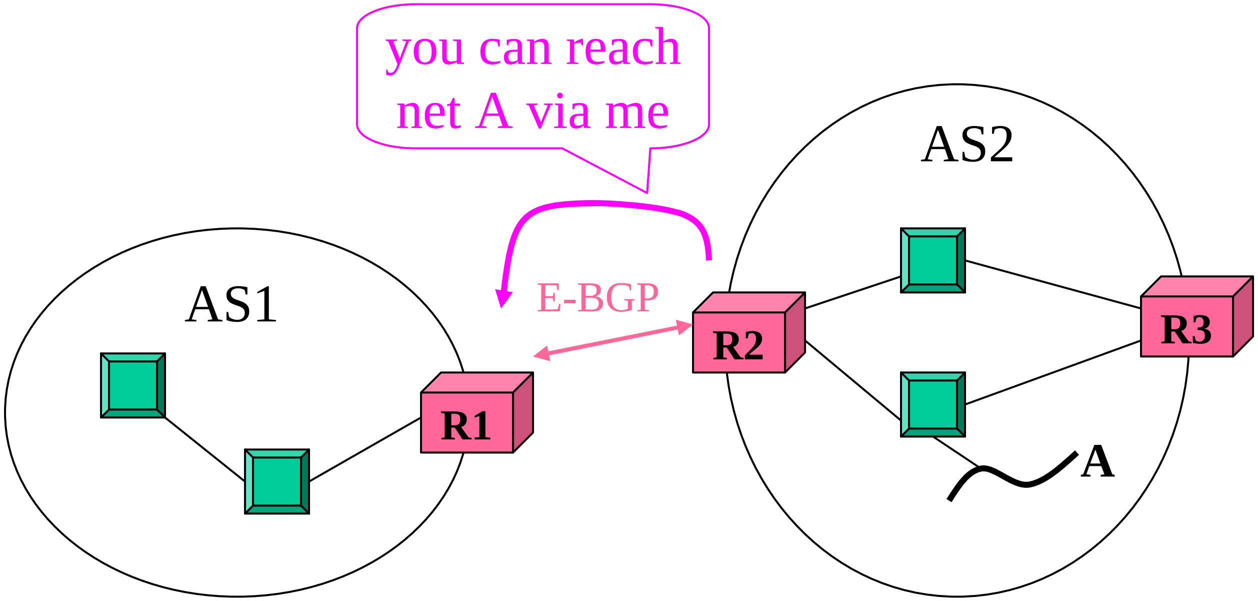


table at R1:

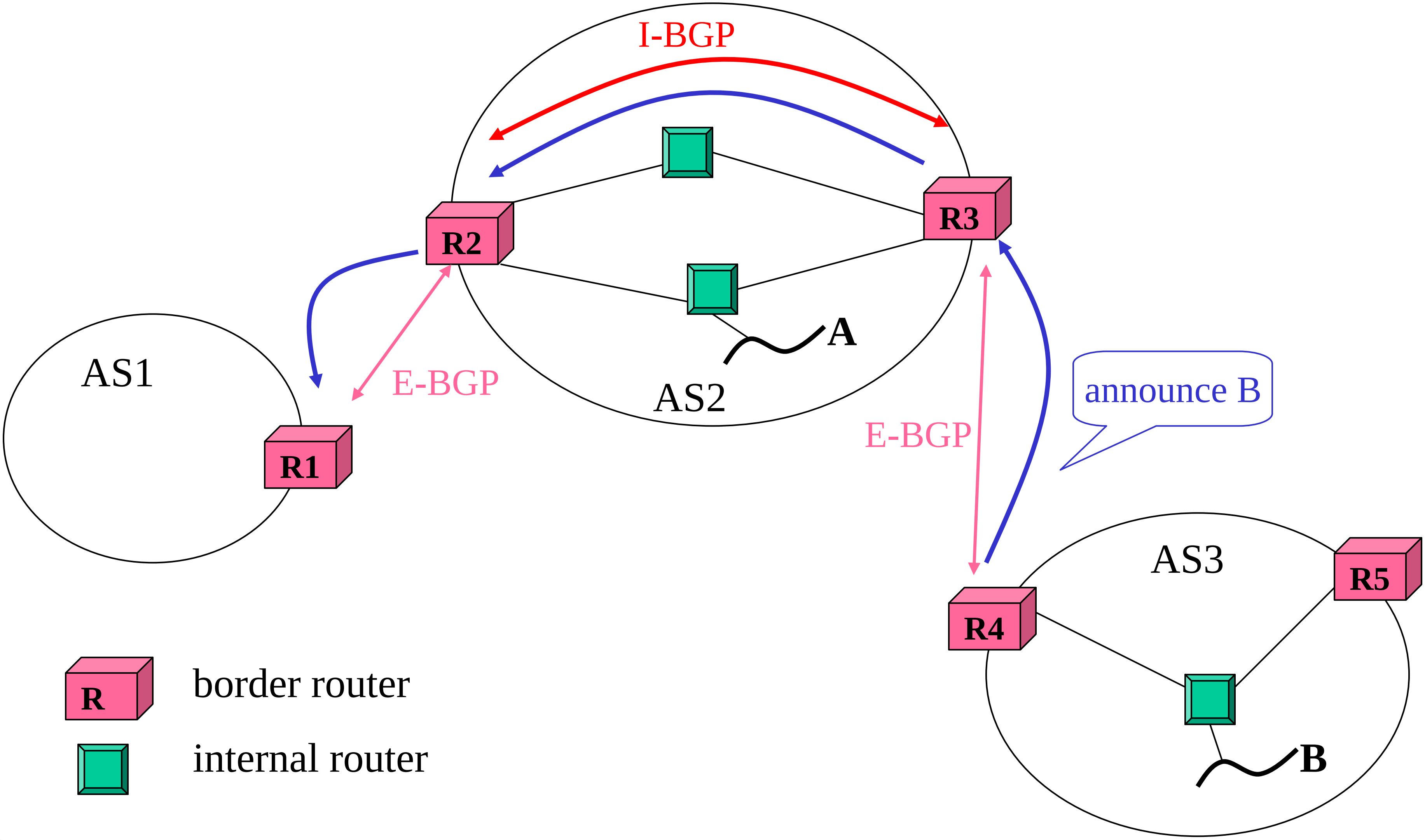
dest	next hop
A	R2

🟥 R border router

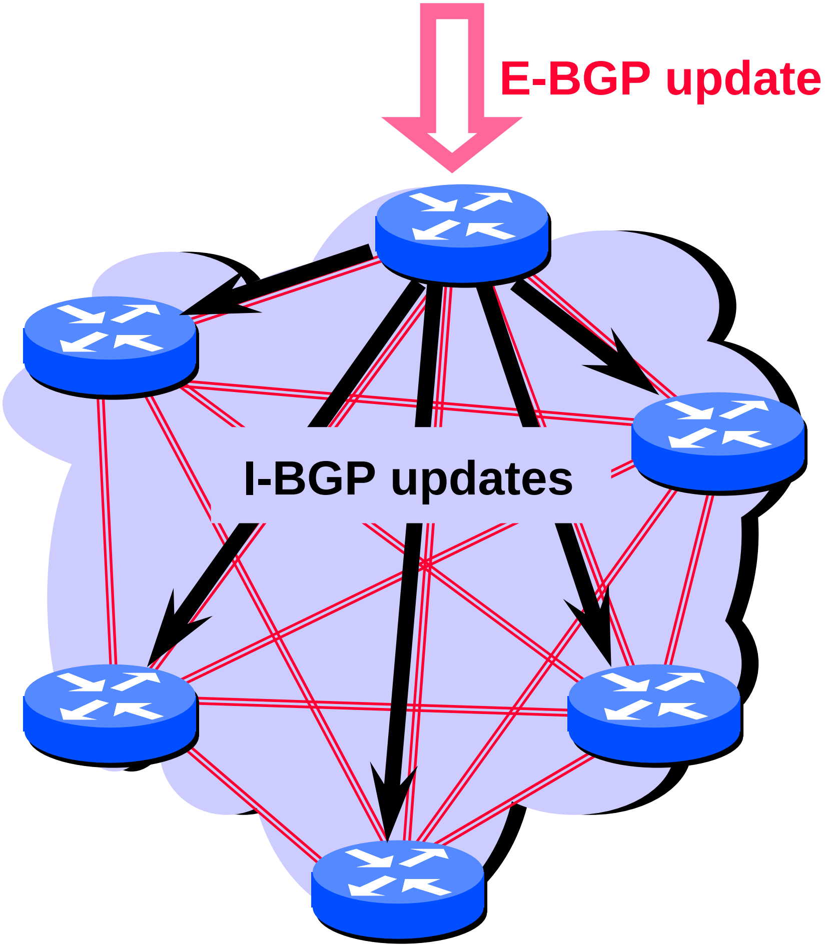
🟩 internal router

Share connectivity information across ASes

Part II: I-BGP, Carrying Info within an AS



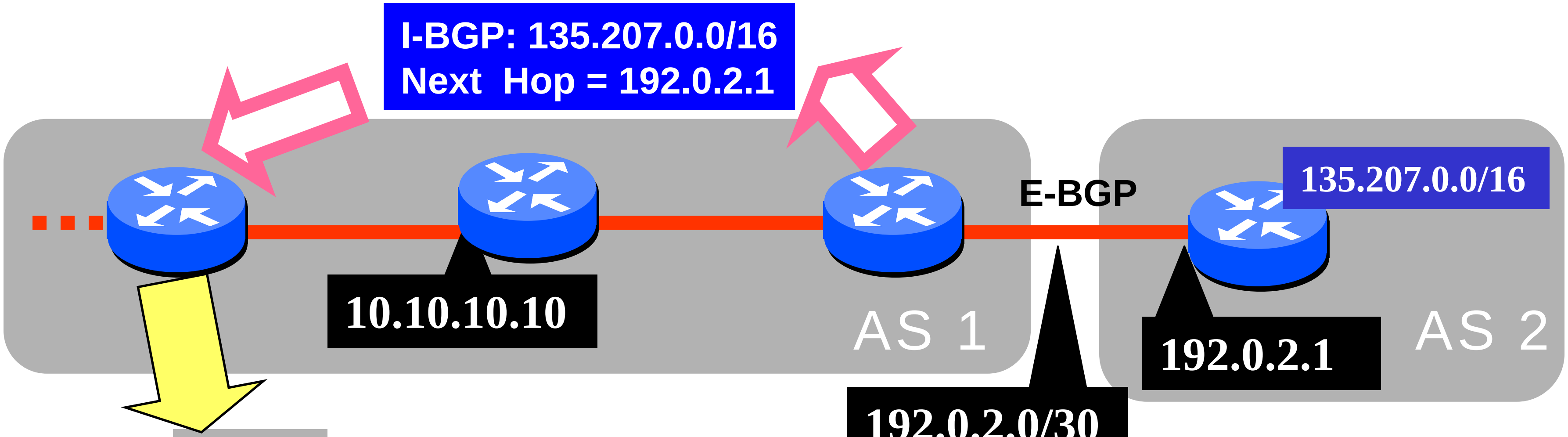
I-BGP



I-BGP neighbors do not announce routes received via I-BGP to other I-BGP neighbors.

- Problem: Injecting external routes into IGP (e.g. OSPF) does not scale and causes BGP policy information to be lost
- I-BGP can be used to disseminate BGP routes to all routers in AS
- BGP route + IGP route suffice to create forwarding table

Join I-BGP with IGP to Create Forwarding Table



IGP

destination	next hop
192.0.2.0/30	10.10.10.10

+

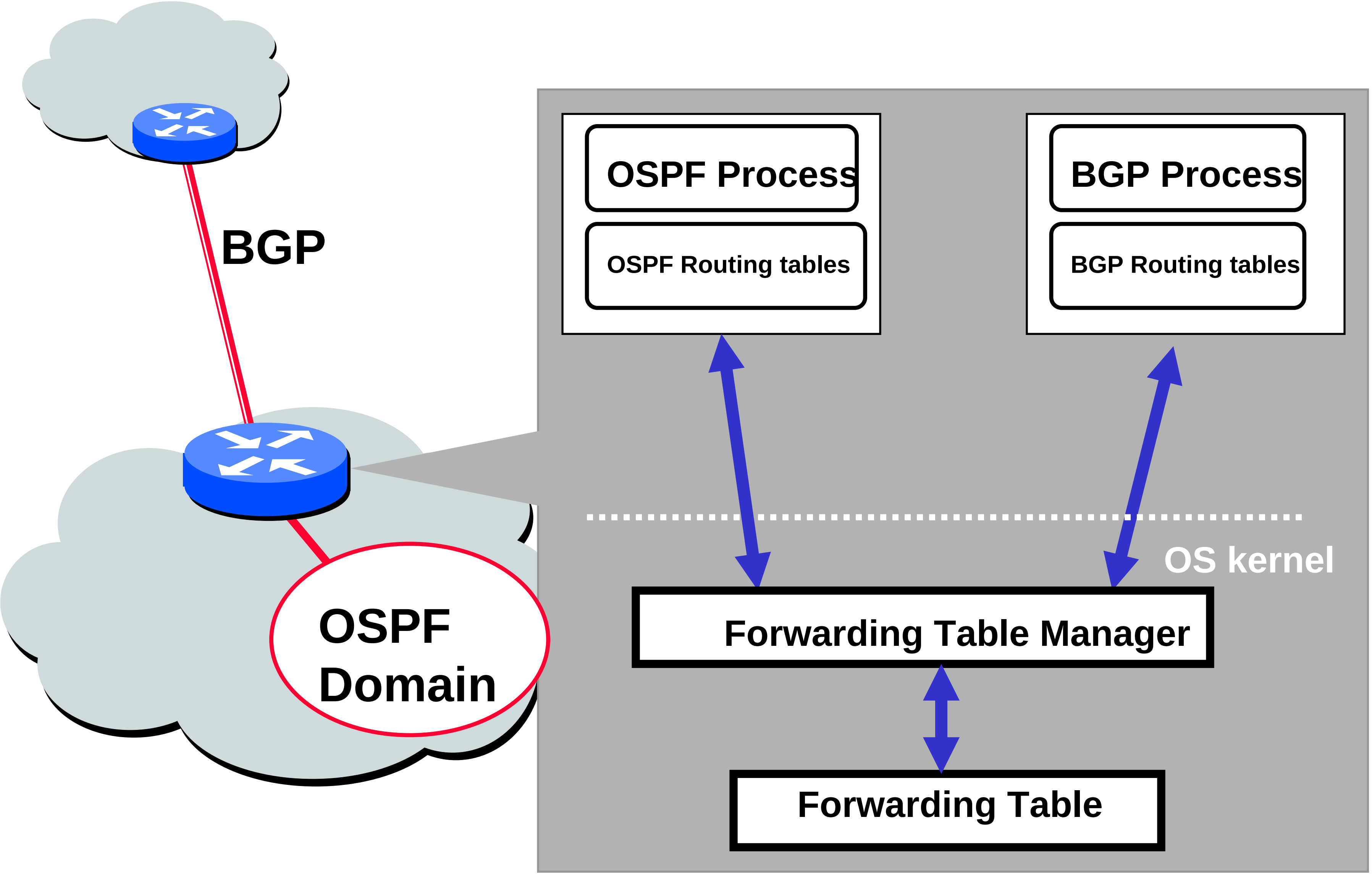
I-BGP

destination	next hop
135.207.0.0/16	192.0.2.1

Forwarding Table

destination	next hop
135.207.0.0/16	10.10.10.10
192.0.2.0/30	10.10.10.10

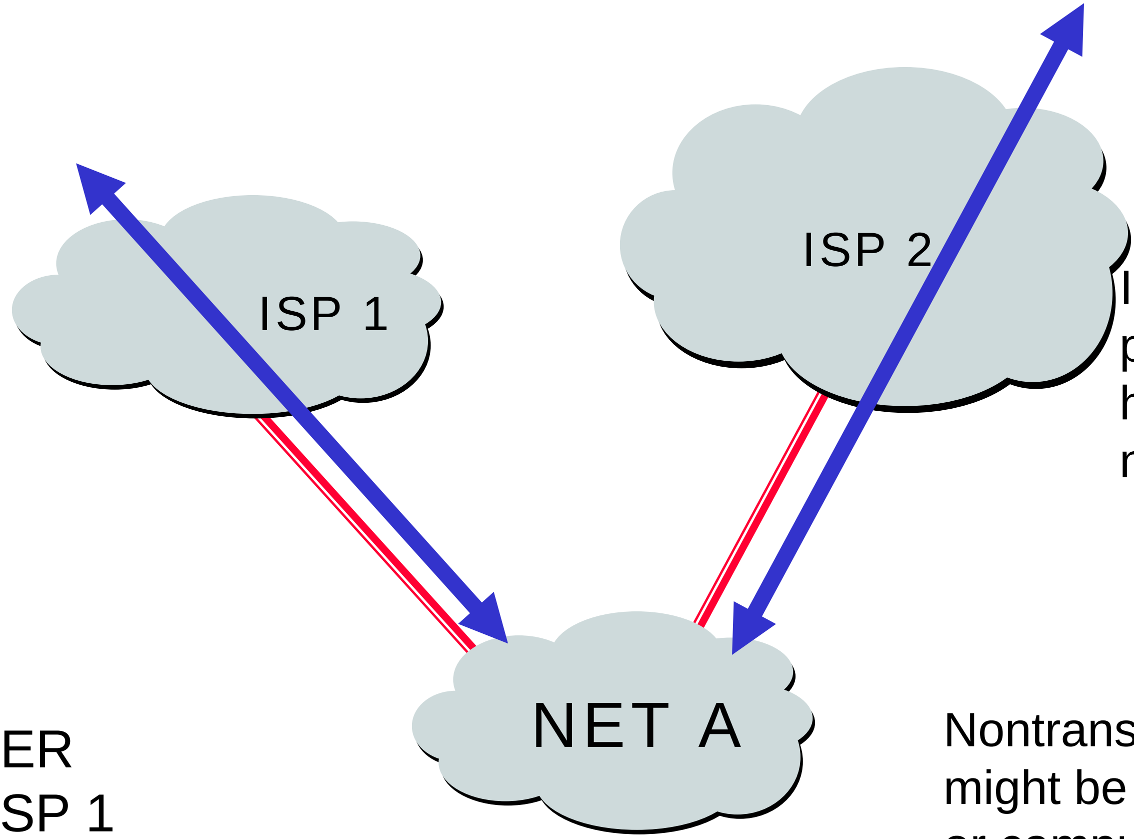
Multiple Routing Processes on a Single Router



Routing between ISPs

- Routing protocol (BGP) contains reachability information (no metrics)
  - Not about optimizing anything
  - All about policy (business and politics)
- Why?
  - Metrics optimize for a particular criteria
  - AT&T’s idea of a good route is not the same as UUnet’s
  - Scale
- What a BGP speaker announces or not announces to a peer determines what routes may get used by whom

Nontransit vs. Transit ASes

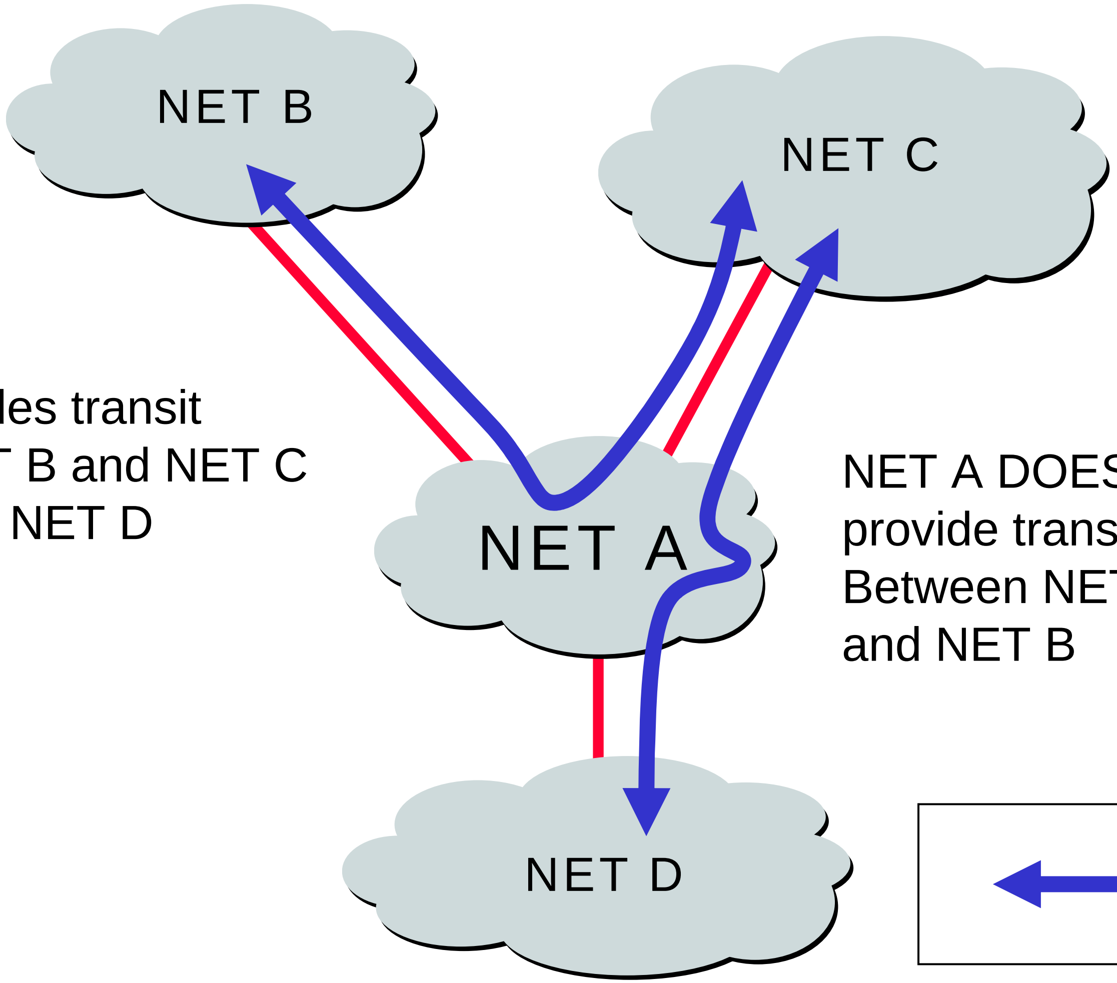


Internet Service providers (often) have transit networks

Traffic NEVER flows from ISP 1 through NET A to ISP 2 (At least not intentionally!)

Nontransit AS might be a corporate or campus network.

Selective Transit

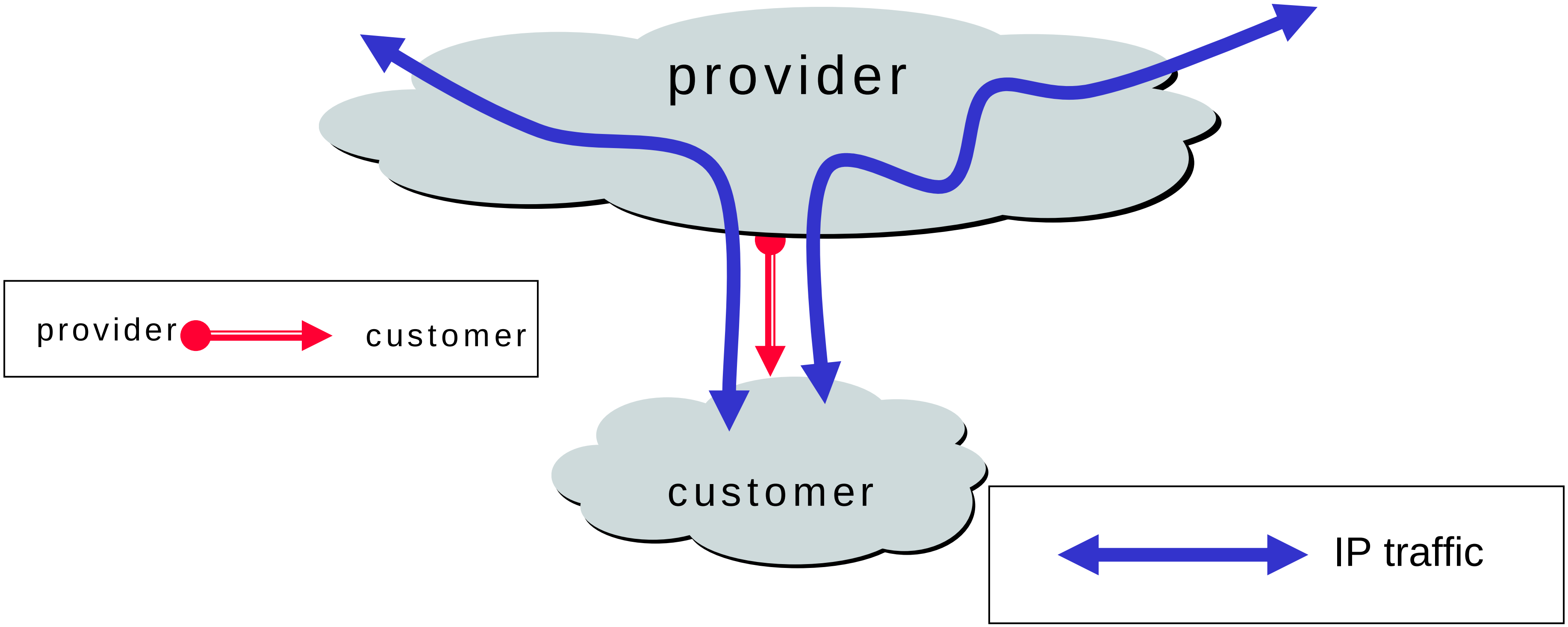


NET A provides transit between NET B and NET C and between NET D and NET C

NET A DOES NOT provide transit Between NET D and NET B

Most transit networks transit in a selective manner...

Customers and Providers



provider

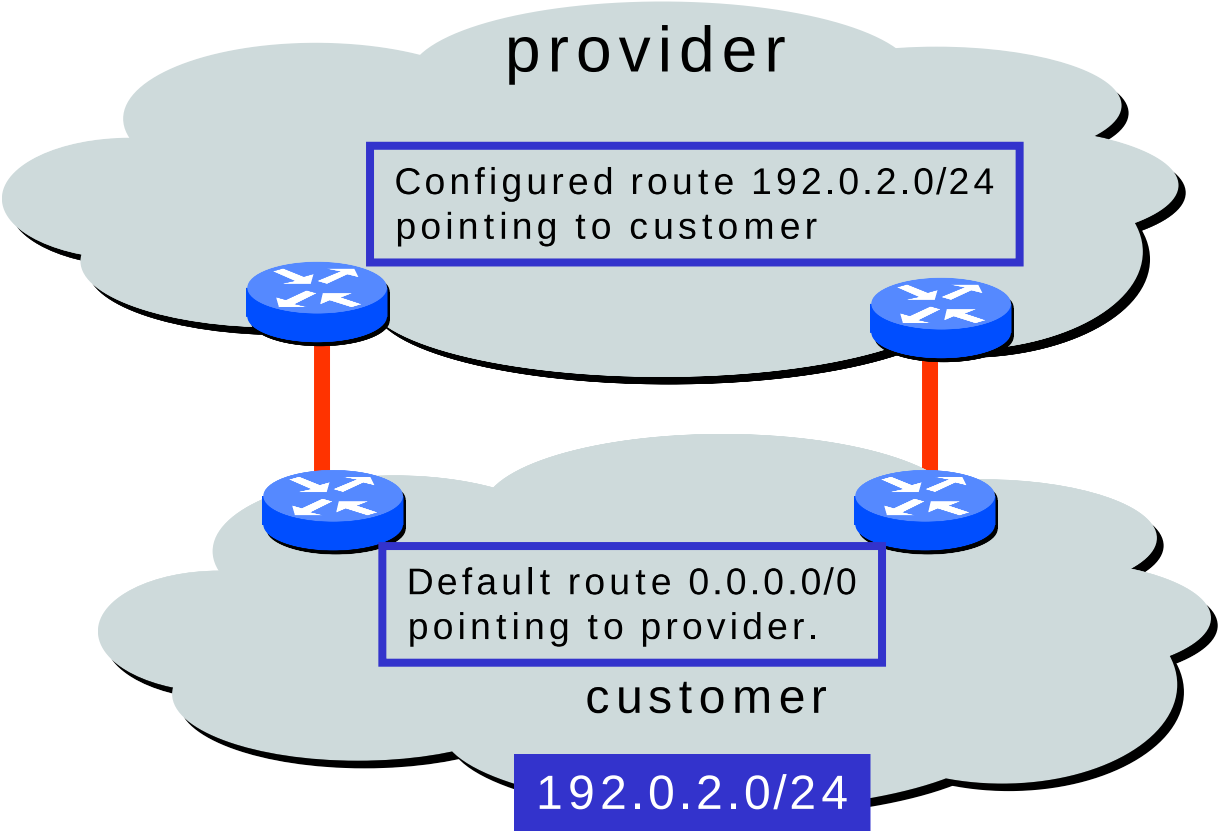
customer

provider customer

IP traffic

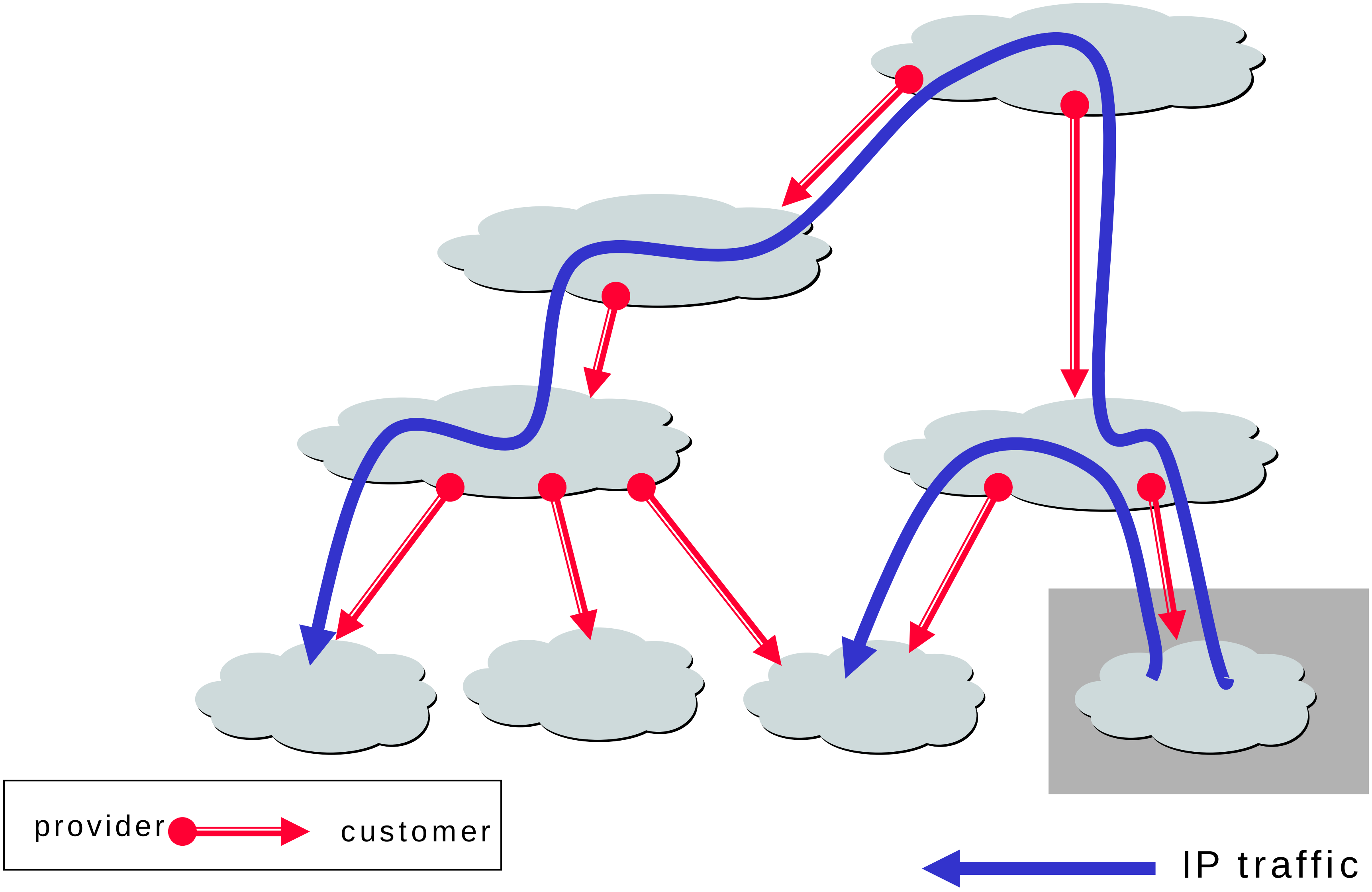
Customer pays provider for access to the Internet

Customers Don't Always Need BGP

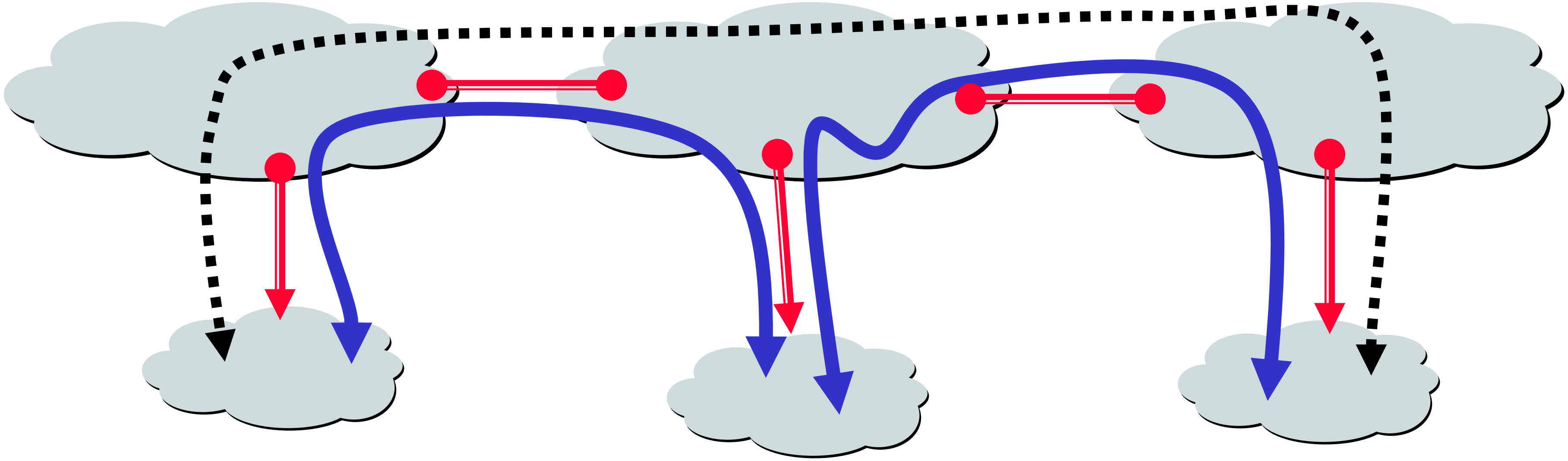


Static routing is the most common way of connecting an autonomous routing domain to the Internet.
This helps explain why BGP is a mystery to many …

Customer-Provider Hierarchy



The Peering Relationship



peer ●═● peer

provider ●═▶ customer

⟷ traffic allowed

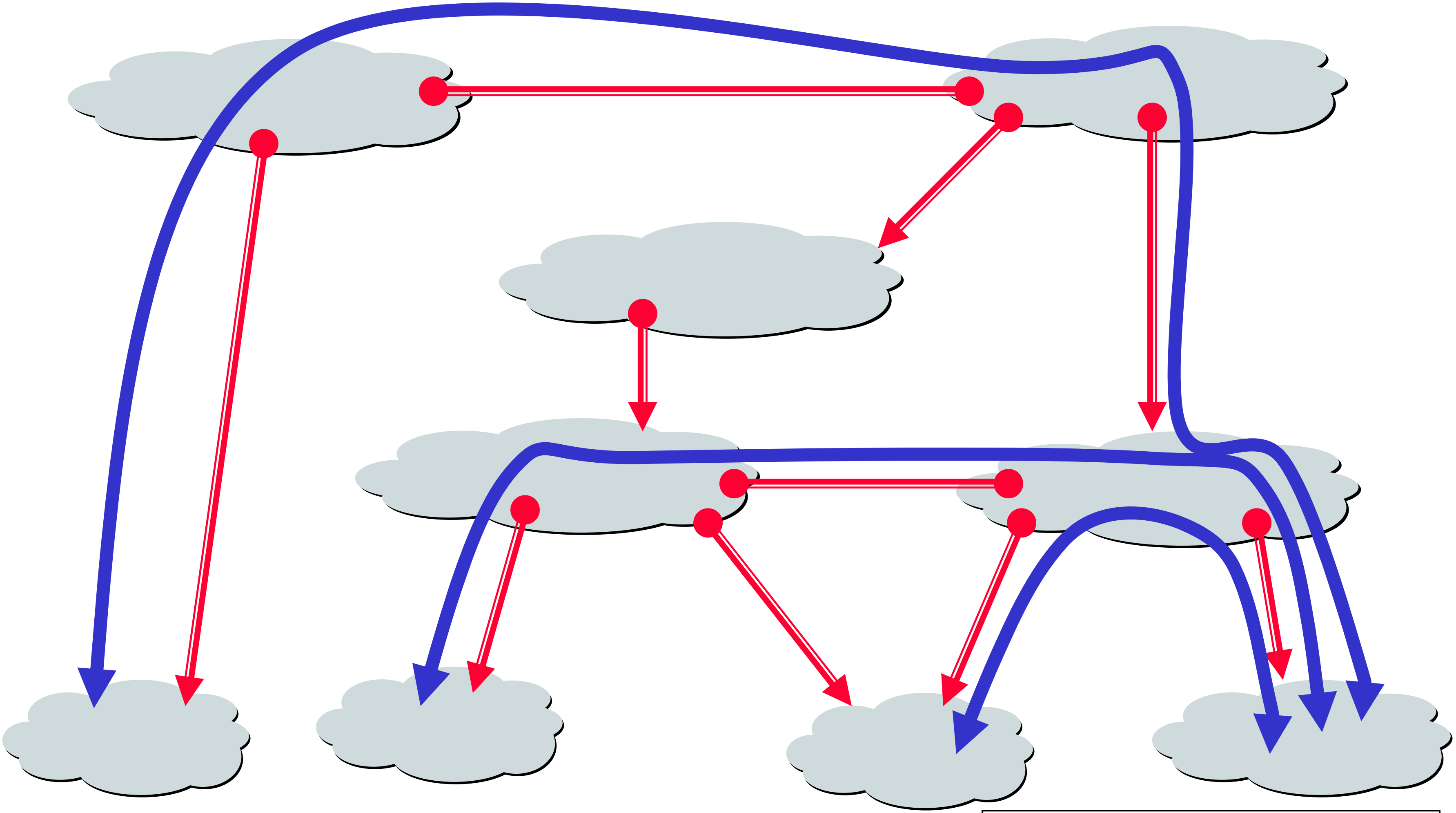
◀┅┅▶ traffic NOT allowed

Peers provide transit between their respective customers

Peers do not provide transit between peers

Peers (often) do not exchange $$$

Peering Provides Shortcuts



Peering also allows connectivity between the customers of “Tier 1” providers.

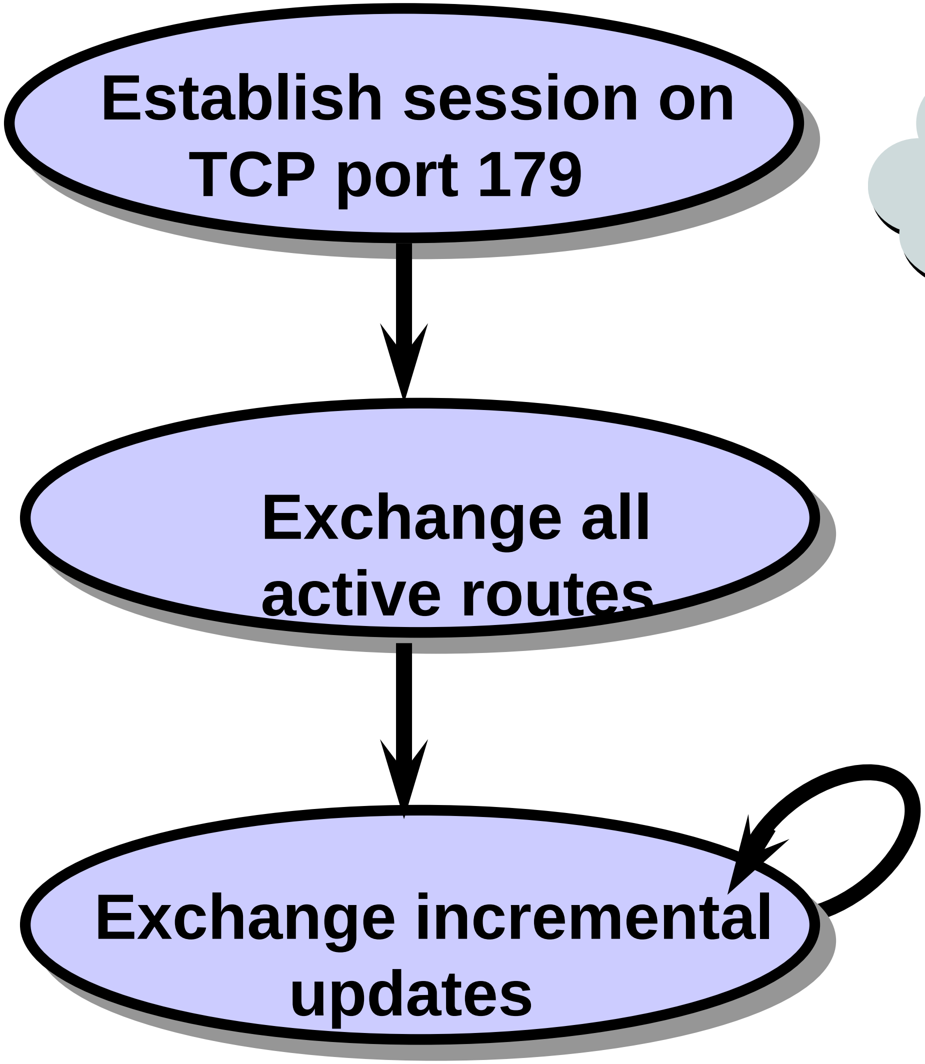
peer ●═● peer

provider ●═▶ customer

BGP: Path Vector Protocol

- Distance vector algorithm with extra information
  - For each route, store the complete path (ASs)
  - No extra computation, just extra storage
- Advantages:
  - can make policy choices based on set of ASs in path
  - can easily avoid loops

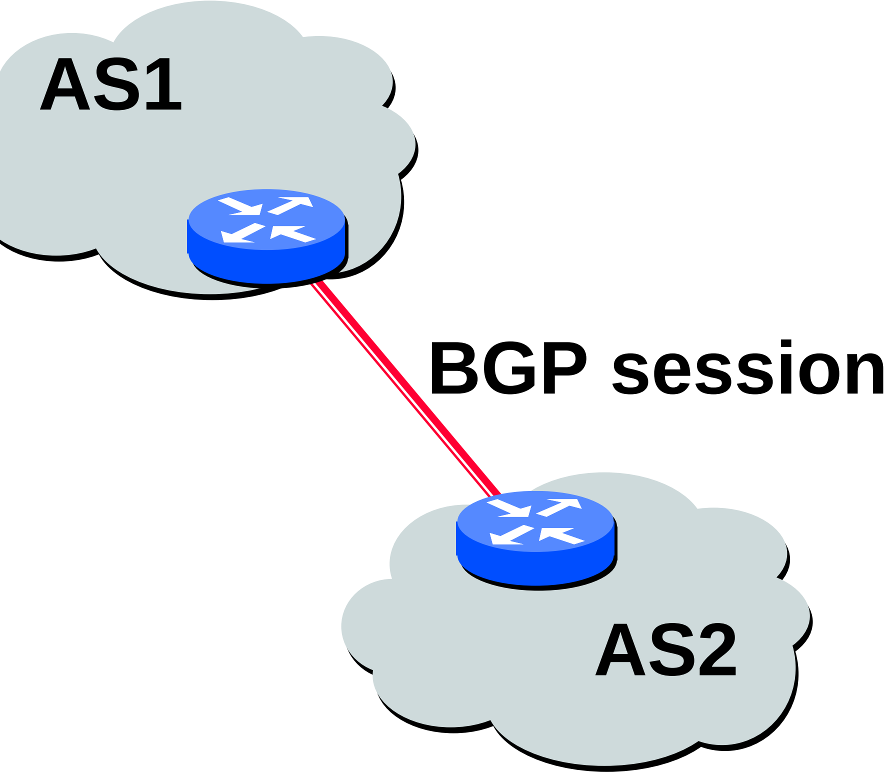
BGP Operations (Simplified)



Establish session on TCP port 179

Exchange all active routes

Exchange incremental updates



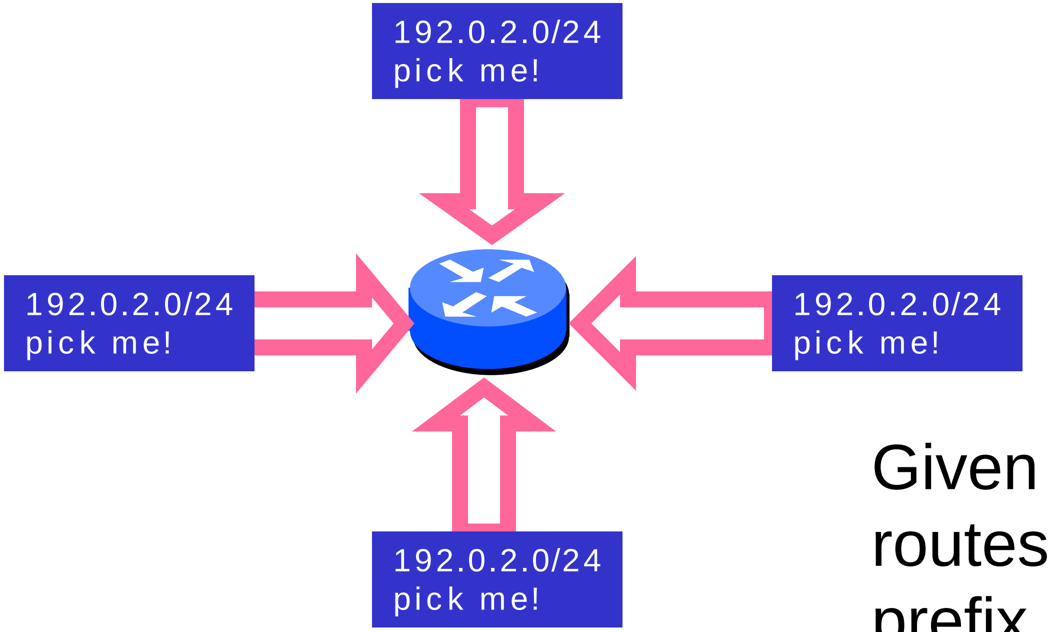
While connection is ALIVE exchange route UPDATE messages

Four Types of BGP Messages

- **Open** : Establish a peering session.
- **Keep Alive** : Handshake at regular intervals.
- **Notification** : Shuts down a peering session.
- **Update** : Announcing new routes or withdrawing previously announced routes.

Announcement
=
prefix + attributes values

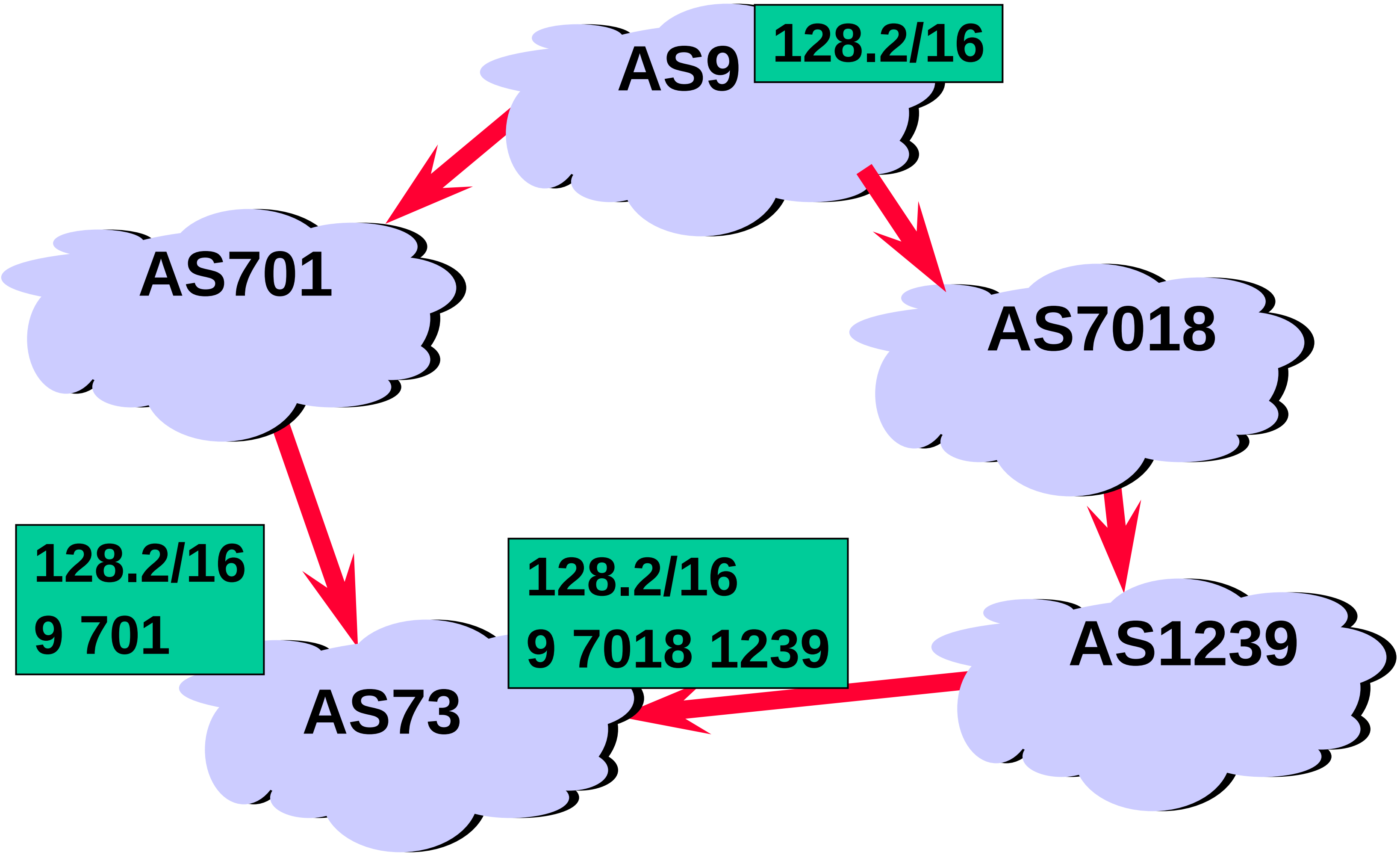
Attributes are Used to Select Best Routes



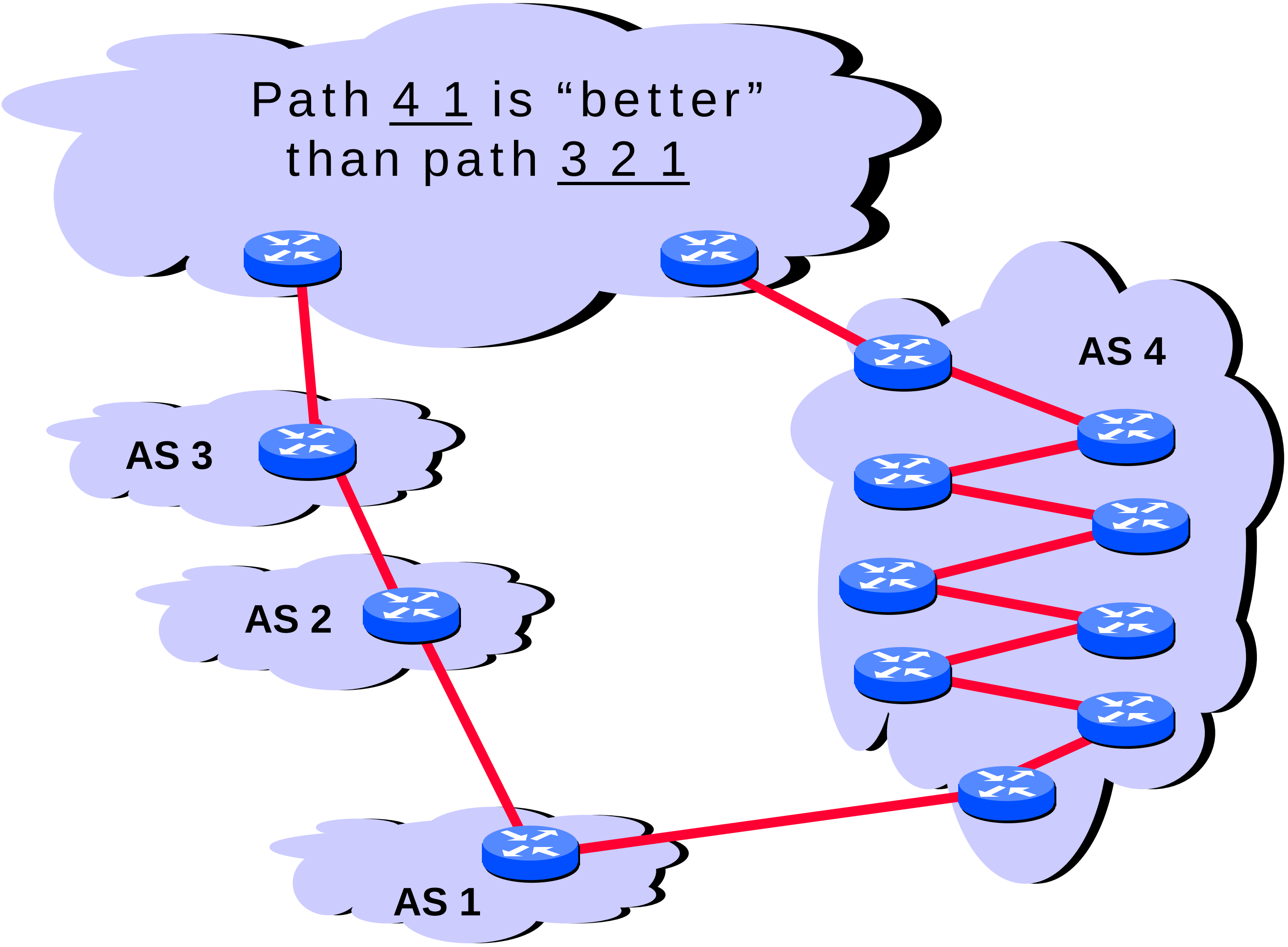
Given multiple routes to the same prefix, a BGP speaker must pick at most one best route

(Note: it could reject them all!)

Example: Multiple AS Paths



Shorter Doesn't Always Mean Shorter



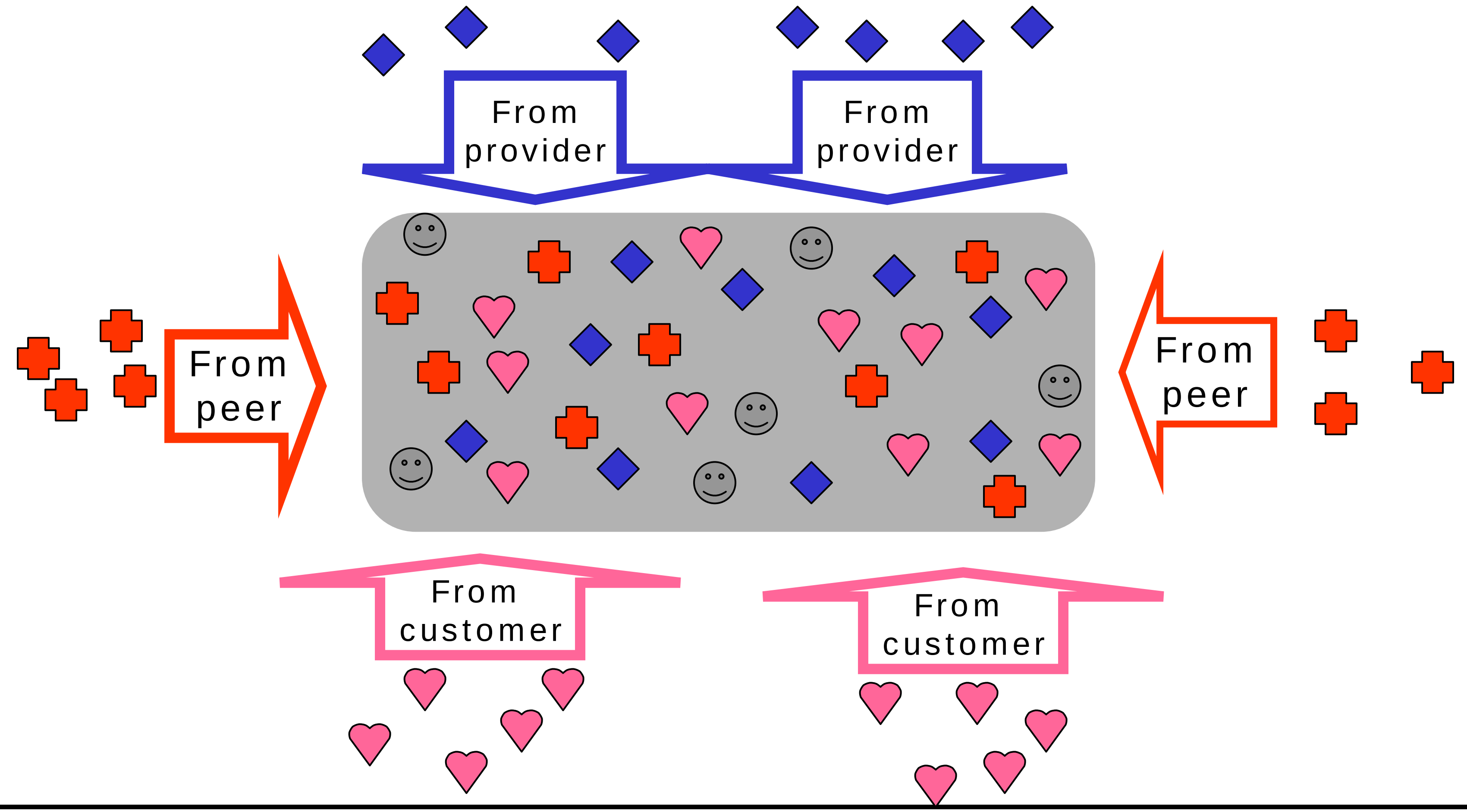
Path 4 1 is "better" than path 3 2 1

Implementing Customer/Provider and Peer/Peer relationships

- What you announce determines what route can be used by whom
- Enforce transit relationships
  - Outbound route filtering
- Enforce order of route preference
  - provider < peer < customer

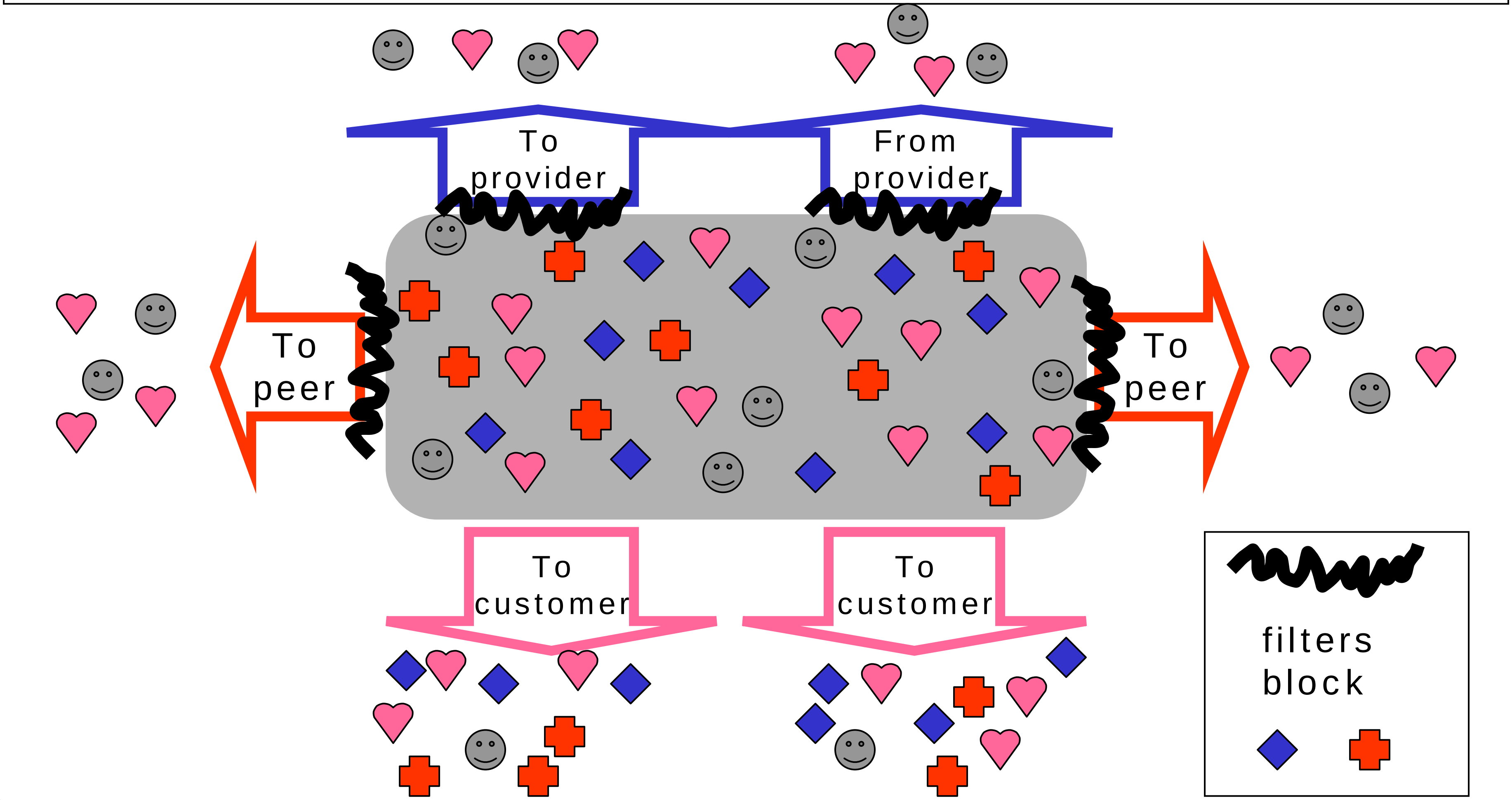
Import Routes

◆ provider route ✚ peer route ♥ customer route ☺ ISP route



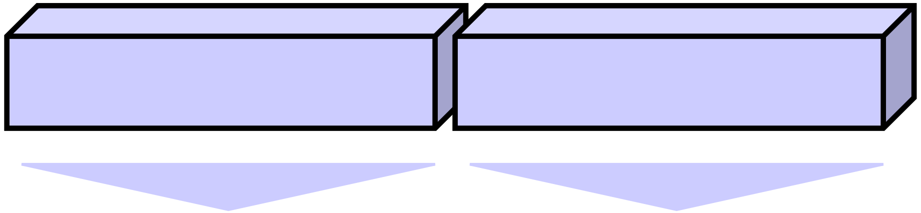
Export Routes

◆ provider route ✚ peer route ♥ customer route ☺ ISP route



How Can Routes be Colored? BGP Communities!

A community value is 32 bits



By convention, first 16 bits is ASN indicating who is giving it an interpretation

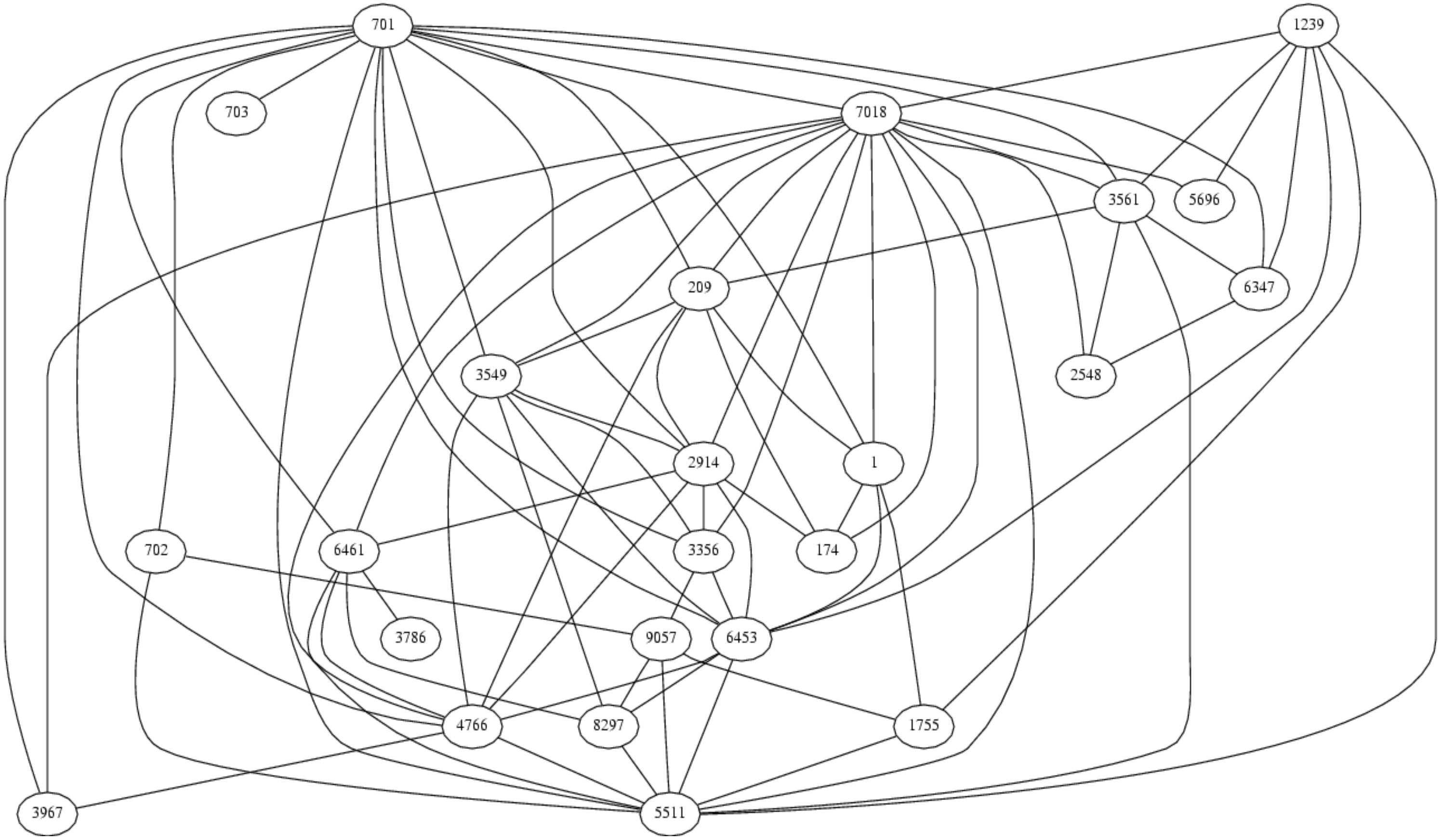
community number

Used for signaling within and between ASes

Very powerful BECAUSE it has no predefined meaning

Community Attribute = a list of community values. (So one route can belong to multiple communities)

Example AS Graph



The <u>subgraph</u> showing all ASes that have more than 100 neighbors in full graph of 11,158 nodes. July 6, 2001. Point of view: AT&T route-server

Does not reflect true topology

BGP Issues

- BGP designed for policy not performance
- Susceptible to router misconfiguration
  - Blackholes: announce a route you cannot reach
- Incompatible policies
  - Solutions to limit the set of allowable policies

More Issues

- Scaling the I-BGP mesh
  - Confederations
  - Route Reflectors
- BGP Table Growth
  - 140K prefixes and growing
  - Address aggregation (CIDR)
  - Address allocation
- AS number allocation and use
- Dynamics of BGP
  - Inherent vs. accidental oscillation
  - Rate limiting and route flap dampening
  - Lots and lots of redundant info
  - Slow convergence time