Lecture 3: Packet Forwarding

CSE 222A: Computer Communication Networks

Alex C. Snoeren

Thanks: Nick Feamster & Mike Freedman

Lecture 3 Overview

- Cerf & Kahn discussion
- The evolution of packet forwarding
- Routing review

IP: The Internet Protocol

- Service mode: *best effort*
  - No guarantees about reliable, in-order, or error-free delivery
  - Enables IP to “run over anything”

Version	HLen	TOS	Length	
Ident			Flags	Offset
TTL		Protocol	Checksum	
SourceAddr				
DestinationAddr				
Options (variable)				Pad (variable)
Data…				

Explosion of Middle Boxes

- Gateway/Router
  - Forward on destination IP address
  - Access control
  - Link scheduling and marking
  - Monitoring traffic
  - Deep packet inspection?
- Firewall
  - Access control on “five tuple” (and more)
- NAT
  - Mapping addresses and port numbers
- Shaper
  - Classify packets
  - Shape or schedule
- Packet sniffer
  - Monitoring traffic

Generalized Data Plane

- Streaming algorithms that act on packets
  - Matching on some bits, taking a simple action
  - … at behest of control and management plane
- Wide range of functionality
  - Forwarding
  - Access control
  - Mapping header fields
  - Traffic monitoring
  - Buffering and marking
  - Shaping and scheduling

Routing

- Goal: efficiently deliver packets between arbitrary hosts in the network
- Key concerns: scalability, performance, robustness
- Key techniques:
  - Hierarchical design
    - » Learn enough to do part of the job, hand task off to someone else
    - » e.g., one technique for inter-domain routing, another for intra-domain
  - Soft state
    - » Learn all the information you need to perform routing
    - » State info not for correctness, just performance optimizations

Forwarding Options

- Source routing (Myrinet)
  - Packet carries path
- Table of global addresses (IP)
  - Stateless routers
- Table of virtual circuits (ATM)
  - Small headers, small tables

- How do hosts/switches learn optimal network routes?
  - Given packet header, how to determine forwarding port

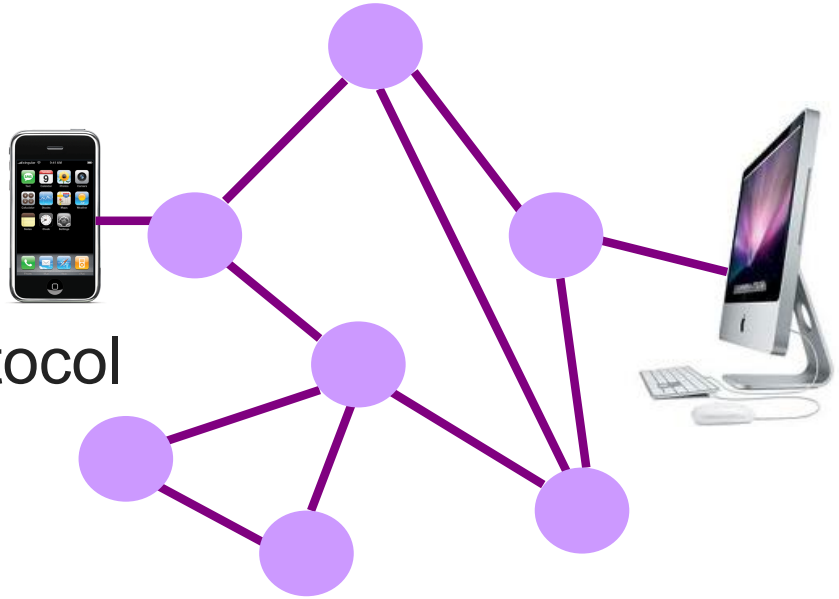
Comparison

	Source routing	**Global addresses**	**Virtual circuits**
Header size	worst	OK ~ large addrs	best
Router table size	none	# of hosts (prefixes)	# of circuits
Forward overhead	best	Prefix matching	Pretty good
Setup overhead	none	none	High
Error recovery	Tell all hosts	Tell all routers	Tear down circuit and reroute

Finding the Hosts

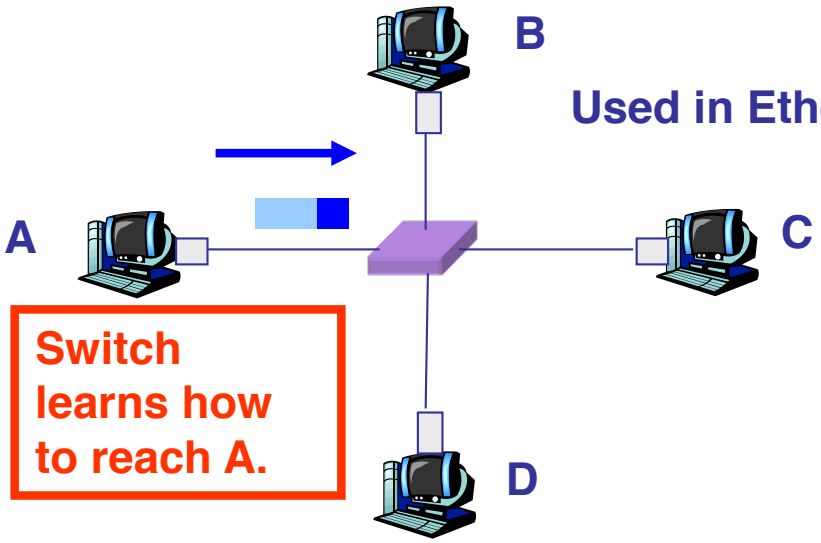
- Building a forwarding table
  - Computing paths between network elements
  - … and figuring out where the end-hosts are
  - … to map a destination address to an outgoing link

- How to find the hosts?
  - Learning/flooding
  - Injecting into routing protocol
  - Dissemination via different protocol
  - Directory service

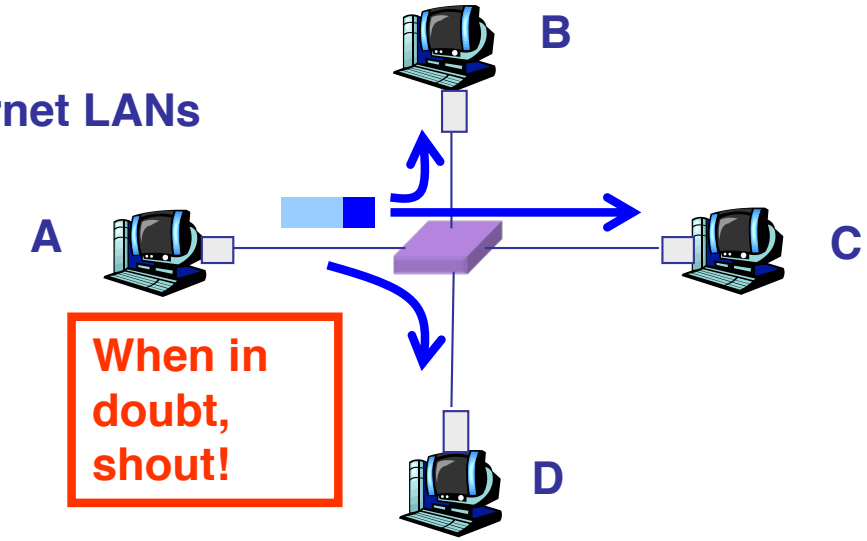


Learning and Flooding

- When a frame arrives
  - Inspect the *source* address
  - Associate address with the *incoming* interface
- When the frame has an unfamiliar destination
  - Forward out all interfaces
  - … except for the one where the frame arrived



Used in Ethernet LANs



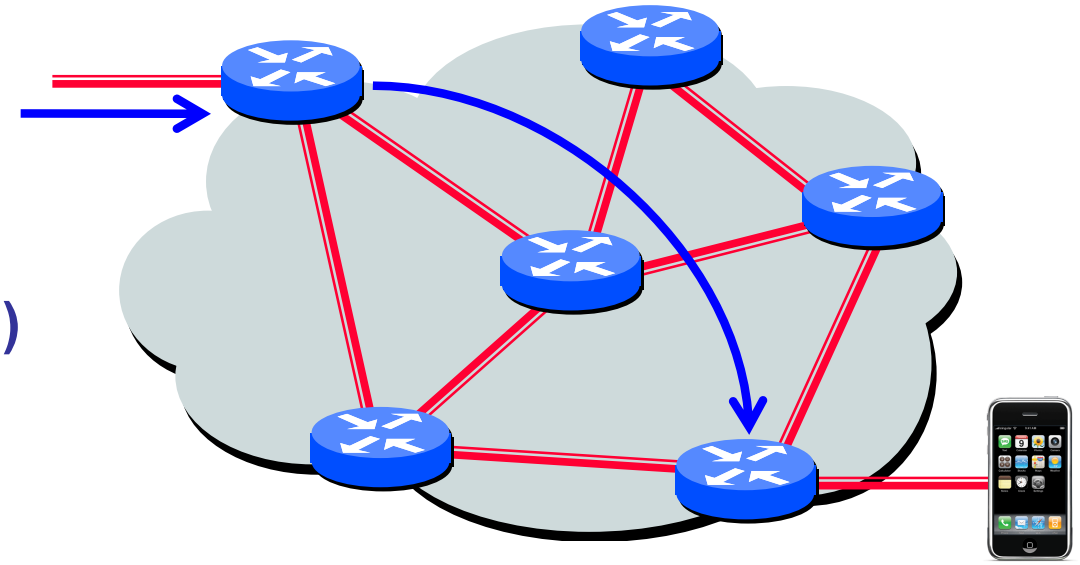
Disseminate w/Other Protocol

- Distribute using another protocol
  - One router learns the route
  - … and shares the information with other routers

disseminate route to other routers

learn a route to d (e.g., via BGP)

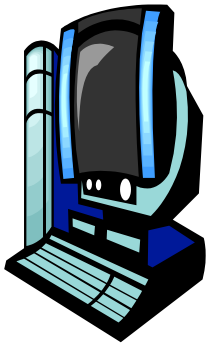
Internal BGP (iBGP) used in backbone networks



Directory Service

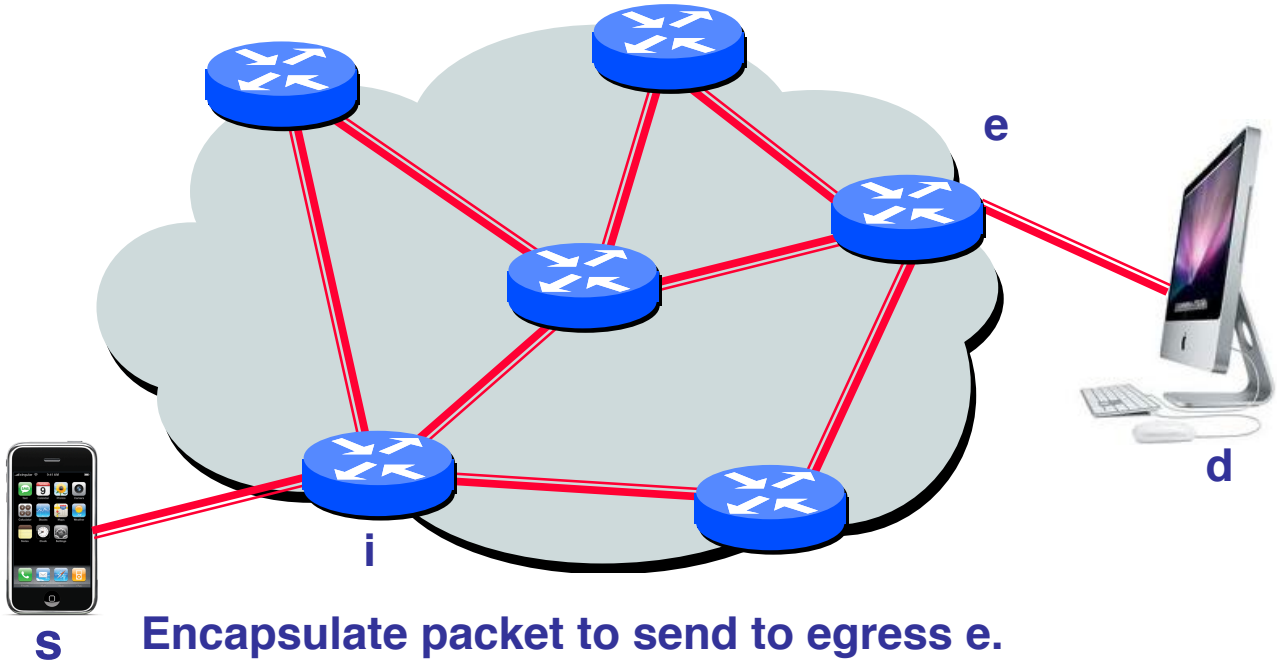
- Contact a service to learn the location
  - Lookup the end-host or subnet address
  - … and learn the label to put on the packet
  - … to get the traffic to the right egress point

directory



"Host d is at egress e"

Used in some data centers



Encapsulate packet to send to egress e.

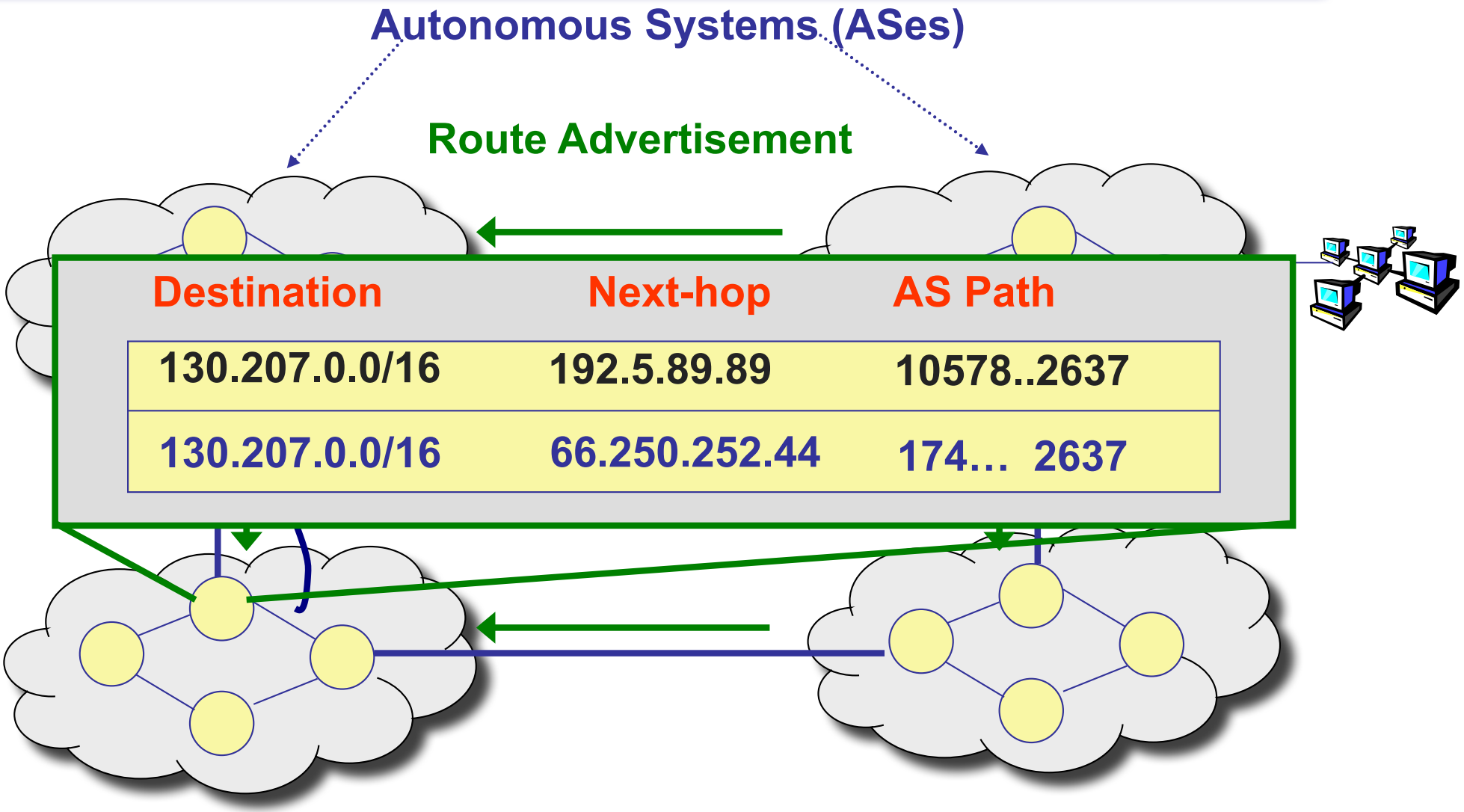
To Each His Own…

- Ethernet LAN: spanning tree, MAC learning, flooding
- Enterprise: link-state routing, injecting subnet addresses
- Backbone: link-state routing inside, path-vector routing with neighboring domains, and iBGP dissemination
- Data centers: many different solutions, still in flux
  - » E.g., link-state routing or multiple spanning trees
  - » E.g., directory service or injection of subnets into routing protocol

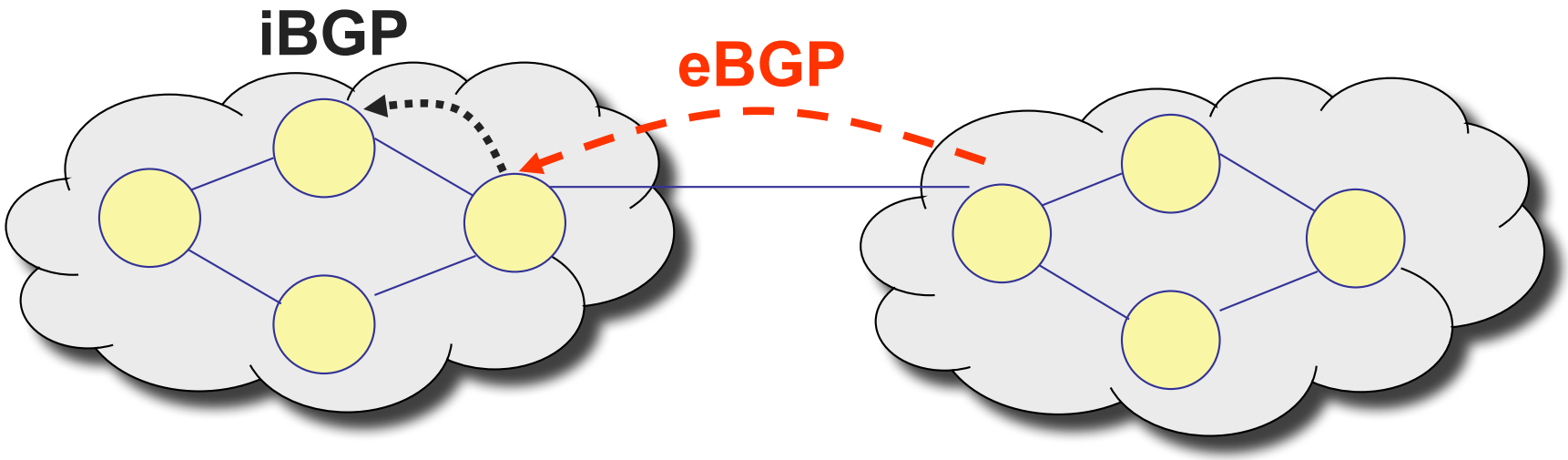
Internet: Hierarchical Routing

- Internet composed of many autonomous systems (AS's)
  - Correspond to administrative domains
- Each AS can choose its own routing algorithm
  - Routing Information Protocol (RIP) used originally
    Part of BSD distribution, distance vector
  - Open Shortest Path First (OSPF) currently most popular
    Link state protocol w/authentication, basic load balancing
- Border Gateway Protocol (BGP) for routing between AS's
  - Default: shortest number of AS's in path
  - Sys admins can express policy control
    Use AS *x* in preference to AS *y*

Internet Routing: BGP



Two Flavors of BGP



- **External BGP (eBGP):** exchanging routes *between* ASes
- **Internal BGP (iBGP):** disseminating routes to external destinations among the routers *within an AS*

Example BGP Routing Table

The full routing table

```
> show ip bgp

 Network          Next Hop        Metric LocPrf Weight Path
*>i3.0.0.0        4.79.2.1             0    110      0 3356 701 703 80 i
*>i4.0.0.0        4.79.2.1             0    110      0 3356 i
*>i4.21.254.0/23  208.30.223.5        49    110      0 1239 1299 10355 10355 i
* i4.23.84.0/22   208.30.223.5       112    110      0 1239 6461 20171 i
```

Specific entry. Can do longest prefix lookup:

```
> show ip bgp 130.207.7.237
BGP routing table entry for 130.207.0.0/16
Paths: (1 available, best #1, table Default-IP-Routing-Table)
  Not advertised to any peer
  10578 11537 10490 2637
    192.5.89.89 from 18.168.0.27 (66.250.252.45)
      Origin IGP, metric 0, localpref 150, valid, internal, best
      Community: 10578:700 11537:950
      Last update: Sat Jan 14 04:45:09 2006
```

Prefix

AS path

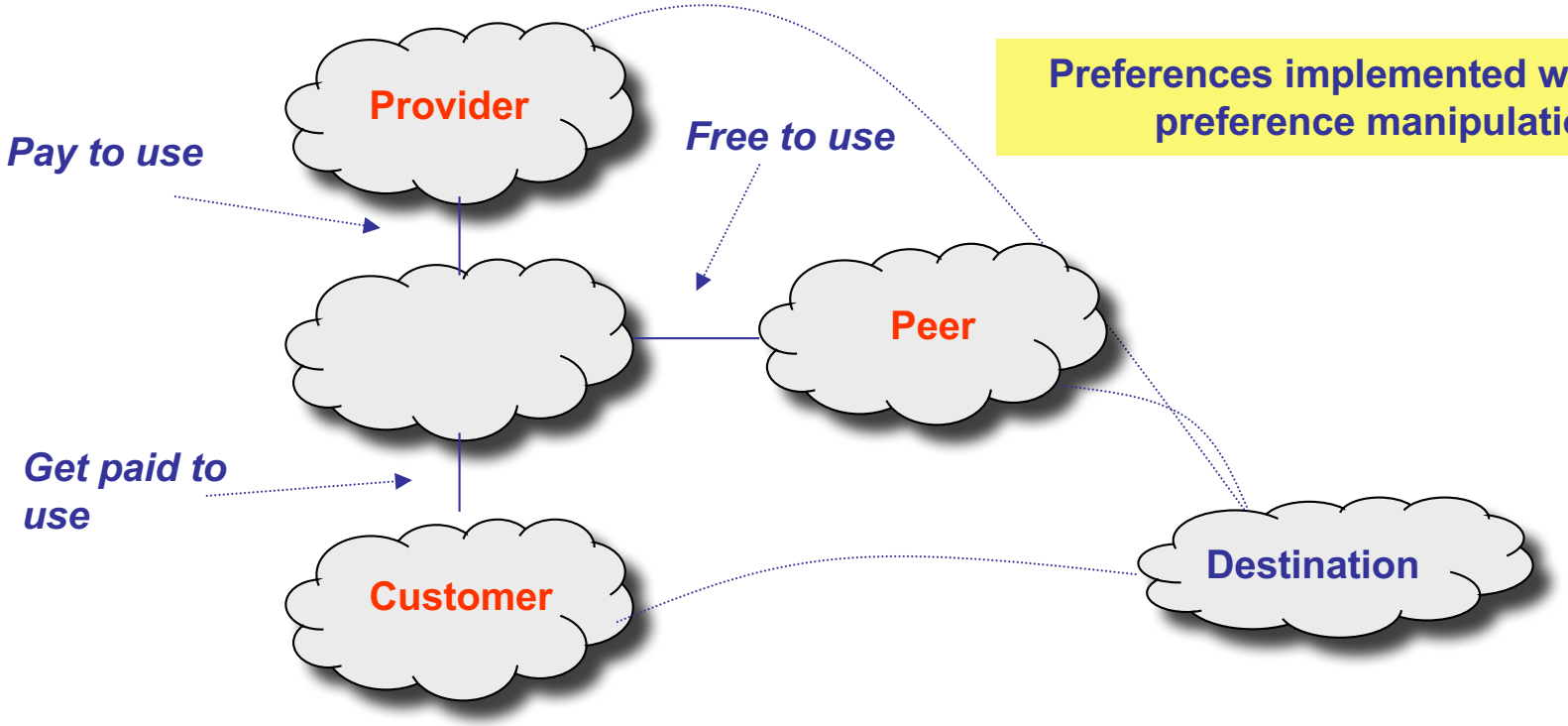
Next-hop

BGP Route Selection

BGP routes have the following attributes, on which the route selection process is based:

- **Local preference:** numerical value assigned by routing policy. Higher values are more preferred.
- **AS path length:** number of AS-level hops in the path
- **Multiple exit discriminator ("MED"):** allows one AS to specify that one exit point is more preferred than another. Lower values are more preferred.
- **eBGP over iBGP**
- **Shortest IGP path cost to next hop:** implements "hot potato" routing
- **Router ID tiebreak:** arbitrary tiebreak, since only a single "best" route can be selected

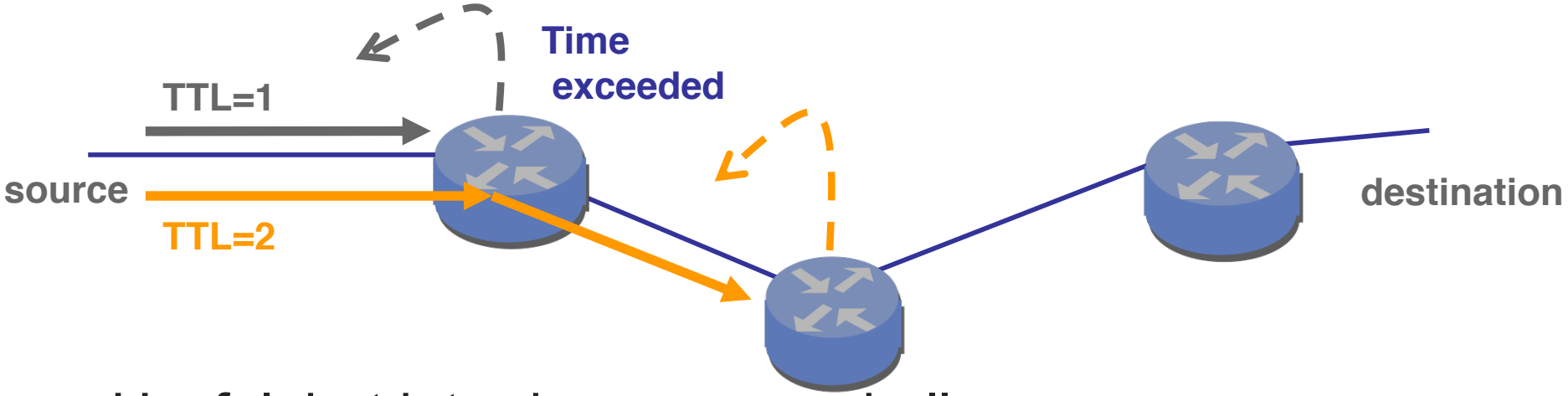
Internet Business Model (Simplified)



- **Customer/Provider:** One AS pays another for reachability to some set of destinations
- **"Settlement-free" Peering:** Bartering. Two ASes exchange routes with one another.

Measurement With Traceroute

- Traceroute tool to measure the forwarding path
  - Send packets with TTL=1, 2, 3…
  - Record the source of the "time exceeded" message



- Useful, but introduces many challenges
  - Path changes
  - Non-participating nodes
  - Inaccurate, two-way measurements

For Next Class…

- Read P&D Chapter 4
- Read and review Paxson ’97
- Keep thinking about term project ideas/groups
  - Initial ideas due next Wednesday
  - Email to me