Lecture 4: Routing

CSE 222A: Computer Communication Networks
Alex C. Snoeren

Thanks: Nick Feamster & Mike Freedman
Lecture 4 Overview

- Cerf & Kahn discussion

- Routing review
  - Alternative designs
  - BGP

- Paxson discussion
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

When in doubt, shout!

Used in Ethernet LANs

Switch learns how to reach A.
Inject into Routing Protocol

- Treat the end host (or subnet) as a node
  - And disseminate in the routing protocol
  - E.g., flood information about where addresses attach

Used in OSPF and IS-IS, especially in enterprise networks

CSE 222A – Lecture 4: Routing
Disseminate w/Other Protocol

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

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

Used in some data centers

“Host d is at egress e”

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 Routing: BGP

Autonomous Systems (ASes)

Route Advertisement

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-hop</th>
<th>AS Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>130.207.0.0/16</td>
<td>192.5.89.89</td>
<td>10578..2637</td>
</tr>
<tr>
<td>130.207.0.0/16</td>
<td>66.250.252.44</td>
<td>174... 2637</td>
</tr>
</tbody>
</table>

CSE 222A – Lecture 4: Routing
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

<table>
<thead>
<tr>
<th>Network</th>
<th>Next Hop</th>
<th>Metric</th>
<th>LocPrf</th>
<th>Weight</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>*&gt;i3.0.0.0</td>
<td>4.79.2.1</td>
<td>0</td>
<td>110</td>
<td>0</td>
<td>3356 701 703 80 i</td>
</tr>
<tr>
<td>*&gt;i4.0.0.0</td>
<td>4.79.2.1</td>
<td>0</td>
<td>110</td>
<td>0</td>
<td>3356 i</td>
</tr>
<tr>
<td>*&gt;i4.21.254.0/23</td>
<td>208.30.223.5</td>
<td>49</td>
<td>110</td>
<td>0</td>
<td>1239 1299 10355 10355 i</td>
</tr>
<tr>
<td>* i4.23.84.0/22</td>
<td>208.30.223.5</td>
<td>112</td>
<td>110</td>
<td>0</td>
<td>1239 6461 20171 i</td>
</tr>
</tbody>
</table>

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
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
Other BGP Attributes

- **Next-hop**: IP address to send packets en route to destination. *(Question: How to ensure that the next-hop IP address is reachable?)*

- **Community value**: Semantically meaningless. Used for passing around “signals” and labelling routes. More in a bit.
Local Preference

- Control over **outbound** traffic
- *Not* transitive across ASes
- Coarse hammer to implement route preference
- Useful for preferring routes from one AS over another (e.g., primary-backup semantics)
Customer expresses provider that a link is a backup
Affords *some* control over inbound traffic
Among routes with highest local preference, select route with shortest AS path length

Shortest AS path != shortest path, for any interpretation of “shortest path”
Hot-Potato Routing

- Prefer route with shorter IGP path cost to next-hop
- *Idea:* traffic leaves AS as quickly as possible

![Diagram showing traffic flow from New York to Dest. via Washington, DC and Atlanta.](image)

**Common practice:** Set IGP weights in accordance with propagation delay (e.g., miles, etc.)
Hot-Potato Dynamics

- Small changes in IGP weights can have significant impact
**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.

Preferences implemented with local preference manipulation
End-to-End Routing Behavior

- Importance of paper
  - Revitalized field of network measurement
  - Use of statistical techniques to capture new types of measurements
  - Empirical findings of routing behavior (motivation for future work)

- Various routing pathologies
  - Routing loops
  - Erroneous
  - Connectivity altered mid-stream
  - Fluttering…
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
Questions

- Why can’t we measure the Internet more directly?
  - What can we do about it?
- Right division of labor between host and network?
  - For path selection
  - For network monitoring
- How do we fix these routing problems?
  - In a decentralized, federated network
  - How to incentivize better network management
For Next Class…

- Read and review Wetherall ’98

- Keep thinking about term project ideas/groups
  - Initial ideas due on Tuesday
  - Email to Siva