Inter-domain Routing
Setting

• Start with simpler goal of *intra-domain routing*

• Routing is the process that all routers go through to calculate the routing tables
  
  • each router knows how to get to every destination in the network
Network as a Graph

• Routing is essentially a problem in graph theory

![Graph Diagram]

- X = router
- \ = link
- 1 = cost
Two Approaches

- **Link state routing:**
  - every node collects a representation of the entire graph and computes shortest paths
  - forwards packets along shortest paths to destination

- **Distance vector routing:**
  - each node knows only about its next hop links
  - each node maintains a vector of costs to all dests
  - periodically exchange with neighbors its routing table
• What are the issues that we have to take into account as we generalize this to a routing protocol for the Internet?

• What should be the goals of an ideal routing protocol for the Internet?
Setting

• BGP is the “inter-domain” routing protocol
  • Each “domain” is a separately administered entity
  • Also referred to as “autonomous systems” (ASes)
  • Each AS might have multiple prefixes (a contiguous set of addresses, e.g., MIT has 18.*.*.* and UW is 128.208.*.*.*)

• Routers route packets based on “longest prefix match”
  • Routing table contains the next hop based on a prefix basis
  • Find the best prefix match and route using it
  • What are the implications of using “longest prefix matching”? 
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
  - Peer-peer
  - Sibling
Customer-Provider Relationship

• Customer needs to be reachable from everyone
  • Provider ensures all neighbors can reach the customer
• Customer needs to reach everyone
• Payments in both directions
• Typically “95th percentile billing”
Peer-Peer relationship

- Peers exchange traffic between customers
  - AS lets its peer reach (only) its customers
  - Often the relationship is settlement-free (i.e., no $$)$
AS Structure

- Top of the Internet hierarchy
  - Has no upstream provider of its own
  - Typically has a large (inter)national backbone
  - Around 10-12 ASes: AT&T, Sprint, Level 3, ...

- Lower-layer providers (tier-2, ...)
  - Provide transit service to downstream customers
  - But need at least one provider of their own

- Stub ASes
  - Do not provide transit service
  - Connect to upstream provider(s)
What is BGP?

- Policy-based path vector routing
  - Path vector: a path of ASes
  - Respect the policies (customer-provider, peer-to-peer, etc.)
    - mechanism for filtering and selecting paths
    - at import and export time
• How powerful is the framework? How would you use import/export policies to influence routing?

• What are the implications of BGP policy-based routing?
# BGP Route Preferences

<table>
<thead>
<tr>
<th>Priority</th>
<th>Rule</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LOCAL_PREF</td>
<td>Highest LOCAL PREF (<a href="#">§4.2.3</a>). E.g., Prefer transit customer routes over peer and provider routes.</td>
</tr>
<tr>
<td>2</td>
<td>AS_PATH</td>
<td>Shortest AS_PATH length (<a href="#">§4.3.5</a>). Not shortest number of Internet hops or delay.</td>
</tr>
<tr>
<td>3</td>
<td>MED</td>
<td>Lowest MED preferred (<a href="#">§4.3.5</a>). May be ignored, esp. if no financial incentive involved.</td>
</tr>
<tr>
<td>4</td>
<td>eBGP &gt; iBGP</td>
<td>Did AS learn route via eBGP (preferred) or iBGP?</td>
</tr>
<tr>
<td>5</td>
<td>IGP path</td>
<td>Lowest IGP path cost to next hop (egress router). If all else equal so far, pick shortest internal path.</td>
</tr>
<tr>
<td>6</td>
<td>Router ID</td>
<td>Smallest router ID (IP address). A random (but unchanging) choice; some implementations use a different tie-break such as the oldest route.</td>
</tr>
</tbody>
</table>
Hot-Potato (Early-Exit) Routing

- Hot-potato routing
  - Each router selects the closest egress point
  - ... based on the path cost in intradomain protocol

- BGP decision process
  - Highest local preference
  - Shortest AS path
  - Closest egress point
  - Arbitrary tie break
Export Policy

- Modify attributes of the active route
  - To influence the way other ASes behave
  - Example: AS prepending

- Artificially inflate AS path length seen by others
  - Convince some ASes to send traffic another way
Prepending will (usually) force inbound traffic from AS 1 to take primary link.
Prepending Doesn’t Always Work

AS 1
provider

192.0.2.0/24
ASPATH = 2

primary
customer
AS 2

192.0.2.0/24

AS 3
provider

192.0.2.0/24
ASPATH = 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
**BGP Communities**

AS 1 provider

AS 2 customer

AS 3 provider

192.0.2.0/24

ASPATH = 2

Customer import policy at AS 3:

- If 3:90 in COMMUNITY then set local preference to 90
- If 3:80 in COMMUNITY then set local preference to 80
- If 3:70 in COMMUNITY then set local preference to 70

AS 3: normal customer local pref is 100, peer local pref is 90

192.0.2.0/24

ASPATH = 2

COMMUNITY = 3:70
Customer Installs Backup

customer sends community that lowers local preference below a provider’s primary backup.
Failure Happens!

Provider A (Tier 1) — Provider B (Tier 1)

Provider C (Tier 2)

customer

customer is happy that backup was installed ...
Primary is Repaired...

Provider A (Tier 1) --- Provider B (Tier 1) --- Provider C (Tier 2)

This is a stable BGP routing!

One “solution” --- reset BGP session on backup link!
• Is BGP secure?
• Is BGP high performant?
Observations

- There is no guarantee that a BGP configuration has a unique routing solution.
- When multiple solutions exist, the (unpredictable) order of updates will determine which one wins.
- There is no guarantee that a BGP configuration has any solution.
- Complex policies (weights, communities setting preferences, and so on) increase chances of routing anomalies.