TOC – Interdomain Routing

- Review
- Why Hierarchical Routing?
- Interconnections
- Big Picture
- Peering and Transit
- Reachability
- BGP
Formulate the routing problem as a Shortest Path Problem
- Link State v/s Distance Vector
- Both work reasonably well in a well engineered network
Why Hierarchical Routing?

- Is a natural way for routing to scale
  - Size
  - Network Administration
  - Governance
- Allows multiple metrics at different levels of the hierarchy
- Exploits address aggregation and allocation
Interconnections

- The internet is an interconnection of unequal networks
- Interconnection arrangements drive
  - the competitive landscape
  - the robustness of the network
  - End-to-end performance
- Interconnection is central to all large networks
  - Voice
  - Data
  - Wireless
  - Cable
Interconnections

- [www.thelist.com](http://www.thelist.com)
  - How many ISP’s in the 415 area code?
    - That start with A-C: about 200…
    - Just DSL that start with A-C: about 80

- In the telephone network
  - How many independent telephone companies in 1894-1902 in the US?
    - 3039 commercial companies, 979 co-operatives
  - By controlling interconnection Bell got rid of them
  - Interconnection is now regulated (CLECs)
The Internet contains a large number of diverse networks
Peering & Transit

- Peering
  - The business relationship whereby ISPs reciprocally provide to each other connectivity to each others’ transit customers

- Transit
  - The business relationship whereby one ISP provides (usually sells) access to all destinations in it’s routing table

William B. Norton, “Internet Service Providers and Peering”
West and East Peer with USNet but they can’t reach each other
By EastNet purchasing transit, EastNet is announced by USNet to USNet Peering and Transit interconnections alike.

...for a (transit) fee of course.

Figure from William B. Norton, “Internet Service Providers and Peering”
Peering & Transit
Benefits of Transit v/s Peering

William B. Norton, “Internet Service Providers and Peering”
Peering & Transit
Moving from Transit to Peering

William B. Norton, “Internet Service Providers and Peering”
Reachability

- Interdomain routing is about implementing policies of reachability
  - Routing efficiency and performance is important, but not essential
- ISPs could be competitors and do not want to share internal network statistics such as load and topology
- Use Border Gateway Protocol (BGP)
  - Border routers communicate over TCP port 179
  - A Path Vector Protocol
    - Communicate entire paths: Route Advertisements
  - A Router Can be involved multiple BGP sessions
BGP – Border Gateway Protocol

- Concept
- I-BGP & E-BGP
- Border Routers
- Sharing Routes
- Message Types
- Issues
- BGP - Protocol
Concept

you can reach net A via me

Share connectivity information across ASes

Table at R1:

<table>
<thead>
<tr>
<th>dest</th>
<th>next hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>R2</td>
</tr>
</tbody>
</table>

TOC – IP – Routing – Types – Interdomain – BGP – Concept
I-BGP and E-BGP

IGP: Interior Gateway Protocol. Example: OSPF

I-BGP

AS1
R1

AS2
R2
R3
A

AS3
R4
R5
B

E-BGP

announce B

border router

internal router

Border Routers

- Border Routers
  - from the same AS speak IBGP
  - from different AS’s speak EBGP
- EBGP and IBGP are essentially the same protocol
  - IBGP can only propagate routes it has learned directly from its EBGP neighbors
  - All routers in the same AS form an IBGP mesh
  - Important to keep IBGP and EBGP in sync
Sharing routes

- One router can participate in many BGP sessions.
- Initially … node advertises ALL routes it wants neighbor to know (could be > 50K routes)
- Ongoing … only inform neighbor of changes
Message types: 4

- Open: Session establishment id exchange
- Notification: exception driven information
- Keep Alive: soft state
- Update: path vector information
Issues

- Advertising a Route
- Routing Table Scaling
- Update Message
Advertising a Route

- One router telling another one
  - The prefix
  - IP address of the next hop
  - Path list of AS’s that the announcement has passed through
    - Since announcement propagates from destination, this yields the path

- No refresh messages required
- The announcing router will follow the path itself
Many small networks
Aggregation hides a lot…
BGP Update Message

- Contains information about
  - New Routes
  - Withdrawn Routes: No longer valid
  - Path Attributes:
    - Path Weights
    - Multiple Exit Discriminators
    - Local Preferences
    - Etc.

- Attribute information allows policies to be implemented
BGP – Protocol

- Path Vector
- Multi-Homing
- Multi-Exit Discriminators
- Routing Process Overview
- Attribute
- Choosing Best Route
- BGP Policies
- Transit vs. Non-Transit AS
- Customer Transit Problem
- BGP and Performance
- Skitter
- Summary
Path Vector

ner-routes>show ip bgp

BGP table version is 6128791, local router ID is 4.2.34.165
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

<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>* i3.0.0.0</td>
<td>4.0.6.142</td>
<td>1000</td>
<td>50</td>
<td>0</td>
<td>701 80 i</td>
</tr>
<tr>
<td>* i4.0.0.0</td>
<td>4.24.1.35</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>i</td>
</tr>
<tr>
<td>* i12.3.21.0/23</td>
<td>192.205.32.153</td>
<td>0</td>
<td>50</td>
<td>0</td>
<td>7018 4264 6468 ?</td>
</tr>
<tr>
<td>* e128.32.0.0/16</td>
<td>192.205.32.153</td>
<td>0</td>
<td>50</td>
<td>0</td>
<td>7018 4264 6468 25 e</td>
</tr>
</tbody>
</table>

- Every route advertisement contains the entire AS path
  - Generalization of distance vector
- Can implement policies for choosing best route
- Can detect loops at an AS level
Multihoming

- Two or more interdomain connections between the same AS’s
- Two or more interdomain connections between a customer and different ISPs
Multiexit Discriminators (MEDs)

One AS influences the decisions of a neighboring AS

- AS_A wants to tell AS_B that network x is closer to router 2 than to router 3
  - Router 2 advertises a smaller MED value for x than Router 3
  - AS_B prefers the path to x that does not go through 6 and 3
  - AS_B does not propagate MEDS from AS_A any further
Routing Process Overview

Import Policy Engine → Decision process → Routes used by router → Export Policy Engine

- Routes received from neighbors
- BGP table
- IP routing table

Choose best route

- accept, deny, set preferences
- forward, not forward set MEDs
Attribute: Local Preference

- Used to indicate preference among multiple paths for the same prefix *anywhere* in the Internet.
- The higher the value the more preferred.
- Exchanged between IBGP peers only. Local to the AS.
- Often used to select a specific exit point for a particular destination.

**BGP table at AS4:**

<table>
<thead>
<tr>
<th>Destination</th>
<th>AS Path</th>
<th>Local Pref</th>
</tr>
</thead>
<tbody>
<tr>
<td>140.20.1.0/24</td>
<td>AS3 AS1</td>
<td>300</td>
</tr>
<tr>
<td>140.20.1.0/24</td>
<td>AS2 AS1</td>
<td>100</td>
</tr>
</tbody>
</table>
Choosing best route

- Choose route with highest LOCAL_PREF
  - Preference-based routing
- If multiple choices, select route with shortest hop-count
- If multiple choices for same neighboring AS, choose path with max MED value
- Choose route based on lowest origin type
  - IGP < EGP < INCOMPLETE
- Among IGP paths, choose one with lowest cost
- Finally use router ID to break the tie.
BGP Policies

- Multiple ways to implement a “policy”
  - Decide not to propagate advertisements
    - I’m not carrying your traffic
  - Decide not to consider MEDs but use shortest hop
    - Hot potato routing
  - Prepend own AS# multiple times to fool BGP into not thinking AS further away
  - Many others…
Transit vs. Nontransit AS

Transit traffic = traffic whose source and destination are outside the AS

**Nontransit AS**: does not carry transit traffic
- Advertise own routes only
- Do not propagate routes learned from other AS’s

**Transit AS**: does carry transit traffic
- Advertises its own routes PLUS routes learned from other AS’s

Customer-Transit Problem

- Assume that the small ISP is a customer of two large ISPs
- If customer ISP does not obey export rules
  - forwards advertisements from one large ISP to another
  - Carries huge volume of transit traffic between two large ISPs
BGP and Performance

- BGP isn’t designed for policy routing not performance
  - Hot Potato routing is most common but suboptimal
  - Performance isn’t the greatest
- 20% of internet paths inflated by at least 5 router hops
- Very susceptible to router misconfiguration
  - Blackholes: announce a route you cannot reach
    - October 1997 one router brought down the internet for 2 hours
  - Flood update messages (don’t store routes, but keep asking your neighbors to clue you in). 3-5 million useless withdrawals!
- In principle, BGP could diverge
  - Various solutions proposed to limit the set of allowable policies
  - Focuses on avoiding “policy cycles”
4/1-4/16 2002
• 1,224,733 IP addresses,
• 2,093,194 IP links,
• 932,000 destinations,
• 70% of globally routable network prefixes;
• 10,999 ASes (84% of ASes),
• 34,209 peering sessions
Plot the AS based on polar co-ordinates \((r, \theta)\):

- \(r = 1 - \log \left(\frac{\text{As degree} + 1}{\text{Max Degree} + 1}\right)\)
  - Higher the degree lower the radius
- \(\theta = \text{longitude of AS headquarters}\)
The Internet is composed of various ASes which use BGP to inter-network themselves.

The Internet uses classless addressing.

BGP as a routing protocol:
- Path-vector based
- Supports route-aggregation
- Supports preferential routing
- Uses Import and Export policies

BGP is the protocol that “holds” the Internet together.
Summary (cont.)

BGP

Transit; Peering Agreements; Customer-Provider