Interdomain Routing and Connectivity

Brighten Godfrey
CS 538 October 1 2013
Choosing paths along which messages will travel from source to destination.
Problems for intradomain routing

- Distributed path finding
- Optimize link utilization (traffic engineering)
- React to dynamics
- High reliability even with failures
- Scale
Problems for interdomain routing

All of intradomain’s problems

Bigger scale

Multiple parties

- No central control
- Conflicting interests
- Greater volume and diversity of attacks

Harder to change architecture

- Intradomain evolution: RIP, ISIS, OSPF, MPLS, OpenFlow, ...
- Interdomain: BGP.
BGP: Border Gateway Protocol

Distance vector variant

- Send incremental changes, not whole vector
- Path vector: Remember path instead of distance

Why path vector?

- Avoid DV’s transient loops; but more importantly...
- **Policy support:** can pick any path offered by neighbors, not necessarily the shortest (Link State cannot)
- **Privacy support:** path choice policy is applied locally, not announced globally

- Q: How much privacy is there?
BGP: The picture at one router

Updates from neighbors

Import policies

Import policies

Import policies

“Adj-RIB-In”

Route selection

Best route for each destination (“Loc-RIB”)

Updates to neighbors (or not)

Export policies

Export policies

Export policies

Forwarding table (FIB) in data plane

prefix

next-hop
BGP is a relatively simple protocol with a few salient features. Each prefix is associated with a set of routing policies, and each router is assigned a single autonomous system (AS) number. When a router receives routing information from one or more outgoing links along shortest path, it uses this information to construct a forwarding table that maps each destination prefix to another. Each router combines the BGP and IGP information to decide which routes to use within its AS. An AS is usually run an Interior Gateway Protocol (IGP) to route traffic within the same AS using internal-BGP (iBGP). In addition, the routers exchange information through BGP sessions with routers in neighboring ASes. These sessions are used to distribute routes learned on these sessions to the routers in the AS and border routers in neighboring ASes. These changes may be established between neighbors or distributed through new advertisements.

This process is extended in many real implementations.

### Route selection process

<table>
<thead>
<tr>
<th>Step</th>
<th>Attribute</th>
<th>Controlled by local or neighbor AS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Highest LocalPref</td>
<td>local</td>
</tr>
<tr>
<td>2.</td>
<td>Lowest AS path length</td>
<td>neighbor</td>
</tr>
<tr>
<td>3.</td>
<td>Lowest origin type</td>
<td>neither</td>
</tr>
<tr>
<td>4.</td>
<td>Lowest MED</td>
<td>neighbor</td>
</tr>
<tr>
<td>5.</td>
<td>eBGP-learned over iBGP-learned</td>
<td>neither</td>
</tr>
<tr>
<td>6.</td>
<td>Lowest IGP cost to border router</td>
<td>local</td>
</tr>
<tr>
<td>7.</td>
<td>Lowest router ID (to break ties)</td>
<td>neither</td>
</tr>
</tbody>
</table>

[Caesar, Rexford, IEEE Network Magazine, 2005]
Common policies

Route selection: prefer customer over peer over provider
Common policies

Route selection: prefer customer over peer over provider

But ... What’s wrong with this picture?
Falsely assumed all routes are exported
Common policies

Route export (most common): to/from customer only ("valley-free")
How does BGP inbound traffic engineering fit with TeXCP? Are they solving the same problem?

How can ISPs perform interdomain outbound TE?

2#1. The sequence of ISPs (AS numbers and/or business names) from the last step.

13030 11537 40387 38
31500 174 40387 38
8928 7132 40387 40387 40387 40387 38
1299 174 40387 38
5413 1299 174 40387 38
6067 174 40387 38
8426 3549 11537 40387 38
19151 11537 40387 38
6939 11537 40387 38
Interconnection: Traditional view

Tier 1’s

Mid-tier

Stub / Leaf

Hierarchical, limited peering at lower tiers
Significant and increasing peering at lower tiers
Significant peering

- Estimated 200,000 peerings just in Europe
- More than 2x as many as non-peering links!

These peerings missed in past measurements

Figure 2: Peering links and visibility in control/data plane (normalized by number of detected P-P links).
Why measurements miss so much

Tier 1’s

Mid-tier

Stub / Leaf

Measurement point ("Looking Glass")
Why measurements miss so much

Tier 1’s

Mid-tier

Stub / Leaf

Measurement point ("Looking Glass")
Why measurements miss so much

Tier 1's

Mid-tier

Stub / Leaf

Measurement point
(“Looking Glass”)
Why measurements miss so much

Not exported!
In common policies, route through peer is not exported to provider
Why measurements miss so much

Tier 1’s

Mid-tier

Stub / Leaf

Measurement point (“Looking Glass”)

To see peer-peer link, both probe source & dest. must be in localized area
What’s the purpose of an IXP?

- “Metcalf’s law”: value of net is $O(n^2)$ when $n$ participants

Why don’t top-tier ISPs peer much at the IXP?
How might router-level interconnection differ from AS-level peering? Would this paper’s conclusions be the same for router-level?
Similarly ... suppose we treat the IXP as an AS “in the middle” of each member AS-to-AS connection

Now how many links are there?

- 396 total members of this IXP, so 396 links
- vs. 50,000 reported in the paper!
- $O(n^2)$ inflation factor for $n$ member ASes

This suggests interesting measurement projects:

- If you care about only the router level, what fraction of the links are observable?
- If you treat the IXP as an AS “in the middle”, what fraction of the links are observable?
What’s to come

Project proposals

• Comments back by Thursday

Next: Part Two of the course: Grand Challenges

• scalability
• complexity: SDN
• reliability
• selfishness
• security & privacy