Interdomain Routing and Connectivity

Brighten Godfrey
CS 538 February 28 2018
Routing

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, Segment Routing, …
• Interdomain: BGP.
Interdomain routing

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”)

Export policies

Export policies

Export policies

Forwarding table (FIB) in data plane

Updates to neighbors (or not)
First, BGP is a relatively simple protocol with a few salient features. It allows each router to advertise prefixes that it learns the internal network topology and compute paths from its policies. Each router combines BGP and IGP information to construct a forwarding table that maps each destination prefix to one or more outgoing links along shortest path-vector. Border (internal) routers as well as other border routers in the same AS usually run an Interior Gateway Protocol (IGP) to learn the internal network topology and compute paths from one router to another. Each router uses BGP update messages to exchange routes between neighboring Autonomous Systems (ASes). Border routers exchange prefixes that reside at the edges of an AS and border routers in neighboring ASes. These sessions are used to exchange routes between neighboring ASes. Border routers also advertise routes over eBGP sessions to produce a set of possible routes to a destination. In the absence of policy, the decision process simplifies policy expression and makes it easier to establish and maintain policy. By changing LocalPref, an operator can force a route with a longer AS path to be chosen over a shorter one. As this allows an ISP to override these suggestions, e.g., by setting the MED to a lower value, the ordering of attributes can complicate router configuration and lead to unforeseen side effects. This process is extended in many real implementations.

<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 business relationships

Provider

- I pay for traffic we exchange

Peer

- Often “settlement-free”, i.e., neither party pays

Customer

- They pay me for traffic we exchange

How might these classifications be used…

- in route selection?
- in route export?
Common policies

Route selection: prefer customer over peer over provider
Route selection: prefer customer over peer over provider

But ... What’s wrong with this picture?

Falsely assumed all routes are exported
Route export (most common): to/from customer only ("valley-free")
**Common policies: summary**

**“Gao-Rexford” policies:**

- Prefer customer > peer > provider
- Export all routes to customers
- Export customer routes to everyone
- (...and export nothing else: “valley-free”)

**Are they used in the real world?**

- “Do you always assign a higher LocalPref to a path through your customer than to a path through your peer or transit provider? (Note: exclude cases where routes through customers are tagged as backup.)” 79% yes
- Does your LocalPref configuration depend only on the next-hop AS? 56% yes
Paper discussion

How does BGP 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.

<table>
<thead>
<tr>
<th>13030</th>
<th>11537</th>
<th>40387</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>31500</td>
<td>174</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>8928</td>
<td>7132</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>40387</td>
<td>40387</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>6067</td>
<td>174</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>40387</td>
<td>40387</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>8426</td>
<td>1299</td>
<td>174</td>
<td>40387</td>
</tr>
<tr>
<td>3549</td>
<td>11537</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>19151</td>
<td>11537</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>6939</td>
<td>11537</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>3303</td>
<td>174</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>3356</td>
<td>7132</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>6667</td>
<td>3549</td>
<td>11537</td>
<td>40387</td>
</tr>
<tr>
<td>44444</td>
<td>3549</td>
<td>11537</td>
<td>40387</td>
</tr>
<tr>
<td>8426</td>
<td>1299</td>
<td>174</td>
<td>40387</td>
</tr>
<tr>
<td>19151</td>
<td>11537</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>6939</td>
<td>11537</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>5413</td>
<td>1299</td>
<td>174</td>
<td>40387</td>
</tr>
<tr>
<td>6762</td>
<td>174</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>14537</td>
<td>3356</td>
<td>7132</td>
<td>40387</td>
</tr>
<tr>
<td>3257</td>
<td>174</td>
<td>40387</td>
<td>38</td>
</tr>
<tr>
<td>13237</td>
<td>3257</td>
<td>174</td>
<td>40387</td>
</tr>
<tr>
<td>6667</td>
<td>3549</td>
<td>11537</td>
<td>40387</td>
</tr>
<tr>
<td>44444</td>
<td>3549</td>
<td>11537</td>
<td>40387</td>
</tr>
</tbody>
</table>
Interconnection: Traditional view

Tier 1’s

Mid-tier

Stub / Leaf

Hierarchical, limited peering at lower tiers
Interconnection: Modern view

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!

Past measurements missed these peerings

![Figure 2: Peering links and visibility in control/data plane (normalized by number of detected P-P links).](image-url)
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

Tier I’s

Mid-tier

Stub / Leaf

Measurement point (“Looking Glass”)

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* source & dest. of the probe 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!
- \( O(n^2) \) peering relationships among \( n \) 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

Next: Part Two of the course: Grand Challenges

• programmability: capturing intent
• reliability
• selfishness
• security & privacy

March 12: Project midterm presentations

• Be ready to present on Monday March 12
• Some groups will present on Wednesday March 14
Two key goals

- Benchmark: Demonstrate concrete progress
- Feedback & discussion with your peers

Content

- What problem are you solving?
- Why has past work not addressed the problem?
- What is your approach for solving it?
- What are your preliminary results & progress?

Logistics

- 10 minutes total: 6:40 min presentation + 4 min discuss
- PechaKucha format: 20 slides x 20 seconds, auto-advance