AS Assignment for Routers

Bradley Huffaker
bradley@caida.org

Amogh Dhamdhere, Marina Fomenkov, kc claffy

CAIDA
University of California
at San Diego, La Jolla, CA

AIMS Workshop -- February 2010
Overview

- motivation
- methodology
- analysis
- conclusions
Which AS, 32 or 12, owns/controls the router a?

<table>
<thead>
<tr>
<th>IP address</th>
<th>120.8.10.23</th>
<th>23.13.32.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>prefix</td>
<td>120.8.10.0/24</td>
<td>23.13.0.0/16</td>
</tr>
<tr>
<td>AS</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>router</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Motivation

- Dual graph
  - a combined router and AS graph
- Dual graph analysis
  - Relationship between AS degree and the AS’s number of routers.
    - how does heuristic assignment affect the inferred number of routers in an AS
- More accurate AS traceroute
  - resolving AS loops
Here is What We Want

Dual Router and AS graph
Router graph with interfaces.
Router graph with prefixes assigned to links.
Router graph with AS assigned to links.
Assigning AS to Routers

Router graph with AS assigned to routers.
We compared the success rates of four different AS assignment heuristics against our ground truth data sets.
Ground Truth

- **ISPs (i)**
  - Tier 1, Tier 2, and five research networks
- **interface sets**
  - $I_i$ interfaces in the address space of ISP$_i$, on routers that do belong to ISP$_i$
  - $\overline{I}_i$ interfaces in the address space of ISP$_i$ on routers that do not belong to ISP$_i$
- **router sets**
  - $R_i$ is the set of routers with interfaces in the address space of ISP$_i$ that do belong to ISP$_i$
  - $\overline{R}_i$ is the set of routers with interfaces in the address space of ISP$_i$ that do not belong to ISP$_i$
- **AS sets**
  - $A_i$ is the set of ASes that do belong to ISP$_i$
  - $\overline{A}_i$ is the set of ASes that do not belong to ISP$_i$
## Ground Truth

<table>
<thead>
<tr>
<th></th>
<th>$R$</th>
<th>$\bar{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>routers owned</td>
<td>routers not owned</td>
</tr>
<tr>
<td>Tier I$^{f,h}$</td>
<td>3,405</td>
<td>2,254</td>
</tr>
<tr>
<td>Tier 2$^h$</td>
<td>241</td>
<td>86</td>
</tr>
<tr>
<td>GEANT$^f$</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>I-Light$^f$</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>Internet 2$^f$</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>National LambdaRail$^f$</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>CANET$^f$</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

$^f$ Organization provided full interface list

$^h$ Organization provided naming heuristic that allowed for inference of $\bar{R}$
Methodology

Data sources

- **Router Graph (MAARS<sup>1</sup>)**
  - 268 million traceroute paths
  - 22 million nodes<sup>2</sup> / 22 million links<sup>3</sup>

- **BGP Data**
  - Oct. 2009
  - 311,230 prefixes

- **AS relationships**
  - Oct. 2009
  - BGP data
  - 148,565 AS relationship pairs

---

<sup>1</sup> router alias resolver
<sup>2</sup> node = set of IPs on same router
<sup>3</sup> link can connect > 2 nodes
## Data Topology

### Interface sets

<table>
<thead>
<tr>
<th></th>
<th>Interface sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_{12}</td>
<td>10.0.1.1, 10.0.2.3, 10.0.1.6</td>
</tr>
<tr>
<td>I_{12}</td>
<td>10.0.1.28</td>
</tr>
</tbody>
</table>

### Router sets

<table>
<thead>
<tr>
<th></th>
<th>Router sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_{12}</td>
<td>b, d, f</td>
</tr>
<tr>
<td>R_{12}</td>
<td>a</td>
</tr>
</tbody>
</table>

### AS sets

<table>
<thead>
<tr>
<th></th>
<th>AS sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_{12}</td>
<td>12</td>
</tr>
<tr>
<td>A_{12}</td>
<td>4, 2, 7</td>
</tr>
</tbody>
</table>

### Routes

<table>
<thead>
<tr>
<th>Route</th>
<th>AS</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>12</td>
<td>single-AS</td>
</tr>
<tr>
<td>b</td>
<td>4, 12</td>
<td>multi-AS</td>
</tr>
<tr>
<td>c</td>
<td>4</td>
<td>single-AS</td>
</tr>
<tr>
<td>d</td>
<td>2, 12</td>
<td>multi-AS</td>
</tr>
<tr>
<td>e</td>
<td>12</td>
<td>single-AS</td>
</tr>
<tr>
<td>f</td>
<td>12, 7</td>
<td>multi-AS</td>
</tr>
</tbody>
</table>

- **we assume** a **has a uninferred interface** which does not belong to 12
- b gets candidate AS from its interface 10.0.1.1 and the link it shares with c.
- f has no interface in I_{12} and I_{12}, so has no known ownership.
AS assignment methods

**Methodology**

<table>
<thead>
<tr>
<th>Single</th>
<th>Election</th>
<th>Neighbor</th>
<th>Customer</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>A \rightarrow A</td>
<td>only one choice</td>
<td>most single AS neighbors</td>
<td>customer AS</td>
<td>smallest degree AS</td>
</tr>
<tr>
<td>Single: only one choice</td>
<td>more links into router’s ISP’s address space</td>
<td>connected to more routers owned by the router’s ISP</td>
<td>- customer’s router uses provider’s address space for the interconnect</td>
<td>- proxy for Customer, large degree AS typically is provider of small degree AS</td>
</tr>
<tr>
<td>Election: most interfaces</td>
<td>most single AS neighbors</td>
<td>- connected to more routers owned by the router’s ISP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor: most single AS neighbors</td>
<td>- connected to more routers owned by the router’s ISP</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Methodology

• primary method
  – assignment is used if it is not ambiguous

• tie-breaker method
  – method with highest success rate on routers for which primary method yields ambiguous results

<table>
<thead>
<tr>
<th></th>
<th>ambiguous</th>
</tr>
</thead>
<tbody>
<tr>
<td>election</td>
<td>no majority AS among links</td>
</tr>
<tr>
<td>neighbor</td>
<td>no majority AS among neighbors</td>
</tr>
<tr>
<td>customer</td>
<td>no unambiguous customer relationship among ASes</td>
</tr>
<tr>
<td>degree</td>
<td>tie between smallest degree ASes</td>
</tr>
</tbody>
</table>
successful assignment:

If router $r$ is known to be owned by ISP$_i$ and method $H(r)$ selects an AS owned by ISP$_i$,

or

if $r$ is known to not be owned by an ISP$_i$ and method $H(r)$ selects an AS not owned by ISP$_i$. 
**Election + Degree** performs best with **80% success** rate.

Tier 1 bias in ground truth reduces accuracy of customer and degree heuristics.

Tie-breaker ambiguous assignments not counted.
• single AS routers
  - all methods successful for $R$ (67% of single AS routers)
  - all methods fail for $\bar{R}$ (33% of single AS routers)

  routers in $\bar{R}$ must have an interface in $A_i$, therefore single AS routers
  only have an AS in $A_i$, making it impossible for any method to select an
  AS in $\bar{A}_i$.

$X$ is ISP$i$'s address space so it maps to $A_i$. $X$ ownership is not known, so is discarded.
Success Rates

- multiple AS routers (28%)
  - **Election + Degree** best with 80% **success** rate.

- single AS routers (72%)
  - all methods successful for $R$ (67% of single AS routers)
  - all methods fail for $\bar{R}$ (33% of single AS routers)

- overall
  - **Election + Degree** best with 70% **success** rate.
Analysis of Dual Topology

Statistical correlation that we can use for topology scaling and generation
Heuristic Effect on AS Router Count

how do different heuristics affect number of inferred routers per AS

Median number of multi-AS routers

AS degree

Election
Customer
Neighbor
Degree

**Neighbor** assigns more nodes to large degree ASes

**Customer** assigns more nodes to small degree ASes
Resolving AS Loops

interface/link path

packet received on D, but response sent from A
Resolving AS Loops

interface/link path

packet received on D, but response sent from A

router path

Using inferred AS assignments resolves apparent AS loop.
Resolved AS Loops

The diagram shows the fraction of traceroute AS path loops resolved for different techniques:

- **Election** resolved 55% of paths.
- **Customer** resolved 2% of loops.
- **Neighbor** resolved the most loops with 63%.
- **Election+Degree** (the combination with the greatest success rate) resolves 62% of AS loops.
- 1~5% of paths contain AS loops, depending on the monitor.
Conclusion

- multiple AS routers
  - Election + Degree best with 80% success rate.
- all routers
  - Election + Degree best with 70% success rate.
- AS loop resolution
  - Election + Degree resolves 62% or AS loops
• More ground truth

• alternative AS assignment heuristics

Bradley Huffaker bradley@caida.org