What we know and what we don’t about the Internet

Dmitri Krioukov
CAIDA
dima@caida.org
X. Dimitropoulos, P. Mahadevan,
M. Fomenkov, B. Huffaker,
Y. Hyun, and kc claffy.

Aveiro, May 15th, 2008
What the Internet does

The Internet was designed for and exists to transfer information packets from $A$ to $B$, where $A$ and $B$ are any two Internet-Protocol- (IP-)talking devices.
### IP packet format

<table>
<thead>
<tr>
<th>+</th>
<th>Bits 0–3</th>
<th>4–7</th>
<th>8–15</th>
<th>16–18</th>
<th>19–31</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Version</td>
<td>Header length</td>
<td>Type of Service (now DiffServ and ECN)</td>
<td></td>
<td>Total Length</td>
</tr>
<tr>
<td>32</td>
<td>Identification</td>
<td>Flags</td>
<td>Fragment Offset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>Time to Live</td>
<td>Protocol</td>
<td></td>
<td>Header Checksum</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td></td>
<td></td>
<td>Source Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>128</td>
<td></td>
<td></td>
<td>Destination Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td></td>
<td></td>
<td>Options</td>
<td></td>
<td></td>
</tr>
<tr>
<td>160 or 192+</td>
<td></td>
<td></td>
<td>Data</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IP addresses

\[ A = 193.137.168.155 \]
\[ B = 192.172.226.78 \]
### IP routes

<table>
<thead>
<tr>
<th>Hop</th>
<th>Time (ms)</th>
<th>Time (ms)</th>
<th>Time (ms)</th>
<th>Destination IP Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;1</td>
<td>2</td>
<td>2</td>
<td>193.137.81.254</td>
</tr>
<tr>
<td>2</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>192.168.255.253</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>193.137.173.254</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>193.136.4.26</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>193.136.1.221</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>193.137.0.30</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>62.40.124.185</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>33</td>
<td>32</td>
<td>62.40.112.146</td>
</tr>
<tr>
<td>9</td>
<td>41</td>
<td>40</td>
<td>40</td>
<td>62.40.112.137</td>
</tr>
<tr>
<td>10</td>
<td>123</td>
<td>124</td>
<td>124</td>
<td>62.40.112.134</td>
</tr>
<tr>
<td>11</td>
<td>130</td>
<td>130</td>
<td>129</td>
<td>216.24.184.85</td>
</tr>
<tr>
<td>12</td>
<td>134</td>
<td>131</td>
<td>130</td>
<td>216.24.186.23</td>
</tr>
<tr>
<td>13</td>
<td>143</td>
<td>144</td>
<td>143</td>
<td>216.24.186.20</td>
</tr>
<tr>
<td>14</td>
<td>167</td>
<td>167</td>
<td>167</td>
<td>216.24.186.8</td>
</tr>
<tr>
<td>15</td>
<td>199</td>
<td>199</td>
<td>198</td>
<td>216.24.186.30</td>
</tr>
<tr>
<td>16</td>
<td>197</td>
<td>197</td>
<td>197</td>
<td>137.164.26.130</td>
</tr>
<tr>
<td>17</td>
<td>203</td>
<td>203</td>
<td>203</td>
<td>137.164.25.5</td>
</tr>
<tr>
<td>18</td>
<td>204</td>
<td>203</td>
<td>203</td>
<td>137.164.27.50</td>
</tr>
<tr>
<td>19</td>
<td>204</td>
<td>205</td>
<td>204</td>
<td>198.17.46.56</td>
</tr>
<tr>
<td>20</td>
<td>203</td>
<td>204</td>
<td>204</td>
<td>192.172.226.78</td>
</tr>
</tbody>
</table>
IP routes
Broadcast media (e.g., ethernet)

Reality

Perception
Autonomous Systems
AS topology

Diagram: A network topology with nodes X, Y, and Z interconnected.
IP routing

- **Intradomain (Interior Gateway Protocols (IGPs))**
  - routing within an Autonomous System (AS)
  - protocols:
    - Open Shortest Path First (OSPF)
    - Intermediate System to Intermediate System (ISIS)
  - Links State (LS) routing protocols

- **Interdomain (Exterior Gateway Protocols (EGPs))**
  - routing between Autonomous Systems (ASs)
  - protocols:
    - Border Gateway Protocol (BGP)
    - Path Vector (PV) routing protocol
Each AS advertises IP addresses that it has
- AS 1930 (U. Aveiro) advertises:
  193.137.168.0 - 193.137.175.255 (193.136.0.0/15)

All neighboring ASs receiving such advertisement re-advertise them to their neighbors after pre-pending their AS numbers

The result is that each AS has a routing entry for (193.136.0.0/15) which looks like:
193.136.0.0/15: AS \(X_1\), AS \(X_2\), ..., AS 1930
The two main sources of the Internet topology data

- **Traceroute data**
  - gives a glimpse of the router topology
    - too many vagaries in IP-to-router resolution
  - gives a view of the AS topology
    - many vagaries in IP-to-AS resolution

- **BGP data**
  - gives another view of the AS topology
    - but there are still some missing links due to sampling biases
Router vs. AS topology

- We do not know the router topology
- We know the AS topology much better
AS relationships and BGP policies

- Each AS link is the relationship (i.e., business, contractual agreement) between the two ASs.
- There are roughly three types of such relationships:
  - customer-provider (c2p)
  - peer-peer (p2p)
  - sibling-sibling (s2s)
- They stem from combinations of the following two BGP route re-advertisement policies:
  - re-advertising to provider or peer, an AS advertises only its own IP addresses and IP routes learnt from its customers.
  - re-advertising to customer or sibling, an AS advertises everything.
- BGP advertisement policy combinations vs. AS relationships:
  - asymmetric combination: c2p
  - symmetric combinations: p2p and s2s
Valid paths

- **uphill**: zero or more links from customer to provider
- **pass**: zero or one link from peer to peer
- **downhill**: zero or more links from provider to customer
- **any number of sibling links** anywhere in the path
Type of Relationship (ToR) problem formulations

- Given a set of BGP paths $P$,
- Extract the undirected AS-level graph $G$.
  - Every edge in $G$ is a link between pair of ASs.
- Assuming edge direction is from customer to provider,
- Direct all edges in $G$ ($2^m$ combinations),
- Inducing direction of edges in $P$,
- Such that the number of invalid paths in $P$ is minimized.
  - Invalid path is a path containing a provider-to-customer link followed by customer-to-provider link
ToR and MAX2SAT

- Split all paths in $P$ into pairs of adjacent links (involving triplets of nodes)
- Perform mapping…
## Mapping to MAX2SAT

<table>
<thead>
<tr>
<th>Edges in $P$</th>
<th>2SAT clause</th>
<th>Edges in $G_{2SAT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i \rightarrow j$</td>
<td>$x_i \lor x_j$</td>
<td>$x_i \rightarrow x_j$</td>
</tr>
<tr>
<td>$i \rightarrow j$</td>
<td>$x_i \lor \overline{x}_j$</td>
<td>$x_i \rightarrow \overline{x}_j$</td>
</tr>
<tr>
<td>$i \rightarrow j$</td>
<td>$\overline{x}_i \lor x_j$</td>
<td>$\overline{x}_i \rightarrow x_j$</td>
</tr>
<tr>
<td>$i \rightarrow j$</td>
<td>$\overline{x}_i \lor \overline{x}_j$</td>
<td>$\overline{x}_i \rightarrow \overline{x}_j$</td>
</tr>
</tbody>
</table>
SDP relaxation to MAX2SAT

\[
\max \quad \frac{1}{4} \sum_{k,l=1}^{2m_1} w_{kl} (3 + v_0 \cdot v_k + v_0 \cdot v_l - v_k \cdot v_l)
\]

s.t. \quad v_0 \cdot v_0 = v_k \cdot v_k = 1, \quad v_i \cdot v_{m_1+i} = -1,
\quad k = 1 \ldots 2m_1, \quad i = 1 \ldots m_1.
Physical interpretation
Infer c2p links using multiobjective optimization

- **Maximize number of invalid paths:**
  - 2-link clauses $w_{kl}(x_k \lor x_l)$

- **Direct along the node degree gradient:**
  - 1-link clauses $w_{kk}(x_k \lor x_k)$
Final form of the generalized problem formulation

\[
\max \quad \frac{1}{4} \sum_{k,l=1}^{2m_1} w_{kl}(3 + v_0 \cdot v_k + v_0 \cdot v_l - v_k \cdot v_l)
\]

s.t.
\[
v_0 \cdot v_0 = v_k \cdot v_k = 1, \quad v_i \cdot v_{m_1+i} = -1, \\
k = 1 \ldots 2m_1, \quad i = 1 \ldots m_1.
\]

\[
w_{kl}(\alpha) = \begin{cases} 
    c_2 \alpha & \text{if } \{kl\} \in P, \\
    c_1(1 - \alpha)f(d_k^-, d_k^+) & \text{if } k = l \leq m_1, \\
    0 & \text{otherwise}.
\end{cases}
\]

\[
f(d_i^-, d_i^+) = \frac{d_i^+ - d_i^-}{d_i^+ + d_i^-} \log(d_i^+ + d_i^-).
\]
AS relationship results

- **Input:** RouteViews, 8-hour interval snapshots between 03/01/05 and 03/05/05
- **Output:**

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>c2p links</th>
<th>p2p links</th>
<th>s2s links</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of links</td>
<td>38,282</td>
<td>34,552</td>
<td>3,553</td>
<td>177</td>
</tr>
<tr>
<td>percentage</td>
<td>100%</td>
<td>90.26%</td>
<td>9.28%</td>
<td>0.46%</td>
</tr>
</tbody>
</table>
### AS hierarchy

<table>
<thead>
<tr>
<th>$\alpha = 0.00$</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.10$</th>
<th>$\alpha = 0.50$</th>
<th>$\alpha = 1.00$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Percentage of invalid paths</strong></td>
<td>12.75%</td>
<td>1.79%</td>
<td>0.69%</td>
<td>0.46%</td>
<td>0.36%</td>
</tr>
<tr>
<td><strong>Top of reachability based hierarchy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AS #</th>
<th>name</th>
<th>degree</th>
<th>dep. wid.</th>
<th>dep. wid.</th>
<th>dep. wid.</th>
<th>dep. wid.</th>
<th>dep. wid.</th>
<th>dep. wid.</th>
<th>dep. wid.</th>
</tr>
</thead>
<tbody>
<tr>
<td>701</td>
<td>UUNET</td>
<td>2334</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>105</td>
<td>0</td>
<td>120</td>
</tr>
<tr>
<td>7018</td>
<td>AT&amp;T</td>
<td>1911</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>1239</td>
<td>Sprint</td>
<td>1703</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>3356</td>
<td>Level 3</td>
<td>1228</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>209</td>
<td>Qwest</td>
<td>1105</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>14551</td>
<td>UUNET</td>
<td>35</td>
<td>128</td>
<td>1</td>
<td>137</td>
<td>2</td>
<td>138</td>
<td>1</td>
<td>151</td>
</tr>
<tr>
<td>13987</td>
<td>IBASIS Inc.</td>
<td>3</td>
<td>1792</td>
<td>955</td>
<td>1802</td>
<td>963</td>
<td>1830</td>
<td>976</td>
<td>1847</td>
</tr>
<tr>
<td>8631</td>
<td>Routing Arbiter</td>
<td>48</td>
<td>108</td>
<td>1</td>
<td>123</td>
<td>1</td>
<td>122</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>23649</td>
<td>Hong Kong Teleport</td>
<td>4</td>
<td>1792</td>
<td>955</td>
<td>1802</td>
<td>963</td>
<td>899</td>
<td>121</td>
<td>916</td>
</tr>
<tr>
<td>4474</td>
<td>Village Communications</td>
<td>2</td>
<td>2747</td>
<td>16136</td>
<td>2765</td>
<td>16118</td>
<td>2806</td>
<td>16077</td>
<td>2818</td>
</tr>
</tbody>
</table>
Phase transition in mean field approximation
Validation

Previous validation efforts
- Gao: AT&T
- SARK: Gao
- Subsequent: SARK/Gao

Our validation
- 38 ASs (5 Tier-1 ISPs, 13 smaller ISPs, 19 universities, and 1 content provider)
- 3,724 links (9.7% of the total)
- 94.2% overall accuracy

<table>
<thead>
<tr>
<th></th>
<th>links</th>
<th>inferred c2p links</th>
<th>inferred p2p links</th>
<th>inferred s2s links</th>
</tr>
</thead>
<tbody>
<tr>
<td>total number of</td>
<td>3,724</td>
<td>3,070</td>
<td>623</td>
<td>31</td>
</tr>
<tr>
<td>number of correct</td>
<td>3,508</td>
<td>2,964</td>
<td>516</td>
<td>28</td>
</tr>
<tr>
<td>percentage of correct</td>
<td>94.2%</td>
<td>96.5%</td>
<td>82.8%</td>
<td>90.3%</td>
</tr>
</tbody>
</table>
Questions in the questionnaire

- For the listed inferred AS relationships, specify how many are incorrect, and what are the correct types of the relationships that we mis-inferred?
- What fraction of the total number of your AS neighbors is included in our list?
- Can you describe any AS relationships, more complex than c2p, p2p, or s2s, that are used in your networks?
Missing links

- 27 (3 tier-1 ISPs) out of 38 answered the second question, too, and provided us with their full AS relationship data: 1,114 links
- Among these, we see only 552 (49.6%):
  - 38.7% out of the 865 (77.6%) p2p links
  - 86.7% out of the 218 (19.6%) c2p links
  - 93.3% out of the 30 (2.7%) s2s links
- Maximum percentage of missing links per node is 86.2% (50% of ASs miss >70% links)
Missing links visualized

Legend:
- Total
- p2p
- c2p
- s2s

Histogram showing True AS adjacencies versus observed AS adjacencies.

Graph showing Observed AS adjacencies against True AS adjacencies.
More complex policies

- Space
- Time
- Prefix
AS taxonomy

- Assign the following six attributes to every AS
  - organization description (IRR data, stop words are filtered out and the rest of words are stemmed)
  - number of customers
  - number of providers
  - number of peers
  - number of advertised IP prefixed
  - size of the advertised IP address space

- Feed this data into a machine learning algorithm (AdaBoost) with a training set of 1200 ASs

- Classify all ASs into the following six categories
  - Large ISPs
  - Small ISPs
  - Customer ASs
  - Universities
  - IXPs
  - NICs
## AS taxonomy results

Classified 95.3% of ASs (non-abstained) with expected accuracy of 78.1%

<table>
<thead>
<tr>
<th></th>
<th>Large ISPs</th>
<th>Small ISPs</th>
<th>Customer ASes</th>
<th>Universities</th>
<th>IXPs</th>
<th>NICs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASes</td>
<td>44</td>
<td>5,599</td>
<td>11,729</td>
<td>877</td>
<td>33</td>
<td>332</td>
</tr>
<tr>
<td>%</td>
<td>0.2</td>
<td>30.1</td>
<td>63.0</td>
<td>4.7</td>
<td>0.2</td>
<td>1.8</td>
</tr>
</tbody>
</table>
AS rank
That’s not all we now about the Internet but it’s pretty much all we know about the Internet AS topology 😊

Thank you!