Geolocation is the identification of the real-world geographic location of Internet ids.

- sinet-1-lo-jmb-702.lsanca.pacificwave.net (207.231.240.135)
- hpr-lax-hpr--sdsc-10ge.cenic.net (137.164.26.33)
- dolphin.sdsc.edu (132.249.31.17)
- piranha.sdsc.edu (198.17.46.8)
- pinot-g1-0-0 (192.172.226.1)
• timeline
• NANOG feedback
• data
• proposed process
## Timeline

<table>
<thead>
<tr>
<th>Month</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan.</td>
<td>public request for feedback to NANOG (12 responses)</td>
</tr>
<tr>
<td>Feb.</td>
<td>discuss feedback at CAIDA's AIMS 2010 workshop</td>
</tr>
<tr>
<td>Mar.</td>
<td>development</td>
</tr>
<tr>
<td>May</td>
<td>run comparison</td>
</tr>
<tr>
<td>Dec.</td>
<td>publish report</td>
</tr>
</tbody>
</table>
Most mentioned Services

• Major Services
  – MaxMind (GeolP, GeoLite)
  – Akamai (EdgePlatform)
  – Google (Google Gears)
  – Digital Envoy (Netacuity)

• Smaller Services
  – Quova (Quova On Demand)
  – IP2Location (IP2Location)
• content localization
• credit card verification
• taxation purposes
• legal terms of service applicability
• ad targeting
• data privacy requirements
• DRM restricted content
• nearest datacenter
Whois data

- whois allocations
  - for each geolocation service, break each block down into largest continuous block of addresses that have the same location
  - whois provides finer breakdown than BGP prefixes

- whois country level “ground truth”ish
  - not really accurate for large organizations
  - registries only provide country as separate field
Whois breakdown data

AFRINIC, APNIC, ARIN, LACNIC, RIPE database dumps
Jan. 29, 2010
Topography Data

- BGP (Routeviews/RIPE)
  - IPv4 Prefix to AS mapping

- ark router graph
  - IPv4 prefixes / hostnames

- CAIDA’s AS relationships
  - classify AS’s into categories
    - large transit provider
    - small transit provider
    - enterprise
    - content provider
<table>
<thead>
<tr>
<th>Organization</th>
<th>Number of IPv4 Addresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier 1</td>
<td>26k</td>
</tr>
<tr>
<td>Tier 2</td>
<td>519</td>
</tr>
<tr>
<td>GEANT</td>
<td>299</td>
</tr>
<tr>
<td>I-Light</td>
<td>265</td>
</tr>
<tr>
<td>Internet 2</td>
<td>317</td>
</tr>
<tr>
<td>National LambdaRail</td>
<td>183</td>
</tr>
<tr>
<td>CANET</td>
<td>96</td>
</tr>
</tbody>
</table>

If you have ground data, please let me know!  
bradley@caida.org
proposed process

steps

1. Run each service against the whois allocations, subdivide until all IP addresses within the suballocation map to the same region.

2. Annotate blocks by their AS class

   \[\text{IP} \rightarrow \text{prefix} \rightarrow \text{AS} \rightarrow \text{class}\]

analysis

- Ground truth
  + Compare against ISP locations
  + Compare against whois country

- AS class
  + How does AS class affect accuracy
  + How does AS class affect the frequency of block subdivision

- Hostnames vs location
  + Check if hostname changes affects geolocation