Autonomous Systems (AS) Introduction and Visualization

Bradley Huffaker
CAIDA
SDSC/UCSD

Data Visualization (CSE 199)
22 January 2016
• introduction
• visualizations
  – AS Core
  – AS Core single AS
  – AS Core single Organization
  – AS Relationships
• data files (URL)
• summary
Internet maps can be grouped into three levels.
1. **IP addresses** that connect devices on to the Internet.
2. **Routers**, machines that route the traffic, interconnect via **IP addresses**.
3. **Autonomous Systems (AS)** are numbers used to route groups of **IP addresses**.
We will focus on the AS layer.
An AS can roughly be thought of as a single organization.

Some companies use multiple ASes.

AS topology (observed January 2016)
- 51,000+ ASes
  - geographic location
  - country of registration
- 418,000+ AS links
  - business relationship
AS Core focuses on visualizing the geographic scope and size of the ASes as a whole.
AS Core data

- AS paths
  - 20150101.paths.bz
- AS longitude
  - 20150101.as2loc.txt.bz
Routes are announced by routers and forwarded toward the collector. So the last AS, the “origin” AS, “owns” (first announces) the prefix.

<table>
<thead>
<tr>
<th>prefix</th>
<th>AS path</th>
<th>origin AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>route 1</td>
<td>5.5.1.0/24</td>
<td>1</td>
</tr>
<tr>
<td>route 2</td>
<td>13.5.1.0/24</td>
<td>1</td>
</tr>
<tr>
<td>route 3</td>
<td>10.0.0.0/16</td>
<td>1</td>
</tr>
<tr>
<td>route 5</td>
<td>9.0.1.0/24</td>
<td>1</td>
</tr>
</tbody>
</table>
Collecting and sharing global routing [Border Gateway Protocol (BGP)] data:

- University of Oregon, Route Views Project
  - [http://www.routeviews.org](http://www.routeviews.org)

- RIPE NCC (Regional Internet Registry for Europe/Middle East)

BGP dump

| TABLE_DUMP2 | 127 1 649600 | B | 157.130.10.233 | 701 | 4.21.103.0/24 | 70 1 3549 46 |
| TABLE_DUMP2 | 127 1 649600 | B | 203.62.252.186 | 122 | 4.21.103.0/24 | 122 1 4637 3549 46 |
| TABLE_DUMP2 | 127 1 649600 | B | 12.0.1.63 | 7018 | 4.21.103.0/24 | 7018 3549 46133 |
The top of the file lists the sources used to generate the file.

```
# source: topology | BGP | 20150104 | routeviews | wide
# source: topology | BGP | 20150105 | routeviews | wide
513 | 3320 | 1299 | 24961 | 13301
28917 | 3356 | 22773 | 16653
28571 | 1251 | 20080 | 6939 | 4766 | 38420
25220 | 1299 | 2914 | 4648 | 4610 | 17746
25091 | 2914 | 286 | 8529
202109 | 1299 | 12778 | 49725
9002 | 9304 | 17408 | 131149
14840 | 6453 | 4755 | 45820 | 22853
```

The actual AS path. `<source>|<hop0>|<hop1>|….|<destination>`

The **source AS** is where the path was observed.

Your view of the topology is limited by the number of sources and origin ASes. Depending on what you are doing with the data, it often helps to filter paths to only sources that have all or almost all origin AS.
Based on the set of AS paths, different types of degrees may be used.

- **global degree:** \{A, C, E, D\}
  - all ASes that connect to the given AS
- **out degree:** \{C, D\}
  - ASes that followed after the given AS in at least one path
- **transit degree:** \{E, D\}
  - ASes that followed after the given AS in at least one path

paths:  
A → B  
B → C  
E → B → D
Based on the set of monitors, you will observe different types of paths and so a different set of degrees.

### AS degree types

**AS** | **global** | **out** | **transit**
--- | --- | --- | ---
1 | 3 | 3 | 0
2 | 1 | 0 | 0
3 | 2 | 0 | 0
4 | 1 | 0 | 0
5 | 2 | 0 | 0
6 | 3 | 2 | 3
7 | 3 | 2 | 2

**Monitors**

- **M1**
  - 1-2
  - 1-5
  - 1-6-3
  - 1-6-7-4

- **M7**
  - 7-3
  - 7-4

**Undirected paths**

- M1: 1-2, 1-5, 1-6-3, 1-6-7-4
- M7: 7-3, 7-4

**Directed paths**

- M1: 1→2, 1→5, 1→6-3, 1→6-7-4
- M7: 7→3, 7→4

**Triplets**

- M1: 1-6-3, 1-6-7-4
- M7: 6-7-4

**Degrees**

- Global degree
- Out degree
- Transit degree
AS paths

paths

Graph

<table>
<thead>
<tr>
<th>AS</th>
<th>name</th>
<th>longitude</th>
<th>neighbors</th>
<th>transit degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>5, 43</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td></td>
<td></td>
<td>1, 12</td>
<td>2</td>
</tr>
</tbody>
</table>
AS geolocation

- **Geolocation**
  identification of real-world geographic location of Internet identifiers

- **Digital Envoy’s Netacuity commercial geolocation server**

- **MaxMind GeoLite is a free service**
  [http://www.maxmind.com/app/geoip_country](http://www.maxmind.com/app/geoip_country)

**Netacuity geographic dump**

<table>
<thead>
<tr>
<th>IP first</th>
<th>IP last</th>
<th>country</th>
<th>state</th>
<th>city</th>
<th>latitude</th>
<th>longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.172.226.0</td>
<td>192.127.226.255</td>
<td>usa</td>
<td>ca</td>
<td>la jolla</td>
<td>32.855</td>
<td>-117.249</td>
</tr>
<tr>
<td>137.164.23.0</td>
<td>137.164.23.255</td>
<td>usa</td>
<td>ca</td>
<td>tustin</td>
<td>33.736</td>
<td>-117.823</td>
</tr>
<tr>
<td>137.164.46.0</td>
<td>137.164.46.255</td>
<td>usa</td>
<td>ca</td>
<td>los angeles</td>
<td>33.973</td>
<td>-118.248</td>
</tr>
<tr>
<td>74.125.49.0</td>
<td>74.125.49.255</td>
<td>usa</td>
<td>il</td>
<td>chicago</td>
<td>41.886</td>
<td>-87.623</td>
</tr>
</tbody>
</table>
An AS’s longitude equals the weighted average of the longitudes assigned by Netacuity to the addresses it announces.

\[ \text{weighted average longitude} = \frac{\sum_{i} \text{block}_i.\text{longitude} \times \text{block}_i.\text{size}}{\sum_{i} \text{block}_i.\text{size}} \]

<table>
<thead>
<tr>
<th>origin AS</th>
<th>prefix</th>
<th>IP block</th>
<th>longitude</th>
<th>weighted average longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>5.5.1.0/24</td>
<td>5.5.1.0 - 5.5.1.255</td>
<td>-103</td>
<td>-103</td>
</tr>
<tr>
<td>1</td>
<td>10.0.0.0/16</td>
<td>10.0.0.0-10.0.127.255</td>
<td>25</td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.0.128.0-10.0.255.255</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>13.5.1.0/24</td>
<td>13.5.1.0-13.5.1.255</td>
<td>-23</td>
<td>-23</td>
</tr>
<tr>
<td>5</td>
<td>9.0.1.0/24</td>
<td>9.0.1.0-9.0.1.255</td>
<td>45</td>
<td>45</td>
</tr>
</tbody>
</table>
The document discusses an AS centroid file, which contains data on the location and number of prefixes announced by Autonomous Systems (AS). The table below shows the format and data content:

<table>
<thead>
<tr>
<th>AS number</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Total Prefixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.023026628203</td>
<td>-88.3649892886005</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>39.9882562979862</td>
<td>-75.4179414754833</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>42.3643803602359</td>
<td>-71.1006790537166</td>
<td>33</td>
</tr>
</tbody>
</table>

The file name is `20150101.as2loc.txt`. The AS number is present, while the continent, country, state, and city fields are empty in this file.
This file contains both organization and ASes. The type changes at every line that starts with “# format:<type>”. Connect between entries with org-id. In this example the AS 46191 belongs to organization Globalinx.

http://www.caida.org/data/as-organizations/
We now have everything we need to build the graph.

<table>
<thead>
<tr>
<th>AS</th>
<th>name</th>
<th>longitude</th>
<th>neighbors</th>
<th>transit degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Level 3</td>
<td>37.5</td>
<td>5, 43</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Symbolics, Inc.</td>
<td>45</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>New York University</td>
<td>-103</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td>Brookhaven Laboratory</td>
<td>-23</td>
<td>1, 12</td>
<td>2</td>
</tr>
</tbody>
</table>
Each node represents a single AS.

ASes with nearby/same degree and longitude will overlap.

<table>
<thead>
<tr>
<th>node:</th>
<th>[1 - \log\left(\frac{\text{degree (AS)} + 1}{\text{maximum degree} + 1}\right)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>node: size</td>
<td>[rac{\text{degree (AS)} + 1}{\text{maximum degree} + 1}]</td>
</tr>
<tr>
<td>node: angle</td>
<td>longitude of the AS’s BGP prefixes</td>
</tr>
<tr>
<td>link: color</td>
<td>node color with smallest degree</td>
</tr>
</tbody>
</table>
ASes are colored based on their transit degree. Since transit degree is a long tailed distribution, it is best to use a logarithmic color scheme.
Link color set by largest degree.

Link color set by average degree.

Link color set by average smallest.
nodes draw over links

unssorted link draw order
draw largest link last
geographic regions as core

- Africa
- Asia
- Europe
- North American
- South American
- Oceania

geo.png
- degree encoded twice (size/radius)
  - could encode additional metrics

- latitude node encode
  - overlap between continents
IPv4 vs IPv6

IPv4 high density in Asia, America, and Europe

IPv6 highest area of density in Europe
IPv4 vs IPv6 cores

IPv4 core primarily in North America

IPv6 core spread between America and Europe
Showing the same placement, but focused on a single AS.
- AS customer cone
  - 20150101.ppdc-ases.bz
- AS longitude
  - 20150101.as2loc.txt.bz
- AS relationships
  - 20150101.as-rel.txt.bz
AS Relationships

provider
• you pay them to transit your traffic
  (XO pays AT&T)

peer
• unpaid exchange of traffic
  (between XO and Google)

customer
• they pay you to transit their traffic
  (UCSD pays XO)
ASes can be organized into a hierarchical structure based on the type of business relationships they form between themselves.

- Customer pays provider for transit.
- Peers do not pay to accept each other's traffic.
ASes at the top of the hierarchy are call Tier-1. Together they form the backbone of the Internet.
AS Relationship (file)

Internet eXchange Points

inferred Tier 1 ASes

# inferred clique: 174 209 701 1239 1299 2828 2914 3257 3320 3356 3549 5511
# IXP ASes: 1200 4635 5507 6695 7606 8714 9355 9439 9560 9722 9989 11670 17819
1|11537|0

AS0 and AS1 are peers  AS0 — AS1

AS0 is a provider of AS1  AS0 ← AS1

AS0 is a customer of AS1  AS0 → AS1

http://www.caida.org/data/as-relationships/
An AS’s customer cone contains the set of ASes we observe the AS announce to its peers or providers. In practice, this is the set of ASes it can reach through its customers.

<table>
<thead>
<tr>
<th>AS</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
</tr>
</tbody>
</table>
### AS Customer Cone (file)

AS Customer Cone members:

<table>
<thead>
<tr>
<th>AS</th>
<th>size</th>
<th>members</th>
</tr>
</thead>
<tbody>
<tr>
<td>2345</td>
<td>1</td>
<td>2345</td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td>14  209  701  1239  1299  2828  2914  3257  3320  3356  3549  5511</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>435  1200  5507  6695  7606  8714  9355  9439  9560  9722</td>
</tr>
</tbody>
</table>

Download file: [20150101.ppdc-ases.txt.bz2](http://www.caida.org/data/as-relationships/20150101.ppdc-ases.txt.bz2)
- Node location and size set by customer cone size
- Node and link color set by relationship to selected AS
- Selected AS size is fixed regardless of degree.
- Draw order: customers, peers, and providers
We can build a list of an organization’s ASes by filtering on its org-id.
An organization’s ASes and their neighbors.

- **black** links connect “sibling” ASes
- Conflicting relationships are resolved: provider, peer, customer.
Information about a AS and its neighbors. Shown without the geographic encoding.
• AS customer cone
  – 20150101.ppdc-ases.bz
• AS country
  – 20150101.as-org2info.txt.bz
• AS relationships
  – 20150101.as-rel.txt.bz
We define an AS’s ranking to be equal to the number of ASes with a customer cone size greater than the AS plus 1. (+1, allows rank to start at 1)

<table>
<thead>
<tr>
<th>rank</th>
<th>AS</th>
<th>customer cone size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>336</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>876</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>91</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>4</td>
</tr>
</tbody>
</table>
| node label       | center: AS number  
|                 | upper left: rank  
|                 | upper right: country flag |
| node vertical position | customer cone size |
| node size | customer cone size |
| link width | customer cone size of neighbor |
| link color | relationship type |

![AS placement diagram](image)
- **AS node placement**
  - overlapping **provider / customers** move along the X axis
  - overlapping **peers** move along the Y axis (distorts cone size mapping)
- **large links drawn over small links**
- **AS always drawn over links**
  - visually disconnects AS and its link
  - but: nodes can be seen and read
Node placement means a node is visible, even if its link is not.

More important, i.e. larger cone sizes, ASes have larger and more readable information.
• redundant encoding of customer cone sizes
• smaller ASes are too small to read
AS are often divided into Transit, Content, or Access depending on the type of companies that operate them. Larger organizations can combine all three types.

- **Transit ASes**, such as AT&T, form the backbone of the Internet by carrying traffic between ASes.
- **Content ASes**, such as Netflix, provide the videos, webpages, and other content.
- **Access ASes**, such as UCSD or Time Warner Cable, are used by smaller organizations to get access to the Internet.
Cases are often divided into Transit, Content, or Access depending on the type of companies that operate them. Larger organizations can combine all three types.

- **Transit ASes**, such as AT&T, form the backbone of the Internet carrying traffic between ASes.
- **Content ASes**, such as Netflix, provide the videos, webpages, and other content.
- **Access ASes**, such as UCSD or Time Warner Cable, are used by smaller organizations to get access to the Internet.
- **important objects drawn last** (i.e. over less important objects)
  - rare objects over common (typically more links then nodes)

- **highlight a few features in a complex graph**

- **not all objects must be readable / distinguishable**
  - it is often enough to be able to get an idea of the frequency
<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
<th>Slide</th>
</tr>
</thead>
<tbody>
<tr>
<td>20150101.paths.txt.bz</td>
<td>all observed AS paths</td>
<td>11</td>
</tr>
<tr>
<td>20150101.as2loc.bz</td>
<td>centroid of AS’s address space</td>
<td>17</td>
</tr>
<tr>
<td>20150101.as-org2info.txt.bz</td>
<td>provides country and organization information</td>
<td>18</td>
</tr>
<tr>
<td>20150101.as-rel.txt.bz</td>
<td>types of business relationships between pairs of ASes</td>
<td>33</td>
</tr>
<tr>
<td>20150101.ppdc-ases.bz</td>
<td>the set of ASes that are customers, or customer’s customers of an AS</td>
<td>36</td>
</tr>
<tr>
<td>20150101.as2type.txt.bz</td>
<td>an AS’s type of business</td>
<td>45</td>
</tr>
</tbody>
</table>
Questions?

Internships:
http://www.caida.org/home/jobs/
Distance from client's centroid and DNS server's geographic location.

Number of DNS clients in a given direction and average distance in that direction.
Geographic sorted bars

- F: San Francisco (US)
- F: New York (US)
- F: Santiago (CL)
- F: Sao Paulo (BR)
- F: Johannesburg (SA)
- F: Tel Aviv (IL)
- F: Brisbane (AU)
- F: Auckland (NZ)

Instances sorted by the number of requests per second.

- F: San Francisco (US)
- F: New York (US)
- F: Sao Paulo (BR)
- F: Santiago (CL)
- F: Johannesburg (SA)
- F: Tel Aviv (IL)
- F: Brisbane (AU)
- F: Auckland (NZ)

Legend:
- Oceania
- Asia
- Africa
- Europe
- S. Amer
- N. Amer
used address space
Code red