Looking for Hypergiants in PeeringDB

Timm Böttger, Felix Cuadrado and Steve Uhlig

timm.boettger@qmul.ac.uk
The term hypergiant

• Term ‘hypergiant’ coined by Labovitz et al.:
  • Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide and Farnam Jahanian. “Internet inter-domain traffic”. SIGCOMM 2010

• Commercial vantage points distributed in 110 networks over 2 years

• Observed shift of traffic
  • Diverted away from Tier-1 and Tier-2 providers
  • Instead directly exchanged between networks

• Leads to ‘Internet flattening’
Hypergiants?

Google

facebook

CLOUDFLARE

NETFLIX

Akamai

twitch

Twitter
The problem with hypergiants

• No definition yet
• No methodology for identification yet
• No complete listing yet
• Resort to listings of examples
• Resort to self-reported evidence from those hypergiants
  • e.g., Google B4, Google Espresso, Facebook Edge Fabric
Main contribution

An *objective* approach to solve the problem of hypergiant classification

Outline:

- Characterisation of organisations as seen through PeeringDB
- Classification of hypergiants based on PeeringDB data
PeeringDB

• Online database containing information on:
  • Internet eXchange points
  • Autonomous Systems

• Community driven, but highly trusted
  • Referred to as authoritative by Google and Netflix
  • Cloudflare provisions filtering rules based on PeeringDB data

• Our snapshot contains 643 IXPs and 6,910 organisations
PeeringDB - Webinterface

Looking for Hypergiants in PeeringDB - Timm Böttger
IXP sizes per region

- Largest IXP ecosystem in Europe
- Followed by Americas and Asia Pacific
- About 50% of capacity from non-local organisations
- Europe dominating number of members
- Only small number of non-local organisations

[Graphs showing port capacity and organisation count for different regions: Europe, North America, South America, Asia Pacific, Australia, Africa, Middle East.]

- Europe: 150 Tbps, 3000 organisations
- North America: 100 Tbps, 1000 organisations
- South America: 50 Tbps, 500 organisations
- Asia Pacific: 10 Tbps, 300 organisations
- Australia: 5 Tbps, 100 organisations
- Africa: 1 Tbps, 20 organisations
- Middle East: 1 Tbps, 10 organisations

About 50% of capacity from non-local organisations.
Peering Organisations

Organisations at one continent: 45% of port capacity, yet 92% of all organisations

Organisations at four continents or more: 38% of port capacity, yet only 1% of all organisations
Peering Organisations

Some of the outbound organisations at one continent: BBC, Hetzner, Strato, VKontakte, Baidu

Organisations at one continent: mostly inbound or balanced traffic profile (>75% port capacity)

Organisations at four continents or more: mostly outbound oriented (Apple, Twitch, Facebook, Google)

Some of the balanced organisations at 4+ continents: Dropbox, AWS, Hurricane Electric, Microsoft
Interim conclusion

• Content providers rely on IXPs to deliver content
• Served to eyeball organisations:
  • Smaller network footprint
  • More local network footprint
  • More inbound oriented footprint

• Relatively small group of organisations gathers significant fraction of port capacity
• Most of them declare themselves as outbound or balanced
What would we need?

• Hypergiants are the ‘biggest of the biggest’
  • Indication of **traffic volume**

• Hypergiants are global
  • Indication of **geographic reach**

• Hypergiants seem to be heavy on content
  • Indication of **traffic balance**
The problem

• Given the three features, classify whether an organisation is a hypergiant
  • Port capacity
  • Geographic reach
  • Traffic profile

• Sounds like something that could be solved with ML techniques

• No labels, so no supervised learning
Intuition

• Hypergiants are ‘the biggest of the biggest’

• They must be different from the crowd somehow
  • On some metric the at the very least

• Use unsupervised learning

• Use k-means (with $k=2$) to actually get labelled data
On the way...
On the way…
“Green group”: 15 organisations (0.2%), but 30%+ of port capacity
## Hypergiants!

<table>
<thead>
<tr>
<th>Organisation</th>
<th>ASN</th>
<th>Continents</th>
<th>Port. Cap.</th>
<th>Traffic Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>714</td>
<td>4</td>
<td>10.960 Tbps</td>
<td>Mostly Outbound</td>
</tr>
<tr>
<td>Amazon</td>
<td>16509</td>
<td>6</td>
<td>9.991 Tbps</td>
<td>Balanced</td>
</tr>
<tr>
<td>Facebook</td>
<td>32934</td>
<td>6</td>
<td>9.840 Tbps</td>
<td>Heavy Outbound</td>
</tr>
<tr>
<td>Google</td>
<td>15169</td>
<td>7</td>
<td>8.741 Tbps</td>
<td>Mostly Outbound</td>
</tr>
<tr>
<td>Akamai</td>
<td>20940</td>
<td>7</td>
<td>7.854 Tbps</td>
<td>Heavy Outbound</td>
</tr>
<tr>
<td>Yahoo</td>
<td>10310</td>
<td>6</td>
<td>5.310 Tbps</td>
<td>Mostly Outbound</td>
</tr>
<tr>
<td>Netflix</td>
<td>2906</td>
<td>7</td>
<td>5.170 Tbps</td>
<td>Mostly Outbound</td>
</tr>
<tr>
<td>Hurricane Electric</td>
<td>6939</td>
<td>7</td>
<td>5.037 Tbps</td>
<td>Balanced</td>
</tr>
<tr>
<td>OVH</td>
<td>16276</td>
<td>4</td>
<td>4.270 Tbps</td>
<td>Heavy Outbound</td>
</tr>
<tr>
<td>Limelight</td>
<td>22822</td>
<td>6</td>
<td>3.840 Tbps</td>
<td>Mostly Outbound</td>
</tr>
<tr>
<td>Microsoft</td>
<td>8075</td>
<td>6</td>
<td>3.680 Tbps</td>
<td>Mostly Outbound</td>
</tr>
<tr>
<td>Twitter</td>
<td>13414</td>
<td>6</td>
<td>3.401 Tbps</td>
<td>Heavy Outbound</td>
</tr>
<tr>
<td>Twitch</td>
<td>46489</td>
<td>5</td>
<td>3.340 Tbps</td>
<td>Heavy Outbound</td>
</tr>
<tr>
<td>Cloudflare</td>
<td>13335</td>
<td>7</td>
<td>3.320 Tbps</td>
<td>Mostly Outbound</td>
</tr>
<tr>
<td>Verizon Digital Media Services</td>
<td>15133</td>
<td>6</td>
<td>3.030 Tbps</td>
<td>Heavy Outbound</td>
</tr>
</tbody>
</table>
Discussion

• “Cloud” hypergiants?
  • Not everyone has a content delivery business, e.g., AWS, Hurricane Electric

• “Local” hypergiants?
  • Not everyone has a global audience, e.g., BBC

• Public vs. private peering

• Stability of results?
  • Organisations grow and change
Summary

• Characterised organisations in PeeringDB
  • Provisioned port capacity, geographic reach, traffic profile

• Observed organisations with different roles and characteristics

• Identified 15 hypergiants using unsupervised learning

• More details to (almost) everything in the paper 😊