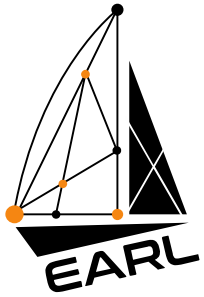
Looking for Hypergiants in PeeringDB

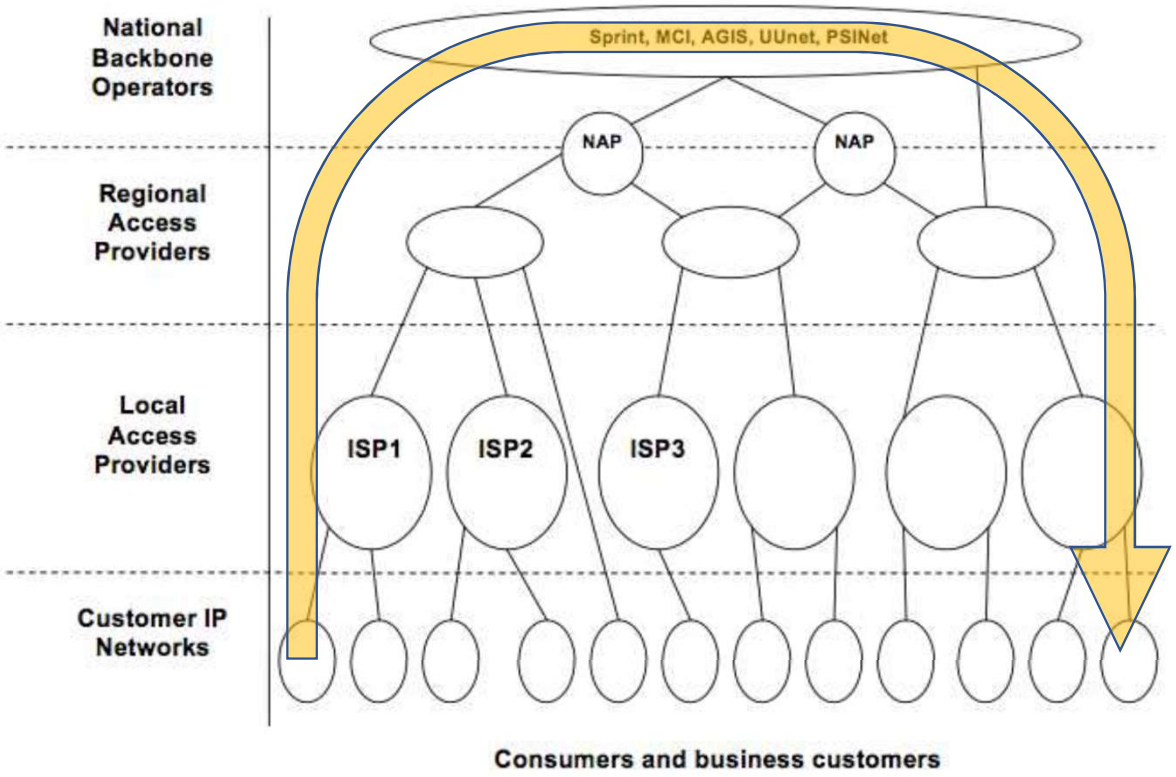
Timm Böttger, Felix Cuadrado and Steve Uhlig

steve.uhlig@qmul.ac.uk



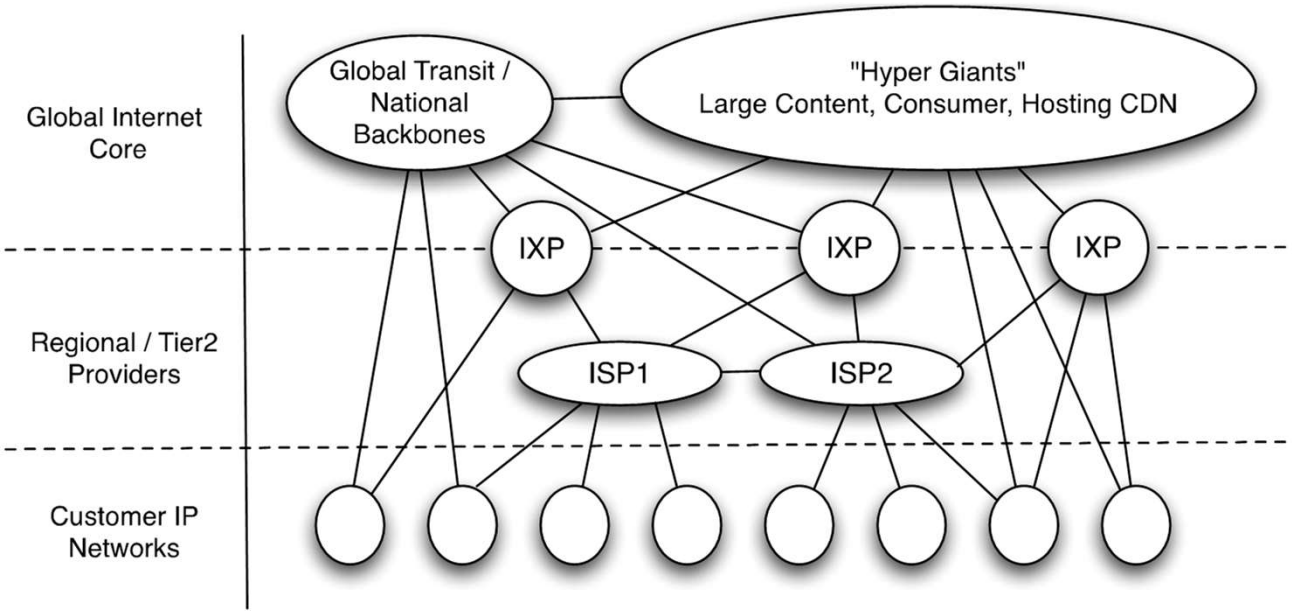
Appeared in T. Böttger, F. Cuadrado, S. Uhlig. *Looking for Hypergiants in PeeringDB.* ACM SIGCOMM Computer Communication Review, July 2018. (Selected as best of CCR for ACM SIGCOMM 2019)
With elements from
T. Böttger, F. Cuadrado, G. Tyson, I. Castro, S. Uhlig. Open connect everywhere: A glimpse at the internet ecosystem through the lens of the Netflix CDN. ACM SIGCOMM Computer Communication Review, January 2018.

Traditional Internet logical topology



Source: C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. ACM SIGCOMM 2010.

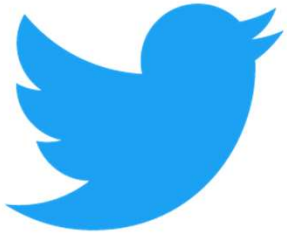
New Internet logical topology



Source: C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. ACM SIGCOMM 2010.

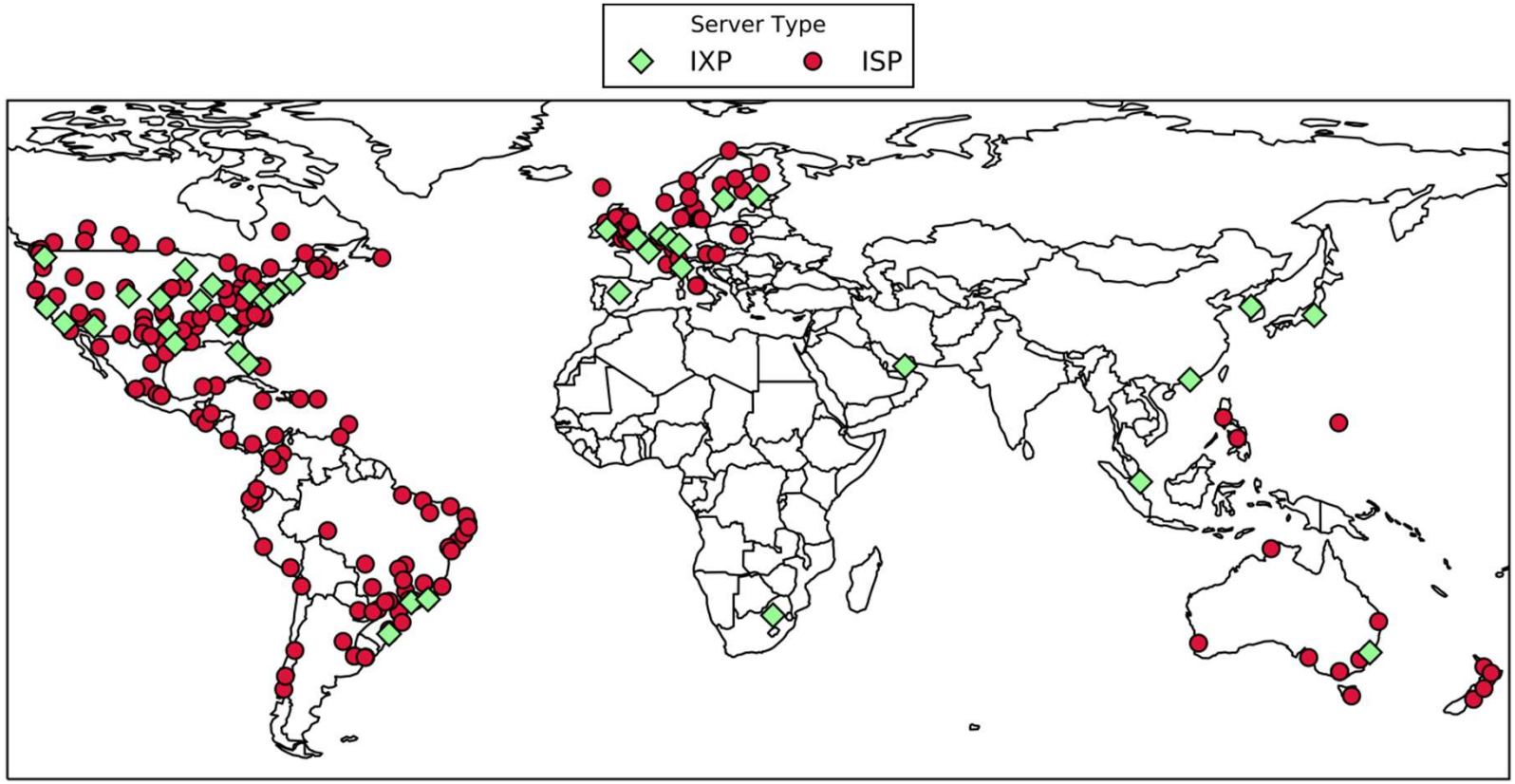
Queen Mary University of London

Large content players in the market



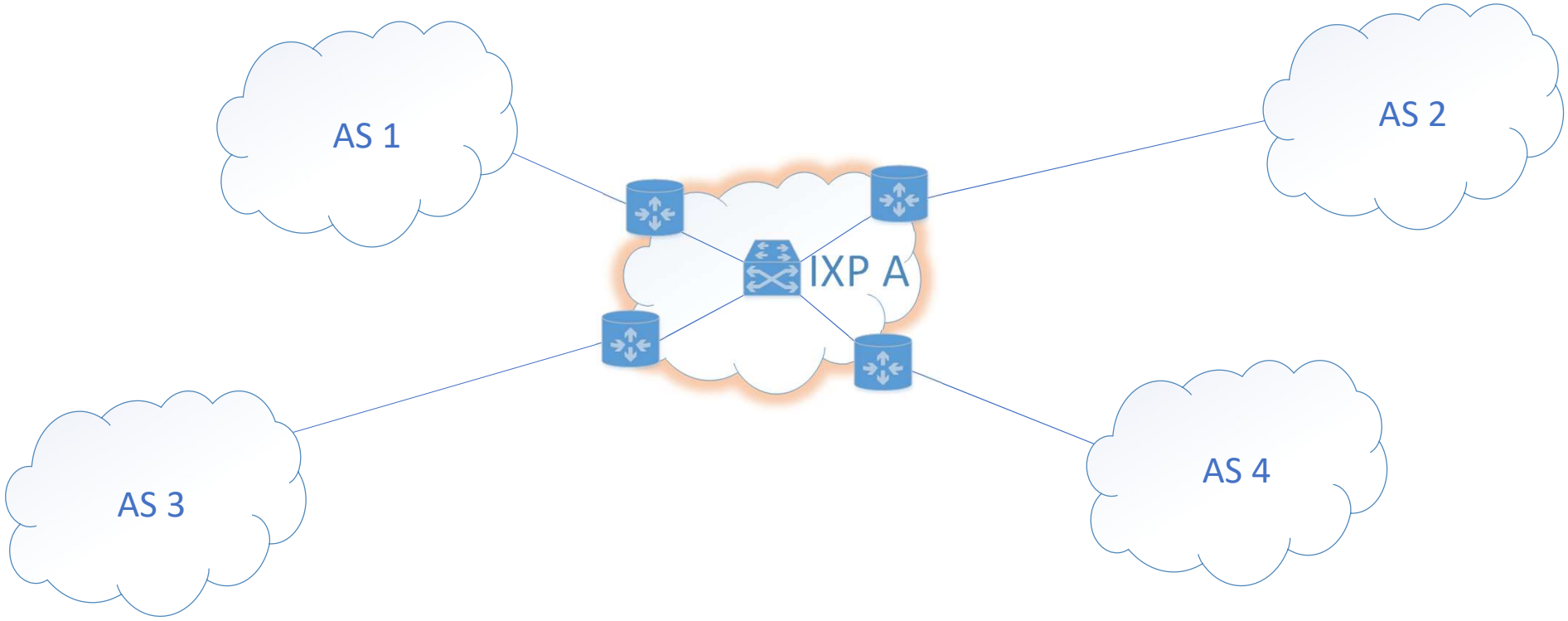
Hypergiants?

Netflix in 2016



T. Böttger, F. Cuadrado, G. Tyson, I. Castro, S. Uhlig. Open connect everywhere: A glimpse at the internet ecosystem through the lens of the Netflix CDN. ACM SIGCOMM Computer Communication Review, January 2018.

Internet eXchange Points



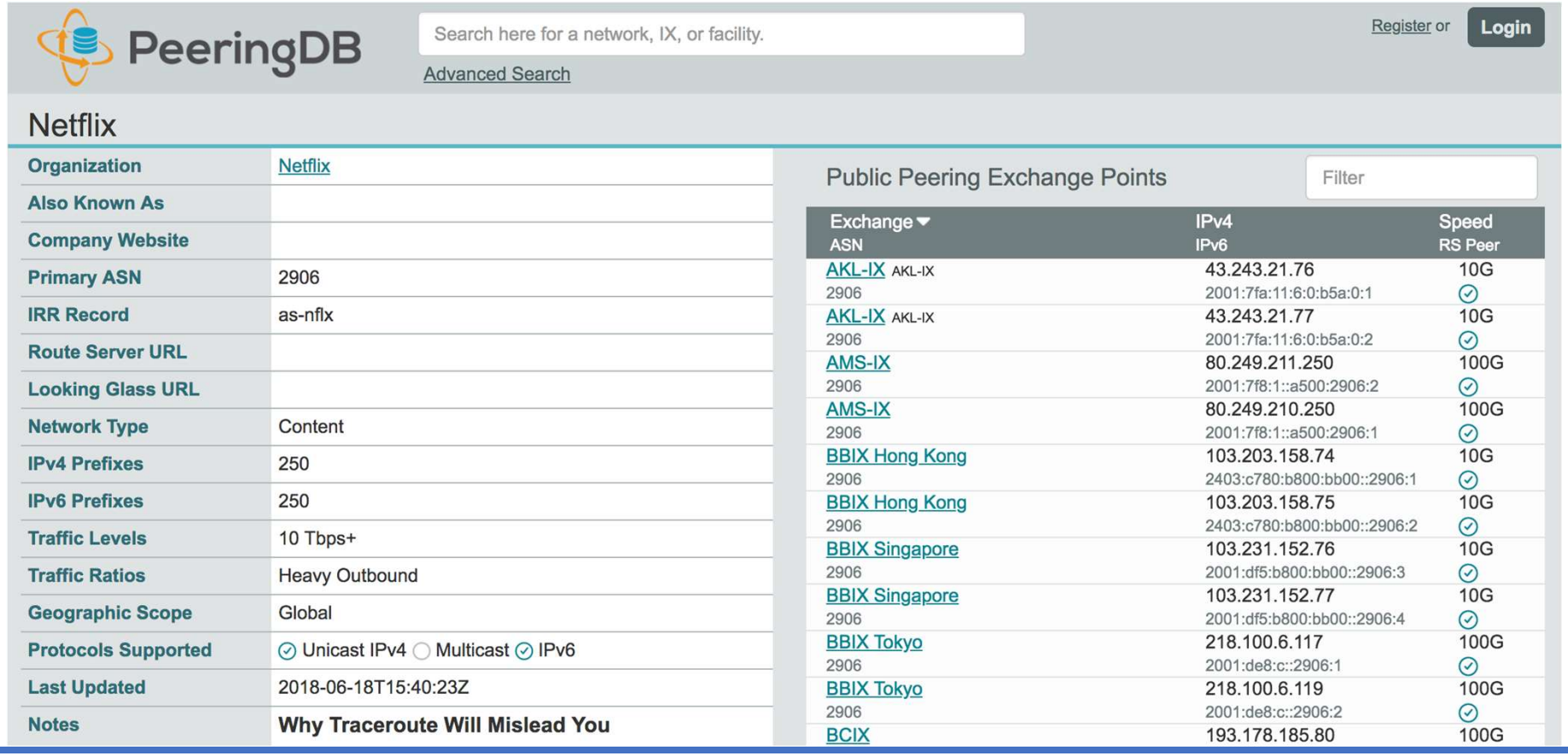
Queen Mary University of London

Examples of IXPs

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**

PeeringDB


PeeringDB
Search here for a network, IX, or facility.
Advanced Search
Register or
Login
Netflix
Organization
Netflix
Also Known As
Company Website
Primary ASN
2906
IRR Record
as-nflx
Route Server URL
Looking Glass URL
Network Type
Content
IPv4 Prefixes
250
IPv6 Prefixes
250
Traffic Levels
10 Tbps+
Traffic Ratios
Heavy Outbound
Geographic Scope
Global
Protocols Supported
Unicast IPv4 Multicast IPv6
Last Updated
2018-06-18T15:40:23Z
Notes
Why Traceroute Will Mislead You
Public Peering Exchange Points
Filter
Exchange
ASN
IPv4
IPv6
Speed
RS Peer
AKL-IX AKL-IX
2906
43.243.21.76
2001:7fa:11:6:0:b5a:0:1
10G
AKL-IX AKL-IX
2906
43.243.21.77
2001:7fa:11:6:0:b5a:0:2
10G
AMS-IX
2906
80.249.211.250
2001:7f8:1::a500:2906:2
100G
AMS-IX
2906
80.249.210.250
2001:7f8:1::a500:2906:1
100G
BBIX Hong Kong
2906
103.203.158.74
2403:c780:b800:bb00::2906:1
10G
BBIX Hong Kong
2906
103.203.158.75
2403:c780:b800:bb00::2906:2
10G
BBIX Singapore
2906
103.231.152.76
2001:df5:b800:bb00::2906:3
10G
BBIX Singapore
2906
103.231.152.77
2001:df5:b800:bb00::2906:4
10G
BBIX Tokyo
2906
218.100.6.117
2001:de8:c::2906:1
100G
BBIX Tokyo
2906
218.100.6.119
2001:de8:c::2906:2
100G
BCIX
193.178.185.80
100G

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

PeeringDB

Search here for a network, IX, or facility.

Advanced Search

Register or Login

Netflix

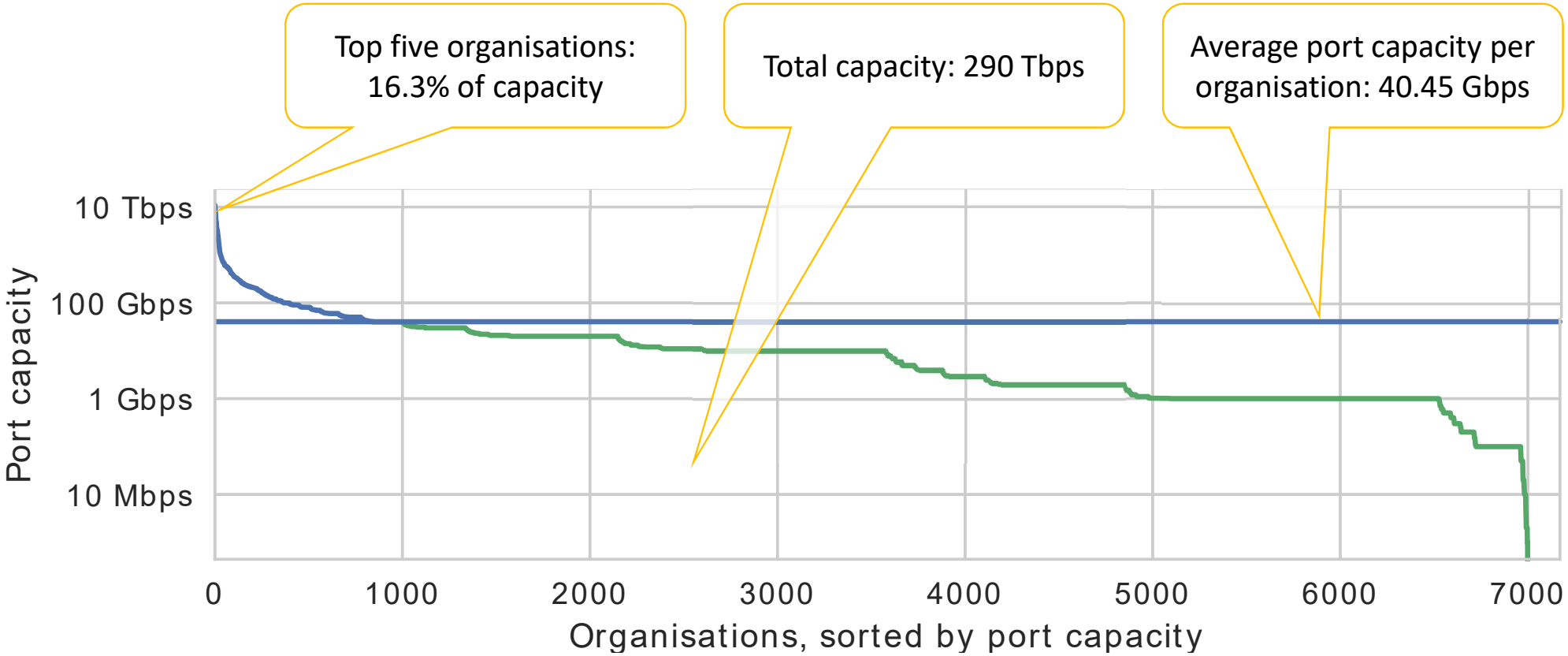
Organization	Netflix
Also Known As	
Company Website	
Primary ASN	2906
IRR Record	as-nflx
Route Server URL	
Looking Glass URL	
Network Type	Content
IPv4 Prefixes	250
IPv6 Prefixes	250
Traffic Levels	10 Tbps+
Traffic Ratios	Heavy Outbound
Geographic Scope	Global
Protocols Supported	✓ Unicast IPv4 ○ Multicast ✓ IPv6
Last Updated	2018-06-18T15:40:23Z
Notes	**Why Traceroute Will Mislead You**

Public Peering Exchange Points

Filter

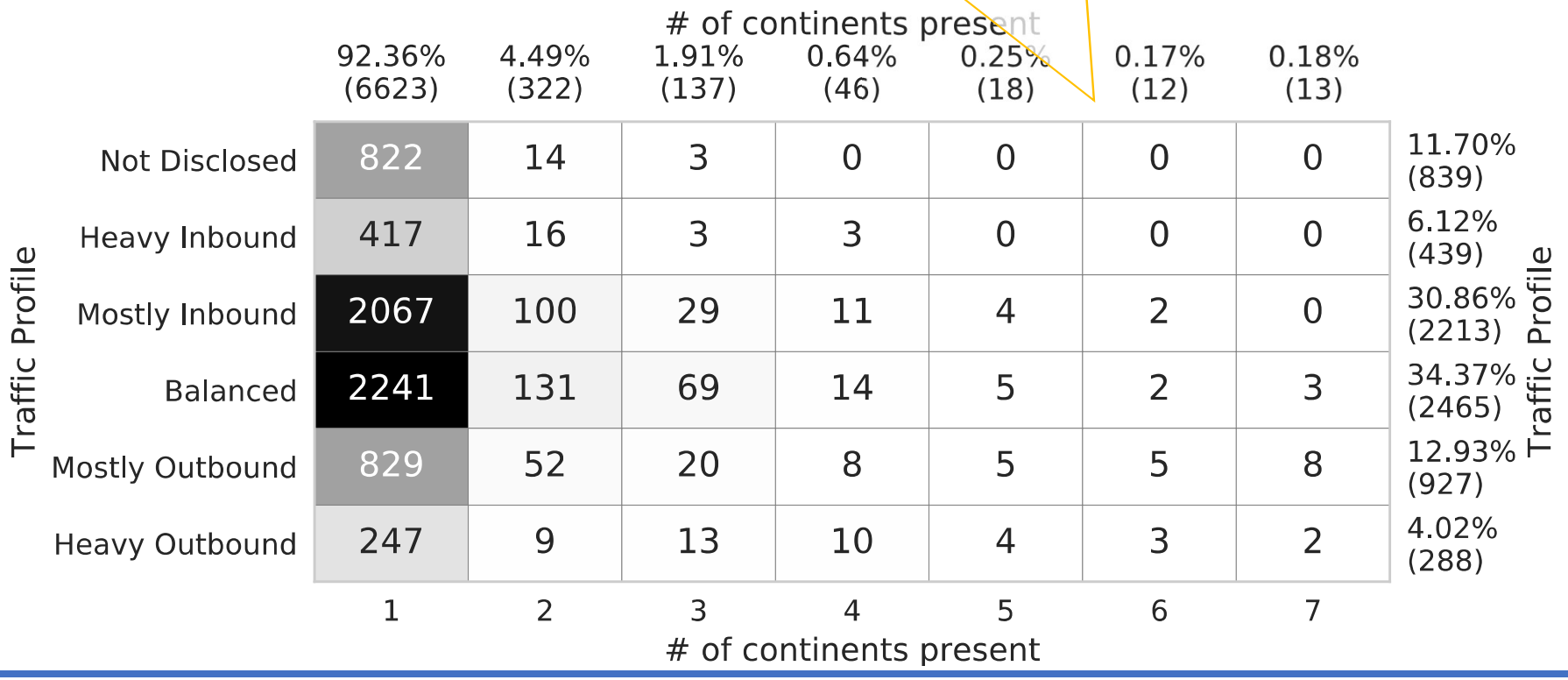
Exchange / ASN	IPv4 / IPv6	Speed / RS Peer
AKL-IX AKL-IX 2906	43.243.21.76 2001:7fa:11:6:0:b5a:0:1	10G ✓
AKL-IX AKL-IX 2906	43.243.21.77 2001:7fa:11:6:0:b5a:0:2	10G ✓
AMS-IX 2906	80.249.211.250 2001:7f8:1::a500:2906:2	100G ✓
AMS-IX 2906	80.249.210.250 2001:7f8:1::a500:2906:1	100G ✓
BBIX Hong Kong 2906	103.203.158.74 2403:c780:b800:bb00::2906:1	10G ✓
BBIX Hong Kong 2906	103.203.158.75 2403:c780:b800:bb00::2906:2	10G ✓
BBIX Singapore 2906	103.231.152.76 2001:df5:b800:bb00::2906:3	10G ✓
BBIX Singapore 2906	103.231.152.77 2001:df5:b800:bb00::2906:4	10G ✓
BBIX Tokyo 2906	218.100.6.117 2001:de8:c::2906:1	100G ✓
BBIX Tokyo 2906	218.100.6.119 2001:de8:c::2906:2	100G ✓
BCIX	193.178.185.80	100G

Port capacities

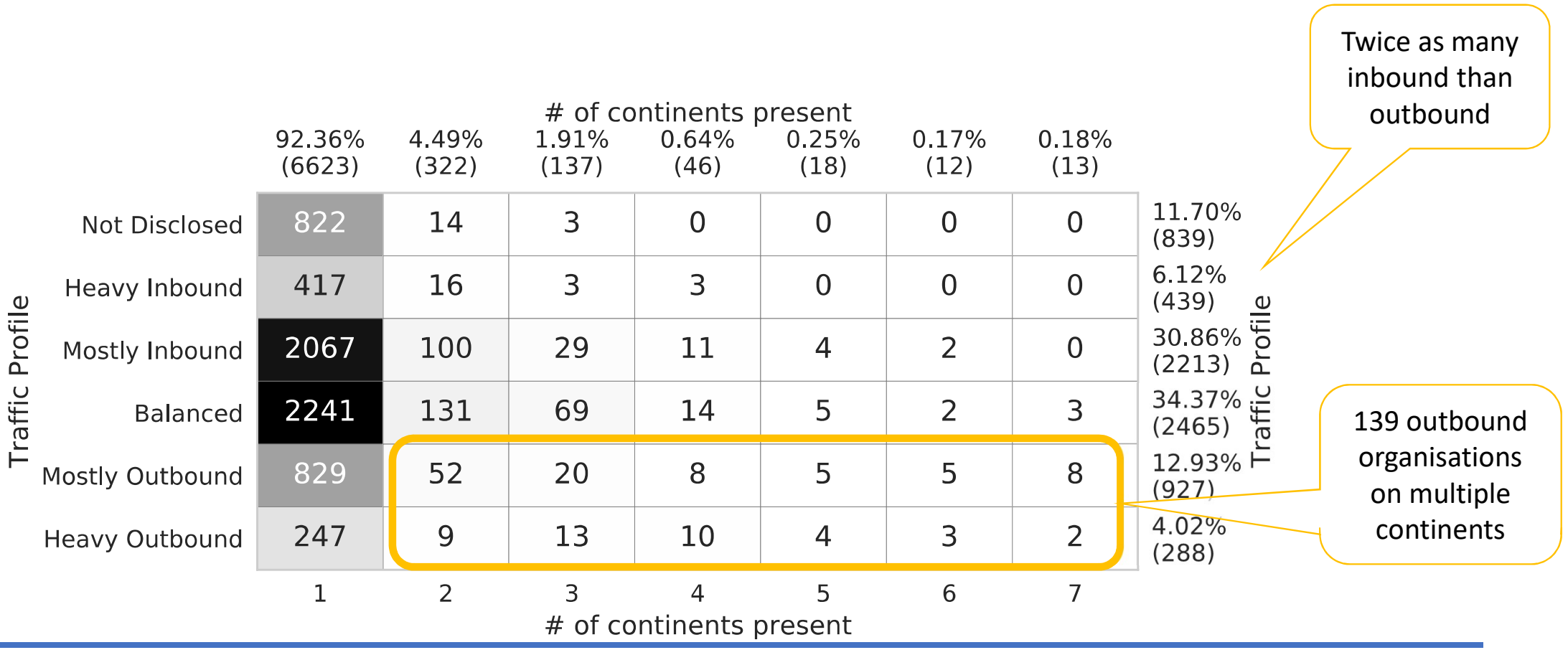


Geographic footprint

Organisations with most port capacity: Apple (4), Twitch (5), Amazon (6), Google (7)



Traffic profiles



Treemaps

- How to visualise hierarchically structured data?
  - Example: File sizes in a filesystem

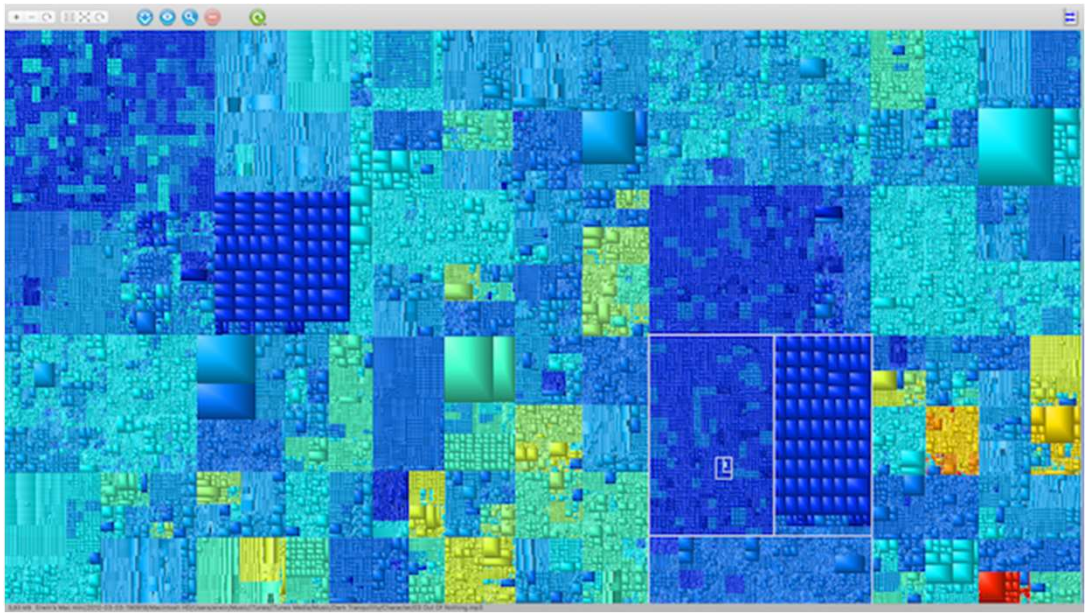
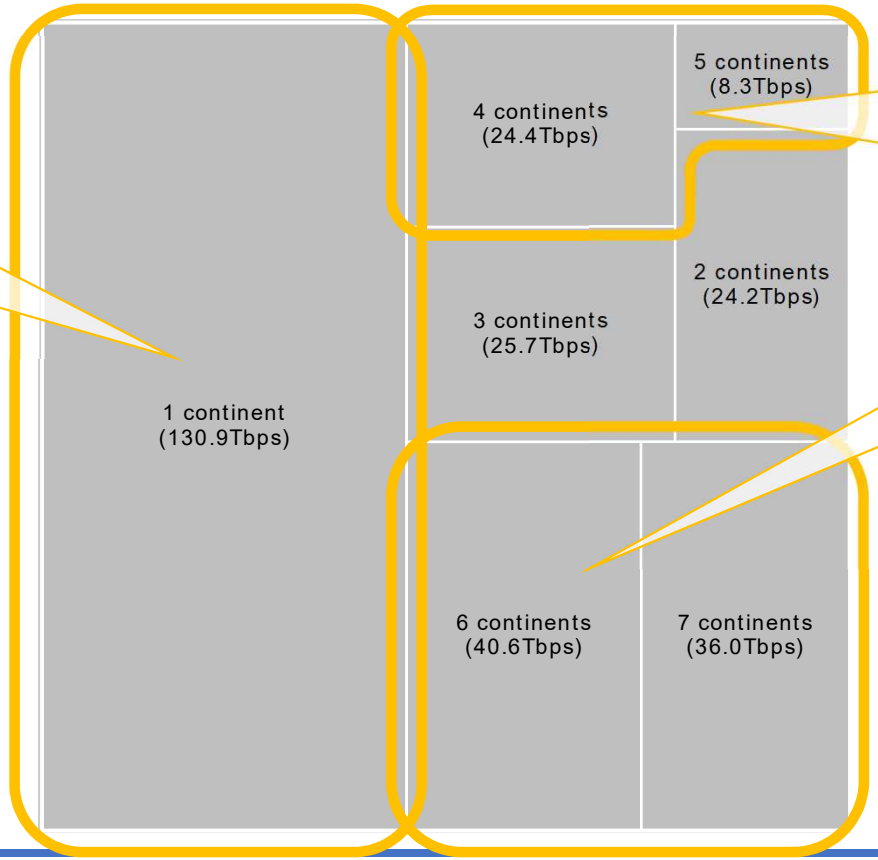


Image source: GrandPerspective http://grandperspectiv.sourceforge.net

The big picture

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

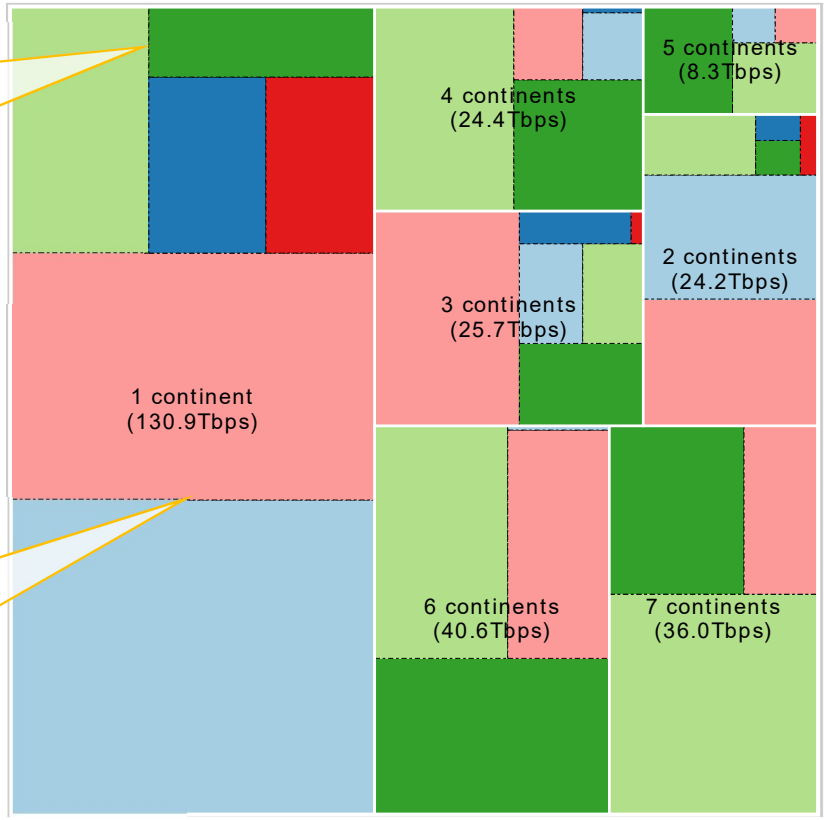


Organisations at four continents or more : 38% of port capacity, yet only 1% of all organisations

The big picture: traffic profile

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 captures significant fraction of port capacity
- Most of them declare themselves as outbound or balanced

So what's a hypergiant?

- Problem: Given the three features, classify whether an organisation is a hypergiant given its
  - Port capacity
  - Geographic reach
  - Traffic profile

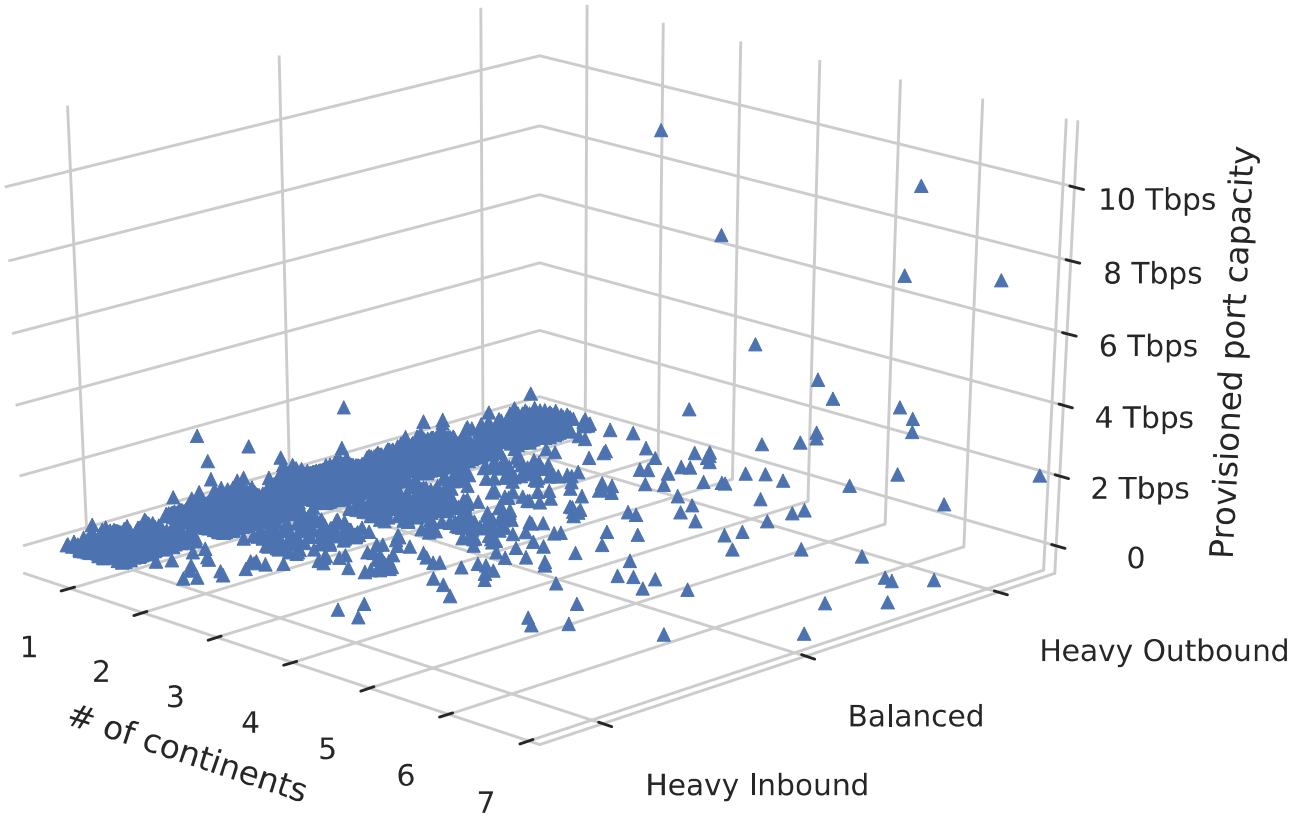
- Sounds like something that could be solved with "ML techniques"

- No labels, so no supervised learning

Intuition

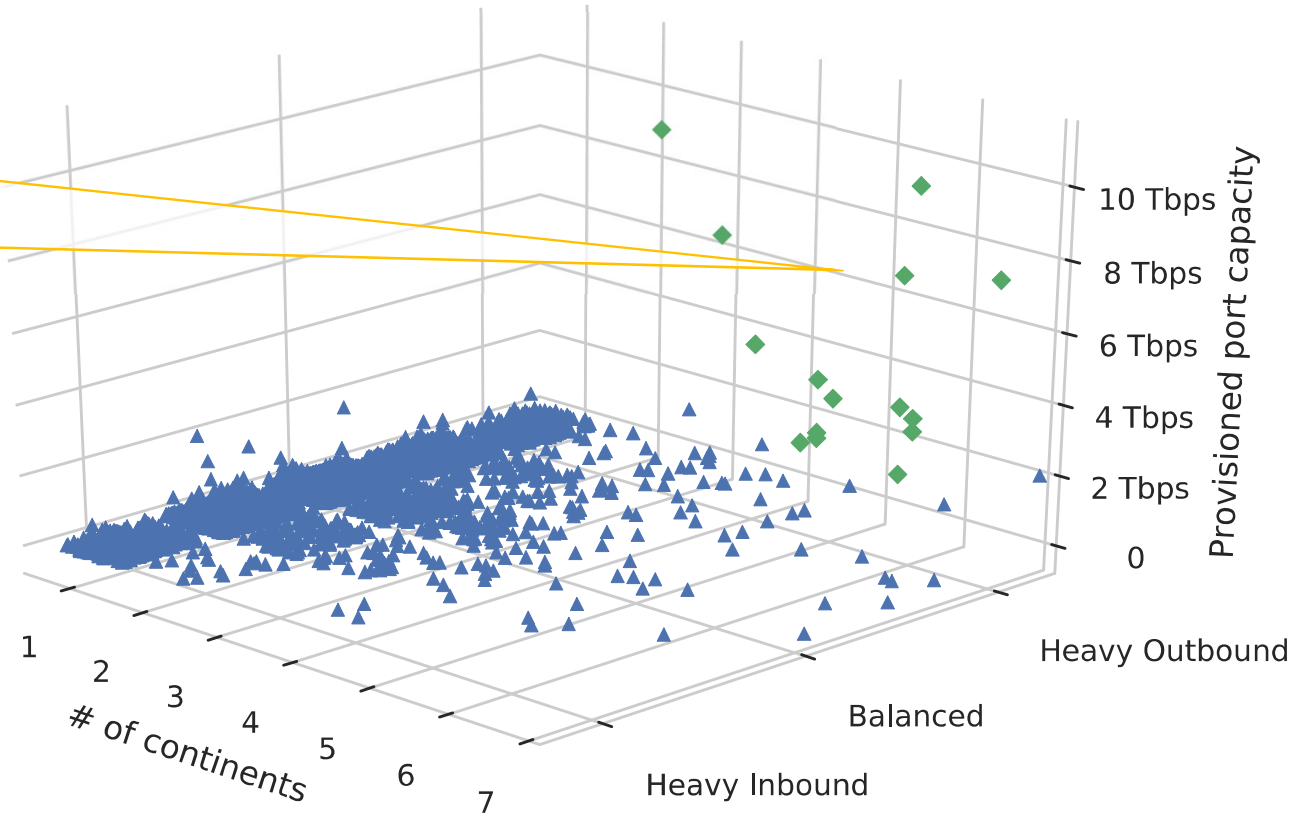
- Hypergiants are ‘the biggest of the biggest’
- They must be different from the crowd somehow
  - On some metric the very least
- Use unsupervised learning
- Use k-means (with k=2) to actually get labelled data

On the way…

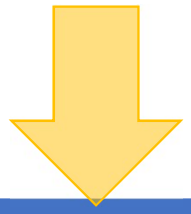
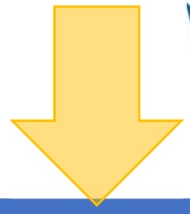

10 Tbps
8 Tbps
6 Tbps
4 Tbps
2 Tbps
0
Provisioned port capacity
1
2
3
4
5
6
7
# of continents
Heavy Inbound
Balanced
Heavy Outbound

Hypergiants!

"Green group": 15 organisations (0.2%), but 30%+ of port capacity



Hypergiants!

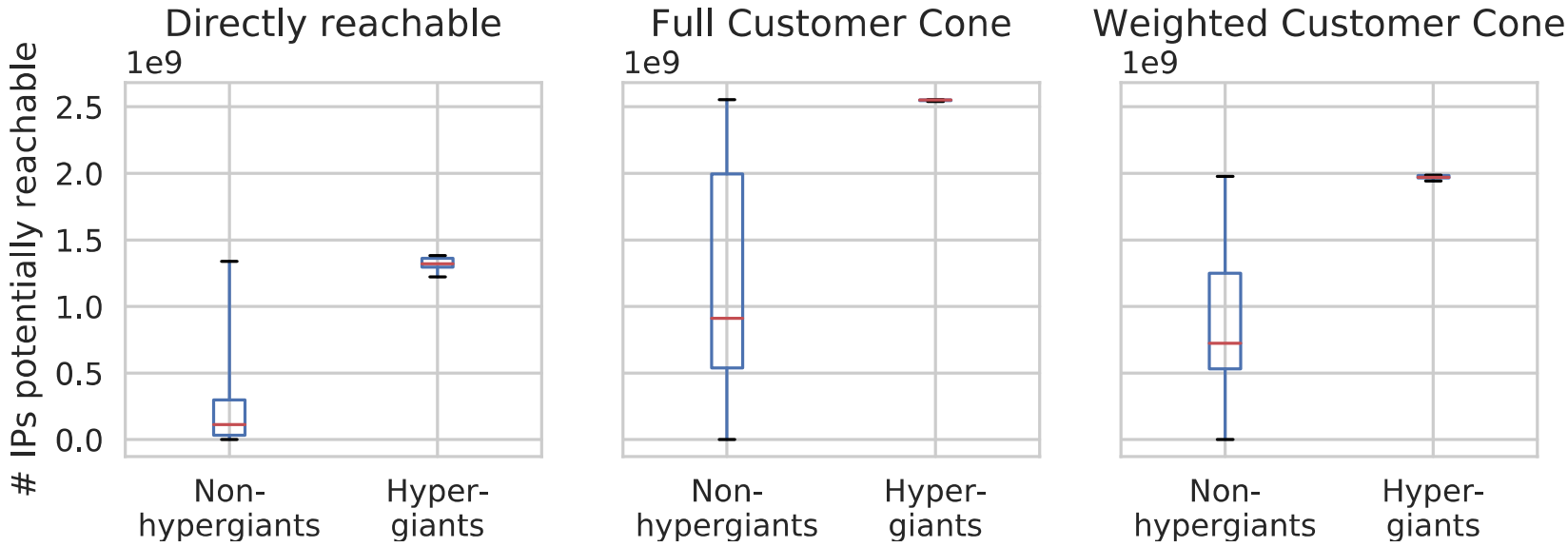




	Organisation	ASN	Continents	Port. Cap.	Traffic Profile
1	Apple	714	4	10.960 Tbps	Mostly Outbound
2	Amazon	16509	6	9.991 Tbps	Balanced
3	Facebook	32934	6	9.840 Tbps	Heavy Outbound
4	Google	15169	7	8.741 Tbps	Mostly Outbound
5	Akamai	20940	7	7.854 Tbps	Heavy Outbound
6	Yahoo	10310	6	5.310 Tbps	Mostly Outbound
7	Netflix	2906	7	5.170 Tbps	Mostly Outbound
8	Hurricane Electric	6939	7	5.037 Tbps	Balanced
9	OVH	16276	4	4.270 Tbps	Heavy Outbound
10	Limelight	22822	6	3.840 Tbps	Mostly Outbound
11	Microsoft	8075	6	3.680 Tbps	Mostly Outbound
12	Twitter	13414	6	3.401 Tbps	Heavy Outbound
13	Twitch	46489	5	3.340 Tbps	Heavy Outbound
14	Cloudflare	13335	7	3.320 Tbps	Mostly Outbound
15	Verizon Digital Media Services	15133	6	3.030 Tbps	Heavy Outbound

Reach of Hypergiants

- Reachable IP addresses of each organisation
- Consider all IXP presences for each organization
- With CAIDA Routeviews and customer cones:
  - Map IXP presences to member ASes (PeeringDB)
  - Then map member ASes to customer cones (CAIDA)
  - Then map customer cones to IP addresses (CAIDA)

Reach of Hypergiants



Hypergiants have top reachability

Reachability alone not enough to differentiate

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
- These hypergiants are among those organisations with biggest IP reachability