Looking for Hypergiants in PeeringDB

Timm Böttger  
Queen Mary University of London  
timm.boettger@qmul.ac.uk

Felix Cuadrado  
Queen Mary University of London  
felix.cuadrado@qmul.ac.uk

Steve Uhlig  
Queen Mary University of London  
steve.uhlig@qmul.ac.uk

ABSTRACT

Hypergiants, such as Google or Netflix, are important organisations in the Internet ecosystem, due to their sheer impact in terms of traffic volume exchanged. However, the research community still lacks a sufficiently crisp definition for them, beyond naming specific instances of them. In this paper we analyse PeeringDB data and derive a set of defining characteristics for hypergiants. To this end, we first characterise the organisations present in PeeringDB, allowing us to identify discriminating properties of the these organisations. We then show that these properties differentiate hypergiants well from other organisations. We conclude this paper by investigating how hypergiants exploit the IXP ecosystem to reach the global IPv4 space.

CCS Concepts

• Networks → Public Internet;  
Very long-range networks;  
Network performance analysis;

Keywords

Hypergiants, PeeringDB, Internet eXchange Points

1. INTRODUCTION

The Internet research community has commonly accepted that a significant fraction of today’s Internet traffic relates to so-called hypergiants like YouTube or Netflix [7]. While their importance is known for some time now, the research community still falls short of a definition of hypergiants. Most evidence on their behaviour and existence is anecdotal, or self-reported but lacking sufficient detail [5, 10, 11]. The current way the community understands and defines hypergiants is mainly by giving examples of them, which we believe is unsatisfactory. This is surprising as hypergiants not only are a massive source of traffic, but they also are one of the driving forces behind the observed flattening of the Internet hierarchy. The reason for the observed flattening indeed is their approach to peering, reaching customers via direct peering links instead of using and paying transit providers. The amount of traffic they carry is so significant that it has shifted traffic away from the traditional hierarchy of the Internet, and asked the research community to revisit its mental model of the Internet [7].

To obtain a better understanding of the role hypergiants play in the Internet, we mine PeeringDB to get a better understanding of the organisations taking part in public traffic exchange at IXPs. Afterwards we identify defining characteristics of hypergiants from the data available in PeeringDB.

In this paper we make the following contributions:

1. We characterise the organisations in PeeringDB, looking at their geographical scope, provisioned port capacity and potential reach.
2. We identify a set of organisations matching the concept of hypergiants, and propose a definition of hypergiants based on these properties.
3. We explore how hypergiants reach the global IPv4 space through the IXP ecosystem.

Code and Data sharing We will make the PeeringDB data snapshot and the program code used for the data analysis in this paper available to the research community, in the hope that this stimulates and facilitates further research.

2. RELATED WORK

In their seminal work, Labovitz et al. [7] were the first to coin the term hypergiant. They observed a shift over time of traffic being diverted away from large Tier-1 and Tier-2 backbone networks and instead being directly exchanged between networks without any intermediary, significantly revisiting the mental model of the Internet from the research community. Our work is motivated by their use of the word hypergiant, which is currently lacking a precise definition. In contrast to their work, we do not use traffic measurements but information within PeeringDB complemented by routing information to define hypergiants.

Previous works have used PeeringDB as an information source, extracting insights about the peering ecosystem and assessing its usability to better understand the Internet ecosystem. Lodhi et al. [8] made a first step in assessing the reliability and thus usability of PeeringDB for Internet research. They assessed the plausibility of PeeringDB data by comparing the information in PeeringDB against Local Internet Registries (LIRs) and BGP data. They found that while the data exhibits some biases, overall it appears to be reliable. They also made a first attempt at characterising the participating organisations. In contrast to our work, their focus is more on an overall assessment of PeeringDB than on analysing a specific type of organisation. Klöti et al. [6] compared the data in PeeringDB against data from other publicly available IXP data sets. They linked together the data sets available from PeeringDB, Euro-IX and PCH to assess their degree of complementarity and completeness. While they found biases in every data set, caused by its
4.1 The Peering Landscape

In this subsection we use the port capacity, geographic footprint and traffic profile dimensions to obtain a basic characterisation of the organisations participating in the public peering ecosystem.

**Port capacity:** Based on PeeringDB data, we extract for each organisation the IXPs it is present at along with the corresponding router port sizes. We then sum up those port sizes to obtain the aggregated provisioned port capacity. The total aggregate port capacity across the data amounts to 262.8 Tbps, with an average port capacity of 39.78 Gbps per organisation.

One would naturally assume that hypergiants should be amongst the organisations with the highest provisioned port capacity. Figure 1 shows the port capacity provisioned by each organisation present in PeeringDB. The distribution of provisioned port capacity is strongly non-uniform, with an average of 39.78 Gbps but a standard deviation of 312.13 Gbps. The figure exposes that a few organisations are responsible for a significant, way above-average port capacity (blue bars), while the overwhelming majority of them declares below-average capacity (green bars). The top five largest organisations represent 19.4% of the total port capacity, the top 68 covering half. In contrast with these massive organisations, the majority provisions significantly less port capacity.

**Geographic footprint:** We now turn to the geographic footprint, by looking at the number of continents\(^1\) where an organisation is present at IXPs. We hypothesise that hypergiants will aim to have global geographic presence, publicly exchanging data in IXPs across multiple continents. Figure 2 shows in its columns the distribution of continent presence in the PeeringDB data set. The largest share of organisations (6,095) are present on only one continent. On the other hand there are only 81 with presence across four or more continents.

**Traffic profile:** Organisations do not only differ in their geographic footprint and total port capacity, but also in their purpose and thus traffic profile. Some, such as content providers, are expected to have a predominantly outbound traffic profile, whereas ISPs connecting eyeballs to the Internet are expected to have an inbound traffic profile. PeeringDB defines five different profiles ranging from (Heavy) Inbound to (Heavy) Outbound, with Balanced in the middle. Organisations not wishing to expose their traffic

---

1PeeringDB recognises the following continents: Africa, Asia Pacific, Australia, Europe, Middle East, North America and South America. We adopt this non-textbook definition of a continent to maintain comparability to other works using PeeringDB data.

<table>
<thead>
<tr>
<th>Entity</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXPs</td>
<td>608</td>
</tr>
<tr>
<td>Organisations (total)</td>
<td>10,246</td>
</tr>
<tr>
<td>Organisations (at IXPs)</td>
<td>6,607</td>
</tr>
<tr>
<td>AS numbers (total)</td>
<td>10,350</td>
</tr>
<tr>
<td>AS numbers (at IXPs)</td>
<td>6,709</td>
</tr>
</tbody>
</table>

Table 1: Number of entities listed in PeeringDB.
profile have the option ‘Not Disclosed’ as well. Figure 2 shows in its rows the traffic profiles of the organisations in the data set. Besides a smaller fraction who hide their traffic profile, we see that the majority are inbound oriented or balanced. There are more than twice as many organisations with an inbound traffic profile than with an outbound profile.

4.2 The Whole Picture

After having discussed the three dimensions in isolation, we now put them together to obtain a more comprehensive picture of the organisations participating in the peering ecosystem as seen from PeeringDB. Figure 3 shows a tree-map combining the three dimensions: continent presence, traffic profile, and aggregate port capacity. In this tree-map, the area of each rectangle is proportional to the aggregated port size it represents. Organisations are first grouped by number of continents (one to seven) at which they maintain IXP presence, enclosed by a white border. The on-print shows the number of continents organisations are present at and aggregated port size of the organisations in each group.

First, we observe that organisations present at a single continent account for 47% of the overall port capacity. The remaining capacity is spread almost evenly across the other groups in terms of continent presence, with roughly 7-9% for each group. This stands in strong contrast to the number of organisations in each group (see Figure 2). While 92.2% of all organisations are present at a single continent, they are only responsible for 47% of the total provisioned port capacity. In contrast, the 1.2% of them with presence on four continents or more are responsible for 36.4% of provisioned port capacity. This implies that the many organisations with a local geographic scope tend to have little port capacity (hence little expected traffic) at IXPs. In contrast, there are a few with large geographic scope, combined with large port capacity (hence large expected traffic) at IXPs.

Second, the organisation composition in each group (in terms of continent presence) differs in terms of traffic profile. Within the single continent group, more than 70% of the port capacity belongs to balanced (29.7%) or inbound dominant (40.6%) organisations. Among the organisations in this group with an outbound traffic profile, we find content and hosting providers with a local audience, like for example BBC, Hetzner, Strato, VKontakte and Baidu.

Looking at the groups representing presence in multiple continents, we see a smaller contribution from inbound traffic profiles to the total port capacity. While inbound dominant organisations still have a notable share in the groups of two, three and four continents, they play no role in the groups of five, six or seven continents. In those groups, organisations with an outbound traffic profile are dominant. In the groups with five, six or seven continents, almost all organisations have a balanced or outbound profile, with the outbound profiles accounting for a significant share of each group. Balanced organisations with presence at four or more continents are those with a data-centric business model, that do not only deliver but also consume content, such as Dropbox, Amazon (AWS), Hurricane Electric and Microsoft.

In this subsection, we have seen how a relatively small group of global organisations gather a substantial amount

![Figure 2: Traffic profile and continent coverage for each organisation. Continent presence means an organisation is present at at least one IXP of this continent. Secondary axes show the distributions of organisations across traffic profiles and continent presence.](image1)

![Figure 3: Distribution of aggregated port sizes over traffic profiles and continent presence. An organisation is present on a continent if it is present at an IXP at this continent. The area of each rectangle is proportional to the aggregated port size it represents. Organisations are grouped by number of continents and then by traffic profile. The on-print depicts the number of continents organisations are present at and aggregated port size of the organisations in each group.](image2)
of port capacity. Moreover, they mostly declare an outbound or balanced traffic profile. Large content providers strive to deliver their content to a global audience of end customers. Based on what we observe in this section, content providers rely on a wide IXP presence to serve traffic to the eyeball organisations (inbound traffic profile) that operate smaller networks with a local footprint. Further, this strong concentration of port capacity strongly hints at hypergiants, which are quite likely to be among them.

4.3 Hypergiants of the Internet

Intuitively, content hypergiants are expected to be heavy on (outbound) traffic, with a large geographic reach to cater for a world-wide customer base. Cloud hypergiants will have similar characteristics, however their traffic profile will be more balanced. In general, we expect hypergiants to fall within the group of organisations with an outbound or balanced traffic profile and presence on many continents.

Mapping these concepts to PeeringDB features, we require hypergiants to have a traffic profile of either 'Balanced', 'Mostly Outbound' or 'Heavy Outbound', and to maintain presence on at least four different continents. We further require a hypergiant to be of a certain size in terms of traffic, i.e., to have an amount of provisioned port capacity above a given threshold. This capacity threshold will determine how many such hypergiants the resulting characterisation will consider. An organisation under this port capacity threshold will not be considered a hypergiant, but a (large) content delivery or cloud player.

For the following, we set this minimum threshold to a rather arbitrary (though empirically derived) value of 100Gbps of provisioned port capacity. Figure 4 shows the 100 organisations with the highest amounts of provisioned port capacity, the blue line and grey area depict our chosen threshold. The exact value of this threshold is not that important. The figure shows that the largest hypergiants, which are the ones arguably having the biggest impact on the Internet, are currently in the range of multiple Tbps, thus well above this threshold.

The continent threshold manages to separate those with huge port capacity and mostly outbound traffic profiles on the right side from those with less port capacity and mostly inbound profiles on the left. In the grey area, we observe two inbound ones as well. Without our requirement on the traffic profile, we would classify those as hypergiants as well. We manually validated those two outliers to be two customer facing ISPs, thus rightfully not fulfilling our hypergiant definition, and reinforcing the discriminating quality of the traffic profile.

Out of the 356 organisations from the data set above the 100 Gbps threshold, 46 (less than 1% of the total data set) fulfill all our criteria for hypergiants. Table 2 lists the fifteen largest organisations in terms of provisioned port capacity, while simultaneously fulfilling our criteria for being a hypergiant. In the whole data set, there only is one other organisation that falls within this range of provisioned port capacity, which for the sake of completeness is shown in the grey row in the table. All fifteen listed are typically identified as hypergiants by the research community. The traffic profile (outbound or balanced) can be used to differentiate content hypergiants from cloud hypergiants. We manually validated that the remaining 31 fulfilling our criteria also match our expectations for a hypergiant. The full list is available in

Figure 4: The top 100 organisations by provisioned port capacity. Blue line and gray area depict our thresholds for hypergiant classification. Note the log-scale of the y-axis.

the supplementary materials, to allow the reader to check.

To further ascertain the validity of our criteria, we now turn towards the remaining 310 organisations over the 100Gbps not classified as hypergiants. We manually inspected the list, and did not find any that should be classified as a hypergiant. They represent either ISPs with inbound traffic profiles, or content providers whose geographic reach limits their potential impact to certain regions only. They thus fail to fulfill the requirement of a hypergiant to be geographically large. This list is also available in the supplementary materials.

5. THE REACH OF HYPERGIANTS

So far, our focus has been on better understanding how the specific information present in PeeringDB would help us come up with a way to define hypergiants. We found out that the geographic presence was a strong aspect differentiating hypergiants from other networks. Combined with the traffic profile and port capacity, this led to a way to rank organisations on PeeringDB that exposes hypergiants.

Now, we slightly shift the focus to IXPs, asking how hypergiants rely on IXPs to build their interconnection footprint. More specifically, we are interested in finding an answer to the question: what it is that hypergiants are looking for with their IXP presence?

Quite naturally, a hypergiant should have a strong interest to reach eyeball IP address space, as they have built their business model around providing services to end users. While this might be less critical to cloud hypergiants that are more focused on hosting networked applications and services, this is definitely very important to content hypergiants like Netflix, who generate all revenue through end-users paying for their services.

The expected importance of IP address space leads us to define the potential reach of an organisation as the number of IP addresses organisations could potentially directly reach through their IXP presence, by peering with all organisations also present at IXPs. To compute this metric, we combine the IXP membership information from PeeringDB with Routeviews routing information from CAIDA [2]. For every organisation, we extract all IXPs it is present at and further extract all ASes which also are present there. We
explaining its unexpected presence whilst not being a hy-
display peering routes from exchanges around the world,
position seven. PCH uses this AS to collect, archive, and
able IP space are hypergiants according to our definition.
the address space. Nine of the top 10
lay in Figure 5) that they can indeed reach the majority of
hypergiants previously identified, we observe (see green over-
96.19% of the address space.

The histogram in Figure 5 shows the fraction of all IPs reachable by a specific organisation through all IXPs of PeeringDB. Note the log-scale of the y-axis.

then use the routing information to map ASes to IPv4 pre-
fixes, and calculate the number of unique IPs covered by
those prefixes.

The histogram in Figure 5 shows the distribution of potential reach among all organisations, normalised by the total address space observed from all IXPs members (not the complete IPv4 address space). The x-axis shows the fraction of IP addresses a specific organisation can potentially reach through all its IXP presence. The y-axis shows the number of organisations in each bin. We again observe that organisations span a continuum: The smallest reach less than 1% of the space, whereas the biggest one, Cloudflare, reaches 96.19% of the address space.

When we focus on reachable IP addresses for the top 15 hypergiants previously identified, we observe (see green overlay in Figure 5) that they can indeed reach the majority of the address space. Nine of the top 10 organisations in reachable IP space are hypergiants according to our definition. The only exception is Packet Clearing House (AS 3856) at position seven. PCH uses this AS to collect, archive, and display peering routes from exchanges around the world, explaining its unexpected presence whilst not being a hypergiant. The fact that hypergiants have a nearly complete

Table 2: Fifteen largest hypergiants by port capacity world-wide. These are the fifteen organisations with an Outbound or Balanced traffic profile, sorted by port capacity. The table also lists in Gray organisations with a different traffic profile but reporting comparable port capacity.

<table>
<thead>
<tr>
<th>Organisation name</th>
<th>ASN</th>
<th>Continents</th>
<th>Port Cap. (rank)</th>
<th>Traffic Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Apple Inc</td>
<td>714</td>
<td>4</td>
<td>10.080 Tbps (1)</td>
<td>Mostly Outb.</td>
</tr>
<tr>
<td>-- Multiplay Sp. z o.o. SK</td>
<td>196729</td>
<td>1</td>
<td>10.000 Tbps (2)</td>
<td>Not Disclosed</td>
</tr>
<tr>
<td>2 Amazon.com</td>
<td>16509</td>
<td>5</td>
<td>8.521 Tbps (3)</td>
<td>Balanced</td>
</tr>
<tr>
<td>3 Facebook</td>
<td>32934</td>
<td>6</td>
<td>8.150 Tbps (4)</td>
<td>Heavy Outb.</td>
</tr>
<tr>
<td>4 Google Inc.</td>
<td>15169</td>
<td>7</td>
<td>7.651 Tbps (5)</td>
<td>Mostly Outb.</td>
</tr>
<tr>
<td>5 Akamai Technologies</td>
<td>20940</td>
<td>7</td>
<td>7.018 Tbps (6)</td>
<td>Heavy Outb.</td>
</tr>
<tr>
<td>6 Yahoo!</td>
<td>10310</td>
<td>6</td>
<td>5.200 Tbps (7)</td>
<td>Mostly Outb.</td>
</tr>
<tr>
<td>7 Netflix</td>
<td>2906</td>
<td>7</td>
<td>5.080 Tbps (8)</td>
<td>Mostly Outb.</td>
</tr>
<tr>
<td>8 Hurricane Electric</td>
<td>6939</td>
<td>6</td>
<td>4.517 Tbps (9)</td>
<td>Balanced</td>
</tr>
<tr>
<td>9 OVH</td>
<td>16276</td>
<td>4</td>
<td>4.180 Tbps (10)</td>
<td>Heavy Outb.</td>
</tr>
<tr>
<td>10 Twitter, Inc.</td>
<td>13411</td>
<td>6</td>
<td>3.471 Tbps (11)</td>
<td>Heavy Outb.</td>
</tr>
<tr>
<td>11 Twitch</td>
<td>46489</td>
<td>5</td>
<td>2.860 Tbps (12)</td>
<td>Heavy Outb.</td>
</tr>
<tr>
<td>12 Cloudflare</td>
<td>13335</td>
<td>7</td>
<td>2.821 Tbps (13)</td>
<td>Mostly Outb.</td>
</tr>
<tr>
<td>13 Microsoft</td>
<td>8075</td>
<td>6</td>
<td>2.720 Tbps (14)</td>
<td>Balanced</td>
</tr>
<tr>
<td>14 Limelight Networks Global</td>
<td>22822</td>
<td>5</td>
<td>2.580 Tbps (15)</td>
<td>Mostly Outb.</td>
</tr>
<tr>
<td>15 Verizon Digital Media Services a</td>
<td>15133</td>
<td>5</td>
<td>2.490 Tbps (16)</td>
<td>Heavy Outb.</td>
</tr>
</tbody>
</table>

a formerly known as EdgeCast.

Figure 5: Histogram of the potential reach of organisations through their IXP presence. The x-axis shows the fraction of all IPs reachable by a specific organisation through all IXPs of PeeringDB. Note the log-scale of the y-axis.

6. DISCUSSION

PeeringDB and the Internet ecosystem: While previous work has already extensively assessed the visibility that PeeringDB provides into the Internet ecosystem [6, 8], one important reason why PeeringDB is so popular is that it allows networks to advertise their presence and willingness to peer at specific locations, especially IXPs. In today’s Internet ecosystem so centred on popular content and applica-
7. SUMMARY

In this paper we combined PeeringDB and Routeviews BGP data to obtain a better understanding of today’s hypergiants. Starting with a characterisation of the organisations taking part in public traffic exchange, we identified the fifteen largest hypergiants and proposed a set of defining characteristics for hypergiants. We showed that hypergiants can be identified using the geographic reach, provisioned port capacity and traffic profile of an organisation. We then explored how those hypergiants make use of IXPs to reach their global customer base. We provided evidence that hypergiants choose IXPs to maximise their reach, showcasing the utility of IXPs for today’s hypergiants. All these steps identified and discussed important characteristics of hypergiants, a set of organisations which has a significant impact on the Internet, due to the massive amount of traffic they are responsible for.

8. REFERENCES


Figure 6: Breakdown of (additionally) reachable IPs through each IXP for each of the fifteen hypergiants. IXPs are depicted by different colours in each bar. For each IXP the bar only shows the number of additional IPs reachable through this IXP on top of the already previously considered IXPs, not the total number of IPs reachable. IXPs are sorted by decreasing reachable IP contribution.

In this paper we focused on coming up with a definition of hypergiants as coined by Labovitz et al. [7]. Because we relied on global reachability as seen through PeeringDB as a way to detect these hypergiants, we limited our study to the largest of them. However, there is a variety of organisations that operate on a less global scale than hypergiants, which move a significant amount of traffic, without global footprint due to the nature of their business, e.g., BBC. Further work into the diversity of hypergiant-like organisations is needed if we are to truly understand the Internet ecosystem, given how much it is driven by content-heavy players.

**Public vs. private:** Despite the unique and rather trustworthy information provided by PeeringDB, it misses an important part of the Internet network interconnection ecosystem, namely private peerings. Some large hypergiants, such as Facebook, rely heavily on private interconnection to deliver their traffic [10, 11]. Fortunately, despite not showing the private part of the network interconnection ecosystem, PeeringDB appears to provide sufficient information to still see the largest hypergiants. However, PeeringDB provides a view that (largely) underestimates the network interconnection ecosystem of the Internet. This bias is similar to the one of the AS-level topology, for which publicly available BGP routing data misses a large fraction of the AS-level connectivity [9], especially due to the rich worldwide IXP ecosystem [3, 4].